Validation of open-source science tools and background model construction in γ -ray astronomy

L. Mohrmann, A. Specovius, D. Tiziani, S. Funk, D. Malyshev, K. Nakashima, and C. van Eldik

Erlangen Centre for Astroparticle Physics, University of Erlangen-Nuremberg, Erwin-Rommel-Str. 1, 91058 Erlangen, Germany
e-mail: lars.mohrmann@fau.de

Received 3 August 2019 / Accepted 10 October 2019

ABSTRACT

In classical analyses of γ -ray data from imaging atmospheric Cherenkov telescopes (IACTs), such as the High Energy Stereoscopic System (H.E.S.S.), aperture photometry, or photon counting, is applied in a (typically circular) region of interest (RoI) encompassing the source. A key element in the analysis is to estimate the amount of background in the RoI due to residual cosmic ray-induced air showers in the data. Various standard background estimation techniques have been developed in the last decades, most of them rely on a measurement of the background from source-free regions within the observed field of view. However, in particular in the Galactic plane, source analysis and background estimation are hampered by the large number of, sometimes overlapping, γ -ray sources and large-scale diffuse γ -ray emission. For complicated fields of view, a three-dimensional (3D) likelihood analysis shows the potential to be superior to classical analysis. In this analysis technique, a spectromorphological model, consisting of one or multiple source components and a background component, is fitted to the data, resulting in a complete spectral and spatial description of the field of view. For the application to IACT data, the major challenge of such an approach is the construction of a robust background model. In this work, we apply the 3D likelihood analysis to various test data recently made public by the H.E.S.S. collaboration, using the open analysis frameworks *ctools* and *Gammapy*. First, we show that, when using these tools in a classical analysis approach and comparing to the proprietary H.E.S.S. analysis framework, virtually identical high-level analysis results, such as field-of-view maps and spectra, are obtained. We then describe the construction of a generic background model from data of H.E.S.S. observations, and demonstrate that a 3D likelihood analysis using this background model yields high-level analysis results that are highly compatible with those obtained from the classical analyses. This validation of the 3D likelihood analysis approach on experimental data is an important step towards using this method for IACT data analysis, and in particular for the analysis of data from the upcoming Cherenkov Telescope Array (CTA).

Key words. methods: data analysis – gamma rays: general

1. Introduction

With the advent of the Cherenkov Telescope Array (CTA, see [Acharya et al. 2018](#)), the field of ground-based γ -ray astronomy is undergoing a major transformation. This is not only because CTA, due to its greatly improved sensitivity with respect to current instruments, will offer a great discovery potential, but also because it will be operated as an open observatory. Among many implications, this entails that an open software package will need to be provided for the high-level analysis of CTA data. There are currently two open-source packages proposed as prototypes for this software package: *ctools*¹ ([Knödlseder et al. 2016](#)) and *Gammapy*² ([Deil et al. 2017](#)).

In this paper, we apply both of these tools³ to the analysis of data from the High Energy Stereoscopic System (H.E.S.S.), one of the current-generation arrays of imaging atmospheric Cherenkov telescopes (IACTs). The motivation for this is twofold. Firstly, we want to support the development of the tools. While they have been validated in a CTA-internal data challenge based on simulated data, they have not been extensively tested on

experimental data so far. Secondly, both *ctools* and *Gammapy* – despite still being in development – already offer analysis techniques that are beyond the capabilities of the standard software packages used for high-level data analysis within the H.E.S.S. Collaboration. In particular, they allow us to perform a three-dimensional (3D), energy-resolved likelihood analysis, which is a major focus of this paper.

In a 3D likelihood analysis, the observed data are described by a combination of spectromorphological (i.e. three-dimensional) models, one for each relevant component in the observed field of view. The models are fitted to the data via a likelihood formalism; the significance of specific components can be determined by means of likelihood ratio tests. This kind of analysis is useful in cases where a “complicated” field of view, with multiple sources or large-scale diffuse emission, prevents standard analysis techniques from working well as they typically rely on a measurement of the residual cosmic-ray background within the observed field of view. One of the major challenges in this approach is the development of an accurate model template for the cosmic-ray background, which strongly depends on the observation conditions. In this paper, we attempt to construct such a model from archival H.E.S.S. observations.

The concept of the 3D likelihood analysis is not new. It was already applied in the analysis of data from the Energetic Gamma Ray Experiment Telescope (EGRET; [Mattox et al. 1996](#)).

¹ <http://cta.irap.omp.eu/ctools>

² <https://gammapy.org>

³ We used the most recent versions of the packages available at the time of writing: *ctools* 1.6.0 and *Gammapy* 0.12.

3D likelihood analysis is also routinely used in the analysis of data from the *Fermi*-LAT satellite (see e.g. [Ackermann et al. 2017](#)), for which changes and uncertainties in observation conditions are much less pronounced, and no strong residual cosmic-ray background is present. Various forms of model-based analyses have also already been pioneered in IACT data analysis. The construction of a background model template from cosmic ray-like events in the observed field of view is described in [Rowell \(2003\)](#) and [Fernandes et al. \(2014\)](#). [Abramowski et al. \(2012\)](#) obtain an estimate for the cosmic-ray background by pairing each observation with one “Off” observation that is free of γ -ray sources and has been taken under similar observation conditions – this approach differs from the one presented in this paper in as much as the background is estimated from a single observation rather than many. A morphological, energy-integrated likelihood analysis with multiple components is presented in [Abdalla et al. \(2018a,b\)](#). Finally, first fully spectromorphological likelihood analyses of H.E.S.S. data have been carried out by [Mayer \(2014\)](#), [Devin \(2018\)](#), and [Ziegler \(2018\)](#).

The outline of the paper is as follows: Sect. 2 gives an overview about H.E.S.S. and the data set utilised in this paper. In Sect. 3, we perform a validation of *ctools* and *Gammapy*. This is achieved by performing several analyses of H.E.S.S. data using standard IACT data analysis techniques, in particular standard techniques that treat the residual background of cosmic ray-induced air showers that are always present in the data. We compare the results of the open-source tools with those obtained with the H.E.S.S. analysis package (HAP), one of the proprietary H.E.S.S. software packages used for data analysis. In Sect. 4, we then introduce a novel method to construct a template model for the residual cosmic-ray background, a key prerequisite for the 3D likelihood analysis technique. We then used *ctools* and *Gammapy* to apply this background model in a 3D likelihood analysis, as presented in Sect. 5, thereby also validating this analysis approach and exploring its capabilities. Finally, we conclude the paper with Sect. 6.

In several parts of this paper, we show results obtained with either one or both of the open-source science tools. We stress that in every case, our intention is to demonstrate that both tools work well and yield results that are compatible with the H.E.S.S. software package HAP, not to perform a comparison between the two.

We release the background model templates that we derived for the data analysed here as supplementary material to this paper, see Appendix D for more information. Furthermore, we make available the results of all spectral fits carried out for this paper in machine-readable format (see Appendix E).

2. The High Energy Stereoscopic System and the H.E.S.S. public test data release

H.E.S.S. ([Aharonian et al. 2006a](#)) is an array of five IACTs, located in the Khomas highland in Namibia ($23^{\circ}16'18''$ S, $16^{\circ}30'00''$ E), at an elevation of 1800 m above sea level. In its first phase (“H.E.S.S. Phase-I”), lasting from 2004 until 2012, the array consisted of four telescopes (CT1-4) with 107 m^2 mirror area each, arranged in a square formation with 120 m side length. In this configuration, H.E.S.S. was able to detect γ rays with energies from $\sim 200\text{ GeV}$ (for observations close to zenith) up to several tens of TeV.

The array was enhanced in 2012 by the addition of a fifth telescope (CT5) with mirror area 612 m^2 in the centre of the array, thus reducing the energy threshold of the instrument to below 100 GeV ([Holler et al. 2015](#)). Data from this second phase

of the experiment (“H.E.S.S. Phase-II”) are not analysed in this paper, however, all presented concepts can in principle also be applied to them.

H.E.S.S. records data in time intervals of usually 28 min, called “observations” or “runs”. The entire H.E.S.S. Phase-I data set consists of 17 712 observations fulfilling basic quality criteria that check for hardware failures (referred to as “detection” criteria, see [Aharonian et al. 2006a](#)), amounting to $\approx 8050\text{ h}$ of observation time. Of these, 15 042 observations (6878 h) additionally pass a stricter set of quality criteria that ensures stable atmospheric conditions (“spectral” criteria). About half of these observations are used in Sect. 4 to construct a background model template.

We analyse data from the first H.E.S.S. public test data release ([Abdalla et al. 2018c](#)) in this paper. Table 1 lists the data sets contained in this release. Here, we use the data sets taken on the Crab nebula, PKS 2155–304 (steady), MSH 15–52, and RXJ1713.7–3946. For each observation, the data consist of a list of recorded events with their reconstructed properties and instrument response functions (effective area, point spread function, and energy dispersion) specific to the observation.

