Open analysis libraries for the high-energy astrophysics

B. Khélifi[†] (APC, Paris)

With inputs from R. Terrier (APC), Gammapy Lead Developers

11th International Fermi Symposium (College Park, Maryland)

[†]: Gammapy Project Manager, VODF convener, one of the HESS DL3 release responsibles, SWH ambassador, etc

HE/VHE γ-ray data analysis

HE astrophysics

Usage mainly driven by HEASARC

- Open format (OGIP)
- Open data
- Open software

Initially, for the X-ray satellites

`FITSIO, FTOOLS`

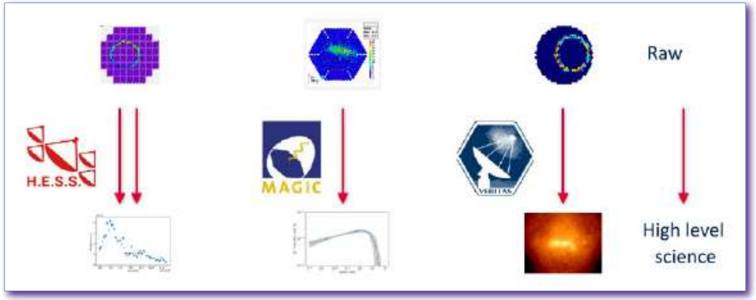
`XANADU, HEASoft, Sherpa, CIAO, XMM-SAS`

Then, for Fermi

`Fermitools, Fermipy, ...`

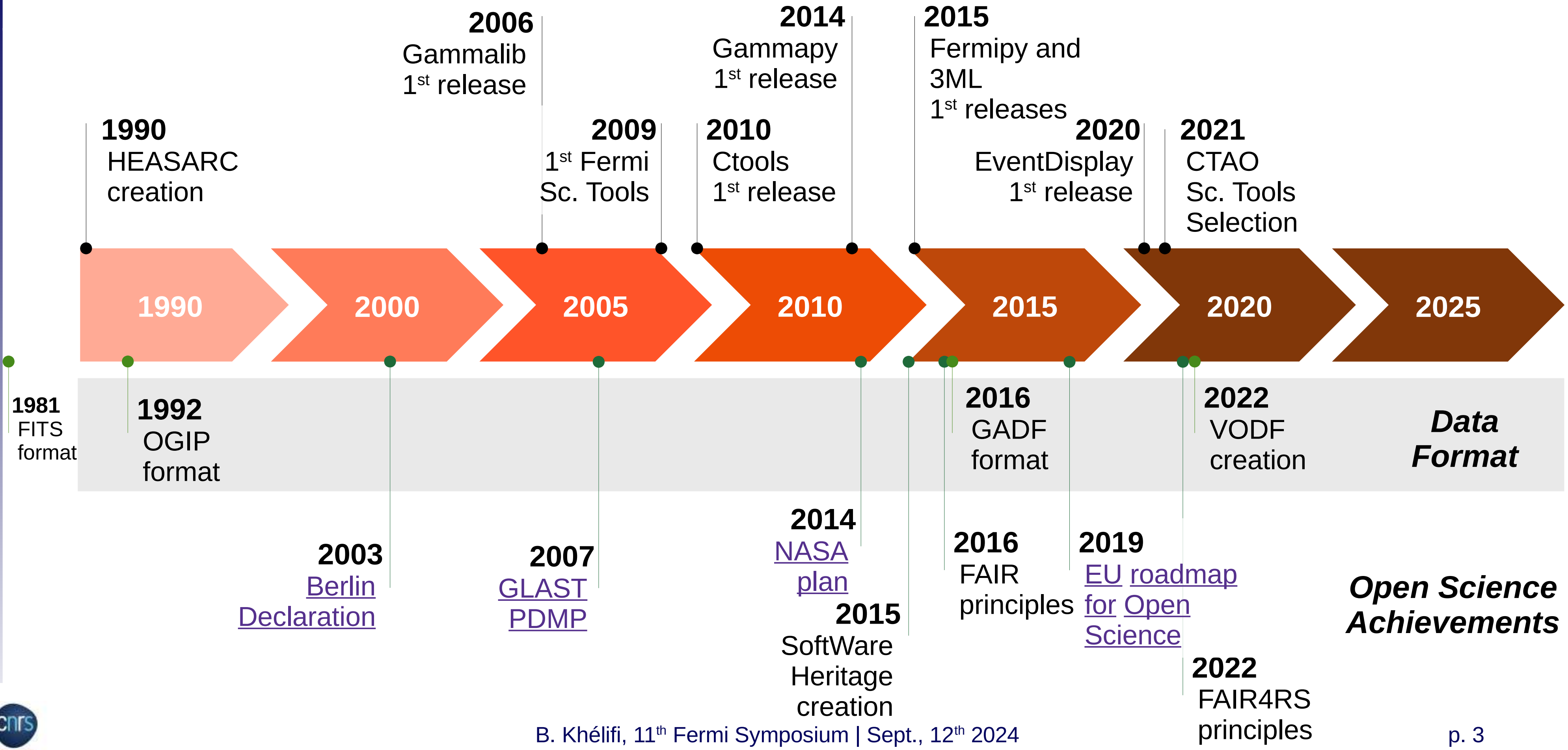
VHE instruments

Proprietary data and formats, and closed software



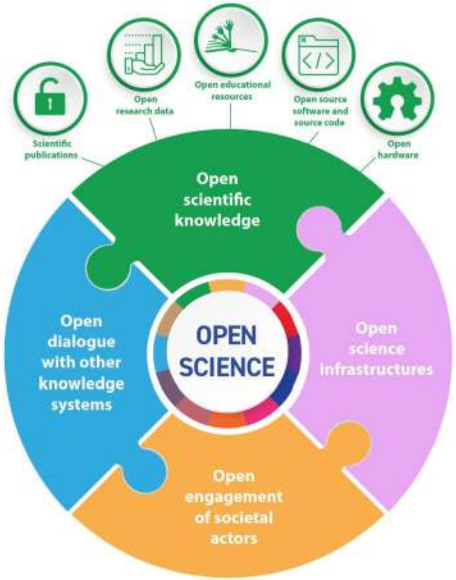
The Fermi driver, the frequent usage of open multi-λ data, the development of the Python ecosystem lead to change of model...

Retrospective on HE/VHE software

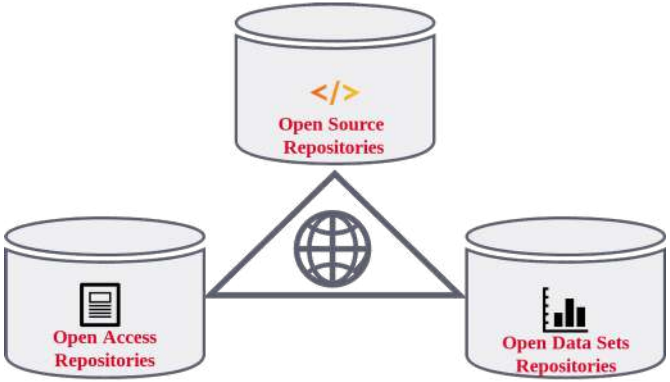


The emergence of Open Science

« Open Science – Making science more accessible, inclusive and equitable for the benefit of all »
UNESCO



The Open Science pillars



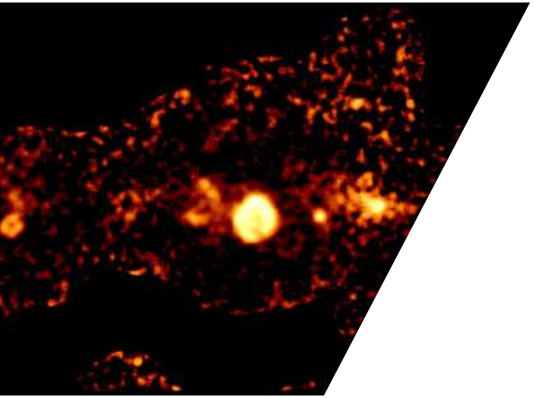
Open software are difficult to separate from open data...

