Search for Light Primordial Black Holes with VERITAS using γ-ray and Optical Observations

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)

in der Wissenschaftsdisziplin
Astroteilchen Physik

eingereicht an der
Mathematisch-Naturwissenschaflichen Fakultät
Institut für Physik und Astronomie
der Universität Potsdam

angefertigt am
Deutschen Elektronen Synchrotron

vorgelegt von

Konstantin Johannes Pfrang

Potsdam, den 26. Oktober, 2022

Diputation erfolgt am: 14. Februar, 2023

Betreuerin:
Dr. Elisa Pueschel

Gutachter / Gutachterinnen:
Dr. Elisa Pueschel
Prof. Dr. Martin Pohl
Prof. Dr. Michele Doro

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-58726
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-587266

Abstract

The Very Energetic Radiation Imaging Telescope Array System (VERITAS) is an array of four imaging atmospheric Cherenkov telescopes (IACTs). VERITAS is sensitive to very-high-energy γ-rays in the range of 100 GeV to > 30 TeV. Hypothesized primordial black holes (PBHs) are attractive targets for IACTs. If they exist, their potential cosmological impact reaches beyond the candidacy for constituents of dark matter. The sublunar mass window $10^{17} \lesssim M/\mathrm{g} \lesssim 10^{23}$ is the largest unconstrained range of PBH masses. This thesis aims to develop novel concepts searching for light PBHs with VERITAS. PBHs below the sublunar window lose mass due to Hawking radiation. They would evaporate at the end of their lifetime, leading to a short burst of γ-rays. If PBHs formed at $\sim 10^{15}$ g, the evaporation would occur nowadays. Detecting these signals might not only confirm the existence of PBHs but also prove the theory of Hawking radiation. This thesis probes archival VERITAS data recorded between 2012 and 2021 for possible PBH signals. This work presents a new automatic approach to assess the quality of the VERITAS data. The array-trigger rate and far infrared temperature are well suited to identify periods with poor data quality. These are masked by time cuts to obtain a consistent and clean dataset which contains about 4222 hours. The PBH evaporations could occur at any location in the field of view or time within this data. Only a blind search can be performed to identify these short signals. This thesis implements a data-driven deep learning based method to search for short transient signals with VERITAS. It does not depend on the modelling of the effective area and radial acceptance. This work presents the first application of this method to actual observational IACT data. This thesis develops new concepts dealing with the specifics of the data and the transient detection method. These are reflected in the developed data preparation pipeline and search strategies. After correction for trial factors, no candidate PBH evaporation is found in the data. Thus, new constraints of the local rate of PBH evaporations are derived. At the 99% confidence limit it is below $< 1.07 \times 10^5\,\mathrm{pc}^{-3}\mathrm{yr}^{-1}$. This constraint with the new, independent analysis approach is in the range of existing limits for the evaporation rate.

This thesis also investigates an alternative novel approach to searching for PBHs with IACTs. Above the sublunar window, the PBH abundance is constrained by optical microlensing studies. The sampling speed, which is of order of minutes to hours for traditional optical telescopes, is a limiting factor in expanding the limits to lower masses. IACTs are also powerful instruments for fast transient optical astronomy with up to $\mathcal{O}$(ns) sampling. This thesis investigates whether IACTs might constrain the sublunar window with optical microlensing observations. This study confirms that, in principle, the fast sampling speed might allow extending microlensing searches into the sublunar mass window. However, the limiting factor for IACTs is the modest sensitivity to detect changes in optical fluxes. This thesis presents the expected rate of detectable events for VERITAS as well as prospects of possible future next-generation IACTs. For VERITAS, the rate of detectable microlensing events in the sublunar range is $\sim 10^{-6}$ per year of observation time. The future prospects for a 100 times more sensitive instrument are at ~ 0.05 events per year.

Kurzfassung

Das Very Energetic Radiation Imaging Telescope Array System (VERITAS) ist ein Instrument mit vier atmosphärischen Cherenkov-Teleskopen (IACTs). VERITAS ist empfindlich für sehr hoch-energetische γ-Strahlung im Bereich von 100 GeV bis $>$ 30 TeV. Hypothetische primordiale Schwarze Löcher (PBHs) sind interessante Ziele für IACTs. Falls sie existieren, könnte ihr potentieller kosmologischer Einfluss über die Möglichkeit, dass sie ein Bestandteil der dunklen Materie sind, hinausgehen. Der größte nicht eingeschränkte Bereich der PBH-Massen ist das sublunare Fenster $10^{17} \lesssim M/\mathrm{g} \lesssim 10^{23}$. Das Ziel dieser Arbeit ist es, neue Konzepte für die Suche nach leichten PBHs mit VERITAS zu entwickeln. Durch die Hawking-Strahlung verlieren PBHs unterhalb des sublunaren Fensters an Masse. Am Ende ihrer Lebenszeit verdampfen diese, was einen kurzen Ausbruch an γ-Strahlung verursacht. Falls PBHs mit $\sim 10^{15}$ g entstanden sind, würde sich dieser Ausbruch in der heutigen Zeit ereignen. Der Nachweis dieser Signale könnte nicht nur die Existenz von PBHs bestätigen, sondern auch die Theorie der Hawking-Strahlung beweisen. In dieser Arbeit werden VERITAS-Daten aus den Jahren 2012 bis 2021 auf mögliche PBH-Signale untersucht. Es wird ein neuer automatisierter Ansatz zur Beurteilung der Qualität der VERITAS-Daten vorgestellt. Die Array-Trigger-Rate und die ferne Infrarot-Temperatur sind gut geeignet, um Zeiträume mit schlechter Datenqualität zu identifizieren. Diese werden maskiert, um einen konsistenten Datensatz zu erhalten, der etwa 4222 Stunden umfasst. Die PBH-Verdampfungen könnten an jeder beliebigen Stelle im Sichtfeld oder zu jeder beliebigen Zeit innerhalb dieser Daten auftreten. Zur Identifizierung dieser kurzen Signale kann nur eine Blindsuche durchgeführt werden. In dieser Arbeit wird eine datengestützte, auf Deep Learning basierende Methode zur Suche nach kurzen vorübergehenden Signalen mit VERITAS implementiert. Die Methode ist nicht von der Modellierung der effektiven Fläche und der radialen Akzeptanz abhängig. Diese Arbeit präsentiert die erste Anwendung dieser Methode mit echten IACT-Beobachtungsdaten. In dieser Arbeit werden neue Konzepte entwickelt, die sich mit den Besonderheiten der Daten und der Methode befassen. Sie spiegeln sich in der entwickelten Datenvorbereitung und den Suchstrategien wider. Nach Korrektur der Versuchsfaktoren wird in den Daten kein Kandidat für PBH-Verdampfung gefunden. Daher wird die lokale Rate von PBH-Verdampfungen auf unter $< 1,07 \times 10^5\,\mathrm{pc}^{-3}\mathrm{yr}^{-1}$ an der 99%-Konfidenzgrenze beschränkt. Dieses Limit, welches mit dem neuen, unabhängigen Analyseansatz erreicht wurde, liegt im Bereich der bestehenden Grenzwerte für die Verdunstungsrate.

In dieser Arbeit wird auch ein alternativer neuer Ansatz für die Suche nach PBHs mit IACTs untersucht. Oberhalb des sublunaren Fensters wird die Existenz von PBHs durch optische Mikrolensing-Studien eingeschränkt. Für niedrige Massen ist die Abtastgeschwindigkeit, die bei herkömmlichen optischen Teleskopen in der Größenordnung von Minuten bis Stunden liegt, ein limitierender Faktor. IACTs sind auch leistungsstarke Instrumente für die schnelle optische Astronomie mit Abtastraten von bis zu $\mathcal{O}$(ns). In dieser Arbeit wird untersucht, ob IACTs das sublunare Fenster mit optischen Mikrolensing-Beobachtungen beschränken könnten. Diese Studie bestätigt, dass die schnelle Abtastgeschwindigkeit eine Ausweitung der Mikrolensing-Suche auf das sublunare Massenfenster ermöglichen könnte. Der begrenzende Faktor für

IACTs ist jedoch die eingeschränkte Empfindlichkeit, um Änderungen im optischen Fluss zu detektieren. In dieser Arbeit werden die erwarteten Raten der nachweisbaren Ereignisse für VERITAS sowie für mögliche zukünftige IACTs der nächsten Generation vorgestellt. Für VERITAS beträgt die Rate der nachweisbaren Microlensing-Ereignisse im sublunaren Bereich $\sim 10^{-6}$ pro Jahr. Die Zukunftsaussichten für ein 100-mal empfindlicheres Instrument liegen bei $\sim 0,05$ Ereignissen pro Jahr.

Contents

List of Figures

List of Tables

1 Introduction

Over 50 years ago, theoretical work proposed the formation of black holes during the Universe's early stages. Since then, these Primordial Black Holes (PBHs) have been involved in many cosmological explanations. Their range of possible masses is enormous. Their predicted existence has motivated experimental work to search for these objects using various methods. Large parts of the mass range are constrained. Yet, their existence could neither be confirmed nor excluded.
The sublunar mass range $10^{17} \lesssim M/\mathrm{g} \lesssim 10^{23}$ which we show in Figure 1.1 is the broadest unconstrained window [1]. PBHs with masses belong this range lose mass due to the Hawking radiation [2]. At the end of their lifetime, the black holes evaporate. If created at $\sim 10^{15}$ g, the evaporation would take place at the current time. This would create an $\mathcal{O}$(s) long burst of very-high-energy (VHE, $E \gtrsim 50\,\mathrm{GeV}$) γ-rays which might be detectable. This might directly prove the existence of PBHs and the theory of Hawking radiation.
In contrast, for PBHs with masses in the sublunar window and above, the radiation is irrelevant. These are a viable candidate for dark matter, which comprises about 26% of the total energy in the Universe [3]. Adjacent to the sublunar mass range, microlensing observations provide the existing constraints for the smallest masses of non-evaporating PBHs. The sampling speed of classical optical instruments typically defines the low-mass end of these studies [4].
The main objective of this thesis is to constrain the PBH abundance with the Very Energetic Radiation Imaging Telescope Array System (VERITAS). It is an array of four imaging atmospheric Cherenkov telescopes (IACTs). By construction, IACTs are specialized optical detectors capable of detecting the $\sim 10\,\mathrm{ns}$ long Cherenkov flashes from VHE γ-ray induced atmospheric air showers.
This thesis particularly aims to answer the following questions:

1. Can we use a novel transient detection method to identify gamma-ray signals of evaporating PBHs in archival VERITAS data?
2. Are fast optical IACT microlensing observations capable of constraining the sublunar mass window?

These two questions tackle the PBH abundance on both sides of the unconstrained sublunar mass window. We present a data-driven deep learning-based transient detection method for VERITAS to answer the first question. It requires a limited number of assumptions about the instrument characteristics and the signals of PBHs.
We divide this thesis into two parts. First, we present the theory and instrumentation of γ-ray astronomy. We search for light primordial black holes with VERITAS in the second part. The bulk of it is dedicated to deploying the novel deep-learning transient detection method used to search for PBH evaporation. In the following, we present a concise outline of this thesis:

Chapter 2 - introduces primordial black holes. We discuss scenarios for their formation and some of their implications. We conclude this

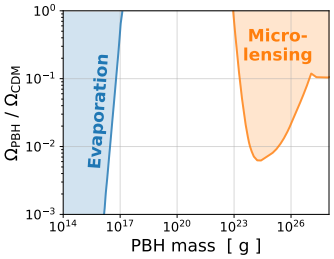


Figure 1.1: The sublunar mass Window of PBHs.

[1]: Carr et al. (2021), 'Constraints on primordial black holes'

[2]: Hawking (1974), 'Black hole explosions?'

[3]: Aghanim et al. (2020), 'Planck 2018 results-VI. Cosmological parameters'

[4]: Griest et al. (2013), 'New Limits on Primordial Black Hole Dark Matter from an Analysis of Kepler Source Microlensing Data'

chapter with a review of existing constraints on the PBH abundance.

Chapter 3 - In this chapter, the concepts of ground-based VHE γ-ray astronomy are presented. We introduce atmospheric air showers and their detection using imaging air Cherenkov telescopes.

Chapter 4 - introduces VERITAS which is one of the currently operating Cherenkov telescope arrays. We highlight the most critical elements of the telescopes and the standard data analysis relevant to this thesis.

Chapter 5 - This chapter sets the stage for the presentation of the novel deep-learning based transient detection method for VERITAS. First, we review the concepts of deep learning which are the core of this approach. Finally, also the strategies for the anomaly detection are introduced. Its core is a recurrent neural network with an encoder-decoder architecture. The network outputs and observations are interpreted statistically using a calibration pipeline.

Chapter 6 - presents the selection and processing of VERITAS data from about nine years of observations which we consider in this thesis. We implement tools to process a bulk of observations automatically. Further, we develop an automatic data quality assessment system. The data processing and selection methods allow the generation of consistent datasets, which are crucial for this thesis.

Chapter 7 - In this chapter, we introduce the expected γ-ray signal from PBH evaporation. We use the introduced parametrization to simulate the detectable signals for VERITAS.

Chapter 8 - presents the newly developed preparation pipeline. It follows the concepts introduced in Chapter 6. The main features are the number of γ-ray counts in time steps and patches of the sky. Also the preparation of the simulated signals is discussed.

Chapter 9 - To give context to the γ-ray counts, auxiliary parameters are required as input to the RNN. These describe the fundamental changes of the data characteristics between different observations. The influences on the observed rates of γ-ray counts are studied to provide relevant auxiliary parameters. With these inputs, we perform the training of the RNN. Finally, the results of the RNN are put into statistical context using the calibration pipeline.

Chapter 10 - In this chapter, we present the results of the search for evaporating PBHs. First, we apply it to the simulated signals to determine the fraction of detected events and calculate the effective search volume. It defines the sensitivity of this analysis. We then scan the selected dataset for PBH bursts and constrain the local rate of PBH evaporation.

Chapter 11 - probes the capabilities of IACTs to constrain the sublunar window with optical microlensing observation. We start with a review of the microlensing formalism, followed by a study of the optical performance of IACTs in this context.

Finally, **Chapter 12** presents the conclusions and outlook of this thesis. We have implemented the data-driven deep learning based transient detection method, a new, independent approach to search for PBH evaporation bursts with VERITAS. We are able to constrain the rate of evaporating PBHs to $< 1.07 \times 10^5\,\mathrm{pc}^{-3}\mathrm{yr}^{-1}$ at the 99% confidence limit. This is in the range of existing limits for the evaporation rate. The prospects for the alternative approach of optical microlensing with IACTs are not encouraging. We expect a rate of detectable events about $10^{-6}\,\mathrm{yr}^{-1}$ with VERITAS.

Theory and Instrumentation for Gamma ray astronomy

Primordial Black Holes 2

The existence of Primordial Black Holes (PBHs) was initially proposed by Zel'dovich in 1967 [5] and Hawking in 1971 [6]. They are a special type of black holes (BHs) which might be created in the early stages of the Universe. They could have been formed at a wide range of BH masses. We review proposed processes of formation in section 2.1.
PBHs are the only objects, theorized or detected, that might emit a detectable Hawking radiation [2]. In section 2.2 we review this emission mechanism. Indeed, PBHs formed at masses smaller than $\sim 10^{15}$ g would have evaporated by today. It also gives rise to a narrow mass window of PBHs that would evaporate nowadays. Among others, evaporating PBHs can provide explanations to the extragalactic and Galactic γ-ray backgrounds [7, 8], the reionization of the pregalactic medium [9] and the annihilation line radiation from the Galactic center [10].
PBHs at much larger masses are unaffected by Hawking radiation. These non-evaporating PBHs are interesting targets on their own. Foremost, they are attractive candidates for the dark matter (DM) that comprises $\sim 26\%$ of the total energy in the universe [3]. We discuss this scenario in more depth in section 2.3. Even if they do not contribute to the DM, they are invoked in numerous cosmological models. For example they might explain the heating of stars within the Galactic disk [11], the seeds of supermassive black holes, the existence of massive compact halo objects [12], and the generation of galaxies [13].
Because of the many implications, PBHs have been focus of intense interest during the past 50 years. Yet, their firm detection is still pending. A large variety of methods aim to constrain the abundance. In this thesis, we search for currently evaporating PBHs and non-evaporating PBHs as DM candidate. We review previous relevant searches and the corresponding limits on the abundance in section 2.3.

[5]: Zel'dovich et al. (1967), 'The Hypothesis of Cores Retarded during Expansion and the Hot Cosmological Model'
[6]: Hawking (1971), 'Gravitationally Collapsed Objects of Very Low Mass'
[2]: Hawking (1974), 'Black hole explosions?'

[7]: Page et al. (1976), 'Gamma rays from primordial black holes.'
[8]: Lehoucq et al. (2009), 'New constraints on the primordial black hole number density from Galactic γ-ray astronomy'
[9]: Belotsky et al. (2015), 'Primordial black holes with mass $10^{16} - 10^{17}$ g and reionization of the Universe'
[10]: Okele et al. (1980), 'Observational consequences of positron production by evaporating black holes'
[3]: Aghanim et al. (2020), 'Planck 2018 results-VI. Cosmological parameters'
[11]: Lacey et al. (1985), 'Massive black holes in galactic halos ?'
[12]: Bean et al. (2002), 'Could supermassive black holes be quintessential primordial black holes?'
[13]: Afshordi et al. (2003), 'Primordial Black Holes as Dark Matter: The Power Spectrum and Evaporation of Early Structures'

2.1 Formation

The Universe after the big bang was in a state of great compression. This high density opens up the chance for various PBH formation scenarios. All formation mechanisms yield a connection between the PBH mass and the horizon mass at formation

$$M_{\mathrm{H}} \sim \frac{c^3 t}{G} \sim 10^{15} \left(\frac{t}{10^{-23}\,\mathrm{s}} \right) \mathrm{s}, \tag{2.1}$$

where t is the time after the big bang, G is the gravitational constant, and c is the speed of light. PBHs can span an enormous range of masses. The earliest time of formation is given by the Planck time 10^{-43} s which yields PBHs at the Planck mass 10^{-5} g. PBHs created 1 s after the big bang, however, have masses of $10^5\,M_\odot$. In the following, we discuss the most frequently studied formation scenarios and the distribution of possible PBH masses.

Collapse of Overdense Regions During the Radiation-Dominated Era

This scenario requires inhomogenities during the radiation-dominated era. Regions with a large overdensity could stop expanding and recollaps [14]. The density contrast of a perturbation with density ρ is described by

$$\delta := \frac{\rho - \bar{\rho}}{\bar{\rho}}, \tag{2.2}$$

where $\bar{\rho}$ is the background energy density. The PBH formation occurs when $\delta_c \leq \delta \leq 1$, where δ_c is the threshold parameter. The value of δ_c is critical as it directly affects the abundance of PBHs. Historically, a first estimate using an analytical argumentation suggested $\delta_c \sim 1/3$. However, more recent numerical and analytical work yield $\delta_c = 0.45$ [15, 16]. The precise value of the thresholds depends on the density profile [17, 18] and the equation of state parameter w [19]. The latter describes the ratio of pressure to energy density. In the radiation dominated era $w = 1/3$. Assuming a Gaussian distribution of the fluctuations with dispersion σ, the fraction of collapsing patches is

$$\beta \sim \mathrm{Erfc}\left(\frac{\delta_c}{\sqrt{2}\sigma}\right), \tag{2.3}$$

where Erfc is the complementary error function.

[14]: Carr et al. (1974), 'Black Holes in the Early Universe'

[15]: Musco et al. (2013), 'Primordial black hole formation in the early universe: critical behaviour and self-similarity'

[16]: Harada et al. (2014), 'Erratum: Threshold of primordial black hole formation [Phys. Rev. D 88, 084051 (2013)]'

[17]: Musco (2019), 'Threshold for primordial black holes: Dependence on the shape of the cosmological perturbations'

[18]: Escrivà et al. (2020), 'Universal threshold for primordial black hole formation'

[19]: Carr (1975), 'The primordial black hole mass spectrum.'

Mass Function of PBHs

The mass of PBHs forming in the radiation-dominated era is [1]

$$M_{\mathrm{PBH}} = \gamma M_{\mathrm{H}} = \gamma \frac{4\pi}{3} \rho R_{\mathrm{H}}^3 = \gamma \frac{c^3 t}{G} \approx 2.03 \times 10^5 \gamma \left(\frac{t}{1\,\mathrm{s}}\right) M_\odot \tag{2.4}$$

where $\gamma < 1$ is a numerical factor whose exact value depends on the details of the gravitation collapse, and R_{H} is the horizon radius. From Equation 2.4, one might expect a monochromatic mass function with $\Delta M_{\mathrm{PBH}} \sim M_{\mathrm{PBH}}$ when PBHs are formed at the same time.
Some scenarios also suggest PBH formation over a prolonged period yielding an extended mass function. But even the formation in single period might have formed extended mass functions. One example might be the collapse from scale-invariant density fluctuations. The spectrum of these perturbations is

$$\mathcal{P}(k) \propto k^{n-1}, \tag{2.5}$$

where k is the wavenumber of fluctuations in Mpc^{-1}, and n is the scalar spectral index. For scale-invariant fluctuation, $n = 1$, the expected mass spectrum is

$$\frac{\mathrm{d}N}{\mathrm{d}M} \propto M^{-\alpha} \quad \text{with} \quad \alpha = \frac{2(1+2w)}{1+w}. \tag{2.6}$$

For a formation during the radiation-dominated era $\alpha = 5/2$.
A much studied effect is the critical collapse that occurs as the density

[1]: Carr et al. (2021), 'Constraints on primordial black holes'

perturbations approach δ_c. This effect leads to an upper cutoff of masses at $\sim M_{\mathrm{H}}$ with a tail to lower masses. For the radiation-dominated era the mass function of PBHs in this scenario is [20, 21]

$$\frac{\mathrm{d}N}{\mathrm{d}M} \propto M^{2.85} \exp\left(-\left(\frac{M}{M_f}\right)^{2.85}\right). \tag{2.7}$$

[20]: Niemeyer et al. (1999), 'Dynamics of primordial black hole formation'
[21]: Carr et al. (2016), 'Constraints on primordial black holes from the Galactic gamma-ray background'

It extends to arbitrary low masses with an exponential cut-off at $M_f \approx M_{\mathrm{H}}$.

