DEPARTMENT OF SPACE RESEARCH
BIRMINGHAM UNIVERSITY
Asterix Data Analysis
ASTERIX User Note 004

USER/004

David J Allan
1 March 1992

Source Searching and Parameterisation

Contents

1 Introduction

This document describes the ASTERIX source searching and parameterisation programs, PSS and EVPSS, and the auxiliary programs to make use of their output.

After an overview of the ASTERIX source searching system, tutorials for both PSS and EVPSS are followed by reference sections where the methods used by each program are presented in depth. Section 6 describes the various ASTERIX programs which use the output of PSS and EVPSS directly, and Section 7 a selection of common processing techniques with examples. Details of numerical algorithms can be found in the Appendices.

The programs documented in this manual are all ASTERIX applications and as such can be run from both DCL and ICL. They are also fully described in the ASTERIX on-line help system, which can be accessed using,

```
> ASTHELP <application_name>
```

USER_004 comes with a number of figures – a command procedure to generate the figures can be created using the following DCL command,

```
$ @AST_DOCS:USER_004_FIGS
Are you using a Postscript Laser printer (Y/N) /N/
Printer queue name /SYS_LASER/ >
Command file written to FIGS.COM
$ @FIGS
Drawing Figure 1...
...
Drawing Figure 3...
```

It is worth checking that the settings for the `PRINT` command within this procedure are correct for your site before executing it.

The remainder of this manual assumes familiarity with the ADAM parameter system, a full description of which can be found in SG/4.

2 Overview

2.1 The PSF System

In most imaging instruments, a combination of telescope and detector spatial response combines to give several detector resolution elements across a point response function. In X-ray instruments, this point response is usually well calibrated before launch, and reasonably stable in flight. Due to the lack of atmosphere, departure from the point response is usually due to the intrinsic surface brightness distribution of non-point sources.

ASTERIX represents this *a priori* information about the instrumental point response using the PSF system (see Appendix A), providing the mechanism by which instrument independent applications may be written. The model used by the system is two dimensional and can handle variation in both energy and time.

This relatively simple situation is in contrast to many forms of ground based imaging, where atmospheric distortions degrade the point response over timescales (until very recently) too short to be corrected. In addition, most optical/infrared detectors can saturate for bright sources, whereas X-ray detectors (which are usually photon counting devices) simply suffer higher deadtime losses. These problems dictate the use of techniques other than that of comparing the observed source surface brightness distribution with a pre-computed model, limiting the usefulness of the PSF system with these data.

2.2 Background Modelling

The PSS (Point Source Searching) program was written for use in the ROSAT project, originally on WFC data. The anticipated nature of the data from the WFC drove the decision to separate the determination of the background (which was expected to contain large gradients) from the detection of the sources (which were expected to be almost exclusively unresolved). This separation is somewhat artificial, as the determination of an accurate background often involves detection of sources, which in turn needs a background. However, the methods of producing backgrounds depend greatly on the nature of the observed objects. While a local background estimation may be sufficient for flat backgrounds containing few point sources, the method will break down when presented with extended sources superimposed on a galaxy cluster emission containing large spatial gradients. The atomic nature of ASTERIX applications enables backgrounds to be derived or fitted in a number of ways by combining simple programs.

2.3 Source Searching

The two forms of ASTERIX data, binned and event datasets, are handled by PSS and EVPSS respectively. They have several features in common which are worth mentioning,

- They share many of the user interface prompts, such as those used for controlling the subset of the input dataset, names of output files, access to point spread functions and others.

- An EXPERT logical keyword is provided with both programs. Specifying this parameter false ensures the absolute minimum of user interaction – usually supplying only the inputs is sufficient for default operation.
- Both have a diagnostic mode, which can be used either to produce evidence of a program's malfunction for the author's benefit, or as a useful supply of extra data for the expert user.
- They share output file format, both using source search results files, or SSDS. These data can be printed, displayed graphically or exported to other software packages.

As binning inevitably involves loss of information, PSS can only ever be regarded as a first pass parameterisation program. Its function is primarily to allow investigation of the properties of sources in an image quickly.

EVPSS on the other hand is intended as a second pass parameterisation program. Current hardware limits the use of this program to relatively small areas of data – typically circles around prospective sources.

2.4 Post Search Processing

Both PSS and EVPSS share the same output format. The prerequisites of such a format are that its structure be sufficiently rich to represent the information which are required to be stored, and that it is easy to export the data to other formats or feed it into other applications. The ASTERIX source search dataset (SSDS) is an HDS file format which permits storage of,

- Bookkeeping information about the parameters used in the various preceding applications used to create the dataset.
- Field information about each source. Consists of the data value or values, errors both symmetric and asymmetric, at multiple levels of confidence, along with units information and other ancillary information. Although certain field names are reserved, extra fields can be written by new applications such as exposure correction program and database software.
- Bulk data relating to the whole dataset, eg. sensitivity maps.

ASTERIX applications exist to print ascii reports of SSDS, export to common database formats and manipulate results graphically.

3 PSS Tutorial

3.1 Definitions

PSS performs a fit of its primary input to a model comprising background plus point spread function (**psf**). At a *fixed* image position PSS finds that scale factor which multiplies the psf to optimise the match of the input data to the model – this factor is the PSS **flux**. The units of this quantity are the same as the units of the input data, whatever they may be.

PSS can process the input data using two statistics, the Cash maximum likelihood statistic, or a correlation assuming gaussian errors. The former is intended for counts images where the errors are purely Poissonian, whereas the latter can be used on any data where the errors *on each pixel* are gaussian.

The PSS **significance** is a measure of the difference in the quality of the fit at the optimum flux compared to that at zero flux, expressed in normal deviates ("sigmas").

3.2 Modes

The results of this fitting process can be used in a number of ways to give different output products. The modes are known by name,

- SEARCH mode. All or part of the input image is searched for peaks above a user specified significance threshold. The parameters describing each source are output.
- PARAM mode. PSS performs the fit at fixed points on the image, and returns the optimum fit parameters at each point.
- UPLIM mode. Similar to PARAM mode, but the fit parameters returned are the upper limits at a user definable level of confidence.
- UPMAP mode. PSS maps the value of the upper limit flux over all or part of the image.

This tutorial section will examine an example session using each mode in the order above. Test data for use in the tutorials can be found in the logical directory `AST_DEMO` – make your own copies using,

```
> copy ast_demo:pss_demo.sdf my_pss_demo
```

3.3 SEARCH mode

When invoking PSS, three parameters have command line positions defined – these are the same for all modes.

```
> PSS <input_image> <background> <results>
```

If the information present in your dataset is sufficient for PSS to work with (as it is in the demo file), then PSS could be run using the following,

```
> PSS MY_PSS_DEMO 1.01 MY_SOURCES \
```

using the ADAM command line token `\` to suppress all prompts for which defaults exist. However, for the purposes of a useful tutorial, the prompts and their uses will be explained.

Start with simply the input image name,

```
> PSS MY_PSS_DEMO
PSS Version 1.6-1
EXPERT - Expert mode /NO/ >
```

and PSS responds with its version number followed by the `EXPERT` parameter prompt. Accepting `NOEXPERT` means PSS will minimise user interaction by assuming sensible defaults for control parameters (usually chosen to minimise processing time). Specifying `EXPERT` true will result in between five and ten extra prompts depending on the PSS mode. Accepting the defaults for all these prompts will give exactly the same behaviour as non-expert mode.

Next PSS requires the source model to be specified,

```
PSF system options :

  ANAL          EXOLE         PWFC          TABULAR       WFC
  XRT_HRI       XRT_PSPC

  POLAR(psf,rbin[,abin])      RECT(psf,xbin[,ybin])

PSF - Choose PSF to use for source model /'WFC'/ >
  Filter Lexan/C/B4C
```

The ASTERIX psf system (see Appendix A) allows the PSS source model to be manipulated in an instrument independent fashion. For datasets produced by an instrument interface a default for `PSF` parameter should be offered. The default in this case is `WFC` – the demonstration is a simulated WFC survey image. On accepting this default the WFC psf routine displays the WFC filter being used.

