Lehrstuhlversuch im SS2020

Das Energiespektrum des Krebsnebels, gemessen mit FACT

Fabian Koch
fabian3.koch@tu-dortmund.de
Nils Breer
nils.breer@tu-dortmund.de
Nicole Schulte
nicole.schulte@tu-dortmund.de

Date of submission: 25.10.2020

TU Dortmund – Fakultät Physik

Contents

1 Theoretical basics and datasamples **3**
- 1.1 The FACT telescope . . . 3
- 1.2 Datasamples . . . 3
- 1.3 Unfolding . . . 4
  - 1.3.1 Naive SVD unfolding . . . 5
  - 1.3.2 Poisson-likelihood-unfolding . . . 5

2 Implementation **5**

3 Analysis **6**
- 3.1 Theta-squared plot and detector significance . . . 6
- 3.2 Migration matrix and unfolding . . . 6
  - 3.2.1 Naive SVD unfolding . . . 8
  - 3.2.2 Poisson-likelihood unfolding . . . 9
- 3.3 Flux calculation . . . 9
- 3.4 Comparison to HEGRA and MAGIC . . . 10

4 Evaluation **11**

5 Anhang **12**

References **12**

1 Theoretical basics and datasamples

1.1 The FACT telescope

FACT (first G-APD Cherenkov Telescope) is the first imaging atmospheric Cherenkov Telescope which uses Geiger-mode avalanche photodiods (G-APD). The main purposes of this telescope is being a benchmark for this kind of technology in Cherenkov astronomy. The second main purpose is to monitor bright and active galactiv nuclei (AGN) in the Terascale. FACT uses the wobble-mode 1 to observe extragalactic sources where the telescope is aimed 0,6° next to source position in order to estimate the background rate in addition to the signal from the source. There is one on-position for the source and 5 off-positions for the background. The reconstructed events in the on-region are labeled "N_{on}" whereas off-region events are labeled "N_{off}". With these parameters, the detector significance is

$$\mathrm{S} = \sqrt{2} \cdot \sqrt{N_{on} \ln\left(\frac{1+\alpha}{\alpha}\left(\frac{N_{on}}{N_{on}+N_{off}}\right)\right) + N_{off} \ln\left((1+\alpha)\left(\frac{N_{off}}{N_{on}+N_{off}}\right)\right)},$$

where α is ratio of the on-region and off-region. In the wobble-mode with one on-postion and five off positions, the ratio is $\alpha = 0.2$.

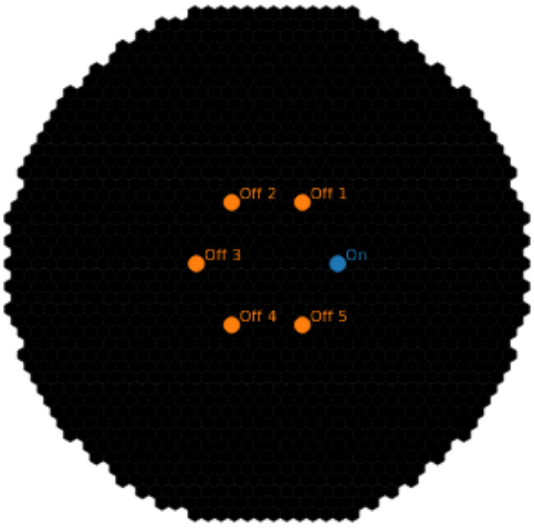


Figure 1: Displayed is the method of the wobble-mode of the FACT detector [1].

1.2 Datasamples

The given datasamples are preprocessed based on the analysis chain typical for data from Cherenkov-Telescopes. For the shower-events, three attributes of the original particles must be reconstructed: the particle class, the energy and the direction to the origin. The particle class is typically a gamma-ray or a hadron, for example a proton. The

energy is reconstructed via regression, most of the time only for gamma-ray candidates. The direction of origin is calculated with a two-dimensional regression in either celestial coordinates or detector coordinates. A typical event is shown in Figure 2.

