Search for hidden AGNs in the Fermi-LAT third source catalog with a data mining approach

Kenji Yoshida
Shibaura Institute of Technology

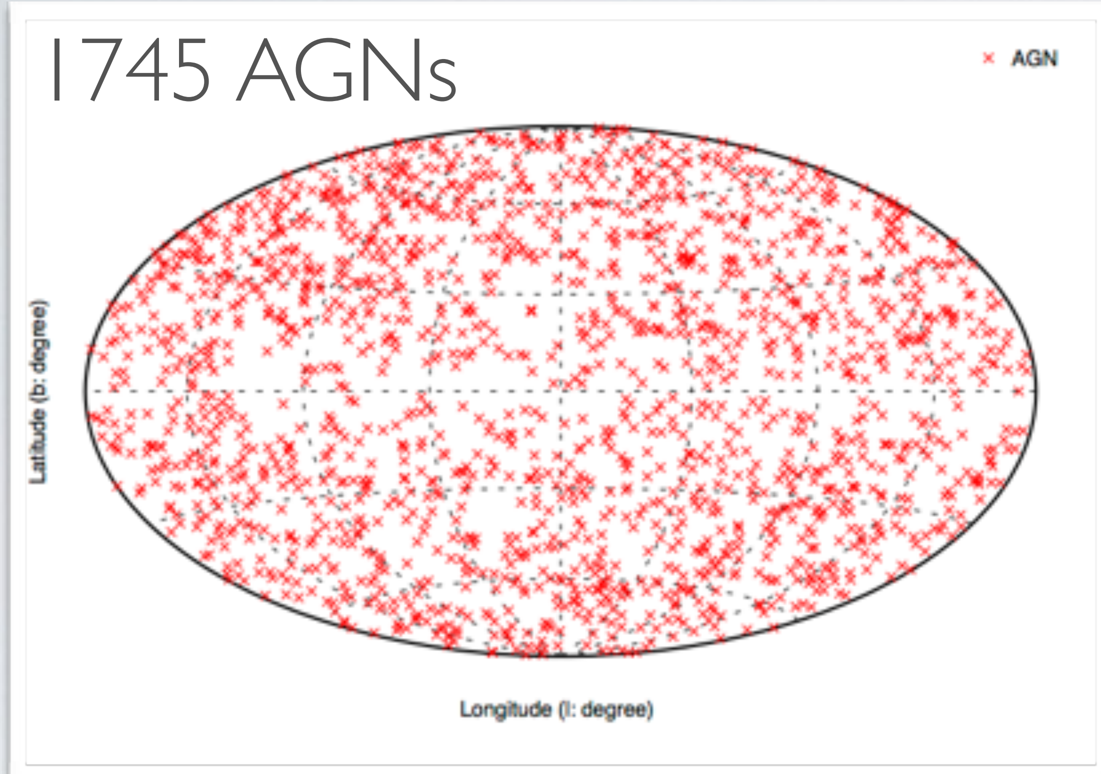
November 8, Patagonia Forum 2016

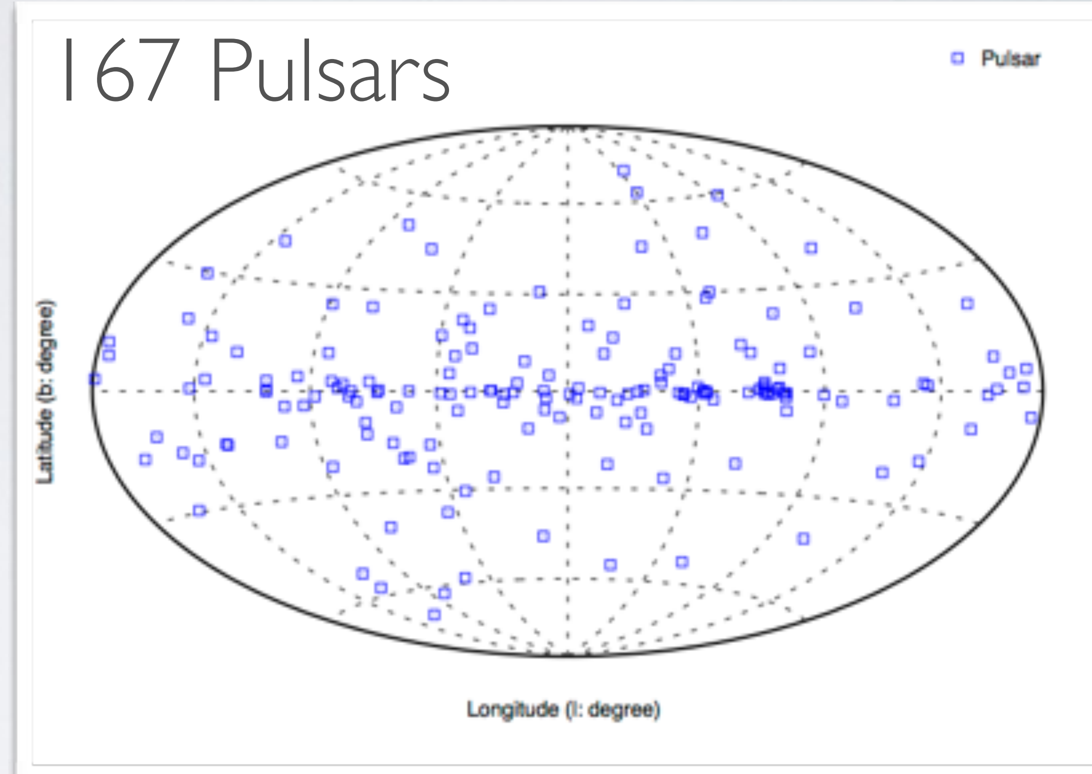
Fermi-LAT 3rd Source Catalog (3FGL)

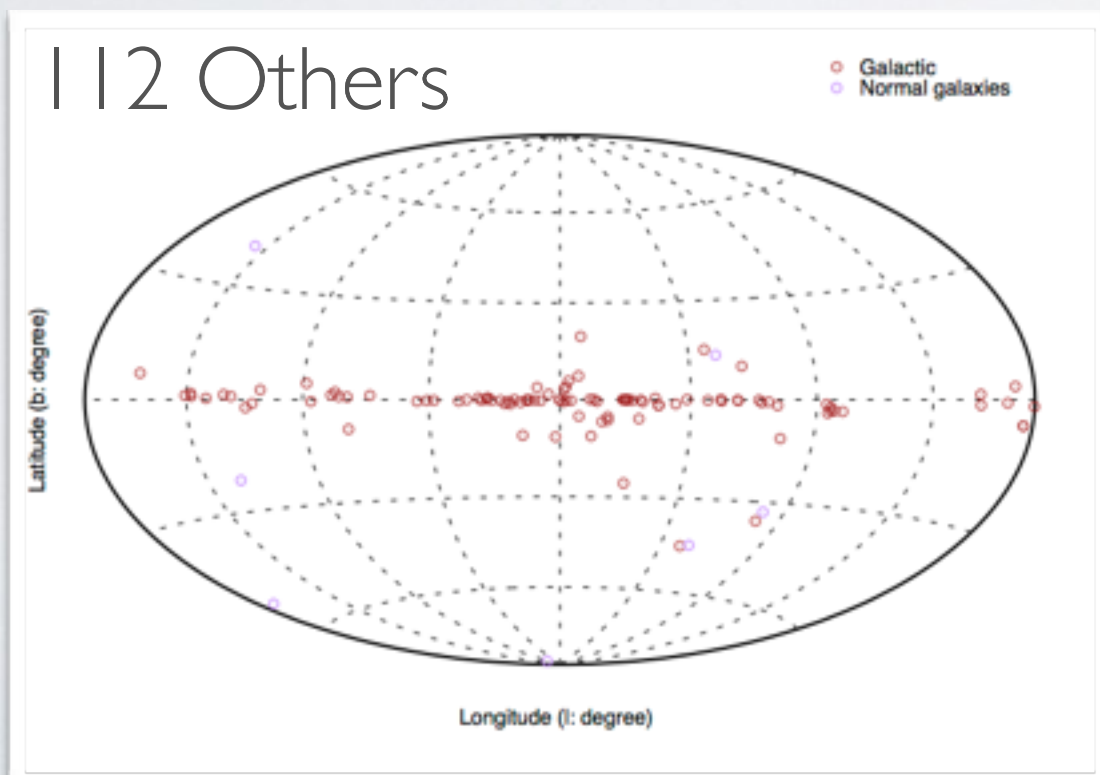
- Fermi-LAT (Large Area Telescope)
  - The principal scientific instrument on the Fermi Gamma Ray Space Telescope spacecraft
  - Energy Range: 0.1GeV — 300GeV
  - Angular resolution: 0.6deg @1GeV
  - Publication of Fermi-LAT 3rd source catalog (3FGL) (2008-2012)

