Universität Potsdam
Mathematisch-Naturwissenschaftliche Fakultät
Institut für Physik und Astronomie

Modelling the Galactic population of very-high-energy γ-ray sources

Arbeit zur Erlangung des akademischen Grades
"doctor rerum naturalium"
(Dr. rer. nat.)
in der Wissenschaftsdisziplin Physik

Eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Potsdam

Von
Constantin Beverly Steppa

1. Gutachter: Prof. Dr. Christian Stegmann
2. Gutachter: Prof. Dr. Martin Karl Wilhelm Pohl
3. Gutachter: Prof. Dr. Christopher van Eldik

Potsdam, den 28. September 2021

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-54947
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-549478

Abstract

The current generation of ground-based instruments has rapidly extended the limits of the range accessible to us with very-high-energy (VHE) γ-rays, and more than a hundred sources have now been detected in the Milky Way. These sources represent only the tip of the iceberg, but their number has reached a level that allows population studies. In this work, a model of the global population of VHE γ-ray sources based on the most comprehensive census of Galactic sources in this energy regime, the H. E. S. S. Galactic plane survey (HGPS), will be presented. A population synthesis approach was followed in the construction of the model. Particular attention was paid to correcting for the strong observational bias inherent in the sample of detected sources. The methods developed for estimating the model parameters have been validated with extensive Monte Carlo simulations and will be shown to provide unbiased estimates of the model parameters. With these methods, five models for different spatial distributions of sources have been constructed. To test the validity of these models, their predictions for the composition of sources within the sensitivity range of the HGPS are compared with the observed sample. With one exception, similar results are obtained for all spatial distributions, showing that the observed longitude profile and the source distribution over photon flux are in fair agreement with observation. Regarding the latitude profile and the source distribution over angular extent, it becomes apparent that the model needs to be further adjusted to bring its predictions in agreement with observation. Based on the model, predictions of the global properties of the Galactic population of VHE γ-ray sources and the prospects of the Cherenkov Telescope Array (CTA) will be presented.

CTA will significantly increase our knowledge of VHE γ-ray sources by lowering the threshold for source detection, primarily through a larger detection area compared to current-generation instruments. In ground-based γ-ray astronomy, the sensitivity of an instrument depends strongly, in addition to the detection area, on the ability to distinguish images of air showers produced by γ-rays from those produced by cosmic rays, which are a strong background. This means that the number of detectable sources depends on the background rejection algorithm used and therefore may also be increased by improving the performance of such algorithms. In this context, in addition to the population model, this work presents a study on the application of deep-learning techniques to the task of γ-hadron separation in the analysis of data from ground-based γ-ray instruments. Based on a systematic survey of different neural-network architectures, it is shown that robust classifiers can be constructed with competitive performance compared to the best existing algorithms. Despite the broad coverage of neural-network architectures discussed, only part of the potential offered by the application of deep-learning techniques to the analysis of γ-ray data is exploited in the context of this study. Nevertheless, it provides an important basis for further research on this topic.

Kurzfassung

Die aktuelle Generation bodengestützter Instrumente hat die Grenzen des uns mit sehr hochenergetischer (very-hig-energy, VHE) γ-Strahlung zugänglichen Bereichs rasch erweitert, so dass inzwischen bereits mehr als hundert Quellen in der Milchstraße entdeckt wurden. Diese Quellen repräsentieren zwar nur die Spitze des Eisbergs, doch ihre Anzahl hat ein Niveau erreicht, das Populationsstudien ermöglicht. In dieser Arbeit wird ein Modell der globalen Population von VHE γ-Strahlungsquellen vorgestellt, das auf den umfassendsten Zensus galaktischer Quellen in diesem Energiebereich, dem H. E. S. S. Galactic plane survey (HGPS), beruht. Bei der Erstellung des Modells wurde ein Populationssynthese-Ansatz verfolgt. Besonderes Augenmerk wurde auf die Korrektur der starken Beobachtungsverzerrung gelegt, die der Stichprobe detektierter Quellen innewohnt. Die für die Schätzung der Modellparameter entwickelten Methoden wurden mit umfangreichen Monte-Carlo-Simulationen validiert und es wird gezeigt, dass sie akkurate Schätzungen der Modelparameter ermöglichen. Mit diesen Methoden wurden fünf Modelle für verschiedene räumliche Verteilungen von Quellen erstellt. Um die Gültigkeit dieser Modelle zu prüfen, werden ihre Vorhersagen für die Zusammensetzung der Quellen innerhalb des Sensitivitätsbereichs des HGPS mit der beobachteten Stichprobe verglichen. Mit einer Ausnahme werden für alle räumlichen Verteilungen ähnliche Ergebnisse erzielt, die zeigen, dass das beobachtete Longitudenprofil und die Quellenverteilung über den Photonenfluss gut mit der Beobachtung übereinstimmen. Bezüglich des Latitudenprofils und der Quellenverteilung über die Winkelausdehnung zeigt sich, dass das Modell weiter angepasst werden muss, um dessen Vorhersagen mit den Beobachtungen in Einklang zu bringen. Auf der Grundlage des Modells werden Vorhersagen über die globalen Eigenschaften der galaktischen Population von VHE γ-Strahlungsquellen und die Perspektiven des Cherenkov Telescope Array (CTA) vorgestellt.

CTA wird unser Wissen über VHE γ-Strahlungsquellen erheblich erweitern, indem es die Detektionsschwelle für den Quellen senkt, vor allem durch einer im Vergleich zu Instrumenten der aktuellen Generation größeren Detektionsfläche. In der bodengebundenen γ-Strahlenastronomie hängt die Empfindlichkeit eines Instruments neben der Detektionsfläche jedoch auch stark von der Fähigkeit ab, Bilder von Luftschauern, die durch γ-Strahlen erzeugt werden, von denen zu unterscheiden, die durch kosmische Strahlung erzeugt werden und einen starken Hintergrund darstellen. Dies bedeutet, dass die Anzahl der detektierbaren Quellen von dem verwendeten Algorithmus zur Hintergrundunterdrückung abhängt und daher möglicherweise auch durch eine Verbesserung der Leistung solcher Algorithmen erhöht werden kann. In diesem Zusammenhang wird in dieser Arbeit zusätzlich zum Populationsmodell eine Studie über die Anwendung von Deep-Learning-Techniken für die Aufgabe der γ-Hadron-Trennung bei der Analyse von Daten von bodengestützten γ-Strahleninstrumenten vorgestellt. Auf der Grundlage einer systematischen Untersuchung verschiedener neuronaler Netzwerkarchitekturen wird gezeigt, dass robuste Klassifikatoren konstruiert werden können, die im Vergleich zu den besten bestehenden Algorithmen eine konkurrenzfähige Leistung aufweisen. Trotz des Umfangs der diskutierten neuronalen Netzwerkarchitekturen wird im Rahmen dieser Studie nur ein Teil des Potenzials ausgeschöpft, das die Anwendung von Deep-Learning-Techniken für die Analyse von Daten in der γ-Astronimie bietet. Dennoch bietet sie eine wichtige Grundlage für weitere Forschungen zu diesem Thema.

Contents

1 Introduction

Each window along the electromagnetic spectrum offers a unique perspective on the universe. In particular, the window associated with very-high-energy (VHE, $E_\gamma \geq 100\,\mathrm{GeV}$) γ-rays allows us to study the origin of cosmic rays and their interaction with the environment. It was one of the last windows to open to astronomers, only gaining practical relevance for source discovery in the 1980s. This is because of the weakness of γ-ray sources and the challenges associated with the indirect observational techniques we currently have to use on Earth if we want to measure the low fluxes of the sources. To date, research in ground-based γ-ray astronomy is primarily driven by the search for new sources and the in-depth analysis of individual, known sources, since for most of them the physical origin of the observed emission is not yet clear. However, with nearly 150 Galactic sources already discovered, we are beginning to get a picture of what the Galactic population of VHE γ-ray sources could possibly look like. Because of the nevertheless small sample of sources and the strong observational bias inherent in it, the current picture is still incomplete and distorted. But with some effort it seems possible to straighten it out and fill in the gaps with the knowledge we have, for example from observations at other wavelengths. The global picture, if correctly reconstructed, can provide guidance in many ways, e.g. in drawing connections between γ-ray sources and their physical origin, in resolving the ambiguity between diffuse emission and unresolved sources, or in optimising observational strategies.

To this end, this work is concerned with the construction of a model of the Galactic population of VHE γ-ray sources. The approach followed here is population synthesis with the aim of deriving the parameters of global distribution functions of the physical properties of the sources directly from the VHE γ-ray observations. To ensure a consistent description of observation and simulation, which is crucial for the population synthesis approach, the data used for the modelling is restricted to the extensive data set of the H.E.S.S. Galactic plane survey (HGPS). Particular attention is paid to correcting for the strong observational bias inherent in the sample of known sources. Two methods have been developed which are shown to provide unbiased estimates of the parameters of the model. The first method used to construct the model is likelihood maximisation, while the second method was developed later to overcome some of the limitations associated with using the likelihood method. This second method makes use of deep-learning techniques and reduces the uncertainties of the parameter estimates.

Another part of this work is devoted to exploring deep-learning techniques as a means to improve the analysis of VHE γ-ray data, in particular the discrimination between the hadronic background and actual γ-ray signals. Since the observational data is dominated by the background, increasing the discrimination power of the analysis leads to improved sensitivity and thus potentially to the discovery of new sources that would complement our picture of the population.

This thesis is structured as follows: In chapter 2 a brief introduction to the astrophysics

underlying the emission of γ-rays is given. This begins with cosmic rays and their interactions with the environment, since cosmic rays are the primary objective in the study of VHE γ-rays. This is followed by a summary of the observed source classes and the results of diffuse emission measurements. In chapter 3 an overview of the imaging atmospheric Cherenkov technique used to observe VHE γ-rays and the data analysis is given using the High Energy Stereoscopic System (H.E.S.S.) as an example. At the end of this chapter, the methodology used and the final results of the H.E.S.S. Galactic plane survey are described. The foundations of deep-learning techniques are laid out in chapter 4, before new algorithms for γ-hadron separation based on these techniques are presented and evaluated in chapter 5. The model of the Galactic population of VHE γ-ray sources developed in this thesis is presented in chapter 6. There, the construction of the model is described in detail, as well as the methodology developed to obtain unbiased estimates for the parameters of the model. For a variety of assumed spatial distributions of the sources, the model predictions are compared with observations. Finally, at the end of this chapter, predictions are made about the global properties of the population. In chapter 7, an improved method for parameter estimation based on deep learning is presented and the results obtained with this method are discussed. This thesis concludes with a summary of the results given in chapter 8.

2 Very-high-energy γ-ray astrophysics

The emission of cosmic γ-rays at very-high energies is the result of non-thermal processes, in particular the interaction of relativistic, charged particles, that are referred to as cosmic rays, with matter or radiation fields. From the observation of this emission we gain unique insights into the astrophysical origin of cosmic rays, the mechanisms that accelerate them up to energies of 10^{20} eV and the interaction of cosmic rays with their environment. γ-rays allow us to trace cosmic rays throughout the Galaxy and beyond. In contrast to that, the origin and path of cosmic rays measured at Earth is obscured by scatters on magnetic fields when these charged particles propagate through interstellar space. A special significance of γ-ray astronomy for astroparticle physics therefore results from its complementary nature to direct measurements of cosmic rays.

In the following section 2.1 a brief overview of cosmic rays is given. This includes a summary of their properties and the principal physical concepts that are considered to give rise to these properties and which cause the emission of γ-rays. Key observations of the Milky Way in the light of VHE γ-rays are then summarised in section 2.2, which includes known sources and the detection of diffuse emission.

2.1 Cosmic rays

The existence of cosmic rays was proven by Victor Hess in 1912 when he gave evidence of the cosmic origin of a previously unknown component of the ionising radiation measured on Earth [1]. This radiation consists mainly of ions, of which $\sim 98\,\%$ are protons and alpha particles and $\sim 1\,\%$ heavier nuclei. Another one percent are electrons and positrons. The energy range of these particles extends over several orders of magnitude, from giga electron volt (GeV) to a few tens of exa electron volt (EeV). The all-particle spectrum of cosmic rays can be approximately described by a power law $\Phi(E) \propto E^{-\alpha}$ with spectral index $\alpha \sim 3$.

Due to the high dynamic range both in energy and flux, with particle fluxes above $1\,\mathrm{m}^{-2}\mathrm{s}^{-1}$ at GeV energies and below $1\,\mathrm{km}^{-2}\mathrm{yr}^{-1}$ for the most energetic particles, different types of instruments must be used to achieve complete coverage of the SED. Balloon-borne experiments like BESS [3], ATIC-2 [4] and CREAM [5] or space-borne instruments like PAMELA [6] and AMS [7] allow direct and thus highly sensitive measurements of cosmic-ray energies and composition. Ground-based experiments such as KASCADE [8], the Telescope Array [9] or the Pierre Auger Observatory [10], on the other hand, provide cosmic-ray measurements by reconstructing properties of primary particles from those extensive air showers (EAS) they induce in the atmosphere (for a detailed review see [11]). These ground-based measurements are less sensitive, but compensate for the statistical limitation at energies above $\sim 1\,\mathrm{TeV}$ that arises in balloon- and space-borne instruments due to practical constraints on their detection area.

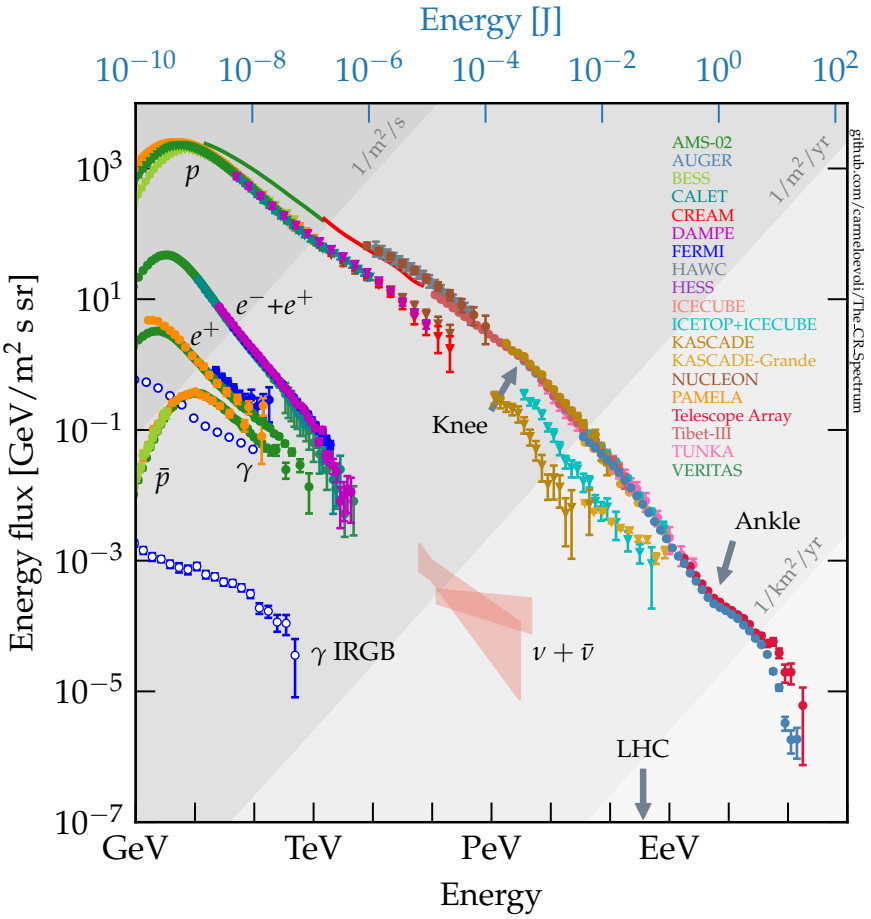


Figure 2.1: *The spectral energy distribution of cosmic rays and some of its components. This plot is composed of the most recent data from various instruments to cover the large dynamical range. Credit: C. Evoli [2].*

In Fig. 2.1, showing the spectral energy flux, the two most prominent features of the cosmic-ray spectrum are indicated, namely the "knee" and the "ankle". Located at $\sim 3\,\mathrm{PeV}$, the knee is characterised by a softening of the spectrum, $\alpha = 2.7 \rightarrow 3.1$ [11]. It is considered to mark the end of the cosmic-ray spectrum for Galactic accelerators. As such, a cut-off of the spectrum as a function of magnetic rigidity is predicted [12] and indeed measurements show a corresponding increase of the relative abundance of heavier elements with increasing energy around the knee [13, 14]. Above the knee a transition from a population dominated by Galactic cosmic rays towards an extragalactic particle population must take place. That is because such energetic particles are less efficiently confined within the Galaxy with its average magnetic field of a few micro Gauss. At $\sim 5\,\mathrm{EeV}$, the observed hardening of the spectrum, $\alpha = 3.3 \rightarrow 2.6$ [15], is referred to as the ankle. The gyro radius of ions with charge Z at this energy is $r_g \simeq 1\,\mathrm{kpc}\, Z^{-1} B_{\mu G}^{-1}$, i.e. larger than the thickness of the Galactic disc. The nature of the ankle is the subject of active debate. Various interactions with intergalactic radiation fields, especially with the cosmic microwave background (CMB), alter the propagated spectrum and composition of cosmic rays. However, both the spectrum and composition of extragalactic cosmic rays are uncertain at their source. If at the source the composition of the high-energy particles resembles that seen in the low-energy Galactic cosmic rays, the ankle

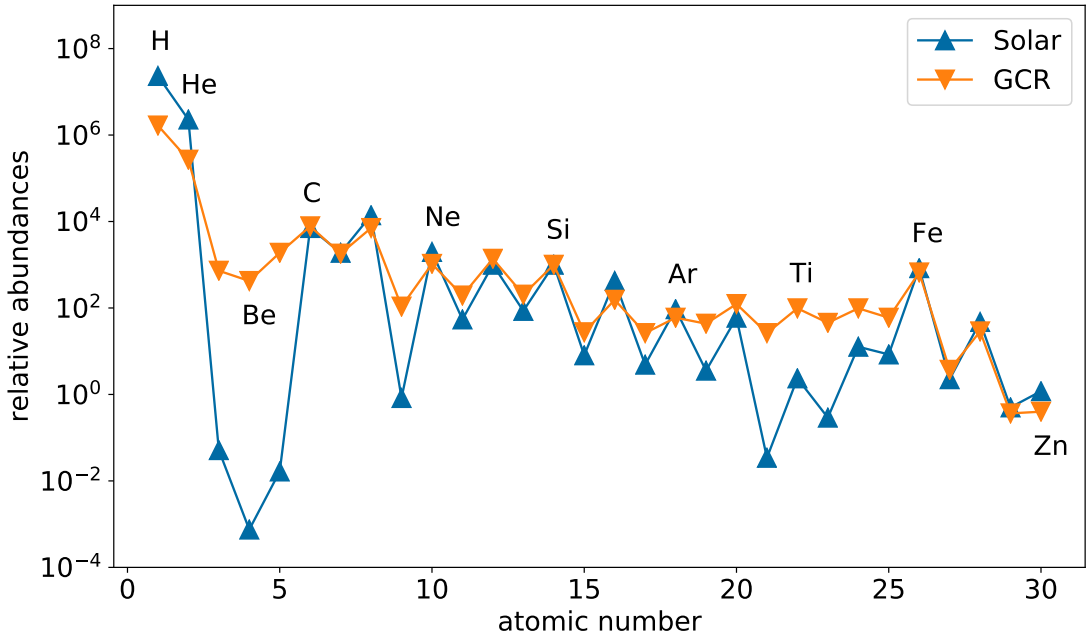


Figure 2.2: *Relative abundances of chemical elements in the solar wind (triangle up) and Galactic cosmic rays (triangle down) with* $Si = 10^3$. *The figure is adopted from [24].*

marks the transition from the Galactic-dominated population to the extragalactic population [16, 17]. Alternatively, for a pure proton composition at the source, the transition would be found at lower energies and the ankle corresponds to a dip in the propagated spectrum as a result of $e^- e^+$ pair production when the cosmic rays interact with photons of the CMB [18].

At the end of the cosmic ray spectrum, a suppression of the particle flux above $4 \cdot 10^{19}$ eV [19, 20] was measured, which is in agreement with predictions by Greisen [21] and Zatsepin & Kuz'min [22] (GZK effect). Assuming an extragalactic cosmic ray flux dominated by protons, pion production is predicted to lead to a spectral cut-off at $\sim 6 \cdot 10^{19}$ eV when cosmic rays interact with CMB photons. However, recent measurements show a heavier composition at the highest energies, suggesting that the steepening of the spectrum is formed by both the GZK effect and the maximum energy achievable in extragalactic sources [23].

Next to their energy distribution, the composition of cosmic rays reveals much about their physics. The composition of Galactic cosmic rays is similar to the one observed in the solar wind as shown in Fig. 2.2. Striking differences are found in the abundances of rare elements such as Lithium, Beryllium and Boron. These so-called secondary cosmic rays are produced through spallation of heavier, primary cosmic rays when these collide with particles of the interstellar medium (ISM). Secondary-to-primary ratios are a vital measure to constrain cosmic-ray propagation models.

The origin of cosmic rays is obscured since their motion is strongly affected by the interaction with turbulences in magnetic fields, that means they do not follow straight lines from the source to the observer. As a consequence, the arrival direction of cosmic rays is isotropic over almost the whole energy range, apart from some evidence of anisotropy at the highest

energies ($> 10^{19}\,\mathrm{eV}$) [25]. However, the emergence of the observed spectrum requires efficient accelerators. Physical considerations about potential acceleration mechanisms guide the search for the sources of cosmic rays.

2.1.1 Acceleration

The smooth spectrum of cosmic rays over many orders of magnitude in energy suggests a universal acceleration mechanism. From observations we know that such a mechanism must (i) naturally give rise to an energy spectrum following a power law with $\alpha = 2-3$, (ii) achieve a maximum particle energy of $10^{20}\,\mathrm{eV}$ and (iii) be efficient enough or potential accelerators abundant enough, respectively, to sustain the observed energy density of cosmic rays in the Galaxy. The primary candidate that is widely discussed in the literature is diffusive shock acceleration (see e.g. [26] or [27] for detailed reviews). Shocks are common in astrophysical environments, for example they are produced by stellar winds [28], supernova remnants [29] or in jets of active galactic nuclei [30]. Furthermore, in agreement with observations, common models of supernovae predict their remnant mass is ejected with a kinetic energy of the order of $10^{51}\,\mathrm{erg}$ for both type Ia and core-collapse supernovae [31–33] . Assuming a supernova rate within the Milky Way of 2-3 per century [34], a fraction $\gtrsim 10\,\%$ of the kinetic energy of the expanding shock of a supernova remnant has to be converted into the kinetic energy of cosmic rays in order to sustain the energy density of the Galactic population of cosmic rays[1]. Diffusive shock acceleration is an elegant explanation of how this transformation can be achieved, while at the same time in the test particle approach the resulting cosmic-ray energy spectrum naturally rises to a power law with index ~ 2.

It has been proposed independently by several authors as a mechanism to accelerate cosmic rays [35–38] and builds upon an original idea by Fermi [39]. Being affect by the Lorentz force, a charged relativistic particle propagating through a medium filled with clouds of plasma is scattered several times by the magnetic field of these clouds. If these clouds also move isotropically with an average speed V, head-on collisions are statistically more frequent, giving rise to an average energy gain of $\left\langle \frac{\Delta E}{E} \right\rangle = \frac{8}{3}\left(\frac{V}{c}\right)^2$ by the particle per scatter. This concept is generally referred to as second-order Fermi acceleration. In shocks, however, the same principle gives rise to a more efficient acceleration mechanism.

Shocks are formed by a stream of highly energetic particles moving supersonically through an ambient medium, which mostly can be considered being at rest. Thereby, the discontinuity at the intersection between the supersonically moving material and the ambient medium is called shock front. The strength of a shock is defined by its Mach number $M = \frac{U}{c_s}$ where U is the speed of the shock and c_s is the speed of sound in the ambient medium. One principal assumption in the vanilla, linear approach to diffusive shock acceleration is that cosmic rays are accelerated in strong shocks $M \gg 1$. Considering the shock front at rest as in the left panel of Fig. 2.3, the speed of plasma in the region ahead of the shock front, the so-called upstream region, is given by $v_u = U$, while conservation laws require that the compressed plasma behind the shock, in the so-called downstream region, moves with $v_d = \frac{1}{4}v_u$. Another assumption is that charged particles are isotropised in the rest frame of both the upstream and downstream region via scattering on magnetic fields (middle and right panel of Fig. 2.3).

[1] Further assuming that the composition and spectrum everywhere in the Galaxy is similar to that observed on Earth.

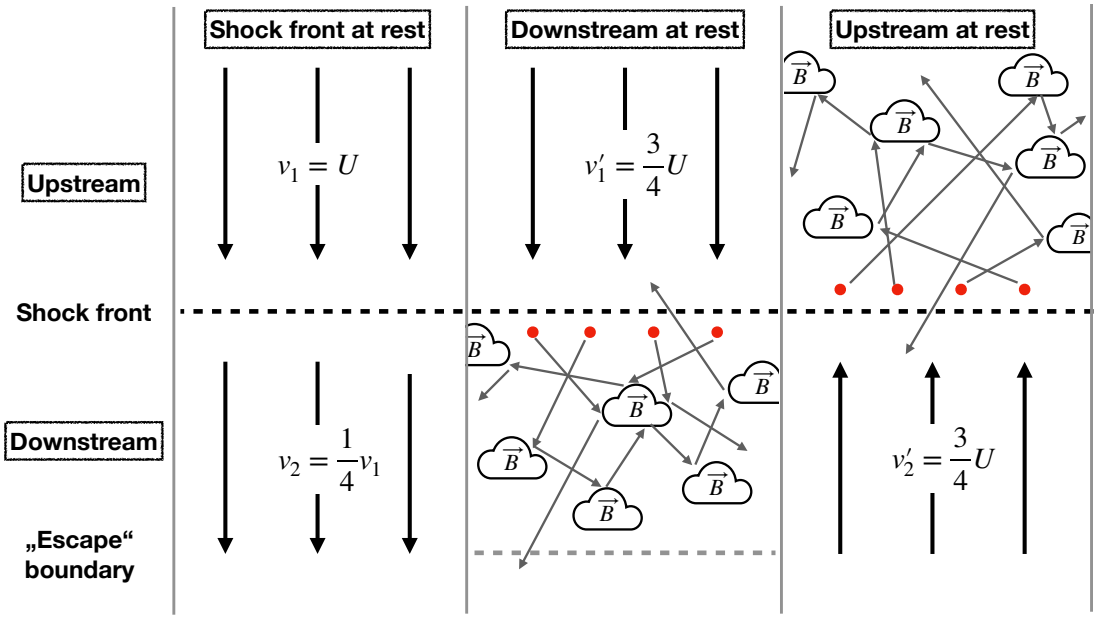


Figure 2.3: *Schematic view of diffusive shock acceleration considered in different rest frames. Left: The rest frame of the shock front.The motion of the plasma up and down stream of the shock is shown by bold arrows. Middle: Rest frame of the downstream. Clouds symbolise turbulences in the magnetic field while dots mark relativistic particles in the shock region. The diffuse motion of those relativistic particles is illustrated by thin arrows. An escape boundary in the rest frame of the downstream region marks the finite size of the acceleration region due to advection. Right: Rest frame of the upstream. Further description is given in the text.*

Thus, relativistic particles crossing the shock front effectively always experience head-on collisions with a plasma moving with a speed $V = \frac{3}{4}U$ towards them. For a round trip, particles on average gain an energy of $\left\langle \frac{\Delta E}{E} \right\rangle = \frac{8}{3}\left(\frac{V}{c}\right)$, i.e. this mechanism is first-order Fermi acceleration.

Via diffuse motion cosmic rays are considered to cross the shock front multiple times, thereby, reaching high energies. Unlike as in the upstream region, in the downstream region particles can be advected away from the shock front. This gives rise to a limited probability of particles to remain in the region of acceleration. Just like the energy gain this probability depends on the shock velocity and the number of shock crossings. Accounting for all these effects, it can be shown that diffusive shock acceleration gives rise to an energy spectrum of the desired form of a power law $N(E)dE \propto E^{-2}dE$. The maximum energy that a particle can achieve by diffusive shock acceleration depends, among other things, on the strength of the magnetic field at the shock, since it determines the diffusion coefficient and thus the size of the acceleration region.

There are several aspects of this simplified model that do not agree with observations to a satisfactory level. For example, taking into account propagation effects, the index of the energy spectrum of cosmic rays injected into the interstellar medium must be closer to $\alpha = 2.5$ to give a result consistent with the spectrum observed on Earth. Moreover, if the efficiency of the energy conversion from the kinetic energy of the shock to the kinetic energy of cosmic

rays is indeed $\geq 10\,\%$, this is accompanied by a non-negligible feedback of these cosmic rays on the dynamics of the magnetic field at the shock. In particular, cosmic rays excite fast-growing non-resonant modes and can thus amplify the initial magnetic field, which in turn affects the maximum achievable energy [40]. Extensive non-linear numerical simulations are being carried out to investigate the physical processes that occur in extreme environments such as shocks in more detail [41]. Not only diffusive shock acceleration is the focus of these studies, but also various other mechanisms that can accelerate cosmic rays to high energies, such as magnetic reconnection [42]. For a recent overview see e.g. [43]. To test the validity of these models, the spectrum of cosmic rays at the source is needed, which cannot be measured directly but is derived from the electromagnetic (and neutrino) emission produced by cosmic rays in the vicinity of the source.

2.1.2 Emission processes

Leptonic component

The spectrum of electromagnetic radiation emitted by cosmic-ray electrons can be broad and convoluted since there are multiple emission processes that may act simultaneously in a typical astrophysical environment. These processes are bremsstrahlung, synchrotron radiation and inverse Compton scattering, the latter being the primary emission channel for VHE γ-rays.

Inverse Compton (IC) refers to the scattering of a low-energy photon off a high-energy electron whereby energy is transferred from the electron to the photon. Scattered photons may stem from the cosmic microwave background (CMB), (inter-)stellar radiation fields or, in some cases, synchrotron radiation from the same population of electrons. A large part of the IC emission spectrum can be derived in the Thomson limit, which holds for photon energies in the centre of momentum frame that are smaller than the electron rest mass, $\gamma\hbar\omega \ll m_e c^2$. In that case, the energy loss of electrons as a result of the scatter is

$$-\left(\frac{\mathrm{d}E}{\mathrm{d}t}\right)_{IC} = \frac{4}{3}\sigma_T\, c\, U_{rad} \left(\frac{v^2}{c^2}\right) \gamma^2\,, \tag{2.1}$$

where σ_T is the Thomson cross section and U_{rad} is the energy density of an isotropic radiation field. For an electron population with a spectral energy distribution (SED) following a power law, $\frac{\mathrm{d}N}{\mathrm{d}E} = k\, E^{-\Gamma_e}$, the differential energy flux of IC radiation also follows a power law but with index $\Gamma_{IC} = \frac{\Gamma_e - 1}{2}$. The SED of IC emission emitted by an electron population with $\Gamma_e = 2$ that up-scatters photons of the CMB, with $T = 2.7\,\mathrm{K}$ and $U_{rad} = 0.25\,\frac{\mathrm{eV}}{\mathrm{cm}^3}$ is shown in the left panel of Fig. 2.4[2]. As described above, the SED follows a power law with a smaller spectral index as compared to the electron population over a wide range of energies. However, at high energies the emission process enters the extreme Klein-Nishina regime, $\gamma\hbar\omega \gg m_e c^2$, which leads to a spectral break at $\sim 100\,\mathrm{TeV}$ in this example. An extensive discussion on IC emission by high-energy electrons can be found in [44]. Furthermore, the simplified assumption that electrons scatter on an isotropic radiation field, which enters Eq. 2.1, does not hold true, for instance, on a global scale for a disc-like galaxy such as the Milky Way. The effect of anisotropic radiation fields and implications for the Galactic diffuse γ-ray emission

[2] All SEDs shown in Fig. 2.4 are calculated with GAMERA in the latest version publicly available on 15th October 2020 (`http://libgamera.github.io/GAMERA/docs/main_page.html`)

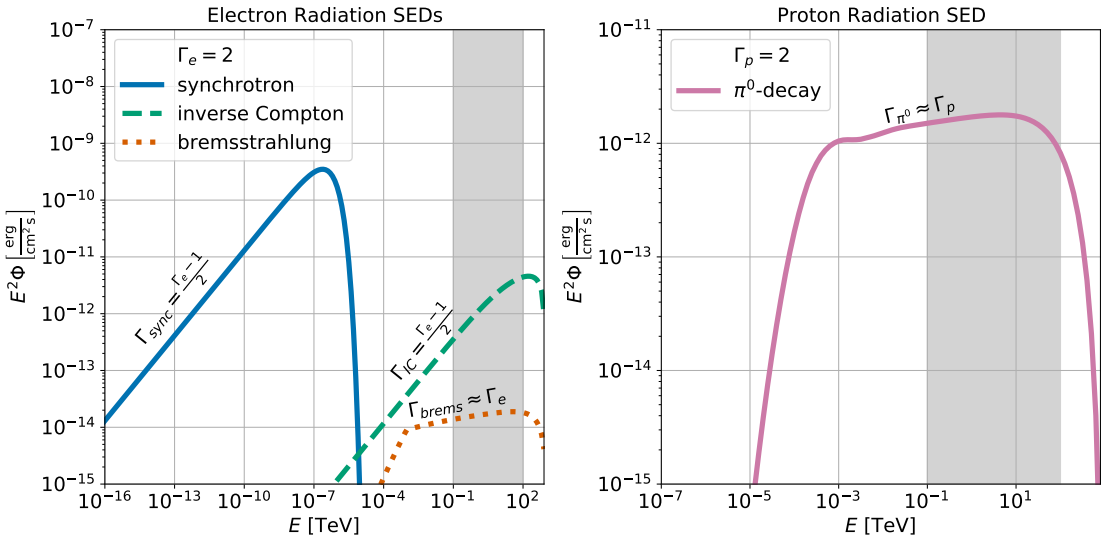


Figure 2.4: *SED of electromagnetic radiation emitted via various emission channels by a population of high-energy particles whose SED follow a power-law with spectral index* $\Gamma = 2$ *in the energy range* $1\,\mathrm{GeV} \leq E \leq 1\,\mathrm{PeV}$. *Left: Emission of an electron population. Right: Emission of a proton population. The VHE regime is marked by a grey band in each of the panels.*

are discussed in [45].

Electrons moving in magnetic fields emit synchrotron radiation owing to the Lorentz force that acts on them and bends their motion into spirals along magnetic field lines. Synchrotron radiation is observed in radio, optical and X-ray bands and is a valuable probe of magnetic field strength and structure. The interaction of electrons with the magnetic field can be described, for instance, by Compton scattering of electrons on virtual photons of the magnetic field [46]. Therefore, synchrotron emission spectra are similar to IC spectra but shifted towards lower energies. As shown in Fig. 2.4, for a power law SED of electrons and magnetic field strenght $B = 10\,\mu\mathrm{G}$ the synchrotron spectra also follows a power law over many orders of magnitude in energy with $\Gamma_{sync} = \frac{\Gamma_e - 1}{2}$. In accordance with the picture of Compton scattering on virtual photons, the equation describing the energy loss of electrons due to synchrotron radiation is of similar form as Eq. 2.1

$$-\left(\frac{\mathrm{d}E}{\mathrm{d}t}\right)_{sync} = \frac{4}{3}\sigma_T\, c\, U_{mag} \left(\frac{v^2}{c^2}\right) \gamma^2\,, \tag{2.2}$$

with the energy density of the radiation field being replaced by the energy densisty of the magnetic field $U_{mag} = \frac{B^2}{2\mu_0}$. As mentioned above, synchrotron photons may also be further boosted from radio and UV to X-rays and γ-rays by the same parent electron population via Compton scattering, a process called synchrotron self-Compton.

Additionally, when traversing matter relativistic electrons experience radiative energy losses as a result of Coulomb interactions. The cross-section for this interaction varies considerably for different media, for example neutral gas or ionised plasma, and can take non-trivial forms [47]. For a neutral gas, the energy loss of relativistic electrons due to bremsstrahlung

is given by [48]

$$-\left(\frac{\mathrm{d}E}{\mathrm{d}t}\right)_{brems} = \frac{Z(Z+1.3)\,e^6\,N}{16\,\pi^3\,\epsilon_0^3\,m_e^2\,c^4\,\hbar} E \left[\ln\left(\frac{183}{Z^{\frac{1}{3}}}\right) + \frac{1}{8}\right], \tag{2.3}$$

where Z is the atomic number and N the particle density of the gas. For a power-law electron spectrum the bremsstrahlung spectrum resembles the SED of the parent electrons over a wide energy range as seen in the example given in Fig. 2.4, where neutral hydrogen is taken as target material with number density $N = \frac{1}{cm^3}$, e.g. a typical value for the ISM. In this example, the VHE emission by IC exceeds bremsstrahlung in this energy range by orders of magnitude. Only if electrons are accelerated and confined in a region of increased matter density, for instance close to the shock front, bremsstrahlung can make a significant contribution to the VHE emission.

Hadronic component

Considering that atomic nuclei are much heavier than electrons, direct γ-ray emission from hadronic cosmic-ray primaries via the afore mentioned emission channels is negligible. Instead, γ-ray emission of this component stems from inelastic scatter with ambient matter via strong interactions. In particular, the dominant emission channel is that of π^0 decay subsequent to p-p collisions

$$p\,p \rightarrow p\,p\,\pi^0 \qquad\qquad \pi^0 \rightarrow 2\gamma\,.$$

From the kinematics of the π^0 decay follows that in a log-log representation the γ-ray spectrum is symmetric with respect to $E_\gamma = \frac{m_\pi c^2}{2} \approx 67.5\,\mathrm{MeV}$ [49], i.e. a distinct feature of the π^0 decay known as the pion bump. For a power-law distribution of cosmic-ray protons the γ-ray spectrum resembles the power law above $\sim 10\,\mathrm{GeV}$ with spectral index $\Gamma_\gamma \approx \Gamma_p$ as shown in the right panel of Fig. 2.4. However, the γ-ray spectrum is shifted towards lower energies with respect to the proton spectrum given that only a fraction $\kappa_\pi \approx 0.17$ of the kinetic energy of protons goes into π^0 production [50]. Furthermore, the spectral index Γ_γ will decrease by $\sim 4\,\%$ per decade in energy [51], which causes the slight slope of the spectrum shown in Fig. 2.4 above $\sim 10\,\mathrm{GeV}$.

Next to neutral pions, charged pions are equally likely produced in inelastic scatter of cosmic rays. These further decay into muons, electrons and neutrinos

$$\pi^+ \rightarrow \mu^+\,\nu_\mu \qquad\qquad \mu^+ \rightarrow e^+\,\bar{\nu}_\mu\,\nu_e$$
$$\pi^- \rightarrow \mu^-\,\bar{\nu}_\mu \qquad\qquad \mu^- \rightarrow e^-\,\nu_\mu\,\bar{\nu}_e\,.$$

Like primary electrons, secondary electrons may add to the γ-ray emission via the afore mentioned channels, giving rise to an additional component in the emission spectrum of cosmic-ray hadrons (not shown in Fig. 2.4). The emission of high-energy neutrinos, on the other hand, is a unique feature of hadronic cosmic-ray interactions.

2.1.3 Propagation

Once cosmic rays have escaped their source, they stream along magnetic field lines, resonantly exciting Alfvén waves. Due to the discrete nature of the sources, the streams are anisotropic

and, hence, streaming instability occurs, i.e. a net growth of the Alfvén waves [52]. Cosmic rays are scattered by these inhomogeneities and isotropised in the reference frame of the waves, so that their motion is effectively described by diffusion. This interaction results in a self-confinement of cosmic rays within the Galaxy. An extensive discussion of the cosmic-ray transport can be found in [53–56]. However, a general picture of the cosmic ray propagation can be derived from the simplified diffusion-loss equation, which for a particle of type i with number density N is written as [57]

$$\frac{\partial N_i}{\partial t} = D\,\nabla^2 N_i + \frac{\partial}{\partial t}\left[b_i(E) N_i\right] + Q_i(E) - \frac{N_i}{\tau_i} + \sum_{j>i} P_{ji}\frac{N_j}{\tau_j} \tag{2.4}$$

In this description an isotropic diffusion with scalar diffusion coefficient D is assumed. Energy losses and gains, e.g. via adiabatic expansion or the afore mentioned emission processes, are summarised by the second term on the right-hand side of Eq. 2.4. In addition, the injection of particles into the Galactic sea of cosmic rays is described by the source term $Q_i(E)$ while contribution from the spallation of heavier elements $j > i$ and losses due to spallation into lighter elements, each with characteristic spallation times τ_j and τ_i respectively, are described by the last two terms. The last term also includes a factor P_{ji} giving the probability of particle type i being the spallation product of particle type j.

Next to the isotropic distribution of cosmic rays, evidence for their diffuse motion is given by the measured ratio of the abundances of secondary and primary cosmic rays. For example, taking the boron-to-carbon ratio with known spallation cross section and an average density of the ISM within the Galactic plane of 1 particle per cubic centimetre, the calculated column density corresponds to a traversed distance of the order of $\sim 1\,\mathrm{Mpc}$. This number is much larger than the size of the Milky Way with a disc radius of $\sim 10\,\mathrm{kpc}$. In addition, radioactive isotopes allow the inference of the residence time of cosmic rays within the Milky Way. From the ratio of the abundances between these isotopes and their decay products, in particular the long-lived radioactive isotope ^{10}Be, a characteristic time of travel can be derived. This estimate is not affected by spallation, since different isotopes are considered to be produced in equal amounts through this process. In the case of ^{10}Be, the time of travel is of the order of 10^7 years. Since cosmic rays move almost at the speed of light, this result, combined with the secondary-to-primary ratio, suggests that cosmic rays spend most of their time in a medium much thinner than the canonical 1 particle per cubic centimetre within the Galactic plane [58]. It is commonly assumed that cosmic rays spend most of their time outside the Galactic disc, i.e. in the thin and highly ionised medium of the Galactic halo. Self-generated inhomogeneities of the magnetic field, driven by streaming cosmic rays, grow faster in the Galactic halo than in the Galactic disc, where the growth of these instabilities is suppressed, for example, due to the friction of plasma with neutral gas. Therefore, diffusion dominates the motion of cosmic rays in the halo, while streaming and advection play a pivotal role for their propagation within the Galactic disc [59].

To obtain a self-consistent picture of the cosmic-ray propagation that is in agreement with the wealth of precision measurements of local cosmic-ray properties [60], sophisticated cosmic-ray propagation codes have been written [61–63] that solve the transport equation numerically. Starting from a given source distribution, e.g that of pulsars, and an initial composition of cosmic rays, e.g. solar, and spectrum, these codes follow the propagation of cosmic rays and their interaction with the ISM by taking into account measured distributions of interstellar gas,

radiation fields and magnetic fields (see e.g. [64] for recent review of the numerical approach). The models that can be probed this way are not only constrained by direct measurements of local cosmic rays, but also from the observation of secondaries such as non-thermal emission. Thereby, diffuse γ-ray emission is used as an important tracer of propagating cosmic rays. Particularly at higher energies, diffuse emission allows more accurate studies of the nucleonic component of cosmic rays since electrons strongly suffer from radiation losses and, therefore, are effectively cooled in close vicinity to their sources.

2.2 Very-high-energy γ-rays

2.2.1 Galactic sources

With the current generation of instruments 229 sources of γ-ray emission at VHE have been detected. There are 86 sources assigned as of extragalactic origin, which are mostly associated with active galactic nuclei. Nearly 150 VHE sources are aligned with the disk of the Milky Way and thus probably of Galactic origin, with 80 VHE sources whose Galactic origin could be established. Most of these sources appear extended ($\sim 60\,\%$)[3]. Next to spatial coincidence, their association with known astrophysical objects is additionally based on spectral or morphology similarity to observations at other wavebands such as radio or X-ray. Due to the limited resolution of current-generation VHE instruments, often multiple associations with objects that lie within the emission region are possible resulting in unidentified sources.

Pulsar wind nebulae

Pulsar wind nebulae (PWNe) are the most frequently identified sources of VHE γ-ray emission. They are formed by ultra-relativistic electrons and positrons which are generated in the magnetospheres of fast spinning neutron stars (pulsars) and powered by their rotational energy. Since pulsars are formed in supernovae, PWNe commonly evolve inside SNRs. Thereby, the expanding winds of relativistic particles form extended nebulae. The dynamical evolution of PWNe can be complex for instance, due to the interaction with the reverse shock of the SNR. A description of their evolution can be found in [65].

VHE γ-rays are produced inside PWNe via inverse Compton scattering of relativistic electrons and positrons on ambient radiation fields. Energy-dependent radiation losses can result in energy-dependent morphologies like seen in the case of HESS J1825-304 (see Fig. 2.5) [66]. Such morphologies are characteristic and used to identify VHE sources as PWNe. Spatial coincidence with synchrotron emission from a PWNe observed in X-ray and radio is another feature to identify this class of sources. Synchrotron emission, however, is believed to be prominent only in young, energetic PWNe while inverse Compton emission can be emitted for several 10^4 years. Missing synchrotron associations possibly render a fraction of VHE-detected PWNe unidentified sources. Another common source of VHE emission are so-called composite SNRs which are spatially coincident with known SNRs and show extended emission with a filled centre. For these sources potential contributions from the SNR and the PWN inside it cannot be resolved, though PWNe are believed to be the dominant VHE emitter of these sources.

[3] Values are taken from TeVCat (http://tevcat2.uchicago.edu; Accessed: 2021-02-23)

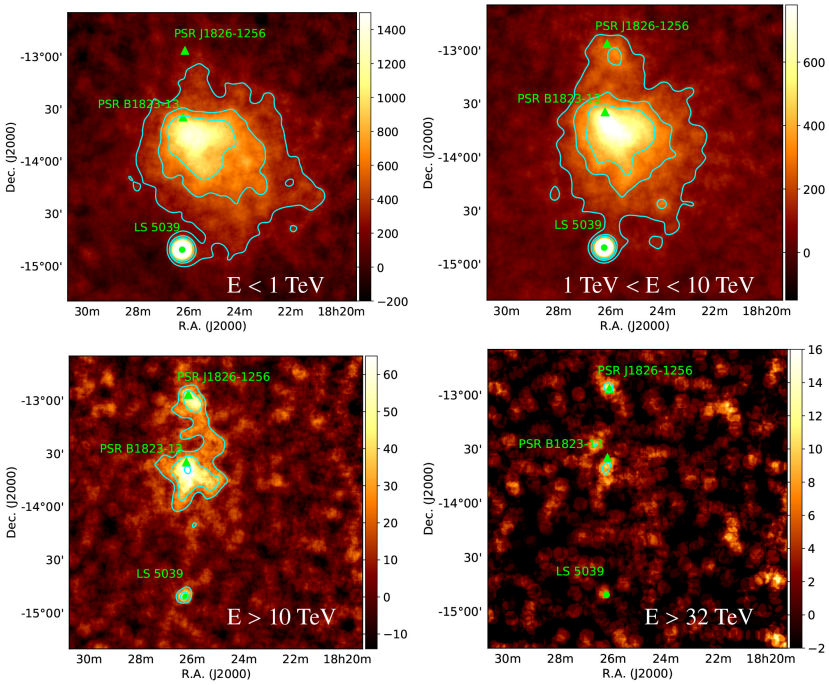


Figure 2.5: *H.E.S.S. γ-ray excess counts map of the HESS J1825–137 region in different energy bands. The extent of HESS J1825–137 around its central pulsar, PSR B1823–13, decreases with increasing energy. Additionally, VHE emission from the microquasar LS 5039 is visible in the lower part of the images. From [67] (Fig. 3).*

Supernova remnants

SNRs are the second most frequently identified sources of VHE γ-ray emission. They are unambiguously identified via a characteristic shell-like structure which is associated with the shock front of an SNR, as shown in Fig. 2.6 for the case of the VHE source RX J1713.7–3946 [68]. Typically, the shell is also visible in radio and X-rays by synchrotron emission of high-energy electrons.

The dynamical evolution of SNRs is well studied (see e.g. [69] and references therein). Following a phase of free, adiabatic expansion, a SNR enters the so-called Sedov-Taylor phase when the mass swept up by the remnant becomes comparable to the mass ejected in the supernova [70, 71]. The peak luminosity at VHE is reached during this second phase, which lasts for some 10^4 years before the transition into the radiative phase takes place and the SNR devolves into the ISM.

