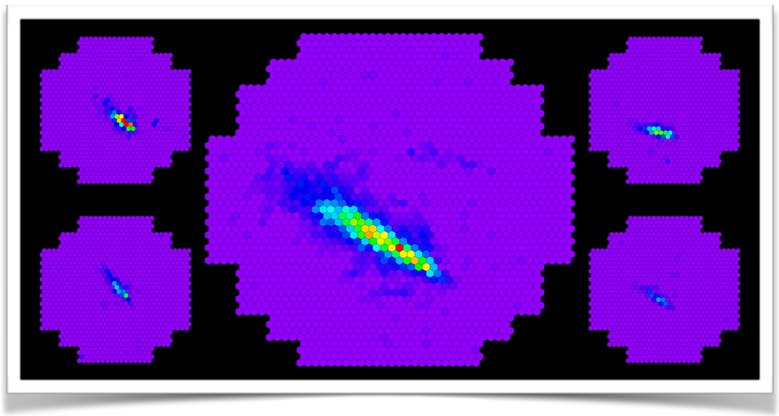
Towards high-level data analysis with open-source tools in H.E.S.S.

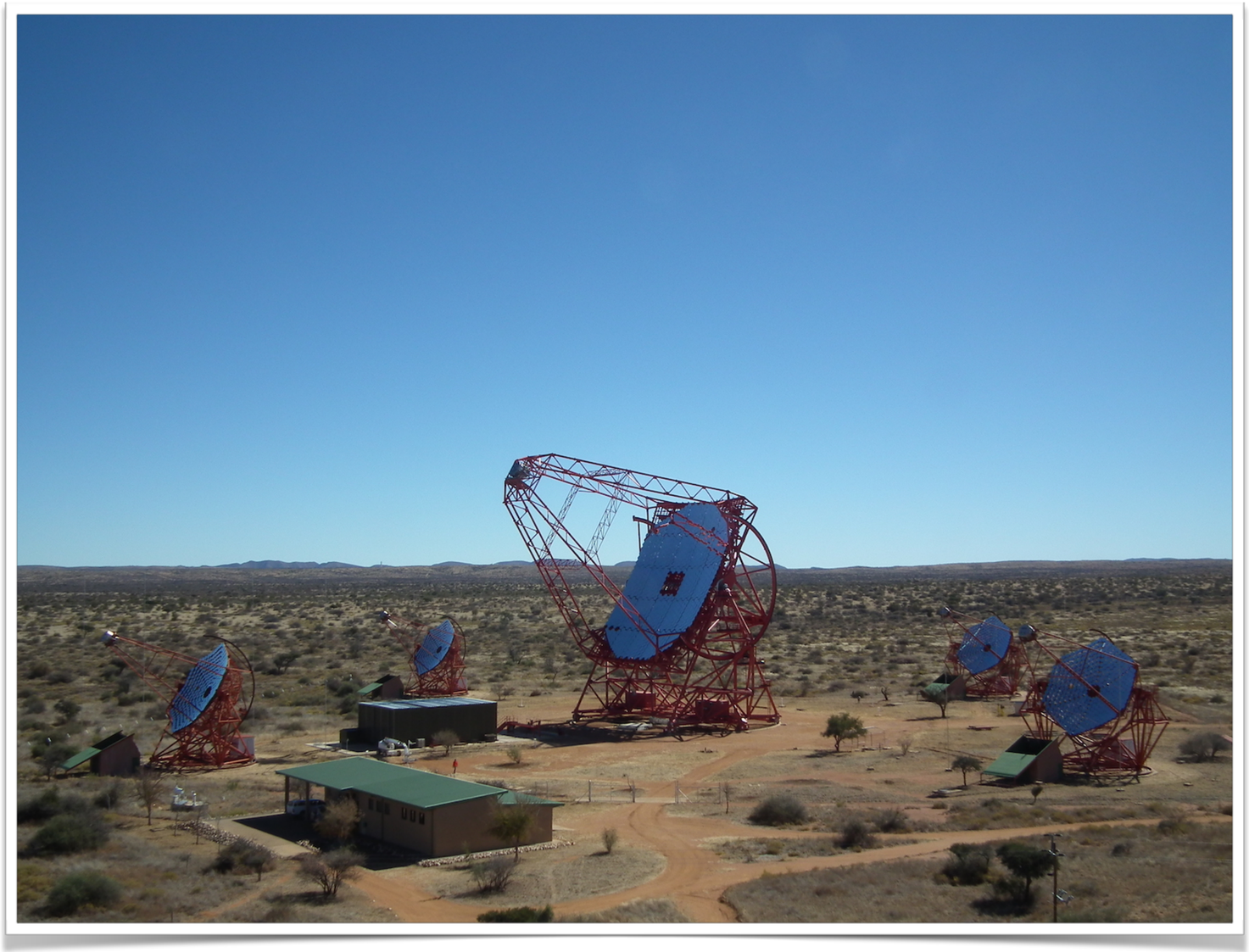
Lars Mohrmann
Erlangen Centre for Astroparticle Physics

PyGamma19
Heidelberg, March 20, 2019

The High Energy Stereoscopic System (H.E.S.S.)

- System of five Imaging Atmospheric Cherenkov Telescopes (IACTs)
- Detects gamma-rays between 20 GeV and 100 TeV
- Located in Namibia
- Four small telescopes in operation since 2004 (“HESS-I”)
- Large telescope added in 2012 (“HESS-II”)





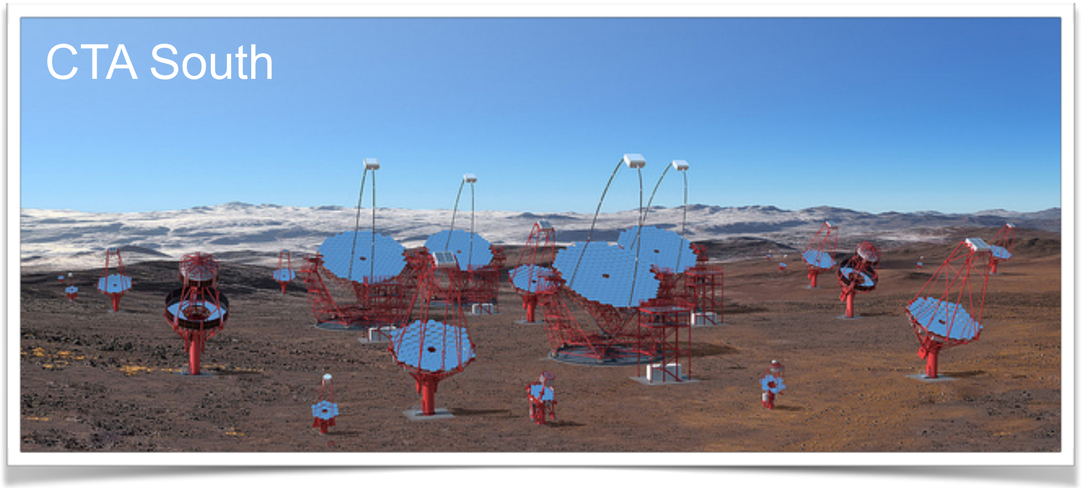
H.E.S.S. data analysis with open-source tools: motivation

- H.E.S.S. has two internal analysis pipelines
  ➤ Why should we move towards new tools?

- Upcoming Cherenkov Telescope Array (CTA)
  - CTA will be an observatory
    ➤ open data, open analysis tools
  - H.E.S.S. data analysis with CTA tools…
    ➤ … helps us prepare for CTA data analysis
    ➤ … lets us profit from CTA developments

- New tools offer new analysis techniques / possibilities
  - 3D likelihood analysis
  - combination of data from different instruments (see presentation by Cosimo yesterday)

- Here: focus on high-level analysis



ctapipe → data calibration, event reconstruction, event classification

ctools cherenkov telescope array

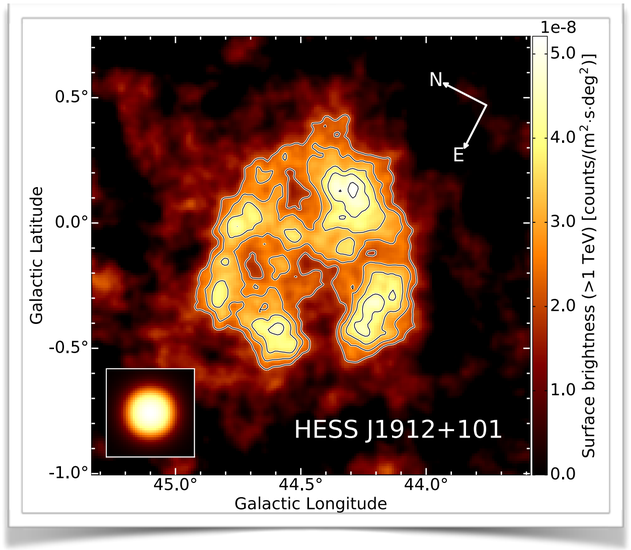
gammapy A Python package for gamma-ray astronomy

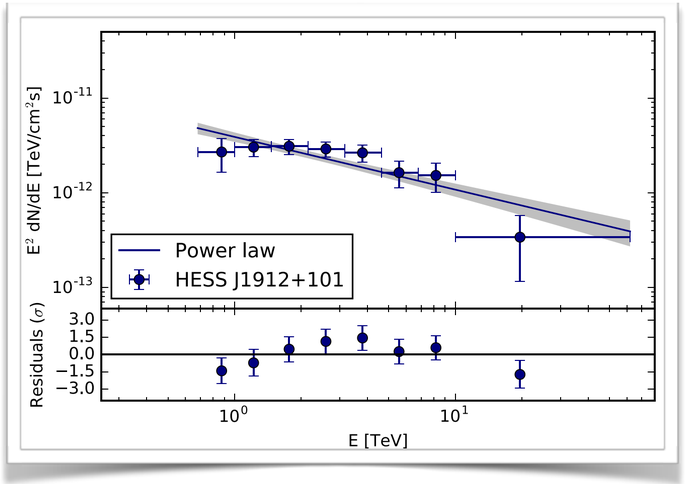
→ “high-level analysis”