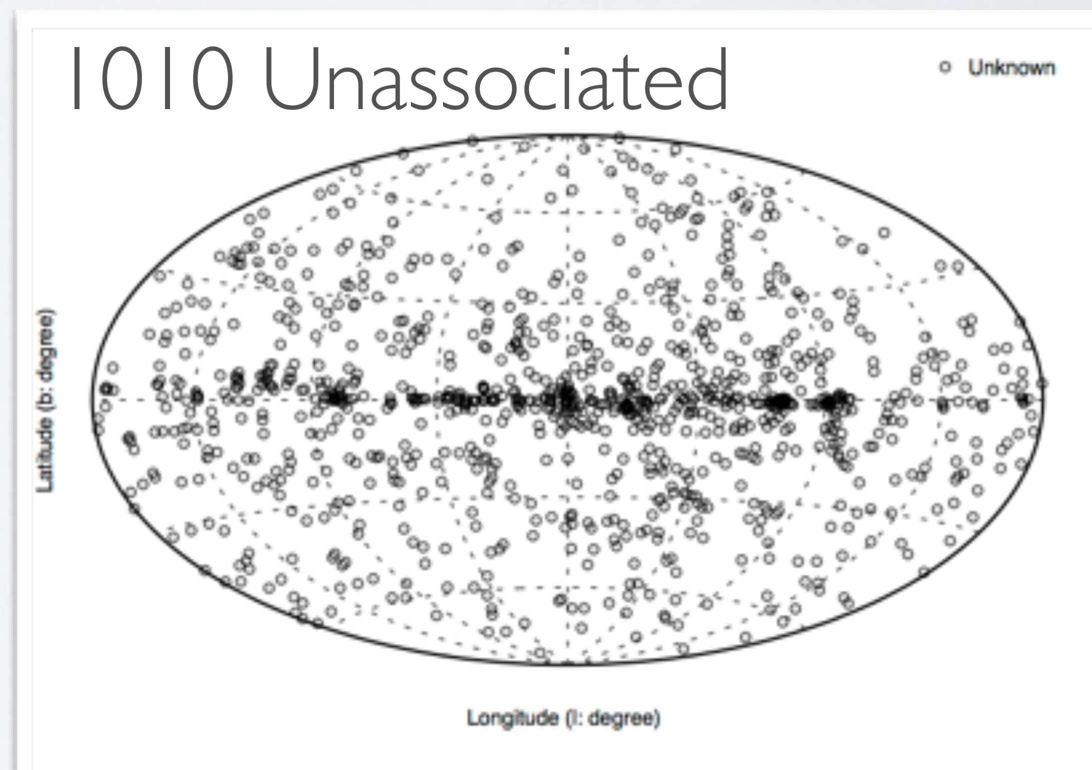
- 3FGL：3034 gamma-ray sources
  - 2024 Identified/Associated gamma-ray sources:
    - 1745 AGNs, 167 Pulsars, and 112 others
  - 1010 Unassociated gamma-ray sources

The 3FGL sources in the Galactic coordinates





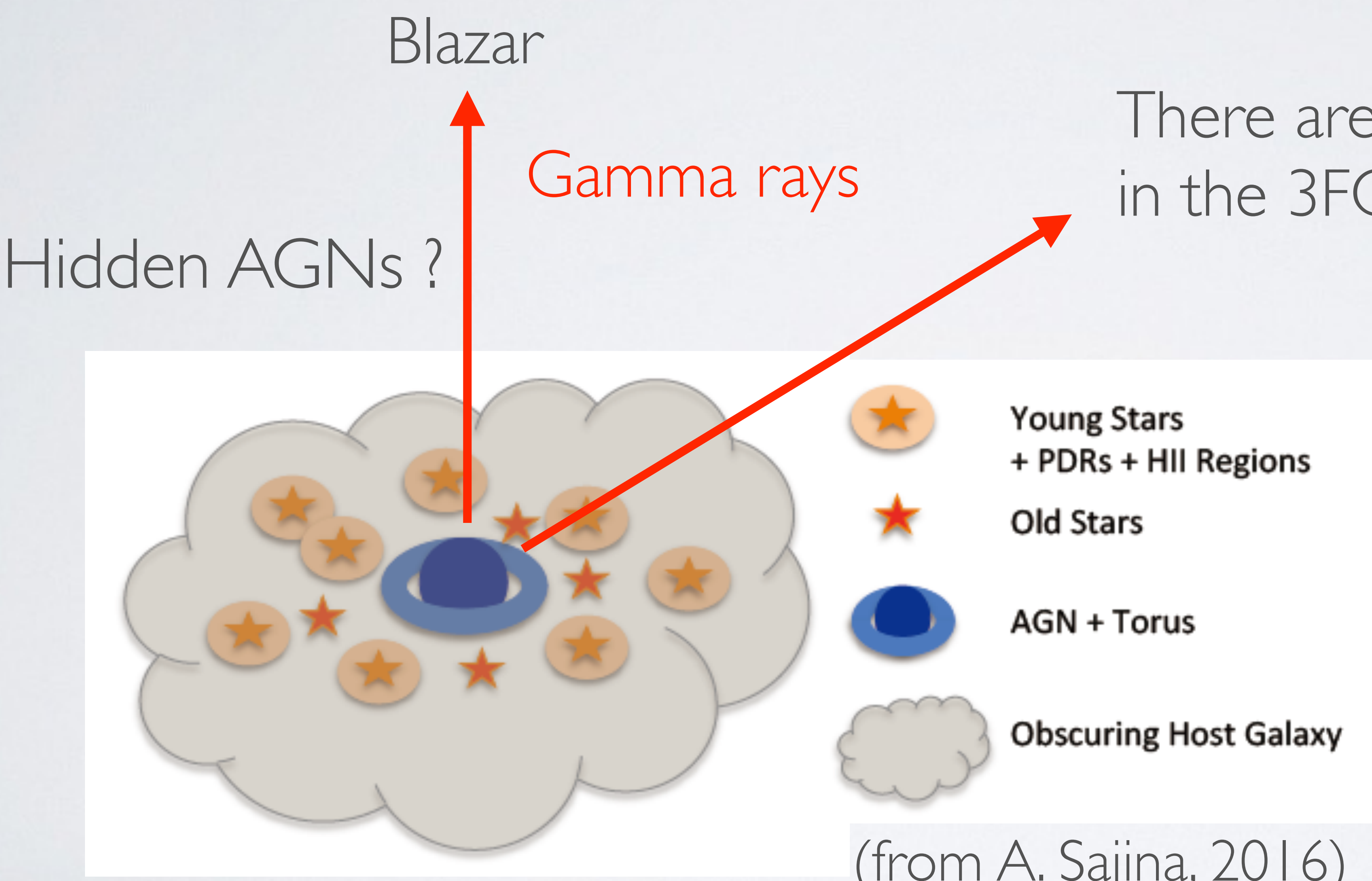




Search for Hidden AGNs

Gamma rays can penetrate obscuring materials of $N_H=10^{24}$ /cm^2

=> It might be possible to search for hidden AGNs with gamma-ray observations.



(from A. Sajina, 2016)

Hidden AGNs look like AGNs for gamma rays?

