Matched filters for source detection in the Poissonian noise regime

I. M. Stewart

Jodrell Bank Observatory, University of Manchester, Macclesfield SK11 9DL, United Kingdom

Received January 0, 0000; accepted January 0, 0000

Abstract. A procedure is described for estimating an optimum kernel for the detection by convolution of signals among Poissonian noise. The technique is applied to the detection of x-ray point sources in XMM-Newton data, and is shown to yield an improvement in detection sensitivity of up to 60% over the sliding-box method used in the creation of the 1XMM catalog.

Key words. Methods: data analysis – Techniques: image processing – X-rays: general

1. Introduction

Over the past few years, CCD cameras on satellite observatories such as ROSAT, ASCA, Chandra and most recently XMM-Newton have generated high-resolution digital images of the x-ray sky which are characterized by relatively faint background (for example, background fluxes of less than 1 count per CCD pixel are often seen in typical-duration XMM-Newton exposures). The number of events per pixel follows a Poisson probability distribution which, at such low flux levels, deviates markedly from its Gaussian bright-end limit.

It seems likely that a significant part of the (non-instrumental) x-ray background is comprised of point sources too faint to be distinguished from one another by present techniques (Mushotzky et al 2000, Hasinger et al 2001). The desire to characterise this population of sources is one reason for attempting to push the sensitivity of x-ray point source detection to the lowest limits allowed by the data.

Several source-detection procedures have been applied to x-ray images, eg: sliding-box (DePonte & Primini 1993; Dobrzycki et al 2000, see also task documentation for the exssas task detect and the SAS task eboxdetect); wavelet (Damiani et al 1997; Starck & Pierre 1998; Pierre et al 2004, see also task documentation for the CIAO task wavdetect and the SAS task ewavelet); maximum-likelihood PSF fitting (Crudace et al 1997; Boese and Doebereiner 2001); and Voronoi tessellation (Ebeling & Wiedenmann 1993). There are also occasional references to use of a ‘matched filter’ technique, but in the x-ray sphere at least this appears to consist just of convolution by the Point Spread Function or PSF (Vikhlinin et al 1995; Alexander et al 2003). A similar technique has also been applied to ASCA data (Ueda et al 1999). As is shown in the present paper, the PSF approaches a true matched filter in the limit of white Gaussian noise but is sub-optimal at low levels of background.

A useful review of source-detection procedures as applied to x-ray images can be found in Valtchanov et al (2001).

Many of these techniques include a step in which the raw image is subjected to a convolution, often as a first step in preparing a detection-likelihood map. Even PSF fitting, in simplified form, can be shown to be equivalent to a convolution (Stetson 1987). Voronoi tessellation seems to be the only technique which cannot readily be understood in this form.

Extraction of source positions and detection likelihoods from the convolved image is in general not simple, because the characteristic width of the PSF of the mirror system is usually larger than the CCD pixel size. In a (raw) image where the PSF is resolved, neighbouring pixels are not statistically independent - detection of some source flux in one pixel makes it more likely to detect it in neighbouring pixels. However, if the null hypothesis (ie, that there are no sources in the field) is assumed, detected counts are just due to background. But the background (by definition) must be slowly varying over the scale of the PSF, otherwise it would be impossible to separate it from the sources; and in the limit of smooth background the event counts in the raw image are statistically independent.

On the other hand, as has already been pointed out, a significant fraction of the cosmological x-ray background (perhaps approaching 100% at energies above 1 keV) actually consists of sources. This seems in direct contradiction to the null hypothesis. It is shown however in appendix A that the bulk of these background sources must be very faint, and that for x-ray telescopes of presently achievable effective areas, one can model the net contribution of this source population by a relatively sparse population of brighter sources superimposed upon
2. Linear signal detection

2.1. General

Suppose we have a parent function \( C \) which is a function of some independent variables \( x \) (e.g., time, position, energy) and which comprises a background \( B \) and signal \( S \) such that

\[
C(x) = B(x) + \alpha S(x - x_0)
\]

(1)

where

\[
\int_{-\infty}^{\infty} dx \ S(x) = 1,
\]

\( \alpha \) is the signal amplitude and \( x_0 \) is a reference point which serves to locate the signal in \( x \) space. Let each experimental measurement of \( C \) return a random variable \( c \) which is distributed according to a probability distribution with an expectation value \( \langle c \rangle = C \). A common problem in signal detection occurs when one one knows (or can estimate) \( B(x) \) and \( S(x) \) \textit{a priori} and one wishes to calculate the most likely values of \( \alpha \) and \( x_0 \) from a set of samples \( c_i \) of \( C \) at different values of \( x \).\(^1\)

In the case of point source detection, \( x \) is a two-coordinate vector which specifies position in the focal plane of a camera, \( S \) is the point spread function (PSF) of the camera, and the samples \( c_{i,j} \) are measurements of flux made on a pixel grid in the focal plane.

As outlined in the introduction, to assess whether there is any signal present in a given channel \( i \) one calculates the probability of the parent background \( B_i \) alone generating either the observed count \( c_i \) or any value higher than this. This is called the probability of the null hypothesis \( (P_{null}) \). A signal is judged to be ‘detected’ in channel \( i \) if the null probability of the measured \( c_i \) falls below a previously selected cutoff value.

It is very often the case in practice that \( S \) extends over several channels. In this case it is usually possible to improve the signal-to-noise ratio in at least one of the channels spanned by the signal, hence the detection sensitivity, by performing a weighted sum of the counts measured over several adjacent channels. In the XMM-Newton case which we are going to consider, \( x \) extends along spatial dimensions \( x \) and \( y \) and energy dimension \( E \). The weighted sum is to be computed for each spatial pixel \((i, j)\), the eventual aim being to produce a map of the null-hypothesis likelihood at each \((i, j)\). For computational purposes the spatial sum is most conveniently expressed as a convolution; the actual expression employed to calculate the weighted sum for each pixel is therefore

\[
c'_{i,j} = \sum_{p=-M}^{M} \sum_{q=-M}^{M} \sum_{k=1}^{N_{pix}} w_{p,q,k} c_{i-p,q,k}.
\]

(2)

For present purposes however it is unnecessary to retain such complication. As far as the statistical analysis goes, equation 2 at any given spatial pixel \((i, j)\) is simply a weighted sum of random variates:

\[
c' = \sum_{i=1}^{N} w_i c_i.
\]

(3)

It is also convenient for the present to assume that the signal is bracketed between \( i = 1 \) and \( i = N \), or, in other words, that the sampled parent function \( C_i \) is given by

\[
C_i = B_i + \alpha S_i.
\]

This is equivalent to assuming that a source is centred on pixel \((i, j)\) of equation 2.

The standard deviation \( \sigma' \) of \( c' \) as given in equation 2 is estimated from the usual error propagation relations to be

\[
\sigma'^2 = \sum_{i=1}^{N} w_i^2 \sigma_i^2
\]

(4)

where \( \sigma_i \) is the standard deviation of \( c_i \).