Alternative Formation Mechanism

A number of PBH formation mechanism were proposed in addition to the above mentioned scenarios. Among these is the PBH formation via collapse of inhomogenities in a matter-dominated era that might arise from slow reheating after the inflation [22]. Also cosmic strings which self-interact and form cosmic loops are proposed. There is a chance of a loop to be completely contained in its Schwarzschild radius [23]. The probability for this formation is constant at every epoch yielding an extended mass function. Also see [24] for a more detailed review of the possible formation mechanism.

[22]: Khlopov et al. (1985), 'Gravitational instability of scalar fields and formation of primordial black holes'
[23]: Hawking (1989), 'Black holes from cosmic strings'
[24]: Carr et al. (2020), 'Primordial Black Holes as Dark Matter: Recent Developments'

2.2 Hawking Radiation

Historically, it was believed that due to their immense gravitational field, black holes could not emit any radiation. The fact that PBHs might be very small triggered an investigation of the quantum properties of black holes. Ultimately, this study by Hawking led to the understanding that black holes might radiate fundamental particles with spin s at a rate of [25]

$$\frac{\mathrm{d}^2N}{\mathrm{d}E\mathrm{d}t} = \frac{\Gamma/2\pi\hbar}{e^x - (-1)^{2s}} n_{\mathrm{dof}}. \tag{2.8}$$

[25]: Heckler (1997), 'Calculation of the Emergent Spectrum and Observation of Primordial Black Holes'

Here $\hbar$ is the reduced Planck constant, n_{dof} are the degrees of freedom of the particle. Γ is the absorption coefficient which describes the fraction of particles absorbed by the BH [26]. The black hole angular momentum and electric charge is radiated away faster than its mass. Accordingly, a significant Hawking radiation is expected only for non-rotating, uncharged black holes. For these, the dimensionless parameter x is

[26]: Hawking (1975), 'Particle creation by black holes'

$$x \equiv \frac{E}{k_{\mathrm{B}} T_{\mathrm{BH}}}, \tag{2.9}$$

where E is the energy of the emitted particle, k_{B} is the Boltzman constant, and the temperature T_{BH} of the black hole only depends on its mass M_{BH} [27]

$$k_{\mathrm{B}} T_{\mathrm{BH}} = \frac{\hbar c^3}{8\pi G M_{\mathrm{BH}}} \approx 1.058 \left(\frac{10^{13}\,\mathrm{g}}{M_{\mathrm{BH}}}\right) \mathrm{GeV}. \tag{2.10}$$

[27]: Ukwatta et al. (2016), 'Primordial Black Holes: Observational characteristics of the final evaporation'

For an emitted particle of rest mass m, the absorption coefficient for $E \gg mc^2$ is in the form of

$$\Gamma(M_{\mathrm{BH}}, E, s) = 27 \left(\frac{x}{8\pi}\right)^2 \gamma_s(x). \tag{2.11}$$

For large x the dimensionless parameter $\gamma_s(x) \to 1$ [28]. The emitted particles result in a decreasing BH mass,

$$\frac{\mathrm{d}M_{\mathrm{BH}}}{\mathrm{d}t} = -\frac{1}{c^2} \sum_i \int_0^\infty \frac{\mathrm{d}^2 N_i}{\mathrm{d}E\mathrm{d}t} E \mathrm{d}E \equiv -\frac{\alpha(M_{\mathrm{BH}})}{M_{\mathrm{BH}}^2}, \tag{2.12}$$

where the sum over i includes all fundamental particle species. The function $\alpha(M_{\mathrm{BH}})$ incorporates all emitted particles and their degrees of freedom [29]. As the mass of the BH decreases and its temperature increases, the list of emitted particles increase. Considering all confirmed particles of the Standard Model, the evolution of α is shown in Figure 2.1.[1]

[28]: MacGibbon et al. (1990), 'Quark- and gluon-jet emission from primordial black holes: The instantaneous spectra'

[29]: Page (1976), 'Particle emission rates from a black hole: Massless particles from an uncharged, nonrotating hole'

1: If they exist, also unknown massive particles beyond the Standard Model might be emitted. These would increase the total value of $\alpha(M_{\mathrm{BH}})$.

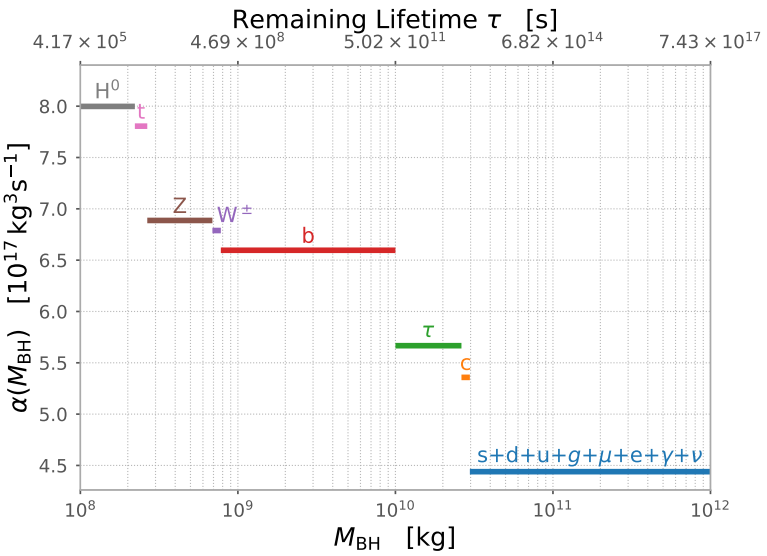


Figure 2.1: The function $\alpha(M_{\mathrm{BH}})$ describes the emission of all fundamental particles. As the BH mass decreases, its temperature rises and thus more particles are emitted. The remaining lifetime τ of the BH is calculated using Equation 2.13. For this, the average value of $\alpha(M_{\mathrm{BH}})$ are calculated in the range $m = M_{\mathrm{BH}}$ to $m = 0\,\mathrm{kg}$. Adapted from [27].

Integrating Equation 2.12 yields the time τ until the BH with mass M_{BH} lost its complete mass

$$\tau = \frac{M_{\mathrm{PBH}}^3}{3\overline{\alpha}}, \tag{2.13}$$

where $\overline{\alpha}$ is the average value of $\alpha(M_{\mathrm{PBH}})$ over the lifetime of the BH. We show values of the remaining lifetime τ in the top axis of Figure 2.1. From this calculation, we find that BHs with $\sim 8.4 \times 10^{14}\,\mathrm{g}$ would evaporate after $\sim 13.8 \times 10^9\,\mathrm{yr}$, corresponding to the age of the Universe [30]. At the final stages ($M_{\mathrm{BH}} \lesssim 2.2 \times 10^8\,\mathrm{kg}$, $\tau \lesssim 0.144\,\mathrm{yr} = 52.6\,\mathrm{d}$) of the BH evaporation $\alpha(M_{\mathrm{PBH}})$ remains constant at

$$\alpha_f = 8 \times 10^{17}\,\mathrm{kg^3 s^{-1}}. \tag{2.14}$$

Using this in Equation 2.12 shows that the emission rate accelerates with M_{BH}^{-2}. Eventually, the BH evaporates completely in a burst of fundamental

[30]: Akrami et al. (2020), 'Planck 2018 results. I. Overview and the cosmological legacy of Planck'

particles.
Equation 2.8 describes the direct emission of fundamental Standard Model particles which are leptons, quarks, and gauge bosons. It also includes a component of directly emitted γ-rays which are important for the highest energy of the photon spectrum. The quarks and gluons undergo fragmentation and hadronization to intermediate states. Eventually, these decay into photons, neutrinos, protons, antiprotons, electrons, and positrons. The most dominant such decay for the photon production is the neutral pion decay $\pi^0 \rightarrow 2\gamma$. A third component comes from the decay of other fundamental particles such as the τ-lepton. However, this component is small compared to the other contributions. We introduce a description of the γ-ray spectrum during the evaporation of a BH in Chapter 7.

2.3 Searches for Primordial Black Holes

In this section we review the current status of searches for PBHs. These works commonly assume that PBHs follow a monochromatic mass function. We discuss the two complementary scenarios of non-evaporating and evaporating PBHs separately. We refer to the standard Λ cold dark matter (ΛCDM) model with the age of the universe $t_0 = 13.8 \times 10^9$ yr and the Hubble constant $H_0 \equiv h \times 100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, where $h = (0.6736 \pm 0.0054)$ [30].

[30]: Akrami et al. (2020), 'Planck 2018 results. I. Overview and the cosmological legacy of Planck'

Constraints on Non-Evaporated Primordial Black Holes

As mentioned above, non-evaporating PBHs are often discussed in the context of dark matter. The existence of DM is well established. Its influences are observed experimentally by various methods. Among them are the rotation velocities of stars in spiral galaxies [31], X-ray emission coming from hot gas in elliptical galaxies [32], and gravitational lensing [33]. The relative densities Ω of different types of matters are typically measured in ratio to the critical density

$$\rho_c = \frac{3H_0}{8\pi G} \approx 0.85 \times 10^{-29}\,\mathrm{g\,cm^{-3}}. \tag{2.15}$$

[31]: Kinney et al. (2000), 'Evidence for universal structure in galactic halos'

[32]: de Paolis et al. (1995), 'Dark Matter in X-Ray-emitting Elliptical Galaxies'

[33]: Clowe et al. (2004), 'Weak-Lensing Mass Reconstruction of the Interacting Cluster 1E 0657-558: Direct Evidence for the Existence of Dark Matter'

Measurements of the cosmic microwave background yield values for the baryonic matter $\Omega_{\mathrm{BM}} = 0.049 \pm 0.001$, for dark matter $\Omega_{\mathrm{CDM}} = 0.264 \pm 0.005$ and for the dark energy $\Omega_{\mathrm{DE}} = 0.685 \pm 0.007$ [3].
While the confirmation of any proposed DM candidate is still pending to date, some of its properties are quite clear. They are non-baryonic objects that possess nonzero masses as they form the bulk of galaxies. DM might be either microscopic or macroscopic. As they exist in the current time, their lifetimes have to be larger then the age of the Universe. They do not interact via the electromagnetic force and thus should be electrically neutral. The state of the art scenario for the formation of cosmic structures such as galaxies is due to a gravitational collapse of DM. This can be best explained by non-relativistic cold DM (CDM) [34]. As discussed above, the most probable PBH formation scenarios are during the radiation-dominated era. Thus, the constraint of the baryonic mass ratio from the big bang nucleosynthesis does not apply to them

[3]: Aghanim et al. (2020), 'Planck 2018 results-VI. Cosmological parameters'

[34]: Peebles (1982), 'Large-scale background temperature and mass fluctuations due to scale-invariant primeval perturbations'

[35]: Cyburt et al. (2003), 'Primordial nucleosynthesis in light of WMAP'

[35]. Therefore, PBHs should be considered non-baryonic. They fulfill all of the above criteria making them valid CDM candidates. An upper limit of the total PBH density is $\Omega_{\rm PBH} \leq \Omega_{\rm CDM}$. It is practical, to express constraints on the density fractional of non-evaporating PBHs in terms of the total DM

$$f(M) \equiv \frac{\Omega_{\rm PBH}(M_{\rm PBH})}{\Omega_{\rm CDM}}. \tag{2.16}$$

[1]: Carr et al. (2021), 'Constraints on primordial black holes'

Figure 2.2 shows current limits on the PBH abundance as a function of $M_{\rm PBH}$. A large number of methods are able to probe different parts of the mass range spanning over 40 orders of magnitudes. For a more detailed discussion of these constraints see [1]. Only three mass windows could have a significant fraction of PBHs: the asteroidal to sublunar range $10^{17} \lesssim M/\mathrm{g} \lesssim 10^{23}$, the intermediate range $10^{1} \lesssim M/M_\odot \lesssim 10^{2}$, and the stupendous mass range $M/M_\odot \gtrsim 10^{11}$. However, the last can not contribute to the DM halo of galaxies, as it it heavier than the total galaxies. In recent years, gravitational wave detections of BH mergers in the range $10-50\,M_\odot$ have drawn much attention to the intermediate mass range. Black holes were initially not expected in this range making PBHs an attractive explanation. However, it is also argued, that the sublunar mass range is more plausible for the PBH formation.

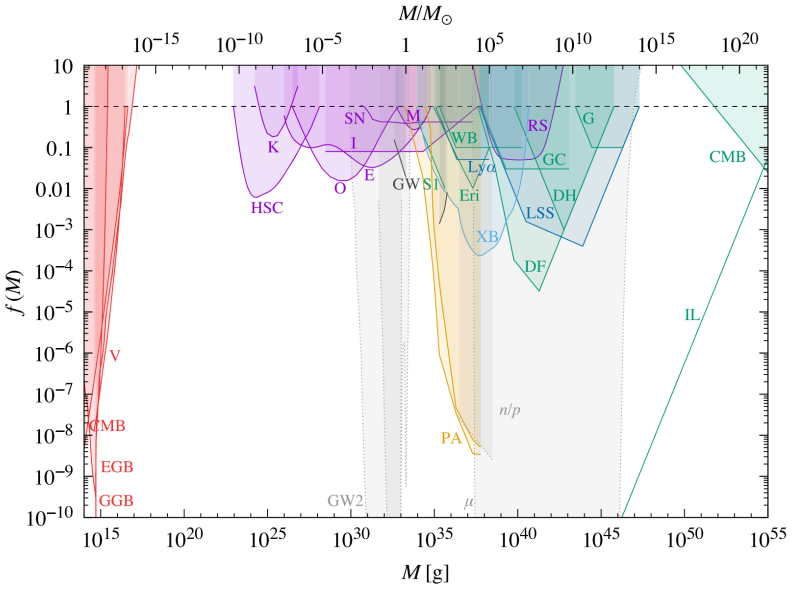


Figure 2.2: Constraints on non-evaporated PBHs in $f(M)$. The colors group together different categories of studies. *Red*: evaporation, *magenta*: lensing, *green*: dynamical effects, *black*: gravitational waves, *light blue*: accretion, *orange*: cosmic microwave background distortions, *dark blue*: large-scale structure, and *grey*: background effects. This graph only includes commonly accepted constraints. The labels within the graph refer to different studies conducted for the PBH abundance. For a more detailed description of these see the original graph in [1] and references therein.

Constraints on Evaporated Primordial Black Holes

[1]: Carr et al. (2021), 'Constraints on primordial black holes'

At the time of formation t_i the fraction of the Universe's mass attributable to PBHs is described by [1]

$$\beta(M) = \frac{M n_{\rm PBH}(t_i)}{\rho(t_i)} \tag{2.17}$$

$$\approx 7.06 \times 10^{-18} \gamma^{-1/2} \left(\frac{h}{0.67}\right)^2 \left(\frac{g_{*\mathrm{i}}}{106.75}\right)^{1/4} \Omega_{\rm PBH}(M) \left(\frac{M}{10^{15}}\right)^{1/2}, \tag{2.18}$$

where $n_{\mathrm{PBH}}(t_i)$ is the number density, $\rho(t_i)$ is the density, $g_{*\mathrm{i}}$ is the relativistic degrees of freedom normalised to its value at 10^{-5} s, h is defined by the Hubble constant above, and γ was first introduced in Equation 2.4. The dependency on Ω_{PBH} implies that limits on the currently existing non-evaporated PBHs directly also constrain β. The quantity β in Equation 2.18 is defined with the combination of $\gamma^{-1/2} g_{*\mathrm{i}}^{1/4} h^2$. The values of $g_{*\mathrm{i}}$ and h are known very precisely, however γ is rather uncertain. It is convenient to define a new parameter to express constraints on evaporated PBHs

$$\beta'(M) = \gamma^{1/2} \left(\frac{h}{0.67} \right)^{-2} \left(\frac{g_{*\mathrm{i}}}{106.75} \right)^{-1/4} \beta(M). \tag{2.19}$$

We show constraints on β' in Figure 2.3. Almost over the complete mass range, the strongest constraints are inferred from observations of the big bang nucleosynthesis (BBN) [36], cosmic microwave background (CMB) anisotropies [37], and extragalactic γ-ray background (EGB) [36]. These limits are based on the assumption of PBHs evaporating due to the Hawking radiation. Without Hawking radiation only the condition $\Omega_{\mathrm{PBH}} \leq \Omega_{\mathrm{CDM}}$, discussed for non-evaporating PBHs above, gives a constraint for these masses. The dashed black line shows this 3σ upper limit using the recent value of Ω_{CDM} mentioned above.

[36]: Carr et al. (2010), 'New cosmological constraints on primordial black holes'

[37]: Acharya et al. (2020), 'CMB and BBN constraints on evaporating primordial black holes revisited'

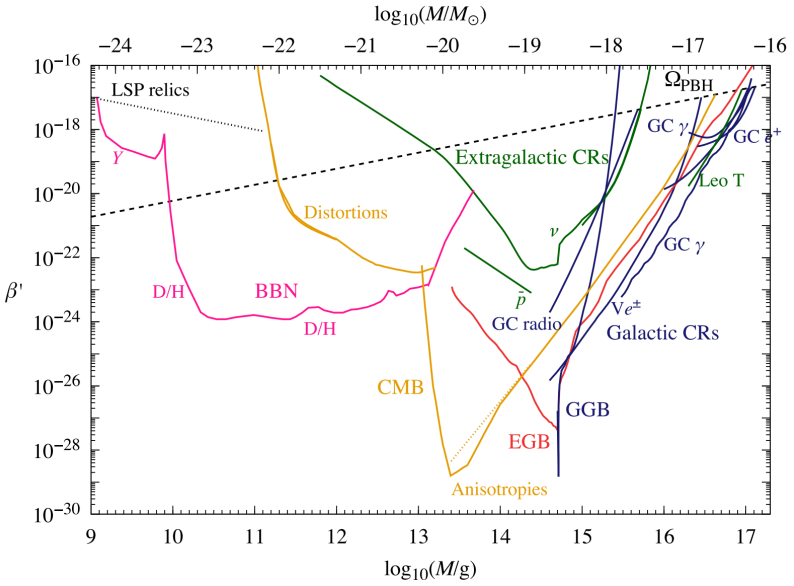


Figure 2.3: Constraints on evaporating PBHs in β'. The colors group together different categories of studies. These works assume the existence of Hawking radiation. If it would not be applied, only the dashed constraints, coming from $\Omega_{\mathrm{PBH}} < \Omega_{\mathrm{CDM}}$, are valid. For a more detailed description of the studies and labels see the original graph in [1] and references therein.

The introduced formalism of Hawking radiation in section 2.2 gives rise to γ-ray signals that might be detectable during the final moments of the evaporation. We introduce a description of these signals in section 7.1. Searches for this emission might not only prove the existence of PBHs but also directly confirm the theory of Hawking radiation. As mentioned above, the probed masses are $\sim 8.4 \times 10^{14}$ g. These works constrain the rate of evaporations per effective search volume V_{eff}. The probed volume are typically small with maximum distances of the PBHs of $r_{\mathrm{max}} = \mathcal{O}(1\mathrm{pc})$. On these scales, the density of PBHs is assumed to be constant. The evaporation rate in the local Galaxy is given by

$$r_{\mathrm{burst}} = \frac{n_{\mathrm{burst}}}{V_{\mathrm{eff}} T_{\mathrm{eff}}}, \tag{2.20}$$

where n_{burst} are the number of detected evaporations and T_{eff} is the effective observation time. These studies search for the integrated γ-ray signals from the evaporation bursts in the final ΔT of the lifetime of the PBH. The integration time ΔT is an instrument specific parameter which largely influence the signal-to-noise ratio. The optimal parameter, yielding the best constraints , is typically determined by scanning various parameters. Most relevant, however, are the best constraints for each experiment.

