Journal of High Energy Astrophysics 44 (2024) 266–278

Contents lists available at ScienceDirect

Journal of High Energy Astrophysics

journal homepage: www.elsevier.com/locate/jheap

# Standardised formats and open-source analysis tools for the MAGIC telescopes data

The MAGIC Collaboration

S. Abe [a], J. Abhir [b], A. Abhishek [c], V.A. Acciari [d], A. Aguasca-Cabot [e], I. Agudo [f], T. Aniello [g], S. Ansoldi [h,ao], L.A. Antonelli [g], A. Arbet Engels [i], C. Arcaro [j], M. Artero [d], K. Asano [a], A. Babić [k], U. Barres de Almeida [l], J.A. Barrio [m], I. Batković [j], A. Bautista [i], J. Baxter [a], J. Becerra González [n], W. Bednarek [o], E. Bernardini [j], J. Bernete [p], A. Berti [i], J. Besenrieder [i], C. Bigongiari [g], A. Biland [b], O. Blanch [d], G. Bonnoli [g], Ž. Bošnjak [k], E. Bronzini [g], I. Burelli [h], G. Busetto [j], A. Campoy-Ordaz [q], A. Carosi [g], R. Carosi [r], M. Carretero-Castrillo [e], A.J. Castro-Tirado [f], D. Cerasole [s], G. Ceribella [i], Y. Chai [a], A. Cifuentes [p], E. Colombo [d], J.L. Contreras [m], J. Cortina [p], S. Covino [g], G. D'Amico [t], V. D'Elia [g], P. Da Vela [g], F. Dazzi [g], A. De Angelis [j], B. De Lotto [h], R. de Menezes [u], M. Delfino [d,ap], J. Delgado [d,ap], F. Di Pierro [u], R. Di Tria [s], L. Di Venere [s], D. Dominis Prester [v], A. Donini [g], D. Dorner [w], M. Doro [j], D. Elsaesser [x], J. Escudero [f], L. Fariña [d], A. Fattorini [x], L. Foffano [g], L. Font [q], S. Fröse [x], S. Fukami [b], Y. Fukazawa [y], R.J. García López [n], M. Garczarczyk [z], S. Gasparyan [aa], M. Gaug [q], J.G. Giesbrecht Paiva [l], N. Giglietto [s], F. Giordano [s], P. Gliwny [o], T. Gradetzke [x], R. Grau [d], D. Green [i], J.G. Green [i], P. Günther [w], D. Hadasch [a], A. Hahn [i], T. Hassan [p], L. Heckmann [i], J. Herrera Llorente [n], D. Hrupec [ab], M. Hütten [a], R. Imazawa [y], K. Ishio [o], I. Jiménez Martínez [i], J. Jormanainen [ac], T. Kayanoki [y], D. Kerszberg [d], G.W. Kluge [t,aq], Y. Kobayashi [a], P.M. Kouch [ac], H. Kubo [a], J. Kushida [ad], M. Láinez [m], A. Lamastra [g], F. Leone [g], E. Lindfors [ac], S. Lombardi [g], F. Longo [h,ar], R. López-Coto [f], M. López-Moya [m], A. López-Oramas [n], S. Loporchio [s], A. Lorini [c], E. Lyard [ae], B. Machado de Oliveira Fraga [l], P. Majumdar [af], M. Makariev [ag], G. Maneva [ag], M. Manganaro [v], S. Mangano [p], K. Mannheim [w], M. Mariotti [j], M. Martínez [d], M. Martínez-Chicharro [p], A. Mas-Aguilar [m], D. Mazin [a,i], S. Menchiari [f], S. Mender [x], D. Miceli [j], T. Miener [m], J.M. Miranda [c], R. Mirzoyan [i], M. Molero González [n], E. Molina [n], H.A. Mondal [af], A. Moralejo [d], D. Morcuende [f], T. Nakamori [ah], C. Nanci [g], V. Neustroev [ai], L. Nickel [x], M. Nievas Rosillo [n], C. Nigro [d,*], L. Nikolić [c], K. Nishijima [ad], T. Njoh Ekoume [d], K. Noda [aj], S. Nozaki [i], Y. Ohtani [a], A. Okumura [ak], J. Otero-Santos [f], S. Paiano [g], D. Paneque [i], R. Paoletti [c], J.M. Paredes [e], M. Peresano [i], M. Persic [h,as], M. Pihet [j], G. Pirola [i], F. Podobnik [c], P.G. Prada Moroni [r], E. Prandini [j], G. Principe [h], C. Priyadarshi [d], W. Rhode [x], M. Ribó [e], J. Rico [d], C. Righi [g], N. Sahakyan [aa], T. Saito [a], F.G. Saturni [g], K. Schmidt [x], F. Schmuckermaier [i], J.L. Schubert [x], T. Schweizer [i], A. Sciaccaluga [g], G. Silvestri [j], J. Sitarek [o], V. Sliusar [ae], D. Sobczynska [o], A. Spolon [j], A. Stamerra [g], J. Strišković [ab], D. Strom [i], M. Strzys [a], Y. Suda [y], S. Suutarinen [ac], H. Tajima [ak], M. Takahashi [ak], R. Takeishi [a], P. Temnikov [ag], K. Terauchi [al], T. Terzić [v], M. Teshima [i,a], S. Truzzi [c], A. Tutone [g], S. Ubach [q], J. van Scherpenberg [i], M. Vazquez Acosta [n], S. Ventura [c], I. Viale [j], C.F. Vigorito [u], V. Vitale [am],

* Corresponding authors. Send offprints requests to: contact.magic@mpp.mpg.de.

https://doi.org/10.1016/j.jheap.2024.09.011
Received 9 June 2024; Received in revised form 8 August 2024; Accepted 24 September 2024
Available online 9 October 2024
2214-4048/© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

I. Vovk [a], R. Walter [ae], M. Will [i], C. Wunderlich [c], T. Yamamoto [an], L. Jouvin [d,at], L. Linhoff [x,*], M. Linhoff [x]

[a] Japanese MAGIC Group: Institute for Cosmic Ray Research (ICRR), The University of Tokyo, Kashiwa, 277-8582 Chiba, Japan
[b] ETH Zürich, CH-8093 Zürich, Switzerland
[c] Università di Siena and INFN Pisa, I-53100 Siena, Italy
[d] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology (BIST), E-08193 Bellaterra (Barcelona), Spain
[e] Universitat de Barcelona, ICCUB, IEEC-UB, E-08028 Barcelona, Spain
[f] Instituto de Astrofísica de Andalucía-CSIC, Glorieta de la Astronomía s/n, 18008, Granada, Spain
[g] National Institute for Astrophysics (INAF), I-00136 Rome, Italy
[h] Università di Udine and INFN Trieste, I-33100 Udine, Italy
[i] Max-Planck-Institut für Physik, D-85748 Garching, Germany
[j] Università di Padova and INFN, I-35131 Padova, Italy
[k] Croatian MAGIC Group: University of Zagreb, Faculty of Electrical Engineering and Computing (FER), 10000 Zagreb, Croatia
[l] Centro Brasileiro de Pesquisas Físicas (CBPF), 22290-180 URCA, Rio de Janeiro (RJ), Brazil
[m] IPARCOS Institute and EMFTEL Department, Universidad Complutense de Madrid, E-28040 Madrid, Spain
[n] Instituto de Astrofísica de Canarias and Dpto. de Astrofísica, Universidad de La Laguna, E-38200, La Laguna, Tenerife, Spain
[o] University of Lodz, Faculty of Physics and Applied Informatics, Department of Astrophysics, 90-236 Lodz, Poland
[p] Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, E-28040 Madrid, Spain
[q] Departament de Física, and CERES-IEEC, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain
[r] Università di Pisa and INFN Pisa, I-56126 Pisa, Italy
[s] INFN MAGIC Group: INFN Sezione di Bari and Dipartimento Interateneo di Fisica dell'Università e del Politecnico di Bari, I-70125 Bari, Italy
[t] Department for Physics and Technology, University of Bergen, Norway
[u] INFN MAGIC Group: INFN Sezione di Torino and Università degli Studi di Torino, I-10125 Torino, Italy
[v] Croatian MAGIC Group: University of Rijeka, Faculty of Physics, 51000 Rijeka, Croatia
[w] Universität Würzburg, D-97074 Würzburg, Germany
[x] Technische Universität Dortmund, D-44221 Dortmund, Germany
[y] Japanese MAGIC Group: Physics Program, Graduate School of Advanced Science and Engineering, Hiroshima University, 739-8526 Hiroshima, Japan
[z] Deutsches Elektronen-Synchrotron (DESY), D-15738 Zeuthen, Germany
[aa] Armenian MAGIC Group: ICRANet-Armenia, 0019 Yerevan, Armenia
[ab] Croatian MAGIC Group: Josip Juraj Strossmayer University of Osijek, Department of Physics, 31000 Osijek, Croatia
[ac] Finnish MAGIC Group: Finnish Centre for Astronomy with ESO, Department of Physics and Astronomy, University of Turku, FI-20014 Turku, Finland
[ad] Japanese MAGIC Group: Department of Physics, Tokai University, Hiratsuka, 259-1292 Kanagawa, Japan
[ae] University of Geneva, Chemin d'Ecogia 16, CH-1290 Versoix, Switzerland
[af] Saha Institute of Nuclear Physics, A CI of Homi Bhabha National Institute, Kolkata 700064, West Bengal, India
[ag] Inst. for Nucl. Research and Nucl. Energy, Bulgarian Academy of Sciences, BG-1784 Sofia, Bulgaria
[ah] Japanese MAGIC Group: Department of Physics, Yamagata University, Yamagata 990-8560, Japan
[ai] Finnish MAGIC Group: Space Physics and Astronomy Research Unit, University of Oulu, FI-90014 Oulu, Finland
[aj] Japanese MAGIC Group: Chiba University, ICEHAP, 263-8522 Chiba, Japan
[ak] Japanese MAGIC Group: Institute for Space-Earth Environmental Research and Kobayashi-Maskawa Institute for the Origin of Particles and the Universe, Nagoya University, 464-6801 Nagoya, Japan
[al] Japanese MAGIC Group: Department of Physics, Kyoto University, 606-8502 Kyoto, Japan
[am] INFN MAGIC Group: INFN Roma Tor Vergata, I-00133 Roma, Italy
[an] Japanese MAGIC Group: Department of Physics, Konan University, Kobe, Hyogo 658-8501, Japan
[ao] Also at International Center for Relativistic Astrophysics (ICRA), Rome, Italy
[ap] Also at Port d'Informació Científica (PIC), E-08193 Bellaterra (Barcelona), Spain
[aq] Also at Department of Physics, University of Oslo, Norway
[ar] Also at Dipartimento di Fisica, Università di Trieste, I-34127 Trieste, Italy
[as] Also at INAF Padova, Italy
[at] Now at IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

## ARTICLE INFO

Dataset link: https://opendata.magic.pic.es/

Dataset link: https://zenodo.org/records/11108473

*Keywords:*
Gamma-ray astronomy
Very high energies
Imaging atmospheric Cherenkov telescopes
Open-source software
Data format
Reproducibility

## ABSTRACT

Instruments for gamma-ray astronomy at Very High Energies ($E > 100\,\mathrm{GeV}$) have traditionally derived their scientific results through proprietary data and software. Data standardisation has become a prominent issue in this field both as a requirement for the dissemination of data from the next generation of gamma-ray observatories and as an effective solution to realise public data legacies of current-generation instruments. Specifications for a standardised gamma-ray data format have been proposed as a community effort and have already been successfully adopted by several instruments. We present the first production of standardised data from the Major Atmospheric Gamma-ray Imaging Cherenkov (MAGIC) telescopes. We converted 166 h of observations from different sources and validated their analysis with the open-source software `Gammapy`. We consider six data sets representing different scientific and technical analysis cases and compare the results obtained analysing the standardised data with open-source software against those produced with the MAGIC proprietary data and software. Aiming at a systematic production of MAGIC data in this standardised format, we also present the implementation of a database-driven pipeline automatically performing the MAGIC data reduction from the calibrated down to the standardised data level. In all the cases selected for the validation, we obtain results compatible with the MAGIC proprietary software, both for the manual and for the automatic data productions. Part of the validation data set is also made publicly available, thus representing the first large public release of MAGIC data. This effort and this first data release represent a technical milestone toward the realisation of a public MAGIC data legacy.