The test data release has been published specifically to support the development of open-source tools like *ctools* and *Gammapy*. It contains data taken on both point-like and extended sources, making it a good choice of data set for this paper. The data are available in FITS format⁴, as specified by [Deil et al. \(2018\)](#), and can be directly processed with *ctools* and *Gammapy*. We note that the data have been processed with an analysis configuration that is no longer state-of-the-art, both in terms of event reconstruction (which is based on Hillas parameters; [Hillas 1985](#)) and γ -hadron separation (which uses the “mean scaled width” parameter; [Aharonian et al. 2006a](#)). This is not a problem, considering that the analysed sources are strong γ -ray emitters and that the main purpose of this paper is the validation of the analysis tools.

3. Validation of standard background estimation techniques

In this section, we demonstrate the capability of the open-source analysis tools *ctools* and *Gammapy* to carry out analyses based on background estimation techniques that are traditionally used in ground-based γ -ray astronomy. We restrict ourselves to the two arguably most widely used techniques, namely the “ring background” and the “reflected background” algorithm (for a detailed description of these algorithms, see [Berge et al. 2007](#)). Both techniques apply aperture photometry, that is, they extract the flux of γ rays from a source by determining the number of registered events in a region of interest, called “on region”, and comparing this to an (appropriately scaled) estimate of the residual cosmic-ray background, obtained from one or multiple “off region(s)” within the observed field of view.

Furthermore, we validate the results obtained with the open-source tools by comparing them with results obtained with the H.E.S.S.-internal analysis software package HAP. In all cases, we find the results to be virtually identical.

We present results obtained with the ring background method in Sect. 3.1, those obtained with the reflected background method in Sect. 3.2. In some cases, we only show results obtained with one of the open-source tools, or for a selection of the available data sets, implying that we obtain the same level of agreement with the other tool or the remaining data sets,

⁴ <https://fits.gsfc.nasa.gov>

Table 1. Data sets in the first H.E.S.S. public test data release (Abdalla et al. 2018c).

Designation	Type	Extension	# Runs	Observation time (h)	Zenith angle (deg)	Publications
Crab nebula	PWN	Point-like	4	1.9	45.4–48.6	(1,2)
PKS 2155–304 (flare)	AGN	Point-like	15	7.0	7.2–50.4	(3,4)
PKS 2155–304 (steady)	AGN	Point-like	6	2.8	22.8–36.8	(5,6,7)
MSH 15–52	PWN	Extended	20	9.1	36.1–40.2	(8)
RX J1713.7–3946	SNR	Extended	15	7.0	16.7–26.3	(9,10,11)
Off runs	–	–	45	20.7	2.5–52.7	–

Notes. “Type” refers to the source type; PWN = pulsar wind nebula; AGN = active galactic nucleus; SNR = supernova remnant. “Extension” specifies whether or not the source can be spatially resolved using this data set. “Publications” lists (selected) H.E.S.S. publications about the sources (usually based on larger data sets). The data sets “PKS 2155–304 (flare)” and “Off runs” are not used in this paper.

References. (1) Aharonian et al. (2006a); (2) Holler et al. (2015); (3) Aharonian et al. (2007); (4) Aharonian et al. (2009); (5) Aharonian et al. (2005a); (6) Abramowski et al. (2010); (7) Abdalla et al. (2017); (8) Aharonian et al. (2005b); (9) Aharonian et al. (2004); (10) Aharonian et al. (2006b); (11) Abdalla et al. (2018d).

Table 2. Settings for analyses applying standard background estimation techniques.

Source	RA (deg)	Dec (deg)	θ^2 (deg ²)	$r_{\min}$ (deg)	$r_{\max}$ (deg)
Crab nebula	83.63	22.01	0.0125	0.5	0.7
PKS 2155–304	329.72	–30.23	0.0125	0.5	0.7
MSH 15–52	228.53	–59.16	0.09	0.5	0.7
RX J1713.7–3946	258.39	–39.76	0.36	0.6	0.8

Notes. RA and Dec give the source position in equatorial coordinates (J2000) as used in the analysis. θ^2 denotes the squared radius of the on region containing the source. $r_{\min}$ and $r_{\max}$ are the inner and outer radius used in the ring background algorithm.

respectively. Table 2 lists the settings used in the analyses (common for all tools).

3.1. Ring background method

The ring background method is typically used to visualise the excess of γ rays attributed to a source, either in the form of a one-dimensional “ θ^2 -plot”⁵ or of two-dimensional sky maps. In both cases, the algorithm starts from a binned map of the events registered in the observation. For each pixel, it then determines an estimate of the residual cosmic-ray background for that pixel by summing up the events in all pixels contained in a ring around the pixel with inner radius $r_{\min}$ and outer radius $r_{\max}$ (cf. Table 2). Pixels around known γ -ray emitters need to be excluded in this process.

The background estimate must be corrected for the different exposure of the pixels in the ring with respect to the pixel under consideration. Since the acceptance (i.e. the probability of detecting an event) of the experiment varies across the field of view, a model of the acceptance is required to apply this correction. For the sake of better comparability, we chose to employ the model utilised in the HAP analysis within the analyses carried out with ctools and Gammapy. The model is based on archival H.E.S.S. data, similar to the model introduced in Sect. 4, but less detailed (e.g. a radial symmetry is assumed). We verified that

⁵ The angular distance between the reconstructed direction of an event and the source is usually denoted with θ .

Fig. 1. Distribution of squared angular distance θ between reconstructed event directions and source position for the Crab nebula, obtained with Gammapy. The blue line shows the shape of the point spread function of the instrument for this data set.

we obtain compatible results when using the background model developed in this paper instead.

A potential excess of γ -ray events can then be determined by subtracting the background estimate from the map of registered events. Similarly, the significance of the excess can be computed.

3.1.1. θ^2 distributions

Especially for point-like sources, a θ^2 distribution is often used to display the excess of γ rays from the direction of the source. In Fig. 1 we show such a distribution for the Crab nebula data set, here obtained with Gammapy. The shape of the point spread function (PSF) for this data set, averaged over all energies and assuming the best-fit energy spectrum for this source (see Sect. 3.2), is illustrated as well. As expected for a point-like source⁶, the distribution follows the shape of the PSF, demonstrating that this instrument response function is processed correctly by Gammapy.

⁶ Even though the Crab nebula has been demonstrated to be extended in very-high-energy γ rays recently (Holler et al. 2017), it can be considered point-like for the data set and analysis configuration used here.

Fig. 2. Significance map for pulsar wind nebula MSH 15–52, in equatorial coordinates (J2000). The map in the background has been derived with Gammapy, the coloured lines display 6, 8, 10, and 12 σ confidence level contours for the corresponding map derived with HAP.

Fig. 3. Significance map for supernova remnant RX J1713.7–3946, in equatorial coordinates (J2000). The map in the background has been derived with cttools, the coloured lines display 3 and 5 σ confidence level contours for the corresponding map derived with HAP.

3.1.2. Sky maps

The γ -ray excess may also be visualised in the form of sky maps, in particular in the case of spatially extended sources. Different quantities can be plotted; here we focus on the common case of sky maps denoting the significance of an excess. As is custom, we smoothed statistical fluctuations by convolving the map of registered events with a top-hat kernel of 0.1 deg radius (see e.g. Abdalla et al. 2018b) and computed the significance following Li & Ma (1983).

We show maps for the two extended sources that are part of the H.E.S.S. public test release data set, MSH 15–52 and RX J1713.7–3946, in Figs. 2 and 3, respectively. In both cases, we plot contour lines of the map derived with the HAP software on top of the map derived with one of the open-source tools, observing an extremely good agreement.

The quality of the description of the cosmic-ray background in the field of view can be judged by inspecting the entry distribution of a significance map. Figure 4 displays the distributions for the map shown in Fig. 3. For a perfectly modelled background, the significance distribution of pixels outside source regions approaches that of a Gaussian distribution (shown by an orange line, for comparison). As for the maps themselves, we observe a very good agreement between the tools.

3.2. Reflected background method

The reflected background method is usually employed to determine the flux of γ rays from a given source, in other words, to extract its spectrum. It requires the observations to be carried out in so-called “wobble mode”, meaning that the source is observed under a small offset with respect to the pointing direction of the telescopes. This allows the definition of off regions that are “reflected” about the pointing direction, meaning that they have the same offset to the pointing direction as the on region encompassing the source. The acceptance can then be assumed to be approximately the same in all regions, leading to reduced systematic uncertainties in the background determination. The extraction of the spectrum then proceeds by determining the

Fig. 4. Significance distribution for RX J1713.7–3946. The histograms are entry distributions of the significance map shown in Fig. 3, obtained with HAP (grey filled histograms) and cttools (green histograms). The distributions are shown for pixels outside source regions (dashed green, dark grey) and for all pixels (solid green, light grey). The orange line shows a Gaussian distribution with zero mean and unity width.

excess of γ -ray events in the on region with respect to the off regions and performing a forward-folding likelihood fit, utilising the instrument response functions (here, the effective area and the energy dispersion matrix). For this paper, we always assumed that the spectrum has the form of a power law,

$$\frac{dN}{dE} = \phi \times \left(\frac{E}{E_0} \right)^{-\Gamma}, \quad (1)$$

where ϕ and Γ are free parameters and E_0 is a normalisation energy that we chose such that the correlation between ϕ and Γ is minimised. We furthermore computed spectral points with a fixed binning of eight flux points per decade of energy, assuming that the spectral index in each bin is equal to that of the fitted power law.

