Identification of high-energy astrophysical point sources via hierarchical Bayesian nonparametric clustering

Andrea Sottosanti

In collaboration with: M. Bernardi, A. R. Brazzale, A. Geringer-Sameth, D. Stenning, R. Trotta, D. A. van Dyk

University of Padova - Department of Statistical Sciences
andrea.sottosanti@unipd.it

UNIVERSITÀ DEGLI STUDI DI PADOVA

Signal source extraction

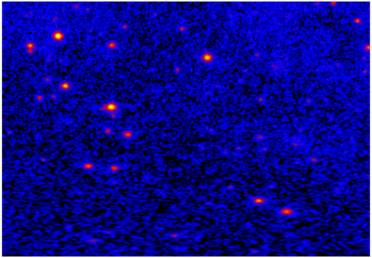


Image credit: NASA/DOE/Fermi LAT Collaboration

- Data are available in form of **photon counts** $i = 1, \ldots, n$ for which we know:

  x_i: galactic longitude,
  y_i: galactic latitude,
  E_i: energy.

Signal source extraction

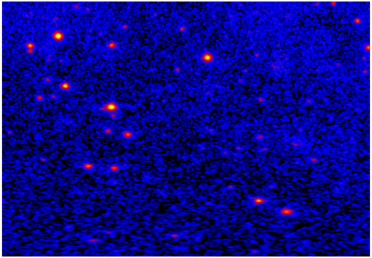


Image credit: NASA/DOE/Fermi LAT Collaboration

- Data are available in form of **photon counts** $i = 1, \ldots, n$ for which we know:

  x_i: galactic longitude,
  y_i: galactic latitude,
  E_i: energy.

- **Main Goal:** locate and quantify the signal of **astronomical sources**.

Signal source extraction

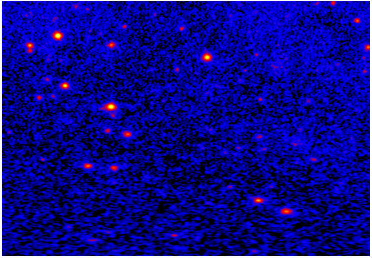


Image credit: NASA/DOE/Fermi LAT Collaboration

- Data are available in form of **photon counts** $i = 1, \ldots, n$ for which we know:

  x_i: galactic longitude,
  y_i: galactic latitude,
  E_i: energy.

- **Main Goal:** locate and quantify the signal of **astronomical sources**.

$$\Downarrow$$

accurately separate the background contamination.

Previous attempts

Single Source Models

- allow to discover a single source at a time;
- work very well on small areas;
- multiple approaches in literature [*Mattox et al. (1996)*, *van Dyk et al. (2001)*, *Protassov et al. (2002)*, *Park et al. (2006)*, *Weisskopf et al. (2007)*, *Knoetig (2014)*,...]

Multiple Source Models

- Allow a simultaneous detection of multiple sources in a map.
- Computationally demanding.
- Require the knowledge of large areas of the background.
- Only few attempts in literature [*Guglielmetti et al. (2009)*, *Primini and Kashyap (2014)*, *Jones et al. (2015)*]

The statistical model

Let us start from the finite mixture model for $\mathbf{x}_i = (x_i, y_i)$:

$$f(\mathbf{x}_i|\Theta) = \delta s(\mathbf{x}_i|\vartheta_s) + (1-\delta)b(\mathbf{x}_i|\vartheta_b), \qquad \delta \sim Beta(\lambda_s, \lambda_b).$$

The statistical model

Let us start from the finite mixture model for $\mathbf{x}_i = (x_i, y_i)$:

$$f(\mathbf{x}_i|\Theta) = \delta s(\mathbf{x}_i|\vartheta_s) + (1-\delta) b(\mathbf{x}_i|\vartheta_b), \qquad \delta \sim Beta(\lambda_s, \lambda_b).$$

The Source Model $s(\cdot|\cdot)$

- the way how photons from a source distribute in the space is known (**Point Spread Function**);
- there is no a priori information on the number of sources and their location in a map.