=> Recognizing AGN candidates in the 3FGL unassociated sources

Data Mining Approaches to Recognizing Source Classes for Unassociated Gamma-ray Sources in Fermi-LAT Source Catalog

- M. Ackermann et al. (2012)

  Boosted Decision Tree & Logistic Regression
- N. Mirabal et al. (2012)

  Random Forest
- M. Doer & M. Errando (2013)

  Neural Network & Random Forest
- J. Lefaucheur et al. (2015)

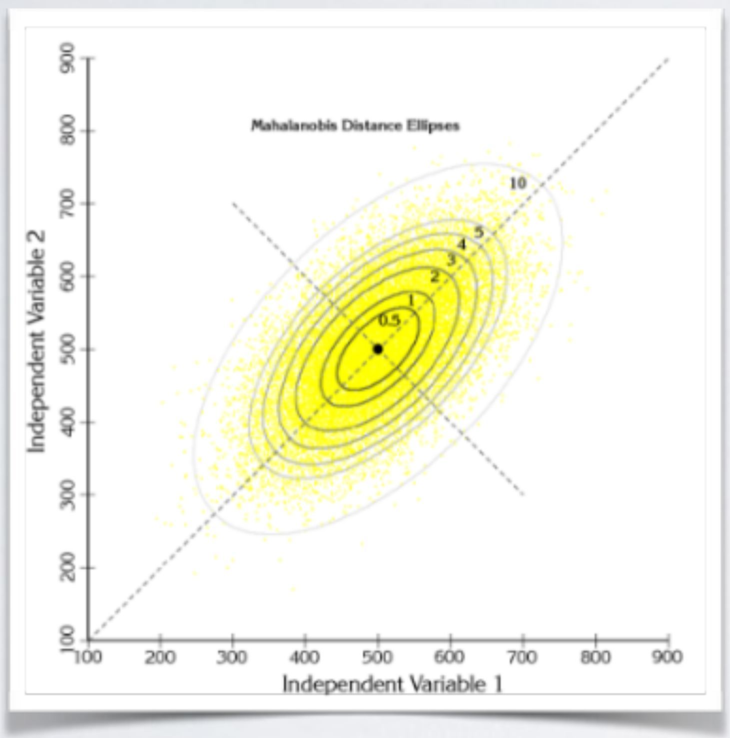
  Boosted Decision Tree & Neural Network

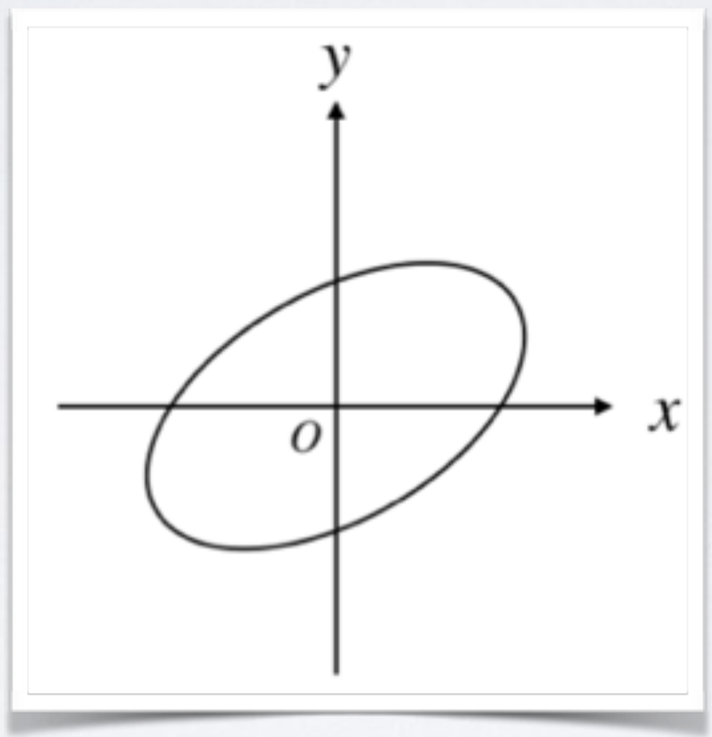
=> Application of a Mahalanobis-Taguchi method to recognizing source classes for unassociated gamma-ray sources in 3FGL

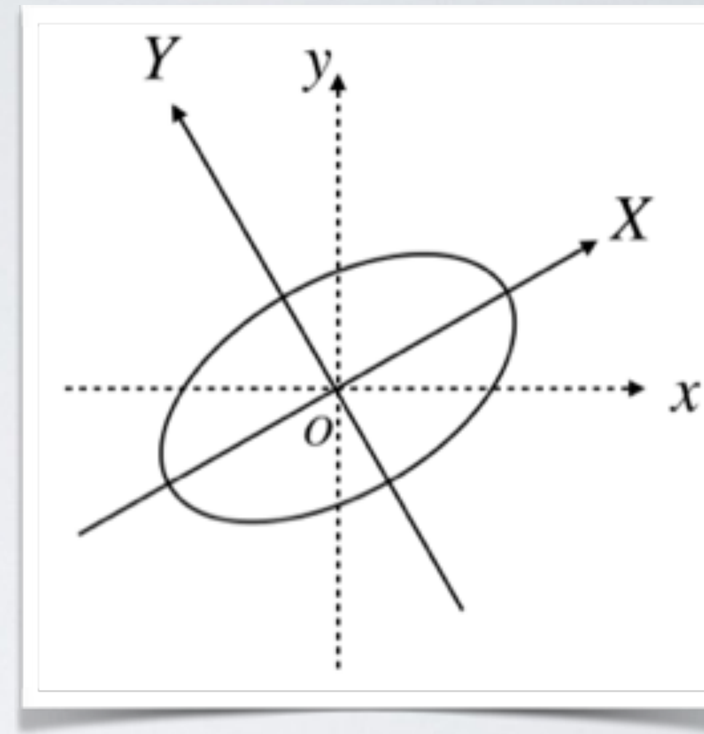
Mahalanobis-Taguchi Method

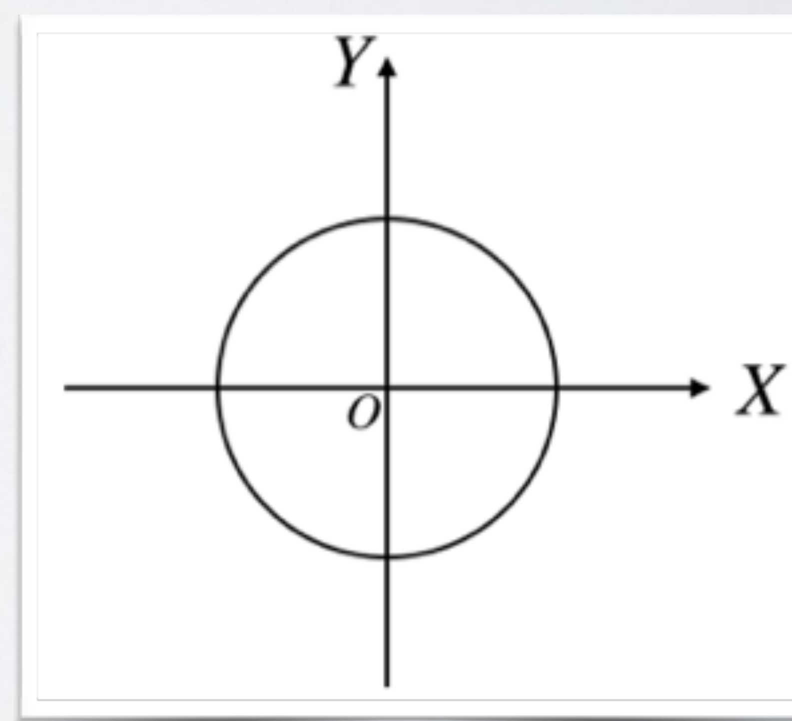
A pattern recognition technique

- Building a unit space from the normal data
- Normalizing comparative data in the unit space
- Deriving the Mahalanobis distance of each source









- Sources similar to the normal data have the Mahalanobis distance close to 1.
- Sources dissimilar to the normal data have the much larger Mahalanobis distance than 1.