In search mode PSS compares the input data and model in a box which is passed over the image. The size of this box is the principle factor controlling both the accuracy of the results, and the speed with which they are produced.

```
Using a psf box of radius 4 pixels
Using a psf constant across the field
```

In non-expert mode PSS chooses a box size which encloses a minimum of 68% of the psf energy. This is a reasonable compromise between accuracy and speed (see Section 4.4). In expert mode, the user has control over this box size. The second line of text indicates that PSS will access the psf only once for the whole field. The alternative, a fully varying psf, is computationally more expensive but required for some detectors (eg. ROSAT PSPC). PSS only chooses a varying psf as default if the user supplies a psf model specification (see Section A.4), thus stating explicitly that the psf varies.

In SEARCH mode a subset of the image may be searched. This may be useful in avoiding areas of bad background subtraction, or simply to cut down the processing time. PSS informs the user of the extent of the image and prompts for a slice.

```
X_CORR axis range is from 29.5 to -29.5 arcmin
Y_CORR axis range is from -29.5 to 29.5 arcmin
SLICE - Section of dataset to search /'*:*,*:*'/ >
```

The default offered is the entire image. To select the central 40 arcminutes box the reponse to this prompt would have been `20:-20,-20:20` (note that the range values must confirm to the direction of increase of the axis values). See Ref 5,"*User Interface,Lists and Ranges*" for a description of the ASTERIX range syntax.

The `SOPT` prompt selects the fitting statistic, the default is always `CASH`,

```
SOPT - Statistic option (CASH,GAUSSIAN) /'CASH'/ >
```

PSS now requires a background model. There are two options, either a file name or a constant value. The latter is a suitable choice here as the tutorial dataset was created using a flat background. The units for such a value are image data units (counts in this case) per pixel.

```
BGND - Background model > 1.01
Using background value 1.01 for entire image...
```

PSS informs you that it is using a constant background.

In SEARCH mode PSS constructs a map of the statistic covering the area selected using the `SLICE` parameter. In non-expert mode, the point spacing of this map is the same as the pixel spacing in the input image.

```
First pass - grid spacing 1 pixels
MAP - Significance map /!/ > sig_map
```

Constructing the significance map is usually the most time consuming part of PSS. Once it is complete, the option to output the map to an external file is given – enter a file name or type `RETURN` to ignore the prompt. If output, the map inherits all the ancillary information associated with the input dataset, enabling significance contours to be overlaid on the original data, for example. The map becomes the default ASTERIX graphics object, and so could be displayed on an already opened device with,

```
> DRAW \
```

The range of the significance map is stated and a threshold for source detection selected.

```
Significance varies from 0 to 26.53978
SIGMIN - Significance threshold /5/ >
```

Once `SIGMIN` is supplied PSS searches for peaks in the map above this threshold – a peak being defined as a point higher than all surrounding points. PSS lists all peaks higher than 80% of the supplied threshold. This 20% factor is allowed on the first pass as the initial search grid will not find the highest part of a source peak unless the position of best source fit lies exactly on a grid point. A second threshold value may be supplied after the first, separated by a comma.

```
SIGMIN - Significance threshold /5/ > 3,5
```

This mechanism allows less peaks below the first threshold to progress to fitting (see below) and possibly extension testing in expert mode, but the second threshold is applied before sources are output.

The list of source candidates is printed,

```
Src    X        Y     Signif

  1   0.38     0.38   26.540
  2  16.62     9.63    9.789
```

The table shows source number, image position and peak significance. PSS now performs a grid search around each source candidate, but with the grid centred on the best source position. This process is repeated a few times to optimise the source position.

```
Second pass
Symmetric flux errors at 68% confidence, and position errors at 90%
```

PSS now finds the errors on the source parameters. The error confidence levels are chosen automatically in non-expert mode. Once the errors have been found PSS prints another source list with the user's significance threshold applied firmly, sorted into increasing Right Ascension.

```
Src      RA          DEC          X         Y      Signif    Flux

  1  00 01 07.7  +00 10 04    16.93     10.07     9.892    95.579
  2  23 59 59.8  -00 00 01    -0.04     -0.01    27.463   358.160

OUT - Source search results file (! for none) /@SRCLIST/ > my_sources
```

The option is also given to output these results to a source search results file (SSDS). This is an HDS file which ASTERIX software uses to store such information. If the printed data is sufficient, reply with `!` to this prompt. Note however, that the SSDS contains all error information and provides a simple mechanism for passing source positions around the ASTERIX system.

3.4 PARAM mode

Parameterise mode can be used to test the value of the optimum flux (that psf scale factor which best matches the data) at any point on the image. To invoke PSS in PARAM mode, specify the `MODE` keyword on the command line,

```
> PSS my_pss_demo 1.01 my_par_results mode=param
```

The `EXPERT` and `PSF` prompts will appear as in the previous tutorial. The first different prompt is `PLIST`, which controls where PSS is to find source positions,

```
PLIST - Source of RA and DECs > MY_SOURCES
2 positions read from SSDS file MY_SOURCES...
```

PSS can get positions from a number of sources (see below), one of which is a SSDS. If any of the positions read do not lie on the image, PSS will warn the user but continue.

The flux and significance are now evaluated at each test position, and the results printed.

```
Src      RA          DEC          X         Y      Signif    Flux

  1  00 01 07.7  +00 10 04    16.93     10.07    10.101    96.441
  2  23 59 59.8  +00 00 01    -0.04      0.01    27.499   354.031
```

Note that despite reading in positions directly from the previous PSS run, the values of flux and significance are different from those final results, by about 1% in flux. This is due to a slight difference in algorithm in the two modes. A time critical part of the error determination in SEARCH mode uses an approximation for the psf which can deviate from the true psf shape by a few percent. These discrepancies should be removed in the near future.

PSS can read in positions from various sources – the source search results file (SSDS) mode is shown above. Positions can also be read from ascii text files. Each line should contain an RA and DEC specification separated by one or more spaces. The valid formats for such specifications are detailed Ref 5,"*User Interface,RA DEC Formats*" topic.

The last alternative is to enter celestial positions manually. This can be achieved in two ways. The first is using `TT:` as a file name. This causes PSS to read lines from `SYS$INPUT` just as if they were read from a text file (and so use `CTRL-Z` to terminate input),

```
PLIST - Source of RA and DECs > TT:
12h43m12.1s -34d56m
190.2       -35.012
2 positions read from ascii file TT:...
```

The alternative is to supply `TERMINAL` in response to the `PLIST` parameter. The user is then repeatedly prompted for RA and DEC.

```
PLIST - Source of RA and DECs > TERMINAL
Enter RA and DEC at prompts, ! to terminate
RA - Right ascension /!/ > 12h43m12.1s
DEC - Declination > -34d56m
RA - Right ascension /!/ > 190.2
DEC - Declination > -35.012
RA - Right ascension /!/ >
2 positions read from TERMINAL...
```

3.5 UPLIM mode

Upper limit mode is identical to PARAM in invocation, with the exception of the `MODE` parameter,

```
> PSS my_pss_demo 1.01 my_uplims mode=uplim
  ...
PLIST - File of RA and DECs > my_sources
2 positions read from SSDS file MY_SOURCES...
Upper limits at 68% confidence
```

In upper limit mode, PSS first finds the optimum flux then finds that positive increment in flux which degrades the fit by a change in χ^2 equivalent to the confidence level requested. The sum of the optimum flux plus this increment gives the upper limit flux. In non-expert mode PSS automatically chooses a confidence level of 68%.

PSS prints out the list of input source positions with the upper flux limit at each point.

```
Src      RA          DEC          X         Y        Flux

  1  00 01 07.7  +00 10 00    16.88     10.00  <  112.426
  2  23 59 59.8  +00 00 04    -0.02      0.07  <  375.872
```

3.6 UPMAP mode

This mode causes the upper limit flux to mapped on an area of the image. Again in non-expert mode PSS automatically chooses a confidence level of 68%. Rather than a significance map, the user is prompted for an upper limits map.

```
MAP - Upper limit flux map /!/ > up_map_68
```

PSS returns control to the environment immediately after creating this map.

