

Dissertation zur Erlangung des akademischen  
Grades eines  
Doktor der Naturwissenschaften  
(Doktor rerum naturalium)

**On the Hunt for Photons:  
Analysis of Crab Nebula Data obtained  
by the First G-APD Cherenkov  
Telescope**

Thomas Fabian Temme  
geboren am 24.02.1987 in Dortmund

September 2016

Lehrstuhl für Experimentelle Physik V  
Fakultät Physik  
Technische Universität Dortmund

Erstgutachter: Prof. Dr. Dr. Wolfgang Rhode  
Zweitgutachter: Prof. Dr. Kevin Kröniger  
Abgabedatum: September 2016

## Abstract

In this thesis the analysis of data of the Crab Nebula obtained by the First G-APD Cherenkov Telescope (FACT) is presented. An analysis chain using modern machine learning methods for energy estimation and background suppression is developed. The generation and application of the machine learning models are validated using Monte Carlo simulated events. The Crab Nebula can be detected as a source of very high energy gamma rays with a significance of  $39.89\sigma$ . Its energy spectrum can be reconstructed between 250 GeV and 16 TeV.

The results of the analysis are used to evaluate the performance of the telescope. The energy bias and energy resolution are evaluated as well as the effective collection area and the sensitivity. The positive energy bias for energies below 1 TeV can be corrected by applying an unfolding method. For higher energies the bias is negligible. The energy resolution is about 22 % for most of the energy range. The effective collection area is monotonously increasing, reaching about  $3 \times 10^4 \text{ m}^2$  around 1 TeV. A sensitivity of 15.5 % of the flux of the Crab Nebula is calculated. The performance values are comparable to the values of current experiments. Taking into account the small reflector surface of FACT, the performance is very promising.

## Kurzfassung

In dieser Dissertation wird die Analyse von Daten des Krebsnebels, aufgenommen durch das First G-APD Cherenkov Telescope (FACT), präsentiert. Zu diesem Zweck wird eine Analyseketten entwickelt, die moderne Machine-Learning-Methoden zur Energieabschätzung und Hintergrundunterdrückung verwendet. Mit Hilfe von Monte-Carlo-simulierten Ereignissen wird die Generierung und Anwendung der Machine-Learning-Modelle validiert. Der Krebsnebel kann mit einer Signifikanz von  $39.89\sigma$  detektiert werden. Sein Energiespektrum kann zwischen 250 GeV und 16 TeV rekonstruiert werden.

Die Ergebnisse der Analyse werden verwendet, um die Performanz des Teleskopes zu evaluieren. Untersucht werden der Bias und die Auflösung der Energie, sowie die effektive Fläche und die Sensitivität. Der positive Bias für Energien unter 1 TeV kann mit Hilfe einer Entfaltungsmethode korrigiert werden. Für höhere Energien ist er vernachlässigbar. Die Energieauflösung beträgt ungefähr 22 % für den größten Teil des Energie Bereiches. Die effektive Fläche steigt monoton an und erreicht einen Wert von ungefähr  $3 \times 10^4 \text{ m}^2$  bei 1 TeV. Die Sensitivität beträgt 15.5 % des Flusses des Krebsnebels. Die Performanzwerte sind vergleichbar mit den Werten aktueller Experimente. In Anbetracht der kleinen Reflektoroberfläche von FACT ist die Performanz sehr vielversprechend.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Multi Messenger Astroparticle Physics</b>	<b>3</b>
2.1	Charged Cosmic Rays	4
2.1.1	Energy Spectrum of Charged Cosmic Rays	5
2.1.2	Air Showers Caused by Charged Cosmic Rays	6
2.2	Gamma Ray Astronomy	9
2.2.1	Air Showers Caused by Very High Energy Gamma Rays	10
2.2.2	Detection Principle of Imaging Air Cherenkov Telescopes	10
2.3	The Crab Nebula	13
<b>3</b>	<b>The First G-APD Cherenkov Telescope</b>	<b>15</b>
3.1	Introduction to FACT	15
3.2	The Camera of FACT	17
3.2.1	SiPMs as photo sensors for FACT	17
3.2.2	The Trigger and Data Acquisition of FACT	20
3.3	Wobble Observation Strategy	22
<b>4</b>	<b>Monte Carlo Simulations</b>	<b>25</b>
4.1	Simulation of IACT Data	25
4.2	Optical System Simulation	25
4.3	SiPM Simulation	26
4.4	Electronic Simulation	26
4.5	Current Development of the Simulation	27
<b>5</b>	<b>Machine Learning Based Data Analysis Chain</b>	<b>29</b>
5.1	Preprocessing Raw Data with FACT-Tools	29
5.1.1	Image Parameters	31
5.2	Machine Learning	33
5.2.1	Bootstrap Method	34
5.2.2	Cross Validation	34
5.3	Energy Estimation with Random Forest Regression	36
5.3.1	Random Forest Regression	36
5.3.2	Performance Evaluation	37

5.4	Background Suppression with Random Forest Classification . . . . .	40
5.4.1	Feature Selection . . . . .	40
5.4.2	Random Forest Classification . . . . .	41
5.4.3	Performance Evaluation . . . . .	42
5.5	Tikhonov Regularized Unfolding of an Energy Spectrum . . . . .	43
<b>6</b>	<b>Analysis of the Crab Nebula with FACT</b>	<b>47</b>
6.1	Data and Monte Carlo Sample . . . . .	47
6.1.1	Data Check . . . . .	48
6.1.2	Simulated Monte Carlo Events . . . . .	50
6.2	Preprocessing with FACT-Tools . . . . .	51
6.2.1	Parameter Distributions for Data and Monte Carlo Events . . . . .	51
6.2.2	Quality Cuts . . . . .	56
6.2.3	Splitting Gamma Data Set . . . . .	59
6.3	Energy Estimation with Random Forest Regression . . . . .	59
6.4	Background Suppression with Random Forest Classification . . . . .	64
6.4.1	Feature Selection . . . . .	64
6.4.2	Training and Validation of the Random Forest Classifier . . . . .	66
6.5	Source Detection of the Crab Nebula . . . . .	70
6.6	Unfolding of the Energy Spectrum of the Crab Nebula . . . . .	73
6.6.1	Preparation of the Unfolding . . . . .	73
6.6.2	TRUEE Test Mode . . . . .	77
6.6.3	TRUEE Pull Mode . . . . .	79
6.6.4	TRUEE Data Mode . . . . .	81
<b>7</b>	<b>Physics Performance of FACT</b>	<b>85</b>
7.1	Event Acceptance of Telescope and Analysis . . . . .	85
7.2	Effective Collection Area . . . . .	88
7.3	Energy Dependent Performance . . . . .	90
7.3.1	Energy Bias and Energy Resolution of FACT . . . . .	90
7.3.2	Energy Threshold . . . . .	93
7.3.3	Energy Spectrum of the Crab Nebula . . . . .	95
7.4	Sensitivity . . . . .	101
<b>8</b>	<b>Summary and Outlook</b>	<b>109</b>
<b>A</b>	<b>Data Check</b>	<b>113</b>
<b>B</b>	<b>Parameter Lists</b>	<b>123</b>
<b>C</b>	<b>Simulation configuration files</b>	<b>129</b>

**Bibliography**

**141**





# 1 Introduction

The first detection of a gravitational wave by the LIGO Observatory [2] on 14. September 2015 opened a brand new observation window in the field of astroparticle physics. Nevertheless the more established observation channels in astroparticle physics, the observation of charged cosmic rays, very high energy gamma rays and astrophysical neutrinos are as important as they were before. The combination of different messenger channels promises an exiting future for astroparticle physics. The current experiments reach sensitivities which enable them to challenge current source models and the next generation experiments like KM3Net or the Cherenkov Telescope Array (CTA) cast their shadows.

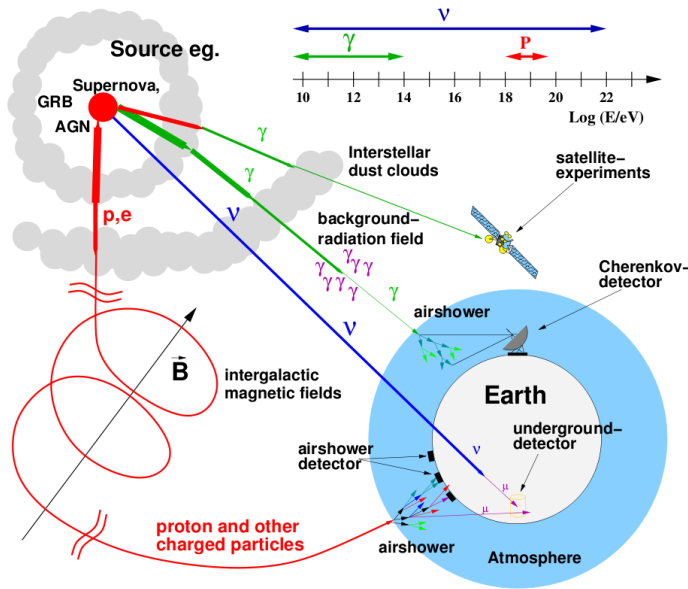
The usage of new photo sensor technology is quite common in the upcoming next generation project CTA. Several prototypes which are currently under development are using new silicon based photo sensors, called Silicon Photo Multipliers (SiPMs) in their cameras. The First G-APD Cherenkov Telescope (FACT) is the first Imaging Air Cherenkov Telescope (IACT) using SiPMs. It acts as a prototype for this sensors in Cherenkov astronomy. Hence the analysis of data obtained by FACT is quite important to establish silicon based photo sensor technology in Cherenkov astronomy. For the long-term monitoring of bright TeV blazars, the primary physics goal of FACT, it is necessary to investigate the performance of the telescope in respect to sensitivity, energy range and resolution.

This thesis presents the analysis of data from the Crab Nebula obtained by FACT. The Crab Nebula is considered to be a "standard candle" in Cherenkov astronomy, due to the high and constant flux of very high energy gamma rays emitted by the Crab Nebula. Hence it is an ideal source to investigate the performance of the telescope. Chapter 2 gives an overview of the current status of multi messenger astroparticle physics. The next chapter 3 gives an introduction to FACT. The Monte Carlo simulations used in the analysis are described in chapter 4. Chapter 5 depicts the machine learning based analysis chain which is used for the analysis of the Crab Nebula data. The analysis itself and its results are presented in chapter 6. These results are used to investigate the performance of FACT and the analysis chain in chapter 7. Chapter 8 sums up the results of this thesis and gives an outlook to future studies resulting from the work of this thesis.



## 2 Multi Messenger Astroparticle Physics

Nowadays there are four major observation windows in astroparticle physics, with individual advantages and disadvantages. Figure 2.1 gives an overview over the three major ones in the field of astroparticle physics. Their sources, the propagation through space and the different detection principles are shown.



**Figure 2.1:** Overview over the different messenger particles in the field of astroparticle physics. The galactic and extragalactic sources, the propagation through space and the different detection principles are shown [12].

Charged cosmic rays, first detected by Victor Hess in the year 1912 [34], are the most frequent ones. They also cover an extremely large energy range of 13 orders of magnitude between  $10^8$  eV and  $10^{21}$  eV [47]. It is not known which sources are able to accelerate cosmic rays to the highest energies. Due to their charge, they are deflected by galactic and intergalactic magnetic fields, losing all information about their origin. Hence the question of the origin of high energy cosmic rays could not be answered yet.

Very high energy gamma rays are not deflected by magnetic fields and therefore can be assigned to an astrophysical source. Unfortunately they are typically absorbed by gas and dust clouds surrounding the source. Only secondary gamma rays produced by the particles of the primary acceleration process interacting with the surrounding environment can be observed. Thus it is difficult to draw conclusions about the acceleration processes in the sources themselves. Also very high energy gamma rays interact with the extragalactic background light, causing a cutoff at the highest energies. The effect increases with the distance of the source.

The big advantage of astrophysical neutrinos is also the major challenge for detecting them. Due to their extremely small cross section they are not absorbed in the surrounding medium of the source or on the way to the earth. Hence conclusions on the acceleration processes in the sources can be drawn from observation of astrophysical neutrinos. But their small cross section also necessarily leads to very large instrumented volumes to observe a statistic significant numbers of neutrinos.

The newest messenger channel in astroparticle physics are gravitational waves. The first detection of a gravitational wave on 14. September 2015 [2] offers a new insight into unknown massive objects, like black holes with masses of about  $30 M_{\odot}$ . The next generation experiments for gravitational waves promise a totally new view on astrophysical phenomena and sources.

This thesis focuses on the analysis of very high gamma rays obtained with FACT. The largest background to this data are charged cosmic rays. The properties of charged cosmic rays are described in detail in section 2.1. Gamma ray astronomy is outlined in section 2.2. The observed source in the analysis performed in this thesis is the Crab Nebula. The Crab Nebula and its properties are described in section 2.3.

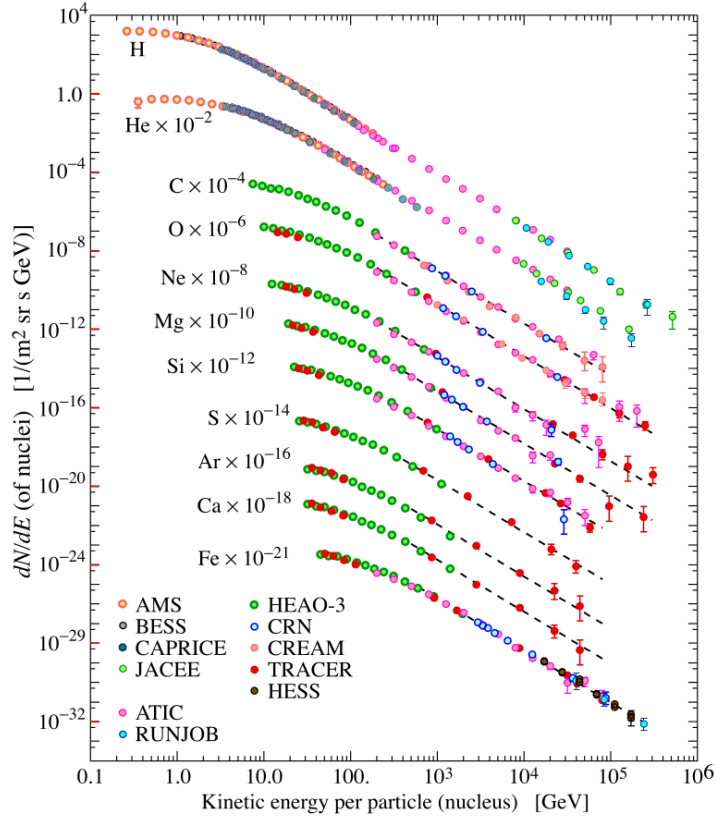
### 2.1 Charged Cosmic Rays

Charged cosmic rays are emitted in astrophysical sources and accelerated to the highest energies mankind ever measured. Up to energies of  $10^{18}$  eV the arrival direction of the particles is isotropic distributed over the sky. Due to their electric charge, the particles are deflected by galactic and intergalactic magnetic fields, making it impossible to infer the origin of the particles from their arrival direction. Only with energies from several EeV the deflection lowers to a few degrees (depending on the distance of the source). Unfortunately the flux at this high energies is drastically low, thus it is difficult to identify possible sources of the ultra high energetic cosmic rays (UHECR). Several tests for anisotropy in the arrival directions

of the UHECR were performed, but no statistically significant evidence could be found [1].

### 2.1.1 Energy Spectrum of Charged Cosmic Rays

The energy spectrum of the charged cosmic rays for energies between 0.1 GeV to  $10^6$  GeV is displayed in figure 2.2.

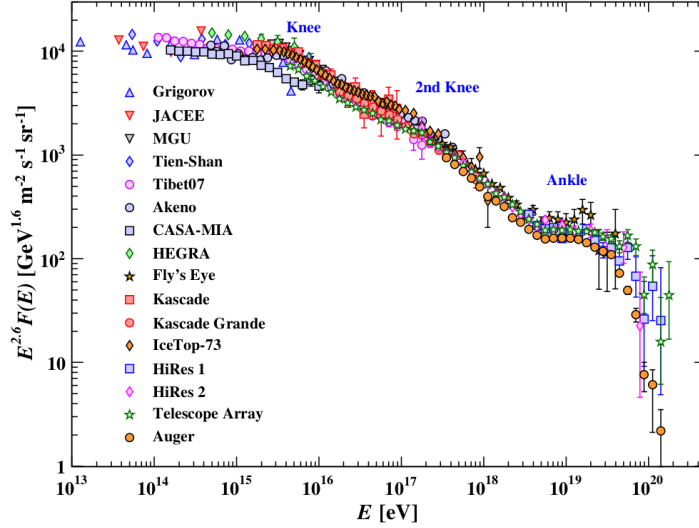


**Figure 2.2:** Energy spectrum of charged cosmic rays between 0.1 GeV to  $10^6$  GeV [47].

Over this energy range the differential flux  $\frac{dN}{dE}$  follows a power law [47]:

$$\frac{dN}{dE} \approx 1.8 \cdot 10^4 \left( \frac{E}{1 \text{ GeV}} \right)^{-2.7} \frac{1}{\text{m}^2 \text{ s sr GeV}} \quad (2.1)$$

For higher energies between  $10^{13}$  eV to  $10^{21}$  eV the spectrum is shown in figure 2.3. The flux is scaled with  $E^{2.6}$  to emphasize the different structures in the spectrum.



**Figure 2.3:** Energy spectrum of charged cosmic rays between  $10^{13}$  eV to  $10^{21}$  eV. The flux is scaled with  $E^{2.6}$  to emphasize the different structures in the spectrum [47].

The differential flux in this energy range also follows a power law, but with changing exponent. The points of changing exponent in the energy spectrum are referred as the first knee ( $4 \times 10^{15}$  eV), the second knee ( $8 \times 10^{16}$  eV) and the ankle ( $10^{18.5}$  eV) [47]. The changing exponent is often explained by different source populations committing to the corresponding energy ranges [47].

### 2.1.2 Air Showers Caused by Charged Cosmic Rays

When a charged cosmic ray particle enters the atmosphere it interacts with the nuclei of the molecules of the atmosphere. A number of baryons and mesons can be created as illustrated in equation 2.2.



The created baryons can again interact with other nuclei, creating more secondary particles according to (2.2). The created mesons, mostly pions, decay and produce muons and neutrinos ((2.3) and (2.4)) and gamma rays (2.5).

$$\pi^+ \rightarrow \mu^+ + \nu_\mu \quad (2.3)$$

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu \quad (2.4)$$

$$\pi^0 \rightarrow \gamma + \gamma \quad (2.5)$$

Some of the muons decay ((2.6) and (2.7)), whereas the gamma rays perform pair productions (2.8), both processes create electrons, positrons and neutrinos.

$$\mu^+ \rightarrow e^+ + \nu_e + \bar{\nu}_\mu \quad (2.6)$$

$$\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu \quad (2.7)$$

$$\gamma \rightarrow e^+ + e^- \quad (2.8)$$

The consecutive interactions of the secondary particles create a cascade of secondary particles (illustrated in figure 2.4) until the mean energy of the particles drops below the energy limits for the interactions.

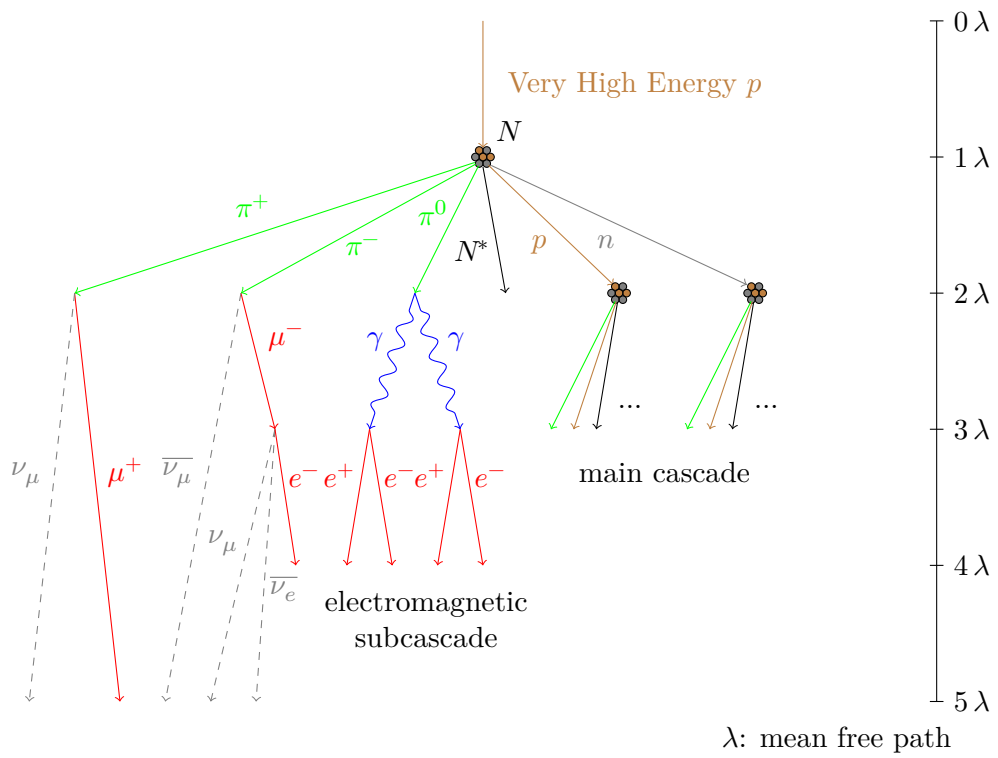
The secondary electrons, positrons and gammas create an electromagnetic subcascade, similar to the cascade caused by very high energy gamma rays, described in section 2.2.1.

The energy of the charged secondary particles is high enough that their velocity  $v$  surpass the speed of light  $c$  in the medium. Thus the particle electrically polarize the medium and a coherent shockwave of radiation is emitted. This effect is called Cherenkov radiation. The opening angle of the light cone in which the Cherenkov light is emitted depends on the refractive index  $n$  of the medium. For air under normal pressure is  $n = 1.00029$  [56]. The opening angle is:

$$\alpha = \arccos\left(\frac{1}{n\beta}\right) = 1.4^\circ \quad (2.9)$$

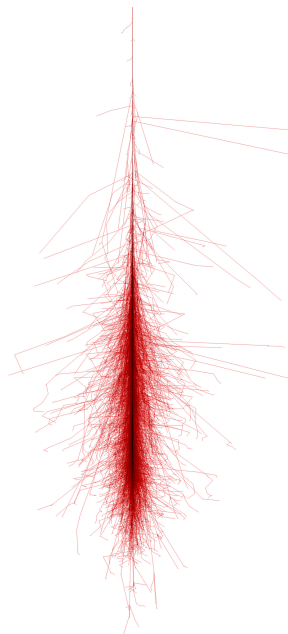
$$\beta = \frac{v}{c} = 1$$

The development of the shower is a stochastic process and figure 2.4 shows only a simplified image of the structure. The simulation of an air shower caused by a 1 TeV primary proton is displayed in figure 2.5.



**Figure 2.4:** Simplified image of the general structure of an air shower caused by charged cosmic rays [57].





**Figure 2.5:** Simulated air shower caused by a 1 TeV proton entering the atmosphere. The simulation was performed with CORSIKA [54].

## 2.2 Gamma Ray Astronomy

A flux of high and very high energy gamma rays is emitted by astrophysical sources and their surroundings. These gamma rays are absorbed by the atmosphere, making it impossible to detect them directly at the surface of the earth. Thus the observation of high energy gamma ray sources is mainly based on satellite experiments like the Fermi-LAT observatory.

Fermi-LAT was able to detect a various number of high energy gamma ray sources. The Fermi-LAT collaboration regularly publishes a catalog of the detected sources, the current Fermi Large Area Telescope Third Source Catalog [3] includes 3033 different sources in the energy range of 100 MeV to 300 GeV.

For the highest energies the flux of the sources decreases, thus the collection area of the satellite experiments is too small to collect a statistically sufficient number of very high energy gamma rays to measure the flux for energies above  $\approx 300$  GeV. For these energies the atmosphere can be used by Imaging Air Cherenkov Telescopes (IACTs) as a detector volume as described in section 2.2.2. This detection principle increases the effective collection area of IACTs for the highest energies enormously in comparison to the collection area of satellites. Thus IACTs are able to detect

fluxes of very high energy gamma ray emitting sources in the range of some GeV up to 100 TeV.

The first source detected by IACTs was the Crab Nebula [61] with the 10 m Whipple telescope in 1989. The first extra galactic sources detected by IACTs follow 3 years later with the detection of TeV photons from the active galaxy Markarian 421 [51]. Currently there are 134<sup>1</sup> sources of very high energy gamma rays known.

### 2.2.1 Air Showers Caused by Very High Energy Gamma Rays

A very high energy gamma ray entering the atmosphere will interact in the Coulomb field of a molecule of the atmosphere. An electron positron pair will be created by pair production (2.10). After propagating the mean free path the created electron and positron will interact again, emitting gamma rays by bremsstrahlung ((2.11) and (2.12)).

$$\gamma \rightarrow e^+ + e^- \quad (2.10)$$

$$e^+ + A \rightarrow e^{+'} + \gamma + A \quad (2.11)$$

$$e^- + A \rightarrow e^{-'} + \gamma + A \quad (2.12)$$

The energy of the photons is high enough to create additional electron positron pairs. The consecutive number of pair creation and bremsstrahlung interactions will create an electromagnetic air shower as it is sketched in figure 2.6. The development of the shower is a stochastic process and figure 2.6 shows only a simplified image of the principal structure. Figure 2.7 presents the simulation of an air shower caused by a 1 TeV primary photon.

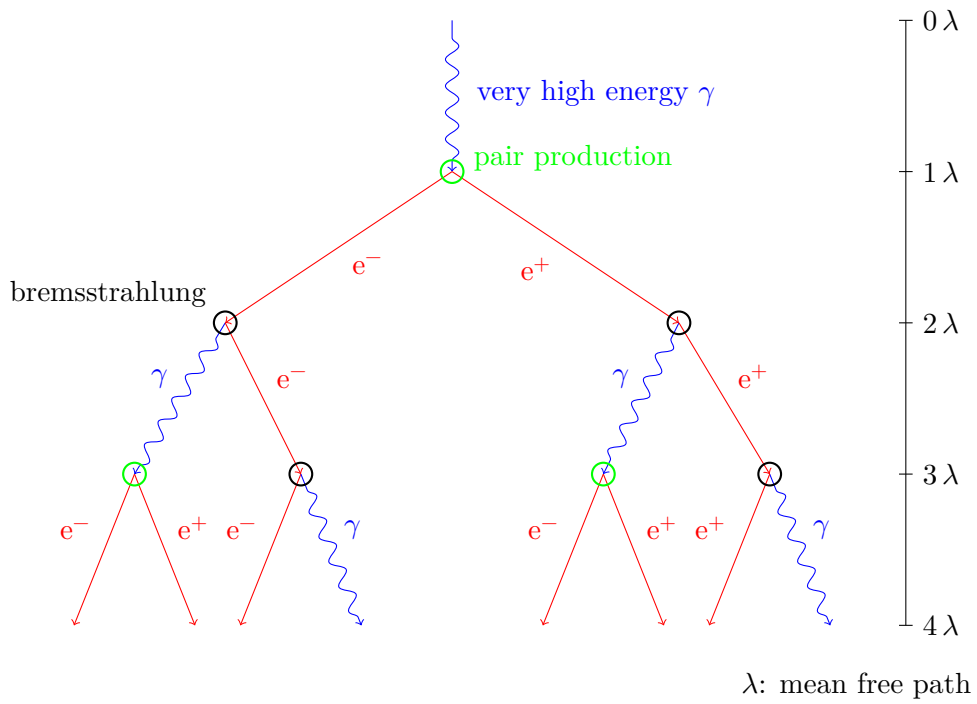
The charged secondary particles in the air shower (mostly electrons and positrons) have an energy high enough to emit Cherenkov light, as described in section 2.1.2.

### 2.2.2 Detection Principle of Imaging Air Cherenkov Telescopes

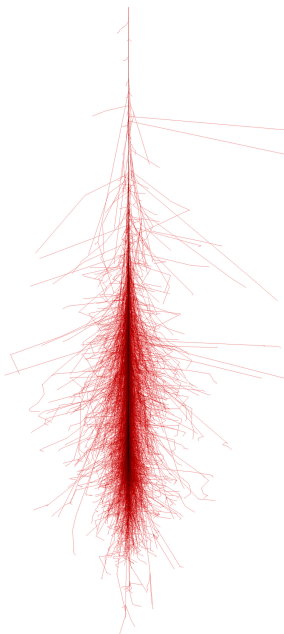
As described in section 2.2.1 very high energy gamma rays entering the atmosphere cause the creation of an air shower. The charged secondary particles in the air shower emit Cherenkov light. Ground based imaging air Cherenkov telescopes (IACTs) reflect the Cherenkov light of the air shower on a photon sensitive camera, taking an image of the air shower projected on the camera.

---

<sup>1</sup>Number taken from the default catalog of the TeVCat online source catalog [60], version 3.4



**Figure 2.6:** Simplified image of the general structure of an air shower caused by very high energy gamma ray [57].



**Figure 2.7:** Simulated air shower caused by a 1 TeV gamma entering the atmosphere. The simulation was performed with CORSIKA [54].

The origin of the primary particles can be reconstructed by performing a back projection of the image of the air shower. Thus enables the detection of very high energy gamma ray sources in the sky. The number of Cherenkov photons emitted in the air shower correlates to the energy of the primary particles. Hence also the energy spectrum of the gamma rays emitted can be reconstructed.

Also charged cosmic rays entering the atmosphere create air showers (see section 2.1.2). These air showers are the main background for IACTs. The ratio of charged cosmic ray particles to very high energy gamma rays is at least 1000:1, resulting in the requirement of a very accurate suppression of the background.

In principle the structure of charged cosmic ray air showers differs from the structure of very high energy gamma ray air showers. The transverse impulse of the secondary particles in cosmic ray air showers is typically larger than for very high energy gamma ray air showers. Thus the image of air showers are typically wider, the edge of the shower is fringed and cosmic ray air showers tend to create more separated islands in the camera plane.

Nevertheless, due to the finite resolution and noise effects, it is a challenging task to suppress the background of the charged cosmic ray air showers. Also the energy reconstruction is not straight forward, due to the large spread in the correlation

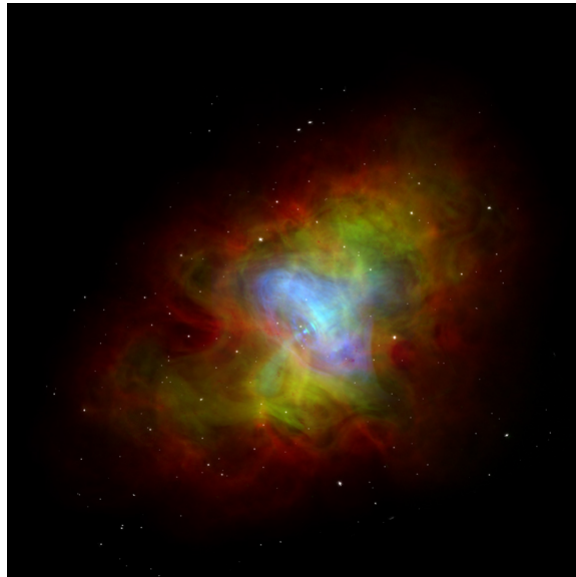
between the number of photons of the air shower and the energy of the primary particles.

Hence, the application of modern data analysis methods, like machine learning algorithms or regularized unfolding methods, offers a great potential to improve the performance of the background suppression and the reconstruction of the energy spectrum.

## 2.3 The Crab Nebula

The Crab Nebula is considered to be the "standard candle" in Gamma Ray Astronomy. The flux of the source in the very high energy gamma ray regime is constant, enabling different experiments to cross-calibrate the instruments by measuring the flux of the Crab Nebula and comparing the results. Thus the major goal of this thesis is the analysis of data from the Crab Nebula obtained by FACT and the investigation of the performance of the telescope in respect to sensitivity, energy range, energy bias and energy resolution.

The Crab Nebula is the remnant of a supernova explosion in the year 1054, recorded by the Chinese astrologer Yang Welt. Figure 2.8 shows a composite image of the Nebula with X-ray, optical and radio measurements.



**Figure 2.8:** Image of the Crab Nebula. The different observations are overlaid. X-Ray observation in blue, optical in green and radio in red [35].

The Nebula contains a pulsar, the Crab pulsar, in its center. The remaining of the exploded star from 1054 is the energy source of the system, emitting a relativistic flow of magnetized plasma in the surrounding shell. This pulsar wind interacts with ambient medium causing optical, radio, X-ray, gamma ray and very high energy gamma ray emission through synchrotron and inverse Compton radiation.

For more details about the current status of experimental observations and theoretical models describing the properties of the Crab Nebula and its pulsar see [22].

## 3 The First G-APD Cherenkov Telescope

### 3.1 Introduction to FACT

The three large Imaging Air Cherenkov Telescope (IACT) experiments MAGIC [58], Veritas [59] and H.E.S.S. [33] are currently dominating the field of gamma ray astronomy. The combination of stereoscopic systems (MAGIC: 2 telescopes, Veritas: 4 telescopes, H.E.S.S.: 5 telescopes) and the large reflective surfaces result in sensitivities outranging the sensitivities of the ancestor experiments (for example HEGRA [5]) and the ones of smaller experiments. All three experiments are using photo multiplier tubes as photo sensors.

Nevertheless, the development of the next generation experiment, the Cherenkov Telescope Array (CTA) [24] is ongoing since several years. The usage of new photo sensor technology, replacing the established photo multiplier tubes, is under consideration.

The First G-APD Cherenkov Telescope (FACT) [29] is the first IACT using Silicon Photo Multiplier Tubes (SiPMs) as photo sensors. Their properties, the electronic and mechanic robustness, the low bias voltage, the homogeneity and their independence from ambient light conditions highly qualify them for usage in gamma ray astronomy. Several prototypes of CTA telescopes are planned with SiPMs as photo sensors.

Thus the experience collected with the operation of FACT and the evaluation of the performance of the FACT telescope is important for the development of the next generation prototype telescopes of CTA. The investigation of the performance of the photo sensors [14] and several results of the long-term monitoring program of FACT [27, 19, 26, 13] already affirm the reliability of SiPMs for Cherenkov Astronomy. The evaluation of the physics performance of FACT is topic of the analysis described in this thesis. The results queue up with the previous results showing that SiPMs are a comparable alternative to photo multiplier tubes. The upcoming next generation of SiPMs with improved performance promises to surpass the performance of photo multiplier tubes.

FACT is located on the Roques de los Muchachos on the canary islands La Palma in Spain on a height of 2200 m above sea level. It was installed on the refurbished



**Figure 3.1:** Image of the First G-APD Cherenkov Telescope in park position during the day.

mount of the former HEGRA CT3 telescope. The reflector consists of 30 single mirrors summing up to a reflective surface of roughly  $9.5 \text{ m}^2$ . The single mirrors are arranged in the so called Davies-Cotton design, each individual mirror placed at its focal distance to the focal point in a sphere-like shape. In May 2014, the reflector was realigned and the arrangement changed to hybrid design between a sphere-like shape and a parabolic shape [45]. The data analyzed in this thesis, was obtained with the Davies-Cotton design.

The camera consists of 1440 pixels each a SiPM. A single SiPM is an array of 3600 solid state Geiger-mode Avalanche Photo Diodes (G-APDs). The camera and its photo sensors are described in more detail in section 3.2. Figure 3.1 shows the telescope in park position during the day.

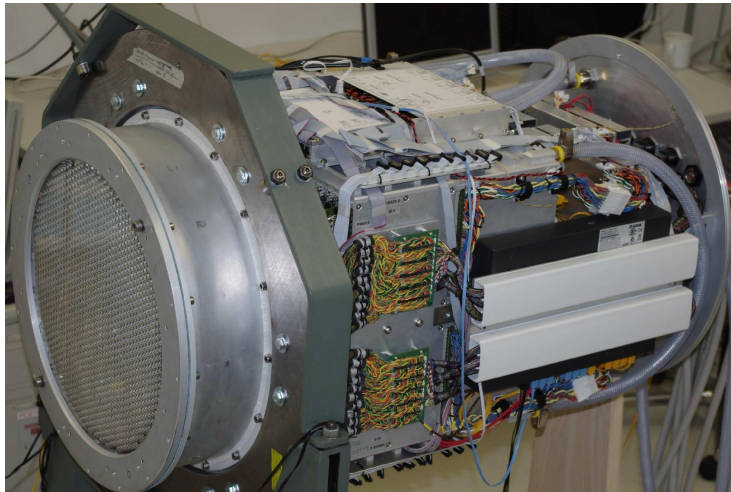
The operation of FACT is segmented in different time blocks, called runs. A standard data taking run with enable physics trigger has a duration of 5 min during the night and 1 min at the beginning and end of the night respectively. The data is taken with the Wobble observation strategy (see section 3.3). After 4 data runs, the Wobble pointing position is changed, pointing to the opposite site of the observed source, to delimit systematic effects. Beside the data runs, dedicated calibration runs are taken, to investigate the properties of the DRS4 data acquisition boards and the observation conditions.

For more details about the concept and design of FACT see [10].



## 3.2 The Camera of FACT

The camera of FACT consists of 1440 pixels. Each pixel combines one SiPM with one light concentrator glued to the SiPM. The edge length of the pixels is 9.5 mm with an opening angle of  $0.11^\circ$ . The field of view of the whole camera is  $4.5^\circ$ . All electronics except for the bias voltage supply are housed inside the camera compartment. This includes the SiPM sensors, the preamplifier and the digitization boards of the readout chain, the trigger unit and trigger master boards and the slow control board. Figure 3.2 shows the camera during its assembly- and test phase.

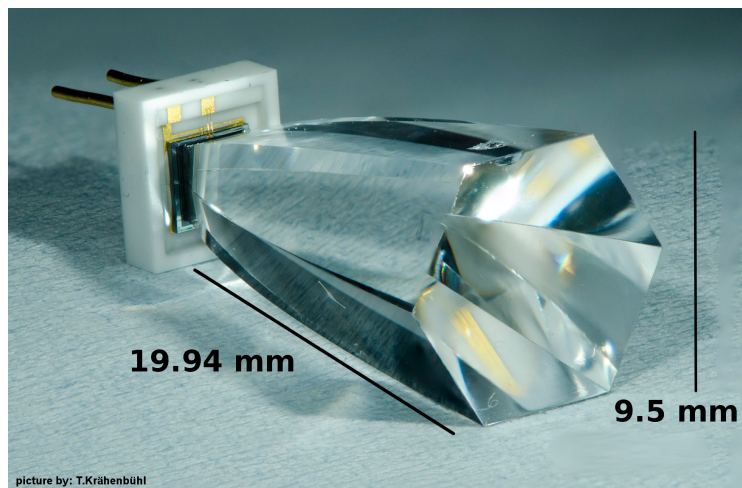


**Figure 3.2:** Image of the camera of FACT during assembly- and test phase. On the left the sensor compartment and the front window can be seen, on the right the electronic compartment with the integrated camera electronics [11].

