



Dark matter searches with MAGIC in dSph with simulation-based inference

— step 1: a generative cosmic-ray background model —

in collaboration with: Miguel Molero Gonzalez, Jorge Martín Camalich

Christopher Eckner (christopher.eckner@iac.es)

UNDARK Postdoctoral Researcher

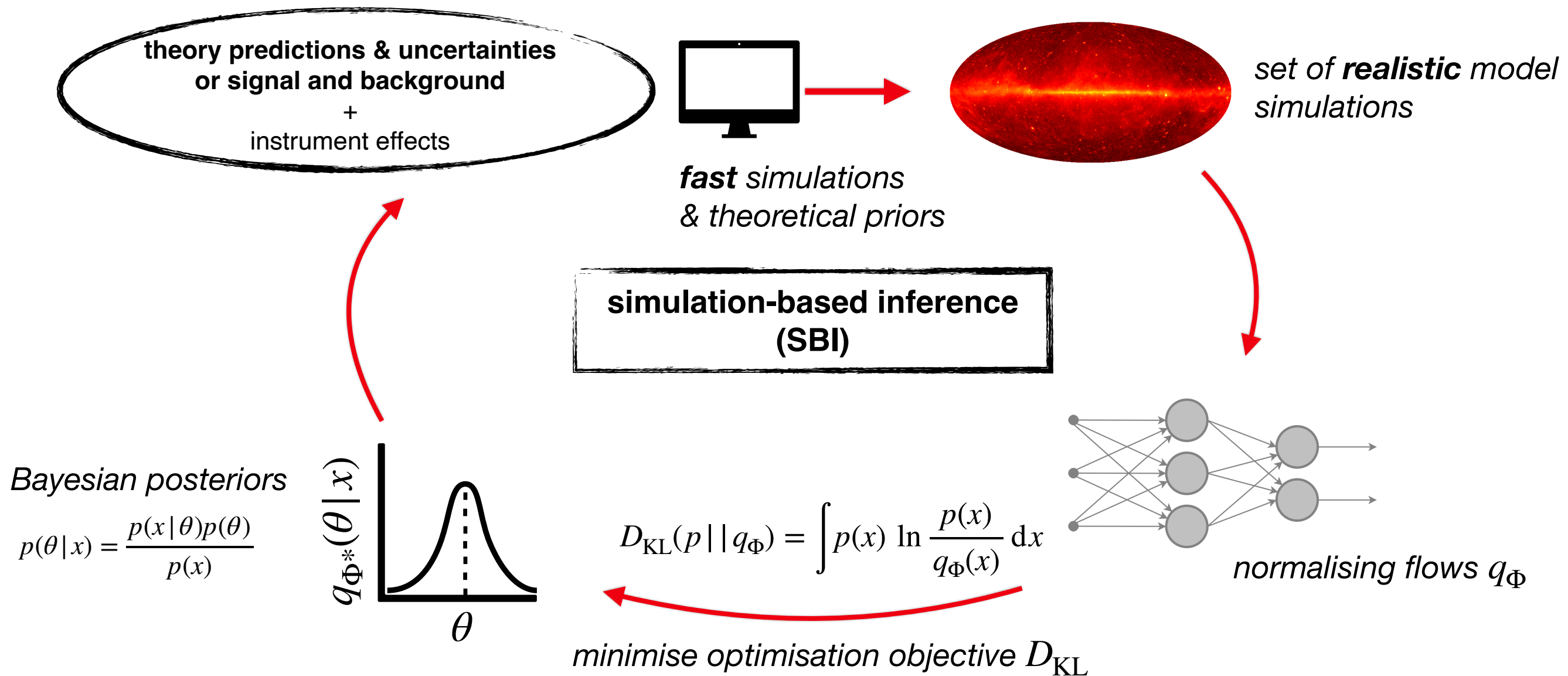
Instituto de Astrofísica de Canarias, Tenerife, Spain

Monthly MAGIC+LST Fundamental Working Group Meeting

24th of September 2026 | online

General scope of the project

Overall objective: Perform a joint analysis of Fermi-LAT and MAGIC dwarf spheroidal galaxy datasets (3D) using **simulation-based inference**.

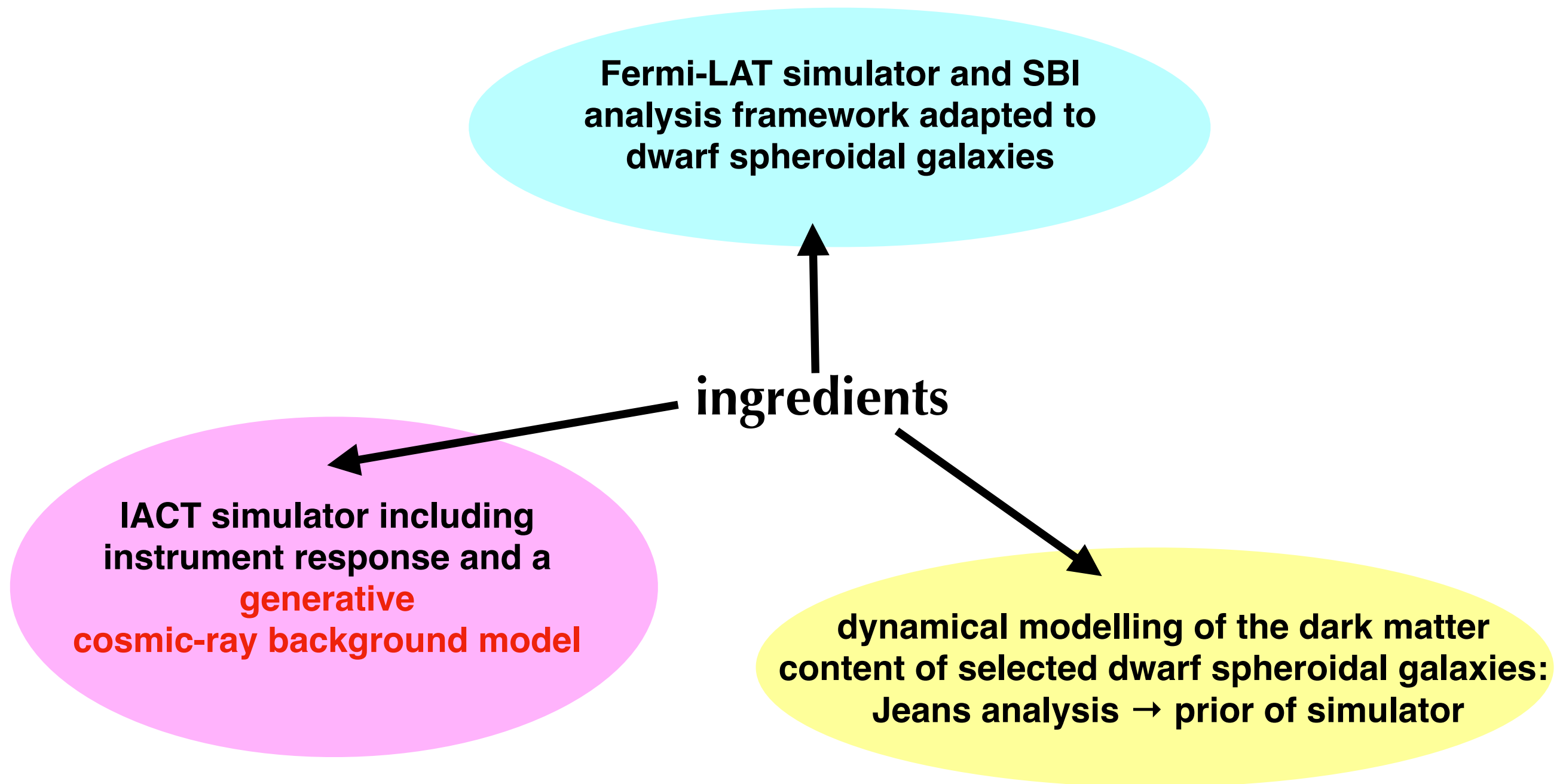


Key benefits:

- **Efficient:** Handles high-dimensional parameter spaces. Marginalisation over nuisance parameters.
- **Robust:** Intrinsic calibration of posteriors.

General scope of the project

Overall objective: Perform a joint analysis of Fermi-LAT and MAGIC dwarf spheroidal galaxy datasets (3D) using simulation-based inference.

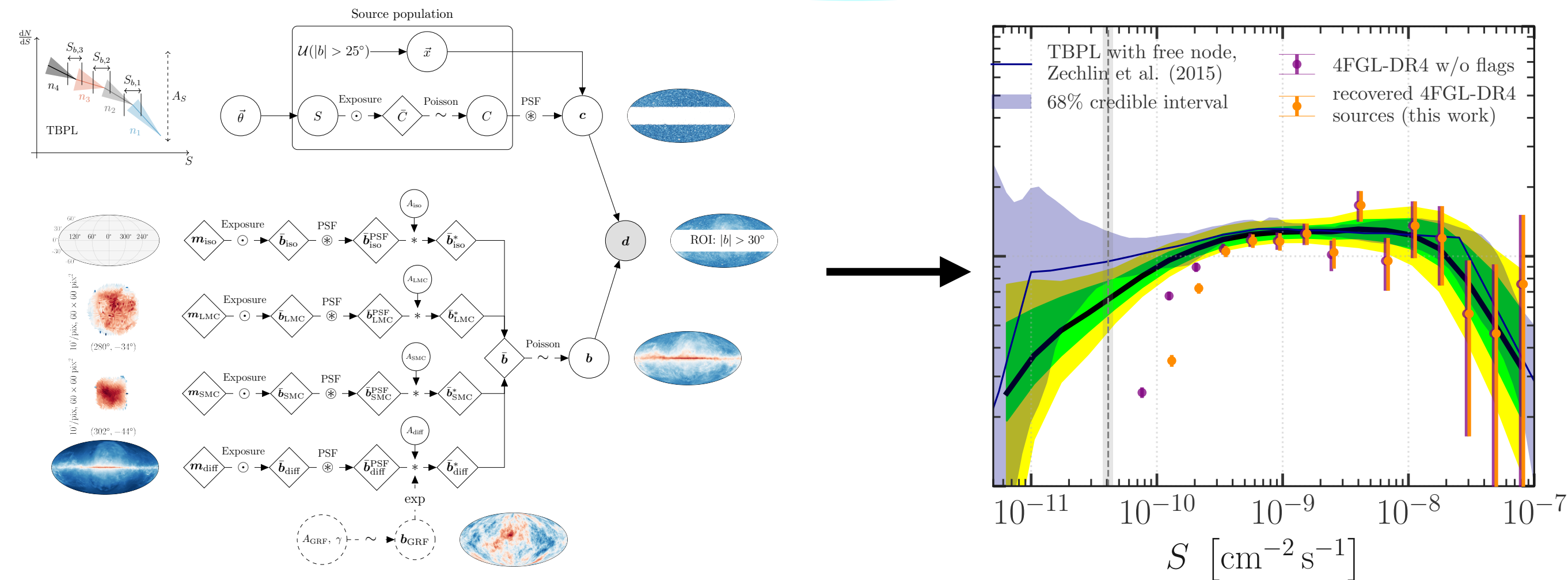


General scope of the project

Overall objective: Perform a joint analysis of Fermi-LAT and MAGIC dwarf spheroidal galaxy datasets (3D) using simulation-based inference.

