

データマイニングによる
FERMI-LAT点源カタログ3FGL
未同定ガンマ線源の識別

吉田健二

芝浦工大 シス理工

Fermi-LAT 3rdソースカタログ(3FGL)

- Fermi-LAT
 - ガンマ線天文衛星Fermi (2008年 — 現在) に搭載された大面積望遠鏡 (Large Area Telescope: LAT)
 - 観測エネルギー範囲: 0.1 GeV — 300 GeV
 - ガンマ線源のカタログを公開:
 - Fermi-LAT 3rdソースカタログ (3FGL) (2008 - 2012年)
- 3FGL: 3034個のガンマ線源
 - 同定ガンマ線源: 2024個
 - 内訳: AGNs 1745個、Pulsars 167個、他 112個
 - 未同定ガンマ線源: 1010個

データマイニング手法を適用した FERMI-LATカタログ未同定ガンマ線源の識別

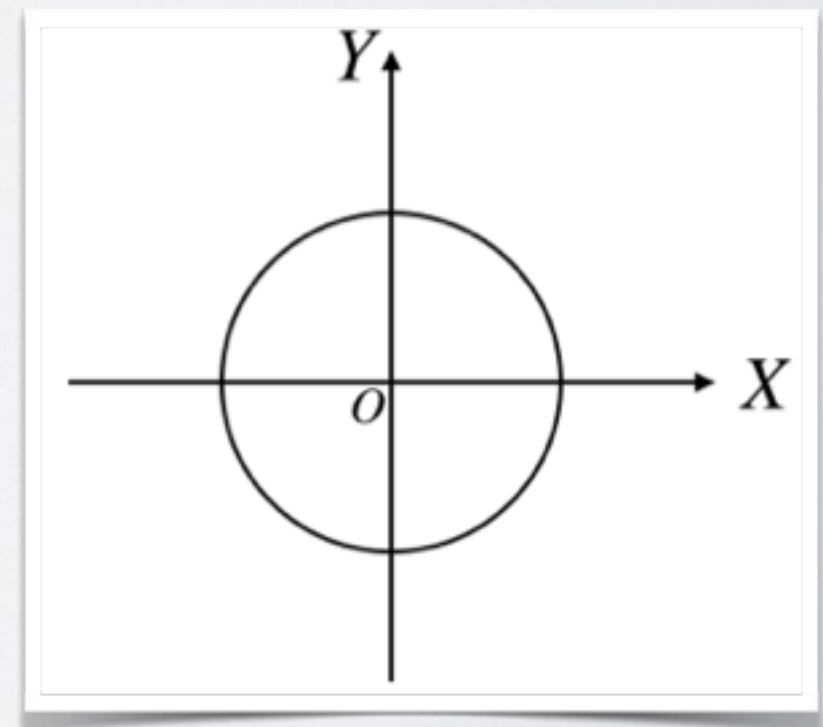
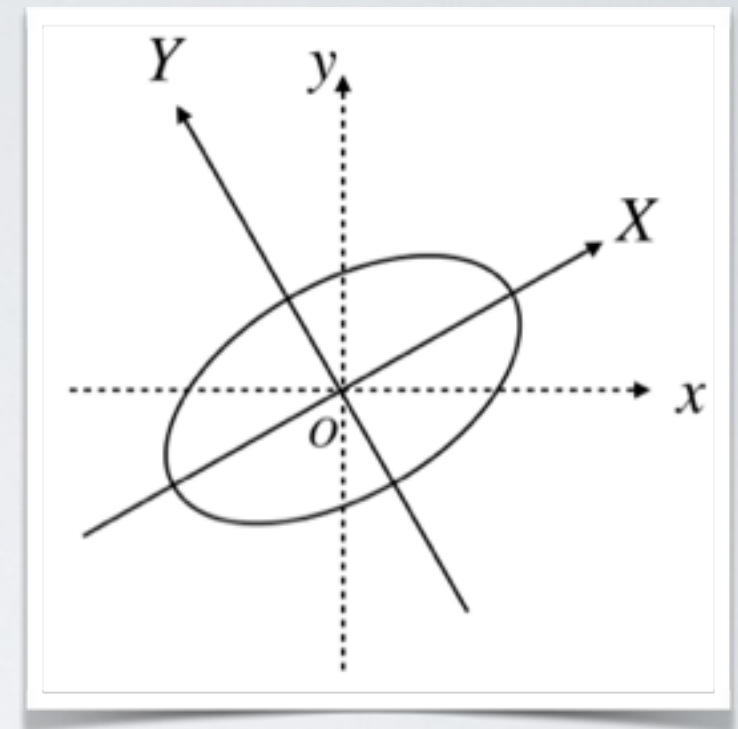
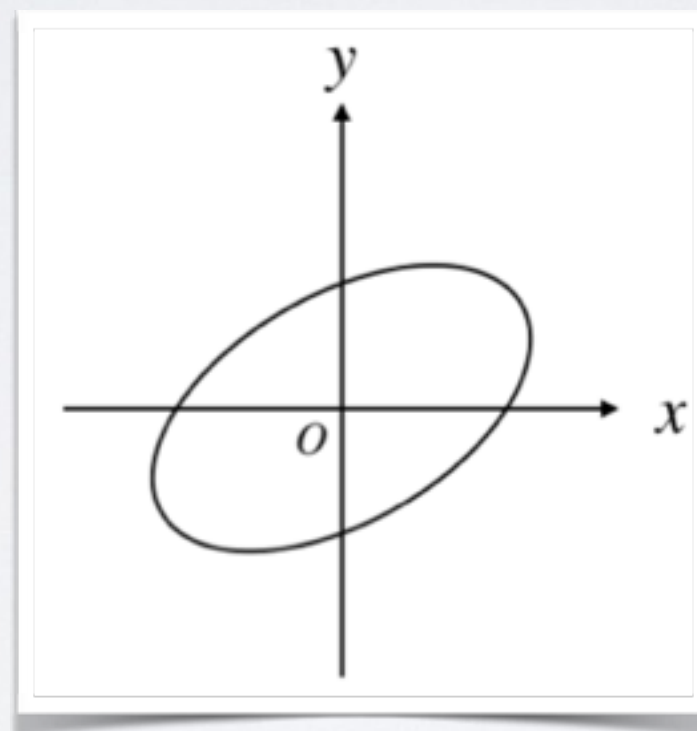
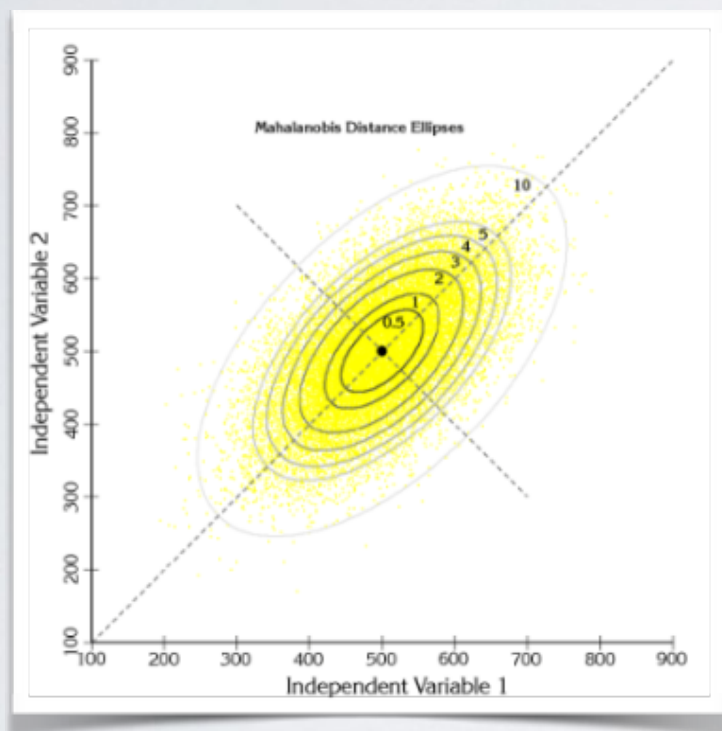
- M. Ackermann et al. (2012)
Boosted Decision Tree & ロジスティック回帰分析
- N. Mirabal et al. (2012)
ランダムフォレスト
- M. Doer & M. Errando (2013)
ニューラルネットワーク & ランダムフォレスト
- J. Lefaucheur et al. (2015)
Boosted Decision Tree & ニューラルネットワーク