### 3.2.1 SiPMs as photo sensors for FACT

SiPMs have several advantages making them suitable for the usage as photo sensors in Cherenkov astronomy. They have a considerable high gain of  $10^5$  to  $10^6$ . They are capable of single photon counting, are electronically and mechanically robust and independent of ambient light conditions. The low bias voltage needed allows for an easy design and operation. The possibility of mass production of SiPMs results in a high homogeneity. The photon detection efficiency is comparable to the one of current photo multiplier tubes and is further improved with new generation of SiPMs developed by the manufacturers. Since start of operation of FACT in October 2011, no aging effect occurred for any of the SiPMs. Thus they are also

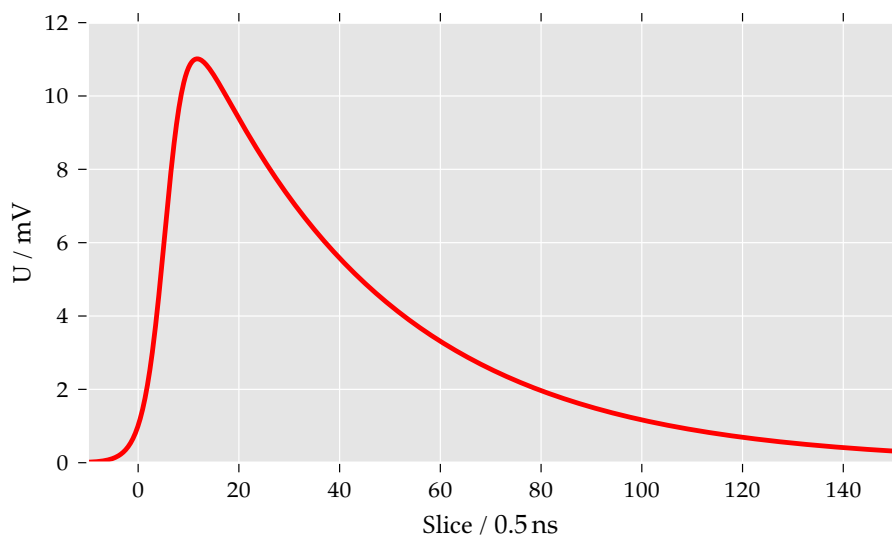
suited for stable and robust long-term operations [52]. Figure 3.3 shows an image of one SiPM and a light concentrator glued to the SiPM.



**Figure 3.3:** Image of a SiPM and a light concentrator glued to the SiPM. Picture taken by Thomas Krähenbühl.

A SiPM is a silicon based photon sensor, consisting of an array of solid state Geiger-mode Avalanche Photo Diodes (G-APDs). A photon entering an individual G-APD can create an electron-hole pair. Due to the bias voltage being operated above the breakdown voltage of the G-APD, the electron-hole pair causes an avalanche of further electron-hole pairs with a gain of about  $10^5$  to  $10^6$ . The individual G-APD is completely discharged (called breakdown) resulting in a constant signal for every photon independent of the incident angle or the energy of the photon. As the G-APD array of the SiPM is operated in parallel, the signal in the SiPM is the sum of the signals of all G-APDs and is linear dependent on the number of triggered G-APD cells, thus the number of registered photons. After the breakdown of an individual G-APD the cell has a short dead time, when it is not able to release another avalanche. For a longer recovery time the cell is recharged, resulting in a lower signal in case of an additional break down. The signal of a single photon in a SiPM has a fast rising edge with a long recovering tail. Figure 3.4 shows the shape of a single photon pulse in a SiPM. The function describing the pulse shape was evaluated by stacking the measurements of a large number of single photons in dark count measurements [14].

There are several effects occurring in SiPMs, which should be discussed. Due to thermal excitation of electron-hole pairs dark counts are created, faking a false photon signal. The rate of the dark counts is typically below 5 MHz, thus below



**Figure 3.4:** Pulse shape caused by a single photon in a SiPM. The fast rising edge when the G-APD is breaking down and the long tail when the G-APD is recovering are visible. The function displayed was evaluated by stacking the signal of single photons in dark count measurements [14].

the rate of photons of the night sky background [44]. Dark counts and night sky background photons add up to the photon noise in the camera.

An avalanche in a G-APD can cause a creation of a photon triggering a neighboring cell. This effect is called crosstalk and causes a higher signal in the SiPM, undistinguishable from an additional photon registered in the sensor. Crosstalk increases the effective gain of a single photon on average, but smears out the resolution for individual photons.

Free charge carriers can be trapped in the G-APD, released with a time delay, causing an additional avalanche delayed in time. These afterpulses occur often at the trailing edge of the original pulse. Their height depends on the time difference between original avalanche and afterpulse, due to the recovery time of the G-APD. Afterpulses add up to noise effects in the pixels and smear out the single photon resolution.

The SiPMs used for the FACT camera are Hamamatsu MPPC S10362-33-50C. They have an active area of  $3\text{ mm} \times 3\text{ mm}$  and consist of 3600 G-APDs.

### 3.2.2 The Trigger and Data Acquisition of FACT

The signals of the SiPMs are processed directly in the camera compartment. First step is the amplifying of the analog signals using a preamplifier board. The signal is then fit in a trigger chain and a digitization chain. For the trigger chain the signals of nine pixels are summed up to a patch signal. Then a cable based clipping is applied to the patch signal, cutting the length of the signal to about 10 ns. A comparator applies the trigger threshold to the patch signal and the comparator signal is fed into one of 40 trigger units. At each trigger unit a N-out-of-4 logic is applied to the signals of 4 patches. The results of the trigger units are compiled into the trigger master, where a N-out-of-40 logic is applied. During standard datataking both N are set to one, resulting that one patch is needed to exceed the trigger threshold, that a trigger signal is raised.

The analog signals in the digitization chain are fed to a so called DRS4 board [53]. These domino ring sampling chips sample the signal with 2 GHz onto an array of 1024 individual cells. In case of a raised trigger signal a subset of 300 DRS4 cells are digitized.

The resulting raw data, the voltage curve of one pixel of one event of the telescope is shown in figure 3.5.

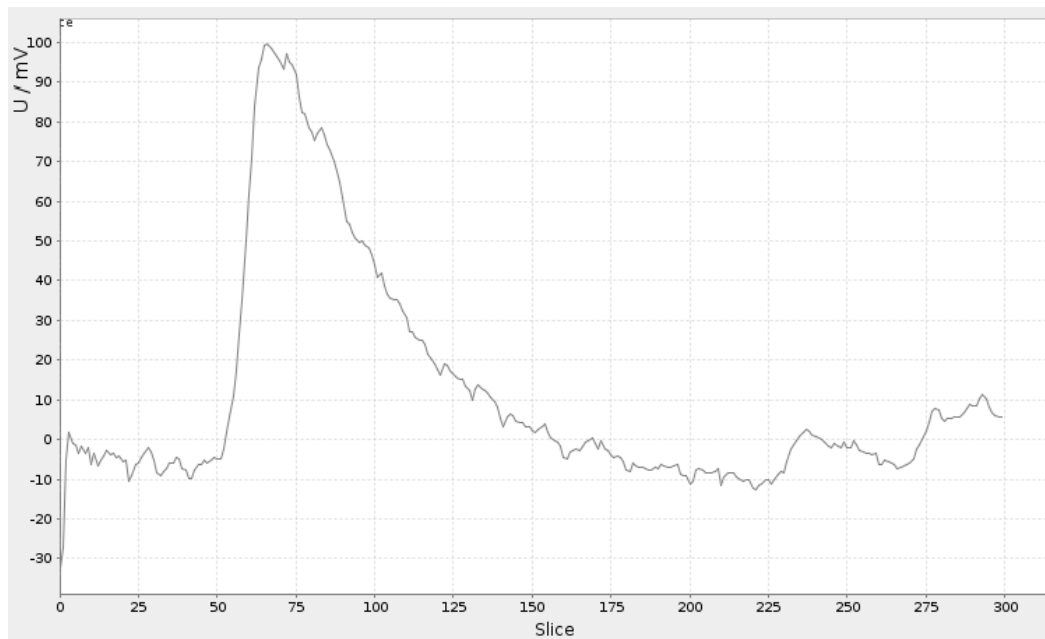
The pulse shape of about 9 to 10 Cherenkov photons is visible. There are some effects adding noise to the raw data.

There is electronically noise on the raw data.

Each DRS4 cell response individually, yielding the necessity to calibrate the data.

The DRS4 can also cause electronic artifacts like spikes and jumps in the voltage curves.

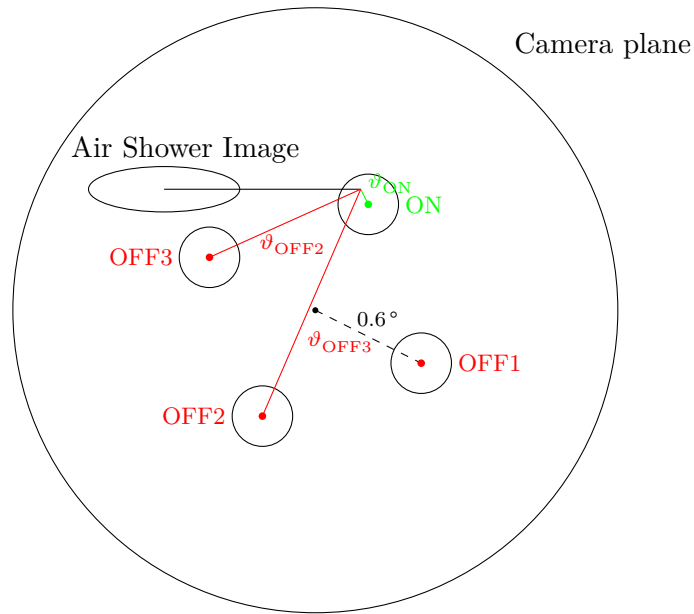
The resulting digitized values are stored in a FITS file [49]. A newly developed compression method is applied to the output files [6].



**Figure 3.5:** Raw data of one event in one pixel. The image is taken from the graphical user interface implemented in FACT-Tools. The voltage curve over the 300 slices (corresponds to 150 ns) sampled by the DRS4 board is shown. The pulse visible is caused by the detection of about 9 to 10 Cherenkov photons.

### 3.3 Wobble Observation Strategy

To obtain measurements of a possible source of very high energy gamma rays and measurements of the corresponding background in parallel the telescope is operated with the so called Wobble observation strategy. Thereby the telescope points  $0.6^\circ$  asides the position of the possible source. The position of the possible source region (called ON region) is now located outside the camera center in the camera plane. Regions for measurements of the background (called OFF regions) can be defined symmetrically to the ON region around the center of the camera. Figure 3.6 visualize the definition of ON and 3 OFF regions for a telescope operated in Wobble mode.



**Figure 3.6:** Illustration of the wobble strategy while observing a source of very high energy gamma rays with an IACT. The ON and 3 OFF regions for signal and background data sets are shown. In addition an image of an air shower and the calculation of  $\vartheta$  for ON and OFF regions is sketched.

For the analysis described in this thesis, 5 OFF regions are defined. For each obtained event the source dependent parameters ( $\alpha$ ,  $\vartheta$ , Distance, CosDeltaAlpha) are calculated for each of the 6 target regions (1 ON, 5 OFF) separately. This is also sketched in figure 3.6. The random forest classifier for the background suppression (see section 5.4) is applied 6 times, with the source dependent parameters changed for each target regions.

The ON and OFF data sets are created by applying a cut in the corresponding confidence. An additional cut in  $\vartheta^2 < 0.2^{\circ 2}$  ensures that no event occurs both in the ON and the OFF set.

The ON data set (also called signal region) contains events from the possible source (called excess events  $N_{\text{exc}}$ ) and background events. The OFF data set is a direct measurement of the distribution of these background events. Therefore for large values of  $\vartheta$  and small confidence values the distributions of the ON and OFF data sets should agree with each other.





## 4 Monte Carlo Simulations

### 4.1 Simulation of IACT Data

For the simulation of air showers caused by cosmic ray particles or very high energy gamma ray particles the shower simulation software CORSIKA (COsmic Ray SIMulations for KAscade) [32] is used. It was originally developed to perform simulations for the extensive air shower experiment KASCADE [37]. Nowadays it is widely used by the whole astroparticle physics community for simulations of extensive air showers initiated by high energy cosmic ray particles. Primary particles can be protons, nuclei up to iron, photons and several more particles. Energies up to  $10^{20}$  eV can be simulated.

CORSIKA tracks the propagation of the primary particle through the atmosphere until it interacts with atmosphere nuclei or decays. The induced air shower with all secondary particles is simulated. The interactions described in section 2.1.2 and section 2.2.1 are simulated. The generation of Cherenkov light is performed and the resulting photons and their arrival information on the ground are stored for the next simulation step.

In the analysis described in this thesis the low energy hadronic interactions are simulated by the interaction package FLUKA [17].

The simulation of the reflector and the camera is performed by the program Ceres within the analysis framework MarsCheobs [20]. It simulates the reflection of the photons on the mirrors, the response of the photo sensors, the behavior of the camera electronics and of the data acquisition system. The resulting output of Ceres are simulated raw data files, similar to the one created by the telescope.

The number of simulated events and the settings used for the simulation in this analysis can be found in section 6.1.2.

### 4.2 Optical System Simulation

The reflector of the FACT telescope is simulated in Ceres. The position and orientation of the single mirrors are obtained from measurements of the mirror

positions and orientations at the real telescope [45].

The reflectivity in dependency of the wavelength of the Cherenkov photos was measured [10]. It is used as a probability function to reject photons hitting the reflector. To correct not simulated higher order optical effects and possible microroughness of the single mirrors, the impact point of the photons on the camera plane after the reflection is smeared with a gaussian function. The standard deviation  $\sigma$  of the gaussian function is called single mirror point spread function.

For photons hitting the camera, the angular dependent transmission of the light concentrators is applied, rejecting photons not going through the light concentrators.

### 4.3 SiPM Simulation

Photons hitting the camera plane and going through the light concentrators are collected in the array of the photo sensors, the SiPMs. The simulation of the SiPMs is microscopic, thus the status of every single cell of each SiPM is evaluated and stored in the simulation. The functionality of the SiPMs is described in section 3.2.1.

In addition to the Cherenkov photons from the air shower simulated in CORSIKA also photons of the night sky background (NSB) and dark counts are generated. NSB photons and dark counts are distributed uniform in time. For simplicity dark counts are also count as (background) photons in the following.

For each photon it is checked if a G-APD cell is triggered, using the wavelength depending photon detection efficiency. The cell number is diced, and the height of the signal, taking dead time and recovery time into account, is simulated. Also possible crosstalk and after pulse signals are simulated. The resulting list of signals in the SiPM is then fed into the electronics simulation (see section 4.4).

### 4.4 Electronic Simulation

Each signal, caused by a photon is inducing a pulse with the shape of a single photon response in the simulated analog signal of the pixel. The pulse shape was evaluated and parameterized using dark count measurements in the real detector [14]. Figure 3.4 shows the parameterized pulse shape of a single photon response. Electronic noise is added to the analog signal. The trigger decision is simulated like in the real telescope. 9 SiPM signals are added to one trigger signal. The trigger signal is clipped and a comparator with the height of the trigger threshold is applied.

Then a N-out-of-4 logic and a N-out-of-40 logic is simulated. If a trigger occurs the digitization of all analog SiPM signals of the camera is simulated. The results of the electronic simulation is stored in FITS files [49], similar to the ones of the telescope.

### 4.5 Current Development of the Simulation

Currently there are several points of the simulation under investigation. Unfortunately not every parameter in the simulation can be measured independently at the telescope. Some can only be measured with limited accuracy, some can not be measured at all.

The current way of improvement of the simulation is to check the agreement of several parameter distributions for different measurement channels. The following channels are currently under investigation:

**Dark Count Measurements:** Give insight into the properties of the SiPMs in the real telescope and the simulation. Also the properties of the real and simulated electronics are influencing these measurements.

**Open Shutter Measurements:** Are more data like measurements. They give insight into the properties of the SiPMs if they are exposed to Night Sky Background.

**Mirrors:** Direct measurements of the light distributions of the whole reflector and of single mirrors on the camera plane [45] can be compared to the simulated light distributions.

**Muons:** The extraction and analysis of Muon events [46] provides an insight into the timing properties and the optical properties of the real and simulated telescope.

**Cosmic Ray Air Showers:** The parameter distributions of the air showers induced by cosmic rays can also be investigated. The dependency of the image parameters to the simulation parameters cannot directly be concluded, but simulation parameters can be varied and the resulting image parameter distributions can be compared.

The simulation has to be tuned in a way that the image parameter distributions of simulated and real cosmic ray particles agree with each other. There was made

large progress in the improvement of the simulation. The distributions of image parameters fit for a large number of parameters over a large range. Nevertheless there are some mismatches which should be further investigated. The parameter distributions and the mismatches are shown and discussed in section 6.2.1.

Possible sources of mismatches are:

**Noise in the Voltage Curves:** Currently only white electronic noise with a constant standard deviation for all pixels is added in the simulation. There could be several different noise effects in the voltage curves of the pixels in the real telescope.

**Mirror simulation:** The light distribution of the whole reflector on the camera plane only agrees in a first order. As the simulation of the reflector is microscopic, though the mirrors are simulated individually, the parameters for each single mirror has to be adapted (for example the single mirror point spread function). Unfortunately finer measurements of the light distribution exist only for the whole reflector, thus it is difficult to adapt the simulation to fit this measurement.

**Timing behavior of the SiPMs:** The variance of the arrival times of the Cherenkov photons of both muon and air shower events differ for the simulation and the real telescope. A first smearing of the arrival time in the simulation was added to let the distributions agree in first order. Further studies have to be made to further increase the agreement.

## 5 Machine Learning Based Data Analysis Chain

The analysis chain used in the analysis described in this thesis is based on the application of machine learning algorithms for several steps of the analysis. First step of the analysis is the preprocessing of the obtained data from the telescope and the camera simulation. The images of the air showers have to be cleared from electronic effects and non Cherenkov photons. The cleaned images are then parameterized, combining the information value in the event to a set of parameters. These preprocessing steps are performed using the software FACT-Tools and are described in section 5.1.

For the next analysis steps machine learning algorithms are used. They are trained using Monte Carlo simulated events and then applied to real data events. By using independent Monte Carlo simulated events as a pseudo data test set, the performance of the algorithms can be evaluated and validated. Section 5.2 outlines the principle of machine learning methods and methods commonly used for their validation.

For the estimation of the primary particles energy a random forest regressor is trained using the image parameters calculated by FACT-Tools. The image parameters are also used by a random forest classifier to suppress the background of cosmic ray events in the event sample. Section 5.3 and section 5.4 respectively explain the machine learning algorithms used for energy estimation and background suppression and the variables commonly used to describe the performance of the algorithms.

The reconstruction of the energy spectrum of the observed source with the unfolding software TRUEE is explained in section 5.5.

### 5.1 Preprocessing Raw Data with FACT-Tools

The preprocessing software FACT-Tools [30] is an extension of the `streams`-framework [15] developed within the scope of the Collaborative Research Center 878 at the TU Dortmund. `streams` provides an xml-based abstraction layer for streaming data analysis processes and enables rapid prototyping. It can be executed on big data stream applications like Spark [63] or Hadoop [62].

The FACT-Tools extend the `streams` library by a number of processors capable of performing the preprocessing analysis of FACT data. The processors are based on the algorithms implemented in PARFACT [57] and are constantly improved within the work of the C3 project of the Collaborative Research Center 876 [23].

For more details about FACT-Tools see [21] and [16]. The different preprocessing steps can be described as follows:

**Calibration:** The raw data of the telescope contains several noise effects of the electronics for which they have to be calibrated. These effects are:

- Different responses (offset, gain and time) of the cells of the DRS4 chips.

- 1 to 4 slices long spikes in the voltage curve.

- Jumps in the voltage curve.

- Time offsets between different pixels.

The different effects are described in [57] and [46].

**Extraction:** The number of Cherenkov photons and their arrival time is extracted from the raw data. For this the peak in the voltage curve is searched for. The integral around the peak is used as an estimator for the number of Cherenkov photons, whereas a fit of a polynomial of third order to the rising edge of the peak is used to estimate the arrival time.

**Cleaning:** The image of the air shower has to be cleaned from night sky background photons. Hence a two-level cleaning with an additional time cleaning is applied to the data. Signal pixels belonging to the air shower are selected, the rest of the pixels are discarded. The two-level cleaning is described in [57]. The time cleaning applied in this analysis is a variant of the one described in [57]. Instead of the difference in time between the mean of all pixels in the air shower to each individual pixel, the pixels selection is based on the difference in time between neighboring pixels in the image of the shower.

**Calculation of Parameters:** The cleaned image of the air shower is parameterized, combining the information value of the air shower image into an image parameter set. The basic concept of the parameterization is the representation of the air shower image with an elliptical shape. This parameterization was introduced by A. M. Hillas in 1985 [36], several other parameters are calculated additionally. The most important image parameters are described in more detail in the following section.

### 5.1.1 Image Parameters

The air shower of a very high energy gamma ray develops into an ellipsoid form due to the fact that the interactions in the shower are forward boosted and particles along the main shower axis have the longest travel distance in the atmosphere. Hence the projection of the image of the air shower in the camera plane has an ellipsoid shape and can be parameterized by the second moments along the two main axes and the orientation angle of the major axis. This parameterization was introduced by A. M. Hillas in 1985 [36], the parameters are called LENGTH for the major axis, WIDTH for the minor axis and DELTA for the orientation angle.

Another important parameter is the sum of all photons in the shower, called SIZE. It strongly correlates with the energy of the primary particle.

A shower with separated islands of pixels in the camera plane is characterized over the NUMBER OF ISLANDS parameter. Typically cosmic ray induced showers tend to have more islands than gamma ray induced showers, due to the larger transverse impulse of the secondary particles.

To characterize air showers collected at the edge of the camera plane, the parameters LEAKAGE and LEAKAGE2 are introduced. LEAKAGE is the ratio between the number of photons in the outer most pixels of the camera and the SIZE of the shower. Similarly LEAKAGE2 is the ratio between the number of photons in the two outer pixels rings and the SIZE of the shower. If air showers are fully contained in the camera LEAKAGE has a value of zero, if air showers consist only of pixels at the edge of the camera LEAKAGE has a value of one.

The ratio between the number of photons in the pixels within the ellipse and the SIZE of the shower is called CONCENTRATION CORE.

The standard deviation over the arrival times ( $\sigma_{\text{arrival time}}$ ) of the Cherenkov photons over the pixels of the shower describes the timing profile of the shower. It is limited by the time resolution of the telescope (see [46] for an investigation of the time resolution).

The reconstruction of the arrival direction of the shower is done with the DISP [39] method. Therefore a reconstructed source position of the shower is assumed to be on the major axis of the shower. The distance DISP between the center of gravity of the shower and the reconstructed source position is calculated by a parameterization using WIDTH and LENGTH:

$$\text{DISP} = c_0 \cdot \left(1 - \frac{\text{WIDTH}}{\text{LENGTH}}\right) \quad (5.1)$$

with the parameter  $c_0$  calculated using gamma Monte Carlo simulated events. The distance in the camera plane between the reconstructed source position and a point of interest  $i$  is called  $\vartheta_i$ . The points of interest are the location of an assumed source of very high energy gamma rays, the ON-position ( $i = \text{ON}$ ) or a background region with no very high energy gamma ray source located, the OFF-position ( $i = \text{OFF}$ ). See section 3.3 for a description of the wobble observation strategy which allows to determine directly parameters for ON and OFF positions.

Figure 5.1 illustrates the definition of the image parameters LENGTH, WIDTH, DELTA, NUMBER OF ISLANDS, DISP and  $\vartheta$ . The obtained number of photons in the pixels are shown color coded. With the set of selected shower pixels  $S$ , the set of pixels within the ellipse  $S_E$ , the one and two outer pixels rings  $L1$  and  $L2$  and the estimated number of Cherenkov photons  $N_{\text{pho},i}$  and their arrival time  $t_{\text{arr},i}$  the parameters SIZE, LEAKAGE, LEAKAGE2, CONCENTRATION CORE and  $\sigma_{\text{arrival time}}$  are defined as:

$$\text{SIZE} = \sum_{i \in S} N_{\text{pho},i} \quad (5.2)$$

$$\text{LEAKAGE} = \frac{1}{\text{SIZE}} \sum_{i \in (L1 \cap S)} N_{\text{pho},i} \quad (5.3)$$

$$\text{LEAKAGE2} = \frac{1}{\text{SIZE}} \sum_{i \in (L2 \cap S)} N_{\text{pho},i} \quad (5.4)$$

$$\text{CONCENTRATION CORE} = \frac{1}{\text{SIZE}} \sum_{i \in S_E} N_{\text{pho},i} \quad (5.5)$$

$$\sigma_{\text{arrival time}} = \sum_{i \in S} (t_{\text{arr},i} - \bar{t}_{\text{arr}})^2 \quad (5.6)$$

A complete list of parameters calculated by FACT-Tools can be found in the appendix B.



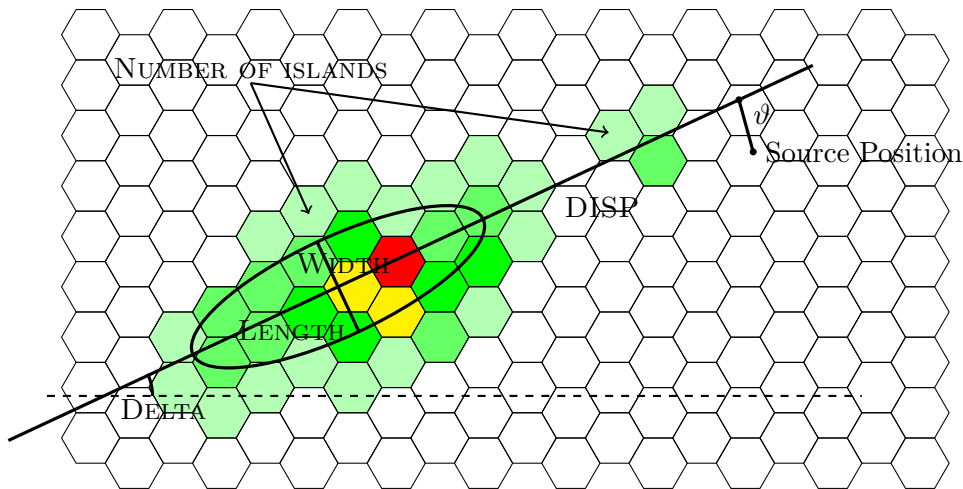


Figure 5.1: Illustration of the image parameters calculated by FACT-Tools.

## 5.2 Machine Learning

Machine learning methods aim to estimate the value of a target variable, called label, of unlabeled events. To do this a machine learning model is built with a training set of labeled events, for example Monte Carlo simulated events. The label can be, for example, the energy (see section 5.3.1) or the particle type of the primary particle (see section 5.4).

With the help of a test set of labeled events the performance of the models can be evaluated. This test set has to be independent, thus the events are not used for the training of the models. As the target variable for which the models are trained for is known in the test set, the quality of the application can be evaluated.

The bootstrap method is a common way to create new event sets  $X_i^*$  from an original event set  $X$ . If  $X$  properly represents the underlying, but unknown probability distribution  $F$ , the  $X_i^*$  sets can be used to estimate the distribution of a variable  $R(X, F)$ . The bootstrap method is described in section 5.2.1.

The independence of the test and training set is important to determine if the effect of overtraining occurs. In case of overtraining, the model is adapted to the training set in such a way that it performs perfectly on the training set. Whereas the application to a test set will perform catastrophically, due to the fact that small variations between test and training set yield very different results by the overtrained model.

The stability of the model and an estimation of the uncertainty of the performance determination can be evaluated if the model is trained and tested on several independent training and test sets. The cross validation method described in section 5.2.2 performs such a validation.

### 5.2.1 Bootstrap Method

The bootstrap method [28] is a common way to create  $m$  new sets  $X_i^*$  with  $n$  events each from an original set  $X$  with  $n$  events.

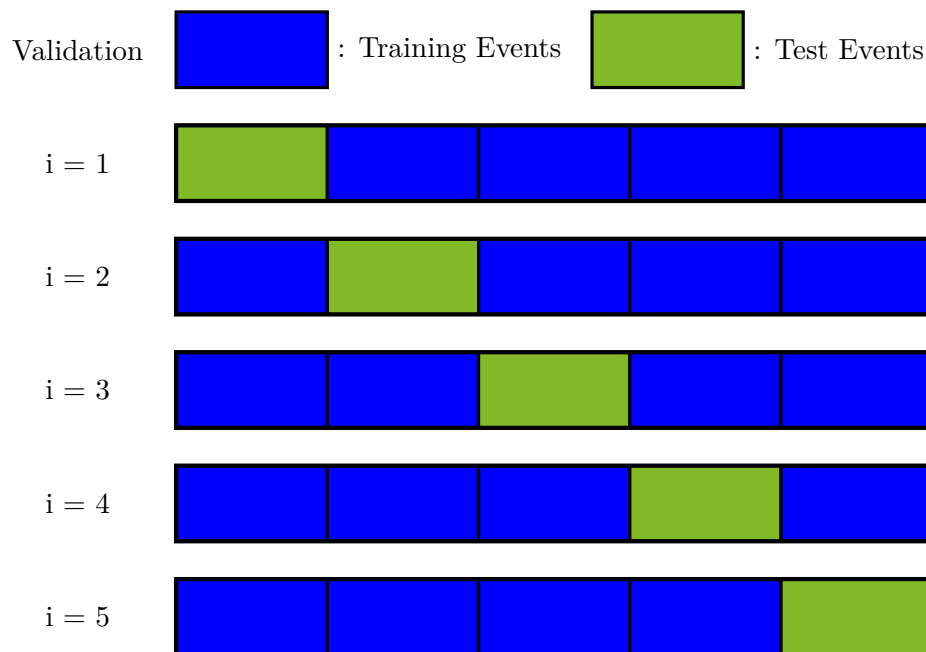
A bootstrapped subset is created by randomly sampling  $n$  events with replacement from the original set. If the original event set properly represents the true probability distribution  $F$  it is based on, the subsets can be used to estimate the distribution of a variable  $R(X, F)$ .

Hence the subsets can be used to evaluate the stability and uncertainty of an application of an arbitrary analysis task on the original event set  $X$ . For example the mean and the uncertainty of the estimation of the performance of a classification task can be calculated by testing  $m$  bootstrapped test events set and calculating the mean and the standard deviation of the performance values.

### 5.2.2 Cross Validation

Another similar method for validating machine learning models is the so called cross validation. An  $m$ -fold cross validation is performed as follows: A Monte Carlo simulated set is separated into  $m$  disjunct subsets. In each validation step the model is trained using the union of  $m - 1$  subsets and tested on the remaining subset. The test set is iterated over all  $m$  subsets. Figure 5.2 illustrates the principle of cross validation.

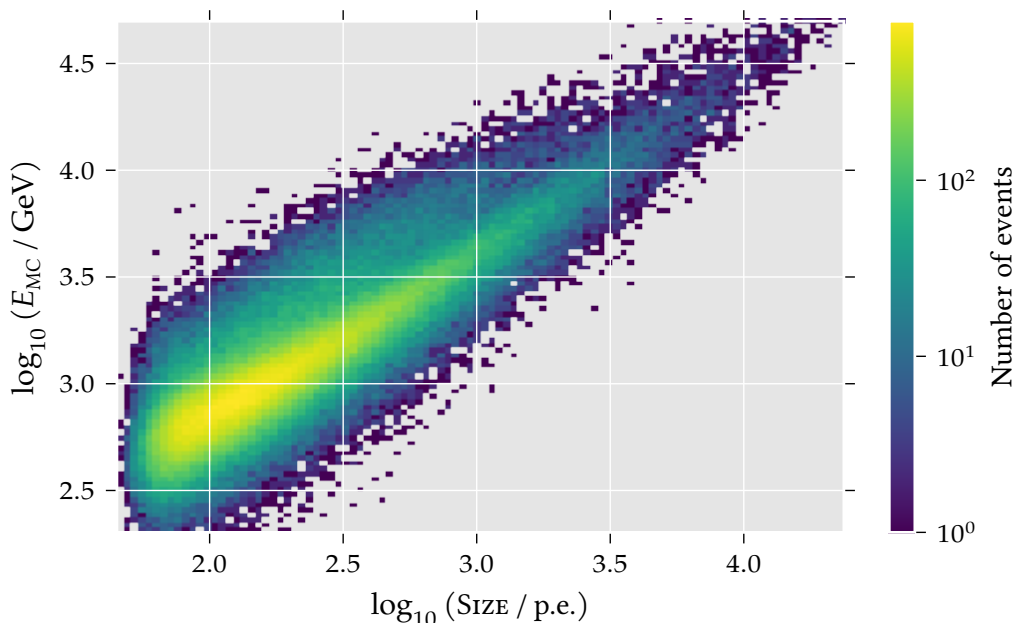
The model is tested on  $m$  independent subsets, the performance of the model can be evaluated and validated. In addition the model is built on  $m$  different training sets though also the model building is validated.



**Figure 5.2:** Illustration of the cross validation method. The labeled event set is separated into  $m$  disjunct subsets. The machine learning model is build on the union of  $m - 1$  subsets and tested on the remaining one. The test set is iterated over all  $m$  subsets, thus the performance of the model and the stability of the model training can be evaluated. In this illustration is  $m = 5$ .

### 5.3 Energy Estimation with Random Forest Regression

As already mentioned in section 5.1.1 the SIZE parameter is strongly correlated to the energy of the primary particle, as shown in figure 5.3.



**Figure 5.3:** Correlation between the SIZE and the true energy  $E_{\text{MC}}$ . SIZE is clearly strongly correlated to the energy. Hence it is often included in an energy estimation algorithm. Shown are background suppressed gamma events.

Nevertheless other parameters also contain information about the energy of the primary particle. Thus an estimation algorithm improves the energy reconstruction of the observed gamma ray events in comparison to the usage of SIZE as an energy estimation. Commonly used in Cherenkov astronomy are a parameterization of the estimated energy [20] or look up tables for the energy [4]. In this analysis a random forest regression is used for energy estimation.

#### 5.3.1 Random Forest Regression

A random forest regression [18] is an ensemble method consisting of  $n$  single decision trees which are individually trained to estimate the target variable. The result of the complete random forest is calculated by averaging over the results of the single decision trees.

A decision tree is built by consecutively separating the training events in different subsets in a way that the variance of the target variable in the subsets is minimized. Each separation is called a node. The separation is repeated until the subsets fulfill a given criterion. Such a criterion can be for example that the subset contains only one event. A subset which is not separated further is called a leaf. The estimation result of this leaf is the average of the target variable of all events in this leaf.

A decision tree gives a result for an unlabeled event by following the nodes for the event until a leaf is reached. The estimation for the target variable for this leaf is used for the unlabeled event.

A single decision tree is relatively unstable, due to the dependency on the training events and tends to be overtrained. To take this into account a random forest adds random effects to the training of the different decision trees. At each node  $j$  features from the total  $k$  features are randomly chosen. Over all  $j$  features the best separation criterion is evaluated and applied. Each tree is trained on a different bootstrapped subset of the training event set. Figure 5.4 illustrates the generation of a decision tree, including the random effects mentioned.

The random forest averages over the results of the individual trees. The resulting estimation of the target variable is stable and performs well. The principle structure of the random forest is illustrated in figure 5.5.

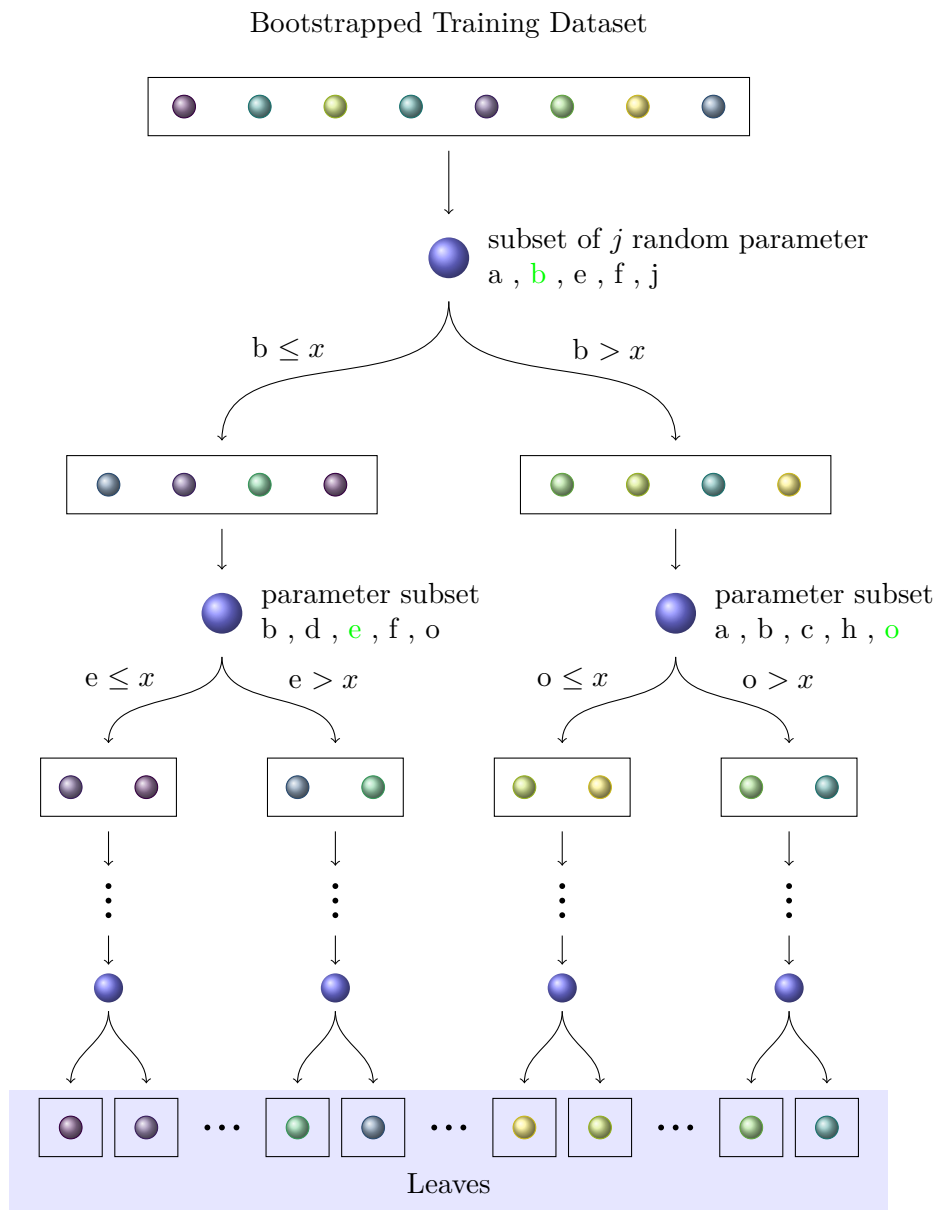
### 5.3.2 Performance Evaluation

The performance of a regression task like the random forest regressor can be evaluated by the coefficient of determination  $R^2$ . It measures how well the estimation  $x_{\text{est},i}$  fits the true value  $x_{\text{true},i}$  distribution.  $R^2$  is evaluated using  $N$  independent test events.

