Classification of Fermi-LAT sources with ML and the puzzle of soft Galactic unassociated sources

ERLANGEN CENTRE FOR ASTROPARTICLE PHYSICS

Dmitry Malyshev

for the Fermi-LAT collaboration

TeVPA, Valencia, 3 – 7 November, 2025

Fermi Gamma-ray Space Telescope

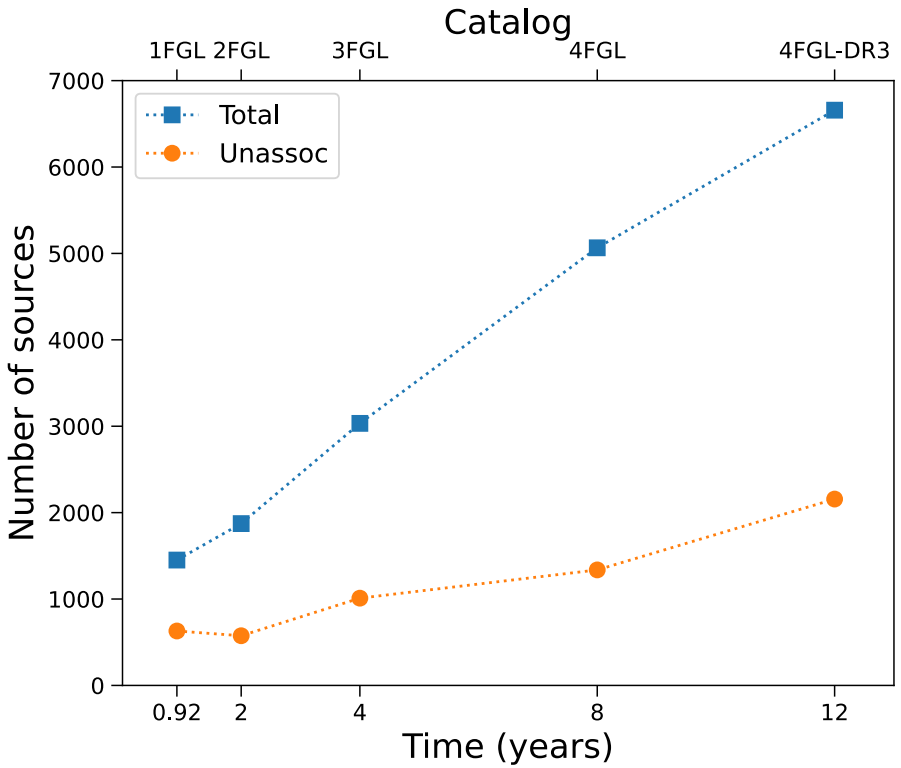
DFG Deutsche Forschungsgemeinschaft German Research Foundation

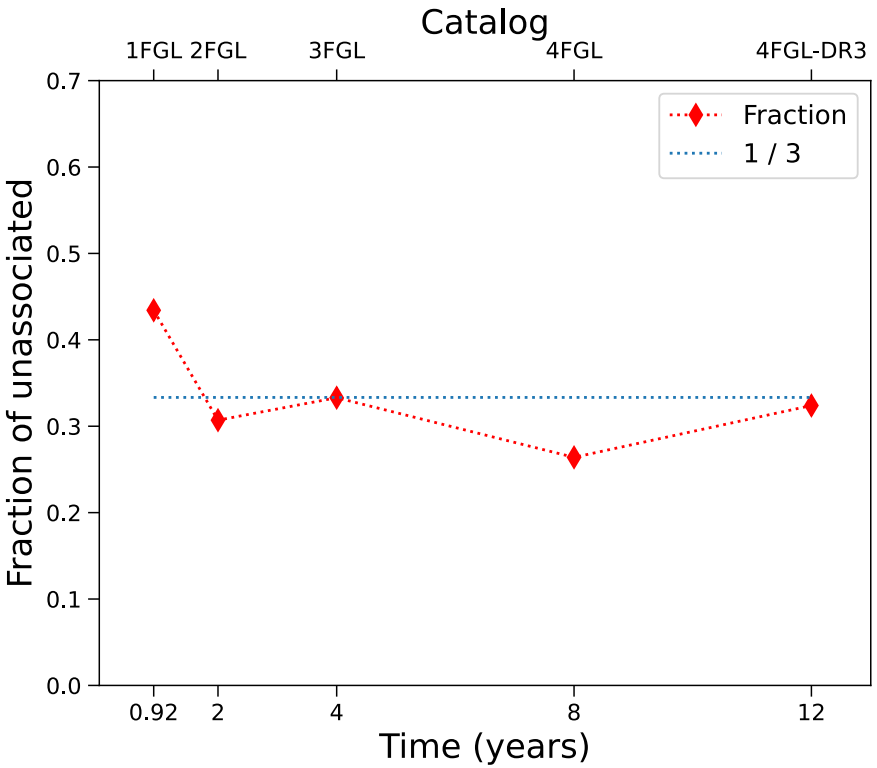
ECAP ERLANGEN CENTRE FOR ASTROPARTICLE PHYSICS

FAU FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Unassociated sources in *Fermi*-LAT catalogs