$$s(\mathbf{x}_i|\mathcal{F}, E_i) = \int \mathrm{PSF}(\mathbf{x}_i|\boldsymbol{\mu}, E_i)\mathcal{F}(d\boldsymbol{\mu}),$$

$$\mathcal{F} \sim \mathcal{DP}(\alpha_s, \mathcal{F}_0), \quad \mathcal{F}_0 : \begin{cases} \mu_x \sim \mathcal{U}(x_m, x_M), \\ \mu_y \sim \mathcal{U}(y_m, y_M). \end{cases}$$

The statistical model

Let us start from the finite mixture model for $\mathbf{x}_i = (x_i, y_i)$:

$$f(\mathbf{x}_i|\Theta) = \delta s(\mathbf{x}_i|\vartheta_s) + (1-\delta) b(\mathbf{x}_i|\vartheta_b), \qquad \delta \sim Beta(\lambda_s, \lambda_b).$$

The Source Model $s(\cdot|\cdot)$

- the way how photons from a source distribute in the space is known (**Point Spread Function**);
- there is no a priori information on the number of sources and their location in a map.

$$s(\mathbf{x}_i|\mathcal{F}, E_i) = \int \mathrm{PSF}(\mathbf{x}_i|\boldsymbol{\mu}, E_i)\mathcal{F}(d\boldsymbol{\mu}),$$

$$\mathcal{F} \sim \mathcal{DP}(\alpha_s, \mathcal{F}_0), \quad \mathcal{F}_0 : \begin{cases} \mu_x \sim \mathcal{U}(x_m, x_M), \\ \mu_y \sim \mathcal{U}(y_m, y_M). \end{cases}$$

The Background Model $b(\cdot|\cdot)$

- complex and completely unpredictable background, which tends to be smoother than the signal of the sources;
- let us define the B-spline kernel $\varphi(\mathbf{x}_i|\boldsymbol{\ell}, \boldsymbol{b}) = \mathcal{B}_4(x_i|\boldsymbol{\ell})\mathcal{B}_4(y_i|\boldsymbol{b})$:

$$b(\mathbf{x}_i|\mathcal{G}) = \int \varphi(\mathbf{x}_i|\boldsymbol{\ell}, \boldsymbol{b})\mathcal{G}(d\boldsymbol{\ell}, d\boldsymbol{b}),$$

$$\mathcal{G} \sim \mathcal{DP}(\alpha_b, \mathcal{G}_0),$$

$$\mathcal{G}_0(\boldsymbol{\ell}) : \begin{cases} \ell_3 \sim \mathcal{U}(x_m, x_M) \\ \ell_j \sim \mathcal{U}(x_m, \ell_{j+1}) \;\; j = 1, 2 \\ \ell_j \sim \mathcal{U}(\ell_{j-1}, x_M) \;\; j = 4, 5. \end{cases}$$

Preventing missclassification

False Positives (Type I Error)

Groups of photons from the background are confounded with point sources.

False Negatives (Type II Error)

The signal from a point source is absorbed by the background model.

Preventing missclassification

False Positives (Type I Error)

Groups of photons from the background are confounded with point sources.

False Negatives (Type II Error)

The signal from a point source is absorbed by the background model.

- Identification constraint:

$$\mathcal{V}(\boldsymbol{\ell}_k) > c, \qquad \mathcal{V}(\boldsymbol{b}_k) > c, \qquad k = 1, 2, \ldots$$

where $\mathcal{V}(\cdot)$ is the variance of a B-spline function [*Carlson (1991)*].

An extension that includes the energy

The previous model can be further extended as

$$f^{ext}(\mathbf{x}_i, E_i|\Theta^{ext}) = \delta s(\mathbf{x}_i|E_i, \mathcal{F})g(E_i|E_{min}, \eta_s) + \\ (1-\delta)b(\mathbf{x}_i|\mathcal{G})g(E_i|E_{min}, \eta_b), \tag{1.8}$$

where $g(\cdot|e, \eta)$ is the density function of a **Pareto distribution** with scale $e = E_{min}$ and shape parameter η.

An extension that includes the energy

The previous model can be further extended as

$$f^{ext}(\mathbf{x}_i, E_i|\Theta^{ext}) = \delta s(\mathbf{x}_i|E_i, \mathcal{F})g(E_i|E_{min}, \eta_s)+ \\ (1-\delta)b(\mathbf{x}_i|\mathcal{G})g(E_i|E_{min}, \eta_b), \tag{1.8}$$

where $g(\cdot|e, \eta)$ is the density function of a **Pareto distribution** with scale $e = E_{min}$ and shape parameter η.