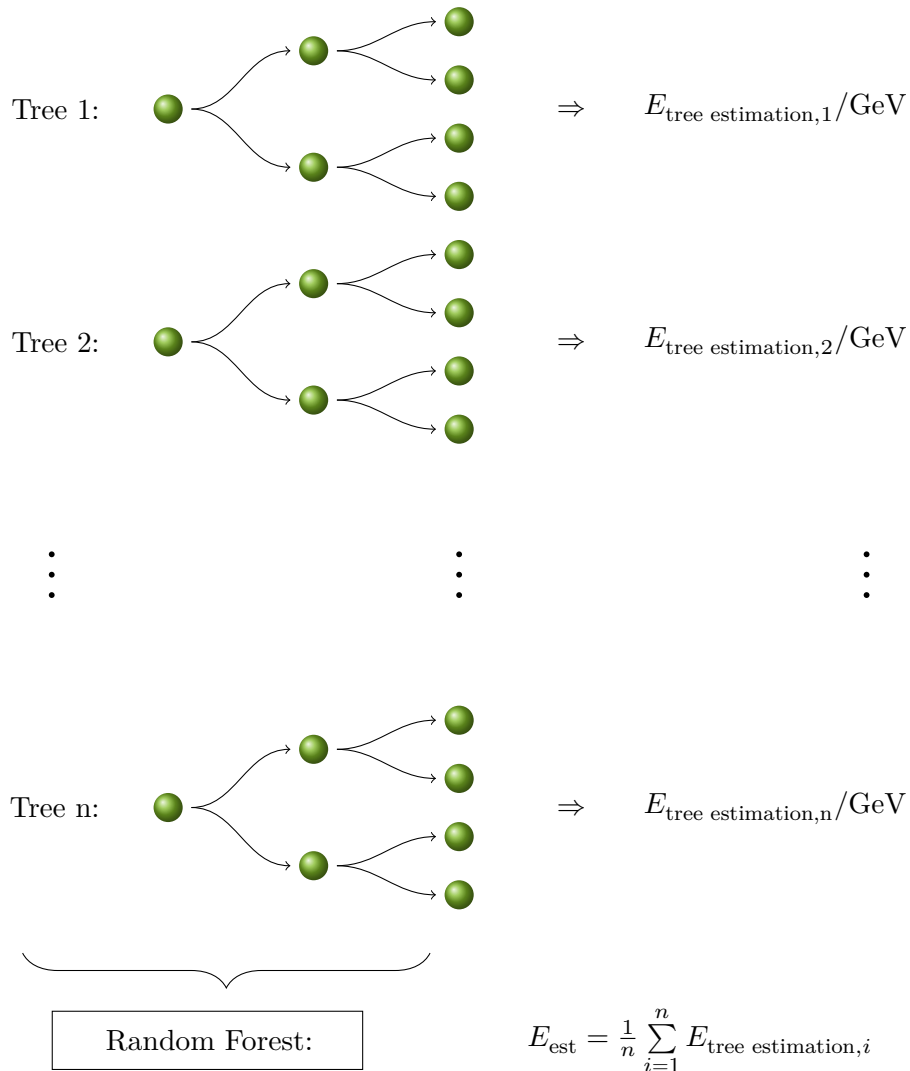
$$R^2 = 1 - \frac{\sum_i^N (x_{\text{true},i} - x_{\text{est},i})^2}{\sum_i^N (x_{\text{true},i} - \bar{x}_{\text{true}})^2} \quad (5.7)$$

$$\text{with: } \bar{x}_{\text{true}} = \sum_i^N x_{\text{true},i}$$

The mean and the uncertainty of  $R^2$  can be evaluated using cross validation.



**Figure 5.4:** Illustration of the generation of a single decision tree. The tree is built using a bootstrapped training event set. At each node a random subset of  $j$  parameters is drawn and the separation criterion, which minimize the variance of the target variable in the subsets, is applied.



**Figure 5.5:** Illustration of the principle structure of a random forest regressor. The target variable (in this example the energy  $E$ ) of an unlabeled event is estimated by evaluating the estimation of  $n$  different decision trees and averaging over the results.

## 5.4 Background Suppression with Random Forest Classification

The ratio between gamma induced air shower events and proton induced air shower events is at best 1000 to 1. Therefore a selection of gamma events by hand is not feasible. An automatic selection, called background suppression, has to be performed.

Commonly used in Cherenkov astronomy are selection cuts in image parameters (for example in [31, 5, 9]) or selection cuts in parameters evaluated by the use of look up tables [4].

In this analysis a random forest classifier is used to perform the background suppression. This allows to combine the information value of a set of image parameters to perform the classification task. Also the random forest does not apply a constant cut in any of the image parameters, the classification is based on a complex cut in the high dimensional image parameter space. This increases the efficiency of the background suppression.

Nevertheless, by using the whole set of image parameters the performance of the random forest classifier can be worse, than by using only a subset with parameters suitable for the background suppression. Hence a feature selection is performed to select the subset of image parameters used in the background suppression.

### 5.4.1 Feature Selection

Instead of using all features for background suppression, a feature selection is performed. There are some advantages provided by a feature selection. The dimension of the feature set is reduced, improving the computational costs of the classification noise. Features not relevant for the background suppression introduce noise in the feature set. Feature selection removes them, thus improving the accuracy of the classification task. The risk of overfitting is decreased by reducing the number of features used.

The first step in feature selection is a user defined feature selection using a priori knowledge of the meaning and importance of features (for example removing event number features from the feature set). Besides this user defined selection exist several feature selection algorithms to select automatically  $k$  features from a set of  $M$  total features. The algorithm used in this analysis is the Minimum Redundancy Maximum Relevance (MRMR) algorithm [25], which is now shortly described.



The MRMR algorithm aims to select a set of features with a maximum relevance towards the target classification label  $l$  and a minimum redundancy within each other. The algorithm evaluates the relevance and redundancy of two variables  $x$  and  $y$  with the mutual information  $I$ :

$$I(x, y) = \sum_{i,j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_j)p(y_j)} \quad (5.8)$$

with the joint probabilistic distributions  $p(x, y)$  and the respective marginal probabilities  $p(x)$  and  $p(y)$ . The first feature  $X$  is selected by the MRMR algorithm, that  $I(X, l)$  is maximal. The following features are selected iteratively. If  $S$  is the feature set of already selected features and  $\Omega_S$  the set of remaining features, the next feature  $Y$  is selected that the mutual information towards the label  $l$  is maximized and the sum of the mutual information towards the features already in  $S$  is minimized. The optimization criteria then is:

$$\max_{Y \in \Omega_S} \left[ I(Y, l) - \frac{1}{|S|} \sum_{X \in S} I(X, Y) \right] \quad (5.9)$$

### 5.4.2 Random Forest Classification

The random forest classifier [18] is very similar to the random forest regressor explained in section 5.3.1. The difference is that the target variable is a binomial variable, the class type (1: signal, 0: background), instead of a continuous variable like the energy. The random forest classifier is an ensemble of decision trees, each trained individually to classify a given event to be a signal or background event.

The training of the decision trees is the same as for the decision trees of the random forest regressor. The training set for each decision tree is a bootstrapped subset of the whole training set and at each node only  $j$  features from the total  $k$  features are randomly chosen. The training events at the nodes are separated in a way that the signal and background class in the subsets are separated as best as possible.

The result of a random forest classifier is the average over the single results of the decision trees. The result is called CONFIDENCE and covers the range from 0 to 1. With a CONFIDENCE value of 0 the random forest model estimates that the event is very likely a background event, with a CONFIDENCE value of 1 the event is very likely to be a signal event.

### 5.4.3 Performance Evaluation

There are several parameters to describe the performance of a classification task like a random forest classifier. They are evaluated using an independent test event set of  $n_1$  signal events and  $n_2$  background events. The mean and uncertainty of this performance determination can be evaluated in a cross validation.

Basic performance values are the true positives  $tp$ , false positives  $fp$ , true negatives  $tn$  and false negatives  $fn$ :

$tp$ : Number of signal events correctly classified as signal events.

$fp$ : Number of background events falsely classified as signal events.

$fn$ : Number of signal events falsely classified as background events.

$tn$ : Number of background events correctly classified as background events.

As these numbers depend on the size of the test set, the relative values recall and precision are defined as:

$$\text{recall} = \text{true positive rate} = \frac{tp}{n_1} \quad (5.10)$$

$$\text{precision} = \frac{tp}{tp + fp} \quad (5.11)$$

The recall describes the ratio of correctly classified signal events to the total number of signal events. The precision describes the ratio of correctly classified signal events to the total number of events classified as signal. Also of interest is the false positive rate, thus the ratio of falsely classified background events to the total number of background events:

$$\text{false positive rate} = \frac{fp}{n_2} \quad (5.12)$$

A well performing classification task has a maximum value for recall and a minimum value for the false positive rate (which means a maximum value for precision). But both values decrease with a larger cut in the CONFIDENCE of the random forest model. To choose a well performing cut in the CONFIDENCE the  $F_\beta$ -score [55]

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}} \quad (5.13)$$

is well suited.  $F_\beta$  takes both, the precision and the recall into account and places a weighting factor  $\beta$  to the precision, to consider an unbalanced signal background ratio as it is the case in this analysis.

## 5.5 Tikhonov Regularized Unfolding of an Energy Spectrum

The energy distribution of the very high energy gamma ray particles emitted by a source is an unknown function  $f(E)$ . The energy  $E$  cannot be measured directly by the telescope, only observables  $y$  which are correlated to  $E$  can be measured. The reconstruction of  $f(E)$  from the measured distributions  $g(y)$  of the observables  $y$  is not trivial. Particles can be absorbed on their way to earth or in the atmosphere. They may not trigger the telescope or are removed from the data set during the different analysis steps. This is called limited acceptance of the detector. In addition due to the finite resolution of the detector, particles with the same energy  $E$  may cause different values for the observables  $y$ . The transformation of the true energy distribution  $f(E)$  to the observables distribution  $g(y)$  is described by a Fredholm integral equation:

$$g(y) = \int A(y, E)f(E)dE + b(y) \quad (5.14)$$

with the response function  $A(y, E)$  including the finite resolution and limited acceptance effects described above.  $b(y)$  describes a known background distribution in the observables  $y$ .

Discretization of the distributions  $g(y)$ ,  $f(E)$  and  $b(y)$  and of the response function  $A(y, E)$  leads to the matrix form of equation (5.14):

$$\vec{g} = \mathbf{A} \cdot \vec{a} + \vec{b} \quad (5.15)$$

with the parameterization of  $f(E)$  by  $m$  cubic B-splines  $p_j$  with the coefficients  $a_j$ . The response matrix  $\mathbf{A}$  can be determined by Monte Carlo simulated events.

Solving (5.15) for the sought-after coefficients  $a_j$  is called unfolding. The straightforward approach is the inversion of  $\mathbf{A}$ . Due to the finite resolution  $\mathbf{A}^{-1}$  contains negative non-diagonal elements causing nonphysical oscillations in the solution for  $f(E)$ . Therefore equation (5.15) is also called an ill-posed problem. To suppress these oscillations a regularization term is added to the unfolding. The resulting equation is solved using a maximum likelihood fit.

The software package TRUEE – Time-dependent Regularized Unfolding for Economics and Engineering problems [43] performs such an unfolding. TRUEE uses a so called Tikhonov regularization which is proportional to the second derivative of the sought-after distribution. Depending of the strength of the regularization term, high order spline coefficients of the sought-after distribution are smoothly cutted away. This can be interpreted as a reduction of the effective number of degrees of freedom ( $ndf$ ). TRUEE offers the possibility to control the strength of the regularization by specifying the number of degrees of freedom. TRUEE also offers the possibility to specify the number of knots ( $nK$ ). Knots are the overlap points of adjacent polynomials of the cubic B-splines used for the parameterization of  $f(E)$ . A set of up to three observables  $y$  can be used to perform the unfolding. The observables should correlate with the sought-after variable and should contain complementary information to increase the precision of the estimation of the sought-after distribution.

The choice of well performing unfolding settings ( $ndf, nK$ ) can be challenging. The strength of the regularization has to be large enough to suppress nonphysical oscillations, but low enough not to smooth away significant features in the sought-after distribution. A large regularization induces a positive correlation between the unfolded data points and may introduce a too high bias. A low regularization results in large fluctuations and uncertainties. This can be evaluated by using Monte Carlo simulated events as pseudo data events. Since the true distribution of the simulated events is known, the agreement between unfolded distribution and true distribution can be calculated. The agreement should be as good as possible. TRUEE offers a test mode to evaluate the best performing unfolding settings by investigating the correlation of the unfolded data points and the agreement between unfolded and true distributions. The test mode and its results for the analysis of this thesis are described in detail in section 6.6.2.

To test the stability of the unfolding settings TRUEE offers a so called pull mode. Test unfoldings with pseudo data event sets are performed several times (each called one pull). The performance of the test unfoldings are evaluated over all pulls. More details and the results of the pull mode performed in the analysis of this thesis can be found in section 6.6.3.

The unfolding of the real data sets is performed in the data mode of TRUEE. Measured background distributions in the observables  $y$  can be subtracted from the signal observables distributions. An acceptance correction to correct for the limited acceptance of the detector and loss of events during the analysis can be performed in the data mode too. Therefore the function which is used to generate the Monte Carlo event distribution has to be specified by the user. The application of the data mode to the data of the Crab Nebula obtained by FACT is described in section 6.6.4.

More details about TRUEE can be found in [43]. The dependency of the unfolding results from the producing spectrum of the Monte Carlo simulated events is investigated in [43]. A moderate deviation in the slope of the spectrum of  $\gamma_{\text{sim}} = \gamma_{\text{real}} \pm 1.0$  is tolerable, thus a-priori knowledge of the slope of the target distribution  $f(E)$  is not necessary.



## 6 Analysis of the Crab Nebula with FACT

This thesis describes the analysis of data of the Crab Nebula obtained by FACT. The goal of this analysis is the detection of a signal from the Crab Nebula in the data and the reconstruction of the Crab Nebula energy spectrum.

The used data and Monte Carlo samples are described in section 6.1. The analysis chain starts with the preprocessing of the events with FACT-Tools (section 6.2). The next step is the energy estimation using a random forest regression, described in section 6.3. To suppress the background in the obtained data set a random forest classification is performed. The training and validation of the random forest model is described in section 6.4. With a classified data set a detection of a signal of the observed source can be performed and the signification of this detection can be evaluated (section 6.5). The unfolding of the energy spectrum is presented in section 6.6.

### 6.1 Data and Monte Carlo Sample

The data of the Crab Nebula sample was obtained by FACT between October 2013 and February 2014. This time window corresponds to the first visibility period of the Crab Nebula after the main hardware and software adaptations during the first commissioning phase were done. To ensure dark night conditions and good observation conditions a data check is applied to the observed data in this time window. The data check is described in detail in section 6.1.1.

Several methods used in this analysis are based on Monte Carlo simulated events. Hence a number of gamma and proton events have to be simulated. The simulation itself is described in chapter 4, the generation of the events in section 6.1.2.

For simplicity real data events taken by the telescope are called data events in the following. Monte Carlo simulated events are called gamma and proton events.

### 6.1.1 Data Check

This analysis focuses on data with low zenith distances during dark night conditions and during good observation conditions. Hence data runs with high zenith distances, too much light or bad observation conditions are removed. This is called data check. The applied data check is based on the data check described in [27], the different types of data check conditions will now be shortly illustrated. Table 6.1 shows the number of files and the total observation time after different types of data check conditions are applied.

**Crab Physics Runs:** So called physics runs are the standard data taking runs with enabled physics trigger. In total nearly 5000 runs in the given time window were taken while observing Crab. They sum up to an observation time of roughly 350 h.

**Zenith Distance:** Low zenith distances are often defined smaller than  $30^\circ$ . Hence the taken data is filtered to obtain only runs with these zenith distance. About 2500 runs remain with an observation time of about 185 h.

**Dark Night:** To ensure dark night conditions all runs with zenith distance of the moon smaller than  $100^\circ$  are excluded. In addition the amount of currents in the pixels is an indirect measurement of the ambient light while observing [14]. Also the trigger threshold set by the system depends on the light conditions. Hence all runs with `MEANCURRENTS` greater or equal  $8 \mu\text{A}$  or with a `TRIGGER THRESHOLD` greater or equal 350 DAC-count are excluded. There are also runs in the `TRIGGER THRESHOLD - MEANCURRENTS` parameter space with bad observation conditions. These runs are excluded by the following condition:

$$\text{TRIGGER THRESHOLD} \geq 14 \frac{\text{DAC-count}}{\mu\text{A}} \cdot \text{MEANCURRENTS} + 265 \text{ DAC-count} \quad (6.1)$$

About 1250 runs with an observation time of about 96 h remain.



**Basic Quality:** The mean trigger rate for the telescope should be around 60 Hz. All runs with a mean trigger rate less or equal 45 Hz or greater or equal 85 Hz are excluded. In addition an effective ontime<sup>1</sup> below 95 % indicates worse observation conditions. These runs are also excluded. About 1180 runs with an observation time of about 90 h remain.

**Analysis Based:** The data check described in [27] is based on the gamma background rate (see section 3.3 for a definition) calculated by an independent analysis chain (Mars-CheObs Analysis [20]). The applied filter takes also the zenith and light condition dependency of the background rate into account. In total 1150 runs with an observation time of 87.63 h remain.

The reference analysis chain Mars-CheObs was updated after the data sample used in this analysis was defined by applying the described data check conditions. Hence the background rate on which the "Analysis Based" data check condition is based on was recalculated. Applying the "Analysis Based" data check condition to the recalculated background rate results in a slightly different data sample. Now 40 of the original 1150 runs are not part of the data sample when using the recalculated background rate and additional 27 runs will be added. The original 1150 runs are used in this analysis.

**Table 6.1:** Number of files and observation time after the application of the different types of data check condition.

Data Check Condition	Number Of Files	Observation Time / h
Crab Physics Runs	4936	346.57
Zenith Distance	2521	184.11
Dark Night	1263	96.08
Basic Quality	1179	89.57
Analysis Based	1150	87.63

The list of the runs and a list of all data check conditions can be found in the appendix A.

<sup>1</sup>ratio of time with enabled trigger to the duration of the data run

### 6.1.2 Simulated Monte Carlo Events

As described in chapter 4 the program CORSIKA simulates air showers caused by very high energy primary particles. For this analysis air showers caused by very high energy gamma rays and air showers caused by very high energy protons are simulated. Next step of the simulation is the program Ceres of the Mars CheObs framework, simulating the reflector and the camera of the telescope. The output of Ceres contains the voltage curves of the simulated pixels and is similar to the output of the data acquisition system.

**Table 6.2:** Simulation parameters for CORSIKA for the proton and gamma simulations.

Simulation Parameter	Proton Simulation	Gamma Simulation
Energy Range	100 GeV - 200 TeV	200 GeV - 50 TeV
Slope of Energy Spectrum	-2.7	-2.7
Maximum impact parameter	400 m	270 m
Viewing Cone	5°	0°
Zenith Distance	0° to 30°	0° to 30°
Number of events	780 046 520	12 000 000

Table 6.2 summarizes the simulation parameters for the CORSIKA simulations for the gamma and proton samples. The slope of the simulated gamma ray spectrum should be similar to the expected slope of the Crab Nebula, but must not have the same value necessarily (see 5.5 for details). The slope of the simulated proton spectrum is equal to the slope of the cosmic ray background in [47] (see equation (2.1)). The maximum impact parameter, the viewing cone and the energy range, has to be large enough that all showers which are able to survive all analysis steps are simulated. The simulated zenith distances cover the same range as for the data sample. The parameters for CORSIKA are specified by so called inputcards. The inputcards used in this analysis can be found in the appendix C.

The parameters used in the reflector and camera simulation in Ceres are specified with a so called rc file. The file used in this analysis can be found in the appendix C. This setting has the internal production number 12 and the parameters are adapted in a way that the simulated proton image parameter distributions fit with the image parameter distributions of the real triggered cosmic ray showers. A comparison of the image parameters of the simulated proton showers and the real showers is shown in section 6.2.

## 6.2 Preprocessing with FACT-Tools

The raw data obtained by the telescope as well as the simulated raw data created by Ceres have to be preprocessed. In this analysis the preprocessing software FACT-Tools [30] in the version 0.14.1 is used. It performs a calibration of the raw data, an extraction of the number and arrival time of the Cherenkov photons per pixel, an image cleaning to identify the air shower image and a parameterization of the shower image last. The output of FACT-Tools is a set of parameters for each event which survives the cleaning. A list of all parameters calculated by FACT-Tools can be found in the appendix B. From the 21 300 000 data events which triggered the telescope 6 920 000 survived the cleaning. For gamma and proton events the ratios are 1 180 000 from 1 920 000 gamma events and 300 000 from 510 000 proton events. The parameter distributions for data events and proton events should agree with each other. For a subset of the parameters the distributions are shown and discussed in section 6.2.1. Based on the image parameter distributions simple quality cuts are chosen to perform a rough background suppression and to get rid of bad reconstructed events or not simulated background events<sup>2</sup>. The quality cuts are described in section 6.2.2.

### 6.2.1 Parameter Distributions for Data and Monte Carlo Events

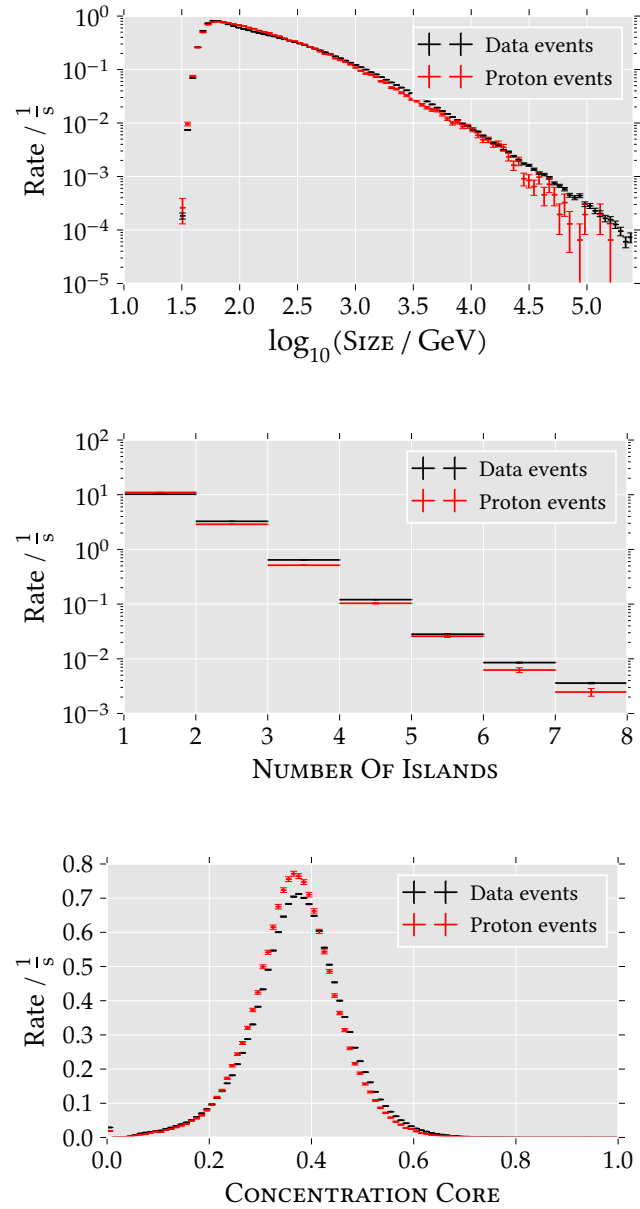
The distributions of a subset of image parameters calculated by FACT-Tools for proton and data events are shown in figures 6.1, 6.2 and 6.3. To remove bad reconstructed events and not simulated background events only events with:

$$\text{NUMBER OF ISLANDS} < 8 \quad \text{NUMBER OF PIXELS IN SHOWER} > 7 \quad (6.2)$$

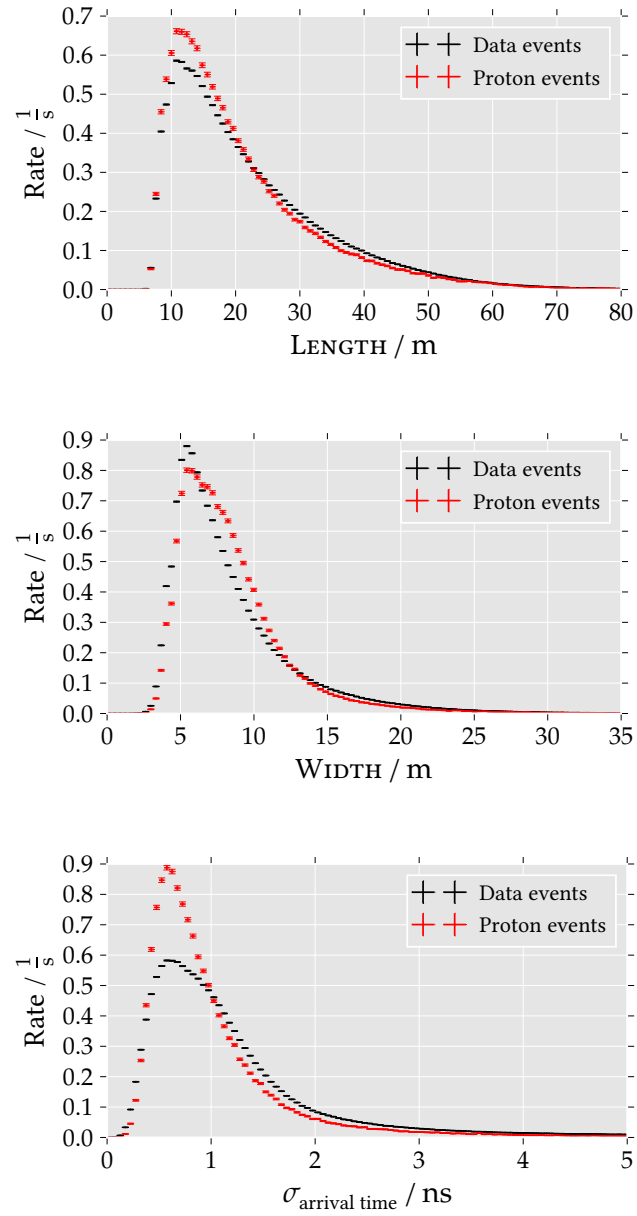
are shown.

---

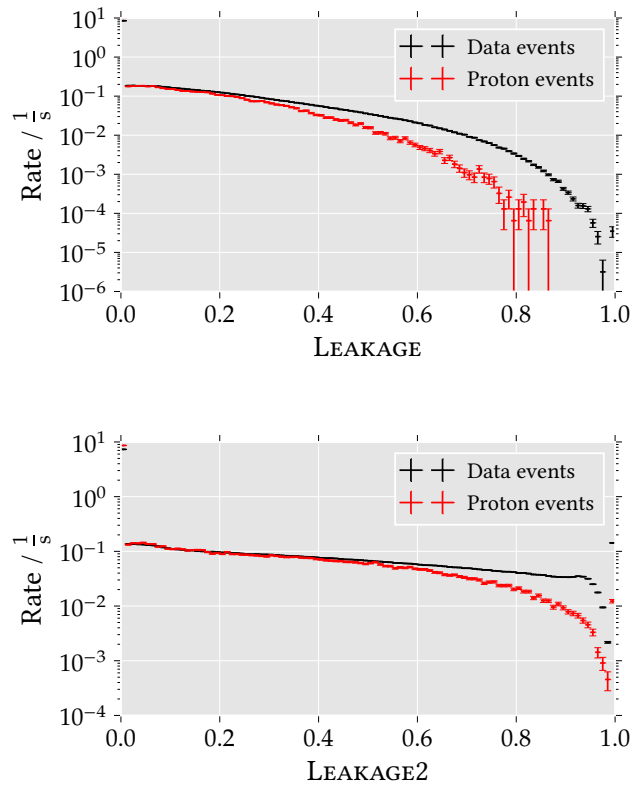
<sup>2</sup>for example highly illuminated data events from a car passing by the telescope



**Figure 6.1:** Rate distributions of the image parameters SIZE, NUMBER OF ISLANDS and CONCENTRATION CORE for proton and data events. The distributions are scaled by the observation time for the data and by the simulated observation time for the proton events.



**Figure 6.2:** Rate distributions of the image parameters LENGTH, WIDTH and the STANDARD DEVIATION OF THE ARRIVAL TIME for proton and data events. The distributions are scaled by the observation time for the data and by the simulated observation time for the proton events.



**Figure 6.3:** Rate distributions of the image parameters LEAKAGE and LEAKAGE2 for proton and data events. The distributions are scaled by the observation time for the data and by the simulated observation time for the proton events.

For comparison the distributions are scaled by the observation time for data events and by the simulated observation time  $T_{\text{sim. Obs}}$  for proton events.  $T_{\text{sim. Obs}}$  is calculated using the measured flux of the cosmic ray spectrum (see equation (2.1)).

$$N_{\text{sim}} = \int_{E_{\text{min}}}^{E_{\text{max}}} \int_0^{A_{\text{max}}} \int_0^{T_{\text{sim. Obs}}} \int_0^{\Omega_{\text{max}}} N_0 E^{-\gamma} d\Omega dT dA dE \quad (6.3)$$

$$\Leftrightarrow T_{\text{sim. Obs}} = \frac{N_{\text{sim}} \cdot (-\gamma + 1)}{A_{\text{max}} \cdot \Omega_{\text{max}} [E^{-\gamma+1}]_{E_{\text{min}}}^{E_{\text{max}}}} \quad (6.4)$$

with

$$N_0 = 1.8 \times 10^4 \frac{\text{nucleons}}{\text{m}^2 \text{ssr GeV}} \quad \gamma = 2.7$$

from [47] and with the simulation parameters (see table 6.2):

$$\begin{aligned} N_{\text{sim}} &= 780\,046\,520 \\ A_{\text{max}} &= \pi \cdot \text{impact}_{\text{max}}^2 = \pi \cdot (400 \text{ m})^2 = 502\,655 \text{ m}^2 \\ \Omega_{\text{max}} &= 2\pi \cdot (1 - \cos(\text{viewing cone})) = 2\pi \cdot (1 - \cos(5^\circ)) = 0.0239 \\ E_{\text{min}} &= 100 \text{ GeV} \quad E_{\text{max}} = 200 \text{ TeV} \\ \Rightarrow T_{\text{sim. Obs}} &= 15\,397.8 \text{ s} \end{aligned}$$

The agreement between data events and proton events for the different parameters is good enough to perform the following analysis steps. Nevertheless there are some mismatches which will be now shortly discussed.

**Length and Width:** There is a small difference between the data and the proton distributions for both LENGTH and WIDTH. Currently there are several points in the simulation under investigation, which can cause these differences (section 4.5). Nevertheless the following analysis steps are performed under the assumption that the differences are small enough to not cause a wrong background suppression. The clear signal of the Crab Nebula as a source of very high energy gamma rays supports this assumption (section 6.5).

**Leakage:** There is a larger rate of data events for higher LEAKAGE values. As there are no gamma events with higher LEAKAGE values (see figure 6.4), this mismatch can be removed by applying a quality cut in LEAKAGE.

**Arrival Time:** There is a mismatch between the distribution of  $\sigma_{\text{Arrival Time}}$  for proton events and data events. The reason is not clear but is investigated currently. As the dependency of other parameters on the arrival time distribution is small enough it is adequate to just use only time independent parameters in the following analysis steps.

**Number of islands:** There is a slightly larger rate of data events with more islands. This is due to the light of a single star in the field of view which is not simulated. As gamma events have less islands than proton events (see 6.5), most of the mismatching events will be removed in the later background suppression.

### 6.2.2 Quality Cuts

To remove bad reconstructed events and not simulated background events (for example car flashes<sup>3</sup>) and to perform a rough background suppression quality cuts are applied to the data and Monte Carlo event sets. They are chosen to cut away only proton events in the parameter distributions. Quality cuts are applied on the following parameters:

Number of pixels in shower, NUMBER OF ISLANDS, WIDTH, LENGTH, LEAKAGE and LEAKAGE2.

Figures 6.4 and 6.5 show the distributions of the parameters for gamma and proton events. The distributions are normalized to an area under the curve of one. The chosen cuts are marked with dashed lines. The cuts are:

$$\text{NUMBER OF PIXELS IN SHOWER} \geq 10 \quad \text{NUMBER OF ISLANDS} < 8 \quad (6.5)$$

$$\text{LENGTH} < 70 \quad \text{WIDTH} < 30 \quad (6.6)$$

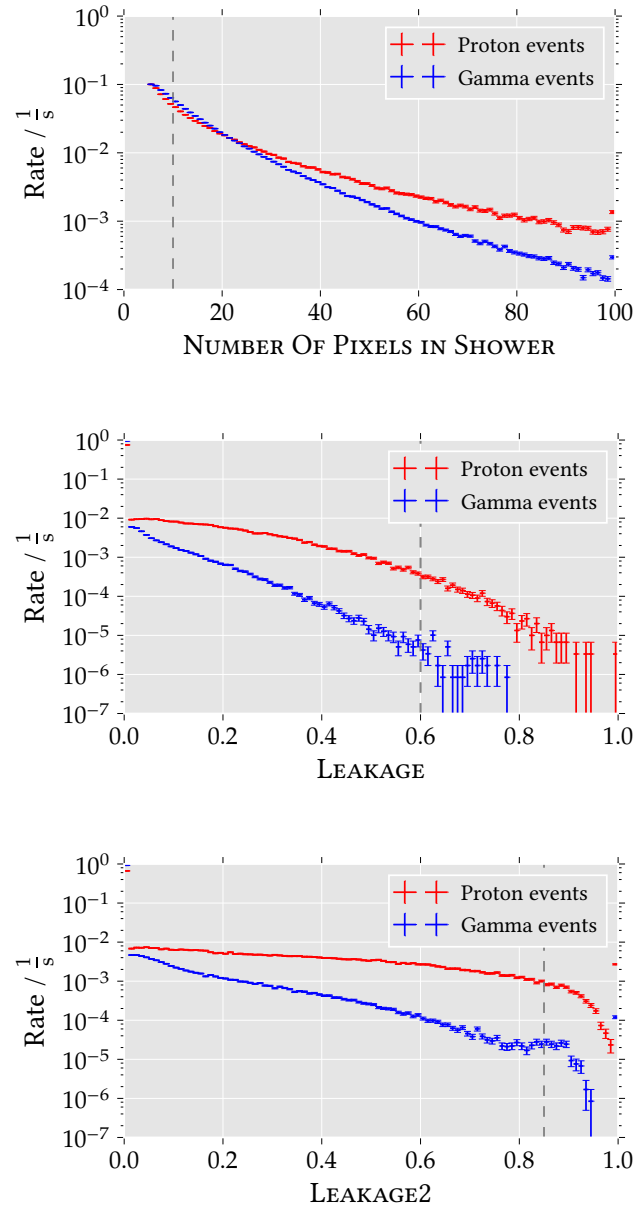
$$\text{LEAKAGE} < 0.6 \quad \text{LEAKAGE2} < 0.85 \quad (6.7)$$

After the application of the quality cuts 3 594 347 data events, 692 836 gamma events and 188 283 proton events remain.

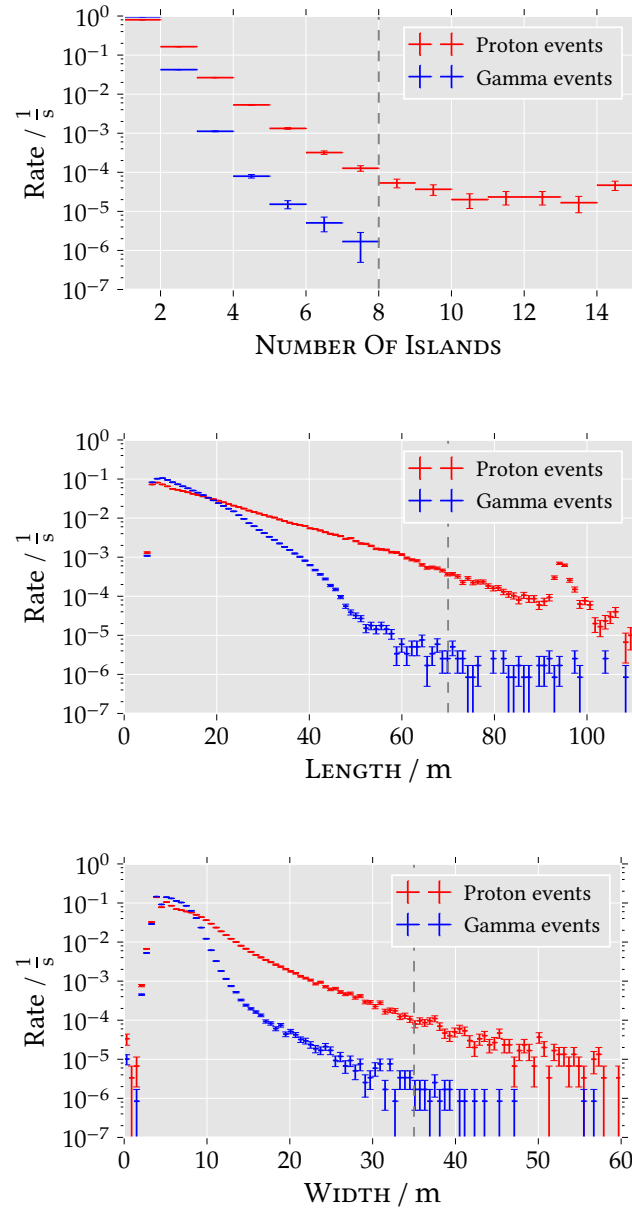
---

<sup>3</sup>Highly illuminated events from a car passing by the telescope and "flashing" the camera





**Figure 6.4:** Normed distributions of the parameters NUMBER OF PIXELS IN SHOWER, LEAKAGE and LEAKAGE2 of proton and gamma events of the image parameters on which the quality cuts are applied. The cuts are marked with gray dashed lines.



**Figure 6.5:** Normed distributions of the parameters NUMBER OF ISLANDS, WIDTH and LENGTH of proton and gamma events of the image parameters on which the quality cuts are applied. The cuts are marked with gray dashed lines.

### 6.2.3 Splitting Gamma Data Set

As described in chapter 5, the following analysis steps, the energy estimation (section 6.3) and the background suppression (section 6.4) are performed by machine learning algorithms. Hence independent test and training Monte Carlo sets for the different steps are needed. Also the unfolding of the energy spectrum with TRUÉE (section 6.6) needs independent test and training sets.

As the target variables in the energy estimation and background suppression (energy of the primary particle and particle type of the primary particle) are independent of each other, the same events can be used for training of these two steps. The Monte Carlo events used in the unfolding have to be background suppressed. Therefore the events used for training of the background suppression cannot be used for the unfolding. Nevertheless the events used in the unfolding can be used to test the energy estimation.

Therefore the gamma event set is splitted into two parts: the *Separation* set (used for training of the energy estimation random forest and the background suppression random forest) and the *Post Separation* set (used for testing the energy estimation random forest and for the unfolding).

