Background Determination and Source Extraction for GALEX Data

Ted K. Wyder

December 16, 2004

The GALEX pipeline utilizes the program SExtractor (Bertin & Arnouts 1996) for detection and photometry of sources in the GALEX imaging data. A general description of SExtractor is given in Bertin & Arnouts (1996) as well as in the SExtractor manual[1]. Another useful reference is the "SExtractor for Dummies" manual written by Benne Holwerda[2]. The following sections describe the application of SExtractor to the GALEX data with particular emphasis on the changes we have made regarding the determination of the background and source detection.

1 Background calculation

We have written a program *poissonbg* which computes a background map for each image as well as a corresponding detection threshold image. Similar to the background calculation implemented in SExtractor, the background is determined in a set of large bins with width specified by the user. The width of the background bins should be selected to be larger than most sources while at the same time small enough to account for variations in the background across the image. In the standard SExtractor background calculation, as described in Bertin & Arnouts (1996), the histogram of pixels within each background bin is clipped iteratively at $\pm 3\sigma$ about its median. If the background value changes by less than 20% after the clipping, then the mean of

[1]Available from http://terapix.iap.fr/soft/sextractor/

[2]Available from his web site http://www-int.stsci.edu/holwerda/se.html

the clipped histogram is assumed as the background. Otherwise, the background value is assumed to be 2.5×median − 1.5× mean. These algorithms implicitly assume that the background count rates are high enough such that Gaussian statistics apply.

Typical backgrounds in high Galactic latitude GALEX fields are $\sim 10^{-4}$ and $\sim 10^{-3}$ photons s^{-1} $arcsec^{-2}$ in the FUV and NUV bands, respectively. With such low count rates, the distribution of count rates in each image is typically quite non-Gaussian, even for relatively long exposures. In order to deal with these low backgrounds, *poissonbg* uses a modified clipping algorithm which makes use of the full Poisson distribution. Before beginning the calculation, any regions of the "counts" image where the relative response (i.e. the exposure time multiplied by a normalized flat field) falls below a user-specified fraction of the maximum is masked out and not included in any of the subsequent computations. Then the mean value of the pixels in each background bin is calculated. For the Poisson distribution, the probability of observing greater than or equal to k events for a mean rate x is given by the incomplete gamma function $P_x(k)$. We compute $P_x(k)$ using the routine *gammp* from *Numerical Recipes in C* (Press et al. 1992). The current pipeline clips the values at a probability of 1.35×10^{-3}, a value equivalent to the probability for a Gaussian at a level of 3σ. As in the standard SExtractor calculation, high pixels are iteratively clipped out until the mean value converges or the maximum number of iterations is reached. In tests on simulated data, the mean tended to converge within just a few iterations. Finally, the value of the background in each bin is divided by the average relative response at that location so that the resulting background map is in units of counts/sec/pixel.

Even when iteratively clipping out high pixels, bright sources can still bias the mean. Therefore, similar to the standard SExtractor calculation, a 5×5 median filter is applied to the background grid points. For those background grid points falling outside the detector boundaries, the median is computed from those grid points within the 5×5 box which do fall on the detector. This has the effect of extending the background grid somewhat beyond the actual detector boundary. When the background grid is linearly interpolated to the full image resolution, this then insures that the background will not artificially drop off at the edge of the detector.

The background determined in this way was found to typically be biased too by a few percent due to the wings of objects contaminting the background. To try and fix this problem, *poissonbg* and SExtractor are run twice.

The first run is done as described above while in the second the SExtractor segmentation map used to mask out pixels detected as sources in the first iteration.

2 Determining the detection threshold

Detection of sources in SExtractor is accomplished by finding contiguous groupings of pixels that have values more than some threshold above the background. Our program *poissonbg* computes a map of the detection threshold for each image using the same grid points as for the background calculation. At each grid point, the program determines the value of k in counts/pixel for which the incomplete gamma function $P_x(k)$ falls below the user-specified detection threshold probability. If the average count rate is above 50 counts (in the integrated image), then the data is assumed to be sufficiently well approximated by a Gaussian, and the detection threshold is computed simply as $r \times \sqrt{x}$, where r is the Gaussian σ corresponding to the detection probability and x is the mean background level. Then using the relative response at that grid point, the detection threshold is converted to units of counts/sec/pixel. As for the background itself, the detection threshold map is upsampled to the full resolution of the data image using bilinear interpolation.