As mentioned above, the exact mechanism of particle acceleration, e.g. shock acceleration, and the corresponding emission channels for VHE γ-rays are still under debate. Knowing that high-energy electrons are present at the shock front, for the majority of the detected SNRs the observed VHE emission can be plausibly explained with inverse Compton scattering of high-energy electrons on ambient photon fields. Only for few sources a firm association of the VHE emission with hadronic interactions of nucleonic cosmic rays with their environment could be established [72–74]. If VHE emission of SNRs is indeed primarily of leptonic origin this could mean that these sources are inefficient cosmic-ray accelerators, which in turn would

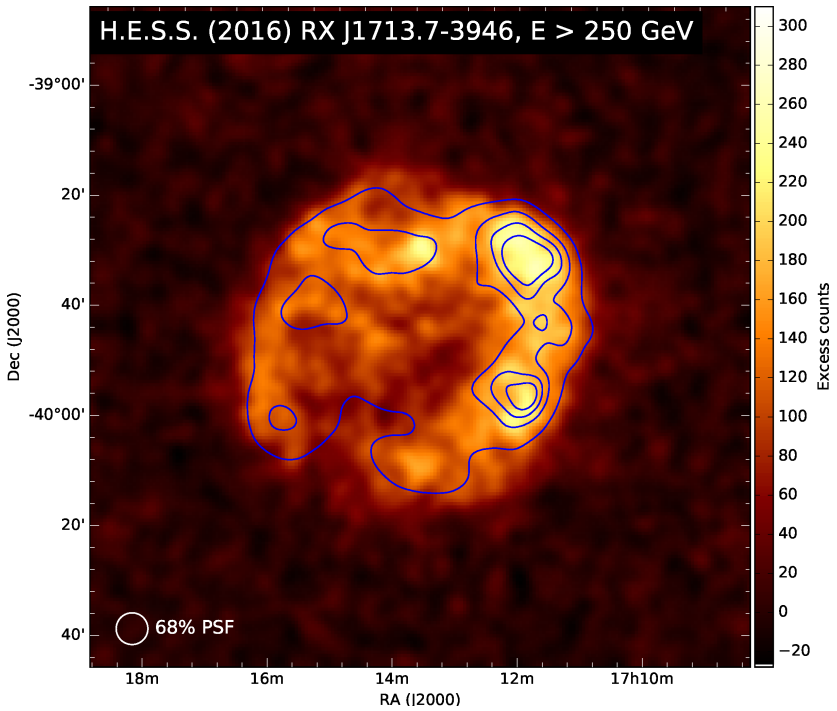


Figure 2.6: *H.E.S.S. γ-ray excess image of RX J1713.7–3946 with overlaid XMM-Newton contours (*1-10 keV*). From [68] (Fig. B.1).*

challenge the standard paradigm of SNRs being the prime source of Galactic cosmic rays. Alternatively, the environment around these SNRs might be simply not dense enough for the production of significant γ-ray emission by cosmic rays. Despite this ambiguity SNRs are considerable sources of VHE γ-rays.

Other sources

Binary systems are another identified class of VHE γ-ray sources that usually show periodic variability locked to their respective orbital phases. Most of these systems consist of a massive star and a compact object, either a pulsar or a stellar black hole. Depending on the nature of the compact object, high-energy particles may be accelerated in jets powered by accretion of matter from the companion onto the compact object [75] or in the magnetosphere of a pulsar [76]. Only recently the binary system η Carina has been shown to emit VHE γ-rays powered by high-energy particles that are accelerated in the colliding winds of the two massive stars constituting the system [77].

Interactions of cosmic rays from an active cosmic-ray accelerator in close proximity or inside a molecular cloud can give rise to considerable VHE γ-ray emission. γ-ray emission from molecular clouds is evidence for the presence of nucleonic cosmic rays and can be used to discriminate between a hadronicly or leptonicly dominated high-energy particle population powered by the nearby accelerator [78].

VHE γ-rays are also observed in the direction of stellar clusters [79,80]. Young stellar clusters can host a number of the afore mentioned objects. Thus, the observed emission might be

emitted by such individual, but yet unresolved sources or it is the result of collective effects of stellar evolution inside these cluster [81].

2.2.2 Diffuse emission

Cosmic rays that escape their source and propagate through interstellar space produce γ-rays through π^0 decay, inverse Compton scattering and bremsstrahlung. This gives rise to diffuse γ-ray emission proportional to the density of cosmic rays, interstellar matter and radiation fields along the line of sight. Thus, diffuse γ-ray emission provides unique insights into cosmic ray propagation and particle transport within the Galaxy. As mentioned in Sec. 2.1.3 the VHE regime is particularly interesting in this regard because, here, diffuse emission primarily stems from hadronic interactions of cosmic rays with the ISM.

While the study of diffuse emission has the potential to give answers to timely questions related to cosmic-ray propagation, for instance the spatial variation of the diffusion coefficient [82], diffuse emission is a subdominant component at VHE and as such hard to detect. Nevertheless, detection of diffuse VHE γ-ray emission has been reported by several ground-based instruments [83–86]. The measured signal exceeds the expectations for the contribution of diffusing cosmic rays in the standard model of cosmic-ray propagation. Considering that, in general, most of the observed VHE radiation comes from distinct sources that are only the closest sources to the observer or the most luminous sources within the Galactic population, a significant fraction of the measured diffuse emission is very likely associated with sources that have not yet been resolved [87]. In order to study different models and parameters of cosmic-ray propagation and be able to distinguish between them, a more precise measurement of diffuse γ-rays is needed, which must include accurate estimates of the contribution of unresolved sources. The model of the population of Galactic sources of VHE γ-rays presented in this work is developed for this purpose.

3 H. E. S. S. & the Galactic plane survey

H. E. S. S. is an experiment dedicated to explore the VHE universe. It is located in the Southern hemisphere on the Khomas highlands of Namibia and is run by an international collaboration of $\sim$ 250 scientists from $>$ 30 member institutes. Given its location, it is particularly well suited for studying Galactic sources of γ-ray emission. In fact, the survey of the Galactic plane was a major design goal that led to the most comprehensive census of Galactic VHE γ-ray sources to date. Its result is summarised in the HGPS catalogue which was released in 2018 [88]. Since this catalogue forms the basis for the population model presented in this work, an overview of the H. E. S. S. experiment and the Galactic plane survey is given in this chapter. This includes a brief introduction to the imaging atmospheric Cherenkov technique and a walkthrough of the complete analysis chain from shower reconstruction to source analysis, which is taken up again in chapter 5 where the study of the application of deep-learning techniques for signal-background discrimination in the analysis of shower images is presented. This chapter concludes with a summary of the catalogue-generation procedure and its outcome.

3.1 H. E. S. S. experiment

As a third generation experiment applying the imaging atmospheric Cherenkov technique (IACT), H. E. S. S. follows into the footsteps of the Whipple experiment, which pioneered this technique [89]. Building upon the experience of its predecessors, H. E. S. S. utilises stereoscopy and, thus, is a multi-telescope instrument. It consists of four identical telescopes with mirrors of 12 m diameter each and a fifth telescope with a mirror of 28 m diameter. A picture of the full array is shown in Fig. 3.1.

Operation started with phase I in 2002. Phase I denotes the first configuration level of the array, consisting of the four smaller telescopes alone (referred to as CT 1-4), which cover the energy range 100 GeV - 100 TeV [90]. During this phase, in the time from 2004 to 2013, H. E. S. S. performed the Galactic plane survey with unprecedented sensitivity.

In 2013 H. E. S. S. entered phase II with the addition of its big telescope (CT 5) to the array. With a mirror area $\sim$ 6 times larger than that of one of the smaller telescopes, the new telescopes is able to detect fainter air showers and, consequently, reduced the energy threshold for the detection of γ-ray signals down to tens of GeV. Furthermore, CT 5 is designed for fast retargeting, reaching any pointing position within 60 seconds [91]. These new capabilities of the instrument were added to comply with a shift in the scientific focus of the H. E. S. S. experiment towards the study of transient phenomena like short-lived γ-ray bursts in particular [92].

Aiming at increasing the rate of hybrid stereo events[1] for an improved reconstruction and,

[1] Events recorded by CT 5 plus at least one of the smaller telescopes

Figure 3.1: *The complete telescope array of H. E. S. S. phase II on the Khomas highlands of Namibia. Credit: C. Föhr.*

generally, increasing the robustness of the smaller telescope, the 14-years-old cameras of the H. E. S. S. I telescopes were successfully replaced by newly designed cameras in a major upgrade campaign in 2016 [93]. In 2020, the camera of CT 5 was replaced by a new model as well. These upgrades are a manifestation of the continuous effort of the H. E. S. S. collaboration to push the limits of ground-based γ-ray astronomy, resulting in significant scientific outcome [94–97]. Therefore, these efforts also find recognition in an approved prolongation of H. E. S. S. operations beyond its determined end of lifetime in 2018.

3.2 Imaging atmospheric Cherenkov technique

3.2.1 Air showers

The atmosphere of Earth is opaque to high-energy γ-rays, which renders their direct detection with ground-based detectors impossible. However, γ-rays of sufficient energy, that is few tens of GeV, initiate cascades of relativistic $e^{\pm}$ and high-energy photons in air that are detectable and called electromagnetic air showers (EAS). When hitting the atmosphere, a primary γ-ray photon produces an electron-positron pair in the Coulomb field of an atomic nucleus. The pair, in turn, emits high-energy photons via bremsstrahlung which, again, produce electron-positron pairs. This avalanche process continues until the energy loss rate of electrons is dominated by ionisation losses rather than by bremsstrahlung. In air this corresponds to a critical energy of $E_c \approx 85\,\mathrm{MeV}$. Generally, this is a fast process, which is why air showers are brief phenomena that last only few tens of nanoseconds.

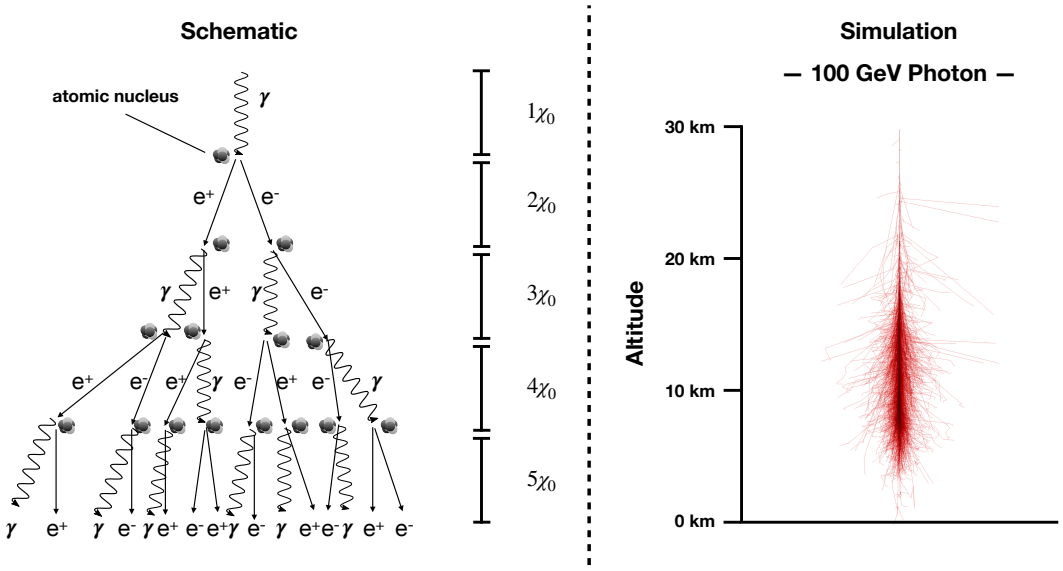


Figure 3.2: *Left: Schematic of an electromagnetic air shower according to the Heitler model. Right: Simulation of an air shower induced by a γ-ray with an energy of* 100 GeV *and a fixed altitude of first interaction at* 30.1 km. *Simulation credit [99].*

The physics of EAS is well known. A simplified model given by Heitler [98] is shown in the left panel of Fig. 3.2. There, a primary characteristic of the shower development is the radiation length χ_0, which is the average distance traversed until a charged particle looses all but $1/e$ of its energy. Typically, during one radiation length electrons emit one or two photons that have sufficient energy ($E > 2m_e c^2$) to create a new pair of relativistic $e^{\pm}$. In the ultra-relativistic limit, the free mean path of photons due to pair production is comparable to χ_0. This model describes well some of the characteristics of EAS, for instance, the maximum number of particles in the shower is proportional to the energy of the primary photon. Consequently, the energy of the primary can be inferred from the measurement of the shower size. For quantitive descriptions of EAS, however, extensive Monte Carlo simulations, such as the air shower shown in the right panel of Fig. 3.2. are utilised. These simulations acknowledge the stochastic nature of the shower development and allow, for example, the changing composition of the atmosphere or the influence of the magnetic field of the Earth to be taken into account over the length of the air shower. Simulations have shown that EAS are generally of compact, ellipsoidal shape. This is because there is little transfer of momentum across the direction of motion of the primary, mainly due to Coulomb scattering. It is found to be a characteristic that can be used to separate γ-ray induced air showers from cosmic-ray induced air showers.

Similar to γ-rays, cosmic rays initiate hadronic cascades in the atmosphere. This kind of air showers represents a sizeable background for the detection of γ-ray signals since the cosmic-ray flux is typically much higher than the flux of γ-rays ($\mathcal{O}(10^4)$). Fortunately, inelastic collisions between cosmic rays and atomic nuclei in air result in some differences between hadronic cascades and purely electromagnetic cascades that allow for a separation. Cosmic rays interact with atomic nuclei via the strong force and, thereby, loose energy by producing

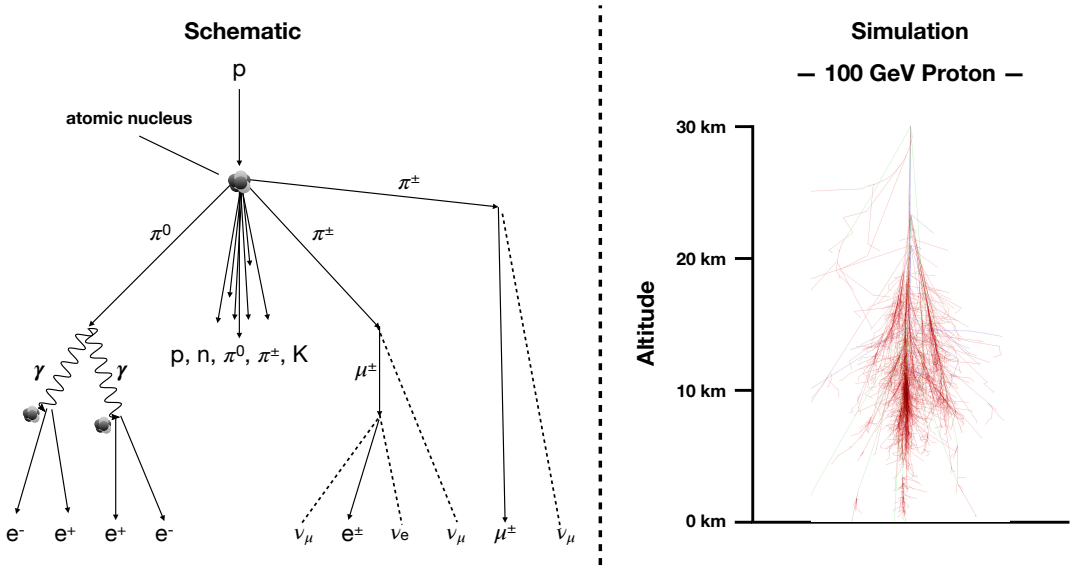


Figure 3.3: *Left: Schematic of an hadronic air shower. Right: Simulation of an air shower induced by a proton with an energy of* 100 GeV *and a fixed altitude of first interaction at* 30.1 km. *Simulation credit [99].*

cascades of new particles along their path of motion. The dominant component of these cascades are pions of which almost equal amounts are π^+, π^- and π^0. As described in Sec. 2.1.2, π^0 nearly instantly decay into two photons, while charged pions successively decay into muons, electrons and neutrinos. Just like primary γ-rays, those photons and electrons subsequently produce electromagnetic cascades as part of the hadronic shower. In contrast to that, muons are a distinct component of cosmic-ray induced showers. Furthermore, during the production of secondary particles in strong interactions, there can be a significant transfer of momentum across the direction of motion of the primary. Thus, hadronic cascades are commonly of wider shape than purely electromagnetic cascades as can be seen by comparing Fig. 3.2 and 3.3. Additionally, heavy nuclei of cosmic ray may experience spallation. Light nuclei that separate from the primary can carry a substantial fraction of the primary's energy and produce pronounced sub-showers next to the main shower. In total, these three features, a wide and irregular shape of the shower, potential sub-showers and muon signatures are exploited to separate cosmic-ray induced showers from γ-ray induced showers.

3.2.2 Cherenkov emission

A relativistic charged particle that moves faster than light in a dielectric medium is a source of coherent light emission. This condition is fulfilled for a large fraction of particles composing air showers. Because of the superluminal speed of those particles, the so called Cherenkov light (after its discoverer P.A. Cherenkov [100]) is emitted like a shock wave in a cone around the direction of motion of the emitting particle. A schematic applying Hyugens' principle is shown in Fig. 3.4. The angle θ at which the light cone is emitted with respect to the motion of the particle only depends on the speed β of the particle and the refractive index n of the medium, $\cos(\theta) = (n\beta)^{-1}$. For air showers, θ varies only little between 1° and 1.2°.

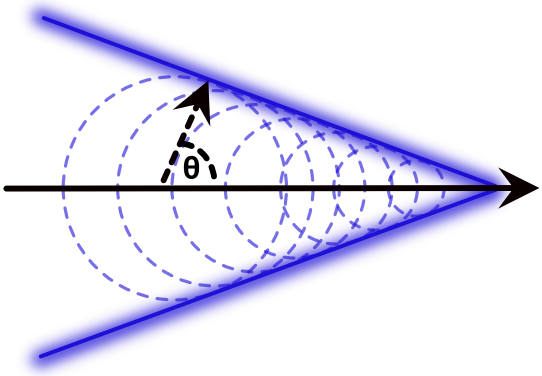


Figure 3.4: *Cherenkov light cone (blue, solid line) from a charged particle moving superluminal in a dielectric medium (black, solid arrow).*

An electromagnetic air shower that strikes the atmosphere from zenith direction, on average, reaches its maximum at 10 km a.s.l. and produces a faint Cherenkov light cone of ~ 120 m radius on ground.

Cherenkov light is emitted in a broad spectrum. The energy loss per unit length $\mathrm{d}x$ and unit emission frequency $\mathrm{d}\omega$ is given by [101]

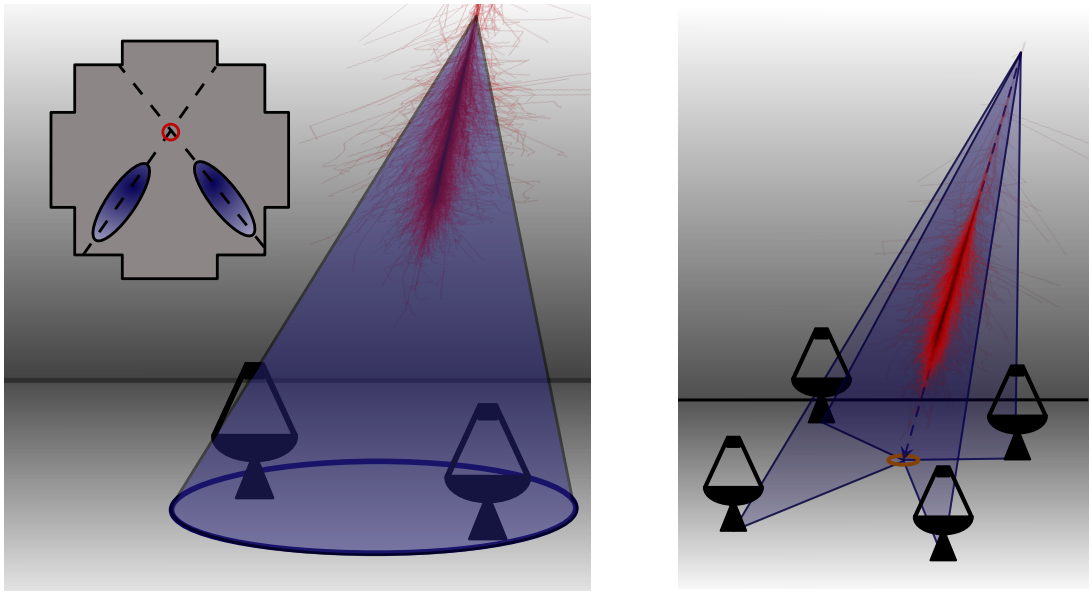
$$\frac{\mathrm{d}E}{\mathrm{d}x\,\mathrm{d}\omega} = \frac{q^2}{4\pi}\omega\mu(\omega)\left(1 - \frac{1}{n^2(\omega)\beta^2}\right) , \tag{3.1}$$

where q is the charge of the particle and μ is the permeability of the medium. In air most of the energy is emitted in the ultraviolet band. However, because of atmospheric absorption the Cherenkov light spectrum of air showers measured by ground-based detectors peaks in the optical blue band.

3.2.3 Detection

The three main properties of γ-ray photons that can be inferred from indirect measurements on ground are their energy, direction and time of arrival. These properties can be deduced from those air showers induced by γ-rays in the atmosphere. Ground-based optical telescopes are utilised to image air showers via their Cherenkov emission (see Fig. 3.5a) and, thus, are referred to as imaging atmospheric Cherenkov telescopes (IACT).

In order to detect γ-rays at energies of 100 GeV and even below, large mirrors are needed to capture the faint Cherenkov emission with a typical yield of Cherenkov photons of $100\,\mathrm{m}^{-2}\mathrm{TeV}^{-1}$ at 2 km a.s.l.. Mirrors with a diameter of $\gtrsim 10$ m are used in all current generation experiments. Through segmentation of the mirrors costs are reduced. With the bulk of air showers extending over 1°-2° in the sky a field of view (FoV) of $> 3°$ is required. The FoV of the H. E. S. S. I telescopes is chosen to be 5° allowing for the observation of extended sources, which are expected to be common in the Galactic plane. Since the evolution of individual showers is subject to statistical fluctuations covering the FoV with pixels that have an angular resolution of 0.2° is sufficient to determine the coarse shape of an air shower and reconstruct the properties of primary particles. The H. E. S. S. I telescopes are equipped with cameras



(a) *Reconstruction of direction* (b) *Reconstruction of impact distance*

Figure 3.5: *Illustration of the imaging atmospheric Cherenkov technique and geometric reconstruction of air-shower parameters. (a) Reconstruction of the direction of arrival via the intersection of shower axes in a common image plane. (b) Reconstruction of the impact distance. The plane defined by the shower direction and the position of a telescope can be derived by the azimuth angle of the major axis of the projection of the shower onto the respective telescope. The intersection of planes from multiple telescopes yield the impact distance. Image adapted from [102].*

consisting of 960 photo-multiplier tubes (PMTs), each with an angular resolution of 0.16°. More details concerning the requirements on the optical system of IACTs and specifics about the H. E. S. S. I telescopes in particular can be found in [90].

The reconstruction of air showers and the discrimination between γ-rays and cosmic rays is largely improved by the usage of stereoscopy [103]. With a typical Cherenkov light cone of 120 m diameter, a spacing of $\sim$ 100 m between telescopes is found to be a good choice for maximising the detection area and providing high coincidence rates. On the other hand, a spacing $>$ 80 m is required to prevent triggering on cosmic-ray muons. The H. E. S. S. I telescopes are arranged on a square with 120 m side length. Elaborate trigger algorithms are applied to reduce background events from cosmic rays and effects related to illumination by the night sky background (NSB). At H. E. S. S. a multi-level trigger scheme is set in place for stereoscopic events [104]. At first, a telescope triggers if a preset number of pixels T_1 in close neighbourhood exceed a threshold level of T_2 in a coincidence interval of 1.3 ns. Typical values for this local trigger are $T_1 = 3$ pixels and $T_2 = 4$ photo electrons. They are adjusted according to the NSB level in the FoV. Stereo events are only recorded to disc if two such local triggers occur coincident in a time interval of 80 ns at the central trigger.

A recorded image contains the signal that arrived at the camera within an integration interval of 16 ns while the sampling resolution is 1 ns. The Cherenkov emission of the majority of air showers is well contained within this interval and the integrated intensity allows for good reconstruction of those showers. However, for the most energetic showers ($E_\gamma > 100$ TeV)

this fixed interval might not capture the whole development of the shower. More information for the reconstruction could be retrieved from a temporally resolved intensity distribution. With the extended capabilities of the upgraded cameras investigations of the potential of using this additional temporal information are explored [105, 106].

3.3 Reconstruction of primary particle properties

3.3.1 Air shower & detector simulation

Ground-based γ-ray astronomy relies heavily on extensive Monte Carlo simulations. In the absence of suitable calibration sources and analytical solutions, simulations are used to characterise air showers and instrument responses on a statistical basis. Simulations of γ-ray and hadron-induced air showers alike deliver their respective intensity distribution of Cherenkov light on ground by following the evolution of the particle cascade in the atmosphere, taking into account particle interactions, Earth's magnetic field and atmospheric conditions. Detector simulations yield the response of the instrument to the incident light, taking into account optical efficiencies, signal processing and NSB noise. Simulations are conducted for multiple pointing positions at varying zenith angles because of the large impact of the atmospheric depth on the intensity distribution observed by IACTs. Additional simulations are required for changing optical efficiencies, for instance due to degrading reflectivity of mirrors, varying NSB noise across the sky or seasonal changes in the atmospheric conditions. In H. E. S. S. two independent shower and detector simulation chains are applied, namely Corsika [107]/sim_telarray [108] and Kaskade [109]/Smash which allows for cross checks of analyses and, thus, additional confidence in their results.

3.3.2 Data calibration

The principal data product of the camera are digital counts per pixel which are generated by an analogue-digital converter (ADC) from the electric current received from the PMT. This signal is a convolution of Cherenkov and NSB photons reaching the telescope with the instrument response which is characterised by an overall Cherenkov-light efficiency and a pixel-based photo-electron efficiency as well as an irreducible electric current in each pixel. The primary task of the data calibration is to relate those digital counts to actual Cherenkov photons.

The irreducible current, or more commonly referred to as pedestal P, is determined by electronic noise, the single-photon resolution of the PMT and NSB noise. The pure electronic noise can be measured when the camera is shielded from external light sources. Taking a series of images this way, the pedestal is characterised by the mean and width of the distribution of ADC counts (per pixel) derived from this series. During observation the pedestal can be significantly altered by contribution from the NSB, i.e. the distribution is broadened. Therefore, the pedestal is computed for each observation, whereby pixels in an image that contain signals from Cherenkov light are discarded from the calculation of the ADC count distribution.

Next to the pedestal, knowledge about the conversion factor from single photo electrons to ADC counts, also called gain, is required to infer the number of measured photons. For this purpose a specific calibration measurement is conducted, called Single PE run. There,

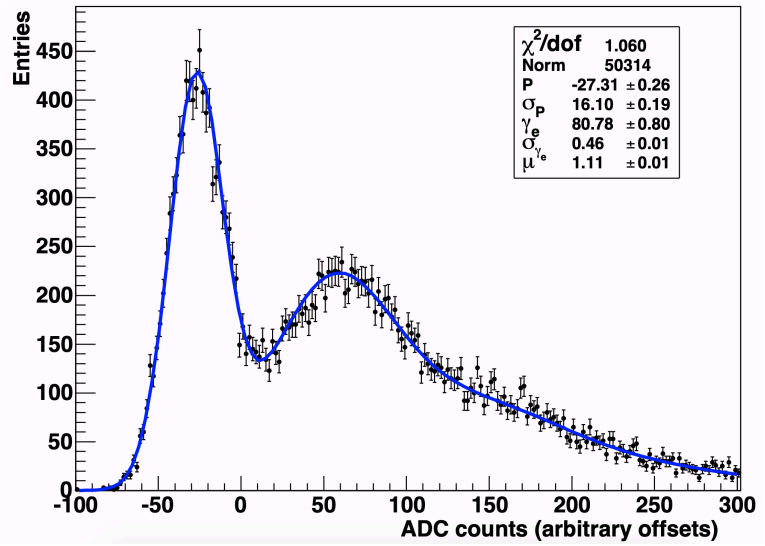


Figure 3.6: *Single photo-electron distribution from measurement (black markers) and fitted solution (blue solid line). Credit: [110].*

the camera is shielded from the night sky and illuminated by an artificial light source that is dim enough for the camera to take images of single photons per pixel. The resulting ADC count distribution is a combination of a Gaussian distribution for the pedestal plus Poisson distributions for a number of photo electrons convolved with a Gaussian characterising the resolution of the PMT ([110], eq. 6). An exemplary distribution is shown in Fig. 3.6 where the first peak corresponds to the pedestal while the second peak stems from single photo electrons. The distance between those two peak positions yield the gain γ_e. While there are two channels available per PMT to increase the dynamical range – a high gain channel for small photon counts and a low gain channel for high photon counts – the gain can be derived only for the high gain channel. For the low gain channel the same conversion factor γ_e as for the high gain channel is applied in addition with the amplification ratio $^{HG}/_{LG}$ between those two channels. This ratio can be derived easily from ADC distributions in the transition region.

Inhomogeneities among pixels within a camera are corrected for via a flat-fielding coefficient. This coefficient FF is derived from dedicated flat-fielding runs for which the camera is homogeneously illuminated by an artificial light source so that relative deviations to a baseline can be calculated. With that, the number of photo electrons A is calculated for the different channels as

$$A_{\mathrm{HG}} = \frac{ADC_{\mathrm{HG}} - P_{\mathrm{HG}}}{\gamma_e} \times FF_{\mathrm{HG}}\,, \qquad A_{\mathrm{LG}} = \frac{ADC_{\mathrm{LG}} - P_{\mathrm{LG}}}{\gamma_e} \times \frac{HG}{LG} \times FF_{\mathrm{LG}}. \tag{3.2}$$

In addition an absolute calibration of the telescopes' quantum efficiency, for instance including the mirror reflectivity, is needed in order relate photo electrons to Cherenkov photons. For this purpose images of atmospheric muons are used [111]. Those muons are frequently produced by hadronic showers and characterised by producing ring like features in camera images. Although the trigger layout is designed to suppress such events still a large enough fraction can be extracted from standard observations. An absolute calibration based on muon rings is possible because the number of Cherenkov photons can be calculated analytically [112].

In addition to deriving calibration parameters, data calibration is also about detecting faulty components such as pixels that are switched off or pixels that are strongly influenced by the night sky background, for example bright stars in the field of view. These can have a significant impact on the reconstruction quality of air showers and therefore need to be handled carefully.

Finally, dedicated pointing runs are performed monthly to ensure correct telescope pointing positions are used for the reconstruction of air showers. For this purpose the telescope is pointed towards bright stars with the camera lid closed. The reflection of those stars on the lid are captured by a CCD camera. With the known position of these stars the pointing of the telescope can be calculated precisely from the CCD camera's images.

3.3.3 Image cleaning

Image cleaning aims at extracting pixels in an image that captured the signal from an air shower rather than the night sky background or electronic noise. Depending on the applied algorithm used to reconstruct the shower, this can improve the result considerably. The standard procedure in H. E. S. S. is a two-level threshold cleaning [113]. More precisely, a high threshold value is applied at the first level to select clusters of pixels, each consisting of at least two neighbouring pixel with each pixel exceeding this threshold. At the second level neighbouring pixels are added to these first-level clusters if they exceed a second, lower threshold. Only clusters comprising at least three pixels after the second level are considered to belong to an air shower while all other pixels are discarded. The reasoning for this procedure is the pronounced spatial and temporal correlation of the signal produced by Cherenkov light from an air shower which is distinct from uncorrelated, randomly distributed signals produced by NSB or electronic noise. The applied thresholds directly affect the sensitivity of the system to low-energy γ-rays. A set of optimised thresholds allows for a trade-off between background rejection efficiency with respect to the night sky background rate in a given field of view and the overall energy threshold of the system. Typical high and low-level pixel threshold are 10 and 5 photo electrons, respectively. Once pronounced shower signatures have been identified with the two-level cleaning algorithm, the restriction of this procedure might be relaxed by adding additional rows of pixels around clusters to include fainter components of the shower.

3.3.4 Estimation of Hillas parameters

After image cleaning a robust analysis of shower images can be based on a parametrisation of the intensity distribution present in those images as suggested by [114]. Using the fact that γ-ray induced showers have ellipsoidal shapes the following characteristics are derived from the intensity distribution:

- the size, i.e. the sum over all pixels,
- the position of its centre of gravity in the camera plane,
- the length (**L**) and width (**W**), i.e. the root mean square along its major and minor axes, respectively,
- the nominal distance **d**, i.e. the length of the line segment connecting the centre of gravity with the centre of the camera plane,
- the angle φ between this line segment and the x-axis of the camera and
- the angle α between this line segment and the major axis.

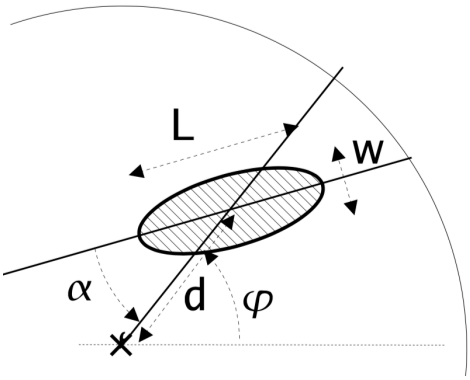


Figure 3.7: *Hillas Parameters as described in the text. Credit: [115].*

These so-called Hillas parameters can be found labeled for a schematic ellipsoidal intensity distribution in the camera plane shown in Fig. 3.7. Hillas parameters are used for discrimination between cosmic rays and γ-rays as well as the reconstruction of primary γ-ray photons. For both task the use of stereoscopic data is of great advantage [103].

Geometrical considerations allow for the estimation of the properties of the primary γ-ray photon. By projecting shower images into a common plane the shower direction and, thus, the direction of the photon is found at the intersection of the major axes of the shower images. The energy of the photon is estimated via the size parameter. The size is determined by the energy deposited into the shower and the impact distance of the shower, that is the distance between the telescope and the intersection of the shower's major axes with the ground. The latter can be derived from the orientation of the main axis of the projection of the shower onto the telescopes (see Fig. 3.5b). Look-up tables, which are generated using air-shower simulation, are then used to relate both parameters to the energy of the primary.

The shapes of the shower images, in particular the lengths and widths, are used to determine the type of the primary particle. For this purpose reduced scaled parameters are introduced, defined as [113]:

$$W_{\mathrm{RS}} = \frac{W - < W >}{\sigma_W}\,, \qquad L_{\mathrm{RS}} = \frac{L - < L >}{\sigma_L}\,. \tag{3.3}$$

Mean and standard deviations are stored in look-up tables for varying sizes, impact distances and zenith angles. Applying stereoscopy, a more robust discrimination between cosmic rays and γ-rays is achievable, i.e. by applying box cuts on mean reduced scaled parameters. For a shower captured by N telescopes these are calculated as

$$W_{\mathrm{MRS}} = \frac{1}{N}\sum_{i=1}^{N} W_{\mathrm{RS},i}\,, \qquad L_{\mathrm{MRS}} = \frac{1}{N}\sum_{i=1}^{N} L_{\mathrm{RS},i}\,. \tag{3.4}$$

Additionally, in order to improve the quality of the reconstruction, it is common to apply a set of preselection cuts. Images of small sizes should be discarded since they are strongly affected

by statistical fluctuations. Shower images with large nominal distances, that is showers close to the edge of the field of view of the camera, can be truncated. Therefore, discarding those images further improves the reconstruction. Finally, the quality increases with the number of images per shower which can be set by requiring a minimum telescope multiplicity. A detailed description of the geometrical analysis of shower images can be found in [103] and references therein. Next to this classic approach, machine learning algorithms, for instance boosted decision trees and neural networks, have also been demonstrated to successfully reconstruct properties of primary particles using Hillas parameters [116, 117].

3.3.5 Fitting shower images

A more precise, yet more elaborate method of reconstructing properties of γ-rays is fitting shower parameters to shower images using image templates. The implementation of such method requires a profound understanding of both air-shower development and detector response. In H. E. S. S., there are two implementations with comparable performance available. One relies on a semi-analytical model of air showers for the generation of image templates [118, 119], the other relies on extensive Monte Carlo simulations [120]. Templates are generated for a range of zenith angles, azimuth angles, primary energies, impact distances and first interaction depths. Interpolation between these high dimensional data points allows for calculating shower images for any shower parameter. This way, a prediction for the number of photo electrons captured by each pixel in the camera can be made and compared with observation. The likelihood for observing a signal s in a pixel i given an expectation of μ photo electrons is determined by a Poisson distribution for the photo electron counts convolved with a Gauss distribution characterising the pixel's photo electron resolution [119]:

$$P_i(s \mid \mu, \sigma_p, \sigma_\gamma) = \sum_n \frac{\mu^n e^{-\mu}}{n!\sqrt{2\pi\left(\sigma_p^2 + n\sigma_\gamma^2\right)}} \exp\left(-\frac{(s-n)^2}{2\left(\sigma_p^2 + n\sigma_\gamma^2\right)}\right). \tag{3.5}$$

The width of the Gaussian is determined by the pedestal width σ_p and the single-photon resolution σ_γ. Assuming independence of all pixels the negative total log-likelihood $-\sum_i \log P_i$ is minimised to find a good estimate of the shower parameters. The goodness of the fit provides additional discrimination power to separate γ-rays from hadrons. Indeed, this method yields excellent results, i.e. with a more accurate direction reconstruction and more efficient background rejection the sensitivity of H. E. S. S. could be improved by a factor of two when compared to the Hillas based reconstruction.

3.4 γ-ray data analysis

Reconstructed particle properties comprise uncertainties that are dealt with on a statistical basis in the analysis of γ-ray data. Besides reconstruction uncertainties, selection effects have to be taken into account as well, for instance at the lower energy threshold the selection bias towards showers with an up-fluctuating intensity distribution. In this section, the steps involved to yield high-level analysis products like flux maps and source spectra are outlined.

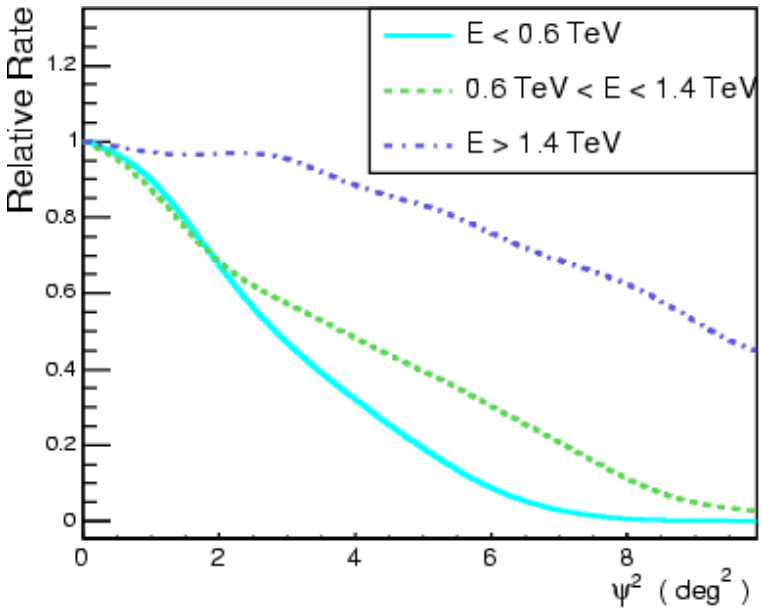


Figure 3.8: *Acceptance of the background as function of the squared angular distance to the pointing position (offset). The so-called acceptance curve is shown for different energy ranges of the primary particle. Credit: [121].*

3.4.1 Instrument response

Instrument response functions (IRFs) relate observed quantities with real ones and are mostly derived from simulations. A principal characteristic is the acceptance of the instrument, which describes the probability of shower events of a certain type to pass selection criteria. These criteria include both trigger and analysis thresholds. Next to these threshold and the analysis applied, the acceptance has many dependencies, most notably the type of the primary particle, its energy and its origin within the field of view, the zenith angle and the NSB rate. The acceptance is calculated for both actual γ-rays and the residual background, i.e. cosmic-ray events passing selection cuts. For γ-rays, it is derived from simulations and used to calculate the true number of incident particles for a given number of observed events. For cosmic rays, precise knowledge of the relative acceptance across the field of view is of importance for the background estimation (see below), but cannot be easily calculated from simulations. This is due to the efficiency of background suppression, which means that the statistics needed to calculate the acceptance accurately involve high computational costs. Instead, the relative acceptance for cosmic rays is commonly derived directly from data, e.g. using regions free of known sources of γ-ray emission. An example of the acceptance curve, i.e. the acceptance as a function of the squared angular distance to the pointing position, is shown in Fig. 3.8. Showing the curves for different energy ranges as in this figure, the energy dependence becomes clear.

Furthermore, to calculate fluxes, the area covered by the instrument must be determined. The area covered by IACTs is much larger than the mirror area of the individual telescopes, but varies depending on the acceptance. In practice, an energy-dependent effective area is derived from simulations. To do this, γ-rays are simulated to cover an area A_{sim} approximately equal to the maximum detection area of the instrument ($\sim 1\,\mathrm{km}^2$ for H. E. S. S.). The

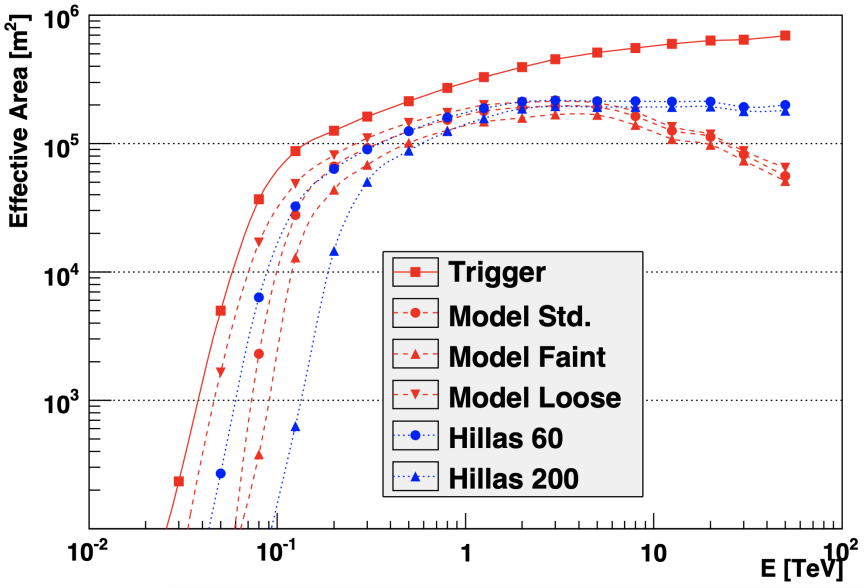


Figure 3.9: *Effective area as function of energy, shown for different analyses and configurations. Credit: [119] (see there for more information).*

ratio of detected events to simulated events, $\epsilon(E) = \frac{N_{det}(E)}{N_{sim}(E)}$ then gives the effective area $A_{eff}(E) = \epsilon(E) A_{sim}$, similar to that shown in Fig. 3.9.

The true energy of real γ-rays is not known, but is estimated by a reconstruction algorithm. The corresponding uncertainty of the reconstruction leads to an energy dispersion, which must be taken into account in the calculation of any physical quantity. With the help of simulations, the energy dispersion is derived from the distribution of true energies over the reconstructed energies of the simulated events. By normalising the corresponding two-dimensional histogram, a conversion matrix $\mathcal{R}(E_{true}, E_{reco})$ is obtained. This matrix allows the calculation of the true energy distribution using the observed distribution of reconstructed energies and vice versa (depending on the normalisation of the matrix).

Similarly, the point-spread function (PSF) φ_{PSF}, is caused by the uncertainty of the reconstructed direction of γ-rays. It is derived from the reconstructed spatial distribution of γ-ray events for a simulated point source. Typically, the PSF is described by an analytical function fitted to the distribution of reconstructed distances from the source position. A more sophisticated analysis may instead include the full two-dimensional distribution to account for inhomogeneities caused, for example, by NSB. The PSF is energy dependent and varies with both zenith angle and offset.

Finally, the observation time T_{obs} is also needed to calculate fluxes. While standard observations with H. E. S. S. are conducted in blocks of 28 minutes, so-called runs, the effective observation time per run is somewhat less. This is due to dead time, i.e. the time during which individual telescopes or the entire system are unable to acquire new events because they are occupied with processing a previous event. Since the acquisition of images does not affect the trigger, the dead time can be derived directly from observations by comparing the trigger rates with the detection rates.

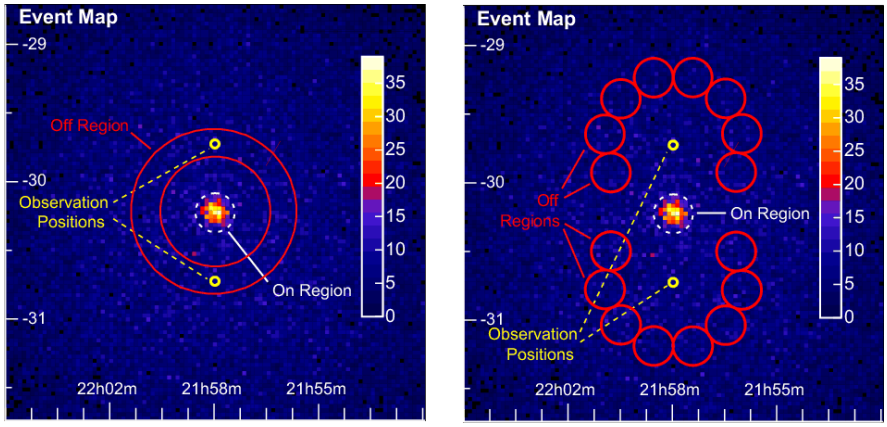


Figure 3.10: *Background-estimation techniques. Left: Ring background method. The off-region is defined as a ring around the source, independently of the pointing positions of the individual runs in the data set. Right: Reflected regions method. Off-regions are determined per run at the same offset as the on-region. Details are given in the text. Credit: [121].*

3.4.2 Background estimation

Background estimation is a critical part of γ-ray analyses. Though there are performant reconstruction and classification algorithms available a residual background component in those events passing selection cuts is inevitable. This background comprises hadronic events, for instance sub-showers initiated by π^0 decays, and showers initiated by cosmic electrons/positrons that are almost indistinguishable from γ-ray events. Most methods derive a background estimate from data within the field of view, exploiting the fact that this component is isotropic. The basic idea is that for an on-source region suitable off-source regions can be found which are exposed to the same level of background as the on-source region but do not contain significant γ-ray emission. Then, the number of background events in the on-source region N_{bkg} can be simply calculated from the number of events in the off-source region N_{off} as $N_{bkg} = \alpha N_{off}$. The factor α accounts for differences between on- and off-regions, which can be their size, exposure or acceptance to cosmic-ray events. An overview and discussion of the different methods can be found in [121]. Here, the two methods relevant for this work are briefly summarised.

The first of these methods is the ring background method. Here the off-region is defined as a ring centred around the source as shown in the left panel of Fig. 3.10. The inner and outer radius of the ring are not predefined, but can be adapted to the specific requirements of the analysis. Exclusion regions, i.e. known regions of γ-ray emission, are cut out of the ring. An advantage of the ring background method is that it can be applied to stacked data sets, which is why it is often used to create sky maps. On the other hand, for the calculation of α it requires an accurate determination of the acceptance, which can vary not only between the on- and off-region, but also within the ring itself. Extracting the background from regions with different offsets introduces an energy dependence that is not easily corrected (cf. Fig. 3.8). Therefore, spectral analyses are usually performed using other techniques for background estimation, such as the reflected region method.

With the reflected-region method, the background is estimated run by run from suitable

off-regions. For a given run, the off-region consists of several regions that have the same shape as the on-region and are arranged on a ring around the pointing position. Thereby, the radius of the ring is determined by the distance of the on-region to the pointing position. An example of the reflected region method is shown in the right panel of Fig. 3.10. Regions that overlap with exclusion regions are also discarded. Since for any given run the acceptance is approximately azimuthally symmetric and both the on- and off-regions are subjected to the same exposure, α is simply the ratio between the size of the on- and off-regions. In particular, this method does not require any correction for the energy dependence of the acceptance.