Summary

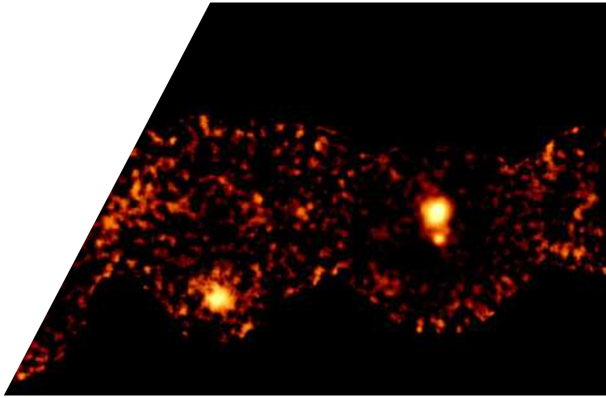
1- Data formats

2- Main open analysis libraries

3- Towards joint multi-instrument analyses:
the interoperability



Data formats

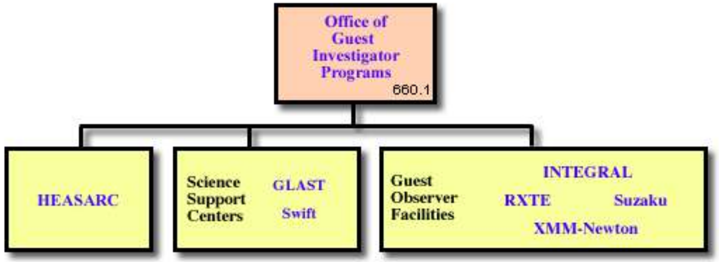


X-rays → GeV: OGIP

Office of Guest Investigator Programs conventions

15 specific format recommendations (link) Corcoran *et al.*, 1995 **ADASS**

- None on the final astrophysical products



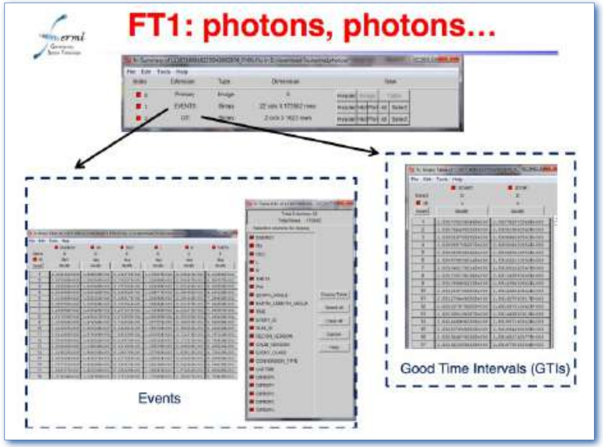
- **Event list:** well described, except errors, no detailed metadata
- **Instrument Response Files** (IRFs):
  - PHA (XSPEC compatible): ie RMF+ARF, ie Edisp+Area, includes statistical errors and systematics, link to the background file, GTI, metadata (e.g. creation, history)
  - PSF (radial or 2D): includes statistical errors, metadata (e.g. creation, history)

Name	Units	Description
TIME	\`s' or \`d'	The time associated with the event
RAWX	\`pixel'	Raw telemetry X position of the event
RAWY	\`pixel'	Raw telemetry Y position of the event
DETX	\`pixel'	Linearized X position of the event on the detector
DETY	\`pixel'	Linearized Y position of the event on the detector
X	\`pixel'	Projected X position of the event on the sky
Y	\`pixel'	Projected Y position of the event on the sky
PHA	\`chan'	Pulse height analyzer' energy channel

HE γ-rays: Fermi-LAT

Formats regulated very early, and afterwards updated

- Project Data Management Plan (PDMP), 2007 (link)
- Science Data Products File Format Document (FFD), 2019 (link)



Definition of Data levels (L0 → L3)

- like for X-rays (HEASARC continuity)
- (raw data)
- **L0**: 'cleaned raw data'
- **L1**: starting point for scientific analyses
- **L2**: result of science analysis tools
- **L3**: catalogs or compendia of DL2 data, including e.g. flux history, sce identification

In addition, **ancillary data**: diffuse galactic interstellar, extragalactic emission models, pulsar ephemerides

Data format

- based on OGIP + specificities (e.g. interstellar emission model, BAT data, LAT LLE)
- specific format for L2 (LC, spectrum) and L3 (catalogs)
- Metadata: s/w name and version, but not the release name, no provenance in data (but in web pages)

Almost FAIR
Light compliance to IVOA standards

VHE γ-rays: GADF

Up to mid-10's, VHE community worked in a totally competitive and closed mode

- All was private
- Except few MoUs around scientific projects

Better results
Interoperability between instruments
Respect of the FAIR principles

Some 'dreamers' worked towards the opening of the VHE astrophysics

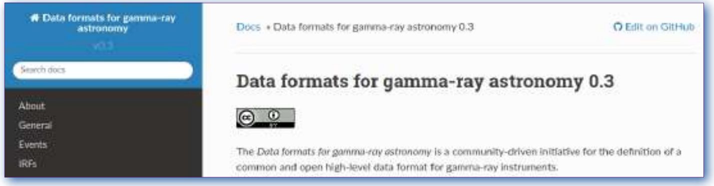
Deil, C., *et al.*, ASTERICS 2016 (link)

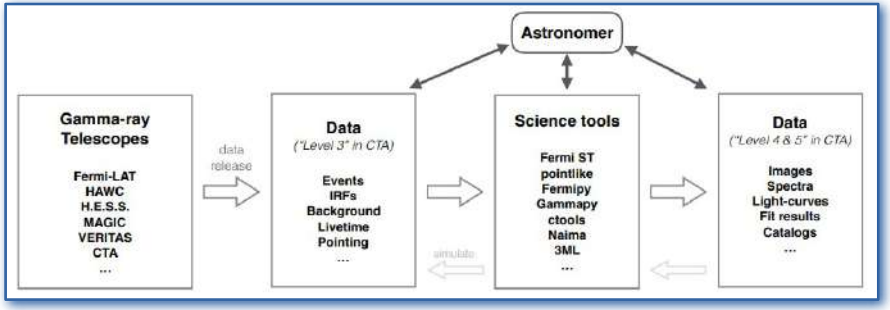
- Data format standardization: open initiative 'Open Gamma-Ray Astro'
- Open Science Tools: Gammapy

