SOURCE DETECTION AND BACKGROUND ESTIMATION WITH BAYESIAN INFERENCE

Fabrizia Guglielmetti[1,2], Rainer Fischer[1], Wolfgang Voges[2], Guenter Boese[2], and Volker Dose[1]

[1]*Centre for Interdisciplinary Plasma Science, Max-Planck-Institut für Plasmaphysik, Garching, Germany.*
F.Guglielmetti@ipp.mpg.de, Rainer.Fischer@ipp.mpg.de, Volker.Dose@ipp.mpg.de
[2]*Max-Planck-Institut für extraterrestrische Physik, Garching, Germany.*
fabrizia@mpe.mpg.de, whv@mpe.mpg.de, gub@mpe.mpg.de

ABSTRACT

A probabilistic technique for the joint estimation of background and sources in high-energy astrophysics is described.
Bayesian inference is applied to gain insight into the coexistence of background and sources through a probabilistic two-component mixture model, which provides consistent uncertainties of background and sources.
The present analysis is applied on ROSAT PSPC data in Survey Mode. A background map is modelled using a Thin-Plate spline. Source probability maps are obtained for each pixel (45 arcsec) independently and for larger correlation lengths, revealing faint and extended sources. Source probability maps are combined for two ROSAT PSPC energy bands, hard (0.5-2.0 keV) and soft (0.1-0.5 keV), and compared with the corresponding source probability maps at the broad energy band (0.1-2.4 keV) and with the ROSAT All-Sky Survey (RASS) catalogues, bright and faint.
The probabilistic method allows for detection improvement of faint extended celestial sources compared to the standard methods applied for the realization of the RASS catalogues.

Key words: data analysis; Bayesian inference; background estimation; source detection.

1. INTRODUCTION

In high-energy astrophysics source detection is limited by the small number of photon counts. The Poisson distribution has to be used for statistical inferences. It becomes difficult to detect sources when they are faint, point-like or extended, and to be distinguished against the background.
The RASS x-ray data were largely analyzed with the Standard Analysis Software System (SASS). The results can be found in the Bright and Faint Source catalogues (Voges et al. 1999, 2000). Nevertheless, SASS is known for lack of sensitivity for faint or extended sources. This is due to the sliding window technique which locally searches for count enhancements relative to the intensity in a surrounding area defining the background intensity. In multiple steps the window width is changed to allow for the detection of extended sources. But faint extended sources and blended faint sources in crowded fields may get lost. One reason is due to the local estimation of the background in a small region around the sliding window which may provide only poor signal-to-noise ratios (S/N). The sources are characterized by fitting the candidate sources in a further step using a Maximum-Likelihood (ML) method (Boese & Doebereiner 2001). The ML method works properly on prominent point-like sources which account for 94% of the sources published in the RASS Bright Source catalogue. The characteristics of the faint sources may not be properly estimated in case of faint extended sources (Voges et al. 1999).
The proposed method using Bayesian probability theory (BPT) estimates the background and sources in a single step neither employing pixel censoring nor using a sliding window technique. The aim is to infer simultaneously a background map for the complete field size ($6.4^\circ \times 6.4^\circ$ in the sky) and a probability for having source intensity in addition to the background intensity in a pixel cell or pixel domain. Note that this method does not make use of background subtraction for source detections. Bayesian inference allows to reason on the basis of sparse data employing additional information independent of the data. The results are given by probability distributions quantifying our state of knowledge. For background estimation and source detection the additional information is the assumption that the background is smooth, e. g. spatially slowly varying compared to source dimensions. To allow for smoothness the background is modelled with a bivariate Thin-Plate spline. The coexistence of background and sources is described with a probabilistic two-component mixture model where one component describes background contribution only and the other component describes background plus source contributions. Each pixel cell (or pixel domain) is characterized by the probability of belonging to one of the two mixture components. For the background spline estimation the photons contained in all pixel cells are considered including pixels containing additional source contributions. The source probability maps produced in pixel domain correlate information with neighbouring pixels for detecting faint and extended