Mahalanobis Distance: *D*

Mahalanobis distance

$$D^2 = (\frac{1}{k}) Z_i^T C^{-1} Z_i$$

where Z_i = standardized vector obtained by values of $X_i (i = 1, \ldots, k)$

$Z_i = (X_i - m_i)/s_i$

X_i =value of i-th characteristic

m_i =mean of i-th characteristic

s_i =standard deviation of i-th characteristic

k =the number of characteristics / variables

T =transpose of the vector

C^{-1} =inverse of the correlation matrix

Characteristics in the 3FGL catalog (1)

Adopting characteristics similar to Ackermann et al. (2012):
Characteristics independent of the source significance

Hardness Ratio

$$HR_{ij} = \frac{\nu F_{\nu_j} - \nu F_{\nu_i}}{\nu F_{\nu_j} + \nu F_{\nu_i}}.$$

i, j: Energy bands
1 (0.1-0.3GeV), 2 (0.3-1GeV), 3 (1-3GeV), 4 (3-10GeV), 5 (10-100GeV)

Fractional Variability

$$\frac{\delta F}{F} = \sqrt{\frac{\Sigma_i (F_i - F_{av})^2}{(N_{int} - 1) F_{av}^2} - \frac{\Sigma_i \sigma_i^2}{N_{int} F_{av}^2} - f_{rel}^2},$$

N_{int} = 48 in 3FGL
F_{av} = Average flux
σ_i = Statistical error of F_i
f_{rel} = Systematic error (2%)

Characteristics in the 3FGL catalog (2)

Spectral Index Γ Best fit photon number power-law index for the following spectra:

$$\frac{\mathrm{d}N}{\mathrm{d}E} = K\left(\frac{E}{E_0}\right)^{-\Gamma}$$

$$\frac{\mathrm{d}N}{\mathrm{d}E} = K\left(\frac{E}{E_0}\right)^{-\alpha-\beta\log(E/E_0)}$$

$$\frac{\mathrm{d}N}{\mathrm{d}E} = K\left(\frac{E}{E_0}\right)^{-\Gamma}\exp\left(\left(\frac{E_0}{E_c}\right)^{b} - \left(\frac{E}{E_c}\right)^{b}\right)$$

Normalized Curvature = Significance Curve/Significance Average

Significance Average: Source significance in σ units over the 100 MeV to 300 GeV band

Significance Curve: Significance in σ units of the fit improvement between power law and the cut-off type spectra

Galactic longitude l、 Galactic latitude b

Recognition as AGNs and non-AGNs

Characteristics

$$HR_{34}, \quad HR_{45}, \quad \log(\frac{\delta F}{F}), \quad \Gamma, \quad \text{Normalized Curvature}, \quad |\ell|, \quad \log(|b|).$$

Building the unit space from the AGN data

Deriving the Mahalanobis distances of AGNs and non-AGNs in the AGN unit space
=> Evaluating the recognition power

Deriving the Mahalanobis distances of the unassociated sources in the AGN unit space
=> Recognizing the unassociated sources as AGNs and non-AGNs

Recognition as Pulsars and non-Pulsars

Characteristics

$$HR_{23} - HR_{34}, \quad HR_{45}, \quad \frac{\delta F}{F}, \quad \ell \times b.$$

Building the unit space from the Pulsar data

Deriving the Mahalanobis distances of Pulsars and non-Pulsars in the Pulsar unit space
=> Evaluating the recognition power

Deriving the Mahalanobis distances of the unassociated sources in the Pulsar unit space
=> Recognizing the unassociated sources as Pulsars and non-Pulsars

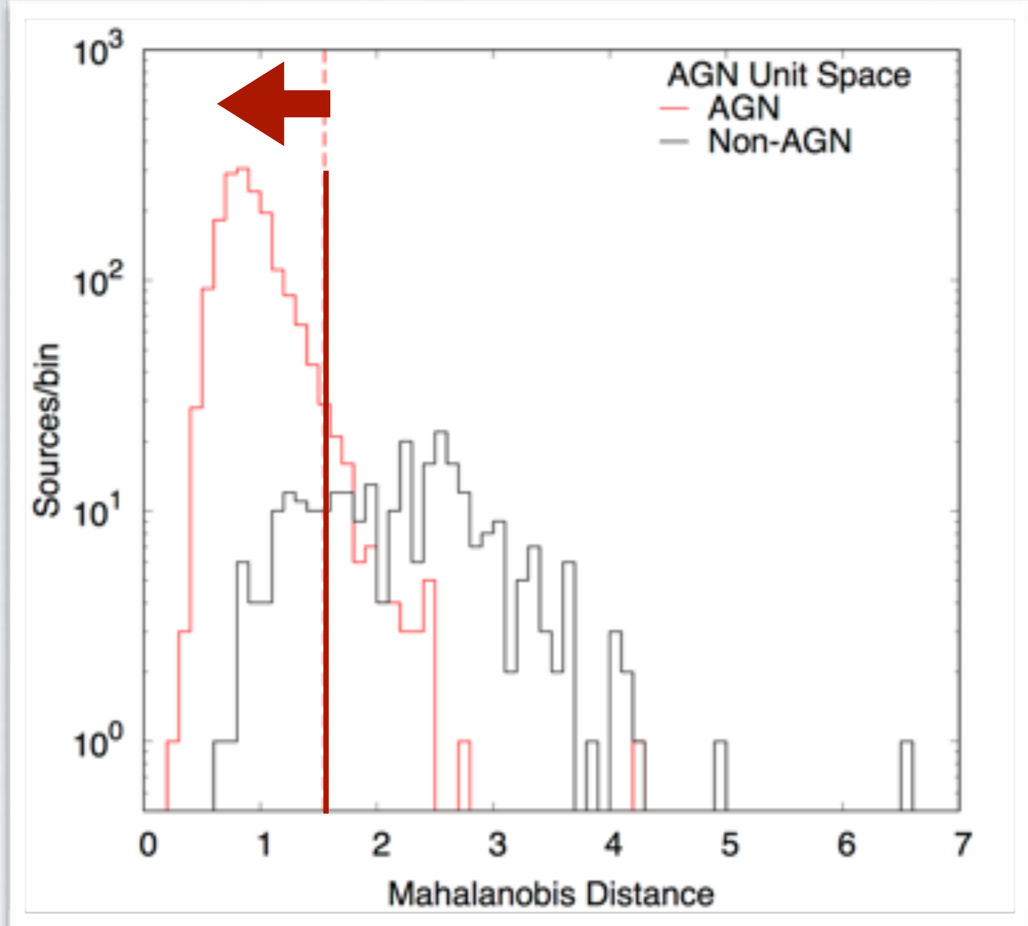
A method for evaluation of the AGN recognition power

Identified/Associated sources: 1745 AGNs, 279 non-AGNs in 3FGL

- Evaluation of the AGN recognition power
  - Use of 4/5 AGN data (1397 AGNs) for building the AGN unit space
  - Use of 1/5 AGN data (349 AGNs) for evaluation
  - Repeating 5 times with different data sets to derive the Mahalanobis distances of 1745 AGNs
- Evaluation of the non-AGN recognition power
  - Deriving the Mahalanobis distances of 279 non-AGNs data in the AGN unit space