High-level analysis of IACT data

- Input: “data level 3”
  - Event lists
    - ➤ After reconstruction and gamma-hadron separation
  - Instrument response functions
    - ➤ Effective area
    - ➤ Point-spread function
    - ➤ Energy dispersion
    - ➤ Background acceptance
  - Open format specifications
    - ➤ https://gamma-astro-data-formats.readthedocs.io
    - ➤ Based on FITS format (standard in astronomy since decades)

- Output:
  - ➤ maps, spectra, …

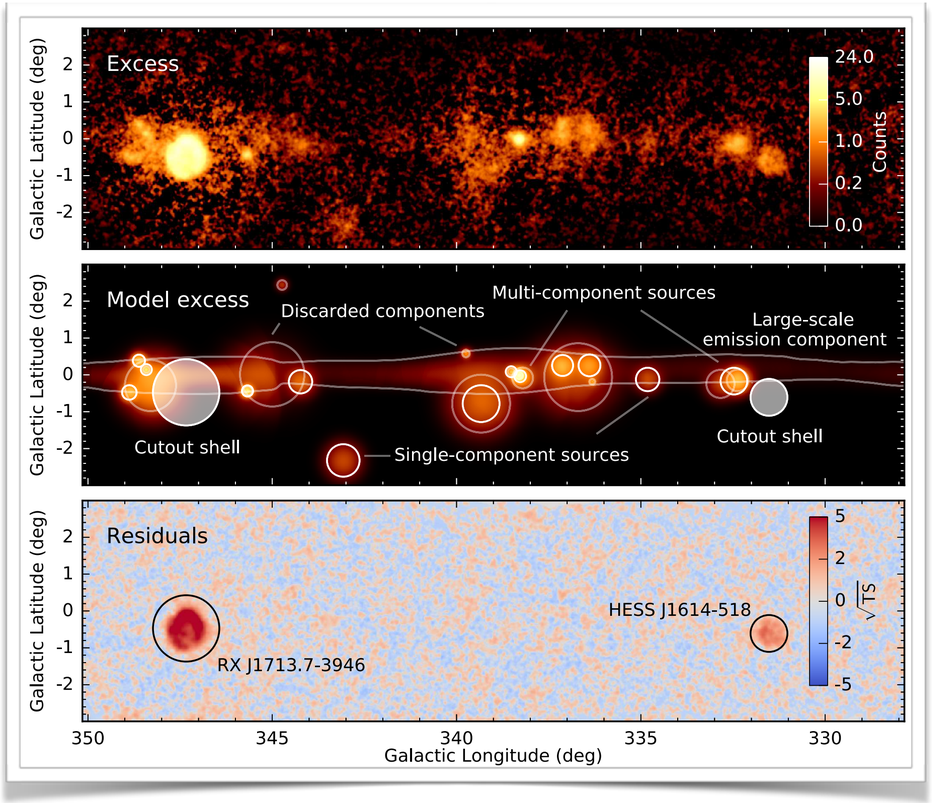




A&A 612, A8 (2018)

H.E.S.S. data analysis with open-source tools: status

- Task group in place since several years, working on:
  - exporter scripts to FITS data format
  - open FITS format specifications
  - analysis validation
- In the past: results obtained with open-source tools viewed with strong scepticism
  - Reliability of FITS data and tools not sufficiently demonstrated
  - → with few exceptions, open-source tools not used for publications
- This is changing now…
  - Tools becoming more and more sophisticated
  - Increasing interest to move towards CTA tools
  - Task group working on increasing confidence in new tools
- In the following:
  - Technical requirements for H.E.S.S. data analysis with open-source tools
  - Public test data release & validation paper



Catalog model construction
for the H.E.S.S. Galactic Plane Survey*
with Gammapy

* A&A 612, A1 (2018)

H.E.S.S. FITS data exporter: observation index

- Current implementation

OBS_ID	RA_PNT	DEC_PNT	ALT_PNT	AZ_PNT	...	OBJECT	LIVETIME	DATE-OBS	TIME-OBS	N_TELS	EVENT_COUNT
	deg	deg	deg	deg	...		s				
int64	float32	float32	float32	float32	...	bytes18	float32	bytes10	bytes12	int64	int64
20136	228.6125	-58.771667	51.487038	195.73102	...	MSH15-52	1521.0269	2004-03-26	02:57:46.184	4	11243
20137	228.6125	-59.771667	49.78384	199.6482	...	MSH15-52	819.2054	2004-03-26	03:28:25.184	4	5693
20151	228.6125	-58.771667	52.835342	190.97171	...	custom	1530.479	2004-03-27	02:28:16.184	4	10722
20275	187.27792	2.552389	53.81757	49.144917	...	3C 273	1525.8447	2004-04-14	19:51:40.184	4	10646
20282	228.6125	-58.771667	52.86866	169.21602	...	MSH 15-5-02	1543.3231	2004-04-14	23:32:13.184	4	11063
20283	228.6125	-59.771667	53.778564	175.77263	...	MSH 15-5-02	1547.8253	2004-04-15	00:02:31.184	4	7628
20301	228.6125	-58.771667	53.160095	171.07347	...	MSH 15-5-02	1534.0626	2004-04-15	23:35:42.184	4	11624
20302	228.6125	-59.771667	53.88368	177.10521	...	MSH 15-5-02	1535.3478	2004-04-16	00:06:02.184	4	11794
20303	228.6125	-58.771667	53.813988	183.91011	...	MSH 15-5-02	1535.1606	2004-04-16	00:36:24.184	4	11803
20322	228.6125	-59.771667	53.086906	170.6292	...	MSH 15-5-02	1529.7925	2004-04-16	23:29:49.184	4	12925
20323	228.6125	-58.771667	53.860382	176.67052	...	MSH 15-5-02	1530.1174	2004-04-17	00:00:08.184	4	13367
20324	228.6125	-59.771667	53.843937	183.57172	...	MSH 15-5-02	1527.8722	2004-04-17	00:30:24.184	4	12822
20325	228.6125	-58.771667	53.043026	189.65836	...	MSH 15-5-02	1529.3948	2004-04-17	01:00:37.184	4	12878
20326	259.29852	-39.762222	71.2878	154.81943	...	RXJ 1713.7-3946	1500.009	2004-04-17	01:32:21.184	4	16880
20327	257.47733	-39.762222	73.265114	173.0282	...	RXJ 1713.7-3946	1497.074	2004-04-17	02:02:53.184	4	16995
20339	201.36667	-42.319168	65.66224	148.75554	...	Cen A	1494.5898	2004-04-17	21:13:47.184	4	15971
20343	228.6125	-58.771667	52.657578	168.30562	...	MSH 15-5-02	1529.0181	2004-04-17	23:15:42.184	4	12590
...	...	...	...	...	...	...	...	...	...	...	...

H.E.S.S. FITS data exporter: events list

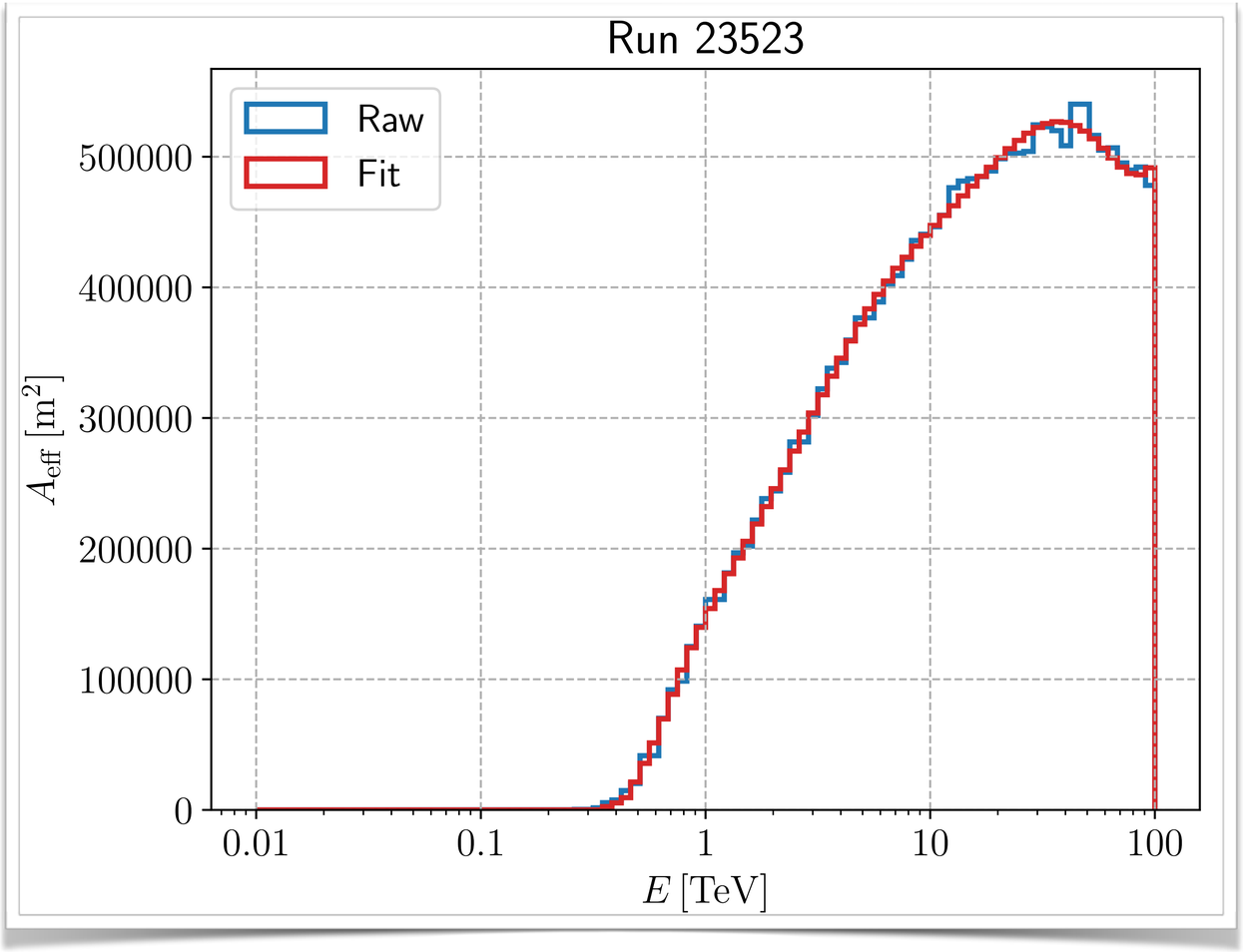
- Current implementation
- Possible improvements
  - Introduce EV_CLASS?
    - ➤ exists for Fermi data
    - ➤ requires update of format specifications
  - …?