Clearly, whatever weight scheme is adopted, the stronger a signal is (i.e., the larger the signal amplitude \( \alpha \)), the higher the probability that it will be detected (the smaller the value of \( P_{null} \)). Ideally we would like to be able to calculate some cutoff value of \( \alpha \) above which the signal definitely would be detected, and below which it definitely would not. It is however not possible to do this, for the reason that the any non-trivial function \( c' \) of the detected counts must itself be a random variable. A

\(^1\) For simplicity I’ll assume in this section that there is only one signal \( S \) to be found, although, since convolution filtering is a linear process, in principle there is no difficulty in detecting many superposed signals provided they have sufficiently different values of \( x_0 \).
given value of $\alpha$ will not always give rise to the same value of $c'$, and any value of $\alpha$ will produce any chosen value of $c'$ within a sufficiently large ensemble. To get around the problem and so to permit the comparison of different choices of weights it is convenient to define a ‘counts amplitude’ $\beta$ such that

$$\beta = \frac{c'-B'}{S'},$$

(5)

where the significance of the primes here is that, for any function $f$,

$$f' = \sum_{i=1}^{N} w_i f_i.$$

It is not hard to show that $(\beta) = \alpha$.

Since any given null-hypothesis probability is associated with a definite value of the weighted sum of counts $c'$, it is also therefore unambiguously associated with a definite value of $\beta$. This allows us to define the detection sensitivity $\beta_{det}$ of any convolution as that value of $\beta$ which is associated with the chosen detection-cutoff value of $P_{null}$.

Clearly it is also the case that, for a given $B_i$ and $S_i$, there will be a set of weights (not necessarily unique) which gives maximum sensitivity, i.e. the smallest possible value of $\beta_{det}$.

2.2. The Gaussian case

For purposes of comparison I’ll briefly reprise the well-studied case where the probability distribution of $c_i$ for each $i$ is Gaussian, with constant $\sigma_i = \sigma$. In this case any weighted sum returns $c'$ values which also have a Gaussian distribution, with

$$\sigma'^2 = \sigma^2 \sum_{i=1}^{N} w_i^2.$$

One can calculate the signal-to-noise ratio $snr$ as follows:

$$snr = \frac{c' - B'}{\sigma'}.$$

The null-hypothesis probability is then given by

$$P_{null}(snr) = 0.5 \left(1 - \text{erf} \left[ snr / \sqrt{2} \right] \right).$$

By differentiating $snr$ with respect to the weights $w_i$ and equating each of the $N$ resulting derivatives to zero one arrives at the well-known result that the optimum set of weights is proportional to the signal itself: i.e

$$w_i = k S_i \forall i,$$

where $k$ is some non-zero constant.

2.3. Weighted-Poisson case with a single weight

In the case where the probability distribution of the observed values $c_i$ is Poissonian, the optimum weights are not so easy to come by, and in general will depend in a non-trivial fashion on the level of background $B_i$.

Consider first the simplest case, in which there is only one weight ($N = 1$ in equation 5). The solution in this case is trivial, but provides a useful mathematical template for the more difficult case in which $N > 1$. Let $c$ be a random (necessarily integer) variable which has a Poisson probability distribution. A second variable $c'$ which is just $c$ multiplied by a single weight $w$ retains the Poisson probability distribution

$$p(c') = \frac{v^c e^{-v}}{c!},$$

(6)

but with $v$ now given by

$$v = \langle c' \rangle = \frac{\langle c' \rangle}{w}.$$

In the null hypothesis, $\langle c \rangle = B$ and thus $\langle c' \rangle = wB = B'$. If we extrapolate from the unscaled Poissonian case it is also clear that the null-hypothesis probability $P_{null}$ is given by

$$P_{null}(c') = 1 - Q \left( \frac{c'}{w}, \frac{\langle c' \rangle}{w} \right) = 1 - Q \left( \frac{c'}{w}, \frac{B'}{w} \right)$$

(7)

where $Q$ is the (complementary) incomplete gamma function, defined by

$$Q(a,x) = \frac{1}{\Gamma(a)} \int_x^{\infty} dt \, e^{-t} t^{a-1}.$$  

$\Gamma$ here represents the gamma function.

2.4. Weighted-Poisson case with $N > 1$ weights

Formally speaking, we are now no longer in the Poissonian regime, since a weighted sum of two or more Poissonian variates does not itself in general have a Poissonian probability distribution. I have not been able to find a closed-form expression for the probability density function in this case. However, two empirical approximations are presented in the present subsection.

2.4.1. The Fay and Feuer approximation

Fay and Feuer (1997) suggested on heuristic grounds that the null-hypothesis probability in this general case might be given to a good approximation by equation 7 with $w$ replaced by an appropriate equivalent weight $w_{equiv}$. Their prescription for $w_{equiv}$

$$w_{equiv} = \frac{\sigma'^2}{\langle c' \rangle},$$

where

$$\langle c' \rangle = \sum_{i=1}^{N} w_i \langle c_i \rangle$$

would
and, from equation \[4\]

\[\sigma^2 = \sum_{i=1}^{N} w_i^2(c_i),\]

is obtained essentially by equating respectively the first and second moments of the single-weight and many-weight probability functions.

In fact when one compares the integrated \(P_{\text{null}}(c')\) function derived from the Fay and Feuer approximation against Monte Carlo data (see figure 1), one sees that the Fay and Feuer curve seems to be displaced too far to high \(c'\). Essentially this is because equation 7 defines a continuous envelope function to the actual discontinuous, stepped \(P_{\text{null}}\) of a single Poisson variate. Consider now a weighted sum of \(N\) Poisson variates, but with all the weights having the single value \(w\): here the sum itself remains a Poisson variate, in which case the true integrated probability distribution remains coarsely stepped as in figure 1 the envelope to this being exactly given by equation 7 with \(w_{\text{equiv}} = w\). The coarseness is because there are no combinations of \(c_i\) which can produce any value of \(c'\) other than \(c' = jw\) for integer \(j\); the steps in the integrated probability distribution \(P_{\text{null}}\) are thus of width \(w\). If we now allow the weights to be randomly perturbed by small amounts, the effect is to smear out the steps: ie, a range of values of \(c'\) close to \(jw\) now become possible. Where the weights \(w_i\) are allowed to become entirely random and independent (and are sufficiently numerous), the coarse steps disappear entirely\(^2\), and the probability curve appears to steer a middle course (eg the dotted line in figure 1) through the coarse steps of the single-weighted Poisson distribution of the same equivalent weight. It seems clear then that the approximation formula suggested by Fay and Feuer could be made more closely applicable to the general case by shifting it towards lower \(c'\) values by an amount \(0.5w_{\text{equiv}}\). The resulting formula is

\[P_{\text{null}}(c') = 1 - Q\left(\frac{c'}{w_{\text{equiv}}} + \frac{3}{2} \frac{B'}{w_{\text{equiv}}}\right).\]  \(8\)

Figure 1 shows an example of a \(P_{\text{null}}\) distribution derived from a Monte Carlo exercise (dotted line). The Monte Carlo ensemble consisted of \(10^6\) values of the weighted sum \(c'\). 25 random weights \(w_i\) were chosen before the start of the exercise: these initially had a uniform probability distribution between 0 and 1 but were then normalized so that they summed to 1. (The only effect of this normalization is to change the x-axis scale.) For each member of the ensemble, a Poisson-random integer \(c_i\) was generated for each of the 25 bins, using a constant background value \(B = 0.3\) counts per bin as the Poisson expectation value. The weighted sum \(c'\) of these 25 random values was made and added to the ensemble. The cumulative probability was then formed by summing, from high towards low values of \(c'\), a 100-bin, normalized histogram of the ensemble values.