4 PSS Reference

This section contains a brief outline of PSS's processing method followed by a number of topics which describe the full functionality of the program.

At the core of PSS are a set of routines which evaluate a statistic comparing model and data. These all provide the same facility to the rest of the PSS application, and are applied in different ways at different stages of processing or in different PSS modes. The facility is the evaluation of the significance surface at a point (x, y) in image coordinates. The mechanism is as follows (See Figure 1 in Appendix D),

- The input image pixel (i, j) containing the point (x, y) is located.
- This pixel defines the centre of the psf box. The size of the box is controlled by the `PSFPIX` parameter is expert mode, otherwise is chosen to include 68% of the psf energy.
- The area of overlap of the psf box and the image grid is found, taking due account of the edges of the latter. Note that any user defined image slice does *not* affect the overlap of the psf box outside the slice.
- The psf values in the psf box are found. If the point (x, y) lies on the exact centre of pixel (i, j), and the parameter `PSFCON` is true, then the psf values are simply retrieved from a storage array. If the former is true, but not the latter, then the psf is evaluated using a call to the psf system. The last alternative, `PSFCON` true and (x, y) lying in the pixel but not at its centre, causes a resampling of the psf by a vector $(x - x_i, y - y_j)$ where (x_i, y_j) is the image coordinate of the centre of pixel (i, j).
- The statistic is evaluated as a summation over the area of overlap of the image and the psf box. The exact algorithms used are specified in Appendix B. Image data points with bad quality are ignored in these summations. Because the significance (x, y) is a summation over the surrounding area, it is possible to produce a significance even if the pixel (i, j) has bad quality. If all the pixels in the overlap region have bad quality, the the significance is set to zero.
- If only a single grid position is called for, the routine sets the optimum flux and any other relevant information in a storage area.

Each routine is actually capable of finding the significance on a grid of image coordinate positions. In this case the psf box is passed over the image one row after another. As can be seen from the method above, this grid need not be centred on the image pixels but savings in psf access time are achieved if this is the case.

The SEARCH mode algorithm uses a statistic routine to calculate the significance map over the entire image region selected by the user. PSS then puts down a small grid around each candidate source with the grid chosen in such a way that the optimum source position lies on a grid vertex. The grid used is just large enough to perform centroiding to derive the next guess at the optimum source location.

PARAM mode invokes a statistic routine to evaluate significance at each test position lying on the image. Only a single grid position is used.

Both UPLIM and UPMAP modes use special versions of the statistic routines which perform the additional task of finding a flux upper limit after the optimum flux has been found.

4.1 Global Properties

There are 2 parameters which control how the whole PSS application appears to the user, without being specific to the source parameterisation process. These are `EXPERT` and `DEV`.

The purpose of `EXPERT` has already been noted, but is repeated here for completeness. `EXPERT` is a logical keyword which is prompted for unless specified on the command line. Its default value is propagated from one invocation of PSS to the next. It toggles the prompting of those parameters which are essentially fine tuning the operation of PSS, ie. those for which PSS can make sensible guesses. Obviously, such things as output file names do not fall within this category.

The `DEV` parameter is a standard ASTERIX parameter, described fully in the online help (Ref 5,"*User Interface,Ascii Output*"). Only data pertinent to the output data products can be redirected by this parameter. Thus, the psf enclosed energy tables are always sent to the user terminal and not to whatever device has been selected using `DEV`. In PSS this parameter has a hidden default of `TERMINAL`, which must be overridden either explicitly on the command line or by forcing parameter prompting using the `PROMPT` keyword.

4.2 Input data

PSS is written to take advantage of the many pieces of information present in an ASTERIX dataset. These include knowledge of the instrument used, the spatial extent and celestial orientation of the image, and presence of bad or missing data.

The ASTERIX binned image format (Ref 3) is a superset of the Starlink NDF structure (Ref 2). The extensions provide the astronomical content lacking in the latter which enable applications to make intelligent processing decisions.

A dataset conforming only to the NDF definitions can be searched quite successfully by PSS, with certain qualifications. The most fundamental assumption is that the image data represent a tangent plane projection of the sky. The ASTERIX standard then defines the axis values as offsets in angular units from the fiducial point of the dataset. In order to interpret the axis units correctly ASTERIX recognises all common angular units and their abbreviations. An additional assumption is made by the psf system that the spatial axes are linear with respect to pixel number. This assumption is necessary as the production of the psf is required to be independent of the binning characteristics of the dataset.

Both the `CASH` and `GAUSSIAN` statistics expect the input data *not* to be background subtracted. This was not the case with older versions of PSS, and so the program can still accept these data. As long as a background subtracted image has the `BGND_SUBTRACTED` flag set in the `MORE.ASTERIX.PROCESSING` structure, then PSS will recognise the fact and add the background model back to create the original data without further ado.

4.3 Background Model

The background model can be supplied in two forms using the `BGND` parameter – either a single numeric value, or a filename. If the latter option is selected then the background image must have dimensions identical to the input image.

4.4 Point Spread Function

Expert mode provides several ways of controlling how PSS uses the source part of its model, the point spread function. The psf may be evaluated only once over the image, or at every point of interest.

```
PSFCON - Assume constant PSF across field /YES/ >
```

The default offered for this parameter is `YES` unless the user has specified a psf model at the `PSF` prompt.

The radius of PSS box used in the statistics is now selected. The case where the psf is constant is displayed thus:

```
Energy fraction   Radius (pixels)

     50%                2.1
     68%                3.9
     90%               15.2
     95%               43.1

PSFPIX - Radius of psf box in pixels /4/ >
```

The default offered is 68% enclosed energy radius. The table is quite different when the psf is allowed to vary. The example below is from a ROSAT XRT image with the PSPC detector, with arcminute bins.

```
Energy fraction    Off-axis angle
                    0    30    90   arcmin

      50%          0.3   1.2   8.1   pixels
      68%          0.4   1.6  10.4   pixels
      90%          0.6   2.3  14.8   pixels
      95%          0.7   2.6  17.4   pixels

PSFPIX - Radius of PSF box in pixels /1,2,10/ >
```

The reason for using a varying psf in this instrument is obvious! PSS offers a default of the 68% enclosed energy radii at the three off-axis angles quoted. Selecting this default means that PSS will vary the size of the box across the field. The box size is found by linear interpolation from the three radii given – bound by the lower and upper values. Alternative responses to this prompt could be,

```
PSFPIX > 1,2,13                ; Use 50% enclosed energy
PSFPIX > 5                     ; Use a constant radius of 5 pixels
```

Any three values may be specified, as long as they increase left to right. It is best to use radii which enclose the same fraction of psf energy with radius, as the fraction of the psf used affects the sensitivity of the parameterisation.

It should be noted that when a varying psf is used in SEARCH mode, PSS is performing a call to the psf system on every pixel in the input image slice. For this to be a practicable method for large image areas the user is advised to use the psf model facility (Appendix A.4).

4.5 Fitting Statistics

4.6 Background Rescaling

PSS works assumes the background model to be correct. For some images however, the model may not be of good quality – it may have been obtained using a simple function fitting approximation for example. If the Cash statistic is being used, PSS will offer the following option in expert mode.

```
RESCALE - Re-scale background estimate /NO/ > Y
```

Selecting this option invokes a derivative of the Cash statistic described in Section B.1. Both the flux and background scaling factor are optimised at every point using this statistic (ie. at every point within the slice in SEARCH mode, or every test position in PARAM mode).

An initial guess for the scale factor is requested.

```
ISCALE - Initial guess for background scale parameter /1/ >
```

Background rescaling is useful where the background is known to be roughly right but needs fine tuning. It will not correct background models containing unrepresentative gradients on spatial scales comparable with the PSS box size.

Needless to say, given the complexity of the algorithm, background rescaling is significantly slower than the standard Cash statistic. The exact time penalty depends how well the background model supplied estimates the true background .