- This modelling approach might look as simplistic and does not properly reflect the high complexity of the data...

An extension that includes the energy

The previous model can be further extended as

$$f^{ext}(\mathbf{x}_i, E_i|\Theta^{ext}) = \delta s(\mathbf{x}_i|E_i, \mathcal{F})g(E_i|E_{min}, \eta_s)+ \\ (1-\delta)b(\mathbf{x}_i|\mathcal{G})g(E_i|E_{min}, \eta_b), \tag{1.8}$$

where $g(\cdot|e, \eta)$ is the density function of a **Pareto distribution** with scale $e = E_{min}$ and shape parameter η.

- This modelling approach might look as simplistic and does not properly reflect the high complexity of the data...
- ...however, we believe it is useful in a first stage of the analysis to explore whether the energy variable helps increasing the detection performance of the model.

Dealing with complex posterior distributions

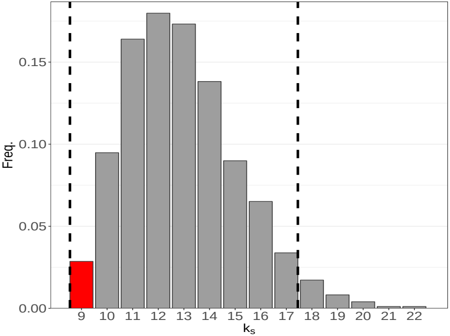


Figure 1: distribution of the number of sources on a simulation experiment

(9 is the true value).

- The posterior distribution of k_s contains the information about the number of sources.

Dealing with complex posterior distributions

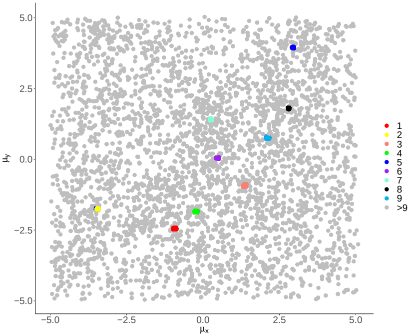


Figure 2: Scatterplot of $\boldsymbol{\mu}_j^{(t)} = (\mu_{jx}^{(t)}, \mu_{jy}^{(t)})$, for $j = 1, \ldots, 12$ and for every iteration t.

- The posterior distribution of k_s contains the information about the number of sources.
- Regions of the map with a large concentration of posterior draws are likely to contain at least a source.

Dealing with complex posterior distributions

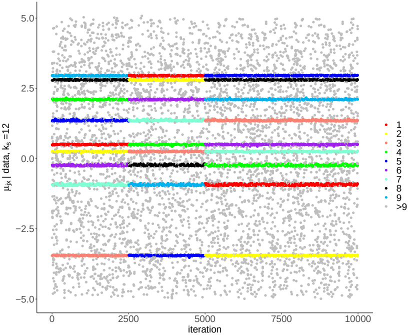


Figure 2: Traceplot of $\mu_{jx}^{(t)}$, for $j = 1, \dots, 12$ and for every iteration t.

- The posterior distribution of k_s contains the information about the number of sources.
- Regions of the map with a large concentration of posterior draws are likely to contain at least a source.
- The posterior distribution of some source locations $\boldsymbol{\mu}_k$ is multimodal.

Dealing with complex posterior distributions

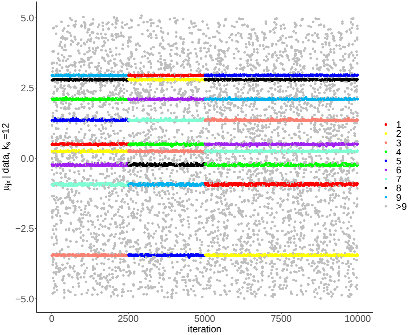


Figure 2: Traceplot of $\mu_{jx}^{(t)}$, for $j = 1, \dots, 12$ and for every iteration t.

- The posterior distribution of k_s contains the information about the number of sources.
- Regions of the map with a large concentration of posterior draws are likely to contain at least a source.
- The posterior distribution of some source locations $\boldsymbol{\mu}_k$ is multimodal.
- We further notice the label switching effect.