3.4.3 Source detection & modelling

Since measurements at very-high energies are dominated by the charged background, statistically robust measures are needed to assess signals from potential sources. Usually, a source is said to be detected if the measured signal within the source region deviates from the background estimate with a significance $S \geq 5\sigma$. For photon counting experiments, in particular for those used in ground-based astronomy, a model-independent estimate of the significance was derived by Li & Ma from the method of statistical hypothesis testing [122]

$$S = \sqrt{2}\left\{ N_{on} \ln\left[\frac{1+\alpha}{\alpha}\left(\frac{N_{on}}{N_{on}+N_{off}}\right)\right] + N_{off}\ln\left[(1+\alpha)\left(\frac{N_{off}}{N_{on}+N_{off}}\right)\right]\right\}^{\frac{1}{2}} . \quad (3.6)$$

Thereby, N_{on} and N_{off} are the photon counts measured in the on- and off-regions as defined, for example, in the previous section with the corresponding normalisation factor α. Alternatively, the significance may also be calculated for some emission model.

In general, an emission model describes the flux, for example of a source, as a function of energy E, position $\vec{r}$ and time t. Since temporal variability has only been measured for a few Galactic sources, the time dependence is neglected here. The model $\Phi(E, \vec{r}\,|\,\vec{\alpha})$ is determined by the parameters $\vec{\alpha}$, which include e.g. the position of the emission maximum, the size of the source or spectral parameters, etc. It is optimised by fitting the predicted photon counts to the observation. The prediction must take into account the response of the instrument. Thus, the expected number of photons n_γ with reconstructed energy E' at the position $\vec{r'}$ on the sky is calculated as

$$n_\gamma(E', \vec{r'}\,|\,\vec{\alpha}) = T_{obs}\int_0^\infty \mathrm{d}E \int \mathrm{d}\vec{r}\,\varphi_{PSF}(\vec{r}, \vec{r'}, E)\, A_{eff}(E)\, \mathcal{R}(E, E')\, \Phi(E, \vec{r}\,|\,\vec{\alpha}) \quad (3.7)$$

As described in section 3.4.1, the IRFs themselves depend on observational parameters that are not explicitly given in Eq. 3.7. For practical purposes, the photon counts are typically analysed in discretised bins in space and energy, so that for the i-th bin one obtains

$$N_{\gamma,i} = \int_{E_{min,i}}^{E_{max,i}} \mathrm{d}E' \int_{A_i} \mathrm{d}\vec{r'}\, n_\gamma(E', \vec{r'})\,. \quad (3.8)$$

In this way, the expected total number of photons in the i-th bin, $\widetilde{N}_i$, can be described by a combination of the expected background $N_{bkg,i}$ and an emission model, which may be an ensemble of m components

$$\widetilde{N}_i = N_{bkg,i} + \sum_{j=1}^{m} N_{\gamma,j,i}\,. \quad (3.9)$$

Since for each bin i the number of detected photons N_i is assumed to be an independent observable following a Poisson distribution, the Cash statistic [123]

$$C = 2 \sum_i \left(\widetilde{N}_i - N_i \ln \widetilde{N}_i \right) \tag{3.10}$$

is commonly used for estimating model parameters and hypothesis tests. For model parameter estimation, calculating $\arg\min_{\vec{\alpha}} C$ is equivalent to the likelihood maximisation. The test statistic, defined as

$$TS = C_0 - C_1 \,, \tag{3.11}$$

equals -2 times the natural logarithm of the likelihood ratio of the null hypothesis with q degrees of freedom (dof) yielding C_0 and the alternative hypothesis with $q+n$ dof yielding C_1. This test statistic follows a χ^2_n distribution with n dof in the limit of a large number of photon counts according to Wilks' theorem [124]. Correspondingly, the statistical significance, e.g. of a source model compared to pure background, is approximated with $\sqrt{|TS|}$.

3.5 Galactic plane survey

The third HGPS catalogue, published in 2018, is the culmination of a decade-long effort to obtain a comprehensive census of VHE γ-ray sources in the inner Galaxy. The main points of its compilation and its results are summarised below. For further details, the interested reader is referred to [88].

3.5.1 Catalogue generation

The survey was conducted from 2004 to 2013 and covers the Galactic longitude range $-110° \leq l \leq 75°$. The entire dataset includes dedicated survey observations focused on the latitude range $-1.8° \leq b \leq 1°$, as well as additional observations spent on specific regions that include sources or source candidates. After a selection according to several quality criteria, e.g. regarding the atmospheric condition, zenith angle or instrument condition, the analysed data set comprises $\sim 2700\,\mathrm{h}$ of observation time, considering only data from the H. E. S. S. phase I telescopes. A fixed procedure was used for the analysis of this data, which includes source detection and characterisation.

The main analysis was performed on sky maps of the studied region using a Cartesian projection and pixels of width 0.02°. In a first step, the background was estimated with an iterative procedure using the ring background method. Adapting to the extended nature of most Galactic sources and correspondingly large exclusion regions, the size of the ring was varied over the different test positions (pixels) to obtain reliable statistics. To prevent contamination of the off-regions with γ-ray emission, the exclusion regions were iteratively adjusted by adding a circle of 0.3° around regions of significant emission ($S \geq 5\sigma$ according to Eq. 3.6) until stable exclusion regions were obtained.

Subsequent analysis steps are based on models of the flux distribution in which the spectral and spatial components were modelled independently

$$\Phi(E, \vec{r}\,|\,\vec{\alpha}) = \phi(E\,|\,\vec{\alpha}_1)\,\Omega(\vec{r}\,|\,\vec{\alpha}_2)\,. \tag{3.12}$$

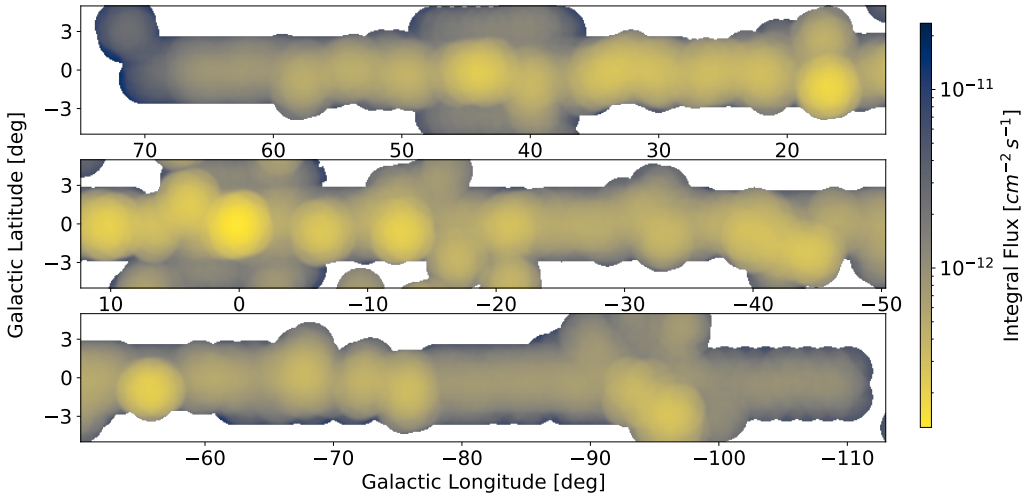


Figure 3.11: *Sensitivity map of the HGPS showing the minimal integral flux above* 1 TeV *contained within a circle of radius* $R_c = 0.2°$ *which is needed to be significantly detected.*

The spectral component was consistently modelled by a power law over the entire field of view

$$\phi(E \,|\, \phi_0) = \phi_0 \left(\frac{E}{E_0} \right)^{-\Gamma} \tag{3.13}$$

with fixed spectral index $\Gamma = 2.3$. From this, the integral flux F was calculated for $E \geq 1\,\text{TeV}$. Using a flat spatial distribution, i.e.

$$\Omega(\vec{r} \,|\, \vec{r}_0, R_c) = \mathcal{H}(R_c - |\vec{r} - \vec{r}_0|) \tag{3.14}$$

where $\mathcal{H}$ denotes the Heaviside function, $\vec{r}_0$ the test position and R_c the correlation radius, two sensitivity maps with $R_c = 0.1°$ and $R_c = 0.2°$ were calculated, giving the minimum flux needed to achieve $S \geq 5\sigma$ according to Eq. 3.6 at the respective pixel position. Thus, the map values correspond approximately to the detection threshold for a point source. The sensitivity map calculated for the correlation radius $R_c = 0.2°$[2] is shown in Fig. 3.11. The sensitivity distribution is primarily determined by the exposure of the different regions and accordingly exhibits considerable spatial variation. This deviation from spatial homogeneity is expected to contribute to the observational bias inherent in the sample of catalogue sources.

Prior to the source analysis, a large-scale emission model was extracted from data. It is used to account for diffuse emission and unresolved sources which otherwise may increase the rate of false source detections or affect the measurement of source extension. The spatial distribution of the large-scale emission was modelled by a one-dimensional Gaussian over Galactic latitude, whereby the amplitude, peak position and width of the flux distribution can vary over Galactic longitude. With a sliding window which covers 20° in longitude and the whole latitude range, the model parameters were fitted to the photon count distribution inside the window but outside exclusion regions using the Cash statistic.

[2] `https://www.mpi-hd.mpg.de/hfm/HESS/hgps/data/hgps_map_sensitivity_0.2deg_v1.fits.gz`

For the detection and characterisation of sources, a two-dimensional Gaussian model was used for the spatial component of the flux distribution

$$\Omega(\vec{r}\,|\,\vec{r}_0,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\vec{r}-\vec{r}_0)^2}{2\sigma^2}} \,. \tag{3.15}$$

Together with the background estimate and the large-scale emission model, the predicted counts per pixel is given by

$$\widetilde{N} = N_{bkg} + N_{LS} + \sum_{j=1}^{m} N_{\gamma,j} \,, \tag{3.16}$$

where $N_{\gamma,j}$ results from individual Gaussian components fitted to the peaks of the emission in the analysed region. The procedure described in section 3.4.3 was used to fit and select the model components. A Gaussian component is included in the list of candidate sources if the test statistic gives a value $TS \geq 30$ compared to the emission model without this component. Isolated components are considered as single γ-ray sources. Overlapping Gaussian components, on the other hand, were manually merged if their peaks are not clearly separable. Then the product of the merging is considered as an individual γ-ray source, whose total flux is calculated from the sum of the fluxes of the merged components. The squared size of the composite source, σ^2, was calculated from the flux-weighted average of the squared sizes of the individual components. Following this approach, regions that include known sources with distinctive morphologies not well described by a Gaussian model, such as the Galactic Centre or shell-type SNRs, were excluded from the source analysis. These sources are still listed in the catalogue, but labeled as *extern* together with additional known sources that were not significantly detected in the HGPS data set.

For all sources derived from the morphology fit, a separate spectral analysis was carried out, which methodologically differs from the map-based procedure described above. However, since the results of this spectral analysis are of little relevance to this work, the procedure is only briefly outlined here for the sake of completeness. A circular region around the source was chosen to extract the spectrum. The counts measured in this region were corrected for possible contamination from nearby sources and large-scale emissions, as well as for a difference in size between the analysis region and the extent of the source. The background within the analysis region was estimated using the reflected regions method. A likelihood maximisation method as described in [125] was then used to fit two spectral models to the source, a power law (Eq. 3.13 with Γ as additional free parameter) and an exponential cut-off power law

$$\phi(E\,|\,\phi_0,\Gamma,E_c) = \phi_0 \left(\frac{E}{E_0}\right)^{-\Gamma} e^{-\left(\frac{E}{E_c}\right)} \,. \tag{3.17}$$

It is shown that the values of the integral flux—which is the relevant quantity for this work—from the map-based estimation are in good agreement with those of the best-fit spectral model (see Fig. 9 in [88]).

3.5.2 Source sample

In total, the HGPS catalogue contains 78 VHE γ-ray sources, of which 16 sources were not significantly detected prior to the survey. For all sources detected in the survey, multi-wavelength information (MWL) was used to establish associations with physical objects.

Specifically, catalogues of known classes of Galactic VHE γ-ray emitters were searched for sources whose positions matches any HGPS source. Among the catalogues searched are the 2FHL [126] and the 3FGL catalogue [127], which contain high-energy sources detected with the *Fermi* Large Area Telescope (*Fermi*-LAT). Furthermore, SNRcat [128] was used, which summarises information on Galactic SNRs and PWNe from observations primarily in the radio and X-ray range, but also from the high-energy range. Finally, the ATNF catalogue [129] was also searched for energetic pulsars. Due to the large point spread function and the extended nature of most HGPS sources, many sources have multiple associations at once. In these cases, additional criteria were used for a firm identification of the corresponding physical source. Depending on the type of source considered, these criteria are matching MWL morphologies (e.g. shell-type SNRs), energy-dependent morphologies (PWNe) or temporal-spectral variability (binaries). In total, the physical nature of 31 sources was determined in this way. PWNe turned out to be the dominant class with 12 identifications. Also 8 sources were identified as composites and are thus additionally associated with PWNe. Furthermore, 8 shell-type SNRs and 3 binaries were identified with confidence. For the majority of sources, either no physical object could be associated (11 sources) or it was not possible to identify the physical origin from several associations based on the applied criteria and the available information. In this context, it was also discussed that at least some of these sources are probably confused, i.e. the observed emission is the combined result of several spatially coincident VHE emitters.

For none of the classes do the corresponding sources show a distinct characteristic in the distributions of the measured quantities integral flux, spectral index and angular extent, with the exception of the binaries, which are exclusively point-like. Most sources (71 %) are extended with a maximum of 1.0° (Vela Junior). Their fluxes cover a range of two orders of magnitude, from $1.25 \cdot 10^{-13}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ to $2.34 \cdot 10^{-11}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. When fitted with a power law, the spectral index of the sources is on average $\Gamma = 2.4 \pm 0.3$, which is compatible with $\Gamma = 2.3$ as used for the map-based analysis of the HGPS data. With knowledge of the physical origin, distance estimates could be obtained from MWL observations for 29 of 31 firmly identified VHE sources. This additional information from the catalogue is used for modelling the Galactic population of VHE γ-ray sources described in chapter 6.

4 Deep-learning techniques

In this work, deep-learning (DL) methods, more specifically convolutional neural networks (CNNs), have been used to approach a set of different problems. This chapter gives a brief overview of the DL techniques which are relevant for this work. The main resource for this chapter has been the monograph by Goodfellow et al. [130]. For a more detailed introduction to the subject, this book is recommended to the interested reader.

DL is a domain of machine learning (ML), which generally deals with computer algorithms that automatically learn from examples how to perform certain tasks. Such algorithms are useful when writing explicit instructions is difficult or impossible for the human programmer. DL is specifically about the application of artificial neural networks (ANNs) and has become increasingly popular in recent years. That is because DL allows constructing complex computer models which, especially for tasks on high-dimensional data, surpass the performance of other ML approaches. State-of-the-art performances are achieved with DL on tasks like computer vision, natural language processing or audio signal processing [131].

On an abstract level, the architecture of DL algorithms is similar to those of other ML algorithms. The algorithms are composed of a data set, a model, a loss function and an optimiser [130]. The model, i.e. the neural network, is trained on examples from the data set. The loss function gives a measure of the discrepancy between the output of the model and some intended output for the given examples. By iteratively adjusting the parameters of the model an optimiser aims at minimising the loss and, thus, maximising the performance of the model. ML algorithms are commonly categorised into supervised learning, unsupervised learning and reinforcement learning. In supervised learning the algorithm is trained on pairs of input and desired output values. Classification models, for instance, are commonly trained on a set of images and corresponding labels. In unsupervised learning, the algorithm is asked to learn descriptive features of the data it is examining. Autoencoders are used, for example, to find a low-dimensional representation of the data that allows the recovery of the original input. Instead of using pairs of input and output values for training, there the input itself becomes the desired output. Reinforcement learning is different from the other two learning schemes as the algorithm interacts with a dynamical environment and seeks to maximise the rewards it receives for the correct behaviour [132].

In this work, supervised learning algorithms have been used exclusively. Thus, unsupervised and reinforcement learning are not covered in the following description of DL algorithms. Also the description of the data sets is postponed to later chapters covering the individual problems that have been addressed with DL.

4.1 Models

Assume there is a function f^* that relates an input $\mathbf{x}$, e.g. an image, to a target y, e.g. the label of this image, such that $y = f^*(\mathbf{x})$. The exact form of f^* is unknown, but is to be inferred from data. The model of a ML algorithm then refers to a parametric function $f(\mathbf{x};\theta)$ whose parameters θ are optimised during the learning process to approximate the objective function $f(\mathbf{x};\theta) \approx f^*(\mathbf{x})$. Different ML algorithms differ in their construction of $f(\mathbf{x};\theta)$ and accordingly in the way they optimise these models. The capacity of a model, i.e., in simple terms, its ability to accurately approximate any complicated function $f^*(\mathbf{x})$, typically increases with the number of parameters. The type of function it can approximate also depends essentially on the structure of the model. Especially for high-dimensional input data, ANNs and their variants are suitable models and provide a flexible architecture that can be adapted to address many different types of problems.

4.1.1 Artificial neural network

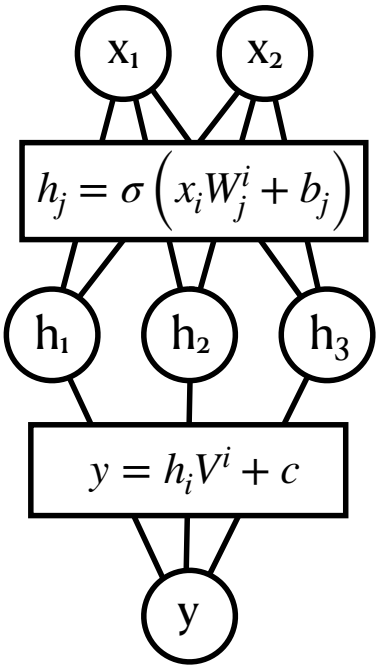


Figure 4.1: *Schematic of a fully connected neural network with one hidden layer.*

The concept of artificial neural networks takes up an idea from neural science, which sees a neuron as a processing and distribution unit for information. A network is formed by many of such interconnected units. Each neuron assigns different weights to its input connections and is activated if the incoming signal corresponds to a specific pattern. When activated, a response is distributed to other neurons in the network, which again filter the signal for specific patterns. While patterns filtered by a single neuron might be simple, within the network more complex patterns arise from the combination of those simple patterns.

Simple ANNs are constructed from layers of neurons. Each neuron of one layer is connected to every neuron in the adjacent layers. Such a network is referred to as a fully connected network (FCN). The mechanism defining a neuron is realised by a combination of linear transformations and activation functions. Beginning with an input vector x_i, where each element i represents the state of a neuron in one layer, the state of the neurons in the next layer of the network h_j is calculated using Einstein notation as

$$h_j = \sigma\left(x_i W_j^i + b_j\right) . \tag{4.1}$$

The elements of the transformation matrix W_j^i and the bias vector b_j are parameters of the model that need to be optimised. The activation function σ introduces a non-linear response for each neuron. Without it, the model could learn to approximate linear functions only. Common examples of such activation functions are hyperbolic tangent or the sigmoid function. The prevailing activation function, however, is the rectified linear unit (ReLU), which is

simply the maximum operation $max(0, x_j)$ [133]. As emphasised above, ANNs are composed of layers of neurons in order to approximate more complex functions. Following the hidden layer h_j, the network is composed of at least one additional layer, the output layer. Using the term hidden layer reflects that the values of h_j are not communicated to the user of the model, who only sees the result of the output layer. The output values of the network are usually calculated by a linear transformation only. In Fig. 4.1 a schematic of a simple two-layer ANN is shown, which is designed to approximate a function $f^* : \mathbb{R}^2 \to \mathbb{R}$. It consists of one hidden layer with three neurons and an output layer with a single neuron which yields the estimate for $f^*(\mathbf{x})$.

The number of neurons per layer and the number of layers define the width and depth, respectively, of the neural network. While any network with three layers is already considered being deep, models that yield state-of-the-art performances on challenging benchmark data sets comprise several, e.g. tens to hundreds, of hidden layers, each with a number of neurons much larger than given in the example in Fig. 4.1 (see e.g. [134]).

4.1.2 Convolutional neural network

CNNs are a variant of neural networks that is very efficient on grid-like data such as images [135]. Those networks comprise at least one layer were the matrix multiplication from the previous section is replaced by a convolution operation. This operation can be thought of as applying a filter on a sliding window over the data to search for local patterns in the input signal.

Mathematically the convolution of two functions, f and k, is defined as

$$(f * k)(x) = \int \mathrm{d}t f(t) k(x - t)\,. \tag{4.2}$$

In DL terminology the function f is commonly referred to as input, the function k as kernel and the output of the convolution as feature map. When working with CNNs, Eq. 4.2 needs to be adapted for a discretised domain space. Thus, the convolution becomes a tensor operation. In the case of images, the input can be considered a three-dimensional tensor $I_{i,m,n}$, where the index m refers to the pixel row, n to the pixel column and i to the input channel, e.g. the three color channels red, green and blue. With a four-dimensional kernel $K_{o,i,r,s}$ each convolution produces a multi-channel feature map. The size of the kernel is typically much smaller than the input. The indices for its rows and columns, r and s respectively, are numbered such that both indices are zero at the kernel's central unit. Its input channel i correspond to that of the input data, the output channel o to that of the feature map. This way, the feature map $Z_{o,m,n}$ is obtained as

$$Z_{o,m,n} = \sum_{i,r,s} I_{i,m+r,n+s} K_{o,i,r,s}\,, \tag{4.3}$$

where the summation takes place solely over valid input indices. An illustration of a CNN is shown in Fig. 4.2. There, the input to the network is a three-dimensional array that is 22 pixel wide, 22 pixel high and has one input channel. In the first layer of the network, this input is convolved with a 3×3 kernel that has four output channels. The result is a feature map that is 20 pixel wide, 20 pixel high and has four channels. Since only valid input indices

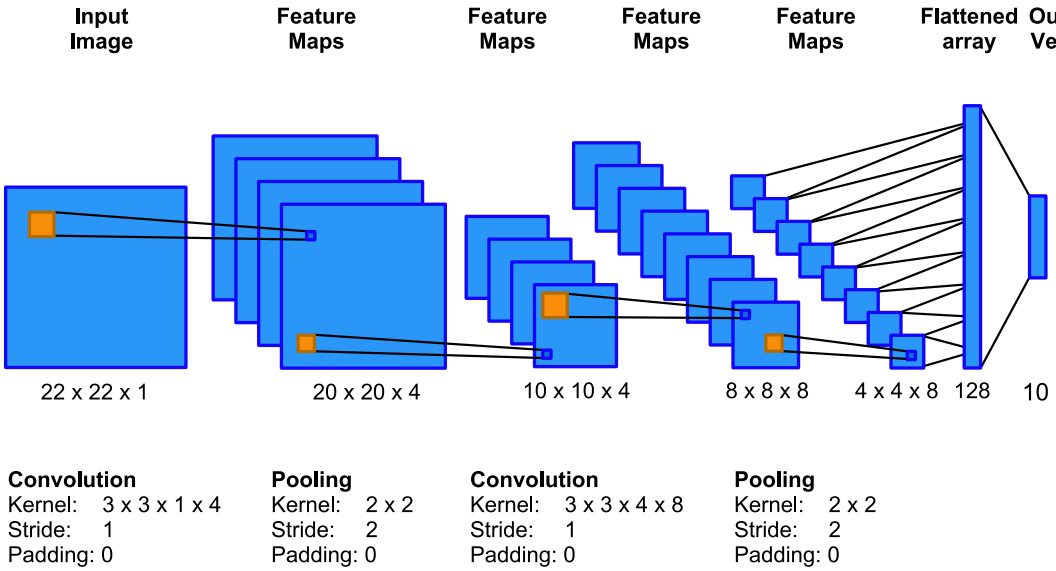


Figure 4.2: *Schematic representation of a convolutional neural network: An image with one input channel is iteratively convolved and pooled with kernels of different sizes, producing feature maps with iteratively decreasing spatial dimensionality and increasing dimensionality in feature space. After the final pooling layer, the feature maps are flattened into a single feature vector that is processed by a fully connected layer to produce a final output vector.*

are allowed in Eq. 4.3, the feature map is reduced in size by the kernel size minus one in each direction. In order to obtain feature maps of the same size as the input, the latter can be padded, e.g. with zeros, prior to the convolution.

Following the convolution, Fig. 4.2 shows the application of a pooling layer. Pooling is a common operation used in CNNs to reduce the size of feature maps, which can otherwise become very large as convolutional layers often increase the number of channels in the feature map. Pooling can be thought of as moving a kernel over the feature map in each of its channels and calculate a local summary statistic for the area covered by the kernel. The summary statistic might be, for example, the sum, the average or the maximum value. By choosing a stride, i.e. the step size with which the kernel is moved, larger than one the size of the feature map is reduced while at the same time local patterns are preserved by the pooling.

After a set of convolution and pooling layers, many CNNs use fully connected layers as described in the previous section as their final layers. Therefore, the elements of the last feature map are flattened into a one-dimensional vector. Often the convolution part and the fully connected part are implemented independently. This structure facilitates the reuse of a pre-trained CNN for new tasks since it is mostly sufficient to only train the much smaller fully connected layers for a new task when the data is similar to the one the CNN has been trained on previously.

The strengths of CNNs are sparse interactions, parameter sharing and translation invari-

ance. Local features are usually restricted to small areas of the input space, especially lower level features such as edges. This means that information outside a given area is irrelevant for the detection of the feature within that area. This fact is exploited by choosing convolution kernels that are smaller than the size of the input. This means that they process only a fraction of the total input data at once, which is called sparse interaction. Moreover, unlike FCNs, where there is one trainable parameter for each pair of neurons between two consecutive layers, in CNNs each trainable parameter is used once for (almost) every such pair of neurons. Because of this parameter sharing, CNNs generally require fewer trainable parameters compared to competitive FCNs. The benefits of this are usually associated with faster training and a less demanding size of the training data set. Finally, since each kernel is moved over the entire input space, CNNs are inherently translation invariant. This is a useful feature for many tasks, such as object detection, where it is important to know whether an object is present or not, but not necessarily where it appears in the input space.

4.2 Loss function

Loss functions measure the discrepancy between the output of the model $y = f(\mathbf{x}, \theta)$ and a target $\hat{y}(\mathbf{x})$. In order to maximise the performance of the model, an optimiser tries to minimise the loss by adjusting the parameters of the model accordingly. The manifold in the parameter space that the optimiser has to probe is determined by the function used to calculate the loss. Consequently, the correct choice of an appropriate loss function for the given task is important for the success of the training of the model.

Suitable loss functions can be derived from the maximum likelihood principle. Cross-entropy is the most common loss function used for classification tasks. The cross-entropy between the data-generating distribution, $\hat{p}_{data}(\mathbf{x})$, and the model distribution, $p_{model}(\mathbf{x}; \theta)$, is defined as

$$H\left(\hat{p}_{data}, p_{model}\right) = -\mathbb{E}_{\mathbf{x}\sim\hat{p}_{data}} \hat{p}_{data}(\mathbf{x}) \log p_{model}(\mathbf{x}; \theta)\,. \tag{4.4}$$

By comparison it can be seen that minimising the cross-entropy corresponds to minimising the negative log-likelihood

$$\theta_{ML} = \arg\min_{\theta} -\mathbb{E}_{\mathbf{x}\sim\hat{p}_{data}} \log p_{model}(\mathbf{x}; \theta) \tag{4.5}$$

since $\hat{p}_{data}(\mathbf{x})$ is independent of θ. Typically, a one-hot encoding is used for classification tasks. That means, for each example i the true class values are given by a vector $\hat{\mathbf{y}}^{(i)}$ of length N_{class} with all elements set to zero except for the element with index c corresponding to the class of the example, which is set to one. The output of the model is a vector $\mathbf{y}^{(i)}$ of the same length. Applying the softmax function

$$\sigma(\mathbf{y})_j = \frac{e^{y_j}}{\sum_{k=1}^{N_{class}} e^{y_k}} \tag{4.6}$$

results in a vector whose elements can be interpreted as the probabilities that the model attributes to each class. Thus, with the definition given in Eq. 4.4, the cross-entropy loss L_{CE} for m samples is calculated as

$$L_{CE} = -\frac{1}{m} \sum_{i=1}^{m} \log \left(\frac{e^{y_c^{(i)}}}{\sum_{k=1}^{N_{class}} e^{y_k^{(i)}}} \right)\,. \tag{4.7}$$

For regression tasks, the mean squared error (MSE) is virtually the default loss function. It is defined as

$$L_{MSE}(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^{m} \left\| y^{(i)} - \hat{y}^{(i)} \right\|^2 . \tag{4.8}$$

The MSE can be derived from the maximum likelihood principle, assuming the true values have Gaussian errors, i.e. $\hat{p}_{data}(\hat{y}|\mathbf{x}) = \mathcal{N}\left(\hat{y}; y, \sigma^2\right)$. The negative log-likelihood of m independent and identically distributed samples from a normal distribution is

$$-\log(\mathcal{L}) = -m\log(\sigma) - \frac{m}{2}\log(2\sigma) - \sum_{i=1}^{m} \frac{\left\| \hat{y}^{(i)} - y^{(i)} \right\|^2}{2\sigma^2} . \tag{4.9}$$

A comparison between Eq. 4.8 and Eq. 4.9 reveals that under the assumption of normal distributed errors the minimisation of the MSE as well corresponds to the minimisation of the negative log-likelihood. One alternative to MSE that is commonly used for regression tasks is the mean absolute error (MAE)

$$L_{MAE}(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^{m} \left| y^{(i)} - \hat{y}^{(i)} \right| . \tag{4.10}$$

This loss function is more robust to outliers, i.e. predictions far from the targets have a less strong impact on the optimisation of the model parameters.

4.3 Optimiser

Most DL models are optimised using stochastic gradient descent (SGD) or variants thereof. Gradient descent means the parameters of the model are updated according to the gradient of the loss function with respect to these parameters

$$\theta \leftarrow \theta - \epsilon \nabla_\theta L\left(f\left(\mathbf{x}, \theta\right), \mathbf{y}\right) \ [1]. \tag{4.11}$$

Thereby, the gradient determines the direction of the update, while the step width of the update is determined by the learning rate ϵ. In order to calculate the gradient, at first, the output of the model $f\left(\mathbf{x}, \theta\right)$ is calculated by propagating the information provided by the input $\mathbf{x}$ forward through the network. Using the chain rule, the gradients are then calculated recursively backwards for all layers of the network starting from the output layer. This task is solved with the back-propagation algorithm (backprop) [136].

In addition, SGD acknowledges that the calculated gradient is an estimate. Most loss functions used in DL can simply be summed over the number of examples, such as the negative log-likelihood. Therefore, the gradient can be well approximated via the average loss calculated over a random sample of input and target values from the dataset. Such a sample is referred to as minibatch. Iterating through the data set in minibatches allows drastically shorter training times in comparison to processing the entire data set at once, while the approximation error of the gradient is modest for typical batch sizes of $\mathcal{O}(100)$.

[1] Read $a \leftarrow b$ as a becomes b.

With deep neural networks the optimiser needs to probe a complex, high-dimensional, non-convex parameter space. There are several aspects on which the success of the optimisation depends. One critical aspect is the learning rate, which initially is set by the user as a hyper-parameter of the model. When the training progresses the learning rate should be adjusted, for instance reduced in order to reach a (local) minimum. More sophisticated variants of SGD introduce moments [137] or automatically adapting learning rates for each individual parameter in order to improve the optimisation process. Among the most popular of these variants are AdaGrad [138], ADADELTA [139] and Adam [140].

Next to the optimiser, several design choices for the network help to improve the optimisation. One issue of deep neural networks is that of vanishing or exploding gradients. The right choice of the activation function is one way to address this issue. For instance, ReLU provides a better alternative to the sigmoid function which easily saturates, thus causes vanishing gradients. Also, the initialisation of the model parameters has significant impact on the success of the optimisation [141]. Batch normalisation, i.e. normalising the input to each layer of the network, has become the standard to address the issue of internal covariate shift [142]. Furthermore, while the depth of a network appears to be critical for learning complex distribution functions, very deep networks can experience a degradation of performance because of vanishing gradients. Residual networks that apply shortcuts with an identity mapping, for example such that Eq. 4.1 becomes

$$h_j = \sigma\left(x_j + x_i W_j^i + b_j\right) , \tag{4.12}$$

have been shown to mitigate this issue and allow for the optimisation of networks with more than a thousand layers [134, 143].

4.4 Training & evaluation

A critical part for any ML algorithm is to ensure that the model generalises well to new, previously unseen input. Typically, the data set used for training is limited and processed iteratively while optimising the model. If the capacity of the model is larger than necessary to fit the data-generating distribution, the model may learn to reduce the loss by relying on features that are unique to that data set instead of using general features of the distribution. This case is referred to as overfitting and manifests in a large generalisation gap, that is the difference between the loss calculated on the set of training examples and another, disjoint set from the same data-generating distribution. The opposite case of overfitting, i.e. underfitting, occurs with models of lower capacity than necessary to fit a given distribution. In that case, the model will not achieve the targeted level of performance.

The aim of ML is to reach low generalisation errors (small loss on unseen samples) which is why this quantity should be monitored during training. For this purpose, the data set is split into a training set and a validation set. While only the training set is used for the optimisation of model parameters, the validation set is used to estimate the generalisation error. The average loss on both sets is calculated for each epoch, i.e. a complete iteration of the data set. The most common case with ANNs is overfitting which gives rise to learning curves similar to the one displayed in Fig. 4.3. There, the learning curves showing the losses

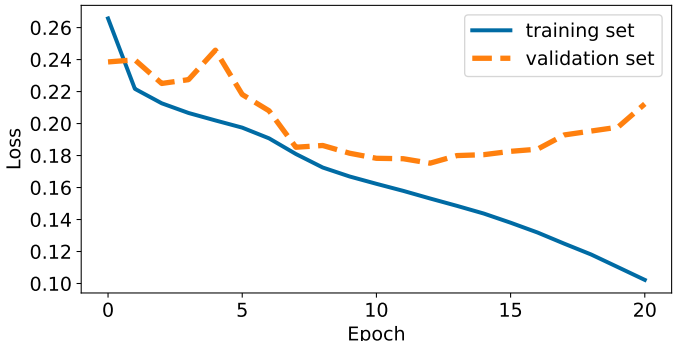


Figure 4.3: *Learning curves showing the losses calculated for the training set and the validation set after each epoch. In the first epochs, the validation loss follows the training loss, except for a short interruption in epoch four. Such interruptions have been observed frequently throughout this work when batch normalisation is applied. After six epochs, the two learning curves diverge continuously due to overfitting.*

calculated on the training set and the validation set begin to diverge at epoch six, when the network starts to increasingly rely on unique features of the training set.

Regularisation is a means of reducing the generalisation error. One prominent regularisation technique is weight decay [144]. Extending Eq. 4.11 to

$$\theta \leftarrow (1-\lambda)\theta - \epsilon \nabla_\theta L\left(f\left(\mathbf{x},\theta\right),\mathbf{y}\right) \ , \tag{4.13}$$

each update pushes the parameters of the model (weights) closer to zero by a factor controlled with the decay parameter λ. This effective generalisation method is assumed to reduce the complexity of the model while preventing it from continuously increasing weights that align with specific patterns. Another prominent technique is dropout [145]. This involves randomly selecting and blocking units and their connections between successive layers during each training step. To a certain extent, this also prevents the network from focusing too much on too specific patterns. Finally, it is common practise to use the parameter configuration that yields the smallest validation loss at run time.

Adjustments of hyperparameters like the learning rate during the training of the model are usually based on the progress of the validation loss. Hence, the estimation of the generalisation error calculated on the validation set becomes biased. An unbiased estimation of the generalisation error is calculated on yet another data set, the test set, with truly unseen samples. However, often the performance of the model at run time is still lower than estimated based on the test set. This is because the target domain is rarely identical to the training domain and ANNs are notoriously susceptible to small differences between the two, e.g. the difference between simulations and reality. The fields that are concerned with this kind of problem are domain adaptation and transfer learning (see e.g. [146] for an introduction).

5 γ-hadron separation with deep learning

5.1 Motivation

By now machine learning is a standard tool for data analysis. In ground-based γ-ray astronomy, it has been used with great success for more than a decade to separate cosmic-ray and γ-ray showers [116, 147, 148]. Boosted decision trees (BDTs), for example, exploit correlations among a number of variables that have high discriminatory power and are therefore more efficient at separating cosmic rays and γ-rays than simple box cuts on these variables. With the advent of deep learning, new tools are available that promise the next leap in performance. Since the data in ground-based γ-ray astronomy is strongly dominated by a background of hadron-induced showers, the impact of the γ-hadron separation algorithm on the overall performance of the instrument and in particular on its sensitivity is significant. Improving the γ-hadron separation algorithm as a means of lowering the detection threshold for γ-ray sources has the potential to increase the sample of known sources and thus our knowledge of the underlying source population. To this end, this work also investigates the performance of DL in γ-hadron separation.

The transition from traditional ML algorithms such as BDTs to DL algorithms represents a paradigm shift. The former rely on human-engineered features to solve a given task. In the case of ground-based γ-ray astronomy, for example, the discriminative power of certain variables, such as the reduced scaled width and length of a shower (see Sec. 3.3.4), can be derived from the physics of the shower and such variables with high discriminative power made available to the ML algorithm. In this way, the algorithm's performance is primarily limited by the human programmer's knowledge of, and ability to access, appropriate discriminative variables. With DL, on the other hand, it is common to provide the algorithm directly with the raw data, e.g. an image, rather than summary statistics as mentioned above, and let the algorithm autonomously derive those attributes that offer the greatest discriminatory power. In this way, the (potential) limitation of traditional ML algorithms is removed and a wider range in the feature space is made accessible.

With neural networks there are many ways how a specific task can be approached. In this work, the focus is on convolutional neural networks, which are the first choice for processing image-like data as generated by IACTs. Using a selection of different models, the potential of CNNs as a tool for γ-hadron separation and the impact of different designs on performance are discussed below. Before that, the next section describes the properties of the data set used to train the CNNs and in particular how the specific data format of the telescope images is handled.

5.2 Data format & data set

Images from IACTs have an unusual format in that their pixels are typically arranged on a hexagonal grid. This layout is particularly cost-effective, i.e. it yields the densest packing on a two-dimensional grid, and offers additional advantages for data analysis such as optimal sampling of circularly band-limited signals and superior angular resolution [149, 150]. On the other hand, the hexagonal layout can be an obstacle for the use of public analysis tools such as the popular DL libraries TensorFlow [151] or PyTorch [152], which are designed to process data sampled on a rectangular grid. Several solutions to this problem have been proposed, most of which focus on converting the data into the correct format, e.g. by interpolation, rebinning or resampling [153, 154]. For a more direct access to the data, an extension to the DL framework PyTorch has been written with significant participation of the author of this thesis, which allows to perform convolution operations on a hexagonal grid, thus taking advantage of the hexagonal layout in the context of applying CNNs [155]. This extension still requires that the data conform to the rectangular layout in order to take advantage of the framework's highly optimised routines. But instead of transforming the data, a special data addressing scheme is applied. Practically, this can be thought of as splitting the hexagonally sampled data into two rectangular arrays, e.g. taking all the even columns and all the odd columns, and shifting the two arrays against each other to bring them into alignment. An example of an image of a H.E.S.S. I telescope before and after applying this addressing scheme is shown in Fig. 5.1. Note that, on the right-hand side, zero-valued pixels are padded to the edges of the camera image in oder to match the rectangular layout. This scheme is the most efficient in terms of data storage given the layout of the cameras and is used throughout the following. For detailed information, see [155].

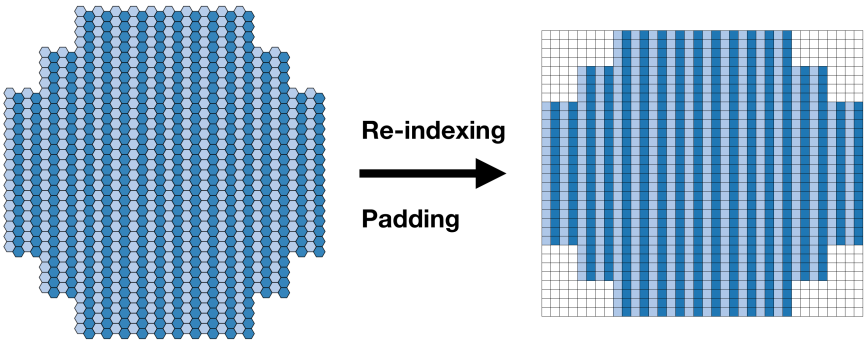


Figure 5.1: *Re-indexing scheme for telescope images. Left: Original hexagonal layout. Right: New rectangular layout after re-indexing. Even columns (dark coloured) are moved up to be aligned with the odd columns (light coloured). Zero-valued pixels (white) are padded at the corners to fill the rectangular shape.*

The study presented here is limited to the analysis of H.E.S.S. I data, i.e. only data acquired with the four 12-m telescopes. For the training and evaluation of the different CNN models, sets of simulated shower images generated with CORSIKA and sim_telarray were used. Unless otherwise stated, each model was trained on 10^6 shower events and validated on $2.5 \cdot 10^5$ shower events. The final performance of the models was calculated on a test set comprising 10^6 simulated shower events. For all sets, the shower events originate equally from γ-rays or protons. γ-rays were simulated in the energy range 20 GeV-100 TeV according to a

power-law distribution with index $\alpha = -2.0$. Similarly, protons in the range 100 GeV-200 TeV were simulated. Note that 20 GeV events are below the energy threshold of the array if it exclusively consists of the smaller telescopes. However, the simulated array includes CT 5 and, thus, has a lower energy threshold. After removing CT 5 from the data, for a fraction of these low-energy events information is still available in images from the smaller telescopes. Although these events are unlikely to trigger the array independently of CT 5, viable information for the analysis of H.E.S.S. I data may still be extracted which is why they were kept in the data set. The direction of arrival of the simulated particles is isotropic within a cone of 2.5° around the pointing position at 180° azimuth and 70° altitude. To account for changes in the optical efficiency of the telescopes over time, e.g. due to a degradation of the reflectivity of their mirrors, there are simulations for different temporal periods that reflect the respective state of the telescopes. The simulations used in this work are internally referenced as phase2b2. Two-stage image cleaning with thresholds of 10 p.e. and 5 p.e. is applied to all shower images. In addition, all images have a telescope multiplicity > 2 after cleaning, which is the only preselection requirement.

5.3 Model architectures

For the purpose of γ-hadron separation, the aim is to construct a model that uses images of the same air shower taken by several telescopes as input and on this basis gives a classification score that allows the user to decide whether the shower is more likely to have been triggered by a γ-ray or by a cosmic ray. With neural networks, there are infinite possibilities to realise such a model. To better compare different models, one can first define subtasks and then categorise the different models based on the order in which these subtasks are executed. One of these subtasks is the extraction of the features of a shower image. This can be done, for example, by a convolutional neural network that provides a low-dimensional representation of the input image in the form of a feature vector. A further task is then the actual classification. This is usually achieved by a fully connected neural network that takes a feature vector as input and calculates a score for each type of particle that could have triggered the air shower. In the specific case of stereo imaging, at least one additional task is required, namely the aggregation of the information provided by the different telescopes. This may involve simple statistical operations such as summation or averaging. Optionally, more elaborate algorithms working with sets or sequences can be used to find and exploit correlations between the different telescope images. In this way, three basic structures for the design of DL models are found, which are shown in Fig. 5.2. In the first structure, the telescope images are first aggregated. Then a feature vector is extracted from the aggregated image, based on which the final classification score is calculated. In the second structure, a feature vector is calculated for each individual telescope image. In an intermediate step, a specialised network such as a long short-term memory (LSTM) [156] or a transformer [157] network can be used to correlate the information before aggregation and final classification. Finally, in the third structure, a classification value is calculated for each individual image and the final decision is then based on the ensemble of these values. An intermediate set or sequence operation may also be included in this structure before the classification network. This proposed scheme for categorising model architectures is useful for discussing the models studied in this work. However, it must be emphasised that it does not cover all possible model architectures. For example, it does not cover networks that consider additional inputs besides telescope images, such as Hillas parameters. Since Hillas parameters are known to have discriminative power,

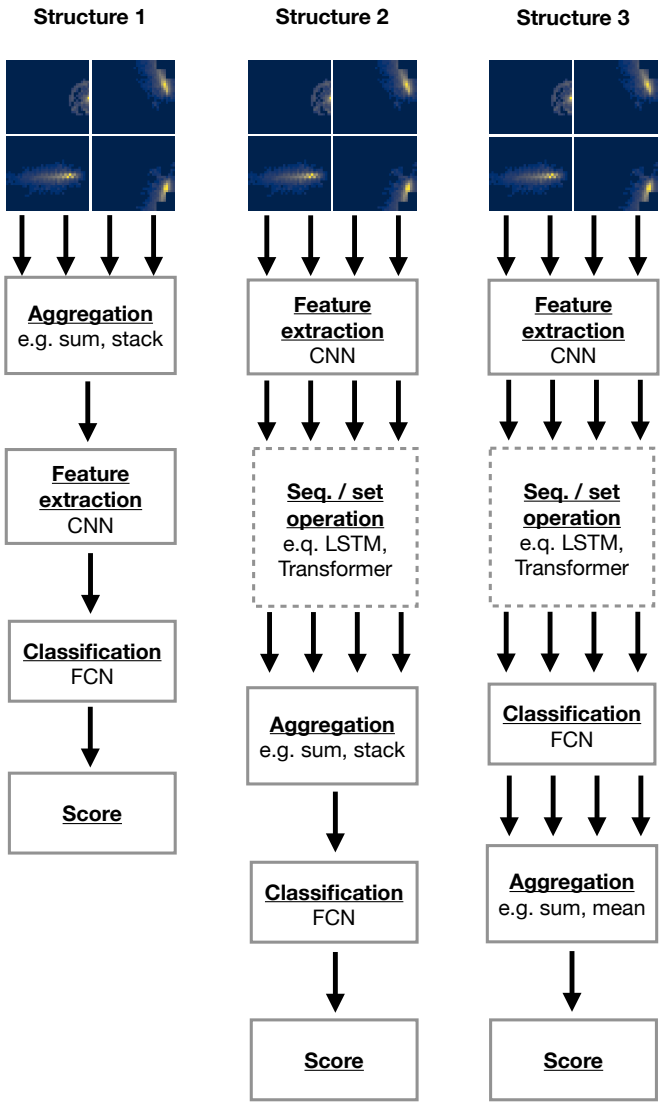


Figure 5.2: *Basic structures of DL models for γ-hadron separation using data from IACTs. Starting from individual telescope images of the same air shower (top), different subtasks are performed to finally obtain a classification score for the considered particle types (bottom). The three structures differ in the order in which the subtasks are performed. Here, the number of arrows corresponds in each case to the number of data streams that are processed and then forwarded by a subprocess. Further details can be found in the text.*

providing them as additional input can help the network achieve its optimal configuration. For the sake of clarity, a discussion of models with additional components such as those involving human guidance is omitted here.