4.7 Oversampling

When evaluating the significance map on the first pass of SEARCH mode its possible to miss the peak of a source because the optimum source position lies between grid points. Although this is not a problem for any image where the psf extends over more than a few pixels, there are exceptions. For example, with a ROSAT PSPC image binned at one arcminute resolution, an on-axis 8σ source centred on a corner of a pixel may have its significance reduced by as much as 3σ if the grid points are located on pixel centres.

One solution to this problem is to lower the the significance threshold on the first iteration. However, this obviously will result in more spurious peaks if the threshold is lowered too far.

A practical alternative for small image areas is to set the value of the `SAMPLE` parameter to greater than unity, which multiplies the density of grid points on the first iteration by that amount on *both* axes. While alleviating the problem above, it has the effect of prolonging the first iteration by the *square* of its value.

4.8 Source Confusion

A limitation of the grid method is that PSS cannot resolve two sources unless their respective peaks are separated by at least one pixel in the significance map. If two sources overlap to some degree then it may be possible for PSS to resolve them simply by increasing the grid resolution using the `SAMPLE` parameter. However, this method assumes that there is a saddle point on the significance surface to be resolved. If this is not the case, then the image should be binned more finely, or a photometry package such as DAOPHOT used instead.

Even if two blended sources can be resolved, the quality of PSS's results for fluxes are likely to be poor. This is simply a consequence of the model used, which assumes that the input image surface brightness can be represented by a background plus *single* point source at every point – this obviously breaks down where sources overlap, where simultaneous knowledge of the contributions from two sources is required. Unfortunately, the process of fitting two sources simultaneously is considerably more difficult than twice that of the single source case, and will have to wait for a future version of PSS to be implemented.

Another aspect of the source confusion problem arises in connection with spurious sources. These occur most often when the model used is deficient in some way, ie. poor background modelling or a psf which does not match the data well, and are often manifested on the wings of bright sources (near which of course a background subtraction program is likely to have more difficulty).

PSS has a mechanism for deciding whether a source is too close to another to be real. After the significance threshold is supplied, PSS finds all the positions of all the peaks in the significance map passing the threshold. These are ranked in descending order of the distance of their nearest neighbour. Source pairs whose separations are closer than the 50hidden `MULREJ` parameter (default 1) are considered to be one source. In this case the less significant (and less bright and/or psf-like) is discarded.

In certain circumstances it is possible that this mechanism appears to be merging peaks separated by small amounts. In this case the value of `MULREJ` should be reduced to make PSS realise they are separate. The opposite problem of multiple detection of a single source can be controlled in the same way, but by increasing the value of `MULREJ`. The latter problem can cause great difficulty if the source subtraction operation is used (Section 4.13).

4.9 Source Extension

In expert SEARCH mode, PSS can test for source extension. The controlling parameter is `EXTEN`,

```
EXTEN - Fit for extension measure /NO/ >
```

If extension testing is enabled, then *after* successfully detecting sources on the first pass, PSS will convolve the point spread function with gaussians of various FWHM and store that width which maximises the significance of the detection.

Note that this method is suitable only for moderately extended sources – the limit is roughly a FWHM more than twice the size of the psf itself. This problem has several causes,

- As the source becomes more extended, it becomes susceptible to removal by background subtraction software.
- As the convolution width becomes large, the significance tends to drop as less of the psf being taken into account. The source can therefore be missed in the first pass. This effect can be be alleviated to a certain extent by two stage thresholding (as decsribed in the tutorial), letting weaker peaks through to fitting and extension testing before applying a more suitable threshold for output.

The solution to these problems is to use the ASTERIX application CREPSF to create a tabular psf. This can be convolved with various functions and used in PSS using the `TABULAR` option at the `PSF` prompt. An example of how to do this can be found in Section 6.8.

Selecting extension testing causes two extra fields to be written to the output SSDS, the extension in arcminutes with error, EXTEN, and the significance of the extension, EXTENSIG. The latter value expresses the probability that the best fit extension is different from zero.

4.10 Specifying Confidence Levels

In SEARCH mode PSS requires confidence levels for the errors on the source parameters, and in UPLIM and UPMAP an upper limit confidence level is required. In non-expert mode these values are chosen automatically, but in expert mode the user can specify them.

Levels can be specified either in percentages (the default) or in normal deviates or sigmas.

```
> 68 %                                 ; 68% confidence
> 68                                   ; Ditto, as % is assumed
> 3 sigma                              ; In normal deviates
> 5 s                                  ; Any abbreviation will do
```

In SEARCH mode multiple confidence levels may be specified. The syntax is as above, with values separated by commas – the two notations can be mixed, for example,

```
> 68 %, 3 sigma, 99.9999%
```

4.11 Source Parameter Errors

In SEARCH mode PSS calculates errors on source flux and position. These errors can be calculated at up to 3 confidence levels for both quantities. The fitting software used by PSS (written by Dick Willingale at Leicester University) supplies both lower and upper bounds on the source parameters. By default PSS returns the average of the two numbers for a given error – this is the 'symmetric' error mode. By specifying the `ASYMMETRIC` logical keyword true, PSS will output both lower and upper bounds to the output file.

Expert mode also allows the number and level of confidence levels to be specified using the `FERL` and `PERL` parameters.

4.12 Restarting PSS

Some of the extended options available in PSS can result in large amounts of cpu use, and are best run in batch. This is hardly ideal for the investigation of the image, as different values for control parameters may be required to be run (eg. search thresholds).

PSS provides the `RESTART` parameter for this eventuality. Provided the user has a valid significance map, PSS can be restarted with this map and the program execution continues from the point where the range of the significance map is written to the terminal. `RESTART` is a hidden parameter, so must either be specified on the command line or prompted for after using `PROMPT`. The following example illustrates the method using the demonstration file,

```
> PSS MY_PSS_DEMO 1.01 MY_SRC ~        ; Create map on first search
    MAP=SIG_MAP \
> PSS MY_PSS_DEMO 1.01 MY_SRC2 ~       ; Try with lower threshold
    EXPERT RESTART MAP=SIG_MAP
    SIGMIN=3.5
```

4.13 Source Subtraction

The last prompt offered by PSS in expert mode is `SSUB`. If a response is given then PSS creates a copy of the input dataset with the sources found in either SEARCH or PARAM modes subtracted. This is performed by simply subtracting the psf scaled by the flux from the input data. It is therefore possible, especially for weak backgrounds, that negative data values will be produced. It would then be necessary to ignore these points using one of the QUALITY applications before source searching was performed on this dataset.

If sources are required only to be ignored, a better technique is to use the SSZAP procedure which is described in Section 6.6. The most suitable use of PSS source subtraction is in removing sources to improve backgrounds. The psf model is unlikely to be sufficiently good that no artifacts remain in the image after subtraction, especially of this source is extended in some non-trvial way.

4.14 Diagnostic Mode

4.15 Limitations

The limitations of PSS arise from deficiencies in either the algorithms used or in the background or source models. Those in the former class have been mentioned in the preceding sections, but can be summarised as source confusion handling and treatment of extended sources. Strategies for working around these limitations have also been discussed.

Bgnd Sensitivity to 20 count source as function of bgnd error

The

Positions on radial symmetry Fluxes only as good as the model

5 EVPSS Reference

EVPSS operates on and is restricted to the use of a purely ASTERIX data format, the event dataset (Ref 3). [1]

[1] In the event that this program might be required to be used on other photon list data, an enterprising user could quite easily import the information into an event dataset using the `LIST` subroutine interface provided in the `ASTLIB` library.

6 Support Programs

Each subsection below gives a brief description of the ASTERIX applications which can make use of SSDS. The parameters used by each program with a one or two line comment are shown in a table for each.

6.1 SSDUMP

INP	Input SSDS containing source positions Takes default from current ASTERIX SSDS
DEV	Ascii output device
HEADER	Output a table header?
HMS	Celestial coordinates in sexigessimal?
ERRORS	Output field error information?

