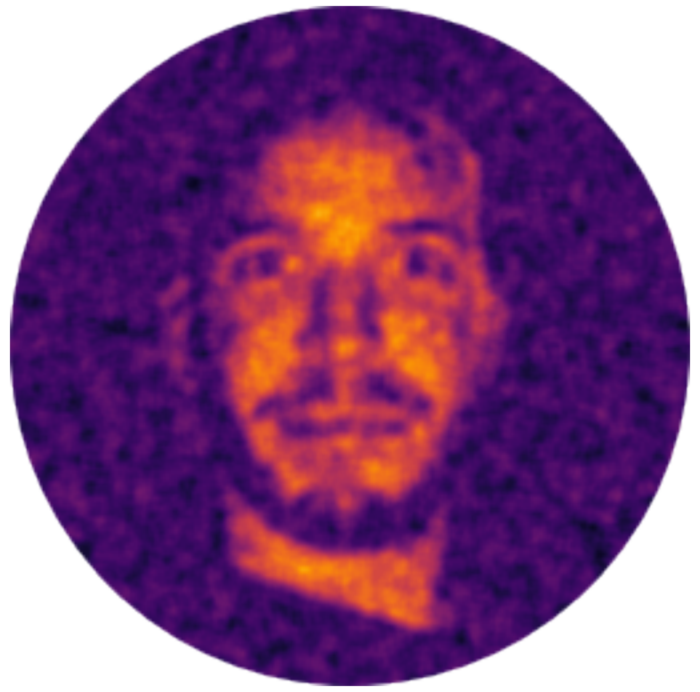
Analysis methods and challenges in TeV gamma-ray astronomy

Topics in Astrostatistics, Seminar, Harvard University
April, 7th 2021

Axel Donath, MPIK Heidelberg

Who am I?



- My name is Axel Donath, I am a **Post-doc at MPIK Heidelberg**
- During my PhD I worked on the **H.E.S.S. Galactic Plane Survey** catalog. Out of this work the Gammapy package evolved. Now I am one of the **lead-developers of Gammapy**
- Interest in the Galactic gamma-rays source population, source detection and catalog production
- General interest in Python based astronomical software: co-organised Python in Gamma-Ray astronomy (e.g. PyGamma 2019), some contributions to the Astropy package
- My Github profile: https://github.com/adonath
- **Thanks a lot for the invitation to present my work to your group!**

Preface

For x-ray astronomers and everyone else…

- Many analysis methods / data model in **gamma-ray astronomy** are **inspired from x-ray astronomy**
- In general the structure of the data is very similar, with **limited counts, Poissonian nature** and requirement to handle the **instrumental response**, such as limited **angular and energy resolution** as well as **non-uniform exposure** and effective detection area
- In detail the structure is certainly different, this probably mostly concerns complexity of spectral and spatial features at high resolution and hadronic **background domination for ground based gamma-ray observatories**
- The **standard statistical analysis method** in gamma-ray astronomy is **binned maximum likelihood fitting** of spatial / spectral models, taking the instrument response into account
- I will not introduce any new "ground breaking" method, but **generalisation of standard methods** and application to larger survey datasets. Also **provide standard methods as tools to a larger community** of science users…

Overview

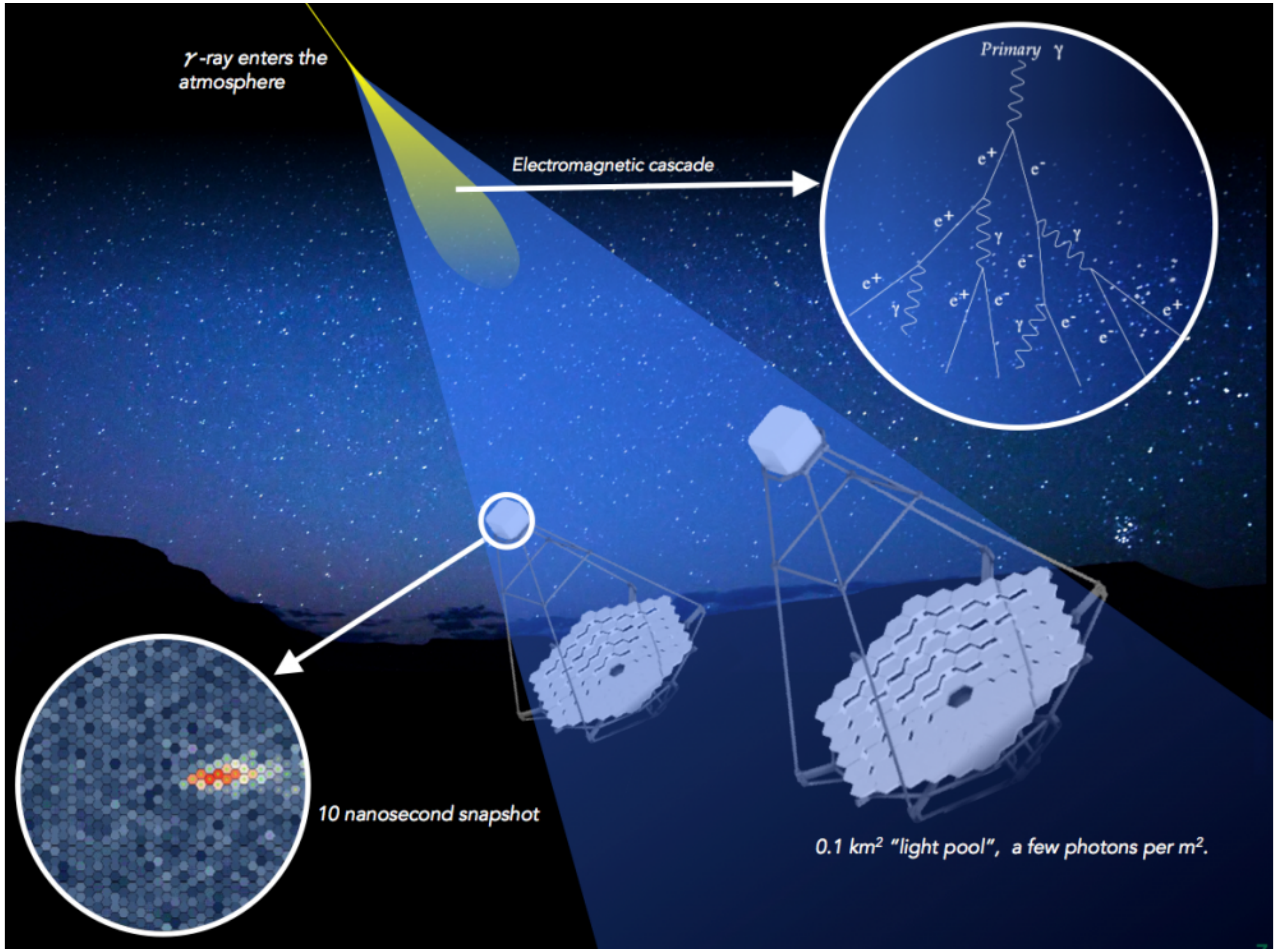
- H.E.S.S. Galactic plane survey & motivation
- Analysis methods
- Gammapy
- Summary / Outlook

H.E.S.S. Galactic plane survey

And general motivation…

Imaging Atmospheric Cherenkov Telescopes

Working principle in a nutshell…

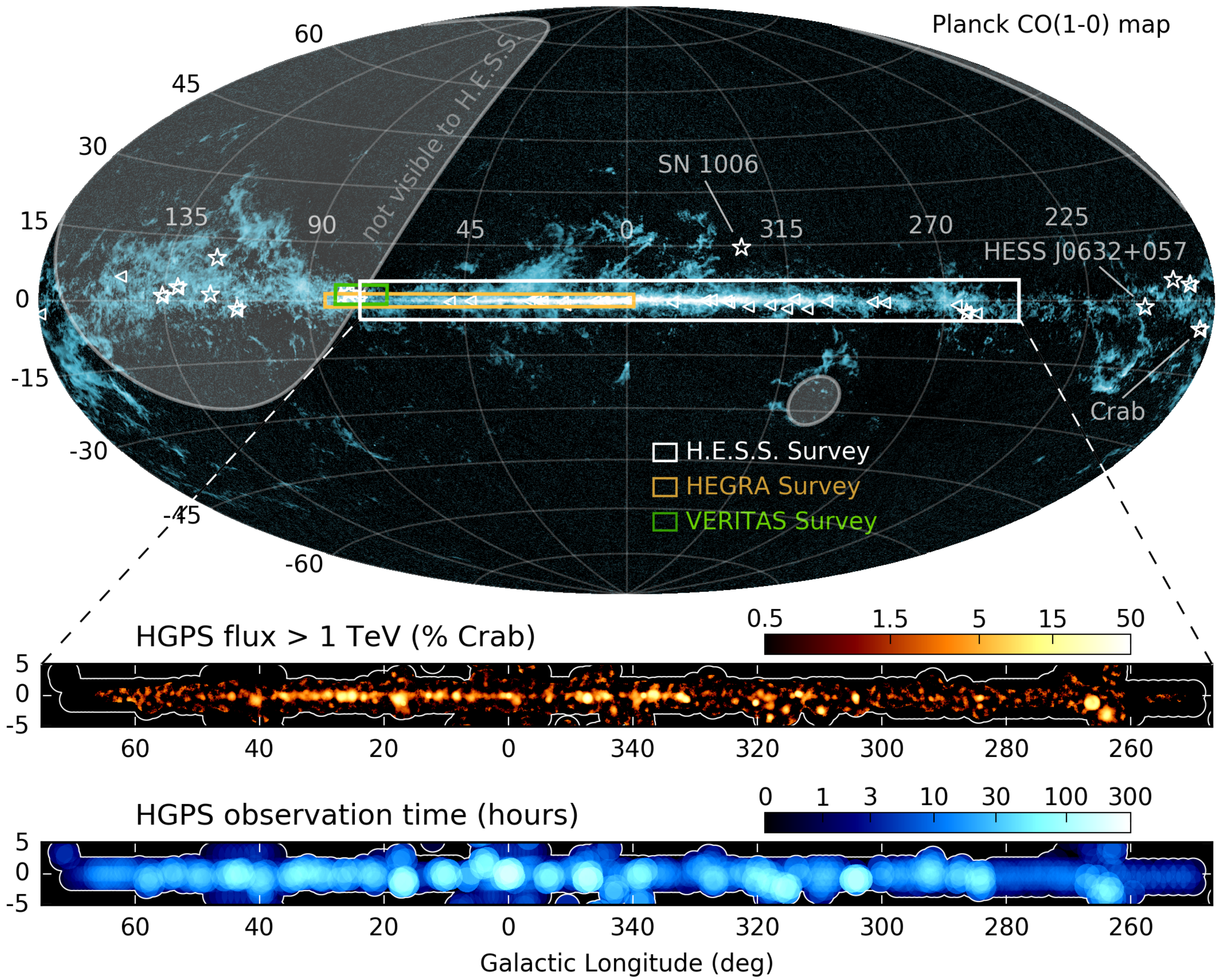


"How CTA Detects Cherenkov Light", taken from www.cta-observatory.org

- **Gamma-ray enters the atmosphere and triggers a "particle shower"**: an electromagnetic cascade of secondary particles
- Secondary particle move faster than the local speed of light and **produce flashes of blue Cherenkov light**
- The nanosecond flashes are **observed by optical telescopes from the ground** with photomultiplier cameras, in the light pool
- If done with multiple telescopes, the **energy and arrival direction** of the incident gamma-ray can be **reconstructed**
- Instrument is pointed for a given period of time to a position one the sky and measures events. This is called an **"observation" or "run"**.

The H.E.S.S. Galactic plane survey

Overview and survey region

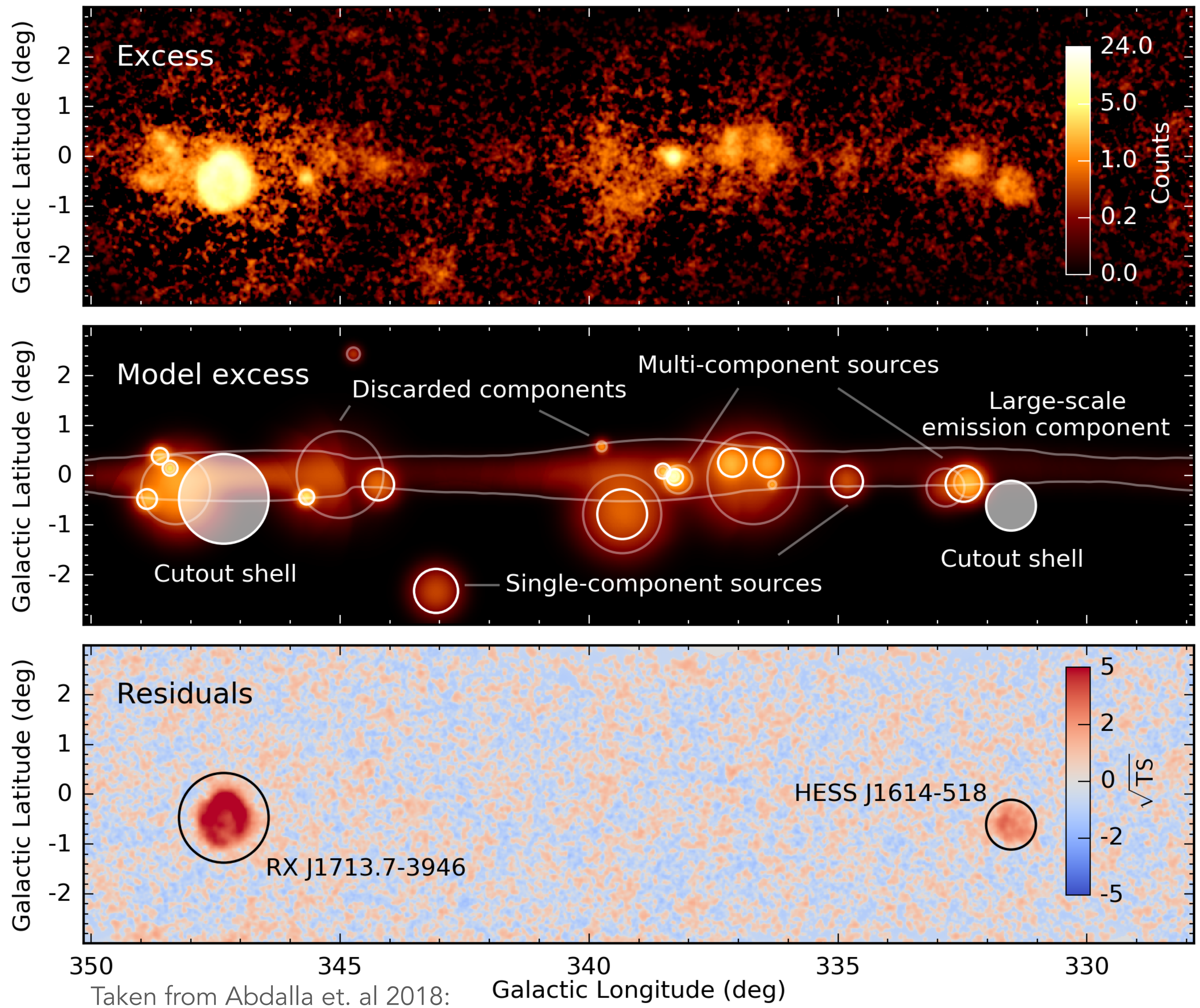


Taken from Abdalla et. al 2018: https://www.aanda.org/articles/aa/full_html/2018/04/aa32098-17/aa32098-17.html

- The **High Energy Stereoscopic System (H.E.S.S.)** located in Namibia is an array of 5 Imaging Atmospheric Cherenkov Telescopes (IACTs)
- "H.E.S.S. Galactic plane survey" (HGPS): observation program conducted from 2003 to 2013, includes 6000 pointed observations amounting to **~3000 hours of observational data**
- Covers an **energy range from 0.2 TeV –100 TeV**
- Covers only a small part of the sky, but the includes most of the known Fermi-LAT high energy sources
- Already obvious: the Milky Way is bright in TeV gamma-rays! Goal to create a catalog form the data…

Source catalog

A multi-gauss representation of the gamma-ray excess

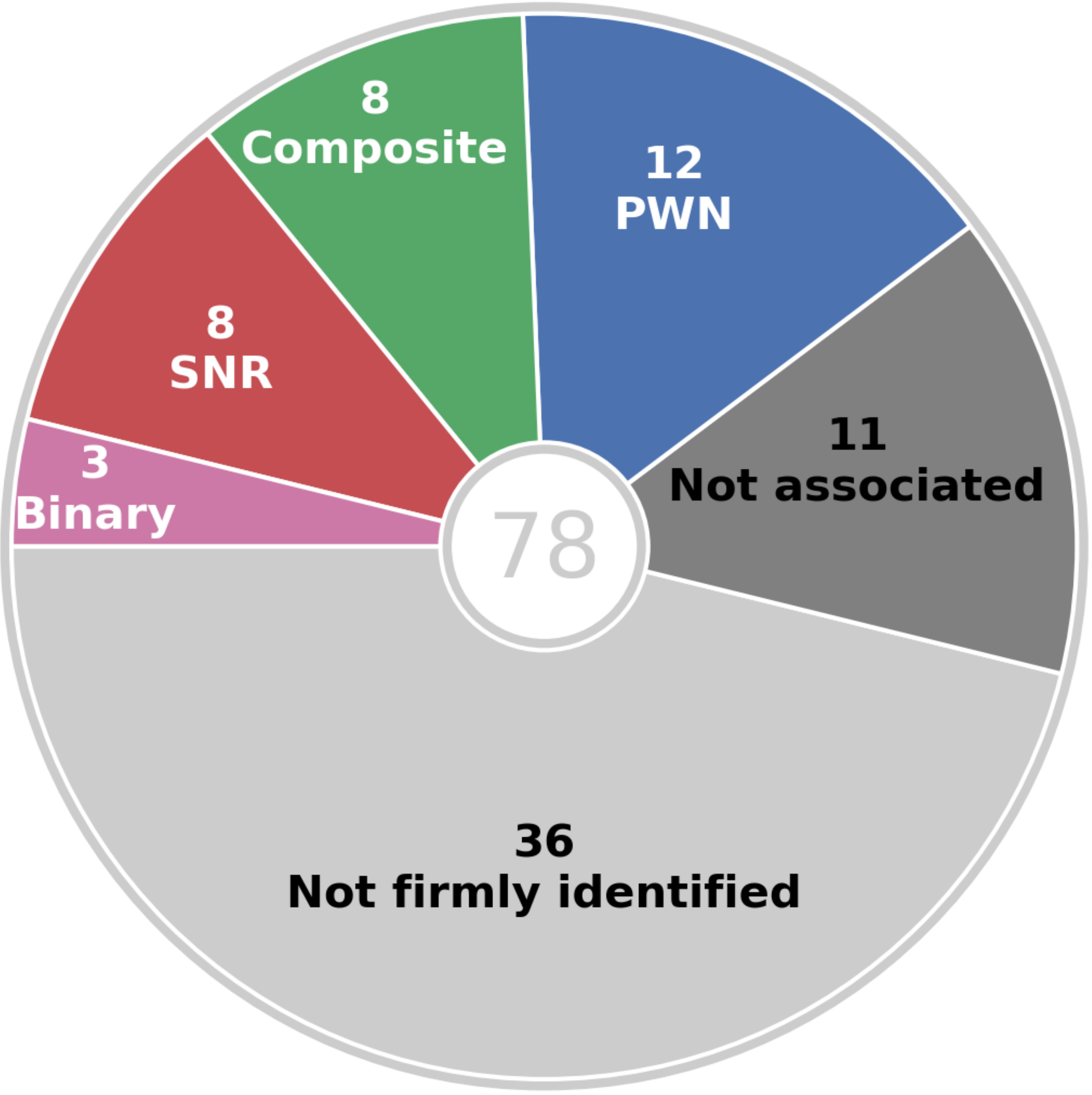


Taken from Abdalla et. al 2018:
https://www.aanda.org/articles/aa/full_html/2018/04/aa32098-17/aa32098-17.html

- Analysis based on **1d and 2d binned Poisson maximum likelihood fitting**
- Model image excess as **superposition of Gaussian components**
- A new Gaussian component was kept in the model, if it improved the global Cash statistics value by TS = 30
- Extended sources can **"decompose" into multiple Gaussian components** and are "merged" by visual inspection
- Exclude complex **shell-like supernova remnants (SNR)** morphologies, **parametric modelling difficult / almost impossible…**
- Some very extended Gaussian components, which could not be attributed to any know astrophysical sources
- Add an **empirical "diffuse band"** derived from the data to account for **unresolved source / true interstellar diffuse emission**
- **Measure spectra for sources** in regions (aperture photometry) defined by the extend / morphology of the source, choose between power-law and exponential-cutoff power-law parametrisation
- Modelling pipeline based on the sherpa package.
- General analysis methods covered in detail later…

Some HGPS results

Association and identification of sources



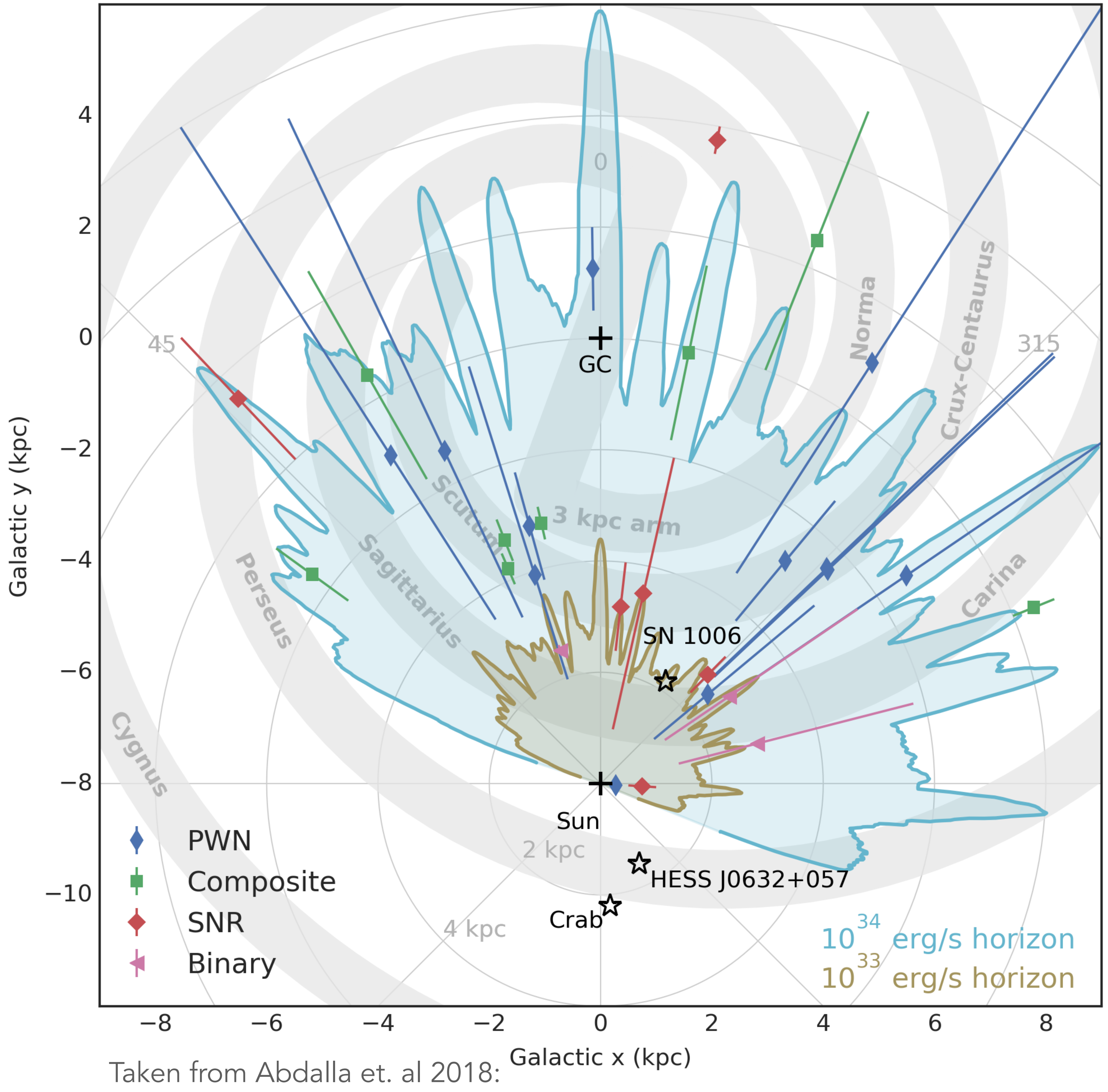
Taken from Abdalla et. al 2018:
https://www.aanda.org/articles/aa/full_html/2018/04/aa32098-17/aa32098-17.html