## 1. Introduction

The free exchange of astronomical data gathered at different wavelengths joined the different branches of astronomy in the so-called multi-wavelength (MWL) domain, demonstrating the scientific potential intrinsic to the study of the emission of sources in different energy ranges. While gamma-ray astronomy at high energies (HE, 100 MeV $< E <$ 100 GeV) adopted the same policy of providing public data and software tools (Barrett, 1995; Pittori, 2013; Band, 2007), the scientific activity of very-high-energy (VHE, $E > 100$ GeV) gamma-ray instruments has been traditionally defined instead by proprietary data and software. The issue of data standardisation in VHE gamma-ray astronomy emerged in the last decade from the decision to build the future Cherenkov Telescope Array Observatory (CTAO), as an open observatory, sharing its observational time and data with the astronomical community (Lamanna et al., 2015; Contreras et al., 2015). Compelling as the case for data standardisation might be for the dissemination of future VHE data, it represents a crucial issue for present instruments as well. The current (third) generation of Imaging Atmospheric Cherenkov Telescopes (IACTs, Hillas 2013; de Naurois and Mazin 2015) has already operated for two decades and has accumulated a wealth of data that cannot be fully explored by the restricted groups of scientists operating these instruments. The full scientific exploitation of these data sets, even beyond the decommissioning of the telescopes, ideally calls for the realisation of public archival data releases, or "data legacies". The adoption of a common format for these data legacies would not only ease their access and usage by the community, since a legacy data release in its native format would require the corresponding proprietary software tool to be maintained and released as well; it would also allow for the exploration and combination of decades of archival gamma-ray observations.

In the context of future and current VHE gamma-ray instruments, the effort to define a standard data format started in the second half of the 2010s with the creation of the *Data Formats for Gamma-ray Astronomy* (GADF) initiative (Deil et al., 2017; Nigro et al., 2021). The GADF is a forum where the specifications of a data format for high-level gamma-ray data are discussed through the `GitHub` workflow.[1] In parallel to the GADF specifications, and likewise motivated by the upcoming CTAO, open-source software for the analysis of gamma-ray data, such as `ctools` (Knödlseder et al., 2016) and `Gammapy` (Donath et al., 2023), were developed. Their data routines were built in compliance with the GADF specifications, making it possible for the first time to analyse standardised VHE gamma-ray data with open-source software. Other noteworthy open-source software initiatives are: the publication of the `EventDisplay` reconstruction and analysis software (Maier and Holder, 2017; Maier et al., 2024) used both for the VERITAS (Very Energetic Radiation Imaging Telescope Array System) and the CTAO data analysis; and the open development of the reconstruction software for the Large-Sized Telescope prototype (LST-1, Lopez-Coto et al. 2024) of the CTAO. Both software, though operating on proprietary data can produce GADF-compliant data. The first prototypical sample of data in the standardised GADF format was publicly released by the High Energy Stereoscopic System (H.E.S.S., H. E. S. S. Collaboration 2018) and was crucial in validating the capabilities of the aforementioned open-source analysis tools (Mohrmann et al., 2019; Knödlseder et al., 2019). The First g-Apd Cherenkov Telescope (FACT) also made all of its Crab Nebula observations public in the GADF format.[2] Nigro et al. (2019), also known as the `joint-crab` project, then demonstrated the possibility of effectively performing combined analyses of gamma-ray data from different instruments adopting the standard data format. Data from Crab Nebula observations performed by the HE satellite *Fermi* Large Area Telescope (LAT) and by all the then-operating IACTs were produced in the GADF format and jointly analysed with `Gammapy`. Beside the already mentioned public H.E.S.S. and FACT data, the VERITAS, and the Major Atmospheric Gamma-ray Imaging Cherenkov (MAGIC) telescopes produced GADF-compliant data specifically for this project. The `joint-crab` dataset thus represented the first joint release of IACT data to the public. Expanding the adoption of the format further, the High Altitude Water Cherenkov (HAWC) Observatory demonstrated that also arrays of particle detectors could produce their gamma-ray data in the GADF format (Albert et al., 2022) and extended the exemplary `joint-crab` combined spectrum measurement to 5 orders of magnitude in energy.

While for the `joint-crab` demonstration only 40 min of MAGIC data were converted to the GADF format and publicly released, the Collaboration, with the long-term objective of a public data legacy, initiated the effort to systematically produce observations in this standardised format. The objective of this paper is to present this endeavour and to make a first public release of MAGIC standardised data. We analyse a total of 166 h of observations from five different sources representing different scientific and technical analysis cases. We validate the analysis of these standardised data with the open-source software `Gammapy` by comparing the results obtained against those produced on the same data sets with the proprietary MAGIC Analysis and Reconstruction Software (`MARS`, Moralejo et al., 2009; Zanin et al., 2013). In this paper, we focus on the validation of the point-like, or one-dimensional, analysis, in which any spatial extension or morphology of the sources is neglected.

This paper is structured as follows: in Sect. 2 we describe the software we developed to convert the MAGIC data from their native format to the standardised one. In Sect. 3 we detail the observations used for this study. In Sect. 4 we illustrate the different analyses performed with `Gammapy` on the standardised data and the comparison with the corresponding results obtained with `MARS`. We conclude in Sect. 5 by providing some perspectives on the future data legacy of the MAGIC telescopes. In Appendix A and Appendix B, we present detailed comparisons of the result of the flux estimation with the two software and investigate the differences produced by the different algorithms adopted. We finally provide public access to part of the dataset adopted for the validation in the section Data availability.

## 2. Conversion of MAGIC data to the standardised format

As proposed in Contreras et al. (2015), one could schematise the progressive data reduction performed by gamma-ray instruments in five different data levels with level 0 representing the raw output of the data acquisition and level 5 scientific results, for example the estimation of the gamma-ray flux of a source. Aiming at facilitating reproducibility and data combination, the GADF provides specifications for the data level 3 (DL3), containing detector- and calibration-independent information that can be directly used to perform a statistical analysis. DL3 data contain two components: a list of events classified as gamma rays with their estimated coordinates, energies and arrival times and a parametrisation of the response of the system, the so-called instrument response function (IRF), necessary to transform the detector information (e.g. counts) into physical quantities (e.g. fluxes). Building on the experience of the public *Fermi*-LAT data,[3] and following the format recommendations for high-energy astrophysical data by NASA's High Energy Astrophysics Science Archive Research Center (HEASARC),[4] the GADF specifications are defined for the Flexible Image Transport System (`FITS`, Wells et al., 1981) file format.

### 2.1. `magic_dl3`

The MAGIC proprietary analysis and reconstruction software, `MARS`, is a `C++` library built on `ROOT` (Brun and Rademakers, 1997) that provides several executables performing the different data-reduction steps,

[1] https://github.com/open-gamma-ray-astro/gamma-astro-data-formats.
[2] https://factdata.app.tu-dortmund.de/.
[3] https://fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/Cicerone/Cicerone_Data/LAT_DP.html.
[4] https://heasarc.gsfc.nasa.gov/docs/heasarc/ofwg/ofwg_recomm.html.

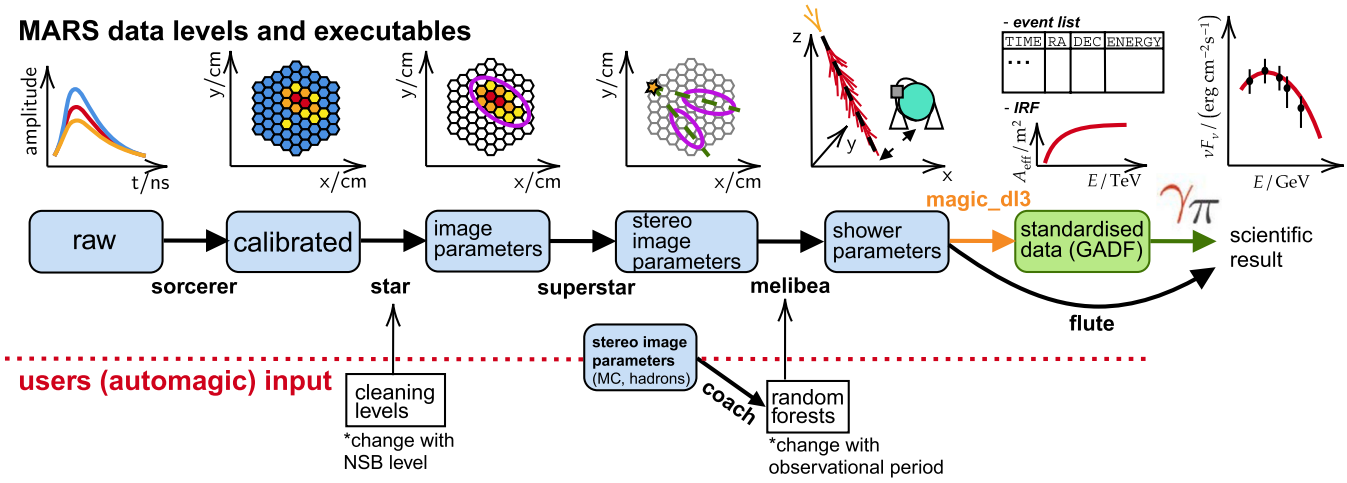


**Fig. 1.** Schematic view of the MAGIC data reduction from raw data to final scientific results. Blue blocks illustrate the different `MARS` proprietary data levels, black arrows connecting them represent the `MARS` proprietary executables reducing them. Data levels take the name of the executable producing them (indicated on the arrows). White boxes are used to represent the data-reduction parameters and inputs provided manually by the user or automatically by the `automagic` pipeline. Green boxes and arrow represent open data and software. The production of GADF-compliant data starts from `melibea` files (containing the reconstructed shower parameters) using the library developed for this paper (orange arrow). Open-source software such as `Gammapy` can be then used to extract scientific results from the standardised data.

schematised in Fig. 1, from raw data to scientific results. Though each data level can be serialised using the `ROOT` file format, none of them exactly corresponds to the GADF DL3 specifications.[5] Thus, we developed a proprietary library, `magic_dl3`, extracting the requested information from the `MARS` files and storing it in the GADF-compliant format. `magic_dl3` is a `C++` library built on `ROOT`, `MARS`, and `CFITSIO` (Pence, 1999) that operates on the `MARS melibea` data level (see Fig. 1). `melibea` files contain reconstructed shower parameters and an estimate of the energy, direction, and arrival time of the primary particle of a set of observed (or simulated) events. Along with these quantities, they specify, per each event, the score of a particle identification algorithm (random forest, Albert et al., 2008). `melibea` data produced from observations are used to extract the first component of the DL3 files, that is the event list, by applying a cut on the random forest score to select a list of "gamma-like" events. The second component of the DL3 files, the IRF, is composed of different functions representing the collection area and the probability distributions of the energy and direction estimators (see Sec. 4.1 and Aleksić et al., 2016, for more details). Histograms representing these IRF components are built from Monte Carlo (MC) events, after applying the same selection cuts adopted for the event list selection. As the `MARS` data reduction down to the `melibea` data level is still necessary to produce DL3 data, in the next section we introduce a tool developed to simplify and automatise the systematic generation of DL3 files.

