Combined Fit Method to Identify a Nearby Cosmic Ray Source in Regions Indicating Anisotropy in Arrival Directions at the Pierre Auger Observatory

von

Daniel Maschmann

Masterarbeit in Physik

vorgelegt der

Fakultät für Mathematik, Informatik und Naturwissenschaften

der

Rheinisch Westfälischen Technischen Hochschule Aachen

im **September 2018**

angefertigt am

III. Physikalischen Institut A

Erstgutachter
Prof. Dr. Martin Erdmann
III. Physikalisches Institut A
RWTH Aachen University

Zweitgutachter
Prof. Dr. Thomas Hebbeker
III. Physikalisches Institut A
RWTH Aachen University

Contents

Chapter 1

Introduction

In 1912, Victor Hess published observation results of radiation measurements he conducted during 6 balloon flights up to an altitude of 5200 m. He concluded that beside radiation originating from radio active material at the Erath's surface an additional unknown component is contributing to the measured flux which increases with altitude. He closed his publication with the announcement to *'concentrate future research on this unknown radiation'* [1]. This *'future research'* rapidly grew during the following decades to a large field of fundamental research and brought him the Nobel Price in 1936 for the discovery of cosmic rays.
Today about 100 years later the energy spectrum of cosmic rays have been probed by numerous experiments over several orders of magnitude up to the highest energies exceeding 10^{20} eV. In 2003 the *Pierre Auger Observatory* started operating with the purpose to investigate the origin and nature of ultra-high energy cosmic rays (UHECRs) which still holds an unsolved question.
Recent works investigating the arrival direction of UHECR at the *Pierre Auger Observatory* favor an astrophysical scenario with nearby sources situated outside our Galaxy some Mpc away [2]. Another work, disconnected from arrival direction measurements, uses a combined method to fit the measured energy and atmospheric depth of extensive air showers induced by UHECRs to a source model. The most common approach is a simple astrophysical scenario assuming a homogeneous and isotropic source distribution injecting a representative mass composition following a power law spectrum with a maximal rigidity cutoff. Sources were characterized by a hard energy spectrum, low maximum injection energy and a heavy chemical composition [3].
In this work we modify the combined fit method by introducing a nearby point source to the homogeneous background. The aim is to establish an analysis which resolves the distance of a nearby source with a significant contribution.
This work is structured as follows: First, cosmic rays are introduced using a phenomenological approach, followed by a summary of observation measurements relevant to this work and possible acceleration mechanisms. Afterwards the simulation of sky regions with and without nearby sources expressed in energy and atmospheric depth measurement is described. Then the fit model is introduced considering a nearby UHECR source over a homogeneous background which extends the basic model in the combined fit. Afterwards the Markov Chain Monte Carlo based fit

algorithm is introduced. Finally, the performance of the method to fit the simulated sky regions is evaluated and discussed in the physical context.

Chapter 2

Detection of Ultra-High Energy Cosmic Rays

About hundred years after the famous balloon experiment of Victor Hess, numerous experiments have been built to better understand the nature of his discovery, the cosmic rays. Today we know that cosmic rays are charged particles, photons, and neutrinos, covering energies of many orders of magnitude. Cosmic rays above an energy of 10^{18} eV are called ultra-high energy cosmic rays (UHECRs) which still present a secret to mankind [4].
Cosmic rays produce extensive air showers in the Earth's atmosphere, which will be explained in this chapter. Furthermore the largest ground based experiment built to investigate UHECRs, the Pierre Auger Observatory, is presented with technical details of its measurement principle.

2.1 Extensive Air Showers

The Earth's atmosphere is constantly hit by cosmic rays which collide with air molecules, creating a cascade of secondary particles which further interact or decay. These cascades were first observed in the 1920s by identifying tracks of a single charged particle splitting into two secondary particles. Later on in the 1930s *Pierre Auger* and his co-workers discovered extensive air showers as they simultaneously measured these secondary particles with an array of detectors, placed in some distance from each other. Using this principle of coincident detection they achieved a measurement of extensive air showers by only instrumenting 1% of the effective area [4].

As the primary particle interacts with the atmosphere, different particle types are created. These can be roughly categorized into three cascade components: the electromagnetic, the hadronic, and the muonic component, which are schematically shown in Fig. 2.1. The hadronic component is created in an early stage of the shower development. Mainly pions and K-mesons are produced and can either decay or re-interact with other air molecules. Charged pions have a decay length of $l_d = \gamma_{\pi^\pm} \times 760\,\mathrm{cm}$, where $\gamma_{\pi^\pm}$ is the gamma factor of the pion, and represent the dominant part of the

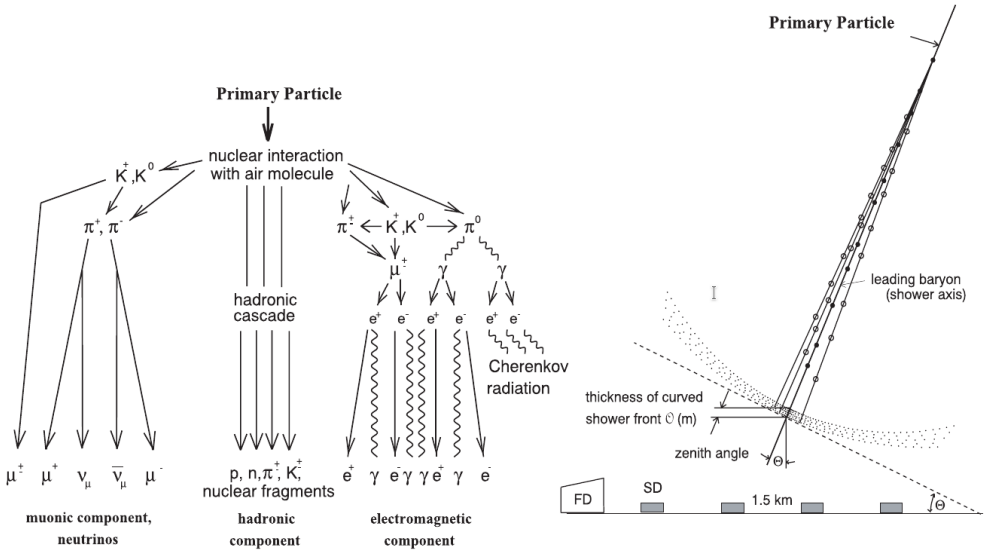


Figure 2.1: *left:* Schematic illustration of the three main components of an extensive air shower (electromagnetic, muonic and hadronic). *right:* Spacial development of an extensive air shower showing the shower front, the zenith angle, and the stations. Taken from [5].

hadronic shower component [4]. The electromagnetic component is the most numerous part and is represented by neutral pions, which directly decay after a mean free path of $l_d = \gamma_{\pi^0} \times 2.51 \times 10^{-6}\,\mathrm{cm}$ corresponding to a mean lifetime of $\tau_{\pi^0} = 10^{-16}\,\mathrm{s}$. The neutral pion decays into two photons which then probably cause electron-positron pair-production: $\gamma \to e^+ + e^-$ [4,5]. The third component, the muonic one, is mainly produced by the charged pion decay $\pi^+ \to \mu^+ + \nu_\mu$ and $\pi^- \to \mu^- + \overline{\nu}_\mu$. Due to relativistic time dilation the muons reach the Earth's surface before decaying, where they are measured with a ground detector [5].

A simple but powerful model of the shower development is the *Heitler Model*. Heitler described a cascade induced by a primary particle which interacts with air molecules and splits into two particles, each carrying half of the initial energy E_0. This is repeated after the interaction length λ where the number of particles at the step n grow to $N = 2^n$ and the energy of each particle is $E = E_0/2^n$. This splitting procedure is shown in Fig. 2.2 and is repeated till the particles reache a critical energy $E_\mathrm{c} = N_0/N_\mathrm{max}$ which is insufficient to create new particles [6].

The production rate of secondary particles depends on the amount of matter which is traversed by the shower. This is described by the *atmospheric depth* which is defined as:

$$X(h_0) = \int_{h_0}^{\infty} \rho(h) dh\,, \tag{2.1}$$

where h is the geometrical height [5].

The longitudinal development of the shower is parametrized by the Gaisser-Hillas formula, which computes the number of produced electrons N_e as a function of at-

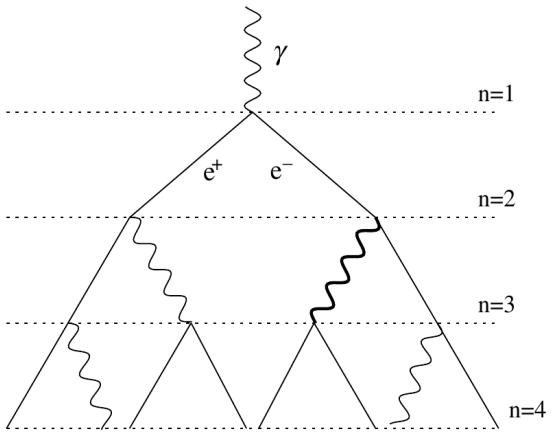


Figure 2.2: Schematic view of the *Heitler Model* to describe an electromagnetic cascade. Photons create an electron-positron pair which then radiate a bremsstrahlung photon. Taken from [7].

mospheric depth [4]:

$$N_e(X) = N_e^{\max}\Big(\frac{X - X_1}{X_{\max} - \lambda}\Big)^{\frac{X_{\max}-\lambda}{\lambda}} \exp\Big(\frac{X_1 - X}{\lambda}\Big) \tag{2.2}$$

where $N_e^{\max}$ is the number of electrons during the shower development denoted as the atmospheric depth $X_{\max}$, λ is the mean free path of the primary particle and X_1 the atmospheric depth of its first interaction. The value of $X_{\max}$ depends on the initial energy of the primary particle E_0 and the critical energy E_c, where the particle loss rate dominates over the particle production rate, and is computed as [5]:

$$X_{\max} = \lambda \frac{\ln(E_0/E_c)}{\ln(2)} \tag{2.3}$$

A semi-empirical approach orientated to the *Heitler Model* showed that the number of muons in an air shower is related to the number of decaying pions. Therefore i is possible to model the energy of the primary particle and additional shower properties, e.g. the elongation rate [7].
Extensive air showers are studied using Monte Carlo based simulations which are compared to shower measurements. For instance, $X_{\max}$ also depends, in addition to the initial particle energy, on its mass [4]. Using superposition models which include the hadronic shower component the cosmic ray mass composition is studied [8].
Nevertheless, we can not directly conclude the energy and composition of the primary particle from individual shower measurements because of fluctuations in the shower development. It is only possible to infer this in average for a reasonably large statistical sample. The fluctuation of the $X_{\max}$ measurement can be described with a probability density which is approximated with the generalized Gumbel distribution [9]:

$$p(X_{\max}) = \frac{\lambda^\lambda}{\sigma\Gamma(\lambda)}\Big(e^{-\lambda z - \lambda e^{-z}}\Big), z = \frac{X_{max} - \mu}{\sigma}. \tag{2.4}$$

The parameters σ, μ and λ describe the shape, location, and the scale of the shower and depend on the cosmic ray energy and mass. These parameters have been obtained empirically by fitting the Gumbel function 2.4 to simulated air showers. The air shower simulation are based on different hadronic interaction models, namely EPOS-LHC, QSHetII-04, and Sibyll 2.1 [9]. The Gumbel distribution is presented in

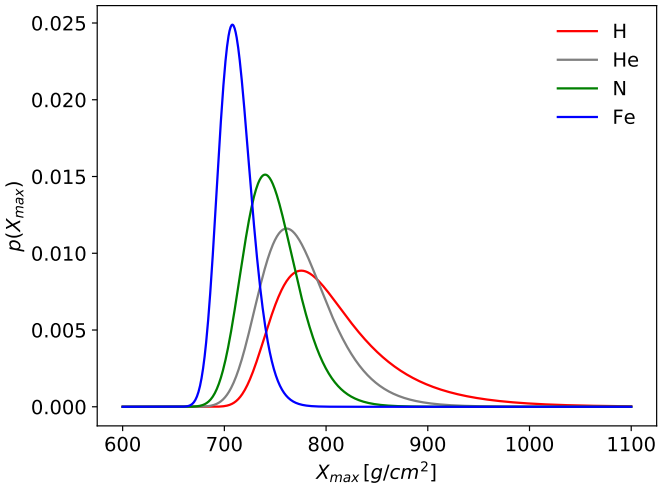


Figure 2.3: The Gumbel distribution giving the probability distribution of measuring a $X_{\max}$ value of a cosmic ray with an energy of 10^{19} eV. The distribution is presented for different elements.

Fig. 2.3 for hydrogen, helium, nitrogen, and iron at the energy of 10^{19} eV using the interaction model EPOS-LHC.

2.2 The Pierre Auger Observatory

The energy spectrum of cosmic rays follows a declining power law, thus the cosmic ray flux decreases towards higher energies. At a certain point the flux is too small to measure cosmic rays directly with satellite experiments. Ground based experiments are needed to detect extensive air showers. The flux of cosmic rays with energies exceeding 10^{18} eV is about three particles per km^2 per steradian per century [4]. To accumulate enough data to explore the nature of UHECRs, detectors have to cover a large amount of surface. The largest ground based experiment for this purpose is the Pierre Auger Observatory. It covers a surface of about 3000 km^2 and is situated in the Argentinian Pampa, in the province of Mendoza. Its average altitude is $\sim$ 1400 m which corresponds to an atmospheric depth for vertical cosmic ray air showers of around 875 g/cm^2, ideal for the detecttion of UHECR showers shortly after their maximum shower development. Furthermore, light pollution from civilization is very small due to the remote location [10].

In Fig. 2.4 the layout of the Pierre Auger Observatory with the detector alignment is

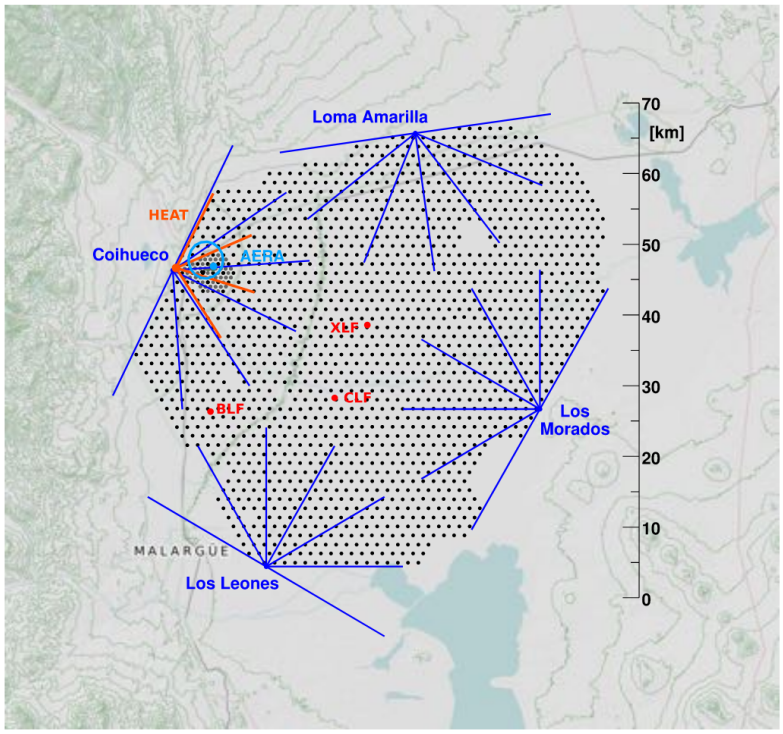


Figure 2.4: Map of the Pierre Auger Observatory [11]. The surface detectors (SD) are marked as black dots, the blue lines are the viewing angle of the fluorescence detectors which are located at four sites around the area covered by SD. Light-blue lines indicate the viewing angle of the low-energy extension HEAT and grey dots mark the more dense SD array with a narrow spacing of 750 m. The blue circle shows the position of AERA.

shown: 1660 *surface detector stations* (SD), consisting of water-Cherenkov detectors, are arranged at a 1.5 km spacing in a triangular grid over the whole area, surrounded by 24 *fluorescence telescopes* (FD) observing the night sky. These two detector types represent the base line detectors. This hybrid design allows cross-calibrations and cross-checking of measurements [4]. Besides the base line detectors additional experiments are placed in the field: The Auger Engineering Radio Array (AERA) measures the radio emission of electromagnetic shower components [12]. The *infill* array is a low-energy extension where the spacing of 61 SD stations is reduced to 750m. The sky above this area is observed by the *High-Elevation Auger Telescope* (HEAT) which covers a lower energy range between 10^{17} eV and 10^{18} eV [10]. To improve the reconstruction of the mass composition, SD stations are currently in the process of beeing equipped with scintillator detectors (*Auger Prime*) [11].

2.2.1 Fluorescence Detector

Charged particles in a cosmic ray air shower excite atmospheric nitrogen, which then emits light in the wavelength band of $300 - 430$ nm [13]. The detection of UHECRs

through the measurement of fluorescence light is a well established technique already used by previous experiments like the Fly's Eye [14]. The Pierre Auger Observatory measures fluorescence light with FD during moonless nights if the weather condition is suitable, thus the duty cycle is approximatly 15 % [10].

Four FD eyes, one shown on the left in Fig. 2.5, are situated at the perimeter of

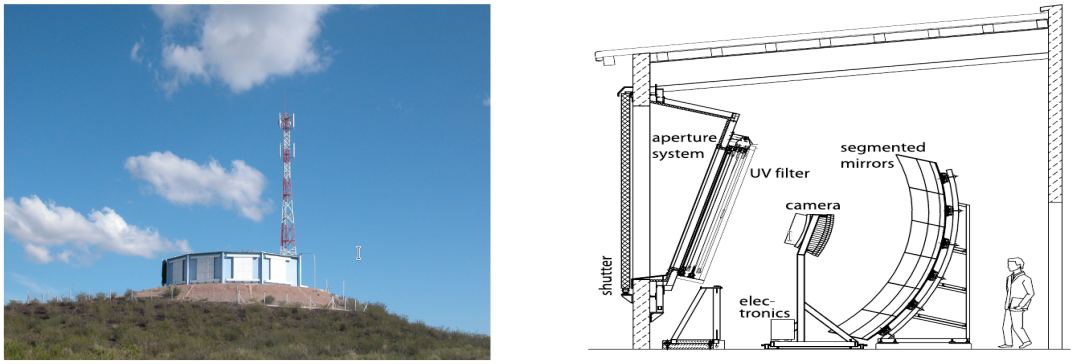


Figure 2.5: *left*: Fluorescence detector station with closed shutter. Taken from [15]. *right*: Illustration of the layout of a flourescense telescope with its main components. Taken from [13].

the detector array, each consisting of 6 fluorescence telescopes with a $30° \times 30°$ field of view. The telescope layout is shown on the right in Fig. 2.5. Each telescope has a camera positioned in the focus of a $10\,\mathrm{m}^2$ segmented mirror. The camera consists of 440 photomultiplier tubes (PMT) arranged in a hexagonal grid. The measured UV-light is digitized every 100 ns and a hierarchical trigger mechanism controls the recording of cosmic ray air showers. The detector reaches a trigger efficiency of 100 % over the whole area for energies exceeding $10^{19}\,\mathrm{eV}$ [13]. An example event, where all four telescopes were triggered by an extensive air shower, is shown in Fig. 2.6 on the left side.

The number of emitted fluorescence photons in a cosmic ray air shower is proportional to the electromagnetic energy loss through charged particles in the atmosphere. This fluorescence light is detected by FD. Using time information of triggered PMTs the longitudinal shower profile can be read out as a track on the telescope camera. The shower shape can be reconstructed by fitting a Gaisser-Hillas function (see equation 2.2) to the measured profile. The total energy of the initial particle can be obtained by integrating the longitudinal profile. Shower components like neutrinos and muons have to be taken into account since they do not emit UV-light and are invisible to the FD detectors. Through the geometrical determination of the shower shape, the arrival direction is reconstructed as shown in Fig. 2.6. The energy resolution for the reconstruction is $\leq$ 10 % with a 14 % systematic energy scale uncertainty [10, 13].

2.2.2 Surface Detector

Each of the *surface detector* stations, shown in Fig. 2.7, contain 12,000 l of pure water. At the top of the tank 3 photomultiplier tubes (PMT) are installed to detect

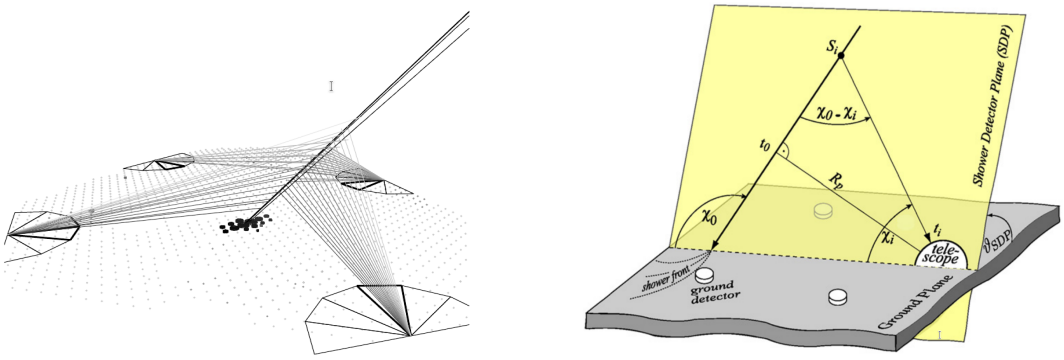


Figure 2.6: *left*: Schematic observation of an UHECR event observed by four FD stations and simoultaneously by SD stations. *right*: Geometry for the reconstruction of the arrival direction and shower detector plane using time information of triggered pixels from FD. Both taken from [13].

Cherenkov light, which is produced when relativistic particles traverse the water with a velocity faster than the speed of light in water. To increase the sensitivity, the inner surface is lined with reflective material. All stations are equipped with two solar panels and two acid-lead car batteries to ensure the energy supply. To communicate with the central data acquisition system a radio transceiver and an antenna is mounted on top of each SD station. The SD stations have a robust design for an estimated lifetime of 20 years [10].