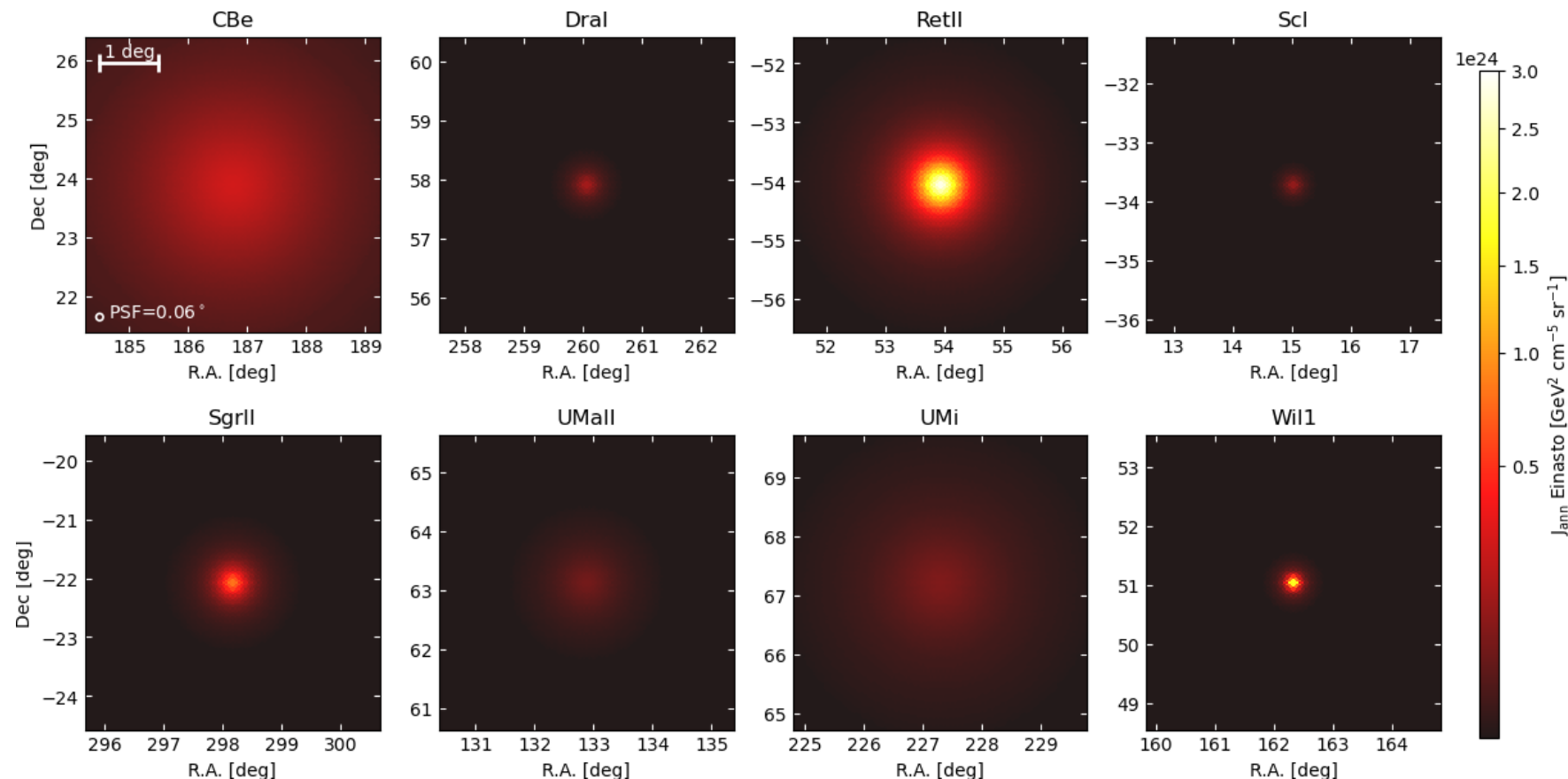
Fermi-LAT simulator and SBI analysis framework adapted to dwarf spheroidal galaxies

[Eckner, Anau Montel, List et al.; *PRD* 112 (2025) 10, 103022]



General scope of the project

Overall objective: Perform a joint analysis of Fermi-LAT and MAGIC dwarf spheroidal galaxy datasets (3D) using simulation-based inference.



[The CTAO Consortium, MNRAS 544 (2025) 3]

**dynamical modelling of the dark matter
content of selected dwarf spheroidal galaxies:
Jeans analysis → prior of simulator**

[critical expertise at the IAC]

General scope of the project

Overall objective: Perform a joint analysis of Fermi-LAT and MAGIC dwarf spheroidal galaxy datasets (3D) using simulation-based inference.

Fermi-LAT simulator and SBI analysis framework adapted to dwarf spheroidal galaxies
[Eckner, Anau Montel, List et al.; *PRD* 112 (2025) 10, 103022]

ingredients

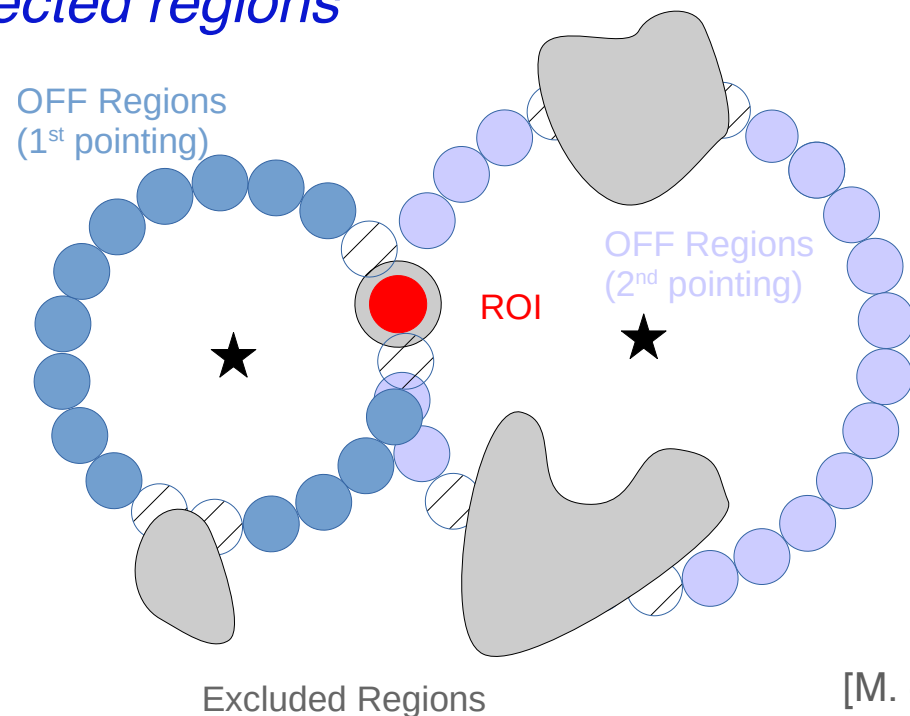
IACT simulator including instrument response and a **generative cosmic-ray background model**
TODAY!

dynamical modelling of the dark matter content of selected dwarf spheroidal galaxies:
Jeans analysis → prior of simulator
[critical expertise at the IAC]

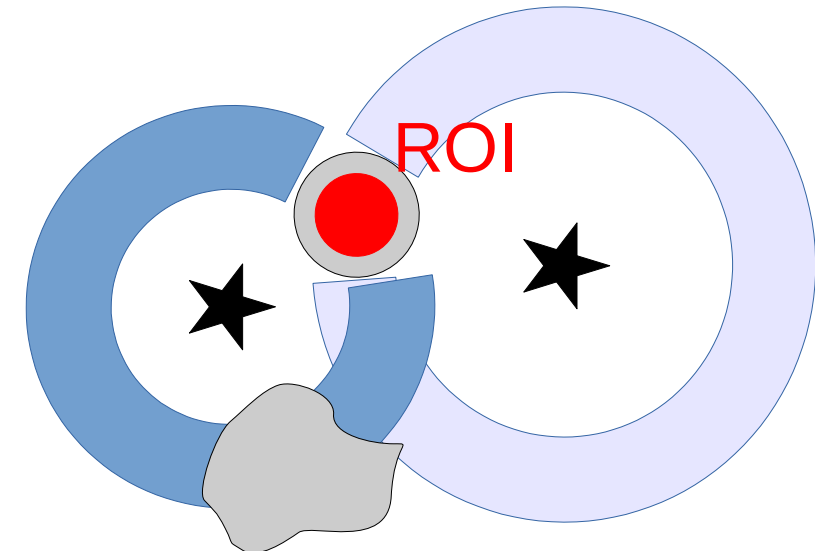
Towards a generative cosmic-ray background model

As you all know, in IACT analyses we may use a variety of established methods to model/estimate the cosmic-ray background in the ROI:

reflected regions



ring background



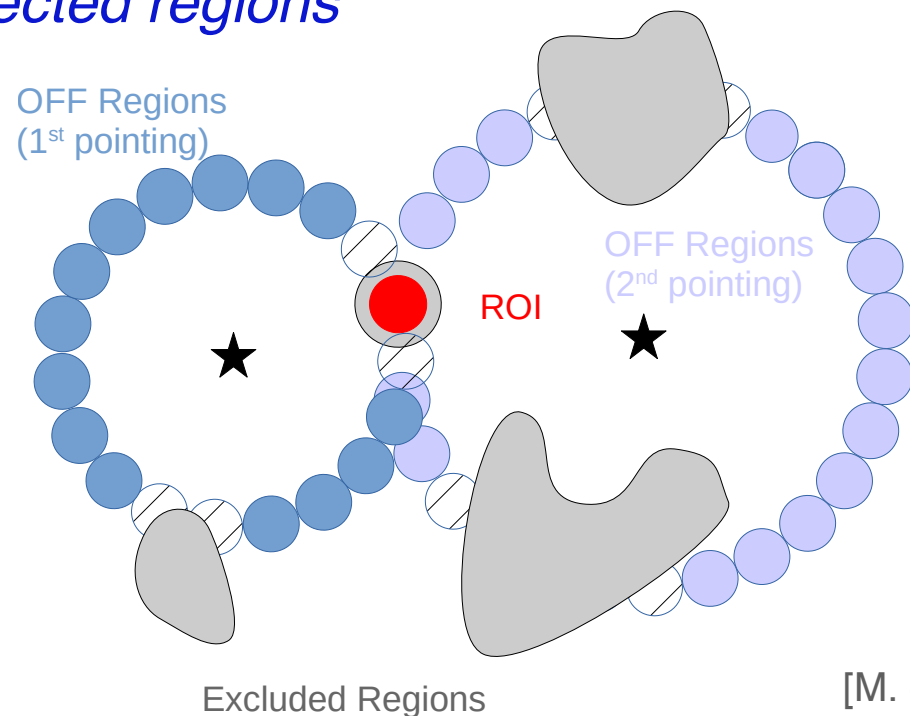
[M. de Naurois; *Universe* 2021, 7, 421]

- Need a signal-free control region at some matched offset
- Yield a number but not a map → mostly 1D analysis
- Fail when the source fills the FoV

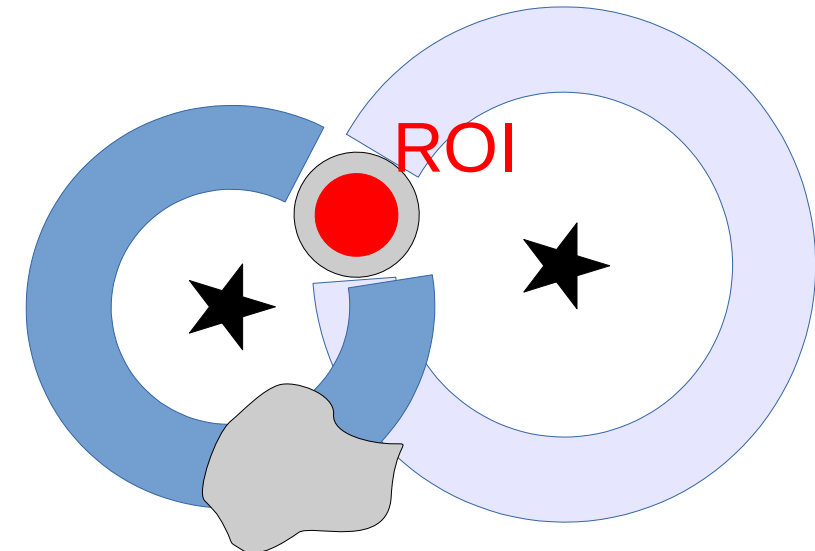
Towards a generative cosmic-ray background model

