PROCEEDINGS OF SCIENCE

ONLINE ICRC 2021
THE ASTROPARTICLE PHYSICS CONFERENCE
Berlin | Germany
37th International Cosmic Ray Conference
12–23 July 2021

FACT - Database-based Analysis and Spectrum Calculations

Bernd Schleicher,[a,*] **A. Arbet-Engels,**[b] **D. Baack,**[c] **M. Balbo,**[d] **A. Biland,**[b] **T. Bretz,**[b,e] **J. Buss,**[c] **D. Dorner,**[a] **L. Eisenberger,**[a] **D. Elsaesser,**[c] **D. Hildebrand,**[b] **R. Iotov,**[a] **A. Kalenski,**[a] **K. Mannheim,**[a] **A. Mitchell,**[b] **D. Neise,**[b] **M. Noethe,**[c] **A. Paravac,**[a] **W. Rhode,**[c] **V. Sliusar**[d] **and R. Walter**[d]

[a] *Universität Würzburg, Institut für Theoretische Physik und Astrophysik, Würzburg, Germany*
[b] *ETH Zurich, Institute for Particle Physics and Astrophysics, Zurich, Switzerland*
[c] *TU Dortmund, Experimental Physics 5, Dortmund, Germany*
[d] *University of Geneva, Department of Astronomy, Versoix, Switzerland*
[e] *also at: III. Physikalisches Institut A, RWTH Aachen University, Aachen, Germany*
E-mail: bernd.schleicher@uni-wuerzburg.de

The First G-APD Cherenkov Telescope (FACT) is located at the Observatory Roque de los Muchachos on the Canary island La Palma. It uses the imaging air Cherenkov technique to detect gamma rays. With the help of silicon based photosensors in the camera, FACT is an ideal instrument to monitor a small sample of sources with a good time coverage. The automatic operation of the telescope allows an increase of the duty cycle of the instrument. An SQL database is part of the automatic analysis chain, where data on an event basis are stored. This way of storing the data has a several advantages. It can provide easy web-access to all data with no need of creating different user accounts for the analysers and without using special software. The data selection is done via simple queries to the database. This allows very flexible and powerful queries with, for example, user-defined time binning or background suppression. By using observed and simulated events, the complete analysis chain can be done, including the calculation of the measured energy spectrum. This could also be implemented to the Quick-Look Analysis to provide spectral information during the night with a low latency.

37th International Cosmic Ray Conference (ICRC 2021)
July 12th – 23rd, 2021
Online – Berlin, Germany

*Presenter

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

PoS(ICRC2021)760

1. Introduction

This contribution summarizes a new approach to analyse data of Cherenkov telescopes. While most analyses work with files containing the data, we present here an alternative approach by handling the event data using a database. In the following, first the structure of the database will be presented. Next, we outline the analysis chain pointing out the differences between the standard approach and the database-based analysis. As example, the calculation of a spectrum is shown. The conclusions summarize the advantages of the database-based analysis.

2. Event Database

The standard FACT analysis uses the Modular Analysis and Reconstruction Software - Cherenkov Observatory edition (MARS - CheObs ed.) [1] to analyse the data. An overview of the analysis can be found in section 3. Therefore the standard structure of the data are run-wise stored files, that are automatically processed in the analysis pipeline. For the majority of FACT data, a data run contains five minutes of data taking. In the beginning and at the end of the night, where the night sky background changes rapidly the time of a run is reduced to one minute. It is assumed, that the night sky background is stable during one run. Additionally to this run-wise stored data, a database is filled automatically during data processing on an event basis. The database is also used to store on a run-wise basis all the auxiliary information at a central location. Therefore, there are different tables storing information at different levels. Most of the auxiliary information are only useful in a context of runs and therefore are linked to the date and the individual run number of this night, or the id of the file, which is a combination of the date and the run number. The event database only contains the information relevant for a single event and the file id to connect it to the run-wise information. The combination of the file id and the event number is unique and used as primary key of this database.

3. FACT Analysis Chain

The FACT analysis chain consists of five main steps. A detailed description for the Quick-Look Analysis (QLA) can be found in [2] and for the offline analysis in [3]. Furthermore, dedicated cuts for light curves and spectra are presented in [4]. The first step is the calibration and signal extraction of the raw data. As second step, the images are cleaned with the help of the amplitude and timing information from each individual pixel. The purpose of this image cleaning is to remove those pixels containing only background noise not connected to the triggered shower. The next analysis step is to calculate image parameters for each single event. Therefore, a statistical analysis of the cleaned image is done and different parameters are calculated. These three steps comprise the low level analysis.