The PMT signals are converted into a digital signal counting in vertical-equivalent-

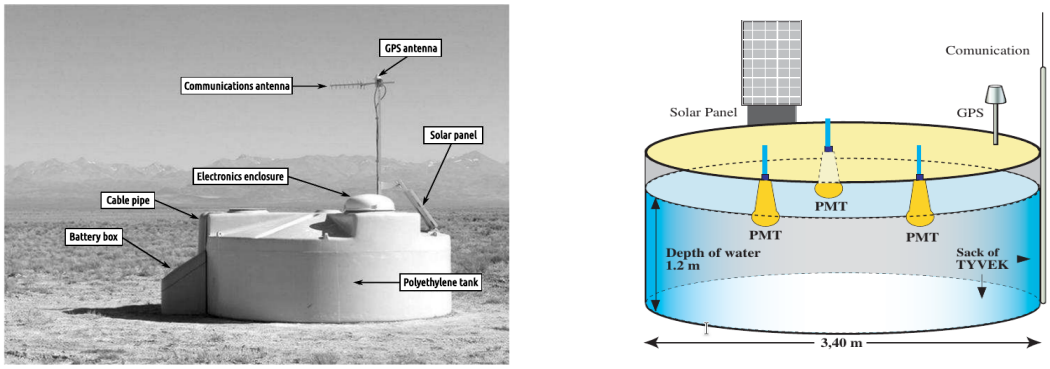


Figure 2.7: *left*: Picture of a SD station with description of main components, taken from [16]. *right*: schematic view of SD station, taken from [5].

muons (VEM), which is the signal that a vertical muon would produce by crossing the detector vertically. A hirarchical trigger system consisting of online and offline trigger separates cosmic ray signals from noise and randomly triggered stations. The trigger system also searches for coincident measurements from neighboring stations and ensures a well defined footprint for later event reconstructions. Above an energy of 3 EeV the *surface detector* becomes fully efficient [16].

When a cosmic ray shower front hits the ground, different stations are triggered

at different times and it is possible through geometrical calculations to reconstruct the arrival direction. Additionally, the shower shape, and therefore the footprint on the ground, varies for different zenith angles, different energies and masses of the primary particle. The reconstruction of the arrival directions has an uncertainty of less than 2^o for an energy above $10^{18.5}$ eV and decreases towards higher ernergies [17]. In combination with the FD reconstruction the total angular resolution depends on the cosmic ray energy and number of triggered stations. Towards higher energies the resolution increases and gets better than 1° for energies exceeding 10^{19} eV [18].

Chapter 3

Observation Measurements

The cosmic ray energy spectrum is surveyed over more than 12 orders of magnitude with numerous experiments revealing features giving hints to the underlying origin and acceleration mechanisms. Especially for the highest energies the composition and arrival direction are still an object of today's research. This chapter will point out the most important measurements of cosmic rays to draw a picture of the current state of research with a special focus on UHECRs and measurements done by the *Pierre Auger Observatory*.

3.1 Energy Spectrum

The flux of cosmic rays depends strongly on the energy and therefore cosmic rays are described by the differential flux which gives the number of particles per surface A, time t, energy E and solid angle Ω [10]:

$$J(E) = -\frac{d^4N}{dE\,dA\,d\Omega\,dt}(E) \tag{3.1}$$

The cosmic ray energy spectrum is shown on the left in Fig. 3.1, observed by different experiments and covering energies from 10^8 to beyond 10^{20} eV. Energies below 10^9 eV are deaccalerated by solar winds, which shields the solar system from these particles [10]. The energy spectrum exceeding energies of 10^{11} eV is described by an almost featureless power law spectrum $J(E) \propto E^{-\gamma}$. Its spectral index γ varies around 3 and has two transition regions. To discuss observations of different parts of the energy spectrum an important aspect is that the particle flux, for example at 10^{11} eV is by sixteen orders of magnitude larger then the flux at $10^{20}\,texteV$. At energies below 10^{15} eV direct measurements are possible with air balloon or satellite experiments, whereas for energies exceeding 10^{15} eV the only way to measure cosmic rays is through detecting extensive air showers [4].

The cosmic ray energy spectrum becomes steeper at energies between 10^{15} and 10^{16} eV: the spectral index changes from $\gamma \sim 2.6$ to 3.1; this transition is the so called 'knee'. The spectral index changes again at energies between 10^{18} and 10^{19} eV to $\gamma \sim 2.6$, which is the so called 'ankle'. At energies above 4×10^{19} eV, the flux is

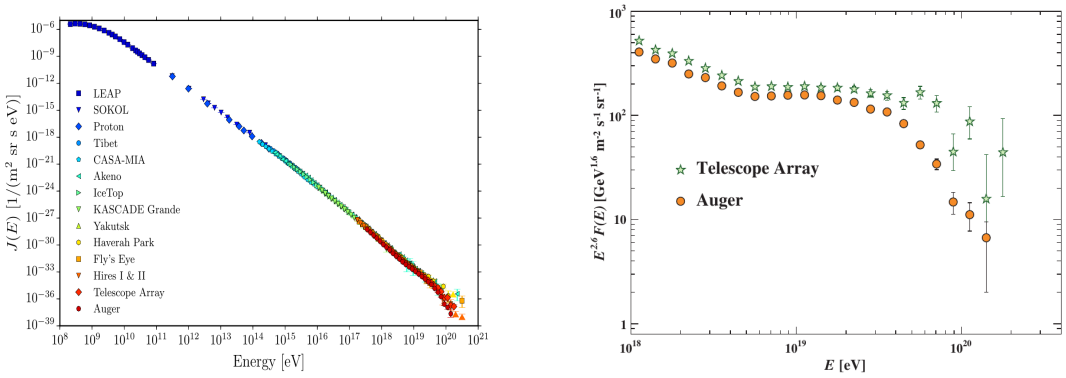


Figure 3.1: *left*: Cosmic ray energy spectrum measured by numerous experiments. Taken from [10] *right*: Spectrum of energies above 10^{15} eV which is measured by air shower experiments and muliplied by E^3 to visualize features in the slope and the abrupt cut-off as described in the text. Taken from [19]

strongly suppressed and indicates a limit to cosmic rays reaching the Earth [10]. These features are from special interest since the transition regions seem to indicate different acceleration mechanisms and propagation effects (discussed in detail in section 4.1). On the right hand side in Fig. 3.1 the tail of the cosmic ray energy spectrum is shown, which is important in this work, since it could be connected to acceleration sites of UHECRs. The shown spectrum is multiplied by E^3 to cancel out the slope and visualize the cut-off.

3.2 Composition

Numerous experiments performed long-term measurements for energies below 10^{15} eV with sophisticated detectors that have a high charge and energy resolution to identify different particle types. The measured flux is shown from different experiments in Fig. 3.2 for all measured elements. The composition is dominated by hydrogen but also heavier nuclei up to iron show a non negligible contribution to the spectrum. By correcting the measured flux for contributions from solar winds or attenuation effects during the propagation it is possible to rule out their composition at the acceleration sites. The results lead to a similar abundance of elements as assumed in the universe with two remarkable exceptions: Hydrogen and Helium exhibit an under-abundance, which poses an unsolved question [4].

For cosmic rays exceeding 10^{15} eV the only way to obtain information on the initial particles is through measuring extensive air showers. As outlined in section 2.1, cosmic ray air showers fluctuate and inferring properties of the primary cosmic ray is only possible with a reasonably large statistical sample. The most sensitive observable related to the composition is the atmospheric depth of the shower maximum $X_{\max}$. Due to the larger cross section of an iron nuclei the air shower develops higher in the atmosphere in comparison to a primary proton at the same energy. In average this leads to a difference of about $100\,\mathrm{g/cm^2}$ in $X_{\max}$. However, shower-to-shower fluctua-

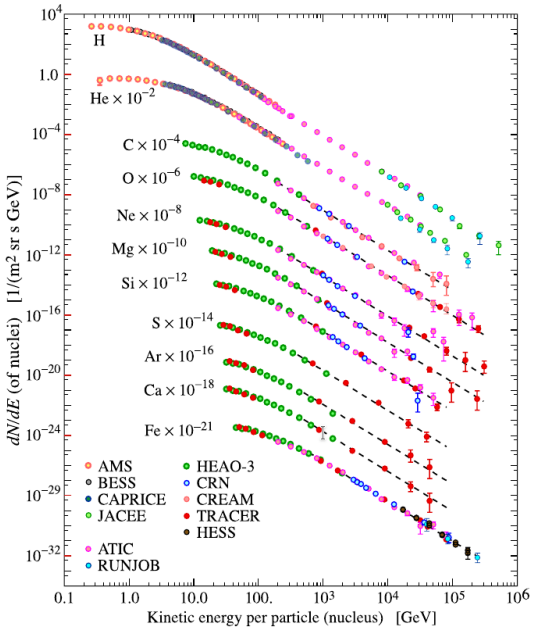


Figure 3.2: Cosmic ray flux measured for different nuclei as function of their kinetic energy. For a better visualization different nuclei have been scaled. Taken from [20]

tions of the order of ~ 20 to $60\,\mathrm{g/cm^2}$ and a typical detector resolution of $\sim 20\,\mathrm{g/cm^2}$ makes it challenging to conclude individual nuclear charges for the primary cosmic ray. This fluctuation in the X_{max} observable can be described by the Gumbel distribution, see 2.1. Nevertheless for larger samples it is feasible to conclude a statistical distribution out of the X_{max} observable. In Fig. 3.3 the mean and the Gaussian width of the X_{max} distribution measured by the *Pierre Auger Observatory* are shown as a function of energy. For comparison with theoretical air shower simulations, predictions for a pure proton and a pure iron composition are displayed. The measurements are compatible with a proton-like composition for energies at around $10^{18}\,\mathrm{eV}$ and shift to a heavier composition towards higher energies. The highest measurements do not exceed $10^{19.5}\,\mathrm{eV}$ since X_{max} is measured by FD which can not provide high statistics due to its low duty-cycle [10]. A current upgrade with scintillator detectors to the SD stations will in future infer further composition information from the shower footprint. The new detectors allow the reconstruction of muons and electromagnetic particles which reveals an additional insight of the shower development [11].

3.3 Anisotropy

To understand the nature of cosmic rays besides the study of energy spectrum and composition the arrival direction is the main observable to identify their sources. However no point sources have been identified yet at any range of ultra-high energies and the observed arrival directions are found to be largely isotropic [10]. Recent studies on arrival directions of UHECRs with the *Pierre Auger Observatory* measured

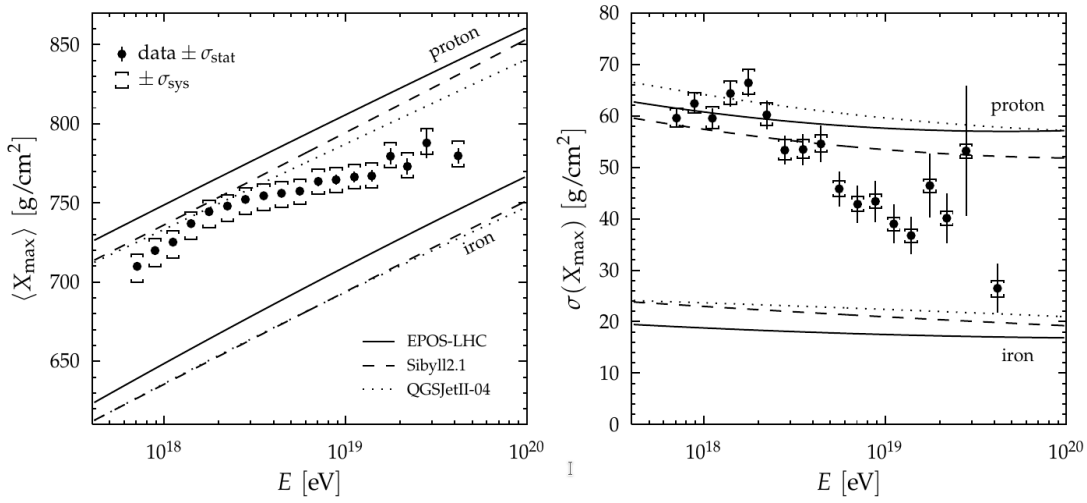


Figure 3.3: *left*: Mean atmospheric depth and *right*: the rms of the shower maximum $X_{\max}$ as a function of energy measured at the *Pierre Auger Observatory*. Additional lines show theoretical predictions for a pure proton and iron composition with different interaction models. Taken from [21]

a significant anisotropy on large scales [22]. Additionally a study in arrival directions on intermediate scales found evidence for correlations between an excess in arrival directions and nearby starburst galaxy positions [2].

3.3.1 Large Scale Anisotropy

In 2017 *The Pierre Auger Collaboration* reported the discovery of a statistical significant access in arrival direction with a dipolar structure towards the right ascension at $\alpha = 100^\circ \pm 10^\circ$ and the declination at $\delta = -24^{\circ+12^\circ}_{-13^\circ}$. Using 3×10^4 events above $8 \times 10^{18}\,\mathrm{eV}$ the *Pierre Auger Observatory* reject isotropy by 5.2σ significance. This anisotropy was modeled by a dipole structure with a $6.5^{+1.3}_{-0.9}\,\%$ amplitude. The normalized event rate is computed as a function of the right ascension, fitted by the first harmonic modulation ($\chi^2/ndf = 10.5/10$) and is shown on the left in Fig. 3.4.
The analysis, detailed in [22], concluded an extra galactic origin due to the miss match with the direction of the galactic center and the dipole position. The 2MRS catalog of infrared detected galaxies exhibits a dipole structure but lays 55° away from the measured dipole. After considering deflections in the galactic magnetic field the 2MRS dipole structure points towards the excess region indicated by arrows on the right in Fig. 3.4. The dipole structure is not significant for energies of $4\,\mathrm{EeV} < \mathrm{E} < 8\,\mathrm{EeV}$. This can be explained with stronger deflection in the galactic magnetic field supporting the hypothesis of an extragalactic origin of UHECRs [22].

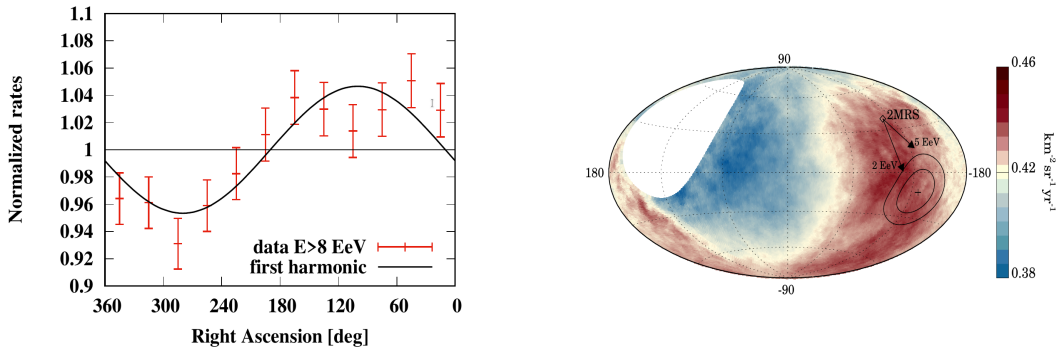


Figure 3.4: *left*: Normalized rate of events as a function of right ascension with a fit of a first-harmonic modulation. The fit shows good agreements with the data: ($\chi^2/ndf = 10.5/10$). *right*: The flux of cosmic rays above 8 EeV in Galactic coordinates. A cross indicates the direction of the reconstructed dipole. The position of the dipole maximum position resulting from the 2MRS catalog is displayed and its deflection in a galactic magnetic field model indicated by arrows. Both taken from [22]

3.3.2 Anisotropy Indication Around Nearby Starburst Galaxies

A more recent analysis found evidence for anisotropy in the arrival directions on an intermediate angular scale in a data set of the *Pierre Auger Observatory* [2]. A Sky model of UHECR sources has been constructed using two distinct catalogs of gamma-ray sources observed by the Fermi Large Area Telescope: active galactic nuclei (AGN) and starburst galaxies (SBG). The individual flux of the sources is approximated by their radio luminosity. For each catalog the model was constructed using two free parameters: the fraction of anisotropy and the search radius to account for deflections since no galactic magnetic field model was used.

A set of 5514 events above an energy of 20 EeV with zenith angles up to 80°, measured before 2017 April 30 by the *Pierre Auger Observatory* was considered in the analysis. The models have been tested by comparing with an isotropic null-hypothesis [2]. Fig. 3.5 shows on the left the test statistic (TS) which quantifies the deviation from the null-hypothesis and was found to be the most significant for the SBG-catalog at an energy threshold of 39 EeV. The parameter scan for this energy threshold using the maximum-likelihood approach is shown on the right of Fig. 3.5 leading to a best fit value of $13^{\circ}{}^{+4^{\circ}}_{-3^{\circ}}$ search radius and an anisotropy fraction of $10 \pm 4\%$. After penalizing for the energy scan and the investigation of one more catalog, the SBG model was found to describe the data better than the null-hypothesis with a significance of 4.0σ for UHECRs exceeding an energy of 39 EeV [2].

In Fig. 3.6 excess maps of UHECRs above 39 EeV are shown. On the right, one finds the excess observed by the *Pierre Auger Observatory* and on the left the modeled excess map using the best fitting parameter for the starburst galaxy scenario is presented. Additionally the three strongest starburst galaxies namely *M83*, *NGC4945*,

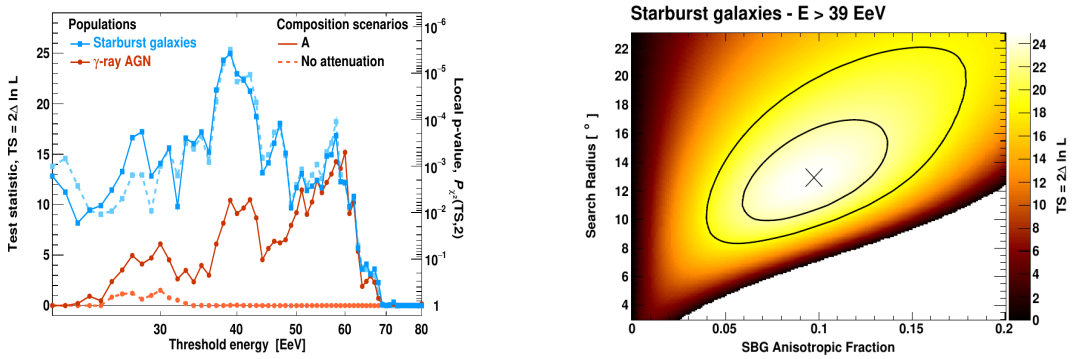


Figure 3.5: *left*: Scan of test statistic for comparison between the AGN and SBG catalog and isotropic expectation as function of the threshold energy. At $E > 39\,\mathrm{EeV}$ the deviation from isotropy maximize for the SGB catalog. *right*: Scan of the best-fit observables of the SBG catalog in the analysis for energies above 39 EeV. Both taken from [2].

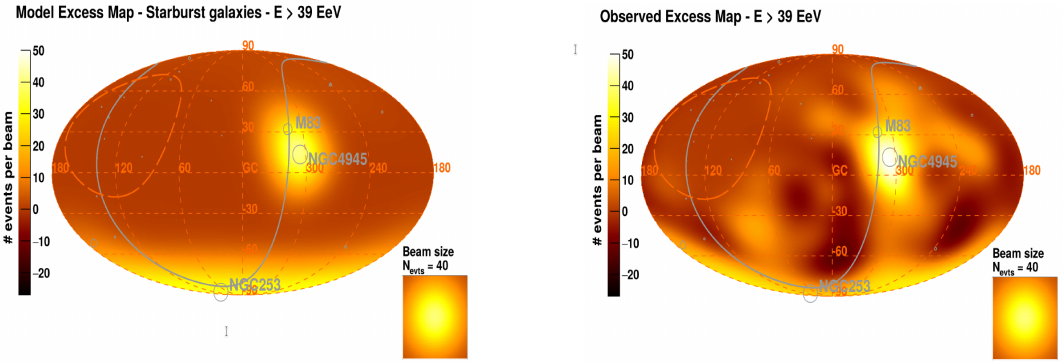


Figure 3.6: *left*: Excess map of cosmic rays reconstructed with the SBG catalog above 39 EeV. *right*: Excess map observed with the *Pierre Auger Observatory*. Both taken from [2]

and *NGC253* are indicated. However, starburst galaxies only show correlation evidence with arrival directions of UHECRs in an analysis where a realistic behavior of the galactic magnetic field is not considered. Moreover, a connection between observed large scale anisotropy and intermediate isotropy has not been identified yet [2].

3.4 Combined Fit Results

An analysis investigating UHECR sources, disconnected from arrival direction measurements, was carried out in [10] and [3]: A combined fit of an astrophysical model to the energy and shower depth measurement, recorded at the *Pierre Auger Observatory*, reveals source properties and constrains theoretical predictions.
The astrophysical model assumes identical UHECR sources uniformly distributed in a comoving volume. The sources have been characterized by a rigidity-depending

acceleration mechanism and an energy spectrum described by a power-law, injecting a nuclei composition of ^1H, ^4He, ^{14}N and ^{56}Fe as representative elements[1]. The injected spectrum has the form [10]

$$J_0(Z_0, E_0) = \phi_0 \cdot a(Z_0) \cdot (E_0)^{-\gamma} \cdot f_{cut}\Big(\frac{E_0/Z_0}{R_{cut}}\Big), \tag{3.2}$$

where ϕ_0 is the absolute flux level of the UHECR source, $a(Z_0)$ represents the abundance of the injected element Z_0 and γ is the spectral index of the energy spectrum. The function f_{cut} is a broken exponential function and describes the cut-off behaviour for maximal accelerated rigidities at the source [10]:

$$f_{cut}\Big(x = \frac{E_0/Z_0}{R_{cut}}\Big) = \Theta(1-x) + \Theta(x-1) \cdot e^{1-x} \tag{3.3}$$

where R_{cut} is the maximal rigidity and Θ the Heaviside function. The propagation through the universe to Earth was simulated with *SimProp* [23] and *CRPropa 3* [3] (for a detailed description of *CRPropa 3* see section 4.2). The composition measurements on Earth have been realized for different hadronic interaction models to generate the $X_{\max}$ observable.