The Gamma Astro Data Format

DOI 10.5281/zenodo.7304668

- Strongly influenced by the Fermi-LAT format (and OGIP) and serialization into FITS
- Same type of data levels: DL3, DL4, DL5
- Full description of the DL3: event list and IRFs



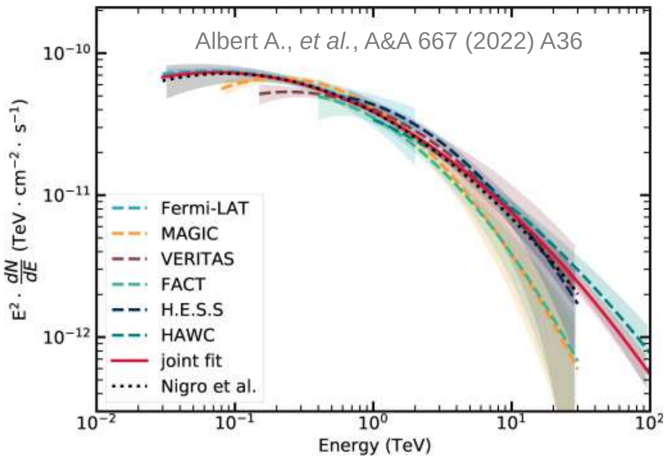


Deil, C., *et al.*, Proc. of Gamma 2016

VHE γ-rays: GADF

Proved to be well suited for VHE needs and can serve as standard

As demonstrated by the joint Crab analysis with `Gammapy`:



Or with astrophysical papers

HGPS, Abdalla H., *et al.*, A&A 612 (2018) A1

But some drawbacks appear

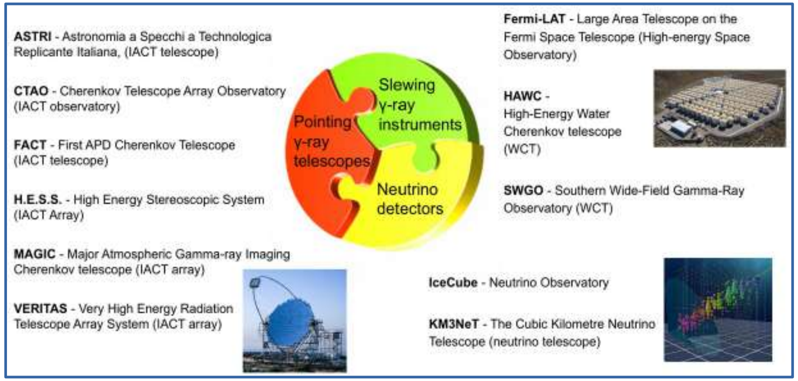
- GADF had no clear organizational structure
- No clear resolution of contentious issues
- No clear roadmap
- Not FAIR enough, no clear standards for DL4+

→ Hand-over to VODF

VHE data: VODF

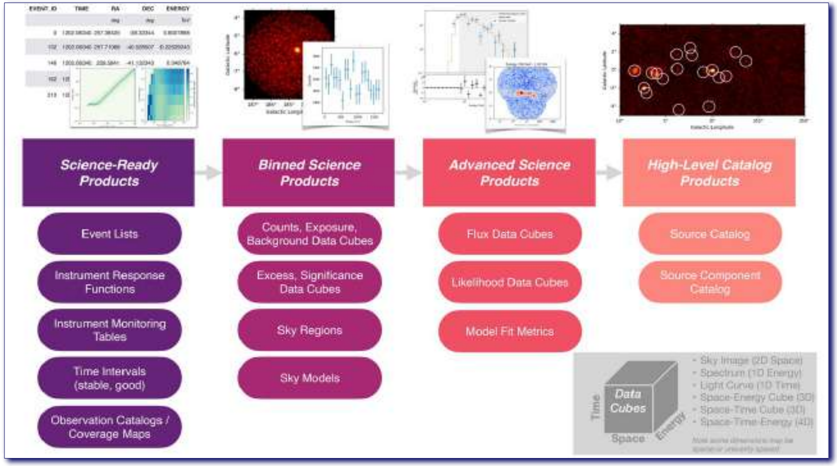
Open Initiative 'Very-high-energy Open Data Format' (link)

- Aims to format VHE data (gamma and neutrino)
- Officially supported by 11 experiments



- Structured with a project organization

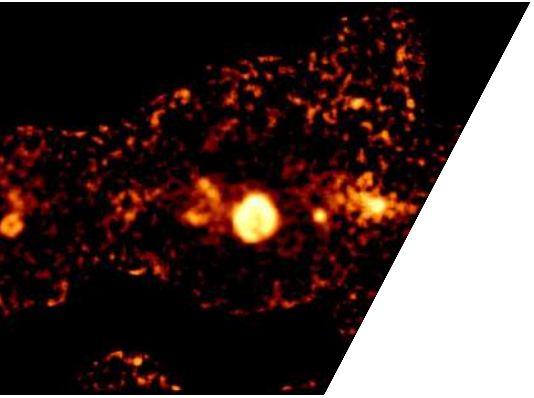
Coordination Committee, Conveners: R. Zanin, B. Khélifi
Lead Editors: K. Kosack, L. Olivera-Nieto, J. Schnabel



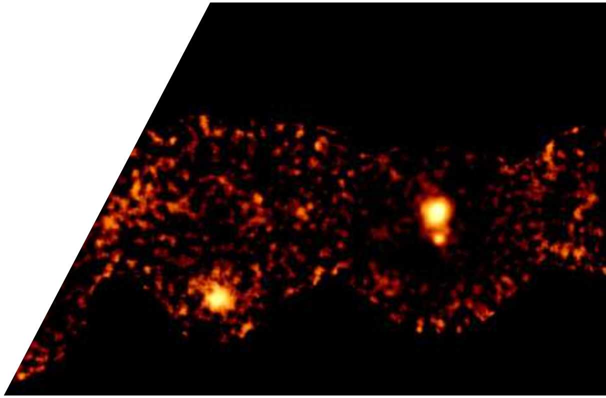
Khélifi, B., *et al.*, Proc. of 38th ICRC (2023)

```
/*************************************************************************
/* FITS FILE:
/*    VODF Level-1 Event Data
/*
/* EXTENSIONS SUMMARY:
/*  IDX  NAME               VER CLASS                TYPE
/*  ---  ------------------ --- -------------------- ----------------
/*    0. EVENTS               0 OGIP.EVENTS          [TableExtension]
/*    1. SOI                  0                      [TableExtension]
/*************************************************************************

/ #######################################################################
/ HDU: EVENTS
/ DESCRIPTION:
/    VODF Level 1 Event List
/ #######################################################################
XTENSION = BINTABLE
EXTNAME  = EVENTS
EXTVER   = 0
```



Main open analysis libraries



Data from counting detectors

HE and VHE experiments detect events

Two analysis approaches:

- **Unfold data** and fit sources model to astrophysical quantities
  - Unfolding makes strong implicit model assumptions
  - End products with uncertain probabilistic distribution
  - Final χ^2 fit leads to large bias even in the high-count regime

Humphrey *et al.*, 2009 **ApJ** 693 822

- **Forward-fold** sources model
  - Clear model assumption
  - Correct treatment of the statistics
  - Possibility to handle instrument systematics