After that, the standard analysis processes the data with run-wise files, while for the database-based analysis, the image parameters of all events are filled into a dedicated table in the database. These image parameters can be used to reconstruct the origin, energy and type of the primary particle of this event. Firstly, the image parameters are used to suppress the background caused by cosmic rays. Of course as this background suppression is based on statistical calculations, the

PoS(ICRC2021)760

single event can not be identified as a gamma or hadronic event, only for a large number of events the identification is valid. With the use of the disp method [5], the origin of the primary particle is estimated and the angular distance to the assumed source position is used to allocate an event to the source region in the sky. To estimate the gamma-ray background, an off-source measurement is needed. Therefore the telescope is operated in wobble mode. This means the telescope does not point directly to the source, but the assumed position of the source is 0.6 degree from the centre of the camera (on-region). In this observation mode, five other positions in the camera (off-regions), with the same angular distance of 0.6 degree to the center of the camera and equally distributed as a hexagon in the camera, are used as off-source measurement. Therefore, the off-regions are analysed the same way as the on-position. To correct for inhomogenities within the camera, the pointing position is exchanged every five minutes. With this kind of analysis, the gamma signal from the source and the background can be measured in parallel. Dependent on the purpose of the study, the significance of an observation or the excess rate can be calculated. The excess rates can be converted to a flux by comparing the excess rate of a given source with the measured excess rate of the Crab Nebula under similar observation conditions like trigger threshold based on the night sky background, or zenith distance. The Crab Nebula is assumed as a constant source at TeV energies and therefore can be used as calibration source. To calculate the energy spectrum of a measurement, dedicated Monte-Carlo simulations are needed to estimate the effective detector area. These simulations include the shower simulation in the atmosphere, including the atmosphere as part of the detector. A detailed description of the spectrum calculation can be found in section 3.2.

3.1 Various Options for the Analysis

As explained in [4], it can be useful to use different background suppression cuts for different purposes. For example, the cuts used for the QLA are optimized for a fast and robust analysis of the data and to search for significant flux enhancements in quasi real-time to initiate automatic alerts in case of a flare. But to use the data for a detailed physics study one better uses the offline analysis, since data are reprocessed after every major change and therefore are fully comparable. For the offline analysis described in [4], the background suppression is optimized for the smallest relative error of the light curve data points. When one wants to use the data energy resolved, for example as differential flux, then, depending on the brightness of the source and its spectrum, it might be useful to use other cuts in the background suppression in order to get more excess events at low energies. The background suppression cuts are mainly dependent on the parameter Size, which is a equivalent to the amount of measured photons of the event and this parameter is proportional to the energy of the primary particle. Therefore, a cut that is survived by more events with low energies can be used for calculating the differential flux and get more events in the low energy range. Depending on the spectral slope of the measured source, the overall significance might be lower with such cuts, but you gain more statistics in the low energy bins, while there is nearly no difference in the intermediate and high energy bins compared to the cuts optimized for light curves. A comparison of the two sets of cuts is shown in [4]. To use the different sets of cuts, in the database-based analysis only one single line in the SQL-query needs to be changed. This nicely illustrates the flexibility of the database-based approach. Also for different binning of a light curve, this flexibility is ideal. Dependent on the source flux, the useful time binning of a light curve can vary between a few minutes, e.g. a bright flaring night of a source (see e.g. [6]), up to yearly or

PoS(ICRC2021)760

even longer time binning, e.g. for sources with a flux just above the detection limit of FACT. Also different kinds of studies might need a different time binning. For simultaneous multi-wavelength data, it can be very useful to have identical time binning with instruments from other wavelength, as shown e.g. in [7].

3.2 Spectrum Calculations

In the following, the steps for calculating the measured spectrum of a source are described. These formulae are used to calculate the spectrum with the use of the database. The differential flux $\Phi(E)$ is defined with E as energy, t as time and A as collection area:

$$\Phi(E) = \frac{dN}{dE \cdot dt \cdot dA} \tag{1}$$

For a constant collection area and time, the differential flux is often referred to as $\Phi(E) = \frac{dN}{dE}$.