Dealing with complex posterior distributions

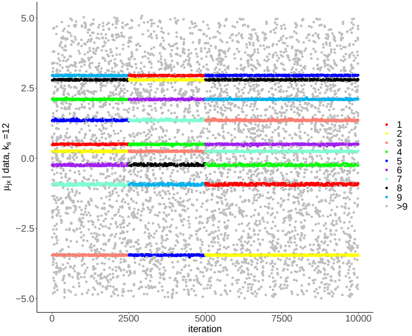
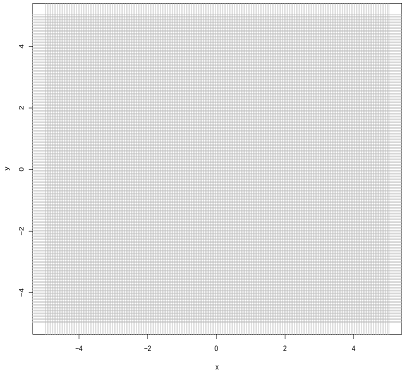


Figure 2: Traceplot of $\mu_{jx}^{(t)}$, for $j = 1, \ldots, 12$ and for every iteration t.

- The posterior distribution of k_s contains the information about the number of sources.
- Regions of the map with a large concentration of posterior draws are likely to contain at least a source.
- The posterior distribution of some source locations $\boldsymbol{\mu}_k$ is multimodal.
- We further notice the label switching effect.
- We develop a **post-processing algorithm** to extract the relevant information from the posterior distribution of $\boldsymbol{\mu}$.

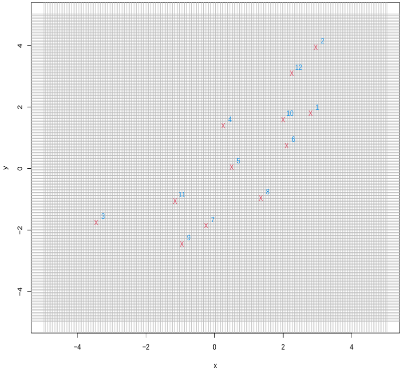
The post-processing algorithm

- Divide the map into small, rectangular pixels;



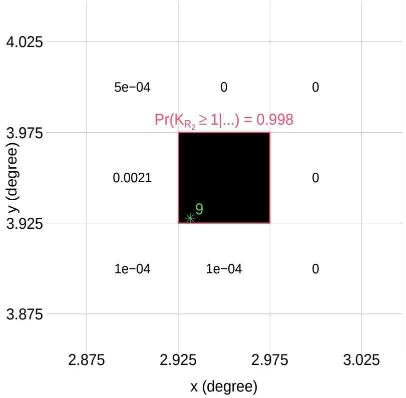
The post-processing algorithm

- Divide the map into small, rectangular pixels;
- Determine the first k^* modal pixels and call them $\mathcal{R}_1, \ldots, \mathcal{R}_{k^*}$;



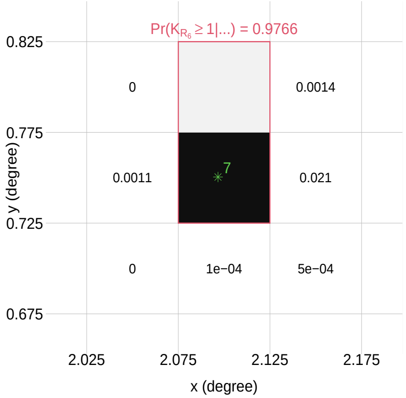
The post-processing algorithm

- Divide the map into small, rectangular pixels;
- Determine the first k^* modal pixels and call them $\mathcal{R}_1, \ldots, \mathcal{R}_{k^*}$;
- For each $\mathcal{R}$, compute $\Pr(K_{\mathcal{R}_m} \geq 1 | \ldots)$, the **probability of containing at least one source**, and eventually enlarge the region until the threshold of 0.95 is reached.



