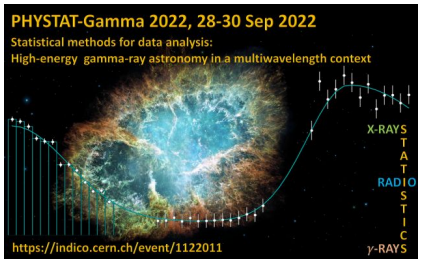
Very-high-energy γ-ray surveys with CTA

Quentin Remy & Jean-Philippe Lenain



PHYSTAT-Gamma 2022, Sept. 28–30, 2022

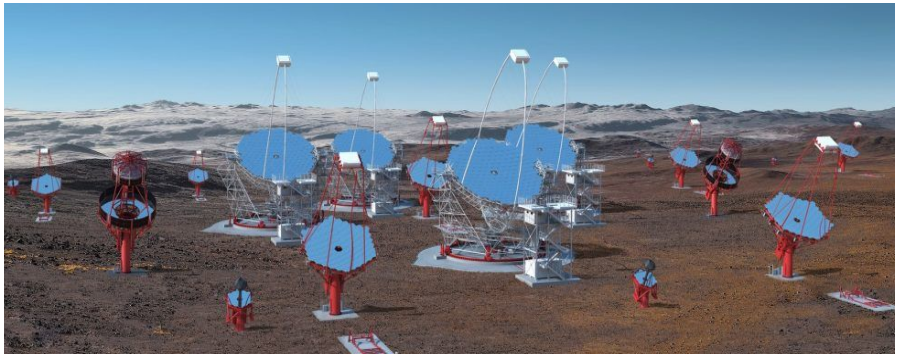
Cherenkov Telescope Array (CTA)

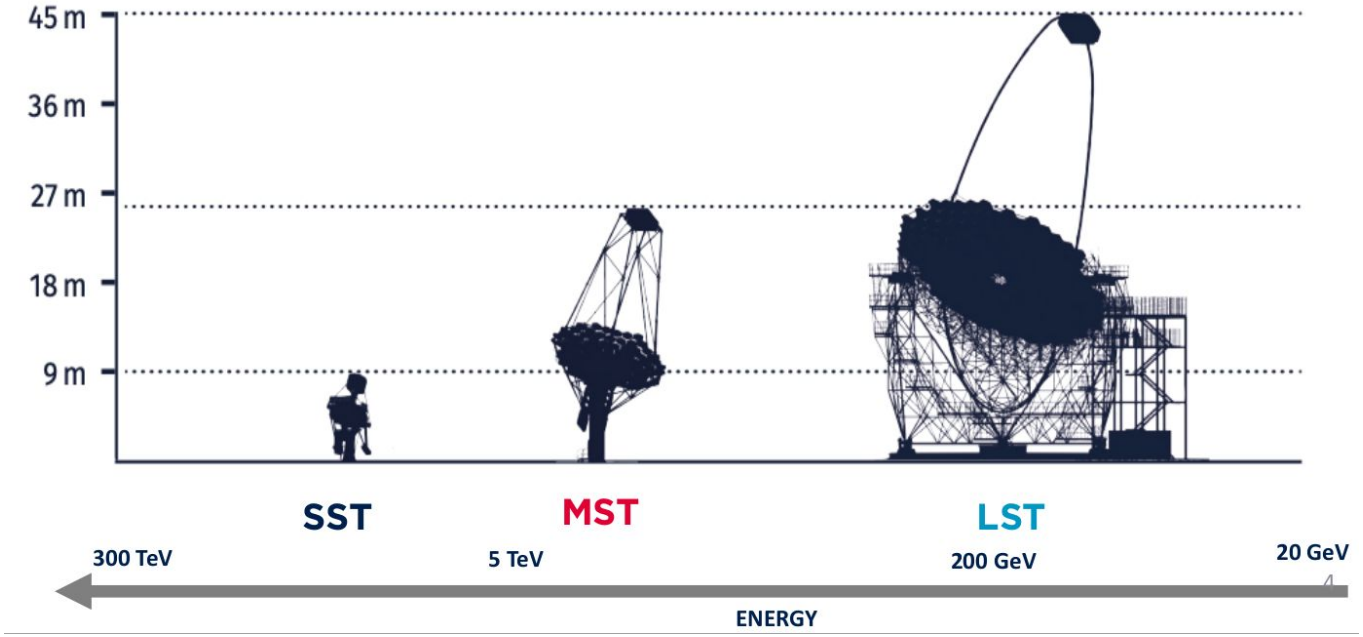
The first ground-based gamma-ray observatory

CTAO sites

- La Palma, Spain:
  9 MSTs + 4 LSTs

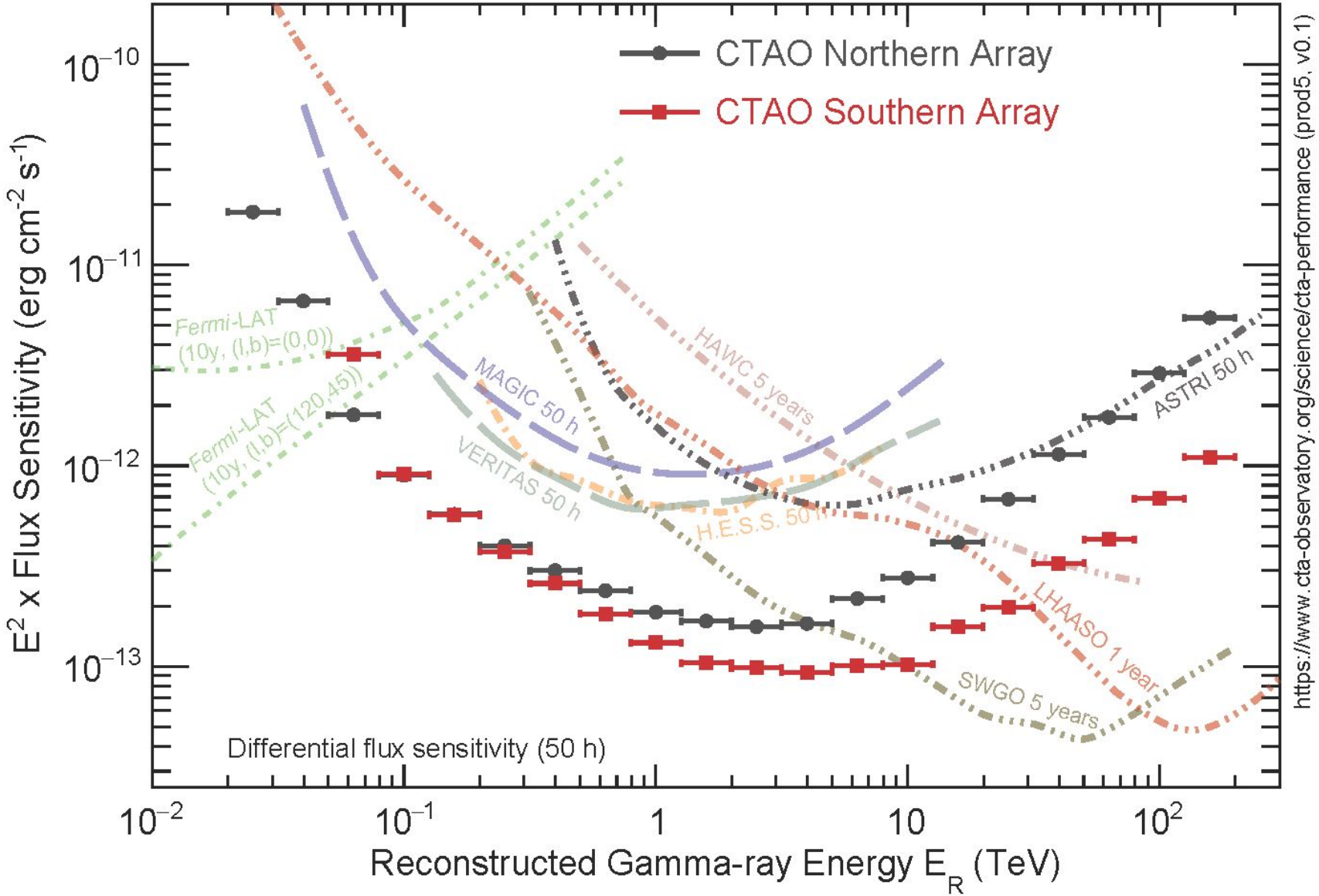
- Paranal, Chile:
  37 SSTs + 14 MSTs + (4 LSTs)