Common algorithms for the libraries: Poisson Log-Likelihood

C-statistic or Cash statistics – bin or unbinned

Cash W. 1979 **ApJ** 228 939
Mattox J.R. *et al.* 1996 **ApJ** 461, 396

HE vs VHE analyses

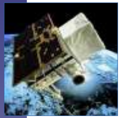
Many similarities

- Forward-Folding in the Poisson regime
- IRFs can be factorized in the similar way
- Use of background models (IRF)
- Need of astrophysical background models

And some differences

- Rare are the use cases of unbinned Log($\mathcal{L}$) analyses for VHE
- For asymmetric VHE arrays or with high geo-magnetic fields, the IRFs have not any more an azimuthal symmetry
- Non-calibrated γ-rays IRFs
- Uncertain VHE CR-background model
- Uncertain VHE pp astrophysical model
- Atmosphere not stable

Global view


Software Name	Official tool of experiment	Supported data format(s)	Able to analyse data from	Language	Associated projects	Interoperability
Fermitools	Fermi (LAT/GBM)	Fermi	Fermi	C++	fermimodels, FermiPy (Gammapy)	–
AGILE Science Tools	AGILE	Private AGILE format	AGILE	Python, C++	–	–
EventDisplay	VERITAS	Private VERITAS format	VERITAS	C++	–	–
Gammapy	CTAO, H.E.S.S.	GADF, OGIP	VHE, UHE, Fermi-LAT, OGIP-Lile	Python	–	Python, Naima, gammapyXray, agnpy, JetSet, Gamera, XSPEC table models
ThreeML	–	OGIP + plugins to support different formats	With additional plugins: HAWC, Fermi-LAT*, OGIP-like, VERITAS, IXPE, Optical telescopes	Python	astromodels	Python, XSPEC models, Naima
gammalib/ctools	–	GADF, OGIP	COMPTEL, Fermi/LAT*, VHE, INTEGRAL/SPI	C++, Python	–	XSPEC models

Data aggregation and reduction

From L1/DL3 to L1.5/DL4

- Data are found, accessed and selected using user's choice
- Selected data are aggregated, IRFs projected, etc

Science tools methods

- **Very specific to an experiment**: importance of the documentation
- **Very sensitive step**, because of the IRFs transformations
- Fermitools and ctools: command lines
- Gammapy and Fermipy: Jupyter nb or YAML configuration files
- 3ML: install the plugin and the associated code (FermiPy, Gammapy, AERIE), learn the methods for this code, and realize the data reduction

Can be quite technical
Can require high computing resources

→ { Schools
Hands-on

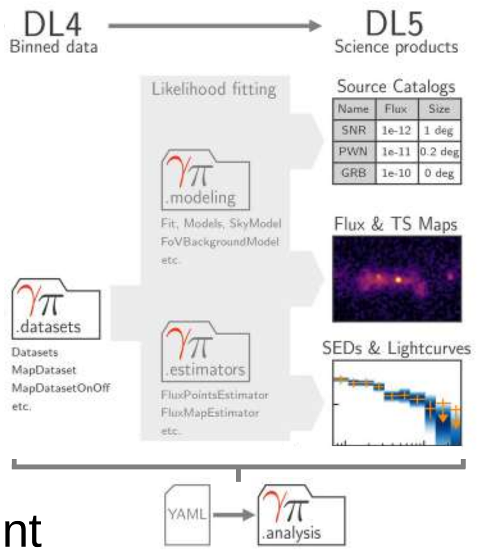
Modelling and fitting

From L1.3/DL4 to L2/DL5

- Description of all sources and background models
- Choice of the fitter, fitting, results inspection

Science tools methods

- **Models** quite different between libraries
  - Non-uniform description
  - Temporal, polarization or user-defined models not always present
  - Templates have their own description
  - Features on priors different
  - Features on operation/arithmetic different
  - Different serialization (w/ or w/o units, XML/YAML)
  - They are not FAIR!

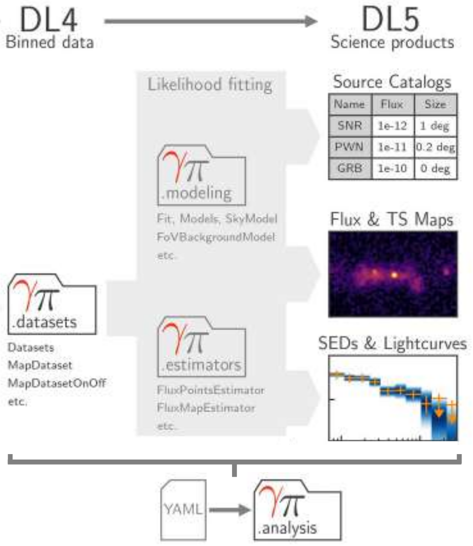


Modelling is a real issue
And even in IVOA!

Modelling and fitting (2)

Science tools methods

- **Fitters**
  - Different methods offered by the libraries
  - Fermitools: 5, FermiPy: 3, ctools: 1, Gammapy: minuit, sherpa (4), scipy (15), 3ML: 9
  - 3ML offers Bayesian samplers
  - Gammapy and 3ML offer Wstat in addition
  - Gammapy does not offer unbinned likelihood
  - 3ML&ctools can not fit extended sources with the Fermi plugin



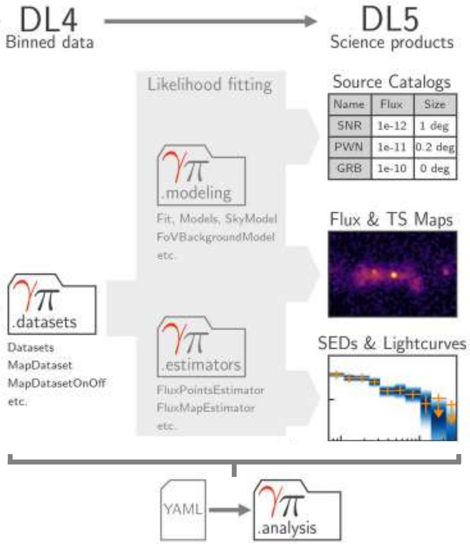
- **Results inspection**
  - Because the goodness of fit is not enough, different graphical methods are offered
  - Fitter options and status not serialized
- **Results serialization**
  - Large variety of serialization (w or w/o likelihood profiles, formats)
  - They are not FAIR (no Provenance, data reduction ignored, etc)

Can require high computing resources (CPU, mem)
Issues associated to the d.o.f.

