A Catalog Pipeline for Sources in the CTA Galactic Plane Survey

Treasures hidden in high energy catalogues
Toulouse, May 2018

Josh Cardenzana for the CTA Consortium
with help from: Jürgen Knödlseder, Luigi Tibaldo (*IRAP*)
Yves Gallant (*LUPM, CNRS/U. Montpellier*)

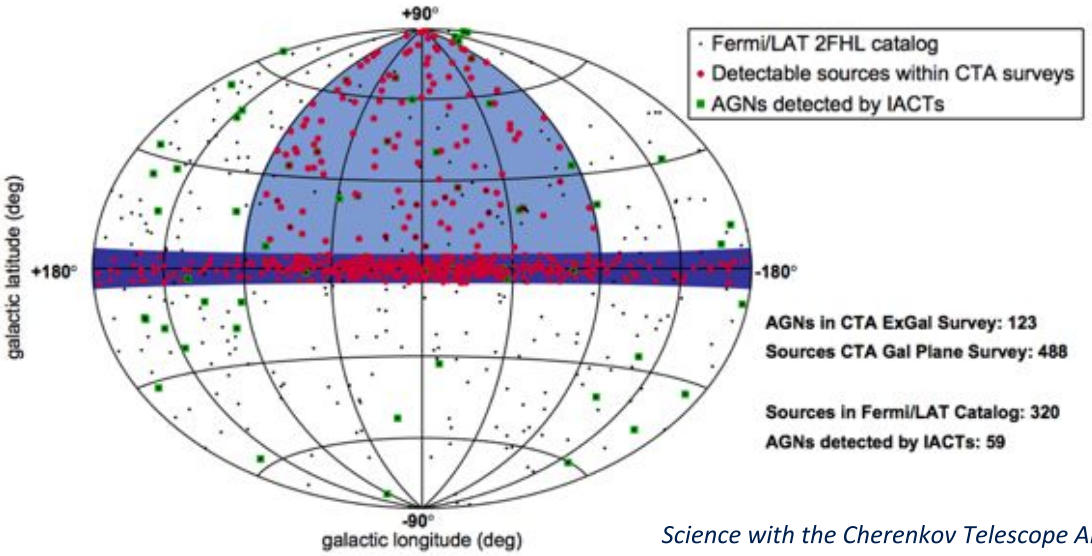
Surveys – Key CTA Science Projects

Extragalactic Survey

- $|l| < 90°$, $b > 5°$
- Better understand TeV population of AGN

Galactic Plane Survey

- $|b| < 5°$
- All longitudes (at varying sensitivity)

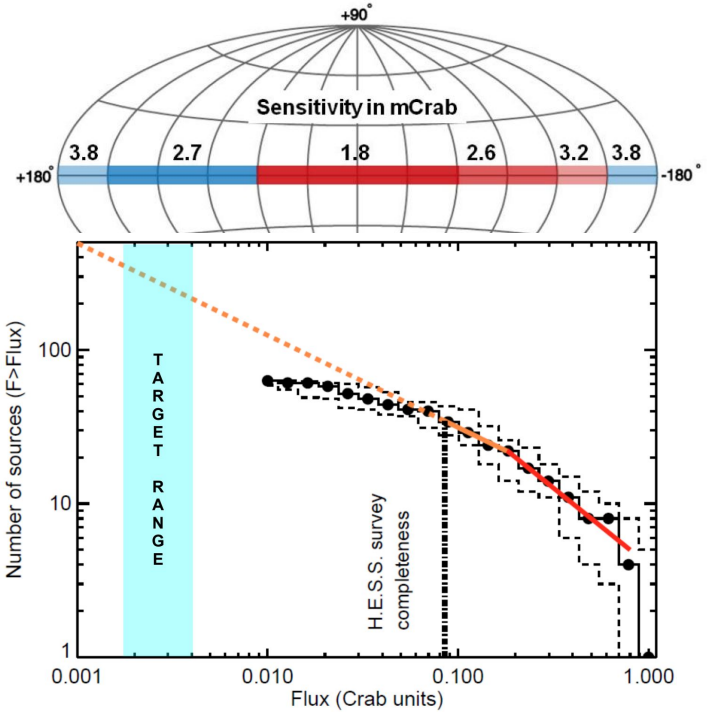


Science with the Cherenkov Telescope Array (2017), arXiv:1709.07997

Galactic Plane Survey (GPS)

