

To cut or not to cut: Data-quality evaluation for the ASTRI Mini-Array

E. Molina,^a L. Foffano,^{b,*} S. Crestan,^c S. Iovenitti,^d S. Lombardi,^{e,f} F. Lucarelli^{e,f} and T. Mineo,^g for the ASTRI Project[†]

^a*Instituto de Astrofísica de Canarias and Dpto. de Astrofísica, Universidad de La Laguna,
C/ Vía Láctea s/n, E-38205 La Laguna, Tenerife, Spain*

^b*INAF - IAPS,
via del Fosso del Cavaliere 100, I-00133 Roma, Italy*

^c*INAF - IASF Milano,
Via Alfonso Corti 12, 20133 Milano, Italy*

^d*INAF - Osservatorio astronomico di Brera,
Via E. Bianchi 46, 23807, Merate, Italy*

^e*INAF - OAR,
Via Frascati 33, I-00078 Monte Porzio Catone (Roma), Italy*

^f*ASI, Space Science Data Center,
Via del Politecnico s.n.c., I-00133 Roma, Italy*

^g*INAF - IASF Palermo,
Via U. La Malfa 153, 90146 Palermo, Italy*

E-mail: emolina@iac.es, luca.foffano@inaf.it

The ASTRI Mini-Array consists of nine Cherenkov telescopes under construction at the Observatorio del Teide, with one of them already performing observations since November 2024. Given the complexity of the analysis of gamma-ray data acquired with Cherenkov telescopes, a proper evaluation of the quality of these data is important to avoid a negative impact on the high-level scientific products. There are several factors that can contribute to reducing the quality of the observations, such as clouds, high humidity or a high dust concentration in the air, among others. In order to take all these factors into account and evaluate their impact on the data, a quality check pipeline has been developed for ASTRI. This contribution describes the pipeline, from its inputs at different data levels to the production of outputs in the form of good-quality lists and diagnostic plots. We also show how setting preliminary quality cuts on the Crab Nebula data taken with the first ASTRI telescope can significantly improve the source detection.

39th International Cosmic Ray Conference (ICRC2025)
15–24 July 2025
Geneva, Switzerland



[†]<http://www.astri.inaf.it/en/library/>

*Speaker

1. Introduction

The ASTRI Mini-Array [1] is an array of nine imaging atmospheric Cherenkov telescopes located at the Observatorio del Teide ($28^{\circ}18'04''$ N, $16^{\circ}30'38''$ W, 2390 m above sea level) in the island of Tenerife, Spain. The array is currently under construction, with one telescope (ASTRI-1) being already in a commissioning phase and taking data since November 2024, and the rest expected to become operational over the second half of 2025 and 2026 [2]. Each telescope follows a modified Schwarzschild-Couder optical configuration [3, 4], with a primary and a secondary mirror with diameters of 4.3 m and 1.8 m, respectively. The camera is made of silicon photomultipliers and has 2368 pixels, covering a field of view (FoV) of 10.5° with an angular pixel size of 0.19° . According to the simulations, the ASTRI Mini-Array is sensitive to gamma rays with energies between 0.3 and 200 TeV, with an energy resolution of 12% and an angular resolution of 0.05° at 10 TeV. The performance of the array remains approximately constant up to an angular distance of 3° from the camera centre. This fact, together with its large FoV, makes ASTRI a great instrument to study extended sources at TeV energies and above. A more detailed description of the array performance can be found in [5].

Together with the nine Cherenkov telescopes responsible for taking scientific data, there are four main auxiliary systems that provide real-time information on the environmental conditions: two weather stations, which measure different weather parameters like humidity, temperature and pressure; the sky quality meter (SQM), which measures the sky brightness in the same direction as the telescopes are pointing to; and the all-sky camera, which provides information on the cloudiness (fraction of the sky covered by clouds). In the future, a LIDAR will also be installed in order to monitor the atmospheric transmission within the FoV. Moreover, the Cherenkov camera also records so-called variance data [6], which provide additional information on the number of stars in FoV and on the effect of the night sky background on the camera pixels. These quantities can be used, among other things, to characterize the environmental conditions or the telescope mispointing.