- Catalog includes in total 78 sources , of which 64 where re-analysed (14 source where excluded because of too complex morphology…)
- Association with objects from other wave-length, based on position: **SNRCat, Green's catalog of SNRs** and energetic pulsars from the **ATNF catalog**
- **31 "identified" objects:** are uniquely associated to a known object
- **11 "dark sources":** are not associated to any object
- **36 not firmly identified sources:** are associated to multiple objects, no firm identification possible
- In total 15 "new", previously unpublished sources

Some more HGPS results

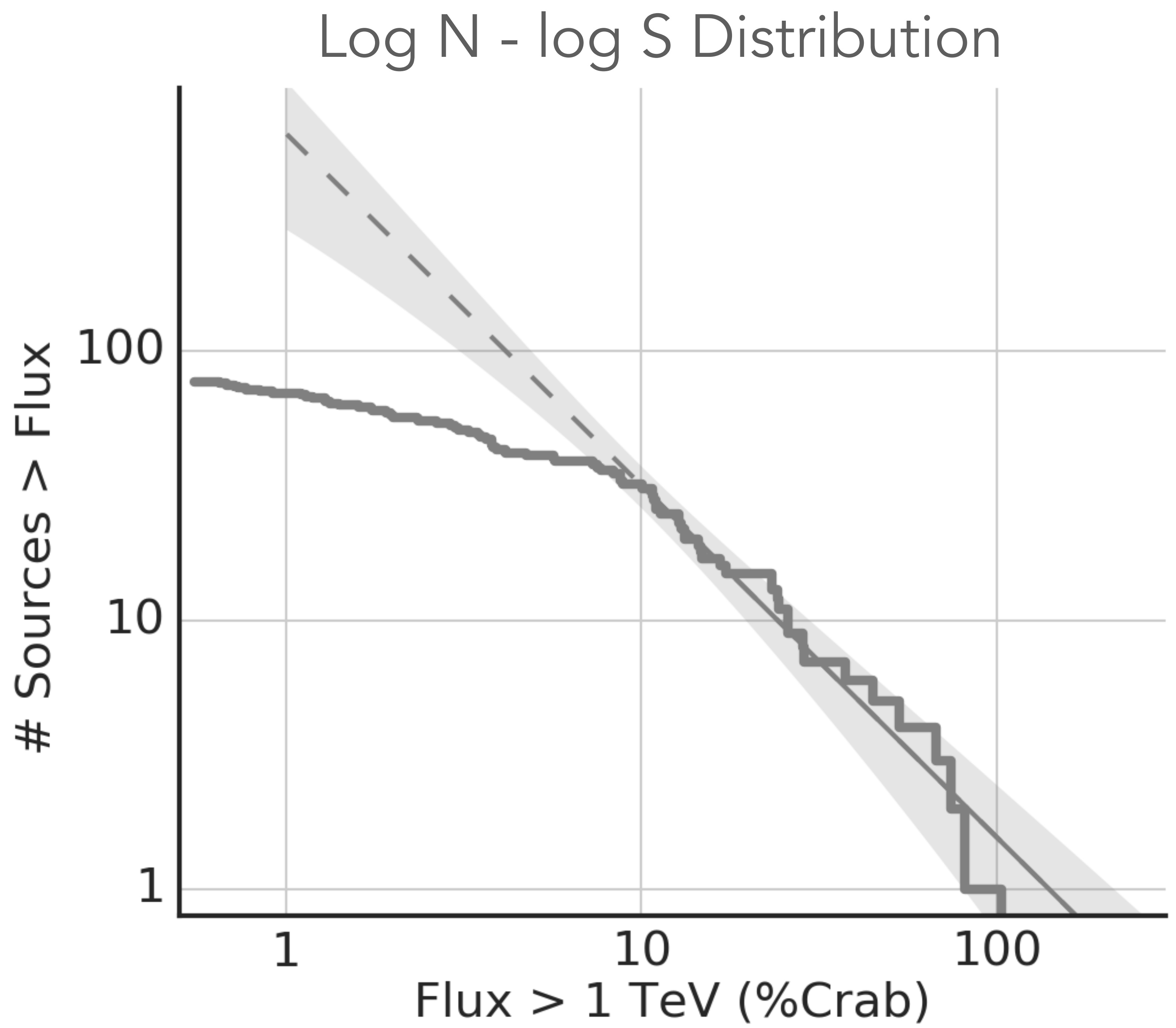
Source distribution in the Milky-Way & “completeness”

Milky way "face on” and survey “horizon”



Taken from Abdalla et. al 2018: https://www.aanda.org/articles/aa/full_html/2018/04/aa32098-17/aa32098-17.html

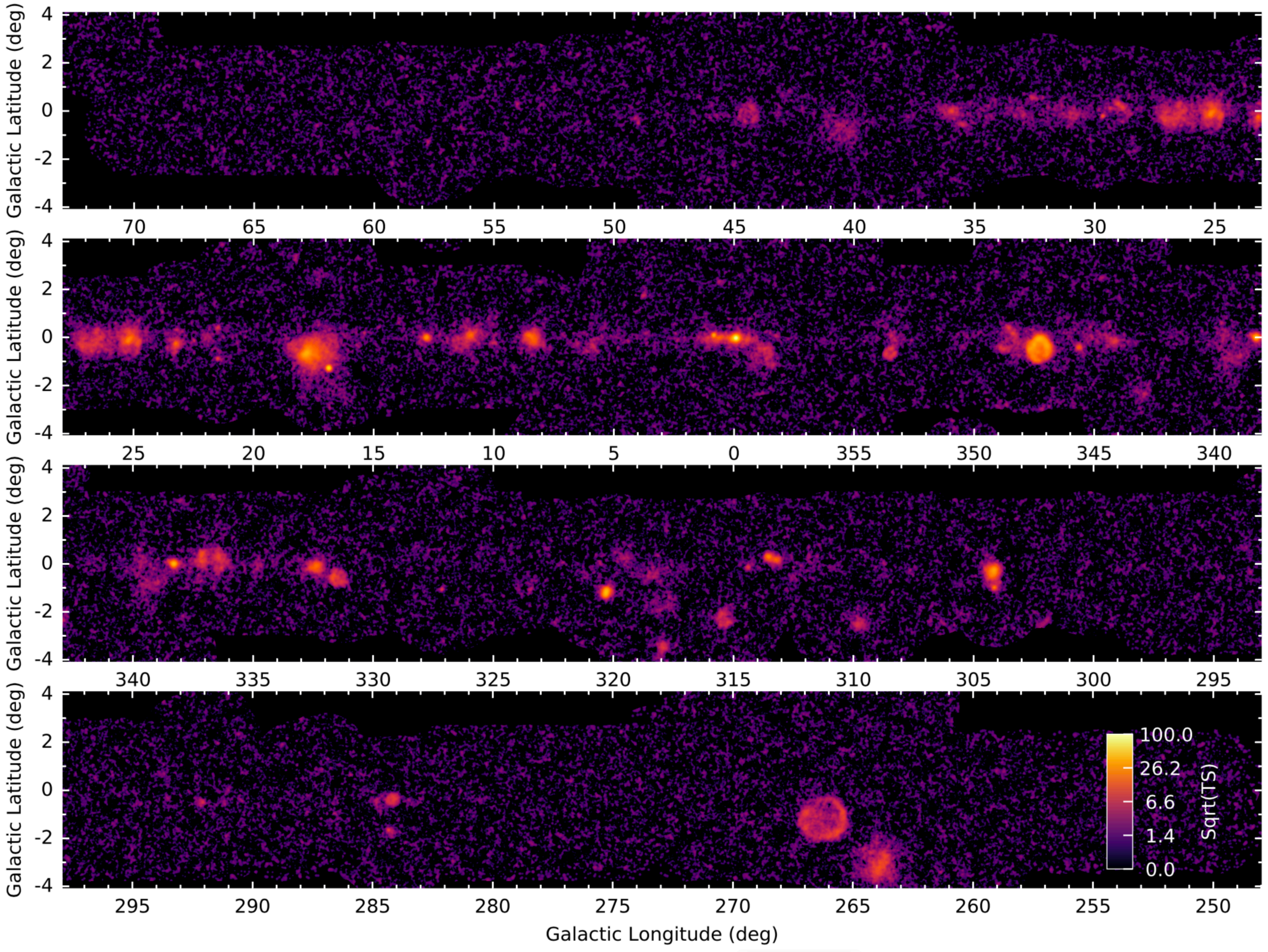
- Identified sources with distances cover only a fraction of the Galaxy
- Horizon of the HGPS survey: the HGPS survey is far from being complete, for weak sources (10^{33} erg /s) it only **covered the solar neighbourhood**
- The extrapolation of the Log N- Log S diagram shows an **expected ~400-600 sources with a flux > 1% of the Crab Nebula** (standard candle)
- In general much more source are expected with a more sensitive instrument: the upcoming **Cherenkov Telescope Array (CTA)**

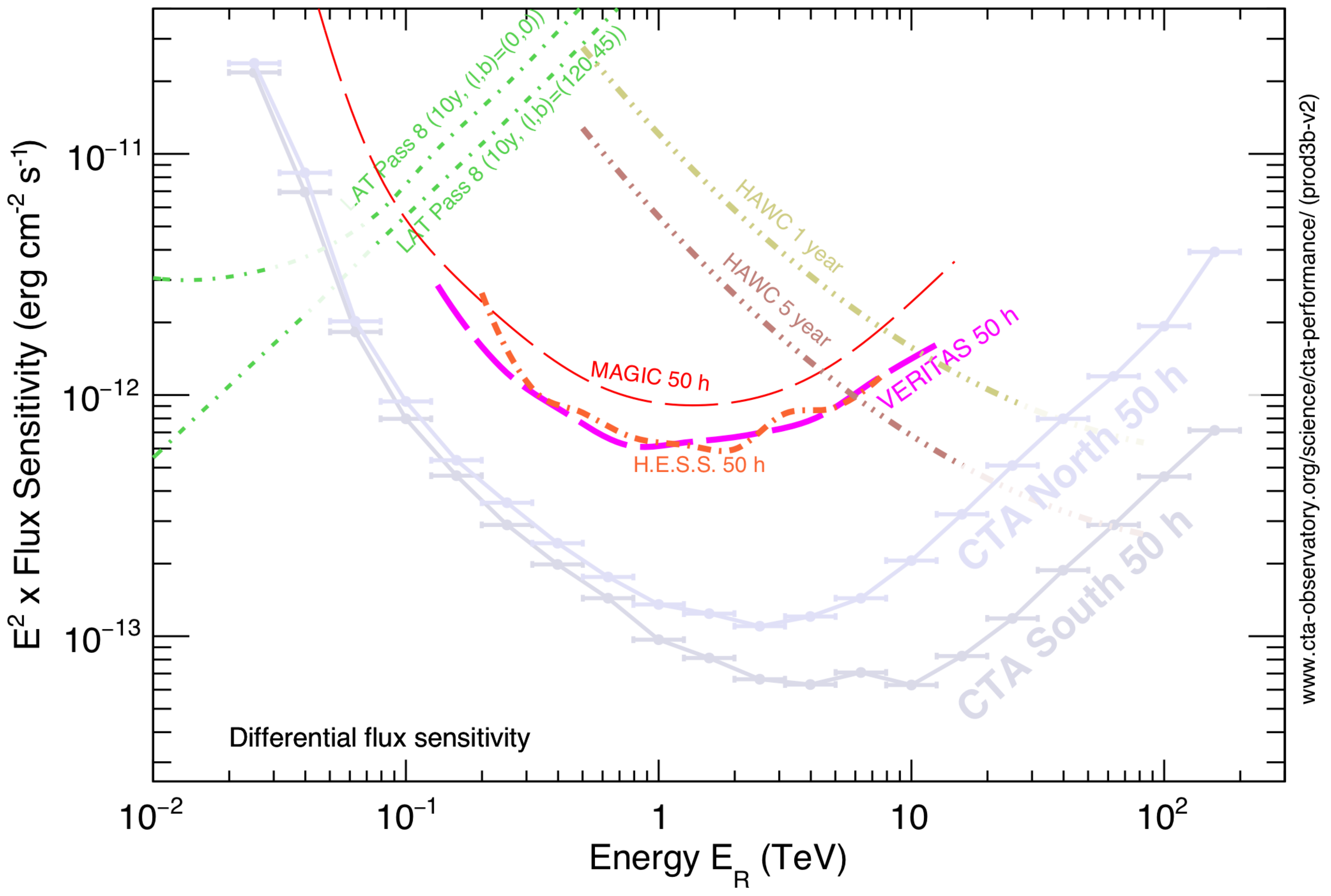


Visual comparison: H.E.S.S. Galactic plane survey

Current view on the Galactic plane

Observation time, ~3000 hours

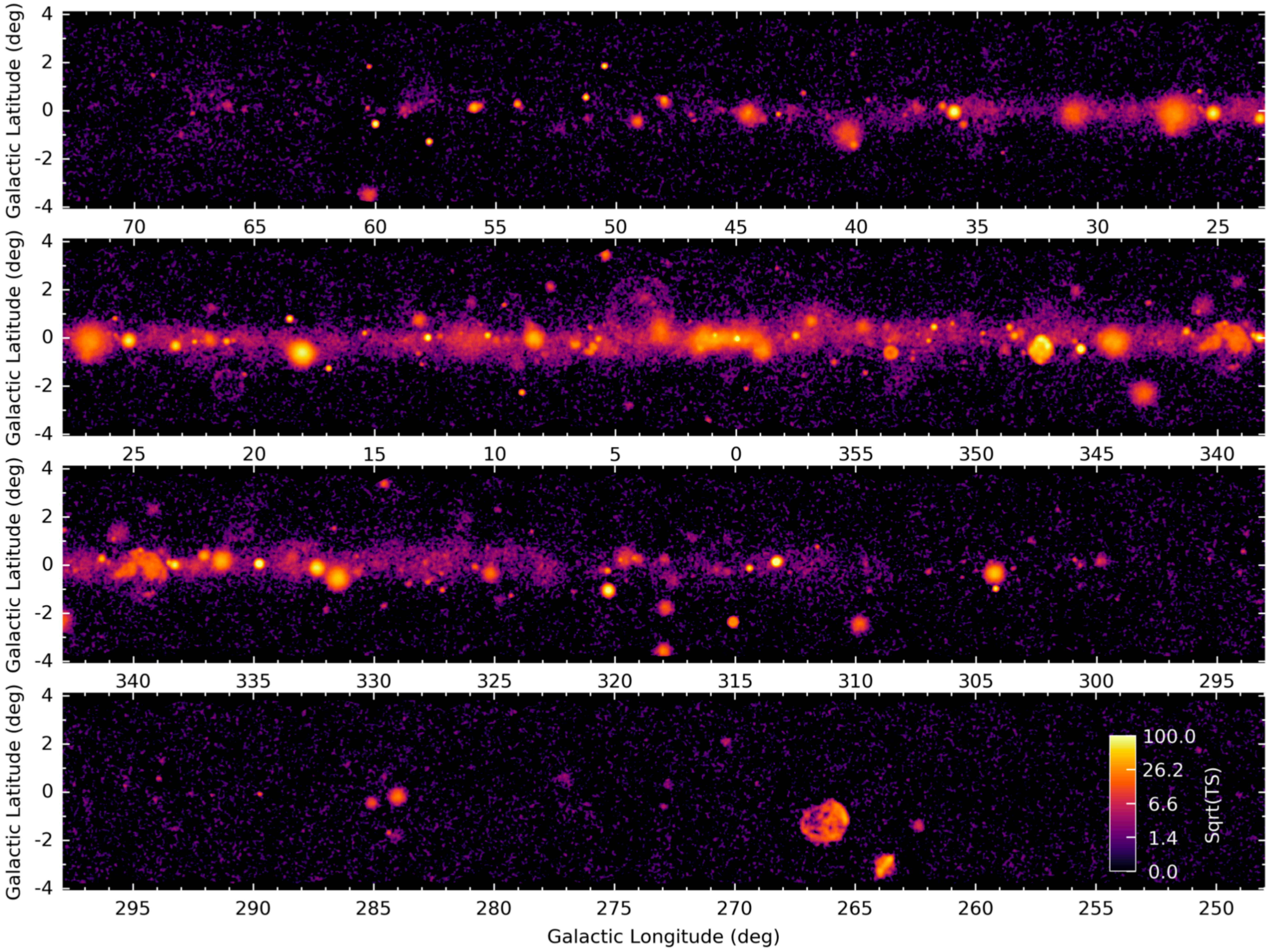


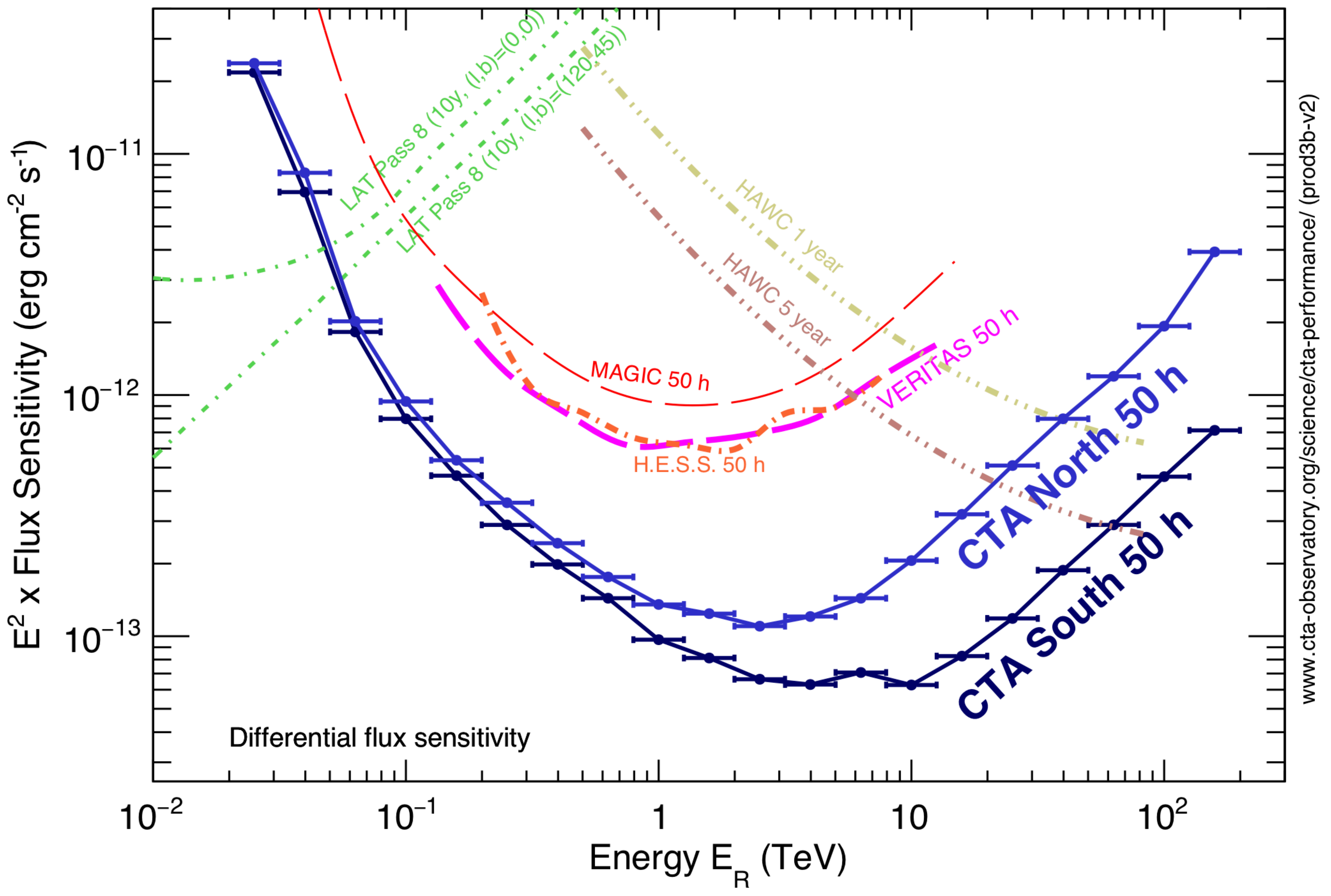


Visual comparison: CTA Galactic plane survey

Expected view on the Galactic plane

Observation time, ~1500 hours





Analysis challenges

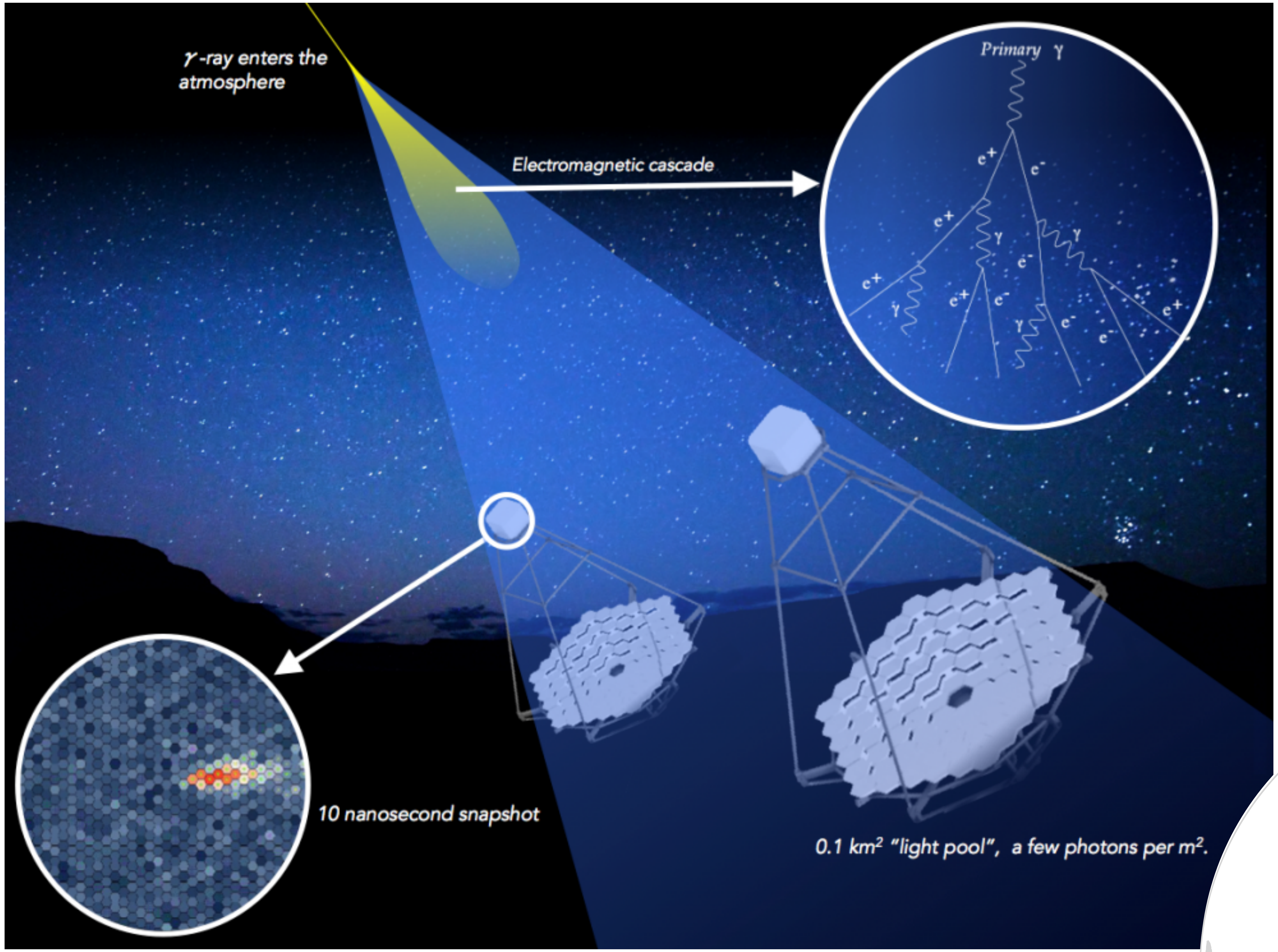
- With the limited angular resolution already with H.E.S.S. a lot of **nearby and "confused" sources**
- **Extended complex morphologies of sources** on the other hand
- Distinct sources **"confused" with true Galactic diffuse emission** and unresolved population of Gamma-ray sources
- **Instrument response varies strongly with energy and offset,** requirement to correct e.g. measured source fluxes for the "containment" of the point spread function and for the flux of extended nearby sources "leaking" into the analysis region
- Need to include **multi-wavelength information** for meaningful association and combining "Gaussian components"
- In general requirement for **more advanced analysis methods and tools** to fully exploit existing and future gamma-ray data
- This motivated us to work on a **new analysis methods and software** for gamma-ray astronomy