Alpha configuration

CTAO performances

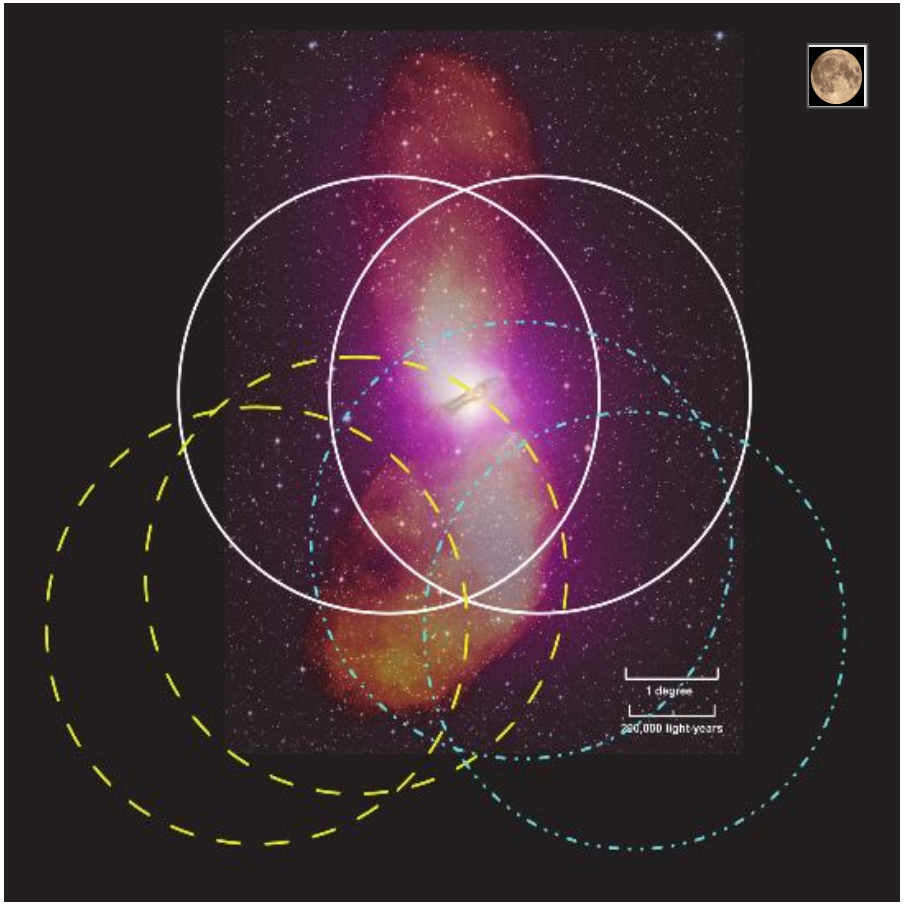


https://www.cta-observatory.org/science/ctao-performance/

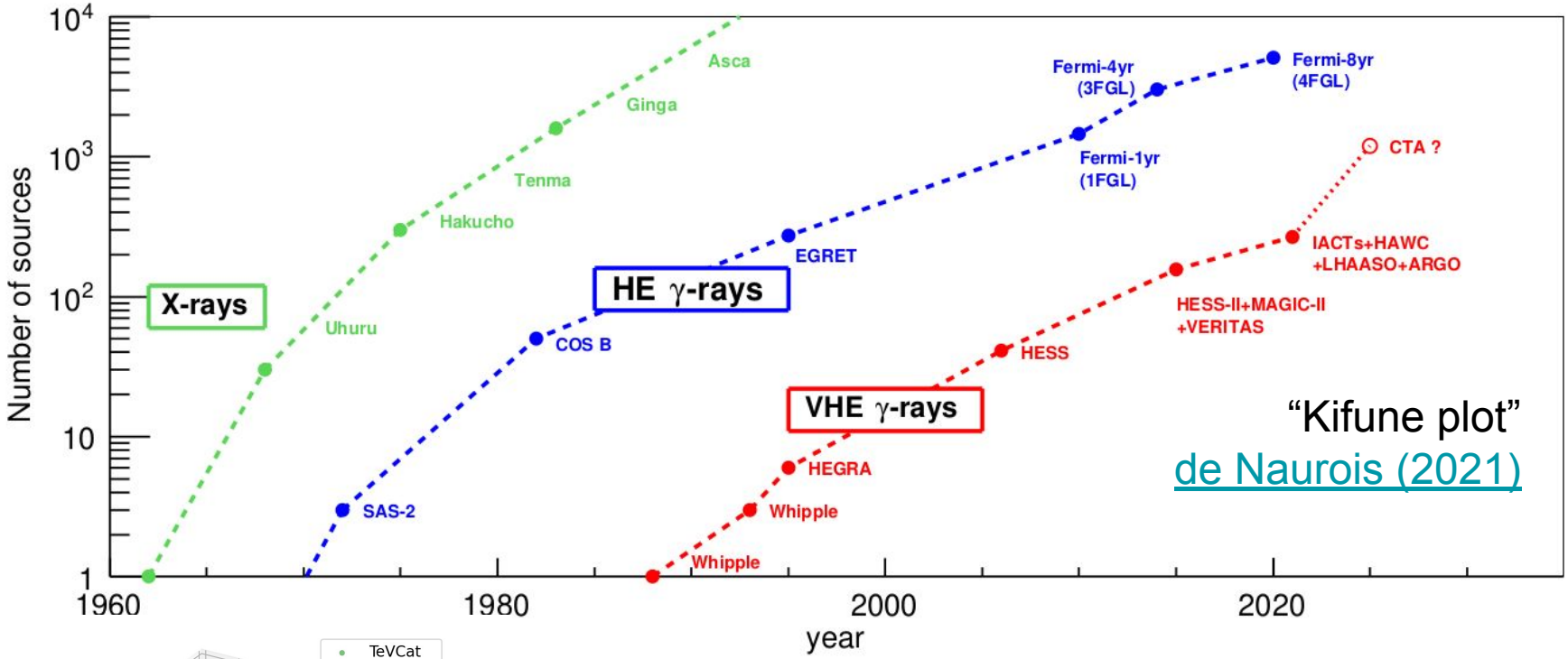
Large field of view

Field of view: Up to ~7-8°

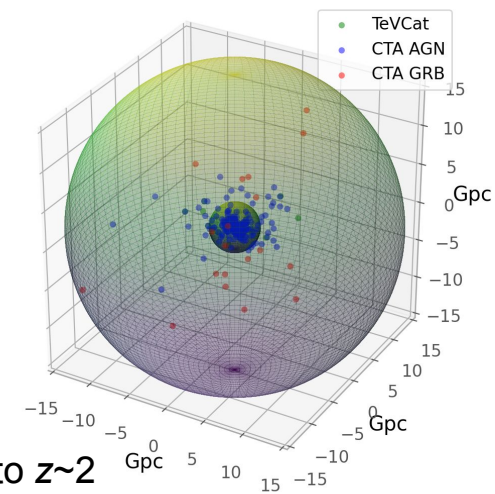
High probability to catch serendipitous sources and transients.



x10 more VHE γ-ray sources every 20 years



"Kifune plot"
de Naurois (2021)



Up to *z*~2

previous IACTs : ~ 1 source per PhD student
CTA : ~ 10x more sources
- dedicated analyses only for a small fraction
- general catalog need to be continuously updated
- more suitable for population studies

Surveys with CTA

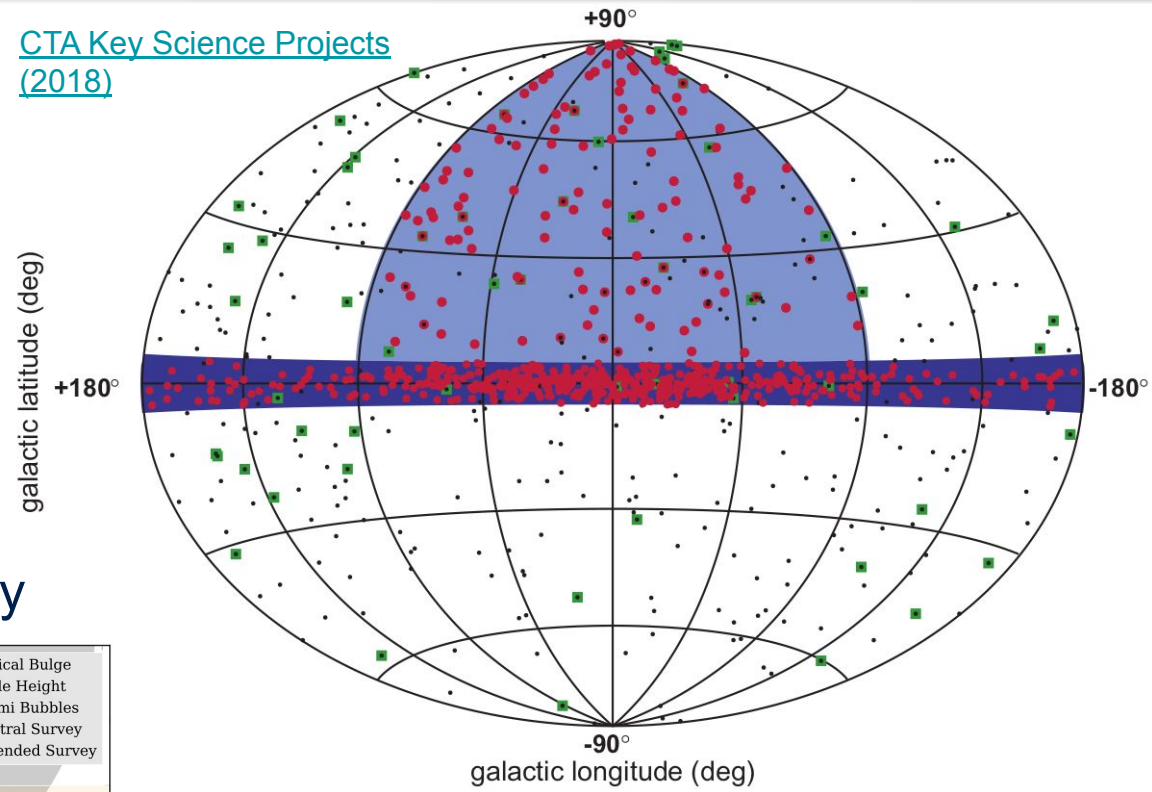
CTA2 large surveys + deep focused regions

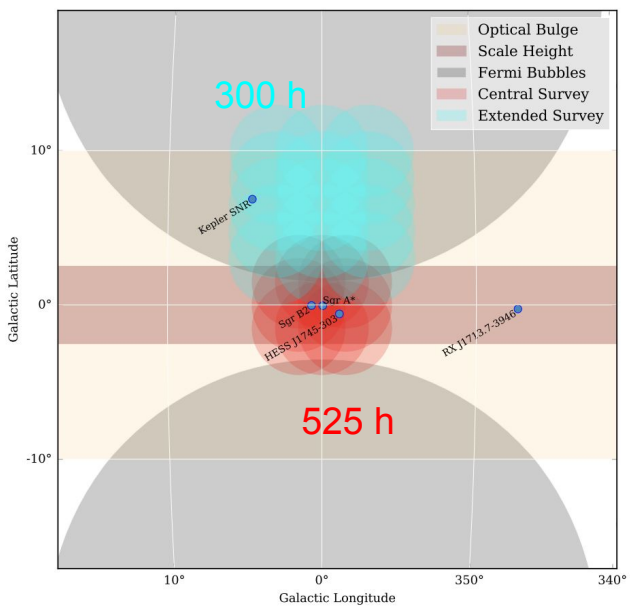
CTA Key Science Projects (2018)

- Galactic plane

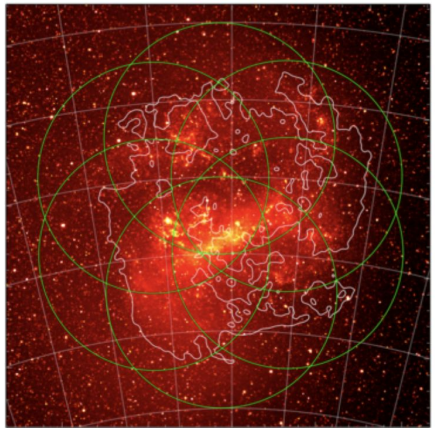
- Extragalactic sky
  25% of the full sky

- Deep Galactic Center survey



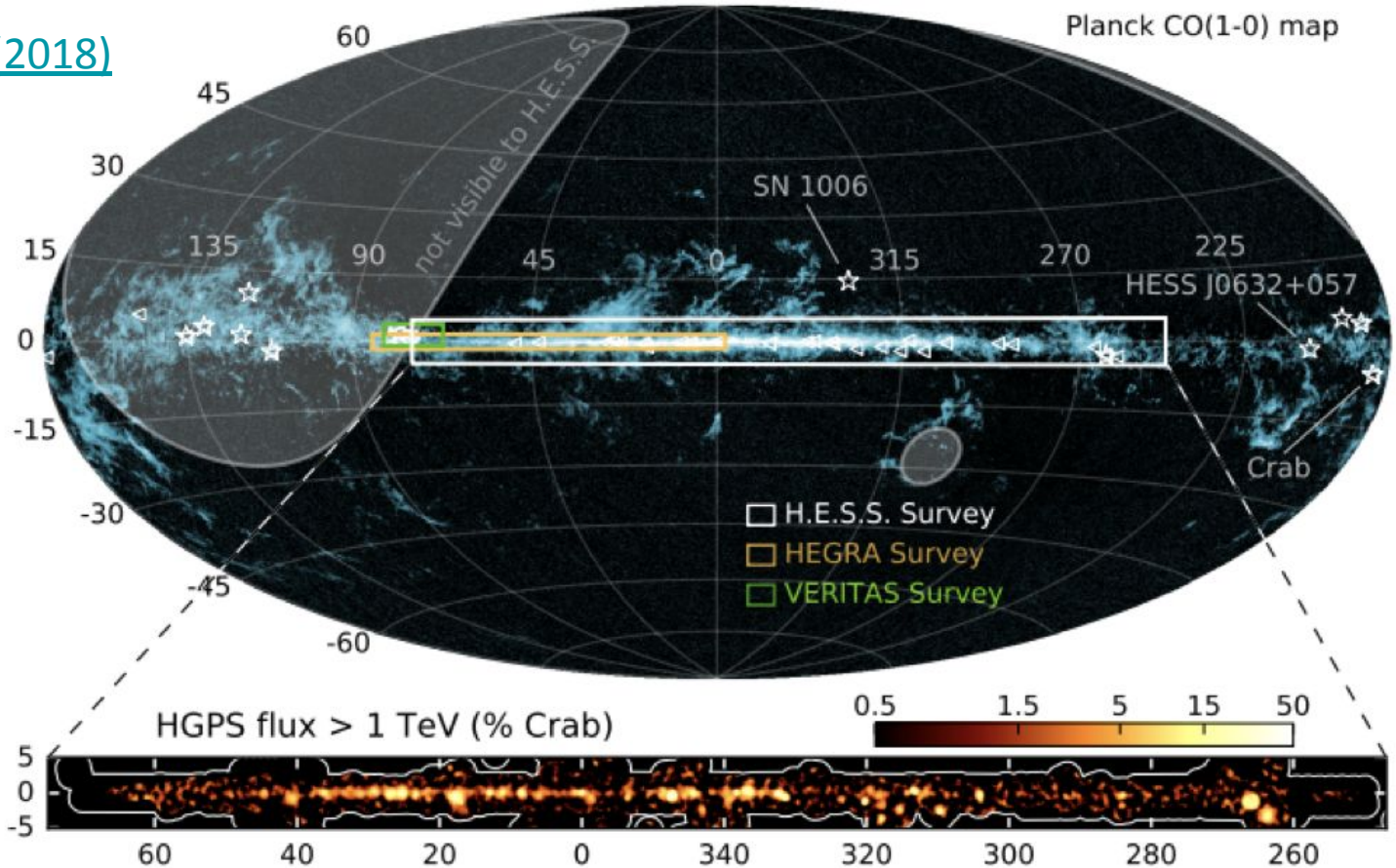


- LMC region



Galactic Plane : VHE γ-ray surveys from IACTs

HESS-GPS (2018)



CTA Galactic Plane Survey (GPS)

5-20x more sensitive than previous surveys

Goals : - unprecedented census of VHE emitters in the entire Galactic plane

- studying diffuse gamma-ray emission
- searching for new and unexpected phenomena

