Access and Analysis of H.E.S.S. data

Bruno Khélifi & Catherine Boisson

IVOA standards for High Energy Astrophysics

June 28th 2023

VHE pointing gamma-ray instruments

Main characteristics of the current detectors

- Energy range: 10s GeV → 10s TeV
- Angular resolution: few arcmin (fct of energy)
- Energy dispersion: bias from 20% to 0% with a resolution from 40% to 5% (fct of energy)
- Pointing instrument with a Field of View of 3-5 deg (diameter)
- Arrays of several telescopes (2→ 5)
- Small duty cycle: ~10% (night with no or small moonlight)

For steady sources, 0(10)h of observations are needed

→ stacked analysis

Imaging Atmospheric Cherenkov Technique

gamma-ray

Atmospheric shower

~ 10 km

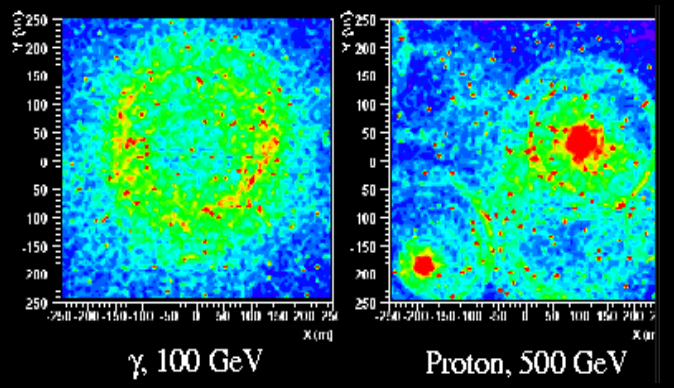
~ 1°

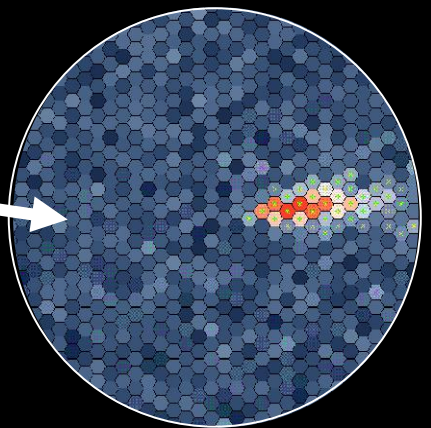
Cherenkov light

~ 120 m

With the stereoscopy:

- Large collection area
- Better gamma/hadron discrimination



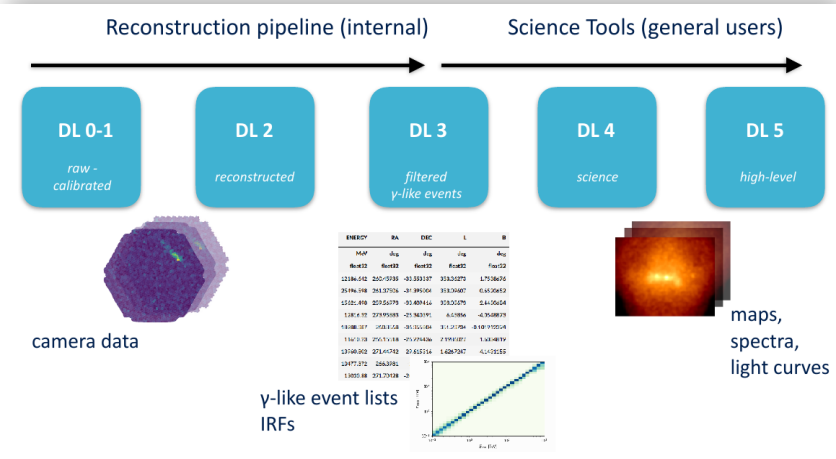


Imaging Atmospheric Cherenkov Technique

IACT features

- The atmosphere is our converter and calorimeter
  - Not stable: density profile and absorption length
- The optical telescopes and their fast camera (GHz sampling) detects the Cherenkov light of the Atmospheric Shower:
  - Indirect measurement
- Use of massive Monte-Carlo simulations to derive the gamma-ray properties
- Sources of “noise” or “background”
  1. Site: Night Sky Background → Hardware triggers
  2. Charged cosmic rays → Stereoscopy (h/w and s/w)
  3. Instrumental → Calibration
  4. Astrophysical → High-level analysis

Modern analysis chain and data levels



Internal reconstruction pipelines for H.E.S.S.