\(^2\) Note that the true probability curve in the case that \(N > 1\) still falls in stepwise fashion, since the possible values of \(c'\) in this case are still discrete; it is just that the steps are much more finely spaced, since the spacing between possible \(c'\) values (and thus also between steps) decreases on average at a rate proportional to \((c')^{-N}\).
and grey lines show respectively the Fay and Feuer and figure 2 was similar to that of figure 1. It consisted of an ensemble of 10 proved to be an acceptable approximation down to at least variates at lower values of age background of 0.1 counts for the null probability shows ratios between probability densities equation 9. The solid black line gives the modified Fay and Feuer approximation of equation 8; the solid grey line, the line of the probability of the null hypothesis, it is desirable to find a better approximation in this region.

2.4.2. The $\chi^2$ approximation

The $\chi^2$-like integrated probability distribution

$$P_{\text{null}}(c^\prime) = Q\left(\frac{B}{w_{\text{equiv}}}, \frac{c^\prime}{w_{\text{equiv}}}{\right)}$$  \hspace{1cm} (9)$$

proved to be an acceptable approximation down to at least $P_{\text{null}} = 10^{-5}$ over a wide range of background values. An example is shown in figure 2.

The Monte Carlo simulation used to generate the results of figure 2 was similar to that of figure 1. It consisted of an ensemble of 10$^7$ weighted sums. The weights in this case constituted a matched filter for detection of sources in images with an average background of 0.1 counts/pixel, the source shape being the on-axis point spread function at 1.25 keV of the XMM-Newton EPIC PN camera.

It might be possible to improve the fit of equation 9 by choosing a different value of $w_{\text{equiv}}$, but this possibility has not been explored.

It is not at present known why the $\chi^2$ cumulative distribution given in equation 9 seems to provide such a good model of the weighted-sum $P_{\text{null}}$.

2.4.3. Consequences of divergence of an approximation

Suppose one inverts a formula such as equation 8 in an attempt to deduce the detection sensitivity at a given value of $P_{\text{null}}$. What in particular goes wrong if the formula is not a good approximation to the true probability distribution? The answer is that the returned sensitivity value is correct, but the assumed $P_{\text{null}}$, which controls the number of false positives expected, will be incorrect. In the example of figure 2a, cutting the source list at an apparent null likelihood of 8 will yield a sensitivity of about 0.31 weighted counts if the Fay and Feuer expression is used. The true null likelihood at that value of $c^\prime$ is that of the Monte Carlo, ie about 6.7. This means that about 3.7 ($= e^{3.7}$) times more false positives will be encountered than expected. The $\chi^2$ formula on the other hand diverges from the Monte Carlo data in the other direction. Use of this formula gives a conservative result, being apparently slightly less sensitive at $c^\prime = 0.36$, with the true null likelihood cutoff of about 8.6 yielding approximately 0.55 fewer false positives than expected. In order to get something like the desired rate of false detections, one must skew the apparent null likelihood cutoff, to about 10 in the Fay and Feuer case and 7.5 in the $\chi^2$ case. The sensitivity in both cases will then be ‘correct’ at about 0.34 weighted counts.

2.5. Optimization of the weights

As described in section 2.4, for a given background $B$ and signal template $S$, any given value of the weighted sum of counts $c^\prime$ is associated with a unique value of the null-hypothesis probability $P_{\text{null}}$. In the preceding subsection two equations 8 and 9 were given which approximate this relation for the situation in which the observed data are random Poisson variates. In order to calculate optimum weights in this case we must first invert an equation of this form to obtain $c^\prime$, then invert equation 9 to obtain the counts amplitude $\beta$.

Inversion of either equation 8 or 9 to obtain $c^\prime$ involves an inversion of the incomplete gamma function $Q$. For the sake of practicality let us define the two inverse functions $Q_1^{-1}$ and $Q_2^{-1}$ as follows: if

$$P = 1 - Q(a, x),$$

then let

$$a = Q_1^{-1}(1 - P, x).$$

and

$$x = Q_2^{-1}(1 - P, a).$$
No closed-form expressions for either \( Q_1^{-1} \) or \( Q_2^{-1} \) seem to be known; for present purposes I have performed the inversions numerically by means of a Ridders-method routine (Ridders 1979) as given in Press et al (2003).

After insertion of the appropriate inverse gamma function, the count amplitude \( \beta \) is thus given for the modified Fay and Feuer approximation by

\[
\beta = \frac{w_{\text{equiv}} \left[ Q_1^{-1} \left( 1 - P_{\text{null}} \right) - \frac{1}{2} \right] - B'}{S'}
\]

and for the \( \chi^2 \) approximation by

\[
\beta = \frac{w_{\text{equiv}} \left[ Q_2^{-1} \left( P_{\text{null}} \right) - \frac{1}{2} \right] - B'}{S'}
\]

The inputs to these formulae are (i) the set of weights, (ii) the value of \( P_{\text{null}} \), and (iii) the background and signal-shape information. Only the first two of these are under our control. Before we may begin to seek for the optimal set of weights, we must choose a value of \( P_{\text{null}} \). The value we choose should be governed by the maximum fraction of false detections we are prepared to tolerate. The desired cutoff value \( P_{\text{det}} \) of null probability can be obtained from the maximum acceptable number \( n_{\text{false}} \) of false detections by dividing by the number of ‘beams’ in a typical image, which is just the image solid angle divided by the ‘beam’ solid angle. In the present case, in which images are convolved before the null hypothesis is tested, the beam solid angle must be some kind of equivalent solid angle of the appropriately convolved PSF. For XMM-Newton at least, where the shape of the PSF varies across the field of view, its average value is probably best estimated via Monte Carlo trials. If however, for the sake of obtaining at least a rough approximation to the ratio between \( P_{\text{det}} \) and \( n_{\text{false}} \), we use the value \( 2.34 \times 10^{-5} \text{ deg}^2 \) calculated in appendix A for the equivalent solid angle of the on-axis EPIC PN PSF as a lower limit to the beam \( \Omega \), we find that a null-probability cutoff of \( \exp(-8.0) \), the value used in the making of the 1XMM catalog, corresponds to at most 2 expected false detections per image.

Let us define \( \beta_{\text{det}} \) as the count amplitude \( \beta \) which, for a given set of weights, corresponds to the chosen value of \( P_{\text{det}} \). \( \beta_{\text{det}} \) can be viewed as the amplitude of a source which is just detectable under these conditions. The optimum weights are then clearly those which yield the smallest value of \( \beta_{\text{det}} \).