### 2.2. `automagic`

Different observing conditions can easily multiply the number of parameters (image cleaning settings, data selection and analysis cuts) and input data sets necessary to perform a `MARS` data reduction. Different MC gamma-ray simulations are produced according to significant changes in the hardware configuration of the telescopes (roughly one per year). Simulated gamma rays, along with observations containing no gamma-ray signal (commonly referred to as hadrons),[6] are used to train the particle identification algorithm. This implies, for analysers considering a data set spanning several years, the necessity to train a number of different algorithms, one per each of the hardware configuration (i.e. MC productions) being considered. While the data reduction would normally start from already computed stereo image parameters, MAGIC developed a procedure to analyse data taken under different moon conditions (described in Ahnen et al., 2017). The latter requires users to start the analysis from the calibrated data level (see Fig. 1) and to perform the cleaning of the shower images with parameters tuned to suppress the pixels illuminated by the diffuse night-sky background (NSB). In case of data taken under different NSB conditions, not only the observations, but also the MC and the hadrons shower images require this tuned cleaning. Therefore the analysis of a data set characterised by different MC productions and NSB conditions requires analysers to handle hundreds of GB or TB of data, and perform a cumbersome number of different low-level data reduction steps. To orchestrate these different processes, we developed `automagic`, a database-supported `python` pipeline that automatically performs all the data reduction steps from calibrated down to `melibea` and DL3 files. The project was started with the aim to simplify individual analyses, but was soon adopted by the MAGIC Collaboration to systematically process, in a reproducible and automatic way, large volumes of data and directly produce high-level data sets. The pipeline runs in the computing cluster of the MAGIC Data Center, at the Port d'Informació Científica (PIC).[7] Its input consists simply of a source name, a given time period, and the data selection cuts (e.g. zenith range, atmospheric transmission, etc.). Having access to the whole archive of MAGIC data, `automagic` localises, in the PIC file system, the calibrated data corresponding to the user selection and creates database tables representing the data files at different stages of data reduction (down to DL3). The pipeline identifies the MC periods within the selected data set and automatically selects the appropriate hadrons and MC data to train the corresponding classification algorithms. It additionally classifies the NSB levels of the observations and configures the

[5] Notice that the `MARS` executable used to produce scientific results, `flute` (Fig. 1), directly performs a data reduction (see Sect. 4.1), without storing the event list information. Additionally, it can compute IRF components only for MCs corresponding to a specific observation configuration (see the discussion in Sect. 3.1.1).

[6] While MC gamma rays are readily available to train the classification algorithm, in order to save the extensive computational time of hadronic shower simulations, hadrons data are manually selected by users examining other observations with no gamma-ray signal that match the observing conditions of the data of interest.

[7] https://www.pic.es/.

**Table 1**
MAGIC data sets used for the DL3 data validation.

| source | period | dark | obs. time | science case | technical case | size (n. of runs) |
|---|---|---|---|---|---|---|
| Crab Nebula | 2011-2012 | yes | 42 h | bright steady source | different offsets | 45 MB (169) |
| Crab Nebula | 2018-2019 | no | 20 h | bright steady source | different moon conditions | 14 MB (120) |
| IC310 | 2012 | yes | 3.5 h | bright hard source | different offsets | 3 MB (11) |
| B0218+357 | 2014 | yes | 2 h | dim soft source | - | 2 MB (7) |
| Mrk421 | 2014 | yes | 42 h | bright variable source | - | 29 MB (176) |
| M15 | 2015-2016 | yes | 57 h | non detection | - | 47 MB (200) |

appropriate settings for the image cleaning (see Fig. 1). Having identified all the input data and the configurations for the different data reduction steps, `automagic` runs them sequentially in the PIC computing nodes. By parallelising the execution of the individual reduction steps and by creating database tables to also track the status of the different jobs, `automagic` is able to manage large data volumes and to perform a reproducible and resource-efficient data reduction. The `automagic` pipeline is proprietary to the Collaboration, as it is designed to work within the MAGIC Data Center at PIC.

## 3. Data samples selected for the validation

Having implemented the tools to produce standardised GADF-compliant DL3 data, we aim to validate that their analysis produces results consistent with those obtained with the MAGIC proprietary software. For the analysis of the standardised data we adopt the open-source software `Gammapy` (`v1.1`, Acero et al., 2021b), the software tool that will also form the basis of the future CTAO analysis software. In this paper, we focus on the validation of the point-like or one-dimensional analysis: signal and background events are extracted from fixed regions in the sky and their distribution as a function of the energy is fitted. This analysis case, suited for individual sources with extension below the instrument's point spread function, is appropriate for most of the MAGIC observations. To test the use cases most commonly considered in a one-dimensional analysis, we choose 6 data sets representing different scientific and technical scenarios. They are presented in Table 1 and described in detail in what follows.

### 3.1. Crab Nebula

The Crab Nebula represents a reference source for VHE gamma-ray astronomy, being the brightest steady source emitting above hundreds of GeV (exceptional gamma-ray flares have been detected only at HE, see Abdo et al., 2011b; Tavani et al., 2011). Indeed, it was the first source to be detected at VHE by Weekes et al. (1989) (see also Bühler and Blandford, 2014, for an overview of the source and its emission mechanisms). We select two different samples of MAGIC observations of the Crab Nebula related to two different technical aspects of the standardised data production we aim at validating.

#### 3.1.1. Observations at different pointing offsets

The first sample consists of 42 h of Crab Nebula observations performed between 2011 and 2012, that we consider to validate two scientific results most commonly obtained in a one-dimensional analysis: spectrum and light curve. These data contain the sample of observations used to estimate the MAGIC performance after the upgrade of the stereoscopic system completed in 2012 (Aleksić et al., 2016). Most of the MAGIC observations are conducted in the so-called wobble mode (Fomin et al., 1994), with the telescopes tracking sky coordinates which are typically 0.4° from the source nominal position. This results in the source having a projected position in the camera plane 0.4° from its centre, which facilitates the background estimation (see Fig. 2). As this configuration represents most of the observations performed by MAGIC, MC gamma rays are commonly simulated with arrival directions at 0.4° from the telescope axis. MCs thus produced are referred to as "ring-wobble" simulations, and are adopted for most of the analyses. As the IRF dependency in the GADF specification is expressed in projected camera coordinates (or offset from the centre, in case of radial symmetry) rather than sky coordinates, we refer to IRFs generated from these MCs as "single-offset", since the offset dependency of the IRF components is restricted to a single value (0.4°). To take into account wobble-mode observations with different pointing offsets, or to study extended sources, "diffuse" MCs are produced, not applying the previous restriction on the direction of the simulated events. We refer to IRFs generated from this MCs as "multi-offset", as the full dependence of the IRF components with the offset from the camera centre is considered (all the IRFs adopted in these tests are radially symmetric in camera coordinates). To test the spectral analysis at several pointing offsets, we first select 30 h of Crab Nebula observations at the standard 0.4° offset and produce single-offset IRFs. We then add to the sample 20 h of Crab Nebula observations with different pointing offsets: $0.20°, 0.35°, 0.40°, 0.70°, 1.00°, 1.40°$, producing multi-offset IRFs. Note that the multi-offset data at 0.40° are a subset of the 30 h single-offset sample, but reprocessed with diffuse MC. The zenith range of the single-offset sample is $[5°, 50°]$, while that of multi-offset sample is $[5°, 35°]$. MC gamma rays are generated with an uniform distribution of cosine of zenith and azimuth. In order to account for the zenith and azimuth dependency of the IRF, MCs are re-weighted according to the effective time distribution in zenith and azimuth of the real events. Proprietary and standardised data sets are composed of observational "runs", corresponding to chunks of 15 to 20 min of data acquisition. The number of runs per data set is indicated in Table 1, along with the size of the standardised data. For a comparison with the proprietary format, in the case of the Crab Nebula single-offset data set, we converted 21 GB of `melibea` files, the reduction factor from `melibea` to DL3 thus being $\sim 500$.

#### 3.1.2. Observations under different moonlight conditions

Another sample of Crab Nebula observations is used to test the reliability of the high-level data produced by the `automagic` pipeline. As already described in Sect. 2.2, `automagic` is designed to produce MAGIC DL3 data automatising the process of data selection and reduction, noticeably simplifying cumbersome analyses as those of observations taken under different moon conditions. To test `automagic`'s capabilities, we consider a sample of 20 h of Crab Nebula observations with four different moon/NSB illumination conditions. The NSB level was measured in units of the mean direct currents generated in the photo-multiplier tubes of the MAGIC I camera, with "dark" observations corresponding to mean currents typically around $1\,\mu A$ (see Ahnen et al., 2017). These moon observations were performed between November 2018 and September 2019, with pointing offset 0.4° and zenith range $[5°, 50°]$. We manually perform the data selection and reduction on this data set while, in parallel, let `automagic` perform the same processes by specifying as the only input of the pipeline the source name and the start and end dates of the observations to be considered. We finally compare the spectra obtained with the manual and the automatic approaches.