Analysis methods

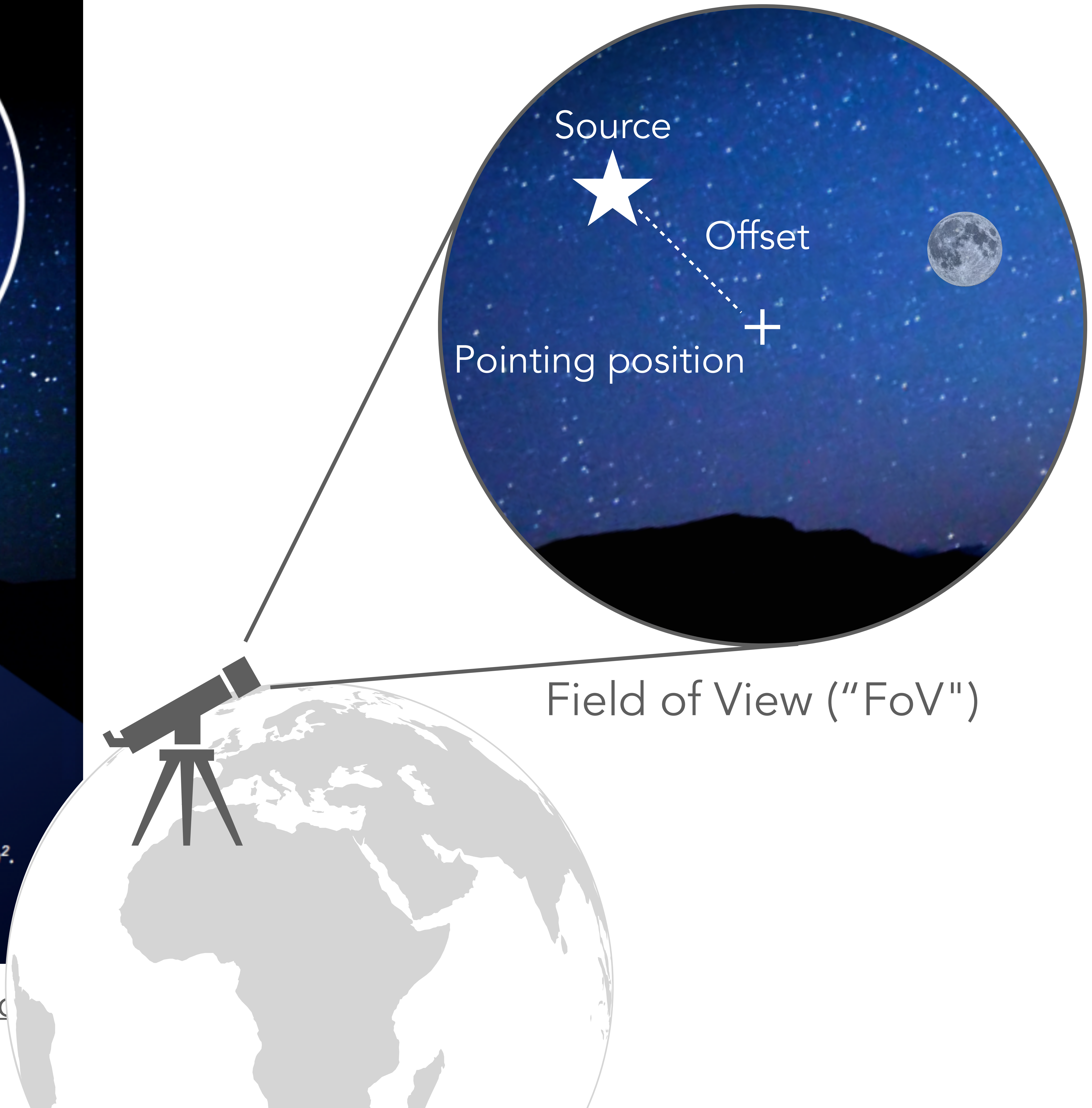
Binned Poisson max. likelihood and joint-likelihood fitting

Imaging Atmospheric Cherenkov Telescopes

Coordinate system & definitions



"How CTA Detects Cherenkov Light", taken from www.cta-observatory.o

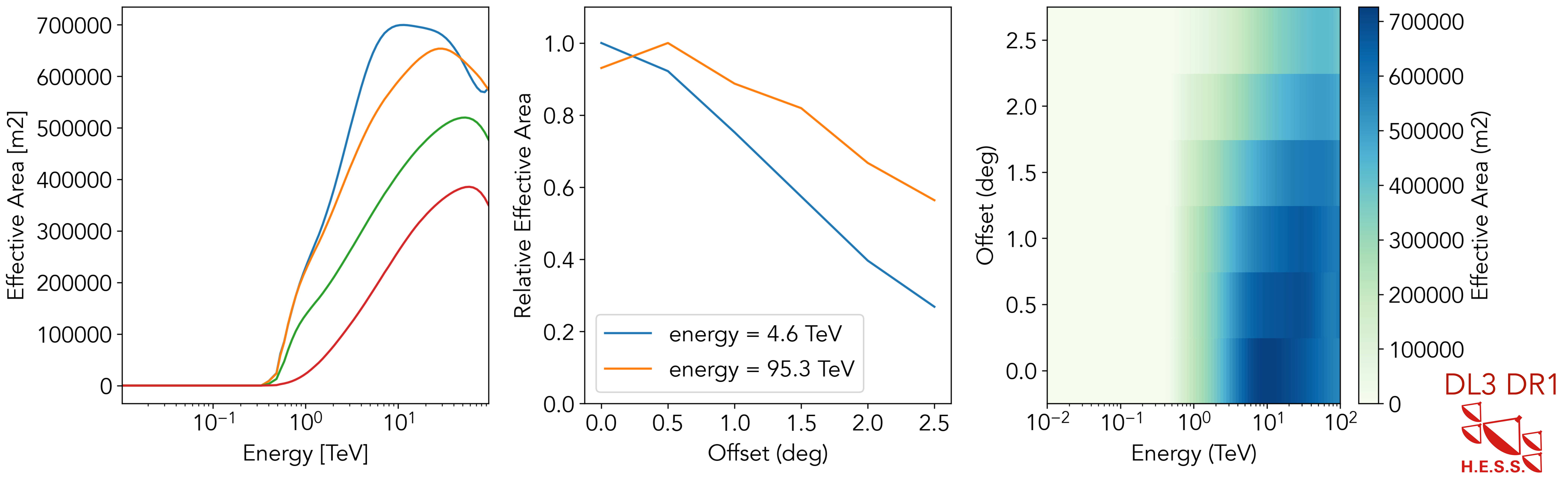


Instrument response

Effective detection area

DL3 → DL4 → DL5

γ-like events | Binned data | Science products

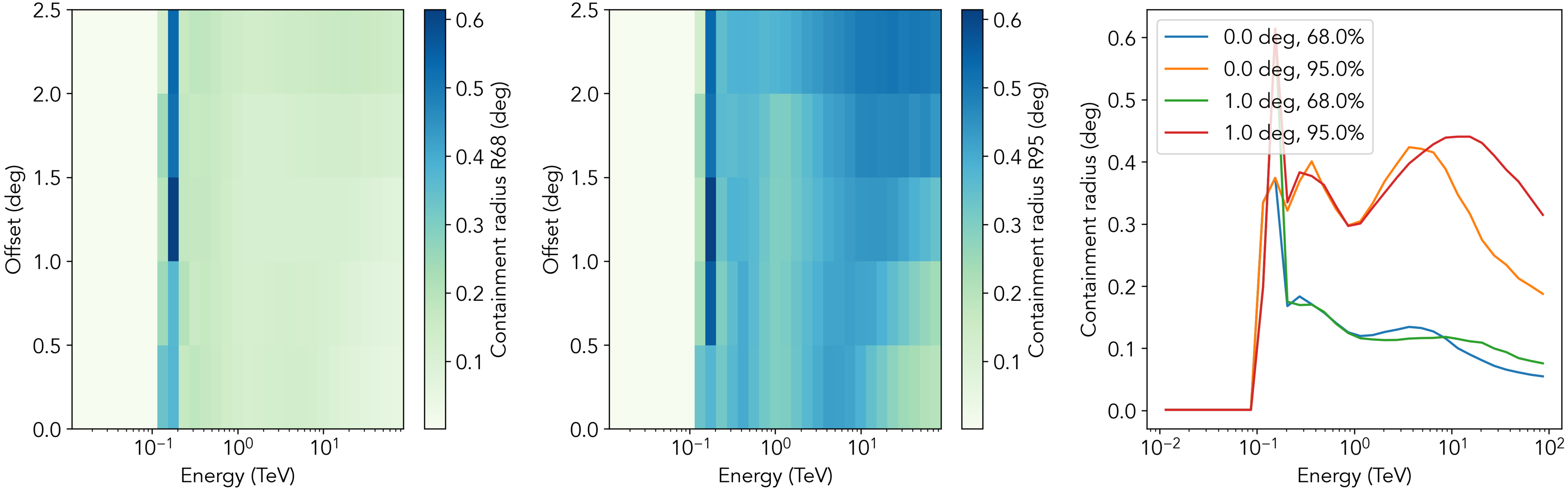


- Effective area: gives number of **expected (true) gamma-like events per area**
- **Varies with offset** from pointing position and true energy
- Estimated from "Monte Carlo" simulations
- Required to measure flux / brightness of gamma-rays sources

Instrument response

Point spread function

DL3 γ-like events → DL4 Binned data → DL5 Science products



DL3 DR1

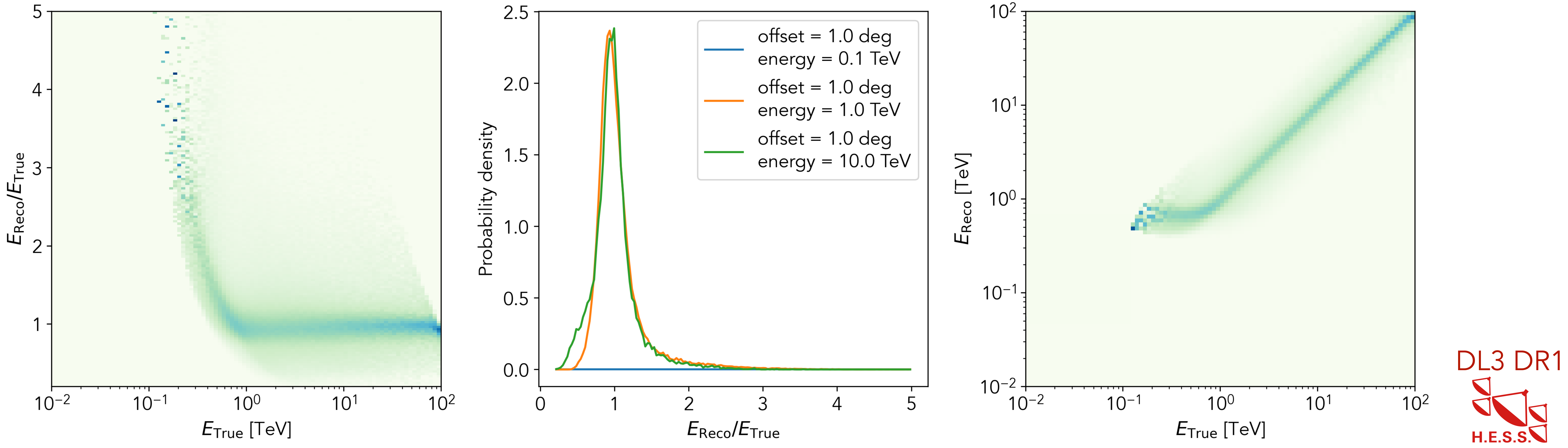
H.E.S.S.

- **Point spread function** (PSF): angular resolution of the instrument, **precision to reconstruct the arrival direction of an event**
- Estimated from "Monte Carlo" simulations and by binning events into offset and true energy
- Typically **stored as "radial profile"** and varies with offset from pointing position and true energy
- There are parametric models of the shape as well e.g. triple Gaussian or King profile, obtained by fitting the shape to the event distributions
- Required to measure extension of Galactics sources & precise flux of point sources

Instrument response

Energy resolution

DL3 γ-like events → DL4 Binned data → DL5 Science products



- **Energy dispersion:** accuracy and precision to reconstruct the energy of an event
- Varies with offset from pointing position and true energy
- Required to measure precise spectra of source, especially at low energies (for IACTS)

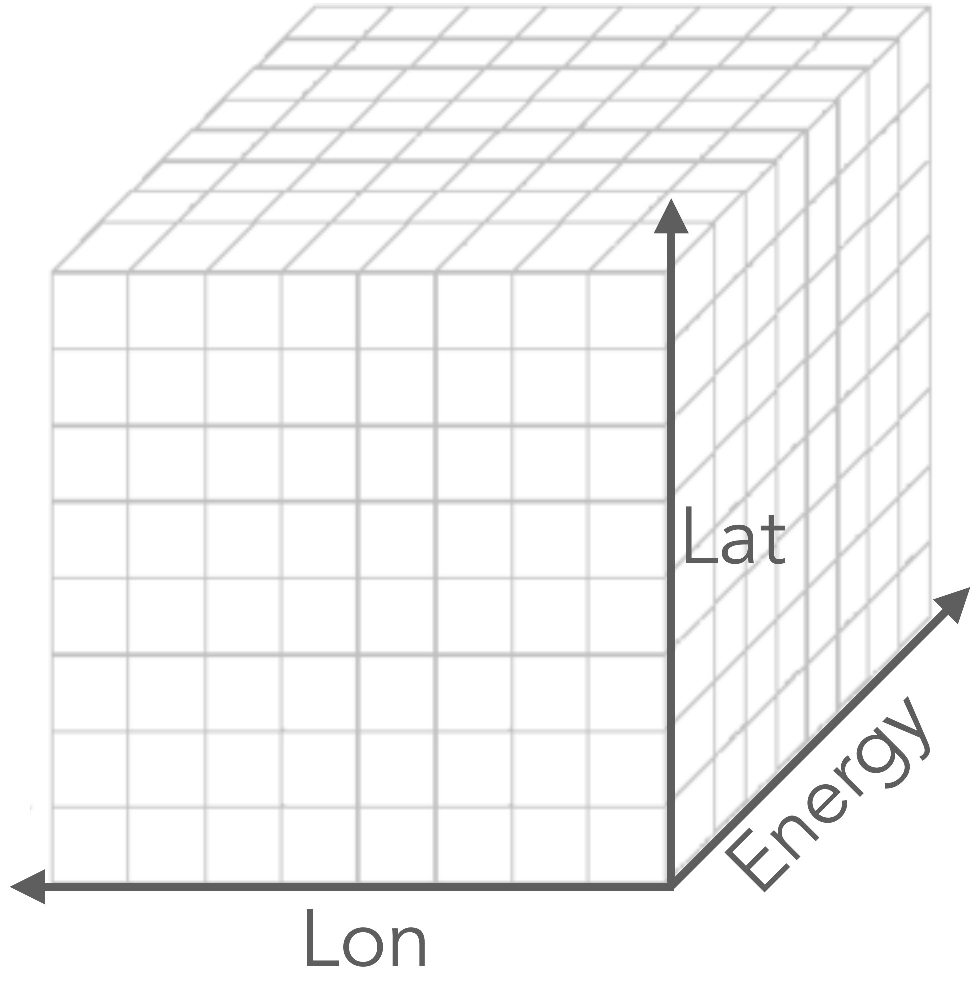
Data reduction

Bin events into ND sky maps

DL3 — γ-like events → **DL4** — Binned data → DL5 — Science products

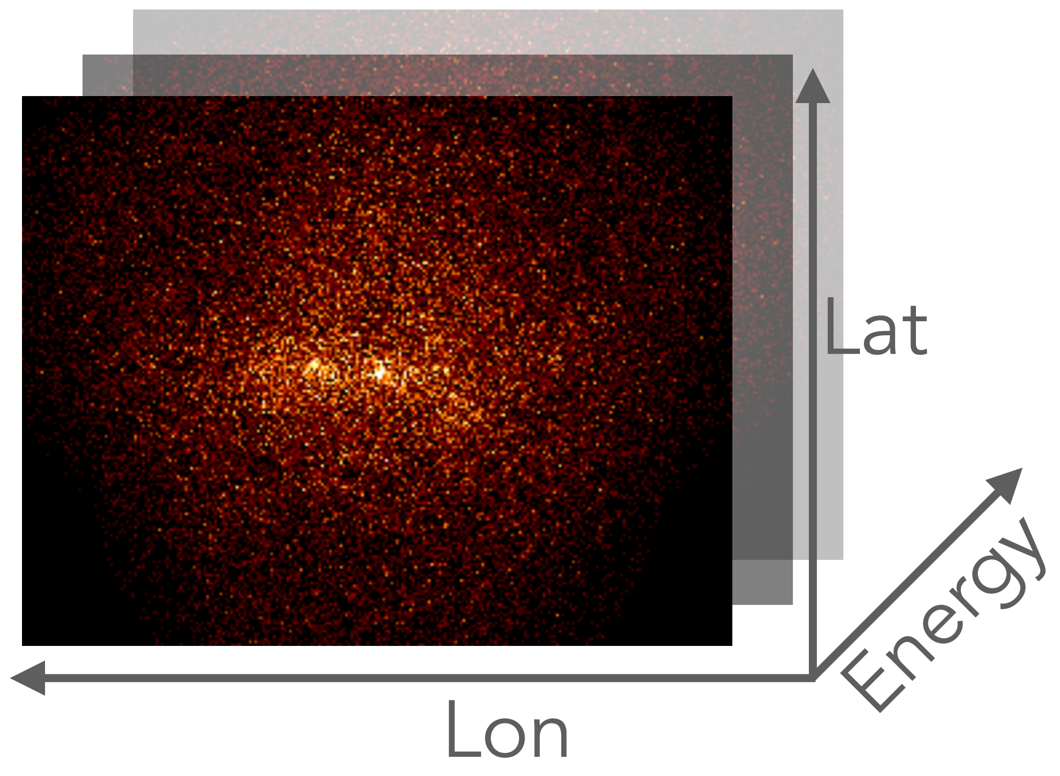
EVENT_ID	TIME	RA	DEC	ENERGY
	s	deg	deg	TeV
int64	float64	float32	float32	float32
5407363825684	123890826.66805482	84.97964	23.89347	10.352011
5407363825695	123890826.69749284	84.54751	21.004095	4.0246882
5407363825831	123890827.23673964	85.39696	19.41868	2.2048872
5407363825970	123890827.79615426	81.93147	20.79867	0.69548655
5407363826067	123890828.26131463	85.98302	21.053099	0.86911184
5407363826095	123890828.41393518	86.97305	21.837437	4.1240892
5407363826128	123890828.52555823	83.40073	19.771587	1.6680022
5407363826168	123890828.6829524	82.25036	19.22003	4.7649446
5407363826383	123890829.53362775	83.18322	22.008213	0.7920148
...	...	...	...	...

Observation and / or time selection



Bin selection: WCS & Energy

Counts (N_{Obs})



Skymap / "Cube"

- Selection of observation "ids" and / or time ranges
- Most general case: selection of spatial & energy binning

Data reduction

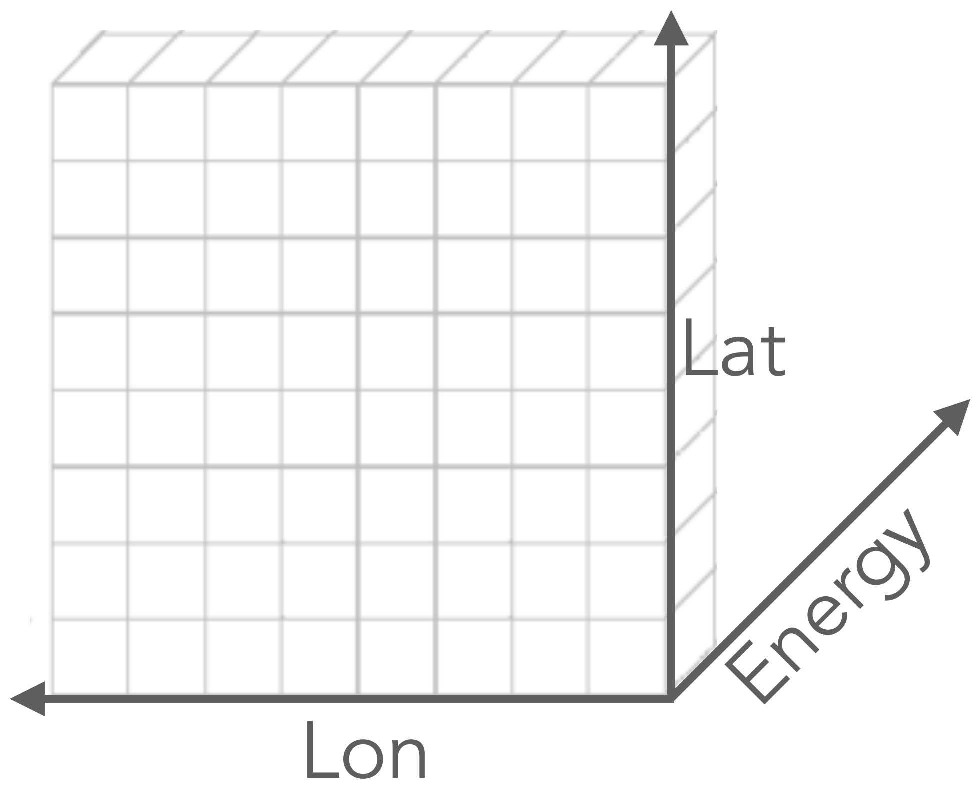
Bin events into ND sky maps

DL3 → DL4 → DL5

γ-like events — Binned data — Science products

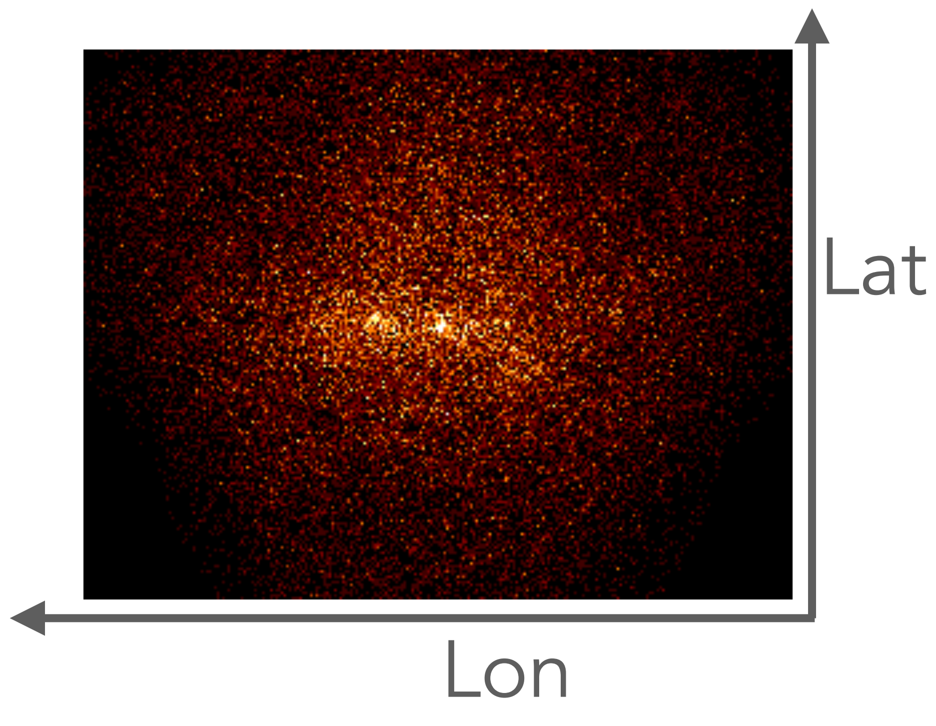
EVENT_ID	TIME	RA	DEC	ENERGY
	s	deg	deg	TeV
int64	float64	float32	float32	float32
5407363825684	123890826.66805482	84.97964	23.89347	10.352011
5407363825695	123890826.69749284	84.54751	21.004095	4.0246882
5407363825831	123890827.23673964	85.39696	19.41868	2.2048872
5407363825970	123890827.79615426	81.93147	20.79867	0.69548655
5407363826067	123890828.26131463	85.98302	21.053099	0.86911184
5407363826095	123890828.41393518	86.97305	21.837437	4.1240892
5407363826128	123890828.52555823	83.40073	19.771587	1.6680022
5407363826168	123890828.6829524	82.25036	19.22003	4.7649446
5407363826383	123890829.53362775	83.18322	22.008213	0.7920148
...	...	...	...	...