In the present study, Powell’s direction-set method as modified by Press et al (Press et al 2003, chapter 10.5) was used to optimize sets of weights by minimizing \( \beta_{\text{det}} \), as defined by either equation (10) or (11) as a function of the weights.

### 3. Detection of x-ray point sources in XMM-Newton data

The EPIC x-ray cameras of XMM-Newton are described in Str¨ uder et al (2001) and Turner et al (2001). There are three cameras: the telescope in each case is similar but two of the CCD detectors comprise 7 chips of MOS type, whereas the third has 12 chips of pn composition. The ‘good’ area of each of the three cameras occupies about 94% (MOS) and 82% (pn) of a 30’ diameter field of view. CCD pixel dimensions are 1.1’ square (MOS) and 4.1’ square (pn).

#### 3.1. The source-detection strategy used for the 1XMM catalog

For each exposure, images in sky coordinates were made in five separate energy bands. The images had square pixels of \( 4 \times 4 \) arcsec dimension. Images were made by transforming the position on the detector of each selected x-ray event into sky coordinates, then binning up the events into the image pixels. The position of each event was dithered within the boundaries of the CCD pixel in which it was detected. Variations over time of the spacecraft attitude were also taken into account.

Source detection was performed on the five images in parallel. The source detection comprised a convolution and detection stage (steps 1 to 3 of the sequence described in the introduction), followed by a source-parameterisation stage (step 4 in the sequence). The detection etc stage was performed by the XMM-Newton SAS task emldetect, the parameterisation by emldetect. Both stages involve the calculation of a detection likelihood; since the value calculated by emldetect is arguably more sensitive than that of emboxdetect, the ideal procedure is to run emboxdetect with a deliberately low detection threshold, submit the resulting long list of source candidates to emldetect and then accept as genuine sources only those for which the emldetect detection likelihood exceeded a second, more reasonable threshold. The emboxdetect threshold should not be so close to the emldetect threshold that the two selections interfere. In 1XMM practice there is some doubt as to whether the two detection thresholds were sufficiently far apart. For this reason, and because I don’t understand enough of the emldetect likelihood calculation to be able to replicate it, I have in the present paper only considered the sliding-box stage of the 1XMM detection procedure.

The first step of this procedure was to make maps, 1 per energy band, of the estimated background in each pixel. The five images and five background maps were then each convolved with a square, 5 × 5 array of unit values.\(^3\) In mathematical form this processing can be represented as follows:

\[
c_{i,j,k} = \sum_{p=-2}^{2} \sum_{q=-2}^{2} c_{i-p,j-q,k}
\]

and

\[
b_{i,j,k} = \sum_{p=-2}^{2} \sum_{q=-2}^{2} b_{i-p,j-q,k}
\]

where \( i \) and \( j \) indicate the position on the image pixel grid and \( k \) refers to the energy band.

\(^3\) Two further rebinning and convolution steps, approximately equivalent to convolution respectively by \( 10 \times 10 \) and \( 20 \times 20 \) arrays, were also performed. However, the principal purpose of these extra steps was to detect extended sources: therefore they can safely be neglected for purposes of the present discussion.
Although a full analysis of the 1XMM source-detection technique is beyond the scope of the present paper, a Monte Carlo simulation of the statistical fluctuation of background was performed in order to check the accuracy of equation 13. An ensemble of $5 \times 10^7$ values of $L$ was accumulated. The procedure for each member of the ensemble was as follows. For each of the 5 bands, a $5 \times 5$ array of Poisson-random integers was generated. To calculate the expectation value $B_{i,j}$ for pixel $(i, j)$ in the $k$th band, the $k$th normalized background weight for the XMM PN camera, as listed in table 1, was multiplied by 0.1 counts/pixel. $c'_{k}^j$ was calculated by summing the 25 counts values of the $k$th band, $B_{i,j,k}$ being of course simply equal to $25 \times B_{i,j,k}$. Likelihoods were then generated for each band by use of equation 7 and summed to give $L$.

The accumulated histogram of the resulting data, shown in figure 3 shows that there is a significant difference between the Monte Carlo results and the prediction of equation 13. Attempts were made to fit a variety of functions to the Monte Carlo data. One must of course fit to the raw histogram data, the adjacent channels of which are statistically independent, rather than to the integrated curve displayed in figure 3. It is not easy to fit a smooth function to the Monte Carlo data. Partly this is because the raw data are rather ‘noisy’, particularly at low values of $L$, a phenomenon which arises because in this range the input counts values to each of the five channels are usually small integers which, when combined, give rise to a sparse distribution in the allowed resulting $L$ values. The effect of this can be seen in the jumpy nature of the accumulated Monte Carlo data shown in figure 3. It is also not straightforward to define a statistic to be minimized in order to generate the fit. A straightforward sum of squared residuals, or even a chi squared sum (ie, sum of squared residuals, each divided by the variance), tends to result in the low-$L$ part of the data dominating the fit. This is undesirable if one is interested in minimizing false detections due to statistical fluctuations in background, in which case intermediate values of $L$ are more important. In the end I chose arbitrarily to minimize a sum of terms

$$Z = \sum_j \left( f_j - y_j \right)^2 / \sigma_j^2$$

where $f$ is the function to be fitted, $y$ represents the Monte Carlo data and $\sigma^2$ is the square of the variance. The fit was however restricted to values of $L$ less than 15, to avoid statistical noise in the Monte Carlo values at high $L$.

There is an infinity of functions one could choose to fit to the data, but in fact a good fit as shown was obtained simply by allowing the first term $\nu/2$ in the $Q$ function in equation 13 to vary. The best fit, shown in figure 3, occurred at $\nu = 5.81$. This however suggests no obvious systematic correction to equation 13 and, since I have at this time no better analysis of the problem to put forward, and since the unmodified equation 13 was in fact used in the 1XMM source detection, I have retained it for comparison with the matched-filter method. It seems clear however that the rate of statistical false detections in 1XMM was probably lower than originally estimated, and thus that a lower detection cutoff could have been used in that survey with-
out penalty. There appears to be no better way to correct estimates of the sensitivity of this method at the moment than by calibrating it via several Monte Carlos performed at different values of background (see eg section 3.2.2).

A final point to note about the 1XMM detection procedure: it did not make use of the fact that, for any given exposure, images from more than one of the XMM-Newton x-ray cameras were usually available. All source detection was performed instead on a camera-by-camera basis. Clearly one would expect a technique which made use of the parallel information to yield an improvement in detection sensitivity.

3.2. Matched filters for XMM-Newton source detection

There are three obvious ways to improve on the 1XMM source-detection procedure by use of matched filters (weighted sums). Firstly, taking one energy band at a time, since we know to a reasonable approximation the point spread function (PSF) of the XMM-Newton telescopes, we could choose in each energy band a set of weights optimized for detecting that shape of signal; secondly, we might do the same thing across the energy bands by the expedient of assuming a common spectrum for the sources; thirdly, we might in similar fashion add together images from the three x-ray cameras of XMM-Newton. Only the first two alternatives are explored in the present study.