In this contribution, we present how a preliminary evaluation of the scientific data quality from the ASTRI-1 telescope is performed using an end-to-end pipeline. The pipeline uses the scientific data themselves, as well as the information provided by the auxiliary systems and the variance data. Section 2 describes the ASTRI-1 observations used for the data-quality study. Section 3 provides a brief technical description of the pipeline and of its main processing steps, including the data selection. A preliminary example of the improved scientific results based on the quality cuts applied with this pipeline is shown in Section 4, and some concluding remarks are given in Sec. 5.

2. Description of the observations

This quality study is based on 200.8 hours of Crab Nebula observations performed between November 2024 and February 2025. The data span four distinct observational periods separated by several days around the full Moon, when the telescope was not operating. A summary of these periods is provided in Table 1.

In order to better estimate the background accounting for the non-uniform sensitivity of the camera, all the observations were done in wobble mode [7], in which the source is located at an offset angle with respect to the camera centre. Four different source positions were normally used

Table 1: Observing times of the Crab Nebula with ASTRI-1 for each of the periods defined by the start and end dates, together with the total observing time. The last column shows the times for the data used in this work, with an offset angle of 0.5° and zenith angles below 30° (see text).

Start date	End date	Observing time (h)	Selected time (h)
2024-11-22	2024-12-09	71.2	20.9
2024-12-27	2025-01-08	56.0	5.9
2025-01-20	2025-02-03	46.6	11.9
2025-02-18	2025-02-27	27.0	13.6
Total		200.8	52.3

for a given offset, two along the right ascension axis and two along the declination axis. Each observation run has the source located in one of these four positions, and consecutive runs alternate between different positions until the cycle is completed and restarted. Typically, a run lasts around 30 min, so the whole cycle takes ~ 2 h to complete.

Since the ASTRI-1 telescope is in a commissioning phase, the observations were done using varying offset angles –ranging from 0.5° to 4.5° – in order to explore the telescope capabilities in different conditions. For the purpose of this work, we only considered observations taken with a 0.5° wobble offset. Moreover, in order to minimize the variation of several quantities with the zenith angle, we restrict our study to the data obtained when the source was at a zenith angle below 30° . The corresponding times for each period that result from applying these conditions are stated in the last column of Table 1, for a total of 52.3 h.

3. The ASTRI data-quality check pipeline

The quality evaluation of the ASTRI-1 data is done within a data-quality check (DQC) pipeline that takes care of all the necessary steps to perform such an evaluation. The pipeline operates with FITS input files at different data levels and obtained by the Cherenkov camera and the auxiliary systems. Files are produced on a run basis, i.e., every observing run has associated its own set of scientific and auxiliary files. Regarding the scientific runs, the pipeline uses the raw information recorded by the camera, as well as some quantities computed at later stages of the data processing, such as the image (Hillas) parameters [8] and the reconstructed energy, arrival direction and *gammanness* of each event (see [5, 9] for more information on the computation of these quantities as part of the global analysis chain). Since the DQC analysis uses processed data that become available only after the scientific observations are completed, the pipeline is typically executed the following morning, once the ASTRI data processing and reduction have been finalized.

The DQC pipeline runs in a computing cluster physically located at the Osservatorio Astronomico di Roma, in Italy. Several configuration parameters are set through a TOML file. Among them, we include the date range of the data to evaluate and the set of quality cuts that will be used in