In order to ensure an accurate reconstruction of the arrival direction of the primary particle, a minimum signal strength is

usually required for each telescope (for the analysis configuration used for the data we analysed here, this signal threshold is set at 80 photo-electrons for each camera image). This signal threshold per telescope translates into a minimum energy of the primary γ ray. γ rays around and below that energy can only be detected (i.e. pass the signal threshold) if an upward fluctuation of the signal occurs. This typically leads to a bias in the reconstructed energy of γ rays at these energies. While this can be corrected for in the extraction of the energy spectrum, it is usually a good measure to define an analysis energy threshold that ensures that the bias is not too large. Here, we required that the bias of the energy reconstruction of γ -ray events incident under an off-set angle corresponding to that of the location of the observed source does not exceed 10%.

We determined the positions of off regions with each of the tools separately, requiring that no regions are placed at the location of known γ -ray sources⁷. We also computed analysis energy thresholds as described above with HAP and Gammapy and applied them in the extraction of the energy spectrum with these tools, respectively. Since ctools in its current version does not provide the possibility to compute energy thresholds based on the energy reconstruction bias, we used the thresholds determined with Gammapy also for the spectrum extraction with ctools. We furthermore note that ctools, unless provided with instrument response functions specifically prepared for point sources, always applies a correction to the effective area based on an assumed leakage of γ -ray events outside the defined source region. This is not desired in the reflected background analysis of the two extended sources studied here (MSH 15–52 and RX J1713.7–3946), since we chose on regions that encompass the sources by a large enough margin such that the leakage is negligible; this is a standard procedure in H.E.S.S. data analysis. For the sake of better comparability, we therefore decided to actively disable this correction by modifying the ctools source code.

We show energy spectra extracted with the reflected background method for the MSH 15–52 and RX J1713.7–3946 data sets in Figs. 5 and 6, respectively. Spectra derived for other sources can be found in Appendix A. Finally, we compare the best-fit parameter values for the normalisation and spectral index of the power law model obtained with the different tools for all sources in Fig. 7. Table B.1 furthermore lists the best-fit parameter values for all analyses. In general, we observe an excellent agreement between the three different tools for all spectra, both concerning the power-law fits and the spectral flux points. The remaining differences give an indication of the systematic uncertainties associated with the implementation of the analysis technique.

It is noteworthy that the spectrum extracted for RX J1713.7–3946 (cf. Fig. 6) does not agree very well with the spectrum published in Abdalla et al. (2018d) at energies below ~ 0.45 TeV. We attribute this discrepancy to the fact that the reflected background method is not well suited for sources with an extent as large as RX J1713.7–3946⁸. Indeed, we were able to find only a single off region for almost all

Fig. 5. Comparison of spectra for MSH 15–52. The spectra shown in red, green, and blue were derived with the reflected background method with HAP, ctools, and Gammapy, respectively. For the HAP analysis, we show the result of the power-law fit in addition. We compute upper limits (95% confidence level) for flux points with a statistical significance of less than two standard deviations. The published spectrum is taken from Aharonian et al. (2005b).

Fig. 6. Comparison of spectra for RX J1713.7–3946. The spectra shown in red, green, and blue were derived with the reflected background method with HAP, ctools, and Gammapy, respectively. For the HAP analysis, we show the result of the power-law fit in addition. We compute upper limits (95% confidence level) for flux points with a statistical significance of less than two standard deviations. The published spectrum is taken from Abdalla et al. (2018d); it uses a power law with exponential cut-off as spectral model.

of the observations in this data set, which makes the analysis susceptible to both statistical and systematic uncertainties that are normally reduced by averaging over multiple off regions. However, our main purpose in this section being the validation of the open-source tools, we note that the tools actually agree well also in this region and do not investigate the discrepancy to the published spectrum further.

4. Development of a background model template

In this section we introduce a procedure to construct a model for the residual cosmic-ray background in arbitrary field of views from archival H.E.S.S. observations. We note that the archival data are proprietary to the H.E.S.S. Collaboration and

⁷ Due to the presence of other sources, it is not possible to find off regions for four of the observations in the RX J1713.7–3946 data set; we excluded these observations from the spectral analysis. Furthermore, the algorithm implemented in ctools did not find off regions for an even larger number of observations. We therefore used the off regions determined with Gammapy for the extraction of the spectrum for RX J1713.7–3946 with ctools.

⁸ A more appropriate method to determine the residual cosmic-ray background has been used in Abdalla et al. (2018d).

Fig. 7. Comparison of spectral fit parameters for all sources. Displayed are the results obtained using the reflected background method with HAP, ctools, and Gammapy. We plot the deviation w.r.t. the results obtained with our standard tool HAP. The error bars denote statistical uncertainties only (68% confidence level).

not publicly available. The procedure is inspired by the work of Mayer (2014), but has been considerably advanced in the course of this work. The constructed model yields the expected background rate as a function of two field-of-view coordinates and the reconstructed energy of the primary particle, which means it is three-dimensional. We present the construction procedure in Sect. 4.1, before we characterise and validate the background model in terms of its spatial and spectral properties in Sect. 4.2. We make available as supplementary material to this paper the background model for the observations that are contained in the first H.E.S.S. public test data release (see Appendix D for more information).

4.1. Construction

We begin with the selection of archival H.E.S.S. observations that are used to construct the model. Here we considered only observations taken during the first phase of H.E.S.S., without the large CT5 telescope. Since we aim to model the residual cosmic-ray background only, we excluded observations taken in the direction of the Galactic plane ($|l| < 60^\circ$, $|b| < 5^\circ$), since we expect contamination from diffuse γ -ray emission there.

Table 3. Background model binning and statistical information.

	$-90^\circ < \phi < 90^\circ$		$90^\circ < \phi < 270^\circ$	
ϑ (deg)	N_{obs}	t_{live} (h)	N_{obs}	t_{live} (h)
0–10	99	44.5	660	301.8
10–20	392	177.8	994	455.6
20–30	650	297.0	1378	632.2
30–40	444	201.4	790	367.6
40–45	300	135.8	242	110.9
45–50	306	140.0	448	204.9
50–55	150	68.4	124	57.3
55–60	61	28.2	25	12.2

Notes. The number of observations N_{obs} used to construct the initial background model for each bin of zenith angle ϑ and azimuth angle ϕ . The corresponding live time t_{live} is listed as well.

We furthermore imposed “spectral” observation quality criteria, rejecting observations that have been taken in the presence of hardware failures or under bad atmospheric conditions (see Aharonian et al. 2006a, for more information). Finally, we used only observations in which all four small telescopes have participated in data taking. This selection yields 7063 observations with a total observation time of ≈ 3240 h, taken between January 21, 2004 and May 15, 2013.

The background rate depends on various observational parameters, requiring us to take these into account and construct a tailored background model for each observation. The construction procedure consists of two steps. We first created an initial model that takes the principal dependencies of the background rate into account. This initial model was subsequently refined in an iterative procedure, thus correcting for less pronounced effects.

4.1.1. Initial model

For the initial model, we first considered the pointing direction of the telescopes in the horizontal (i.e. local) coordinate system. This is motivated by the strong dependence of the background rate on the zenith angle of the observation (cf. the characterisation of the model in Sect. 4.2.1). The dependence on the azimuth angle (due to the Earth’s magnetic field), albeit less strong, could easily be incorporated into the model at this stage as well.

To take these effects into account, we grouped the observations in bins of the average zenith angle (ϑ) and azimuth angle (ϕ) of their pointing direction and constructed an initial model for each of these bins. As the background rate does not vary strongly with azimuth angle, it is sufficient to use only two bins for this parameter ($-90^\circ < \phi < 90^\circ$ and $90^\circ < \phi < 270^\circ$). The dependence of the background rate on the zenith angle is much stronger, in particular as the zenith angle increases. Here we used eight bins for this parameter, as listed in Table 3. The table furthermore lists the resulting number of observations (and corresponding observation time) available in each of the bins.

Further parameters that affect the background rate and could thus be considered in the construction of the model include for example the date of the observation (accounting for efficiency loss over time) or the atmospheric conditions during the observation. However, a too fine separation of observations into bins can lead to an insufficient number of observations per bin, resulting in too large statistical uncertainties. We therefore chose not to

Fig. 8. Illustration of background model construction. We exemplarily show the model for the bin with azimuth angle $90^\circ < \phi < 270^\circ$ and zenith angle $20^\circ < \theta < 30^\circ$, for energies between 0.8 and 1.1 TeV. *a*: sum of the number of registered events in each observation, excluding events with a reconstructed direction close to known γ -ray sources. *b*: effective exposure time of each spatial pixel. The effective exposure time is the summed observation time of all observations, corrected for the exclusion of events from regions around known γ -ray sources. *c*: averaged background rate, given by the number of registered events divided by effective exposure time, energy interval and solid angle. *d*: averaged background rate after the application of a spline-based smoothing algorithm. All vertical axis labels refer to the same colour bar.

add further dimensions to the initial model but rather to incorporate a correction for these effects in a later step (see Sect. 4.1.2).