Observation and / or time selection



Bin selection: WCS & 1 Energy Bin

Counts (N_{Obs})



Skymap

- Image analysis is handled as a “cube” with one energy bin

Data reduction

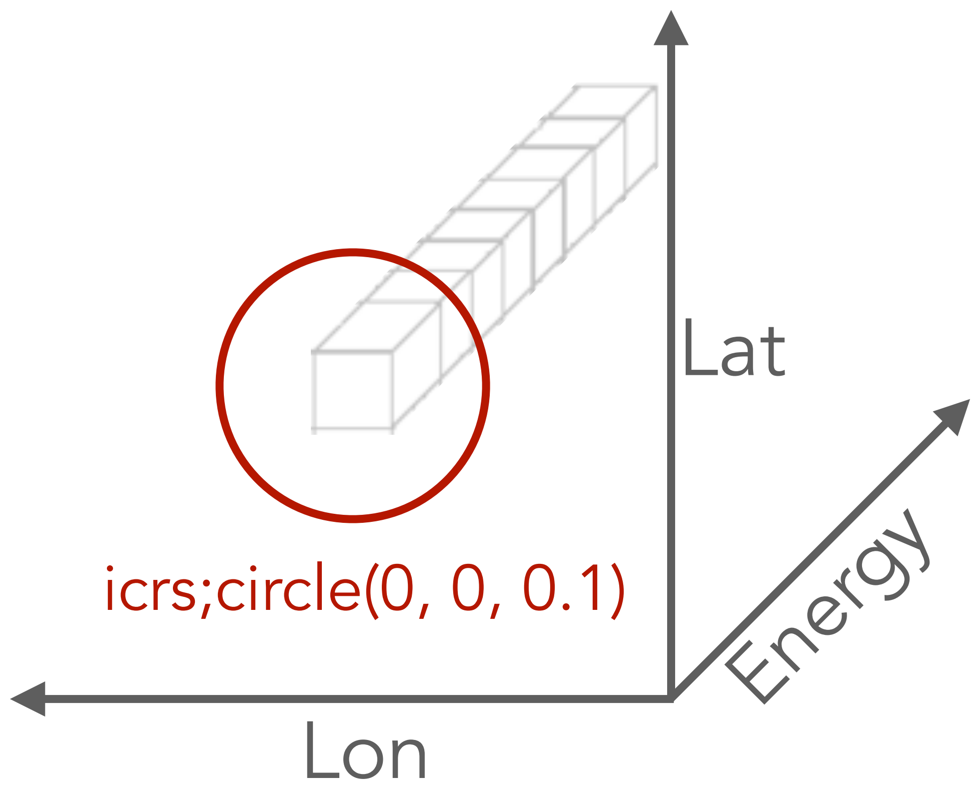
Bin events into ND sky maps

DL3 → DL4 → DL5

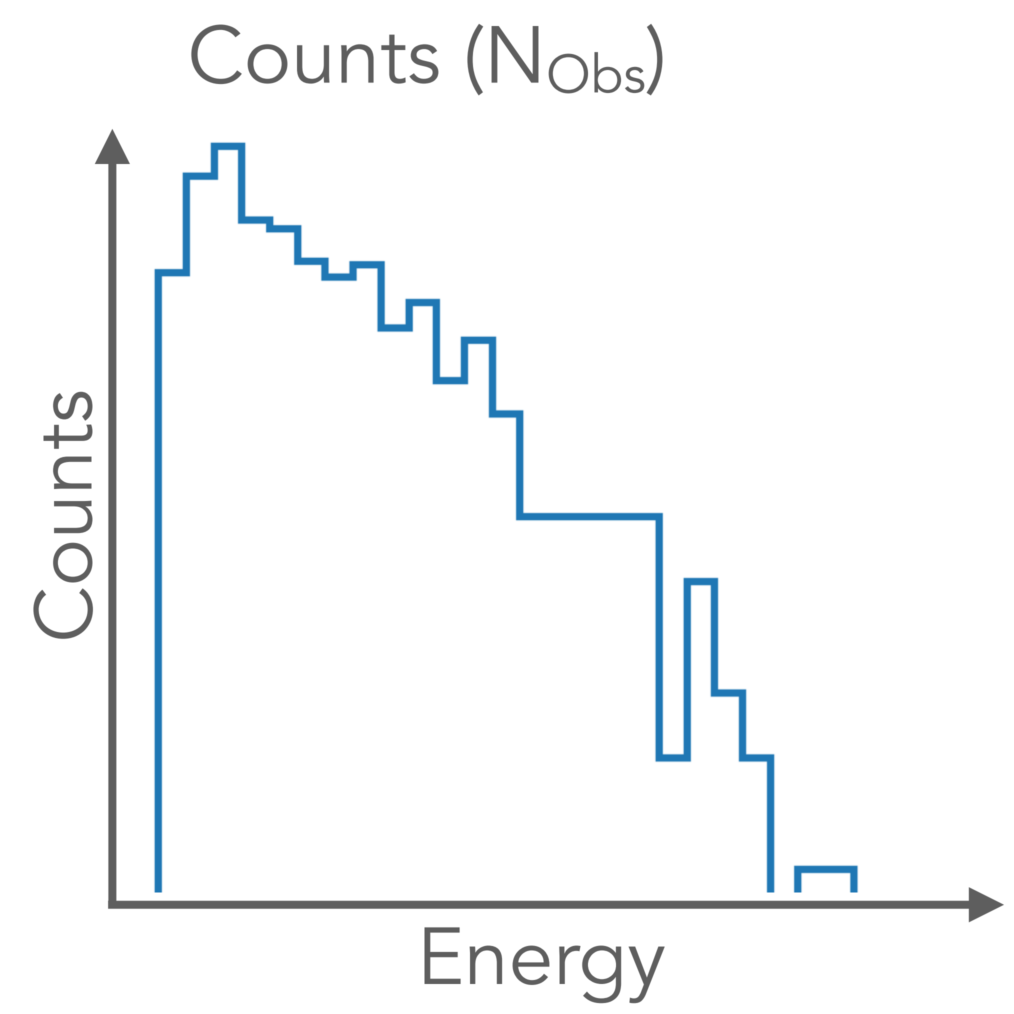
γ-like events | Binned data | Science products

EVENT_ID	TIME	RA	DEC	ENERGY
	s	deg	deg	TeV
int64	float64	float32	float32	float32
5407363825684	123890826.66805482	84.97964	23.89347	10.352011
5407363825695	123890826.69749284	84.54751	21.004095	4.0246882
5407363825831	123890827.23673964	85.39696	19.41868	2.2048872
5407363825970	123890827.79615426	81.93147	20.79867	0.69548655
5407363826067	123890828.26131463	85.98302	21.053099	0.86911184
5407363826095	123890828.41393518	86.97305	21.837437	4.1240892
5407363826128	123890828.52555823	83.40073	19.771587	1.6680022
5407363826168	123890828.6829524	82.25036	19.22003	4.7649446
5407363826383	123890829.53362775	83.18322	22.008213	0.7920148
...	...	...	...	...

Observation and / or time selection



Bin selection: Region & Energy



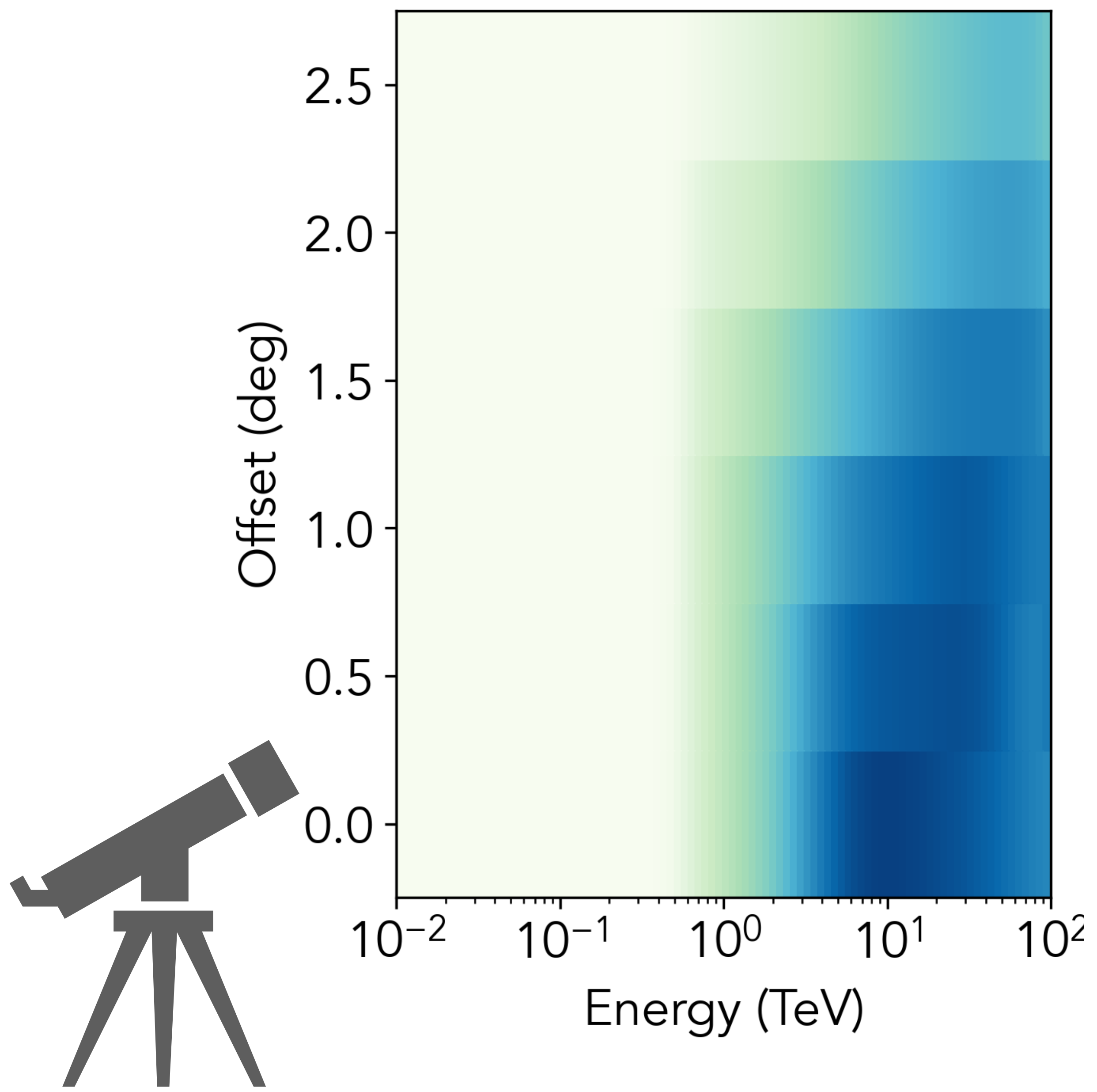
Spectrum

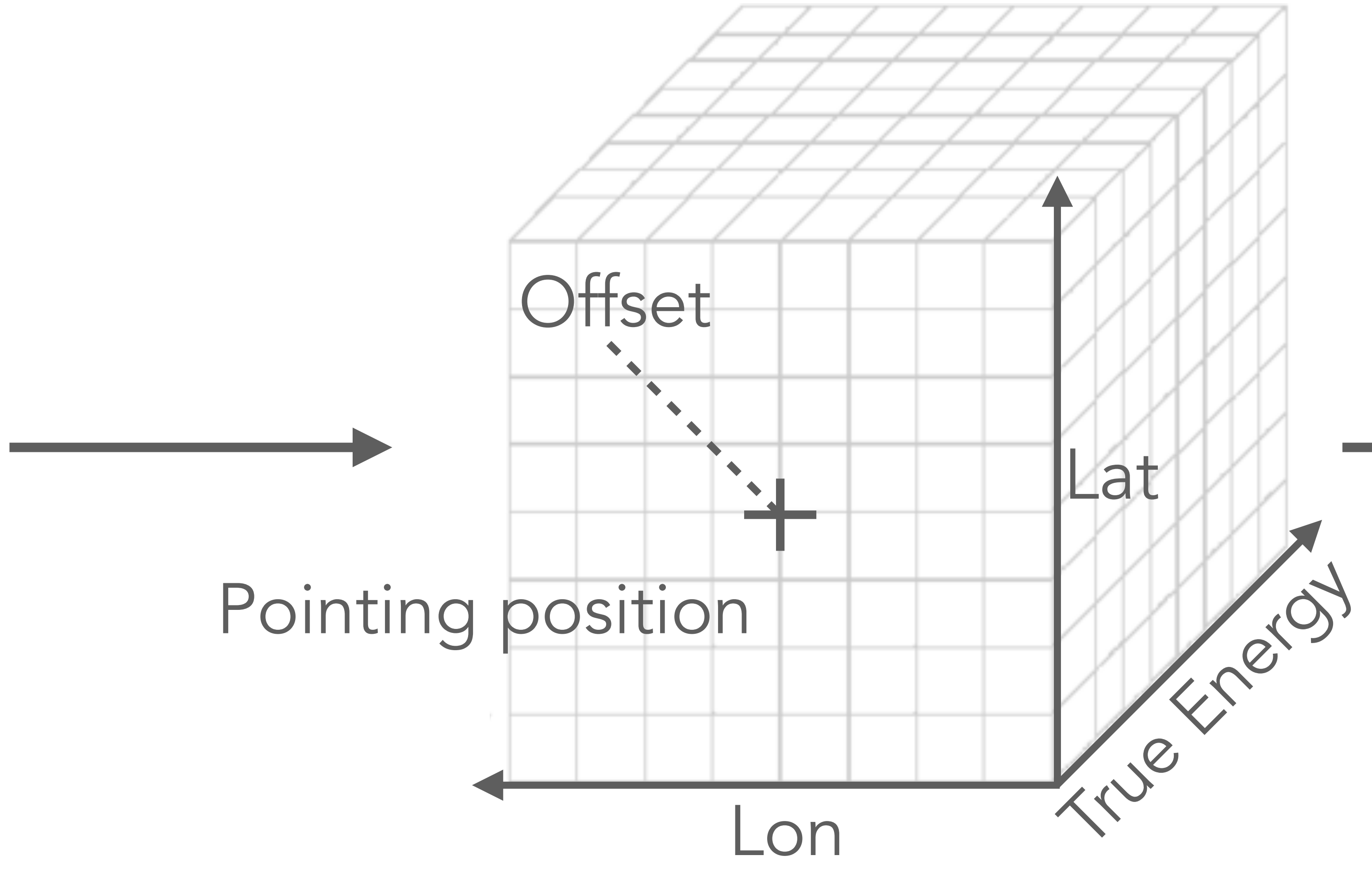
- Spectral analysis is handled as a "cube" with one spatial bin

Data reduction

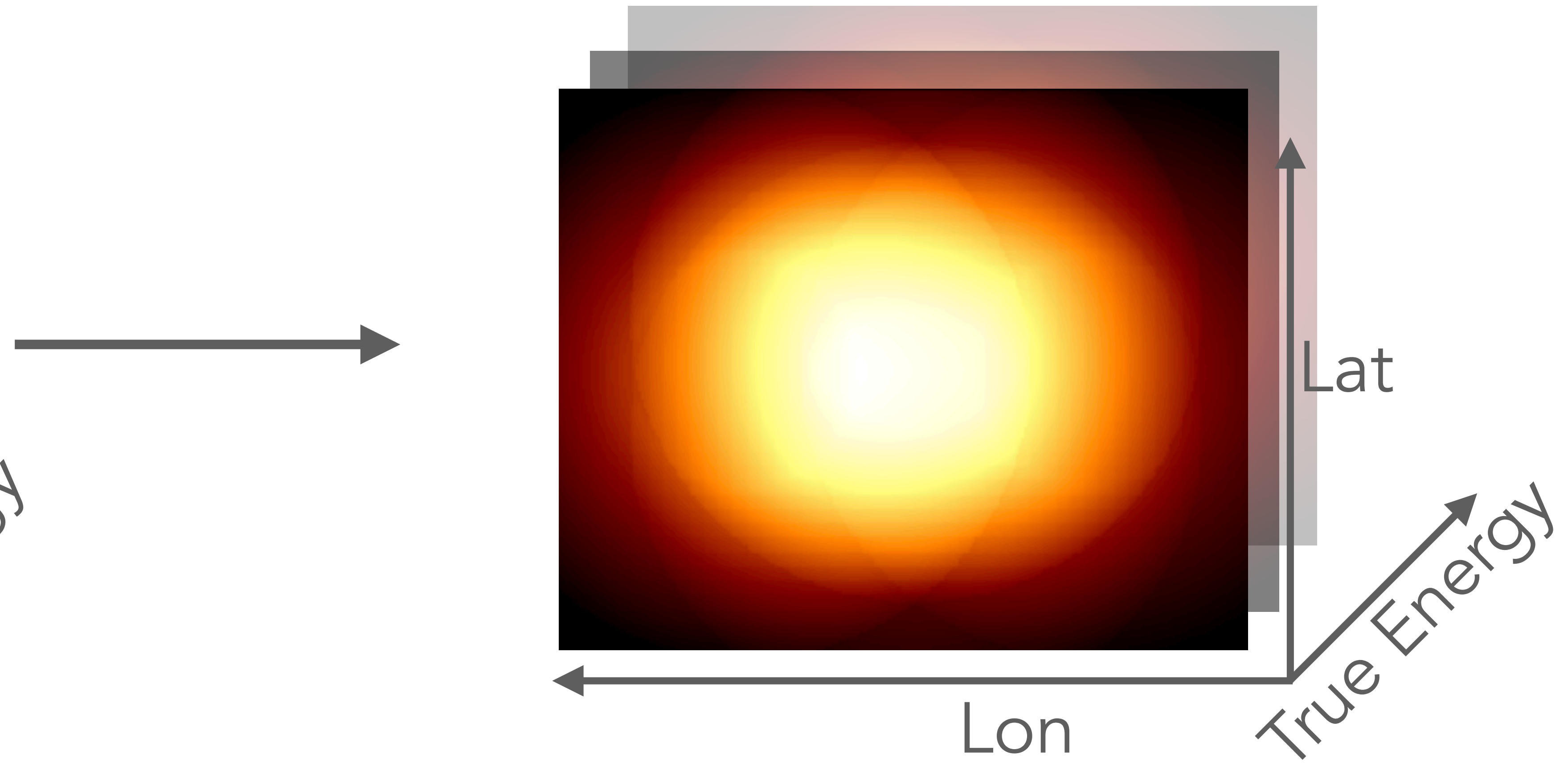
Reproject IRFs to ND sky maps

DL3 γ-like events → DL4 Binned data → DL5 Science products





Bin selection: WCS & Energy



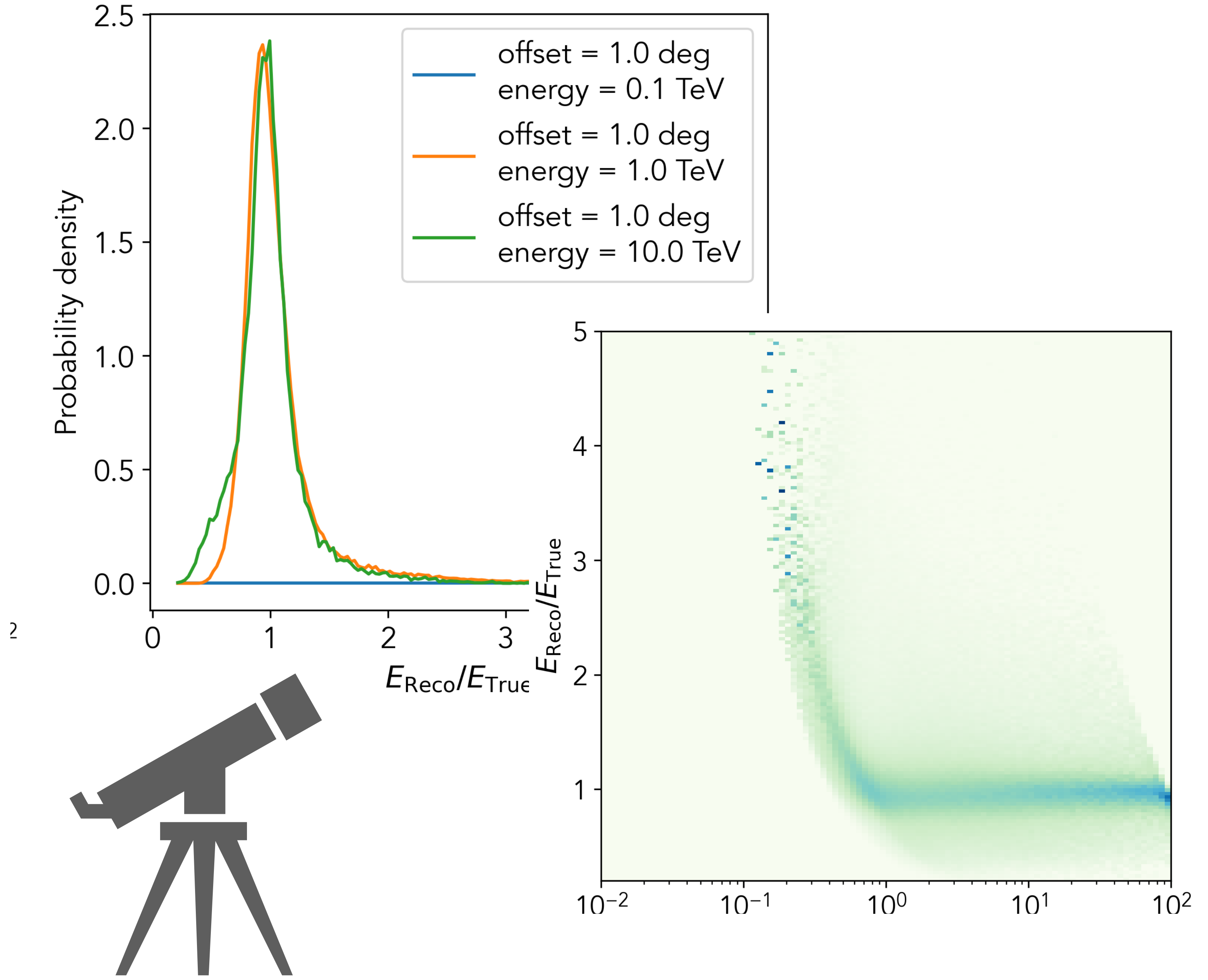
- Interpolate effective area on the chosen spatial and true energy binning
- Multiply with the (dead time corrected) observation time, to get an exposure map