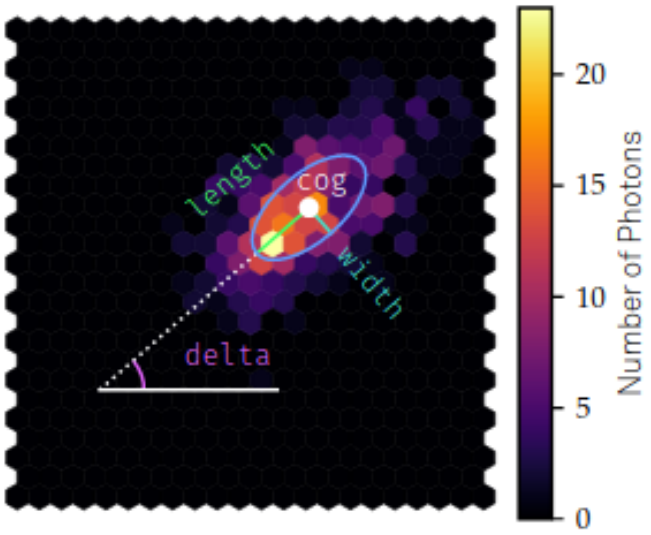


Figure 2: Displayed is a typical event with the number of photons stored in each bin. A principal component analysis is additionally displayed to visualize, how the method obtains the parameters `width` and `length` [1].

The standard deviations `width` and `length` are calculated from a principal component analysis. The middle of the event is called `cog` and the orientation is defined through angle `delta` with regards to the x-axis.
To determine the particle class, the energy and the direction of the origin as well as the detector acceptance, complex simulations are needed to generate the training data. The software `CORSIKA` was used to generate the airshower, cherenkov-production as well as particle propagation. After that, the software saves the Cherenkov-light detected. To simulate the FACT telescope, the software `CERES` was used. Using machine learning techniques, three datasamples were generated:

- gammas from point-like source, observed with wobble mode
- gammas coming from random directions
- protons coming from random directions

1.3 Unfolding

In physical measurements, the interesting parameters are almost never measured directly. In the astrophysics department energy deposits are used to yield the original attributes of the particles. This is called an "inverse problem" and can be described by the equation

$$g(y) = \int A(y,x) f(x) \mathrm{d}x + b(y)\,, \tag{1}$$

where f(x) is the probability density function of interest, depending on the physical parameter x. The function g(y) is the probability density function of the measured parameter y and b(y) is the background. A(y,x) is the concolution core represented by the detector. To solve this problem, unfolding techniques are used.
In a discrete form, equation (1) can be written as

$$\boldsymbol{g} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{b}\,. \tag{2}$$

Here, $\boldsymbol{g}$ is the histogram of the estimated gamma-energies and $\boldsymbol{f}$ is the histogram of the true gamma-energies. $\boldsymbol{A}$ is the migration matrix constructed from the estimated and true gamma energies and $\boldsymbol{b}$ is the background derived from the off-regions.

1.3.1 Naive SVD unfolding

An easy method to invert equation (1), is to invert the matrix A. In the case of non-quadratic matrices, the Moore-Penrose-Pseudoinverse needs to be calculated using

$$\hat{\boldsymbol{f}} = \boldsymbol{A}^{+}(\boldsymbol{g} - \boldsymbol{b})\,. \tag{3}$$

This solution is equivalent to the method of least squares.

1.3.2 Poisson-likelihood-unfolding

If $\boldsymbol{g}$ follows a poisson distribution, a maximum-likelihood fit can be performed. The probability to measure g_i is

$$\mathrm{P}(g_i) = \mathcal{P}(g_i, \lambda_i) \tag{4}$$

with

$$\boldsymbol{\lambda} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{b} \tag{5}$$

minimizing the negative log-likelihood yields

$$-\ln(\mathcal{L}) = \sum_{i=1}^{M} = -g_i \ln(\lambda_i) + \lambda_i \tag{6}$$

the estimator for $\boldsymbol{f}$ is then

$$\hat{\boldsymbol{f}} = \mathrm{argmin}(-\ln(\mathcal{L}(\boldsymbol{f}|\boldsymbol{A}, \boldsymbol{g}, \boldsymbol{b})))\,. \tag{7}$$

2 Implementation

First a data selection is implemented including only events which have a probabilty of at least 80 % to originate from a gamma-ray. Only the events with an angular distance to the position of the source of $\theta^2 \leq 0{,}025$ are taken into consideration. Afterwards, a theta-squared plot is made and the detector significance is calculated. Then, the

migration matrix of the random-forest-regressor with the estimated and true gamma energies in an energy range from 500 GeV to 15 TeV is created.
Then, the unfolding for the crab nebula measurement with the naive SVD and the poisson-likelihood method is done. At last, the calculated flux for the unfolded data is compared to the HERA and MAGIC measurements.