We constructed the model in a field-of-view coordinate system that is aligned with the horizontal coordinate system, but rotated such that the pointing position of the corresponding observation lies at the equator at coordinates ($l = 0, b = 0$). In this system, the longitude axis points in the direction of decreasing azimuth angles, and the latitude axis points in the direction of increasing altitude angles (or decreasing zenith angles, respectively). This allowed us to compute an average background rate from all observations in each bin of θ and ϕ , even if their pointing positions in equatorial coordinates are different. We obtained an averaged rate in a square grid with a side length of 7.5° and spatial bins of 0.1° size. The energy axis is divided into 20 logarithmically spaced bins between 100 GeV and 100 TeV. In the computation, we discarded events from directions close to known γ -ray sources, applying a corresponding correction to the exposure time per spatial bin. The construction procedure is illustrated in Figs. 8a–c.

In the next step, we transformed the model into a field-of-view coordinate system that is aligned with the equatorial coordinate system. Similar to the altitude-azimuth-aligned system, this system is centred on the pointing position, but rotated such that the longitude and latitude axes are aligned with the right-ascension and declination coordinate, respectively. `ctools` requires any background model to be specified in this system, and `Gammapy` accepts models defined in this system as well. Finally, we smoothed the initial model using a two-dimensional cubic spline function to remove statistical fluctuations that are inevitably introduced by the procedure described above (see Fig. 8d⁹).

4.1.2. Refined model

In order to account for further observation-specific parameters in the background model construction, we need to assign an initial model to each individual observation. Here, we start by simply grouping all observations in the same azimuth and zenith angle

bins as used in the background model construction (cf. Table 3), and selecting for each observation the initial model that was constructed from observations falling into the same bin.

To assess the quality of the background model, and to apply corrections, we performed a likelihood fit of the model for each observation to the observed data, masking regions that contain known γ -ray sources. To avoid being dominated by bins at low energies, where statistics are large, we performed a fit in each energy bin separately, fitting the model normalisation in each bin. We did not perform a fit for energy bins that are below the energy threshold computed for the observation. We then obtained a single normalisation value for each observation by averaging over all energy bins.

Figure 9a shows the average background model normalisation obtained in this way for all observations that have been used to construct the model itself, as a function of the zenith angle θ of the observation. It is evident that the procedure of selecting an initial model simply based on the zenith angle bin leads to jumps in the average fitted normalisation at the boundaries between bins. We therefore proceeded to performing a linear interpolation of the predicted background rate between adjacent zenith angle bins to assign a model to each observation. Re-performing the likelihood fit for all observations with this model, we observe that the average fitted background normalisation no longer depends on the zenith angle, as shown in Fig. 9b.

In Fig. 10, we show the average fitted background normalisation as a function of the so-called transparency coefficient. This coefficient is computed based on the trigger rate of the telescopes and describes the optical transparency of the atmosphere (in arbitrary units), with larger values implying a more transparent atmosphere (for more details see Hahn 2014). Unsurprisingly, the fitted background normalisation is correlated with the transparency coefficient (Spearman correlation coefficient 0.57), reflecting the fact that a decreased absorption of Cherenkov photons in the atmosphere leads to an increased rate of triggered background events. The dependence can be fitted well with a linear function, as indicated in the figure. Aiming for a more accurate description of the background rate, we used this function as a correction in the construction of the initial model as well as in the assignment of the initial model to individual observations. After this correction, the fitted background normalisation is no longer correlated with the transparency coefficient.

⁹ We note that, for illustrative purposes, the figure displays the smoothed model still in the altitude-azimuth-aligned field-of-view coordinate system, whereas we actually apply the smoothing algorithm *after* transforming into the RA/Dec-aligned field-of-view coordinate system.

Fig. 9. Average fitted background model normalisation as a function of zenith angle θ : (a) without zenith-angle interpolation; (b) with zenith-angle interpolation. The normalisation averaged over energy bins is shown for all observations used in the model construction. Boundaries between zenith angle bins as used in the construction of the model are marked by the black dashed lines. The blue markers show the individual observations, whereas the red data points denote the mean and standard deviation in bins of θ .

Fig. 10. Average fitted background model normalisation as a function of transparency coefficient. The grey points denote all observations taken with four active telescopes. Observations fulfilling “spectral” quality criteria in addition are shown in orange. The observations that were used to construct the background model are depicted in blue. The black, dashed line shows the fit function that is used to correct for the atmospheric transparency.

As a last step, we applied a correction factor for the so-called “optical phase” of each observation (see Table 4). Optical phases are specific periods in time that have been defined in order to account for the varying optical efficiency of the telescopes, for example due to mirror degradation¹⁰. One set of Monte Carlo

Table 4. Correction factors for different “optical phases”.

Phase	Start	Correction factor
1	2004-01-21	1.145
1b	2004-05-26	0.994
1c	2007-07-03	0.948
1c1	2010-04-27	0.960
1c2	2010-10-17	1.047
1c3	2011-04-14	1.082
1d	2011-11-12	1.004

Notes. Correction factor applied for each optical phase. The start date of each phase is given, with each phase lasting up to the start of the subsequent one. See main text for details.

Fig. 11. Distribution of fitted background normalisations. The histograms display the fitted normalisation values for all energy bins, i.e. the normalisations are not averaged for each observation here. The grey histogram shows the distribution for the final model and all 4-telescope observations. The coloured histograms display the distribution for the observations that have been used to construct the model, without corrections (red), with interpolation between zenith angle (θ) bins (orange), with a correction for atmospheric transparency (blue), and with a correction for optical phases (green). We give the standard deviation of the distributions in the upper right corner of the plot.

simulations for the generation of instrument response functions has been generated by the H.E.S.S. Collaboration for each optical phase. We observe a slight bias of the fitted background normalisations with respect to the expected value of 1 for some of the phases, which we attribute to imperfections in the Monte Carlo simulations. Using the average fitted background normalisation for each phase (listed in Table 4) as correction factor, we were able to eliminate this bias.

We illustrate the improvement in accuracy achieved by the refinement procedure in Fig. 11. While the distribution of fitted background normalisations has a standard deviation of 15% for the initial model (considering only observations used in the model construction), we obtain a width of 9% for the refined model. Including also observations that were not used to construct the model, the standard deviation increases to 12%. This is expected, considering that observations not fulfilling spectral quality criteria or taken in the direction of the Galactic plane cannot be perfectly described by the background model.

Finally, we note that the background model often fails to describe the data well close to the energy threshold of the instrument, where the variation of the background rate with energy is large and strongly dependent on the specific observation

¹⁰ The series of comparatively short periods in 2010 and 2011 are motivated by an exchange of the mirrors of the H.E.S.S. telescopes performed during that time.

Fig. 12. Background model visualisation. We show the background model in the field-of-view coordinate system, in four different energy bins. Here, the model for the bin with azimuth angle $90^\circ < \phi < 270^\circ$ and zenith angle $20^\circ < \theta < 30^\circ$ is displayed. The rate in all energy bins but the first has been multiplied by the factor indicated in the figure, to allow for a common colour scale.

conditions. Therefore, it is usually necessary to apply an analysis energy threshold, thus restricting the analysis to energies where the background model describes the cosmic-ray background well.

4.2. Characterisation and validation

4.2.1. Characterisation

Figure 12 shows a visualisation of the spatial shape of the final model in different energy bins. We show the model in the altitude-azimuth aligned field-of-view coordinate system here, that is, before the assignment to a specific observation. The background rate is clearly asymmetric with respect to the y-coordinate in the first energy bin shown. This reflects the altitude-angle (or zenith-angle) dependence of the background rate, since the y-coordinate is aligned with the altitude coordinate in the chosen coordinate system. Around 1 TeV, the shape is symmetric and peaked at the centre (i.e. at the location of the pointing). For higher energy bins, we observe an increase of the background rate at large offset angles, leading to a ring-shaped distribution at the highest energies. This is due to the poor rejection power for cosmic-ray background events obtained with the analysis configuration that has been employed to prepare the data used here; we observe this feature to be much less pronounced for an analysis configuration with better γ -hadron separation (e.g. as in Ohm et al. 2009).

We show the spectral shape of the final model in Figs. 13 and 14 (again before the assignment to a specific observation). We observe that the spectral shape is close, but not identical, to the shape of the primary cosmic-ray spectrum, which follows a power law $\propto E^{-2.7}$ in good approximation at the energies relevant here. The discrepancies can be attributed to a dependence of the effective area on the energy, caused for example by an energy-dependent event selection efficiency.

Figure 13, which shows the background model spectrum for different zenith angle bins, illustrates that the energy threshold increases with increasing zenith angle. This reflects the increased

Fig. 13. Background model energy spectrum in different bins of zenith angle θ , for azimuth angles $90^\circ < \phi < 270^\circ$. Shown is the rate integrated in a circle around the pointing position with radius 2.5° . To enhance features, the vertical axis is multiplied by $E^{2.7}$.

absorption of Cherenkov photons due to, on average, larger distances between the telescopes and the air shower for larger zenith angles. It is also evident that the effective area of the telescopes increases with increasing zenith angles, leading to a larger rate of background events. This well-known effect can be understood when considering that air showers that are incident under a large angle illuminate with Cherenkov photons a larger area on the ground, thus increasing the probability that enough telescopes trigger the event.