Survey of the Galactic plane to address:

- Physics & census of Galactic gamma-ray source populations
  (*SNR, PWNe, binaries, etc…*)
- Identifying possible PeVatron candidates
- Characterize the diffuse Galactic gamma-ray emission
- Study the origin of cosmic rays



Science with the Cherenkov Telescope Array (2017), arXiv:1709.07997

Method Overview

Pipeline for generating catalog of sources from survey data
(*built on GammaLib and ctools*)

The gist:

- Input data and background model
- Returns fully parameterized list of additional sources in that data
- Provide tools for assessing the results of the analysis

Additional Resources:

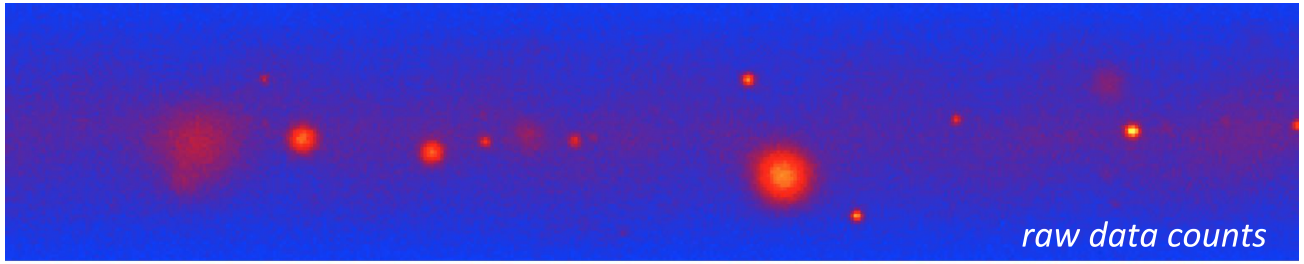
- Tools for source detection
- Tools for iterative source fitting
- Macros for studying results

Source Detection (*finding seeds*)

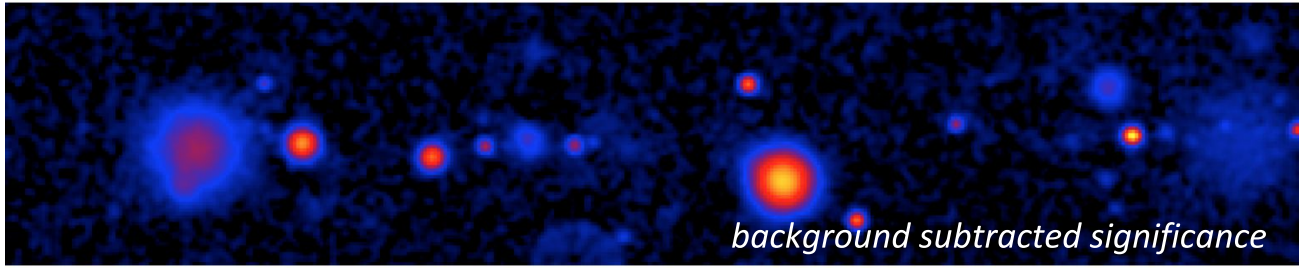
How it works:

- Compute a significance map based on observed & predicted counts
  - depending on region size & model type, this can take a lot of time
- Apply technique based on the SExtractor method (https://doi.org/10.1051/aas:1996164)
  - identifies individual pixels above a given threshold (typically >3σ)
  - connect nearby pixels to identify “objects” (these are individual sources)
  - done in a single pass over the map (i.e. it’s pretty fast)
  - de-blend objects to detect overlapping sources
  - **Tuning parameters:** significance threshold, pixels per object, deblend levels

