Robustly Dissecting the Gamma-Ray Sky at High Latitudes with Simulation-Based Inference

Christopher Eckner[1,2], Noemi Anau Montel[3], Florian List[4], Francesca Calore[2] and Christoph Weniger[5]

[1]University of Nova Gorica-CAC, Slovenia [2]LAPTh, CNRS, France [3]Max Planck Institute for Astrophysics, Germany [4]University of Vienna, Austria [5]GRAPPA, University of Amsterdam, Netherlands

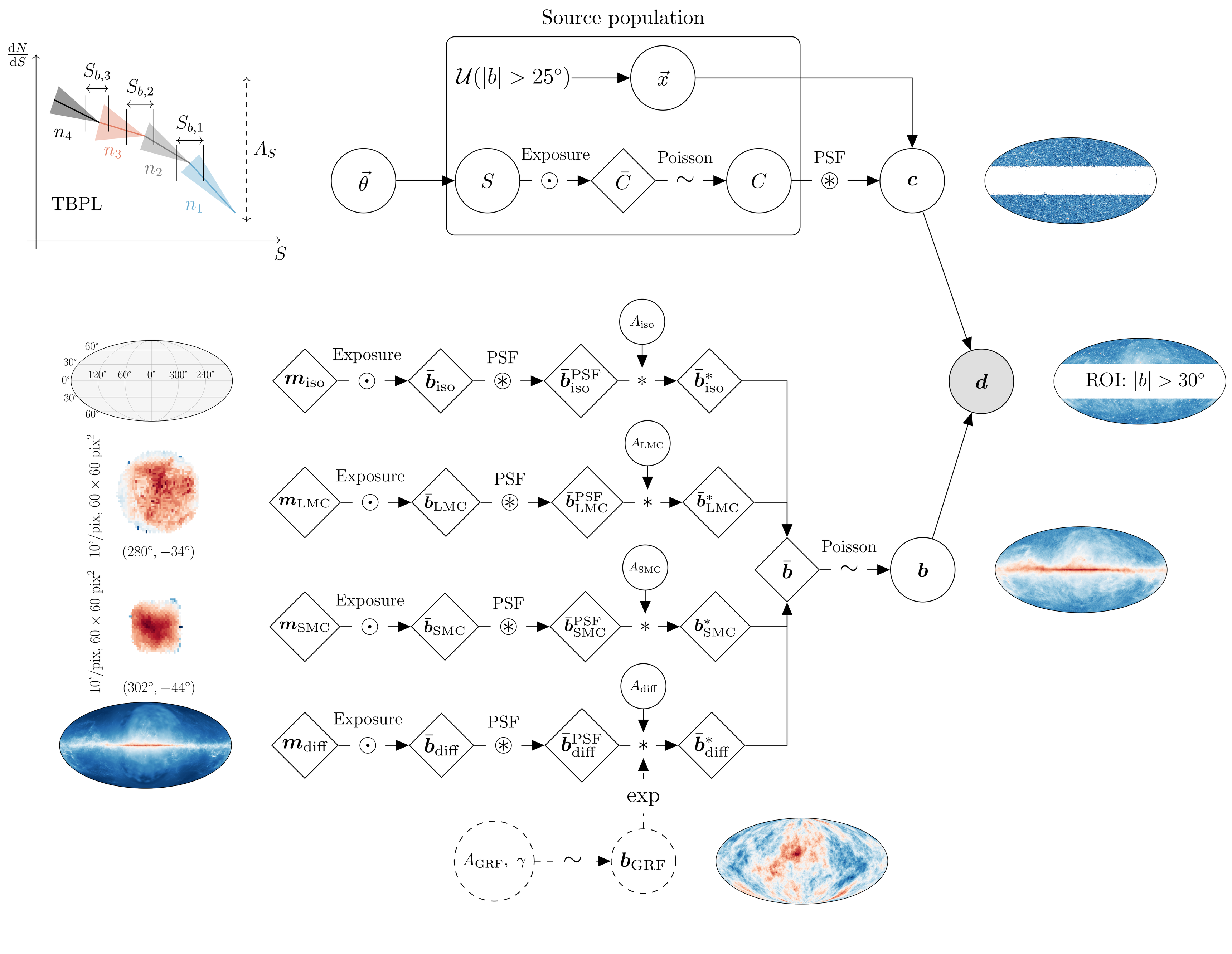
Motivation

- The gamma-ray sky seen by the *Fermi* Large Area Telescope (LAT) at GeV energies is comprised of a multitude of Galactic and extragalactic source populations as well as diffuse emissions.
- Additional exotic signatures like pair-annihilating thermal dark matter are typically a sub-dominant contribution requiring a very good knowledge of all astrophysical gamma-ray emissions.
- As shown with a *toy setup* in [1], simulation-based inference (SBI) allows for a comprehensive treatment of source detection and parameter inference.
- This framework extracts at the same time information from the detected and sub-threshold parts of source populations thereby accounting for detection biases.
- No external high-level data products like source catalogues that rely on different data selection criteria, assumptions and model simplifications are needed.