=> The similar evaluation method for the Pulsar recognition power

Mahalanobis Distances of the Identified/Associated Sources — AGN Unit Space —



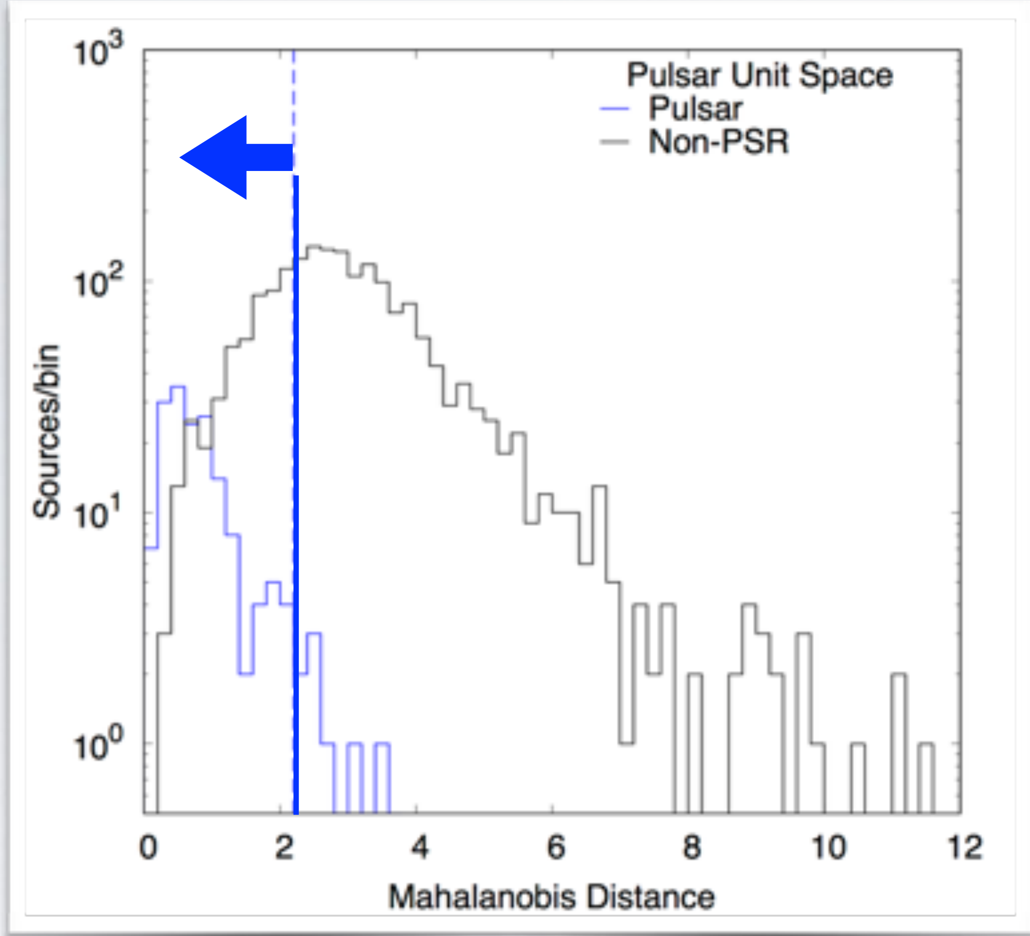
$D_A < 1.16$
- 80% AGNs
- 7.2% (20/279) non-AGNs

$D_A < 1.55$
- 95% AGNs
- 22.2% (62/279) non-AGNs

Ranking of the characteristics

Variable	Gain (db)
log(|b|)	4.541
|l|	3.206
log(δF/F)	2.750
HR_{45}	2.605
Γ	1.181
Norm. Curvature	0.990
HR_{34}	0.421

Mahalanobis Distances of the Identified/Associated Sources — Pulsar Unit Space —



D_P<1.16
80% Pulsars
4.5% (84/1857) non-Pulsars

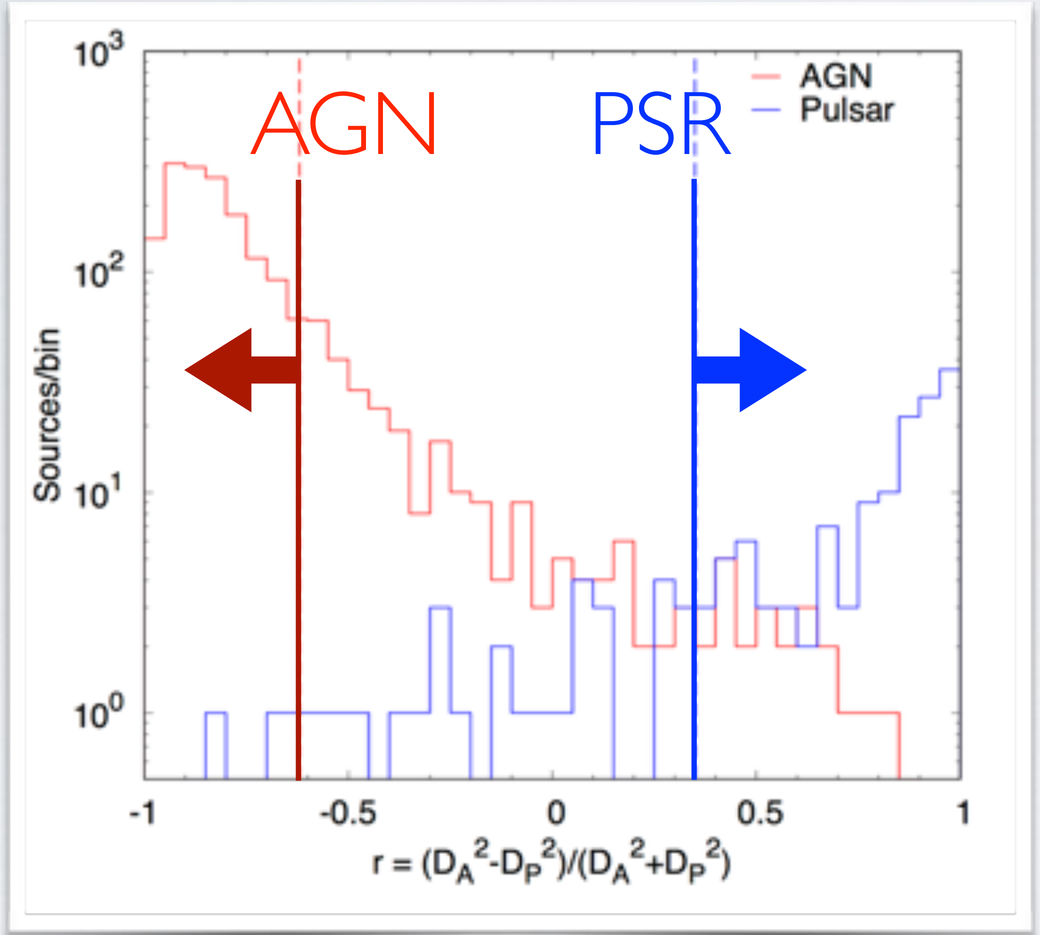
D_A<2.20
95% Pulsars
26.4% (490/1857) non-Pulsars