Modelling and fitting (3)

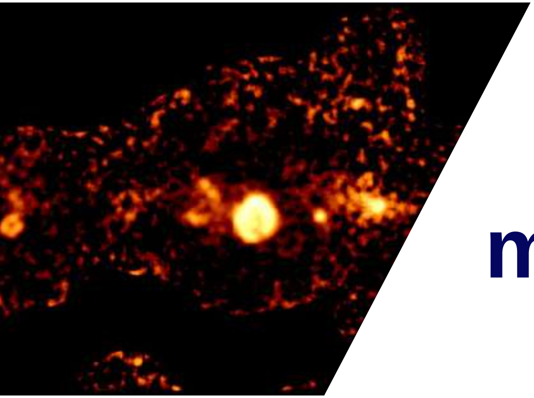
Generation of astrophysical products

Beyond the modelling and fitting results, one has also light curves, maps, time domain quantities (PSD, excess variance, etc), flux ‘points’, spectro-morphological properties, etc

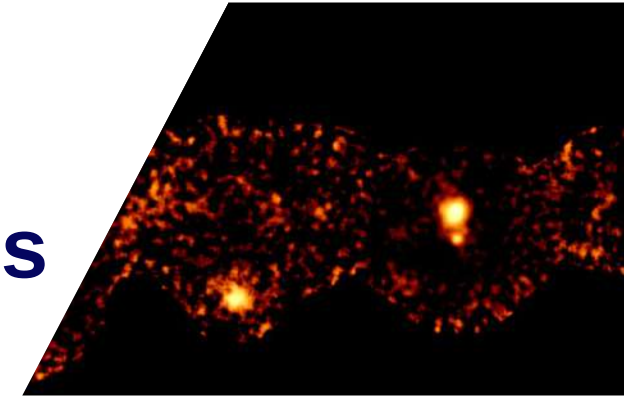


- **Dedicated methods**
  - Inside the libraries or inside their ecosystem
  - Large variety → exploration of the documentation
- **Results**
  - Badly serialized (w/ or w/o units, format)
  - Issue of reproducibility
  - They are not FAIR!

No data model in IVOA:
Need to invent our HE standards



Towards joint
multi-instrument analyses

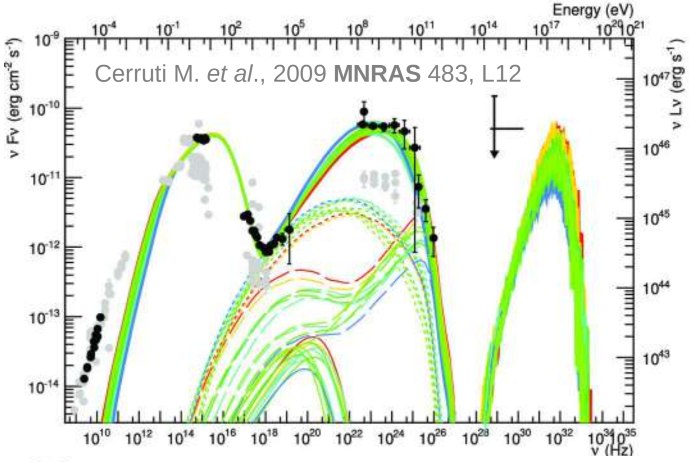


Multi-wavelength and multi-messenger astrophysics

HE astrophysics

- Provision of more open data with increasing quality: photons, neutrinos, GW, CRs
- Leads to more and more MWL & MM studies

Increased community expectations towards the analysis tools



(b) Lepto-hadronic modeling of TXS 0506+056

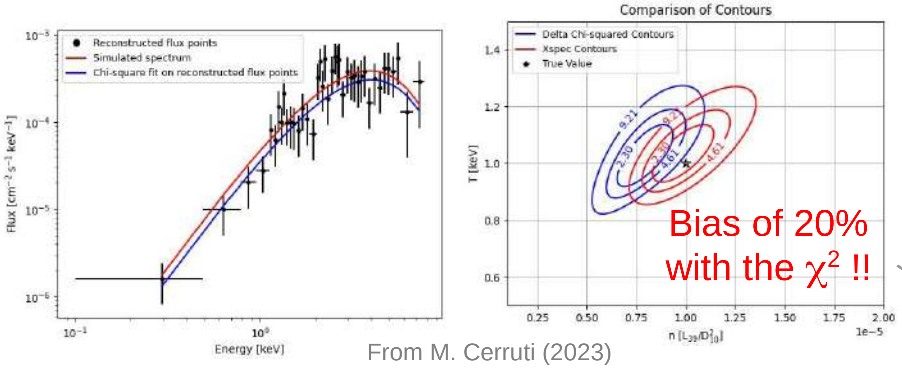
Open science tools can offer the feature of multi-instrument fit

- Even if beyond the scope of instrument specific software
- The IRFs of the HE/VHE instruments can be identically factorized
- Such libraries handle correctly statistics (in contrary to basic χ^2 fits)

Challenges for multi-instrument analyses

Rigorous data analyses

Ex: simulation of a BB seen by Swift with XSPEC



From M. Cerruti (2023)

Proper combination of the dataset statistics... and profit of the experiment's accuracy!

Humphrey *et al*., 2009 **ApJ** 693 822

- And uses of priors to deal with possible systematics between instruments

Open standards on data format

- Need of very well documented, following FAIR principles
  - Use of community standards (OGIP, GADF)
- Otherwise, needs plugins (ex: 3ML created `HAL` to use HAWC data)
- If formats have limitations, needs adapters (ex: Gammapy created `gammapyXray` to handle the limitations of the OGIP format)

DOI 10.5281/zenodo.7092736

Challenges for multi-instrument analyses

Software interoperability

- To **add features**: the Python eco-system helps a lot !
  - Ex: Phase computation (`pint`), multiprocessing (ex: `ray`), AGN time lags (`pyPETaL`)
  - Community-driven projects are in the core of the Open Science
- To **use astrophysical models of source**
  - Gammapy can be used with `Naima`, `GAMERA`, `AgnPy`, `JetSet` for the source models
  - Still need format standards for predictions from heavy theoretical models (3DRMHD)

Workshop on numerical multi-messenger, 2024 (APC) - link

Need of increased computing resources

- **Speed**: multi-core, clusters, GPU
  - Libraries offer that, but it is not enough! Gammapy is evaluating the use of `JAX`
- **Memory**: avoid the saturation of the RAM
  - Making joint fit with many instrument datasets is very demanding
  - Need memory-saving of internal data (caching, sparse arrays, generator, AoB)
  - Ex: Use of CNN trained on model predictions to reduce template's size

Sahakyan N. *et al.*, 2024 **ApJ** 971 70

Challenges for multi-instrument analyses

Software interoperability

- Development of **science analysis web platform**
  - Web interface to allow MWL & MM data browser, and sometime data analysis or data fitting with astrophysical models

  - ASI: Space Science Data Center (link) - access to data and catalogs
  - ISDC/CNRS: MMODA (link) - access to data and L1/DL3 analysis
  - ICRANet Armenia: MMDC (link) - catalogue access and SED reconstituion
  - HEASARC:
    - SciServer (link) - access to HEASARC data and data analysis with HEAsoft (radio → Fermi)
    - HERA (link) - access to HEARARC data and X-ray data analysis
    - VO services - catalogs, images, spectra (infrastructure services: registry, validation, monitoring)

Jaffe et al., 2022 IVOA (link)

  - Leicester Univ: Swift-XRT product building
  - Etc

Challenges for multi-instrument analyses

Software interoperability

- Development of **science analysis web platform**
  - Library to download and run on any system/cluster
    - Need easy multi-processing configuration (`fermiPy/gtapps_mp`, `ctools`, `3ML`, `Gammapy` have that)
  - Data to download : VO interface by the platform or the science tools)
    - `astropy/pyvo`, `astroquery`
    - But HE IRFs are not foreseen in VO data models → Creation of an High Energy Interest Group (HEIG) In next IVOA/InterOp meeting (Nov., Malta)?
  - Configurable data analysis : using YAML/XML files
  - Creation of FAIR astrophysical data
    - And are the produced data understandable? Reproducible?
    - Analysis software need the Provenance (IVOA link)