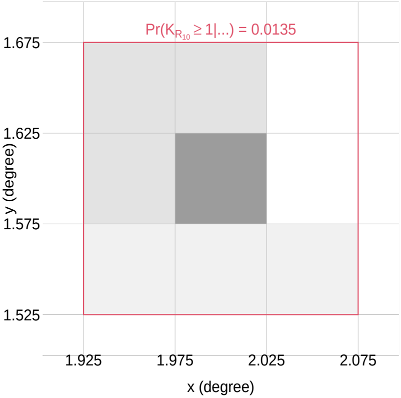
The post-processing algorithm

- Divide the map into small, rectangular pixels;
- Determine the first k^* modal pixels and call them $\mathcal{R}_1, \ldots, \mathcal{R}_{k^*}$;
- For each $\mathcal{R}$, compute $\Pr(K_{\mathcal{R}_m} \geq 1|\ldots)$, the **probability of containing at least one source**, and eventually enlarge the region until the threshold of 0.95 is reached.



The post-processing algorithm

- Divide the map into small, rectangular pixels;
- Determine the first k^* modal pixels and call them $\mathcal{R}_1, \ldots, \mathcal{R}_{k^*}$;
- For each $\mathcal{R}$, compute $\Pr(K_{\mathcal{R}_m} \geq 1 | \ldots)$, the **probability of containing at least one source**, and eventually enlarge the region until the threshold of 0.95 is reached.



Simulation experiments

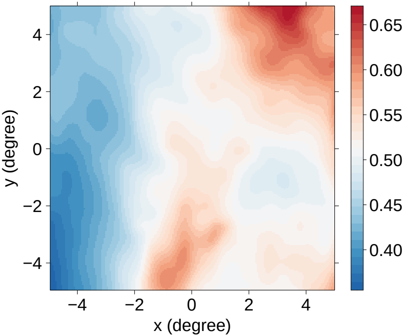
- Simulate the number of photons from a source s with location $\boldsymbol{\mu}_s$ and power-law spectrum with parameters $(F_{0,s}, \varrho_s)$

$$\text{power-law}_s = F_{0,s} \left(\frac{E}{1\mathrm{GeV}} \right)^{-\varrho_s}$$

$F_{0,s}$ and ϱ_s are simulated according to Abdo et al. (2010).

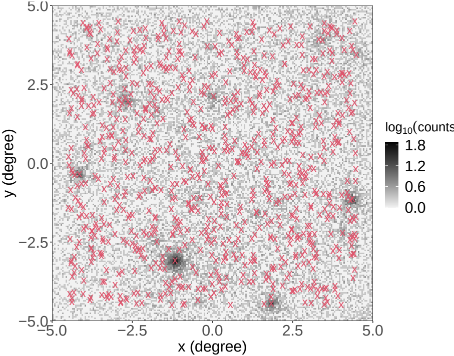
Simulation experiments

- Simulate the number of photons from a source s with location $\boldsymbol{\mu}_s$ and power-law spectrum with parameters $(F_{0,s}, \varrho_s)$
- Simulate the background contamination from the model of Acero et al. (2015).



Simulation experiments

- Simulate the number of photons from a source s with location $\boldsymbol{\mu}_s$ and power-law spectrum with parameters $(F_{0,s}, \varrho_s)$
- Simulate the background contamination from the model of Acero et al. (2015).
- We simulated 20 different maps, each of which with more than 800 sources.



The *Signal-to-Noise Ratio*

- We define as *signal-to-noise ratio* of a source the quantity

$$R_s = \sum_{i,j,k} \frac{\Lambda^s(i,j,k;\boldsymbol{\mu}_s, F_{0,s}, \varrho_S)}{\Lambda^b(i,j,k)},$$

where i,j,k refers to the ij-th spatial pixel and the k-th energy pixel,

$$\Lambda^s(i,j,k;\boldsymbol{\mu}, F_0, \varrho) = \mathrm{PSF}(i,j|k,\boldsymbol{\mu}_s) \cdot F_{0,s} \left(\frac{E_k}{1\mathrm{GeV}}\right)^{-\varrho_s} \cdot \epsilon(E_k),$$

$\epsilon(E_k)$ is the exposure, and $\Lambda^b(i,j,k)$ is the background model of Acero et al.