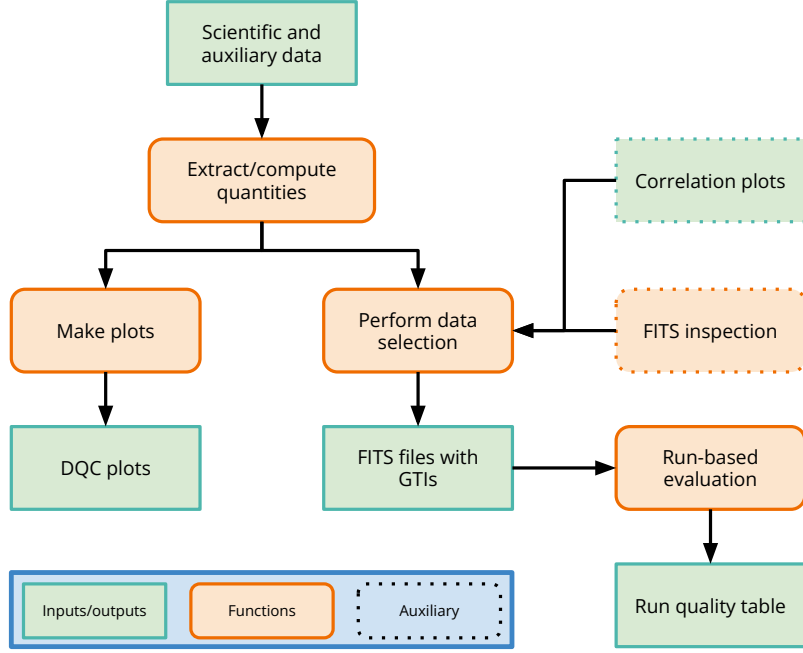


Figure 1: Schematic representation of the the ASTRI DQC pipeline. Orange boxes represent the steps performed by the pipeline, while green boxes are the inputs and outputs of each step. Dotted boundary lines are used for auxiliary functions and inputs not belonging to the automatic pipeline itself, but that are employed for the data-quality selection step.

the selection procedure. Once executed, the pipeline automatically performs a series of sequential steps that end up in the production of different outputs ready to use in later stages of the ASTRI data analysis. Specifically, three kinds of outputs are generated: a list of good-quality runs, a list of good time intervals (GTIs) within single runs, and diagnostic plots for human inspection (see examples of the latter in Fig. 2). A schematic representation of the DQC pipeline execution is provided in Fig. 1.

4. Preliminary results

Here we present preliminary results from the analysis of the Crab Nebula data described in Sec. 2, highlighting the improved outcomes obtained by applying the quality cuts defined in this work. These results were obtained by using the 52.3 h of data listed in Table 1. The definition of the proper quantities that characterize the quality of the data and the study of their effect on the high-level scientific results is an ongoing work. Currently, this data-quality selection is based on an empirical approach applied to the Crab Nebula observations, and will evolve in the next months as more data from different FoVs and telescopes become available. Therefore, the results presented here should be taken as very preliminary.

In order to select the parameters that best define the quality of the data, we studied the correlations between different quantities. In particular, we used the event rate of the reduced (and cleaned) data as a reasonably good indicator of the data-taking conditions, and checked how other