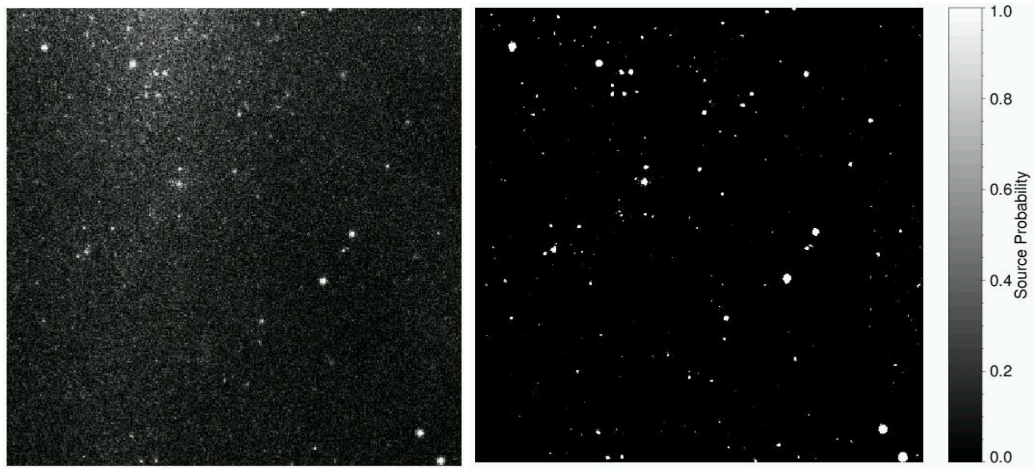

Figure 1. On the left, **RS930625** *ROSAT PSPC in Survey Mode field, broad energy band. The image center is located at $\alpha = 17^h 49^m 5^s$, $\delta = +61°52'30''$. North is up, and east is to the left. The field of view corresponds to $6.4° \times 6.4°$ in the sky. The observatory's exposure time ranges from 1693.1 sec to 13475.5 sec. On the right, the corresponding source probability map estimated with BPT accounting for the width of the instrumental point spread function and utilizing gaussian weights.*

sources. Two methods have been developed so far either correlating information in a box or employing gaussian weighting.
This technique is applied on a data sample coming from the ROSAT PSPC in Survey Mode (0.1-2.4 keV). The observatory's exposure time and point spread function (PSF) have been properly accounted for. Source probability maps for the three ROSAT PSPC energy bands have been produced. Soft and hard energy bands are combined generating final source maps at different correlation lengths. These source detections are finally compared with the corresponding source probability maps at the broad energy band and with the ROSAT All-Sky Survey (RASS) catalogues (Voges et al. 1999, 2000).

2. METHOD

Given the observed data set $D = \{d_{ij}\} \in \mathbf{N}_0$, where d_{ij} is photon counts in pixel cell ij, two complementary hypotheses arise:

$$\left\{ \begin{array}{l} B_{ij} : d_{ij} = b_{ij} + \epsilon_{ij} \\ \overline{B}_{ij} : d_{ij} = b_{ij} + s_{ij} + \epsilon_{ij} \end{array} \right.$$

Hypothesis B_{ij} specifies that d_{ij} consists only of background b_{ij} spoiled with noise ϵ_{ij}. Hypothesis $\overline{B}_{ij}$ specifies the case where additional source intensity s_{ij} contributes to the background.
Additional assumptions are that no negative values for signal and background amplitudes are allowed and that the background is smoother than the signal. This is achieved by modelling the background count rate with a bivariate Thin-Plate spline where the supporting points are chosen sparsely to ensure that sources can not be fitted. The spline fits the background component whereas count enhancements classify pixel (domains) with source contributions.