We present our **results [2]** obtained from analysing **high-latitude data of the *Fermi* LAT** to infer the **source-count distribution of (mostly) extragalactic point-like sources** in combination with **a catalogue of sources** detected by our SBI method. The analysis exploits gamma rays events from 1 to 10 GeV binned into a single image of the high-latitude gamma-ray sky.

A *Fermi*-LAT simulator of the high-latitude sky

- **Hierarchical Bayesian Model:** At latitudes $|b| \geq 30°$ well-defined astrophysical expectations in terms of gamma-ray emissions: extragalactic objects (blazars; Large and Small Magellanic Cloud), a few Galactic sources (pulsars), the diffuse Milky Way foreground and a diffuse isotropic background (IGRB). We introduce variations of the diffuse foreground via Gaussian random fields.



Assessing the degree of model mis-specification via anomaly detection

- We probed the realism of our simulated gamma-ray sky in relation to the observations of the *Fermi* LAT via anomaly detection. We employed the so-called One-Class Deep Support Vector Data Description (SVDD) method [3].
Idea: Map high-dimensional data onto a predefined lower-dimensional manifold. New target data points that significantly deviate from this manifold are identified as anomalies.

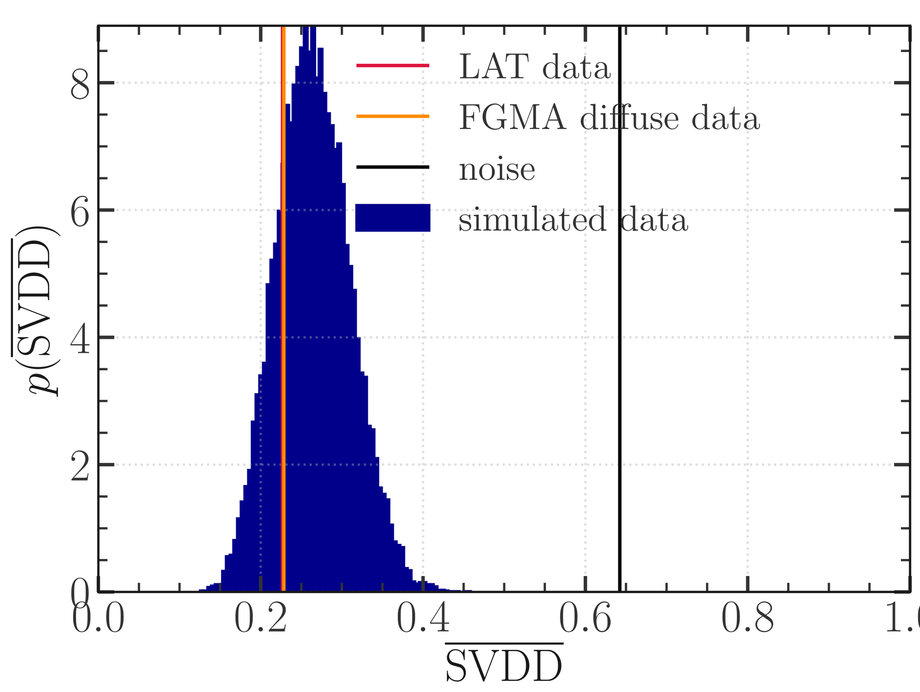
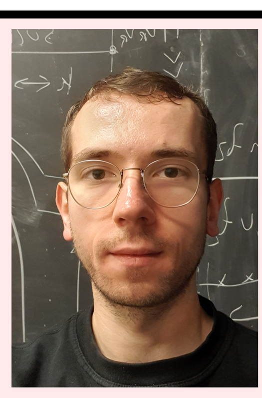