EVENT_ID	TIME	RA	DEC	ENERGY
	s	deg	deg	TeV
int64	float64	float32	float32	float32
1808181231761	101962602.82030201	229.23874	-58.34167	0.55890286
1808181231778	101962602.86233354	226.88606	-58.92101	0.6710115
1808181231793	101962602.91234565	230.85834	-57.98291	5.2125587
1808181231871	101962603.16041279	229.7445	-60.32017	0.5505927
1808181231873	101962603.16430616	227.88457	-60.650185	0.4242001
1808181231920	101962603.32223964	230.81284	-59.07336	0.52716583
1808181231987	101962603.52126741	230.767	-57.855816	0.82927954
1808181231989	101962603.5271964	225.2312	-60.795094	1.966491
1808181232037	101962603.71615148	232.74625	-58.818344	0.9973961
1808181232159	101962604.136518	230.62297	-58.39907	0.37792137
1808181232280	101962604.54519558	227.16501	-59.68531	0.39711657
1808181232297	101962604.60105062	228.14122	-60.638012	0.44185287
1808181232314	101962604.6587925	229.23718	-58.540577	0.39222783
1808181232335	101962604.712075	230.52957	-57.761356	0.36581835
1808181232336	101962604.7151413	227.86511	-57.4885	0.45009688
1808181232344	101962604.73378491	229.66359	-59.745564	0.420972
1808181232473	101962605.1321826	226.64604	-59.21028	0.4415344
...	...	...	...	...

(n.b. more event information is available in principle,
but currently not included by default)

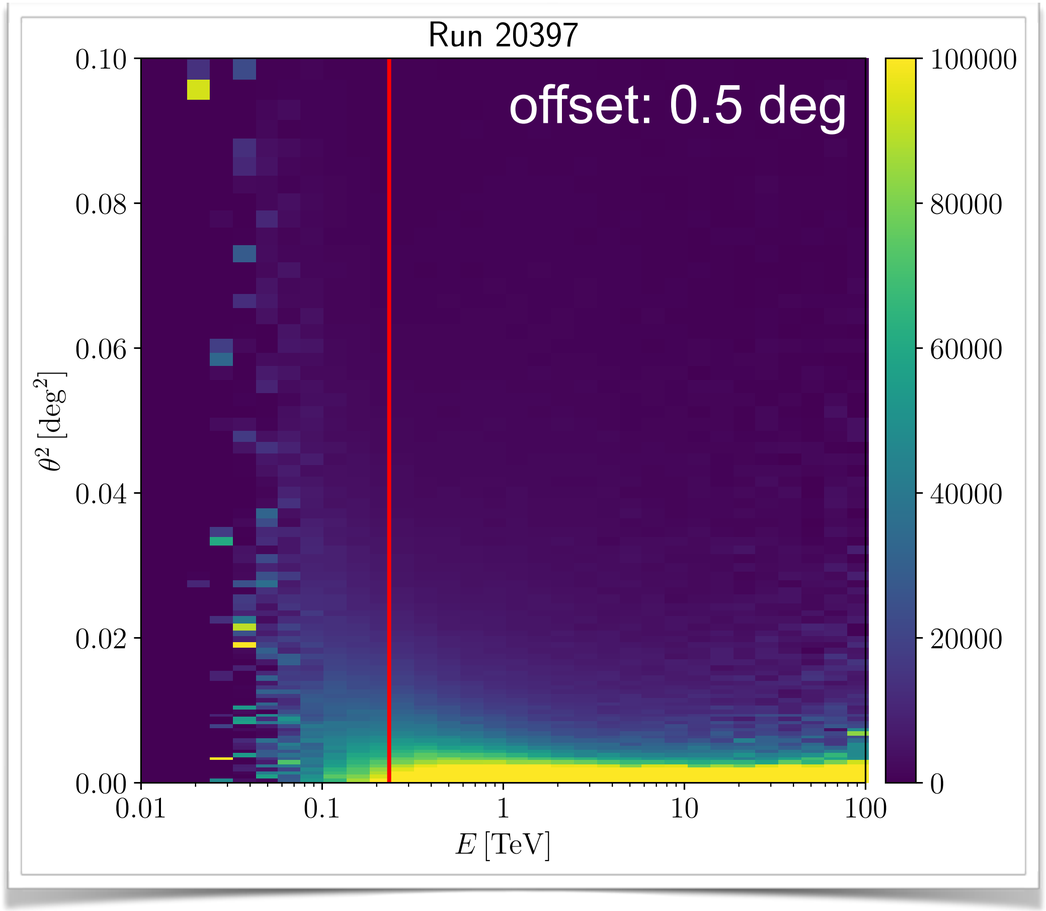
H.E.S.S. FITS data exporter: effective area

- Current implementation
  - 6 offset angles: [0.0, 0.5, 1.0, 1.5, 2.0, 2.5] deg
  - Energy range: [0.01 - 100] TeV
  - 24 bins / decade in energy
  - Smoothed with polynomial fit
- Problems
  - No automated check for robustness of smoothing fit
- Possible improvements
  - Implement automatic checks
  - Use better smoothing algorithms



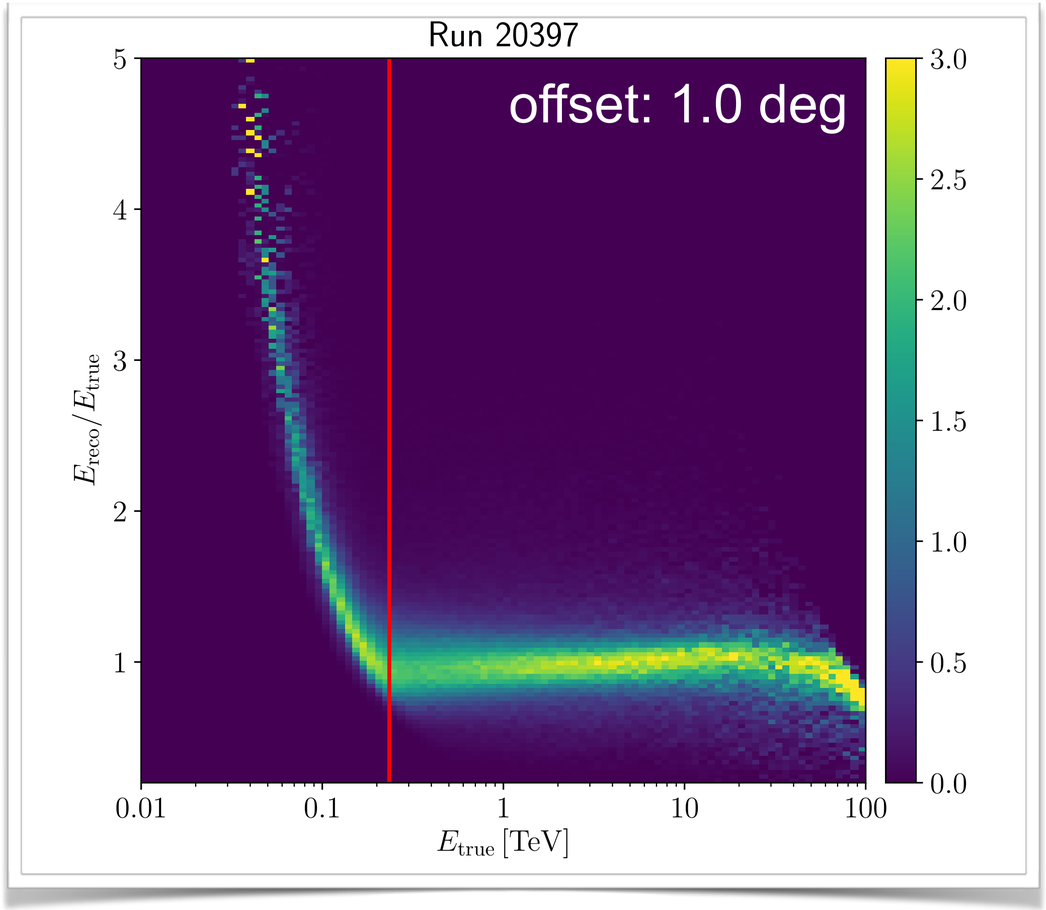
H.E.S.S. FITS data exporter: point-spread function

- Current implementation
  - 6 offset angles: [0.0, 0.5, 1.0, 1.5, 2.0, 2.5] deg
  - Energy range: [0.01 - 100] TeV
  - 8 bins / decade in energy
  - 144 bins in angle (fine → coarse)
  - No smoothing applied
- Problems
  - Statistical noise
- Possible improvements
  - Tested analytical descriptions (King profile, triple Gaussian)
    ➤ have their own issues, generally not a better description
  - Denoising / smoothing algorithms
    ➤ “Anisotropic smoothing” (proposal by Axel)?
    ➤ Kernel density estimation?