In this work, several models have been probed that cover all the proposed structures. The common basis of these models is the CNN used as feature extractor. Within the context of this work, many different CNN designs with, for example, different numbers of convolutions

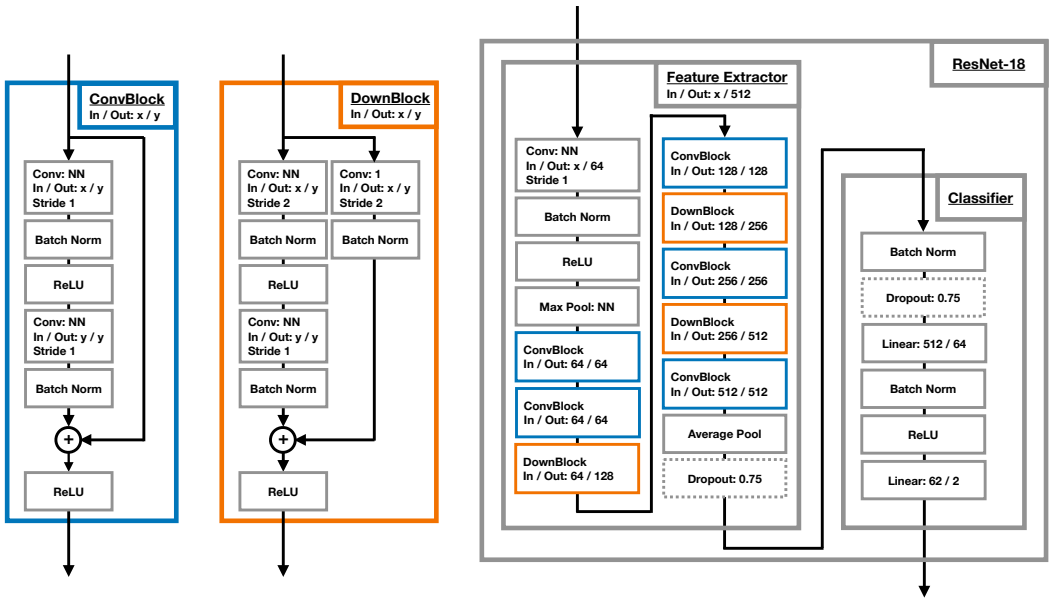


Figure 5.3: *Schematic representation of the ResNet-18 architecture. Left: Convolutional block with residual connections. Middle: Convolutional block with residual connections and downsampling. Right: Full ResNet-18 architecture divided into a feature extractor composed of convolutional blocks and a subsequent fully connected network as the final classifier. Individual convolution blocks are differ by the size of the convolution kernel, the number of input and output channels and the size of the stride. Linear operations (fully connected layers) are determined by the number of input and output channels. Max Pooling may differ in the size of the kernel, while a fixed stride of size 2 is applied.*

and pooling layers or kernel sizes were tested and their effects on the γ-hadron separation were investigated. The selection presented here is consistently based on the so-called ResNet-18 architecture [134], which performed best among all tested network designs. The detailed structure of the ResNet-18 variant used is shown in Fig. 5.3. Its main feature is residual connections, which are used to mitigate the effect of vanishing gradients for successful training of deep networks. The left and middle panels of Fig. 5.3 show how residual connections are implemented in dedicated substructures. The difference between these structures is the downsampling of the input in the structure called DownBlock (middle panel). This is achieved by using a convolution kernel of size one and stride two. While the width and height of the input is reduced, at the same time the number of channels in the output is doubled in relation to the input channels, as can be seen in the right panel of Fig. 5.3, which shows the implementation of these substructures in the overall architecture of ResNet-18. It can also be seen there that the ResNet-18 model consists of two separable parts. The first part consists of convolutional operations and corresponds to the feature extraction task in Fig. 5.2. The second part is a fully connected network that performs the task of classification.

Furthermore, in this figure, the term *NN* used in the representation of convolutional layers is an acronym for *next neighbour*. It describes the size of the convolution kernel consistently for hexagonal or rectangular convolution. For a hexagonal convolution, this means that the kernel covers a central pixel and its six adjacent pixels, while for a rectangular convolution it refers to a standard 3×3 kernel. With the data format chosen, both hexagonal and rectangu-

lar convolution can be applied to the data. For the former, the PyTorch extension mentioned earlier, HexagDLy, is used [155]. While this extension allows accurate computations with hexagonally sampled data, that is unaffected by transformation artefacts, its current implementation is not as computationally efficient as the underlying rectangular convolution operations. For this reason, all models presented below use rectangular convolutions. Following the comparison of the performance of the different models on simulations presented in the next section, the model with the best performance is compared with its hexagonal equivalent to evaluate the differences.

As for the classifier, it can be seen in Fig. 5.3 that the output of the FCN is a two-dimensional vector. Each dimension of this vector corresponds to one of the two classes under consideration, and the numerical values of its elements are the scores assigned by the network to the respective class. Applying the softmax function (Eq. 4.6) to this vector, the scores are converted into estimates of the class-wise probabilities. With binary classifiers, events are completely characterised by the probability of a single class, in this case γ-ray showers, which will be referred to as ζ in the following.

The discussion on the performance of neural networks will be based on seven different models, each following one of the three proposed structures. The details of each model are described below. For a quick overview, a summary of the model architectures is given in Tab. 5.1.

Structure 1 Two models were examined that follow structure 1 shown in Fig. 5.2. In the first of these models, called **mSumConv**, summation is used to aggregate the shower images. The classification result is then calculated using ResNet-18, as shown in Fig. 5.3. This model is particularly efficient as only one image per shower event needs to be processed. The disadvantage of this model is that subtle features of the showers present in the individual telescope images may be lost in the merged result. The second model is similar to mSumConv, but instead of the sum, all images are stacked. Accordingly, this model is called **mStackConv**. The CNN is set to accept input data with four channels, one for each telescope[1]. In this way, all telescope-specific information can be evaluated by a network. Computationally, this model is still efficient, although the number of model parameters increases slightly due to the increased number of input channels. One disadvantage of this design is that a comparatively large number of pixels containing no information must be processed by the network. This is because the majority of all shower events have a telescope multiplicity of two, leaving entire input channels empty. Furthermore, with a potentially larger telescope array such as the future Cherenkov Telescope Array (CTA) [158], the model would become inefficient in terms of memory usage.

Structure 2 Four models following the proposed structure 2 were studied. In general, this structure is more computationally intensive than the previous one. On the other hand, it is more memory-efficient than mStackConv because only the images that actually contain information need to be processed. In addition, each individual image can be processed in parallel. This feature of the architecture also makes it interesting for potential use with larger telescope arrays. By providing the CNN with individual telescope images instead of individual shower events, the size of the dataset used for training is effectively increased by a factor ~ 2.8, which corresponds to the average telescope multiplicity. This can lead to higher

[1] CNNs can handle image-like data with any (fixed) number of input channels

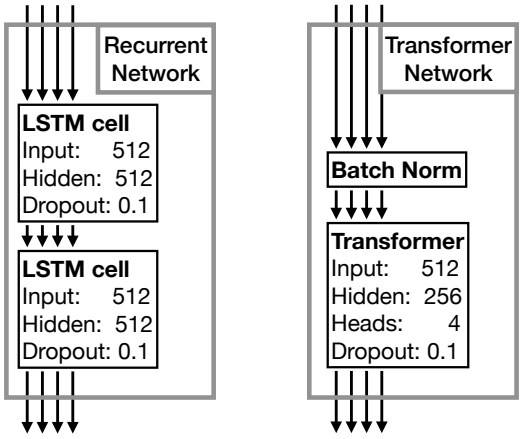


Figure 5.4: *Two structures referring to the sequence / set operation in Fig. 5.2. Left: A recurrent neural network as used in model mConvRec. The details of the LSTM cell are given in [159]. Right: A transformer network as used in model mConvTrans. The details of the transformer cell are given in [157].*

accuracies or lower generalisation errors. Depending on the aggregation function, correlations between telescope images are exploited more or less effectively. The first two studied models following this structure are similar to those discussed previously. In the model **mConvSum**, the feature vectors are simply summed up before the classification task is applied. In the model **mConvCat**, the feature vectors are concatenated. The classifier is adapted accordingly to use a vector as input that is four times as large as the vector used in mConvSum. In this way, connections between two or more components of the feature vectors can be explored individually by the network, in contrast to the sum approach. On the other hand, in this approach, the size of the feature vector is tied to the total size of the telescope arrays. Another advantage of summing as an aggregation function, besides more efficient memory use, is that the model becomes permutation equivariant with respect to the telescope arrangement in the input. For tasks such as γ-hadron separation, where the relative position and orientation of the telescopes is not expected to be information with high discrimination power, permutation equivariance can help the network to generalise from the training sample to new inputs and increase its robustness to distortions in the input distribution.

The other two networks that follow structure 2 use an additional network substructure between feature extraction and aggregation. One of these is called **mConvRec** and uses a recurrent neural network (RNN). RNNs perform sequence-to-sequence operations, with the main application being language modelling. In an RNN, each new input changes the state of the network, which affects subsequent computations. In this way, the output y_i corresponding to the input x_i at step i in a sequence of inputs depends on all previously processed inputs $y_i = f(x_i \,|\, x_1, x_2, ..., x_{i-1})$. It has already been shown that using RNNs to exploit correlations between telescope data allows the construction of DL models that outperform classical γ-hadron separation algorithms [160,161]. For the model tested here, LSTM layers were used to construct the RNN. The main property of the LSTM is that it both adds information to the output stream based on the new and previous inputs and subtracts information that does not match the new input. For details of the realisation of an LSTM, the interested reader is referred to [159]. In mConvRec, a two-layer LSTM network is used as shown in the left panel

of Fig. 5.4. The output sequence $[y_1, ..., y_N]$ is summed before being passed to the classifier. An important aspect in the application of RNNs is the order of the input sequence, which has a significant impact on the performance of the network [162]. While there is no natural order for the arrangement of the telescope images, it has been shown that ordering by shower size (integral intensity) gives good results [160], which has been adopted here.

As an alternative to a RNN, the **mConvTrans** model is based on a transformer network. For many tasks, especially in the field of language modelling, the most advanced neural networks are based on transformers, outperforming any RNN variant. The most important building block of transformers is the mechanism of self-attention. Self-attention relates the information present in a vector v_i to all vectors v_j in a set $\{v_1, ..., v_N\}$ simultaneously via the dot product and learnable representations of these vectors. The details are given in [157]. Unlike RNNs, transformers operate on sets and are intrinsically permutation equivariant. In language modelling, positional encodings of the input are used to make the transformer aware of the sequential order of the set elements. Such encodings could also be used for the position of telescopes in the array. In the model mConvTrans, however, the permutation equivariance of the transformer is to be exploited, which means that no positional encoding of the input is used. The structure of the transformer network used is shown in the right panel of Fig. 5.4. The set of output vectors $\{y_1, ..., y_N\}$ is summed as in mConvRec before being passed to the classifier.

Structure 3 Finally, another model was investigated, following structure 3 in Fig. 5.2. In this model, referred to as **mSingle**, all telescope images are classified individually, i.e. each image is treated as an independent observation. From the assumed independence, it follows that the likelihood of a shower being triggered by a γ-ray can be calculated over the n telescopes as $\mathcal{L}_\gamma = \prod_{i=1}^{n} \zeta_i$. Accordingly, the likelihood of a proton shower is calculated as $\mathcal{L}_p = \prod_{i=1}^{n} 1 - \zeta_i$. The final aggregation of the results is then to approximate the probability of the shower being triggered by a γ-ray as

$$\zeta = P(\gamma \,|\, \{\zeta_1, ..., \zeta_n\}) = \frac{\mathcal{L}_\gamma}{\mathcal{L}_\gamma + \mathcal{L}_p} \,. \tag{5.1}$$

This model is primarily used as a basis to check whether the other proposed networks effectively exploit the correlations between the telescope images.

Next to the architecture of the network, the preprocessing of the data also has a great impact on the final performance of the model. Image cleaning and the application of the addressing scheme, as discussed previously, can be considered part of the preprocessing. It is also common to standardise the input to a neural network to a fixed mean and standard deviation so that the output is less affected by shifts in the input distribution with respect to the training sample. However, among several alternative schemes, it has been found that the best results in terms of performance and stability of γ-hadron separation are obtained when each channel of the input to the CNN feature extractor is normalised to the maximum value of the intensity in that channel. This scheme allows for a fast calculation and can be performed in real time. It also mitigates to some extent the effects of varying optical efficiencies between simulations and observations, with the exception of threshold effects such as triggering upward fluctuating low-energy showers.

Table 5.1: *Overview of the neural-network architectures of the seven models discussed in this work. Each module refers to a building block or operation associated with a certain subtasks. Details are given in the text.*

Model (Structure)	1. Module	2. Module	3. Module	4. Module
mSumConv (1)	Summation	CNN	FCN	-
mStackConv (1)	Stacking	CNN	FCN	-
mConvSum (2)	CNN	Summation	FCN	-
mConvCat (2)	CNN	Concatenation	FCN	-
mConvRec (2)	CNN	RNN	Summation	FCN
mConvTrans (2)	CNN	Transformer	Summation	FCN
mSingle (3)	CNN	FCN	Joint Probability	-

5.4 Performance on Monte Carlo simulations

All models described above were built using the DL library PyTorch and trained on a system specified in Tab. 5.2. The models mSumConv and mStackConv were trained end-to-end using stochastic gradient descent with a momentum of 0.9 to optimise the models based on cross-entropy loss. The initial learning rate of 10^{-3} was reduced by a factor of 0.5 each time the learning progress stalled for one epoch. In addition, a weight decay by a factor of $4 \cdot 10^{-3}$ was applied. In the same way, the model mSingle was trained on single shower images without the final aggregation, which is only applied in the final evaluation of the trained model.

Table 5.2: *Specifications of the computational system used to obtain the presented results.*

Component	Specification
OS	Ubuntu 18.10
CPU	Intel(R) Core(TM) i7-7800X
GPU	Nvidia GeForce GTX 1080 Ti
Memory	4×16 GB DDR4, 2400 Hz
Storage	256 GB SSD
	1000 GB HDD

All models following structure 2 in Fig. 5.2 are built on the feature vectors derived from the trained model mSingle, i.e. the parameters of the CNN were not further optimised. The Adam optimiser was used to train the subsequent components of the respective network. For training these models, the learning rate was first increased linearly over several training samples up to a maximum value of 10^{-4}, and then periodically varied around a slowly decreasing value. This method is commonly known as cosine warm-up annealing. After each training epoch, the accuracy of the network prediction was calculated on the validation set for a classification threshold of $\zeta_{thresh} = 0.5$. Only the network configuration that achieved the highest accuracy was stored. If there was no improvement in validation accuracy for five epochs, the training was stopped. Usually, the training was stopped in this way after less

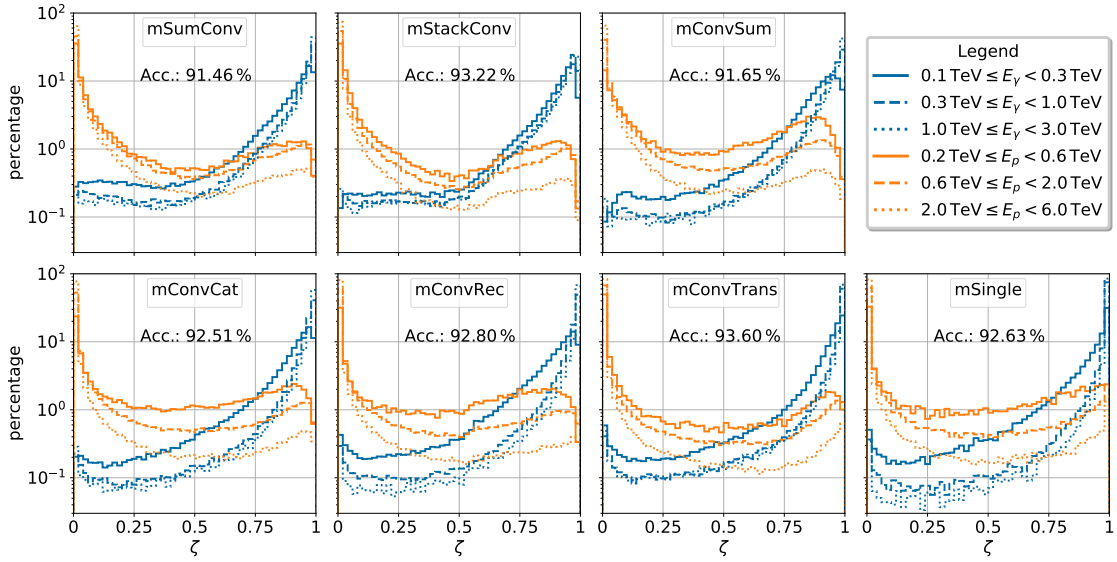


Figure 5.5: *The ζ distribution of γ-rays (blue) and protons (orange) derived with the various models on the simulated test set in three different energy bands. The accuracy over all energies is given for a classification threshold of $\zeta_{thresh} = 0.5$.*

than 20 epochs. With their final configurations, all models reach a high accuracy on the test set which varies between 91.5 % (mSumConv) and 93.6 % (mConvTrans).

For a more detailed inspection of the performance of the models, Fig. 5.5 shows the ζ distribution of γ-rays and protons separately in three different energy bands. Each panel in the figure corresponds to one model. There is a clear energy dependence of the ζ distribution seen in all models, although the strength of this dependence varies from model to model. Low-energy events are generally less well classified. The main reason is that air showers become smaller and fainter towards lower energies. Therefore, the images of these air showers provide less information to the network for its prediction. Additionally, for the same reason fewer air showers are capable of triggering the telescope array at the lower end of the simulated energy range. Consequently, a smaller fraction of these events is present in the training set, so the network is not equally well optimised for these events compared to higher energy events. More precisely, the fraction of γ-rays in the data sets peaks at $\sim 200\,\mathrm{GeV}$ and decreases faster towards lower energies than towards higher energies. As a result, Fig. 5.5 shows a large performance gap in the classification of γ-rays in all models between the lowest energy bin at $0.1\,\mathrm{TeV} \leq E_\gamma \leq 0.3\,\mathrm{TeV}$ and the two energy bins at higher energies, $0.3\,\mathrm{TeV} \leq E_\gamma \leq 1\,\mathrm{TeV}$ and $1\,\mathrm{TeV} \leq E_\gamma \leq 3\,\mathrm{TeV}$, while the latter show similar performances. Compared to γ-rays, the decrease of the performance of the classifier from high energies towards low energies is more pronounced for protons. A reason might be that the characteristic complex structures caused by hadronic interactions are not well resolved at low energies due to the small shower sizes in the images. Generally, protons are predicted to be γ-like more often than γ-rays are predicted to be proton-like in all models. Since hadronic showers can produce electromagnetic sub-showers identical to those of γ-rays, such a sub-shower is most likely the dominant observed component in the misclassified proton events.

Overall, the ζ distributions of the different models are very similar. Only the models

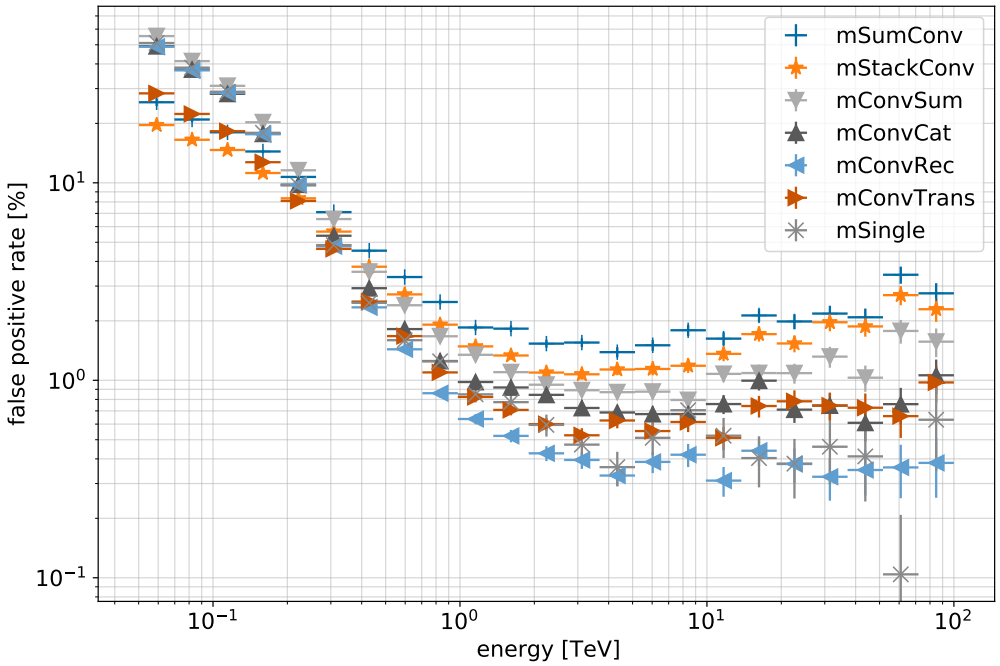


Figure 5.6: *False positive rates, i.e. the fraction of protons with $\epsilon_\gamma \leq \epsilon_{\gamma,max}$, as a function of energy for $\epsilon_{\gamma,max} = 0.8$. All markers provide vertical error bars representing the uncertainties of the FPR estimate and horizontal error bars corresponding to the width of the respective energy bin. Note that protons of twice the energy defined by each bin are used to account for the differences in the estimation of energies for γ-rays and cosmic rays.*

mSumConv and mStackConv stand out with their comparably weak energy dependence. In terms of accuracy, the models mConvTrans and mStackConv reach the best scores. However, the accuracy is not an ideal measure of the γ-hadron separation performance because it is based on a fixed ζ threshold and does not take into account the energy dependence of the network prediction and the class distribution in the target domain, which is not balanced but largely dominated by cosmic rays. With the BDT classifier in H. E. S. S., this issue is addressed by translating the output of the classifier into the corresponding γ-ray efficiency in different energy bands. For a given energy band and corresponding differential ζ distribution of γ-ray events $\frac{\mathrm{d}N_\gamma}{\mathrm{d}\zeta}$, the γ-ray efficiency is defined as

$$\epsilon_\gamma(\zeta) = \frac{\int_{\zeta}^{\zeta_{max}} \mathrm{d}\zeta' \frac{\mathrm{d}N_\gamma}{\mathrm{d}\zeta'}}{\int_{\zeta_{min}}^{\zeta_{max}} \mathrm{d}\zeta' \frac{\mathrm{d}N_\gamma}{\mathrm{d}\zeta'}} \,. \tag{5.2}$$

The range $[\zeta_{min}, \zeta_{max}]$ in Eq. 5.2 is determined by the range of classification scores, i.e. in the case of the neural networks $[0, 1]$. This approach allows for setting a single threshold $\epsilon_{\gamma,max}$ that yields a constant fraction of γ-ray events in each energy band. On the other hand, since the ζ distribution of γ-rays and protons have different energy dependencies, the false positive rate (FPR) for a given $\epsilon_{\gamma,max}$, that is the fraction of protons with $\epsilon_\gamma \leq \epsilon_{\gamma,max}$, is again energy dependent. As a more application-related evaluation of the performance, the FPR as a function of energy for $\epsilon_{\gamma,max} = 0.8$ is shown in Fig. 5.6 for all DL models. The true simulated energy of the events was used to create this figure. In practice, the energy for each event is estimated by considering a γ-ray shower. While these estimates are accurate for

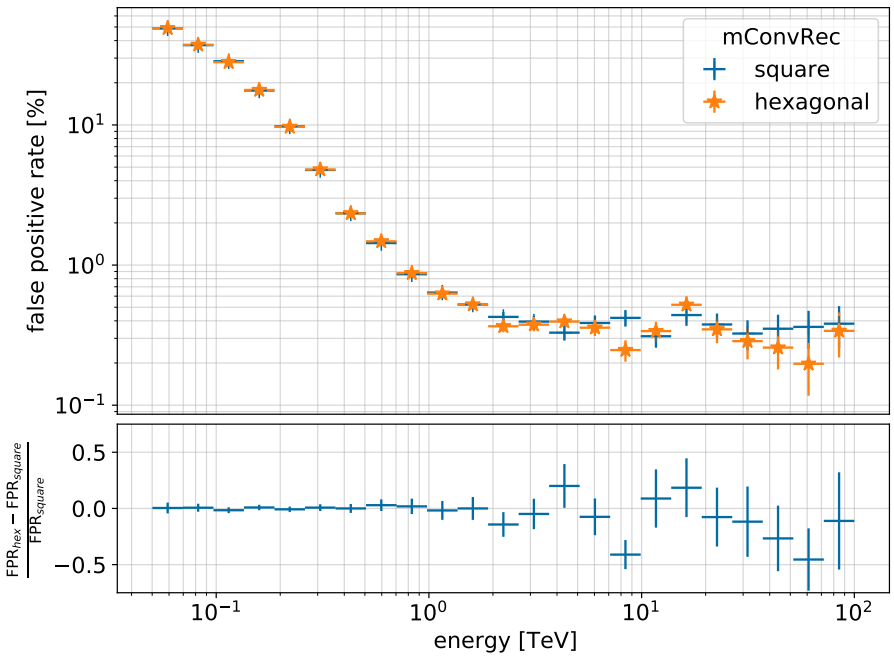


Figure 5.7: *Top: False positive rates over energy obtained with mConvRec for rectangular (cross) and hexagonal (star) convolutions. Bottom: Relative difference with respect to the model variant using square convolutions.*

actual γ-ray showers, the reconstructed energy for protons is on average smaller than their true energy. For this reason, protons with twice the energy given by a bin defining the energy range of γ-rays were considered.

At low energies the classification is generally more difficult such that the FPR is high with all models, e.g. $\gtrsim 10\,\%$ below 200 GeV. Towards higher energies, the FPR steadily decreases until ~ 3 TeV before almost all models show again an increase of the FPR. Next to fewer trainings samples, corresponding to the power-law distribution of the simulated events, the worsening of the performance at high energies can be likely attributed to truncation of shower images, which becomes more frequent as the size of the showers grows. Only the models mConvRec and mSingle appear to be unaffected by these effects and show a relatively constant FPR of 0.4 % above 3 TeV. In the energy range above 400 GeV both models perform best, with mConvRec yielding only somewhat lower FPRs than mSingle, e.g. a 30 % lower FPR around the core energy of 1 TeV. Below 400 GeV there is a turning point such that the models mStackConv and mConvTrans show the best results. Considering the peak of the event distribution around 200 GeV, these two networks in particular seem to have adjusted to the energy-dependent event distribution of simulated air showers in the training set, which explains that both achieve the highest accuracies based on the threshold $\zeta_{thresh} = 0.5$. Different results may be achieved if the networks are optimised for specific energy bands or more elaborate networks are used that, for instance, take additional input into account.

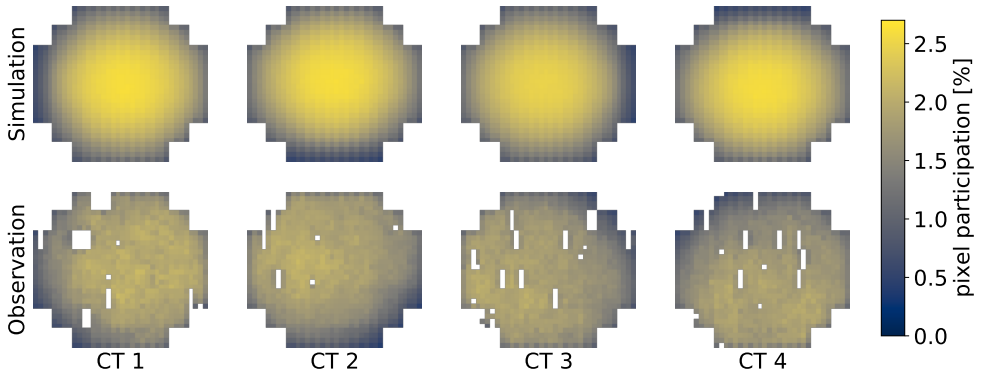


Figure 5.8: *Pixel participation fractions in simulations (top row) and one particular observation run (bottom row). Each column represents one H.E.S.S. I telescope. For the rectangular representation of the telescope images, the addressing scheme as described in Sec. 5.2 is used. Empty (white) pixels mean that there is no signal over the entire data set.*

Another aspect to be discussed here is the influence of the data geometry in conjunction with the convolution operations applied. So far, all the results presented have been obtained by applying standard convolutional operations with square kernels within the ResNet-18 architecture, which do not take into account the hexagonal symmetry of the shower images. To investigate whether the results could be improved, networks were trained using the same architecture but with all rectangular convolutions replaced by their hexagonal equivalent, as described in section 5.2. Figure 5.7 shows a comparison of the FPRs obtained with rectangular and hexagonal convolutions, exemplary for the model mConvRec. Analogous to Fig. 5.6, the upper panel of Fig. 5.7 shows the FPRs of the two variants of the model mConvRec as a function of energy. The relative difference with respect to the variant with rectangular convolutions is shown in the bottom panel. This difference is well fitted by a constant compatible with zero ($\chi^2/\mathrm{ndf} = 18.54/21$). The result is consistent with results obtained with other models and demonstrates that the capacity of the ResNet-18 architecture for the task of γ-hadron separation is sufficient to compensate for the symmetry breaking caused by the addressing scheme used to reformat the data. Furthermore, this result suggests that neither the application of hexagonal convolution nor a specific choice of data transformations, such as interpolation from a hexagonal to a rectangular grid, can improve the performance of a DL classifier. For other tasks, however, such as reconstructing the direction of arrival, exploiting the higher rotational symmetry of the hexagonal geometry of the shower images with appropriate convolution operations may prove beneficial. Nevertheless, with the result of this comparison, the following discussion on DL classifiers will further focus on models based on rectangular convolutions.

5.5 Performance on observations

The performance of neural network should not be judged solely based on simulations. A common issue with neural networks is their susceptibility to differences in the underlying distribution function of the input data between the training domain and the target domain, sometimes resulting in drastic performance degradation. Such differences naturally occur be-

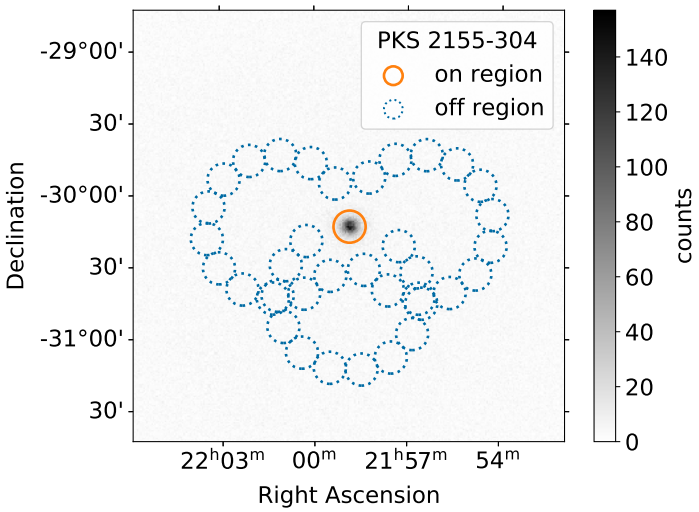


Figure 5.9: *Sky map of all events recorded during an observation of PKS 2155-304 in a high flaring state. The bright point-like source stands out clearly from the background. Circles with a squared radius of* $0.0125\,\mathrm{deg}^2$ *show the on-region (solid) and off-regions (dotted) defined with the reflected region method for the three analysed observation runs.*

tween simulations (training domain) and real-world samples (target domain), due to missing details or variation in the simulations. In ground-based γ-ray astronomy, data analysis relies heavily on simulations, as many tools and the IRFs are calibrated using simulations. At the same time, only a small fraction of the parameter space covered by observations can actually be simulated. For example, the composition of the particles and their spectra are only simulated for some representative parameter values for reasons of efficiency. The same applies to the condition of the atmosphere. The biggest impact on the input distribution, however, is expected from the differences between the simulated and the actual state of the instrument, as well as from the changing altitudes and NSB rates depending on the observed regions on the sky. A particular example of such a difference is shown in Fig. 5.8. This figure shows the average participation fraction of each pixel in the camera per shower event recorded by that camera. Each column in the figure represents one of the H.E.S.S. I telescopes. The top row shows the distribution for the simulated events used to train the networks, the bottom row shows the distribution for an observation run targeted at the active galactic nucleus PKS 2155-304 [163]. The pointing position of the telescopes for the observation run is similar to the simulated position and the NSB rates are low because the extragalactic target is away from the Galactic plane. The noticeable difference is instrumental, namely missing pixels in the cameras as seen in the real-world example. These are pixels that were switched off due to malfunctions during the observation or were removed from the data during the calibration of the data. A variable number of these so-called "broken pixels" is common in observations. They are an obvious example of how real-world data can be distorted compared to simulations of ideal shower images. Other differences can be more subtle. A suitable classifier must prove robust to such distortions.

To test the performance of the various models on the target domain, data from three observation runs targeted at PKS 2155-304 during a high flaring state [163] have been analysed.

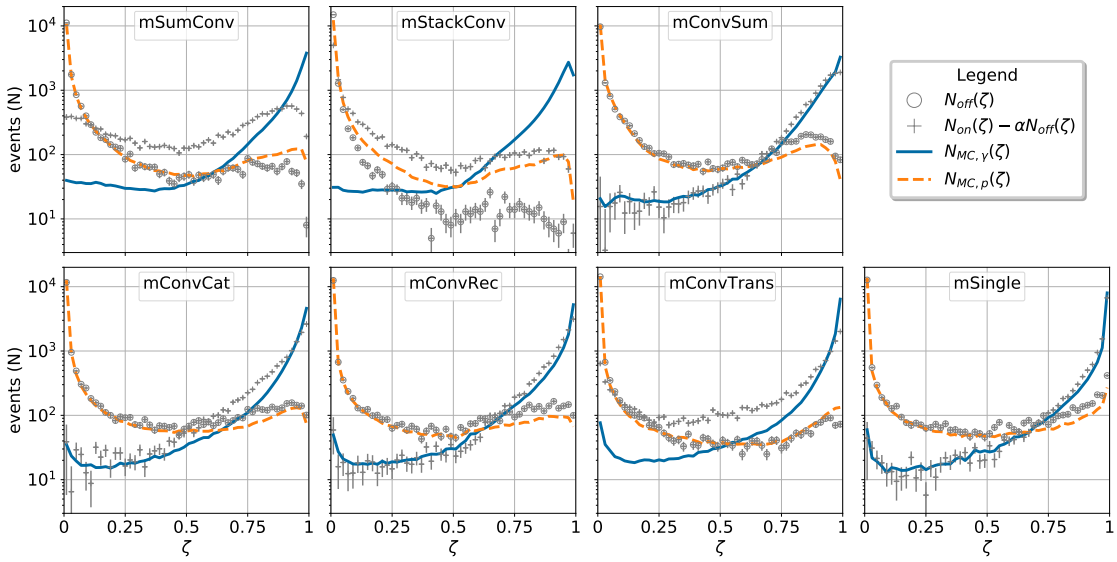


Figure 5.10: *Comparison of the ζ distributions between simulations and real observations. The distribution obtained from events from the off-regions of the observational runs ($N_{off}(\zeta)$, circle marker) shall be compared with the distribution obtained from simulated protons ($N_{MC,p}(\zeta)$, dashed line). Similarly, the distribution of excess events ($N_{on}(\zeta) - \alpha N_{off}(\zeta)$, cross marker) shall be compared with the distribution obtained from simulated γ-rays ($N_{MC,\gamma}(\zeta)$, solid line). Each panel represents a different model.*

Besides the distribution of broken pixels as shown in Fig. 5.8, a known difference with respect to simulations concerns the spectrum of the source which is much steeper than the simulated γ-ray spectrum, i.e. a broken power law with spectral index $\Gamma_1 = 2.71$ below an energy of 430 GeV and $\Gamma_2 = 3.53$ above that energy. The integral flux above 200 GeV is reported to be ~ 7 times that of the Crab Nebula (the common reference in VHE γ-ray astronomy), which gives high statistics and allows for detailed comparisons with simulations. In the following, the ζ distribution of simulated and real cosmic rays as well as for simulated and real γ-rays shall be compared. For this purpose, all recorded shower events of the observation runs have been reconstructed with ImPACT [120] and on- and off-regions have been defined using the reflected regions method. The corresponding sky map is shown in Fig. 5.9. Since the off-region presumably contains only cosmic-ray events, the ζ distribution of these events is expected to match the ζ distribution of the simulated protons. To extract the ζ distribution of the real γ-rays, the distribution was first calculated for all events in the on-region and then the distribution calculated for the off-region, scaled by the α factor (see section 3.4.2), was subtracted. To clarify that the result of this procedure is only an estimate of the distribution of γ-rays, it is more accurately referred to as the distribution of excess events. For the comparison between simulation and observation shown in Fig. 5.10, for each particle type the ζ distribution of the simulated events is scaled to match the number of events in the observation. Each panel of the figure shows the result for a different DL model. There it can be seen that some models do not respond well to the change of the target domain. Especially the models that show good discriminatory power in simulations at low energies (cf. Fig. 5.6), namely mSumConv, mStackConv and mConvTrans, show great difficulties in dealing with the distortions present in the observational data. While the models mSumConv and mConvTrans show at least a good agreement between the distributions of the cosmic-ray

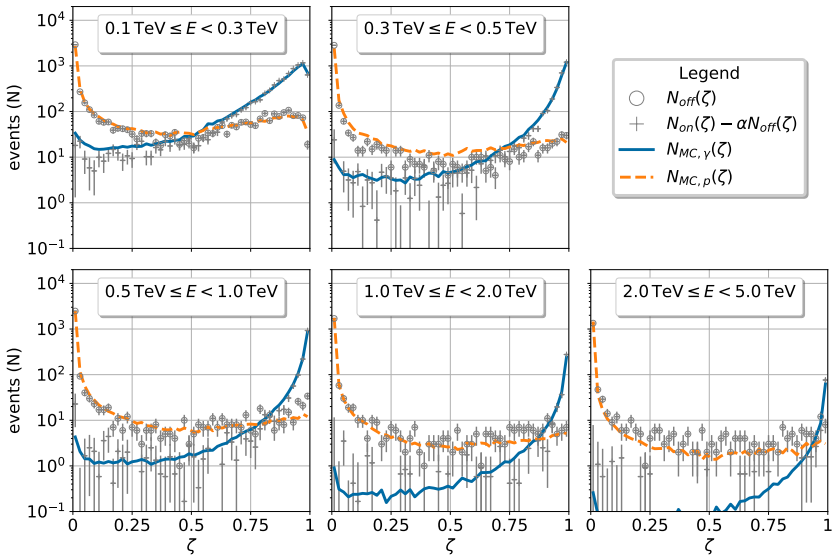


Figure 5.11: *Analogous to Fig. 5.10, a comparison of the ζ distributions between simulations and real observations in different energy intervals for the model mConvRec.*

background from the off-region and the simulated protons, all three models show large discrepancies between the expected distribution of γ-ray events and the distribution of excess events in the observational data. Therefore, these three models are clearly unsuitable for the actual analysis of γ-ray data. Small differences are also observed in the mConvCat model. Even these differences in the ζ distribution of γ-ray events could already be large enough to influence the calculation of the fluxes. In contrast to that, the three models mConvSum, mConvRec and mSingle all show good agreement between simulation and observation for both γ-rays and cosmic rays. In fact, an energy-dependent comparison, as exemplified in Fig. 5.11 for mConvRec, shows excellent agreement. To account for the fact that the estimated energies for observation are compared with the simulated true energies, in this figure the energy of the simulated protons corresponds to twice the energy of a given energy bin (as in Fig. 5.6). Energy-dependent features such as the small dip in the distribution of excess events towards high ζ values at low energies ($0.1\,\text{TeV} \leq E \leq 0.3\,\text{TeV}$) are well matched by the distribution of simulated γ-rays. In general, the same level of agreement is observed for mConvSum and mSingle. However, since the mConvRec model has been shown to perform best among these three models, it is used in the following section to compare the DL classifier with the standard BDT classifier.

5.6 Comparison with the BDT classifier

For the comparison of the DL model with a state-of-the-art classifier, the classifier described in [116] was chosen, which is based on boosted decision trees and used in one of the two H. E. S. S. analysis chains. This BDT classifier was trained in energy bands and the standard output of the analysis is the γ-ray efficiency ϵ_γ, which corresponds to the ζ value calculated by the classifier for each event. Accordingly, for the DL classification results ϵ_γ was also calculated in energy bands corresponding to those used for the BDT classifier (those also used in Fig. 5.11). In order to put the DL classification on the same basis as the BDT classification, a preselection of events in the test data set and the set of observational data

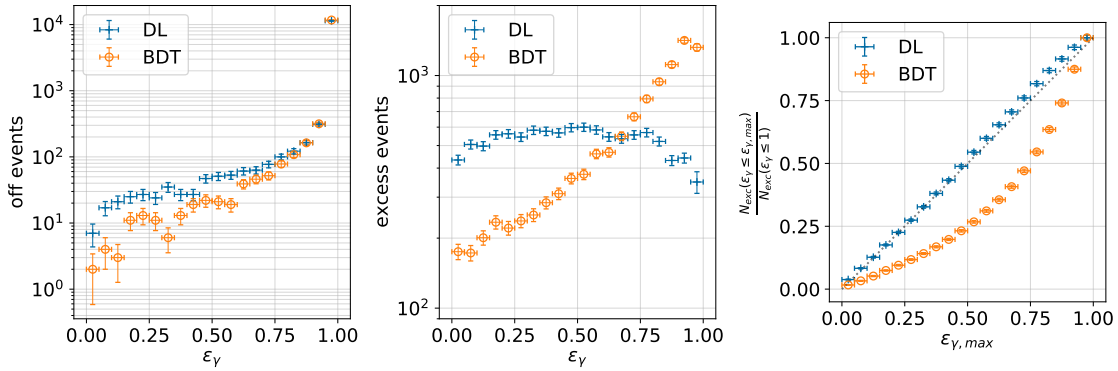


Figure 5.12: *Distribution of γ-ray efficiencies in observational data obtained with the BDT classifier (circles) and the DL classifier (crosses). Left: Distribution of events in the off-region. Middle: Distribution of excess events. Right: Normed cumulative distribution of excess events.*

was applied, corresponding to the standard analysis configuration as stated in [116]. In this preselection, images with a size of less than 60 p.e. (photo electrons) or a distance between the centre-of-gravity of the Hillas ellipse and the camera centre of more than 0.525 m are rejected. Following these cuts, those events which have a telescope multiplicity of at least two are selected for further analysis. A significant fraction of events ($\sim 30\,\%$) is removed from the data with this preselection. On the other hand, it reduces the uncertainties connected to fluctuations in shower images with small sizes and truncation of showers at the edge of the camera, thus, improving the reconstruction of air showers including the classification of the primary particle. In this way, the accuracy of the model mConvRec is enhanced from 92.8 % to 94.6 %. Furthermore, the standard analysis configuration for point-like source implies the choice of a circular analysis region with a squared radius of $\theta^2 = 0.0125\,\mathrm{deg}^2$ centred on the target, which was adopted here (see Fig. 5.9).

The distributions of γ-ray efficiencies in observational data is shown in Fig. 5.12 for both the BDT classifier and the DL classifier. The left panel shows the distributions derived in the off-region, i.e. only cosmic-ray events. As expected, both distributions increase continuously with increasing γ-ray efficiency. Thereby, the distribution obtained with the DL classifier is consistently above that obtained with the BDT classifier for $\epsilon_\gamma \leq 0.85$. This means that for a typical threshold of $\epsilon_{\gamma,max} = 0.84$, as used in the standard analysis configuration, a larger background contamination must be expected with the DL classifier. The middle panel of Fig. 5.12 shows the distribution of γ-ray efficiencies derived from excess events. Here, the distribution obtained with the BDT classifier increases continuously, as in the case of the background events, while the DL classifier gives an almost flat distribution, which is to be expected according to the definition of the γ-ray efficiency, provided that the excess events indeed represent γ-rays. The relation between the expectation regarding the distribution of γ-rays and the observation of excess events is highlighted in the right panel of Fig. 5.12. There, the normalised cumulative distribution of γ-ray efficiencies for excess events is shown. From the definition of the γ-ray efficiency it follows that $\epsilon_{\gamma,max}$ corresponds to the fraction of γ-rays selected regardless of their energy. This is shown by the dotted line, which agrees well with the cumulative distribution of excess events obtained with the DL classifier. This again shows that the DL classifier generalises well to new data across all energies. In contrast, the distribution obtained with the BDT classifier deviates significantly from this

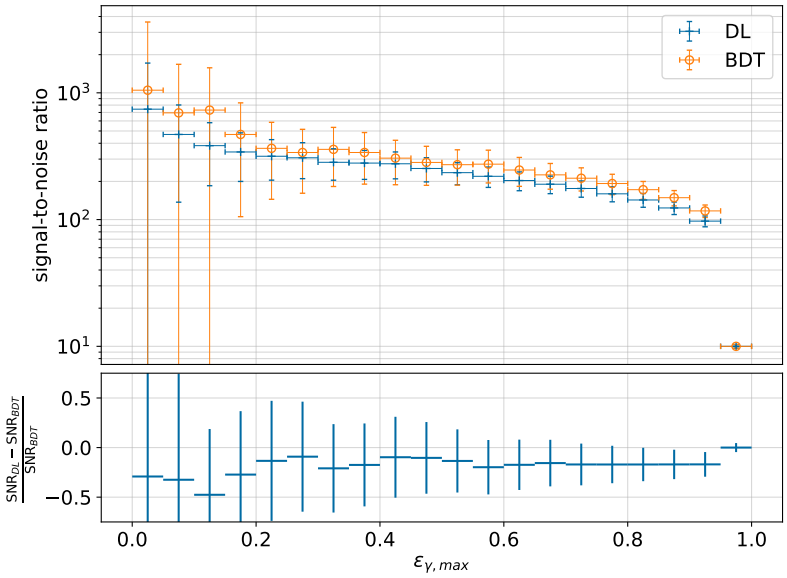


Figure 5.13: *Signal-to-noise ratio (SNR) for the BDT classifier (circles) and the DL classifier (crosses). Top: SNR as function of the γ-ray efficiency cut. Bottom: Relative difference with respect to the BDT classifier.*

expectation, indicating problems in the analysis of the data that lead to misclassification of events. Using observational data only, no statement can be made as to whether the problem concerns only the classification of γ-ray events or also cosmic-ray events and whether it is an energy-dependent effect. This is a question that should be followed up and requires a detailed comparison of the BDT classification between simulation and observation. However, this is not a topic addressed in this work.

With this issue, the comparison of the signal-to-noise ratio (SNR) as a function of $\epsilon_{\gamma,max}$ between the two classifiers shown in Fig. 5.13 should therefore be taken with caution. The shown SNR is highest for the most restrictive cuts at the expense of signal strength, i.e. for low $\epsilon_{\gamma,max}$. It decreases with increasing thresholds as the background increases faster than the signal. For a strong source like PKS 2155-304 in this flaring state, background rejection is not a big concern. For more common use cases, however, a trade-off between SNR and signal strength must be made. In this example, it can be seen that the SNR for the DL classification for almost any value $\epsilon_{\gamma,max}$ is about 20 % below the value obtained with the BDT classifier. However, considering the problem of misclassification of events with the BDT classifier, no firm conclusions can be drawn. This comparison only shows that DL classifiers achieve performances at least close to that of the BDT classifier.

5.7 Outlook

The study presented in this chapter on the application of DL models as γ-hadron classifiers has shown the potential of such models as robust classifiers with competitive performances.

Although this study is relatively extensive in terms of the neural-network architectures studied and the impact of various design choices on the final performance, it is far from conclusive. As for the robustness of the classifier to shifts in the input distribution, only the effect of broken pixels was considered here. Another aspect that needs to be investigated is the robustness of the classifier to different NSB levels before any application of DL models to γ-ray analyses can be considered. In the comparison presented here, the tested DL models were also not found to be superior to the BDT classification in terms of signal-to-noise ratio. However, the comparison is known to be inaccurate, as it has also been shown that the BDT classifier misclassifies events. On the other hand, different studies claim that DL models perform better than BDT models [160, 161]. A common difference between these two studies and the present study is the preprocessing of the data, especially the conversion of the hexagonal air-shower images into a rectangular format. Here it has been shown that even accurate hexagonal convolution does not give any improvements, so no improvement can be expected from the transformation of the image geometry itself either. Instead, the applied transformation in both of the aforementioned studies also involves upscaling the image resolution. This gives the convolutional neural network more scope to explore subtle features of air-shower images that could contribute to better classification. Additionally, in [161] Hillas parameters were given to the model as additional input. Although none were presented here, similar network architectures were tested during the course of this work, but no benefit of such additional input was found. Another aspect that is of great importance in the optimisation of DL models and could be responsible to some extent for the observed performance differences is the correct choice of hyperparameters and the initialisation of the network parameters.

Standard classifiers and the presented DL models very effectively exploited differences in the geometry of the intensity distribution of air showers of γ-rays and cosmic rays. In this way, they approach the practical limits of this method, which is defined by an irreducible background of hadronic showers that transfer most of their energy into an electromagnetic channel within the first interactions with the atmosphere and are therefore indistinguishable from γ-ray-induced air showers [164]. Besides the spatial distribution, modern IACTs are also able to provide information about the temporal evolution of an air-shower signal in the camera plane with nanosecond resolution. One way to use this information for γ-hadron separation is to discard or attenuate the cuts applied in image cleaning. The timing information allows the coherent signals of the air shower to be better distinguished from the random signals of the NSB, so that, especially at low energies, a larger part of the signal of an air shower can be used for the reconstruction of the primary particle than would be possible if restrictive cuts were applied in the image cleaning to reduce the contribution of the NSB. Without much effort, it would be possible to integrate the processing of this additional information into the architecture of a neural network. In general, it becomes clear that neural networks are a flexible analysis tool and offer many more possibilities for using the rich data collected by modern ground-based telescopes than could be covered in a single study. Therefore, they should be further explored in this context. The study presented here lays foundations that can be helpful for such further research.