As for the acceptance correction of the unfolded energy spectrum (section 5.5) the number of simulated events is needed, this number has to be noted for the splitted event sets. Table 6.3 lists the number of measured events and the number of simulated events for the two gamma event sets.

**Table 6.3:** Split of the gamma event set into two independent event sets. The number of measured events in each set and the corresponding number of simulated events is listed.

Set	Number of measured events	Number of simulated events
<i>Separation</i>	250 000	4 330 030
<i>Post Separation</i>	442 836	7 669 970

## 6.3 Energy Estimation with Random Forest Regression

To estimate the energy of the primary particle a random forest regressor is used. This regression method is described in section 5.4. In this analysis the implementation of the random forest regressor in the SciKit Learn [48] python package is used. The settings are listed in table 6.4.

The following parameters are used:

- SIZE
- WIDTH
- LENGTH
- M3TRANS
- M3LONG
- CONCCORE
- M3L
- M3T
- CONCENTRATION\_ONEPIXEL
- CONCENTRATION\_TWOPIXEL
- LEAKAGE
- LEAKAGE2
- CONCCOG
- NUMISLANDS
- NUMPIXELINSHOWER
- PHCHARGECHOWER\_MEAN
- PHCHARGECHOWER\_VARIANCE
- PHCHARGECHOWER\_MAX

**Table 6.4:** Settings for the random forest regression used for an estimation of the energy of the primary particle

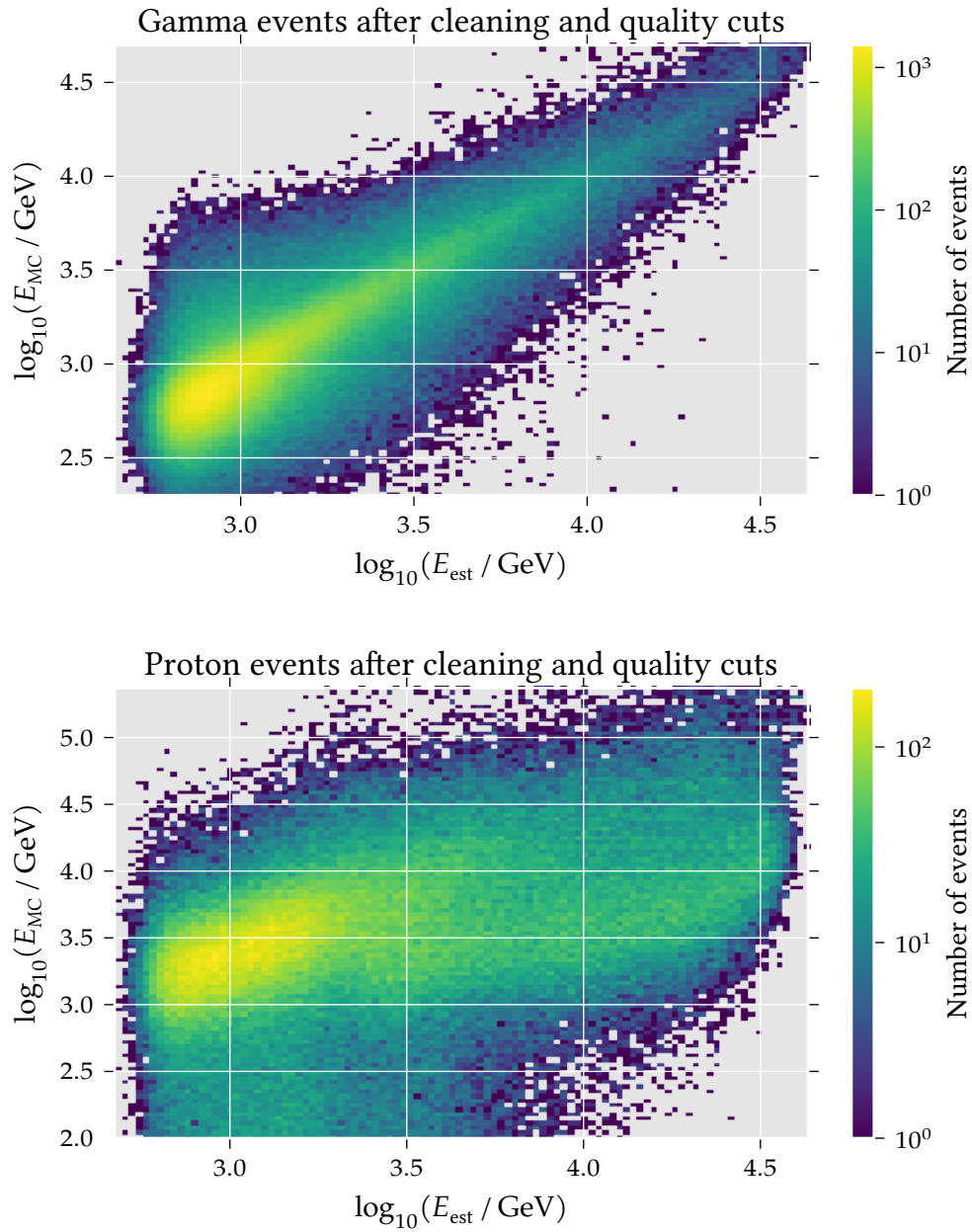
Setting	Value
Number of trees	100
Number of features per node	5
Number of total features	18
Minimum sample split	3
Number of training events	200 000
Number of tests events per validation	50 000

The performance is evaluated in a 5-fold crossvalidation. The coefficient of determination  $R^2$  for the trained random forest regression model is:

$$R^2 = 0.807 \pm 0.002 \quad (6.8)$$

The trained random forest regressor is applied to the *Post Separation* gamma set<sup>4</sup> and the correlation between the true energy  $E_{MC}$  and the estimated energy  $E_{est}$  is evaluated. This is shown in figure 6.6 (top). A linear correlation is clearly visible, but the spread around this correlation is quite wide.

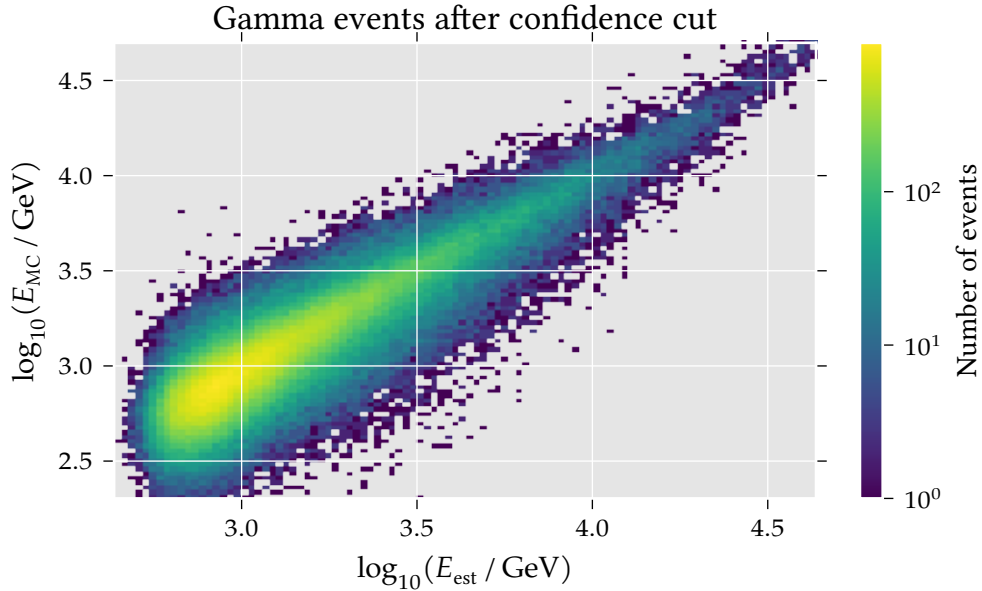
<sup>4</sup>The set is not separated yet. The *Post Separation* set is used as a test set, cause it was not used for training of the random forest regressor



**Figure 6.6:** Correlation between simulated energy and estimated energy for gamma (top) and proton (bottom) events. Shown are all events which survived the cleaning and the quality cuts.

The random forest regressor is trained on gamma events. Therefore it is expected that the performance for proton events is worse. Figure 6.6 (bottom) shows the correlation for proton events. Although a linear correlation is also visible, the spread is much larger and there is also a large bias towards smaller estimated energies. As the energy of proton events is not a question in this analysis, the performance of the energy estimation may be so poor for proton events.

After performing the background suppression (see section 6.4) only gamma events which are distinguishable from proton events by the random forest classifier remain in the gamma event set. Figure 6.7 shows the correlation between  $E_{\text{MC}}$  and  $E_{\text{est}}$  for gamma events which survive the background suppression. The performance of the energy estimation clearly improves. There are two features visible which are now discussed in more detail.

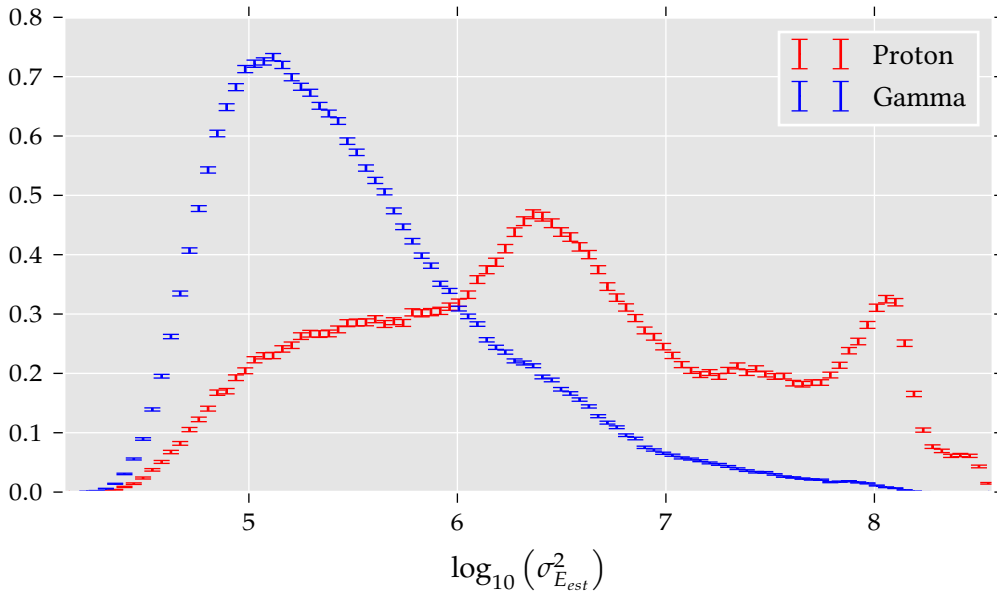


**Figure 6.7:** Correlation between simulated energy and estimated energy for gamma events. Shown are all events after the CONFIDENCE cut for the unfolding (see section 6.6).

If the information value of the parameters of an event is small (which is often the case for low energy events) the estimation of the random forest regressor tends to the most frequent energy value in the training event set. The energy distribution of the training event set has its maximum at 724 GeV (see table 7.2). Hence the spread of the correlation between  $E_{\text{MC}}$  and  $E_{\text{est}}$  around 724 GeV is quite wide.

Low energy events only trigger the telescope when they deposit a statistically unusual large amount of light in the camera. For the random forest regressor the signature of these events is similar to the signature of events with a higher energy. Hence they are estimated to have a higher energy, the dependency between  $E_{MC}$  and  $E_{est}$  steepens for Monte Carlo energies below  $\approx 600$  GeV. This effect is clearly visible as a positive energy bias for small energies (section 7.3.1).

As the random forest regression is an ensemble based regression method which calculates the average of all single tree estimations, also the variance over the single tree estimations can be calculated. The random forest is trained using gamma events, hence it is expected that the variance for proton events is larger than for gamma events. Figure 6.8 shows the distributions of the variance for gamma and proton events. The distributions are distinguishable from each other, hence the variance of the estimated energy can be used as a parameter in the background suppression section 6.4).



**Figure 6.8:** Distribution of the variance of the estimated energy for gamma and proton events. It is clearly visible that the variance for proton events is larger than for gamma events, thus the variance is suitable for the background suppression.

## 6.4 Background Suppression with Random Forest Classification

### 6.4.1 Feature Selection

As a first step of the feature selection for the background suppression all time dependent features are removed from the event sets (see section 6.2.1). Also all features which are known to be uncorrelated to the type of the primary particle are removed. For example the event number, the time of the event, or the position of the center of gravity.

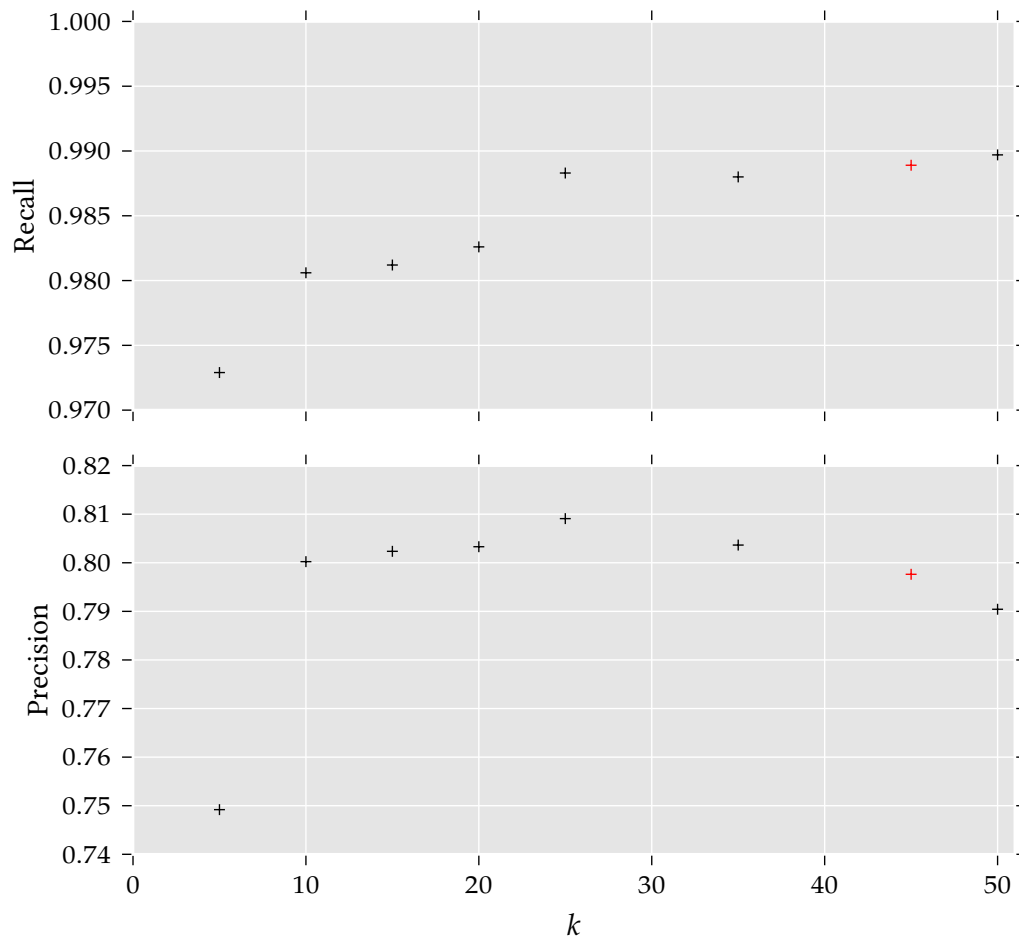
Several new features are created from the already existing ones. A list of the new generated features can be found in the appendix B. The final feature set for the background suppression is selected by a minimum redundancy maximum relevance method (MRMR) as described in section 5.4.1. The MRMR method selects the features from 53 features in total.

To choose the number of features selected by the MRMR algorithm the stability of the feature selection and the performance of the background suppression for different numbers of selected features is evaluated. The stability of the feature selection with the MRMR algorithm is evaluated by performing 10 feature selections of  $k$  features for each  $k$  between 1 and 50. The different selections are performed on bootstrapped event subsets. For each  $k$  the Jaccard index  $J$  [40] and the Kuncheva index  $I_C$  [38] of the 10 feature sets are calculated. A stable selection results in feature sets independent of the event subset. Thus the different feature sets should not differ too much from each other, the indices should be near one. The feature selection for this analysis is stable ( $J, I_C > 0.93$ ) for all numbers of selected features between 1 and 50.

The performance for different number of selected features is evaluated by investigating the performance of different random forest models which use  $k$  features selected by MRMR.  $k$  covers the range between 5 and 50. The random forest models are all trained with 100 trees and 95 000 gamma and 95 000 proton events as training events. The number of features per node used is  $\lfloor \log(k) + 1 \rfloor$ , which is the default value for  $k$  features. The recall and precision are calculated by applying a CONFIDENCE cut of 0.9. Figure 6.9 shows the recall and the precision of the random forest models against  $k$ .

For the background suppression  $k$  was chosen to have the maximum recall for an acceptable precision of at least 0.795, resulting in  $k = 45$ . All 53 features available for the MRMR algorithm are listed in the appendix B, the 45 selected features are marked there as well.





**Figure 6.9:** Recall and precision for random forest models trained with  $k$  features selected by the MRMR algorithm. The chosen  $k$  for the following background suppression random forest model is marked in red.

### 6.4.2 Training and Validation of the Random Forest Classifier

For the background suppression a random forest classifier as described in section 5.4 is used. The implementation of the Weka-Random Forest in the weka extension of RapidMiner [42] is used. Table 6.5 lists the settings of the random forest classifier.

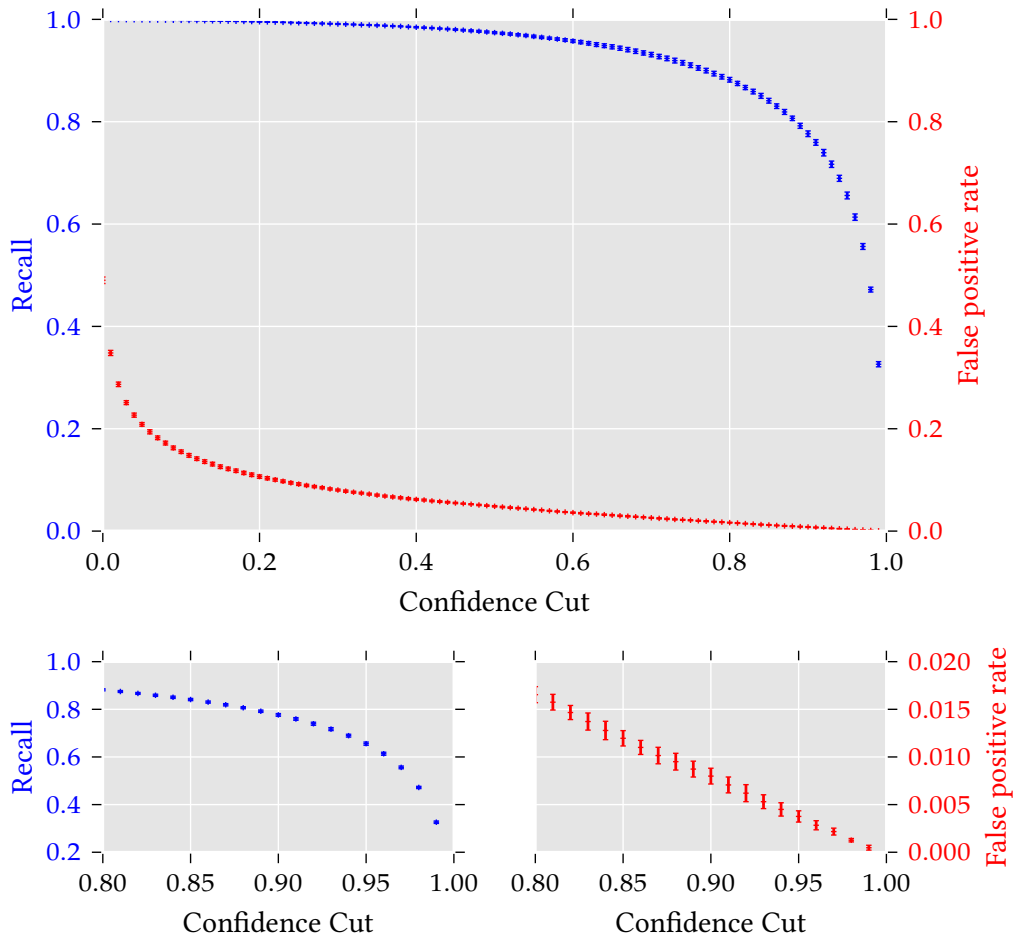
**Table 6.5:** Settings for the random forest classifier used for the background suppression.

Setting	Value
Number of trees	100
Number of features per node	7
Number of total features	45
Number of gamma training events	95 000
Number of proton training events	95 000

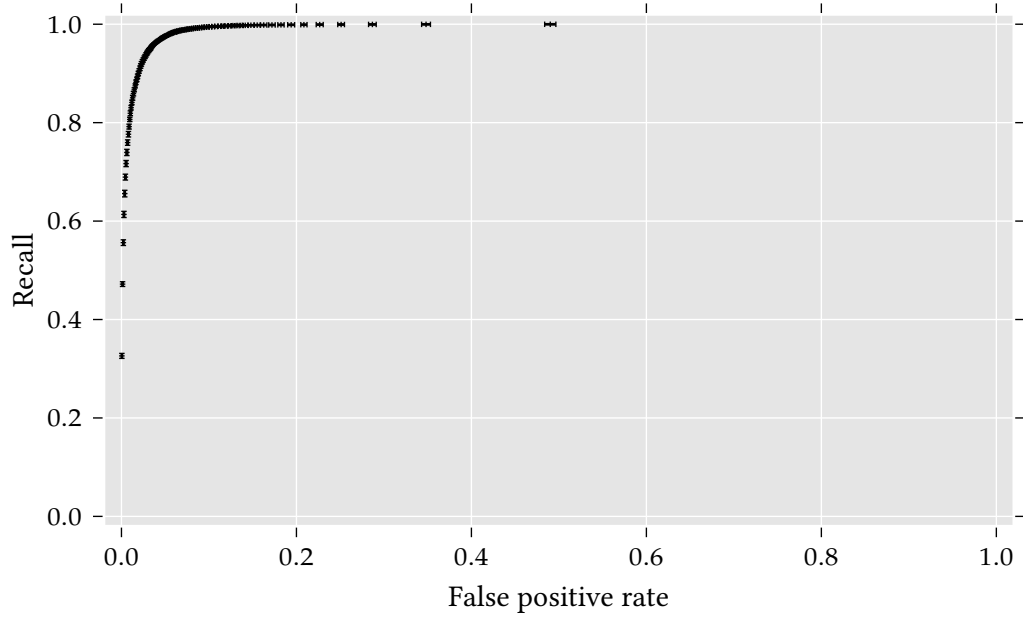
The random forest classifier is trained and tested using a 10-fold cross validation. The performance and the stability of the model building can be evaluated by the results of the test event sets in the cross validation. The cross validation method is described in section 5.2.2 and the performance evaluation of a random forest classifier in section 5.4.3. Figure 6.10 shows the recall and false positive rate in dependency of the applied CONFIDENCE cut. In the lower part of the figure the relevant range of CONFIDENCE cuts is shown in more detail. As expected, both values decrease for higher CONFIDENCE cuts. The lowest possible CONFIDENCE cut (and therefore the highest recall) with an acceptable false positive rate has to be chosen for the background suppression. The small uncertainties of the efficiencies are indicators for a stable model building, no overtraining occurs.

The receiver operating characteristics (roc) curve displays the recall against the false positive rate. A perfect classification task would have a step function at a false positive rate of zero, the roc curve of a task which classifies randomly is an angle bisector. A suitable CONFIDENCE cut for the background suppression should be in the upper left edge of the roc curve. The area under the roc curve (AUC) is a measure of the overall classifier performance. The roc curve for the trained random forest classifier is displayed in figure 6.11. The area under the curve is:

$$\text{AUC} = 0.9927 \quad (6.9)$$



**Figure 6.10:** Recall and false positive rate in dependency of the applied CONFIDENCE cut. The relevant CONFIDENCE cut range is shown in more detail below.

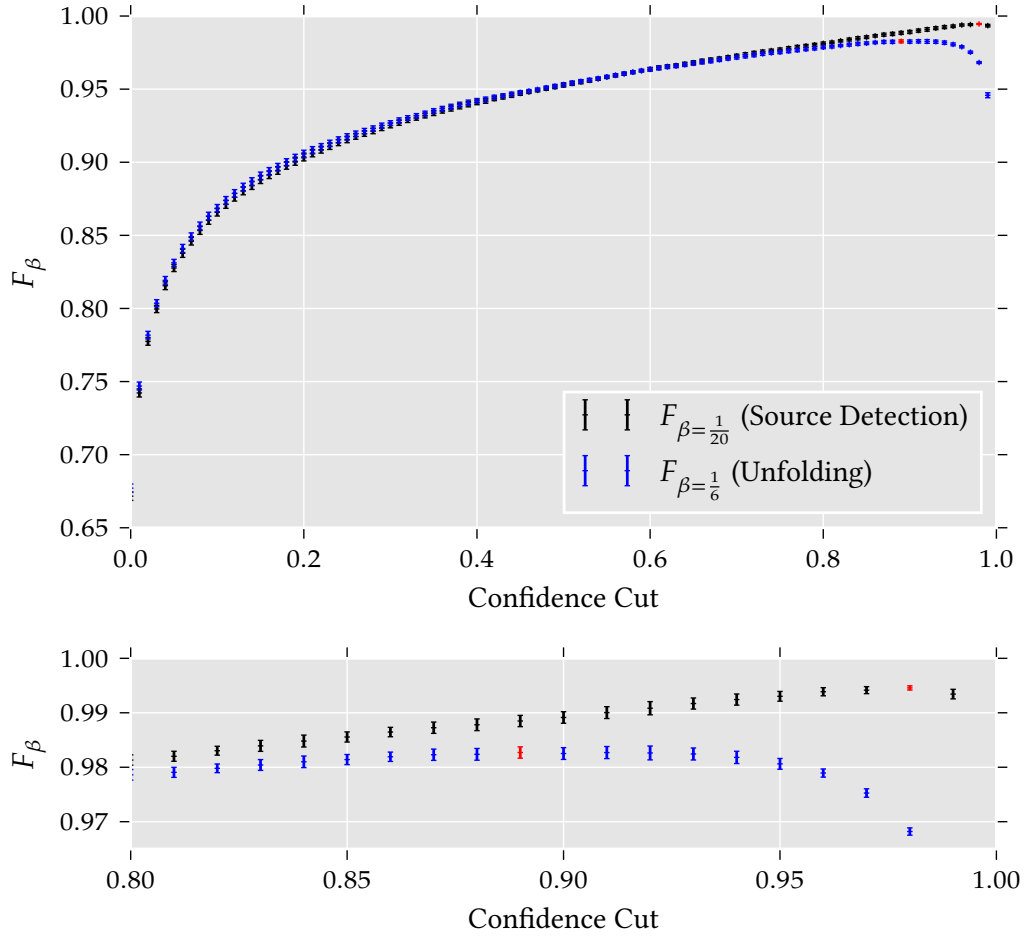


**Figure 6.11:** Receiver operating characteristics curve for the random forest used in the background suppression. The area under the roc curve is 0.9927.

For choosing the CONFIDENCE cut for the background suppression the  $F_\beta$ -factor [55]

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}} \quad (6.10)$$

is used. For the detection of a very high energy gamma ray source the  $F_\beta$ -score for  $\beta = 1/20$  is calculated. For the unfolding of the energy spectrum a larger number of events is helpful, whereas the requirement to the precision in the event set can be lowered. Hence for the unfolding event set the  $F_\beta$ -score for  $\beta = 1/6$  is calculated. Figure 6.12 shows both  $F_\beta$ -scores. The relevant range is shown in more detail below. The CONFIDENCE cuts for both sets are chosen so that the corresponding  $F_\beta$ -score has its maximum. The chosen values are marked in red in figure 6.12. Table 6.6 lists the performance values of the classification for the two event sets.



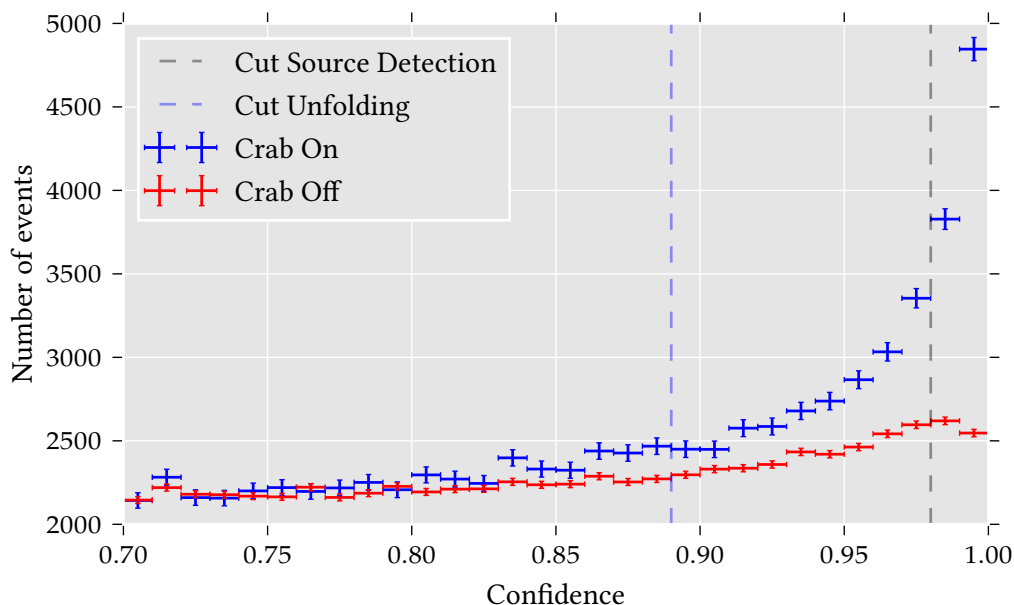
**Figure 6.12:**  $F_\beta$ -score in dependency of the applied CONFIDENCE cut. Shown are  $\beta = \frac{1}{6}$  and  $\beta = \frac{1}{20}$ . The relevant range of CONFIDENCE cuts is shown in more detail. The CONFIDENCE cuts with maximum of  $F_\beta$  are marked in red.

**Table 6.6:** Performance values for the two different CONFIDENCE cuts, chosen for the two event sets.

Performance value	Source Detection Event Set	Unfolding Event Set
Confidence cut	0.98	0.89
Recall	$0.472 \pm 0.005$	$0.792 \pm 0.005$
False positive rate	$0.0013 \pm 0.0002$	$0.009 \pm 0.001$
Precision	$0.9973 \pm 0.0004$	$0.989 \pm 0.001$
$F_\beta$	$0.9946 \pm 0.0004$	$0.983 \pm 0.001$

## 6.5 Source Detection of the Crab Nebula

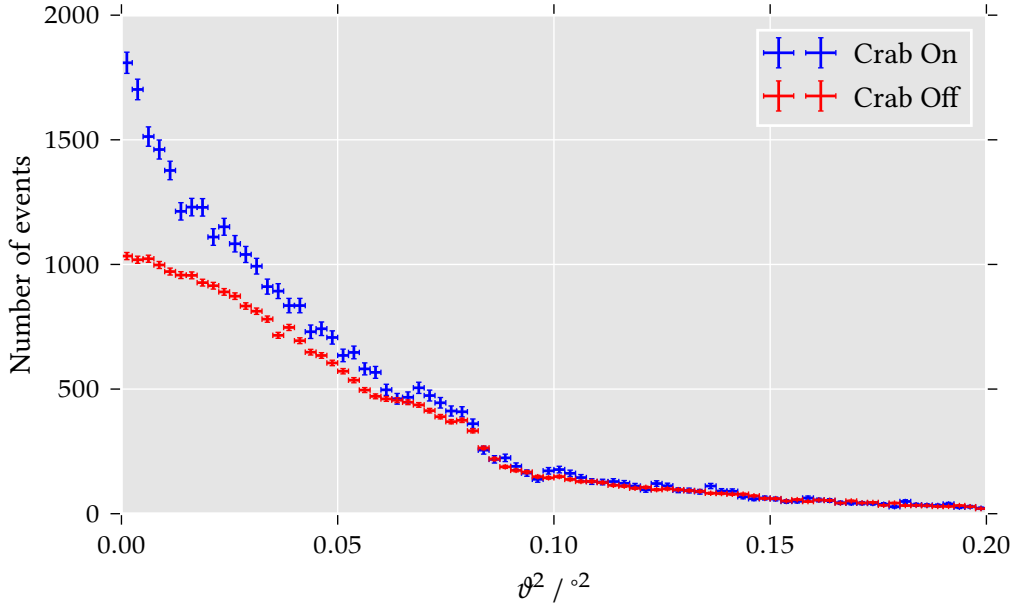
After performing the random forest classification each event is assigned to the ON or OFF data sets with a corresponding CONFIDENCE value (see section 3.3). Figure 6.13 shows the CONFIDENCE distributions for both data sets. The CONFIDENCE cuts chosen for the source detection and the unfolding are marked with vertical dashed lines. A clear excess of the ON data set over the OFF data set is visible for both cuts.



**Figure 6.13:** CONFIDENCE distribution for ON and OFF data of the Crab Nebula. A clear excess of the ON data set over the OFF data set for the two marked CONFIDENCE cuts is visible.

In common analysis of very high energy gamma ray sources with imaging air Cherenkov telescopes the source detection is based on a cut in  $\vartheta^2$ . In this analysis  $\vartheta$  is included in the parameters used by the random forest classifier and the CONFIDENCE cut replaces the  $\vartheta^2$ -cut. For comparison the  $\vartheta^2$  distributions for both CONFIDENCE cuts are shown in figure 6.14 (CONFIDENCE cut 0.89) and figure 6.15 (CONFIDENCE cut 0.98). In all three figures (6.13, 6.14 and 6.15) the ON and OFF data event sets merge for low CONFIDENCE values, high  $\vartheta^2$  values respectively, as expected.

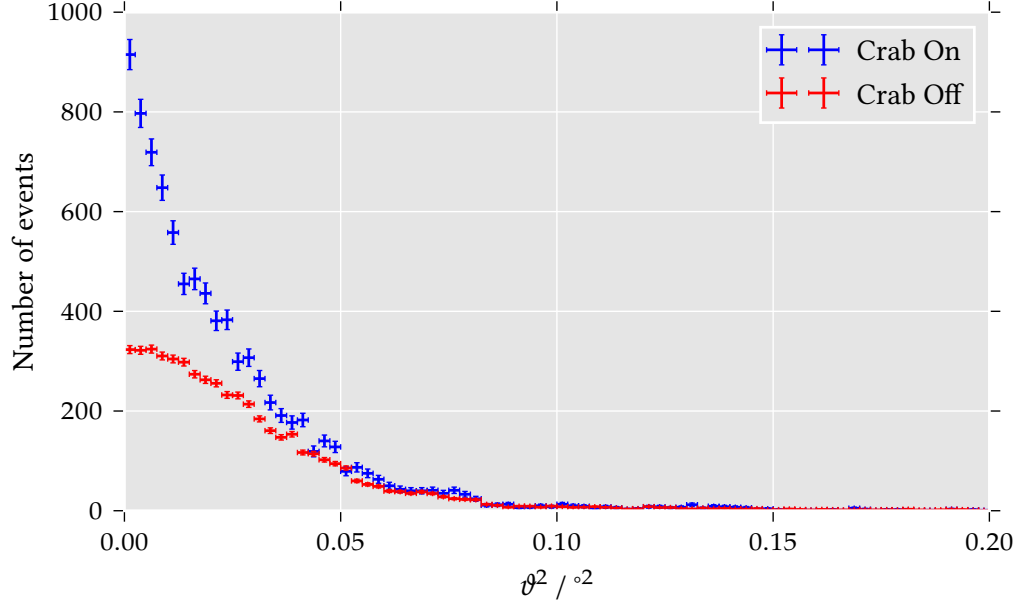
To calculate the significance of the detection the significance after [41] is used.



**Figure 6.14:**  $\chi^2$  distribution for ON and OFF data of the Crab Nebula after applying a CONFIDENCE cut of 0.89 (used in Unfolding).

$$\text{Significance} = \sqrt{2} \cdot \left[ N_{\text{ON}} \cdot \ln \left( \frac{(1 + \alpha) \cdot N_{\text{ON}}}{\alpha \cdot (N_{\text{ON}} + N_{\text{OFF}})} \right) + N_{\text{OFF}} \cdot \ln \left( \frac{(1 + \alpha) \cdot N_{\text{OFF}}}{N_{\text{ON}} + N_{\text{OFF}}} \right) \right]^{\frac{1}{2}} \quad (6.11)$$

Table 6.7 lists the number of events and the significance of the detection of the signal of the Crab Nebula for both sets. The number of excess events  $N_{\text{exc}} = N_{\text{ON}} - \alpha N_{\text{OFF}}$  is also listed.



**Figure 6.15:**  $\chi^2$  distribution for ON and OFF data of the Crab Nebula after applying a CONFIDENCE cut of 0.98 (used in source detection).

**Table 6.7:** Number of events and significance of the detection of the signal of the Crab Nebula for the source detection event set and the unfolding event set.

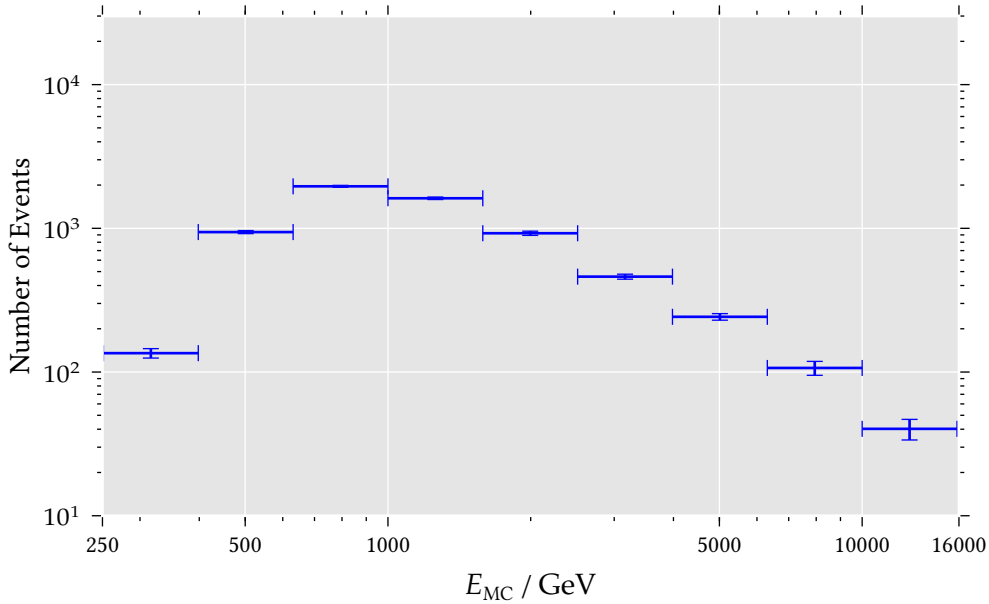
	Source detection	Unfolding
$N_{\text{on}}$	8674	33 405
$N_{\text{off}}$	25 831	134 711
$N_{\text{exc}}$	3507.8	6462.8
Significance ( $\alpha = 0.2$ ) / $\sigma$	39.89	34.41
$T_{\text{Obs.}} / \text{h}$		87.63
Sensitivity / $\sigma/\sqrt{\text{h}}$	4.26	3.68



## 6.6 Unfolding of the Energy Spectrum of the Crab Nebula

### 6.6.1 Preparation of the Unfolding

For the reconstruction of the energy spectrum of the Crab Nebula the unfolding software TRUUEE [43] is used (section 5.5). As a preparation for the unfolding, the binning of the unfolded energy spectrum is determined. As the energy resolution is around 20% for most of the energy range (see section 7.3.1) the bin width for the unfolded energy spectrum is chosen to be 20% or 0.2 in a ten-based logarithmic scale. To choose the energy range a 10 times bootstrapped subset of  $N_{\text{exc}} = 6463$  gamma events is drawn. Figure 6.16 shows the distribution of  $E_{\text{MC}}$  with the mean and the uncertainty from the bootstrapping.