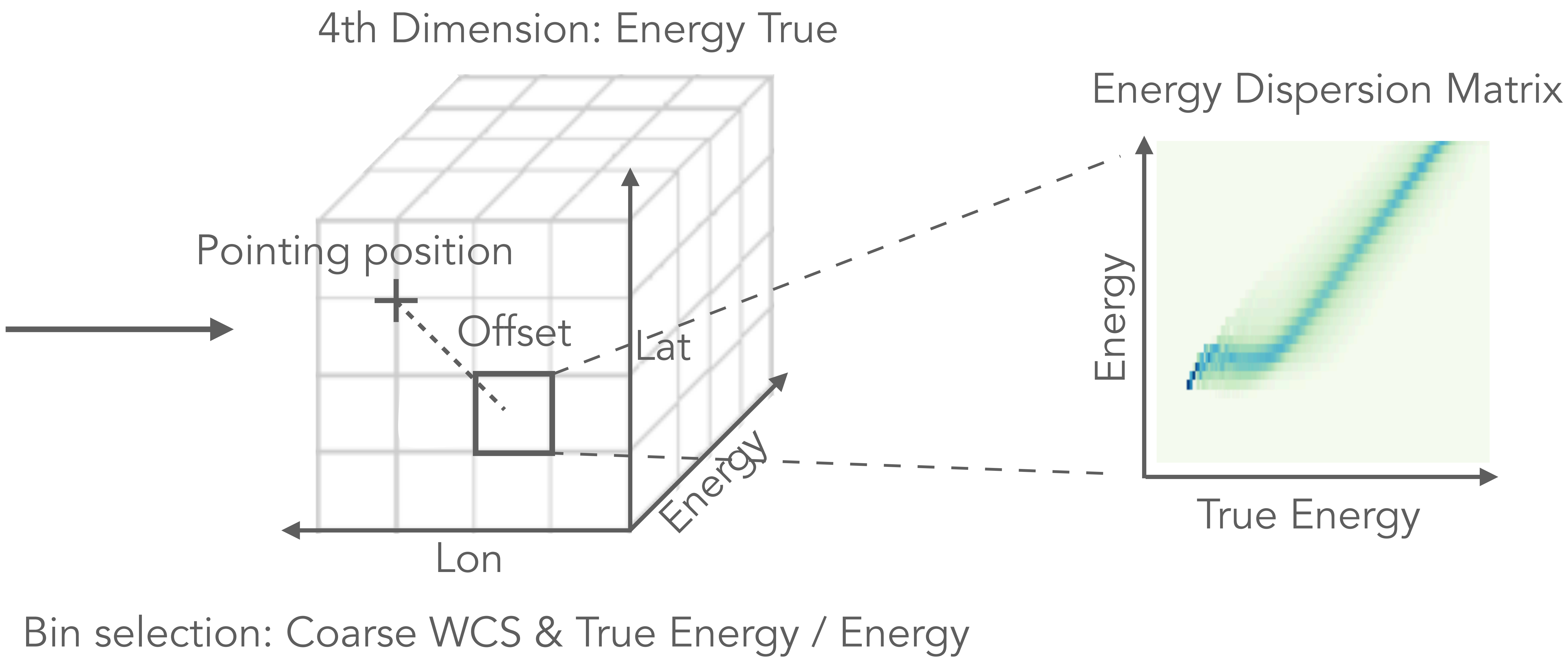
Data reduction

Energy dispersion

DL3 → DL4 → DL5

γ-like events | Binned data | Science products



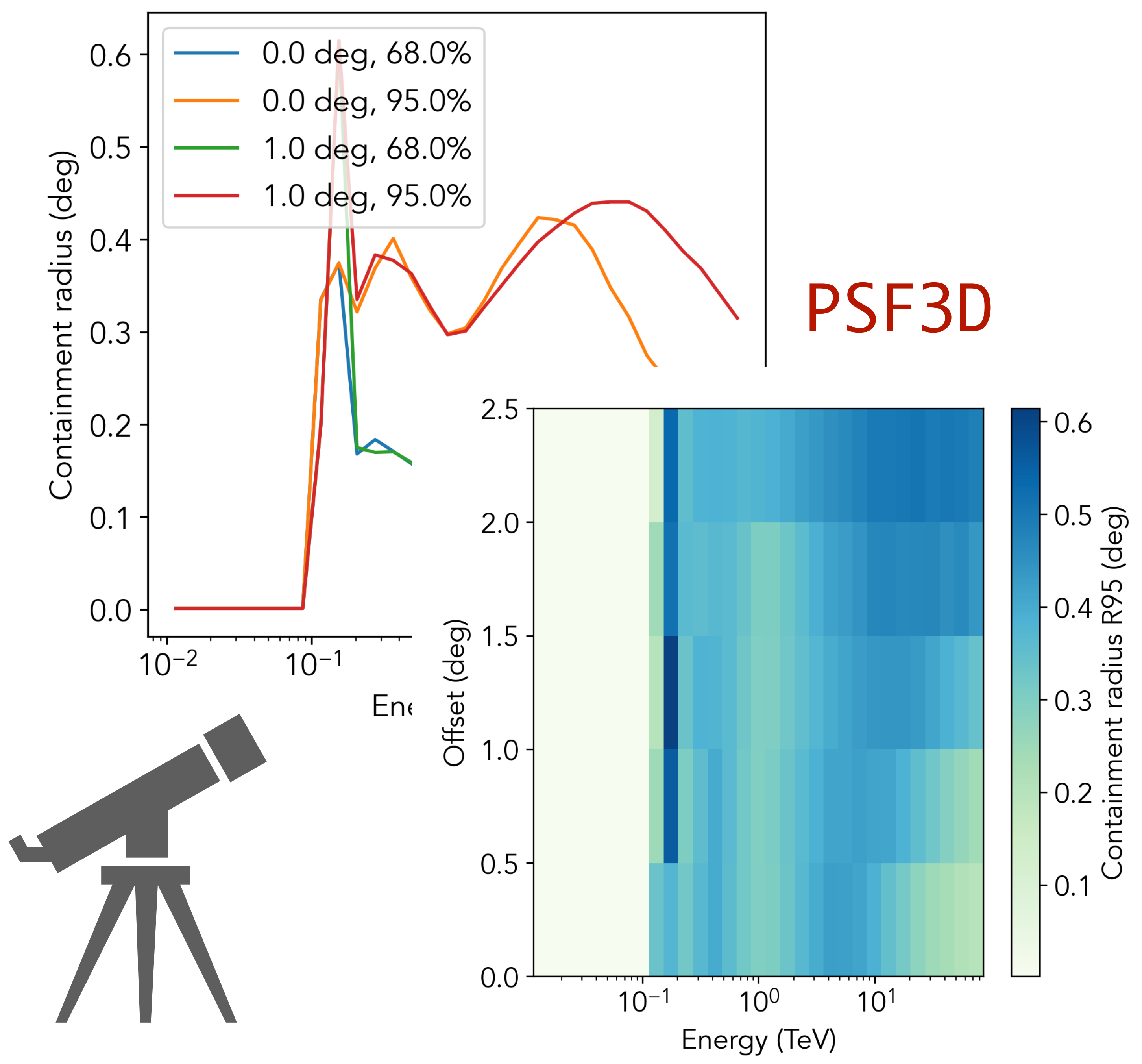


- Energy dispersion varies with offset: if multiple sources are fitted during the analysis a “map” of RMF matrices is required
- Typically computed on a WCS with coarser bins (~0.2 deg)
- During fitting / model evaluation the corresponding matrix for a given model component is extracted from the map and applied using matrix multiplication

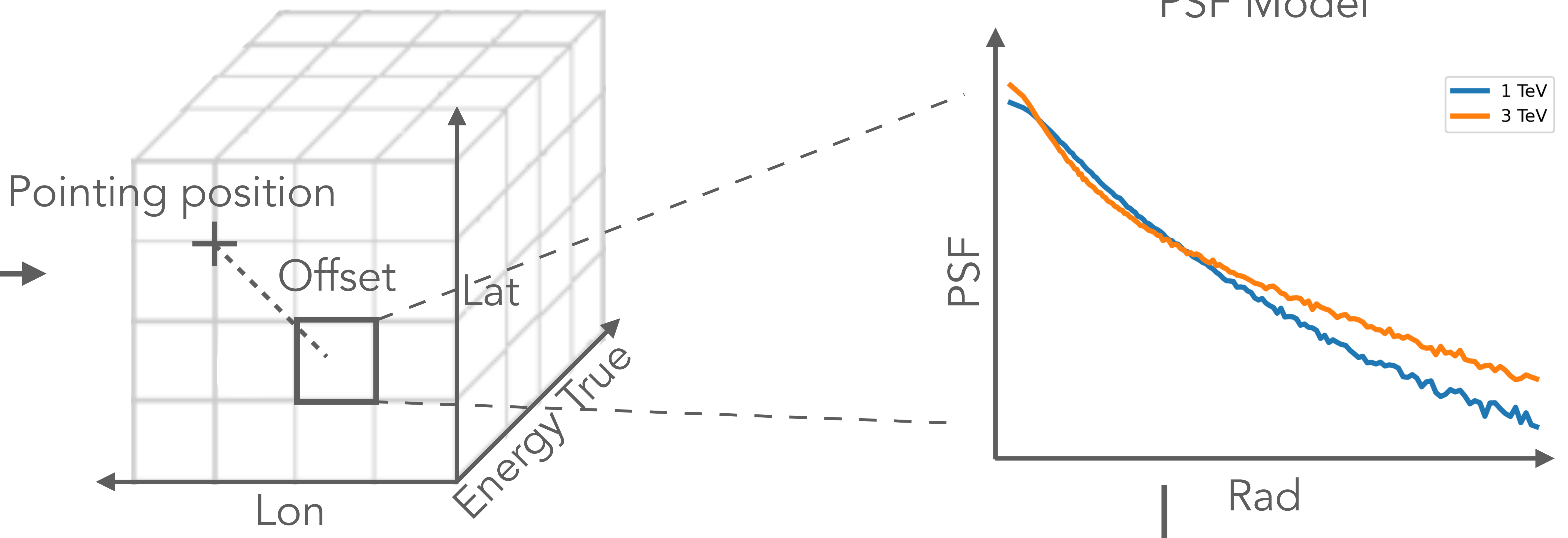
Data reduction
Point spread function

DL3 → DL4 → DL5

γ-like events — Binned data — Science products

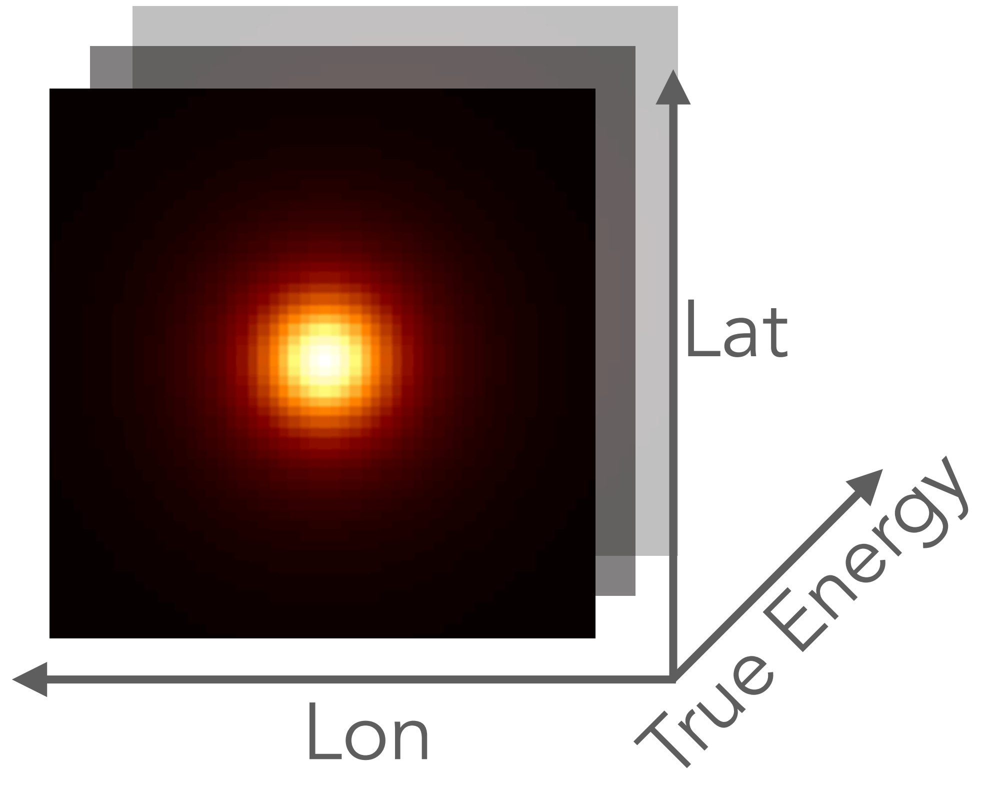


4th Dimension: Rad



Bin selection: Coarse WCS, True Energy & Rad

PSF Kernel



- PSF varies with offset: if multiple sources are fitted during the analysis a "map" of PSFs is required
- Typically computed on a WCS with coarser bins (~0.2 deg)
- During fitting / model evaluation the corresponding PSF the current position of a model component is extracted from the map
- From the radial profile a 3D "kernel" is computed and applied using convolution

Data model

Predict number of counts

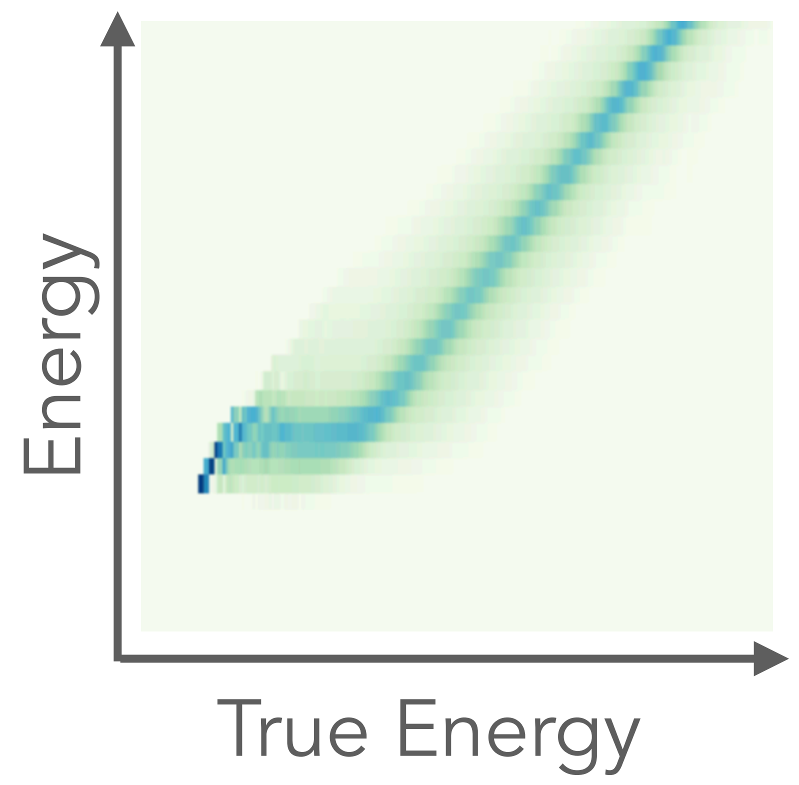
DL3 → DL4 → DL5

γ-like events → Binned data → Science products

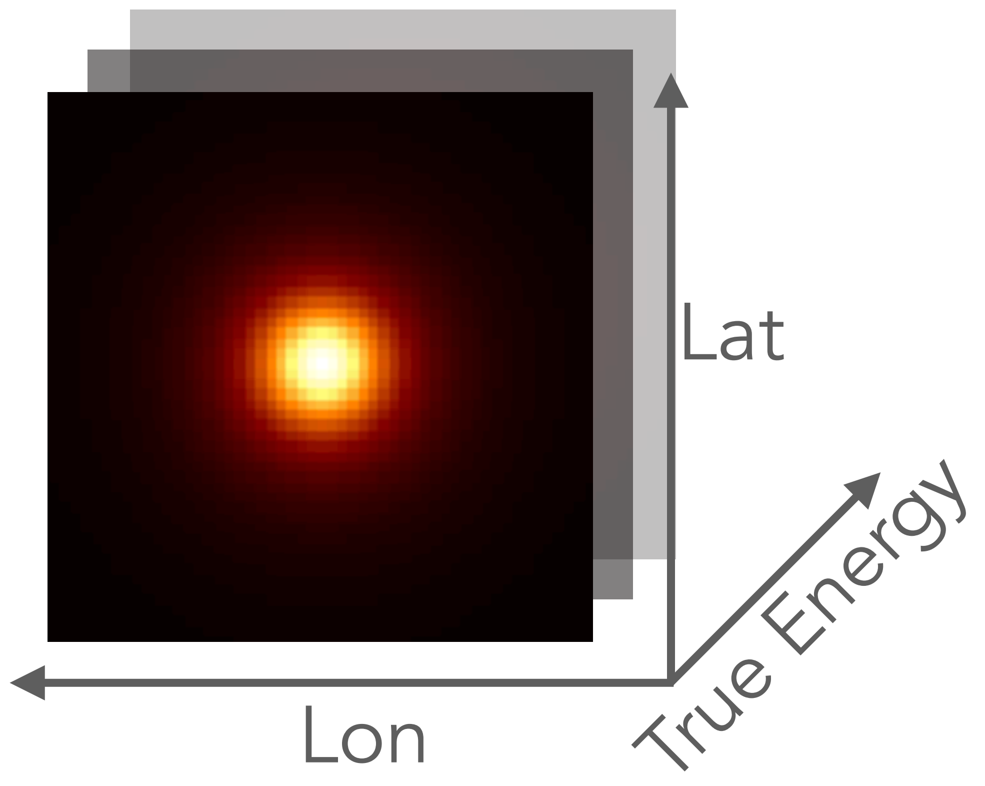
$$N_{Pred,Src} = \mathrm{EDISP}_{Src}(\mathrm{PSF}_{Src}(\mathscr{E}_{Src} \cdot f_{Src}))$$

$$f_{Src} = f_{Spectral}(E) \cdot f_{Spatial}(E, l, b) \cdot f_{Temporal}(t)$$

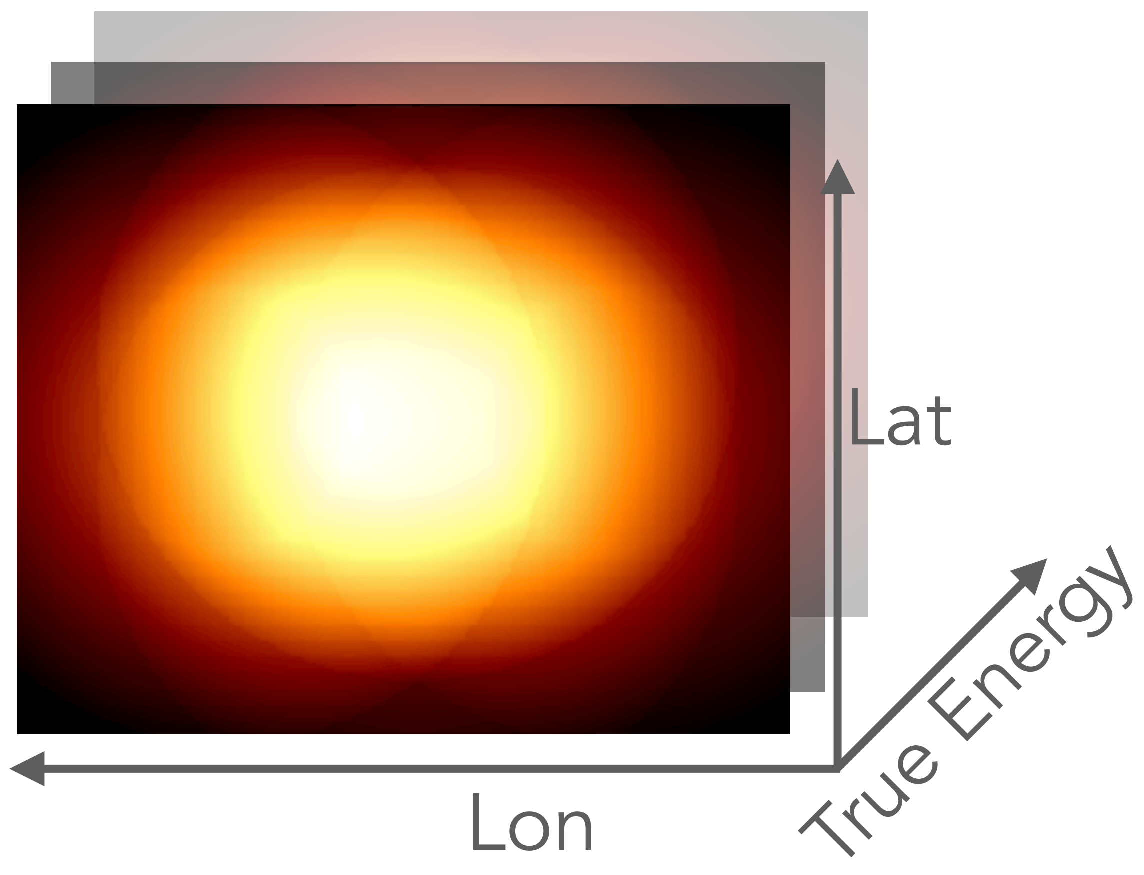
Energy Dispersion Matrix



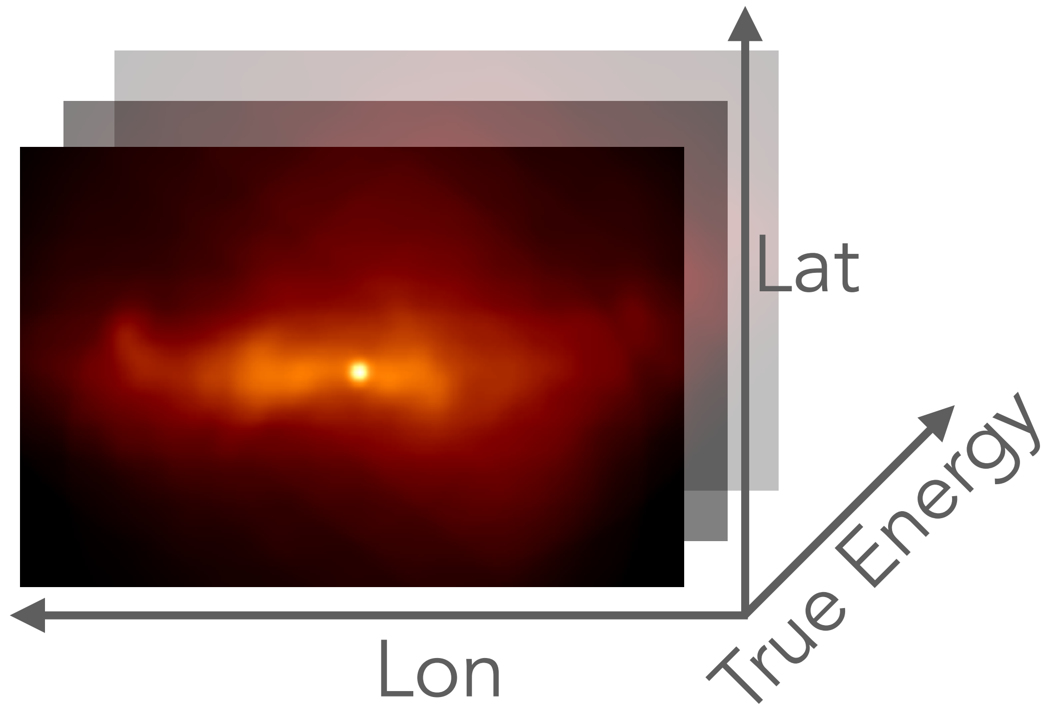
PSF Kernel



Exposure
(eff. area x lifetime)