6 Model of the population of Galactic VHE γ-ray sources

Modelling the population of Galactic VHE γ-ray sources is a timely issue. Following the detection of diffuse VHE γ-ray emission in the Galaxy [83–86], the question of the contribution of unresolved sources still remains. In particular, resolving spatial variations of the contribution from unresolved sources and propagating cosmic rays has the potential to yield valuable insights into the physical mechanisms of cosmic-ray propagation. A population model can be used to address this question. In addition, it can be expected that γ-ray analyses should yield more robust results when estimations of the contribution of emission from unresolved sources within an analysed region of the sky are included. Furthermore, simulations of entire samples of sources based on such a population model may also be used to optimise observation strategies and analysis techniques.

Essential parts of this chapter have been presented in [165]. Here, the modelling approach is described in more detail. The estimation of the model parameters was adjusted so that the uncertainties in the measurement of source properties are taken into account. Confidence intervals for the estimated model parameters are also given.

This work is not the first endeavour to model the population of Galactic VHE γ-ray sources. In previous studies, the population was investigated using different means and corresponding assumptions. A simple approach is to fit a power law to the cumulative distribution of detected sources over their integral flux (typically presented with a double logarithmic scaling as $\log(N)$-$\log(F)$ distribution) within the so-called range of completeness. For a spatially coherent luminosity function, the fitted power law can be extrapolated to some targeted flux level outside the range of completeness to infer the number of sources corresponding to that flux level [87, 166]. As the name suggests, the range of completeness defines the flux interval for which the sample of observed sources is expected to be complete. Another approach is population synthesis. Thereby, populations are simulated based on a parametric model whose parameters are adjusted such as to give synthetic source samples in the range of source detectability (for a given instrument) which are compatible with observations. This way, the populations of, for example, SNRs [167, 168] and PWNe [169] at very-high energies have been studied with models for the distribution of source properties derived either from theoretical considerations or from observations.

The population synthesis approach was also used for the model presented here. The basis of the model is the HGPS catalogue. Compared to previous studies, no emphasis has been put on a certain class of sources. Instead, both the HGPS sample and the simulated populations are treated as sets of generic γ-ray sources with common distribution functions. In this way, the model allows the prediction of global properties of the population and the estimation of e.g. the emission of unresolved sources or the total number of sources detectable with future instruments such as CTA. An accurate treatment of the observational bias is fundamental

to the population synthesis approach. Using the extensive supplements provided with the HGPS catalogue, in particular the sensitivity map, this study improves two aspects of bias correction that have been neglected so far. First, the HGPS sensitivity is inhomogeneous, i.e. the sensitivity varies greatly over different regions of the sky, as shown in Fig. 3.11. Second, the sensitivity is a function of the angular extent of a source. This last aspect affects not only studies based on the population synthesis approach, but also those based on fits to the $\log(N)$-$\log(F)$ distribution in the range of completeness. This is because most of the time the range of completeness is defined only for point-like sources, whereas most detected sources appear to be extended.

6.1 Model parameters

The sample of discovered VHE sources within the Milky Way is fairly heterogeneous, but our knowledge of the population characteristics of individual source classes is limited. For the model developed here, it is assumed that the joint distribution of source properties for the entire population can be effectively described by parametric functions with a small number of parameters. This approach is complementary to detailed source modelling, which typically involves fitting a comparatively large number of parameters. At this point, the model comprises distribution functions for the bolometric luminosity $L_{E>1\,\mathrm{TeV}}$ (henceforth only L), radius R and location of sources $\mathbf{x}$. These are the three physical parameters that relate to the observational parameters integral flux $F_{E>1\,\mathrm{TeV}}$ (henceforth only F), angular extent σ and position on the sky (l, b), which in turn determine the detectability of a source in the HGPS. Note that another simplification of the model is to treat sources as spherically symmetric objects such that the radius is sufficient to describe their physical extent. While a number of VHE sources exhibit complex morphologies, modelling these for all sources in the entire population is beyond the scope of this thesis. The conversion between the radius of a source and its angular extent, which in the HGPS catalogue is given by the 68% containment radius of a spherically symmetric 2D-Gaussian, is calculated as $\tan(\sigma) = \frac{R}{d}$ for a source at distance d. Sources in the catalogue with complex morphologies, for instance shell-like structures, are treated the same way, in which the shell radius is taken as σ.

6.1.1 Spatial distribution

Models are constructed for a set of five different spatial distributions with fixed parameters and compared against each other. The set was chosen to cover a wide range of reasonable spatial distributions, including axially symmetric and non-axially symmetric distributions.

The two dominant classes among the identified sources are PWNe and SNRs. Correspondingly, models for the spatial distribution of pulsars and SNRs are included in the set of probed distributions. Both are axially symmetric, i.e. the source density ρ is only a function of Galactocentric distance r and Galactic height z

$$\rho(r, z) = \left(\frac{\left(r + r_{off}\right)}{\left(R_\odot + r_{off}\right)}\right)^{\alpha} \exp\left(-\beta \frac{(r - R_\odot)}{\left(R_\odot + r_{off}\right)}\right) \exp\left(-\frac{|z|}{z_0}\right) . \tag{6.1}$$

In Eq. 6.1, the distance between the Galactic centre and the Sun $R_\odot$ enters into the scaling factor for the Galactocentric distance. The Galactic height is scaled by a characteristic scale height z_0. Furthermore, the form of the distribution in radial direction is determined by

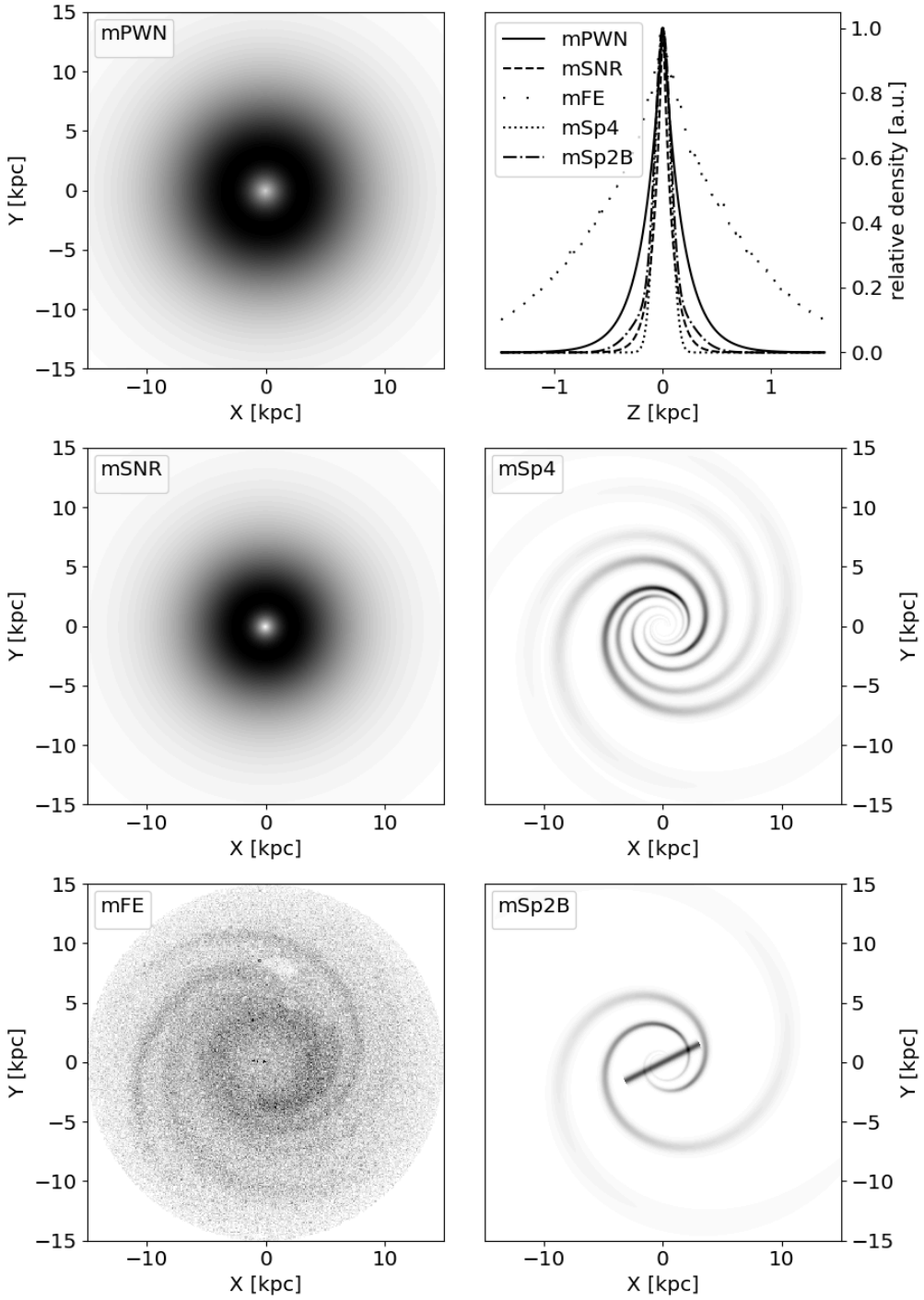


Figure 6.1: *Spatial distribution of sources in the Galactic plane at* $z = 0\,\mathrm{kpc}$ *and corresponding profiles over Galactic height for the five different models. The coordinate system is oriented so that the position of the Sun is at* ($x = 0\,\mathrm{kpc}$, $y = 8.5\,\mathrm{kpc}$, $z = 0\,\mathrm{kpc}$). *Note that the distribution function for mFE was evaluated at random positions due to computational constraints, while the distribution function for all other models was evaluated on a high-resolution spatial grid. In addition, mFE includes several features of the local ISM that exhibit lower density regions. One of these is the local hot bubble [170], which leads to the larger bright spot near the Sun in this figure.*

the shape parameter α and the rate parameter β. A possibly non-vanishing density at the Galactic centre position is accounted for with the parameter r_{off}. For the distribution of SNRs, which is referred to as mSNR, the parameters are taken from [171, 172]. Likewise, the distribution of pulsars is described with parameters as given in [173, 174] and referred to as mPWN. For both distributions the corresponding parameters are listed in Table 6.1. As shown in Fig. 6.1, both the density distribution over the Galactic plane and the density profile over Galactic height for mPWN is considerably broader than for mSNR.

Table 6.1: *Parameter values for the source spatial distribution of mPWN & mSNR corresponding to Eq. 6.1.*

Model	$R_\odot$ [kpc]	r_{off} [kpc]	α	β	z_0 [kpc]
mSNR	8.5	0	1.09	3.87	0.083
mPWN	8.5	0.55	1.64	4.01	0.18

Progenitors of VHE γ-ray sources, i.e. massive stars, have short lifetimes. They occur most commonly within young stellar populations that, on cosmic time scales, have formed only recently in dense regions of gas and dust. This connection between γ-ray sources and dense regions, gives rise to the assumption that the distribution of γ-ray sources could be correlated with the density distribution of the ISM within the Galaxy. The Milky Way is known to be a spiral galaxy with ISM concentrated in its spiral arms. Following this line of argument, the set of spatial distributions probed in this thesis contains distributions with spiral-arm structure as well. Since there is no consensus about the exact structure of the Milky Way, three different distributions with notable variation are used. The choice of these distributions follows the study by [175].

The first of these distributions represents a pronounced four-arm spiral structure and is denoted mSp4. It is adopted from [176]. For each spiral arm the density distribution is described by

$$\rho(r,\phi,z) = \sum_{i=1}^{4} A_i \, exp\left(-\frac{1}{\delta^2}\left(\phi - \frac{\ln\left(\frac{r}{a_i}\right)}{\beta_i}\right)^2\right) \cdot exp\left(-\frac{|r-R|}{\sigma_r}\right) exp\left(-\frac{z^2}{2\sigma_{z,2}^2}\right) . \quad (6.2)$$

In radial direction, each spiral arm is defined by a scale length σ_r and a local maximum at R. In azimuth direction, a spiral arm is defined by a scale length δ and two constants, β_i determining the pitch angle of the spiral and a_i giving its orientation. Also in contrast to Eq. 6.1, the distribution perpendicular to the Galactic plane depends on the square of the Galactic height, not its absolute value, but is likewise scaled by a characteristic scale height $\sigma_{z,2}$. The contribution of each arm is determined by a weighting parameter A_i. Best fitting values to C II cooling line measurements that have been used in [176] to trace the ISM density are adopted for mSp4 and listed in Table 6.2. The result is a density distribution with the narrowest profile over Galactic height for all five models and high concentrations within the spiral arms, but vanishing contributions from inter-arm regions (see Fig. 6.1).

Table 6.2: *Parameter values of the spiral arms used in mSp4 & mSp2B. Values are taken from [176], with the four-arm model fitted to the ISM distribution traced by C II emission.*

Spiral Arm	β_i	a_i	R	σ_r		$\sigma_{z,2}$	δ	A_i
				$(r < R)$	$(r > R)$			
			[kpc]	[kpc]	[kpc]	[kpc]	[deg]	
Sagittarius-Carina	0.242	0.246	2.9	0.7	3.1	0.070	15	169
Scutum-Crux	0.279	0.608	2.9	0.7	3.1	0.070	15	266
Perseus	0.249	0.449	2.9	0.7	3.1	0.070	15	339
Norma-Cygnus	0.240	0.378	2.9	0.7	3.1	0.070	15	176

The second distribution with spiral structure is based on the Galactic free electron density that has been inferred from pulsar dispersion measurements [177] and is denoted mFE. This distribution as well comprises four spiral arms. However, the spiral structure is less pronounced, with inter-arm regions having higher densities than in mSp4. Compared to all other models, mFE exhibits a notably broader distribution over Galactic height. To calculate local densities this work makes use of the numerical code provided by [177][1].

Another distinct distribution is denoted mSp2B and represents a two-arm spiral structure. In addition, mSp2B comprises a central bar, whose existence in our Galaxy has been indicated by Spitzer data [178] and more recently confirmed by Gaia data [179]. This distribution is taken from [180] and includes the Scutum-Crux and Perseus arms as in mSp4. The additional component for the Galactic bar is described by

$$\rho_{bar}(r,\phi,z) = \begin{cases} A_{bar}\ \exp\left(-\frac{z^2+r^2\left(\sin(\phi)-\cos(\phi)\sin(\theta)\right)^2}{\sigma_{z,1}^2}\right), & \text{if } r < l_b \\ 0, & \text{otherwise} \end{cases} \tag{6.3}$$

and has a radial extent $l_b = 3.5\,\mathrm{pc}$. The rotation angle relative to the solar-galactic centre line is $\theta = 30.0\,\mathrm{deg}$. The characteristic scale height of this component is $\sigma_{z,1} = 0.31\,\mathrm{pc}$. The value of A_{bar} is chosen such that the bar is equally contributing to the source density as the spiral arms ($A_{bar} = 364$). The choice of this weighting factor is arbitrary as the decision to include a Galactic bar component in mSp2B but not in mSp4. It does not reflect any physical motivation, but rather the intention to investigate the effect of different features with individually different manifestations with a limited set of distribution functions. The broader distribution in the profile over Galactic height observed for mSp2B in Fig. 6.1 in comparison to mSp4 is due to the contribution of this additional Galactic bar component.

These modelled distributions can be compared with those 29 sources for which an estimate of their distances to the observer is available, as is done in Fig. 6.2. There, the source densities ρ along the line of sight for each of these sources are shown for all five models. The distances shown are constrained to a cylindrical volume defined by a Galactocentric distance $r = 15\,\mathrm{kpc}$ and Galactic height $|z| = 1\,\mathrm{kpc}$. For better visibility, each of the source distributions is scaled so that $\rho_{max} = 1$ along the line of sight. The best estimate of the source position, as taken

[1]The code is available at `http://www.astro.cornell.edu/~cordes/NE2001/`.

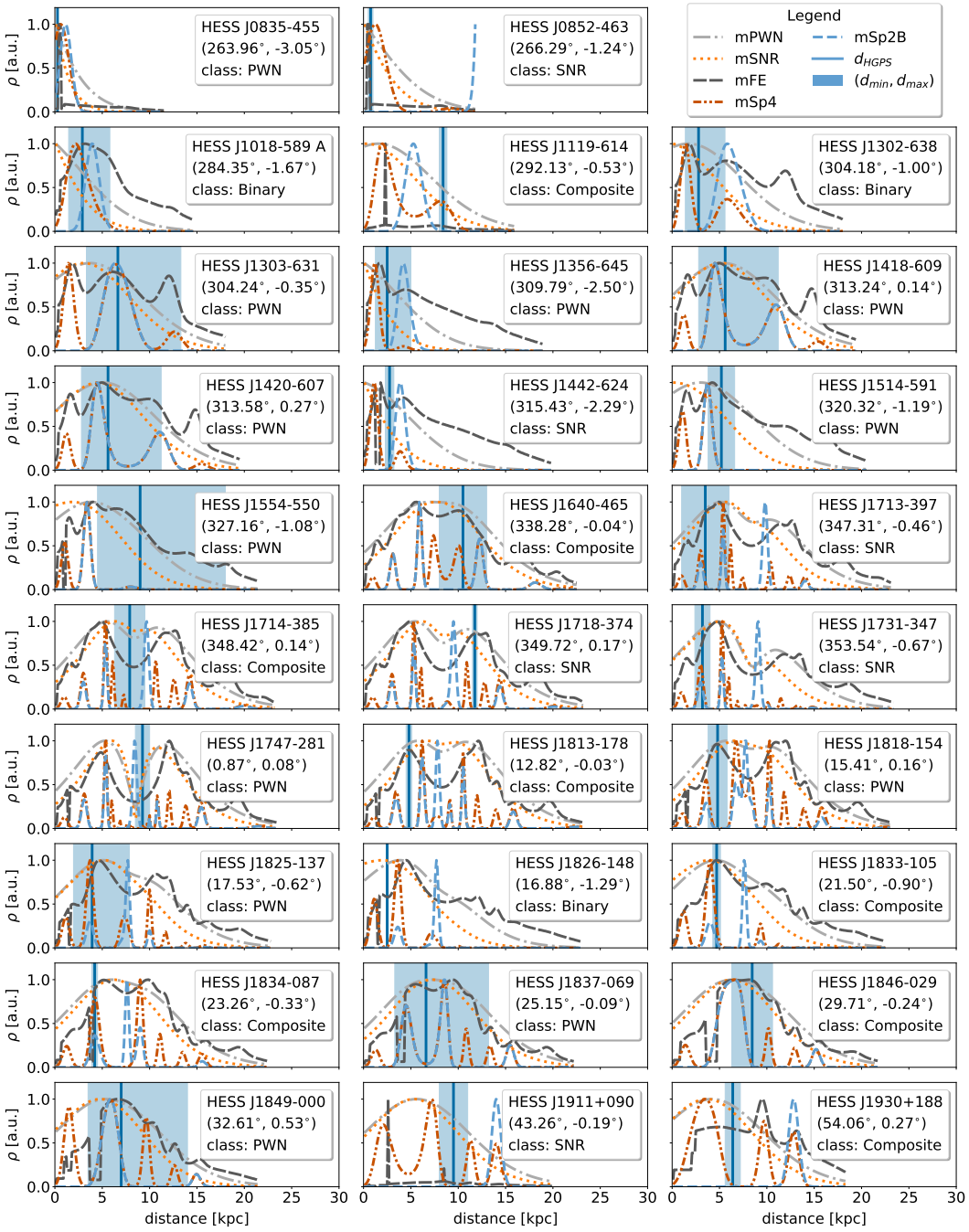


Figure 6.2: *Comparison of sources with known distances to the observer and the source densities along the line of sight to these sources as derived from the models. The distance of each source as given in the HGPS catalogue is marked by a vertical line and a shaded area representing the interval between the minimum and maximum estimated distances. The source density ρ along the line of sight is normalised so that $\rho_{max} = 1$. The sampled distances along the line of sight are constrained to a cylindrical volume defined by a Galactocentric distance $r = 15\,\mathrm{kpc}$ and Galactic height $|z| = 1\,\mathrm{kpc}$ ($|z| = 2\,\mathrm{kpc}$ for mFE). For each source, the position in Galactic coordinates (lon, lat) and the class of the associated astronomical object are given in the legend.*

from the HGPS catalogue, is marked by a vertical line together with a shaded area representing the interval between an estimated minimum and maximum source distance (also from the HGPS catalogue). The models mPWN, mSNR and mFE show broad distributions that in most cases agree well with the best estimate of the source positions. For the models mSp4 and mSp2B, the best position estimate yields as many sources within or in close proximity to a spiral arm region as in supposedly empty inter-arm regions. No dependence is observed between the class of a source and its proximity to a spiral arm region. However, for 24 of the 29 sources, the interval between minimum and maximum distances overlaps with at least one spiral arm. Although these large uncertainties, combined with the small sample size, prevent us from making a sound statement about the true spatial distribution of sources, this comparison shows that in most cases the distribution of known sources is consistent with each choice of spatial distribution described above.

For each spatial model, the results shown below are derived from a fixed set of simulated source positions S_{sim}. For all but the spatial model mFE, the set contains 40 million uniformly randomly distributed points in a cylinder with radius $r = 15\,\mathrm{kpc}$ and height $|z| = 1\,\mathrm{kpc}$, giving a mean distance between points of 33 pc. Each point is then weighted according to the source density given by the tested model. To account for the broader distribution of source density over Galactic height as defined by the model mFE compared to the other spatial models (see Fig. 6.1), the volume of the cylinder representing the Galactic disk is increased such that $|z| = 2\,\mathrm{kpc}$. Despite the increased volume of the cylinder, the number of points in space for mFE is reduced to 8 million, giving a mean distance between points of 70 pc. This is due to the computational limitations of the used implementation of the model.

6.1.2 Luminosity & radius distribution

The dynamical evolution of Galactic γ-ray sources is subject to complex interactions of these sources with their environment. It is evident that both the luminosity and the size of individual sources depend on many physical parameters, e.g. the energy input of the central source, the strength of the surrounding magnetic field and the density of the ISM or ambient radiation fields. While these parameters can be and have been derived for a number of sources using multi-wavelength observations, there are few constraints on the overall distributions of all of these parameters in the direct vicinity of γ-ray sources throughout the Galaxy. Therefore, parametric functions are used in this work to effectively describe the joint distribution of source luminosities and radii that emerge from the combined distributions of the underlying physical parameters.

The sample of known sources is still too small to derive the exact shape of these joint distribution functions. Theory does not provide strong constraints either, but can give some guidance. For a population of sources whose γ-ray emission is driven by the spin-down energy of pulsars, $\dot{E}$, such that for a single source $L_\gamma \propto \dot{E}^{-\beta}$, [181] derives, for example, for the distribution of luminosities $N(L_\gamma) \propto L_\gamma^\alpha$ with $\alpha = -\frac{1+2\beta}{2\beta}$. For pulsars, values of α between -3 and -1.5 are reasonably explained by various emission models that determine the parameter β (see [181] and references therein). A power law for the distribution of the source sizes can also be motivated by considering SNRs, for example. SNRs emit photons at very high energies, especially during the Sedov-Taylor phase. There, the radius of the shock front evolves like $r \propto t^{\frac{2}{5}}$. Assuming that the population represents a snapshot of sources at different times in their evolution, the distribution of source sizes in the population should reflect

just this evolution of individual sources over time. Although these two examples motivate simple distribution functions, the fact that they refer to two different classes of γ emitters emphasises the heterogeneity of the population. This implies that the actual distribution function is probably anything but trivial. However, as a working hypothesis, the following assumptions were made in this work

- Source properties are independent of a source's location within the Galaxy, i.e. the distributions of luminosities and radii are separately modelled from the spatial distribution of sources.
- The distributions of luminosities and radii are independent of each other.
- Both the distribution of luminosities and the distribution of radii follow power laws.

This way, the joint probability density function (PDF) $P(L, R)$ is written as

$$P(L, R) = N \left(\frac{L}{L_0}\right)^{\alpha_L} \left(\frac{R}{R_0}\right)^{\alpha_R} , \tag{6.4}$$

with scaling factors L_0, R_0 and a normalisation factor N which depends on the boundaries of L and R.

These boundaries are derived from sources with known distances in the HGPS sample. For extended sources the radius is calculated directly. Point-like sources can be used to further constrain the range with upper limits. For these point-like sources, a radius is either calculated based on the upper limit of their angular extent whenever available or based on the systematic uncertainty of the size of the PSF (0.03°) alternatively. This way, the range of radii is defined to be $10^{0.1}\,\mathrm{pc} \leq R \leq 10^{1.7}\,\mathrm{pc}$, i.e. almost two orders of magnitude. The range of luminosities covers almost three orders of magnitude, $10^{32.1}\,\mathrm{ph\,s^{-1}} \leq L \leq 10^{34.8}\,\mathrm{ph\,s^{-1}}$ (*cf.* with a dynamical range of three orders of magnitude chosen by [181] and five orders of magnitude chosen by [127]). The scale factors in Eq. 6.4 are set to $R_0 = 1\,\mathrm{pc}$ and $L_0 = 10^{34}\,\mathrm{ph\,s^{-1}}$. This leaves α_L and α_R as the only two free parameters of the model.

Regarding the luminosity, throughout this work, the VHE photon rate, that is the number of photons above 1 TeV per second, is used as proxy for the luminosity of a source, corresponding to the integral photon flux from which it is derived. For a conversion to the more familiar unit of erg/s, the photon rate can be simply scaled with a characteristic mean energy per photon. However, this mean energy per photon depends on the spectral shape of the sources and it is refrained from invoking an additional assumption about this spectral shape on the presented results. Just to give an idea about the magnitude of this quantity, the mean energy per photon is $3.52\,\frac{\mathrm{erg}}{\mathrm{photon}}$ if the sources are assumed to follow a power law spectrum in the energy range 1 TeV-10 TeV with a mean spectral index of -2.4 as found for the HGPS source sample.

6.2 Observation bias of the HGPS sample

The data used for the generation of the HGPS catalogue translates into a spatially inhomogeneous sensitivity. This is illustrated in Fig. 6.3 showing the detection horizon for point-like sources with a luminosity of $10^{33}\,\frac{\mathrm{photons}}{\mathrm{s}}$ on top of a source distribution according to the

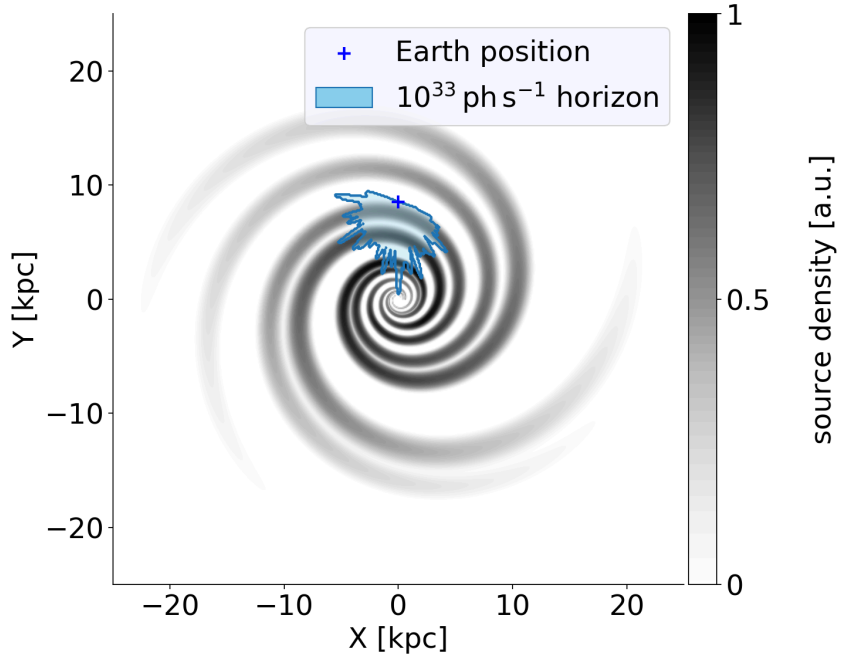


Figure 6.3: *Sensitivity range of the HGPS at Galactic latitude $b = 0°$ for point-like sources with an intrinsic luminosity of 10^{33} $\frac{\text{photons}}{\text{s}}$. The spiral structure depicts an assumed source density according to model mSp4.*

spiral-arm model mSp4. In this figure it can be seen that the sensitivity is increased especially tangentially to the spiral arms. This means that the assumption of a homogeneous sensitivity with a global average value would underestimate the actually explored proportion of the population for an mSp4 distribution. Similar effects arise for the other spatial distribution models.

Two sensitivity maps are provided with the HGPS catalogue which contain the spatially varying thresholds for the minimal detectable flux of point-like sources. One map is generated from an analysis applying a 0.1° correlation radius, the other one likewise for a correlation radius of 0.2°. The larger correlation radius corresponds to the average angular extent of sources detected in the HGPS. The smaller correlation radius is appropriate for point-like source in accordance with an average PSF of $\sigma_{PSF} = 0.08°$. For a fixed background level the sensitivity decreases with the angular extent of a source, i.e. the minimal detectable flux scales as $F_{min}(\sigma_{source}) \propto \sqrt{\sigma^2_{source} + \sigma^2_{PSF}}$ [88]. While the threshold for point-like sources that is found in the sensitivity map with correlation radius 0.1° is expectedly lower than the threshold given in the map with correlation radius 0.2°, there is no smooth transition, e.g. with the scaling factor for source extent, between both. For a consistent treatment of all simulated sources in the population synthesis, thresholds are taken only from the map with 0.2° correlation radius in accordance with the observation that most HGPS sources are extended.

Next to the spatial variation of the sensitivity, specifically its dependency on Galactic longi-

tude l and latitude b, and its scaling with the angular extent of sources, another factor has to be taken into account for an accurate estimation of the observational bias. The limited field of view of IACTs in addition to the ring background method used in the HGPS for estimating the background from within the field of view renders the detection of sources larger than 1° impossible. This results in a threshold described by

$$F_{min}(l, b, \sigma_{source}) = \begin{cases} F_{min,0}(l, b)\sqrt{\frac{\sigma^2_{source}+\sigma^2_{PSF}}{\sigma^2_{PSF}}}, & \sigma_{source} \leq 1^\circ \\ \infty, & \sigma_{source} > 1^\circ \end{cases} \tag{6.5}$$

This description gives an approximation that is as close as possible to the actual sensitivity of the HGPS with the data available. However, there are some discrepancies which are already discussed in [88]. In particular, the detection threshold for the sources listed in the catalogue is a TS value of 30, which corresponds to a significance of 5.5σ for an isolated, point-like source, compared to a threshold of 5.0σ used in the sensitivity maps. Furthermore, the null hypothesis used for source detection in the HGPS includes a large-scale diffuse emission model in addition to the background model, while only the background model was used for the calculation of the sensitivity maps. That is why the values provided by these maps must partly be considered as too optimistic. However, no information on their uncertainties is provided.

When using the sensitivity map with correlation radius 0.2° and applying Eq. 6.5 to the HGPS sample, 62 out of the 78 sources surpass the minimal flux threshold. These sources are listed in Table 6.3. As can be seen in the table, most of them are extended. All but three of the sources that do not exceed the threshold are point-like sources. Two of these three extended sources are marked as *extern*, i.e. they were detected with a different analysis chain and therefore the specified flux threshold does not apply. The other extended source is HESS J1813-126, an unidentified source located at a comparatively high latitude ($b = 2.49^\circ$), where the exposure is typically low. However, with an extent of 0.21° it should nevertheless fall within the specified sensitivity range. This discrepancy could be due to estimating the background only at the centre of a given field of view as done for the generation of the maps with 0.2° correlation radius, leading to gradients in the exposure [88].

Table 6.3: *List of HGPS sources exceeding the threshold defined by Eq. 6.5 using the minimum flux values from the sensitivity map for a* 0.2° *correlation radius. All source attributes were taken from [88]. Errors or pairs of minimum and maximum values are given in brackets, except for those external sources where this information is not available. Sources that additionally meet the criteria for the likelihood estimation (see section 6.3) are highlighted.*

H. E. S. S. Name	Class	Integral Flux 10^{-12} [cm^{-2} s^{-1}]	Extent 10^{-1} [deg]	Distance [kpc]
HESS J0835-455	PWN	15.36 (0.53)	5.85 (0.52)	0.28 (0.25, 0.30)
HESS J0852-463	SNR	23.39 (2.35)	10.00 (-)	0.75 (0.50, 1.00)
HESS J1018-589 B	-	0.83 (0.17)	1.50 (0.26)	-
HESS J1023-575	-	2.56 (0.17)	1.66 (0.09)	-
HESS J1026-582	-	0.69 (0.19)	1.30 (0.39)	-

Table 6.3: *(continued)*

H. E. S. S. Name	Class	Integral Flux 10^{-12} [cm^{-2} s^{-1}]	Extent 10^{-1} [deg]	Distance [kpc]
HESS J1119-614	Composite	0.87 (0.13)	0.98 (0.14)	8.40 (8.00, 8.80)
HESS J1302-638	Binary	0.40 (0.05)	-	2.79 (1.39, 5.58)
HESS J1303-631	PWN	5.26 (0.27)	1.77 (0.15)	6.65 (3.33, 13.30)
HESS J1356-645	PWN	5.53 (0.53)	2.31 (0.20)	2.50 (1.25, 5.00)
HESS J1418-609	PWN	3.01 (0.31)	1.08 (0.11)	5.60 (2.80, 11.20)
HESS J1420-607	PWN	3.28 (0.24)	0.81 (0.06)	5.61 (2.81, 11.22)
HESS J1427-608	-	0.74 (0.10)	-	-
HESS J1442-624	SNR	2.44 (0.67)	1.90 (0.30)	2.75 (2.30, 3.20)
HESS J1457-593	-	2.50 (0.40)	3.32 (0.45)	-
HESS J1458-608	-	2.44 (0.30)	3.73 (0.31)	-
HESS J1503-582	-	1.89 (0.28)	2.82 (0.33)	-
HESS J1507-622	-	2.99 (0.31)	1.75 (0.17)	-
HESS J1514-591	PWN	6.43 (0.21)	1.45 (0.26)	5.20 (3.80, 6.60)
HESS J1614-518	-	5.87 (0.42)	1.80 (0.10)	-
HESS J1616-508	-	8.48 (0.44)	2.32 (0.35)	-
HESS J1626-490	-	1.65 (0.33)	1.97 (0.35)	-
HESS J1632-478	-	2.93 (0.51)	1.82 (0.20)	-
HESS J1634-472	-	2.90 (0.37)	1.74 (0.13)	-
HESS J1640-465	Composite	3.33 (0.19)	1.10 (0.34)	10.50 (8.00, 13.00)
HESS J1646-458	-	5.48 (0.46)	5.03 (0.30)	-
HESS J1702-420	-	3.91 (0.65)	1.97 (0.25)	-
HESS J1708-410	-	0.88 (0.09)	0.64 (0.06)	-
HESS J1708-443	-	2.28 (0.32)	2.79 (0.31)	-
HESS J1713-381	-	0.65 (0.13)	0.92 (0.17)	-
HESS J1713-397	SNR	16.88 (0.82)	5.00 (-)	3.50 (1.00, 6.00)
HESS J1714-385	Composite	0.25 (0.05)	-	7.90 (6.30, 9.50)
HESS J1718-385	-	0.80 (0.14)	1.15 (0.15)	-
HESS J1729-345	-	0.86 (0.17)	1.89 (0.31)	-
HESS J1731-347	SNR	2.01 (0.15)	2.70 (0.20)	3.20 (2.40, 4.00)
HESS J1745-290	-	1.70 (0.08)	-	-
HESS J1745-303	-	0.94 (0.21)	1.79 (0.20)	-
HESS J1746-285	-	0.15 (0.05)	-	-
HESS J1746-308	-	0.68 (0.22)	1.62 (0.36)	-
HESS J1747-248	-	0.29 (0.05)	0.61 (0.12)	-
HESS J1747-281	PWN	0.60 (0.13)	-	9.25 (8.50, 10.00)
HESS J1800-240	-	2.44 (0.35)	3.16 (0.39)	-
HESS J1804-216	-	5.88 (0.27)	2.43 (0.34)	-
HESS J1809-193	-	5.27 (0.29)	4.04 (0.48)	-
HESS J1813-178	Composite	1.98 (0.15)	0.49 (0.04)	4.80 (4.50, 5.10)
HESS J1825-137	PWN	18.41 (0.56)	4.61 (0.32)	3.93 (1.97, 7.86)
HESS J1826-130	-	0.86 (0.17)	1.52 (0.21)	-
HESS J1826-148	Binary	1.28 (0.04)	-	2.50 (2.40, 2.60)

Table 6.3: *(continued)*

H. E. S. S. Name	Class	Integral Flux 10^{-12} [cm^{-2} s^{-1}]	Extent 10^{-1} [deg]	Distance [kpc]
HESS J1828-099	-	0.43 (0.07)	-	-
HESS J1833-105	Composite	0.39 (0.07)	-	4.70 (4.30, 5.10)
HESS J1834-087	Composite	3.34 (0.24)	2.10 (0.37)	4.20 (3.90, 4.50)
HESS J1837-069	PWN	12.05 (0.45)	3.55 (0.31)	6.60 (3.30, 13.20)
HESS J1841-055	-	10.16 (0.42)	4.08 (0.33)	-
HESS J1843-033	-	2.88 (0.30)	2.39 (0.63)	-
HESS J1846-029	Composite	0.45 (0.05)	-	8.45 (6.30, 10.60)
HESS J1848-018	-	1.74 (0.35)	2.48 (0.32)	-
HESS J1849-000	PWN	0.53 (0.09)	0.90 (0.15)	7.00 (3.50, 14.00)
HESS J1852-000	-	1.30 (0.25)	2.78 (0.42)	-
HESS J1857+026	-	3.77 (0.40)	2.59 (0.56)	-
HESS J1858+020	-	0.53 (0.11)	0.79 (0.16)	-
HESS J1908+063	-	6.53 (0.50)	4.86 (0.27)	-
HESS J1912+101	-	2.49 (0.35)	3.20 (0.20)	-
HESS J1923+141	-	0.78 (0.15)	1.20 (0.19)	-

For a population synthesis approach like the one followed here, it is essential that observations and simulations are described consistently. This work relies on the description of the HGPS sensitivity given by Eq. 6.5 and corresponding values from the sensitivity map for a 0.2° correlation radius. This implies that only the sources listed in table 6.3 (or a specific subset of them) can be used for the following analysis. This sample includes all classes of firmly identified sources. These sources provide distance estimates that can be used to calculate luminosities and radii or upper limits for radii.

6.3 Parameter estimation

6.3.1 Method

All parameters of the model but the two indices α_L and α_R in Eq. 6.4 are fixed. One way to estimate these two parameters is to tune them such that they maximise the likelihood for deriving the observed number of sources which are detected with the HGPS. The number of detected sources depends on the modelled distributions $P(\mathbf{x})$ and $P(L, R)$ as follows

$$N_{det}(L,R) = N_{FoV}\, P(L,R) \int_0^\infty \mathrm{d}F \int_0^\infty \mathrm{d}\sigma \int_{V_{FoV}} \mathrm{d}\mathbf{x}\, P(\mathbf{x}) \cdot \mathcal{H}\left(F - F_{min}(l(\mathbf{x}), b(\mathbf{x}), \sigma)\right) \cdot \delta\left(F - \frac{L}{4\pi d^2(\mathbf{x})}\right) \cdot \delta\left(\sigma - \tan^{-1}\left(\frac{R}{d(\mathbf{x})}\right)\right), \quad (6.6)$$

where N_{FoV} is the number of all sources in the field of view covering the volume V_{FoV}. This number of sources is unknown and therefore enters the likelihood estimation as an additional free parameter. Furthermore, the distance to the observer in Eq. 6.6 is denoted by d. With the expression inside the Heaviside function $\mathcal{H}$, the cut with respect to the minimal flux

threshold finds its way into the equation, while the expressions inside the two Dirac delta functions δ provide the relation between flux and luminosity and angular extent and radius respectively. The integral in Eq. 6.6 describes the sensitivity of the HGPS and, thus, determines its observational bias. In the following it is summarised as a correction function $C(L, R)$.

For the estimation of the model parameters, the probed $L \times R$ phase space is split into bins of size 0.1×0.1 in $\log_{10}$ space. With Eq. 6.6 the number of detected sources $N_{pred,i}$ can be predicted for each bin i

$$N_{pred,i} = N_{FoV} \int_{L_{min,i}}^{L_{max,i}} \mathrm{d}L \int_{R_{min,i}}^{R_{max,i}} \mathrm{d}R \, P(L, R) \, C(L, R) \, . \tag{6.7}$$

The number of actually detected sources per bin $N_{true,i}$ is expected to follow a Poissonian distribution $P_{\lambda_i}(N_{true,i})$ with $\lambda_i = N_{pred,i}$. Thus, the model parameters are found by minimising the negative log-likelihood

$$\mathcal{L} = -\sum_i \log\left(P_{N_{pred,i}}\left(N_{true,i}\right)\right) \, . \tag{6.8}$$

Since this approach requires that the luminosities and radii of the detected sources are known, only a subset of the HGPS sample can be used. First, point-like sources are rejected because a measured angular extent is needed to infer the radius of the source.The minimum measurable extent is not well defined for the HGPS. An extent is claimed if the measured extent σ_{source} satisfies the condition $\sigma_{source} - 2\Delta\sigma_{source} > \sigma_{sys}$, where $\Delta\sigma_{source}$ is the uncertainty of the measured extent and σ_{sys} is the systematic error of the size of the PSF [88]. In this work, a fixed value for the minimum angular extent is used instead to allow consistent treatment of data and simulations. This threshold is set to the mean size of the PSF. This is a conservative choice and reduces the sample from 62 sources to 48 sources. Second, a distance estimate is required for the calculation of both luminosity and radius. Distance estimates are available for only few sources that have associations with astronomical objects that are known from observations in other wavelengths. This limitation further reduces the sample down to 16 sources, which are highlighted in Table 6.3. This subset comprises nine PWNe, four SNRs and three composite objects, but no binary since all of the three binaries included in the HGPS are point-like.

In addition to the observational bias inherent to the HGPS catalogue, this selection alters the bias present in the analysed sample. For the likelihood estimation, the exclusion of point-like sources is accounted for by setting the sensitivity threshold for sources that fall below σ_{PSF} to infinity, i.e. adjusting Eq. 6.5 as

$$F_{min}^{\text{extended}}(\sigma_{source}) = \begin{cases} \infty, & \sigma_{source} < \sigma_{PSF} \\ F_{min,0}\sqrt{\frac{\sigma_{source}^2 + \sigma_{PSF}^2}{\sigma_{PSF}^2}}, & \sigma_{PSF} \leq \sigma_{source} \leq 1^\circ \\ \infty, & 1^\circ < \sigma_{source} \, . \end{cases} \tag{6.9}$$

The association of a source with an astronomical object, which yields the distance estimate, is assumed to be independent of the position, luminosity and radius of the source. When

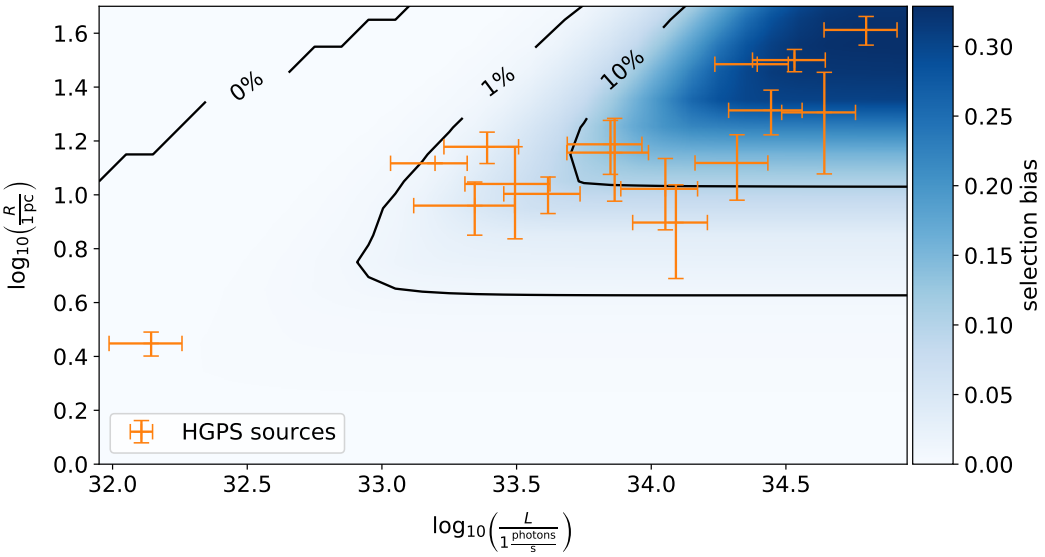


Figure 6.4: *Selection bias as defined by* $\frac{16}{48} \cdot C(L_i, R_i)$ *and calculated for a source spatial distribution following mSp4. HGPS sources that are selected for the likelihood estimation are represented by crosses. Errors on the luminosities and radii of sources are derived from errors given on their fluxes and extents if available. Errors on the distance estimate are not considered here.*

calculating $N_{pred,i}$ according to Eq. 6.7 and Eq. 6.9, this additional selection is, therefore, accounted for by simply multiplying the result with a constant factor of $16/48$.

With this description of the selection bias, the expected number of sources according to Eq. 6.7 has been approximated numerically in the following way. The correction function $C(L, R)$ is estimated, first, by assigning each source of the set of simulated sources S_{sim} (see Sec. 6.1.1) the same luminosity L and source radius R and calculating the corresponding fluxes, angular extents, and positions on the sky as they would be observed at Earth. Then, with the threshold given by Eq. 6.9, the sources that would be detected in the HGPS are selected to form the subset $S_{det}(L, R) \subset S_{sim}$. For a source density ρ corresponding to an assumed spatial model the correction function is calculated as

$$C(L, R) = \sum_{\mathbf{x} \in S_{det}(L,R)} \rho\left(\mathbf{x}\right) / \sum_{\mathbf{x} \in S_{sim}} \rho\left(\mathbf{x}\right) . \tag{6.10}$$

This numerical approach is a computationally expensive procedure. Since the correction function is only slowly varying, it is treated as a constant within each bin i and evaluated only at the bin centre. In this way, the expected number of sources per bin is finally calculated as

$$N_{pred,i} = N_{FoV} \cdot \frac{16}{48} \cdot C(L_i, R_i) \int_{L_{min,i}}^{L_{max,i}} dL \int_{R_{min,i}}^{R_{max,i}} dR\, P(L, R)\,, \tag{6.11}$$

The selection bias defined by $\frac{16}{48} \cdot C(L_i, R_i)$ is shown in Fig 6.4 for the spatial model mSp4. There it can be seen that the detectability of a source generally increases with its luminosity. For small radii a large fraction of distant sources is disregarded because of the selection of extended sources. With large radii, on the other hand, the detection threshold for nearby sources increases with a cut-off of sources exceeding an angular extend of 1°. For a given luminosity both effects give rise to an optimal radius for source detectability which shifts

towards larger values with increasing luminosities. Overall the selection bias exhibit a diagonal shape. The correction functions derived for the other spatial models exhibit similar shapes. In addition to the selection bias, the distribution of sources from the HGPS used for the likelihood estimation is shown in Fig 6.4. Errors that are shown for individual sources are derived from the measured errors on their flux and angular extent if these are provided in the HGPS catalogue. Errors on the distance estimate are not considered in this plot. It can be seen that the distribution of these sources roughly follows the diagonal shape of the estimated selection bias.

6.3.2 Monte Carlo verification

The capability of this likelihood approach to reconstruct properties of the parent population of sources from a small sample of sources observed under the conditions of the HGPS has been studied by means of Monte Carlo simulations. Populations of Galactic VHE γ-ray sources are simulated by randomly sampling N points from S_{sim}, which are weighted according to one of the spatial models defined above. Subsequently, each of these hypothetical sources is assigned a luminosity and radius according to the distribution defined by Eq. 6.4 with some fixed parameters. In order to match the HGPS sample, the total number of sources of the synthetic population N is adjusted for a each tuple of parameters (α_L, α_R) such as to yield on average 48 extended sources when the sensitivity threshold defined by Eq. 6.9 is applied. From those sources that surpass the threshold, 16 are randomly selected to reconstruct the input parameters.