=> Mahalanobis-Taguchi法を適用したFermi-LAT 3rdソースカタログ未同定ガンマ線源の識別

Mahalanobis-Taguchi法

パターン認識法の一つ

- 基準データの特徴量から基準空間を作成
- 比較データを基準空間により正規化
- 各データのマハラノビス距離を算出



- マハラノビス距離が1に近いほど基準データに似ている
- マハラノビス距離が大きいほど基準データに似ていない

マハラノビス距離 D の計算法

Mahalanobis distance

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

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

$$Z_i = (X_i - m_i) / s_i$$

X_i = value of i -th characteristic

m_i = mean of i -th characteristic

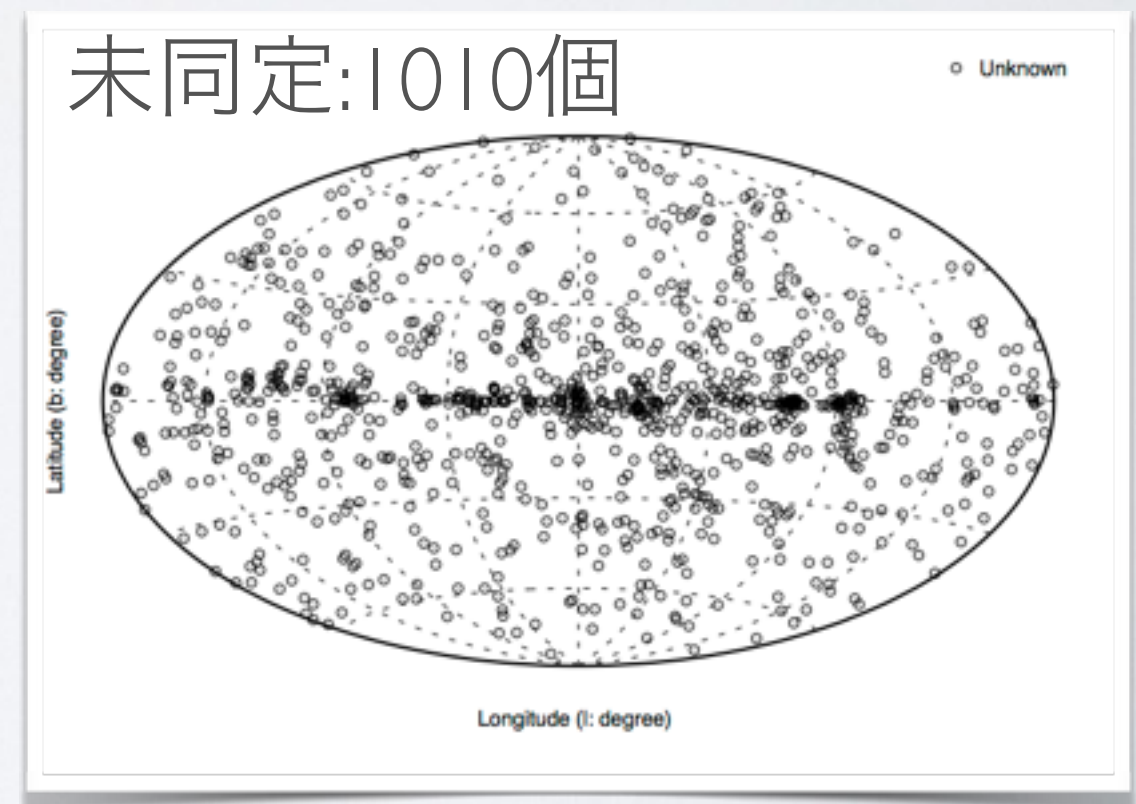
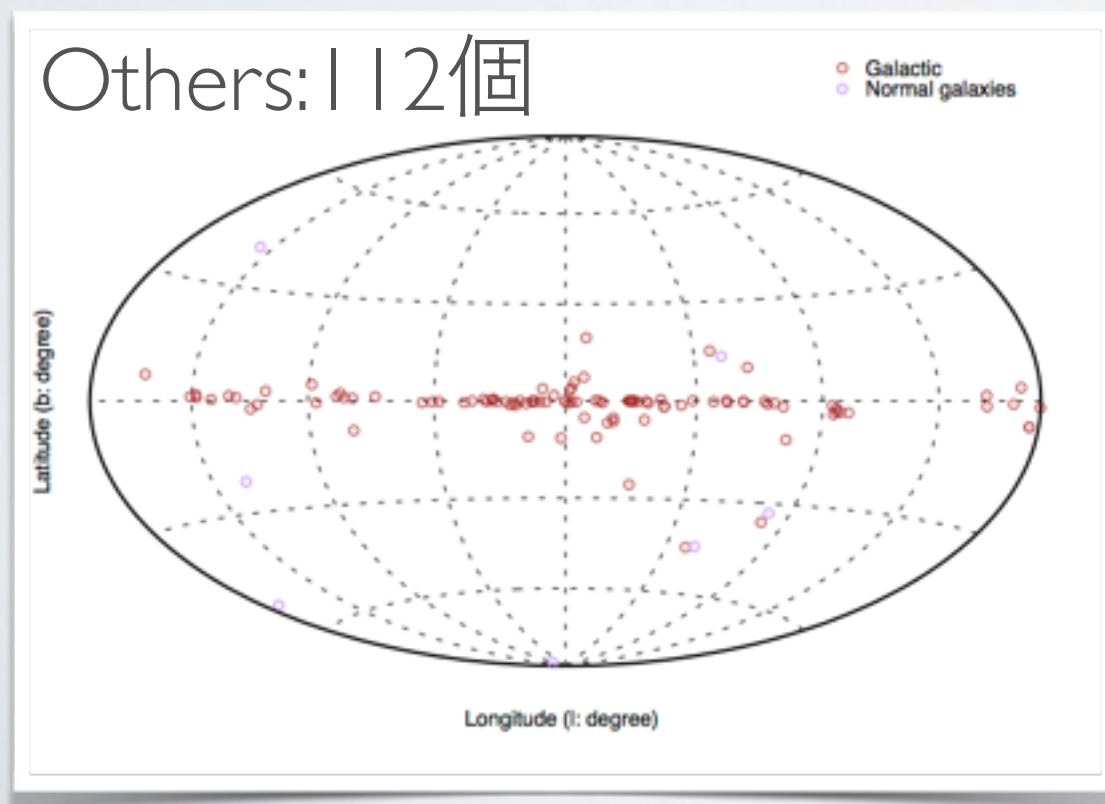
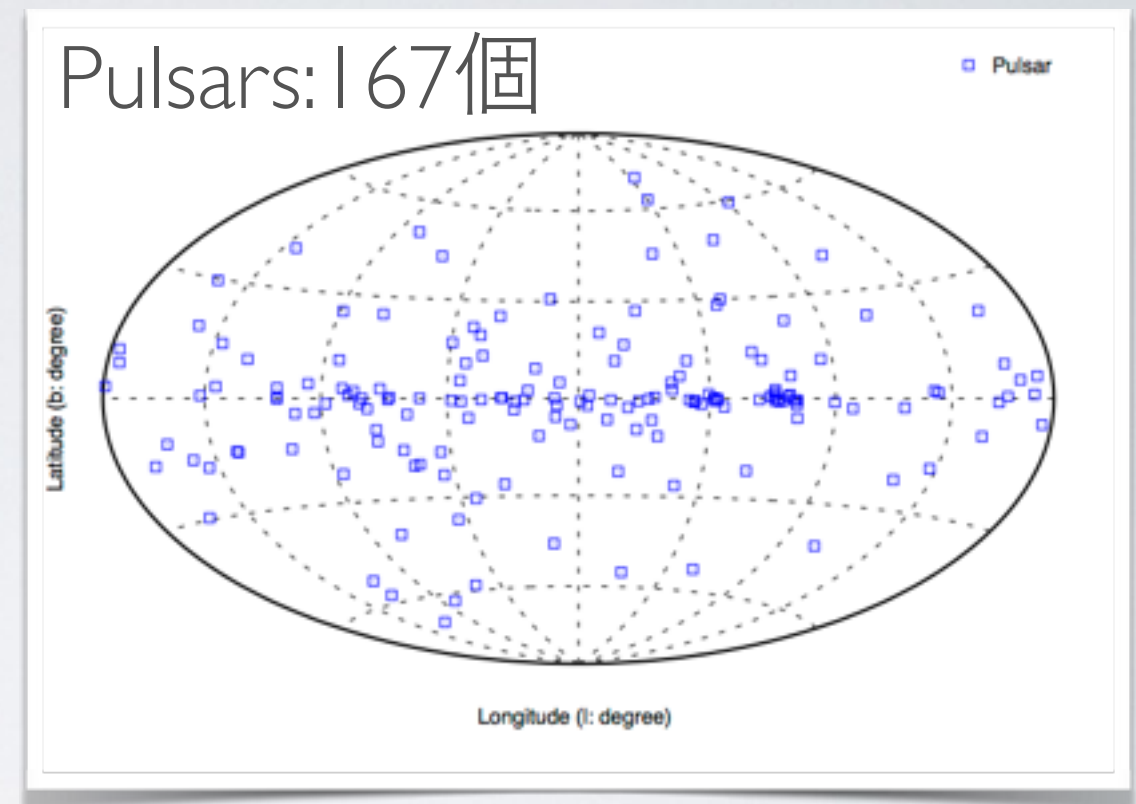
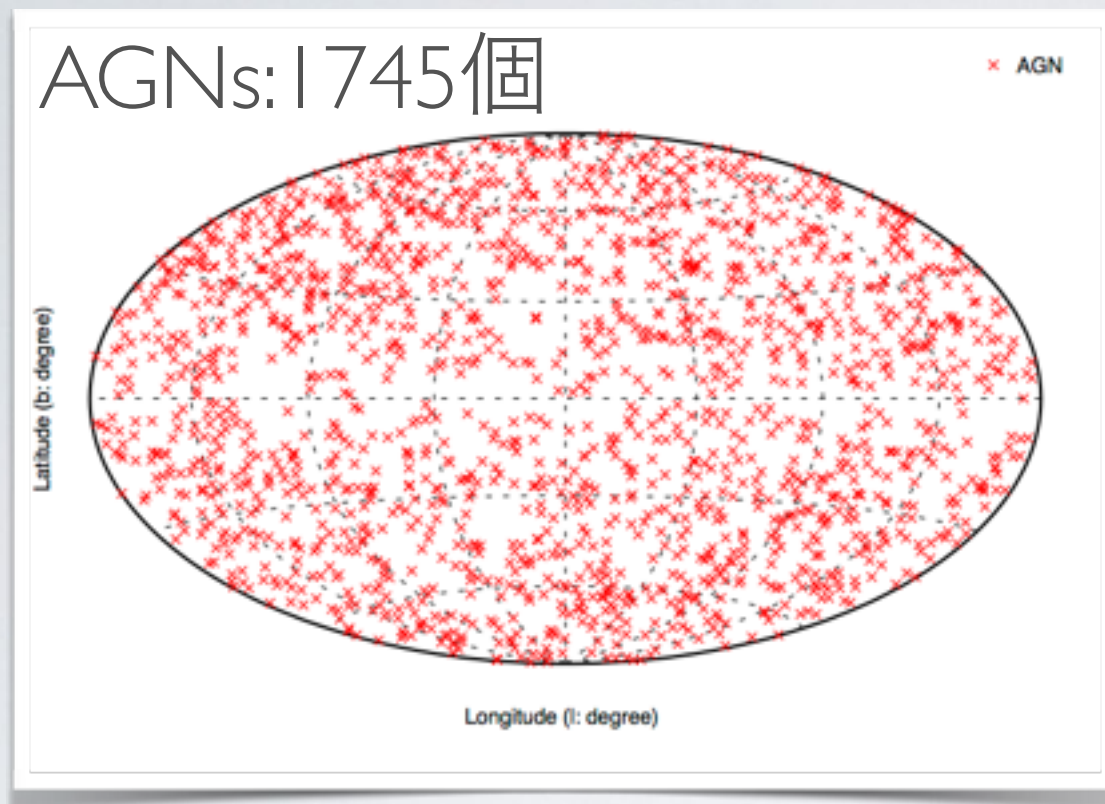
s_i = standard deviation of i -th characteristic