The flux of an energy interval ΔE is

$$\Phi(\Delta E) = \frac{1}{A_0 \cdot \Delta T} \frac{N(\Delta E)}{\epsilon(\Delta E) \cdot \Delta E} \tag{2}$$

with the effective observation time $\Delta T = \sum \delta t_i$, the total area A_0 and the efficiency $\epsilon(\Delta E)$. The collection area and the efficiency are only available for simulated data. A_0 is the production area of the simulations and the efficiency $\epsilon(\Delta E)$ is dependent on the analysis chain. The efficiency is defined by the ratio between the number of excess events $N_{\text{exc}}(\Delta E)$ recognised by the analysis and the number of simulated events $N_0(\Delta E)$ per energy bin. The excess events are those events that are classified by the analysis chain to be gamma particles from the source. The effective area A_{eff} is therefore defined as $A_{\text{eff}} = A_0 \cdot \epsilon(\Delta E)$.

The number of excess events is calculated as

$$N_{\text{exc}} = N_{\text{sig}} - \hat{N}_{\text{bg}}. \tag{3}$$

N_{sig} is the number of signal events. These are the events identified as potential gamma events from the direction of the source and $\hat{N}_{\text{bg}}$ is the number of background events, normalized by the number of off-regions. The number of background events is calculated the same way as for the signal events, the only difference is using an off-position in the camera. As this is done for five off-positions, the number of background events has to be normalized by $\frac{1}{5}$. The statistical error of the excess events $\sigma^2(N_{\text{exc}})$ are thus calculated as

$$\sigma^2(N_{\text{exc}}) = \left(\frac{dN_{\text{exc}}}{dN_{\text{sig}}}\right)^2 \cdot \sigma^2(N_{\text{sig}}) + \left(\frac{dN_{\text{exc}}}{d\hat{N}_{\text{bg}}}\right)^2 \cdot \sigma^2(\hat{N}_{\text{bg}}) = \sigma^2(N_{\text{sig}}) + \frac{1}{5^2}\sigma^2(N_{\text{bg}}). \tag{4}$$

The uncertainties are given by the Poisson error $\sigma(N_{\text{sig, bg}}) = \sqrt{N_{\text{sig, bg}}}$ for the number of background and signal events.

Since the number $N(\Delta E, \Delta\Theta)$ of events in an energy bin $E \in \Delta E = [E_{\text{min}}; E_{\text{max}}]$ and a zenith angle interval $\Theta \in \Delta\Theta = [\Theta_{\text{min}}; \Theta_{\text{max}}]$ are dependent on this two quantities, it is useful to calculate a weighted number of events $N'(\Delta E, \Delta\Theta)$, defined as

$$N'(\Delta E, \Delta\Theta) = \sum_{i=0\ldots N}^{E\in\Delta E} \sum_{j=0\ldots N}^{\Theta\in\Delta\Theta} w(E_i, \Theta_j) = \sum_{i=0\ldots N}^{E\in\Delta E} \sum_{j=0\ldots N}^{\Theta\in\Delta\Theta} \rho(E_i)\tau(\Theta_j) = \sum_{i=0\ldots N}^{E\in\Delta E} \rho(E_i) \sum_{j=0\ldots N}^{\Theta\in\Delta\Theta} \tau(\Theta_j) \tag{5}$$

PoS(ICRC2021)760

The spectral weight $\rho(E)$ adapts the spectral shape of the simulated spectrum to the measured spectrum of the source and the zenith weight $\tau(\Theta)$ adapts the observation time versus the zenith angle. Both weights are independent from each other. The uncertainty of a sum of weights $\sigma^2(N')$ like this is defined as:

$$\sigma^2(N') = \sum_N [\rho_i \cdot \sigma(\tau_i) + \tau_i \cdot \sigma(\rho_i)]^2 \tag{6}$$

The second term of this sum will be zero, because with the well defined energy, the error $\sigma(\rho_i)$ of the spectral weight is zero and the overall error calculates as

$$\sigma^2(N') = \sum_N \rho_i^2 \cdot \sigma^2(\tau_i). \tag{7}$$

The spectral and zenith weight can be defined as followed:

$$\rho(E) = \rho_0 \frac{\Phi_{\mathrm{src}}(E)}{\Phi_0(E)} \tag{8}$$

$$\tau(\Delta\Theta) = \tau_0 \frac{\Delta T(\Delta\Theta)}{N(\Delta\Theta)} \tag{9}$$