To take a detailed look at the performance of the method, $\sim$ 1500 synthetic populations were simulated, based on the spatial distribution model mPWN and with fixed parameters $\alpha_L = -2$ and $\alpha_R = 0$. According to these model parameters, each synthetic population consists of 5800 hypothetical sources. Their luminosity and radius distributions are shown in Fig. 6.5 in the top and middle panels, respectively. The solid line represents the median over all populations, the shaded area the interquartile range. Similarly, the distributions of sources exceeding the sensitivity threshold of the HGPS are represented by the dashed lines. As an effect of the observational bias, their distribution deviates clearly from the distribution of the global population. Nevertheless, with a sample of only 16 detectable sources, the method presented here allows an unbiased reconstruction of the input parameters. This can be seen in the distributions of the three reconstructed parameters shown in the lower panels of Fig. 6.5, which are all centred on their respective true values. Accordingly, the estimated luminosity and radius distributions (dotted lines in the top and middle panels), which are based on the reconstructed parameters, agree well on average with those of the input distribution.

To evaluate the method on a larger scale, synthetic populations were simulated and then reconstructed for a number of different (α_L, α_R) pairs. More specifically, for each spatial distribution model $5 \cdot 10^4$ populations were generated, where the input parameters $\alpha_{L,true}$ and $\alpha_{R,true}$ were taken independently at random from a uniform distribution in the interval (-3, 2). For the luminosity function, the distribution of the reconstruction error, defined as $\Delta\alpha_L = \alpha_{L,true} - \alpha_{L,reco}$, over the reconstructed value $\alpha_{L,reco}$ is shown in the left panels of Fig. 6.6 for all spatial models. This distribution is shown for four different intervals of the paired $\alpha_{R,reco}$ values. Regardless of the $\alpha_{R,reco}$ interval or the spatial distribution model, the mean value of the reconstruction error $\Delta\alpha_L$, represented by the lines, is close to zero for $\alpha_{L,reco}$ in the interval (-2.5, 1), i.e. for most of the parameter range. Deviations from a

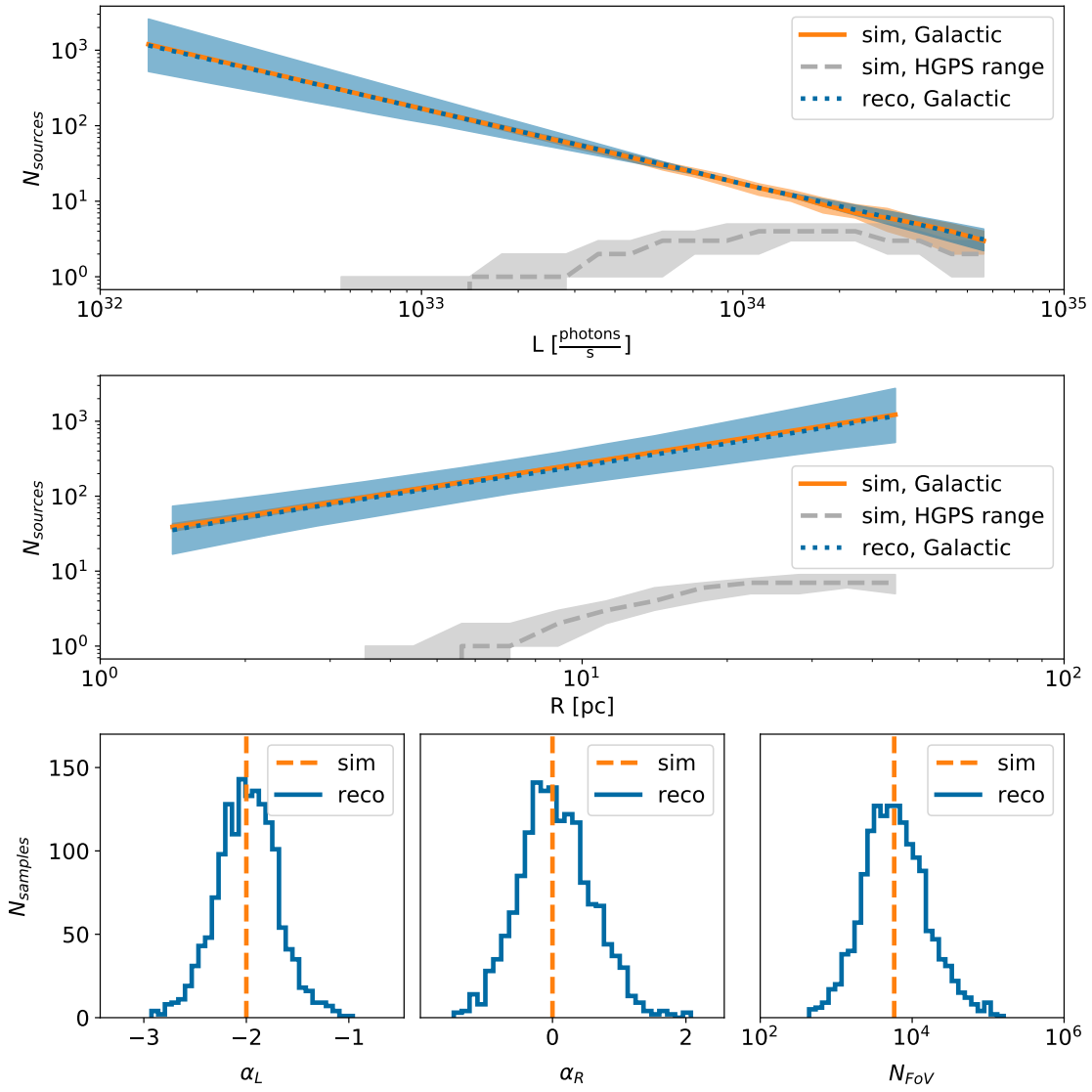


Figure 6.5: *An example of the reconstruction of model parameters for mPWN with the likelihood approach. Top: Luminosity distribution of simulated sources (solid line), the subset of detectable sources (dashed line) and estimated source distribution based on the reconstructed parameters (dotted line). Middle: Radius distribution corresponding to the top panel. Bottom: Distribution of reconstructed parameters for fixed input values. More details given in the text.*

zero-centred reconstruction error only occur at the edges of the probed interval and are an expected statistical effect. The uncertainty of the reconstruction value is represented by the shaded area showing the standard deviation of the reconstruction error for a given reconstruction value. Consistent for all spatial models, the uncertainty increases with increasing $\alpha_{L,reco}$ values. There is a weak dependence on the paired $\alpha_{R,reco}$ values, i.e. smaller uncertainties are observed for the two intervals around $\alpha_{R,reco} = -0.5$ than for the two intervals at the outer parameter range. For all variations, the uncertainty of the reconstruction error of α_L for $-2 \leq \alpha_L \leq 1$ spans a range between 0.27 and 0.55. For the radius function, Fig. 6.6 also shows the distribution of the reconstruction error of $\Delta\alpha_R$ as a function of $\alpha_{R,reco}$ in the panels on the right. The overall characteristic of this distribution is the same as for $\alpha_{L,reco}$.

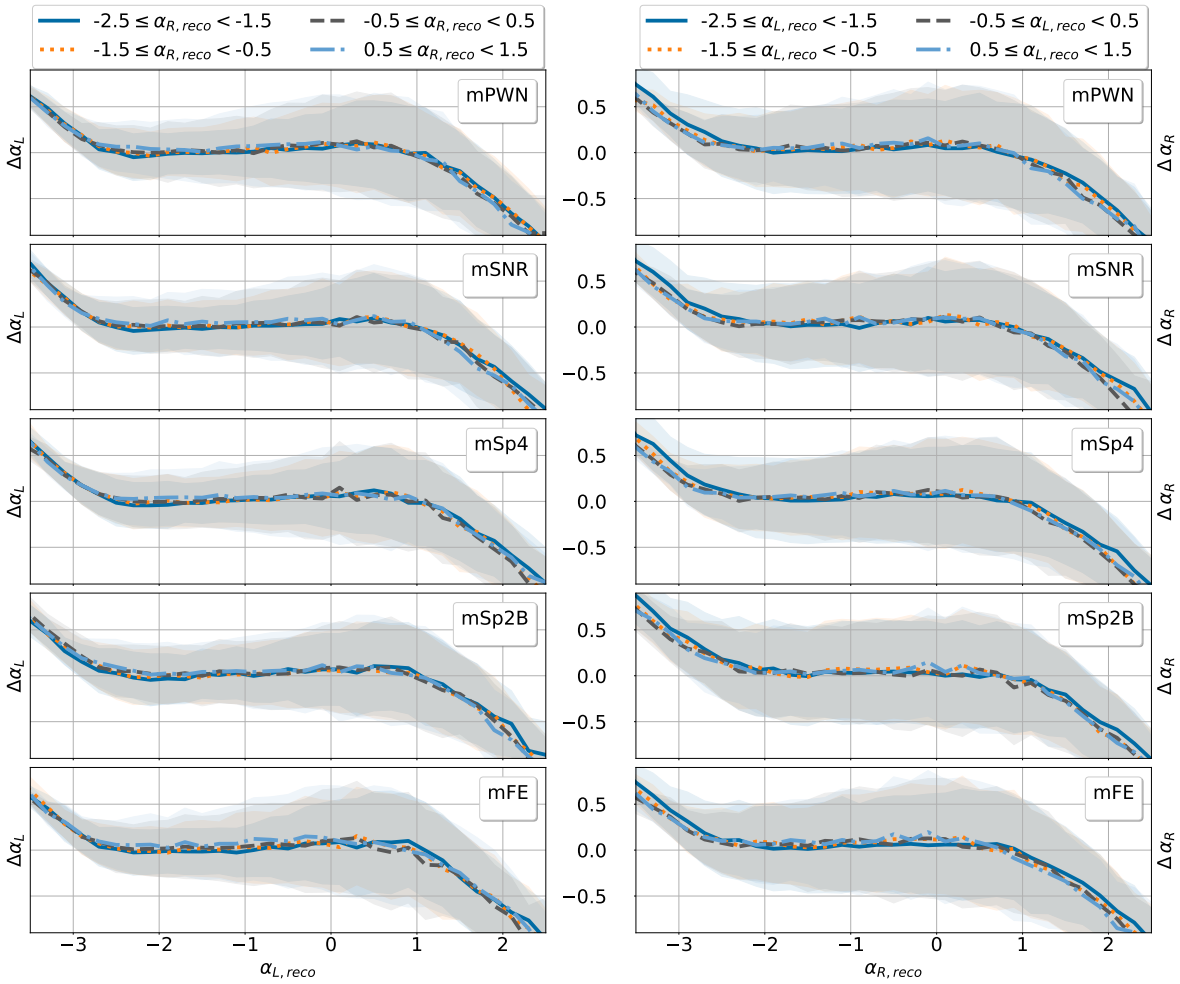


Figure 6.6: *The errors of the reconstructed parameters* $\Delta\alpha_L = \alpha_{L,true} - \alpha_{L,reco}$ *(left) and* $\Delta\alpha_R = \alpha_{R,true} - \alpha_{R,reco}$ *(right), calculated with the likelihood method. Details are given in the text.*

Only the uncertainties are larger and lie between 0.42 and 0.62 for α_R for $-2 \leq \alpha_R \leq 1$.

These tests show that the method presented in this section is well suited to reconstruct the properties of a population following the proposed distribution functions for the luminosity and radius of the sources. In particular, the method correctly accounts for the observational bias inherent in the HGPS. It is shown that the reconstructed values of α_L and α_R are unbiased over a large parameter space. However, the uncertainties in the reconstruction values are considerable, which is not surprising given the small sample of sources used for the reconstruction.

6.3.3 Result

After validating the method, it was applied to the HGPS sample to derive the parameters of the luminosity and radius functions of the Galactic population of VHE γ-ray sources. To account for both the uncertainty in the input values and the uncertainty arising from the reconstruction method, the following procedure was applied. First, bootstrapping is applied to generate a set of 1000 samples from the original HGPS sample, where the values of source

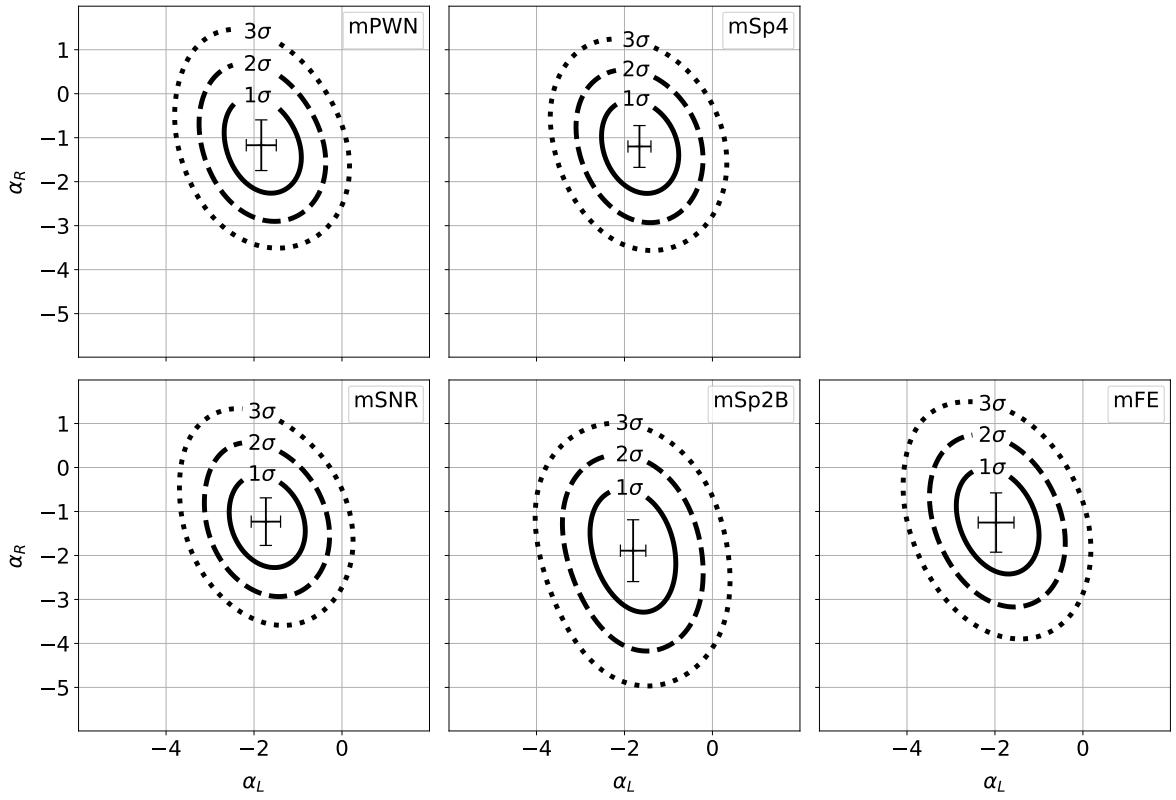


Figure 6.7: *Best estimates and confidence intervals of α_L and α_R as calculated with the likelihood approach. The position of the cross marks the best estimate values. Error bars reflect the uncertainty corresponding to the errors of the measured source properties. Lines show the 1σ, 2σ and 3σ confidence intervals for the true values of α_L and α_R.*

properties in these samples vary according to the uncertainties in their measured values in the original sample. In this process, the flux values are sampled from a Gaussian distribution centred on the respective measured value and having a variance that is the square sum of the statistical error and a systematic error of 30 % of the flux, as proposed in [88]. Similarly, the angular extents are sampled from Gaussian distributions. Here, a fixed value of 0.03°, which is the average uncertainty of the PSF size, is used as the systematic error. The HGPS catalogue does not provide errors for distance estimates, but minimum and maximum values of the range in which the source can be located. In general, these do not correspond to normal distributed errors around the best estimate. To account for the uncertainty in the source distances, a conservative approach was taken in which the distances were taken from a uniform distribution within the interval defined by the minimum and maximum values of the distance estimates. Luminosities and radii were calculated from the randomly sampled fluxes, extents and distances and used to reconstruct the model parameters. The mean and standard deviation of the reconstructed parameters thus derived for the different spatial distribution models are shown in Fig. 6.7 by the position of the marker and the size of the error bars, respectively.

In a second step, a kernel density estimation is applied to the set of reconstructed values to derive confidence intervals for the true values of α_L and α_R, taking into account the uncertainty resulting from the reconstruction method. A Gaussian kernel was used for this

purpose. The variance of the Gaussian kernel is determined by the quadratic sum of the reconstruction errors of α_L and α_R, which correspond to the respective values derived from the sample mean in the previous step. Confidence intervals were calculated from the resulting probability density function. Contours of the 1σ, 2σ and 3σ confidence intervals are also shown in Fig. 6.7. The numerical values of the best estimates for α_L and α_R and the errors derived from the 1σ confidence interval are given in Table 6.4.

Table 6.4: *Reconstructed values of α_L & α_R using the likelihood method. Errors are derived from the 1σ confidence interval as shown in Fig. 6.7.*

Model	α_L	α_R
mPWN	-1.83 ± 0.85	-1.17 ± 1.09
mSNR	-1.73 ± 0.84	-1.23 ± 1.03
mSp4	-1.66 ± 0.86	-1.20 ± 1.04
mSp2B	-1.80 ± 0.94	-1.89 ± 1.41
mFE	-1.97 ± 0.90	-1.25 ± 1.16

These values are little affected by the choice of the boundaries of the range of luminosities and the range of radii. Different boundaries were investigated and lead to deviations of the reconstruction values that are smaller than the reconstruction error. The largest effect was seen for the upper bound of the luminosity. However, this upper bound is well constrained. According to the HGPS sensitivity, a source with $L = 10^{34.8} \frac{\text{photons}}{\text{s}}$ within the FoV has a detection probability close to one, i.e. it is unlikely that there is a Galactic source with an even higher luminosity missed by the HGPS.

The reconstructed value of the index of the luminosity function α_L lies in the range between -1.64 and -1.93 for the different spatial models. In comparison to this result, previous studies on the VHE source population, which do not take into account source sizes or sensitivity inhomogeneity, have estimated the luminosity function to follow a power law with a harder index (e.g. $-1.5 \leq \alpha_L \leq -1.0$; [87]). However, with the large uncertainty of ± 0.88 on average, these results are still compatible with each other. With the exception of the spatial model $mSp2B$, the reconstructed parameter values of the index of the radius function α_R are remarkably consistent among the other four spatial models. The mean value is $\langle \alpha_R \rangle = -1.23$ and, thus, lower than one could expect, for example, for a population of expanding SNRs (see Sec. 6.1.2). Again the uncertainties on the reconstructed values are large, i.e. larger than ± 1. In the next two section the implications of these results for a population of VHE sources are discussed.

6.4 Comparison with observable quantities

At first, predictions of the model for sources in the range of the HGPS sensitivity are compared with the sample of detected sources based on the observable quantities flux, angular extent and position on the sky. For this purpose 3000 synthetic populations were simulated for each spatial model. The total number of sources was chosen such that on average the same number of sources as in the catalogue (62) suffice the detection criterion of the HGPS. The

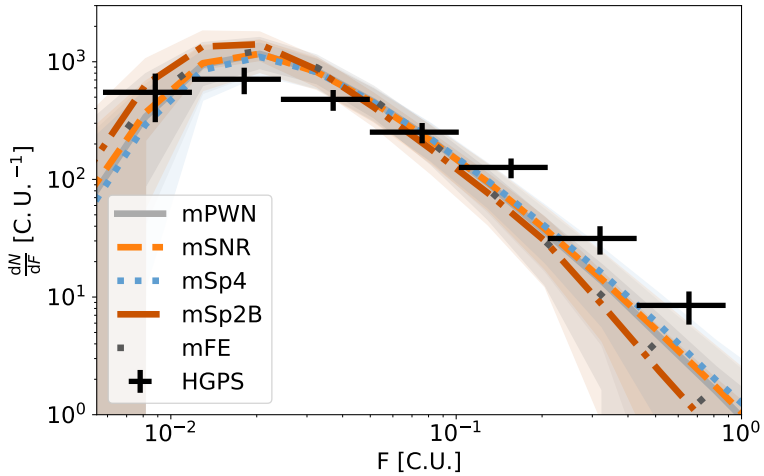


Figure 6.8: *Comparison of the source differential distribution over flux* $\frac{dN}{dF}$ *between synthetic source populations and the HGPS sample. Lines represent the mean over model realisations, shaded region represent the standard deviation. Marker with error bars represent the distribution of HGPS sources. Horizontal bars represent the bin sizes and the vertical bars represent the bootstrapped derived errors on the differential source count.*

total numbers for the different spatial models are listed in Tab. 6.6 and discussed in Sec. 6.5.1. Synthetic sources that fall into the HGPS sensitivity range are then selected to derive the distribution of observable quantities.

6.4.1 Flux & angular extent

The flux distribution is commonly used to describe the sample of detectable sources and to evaluate population models. In addition, the distribution of the angular extent is a valuable measure for this purpose. Both quantities are closely connected due to their dependence on the distance between sources and the observer and, thus, the spatial distribution of sources. Instrumental selection effects, such as the dependency of the sensitivity on the angular extent, additionally can have a notable effect on both distributions. Furthermore, an intrinsic correlation between the luminosity and radius of sources may affects the distributions of these two observable quantities.

The flux distribution, more precisely the differential source number over the flux $\frac{dN}{dF}$ for all sources in the sensitivity range, is shown in Fig. 6.8. The flux is given in C.U., i.e. the integral flux above 1 TeV of the Crab Nebula ($2.66 \cdot 10^{-11}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$). For each spatial model, the mean over the various model realisations is shown by a line, while the standard deviation is shown as a shaded area. The distribution for the HGPS sources is shown by markers with error bars. Here, the horizontal bars represent the bin sizes and the vertical bars represent the bootstrapped derived errors on the differential source count. Although not in perfect agreement, the flux distribution of the simulations is close to observation. There are only minor differences between the different spatial distribution models. Overall, the models give slightly tilted distributions compared to the observation, so that there are too many sources at low fluxes ($< 5 \cdot 10^{-2}$ C.U.) and too few at higher fluxes ($> 1 \cdot 10^{-1}$ C.U.). A large reconstruction error in the model parameters or an oversimplified model of the luminosity and radius function may be a reason for this mismatch. As for the influence of angular extent,

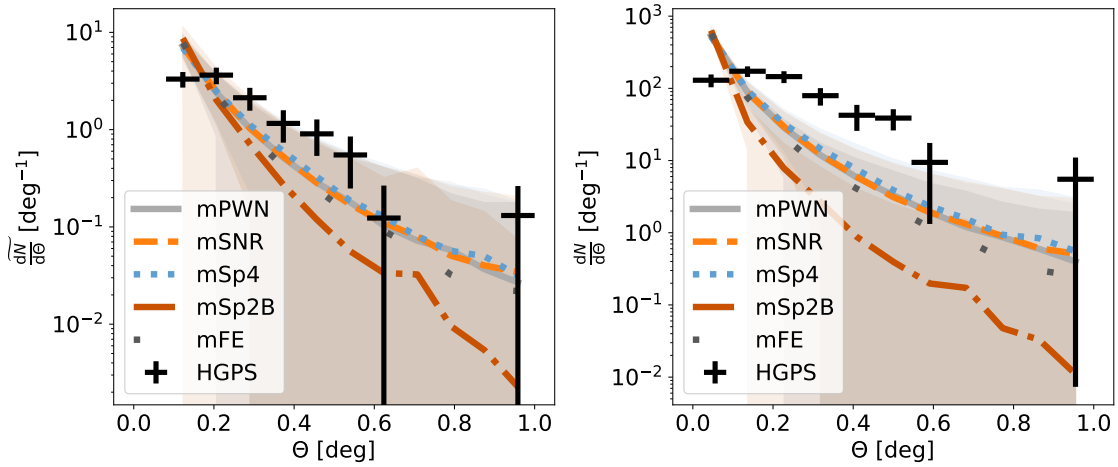


Figure 6.9: *Comparison of the source differential distribution over angular extent* $\frac{\mathrm{d}N}{\mathrm{d}\theta}$ *between synthetic source populations and the HGPS sample as in Fig. 6.8. Left: Normalised distribution of extended sources in the sensitivity range. Right: Non-normalised distribution of all sources in the sensitivity range.*

the large fraction of point-like sources in the simulations (see below) can explain to some extent the overabundance of faint sources, since the detection threshold for point-like sources is lower than for extended sources. On the other hand, the larger fraction of extended sources in the HGPS increases the chances for source confusion. Therefore, the flux of some of these extended sources could be overestimated due to confusion with another source. In addition, the rather optimistic estimate of the detection threshold based on the sensitivity maps of the HGPS (see section 6.2) certainly contributes to this discrepancy between simulation and observation.

In Fig. 6.9 the distribution of angular extent defined as differential source count over extent, $\frac{\mathrm{d}N}{\mathrm{d}\Theta}$, is shown. The left panel depicts the normalised distribution for sources larger than the PSF from simulations and observations, which is denoted $\widetilde{\frac{\mathrm{d}N}{\mathrm{d}\Theta}}$. It is clear that the distribution from the simulations fall too steeply compared to the observation, especially the one for the spatial model mSp2B because of the large distance to the closest spiral arm ($\sim 3\,\mathrm{kpc}$). As a result, the total sample of simulated sources in the HGPS sensitivity range includes many more point-like sources than are observed. This can be seen in the right panel of Fig. 6.9, which shows the non-normalised distribution for all sources in the HGPS sensitivity range. Since the total number of sources for the synthetic populations is adjusted to give (on average) the same total number of sources in the sensitivity range, i.e. not the same number of extended sources, there are too few extended sources in the simulated sample. Compared to the catalogue source sample, where a fraction of 77 % of the sources appear to be larger than the PSF, this is only true for a fraction of $< 27\,\%$ in the simulations. Possible reasons for this discrepancy are manifold. First, the reconstructed parameters of the radius function are poorly constrained. With the small number of sources available for estimating the model parameters, the estimated value of α_R may well not be close to the true value. Also, the modelled distribution itself might be oversimplified, e.g. by assuming independence between source radius and luminosity. In both cases, more accurate parameter estimation is needed to improve the model. Other effects are inherent to the HGPS. Apart from the known

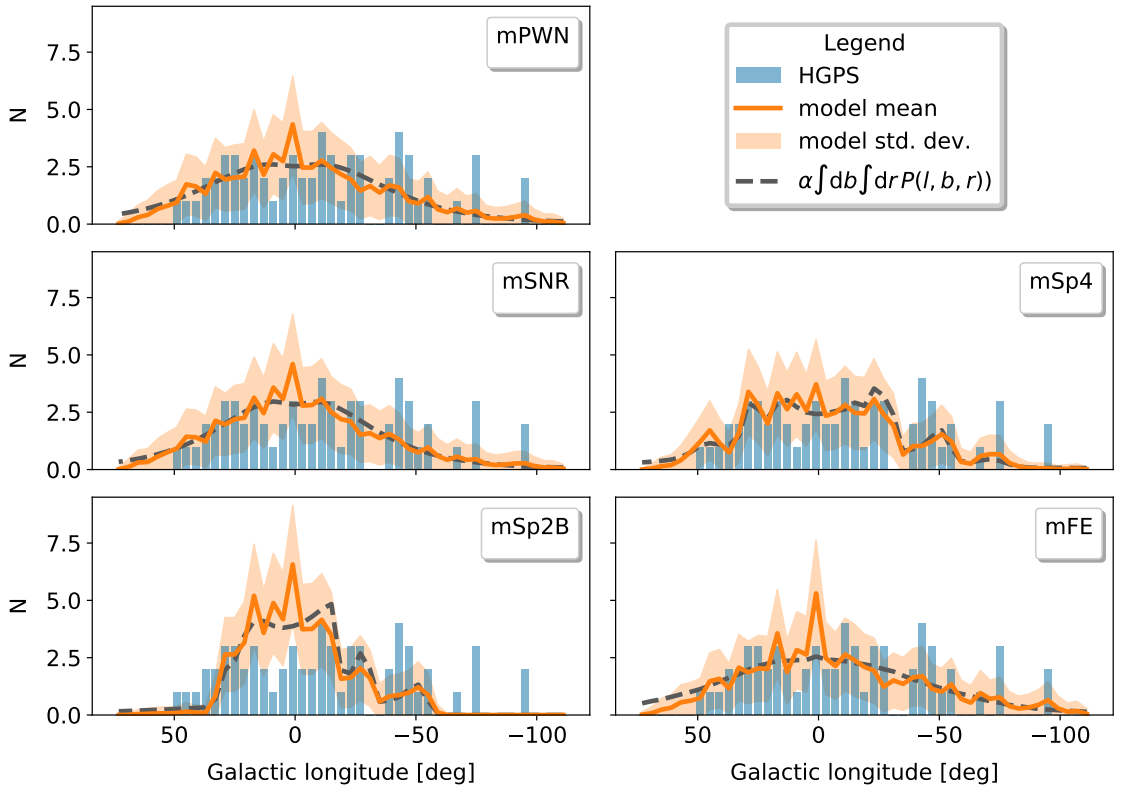


Figure 6.10: *Comparison of the source distribution over Galactic longitude. Description given in the text.*

point-like sources in the catalogue, which are not considered here (see section 6.2), an even larger number of point-like sources could be present in the HGPS but not detected as such, i.e. they are obscured in extended sources due to source confusion [182] or detected and later merged with an overlapping source. To properly account for these effects, synthetic populations must be analysed with a replication of the procedure used for the analysis of the HGPS data, including background and diffuse emission models. While the effect of source confusion on source detectability is worth investigating given the relationship between angular extent and observational bias, this is beyond the scope of this work and is deferred to a later study.

6.4.2 Spatial distribution of sources

The source distributions in Galactic longitude is shown in Fig. 6.10. There, bars show the distribution of the HGPS sources in the sensitivity range. Solid lines represent the mean over the model realisations, the shaded area similarly represents the standard deviation. Dashed lines represent the source density in the field of view, integrated over latitude and source distance and normalised to the number of HGPS sources in the sensitivity range. Thus, differences between the solid line and the dashed line are attributed to the inhomogeneity of the HGPS sensitivity and, thus, visualise its impact on the observed source distribution. All models except mSp2B agree well with the observed distribution. This is confirmed quantitatively using the Kolmogorov-Smirnov statistic. The corresponding p-values are listed in Tab. 6.5, which all exceed at least a 5 % significance threshold for the other four models. For mSp2B, the longitude distribution appears to be somewhat too narrow, falling off too steeply

in the outer parts of the Galactic plane. The result for mSp2B suggests that if the ISM density distribution used as a proxy for the distribution of regions with high star formation rates indeed follows the assumed shape, the VHE source population must include at least one source class evolving outside these regions. Furthermore, all models except mSp4 tend to overpredict the actual source distribution in the central regions, where both source density and exposure in the HGPS are highest. Due to the high source density, source confusion is likely to influence the source count in this region. In addition, the presence of bright diffuse emission in the Galactic ridge region [183] leads to an increased background level that may well cause a deficit of detected sources in this region. While the inclusion of these effects in a future iteration will certainly improve the model, a direct comparison with the HGPS in this particular region must be taken with caution, as it was not analysed with the HGPS pipeline due to the complex morphology, but the results were taken from an external analysis. Interestingly, the source density defined by the mSp4 model agrees well with the observed distribution of sources, considering the position of the peaks in both distributions. Assuming that most of the VHE γ-ray sources are also cosmic-ray sources and that their spatial distribution indeed follows the spiral structure of the Galaxy, corresponding effects on the spatial distribution of cosmic rays would be noticeable and particularly pronounced in the ratio of primary to secondary cosmic rays. In particular, the measured distribution on Earth would not be representative for the entire Milky Way [175]. However, other models provide equally good or even better matches for the source spatial distribution when considering only the simulated sources within the sensitivity range of the HGPS. In fact, due to the small number of sources in combination with the inhomogeneity of the sensitivity, an accurate assessment of the true source distribution is currently not possible.

Table 6.5: *P-values from the comparison of the source distributions over Galactic longitude and latitude with the Kolmogorov-Smirnov test.*

Model	p-value	
	longitude	latitude
mPWN	$2.43 \cdot 10^{-1}$	$1.58 \cdot 10^{-4}$
mSNR	$9.88 \cdot 10^{-2}$	$2.74 \cdot 10^{-5}$
mSp4	$1.09 \cdot 10^{-1}$	$1.93 \cdot 10^{-5}$
mSp2B	$2.72 \cdot 10^{-3}$	$4.88 \cdot 10^{-5}$
mFE	$4.48 \cdot 10^{-1}$	$3.48 \cdot 10^{-5}$

The source distributions in Galactic latitude are shown in Fig. 6.11. There are two prominent deviations between models and observations, namely a clear asymmetry towards negative latitude in the distribution of observed sources in conjunction with a more pronounced peak at latitude 0° which is not recovered by any of the source spatial distribution models. Accordingly, judging from the Kolmogorov-Smirnov statistics, none of the models is compatible with the observation (see Tab. 6.5). Although mFE gives the best fit to the longitude distribution, this model in particular gives a rather flat source distribution over latitude, which is in contrast to the observation. Since the latitude profile is averaged over a large longitude range, it is to be expected that source confusion has little influence on the distribution of observed sources. The sensitivity of the HGPS itself, averaged over the entire longitude range, shows

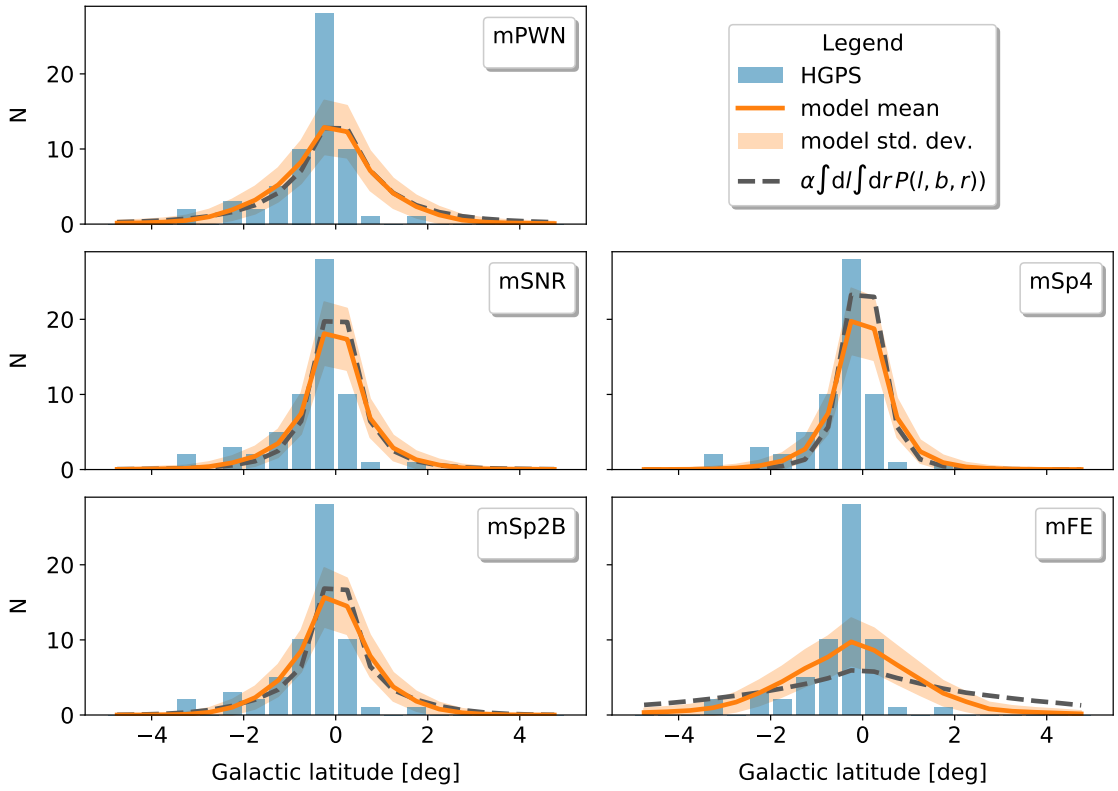


Figure 6.11: *Comparison of the source distribution over Galactic latitude as in Fig. 6.10.*

no pronounced asymmetry. This means that the reason for the asymmetry in the distribution of the observed sources must be sought elsewhere. Such an asymmetry has also been observed for pulsars and can be attributed to the warp of the Milky Way [184]. The link to PWNe suggests that the warp of the Galaxy also contributes to the observed asymmetry in the latitudinal distribution of VHE γ-ray sources. Moreover, all adapted spatial distributions were originally derived assuming that the Sun is at a Galactic height of $z = 0\,\mathrm{pc}$, while its true position is closer to $z = 20\,\mathrm{pc}$ above the Galactic plane [185]. The known physical sources of VHE γ-rays, such as SNRs and PWNe, are associated with star-forming regions and are thus tracers of the thin Galactic disk. If the observed VHE γ-ray sources are indeed located within the thin Galactic disk, the slight offset of the Sun could provide an explanation for the observed asymmetry in the latitudinal source distribution. Both the Galactic warp and the position of the Sun above the Galactic plane thus provide good starting points for improvements to the spatial model and motivate possible, more detailed studies.

6.5 Global properties of the Galactic source population

The comparison between simulations and observations in the previous section shows that although the model does not yet provide a perfect match, it is already approaching a reasonable description of the population. Despite the limited predictive power, the study of the derived properties of the entire population is nevertheless informative. While the different source spatial distribution models have been shown to almost equally well describe the source sample in the sensitivity range of the HGPS, their influence on the global properties of the

population will be the focus of this section.

6.5.1 Number of sources

The total number of VHE γ-ray sources in the Galaxy was estimated based on the number of detected HGPS sources in the sensitivity range, as described in Sec. 6.2, $N_{det} = 62$. With population synthesis, for a given total number of sources in the population N, an expected number of detectable sources $\tilde{n}$ can be calculated and in this way a mapping $f : N \to \tilde{n}$ can be obtained. The likelihood of N_{det} is given by the Poissonian

$$\mathcal{P}_{\tilde{n}}(N_{det}) = \frac{\tilde{n}^{N_{det}} e^{-\tilde{n}}}{N_{det}!}\,, \tag{6.12}$$

so that the probability of the total number of sources can be calculated as $P(N) = \mathcal{P}_{f(N)}(N_{det})$. The estimated number of sources in the Galaxy, i.e. $\arg\max_{N\in\mathbb{N}} P(N)$, for the various spatial distributions are listed in Tab. 6.6 together with errors derived from the 1σ confidence interval.

Table 6.6: *Estimates of the properties of the Galactic VHE γ-ray population. The total number of sources and its errors were calculated as described in the text. The combined luminosity of all sources is almost symmetrically distributed over the different realisations. Therefore, it is well characterised by the mean and standard deviation over the model realisations listed here. In contrast, the median and quartile deviation are used as a more robust description of the total flux, whose distribution is strongly influenced by outliers originating from (rare) nearby sources.*

Model	Sources / Error	Integral Flux / Error 10^{-10} [cm^{-2} s^{-1}]	Luminosity / Error 10^{36} [s^{-1}]
mPWN	1867 / $^{+245}_{-226}$	6.61 / 1.64	2.14 / 0.17
mSNR	1002 / $^{+131}_{-121}$	4.19 / 1.24	1.50 / 0.15
mSp4	785 / $^{+103}_{-94}$	3.02 / 0.63	1.43 / 0.15
mSp2B	988 / $^{+130}_{-119}$	1.67 / 0.17	1.22 / 0.13
mFE	6681 / $^{+880}_{-809}$	13.72 / 1.44	5.54 / 0.23

These numbers vary considerably among the probed source spatial distributions, ranging from 785 sources (mSp4) up to 6681 sources (mFE). Although sources are treated generically as VHE γ-ray emitters in this model, that is no source type is explicitly assumed, a source count as high as seen for model mFE is challenging for the paradigm that SNRs and PWNe are the dominant source classes of VHE γ-ray emission. For instance, with a Galactic supernova rate of one per 40 yr [34] a source count of 7000 implies a maximum age of emitters of $\sim 3 \cdot 10^5$yr. Interestingly, the models mSNR, mSp4, and mSp2B yield very similar results regarding the total number of sources. The similarity between model mSNR and mSp4 is also apparent in the cumulative source distribution over flux (log(N)-log(F)) as shown in Fig. 6.12. In this figure, the distribution of all sources in the HGPS catalogue is compared with the distributions derived with the different spatial models for the entire Galactic population, including non-detectable sources and sources outside the FOV of the HGPS. In Sec. 6.4.1 it was discussed that the models underestimate the number of high-flux sources ($F > 0.1$ C.U.). From Fig. 6.12 it can be seen that all spatial distributions except mSp2B provide high-flux

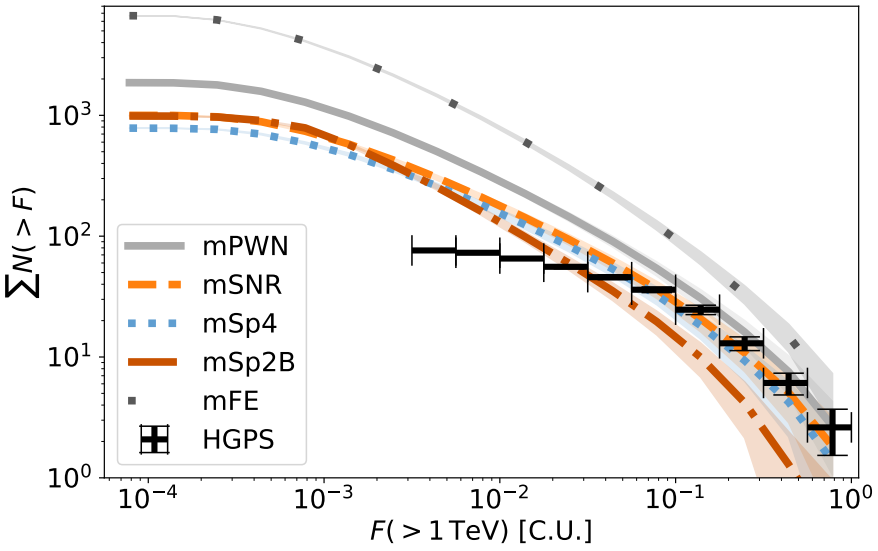


Figure 6.12: *Cumulative source distribution over flux (log(N)-log(F)). The HGPS source distribution is shown by markers with horizontal error bars depicting the bin width and vertical error bars showing the error account for the measurement error on the flux and are derived from bootstrapping. Lines represent the distribution of the entire Galactic source population averaged over the different realisations of the respective model. Shaded regions represent the corresponding standard deviations..*

sources outside the sensitivity range of the HGPS, so that the level of the HGPS is met or exceeded. In particular, broad source spatial distributions such as mPWN and mFE yield a large number of high-flux sources, since the presence of sources close to the observer is increased in these distributions. On the other hand, mSp2B with a minimum source distance of $\sim 3\,\mathrm{kpc}$ generally does not yield a sufficiently high number of sources above 0.03 C.U. to comply with the HGPS distribution.

The source spatial distribution also affects the completeness range of the HGPS, as can be seen in Fig. 6.13. The figure shows the ratio of sources that lie within the HGPS sensitivity range to all sources in the FoV for a varying minimum flux level, with the lines representing the median across the model realisations and the shaded areas representing the interquartile range. In general, this ratio, representing HGPS completeness, decreases with lower flux levels as fainter sources are less likely to be detected. For mPWN, mSNR, mSP4 and mFE, the completeness generally falls below 100 %, starting at 0.2 C.U.. For the two broad spatial distributions, mFE and mPWN, the completeness decreases faster than for the other two. This can also be explained by the increased probability of the presence of sources close to the observer. Since these sources can have a large angular extent, the probability of detection decreases for IACTs, such as. H. E. S. S.. This explanation is supported by the observations that the distribution with the largest fraction of point-like sources, mSp2B, has the highest degree of completeness among all spatial distributions, falling below 100 % only at a flux of 0.08 C.U.. Although these values are not exact due to the imperfections of the model, this comparison highlights the importance of both angular extent and source spatial distribution when studying the overall population, especially when assessing the completeness range of the catalogue, and cautions against the unconditional assumption of completeness up to values of 0.1 C.U., as is often seen in the literature.

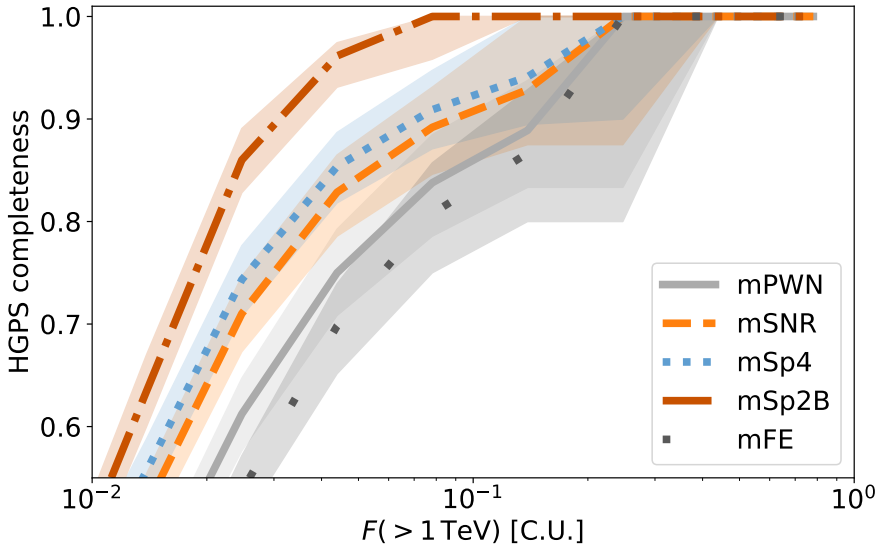


Figure 6.13: *Fraction of detectable sources in the FOV of the HGPS cumulatively over source flux. The shaded regions represent the interquartile range; the lines depict the median.*

Further along this line, one can use the model to estimate the number of sources that exceed the absolute angular extent threshold of 1°. Disregarding mSp2B here, a minimum number of such extended sources within the FoV is given by mSp4 with a mean across model realisations of 1.26 ± 1.10 and a maximum number of 1.82 ± 1.32 by mPWN, i.e. all models give similar results. Extending this estimate to the whole population yields a slightly increased value of 1.86 ± 1.36 for mSp4 (similar to mSNR), a larger increase for mPWN to 8.28 ± 2.91 and a maximum number of 20.04 ± 4.46 given by mFE. In order to study this regime of widely extended sources, either other data analysis techniques or other observational techniques, for example with water Cherenkov telescopes that can use a large FoV, must be exploited. Indeed, two extended sources, Geminga [186, 187], which lies within the FoV of the HGPS, and 2HWC J0700+143 [187], were detected in this way. While this observation is in good agreement with the predictions for mSp4 and mSNR, it is already challenging for mPWN. Since this estimate of very extended sources is independent of instrumental effects, and since it has also been shown that the model generally underestimates the number of extended sources, the above values can be considered as lower limits.

The model was also used to predict the number of sources detectable with the next generation of IACTs, CTA. For this purpose, a homogeneous point-source sensitivity of 2 mCrab was assumed in the longitude range $|l| < 60°$ and latitude range $|b| < 2°$ [188]. With a slightly improved PSF of 0.05° at 1 TeV[2], the same decrease in sensitivity with angular extent as for the HGPS is applied (Eq. 6.5), except for the cut on sources larger than 1° due to the comparatively larger FoV of the instrument. In this way, the predicted number of detectable sources is $\sim$ 570 for mFE or in the range of 200 - 300 for the other four models. Since 70 sources were detected in the HGPS in the region to be observed by the Galactic plane CTA scan, the CTA sample according to these models would increase the current source sample by

[2]Value taken from Monte Carlo simulations of the CTA Observatory performance (prod3b-v2), `https://www.cta-observatory.org/science/ctao-performance`

a factor between 3 - 8. These derived numbers for the sensitivity range of CTA are consistent with those discussed in the literature (see e.g. [188] and references therein). The numbers given here are for the chosen model parameters and the limits chosen for e.g. the luminosity range of the model. The implications for investigating a wider dynamic range, as would be possible with CTA, and changes to the model parameters are discussed in the section 6.6.