### 3.2. QSO B0218+357 and IC310

We include the jetted active galactic nuclei QSO B0218+357 and IC310 in the validation sample to test the spectrum estimation in case of a soft and hard gamma-ray source, respectively. Given the non-negligible absorption of their gamma-ray emission by the extragalac-

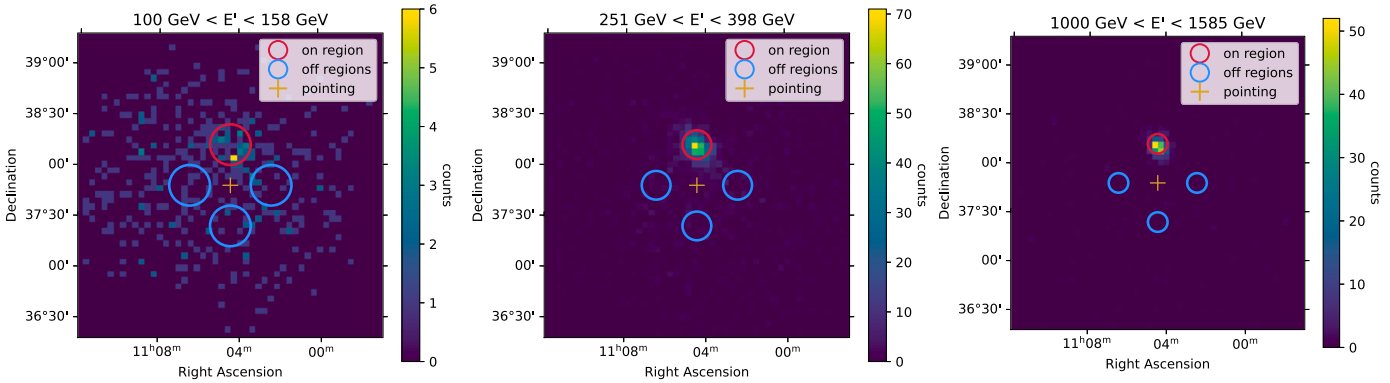


**Fig. 2.** Energy-dependent signal and background estimation. Events from a single run of Mrk421 observations. The three different panels show the distribution of reconstructed coordinates of gamma-ray candidates in three different bins of estimated energy, in order to illustrate how the size of the on and off regions varies with energy.

tic background light (EBL, Cooray 2016), we also compare the EBL absorption treatment in `MARS` and `Gammapy`. QSO B0218+357 is a gravitationally-lensed blazar located at redshift $z = 0.944$ (Cohen et al., 2003). We consider data from July 2014 corresponding to the detection in VHE of the delayed component of a gamma-ray flare first observed at HE by *Fermi*-LAT (Ahnen et al., 2016). The source, displaying a very soft gamma-ray spectrum (with measured spectral index $\sim -4$), was observed for 2 h in the zenith range $[5°, 50°]$ with the standard pointing offsets of 0.4°. IC310 is a radio galaxy located at a redshift $z = 0.0189$ (Bernardi et al., 2002), embedded in the Perseus cluster. We consider the data set corresponding to the November 2012 gamma-ray flare described in Aleksić et al. (2014). This contains the source brightest and hardest gamma-ray emission. IC310 was observed for 5 h in the zenith range $[5°, 35°]$ with two different pointing offsets of 0.4° and 0.94° (due to a pointing strategy optimised to include at once different sources of the Perseus cluster in the MAGIC field of view).

### 3.3. Mrk421

To test the estimation of the gamma-ray flux of a very variable source, we consider Mrk421, one of the brightest blazars observed at all wavelengths (Abdo et al., 2011a). We selected 42 h of observations from 2014, already presented in Acciari et al. (2020), that we use to test the light curve estimation in the case of a very fast (sub-hour time scale) flux variability. The source was observed with 0.4° pointing offset, in the zenith range $[5°, 70°]$.

### 3.4. M15

Another computation that can be performed in a one-dimensional analysis is the estimation of upper limits on the flux of a undetected target. For this analysis case, we select 57 h of observation of M15, a globular cluster whose potential VHE emission was investigated by MAGIC in MAGIC Collaboration (2019). The source was observed with 0.4° pointing offset, in the zenith range $[5°, 50°]$.

## 4. Validation

The process that goes from high-level data, standardised or proprietary, to scientific results (i.e. the last arrow in Fig. 1) typically consists of two parts: a data reduction process that produces binned data from event lists and IRFs[8]; and a statistical analysis, that uses the latter to estimate fluxes. Validating the results of the analysis of standardised data with `Gammapy` against `MARS` therefore implies comparing both data reduction and statistical algorithms. In this section, we provide a brief theoretical background of the algorithms used for reduction and analysis, before moving on to the results of their validation.

### 4.1. Point-like analysis

In the most general case, when analysing gamma-ray observations, one would like to estimate the flux of one or more sources in the field of view (FoV). An analysis accounting for the position and morphology of different sources in the FoV is referred to as spectro-morphological or three-dimensional (two sky coordinates and the energy of the events are the dependencies considered for binning the data and for the emission model to be fitted). In this type of analysis, an estimate of the gamma-ray background (i.e. hadronic showers misidentified as gamma rays) over the whole FoV is required. When considering instead a single source with negligible extension, the background can be directly estimated from the observation itself with aperture photometry techniques (Berge et al., 2007). While the signal is measured from a region referred to as "on", centred on the source, the background is commonly extracted from one or more "off" regions, often symmetric to the on region with respect to the camera centre, as illustrated in Fig. 2. As this analysis integrates out all spatial dependencies and considers only the distributions of events as a function of energy, it is commonly referred to as "point-like" or "one-dimensional" analysis. Moreover, the sizes of the on and off regions can be tuned to reflect the improvement of the angular resolution of the telescopes with energy.[9] As part of this work, this energy-dependent signal and background extraction, illustrated in the different panels of Fig. 2, was implemented in `Gammapy` and made available since `v1.0` (Acero et al., 2021a).

In the one-dimensional analysis, in order to estimate the spectrum of a source, an analytical model is considered to represent the differential flux of the source as a function of true energy

$$\frac{\mathrm{d}\phi}{\mathrm{d}E}(E;\boldsymbol{\theta})\,[\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{TeV}^{-1}], \tag{1}$$

where $\boldsymbol{\theta}$ is the set of parameters specifying the analytical model. They can be estimated through a statistical procedure, e.g. a likelihood maximisation. In order to do so, as the result of the signal/background estimation is a histogram of counts vs (estimated) energy (see e.g. Fig. 3), one needs to translate the analytical flux model into "predicted" source counts. This is done by folding the assumed differential flux model with the IRF. In case of an observation that can be characterised by a single

[8] Unbinned analysis considering the measured quantities of individual events can also be considered, see Sect. 1.2.5 of Malyshev and Mohrmann (2023).

[9] The size of the on and off regions can be computed from MC simulations by selecting, across different energy bins, the radius of the region containing a common fraction of events simulated from a point-like source.

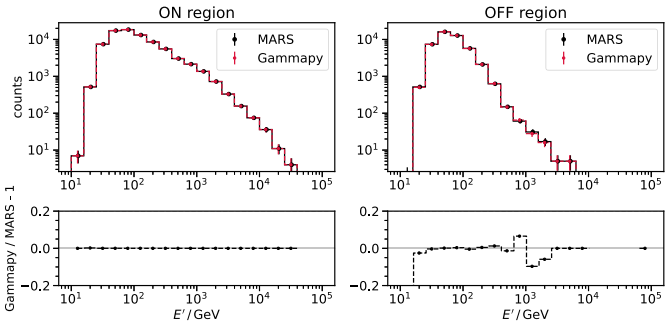


**Fig. 3.** Comparison of the observed number of counts in the on and off regions with the two software using the proprietary (MARS) and the standardised files (Gammapy), respectively.

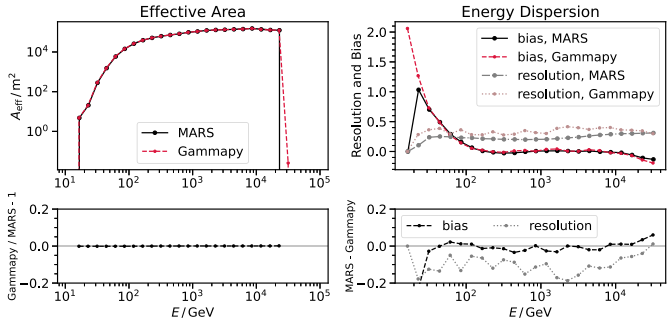


**Fig. 4.** *Left*: comparison of the effective area stored in the proprietary and standardised files. *Right*: comparison of bias and resolution of the energy dispersion (see Section 4.4 in Aleksić et al. 2016 for their definition) stored in the proprietary and standardised files.

IRF (i.e. stable observing conditions) the number of gamma-ray events in the $k$-th bin of estimated energy is given by

$$\begin{aligned} g_k(\boldsymbol{\theta}) &= t_{\mathrm{eff}} \int_{E'_k}^{E'_{k+1}} \mathrm{d}E' \int_0^{\infty} \mathrm{d}E\, \mathrm{IRF}(E'|E)\, \frac{\mathrm{d}\phi}{\mathrm{d}E}(E;\boldsymbol{\theta}) \\ &= t_{\mathrm{eff}} \int_{E'_k}^{E'_{k+1}} \mathrm{d}E' \int_0^{\infty} \mathrm{d}E\, A_{\mathrm{eff}}(E)\, R(E'|E)\, \frac{\mathrm{d}\phi}{\mathrm{d}E}(E;\boldsymbol{\theta}), \end{aligned} \tag{2}$$

where, $t_{\mathrm{eff}}$ is the effective time of the observations and, in the second line, we have factorised the IRF into two components: effective area, $A_{\mathrm{eff}}(E)$, and energy dispersion, or migration matrix, $R(E'|E)$, expressing the probability density function of the energy estimator. We have adopted the convention of expressing estimated (measured) quantities as primed. Observed and predicted counts can be combined in a likelihood assuming a Poissonian distribution of the on and off events in each of the $n_{E'}$ bins in estimated energy considered for the analysis

$$\begin{aligned} &\mathcal{L}(\boldsymbol{\theta}|\{N_{\mathrm{on},k}, N_{\mathrm{off},k}\}_{k=1,\dots,n_{E'}}) = \\ &\prod_{k=1}^{n_{E'}} \mathrm{Pois}(g_k(\boldsymbol{\theta}) + b_k|N_{\mathrm{on},k})\mathrm{Pois}(\alpha\, b_k|N_{\mathrm{off},k}), \end{aligned} \tag{3}$$

where $N_{\mathrm{on},k}$ and $N_{\mathrm{off},k}$ are the number of events observed in the on and off regions, respectively, in the $k$-th estimated energy bin. The likelihood is maximised by varying the values of the model parameters, $\boldsymbol{\theta}$, and consequently the number of predicted source counts in each energy bin, $g_k$; the number of predicted background counts, $b_k$, is instead commonly treated as a nuisance parameter (more details can be found in Appendix A. of Piron et al., 2001). $\alpha$ is a factor taking into account the different exposures of the on and off regions (e.g. in the case of Fig. 2, $\alpha = 3$ as there are three off regions of assumed equal acceptance). To obtain a result from data sets corresponding to different observing conditions (i.e. IRFs) or even different instruments, their likelihood terms as in Eq. (3) can be factored.

### 4.2. Validation of the data reduction

The first step to validate the analysis of standardised data with Gammapy consists in examining the results of the data reduction, that is histograms of the "on" and "off" counts and of the IRF components[10] to be used for statistical analyses (e.g. for the likelihood maximisation previously described). We illustrate, in Fig. 3 and 4 the on and off counts and the IRF components, respectively, obtained with the standardised and proprietary pipelines. Though we illustrate results only for the single-offset Crab Nebula data sample, we observe an almost exact agreement in most of the cases. The difference in the last value of the effective area is an effect of Gammapy's interpolation; differences in the energy dispersion values are due to the slightly different formats adopted by the two pipelines (while a two-dimensional histogram of $E'$ vs $E$ is stored in MARS, a histogram of $E'/E$ vs $E$ is stored in the DL3 files according to the GADF specs).