3.2.1. Examples of matched filters for 1 energy band

We take a square array of dimension \(2N + 1\). Let us assume a parent function \(C_{i,j}\) as follows:

\[
C_{i,j} = B_{i,j} + \alpha S_{i,j}, \quad \forall i, j \in [-N, N]
\]

with the normalization condition

\[
\sum_{i=-N}^{N} \sum_{j=-N}^{N} S_{i,j} = 1.
\]

For simplicity, the background \(B\) is assumed to be constant across the array. Let \(S_{i,j}\) be the PSF on the optic axis of the PN telescope of XMM-Newton, at an energy of 1.25 keV (the mean energy of band 2 as defined in the 1XMM catalog (Watson et al 2003)). The PSF model used in the present exercise was originally calculated via a ray-tracing approach (Gondoin et al 1998), and is the same that was used to determine positions and fluxes of the sources in 1XMM. In the present exercise the PSF was centred on the middle of the \((i, j) = (0, 0)\) pixel.\(^4\)

We assume that we have an array of measured count values \(c_{i,j}\) across the array, each of which is a random, Poisson variable, with \((c_{i,j}) = C_{i,j}\). We make a weighted sum of the \(c_{i,j}\) as follows:

\[
c' = \sum_{i=-N}^{N} \sum_{j=-N}^{N} w_{i,j} c_{i-j,
\]

where \(w_{i,j}\) is a function of \(i, j\) which is positive for the central pixel, and zero elsewhere.

\(^4\) In practice, the centre of the PSF for any real source can of course be located anywhere within the ‘central’ pixel. A more rigorous procedure would take this into account. However, because it is not clear how best to do this, the simpler assumption has been made for the present study.

and likewise for \(B'\).

The two cases we want to compare are, firstly, the 1XMM-like case in which the \(w_{i,j} = 1\) and \(N = 2\); secondly the case in which the \(w_{i,j}\) are optimized according to the procedure described in section 2.5. But before the second procedure can be used, a value of \(N\) must be chosen. It is not hard to see that, in principle, the larger the \(N\) the better. To see this, suppose we compare two convolvers: one optimised on a \((2N + 1)\)-square array, the other on a \((2(N - 1) + 1)\)-square array. The optimized values of the small convolver can be thought of as a subset of the possible values of the large convolver; one just sets the extra ring of pixels to zero. So, if the smaller convolver were a better source-finder than the large, the optimisation routine would have set the outer pixel values to zero automatically, giving rise to the same sensitivity of detection with the large as with the small. Thus a small optimized convolver can never be better than a large one; and the only limiting factor to \(N\) becomes computational practicality. A value of \(N = 4\) for the optimized convolvers has been chosen as the largest value which can be processed in the 5-band case (see section 3.2.2) in a reasonable time.

The difference in size between the 1XMM and the optimized convolvers makes it more difficult to compare their efficiency. It seems a bit pointless to hobble the optimized convolver by restricting its size to the 1XMM \(5 \times 5\) - after all, the whole aim of the exercise is to achieve the maximum practical improvement in sensitivity. It might be argued though that the discrepant sizes are unfair to the 1XMM ‘box’ convolver - if bigger convolvers are sometimes better, might not much of the gain in going from a \(5 \times 5\) box-type to a \(9 \times 9\) optimized simply be due to the increase in size? It turns out however that box-type convolvers larger than \(5 \times 5\) perform uniformly worse than the \(5 \times 5\), as is shown in figure 3. In fact, if we go in the other direction, to a \(3 \times 3\) box, we get a slightly improved performance at all but the lowest background fluxes. Hence one can conclude that the comparison between convolvers of different size is probably being rather kind to the 1XMM algorithm than otherwise.

Some examples of optimized \(9 \times 9\) kernels \(w_{i,j}\) are compared to the PSF \(S_{i,j}\) in figure 5. One would expect that \(w \to 1\) as \(B \to 0\) (all the counts are equally valuable) and \(w \to S\) as \(B \to \infty\) (Gaussian limit). This is consistent with the form of the high- and low-background kernels shown in the figure.

The quantity which should be compared is the counts amplitude \(P_{\text{det}}\), whereas \(P_{\text{null}}\) is the null-probability probability. This is consistent with the form of the high- and low-background kernels shown in the figure.

The quantity which should be compared is the counts amplitude \(P_{\text{det}}\) as given in for example in equation 9. at which the signal is just detectable - that is, at which the resulting hypothesis probability is just equal to some previously decided cutoff \(P_{\text{det}}\). Since the 1XMM sources were detected at a probability cutoff of \(P_{\text{null}} = 0.005\) (equivalent to about a 4.3-sigma detection), this is the value that was chosen for the present exercise.
Fig. 4. Sensitivity of different convolvers as the array size is varied. The array side length is \(2N + 1\). The squares represent the 1XMM ‘box’ convolver. The optimized convolver is represented by both crosses and filled triangles: the crosses show the estimates obtained by using the modified Fay and Feuer approximation (equation 10), whereas the triangles show the estimates from the chi-squared formula (equation 11). Those cases in which the sensitivity of the optimized convolver appears to be worse than the box-type, and where the two approximate formulae for \(\beta\) give widely differing results, are because the approximations have not here been corrected by Monte Carlos as in figure 6.

Fig. 5. Contour plots of the XMM-Newton PSF compared to two corresponding optimized convolution kernels. a) The on-axis PSF for the XMM-Newton EPIC PN camera at 1.25 keV, rebinned to 4 arcsec square pixels. b) Optimized convolution kernel for background \(B = 10^{-4}\) counts/pixel. c) Same, but optimized for background \(B = 10\) counts/pixel. All arrays were normalized before plotting. The main contour separation (solid contours) is 0.01; the lowest of these intervals has again been divided into five (dashed contours).

Comparing figures 6 and 8 shows that the degree of improvement to be gained through the use of matched filters is approximately constant across the field of view. Also, whatever the method used, the sensitivity decreases by about 15% towards the edge of the field of view. This is because the less compact the PSF, the more difficult it is to sequester source from background counts - the effective background counts involved with the source are higher. For similar reasons, a decrease in sensitivity is seen in the detection of extended (i.e., non-pointlike) sources. Use of on-axis weights for the off-axis signal degrades the sensitivity by about 10% over the whole range of background. For values of background lower than about 0.1 counts/pixel the sensitivity using this un-matched filter becomes nominally worse than that achievable via the box-

approximations, as well as four points (diamonds) where the true \(P_{\text{null}}\) has been estimated via Monte Carlos of \(10^6\) iterations each.

As described in the introduction, several of the x-ray source detection procedures to be found in the literature include a step in which the raw images are convolved with the telescope PSF. The PSF is an optimum convolver in the Gaussian limit but may be expected to depart from the ideal at low values of background flux. To investigate this, the detection sensitivity obtained by use of the PSF as convolver was calculated at four levels of background. The resulting sensitivities, corrected via use of Monte Carlos to estimate \(P_{\text{null}}\), are plotted as black squares on fig 6.