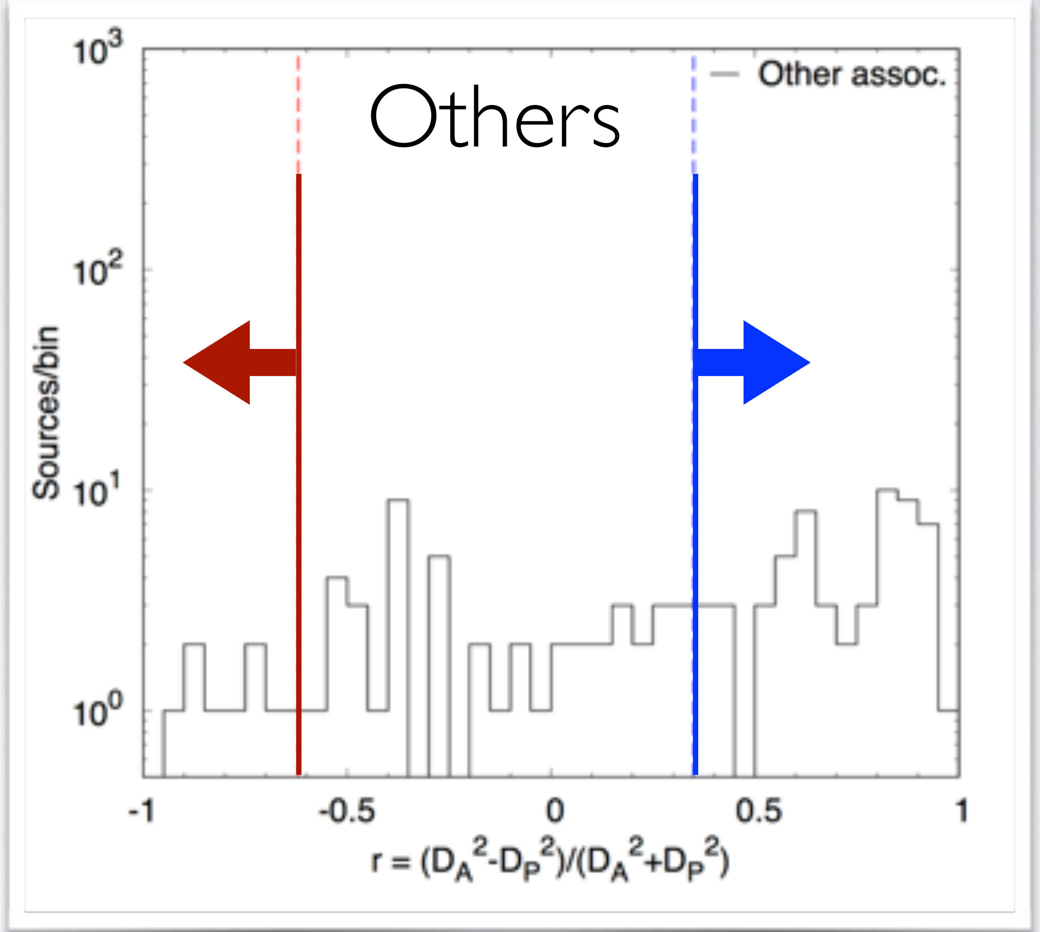
Ranking of the characteristics

Variables	Gain (db)
δF/F	11.372
HR_{23}-HR_{34}	9.347
HR_{45}	-0.877
l x b	-5.935

Recognition Power for the Identified/Associated Sources

Mahalanobis Distance in AGN Unit Space = D_A
Mahalanobis Distance in Pulsar Unit Space = D_P => $r = \frac{D_A^2 - D_P^2}{D_A^2 + D_P^2}$





D_A<1.55 & r<-0.62

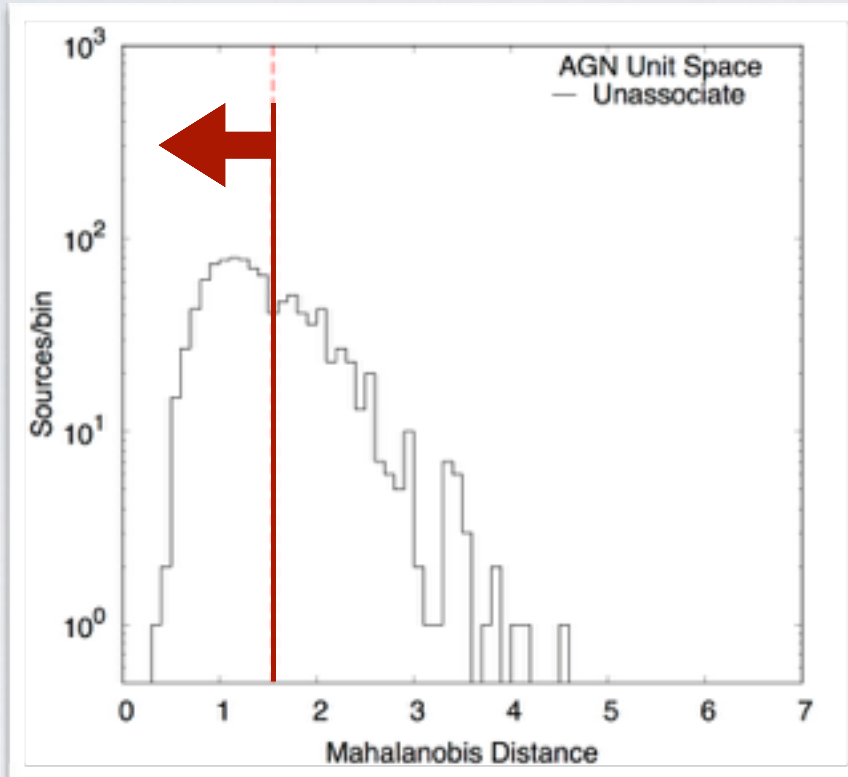
AGN 80.5% (1405/1745), Pulsar 1.2% (2/167), Others 7.1% (8/112)

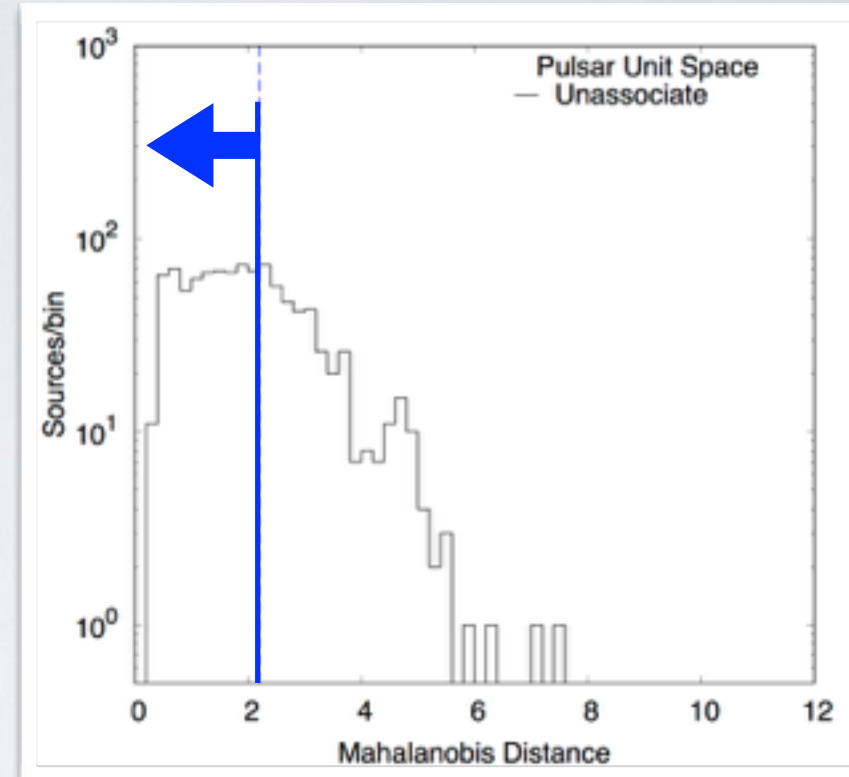
D_P<2.20 & r>0.35

Pulsar 80.8% (135/167), AGN 1.3% (22/1745), Others 49.1% (55/112)