For many of the GALEX fields, the detection threshold computed in this way corresponds to a fairly bright surface brightness. SExtractor has the option of first convolving the data with some kernel before detection, allowing us to employ a fainter detection threshold. (This is called "filtering" in the SExtractor manual). An example of what happens when Poisson-distributed data is convolved with a Gaussian kernel is shown in Figure 1. The left-hand panel shows the distribution of pixels in an artificial image with a uniform background while the right-hand panel shows the distribution of pixels in that image after convolution with a Gaussian kernel with FWHM of 3 pixels (or $\sigma = 1.3$). The red line shows the Poisson distribution for the same mean but assuming that the value in each pixel is averaged over an area $4\pi\sigma^2$ pixels. Thus, as expected, convolution with a properly normalized kernel retains the same mean but reduces the variance.

If the convolution in SExtractor is to be turned on, then *poissonbg* computes the detection threshold map as before, except that when determining the number of counts at which the probability falls below the specified level,

Table 1. *Poissonbg* parameters for non-DIS fields

parameter	value	description
sigmaclip	3.0	Probability, expressed in Gaussian sigma, used in clipping out high pixels
threshold (NUV)	2.0	Probability, expressed in Gaussian sigma, used in
threshold (FUV)	2.5	computing the detection threshold map
area_detect	$4\pi\sigma^2$	Effective area of Gaussian convolution kernel with width σ
backsize	128	Size of background grid in pixels
filtersize	5	Size of median filter applied to the background grid
maxiter	10	Maximum number of iterations allowed in background calculation
response_limit	0.2	Low response limit, expressed as a fraction of the maximum

the signal is assumed to be averaged over an area determined by the parameter area_detect. As discussed above this effective area is given by $4\pi\sigma^2$. For all of the non-DIS GALEX data, the data have been convolved with a FWHM=4.5″ kernel, corresponding to a value for area_detect of 20.4 pixels. For the DIS a smaller convolution kernel is used with FWHM=3.5″. The value of the detection threshold as a function of the mean background for the non-DIS value of area_detect and for a probability if 1.34×10^{-3} is shown in Figure 2 as the solid line. The dashed line in the figure shows the corresponding threshold computed assuming Gaussian statistics and the same values for the detection probability and area_detect. For high count rates, the Poisson calculation converges to the Gaussian case, as expected, while for low counts rates the Poisson threshold is larger then the Gaussian value. Approximate background count rates for each of the GALEX imaging surveys are indicated along the bottom of the Figure.

All of the parameters used by *poissonbg* with the current default values are listed in Tables 1 and 2 for non-DIS and DIS data, respectively.

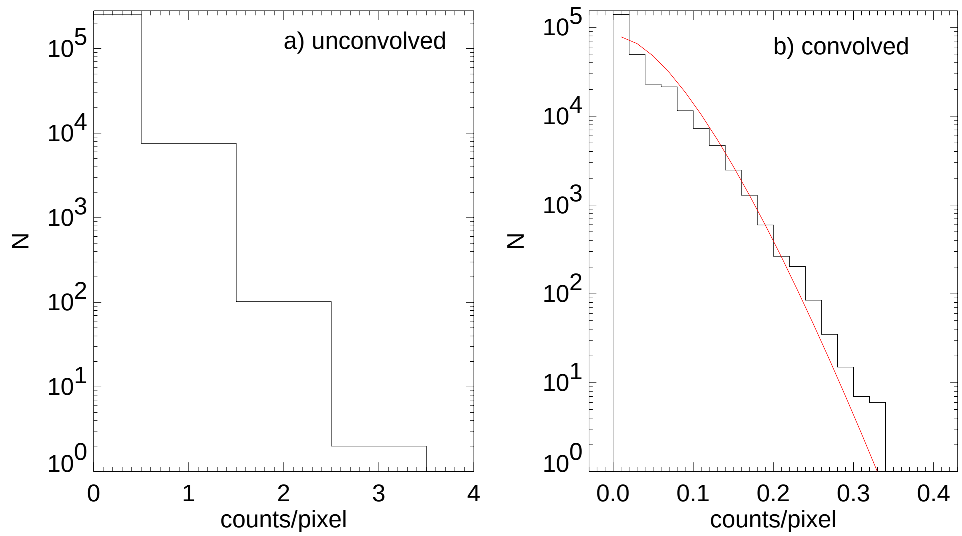

Figure 1 (a) The distribution of pixel values in an artificial 512 × 512 image with a uniform background count rate of 0.03 counts/sec/pixel, a typical count rate in an FUV AIS exposure. (b) The distribution after convolution with a Gaussian kernel with FWHM=3 pixels. The red line is the Poisson distribution calculated from the mean background rate and assuming the signal is averaged over an effective area of $4\pi\sigma^2$.