6.5.2 Flux

The dominant contribution to the measured Galactic VHE flux stems from nearby or exceptionally bright sources. Based on the given model, these distinguished sources, which constitute the HGPS sample, are expected to comprise only 1 %-10 % of all Galactic sources. The total flux of the Galactic population, integrated over the whole sky, was calculated and lies in the range $(3.0\text{-}13.7)\cdot 10^{-10}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ (see Tab 6.6), disregarding the result of mSp2B, which falls even below the flux of the HGPS sample. The stated values of the other distributions may also underestimate the true value, as bright sources in the HGPS sensitivity range are generally underrepresented in the simulations.

Figure 6.14 gives an impression of what the population might look like. The top panel shows a sky map of the differential flux at 1 TeV for the HGPS sample. All sources are represented as two-dimensional Gaussians. Each panel below likewise shows a synthetic population derived from a different source spatial distribution. HGPS sources were included in the simulations by substituting those simulated sources that have the minimum distance to an observed source or that are closest to the line of sight when the distance of the HGPS source to the observer is unknown. Luminosities and radii were then redistributed among the remaining simulated sources according to the model in such a way that no additional source falls within the sensitivity range of the HGPS. In this way, the simulations presented agree with both the observation and the model. For all sources a power law with spectral index 2.3 was assumed.

The sky maps displayed change, sometimes drastically, between the different distributions. For mSNR and mSp4, a narrow band of mainly point-like and some faint extended sources along the Galactic plane is added to the HGPS sample. In comparison, the contribution of undetected sources for mSp2B appears to be confined in the range $-60^\circ \leq l \leq 50^\circ$ and there are generally fewer extended sources than in the two previously mentioned source distributions. On the other hand, the distribution here is broader in the central region, i.e. it extends to higher latitudes due to the contribution of the Galactic bar. In contrast to these three distributions, the sky maps for mPWN and mFE are characterised by a large fraction of extended sources outside the HGPS sensitivity range across the FoV, which in combination give the appearance of broad diffuse emission. At $l = 350^\circ$, a large source is seen in the distribution of mPWN that would remain undetected according to the HGPS sensitivity. This demonstrates the effect of the instrument's limitation to detect sources larger than 1°. Instead, this source would contribute to the background estimate from within the FoV, which would accordingly be too large. Furthermore, with a source distribution like mFE, any (alternative) background estimate is challenged, as there is virtually no region free of sources up to large latitudes ($|b| > 3^\circ$). In this context, it is also worth noting that with each distribution studied in this work, the HGPS sample is expected to be affected by source confusion.

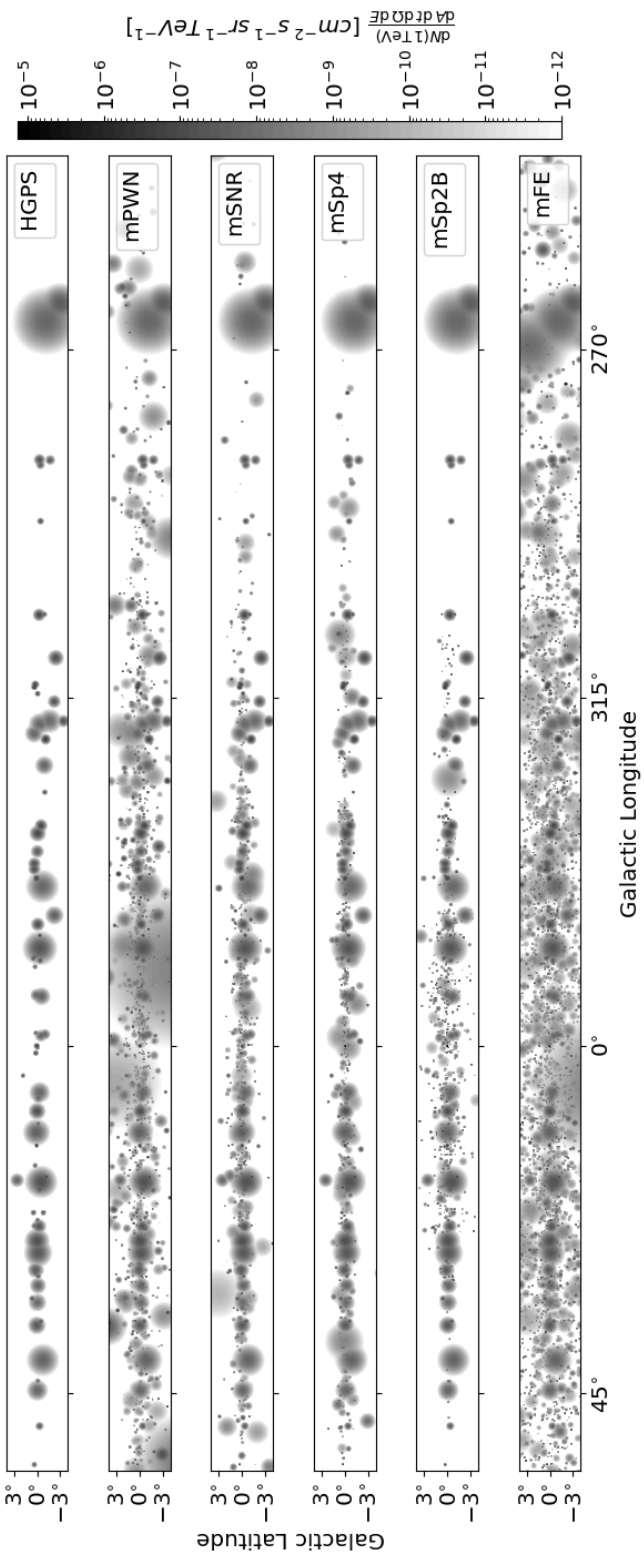


Figure 6.14: *Flux maps for various source samples using Gaussian models defined by flux, extent and position to represent the sources. The top panel shows the HGPS source sample. Below are flux maps for different realisations of Galactic source populations derived from the different spatial distribution models. There, simulated sources within the sensitivity range of the HGPS have been replaced by the observed sources.*

These examples also show the importance of an appropriate model of the Galactic source population to assess the measurement of diffuse emission. Within the FoV of the HGPS, sources that satisfy the detection criterion described in Sec. 6.2 are estimated to account for 40 % (mFE) to 67 % (mSp4) of the flux of the Galactic source population. This implies that a significant contribution to the total flux can be expected from as yet unresolved sources. The H. E. S. S. measurement of the large-scale diffuse emission in the HGPS from regions that do not contain significant γ-ray emission yields a comparable number of $\sim 28\,\%$ of the total measured VHE emission in the large-scale diffuse emission [84]. However, these numbers are not directly comparable. First, the detection criterion used for the estimation does not take into account all sources detected with the HGPS, i.e. the detection threshold is overestimated for some sources and thus the flux of the HGPS sample is underestimated. Second, the sky regions from which the estimate and the measurement are derived are not identical. While the model allows for the simple removal of detectable sources, complex exclusion regions were applied to the H. E. S. S. measurement to exclude the contribution of sources, most of which cluster at small values of Galactic latitude. Nevertheless, this comparison gives an indication that unresolved sources may well be the dominant component of the measured diffuse emission, in particular if pulsars or PWNe are the prevailing class of VHE γ-ray sources.

6.5.3 Luminosity

Using the mean photon energy of $3.52\,\mathrm{erg\,ph^{-1}}$ from Sec. 6.1.2, the total luminosity of the Galactic VHE source population is estimated to lie in the range $(5.0\text{-}19.5)\cdot 10^{36}\,\mathrm{erg\,s^{-1}}$ (disregarding mSp2B). Assuming that γ-ray sources are the dominant contribution to the overall VHE γ-ray luminosity of the Milky Way and that the diffuse emission originating from propagating cosmic rays adds only a small contribution, these values can be compared with the total luminosity at megaelectronvolt and gigaelectronvolt energies ($\sim 3 \cdot 10^{38}\,\mathrm{erg\,s^{-1}}$ and $\sim 8 \cdot 10^{38}\,\mathrm{erg\,s^{-1}}$, respectively; [189]). The γ-ray luminosity of the Milky Way at VHE turns out to be one to two orders of magnitude lower than in those two lower energy bands. This demonstrates that the presented model is compatible within the available energy budget constraints and indicates a drop in luminosity between the HE and VHE ranges.

6.6 Discussion

In this work, an attempt was made to describe the population of Galactic VHE γ-ray sources by suitable distribution functions of position, luminosity and size. The addition of the mostly neglected size distribution was done to account for the observational bias and its dependence on the angular extent when evaluating the sample of observed sources. Comparison of the model with observations has confirmed the need for this component. For example, the results presented in Sec. 6.5.2 show that the fraction of detected sources among all sources within the FoV of the HGPS above a given flux depends on the spatial distribution of VHE sources and their size distribution. This means that the specification of a completeness range is not unconditional and exclusively bound to a flux value. Thus, an assessment of the Galactic source population based on the assumption of a complete sample of sources without taking both of these distributions into account may not be reliable.

With the exception of mSp2B, no source spatial distribution is preferred over another. In

particular, the small number of sources combined with the inhomogeneity of the HGPS sensitivity make each spatial distribution model equally suitable for describing the distribution of observed sources over longitude. Conversely, each of the spatial distributions studied is equally ill-suited for describing the observed source distribution over latitude. The observed asymmetry provides a clue for improving the model, for example by considering the warp of the Milky Way or the position of the Sun above the Galactic plane. Distinguishing between different spatial distribution models will likely be possible with larger samples covering a larger volume of the Galaxy and obtained from future instruments such as CTA. Constraints on the spatial distribution may also be derived from the measurement of diffuse emission. For example, if the measured spectrum of diffuse emission allows an association with unresolved sources, the distribution of emission over latitude can provide valuable insight into the underlying source distribution, as demonstrated in Fig. 6.14.

With regard to the distribution over angular extent, additional work needs to be done to reconcile model and observation. On the other hand, rough agreement is already achieved for the distribution over flux, but not yet at a statistically robust level. On closer inspection, the model-derived distribution over flux tends to be too steep compared to observation, regardless of the spatial distribution, and overestimates the contribution of sources below 0.1 C.U. and accordingly yields too few sources above this value. Similarly, the source distribution over angular extent derived from the model is too steep and, in contrast to observation, favours a large proportion of point-like sources in the sample of sources within the sensitivity range of the HGPS. Since the complexity of the model is increased by accounting for the effect of angular extent, there are several issues that need to be considered to address these discrepancies. However, the primary problem does not concern the model, but the method used to estimate the parameters of the model. While it has been shown to provide unbiased estimates, the maximum likelihood approach presented above yields large uncertainties on these estimates due to the small number of sources that can be considered. Without sufficient constraints on the parameter space, the validity of the model can hardly be judged. Therefore, developing an alternative method that is able to exploit a larger set of known sources and thus reduce the uncertainties of the estimated parameters would be the first step towards solving the problem. In the case of improved or tightly constrained estimates, it may well turn out that the heuristically chosen luminosity and radius functions are too simple and an adjustment is needed to match the observation. In addition, source confusion affects the sample of observed sources in almost all aspects, i.e. the absolute number of sources, their flux and angular extent. It has been shown that the extent to which the sample is affected depends strongly on the spatial distribution of the sources. Correctly accounting for source confusion is a challenging task and is deferred to a future study. Finally, another aspect of the applied population synthesis approach that affects the distribution over flux and angular extent is the discrepancy between the provided estimate of the sensitivity and the actual sensitivity of the HGPS, which was discussed in Sec. 6.2. However, with the available information, this problem can hardly be solved.

7 Improved estimation of model parameters with deep learning

While the likelihood method presented in the previous chapter allows for an unbiased reconstruction of the model parameters, taking into account the complex observational bias inherent in the HGPS sample, it suffers from two limitations, namely its dependence on available distance estimates and the resolution of the angular extent of a source. This limits the analysis to a small subsample of the known sources and leads to large uncertainties in the reconstructed parameters. One way to mitigate this limitation is to provide estimates for the missing values. For example, for a given source with unknown distance a probability function over distances can be derived from the source spatial distribution model. Correspondingly, this translates into a probability function over luminosity and radius. Following this approach, the combined distribution of estimated luminosities and radii for the whole HGPS sample is complex (as shown below) and difficult to describe analytically. An even greater challenge is finding a suitable function that allows inference of model parameters from this distribution. For such task, machine learning is an ideal tool and is used here.

In Sec. 7.1 the method used for the inference of model parameters is explained in detail, starting with the approach how missing source parameters are estimated, followed by the generation of a data set and finally the machine-learning model. In Sec.7.2 the performance of the method is evaluated. The results derived from the HGPS sample are presented in Sec. 7.3 and discussed in Sec. 7.4.

7.1 Method

Similar to the likelihood method, this method performs the reconstruction of the model parameters based on a two-dimensional, binned distribution of luminosities L and radii R derived from the sample of observed sources. In contrast to the previous method, the distribution contains information about point-like sources as well as sources without known distance. This means that four different cases have to be considered with respect to the contribution of each source to this distribution. The first case is an extended source with known distance. It provides all the information needed to calculate its luminosity and radius, namely the flux F, the angular extent σ and the distance r. The contribution of such a single source j to the content of bin i of the luminosity-radius distribution I is given by

$$I_{ij} = \int_{L_{min,i}}^{L_{max,i}} \mathrm{d}L \int_{R_{min,i}}^{R_{max,i}} \mathrm{d}R\, \delta\left(F_j - \frac{L}{4\pi r_j^2}\right) \cdot \delta\left(\sigma_j - \tan^{-1}\left(\frac{R}{r_j}\right)\right) . \tag{7.1}$$

The second case is an extended source whose distance is unknown. In this case, the distance is estimated using a PDF derived from the modelled source distribution along the

line of sight (cf. Fig. 6.2). The contribution of this source j to the content of bin i of the luminosity-radius distribution I is then calculated as

$$I_{ij} = N \int_{L_{min,i}}^{L_{max,i}} \mathrm{d}L \int_{R_{min,i}}^{R_{max,i}} \mathrm{d}R \int_{0}^{r_{max,j}} \mathrm{d}r\, P(l_j, b_j, r) \cdot \delta\left(F_j - \frac{L}{4\pi r^2}\right) \cdot \delta\left(\sigma_j - \tan^{-1}\left(\frac{R}{r}\right)\right), \tag{7.2}$$

where the source density is given in Galactic coordinates $P(l, b, r)$. The maximum distance $r_{max,j}$ is determined by the assumed geometry of the Galaxy, e.g. a cylinder with finite volume. To ensure that each source is contributing equally to the luminosity-radius distribution of the sample, Eq. 7.2 is normalised by a factor N so that $\sum_i I_{ij} \stackrel{!}{=} 1$.

The third case is sources whose distance are known but which appear to be point-like. In the HGPS, any source with an angular extent smaller than the systematic uncertainty in the size of the PSF, 0.03°, is said to be point-like. Conservatively, one can assume that the actual angular extent of a point-like source lies with uniform probability somewhere in the interval (0°, 0.03°)[1]. Since the distance is known, this implies a uniform probability distribution for the radius of the source. Here, the upper bound of the radius interval is determined by the systematic uncertainty in the size of the PSF and the source distance, while the lower bound is shifted upwards to the minimum source radius defined for the radius distribution function of the population model. This finally results in

$$I_{ij} = N \int_{L_{min,i}}^{L_{max,i}} \mathrm{d}L \int_{R_{min,i}}^{R_{max,i}} \mathrm{d}R\, \delta\left(F_j - \frac{L}{4\pi r_j^2}\right) \cdot \mathcal{U}\left(R, \left(R_{min}, r_j \tan(\sigma_j)\right)\right), \tag{7.3}$$

where $\mathcal{U}\left(R, (\cdot, \cdot)\right)$ denotes the uniform distribution over R in the specified interval $(\cdot, \cdot)$. Again, a normalisation factor N ensures an equally weighted contribution from all sources in the sample.

Finally, there are sources that neither appear to be extended nor have a distance estimate. For such sources, their intrinsic luminosity and radius are maximally uncertain and are estimated by combining Eq. 7.2 & Eq. 7.3 in such a way that

$$I_{ij} = N \int_{L_{min,i}}^{L_{max,i}} \mathrm{d}L \int_{R_{min,i}}^{R_{max,i}} \mathrm{d}R \int_{0}^{r_{max,j}} \mathrm{d}r\, P(l_j, b_j, r) \cdot \delta\left(F_j - \frac{L}{4\pi r^2}\right) \cdot \mathcal{U}\left(R, \left(R_{min}, r \tan(\sigma_j)\right)\right). \tag{7.4}$$

In Fig. 7.1 the four different cases are shown. For this purpose, a hypothetical source was placed in the direction of the Galactic centre at a distance of 5.5 kpc. Its intrinsic luminosity is $10^{34}\ \frac{\text{photons}}{\text{s}}$ and the radius is set either to $R = 20$ pc, corresponding to an angular extent of 0.2°, or to $R \leq 2.8$ pc for a source consistent with being point-like. The upper left panel shows the first case, a source whose distance and angular extent are known. After applying Eq. 7.1, one sees that a single bin contains all the information about this source. For the second case, a source with known extent but unknown distance, Eq. 7.2 is applied. The result is shown

[1] More precise upper limits for the angular extent of some of the point-like HGPS sources are given in the catalogue. These are somewhat higher than the systematic uncertainty in the size of the PSF. However, for a consistent treatment of all sources and in particular for a consistent treatment of observations and simulations, a value of 0.03° is uniformly adopted

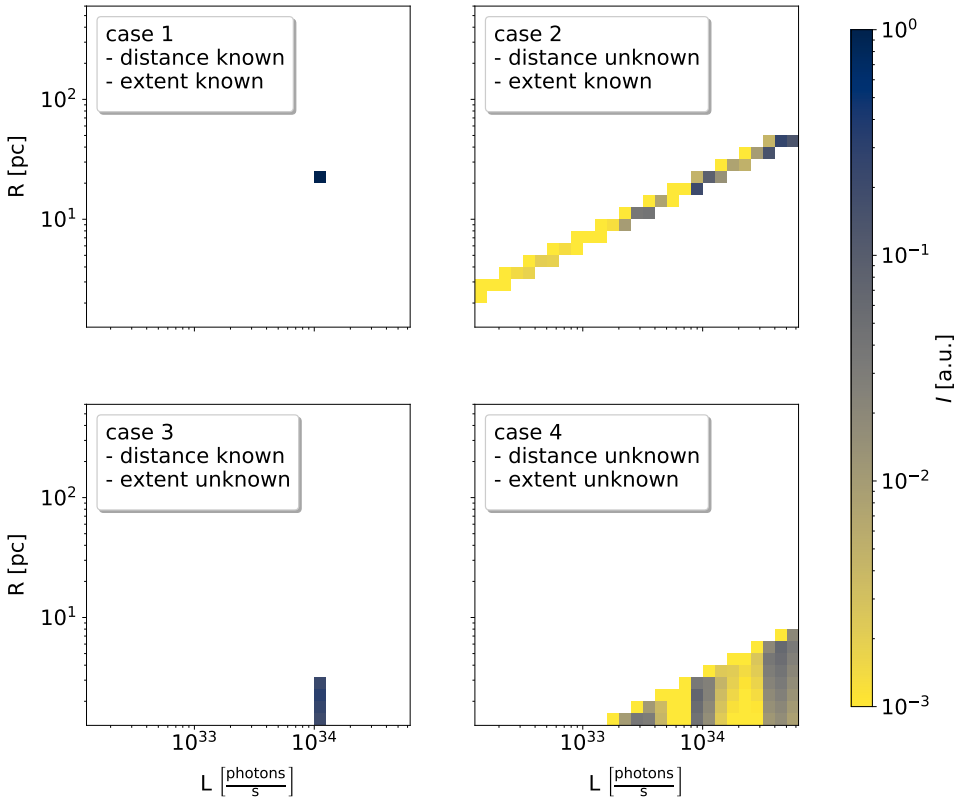


Figure 7.1: *The four cases of estimating the contribution of a single source to the luminosity-radius distribution of the sample. The hypothetical source is located in the direction of the Galactic centre at a distance of* 5.5 kpc*. Its intrinsic luminosity is* 10^{34} photons s^{-1}*. In case its angular extent is known (upper panels), its radius was set to* 20 pc*, otherwise the radius is* < 2.8 pc*. For more information see the text.*

in the upper right panel. There, a characteristic diagonal line can be seen, which reflects the uncertainty of the luminosity and radius of the source. Along this diagonal line, those regions of the parameter space are weighted more heavily which correspond to the source being located in a region of high source density. In this example, the spatial model mSp4 is used and the source was placed in the region of the Norma arm. For this reason, luminosities and radii close to the true values are given large weights. Case three is shown in the lower left panel. Since the distance is known, the luminosity is fixed, but the uncertainty of the radius, due to the lack of knowledge of the angular extent, results in a characteristic vertical line in this two-dimensional distribution. Finally, we see that sources that have neither distance nor angular extent cover a large area of the phase space, as shown in the lower right panel. Like for case two, regions of the parameter space are weighted more heavily which correspond to the source being located in a region of high source density.

Taking the 62 sources of the HGPS sample that surpass the the sensitivity threshold de-

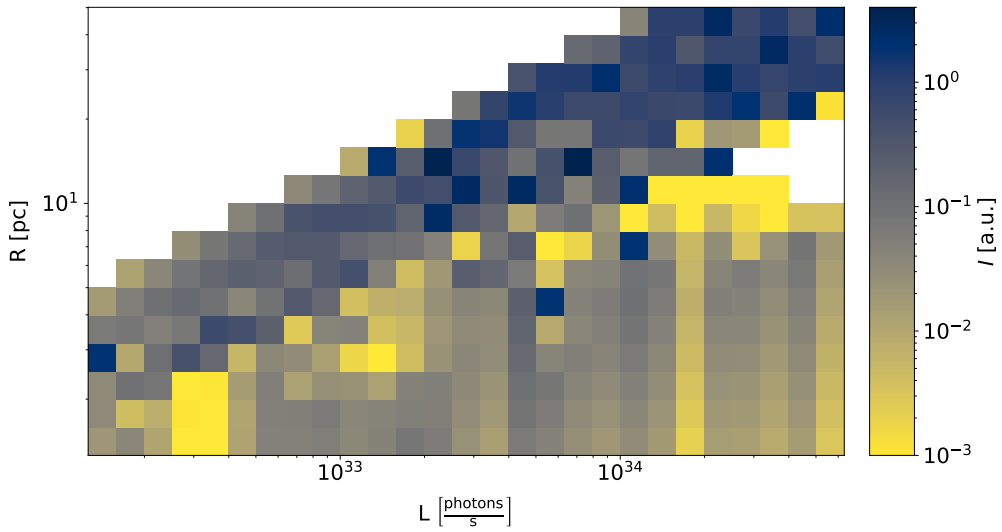


Figure 7.2: *Luminosity-radius distribution of the sample of 62 HGPS sources that surpass the sensitivity threshold derived from the map with* 0.2° *correlation radius. Estimates for sources with missing information were performed on the basis of the spatial model mSp4.*

rived from the map with 0.2° correlation radius (see Table 6.3) and combining the available information from all these sources, i.e. $I = \sum_j I_j$, results in a complex luminosity-radius distribution as shown in Fig. 7.2. For the distribution shown here, estimates were made for sources with missing information based on the spatial model mSp4. While no contribution from a source from case 3 appears clearly, a visual distinction is possible for the contributions of sources from cases 1, 2 and 4. It is obvious that this distribution contains much more information than can be derived from the extended sources with known distances alone (e.g. bins with values ≥ 1). However, its complex structure defies a simple analytical description as used in the previous section. On the other hand, it is straightforward to reproduce such a distribution with population synthesis.

The characteristics of the observed distribution are shaped by the global source distribution and the observational bias as defined by Eq. 6.5. In order to reproduce the observed distribution, synthetic source populations are simulated for a range of model parameters. The procedure is the same as outlined in Sec. 6.3.2, except that the total number of sources of the population is adjusted such as to yield on average 62 sources when Eq. 6.5 is applied. This adjustment is necessary because equation 6.5 does not cut on point-like sources and therefore the number of sources that fall within the sensitivity range can change considerably compared to the previously described procedure, depending on the proportion of sources whose angular extent cannot be resolved. For example, a population that contains a large amount of luminous sources, but most of which appear to be point-like due to small radii, must have a large total number of sources to reach 48 sources that exceed the threshold defined by Eq. 6.9. On the other hand, if the threshold is relaxed to allow point-like sources, the same population with the same total number of sources will most likely significantly exceed the required number of 62 detectable sources. Furthermore, to ensure that the synthetic samples provide as much information as the HGPS sample, the same number of sources (23) are assigned distance information. In particular, 17 extended sources are assigned distance information, provided the sample contains a sufficiently large number of extended sources.

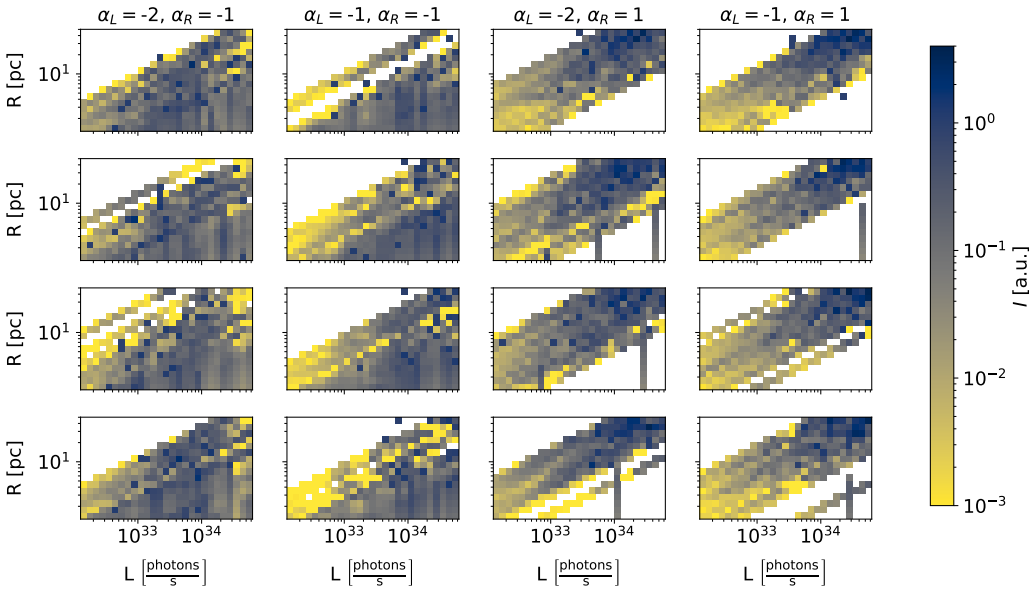


Figure 7.3: *Estimates of the luminosity-radius distributions of synthetic source populations. The simulations are based on model mSp4 and obtained with varying parameters of the luminosity function and radius function.*

Apart from this boundary condition, the selection of sources being assigned distance information is random.

Examples of luminosity-radius distributions derived from synthetic source populations with varying model parameters are shown in Fig. 7.3. Each column shows four random samples derived from the same pair of model parameters (α_L, α_R). The samples shown in the two left columns were derived from a radius function with $\alpha_R = -1$, the ones on the right with $\alpha_R = +1$. Regarding the luminosity function, $\alpha_L = -2$ was used to derive the samples shown in columns one and three, while $\alpha_L = -1$ was used for the samples in columns two and four. A dependence of the shape of the distribution on these model parameters can be seen. With larger α_L values, the fraction of high luminosity sources in the sample is increased. Visually, this is particularly evident in the distribution of sources with known luminosity and radius (dark dots), which is consistently shifted towards high luminosities for all samples derived with $\alpha_L = -1$. Similarly, the fraction of extended sources is increased for larger α_R values, as expressed in these examples by a much smaller number of vertical lines originating from point-like sources in the samples shown on the right compared to those shown on the left. These properties can be exploited to relate the observed luminosity-radius distribution to the parameters of the underlying luminosity and radius functions of the population. Convolutional neural networks are particularly suitable for this type of task. By training on a set of simulated luminosity-radius distributions derived from synthetic populations, a network implicitly learns the impact of the observational bias on the observed sample and how to correct for it in its prediction of the parameters describing the entire population.

To achieve good reconstruction performance, a large dataset is required to train the CNN.

Moreover, the parameter space covered by the dataset should contain the true values of the population, as CNNs are excellent at interpolating between known values but perform poorly when they have to extrapolate. A suitable parameter space could be constrained by the previous results. However, this could bias the results obtained with this method. Therefore, a data set was generated that includes a fairly large parameter space with values for both α_L and α_R ranging from -3 to 2. For data generation, these values for pairs of (α_L, α_R) were randomly sampled independently from a uniform distribution in the specified interval. Synthetic populations with an appropriate number of sources and corresponding luminosity-radius distributions from the sample of sources within the HGPS sensitivity range were calculated for each pair as described above. In this way, 130,000 samples were generated for a given spatial distribution model, of which 60,000 are used as a training sample, 20,000 as a validation sample and the remainder for evaluation of the method. Since the calculation of the luminosity-radius distribution is computationally expensive, especially the calculation of the distance PDF along the line of sight for many sources, while the software implementation used for the spatial distribution model mFE is slow, no dataset was generated for this model.

For the CNN model used here, the same architecture as the one described in Chap. 5, ResNet-18, was chosen. For the task at hand, the number of input channels was adjusted to one. The prediction of the model is a vector (α_L, α_R), which is compared with the true values of the model parameters. Note that unlike the likelihood fit used in the previous section, the number of sources in the population is not predicted by the model. To optimise the model, the mean absolute error is used as the loss function and the stochastic gradient descent as the optimisation algorithm. Originally, the mean squared error was used to train CNNs, but since it performed worse, in particular its use led to a considerable reconstruction bias, it was discarded as a suitable loss function for this task. The performance of the final model is discussed in the next section.

7.2 Monte Carlo verification

The verification of this method is done using the distribution of the reconstruction error $\Delta\alpha_{L/R} = \alpha_{L/R,\,true} - \alpha_{L/R,\,reco}$ as a function of the reconstructed value $\alpha_{L/R,\,reco}$, analogous to what was described in Sec. 6.3.2 for the likelihood approach. The distributions for the parameters of the luminosity function and the radius function derived with the four spatial models for which this method was applied are shown in Fig. 7.4. For α_L, the general shape of the distributions is similar to those derived with the likelihood approach, but the uncertainties of the reconstructed values derived with the CNN approach are noticeably smaller. Again, the reconstructed values are unbiased over most of the parameter space, regardless of the range of paired α_R. Expected deviations from a zero-centred error distribution occur at the edges of the probed parameter space, but are less pronounced than with the likelihood method. Here, the uncertainty of the reconstruction also increases with the reconstruction value and spans a range between 0.14 and 0.39 for $-2.5 \leq \alpha_L \leq 1$ (compare 0.23 and 0.53 for the likelihood approach). A larger improvement is obtained for the reconstruction error of the parameter α_R. Again, the reconstruction value is unbiased over most of the parameter space, independent of the range of the paired α_L. The uncertainties of the reconstruction values reduce to values between 0.14 and 0.26 for $-2.5 \leq \alpha_R \leq 1$ (compare 0.31 and 0.48 with the likelihood approach). Smaller uncertainties are observed around the middle of the probed parameter space ($-1 \leq \alpha_R \leq 0$) consistently for all spatial distribution models.

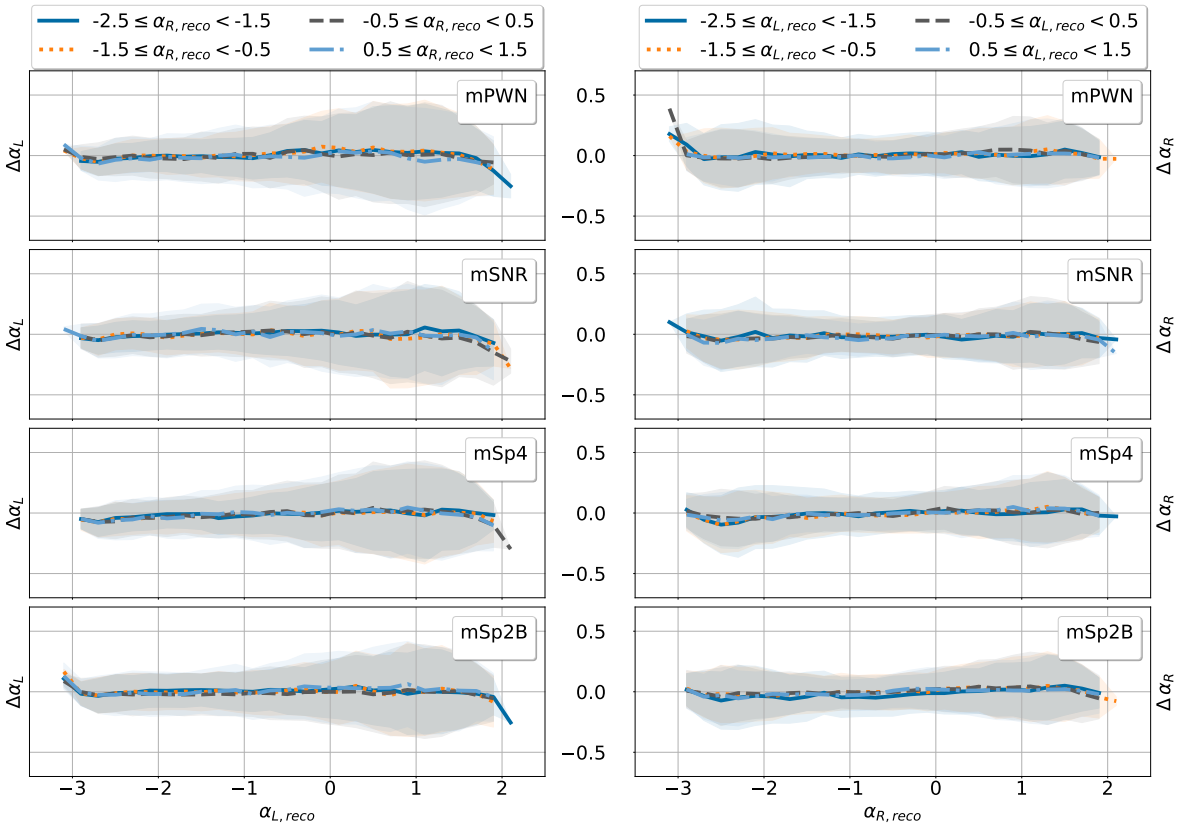


Figure 7.4: *The errors of the reconstructed parameters α_L (left) and α_R (right) when calculated with the deep-learning method. Details are given in the text.*

Judging from the result of this Monte Carlo study, the CNN approach clearly outperforms the likelihood approach in terms of the uncertainty on reconstructed parameters. In particular, the reconstruction of the parameter of the radius function is improved by using information from all sources within the sensitivity range, including those that appear point-like.

7.3 Result

To estimate the parameters of the luminosity function and the radius function, the same procedure as in section 6.3.3 was used. Here, the value of the upper limit for the extent of the point-like sources, i.e. 0.03°, was not varied during bootstrapping. The reconstructed values and their 1σ uncertainties are listed in Table 7.1. Again, the results among the spatial distributions are fairly consistent with average values of $\langle\alpha_L\rangle = -2.37$ and $\langle\alpha_R\rangle = 0.19$. The results including the two-dimensional uncertainty intervals are also shown in Fig 7.5. There, the bold crosses represent the best estimates and the errors corresponding to the uncertainties of the measured quantities of the sources as derived from bootstrapping. The bold lines show the contours of the final uncertainties of the estimate after taking into account the systematic errors of the reconstructed values. For comparison, the same is shown for the likelihood results by thin contour lines and thin crosses. For all spatial models, the new results are systematically shifted towards lower α_L values and higher α_R values, which

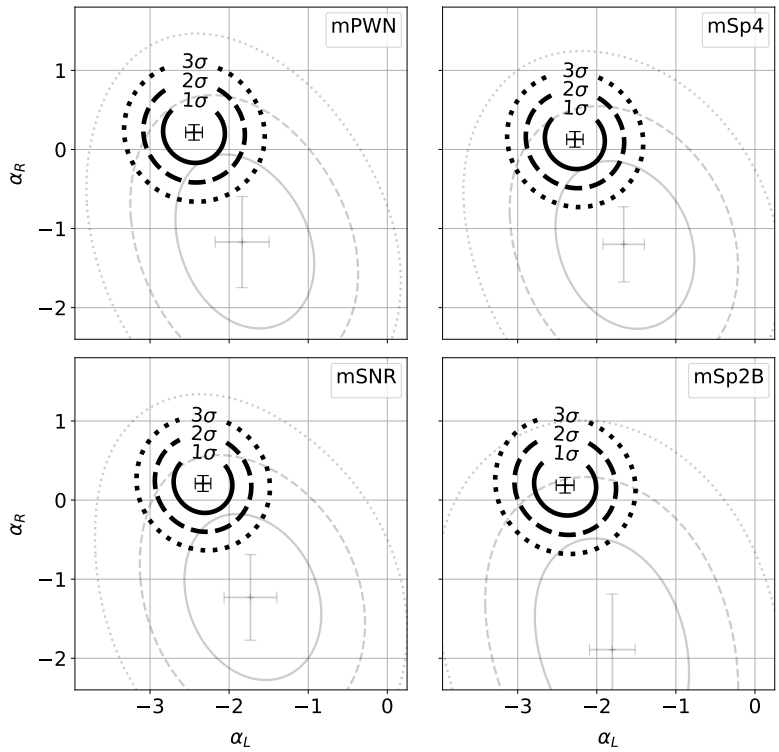


Figure 7.5: *Best estimates and confidence intervals of α_L and α_R calculated with the DL approach (bold lines) and comparison with the results of the likelihood approach (thin lines). The position of the cross marks the best estimate values. Error bars reflect the uncertainty corresponding to the errors of the measured source properties. Lines show the 1σ, 2σ and 3σ confidence intervals for the true values of α_L and α_R.*

however all fall within the 95 % confidence interval obtained with the likelihood approach. The new confidence intervals obtained with the machine-learning approach are much smaller compared to the previous results. The ensemble derived from bootstrapping also has a lower variance compared to the previous results, as can be seen from the sizes of the error bars.

Table 7.1: *Reconstructed values of α_L & α_R using deep learning. Errors are derived from the 1σ confidence interval as shown in Fig. 7.5.*

Model	α_L	α_R
mPWN	-2.45 ± 0.39	0.22 ± 0.38
mSNR	-2.33 ± 0.37	0.21 ± 0.37
mSp4	-2.28 ± 0.38	0.13 ± 0.37
mSp2B	-2.40 ± 0.39	0.19 ± 0.38

The new values for α_R are much closer to the expectation that follows the naive assumption of a population of sources that all expand similar to the shock front of a single SNR (see Sec. 6.1.2). On the other hand, the newly found indices of the luminosity function are in clear contradiction with previous findings (e.g. [87]). The implication of these results are discussed

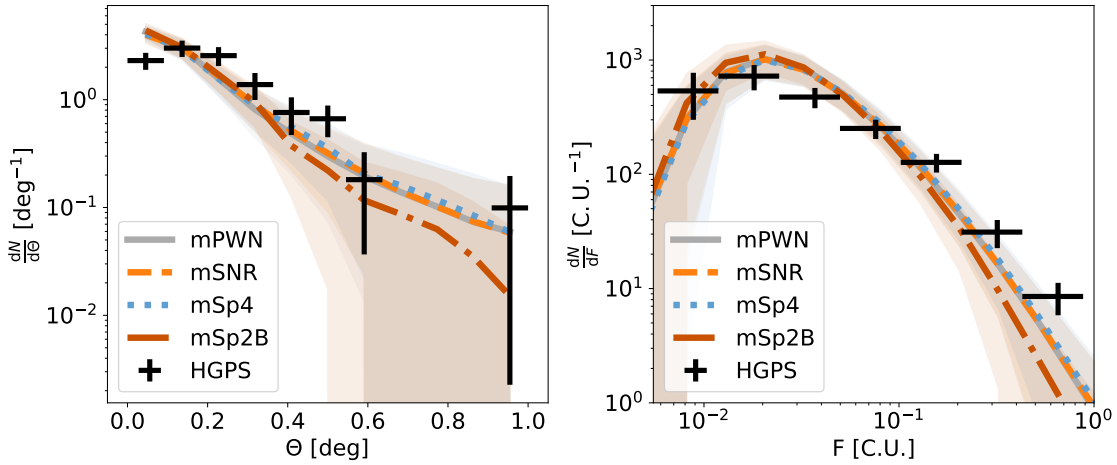


Figure 7.6: *Comparison of the source differential distribution over flux (right) and angular extent (left) between synthetic source populations and the HGPS sample as in Fig. 6.8, with reconstructed model parameters from the CNN.*

in the following section.

7.4 Discussion

First, the effect of the new values for the parameters of the luminosity and radius functions on the distributions of sources in the sensitivity range of the HGPS are examined. Since the new results only concern the luminosity and radius function of the model, no changes in the distribution over Galactic longitude and latitude are observed. Instead, significant changes are obtained in the distribution over angular extent as shown in the left panel of Fig. 7.6. While the distribution obtained with mSp2B still falls off too steeply towards large extents, the other models with the new parameter values provide a qualitatively better description of the observation. More precisely, above 0.2°, the predicted distributions became flatter such that the average values over model realisations derived with mPWN, mSNR and mSp4 are much closer to observation with a large overlap of the uncertainty regions. Quantitatively, however, predicted and observed distributions still differ significantly. The most striking difference is seen in the case of point-like sources, for which the model gives predictions that are larger than the observation by a factor of two. At this point, source confusion is again likely to contribute to this discrepancy. Some sources that appear point-like when observed in isolation, as assumed in the model, may appear extended when confused with a larger source or several smaller sources, which is to be expected, for example, in crowded regions such as towards the Galactic centre. This effect would reduce the number of apparent point-like sources in the simulated observation. On the other hand, if some of the observed extended sources are in fact the product of several smaller sources, taking source confusion into account in the modelling will shift the radius distribution towards smaller radii, i.e. towards smaller α_R. This would in turn increase the number of expected point-like sources. An accurate implementation of source confusion in the modelling is necessary to test whether possibly one of these two effects dominates over the other. The additional difficulty with such an implementation is that in the creation of the HGPS catalogue Gaussian components were manually merged into sources without applying well-defined criteria.

Compared to the previous results, the source distribution over flux, as shown in the right panel of Fig. 7.6, seems relatively unaffected by the change in the luminosity function with $\Delta\alpha_L \geq 0.6$ (cf. Fig. 6.8) and still provides good agreement. Compared to observation, the abundance of faint sources in the model predictions mirrors to some extent the abundance of point-like sources. The reason for this is that the detection threshold for point-like sources is lower than for extended sources. Therefore, in the context of the source distribution over the angular extent, a comparatively larger contribution of faint, point-like sources compared to observation can be expected from the model. Overall, the new deep-learning based estimation technique proves to be a valuable tool, judging by the improvements observed here between the model predictions and the actual distribution of HGPS sources.

As for the global properties of the populations, the newly found model parameters cause more drastic changes. The stronger weighting of low-luminosity sources resulting from a smaller α_L leads to a large increase in the number of sources in the total population required to obtain the observed number of sources in the HGPS sensitivity range, i.e. $9.2 \cdot 10^3$ for mSp4 and $2.6 \cdot 10^4$ for mPWN. In the context of the results obtained with mFE, it has already been discussed in Sec. 6.5.1 that a population of 7000 γ-ray sources would challenge the standard paradigm of SNRs and PWNe as the dominant source classes. These new results are in even greater tension with that paradigm. However, with its prediction of the total flux within the FoV of the HGPS exceeding that of all known sources by a factor of 3-6, it is rather the model that is challenged by the observation. This is because the observation clearly shows that the emission of VHE γ-rays is dominated by few nearby or exceptionally luminous sources, and that the flux in sources not yet resolved contributes only to a small extent to the total flux. Since the new, much more accurate estimates of the model parameters have led to a significantly improved description of the source distribution in the sensitivity range of the HGPS, it must be assumed that the remaining discrepancies are due to unaccounted instrumental effects in the description of the sensitivity range and accordingly in the estimation of the observational bias, or to the fact that the distribution functions of the source properties are not sufficiently well described in the presented model.

Some starting points for further improvements of the model have already been mentioned. As for the instrumental effects that have not been taken into account, source confusion is a significant problem that needs to be addressed in order to make more accurate predictions. Also, the effects of the large-scale diffuse emission model and the differing estimates of the sensitivity used in creating the sky map and extracting sources for the catalogue have not been considered in the modelling presented here. Ideally, all these components would be accounted for in a detailed simulation of the survey and a full replication of the catalogue generation. Apart from the computational costs that would be associated with applying this procedure to hundreds of synthetic source populations, a realisation of such a procedure fails because not all necessary information is provided, e.g. the background estimates or the IRFs used for the analysis of the data. At least for reasons of reproducibility, future surveys will hopefully provide all this information. At present, however, the development of an emulation of the HGPS pipeline that approximates its results reasonably well is needed to further improve the estimation of the observational bias obtained in this work. As for the description of the source properties, simple one-parametric distribution functions, namely power laws, were used here as a working hypothesis for the luminosities and radii of the sources. While with

the large uncertainties of the parameter estimates derived with the likelihood approach any result obtained with a more complex distribution function with a larger number of parameters would have had little significance, the situation has changed with the deep-learning approach. Now various hypotheses can be tested with some confidence. Nevertheless, more suitable distribution functions must first be found. One idea would be to first simulate the properties for individual sources of known classes for a wide range of reasonable parameters that determine their evolution. Subsequently, parametric functions would have to be found that describe the luminosity and radius distribution of the simulated ensemble and its variation with the simulation parameters. In this way, possible correlations between luminosities and radii could also be revealed and included in the model. Going one step further and extending the model to include the spectral properties of the sources would be the logical next step, thus extending the population model to neighbouring energy bands. In this way, the additional information provided by a multi-wavelength approach could be exploited, e.g. with the large statistics at high energies in the *Fermi*-LAT data or the wide-FoV observations with HAWC. Furthermore, spectral information offers the possibility of an informative comparison of this model of the entire population of VHE γ-ray sources with theoretical models of individual source classes. In addition, the spectral information would also be useful in using this model as a template for unresolved sources in data analyses. Finally, it was quantitatively shown in this work that the asymmetry in the latitudinal profile of the observed sources is not due to an inhomogeneous sensitivity, but that the reason must be sought elsewhere. The warp of the Galaxy and the position of the Sun above the Galactic plane were identified as two effects that could lead to such an asymmetry and will therefore be taken into account in the next iteration of the population model.

8 Summary

In this work, novel techniques based on Monte Carlo simulations and recent advances in machine learning have been developed to address two prominent challenges in ground-based γ-ray astronomy, namely the separation of signals from the hadronic background in the analysis of air-shower images of IACTs and the extraction of global properties of the population of VHE γ-ray sources from a strongly biased sample of detected sources. Starting with a detailed study of the potential of deep-learning techniques to improve the γ-hadron separation in IACT data as a means to lower the detection threshold for VHE γ-ray sources, the focus of this work is on the construction of a model of the population of VHE γ-ray sources. A population synthesis approach was followed, with a special emphasis on correcting for the observational bias. In particular, the estimation of the parameters of the model took into account the spatial inhomogeneity of the sensitivity of the HGPS and its dependence on the source extent, both of which were often neglected in previous studies.

In Chap. 5, a classification of different neural network architectures suitable for the analysis of air-shower images was introduced, and on this basis a set of seven different deep-learning models was constructed. The performance of these models on the γ-hadron separation task was evaluated using both Monte Carlo simulations and real observations. All models showed high accuracies on simulations, e.g. $\gtrsim 92\,\%$, and similar energy dependencies of their classification results. In the analysis of observational data, only three models were found to be robust to distortions of the input distribution with respect to the training data consisting of simulated air-shower images, namely to missing pixel information and a different γ-ray spectrum. The best performing model was a combination of a convolutional neural network and a recurrent neural network. A comparison with a standard air-shower classifier based on boosted decision trees showed that the neural network is more robust when analysing observational data and achieves a competitive signal-to-noise ratio. In addition, software was written to apply convolution on hexagonally sampled data, which allowed testing whether the performance of these models increases when the specific geometry of the input data is taken into account in the convolution operations. A comparison between models using the same network architecture but either hexagonal or standard rectangular convolution showed that the achievable performance is the same. This result argues against the need for such specialised software for γ-hadron separation and, alternatively, the need to transform air-shower images into a rectangular geometry. After the detailed investigation of different network architectures for the analysis of air-shower images presented in this work, the next step in future efforts to use neural networks to improve the γ-hadron separation, and thus possibly increase the number of detectable sources, would be to add additional, previously unaccounted for information, such as the temporal evolution of an air shower signal.