Source Model



Max. likelihood fitting

"Fermi-LAT style" analysis

DL3 → DL4 → DL5

γ-like events — Binned data — Science products

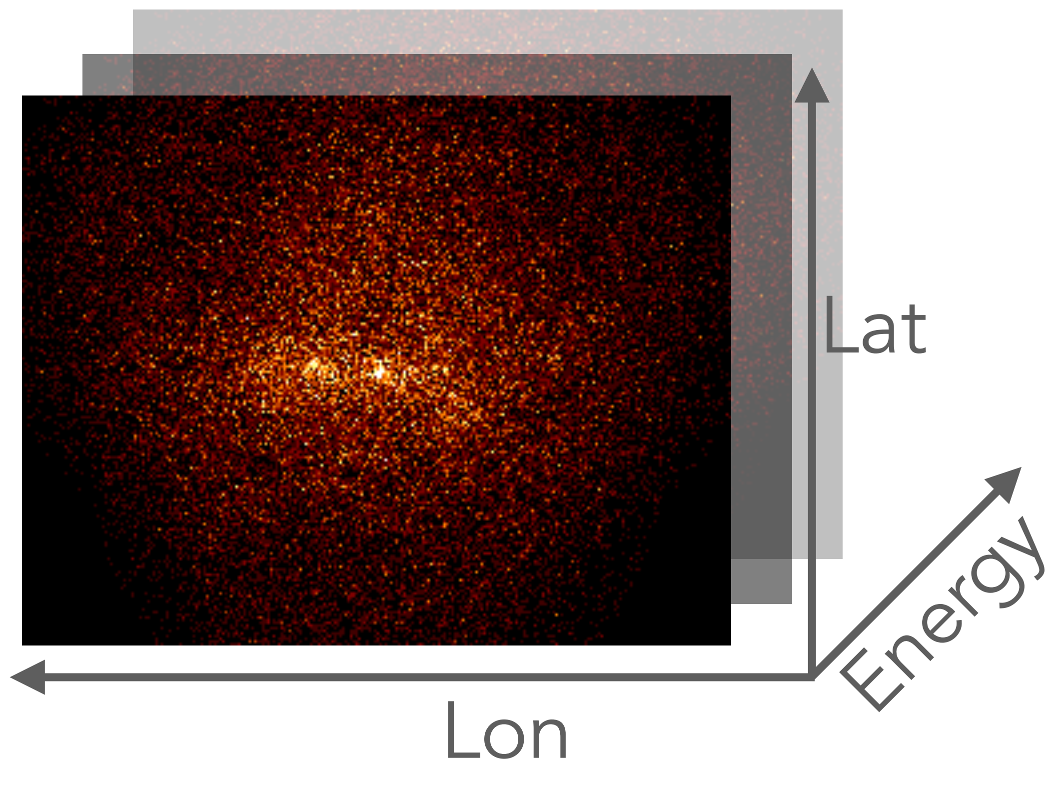
"Cash" statistics: summed over all "bins"

$$\mathscr{C} = 2\sum_i N^i_{Pred} - N^i_{Obs} \cdot \log N^i_{Pred}$$

$$N_{Pred} = N_{Bkg} + \sum_{Src} N_{Pred,Src}$$

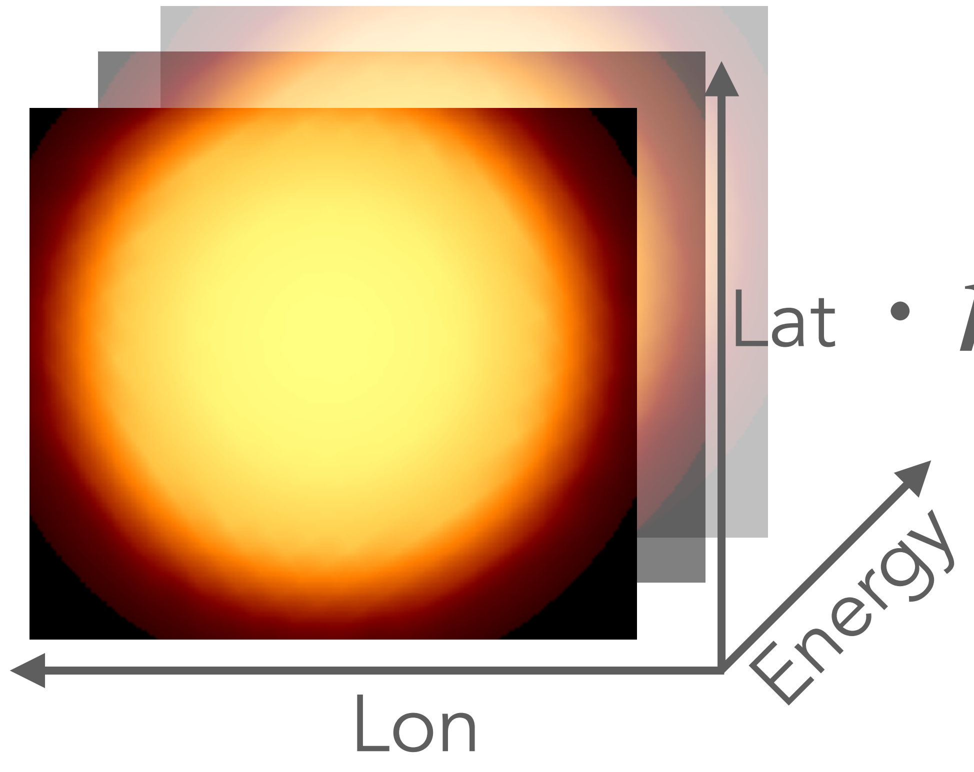
Counts

$N_{Obs} =$



Bkg Template

$N_{Bkg} =$



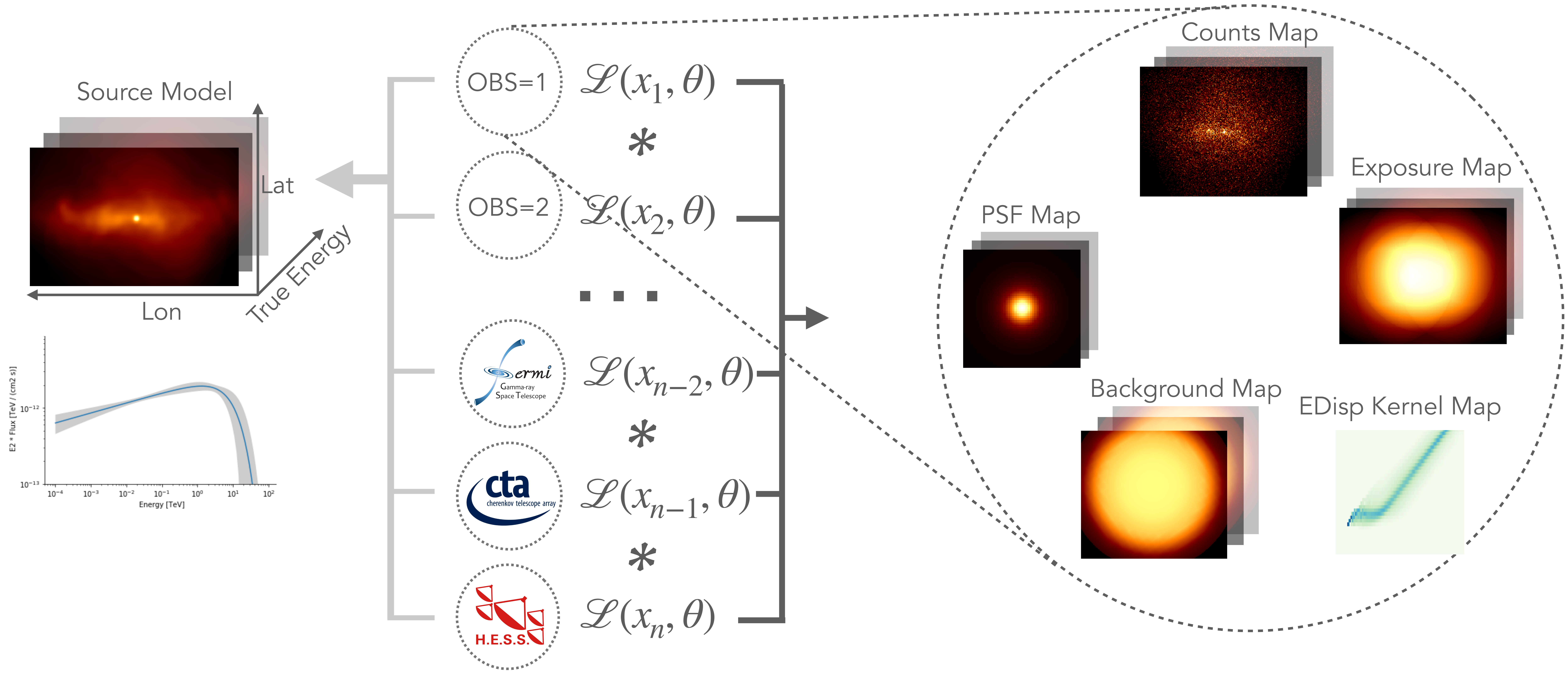
$\cdot\, n_0 \cdot \left(\frac{E}{E_0}\right)^{-\Gamma}$

- Predicted counts (as defined on the previous slide) are **computed per model component** ("source") and summed
- A **"global" background model** with "correction parameters" (like Fermi-LAT analysis) is added
- Summed per bin **fit statistics= - 2 log(L)** computed and **optimised**

Joint likelihood

Combining data

DL3 γ-like events → **DL4** Binned data → **DL5** Science products

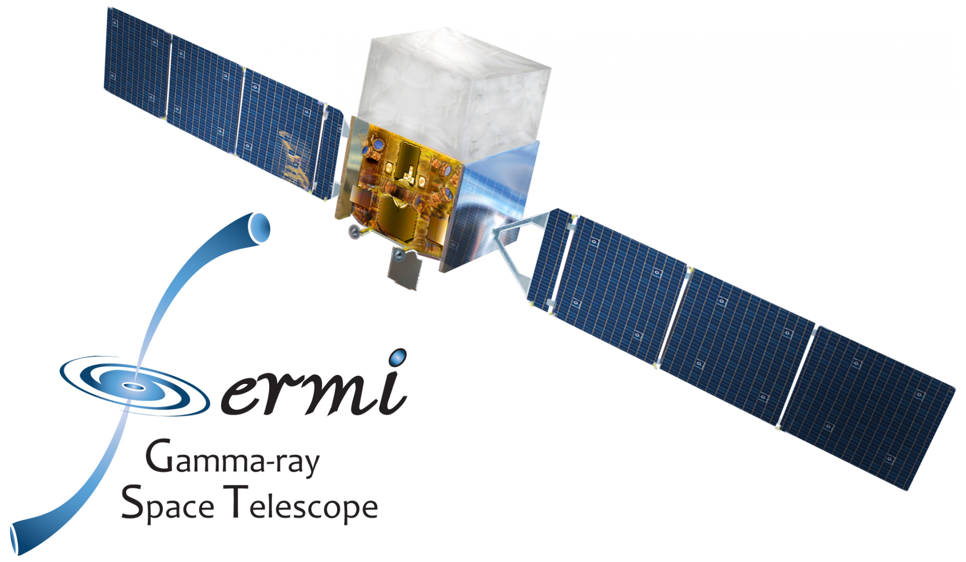


Gammapy

A python package for gamma-ray astronomy

Gammapy overview

Input data and general concept



Common FITS based data formats, inspired from Fermi-LAT:
https://gamma-astro-data-formats.readthedocs.io/en/latest/

EVENT_ID	TIME	RA	DEC	ENERGY
	s	deg	deg	TeV
int64	float64	float32	float32	float32
5407363825684	123890826.66805482	84.97964	23.89347	10.352011
5407363825695	123890826.69749284	84.54751	21.004095	4.0246882
5407363825831	123890827.23673964	85.39696	19.41868	2.2048872
5407363825970	123890827.79615426	81.93147	20.79867	0.69548655
5407363826067	123890828.26131463	85.98302	21.053099	0.86911184
5407363826095	123890828.41393518	86.97305	21.837437	4.1240892
5407363826128	123890828.52555823	83.40073	19.771587	1.6680022
5407363826168	123890828.6829524	82.25036	19.22003	4.7649446
5407363826383	123890829.53362775	83.18322	22.008213	0.7920148
...	...	...	...	...

$\gamma\pi$

DL3 —Data reduction / binning→ **DL4** —Model fitting→ **DL5**

DL3
- Event lists
- Raw IRFs

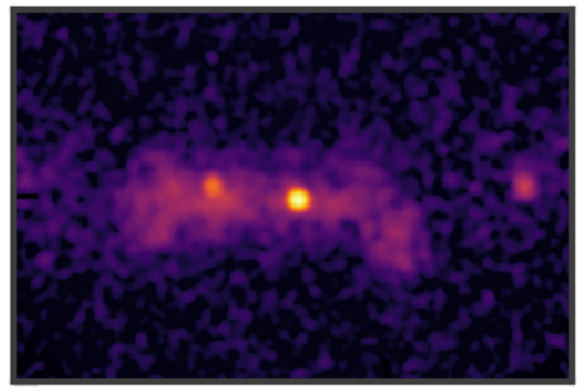
DL4
- Data binned in "Maps" using WCS, HPX or spectra
- Reduced IRFs e.g. PSF convolution kernel

DL5
- Source catalogs
- Flux points, SEDs
- Extension measures etc.

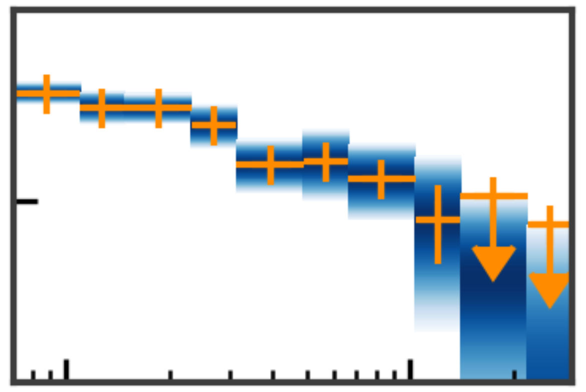
Source Catalogs

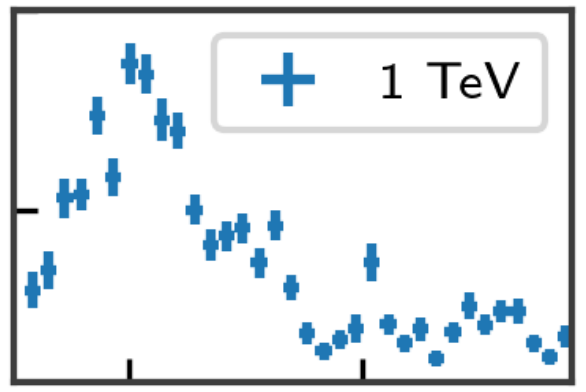
Name	Flux	Size
SNR	1e-12	1 deg
PWN	1e-11	0.2 deg
GRB	1e-10	0 deg

Flux & TS Maps



SEDs & Lightcurves





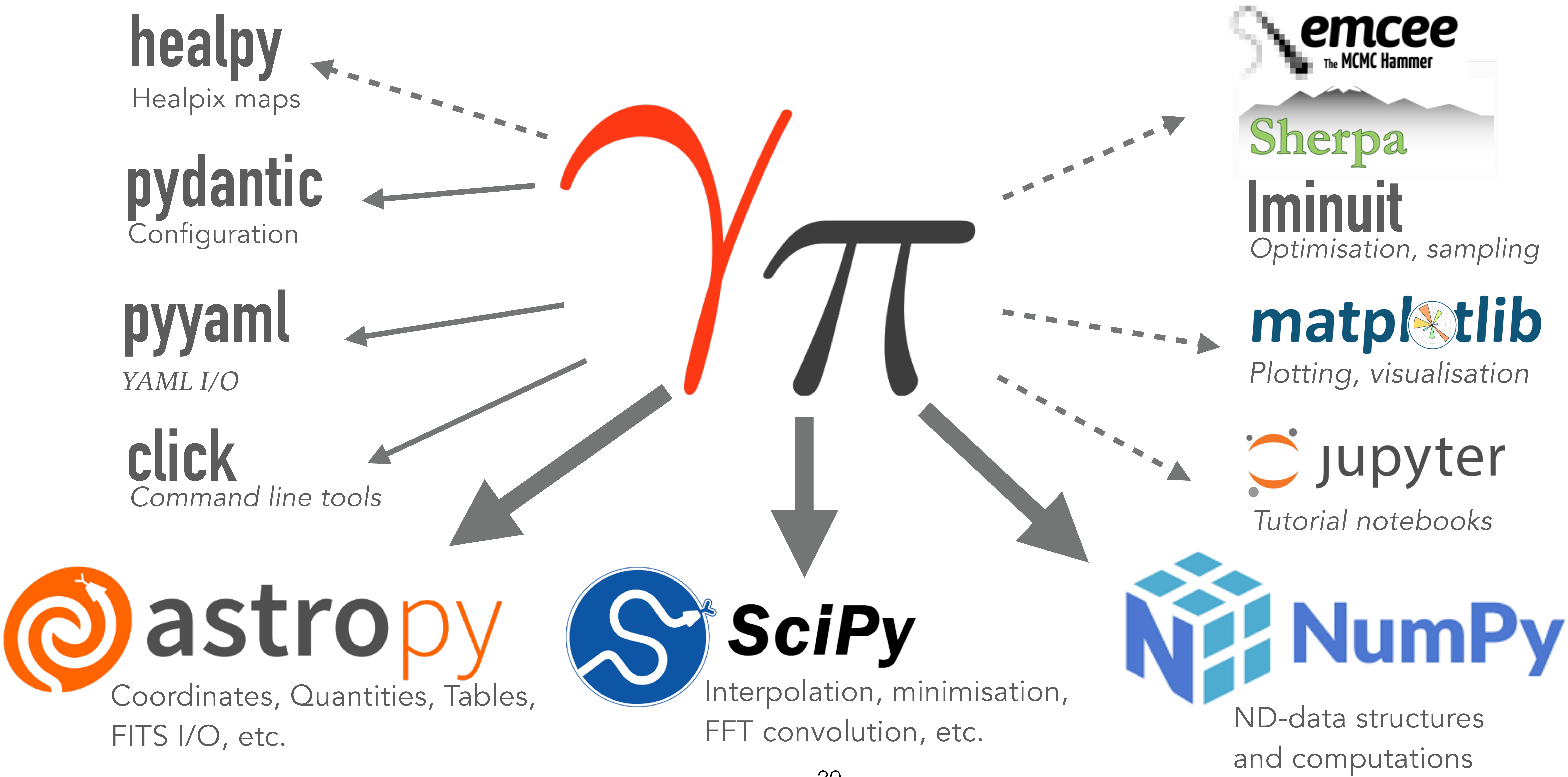
29

Dependencies

Optional dependencies: bring in useful functionality

Optional dependencies - - - - - - - ->

Required dependencies ———————>



Gammapy development

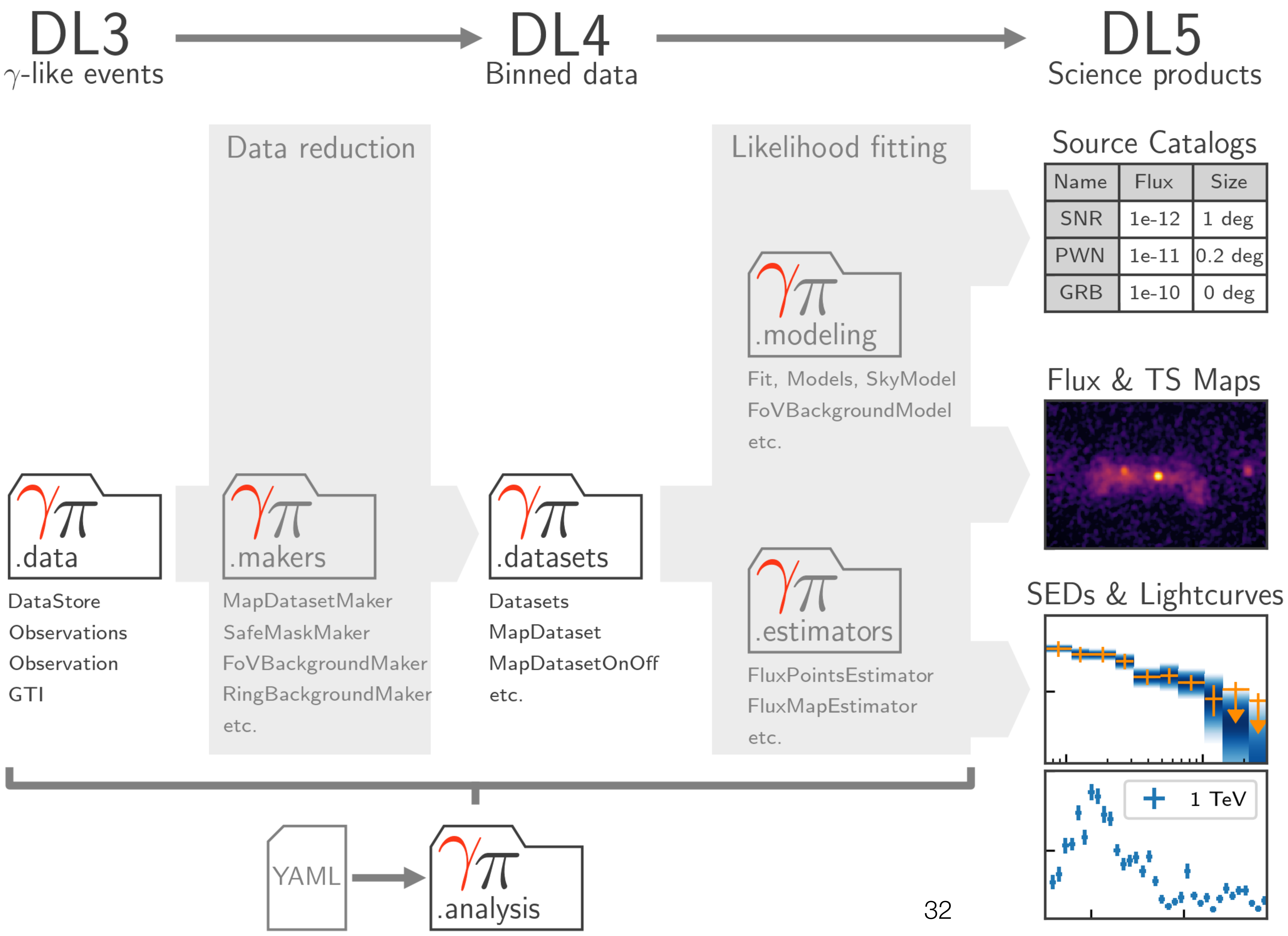
Setup, repository & workflow

- **Openly developed on GitHub**: https://github.com/gammapy/gammapy, using standard "multi branch GitHub" workflow: new features are developed on the side in branches and merged via pull request (PR) into the master branch
- Each PR is reviewed by a more experience developer
- **Continuous integration** via GitHub actions: testing of each PR
- Automatic build and **deployment of docs** to https://docs.gammapy.org/stable
- **Validation and benchmarks** run daily: https://github.com/gammapy/gammapy-benchmarks
- Uses pytest for testing and sphinx for documentation
- Gammapy **developer meeting and co-working day every Friday**, co working weeks and coding sprints
- Currently only minor versions, started to do bug fix releases as "training", working towards a **LST v1.0 version and paper this year**