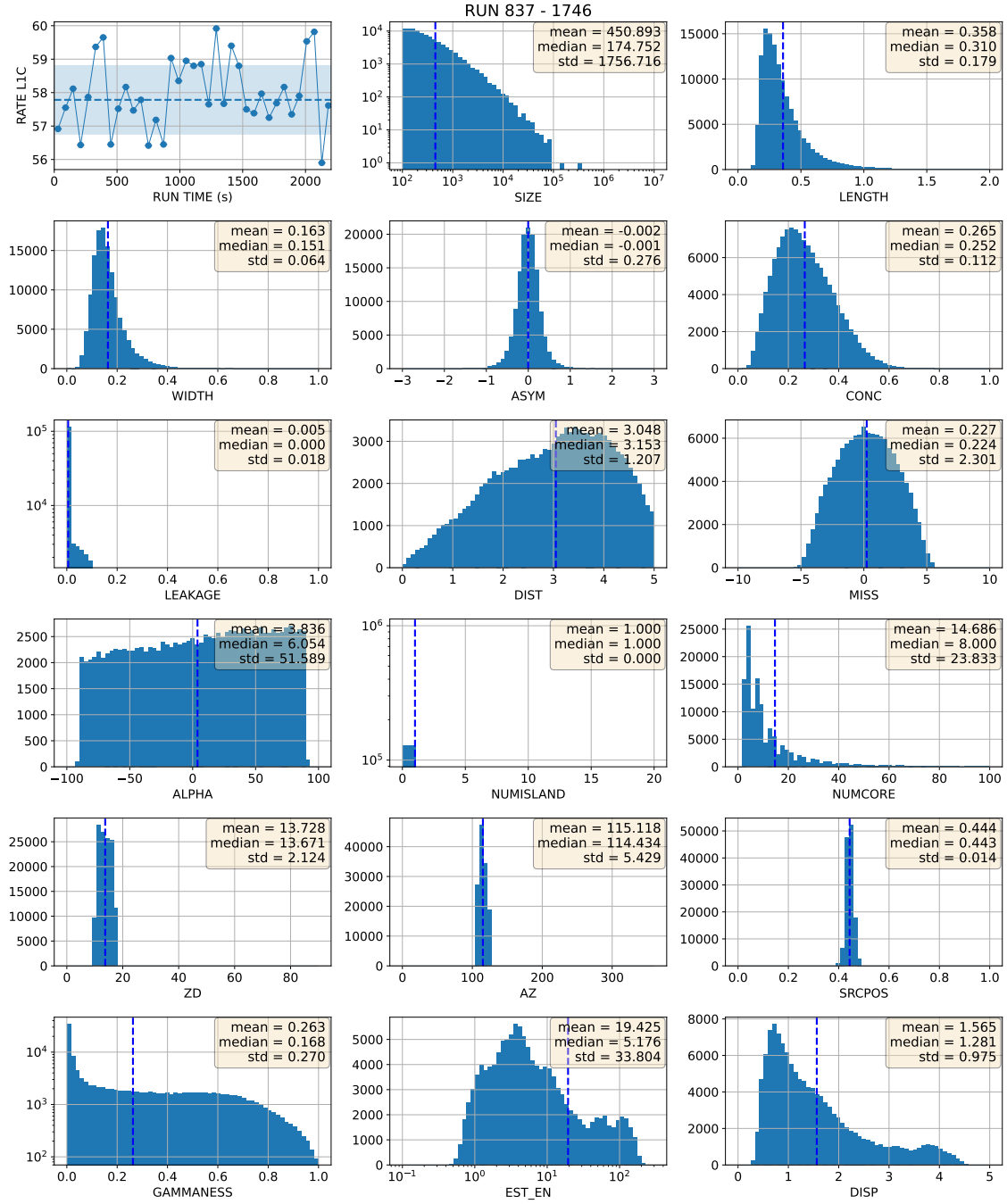


Figure 2: Representative example of some of the plots produced by the DQC pipeline for a given observing run. The intra-run evolution of the event rate is shown in 1-min time bins, together with the histograms of several quantities.

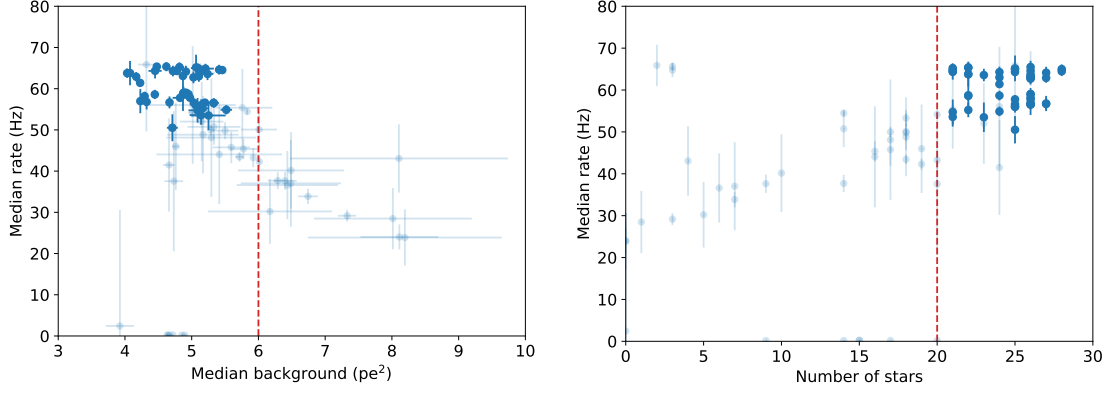


Figure 3: Dependency between the rate of the cleaned scientific data runs with the background (left) and number of detected stars (right) computed from variance data. The error bars represent the standard deviation of each quantity within the run. The highlighted points are those fulfilling the 3 quality cuts described in Sec. 4. The dashed red lines represent the cuts in the background and number of stars in the left and right panels, respectively.