The likelihood distributions for the two hypotheses are

$$p(d_{ij} \mid B_{ij}, b_{ij}) = \frac{b_{ij}^{d_{ij}}}{d_{ij}!} e^{-b_{ij}}, \qquad (1)$$

$$p(d_{ij} \mid \overline{B}_{ij}, b_{ij}, s_{ij}) = \frac{(b_{ij} + s_{ij})^{d_{ij}}}{d_{ij}!} e^{-(b_{ij}+s_{ij})}. \qquad (2)$$

For background estimation we marginalize over the signal in eqn. 2 according to the sum rule of BPT. The prior distribution over the signal is chosen to be exponential, accordingly to the Maximum Entropy principle, $p(s_{ij} \mid \lambda) = \exp\{-s_{ij}/\lambda\}/\lambda$, assuming we know only the average value of the source intensity λ over the complete field. The prior probability for the two complementary hypotheses is chosen to be: $p(B_{ij}) = \beta$ and $p(\overline{B}_{ij}) = 1 - \beta$, independent on ij.
Within the framework of BPT the hyperparameters (nuisance parameters) λ and β have to be marginalized. Alternatively and not quite rigorous in the Bayesian sense, the hyperparameters can be estimated from the marginal posterior probability density function, where the background and source parameters are integrated out.

$$\max_{\beta,\lambda} p(\beta, \lambda \mid D) \rightarrow \beta^*, \lambda^*. \qquad (3)$$

Since we do not know if a certain pixel contains purely background or additional signal, the likelihood for the mixture model is

$$p(D \mid b, \lambda^*) = \qquad (4)$$

$$\prod_{ij} [\beta^* \cdot p(d_{ij} \mid B_{ij}, b_{ij}) + (1 - \beta^*) \cdot p(d_{ij} \mid \overline{B}_{ij}, b_{ij}, \lambda^*)]$$

where, with the incomplete gamma function $\Gamma(\cdot, \cdot)$,

$$p(d_{ij} \mid \overline{B}_{ij}, b_{ij}, \lambda^*) =$$

$$\frac{e^{b_{ij}/\lambda^*}}{\lambda^*(1 + \frac{1}{\lambda^*})^{d_{ij}+1}} \cdot \frac{\Gamma[(d_{ij} + 1), b_{ij}(1 + \frac{1}{\lambda^*})]}{\Gamma(d_{ij} + 1)}$$

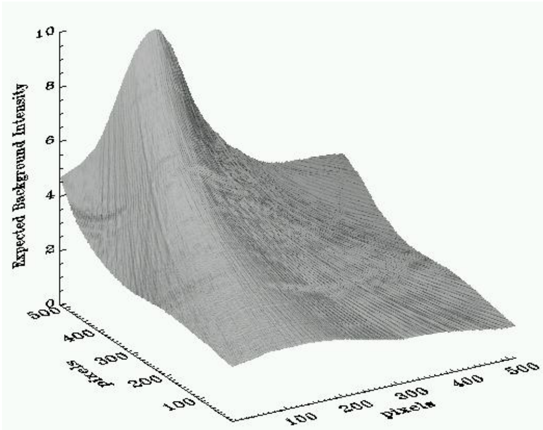

Figure 2. Background map extracted from RS930625 field (fig. 1) with a bivariate spline multiplied with the observatory's exposure time.

is the Poisson distribution marginalized over s_{ij}.
The posterior distribution is according to Bayes theorem proportional to the product of the mixture likelihood and the prior. Its maximum with respect to b gives an estimate of the background map which consists of the Thin-Plate spline multiplied with the observatory's exposure map.
The probability of having source contribution in pixel cells or domains is

$$p(\overline{B}_{ij} \mid d_{ij}) \simeq p(\overline{B}_{ij} \mid d_{ij}, b^*_{ij}, \lambda^*) = \tag{5}$$

$$\frac{p(\overline{B}_{ij}) \cdot p(d_{ij} \mid \overline{B}_{ij}, b^*_{ij}, \lambda^*)}{p(\overline{B}_{ij}) \cdot p(d_{ij} \mid \overline{B}_{ij}, b^*_{ij}, \lambda^*) + p(B_{ij}) \cdot p(d_{ij} \mid B_{ij}, b^*_{ij})}.$$