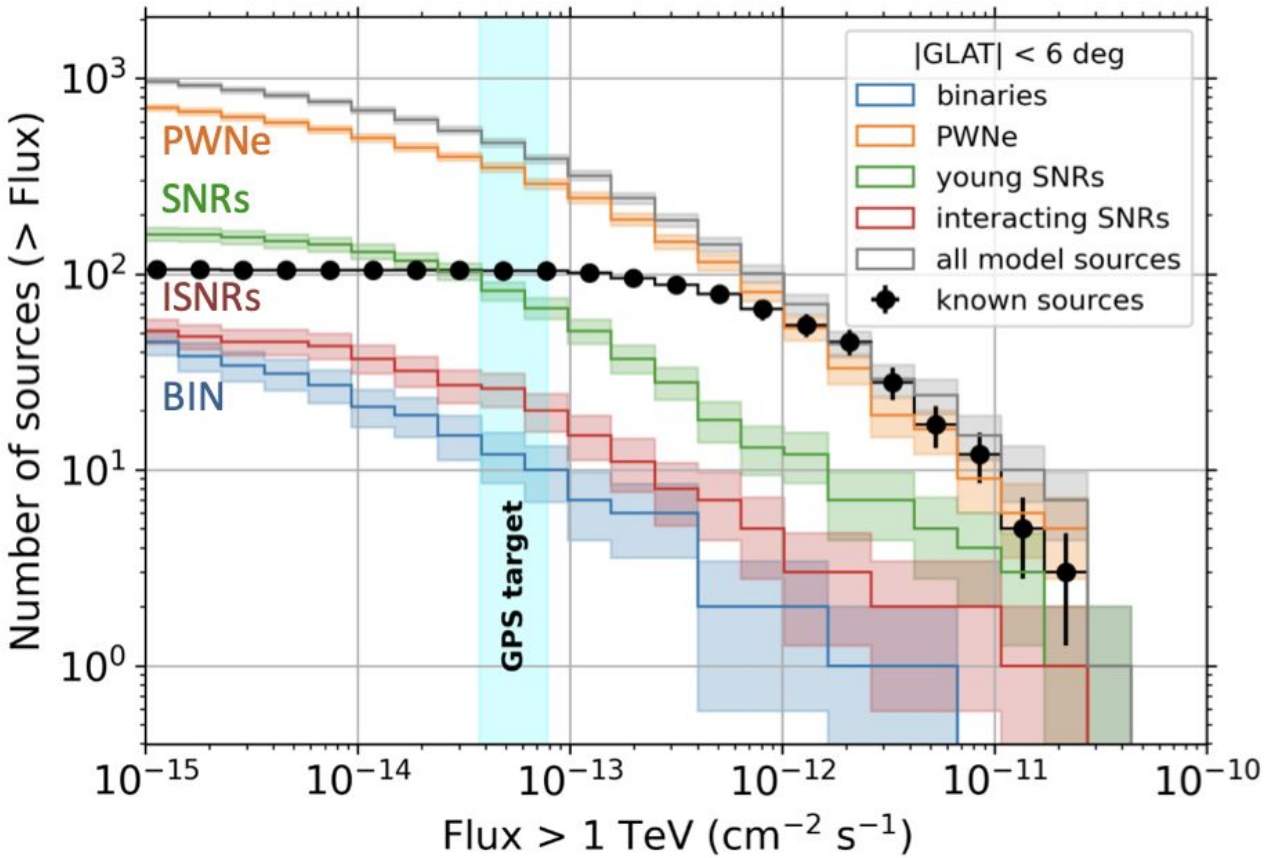
GPS simulation : sky model

Known sources

- IACTS sources compilation (gamma-cat.readthedocs.io)
- Fermi-LAT 3FHL
- 2HAWC

Source population synthesis based on physical modelling

- binaries (Dubus et al. 2017)
- pulsar wind nebulae (Fiori et al. 2021)
- supernova remnants: young and interacting with interstellar medium (Cristofari et al. 2017, Rice et al. 2016)

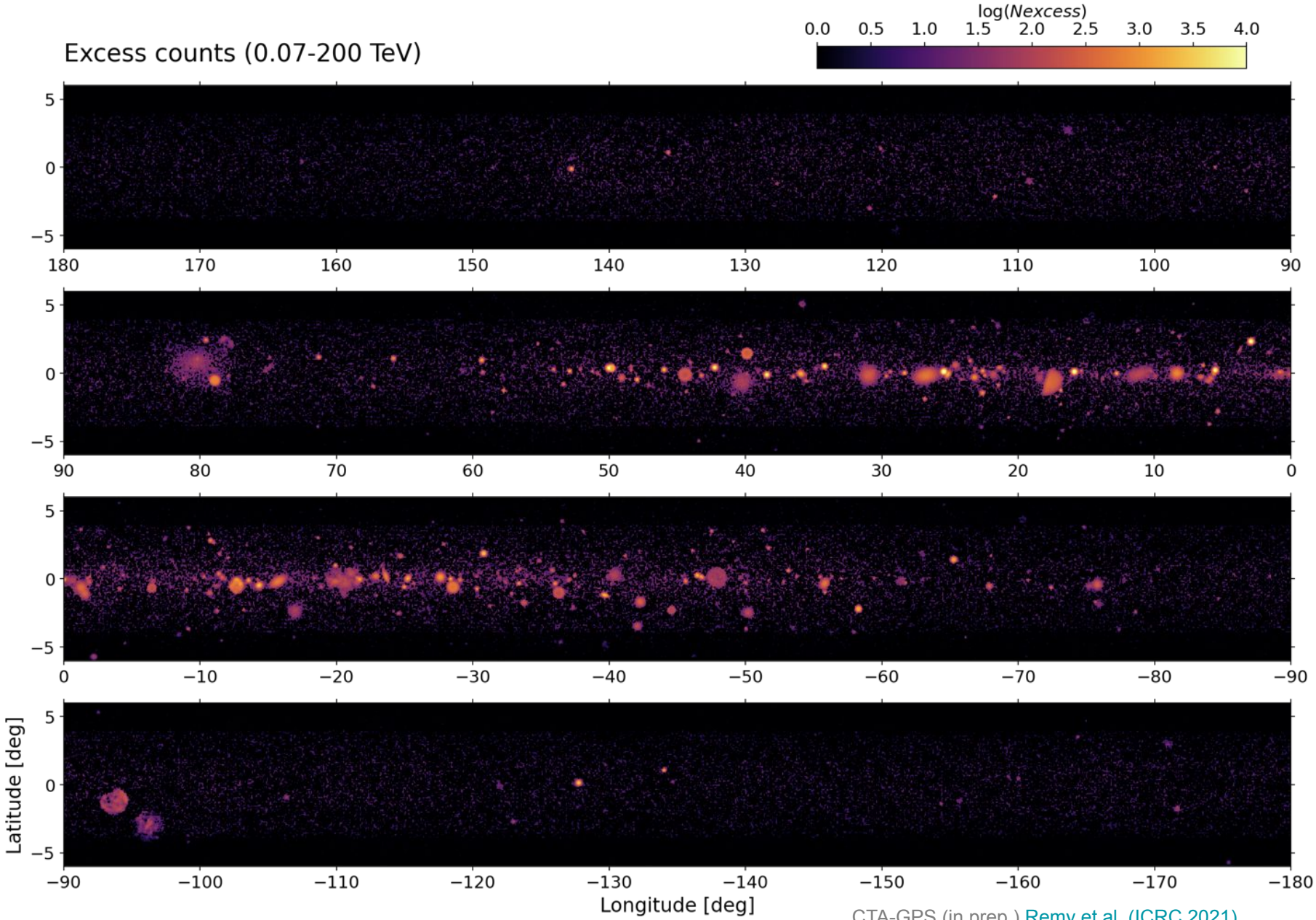


CTA-GPS (in prep.)
Remy et al. (ICRC 2021)

Interstellar emission

- Galactic Ridge - Fermi-bubbles
- minimal model for gamma-ray emission from Galactic cosmic rays using DRAGON cosmic-ray propagation code

GPS simulation



CTA-GPS (in prep.) Remy et al. (ICRC 2021)

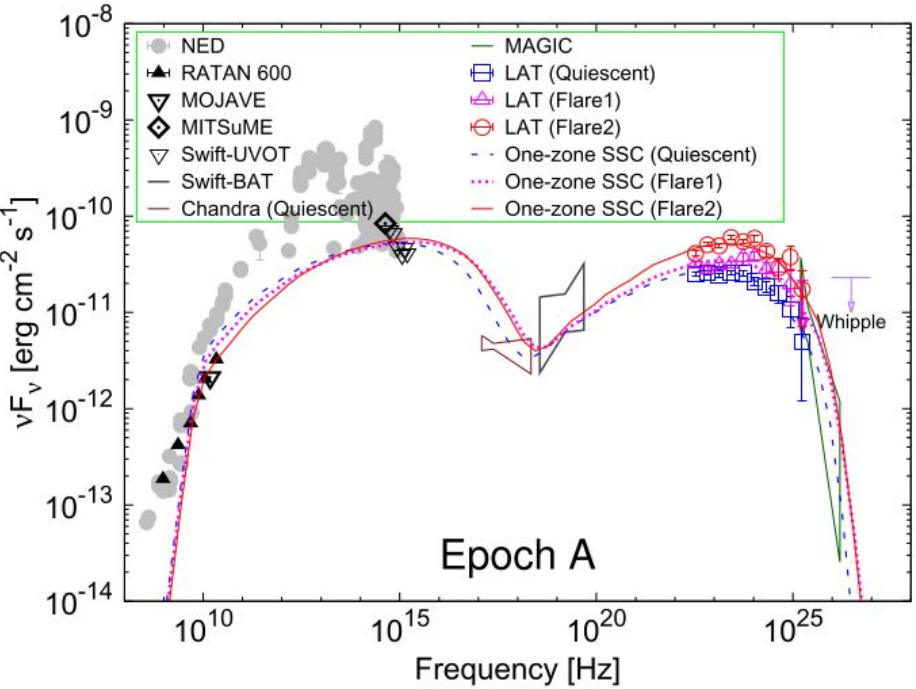
Statistics and data analysis challenges

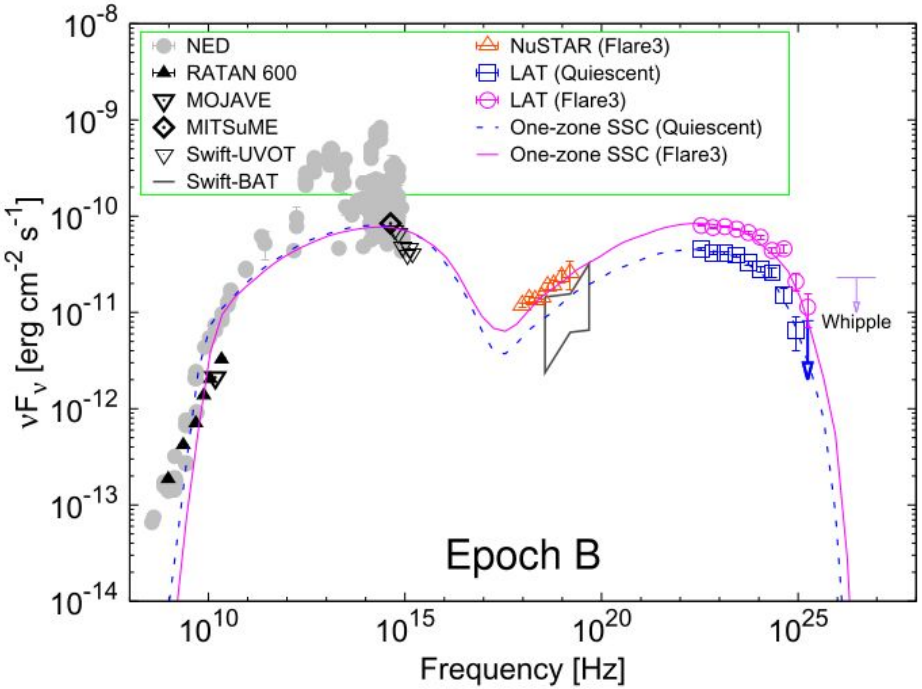
A few open questions on the exploitation strategies

- Variability
- Real-time analysis
- Source confusion
- Extended sources modelling
- Catalog cross-matches and associations by chance
- Instrumental and astrophysical backgrounds modelling

Variability

Active galactic nuclei show different variability patterns, depending on source class (emission mechanism) and flare occurrence