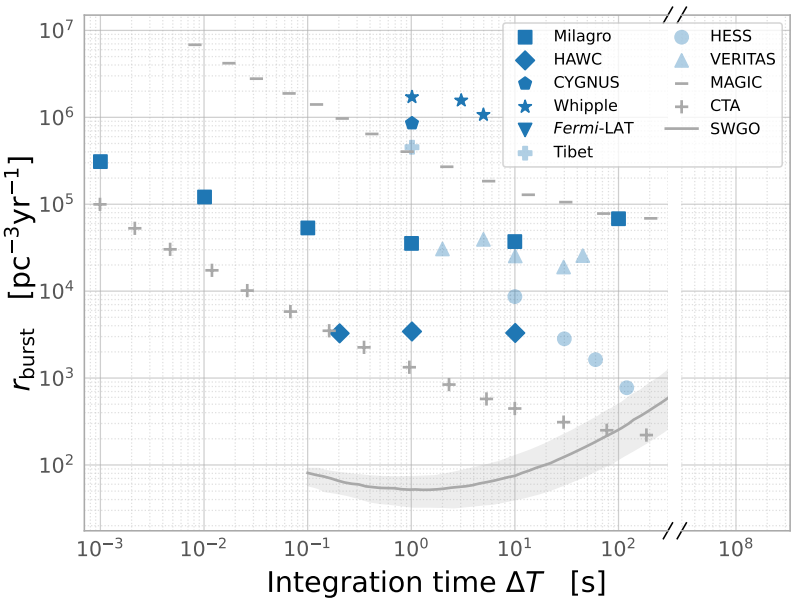


Figure 2.4: Constraints on burst rate of PBHs. Dark blue data points are peer-reviewed published results while light blue shows limits published in proceedings. These come from Milagro [38], HAWC [39], CYGNUS [40], Whipple [41], *Fermi*-LAT [42], Tibet air shower array [43], H.E.S.S. [44], and VERITAS [45]. Published prospects are shown in grey for MAGIC, CTA [46], and 10 years of observations with SWGO [47]. Adjusted from [48].

A current status of searches is shown Figure 2.4. We review the different detector types for γ-ray astronomy below in Chapter 3. The blue data points show existing constraints, where the dark blue highlights peer-reviewed publications. The grey data points illustrate prospects for future studies. Detectors such as Milagro, HAWC, *Fermi*-LAT, Tibet air shower array and SWGO benefit from their large fields of view and dense duty cycle. However, r_{max} is small for in these works (e.g. $r_{\text{max}} = 0.02\,\text{pc}$ for *Fermi*-LAT [42]). The most constraining wide FoV upper limits are obtained with 3 yrs of data from HAWC with $\Delta T = 10\,\text{s}$ at $\sim 3.4 \times 10^3\,\text{pc}^{-3}\text{yr}^{-1}$ [39]. Prospects for 10 yrs of observations with the Southern Wide field of view Gamma-ray Observatory (SWGO) yield an improvement of more than one magnitude [47]. Imaging Air Cherenkov Telescopes such as Whipple, H.E.S.S., VERITAS, MAGIC and CTA on the other hand have significantly smaller FoVs. However, the depth r_{max} is in the order of pc. The best IACT upper limits are currently set with 4924 h of H.E.S.S. data to $< 527\,\text{pc}^{-3}\text{yr}^{-1}$ [44]. As ΔT is not a parameter of interest, it is also the overall strongest constrain on the evaporation rate of PBHs.

[42]: Ackermann et al. (2018), 'Search for Gamma-Ray Emission from Local Primordial Black Holes with the *Fermi* Large Area Telescope'

[39]: Albert et al. (2020), 'Constraining the local burst rate density of primordial black holes with HAWC'

[47]: López-Coto et al. (2021), 'Prospects for the observation of Primordial Black Hole evaporation with the Southern Wide field of view Gamma-ray Observatory'

[44]: Tavernier et al. (2021), 'Limits on primordial black hole evaporation from H.E.S.S. observations.'

Very-High-Energy γ-ray Astronomy

3

In the previous chapter we discussed how PBHs could directly and indirectly emit photons during their final moments of evaporation. This burst of photons might be detectable at the high-energetic edge of the electromagnetic spectrum. The most stringent constraints on the evaporation rate are obtained with γ-ray observatories.

Space bound instruments can directly detect γ-rays hitting the detector. The effective collection areas of the detectors are small ($< 1\,\mathrm{m}^3$). Thus, they are only sensitive up to a few 100 GeV.

Due to the opacity of the atmosphere, γ-rays can not directly be measured on the earth's surface. They interact in the atmosphere inducing atmospheric air showers consisting of particles propagating towards the earth's surface. The products of these showers can be detected which allows indirect conclusions about the initial γ-ray. We review the evolution of atmospheric air showers in section 3.1.

Only very-high-energy (VHE, $E \gtrsim 50\,\mathrm{GeV}$) γ-rays induce sufficiently large showers for a detection at earth with the current generation of instruments. The main advantage of the ground-based γ-ray detection are the effective photon collection areas up to about $10^5\,\mathrm{m}^2$. The size of the particle shower scales with the energy of the primary particle the energy range allowing the detection of VHE particles. This makes them very powerful instruments to search for PBH evaporation busts. As shown in Figure 2.4, the most stringent existing limits are coming from this family of experiments.

One type of ground-based γ-ray detectors are the wide field-of-view detectors. They are typically based on the water Cherenkov technique or scintillation counters located at high altitudes of $\sim 4000\,\mathrm{m}$ a.s.l.. Existing experiments are sensitive to events above $\gtrsim 10\,\mathrm{TeV}$. Further they provide an enormous field of view with radius about 45°. The robust detectors can also operate under moonlight enabling a dense duty cycle. However, the reconstruction of the properties of the initial event from the recorded information is challenging. The currently and previously operating experiments based on water Cherenkov technique are HAWC [49], MILAGRO [50]. The Tibet air shower array [51] is based on silicon scintillation counters. The SWGO [52] will be a next-generation experiment based on the water Cherenkov technique.

The alternative approach of ground-based γ-ray astronomy is the imaging air Cherenkov method. Imaging atmospheric Cherenkov telescopes (IACTs) measure the Cherenkov light which is emitted when relativistic charged particles propagate through the atmosphere. This work studies the detection of PBHs with the VERITAS [53] experiment that is based on this technique. We discuss this approach in more depth in section 3.2. Other currently operating experiments using this technique are HESS [54] and MAGIC [55, 56]. These will be succeeded by CTA [57] in the future. Recently, also hybrid systems such as LHASSO [58] were developed. These can combine the strengths of the individual methods.

[49]: Abeysekara et al. (2012), 'On the sensitivity of the HAWC observatory to gamma-ray bursts'

[50]: Yodh (1997), 'The MILAGRO gamma ray observatory'

[51]: Amenomori et al. (2008), 'Tibet air shower array: results and future plan'

[52]: Hinton et al. (2022), 'The Southern Wide-field Gamma-ray Observatory: Status and Prospects'

[53]: Holder et al. (2006), 'The first VERITAS telescope'

[54]: Aharonian et al. (2006), 'Observations of the Crab nebula with HESS'

[55]: Aleksić et al. (2016), 'The major upgrade of the MAGIC telescopes, Part I: The hardware improvements and the commissioning of the system'

[56]: Aleksić et al. (2016), 'The major upgrade of the MAGIC telescopes, Part II: A performance study using observations of the Crab Nebula'

[57]: Actis et al. (2011), 'Design concepts for the Cherenkov Telescope Array CTA: an advanced facility for ground-based high-energy gamma-ray astronomy'

[58]: Cao et al. (2021), 'Ultrahigh-energy photons up to 1.4 petaelectronvolts from 12 γ-ray Galactic sources'

3.1 Atmospheric Air Showers

IACTs are designed for detecting the Cherenkov light emitted from VHE air showers. In subsection 3.1.1, we discuss the mechanism of Cherenkov radiation. Eventually, the detected Cherenkov shower images allow conclusions about the properties of the inducing event. Beyond the energy and origin, especially the particle type is crucial. The target are γ-rays which are obscured by a large component of cosmic-ray induced air showers. In the subsection 3.1.2 and subsection 3.1.3 we discuss the shower development for both initial particle types.

3.1.1 Cherenkov Radiation

A charged particle passing through a dielectric medium excites the surrounding molecules. As they return to the ground state, they re-emit this energy in form of photons. These photons move at the speed of light in the medium $v_c = c/n$, where c is the speed of light in the vacuum and $n > 1$ is the refraction index of the medium. In case particles moving at speeds $v_p < v_c$, the emitted radiation is spherically symmetric and thus cancels out. However, high energetic particles can move faster than the speed of light, $v_p > v_c$. In this case, the emission gets asymmetric and interferes constructively leading to a cone of light. The opening angle of this cone is given by $\theta = \arccos(1/\beta n)$, where $\beta = v_p/c$ is the speed of the particle in units of c. The value of n scales with the density of the atmosphere. Thus, the opening angle increases towards the earth surface up to $\theta \approx 1.3°$ at sea level.
The frequency spectrum of the Cherenkov radiation is described by the Frank-Tamm formula [59]

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

where x is the unit length traveled by the particle, ω is the frequency of the light, q is the charge of the particle, $\mu(\omega)$ is the permeability, and $n(\omega)$ is the refraction index. The photon density of the emitted Cherenkov radiation in the atmosphere peaks in the near-ultra violet (near-UV) at wavelengths around 350 nm.

[59]: Tamm et al. (1937), 'Coherent radiation of fast electrons in a medium'

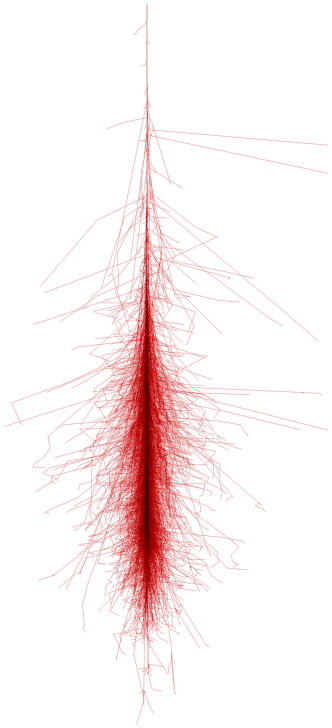


Figure 3.1: Air shower induced by a 100 GeV photon simulated in Corsika projected in the xz-axis. This shower mainly consists of the photons, electrons and positrons (red tracks). [60].

3.1.2 γ-ray Induced Air Showers

In the atmosphere, VHE γ-rays decay via pair production in the presence of the electromagnetic (EM) field of the nuclei. This produces a pair of electron and positron, $\gamma \rightarrow e^+ + e^-$. The relativistic $e^\pm$ pairs subsequently produce photons through Bremsstrahlung. If the energy of these photons is sufficient, they themselves can undergo pair production. This alternating process of pair production and Bremsstrahlung yields a cascade of EM particles. The $e^\pm$ also undergo Coulomb scattering with the molecules in the atmosphere. It causes a broadening of the overall particle cascade. While propagating through the atmosphere, the $e^\pm$ lose energy due the Bremsstrahlung and ionization of the medium. Below the threshold energy $E_{\text{thresh}} \approx 84$ MeV, the ionization becomes the dominating loss. Once this point is reached, the particle production falls

off.
We show an example of a γ-ray induced air shower in Figure 3.1. The primary energy is 100 GeV. Each trajectory of a particle is visualized in the xz-projection. As discussed above, the EM showers consist of photons, electrons and positions. These particle tracks are visualized by the red lines.

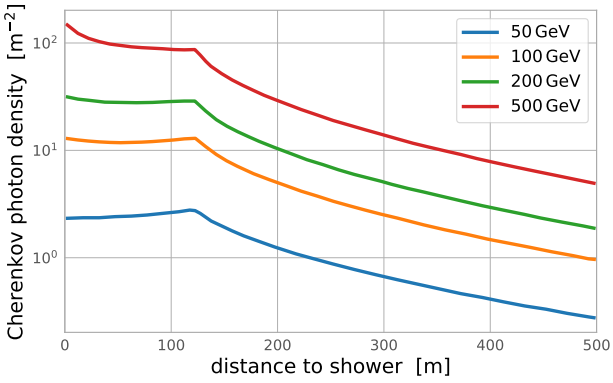


Figure 3.2: Density of Cherenkov photons at an altitude of 1800 m a.s.l. for vertical γ-ray induced showers for different primary energies. Adapted from [61].

The ionization caused by relativistic $e^{\pm}$ in the air cause the characteristic cone of Cherenkov radiation. We show the density of Cherenkov photons at an altitude of 1800 m a.s.l. in Figure 3.2 as a function of the distance to the projected impact point of shower axis on the surface. The four graphs denote four different energies of the initial γ-ray. Up to about 130 m, the density is roughly constant independent of the primary energy. Further away, it is dropping rapidly. The detection of these signals using IACTs is described in section 3.2.

3.1.3 Cosmic-ray Induced Air Showers

As mentioned above, the cosmic-ray induced showers are a huge component of background for ground-based γ-ray observatories. The cosmic-rays themselves are dominated by hadrons, predominantly protons, and a smaller population of cosmic electrons. These electrons also induce EM air showers which are very similar in their evolution to the γ-ray showers. Due to energy losses from inverse Compton scattering and synchrotron radiation, the electron flux drops steeply at around 1 TeV.
Contrary to EM cascades, for which the possible interactions are limited, the hadrons interact via the strong force with the particles of the atmosphere. This produces secondary hadrons as well as neutral and charged pions. The mean lifetime of the neutral π^0 is $\mathcal{O}(0.1\,\text{fs})$ [62], causing a rapid decay $\pi^0 \rightarrow e^+ + e^-$. These $e^{\pm}$ induce a EM component of the shower. Due to the longer lifetime, relativistic $\pi^{\pm}$ interact in the atmosphere producing further charged particles, mesons, muons, or neutrinos before decaying. This leads to a hadronic component of the air shower.
Due to the large amount of possible interactions, the shower development is less regular compared to EM showers. The produced pions typically have larger transverse momenta. Thus the general spread of the hadronic showers is expected to be broader. These effects are also represented in the emitted Cherenkov radiation from hadronic showers: The typical signal on the ground is wider and less regular. However, the EM component of the showers might also cause an appearance very similar to γ-ray

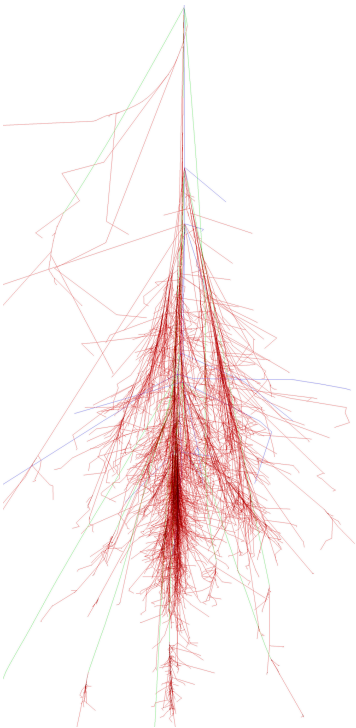


Figure 3.3: Air shower induced by a 100 GeV proton simulated in Corsika projected in the xz-axis. Besides the EM component (red tracks), also muons (green) and hadrons (blue) are present. [60].

[62]: Group et al. (2020), 'Review of Particle Physics'

induced showers.
Figure 3.3 shows an example of a simulated proton-induced air shower with $E_{\text{init}} = 1\,\text{TeV}$. Contrary to the EM shower in Figure 3.1, also hadron and muon components are present that are illustrated by the blue and green tracks. These cause a larger lateral spread of the particles and more irregularity.

3.2 Imaging Air Cherenkov Telescopes

[63]: (1953), 'Light Pulses from the Night Sky associated with Cosmic Rays'

The first detection of Cherenkov light produced by cosmic rays [63] turned out to be the start of a new field, the Cherenkov astronomy. About 70 years later, it has matured and provided many exciting insights to the VHE universe. IACTs are designed to detect the near-UV flashes of Cherenkov light from air showers. Effectively, they are optical telescopes covering the optical blue and UV light. However, the requirements significantly differ from standard optical telescopes:

[64]: Holder (2015), 'Atmospheric Cherenkov Gamma-Ray Telescopes'

reflectors Very large mirror areas are needed to detect a large amount of Cherenkov photons, which directly influences the lower energy threshold of detectable events. Traditional optical telescopes have strict requirements on the optical point-spread function (oPSF). For IACTs however, an oPSF in the order of arcminutes is sufficient, which can be achieved using a grid of individual mirror facets [64].

field-of-view Cherenkov signals from extensive air showers have an extension of up to few degrees. The field-of-view of IACTs needs to be at least of the same order to record these images. This requirement is fulfilled using small focal lengths typically around $f/D \sim 1.0$.

time resolution The Cherenkov signals from individual air showers have a duration of about 10 ns. IACTs require $\mathcal{O}$(ns) sampling speed to record the evolution of these showers.

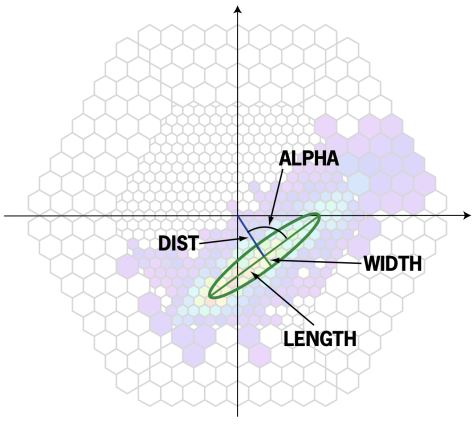


Figure 3.4: Image parameters of γ-ray images. [65].

With their $\mathcal{O}$(ns) sampling, IACTs record the evolution of the intensity over multiple samples. Integrating these traces yields the total charge per pixel. To identify the pixels with relevant Cherenkov signals, a image cleaning procedure is applied.[1] For γ-rays, the signal shape in the camera is approximately elliptical. This motivates their characterization using the Hillas parameters [66]. They consist of

1: See more details on the VERITAS data analysis in section 4.2.

[66]: Hillas (1985), 'Cerenkov Light Images of EAS Produced by Primary Gamma Rays and by Nuclei'

- *length* l and *width* w of the ellipse

- *rotation angle* α of the ellipse
- angular *distance* and *azimuth* angle describing the location of the image's center of gravity relative to the camera center in polar coordinates
- integrated charge of all pixels aver image cleaning *size s*

These Hillas parameters are visualized in Figure 3.4. The projected major-axis of the images points in the direction of the shower origin. An array of IACTs can record images of the same shower from different angles. This triangulation allows to estimate the primary particle's origin and core position, which is the projected impact point on the ground. The impact distance R described the separation of each telescope to the core position. We show an illustration of this stereoscopic approach for IACTs in Figure 3.5.

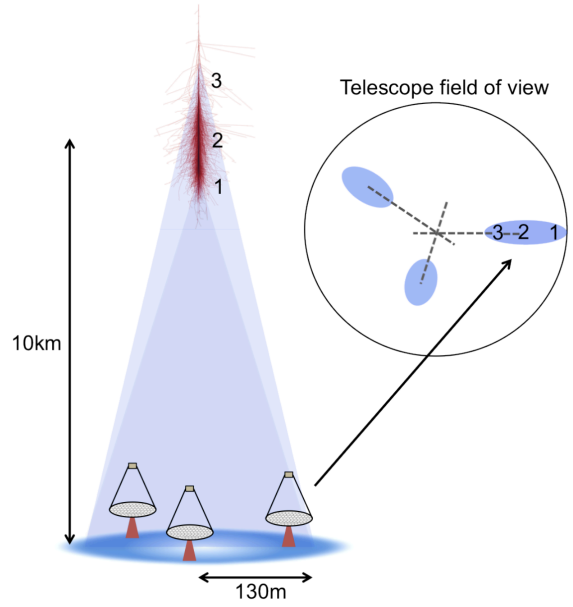


Figure 3.5: Illustration of the stereoscopic IACT technique. [64].

The stereoscopic parameters describe the joint set of images. Especially, these are the *mean reduced scaled width* ($mscw$)

$$mscw = \frac{1}{N_{\text{images}}} \sum_{i=1}^{N_{\text{images}}} \frac{w_i - w_{\text{MC}}(R,s)}{\sigma_{w,\text{MC}}(R,s)}, \tag{3.2}$$

and *mean reduced scaled length* ($mscl$)

$$mscl = \frac{1}{N_{\text{images}}} \sum_{i=1}^{N_{\text{images}}} \frac{l_i - l_{\text{MC}}(R,s)}{\sigma_{l,\text{MC}}(R,s)}, \tag{3.3}$$

where w_i and l_i are the widths and lengths of the images. These are compared to prediction from Monte Carlo (MC) simulated γ-ray showers in function of R and s. The differences are normalized to the expected spread derived from simulations $\sigma_{w,\text{MC}}$ and $\sigma_{l,\text{MC}}$, respectively. These stereoscopic parameters provide crucial information that helps to inferring the sought for primary particle attributes. In particular, $mscw$ and $mscl$ are powerful measurements to discriminate between hadronic and EM showers. We describe these reconstructions for VERITAS in subsection 4.2.3 and subsection 4.2.4.

In general, the performance of IACT arrays is improved with the total number of telescopes. This is due to the additional information in the event reconstruction. However, the limiting factor for the effective collection area is not defined by the total mirror size but the shower characteristic instead. The showers can be recorded from anywhere within the Cherenkov light cone. As shown in Figure 3.2, the cone area at ground level is about 130 m.