3 Running SExtractor

When running SExtractor, we use two images as input: one for detection and the second for photometry. For the detection image we use the ratio of the background-subtracted data image to the detection threshold map. All pixels in this ratio map which rise above the detection threshold will have values greater than one and we have therefore set the parameters in SExtractor so that it will consider all of these pixels as possible detections. The image used for photometry is simply the background subtracted data image. A list of the SExtractor parameters is given Table 2 for one of the MIS fields. Except for the GAIN parameter, which is set equal to the exposure time, and ANALYSIS_THRESH, the remaining parameters are the same for all data sets.

SExtractor makes a few different measurements of the total flux of a

Table 2. *Poissonbg* parameters for DIS fields

parameter	value
sigmaclip	3.0
threshold (NUV)	3.0
threshold (FUV)	3.0
area_detect	$4\pi\sigma^2$
backsize	64
filtersize	3
maxiter	10
response_limit	0.2

source. See the SExtractor manual or Bertin & Arnouts (1996) for more detailed descriptions. MAG_APER corresponds to the flux measured through circular apertures with radii given by the value of the PHOT_APERTURES parameter (currently 5 pixels). The sum over all the pixels rising above the detection threshold is stored as the isophotal magnitude MAG_ISO. Sextractor makes an attempt to account for the light lost outside of the isophotal aperture by assuming that each object can be approximated by an axisymmetric Gaussian and finding the fraction of light falling below the detection threshold. This is output as the corrected isophotal magnitude MAG_ISOCOR. The MAG_AUTO magnitude corresponds to the flux measured within an elliptical aperture, as described in Kron (1980). The position angle and axis ratio of the elliptical aperture are determined from the second order moments of the object's profile. The size of the ellipse is scaled by kr_1 where r_1 is the Kron radius for that object and k is a constant set by the first number of the PHOT_AUTOPARAMS parameter. The Kron radius r_1 is defined as the first moment of the object's radial profile. The semi-major axis of the NUV MAG_AUTO ellipse in pixels can be calculated as NUV_KRON_RADIUS $\times$ NUV_A_IMAGE with a similar expression for the FUV. The major to minor axis ratio is given by the NUV_ELONGATION column whereas the position angle in degrees east of north is given by NUV_THETA_J2000.

Finally, SExtractor outputs a magnitude called MAG_BEST. This de-

Table 3. SExtractor Parameters

parameter	non-DIS	DIS
ANALYSIS_THRESH	0.0011737	4.4774e-05
BACKPHOTO_TYPE	GLOBAL	GLOBAL
BACK_TYPE	MANUAL	MANUAL
BACK_VALUE	0	0
CLEAN	Y	Y
CLEAN_PARAM	1	1
DEBLEND_MINCONT	0.005	5e-05
DEBLEND_NTHRESH	32	32
DETECT_MINAREA	10	8
DETECT_THRESH	1	1
DETECT_TYPE	CCD	CCD
FILTER	Y	Y
FILTER_NAME	gauss_30_7x7.conv	gauss_23_7x7.conv
GAIN	1704	44668.2
MASK_TYPE	CORRECT	CORRECT
MEMORY_BUFSIZE	1024	1024
MEMORY_OBJSTACK	2000	2000
MEMORY_PIXSTACK	1500000	1500000
PHOT_APERTURES	2,3,5,8,12,17,23	2,3,5,8,12,17,23
PHOT_AUTOPARAMS	2.5,3.5	2.5,3.5
PHOT_FLUXFRAC	0.2,0.5,0.8,0.9,0.95	0.2,0.5,0.8,0.9,0.95
PIXEL_SCALE	1.5	1.5
SATUR_LEVEL	999999999	999999999
SEEING_FWHM (FUV)	5	5
SEEING_FWHM (NUV)	5.5	5.5
STARNNW_NAME	default.nnw	default.nnw
THRESH_TYPE	ABSOLUTE	ABSOLUTE
WEIGHT_GAIN	N	N
WEIGHT_IMAGE	NONE	NONE
WEIGHT_TYPE	NONE	NONE

faults to the value of the Kron magnitude if neighboring objects are estimated to bias the magnitude by less than 0.1. Otherwise, the corrected isophotal magnitude is used. The FUV_MAG and NUV_MAG columns in the GALEX catalogs by default use the MAG_AUTO measurements.

Despite its name, MAG_BEST is not necessarily the best choice for most applications since the measurement for all objects are not made in a consistent way. For resolved sources, the MAG_AUTO measurement is probably the most appropriate choice while MAG_APER would be better suited for measurements of unresolved sources.