quantities changed with it. We recall that we limit our study to zenith angles below 30° , where the change of rate with the pointing position is small compared to other effects. The result of this study was the selection of three different cuts that improve the quality of the data, each of them taking care of different aspects. These cuts are always considered at a run level, leaving the GTI-based data selection as future work:

1. **Median background from the variance data below 6 squared photo-electron.** This cut ensures the usage of data taken under dark conditions, removing observations performed during moon time or twilight.
2. **Number of detected stars within the FoV higher than 20.** For Crab Nebula observations, this corresponds to more than $\sim 75\%$ of the stars detected in optimal conditions. Lower values of the number of stars indicate poor weather conditions, mainly owing to the presence of clouds or high humidity.
3. **Relative standard deviation of the cleaned rate below 10%.** This condition is set in order to get rid of runs with high internal variability due to different reasons (clouds passing by, moon rising or setting, electronic problems,...).

Figure 3 shows how the median run rate of the cleaned data varies with the background and the number of detected stars computed from the variance data. As seen from the highlighted points in the figure, setting the cuts described above ensures a stable rate above 50 Hz, which is a good indicator regarding the quality of the observations. In order to confirm the adequate data-taking conditions of the selected runs, we also performed further checks (not shown here for simplicity) by visually inspecting the evolution of some quantities in the (intra-run) diagnostic plots produced by the DQC pipeline.

As a preliminary evaluation of the performance of the quality study, we produced θ^2 detection plots of the Crab Nebula for both the runs selected (Fig 4, left panel) or rejected (Fig 4, right panel)

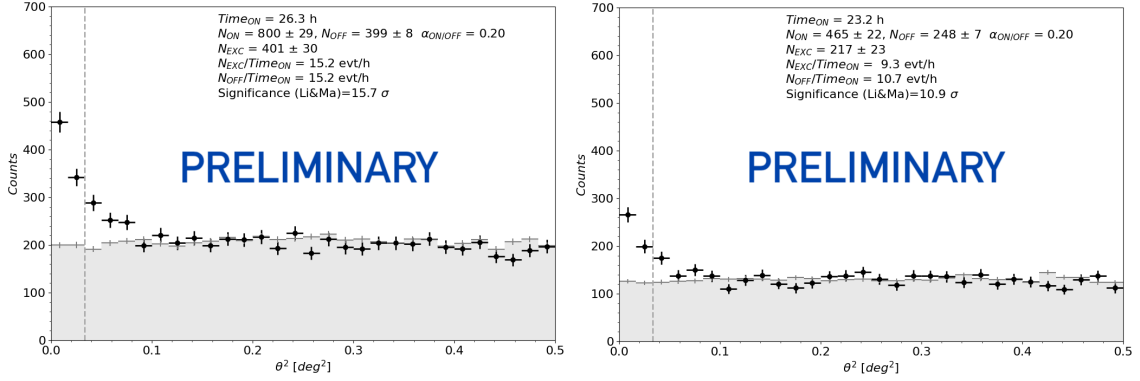


Figure 4: Detection θ^2 plots for the Crab Nebula data passing the quality cuts (left) and those rejected by the data-quality selection (right). The black crosses represent the number of (ON) events coming from a circular region centred around the source, while the grey area are the background (OFF) events computed from a symmetric region with respect to the camera centre. The vertical dashed line at 0.033 deg² represents the limit of the region from which the number of ON and OFF events are taken in order to compute the detection significance of the source.

by the three cuts above. We can clearly see an improvement of the Li&Ma detection significance [10] of the good-quality data, 15.7σ , with respect to the bad-quality ones, 10.9σ , with similar observing times of 26.3 h and 23.2 h, respectively¹. We can further evaluate this improvement by computing the rate of excess events in each case, which is 15.2 ± 1.1 events/h for the good-quality data, and 9.3 ± 1.0 events/h for the bad-quality data, resulting in an enhancement of 63%. Given that both θ^2 plots were produced using the same data analysis configuration, we can attribute the observed improvement to the applied quality cuts.