GitHub Actions

Gammapy API

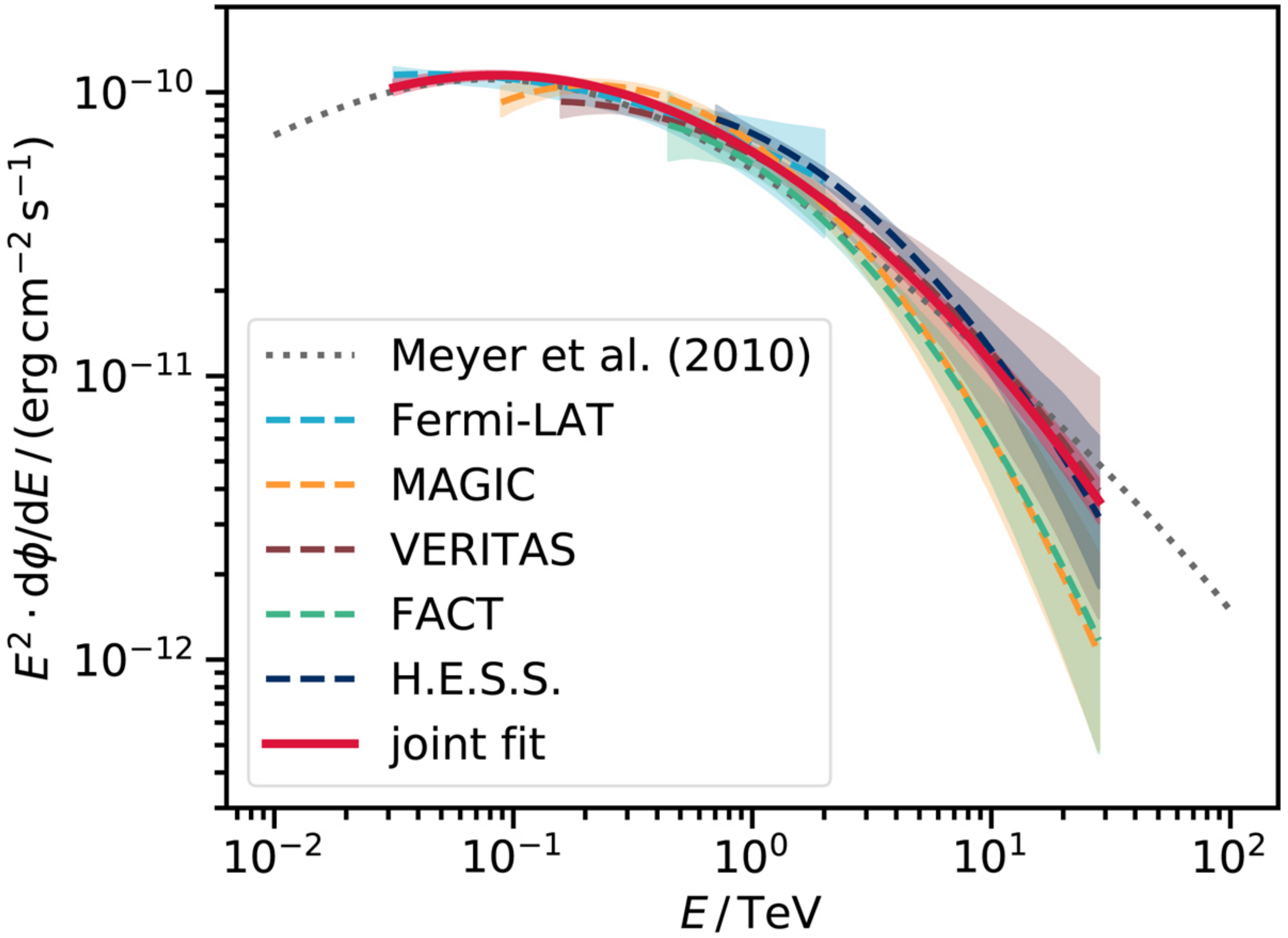
Subpackage structure and features

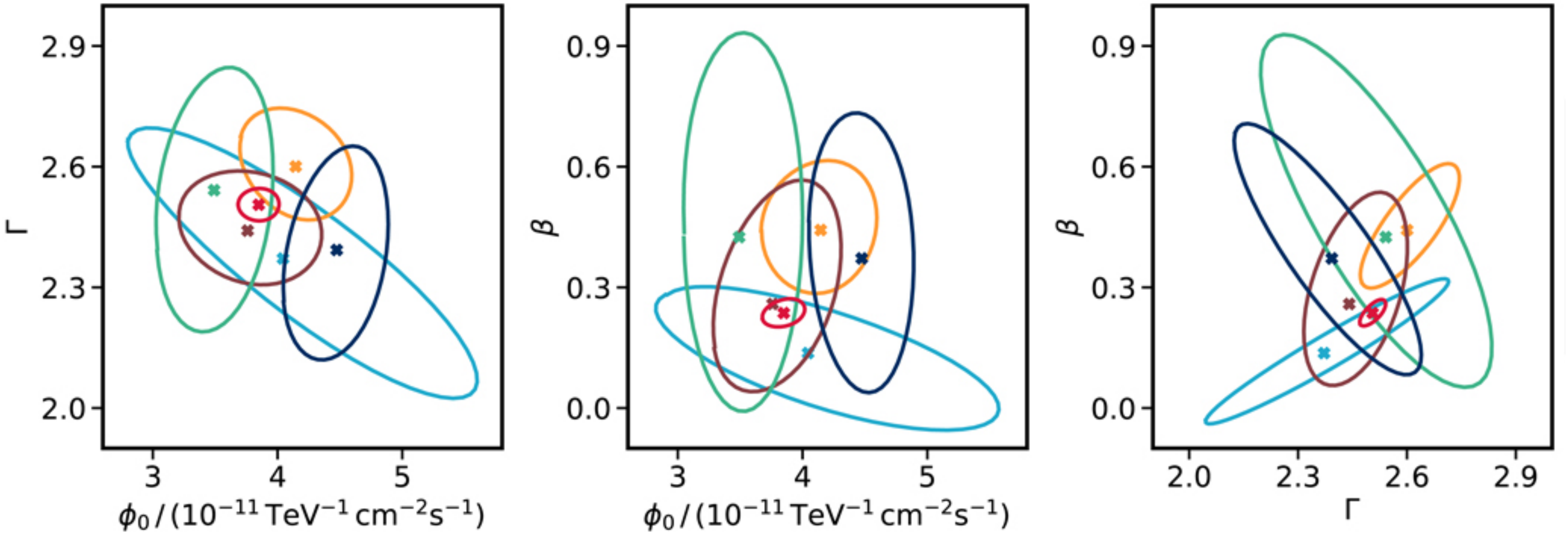


Name	Flux	Size
SNR	1e-12	1 deg
PWN	1e-11	0.2 deg
GRB	1e-10	0 deg

Application example I: "joint Crab" analysis

Combining data from different instruments



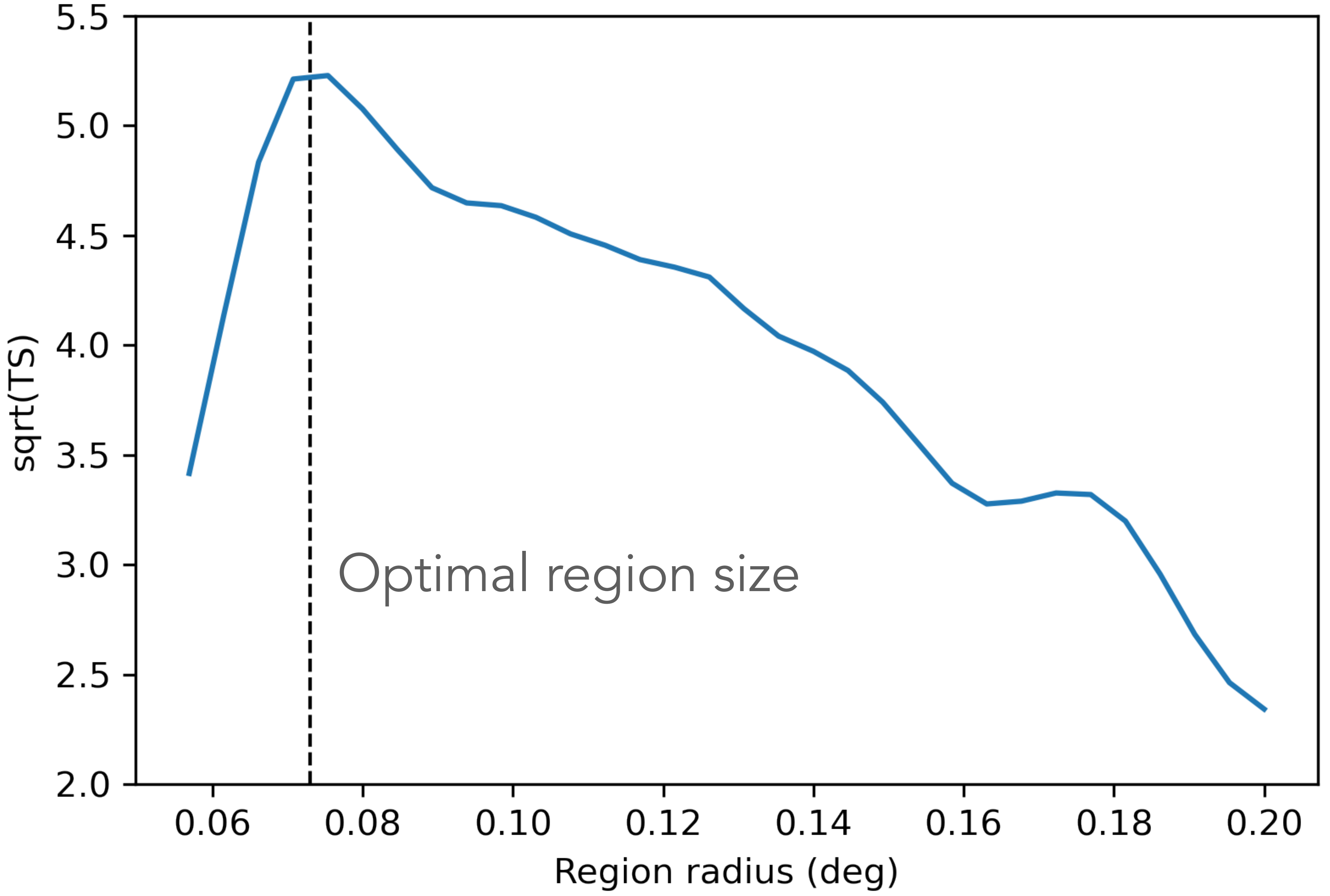


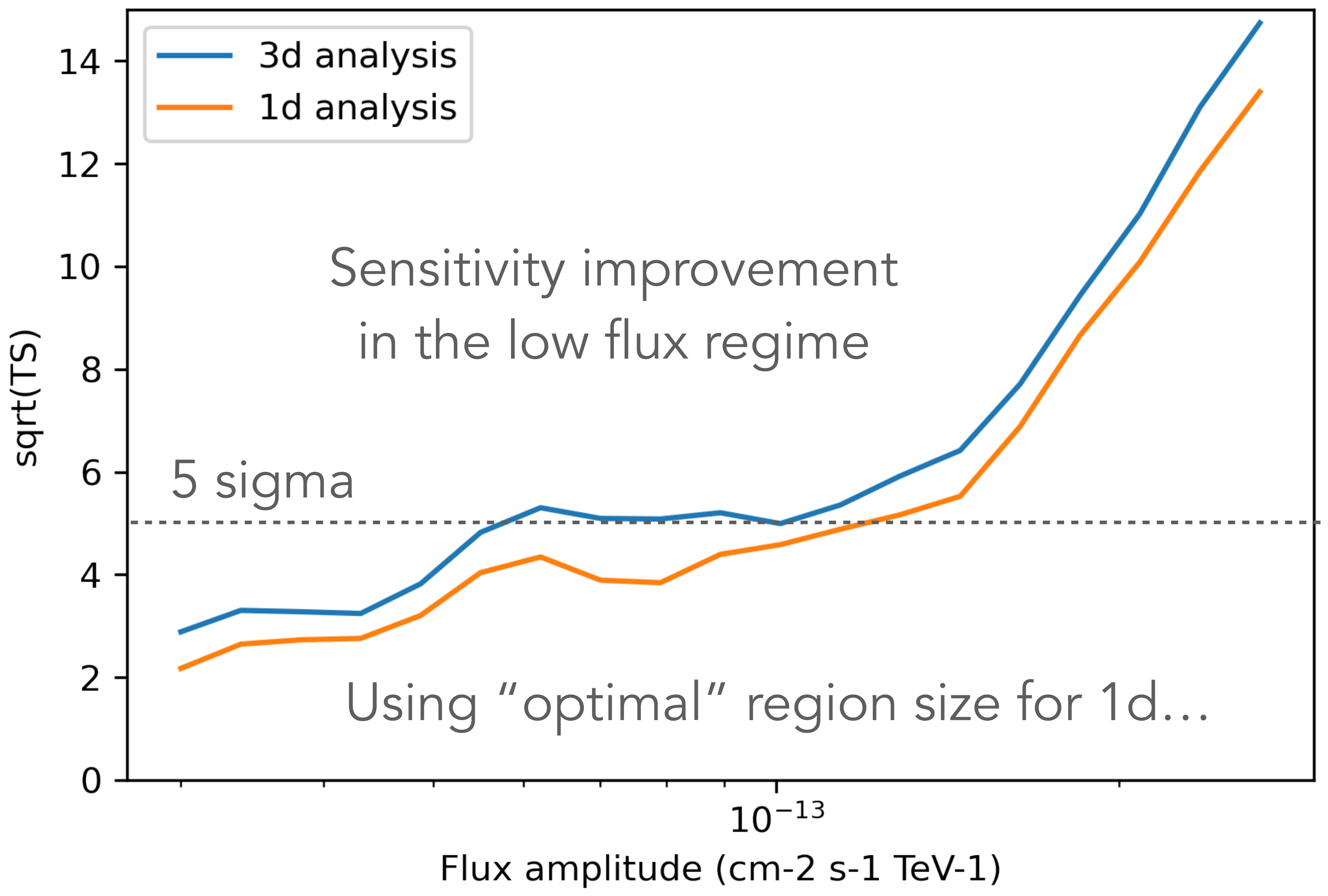
Taken from Nigro et al 2019:
https://www.aanda.org/articles/aa/full_html/2019/05/aa34938-18/aa34938-18.html

- Proof of concept for a multi-instrument analysis **using Gammapy v0.12**
- **Spectral analysis of the Crab Nebula** by fitting a log-parabola model. Includes data from H.E.S.S., MAGIC, VERITAS, FACT and Fermi-LAT instruments
- Data is combined at a likelihood level using a "joint-likelihood" approach: the **likelihood is evaluated per instrument / dataset** and the individual likelihoods are multiplied to the global likelihood
- The combination of data nicely **improves the statical error and leads to smaller correlations between the spectral parameters**, however requires treating systematic uncertainties of the instruments
- In the meantime data can be combined using a full "3D likelihood" as well

Application example II: 3d point source analysis

Mini study: 1d region based vs. 3d PSF fitting analysis





- Because of the limited angular resolution the significance / **sensitivity of a point source analysis** based on a region **varies with the size of the region**. In this case the optimal region size corresponds to ~80% containment radius of the PSF. But depends on shape and variation with energy of the PSF
- Even using the optimal region size the **"3d" analysis offers an improved sensitivity** for point sources because it takes into account more information on the instrument
- Possibility to **separate close by point sources** based on spectral information as well

Summary

- The H.E.S.S. Galactic plane survey showed that gamma-ray astronomy is entering the **“aera of catalogs”,** i.e. gamma-ray sources are not treated as individual phenomena, but **populations of sources** can be studied
- “Classical” gamma-ray analysis methods such as **region and image based analysis are not sufficient** to deal with challenges such as source confusion and complex varying IRFs.
- A more complete way to treat the spatial and spectral information simultaneously is a combined **spectro-morphological analysis.** It **o**ffers the possibility to better resolve close-by sources based on spectral information as well as separate “local” sources from diffuse emission based on spectral information
- Statistics and energy range of a measurement can be improved by **combining data from different instruments** and therefore improve statistical errors and constraints of spectral models
- There is a requirement for **analysis software** that implements this functionality and makes it available to a larger community of gamma-ray astronomers, which lead to creating the **Gammapy Python package, which is now a candidate for the CTA science tools as well**

Outlook

- The simultaneous spectral and spatial analysis of TeV gamma-ray data is not yet available since a long time in the gamma-ray community, so not many studies have been published yet. However there are many studies being worked on…
- The next Gammapy release, scheduled for May, will be the **release candidate for a v1.0 version with a stable API.** Together with the In addition there along with a **paper describing the software**
- **Using existing H.E.S.S. data measurements are already repeated using Gammapy** and combined spectro-morphological analysis, which also provides the simpler solution to **studying e.g. energy dependent morphology** of gamma-ray sources
- **Gammapy is a candidate for the CTA science tools** and is already used for simulated CTA data such as the CTA Galactic plane survey
- Gammapy is being **evaluated for use in other IACTs** as well, such a **VERITAS and MAGIC**
- Work has started to evaluate the use of **Gammapy for ground based water Cherenkov observatories such as HAWC** as well
- This opens up the possibility for many future **combined measurements using e.g. Fermi-LAT, H.E.S.S., VERITAS, MAGIC and HAWC data** and increase the statistics as well as energy range of existing measurements considerably

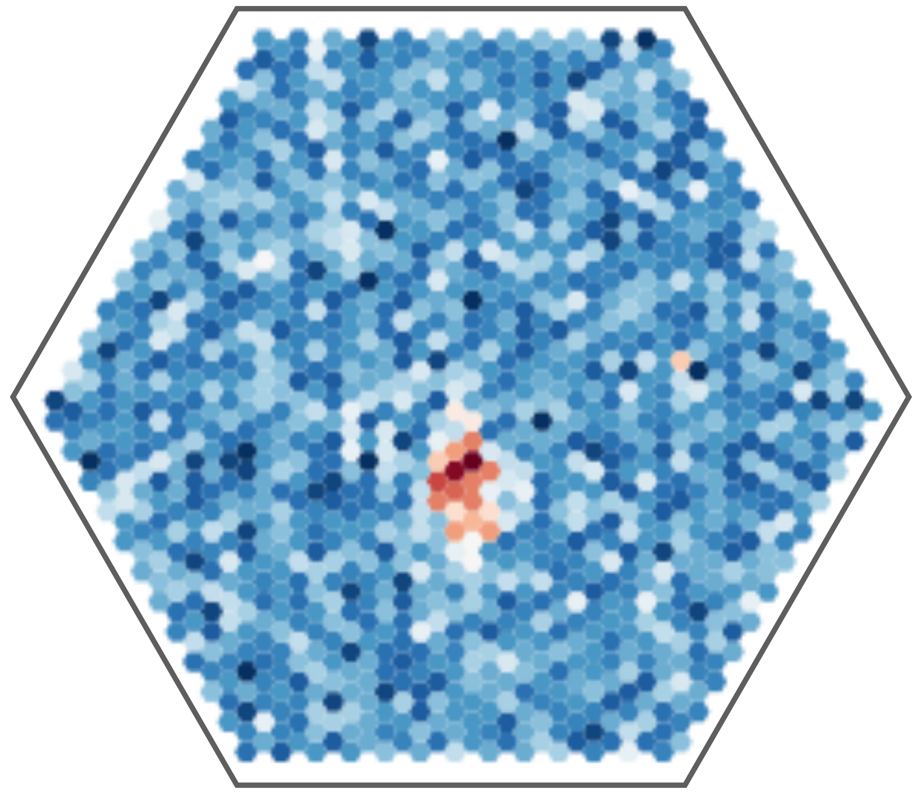
Backup slides

Event lists

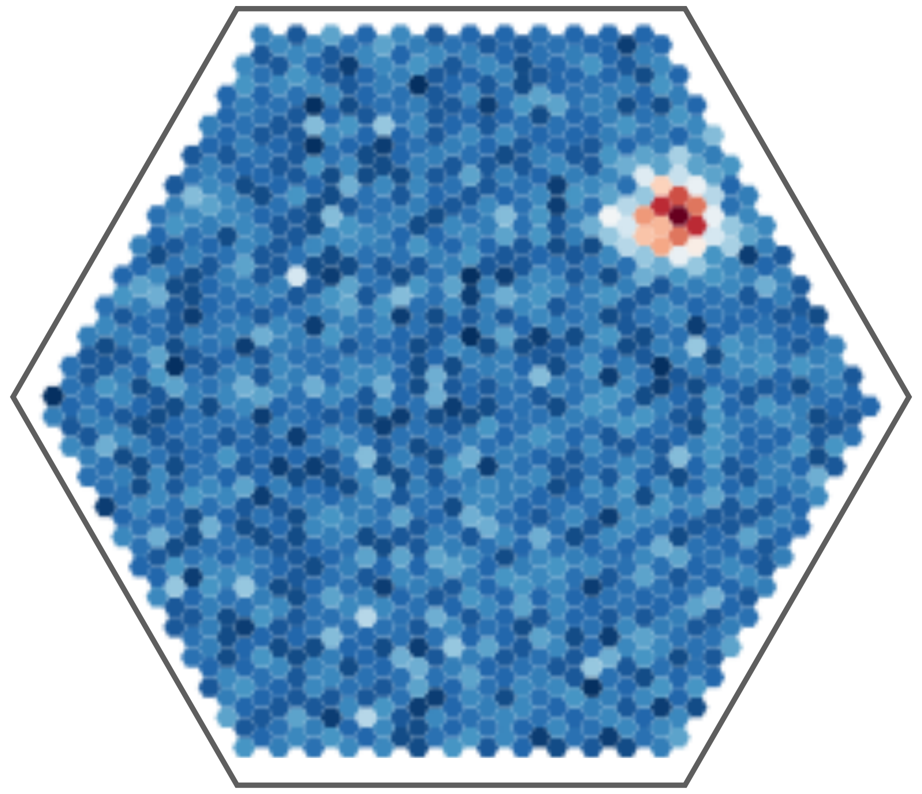
Reconstruction &
Gamma/Hadron Separation

DL3 → DL4 → DL5

γ-like events — Binned data — Science products



Telescope I Camera



Telescope II Camera

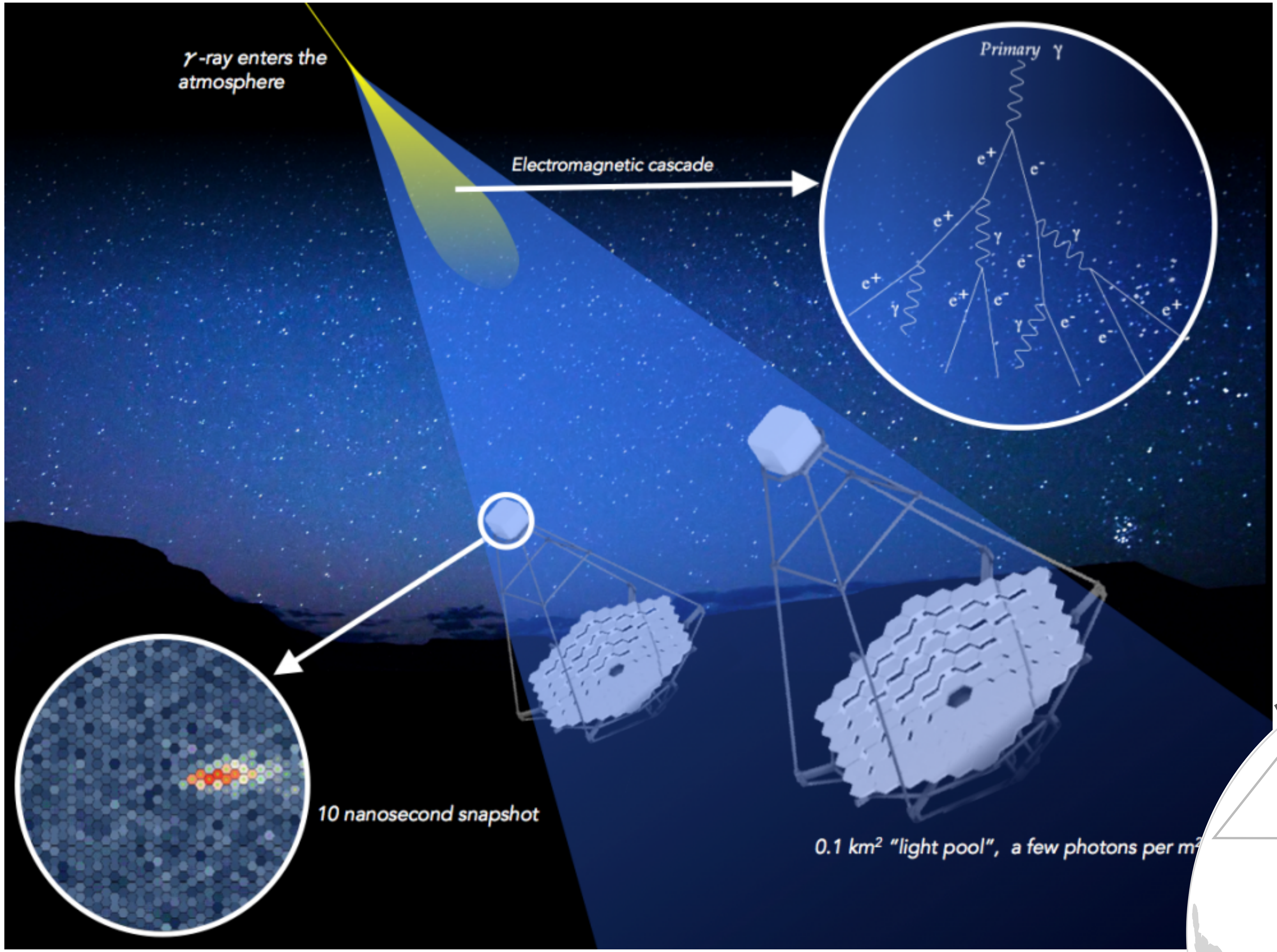
- **Goal:** estimate properties of the event: **energy, arrival time, arrival direction,** possibly reconstruction quality (“event class”)
- Separation between Gamma-like and Hadron-like events
- Only “Gamma-like” events are kept in the list
- However it is still **>90% background**, i.e. hadronic (or electron) events “mis-classified” as Gamma events (**requires additional background estimation**, comes later…)

List of gamma-like events

EVENT_ID	TIME	RA	DEC	ENERGY
	s	deg	deg	TeV
int64	float64	float32	float32	float32
5407363825684	123890826.66805482	84.97964	23.89347	10.352011
5407363825695	123890826.69749284	84.54751	21.004095	4.0246882
5407363825831	123890827.23673964	85.39696	19.41868	2.2048872
5407363825970	123890827.79615426	81.93147	20.79867	0.69548655
5407363826067	123890828.26131463	85.98302	21.053099	0.86911184
5407363826095	123890828.41393518	86.97305	21.837437	4.1240892
5407363826128	123890828.52555823	83.40073	19.771587	1.6680022
5407363826168	123890828.6829524	82.25036	19.22003	4.7649446
5407363826383	123890829.53362775	83.18322	22.008213	0.7920148
...	...	...	...	...