- Using private data (observations and MC) and private s/w
- Output for each (30min) observation:
  - List of gamma-ray candidates with (Energy, RA, Dec, Time)
  - 4 Instrument Responses Files (IRFs):
    Effective Area, PSF, Energy dispersion, Background Model
- Massive reduction of the data volume
  - 30min of raw data ~50Go Versus 16yrs of 'DL3' ~43Go

H.E.S.S. data levels

Data Levels: using the Gamma Astro Data Format

GADF: C. Deil et al., arXiv:1610.01884 - Documentation here

- DL3: gamma-like event, the 4 IRFs and auxiliary data
- DL4: binned data
- DL5: astrophysical products
- DL6: catalogues
- And two index files (storage)

Serialisation into FITS files

H.E.S.S. data levels: DL3

- **Gamma-like events**

  A list (mandatory+optional col.)

  PS: from real data or MC

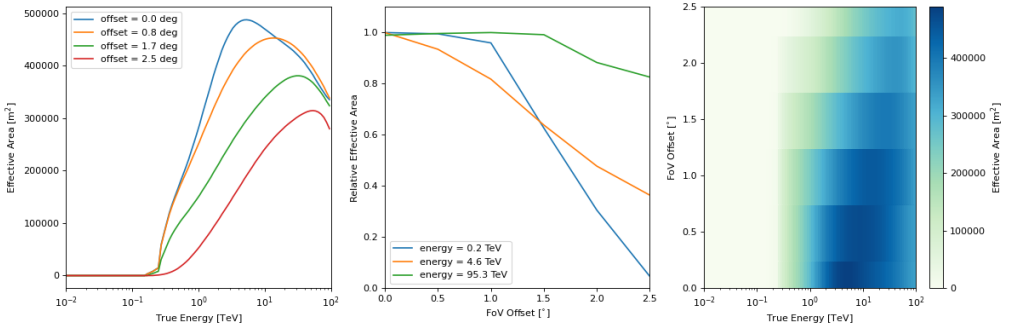
EVENT_ID	TIME s	RA deg	DEC deg	... ...	DETX deg	DETY deg	MC_ID	
1	664502403.0454683	-92.63541	-30.514854	...	-0.9077294	-0.2727693		2
2	664502405.2579999	-92.64103	-28.262728	...	1.3443842	-0.2838398		2
3	664502408.8205513	-93.20372	-28.599625	...	1.0049409	-0.7769775		2
4	664502409.0143764	-94.03383	-29.269627	...	0.32684833	-1.496021		2
5	664502414.8090746	-93.330505	-30.319725	...	-0.716062	-0.8733348		2

- **Effective Area**

  3 dimensions:

  FoV offset, Etrue, Area

  PS: from MC

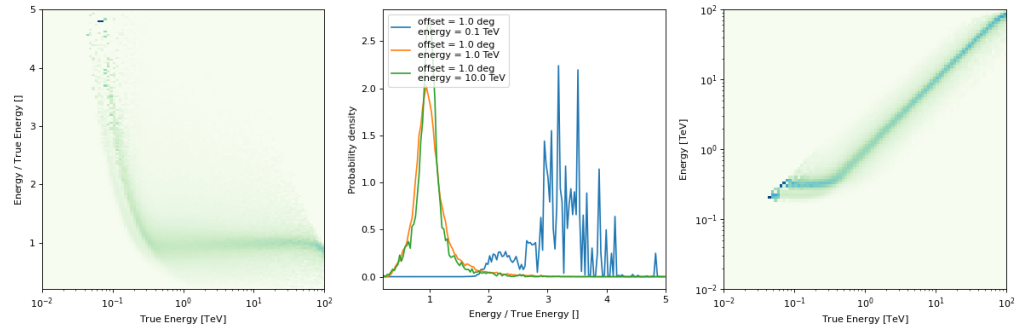


- **Energy Dispersion**

  3 dimensions:

  FoV offset, Etrue, Ereco/Etrue

  PS: from MC

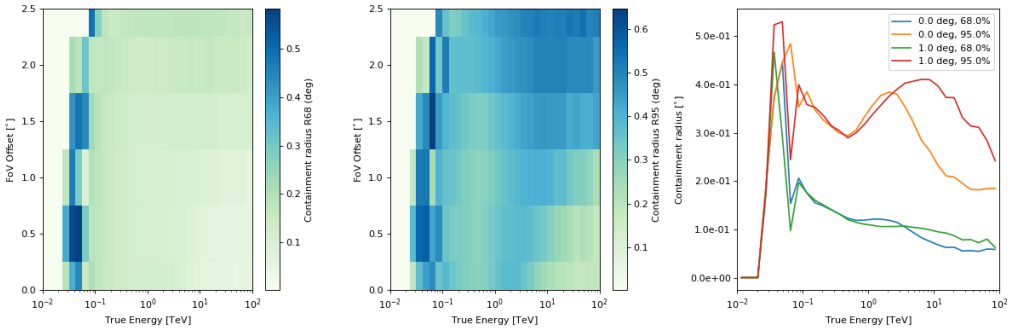


H.E.S.S. data levels: DL3

- **PSF**

  4 dimensions:
  FoV offset, Etrue, pdf(θ)

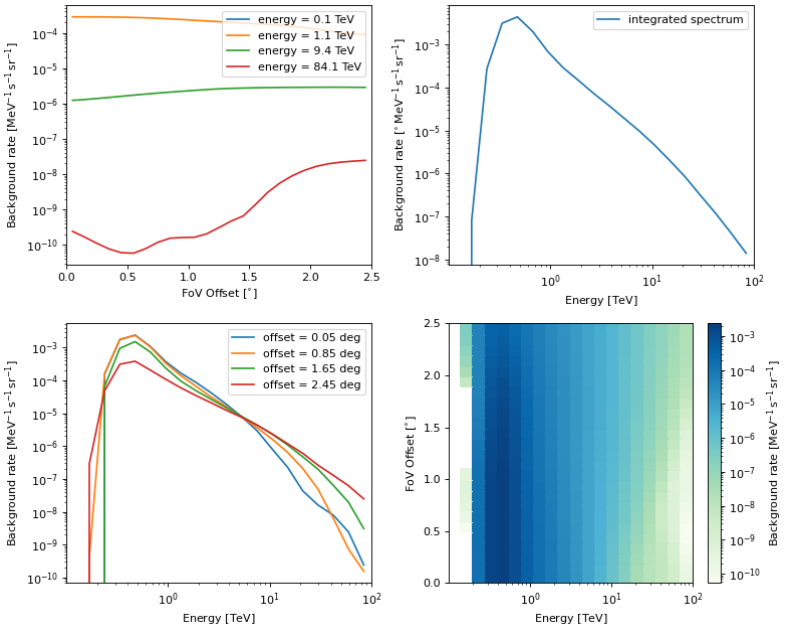
  PS: from MC



- **Background**

  3 dimensions:
  FoV offset, Ereco, Rate

  PS: from MC



- **Auxiliary information**

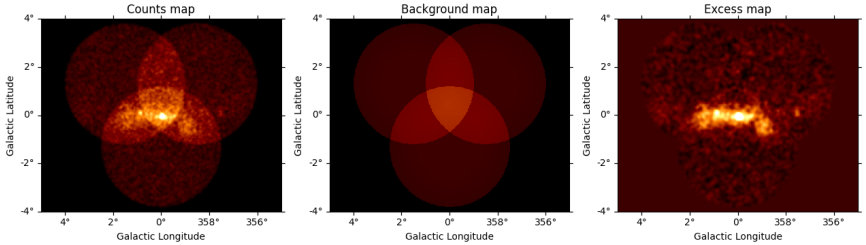
  e.g. time references, some provenance, pointing, “stable time interval”, livetime

H.E.S.S. data levels: DL3 index files

- **Observation index table**
  - Provides a kind of DB of the properties of the observations (e.g. pointing, quality, sub-array, obs time)
  - Improve the speed of the data selection
  - In GADF, this table is not mandatory as all the stored data are in each DL3 event files
- **HDU index table**
  - Provides the location of the data files
  - Offers flexibility on the repository organisation
    - Location of the storage of IRF files (within or not the event files)
  - Ultimately associated the files between events and IRFs
  - In GADF, this is the mandatory table! (a data model choice)