3 Analysis

3.1 Theta-squared plot and detector significance

After applying the event selection descriped in the implementation, the on- and off-events are displayed to find a cut-off value. The theta-squared plot for the selected data as well as the cut-off is shown in Figure 3.

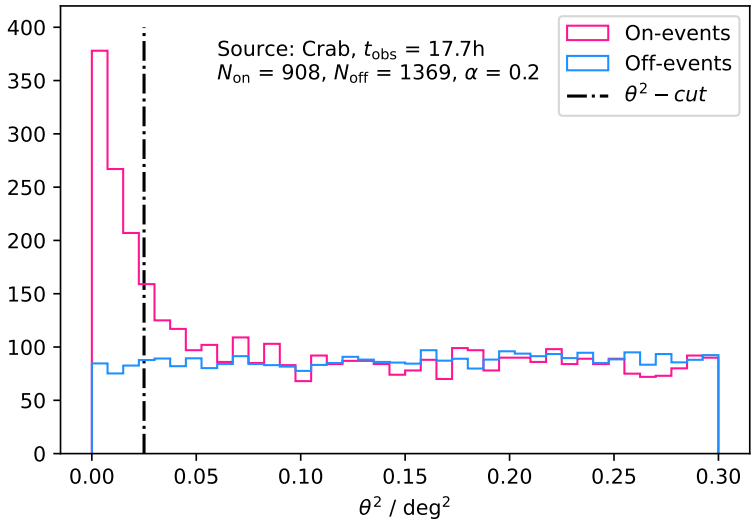


Figure 3: Theta-squared plot for on-region and off-region data.

It can be seen, that the off-region events are evenly distributed throughout the entire angular space and the on-region events peak at small distances to the source, as expected. The detector-significance for the given parameters in Figure 3 results in

$$S = 26.2759\,. \tag{8}$$

Further evaluation is needed to determine weather this value is adequate.

3.2 Migration matrix and unfolding

To calculate the migration matrix, the estimated values `gamma_energy_prediction` of the `gamma_test_dl3.hdf5` sample and the true-values `corsika_event_header_total_energy`

are used. The inverse problem is better conditioned if the binning is made with regards to the measured parameter, not in the unfolded. Therefore the bins range from 500 GeV to 15 TeV on a logarithmic scale. An extra bin at 50 TeV is established to catch events wth a higher energy than the upper threshold. This was done for both the estimated and the true values. The normed migration matrix is shown in Figure 4.
It can be seen, that the correlation is quite close to one.

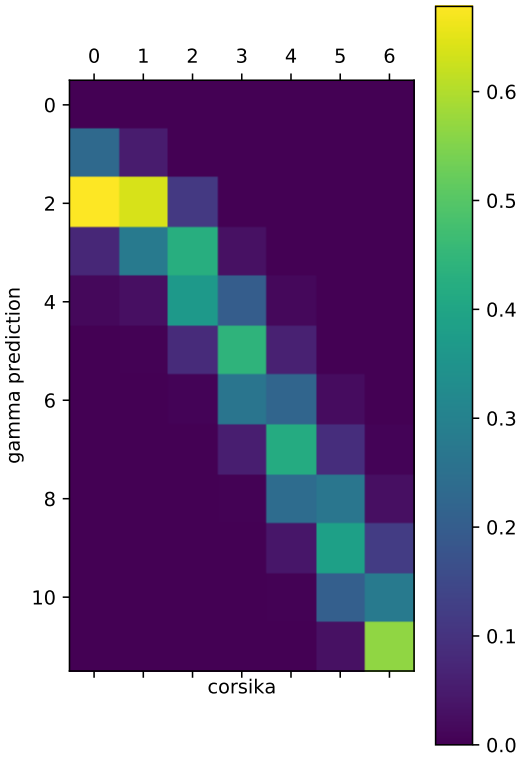


Figure 4: The migrationmatrix of the random-forest regressor.