4 VERITAS

In Chapter 3 we discussed the concepts of ground-based γ-ray astronomy. This thesis employs data from the Very Energetic Radiation Imaging Telescope Array System (VERITAS). It is one of the currently operating IACTs. We describe its specifics and performance in section 4.1. In section 4.2, we review the steps of the standard reconstruction, which are relevant for this thesis.

4.1 The VERITAS Telescopes

The VERITAS instrument is an array for four Imaging Atmospheric Cherenkov Telescopes. It is located at the *Fred Lawrence Whipple Observatory* (31°40′30″ N, 110°57′07″ W) near Amado, Arizona, at an altitude of 1270 m.a.s.l. The construction of the first telescope started in 2005. The inauguration of the entire array took place in 2007. Figure 4.1 shows a photograph of the observation site with all four telescopes.

Figure 4.1: Image of the basecamp of the Fred Lawrence Whipple Observatory with the VERITAS telescopes. Photograph taken from [67].

4.1.1 Upgrades

Since 2007, VERITAS is observing the very-high energy ($E_\gamma > 100\,\mathrm{GeV}$) γ-ray sky [68]. The VERITAS collaboration steadily improved the sensitivity of the instrument during its lifetime. Development took place on the analysis techniques (e.g. see [69]) as well as on the hardware [70]. VERITAS operated during three major epochs, separated by two significant upgrades [71]. In its initial configuration, VERITAS operated from 2007 to 2009. We refer to this epoch as *V4*.
In the summer of 2009, the relocation of telescope T1 started the epoch *V5*. The T1 telescope in its new location is the foreground telescope in Figure 4.1. The updated array geometry resulted in a notable improvement of the significance.
The second major upgrade took place in the summer of 2012. After this upgrade, the new photomultiplier tube cameras provided higher

[68]: Holder et al. (2008), 'Status of the VERITAS Observatory'

[69]: Krause et al. (2017), 'Improved γ/hadron separation for the detection of faint γ-ray sources using boosted decision trees'

[70]: McCann et al. (2010), 'A new mirror alignment system for the VERITAS telescopes'

[71]: Kieda (2011), 'Status of the VERITAS Upgrade'

quantum efficiency. Furthermore, a new L2-trigger was implemented. We discuss details of the trigger system in subsection 4.1.3. Since 2012, VERITAS is operating in this *V6* epoch.
In this thesis, we employ data from the V6 epoch. It corresponds to the period with the highest overall sensitivity. Below, we review the instrument specifications and performance for this epoch.

4.1.2 Optics and Camera

The four telescopes of VERITAS employ the same configuration. Each telescope has a reflecting area with 12 m diameter in the Davies-Cotton design [72]. It consists of 345 individually adjustable hexagonal mirrors and has a focal length of 12 m. For the elevation range of typical observations ($> 40°$), measurements of the *optical point spread function* (OPSF) typically yield around $\sim 0.08°$ for a 0.68% containment.
Each telescope employs a *photomuliplier tube* (PMT) camera with 499 pixels at the focal plane. Each PMT has a hexagonal Winston cone put on. They reduce the dead space between the pixels and limit the contamination from stray light [73]. The cameras cover a total *field of view* (FoV) of a diameter 3.5°. Each pixel covers a patch with a diameter of 0.15° of the total FoV. In nominal operation, the PMTs are operated with an average high voltage (HV) of ~ 1000 V.

[72]: Davies et al. (1957), 'Design of the quartermaster solar furnace'

[73]: Winston et al. (1971), 'Retinal Cone Receptor as an Ideal Light Collector'

4.1.3 Data Acquisition and Trigger System

The PMT signals are passed to a pre-amplifier board and sent to an 8-bit *flash analog-to-digital converter* (FADC) for data acquisition. It digitizes the signals at a rate of 500 MHz and stores the pulses temporarily in a buffer. Each pulse step increments a 2 ns interval. Deploying high- and low-gain amplifiers increases the dynamic range of the FADC. The low-gain signal is digitized for pulses in which the high-gain exceeds the dynamic range. It is impossible and unnecessary to store the continuous data stream during operation. A trigger system discriminates signals of Cherenkov air-showers from random fluctuations of the night-sky background (NSB) light and hadronic showers. VERITAS employs three trigger levels. The digitized signals are read out and stored if they meet the criteria for all three levels.

L1 The first trigger is applied to the individual pixels. The L1 requires a pulse from a PMT to exceed a specific threshold. It deploys a *constant-fraction-discriminator* (CFD) that ensures a stable trigger timing [74]. The pulses after the CFDs are sent to the level-two trigger.

L2 The second trigger is a pattern trigger on the telescope level. Cherenkov light coming from γ-ray induced air showers reaches the camera in spatial and temporal coincidence. By contrast, the position and timing of the L1 triggers due to fluctuations of the NSB are uncorrelated. The L2 requires three neighboring PMTs to receive an L1 trigger within a 5 ns time window in a telescope. This level significantly suppresses accidental triggers due to the NSB fluctuations.

[74]: Hall et al. (2003), 'Veritas CFDs'

L3 The third and final trigger is an array trigger. It requires coincidence L2 triggers from at least two telescopes within a 50 ns time window. The muons in hadronic-induced air showers produce Cherenkov light that can pass the L2 trigger. However, the spatial separation of the telescopes makes it unlikely that the signal of single muons triggers more than one telescope. The L3 trigger suppresses the muon triggers by a factor of ~ 10.

When an event fulfills all trigger conditions, parts of the buffered pulses are written to the disk. The typical readout window is 16 samples. This corresponds to a pulse duration of 32 ns.

4.1.4 Weather Monitoring

For the science analyses, an assessment of the data quality is crucial. Clouds increase the opacity of Cherenkov light in the atmosphere. Thus, they have a critical effect on the detection efficiency of gamma rays.
In order to identify these, three *far-infrared* (FIR) cameras on the observation site monitor the sky conditions. These record the sky temperature during observations. One camera constantly monitors the temperature at the zenith. The other two are attached to the telescopes T1 and T3 and point in the same direction as these telescopes. Thus, they observe the sky temperature directly in the center of the FoV. When clouds pass, the temperature increases compared to the clear night sky. These signatures reveal periods with clouds. For scientific analyses, this data is discarded. We use the FIR time series in Chapter 6 to identify periods that are affected by clouds.

4.2 Data Analysis

As described in section 3.2, IACTs provide indirect detection of γ-rays via the Cherenkov-radiation produced by EM cascades in the atmosphere. The properties of the primary target need to be inferred from the recorded information. In the following, we review the most critical steps of this process. In this work, we use the VERITAS standard analysis package EventDisplay in version v487.

4.2.1 Calibration and Signal Extraction

During each observing night, the so-called *flasher* runs are recorded. For these, pulses of lasers illuminate all pixels. The analysis pipeline determines the conversion factors of the amplitudes in the PMTs to the number of photoelectrons (p.e.) and the timing offsets.
Due to the NSB, the recorded pulses include an inevitable baseline. During data taking, forced triggers are injected at 1 Hz [75]. These record the so-called pedestal events that do not contain Cherenkov signals. They are used to estimate the baselines for each pixel. The variance of the pedestal values of all pixels scales with the NSB level. A more significant variance corresponds to a brighter background. This value is called the pedestal variance ped_var. It is used later for image cleaning.
Only the samples of the 32 ns pulses that contain the $\sim$ 10 ns long

[75]: Daniel (2008), 'The VERITAS standard data analysis'

Cherenkov signals are from interest. EventDisplay uses a double pass method for extracting the relevant part of the signal. The first pass provides an estimated arrival time of the Cherenkov signal for each pixel. An integration window of 6 samples is used during the second pass that integrates the relevant parts of the pulses. The subtraction of the baseline from the pedestal removes the contribution of the noise. The integrated charges for all individual pixels yield the image of the shower.

4.2.2 Image Cleaning and Characterization

A cleaning procedure is applied to identify the pixels with Cherenkov signals from γ-ray-induced showers. This procedure can remove triggered images caused by noise or background events. However, γ-like background events cause images similar to γ-rays. We describe the advanced background rejection mechanism in subsection 4.2.4.
EventDisplay uses a two-level filter with thresholds q_1 and q_2, where $q_1 > q_2$. The total signal depends on the NSB, so the variable thresholds incorporate the pedestal variance. First, the *image pixels* that have integrated signals $q_1 > 5\,\text{ped\_var}$ are identified. Next, the *border pixels* are searched adjacent to the image pixels. Their signals are $q_2 > 2.5\,\text{ped\_var}$. This procedure is robust under a wide range of conditions [76].
In section 3.2, we introduced the Hillas parameters to characterize the Cherenkov images of air showers. These are determined after image cleaning with a log-likelihood fit of a two-dimensional normal distribution to the pixels. Extrapolation beyond the camera edge allows recovering partially contained images. This procedure increases the sensitivity for events with high energies that typically correspond to extended images in the camera and events at large offsets to the camera center.

[76]: Maier et al. (2018), 'Eventdisplay: An Analysis and Reconstruction Package for Ground-based Gamma-ray Astronomy'

4.2.3 Stereo Parameters

After the image characterization, the *origins* of the γ-rays is reconstructed. Under ideal conditions, the major-axes of two individual images of the same shower intersect at the origin of the γ-ray. The reconstructed origin of an event is the mean of the intersection points of all image pairs. In this approach, intersections of image pairs with larger sizes, higher eccentricity, and larger relative angles between the major-axes have higher weights. This procedure also yields the *impact parameter*. It is the distance to the center of the VERITAS array at which the projected shower axis hits the ground.
To reconstruct the energy of the primary γ-ray, look-up tables (LUTs) are generated from events simulated with a Monte Carlo approach. The VERITAS Collaboration employs CORSIKA (Cosmic Ray Simulation for KASCADE) to simulate the development of the γ-induced air showers [77]. The response of the VERITAS detector is simulated with GrOptics [78] and CARE [79]. In this thesis, we use LUTs based on CARE. They map the energy of the simulated events to directly observable parameters such as the size, the impact, the azimuth and elevation angle, the NSB brightness, and the array configuration. The average reconstructed energy of all individual images determines the reconstructed stereo energy.

[77]: Heck et al. (1998), *CORSIKA: a Monte Carlo code to simulate extensive air showers.*
[78]: Okumura et al. (2011), 'Development of Non-sequential Ray-tracing Software for Cosmic-ray Telescopes'
[79]: Otte (Visited on 19 Apr. 2022), *CARE website*

4.2.4 Background Suppression

A huge component of cosmic-ray-induced background showers dominates the total number of recorded events and conceal the sought-for γ-ray events. However, the disparities in the shower development make them distinguishable by the camera images. *Boosted Decision Trees* (BDTs) are optimized to identify the differences in Monte Carlo generated γ-ray events and hadronic events from actual observations. These differences are manifested in the mean-scaled width and length, the height of the maximum Cherenkov emission, the impact parameter, and the size of the second-largest image.
EventDisplay provides three sets of BDT cuts in the standard analysis: *soft*, *moderate*, *hard*. They optimize the detection of γ-ray sources based on their spectra. The soft cuts are most suited for sources with a soft spectrum, meaning a spectral index $\Gamma \lesssim -3.5$. The moderate cuts optimize the analysis for sources with $-3.5 \lesssim \Gamma \lesssim -2.5$. Finally, the best performance for sources with $\Gamma \gtrsim -2.5$ is obtained with the hard BDT cuts. The softer the BDT cuts, the more of the lower-energetic events are kept in the analysis. We define the energy threshold as the energy for which the energy bias reaches $< 10\%$. See details on the definition of the energy bias in subsection 4.2.5. For typical VERITAS observations at an zenith angle of 20°, the soft BDT cuts yield an energy thresholds of $\sim 170\,\mathrm{GeV}$, the moderate cuts $\sim 205\,\mathrm{GeV}$, and the hard cuts $\sim 350\,\mathrm{GeV}$ [69]. The exact energy threshold depends on the observing conditions and instrument configuration.

[69]: Krause et al. (2017), 'Improved γ/hadron separation for the detection of faint γ-ray sources using boosted decision trees'

4.2.5 Instrument Response Characterization

For the traditional analysis methods, *Instrument Response Functions* (IRFs) summarize the characteristics of the instrument and the data analysis. They describe the response of the reconstructed data to showers of known physical properties. They are usually calculated based on Monte-Carlo simulations. The IRFs allow interpreting the physical meaning of the instrument-specific results. The VERITAS performance is described by the following IRFs:

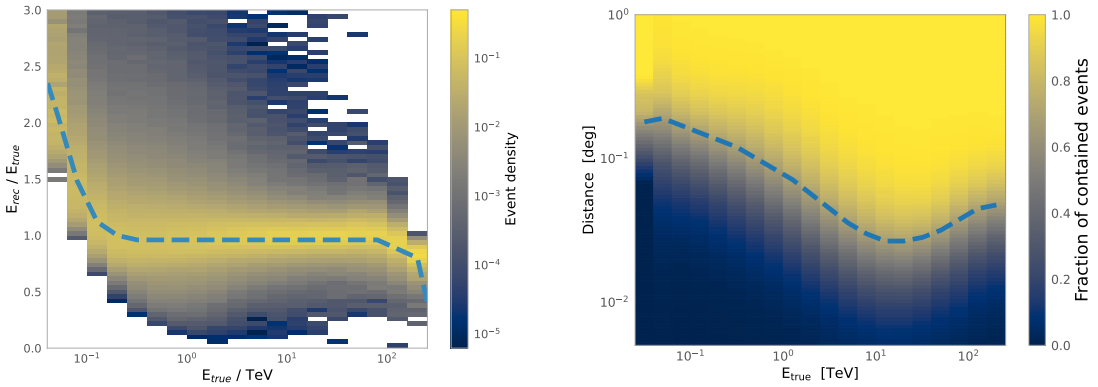
EA The *Effective Area* describes the energy-dependent effective collection area of the γ-rays. It is the product of the collection area of VERITAS and the detection efficiency for γ-ray events. It depends on the observing conditions and analysis settings and reaches up to $\sim 10^5 \mathrm{m}^2$.

ED The *Energy Dispersion* is a mapping between the reconstructed E_{rec} and true energy E_{true} of the γ-rays as function of E_{true}. It is displayed in Figure 4.2a. The energy-dependent width of the distribution is called the energy resolution. It describes the uncertainty of the reconstructed energy. Depending on E_{true}, the distribution does not center at the ideal value of 1. We define the energy bias b as a deviation of the median from 1. The blue dashed line in Figure 4.2a illustrates this bias.

PSF The *Point Spread Function* describes the uncertainty of the reconstructed origin. The PSF is given by the radius $r_{0.68}$ at which 68% of all γ-rays from a point source are contained. It depends on the observing conditions, array configuration, and the energy of the

events. The dashed blue line in Figure 4.2b shows an example of the PSF. For most observations $0.05° < r_{0.68} < 0.25°$.

RA The *Radial Acceptance* describes the detection efficiency as a function of the distance to the center of the FoV. In contrast to other IRFs, it is estimated from recorded data. It is roughly constant within the inner $\sim 0.5°$ of the FoV and decreases further out.



(a) Event density of reconstructed energy relative to the true energy $E_{\text{rec}}/E_{\text{true}}$ as function of the true energy E_{true}. The width of this energy dispersion describes the uncertainty of the energy reconstruction. The energy-dependent median of this distribution is illustrated by the dashed blue line. The difference of the median from 1 is called energy bias b.

(b) Fractional containment of events as a function of the reconstructed energy and the radius around the origin. Each energy column is independently normalized to 1. The blue dashed line is the energy-dependent 68% containment radius $r_{0.68}$ which is the definition of the PSF.

Figure 4.2: Example of the PSF and energy dispersion. These graphs are valid for the center of the field-of-view for an observation taken at 79° elevation, 146° azimuth, and a pedestal variance of 7.25.

The transient detection we implement in this thesis does not rely on the instrument characterization. Instead, a deep neural network infers the instrument characteristics from the data itself. We discuss the concept of this novel approach in depth in Chapter 5. As part of the data preparation, we use the PSF and ED, examples of which are displayed in Figure 4.2

4.2.6 Throughput Calibration

The optical throughput is the product of all factors of the camera and instrument such as the quantum efficiency of PMTs and the mirror reflectivity. The aging processes of the VERITAS instrument influence the performance and reduces the throughput. This throughput is calculated from dedicated calibration data. The standard event reconstruction scales the signals of each pixel according to the monitored throughput [80]. For this, the V6 epoch is subdivided into minor epochs which applies different throughput corrections. This approach allows consistent reconstruction of the recorded events over the complete V6 epoch. However, the scaling does not recover faint events close to the energy threshold. Consequently, the energy threshold of VERITAS increases during the V6 epoch.

[80]: Adams et al. (2022), 'The throughput calibration of the VERITAS telescopes'

Search for light Primordial Black Holes

Deep-Learning-Based Transient Detection Method for VERITAS

5

Astrophysical transient events at very-high energies provide insight into various fundamental phenomena. These include previously detected transient phenomena such as γ-ray bursts or flaring blazars and hypothesized transient phenomena such as the evaporation of primordial black holes. PBH evaporation can neither be detected in other wavelengths nor be predicted. Thus, in contrast to many other transient objects, studies need to deal with serendipitous locations and times. In a consequence, a blind search for these events is the only option. In this thesis, we present the implementation of a new transient detection approach for VERITAS based on deep learning methods.
We review the concepts of deep learning methods in section 5.1. We introduce the general concepts and expand to networks suited to deal with sequential data. These are the heart of the search for PBHs implemented in this thesis. In section 5.2, we review previous applications of deep learning methods in astronomy. We present the deep learning transient search concepts in section 5.3. We also give an overview of the total analysis pipeline. This provides context to the discussion of individual steps of the pipeline. Later chapters cover these separate parts in more depth.

5.1 Deep Learning

Deep learning is a machine learning method based on *artificial neural networks* (ANNs) with many layers. In recent years, *deep neural networks* (DNNs) have often exceeded more conventional machine learning methods in terms of performance, scalability, and flexibility [81]. The universal approximation theorem shows that sufficiently large ANNs can approximate any continuous function [82, 83]. DNNs consist of multiple layers of neurons. Each neuron collects inputs and merges these into new output. The layers of the DNN represent different levels of abstraction. Internally, it generates the required representations from the raw data. During training, it automatically discovers appropriate features from the examples. This flexibility minimizes the assumptions required by the human designer and can improve the results [84].
In subsection 5.1.1, we review the basic concepts of deep learning. We also discuss the approach of optimizing these during training. In this thesis, we work with sequential data. A specific type of DNNs, called recurrent neural networks, is particularly suited for this input type. We introduce these in subsection 5.1.2. As discussed below, they naively have difficulty in learning long-term trends. We use an advanced sub-type of these, called long short-term memory cells, to solve this. We introduce this method in subsection 5.1.3.