### 4.3. Validation of the statistical results

Having validated the data reduction, we move on to examine the statistical results that can be obtained with the different scientific and technical cases in our data samples.

#### 4.3.1. Spectrum and light curve of the Crab Nebula observed at different offsets

We use the single- and multi-offset data sets described in Sect. 3.1.1 to estimate the spectrum of the Crab Nebula. We assume a log-parabolic spectral model

$$\frac{\mathrm{d}\phi}{\mathrm{d}E}(E;\phi_0,\alpha,\beta,E_0) = \phi_0 \left(\frac{E}{E_0}\right)^{-\alpha-\beta\log_{10}\left(\frac{E}{E_0}\right)}, \tag{4}$$

fixing the arbitrary reference energy $E_0$ to 1 TeV. We obtain an estimate of the amplitude $\phi_0$, and of the two spectral indices, $\alpha$ and $\beta$, by maximising the likelihood in Eq. (3). In Fig. 5 we show the spectral energy distribution (SED) obtained from the single- and multi-offset samples and provide a detailed comparison of the estimated spectral parameters and their errors in Fig. A.11. We observe an excellent agreement between the results of the standardised and proprietary pipelines. We consider this result as the main focus of our validation of the spectrum estimation, as MARS and Gammapy implement the same likelihood method. As a secondary check, we also compute spectral data points, or "flux points". For this estimation, the methods adopted by the two pipelines differ. Gammapy computes flux points by re-perfoming the likelihood fit considering only the events with energies within the given $k$-th estimated energy bin. Starting from the best-fit model of the broadband fit, only the amplitude $\phi_0$ is re-fitted, thus returning a flux measurement in each energy bin.[11] MARS instead performs an unfolding (Albert et al., 2007) procedure, migrating the excess events in their true energy bins accounting for the energy dispersion. The flux is then directly estimated

[10] IRF components are stored in the DL3 files as tables of values of the response at different offsets from camera centre and at different true energies. These tables are interpolated by Gammapy's data routines (Weiser and Zarantonello, 1988) thus obtaining functions that can be evaluated according to the offset and energy binning chosen for a particular analysis.

[11] We remark that despite the flux points computation is performed considering only the events in a given estimated energy bin, they are represented, in all the plots, as bins in true energy.

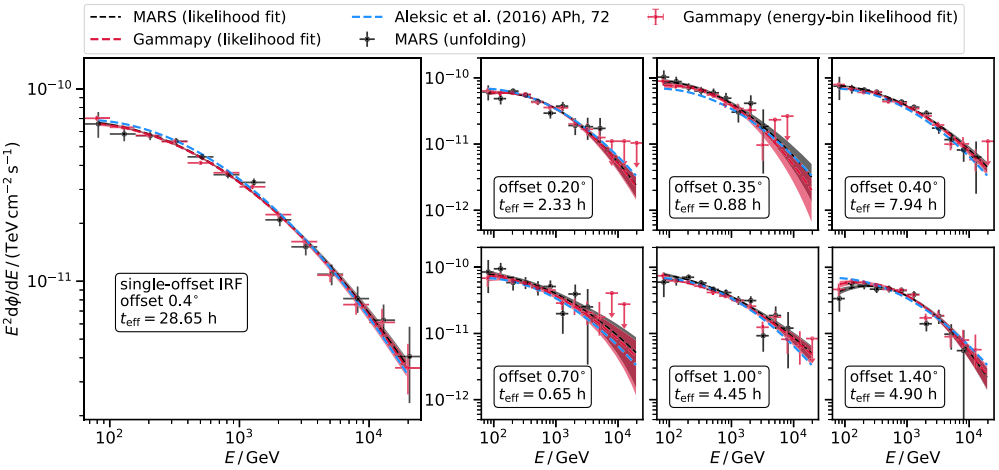


**Fig. 5.** SED of Crab Nebula estimated at different camera offsets and with different MC productions. Dashed lines represent the result of the likelihood fit considering all the events, the bands represent the uncertainty on this results. Points represent a flux estimation considering only the events in a given energy bin, they are computed with different methods by the two software (see main text). The spectrum obtained in Aleksić et al. (2016) from the same data set is added for reference.

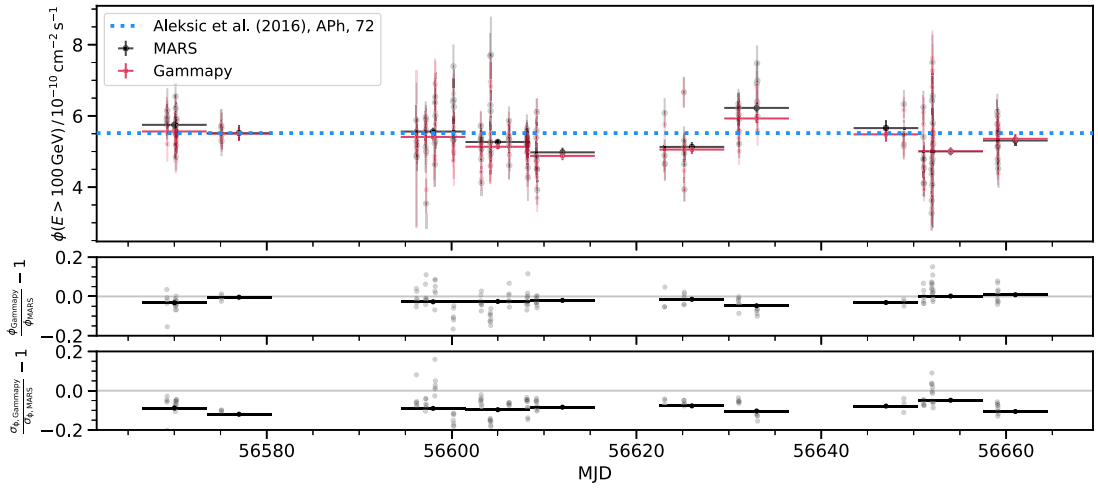


**Fig. 6.** LC of the Crab Nebula obtained with the single- and multi-offset data sets. Transparent points correspond to run-wise binning, solid points to weekly binning. The blue dotted lines represents the measurement obtained by MAGIC in Acciari et al. (2020).

from the migrated excess events. Despite the different methods adopted, we observe a good agreement also in the estimated flux points.

We also employed the single- and multi-offset Crab Nebula data sets to compute the light curve (LC). In this case, the software implement a different estimation of the integral flux $\phi(E > E_{\mathrm{min}}) / [\mathrm{cm}^{-2}\,\mathrm{s}^{-1}]$ in a given time bin. `Gammapy` performs the likelihood maximisation described in Sect. 4.1, considering all the events within a given time bin, and (as for the spectral flux points) re-fitting only the amplitude of the spectrum obtained from the broad-band fit. `MARS` instead computes the LC flux points directly from the number of excess events, without performing a likelihood maximisation (the actual computation is described in detail in Appendix B). The LCs obtained with the two software are presented in Fig. 6, where we show both a run-wise (transparent) and a weekly (solid) time binning. Despite the different methods adopted to estimate the integral flux, we observe a very good agreement between the two pipelines: the estimated flux values display differences between 20% for the run-wise binning and less than 5% for the weekly binning. Uncertainties on the estimated fluxes display deviations within 20%, with those computed by `Gammapy` about 10% smaller than those estimated with `MARS`. The differences observed in the estimation of integral fluxes and their uncertainties are further investigated in Appendix B, where we replicate the LC computation performed by `MARS` with `Gammapy`.

#### 4.3.2. *Spectrum of the Crab Nebula observed under different moonlight conditions*

To validate the `automagic` pipeline, we use the Crab Nebula observation taken under different moon conditions described in Sect. 3.1.2 and compare the spectra estimated using the DL3 data produced both manually and automatically. We classify the data in NSB levels according to the direct current measured in the photomultipliers of the MAGIC I telescope (according to the prescriptions of Table 1 of Ahnen et al., 2017). We consider NSB levels from 1 to 8 (in units of dark NSB level) and group them into four larger bins: NSB 1-2, NSB 2-3, NSB 3-5, and NSB 5-8. Fig. 7 shows the spectra estimated for the four different NSB bins, along with the quantity of data selected per each level, specified through the data set effective time. Both the manual and `automagic` high-level analyses are performed with DL3 data and `Gammapy`. For what concerns the data selection, the manual and automatic procedures select and classify similar amounts of data in each NSB bin, with discrepancies of a few minutes due to different data selection procedures. We observe a very good agreement between the spectra obtained with the two approaches, for all the considered moon levels A detailed comparison of the estimated spectral parameters and their errors is provided in Fig. A.11. We notice that, despite the tuned image cleaning, in the bins with the highest NSB levels (3-5 and 5-8) the spectrum of the Crab Nebula remains slightly underestimated with respect to the reference. This

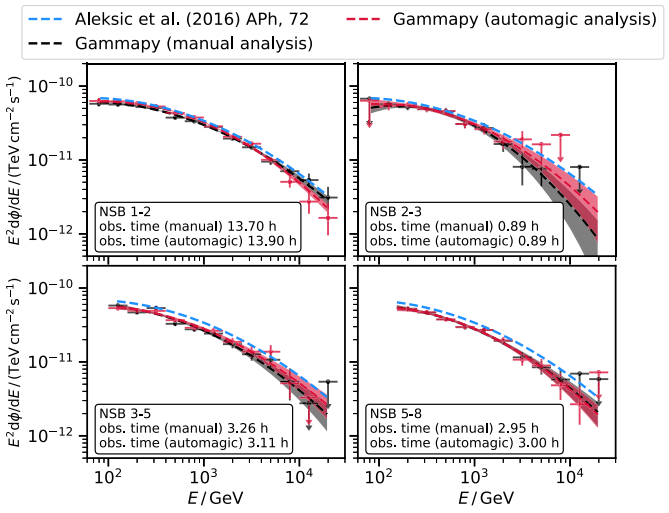


**Fig. 7.** SED of Crab Nebula obtained using a manual and an automated data selection and reduction. The quantity of data selected per each moon level by the two analyses is expressed through the data set effective time. Each NSB level indicates the intensity of the moon illumination.

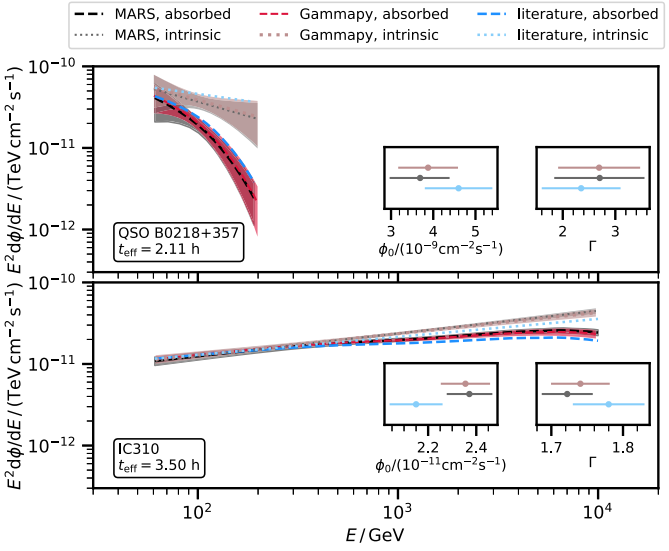


**Fig. 8.** SEDs of QSO B0218+357 (*top*) and IC310 (*bottom*). Lighter colours represent the estimated spectra without the effect of the EBL absorption. The inlets represent the estimated spectral parameters.

behaviour is also observed in the moon performance study (see Fig. 10 of Ahnen et al., 2017, and the systematics evaluation therein).