As you all know, in IACT analyses we may use a variety of established methods to model/estimate the cosmic-ray background in the ROI:

reflected regions

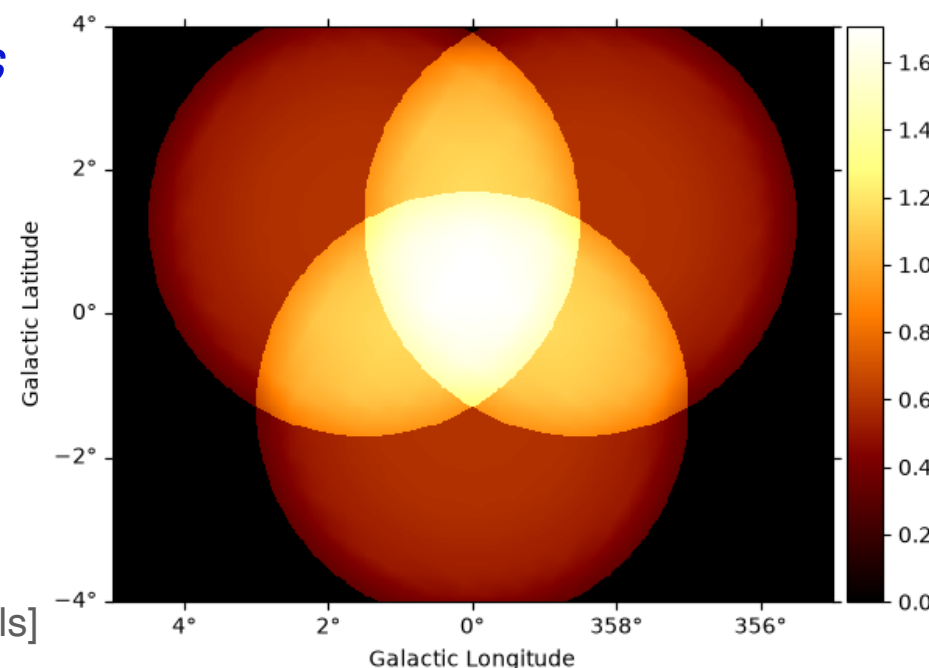


ring background



[M. de Naurois; *Universe* 2021, 7, 421]

template-based methods



[credit: gammapy tutorials]

- Needs a background model over the whole FoV (and full enclosure IRFs for 3D analyses)
- DL3 carries no uncertainty on that model

Towards a generative cosmic-ray background model

As you all know, in IACT analyses we may use a variety of established methods to model/estimate the cosmic-ray background in the ROI:

reflected regions



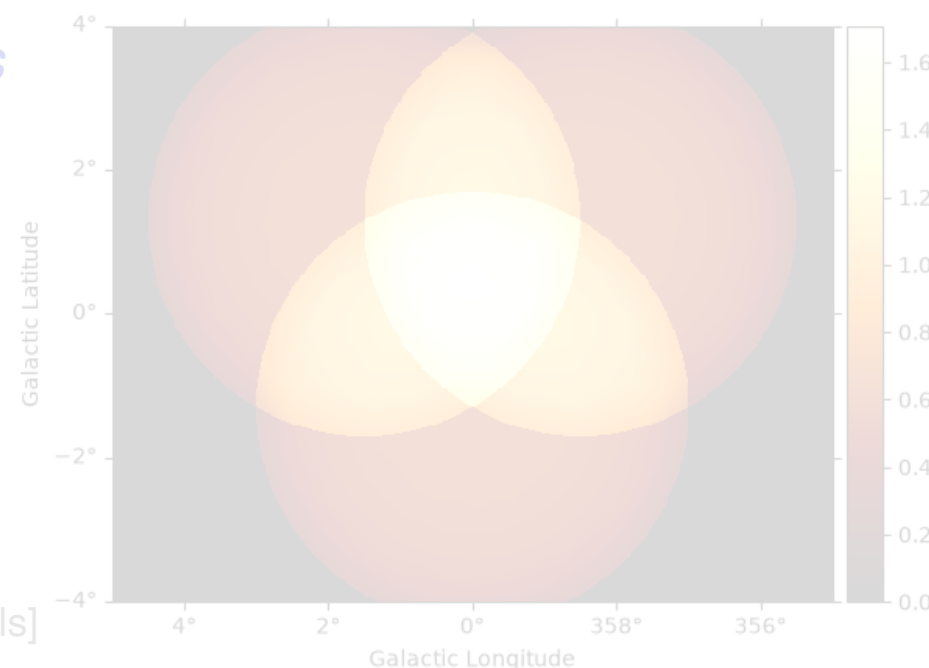
ring background



A machine-learning-driven different approach:

A background model evaluable anywhere in the camera frame and energy, conditioned on the observing conditions, with a calibrated uncertainty a source fit/analysis can propagate.

template-based methods



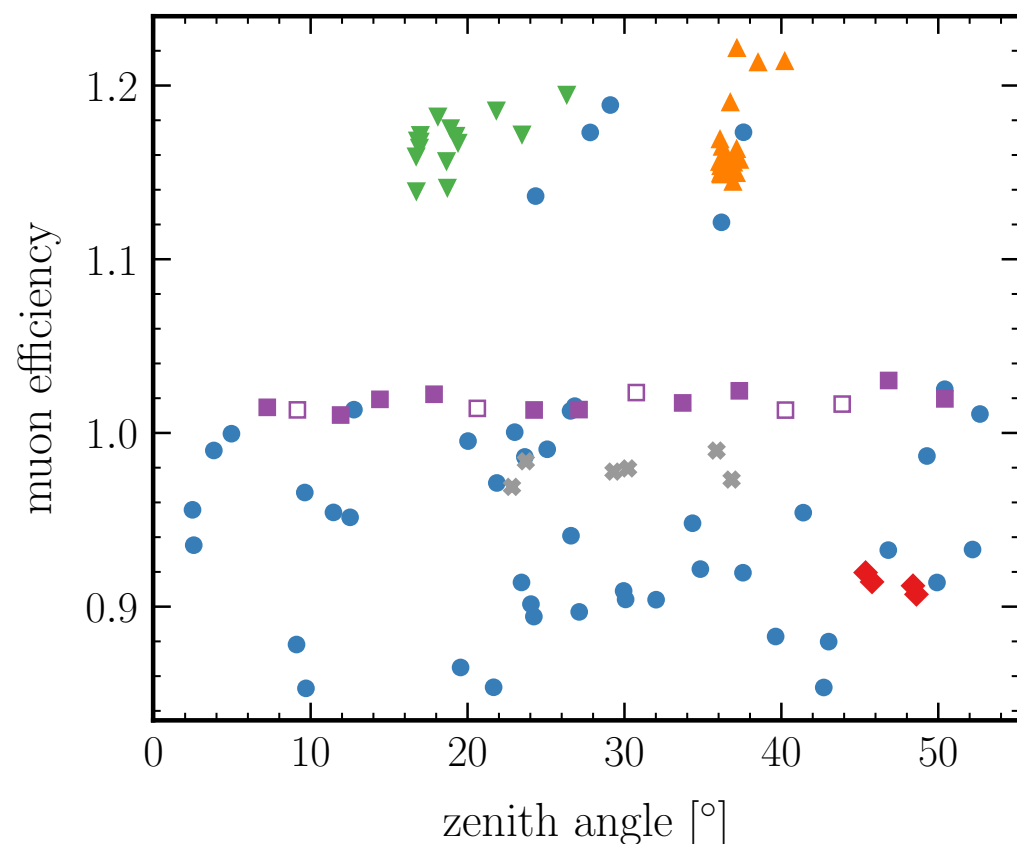
[credit: gammapy tutorials]

- Needs a background model over the whole FoV (and full enclosure IRFs for 3D analyses)
- DL3 carries no uncertainty on that model

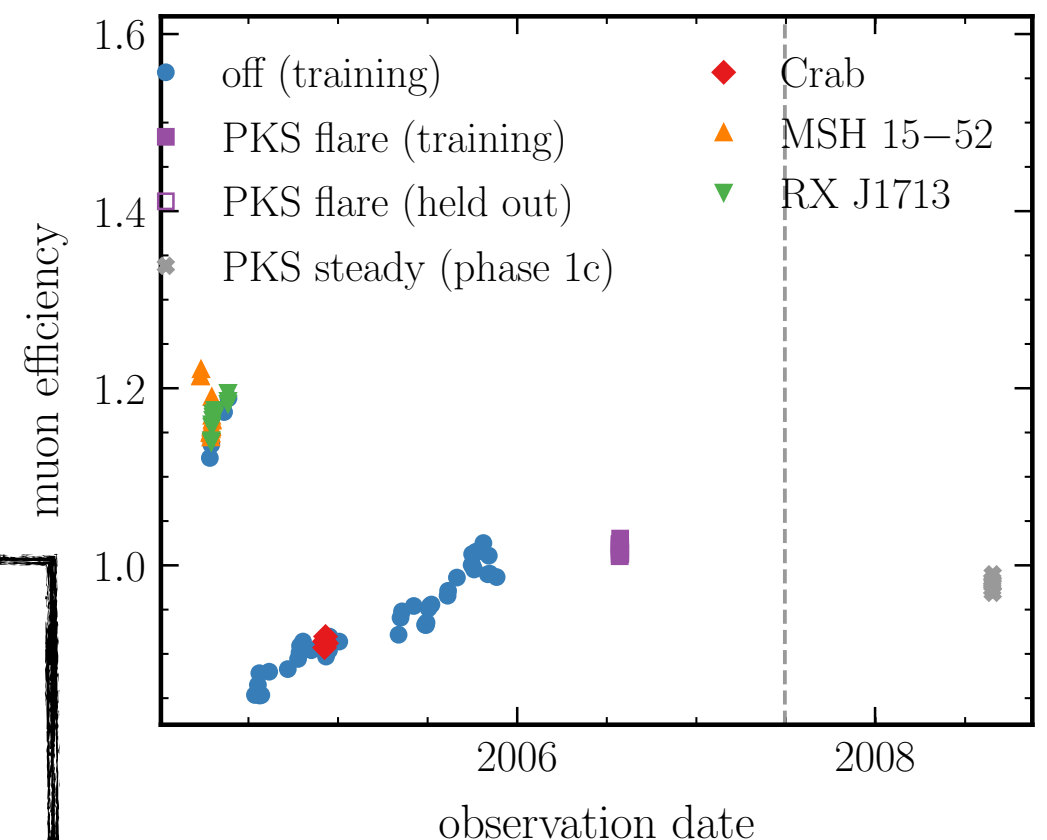
Towards a generative cosmic-ray background model

The prototype for understanding if it works will use the public H.E.S.S. dataset (because it contains dark patches; 4 telescopes + standard Hillas cuts):

dataset	runs	live [h]	zenith [°]	muon eff.	E_{safe} [TeV]	dates
Off data	45	19.23	2.5–52.7	0.853–1.189	0.25–1.32	04/2004–11/2005
PKS 2155–304 flare	15	6.58	7.2–50.4	1.010–1.030	0.25–1.07	07/2006
PKS 2155–304 steady	6	2.61	22.8–36.8	0.969–0.990	0.38–0.62	08/2008
Crab	4	1.75	45.4–48.6	0.907–0.920	0.66–0.87	12/2004
MSH 15–52	20	8.33	36.1–40.2	1.145–1.221	0.38–0.41	03–04/2004
RX J1713.7–3946	15	6.29	16.7–26.3	1.139–1.195	0.20–0.25	04–05/2004



Nothing about
atmospheric
conditions at
observation time.