[81]: Abiodun et al. (2018), 'State-of-the-art in artificial neural network applications: A survey'

[82]: Cybenko (1989), 'Approximation by superpositions of a sigmoidal function'
[83]: Hornik (1991), 'Approximation capabilities of multilayer feedforward networks'

[84]: Lecun et al. (1998), 'Gradient-based learning applied to document recognition'

5.1.1 Deep Learning Basics

We show an example of a *feed-forward neural network* in Figure 5.1. This network propagates information only from the inputs in the direction of the outputs. The example consists of an *input layer* (0) two *hidden layers* ($l-1$ and l) and *output layer* L. Each layer $l \in \{0, \ldots, L\}$ has a number of neurons $k_0 = n, k_1, \ldots, k_L = m$ which values are $u_i^{[l]}$. The neurons are fully connected to all neurons in the adjacent layers. The weights between the i-th neuron in layer $l-1$ and the j-th layer in l are denoted by $w_{ij}^{[l]}$. Each neuron is associated with a bias $b_i^{[l]}$, $i \in \{1, \ldots, k_l\}$. With these notations, the inputs $z_j^{[l]}$ to neuron j in layer l, is the linear combination of the neuron values in the previous layer

$$z_j^{[l]} = \sum_{i=1}^{k_{l-1}} w_{ij}^{[l]} u_i^{[l-1]} + b_j^{[l]}. \tag{5.1}$$

We can rewrite this in matrix form

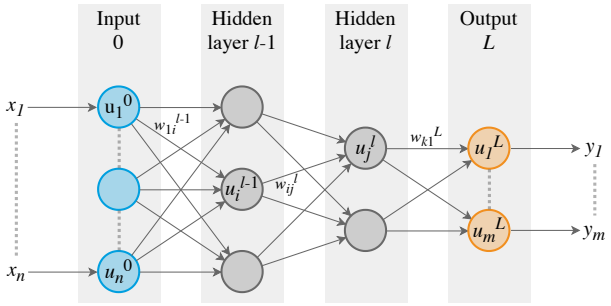


Figure 5.1: Schematic illustration of a feed-forward neural network with layers $l \in \{0, \ldots, L\}$. The input $\mathbf{x}$ is passed to $l = 0$. The neurons of adjacent layers are connected by weights $\mathbf{W}$. The value of the i-th neuron in of a layer is $u_i^{[l]}$. The outputs $\mathbf{y}$ are the values of the neurons in the last layer L. Adapted from [85, 86].

$$\mathbf{z}^{[l]} = \mathbf{W}^{[l]} \mathbf{u}^{[l-1]} + \mathbf{b}^{[l]}, \tag{5.2}$$

where $\mathbf{W}^{[l]}$ is the matrix of weights between the layers $l-1$ and l, and $\mathbf{z}^{[l]}$, $\mathbf{u}^{[l]}$ and $\mathbf{b}^{[l]}$ are vectors. This linear combination $\mathbf{z}^{[l]}$ is passed to the non-linear *activation function* $\sigma^{[l]}$ to calculate the output values $\mathbf{u}^{[l]}$ of the nodes

$$\mathbf{u}^{[l]} = \sigma^{[l]}(\mathbf{z}^{[l]}). \tag{5.3}$$

[87]: Nair et al. (2010), 'Rectified Linear Units Improve Restricted Boltzmann Machines'
[88]: Shrestha et al. (2019), 'Review of Deep Learning Algorithms and Architectures'

The *logistic function*, the *hyperbolic tangent*, and the *rectified linear unit* (ReLU) are common alternatives for the activation function [87, 88]. We illustrate these functions in Figure 5.2. The computation of partial derivatives is essential for the training of deep neural networks. The outputs $\mathbf{y}$ of the network are the activated values of the final layer $\mathbf{y} = \mathbf{u}^{[L]}$. The training process of DNNs updates the weights $\mathbf{W}$ and biases $\mathbf{b}$ to minimize the cost function C. In supervised learning, the ideal result of the DNN is the target $\hat{\mathbf{y}}$. A common choice for the cost function is the root mean square error (RMSE) between the target and the predictions of the DNN

$$C(\hat{y}, y) = \sum_{i=1}^{m} \frac{(\hat{y}_i - y_i)^2}{m}, \tag{5.4}$$

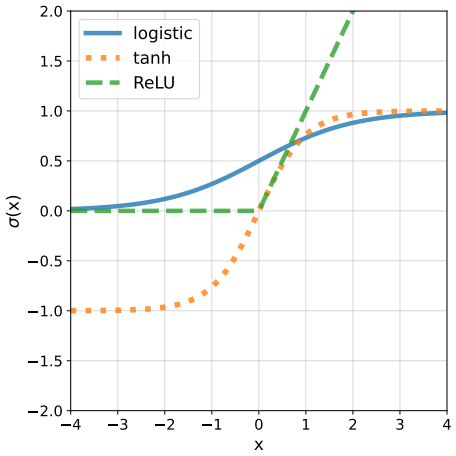


Figure 5.2: Common activation functions for deep learning.

where m is the number of outputs. Efficient optimization of the parameters of DNNs is possible using backpropagation [89]. In literature, the partial derivative

$$\boldsymbol{\delta}^{[l]} = \frac{\partial C}{\partial \mathbf{z}^{[l]}} \tag{5.5}$$

is often called the *error*. The error for neuron i in the output layer L is

$$\delta_i^{[L]} = \frac{\partial C}{\partial u_i^{[L]}} \sigma'(z_j^{[L]}), \tag{5.6}$$

where σ' is the derivative of the activation function of layer L. The first term, $\partial C / \partial u_i^{[L]}$, measures how fast the cost changes as a function of the output of this neuron. The second term, $\sigma'(z_j^{[L]})$, describes the change of the activation function at position $z_j^{[L]}$. The error can be rewritten in matrix form

$$\delta^{[L]} = \frac{\partial C}{\partial \mathbf{u}^{[L]}} \odot \sigma'(\mathbf{z}^{[L]}), \tag{5.7}$$

where $\odot$ is the *Hadamard operator* for the element-wise product. This error is iteratively propagated back to the earlier layers using

$$\boldsymbol{\delta}^{[l-1]} = \left[\boldsymbol{\delta}^{[l]}\right]^\intercal \mathbf{W}^{[l]} \odot \sigma'(\mathbf{z}^{[l-1]}). \tag{5.8}$$

The gradients of the trainable parameters, the weights $\mathbf{W}$ and biases $\mathbf{b}$, are given by

$$\frac{\partial C}{\partial \mathbf{b}^{[l]}} = \boldsymbol{\delta}^{[l]} \tag{5.9}$$

$$\frac{\partial C}{\partial \mathbf{W}^{[l]}} = \boldsymbol{\delta}^{[l]} \odot \mathbf{u}^{[l-1]}. \tag{5.10}$$

The training process updates these parameters in the opposite direction of their gradients. This is the concept of the *Gradient Descent Algorithm* [90]. The new value at iteration t for the variable x, where x can be

[89]: Rumelhart et al. (1986), 'Learning representations by back-propagating errors'

[90]: Bottou et al. (1991), 'Stochastic gradient learning in neural networks'

weights or biases, is

$$x_t = x_{t-1} - \alpha \frac{\partial C}{\partial x_{t-1}}. \tag{5.11}$$

α is the learning rate and determines the rate at which the parameters are updated. Eventually, this algorithm can minimize the cost function C. The Adam optimizer provides a possible advancement of this method [91]. It calculates individual learning rates for different parameters based on the first and second moments of the gradients.

[91]: Kingma et al. (2015), 'Adam: A Method for Stochastic Optimization'

A typical problem in training DNNs with many layers is the *vanishing gradient problem*. Backpropagation calculates the error of the first layers as the product of the partial derivatives of the activation function. Some activation functions, such as the logistic function or hyperbolic tangent, have small derivatives. Thus, the calculated product of many small values causes vanishing gradients. The training does not update parameters with vanishing gradients. This issue is addressed using *batch normalization* [92], *residual networks* [93], or an activation function with larger derivatives such as ReLU.

[92]: Ioffe et al. (2015), 'Batch normalization: Accelerating deep network training by reducing internal covariate shift'

[93]: He et al. (2016), 'Deep Residual Learning for Image Recognition'

5.1.2 Recurrent Neural Networks

The neurons in *recurrent neural networks* (RNNs) form a circle. The results of the output layer are fed back as input to the network. This feedback loop carries information from previous states and can influence the upcoming outputs. This varies from the previously discussed feed-forward network. RNNs are particularly suited to deal with sequential data efficiently. Previous work showered their outstanding performance in various fields such as natural language processing [94–96], translation [97], and time series analysis and forecasting [98].

[94]: Graves et al. (2014), 'Towards End-to-End Speech Recognition with Recurrent Neural Networks'

[95]: Song et al. (2017), 'Translation of Natural Language Query Into Keyword Query Using a RNN Encoder-Decoder'

[96]: Young et al. (2018), 'Recent Trends in Deep Learning Based Natural Language Processing [Review Article]'

[97]: Sutskever et al. (2014), 'Sequence to sequence learning with neural networks'

[98]: Hewamalage et al. (2021), 'Recurrent Neural Networks for Time Series Forecasting: Current status and future directions'

We show an example of a deep RNN in Figure 5.3. It consists of a DNN with L layers with the cyclic connection. The input $\mathbf{X}$ and output $\mathbf{Y}$ represents sequences of T steps. The internal DNN calculates the *hidden state* which is an abstract representation of the current state of the RNN. It carries information about previous entries in the RNN. Thus, the new input X_t at each sequence step t is interpreted in the context of the previous steps $\{1, \ldots, t-1\}$.

On the right side, we show the RNN in the unrolled representation. It illustrates the cyclic passages of information that represent each step t. It is important to note that the RNN consists of only one DNN shared at all steps. The hidden state $\mathbf{h}_{t-1}^{[l]}$ of each layer l of the DNN is used to calculate the hidden state $\mathbf{h}_t^{[l]}$ of the following step. Furthermore, it is also passed to the following layer in the same step. In matrix form, the hidden state at step t and layer l is

$$\mathbf{z}_t^{[l]} = \mathbf{h}_{t-1}^{[l]}\mathbf{W} + \mathbf{h}_t^{[l-1]}\mathbf{U} + \mathbf{b}_h^{[l]} \tag{5.12}$$

$$\mathbf{h}_t^{[l]} = \sigma(\mathbf{z}_t^{[l]}), \tag{5.13}$$

where $\mathbf{U}$ are the weights between the individual layers of the DNN, $\mathbf{W}$ are weight for the hidden state of the previous step, and $\mathbf{b}$ is the bias. Here we use $\mathbf{h}_t^{[0]} = \mathbf{x}_t$ to indicate the input to the first layer of the DNN. In contrast to Equation 5.3, the RNN has the additional term $\mathbf{h}_{t-1}^{[l]}\mathbf{W}$.

It carries information from one sequence step to the subsequent. The output at each step is calculated from the final hidden state at layer L with

$$\mathbf{y}_t = \mathbf{h}_t^{[L]}\mathbf{V} + \mathbf{b}_v, \tag{5.14}$$

where the weights $\mathbf{V}$ and biases $\mathbf{b}_v$ are additional model parameters.

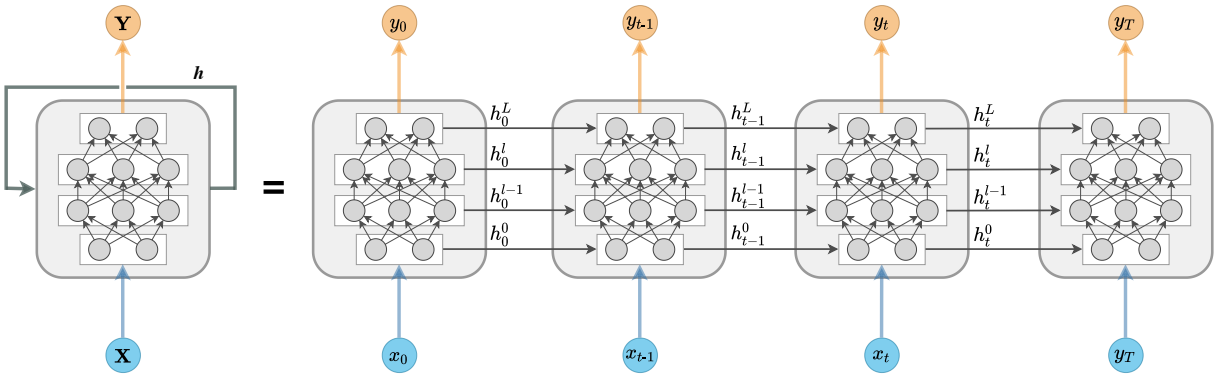


Figure 5.3: Schematic illustration of a deep recurrent neural network. The *left* side shows a compact illustration of all sequence steps. It has input vector $\mathbf{X}$ and output vector $\mathbf{Y}$ representing all time steps. We show an example of the DNN with four layers. At each step, the hidden state $\mathbf{h}$ is an additional input. It forms a cyclic connection from the network to itself. The illustration on the *right* highlights the nature of the RNN to represent sequential data. Each of the individual cells represents the identical DNN. However, the inputs $\mathbf{x}_t$ change at each step. Also, hidden states $\mathbf{h}_t$ are updated at each iteration. The hidden state $\mathbf{h}_t^l$ of layer l is passed to the next layer $l+1$ and to the same layer of the subsequent step $t+1$.

In principle, the RNN can be trained using the backpropagation approach discussed above. The total cost function is the sum of costs C_t of the individual steps in the sequence

$$C = \sum_{t=0}^{T} C_t(\hat{y}_t, y_t). \tag{5.15}$$

We illustrate the backpropagation in Figure 5.4. The forward pass (solid grey arrows) calculates the outputs $\mathbf{Y}$. The dashed red arrows illustrate the propagation path during the backward pass. The errors are not only backpropagated through the network but also backward in the sequence from step t to $t-1$. As RNNs are often used to represent time series, this process is in literature often referred to as *backpropagation through time* (BPTT).
The gradients of the total loss C concerning the weights $\mathbf{W}$ are

$$\frac{\partial C}{\partial \mathbf{W}} = \sum_{t=0}^{T} \frac{\partial C_t}{\partial \mathbf{W}} = \sum_{t=0}^{T} \sum_{k=0}^{t} \frac{\partial C_t}{\partial \hat{\mathbf{y}}_t} \frac{\partial \hat{\mathbf{y}}_t}{\partial \mathbf{h}_t} \frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_k} \frac{\partial \mathbf{h}_k}{\partial \mathbf{W}}, \tag{5.16}$$

where $\partial \mathbf{h}_t / \partial \mathbf{h}_k$ is calculated using the chain rule

$$\frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_k} = \prod_{j=k}^{t-1} \frac{\partial \mathbf{h}_{j+1}}{\partial \mathbf{h}_j} = \frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_{t-1}} \frac{\partial \mathbf{h}_{t-1}}{\partial \mathbf{h}_{t-2}} \cdots \frac{\partial \mathbf{h}_{k+1}}{\partial \mathbf{h}_k}. \tag{5.17}$$

[99]: Hochreiter (1991), *Untersuchungen zu dynamischen neuronalen Netzen*

This hidden state ratio provides a challenge during training [99]. The loss at the first step is proportional to the product of $T-1$ factors in the

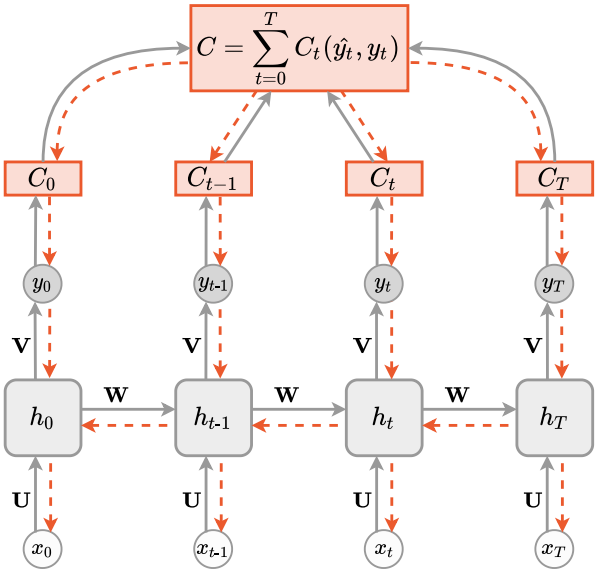


Figure 5.4: Backpropagation in recurrent neural network. The solid grey arrows illustrate operations during the forward pass. The dashed red arrows show the backpropagation path. The errors are propagated back through the network (vertical arrows) and in time (horizontal arrows).

structure

$$\frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_{t-1}} = \sigma'(\mathbf{W}\mathbf{h}_{t-1} + \mathbf{U}\mathbf{x}_t)\mathbf{W}. \tag{5.18}$$

In the case of large weights $\mathbf{W}$, we get

$$\left\| \frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_{t-1}} \right\|_2 > 1, \tag{5.19}$$

and the gradient exponentially approaches infinity. This problem is referred to as the *exploding gradient*. Similarly, when the derivative of the activation is smaller than 1 and $\mathbf{W}$ is also small,

$$\left\| \frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_{t-1}} \right\|_2 < 1, \tag{5.20}$$

and the gradient approaches 0. In this case, we face a *vanishing gradient*. These effects make it challenging to learn long-term dependencies with classical RNNs [100]. This has motivated numerous works to avoid these problems. Among them are *gradient clipping* [101], *non-saturating activation function* [102], *modified propagation paths* of gradients [103], and *gating* mechanisms [104, 105].

[100]: Bengio et al. (1994), 'Learning long-term dependencies with gradient descent is difficult'

[101]: Pascanu et al. (2013), 'On the difficulty of training recurrent neural networks'

[102]: Chandar et al. (2019), 'Towards non-saturating recurrent units for modelling long-term dependencies'

[103]: Arpit et al. (2019), 'h-detach: Modifying the LSTM gradient towards better optimization'

[104]: Hochreiter et al. (1997), 'Long short-term memory'

[105]: Cho et al. (2014), 'Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation'

5.1.3 Long Short-Term Memory

The network we use in this thesis is based on the *Long Short-Term Memory* (LSTM) [104]. It uses the approach of gating mechanisms to evade vanishing and exploding gradients of standard RNNs. It can efficiently deal with long-term dependencies in sequences.
Instead of only one DNN, it consists of four networks interacting. We illustrate them with the grey rectangles in the schematic illustration of an LSTM in Figure 5.5. Each of these networks may be a DNN with several layers itself. DNNs labeled with σ are activated with the logistic function, while the other network uses the hyperbolic tangent activation function. The black squares indicate element-wise operations on

tensors, multiplication in the case of the Hadamard operator, or addition otherwise. This illustration shows step t in the unrolled representation. From the previous step, the hidden state $\mathbf{h}_{t-1}$ and the cell state $\mathbf{c}_{t-1}$ are passed forward. We highlight the flow of the hidden and cell states in the LSTM by the dashed and dotted grey rectangles, respectively. Three gates control the values which are passed from one step to the next. Tthe *forget gate* (orange) and the *input gate* (red) modify $\mathbf{c}$. The *output gate* (blue) updates the hidden state. The output of the LSTM sequence step is computed from the updated $\mathbf{h}$. In contrast to the basic RNN, controlling the propagation through gates causes the information to flow more easily over many sequence steps.

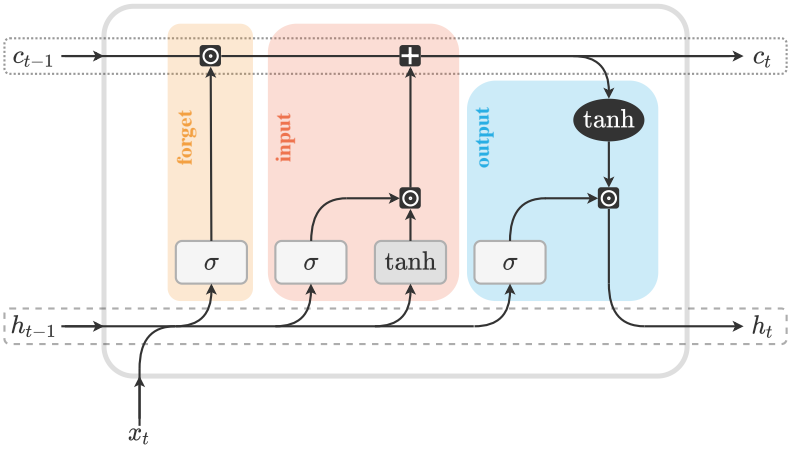


Figure 5.5: Gates in Long Short-Term Memory network. The cell state c and hidden state h are modified in the LSTM according to the forget, input, and output gates. These gates consist of a total of four DNNs that interact with each other.

Concatenating the hidden state $\mathbf{h}_{t-1}$ from the previous step H with the input at this step $\mathbf{x}_t$ yields the total input

$$\mathbf{m}_t = [\mathbf{h}_{t-1}, \mathbf{x}_t] \,. \tag{5.21}$$

First, this input is passed to the forget gate. The goal is to decide which information from the previous cell state is not required in upcoming steps. The gate removes this redundant information from the cell state. The DNN calculates the vector

$$\mathbf{f}_t = \sigma \left(\mathbf{W}_f \mathbf{m}_t + \mathbf{b}_f \right) . \tag{5.22}$$

It contains one number for each entry in the cell state. Using the combined input of the $\mathbf{h}_{t-1}$ and $\mathbf{x}_t$, it takes into account the earlier and new information. With the choice of the logistic activation function, the values of $\mathbf{f}_t$ are in the interval $[0, 1]$. The element-wise multiplication updates the cell state. Elements with $f_i \sim 0$ suppress the information. Contrary, values with $f_i \sim 1$ preserve it for future steps.
The input gate decides which new information to add to the cell state in the this step. It consists of two DNNs that both take $\mathbf{m}_t$ as input. The DNN with the hyperbolic tangent activation function calculates

$$\tilde{\mathbf{c}}_t = \tanh \left(\mathbf{W}_C \mathbf{m}_t + \mathbf{b}_C \right) . \tag{5.23}$$

This vector $\tilde{\mathbf{c}}_t$ contains the candidate values that might be added to the cell state. The network with the logistic activation function calculates

$$\mathbf{i}_t = \sigma \left(\mathbf{W}_i \mathbf{m}_t + \mathbf{b}_i \right) . \tag{5.24}$$

Its values are in the range $[0, 1]$. $\mathbf{i}_t$ filters which entries of $\tilde{\mathbf{c}}_t$ are added for the cell state. The forget and input gates together update the cell state

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i_t} \odot \tilde{\mathbf{c}}_t . \tag{5.25}$$

The updated $\mathbf{c}_t$ is passed to the next step of the sequence. It is also used to calculate the new hidden state $\mathbf{h}_t$ in the output gate. The DNN with logistic activation function calculates the vector

$$\mathbf{o}_t = \sigma \left(\mathbf{W}_o \mathbf{m}_t + \mathbf{b}_o \right) . \tag{5.26}$$

It is a filter to determine the relevant entries of $\mathbf{c}_t$ to the output. The tanh function brings the values of the cell state into the $[-1, 1]$ range. Finally, multiplication with the filter yields the updated hidden state

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t). \tag{5.27}$$

The critical factors for exploding and vanishing gradients during BPTT in standard RNNs are in Equation 5.18. For LSTMs these factors are modified into the form of

$$\frac{\partial \mathbf{c}}{\partial \mathbf{c}_{t-1}} = \frac{\partial \mathbf{f}_t}{\partial \mathbf{c}_{t-1}} \mathbf{c}_{t-1} + \mathbf{f}_t + \frac{\partial \mathbf{i}_t}{\partial \mathbf{c}_{t-1}} \tilde{\mathbf{c}}_{t-1} + \frac{\partial \tilde{\mathbf{c}}_t}{\partial \mathbf{c}_{t-1}} \mathbf{i}_t . \tag{5.28}$$

The forget gate predominantly controls the information passed at each time step. Along with the fact that Equation 5.28 consists of four additive terms, it can balance the gradients in the LSTM. This way, LSTMs can efficiently avoid the exploding and vanishing gradient problems.