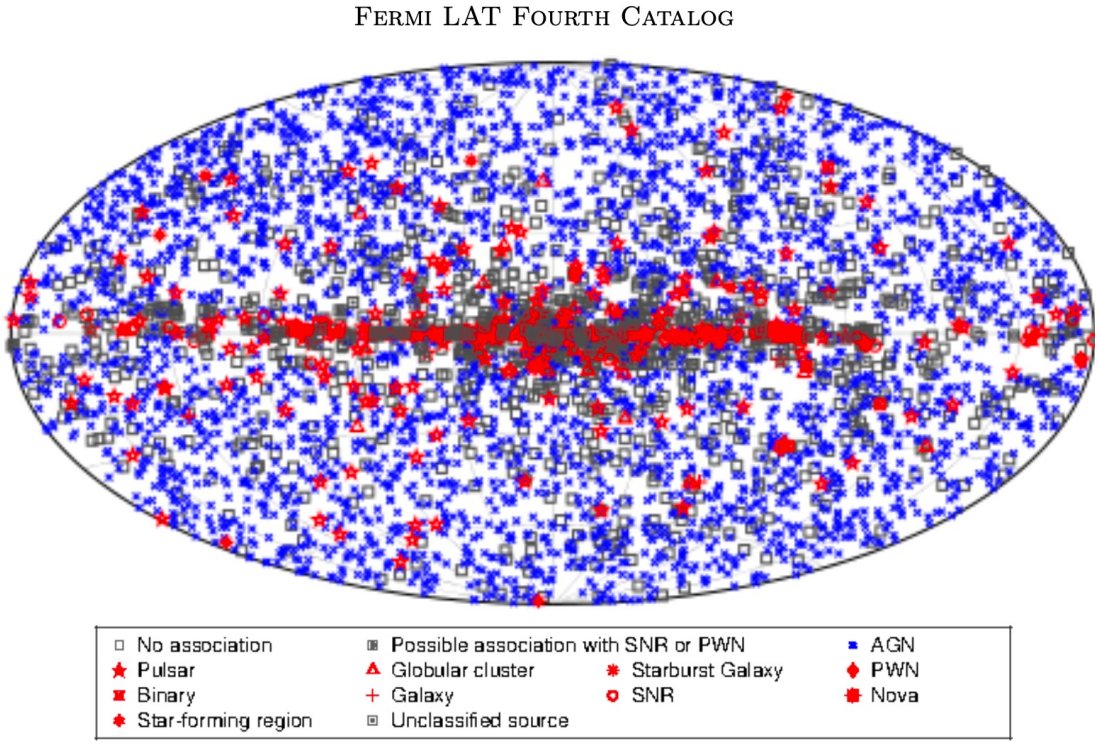
ERLANGEN CENTRE FOR ASTROPARTICLE PHYSICS

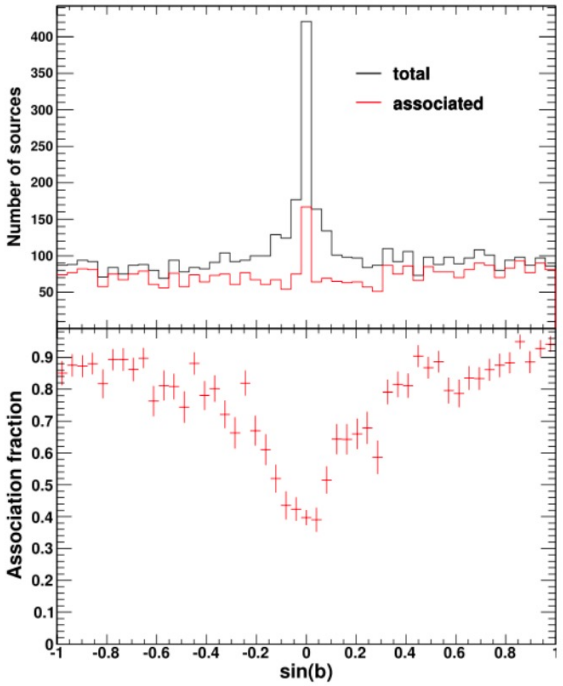




- In Fermi-LAT catalogs, both the total number of sources and the number of unassociated sources grow as a function of time
- The fraction of unassociated sources in *Fermi*-LAT catalogs is remarkably constant ~ 1 / 3
- This is surprising given the significant effort in follow up searches!

Most of unIDs are in the Galactic plane

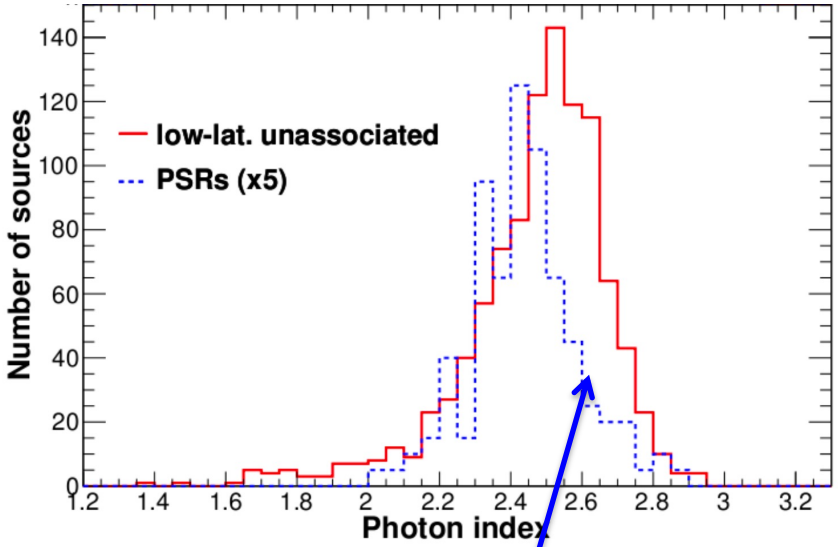




4FGL paper, Abdollahi et al., ApJS 247 (2020)