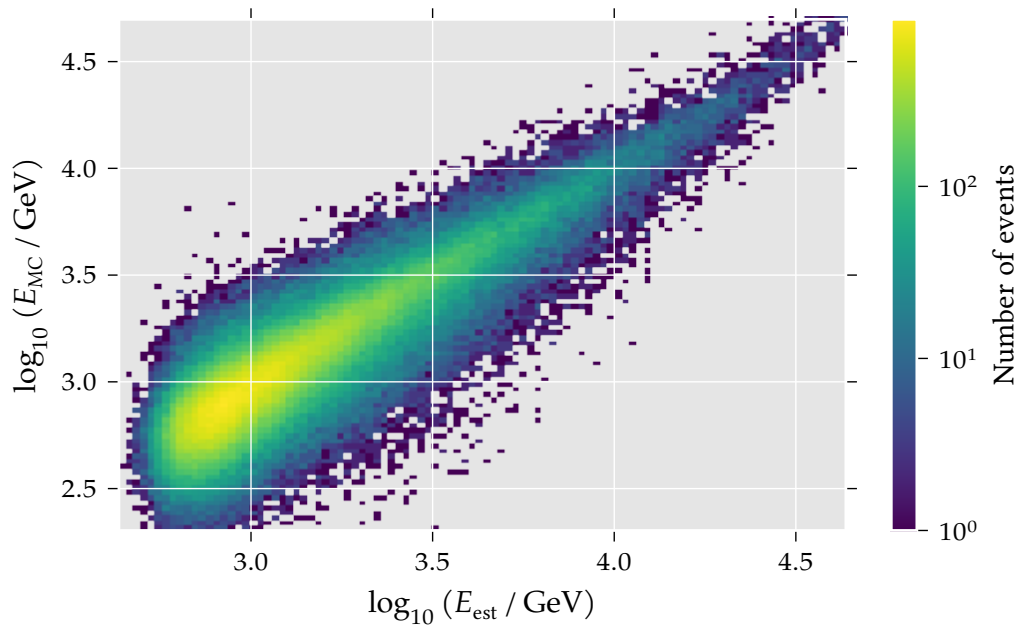


**Figure 6.16:** Distribution of the MC energy of 6436 gamma events. The number of events is the same as the number of excess events. The mean and the uncertainty is evaluated by bootstrapping. It is clearly visible that each energy bin is populated with a sufficient number of events to unfold the energy spectrum.

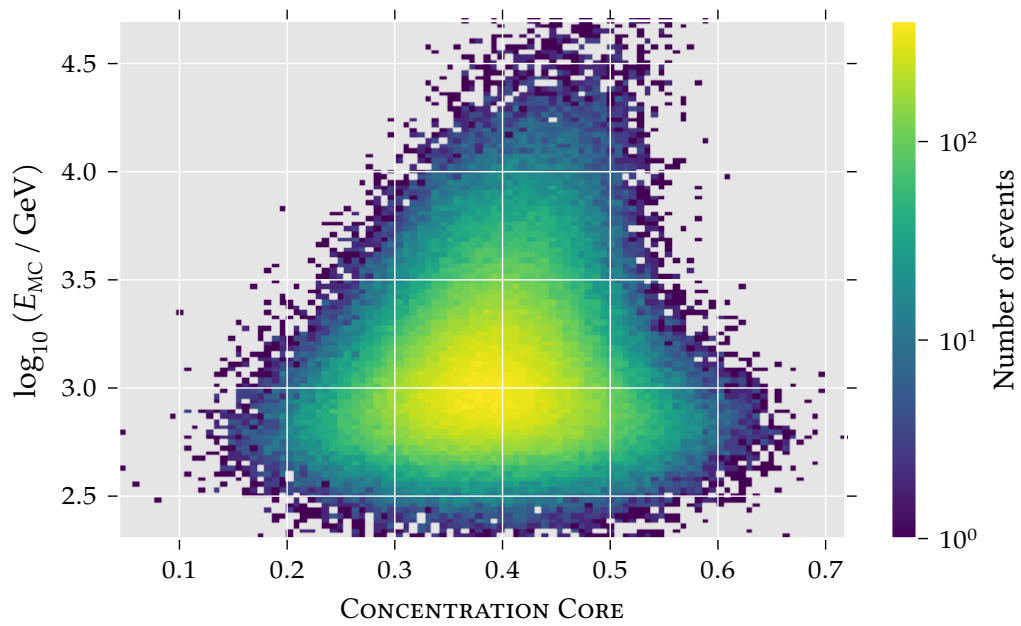
In each bin between 250 GeV and 16 TeV there are at least 30 events which is enough to unfold the corresponding bins. This leads to 9 bins with a bin width of 0.2 in ten-based logarithmic scale between 250 GeV and 16 TeV.

The observables used in the unfolding should have a clear correlation to the true

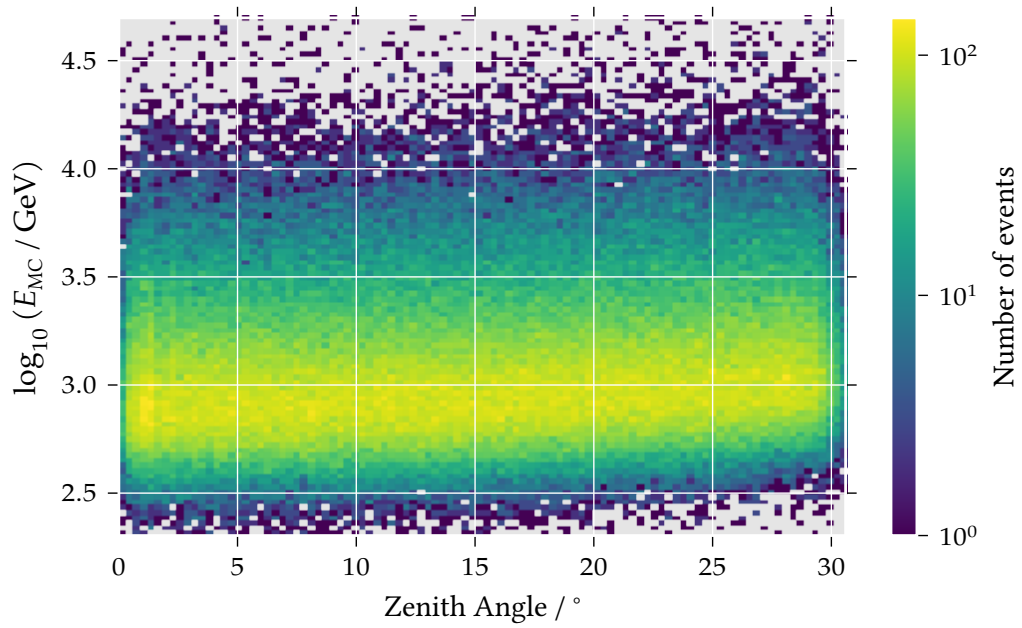
energy  $E_{\text{MC}}$ . The correlation of the chosen observables to the true energy are shown in the figures 6.17, 6.18 and 6.19. The best correlation with  $E_{\text{MC}}$  can be found with the estimated energy  $E_{\text{MC}}$ . Also the CONCENTRATION CORE shows a correlation to  $E_{\text{MC}}$ . In addition it is relative independent to  $E_{\text{est}}$ , though the CONCENTRATION CORE adds additional information to the unfolding of the true energy distribution. The last observable used is the zenith angle. Although it does not show a correlation with the true energy directly, the correlation of the other observables to the true energy is different for different zenith angles. Taking the zenith angle into the unfolding process adds more information and improves the reconstruction of the energy spectrum.



**Figure 6.17:** Correlation between the estimated energy  $E_{\text{est}}$  and the true energy  $E_{\text{MC}}$ . There is a clear linear dependency making  $E_{\text{est}}$  suitable for the unfolding.



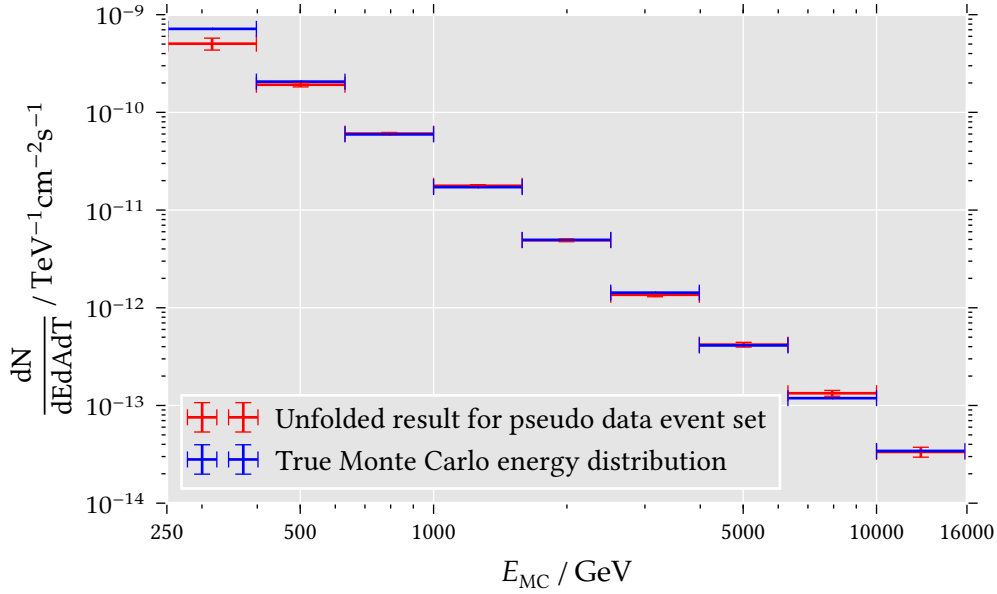
**Figure 6.18:** Correlation between the CONCENTRATION CORE and the true energy  $E_{MC}$ . There is a dependency visible, thus CONCENTRATION CORE adds additional information about the true energy to the unfolding.



**Figure 6.19:** Correlation between the zenith angle and the true energy  $E_{MC}$ . A direct dependency is not visible, but it is expected that the correlation between the true energy and the observables measured by the telescope depends on the zenith angle. Though taking the zenith angle into the unfolding improves the reconstruction of the energy spectrum.

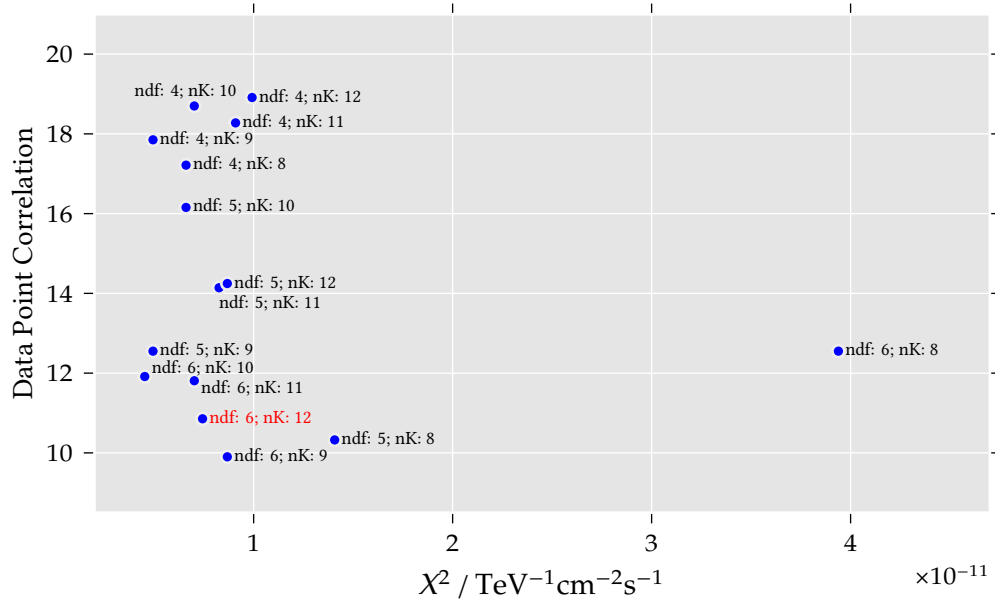
### 6.6.2 TRUEE Test Mode

In the test mode of TRUEE the gamma event set is separated into a pseudo data event set and a training event set. The response matrix is built using the training event set. The pseudo data event set is unfolded and the result is compared to the true Monte Carlo energy distribution. An example of one of this test unfoldings is shown in figure 6.20.



**Figure 6.20:** One example test unfolding with the unfolded result of a pseudo data event set and the Monte Carlo energy distribution of this event set.

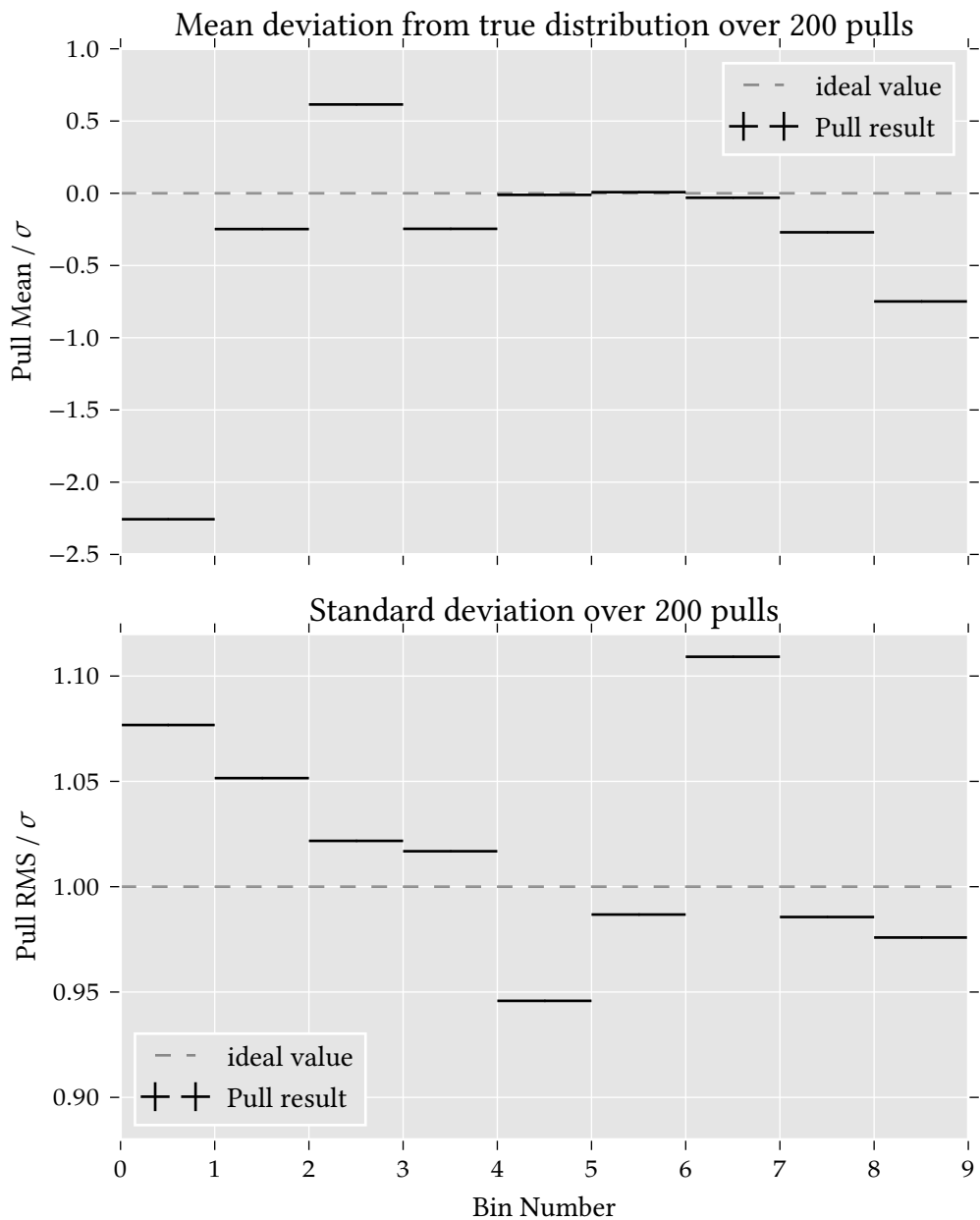
For different settings a test unfolding is performed and the  $X^2$  between the unfolded result and the true distribution and the data point correlation between the unfolded bins are calculated. Well performing settings have minimal values in both parameters. The settings are defined by the two parameters number of degrees of freedom (ndf) and number of knots (nK). In most cases the different settings build a l-shaped curve in the  $X^2$  - data point correlation parameter space. A setting on the left and lower edge of the curve should be chosen. Figure 6.21 shows the  $X^2$  - data point correlation parameter space for different settings. The chosen setting (number of degrees of freedom 6, number of knots 12) is marked in red.



**Figure 6.21:** So called L-Curve calculated by TRUEE. The different settings (ndf: number of degrees of freedom, nK: number of knots) are used in test mode with a pseudo data set. The  $\chi^2$  between the true energy distribution and the unfolded result and the data point correlation between the unfolded bins are calculated. Settings with minimal values in both parameters should be chosen. The setting used in this analysis is marked red.

### 6.6.3 TRUEE Pull Mode

In the pull mode of TRUEE the test unfolding of a pseudo data event set is performed several times (each time is called one pull) to check the stability of the evaluated performance. In this analysis 200 pulls were performed. For each energy bin the so called pull mean and pull RMS are calculated. The pull mean is the average over all pulls of the difference between the unfolded result and the true energy in units of the uncertainty of the unfolded results. The pull RMS is the standard deviation in units of the uncertainty of the unfolded results over all pulls. Both values are shown in figure 6.22. For a stable unfolding the pull mean should be zero and the pull RMS should be one. The pull RMS fulfills this condition very well. The pull mean is between  $-1\sigma$  and  $1\sigma$  for all bins except the first one. As the first bin is quite difficult to unfold, due to the small statistic in the training gamma event set and the typically small showers in this energy bin, a pull mean of  $-2.25\sigma$  for this bin is acceptable.

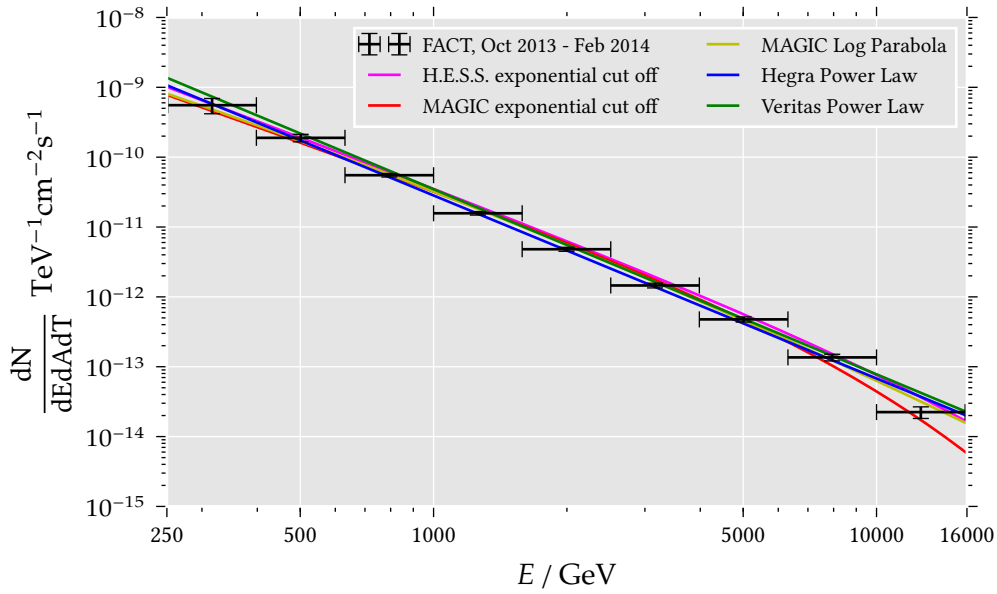


**Figure 6.22:** The results of the pull mode of TRUEE. The deviation of each individual pull result to the true distribution (in units of standard deviations  $\sigma$ ) is calculated and averaged over all pulls (top). With a perfect unfolding the average deviation (called pull mean) should be zero. The standard deviation (in units of the uncertainty, called pull RMS) of the unfolded results over all pulls is calculated (bottom). For a stable unfolding the the pull RMS should be one.



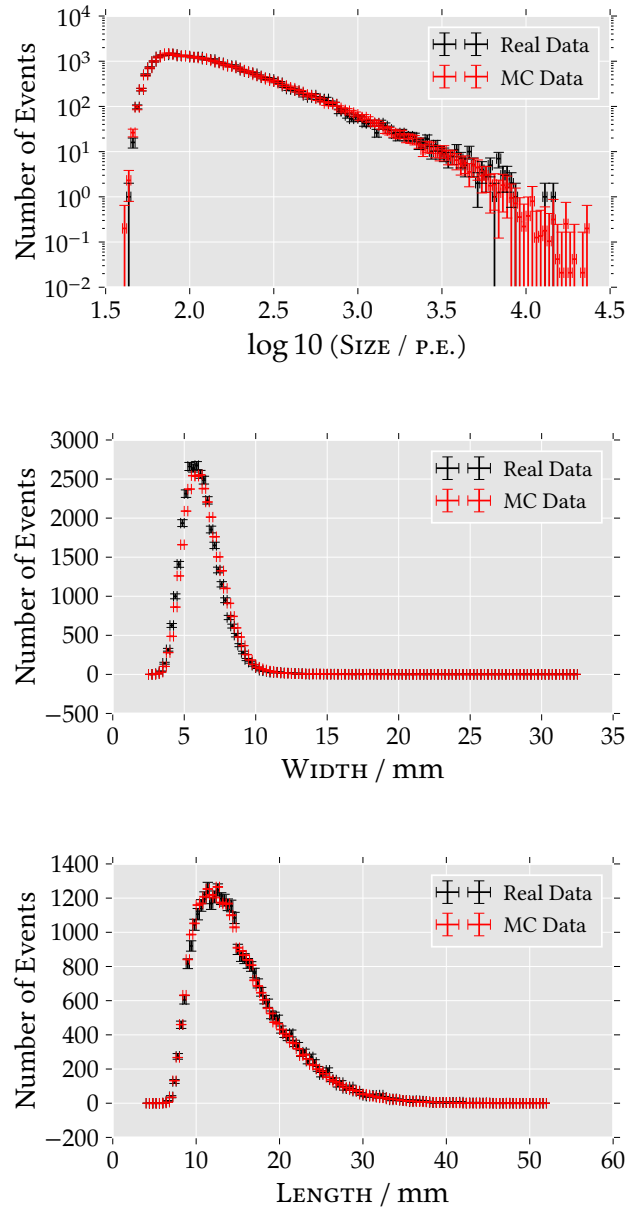
## 6.6.4 TRUEE Data Mode

In the data mode of TRUEE the real data event sets are unfolded. The OFF data set is used as a background data event set to subtract the background distributions from the ON data distributions. The result of the unfolding, the energy spectrum of the Crab Nebula measured by FACT between October 2013 and February 2014, is shown in figure 6.23. For comparison several fluxes measured by other experiments are also displayed. The FACT results are in good agreement with the results of the other experiments within the uncertainties. The measured fluxes for each energy bin and the corresponding uncertainties are listed in table 6.8.



**Figure 6.23:** The unfolded energy spectrum of the Crab Nebula, measured by FACT between October 2013 and February 2014. For comparison the fits of the Crab Nebula from measurements of HEGRA [5], MAGIC [7], Veritas [9] and H.E.S.S [4] are shown in addition.

To check the correctness of the unfolding TRUEE offers so called check unfolding tests. Thereby distributions of parameters not used in the unfolding are compared between gamma events and data events. The gamma distributions are reweighted according to the unfolded energy distribution of the data events. The distributions should be in agreement. Figure 6.24 shows the check unfolding plots for the parameters SIZE, WIDTH and LENGTH. A small mismatch for the parameter WIDTH is visible, SIZE and LENGTH are in good agreement.



**Figure 6.24:** Distribution of the parameters SIZE, WIDTH and LENGTH for reweighted gamma events and data events. A proper unfolding is indicated by the agreement of the distributions.

**Table 6.8:** Overview over the unfolded energy spectrum of the Crab Nebula measured by FACT.

Bin	$E_{\min} / \text{GeV}$	$E_{\max} / \text{GeV}$	$\frac{dN}{dE dA dT}$	$\text{TeV}^{-1} \text{cm}^{-2} \text{s}^{-1}$
0	251	398	$(5.55 \pm 1.37) \times 10^{-10}$	
1	398	631	$(1.89 \pm 0.23) \times 10^{-10}$	
2	631	1000	$(5.52 \pm 0.32) \times 10^{-11}$	
3	1000	1585	$(1.57 \pm 0.09) \times 10^{-11}$	
4	1585	2512	$(4.81 \pm 0.32) \times 10^{-12}$	
5	2512	3981	$(1.46 \pm 0.11) \times 10^{-12}$	
6	3981	6310	$(4.76 \pm 0.40) \times 10^{-13}$	
7	6310	10 000	$(1.36 \pm 0.15) \times 10^{-13}$	
8	10 000	15 849	$(2.24 \pm 0.42) \times 10^{-14}$	



## 7 Physics Performance of FACT

### 7.1 Event Acceptance of Telescope and Analysis

One of the performance values of the telescope and the analysis is the acceptance of events for each analysis step. Table 7.1 lists the accepted number of events for the data sample and the simulated proton and gamma samples for the individual steps of the analysis chain.

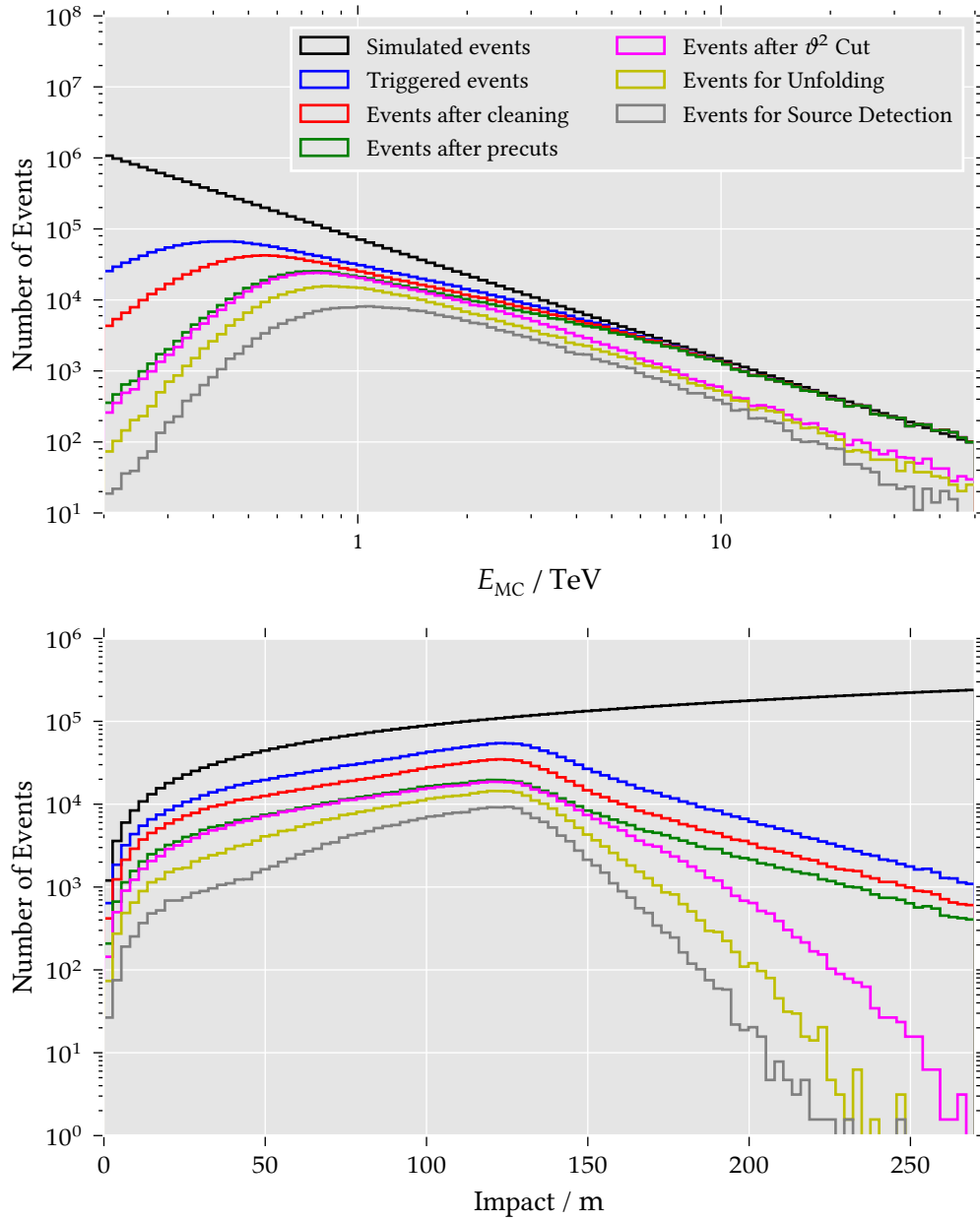
**Table 7.1:** Number of events after the different steps of the analysis chain for the data sample, for the simulated proton sample and for the simulated gamma sample

Number events after	Simulation	Trigger	Cleaning
Data		21 317 026	6 915 916
Proton	780 046 520	509 652	299 858
Gamma	12 000 000	1 915 336	1 183 121
Number events after	Quality cuts	CONFIDENCE cut	
		Unfolding	Source Detection
Data On		33 405	8674
Data Off	3 594 347	134 711	25 831
Proton	188 283		
Gamma	692 836	216 528	115 704

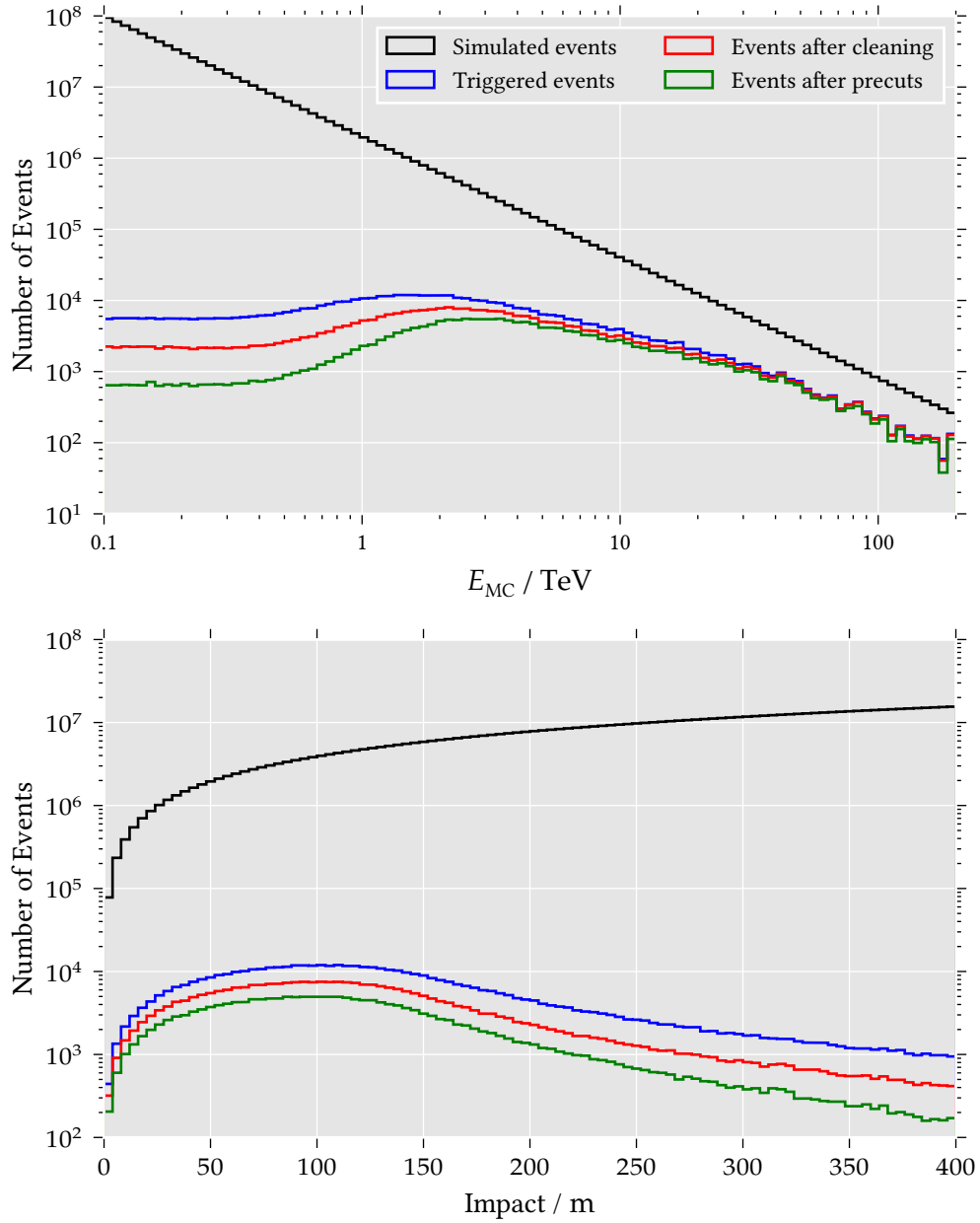
With the help of simulated events the acceptance in dependency of the true energy and the true impact of the primary particle can be evaluated. Figure 7.1 shows these dependencies for gamma events, whereas figure 7.2 shows it for proton events.

For the source detection and the unfolding gamma sets, the energy distribution and the impact distribution are covered completely by the simulated energy range and impact range. It is also visible that gamma events with an energy larger than about 2 TeV are lost between the precuts set and the  $\vartheta^2$  Cut set.

For both proton and gamma events the acceptance of low energy events decreases with decreasing energy. In addition also the acceptance of events with high impact decreases with increasing impact.



**Figure 7.1:** Distributions of the energy and the impact of the primary particle of the simulated gamma events after the different analysis steps.



**Figure 7.2:** Distributions of the energy and the impact of the primary particle of the simulated proton events after the different analysis steps.

## 7.2 Effective Collection Area

A more formal description of the acceptance of very high energy gamma rays by the telescope and the analysis chain can be given with the effective collection area  $A_{\text{eff}}$ . The effective collection area describes a fictive area for which the telescope is sensitive to any very high energy gamma ray flux going through this area. The interpretation of  $A_{\text{eff}}$  is independent of the telescope and the analysis chain, making it possible to compare the performance to other experiments.  $A_{\text{eff},i}$  is given by the ratio of accepted events  $N_i$  to the total number of simulated events  $N_{\text{sim},i}$  per energy bin  $i$ :

$$A_{\text{eff},i} = \frac{N_i}{N_{\text{sim},i}} A_{\text{sim}} \quad (7.1)$$

with

$$N_{\text{sim},i} = N_{\text{sim}} \frac{\int_{E_{\text{low},i}}^{E_{\text{high},i}} E^{-\gamma} dE}{\int_{E_{\text{max}}}^{E_{\text{min}}} E^{-\gamma} dE} \quad (7.2)$$

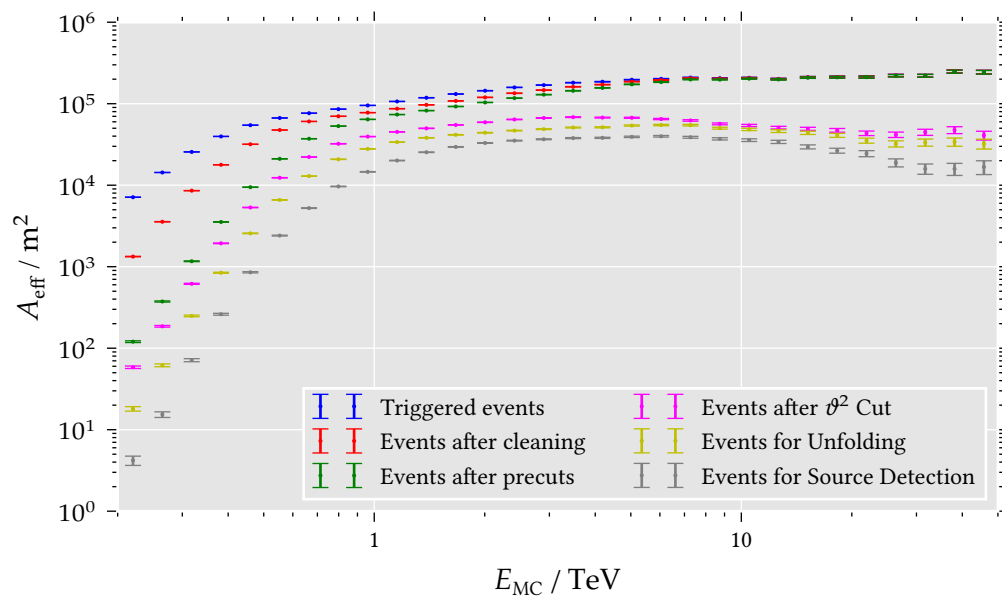
$$A_{\text{sim}} = \pi \cdot \text{impact}_{\text{sim,max}}^2 \quad (7.3)$$

follows for  $A_{\text{eff},i}$ :

$$A_{\text{eff},i} = \frac{N_i}{N_{\text{sim}}} \cdot \frac{(E_{\text{max}}^{-\gamma+1} - E_{\text{min}}^{-\gamma+1})}{(E_{\text{high},i}^{-\gamma+1} - E_{\text{low},i}^{-\gamma+1})} \cdot \pi \cdot \text{impact}_{\text{sim,max}}^2 \quad (7.4)$$

With the simulation parameters from table 6.2 the effective collection area can be calculated for all steps of the analysis chain. The result is shown in figure 7.3.





**Figure 7.3:** Effective collection area  $A_{\text{eff}}$  in dependency of the true energy  $E_{\text{MC}}$  for the different analysis steps.