Figure 14 shows the background model spectrum for different offset angles Ψ with respect to the centre of the field of view. Here, we observe again the feature that, at high energies, the background rate is larger at high offset angles than at the centre.

4.2.2. Validation

Before applying the constructed background model in data analysis, we performed a general validation of the model by comparing it to archival H.E.S.S. data. This procedure is similar to that outlined in Sect. 4.1.2, where we already fitted the normalisation of the model to archival observations in separate energy bins. Here, we adapted this fit such that it resembles more the utilisation of the model in the data analysis with ctools or Gammapy. These tools currently offer the possibility to fit a model normalisation (across all energy bins) and a spectral “tilt”, that is to say, a parameter δ that modifies the predicted background rate R at energy E as

$$R' = R \cdot (E/E_0)^{-\delta}, \quad (2)$$

where $E_0 = 1$ TeV is a reference energy. Performing a fit of these parameters to all archival H.E.S.S. Phase-I observations, we obtained the parameter distributions displayed in Fig. 15.

Considering only observations used in the construction of the model, we obtain again a normalisation distribution with a standard deviation of 9% (cf. Fig. 11). This, together with the narrow

Fig. 14. Background model energy spectrum for zenith angles $20^\circ < \theta < 30^\circ$ and azimuth angles $90^\circ < \phi < 270^\circ$. Shown is the rate integrated in concentric rings of equal area around the pointing position. To enhance features, the vertical axis is multiplied by $E^{2.7}$.

distribution obtained for the spectral tilt parameter (width 0.05), demonstrates the validity of the background model in terms of its normalisation and spectral shape for a very large and diverse set of observations.

Evaluating the validity of the prediction of the spatial shape of the model is slightly more complicated, owing to the fact that the spatial shape is not easily parametrisable. For single observations, the validity can be checked, for instance, by inspecting the predicted and observed rate for one-dimensional slices; an example is shown in Fig. 16. We observe a good agreement between the prediction of the fitted model and the measured data, as well as between the model prediction and the prediction of the ring background method in this case.

For a more quantitative evaluation that can also be applied to many observations, we performed a χ^2 test for these one-dimensional slices, both along the right-ascension and declination axis. Since the model can only be expected to give a good prediction for regions that are free of γ -ray sources, we excluded regions that contain known sources from the χ^2 computation. The χ^2 test yields a p-value that is expected to follow a flat distribution if the model describes the data perfectly. We show the distributions obtained for slices along both axes in the left panels of Fig. 17, observing only slight deviations from a flat distribution. The right panels display distributions of the corresponding significance values in terms of standard deviations of a normal distribution. Again, we observe only a small deviation from the expectation of a Gaussian distribution with zero mean and unity width, concluding that also the spatial shape of the cosmic-ray background is described well by our model. Inspecting closer those observations with the smallest χ^2 probabilities, we found that sometimes hardware malfunctions or bad atmospheric conditions are responsible for the bad agreement. However, we were not able to identify single causes that are responsible for a bad agreement in a majority of the observations. We note that it is possible to utilise the result of the χ^2 test as an additional quality criterion for the analysis, that is, to discard observations for

which the agreement of the background model fit is particularly bad.

5. Validation of the 3D likelihood analysis

We applied the 3D likelihood analysis, using *ctools* and *Gammapi*, to the same data sets that we have used for the validation of the analysis tools in Sect. 3. In each case, we described the cosmic ray-induced background using the background model template introduced in the previous section. Section 5.1 provides a description of the analysis details. We present the analysis results in Sect. 5.2.

5.1. Analysis description

The analysis is based on a fit of (typically multiple) models to the observed data. In this paper, we always use one model (per observation) for the residual cosmic-ray background and one model for the analysed source. The fit proceeds via an optimisation of a likelihood function that expresses the agreement between the model prediction and the observed data, taking into account the instrument response functions (IRFs), that is, the effective area, the point spread function (PSF), and the energy dispersion matrix. The likelihood function depends on the reconstructed direction and energy of the observed events, meaning that it has three dimensions.

5.1.1. General settings

We performed the likelihood analysis with the two tools in a conceptionally different way: while we carried out an unbinned likelihood fit with *ctools*, we used a binned likelihood fit in *Gammapi*. The former uses probability density functions for the IRFs to determine a likelihood for each observed event, while the latter is a forward-folding fit employing a Poisson likelihood in each bin. The reason for this choice is that, on the one hand, an unbinned fit is not possible yet with *Gammapi*, while on the other hand, a binned fit including full treatment of the energy dispersion is computationally extremely intensive with the current version of *ctools* at least for parts of the data sets analysed here. Both methods yield identical results for large statistics and not-too-coarse binning.

With both tools, we employed a “joint” fit, that is, we calculated a likelihood value for each observation of the data set and multiplied these values to obtain the final likelihood value. This is opposed to a “stacked” fit, for which the measured data are summed over all observations and a model prediction for the entire data set, based on averaged IRFs, is obtained, leading to only one likelihood value. The joint fit, not relying on an averaged description of the instrument, is generally expected to lead to a more accurate model for the observed data. However, since there is typically at least one free fit parameter for the background model template of each observation in a joint fit, it can have many more free parameters than the corresponding stacked fit for the same data set. This is still feasible for the data sets analysed here, where the maximum number of observations in a data set is 20 (for MSH 15–52). We expect that for considerably larger data sets, the fit becomes computationally prohibitive due to the large number of free parameters, calling for more elaborate analysis strategies.

For the binned fit with *Gammapi*, as well as for the generation of result sky maps, we utilised spatial pixels with 0.02° side length. The energy axis is defined with eight logarithmically spaced bins per decade in energy.

Fig. 15. Distribution of fitted background normalisations and spectral tilts (δ , cf. Eq. (2)). We show results for all 4-telescope observations (grey, filled histograms) and those used in the construction of the background model (green histograms). The mean (μ) and standard deviation (σ) of the distributions are indicated.

Fig. 16. Number of events for observation 47829 at final selection level, as a function of right-ascension (J2000). The observation is part of the PKS 2155–304 (steady) data set. We show all events with declination $|\delta - \delta_{\text{pnt}}| < 0.7$ deg, where δ_{pnt} is the declination coordinate of the pointing position. Black points are the measured data; the prediction of the fitted background model is shown as a blue line. The grey-shaded area marks an exclusion region around the position of PKS 2155–304. The text box in the upper left corner denotes the applied energy threshold as well as the result of a χ^2 test for the agreement between the model and the measured data. We show in addition the number of background events obtained from the ring background method for this observation (red line, cf. Sect. 3.1).

In the analysis of IACT data, the largest systematic uncertainties in the description of the residual cosmic-ray background typically occur at large offset angles, at the edges of the observed field of view, and at the lowest energies, close to the instrument threshold. We therefore restricted the analysis to events within a maximum offset angle, Ψ_{max} , and above an energy threshold value, E_{thr} . We chose Ψ_{max} such that the observed source

as well as sufficiently large areas free of known γ -ray sources (the latter providing constraints for the background model template) are enclosed inside the selected region for all observations. We computed the energy threshold for each observation separately, defining it as $E_{\text{thr}} = \max(E_{\text{thr}}^{\text{bias}}, E_{\text{thr}}^{\text{bkg}})$. Here, $E_{\text{thr}}^{\text{bias}}$ denotes a threshold value that ensures that the bias of the energy reconstruction does not exceed 10% for events within the maximum offset angle; this requirement is similar to that imposed in the spectrum extraction with the reflected background method (cf. Sect. 3.2). In addition, we introduced a threshold value $E_{\text{thr}}^{\text{bkg}}$ for the background model of each observation, conservatively defined as the upper edge of the energy bin with the highest predicted background rate. This was necessary here because many of the analysed observations have been taken very early in the operation time of H.E.S.S., when the optical efficiency of the telescopes was very high, leading to a comparatively low $E_{\text{thr}}^{\text{bias}}$. These observations are sometimes not described well close to the threshold by the background model, which is constructed from observations that, on average, have been conducted with lower optical efficiency of the telescopes. Thus, $E_{\text{thr}}^{\text{bkg}}$ is larger than $E_{\text{thr}}^{\text{bias}}$ for most of the observations analysed here. This is not the case in analyses of data sets recorded under less exceptional conditions, which can likely proceed without this additional restriction. We list the resulting range of values for E_{thr} , as well as the employed values of Ψ_{max} , for each data set in Table 5.

5.1.2. Background and source models

As model for the residual cosmic-ray induced background, we used the background model template developed in Sect. 4. This template, specific for each observation, can be modified in the fit by two parameters: a global normalisation factor, and a spectral “tilt”, as defined previously in Eq. (2).

Like in Sect. 3.2, we always used a simple power law as spectral model for the source, with the flux normalisation and spectral index as free fit parameters (see Eq. (1)). For the estimation of flux points, we re-performed the analysis for each of the energy bins, fixing the spectral index, the background tilt parameters, and the parameters of the spatial source model to their best-fit values, but leaving free the background and source normalisation parameters.

Fig. 17. Distributions of χ^2 statistic values. We show results for all 4-telescope observations (grey, filled histograms) and those used in the construction of the background model (green histograms). Distributions are shown for a slice along the right-ascension (*top*) and declination (*bottom*) axis (see main text for details). Displayed is the χ^2 probability (*left*) and the corresponding significance in terms of standard deviations of a normal distribution (*right*). The fitted mean (μ) and width (σ) of the significance distributions are indicated as well.