k = the number of characteristics / variables

T = transpose of the vector

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

3FGLカタログにおけるガンマ線源



3FGLカタログにおける特徴量

Ackermann et al. (2012)と同様の特徴量を採用：

Source significanceに依存しない特徴量

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

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

$N_{int} = 48$ in 3FGL

$F_{av} =$ 平均Flux

$\sigma_i = F_i$ の統計誤差

$f_{rel} =$ 系統誤差(2%)

Spectral Index

Γ

銀経 \wedge 銀緯 \sphericalangle

Signif.-Curve/Signif.-Avg.

AGNとNON-AGNの識別

使用した特徴量

$$HR_{34}, HR_{45}, \log\left(\frac{\delta F}{F}\right), \Gamma, \text{Signif.Curve/Signif.Avg}, \ell^2, \frac{1.0}{|b|}.$$

AGNデータを用いて基準空間を作成

AGN基準空間上でAGNs、non-AGNsのマハラノビス距離を導出
=> 識別力を評価

AGN基準空間上で未同定ガンマ線源のマハラノビス距離を導出
=> AGNとnon-AGNに識別

PULSARとNON-PULSARの識別

使用した特徴量

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

Pulsarデータを用いて基準空間を作成

Pulsar基準空間上でPulsars、non-Pulsarsのマハラノビス距離を導出

=> 識別率を評価

Pulsar基準空間上で未同定ガンマ線源のマハラノビス距離を導出

=> Pulsarとnon-Pulsarに識別

識別力の評価方法 (AGN基準空間)

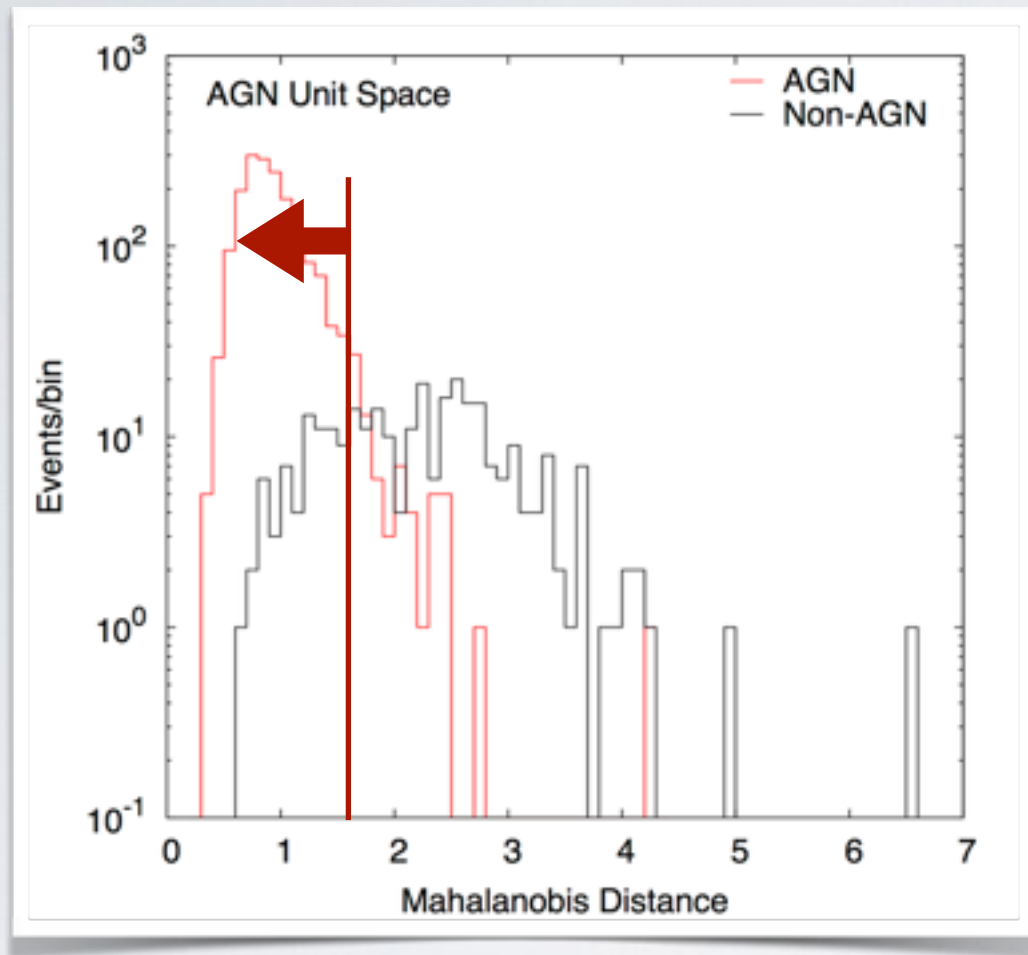
同定ガンマ線源：1745 AGNs、279 non-AGNs in 3FGL

- AGN天体の識別力の評価
 - 基準空間の作成にAGNsの $4/5 = 1397$ AGNsを使用
 - 評価にAGNsの $1/5 = 349$ AGNsを使用
 - 別のデータセットにして5回繰り返し、1397個のAGNsのマハラノビス距離を導出
- non-AGN天体の識別力の評価
 - 279個のnon-AGNsデータをAGN基準空間に適用して、マハラノビス距離を導出

=> Pulsar基準空間の場合も同様に処理

3FGL同定ガンマ線源のマハラノビス距離分布

— AGN基準空間 —



$D_A < 1.17$ に80%のAGN天体
 7.2%(279中20個) のnon-AGN天体
 $D_A < 1.56$ に95%のAGN天体
 22.2%(279中62個) のnon-AGN天体