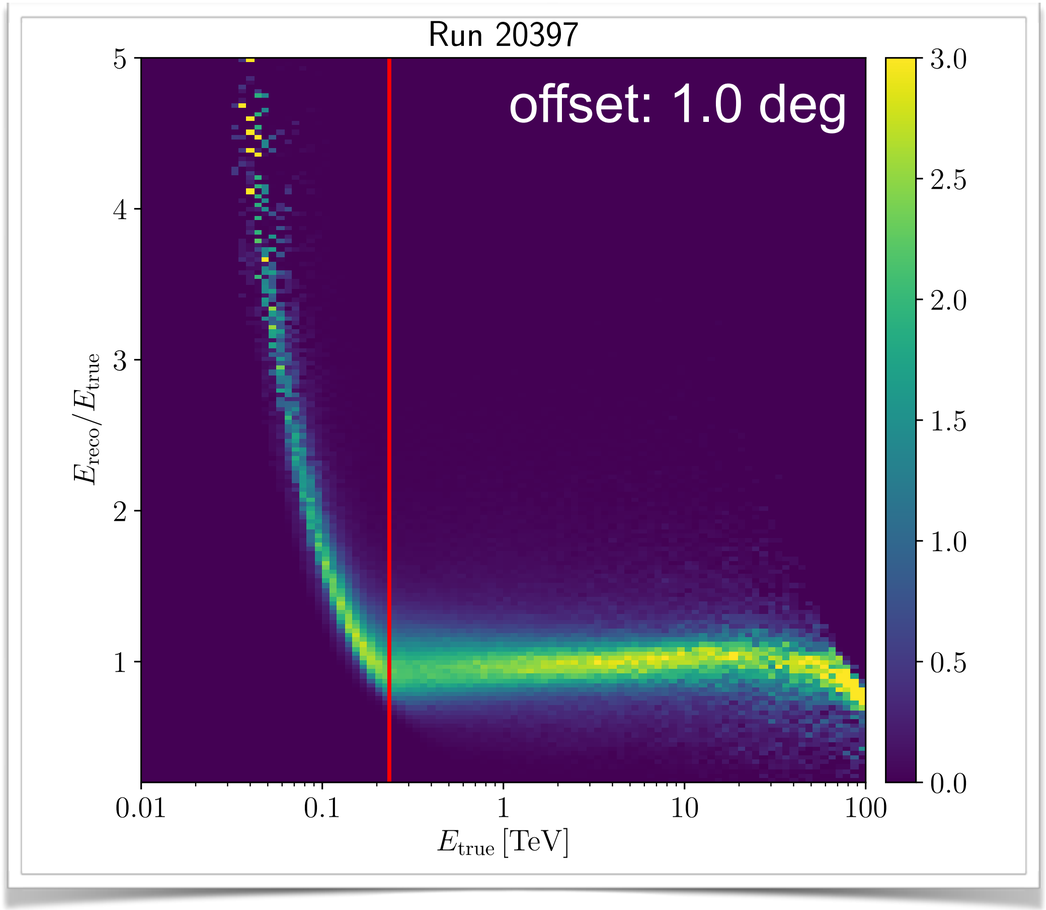
H.E.S.S. FITS data exporter: energy dispersion

- Current implementation
  - 6 offset angles: [0.0, 0.5, 1.0, 1.5, 2.0, 2.5] deg
  - Energy range: [0.01 - 100] TeV
  - 24 bins / decade in energy
  - 160 bins in E_{reco} / E_{true} (bin width of 3%)
  - No smoothing applied
- Problems
  - Statistical noise
    (n.b. this has even been an issue in the CTA 1DC!)
- Possible improvements
  - Denoising / smoothing algorithms
    (very likely similar solution as for the PSF)



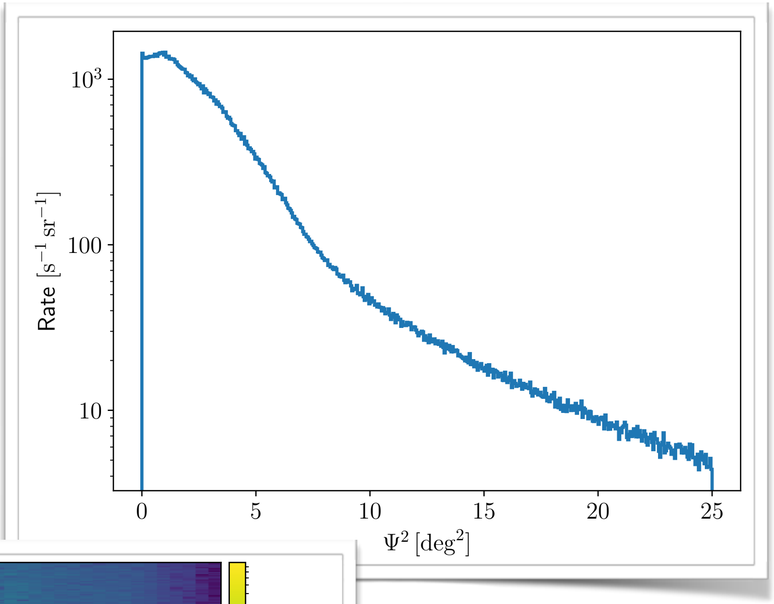
H.E.S.S. FITS data exporter: safe energy threshold

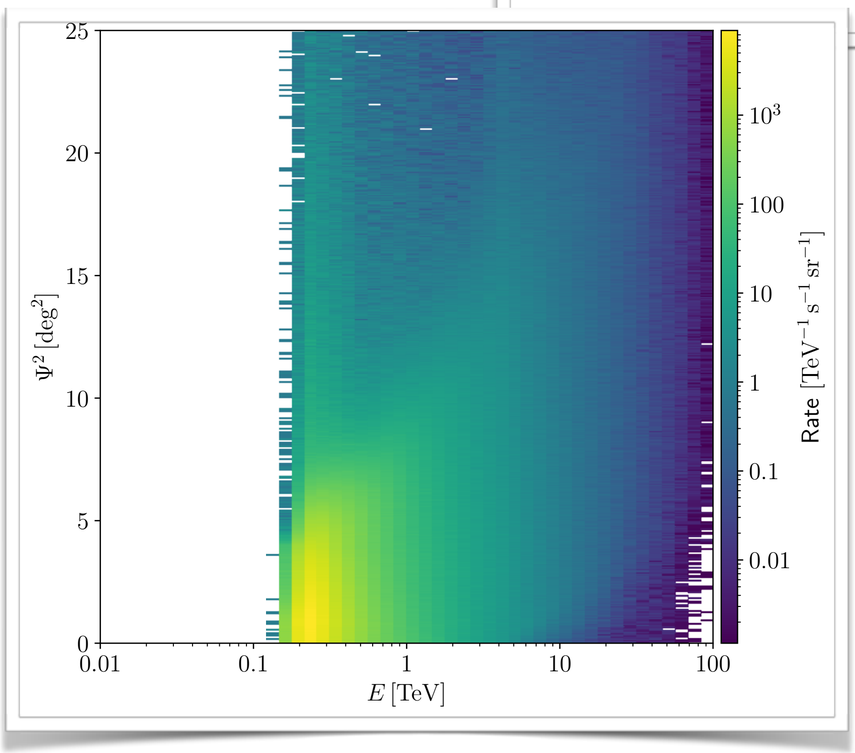
- Current implementation
  - One value, specified for each observation
  - Energy bias <10% for source at 1 deg offset from pointing
- Problems
  - Source offset is often not 1 deg
  - Analysis may require acceptable energy bias across entire field-of-view
  - Specify validity of IRFs or recommended analysis threshold?
- Possible improvements
  - Offset-dependent thresholds
    ➤ would solve first two problems
    ➤ not foreseen in format specifications so far
  - Give values for IRF validity and recommended analysis threshold?



H.E.S.S. FITS data exporter: background acceptance

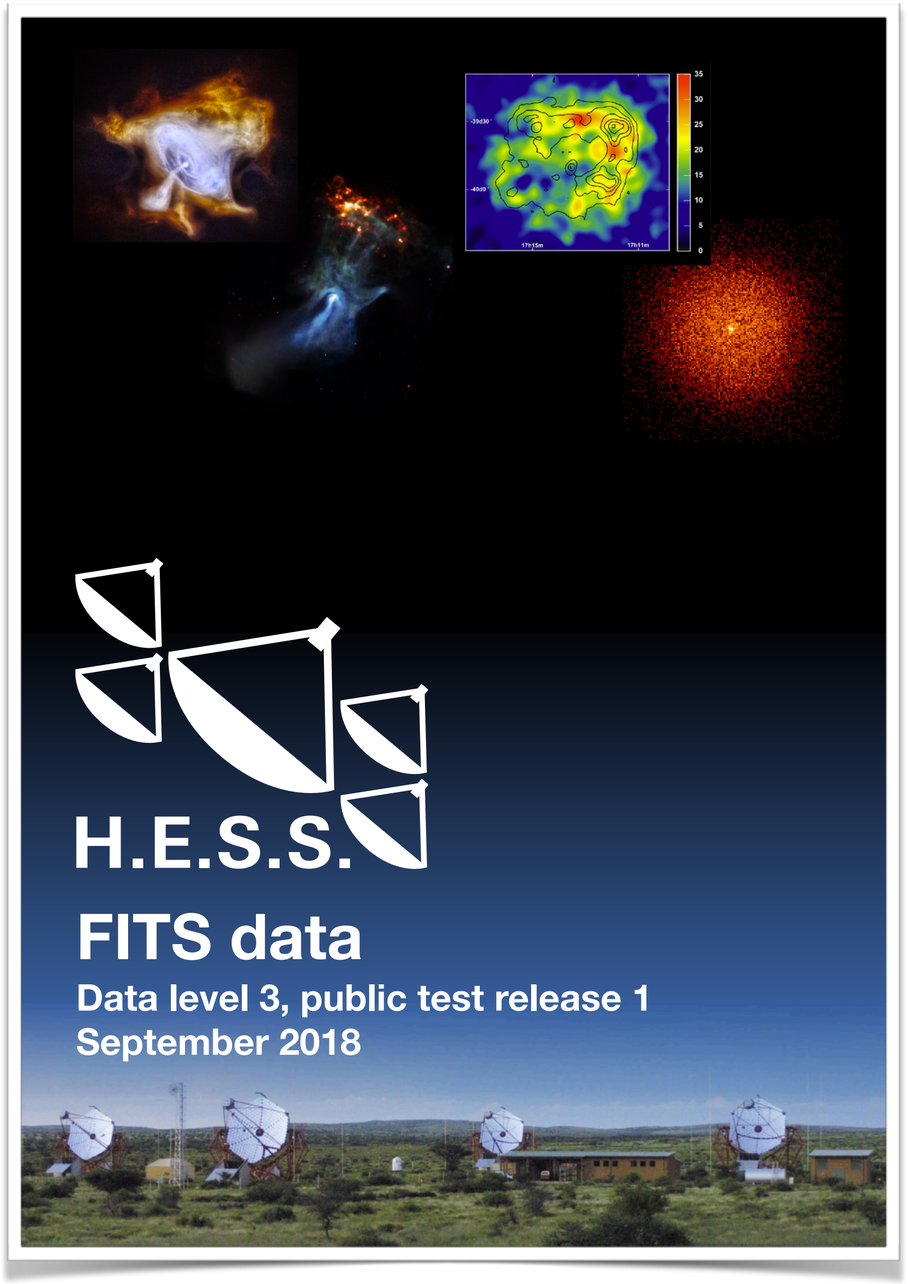
- Current implementation
  - None
- Open questions
  - Assumption of radial symmetry w.r.t. pointing position (2D vs 3D model)
  - Required accuracy (heavily depends on analysis…)
  - Coordinate system definition for 3D model
    ➤ ctools expects RA/Dec-aligned system
    ➤ Gammapy expects Alt/Az-aligned system
- Short-term solution
  - Provide simple 2D acceptance for several zenith angle bins
  - Development of a more sophisticated model is underway (more later)