In Figure 5 the energy distributions for both the on-region and the off-region are displayed. It can be seen, that very few events are more energetic than 10 TeV so the "overflow bin" at 50 TeV is not that relevant but not negligable either.

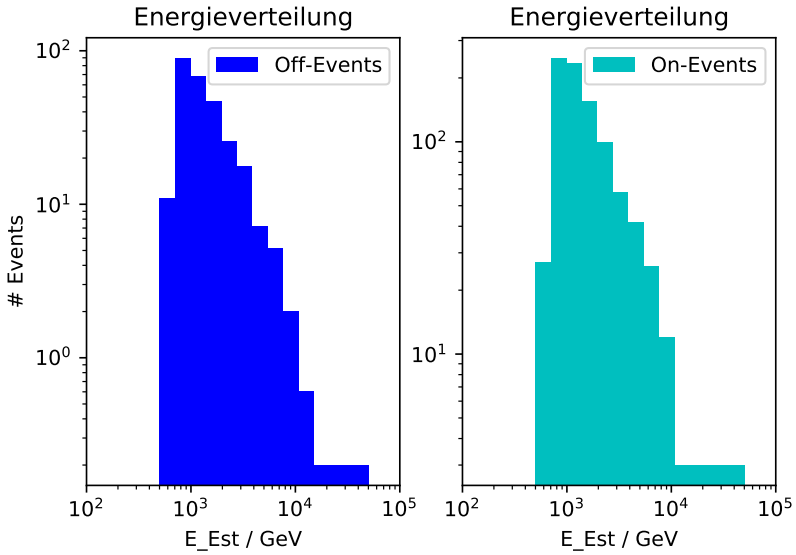


Figure 5: energy distributions for on- and off-regions.

3.2.1 Naive SVD unfolding

Following equation (3), the histograms for the background $\boldsymbol{b}$ is taken from the off-region events and the histogram for the estimated gamma-energies is yielded from the on-regions. With `scipy` and `uncertainties` the pseudoinverse matrix was calculated and the unfolded energies calculated as shown in figure 6.

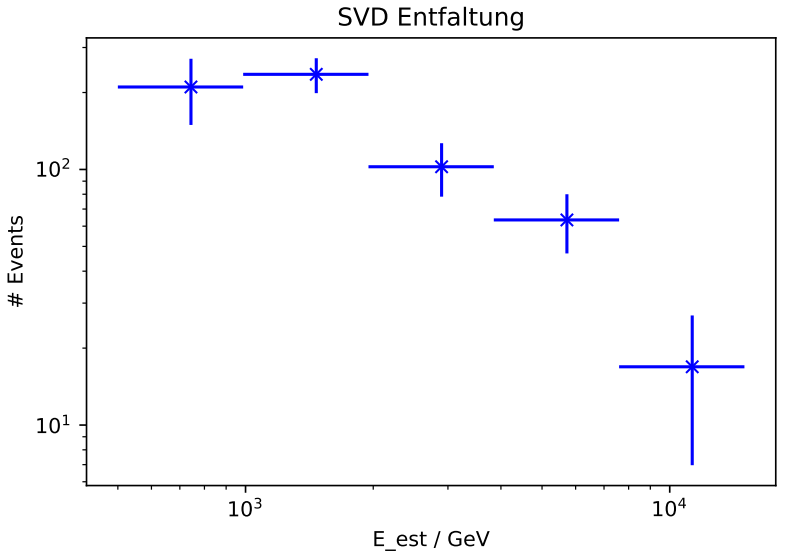


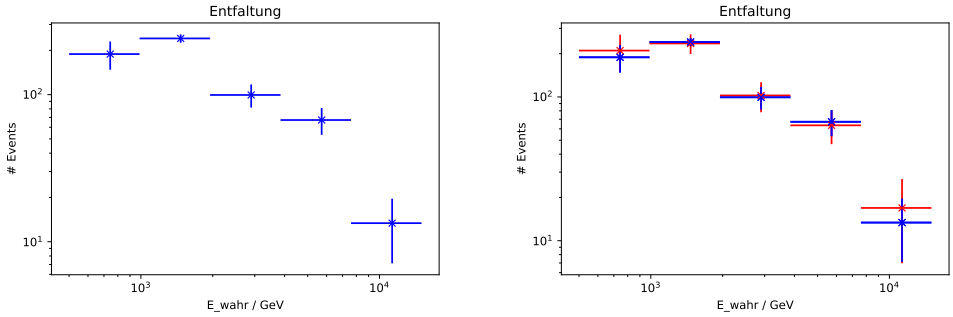
Figure 6: unfolded energies with naive SVD.