Several conclusions can be drawn from figure 6. Firstly, for this form of signal, the Fay and Feuer approximation appears to be unserviceable for all but the highest values of background. The \(\chi^2\) formula performs much better, accurately representing the true data down to about 0.2 counts/pixel background, below which it begins to diverge. Clearly though the matched-filter approach yields a better sensitivity than the 5\(\times\)5 box convolver at all values of background, apparently asymptoting to about 25% better at high background, the advantage decreasing to zero at low. The PSF appears to be a useful approximation to the optimum convolver for background levels greater than about 0.03 counts s\(^{-1}\).

The PSF of the XMM-Newton EPIC cameras becomes azimuthally distorted with distance from the centre of the field of view. It is therefore of interest to repeat the above exercise for an example of the off-axis PSF. In the present case, a PSF at 850 arcsec from the edge of the optical axis (94% of the radius of the field-of-view of the EPIC cameras), at an azimuth of 45\(^\circ\), was arbitrarily chosen. All other variables were retained unchanged. No Monte Carlos were performed in this case, and only the results of the \(\chi^2\) formula were used. The results can be seen in figures 7 and 8.

Comparison between figures 6 and 8 shows that the degree of improvement to be gained through the use of matched filters is approximately constant across the field of view. Also, whatever the method used, the sensitivity decreases by about 15% towards the edge of the field of view. This is because the less compact the PSF, the more difficult it is to sequester source from background counts - the effective background counts involved with the source are higher. For similar reasons, a decrease in sensitivity is seen in the detection of extended (i.e., non-pointlike) sources. Use of on-axis weights for the off-axis signal degrades the sensitivity by about 10% over the whole range of background. For values of background lower than about 0.1 counts/pixel the sensitivity using this un-matched filter becomes nominally worse than that achievable via the box-
The minimum detectable counts amplitude $\beta_{\text{det}}$, plotted as a function of background flux $B$. The solid line represents the results obtained by convolving with a $5 \times 5$ unit array, as used in the creation of the 1XMM catalog. The remainder of the plot refers to a weighted-sum scheme in which the weights are optimized to the on-axis PSF of the XMM-Newton EPIC PN camera at 1.25 keV, binned up on a 9 by 9 grid of 4 arcsec square pixels. The dash-dot line shows the sensitivity predicted by the modified Fay and Feuer approximation formula ($\beta_{\text{det}}$ given by equation 10). The dashed line gives the prediction of the $\chi^2$ formula ($\beta_{\text{det}}$ given by equation 11). The diamonds are samples of the ‘true’ sensitivity of the matched filter method as derived from Monte Carlo experiments. The black squares, also Monte Carlo corrected, are the result of convolving with the PSF instead of the optimized convolver.

The variation in the shape of the PSF in XMM-Newton EPIC images puts some practical difficulties in the way of source detection via matched filters, because one has, in effect, to employ a different convolving kernel for each pixel of the image. This was the method adopted by Ueda et al (1999). An approximation to this which still allows one to use the useful mathematical properties of convolution is to divide the image into patches, convolve each patch separately, then add the results. Vikhlinin et al (1995) used this technique. The XMM-Newton SAS (Gabriel et al 2004) task assmooth (Stewart 2004) can also perform such a piece-wise convolution.

3.2.2. Matched filters for 5 energy bands

In this case it is no longer possible to invert the 1XMM and matched-filter methods in the same way, since the summed-likelihood approach used to find 1XMM sources in 5 bands (described in section 3.1) cannot be expressed as a weighted sum of Poissonian integers. However, equations 5, 6 and 13 taken together amount to a relationship between the counts amplitude $\beta$ and the null-probability $P_{\text{null}}$: hence one can numerically invert this relationship to obtain the sensitivity $\beta_{\text{det}}$ which corresponds to the cutoff probability $P_{\text{det}}$, which is left at $e^{-8}$ as before.

In order to calculate a matched filter for summing images in several energy bands, one must know the relative strength of the signal in each band, which amounts to knowing the source spectrum. Where sources have a variety of spectral shapes, as is the case for cosmic sources of x-rays, the matched-filter technique can only be optimized for a single class of sources at a time. The performance of the matched filter against an unmatched spectrum is examined toward the end of the present section.

For purposes of the present study, weights were optimized to detect x-ray sources with an absorbed power-law spectrum having a photon index 1.7 and a HI column density of $3.0 \times 10^{20}$ cm$^{-2}$. The relative count rates in each band were obtained, via the program xspec, by folding this spectrum with the on-axis effective area function of the XMM-Newton PN camera. These
Table 1. Source and background weights for the XMM-Newton PN camera.

| Low | High  | Background weights: | Source weights: | Hard source: |
|-----|-------|---------------------|----------------|-------------|
| 0.2 | 0.5   | 0.0842              | 0.1208         | 0.0023      |
| 0.5 | 2.0   | 0.1733              | 0.3763         | 0.0193      |
| 2.0 | 4.5   | 0.0941              | 0.0943         | 0.2993      |
| 4.5 | 7.5   | 0.0941              | 0.0350         | 0.6172      |
| 7.5 | 12.0  | 0.1485              | 0.0094         | 0.0619      |

Table 2. Values of $v$ obtained by fitting $Q(v/2, L)$ to Monte Carlo distributions of $L$ at different background levels.

| Nett background (counts/pixel) | Fitted $v$ |
|-------------------------------|------------|
| 0.01                          | 4.31       |
| 0.032                         | 4.21       |
| 0.1                           | 5.81       |
| 0.32                          | 7.26       |
| 1.0                           | 8.22       |
| 3.2                           | 8.90       |
| 10.0                          | 9.35       |

As noted in section 3.2.1, a value of $N = 4$ for the size of the optimized convolvers was chosen as giving the largest convolver which was still practical to compute. The optimization with 5 bands is observed to be a little slower than for the single-band case, understandable because there are are now $5 \times (2N + 1)^2 = 405$ weights which must be optimized in parallel.

The two approaches are compared in figure 9. Before commenting on the difference in sensitivity between the two methods we ought to make sure we have an accurate measure of those sensitivities. As shown via a Monte Carlo experiment in section 3.1, equation (13) appears to be a poor approximation at values of background less than about 1.6 counts/pixel: that is, for each value of background, a Monte Carlo ensemble was generated, a cumulative distribution of the ensemble values $L$ was calculated, and finally a function $Q(v/2, L)$ was fitted to this distribution curve. The resulting values of $v$ are tabulated in table 2. The ‘canonical’ value of $v$ in this case is 10 ($= 2 \times 5$ bands).

Sensitivity values obtained by replacing the 5 in equation (13) by the appropriate value of $v/2$ are shown by crosses in figure 9. Clearly, use of the unmodified formula exacts a significant sensitivity penalty (because of over-estimation of the rate of false positives) at values of background less than about 1.0 counts/pixel.

As was seen in section 3.2.1, the $\chi^2$ formula used to approximate the null-probability distribution of weighted-Poisson-sum values also tends to diverge from the true distribution at low values of background. A similar set of Monte Carlo corrections was therefore performed for the weighted-sum data. Fortuitously, for the 5-band signal presently chosen, the corrections appear to be insignificant. The reader is cautioned not to expect this to be the case for all signal shapes.