#### 4.3.3. Estimation of soft and hard gamma-ray spectra, EBL absorption

We estimate the spectra of the jetted AGN QSO B0218+357 and IC310 assuming a power law function

$$\frac{\mathrm{d}\phi}{\mathrm{d}E}(E;\phi_0,\Gamma,E_0)=\phi_0\left(\frac{E}{E_0}\right)^{-\Gamma}, \tag{5}$$

with the reference energy fixed to 100 GeV for QSO B0218+357 and to 1 TeV for IC310. We consider, in both software, the effect of the EBL absorption according to the model of Domínguez et al. (2011). Fig. 8 displays the results of the spectrum estimation for both sources. We observe a very good agreement between the two pipelines for both spectral types and for both distances (and absorption factors) considered.

#### 4.3.4. Light curve of a highly-variable source

We perform a light curve estimation using the Mrk421 data set described in Sect. 3.3. Remarking the different method to estimate the LC described in Sect. 4.3.1, we observe a good agreement between the two software, as illustrated in Fig. 9, with the estimated flux values differing by 20% and the estimated uncertainites showing a similar deviation.

#### 4.3.5. Upper limits on the flux of a non-detected object

As a last example of the estimations commonly performed in a one-dimensional analysis, we compute 95% confidence level upper limits (ULs) on the gamma-ray flux of M15 using both `MARS` and `Gammapy`. By considering a likelihood ratio test based on Eq. (3), `MARS` estimates a confidence level on the number of predicted counts $g$, that is then converted into a flux upper limit. For the likelihood profiling, the Rolke and López (2001) method is applied. `Gammapy` computes instead the upper limit on the flux using the likelihood in Eq. (3) to estimate a one-sided confidence interval on the amplitude parameter, $\phi_0$. As illustrated in Fig. 10 a good agreement between the two methods is found, with deviations below 30%. We remark that typically `MARS` considers a 30% systematic uncertainty on the efficiency of the detected gamma rays (not applied in this particular comparison). The inclusion of this systematic uncertainty results in a $\sim 20\%$ higher flux UL.

## 5. Conclusion

In this paper, we presented the systematic effort to convert observations of the MAGIC telescopes to the standardised GADF data format and to validate their analysis with the open-source software `Gammapy`. We considered six data sets representing different scientific and technical analysis cases. We focus our attention on the one-dimensional analysis and reproduce with the standardised data and `Gammapy` the computations commonly performed in such an analysis: estimation of a spectrum, of a light curve, and of upper limits on the flux in case of a non detection. In all the cases considered, even when different flux estimation methods are adopted, we observe a good agreement, within 20%, between the results of the MAGIC proprietary analysis chain and the standardised data analysed with `Gammapy` (except for upper limits where we observe deviations up to 35%). We demonstrate in Appendix B that when replicating the same flux estimation algorithms of `MARS` in `Gammapy`, differences in the estimated values and uncertainties reduce to below 5%. As the realisation of any future MAGIC data legacy will require the systematic conversion of large amounts of data, we presented in this publication also the implementation of `automagic`, a database-supported pipeline whose purpose is to produce standardised data automatising the process of data selection and reduction in case of different observing periods and conditions. We devoted part of the validation to cross-check the result of the automated pipeline against those of a manual analysis.

Considering future prospects from the technical point of view, the next verification to be performed using the standardised data would be that of the spectro-morphological or three-dimensional analysis. A study of the gamma-ray background of the MAGIC telescopes, mandatory for such an analysis, is already underway (see Mender et al., 2023) and we will present this validation in a follow-up publication.

From the point of view of the community, the validation conducted in this paper makes us confident that the standardised data correctly encapsulate the information in the proprietary one, and that the open-source analysis tools produce results consistent with those obtained with the proprietary software. This process already granted Collaboration members the possibility to adopt standardised data and open-source analysis tools for their analyses. In the hope to extend this possibility to the community in the near future, and to already encourage the exploration of MAGIC data, we make all the Crab Nebula standardised data used in this paper publicly available. This represents the first major release of MAGIC data to the public.

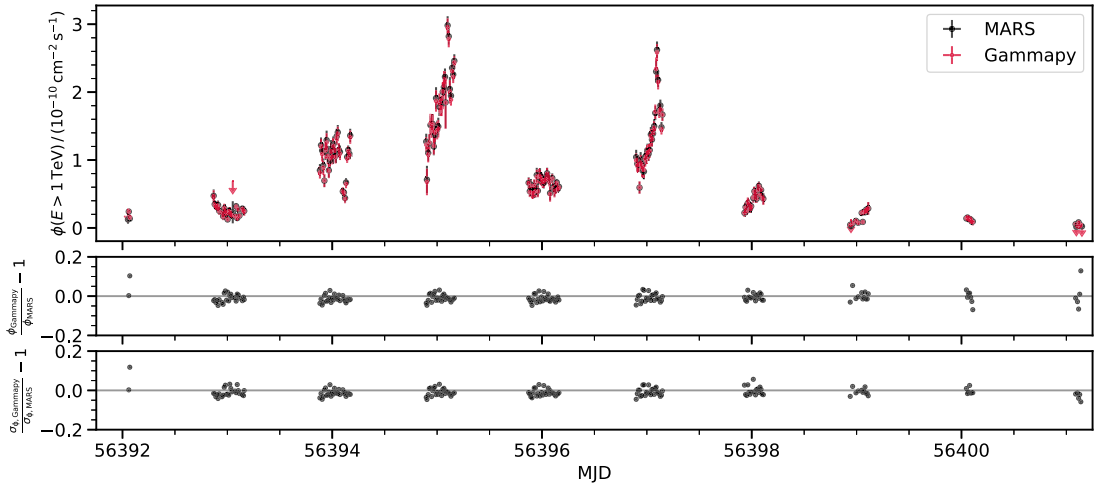


**Fig. 9.** Run-wise LC of Mrk421 during April 2013 computed with `MARS` and `Gammapy`.

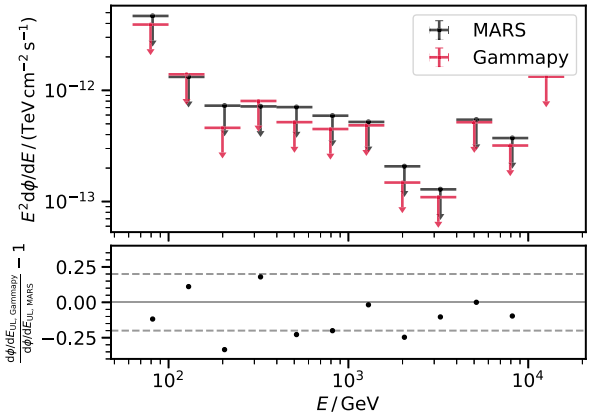


**Fig. 10.** UL on the gamma-ray flux of M15 computed with `MARS` and `Gammapy`.

This work represents the first milestone in the realisation of the MAGIC data legacy: future data releases will follow this publication, inaugurating the public scientific exploitation of two decades of observations. We remark that this scientific exploitation can be conducted in an accessible and fully-reproducible manner thanks to the efforts already devoted by the community to the data standardisation and to the development of open-source analysis tools. The production of VHE gamma-ray data legacies, and the consequent possibility to combine decades of archival gamma-ray data, could deliver a closing statement from this current generation of instruments on several open questions in high-energy astrophysics. This would be of fundamental importance in deciding the most profitable scientific avenues to be pursued with the next-generation of observatories.

## CRediT authorship contribution statement

**Cosimo Nigro**: conceptualisation, methodology, software, validation, formal analysis, resources, data curation, writing - original draft, writing - review & editing, visualisation. **Lena Linhoff**: conceptualisation, software, validation, resources, data curation, writing - original draft. **Simone Mender**: software, validation, resources, data curation. **Jan Lukas Schubert**: software, validation, resources, data curation. **Maximilian Linhoff**: software, validation. **Julian Sitarek**, **Marcel Strzys**, **Ivana Batković**, **Mireia Nievas Rosillo**: resources, data curation. The rest of the authors have contributed in one or several of the following ways: design, construction, maintenance and operation of the instrument(s) used to acquire the data; preparation and/or evaluation of the observation proposals; data acquisition, processing, calibration and/or reduction; production of analysis tools and/or related Monte Carlo simulations; overall discussions about the contents of the draft, as well as related refinements in the descriptions.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

We would like to acknowledge the early MAGIC data standardisation efforts by Tarek Hassan and Léa Jouvin. We would like to thank also Regis Terrier, Axel Donath, and all the `Gammapy` developers for the support they provided during the analysis development and validation.

We would like to thank the Instituto de Astrofísica de Canarias for the excellent working conditions at the Observatorio del Roque de los Muchachos in La Palma. The financial support of the German BMBF, MPG and HGF; the Italian INFN and INAF; the Swiss National Fund SNF; the grants PID2019-104114RB-C31, PID2019-104114RB-C32, PID2019-104114RB-C33, PID2019-105510 GB-C31, PID2019-107847RB-C41, PID2019-107847RB-C42, PID2019-107847RB-C44, PID2019-107988 GB-C22, PID2022-136828NB-C41, PID2022-137810NB-C22, PID2022-138172NB-C41, PID2022-138172NB-C42, PID2022-138172NB-C43, PID2022-139117NB-C41, PID2022-139117NB-C42, PID2022-139117NB -C43, PID2022-139117NB-C44 funded by the Spanish MCIN/AEI/10. 13039/501100011033 and “ERDF A way of making Europe”; the Indian Department of Atomic Energy; the Japanese ICRR, the University of Tokyo, JSPS, and MEXT; the Bulgarian Ministry of Education and Science, National RI Roadmap Project DO1-400/18.12.2020 and the Academy of Finland grant nr. 320045 is gratefully acknowledged. This work was also been supported by Centros de Excelencia “Severo Ochoa” y Unidades “María de Maeztu” program of the Spanish MCIN/AEI/10.13039/501100011033 (CEX2019-000920-S, CEX2019-000918-M, CEX2021-001131-S) and by the CERCA institution and grants 2021SGR00426 and 2021SGR00773 of the Generalitat de Catalunya; by the Croatian Science Foundation (HrZZ) Project IP-2022-10-4595 and the University of Rijeka Project uniri-prirod-18-48; by the Deutsche Forschungsgemeinschaft (SFB1491) and by the Lamarr-Institute for Machine Learning and Artificial Intelligence; by the Polish Ministry of Education and Science grant No. 2021/WK/08; and by the Brazilian MCTIC, CNPq and FAPERJ.