H.E.S.S. data levels: DL4

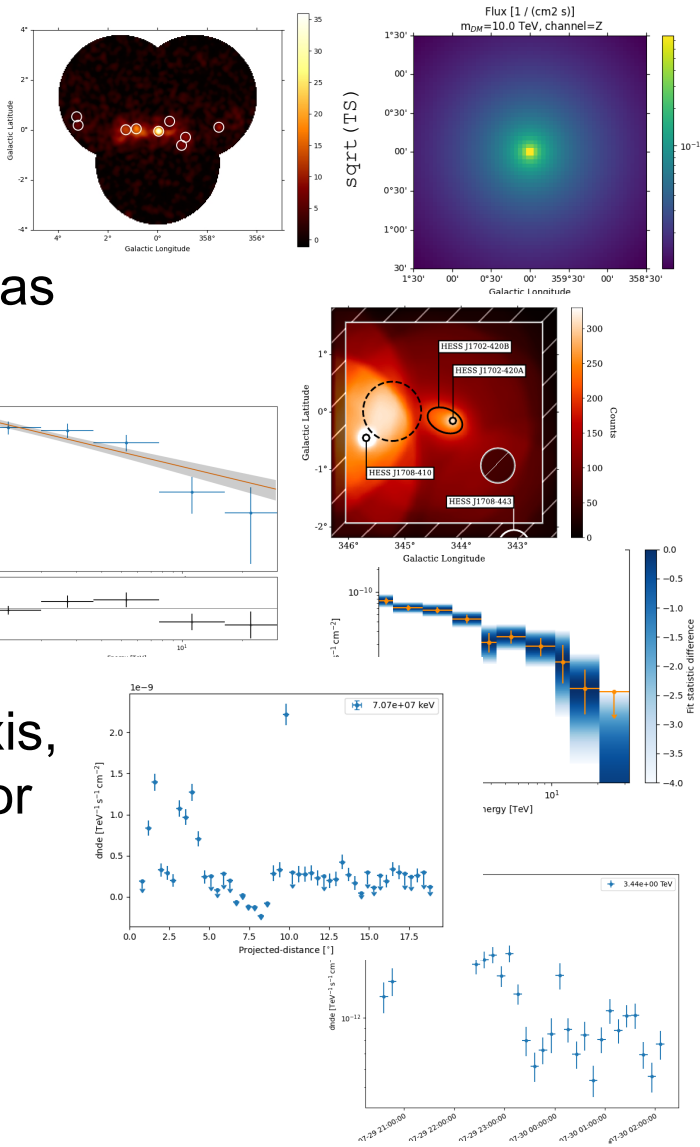
- **Binned data**
  - Bound a collection of “maps” needed for the analysis (e.g. exposure, bkg, edisp, psf, acceptance), some analysis masks (from the irfs or the users), etc



  - Using the user geometry: WCS, HEALPIX or 1 dimension with its own binning (spatial or spectral) → Reprojection of the IRFs
  - Can have 2 or more dimensions (e.g. energy, time)
  - Intermediate products for the user analysis (serialization possible): Not have the vocation to be published (but shared)
  - In GADF: not completely defined (map axis, dimensions, metadata)

H.E.S.S. data levels: DL5

- **Astrophysical data**
  - Sky maps: flux map, significance map, exposure map
  - On a given region: spectrum, flux points (as function of energy, time, distance)
  - Can have several dimensions (>3)
  - Final products of a user (serialization possible)
  - In GADF: not completely defined (map axis, dimensions, metadata, correlated errors or correlation matrix, source model)
  - In Gammapy, the source models can be serialised in YAML



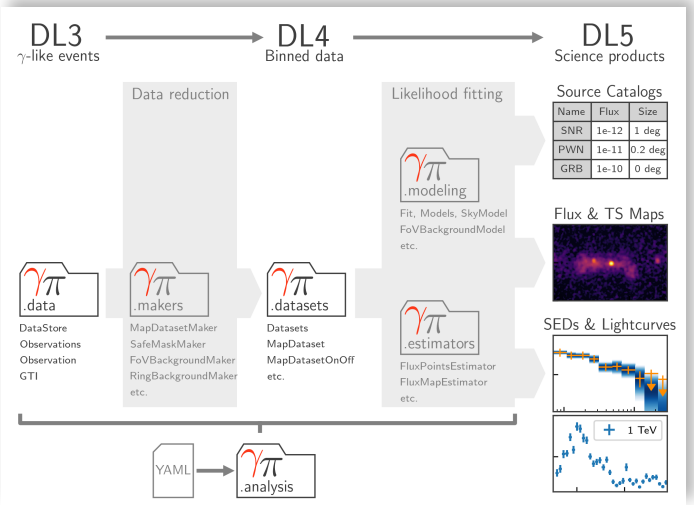
(H.E.S.S.) data levels: DL6

- **Catalogue data**
  - No real format so far, in H.E.S.S., GADF
  - In Gammapy, any dictionary can be serialised in YAML, but… this is not FAIR enough!