Recognition of the Unassociated Sources

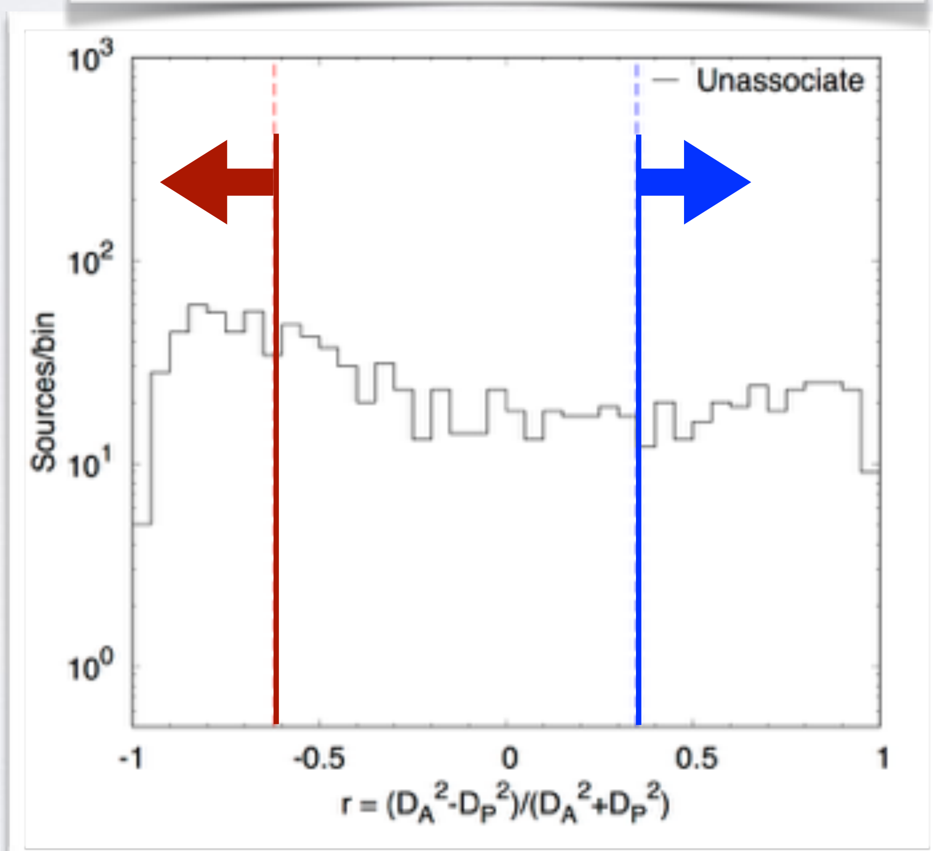
Distributions of Mahalanobis Distances in AGN and Pulsar unit spaces





Distributions of

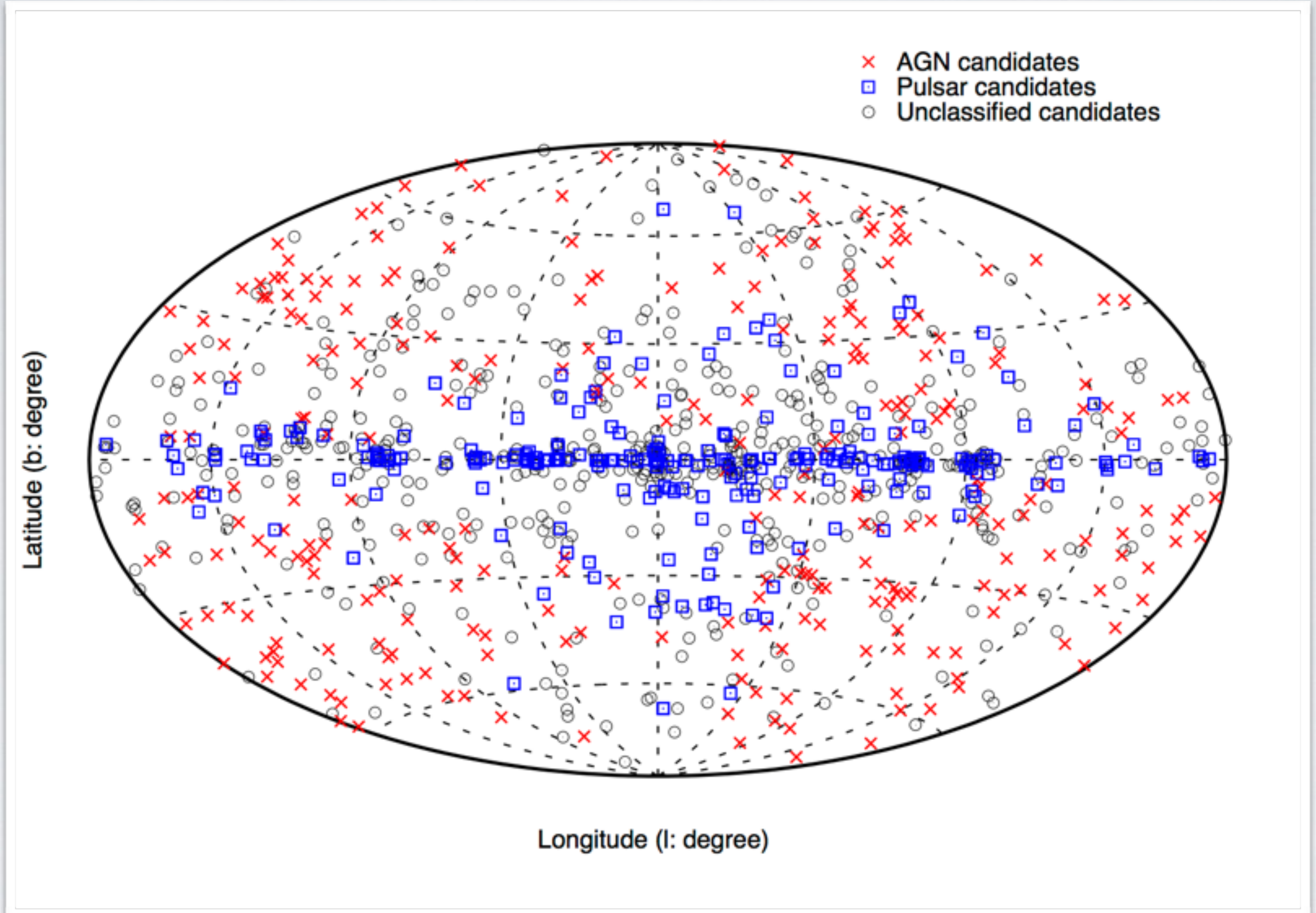
$$r = \frac{D_A^2 - D_P^2}{D_A^2 + D_P^2}$$



D_A<1.55 & r<-0.62 => recognized to be AGN

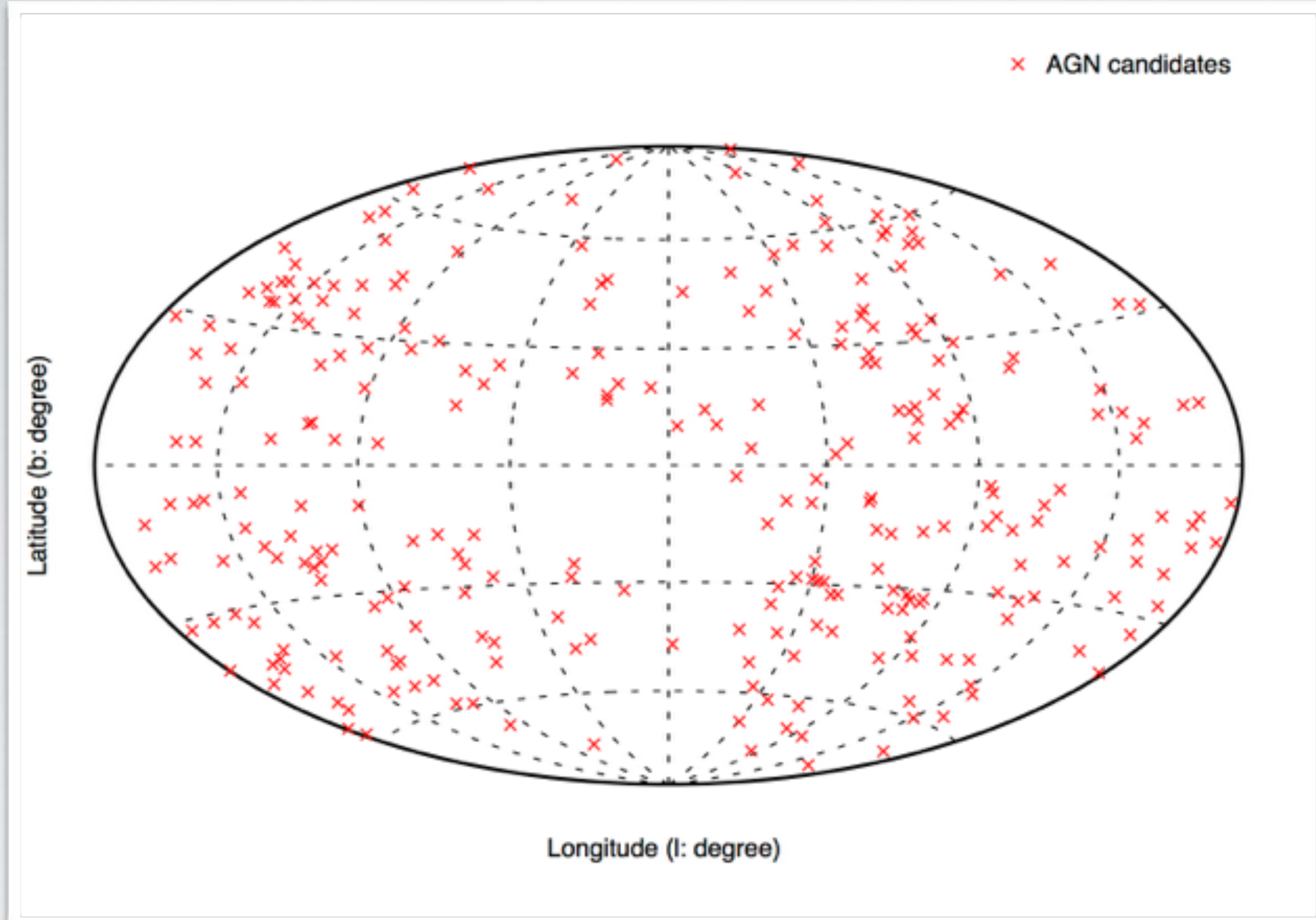
D_P<2.20 & r>0.35 => recognized to be Pulsar