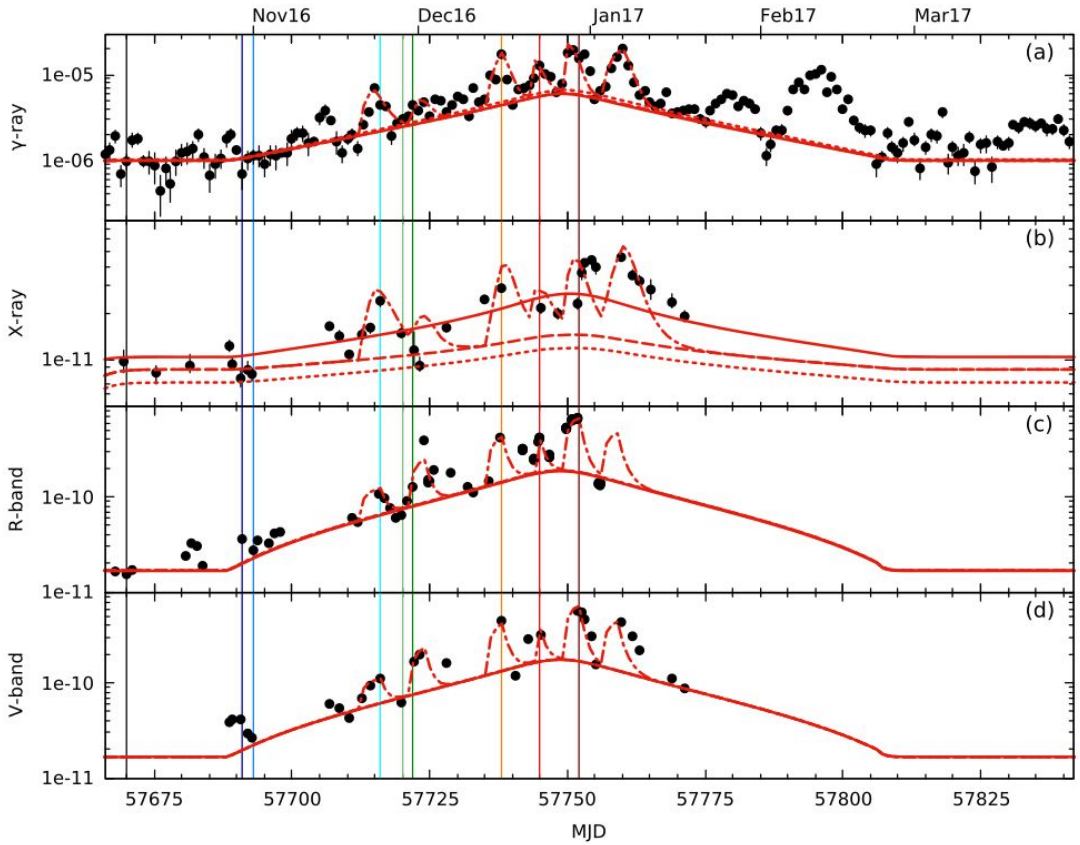


Tanada et al. (2018)

Variability

Active galactic nuclei show different variability patterns, depending on source class (emission mechanism) and flare occurrence

- Long trends
- Short, intense flares

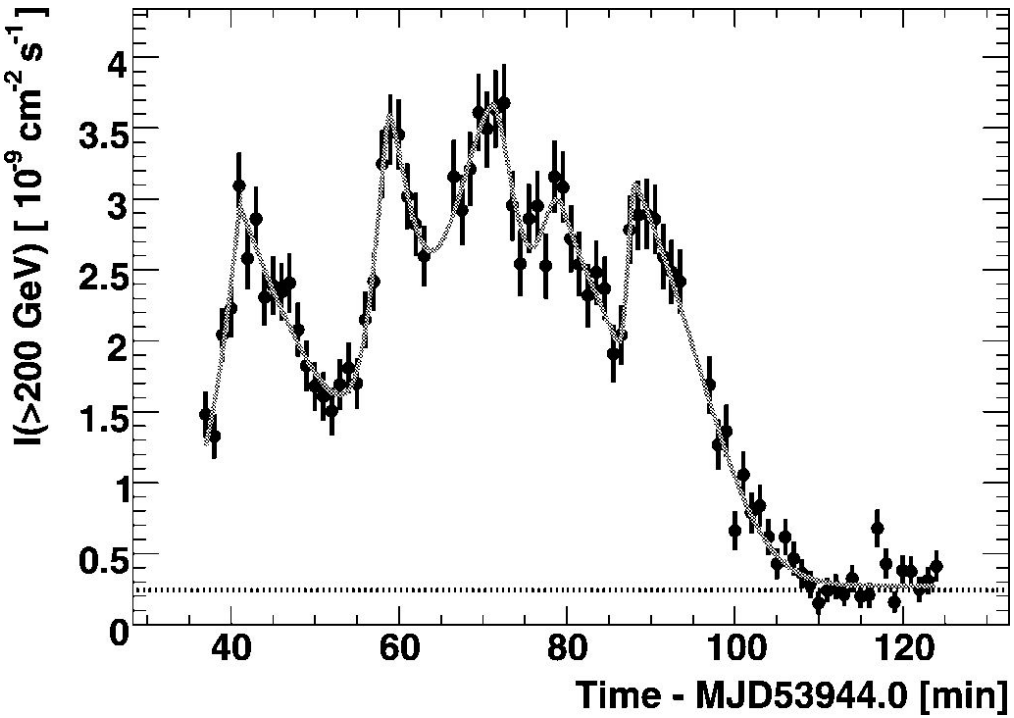


CTA 102 (FSRQ), Zacharias et al. (2019)

Variability

Active galactic nuclei show different variability patterns, depending on source class (emission mechanism) and flare occurrence

- Long trends
- Short, intense flares

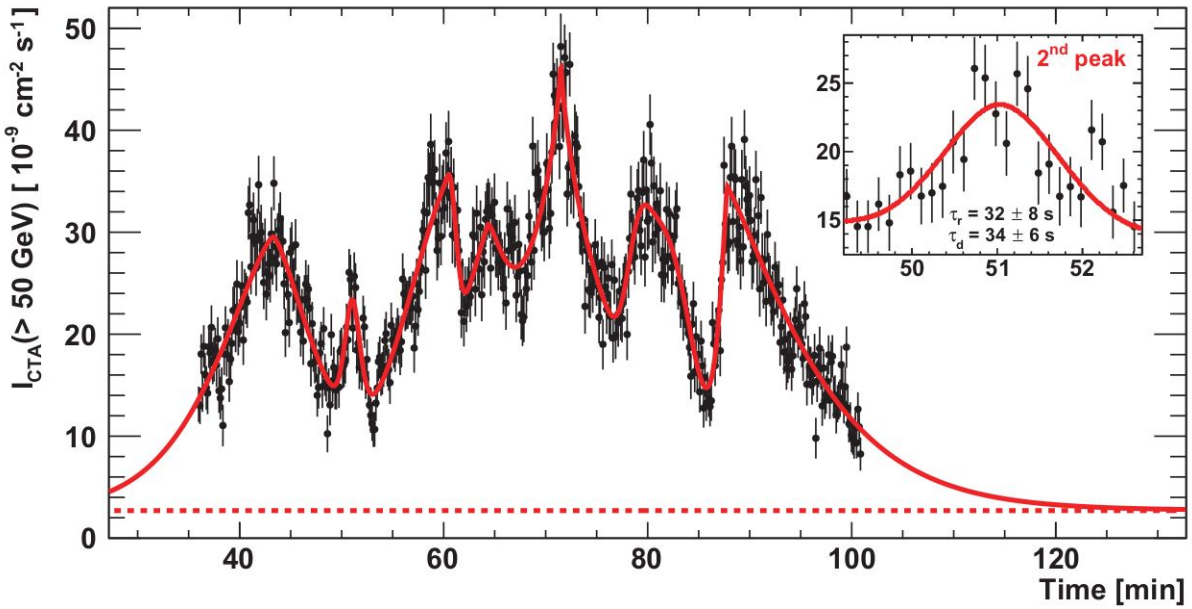


Aharonian et al. (2007)

Variability

Active galactic nuclei show different variability patterns, depending on source class (emission mechanism) and flare occurrence

- Long trends
- Short, intense flares



The CTA Consortium (2019)

Time-series analysis

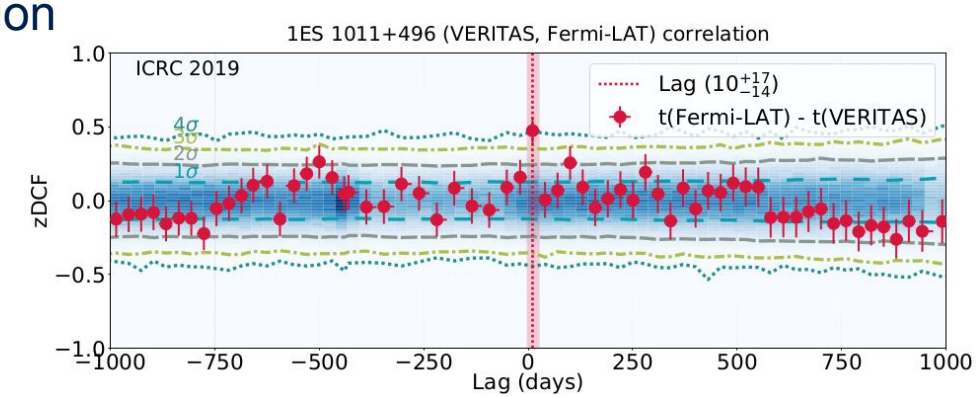
- Depending on emission mechanism and wavelength bands, multi-band spectra can be:
  - perfectly correlated
    ex.: expectation between X-ray and VHE within SSC framework (same leptons probed in both bands)
  - mildly correlated
    ex.: External inverse Compton emission in FSRQs, hadronic emission processes
  - not correlated:
    ex.: VHE “orphan” flare

- Multi-wavelength cross-correlation

Ex.: Unevenly sampled data

DCF: Discrete Correlation Function
(Edelson & Krolik, 1988)

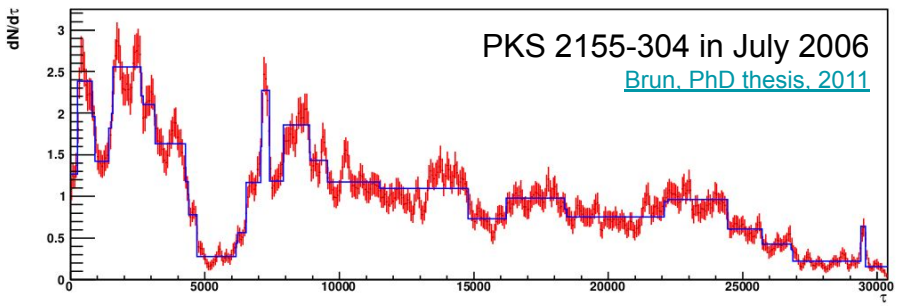
ZDCF: z-transformed Discrete Correlation Function to handle large time gaps.
(Alexander, 1997)



Gueta, O. (for the VERITAS collaboration, ICRC2019)

Time-series analysis

- Burst searches:
  - At known target position
  - Serendipitous flares within the field of view
  - Examples:
    - Bayesian blocks (Scargle, 1998)
    - exp-test (Prahl, 1999)



Poisson regime
PDF:

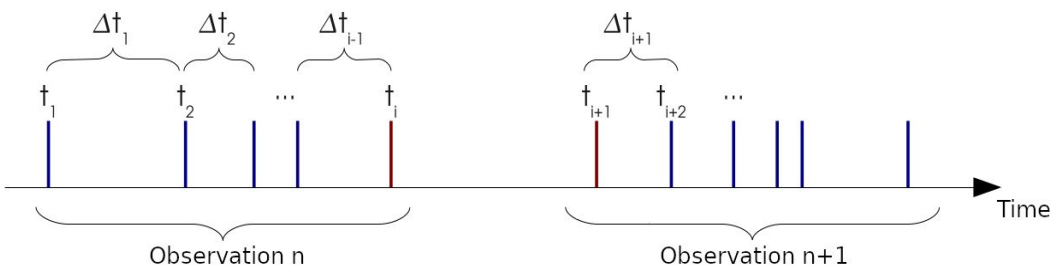
$$f_C(\Delta t) = \frac{1}{C} \cdot \exp\left(-\frac{\Delta t}{C}\right)$$