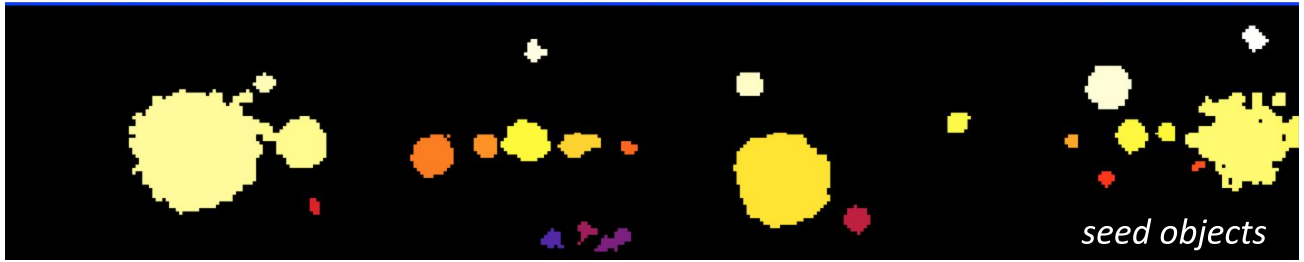
Source Detection (*finding seeds*)



Raw data counts map



Significance map after accounting for expected background events and interstellar emission



Detected seeds input to fitting algorithm (*colored by seed ID*)

Source Fitting

How it works:

- Take seeds from source detection
  - Each source starts as point source with sub-sensitivity flux
- Iteratively fit all sources:
  - Fit source
    - If source is new, test extension (*disk*, *Gaussian*)
  - Evaluate fit validity (*reasonable parameters & extension*)
  - Do this for every source
- Loop until all sources have converged ($\Delta TS < 10$)
  - Remove insignificant sources (*final* $TS < 10$)
- Test for spectral curvature
- Reoptimize all parameters

cherenkov telescope array

Testing the Pipeline

Testing on Simulation of GPS

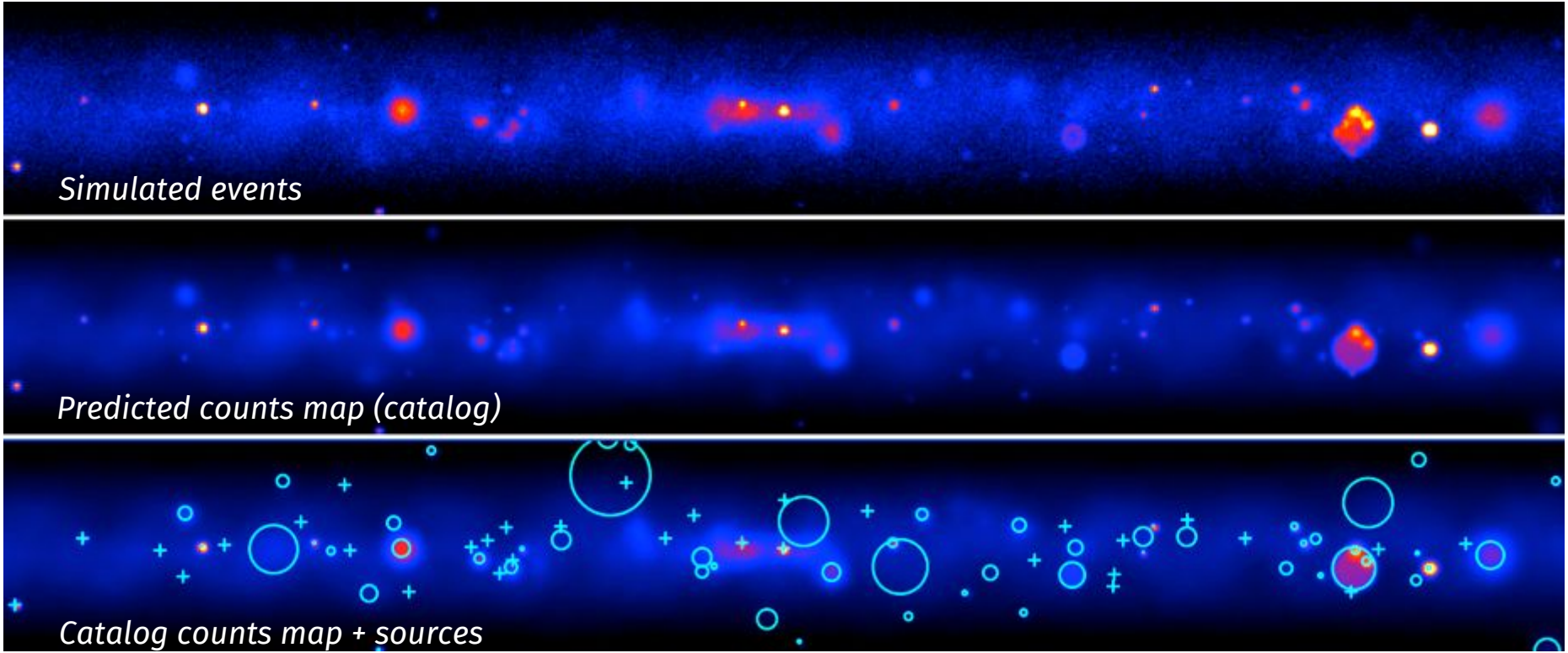
Purpose:

- Stress test analysis methods on mock-up “data”
- Simulate (*somewhat realistically*) the CTA GPS data will obtain
  - Includes sources modeled based on existing observations

Caveat:

- ***The simulation should only be considered as a means to test software tools and not a representation of what CTA will ultimately see.***

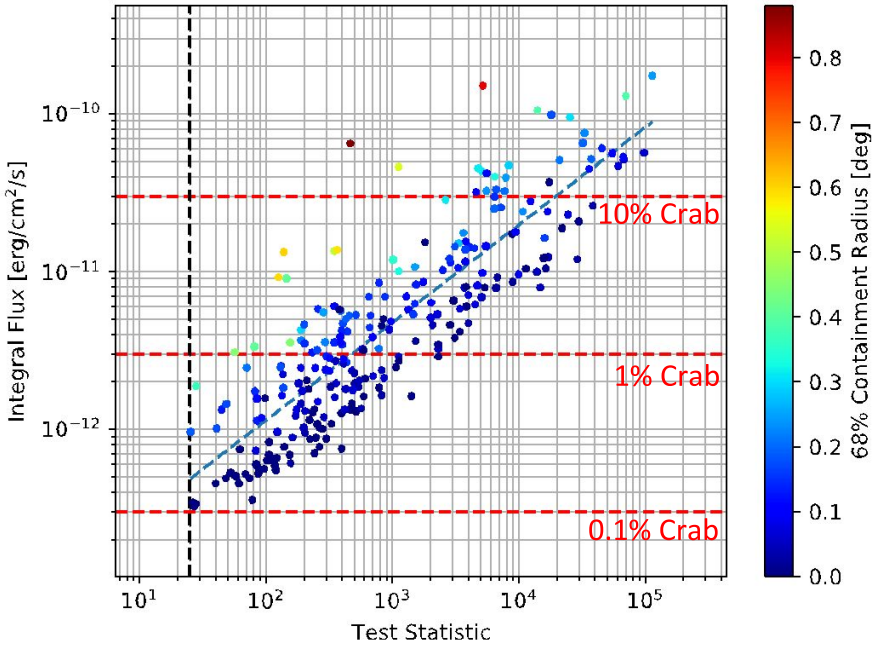
Testing on Simulation of GPS



Source Distribution – TS vs. Flux

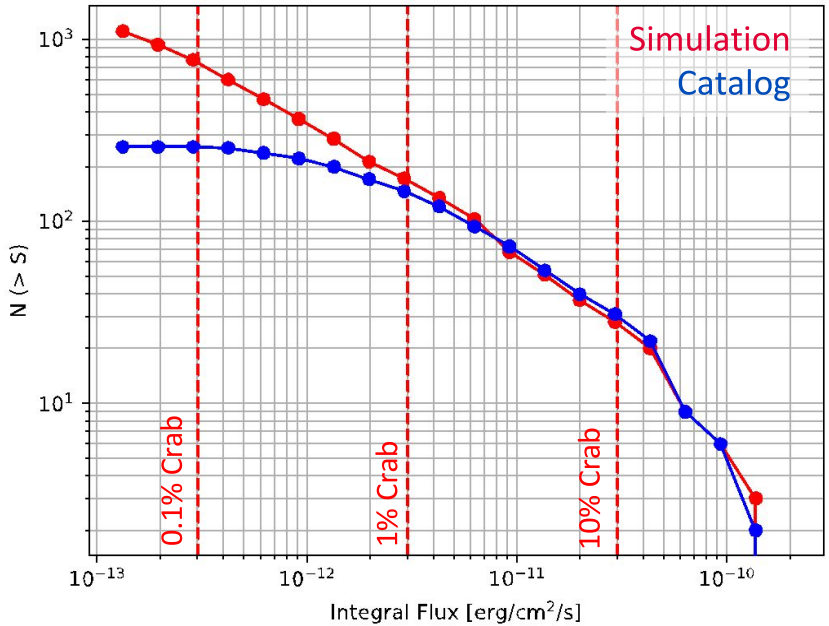
Noteable features:

- Trend between detection significance & flux
- Appear to miss sources at low flux
  - Optimize detection parameters?
  - Optimize energy threshold for detection