High-level analysis for H.E.S.S.

The open software Gammapy, official analysis tool

- This library is used to produce DL5/6 products from the DL3 data



- Several DL3 productions are made under the GADF format
  - Stored on the 2 H.E.S.S. data centres
  - Everyone can make a copy into their PC
- Gammapy accesses to the data with a specific API, using the index files and loading the DL3 files when necessary

Future of the H.E.S.S. analysis

Preparation of the H.E.S.S. legacy archive

- New low-level analysis is under development
  - With new calibration, new simulations, new reconstruction and discrimination methods
- This will lead to the production of
  - new event list (with event types),
  - new MC-based IRFs (with validity cube or error cube?)
- The determination of new background models is still pending
  - But an extra dimension might be added: FoV offset → FoV (X, Y)
- This new data release will use the up-to-date data format
- The choice of the archive type has not yet been discussed
  - Data centre repository,
  - What about a TAP server in addition?

Future of the H.E.S.S. data release

Evolution of the GADF format

- The main actors of the GADF initiative have identified some limitations
  - On the format: e.g. definition of multi-dimensional maps, serialisation of source models, definition of the metadata at each data level, missing specification for DL4+
  - On the organisation: no clear organisation, supporting experiments
- The under-construction major astroparticle observatories will have public data releases
  - The current GADF format is not perfectly in line with the requirements
- Some of us decided to contact the main astroparticle experiments to built a new initiative aiming to create a new VHE data format:
  - Respecting more closely the FAIR principles,
  - Following as much as possible the IVOA recommendations, in particular from the recently created HE group

Future of the H.E.S.S. data release (2)

Creation of a new open initiative:

Very-high-energy Open Data Format (VODF)

https://vodf.readthedocs.io/en/latest/index.html

- Officially supported by eleven VHE experiments:
  - ASTRI, CTAO, FACT, Fermi-LAT, HAWC, H.E.S.S., IceCube, KM3NeT, MAGIC, SWGO, VERITAS
- Common open format for VHE gamma-ray and neutrino data
  - Pointing gamma-ray array, Slewing gamma-ray array, Neutrino detector
- Aims:
  - Compliant with the requirements of the certified data repositories (e.g. EOSC, SDC), respecting more closely the FAIR principles,
  - Following as much as possible the FAIR principles and the IVOA standards (metadata and provenance)
  - Formatting of the DL3 (science ready) up to DL6 (catalogs), with the common data structures (general metadata, time format, coordinates, N-dim maps)

Future of the H.E.S.S. data release (2)

Very-high-energy Open Data Format (VODF)

VODF
very-high-energy open data format

- Clear organisation:
  - Governance document
  - Steering committee of one official delegate of each experiment
  - 3 Lead Editors: K. Kosack (IACT), L. Olivera-Nieto (WCD), J. Schnabel (Neutrino)
  - 2 conveners: R. Zanin, B. Khélifi
- Operation
  - Using inputs from observatories and individuals, discussion on features and agreement by consensus
  - The LE animate the discussion and validate the modifications
  - The major change will be validated by the SC
- Perspective:
  - A preliminary version will be based on the CTAO extension of the GADF format and on the HAWC experience
  - Next versions will respect the FAIR principles and will try to follow the IVOA recommendations
- Open questions:
  - What about the source models?
  - Why not “opening” the serialization format? (e.g. ASDF)

Summary and conclusions

- **H.E.S.S. dataflow and analysis**
  - Low-level and high-level analysis using private data and software
  - With the arrival of CTA, use of open tools (Gammapy) using an open data format (GADF)
  - The high-level analysis scheme appears to be similar to gamma-ray WCD and neutrino detectors

- **Construction of open observatories**
  - In phase with the Open Science movement
  - Exacerbated need of joint MWL/MM analyses
  - Data will be accessible on certified repositories

- **Future of the VHE data format: VODF**
  - Creation of this initiative to establish a common data format for gamma rays and neutrino
  - Respect of the FAIR principles
  - Follow as much as possible the IVOA reco, and participation to the IVOA work