Model definition and training

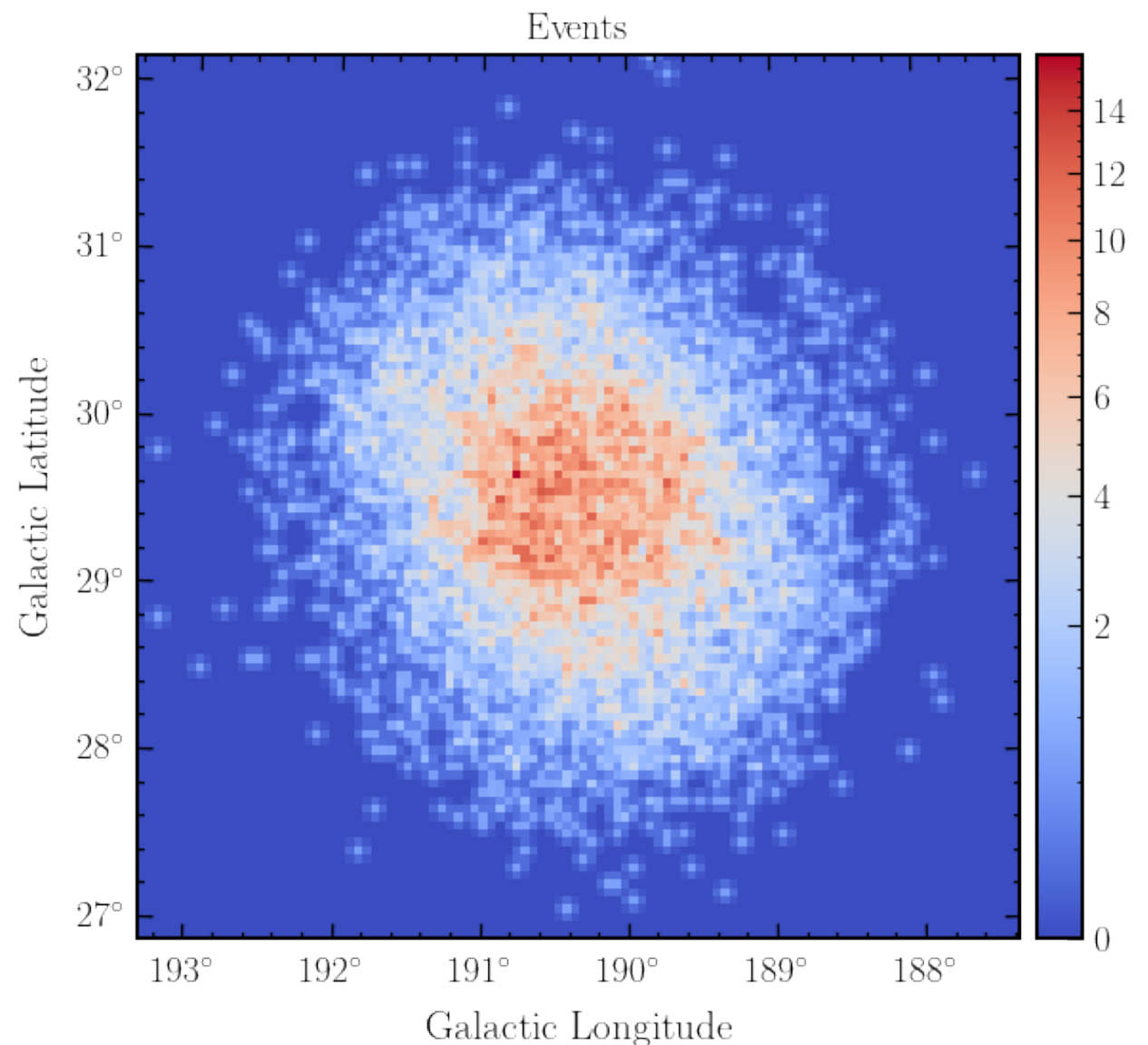
Definition of the model

In the field-of-view frame, the cosmic-ray background is nothing but an **inhomogeneous Poisson point process with position and energy-dependent intensity λ** .

→ However, everything needs to be conditioned on the observation properties.

$$\lambda(u | c) = \rho(c) \times p(u | c)$$

- $u = (x, y, \log_{10} E_{\text{reco}}) \rightarrow$ 3D info
- c : observation conditions
- ρ : one total rate per run
- p : normalised acceptance shape



Observation conditions for H.E.S.S. dataset:

$c = (\cos \theta_{\text{zen}}, \sin \varphi_{\text{az}}, \cos \varphi_{\text{az}}, \varepsilon_{\text{opt}})$, **muon efficiency** ε_{opt} : related to camera throughput; **zenith angle** θ_{zen} : impacts threshold and rate via traversed atmospheric column; **azimuth** φ_{az} : its angle to the shower broadens images due to the geomagnetic field

Definition of the model

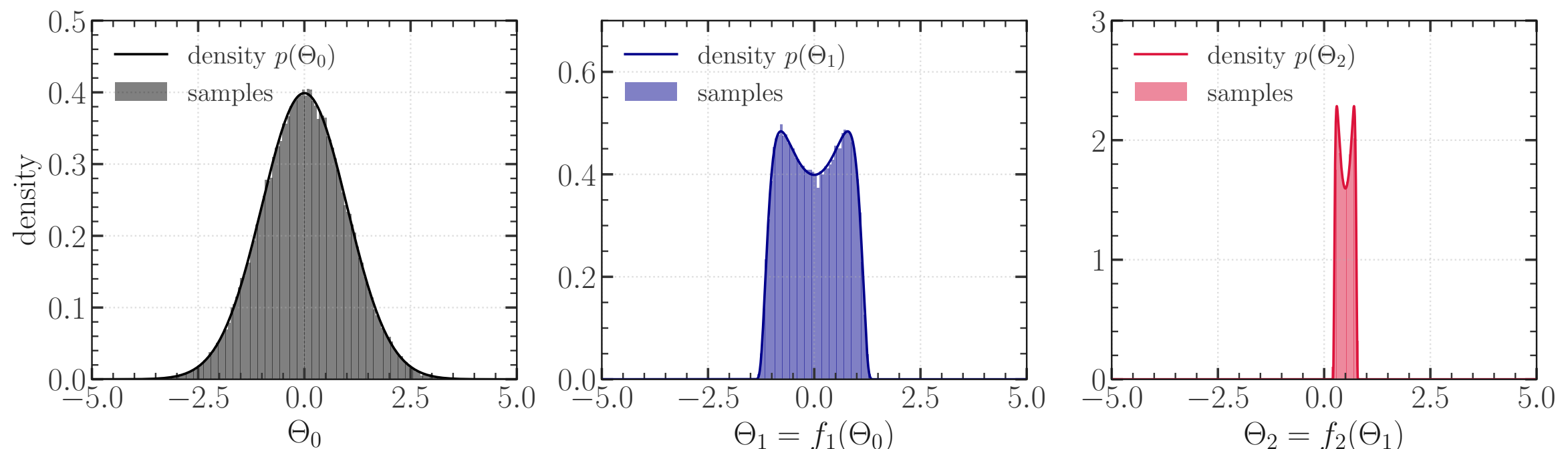
The acceptance shape p will be modelled via a **normalising flow** and conditioned on the observation conditions. The rate ρ is its own feed-forward network.

→ Density distribution answering the question what is the probability to observe an event with this energy in this field-of-view position under these conditions?

What is a normalising flow?

A neural network architecture that transforms a simple base distribution into a complex one via a chain of invertible, differentiable maps. Given $\Theta_0 \sim p(\Theta_0)$ (e.g. standard normal), define:

$$x = \Theta_K = f_K \circ \dots \circ f_1(\Theta_0)$$
$$\log p(x) = \log p(\Theta_0) - \sum_{k=1}^K \log \left| \det J_{f_k}(\Theta_{k-1}) \right|$$



Training and calibration of the model

What do we want the generative model to do?

→ The Poisson mean μ at any infinitesimal volume δ in the FoV coincides with the observed Poisson scattered events in this volume element.

$$\mu = \int_{\delta} \lambda(u | c) du \quad \xrightarrow[\text{with } N_k \text{ events:}]{\text{per run } k} \quad \ln L = \ln \mu_k - \underbrace{\sum_i^{N_k} \ln [\rho(c_k) p(u_i | c_k)]}_{\text{minimise}}$$

Implementation of systematic uncertainty via calibration:

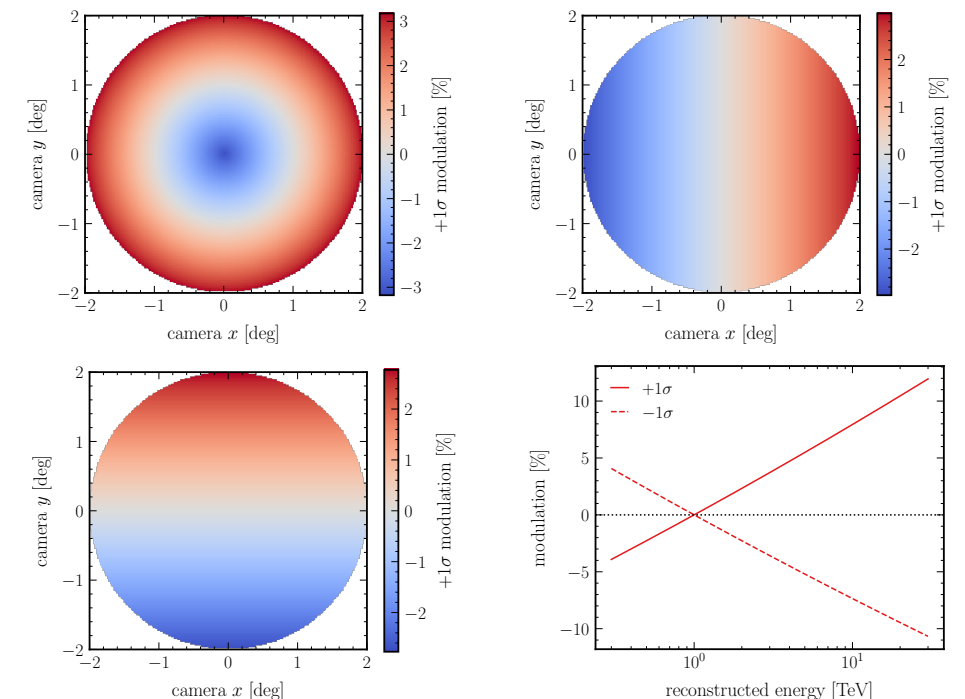
→ Allow for additional scatter of the of the trained intensity λ to accommodate systematics.

→ This it what makes it a probabilistic generative model!

$$\lambda_z = \lambda \exp(B(u) \cdot z_k), \quad z_k \sim \mathcal{N}(0, \Sigma_k),$$

$$B = \left[1, \frac{r}{R_0} - 1, \frac{x}{R_0}, \frac{y}{R_0}, \frac{E - \bar{E}}{\sigma_E} \right]$$

Estimate the residual scatter from runs that were not part of the training dataset!