The *Signal-to-Noise Ratio*

- We define as *signal-to-noise ratio* of a source the quantity

$$R_s = \sum_{i,j,k} \frac{\Lambda^s(i,j,k;\boldsymbol{\mu}_s, F_{0,s}, \varrho_S)}{\Lambda^b(i,j,k)},$$

where i, j, k refers to the ij-th spatial pixel and the k-th energy pixel,

$$\Lambda^s(i,j,k;\boldsymbol{\mu}, F_0, \varrho) = \mathrm{PSF}(i,j|k,\boldsymbol{\mu}_s) \cdot F_{0,s} \left(\frac{E_k}{1\mathrm{GeV}}\right)^{-\varrho_s} \cdot \epsilon(E_k),$$

$\epsilon(E_k)$ is the exposure, and $\Lambda^b(i,j,k)$ is the background model of Acero et al.

- The larger is R_s, the more intensive is the signal of the source with respect to the background.

Simulation experiments - Results

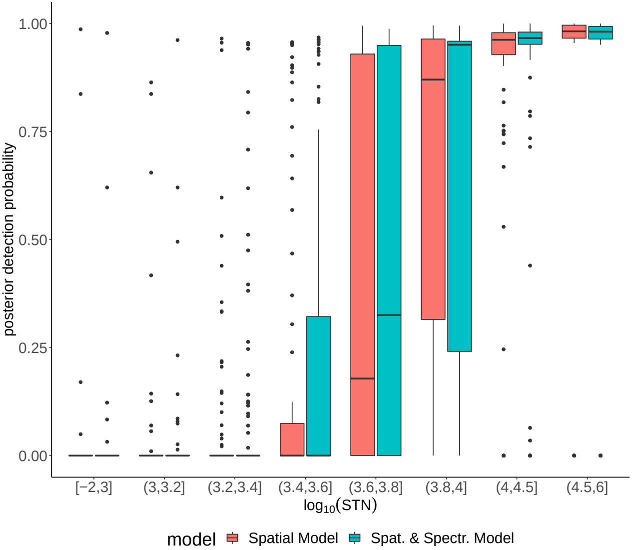
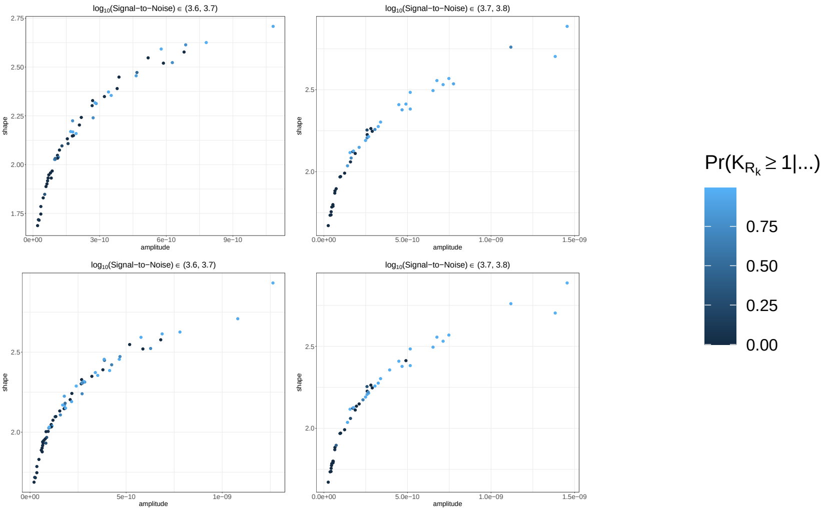


Figure 1: Results obtained over the 20 simulated maps.

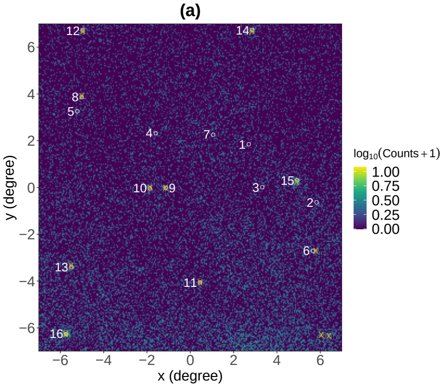
Simulation experiments - Results



Top: Spatial Model results. Bottom: Spatial-and-Spectral Model results

Application to *Antlia 2* Fermi LAT data

- $\sim$22.000 photons available within the energy range $[0.5\text{GeV}, 300\text{GeV}]$;
- The background contamination from our galaxy is visible in the bottom of the image;
- 16 sources are known to be in this area from 4FGL catalogue.