As mentioned in section 2.3, one may compensate for the distortion in the true null likelihood (thus in the rate of false positives due to background fluctuations) by changing the value of nominal null likelihood used to sort ‘sources’ from ‘non-sources’. Such compensating values of null likelihood for seven values of background, evenly-spaced on a logarithmic scale, are plotted in figure 10.
Fig. 10. Nominal values of detection likelihood which should be chosen in order to obtain a true detection likelihood of 8. The crosses show the results relevant to the 1XMM procedure; diamonds refer to the matched-filter method (\( \chi^2 \) formula).

Increasing gradually to about 1.2 at the lowest background value plotted.

A hasty comparison of figures 6 and 9 may lead one to conclude that there is, paradoxically, not much advantage to be gained by detecting sources over 5 bands rather than 1. However, the vertical scales of the two graphs are not equivalent, because they refer to different signal shapes. To make a proper comparison one would need to multiply all the background values of figure 9 by 0.1733 (the proportion of the total background found in energy band 2) and, for like reasons, all the amplitude values by 0.3763. Comparison with figure 6 then reveals an improvement in sensitivity for 5- versus 1-band detection by about a factor of 2 for the matched-filter method and 1.5 for the 1XMM method.

As mentioned earlier, a filter which has been optimized to detect sources of a particular spectrum may not perform well in the detection of sources with very different spectra. To investigate this, the exercise of figure 6 was repeated. The filter was optimized as before for a source having weights listed in column 4 of table 1 but the sensitivity values were calculated under the assumption that the source weights came from column 5. The results are plotted in figure 11.

It is apparent that the sensitivity of the matched filter to the hard-spectrum source is significantly degraded - indeed by a factor of 2 - whereas the detection efficiency of the 1XMM procedure is if anything slightly improved. Although this hard spectrum is the worst case likely to be encountered in practice, the results suggest the desirability of using a non-matched procedure in parallel. It is however also worth noting in passing that, regardless of whether the detection technique is matched to a particular source spectrum or not, some spectrum must be assumed in order to calculate any kind of multi-band sensitivity value.

3.2.3. A check on the sensitivity results

So far we have been ‘working backwards’, inverting expressions to obtain estimates of the minimum signal detectable under a variety of conditions. As a check on this procedure, a ‘forwards’ Monte Carlo experiment was performed as follows. Random data were generated from parent distributions of the form of equation 1 for a sequence of values of the amplitude \( \alpha \). The data were generated in 5 bands, using the PN background ratios tabulated in table 1 with a net background flux of 1.0 counts/pixel. The usual PSFs provided the signal appropriate to each band. An ensemble of 10^4 mini images was accumulated at each of 200 equally-spaced values of \( \alpha \). Each 5-band image stack was submitted to both the 1XMM and the matched-filter source detection procedures. The detection frequencies as functions of \( \alpha \) are compared in figure 12. From figure 9 one would expect that signals of \( \alpha > \beta_{\text{det}} \) for that background flux are occasionally detected, and signals with \( \alpha < \beta_{\text{det}} \) are occasionally not, the results of this Monte Carlo are consistent with the earlier analysis.

4. Conclusions

It is not possible to detect with certainty a signal superimposed on a noisy background where there is some non-zero probability that a combination of random background values can mimic the signal: the best that can be done is to calculate a detection probability - or its complement, the null or non-detection probability. The aim of any method to enhance signal detectability is therefore to decrease the null probability of a signal of any given amplitude, or, equivalently, to decrease the amplitude at which a signal generates a given value of null probability. This
must be achieved by performing some transform upon the set of measured values of signal plus background plus noise.

The main thrust of the present paper has been to examine that subset of transforms which may be expressed as discrete convolutions of the input data. There appears to be nothing new to say about the case in which the probability distribution of the noise value in any given data channel is Gaussian; because of this, attention has been further restricted to the Poissonian regime. Two independent approximations to the null-probability distribution for a convolution of Poissonian data have been described and compared; such approximate formulas allow one to optimize the convolution for detection of signals of any specified shape without performing a Monte Carlo experiment. The vertical scale gives the fraction of 'sources' at each value of \( \alpha \) which were 'detected'.

![Figure 12](image.png)

**Fig. 12.** Detection sensitivities of the 1XMM (solid line) and matched-filter (dashed line) methods are compared via a Monte Carlo experiment. The vertical scale gives the fraction of 'sources' at each value of \( \alpha \) which were 'detected'.

A natural extension of the convolver method as applied to x-ray source detection is to allow one to correctly add together overlapping images. One's first impulse in this situation is simply to add the images without weighting, but (certainly in the XMM-Newton case) because the background rate can vary with time, separate images may have different source/background ratios and should therefore be weighted accordingly. The theory described in this paper allows one to estimate the optimum weights for such combinations.
The author hopes to make the matched-filter method available in the edetect package of the v-6.5 release of the XMM SAS.

Appendix A: The x-ray background

As mentioned in the introduction, there is good evidence that the bulk of the cosmological x-ray background consists of numerous faint discrete sources. The present paper assumes however that the x-ray sky may be modelled by sparse, relatively bright sources (though not all necessarily detectable at any given exposure duration) upon a relatively smooth background. It is therefore of interest to see how consistent this model is with our current understanding of the real sky.

The fundamental quantity to deal with is the probability distribution of detected flux, where the ensemble is taken to consist of measurements at random directions in the sky; non-source background is neglected and it is assumed that each measurement is free from other sorts of variation, eg the Poissonian detection noise we have been discussing so far. The connection between the distribution of source flux and the distribution of detected source counts depends in a mathematically complicated way on the ‘beam shape’, or PSF in our case. Useful treatments of the topic can be found in Condon and Scheuer (1974).

For present purposes it is enough to consider the standard deviation of such a probability distribution. Let us consider the case in which the distribution of source flux $n(S)$ obeys a power law, viz:

$$n(S) = kS^{-\gamma}.$$  

Condon derives the following formula for standard deviation $\sigma$ in this case:

$$\sigma = \left( \frac{k\Omega_{c}C_{max}^{3-\gamma}}{3-\gamma} \right)^{1/2},$$  

where $C_{max}$ is some upper limit on the detected source counts and $\Omega_{c}$, the equivalent solid angle of the PSF, is defined as

$$\Omega_{c} = \int [psf(\theta,\varphi)]^{-1} d\Omega.$$  

Note that $\sigma$ is unbounded for the entire probability distribution (ie, as $C_{max} \to \infty$).