5. Conclusions

The data-quality evaluation of the ASTRI Mini-Array is done through an end-to-end pipeline that takes care of performing a series of steps required to conduct such evaluation and produce different outputs to be used at further steps of the ASTRI data analysis. In this work, the DQC pipeline was applied to a subset of the Crab Nebula data obtained during the commissioning phase of the ASTRI-1 telescope, for a total of 52.3 h. The results of the data-quality selection, even if indicative and based on a preliminary choice of the quality parameters, already suggest a significant improvement in the signal obtained for the good-quality data with respect to the bad-quality ones, pointing at the importance of the data selection as part of the overall analysis of ASTRI data. In this direction, further work is being performed in order to refine and improve the selection procedure.

Acknowledgments

This work was conducted in the context of the ASTRI Project. We gratefully acknowledge support from the people, agencies, and organisations listed here: <http://www.astri.inaf.it/>

¹The 2.8 h discrepancy between the sum of these times and the 52.3 h reported in Sec.2 is explained by the rejection of some individual events by the data reduction chain, before producing the θ^2 plots.

[en/library/](#). EM acknowledges support from the Agencia Estatal de Investigación from the Spanish Ministerio de Ciencia, Innovación y Universidades (MCIU/AEI) under grant 630.

References

- [1] S. Scuderi, A. Giuliani, G. Pareschi, G. Tosti, O. Catalano, E. Amato et al., *The ASTRI Mini-Array of Cherenkov telescopes at the Observatorio del Teide*, *Journal of High Energy Astrophysics* **35** (2022) 52 [2208.04571].
- [2] G. Pareschi, *Status of the ASTRI Mini-Array Gamma-Ray Experiment*, in *39th International Cosmic Ray Conference (ICRC2025, these proceedings)*, vol. 39 of *International Cosmic Ray Conference*, July, 2025.
- [3] V. Vassiliev, S. Fegan and P. Brousseau, *Wide field aplanatic two-mirror telescopes for ground-based γ -ray astronomy*, *Astroparticle Physics* **28** (2007) 10 [astro-ph/0612718].
- [4] G. Sironi, *Aplanatic telescopes based on Schwarzschild optical configuration: from grazing incidence Wolter-like x-ray optics to Cherenkov two-mirror normal incidence telescopes*, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, S.L. O'Dell and G. Pareschi, eds., vol. 10399 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, p. 1039903, Sept., 2017, DOI.
- [5] S. Vercellone, C. Bigongiari, A. Burtovoi, M. Cardillo, O. Catalano, A. Franceschini et al., *ASTRI Mini-Array core science at the Observatorio del Teide*, *Journal of High Energy Astrophysics* **35** (2022) 1 [2208.03177].
- [6] A. Segreto, O. Catalano, M.C. Maccarone, T. Mineo and A. La Barbera, *Calibration and monitoring of the ASTRI-Horn telescope by using the night-sky background measured by the photon-statistics ("variance") method*, in *36th International Cosmic Ray Conference (ICRC2019)*, vol. 36 of *International Cosmic Ray Conference*, p. 791, July, 2019, DOI [1909.08750].
- [7] V.P. Fomin, S. Fennell, R.C. Lamb, D.A. Lewis, M. Punch and T.C. Weekes, *New methods of atmospheric Cherenkov imaging for gamma-ray astronomy. II. The differential position method*, *Astroparticle Physics* **2** (1994) 151.
- [8] A.M. Hillas, *Cerenkov Light Images of EAS Produced by Primary Gamma Rays and by Nuclei*, in *19th International Cosmic Ray Conference (ICRC19)*, Volume 3, F.C. Jones, ed., vol. 3 of *International Cosmic Ray Conference*, p. 445, Aug., 1985.
- [9] S. Crestan and S. Lombardi, *ASTRI-1: Early data and performance highlights*, in *39th International Cosmic Ray Conference (ICRC2025, these proceedings)*, vol. 39 of *International Cosmic Ray Conference*, July, 2025.
- [10] T.P. Li and Y.Q. Ma, *Analysis methods for results in gamma-ray astronomy.*, *Astrophysical Journal* **272** (1983) 317.