Generative model from the H.E.S.S. dataset

Training pool:

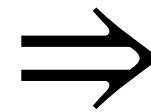
45 dark patches + 10 runs for PKS 2105-344 in flaring mode
→ PKS masked at 0.5° in camera FoV; position followed over time

Calibration pool:

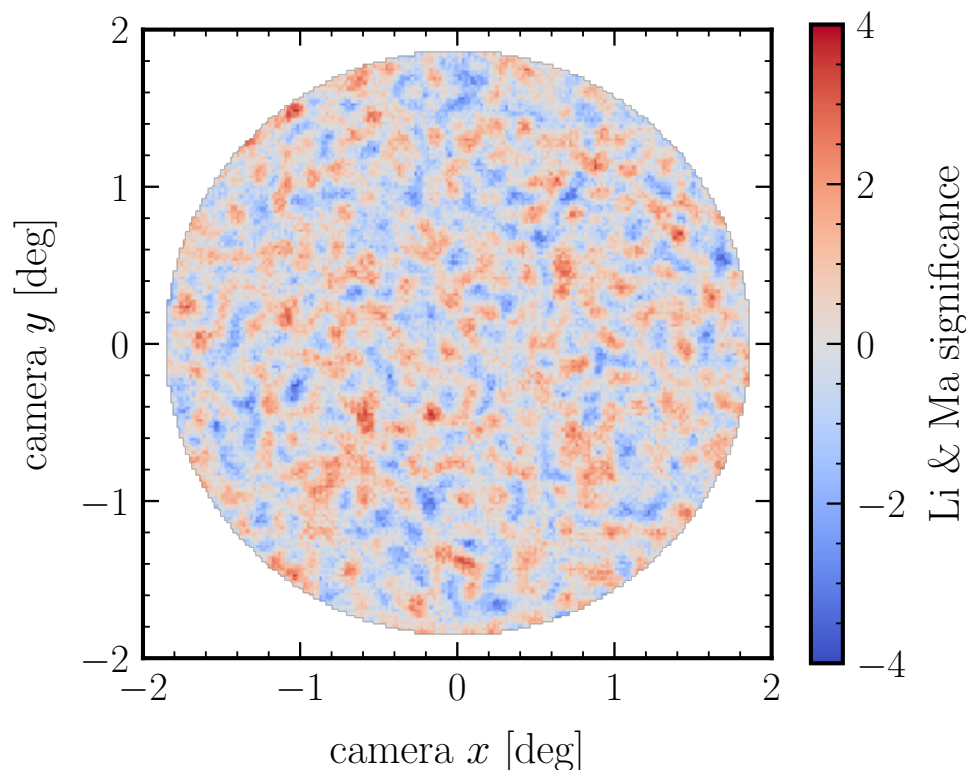
5 runs for PKS 2105-344 in flaring mode
→ PKS masked at 0.5° in camera FoV; position followed over time

Calibration:

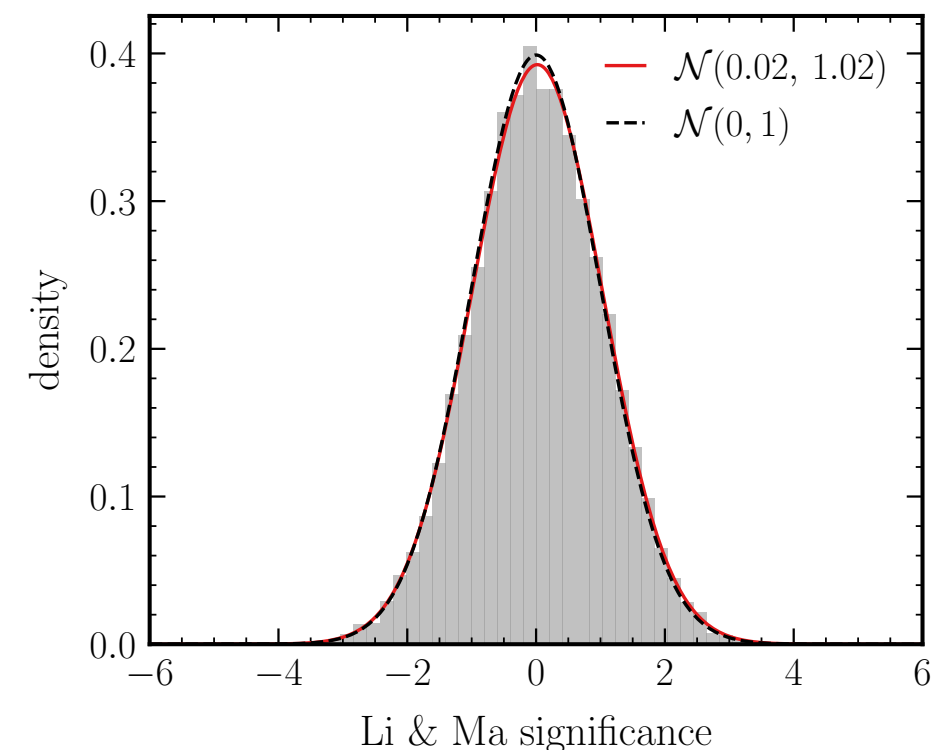
mode	1σ
normalisation	15.18 % 95 %: 11.96–17.95 %
radial	3.13 %
x / y gradient	1.48 / 1.38 %
spectral tilt	3.82 %



Given its conditions, a run's rate is still uncertain by $\sim 15\%$ and redrawn per run in the simulator.



**background
residuals on stacked
calibration runs**



The generative model performs as it should.

Application to source cases: Crab Nebula and MSH 15-52

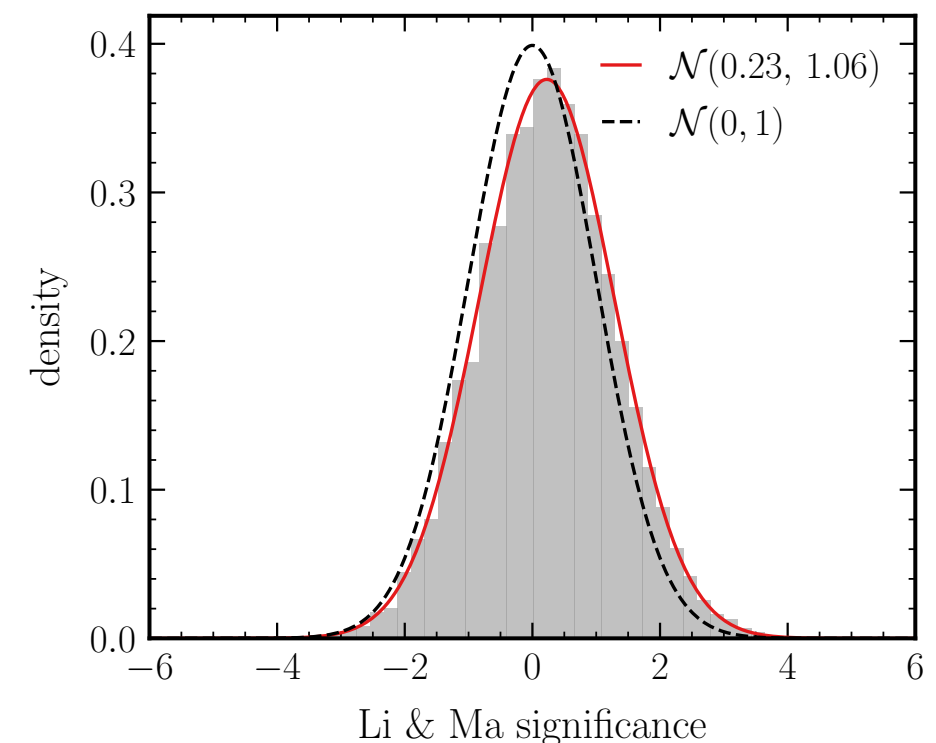
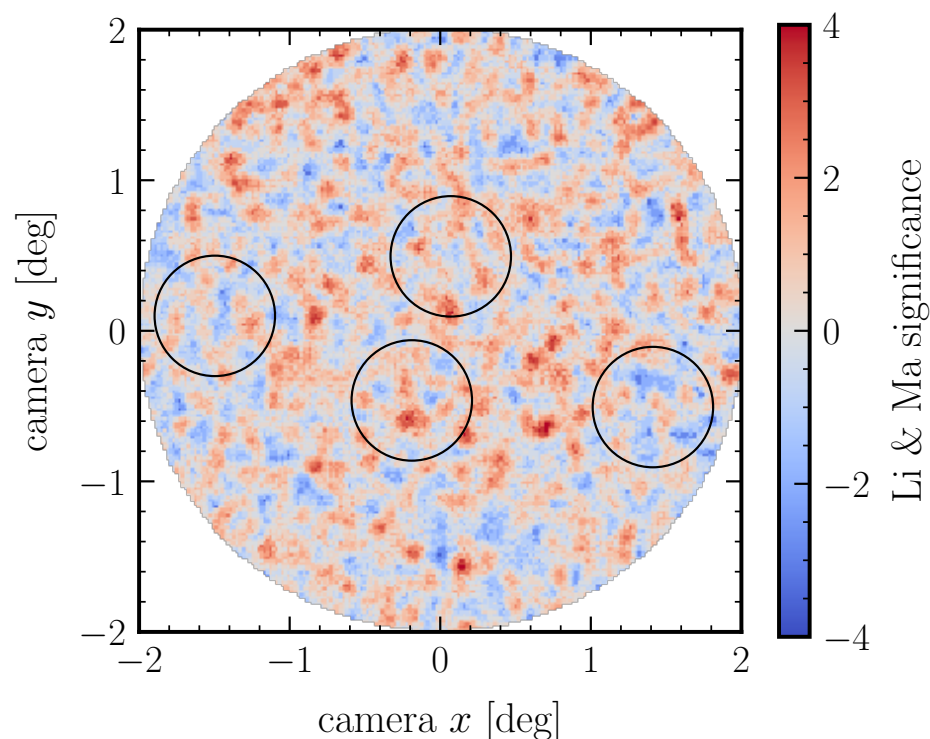
The case of the Crab Nebula in the H.E.S.S. data

run	zen.[°]	ε_{opt}	E_{safe}	offset
23523	48.6	0.907	0.87	0.5°
23526	45.8	0.914	0.71	0.5°
23559	45.4	0.920	0.66	1.5°
23592	48.4	0.912	0.87	1.5°

4–8 Dec 2004, 1.75 h live; E_{safe} in TeV

- Not in the training pool
- Mask 0.4° (worst-PSF need: 0.35°)
- Common energy band for > 0.871 TeV
- Inside the condition range but only 15% of training runs at or above its zenith
→ too few runs to train on the field

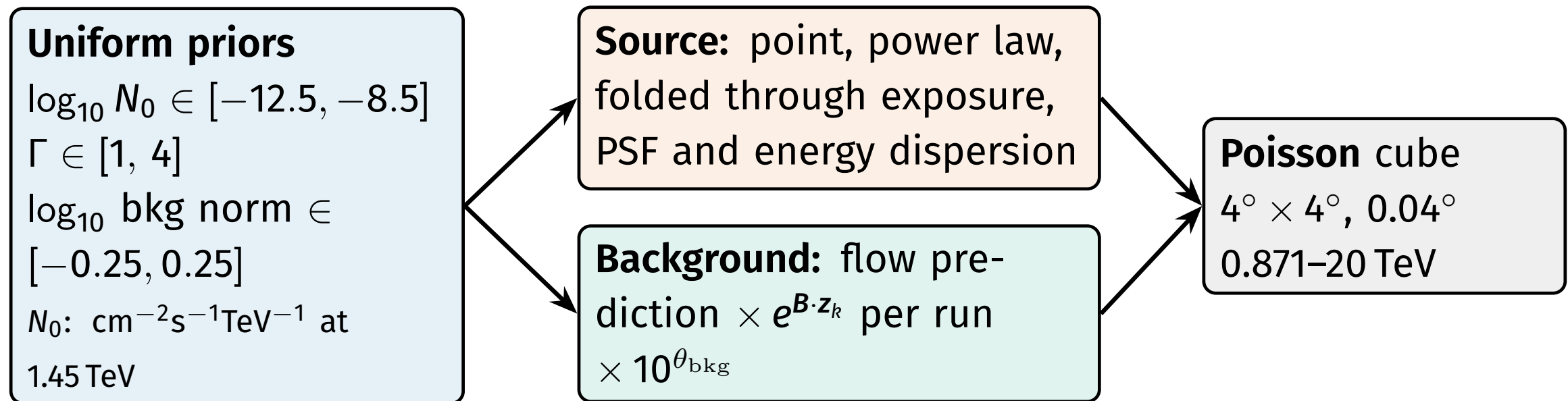
Background residuals on stacked Crab runs:



The total rate in all four runs lies above the prediction: We need a free floating background norm in the analysis, i.e., $\rho(c)$ with the available data is not sufficient for total rate prediction.