Figure 1: Distribution of the averaged anomaly score ($\overline{\mathrm{SVDD}}$) derived from training dataset in comparison to the target dataset with an alternative diffuse Milky Way foreground, *Fermi*-LAT sky, a target dataset with an alternative diffuse Milky Way foreground, and Gaussian noise.

References

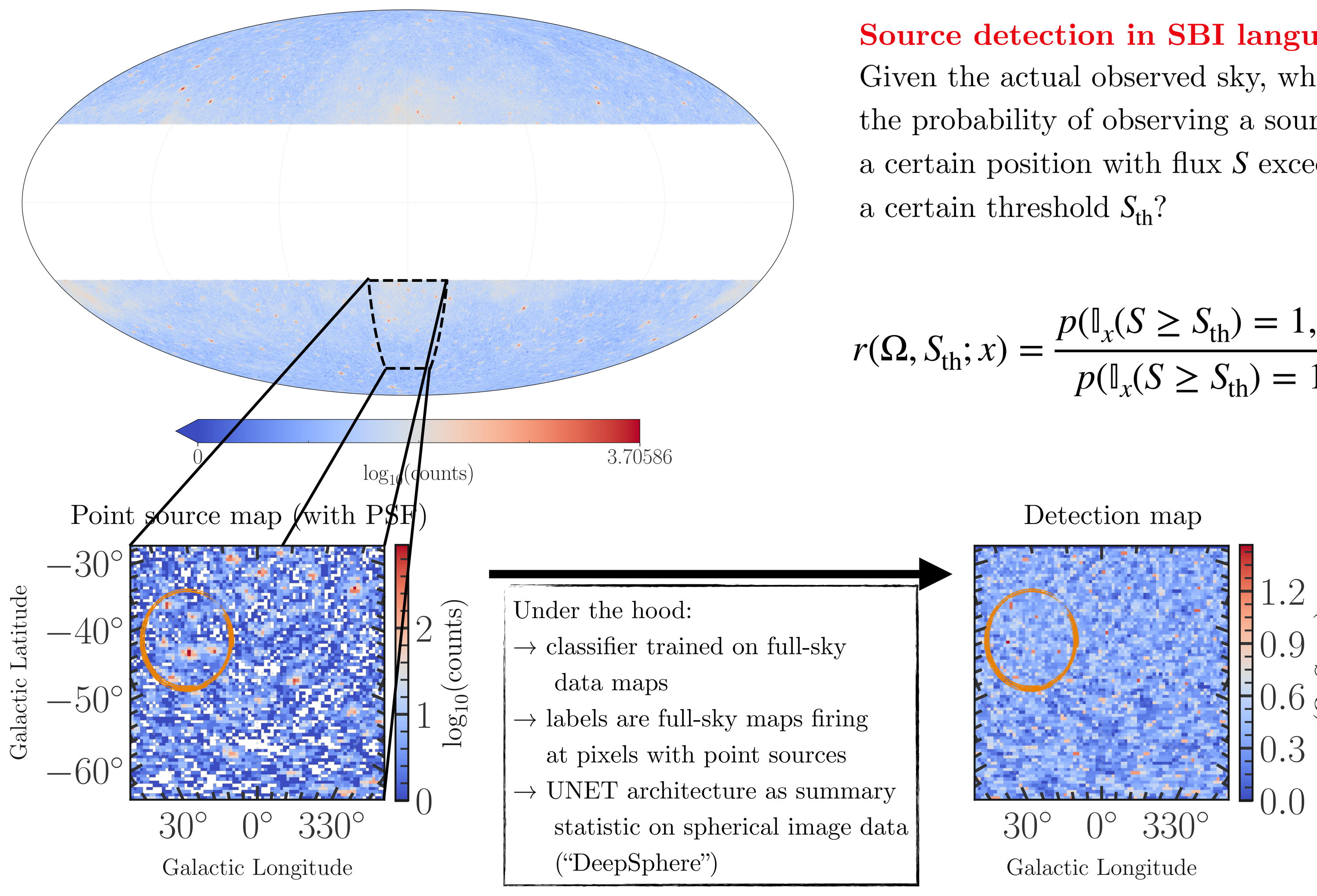
[1] N. Anau Montel & C. Weniger, NeurIPS 2022, arXiv: 2211.04291
[2] C. Eckner et al., arXiv:2505.02906
[3] S. Caron et al., SciPost Phys. 12 (2022) 077
[4] N. Perraudin et al., Astron. Comput. 27 (2019) 130-146
[5] https://github.com/deepsphere/deepsphere-pytorch
[6] The Fermi-LAT Collaboration, ApJ 799 (2015) 86