Exp-test estimator:

$$M = \frac{1}{N} \sum_{\Delta T_i < C^*} \left(1 - \frac{\Delta T_i}{C^*}\right)$$

with: $\{\Delta T_i\}_{i=1\ldots N} := \{(T_{i+1} - T_i)\}_{i=1\ldots N}$

$$\overline{\Delta T} =: C^*$$

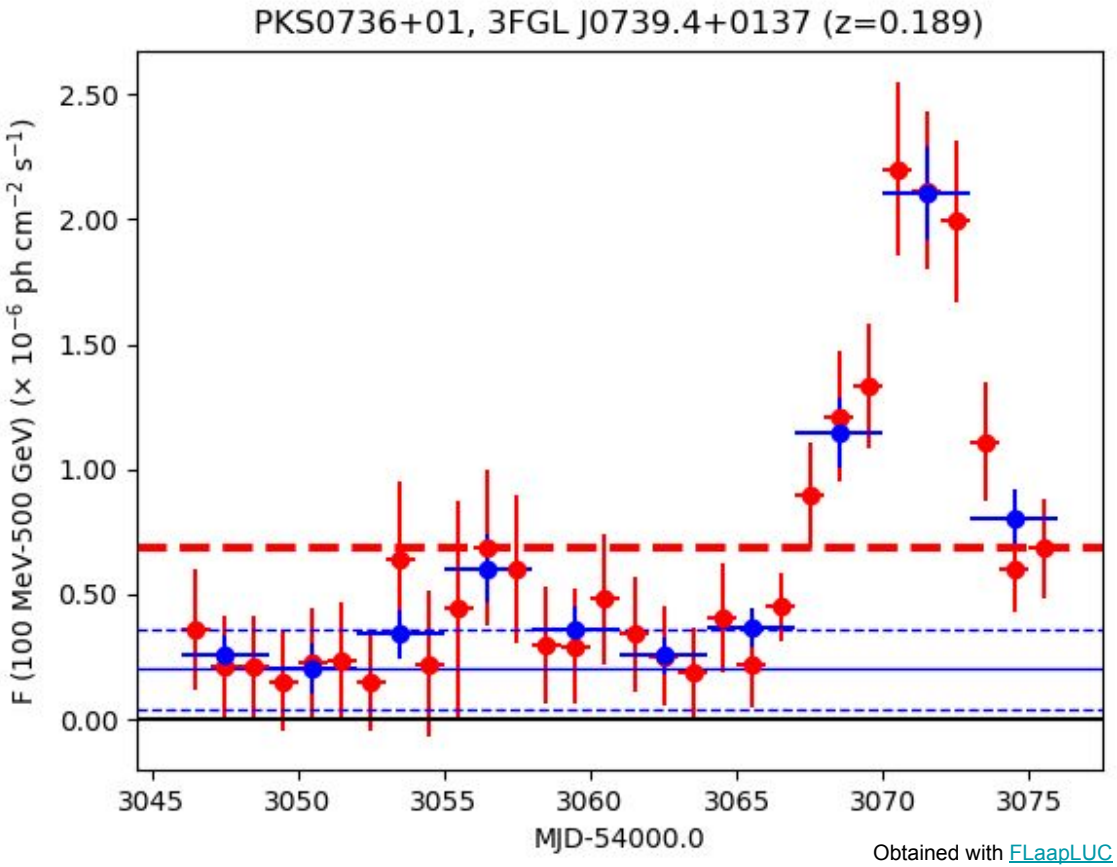


Brun, PhD thesis, 2011

Time-series analysis

How to identify a flare in (near-)real time ?

- A posteriori flare identification:

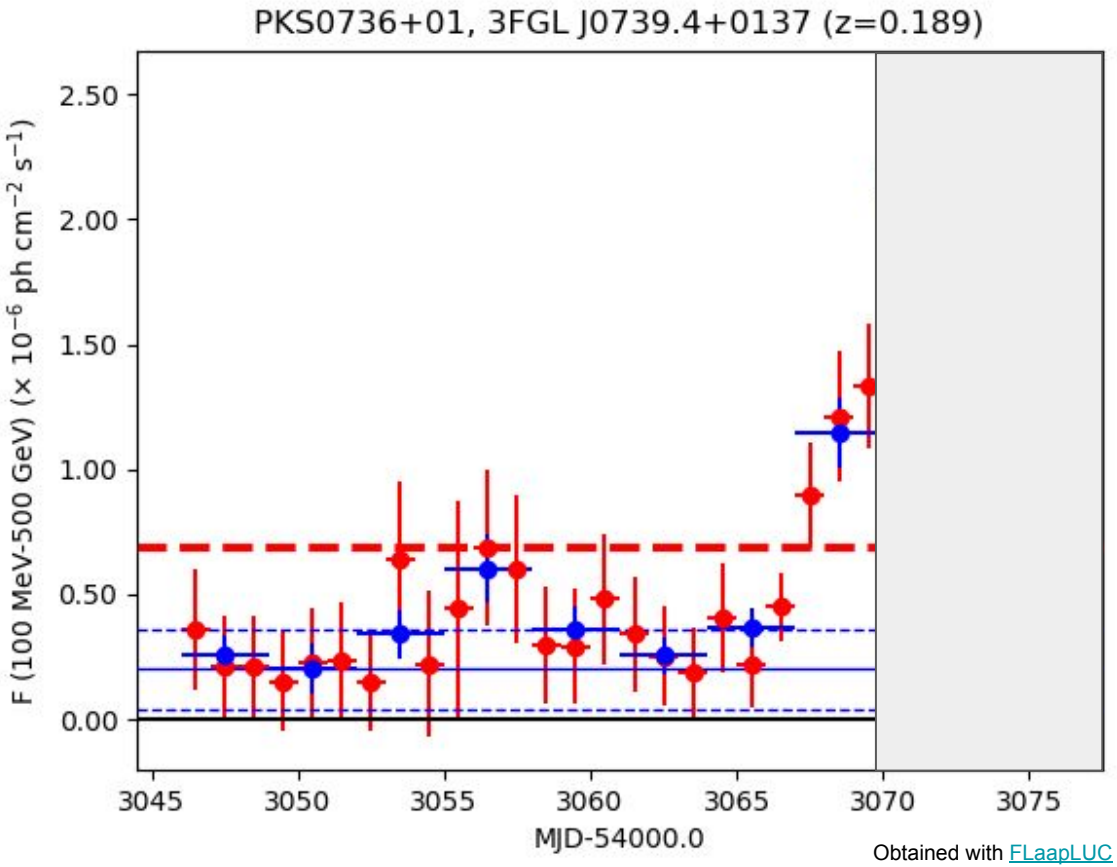


Obtained with FLaapLUC
Lenain (2018)

Time-series analysis

How to identify a flare in (near-)real time ?

- How to assess a significant activity increase when a flare is building up ?



Obtained with FLaapLUC
Lenain (2018)

Time-series analysis

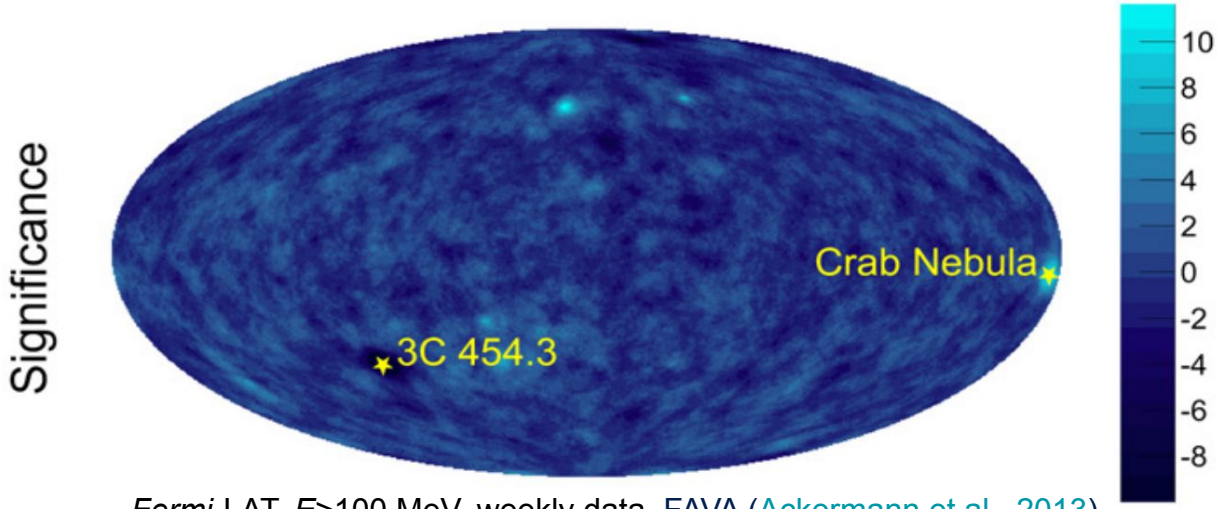
How to identify a flare in (near-)real time ?

- How to assess a significant activity increase when a flare is building up ?

Event clustering analysis: e.g. FAVA (Ackermann et al., 2013) for *Fermi*-LAT data.
Compare:

$$N^{exp}(\phi,\theta) = \sum_{E:j=1..12} \sum_{\alpha:i=1..4} N^{tot}_{i,j}(\phi,\theta) \times \frac{\epsilon^{week}_{i,j}(\phi,\theta)}{\epsilon^{tot}_{i,j}(\phi,\theta)}$$

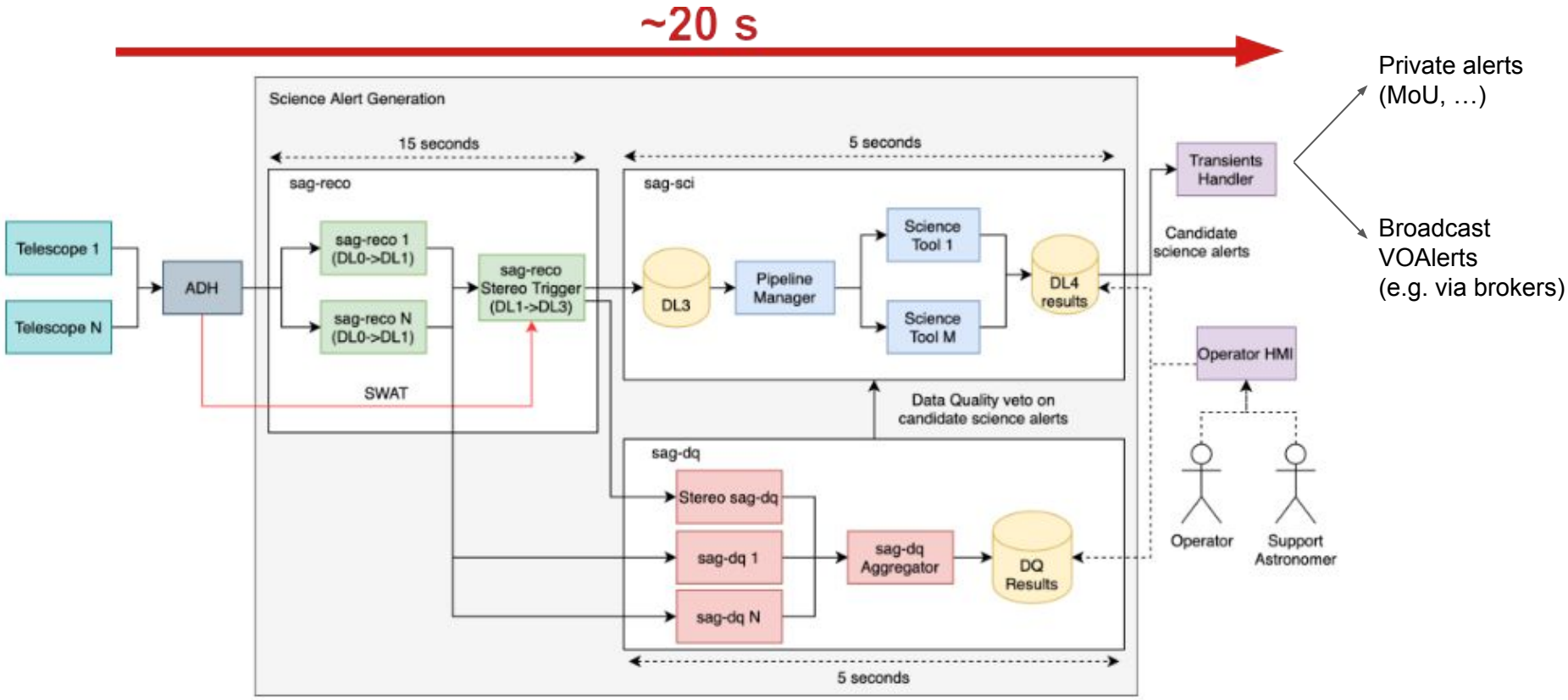
with archival data.
Or DBSCAN in 3D (*x*, *y*; *t*) (e.g. Tramacere & Vecchio, 2012)



Fermi-LAT, *E*>100 MeV, weekly data, FAVA (Ackermann et al., 2013)

Time-domain astronomy & brokers

CTA Science Alert Generation system



Bulgarelli et al. (2022, ICRC2021)

GPS catalog : detection criterion

Detectability criterion:

$$TS_{null} = 2\,\Delta\ln(L) > 25$$

with $\Delta\ln(L)$ the log-likelihood difference between the cases with and without the source

Threshold in Significance rather than TS to account for degree of freedom in different models ?
assuming TS follows a Chi2 distribution :

```
Significance = sqrt(chi2.isf(chi2.sf(TS, DoF), 1)
```

Model selection:
minimal AIC between point-like/shell/generalized gaussian

$$AIC = 2k - 2\ln(L)$$

Small sample limit if few excess counts ? AICc = AIC + 2k(k+1)/(n-k-1) if n/K <40 ?