Using a Bayesian approach with a Markov Chain Monte Carlo method, the spectral

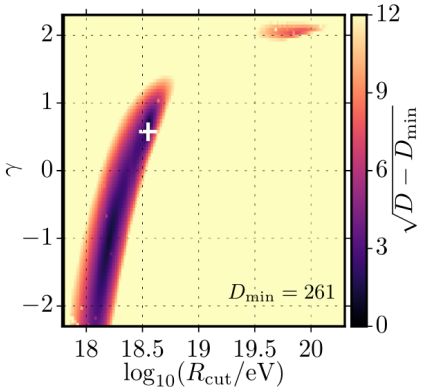


Figure 3.7: Scan in γ-R_{cut} plane of the goodness-of-fit D . The color code displays a pseudo standard deviation $n_{sigma} = \sqrt{D - D_{min}}$ according to the best fit value at D_{min}. Taken from [10].

index γ, the maximum rigidity R_{cut} and the chemical fractions at the source are fitted. A scan of the likelihood space for discrete values in γ and R_{cut} is shown in Fig. 3.7 and reveals the characteristic behavior of the combined fit. A valley-like structure is visible which is a general feature of the combined fit method [3]. Another feature is the visibility of two minimums: the global minimum at $\gamma \approx 1$ and a low rigidity-cut of $R_{cut} \approx 10^{18.5}$ eV and a second local minimum at $\gamma \approx 2$ and $R_{cut} \approx 10^{19.7}$ eV. Table 3.1 shows the the best fit values from [10] and Fig. 3.8 presents the best-fit-values

[1] In [3] 5 elements have been injected to enhance the goodness-of-fit by adding ^{28}Si

Parameter	$\hat{\theta}$	$\hat{\theta}$
	First minimum	Second minimum
$log_{10}(R_{cut}/eV)$	18.56	19.88
γ	0.62	2.03
$H[\%]$	0.1	0.0
$He[\%]$	0.1	0.3
$N[\%]$	98.5	92.0
$Fe[\%]$	1.2	7.7

Table 3.1: Best fit parameter results from the baseline scenario specified in [10]. Displayed are the global maximum and a second local maximum found in Fig. 3.7.

as simulated measurements of the energy spectrum in contrast to measurements from the *Pierre Auger Observatory*.

The Combined fit was carried out using different scenarios and interaction models

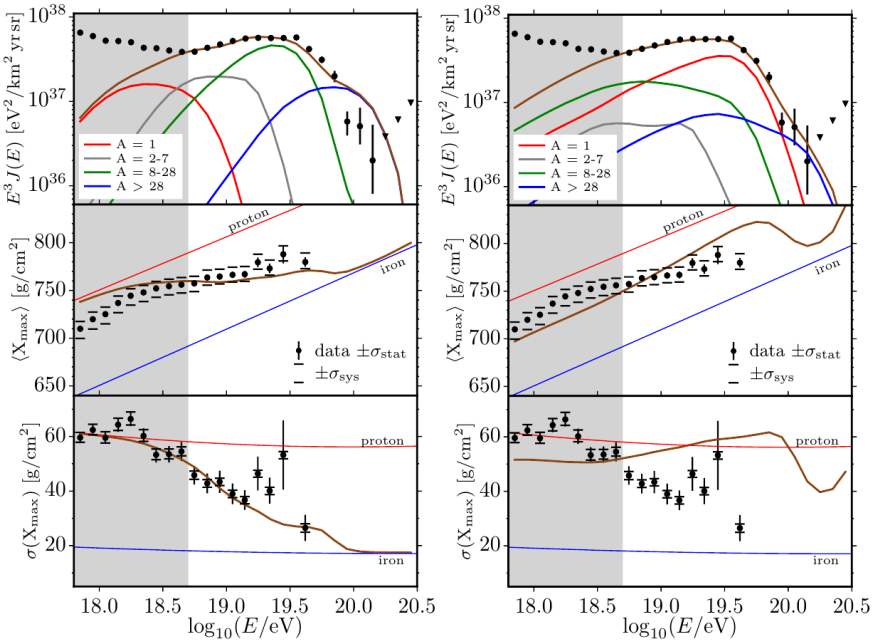


Figure 3.8: Comparison of energy spectrum and moment of the $X_{\max}$ observable measured at the *Pierre Auger Observatory*. Reconstruction with the best fit results for the global minimum on the *left* and the local minimum on the *right*. Taken from [10].

leading to slightly different results then presented in Table 3.1. But the analysis concludes a constrain in the γ and R_{cut} parameter as shown in Fig. 3.7 with low maximum injection energies and a chemical composition dominated by heavy elements [3,10,24].

Chapter 4

Origin and Propagation

This chapter will give a brief introduction about our current knowledge of possible sources. Furthermore the propagation from possible sources towards the earth will be discusses presenting the state-of-the-art simulation tool *CRPropa3*, taking all known interaction and effects into account.

4.1 Acceleration and Possible Sources

The highest measured energies of cosmic rays exceed 10^{20} eV which is about 16 J and is still an object of today's research [10]. Especially the acceleration mechanism of UHECRs is still unsolved. However, some well founded theories exist about possible acceleration mechanisms and at lower energies significant results have been made during the last twenty or thirty years [4]. Today we believe that cosmic rays arriving at Earth with an energy below 10^{15} eV, which is the energy range below the so called 'knee', are accelerated in supernova remnants inside our Galaxy. Cosmic rays with an energy between the so called 'ankle' and the 'knee' which corresponds to energies between 10^{16} eV and 10^{18} eV are assumed to originate from either galactic or extragalactic sources. Evidence favors that the highest cosmic ray energies above 10^{19} eV are originating from sources outside our Galaxy [4, 22].
Theories developed to explain the origin of UHECRs can be categorized into two fundamentally different approaches: '*top-down*' models and '*bottom-up*' models. In the '*top-down*' model the basic idea is that the cosmic rays we observe are decay products of an initial very massive particles with a mass of around 10^{25} eV. These models lead to distinct features like a flat injection spectrum and a specific particle composition since the decay chain leads to elementary particles and nucleons. In detail this would imply that the gamma-ray and neutrino flux would be thirty times higher than the nucleon flux [4]. This is clearly disfavored by measurements of the longitudinal air shower profile by the *Pierre Auger Collaboration* [25].
To study theories which uses the *bottom-up* approach describing the acceleration of standard model particles, one has to differentiate between two approaches: single shot and stochastic acceleration. To achieve high enough energies with a single shot, strong electric or magnetic fields are needed. This can be found in *Magnetars* which

are young rapidly rotating neutron stars with a strong magnetic field. These objects are potentially able to accelerate cosmic rays to the highest energies we observe on Earth. The energy, accelerated particles achieve, depends on the magnetic field strength and thus the rotational speed of the *magnetar* which decreases with time leading to an energy spectrum following a spectral index of $\gamma \approx 1$ [10].
Enrico Fermi first introduced the idea of stochastic acceleration for charged particles which scatter many times on massive clouds with turbulent magnetic fields. The clouds move randomly with a velocity v_{cl} and depending on the direction the particle enters the cloud, it can lose or gain energy (detailed calculation in [4]). It comes out that the average energy gain of the particle is then proportional to v_{cl}^2 [4].
A more efficient acceleration mechanism is when the particle encounters a shock front. A shock front is a wave which propagates with a higher speed than its speed of sound. In this case the acceleration is proportional to v_{cl}. Shock fronts are assumed to occur in supernova remnants, jets of active galactic nuclei, gamma ray bursts in starburst galaxies, or in accretions shock around massive galaxy cluster [10].
In the case of stochastic acceleration the particle at each acceleration cycle has an energy gain ΔE and thus the relative gain $\xi := \frac{\Delta E}{E}$. After n acceleration cycles the particle energy will be:

$$E_n = E_0(1+\xi)^n \tag{4.1}$$

At each cycle the particle has a certain probability P_{esc} to escape the region occupied by the magnetic field. The number of particles exceeding a certain energy E_n is proportional to those who remain in the acceleration system:

$$N(>E_n) = N_0 \sum_n^{\infty} (1-P_{\mathrm{esc}})^n \tag{4.2}$$

This ansatz leads to a power law spectrum:

$$N(>E_n) \propto \left(\frac{E_n}{E_0}\right)^{1-\gamma} \tag{4.3}$$

with $\gamma = 1 - \ln(1-P_{\mathrm{esc}})/\ln(1+\xi)$ [10]. For non-relativistic shock acceleration the spectral index is $\gamma \approx 2$ and $\gamma \approx 2.2$ to 2.3 for relativistic shock acceleration.
An acceleration site needs to preserve the particles inside the magnetic field region. This leads to a maximal kinetic energy the particle can gain before leaving the site depending on its size and magnetic field strength: $E_{\mathrm{max}} = e \cdot Z \cdot B \cdot R$, where R is the linear dimension of the site, B its magnetic field strength and Z the charge number of the particle. The Hillas criterion is based on the gyroradius of the particle to estimate the size and magnetic field strength to constrain possible sources:

$$\left(\frac{B}{\mathrm{G}}\right)\left(\frac{R}{\mathrm{pc}}\right) > \frac{0.2}{\beta_{\mathrm{sc}} Z}\left(\frac{E}{10^{20}\,\mathrm{eV}}\right) \tag{4.4}$$

β_{sc} denotes the average velocity of the scattering center. The graph shown in Fig. 4.1 often referred to as the 'Hillas plot', is presents the constraints for the magnetic field strength and gyroradius expected for a maximal acceleration of cosmic rays. Additional different astrophysical and cosmological objects are displayed according to their size and magnetic field strength [4].

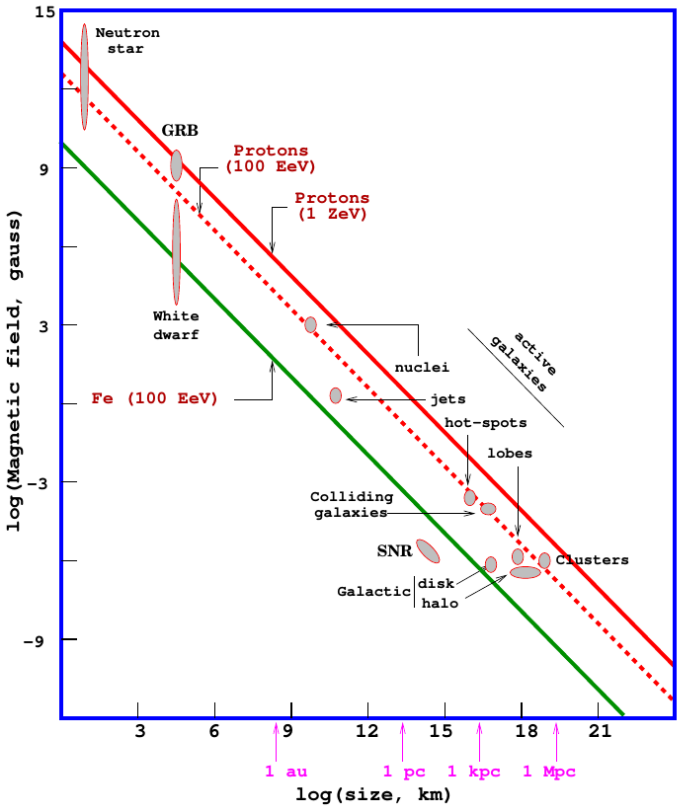


Figure 4.1: The Hillas diagram visualizes the hillas criterion (c.f. equation 4.4) to accelerate a proton or a iron nuclei up to an energy of 10^{20} eV.The possible accelerator sites are marked according to their magnetic field strength and their size. Taken from [26].

4.2 Propagation Simulation with CRPropa3

Once cosmic rays leave the acceleration site and propagate through the universe towards Earth they interact with the cosmic photon background, are deflected in magnetic fields according to their charge and cosmological effects appear on large propagation scales. *CRPropa3* is a public available state of the art simulation tool which considers all known significant effects to model the effects on cosmic rays during propagation. With this framework astrophysical predictions for UHECRs inside and outside our galaxy can be studied using high-performance computing [27]. In order to introduce all significant effects this section will give a brief overview about the knowledge of cosmic ray propagation and shows simulation done with *CRPropa3* according to the discussed effects.

4.2.1 Cosmic Photon Background

Even the largest voids of our universe are filled with photons at different wavelength, which can interact with cosmic rays. Fig. 4.2 shows models of different wavelength bands from the radio photons up to UV-light. The cosmic radio background (CRB)

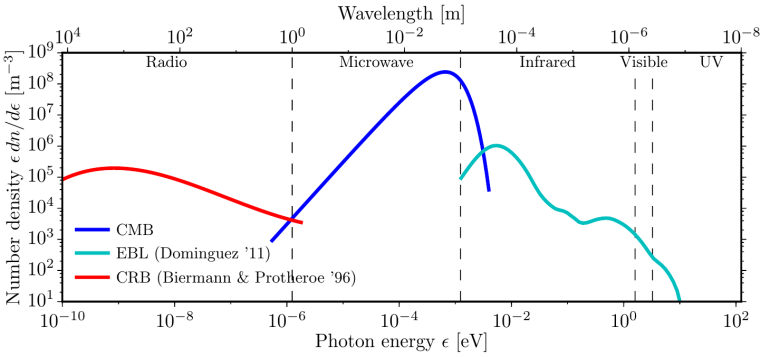


Figure 4.2: Models representing the measured Energy spectrum of the cosmic photon background: The cosmic radio background(CRB), the cosmic microwave background (CMB) and the extra galactic background light (EBL). taken from [10]

originates from synchroton radiation of relativistic electrons produced in radio loud galaxies but also in normal galaxies. Models which are based on measurements predict a photon density of greater than $n_{\mathrm{CRB}} = 10^5 \mathrm{m}^{-3}$ [10].
The dominant radiation field is the cosmic microwave background at a temperature of $T_0 = 2.7255\,K$ and a current energy density of $\epsilon_{\mathrm{CMB}} = 0.2606\,\mathrm{MeV\,m^{-3}}$. The number density of CMB photons in the universe is large: $n_{\mathrm{CMB}} = 4.107 \times 10^8 \mathrm{mm}^{-3}$. In our standard model of cosmology the CMB is a diffuse radiation originating from the epoch of recombination at a redshift of $z \approx 1090$, which corresponds to 0.37 Myr after the big bang. At this time free electrons and protons bounded and formed hydrogen and the universe became translucent for photons. The CMB radiation is described by a perfect black body spectrum [28].
Outside our Galaxy space is also filled with photons emitted by stars: the extra galactic background light (EBL). The number density of this radiation is much smaller than the CMB but plays still a roll for cosmic rays since it limits the propagation of UHECR [10]. A more detailed description of particular interactions is given in the next section.

4.2.2 Photonuclear Processes

Cosmic rays propagating through the universe interact with the cosmic photon background which is strongly blue-shifted in the frame of the cosmic rays and thus reaches energies where photonuclear processes occur. The dominant effects cosmic rays undergo by interacting with the photon background are electron pair production, pion production and photodisintegration. In *CRPropa3* the interaction length of each process is numerically calculated using models of the cosmic photon background specified

in 4.2.1 [10].
The process with the lowest energy threshold but the highest cross section is electron pair production by hadrons which is described by the Bethe-Heitler process: a photon scatters with a cosmic ray nucleus producing an electron-positron-pair, which carries a fraction of the cosmic ray energy. Another photonuclear process is photodisintegration where a background photon excites the nucleus to an unstable state. The nucleus de-excites again by emitting a high energetic photon and protons, neutrons or α-particles which also carry away a fraction of the primary nucleon energy. For heavy primary particles like iron also photofission becomes relevant [10].
At a threshold energy of $E \approx 145\,\mathrm{MeV}$ pion production via Δ resonance starts. A proton scatters with a photon and produces a Δ^+ baryon which then decays into a proton or a neutron plus a pion [10]:

$$p + \gamma \rightarrow \Delta^+ \rightarrow p + \pi^0$$

$$p + \gamma \rightarrow \Delta^+ \rightarrow n + \pi^+$$

The highest number density of photons in the extra galactic space is represented by the CMB with about 400 photons per cm^3 at an energy of $E_\gamma^{\mathrm{CMB}} = 6.34 \times 10^{-4}\,\mathrm{eV}$ [28]. With a proton mass of 938 MeV and the mass of the Δ barion of 1232 MeV [29] the minimum threshold energy for this process is:

$$E_{\mathrm{min}\Delta} = \frac{m_\Delta^2 - m_p^2}{4E_\gamma^{\mathrm{CMB}}} \approx 2.52 \times 10^{20}\,\mathrm{eV} \tag{4.5}$$

Due to this effect the universe becomes opaque at energies around $10^{20}\,\mathrm{eV}$ since the

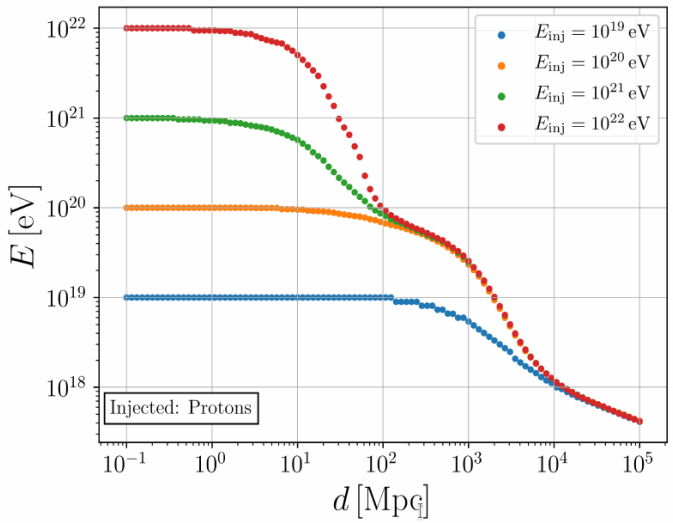


Figure 4.3: The mean energy of protons as a function of traveled distance. At higher energies the energy loss is dominated by pion production. Taken from [30]

Δ-resonance cross section is dominating [4]. This effect apply already at lower energies

due to the long tail of the energy distribution of the CMB. This describes a cut-off for energies we can observe which was first predicted by Greisen, Zatsepin, and Kuzmin and is known as the GZK-limit [31]. This limit is visualized in Fig. 4.3 where protons at different energies have been simulated with *CRPropa3*. The rapid energy loss after already small distances for protons over 10^{20} eV is the Δ-resonance dominating the cross section. However in the presented simulation all known significant interactions have been included as well as cosmological effects.

4.2.3 Deflection in Magnetic Fields

Charged cosmic rays are deflected in magnetic fields which complicates the identification of their sources since we neither know the charges, nor exact magnetic fields inside or outside our galaxy. Particles affected by a magnetic field are deflected with a gyroradius of:

$$r_{\mathrm{g}} = \frac{p}{|q|B_{\perp}} \simeq \frac{E/c}{ZeB_{\perp}} \tag{4.6}$$

where $B_{\perp}$ is the perpendicular component of the magnetic field. Nuclei with higher charge are more effected since the deflection depends on the rigidity $R = E/Z$ of the cosmic ray [10].
To simulate the propagation of cosmic rays through a magnetic field different models are established categorized into two types: the galactic magnetic field (GMF) and the extra galactic magnetic field (EGMF). The most sophisticated GMF model today is the JF12 model [32] which describes a magnetic field following the spiral arm structure of our Galaxy and an additional irregular field which acts as a random component. On the left in Fig. 4.4 the mean deflection of particles is computed as a function of

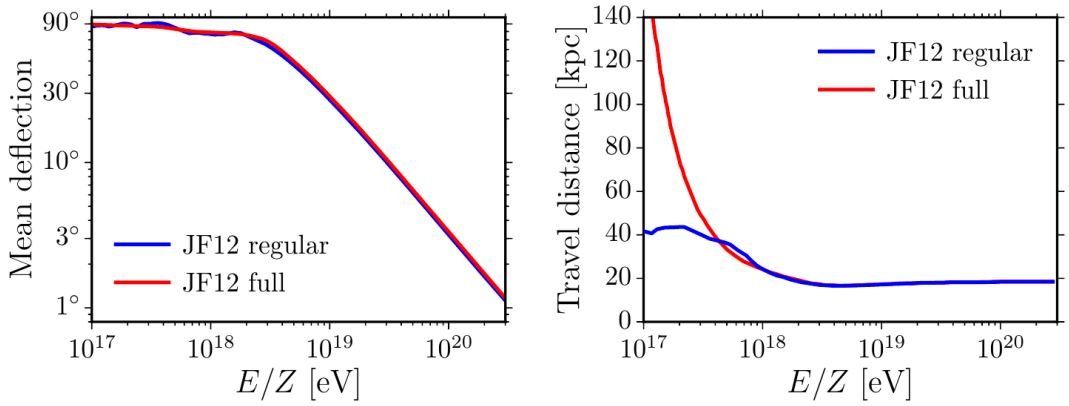


Figure 4.4: *left*: Mean deflection of extra galactic cosmic rays as a function of the rigidity $R = E/Z$ simulated with the JF12 GMF model. *right*: The mean traveled distance inside our Galaxy. In both plots the blue line describ the regular field component following the spiral structure of our galaxy and the red line with an additional random component for a turbulent magnetic field. Both taken from [10].

the rigidity. On the right the mean traveled distance inside our Galaxy as a function

of the rigidity is shown. Cosmic rays at the highest energies travel on straight lines whereas smaller energies are strongly deflected which makes it difficult to reconstruct their origin.
We know less about the EGMF than about the GMF which is modeled using radio measurements. The only way to study EGMF fields is through magnetohydrodynamic simulations which differ in their predictions. EGMF models vary in their prediction [10]: e.g. the deflection expectations of a 40 EeV proton traveling over a distance of 110 Mpc are less than 1° according to *Dolag et al.* [33] or could could be at least 45° according to *Miniati et al.* [34].

Chapter 5

Simulation of Sky Regions with Nearby Source Contribution

A recent Analysis investigating in UHECR sources revealed a dipolar structure and favor an astrophysical scenario with acceleration sites situated outside our galaxy [22]. Furthermore, other studies found an evidence for anisotropy correlation with nearby sources which are located some Mpc away from our Galaxy [2]. In this work we take these results as a starting point and construct a scenario with sky regions exhibiting a nearby UHECR source. This chapter describes the setup of a model universe containing extragalactic UHECR sources, considering cosmological effects and simulate the UHECR propagation through the universe using *CRPropa3* [27].