SSDUMP outputs the contents of a source search results file in ascii form. The output device can be controlled using the standard ASTERIX `DEV` parameter (Ref 5,"*User Interface,Ascii Output*"), which allows the console, printer or file to be chosen. For example,

```
> SSDUMP my_sources PRINTER \
SSDUMP Version 1.6-0
Output to PRINTER
```

By default SSDUMP generates a header listing the files searched and the important parameters for each column in the listing. This can be turned off by specifying the `HEADER` keyword false. Another keyword, `ERRORS` controls whether the field errors are printed. The `HMS` logical keyword toggles between decimal and sexigessimal output of celestial coordinates. These toggles permit most non-numeric text to be excluded from the output text, this facilitating import into other software.

6.2 SSCARIN

INP	Input SSDS containing source positions Takes default from current ASTERIX SSDS
OUT	Output SCAR catalogue name
SORT	Sort records into standard SCAR order?

SSCARIN performs the export of the SSDS specified by `INP` to a SCAR format binary catalogues specified by `OUT`. Note that the output file name should not be preceded by a directory specification, nor should a file extension be supplied. Two files are created called `<OUT>.DAT` and `DSCF<OUT>.DAT` where `<OUT>` is the value of the `OUT` parameter. For example,

```
> SSCARIN my_sources my_cat \
SSCARIN Version 1.5-1
> CAR_REPORT my_cat 2 NOSELECT \
```

The `SORT` logical parameter (default `TRUE`) controls whether the output file is sorted in the standard SCAR fashion, ie. descending DEC, ascending RA. If the file is sorted, then immediate cross-correlation with most SCAR catalogues will be possible. If `SORT` is specified `FALSE` then a field called NUMBER is created which is the ordinal number of each source in the original SSDS. This field is required as a SCAR catalogue *must* be sorted by some field.

6.3 SSANOT

INP	Input graphics dataset to be annotated Takes default from current ASTERIX graphics dataset
LIST	Input SSDS containing source positions Takes default from current ASTERIX SSDS
MARKER	Marker symbol to mark source position Hidden default is 2 (+)
BOLD	Boldness of marker symbol to be plotted
ERROR	Draw error circles if present

SSANOT is the interface between SSDS and the ASTERIX presentation graphics system. It annotates a specified ASTERIX graphics dataset with information contained in an SSDS, eg.

```
> SSANOT my_pss_demo my_sources \
```

Annotations created by SSDS can be canceled using,

```
> ANOTATE CANCEL INP=dataset \
```

or

```
> ANOTATE CANCEL \
```

assuming the current ASTERIX graphics dataset is defined. Note, however, that this command cancels *all* annotations, and not just those created by SSANOT.

6.4 IMARK

CURR	Mark current position? Hidden default `N`
LIST	Source list file
SYMBOL	Marker symbol, default is 2 (+)
COLOUR	Symbol colour, default is 1 (white)

The IMARK command is part of the ASTERIX image processing system. One of its modes of operation enables an SSDS to be used as the source of celestial positions (thus enabling images with different orientations from that which was searched to be annotated). Control over the type and colour of the plot symbol can be exercised. For example,

```
> ISTART my_pss_demo vws
> IDISPLAY
> IMARK MY_SOURCES SYMBOL=4 COLOUR=2
```

marks each source position with a red open circle.

6.5 SSMERGE

INP	Wildcard specification of SSDS to merge Takes default from current ASTERIX graphics dataset
OUT	Name of output SSDS Becomes current ASTERIX SSDS

The function of SSMERGE is purely to collect source search results files together in to one output file, for example,

```
> SSMERGE ps_%%%% merged_ps \
SSMERGE Version 1.5-0
Will merge 23 input files
Merging files 1 to 10
Merging files 11 to 20
Merging files 21 to 23
```

It is important to realise what this application *does not* do,

- It does not remove multiple detections of the same source. All sources in all the input files appear in the output file.
- It cannot accept SSDS with field errors of different forms in the input files.
- Only those fields present in all inputs appear in the output.

SSMERGE is especially useful for collating source search runs produced as a result of simulation runs, where many SSDS are likely to be produced.

6.6 SSZAP

SSZAP is a procedure and hence all parameters must always be specified on the command line. The syntax to invoke it is,

```
> SSZAP <dataset> <ssds> <radius> <qmode>
```

SSZAP applies the quality operation specified by `<qmode>` to all pixels in `<dataset>` within a radius `<radius>` of the points in space defined by the positions of the sources in `<ssds>`. The value of `<radius>` should be specified in the spatial axis units of `<dataset>`. `<qmode>` should be any one of the quality operations supported by the CQUALITY application.

The first point to note is that `<dataset>` need not be the dataset which produced `<ssds>` – in fact it can be *any* n-dimensional binned dataset as long as the pointing direction is the same as that which produced `<ssds>`. Consider the following example,

```
> XSORT ... OUT=cube                ; Create data cube from instrument
                                    ; interface
> PROJECT cube image 3              ; Collapse cube to image
> PSS image 0.5 sources \           ; Source search the image using
                                    ; a noddy background value
> SSZAP cube 0.02 sources IGNORE    ; Ignore data in cube within 0.02
                                    ; degrees of a source
> INTERP cube intcube ...           ; Patch over data where sources were
> PROJECT intcube better_bgnd 3     ; Create new bgnd model
> SSZAP cube 0.02 sources RESTORE   ; Reset quality in cube
> PSS image better_bgnd ...         ; Improved source search
```

6.7 EVSIM

OUT	Name of output event dataset Becomes current ASTERIX event dataset
BMODEL	Name of a 2D background model Default is `!`, ie. no model
BACK	Number of counts in background
FIELDSIZE	Field size in arcminutes
PIXSIZE	Instrument resolution in arcminutes
SEED	Seed for random number generator
SOURCEC	Source counts – separate by commas
SOURCEP	Source positions
PSFCON	Psf constant across the field
PSF...	The PSF system parameters

EVSIM generates simulated event datasets which can subsequently be binned into images. The program combines a background model with source models defined using the PSF system, although either may be used individually.

6.8 CREPSF

OUT	Name of output tabular psf Becomes current ASTERIX binned dataset
PIXSIZE	Pixel size in arcminutes
RADIUS	Half width of field in pixels
PSF...	The PSF system parameters
MJD	Modified Julian Date for dataset Use for time dependent psfs
X0,Y0	The image position for psf evaluation The default is the image coordinate $(0,0)$

CREPSF generates a two dimensional binned dataset containing a psf of the user's choice. The user controls the bin and image sizes, and the psf to be used. At the moment, this application cannot produce tabular datasets with spatially varying psfs.

There are two main uses of CREPSF. The first is as a method of speeding up multiple invocations of programs using the same psf, which is slow to compute. For example, the BSUB application requires a psf template which enclosed 90% to 95% of the psf energy. At the WFC survey image size of 1 arcminute this required a psf more than 30 pixels across, which took of order 30 seconds to compute. The following sequence of commands generated a psf which could be used for both BSUB and PSS,

```
> CREPSF OUT=PSF_S1A PSF=WFC ~      ; Create tabular psf
    AUX=S1A PIXSIZE=1 RADIUS=15 \
> BSUB IMAGE PSF=TABULAR ~          ; Invoke background subtraction
    MASK=PSF_S1A ...                ; using tabular psf
```

The second use is in searching or parameterising extended sources. As mentioned in Section 4.9 PSS is limited in its ability to search for very extended sources. An alternative to using PSS source extension testing mechanism (which can only be used in SEARCH and PARAM modes anyway) is to use CREPSF and (say) SMOOTH to create a test source profile. The example below creates the extended spread function `EXTEN_SF` by convolving the point response produced by CREPSF with a top-hat function (the source's supposed intrinsic surface brightness distribution),

```
> CREPSF OUT=POINT_SF PSF=WFC ~     ; Create tabular instrument psf
    AUX=S1A PIXSIZE=1 RADIUS=15 \
> SMOOTH POINT_SF EXTEN_SF ~        ; Smooth with top-hat
    MSK_MASK=TOP MSK_WIDTH=5 ..
> PSS ... MODE=UPLIM PSF=TABULAR ~ ; Upper limit using extended
    MASK=EXTEN_SF                   ; spread function
```

and uses it to find the upper limit to the source intensity using PSS.