References

Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393

Kron, R. G. 1980, ApJS, 43, 305

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical Recipes in C*, (Cambridge: Cambridge University Press)

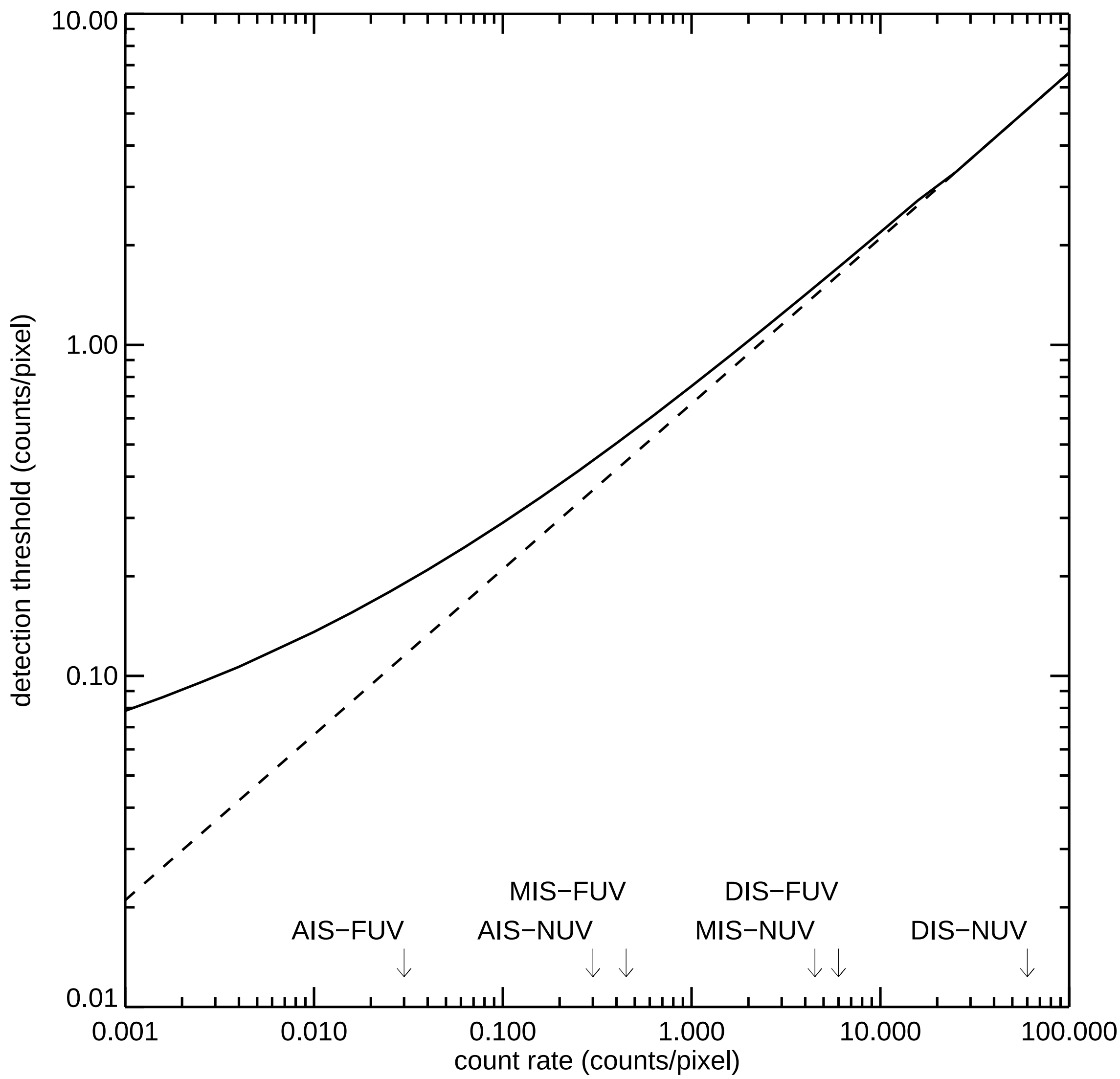

Figure 2 The Poisson detection threshold as a function of the background count rate in counts/pixel is plotted as the solid line for a detection probability of 1.34×10^{-3} (equivalent to 3σ for a Gaussian). The dashed line indicates the detection threshold assuming Gaussian statistics. Both curves assume that the data are being convoled with a Gaussian with a FWHM of 3 pixels.