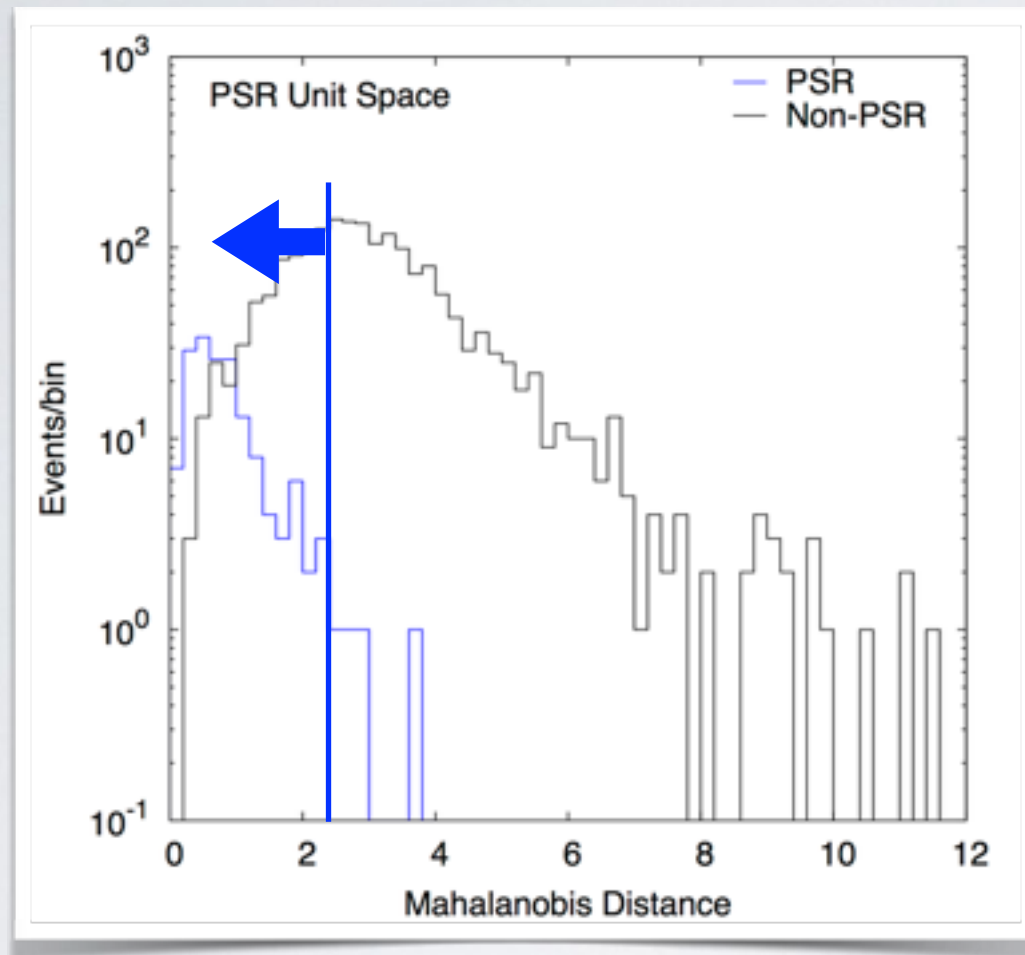
特徴量の重要性ランク

特徴量	使用しない場合：SN(db)	使用した場合：SN(db)
$1.0/ b $	-2.248	2.034
$ ^2$	-1.760	1.592
$\log(\delta F/F)$	-1.408	1.274
HR ₄₅	-1.196	1.083
Γ	-0.690	0.625
Sig.Curve/ Sig.Avg	-0.423	0.382
HR ₃₄	0.004	-0.004

3FGL同定ガンマ線源のマハラノビス距離分布

— PULSAR基準空間 —

特徴量の重要性ランク



$D_P < 1.17$ に80%のPulsar天体
 4.6%(1857中85個) のnon-Pulsar天体
 $D_P < 2.18$ に95%のPulsar天体
 25.8%(1857中479個) のnon-Pulsar天体

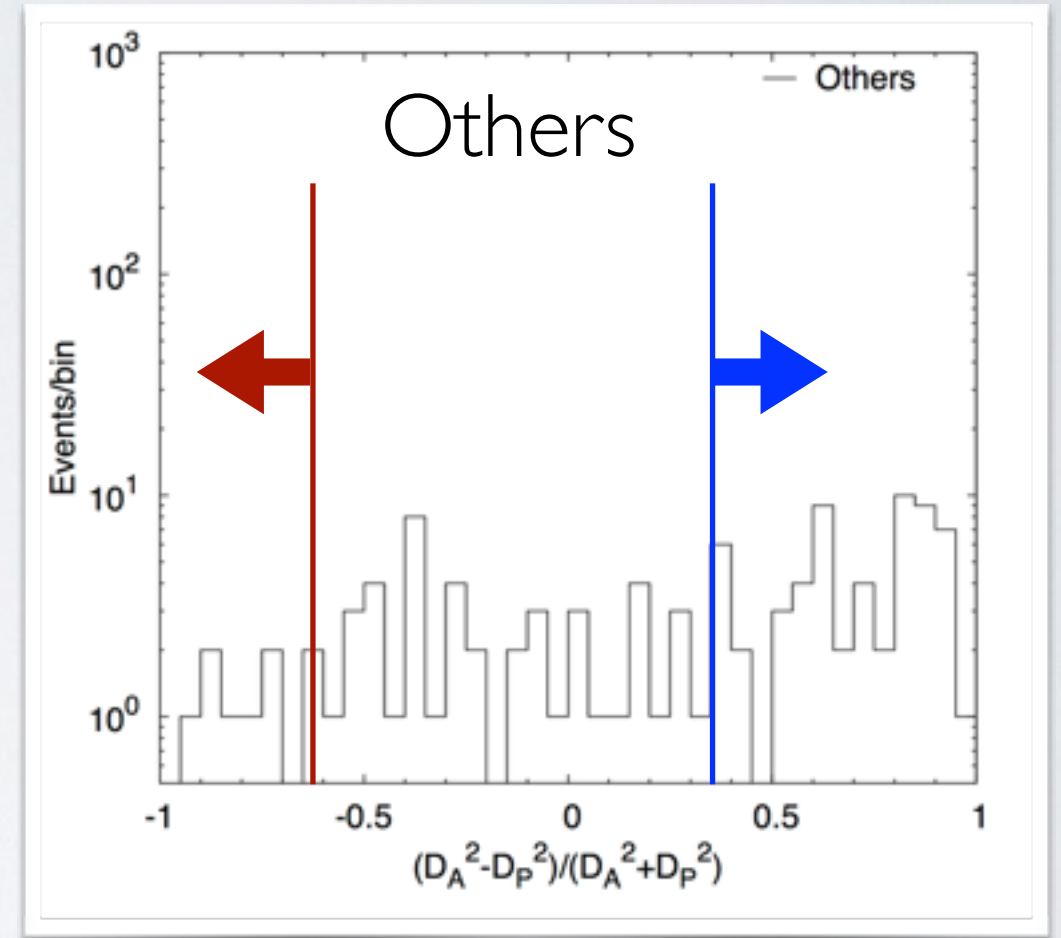
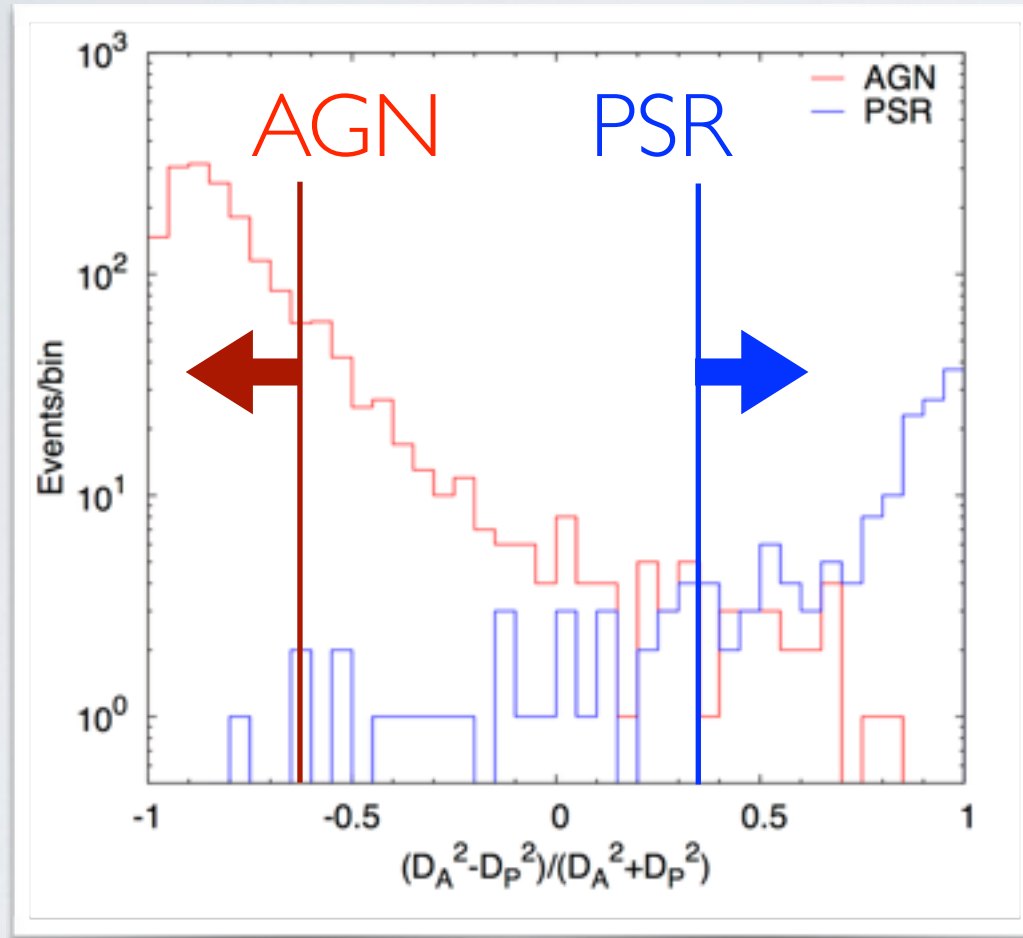
特徴量	使用しない場合：SN(db)	使用した場合：SN(db)
$\delta F/F$	-7.237	4.135
HR ₂₃₋₃₄	-5.948	3.399
HR ₄₅	0.558	-0.319
$l \times b$	3.777	-2.158