GPS catalog : association criterion

Matching criterion : detected object *vs* simulated source

(unique association)

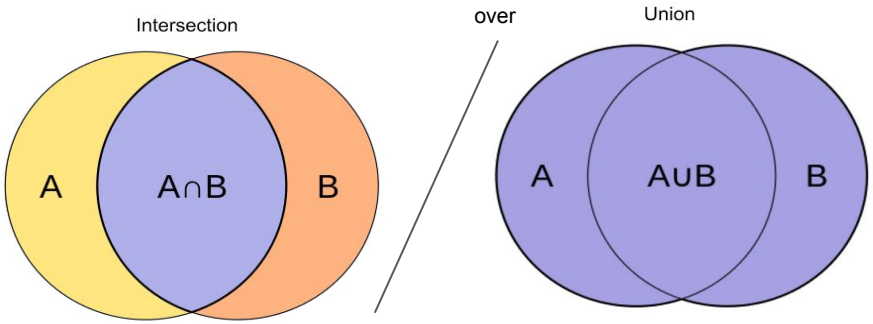
Intercenter distance :

$$d_c < 0.1^{\circ} + 0.3 \times R_{object}$$

and best :

$$SF_{overlap} = \frac{S_{object \cap source}}{S_{object \cup source}}$$

Jaccard coefficient



Test also spectral match not only morphology ?

Account for parameter errors or expected differences with wavelength ?

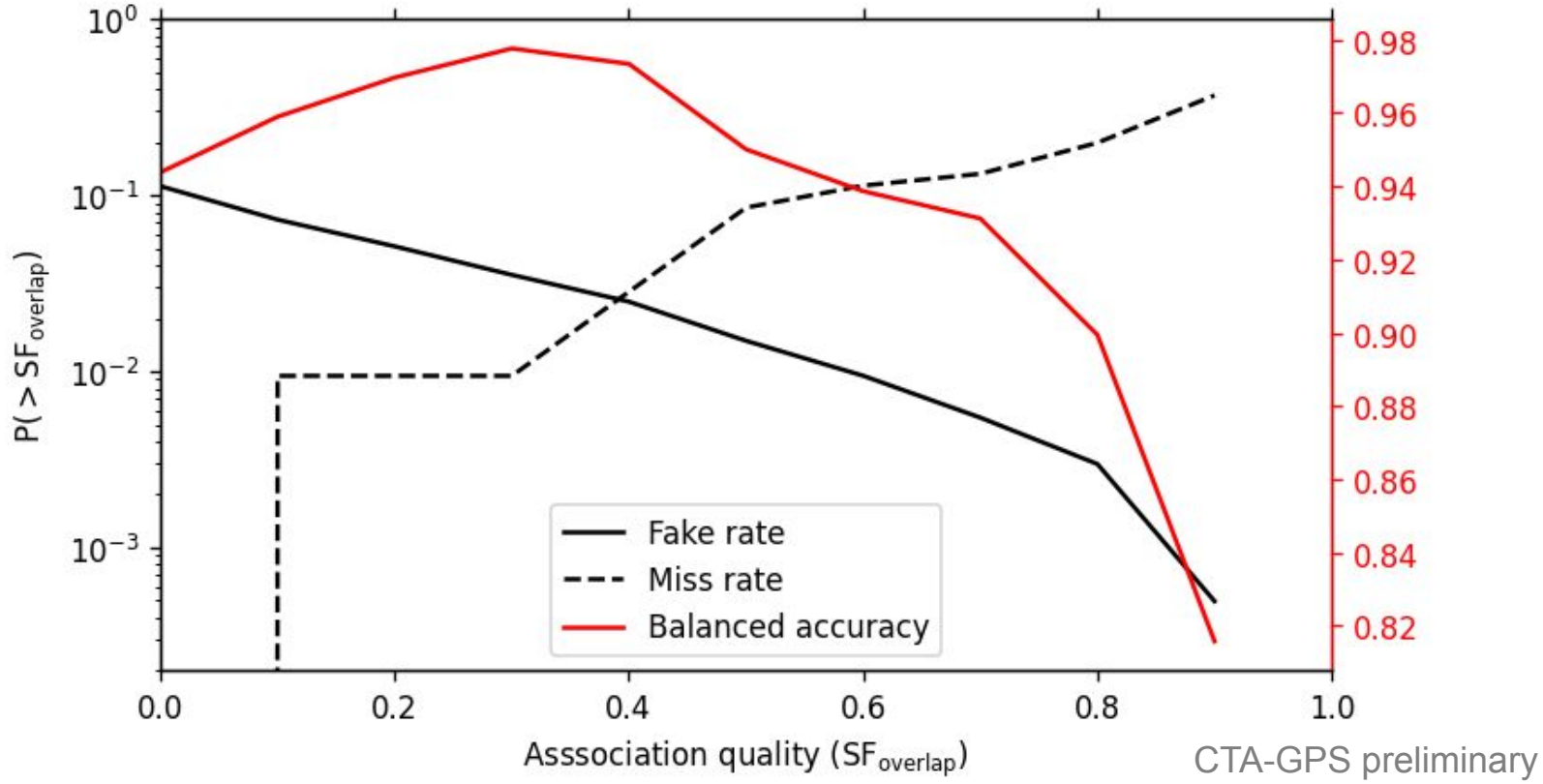
Different criterion for different MWL catalogs or sources populations ?

GPS catalog : associations by chance ?

Catalog bootstrap

Fix known sources and shuffle others several times

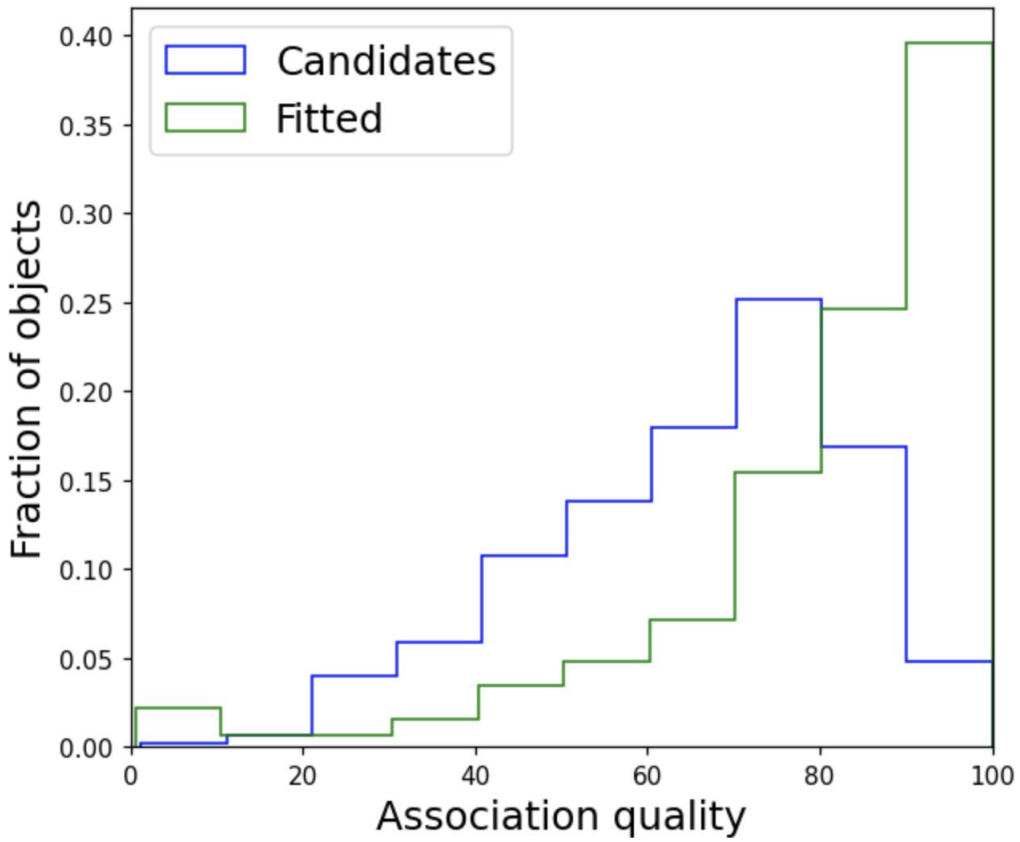
- Fake rate : false positive / true negative
- Miss rate : false negative / true positive
- Balanced Accuracy : $BA = 1 - \frac{P_{fake}+P_{miss}}{2}$



GPS catalog : associations by chance ?

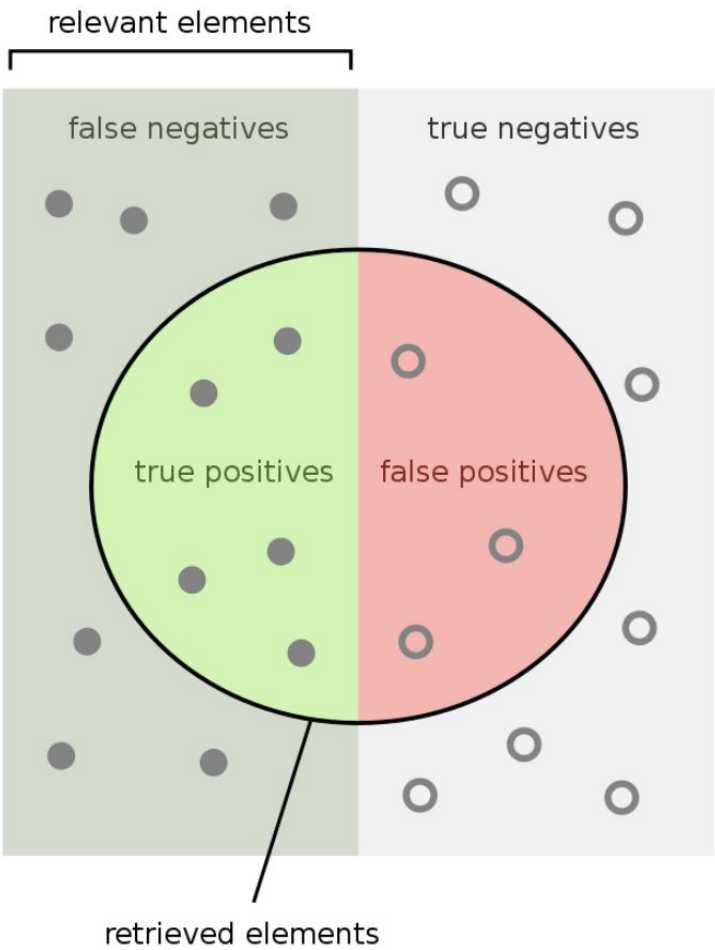
- High association quality so low probability of association match by chance
  => Mostly good association in fitted catalog

- Objects with low TS but high association quality to be included in public catalog ?
  Valuable targets for deeper observations ?



CTA-GPS preliminary

GPS catalog : purity and completeness



Purity

How many retrieved items are relevant?

Precision =

Completeness

How many relevant items are retrieved?

Recall =

Metric to tune catalog algorithm maximizing both ?