Application to *Antlia 2* Fermi LAT data - background fitting

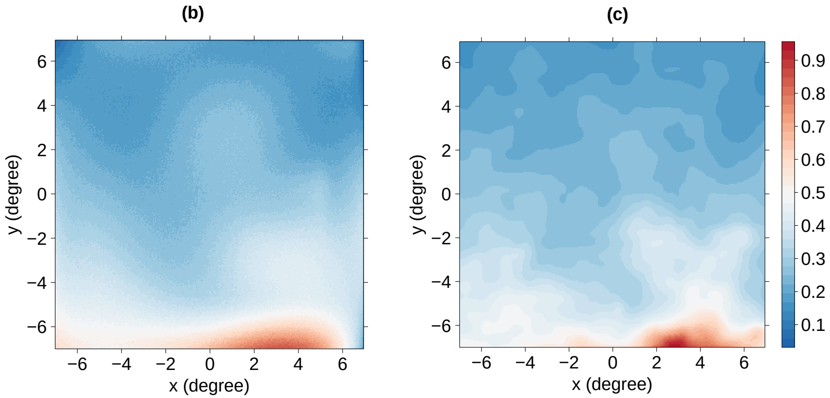
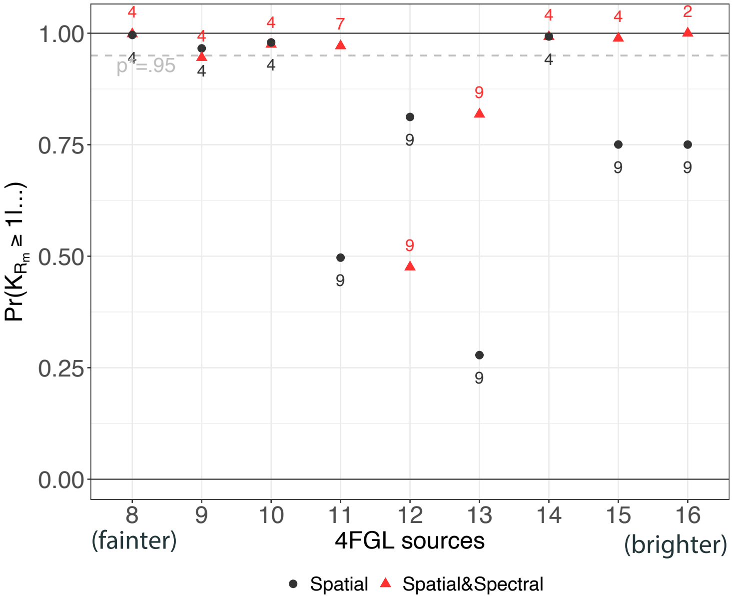


Figure 2: Left: expected posterior background density. Right: Acero et al. (2015)'s background model.

Application to *Antlia 2* Fermi LAT data - source detection



Bibliography

- Acero, F., Ackermann, M., Ajello, M., Albert, A., Atwood, W., Axelsson, M., Baldini, L., Ballet, J., Barbiellini, G., Bastieri, D. *et al.* (2015) Fermi large area telescope third source catalog. *The Astrophysical Journal Supplement Series* **218**(2), 23–63.
- Acero, F., Ackermann, M., Ajello, M., Albert, A., Baldini, L., Ballet, J., Barbiellini, G., Bastieri, D., Bellazzini, R., Bissaldi, E. *et al.* (2016) Development of the model of galactic interstellar emission for standard point-source analysis of fermi large area telescope data. *The Astrophysical Journal Supplement Series* **223**(2), 26–48.
- Carlson, B. C. (1991) B-splines, hypergeometric functions, and Dirichlet averages. *J. Approx. Theory* **67**(3), 311–325.
- Frühwirth-Schnatter, S. (2011). Label switching under model uncertainty. *Mixtures: Estimation and Application*, 213-239.
- Gaetan, C., and Guyon, X. (2010). *Spatial statistics and modeling* (Vol. 90). New York: Springer.
- Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005). Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association*, **100**(471), 1021-1035.