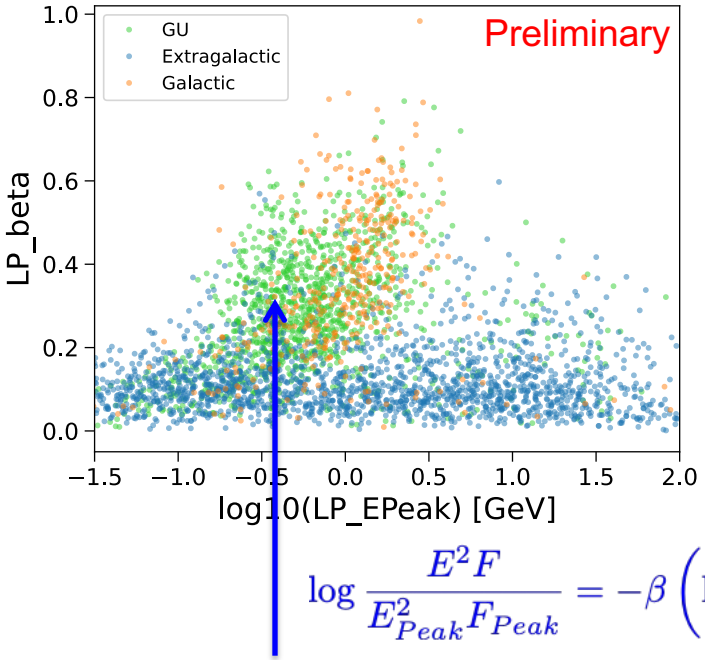
- ~60% of sources within |b| < 10° are unIDs:
  - 1293 / 2124 (in 4FGL-DR4)
- ~50% of all unID sources are within |b| < 10°:
  - 1293 / 2575 (in 4FGL-DR4)

Galactic unassociated (GU) sources



4FGL-DR3 paper,
Abdollahi et al.,
ApJS 260 (2022)

Power-law index of GU sources is softer than that of pulsars

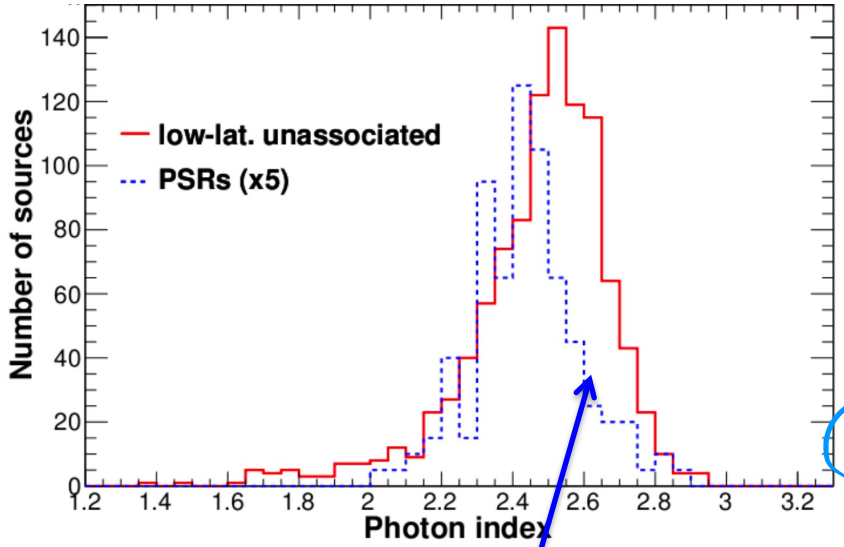


$$\log \frac{E^2 F}{E_{Peak}^2 F_{Peak}} = -\beta \left(\log \frac{E}{E_{Peak}} \right)^2$$