This unfolding method leads to an estimator for $\hat{\boldsymbol{f}}$ of:

$$\hat{\boldsymbol{f}} = \begin{pmatrix} 210.21 \pm 60.88 \\ 235.61 \pm 36.74 \\ 102.47 \pm 24.21 \\ 63.46 \pm 16.53 \\ 16.90 \pm 9.94 \end{pmatrix} \tag{9}$$

3.2.2 Poisson-likelihood unfolding

The poisson-likelihood unfolding was calculated using methods from `scipy`. The minimization was done using the `LBFGSB` minimization function from `scipyoptimize`.
In Figure 7b the poisson-likelihood unfolding is shown in blue and the naive SVD in red. They clearly show very similar results, therefore the flux can be calculated with unfolded data by any of these methods.



(a) Unfolded data using poisson-likelihood unfolding.

(b) Comparison between both unfolding techniques.

The poisson-likelihood method results in an esimator $\hat{\boldsymbol{f}}$ of

$$\hat{\boldsymbol{f}} = \begin{pmatrix} 188.75 \pm 40.98 \\ 241.08 \pm 15.70 \\ 99.47 \pm 17.74 \\ 67.21 \pm 13.92 \\ 13.37 \pm 6.24 \end{pmatrix} \tag{10}$$

3.3 Flux calculation

The flux is calculated with

$$\phi_i = \frac{\hat{f}_i}{A_{\text{eff},i} \Delta \text{E}_i t_{\text{obs}}} \tag{11}$$

where $\Delta \mathrm{E}_i$ is the width of the energy bin, t_{obs} is the duration of observation, $A_{\mathrm{eff},i}$ is calculated from

$$A_{\mathrm{eff},i} = \frac{N_{\mathrm{selec},i}}{N_{\mathrm{sim},i}} \cdot A\,. \tag{12}$$

The effective detector area is multiplied by 0.7 since only 70 % of the data is taken. The flux is calculated with the unfolded data both via both unfolding methods and shown in figure 8. Both true-energies are pretty close together.

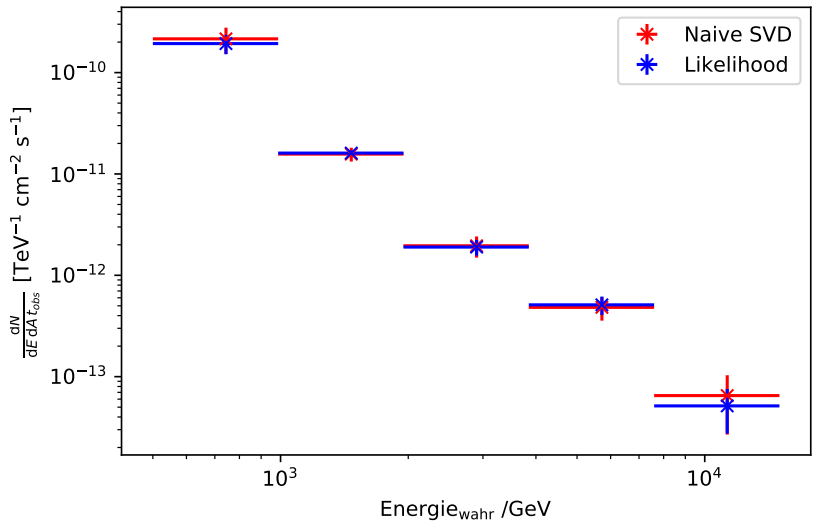


Figure 8: flux calculation with naive SVD and poisson-likelihood against true-energy.