Bibliography

- Guglielmetti, F., Fischer, R. and Dose, V. (2009) Background-source separation in astronomical images with Bayesian probability theory - I. The method. *Monthly Notices of the Royal Astronomical Society* **396**(1), 165–190.
- Gupta, A. C., Tripathi, A., Wiita, P. J., Kushwaha, P., Zhang, Z., & Bambi, C. (2019). Detection of a quasi-periodic oscillation in γ-ray light curve of the high-redshift blazar B2 1520+31. *Monthly Notices of the Royal Astronomical Society*, **484**(4), 5785-5790.
- Jones, D. E., Kashyap, V. L. and van Dyk, D. A. (2015) Disentangling overlapping astronomical sources using spatial and spectral information. *The Astrophysical Journal* **808**(2), 137–160.
- Kelly, B. C., Bechtold, J. and Siemiginowska, A. (2009) Are the variations in quasar optical flux driven by thermal fluctuations? *The Astrophysical Journal* **698**(1), 895-910.
- Knoetig, M. L. (2014) Signal discovery, limits, and uncertainties with sparse on/off measurements: An objective Bayesian analysis. *The Astrophysical Journal* **790**(2), 106–113.
- Malsiner-Walli, G., Frühwirth-Schnatter, S. and Grün, B. (2016) Model-based clustering based on sparse finite gaussian mixtures. *Statistics and Computing* **26**(1), 303–324.

Bibliography

- Mattox, J. R., Bertsch, D., Chiang, J., Dingus, B., Digel, S., Esposito, J., Fierro, J., Hartman, R., Hunter, S., Kanbach, G. *et al.* (1996) The likelihood analysis of EGRET data. *The Astrophysical Journal* **461**, 396–407.
- Meyer, L., Witzel, G., Longstaff, F. and Ghez, A. (2014) A formal method for identifying distinct states of variability in time-varying sources: Sgr A* as an example. *The Astrophysical Journal* **791**(1), 24-32.
- Ornstein, L. S. and Uhlenbeck, G. E. (1930) *On the theory of the brownian motion. Physical Review (Series I)* **36**, 823-841.
- Park, T., Kashyap, V. L., Siemiginowska, A., van Dyk, D. A., Zezas, A., Heinke, C. and Wargelin, B. J. (2006) Bayesian estimation of hardness ratios: Modeling and computations. *The Astrophysical Journal* **652**(1), 610–628.
- Primini, F. A. and Kashyap, V. L. (2014) Determining x-ray source intensity and confidence bounds in crowded fields. *The Astrophysical Journal* **796**(1), 24–37.
- Protassov, R., van Dyk, D. A., Connors, A., Kashyap, V. L. and Siemiginowska, A. (2002) Statistics, handle with care: detecting multiple model components with the likelihood ratio test. *The Astrophysical Journal* **571**(1), 545–559.

Bibliography

- Ramakrishnan, V., Hovatta, T., Nieppola, E., Tornikoski, M., Lähteenmäki, A. and Valtaoja, E. (2015) Locating the γ-ray emission site in Fermi/LAT blazars from correlation analysis between 37GHz radio and γ-ray light curves. *Monthly Notices of the Royal Astronomical Society* **452**(2), 1280-1294.
- Robotham, A. S. G., Davies, L. J. M., Driver, S. P., Koushan, S., Taranu, D. S., Casura, S., & Liske, J. (2018). ProFound: source extraction and application to modern survey data. *Monthly Notices of the Royal Astronomical Society*, **476**(3), 3137-3159.
- Sobolewska, M. A., Siemiginowska, A., Kelly, B. C. and Nalewajko, K. (2014) Stochastic modeling of the Fermi/Lat γ-ray blazar variability. *The Astrophysical Journal* **786**(2), 143-156.
- van Dyk, D. A., Connors, A., Kashyap, V. L. and Siemiginowska, A. (2001) Analysis of energy spectra with low photon counts via Bayesian posterior simulation. *The Astrophysical Journal* **548**(1), 224–243.
- Weisskopf, M. C., Wu, K., Trimble, V., O'Dell, S. L., Elsner, R. F., Zavlin, V. E. and Kouveliotou, C. (2007) A Chandra search for coronal X-rays from the cool white dwarf gd 356. *The Astrophysical Journal* **657**(2), 1026–1036.