5.1 Source Properties

5.1.1 Source Density

The universe we are simulating is homogeneous and isotropic containing identical UHECR sources. The source density ρ_s is constant in a comoving volume and all sources have the same absolute luminosity. We set up a sphere with the radius R in which we distribute these sources uniformly. We place an observer in the center, representing our Galaxy. Depending on the source density the mean distance to the closest source can vary. We simulated sets of random universes with ρ_s between $10^{-1}\,\mathrm{Mpc}^{-3}$ and $10^{-7}\,\mathrm{Mpc}^{-3}$ and computed the distance to the closest source D_min. In Fig. 5.1 we show the median and the 68 % quantile of D_min for 1,000 simulation sets of each ρ_s value.
The closest source represents a distinguished direction: by zooming into the closest distances of the model universe the distribution of sources is not homogeneous and isotropic anymore. This difference between the small scale and the large scale is well known for matter distribution in cosmology [28]. Recent results favor a significant contribution of nearby sources within 10 Mpc (i.e. [2]) and therefore we selected a source density of $\rho_s = 10^{-3}\mathrm{Mpc}^{-3}$ which fulfills this requirement(c.f. Fig. 5.1). In Fig. 5.2 the difference between nearby region and larger distances is visualized by

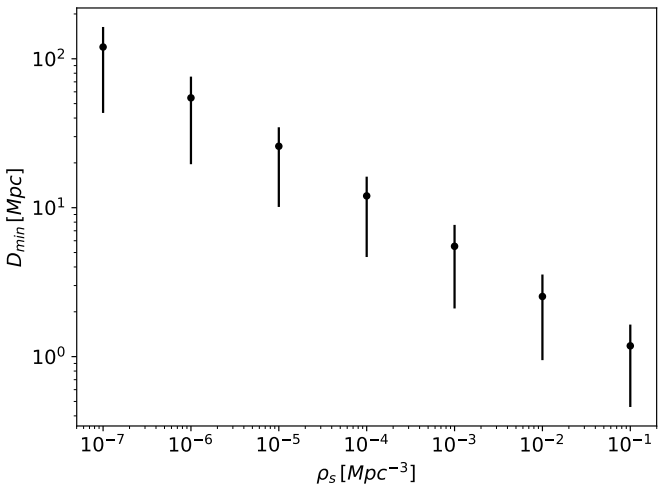


Figure 5.1: We show the distance to the closest source $D_{\min}$ in an isotropic and homogeneous universe. We compute the median and the 68 % quantile of 1000 ralisation for each presented source density ρ_{s}.

projecting the position of uniform distributed sources on the x-y-plane for different scales.

The idea of this approach is that the nearest source has a higher apparent luminosity in comparison to the sources further away and thus forming a significant signal which is the base for a later analysis. The number of sources inside a sphere shell of size ΔR for an isotropic universe is:

$$N_{\mathrm{s}} = \int_{R}^{R+\Delta R} \int_{0}^{\pi} \int_{0}^{2\pi} \rho_{\mathrm{s}} r^2 \sin(\theta) dr\, d\theta\, d\phi \tag{5.1}$$

which than tends for larger distances to

$$N_{\mathrm{s}} \approx \rho_{\mathrm{s}} 4\pi R^2 \Delta R \tag{5.2}$$

This limit holds not for small scales which are dominated by random fluctuations, as visible in Fig. 5.2 with one nearby source at 4 Mpc followed by the next sources further away.

5.1.2 Cosmic Ray Injection

Each simulated point source is considered to inject UHECR isotropic. The acceleration mechanism is motivated on theoretical predictions and experimental constraints. The modeled sources follow a so called *'bottom-up'* acceleration process where standard model particles are accelerated up to a specific energy (c.f. 4.1). We assume each source to be identical following a power law energy spectrum at the source with a maximal rigidity dependent acceleration energy. The maximal energy is described through a cut-off function with the shape of a broken exponential cut-off as shown

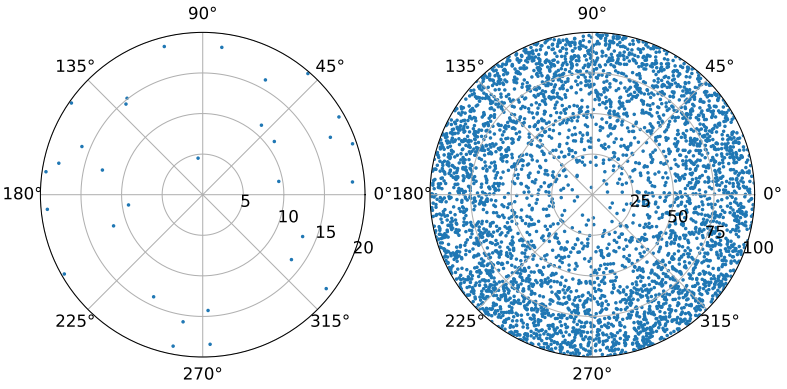


Figure 5.2: Visualization of an isotropic and homogeneous universe which is projected on the x-y-plane. The radial axis denotes the radius r in Mpc and the azimuth angular is the ϕ coordinate. On the *left* sources up to a distance of $r = 20\,\mathrm{Mpc}$ are displayed, whereas on the *right* sources till $r = 100\,\mathrm{Mpc}$ are shown.

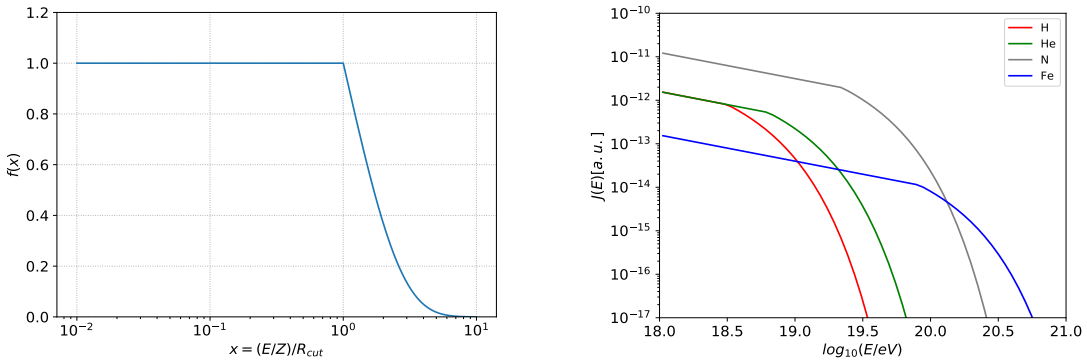


Figure 5.3: *left*: the cut-off function describing the maximal injected energy. *right*: the injected spectrum following 5.3 using parameters specified in 5.1.2.

in Fig. 3.3. The sources inject 4 representative elements following approximately logarithmic mass groups ^{1}H, ^{4}He, ^{14}N and ^{56}Fe. The spectrum has the form [3, 10, 24]:

$$J(E_{\mathrm{inj}}, Z_{\mathrm{inj}}) = \phi_0 \left(\frac{E_{\mathrm{inj}}}{\mathrm{EeV}}\right)^{-\gamma} a(Z^j_{\mathrm{inj}}) \cdot \left(\Theta\left(1 - \frac{E^i_{\mathrm{inj}}/Z^j_{\mathrm{inj}}}{R_{\mathrm{cut}}}\right) + \Theta\left(\frac{E^i_{\mathrm{inj}}/Z^j_{\mathrm{inj}}}{R_{\mathrm{cut}}} - 1\right) \cdot e^{1 - \frac{E^i_{\mathrm{inj}}/Z^j_{\mathrm{inj}}}{R_{\mathrm{cut}}}}\right) \tag{5.3}$$

where ϕ_0 is the total flux level of the cosmic rays, E_{inj} and Z_{inj} are the injected energy and injected elements, respectively and R_{cut} is the maximal rigidity. The exact values we use to model the source energy spectrum are orientated at the best fit values resulting from the baseline scenario of the combined fit method in [3, 10, 24]. We take the global minimum found in the fitting procedure in [10]. Furthermore, we

use a slightly lighter composition.

$$\gamma = 0.6$$
$$log_{10}(R_{\mathrm{cut}}/\mathrm{eV}) = 18.5$$
$$f_{\mathrm{H}} = 10\,\%,\ f_{\mathrm{He}} = 10\,\%,\ f_{\mathrm{N}} = 79\,\%,\ f_{\mathrm{Fe}} = 1\,\%$$

The injected spectrum is shown on the right in Fig. 5.3. At lower energies nitrogen dominates the spectrum whereas due to the rigidity cut-off at higher energies iron dominates. This is also compatible with the X_{max} measurements on Earth which favors heavier elements towards higher energies [21].

5.2 Propagation Horizon

Due to attenuation effects during the propagation cosmic rays lose energy and due to the GZK-cutoff higher energies are not able to reach the Earth from far away [4]. Using *CRPropa3* we inject particles at different distances to investigate the maximal propagation distance for different energies. We choose protons and iron at energies of 10^{23} eV and compute the distance from which we can receive energies exceeding 10^{18} eV , 10^{19} eV and 10^{21} eV. Fig. 5.4 shows the fraction of particles exceeding these

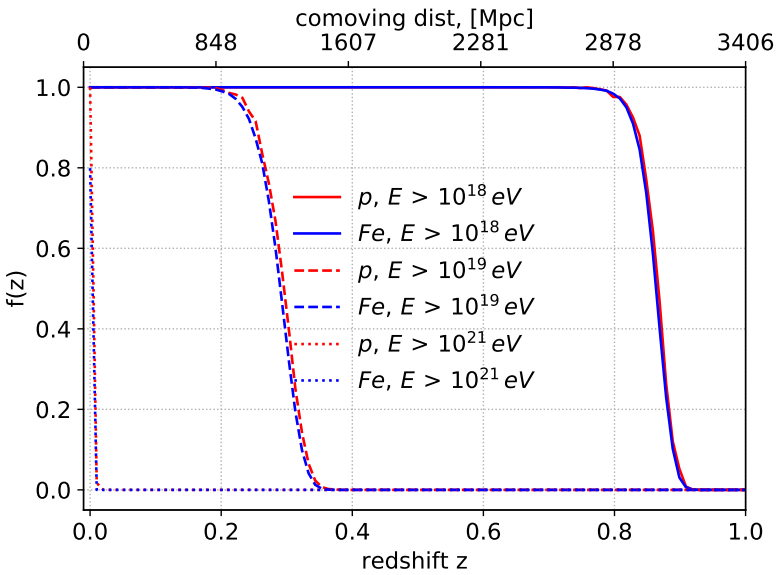


Figure 5.4: Fraction of particles propagating over distance, exceeding a certain energy, from initial ^{1}H and ^{56}Fe.

energies defining a propagation horizon. Particles with an energy above 10^{21} eV loose energy fast due to the GZK-effect (see section 4.2.2) which is described by a very low propagation Horizon. The propagation horizon for particles with an energy above 10^{18} eV exceeds $3,000$ Mpc. We determine our minimum energy to 10^{19} eV which can originate from sources at a maximal redshift of $z \approx 0.35$. This choice of a lower energy

limit is motivated by recent anisotropy studies which consider minimum cosmic ray energy in this order of magnitude [2, 22].

5.3 Cosmological Effects

We setup our model universe in a cosmological context where effects due to the expansion of the universe are contributing. We assume a flat ΛCDM universe with a Hubble constant of $H_0 \approx 67.3\,\mathrm{km/s/Mpc}$ the present baryonic and dark matter density of $\Omega_M \approx 0.315$ and the dark energy density of $\Omega_\Lambda \approx 0.685$. This is the same cosmological model used for the simulation tool *CRPropa3* [27] and other combined fit method [10]. The space time of our universe is described by the Robertson Walker metric, where we assume flat space:

$$ds^2 = -c^2dt^2 + a(t)^2[dr^2 + r^2d\Omega^2] \tag{5.4}$$

Here t is called the cosmological proper time which is measured by an observer who sees the universe expanding around him. The spacial coordinates r, ϕ and θ are the comoving coordinates which describe the position of a point in space. The scale factor $a(t)$ describes how a distance expands or contracts with time. Today which is denoted as t_0 the scale factor is set to 1. This means that the comoving distance is coinciding with the proper distance at the present time [28].
By observing light[1] we get information from a source that emitted a photon at a certain cosmological time t_{em}, whereas we observe the photon at a later cosmological time t_{rec}. Since the Universe expands the proper distance to the source has changed and the source was closer at the moment the photon was injected. This observation effect is measured in redshift which is defined as:

$$1 + z = \frac{a(t_{\mathrm{rec}})}{a(t_{\mathrm{em}})} \tag{5.5}$$

The comoving distance describes the distance this source has at the time the photon is observed, like we would measure the distance with a ruler. In fact the comoving distance factors out the expansion of the universe since it gives us a distance which does not change in time [35].
Cosmic rays propagate over a certain distance and interact with the photon background, therefore it is important to know the distance the cosmic ray has actually traveled. The look-back distance, often referred as the light-travel distance, is the time a photon took to reach the observer multiplied by the speed of light. This distance is depending on the model which describes how fast the universe is expanding and leads to the look back distance defined as:

$$d_L(z) = \frac{c}{H_0}\int_0^z \frac{dz'}{(1+z')\sqrt{\Omega_{\mathrm{M}}(1+z')^3 + \Omega_\Lambda}} \tag{5.6}$$

[1] In our analysis we are considering cosmic rays which propagate with approximately the speed of light.

where H_0 is the Hubble constant today, Ω_{M} the mass density inside our universe and Ω_Λ the density of dark energy [2] [35]. This effect is implemented in the simulation software *CRPropa3* which also consider adiabatic energy loss of the cosmic rays due to expansion [10].
Another important aspect is that the expansion of the universe changes also the apparent luminosity of the source. We consider a proper surface: $A_p = 4\pi r^2$. A photon emitted at a sources carries its own wavelength as a proper distance which is $c(\delta t_{\mathrm{em}})$ at the source. When the photon is detected the proper wavelength is in terms of the metric $c(\delta t_{\mathrm{rec}})$. Due to expansion this proper distance was stretched following the relation:

$$c(\delta t_{\mathrm{rec}}) = c(\delta t_{\mathrm{em}})(1+z) \tag{5.7}$$

The flux is calculated as particles crossing a surface per time and therefore we have to consider the stretched surface to calculate the flux of a distant source:

$$f = \frac{L}{4\pi r^2 (1+z)^2} \tag{5.8}$$

where L is the absolute luminosity of a source, and the luminosity distance is defined as: $d_L = r(1+z)$. This aspect is crucial in cosmology as on large scales the flux

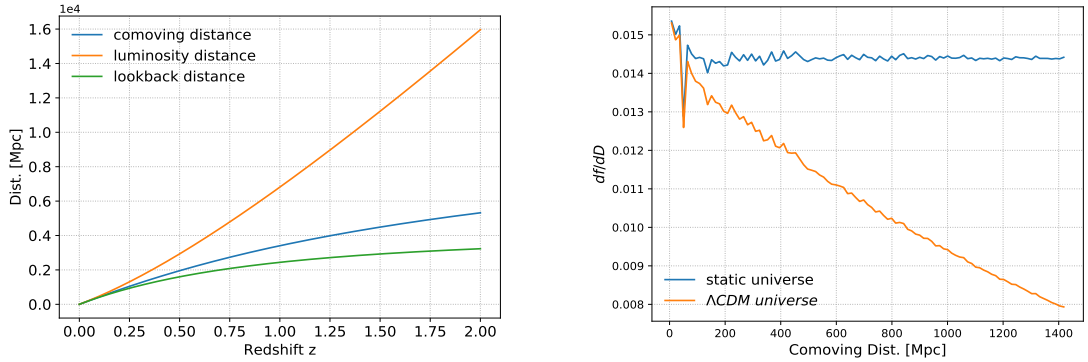


Figure 5.5: *left*: The comoving, luminosity and look-back distance is shown as function of the redshift. *right*: The flux contribution from different distances originating from a homogeneous source distribution in a static universe and in an expanding universe.

rapidly drops and the distance to a standard candle would be overestimated by the factor(1 + z) [28].
The behavior of the three different cosmological distances we introduced here are shown on the left in Fig. 5.5. On the right the flux contribution per distance interval is computed for the case of a static universe and for the case of an expanding universe described by the ΛCDM model. In the continuous limit the distribution of the source contribution originating from different distances is described by a uniform distribution $\mathcal{U}(d)$ whereas for the ΛCDM model is described by $\mathcal{U}(d)/(1+z)^2$. This effect is considered in the present work to describe the relation between flux and distance in a realistic manner.

[2] The radiation density and the curvature was neglected since we assume a flat ΛCDM universe.

5.4 One-Dimensional Model of Sky Region

Considering all effects and constraints discussed above we are able to setup a Monte Carlo simulation of sky region with nearby sources. We construct a universe containing identical sources with a source density of $\rho_s = 10^{-3}\text{Mpc}^{-3}$. We place our galaxy in the center as an observer and consider sources up to a redshift of $z = 0.35$ which corresponds to a comoving distance of $d_c = 1425\,\text{Mpc}$. We select 3 sky regions around the direction of the 3 closest sources with an radius of $\alpha = 0.3\,\text{rad}$. The radius is orientated by recent works investigating nearby sources using a search radius to mimic deflections of cosmic rays in magnetic fields [2]. With the setup of a search

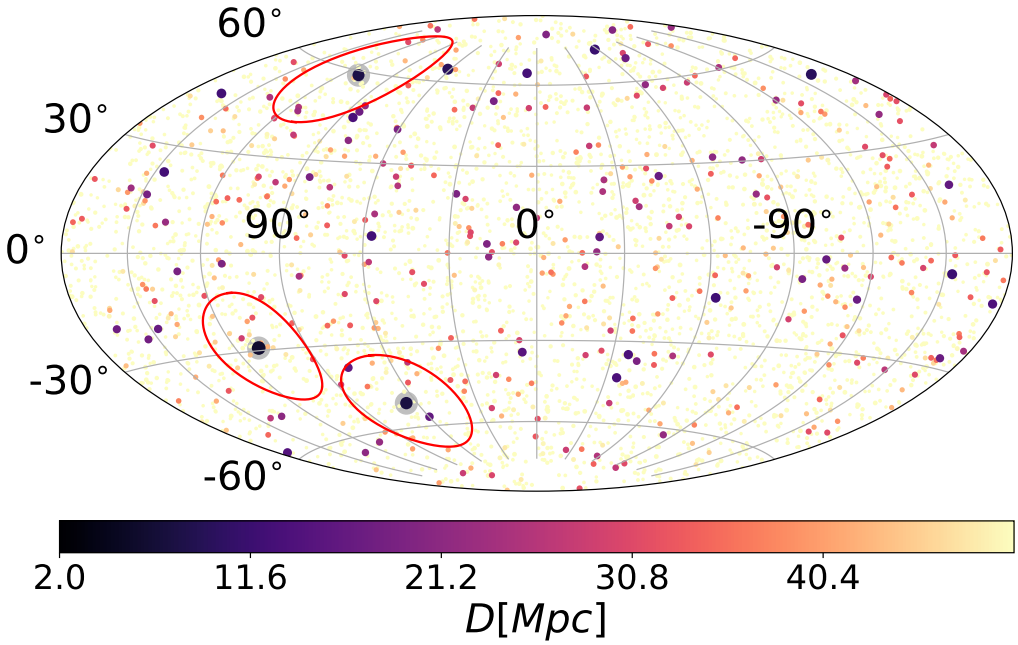


Figure 5.6: Visualization of the source positions with their distance represented in the circle size and color code. The three closest sources are highlighted by gray circles and the ROIs are displayed with red circles.

radius around a nearby source we would also record the flux from contributors further away but a significant fraction would originate from the nearby source. One scenario is visualized in Fig. 5.6 where the nearest sources are highlighted and the region of interest (ROI) are marked by red circles.

To study characteristics of the nearby source in the on-target sets we construct sets without a nearby source for comparison. We set up a off-target set which is the part of the sky outside the ROIs. Since we selected already the 3 closest sources in the simulated universe the closest source in the off-target remains further away. This is shown in Fig. 5.7 where we counted the closest source for all 3 ROI and outside the ROI for $1,000$ scenarios. It is visible that the mean distance of the ROI with the closest source is a few mega parsec closer than the nearest source outside the ROI.

To simulate the cosmic ray flux of all sources on Earth, we have to calculate the

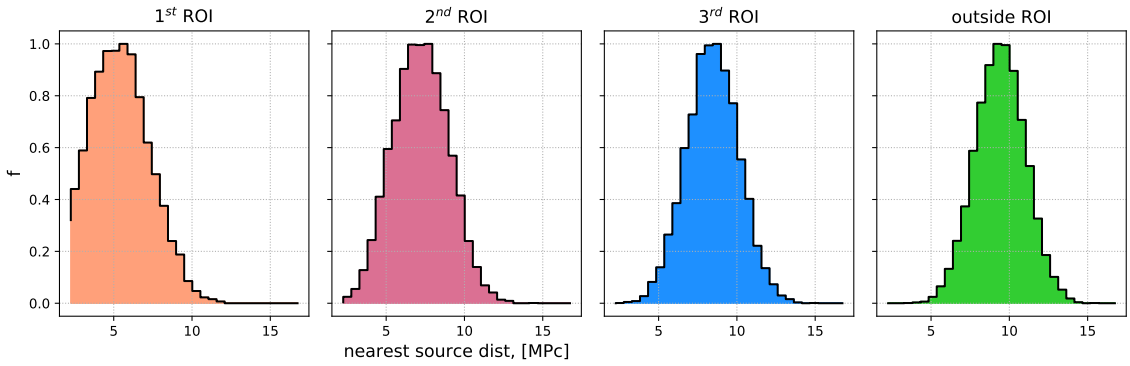


Figure 5.7: Distance of the nearest Source in the three ROIs and outside the ROIs for 1,000 Monte Carlo sets.