For more information on the project, come and talk to this guy



or have a look at our paper!

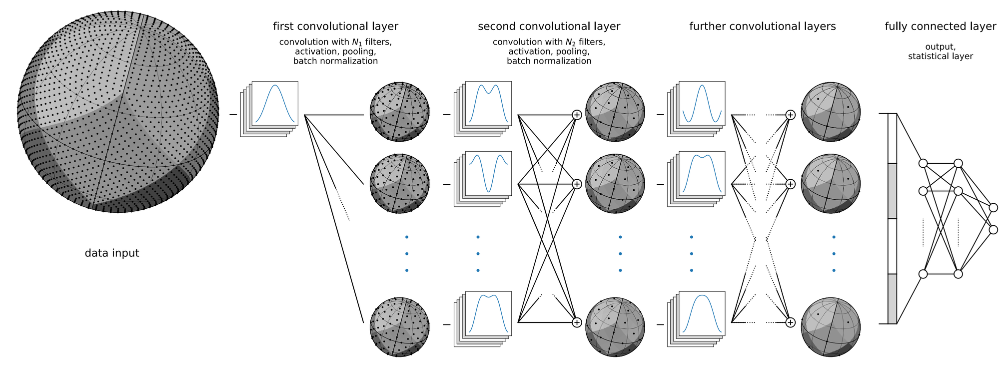
Source Detection with Neural Ratio Estimation



Source detection in SBI language: Given the actual observed sky, what is the probability of observing a source at a certain position with flux S exceeding a certain threshold S_{th}?

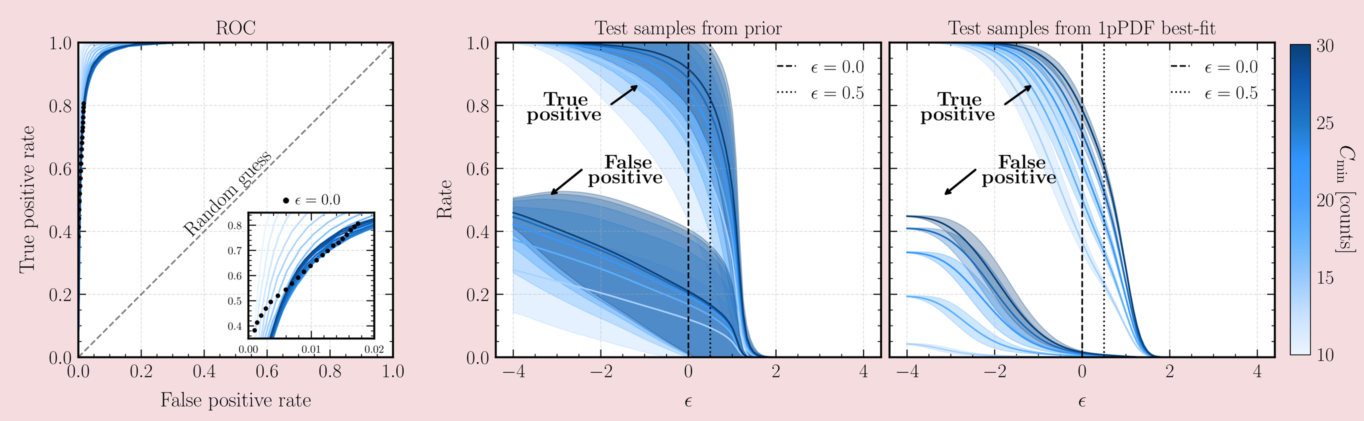
$$r(\Omega, S_{th}; x) = \frac{p(\mathbb{I}_x(S \geq S_{th}) = 1, \Omega \mid x)}{p(\mathbb{I}_x(S \geq S_{th}) = 1, \Omega)}$$