Note that it is important that the normalisation of the psf remains correct, ie. sum of the values in the array would be unity if it was big enough. Failure to ensure this will result in incorrect results out of most programs using the ASTERIX psf system.

Once the tabular psf has been created it can be used in any ASTERIX application which uses the psf system, including the event simulator EVSIM. This is useful as perhaps the most reliable method of measuring the significance of extended source features is by testing the processing method on simulated data, with and without the supposed feature.

6.9 WFCSPEC

WFCSPEC is for use only with ROSAT WFC data.

INP	Name of input dataset. May be an SSDS
OUT	The output spectrum name
SRC	The source number to choose in the SSDS

Creates a Wide Field Camera 'spectrum' (which has only one channel) complete with energy response.

The program takes as input either a WFC datafile (an NDF) containing only a single data value (it does not have to be 1-dimensional though), or an SSDS. In the latter case, all sources in the file will be listed, and the user selects which is to be used.

6.10 XPSSCORR

XPSSCORR is for use only with ROSAT PSPC data.

INP	Name of input SSDS
MEAN_EN	Mean photon energy in searched image
RTNAME	Root name for calibration files
RESPFILE	Name of detector matrix file

Converts raw flux in an SSDS counts per second by correcting for vignetting as a function of both position and energy, and presence of wires. The input SSDS is updated with the new information.

7 Techniques

7.1 Bad data

Quality handling

7.2 Simulation

7.3 Upper limits

As an object for further discussion it is helpful to have a copy of the upper limits map produced by PSS for the demonstration image. A copy can be created using the command,

```
> PSS ast_demo:pss_demo 1.01 mode=upmap map=demo_map NOEXPERT \
```

which creates an upper limit flux map called `DEMO_MAP` at 68% confidence. Figure 3 is copy of this image contoured at 10, 20, 30, 40, 50, 75, 100, 150, 200, 250, 300 and 350 counts.

Two features of the upper limit map are worth noting immediately,

- The two simulated sources stand out strongly. It is apparent that these sources (FWHM about 3 arcminutes) contaminate the surrounding data for quite some distance
- The remainder of the image is clumpy on a scale similar to that of the psf.

The first can be explained in terms of the PSS algorithm. As stated previously PSS has a simple data model, consisting soley of a background plus one source. When finding upper limits the source part of the model is at the position of the requested upper limit (or the pixel centre in upper limit mapping mode), and hence any nearby strong source makes this a poor representation of the data. To find upper limits near bright sources using PSS it is necessary to either subtract the contaminating source (Section 4.13), or to ignore it (Sections 6.6 and 7.4).

The upper limit flux map is clumpy simply because of local variations in the background. Test positions in an image sited in areas deficient in counts with respect to the local background constrain a necessarily positive source flux more tightly than positions in areas with excess counts. The justification for this statement can be found in Appendix **??**.

The restricted scope of the above analysis must be emphasised. The upper limit flux map is the flux of a *point* source whose surface brightness is exactly represented by the psf system, at a confidence level 68%. This raises the question of what is required by a particular upper limits analysis – the answers to two questions should be clear in the mind of anyone using PSS in UPLIM mode,

- Is the test source position known well? The definition of 'well' depends on the data, but basically relates uncertainties in a test position to the scale of clumpiness in the upper limit flux surface, which in turn depends on the characteristic size of the psf. There are two possible sources of uncertainty. The first is in the (assumed) X-ray data, where uncertainties are usually limited by residual errors in spacecraft attitude reconstruction. These errors may be manifested as either systematic shifts across a field or simply as a random positional blurring of the events constituting an image. The sizes of these errors are of order 10 and 5 arcseconds for the WFC and PSPC respectively – about one third of the smallest pixel resolution for both detectors. The second source of error is that in the test position. This is generally smaller than the first source of error for wavebands other than X-ray or γ-ray, and can be generally regarded as zero for stellar data. Combine estimates of your two sources of error in quadrature and compare with the FWHM of the psf being used. If the former is smaller than the latter by a factor of ~ 3 then using UPLIM mode is acceptable. This is simply because possible errors in the registration of the test position on the image are much smaller than the scales over which the upper limit surface is varying significantly. If the condition is *not* satisified then UPMAP mode should be used to map the upper limit surface over a small area around the test position, and the upper limit taken to the maximum value in that area.

- Is the test source psf well known? In the case where the test is being performed on an assumed point source this question reduces to one of whether the psf system as accurately modelling the psf. This topic is addressed in Appendix A.1. Where upper limits are required for sources of unknown surface brightness distribution, the only technique avaliable using PSS is to construct various likely surface brightness models (see Section 6.8) and use whatever mode has been selected on the basis of the answer to previous question for these different models.

Practical examples may serve to clarify the above. Suppose we had a poorly constrained source position corresponding to image position $(-20, -20)$ in Figure 3, with an error circle of radius 5 arcmin. By our first criterion we should use UPMAP mode, as our uncertainity is comparable to if not greater than the clumpiness in the upper limit surface. The largest upper limit flux within the circle about 26 counts. A supposed stellar object at the same position with an uncertainty of say 1 arcsecond allows use of UPLIM mode and would give a result of only 2.5 counts.

7.4 Using data QUALITY

7.5 Interpreting Changes in Significance

7.6 Iterative Methods

8 References

Cash, Webster, 1979 (*Ap.J.*,**228**,939)

Currie, M.J., Wallace, P.T. & Warren-Smith, R.F., 1989 *SGP/38 Starlink Standard Data Structures*

Ponman, T.J. PROG_002 *Asterix Standard Structures and Conventions*

Lampton, Margon & Bowyer (*Ap.J.*,**208**,177)

ASTERIX online help facility.

A The PSF System

The ASTERIX psf system is used by instrument independent applications requiring access to point spread function information, which is inherently instrument specific. The psf uses three ADAM parameters, `PSF`, `MASK` and `AUX`, although not all three are needed by all psfs. The `PSF` parameter is the most important, controlling the psf to be used,

```
PSF system options :

  ANAL          EXOLE         PWFC          TABULAR       WFC
  XRT_HRI       XRT_PSPC

  POLAR(psf,rbin[,abin])      RECT(psf,xbin[,ybin])

PSF - Choose PSF to use for source model /'WFC'/ >
```

Note that the exact prompt displayed with `PSF` may vary from application to application, as the use to which the data is to be put may also vary. The seven lines preceding the prompt show the options available in the psf system, in two groups. The first group contains the primitive psf names, `ANAL`, `EXOLE` and so on. These are the names of instrument psfs plus `TABULAR` and `ANAL`. To invoke one of this group, simply enter its name at the `PSF` prompt. A default may be supplied if the dataset was produced by an ASTERIX instrument interface or the event simulator, EVSIM.

The second group shows the allowed syntax for psf model specifications.

A.1 Instrument Psfs

The PSF system supplies the following instrument specific psfs,

- **EXOLE** The EXOSAT LE detector

- **PWFC** The ROSAT WFC (detector 2). The PWFC psf accesses the WFC online calibration database to construct the spatially varying psf. The psf is evaluated at a single energy which is taken to be the peak of the bandpass of the filter used. If the filter cannot be found from the dataset, then the `AUX` parameter is used to prompt for it.

- **WFC** As PWFC but for survey data. The WFC survey psf is a weighted supposition of the pointed response sampled at a large number of field positions. Note that EVPSS should not use this psf, *even for survey data*, as the detector position is known for every photon.

- **XRT_HRI** The ROSAT HRI detector. This psf combines the essentially constant detector response with a radial model of the off-axis telescope response. The latter effect is not significant due to the small field of view of this detector

- **XRT_PSPC** The ROSAT PSPC1 and 2 detectors. The current model is a simple gaussian, whose width is a function of both off-axis angle and photon energy. The `AUX` prompt controls an assumed mean photon energy for the dataset. The current model is rather poor outside the support ring, and even inside does not model the mirror scattering properly. A more accurate model should be avaliable soon (March 92), which will incorporate a 3 component model consisting of a gaussian core, a Lorentzian mirror scattering component and an exponential fit to the wings caused by the deeper penetration of the detector by high energy photons.