Simulation-based inference approach to the Crab

The approach needs first a **physics simulator** (generative CR model + gammapy simulation of Crab from spectral and spatial model) with **parameter priors**:



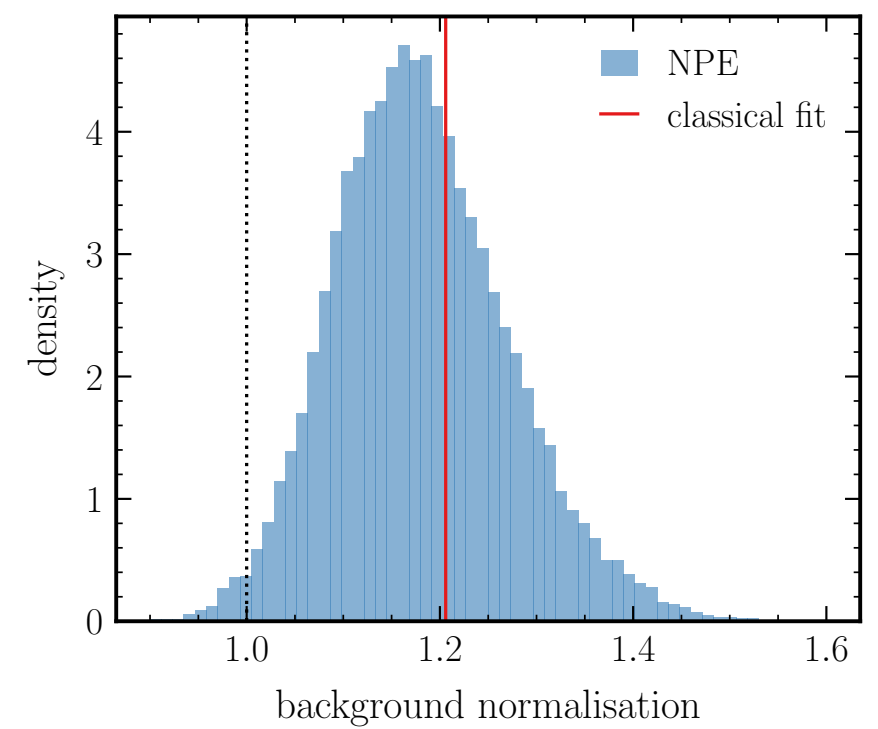
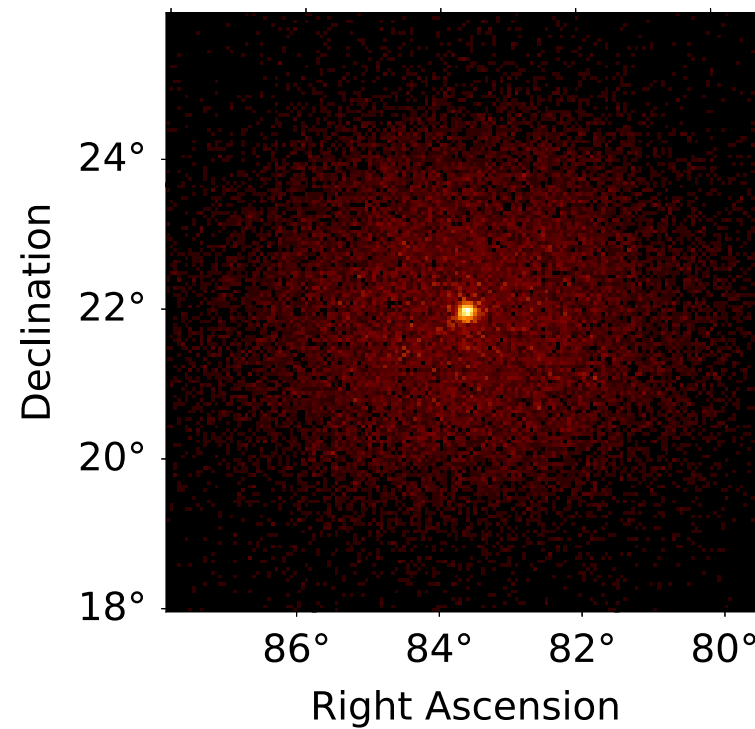
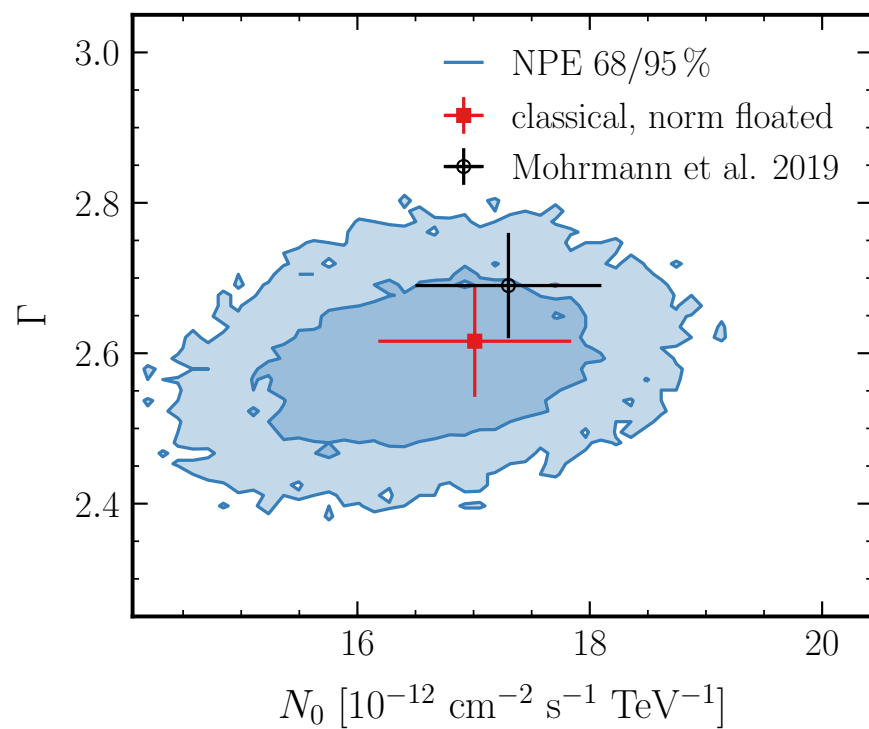
The inverse problem/parameter inference comes via a trainable normalising flow representing the posteriors: **neural posterior estimation**.

Train a conditional density $q(\theta | x)$ on 10,000 simulated pairs; applied to the observed cube it returns the posterior, with z_k and Poisson noise marginalised because they vary across the training set.

→ *for the experts: also requires an embedding network that generates a summary statistic from the input cubes; here a CNN is sufficient.*

Simulation-based inference approach to the Crab

Results in terms of Bayesian posteriors:



NPE background norm

1.17 [1.09, 1.27] (median [68 %])

gammapy

Classical 3D fit, norm floated

$\Gamma = 2.616 \pm 0.074,$

$N_0 = 17.01 \pm 0.83,$

norm 1.206

Mohrmann et al. 2019

$\Gamma = 2.69 \pm 0.07,$

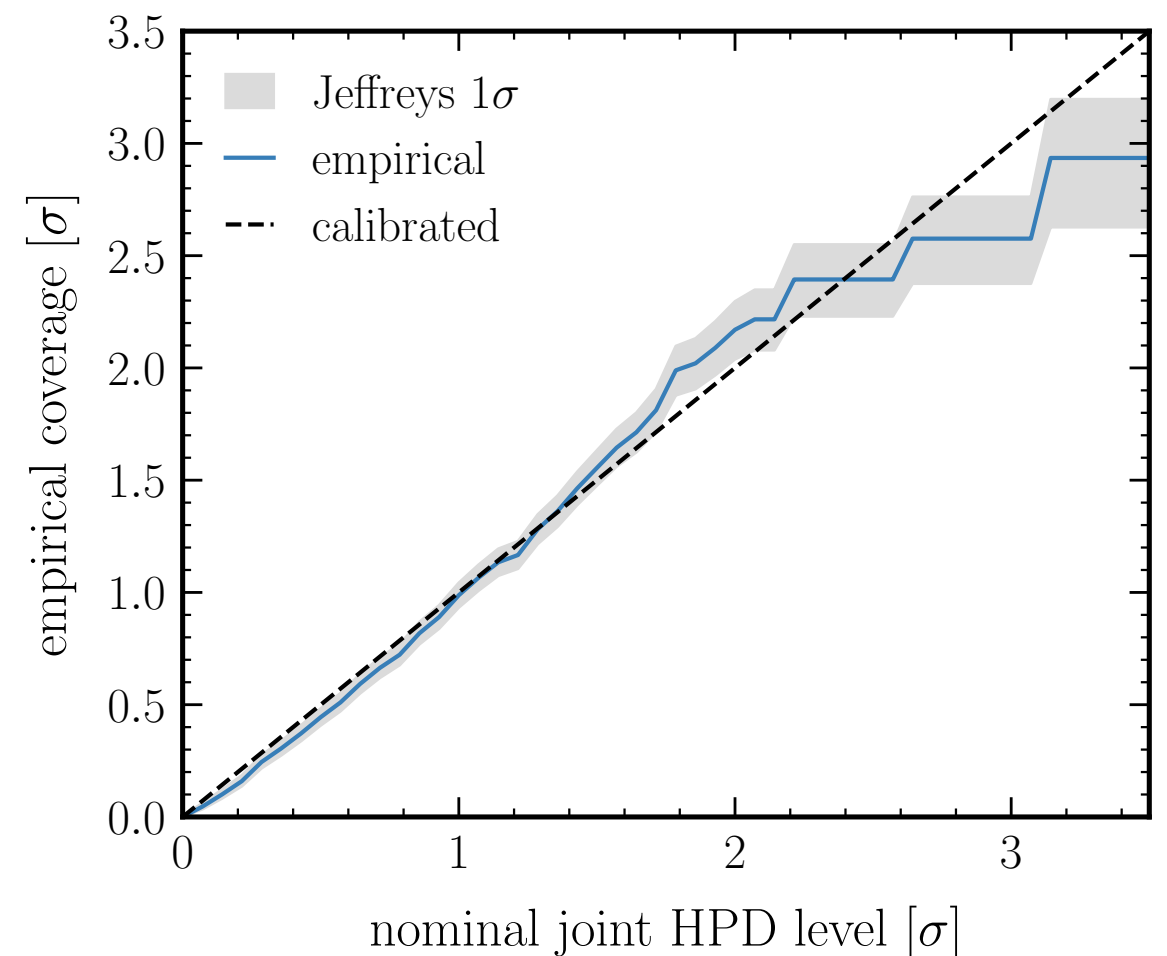
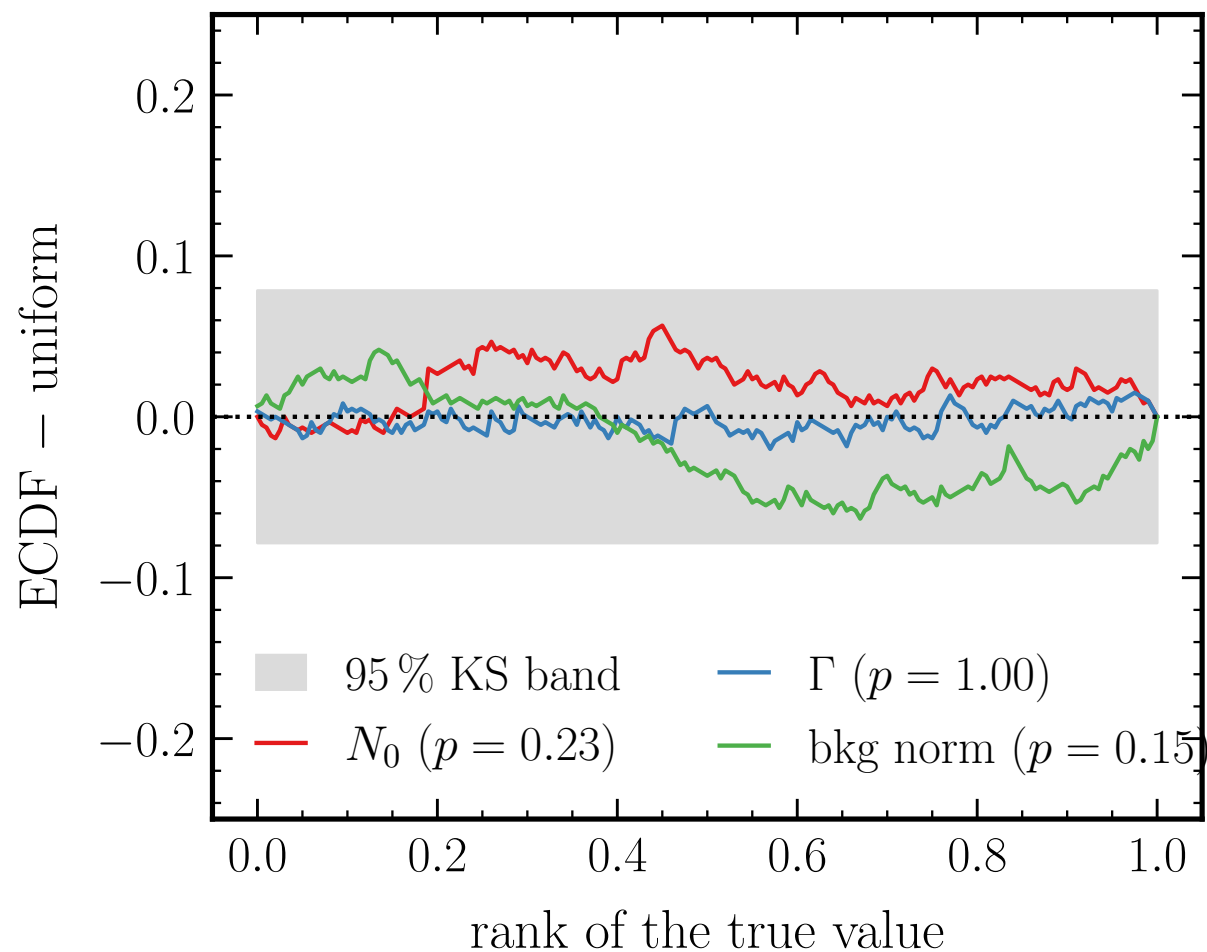
$N_0 = 17.3 \pm 0.8$