H.E.S.S. first public test data release

- https://www.mpi-hd.mpg.de/hfm/HESS/pages/dl3-dr1
  https://zenodo.org/record/1421099
  https://arxiv.org/abs/1810.04516
- Released in September 2018
  (multi-year effort!)
- Provides “real” IACT data in
  FITS format for the first time
- Small dataset, for testing purposes
- Aims
  - Support development of open-source tools and format specifications
  - Provide IACT data for non-experts
  - Prepare analysis of H.E.S.S. FITS data

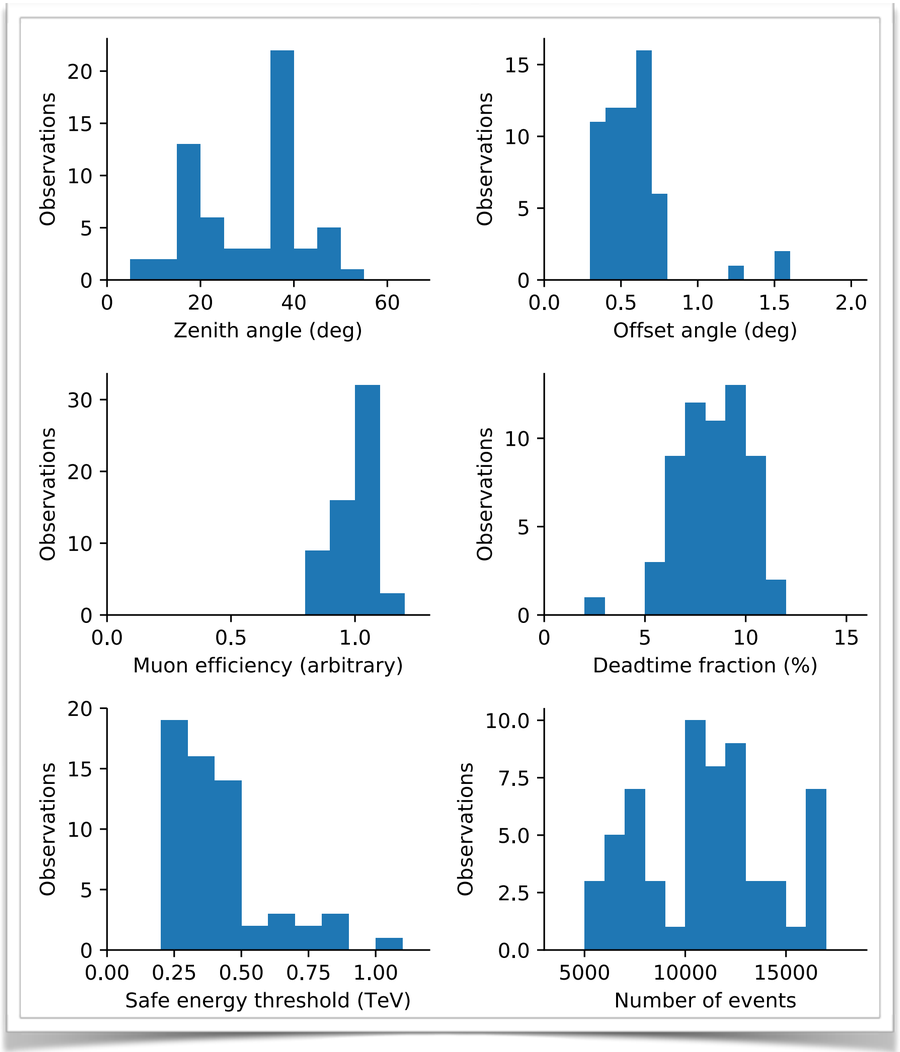


H.E.S.S. first public test data release

- Data sets

Source	Type	Size	Livetime	Period
Crab nebula	PWN	Point-like	1.9 h	2004
PKS 2155-304	AGN	Point-like	9.8 h	2006/2008
MSH 15-52	PWN	Small	9.1 h	2004
RX J1713.7-3946	SNR	Large	7.0 h	2004
Off data	-	-	20.7 h	2004/2005

- Notes
  - All data were taken with the HESS-I array (i.e., with the four small telescopes)
  - “std” configuration: “Hillas-style” analysis
    ➤ no sophisticated gamma-hadron separation or reconstruction techniques
  - Quality of instrument response functions not always optimal
  - Use for scientific publications not permitted



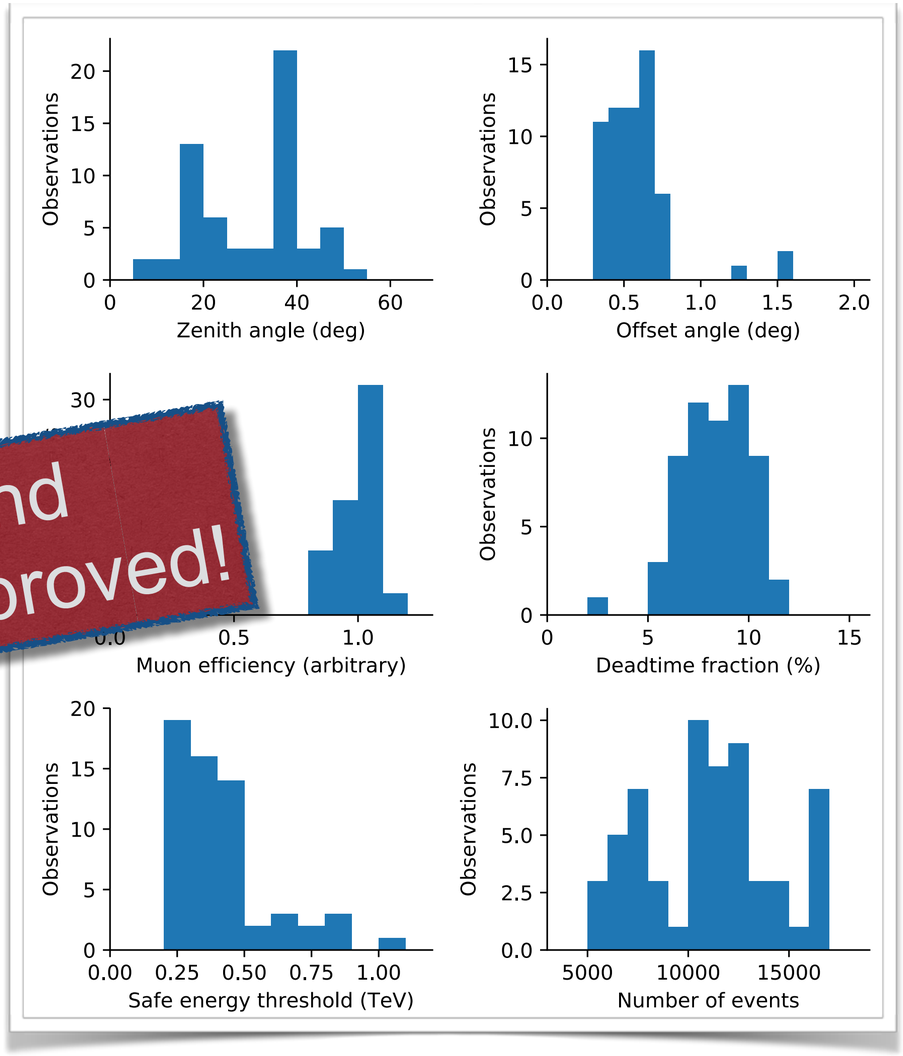
H.E.S.S. first public test data release

- Data sets

Source	Type	Size	Livetime	Period
Crab nebula	PWN	Point-like	1.9 h	2004
PKS 2155-304	AGN	Point-like	9.8 h	2006/2008
MSH 15-52	PWN	Small	9.1 h	2004
RX J1713.7-3946	SNR	Large	7.0 h	
Off data	-			

- Notes
  - All data were taken wi… …st; with the four small telescopes)
  - "std" configuration: "Hil… style" analysis
    ➤ no sophisticated gamma-hadron separation or reconstruction techniques
  - Quality of instrument response functions not always optimal
  - Use for scientific publications not permitted

Please, make use of this and let us know what should be improved!