Table 5. Event selection criteria for 3D likelihood analysis.

Source	$\Psi_{\max}$ (deg)	E_{thr} (TeV)
Crab nebula	2.0 ^(a)	0.750–0.825
PKS 2155–304	1.5	0.422–0.619
MSH 15–52	1.5	0.422–0.562
RX J1713.7–3946	1.5 ^(b)	0.316–0.422

Notes. $\Psi_{\max}$ denotes the maximum event offset angle; the last column gives the minimum and maximum applied energy threshold of all observations in the data set. ^(a)We use the larger value of 2° for the Crab nebula because two (out of four) observations of this data set have been taken with an offset angle of 1.5° w.r.t. the source position. ^(b)The data set for RX J1713.7–3946 contains one observation with a considerably larger offset between the pointing direction and the source w.r.t. the remaining observations. For this observation, the source is not fully enclosed in the selected region. This leads to a slight loss in sensitivity, which we accept here for the sake of a consistent handling of the observations in the analysis.

The different morphologies of the sources analysed in this paper call for different spatial source models. For the Crab nebula and PKS 2155–304, we employed the model of a point-like source, with the two source coordinates as free parameters. The morphology of MSH 15–52 is still simple enough to be able to use an analytical spatial model for this source as well; we used an elliptical disk model with five free parameters (the source coordinates and the major axis, eccentricity, and position angle of the ellipse) here. In contrast, the complex morphology of RX J1713.7–3946 (cf. Fig. 3) prohibits the use of an analytical model. We have therefore developed a procedure to generate an “excess template” as spatial source model. As the name suggests, this template represents a map of the excess of events that can be attributed to γ -ray emission from the source. We derived it by fitting only the background model template to the observations, excluding a region around the source from the fit, and subtracting the resulting best-fit background model from the observed data. We note that the template – being derived from the data themselves – is subject to the same statistical

fluctuations as the observed data, implying that this approach can in principle lead to a bias of the fitted parameters. In an attempt to minimise such a bias, we smoothed the excess map using a two-dimensional cubic spline function, thus reducing the statistical fluctuations. Finally, we clipped the derived template map at zero, removing negative entries. We furthermore note that, since the excess template is derived from observed data, it need not be convolved with the PSF in the fit; this is possible currently with Gammapy but not with cttools. We performed an analysis with the excess template approach not only for RX J1713.7–3946, but also for MSH 15–52, being able to compare to the results obtained with the elliptical disk model in that case.

5.2. Results

We summarise the results of all 3D likelihood analyses in Table B.1. Furthermore, Fig. 18 shows a comparison of the best-fit parameter values of the power law model for all sources. The spectrum obtained with the 3D likelihood analysis is steeper (i.e. the spectral index Γ is larger) than that obtained with the reflected background method in all cases. Comparing the spectra in detail, we find that this is partly due to an improved sensitivity of the 3D likelihood analysis at high energies, where we obtain only flux upper limits because the excess of γ -ray events is not significant. Here, the reflected background method suffers from poor statistics in the off regions as well, leading to an inaccurate background estimate and hence bad sensitivity. In contrast, the cosmic-ray background model employed in the 3D likelihood analysis, being derived from many observations, is afflicted less by this problem. Consequently, the 3D likelihood analysis yields slightly steeper spectra, and more constraining upper limits at high energies. Furthermore, the fact that we apply slightly higher energy thresholds in the 3D likelihood analyses can also explain part of the discrepancy, in particular if the intrinsic source spectrum is not a true power law. We note that the deviation between the two methods is at most ~ 0.1 for the spectral index and $\sim 20\%$ for the flux normalisation; this is within the systematic uncertainties on these parameters that are usually quoted by H.E.S.S. (see e.g. Aharonian et al. 2006a).

Fig. 18. Comparison of spectral fit parameters for all sources, obtained using 3D likelihood analysis with ctools and Gammapy. We plot the deviation w.r.t. the results obtained with the reflected background method and our standard tool HAP. The error bars denote statistical uncertainties only (68% confidence level).

In the following, the results obtained for the different data sets are discussed in detail. Due to space restrictions, we can show spectra and maps only for a selection of the analyses here; all remaining plots can be found in Appendix C. Furthermore, we always only show maps derived with one of the open-source tools, implying that the corresponding maps derived with the other tool are qualitatively and quantitatively compatible.

5.2.1. Crab nebula and PKS 2155–304

We show the energy spectra that we obtain for PKS 2155–304 and the Crab nebula in Figs. 19 and C.1, respectively. Figures 20 and C.2 show the respective residual significance maps, together with the corresponding entry distributions. The computation of the significance maps starts by convolving the best-fit model prediction and the observed data (both integrated over energy) with a top-hat kernel of 0.1° radius, thus smoothing statistical fluctuations (cf. Sect. 3.1). We then calculate the significance of the observed data under the hypothesis of the best-fit model

Fig. 19. Comparison of spectra for PKS 2155–304. The spectra shown in green and blue were derived using a 3D likelihood analysis with ctools and Gammapy, respectively. The butterflies show the fitted power laws. The results are compared to those obtained with the reflected background method using the HAP software (in red). We compute upper limits (95% confidence level) for flux points with a statistical significance of less than two standard deviations.

Fig. 20. Residual significance map for PKS 2155–304, in equatorial coordinates (J2000). Results derived with the 3D likelihood analysis with Gammapy are displayed. We apply a convolution with a top-hat kernel of 0.1° radius to reduce statistical fluctuations. The position of PKS 2155–304 is indicated by the "x". Upper panel: distribution of significance values, together with the fit of a normal distribution (red line).

Fig. 21. Counts map (a) and best-fit model map (b) for MSH 15–52, in equatorial coordinates (J2000). Results derived with the 3D likelihood analysis with `ctools` using an elliptical disk as source model are displayed. The maps have been integrated over all energy bins contributing to the fit and smoothed using a Gaussian kernel with a width of 0.08° . The small panels show projections onto the two spatial axes; the separate components are indicated as well here.

following [Li & Ma \(1983\)](#), adopting the limiting case of a perfectly known number of “off” counts (corresponding to the best-fit model prediction in our case).

The data sets of the Crab nebula and PKS 2155–304 comprise few observations and are hence dominated by statistical rather than systematic uncertainties. Hence, unsurprisingly, the likelihood analysis works well for these data sets, yielding results that are highly compatible with those obtained with standard analysis techniques. The significance maps are governed entirely by statistical fluctuations, indicating an almost perfect description of the analysed source as well as the residual background in the field of view.

5.2.2. MSH 15–52

Figure 21 shows sky maps of the observed data as well as the best-fit model obtained with the 3D likelihood analysis for MSH 15–52, using the elliptical disk as spatial model for the source. We smoothed the maps using a Gaussian kernel with a width of 0.08° , which approximately corresponds to the size of the PSF for the data sets analysed here. We observe a very good agreement between the two maps, indicating that both the cosmic-ray background and MSH 15–52 are described well by the fitted models. The slight disagreement between the background model prediction and the data in the western part of the maps does not seem to affect the fitted source model. This interpretation is supported by the essentially featureless significance map for this analysis, which we show in Fig. C.3

We show the spectrum obtained for MSH 15–52 with the 3D likelihood analysis using an elliptical disk model in Fig. 22. In

accordance with the maps, we observe an excellent agreement with the spectrum derived with the reflected background method, as well as with the published spectrum ([Aharonian et al. 2005b](#)).

We show the counts and model maps, the significance map, and the spectrum derived for MSH 15–52 using the 3D likelihood analysis with an excess template rather than an elliptical disk model in Figs. C.4–C.6, respectively. The results are highly compatible with those obtained with the elliptical disk model, giving us confidence that the procedure of generating a model template from the excess map is valid and can also be applied to the analysis of RX J1713.7–3946.

5.2.3. RX J1713.7–3946

Figure 23 shows the significance map derived for the 3D likelihood analysis of RX J1713.7–3946. The grey regions contain known γ -ray sources or bright stars and were masked in the fit^{11,12}. We observe two noticeable features. First, there is a residual positive excess to the south-east of RX J1713.7–3946. This could be caused by a true excess of γ rays (e.g. from unresolved sources, or of diffuse nature) as well as by an imperfect model of the cosmic-ray background, likely a combination

¹¹ In principle, the likelihood analysis offers the possibility to model all sources in the field of view. A full analysis of the field of view containing RX J1713.7–3946 is however beyond the scope of this paper.

¹² The exclusion of sub-regions from the fit is not easily possible in an unbinned fit with the version of `ctools` that we have used. Hence, no regions were excluded in the `ctools` fit. Performing the `Gammapy` fit without exclusion regions as well, we find that the results are altered by less than 5%.

Fig. 22. Comparison of spectra for MSH 15–52. The spectra shown in green and blue were derived using a 3D likelihood analysis with *ctools* and *Gammapy*, respectively, employing an elliptical disk as source model. The butterflies show the fitted power laws. The results are compared to those obtained with the reflected background method using the HAP software (in red). We compute upper limits (95% confidence level) for flux points with a statistical significance of less than two standard deviations. The published spectrum is taken from [Aharonian et al. \(2005b\)](#).

of both. Second, the region covered by the source model template is almost free of statistical fluctuations. This is an artefact of the generation of the source model from the excess map (cf. Sect. 5.1.2); although we applied a smoothing algorithm, the excess template necessarily is subject to the same statistical fluctuations as the data it is derived from.