3FGL同定ガンマ線源に対する識別力

AGN基準空間でのマハラノビス距離= D_A

Pulsar基準空間でのマハラノビス距離= D_P

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



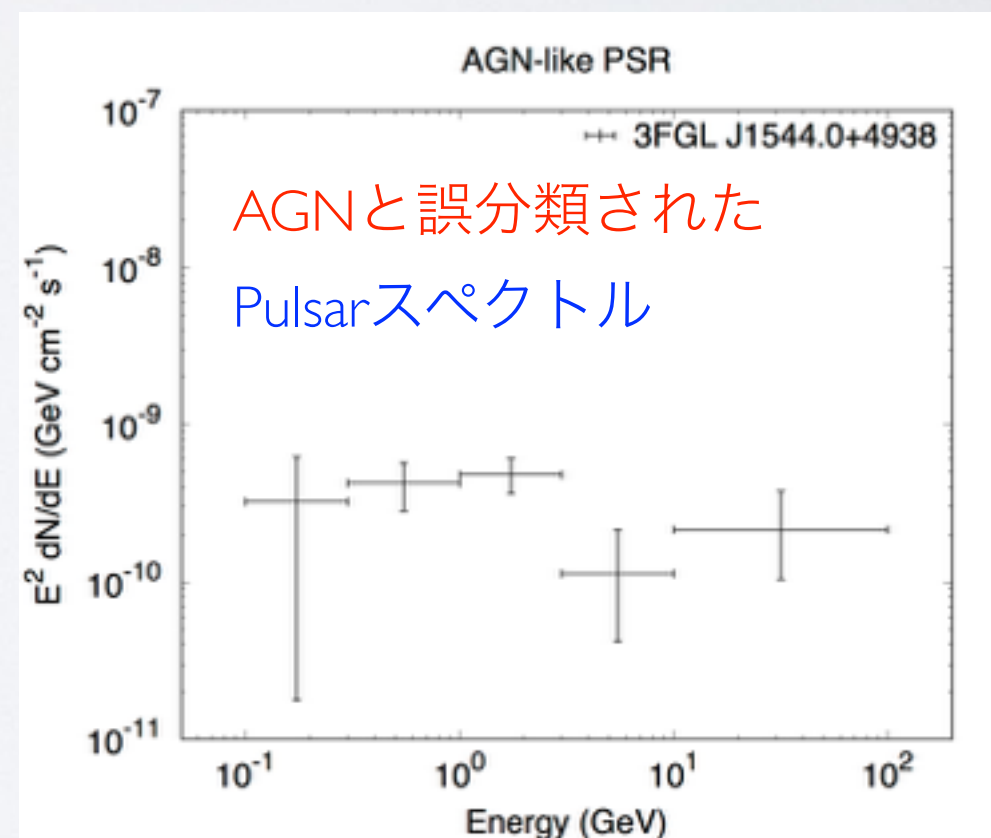
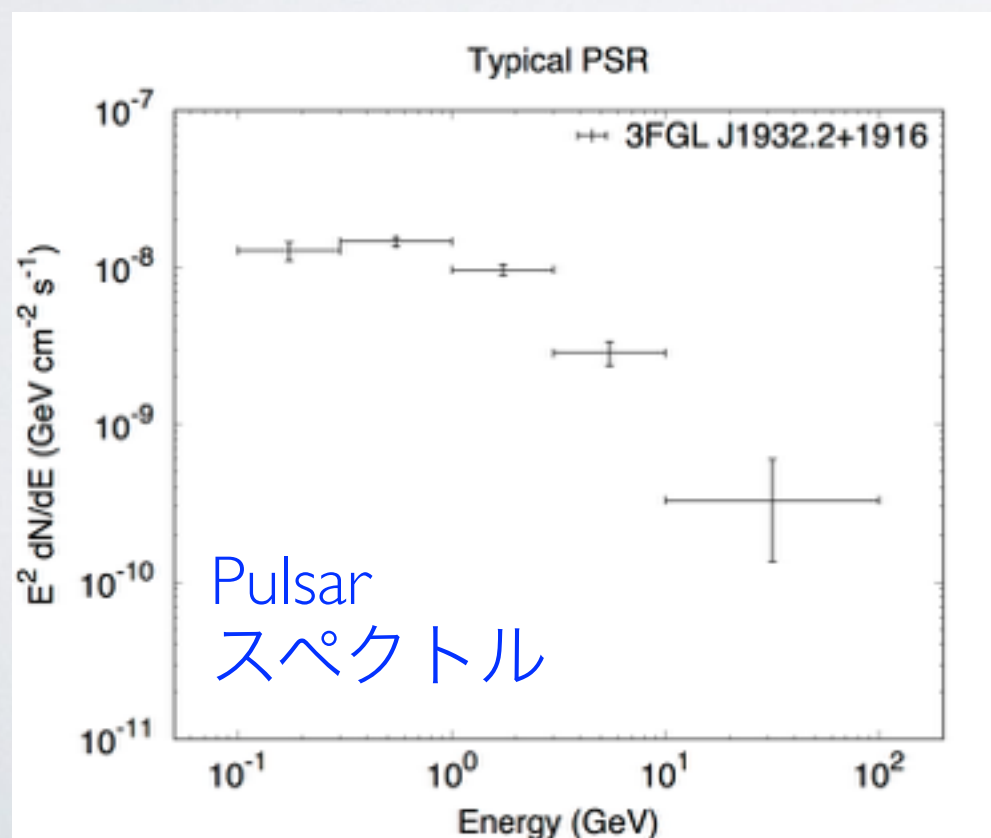
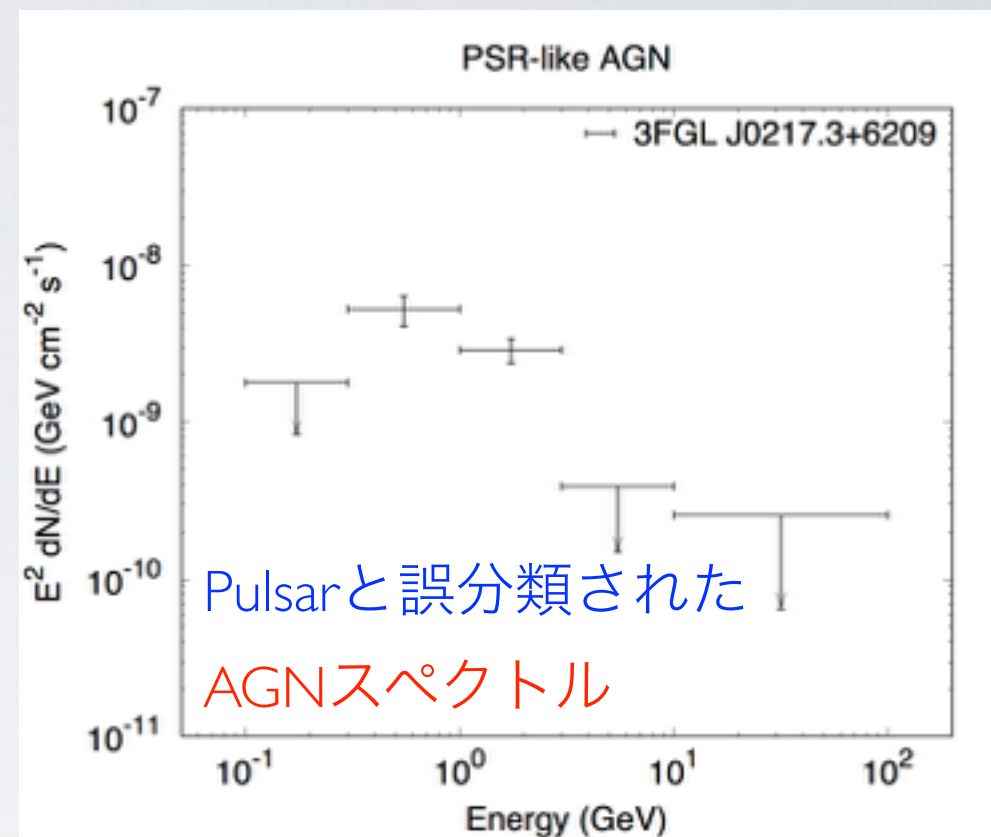
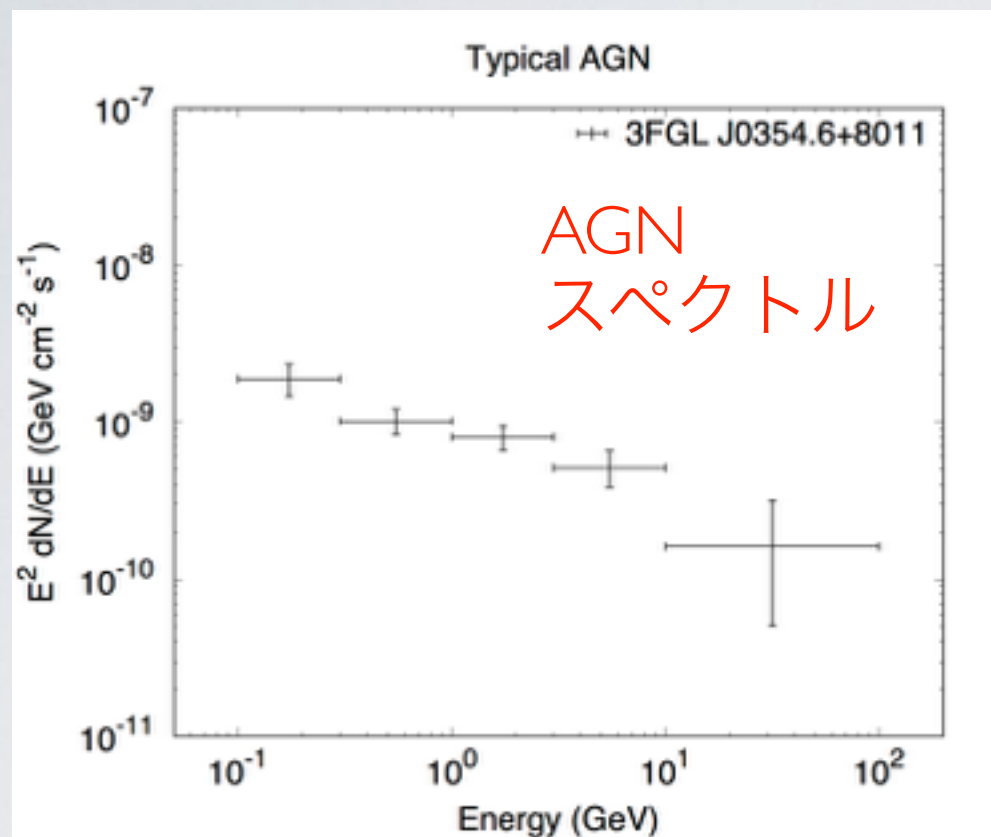
$$D_A < 1.56 \ \& \ r < -0.62$$