Comparison to Simulation – logN-LogS

- Appear to miss sources at low flux
  - May need looser detection criteria
  - Alternative source detection methods



Comparison to Simulation

Each source is compared to simulated sources by:

- Centroid position & 68% containment radius (*closest match => association*)
- No consideration for flux (*to avoid biases*)

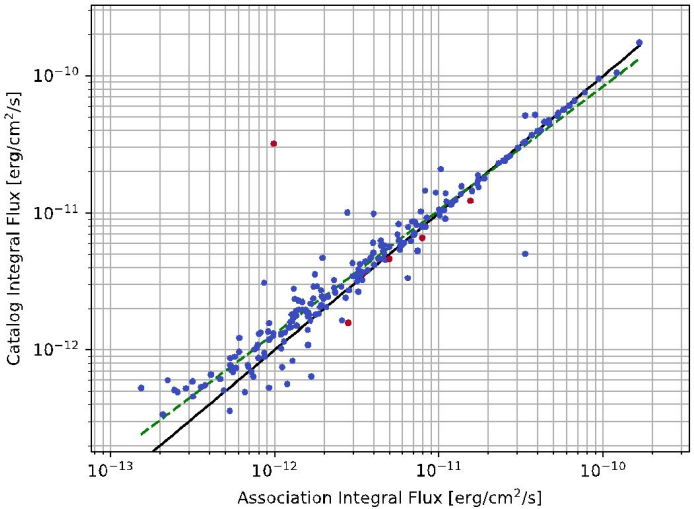
Results:

- Associated: 80.6%
- Unassociated: 19.4%
  - Spurious sources
    (*e.g. fluctuations in diffuse gamma-ray background*)
  - Source confusion
    (*e.g. multiple overlapping sources detected as single source*)