Current libraries should respect more the FAIR4RS principles

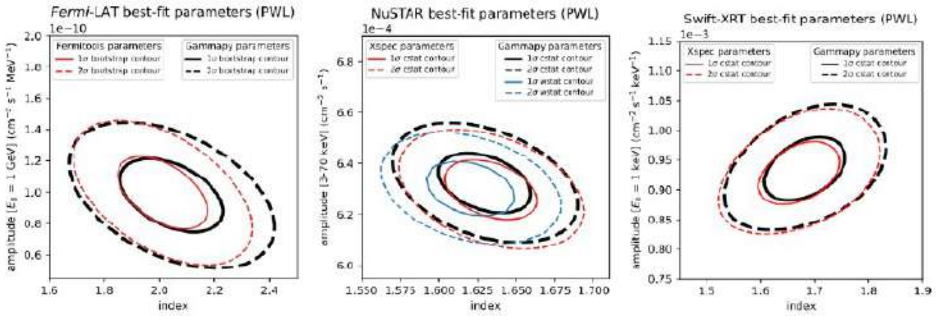
An example of achievements

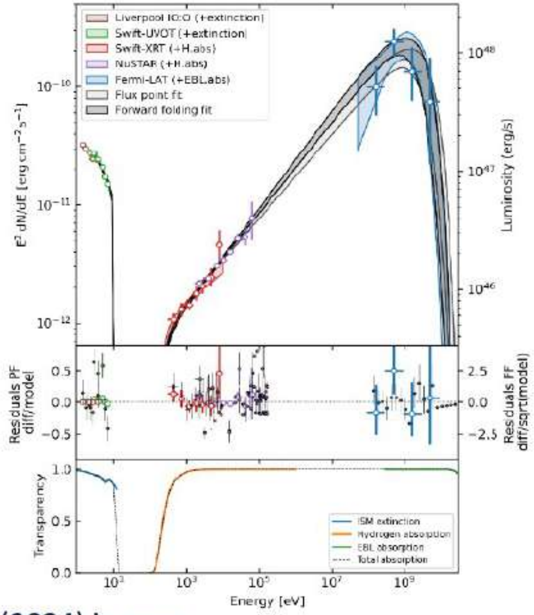
Multi-instrument analysis over 10 decades

OP 313 campaign

- Joint forward fitting fit from eV to 10^{10} eV with Gammapy:
  Liverpool OT, Swift-UVOT, Swift-XRT, NUSTAR, Fermi-LAT

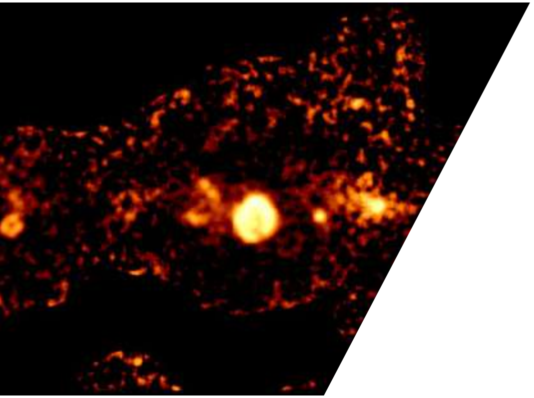
- Flux points lose some stat. information (e.g UL)
- Full forward fit provides more accurate results
- Gammapy facilitates the distribution and reproducibility of the results



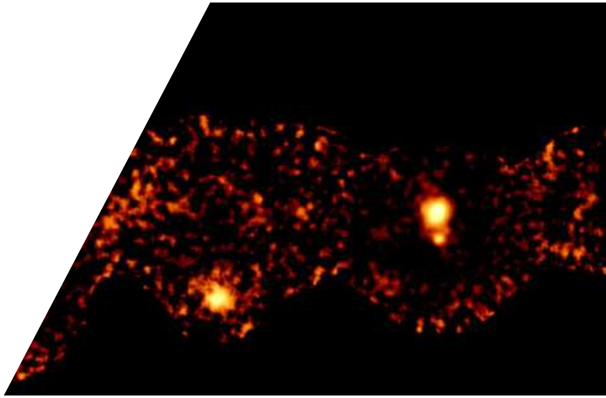


From M. Nievas et al (2024) in prep.

From R. Terrier, Gamma2024



Conclusions



Personal summary

The major impact of HEASARC and Fermi coll.

- Creation of data format standards
- Creation of open science tools
- Willing to participate to shared projects (technical and scientific)

Impact of the Open Science in the HE domain

- Many **community driven projects** and with an open development scheme
  - On HE γ-rays software, data formats, multi-purposes libraries (ex: `astropy`)
  - The Python revolution had a major impact
- **Open research software**
  - Both the hardware and the software render successful an experiment
  - Finding resources
    - User training, and developer formation
  - Valorization, recognition and evaluation are topics of the Open Science
    - Cite properly the software (and not only their paper)
- **The challenges of multi-instrument analyses**
  - Data formats, handling the need of increased computing resources (CPU/GPU, mem), reproducibility
  - Community should support the data scientists in their carrier

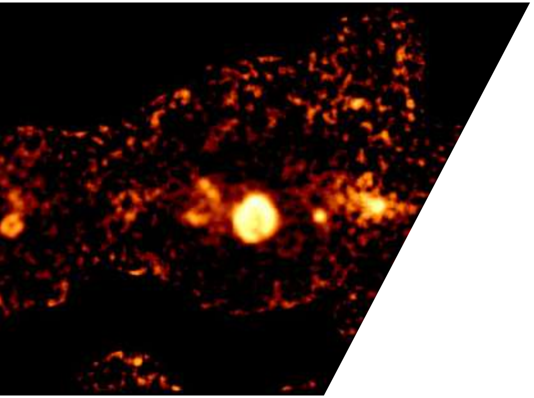

PIG 24 - Authorship policy

Abstract

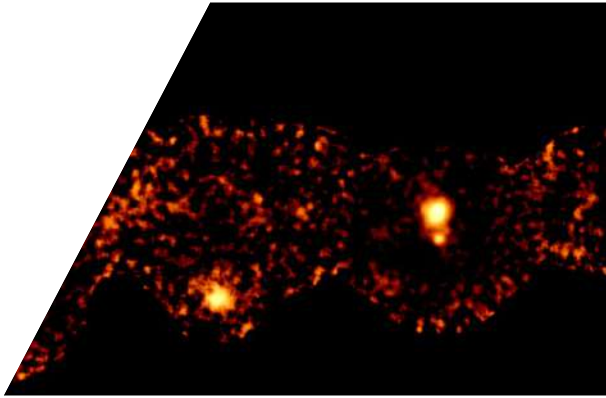
Thank you for your attention

Fermi Symposium 2024


A view of the Universe



Annexes



Poisson Log-Likelihood

Common algorithms for the libraries: Poisson Log-Likelihood

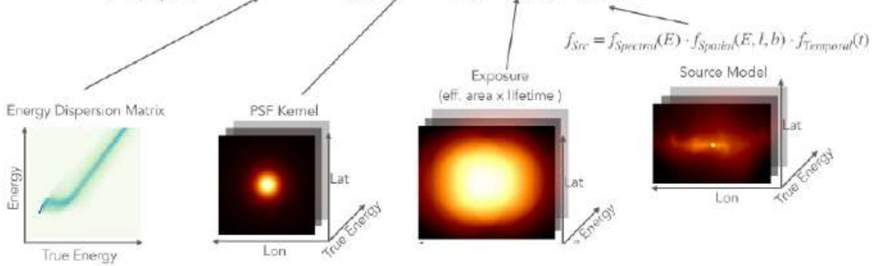
"Cash statistics": summed over all "bins"

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

i: spectral channels or 3D voxels

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

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



- Bins in the spectral, spatial, temporal domain

→ Need of a "global" background model template with "correction parameters"

→ Need of the "signal" IRFs and source models

Most of the time,
Identical factorization of the IRFs
for X-rays → UHE & neutrino exp.