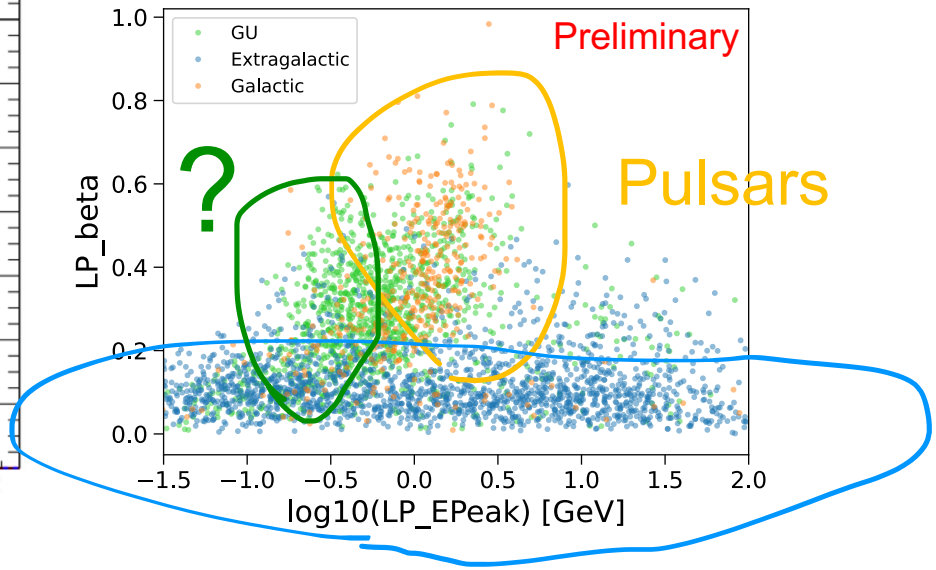
GUs have curved spectra with the peak in SED lower than that of Galactic sources (mostly pulsars)

- There seems to be a subpopulation of sources distinct from known Galactic and extragalactic sources

Galactic unassociated (GU) sources



4FGL-DR3 paper,
Abdollahi et al.,
ApJS 260 (2022)

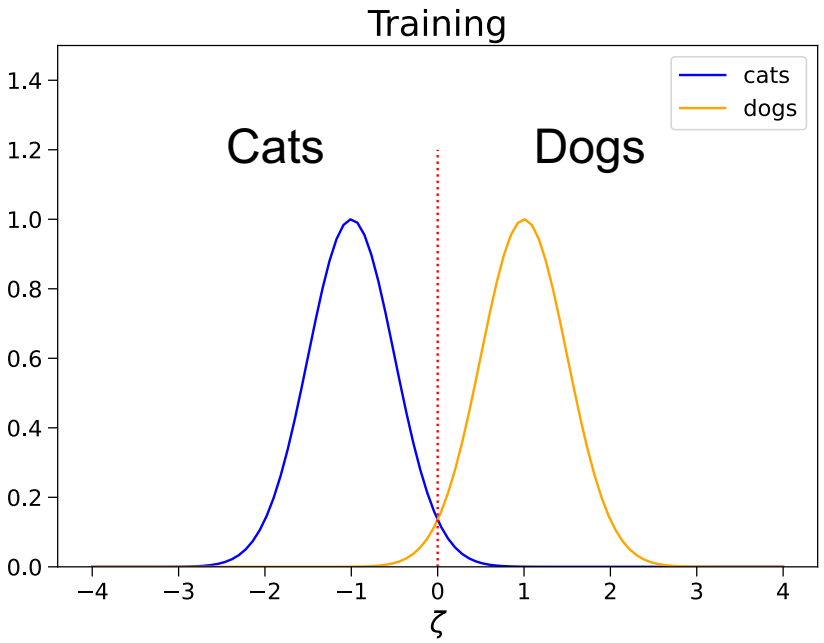


Power-law index of GU sources is softer than that of pulsars

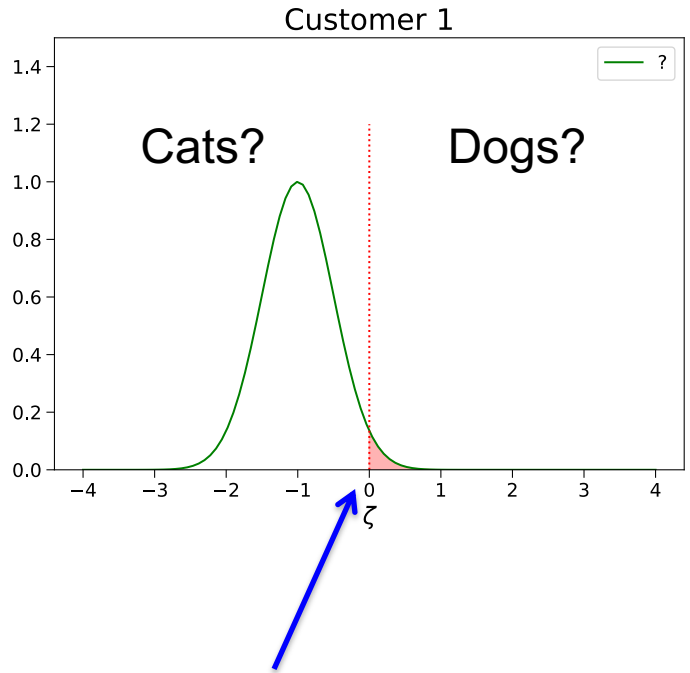
- AI can solve all problems in the world
  - Can we use it to learn something about GUs?