The population model, which is based on the data from the H. E. S. S. Galactic plane survey, was presented in Chap. 6. Next to the source luminosity distribution function, the model also comprises the distribution function of source radii and the source spatial distribution function. For the former two, simple power laws were assumed with free indices. For the

source spatial distribution five different models with fixed parameters were probed, comprising axially symmetric distributions, e.g. following the distribution of supernova remnants or pulsars, and non-symmetric distributions, namely spiral structures as observed for the interstellar medium. A likelihood maximisation procedure for the estimation of the model parameters was presented and shown to give unbiased estimates. As a result of applying this procedure to the HGPS source sample, for the indices of the luminosity distribution function estimates between -1.97 and -1.66 were obtained and, likewise, estimates between -1.89 and -1.17 for the indices of the radius distribution function.

The predictions for the distribution of source samples in the sensitivity range of the HGPS already are in part in fairly good agreement with the HGPS sample distribution, especially with the longitudinal profile and source distribution over flux. As for the source distribution over the angular extent, the model consistently yields too many point-like sources. To some extent, this difference between simulations and observations is probably due to source confusion, which is not accounted for in the model presented. There are also differences in the latitudinal distribution between the model and the HGPS sample, in particular the asymmetry observed in the HGPS sample is not reproduced. With the procedure used in this work, this result suggests that this asymmetry is not an effect of an inhomogeneous sensitivity, but is due to an inherent feature of the global source spatial distribution, such as the warp of the Galaxy, or due to an unaccounted observational effect, such as the height of the Sun above the Galactic plane. With the available data and the prediction made, only the tested two-arm spiral model could be ruled out as a possible source distribution.

Predictions were also made about the global properties of the population and the prospects for CTA. The total number of sources was estimated to range from 780 to 6700. The upper end of this range is in contrast to the assumption that supernova remnants and pulsar wind nebulae are the dominant class of VHE γ-ray sources. In addition, it has been estimated that CTA will be able to detect 200 to 600 sources, which would greatly increase our knowledge of the source population. However, this estimate does not take into account the impact of source confusion or the possibility of detecting new classes of sources that are below the detection limit of the current generation instruments. The total flux within the field of view of the HGPS and the contribution of unresolved sources were also estimated. For all spatial distributions of sources, the unresolved sources account for more than 30 % of the total flux of the population. This result is comparable to the result of diffuse emission measurements, demonstrating the importance of the unresolved sources as a major component of the measured diffuse emission. The total luminosity of VHE γ-ray sources was predicted to be in the range of $(5.0\text{-}19.5)\cdot 10^{36}\,\mathrm{erg\,s^{-1}}$, which is two orders of magnitude lower than the estimates in the MeV and GeV energy range for comparison.

In order to reduce the uncertainties in the estimates of model parameters, a novel estimation procedure based on deep learning was developed and presented in Chap. 7. Using extensive Monte Carlo simulations, it was shown that this method also provides unbiased estimates for model parameters, but with much smaller uncertainties. Using this method, new model parameters were found that deviate noticeably from the original estimates, but still fall within their 95 % confidence intervals. For the luminosity, the new estimates are in the range -2.45--2.33 with a 68 % confidence interval of ± 0.38. For the radius distribution, the estimates are in the range 0.13-0.22 with an uncertainty of ± 0.38. The predictions for the distribution of

sources falling within the HGPS sensitivity range have been clearly improved with these new parameters and provide a better match to the distributions of HGPS sources. On the other hand, a prediction of the total flux that is a factor of 3-6 larger than the combined flux of all HGPS sources is not consistent with observation. Therefore, it was concluded that the source distribution functions need to be adjusted and additional observational effects such as source confusion need to be included to further improve the model. For example, source confusion could be included in the population synthesis with a suitable emulation of the whole HGPS analysis procedure, although the difficulty in realising such an emulation lies in the addition of missing information such as the background model and the IRFs used. More accurate source distribution functions could possibly be derived from detailed source models and their predictions about the properties of an ensemble of sources with different initial conditions. Together with the adjustments to the source spatial distribution mentioned above, these are starting points for improvements in a future iteration of the population model.

List of Figures

List of Tables

Bibliography

[1] V.F. Hess, *Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten*, *Physik. Zeitschr.* **13** (1912) 1084.

[2] C. Evoli, *The cosmic-ray energy spectrum*, Dec., 2020. 10.5281/zenodo.4396125.

[3] Y. Shikaze, S. Haino, K. Abe, H. Fuke, T. Hams, K.C. Kim et al., *Measurements of 0.2 to 20 GeV/n cosmic-ray proton and helium spectra from 1997 through 2002 with the BESS spectrometer*, *Astroparticle Physics* **28** (2007) 154 [astro-ph/0611388].

[4] A.D. Panov, J.H. Adams, H.S. Ahn, G.L. Bashinzhagyan, J.W. Watts, J.P. Wefel et al., *Energy spectra of abundant nuclei of primary cosmic rays from the data of ATIC-2 experiment: Final results*, *Bulletin of the Russian Academy of Sciences, Physics* **73** (2009) 564 [1101.3246].

[5] Y.S. Yoon, H.S. Ahn, P.S. Allison, M.G. Bagliesi, J.J. Beatty, G. Bigongiari et al., *Cosmic-ray Proton and Helium Spectra from the First CREAM Flight*, *ApJ* **728** (2011) 122 [1102.2575].

[6] O. Adriani, G.C. Barbarino, G.A. Bazilevskaya, R. Bellotti, M. Boezio, E.A. Bogomolov et al., *Time Dependence of the Proton Flux Measured by PAMELA during the 2006 July-2009 December Solar Minimum*, *ApJ* **765** (2013) 91 [1301.4108].

[7] M. Aguilar, D. Aisa, B. Alpat, A. Alvino, G. Ambrosi, K. Andeen et al., *Precision Measurement of the Proton Flux in Primary Cosmic Rays from Rigidity 1 GV to 1.8 TV with the Alpha Magnetic Spectrometer on the International Space Station*, *Phys. Rev. Lett.* **114** (2015) 171103.

[8] KASCADE Collaboration, T. Antoni, W.D. Apel, F. Badea, K. Bekk, K. Bernlöhr et al., *A non-parametric approach to infer the energy spectrum and the mass composition of cosmic rays*, *Astroparticle Physics* **16** (2002) 245 [astro-ph/0102443].

[9] J. Matthews and Telescope Array Collaboration, *Highlights from the Telescope Array Experiment*, in *35th International Cosmic Ray Conference (ICRC2017)*, vol. 301 of *International Cosmic Ray Conference*, p. 1096, Jan., 2017.

[10] A. Aab, P. Abreu, M. Aglietta, J.M. Albury, I. Allekotte, A. Almela et al., *Measurement of the cosmic-ray energy spectrum above 2.5 $\times 10^{18}$ eV using the Pierre Auger Observatory*, *Phys. Rev. D* **102** (2020) 062005 [2008.06486].

[11] A. Haungs, H. Rebel and M. Roth, *Energy spectrum and mass composition of high-energy cosmic rays*, *Reports on Progress in Physics* **66** (2003) 1145.

[12] B. Peters, *Primary cosmic radiation and extensive air showers*, *Il Nuovo Cimento* **22** (1961) 800.

[13] M. Aglietta, B. Alessandro, P. Antonioli, F. Arneodo, L. Bergamasco, M. Bertaina et al., *The cosmic ray primary composition in the "knee" region through the EAS electromagnetic and muon measurements at EAS-TOP*, *Astroparticle Physics* **21** (2004) 583.

[14] T. Antoni, W. Apel, A. Badea, K. Bekk, A. Bercuci, J. Blmer et al., *Kascade measurements of energy spectra for elemental groups of cosmic rays: Results and open problems*, *Astroparticle Physics* **24** (2005) 1.

[15] The Pierre Auger Collaboration, A. Aab, P. Abreu, M. Aglietta, E.J. Ahn, I.A. Samarai et al., *The Pierre Auger Observatory: Contributions to the 34th International Cosmic Ray Conference (ICRC 2015)*, *arXiv e-prints* (2015) arXiv:1509.03732 [1509.03732].

[16] D. Allard, E. Parizot and A.V. Olinto, *On the transition from galactic to extragalactic cosmic-rays: Spectral and composition features from two opposite scenarios*, *Astroparticle Physics* **27** (2007) 61 [astro-ph/0512345].

[17] M. Unger, G.R. Farrar and L.A. Anchordoqui, *Origin of the ankle in the ultrahigh energy cosmic ray spectrum, and of the extragalactic protons below it*, *Phys. Rev. D* **92** (2015) 123001 [1505.02153].

[18] R. Aloisio, V. Berezinsky, P. Blasi, A. Gazizov, S. Grigorieva and B. Hnatyk, *A dip in the UHECR spectrum and the transition from galactic to extragalactic cosmic rays*, *Astroparticle Physics* **27** (2007) 76 [astro-ph/0608219].

[19] R.U. Abbasi, T. Abu-Zayyad, M. Allen, J.F. Amman, G. Archbold, K. Belov et al., *First Observation of the Greisen-Zatsepin-Kuzmin Suppression*, *Phys. Rev. Lett.* **100** (2008) 101101 [astro-ph/0703099].

[20] J. Abraham, P. Abreu, M. Aglietta, C. Aguirre, D. Allard, I. Allekotte et al., *Observation of the Suppression of the Flux of Cosmic Rays above* $4\times10^{19}\,eV$, *Phys. Rev. Lett.* **101** (2008) 061101 [0806.4302].

[21] K. Greisen, *End to the cosmic-ray spectrum?*, *Phys. Rev. Lett.* **16** (1966) 748.

[22] G.T. Zatsepin and V.A. Kuz'min, *Upper Limit of the Spectrum of Cosmic Rays*, *Soviet Journal of Experimental and Theoretical Physics Letters* **4** (1966) 78.

[23] The Pierre Auger Collaboration collaboration, *Features of the energy spectrum of cosmic rays above* 2.5×10^{18} eV *using the pierre auger observatory*, *Phys. Rev. Lett.* **125** (2020) 121106.

[24] M. Israel, J. George, R. Mewaldt and G. deNolfo, "An overview of cosmic-ray elemental composition." http://www.srl.caltech.edu/ACE/ACENews/ACENews83.html, 2004.

[25] The Pierre Auger Collaboration, A. Aab, P. Abreu, M. Aglietta, I.F.M. Albuquerque, I. Allekotte et al., *The Pierre Auger Observatory: Contributions to the 35th International Cosmic Ray Conference (ICRC 2017)*, *arXiv e-prints* (2017) arXiv:1708.06592 [1708.06592].

[26] L.O. Drury, *REVIEW ARTICLE: An introduction to the theory of diffusive shock acceleration of energetic particles in tenuous plasmas*, *Reports on Progress in Physics* **46** (1983) 973.

[27] R. Blandford and D. Eichler, *Particle acceleration at astrophysical shocks: A theory of cosmic ray origin*, *Physics Reports* **154** (1987) 1.

[28] R. Weaver, R. McCray, J. Castor, P. Shapiro and R. Moore, *Interstellar bubbles. II. Structure and evolution.*, *ApJ* **218** (1977) 377.

[29] L. Woltjer, *Supernova Remnants*, *ARA&A* **10** (1972) 129.

[30] R. Blandford, D. Meier and A. Readhead, *Relativistic Jets from Active Galactic Nuclei*, *ARA&A* **57** (2019) 467 [1812.06025].

[31] S.E. Woosley, R.E. Taam and T.A. Weaver, *Models for Type I Supernova. I. Detonations in White Dwarfs*, *ApJ* **301** (1986) 601.

[32] H.A. Bethe, *Supernova mechanisms*, *Rev. Mod. Phys.* **62** (1990) 801.

[33] H.-T. Janka, *Explosion Mechanisms of Core-Collapse Supernovae*, *Annual Review of Nuclear and Particle Science* **62** (2012) 407 [1206.2503].

[34] G.A. Tammann, W. Loeffler and A. Schroeder, *The Galactic Supernova Rate*, *ApJS* **92** (1994) 487.

[35] W.I. Axford, E. Leer and G. Skadron, *The Acceleration of Cosmic Rays by Shock Waves*, in *International Cosmic Ray Conference*, vol. 11 of *International Cosmic Ray Conference*, p. 132, Jan., 1977.

[36] G.F. Krymskii, *A regular mechanism for the acceleration of charged particles on the front of a shock wave*, *Akademiia Nauk SSSR Doklady* **234** (1977) 1306.

[37] A.R. Bell, *The acceleration of cosmic rays in shock fronts - I.*, *MNRAS* **182** (1978) 147.

[38] R.D. Blandford and J.P. Ostriker, *Particle acceleration by astrophysical shocks.*, *ApJ* **221** (1978) L29.

[39] E. Fermi, *On the Origin of the Cosmic Radiation*, *Physical Review* **75** (1949) 1169.

[40] A.R. Bell, *Turbulent amplification of magnetic field and diffusive shock acceleration of cosmic rays*, *Monthly Notices of the Royal Astronomical Society* **353** (2004) 550.

[41] D. Caprioli and A. Spitkovsky, *Simulations of Ion Acceleration at Non-relativistic Shocks. I. Acceleration Efficiency*, *ApJ* **783** (2014) 91 [1310.2943].

[42] Y. Matsumoto, T. Amano, T.N. Kato and M. Hoshino, *Stochastic electron acceleration during spontaneous turbulent reconnection in a strong shock wave*, *Science* **347** (2015) 974 [https://science.sciencemag.org/content/347/6225/974.full.pdf].

[43] M. Pohl, M. Hoshino and J. Niemiec, *PIC simulation methods for cosmic radiation and plasma instabilities*, *Progress in Particle and Nuclear Physics* **111** (2020) 103751 [1912.02673].

[44] G.R. Blumenthal and R.J. Gould, *Bremsstrahlung, Synchrotron Radiation, and Compton Scattering of High-Energy Electrons Traversing Dilute Gases*, *Reviews of Modern Physics* **42** (1970) 237.

[45] I.V. Moskalenko and A.W. Strong, *Anisotropic Inverse Compton Scattering in the Galaxy*, *ApJ* **528** (2000) 357 [astro-ph/9811284].

[46] J.D. Jackson, *Classical Electrodynamics, 3rd Edition* (1998).

[47] H.W. Koch and J.W. Motz, *Bremsstrahlung Cross-Section Formulas and Related Data*, *Reviews of Modern Physics* **31** (1959) 920.

[48] H. Bethe and W. Heitler, *On the Stopping of Fast Particles and on the Creation of Positive Electrons*, *Proceedings of the Royal Society of London Series A* **146** (1934) 83.

[49] F.W. Stecker, *The Cosmic Gamma-Ray Spectrum from Secondary-Particle Production in the Metagalaxy*, *ApJ* **157** (1969) 507.

[50] F.A. Aharonian and A.M. Atoyan, *Broad-band diffuse gamma ray emission of the galactic disk*, *A&A* **362** (2000) 937 [astro-ph/0009009].

[51] E. Kafexhiu, F. Aharonian, A.M. Taylor and G.S. Vila, *Parametrization of gamma-ray production cross sections for p p interactions in a broad proton energy range from the kinematic threshold to PeV energies*, *Phys. Rev. D* **90** (2014) 123014 [1406.7369].

[52] R. Kulsrud and W.P. Pearce, *The Effect of Wave-Particle Interactions on the Propagation of Cosmic Rays*, *ApJ* **156** (1969) 445.

[53] J. Skilling, *Cosmic Rays in the Galaxy: Convection or Diffusion?*, *ApJ* **170** (1971) 265.

[54] J. Skilling, *Cosmic ray streaming - I. Effect of Alfvén waves on particles.*, *MNRAS* **172** (1975) 557.

[55] J. Skilling, *Cosmic ray streaming - II. Effect of particles on Alfvén waves.*, *MNRAS* **173** (1975) 245.

[56] J. Skilling, *Cosmic ray streaming - III. Self-consistent solutions.*, *MNRAS* **173** (1975) 255.

[57] M.S. Longair, *High Energy Astrophysics* (2014).

[58] M. Garcia-Munoz, G.M. Mason and J.A. Simpson, *The age of the galactic cosmic rays derived from the abundance of ^{10}Be.*, *ApJ* **217** (1977) 859.

[59] J.A. Holmes, *The confinement of galactic cosmic rays by Alfvén waves.*, *MNRAS* **170** (1975) 251.

[60] P.D. Serpico, *Entering the cosmic ray precision era*, *Journal of Astrophysics and Astronomy* **39** (2018) 41.

[61] A.W. Strong and I.V. Moskalenko, *Propagation of Cosmic-Ray Nucleons in the Galaxy*, *ApJ* **509** (1998) 212 [astro-ph/9807150].

[62] C. Evoli, D. Gaggero, D. Grasso and L. Maccione, *Cosmic ray nuclei, antiprotons and gamma rays in the galaxy: a new diffusion model*, *Journal of Cosmology and Astroparticle Physics* **2008** (2008) 018 [0807.4730].

[63] R. Kissmann, *PICARD: A novel code for the Galactic Cosmic Ray propagation problem*, *Astroparticle Physics* **55** (2014) 37 [1401.4035].

[64] M. Hanasz, A.W. Strong and P. Girichidis, *Simulations of cosmic ray propagation*, *Living Reviews in Computational Astrophysics* **7** (2021) 2 [2106.08426].

[65] B.M. Gaensler and P.O. Slane, *The Evolution and Structure of Pulsar Wind Nebulae*, *ARA&A* **44** (2006) 17 [astro-ph/0601081].

[66] F. Aharonian, A.G. Akhperjanian, A.R. Bazer-Bachi, M. Beilicke, W. Benbow, D. Berge et al., *Energy dependent γ-ray morphology in the pulsar wind nebula HESS J1825-137*, *A&A* **460** (2006) 365 [astro-ph/0607548].

[67] H. E. S. S. Collaboration, H. Abdalla, F. Aharonian, F. Ait Benkhali, E.O. Angüner, M. Arakawa et al., *Particle transport within the pulsar wind nebula HESS J1825-137*, *A&A* **621** (2019) A116 [1810.12676].

[68] H. E. S. S. Collaboration, H. Abdalla, A. Abramowski, F. Aharonian, F. Ait Benkhali, A.G. Akhperjanian et al., *H.E.S.S. observations of RX J1713.7-3946 with improved angular and spectral resolution: Evidence for gamma-ray emission extending beyond the X-ray emitting shell*, *A&A* **612** (2018) A6 [1609.08671].

[69] S.P. Reynolds, *Supernova remnants at high energy.*, *ARA&A* **46** (2008) 89.

[70] G. Taylor, *The Formation of a Blast Wave by a Very Intense Explosion. I. Theoretical Discussion*, *Proceedings of the Royal Society of London Series A* **201** (1950) 159.

[71] L.I. Sedov, *Similarity and Dimensional Methods in Mechanics* (1959).

[72] G. Morlino and D. Caprioli, *Strong evidence for hadron acceleration in Tycho's supernova remnant*, *A&A* **538** (2012) A81 [1105.6342].

[73] M. Ackermann, M. Ajello, A. Allafort, L. Baldini, J. Ballet, G. Barbiellini et al., *Detection of the Characteristic Pion-Decay Signature in Supernova Remnants*, *Science* **339** (2013) 807 [1302.3307].

[74] A.U. Abeysekara, A. Archer, W. Benbow, R. Bird, R. Brose, M. Buchovecky et al., *Evidence for Proton Acceleration up to TeV Energies Based on VERITAS and Fermi-LAT Observations of the Cas A SNR*, *ApJ* **894** (2020) 51 [2003.13615].

[75] V. Bosch-Ramon, G.E. Romero and J.M. Paredes, *A broadband leptonic model for gamma-ray emitting microquasars*, *A&A* **447** (2006) 263 [astro-ph/0509086].

[76] A. Sierpowska-Bartosik and D.F. Torres, *Pulsar wind zone processes in LS 5039*, *Astroparticle Physics* **30** (2008) 239 [0801.3427].

[77] H. E. S. S. Collaboration, H. Abdalla, R. Adam, F. Aharonian, F. Ait Benkhali, E.O. Angüner et al., *Detection of very-high-energy γ-ray emission from the colliding wind binary η Car with H.E.S.S.*, *A&A* **635** (2020) A167 [2002.02336].

[78] H. E. S. S. Collaboration, A. Abramowski, F. Aharonian, F. Ait Benkhali, A.G. Akhperjanian, E.O. Angüner et al., *H.E.S.S. detection of TeV emission from the interaction region between the supernova remnant G349.7+0.2 and a molecular cloud*, *A&A* **574** (2015) A100 [1412.2251].

[79] F. Aharonian, A.G. Akhperjanian, A.R. Bazer-Bachi, M. Beilicke, W. Benbow, D. Berge et al., *Detection of extended very-high-energy γ-ray emission towards the young stellar cluster Westerlund 2*, *A&A* **467** (2007) 1075 [astro-ph/0703427].

[80] A. Abramowski, F. Acero, F. Aharonian, A.G. Akhperjanian, G. Anton, A. Balzer et al., *Discovery of extended VHE γ-ray emission from the vicinity of the young massive stellar cluster Westerlund 1*, *A&A* **537** (2012) A114 [1111.2043].

[81] A.M. Bykov, *Particle Acceleration and Nonthermal Phenomena in Superbubbles*, *Space Sci. Rev.* **99** (2001) 317.

[82] D. Gaggero, A. Urbano, M. Valli and P. Ullio, *Gamma-ray sky points to radial gradients in cosmic-ray transport*, *Phys. Rev. D* **91** (2015) 083012.

[83] A.A. Abdo, B. Allen, T. Aune, D. Berley, E. Blaufuss, S. Casanova et al., *A Measurement of the Spatial Distribution of Diffuse TeV Gamma-Ray Emission from the Galactic Plane with Milagro*, *ApJ* **688** (2008) 1078 [0805.0417].

[84] A. Abramowski, F. Aharonian, F. Ait Benkhali, A.G. Akhperjanian, E.O. Angüner, M. Backes et al., *Diffuse Galactic gamma-ray emission with H.E.S.S.*, *Phys. Rev. D* **90** (2014) 122007 [1411.7568].

[85] B. Bartoli, P. Bernardini, X.J. Bi, P. Branchini, A. Budano, P. Camarri et al., *Study of the Diffuse Gamma-Ray Emission from the Galactic Plane with ARGO-YBJ*, *ApJ* **806** (2015) 20 [1507.06758].

[86] M. Amenomori, Y.W. Bao, X.J. Bi, D. Chen, T.L. Chen, W.Y. Chen et al., *First Detection of sub-PeV Diffuse Gamma Rays from the Galactic Disk: Evidence for Ubiquitous Galactic Cosmic Rays beyond PeV Energies*, *Phys. Rev. Lett.* **126** (2021) 141101 [2104.05181].

[87] S. Casanova and B.L. Dingus, *Constraints on the TeV source population and its contribution to the galactic diffuse TeV emission*, *Astroparticle Physics* **29** (2008) 63 [0711.2753].

[88] H. E. S. S. Collaboration, H. Abdalla, A. Abramowski, F. Aharonian, F. Ait Benkhali, E.O. Angüner et al., *The H.E.S.S. Galactic plane survey*, *A&A* **612** (2018) A1 [1804.02432].

[89] T.C. Weekes, M.F. Cawley, D.J. Fegan, K.G. Gibbs, A.M. Hillas, P.W. Kowk et al., *Observation of TeV Gamma Rays from the Crab Nebula Using the Atmospheric Cerenkov Imaging Technique*, *ApJ* **342** (1989) 379.

[90] K. Bernlöhr, O. Carrol, R. Cornils, S. Elfahem, P. Espigat, S. Gillessen et al., *The optical system of the H.E.S.S. imaging atmospheric Cherenkov telescopes. Part I: layout and components of the system*, *Astroparticle Physics* **20** (2003) 111 [astro-ph/0308246].

[91] P. Hofverberg, R. Kankanyan, M. Panter, G. Hermann, W. Hofmann, C. Deil et al., *Commissioning and Initial Performance of the H.E.S.S. II Drive System*, in *International Cosmic Ray Conference*, vol. 33 of *International Cosmic Ray Conference*, p. 3092, Jan., 2013 [1307.4550].

[92] D. Lennarz, P.M. Chadwick, W. Domainko, R.D. Parsons, G. Rowell and P.H.T. Tam, *Searching for TeV emission from GRBs: the status of the H.E.S.S. GRB programme*, *arXiv e-prints* (2013) arXiv:1307.6897 [1307.6897].

[93] T. Ashton, M. Backes, A. Balzer, D. Berge, J. Bolmont, S. Bonnefoy et al., *A NECTAr-based upgrade for the Cherenkov cameras of the H.E.S.S. 12-meter telescopes*, *Astroparticle Physics* **118** (2020) 102425 [2001.04510].

[94] HESS Collaboration, A. Abramowski, F. Aharonian, F.A. Benkhali, A.G. Akhperjanian, E.O. Angüner et al., *Acceleration of petaelectronvolt protons in the Galactic Centre*, *Nature* **531** (2016) 476 [1603.07730].

[95] H. E. S. S. Collaboration, *Resolving the Crab pulsar wind nebula at teraelectronvolt energies*, *Nature Astronomy* **4** (2020) 167 [1909.09494].

[96] H. Abdalla, R. Adam, F. Aharonian, F. Ait Benkhali, E.O. Angüner, M. Arakawa et al., *A very-high-energy component deep in the γ-ray burst afterglow*, *Nature* **575** (2019) 464 [1911.08961].

[97] H. E. S. S. Collaboration, H. Abdalla, R. Adam, F. Aharonian, F. Ait Benkhali, E.O. Angüner et al., *Resolving acceleration to very high energies along the jet of Centaurus A*, *Nature* **582** (2020) 356 [2007.04823].

[98] W. Heitler, *The Quantum Theory of Radiation*, Dover Pubn Inc, 3 ed. (1984).

[99] F. Schmidt and J. Knapp, "Corsika shower images." https://www-zeuthen.desy.de/~jknapp/fs/showerimages.html, 2005.

[100] P.A. Čerenkov, *Visible Radiation Produced by Electrons Moving in a Medium with Velocities Exceeding that of Light*, *Physical Review* **52** (1937) 378.

[101] I.M. Frank and I. Tamm, *Coherent visible radiation of fast electrons passing through matter*, *C. R. Acad. Sci. URSS* **14** (1937) 109.

[102] M. De Naurois, *Very High Energy astronomy from H.E.S.S. to CTA. Opening of a new astronomical window on the non-thermal Universe*, habilitation à diriger des recherches, Université Pierre et Marie Curie - Paris VI, Mar., 2012.

[103] F.A. Aharonian, W. Hofmann, A.K. Konopelko and H.J. Völk, *The potential of ground based arrays of imaging atmospheric Cherenkov telescopes. I. Determination of shower parameters*, *Astroparticle Physics* **6** (1997) 343.

[104] S. Funk, G. Hermann, J. Hinton, D. Berge, K. Bernlöhr, W. Hofmann et al., *The trigger system of the H.E.S.S. telescope array*, *Astroparticle Physics* **22** (2004) 285 [astro-ph/0408375].

[105] G. Giavitto, S. Bonnefoy, T. Ashton, M. Backes, A. Balzer, D. Berge et al., *Performance of the upgraded H.E.S.S. cameras*, in *35th International Cosmic Ray Conference (ICRC2017)*, vol. 301 of *International Cosmic Ray Conference*, p. 805, Jan., 2017, DOI [1708.04550].

[106] J. Zorn, *Sensitivity Improvements of Very-High-Energy Gamma-Ray Detection with the Upgraded H.E.S.S. I Cameras using Full Waveform Processing*, in *36th International Cosmic Ray Conference (ICRC2019)*, vol. 36 of *International Cosmic Ray Conference*, p. 834, July, 2019 [1908.04620].

[107] D. Heck, J. Knapp, J.N. Capdevielle, G. Schatz and T. Thouw, *CORSIKA: a Monte Carlo code to simulate extensive air showers.* (1998).

[108] K. Bernlöhr, *Simulation of imaging atmospheric Cherenkov telescopes with CORSIKA and sim_ telarray*, *Astroparticle Physics* **30** (2008) 149 [0808.2253].

[109] M.P. Kertzman and G.H. Sembroski, *Computer simulation methods for investigating the detection characteristics of TeV air Cherenkov telescopes*, *Nuclear Instruments and Methods in Physics Research A* **343** (1994) 629.

[110] F. Aharonian, A.G. Akhperjanian, K.M. Aye, A.R. Bazer-Bachi, M. Beilicke, W. Benbow et al., *Calibration of cameras of the H.E.S.S. detector*, *Astroparticle Physics* **22** (2004) 109.

[111] N. Leroy, O. Bolz, J. Guy, I. Jung, I. Redondo, L. Rolland et al., *Calibration Results for the First Two H.E.S.S. Array Telescopes*, in *International Cosmic Ray Conference*, vol. 5 of *International Cosmic Ray Conference*, p. 2895, July, 2003.

[112] G. Vacanti, P. Fleury, Y. Jiang, E. Paré, A.C. Rovero, X. Sarazin et al., *Muon ring images with an atmospheric Čerenkov telescope*, *Astroparticle Physics* **2** (1994) 1.

[113] F. Aharonian, A.G. Akhperjanian, A.R. Bazer-Bachi, M. Beilicke, W. Benbow, D. Berge et al., *Observations of the Crab nebula with HESS*, *A&A* **457** (2006) 899 [astro-ph/0607333].

[114] A.M. Hillas, *Cerenkov Light Images of EAS Produced by Primary Gamma Rays and by Nuclei*, in *19th International Cosmic Ray Conference (ICRC19), Volume 3*, vol. 3 of *International Cosmic Ray Conference*, p. 445, Aug., 1985.

[115] M. de Naurois, *Analysis methods for Atmospheric Cerenkov Telescopes*, *arXiv e-prints* (2006) astro [astro-ph/0607247].

[116] S. Ohm, C. van Eldik and K. Egberts, *γ/hadron separation in very-high-energy γ-ray astronomy using a multivariate analysis method*, *Astroparticle Physics* **31** (2009) 383 [0904.1136].

[117] T. Murach, M. Gajdus and R.D. Parsons, *A Neural Network-Based Monoscopic Reconstruction Algorithm for H.E.S.S. II*, *arXiv e-prints* (2015) arXiv:1509.00794 [1509.00794].

[118] S. Le Bohec, B. Degrange, M. Punch, A. Barrau, R. Bazer-Bachi, H. Cabot et al., *A new analysis method for very high definition imaging atmospheric Cherenkov telescopes as applied to the CAT telescope.*, *Nuclear Instruments and Methods in Physics Research A* **416** (1998) 425 [astro-ph/9804133].

[119] M. de Naurois and L. Rolland, *A high performance likelihood reconstruction of γ-rays for imaging atmospheric Cherenkov telescopes*, *Astroparticle Physics* **32** (2009) 231 [0907.2610].

[120] R.D. Parsons and J.A. Hinton, *A Monte Carlo template based analysis for air-Cherenkov arrays*, *Astroparticle Physics* **56** (2014) 26 [1403.2993].

[121] D. Berge, S. Funk and J. Hinton, *Background modelling in very-high-energy γ-ray astronomy*, *A&A* **466** (2007) 1219 [astro-ph/0610959].

[122] T.P. Li and Y.Q. Ma, *Analysis methods for results in gamma-ray astronomy.*, *ApJ* **272** (1983) 317.

[123] W. Cash, *Parameter estimation in astronomy through application of the likelihood ratio.*, *ApJ* **228** (1979) 939.

[124] S.S. Wilks, *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, *The Annals of Mathematical Statistics* **9** (1938) 60 .

[125] F. Piron, A. Djannati-Atai, M. Punch, J.P. Tavernet, A. Barrau, R. Bazer-Bachi et al., *Temporal and spectral gamma-ray properties of Mkn 421 above 250 GeV from CAT observations between 1996 and 2000*, *A&A* **374** (2001) 895 [astro-ph/0106196].

[126] M. Ackermann, M. Ajello, W.B. Atwood, L. Baldini, J. Ballet, G. Barbiellini et al., *2FHL: The Second Catalog of Hard Fermi-LAT Sources*, *ApJS* **222** (2016) 5 [1508.04449].

[127] F. Acero, M. Ackermann, M. Ajello, A. Albert, W.B. Atwood, M. Axelsson et al., *Fermi Large Area Telescope Third Source Catalog*, *ApJS* **218** (2015) 23 [1501.02003].

[128] G. Ferrand and S. Safi-Harb, *A census of high-energy observations of Galactic supernova remnants*, *Advances in Space Research* **49** (2012) 1313 [1202.0245].

[129] R.N. Manchester, G.B. Hobbs, A. Teoh and M. Hobbs, *The Australia Telescope National Facility Pulsar Catalogue*, *AJ* **129** (2005) 1993 [astro-ph/0412641].

[130] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press (2016).

[131] Y. Lecun, Y. Bengio and G. Hinton, *Deep learning*, *Nature* **521** (2015) 436.

[132] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra et al., *Playing Atari with Deep Reinforcement Learning*, *arXiv e-prints* (2013) arXiv:1312.5602 [1312.5602].

[133] X. Glorot, A. Bordes and Y. Bengio, *Deep sparse rectifier neural networks*, in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, G. Gordon, D. Dunson and M. Dudk, eds., vol. 15 of *Proceedings of Machine Learning Research*, (Fort Lauderdale, FL, USA), pp. 315–323, JMLR

Workshop and Conference Proceedings, 11–13 Apr, 2011, http://proceedings.mlr.press/v15/glorot11a.html.

[134] K. He, X. Zhang, S. Ren and J. Sun, *Deep Residual Learning for Image Recognition*, *arXiv e-prints* (2015) arXiv:1512.03385 [1512.03385].

[135] Y. LeCun et al., *Generalization and network design strategies*, *Connectionism in perspective* **19** (1989) 143.

[136] D.E. Rumelhart, G.E. Hinton and R.J. Williams, *Learning representations by back-propagating errors*, *Nature* **323** (1986) 533.

[137] B. Polyak, *Some methods of speeding up the convergence of iteration methods*, *USSR Computational Mathematics and Mathematical Physics* **4** (1964) 1.

[138] J. Duchi, E. Hazan and Y. Singer, *Adaptive subgradient methods for online learning and stochastic optimization*, *Journal of Machine Learning Research* **12** (2011) 2121.

[139] M.D. Zeiler, *ADADELTA: An Adaptive Learning Rate Method*, *arXiv e-prints* (2012) arXiv:1212.5701 [1212.5701].

[140] D.P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, *arXiv e-prints* (2014) arXiv:1412.6980 [1412.6980].

[141] X. Glorot and Y. Bengio, *Understanding the difficulty of training deep feedforward neural networks*, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Y.W. Teh and M. Titterington, eds., vol. 9 of *Proceedings of Machine Learning Research*, (Chia Laguna Resort, Sardinia, Italy), pp. 249–256, PMLR, 13–15 May, 2010, https://proceedings.mlr.press/v9/glorot10a.html.

[142] S. Ioffe and C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift*, *arXiv e-prints* (2015) arXiv:1502.03167 [1502.03167].

[143] K. He, X. Zhang, S. Ren and J. Sun, *Identity Mappings in Deep Residual Networks*, *arXiv e-prints* (2016) arXiv:1603.05027 [1603.05027].

[144] S. Hanson and L. Pratt, *Comparing biases for minimal network construction with back-propagation*, in *Advances in Neural Information Processing Systems*, D. Touretzky, ed., vol. 1, Morgan-Kaufmann, 1989, https://proceedings.neurips.cc/paper/1988/file/1c9ac0159c94d8d0cbedc973445af2da-Paper.pdf.

[145] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, *Dropout: A simple way to prevent neural networks from overfitting*, *Journal of Machine Learning Research* **15** (2014) 1929.

[146] W.M. Kouw and M. Loog, *An introduction to domain adaptation and transfer learning*, *arXiv e-prints* (2018) arXiv:1812.11806 [1812.11806].

[147] J. Albert, E. Aliu, H. Anderhub, P. Antoranz, A. Armada, M. Asensio et al., *Implementation of the Random Forest method for the Imaging Atmospheric Cherenkov Telescope MAGIC*, *Nuclear Instruments and Methods in Physics Research A* **588** (2008) 424 [0709.3719].

[148] M. Krause, E. Pueschel and G. Maier, *Improved γ/hadron separation for the detection of faint γ-ray sources using boosted decision trees*, *Astroparticle Physics* **89** (2017) 1 [1701.06928].

[149] R. Mersereau, *The processing of hexagonally sampled two-dimensional signals*, *Proceedings of the IEEE* **67** (1979) 930.

[150] R.C. Staunton and N. Storey, *A Comparison Between Square and Hexagonal Sampling Methods for Pipeline Image Processing*, in *Optics, Illumination, and Image Sensing for Machine Vision IV*, D.J. Svetkoff, ed., vol. 1194, pp. 142 – 151, International Society for Optics and Photonics, SPIE, 1990, https://doi.org/10.1117/12.969847.

[151] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro et al., *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*, *arXiv e-prints* (2016) arXiv:1603.04467 [1603.04467].

[152] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan et al., *Pytorch: An imperative style, high-performance deep learning library*, in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett, eds., pp. 8024–8035, Curran Associates, Inc. (2019), http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[153] Q. Feng, T.T.Y. Lin and VERITAS Collaboration, *The analysis of VERITAS muon images using convolutional neural networks*, in *Astroinformatics*, M. Brescia, S.G. Djorgovski, E.D. Feigelson, G. Longo and S. Cavuoti, eds., vol. 325, pp. 173–179, June, 2017, DOI [1611.09832].

[154] T. Lukas Holch, I. Shilon, M. Büchele, T. Fischer, S. Funk, N. Groeger et al., *Probing Convolutional Neural Networks for Event Reconstruction in γ-Ray Astronomy with Cherenkov Telescopes*, *arXiv e-prints* (2017) arXiv:1711.06298 [1711.06298].

[155] C. Steppa and T.L. Holch, *Hexagdly - processing hexagonally sampled data with cnns in pytorch*, *SoftwareX* **9** (2019) 193.

[156] S. Hochreiter and J. Schmidhuber, *Long Short-Term Memory*, *Neural Computation* **9** (1997) 1735.

[157] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez et al., *Attention Is All You Need*, *arXiv e-prints* (2017) arXiv:1706.03762 [1706.03762].

[158] B.S. Acharya, M. Actis, T. Aghajani, G. Agnetta, J. Aguilar, F. Aharonian et al., *Introducing the CTA concept*, *Astroparticle Physics* **43** (2013) 3.

[159] F.A. Gers, N.N. Schraudolph and J. Schmidhuber, *Learning precise timing with lstm recurrent networks*, *J. Mach. Learn. Res.* **3** (2003) 115?143.

[160] I. Shilon, M. Kraus, M. Büchele, K. Egberts, T. Fischer, T.L. Holch et al., *Application of deep learning methods to analysis of imaging atmospheric Cherenkov telescopes data*, *Astroparticle Physics* **105** (2019) 44 [1803.10698].

[161] R.D. Parsons and S. Ohm, *Background rejection in atmospheric Cherenkov telescopes using recurrent convolutional neural networks*, *European Physical Journal C* **80** (2020) 363 [1910.09435].

[162] O. Vinyals, S. Bengio and M. Kudlur, *Order Matters: Sequence to sequence for sets*, p. arXiv:1511.06391, Nov., 2015 [1511.06391].

[163] F. Aharonian, A.G. Akhperjanian, A.R. Bazer-Bachi, B. Behera, M. Beilicke, W. Benbow et al., *An Exceptional Very High Energy Gamma-Ray Flare of PKS 2155-304*, *ApJ* **664** (2007) L71 [0706.0797].

[164] G. Maier and J. Knapp, *Cosmic-ray events as background in imaging atmospheric Cherenkov telescopes*, *Astroparticle Physics* **28** (2007) 72 [0704.3567].

[165] C. Steppa and K. Egberts, *Modelling the Galactic very-high-energy γ-ray source population*, *A&A* **643** (2020) A137 [2010.03305].

[166] M. Renaud, *Latest results on Galactic sources as seen in VHE gamma-rays*, *arXiv e-prints* (2009) arXiv:0905.1287 [0905.1287].

[167] S. Funk, J.A. Hinton, G. Hermann and S. Digel, *A future very-high-energy view of our Galaxy*, in *American Institute of Physics Conference Series*, F.A. Aharonian, W. Hofmann and F. Rieger, eds., vol. 1085 of *American Institute of Physics Conference Series*, pp. 886–889, Dec., 2008, DOI [0901.1885].

[168] P. Cristofari, S. Gabici, T.B. Humensky, M. Santander, R. Terrier, E. Parizot et al., *Supernova remnants in the very-high-energy gamma-ray domain: the role of the Cherenkov telescope array*, *MNRAS* **471** (2017) 201 [1709.01102].

[169] G. Dubus, J.L. Contreras, S. Funk, Y. Gallant, T. Hassan, J. Hinton et al., *Surveys with the Cherenkov Telescope Array*, *Astroparticle Physics* **43** (2013) 317 [1208.5686].

[170] W.T. Sanders, W.L. Kraushaar, J.A. Nousek and P.M. Fried, *Soft diffuse X-rays in the southern galactic hemisphere.*, *ApJ* **217** (1977) L87.

[171] D.A. Green, *Constraints on the distribution of supernova remnants with Galactocentric radius*, *MNRAS* **454** (2015) 1517 [1508.02931].

[172] J.-W. Xu, X.-Z. Zhang and J.-L. Han, *Statistics of Galactic Supernova Remnants*, *Chinese Journal of Astronomy & Astrophysics* **5** (2005) 165.

[173] I. Yusifov and I. Küçük, *Revisiting the radial distribution of pulsars in the Galaxy*, *A&A* **422** (2004) 545 [astro-ph/0405559].

[174] D.R. Lorimer, A.J. Faulkner, A.G. Lyne, R.N. Manchester, M. Kramer, M.A. McLaughlin et al., *The Parkes Multibeam Pulsar Survey - VI. Discovery and timing of 142 pulsars and a Galactic population analysis*, *MNRAS* **372** (2006) 777 [astro-ph/0607640].

[175] R. Kissmann, M. Werner, O. Reimer and A.W. Strong, *Propagation in 3D spiral-arm cosmic-ray source distribution models and secondary particle production using PICARD*, *Astroparticle Physics* **70** (2015) 39 [1504.08249].

[176] T.Y. Steiman-Cameron, M. Wolfire and D. Hollenbach, *COBE and the Galactic Interstellar Medium: Geometry of the Spiral Arms from FIR Cooling Lines*, *ApJ* **722** (2010) 1460.

[177] J.M. Cordes and T.J.W. Lazio, *NE2001.I. A New Model for the Galactic Distribution of Free Electrons and its Fluctuations*, *arXiv e-prints* (2002) astro [astro-ph/0207156].

[178] R.A. Benjamin, E. Churchwell, B.L. Babler, R. Indebetouw, M.R. Meade, B.A. Whitney et al., *First GLIMPSE Results on the Stellar Structure of the Galaxy*, *ApJ* **630** (2005) L149 [astro-ph/0508325].

[179] F. Anders, A. Khalatyan, C. Chiappini, A.B. Queiroz, B.X. Santiago, C. Jordi et al., *Photo-astrometric distances, extinctions, and astrophysical parameters for Gaia DR2 stars brighter than G = 18*, *A&A* **628** (2019) A94 [1904.11302].

[180] M. Werner, R. Kissmann, A.W. Strong and O. Reimer, *Spiral arms as cosmic ray source distributions*, *Astroparticle Physics* **64** (2015) 18 [1410.5266].

[181] A.W. Strong, *Source population synthesis and the Galactic diffuse gamma-ray emission*, *Ap&SS* **309** (2007) 35 [astro-ph/0609359].

[182] L. Ambrogi, E. De Oña Wilhelmi and F. Aharonian, *On the potential of atmospheric Cherenkov telescope arrays for resolving TeV gamma-ray sources in the Galactic plane*, *Astroparticle Physics* **80** (2016) 22 [1603.04365].

[183] F. Aharonian, A.G. Akhperjanian, A.R. Bazer-Bachi, M. Beilicke, W. Benbow, D. Berge et al., *Discovery of very-high-energy γ-rays from the Galactic Centre ridge*, *Nature* **439** (2006) 695 [astro-ph/0603021].

[184] I. Yusifov, *Pulsars and the Warp of the Galaxy*, in *The Magnetized Interstellar Medium*, B. Uyaniker, W. Reich and R. Wielebinski, eds., pp. 165–169, Feb., 2004 [astro-ph/0405517].

[185] M.T. Karim and E.E. Mamajek, *Revised geometric estimates of the North Galactic Pole and the Sun's height above the Galactic mid-plane*, *MNRAS* **465** (2017) 472 [1610.08125].

[186] A.A. Abdo, B. Allen, D. Berley, S. Casanova, C. Chen, D.G. Coyne et al., *TeV Gamma-Ray Sources from a Survey of the Galactic Plane with Milagro*, *ApJ* **664** (2007) L91 [0705.0707].

[187] A.U. Abeysekara, A. Albert, R. Alfaro, C. Alvarez, J.D. Álvarez, R. Arceo et al., *The 2HWC HAWC Observatory Gamma-Ray Catalog*, *ApJ* **843** (2017) 40 [1702.02992].

[188] Cherenkov Telescope Array Consortium, B.S. Acharya, I. Agudo, I. Al Samarai, R. Alfaro, J. Alfaro et al., *Science with the Cherenkov Telescope Array* (2019), 10.1142/10986.

[189] A.W. Strong, T.A. Porter, S.W. Digel, G. Jóhannesson, P. Martin, I.V. Moskalenko et al., *Global Cosmic-ray-related Luminosity and Energy Budget of the Milky Way*, *ApJ* **722** (2010) L58 [1008.4330].

Acknowledgement/Danksagung

Die Zeit während meiner Doktorarbeit war eine lehrreiche, vor allem aber auch eine freudige Zeit, an die ich mich gern zurück erinnern werde. Dies lag insbesondere an den Menschen, mit denen ich zusammenarbeiten durfte und die mich auf diesem Weg begleitet haben. Diesen Menschen gilt mein Dank.

Zuerst möchte ich mich bei Prof. Dr. Christian Stegmann bedanken, der mich während meines Studiums mit seiner Begeisterung und seinem Enthusiasmus für die Astroteilchenphysik angesteckt hat und es mir seitdem ermöglicht hat, mich ebenfalls in diesem Forschungsfeld zu betätigen.

Mein Dank gilt auch Prof. Dr. Martin Pohl und Prof. Dr. Christopher van Eldik dafür, dass sie sich bereit erklärt haben diese Arbeit zu begutachten.

Ganz besonders möchte ich mich bei meiner Mentorin Dr. Kathrin Egberts bedanken. Neben ihrer Expertise, von der ich viel während unserer Zusammenarbeit gelernt habe, schätzte ich ihr Engagement in der Förderung meiner persönlichen Weiterentwicklung als Wissenschaftler und den freundschaftlichen, persönlichen Umgang miteinander, der die ganze Gruppe an der Universität Potsdam prägte und Grund für schöne Zeit dort war. Vielen Dank dafür!

Ich hatte außerdem das Glück eine Vielzahl netter Menschen als Kollegen kennenzulernen. Für stets unterhaltsame Pausengespräche danke ich allen in der Astro-Gruppe an der Universität Potsdam sowie der Gamma-Gruppe am DESY. Persönlich möchte ich mich bei Clemens, Tim und Victor, die mir darüber hinaus zu guten Freunden wurden, für eine tolle Zeit auch außerhalb der Arbeit bedanken.

Auch wenn ich viel Freude während meiner Doktorarbeit hatte, war es nicht immer einfach. Der Rückhalt in meiner Familie und die Liebe meiner Frau Laura waren mir eine wichtige Stütze. Euch gilt mein größter Dank!

Erklärung

Hiermit erkläre ich, dass ich diese Arbeit selbstständig angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Direkt oder dem Sinn nach übernommene Gedanken aus anderen Werken sind als solche kenntlich gemacht.

Die Arbeit wurde weder einer anderen Prüfungsbehörde vorgelegt noch veröffentlicht.

Ort, Datum