In these formulae, Φ_0 and Φ_{src} are the simulated spectrum and the source spectrum and $N(\Delta\Theta)$ is the number of produced events in the zenith distance interval $\Delta\Theta$ and $\Delta T(\Delta\Theta)$ is the total observation time in this bin. The quantities ρ_0 and τ_0 are normalization constants, that will be shortened in the next steps. The remaining error of the weights is defined by the error of the zenith weight τ:

$$\sigma^2(\tau) = \left[\frac{d\tau}{d\Delta T}\sigma(\Delta T)\right]^2 + \left[\frac{d\tau}{dN}\sigma(N)\right]^2 = \tau\left[\left(\frac{\sigma(\Delta T)}{\Delta T}\right)^2 + \left(\frac{\sigma(N)}{N}\right)^2\right]^2 \tag{10}$$

The relative error of the total observation time is mainly caused by the measured dead-time of the data acquisition and is about $\frac{\sigma(\Delta T)}{\Delta} \approx \frac{1\,\mathrm{s}}{5\,\mathrm{min}}$. The relative error of the number of events is given by the Poisson error $\frac{\sigma(N)}{N} = \frac{1}{\sqrt{(N)}}$. The efficiency $\epsilon(\Delta E, \Delta\Theta)$ of the energy interval ΔE and the zenith distance interval $\Delta\Theta$ for the weighted events is similarly as above define as:

$$\epsilon(\Delta E, \Delta\Theta) = \frac{N'_{\mathrm{exc}}(\Delta E, \Delta\Theta)}{N'_0(\Delta E, \Delta\Theta)} \tag{11}$$

The differential flux in an energy interval ΔE and the zenith distance interval $\Delta\Theta$ can be calculated with the weighted efficiency ϵ as:

$$\Phi(\Delta E, \Delta\Theta) = \frac{1}{A_0 \cdot \Delta T}\frac{N_{\mathrm{exc,measured}}}{\epsilon \cdot \Delta E} = \frac{N_{\mathrm{exc,measured}}}{N'_{\mathrm{exc,simulated}}}\frac{N'_{0,\mathrm{simulated}}}{A_0 \Delta T \Delta E} \tag{12}$$

With the use of Gaussian error propagation, the error in a given energy interval ΔE of the differential flux can be calculated as:

$$\sigma^2(\Phi) = \Phi^2\left[\left(\frac{\sigma(N_{\mathrm{exc,measured}})}{N_{\mathrm{exc,measured}}}\right)^2 + \left(\frac{\sigma(N'_{\mathrm{exc,simulated}})}{N'_{\mathrm{exc,simulated}}}\right)^2 + \left(\frac{\sigma(N'_{0,\mathrm{simulated}})}{N'_{0,\mathrm{simulated}}}\right)^2\right] \tag{13}$$