Details of mixture modelling in the framework of BPT can be found in von der Linden et al. (1999) and Fischer et al. (2000).
The instrumental PSF (in pointing mode) changes with photon energy, getting broader at higher energies. It is characterized by three components: gaussian, lorentzian and exponential (see Boese et al. 2000). The gaussian component has a standard deviation (std) smaller than 45 arcsec at higher energies. Since we are interested in detecting faint and extended sources, we approximate the PSF to a gaussian capturing the mass of the PSF. We consider correlations larger than 2 arcmin for detecting faint and extended sources: in case of gaussian weighting a minimum std of 1 arcmin was used, which corresponds to a 3 by 3 pixels domain for the box filtering.
Source probability maps are obtained for the three energy bands (broad, hard, soft) separately. BPT results of soft and hard bands are combined in a unique source probability map, whereas SASS uses the Maximum Likelihood sources of the three bands.
The resulting combined source probability maps are then compared with the broad energy band and with RASS Bright Source Catalogue (BSC) and Faint Source Catalogue (FSC).

3. RESULTS

The method previously described has been applied on simulated data and on ROSAT PSPC in Survey Mode data. We are presenting here as an example the ROSAT PSPC field RS930625 in the broad energy band ($0.1 - 2.4$ keV) located at the ecliptic polar region (see fig. 1 left). The field contains on the average 2.9 photon counts per pixel. The observatory's exposure time varies largely in this field (1693.1-13475.5 sec): brighter area in the image corresponds to larger satellite's exposure.
The sources detected from the ROSAT field RS930625 and treated with BPT are displayed on the right image in fig. 1. The resulting source probability map has been obtained correlating information with neighbouring pixels using gaussian weighting with a std including the width of the instrumental PSF. Point-like and extended sources are clearly quantified in terms of probabilities. Sources are distinctly detected also at the edge of the field.
The corresponding background map estimated from the selected ROSAT field is displayed in fig. 2. The background is assessed through the Thin-Plate spline with four supporting points and multiplied with the observatory's exposure time. The background map shows the prominent variation due to the heterogeneous satellite exposure time. Nevertheless the background map results are very smooth and remain stable also at the edge of the field. For this image the value of β^* (see eq. 3) is very close to 1, which tells us that most of the pixels belong to the background, as one can see on the right of fig. 1. Considering that the ROSAT PSPC instrumental background is known to be very low, we infer that the background map contains information of the diffuse cosmic x-ray emission only.

The ROSAT image in the hard and soft energy bands

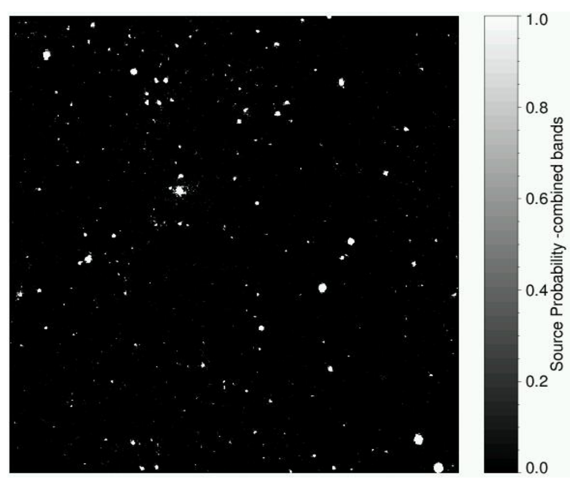

Figure 3. Combined source probability: hard (0.5-2.0 keV) and soft (0.1-0.5 keV) energy bands.

were analysed correspondingly. The background maps show that the soft cosmic x-ray count-rate is nearly twice as large as in the hard band. The broad background is higher then the soft and hard backgrounds. The resulting source probability maps in these two energy bands are combined producing conclusive source probability maps at different correlation lengths (fig. 3). The combined