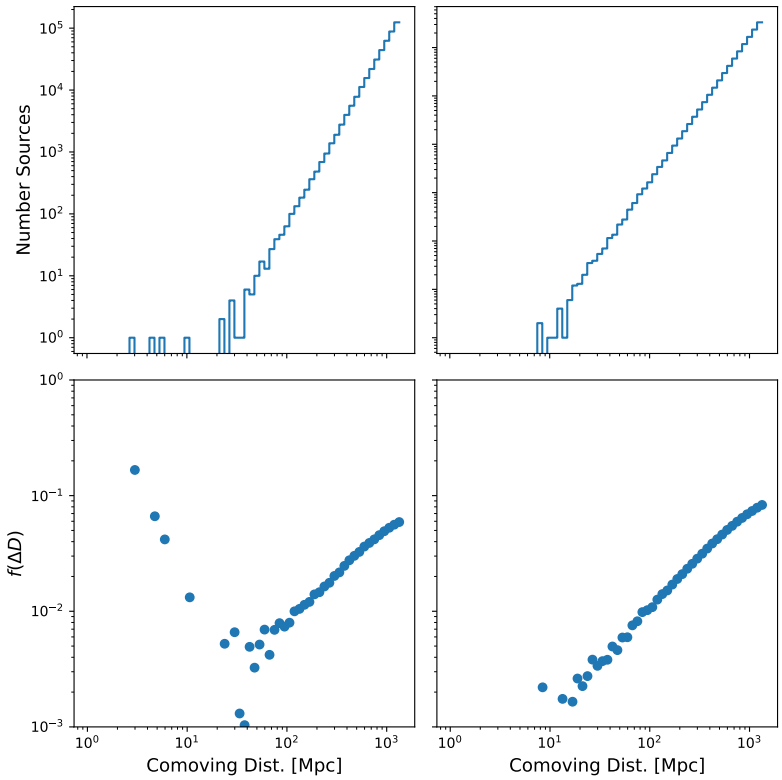
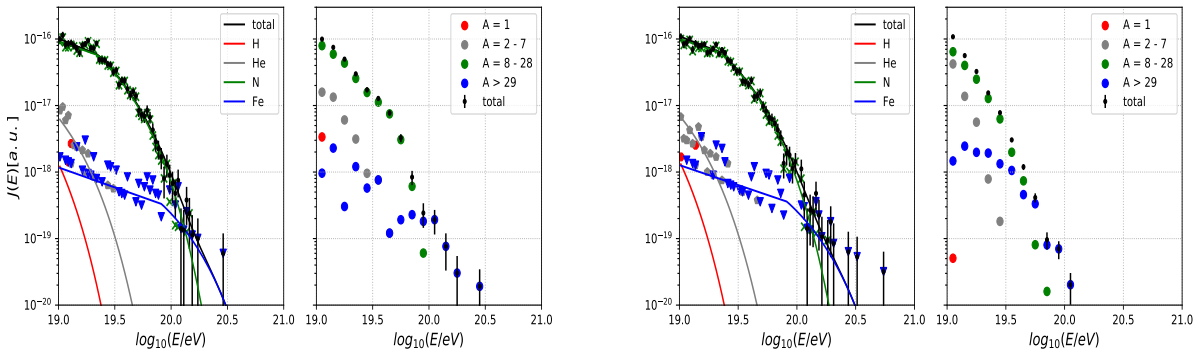


Figure 5.8: *top* histogram of source distance using a logarithmic binning. *bottom* the flux contribution of each distance bin normed to 1. On the *left* row we show a on-target set, whereas on the *right* a off-target set.

relative flux which each source is contributing. Therefore, all sources inside the selected ROI contribute with their individual fraction to the analyzed set of cosmic rays, where only the distance information is left. We choose a logarithmic binning for the source distances and calculate the flux for each distance according to the luminosity distance as explained in 5.3. Each calculated flux is multiplied by the

number of counted source inside the according distance bin. The number of sources filled into distance bins and the contributing flux of each distance bin is shown for one example scenario in Fig. 5.8. Since the flux decreases according to $1/r^2$ nearby sources have a larger contribution than sources from further away. Note that the increasing flux for larger distances is due to the increasing bin size, which means that at higher distances more sources are counted in one bin. The contribution in Fig. 5.8 are already normed to 1 to calculate the relative contribution from each distance. In the displayed example the first 3 sources are at 2.99, 4.74 and 5.97 Mpc contributing with a combined source fraction of $\approx 27\,\%$.

The next step is to inject a spectrum described in 5.3 with the parameters 5.1.2



(a) signal set (b) null-hypothesis

Figure 5.9: We present the injected spectrum on the *left* and the simulated spectrum on Earth on the *right* for the on-target set and 5.9b the off-target set. For both sets we injected $2,000$ particles.

according to the flux for each distance bin. In total we inject $2,000$ particles for each set where the on-target set is described by the sum of the 3 ROI and the off-target set corresponds to the region outside the ROI. We simulate $1,000$ on-target sets and $1,000$ off-target sets.

We simulate the propagation with *CRPropa3*, described in detail in section 4.2. The

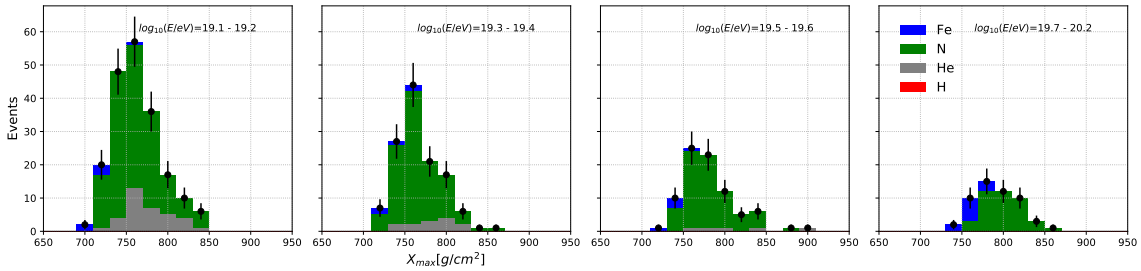


Figure 5.10: $X_{\max}$ distribution for different energy bins, drawn from the Gumbel distribution (see equation 2.4) fitted to the interaction model EPOS-LHC (see section 2.1). The $X_{\max}$ observable is presented additional for different mass groups.

Simulation returns the cosmic ray energy and exact composition which reaches the

earth. In Fig. 5.9 we present the injected and the measured energy spectrum for one on-target set and one off-target set.
To mimic measurements of the atmospheric depth at the Pierre Auger Observatory, we draw $X_{\max}$ values from the Gumbel distribution (see equation 2.4). The $X_{\max}$ observable for different energy bins is presented in Fig. 5.10 for one on-target set.
Finally, we simulated scenarios of Sky regions with and without a nearby sources. The resulting simulations are saved as an energy spectrum and $X_{\max}$ measurement like the observable of the Pierre Auger Observatory. These simulation sets will be investigated by an analysis explained in the next chapter.

Chapter 6

Analysis Method

In the previous chapter we described the construction of a Monte Carlo universe with a homogeneous distributed UHECR source population generating an energy and atmospheric depth measurement on Earth. In this chapter we describe a simple astrophysical model consisting of a homogeneous source background and a nearby point source. We then introduce a statistical evaluation method resolving a nearby source by their typical energy and shower depth signature.

6.1 Fit Model

We model the cosmic ray flux with *CRPropa3* on Earth assuming a simple astrophysical scenario: We simulate homogeneous distributed UHECR sources in one dimension between a minimum distance of $d_{\mathrm{min}} = 2\,Mpc$ and a maximum distance d_{max} corresponding to the redshift $z_{\mathrm{max}} = 0.35$. In the case of homogeneous distributed sources, the distance where cosmic rays are injected is drawn from a Uniform distribution multiplied by $(1+z)^{-2}$ to compensate expansion effects, as described in 5.3. In addition we simulate a point source at the distance d_{p} between 1 and 100 Mpc which describes a nearby source. The flux on Earth is measured in counts of cosmic rays and therefore we can simulate the two source scenarios separately and add the fluxes afterwords.

6.1.1 Source Model

The source model is described by the same acceleration mechanism like it was used to create the Monte Carlo truth in 5.1.2 and in the fit model of [3, 10, 24]: We assume UHECR accelerators described by power law energy spectrum, injecting 4 representative elements ^{1}H, ^{4}He, ^{14}N and ^{56}Fe with a maximal rigidity dependent acceleration energy. In the following we will describe a fast simulation technique to model a cosmic ray flux prediction on Earth for different source parameters.

6.1.2 Cosmic Ray Flux Simulation

The propagation through the universe is simulated using *CRPropa3*. The cosmic ray flux is modified during propagation in a non trivial way (see section 4.2) and the simulations with high statistics are computational expensive. Therefore we only simulate the cosmic ray flux once with a flat spectrum and make use of a re-weighting technique to modify the cosmic ray flux.
We inject a spectrum following $J(E_{\mathrm{inj}}) \propto E_{\mathrm{inj}}^{-1}$ with a minimum energy of $10^{19}\,\mathrm{eV}$ and a maximal energy of $Z_0 \times 10^{21.5}\,\mathrm{eV}$. We simulate this spectrum for a homogeneous source distribution and for a point source at different distances. For each simulation we inject the four representative elements introduced above with $4,000,000$ particles, each, to guarantee a statistical stability. The point source is simulated in 40 logarithmic steps between 1Mpc and 100 Mpc.
The propagation simulation with *CRPropa3* returns the flux and composition on Earth. To mimic the measurements from the Pierre Auger Observatory we sample the X_{max} observable from the Gumbel distribution assuming the EPOS-LHC interaction model (see section 2.1). To have an accessible simulation we bin the simulated cosmic rays before and after the simulation: At the source we use a binning with a bin width of $\log_{10}(E_{\mathrm{inj}}/\mathrm{eV}) = 0.025$ to avoid binning effects after re-weighting. On Earth we use an energy bin width of $\log_{10}(E_{\mathrm{det}}/\mathrm{eV}) = 0.1$ and $20\,\mathrm{g/cm^2}$ for the X_{max} observable which is similar to the detector resolution [10]. For the simulated point source we add one dimension to the histogram for the different distances. We finally construct the histogram for the homogeneous distributed sources as:

$$N_{\mathrm{H}} = N_{\mathrm{H}}(Z_{\mathrm{inj}}^i, E_{\mathrm{inj}}^i, E_{\mathrm{det}}^k, X_{\mathrm{max}}^l) \tag{6.1}$$

and for the point sources as:

$$N_{\mathrm{P}} = N_{\mathrm{P}}(\mathrm{d}^m, Z_{\mathrm{inj}}^i, E_{\mathrm{inj}}^i, E_{\mathrm{det}}^k, X_{\mathrm{max}}^l) \tag{6.2}$$

The upper index denotes the corresponding bin. In the next step we introduce a re-weighting technique to model the injected spectrum and compute the predicted cosmic ray flux on Earth.

6.1.3 Event Weighting

The model we are creating has two components: The flux injected by homogeneous distributed UHECR sources and a single point source. The predicted number of events on Earth depends on the source parameters $\gamma, R_{\mathrm{cut}}, a(Z_{\mathrm{inj}})$ and the distance d_{P} in case of the point source. The re-weighting technique is orientated at the fast simulation procedure used in [10, 30]. We can write the predicted number of cosmic rays events using the simulated histograms 6.1 for the homogeneous source distribution:

$$M_{\mathrm{H}}(E_{\mathrm{det}}^k, X_{\mathrm{max}}^l) = \sum_i \sum_j w(E_{\mathrm{inj}}^i, Z_{\mathrm{inj}}^j)\, N_{\mathrm{H}}(Z_{\mathrm{inj}}^i, E_{\mathrm{inj}}^j, E_{\mathrm{det}}^k, X_{\mathrm{max}}^l) \tag{6.3}$$

and using 6.2 for the point source at a distance d_{P}:

$$M_{\mathrm{P}}(E^k_{\mathrm{det}}, X^l_{\mathrm{max}}) = \sum_i \sum_j \sum_m w(E^i_{\mathrm{inj}}, Z^j_{\mathrm{inj}})\, f_{\mathrm{dist}}(d_{\mathrm{P}})\, N_{\mathrm{P}}(d^m, Z^i_{\mathrm{inj}}, E^j_{\mathrm{inj}}, E^k_{\mathrm{det}}, X^l_{\mathrm{max}}), \tag{6.4}$$

where

$$w(E^i_{\mathrm{inj}}, Z^j_{\mathrm{inj}}) = \left(\frac{E^i_{\mathrm{inj}}}{\mathrm{EeV}}\right)^{-\gamma+1} a(Z^j_{\mathrm{inj}}) \cdot \left(\Theta\Big(1 - \frac{E^i_{\mathrm{inj}}/Z^j_{\mathrm{inj}}}{R_{\mathrm{cut}}}\Big) + \Theta\Big(\frac{E^i_{\mathrm{inj}}/Z^j_{\mathrm{inj}}}{R_{\mathrm{cut}}} - 1\Big) \cdot e^{1 - \frac{E^i_{\mathrm{inj}}/Z^j_{\mathrm{inj}}}{R_{\mathrm{cut}}}}\right), \tag{6.5}$$

and

$$f_{\mathrm{dist}}(d_{\mathrm{P}}) = \frac{d_+ - d_{\mathrm{P}}}{d_+ - d_-}\, \delta_{d^m, d_-} + \frac{d_{\mathrm{P}} - d_-}{d_+ - d_-}\, \delta_{d^m, d_+} \tag{6.6}$$

d_{P} denotes the position of the point source, δ is the Kronecka Delta, d_+ is the next greater and d_- the next smaller simulated discrete distance in the histogram N_{P}. Note that we added 1 to the spectral index of the re-weight function to account for the index of the pre-simulated energy spectrum.
The two histograms M_{H} and M_{P} are re-weighted individually and combined afterwards. According to the distance of the nearby sources in the Monte Carlo universe the total flux fraction of the nearby sources varies. This is modeled by introducing the source fraction f_{s} to the superposition:

$$M_{\mathrm{total}}(E^k_{\mathrm{det}}, X^l_{\mathrm{max}}) = (1 - f_{\mathrm{s}}) \cdot M_{\mathrm{H}}(E^k_{\mathrm{det}}, X^l_{\mathrm{max}}) + f_{\mathrm{s}} \cdot M_{\mathrm{P}}(E^k_{\mathrm{det}}, X^l_{\mathrm{max}}) \tag{6.7}$$

This is the predicted cosmic ray flux measured on Earth which is shown exemplary in Fig. 6.1 for different source parameters which are inspired by results of the combined fit method in [10] and an arbitrary chosen nearby source distance and source fraction.

6.2 Statistical Evaluation Method

The aim of the present analysis is to estimate the free parameter of the fit model for the Monte Carlo sets developed in chapter 5. Bayesian statistics provides a frame to evaluate a probability model given a data set, or in our case a Monte Carlo truth, under the assumption of a certain degree of belief. This degree of belief can be based on prior knowledge such as previous experiments or theoretical assumption [36]. In the following we explain how we can evaluate in a Bayesian context the created Monte Carlo truth with the model above.

6.2.1 Bayesian Inference

Bayesian inference is a method to infer a parameter θ assuming a certain model H in terms of a probability. The probability of a certain set of model parameters is conditional to the observation denoted as 'data' and is written as $P(\theta|\mathrm{data}, H)$. This makes a probability statement about θ for a given set of data and is called the

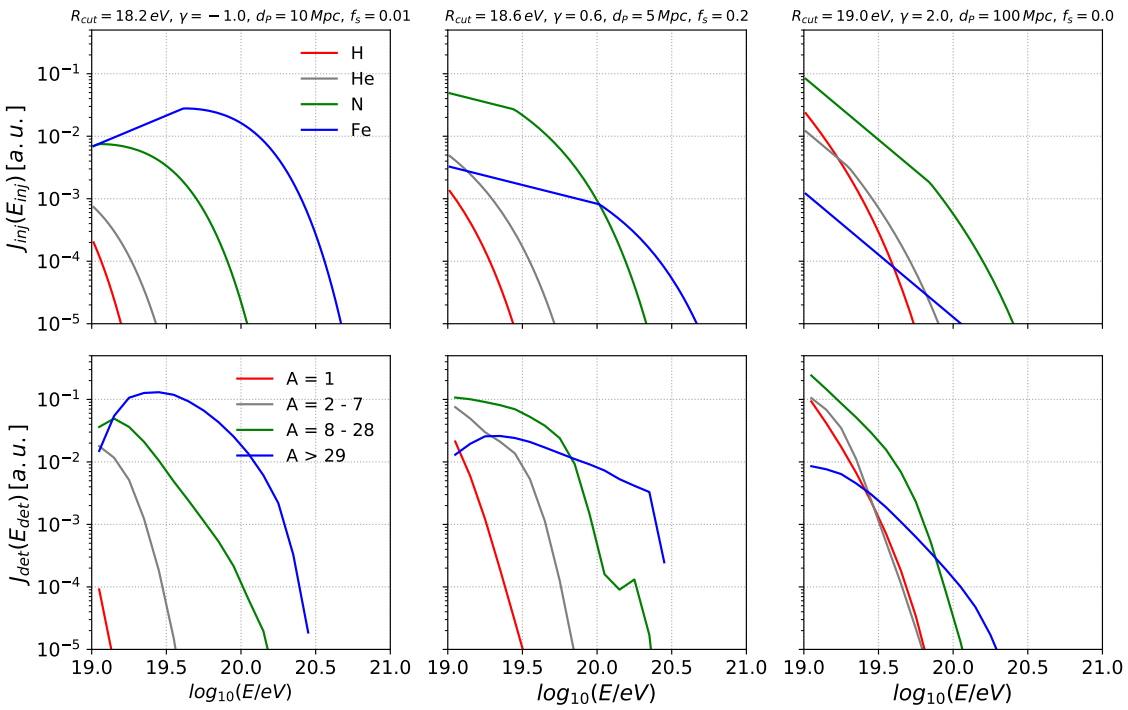


Figure 6.1: We show the injected (*top row*) and the expected energy spectrum on Earth (*bottom row*) for three different source scenarios.

posterior density. Bayes' theorem describes how this *posterior* probability depends on the measured data and the stated model H:

$$P(\theta|\text{data}, H) = \frac{P(\theta) \times P(\text{data}|\theta, H)}{P(\text{data}|H)} \tag{6.8}$$

The *prior* distribution $P(\theta) := \pi(\theta)$ expresses our state of knowledge about the parameter set θ before we have analyzed the data. $P(\text{data}|\theta, H)$ is the conditional probability of observing the data under a particular choice of parameters θ. This is also called the *likelihood* $\mathcal{L}(\theta) =: P(\text{data}|\theta, H)$. In the denominator we have the *Bayesian evidence* $P(\text{data}|H) = \int P(\theta|H)P(\text{data}|\theta, H)d\theta$ which is the distribution of the observed data marginalized over the assumed parameters and is therefore also called the *marginalized likelihood* [37]. This distribution is not depending on θ and can be disregarded for fitting the model parameter to the data. Therefore we can write the *posterior* distribution as:

$$P(\theta|\text{data}, H) \propto \pi(\theta) \times \mathcal{L}(\theta) \tag{6.9}$$

This provides a probability, measured for the specific parameters, in the given light of the data [36].
Bayesian inference is the basis to fit a model to a given set of data providing a probability distribution on the parameters of the model [37]. Once we get the posterior distribution we can conclude on the parameter: The maximum of the posterior

distribution is the most probable value working as a point estimator

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ P(\theta|\text{data}, H) \tag{6.10}$$

This point in the parameter space is not necessary a representative value and more often the posterior mean is taken:

$$\langle\theta\rangle = \int \theta P(\theta|\text{data}, H) d\theta \tag{6.11}$$

which indicates the average value of the estimated parameter θ [10].
In the following we formulate the likelihood function for the cosmic ray energy E and the shower depth X_{max}.

6.2.2 Likelihood

The likelihood function for a combined fit method evaluates both, the measured energy spectrum and X_{max} value.
For the energy measurement the number of events in a specific energy bin is described by a counting experiment. This is described by the Poisson distribution and therefore the likelihood of the spectrum reads:

$$\mathcal{L}_{\text{spec}} = \prod_{\text{e}} \frac{(\lambda_{\text{e}})^{k_{\text{e}}}}{k_{\text{e}}!} e^{-\lambda_{\text{e}}} \tag{6.12}$$

where λ_{e} is the expected number of counts and k_{e} is the number of measured cosmic rays in the respective energy bin e [10].
The events in one energy bin are further specified according to the X_{max} observable. Therefore, $k_{\text{e,x}}$ are the measured cosmic rays in the energy bin e and the X_{max} bin x and $p_{\text{e,x}}$ is the probability to find a cosmic ray in this bin. Thus, the total number in one energy bin $n_{\text{e}} = \sum_{\text{x}} k_{\text{e,x}}$ is also a measurement for the cosmic ray flux. This information is already contained in the likelihood function of the spectra and therefore the number of events in one energy bin is constant for the X_{max} observable. The distribution in X_{max} bins with a fixed total number of events is described by a multinomial distribution and therefore the likelihood according to the X_{max} observable is the product of all multinomial distribution describing an energy bin:

$$\mathcal{L} = \mathcal{L}_{X_{\text{max}}} = \prod_{\text{e}} n_{\text{e}}! \prod_{\text{x}} \frac{1}{k_{\text{e,x}}!} (p_{\text{e,x}})^{k_{\text{e,x}}} \tag{6.13}$$

Since the energy E and the shower depth X_{max} are two independent observable the total likelihood is provided by the product [10]:

$$\mathcal{L} = \mathcal{L}_{\text{spec}} \times \mathcal{L}_{X_{\text{max}}} \tag{6.14}$$

For computational reasons it is often easier to calculate with sums instead of products and therefore it can be an advantage to compute the logarithm if the likelihood

function. Since the logarithm is a monotone function the proportionality sign in equation 6.9 remains valid [38]. In our case the logarithm of the likelihood function is also an advantage since too large numbers can cause a bin-overflow in sampling algorithms. The logarithmic likelihood for the spectrum reads:

$$\log(\mathcal{L}_{\mathrm{Spec}}) = \sum_{\mathrm{e}} k_{\mathrm{e}} \log(\lambda_{\mathrm{e}}) - \log(k_{\mathrm{e}}!) - \lambda_{\mathrm{e}} \tag{6.15}$$

and for the $X_{\max}$ observable:

$$\log(\mathcal{L}_{X_{max}}) = \sum_{\mathrm{e}} k_{\mathrm{e}} - \log(k_{\mathrm{e}}!) + \sum_{x} -\log(k_{\mathrm{e,x}}!) + k_{\mathrm{e,x}} \log(p_{\mathrm{e,x}}) \tag{6.16}$$

and for the combined likelihood:

$$\log(\mathcal{L}_{\mathrm{Spec}+X_{\max}}) = \sum_{\mathrm{e}} k_{\mathrm{e}} \log(\lambda_{\mathrm{e}}) - \lambda_{\mathrm{e}} + \sum_{x} -\log(k_{\mathrm{e,x}}!) + k_{\mathrm{e,x}} \log(p_{\mathrm{e,x}}) \tag{6.17}$$

6.3 Fitting Technique

The likelihood function provides a tool that evaluates a certain model in the light of the Monte Carlo truth. As described in the 6.1 we have 8 free parameters in our fit model and we do not have any a prior knowledge about the behavior of the likelihood. A brute forth approach which scans the parameter space to find the maximum of the likelihood, according to 6.10, would be computational intensive and unfeasible with a reasonable fine parameter step size. At this point Bayesian statistics provides possibilities to explore the parameter space using the likelihood function.