Fowlkes-Mallows index:

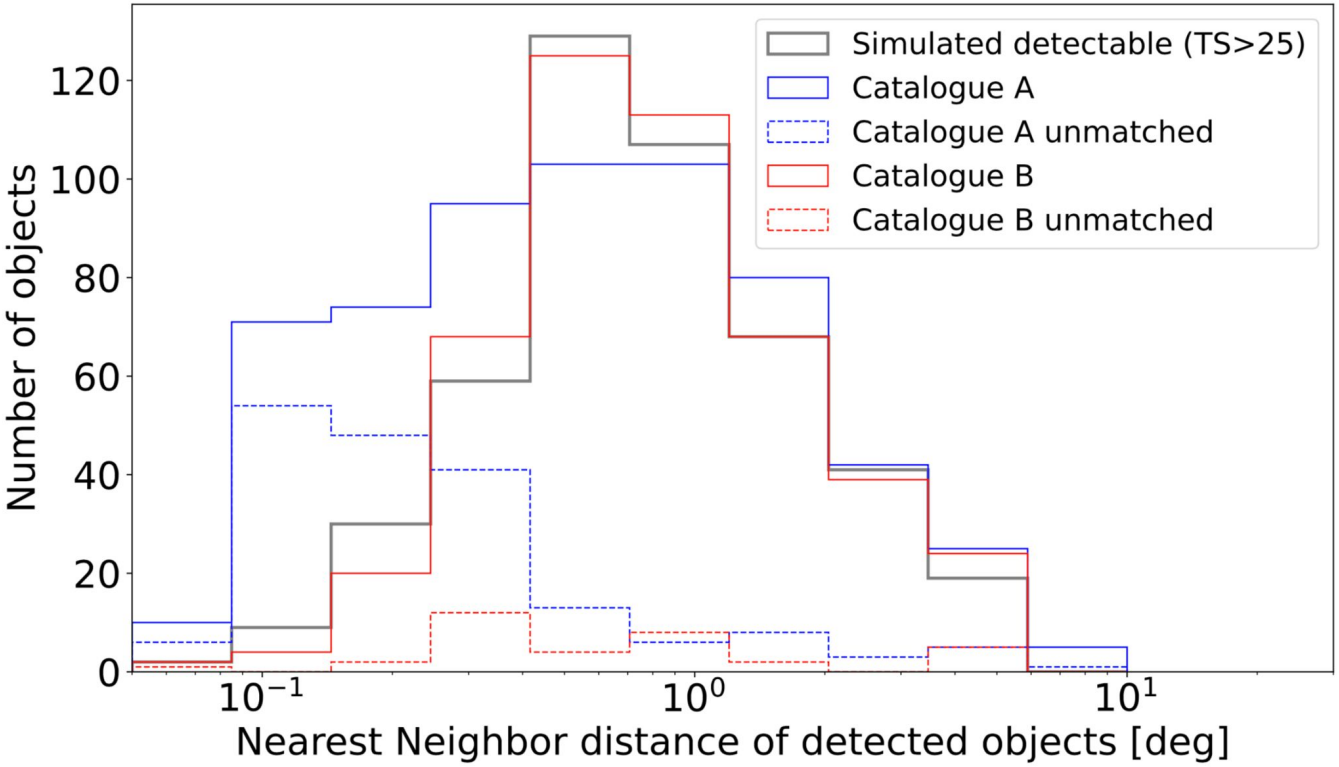
$$FM = \sqrt{\text{precision} \times \text{recall}}$$

GPS catalog : expected detections

CTA-GPS : Detections with TS >25 for E = 0.07-200 TeV :

Name	PWN	SNR	ISNR	BIN	Known	No-match	Total	f_{match} (Purity)	f_{reco} (Completeness)
Simulated detectable	294	37	24	10	134	-	499	-	-
Catalogue A	241	16	20	10	111	169	567	0.70	0.80
Catalogue B	257	31	14	10	122	36	470	0.92	0.87

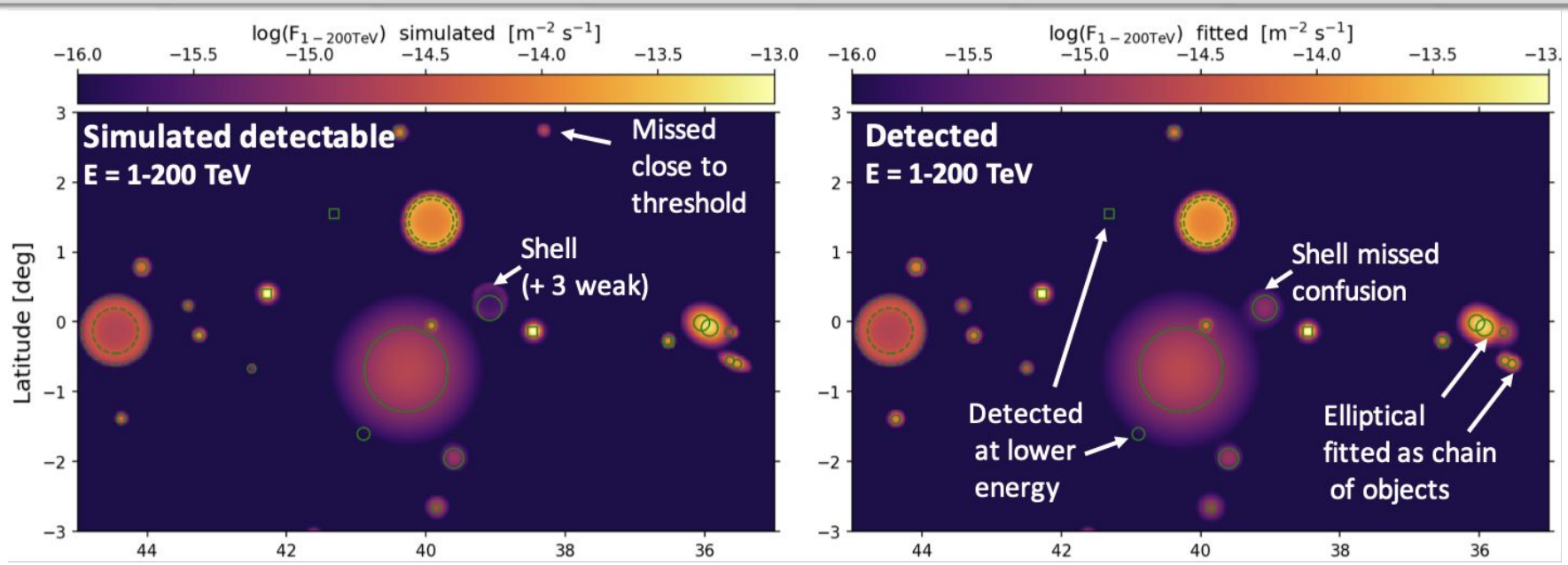
CTA-GPS (in prep.) Remy et al. (ICRC 2021)

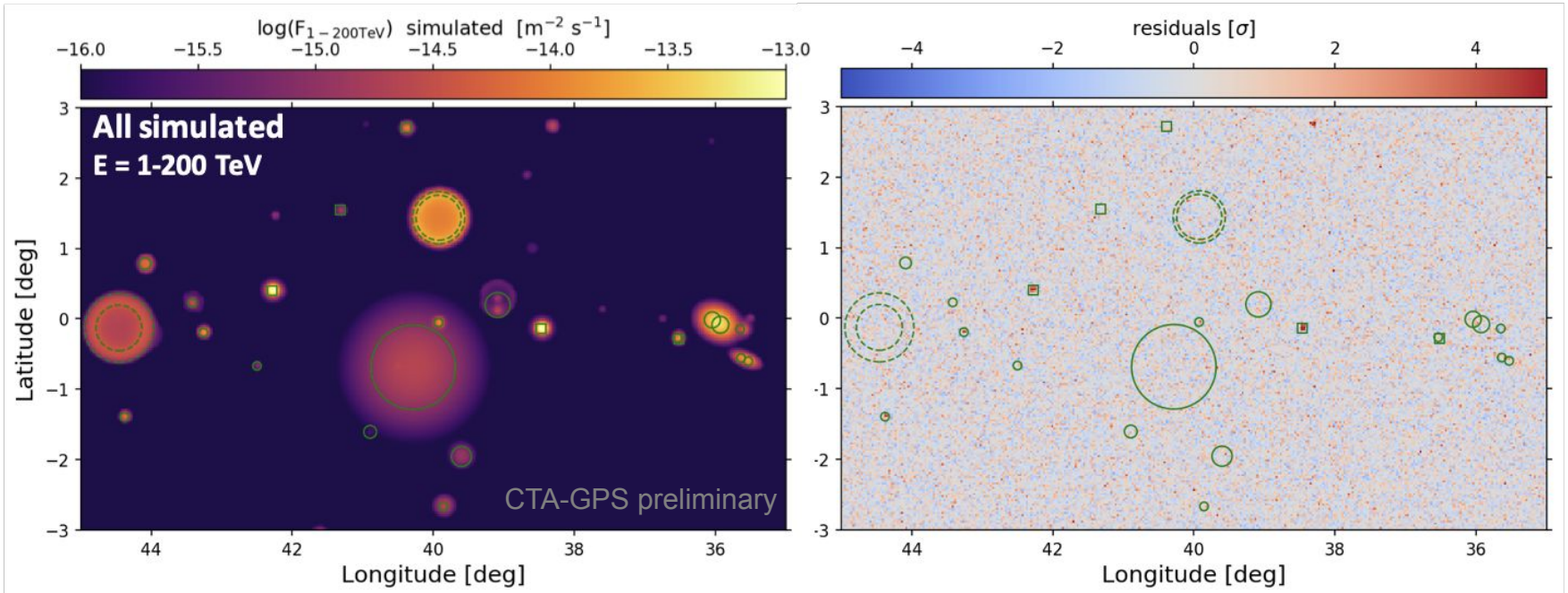


- Most of detectable sources are detected and only few spurious detections

- Spurious detections mostly from source confusion and fragmentation effects

True, confused, and missed detections

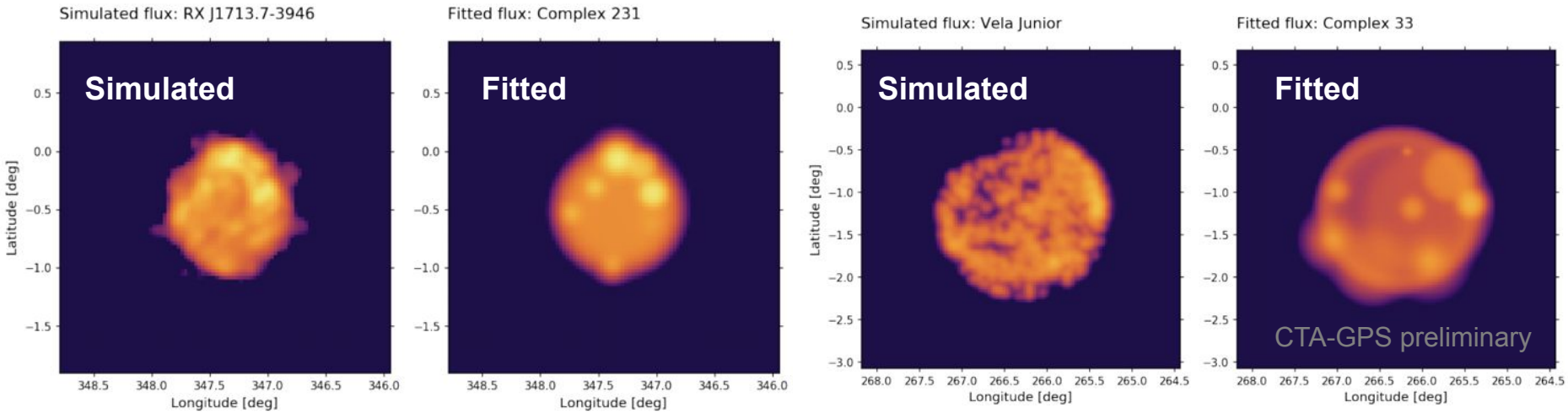




How to model complex sources ?

- Complex sources are fragmented in multiples objects with parametric models

  How to identify complex sources sub-structures as a single entity ?