5.2 Deep Learning in Astronomy

Recently, deep learning methods attracted considerable attention in astronomy. In ground-based γ-ray astronomy, previous work proved its potential for event reconstruction. These works frequently use convolutional neural networks (CNNs). These DNNs are suited to deal with data in the form of images. As discussed in subsection 4.1.3, the raw data of triggered events are the sequences of samples for each pixel. The samples from all camera pixels form a sequence of images of the air showers. Recurrent CNNs (CRNNs) can deal with sequences of images [106]. Previous studies used these CRNNs for the background suppression of IACT data [107–110]. Similarly, approaches also exist for the energy and direction reconstruction [111, 112].

In the broader context, astronomers often use deep learning to identify and classify various sources. Feed-forward networks, CNNs, and RNNs all show enormous potential for these tasks [113–116]. The ideal choice usually depends on the input data type in each survey. These methods can consider time series, morphology, or energy spectra for classification [117]. However, it is challenging to interpret the classification results as statistical significance [118].

Deep learning also proved outstanding capabilities in identifying transient astrophysical sources. The critical time dependency for these sources usually makes RNNs a suited method [119]. Frequently, studies use bidirectional RNNs that can access information from the future and the past at any time step [120, 121]. These are suited for identifying transient

[106]: Liang et al. (2015), 'Recurrent convolutional neural network for object recognition'

[107]: Postnikov et al. (2019), 'Gamma/Hadron Separation in Imaging Air Cherenkov Telescopes Using Deep Learning Libraries TensorFlow and PyTorch'

[108]: Parsons et al. (2020), 'Background rejection in atmospheric Cherenkov telescopes using recurrent convolutional neural networks'

[109]: Nieto Castaño et al. (2019), 'CTLearn: Deep Learning for Gamma-ray Astronomy'

[110]: Spencer et al. (2021), 'Deep learning with photosensor timing information as a background rejection method for the Cherenkov Telescope Array'

[111]: Miener et al. (2021), 'IACT event analysis with the MAGIC telescopes using deep convolutional neural networks with CTLearn'

[112]: Nieto et al. (2021), 'Reconstruction of IACT events using deep learning techniques with CTLearn'

[113]: Rezaei et al. (2022), 'DECORAS: detection and characterization of radio-astronomical sources using deep learning'

[114]: Rabeendran et al. (2021), 'A Two-stage Deep Learning Detection Classifier for the ATLAS Asteroid Survey'

[115]: Becker et al. (2020), 'Scalable end-to-end recurrent neural network for variable star classification'

[116]: Möller et al. (2019), 'SuperNNova: an open-source framework for Bayesian, neural network-based supernova classification'

[117]: Finke et al. (2021), 'Classification of Fermi-LAT sources with deep learning using energy and time spectra'

[118]: Gebhard et al. (2019), 'Convolutional neural networks: A magic bullet for gravitational-wave detection?'

[119]: Duev et al. (2019), 'Real-bogus classification for the zwicky transient facility using deep learning'

[120]: Charnock et al. (2017), 'Deep Recurrent Neural Networks for Supernovae Classification'

[121]: Hinners et al. (2018), 'Machine Learning Techniques for Stellar Light Curve Classification'

objects with fully recorded light curves. However, it does not conform to the real-time detection of transients. Contrary uni-directional RNNs only use inputs from time steps previous to potential transient signals. This is a powerful method for the real-time detection of transient sources [122]. A mismatch between the training and inference data can cause problems for DNNs. An analysis for source detection must include all systematic effects during the training. Also, the composition of the training data, i.e., the proportion of different effects, influences the performance. Traditionally, it requires careful modeling of the detector response under all possible observing conditions. An alternative approach is using data-driven methods. Rather than training on a synthetic dataset, they employ actual observational data. Data-driven approaches can be utilized for unbiased anomaly detection [123, 124].

[122]: Muthukrishna et al. (2019), 'RAPID: early classification of explosive transients using deep learning'

A novel method for anomaly detection in IACT observations was proposed in [125]. It is a data-driven approach that utilizes an LSTM architecture. It couples the outputs of the DNN to a calibration pipeline. This pipeline assigns a statistical significance to the results. In simulated data, this method yields promising results for identifying low-luminosity γ-ray bursts.

[123]: Zhang et al. (2020), 'Data-Driven Anomaly Detection Approach for Time-Series Streaming Data'

[124]: Pimentel et al. (2014), 'A review of novelty detection'

[125]: Sadeh (2020), 'Data-driven Detection of Multimessenger Transients'

5.3 Search Strategies

This thesis uses deep learning methods to search for PBH evaporation signals in archival VERITAS data. Our approaches are an extension of the methods developed in [125]. They are developed to be used for real time transient detection and for searching for signals in archival data. Thus, it utilizes uni-directional RNNs. In this thesis, we implement an anomaly detection to search for the expected transient signals. This work is the first application of this method on actual IACT data. Compared to previous work dealing with simulated data, we need to define new search strategies to perform the search on the complete dataset. In the following, we introduce the concepts of this search. More details are also given in the corresponding chapters later in this thesis.

[125]: Sadeh (2020), 'Data-driven Detection of Multimessenger Transients'

The core of the method is an RNN using an LSTM architecture that we illustrate in Figure 5.6. It is similar to the architecture discussed in [125]. The transient detection is implemented in the python package `trans_finder`. It uses the implementation of LSTM networks in the popular packages TensorFlow [126] and Keras [127]. We decompose the RNN into an encoder and decoder phase with τ_{enc} and τ_{dec} steps, respectively. In total, the RNN represents a time series with $\tau_{\text{RNN}} = \tau_{\text{enc}} + \tau_{\text{dec}}$ steps. The inputs $\mathbf{I}$ are the number of event counts at each time step. In this theses, we consider events after the standard VERITAS background suppression which we discussed in subsection 4.2.4. The counts are separated into several energy bins η.

[125]: Sadeh (2020), 'Data-driven Detection of Multimessenger Transients'

[126]: Martín Abadi et al. (2015), *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*

[127]: Chollet et al. (2015), *Keras*

The encoder steps represent the background interval. It is a sequence of τ_{enc} event counts representing the time series before a transient signal occurs. The LSTM updates the hidden state at each time step. After the last encoder step τ_{enc}, it represents an encoded description of the background counts.

We duplicate this hidden state and use it as the input to the time steps of the decoder. The decoder uses the abstract information provided by the encoder to predict the background counts $\mathbf{B}$. This prediction estimates

the expected counts assuming no transient signal occurred. The hidden state passes additional information about previous predictions from one step to the next. This information helps to model also predictions with trends spanning several steps.
In this thesis, we add a probabilistic layer to the background predictions. We implement it with the `TensorFlow Probability` package. With this addition, the network does not directly estimate the value for the background. Instead, it optimizes a probability distribution to describe the input data. We use a normal distribution characterized by the mean and uncertainty. With this approach, we directly have an estimate for the uncertainty of the predicted counts. To evaluate the network, we sample from the estimated distribution. This sampling can also be repeated several times. The resulting predictions for the background counts varies according to the scale of he distribution.
A transient signal **T** causes the observed number of γ-like events to be **S** = **B** + **T**. We can detect the transient source when **S** is significantly higher than the expected background. We define a test statistic that is sensitive to increased signals. A calibration pipeline converts it into a statistical significance. We discuss more details below.

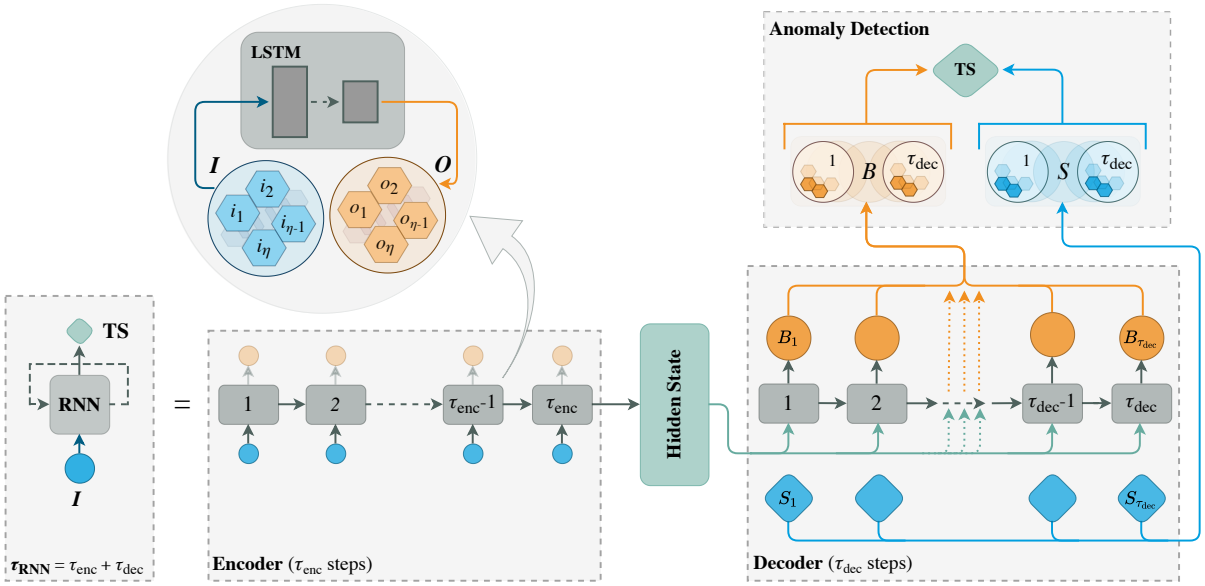


Figure 5.6: Schematic illustration of the RNN architecture implemented in `trans_finder`. The network utilizes an LSTM that consists of an encoder and decoder phase with τ_{enc} and τ_{dec} steps, respectively. The RNN represents a time series with a total of τ_{RNN} steps. During the encoder steps, the RNN calculates the hidden state. It carries encrypted information about the context of the background data. The decoder is the period during which the method can detect a potential transient signal. We use the hidden state of the encoder as input to the decoder. The RNN predicts the expected background counts **B** during the decoder sequence. We define a test statistic (TS) from the estimated and measured counts **S**. The TS increases as the signal counts rise compared to the background. A calibration pipeline transforms the TS into a statistical significance. Adapted from [125].

The RNN is used in various ways depending on the run phase. We illustrate the analysis pipeline in Figure 5.7. It is subdivided into three main run phases, *training*, *calibration*, and *inference*. Each of these phases employs actual observational data as input. In Chapter 6, we implement a tool that automatically assesses the data quality for VERITAS. We select the suited data for each run phase with this python package. Each of the phases requires a specific data preparation pipeline. For these, we implement additional run phases *prep_train*, *prep_calib*, and *prep_infer*. The general data pre-processing is common for all run phases. It generates the time series τ_{RNN} that are the inputs to the RNN. The series' main features

are the number of γ-like events during each time step. As mentioned above, we separate these into different energy bins. Furthermore, we also define multiple regions of interest (ROIs) within the FoV of VERITAS. As we discuss below in section 8.1, the size of these ROIs can increase the signal-to-noise ratio for the expected transient sources.

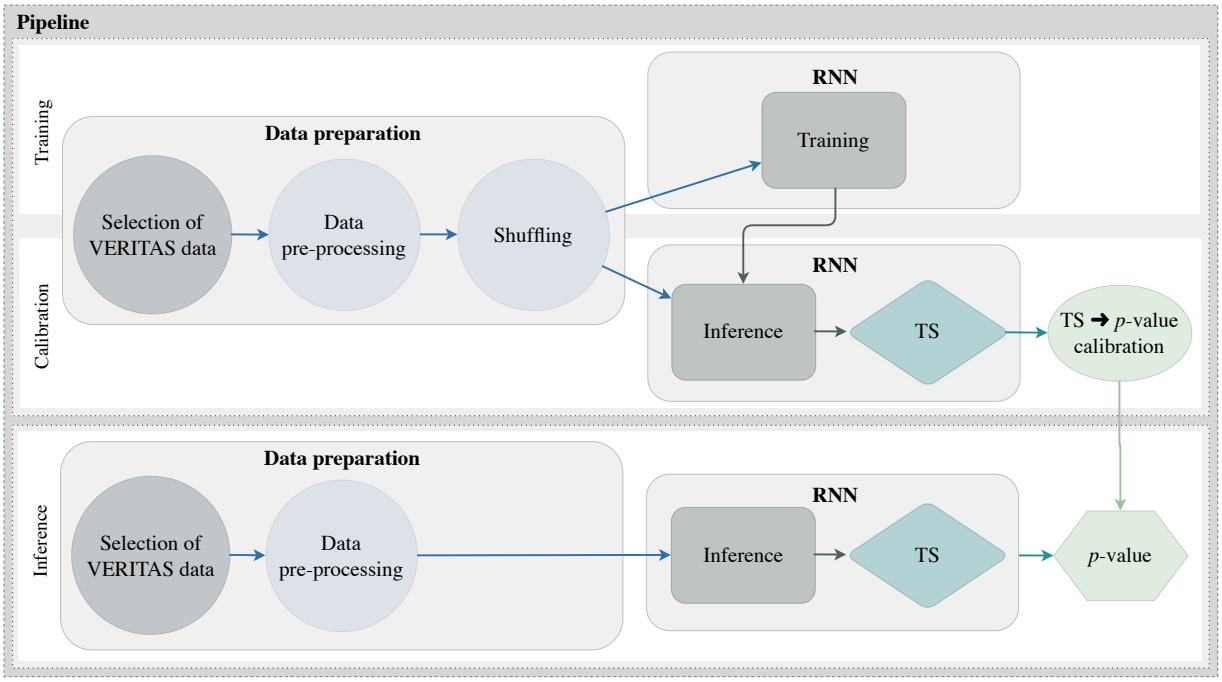


Figure 5.7: Illustration of the data preparation and analysis pipeline. The three main phases are *training*, *calibration* and *inference*. The pre-processing converts the reconstructed γ-like event lists of VERITAS into the required input to the RNN. For the training and calibration phase, a background sample is required, i.e., without possible transient sources. The shuffling algorithm removes these possible contaminants from the dataset. During the training phase, the RNN learns to predict the background counts during the decoder steps. The trained network is applied to the background dataset during the calibration phase. It yields the TS distribution for the background data. This distribution gives the frequency a certain TS value is reached under the background hypothesis. We use it to map the TS to a p-value. In the inference phase, the search for transient signals is performed. The trained RNN evaluates the unscrambled data, and the resulting TS values are converted to p-values using the calibration. Adapted from [125].

The first run phase is the *training* of the LSTM-RNN. We require a dataset without transient signals to learn predicting the background counts during the decoder steps. For this data-driven approach, we generate this background dataset using a shuffling algorithm on real data. It smears out transient contributions on the timescales we investigate in this thesis. Details of the implementation are discussed in Chapter 8. The advantage of this data-driven method is that it does not rely on modeling IRFs. Instead, the RNN extracts the system's characteristics from the training data. In order to do this efficiently, we require additional parameters besides the γ-ray counts. These characterize the relevant changes of the observing condition, which the network can learn during training. These auxiliary parameters are selected from a study of the correlation of the parameters to the observed event rates. We describe this in more detail in section 9.1.

As discussed in section 5.1, during training, the parameters of the LSTM are adjusted to minimize the cost function. Working with the probabilistic network, the cost to optimize is the negative logarithmic probability.[1] We apply the Adam optimizer [91] in this work. The shuffled training data represents sequences of pure background. The trained RNN can predict the expected number of background γ-ray counts. With the auxiliary

1: See also section 9.2 for more details.

[91]: Kingma et al. (2015), 'Adam: A Method for Stochastic Optimization'

parameters, the RNN can deal with changing observing conditions.
During the *calibration* phase, we work with a background dataset generated in the same way as for the training. We use the trained network to predict the background for these series. At each decoder time step $\tau \in \{1, \ldots, \tau_{\text{dec}}\}$, the predicted background and signal counts consist of one number per energy bin η. We denote these as $B(\tau, \eta)$ and $S(\tau, \eta)$, respectively. Further, we define each energy bin's fundamental test statistic (TS)

$$\text{TS}(\eta) = \sum_{\tau=1}^{\tau_{\text{dec}}} \frac{S(\tau, \eta) - B(\tau, \eta)}{\sqrt{B(\tau, \eta) + 1}}, \tag{5.29}$$

where the 1 in the denominator is added to stabilize the TS for small (or zero) background counts. It increases when the relative difference of the signal to the background is higher. Thus, the TS measures the strength of the transient signal. However, it can not directly be interpreted as statistical significance. The background dataset determines how frequently a specific TS value is reached under the background conditions. Comparing the estimated and actual background directly considers the precision of the RNN predictions. This frequency yields a mapping of TS(η) to a p-value. The p-value describes the probability that a pure background fluctuation caused this observation. The p-value is between 0 and 1. A smaller p-value corresponds to a higher statistical significance.
This mapping from the TS to the p-value depends on the background characteristics. Primarily, the relative standard deviation depends on the number of background counts λ. With a simplified assumption of a constant Poisson distribution, the expected uncertainty is $\sigma = 1/\sqrt{\lambda}$. Furthermore, the quality of the estimate of the background by the RNN can vary for different observing conditions. An accurate calibration needs to consider these effects. We split the observational data into *meta bins*. These account for the most significant changes in the background rates. We then run the calibration independently within each of these bins.
When a signal contribution is found in various individual p-values, combining these into one p-value can increase the overall detection significance. We can construct a new test statistic TS_{combo} as a combination of individual p-vadiscardedlues in the form of

$$\text{TS}_{\text{combo}} = -\sum_i \log p_i, \tag{5.30}$$

where i denote the individual p-value. Using the background dataset, we can map TS_{combo} again to a probability p_{combo}. We construct a calibration pipeline that uses this combination on several levels. These steps combine the TS values from the individual energy bins and different ROIs. We discuss more details of this pipeline in section 9.3.
During the *inference* phase, we use unscrambled input data to conduct the search for transients. The RNN gives the prediction of the background counts. With this estimate, the TS is calculated and mapped to a significance of detection. Here, also the corresponding meta bins are considered. During this phase, we want to investigate the entire dataset. This also includes edge cases at the beginning and the end of observing runs. As transient signals only can be detected during the decoder, we add padding data consistent with the background. These padding steps

fill the required steps to investigate the whole run. Furthermore, we do not know a priori when to expect a transient signal. In order to search consistently within the complete data, we implement a sliding window. This sliding window of length τ_{RNN} moves through complete data and generates the input sequences. In this approach, the method checks each sequence for transient signals. We discuss the details of the padding and the sliding window in section 8.3.
The detection efficiency if crucial for constraining the rate of PBH evaporation. For this calculation, we simulate signals of PBH evaporations. Folding these with the VERITAS IRFs yields the potentially detectable signals. We generate one set of simulations for each meta bin. This guarantees that the full range of observed parameters is considered. We discuss the details of the simulations in Chapter 7. In `trans_finder`, we add a run phase *prep_sims* in which we apply the same data preparation to bin the simulated events in space, energy, and time. As we have methods to generate a shuffled background dataset from actual data, we do not need to simulate the background. In the *inject_sims* phase, we superimpose the prepared simulated signals to the background from shuffled data. Finally, we run the inference for the simulated data in the *infer_sims* run phase. Compared to the normal inference phase, additional information about the simulated signals is written to the output. With this additional information, we estimate the detection efficiency as a function of the source parameters.

Data Processing and Selection 6

The processing and validation of the data quality are critical point of the analysis. Existing tools provide the possibility to perform these manually. This thesis performs an analysis on large set of observing runs recorded by VERITAS during several years of science operations. On these scales the manual standard processing and data quality monitoring tools are difficult to use. Thus we develop automatic approached for these tasks that can work for large set of observations. First we discuss the scope of this analysis in section 6.1 and how we automatize the data processing in section 6.2. The new approach for the automatic data quality monitoring is presented in section 6.3.

6.1 Analysis Scope

In subsection 4.1.1 we discussed the major upgrades of the instrument. In this thesis, we consider all data which was recorded during the latest major epoch V6. The first data in this period is recorded on September 15, 2012 and the latest data included on May 4, 2021. Only data which was recorded for scientific use is analyzed. In total 20864 science runs were recorded during this period. This data is analyzed using the standard VERITAS event reconstruction methods. The approaches of the data processing and selections are discussed below.

6.2 Data Processing

The data preparation for the input to the neural network for the anomaly detection, which is described in Chapter 8, requires the reconstructed list of events. Previously, we described the relevant reconstruction methods of the VERITAS analysis package EventDisplay v487. The individual analysis steps are divided into various subprograms.

`download`	Download raw data from archive server.
`evndisp`	Calibrate and parametrize the camera images.
`mscw_energy`	Calculate stereo parameters.
`anasum`	Apply background suppression.