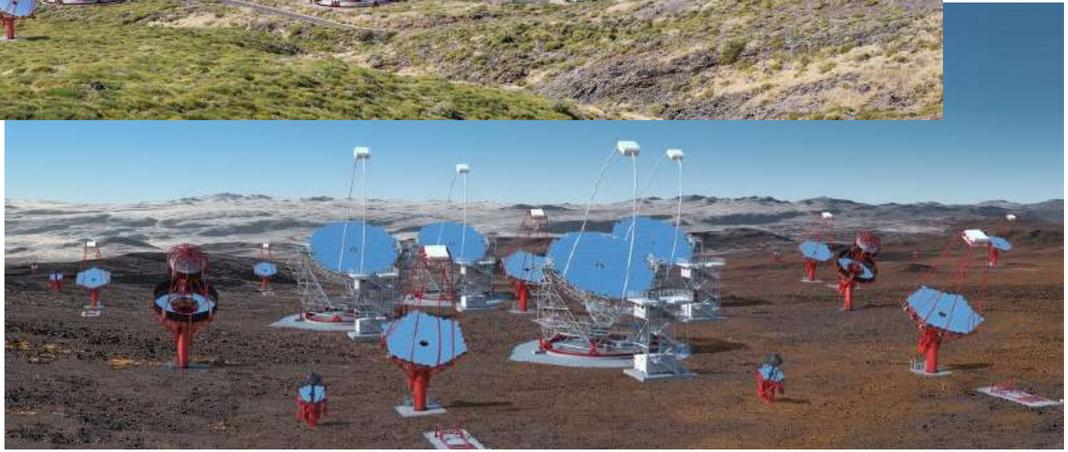
Cherenkov Telescope Array

Gamma-ray observatory

- First Open VHE observatory
- O(10)GeV → 200TeV
- PSF: 0.3° → 0.02°
- Observations with FoV of 3°-10° during dark mights

Full sky observations

Under construction



© CTAO

Array of the Northern site (La Palma)

- 9 MSTs (88 m²), 4 LSTs (400 m²)

Array of the Southern site (Paranal)

- 2/3 LSTs (400 m²), 14 MSTs (88 m²), 42 SSTs (5 m²)

Southern Wide-field Gamma-ray Observatory

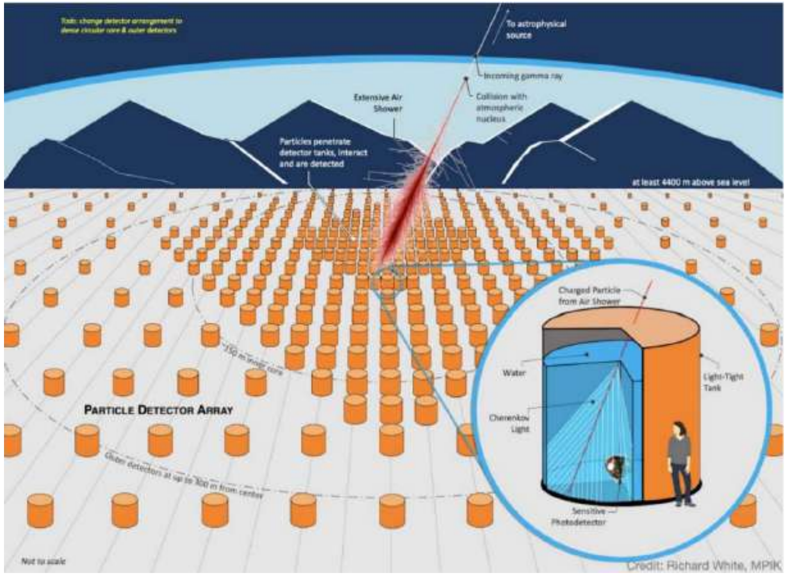
Project of new UHE observatory

- O(100)GeV → PeV
- PSF: 1°-2° → 0.04°
- FoV of 60° during day and nights

Under Design Study

M1	R&D Phase Plan Established
M2	Science Benchmarks Defined
M3	Reference Configuration & Options Defined
M4	Site Shortlist Complete
M5	Candidate Configurations Defined
M6	Performance of Candidate Configurations Evaluated
M7	Preferred Site Identified
M8	Design Finalised
M9	Conceptual Design Report Complete

Water Cherenkov Detector
Ex: HAWC, LHAASO



Selected site (Aug. 24): Atacama Astronomical Park, Chili

The project Gammapy

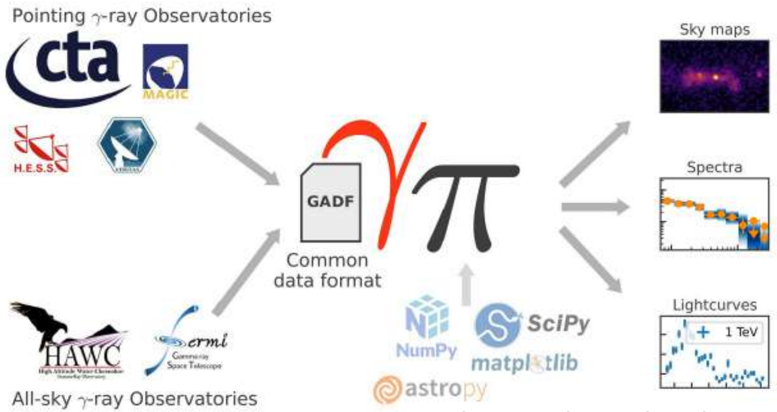
A **Python** package for **gamma-ray** astronomy

Open Python analysis library

- Uses data written in the GADF format
- Inserted into the Python ecosystem
- Respecting the FAIR4RS principles
- Making multi-instrument joint analyses

Open Research Software

- Open contributions within an open organisation with an open governance
- Reference library for the VHE gamma-ray astronomy
  - Selected as core library of the open Science Analysis Tool of CTAO
- Well recognized : jury's prize of the CoSO (2022)



Donath, A., *et al.*, A&A (2023) **678** A157

FAIR4RS principles

As with the FAIR Guiding Principles, the FAIR4RS Principles (2022) are intended to be aspirational. The application of the FAIR4RS Principles is the responsibility of the owners (who are often the creators) of the software, not the users.