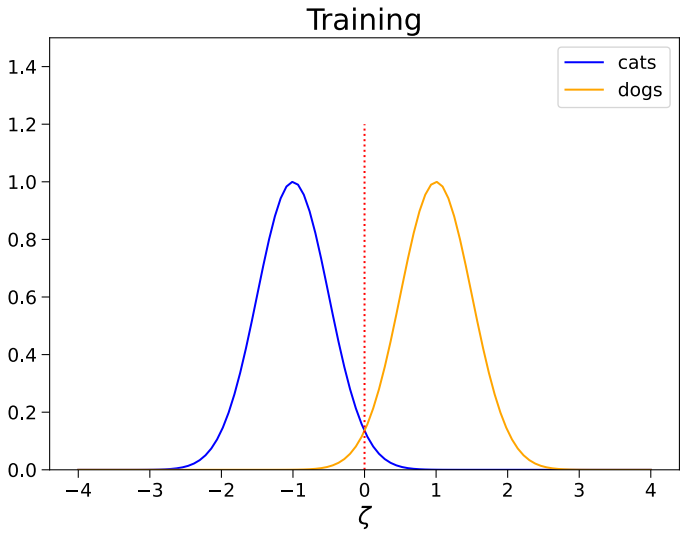
Dataset shift

- There is a subtlety in the application of ML for the analysis of gamma-ray sources due to difference in the distribution of training data (associated sources) and target data (unassociated sources)
  - This is known as dataset shift in ML literature
  - Let's look at a toy example
    - Suppose we have a classifier that can separate very well cats and dogs (aka AGNs and pulsars in astrophysics)



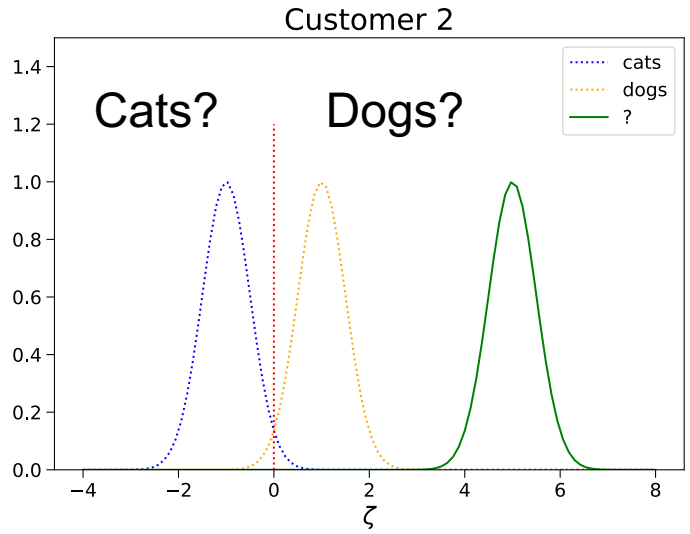
Cats & Dogs company, case 1

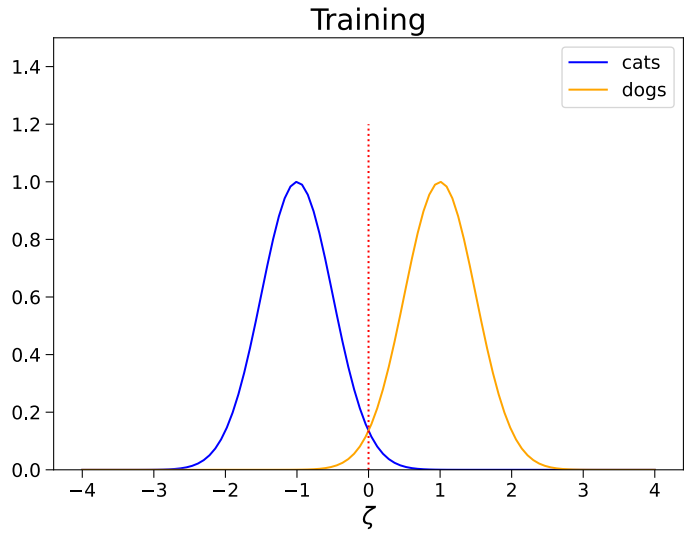




- What are these animals?
  - High probability dogs
  - More likely to be dogs
  - More likely to be cats
  - High probability cats