Groups of nested objects merged into a single component if sub-structures don't have :

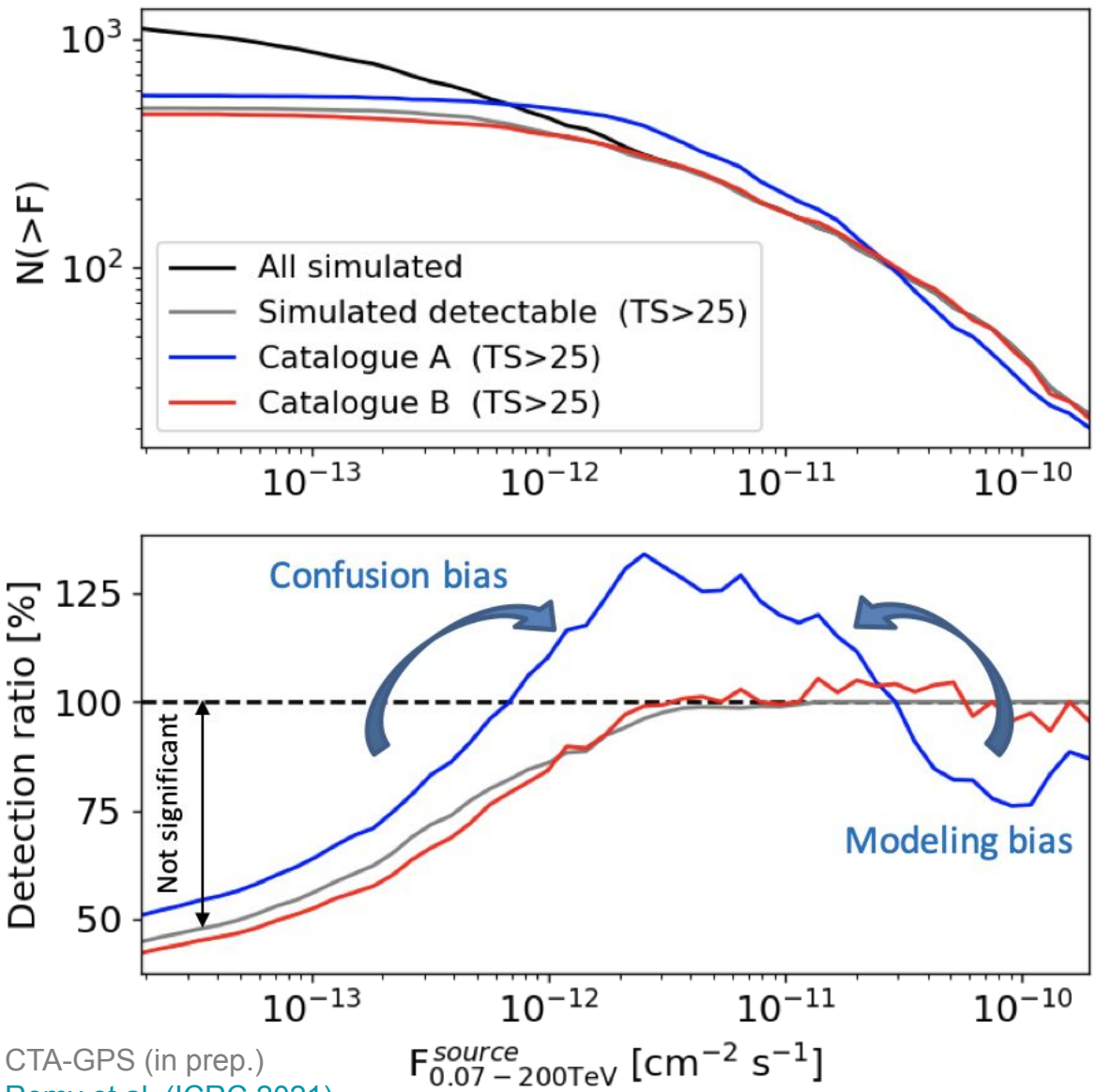
- a different association in known sources
- an obvious deviation in spectral parameters from mean values of the group

- Dedicated analysis often rely on both MWL templates and parametric models

  How to statistically compare fully parametric models to templates ?

  Detection significance, degree of freedom and Information entropy for templates ?

Detection biases



CTA-GPS (in prep.)
Remy et al. (ICRC 2021)

- Confusion bias: enhanced flux and TS of sources near threshold due to sources below threshold (affects both catalogues)

- Modeling bias: complex sources fragmented in multiple detections of lower flux (mainly in Catalogue A)

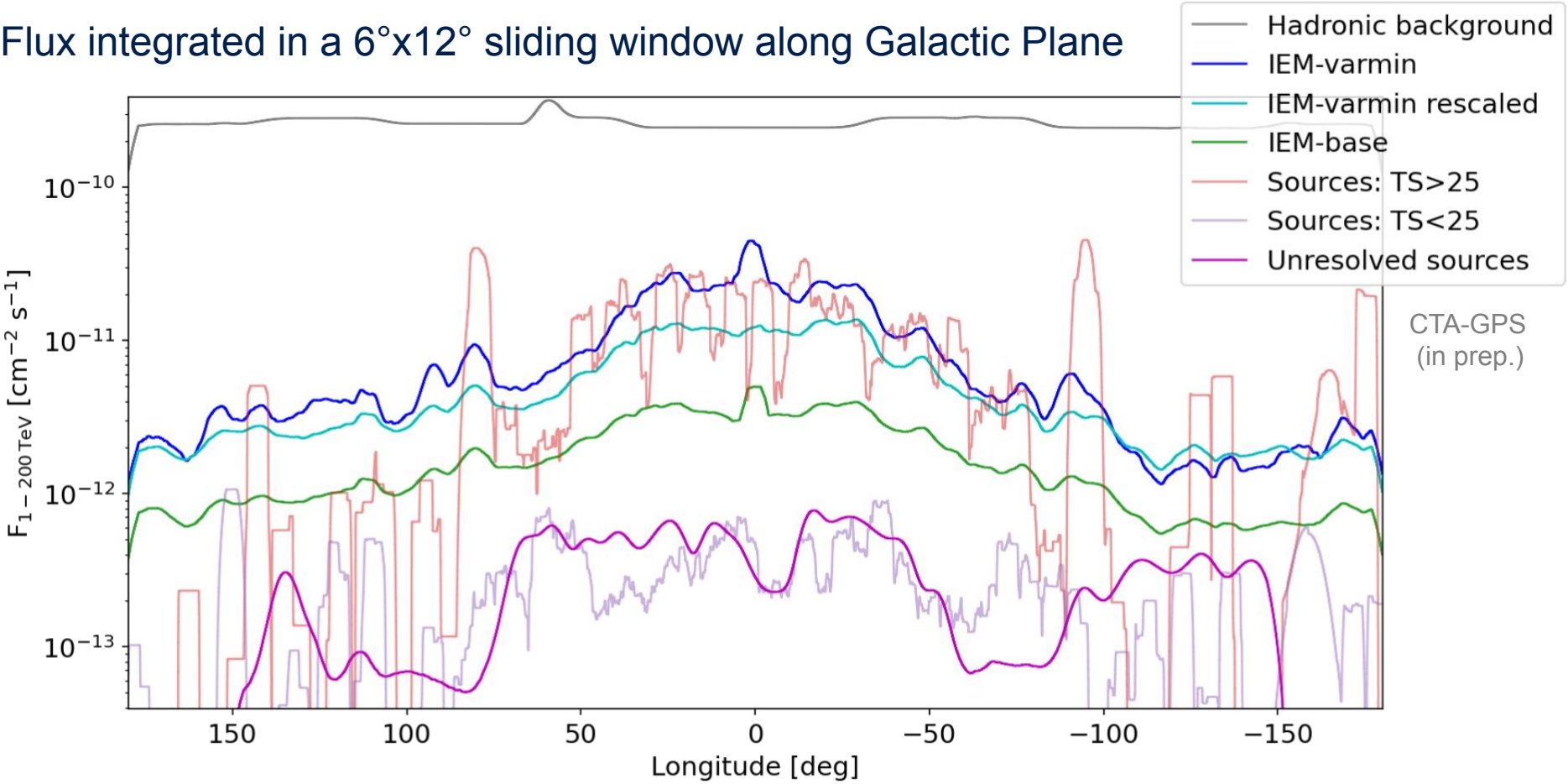
- Catalogue B tests shells, ellipticals and merges templates sub-structures => better estimate of flux and number of objects

- Catalogue B tends to the expected detections

- Exact simulated backgrounds used so systematic effects underestimated

Backgrounds and sources on large scales

Flux integrated in a 6°x12° sliding window along Galactic Plane



- Few percent error on hadronic background can lead to large error on diffuse emission

- Large systematic uncertainties on interstellar emission models (IEM) :
Total gas column density estimates, Cosmic-Ray propagation modeling,
Large scale Inverse Compton contribution, unresolved sources, …

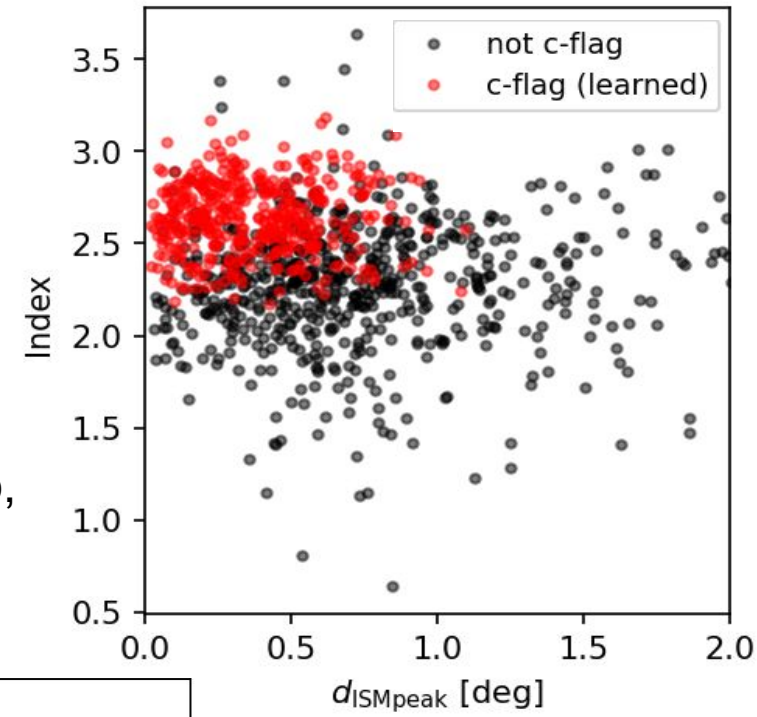
Fake or biased sources due to diffuse emission ?

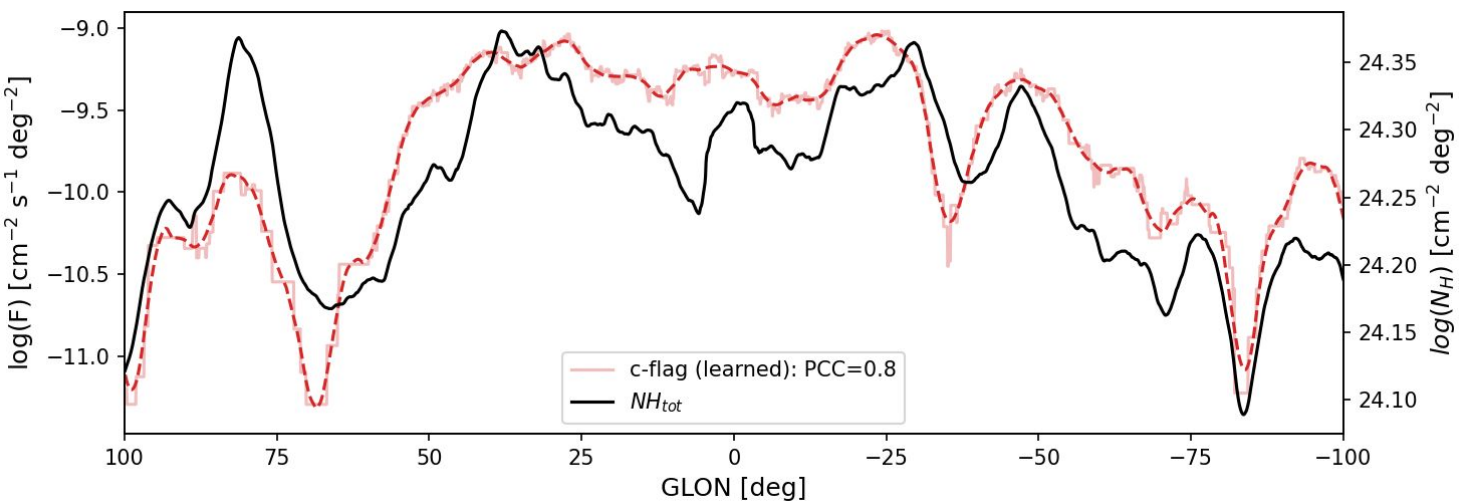
Examples from Fermi-LAT 4FGL catalog :

- flag sources below TS threshold or change in flux with alternative diffuse model
  Check significance and parameters deviation ?
  Anomaly detection to rank sources ?

- "c"-flag : visual screening for diffuse features
  Astronomer's eye selection
  meaning ? reproducibility ? scalability ?

Semi-supervised classifier to reproduce "c"-flag ?
features : spectral parameters, distance to gas clump, differential flux ratio between source and diffuse





Larger correlation with gas for "c"-flag sources

Summary : a few open questions

- Variability

  MWL variability correlation, if any, strongly depends on wavebands and emission mechanisms. How to to optimize ?

- Real-time analysis

  Inexpensive and sensitive way to catch flares/transients. More optimal way ?

- Source confusion and extended sources modelling

  Catalogs objects are significant excess parameterized, not necessarily individual sources

  How to identify complex sources sub-structures as a single entity ?

  Model selection and significance estimates with templates + parametric models ?

- Catalog cross-matches and associations by chance

  Morphological criterion only or also spectral ?

  Account for parameter errors or expected differences with wavelength ?

  Different criterion for different catalogs or populations ?

- Instrumental and astrophysical backgrounds modelling

  Quantify systematic uncertainties on backgrounds ?

  Estimate probability for a source to be fake or biased due to imperfect background ?