## 7.3 Energy Dependent Performance

The energy dependent performance of the telescope and the analysis chain is described by the energy bias, the energy resolution and the energy threshold. The energy spectrum of the Crab Nebula gives information about the performance of the telescope and the analysis chain since the Crab Nebula is considered to be the "standard candle" in very high energy gamma ray astrophysics. Therefore the energy bias and energy resolution will be evaluated in section 7.3.1. Next the energy threshold for the different analysis steps is evaluated (section 7.3.2). Finally the energy spectrum of the Crab Nebula from section 6.6 is evaluated. Three power law like functions are fitted to the flux of the Crab Nebula (section 7.3.3).

### 7.3.1 Energy Bias and Energy Resolution of FACT

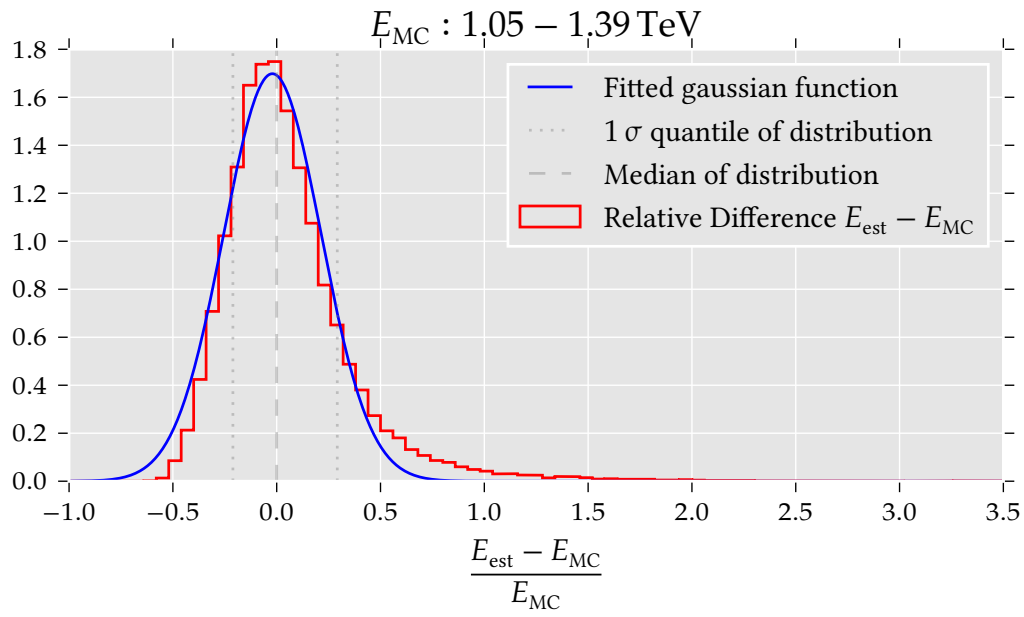
To evaluate the energy bias and energy resolution of FACT the correlation between the estimated energy  $E_{\text{est}}$  and the true energy  $E_{\text{MC}}$  for gamma events is investigated. The correlation is shown in figure 6.6 for all events which survive the cleaning and the quality cuts and in figure 6.7 for all events which survive the CONFIDENCE cut used for the unfolding.

For each bin in  $E_{\text{MC}}$  the distribution of the relative difference between  $E_{\text{MC}}$  and  $E_{\text{est}}$  ( $\frac{E_{\text{est}} - E_{\text{MC}}}{E_{\text{MC}}}$ ) is calculated. In literature (for example in [8]) the energy bias and energy resolution are determined by fitting a gaussian function to the distribution.

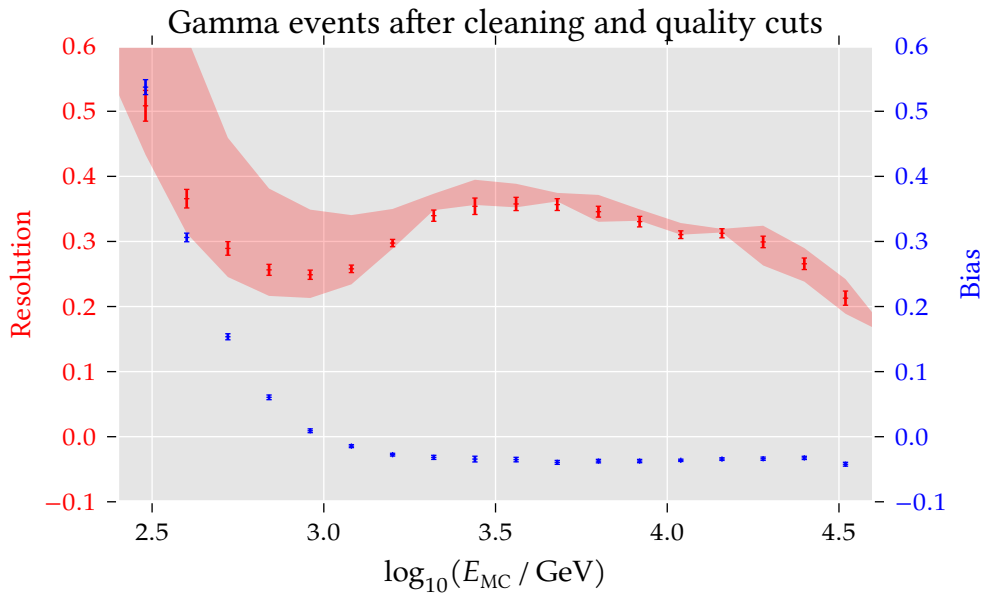
The mean of the fitted gaussian function is interpreted as the energy bias, the standard deviation of the gaussian function is interpreted as the energy resolution. Nevertheless the distribution has non-gaussian tails and the fitted function is not sensitive to these tails. Therefore also the differences between the median of the distribution and the 15.865% quantile and the 84.135% quantile are calculated. These quantiles represent the  $1\sigma$  regions around the median. The differences are interpreted as a band estimation for the energy resolution. Both the fitted gaussian function and the median and the  $1\sigma$  quantiles are shown for one bin in  $E_{\text{MC}}$  in figure 7.4.

Figure 7.5 shows the calculated energy bias and energy resolution for gamma events which survived the cleaning and the quality cuts.

The energy bias and energy resolution for the unfolding events set, though after the application of the CONFIDENCE cut for the unfolding, is shown in figure 7.6. The bias and resolution is improved after applying the CONFIDENCE cut. A large positive energy bias can be seen for low energies. This is already described in section 6.3.

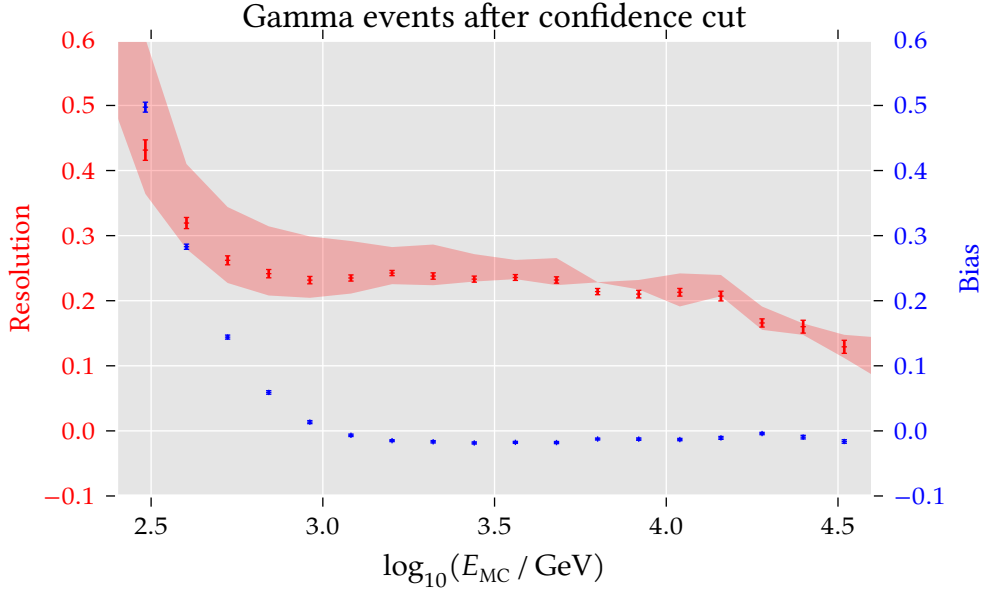


**Figure 7.4:** Distribution of the relative difference  $\frac{E_{est} - E_{MC}}{E_{MC}}$  for events with  $E_{MC}$  between 1.05 TeV and 1.39 TeV. A gaussian function is fitted to the distribution. The mean and the standard deviation of the function are interpreted as the energy bias and energy resolution in this energy bin. As the function is not sensitive to the non-gaussian tails of the distribution also the median and the 1  $\sigma$  quantiles are calculated. Shown are events after the CONFIDENCE cut for unfolding.



**Figure 7.5:** Energy bias and energy resolution for gamma events which survived the cleaning and the quality cuts. The values are evaluated by the mean and standard deviation of a gaussian function fitted to the relative difference between  $E_{\text{est}}$  and  $E_{\text{MC}}$  in each energy bin in  $E_{\text{MC}}$ . The band for the energy resolution is evaluated by calculating the difference between the median and the  $1\sigma$  quantiles of the distributions of the relative difference.

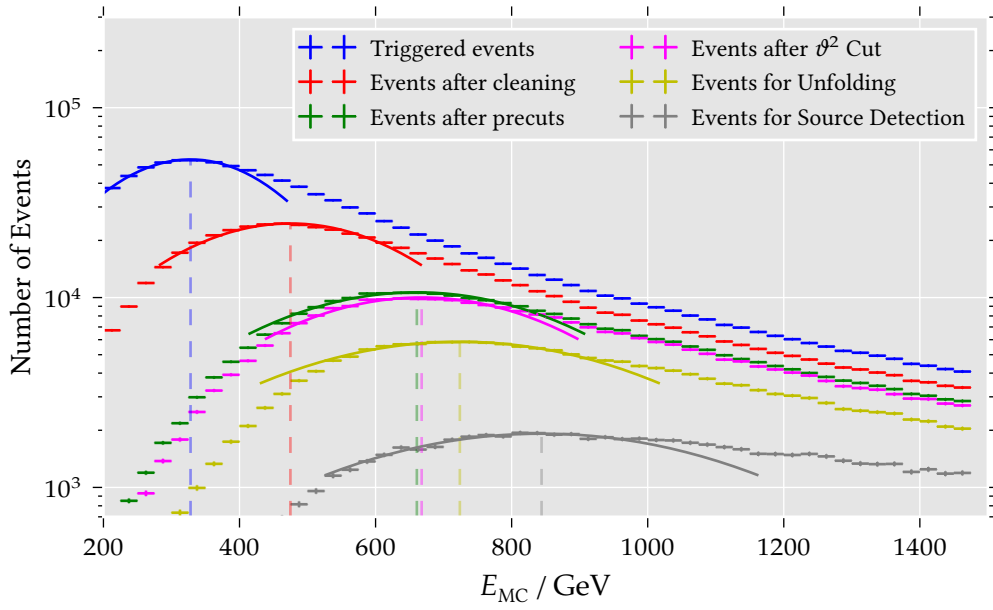
Around 1 TeV the bias decreases to almost zero and converges to a small negative value of  $\approx -0.015$ . The energy resolution reaches roughly 22% at around 600 GeV. With energies from 6 TeV up the resolution further improves to 20% and finally down to 13% for the highest energies. The energy bias and energy resolution contain all effects of the telescope and the analysis chain.



**Figure 7.6:** Energy bias and energy resolution for gamma events after applying the CONFIDENCE cut for unfolding. The values are evaluated by the mean and standard deviation of a gaussian function fitted to the relative difference between  $E_{\text{est}}$  and  $E_{\text{MC}}$  in each energy bin in  $E_{\text{MC}}$ . The band for the energy resolution is evaluated by calculating the difference between the median and the  $1\sigma$  quantiles of the distributions of the relative difference.

### 7.3.2 Energy Threshold

The energy threshold is defined as the maximum of the distribution of the true energy  $E_{\text{MC}}$ . To calculate this maximum a gaussian function is fitted to the distribution, the mean of the function is taken as the energy threshold. While the energy threshold is an easy value to describe the performance of the telescope and the analysis chain, the more complex shape of the distribution has to be taken into account to fully investigate the performance. Figure 7.7 shows the distributions of  $E_{\text{MC}}$  for the different analysis steps.



**Figure 7.7:** Distribution of the true energy  $E_{MC}$  for the different analysis steps. The energy threshold is defined as the maximum of this distribution. It is determined by fitting a gaussian function to the distribution and taking the mean of the function as the energy threshold.

The fitted gaussian function and the mean of the gaussian function are marked, too. Table 7.2 lists the energy thresholds for the different analysis steps. The standard deviation  $\sigma$  of the fitted gaussian function is listed as an estimator for the width of the energy distribution.

**Table 7.2:** Energy thresholds for the different analysis steps. Also the standard deviation  $\sigma$  of the fitted gaussian function is listed as an estimator for the width of the energy distribution.

Analysis Step	Energy threshold / GeV	$\sigma$ / GeV <sup>2</sup>
Triggered events	$328.0 \pm 1.7$	$141.4 \pm 3.5$
Events after cleaning	$474.8 \pm 2.9$	$191.5 \pm 5.2$
Events after precuts	$660.7 \pm 3.3$	$245.7 \pm 8.2$
Events after $\vartheta^2$ Cut	$667.8 \pm 4.0$	$228.7 \pm 7.6$
Events for Unfolding	$723.9 \pm 3.5$	$292.1 \pm 10.4$
Events for Source Detection	$844.0 \pm 7.7$	$316.9 \pm 19.4$

### 7.3.3 Energy Spectrum of the Crab Nebula

The energy spectrum of the Crab Nebula is reconstructed using a Tikhonov Regularized unfolding with the software TRUEE as described in section 6.6. The achievable energy spectrum for the 87.63h long data set is shown in figure 6.23. It ranges from 250 GeV to 16 TeV and is in good agreement with measurements of other experiments within the uncertainties.

In literature the differential energy spectrum of the Crab Nebula is described by different power-law like functions. A simple power law (7.5) is used in [5] and [9], a power law with an exponential cut off (7.6) in [7] and [4] and a log-parabola function (7.7) in [7].

$$\frac{dN}{dEdAdT} = f_0 \cdot \left(\frac{E}{E_0}\right)^{-\alpha} \quad (7.5)$$

$$\frac{dN}{dEdAdT} = f_0 \cdot \left(\frac{E}{E_0}\right)^{-\alpha} \exp\left(-\frac{E}{E_C}\right) \quad (7.6)$$

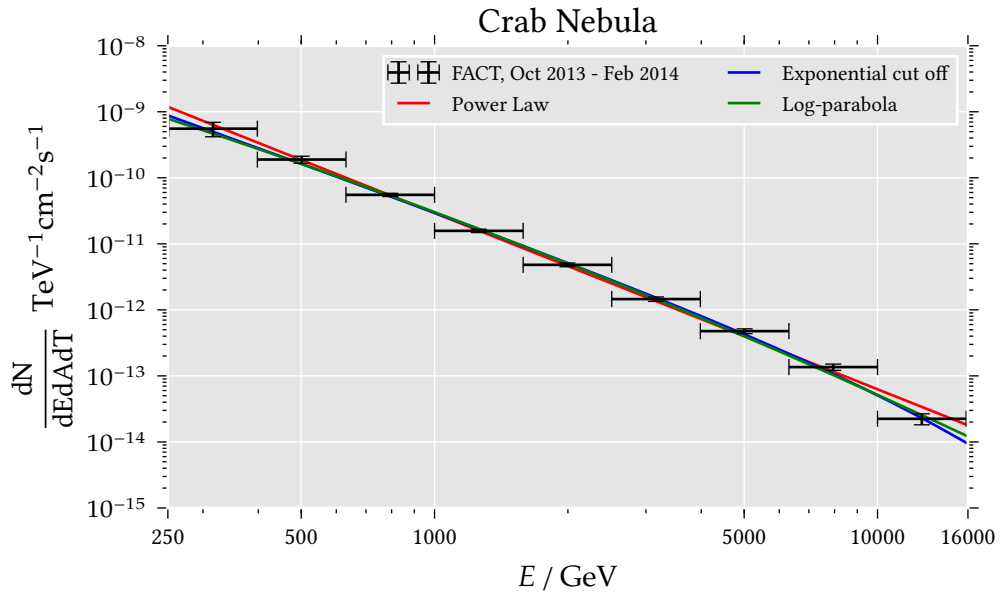
$$\frac{dN}{dEdAdT} = f_0 \cdot \left(\frac{E}{E_0}\right)^{-\alpha + \beta \log\left(\frac{E}{E_0}\right)} \quad (7.7)$$

$(E_0 = 1 \text{ TeV})$

All functions are fitted to the unfolded data points  $\{\vec{x}, \vec{y}\}$ , using a least-square minimization of the square of the residual vector  $\vec{r} = \vec{y} - f(\vec{x}, \vec{a})$ . The uncertainty and the correlation of the data points are taken into account by using the inverse of the covariance matrix ( $\mathbf{Cov}$ ) calculated by TRUEE as the weight matrix. The minimization term is:

$$X^2(\vec{a}) = \vec{r} \times \mathbf{W} \times \vec{r}^T \quad (7.8)$$

with  $\mathbf{W} = \mathbf{Cov}^{-1}$ . The minimization of  $X^2$  yields the best fit parameters  $\vec{a}_0$ . The fitted functions and the unfolded data points are shown in figure 7.8.



**Figure 7.8:** The unfolded energy spectrum of the Crab Nebula measured by FACT. Several power law-like functions are fitted to the data points. The power law with exponential cut off and the log-parabola function describe the data well.

With the  $X^2$  and the number of degrees of freedom ( $ndf$ ) of the best fit parameters the probability of the fit ( $Q$ ) can be calculated [50]:

$$Q = 1 - P\left(\frac{ndf}{2}, \frac{X^2}{2}\right) \quad (7.9)$$



with the incomplete gamma function  $P(a, x)$ .  $X^2$ ,  $ndf$  and  $Q$  for the three fitted functions are:

$$\text{Power Law : } \quad X^2 = 29.8 \quad ndf = 7 \quad Q = 0.0001 \quad (7.10)$$

$$\text{Exponential cut off : } \quad X^2 = 10.6 \quad ndf = 6 \quad Q = 0.1024 \quad (7.11)$$

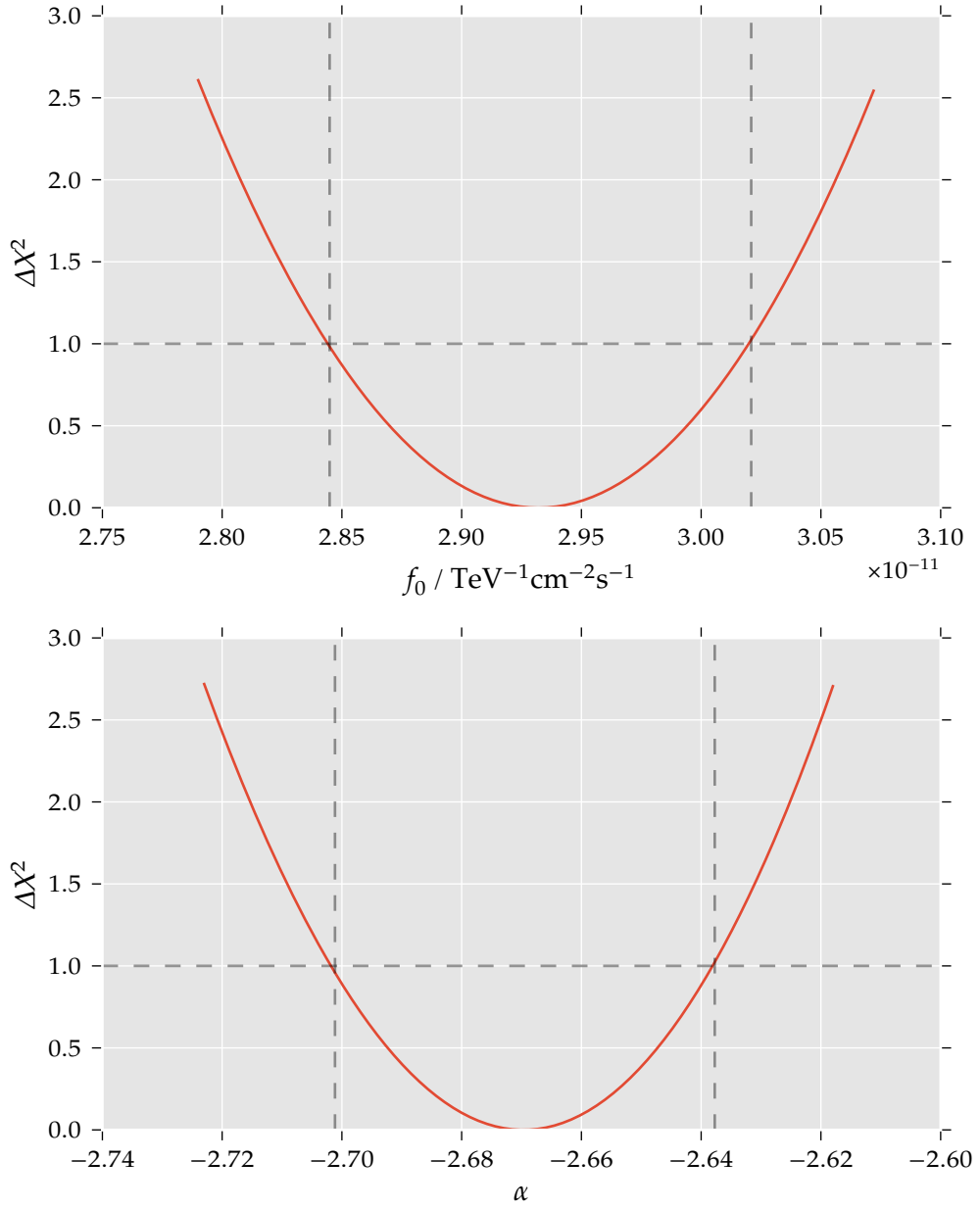
$$\text{Log-Parabola : } \quad X^2 = 13.6 \quad ndf = 6 \quad Q = 0.0345 \quad (7.12)$$

The power law function is rejected with a very low fit probability. The best fitting function is the exponential function with cut off. The probability for both the exponential function and the log-parabola function are quite low, due to the fact, that no systematic uncertainties are taken into account.

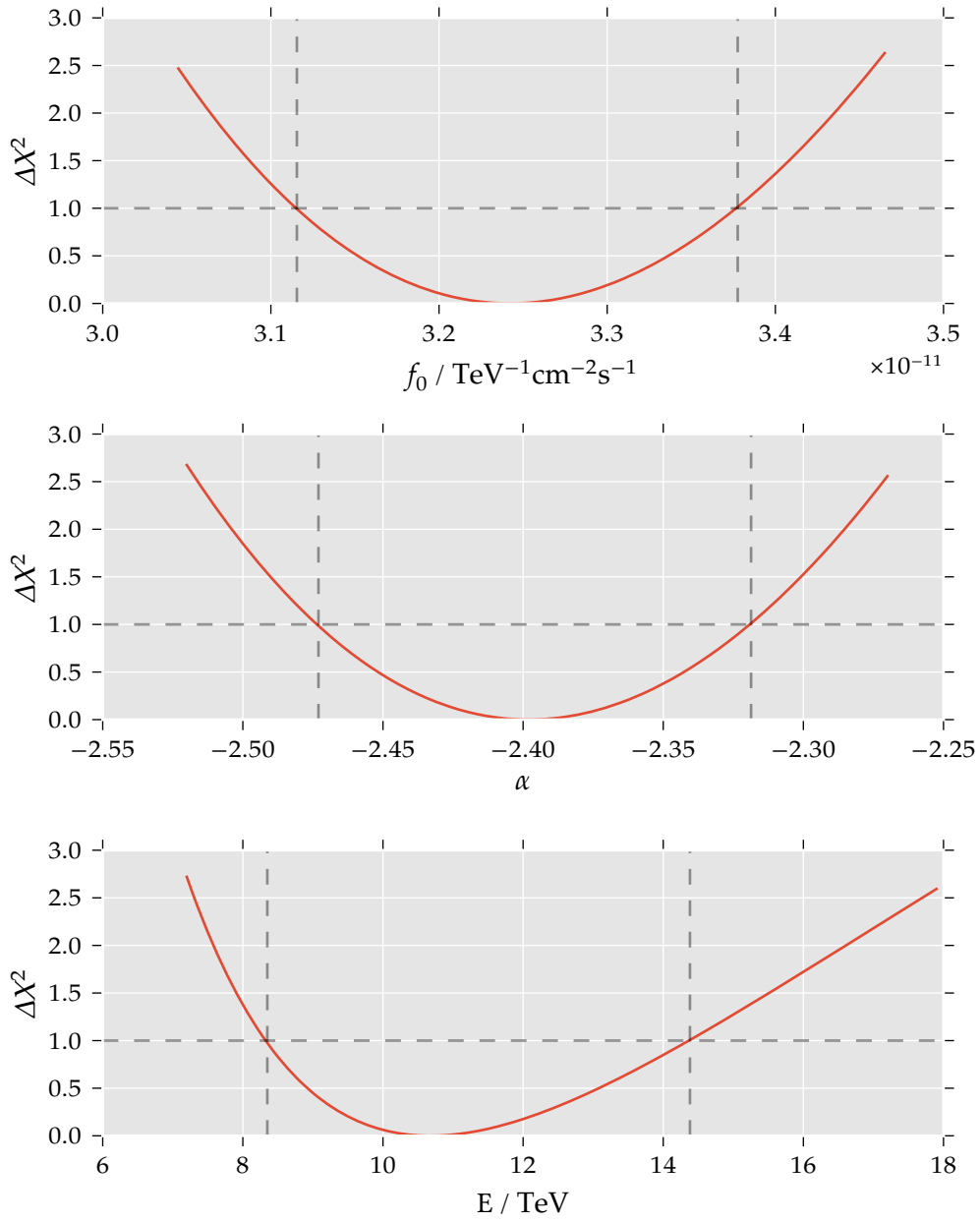
The uncertainties of the fitted parameters can be evaluated by performing a scan in the  $X^2$  parameter space. The scan is performed for each individual parameter  $j$ . While  $a_j$  is hold fix,  $X^2$  is minimized yielding into a parameter set  $\vec{a}_k$  with the fixed parameter  $a_j$ . The difference  $\Delta X^2 = X^2(\vec{a}_k) - X^2(\vec{a}_0)$  between the  $X^2$  of the best fit parameters and the new parameter set is calculated. The  $1\sigma$  confidence uncertainty of the parameter  $a_{j,0}$  results from the values of  $a_j$  where  $\Delta X^2 = 1$ . For more details about least-square minimization and uncertainty estimation of fit parameters see [50, pp. 656 - 698]. Figure 7.9 shows the scan for the parameters of the power law function, figure 7.10 for the parameters of the power law with an exponential cut off and figure 7.11 for the parameters of the log-parabola function. The values for  $\Delta X^2 = 1$  are marked with lines. The best fit values and their uncertainties of the parameters are listed in table 7.3.

**Table 7.3:** Fit parameters of the fitted functions to the unfolded differential energy spectrum. The best fit parameters and the uncertainties are listed.

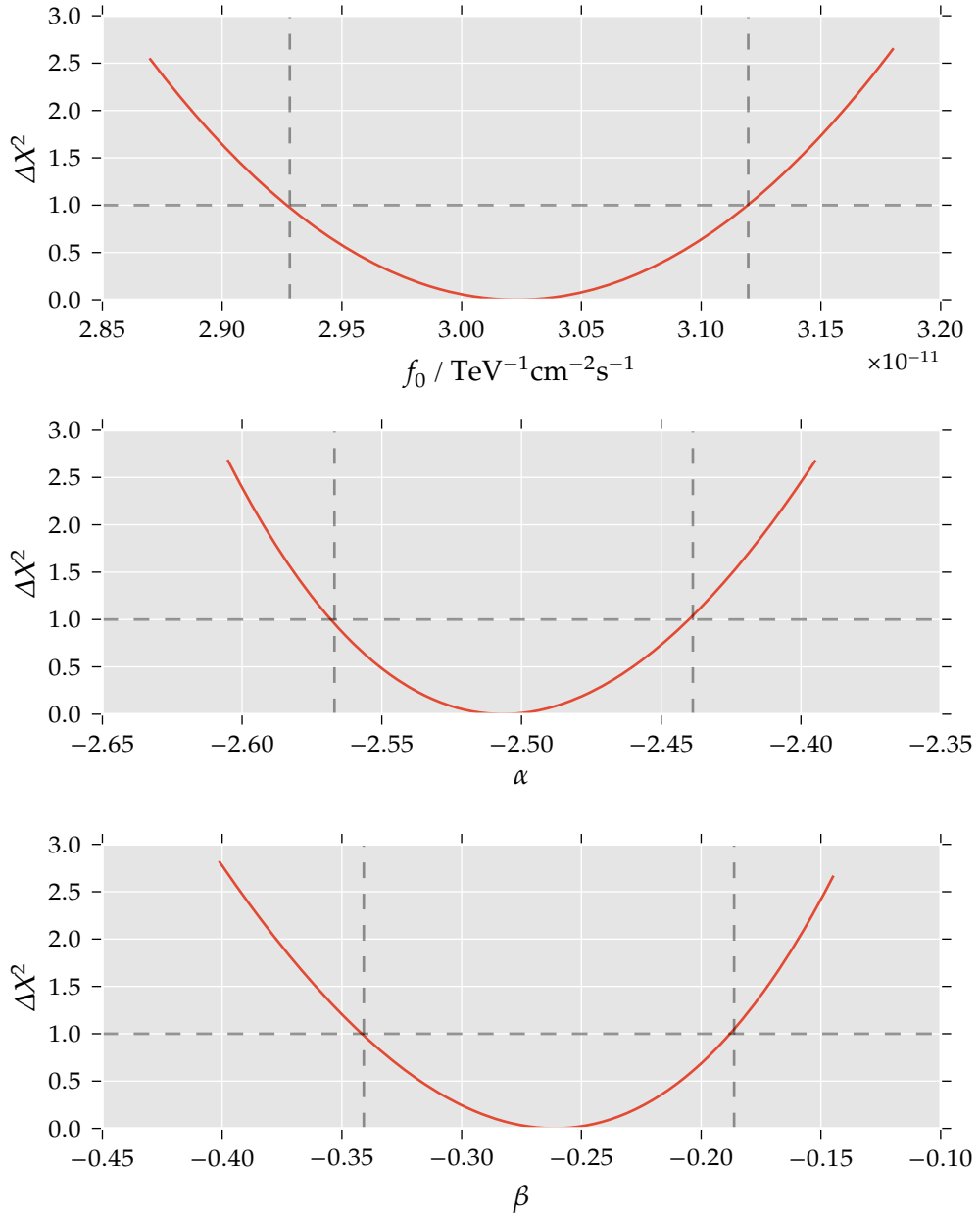
Function	$f_0 / \text{TeV}^{-1}\text{cm}^{-2}\text{s}^{-1}$	$\alpha$	$\beta$	$E_C / \text{TeV}$
Power Law	$(2.93_{-0.09}^{+0.09}) 10^{-11}$	$-2.67_{-0.03}^{+0.03}$	-	-
Exponential cut off	$(3.24_{-0.13}^{+0.14}) 10^{-11}$	$-2.40_{-0.07}^{+0.08}$	-	$10.68_{-2.33}^{+3.71}$
Log-parabola	$(3.02_{-0.09}^{+0.10}) 10^{-11}$	$-2.51_{-0.06}^{+0.07}$	$-0.26_{-0.08}^{+0.07}$	-



**Figure 7.9:** The distribution of  $\Delta X^2$  for each parameter of the power law function.  $\Delta X^2 = 1$  defines the  $1\sigma$  confidence region for the parameters.



**Figure 7.10:** The distribution of  $\Delta X^2$  for each parameter of the power law with an exponential cut off function.  $\Delta X^2 = 1$  defines the  $1\sigma$  confidence region for the parameters.



**Figure 7.11:** The distribution of  $\Delta X^2$  for each parameter of the log-parabola function.  $\Delta X^2 = 1$  defines the  $1\sigma$  confidence region for the parameter.

## 7.4 Sensitivity

The sensitivity  $S$  of a telescope and the analysis chain is defined as the flux value of a source which can be detected with a significance of  $5\sigma$  within 50 h of observation time. The significance  $S$  of a detection measured  $N_{\text{ON}}$  events in the ON-region and  $N_{\text{OFF}}$  events in the OFF-region with the norm factor  $\alpha$  between ON- and OFF-region is defined by:

$$S = \frac{N_{\text{ON}} - \alpha N_{\text{OFF}}}{\sqrt{N_{\text{ON}} + \alpha^2 N_{\text{OFF}}}} \quad (7.13)$$

The significance can be scaled to 50 h of observation time, by scaling the number of events  $N_i$  to 50 h:

$$S_{50 \text{ h}} = \sqrt{\frac{50 \text{ h}}{T_{\text{Obs.}}}} \cdot s \quad (7.14)$$

Now the relative sensitivity  $S_{\text{rel}}$  is the ratio of excess events  $N_{\text{excess}}$  for which  $S_{50 \text{ h}}$  is  $5\sigma$ :

$$N_{\text{excess}} = N_{\text{ON}} - \alpha N_{\text{OFF}} \rightarrow S_{\text{rel}} \cdot N_{\text{excess}} \quad (7.15)$$

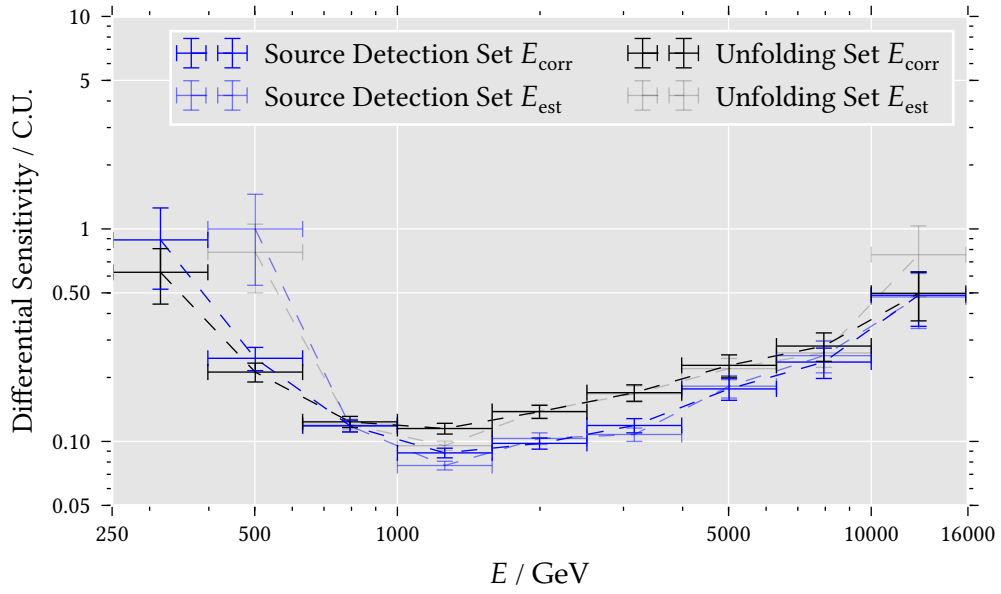
$$S_{50 \text{ h}} = 5\sigma \quad (7.16)$$

$$S_{\text{rel}} = \frac{T_{\text{Obs.}/\text{h}}}{4N_{\text{excess}}} \cdot \left( 1 \pm \sqrt{1 + \frac{8}{T_{\text{Obs.}/\text{h}}} \cdot (1 + \alpha) \cdot \alpha \cdot N_{\text{off}}} \right) \quad (7.17)$$

The relative sensitivity  $S_{\text{rel}}$  describes the minimum fraction of the flux of the Crab Nebula which is detectable by the telescope and the analysis chain with  $5\sigma$  in 50 h. Hence it is given in so called Crab Units (C.U.). For the source detection set and the unfolding the relative integral sensitivities over the whole energy range are:

$$\text{Source Detection: } S_{\text{rel}} = (0.155 \pm 0.005) \text{ C.U.} \quad (7.18)$$

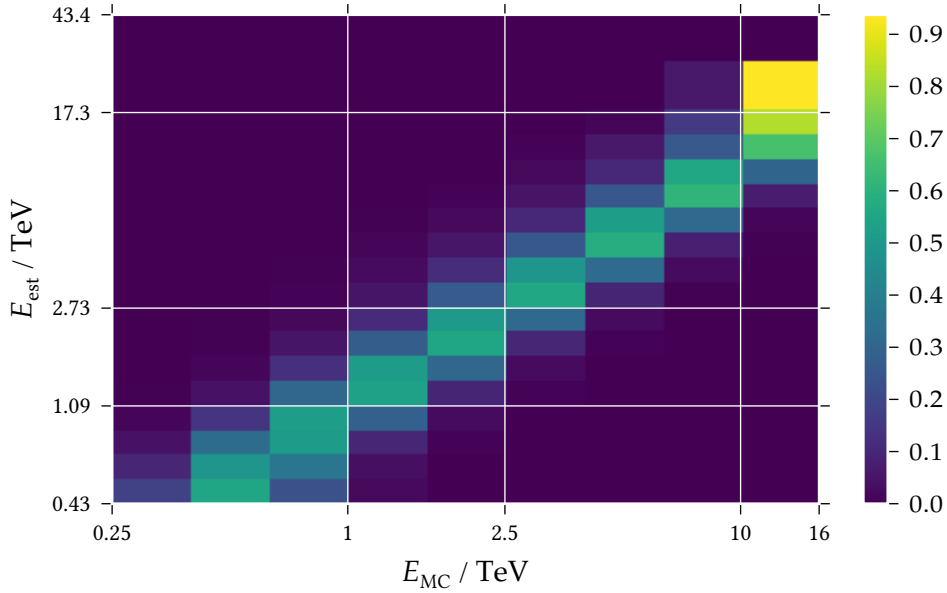
$$\text{Unfolding: } S_{\text{rel}} = (0.188 \pm 0.006) \text{ C.U.} \quad (7.19)$$



**Figure 7.12:** Differential relative sensitivity in Crab Units of the telescope and the analysis chain in dependency of the energy  $E$ . The sensitivity is shown for the source detection set and for the unfolding set. The sensitivity is shown against the estimated energy  $E_{\text{est}}$  and against the corrected energy  $E_{\text{corr}}$  using a migration matrix between  $E_{\text{est}}$  and  $E_{\text{corr}}$ . It is clear that due to the large energy bias of  $E_{\text{est}}$  for small energies the sensitivity is overestimated for small energies using  $E_{\text{est}}$ .

Equation (7.17) can be calculated for each energy bin individual, resulting in the differential relative sensitivity. Figure 7.12 shows the differential relative sensitivity for the source detection set and the unfolding set.