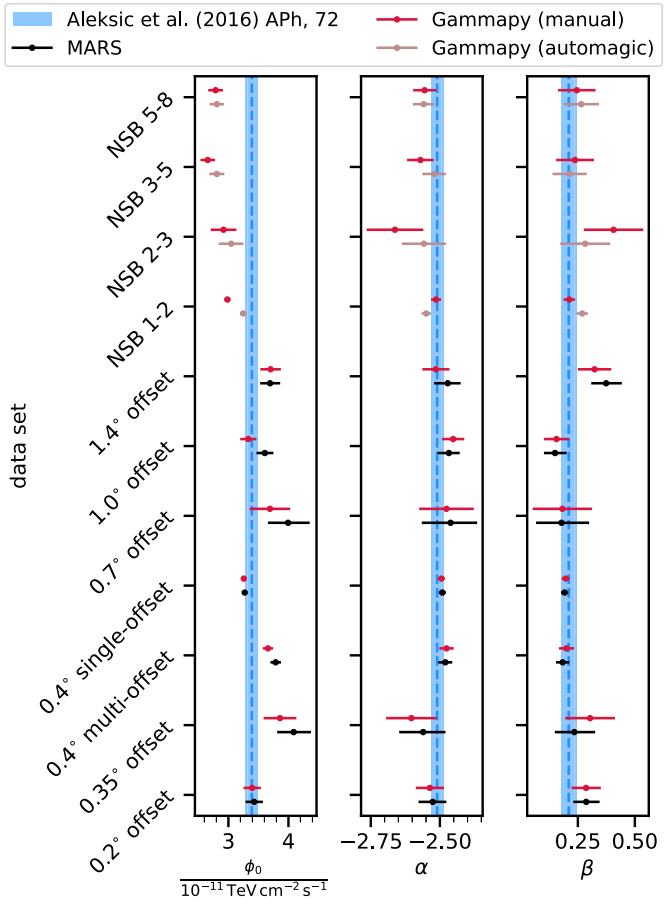


**Fig. A.11.** Log-parabola (Eq. (4)) parameters obtained from the likelihood fit performed with both software for all the Crab Nebula data sets in this paper. The values obtained in Aleksić et al. (2016) from the same data set are represented as a blue band for comparison.

## Appendix A. Comparison of likelihood results for the Crab Nebula data

In Fig. A.11, we illustrate the parameters obtained fitting the log-parabola spectrum to all Crab Nebula data in our validation sample, along with their errors. We also draw, for comparison, the parameters obtained in the performance paper (Aleksić et al., 2016) using MARS and considering a sub sample of the single-offset data set.

## Appendix B. Detailed comparison of the integral flux computation with both software

In order to investigate the systematic 10% discrepancy observed between the uncertainties on the integral flux estimated by Gammapy and MARS (see Fig. 6), we replicate the integral flux computation performed by MARS using the Gammapy routines. Let $E_{\mathrm{min}}$ be the energy above which we would like to estimate the integral flux in a given time bin. The latter is computed by MARS as

$$\Phi(E > E_{\mathrm{min}}) = \frac{N_{\mathrm{ex}}(E' > E_{\mathrm{min}})}{\tilde{A}_{\mathrm{eff}}(E' > E_{\mathrm{min}})\, t_{\mathrm{eff}}}, \tag{B.1}$$

where $N_{\mathrm{ex}}(E' > E'_{\mathrm{min}})$ is the total number of excess events ($N_{\mathrm{on}} - \alpha N_{\mathrm{off}}$, see Sec. 4.1) with estimated energy above $E_{\mathrm{min}}$, $t_{\mathrm{eff}}$ is the effective time, and $\tilde{A}_{\mathrm{eff}}(E' > E_{\mathrm{min}})$ is an effective area modified to take into account the energy dispersion. It is computed as

$$\tilde{A}_{\mathrm{eff}}(E' > E_{\mathrm{min}}) = \frac{N_{\gamma,\mathrm{after\,cuts}}(E' > E_{\mathrm{min}})}{N_{\gamma,\mathrm{simulated}}(E > E_{\mathrm{min}})} A_{\mathrm{simulated}}, \tag{B.2}$$

where in the numerator we have the total number of simulated events with estimated energy above $E_{\mathrm{min}}$ that survive the analysis cuts and in the denominator we have the total number of simulated events with true energy above $E_{\mathrm{min}}$. $A_{\mathrm{simulated}}$ represents the total simulated area. Obtaining the quantities in Eq. (B.2) would require having access to the information relative to the individual simulated events, which is no longer available at the reduction level of the DL3 data. We therefore make use of the IRF components stored in the DL3 files to obtain such an effective area. Given an assumed spectral model, $\frac{\mathrm{d}\phi}{\mathrm{d}E}$ that we take to be the best-fit model obtained from the likelihood maximisation (Eq. (3)), we compute the effective area above a certain estimated energy as

$$\tilde{A}_{\mathrm{eff}}(E' > E_{\mathrm{min}}) = \frac{\frac{\mathrm{d}N}{\mathrm{d}t}(E' > E_{\mathrm{min}})}{\int\limits_{E > E_{\mathrm{min}}} \frac{\mathrm{d}\phi}{\mathrm{d}E}\mathrm{d}E}, \tag{B.3}$$

where $\frac{\mathrm{d}N}{\mathrm{d}t}(E' > E_{\mathrm{min}})$ is the rate of gamma rays with estimated energies $E' > E_{\mathrm{min}}$. The rate in a given true energy bin can be obtained as

$$\frac{\mathrm{d}N}{\mathrm{d}t}(E_i) = \int\limits_{\Delta E_i} A_{\mathrm{eff}}(E)\frac{\mathrm{d}\phi}{\mathrm{d}E}\mathrm{d}E \tag{B.4}$$

where $A_{\mathrm{eff}}(E)$ is the effective area as a function of true energy provided in the GADF-compliant IRF components and $i$ is an index running over the true energy bins, while $\Delta E_i$ the true energy bin width. To obtain the rate above an estimated energy, we multiply Eq. (B.4) with the migration matrix $R(E'|E)$

$$\frac{\mathrm{d}N}{\mathrm{d}t}(E' > E_{\mathrm{min}}) = \sum_{j\,:\,E' > E_{\mathrm{min}}} \frac{\mathrm{d}N}{\mathrm{d}t}(E_i) R_{ij}, \tag{B.5}$$

where now $j$ is an index that runs over the estimated energy bins and we have represented the energy migration as an $i \times j$ matrix. Using the Gammapy routines, we can compute the effective area in Eq. (B.3) from the IRF components in the DL3 files ($A_{\mathrm{eff}}(E)$ and $R(E'|E)$). The integral flux is then obtained, according to Eq. (B.1), by considering the total number of excess events counted in an observation and its effective time. For the uncertainty on the flux we neglect any uncertainty on the effective time and effective area and consider only the uncertainty on the number of excess events, obtained as $\sigma(N_{\mathrm{ex}}) = \sqrt{N_{\mathrm{on}} + \alpha^2 N_{\mathrm{off}}}$.

In the left panels of Fig. B.12 we illustrate the factors in Eq. (B.1) obtained with both software for two consecutive nights in the single-offset Crab Nebula data set. In the bottom panel, we illustrate the fluxes obtained by MARS and by Gammapy, with the latter using both Eq. (B.1) and the result of the likelihood maximisation. In the right panels of the figure, we illustrate the deviations of the flux values and uncertainties estimated by Gammapy from those estimated by MARS, considering the whole Crab Nebula single-offset data set. As one can see, when adopting the same method as MARS, flux values and uncertainties estimated with Gammapy differ by less than 5% from those estimated by MARS. We are thus confident that the 10% underestimation of the integral flux uncertainty observed when using gammapy's default likelihood routine is due to the different approaches adopted for the estimation and not to any issue with the standardised DL3 data.

## Data availability

As the purpose of this paper is not only to present the standardisation effort, but also to initiate the dissemination of the MAGIC standardised data, we make all the Crab Nebula DL3 data used in this article publicly available both via the MAGIC data centre at PIC (https://opendata.magic.pic.es/, under "MAGIC DL3 Public Data Release 1") and via zenodo (MAGIC Collaboration, 2024) (https://zenodo.org/records/11108473). Differently than previous releases of DL3 data (e.g. Nigro et al., 2019), we release the Crab Nebula data sets with a permissive license

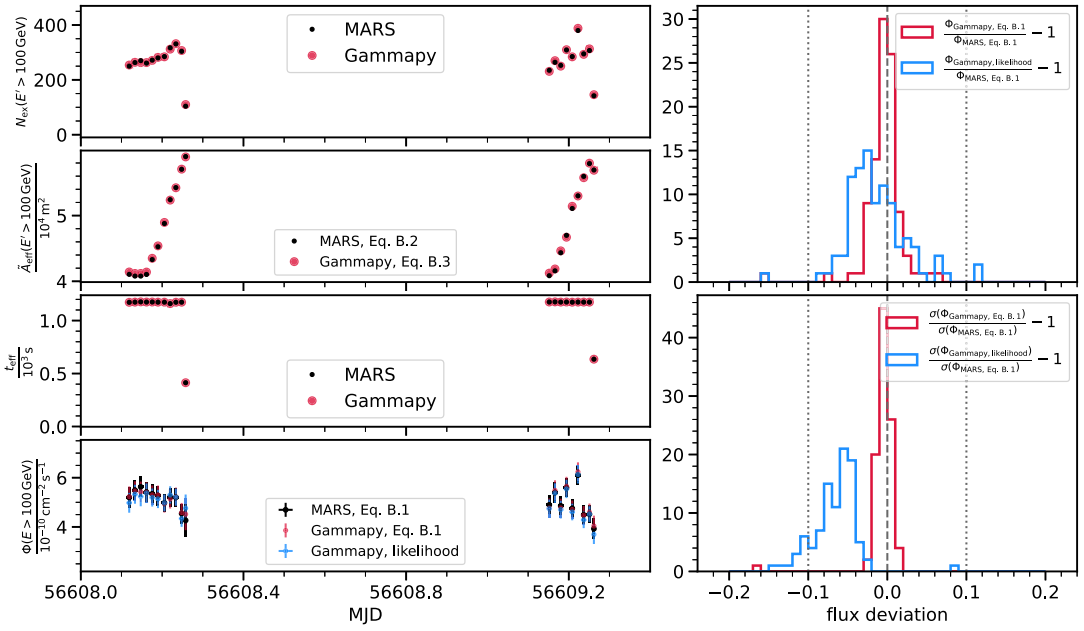


**Fig. B.12.** Replicating the `MARS` integral flux estimation with `Gammapy`. In the left panels, we compare the factors used for the integral flux estimation and the fluxes estimated with both software. For `Gammapy`, we consider both a re-implementation of `MARS` flux estimation (red) as well as its default likelihood-based estimation (blue). For simplicity, only two consecutive nights of the single-offset Crab Nebula data set are shown. In the right panels, we illustrate the deviations of the flux values and uncertainties estimated by `Gammapy` with both method (MARS-like and likelihood-based) from those estimated by `MARS` considering the whole Crab Nebula single-offset data set.

(Creative Commons Attribution, CC BY, https://creativecommons.org/licenses/by/4.0/) that allows for their usage for scientific investigations.