Each step relies on the previous results. Existing analysis scripts run the analysis for runs based on lists of observing runs. The generation of the run lists is the duty of the analyzers. This is a well established approach for the standard VERITAS analysis which usually has a limited list of runs for specific sources. However, in this analysis all science runs within almost nine years of operations are taken into consideration. On these scales, the standard approach is not practicable as we deal with a large list of runs and many different sources. We develop `bulk_processing` which is a collection of *python* and *bash* scripts to perform the VERITAS analysis on large scales. Its objective is to automatically generate generate run-lists and manage running the different phases of the analysis. When

running the scripts, a query of the VERITAS database is performed. This yields a relevant science runs within the considered period. The paths to the directories in which the raw data and analysis files are expected are specified. By inspecting these directories, it we collect, which files are currently on the disk. We then can decide, which analysis steps are possible to start for which runs. If no files from one run are currently on disk, the raw data is downloaded from the archive server. Otherwise, the succeeding analysis step can be started until the anasum stage is reached. For the analysis stages, also the log-files are scanned to identify possible errors during the analysis. In the case of errors at a particular stage, the analysis does not proceed further for this observing run. This can happen due to corrupted files, missing information in the database, or other unexpected events during the analysis. For the sake of time, we exclude these runs from this analysis. With these disk-checks we generate lists of runs that are submitted for each of the analysis stages. These run-lists are a subsets of the initial list. Checking the disks and starting the analysis stages is an iterative process until all files reached the final analysis stage. The download for the required runs is started during the first iteration only and runs in the background. The total amount of runs that are analyzed is shown in Table 6.1.

Table 6.1: The number of runs for each analysis stage. These still contain runs with errors that are excluded at each following stages.

stage	number of runs
total	20864
`evndisp`	19329
`mscw_energy`	17679
`anasum`	17599

VERITAS recorded 20864 science runs during the period considered in this thesis. These also include runs with bad data quality, especially also runs labeled as `do_not_use`[1]. Besides bad weather, this can also hint at corrupted observations or missing data. Thus we expect a significant part of these to fail. A total of 17619 runs are processed successfully to the final stage of EventDisplay `anasum`. We use soft BDT cuts for the analysis, which provide the lowest energy threshold. The final list of runs is converted to the DL3 format [128]. It is a VERITAS independent data format and thus suited for this analysis. Besides the event list, it also contains the VERITAS-specific instrument response functions IRFs for this observation. These are used later in the simulations of PBH signals described in section 7.2. In total, 16773 runs were converted successfully into the DL3 format. The majority of the missing 846 runs are lost due to observing conditions outside the IRF axes. These runs have special observing conditions, which would not be selected for this analysis later on. The careful selection of observing runs is described in the section below.

1: This general assessment of a run denotes major problems during the data taking. See also section 6.3.

[128]: Deil et al. (2017), 'Open high-level data formats and software for gamma-ray astronomy'

6.3 Data Selection

IACTs are highly complex instruments. Various effects influence their performance. The hardware setting, such as the number of participating telescopes, is a crucial parameter that might change between one observing run and the next. Also, the observing parameters such as the elevation of the pointing position influence the performance. Furthermore,

hardware issues or bad weather can influence a run in total or partly. It is crucial to select observations that correspond to the instrument's stable performance in this work. It requires a primary selection based on the instrument setup and an assessment of the data quality over time. Time periods that are affected by uncontrolled influences need to be excluded from this analysis. It might require excluding total observing runs to achieve this. However, when only part of an observing run is affected, good periods can still be used for the science analysis. In these cases, it is possible to specify time cuts that mask the parts of the runs with poor conditions. With this approach, we do not need not to exclude data taken under stable observing conditions needlessly.

The VERITAS database stores information associated with the instrument's operations. It keeps track of the instrument's settings and the observing conditions during the operation. The database parameters listed in Table 6.2 play a crucial role in the selection of the data for this work. Among these is the type of observation (science, calibration, or engineering run), the start and end times of the runs, the high voltage setting, and the operating telescopes.

The observers who operate the instrument during observations add further information. They are responsible for manually assessing the weather conditions and noting them in the database. Here *A* denotes perfect observing conditions while *D* is the worst possible ranking.[2] Furthermore, they can add comments when they notice unexpected behavior or events during the observing runs. At the end of each night, iterating DQM-shifters perform a manual DQM. They add information to the VERITAS database, such as the status and comments about data quality. The status is a general assessment of the runs and can be one of the following options:

2: The far infrared temperature introduced in subsection 4.1.4 provides an independent monitoring of the weather conditions. As described below, both are used in the data selection.

`good_run`	run can be directly used in science analysis
`minor_problems`	minor problems encountered during operation but no further actions required
`needs_adjustments`	run can be used for science analysis after adjustments, e.g. remove part of the observation
`do_not_use`	observation should not be used for science analysis

They also have the option to suggest time cuts if necessary. While this information provides essential information about the data quality, the frequent change of the shifters can lead to inconsistent treatment of typical influences. Thus these alone are not sufficient for the data selection for this thesis. Instead, we require a more consistent approach to assess the data quality and set necessary time cuts.

This work considers all 16773 observing runs between September 2012 and May 2021 that were processed successfully and converted into the DL3 file format. Performing a manual data check is not feasible on such a large dataset. It would be very time-consuming and might still be a significant source of inconsistency. Instead, we develop a novel automatic DQM system for this work. `vts_datacheck` is a python package that assesses the quality of the runs based on information available in the VERITAS internal database. On one hand, we collect all the required information about the runs listed in Table 6.2. We preselect runs based on these instrument settings and existing DQM information. On the other hand, we add advanced diagnostics to identify periods with good and bad data quality. We automatically determine new time cuts to

Table 6.2: Basic database information that is used to identify the basic instrument setup and observing conditions for a given observation. These parameters provide a preselection of the observing runs.

Entry	Value	Comment
observation type	science	general instrumen settings
high-voltage	$\simeq 1000\,\mathrm{V}$	
operating telescopes	1, 2, 3, 4	
Weather	*train*: A *infer*: A, B	filled by observer
observer comments	*train*: none *infer*: —	
status	good_run	filled by DQM shifter
DQM comments	*train*: none *infer*: —	
DB time-cuts	none	

consistently recover as much of the stable observation periods as possible. We avoid judging the data quality based on high-level results, such as the background event rates. These are directly used in the transient detection method. Thus, a data selection based on this information would come with the risk of introducing a bias in the analysis.
The pre-selection of runs is summarized in Table 6.2. We only include science runs in which the instrument operated with all four telescopes and nominal high-voltage settings. Further, we only consider runs that are labeled *good_run* in this work. We use different sets of criteria for the weather label, the comments, and the time-cuts in the database for different phases of the deep-learning transients search. These database entries do not necessarily correspond to effects on the data quality for these runs. Nonetheless, they can still indicate abnormal behavior or external influences. Especially for the training and calibration phases require a dataset with good observing conditions. Otherwise, the reduced background event rates might bias the method. Thus, more severe selection cuts are applied for these phases to have a high efficiency of removing these. Primarily, we only include runs marked with perfect weather label (*A*), no comments in the database, and no time-cuts in the database in these phases. When we perform the actual search for PBH signals in the inference phase, we use less conservative selection criteria. As indicated in Table 6.2 we also include data with *B* weather and DQM or observer comments in the database. The following DQM assessment still identifies typical influences on this dataset and sets time cuts accordingly. However, even including periods with not perfect conditions is not as harmful in the inference phase. In these observations, the background event rates are lower and thus cannot create a false positive detection.
This preselection provides a good starting point for the analysis. However, it is still possible that the dataset contains periods with unstable conditions. Typical influences on the data quality are

- spikes in the rates (e.g., due to illumination of the camera)
- clouds passing the field of view during operation
- background illumination changes (e.g., due to Moonrise).

These periods should be excluded using time-cuts such that the good parts of the runs can still be used in the analysis. The database contains time series data which allows assessing the quality of the observations

as a function of time. Among these are

- L3 array trigger rate r_{L3}
- Far infrared (FIR) temperature of the sky T_{FIR}
- elevation and azimuth of the pointing position of the array.

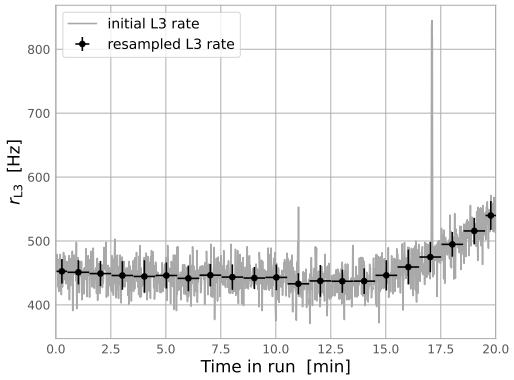
Furthermore, we add information from an external database about the position and illumination of the Moon during each observation. The previously discussed L3 rate is stored with a frequency of 1 Hz in the database. As many underlying effects are folded into it, the L3 rate is a good proxy for the stable performance of the instrument. During an observing run, the L3 rate can change for expected reasons such as a change in the elevation of the pointing position. However, if the trigger rate is not stable during the run, i.e., it contains spikes or has a significant gradient, it can also indicate problems during the observation. These include hardware issues, bad weather or external influences such as an flashing of the telescopes. Thus, it provides an excellent starting point to search for these impacts.
The FIR temperature measures the sky's temperature in the same direction as the telescopes are pointing. As clouds have a higher temperature than the clear sky, the FIR temperature can specifically help to identify these as they pass the field of view (FoV) The Moon and pointing information provide additional information to help understand changes in the L3 rates. During the Moonrises and sets, the brightness of the night sky background changes, which influences the trigger rates. We implement an evaluation of these time series to identify the periods in which the instrument's performance might be affected by typical effects.
The runs are first checked for influences by spikes or drops, next for clouds that affect the data quality and changes due to the Moonrise or set. Finally, we validate that the periods that survive these checks are consistent with the expectation for stable observing conditions.

6.3.1 Spike Detection

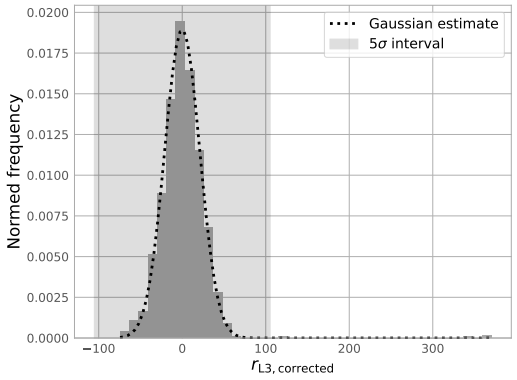
Spikes or drops are a sharp increase or decrease of the L3 rate in a short $\mathcal{O}$(s) time scale. These might be caused by a sudden illumination (e.g., from headlights) of at least one of the cameras. They affect the sensitivity to γ-like events in this period. The short time effects of spikes in the L3 rate can be superimposed to L3 changes on slower $\mathcal{O}$(min) time scales. Thus, we first apply a correction for slow effects to the L3 rate. We resample the L3 rate to 60 seconds and calculate the median rate $med(r_{\mathrm{L3}})_i$ in each of these new bins i. This median is subtracted from the second-wise initial samples within the new bin

$$r_{\mathrm{L3,corrected}} = r_{\mathrm{L3}} - med(r_{\mathrm{L3}})_i \tag{6.1}$$

Figure 6.1 illustrates this procedure. The grey curve in Figure 6.1a shows the initial data, which has at least one clear spike and a slower change in the L3 rate at the end of this observing run. The graph in Figure 6.1b shows the distribution of the L3 rate corrected by the minute-wise median. Except for very rapid changes (i.e., spikes and drops), the corrected L3 rate roughly follows a normal distribution $\mathcal{N}(\mu, \sigma)$. We estimate this



(a) Initial L3 rate during a run that contains a spike that is superimposed to a slower change. The resampled distribution within 60 seconds is illustrated by the black markers.



(b) Distribution of corrected L3 rate $r_{L3,corrected}$ within the run. The dashed line is the estimate by the normal distribution. The grey shaded area shows the 5 sigma containment. Any samples outside of this regime are considered as outliers.

Figure 6.1: Example of the L3 rate correction for the spike detection and the definition of outliers from this distribution.

distribution with

$$\mu = q_{0.5} \tag{6.2}$$

$$\sigma_{L3} = q_{0.84} - q_{0.16}, \tag{6.3}$$

where q_i are the i-quantiles of the $r_{L3,corrected}$ distribution. In the next step, we define *outliers* as an $5\sigma_{L3}$ deviation from the median of this distribution. The shaded grey area of Figure 6.1b shows this 5σ interval. Each outlier denotes the L3 rate in a one-second window. We control the maximum total effect on the analysis by the number of consecutive outliers that define a spike or drop. For this thesis, we exclude spikes that last longer than 10 % of the time steps of the analysis. This corresponds to a minimum duration of 2 sec. See also Chapter 8 for the definition of the time steps. We add time cuts around each detected spike or drop with a buffer of 15 sec. The result of this procedure for the example shown above is illustrated in Figure 6.2. One period with three consecutive outliers was identified. These are covered by the time cut, which spans the orange shaded area.

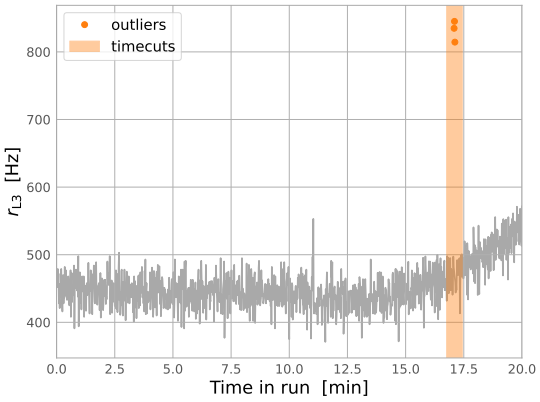


Figure 6.2: The orange circles are the three consecutive outliers that are identified. The shaded grey area illustrated the set time cuts around this spike. It includes an additional 15 seconds before and after the outliers.

6.3.2 Cloud Detection

During operations, clouds are the most crucial weather effects that can change the data quality on $\mathcal{O}$(min) timescales. They are relatively opaque to the Cherenkov light, which can drastically reduce the rate of γ-like events. The strength of the effect of clouds depends on the altitude of the cloud and the energy of the primary particles. If the shower maximum is above the clouds, the Cherenkov light is absorbed and thus the sensitivity of VERITAS decreases. The altitude of the shower maximum is lower for high primary energies [129]. In general, γ rays with energies of view TeV are little affected by clouds at high altitude of $\sim$ 10 km. The lower the cloud is, the larger is the fraction of absorbed Cherenkov photons [130]. At these primary energies and cloud altitudes of 7 km already $\sim$ 40% of photons are diminished. This effect is more significant for lower primary particles [131].
For this thesis, any data affected by clouds must be excluded to select periods with stable conditions. As mentioned above, an increase in the FIR temperature can indicate a cloud. We employ data from FIR camera attached to the T1 telescope that monitors the temperate in the FoV of VERITAS. A reduced L3 trigger rate is expected due to the increased opacity. Clouds that affect the data quality, are detectable in the FIR temperature and the L3 rate simultaneously. The effect of clouds on these two observables over the course of an observing run is illustrated in Figure 6.3.

[129]: Schoorlemmer et al. (2019), 'Characteristics of extensive air showers around the energy threshold for ground-particle-based γ -ray observatories'

[130]: Sobczyńska et al. (2014), 'Influence of clouds on the parameters of images measured by IACT at very high energies'

[131]: Dorota et al. (2020), 'An analysis method for data taken by Imaging Air Cherenkov Telescopes at very high energies under the presence of clouds'

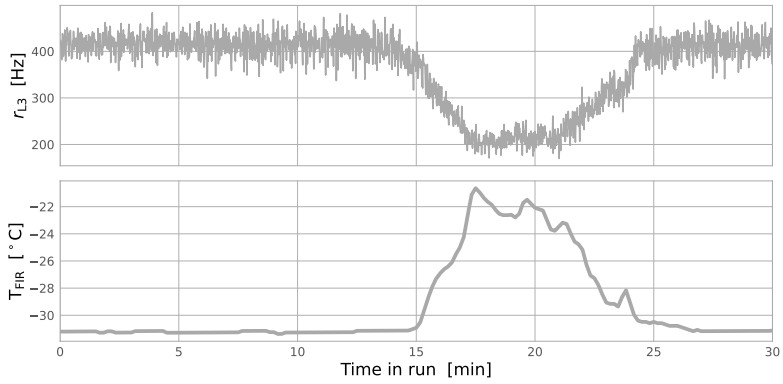
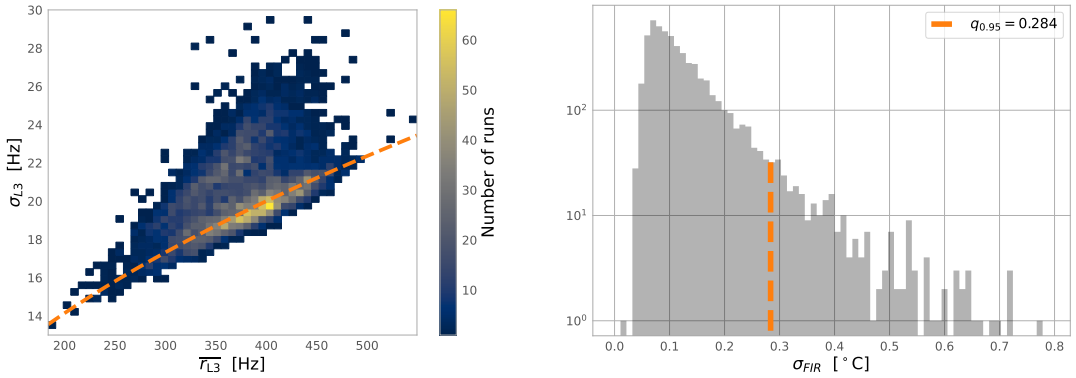


Figure 6.3: Example of the effect of clouds on the L3 rate and FIR temperature.

As the absolute scales and the expected variances are different for both metrics, we first norm them independently. The distributions are shifted by their corresponding medians $med(r_{L3})$ and $med(T_{FIR})$ which corrects for the different scales. To compare the variations of the distributions, we normalize them to the expected variance for good data. We select a list of 5975 observing runs labeled as `good_run` with stable L3 rates, no database comments, A weather, no database time cuts, and run duration above 10 minutes. This corresponds to the preselection for training the training dataset described below in subsection 6.3.5.
The distribution of the L3 rate and the standard deviation for the selected runs are illustrated in Figure 6.4a. For runs with stable observing conditions, the L3 rate follows a Poisson distribution. Thus, its expected standard deviation is a good proxy for normalization $n_{L3} = \sigma_{L3,expected} = \sqrt{r_{L3}}$. The orange dashed curve in the graph illustrates this theoretical expectation. The most frequent runs are in good agreement with this expectation. However, many runs also have up to $\sim$ 30% percent higher standard



(a) Distribution of L3 rate r_{L3} and its standard deviation σ_{L3}. The orange dashed curve indicates the expected value $\sigma_{L3} = \sqrt{r_{L3}}$ for a pure stochastic process.

(b) Distribution of the standard deviation of the FIR temperatures. The orange dashed line is the 95% quantile used for the normalization.

Figure 6.4: Distributions of L3 rate and FIR temperature for the 5975 selected runs to determine the normalizations of these parameters.

deviation. These are runs in which the L3 rate changes during the run. These happens due to changes in the NSB level or the elevation of the pointing position. These effects are handled in the VERITAS standard reconstruction methods and thus the runs are considered suitable for science analysis.

The FIR temperature adds independent measurements for the cloud detection. It helps to discriminate whether the increased σ_{L3} is due to the expected variations or rather because of clouds within the FoV. In order to estimate the standard deviation of the FIR rate σ_{FIR}, we calculate the root mean squared (RMS) of all runs in the run list. The RMS can be used as an approximation of the standard deviation. The distribution of σ_{FIR} is given in Figure 6.4b. Due to the restrictive selection cuts for this run list, the majority has good observing conditions. We rarely expect effects due to clouds in this dataset. We select the 0.95 quantile $q_{FIR,0.95}$ of this distribution for the normalization n_{FIR}. Selecting $q_{FIR,0.95}$ instead of $q_{FIR,0.68}$ ensures that good runs are not removed unnecessarily later on. Clouds are expected to cause significant changes in the FIR temperature and thus will still be identified using the higher quantile. With this normalization, we have defined scales of σ_{FIR} which we expect for observations without effects by clouds.