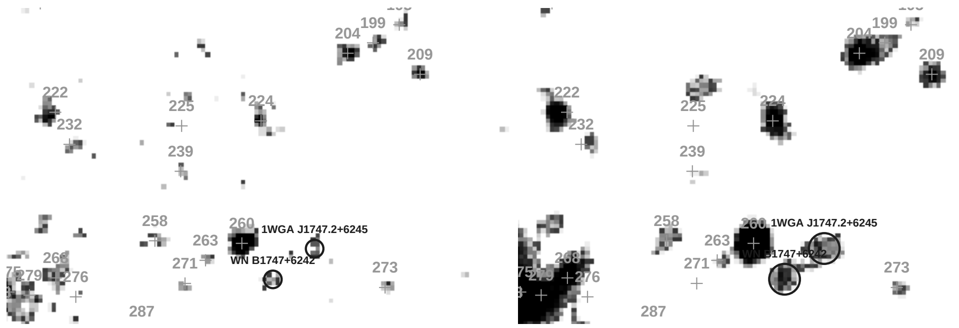

Figure 4. Source probability maps compared with SASS detections (overplotted): zoomed images of fig. 3 (with different color coding) showing a width of 100 arcmin. The left image has been produced with a std of 2 arcmin, while the one on the left with 5 arcmin. In the foreground labelled and with crosses are RASS objects, labelled and with circles are counterparts found utilizing the NASA/IPAC extragalactic database. Circles on the left image cover a radius smaller than 2 arcmin, on the right image larger than 3 arcmin.

source probability maps at different correlation lengths are compared with the broad energy band and the RASS catalogues. For the broad energy band, the combined source probability images from soft and hard bands contain more sources then the one resulting from the broad band because of the higher background in the broad band. Bright point-like and extended sources belonging to the RASS Bright Source Catalogue (BSC) have been clearly detected also with our technique. Most of the sources belonging to the RASS Faint Source Catalogue (FSC) have been also clearly detected with a probability larger than 90%.

New faint and extended sources have been found. As an example fig. 4 shows a zoomed portion of fig. 3. On the left, the image is covering a correlation length larger than the width of the instrumental PSF, i.e. 2 arcmin. The figure on the right is at the same location but obtained at larger correlation length (5 arcmin). The objects labelled and with a cross belong to the RASS FSC. Objects that were not detected on the left figure are then recovered correlating information with neighbouring pixels.

Two sources with no counterparts in the RASS catalogues are shown (see objects labelled and inside circles). Comparing our results with the NASA/IPAC extragalactic database, we found an x-ray source, 1WGA J1747.2+6245, listed in the WGA catalog (upper circle). This source has been found at 0.6 arcmin far from the position suggested by the described probabilistic method, that means that the difference between the position of our detection and of 1WGA J1747.2+6245 is less than 1 pixel ($45''$).

The second object, lower circle, may be represented by the radio source WN B1747+6242, listed in the WENSS North radio survey. This potential x-ray source has been found at a distance of 1.9 arcmin (= 2.5 pixels) from the circle center.

4. CONCLUSIONS AND FURTHER PROSPECTS

Bayesian probability theory allows to estimate background maps and to detect sources in a single step providing consistent uncertainties of background and sources. The source probability is evaluated for single pixels as well as for pixel domains (box filter or gaussian weighting) to enhance detection of weak and extended sources. We employ a probabilistic mixture model for separating the background from the sources. The detection sensitivity is enhanced compared to SASS. It results in the detection of faint sources listed in the RASS FSC and in the detection of very faint and extended sources not listed in the RASS catalogues. A more elaborate comparison with SASS results will be published in a forthcoming paper.

REFERENCES

Boese F. G., Doebereiner S., 2001, *A&A*, 370, 649

Boese F. G., 2000, *A&AS*, 141, 507

Fischer R., Dose V., 2002, *AIPC*, 617, 143

Fischer R., Hanson K. M., Dose V., von der Linden W., 2000, *PRE*, 61.1152

Guglielmetti F., Voges W., Fischer R., Boese G., Dose V., 2004, ADASS XIII, *ASP Conference Series*, Vol. 314, in print.

Voges W., et al., 2000, *IAUC*, 7432

Voges W., et al., 1999, *A&A*, 349, 389

von der Linden W., Dose V., Padayachee J., Prozesky V., 1999, *PRE*, 59.6527