- **Neural Network framework:** We use the `DeepSphere` architecture [4] and HEALPix routines translated into `pytorch` syntax [5].

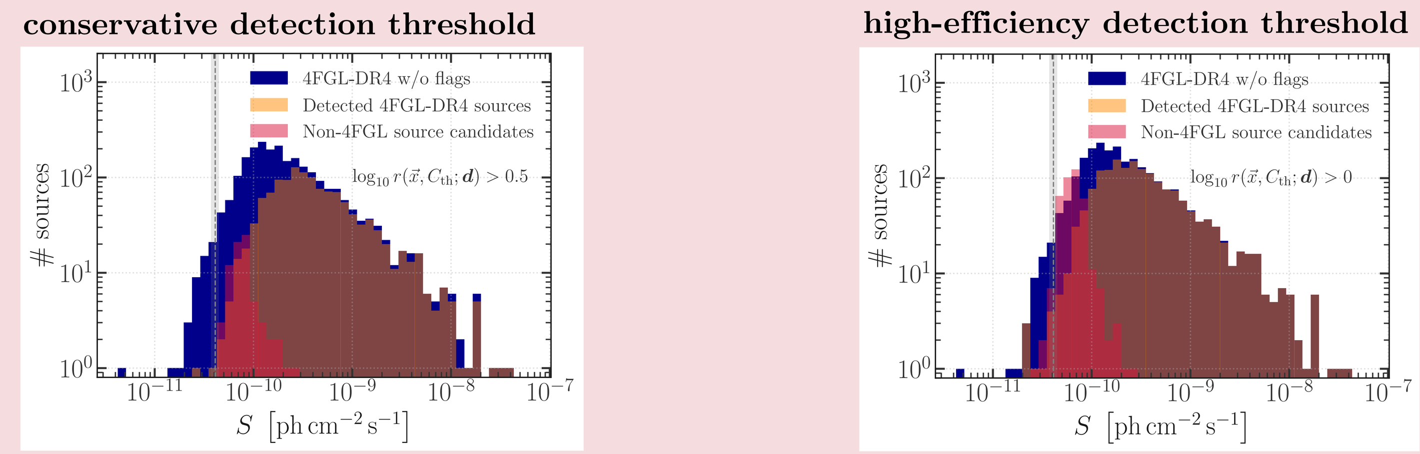


Results: Inferring the high-latitude source-count distribution

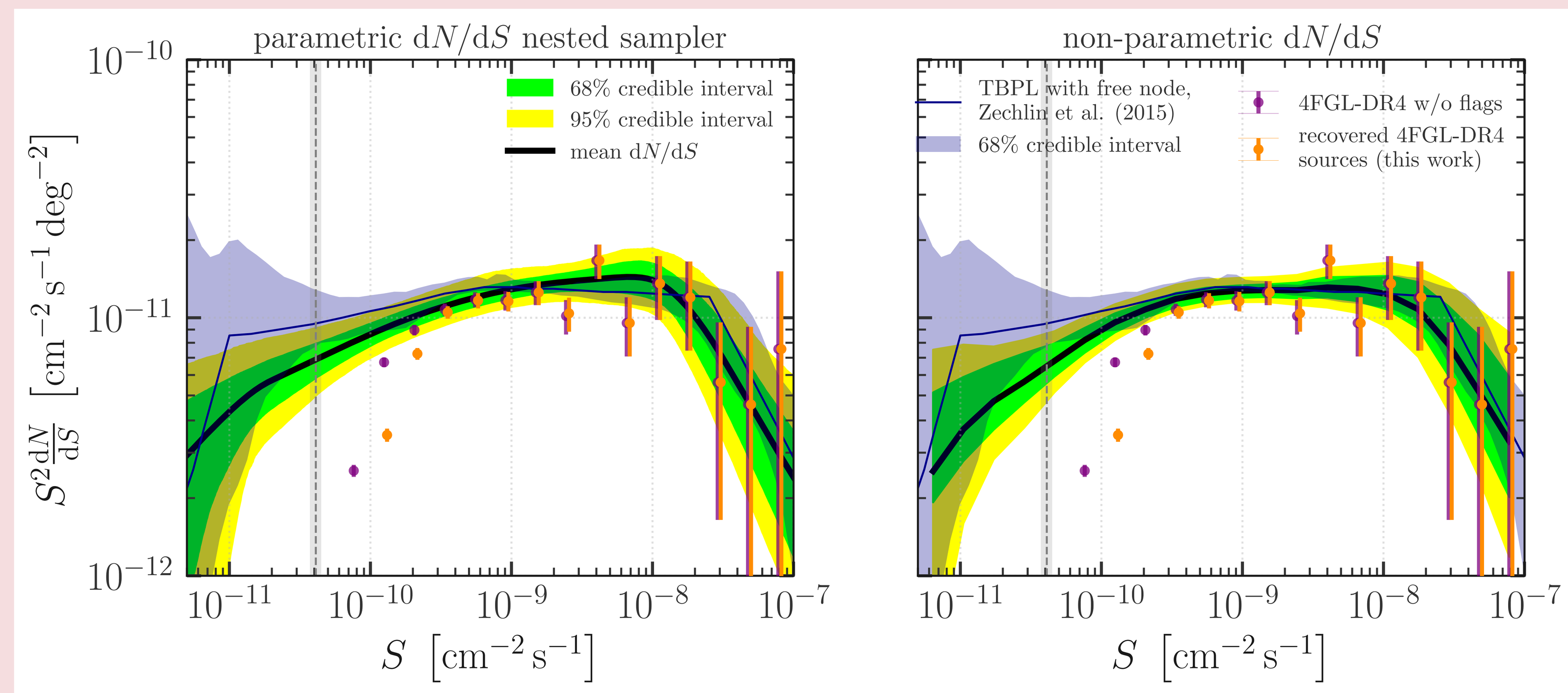
- **Detection efficiency calibration:** Performed on simulated data. It allows us to **recover up to 98% of the brightest gamma-ray sources** detected with traditional methods (4FGL-DR4 source catalogue [6]).



conservative detection threshold | high-efficiency detection threshold



- **Inference of the source-count distribution's profile:** We employ two distinct approaches to reconstruct the high-latitude source-count distribution: *(i)* **parametrically** using autoregressive neural ratio estimation to obtain the joint posterior with subsequent nested sampling and, *(ii)*, **non-parametrically** inferring the flux of the profile per flux bin directly. **Both approaches yield consistent results.**



Conclusions and outlook

Our work demonstrates that simulation-based inference is a robust tool capable of performing **source detection** in noisy datasets such as gamma-ray observations as well as **parameter inference** regarding physically relevant observables as the high-latitude source-count distribution.
Our framework for gamma-ray simulations and inference lies the foundations for future applications to more detailed physics questions like the composition of the IGRB. This requires:

- Extension to the handling of **multiple energy bins**, i.e. multiple image input channels to capture the spectral dependence of the occurring components.
- In parallel to source detection, we implement **source classification to distinguish multiple gamma-ray source populations**.

christopher.eckner@ung.si