PoS(ICRC2021)760

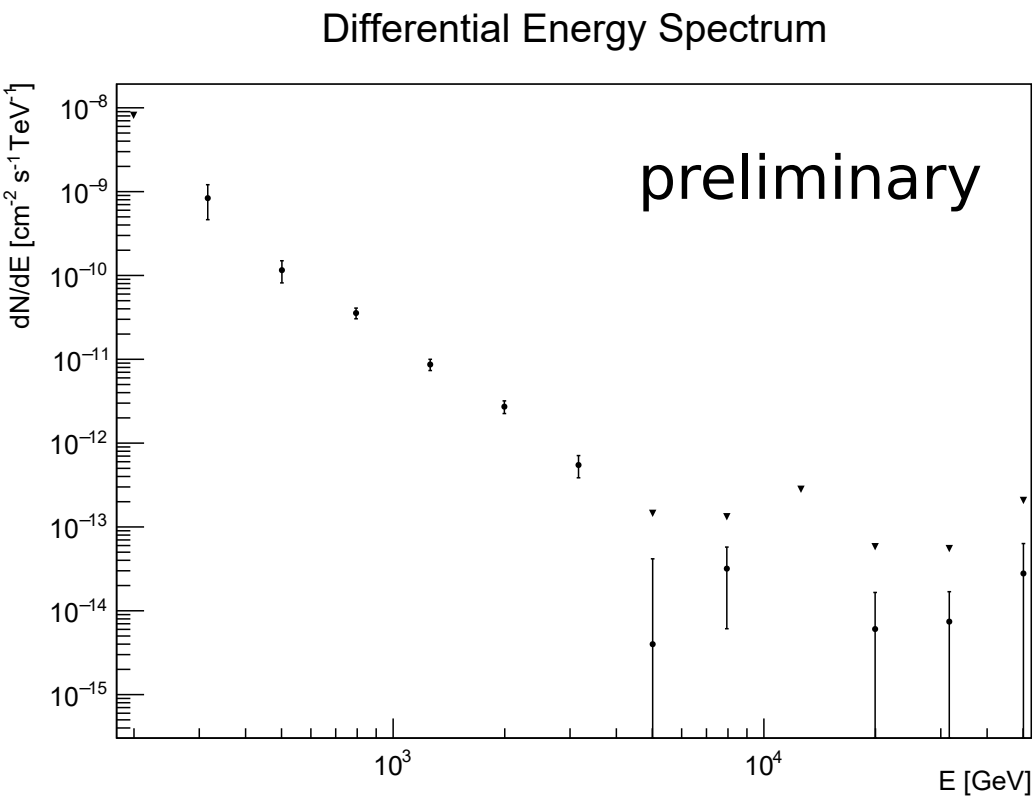

Figure 1: Differential energy spectrum of the Blazar Mrk 421 for data from 11.01.2019 until 14.01.2019. The black triangles show upper limit calculations using the Rolke method. For the highest energies, the errors of the data points increase rapidly since the number of events per energy bin is very low.

In Figure 1, the result of a spectrum calculation is shown for data of Mrk 421. The time range of the data starts at 11.01.2019 after midnight and ends at the morning of the 14.01.2019. Only data with a trigger threshold of less than 560 DAC-counts and zenith distance smaller than 30 deg are selected, and an automatic data check is applied [8] to exclude data that are taken under not perfect conditions like clouds or dust. One can see that the data points for the highest energies (> 3 TeV) have significantly larger error bars than the other ones. The reason for that is the limited number of events with these energies. Therefore the number of events per bin is so low that this points should not be taken into account. Instead, the black triangles show upper limits calculated with the Rolke method [9].

4. Conclusion: Database-based Analysis – A Flexible and User-Friendly Approach

The analysis examples discussed above illustrate the flexibility of the database-based analysis approach. In the following, we summarize the advantages of storing the event data in a database instead of files.

Querying the events from a database removes a lot of complexity from the process and provides the user a fast and very simple way to access and explore the data. Therefore, it is very easy to get started with analysing the data providing more time for high level studies and the physics

PoS(ICRC2021)760

interpretation. The preparatory work is reduced to installing a MySQL client on a computer with a network connection and acquiring some very basic knowledge in SQL queries. With that setup, a user can start exploring the database of FACT and start to analyze data immediately. The users then can choose if they want to do the full high-level analysis with one SQL-query, or if they prefer to select and download data for further local processing. For example, they can choose data based on parameters like date, source or zenith distance, or one of the more parameters available for the data selection. Also a combination of selections can easily be chosen. The result is a list of events, where the user can also choose which parameters to be extracted. The output can be stored in a file format of the user's choice and then processed by any tool. As example, for a background suppression only some basic image parameters are needed and therefore the resulting file size on the user's computer is very small compared to the standard files containing any parameter available. The users have also free choice of the programming language which they use for further analysis. Alternatively, the users can do the complete high level analysis as one (eventually rather complex) query instead of using different programs and processing the data step by step as it is done in the standard analysis. This one step solution also has the advantage that only a very small amount of data has to be stored locally on the user's machine. Also the access to all available data can be structured very easy, because there are no individual user accounts needed to give access to the data. A read-only database account is sufficient to get the data from the database. The easy access without specific user accounts is very useful for students to get started working as fast as possible. It is also ideal to grant universal access to public data. Another possibility is to use web pages to connect to the database. The time range and source can be chosen on the web interface and a list of light curve data points is delivered as a csv file. To reduce the calculation time of such data requests, it is useful to also store the results of the Quick-Look Analysis (QLA) and other predefined offline analyses (see details in [2], [3] and [4]) in a separate table of the database on a run-basis. This allows to just access the light curve results there and bin them accordingly to the request.

At the moment, the major instruments push for the use of data-level 3 (dl3) and analysing them with software developed for this kind of data, like for example gammapy [10, 11] or ctools [12]. In this approach, the data are provided as an event list and the characteristics of an instrument are given via a predefined, fixed instrument-response-function (IRF). The database approach could used to provide the dl3 data and the IRF. With simple queries, it would be possible to provide custom-made IRFs for a specific request taking into account any possible set of cuts on the data. A further step to more user friendliness would be to provide a website, where one can easily request the custom-made dl3 data and IRFs.