AGN 80.7% (1408 / 1745) 、 Pulsar 0.6% (1 / 167) 、
Others 7.1% (8 / 112)

$$D_P < 2.18 \ \& \ r > 0.35$$

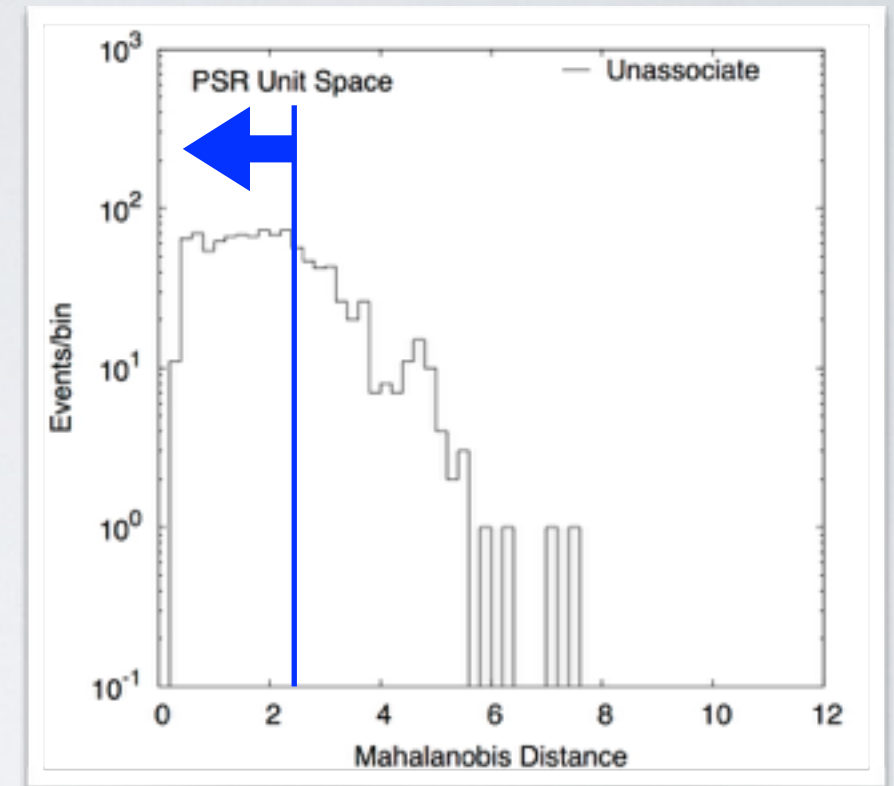
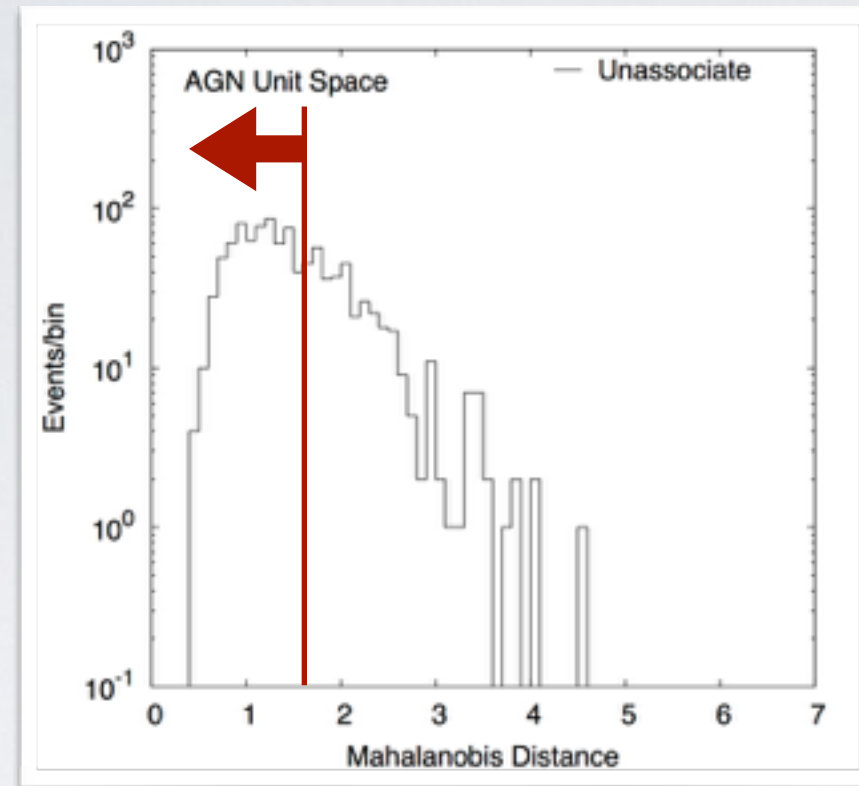
Pulsar 80.8% (135 / 167) 、 AGN 1.1% (20 / 1745) 、
Others 50.9% (57 / 112)