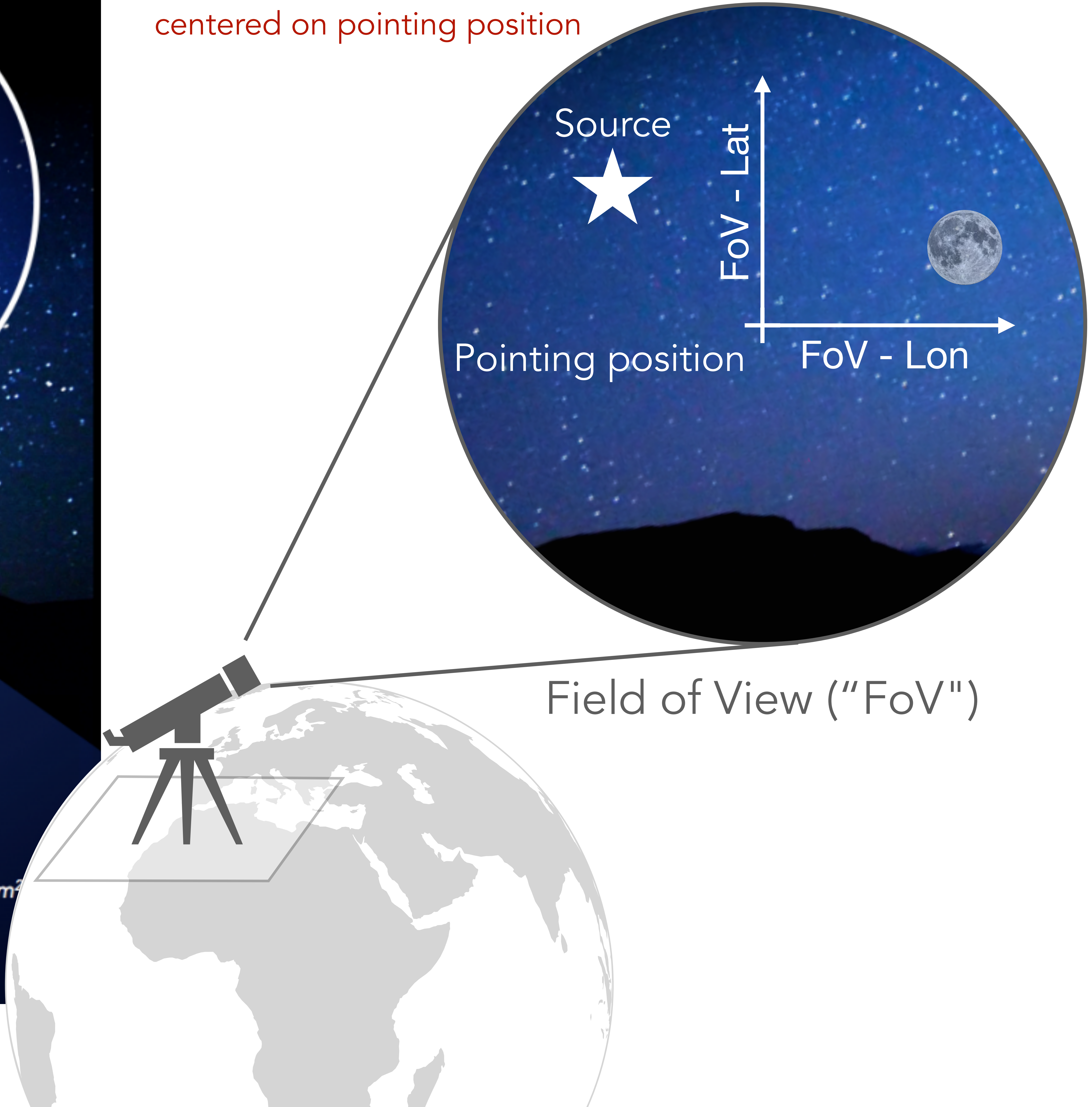
IACT Observation

"FoV coordinates" and conventions



"How CTA Detects Cherenkov Light", taken from www.cta-observatory.org

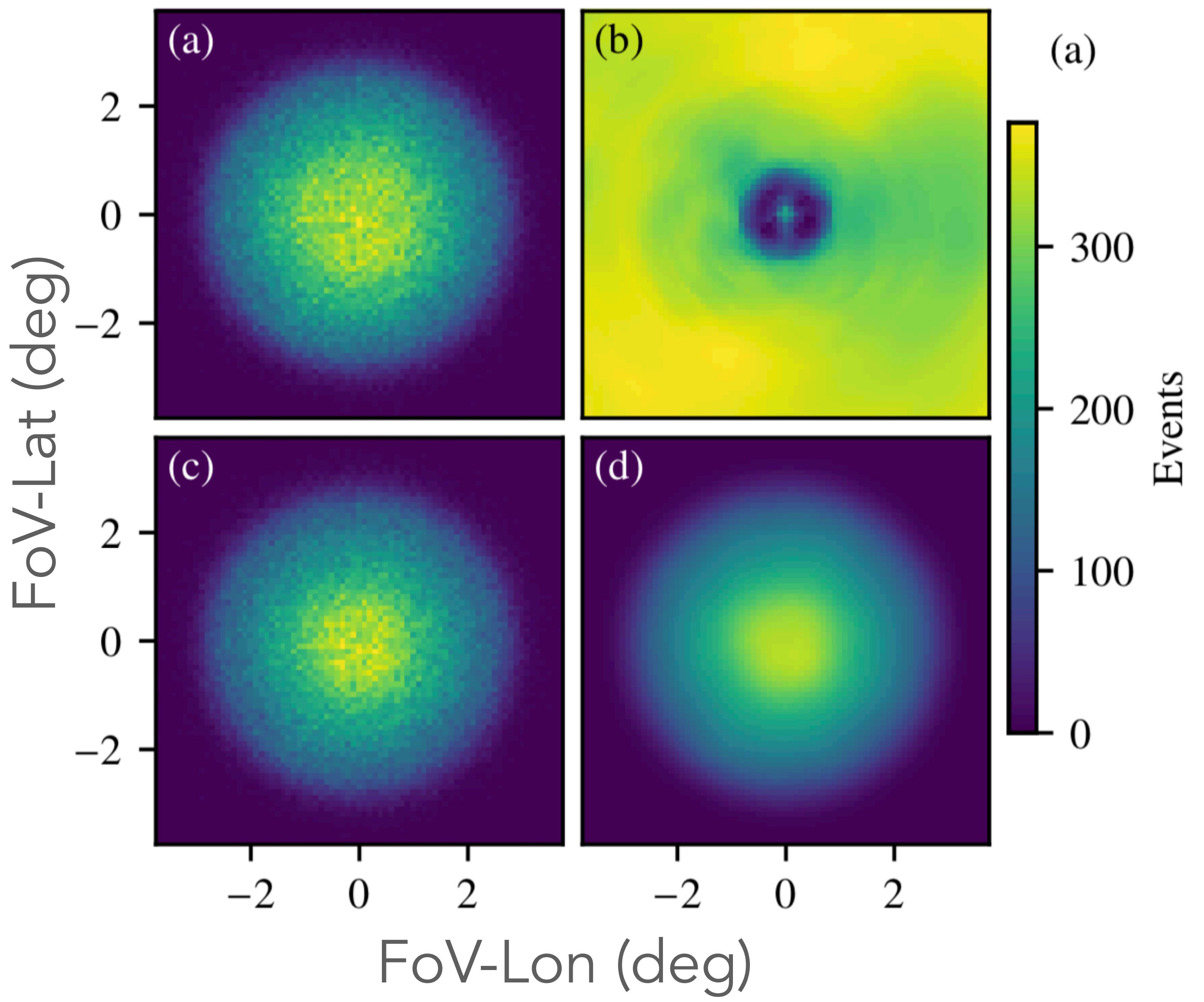
Aligned with horizontal ("Alt-Az") system, centered on pointing position

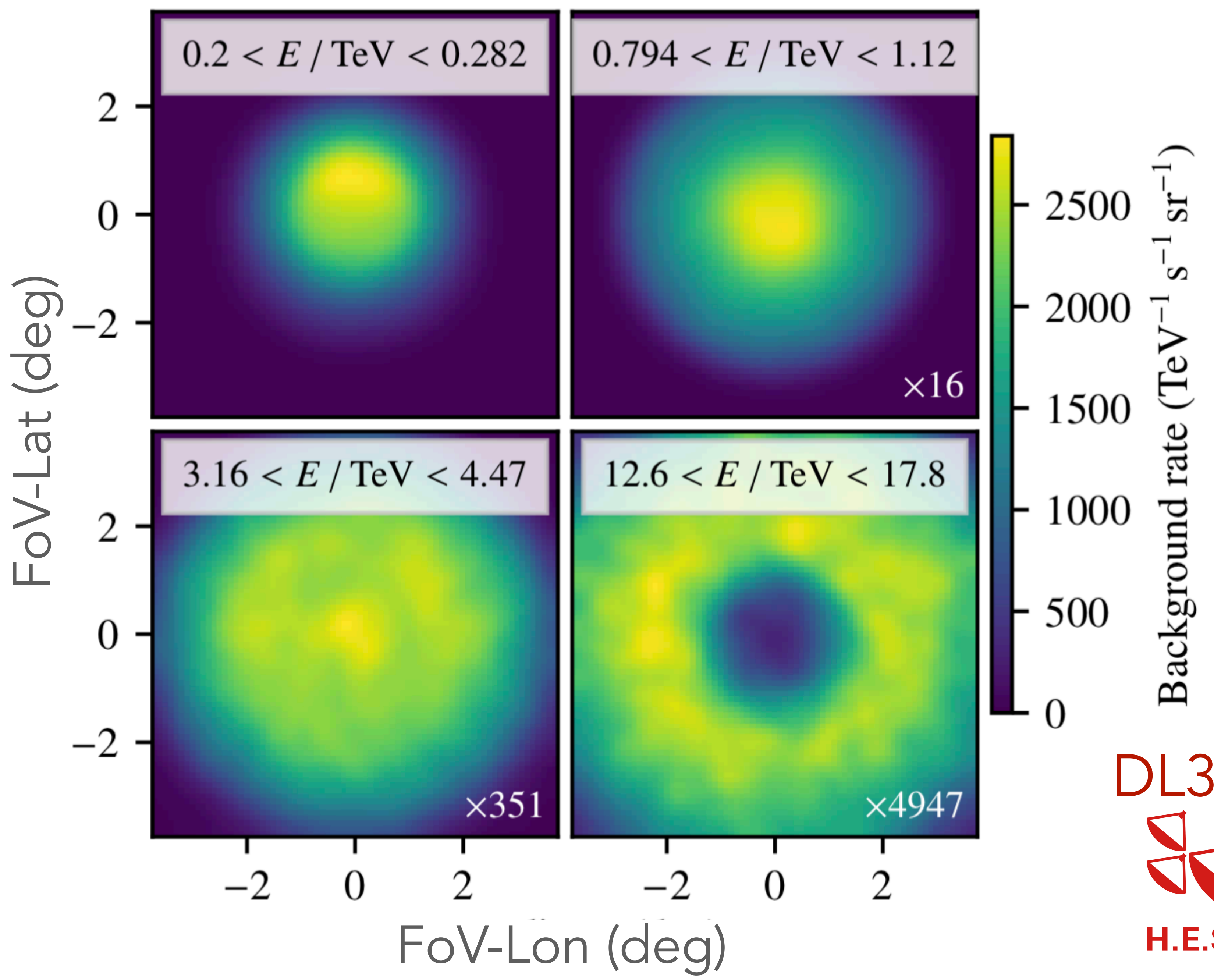


Instrument response

Residual Hadronic Background

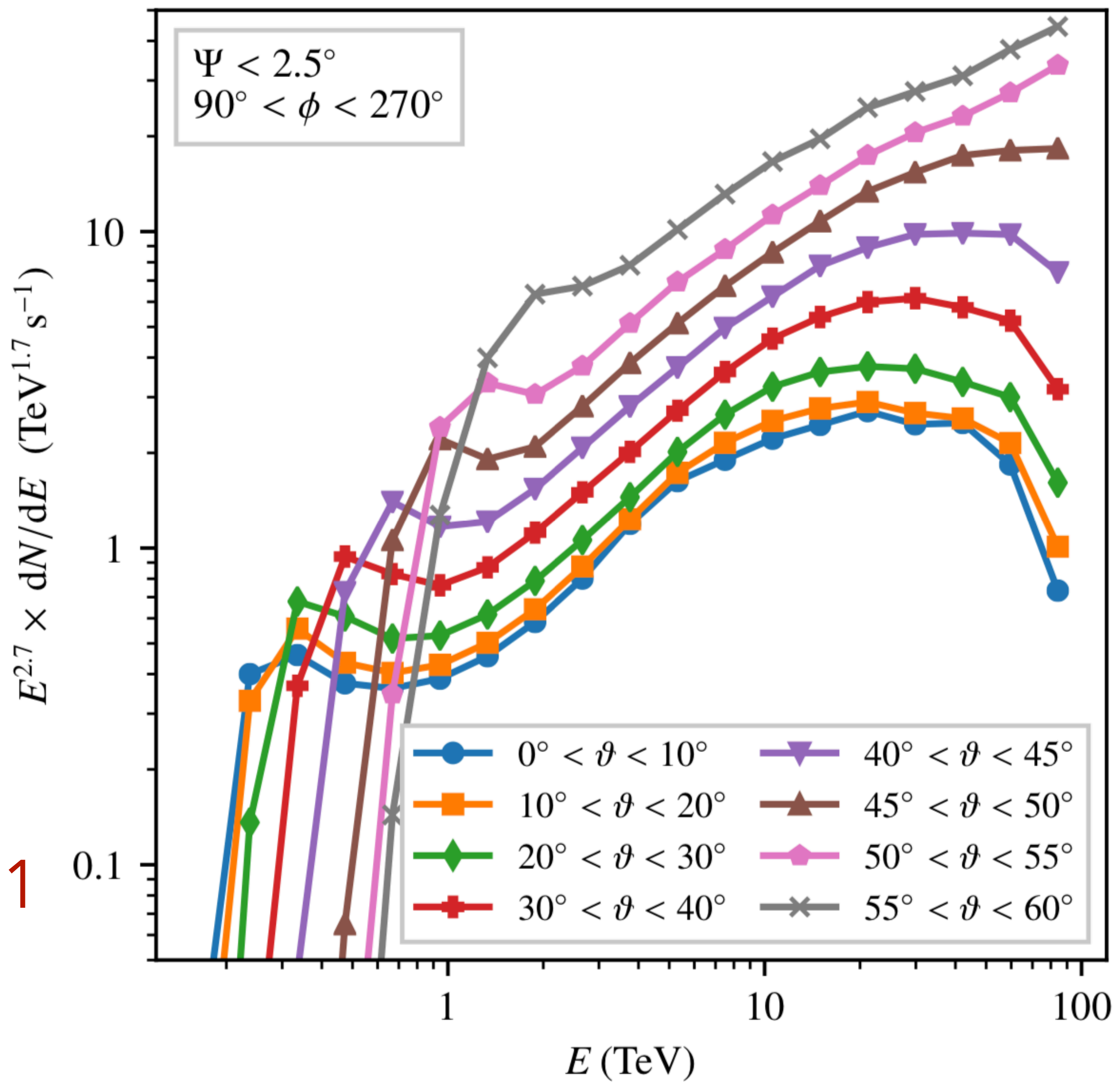
DL3 γ-like events → DL4 Binned data → DL5 Science products





DL3 DR1

H.E.S.S.

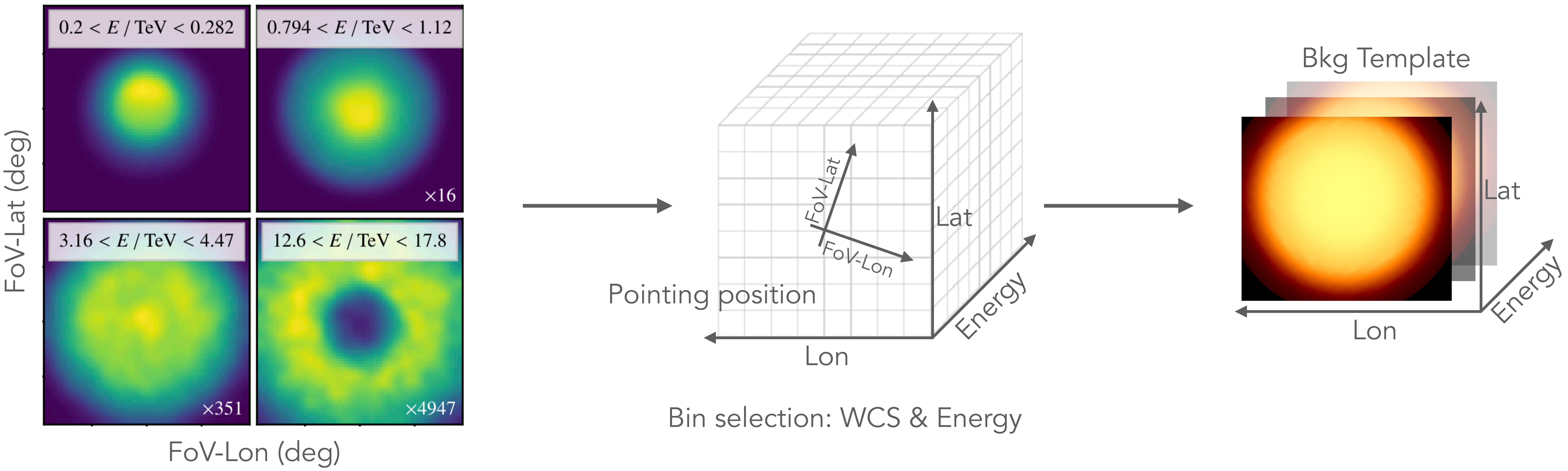


- Background model (sometime called “acceptance”…) derived from many “OFF” observations
- Instrument is pointed to an “empty” (without known gamma-ray sources..) region in the sky
- Done many times observations are grouped with similar observation conditions, mostly number of telescopes and bands of zenith angle
- Taken from: L.Mohrmann et al. https://ui.adsabs.harvard.edu/abs/2019arXiv191008088M

Data reduction
Background

DL3 → DL4 → DL5

γ-like events · Binned data · Science products



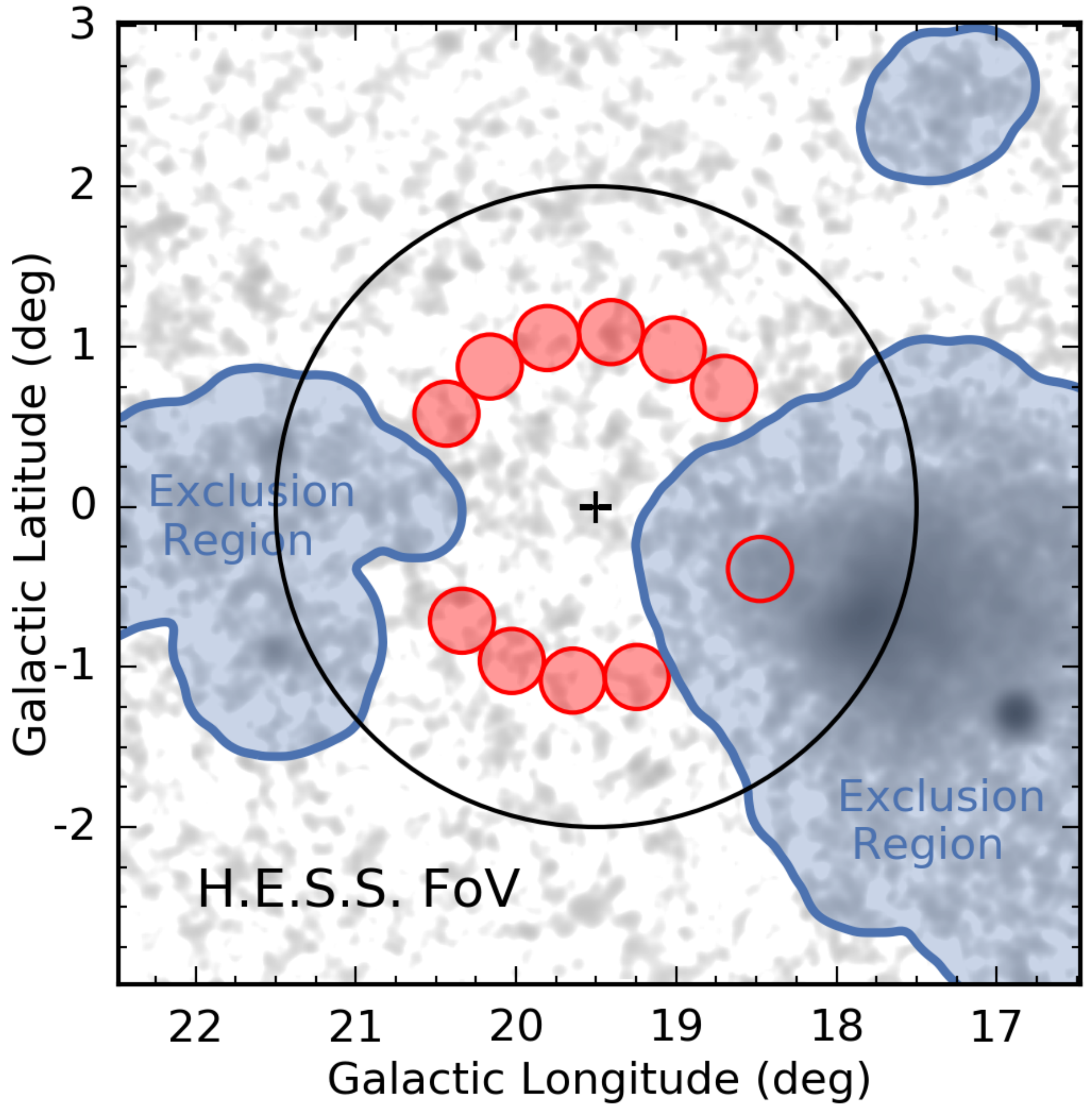
- Background rate (events / s / sr / TeV) is reprojected on the sky and multiplied with solid angle, observation time and integrated in energy
- For the later analysis the "template" is combined with a parametric correction such as norm and spectral "tilt"

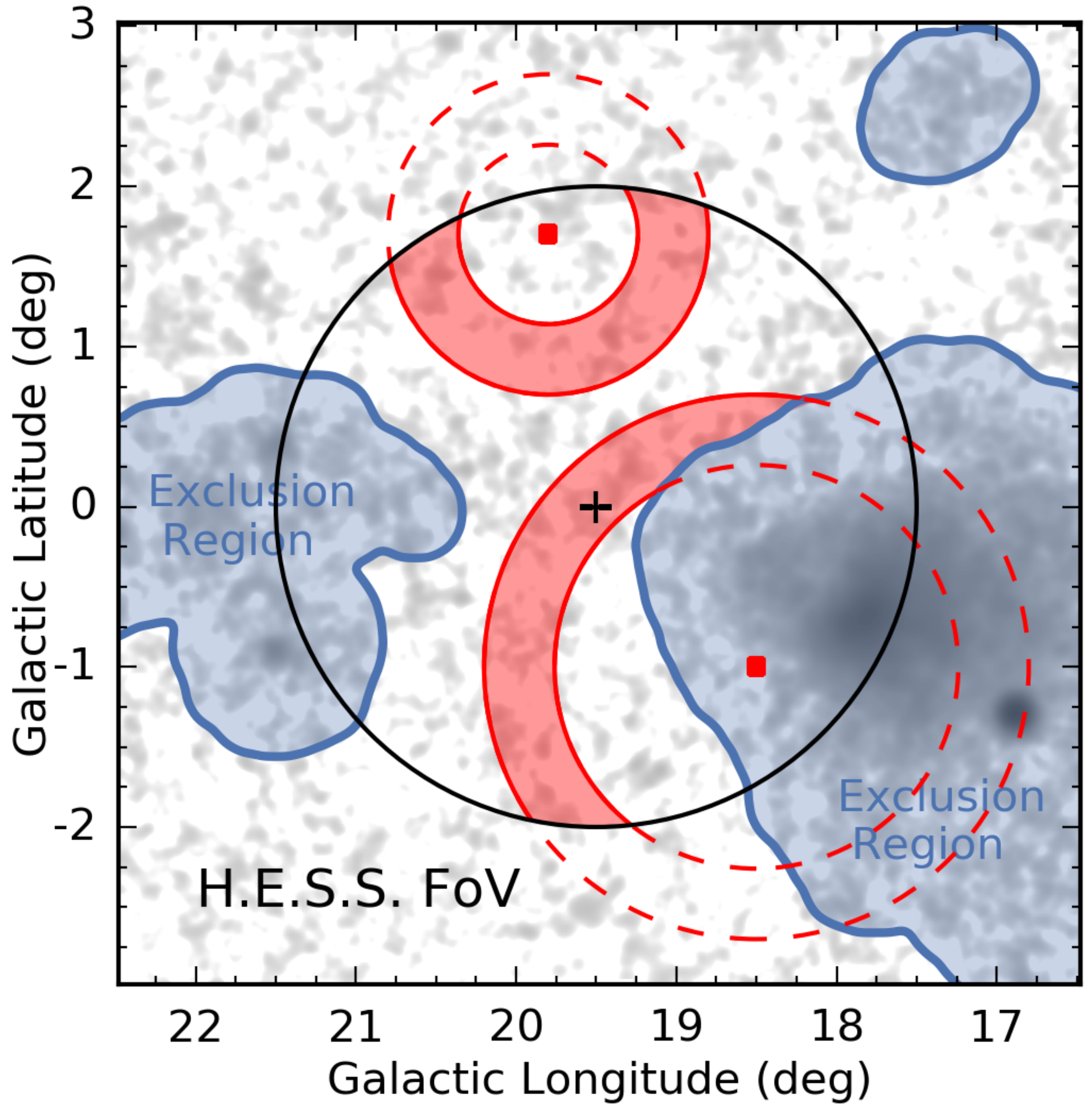
"Classical" Background

Measure "Off" counts from data

DL3 → DL4 → DL5

γ-like events — Binned data — Science products





- "Reflected regions": used for spectral analysis, find "off regions" reflected from the pointing position, ignore exclusion regions
- "Ring background": mostly for image analysis, derive "off counts" per pixel from ring convolved counts, ignore exclusion regions. There is a variant called "adaptive ring", while enlarges the ring to achieve better statistics.
- See e.g. Berge et.al. https://ui.adsabs.harvard.edu/abs/2007A%26A...466.1219B/abstract