The retrieved posterior encompasses previous literature results and the standard gammapy fit (using the IRF background model). The CR background normalisation is exactly at the level the generative model failed to predict.

Simulation-based inference approach to the Crab

There are ways to validate the calibration of the obtained posteriors via SBI-specific methods (on simulations !): simulation-based calibration and coverage tests.

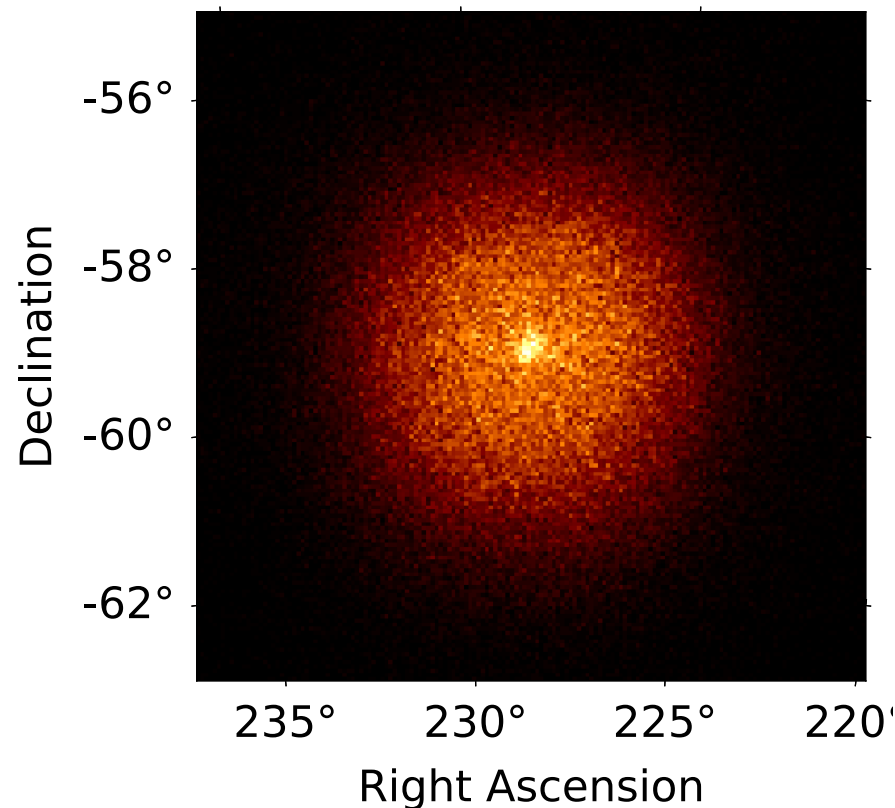
Basic idea: Prepare a large number of calibration simulations with parameters and datasets disjoint from the training sample. For each, get the posterior and draw a sufficiently large number of samples. Check that the desired statistical properties hold.



A calibrated posterior does not imply the absence of a reality gap between the simulator and the real data!

Simulation-based inference approach to MSH 15-52

More interesting object as it is extended and offers many more runs in the H.E.S.S. public dataset.



Signal model:

power law and elliptical disc with variable centre;
parameters priors:
RA $[228.35, 228.75]^\circ$, Dec $[-59.35, -58.99]^\circ$;
extension along main axis: $a \in [0.08, 0.35]^\circ$;
eccentricity and orientation angle: (e_1, e_2) on the disc
 $|e| \leq 0.95$,
spectral index: $\alpha \in [1.5, 3.5]$;
normalisation: $\log_{10} \Psi_0 \in [-13, -10]$

Side note: the typical disc parameters e and ϕ are bad parameterisations for a normalising flow. Writing the shape as $e_1 = e \cos 2\phi$, $e_2 = e \sin 2\phi$ instead of (e, ϕ) removes these pathologies.

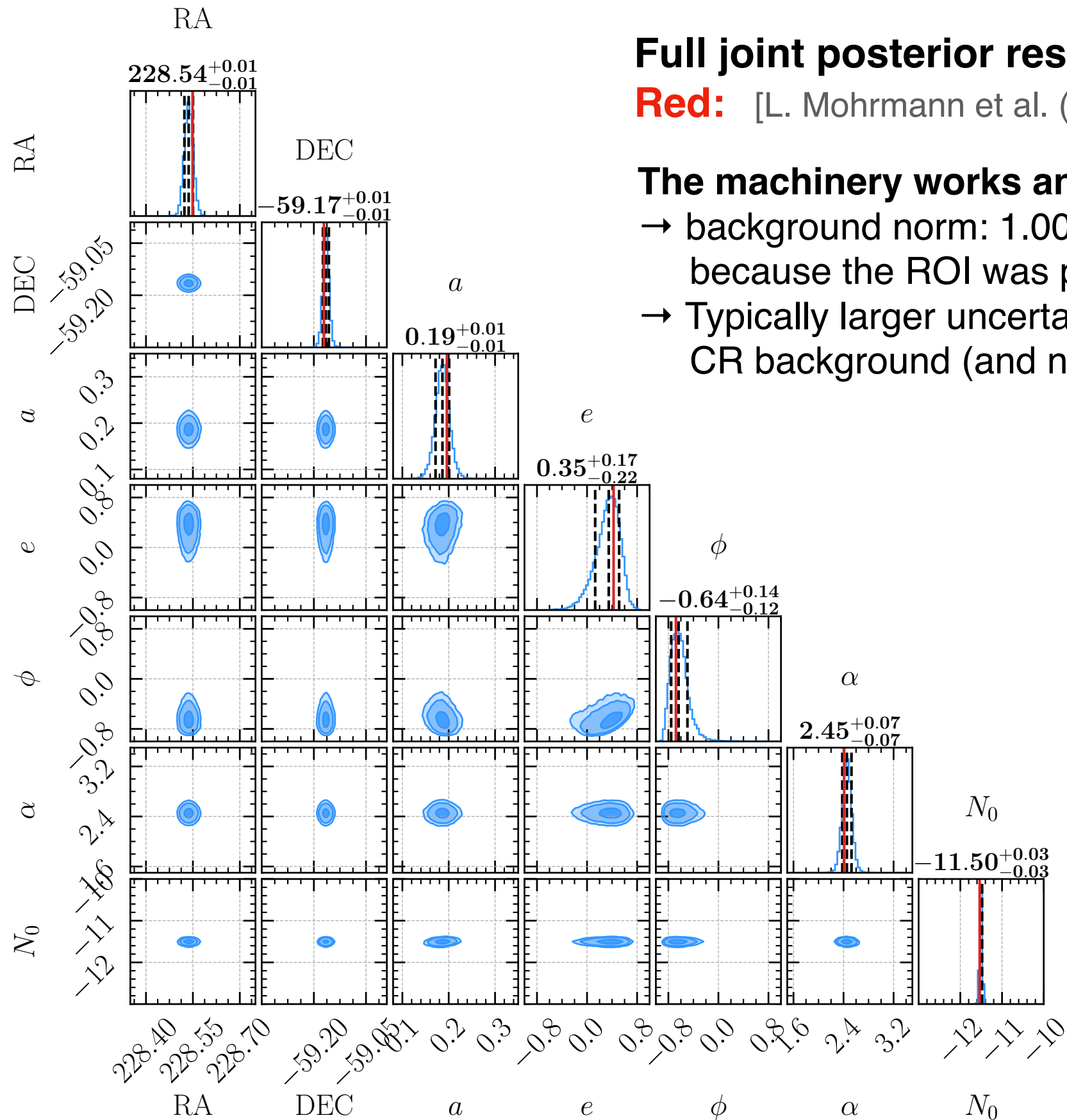
background model:

generative CR background model also trained on the MSH runs by masking the source at 0.5° (retaining some for calibration) and free-floating norm.

embedding network and training:

20,000 samples with finer CNN to encode the 3D cubes

Simulation-based inference approach to MSH 15-52

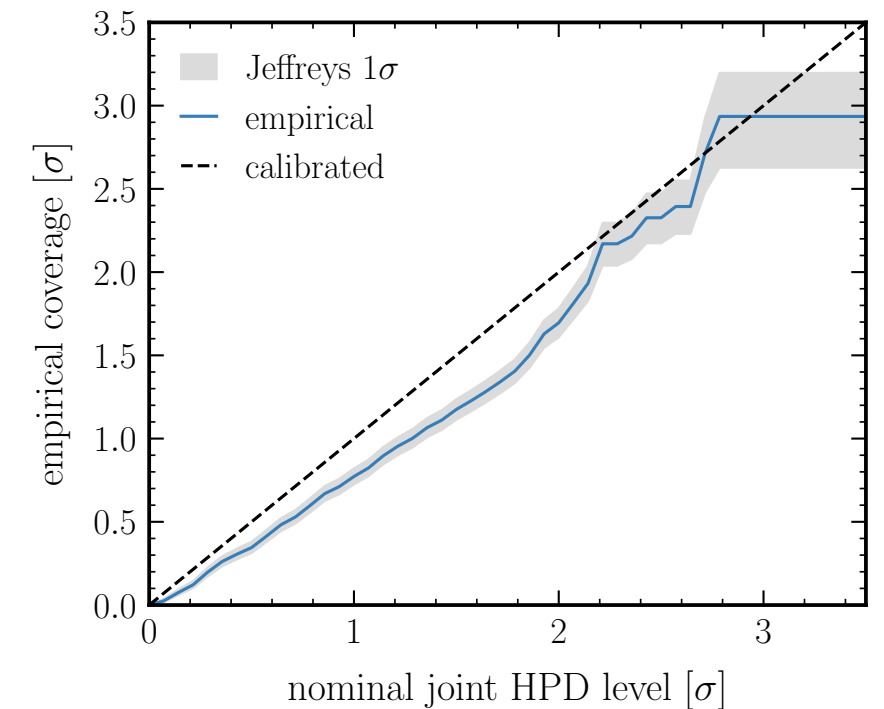


Full joint posterior results.