6.3.1 Markov Chain Monte Carlo

In general a Markov Chain is a sequence of possible events. The transition from one event to another is only depending on the probability of the event attained in the previous one [39]. A Markov Chain Monte Carlo (MCMC) is working by random sampling using a Markov chain for the transition probability. If the likelihood of the Markov Chain would be isotropic, thus all the transition probabilities would be equal, the MCMC would describe a uniform random walk. The trick of using the MCMC method is to construct a Markov Chain with transition probabilities following a target distribution which is in our case the posterior distribution. For large number of sample steps the MCMC would then converge to the target distribution [36].
When we sample from a posterior distribution we are only interested in regions of the parameter space with a high probability density.

6.3.2 Metropolis Hastings Algorithm

In 1953 Metropolis developed an algorithm to sample efficiently from the posterior distribution [40] which was later generalized from Hastings in 1970 [41] and is known today as the *Metropolis-Hastings Algorithm*.

The *Metropolis-Hastings Algorithm* is an iterative procedure which accepts or rejects proposed random steps in the parameter space. $P(\theta|\text{data})$ is the posterior distribution we want to sample from and $J_t(\theta^*|\theta^{t-1})$ is the proposal distribution from which we sample each step. We sample a proposal point in the parameter space θ^* at a time t. It is not clear which proposal distribution should be used in general and therefore a normal distribution is used by default. In the first sampling algorithm from 1953 J_t has to be a symmetric distribution at any time which is in the final *Metropolis-Hastings algorithm* not the case any more.
The algorithm is described by the following steps:

1. We start at a point θ^0 for which $P(\theta^0|\text{data}) > 0$
2. We sample a proposal step θ^* from the proposal distribution $J_t(\theta^*|\theta^{t-1})$
3. calculate the ratio
$$r = \frac{P(\theta^*|\text{data})/J_t(\theta^*|\theta^{t-1})}{P(\theta^{t-1}|\text{data})/J_t(\theta^{t-1}|\theta^*)}$$
4. Now we generate a uniform random number $u_t \in [0, 1]$
5. We accept or reject the proposed step following:

   if $u_t < r$ we set $\theta^t = \theta^*$

   otherwise we reject the proposal and set $\theta^t = \theta^{t-1}$
6. return to 2. till the number of sampling steps is reached.

The sequence of iteration $\theta^1, \theta^2, ...$ converges to the target distribution $P(\theta|\text{data})$ as proven in detail in [37]. The number of iterations needed to effectively estimate the target distribution depends on the complexity of the model and its number of parameters [36].

6.3.3 Prior Distribution

The prior distributions comprises the knowledge we have before looking into the data or expresses a degree of belief. Therefore we summarize all information we have into the priors, e.g. we constrain the distance parameter to only positive distances. All our constraints are summarized in Table 6.1. We orientate the allowed range of parameters on the choice made in [10]. In particular we set $\gamma \in [-3, 3]$ that our analysis is able to fit sources following Fermi acceleration or even more exotic acceleration mechanism where the source emissivity increase towards higher energies. The constraints to the R_{cut} parameter is orientated to the simulation range chosen in 6.1.2 where we set the minimum cut-off rigidity to $R_{\text{cut}} = 17.5\,\text{eV}$ and the maximum to $R_{\text{cut}} = 20.5\,\text{eV}$ to be able to model a smooth cot-off shape with the pre-simulated spectrum as described in 6.1.3. We choose the nearby source distance to be between 1 and 100 Mpc which describes the region where we would expect a nearby source and a transition range to distances where the source would be indistinguishable from

Source parameter	Prior distribution
Spectral index	$\pi(\gamma) = \text{Uniform}(-3, 3)$
Maximum rigidity	$\pi(\log_{10}(R_{\text{cut}}/\text{eV})) = \text{Uniform}(17.5, 20.5)$
Nearby source distance	$\pi(d_{\text{P}}/\text{Mpc}) = \text{Uniform}(1, 100)$
Source fraction	$\pi(F_{\text{s}}) = \text{Uniform}(0, 1)$
Elemental fraction	$\pi(a(Z_{\text{inj}})) = \text{Unit simplex}$

Table 6.1: Overview of the parameter for the fit model introduced in 6.1. All constrains are summarized in the chosen prior distribution.

the background. The source fraction f_{s} can vary between 0 and 1 whereas $f_{\text{s}} = 0$ corresponds to the isotropic hypothesis and $f_{\text{s}} = 1$ would be only single point source. Since we do not have further information about the parameters we describe them by a uniform distribution.
The elemental fraction is the only parameter which has an additional constrain. We inject 4 elements each with a certain fraction a_i, $i \in 1, 2, 3, 4$ which has to sum up to one: $\sum_i a_i = 1$. Since we have no further constrains and leave all possible combinations equal, the set of element fraction is described by a unit simplex. We construct a unit simplex by drawing 3 random variables between 0 and 1 and sort them:

$$x_1 \leq x_2 \leq x_3 \tag{6.18}$$

We now select the elemental fraction as follows:

$$\begin{aligned} a_1 &= x_1 \\ a_2 &= x_2 - x_1 \\ a_3 &= x_3 - x_2 \\ a_4 &= 1 - x_3 \end{aligned}$$

This sums up to 1 and provides a uniform distribution on a 4 dimensional simplex as proven in [42].

6.3.4 Sampled Posterior Distribution and Convergence Check

We described how we construct a model of an astrophysical scenario and formulate the likelihood function to evaluate a Monte Carlo truth. Furthermore we introduced the *Metropolis-Hastings algorithm* to sample from the posterior distribution. Before we can fit and evaluate the created Monte Carlo truth, we need to know how many sampling steps are necessary to efficiently estimate the posterior distribution and how we know if the sampling converged.
A convergence check for MCMC sampling is provided by the Gelman-Rubin diagnostic [43]: We sample from the posterior distribution in 10 independent runs, each starting at a random point in the parameter space. The diagnostic combines the variances of each trace and between all traces and formulates a convergence limit expressed in the Gelman-Rubin coefficient $\hat{R}$. A detailed calculation of the coefficient is given in

the appendix A.1.1. The diagnostic indicates a convergence if $\hat{R}$ tends to 1 for each parameter. We require a value of $\hat{R} \leq 1.05$ which represents a convergence within 5 % uncertainty [43, 44].
In this work we use the software package **PyMC**, version 3.6 [45], which provides beside several MCMC algorithms the *Metropolis-Hastings Algorithm*, used in this work. To visualize the sampling procedure we show in Fig. 6.2 the distance as an example parameter of our fit model and the corresponding likelihood value of each sample step. We performed 10 independent runs of sampling and displayed all in

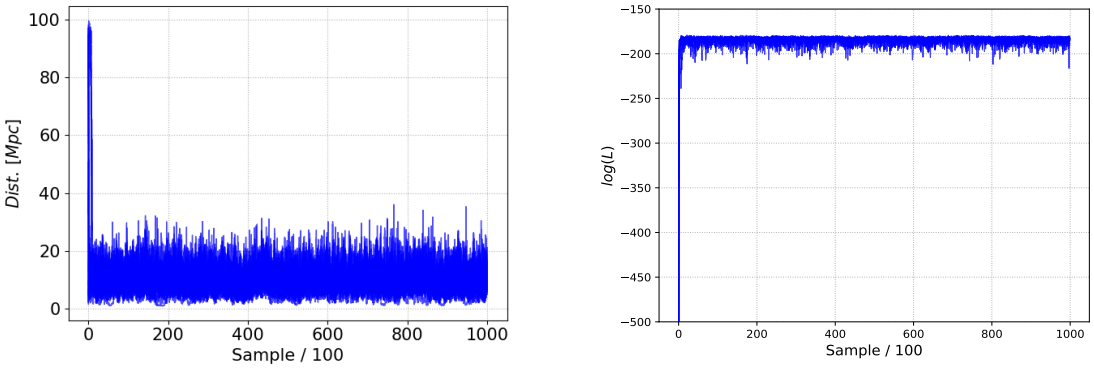


Figure 6.2: *left*: Exemplary the distance parameter of the fit model and *right* the log-likelihood value for ten independent runs. After a burn in phase the Markov Chain stays in a region of high posterior probability.

Fig. 6.2. The sample algorithm started at random points in the parameter space and converges fast into the region with the higher probability where the Markov Chain reached a stationary behavior. The First phase of the sampling procedure is the *burn-in phase* where the Markov chain has not found the region with the highest probability. These samples are discarded afterwards and the rest, which is called the *draw phase* is used to approximate the posterior distribution. The number of steps needed for both phases can vary for different problems and has to be set by hand [44]. For our Model we use 100, 000 samples for the *burn-in phase* to make sure that the Markov Chain reached a stable point and 100, 000 samples to draw from the posterior distribution. This was done in 10 independent chains which are combined for an analysis presented in the next chapter.

Chapter 7

Evaluation of combined Fit in Sky Regions

In this work we constructed simulated universes with homogeneous distributed UHECR sources and selected sky regions with and without nearby UHECR sources. Thus the sky regions contain typical energy and shower depth pattern that we try to resolve. We developed a fit model describing a homogeneous UHECR source distribution with a nearby source to predict the source properties encoded in 7 parameters. Furthermore, we use a statistical fitting procedure based on Markov Chain Monte Carlo method to evaluate the fit model on the constructed Monte Carlo (MC) sets. In this Chapter we will present the results of the combined fit and discuss a physical interpretation.

7.1 Combined Fit

The fit model (in detail described in section 6.1) has 7 free source parameters: the spectral index γ, the cutoff rigidity R_{cut}, the distance of the nearby source d_{P}, its source fraction f_{s} and the 4 elemental fractions $a(Z_0)$, described in 3 fit parameters. We sample from the Posterior distribution $P(\theta|MC)$ using the *Metropolis-Hastings Algorithm* (detailed in section 6.2) exemplary for one specific on-target and one off-target set. The on-target set contains the closest source at 2.99 Mpc with a source fraction of 27 % whereas the off-target set has the closest source at 9.46 Mpc with an according source fraction of 0.4 %. This are representative values for the source distances of our simulated sets, as can be seen in Fig. 5.7
Using the combined fit method we sample $100,000$ steps in 10 individual chains from the posterior distribution in 7 parameters from which we can infer the most probable value and the posterior mean as described in section 6.2. In Table 7.1 we present the fitted parameter values for the on-target and the off-target set. Some of the best fit values differ from the injected parameters specified in equation 5.1.2 and the posterior mean indicates a broad distribution in some parameters.
We already see from the best fit value and the posterior mean, that the injected maximal cutoff rigidity of $\log_{10}(R_{\mathrm{cut}}/\mathrm{eV}) = 18.5$ is found by the fitting procedure. The

	On-target set		Off-target set		MC
Parameter	Best fit	Posterior mean	Best fit	Posterior mean	True value
	$\hat{\theta}$	$\langle\theta\rangle \pm \sigma$	$\hat{\theta}$	$\langle\theta\rangle \pm \sigma$	θ_{MC}
γ	-0.19	0.55 ± 0.69	0.57	0.65 ± 0.18	0.6
$\log_{10}(R_{\mathrm{cut}}/\mathrm{eV})$	18.34	18.47 ± 0.14	18.42	18.45 ± 0.04	18.5
d_{P}[Mpc]	3.1	3.0 ± 1.0	70.8	79.1 ± 19.0	3.0/9.5
f_{s}	0.32	0.54 ± 0.21	0.04	0.04 ± 0.03	0.27/0.004
H[%]	47.3	21.3 ± 17.8	44.2	23.8 ± 17.1	10
He[%]	13.1	23.6 ± 16.3	28.7	30.9 ± 13.7	10
N[%]	39.5	54.2 ± 20.3	26.8	44.5 ± 14.0	79
Fe[%]	0.1	0.8 ± 0.8	0.3	0.6 ± 0.3	1

Table 7.1: Results for the parameters of the fit model for one on-target set and one off-target set: the best fit value $\hat{\theta}$ and the posterior mean $\langle\theta\rangle \pm \sigma$ with the region of the highest posterior density. The posterior distributions are visualized in Fig. 7.3 , 7.5, 7.4 and B.2

spectral index γ which is initially set to 0.6 for the Monte Carlo sets is reconstructed for the off-target set quiet well with $\gamma_{\mathrm{off}} = 0.57$, whereas the fit for the on-target set finds a best fit value at $\gamma_{\mathrm{on}} = -0.19$. Furthermore predicts the fit a significant nearby source for the on-target set whereas the off-target set favors a distant point source with a low signal fraction. In the latter case the point source would be indistinguishable from the homogeneous background. In both Monte Carlo sets the fit fluctuates in the mass composition despite the iron fraction which is in both cases smaller than 1 %.

In Fig. 7.1 we present the observed energy spectrum of the Monte Carlo sets and

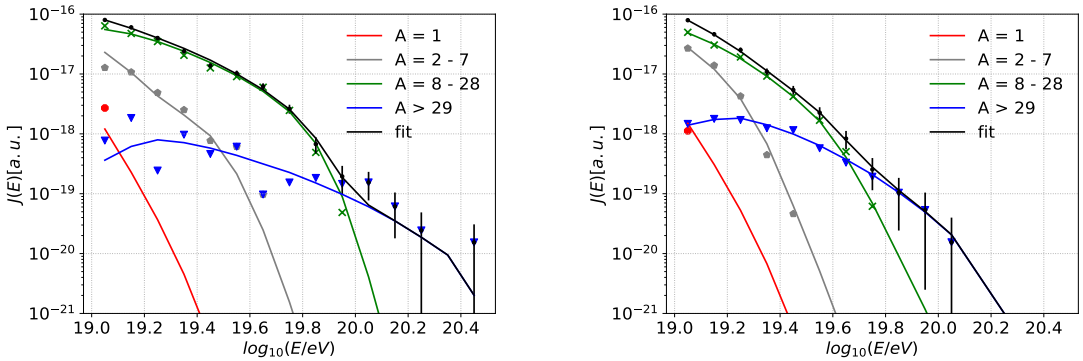


Figure 7.1: The energy spectrum of the Monte Carlo sets with the Poisson uncertainties and the prediction of the fit model for the best fit values which are given in Table 7.1. For better comparison, the model prediction and the Monte Carlo set are further specified in 4 mass groups. On the *left* we show the results for the on-target set and on the *right* for the off-target set.

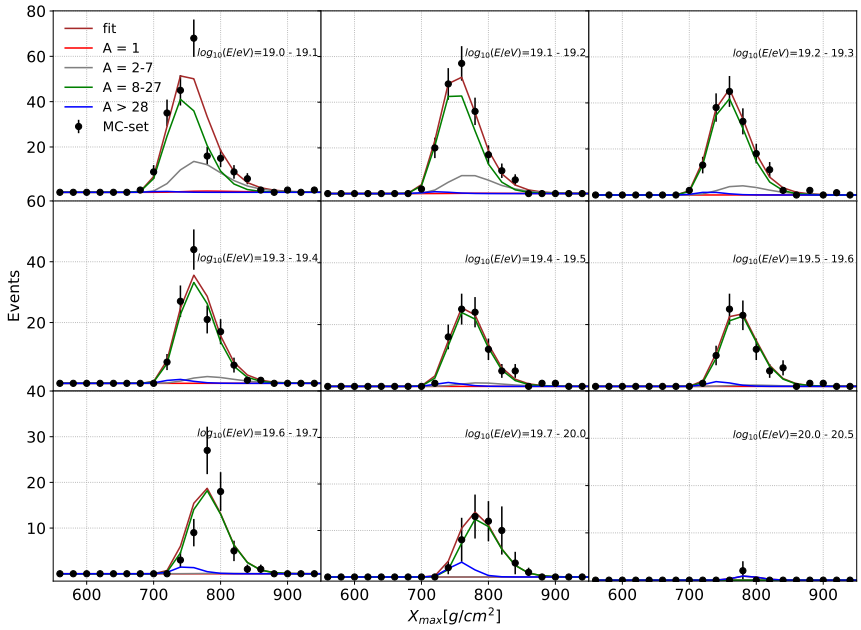


Figure 7.2: The $X_{\max}$ distribution of the on-target set for the indicated energy bins. The fit model prediction for the best fit values from Table 7.1 are shown separated in 4 mass groups.

the predicted flux from the fit model for the best fit value. We also show the predicted $X_{\max}$ observable as presented in Fig. 7.2 for the signal, and the predicted $X_{\max}$ observable for the off-target set can be found in the appendix B.1. The predicted observables match the Monte Carlo sets. In both cases we observe a composition dominated by a mass group between A = 8 and A = 28 up to an energy of $10^{19.9}$ eV for the on-target set and $10^{19.6}$ eV for the off-target set. After this threshold the mass composition is represented by higher masses A > 29 which can only originate from injected iron. Masses of A < 8 are in both sets under abundant which is due to the low rigidity cut at the source and the fact that secondary particles from particle decays are under the energy threshold of 10^{19} eV. We also observe a steeper energy spectrum for the off-target set and higher absolute energies in the on-target sets. This could be an effect of an significant contribution from the nearby source, since on smaller distances cosmic rays lose less energy during propagation.

7.1.1 Posterior Distributions

To gain more information from the posterior distribution we visualize it for the different parameters. Therefore we bin the sampled parameter values in 50 bins with a bin size according to their spread. We use a linear binning adapted to the range of the

parameters γ, R_{cut}, d_{P} and f_{s}, whereas for the elemental fractions $a(Z_0)$ we used a fixed logarithmic binning between 0 and 100 %. We show the marginalized Posterior distributions for γ, R_{cut}, d_{P} and f_{s} in Fig. 7.3 for the case of the on-target set presented above. We additionally present 2 dimensional histograms of two parameters to identify correlations. We compute the true parameter values from the Monte Carlo set and the best fit value $\hat{\theta}$.

In Fig. 7.3 a broad distribution for the γ parameter is visible showing a double-peak

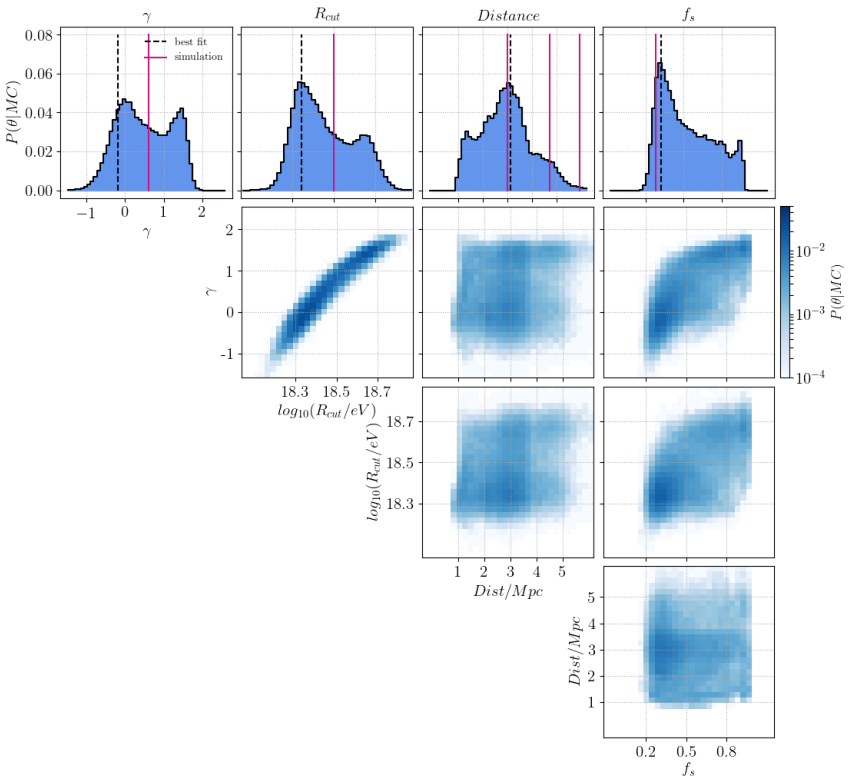


Figure 7.3: The marginalized posterior distribution for the parameters γ, R_{cut}, d_{P} andf_{s} for one on-target set presented in Table 7.1. We show the one dimensional histogram for each parameter in the *top* row and display the best fit value $\hat{\theta}$ and the true parameter value injected in the Monte Carlo set. We also show pair-wise joint distribution to visualize correlations among the parameters.

structure indicating a second region of high posterior probability. This feature is also visible in the one dimensional histogram of R_{cut}. Furthermore, we observe a strong correlation between γ and R_{cut} which was already described by previous combined fit methods as a general feature of this method (see section 3.4). Additionally, we can conclude that the distance parameter shows a high posterior density for nearby sources and the source fraction f_{s} favors a significant contribution. We also observe

a correlation between f_s and γ which can be interpreted that a strong nearby source can compensate a steeper spectrum caused by attenuation effects during propagation.