AGNとPULSARのスペクトル例



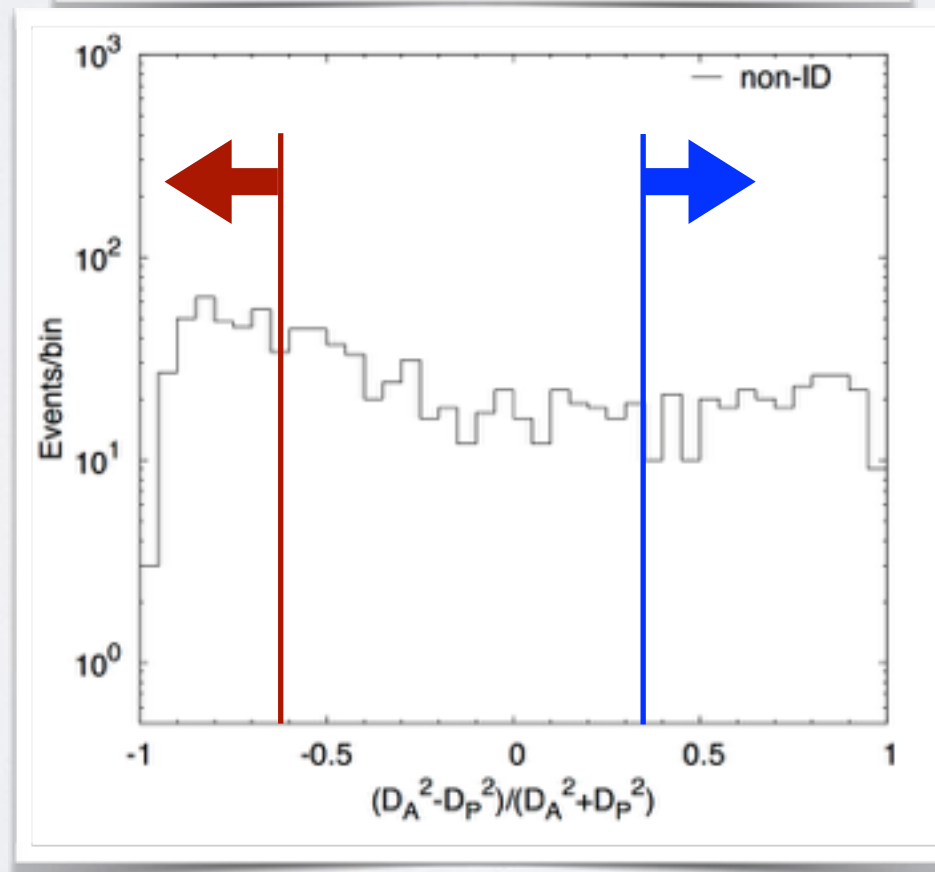
3FGL未同定ガンマ線源の識別

未同定ガンマ線源
マハラノビス距離
の頻度分布



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

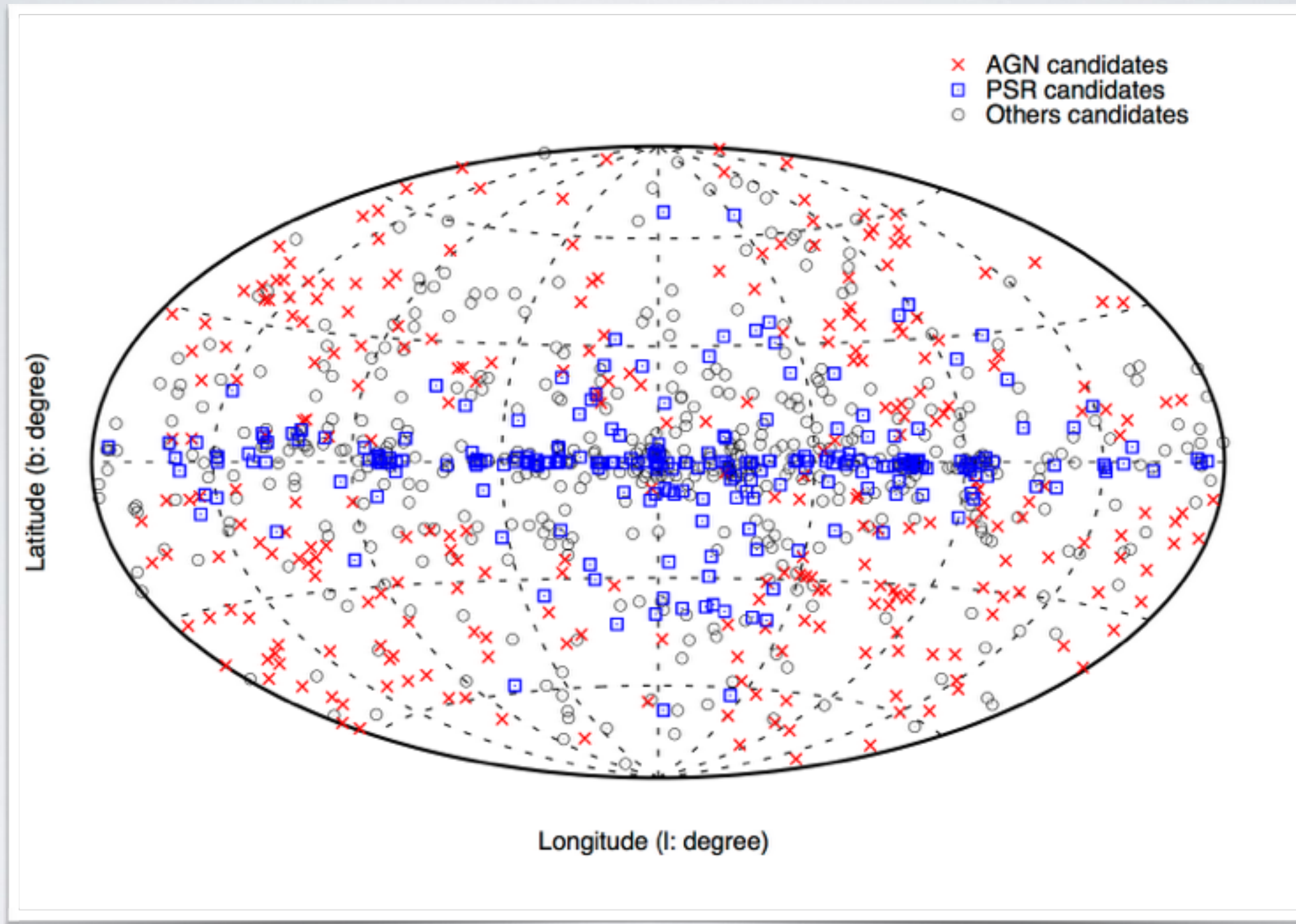
の頻度分布



$D_A < 1.56$ & $r < -0.62$
=> AGNと識別

$D_P < 2.18$ & $r > 0.35$
=> Pulsarと識別

3FGL未同定ガンマ線源の識別結果



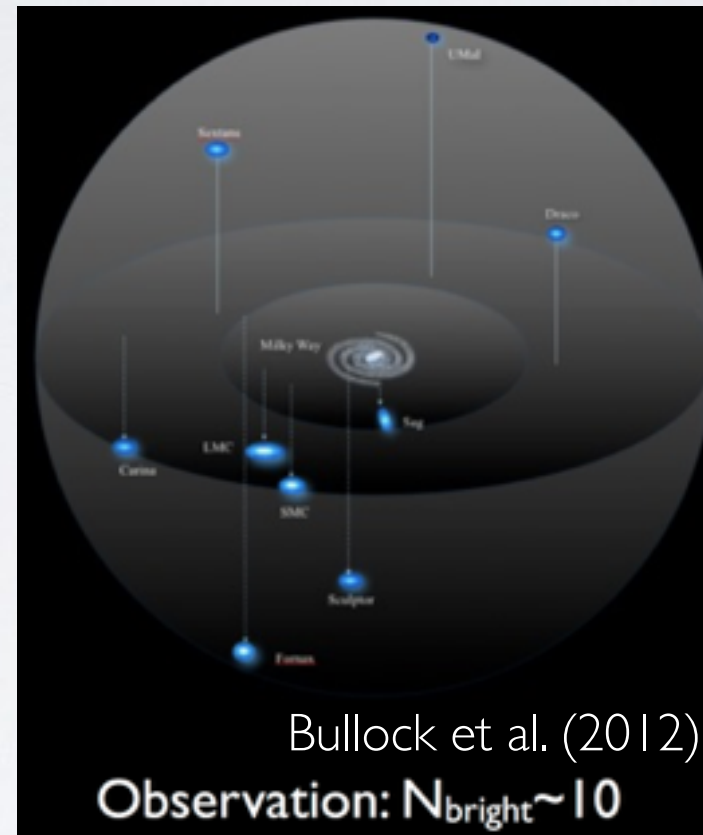
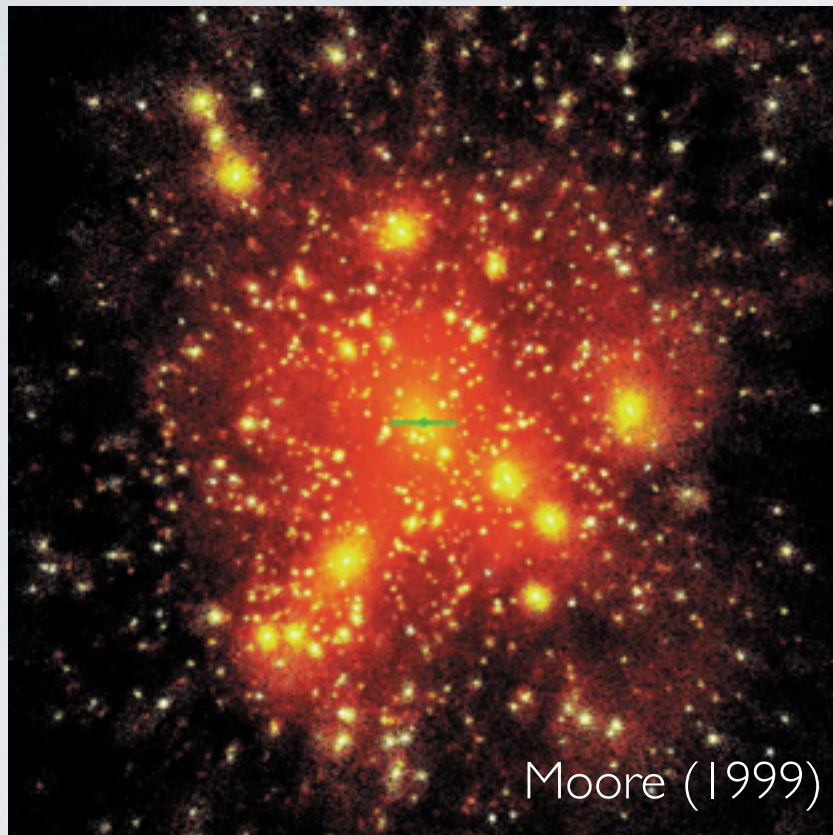
AGN候補：277個、Pulsar候補：245個、他：488個

暗黒物質サブハロー

矮小銀河・サブハローの個数の問題

数値計算： $N \gg 1000$

観測： $N \sim 10$



数値計算によるサブハローの個数が、銀河系内に観測される矮小銀河の数より桁違いに多い問題

- 見えないサブハロー「暗黒物質サブハロー」があるのでは？
=> WIMP暗黒物質の対消滅でガンマ線が発生
- 3FGL未同定ガンマ線源の中に ~ 10 個の暗黒物質サブハローが含まれる可能性の指摘 (Bertoni et al. 2015)