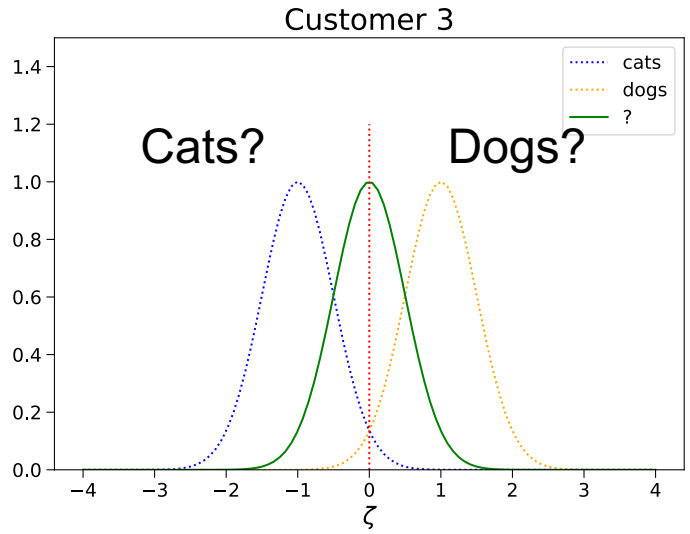
Cats & Dogs company, case 2

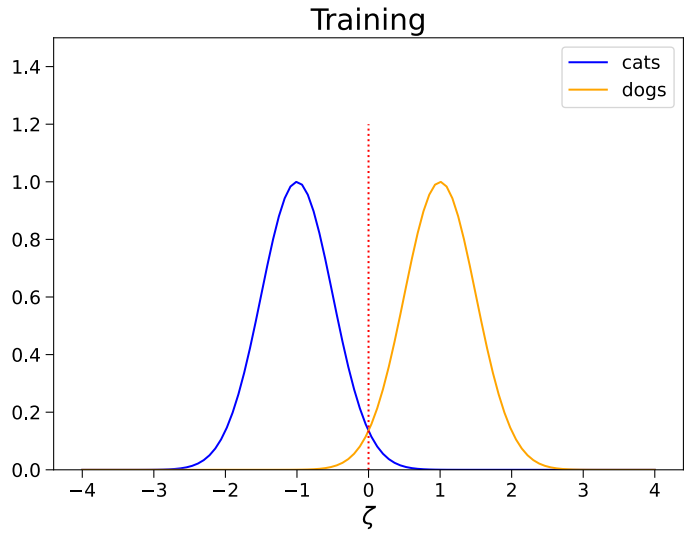




- Now the algorithm is almost 100% confident that these are dogs
  - It turns out that it's 100% wrong
  - These are dinosaurs
- You start to get worried: maybe you should allow only balanced datasets?

Cats & Dogs company, case 3



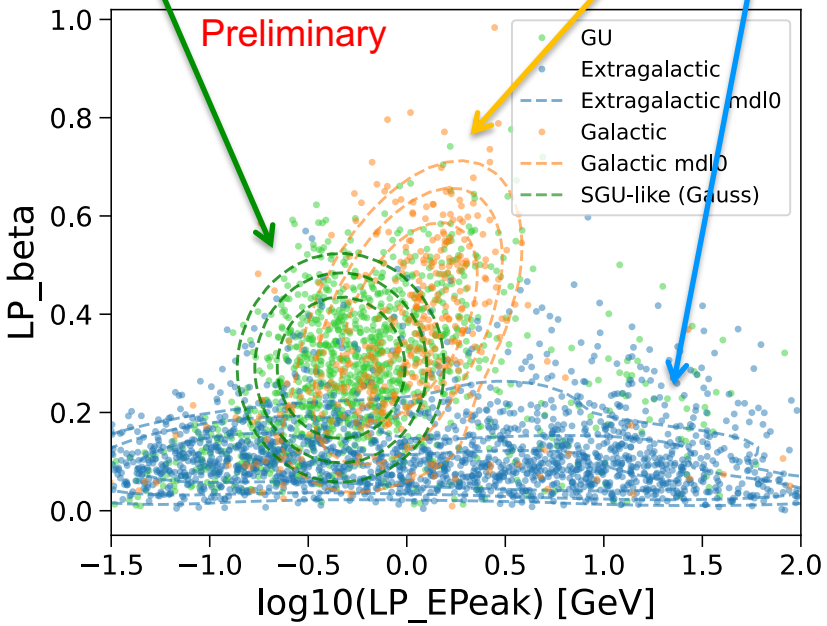


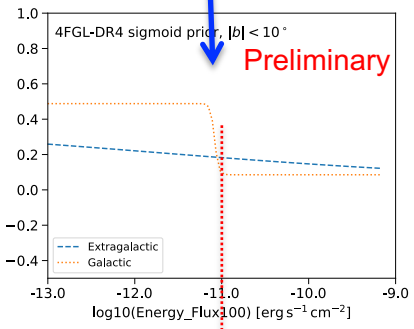
- Now the prediction is perfectly balanced and still 100% wrong
- These are racoons.
- If your model cannot describe the data, you should change the model, not the data!
  - otherwise the Cats & Dogs company will go bankrupt
  - or your paper on AGNs and pulsars will be rejected from a journal

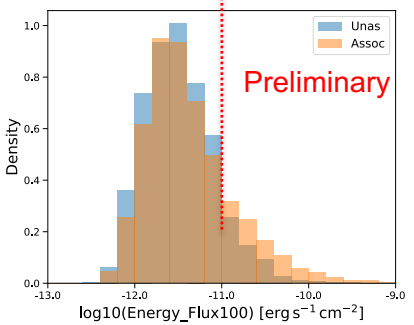
Mathematical model

- Define a probability distribution function (PDF) for unassociated sources as a mixture of known components and a new component (modeled as a Gaussian in the feature space):

$$p_{\mathrm{unas}}(x) = G(x) + \sum_{k=(\mathrm{Gal},\,\mathrm{egal})} p_{\mathrm{assoc}}(x|k)\pi_k(\log W)$$