As for $n(S)$, the measurements described in Mushotzky et al (2000) are consistent with a ‘two-slope’ model in which, for energies from 0.5 to 2 keV (corresponding to energy band 2 of 1XMM), $\gamma$ has the value 1.7 below $S = 7 \times 10^{-15}$ erg cm$^{-2}$ s$^{-1}$ and 2.5 above it. The $k$ values can be evaluated from

$$k_{faint} = k_{bright} \frac{S^{1-\gamma_{faint}}}{1-\gamma_{faint}} = k_{bright} \frac{S^{1-\gamma_{bright}}} {1-\gamma_{bright}} \sim 200 \text{ deg}^{-2}$$

where $S$ is the flux at the ‘knee’.\n
Which $\gamma$ to choose? 200 sources per square degree, the integrated number density at the knee, yields about 30 sources per EPIC PN field. This is quite a typical number of serendipitous sources to find in such fields: hence one can say that the PN camera will, in a typical exposure, detect sources fainter than the flux at the knee. I therefore choose a value of 1.7 for $\gamma$. Substitution of this value into equation (A.2) yields $2.34 \times 10^{-5}$ deg$^{-2}$ for the equivalent area of the on-axis, band 2 EPIC PN PSF.

At this point it is convenient to convert fluxes $S$ to counts $C$. The highest value of background considered in the present paper is 10 counts pixel$^{-1}$. By use of the same histogram technique as described in section 3.2.2 one may deduce that the typical background count rate for EPIC PN in 1XMM band 2 is about $2 \times 10^{-5}$ counts pixel$^{-1}$ s$^{-1}$. A background value of 10 counts pixel$^{-1}$ thus corresponds to an exposure time of $5 \times 10^{5}$ s, which is 5 times longer than the longest (PN) exposure time in the 1XMM catalog. Some multi-epoch observations made with XMM-Newton may approach this duration however; hence I take it as a reasonable upper limit to practical XMM-Newton observations. The flux to count-rate conversion factor for 1XMM band 2 (calculated in the same exercise as the source weights tabulated in table 1) is $7.50 \times 10^{11}$ counts erg$^{-1}$ cm$^{-2}$; the ‘knee’ in the Mushotzky soft-band logN-logS diagram thus falls, for a $5 \times 10^{5}$ s exposure time, at $\sim 2600$ counts in 1XMM band 2 and (making use of the source weights in table 1) 7000 counts in the total band. Comparison of these numbers with figure 9 shows that even the 1XMM algorithm could detect sources 2 orders of magnitude fainter than the ‘knee’ at this exposure duration; clearly the estimate in the previous paragraph that XMM-Newton is capable of seeing far past the ‘knee’ in the logN-logS curve is correct.

One finds that $k_{faint}$ in these units works out to be $6.88 \times 10^{4}$. We are now in a position to use equation (A.1) to calculate the ‘standard deviation’ $\sigma$ of the noise in the ensemble of measurements. Equation (A.1) evaluates to

$$\sigma = 1.11 \times C_{max}^{0.65}.$$  

It only remains to select a value for $C_{max}$. Suppose we choose $C_{max} = 47$ counts, which from figure 9 is the detection sensitivity obtainable (under the assumption that the background is flat) at this exposure length using the matched-filter algorithm described in the present paper. $\sigma$ evaluates to ~13 counts, which is several times larger than the standard deviation $\sqrt{10}$ of the Poisson noise. However, the logN-logS model indicates that the total source density at this counts value is 6640 deg$^{-2}$, which is still only 0.16 sources per PSF equivalent area $\Omega_{c}$. The counts value at which one expects 1 source in total per $\Omega_{c}$ is 3.3; using this for $C_{max}$ yields 2.4 counts for $\sigma$ instead, just lower than the Poisson noise. Thus we may conclude that, for XMM-Newton exposures of total duration up to $5 \times 10^{5}$ s in length, the x-ray sky may still be modelled to acceptable accuracy by a flat background with superposed sources. It is clear though that the next generation of x-ray telescopes may not have life so easy.

Finally, a word about background estimation. So far in the present paper it has been assumed that the background contribution is known. Background can be difficult to estimate,
though - to calculate the background one must first excise or avoid the sources, but it is difficult to find sources without first having a knowledge of the background. However, one can conceive of an iterative process in which the detectable sources are gradually detected and excised in parallel with improvements in the background estimate. In the present case that would still leave a population of sources which are too faint to be detected but which are brighter than the confusion level calculated in the preceding paragraph. The sum of these faint sources will bias the background estimate upward. The average counts deg$^{-2}$ contributed by these sources is

\[ \phi = \int_{C_{\text{conf}}}^{C_{\text{det}}} dC \ C \ n(C). \]

For a single-power-law \( n = kC^{-\gamma} \) this gives

\[ \phi = \frac{k}{2 - \gamma} \left( C_{\text{det}}^{2 - \gamma} - C_{\text{conf}}^{2 - \gamma} \right). \]

If we use \( 6.88 \times 10^4 \) for \( k \), 1.7 for \( \gamma \), 47 for \( C_{\text{det}} \) and 3.3 for \( C_{\text{conf}} \), \( \phi \) evaluates to \( 4 \times 10^5 \) counts deg$^{-2}$, or 0.5 counts pixel$^{-1}$ for 4 arcsec square pixels. A similar calculation can of course be performed for any other exposure duration.

Acknowledgements. I would like to thank Silvia Mateos for calculating the source weights in table[1].

References

Alexander, D. M., Bauer, F. E., Brandt, W. N. et al 2003, Ap. J., 126, 539
Bevington, P. R. & Robinson, D. K. 1992, Data Reduction and Error Analysis for the Physical Sciences, 2nd ed. (WCB/McGraw-Hill, Boston)
Boese, F. G. & Doebereiner, S. 2001, A&A 370, 649
Cash, W. 1979, Ap. J., 228, 939
Condon, J. J. 1974, Ap. J., 188, 279
Crudace, R. G., Hasinger, G. R. & Schmitt, J. H. 1987, Astronomy from Large Databases, ed. F. Murtagh & A. Heck, p177.
Damiani, F., Maggio, A., Micela, G. & Sciortino, S. 1997, Ap.J., 483, 350
DePonte, J., & Primini, F. A. 1993, Astronomical Data Analysis Software and Systems II, ASP Conference Series 52, 425
Dobrzycki, A., Jessop, H., Calderwood, T. J. & Harris, D. E. 2000, Bull. Am. Astron. Soc. 32, 1228
Ebeling, H. & Wiedenmann, G. 1993, Phys. Rev. 47, 704
Fay, M. P. & Feuer, E. J. 1997, Statistics in Medicine, 16, 791
Gabriel, C., Denby, M., Fyfe, D. J. et al 2004, Astronomical Data Analysis Software and Systems XIII, ASP Conference Series, 314, 359, ed. F. Ochsenbein, M. Allen & D. Egret.
Gondoin, P., Aschenbach, B. R., Beijersbergen, M. W. et al 1998, Proc. SPIE, 3444, 278
Hasinger, G., Altieri, B., Arnaud, M. et al 2001, A&A 365, L45
Mushotzky, R. F., Cowie, L. L., Barger, A. J. & Arnaud, K. A. 2000, Nature 404, 459
Pierre, M., Valtchanov, I., Altieri, B. et al 2004, J. Cosmol. Astropart. Phys. 9, 1
Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. 2003, Numerical Recipes in Fortran 77: the Art of Scientific Computing, 2nd ed., vol. 1 (Cambridge University Press, Cambridge UK)