3.4 Comparison to HEGRA and MAGIC

The fitting function used for the HEGRA and MAGIC data is defined as

$$f(x, a, b, c, d) = a \left(\frac{x}{b}\right)^{\left(-c + d \cdot \ln\left(\frac{x}{b}\right)\right)} \tag{13}$$

The parameters for both measurements are shown in table 1. The plot showing the comparison is shown in figure 9.

Table 1: Fitted values for the flux of the HEGRA [2] and MAGIC [3] experiment.

Experiment	$a \,/\, \frac{1}{\mathrm{TeV\,cm^2}} \times 10^{-11}$	b / TeV	c	d
MAGIC	$3{,}23 \pm 0{,}03$	1	$2{,}47 \pm 0{,}01$	$-0{,}24 \pm 0{,}01$
HEGRA	$2{,}83 \pm 0{,}6$	1	$2{,}62 \pm 0{,}05$	0 ± 0

Our analysis is compatible with the one of HEGRA and MAGIC for energies larger than 1 TeV. The first value is slightly higher than in the other measurements but the scale is

quite small, so is the error. Even for high energies our results are comparable to the ones from HEGRA. The flux calulated by MAGIC branches off of the HEGRA measurement but is still in agreement with our results.

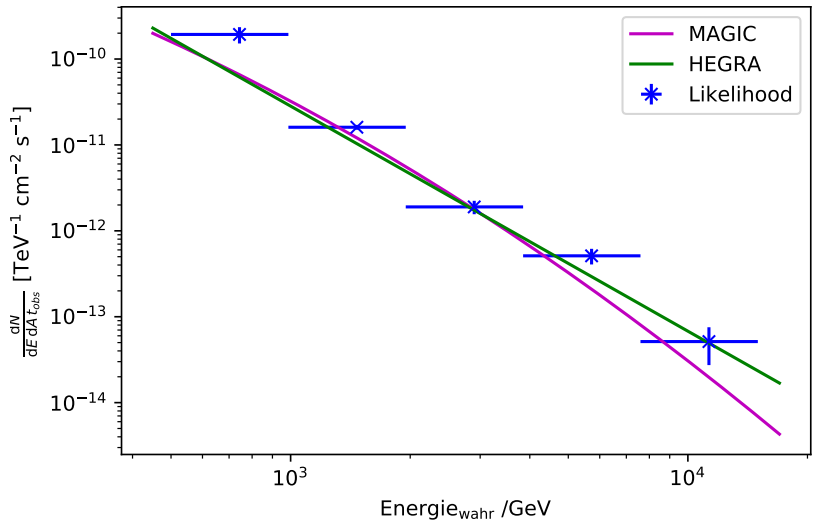


Figure 9: Comparison of our flux measurements with HEGRA and MAGIC.

4 Evaluation

The comparison of the unfolding methods Naive SVD and Poisson-Likelihood demonstrates that similar results can be achieved using either one. The flux calculation shows almost perfect agreement between the two methods in the first four bins. Only in the last bin, a larger uncertainty in both methods as well as a lager discrepancy compared to the other bins is noticable. This may be related to the fact, that this bin contains fewer entries and therefore rises larger uncertainties when calculating the error on the bin.
Comparing the unfolded data from this course to HERA and MAGIC results, very good agreement to the HERA results and good agreement to the MAGIC results can be found. Both curves lie well within the uncertainty of the unfolded data except for the first bin, for which MAGIC does no longer fit within the uncertainty. All in all though it can be said, that the unfolded data fits well with the results from the other collaborations.
Systematic uncertainties may arise from detector simulation or the cut on the theta-value, but seeing the agreement of the data to results from other collaborations, these uncertainties will be minor.

5 Anhang

References

[1] Astroparticle physics department TU Dortmund, *FP2 Teilchenphysik - Das Energiespektrum des Krebsnebens, gemessen mit FACT*

[2] F. Aharonian et al., *The Crab Nebula and Pulsar between 500 GeV and 80 TeV: Observations with the HEGRA stereoscopic air Cherenkov telescopes*, APJ 64 (2004) 897-913

[3] J. Aleksic et al., *Measurement of the Crab Nebula spectrum over three decades in energy with the MAGIC telescopes*, JHEAp 5-6 (2015) 30-38