Model optimization

ERLANGEN CENTRE FOR ASTROPARTICLE PHYSICS

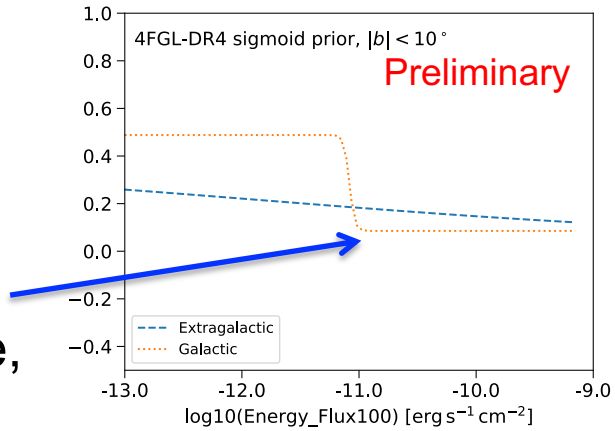
- Determine the parameters of the Gaussian (for the new component) and modulation of the known classes of astrophysical sources by maximizing the unbinned Poisson log likelihood

$$\log L = \sum_{i \in \mathrm{unas}} \log(p_{\mathrm{unas}}(x_i)) - N_{\mathrm{unas}} \int p_{\mathrm{unas}}(x) dx$$

- The modulation as a function of log of energy flux is parameterized as a sigmoid plus a constant:

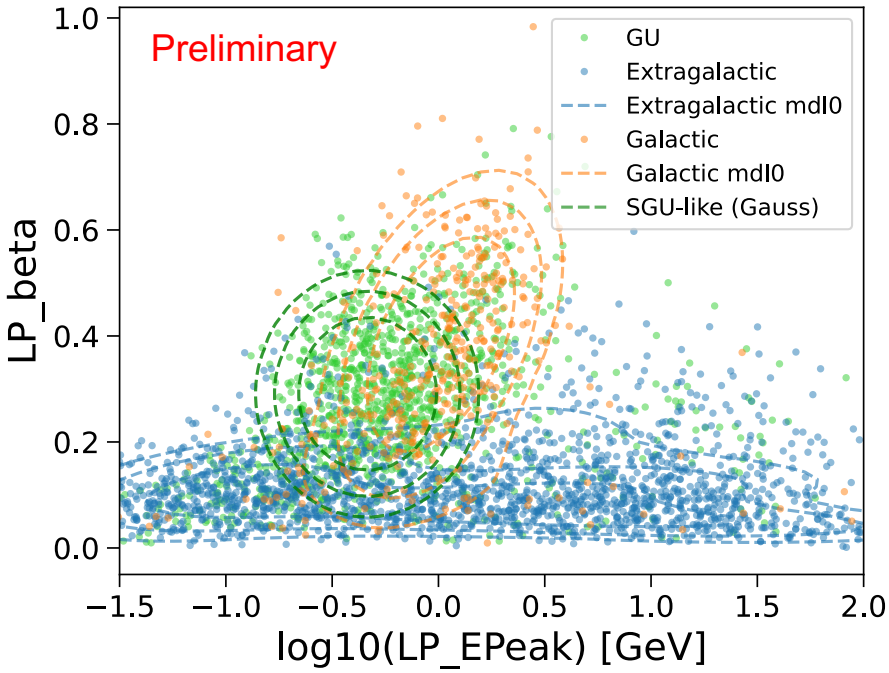
$$\sigma(\log W) = \frac{a}{1 + e^{(\log W - b)/c}} + d$$

- In particular, the model wants to suppress the bright Galactic sources (pulsars, PWNe, SNRs)



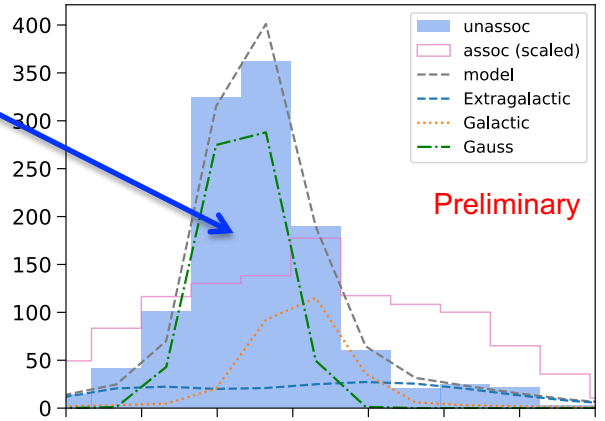
Data selection

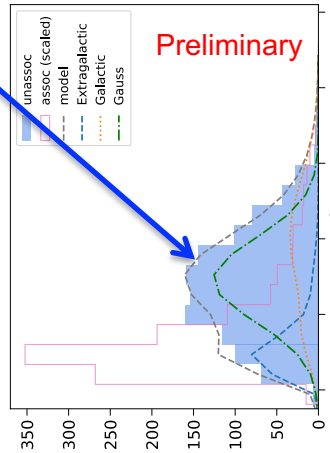
- 4FGL-DR4 (v34) catalog
- Three features (no coordinate features):
  - 'log10(Energy_Flux100)', 'LP_beta', 'log10(Epeak)'
- Two classes (no bcu or spp sources):
  - Galactic: psr, hmb, sfr, snr, pwn, gc, gal, bin, msp, lmb, glc, nov;
  - Extragalactic: bll, sbg, rdg, css, ssrq, fsrq, sey, nlsy1, agn.
- Training:
  - All sky
- Target:
  - unIDs (unas + unk) within $|b| < 10^{o}$