暗黒物質サブハロー候補の探索

高銀緯 $|b| > 5^\circ$ でマハラノビス距離 D の大きい ($D_A > 2.47$ && $D_P > 3.22$) ガンマ線源

- 同定ガンマ線源 2014個 \Rightarrow 0個
- 未同定ガンマ線源 1010個 \Rightarrow 4個

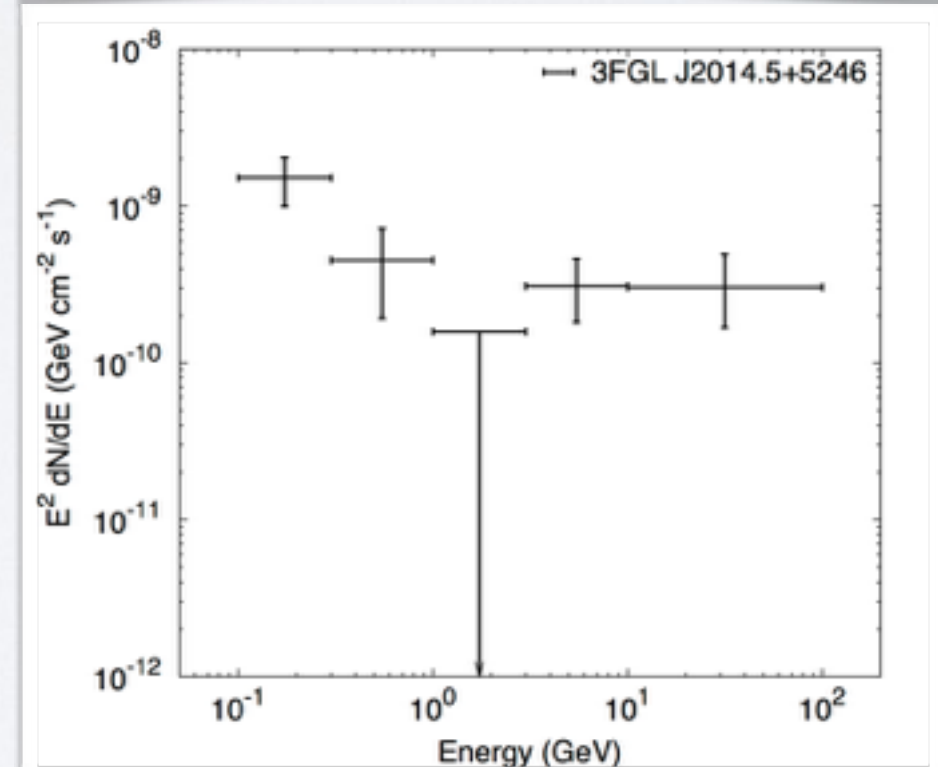
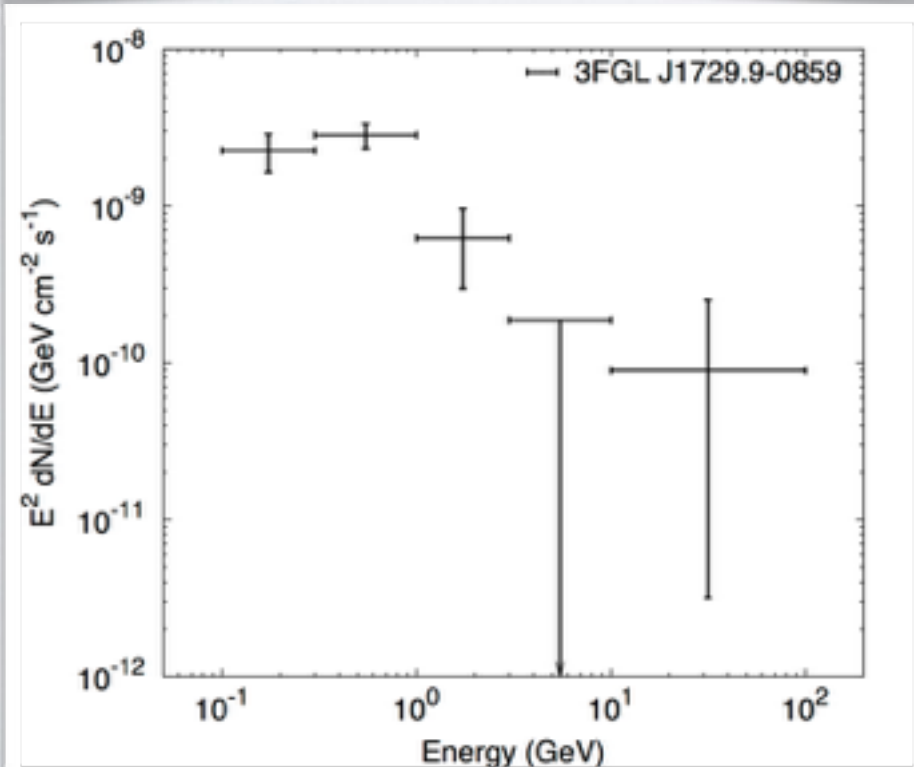
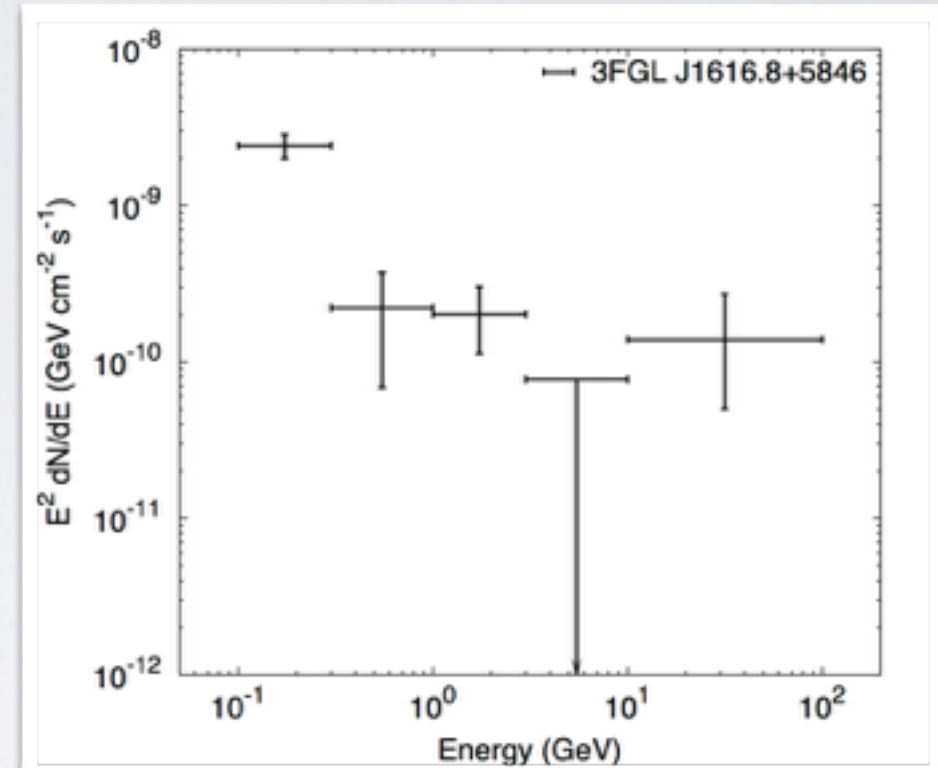
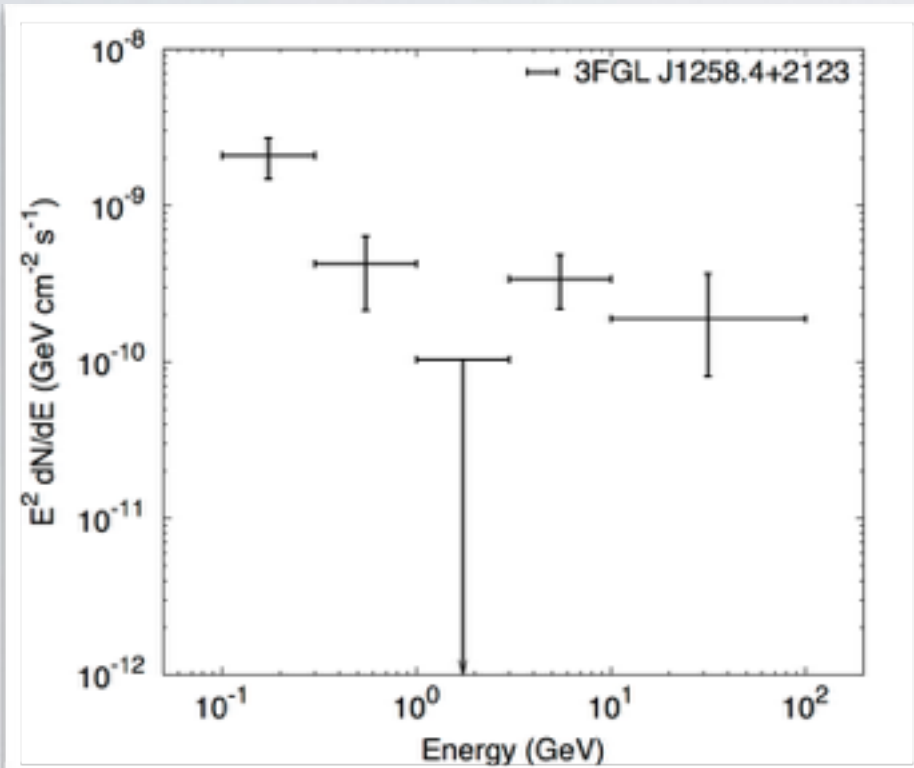
同定と未同定ガンマ線源での有意差：1.2%

\Rightarrow 未同定ガンマ線源の中に同定できないような種類の天体？

暗黒物質サブハロー候補？

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

高銀緯 $|b| > 5^\circ$ でマハラノビス距離 D の大きい ($D_A > 2.47$ &&
 $D_p > 3.22$) 4個の未同定ガンマ線源スペクトル



まとめ

- 3FGLガンマ線源をMahalanobis-Taguchi法を用いてAGN、Pulsar、その他に識別：
 - AGN識別率80.7%でPulsar混入率0.6%
 - Pulsar識別率80.8%でAGN混入率1.1%

=> 従来の結果 (e.g. 各々3%、1%、M.Ackermann et al. 2012)
と同等以上
- 高銀緯 $|b| > 5^\circ$ で暗黒物質サブハロー候補の探索：
 - マハラノビス距離の大きい4個の未同定ガンマ線源