H.E.S.S. “FITS validation paper”

- Technical paper with two main goals:
  - Validation of FITS data and open-source tools
  - Exploration of template-based, 3D (“cube-style”) likelihood analysis
- Strategy:
  - Perform “standard” high-level analyses with open-source tools and compare to H.E.S.S.-internal analysis software
  - Develop new background acceptance model and apply it in 3D likelihood analysis
  - Use public test data
    → well-understood sources
    with different morphologies

Towards the usage of open-source science tools in very-high-energy γ-ray astronomy: development of a background model and validation of the tools using public H.E.S.S. data

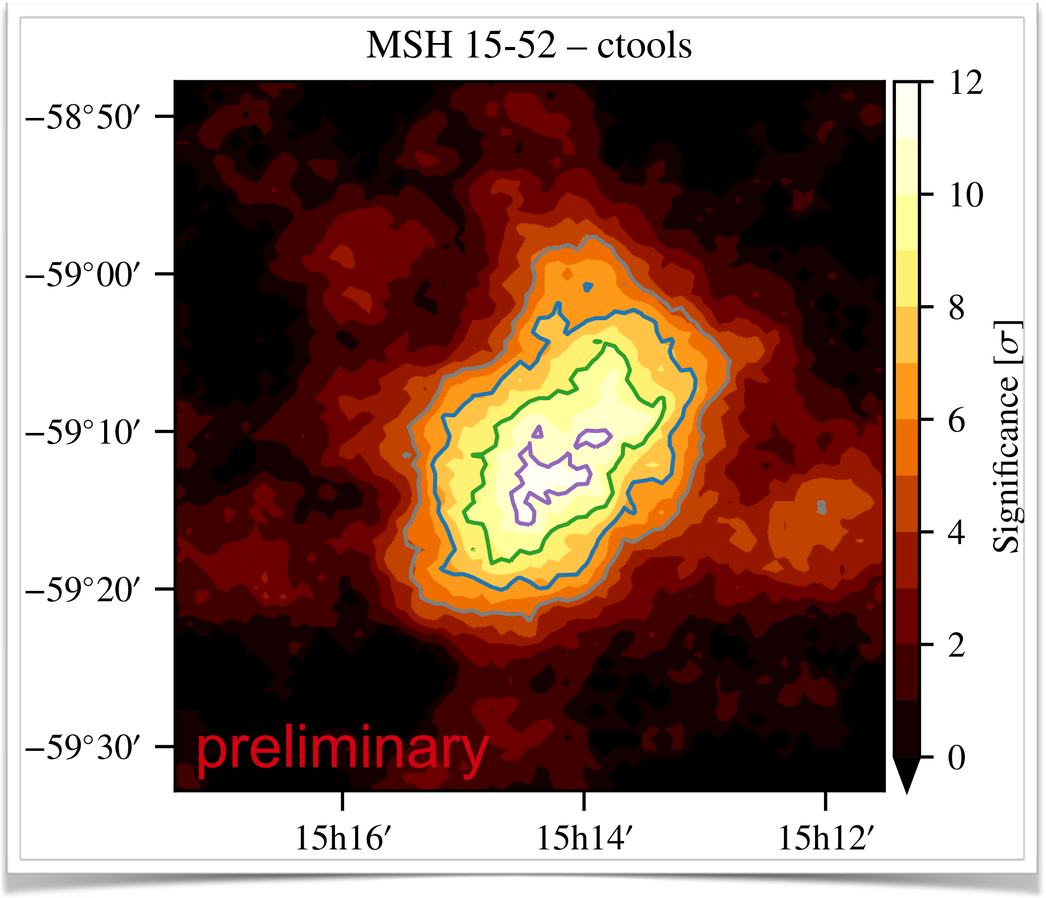
ABSTRACT

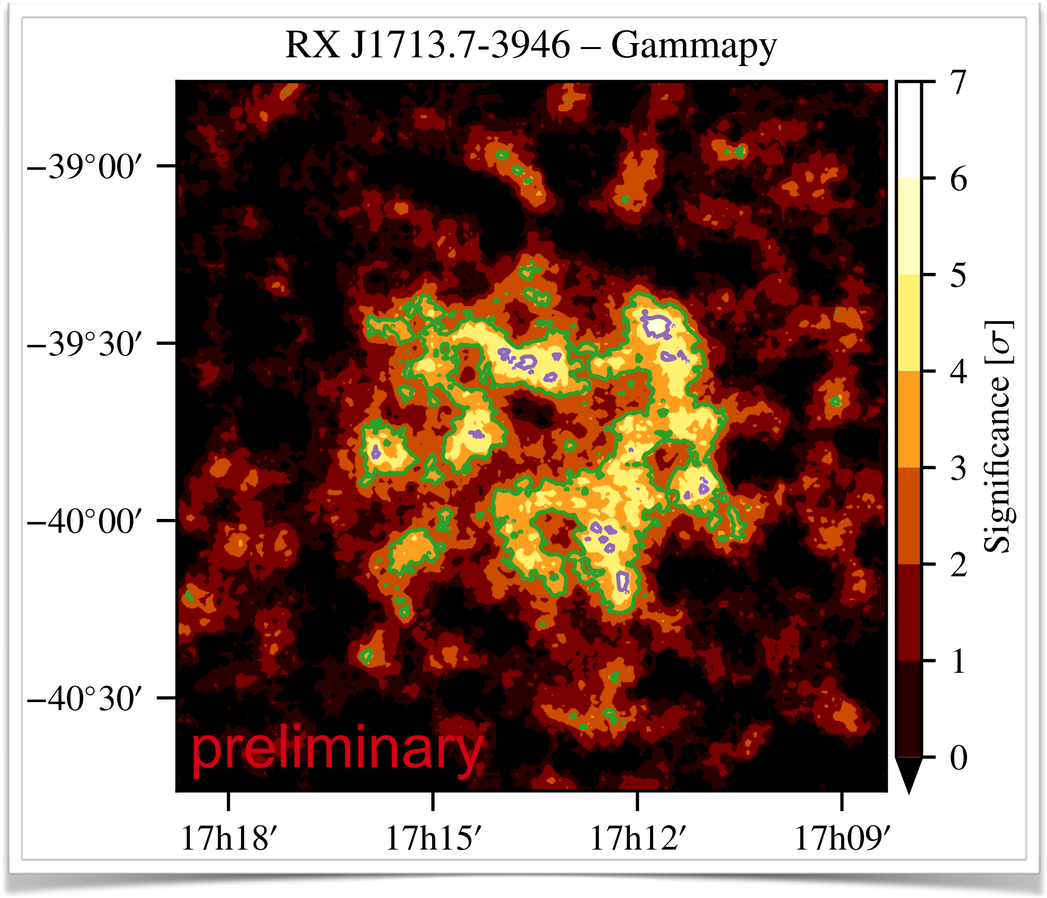
In classical analyses of gamma-ray data from Imaging Atmospheric Cherenkov Telescopes (IACT) such as H.E.S.S., aperture photometry, i.e. photon counting, is applied in a (typically circular) region of interest (RoI) encompassing the source. A key element in the analysis is to estimate the amount of background in the RoI due to residual cosmic ray-induced air showers in the data. Various standard background estimation techniques have been developed in the last decades, most of them relying on a measurement of the background from source-free regions within the observed field-of-view. However, in particular in the Galactic plane, source analysis and background estimation are hampered by the large number of (sometimes overlapping) gamma-ray sources and large-scale diffuse gamma-ray emission.
For complicated field-of-views, template analysis might be superior to classical analysis. In the template approach, a spectromorphological model template, consisting of one or multiple source components and a background component, is fit to the gamma-ray data, resulting in a complete spectral and spatial description of the field-of-view. For the application to IACT data, the major challenge of such approach is the construction of a robust background model.
In this work, we apply template analyses to various test data recently made public by the H.E.S.S. collaboration, using the open analysis frameworks *ctools* and *Gammapy*. First, we show that, when using these tools in a classical analysis approach and comparing to the proprietary H.E.S.S. analysis framework, virtually identical high-level analysis results, such as field-of-view maps, spectra and lightcurves, are obtained. We then describe the construction of a generic background model from data of H.E.S.S. observations, and demonstrate that template analysis using this background model yields high-level analysis results that are highly compatible with those obtained from the classical analyses. This validation of the template analysis approach on experimental data is an important step towards using this method for IACT data analysis, and in particular for the analysis of data from the upcoming Cherenkov Telescope Array (CTA).

- Important not only to increase confidence in open-source tools within H.E.S.S. Collaboration
  - Also demonstrates capability of tools to work well with “real” IACT data
    (in contrast to the CTA data challenge, which used simulated data)

H.E.S.S. “FITS validation paper”: maps

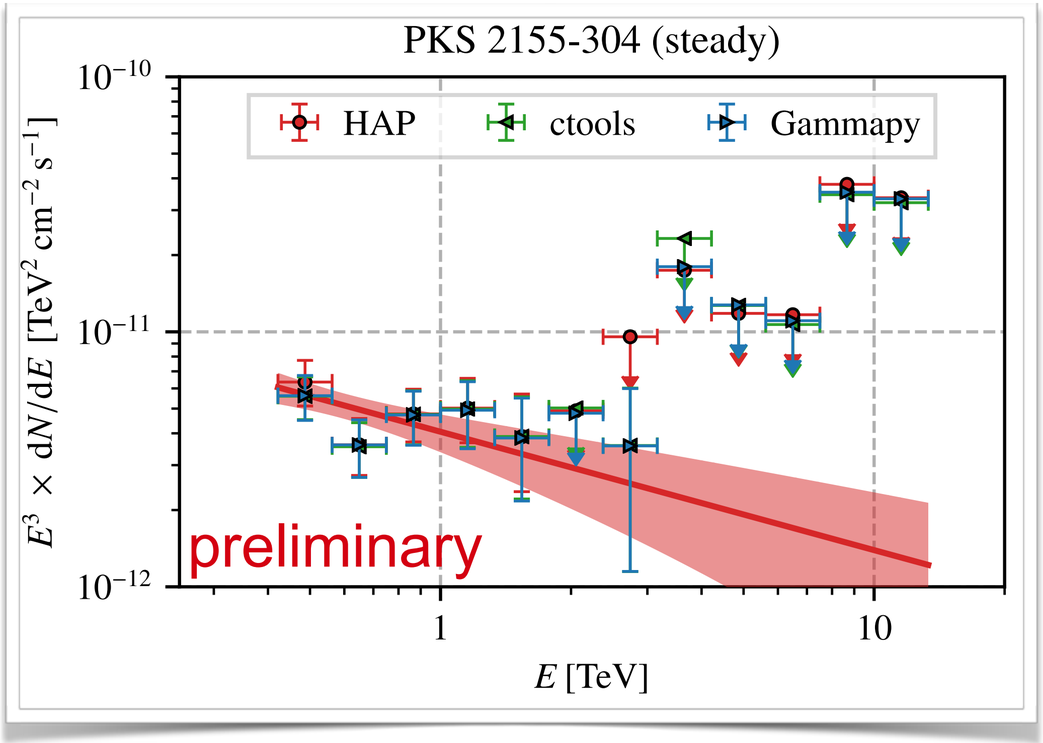
- Significance maps derived with *ring background* method
  - Background map: open-source tool
  - Contour lines: H.E.S.S.-internal analysis software