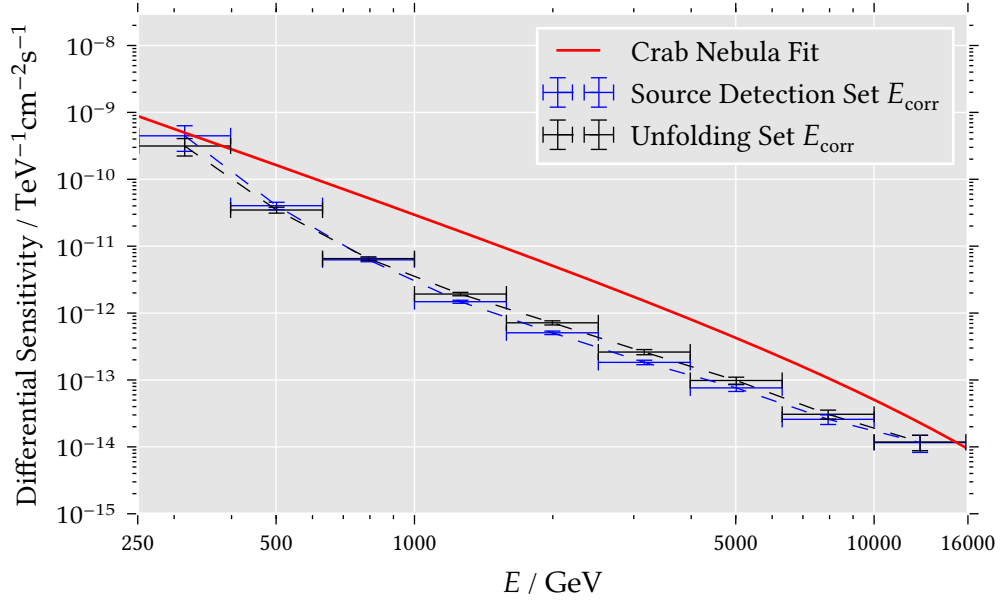
The data is binned in the estimated energy  $E_{\text{est}}$  using the same bins as in the unfolding of the energy spectrum. Due to the large energy bias for small energies,  $E_{\text{est}}$  does not describe the energy reconstruction well, resulting in an overestimated sensitivity for small energies. To correct this, the bin entries of the distribution of  $E_{\text{est}}$  are corrected using the migration matrix  $\mathbf{M}$  evaluated by simulated gamma events.  $\mathbf{M}$  is the correlation matrix between  $E_{\text{est}}$  and  $E_{\text{MC}}$  with each row normalized to one.  $\mathbf{M}$  has  $N$  bins in  $E_{\text{est}}$  and  $M$  bins in  $E_{\text{MC}}$ . In this analysis  $N$  is 20 and the  $M$  bins in  $E_{\text{MC}}$  are equal to the bins in the unfolding of the energy spectrum. Figure 7.13 shows the migration matrix for the unfolding set.



**Figure 7.13:** Migration matrix  $\mathbf{M}$  between  $E_{\text{est}}$  and  $E_{\text{MC}}$  evaluated with the help of simulated gamma events.

To calculate the distribution of the corrected energies  $E_{\text{corr}}$  an event  $i$  with estimated energy  $E_{\text{est},i}$  adds the entries of the row  $k$  of  $\mathbf{M}$  to the distribution of  $E_{\text{corr}}$  where  $E_{\text{est},i}$  is in the  $k$ -th bin of the  $N$  bins in  $E_{\text{est}}$  of  $\mathbf{M}$ . Figure 7.12 shows also the differential relative sensitivity in dependency of  $E_{\text{corr}}$ . The absolute sensitivity can be evaluated by scaling the relative sensitivity with the flux of the Crab Nebula. To calculate the flux of the Crab Nebula, the fitted power law with exponential cut

off function (see section 7.3.3) is integrated over the individual energy bins. The differential absolute sensitivity is shown in figure 7.14. For comparison the fitted power law with exponential cut off function of the Crab Nebula flux is displayed also. Both, the differential relative sensitivity and the differential absolute sensitivity for the telescope and the analysis chain are listed in table 7.4 for the energy bins of the unfolded energy spectrum.



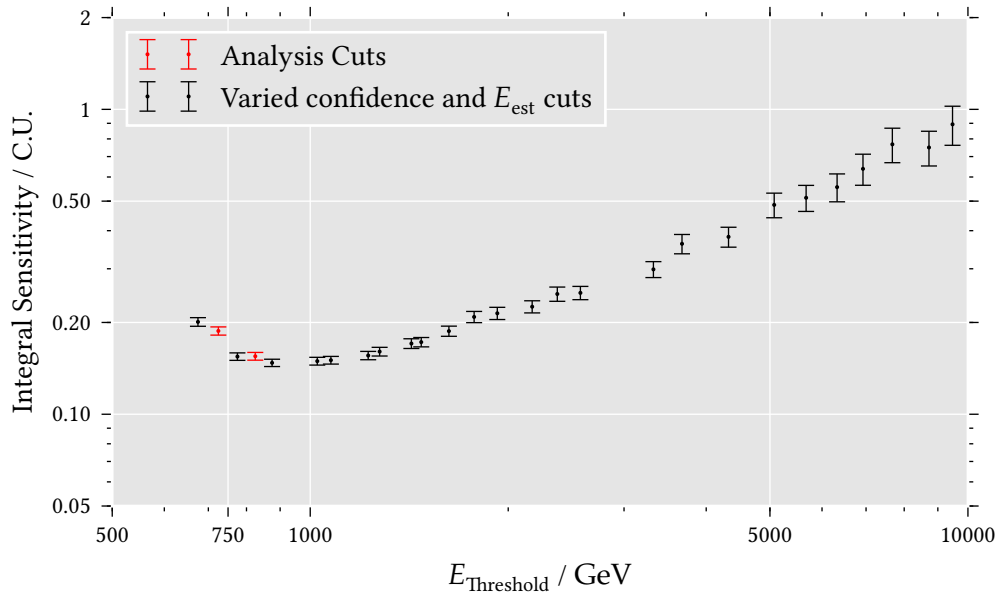
**Figure 7.14:** Differential sensitivity of the telescope and the analysis chain in dependency of the energy  $E$ . The relative sensitivity is scaled with the flux of the Crab Nebula in each energy bin. The sensitivity is shown for the source detection set and for the unfolding set.



**Table 7.4:** Differential sensitivity of the telescope and the analysis chain. For the source detection set and the unfolding set the relative sensitivity  $S_{\text{rel}}$  and the absolute sensitivity  $S_{\text{abs}}$  are displayed for the energy bins of the unfolded energy spectrum.

Source Detection Set			
$E_{\text{low}}$	$E_{\text{high}}$	$S_{\text{rel}} / \text{C.U.}$	$S_{\text{abs}} / \text{TeV}^{-1}\text{cm}^{-2}\text{s}^{-1}$
251	398	$0.9 \pm 0.4$	$(4.5 \pm 1.8) \times 10^{-10}$
398	631	$0.246 \pm 0.031$	$(4.0 \pm 0.5) \times 10^{-11}$
631	1000	$0.118 \pm 0.008$	$(6.3 \pm 0.4) \times 10^{-12}$
1000	1585	$0.088 \pm 0.005$	$(1.48 \pm 0.08) \times 10^{-12}$
1585	2512	$0.098 \pm 0.006$	$(5.07 \pm 0.31) \times 10^{-13}$
2512	3981	$0.119 \pm 0.009$	$(1.83 \pm 0.14) \times 10^{-13}$
3981	6310	$0.177 \pm 0.021$	$(7.6 \pm 0.9) \times 10^{-14}$
6310	10 000	$0.24 \pm 0.04$	$(2.6 \pm 0.4) \times 10^{-14}$
10 000	15 849	$0.49 \pm 0.14$	$(1.16 \pm 0.33) \times 10^{-14}$
Unfolding Set			
$E_{\text{low}}$	$E_{\text{high}}$	$S_{\text{rel}} / \text{C.U.}$	$S_{\text{abs}} / \text{TeV}^{-1}\text{cm}^{-2}\text{s}^{-1}$
251	398	$0.63 \pm 0.18$	$(3.1 \pm 0.9) \times 10^{-10}$
398	631	$0.212 \pm 0.022$	$(3.47 \pm 0.35) \times 10^{-11}$
631	1000	$0.124 \pm 0.008$	$(6.5 \pm 0.4) \times 10^{-12}$
1000	1585	$0.115 \pm 0.007$	$(1.93 \pm 0.11) \times 10^{-12}$
1585	2512	$0.138 \pm 0.010$	$(7.2 \pm 0.5) \times 10^{-13}$
2512	3981	$0.170 \pm 0.015$	$(2.61 \pm 0.23) \times 10^{-13}$
3981	6310	$0.228 \pm 0.027$	$(9.8 \pm 1.2) \times 10^{-14}$
6310	10 000	$0.28 \pm 0.04$	$(3.1 \pm 0.5) \times 10^{-14}$
10 000	15 849	$0.50 \pm 0.13$	$(1.18 \pm 0.31) \times 10^{-14}$

To evaluate the dependency of the integral sensitivity from the energy threshold a scan with different cuts in the CONFIDENCE and in the estimated energy  $E_{\text{est}}$  is performed. For a combination of CONFIDENCE cut and estimated energy cut, the integral sensitivity is calculated. The same cuts are applied to the gamma events set and the maximum of the true energy distribution of the remaining events set is taken as the energy threshold. The results are displayed in figure 7.15 and listed in table 7.5.



**Figure 7.15:** Integral sensitivity of the telescope and the analysis chain in dependency of energy threshold  $E$ . The sensitivity for the source detection set and the unfolding sets are marked in red. The sensitivities for different energy thresholds are evaluated by varying CONFIDENCE cuts and cuts in  $E_{\text{est}}$ .

**Table 7.5:** Integral sensitivity of the telescope and the analysis chain for different energy thresholds. For the unfolding set (CONFIDENCE cut 0.89) and the source detection set (CONFIDENCE cut 0.98) the integral sensitivities are given first. The integral sensitivities for different energy thresholds are evaluated by varying the CONFIDENCE cut and a cut in the estimated energy  $E_{\text{est}}$ . The cuts are listed in the table. The corresponding energy threshold is determined by the maximum of the true energy distribution of a gamma events set with the same applied cuts.

$E_{\text{Threshold}} / \text{GeV}$	Confidence Cut	$E_{\text{est}}$ Cut / GeV	$S_{\text{rel}} / \text{C.U.}$
725	0.89		$0.188 \pm 0.006$
825	0.98		$0.155 \pm 0.005$
675	0.84	501	$0.201 \pm 0.007$
775	0.97	501	$0.155 \pm 0.004$
875	0.97	708	$0.147 \pm 0.004$
1025	0.98	794	$0.149 \pm 0.004$
1075	0.97	891	$0.150 \pm 0.004$
1225	0.98	1000	$0.156 \pm 0.005$
1275	0.98	1120	$0.160 \pm 0.005$
1425	0.99	1120	$0.171 \pm 0.006$
1475	0.98	1260	$0.172 \pm 0.006$
1625	0.98	1410	$0.187 \pm 0.007$
1775	0.98	1580	$0.209 \pm 0.009$
1925	0.99	1580	$0.214 \pm 0.010$
2175	0.98	1780	$0.225 \pm 0.010$
2375	0.99	2000	$0.248 \pm 0.013$
2575	0.97	2240	$0.250 \pm 0.013$
3325	0.97	2820	$0.299 \pm 0.018$
3675	0.98	3160	$0.362 \pm 0.026$
4325	0.97	3550	$0.382 \pm 0.029$
5075	0.97	4470	$0.49 \pm 0.05$
5675	0.97	5010	$0.51 \pm 0.05$
6325	0.97	5620	$0.56 \pm 0.06$
6925	0.98	5620	$0.64 \pm 0.07$
7675	0.94	7080	$0.77 \pm 0.10$
8725	0.97	7080	$0.75 \pm 0.10$
9475	0.97	7940	$0.89 \pm 0.13$



## 8 Summary and Outlook

The analysis of data of the Crab Nebula obtained by the imaging air Cherenkov telescope FACT is presented in this thesis. In 89.7 h observation time the Crab Nebula is detected as a source of very high energy gamma rays with a significance of  $39.89\sigma$ .

Modern machine learning methods are applied to perform several tasks of the analysis. The background suppression is performed by a random forest classifier, the estimation of the energy of the primary particle by a random forest regressor. The tuning of the algorithms and the validation of the model building is based on Monte Carlo simulated events. This ensures that no signal is artificially generated and allows to estimate the performance of the methods on the telescope data directly.

The energy spectrum of the Crab Nebula is reconstructed using a Tikhonov regularized unfolding. Again the tuning and validation of the unfolding is based on Monte Carlo simulated events. An energy spectrum ranging from 250 GeV up to 16 TeV could be reconstructed. It is in good agreement with the measurements of other experiments.

The HEGRA array, consisting of five telescopes, each one comparable in reflector surface to the FACT telescope, achieved an energy spectrum between 316 GeV and 100 TeV with a data set of 384.86 h [5]. The achieved lower edge of the reconstructed energy spectrum in this thesis outranges the one of the nearly five times larger HEGRA telescope array. This supports the usability of the SiPM photo sensor technology which is used in FACT for the first time in the Cherenkov astronomy. The performance for higher energies is ultimately limited by the observation time. Nevertheless, the evaluation of the analysis presented in this thesis shows that for higher energies the performance could be improved for example with an enhanced calculation of the  $\vartheta$  image parameter for higher energies.

Three functions are fitted to the unfolded data points, to describe the energy spectrum of the Crab Nebula. A simple power law function is rejected by a very small fit probability. A power law with exponential cut off is favored by the unfolded data points in this analysis. A log parabola function can describe the data points also. Both functions can be found in literature ([8] and [4]), describing the energy spectrum of the Crab Nebula.

The results of the analysis are used to evaluate the performance of the telescope and the analysis chain in respect to energy bias, energy resolution, effective area and sensitivity.

For small energies the estimation of the energy of the primary particle has an expected bias. This bias can be corrected by applying an unfolding method. With energies starting from 1 TeV the bias is negligible. The energy resolution reaches about 22 % at around 600 GeV. With energies from 6 TeV upwards the resolution further improves to 20 % and then finally declines down to 13 % for the highest energies. These values are only a little bit larger than for the quite larger MAGIC telescopes [8].

The effective area and the sensitivity of both the telescope and the analysis chain is evaluated. The effective area is monotonously increasing, reaching about  $3 \times 10^4 \text{ m}^2$  around 1 TeV. An improvement for higher energies can be expected by enhancing the  $\vartheta$  calculation. The sensitivity is 15.5 % of the flux of the Crab Nebula. This means a flux corresponding to 15.5 % of the flux of the Crab Nebula can be detected with a significance of  $5\sigma$  within an observation time of 50 h. The differential sensitivity yielding the best sensitivity for the energy range of 1 TeV to 1.6 TeV with a sensitivity of 8.8 % of the Crab Nebula flux being detectable with  $5\sigma$  within 50 h.

To summarize the performance of the telescope, its photo sensors and the analysis chain is comparable to other IACTs. Taking into account the small reflector surface this is quite impressive. Several prototypes of the next generation experiment, the Cherenkov Telescope Array, are planned with SiPMs as photo sensors. The results of the analysis, presented in this thesis promise an improved performance of SiPM based imaging air Cherenkov telescopes in comparison to photo multiplier tubes based ones.

Further improvements for the analysis can be done. Several points in the simulations can be investigated, improving the agreement between simulation and real telescope data. Next to a better description of the telescope by the simulation, this enables to use more image parameters, for example timing based parameters in the analysis. An improvement of the performance at higher energies can be achieved by improving the calculation of  $\vartheta$  for higher energy events. Another alternative is to perform dedicated analysis for high and low energy events. Also the work on the preprocessing analysis steps, the extraction, cleaning or parameterization, promises to improve the performance of the analysis. Within the Collaborative Research Center 876 at the TU Dortmund the preprocessing software FACT-Tools is further improved. In addition, uncommon analysis ideas, for example the application of machine learning algorithms for preprocessing analysis tasks, are evaluated.

---

FACT also offers the possibility to perform systematic studies of the properties of the SiPMs. This yields a large impact on the Cherenkov astronomy in respect to the planned SiPM based telescopes.

To summarize the results of the analysis presented in this thesis show the impressive performance of FACT, the first imaging air Cherenkov telescope using SiPMs, in respect to energy range, energy bias, energy resolution, effective area and sensitivity.





## A Data Check

### Data Check Conditions

List of the data check conditions.

```
fSourceKey = 5
AND fRunTypeKey = 1
AND fNight BETWEEN 20130630 AND 20140205
AND fZenithDistanceMax < 30
AND fMoonZenithDistance > 100
AND fThresholdMinSet < 350
AND fCurrentsMedMeanBeg < 8
AND fCurrentsMedMeanBeg < 20
AND fThresholdMinSet < 400
AND fThresholdMinSet < (14 * fCurrentsMedMeanBeg + 265)
AND fTriggerRateMedian < 85
AND fTriggerRateMedian > 40
AND fEffectiveOn > 0.95
AND fNight > 20120420
AND NOT fNight IN (20120406,20120410,20120503)
AND NOT fNight BETWEEN 20121206 AND 20130110
AND -0.085 < (fNumEvtsAfterBgCuts/5-fNumSigEvts)/
fOnTimeAfterCuts - pow(0.753833 * cos(Radians(
fZenithDistanceMean)), 7.647435) * exp(-5.753686*pow(Radians(
fZenithDistanceMean),2.089609)) - pow((if(isnull(
fThresholdMinSet),fThresholdMedian,fThresholdMinSet)-329.4203)
,2) * (-0.0000002044803) AND (fNumEvtsAfterBgCuts/5-
fNumSigEvts)/fOnTimeAfterCuts - pow(0.753833 * cos(Radians(
fZenithDistanceMean)), 7.647435) * exp(-5.753686*pow(Radians(
fZenithDistanceMean),2.089609)) - pow((if(isnull(
fThresholdMinSet),fThresholdMedian,fThresholdMinSet)-329.4203)
,2) * (-0.0000002044803) < 0.25
```

## Data Run List

List of the data runs resulting from the application of the data check conditions.

Number	Run names				
1 - 5	20140203_110	20140202_104	20140202_105	20140202_106	20140202_107
6 - 10	20140202_110	20140202_111	20140202_112	20140202_091	20140202_092
11 - 15	20140202_093	20140202_094	20140201_083	20140201_084	20140201_087
16 - 20	20140201_088	20140201_089	20140201_090	20140201_100	20140201_063
21 - 25	20140201_101	20140201_064	20140201_102	20140201_065	20140201_103
26 - 30	20140201_075	20140201_106	20140201_076	20140201_107	20140201_077
31 - 35	20140201_108	20140201_078	20140201_081	20140201_082	20140130_036
36 - 40	20140130_067	20140130_105	20140130_039	20140130_077	20140130_108
41 - 45	20140130_040	20140130_078	20140130_109	20140130_079	20140130_110
46 - 50	20140130_042	20140130_080	20140130_113	20140130_052	20140130_083
51 - 55	20140130_053	20140130_084	20140130_054	20140130_085	20140130_055
56 - 60	20140130_086	20140130_058	20140130_089	20140130_059	20140130_090
61 - 65	20140130_060	20140130_091	20140130_061	20140130_092	20140130_033
66 - 70	20140130_064	20140130_102	20140130_034	20140130_065	20140130_103
71 - 75	20140130_035	20140130_066	20140130_104	20140128_054	20140128_085
76 - 80	20140128_057	20140128_088	20140128_058	20140128_089	20140128_059
81 - 85	20140128_090	20140128_060	20140128_091	20140128_031	20140128_063
86 - 90	20140128_101	20140128_032	20140128_064	20140128_102	20140128_033
91 - 95	20140128_065	20140128_103	20140128_034	20140128_066	20140128_104
96 - 100	20140128_037	20140128_076	20140128_107	20140128_038	20140128_077
101 - 105	20140128_108	20140128_039	20140128_078	20140128_109	20140128_040
106 - 110	20140128_079	20140128_112	20140128_051	20140128_082	20140128_113
111 - 115	20140128_052	20140128_083	20140128_114	20140128_053	20140128_084

Number	Run names				
116 - 120	20140127_033	20140127_062	20140127_093	20140127_034	20140127_065
121 - 125	20140127_103	20140127_035	20140127_066	20140127_104	20140127_038
126 - 130	20140127_067	20140127_105	20140127_039	20140127_068	20140127_106
131 - 135	20140127_040	20140127_078	20140127_109	20140127_079	20140127_110
136 - 140	20140127_042	20140127_080	20140127_111	20140127_052	20140127_081
141 - 145	20140127_112	20140127_053	20140127_084	20140127_115	20140127_054
146 - 150	20140127_085	20140127_116	20140127_055	20140127_086	20140127_117
151 - 155	20140127_058	20140127_087	20140127_090	20140127_060	20140127_091
156 - 160	20140127_061	20140127_092	20140125_022	20140125_032	20140125_033
161 - 165	20140125_035	20140125_036	20140125_039	20140125_040	20140125_041
166 - 170	20140125_042	20140125_016	20140125_019	20140125_020	20140125_021
171 - 175	20140121_040	20140121_069	20140121_041	20140121_070	20140121_042
176 - 180	20140121_073	20140121_044	20140121_047	20140121_048	20140121_049
181 - 185	20140121_051	20140121_061	20140121_024	20140121_062	20140121_034
186 - 190	20140121_063	20140121_035	20140121_064	20140121_036	20140121_067
191 - 195	20140121_037	20140121_068	20140106_143	20140106_144	20140106_145
196 - 200	20140106_146	20140105_132	20140105_133	20140105_134	20140105_135
201 - 205	20140105_145	20140105_146	20140105_147	20140105_119	20140105_148
206 - 210	20140105_120	20140105_151	20140105_121	20140105_152	20140105_122
211 - 215	20140105_153	20140105_123	20140105_126	20140105_127	20140105_128
216 - 220	20140105_129	20140104_107	20140104_138	20140104_108	20140104_139
221 - 225	20140104_109	20140104_140	20140104_112	20140104_150	20140104_113
226 - 230	20140104_151	20140104_114	20140104_152	20140104_115	20140104_153
231 - 235	20140104_125	20140104_126	20140104_127	20140104_128	20140104_100
236 - 240	20140104_131	20140104_101	20140104_132	20140104_102	20140104_133
241 - 245	20140104_134	20140104_106	20140104_137	20140103_087	20140103_125
246 - 250	20140103_097	20140103_126	20140103_098	20140103_129	20140103_099

Number	Run names				
251 - 255	20140103_130	20140103_100	20140103_131	20140103_103	20140103_132
256 - 260	20140103_104	20140103_135	20140103_073	20140103_136	20140103_074
261 - 265	20140103_106	20140103_137	20140103_075	20140103_109	20140103_138
266 - 270	20140103_078	20140103_110	20140103_148	20140103_079	20140103_111
271 - 275	20140103_149	20140103_080	20140103_112	20140103_150	20140103_084
276 - 280	20140103_122	20140103_151	20140103_085	20140103_123	20140103_152
281 - 285	20140103_086	20140103_155	20140102_107	20140102_138	20140102_073
286 - 290	20140102_110	20140102_148	20140102_074	20140102_111	20140102_149
291 - 295	20140102_079	20140102_112	20140102_150	20140102_080	20140102_113
296 - 300	20140102_151	20140102_081	20140102_123	20140102_154	20140102_085
301 - 305	20140102_124	20140102_155	20140102_086	20140102_125	20140102_156
306 - 310	20140102_087	20140102_126	20140102_088	20140102_129	20140102_130
311 - 315	20140102_099	20140102_131	20140102_101	20140102_132	20140102_104
316 - 320	20140102_135	20140102_105	20140102_136	20140102_106	20140102_137
321 - 325	20140101_085	20140101_116	20140101_154	20140101_087	20140101_126
326 - 330	20140101_157	20140101_088	20140101_127	20140101_089	20140101_128
331 - 335	20140101_090	20140101_129	20140101_101	20140101_132	20140101_102
336 - 340	20140101_133	20140101_103	20140101_134	20140101_073	20140101_104
341 - 345	20140101_135	20140101_074	20140101_107	20140101_138	20140101_075
346 - 350	20140101_108	20140101_139	20140101_076	20140101_109	20140101_140
351 - 355	20140101_079	20140101_110	20140101_141	20140101_080	20140101_113
356 - 360	20140101_151	20140101_081	20140101_114	20140101_082	20140101_115
361 - 365	20140101_153	20131231_113	20131231_144	20131231_116	20131231_147
366 - 370	20131231_085	20131231_117	20131231_148	20131231_086	20131231_118
371 - 375	20131231_087	20131231_119	20131231_090	20131231_122	20131231_091
376 - 380	20131231_123	20131231_092	20131231_124	20131231_093	20131231_125
381 - 385	20131231_096	20131231_135	20131231_097	20131231_136	20131231_098

Number	Run names				
386 - 390	20131231_137	20131231_099	20131231_138	20131231_110	20131231_141
391 - 395	20131231_111	20131231_142	20131231_112	20131231_143	20131204_137
396 - 400	20131204_168	20131204_206	20131204_138	20131204_169	20131204_141
401 - 405	20131204_179	20131204_142	20131204_180	20131204_143	20131204_181
406 - 410	20131204_144	20131204_182	20131204_154	20131204_185	20131204_117
411 - 415	20131204_155	20131204_186	20131204_118	20131204_156	20131204_187
416 - 420	20131204_119	20131204_157	20131204_188	20131204_129	20131204_160
421 - 425	20131204_191	20131204_130	20131204_161	20131204_192	20131204_131
426 - 430	20131204_162	20131204_193	20131204_132	20131204_163	20131204_194
431 - 435	20131204_135	20131204_166	20131204_204	20131204_136	20131204_167
436 - 440	20131204_205	20131201_079	20131201_080	20131201_082	20131201_092
441 - 445	20131201_093	20131201_105	20131201_116	20131130_128	20131130_116
446 - 450	20131130_117	20131130_118	20131130_119	20131130_122	20131130_123
451 - 455	20131130_126	20131130_127	20131128_175	20131128_143	20131128_176
456 - 460	20131128_144	20131128_179	20131128_145	20131128_180	20131128_148
461 - 465	20131128_181	20131128_149	20131128_182	20131128_150	20131128_192
466 - 470	20131128_154	20131128_193	20131128_155	20131128_194	20131128_156
471 - 475	20131128_195	20131128_198	20131128_168	20131128_199	20131128_169
476 - 480	20131128_200	20131128_170	20131128_201	20131128_173	20131128_174
481 - 485	20131126_131	20131126_132	20131126_133	20131126_134	20131126_144
486 - 490	20131126_145	20131126_146	20131126_147	20131126_148	20131126_127
491 - 495	20131112_222	20131112_225	20131112_226	20131112_227	20131112_228
496 - 500	20131112_213	20131112_215	20131112_216	20131112_219	20131112_220
501 - 505	20131111_206	20131111_216	20131111_217	20131111_218	20131111_219
506 - 510	20131111_191	20131111_222	20131111_192	20131111_223	20131111_193
511 - 515	20131111_224	20131111_194	20131111_225	20131111_197	20131111_228
516 - 520	20131111_198	20131111_199	20131111_200	20131111_203	20131111_204

Number	Run names				
521 - 525	20131111_205	20131110_182	20131110_213	20131110_183	20131110_214
526 - 530	20131110_186	20131110_224	20131110_187	20131110_225	20131110_188
531 - 535	20131110_226	20131110_189	20131110_227	20131110_199	20131110_230
536 - 540	20131110_200	20131110_231	20131110_201	20131110_202	20131110_174
541 - 545	20131110_205	20131110_175	20131110_206	20131110_176	20131110_207
546 - 550	20131110_177	20131110_208	20131110_180	20131110_211	20131110_181
551 - 555	20131110_212	20131109_162	20131109_200	20131109_231	20131109_172
556 - 560	20131109_203	20131109_234	20131109_173	20131109_204	20131109_235
561 - 565	20131109_174	20131109_205	20131109_175	20131109_206	20131109_178
566 - 570	20131109_209	20131109_179	20131109_210	20131109_180	20131109_211
571 - 575	20131109_150	20131109_181	20131109_212	20131109_153	20131109_184
576 - 580	20131109_222	20131109_154	20131109_185	20131109_223	20131109_155
581 - 585	20131109_186	20131109_224	20131109_156	20131109_187	20131109_225
586 - 590	20131109_159	20131109_197	20131109_228	20131109_160	20131109_198
591 - 595	20131109_229	20131109_161	20131109_199	20131109_230	20131108_180
596 - 600	20131108_211	20131108_181	20131108_212	20131108_184	20131108_222
601 - 605	20131108_153	20131108_185	20131108_223	20131108_154	20131108_186
606 - 610	20131108_224	20131108_155	20131108_187	20131108_225	20131108_156
611 - 615	20131108_197	20131108_228	20131108_159	20131108_198	20131108_229
616 - 620	20131108_160	20131108_199	20131108_230	20131108_161	20131108_200
621 - 625	20131108_231	20131108_162	20131108_203	20131108_172	20131108_204
626 - 630	20131108_173	20131108_205	20131108_174	20131108_206	20131108_178
631 - 635	20131108_209	20131108_179	20131108_210	20131107_151	20131107_182
636 - 640	20131107_220	20131107_152	20131107_183	20131107_221	20131107_155
641 - 645	20131107_193	20131107_224	20131107_156	20131107_194	20131107_225
646 - 650	20131107_157	20131107_195	20131107_226	20131107_158	20131107_196
651 - 655	20131107_227	20131107_168	20131107_199	20131107_169	20131107_200

Number	Run names				
656 - 660	20131107_170	20131107_201	20131107_171	20131107_202	20131107_143
661 - 665	20131107_174	20131107_205	20131107_144	20131107_175	20131107_206
666 - 670	20131107_145	20131107_176	20131107_207	20131107_146	20131107_177
671 - 675	20131107_208	20131107_149	20131107_180	20131107_218	20131107_150
676 - 680	20131107_181	20131107_219	20131106_175	20131106_206	20131106_176
681 - 685	20131106_207	20131106_179	20131106_210	20131106_180	20131106_211
686 - 690	20131106_181	20131106_212	20131106_182	20131106_213	20131106_154
691 - 695	20131106_185	20131106_223	20131106_155	20131106_186	20131106_224
696 - 700	20131106_156	20131106_187	20131106_225	20131106_157	20131106_188
701 - 705	20131106_226	20131106_160	20131106_198	20131106_229	20131106_161
706 - 710	20131106_199	20131106_235	20131106_162	20131106_200	20131106_237
711 - 715	20131106_163	20131106_201	20131106_173	20131106_204	20131106_174
716 - 720	20131106_205	20131105_164	20131105_206	20131105_237	20131105_174
721 - 725	20131105_207	20131105_238	20131105_175	20131105_208	20131105_239
726 - 730	20131105_176	20131105_209	20131105_240	20131105_177	20131105_212
731 - 735	20131105_180	20131105_213	20131105_181	20131105_214	20131105_182
736 - 740	20131105_215	20131105_183	20131105_225	20131105_186	20131105_226
741 - 745	20131105_187	20131105_227	20131105_188	20131105_228	20131105_189
746 - 750	20131105_231	20131105_155	20131105_199	20131105_232	20131105_156
751 - 755	20131105_201	20131105_233	20131105_162	20131105_205	20131105_234
756 - 760	20131104_187	20131104_218	20131104_188	20131104_219	20131104_220
761 - 765	20131104_190	20131104_221	20131104_162	20131104_193	20131104_231
766 - 770	20131104_163	20131104_194	20131104_232	20131104_164	20131104_195
771 - 775	20131104_233	20131104_165	20131104_196	20131104_243	20131104_168
776 - 780	20131104_206	20131104_244	20131104_169	20131104_207	20131104_170
781 - 785	20131104_208	20131104_171	20131104_209	20131104_181	20131104_212
786 - 790	20131104_182	20131104_213	20131104_183	20131104_214	20131104_184

Number	Run names				
791 - 795	20131104_215	20131103_166	20131103_168	20131103_178	20131103_134
796 - 800	20131103_135	20131103_136	20131103_137	20131103_140	20131103_143
801 - 805	20131103_153	20131102_180	20131102_190	20131102_191	20131102_192
806 - 810	20131101_183	20131101_214	20131101_184	20131101_216	20131101_185
811 - 815	20131101_188	20131101_189	20131101_190	20131101_191	20131101_163
816 - 820	20131101_201	20131101_164	20131101_202	20131101_165	20131101_203
821 - 825	20131101_166	20131101_204	20131101_176	20131101_207	20131101_177
826 - 830	20131101_208	20131101_178	20131101_209	20131101_179	20131101_210
831 - 835	20131101_182	20131101_213	20131031_110	20131031_113	20131031_083
836 - 840	20131031_114	20131031_084	20131031_115	20131031_085	20131031_116
841 - 845	20131031_088	20131031_089	20131031_090	20131031_091	20131031_094
846 - 850	20131031_095	20131031_096	20131031_097	20131031_107	20131031_108
851 - 855	20131031_109	20131029_157	20131029_158	20131029_161	20131029_162
856 - 860	20131014_213	20131014_214	20131014_215	20131014_216	20131014_219
861 - 865	20131014_220	20131014_190	20131014_221	20131014_191	20131014_222
866 - 870	20131014_194	20131014_223	20131014_195	20131014_224	20131014_196
871 - 875	20131014_225	20131014_197	20131014_200	20131014_201	20131014_202
876 - 880	20131014_203	20131013_143	20131012_168	20131012_200	20131012_230
881 - 885	20131012_170	20131012_201	20131012_231	20131012_171	20131012_202
886 - 890	20131012_174	20131012_205	20131012_175	20131012_206	20131012_176
891 - 895	20131012_207	20131012_177	20131012_208	20131012_180	20131012_218
896 - 900	20131012_181	20131012_219	20131012_182	20131012_220	20131012_183
901 - 905	20131012_221	20131012_193	20131012_224	20131012_194	20131012_226
906 - 910	20131012_195	20131012_227	20131012_196	20131012_228	20131012_199
911 - 915	20131012_229	20131011_197	20131011_181	20131011_182	20131011_183
916 - 920	20131011_184	20131011_201	20131011_202	20131011_203	20131011_204
921 - 925	20131011_214	20131011_175	20131011_215	20131011_176	20131011_216



Number	Run names				
926 - 930	20131011_177	20131011_217	20131011_178	20131011_194	20131011_218
931 - 935	20131011_195	20131011_221	20131011_196	20131011_222	20131010_182
936 - 940	20131010_212	20131010_185	20131010_213	20131010_155	20131010_186
941 - 945	20131010_214	20131010_156	20131010_187	20131010_215	20131010_157
946 - 950	20131010_188	20131010_216	20131010_160	20131010_198	20131010_161
951 - 955	20131010_199	20131010_162	20131010_200	20131010_163	20131010_202
956 - 960	20131010_173	20131010_205	20131010_174	20131010_206	20131010_175
961 - 965	20131010_176	20131010_208	20131010_179	20131010_209	20131010_180
966 - 970	20131010_210	20131010_181	20131010_211	20131009_208	20131009_202
971 - 975	20131009_224	20131009_218	20131009_177	20131009_225	20131009_219
976 - 980	20131009_226	20131009_220	20131009_227	20131009_221	20131009_228
981 - 985	20131009_180	20131009_229	20131009_181	20131009_230	20131009_182
986 - 990	20131009_231	20131009_183	20131009_232	20131009_193	20131009_233
991 - 995	20131009_194	20131009_196	20131009_205	20131009_199	20131009_206
996 - 1000	20131009_200	20131009_207	20131009_201	20131008_231	20131008_232
1001 - 1005	20131008_233	20131008_177	20131008_234	20131008_178	20131008_235
1006 - 1010	20131008_181	20131008_236	20131008_200	20131008_183	20131008_201
1011 - 1015	20131008_184	20131008_202	20131008_194	20131008_203	20131008_195
1016 - 1020	20131008_219	20131008_196	20131008_220	20131008_197	20131008_225
1021 - 1025	20131008_206	20131008_226	20131008_207	20131008_227	20131008_208
1026 - 1030	20131008_228	20131008_209	20131007_210	20131007_185	20131007_211
1031 - 1035	20131007_186	20131007_213	20131007_187	20131007_223	20131007_188
1036 - 1040	20131007_224	20131007_225	20131007_226	20131007_229	20131007_230
1041 - 1045	20131007_232	20131007_233	20131007_198	20131007_234	20131007_200
1046 - 1050	20131007_235	20131007_201	20131007_205	20131007_207	20131007_182
1051 - 1055	20131006_232	20131006_233	20131006_234	20131006_235	20131006_236
1056 - 1060	20131006_198	20131006_237	20131006_199	20131006_185	20131006_238

Number	Run names				
1061 - 1065	20131006_200	20131006_187	20131006_239	20131006_201	20131006_188
1066 - 1070	20131006_240	20131006_206	20131006_182	20131006_207	20131006_210
1071 - 1075	20131006_211	20131006_212	20131006_213	20131006_224	20131006_229
1076 - 1080	20131005_185	20131005_234	20131005_195	20131005_235	20131005_196
1081 - 1085	20131005_197	20131005_198	20131005_201	20131005_202	20131005_203
1086 - 1090	20131005_204	20131005_207	20131005_227	20131005_208	20131005_228
1091 - 1095	20131005_220	20131005_209	20131005_229	20131005_221	20131005_210
1096 - 1100	20131005_230	20131005_222	20131005_182	20131005_231	20131005_223
1101 - 1105	20131005_183	20131005_232	20131005_224	20131005_184	20131005_233
1106 - 1110	20131004_205	20131004_215	20131004_176	20131004_206	20131004_216
1111 - 1115	20131004_177	20131004_207	20131004_217	20131004_180	20131004_208
1116 - 1120	20131004_218	20131004_181	20131004_186	20131004_219	20131004_182
1121 - 1125	20131004_187	20131004_220	20131004_183	20131004_188	20131004_189
1126 - 1130	20131004_199	20131004_200	20131004_202	20131004_212	20131004_213
1131 - 1135	20131004_214	20131004_175	20131003_188	20131003_189	20131003_190
1136 - 1140	20131003_200	20131003_201	20131003_202	20131003_203	20131003_229
1141 - 1145	20131003_212	20131003_230	20131003_206	20131003_207	20131003_208
1146 - 1150	20131003_209	20131003_216	20131003_217	20131003_218	20131003_219

## B Parameter Lists

### FACT-Tools Parameter List

List of parameters calculated by FACT-Tools The parameters used in the MRMR algorithm are marked as well as the parameters selected by the MRMR algorithm.