A.2 ANALytic Psfs

The `ANAL` option permits one of a number of analytic functional forms to be chosen for a psf, using the `MASK` parameter,

```
PSF definitions available

GAUSSIAN   - Gaussian response    TRIANGLE   - Triangular response
TOPHAT     - Circular top-hat     FLAT_TRI   - Triangle with flat top

MASK - Name of profile to use (select from above) /'GAUSSIAN'/ >
```

The default is always `GAUSSIAN`. The spatial scale of the analytic form is established using the `AUX` parameter. The nature of the data required varies with the `MASK` choice,

```
Dataset pixels are 1 arcmin square   ; User is informed of dataset
                                     ; pixel size.

AUX - Gaussian FWHM in arcmin /1/ >  ; GAUSSIAN prompt

AUX - Tophat full width in arcmin >  ; TOPHAT prompt

AUX - Triangle zero-point full       ; TRIANGLE prompt
                width in arcmin >

Please supply full-width at zero
 response (FWZR) and full-width
 half-max (FWHM)
AUX - FWZR and FWHM in arcmin >      ; FLAT_TRI prompt
```

Note that in all the above cases the geometrical information is supplied in terms of the axis units of the dataset if present (pixels otherwise). These data are sufficient to define the normalisation of the psf in each case, as the integrated total is always unity.

A.3 Tabular Psfs

The `TABULAR` option reads in a dataset created by the CREPSF application (see Section 6.8). The psf system uses the `MASK` parameter to read in the name of the dataset,

```
PSF - Choose PSF to use for source model /'WFC'/ > TABULAR
MASK - Name of a 2D dataset containing psf > POINT_SF
```

The `AUX` parameter is not used. Note that the psf system does not check that the binning of the tabular psf is compatible with an image being searched or background subtracted. The solution to this shortcoming is not simple, as there are occasions when the binning does not matter, and the TABULAR psf routine has insufficient information to decide this.

A.4 PSF models

The practical application of psf models has already been alluded to (see Sections 3.3 and 4.4). The advantage of psf models over simple psf access is purely one of processing time. Injudicious selection of parameters can result in either poor representation of the spatial variation of the psf or even *more* processing time being expended than with simple access.

The model psfs should be used in conjunction with applications which have the facility to support spatially variable psfs,

Application	Psf Access Method
PSS, EVPSS	Either once per field or at every image position Controlled by `PSFCON` parameter
EVSIM	Once per field or once per simulated source position Controlled by `PSFCON` parameter

When using PSS or EVPSS it is almost always worthwhile using a model to represent variable psfs, as access is so intensive. The psf access time for EVSIM will only dominate for large numbers of sources, so the model option is only worthwhile in that situation. All other applications using the psf system cannot benefit from the use of model options.

The two model options work on a similar principle. The model specifications define a grid of a finite number of psfs. Each access to the psf system by an application at a given image position then extracts the psf from the grid bin containing that position. Psfs are are only calculated as they are needed, so not even the number of psfs in the grid need be calculated. The first argument of each model specification is an instrument psf name from the first display group.

The `POLAR` option sets up a grid whose bin boundaries form lines of constant radius and/or azimuth in polar coordinates. The most simple form of a polar psf grid defines only regularly spaced radial bins. For example,

```
POLAR(XRT_PSPC,0.04)
```

defines a grid of the ROSAT PSPC psf where the radial bin boundaries are 0.04 degrees apart (floating point values in model specifications are *always* in image axis units). The psf system automatically finds the number of radial bins from the image centre to the extreme corner. The first complication is to allow non-regular radial bins,

```
POLAR(XRT_PSPC,0.05:0.1:0.2:0.5)
```

This example defines bin boundaries placed at to 4 specified radii (see Ref 5,"*User Interface,Lists and ranges*" for a description of bin boundary syntax). The psf system supplies additional boundaries of zero and infinity. The third optional argument for a polar model is an integer, the number of azimuthal bins. By default this is one, but any number can be specified. The specification,

```
POLAR(XRT_PSPC,0.05:0.1:0.2:0.5,4)
```

defines 4 azimuthal bins. The boundary of the first starts at the algebraic angle zero degrees, proceeding round the image centre at intervals of $\frac{360}{n}$ degrees. A special case is the central radial bin, which *always* has only one azimuthal bin. This prevents subdivision of the image centre into very small grid bins ensuring continuous psf data across this area (where there is greater likelihood of there being a source).

The `RECT` option sets up a grid whose bin boundaries form lines in constant image ordinate or abscissa. Again the third argument is optional, but in the `RECT` option the default is to assume in the same specification as defined in the second argument. Again there is a regular and an irregular form,

```
RECT(PWFC,5)
```

defines a regular grid of spacing 5 arcminutes in both axes, whereas

```
RECT(PWFC,-60:-40:-20:-10:0:10:20:40:60)
```

defines an irregularly spaced grid. The third argument, if specified, has the same syntax as the second.

The precise grid to define depends on the instrument, some guidance can be found in Ref 5,"*User Interface,Psf System*". Once the psf model has been specified, the `MASK` and `AUX` prompts may appear to get additional information for the instrumental psf.

B PSS Algorithms

The following sections use common terminology to refer to the various data used by PSS.

$d_{i,j}$	The regularly spaced input data (ie. the data from the `INP` dataset. The indices (i,j) can be thought of as pixel numbers
$\sigma^2_{i,j}$	The variance of the input data
$b_{i,j}$	The background model data values. The $b_{i,j}$ match the $d_{i,j}$ one-to-one.
$p_{i,j;x,y}$	The point spread function (psf) amplitude integrated over pixel (i,j) arising from a point source at image position (x,y).
$m_{i,j}$	The PSS model value at pixel (i,j). See below for the two forms this can take.

The normal model consists of a known background plus unknown contribution from a point source,

$$m_{i,j} = A_{x,y} p_{i,j;x,y} + b_{i,j} \tag{1}$$

where $A_{x,y}$ is the flux at image position (x,y). Background rescaling allows one more degree of freedom by allowing the background normalisation to change at each point in the image,

$$m_{i,j} = A_{x,y} p_{i,j;x,y} + \beta_{x,y} b_{i,j} \tag{2}$$

where $\beta_{x,y}$ is the unitless background rescaling factor.

B.1 CASH

Selecting the `CASH` option at the `SOPT` prompt selects the Cash maximum likelihood statistic (Ref 1) as the method of comparing data and model.

$$C_{x,y} = 2 \sum_{i,j} [m_{i,j} - d_{i,j} \log m_{i,j} + \log(d_{i,j}!)] \tag{3}$$

where (x,y) is the image position, $d_{i,j}$ are the regularly spaced Poisson distributed data and $m_{i,j}$ the model.

Using the normal model (1), and ignoring the $\log(d_{i,j}!)$ term in (3), which is independent of the parameters (and hence constant for a given dataset), we minimise,

$$\tilde{C}_{x,y} = 2 \sum_{i,j} [A p_{i,j;x,y} + b_{i,j} - d_{i,j} \log(A p_{i,j;x,y} + b_{i,j})] \tag{4}$$

the summation being performed over all pixels for which p is significantly greater than zero (see discussion below). At each (x,y) the statistic is evaluated repeatedly to find the optimum flux $\hat{A}$, when,

$$\frac{\partial \tilde{C}}{\partial A} = \sum_{i,j} p_{i,j;x,y} - \sum_{i,j} \frac{d_{i,j} p_{i,j;x,y}}{A p_{i,j;x,y} + b_{i,j}} = 0 \tag{5}$$