- **F** Software, and its associated metadata, is easy for both humans and machines to find
- **A** Software, and its metadata, is retrievable via standardized protocols
- **I** Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.
- **R** Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software)

H.E.S.S. and XMM-Newton

γπ Multi-instrument analysis examples

Joint X-ray and γ-ray fits

First approach:

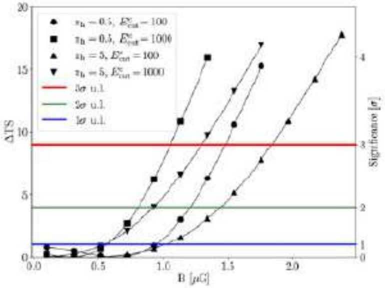
- Read OGIP spectra (1D DL4) produced by X-ray telescopes and fit

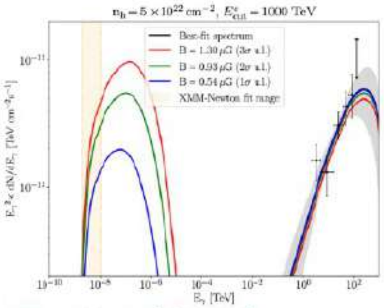
Second approach:

- Read X-ray events, IRFs and create 3D DL4 dataset

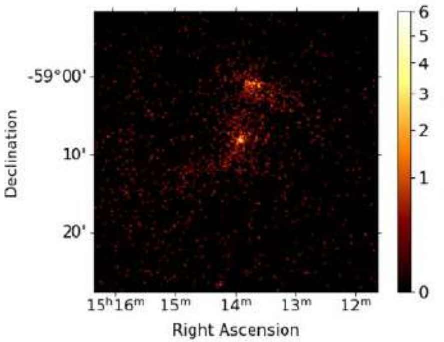
gammapyXray package DOI 10.5281/zenodo.7092736

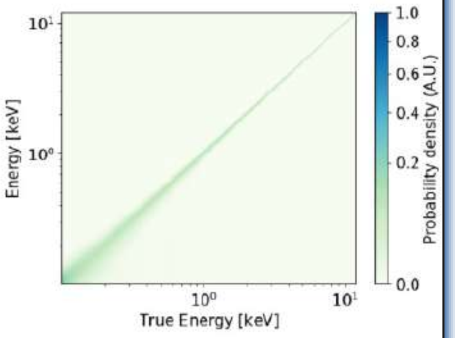
See e-ROSITA converter in K. Egg poster





Giunti et al (2022)





19

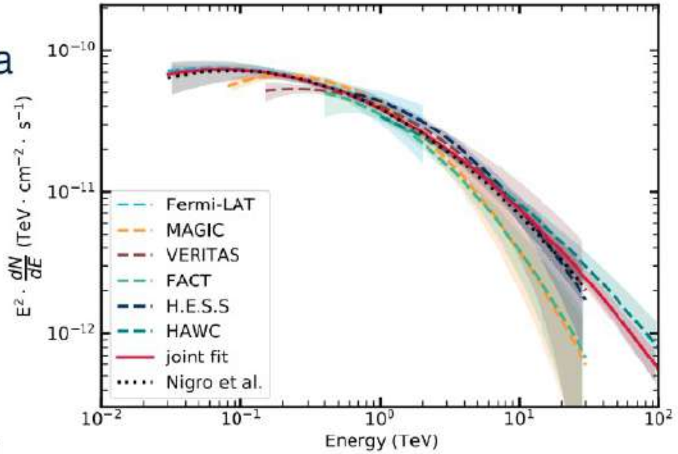
From R. Terrier, Gamma2024

Crab nebula: HE → UHE

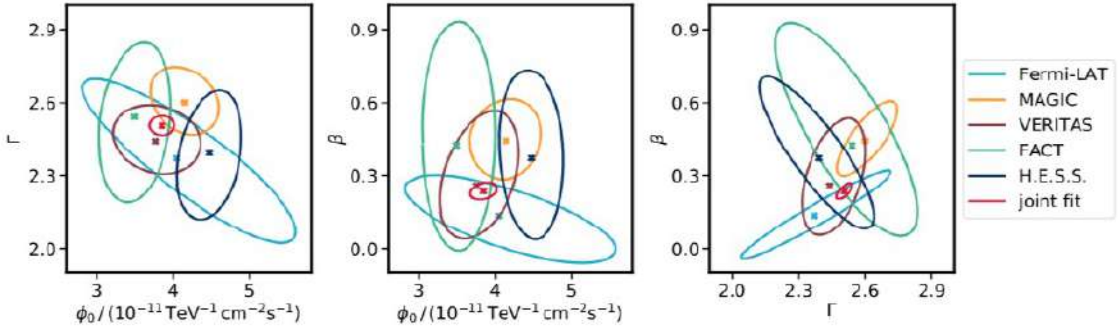
An example of multi-instrument analysis

Combining Crab observations

- Joint point-like 1D spectral analysis of the Crab nebula
- 5 different instruments over 3 decades in energy
  - Simple log-parabola & physical inverse Compton model
  - Modeling of some systematic uncertainties
- Fully reproducible analysis



+HAWC data Albert et al (2022)



Nigro et al. 2019

17

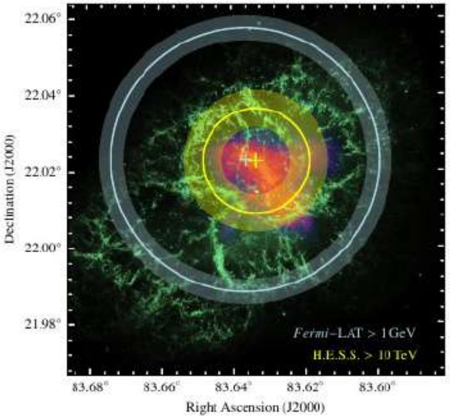
From R. Terrier, Gamma2024

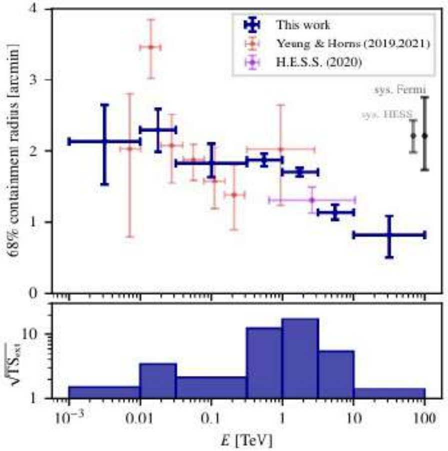
MWL extension of the Crab

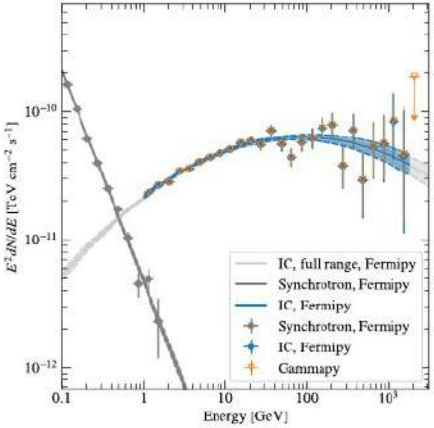
Multi-instrument analysis examples

Spectrum and extension of the inverse-Compton emission of the Crab Nebula from a combined *Fermi*-LAT and H.E.S.S. analysis [HESS Collab. 2024]

Gammapy has the advantage that it enables a combination of the *Fermi*-LAT data with that of H.E.S.S., and it offers the possibility to include customised physical models in the analysis.







From R. Terrier, Gamma2024