Parameter	Used in MRMR	Selected by MRMR
numPixelInShower	x	x
Size	x	x
Length	x	x
Width	x	x
numIslands	x	x
Delta		
COGx		
COGy		
m3l	x	
m3t	x	x
M3Long	x	
M3Trans	x	x
M4Long	x	x
M4Trans	x	x
Disp	x	
Concentration_onePixel	x	x
Concentration_twoPixel	x	x
ConcCore	x	x
concCOG	x	x
Leakage	x	x
Leakage2	x	x
Slope_long		
Slope_trans		
Slope_spread		
Slope_spread_weighted		
Timespread		

## B Parameter Lists

---

Parameter	Used in MRMR	Selected by MRMR
Timespread_weighted		
timeGradientSlope		
timeGradientSlope_err		
timeGradientIntercept		
timeGradientIntercept_err		
timeGradientSSE		
sourcePosition		
Alpha	x	
Distance	x	x
CosDeltaAlpha	x	
Theta		
Theta_recPos		
NIGHT		
RUNID		
EventNum		
UnixTimeUTC		
AzTracking		
ZdTracking		
AzPointing		
ZdPointing		
AzSourceCalc		
ZdSourceCalc		
photonchargeMean		
arrivalTimeMean		
phChargeShower_mean		
phChargeShower_max		
phChargeShower_min		
phChargeShower_kurtosis		
phChargeShower_variance		
phChargeShower_skewness		
arrTimeShower_mean		
arrTimeShower_max		
arrTimeShower_min		
arrTimeShower_kurtosis		
arrTimeShower_variance		
arrTimeShower_skewness		
maxSlopesShower_mean		
maxSlopesShower_max		
maxSlopesShower_min		

---

Parameter	Used in MRMR	Selected by MRMR
maxSlopesShower_kurtosis		
maxSlopesShower_variance		
maxSlopesShower_skewness		
arrTimePosShower_mean		
arrTimePosShower_max		
arrTimePosShower_min		
arrTimePosShower_kurtosis		
arrTimePosShower_variance		
arrTimePosShower_skewness		
maxSlopesPosShower_mean		
maxSlopesPosShower_max		
maxSlopesPosShower_min		
maxSlopesPosShower_kurtosis		
maxSlopesPosShower_variance		
maxSlopesPosShower_skewness		
maxPosShower_mean		
maxPosShower_max		
maxPosShower_min		
maxPosShower_kurtosis		
maxPosShower_variance		
maxPosShower_skewness		

---

## Parameter Generation List

List of parameters generated after the preprocessing analysis with FACT-Tools. The generation function is listed, as well as if the parameters is used in the MRMR algorithm and if the parameter is selected by the MRMR algorithm.

Parameter	Generation function	Used in MRMR	Selected by MRMR
MaxPhotonCharge	Concentration_onePixel*Size	x	
logSize	log(Size)	x	x
logLength	log(Length)	x	x
logMaxPhotonCharge	log(Concentration_onePixel*Size)	x	
logabs(m3trans)	log(abs(M3Trans))	x	x
logabs(m3long)	log(abs(M3Long))	x	x
logConc1	log(Concentration_onePixel)	x	x
Area	pi*Width*Length	x	x
logWidth	log(Width)	x	x
logM4Long	log(M4Long)	x	x
logM4Trans	log(M4Trans)	x	x
logabs(m3t)	log(abs(m3t))	x	x
logabs(m3l)	log(abs(m3l))	x	x
ThetaDeg	Theta/9.5*0.11	x	x
SizeArea	Size/Area	x	x
Conc1area	Concentration_onePixel*Area	x	x
SizeConc1	Size/Concentration_onePixel	x	x
LengthNumberShowerPixel	Length/numPixelInShower	x	x
Conc1NumberShowerPixel	Concentration_onePixel/numPixelInShower	x	x
logsizewidthlength	log(Size)/(Width*Length)	x	x
sign(M3long)	M3Long*sgn(CosDeltaAlpha)	x	x
logConc1area	log(Concentration_onePixel*Area)	x	x

Parameter	Generation function	Used in MRMR	Selected by MRMR
logConc1NumberShowerPixel	$\log(\text{Concentration\_onePixel}/\text{numPixelInShower})$	x	x
logLengthNumberShowerPixel	$\log(\text{Length}/\text{numPixelInShower})$	x	
logLengthWidth	$\log(\text{Length}/\text{Width})$	x	x
logSizeArea	$\log(\text{Size}/\text{Area})$	x	x
logArea	$\log(\text{Area})$	x	x
AreaSizeCutVar	$\text{Area}/(\log(\text{Size}))^{**2}$	x	x
sign(m3l)	$m3l * \text{sgn}(\text{CosDeltaAlpha})$	x	x
ThetaDeg2	$\text{ThetaDeg} * \text{ThetaDeg}$	x	x
E_RF	Generated by random forest regressor	x	x
var_E_RF	Generated by random forest regressor	x	x





## C Simulation configuration files

### CORSIKA Inputcard for protons

Inputcard for CORSIKA for producing proton events. The entries in curved brackets are filled by an automation production script. Theta is varied between 0° and 30° in 1° steps.

```
RUNNR      {run_number}
EVTNR      1
NSHOW      1500
SEED       {seed1} 0 0
SEED       {seed2} 0 0
SEED       {seed3} 0 0
PRMPAR     14
ERANGE     100.0 200000.0
ESLOPE     -2.700
THETAP     {theta} {theta}
PHIP       0.000 0.000
VIEWCONE   0.000 5.000
FIXCHI     0
OBSLEV     220000.0
MAGNET     30.3 24.1
ARRANG     -7.0
ATMOSPHERE 11 T
ATMLAY     775000.0 1650000.0 5000000.0 10500000.0
RADNKG     20000.0
ECUTS      0.3 0.3 0.02 0.02
ECTMAP     10000.0
MUADDI     F
MUMULT     T
CWAVLG     290.0 900.0
CSCAT      20 0.0 40000.0
CERSIZ     1.0
CERFIL     T
CERTEL     1
           0.0 0.0 0.0 0.0 0.0 500.0 500.0
LONGI      F 20.0 F F
MAXPRT     0
```

```
PAROUT   F F
DATBAS   F
EXIT
```

## Corsika Inputcard for gamma

Inputcard for CORSIKA for producing gamma events. The entries in curved brackets are filled by an automation production script. Theta is varied between 0° and 30° in 1° steps.

```
RUNNR     {run_number}
EVTNR     1
NSHOW     3000
SEED      {seed1} 0 0
SEED      {seed2} 0 0
SEED      {seed3} 0 0
PRMPAR    1
ERANGE    200.0 50000.0
ESLOPE    -2.700
THETAP    {theta} {theta+1}
PHIP      0.000 0.000
VIEWCONE  0.000 0.000
FIXCHI    0
OBSLEV    220000.0
MAGNET    30.3 24.1
ARRANG    -7.0
ATMOSPHERE 11 T
ATMLAY    775000.0 1650000.0 5000000.0 10500000.0
RADNKG    20000.0
ECUTS     0.3 0.3 0.02 0.02
ECTMAP    10000.0
MUADDI    F
MUMULT    T
CWAVLG    290.0 900.0
CSCAT     1 0.0 27000.0
CERSIZ    1.0
CERFIL    T
CERTEL    1
          0.0 0.0 0.0 0.0 0.0 500.0 500.0
LONGI     F 20.0 F F
MAXPRT    0
PAROUT    F F
DATBAS    F
EXIT
```

---

## Ceres rc-file

Configuration file for the reflector and camera simulation program Ceres.

MLog.VerbosityLevel: 4

BinningImpact.Raw: 40 0 1000  
BinningTrigPos.Raw: 300 -25 275

RandomNumberGenerator: TRandom3

MSimPointingPos.OffTargetDistance: 0.6

Reflector.Constructor: MReflector  
Reflector.FileName: fact-reflector\_first\_light\_untill\_May2014.txt  
Reflector.SetSigmaPSF: 2.0

MGeomCam.Constructor: MGeomCamFACT();

MSimAPD.Type: 0  
MSimAPD.NumCells: 60  
MSimAPD.DeadTime: 3.0  
MSimAPD.RecoveryTime: 8.75  
MSimAPD.CrosstalkCoefficient: 0.1  
MSimAPD.AfterpulseProb1: 0.14  
MSimAPD.AfterpulseProb2: 0.11  
MSimExcessNoise.ExcessNoise: 0.096

MirrorReflectivity.FileName:  
MirrorReflectivity\_Lustermann\_FACT\_bearbeitet.txt  
PhotonDetectionEfficiency.FileName: fact-pde-1.4V.txt  
ConesAngularAcceptance.FileName: fact-cones-angular-  
acceptance.txt  
ConesTransmission.FileName:  
Transmittance\_1439Cones\_FACT\_bearbeitet.txt

AdditionalPhotonAcceptance.Function.Name: 0.85  
AdditionalPhotonAcceptance.Function.Npx: 100  
AdditionalPhotonAcceptance.Function.Xmin: 290  
AdditionalPhotonAcceptance.Function.Xmax: 900

MSimRandomPhotons.FrequencyFixed: 0.004  
MSimRandomPhotons.FileNameNSB: resmc/night-sky-la-palma.txt  
MSimRandomPhotons.FrequencyNSB: 0.0

MSimTrigger.FileNameRouteAC: fact-trigger-sum.txt

## *C Simulation configuration files*

---

```
MSimTrigger.DiscriminatorThreshold: -192.387
MSimTrigger.CableDelay: 21.0
MSimTrigger.CableDamping: -0.96
MSimTrigger.CoincidenceTime: 0.5

MSimCalibrationSignal.NumPhotons: 24
MSimCalibrationSignal.NumEvents: 1000

IntendedPulsePos.Val: 26

MRawRunHeader.SamplingFrequency: 2000
MRawRunHeader.NumSamples: 300
MRawRunHeader.NumBytesPerSample: 2
MRawRunHeader.FadcResolution: 12

MSimCamera.DefaultOffset: -1850.0
MSimCamera.DefaultNoise: 2.8125
MSimCamera.DefaultGain: 22.553
MSimCamera.ACFudgeFactor: 0.3136
MSimCamera.ACTimeConstant: 20

PulseShape.Function.Name: (1.239*(1-1/(1+exp((0.5*x-2.851)
/1.063))))*exp(-(0.5*x-2.851)/19.173))
PulseShape.Function.Npx: 310
PulseShape.Function.Xmin: -10
PulseShape.Function.Xmax: 300

Cut.Inverted: yes
Cut.Condition: MHillas.fSize>10.0
ContEmpty3.Condition: MPhotonEvent.GetNumPhotons<10

MFixTimeOffset.FileName: resmc/fact/pixel_delays_ALL_ZERO.csv

ResidualTimeSpread.Val: 0.0
GapdTimeJitter.Val: 1.5
```

## List of Figures

2.1	Overview over the different messenger particles in the field of astroparticle physics. The galactic and extragalactic sources, the propagation through space and the different detection principles are shown [12]. . . . .	3
2.2	Energy spectrum of charged cosmic rays between 0.1 GeV to $10^6$ GeV [47]. . . . .	5
2.3	Energy spectrum of charged cosmic rays between $10^{13}$ eV to $10^{21}$ eV. The flux is scaled with $E^{2.6}$ to emphasize the different structures in the spectrum [47]. . . . .	6
2.4	Simplified image of the general structure of an air shower caused by charged cosmic rays [57]. . . . .	8
2.5	Simulated air shower caused by a 1 TeV proton entering the atmosphere. The simulation was performed with CORSIKA [54]. . . . .	9
2.6	Simplified image of the general structure of an air shower caused by very high energy gamma ray [57]. . . . .	11
2.7	Simulated air shower caused by a 1 TeV gamma entering the atmosphere. The simulation was performed with CORSIKA [54]. . . . .	12
2.8	Image of the Crab Nebula. The different observations are overlayed. X-Ray observation in blue, optical in green and radio in red [35]. . . . .	13
3.1	Image of the First G-APD Cherenkov Telescope in park position during the day. . . . .	16
3.2	Image of the camera of FACT during assembly- and test phase. On the left the sensor compartment and the front window can be seen, on the right the electronic compartment with the integrated camera electronics [11]. . . . .	17
3.3	Image of a SiPM and a light concentrator glued to the SiPM. Picture taken by Thomas Krähenbühl. . . . .	18
3.4	Pulse shape caused by a single photon in a SiPM. The fast rising edge when the G-APD is breaking down and the long tail when the G-APD is recovering are visible. The function displayed was evaluated by stacking the signal of single photons in dark count measurements [14]. . . . .	19

3.5	Raw data of one event in one pixel. The image is taken from the graphical user interface implemented in FACT-Tools. The voltage curve over the 300 slices (corresponds to 150 ns) sampled by the DRS4 board is shown. The pulse visible is caused by the detection of about 9 to 10 Cherenkov photons. . . . .	21
3.6	Illustration of the wobble strategy while observing a source of very high energy gamma rays with an IACT. The ON and 3 OFF regions for signal and background data sets are shown. In addition an image of an air shower and the calculation of $\vartheta$ for ON and OFF regions is sketched. . . . .	22
5.1	Illustration of the image parameters calculated by FACT-Tools. . . .	33
5.2	Illustration of the cross validation method. The labeled event set is separated into $m$ disjunct subsets. The machine learning model is build on the union of $m - 1$ subsets and tested on the remaining one. The test set is iterated over all $m$ subsets, thus the performance of the model and the stability of the model training can be evaluated. In this illustration is $m = 5$ . . . . .	35
5.3	Correlation between the SIZE and the true energy $E_{MC}$ . SIZE is clearly strongly correlated to the energy. Hence it is often included in an energy estimation algorithm. Shown are background suppressed gamma events. . . . .	36
5.4	Illustration of the generation of a single decision tree. The tree is built using a bootstrapped training event set. At each node a random subset of $j$ parameters is drawn and the separation criterion, which minimize the variance of the target variable in the subsets, is applied. . . . .	38
5.5	Illustration of the principle structure of a random forest regressor. The target variable (in this example the energy $E$ ) of an unlabeled event is estimated by evaluating the estimation of $n$ different decision trees and averaging over the results. . . . .	39
6.1	Rate distributions of the image parameters SIZE, NUMBER OF ISLANDS and CONCENTRATION CORE for proton and data events. The distributions are scaled by the observation time for the data and by the simulated observation time for the proton events. . . . .	52
6.2	Rate distributions of the image parameters LENGTH, WIDTH and the STANDARD DEVIATION OF THE ARRIVAL TIME for proton and data events. The distributions are scaled by the observation time for the data and by the simulated observation time for the proton events. . . . .	53

6.3	Rate distributions of the image parameters LEAKAGE and LEAKAGE2 for proton and data events. The distributions are scaled by the observation time for the data and by the simulated observation time for the proton events. . . . .	54
6.4	Normed distributions of the parameters NUMBER OF PIXELS IN SHOWER, LEAKAGE and LEAKAGE2 of proton and gamma events of the image parameters on which the quality cuts are applied. The cuts are marked with gray dashed lines. . . . .	57
6.5	Normed distributions of the parameters NUMBER OF ISLANDS, WIDTH and LENGTH of proton and gamma events of the image parameters on which the quality cuts are applied. The cuts are marked with gray dashed lines. . . . .	58
6.6	Correlation between simulated energy and estimated energy for gamma (top) and proton (bottom) events. Shown are all events which survived the cleaning and the quality cuts. . . . .	61
6.7	Correlation between simulated energy and estimated energy for gamma events. Shown are all events after the CONFIDENCE cut for the unfolding (see section 6.6). . . . .	62
6.8	Distribution of the variance of the estimated energy for gamma and proton events. It is clearly visible that the variance for proton events is larger than for gamma events, thus the variance is suitable for the background suppression. . . . .	63
6.9	Recall and precision for random forest models trained with $k$ features selected by the MRMR algorithm. The chosen $k$ for the following background suppression random forest model is marked in red. . . .	65
6.10	Recall and false positive rate in dependency of the applied CONFIDENCE cut. The relevant CONFIDENCE cut range is shown in more detail below. . . . .	67
6.11	Receiver operating characteristics curve for the random forest used in the background suppression. The area under the roc curve is 0.9927. 68	
6.12	$F_\beta$ -score in dependency of the applied CONFIDENCE cut. Shown are $\beta = \frac{1}{6}$ and $\beta = \frac{1}{20}$ . The relevant range of CONFIDENCE cuts is shown in more detail. The CONFIDENCE cuts with maximum of $F_\beta$ are marked in red. . . . .	69
6.13	CONFIDENCE distribution for ON and OFF data of the Crab Nebula. A clear excess of the ON data set over the OFF data set for the two marked CONFIDENCE cuts is visible. . . . .	70
6.14	$\vartheta^2$ distribution for ON and OFF data of the Crab Nebula after applying a CONFIDENCE cut of 0.89 (used in Unfolding). . . . .	71

6.15	$\vartheta^2$ distribution for ON and OFF data of the Crab Nebula after applying a CONFIDENCE cut of 0.98 (used in source detection). . . .	72
6.16	Distribution of the MC energy of 6436 gamma events. The number of events is the same as the number of excess events. The mean and the uncertainty is evaluated by bootstrapping. It is clearly visible that each energy bin is populated with a sufficient number of events to unfold the energy spectrum. . . . .	73
6.17	Correlation between the estimated energy $E_{\text{est}}$ and the true energy $E_{\text{MC}}$ . There is a clear linear dependency making $E_{\text{est}}$ suitable for the unfolding. . . . .	74
6.18	Correlation between the CONCENTRATION CORE and the true energy $E_{\text{MC}}$ . There is a dependency visible, thus CONCENTRATION CORE adds additional information about the true energy to the unfolding. . . . .	75
6.19	Correlation between the zenith angle and the true energy $E_{\text{MC}}$ . A direct dependency is not visible, but it is expected that the correlation between the true energy and the observables measured by the telescope depends on the zenith angle. Though taking the zenith angle into the unfolding improves the reconstruction of the energy spectrum. . . . .	76
6.20	One example test unfolding with the unfolded result of a pseudo data event set and the Monte Carlo energy distribution of this event set. . . . .	77
6.21	So called L-Curve calculated by TRUEEE. The different settings (ndf: number of degrees of freedom, nK: number of knots) are used in test mode with a pseudo data set. The $X^2$ between the true energy distribution and the unfolded result and the data point correlation between the unfolded bins are calculated. Settings with minimal values in both parameters should be chosen. The setting used in this analysis is marked red. . . . .	78
6.22	The results of the pull mode of TRUEEE. The deviation of each individual pull result to the true distribution (in units of standard deviations $\sigma$ ) is calculated and averaged over all pulls (top). With a perfect unfolding the average deviation (called pull mean) should be zero. The standard deviation (in units of the uncertainty, called pull RMS) of the unfolded results over all pulls is calculated (bottom). For a stable unfolding the the pull RMS should be one. . . . .	80
6.23	The unfolded energy spectrum of the Crab Nebula, measured by FACT between October 2013 and February 2014. For comparison the fits of the Crab Nebula from measurements of HEGRA [5], MAGIC [7], Veritas [9] and H.E.S.S [4] are shown in addition. . . . .	81
6.24	Distribution of the parameters SIZE, WIDTH and LENGTH for reweighted gamma events and data events. A proper unfolding is indicated by the agreement of the distributions. . . . .	82



7.1	Distributions of the energy and the impact of the primary particle of the simulated gamma events after the different analysis steps. . . . .	86
7.2	Distributions of the energy and the impact of the primary particle of the simulated proton events after the different analysis steps. . . . .	87
7.3	Effective collection area $A_{\text{eff}}$ in dependency of the true energy $E_{\text{MC}}$ for the different analysis steps. . . . .	89
7.4	Distribution of the relative difference $\frac{E_{\text{est}} - E_{\text{MC}}}{E_{\text{MC}}}$ for events with $E_{\text{MC}}$ between 1.05 TeV and 1.39 TeV. A gaussian function is fitted to the distribution. The mean and the standard deviation of the function are interpreted as the energy bias and energy resolution in this energy bin. As the function is not sensitive to the non-gaussian tails of the distribution also the median and the $1\sigma$ quantiles are calculated. Shown are events after the CONFIDENCE cut for unfolding. . . . .	91
7.5	Energy bias and energy resolution for gamma events which survived the cleaning and the quality cuts. The values are evaluated by the mean and standard deviation of a gaussian function fitted to the relative difference between $E_{\text{est}}$ and $E_{\text{MC}}$ in each energy bin in $E_{\text{MC}}$ . The band for the energy resolution is evaluated by calculating the difference between the median and the $1\sigma$ quantiles of the distributions of the relative difference. . . . .	92
7.6	Energy bias and energy resolution for gamma events after applying the CONFIDENCE cut for unfolding. The values are evaluated by the mean and standard deviation of a gaussian function fitted to the relative difference between $E_{\text{est}}$ and $E_{\text{MC}}$ in each energy bin in $E_{\text{MC}}$ . The band for the energy resolution is evaluated by calculating the difference between the median and the $1\sigma$ quantiles of the distributions of the relative difference. . . . .	93
7.7	Distribution of the true energy $E_{\text{MC}}$ for the different analysis steps. The energy threshold is defined as the maximum of this distribution. It is determined by fitting a gaussian function to the distribution and taking the mean of the function as the energy threshold. . . . .	94
7.8	The unfolded energy spectrum of the Crab Nebula measured by FACT. Several power law-like functions are fitted to the data points. The power law with exponential cut off and the log-parabola function describe the data well. . . . .	96
7.9	The distribution of $\Delta X^2$ for each parameter of the power law function. $\Delta X^2 = 1$ defines the $1\sigma$ confidence region for the parameters. . . . .	98
7.10	The distribution of $\Delta X^2$ for each parameter of the power law with an exponential cut off function. $\Delta X^2 = 1$ defines the $1\sigma$ confidence region for the parameters. . . . .	99

7.11	The distribution of $\Delta X^2$ for each parameter of the log-parabola function. $\Delta X^2 = 1$ defines the $1\sigma$ confidence region for the parameter.	100
7.12	Differential relative sensitivity in Crab Units of the telescope and the analysis chain in dependency of the energy $E$ . The sensitivity is shown for the source detection set and for the unfolding set. The sensitivity is shown against the estimated energy $E_{\text{est}}$ and against the corrected energy $E_{\text{corr}}$ using a migration matrix between $E_{\text{est}}$ and $E_{\text{corr}}$ . It is clear that due to the large energy bias of $E_{\text{est}}$ for small energies the sensitivity is overestimated for small energies using $E_{\text{est}}$ .	102
7.13	Migration matrix $\mathbf{M}$ between $E_{\text{est}}$ and $E_{\text{MC}}$ evaluated with the help of simulated gamma events. . . . .	103
7.14	Differential sensitivity of the telescope and the analysis chain in dependency of the energy $E$ . The relative sensitivity is scaled with the flux of the Crab Nebula in each energy bin. The sensitivity is shown for the source detection set and for the unfolding set. . . . .	104
7.15	Integral sensitivity of the telescope and the analysis chain in dependency of energy threshold $E$ . The sensitivity for the source detection set and the unfolding sets are marked in red. The sensitivities for different energy thresholds are evaluated by varying CONFIDENCE cuts and cuts in $E_{\text{est}}$ . . . . .	106

## List of Tables

6.1	Number of files and observation time after the application of the different types of data check condition. . . . .	49
6.2	Simulation parameters for CORSIKA for the proton and gamma simulations. . . . .	50
6.3	Split of the gamma event set into two independent event sets. The number of measured events in each set and the corresponding number of simulated events is listed. . . . .	59
6.4	Settings for the random forest regression used for an estimation of the energy of the primary particle . . . . .	60
6.5	Settings for the random forest classifier used for the background suppression. . . . .	66
6.6	Performance values for the two different CONFIDENCE cuts, chosen for the two event sets. . . . .	69
6.7	Number of events and significance of the detection of the signal of the Crab Nebula for the source detection event set and the unfolding event set. . . . .	72
6.8	Overview over the unfolded energy spectrum of the Crab Nebula measured by FACT. . . . .	83
7.1	Number of events after the different steps of the analysis chain for the data sample, for the simulated proton sample and for the simulated gamma sample . . . . .	85
7.2	Energy thresholds for the different analysis steps. Also the standard deviation $\sigma$ of the fitted gaussian function is listed as an estimator for the width of the energy distribution. . . . .	95
7.3	Fit parameters of the fitted functions to the unfolded differential energy spectrum. The best fit parameters and the uncertainties are listed. . . . .	97
7.4	Differential sensitivity of the telescope and the analysis chain. For the source detection set and the unfolding set the relative sensitivity $S_{\text{rel}}$ and the absolute sensitivity $S_{\text{abs}}$ are displayed for the energy bins of the unfolded energy spectrum. . . . .	105

7.5 Integral sensitivity of the telescope and the analysis chain for different energy thresholds. For the unfolding set (CONFIDENCE cut 0.89) and the source detection set (CONFIDENCE cut 0.98) the integral sensitivities are given first. The integral sensitivities for different energy thresholds are evaluated by varying the CONFIDENCE cut and a cut in the estimated energy  $E_{\text{est}}$ . The cuts are listed in the table. The corresponding energy threshold is determined by the maximum of the true energy distribution of a gamma events set with the same applied cuts. . . . . 107

## Bibliography

- [1] A. Aab et al. “Searches for Anisotropies in the Arrival Directions of the Highest Energy Cosmic Rays Detected by the Pierre Auger Observatory”. In: *The Astrophysical Journal* 804.1 (2015), p. 15. URL: <http://stacks.iop.org/0004-637X/804/i=1/a=15>.
- [2] B. P. Abbott et al. “Observation of Gravitational Waves from a Binary Black Hole Merger”. In: *Phys. Rev. Lett.* 116 (6 Feb. 2016), p. 061102. DOI: 10.1103/PhysRevLett.116.061102. URL: <http://link.aps.org/doi/10.1103/PhysRevLett.116.061102>.
- [3] F. Acero et al. “Fermi Large Area Telescope Third Source Catalog”. In: *The Astrophysical Journal Supplement Series* 218, 23 (June 2015), p. 23. DOI: 10.1088/0067-0049/218/2/23. arXiv: 1501.02003 [astro-ph.HE].
- [4] F. Aharonian et al. “Observations of the Crab nebula with HESS”. In: *Astronomy and Astrophysics* 457.3 (2006).
- [5] F. Aharonian et al. “The Crab Nebula and Pulsar between 500 GeV and 80 TeV: Observations with the HEGRA Stereoscopic Air Cerenkov Telescopes”. In: *ApJ* 614.897 (2004).
- [6] M.L. Ahnen et al. “Data compression for the first G-APD Cherenkov Telescope”. In: *Astronomy and Computing* 12 (2015), pp. 191–199. ISSN: 2213-1337. DOI: <http://dx.doi.org/10.1016/j.ascom.2015.06.007>. URL: <http://www.sciencedirect.com/science/article/pii/S2213133715000670>.
- [7] J. Aleksić et al. “Measurement of the Crab Nebula spectrum over three decades in energy with the {MAGIC} telescopes”. In: *Journal of High Energy Astrophysics* 5–6 (2015), pp. 30–38. ISSN: 2214-4048. DOI: <http://dx.doi.org/10.1016/j.jheap.2015.01.002>. URL: <http://www.sciencedirect.com/science/article/pii/S2214404815000038>.
- [8] J. Aleksić et al. “The major upgrade of the {MAGIC} telescopes, Part II: A performance study using observations of the Crab Nebula”. In: *Astroparticle Physics* 72 (2016), pp. 76–94. ISSN: 0927-6505. DOI: <http://dx.doi.org/10.1016/j.astropartphys.2015.02.005>. URL: <http://www.sciencedirect.com/science/article/pii/S0927650515000316>.

- [9] E. Aliu et al. “A search for enhanced very high energy gamma-ray emission from the 2013 March Crab Nebula flare”. In: *ApJ* 781.L11 (2014).
- [10] H. Anderhub et al. “Design and operation of FACT – the first G-APD Cherenkov telescope”. In: *JINST* 8.06 (2013), P06008.
- [11] H. Anderhub et al. “Electronics for the camera of the First G-APD Cherenkov Telescope (FACT) for ground based gamma-ray astronomy”. In: *Journal of Instrumentation* 7.01 (2012), p. C01073. URL: <http://stacks.iop.org/1748-0221/7/i=01/a=C01073>.
- [12] Michael Backes. “Long-term observations of the TeV blazar 1ES 1959+650, Temporal and spectral behavior in the multi-wavelength context”. PhD thesis. TU Dortmund, Dec. 2011.
- [13] Adrian Biland et al. “FACT - Status and Experience from Three Years Operation of the First SiPM Camera”. In: *Proceedings of the 34th ICRC, 865, The Hague, 2015*. 1032. 2015.
- [14] A. Biland et al. “Calibration and performance of the photon sensor response of FACT — the first G-APD Cherenkov telescope”. In: *JINST* 9.10 (2014), P10012. URL: <http://stacks.iop.org/1748-0221/9/i=10/a=P10012>.
- [15] Christian Bockermann. “Mining Big Data Streams for Multiple Concepts”. PhD thesis. TU Dortmund, Nov. 2015.
- [16] Christian Bockermann et al. “Online Analysis of High-Volume Data Streams in Astroparticle Physics”. In: *Machine Learning: ECML 2015, Industrial Track*. Springer Berlin Heidelberg, 2015.
- [17] T.T. Böhlen et al. “The FLUKA Code: Developments and Challenges for High Energy and Medical Applications”. In: *Nuclear Data Sheets* 120 (2014), pp. 211–214. ISSN: 0090-3752. DOI: <http://dx.doi.org/10.1016/j.nds.2014.07.049>. URL: <http://www.sciencedirect.com/science/article/pii/S0090375214005018>.
- [18] Leo Breiman. “Random Forests”. English. In: *Machine Learning* 45.1 (2001), pp. 5–32. ISSN: 0885-6125. DOI: 10.1023/A:1010933404324. URL: <http://dx.doi.org/10.1023/A:1010933404324>.
- [19] T. Bretz et al. “FACT - The First G-APD Cherenkov Telescope: Status and Results”. In: *Proceedings of the 33rd International Cosmic Ray Conference* 682 (Aug. 2013). arXiv: 1308.1512 [astro-ph.IM].
- [20] T. Bretz, D. Dorner, et al. “MARS - CheObs ed.: A flexible software framework for future Cherenkov telescopes”. In: *Proceedings of the 11th ICATPP Conference*. 2009.

- 
- [21] Kai Brügge et al. “FACT-Tools: Streamed Real-Time Data Analysis”. In: *Proceedings of the 34th ICRC, 865, The Hague, 2015*. 865. 2015.
- [22] R. Bühler and R. Blandford. “The surprising Crab pulsar and its nebula: a review”. In: *Reports on Progress in Physics* 77.6 (2014), p. 066901. URL: <http://stacks.iop.org/0034-4885/77/i=6/a=066901>.
- [23] *Collaborative Research Center SFB 876 - Providing Information by Resource-Constrained Data Analysis*. URL: <http://sfb876.tu-dortmund.de>.
- [24] *CTA - Cherenkov Telescope Array*. URL: <https://portal.cta-observatory.org/Pages/Home.aspx>.
- [25] C. Ding and H. Peng. “Minimum redundancy feature selection from microarray gene expression data”. In: *Bioinformatics Conference, 2003. CSB 2003. Proceedings of the 2003 IEEE*. Aug. 2003, pp. 523–528. DOI: 10.1109/CSB.2003.1227396.
- [26] Daniela Dorner et al. “FACT – TeV Flare Alerts Triggering Multi-Wavelength Observations”. In: *Proceedings of the 34th ICRC, 865, The Hague, 2015*. 704. 2015.
- [27] D. Dorner et al. “FACT - Long-term Monitoring of Bright TeV-Blazars”. In: *Proceedings of the 33rd International Cosmic Ray Conference* 686 (2013).
- [28] B. Efron. “Bootstrap Methods: Another Look at the Jackknife”. In: *Ann. Statist.* 7.1 (Jan. 1979), pp. 1–26. DOI: 10.1214/aos/1176344552. URL: <http://dx.doi.org/10.1214/aos/1176344552>.
- [29] *FACT - The First G-APD Cherenkov Telescope*. URL: <https://fact-project.org/>.
- [30] fact-project. *FACT-Tools Repository*. URL: <https://github.com/fact-project/fact-tools>.
- [31] J. Grube et al. “Observations of the Crab Nebula with the Whipple 10 m Telescope”. In: *Proceedings of the 30th ICRC*. Sept. 2007. arXiv: 0709.4300.
- [32] D. Heck et al. *CORSIKA: a Monte Carlo code to simulate extensive air showers*. Feb. 1998.
- [33] *H.E.S.S. - High Energy Stereoscopic System*. URL: <https://www.mpi-hd.mpg.de/hfm/HESS/>.
- [34] Viktor Franz Hess. “Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten”. In: *Physikalische Zeitschrift XIII* (1912).
- [35] Hester et al. *X-ray, Optical & Radio Composite image of the Crab Nebula*. X-Ray: NASA/CXC/ASU/ , Optical: NASA/HST/ASU/ , Radio: NRAO/AUI/NSF. URL: <http://chandra.harvard.edu/photo/2002/0052/more.html>.

- [36] A. M. Hillas. “Cerenkov light images of EAS produced by primary gamma”. In: *International Cosmic Ray Conference 3* (Aug. 1985), pp. 445–448.
- [37] *KASCADE - Karlsruhe Shower Core And Array Detector*. URL: <https://kcdc.ikp.kit.edu/>.
- [38] Ludmila I. Kuncheva. “A Stability Index for Feature Selection”. In: *Proceedings of the 25th Conference on Proceedings of the 25th IASTED International Multi-Conference: Artificial Intelligence and Applications*. AIAP’07. Innsbruck, Austria: ACTA Press, 2007, pp. 390–395. URL: <http://dl.acm.org/citation.cfm?id=1295303.1295370>.
- [39] R. W. Lessard et al. “A new analysis method for reconstructing the arrival direction of TeV gamma rays using a single imaging atmospheric Cherenkov telescope”. In: *Astroparticle Physics* 15 (Mar. 2001), pp. 1–18. DOI: 10.1016/S0927-6505(00)00133-X. eprint: astro-ph/0005468.
- [40] Miachel Levandowsky and David Winter. “Distance between Sets”. In: *Nature* 234 (Nov. 1971), pp. 34–35.
- [41] T.-P. Li and Y.-Q. Ma. “Analysis methods for results in gamma-ray astronomy”. In: *Astrophysical Journal, Part 1* 272 (Sept. 1983), pp. 317–324. DOI: 10.1086/161295.
- [42] Ingo Mierswa et al. “YALE: Rapid Prototyping for Complex Data Mining Tasks”. In: *Proceedings of the 12th ACM SIGKDD*. 2006.
- [43] N. Milke et al. “Solving inverse problems with the unfolding program TRUEE: Examples in astroparticle physics”. In: *NIM A* 697 (2013), pp. 133–147.
- [44] R. Mirzoyan and E. Lorenz. “Measurement of the Night Sky Light Background at La Palma”. In: *MPI-PhE/94-35* (1994).
- [45] Sebastian Müller. “Clear Sight in Cherenkov-Astronomy. Investigations of Reflector Geometry and Mirror Alignment for the FACT-Telescope”. MA thesis. TU Dortmund, 2014.
- [46] Maximilian Nöthe. “Ermittlung des zeitlichen und optischen Auflösungsvermögens des First G-APD Cherenkov Telescope mithilfe von Myon-Ringen”. MA thesis. TU Dortmund, 2016.
- [47] K. A. Olive et al. “Review of Particle Physics”. In: *Chin. Phys.* C38 (2014 and 2015 update), p. 090001. DOI: 10.1088/1674-1137/38/9/090001.
- [48] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.
- [49] Pence, W. D. et al. “Definition of the Flexible Image Transport System (FITS), version 3.0”. In: *A&A* 524 (2010), A42. DOI: 10.1051/0004-6361/201015362. URL: <http://dx.doi.org/10.1051/0004-6361/201015362>.



- 
- [50] William H. Press et al. *Numerical Recipes in C (2Nd Ed.): The Art of Scientific Computing*. New York, NY, USA: Cambridge University Press, 1992. ISBN: 0-521-43108-5.
- [51] M. Punch et al. “Detection of TeV photons from the active galaxy Markarian 421”. In: *Nature* 358 (Aug. 1992), p. 477. DOI: 10.1038/358477a0.
- [52] D Renker and E Lorenz. “Advances in solid state photon detectors”. In: *Journal of Instrumentation* 4.04 (2009), P04004. URL: <http://stacks.iop.org/1748-0221/4/i=04/a=P04004>.
- [53] Stefan Ritt, Roberto Dinapoli, and Ueli Hartmann. “Application of the {DRS} chip for fast waveform digitizing”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 623.1 (2010). 1st International Conference on Technology and Instrumentation in Particle Physics, pp. 486–488. ISSN: 0168-9002. DOI: <http://dx.doi.org/10.1016/j.nima.2010.03.045>. URL: <http://www.sciencedirect.com/science/article/pii/S0168900210006091>.
- [54] F. Schmidt and J. Knapp. *CORSIKA Shower Images*. 2005. URL: <https://www.ikp.kit.edu/corsika/>.
- [55] M. Sokolova et al. “Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation”. In: *Proceedings of the 19th Australian Joint Conference on AI: Advances in Artificial Intelligence*. 2006.
- [56] Horst Stöcker. *Taschenbuch der Physik*. Vol. 5. korrigierte Auflage. ISBN 3-8171-1721-3. Wissenschaftlicher Verlag Harri Deutsch, 2004.
- [57] Fabian Temme. “Analysis of Crab Nebula Data using PARFACT, a newly Developed Analysis Software for the First G-APD Cherenkov Telescope”. diploma thesis. TU Dortmund, 2013.
- [58] *The MAGIC Telescopes*. URL: <https://magic.mpp.mpg.de/>.
- [59] *Veritas - Very Energetic Radiation Imaging Telescope Array System*. URL: <http://veritas.sao.arizona.edu/>.
- [60] S. P. Wakely and D. Horan. “TeVcat: An online catalog for Very High Energy Gamma-Ray Astronomy”. In: *International Cosmic Ray Conference* 3 (2008), pp. 1341–1344.
- [61] T. C. Weekes et al. “Observation of TeV gamma rays from the Crab nebula using the atmospheric Cerenkov imaging technique”. In: *Astrophysical Journal, Part 1* 342 (July 1989), pp. 379–395. DOI: 10.1086/167599.
- [62] Tom White. *Hadoop: The Definitive Guide*. 1st. O’Reilly Media, Inc., 2009. ISBN: 0596521979, 9780596521974.

## Bibliography

---

- [63] Matei Zaharia et al. “Spark: Cluster Computing with Working Sets”. In: *Proceedings of the 2Nd USENIX Conference on Hot Topics in Cloud Computing*. HotCloud’10. Boston, MA: USENIX Association, 2010, pp. 10–10. URL: <http://dl.acm.org/citation.cfm?id=1863103.1863113>.

## Acknowledgment

Writing this thesis would not have been possible without help of a number of people. Over the last years they supported me in every way I could imagine and stood by my side. I owe all people a debt of gratitude, who enabled me to write this thesis.

First of all, I would like to thank Prof. Dr. Dr. Wolfgang Rhode for giving me the opportunity to participate in the very interesting field of astroparticle physics. The support I witnessed, as well the chance to attend conferences with talk and posters as working independently on my research topics, was great.

I would like to thank Prof. Dr. Kevin Kröniger for his time and efforts, being the second assessor of this thesis.

I thank the astroparticle physics group E5b at the TU Dortmund for endless interesting scientific (and non-scientific) discussions and a work environment which I will miss in future.

The FACT collaboration members are acknowledged for operating a great project. I am grateful to be part of such a motivating and inspiring collaboration.

Special thanks goes out to Kai Brügge, Lena Linhoff, Maximilian Meier, Anke Temme and Dr. Karl-Heinz Temme for proofreading this thesis.

Last but not least I would like to express my gratitude to my family for supporting me in every decision and every step taking in my life. I could not wish to have a better family.