The marginalized posterior distribution for the off-target set, we presented above,

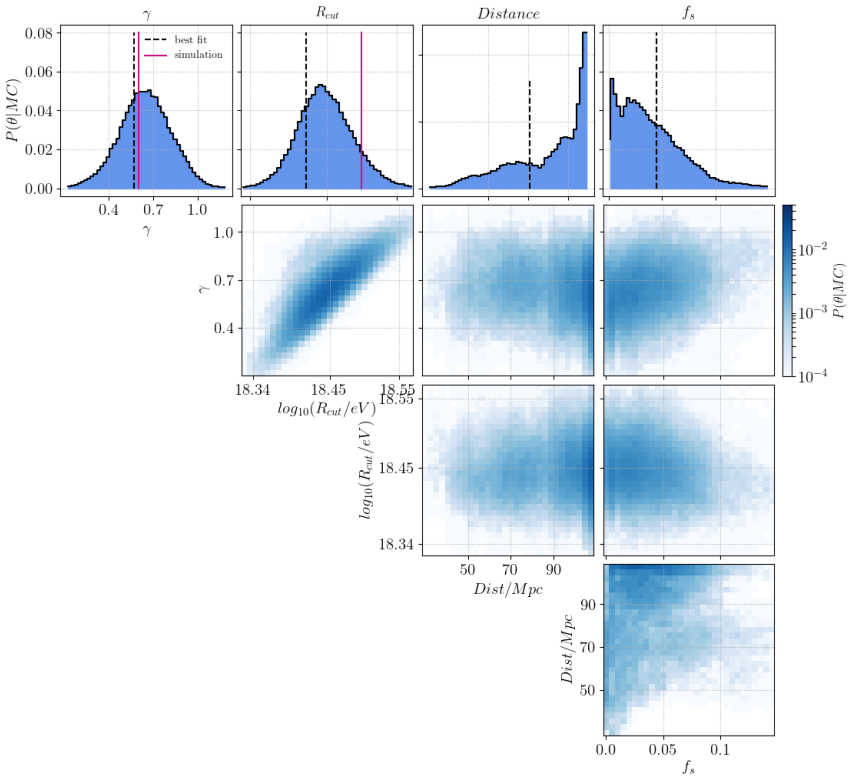


Figure 7.4: The marginalized posterior distribution as described in Fig. 7.3 for the case of the isotropic set.

is shown in Fig. 7.4. In comparison to the presented on-target set we observe in this case a smoother marginalized posterior distribution for the γ and R_{cut} parameters with no double peak structures. But also for this set we see a strong correlation between these two parameters. In contrast to the on-target set the distance parameter in combination with the source fraction clearly disfavors a nearby source with a significant contribution. This is manifested in a low source fraction and a wide spread of sampled distances favoring distances above 50 Mpc which can be neglected according to the homogeneous source distribution. The off-target set is well described by the fit model since the combined fit found the source parameters correctly in contrast to the on-target set where marginalized posterior distribution shows an ambiguous structure. Nevertheless is the gain of this result that the combined fit method developed in this work is capable to find a nearby source correctly.

The marginalized posterior distribution and the according correlation maps for the

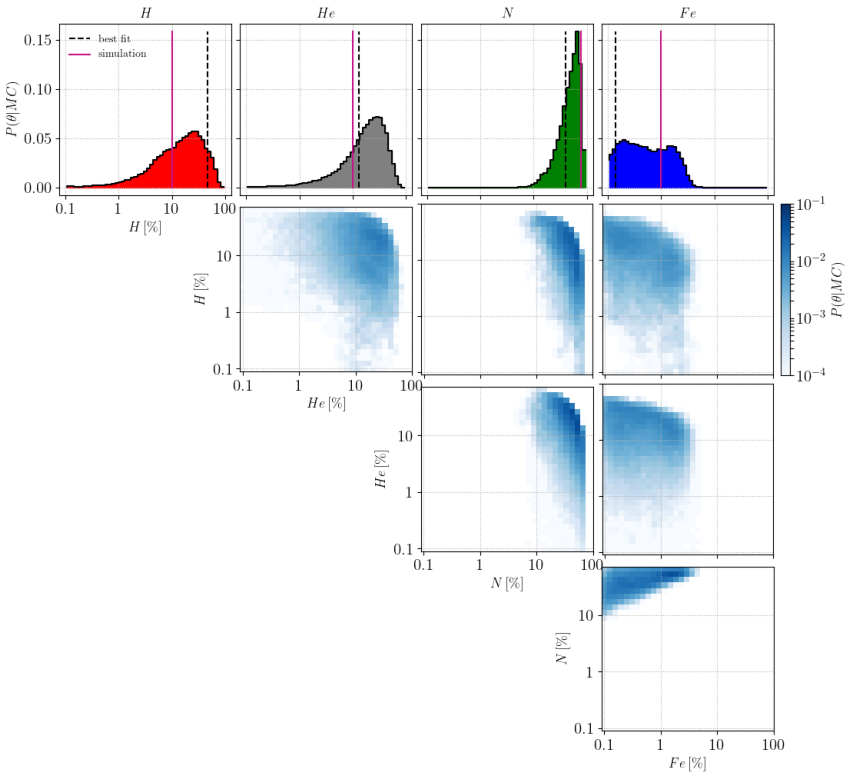


Figure 7.5: We show the marginalized posterior distribution of the on-target set for the mass fraction (for all parameter results see Table 7.1). In the *top* row we show the one dimensional histograms for each element fraction with the best fit value $\hat{\theta}$ and the true value injected in the Monte Carlo simulation. We also show the pair-wise joint distributions to visualize correlations among the different fractions.

mass composition is shown for the on-target set in Fig. 7.5 and for the null hypothesis in the appendix B.1. The marginalized posterior distributions for the elemental fraction is similar for the off-target set and the on-target set. In both cases we observe a broad posterior distribution for ^{1}H, ^{2}He and ^{7}N with the highest fraction of the latter. The fraction of iron at the source is always small, favoring a contribution around 1 %.

In Fig. 7.6 we show the correlation maps for the on-target set between the sampled distance and the elemental fraction to investigate for correlation between these two variables. It turns out that there is no visible correlation also for the off-target set as shown in the appendix B.1. This is quiet unexpected since heavier elements decay during propagation and therefore a correlation is expected. The absence of a visible dependency can be explained for the small sampled distance range for the on-target set where these effects are not significant. In the case of the off-target set the posterior distribution is sampled for a source fraction around 0 which makes the distance parameter redundant and no correlation with the mass composition would be visible.

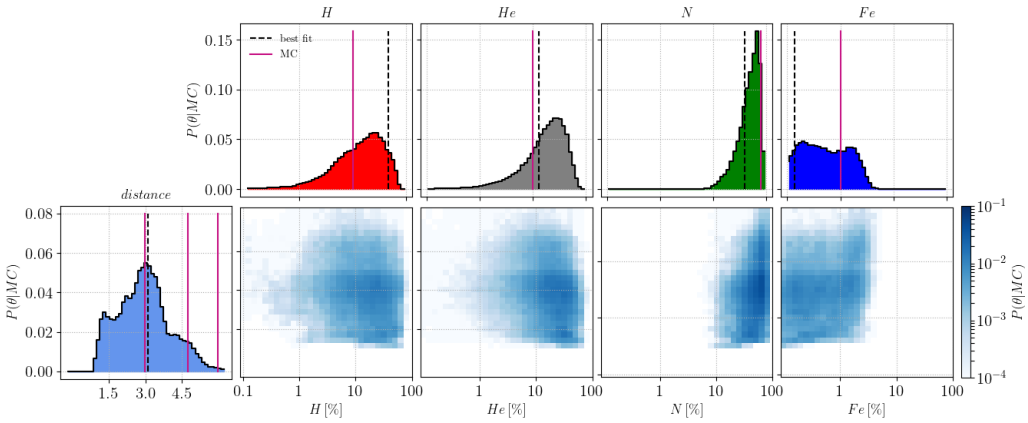


Figure 7.6: We show pair-wise joint distributions between the point source distance d_{P} and the elemental fractions of the on-target set to visualize correlations. Additional marginalized posterior distributions are shown in Fig. 7.3 and 7.5. For the case of the off-target set we show this posterior distribution in B.3.

7.1.2 Likelihood Scan

In the previous paragraph the marginalized posterior distributions revealed already information about correlation between fit parameters. We now want to gain more information about the behavior of the likelihood function by scanning different parameters for the best fit value. With this approach we can also visualize dependencies in the parameter space which are not considered by a Markov Chain Monte Carlo if the posterior density is too low. Therefore we take fixed values for two parameters and fit the remaining 5 free parameter of the fit model. We maximize the likelihood value with the *basin-hopping* approach which finds a global maximum[1] using iteratifly random perturbations of an local maximisation algorithm [46]. This algorithm is implemented in the python package **SciPy** which is an open source software for mathematics, science, and engineering [47].
According to Wilks theorem a nested model with a high number of samples approaches a chi-squared distribution [48]. Thus we can use the $n_\sigma = \sqrt{L_{\mathrm{max}} - L}$ as a pseudo standard deviation to estimate the topology of the two scanned parameters. Here L_{max} is the maximal likelihood value found by the scan and L the best fit value found at each scanning point.

In Fig 7.7 we present the likelihood scan for the γ and R_{cut} parameters for the on-target set, where we observe a global minimum at $\gamma = -0.19$ and $R_{\mathrm{cut}} = 18.34$ as we found with the combined fit for all 7 parameters. Further more we see a valley like

[1]Note that in our case we are searching a maximum in the likelihood space, whereas in the literature, refering to an optimization problem, the term 'minimization' is commom. The likelihood is often computed as a deviance which describes a likelihood ratio and turns the maximum-likelihood-method into a minimisation probme.

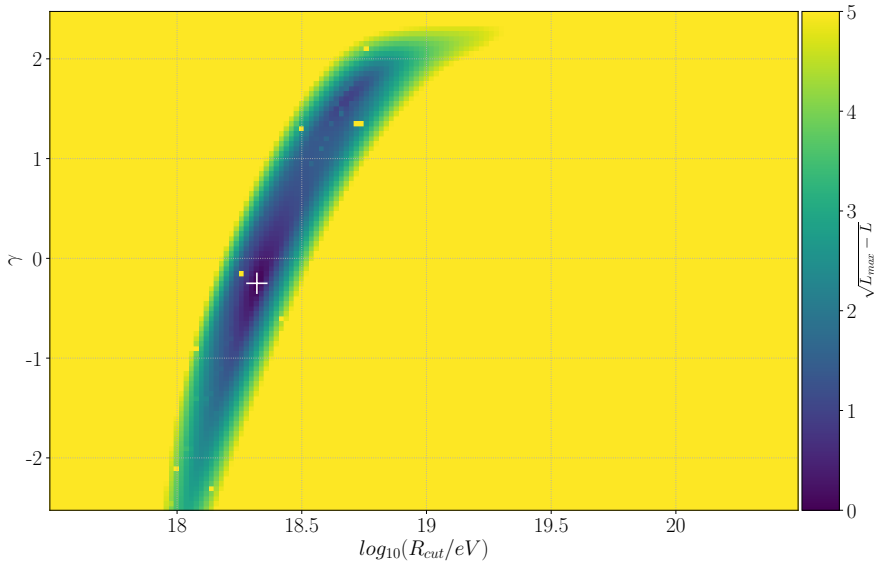


Figure 7.7: Scan of the likelihood function in the (γ, R_{cut})-plane. The color code denotes the pseudo standard deviance $n_\sigma = \sqrt{L_{\text{max}} - L}$ with respect to the optimal fit value which is denoted by an white cross. The optimum is found at $\gamma = -0.19$ and $R_{\text{cut}} = 18.34$. A second local minimum is situated at $\gamma = 1.47$ and $R_{\text{cut}} = 18.65$.

structure which was reported in [3, 10, 24] as a feature of the combined fit and is the same correlation we observed with the marginalized posterior distribution in Fig 7.3. We present the same likelihood scan for the off-target set in the appendix B.1.
We also scan the point source distance d_{P} and its total fraction f_{s} which is presented in Fig. 7.8 for the on-target and off-target set. This scan revealed topological features explaining the behavior of the Markov Chain Monte Carlo in 7.1.1. The scan of the on-target set describes the region for a nearby sources and a significant flux contribution by a high likelihood value. The scan finds the global minimum at $d_{\text{P}} = 3\,\text{Mpc}$ and a source fraction of $f_{\text{s}} = 0.32$ which coincides with the results of the Markov Chain Monte Carlo method. Furthermore a distant source beyond 10 Mpc or a too small source fraction is disfavored. For the off-target set the topology of the likelihood space is the opposite as for the on-target set: The region for a strong contributing nearby source is disfavored and a source fraction approaching 0 is preferred or the point source situated at large distances. The global minimum is situated at $d_{\text{P}} = 71\,\text{Mpc}$ and $f_{\text{s}} = 0.16$ which is in the region of the best fit value found with the combined fit above.

7.2 Distance Scan

By scanning the distance parameter, we want to statistically resolve a point source and give an estimation of the expected sensitivity of this method.

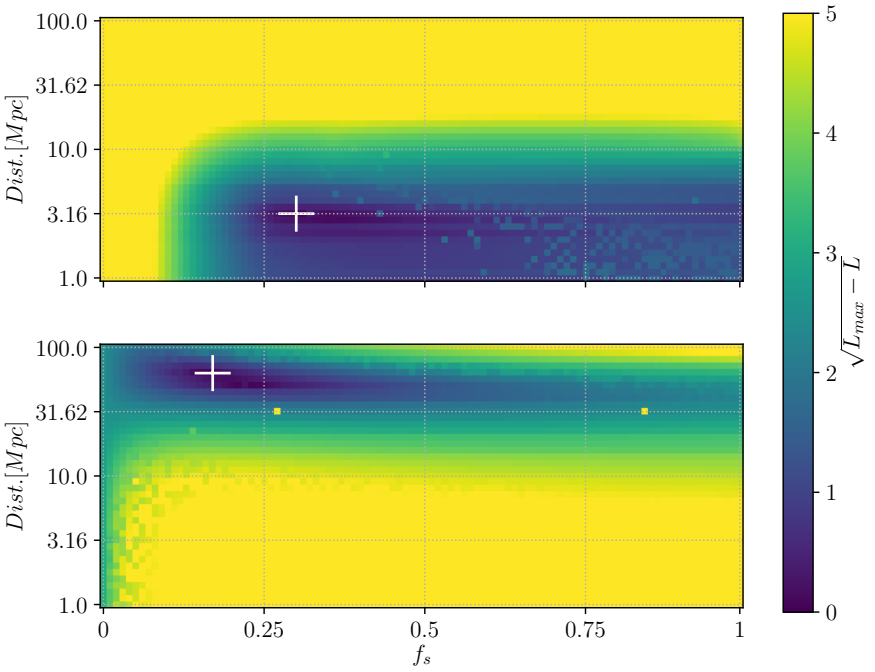


Figure 7.8: We show the same scan setup as for Fig. 7.7 for the fit model parameters d_{P} and f_{s}. In the *top* we show the likelihood scan for the on-target set and for the off-target set in the *bottom*

7.2.1 Likelihood Ratio

By fixing the point source at different distances we want to know how much better the fitted distance is in comparison to an null hypothesis. Therefore we formulate the null hypothesis as the best fit value for a fit model without a nearby source. This is accomplished by setting $f_{\mathrm{s}} = 0$. We compute the likelihood ratio:

$$Q = 2 \cdot \log\Big(\frac{L_{\mathrm{d}}}{L(f_{\mathrm{s}} = 0)}\Big) \tag{7.1}$$

In Fig. 7.9 we show the median likelihood ratio and the 68 % quantile for 1000 on-target sets and 1000 off-target sets. We see a clear separation for distances up to 10 Mpc. In this range the on-target set has a significant higher likelihood ratio than the off-target sets since the nearby source is better described by a fit model considering a nearby source component. For larger distances we observe that the off-target sets can be better described with a point source than with the null hypothesis. In this range the likelihood ratio drops for the on-target sets and approaches the case of an isotropic fit model.

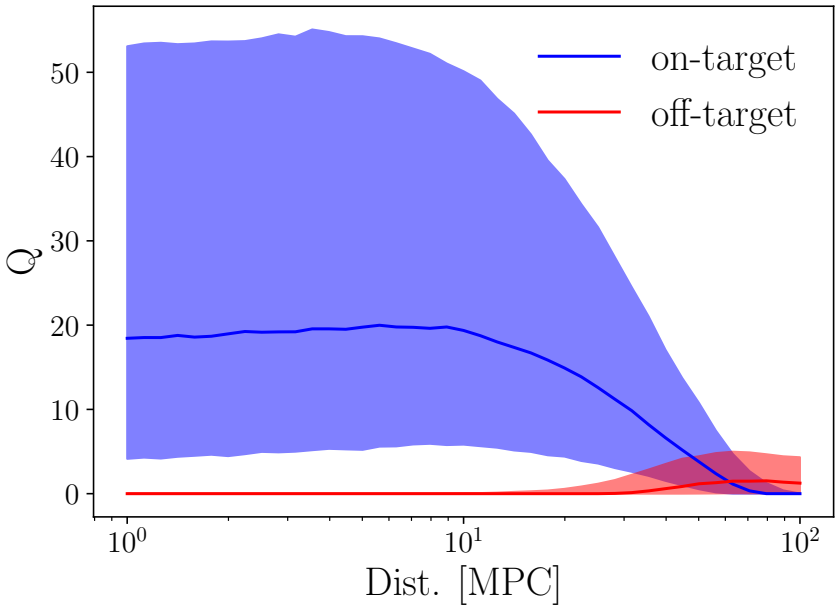


Figure 7.9: We show the likelihood ratio Q for the best fit value for different distances of the point source specified in the fit model 6.1. We compute the median and the 68 % quantile for 1000 on-target sets and 1000 off-target sets.

7.2.2 Expected Sensitivity

We have seen that the combined fit method developed in this work is in principle able to distinguish between on-target and off-target sets. To quantify this method for an investigation on sky regions we have to assess the statistical stability. Therefore we compute the best fit value from the distance scan and fill its likelihood ratio Q into a histogram for the on-target and off-target sets. These two distributions are shown on the right in Fig. 7.10. We also compute the median of Q for the on-target sets and count the sets of off-target sets which lay above this value. It turns out that from 1000 sets exactly 1 set lays above this value which leads us to p-value of $P(Q_{\text{iso}} > \text{median}_{\text{signal}}) = 0.001$. On the left in Fig. 7.10 we show the distance distribution of the best fit value. For the on-target sets closer regions are preferred, whereas the distribution tends to distant positions for off-target sets.

The distribution of the likelihood value for off-target sets fluctuates at higher ratios due to its low statistic thus our calculation of the p-value is not precise. Therefore we fit an exponential function to the tail of the distribution to estimate the integral laying above the median of the on-target set as shown in on the left in Fig. 7.11. This approach provides a conservative p-value estimation of $P(Q_{\text{iso}} > \text{median}_{\text{signal}}) = 0.002$.

We finally estimate the range where we would expect p-values if the nature follows our model. Therefore we compute beside the median also the 68 % quantile and calculate the p-value for this range using the fitted estimation which is shown in

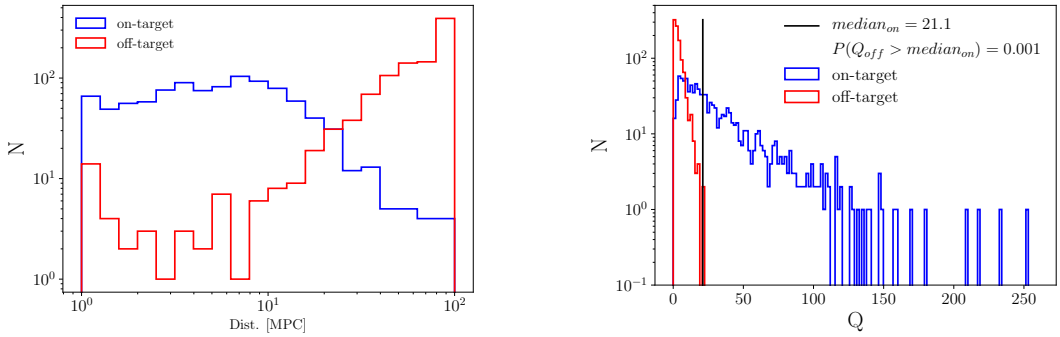


Figure 7.10: *left*: Distribution of distances of the best fit value for 1000 signal and 1000 off-target sets. *right*: Distribution of the likelihood ratio for the signal and the off-target sets. we also display the median of the on-target sets to calculate the p-value for an hypothesis test.

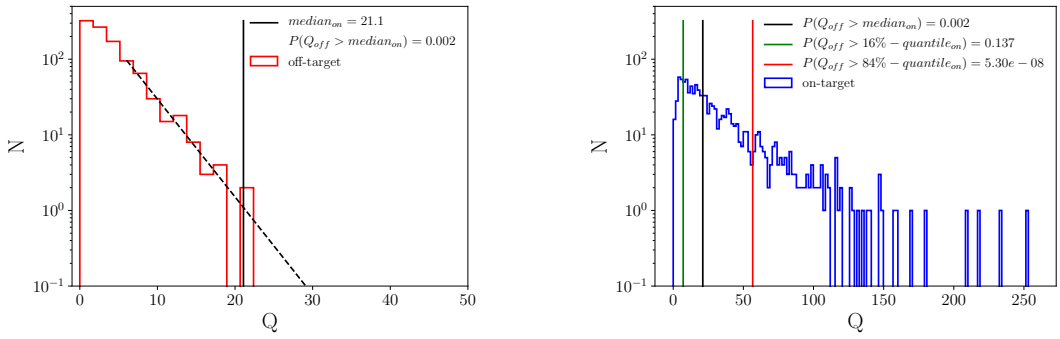


Figure 7.11: *left*: Likelihood ratio distribution for the off-target sets and the median for the on-target sets. To estimate the p-value we fitted an exponential function to the last bins. *right*: We show the likelihood ratio distribution for 1000 on-target sets and the median together with the 68 % quantile to calculate the expected sensitivity.

Fig. 7.11. Therefore we get for the lower limit of the 68 % quantile a p-value of 0.137 and for the upper limit a p-value of 5.3×10^{-8} for the upper limit. We have to note that these values are based on an approximation. In the case of an investigation on data we would have to generate more off-target sets for a precise p-value estimation.

7.3 Discussion

In this chapter we presented the performance of the nearby source detection using a combined fit. We have shown that the fit is able to resolve a nearby point source in individual cases. We showed with a likelihood scan that the combined fit method is capable to reproduce features which have been observed in previous works, i.e. [3, 10]. Furthermore, we performed a scan in the distance variable and the source fraction which reveals topological properties distinguishing between Monte Carlo sets with

and without a nearby source.
To investigate the statistical stability of resolving a point source we scanned the distance parameter for 1000 on-target and 1000 off-target sets and computing the likelihood ratio as an information of how much better the model with the point source describes the Monte Carlo set in comparison to the null hypothesis of no point source. It turns out that the method predicts an expected sensitivity of the p-value to discard the off-target case between 0.137 and 5.3×10^{-8} with a median of 0.002. This p-values are calculated by fitting an exponential function to the tail of the background distribution, which is necessary due to low statistics.
Summarizing, we developed a promising method to investigate nearby source distances for sky regions using the energy and shower depth pattern which is an alternative approach to investigation only considering arrival directions of UHECR.