Recognition of the unassociated sources in 3FGL



274 AGN candidates, 247 Pulsar candidates, 489 others candidates

AGN candidates in the 3FGL unassociated sources



Contamination rate in the identified/associated sources of 3FGL:

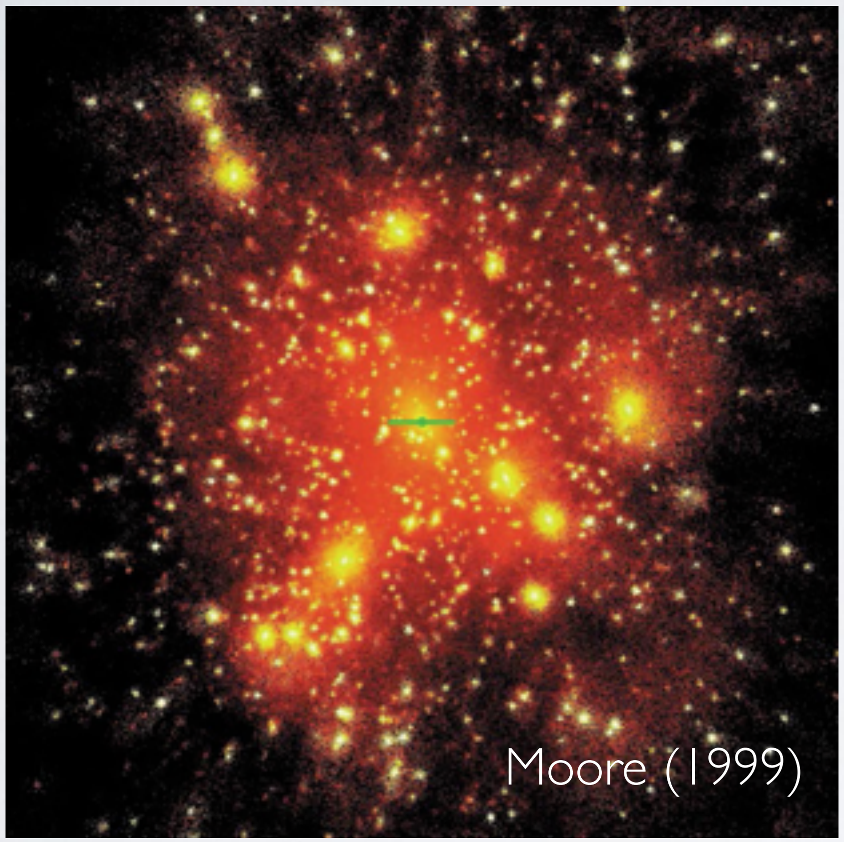
- Pulsar: 2/167
- Starburst galaxy: 4/4
- Pulsar wind nebula: 3/60
- Globular cluster: 1/15

=> Hidden AGNs in these AGN candidates ?

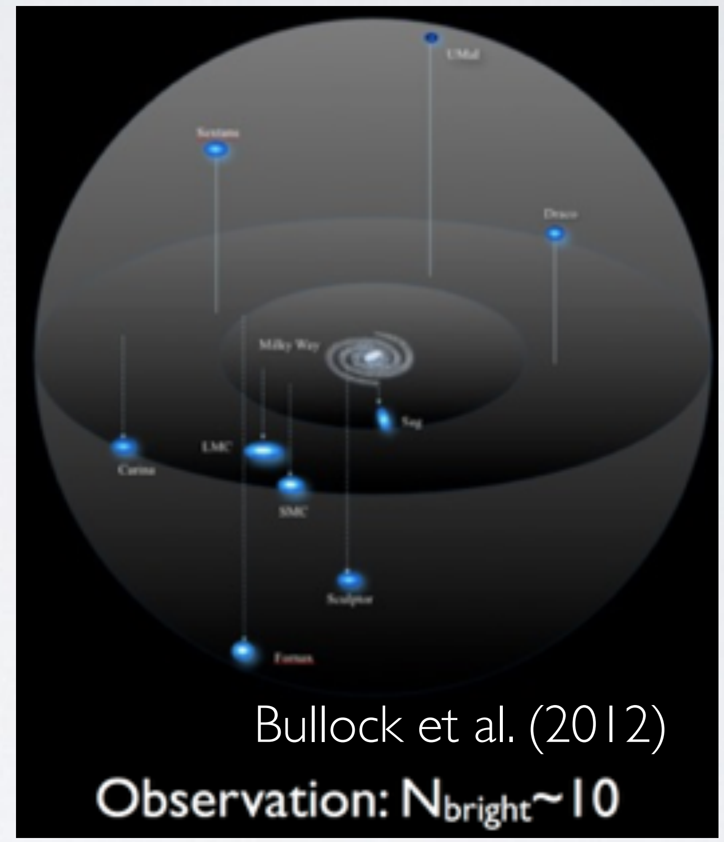
The Other topic: Dark Matter Sub-Halo

"Missing Dwarf Satellite Galaxies Problem"

Simulation：N>>1000



Observation：N~10



- Dark matter sub-halos ?
  => annihilated gamma rays from WIMP dark matter
- The 3FGL might contain on the order of ~10 sources which are dark matter sub-halos (Bertoni et al. 2015)

Search for dark matter sub-halos

At high Galactic latitude $|b|>5°$,
the large Mahalanobis distances of $D_A>2.47$ and $D_P>3.22$

- Identified/Associated gamma-ray sources: 0/2014
- Unassociated gamma-ray sources: 4/1010

=> Outlier sources in the 3FGL ?

Dark matter sub-halo candidates ?

Source	ℓ	b	D (AGN)	D (pulsar)
3FGLJ1258.4+2123	-41.094	84.038	2.989	4.893
3FGLJ1616.8+5846	89.516	42.688	2.591	5.097
3FGLJ1729.9-0859	12.518	13.576	2.791	4.575
3FGLJ2014.5+5246	87.889	9.900	2.973	4.823

Summary

- Recognizing the 3FGL gamma-ray sources to be AGNs, Pulsars, and the others with a Mahalanobis-Taguchi method:
  - Efficiency rate for AGNs = 80.5%, error rate for pulsars = 1.2%
  - Efficiency rate for Pulsars = 80.8%, error rate for AGNs = 1.3%
    => as well or better than the previous results (e.g. 3% and 1% for each error rate, M.Ackermann et al. 2012)
- Search for hidden AGNs
  - Hidden AGNs might be included in the 274 AGN candidates of the 3FGL unassociated sources.
- Search for dark matter sub-halos at high Galactic latitude |b|>5°
  - 4 unassociated sources with large Mahalanobis distance