Finally, we refer to the data and model maps and the spectrum for RX J1713.7–3946 in Figs. C.7 and C.8, respectively. Similarly as for the results obtained with the reflected background method, we observe a disagreement to the published spectrum ([Abdalla et al. 2018d](#)) at energies below ~ 0.45 TeV, that is, for the first derived flux point. The disagreement is reduced with respect to the standard method, indicating that the 3D likelihood analysis is a better choice of analysis method for this source. That a deviation to the published spectrum remains is likely due to a systematic problem with the (relatively small) data set analysed here.

6. Conclusion

In this paper, we demonstrate the applicability of the open-source high-level γ -ray analysis tools *ctools* and *Gammapy* to experimental data recorded by the H.E.S.S. system of Cherenkov telescopes. In Sect. 3, we show that when applying standard, well-tested analysis techniques, *ctools* and *Gammapy* are able to exactly reproduce results obtained with the proprietary H.E.S.S. analysis software. We then focus on the 3D likelihood analysis, an analysis approach that is new in the field of very-high-energy γ -ray astronomy. First, in Sect. 4, we introduce a procedure to construct a model template for the residual cosmic ray-induced background, one of the key ingredients for the 3D likelihood analysis approach. Furthermore, we characterise the principal features of the resulting background model and perform a general validation. Finally, in Sect. 5, we apply the 3D likelihood analysis to H.E.S.S. data. We obtain results that are highly compatible with those derived with standard analysis techniques,

Fig. 23. Residual significance map for RX J1713.7–3946, in equatorial coordinates (J2000). Results derived with the 3D likelihood analysis with *Gammapy* are displayed. The position of RX J1713.7–3946 is indicated by the “x”; the dashed line shows the size of the excess template. *Upper panel:* distribution of significance values, together with the fit of a normal distribution (red line).

thus demonstrating the validity of the background model as well as the analysis approach itself.

We note that the data set used for analysis verification in this paper (cf. Sect. 2) has various limitations: it comprises only few observations taken on relatively strong γ -ray sources; the corresponding fields of view do not require the modelling of multiple source components; and the analysis configuration utilised to process the data is no longer state-of-the-art. However, we are confident that the analysis concept can also be applied to larger data sets, more intricate fields of view, and data processed with up-to-date analysis configurations; first successful attempts have been made by [Mayer \(2014\)](#), [Devin \(2018\)](#), and [Ziegler \(2018\)](#). Furthermore, we remark that particularly complicated sky regions, such as for example the Galactic centre region, cannot be properly analysed with traditional analysis techniques at all, calling for new approaches like the one presented here. Nevertheless, we expect that further studies will be necessary for this. For instance, the question up to which level of precision the residual cosmic-ray background can be modelled using the approach described here for deep observations is important, but beyond the scope of this paper.

That we have successfully validated the application of *ctools* and *Gammapy* to the analysis of H.E.S.S. data is important not

only for the H.E.S.S. experiment, but also for the upcoming Cherenkov Telescope Array. First, it shows that the development of the analysis tools that could be used for CTA is progressing well; both packages – while still in development – can already be considered mature now. Second, this work also paves the way for the application of the 3D likelihood analysis to CTA data. CTA will have greatly improved sensitivity with respect to current instruments and is thus expected to discover many new sources of γ rays. Therefore¹³, it will benefit from the 3D likelihood analysis approach, which is designed to simultaneously analyse multiple components in the observed field of view¹⁴.

Acknowledgements. We would like to thank the H.E.S.S. Collaboration for allowing us to use their internal analysis software package and their archival data; we used the latter to construct and validate the background model. Furthermore, we acknowledge the preparatory work of the members of the H.E.S.S. FITS task group, which have prepared the H.E.S.S. public test data release. We thank the developer teams of the *ctools* and *Gammapy* packages for helpful support during the preparation of this paper. This research made use of *ctools*, a community-developed analysis package for Imaging Air Cherenkov Telescope data. *ctools* is based on *GammaLib*, a community-developed toolbox for the scientific analysis of astronomical γ -ray data. This research made use of *Gammapy*, a community-developed core Python package for γ -ray astronomy. This research made use of *Astropy*, a community-developed core Python package for Astronomy (Robitaille et al. 2013; Price-Whelan et al. 2018). The plots shown in this paper have been produced with the *matplotlib* package (Hunter 2007). The authors acknowledge support by the German Federal Ministry of Education and Research under grant ID 05A17WE1. The authors gratefully acknowledge the compute resources and support provided by the Erlangen Regional Computing Center (RRZE).

References

- Abdalla, H., Abramowski, A., Aharonian, F., et al. 2017, *A&A*, **598**, A39
 Abdalla, H., Abramowski, A., Aharonian, F., et al. 2018a, *A&A*, **612**, A9
 Abdalla, H., Abramowski, A., Aharonian, F., et al. 2018b, *A&A*, **612**, A1
 Abdalla, H., Abramowski, A., Aharonian, F., et al. 2018c, <https://doi.org/10.5281/zenodo.1421098>
 Abdalla, H., Abramowski, A., Aharonian, F., et al. 2018d, *A&A*, **612**, A6
 Abramowski, A., Acero, F., Aharonian, F., et al. 2010, *A&A*, **520**, A83
 Abramowski, A., Acero, F., Aharonian, F., et al. 2012, *A&A*, **548**, A38
 Acharya, B. S., Agudo, I., Al Samarai, I., et al. 2018, *Science with the Cherenkov Telescope Array* (World Scientific Publishing)
 Ackermann, M., Ajello, M., Albert, A., et al. 2017, *ApJ*, **840**, 43
 Aharonian, F. A., Akhperjanian, A. G., Aye, K.-M., et al. 2004, *Nature*, **432**, 75
 Aharonian, F., Akhperjanian, A. G., Aye, K.-M., et al. 2005a, *A&A*, **430**, 865
 Aharonian, F., Akhperjanian, A. G., Aye, K.-M., et al. 2005b, *A&A*, **435**, L17
 Aharonian, F., Akhperjanian, A. G., Bazer-Bachi, A. R., et al. 2006a, *A&A*, **457**, 899
 Aharonian, F., Akhperjanian, A. G., Bazer-Bachi, A. R., et al. 2006b, *A&A*, **449**, 223
 Aharonian, F., Akhperjanian, A. G., Bazer-Bachi, A. R., et al. 2007, *ApJ*, **664**, L71
 Aharonian, F., Akhperjanian, A. G., Anton, G., et al. 2009, *A&A*, **502**, 749
 Berge, D., Funk, S., & Hinton, J. 2007, *A&A*, **466**, 1219
 Deil, C., Zanin, R., Lefaucheur, J., et al. 2017, *Proc. 35th Int. Cosmic Ray Conf.*, No. 766
 Deil, C., Wood, M., & Hassan, T. 2018, <https://zenodo.org/record/1409831>
 Devin, J. 2018, PhD Thesis, Université de Montpellier, France
 Fernandes, M. V., Horns, D., Kosack, K., Raue, M., & Rowell, G. 2014, *A&A*, **568**, A117
 Hahn, J., de los Reyes, R., Bernlöhr, K., et al. 2014, *Astropart. Phys.*, **54**, 25
 Hillas, A. M. 1985, *Proc. 19th Int. Cosmic Ray Conf.*, **3**, 445
 Holler, M., Berge, D., van Eldik, C., et al. 2015, *Proc. 34th Int. Cosmic Ray Conf.*, No. 847
 Holler, M., Berge, D., Hahn, J., et al. 2017, *Proc. 35th Int. Cosmic Ray Conf.*, No. 676
 Hunter, J. D. 2007, *Comput. Sci. Eng.*, **9**, 90
 Knödseder, J., Mayer, M., Deil, C., et al. 2016, *A&A*, **593**, A1
 Li, T., & Ma, Y. 1983, *ApJ*, **272**, 317
 Mattox, J. R., Bertsch, D. L., Chiang, J., et al. 1996, *ApJ*, **461**, 396
 Mayer, M. 2014, PhD Thesis, Universität Potsdam, Germany
 Ohm, S., van Eldik, C., & Egberts, K. 2009, *Astropart. Phys.*, **31**, 383
 Price-Whelan, A. M., Sipőcz, B. M., Günther, H. M., et al. 2018, *AJ*, **156**, 123
 Robitaille, T. P., Tollerud, E. J., Greenfield, P., et al. 2013, *A&A*, **558**, A33
 Rowell, G. P. 2003, *A&A*, **410**, 389
 Ziegler, A. 2018, PhD Thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

¹³ <http://www.astropy.org>

¹⁴ <http://www.matplotlib.org>

Appendix A: Reflected background method spectra for point sources

Fig. A.1. Comparison of spectra for the Crab nebula. The spectra shown in red, green, and blue were derived with the reflected background method with HAP, ctools, and Gammapy, respectively. For the HAP analysis, we show the result of the power-law fit in addition. We compute upper limits (95% confidence level) for flux points with a statistical significance of less than two standard deviations. The published spectrum is taken from [Aharonian et al. \(2006a\)](#); it uses a power law with exponential cut-off as spectral model.