## References


Abdo, A.A., et al., 2011a. Fermi large area telescope observations of Markarian 421: the missing piece of its spectral energy distribution. Astrophys. J. 736, 131. https://doi.org/10.1088/0004-637X/736/2/131. arXiv:1106.1348.

Abdo, A.A., et al., 2011b. Gamma-ray flares from the Crab Nebula. Science 331, 739. https://doi.org/10.1126/science.1199705. arXiv:1011.3855.

Acciari, V.A., et al., 2020. Unraveling the complex behavior of Mrk 421 with simultaneous X-ray and VHE observations during an extreme flaring activity in 2013 April. Astrophys. J. Suppl. Ser. 248, 29. https://doi.org/10.3847/1538-4365/ab89b5. arXiv:2001.08678.

Acero, F., et al., 2021a. `gammapy`: v0.2.0. Zenodo. https://doi.org/10.5281/zenodo.6552377. https://zenodo.org/record/6552377.

Acero, F., et al., 2021b. `gammapy`: v1.1. Zenodo. https://doi.org/10.5281/zenodo.8033275. https://doi.org/10.5281/zenodo.7734804.

Ahnen, M.L., et al., 2016. Detection of very high energy gamma-ray emission from the gravitationally lensed blazar QSO B0218+357 with the MAGIC telescopes. Astron. Astrophys. 595, A98. https://doi.org/10.1051/0004-6361/201629461. arXiv:1609.01095.

Ahnen, M.L., et al., 2017. Performance of the MAGIC telescopes under moonlight. Astropart. Phys. 94, 29–41. https://doi.org/10.1016/j.astropartphys.2017.08.001. arXiv:1704.00906.

Albert, A., et al., 2022. Validation of standardized data formats and tools for ground-level particle-based gamma-ray observatories. Astron. Astrophys. 667, A36. https://doi.org/10.1051/0004-6361/202243527. arXiv:2203.05937.

Albert, J., et al., 2007. Unfolding of differential energy spectra in the magic experiment. Nucl. Instrum. Methods Phys. Res. A 583, 494–506. https://doi.org/10.1016/j.nima.2007.09.048. arXiv:0707.2453.

Albert, J., et al., 2008. Implementation of the random forest method for the imaging atmospheric Cherenkov telescope MAGIC. Nucl. Instrum. Methods Phys. Res. A 588, 424–432. https://doi.org/10.1016/j.nima.2007.11.068. arXiv:0709.3719.

Aleksić, J., et al., 2014. Black hole lightning due to particle acceleration at subhorizon scales. Science 346, 1080–1084. https://doi.org/10.1126/science.1256183. arXiv:1412.4936.

Aleksić, J., et al., 2016. The major upgrade of the MAGIC telescopes, Part II: a performance study using observations of the Crab Nebula. Astropart. Phys. 72, 76–94. https://doi.org/10.1016/j.astropartphys.2015.02.005. arXiv:1409.5594.

Band, D.L., 2007. Gamma-ray large area space telescope (GLAST) project data management plan. available online at https://fermi.gsfc.nasa.gov/ssc/data/policy/PDMP.pdf.

Barrett, P., 1995. The Compton gamma-ray observatory archive. In: Egret, D., Albrecht, M.A. (Eds.), Information & on-Line Data in Astronomy, p. 15.

Berge, D., Funk, S., Hinton, J., 2007. Background modelling in very-high-energy $\gamma$-ray astronomy. Astron. Astrophys. 466, 1219–1229. https://doi.org/10.1051/0004-6361:20066674. arXiv:astro-ph/0610959.

Bernardi, M., et al., 2002. Redshift-distance survey of early-type galaxies. I. The ENEARc cluster sample. Astron. J. 123, 2990–3017. https://doi.org/10.1086/340463. arXiv:astro-ph/0203022.

Brun, R., Rademakers, F., 1997. ROOT — an object oriented data analysis framework. Nucl. Instrum. Methods Phys. Res. A 389, 81–86. https://doi.org/10.1016/S0168-9002(97)00048-X.

Bühler, R., Blandford, R., 2014. The surprising Crab pulsar and its nebula: a review. Rep. Prog. Phys. 77, 066901. https://doi.org/10.1088/0034-4885/77/6/066901. arXiv:1309.7046.

Cohen, J.G., Lawrence, C.R., Blandford, R.D., 2003. The redshift of the lensed object in the Einstein ring B0218+357. Astrophys. J. 583, 67–69. https://doi.org/10.1086/344837. arXiv:astro-ph/0209457.

Contreras, J.L., et al., 2015. Data model issues in the Cherenkov telescope array project. In: 34th International Cosmic Ray Conference (ICRC2015), p. 960. arXiv:1508.07584.

Cooray, A., 2016. Extragalactic background light measurements and applications. R. Soc. Open Sci. 3, 150555. https://doi.org/10.1098/rsos.150555. arXiv:1602.03512.

de Naurois, M., Mazin, D., 2015. Ground-based detectors in very-high-energy gamma-ray astronomy. C. R. Phys. 16, 610–627. https://doi.org/10.1016/j.crhy.2015.08.011. arXiv:1511.00463.

Deil, C., et al., 2017. Open high-level data formats and software for gamma-ray astronomy. In: 6th International Symposium on High Energy Gamma-Ray Astronomy, 070006. arXiv:1610.01884.

Domínguez, A., et al., 2011. Extragalactic background light inferred from AEGIS galaxy-SED-type fractions. Mon. Not. R. Astron. Soc. 410, 2556–2578. https://doi.org/10.1111/j.1365-2966.2010.17631.x. arXiv:1007.1459.

Donath, A., et al., 2023. Gammapy: a python package for gamma-ray astronomy. Astron. Astrophys. 678, A157. https://doi.org/10.1051/0004-6361/202346488. arXiv:2308.13584.

Fomin, V.P., et al., 1994. New methods of atmospheric Cherenkov imaging for gamma-ray astronomy. I. The false source method. Astropart. Phys. 2, 137–150. https://doi.org/10.1016/0927-6505(94)90036-1.

H. E. S. S. Collaboration, 2018. H.E.S.S. first public test data release. arXiv e-prints. arXiv:1810.04516.

Hillas, A.M., 2013. Evolution of ground-based gamma-ray astronomy from the early days to the Cherenkov telescope arrays. Astropart. Phys. 43, 19–43. https://doi.org/10.1016/j.astropartphys.2012.06.002.

Knödlseder, J., et al., 2016. GammaLib and ctools. A software framework for the analysis of astronomical gamma-ray data. Astron. Astrophys. 593, A1. https://doi.org/10.1051/0004-6361/201628822. arXiv:1606.00393.

Knödlseder, J., et al., 2019. Analysis of the H.E.S.S. public data release with ctools. Astron. Astrophys. 632, A102. https://doi.org/10.1051/0004-6361/201936010. arXiv:1910.09456.
Lamanna, G., et al., 2015. Cherenkov telescope array data management. In: 34th International Cosmic Ray Conference (ICRC2015), p. 947. arXiv:1509.01012.
Lopez-Coto, R., et al., 2024. cta-observatory/cta-lstchain. Zenodo. https://doi.org/10.5281/zenodo.6344673.
MAGIC Collaboration, 2019. Deep observations of the globular cluster M15 with the MAGIC telescopes. Mon. Not. R. Astron. Soc. 484, 2876–2885. https://doi.org/10.1093/mnras/stz179. arXiv:1901.04367.
MAGIC Collaboration, 2024. MAGIC data level 3 (DL3) public data release 1 (PDR1). Zenodo. https://doi.org/10.5281/zenodo.11108473.
Maier, G., Holder, J., 2017. Eventdisplay: an analysis and reconstruction package for ground-based gamma-ray astronomy. In: 35th International Cosmic Ray Conference (ICRC2017), p. 747. arXiv:1708.04048.
Maier, G., et al., 2024. Eventdisplay: an analysis and reconstruction package for VERITAS. Zenodo. https://doi.org/10.5281/zenodo.11096726.
Malyshev, D., Mohrmann, L., 2023. Analysis methods for gamma-ray astronomy. In: Handbook of X-Ray and Gamma-Ray Astrophysics, p. 137.
Mender, S., et al., 2023. Computing sky maps using the open-source package Gammapy and MAGIC data in a standardized format. arXiv e-prints. arXiv:2302.11285. https://doi.org/10.48550/arXiv.2302.11285.
Mohrmann, L., et al., 2019. Validation of open-source science tools and background model construction in $\gamma$-ray astronomy. Astron. Astrophys. 632, A72. https://doi.org/10.1051/0004-6361/201936452. arXiv:1910.08088.
Moralejo, A., et al., 2009. MARS, the MAGIC analysis and reconstruction software. arXiv e-prints. arXiv:0907.0943. arXiv:0907.0943.
Nigro, C., Hassan, T., Olivera-Nieto, L., 2021. Evolution of data formats in very-high-energy gamma-ray astronomy. Universe 7, 374. https://doi.org/10.3390/universe7100374. arXiv:2109.14661.
Nigro, C., et al., 2019. Towards open and reproducible multi-instrument analysis in gamma-ray astronomy. Astron. Astrophys. 625, A10. https://doi.org/10.1051/0004-6361/201834938. arXiv:1903.06621.
Pence, W., 1999. CFITSIO, v2.0: a new full-featured data interface. In: Mehringer, D.M., Plante, R.L., Roberts, D.A. (Eds.), Astronomical Data Analysis Software and Systems VIII, p. 487.
Piron, F., et al., 2001. Temporal and spectral gamma-ray properties of <ASTROBJ>Mkn 421</ASTROBJ> above 250 GeV from CAT observations between 1996 and 2000. Astron. Astrophys. 374, 895–906. https://doi.org/10.1051/0004-6361:20010798. arXiv:astro-ph/0106196.
Pittori, C., 2013. The AGILE data center at ASDC. In: American Astronomical Society Meeting Abstracts, vol. 221, 439.03.
Rolke, W.A., López, A.M., 2001. Confidence intervals and upper bounds for small signals in the presence of background noise. Nucl. Instrum. Methods Phys. Res. A 458, 745–758. https://doi.org/10.1016/S0168-9002(00)00935-9. arXiv:hep-ph/0005187.
Tavani, M., et al., 2011. Discovery of powerful gamma-ray flares from the Crab Nebula. Science 331, 736. https://doi.org/10.1126/science.1200083. arXiv:1101.2311.
Weekes, T.C., et al., 1989. Observation of TeV gamma rays from the Crab Nebula using the atmospheric Cerenkov imaging technique. Astrophys. J. 342, 379. https://doi.org/10.1086/167599.
Weiser, A., Zarantonello, S.E., 1988. A note on piecewise linear and multilinear table interpolation in many dimensions. Math. Comput. 50, 189–196. https://doi.org/10.1090/S0025-5718-1988-0917826-0. https://www.ams.org/journals/mcom/1988-50-181/S0025-5718-1988-0917826-0/S0025-5718-1988-0917826-0.pdf.
Wells, D.C., Greisen, E.W., Harten, R.H., 1981. FITS - a flexible image transport system. Astron. Astrophys. Supplement 44, 363.
Zanin, R., et al., 2013. MARS, the MAGIC analysis and reconstruction software. In: International Cosmic Ray Conference, p. 2937.