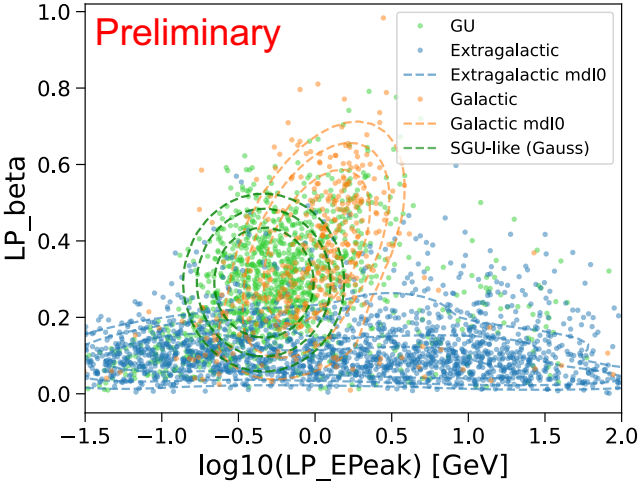


Properties of the new component

- Lower Epeak relative to Galactic sources
- Smaller curvature than for the Galactic sources but larger than for the extragalactic ones



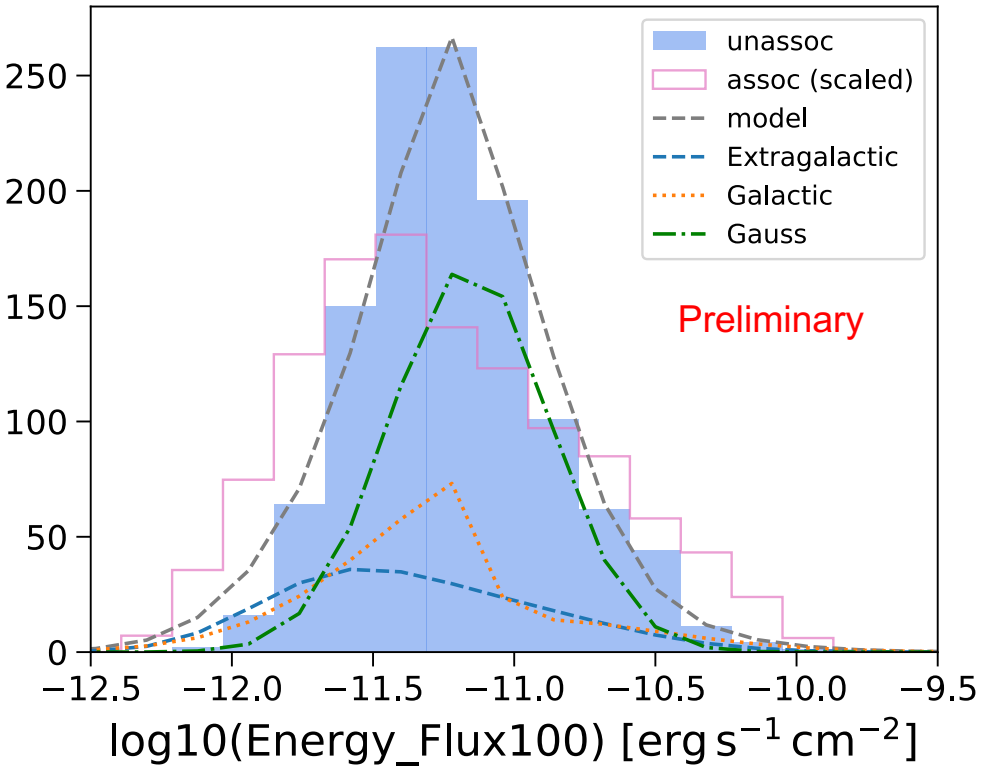




Are SGU-like sources bright?

ERLANGEN CENTRE FOR ASTROPARTICLE PHYSICS

- The energy flux of SGU-like sources is generally larger than for the unassociated sources attributed to Galactic and extragalactic components of the model



Nature of (S)GUs?

ERLANGEN CENTRE FOR ASTROPARTICLE PHYSICS

- The analysis above suggests that about half of GUs have a different distribution in the space of spectral parameters compared to the known classes of gamma-ray sources
  - What is the nature of these sources?
- There is a paper in preparation by the *Fermi* LAT collaboration with a comprehensive study of the GU sources. It contains:
  - An analysis presented above but for 4 classes (dominated by BLLacs, FSRQs, pulsars and MSPs) rather than 2 classes. Results are similar.
  - Searches for counterparts
    - Binaries, pulsars, star-forming regions etc.
    - Multiwavelength study of bright GUs
  - Mismodeled diffuse emission
    - More details on mismodeled diffuse emission: presentation by Jean Ballet at the Gamma 2024 conference
- Details about the project in general:
  - presentation by Benoit Lott at the 11th Fermi symposium

Conclusions

- There is an evidence for a new component among the unassociated *Fermi*-LAT sources with a distribution of spectral parameters different from the distributions of known classes of gamma-ray sources
- The origin of this component is not known. Possible explanations include:
  - Mismodeled diffuse emission, e.g., missing gas?
  - Sub-population of an existing class, e.g., pulsars or MSPs?
  - A “new” population of sources, e.g., young star clusters (Peron+ 2024)
- These explanations cannot account for all SGUs
  - The quest for the origin of SGUs is still open!