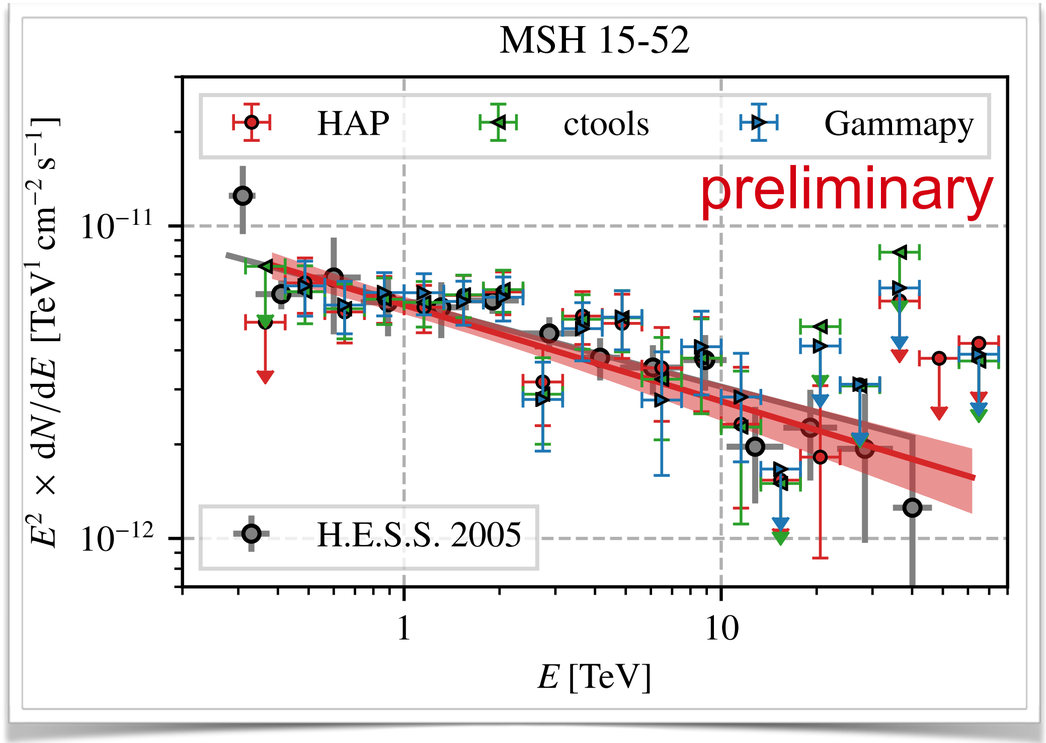




H.E.S.S. “FITS validation paper”: spectra

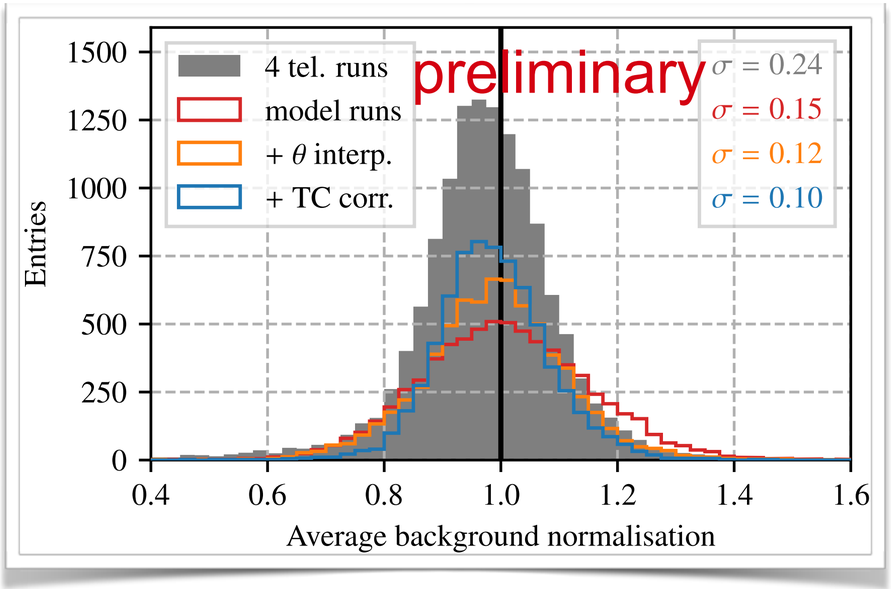
- Energy spectra derived with *reflected background* method
  - Results virtually identical when using the same algorithms
  - Compatible with published spectra

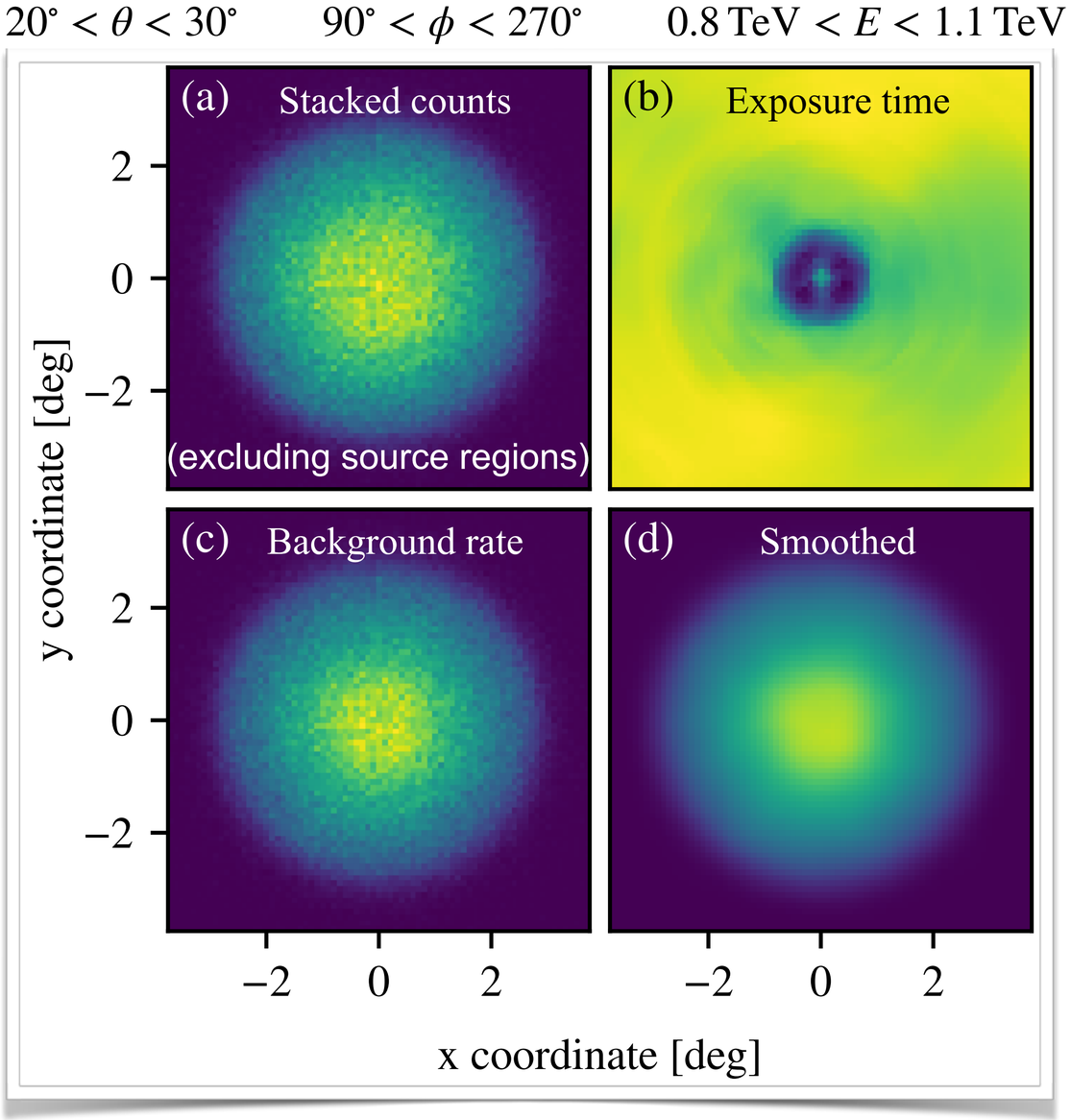




H.E.S.S. “FITS validation paper”: background model

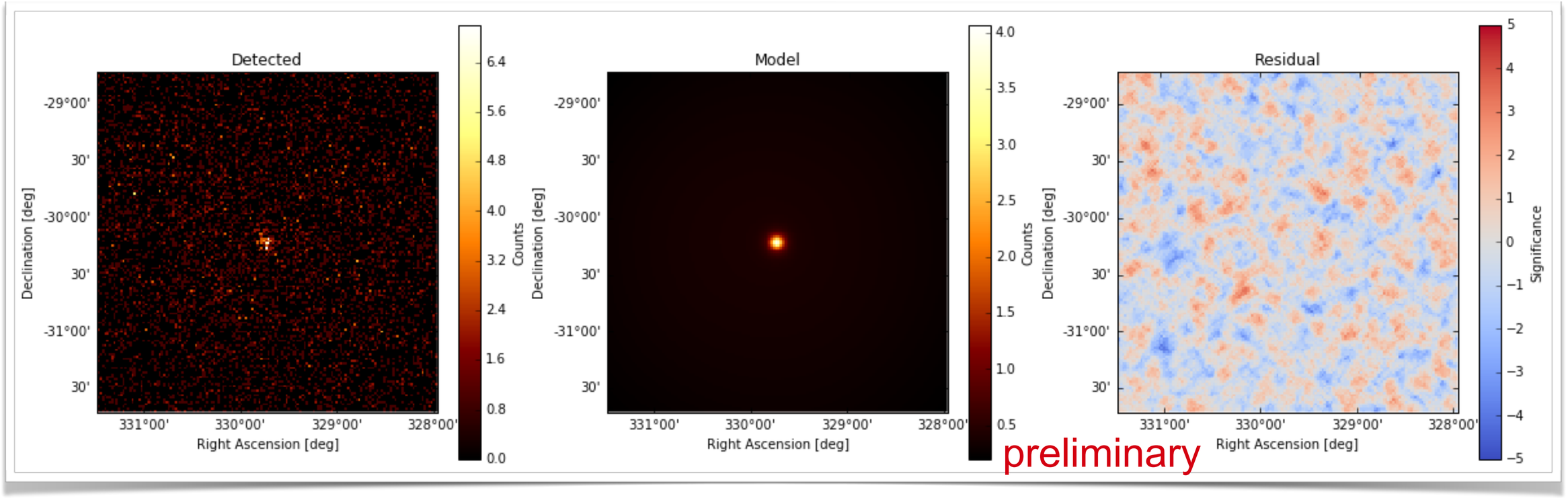
- Strategy
  - Determine background from archival data
  - Bin observations according to zenith / azimuth of pointing
  - Average rate in field-of-view coordinate system (excluding contributions from source regions)
  - Transform to celestial coordinate system
  - Apply smoothing algorithm
  - Correct for precise zenith angle and atmospheric transparency