Red: [L. Mohrmann et al. (2019), A&A 632, A72]

The machinery works and recovers the literature results.

- background norm: 1.005 [0.969, 1.040] (perfect match because the ROI was part of the training data).
- Typically larger uncertainties due to included systematics of CR background (and not being a likelihood fit).

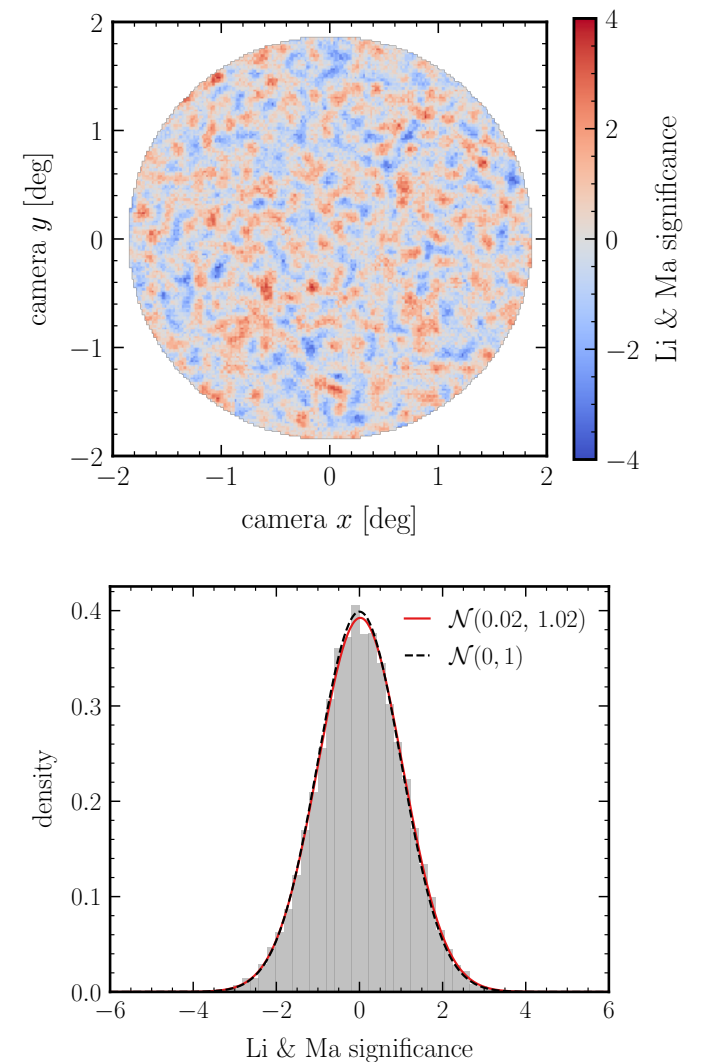


Slightly overconfident results but repairable by selecting best calibrated training epoch.

Summary and outlook

What are my take-home messages?

1. It is possible to “learn” the cosmic-ray background of a ground-based gamma-ray telescope including an estimate of the systematic uncertainties.
2. The H.E.S.S. public data is not sufficient to show that this method can also learn to predict the total rate of the background per observation.
3. A Bayesian analysis of the Crab and MSH employing simulation-based inference reproduces literature results with systematics-inflated uncertainties.



What does this mean for our intended MAGIC dwarf spheroidal analysis?

→ All good, MAGIC has dark patch observations and Miguel produced the first DL3 datasets.

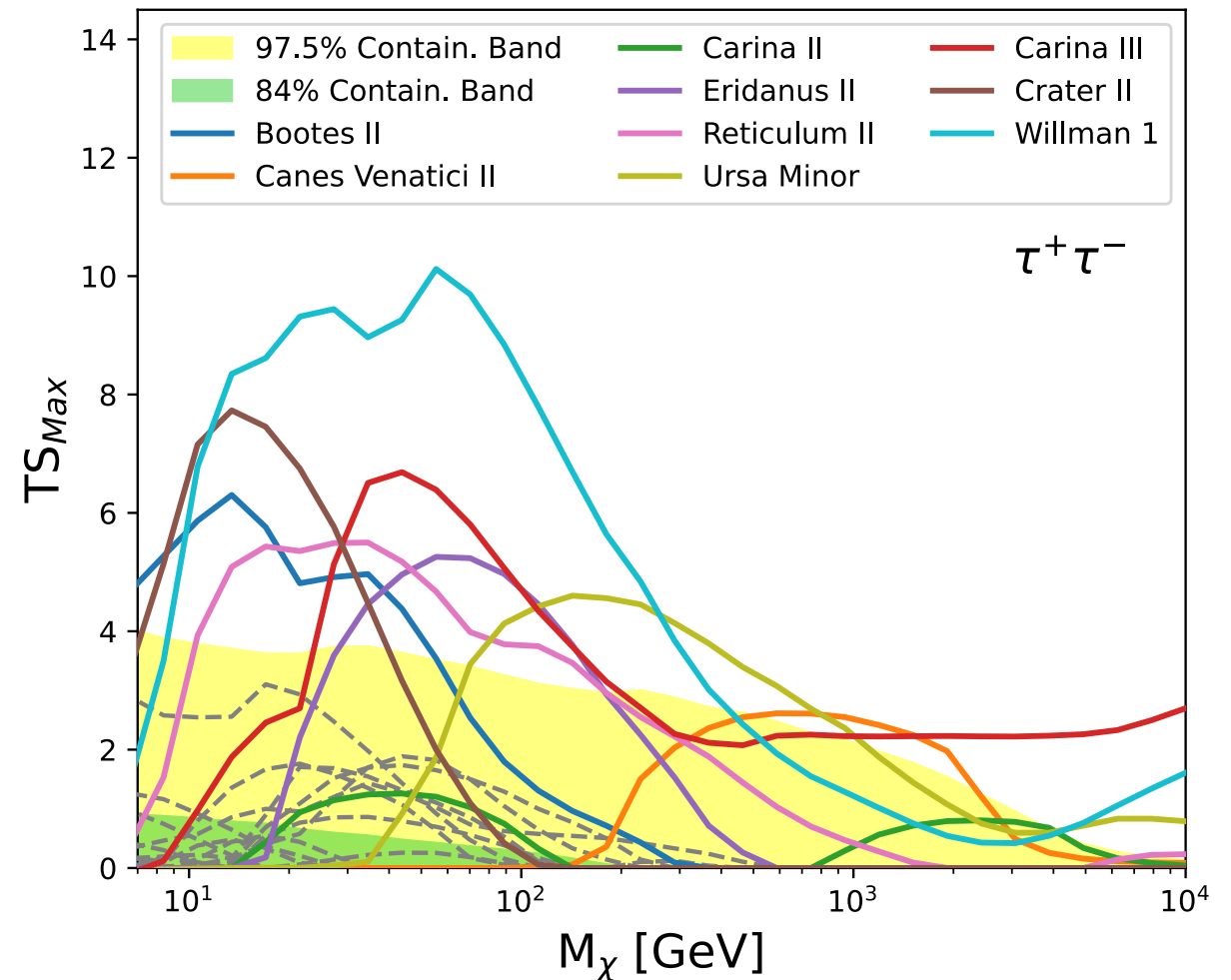
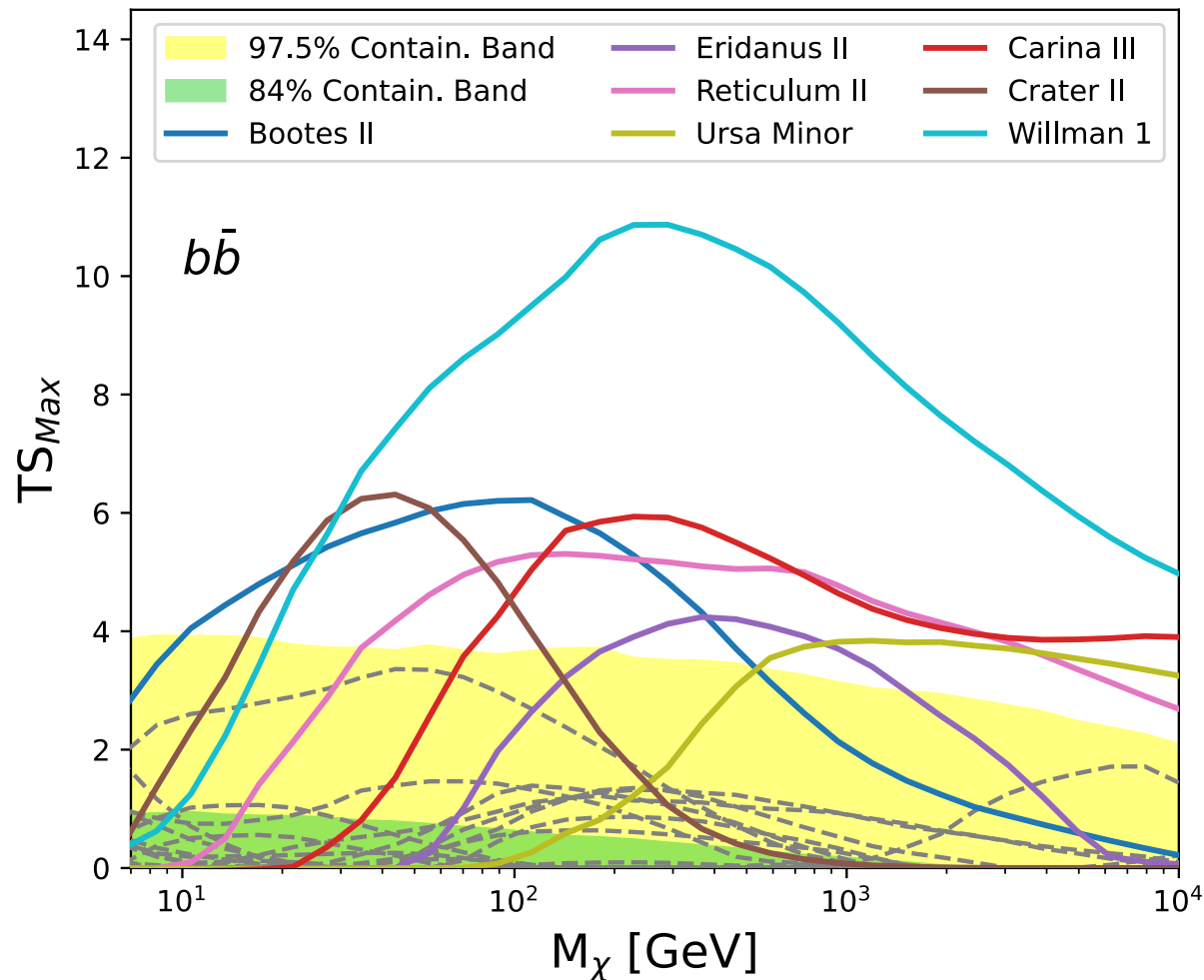
Interesting question: Does adding atmospheric conditions (LIDAR measurements) per run help to predict the total background rate?

Backup slides

Gamma-ray excess in dwarf spheroidals by the LAT

Motivation to address this question: There is a persistent hint of an excess towards dwarf spheroidal galaxies in Fermi-LAT data.

[A. Circiello et al., *ApJ.Lett.* 1004 (2026) 2, L33]



The relevant energy range (and target dwarfs) of the excess emission is rather aligned with the LAT but additional information from MAGIC observations certainly add constraining power when combined jointly.

A list of suitable dwarf spheroidal galaxies

[The CTAO Consortium, MNRAS 544 (2025) 3]