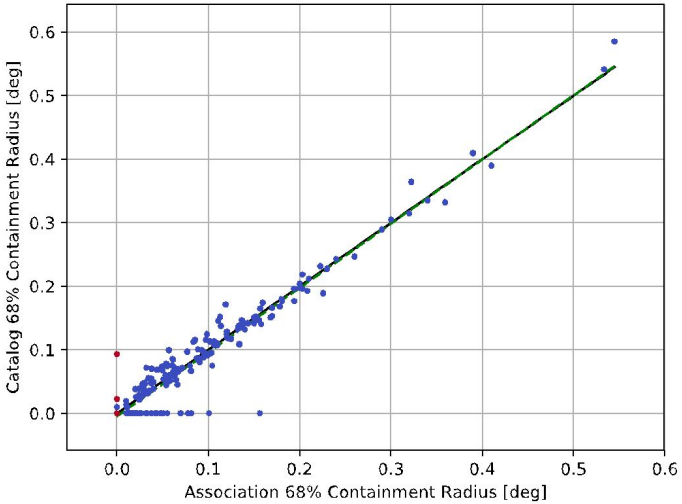
Comparison to Simulation

- Integrated source flux (*0.1 – 100 TeV*)
- Red dots denote variable sources (*binaries* & *pulsars*)

Flux



Radius

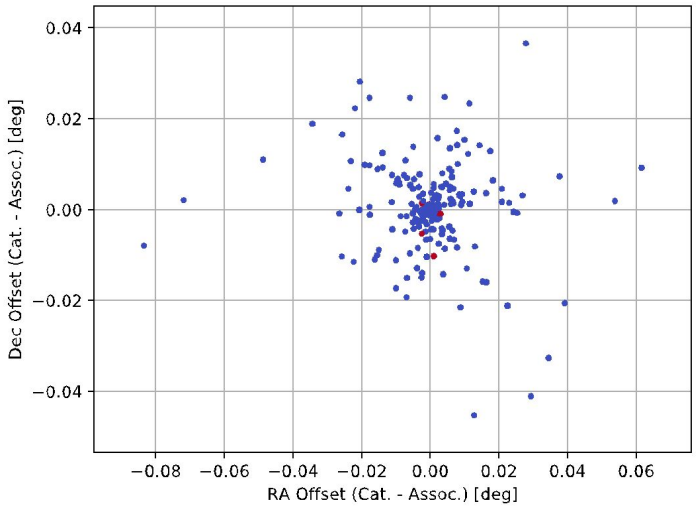


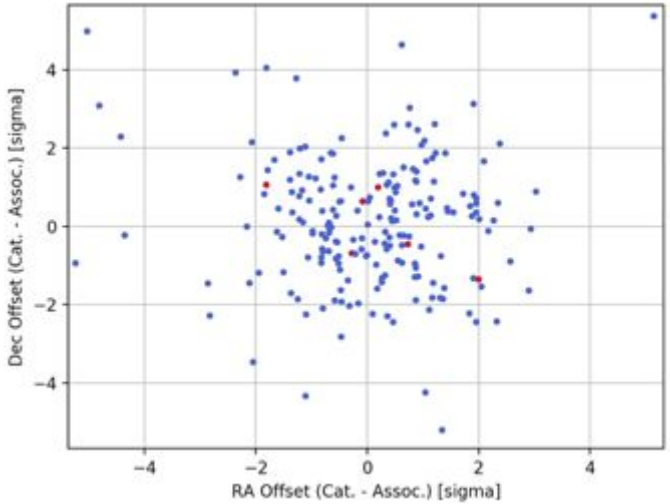
Black line:
1:1 relation
Green line:
Fit to data

Comparison to Simulation - Centroid

Distribution of catalog RA,Dec vs. association RA,Dec:

- Most points within 1 bin width (0.02°)
- Most sources reasonably close to the corresponding association



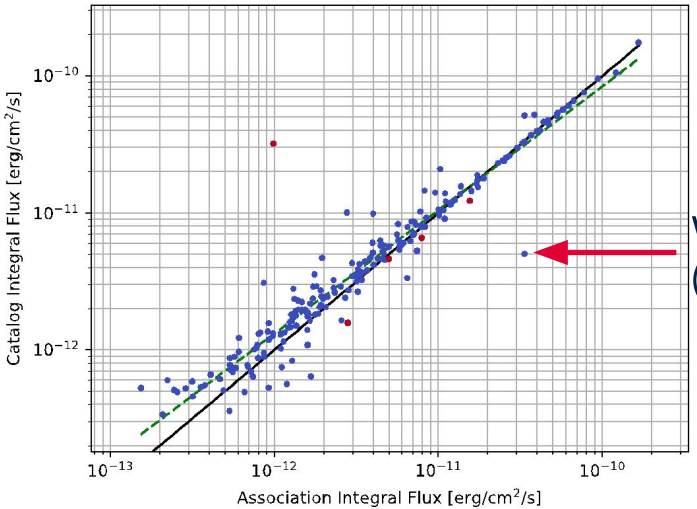


Source Confusion:

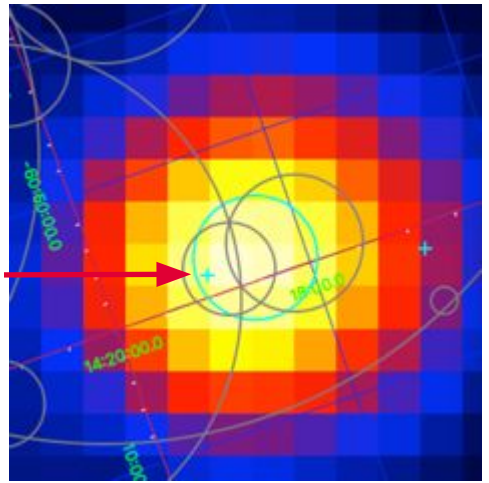
- Overlapping sources
- Properly modeling extended sources

Overlapping sources

- Another large outlier in the flux plot is actually a multiple association



What about here?
(it's the point source)

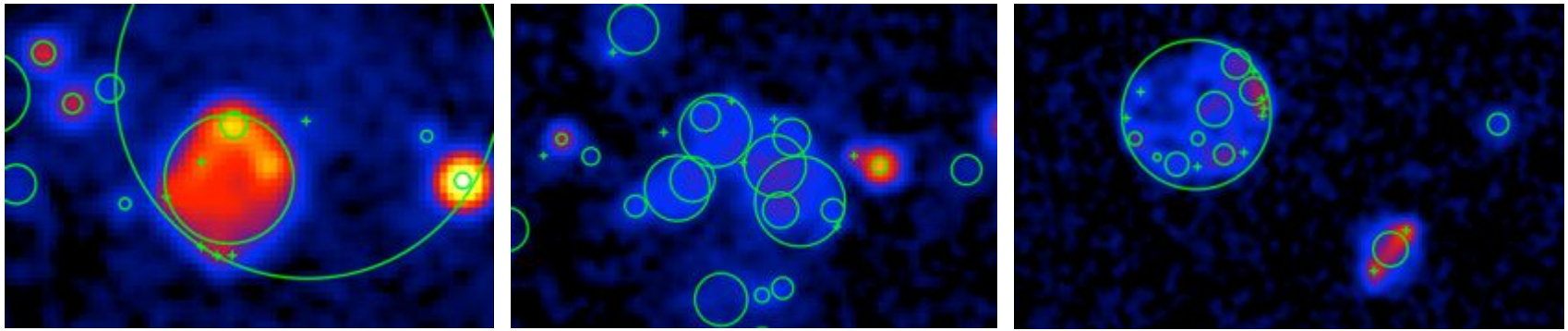


Cyan: Catalog
Gray: Simulated

Issue – Diffuse sources

Multiple detections are a problem for extended structures

- Most sources will NOT be well described as a ‘Gaussian’ or ’disk’
- How to identify these objects as the same object?



Summary

Latest work:

- Framework exists for detecting & characterizing sources in CTA data
- On the right track (*also not the only survey tools being developed*)

Things to be Mindful of:

- Cannot try every model combination:
  - Limits on computation time, degrades significances
- Extended, diffuse sources are complicated
  - Can reconstruct as multiple overlapping sources (*also seen in HGPS*)
  - Need to statistically assess how likely these are to be the same source

Open questions:

- Do uncertainties in gamma-ray background impact detection/characterization?
  - Interstellar gamma-ray background + unresolved sources
  - Can be tested through simulations
- Investigate source detection optimizations
  - Adjust energy range, detection parameters, etc…
- MWL association of sources with other catalogs