Acknowledgements The important contributions from ETH Zurich grants ETH-10.08-2 and ETH-27.12-1 as well as the funding by the Swiss SNF and the German BMBF (Verbundforschung Astro- und Astroteilchenphysik) and HAP (Helmoltz Alliance for Astroparticle Physics) are gratefully acknowledged. Part of this work is supported by Deutsche Forschungsgemeinschaft (DFG) within the Collaborative Research Center SFB 876 "Providing Information by Resource-Constrained Analysis", project C3. We are thankful for the very valuable contributions from E. Lorenz, D. Renker and G. Viertel during the early phase of the project. We thank the Instituto de Astrofísica de Canarias for allowing us to operate the telescope at the Observatorio del Roque de los Muchachos in La Palma, the Max-Planck-Institut für Physik for providing us with the mount of the former HEGRA CT3 telescope, and the MAGIC collaboration for their support.

PoS(ICRC2021)760

References

[1] T. Bretz and D. Dorner, *MARS - CheObs ed. – A flexible Software Framework for future Cherenkov Telescopes*, in *Astroparticle, Particle and Space Physics, Detectors and Medical Physics Applications*, C. Leroy, P.-G. Rancoita, M. Barone, A. Gaddi, L. Price and R. Ruchti, eds., pp. 681–687, Apr., 2010, DOI.

[2] D. Dorner, M.L. Ahnen, M. Bergmann, A. Biland, M. Balbo, T. Bretz et al., *FACT - Monitoring Blazars at Very High Energies*, *arXiv e-prints* (2015) arXiv:1502.02582 [1502.02582].

[3] D. Dorner, FACT Collaboration, J. Adam, L.M. Ahnen, D. Baack, M. Balbo et al., *FACT - Time-Resolved Blazar SEDs*, in *35th International Cosmic Ray Conference (ICRC2017)*, vol. 301 of *International Cosmic Ray Conference*, p. 608, Jan., 2017.

[4] M. Beck, A. Arbet-Engels, D. Baack, M. Balbo, A. Biland, M. Blank et al., *FACT - Probing the Periodicity of Mrk 421 and Mrk 501*, in *36th International Cosmic Ray Conference (ICRC2019)*, vol. 36 of *International Cosmic Ray Conference*, p. 630, July, 2019.

[5] R.W. Lessard, J.H. Buckley, V. Connaughton and S. Le Bohec, *A new analysis method for reconstructing the arrival direction of TeV gamma rays using a single imaging atmospheric Cherenkov telescope*, *Astroparticle Physics* **15** (2001) 1 [astro-ph/0005468].

[6] C. Romoli, N. Chakraborty, D. Dorner, A. Taylor and M. Blank, *Flux Distribution of Gamma-Ray Emission in Blazars: The Example of Mrk 501*, *Galaxies* **6** (2018) 135 [1812.06204].

[7] B. Schleicher, A. Arbet-Engels, D. Baack, M. Balbo, A. Biland, M. Blank et al., *Fractional Variability—A Tool to Study Blazar Variability*, *Galaxies* **7** (2019) 62.

[8] A. Arbet-Engels, D. Baack, M. Balbo, A. Biland, M. Blank, T. Bretz et al., *The relentless variability of Mrk 421 from the TeV to the radio*, *Astronomy and Astrophysics* **647** (2021) A88 [2101.10651].

[9] W.A. Rolke, A.M. López and J. Conrad, *Limits and confidence intervals in the presence of nuisance parameters*, *Nuclear Instruments and Methods in Physics Research A* **551** (2005) 493 [physics/0403059].

[10] C. Deil, R. Zanin, J. Lefaucheur, C. Boisson, B. Khelifi, R. Terrier et al., *Gammapy - A prototype for the CTA science tools*, in *35th International Cosmic Ray Conference (ICRC2017)*, vol. 301 of *International Cosmic Ray Conference*, p. 766, Jan., 2017 [1709.01751].

[11] C. Nigro, C. Deil, R. Zanin, T. Hassan, J. King, J.E. Ruiz et al., *Towards open and reproducible multi-instrument analysis in gamma-ray astronomy*, *Astronomy and Astrophysics* **625** (2019) A10 [1903.06621].

[12] J. Knödlseder, M. Mayer, C. Deil, J.B. Cayrou, E. Owen, N. Kelley-Hoskins et al., *GammaLib and ctools. A software framework for the analysis of astronomical gamma-ray data*, *Astronomy and Astrophysics* **593** (2016) A1 [1606.00393].

PoS(ICRC2021)760