Chapter 8

Conclusion

One of the central unsolved questions in the field of astroparticle physics is the origin of ultra-high energy cosmic rays (UHECRs) observed on Earth. Recent studies at the *Pierre Auger Observatory* favor an astrophysical scenario with sources of UHECRs situated outside our Galaxy [22]. Another analysis found evidence for a correlation in the arrival directions of cosmic rays with starburst galaxies situated some Mpc away [2]. In this work we investigate the question whether the distance of the sources can be determined by the measured energy and shower depth by extending a combined fit method [3].
A previous analysis established a method to reveal the source properties by assuming a homogeneous and isotropic source distribution. Sources described by a hard energy spectrum, low maximum injection energies, and a heavy chemical composition are favored [3].
In this work we modified the combined fit method by adding a point source over a homogeneous background in order to resolve the distance and flux contribution of this source. A statistical fitting procedure based on a Markov Chain Monte Carlo method was implemented to investigate the source parameters. The new fit method was tested with a Monte Carlo study of random universes following a fixed source density describing a distinguished direction of the closest source. Here, we were able to determine the same constraints found by previous works on the combined fit, forming a reliable basis for new results.
It is found that the extended combined fit method can resolve the distance of a nearby source and distinguish between a scenario of sky regions exhibiting a nearby source (on-target) and an homogeneous background (off-target). To statistically quantify the performance of this method we scanned the distance parameter of the fit model for 1000 on-target and 1000 off-target Monte Carlo sets computing the likelihood ratio of the best fit value at the scanned distance and a null hypothesis described by no nearby source.
Thus, if the nature follows our physical motivated model, we are able to identify a point source in a sky region and disfavor a homogeneous background with a median p-value of $2 \cdot 10^{-3}$ by using the energy and shower depth pattern.

Appendix A

Derivation of formulas

A.1 Detailed Derivation of Formula

A.1.1 Gelman-Rubin Diagnostics

We calculate the Gelman-Rubin coefficient for the parameter traces θ_{ij} where i denotes the chain and j the individual sample steps inside a chain. In total we have m chains and n samples per chain. The calculation follows [44].
We calculate the Gaussian variance of each chain

$$s_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\theta_{ij} - \overline{\theta}_j)^2, \tag{A.1}$$

the mean of the variance is

$$W = \frac{1}{m}\sum_{j=1}^{m} s_j^2, \tag{A.2}$$

the next step we calculate the total mean of the parameter θ over all chains

$$\overline{\overline{\theta}} = \frac{1}{m}\sum_{j=1}^{m}\overline{\theta}_j. \tag{A.3}$$

The parameter B quantifies the spread between the chains

$$B = \frac{n}{m-1}\sum_{j=1}^{m}(\overline{\theta}_j - \overline{\overline{\theta}})^2 \tag{A.4}$$

Finally, the Gelman-Rubin coefficients are computed as

$$\hat{Var}(\theta) = (1 - \frac{1}{n}) \cdot W + \frac{1}{n} \cdot B \tag{A.5}$$

$$\hat{R} = \sqrt{\frac{\hat{Var}(\theta)}{W}} \tag{A.6}$$

Appendix B

Additional Plots

B.1 Additional Evaluation Plots

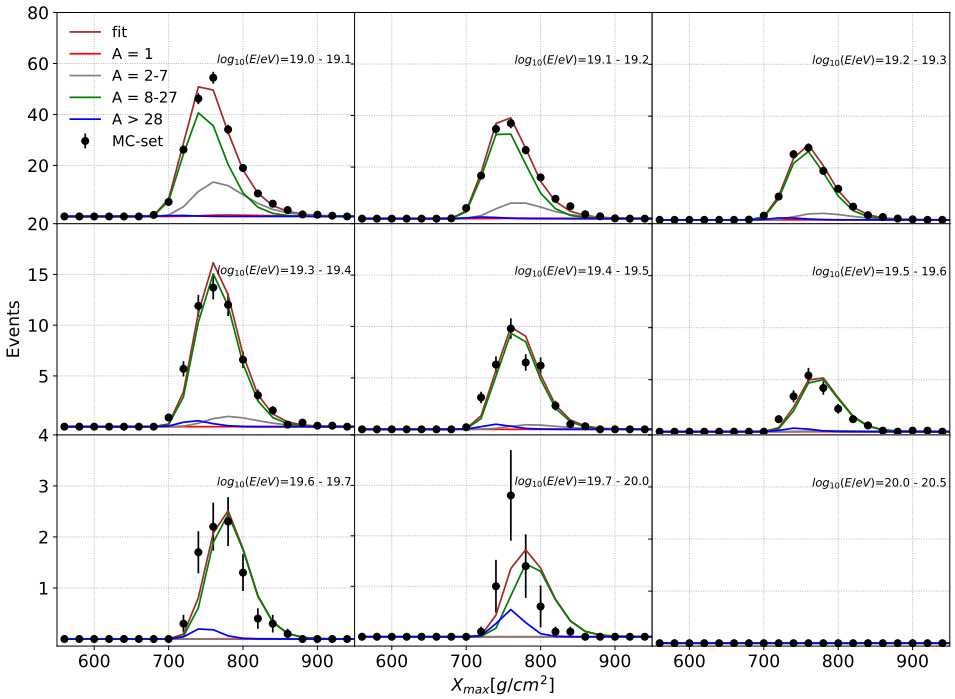


Figure B.1: The $X_{\max}$ distribution of the off-target set for the indicated energy bins. The fit model prediction for the best fit values from Table 7.1 are shown separated in 4 mass groups.

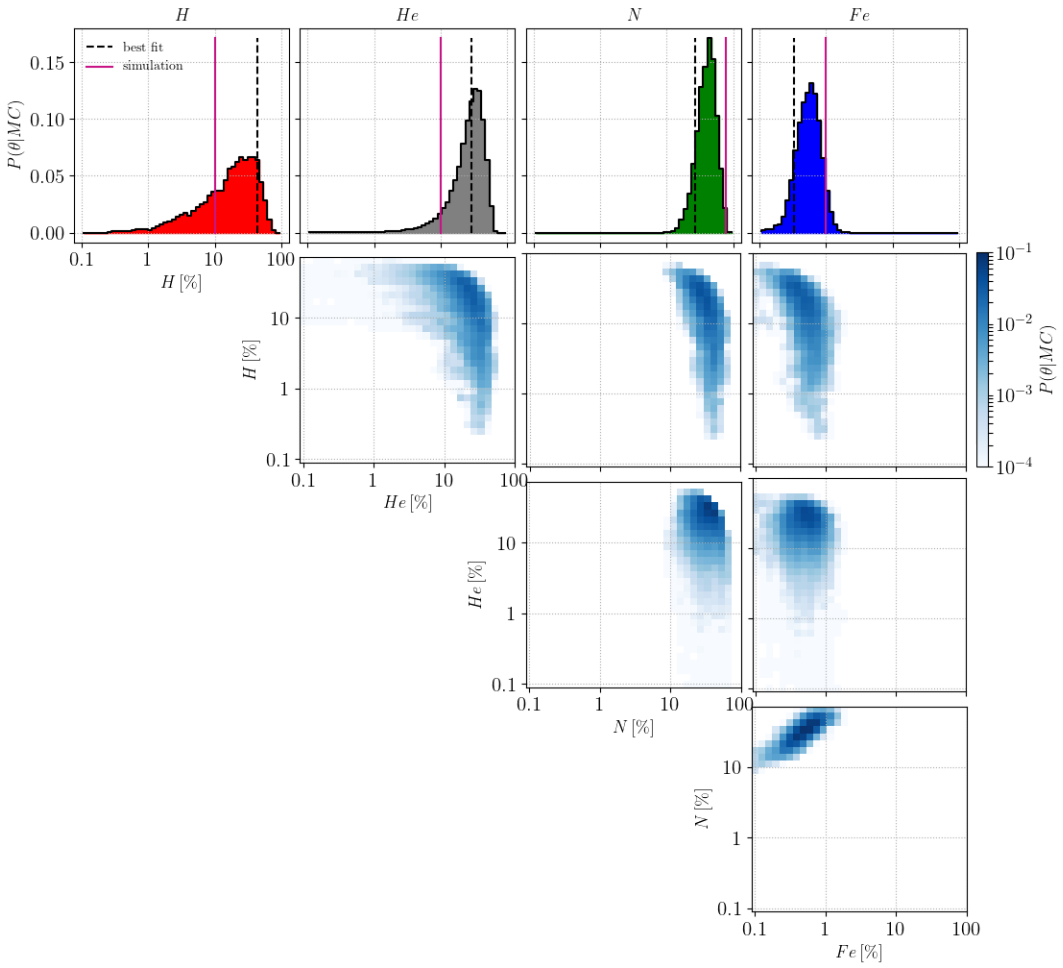


Figure B.2: The marginalized posterior distribution as described in Fig. 7.5 for the case of the isotropic set.

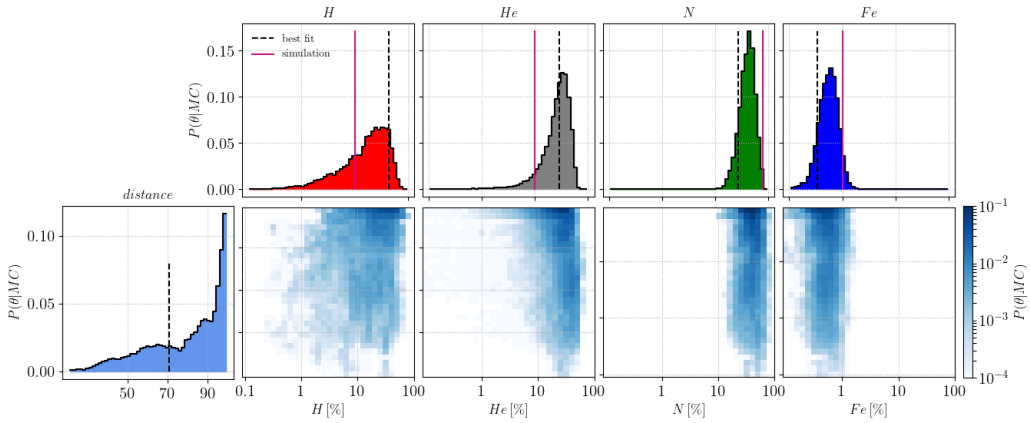


Figure B.3: The marginalized posterior distribution as described in Fig. 7.6 for the case of the isotropic set.

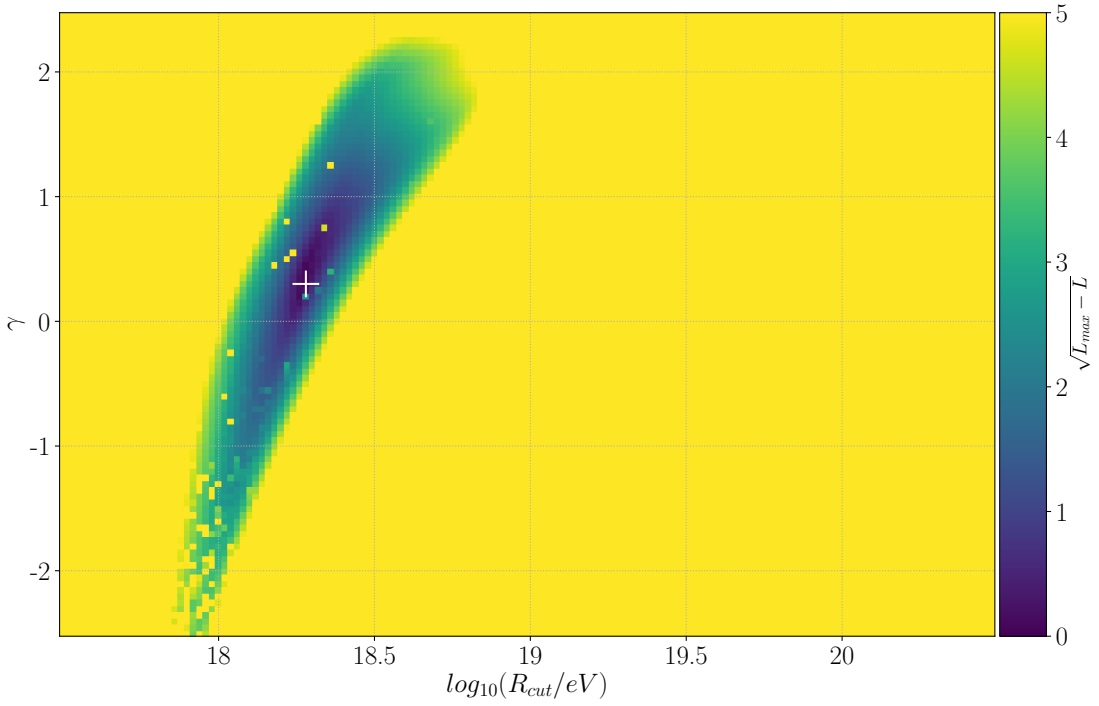


Figure B.4: We show the same scan setup as for Fig. 7.7 for the case of the off-target set. The optimum is found at $\gamma = 0.34$ and $R_{\mathrm{cut}} = 18.30$ which corresponds to the case of the combined fit presented in Table 7.1.

Bibliography

[1] V. F. Hess, "Über beobachtungen der durchdringenden strahlung bei sieben freiballonfahrten," *Phys. Z.*, vol. 13, pp. 1084–1091, 1912.

[2] A. A. et al., "An indication of anisotropy in arrival directions of ultra-high-energy cosmic rays through comparison to the flux pattern of extragalactic gamma-ray sources," *The Astrophysical Journal Letters*, vol. 853, no. 2, p. L29, 2018.

[3] A. A. et al., "Combined fit of spectrum and composition data as measured by the pierre auger observatory," *Journal of Cosmology and Astroparticle Physics*, vol. 2017, no. 04, p. 038, 2017.

[4] T. Stanev, *High Energy Cosmic Rays*. Springer Praxis Books, Springer Berlin Heidelberg, 2010.

[5] B. G. Keilhauer, *Investigation of atmospheric effects on the development of extensive air showers and their detection with the Pierre Auger Observatory*. PhD thesis, Karlsruhe, Forschungszentrum, 2004.

[6] W. Heitler, *The Quantum Theory of Radiation*. The International Series of Monographs on Physies, The Clarendon Press, 1936.

[7] J. Matthews, "A heitler model of extensive air showers," *Astroparticle Physics*, vol. 22, no. 5, pp. 387 – 397, 2005.

[8] K.-H. Kampert and M. Unger, "Measurements of the cosmic ray composition with air shower experiments," *Astroparticle Physics*, vol. 35, pp. 660–678, May 2012.

[9] M. D. Domenico, M. Settimo, S. Riggi, and E. Bertin, "Reinterpreting the development of extensive air showers initiated by nuclei and photons," *Journal of Cosmology and Astroparticle Physics*, vol. 2013, no. 07, p. 050, 2013.

[10] D. S. Walz, *Constraining models of the extragalactic cosmic ray origin with the Pierre Auger Observatory*. Dissertation, RWTH Aachen University, Aachen, 2016. Veröffentlicht auf dem Publikationsserver der RWTH Aachen University.

[11] A. e. a. Aab, "The pierre auger observatory upgrade - preliminary design report," *ArXiv:1604.03637*, 2016.

[12] F. G. Schröder, "Radio detection of high-energy cosmic rays with the auger engineering radio array," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 824, pp. 648 – 651, 2016. Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors.

[13] J. A. et. al, "The fluorescence detector of the pierre auger observatory," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 620, no. 2, pp. 227 – 251, 2010.

[14] R. Baltrusaitis, G. Cassiday, R. Cooper, B. Dawson, J. Elbert, B. Fick, D. Liebing, E. Loh, P. Sokolsky, and D. Steck, "The fly's eye detector: Present and future," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 264, no. 1, pp. 87 – 92, 1988.

[15] "The pierre auger cosmic ray observatory," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 798, pp. 172 – 213, 2015.

[16] I. A. et. al, "The surface detector system of the pierre auger observatory," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 586, no. 3, pp. 409 – 420, 2008.

[17] S. Fliescher, *Antenna devices and measurement of radio emission from cosmic ray induced air showers at the Pierre Auger Observatory*. PhD thesis, Aachen, 2012. Aachen, Techn. Hochsch., Diss., 2011.

[18] C. Bonifazi, "The angular resolution of the pierre auger observatory," *Nuclear Physics B - Proceedings Supplements*, vol. 190, pp. 20 – 25, 2009. Proceedings of the Cosmic Ray International Seminars.

[19] M. e. a. Tanabashi, "Review of particle physics," *Phys. Rev. D*, vol. 98, p. 030001, Aug 2018.

[20] K. A. Olive *et al.*, "Review of Particle Physics," *Chin. Phys.*, vol. C38, p. 090001, 2014.

[21] A. e. a. Aab, "Depth of maximum of air-shower profiles at the pierre auger observatory. i. measurements at energies above $10^{17.8}$ eV," *Phys. Rev. D*, vol. 90, p. 122005, Dec 2014.

[22] A. Aab *et al.*, "Observation of a Large-scale Anisotropy in the Arrival Directions of Cosmic Rays above 8×10^{18} eV," *Science*, vol. 357, no. 6537, pp. 1266–1270, 2017.

[23] R. Aloisio, D. Boncioli, A. di Matteo, A. F. Grillo, S. Petrera, and F. Salamida, "Simprop v2r4: Monte carlo simulation code for uhecr propagation," *Journal of Cosmology and Astroparticle Physics*, vol. 2017, no. 11, p. 009, 2017.

[24] D. W. for the Pierre Auger Collaboration., "The Pierre Auger Observatory: Contributions to the 35th International Cosmic Ray Conference (ICRC 2017)," *ArXiv e-prints*, Aug. 2017.

[25] J. A. et al., "Upper limit on the cosmic-ray photon flux above 1019ev using the surface detector of the pierre auger observatory," *Astroparticle Physics*, vol. 29, no. 4, pp. 243 – 256, 2008.

[26] Z. Fodor, S. D. Katz, A. Ringwald, and H. Tu, "Strong Neutrino-Nucleon Interactions at Ultrahigh Energies as a Solution to the GZK Puzzle," in *The Tenth Marcel Grossmann Meeting. On recent developments in theoretical and experimental general relativity, gravitation and relativistic field theories* (M. Novello, S. Perez Bergliaffa, and R. Ruffini, eds.), p. 741, Feb. 2006.

[27] R. Alves Batista, A. Dundovic, M. Erdmann, K.-H. Kampert, D. Kuempel, G. Müller, G. Sigl, A. van Vliet, D. Walz, and T. Winchen, "CRPropa 3 a public astrophysical simulation framework for propagating extraterrestrial ultra-high energy particles," *jcap*, vol. 5, p. 038, May 2016.

[28] B. Ryden, *Introduction to Cosmology*. Cambridge University Press, 2016.

[29] J. B. et al., "Review of particle physics," *Phys. Rev. D*, vol. 86, p. 010001, Jul 2012.

[30] E. Jonas Elias, *Investigation of Propagation Distances of Extragalactic Nuclei with CRPropa Simulations*. Bachelor thesis, RWTH Aachen University, Aachen.

[31] G. T. Zatsepin and V. A. Kuz'min, "Upper Limit of the Spectrum of Cosmic Rays," *Soviet Journal of Experimental and Theoretical Physics Letters*, vol. 4, p. 78, Aug. 1966.

[32] R. Jansson and G. R. Farrar, "The galactic magnetic field," *The Astrophysical Journal Letters*, vol. 761, no. 1, p. L11, 2012.

[33] K. Dolag, D. Grasso, V. Springel, and I. Tkachev, "Constrained simulations of the magnetic field in the local universe and the propagation of ultrahigh energy cosmic rays," *Journal of Cosmology and Astroparticle Physics*, vol. 2005, no. 01, p. 009, 2005.

[34] G. Sigl, F. Miniati, and T. A. Ensslin, "Ultrahigh energy cosmic rays in a structured and magnetized universe," *Phys. Rev. D*, vol. 68, p. 043002, Aug 2003.

[35] D. W. Hogg, "Distance measures in cosmology," *ArXiv Astrophysics e-prints*, May 1999.

[36] D. Sivia and J. Skilling, *Data Analysis: A Bayesian Tutorial*. Oxford science publications, OUP Oxford, 2006.

[37] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis, 2013.

[38] M. Erdmann and T. Hebbeker, *Experimentalphysik 5: Moderne Methoden der Datenanalyse Physik Denken*. Springer-Lehrbuch, Springer Berlin Heidelberg, 2013.

[39] P. Gagniuc, *Markov Chains: From Theory to Implementation and Experimentation*. Wiley, 2017.

[40] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines," , vol. 21, pp. 1087–1092, June 1953.

[41] W. K. Hastings, "Monte carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.

[42] N. A. Smith and R. W. Tromble, "Sampling Uniformly from the Unit Simplex," *Department of Computer Science / Center for Language and Speech Processing, Johns Hopkins University*, 2004.

[43] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Statist. Sci.*, vol. 7, pp. 457–472, 11 1992.

[44] F. Liang, C. Liu, and R. Carroll, *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples*. Wiley Series in Computational Statistics, Wiley, 2010.

[45] J. Salvatier, T. V. Wiecki, and C. Fonnesbeck, "Probabilistic programming in python using pymc3," *PeerJ Computer Science*, vol. 2, p. e55, Apr. 2016.

[46] D. J. Wales and J. P. K. Doye, "Global optimization by basin-hopping and the lowest energy structures of lennard-jones clusters containing up to 110 atoms," *The Journal of Physical Chemistry A*, vol. 101, no. 28, pp. 5111–5116, 1997.

[47] E. Jones, T. Oliphant, P. Peterson, *et al.*, "SciPy: Open source scientific tools for Python," 20018.

[48] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *Ann. Math. Statist.*, vol. 9, pp. 60–62, 03 1938.