B.1.1 Detection Significance

Since $\tilde{C}$ is χ_1^2 distributed about its minimum with respect to A (at a fixed point), the probability of $A \leq 0$ can be evaluated by performing a one-tailed χ_1^2 test on the statistic. Given the best-fit value $\hat{A}$, the detection significance is evaluated by noting that that $A = 0$ falls on the probability P confidence contour, such that

$$\chi_1^2(P) = \tilde{C}_0 - \tilde{C}_{min} \tag{6}$$

The significance of the detection is then the deviate z_{norm} of a one-tailed normal distribution (one-tailed as negative A are regarded as un-physical) such that

$$\int_0^{z_{norm}} erf(x)dx = \frac{P}{2} \tag{7}$$

For small values of χ_1^2, P can be readily found using a NAG routine and converted into the a normal deviate. However, for very small P (or equivalently, large z_{norm}) the NAG routine has overflow problems. The following asymptotic expansion is useful,

$$P(|z| > x) = \frac{2e^{-x^2/2}}{x\sqrt{2\pi}} \left\{1 - \frac{1}{x^2} + \frac{3}{x^4} - \frac{3.5}{x^6} + \frac{3.5.7}{x^8} - \cdots\right\} \tag{8}$$

B.1.2 Background Rescaling

In background rescaling mode, PSS's model is (2) and (3) becomes,

$$\tilde{C}_{x,y} = 2\sum_{i,j}[Ap_{i,j;x,y} + \beta b_{i,j} - d_{i,j}\log(Ap_{i,j;x,y} + \beta b_{i,j})] \tag{9}$$

This is minimised to find both $\hat{A}$ and $\hat{\beta}$ at each (x, y) using the simultaneous conditions

$$\frac{\partial \tilde{C}}{\partial A} = \sum_{i,j} p_{i,j;x,y} - \sum_{i,j} \frac{d_{i,j}p_{i,j;x,y}}{Ap_{i,j;x,y} + \beta b_{i,j}} = 0 \tag{10}$$

$$\frac{\partial \tilde{C}}{\partial \beta} = \sum_{i,j} b_{i,j} - \sum_{i,j} \frac{d_{i,j}b_{i,j}}{Ap_{i,j} + \beta b_{i,j}} = 0 \tag{11}$$

The minimum is found iteratively using the second order scheme,

$$\begin{bmatrix} \hat{A}' \\ \hat{\beta}' \end{bmatrix} = \begin{bmatrix} \hat{A} \\ \hat{\beta} \end{bmatrix} - \begin{bmatrix} \tilde{C}_{AA} & \tilde{C}_{A\beta} \\ \tilde{C}_{A\beta} & \tilde{C}_{\beta\beta} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{C}_A \\ \tilde{C}_\beta \end{bmatrix} \tag{12}$$

where $\tilde{C}_X$ denotes differentiation with respect to X.

B.2 GAUSSIAN

The search statistic used by PSS is

$$\tilde{S}(x,y) = \frac{\sum_{i,j} p(i,j;x,y)d(i,j)/\sigma^2(i,j)}{\sqrt{\sum_{i,j} p(i,j;x,y)^2/\sigma^2(i,j)}}, \tag{13}$$

where $d(i,j)$ is the data value in pixel (i,j), $\sigma^2(i,j)$ is its variance, and $p(i,j;x,y)$ is the point spread function (psf) amplitude integrated over pixel (i,j) arising from a point source at position (x,y). The form of this psf could in general be position-dependent. The pixels (i,j) are assumed to be regularly spaced and the summations in (13) extend over all pixels for which p is significantly greater than zero (see discussion below).

The numerator in equation (13) is just the weighted cross-correlation of the data with the psf, whilst the denominator is its standard deviation – hence $\tilde{S}$ is a measure of source significance. The search threshold should obviously be set to avoid large numbers of spurious detections. For a threshold value S_{th}, one expects to get typically $n = P(z > S_{th})\Psi_{\rm field}/\Psi_{\rm psf}$ spurious detections, where $\Psi_{\rm field}$ is the area of the field being searched, $\Psi_{\rm psf}$ is the area under the psf, and $P(z > S_{th})$ is the integrated probability under the upper tail of a unit normal distribution, beyond the value S_{th}; e.g. $P = 1.35 \times 10^{-3}, 3.2 \times 10^{-5}, 2.9 \times 10^{-7}$ for $S_{th} = 3, 4, 5$ respectively.

This all assumes that any background has been removed from the data d – if not then it will contribute to the value of $\tilde{S}$ and will increase the incidence of spurious detections.

B.2.1 Selection of psf box

Evaluation of the summations in equation (13) is the most time consuming component of the source search. It is therefore important to limit the summations to a suitable range – an overlarge range may result in excessive processing time, whilst too small a range will cause some source flux to be omitted from the cross-correlation, reducing the detection significance. The latter effect may be quantified by considering the case where a point source of strength A is located, so that $d(i,j) \approx Ap(i,j;x,y)$. Substituting this in (13), the detection significance should be

$$\tilde{S} = \frac{A\sum_{i,j} p_{i,j}^2/\sigma_{i,j}^2}{\sqrt{\sum_{i,j} p_{i,j}^2/\sigma_{i,j}^2}} = A\sqrt{\sum_{i,j} p_{i,j}^2/\sigma_{i,j}^2}, \tag{14}$$

where the x,y indices have been suppressed for simplicity.

Consider the common case of an unscaled (i.e. counts) image with Poissonian errors. In the limit of a strong source, such that background is insignificant, $\sigma_{i,j}^2 \approx Ap_{i,j}$ and (14) reduces to

$$\tilde{S} \approx \sqrt{A}\sqrt{\sum_{i,j} p_{i,j}}. \tag{15}$$

Hence, setting the range of i, j such that $\alpha\%$ of the psf is excluded would lead to a loss of significance of $\approx \frac{\alpha}{2}\%$ (e.g. a box extending to 90% enclosed energy would include about 95% of the significance).

In the weak source limit, $\sigma^2_{i,j}$ is dominated by background counts. Assuming these to be constant over the region, so that $\sigma^2_{i,j} = \sigma^2$, equation (14) becomes

$$\tilde{S} \approx \frac{A}{\sigma}\sqrt{\sum_{i,j} p^2_{i,j}}. \tag{16}$$

In this limit, the loss in sensitivity due to truncation is much smaller than for a strong source, since $\sum_{i,j} p^2_{i,j}$ is dominated by the central pixels where $p_{i,j}$ is largest.

In practice, then, there is rarely any need to extend the psf box beyond the 90% enclosed energy radius, and a somewhat smaller radius will often give acceptable results.

B.2.2 Source flux estimates

Source searching can be viewed as a fitting operation, in which the psf profile is matched to the observed flux distribution. The presence of a source is indicated by a good fit with significantly non-zero amplitude. The weighted least squares solution to the fitting problem entails minimising the chi-squared distributed statistic

$$C = \sum_{i,j} [A p_{i,j;x,y} - d_{i,j}]^2 / \sigma^2_{i,j}, \tag{17}$$

where A is the source strength. For a given trial source position (x, y) the optimum estimate of A is that which gives $\partial C/\partial A = 0$, i.e.

$$\hat{A} = \frac{\sum_{i,j} p_{i,j;x,y} d_{i,j} / \sigma^2_{i,j}}{\sum_{i,j} p^2_{i,j;x,y} / \sigma^2_{i,j}} = \frac{\tilde{S}}{\sqrt{\sum_{i,j} p^2_{i,j;x,y} / \sigma^2_{i,j}}}. \tag{18}$$

When the errors are Gaussian distributed, this can also be shown to be the maximum likelihood (ML) estimate of the source flux. The variance of this estimate is

$$\mathrm{var}(\hat{A}) = \frac{\mathrm{var}(\tilde{S})}{\sum_{i,j} p^2_{i,j;x,y} / \sigma^2_{i,j}} = \frac{1}{\sum_{i,j} p^2_{i,j;x,y} / \sigma^2_{i,j}}. \tag{19}$$

B.2.3 Flux and Positional Errors

C EVPSS Algorithms

D Figures

- Schematic of PSS method. See Section 4.

- Shows contours of significance of a change in significance (y axis) as a function of significance (x-axis). See Section 7.5 for details. Contours are at 1, 3, 5, 7, 10, 15, 20 and 30 sigma.

- A 68% upper limit flux map of the PSS demonstration image, `AST_DEMO:PSS_DEMO`. The contours are at 10, 20, 30, 40, 50, 75, 100, 150, 200, 250, 300 and 350 counts. See Section 7.3 for further details.