H.E.S.S. “FITS validation paper”: 3D likelihood analysis

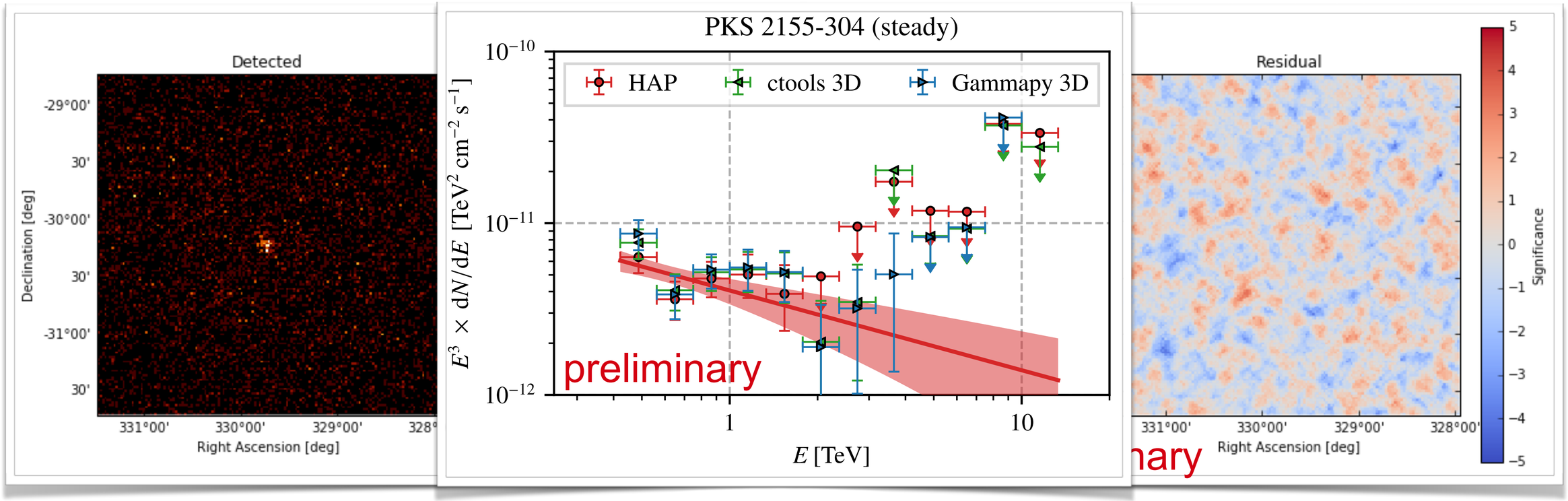
- Application example: PKS 2155-304



- Extended sources are work in progress…

H.E.S.S. “FITS validation paper”: 3D likelihood analysis

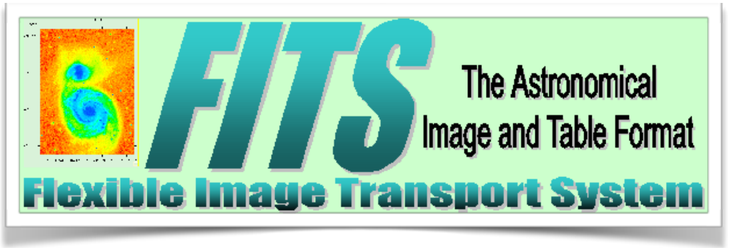
- Application example: PKS 2155-304

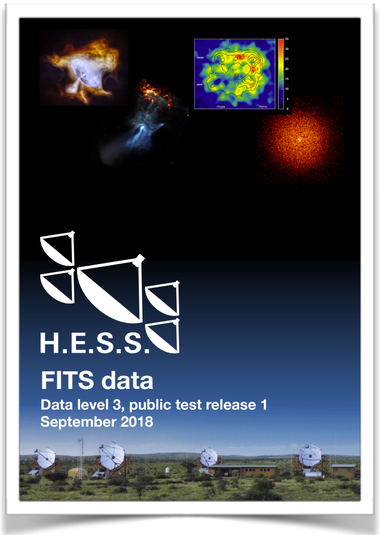


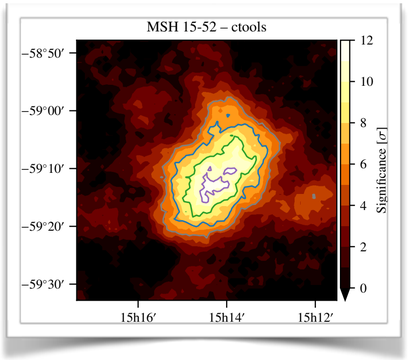
- Extended sources are work in progress…

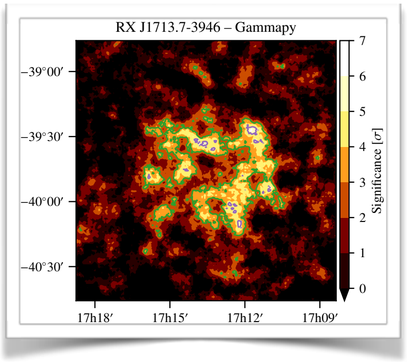
Summary

- H.E.S.S. data analysis with open-source tools: motivation and status
- Technical requirements
  - Open data format specifications
  - Status and problems of FITS exporter
  - Ideas for improvements (IRF quality, energy thresholds, background acceptance)
- What are we doing to make H.E.S.S. data analysis with open-source tools reality?
  - H.E.S.S. public test data release
    - small release of data taken on well-known sources
    - provides IACT data in FITS format for the first time
    - support development of format specifications and open-source analysis tools
  - H.E.S.S. FITS validation paper
    - validate results obtained with open-source analysis tools
    - explore 3D likelihood analysis provided by open-source tools
    - coming soon!



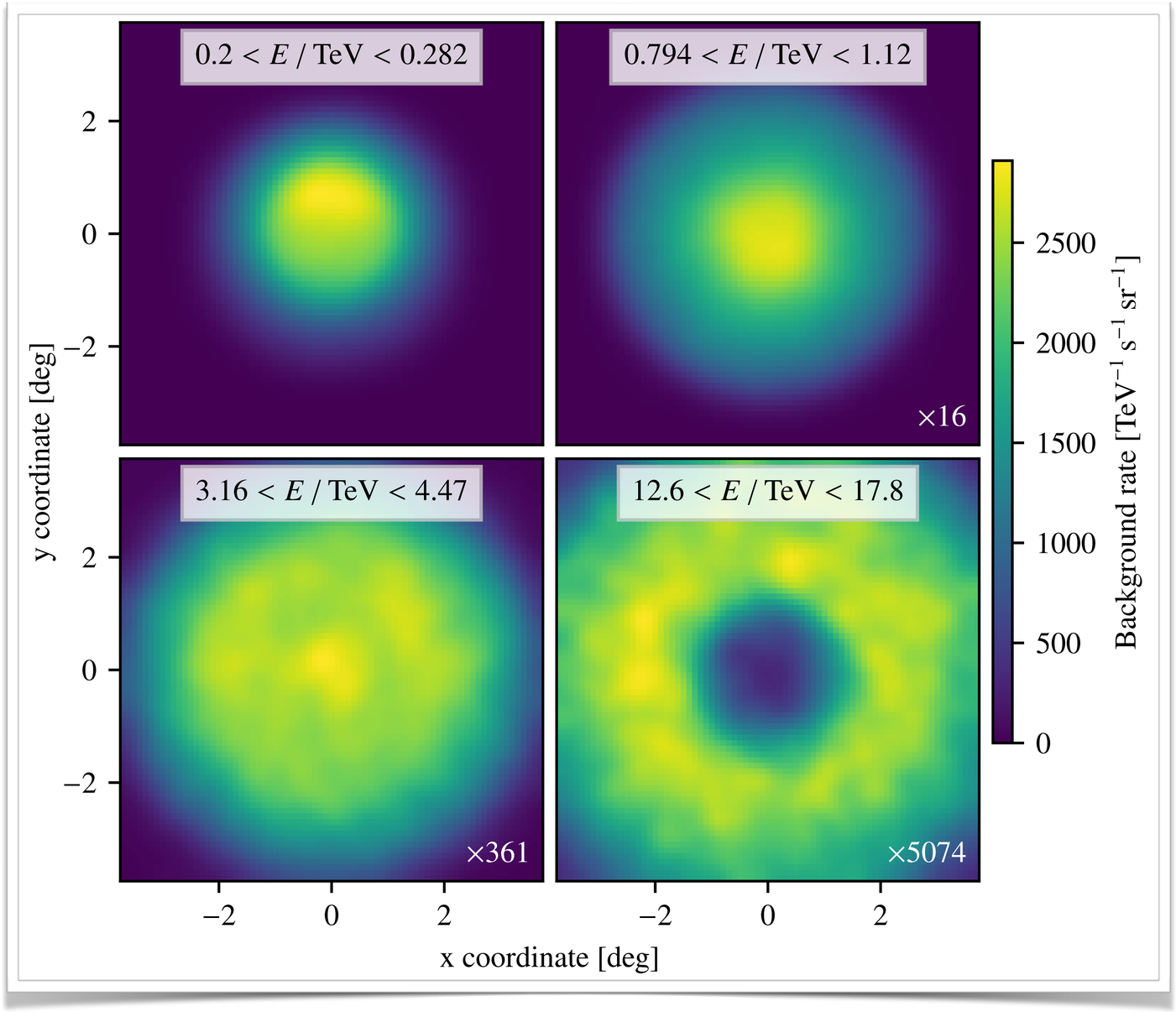






Backup slides

- Background model at different energies



Backup slides

- Correction for atmospheric transparency