We show the spectra extracted with the reflected background method for the Crab nebula and PKS 2155–304 in Figs. A.1 and A.2, respectively. Since PKS 2155–304 is a variable source, we cannot meaningfully compare our results with the literature in this case.

Fig. A.2. Comparison of spectra for PKS 2155–304. The spectra shown in red, green, and blue were derived with the reflected background method with HAP, ctools, and Gammapy, respectively. For the HAP analysis, we show the result of the power-law fit in addition. We compute upper limits (95% confidence level) for flux points with a statistical significance of less than two standard deviations.

Appendix B: Fit results

Table B.1 lists the results of the spectral fits performed with the classical analysis approach (reflected background method, cf. Sect. 3), as well as the results obtained with the 3D likelihood analyses described in Sect. 5 (marked with the annotation “3D” in the table).

Table B.1. Comparison of best-fit parameter values obtained in all analyses.

Analysis	Source model	RA (J2000) (deg)	Dec (J2000) (deg)	a (deg)	e	PA (deg)	E_0 (TeV)	ϕ ($10^{-12} \text{ cm}^{-2} \text{ s}^{-1} \text{ TeV}^{-1}$)	Γ
Crab nebula									
HAP	–	–	–	–	–	–	1.45	15.9 ± 0.8	2.59 ± 0.06
ctools	–	–	–	–	–	–	1.45	16.2 ± 0.9	2.64 ± 0.08
Gammapy	–	–	–	–	–	–	1.45	16.3 ± 0.9	2.63 ± 0.07
ctools 3D	Point source	83.619 ± 0.003	22.025 ± 0.002	–	–	–	1.45	17.0 ± 0.8	2.66 ± 0.07
Gammapy 3D	Point source	83.619 ± 0.003	22.025 ± 0.003	–	–	–	1.45	17.3 ± 0.8	2.69 ± 0.07
PKS 2155–304									
HAP	–	–	–	–	–	–	0.65	18.0 ± 2.1	3.46 ± 0.24
ctools	–	–	–	–	–	–	0.65	17.4 ± 2.1	3.52 ± 0.28
Gammapy	–	–	–	–	–	–	0.65	17.5 ± 2.1	3.46 ± 0.25
ctools 3D	Point source	329.725 ± 0.007	-30.227 ± 0.006	–	–	–	0.65	18.5 ± 1.9	3.48 ± 0.22
Gammapy 3D	Point source	329.728 ± 0.006	-30.225 ± 0.006	–	–	–	0.65	22.1 ± 2.2	3.50 ± 0.23
MSH 15–52									
HAP	–	–	–	–	–	–	1.4	2.63 ± 0.16	2.31 ± 0.06
ctools	–	–	–	–	–	–	1.4	2.84 ± 0.16	2.34 ± 0.07
Gammapy	–	–	–	–	–	–	1.4	2.70 ± 0.16	2.35 ± 0.06
ctools 3D	Elliptical disk	228.553 ± 0.011	-59.174 ± 0.006	0.194 ± 0.010	0.794 ± 0.044	151 ± 5	1.4	3.07 ± 0.15	2.42 ± 0.06
Gammapy 3D	Elliptical disk	228.549 ± 0.011	-59.171 ± 0.007	0.196 ± 0.007	0.801 ± 0.038	151 ± 6	1.4	2.91 ± 0.14	2.41 ± 0.05
ctools 3D	Excess template	–	–	–	–	–	1.4	3.17 ± 0.15	2.41 ± 0.06
Gammapy 3D	Excess template	–	–	–	–	–	1.4	2.95 ± 0.13	2.39 ± 0.05
RX J1713.7–3946									
HAP	–	–	–	–	–	–	1.15	12.6 ± 0.7	2.16 ± 0.05
ctools	–	–	–	–	–	–	1.15	13.5 ± 0.7	2.18 ± 0.05
Gammapy	–	–	–	–	–	–	1.15	12.7 ± 0.7	2.16 ± 0.04
ctools 3D	Excess template	–	–	–	–	–	1.15	12.9 ± 0.5	2.22 ± 0.04
Gammapy 3D	Excess template	–	–	–	–	–	1.15	13.3 ± 0.5	2.21 ± 0.04

Notes. E_0 , ϕ , and Γ are parameters of the power law model as defined in Eq. (1). a , e , and “PA” denote the major axis, eccentricity and position angle of the elliptical disk model used for MSH 15–52, respectively. Results obtained with the 3D likelihood analysis approach are annotated with “3D”.

Appendix C: Additional maps and spectra for the 3D likelihood analysis

Fig. C.1. Comparison of spectra for the Crab nebula. The spectra shown in green and blue were derived using a 3D likelihood analysis with ctools and Gammapy, respectively. The butterflies show the fitted power laws. The results are compared to those obtained with the reflected background method using the HAP software (in red). The published spectrum is taken from [Aharonian et al. \(2006a\)](#); it uses a power law with exponential cut-off as spectral model.

Fig. C.2. Residual significance map for the Crab nebula, in equatorial coordinates (J2000). Results derived with the 3D likelihood analysis with Gammapy are displayed. We apply a convolution with a top-hat kernel of 0.1° radius to reduce statistical fluctuations. The position of the Crab nebula is indicated by the “x”. *Upper panel:* distribution of significance values, together with the fit of a normal distribution (red line).

Fig. C.3. Residual significance map for MSH 15–52, in equatorial coordinates (J2000). Results derived with the 3D likelihood analysis with ctools using an elliptical disk as source model are displayed. We apply a convolution with a top-hat kernel of 0.1° radius to reduce statistical fluctuations. The position of MSH 15–52 is indicated by the “x”. *Upper panel:* distribution of significance values, together with the fit of a normal distribution (red line).

We show the spectrum and significance map obtained with the 3D likelihood analysis for the Crab nebula in Figs. C.1 and C.2, respectively.

Figure C.3 shows the significance map for the 3D likelihood analysis of MSH 15–52 with an elliptical disk model.

Figures C.4–C.6 display the results of the 3D likelihood analysis of MSH 15–52 with an excess template model.

Finally, we show the data and model maps and the spectrum derived with the 3D likelihood analysis for RX J1713.7–3946 in Figs. C.7 and C.8, respectively.

Fig. C.4. Counts map (a) and best-fit model map (b) for MSH 15–52, in equatorial coordinates (J2000). Results derived with the 3D likelihood analysis with *ctools* using an excess template as source model are displayed. The maps have been integrated over all energy bins contributing to the fit and smoothed using a Gaussian kernel with a width of 0.08° . The small panels show projections onto the two spatial axes; the separate components are indicated as well here.

Fig. C.5. Residual significance map for MSH 15–52, in equatorial coordinates (J2000). Results derived with the 3D likelihood analysis with ctools using an excess template as source model are displayed. We apply a convolution with a top-hat kernel of 0.1° radius to reduce statistical fluctuations. The position of MSH 15–52 is indicated by the “x”. *Upper panel:* distribution of significance values, together with the fit of a normal distribution (red line).

Fig. C.6. Comparison of spectra for MSH 15–52. The spectra shown in green and blue were derived using a 3D likelihood analysis with `ctools` and `Gammapy`, respectively, employing an excess template as source model. The butterflies show the fitted power laws. The results are compared to those obtained with the reflected background method using the `HAP` software (in red). The published spectrum is taken from [Aharonian et al. \(2005b\)](#).

Fig. C.7. Counts map (a) and best-fit model map (b) for RX J1713.7–3946, in equatorial coordinates (J2000). Results derived with the 3D likelihood analysis with Gammapy using an excess template as source model are displayed. The maps have been integrated over all energy bins contributing to the fit and smoothed using a Gaussian kernel with a width of 0.08° . The small panels show projections onto the two spatial axes; the separate components are indicated as well here.

Fig. C.8. Comparison of spectra for RX J1713.7–3946. The spectra shown in green and blue were derived using a 3D likelihood analysis with ctools and Gammapy, respectively, employing an excess template as source model. The butterflies show the fitted power laws. The results are compared to those obtained with the reflected background method using the HAP software (in red). The published spectrum is taken from Abdalla et al. (2018d); it uses a power law with exponential cut-off as spectral model.

Appendix D: Supplementary material: background model templates

We make available the three-dimensional templates for the residual cosmic-ray background that we derive in this paper for all observations that are part of the first public H.E.S.S. test data release (see Abdalla et al. 2018c). We note that the usage of the public H.E.S.S. test data set is subject to the terms of use that are distributed together with the data, in particular “no scientific publications may be derived from the data”.

The material can be found at the following URL: https://github.com/lmohrmann/hess_ost_paper_material

Appendix E: Supplementary material: machine-readable tables of spectral results

We release as ASCII text files the results of all spectral fits carried out in this paper. Both the results of the fitted power-law models as well as extracted spectral flux points are available for each of the analysis tools that we have used (i.e. the H.E.S.S.-internal analysis software program HAP and the open-source packages ctools and Gammapy).

The material can be found at the following URL: https://github.com/lmohrmann/hess_ost_paper_material