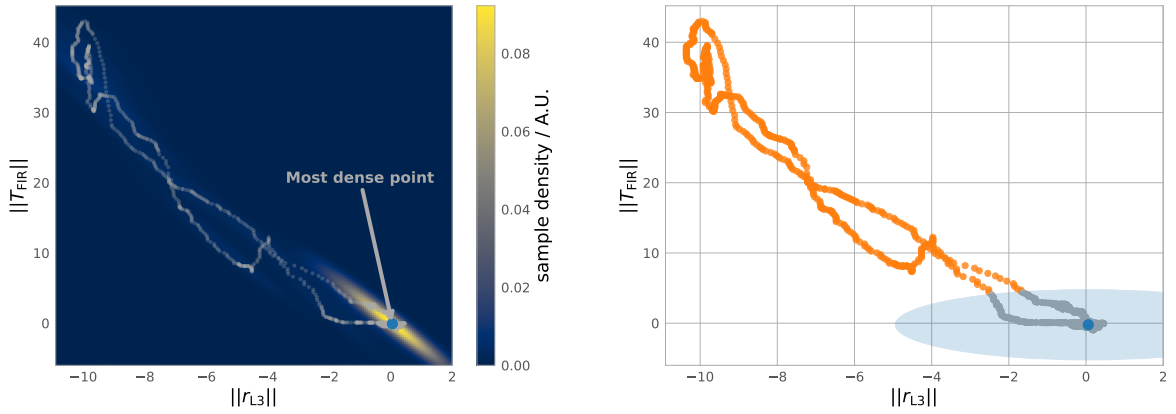
The transformed distributions are

$$\|r_{L3}\| = \frac{r_{L3} - med(r_{L3})}{n_{L3}} = \frac{r_{L3} - med(r_{L3})}{\sqrt{r_{L3}}} \tag{6.4}$$

for the L3 rate, and

for the FIR temperature. Due to the normalization to the expectations for good observation, $\sqrt{r_{L3}}$ and $q_{FIR,0.95}$, the changes in the normalized parameters can be investigated simultaneously. We interpolate the normalized FIR temperature to the same frequency of 1 Hz as the L3 values. We then investigate the 2-dimensional space of $\|r_{L3}\|, \|T_{FIR}\|$ for influences of clouds. The new samples are displayed by the grey points in

Figure 6.5a. During periods with good observing conditions, we expect a



(a) The grey points show the normalized samples in the $\|r_{L3}\|$, $\|T_{FIR}\|$ phase space. They are superimposed to the density estimate with the KDE. The most dense point in this plot is highlighted by the blue point.

(b) The blue shaded area is the circle with radius 5 around the most dense point. All orange samples are outside of this circle and thus considered outliers.

Figure 6.5: Detection of outliers in the normalized phase space of L3 rate and FIR temperature.

stable L3 rate and FIR temperature. On the other hand, clouds cause large variations in both metrics. As a consequence, the good periods cause a dense point in the $\|r_{L3}\|$, $\|T_{FIR}\|$ phase space. With the large spread, the cloudy periods have more spaced points and thus a lower density. We want to detect outliers from this good, dense point as candidates for the influence of clouds. A Kernel density estimate (KDE) is applied to determine the densest point in the phase space. [132] The selection of bandwidth h can significantly impact the results. As a rule of thumb we use the *Scott's rule* [133] which sets the bandwidth to

$$h = n^{-1/d+4}\,\mathrm{Cov}(\|r_{L3}\|, \|T_{FIR}\|), \tag{6.5}$$

with the number of data points n, the number of dimensions $d = 2$, and the covariance matrix $\mathrm{Cov}(\|r_{L3}\|, \|T_{FIR}\|)$. The colored map in Figure 6.5a shows the result density estimate with the KDE. The blue point close to (0, 0) is showing the densest point in the data. We consider all periods within a radius of 5 around the densest point as good time intervals. This interval corresponds to the blue shaded area in Figure 6.5b. Other data points are defined as outliers and illustrated in orange. For these, we add candidate time cuts with a buffer of 30 sec before and after outliers. These are shown by the orange shaded area in Figure 6.6.

As described above, we expect a simultaneous increase in the FIR temperature and a decrease in the L3 trigger rate for clouds that affect the performance. To ensure that the previous selection picked up a cloud, we check if both parameters are affected by testing their correlation. In a sliding window of 30 sec, Spearman's rank correlation coefficient r_s [134] is calculated. It is a metric for the correlation of T_{FIR} and r_{L3} within these intervals. We expect the most significant anti-correlation during periods that are affected by clouds. However, especially the L3 rate has a large noise and thus identifying clouds directly from r_s is difficult. We modify the correlation coefficient to

$$r_{s,\mathrm{mod}} = r_s \mathrm{abs}((\|r_{L3}\|)\mathrm{abs}((\|T_{FIR}\|). \tag{6.6}$$

[132]: Silverman (1986), *Density Estimation for Statistics and Data Analysis*

[133]: Scott (1992), *Multivariate density estimation : theory, practice, and visualization*

[134]: Spearman (1904), 'The Proof and Measurement of Association Between Two Things'

This modified metric suppresses the correlations during the regular intervals as the normalized values are close to 0. Thus, we expect the most significant excesses during phases affected by clouds. For the

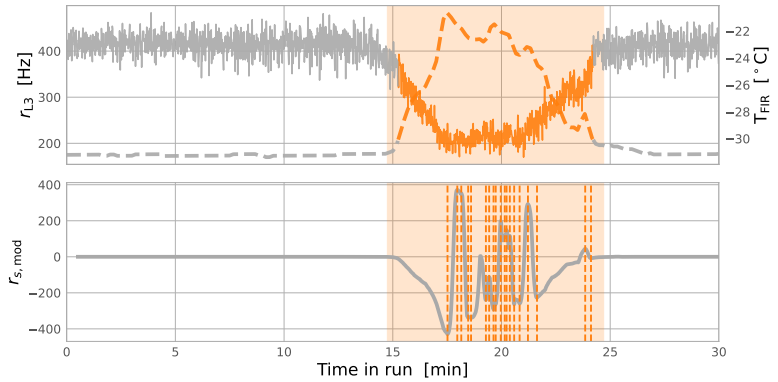


Figure 6.6: *Top:* Solid line is the L3 during the run, and the dashed curve illustrates the FIR temperature. The outliers are highlighted by orange points. The orange shaded area illustrates the candidate time cut to remove the period affected by clouds. *Bottom:* Modified correlation of the L3 rate and the FIR temperature. The vertical dashed lines display the detected spikes. As these are within the candidate time cut, we can confirm clouds as the origin of these changes.

example run above, $r_{s,\mathrm{mod}}$ is given in the lower graph of Figure 6.6. We apply a peak finding algorithm in the $r_{s,\mathrm{mod}}$ time series to identify points with a relevant correlation. We use `find_peaks` implemented in `SciPy` version `1.8.0` [135]. For this application we set the minimum prominence of the peaks to 3 and the minimum height 5. With these parameters only significant, individual peaks are identified. The found peaks correspond to the vertical dashed lines in that graph. For each candidate of a cloudy period, we require that at least one of the peaks falls within it. Otherwise, either the trigger rate or the FIR temperature is not affected. In this case, we can not confirm clouds as the reason for the outliers. In this case, we remove this time cut from the candidates. Later on, these are reevaluated to decide if they have overall good conditions or if they should be removed. This final validation is described in subsection 6.3.4.

[135]: Virtanen et al. (2020), 'SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python'

6.3.3 Changes of NSB

The VERITAS standard analysis package EventDisplay uses variable image cleaning thresholds to deal with different night-sky background levels during observations. However, the brighter the background, the higher the cleaning thresholds are, and the energy threshold of the instrument raises. Particularly Moon rises or falls can cause a rapid change in the NSB level and thus affect the low-energy performance within a run. Occasionally, also observations are affected by a ending sunset or beginning sunrise. An increase or decrease in the L3 trigger rate can indicate these. It is also possible that an observing run is affected only partly by this effect. In principle, the deep learning transient detection method could also perform well for these cases. It is possible to control these trends during the analysis with the correct parameters. However, implementing the required changes and validating the performance for these particular cases is beyond the scope of this work. Thus these should be removed for this analysis.
If a run is affected in total, it is easy to identify it because of a constantly rising or falling trigger rate. These runs are identified by the final validation described in subsection 6.3.4. However, we include stable periods in the analysis if runs are only partly affected. We implement a diagnostic to detect the changes in the slope of the L3 rate for these cases.

We fit the L3 rate with a linear function

$$r_{1,\mathrm{L3}}(c, m) = c + m * t_\mathrm{r} \tag{6.7}$$

that has $p_1 = 2$ free parameters and a piece-wise linear function

$$r_{2,\mathrm{L3}}(c, m_1, m_2, t_\mathrm{break}) = \begin{cases} c + m_1 * t_\mathrm{r} & , t_\mathrm{r} < t_\mathrm{b} \\ (c + m_1 * t_\mathrm{b}) + m_2 * (t_\mathrm{r} - t_\mathrm{b}) & , t_\mathrm{r} \geq t_\mathrm{b} \end{cases} \tag{6.8}$$

that has $p_2 = 4$ free parameters. t_r is the time relative to the beginning of the run in minutes and t_b is the time of the break.

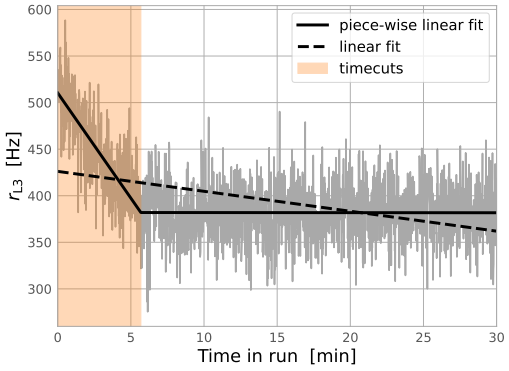


Figure 6.7: Example of the L3 rate due to a Moonset. As the NSB level decreases, the L3 rate does as well. The solid black line, which is the piece-wise linear fit, describes the curve significantly better than the linear fit. Thus, we exclude the steeper part of this fit with a time cut.

These two fits are illustrated by the solid and dashed black lines in Figure 6.7. With its flexible break-point, the piece-wise linear fit can improve describe the L3 rate better if the observations contains a point in time at which the slope has changed. If a bright Moonrise appears during the observation, such a break is expected. As the linear fit is a nested model of the piece-wise linear fit, the F-test can be used to compare the null hypothesis $H_0 : \sigma_2^2 = \sigma_1^2$ against the alternative hypothesis $H_1 : \sigma_2^2 < \sigma_1^2$ [136]. The F statistic is given by

[136]: Box (1953), 'Non-Normality and Tests on Variances'

$$F = \frac{\mathrm{RSS}_1 - \mathrm{RSS}_2}{\mathrm{RSS}_2} \frac{n - p_1}{p_2 - p_1}, \tag{6.9}$$

where RSS_1 and RSS_2 are the residual sum of squares of the two models $r_{1,\mathrm{L3}}$ and $r_{2,\mathrm{L3}}$ and n is the number of data-points for the fit. For the null-hypothesis the test statistic follows the F-distribution with $p_2 - p_1$ and $n - p_2$ degrees of freedom. If the piece-wise linear fit is preferred, we expect a F-value considerably above 1, and we can reject the null hypothesis. We set the threshold for this to 5σ. After a significant detection of a break-point, the part with the higher slope is considered the period affected by a rising NSB level and removed from the dataset. The other period is a candidate for good data quality but can still include a significant slope in the L3 rate. Thus, this is checked as well in the final validation.

6.3.4 Final Quality Evaluation

Above we developed the diagnostics to identify typical influences to the data quality. The remaining periods of the observing runs are candidates for observations with good data quality. However, in the large dataset

the effects can occur with many different appearances. Furthermore, it is possible that further events appear which are not accounted before. These are especially hardware issues such as the loss of data during the readout. Thus, we reevaluate all candidate periods. In contrast to the previous diagnostics, the goal is a general assessment of the periods. It is supposed to identify deviation from the expected, stable performance.
First, we remove periods shorter than 5 minutes from the candidates. For all other periods, we test if the slope of the L3 rate is roughly constant. A linear function

$$f(t_\mathrm{r}) = c + m * t_\mathrm{r}, \tag{6.10}$$

where m is the slope and c is a constant factor, is fitted to the L3 rate. We include a tolerance to control how much slope is still acceptable. It depends on the run duration and the expected standard deviation for a stochastic process and is given by

$$m_\mathrm{acc} = \frac{\sqrt{\overline{r_\mathrm{L3}}}}{\Delta t}, \tag{6.11}$$

where $\Delta t = t_\mathrm{max} - t_\mathrm{min}$ is the total duration of the run. t_max and t_min are the end and start times of the period, respectively. For nominal runs the duration is $\Delta t \sim 20\,\mathrm{min}$ and the rate is $\overline{r_\mathrm{L3}} \sim 400\,\mathrm{Hz}$. This yields an accepted slope of $m_\mathrm{acc} \sim 1\,\mathrm{Hz/min}$. We perform a likelihood test to determine the agreement of m to the closest value in this interval. With the assumption of a stochastic process, the L3 rate has a normally distributed noise. We define a linear curve

$$f_\mathrm{ref}(t_\mathrm{r}) = c_\mathrm{ref} + m_\mathrm{ref} * t_\mathrm{r} = c_\mathrm{ref} + \begin{cases} -m_\mathrm{acc} t_\mathrm{r} & , m < -m_\mathrm{acc} \\ m t_\mathrm{r} & , -m_\mathrm{acc} < m < m_\mathrm{acc} \\ m_\mathrm{acc} t_\mathrm{r} & , m > m_\mathrm{acc} \end{cases} \tag{6.12}$$

as reference, where c_ref is optimized to give the optimal fit to the distribution for the fixed slope. If m is within the interval, the reference curve f_ref is identical with the optimal fit f. We can test the hypothesis $H_0 : |m| \leq |m_\mathrm{acc}|$ against $H_1 : |m| > |m_\mathrm{acc}|$ by writing the log-likelihood ratio as

$$-2 \log \Lambda = \frac{1}{\sigma^2} \left(\sum_{i=1}^{n} (r_{\mathrm{L3},i} - f(t_{\mathrm{r},i}))^2 - \sum_{i=1}^{n} (r_{\mathrm{L3},i} - f_\mathrm{ref}(t_{\mathrm{r},i}))^2 \right), \tag{6.13}$$

where $r_{\mathrm{L3},i}$ are the measured data points of the L3 rate at the times $t_{\mathrm{r},i}$, and $f_\mathrm{ref}(t_{\mathrm{r},i})$ and $f(t_{\mathrm{r},i})$ are the linear curves evaluated at the same points in time. With the 1 Hz sampling of the L3 rate, there are at least 300 data points for each tested period. Thus it is valid to apply Wilks' theorem, meaning that the test statistic of $-2 \log \Lambda$ follows the χ^2-distribution with 1 degree of freedom. After evaluating the χ^2-distribution to calculate the p-value, we use the normal distribution to calculate the significance at which we can reject H_0. We remove any period with a significance above 5σ from the good periods.

6.3.5 Selection Summary

The diagnostic above makes it possible to assess the data quality of runs and select periods in which the data quality is stable. First we select all observation taken either in mode *on*, *off*, or *wobble*. These modes corresponding to a fixed position in the ICRS frame and are suited for this analysis. In this thesis we only use observation field without a significant contribution from the galactic plane. We discuss more details on this selection in below in subsection 9.1.2. Further, we select all processed runs with $\sec(\overline{\theta_z}) < 1.4$, where $\overline{\theta_z}$ is the mean zenith angle of the pointing position during a run. This corresponds to a mean elevation of $\alpha > 44.41°$. It selects the majority of observations that also provide the best sensitivity and reduces the total parameter range. All other runs are completely removed. We only consider data for which all four VERITAS telescopes are marked as good telescopes. Also, only runs in which at least 600 s left as good data after the automatic DQM are kept. Further selection criteria are according to preselection summarized in Table 6.2. Table 6.3 summarizes the selection of the total runs for the cuts common to all run phases.

Common selection criteria

	condition	N_{runs}
total candidate runs		16773
Mode	on, off, wobble	16181
galactic	False	16429
Status	good_run	11770
DB time-cuts	none	13496
UsableDuration	> 600 s	14941
good telescopes	[1, 2, 3, 4]	15538
$\sec(\overline{\theta_z})$	< 1.4	16198

Table 6.3: Common selection criteria for all run phases. The number of runs N_{runs} illustrate all runs that survive each individual cut. A summary of runs that pass all criteria is listed in Table 6.5.

As discussed in page 44, we use different preselection for the training and calibration phase compared to the inference phase. The overview of the run selection criteria with these different conditions are summarized in Table 6.4.

Different selection criteria

	train, calib		*inference*	
	condition	N_{runs}	condition	N_{runs}
Weather	A	11139	A, B	14779
observer comments	none	13189	—	
DQM comments	none	11359	—	

Table 6.4: Different selection criteria for the run phases. Equivalent to Table 6.3 for the criteria in which different cuts are applied for the training, calibration and the inference run phase. A summary of runs that pass all criteria is listed in Table 6.5.

The total selection of runs is given by these that pass all the cuts listed in Table 6.3 and Table 6.4. Table 6.5 shows an overview of the total result of the data selection. Runs excluded entirely or to large parts are removed from these run lists by requiring 10 usable minutes during the preselection. We note that even in the selected training and calibration dataset, which uses a conservative preselection, time cuts were set for 300 runs to exclude problematic periods in the data quality. It proves the importance of the automatic data quality assessment developed in this thesis. The upper graph of Figure 6.8 shows the standard deviation L3 rate σ_{L3} for all periods selected for training, and the bottom graph is the

Table 6.5: Overview of data selection for the training, calibration and prediction phase of the analysis.

	train, calib	**inference**
runs after preselection	5725	9496
total duration	2597.2 h	4287.9 h
duration after time cuts	2567.0 h	4222.6 h

same for the prediction dataset. Runs entirely removed from the run lists are not displayed. For the runs that have parts of it excluded by a time cut, we also calculate σ_{L3} only in the periods selected for the analysis. The orange arrows indicate the differences of using the full runs to the good periods only. We note that a couple of runs in each dataset have larger $\sigma_{L3} > 30$ Hz even after the time cuts. These are checked manually and validated that no systematic effect was missed.

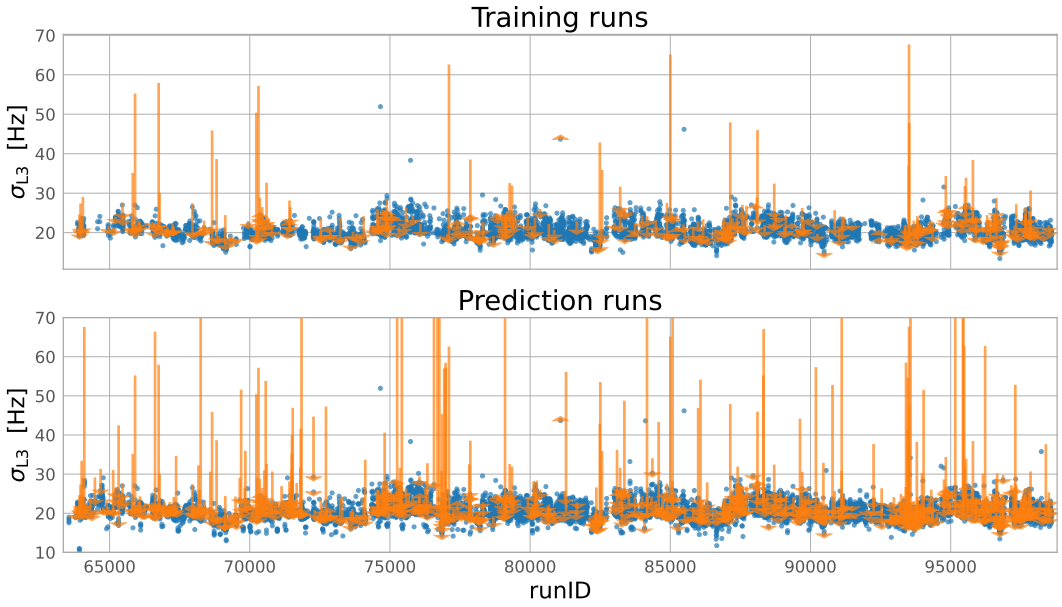


Figure 6.8: Standard deviation of the L3 rate σ_{L3} for the training and prediction datasets. The blue points only include the periods which are not removed by the automatic time cuts. If time cuts are set, the change in σ_{L3} due to these are illustrated by the orange arrows.

These distributions prove the advantage of the automatic data quality assessment developed above. For the prediction dataset, the preselection was loosened which increased the total duration by 65%. This also increases the risk of including periods with unstable data conditions. However, the automatic DQM using is able to identify these intervals and exclude them efficiently. With less complex data selection feasible on large run lists, a significant fraction of these runs would have been excluded entirely from the analysis. Furthermore, these runs are selected consistently and thus don't include selection biases.
All relevant entries associated with the data quality of the runs as well are saved in a `hdf5` file. One one hand these are the basic results of the DB query such the start and end times of the observations and the entries in listed Table 6.2. On the other hand also the results of the automatic DQM are stored, especially the time cuts as well as the reasons for it. We also store locally the time-series data from the DB such as the L3 rate, FIR temperature, elevation and azimuth. These are later reused in Chapter 8 for the data preparation.

6.3.6 Validation

The selection described in subsection 6.3.5 makes use of the diagnostics described above. Especially the automatic selection of time cuts is novel in this approach. In Figure 6.8 we showed that effects on the L3 trigger rate are identified and successfully removed. As discussed before, we use the L3 rate as an independent proxy of the data quality. A focus on the runs for which parts are excluded is given in Figure 6.9. We only display runs with timecuts and show the σ_{L3} before and after timecuts by the gray squares and blue circles, respectively. All periods that cause a large σ_{L3} are removed.

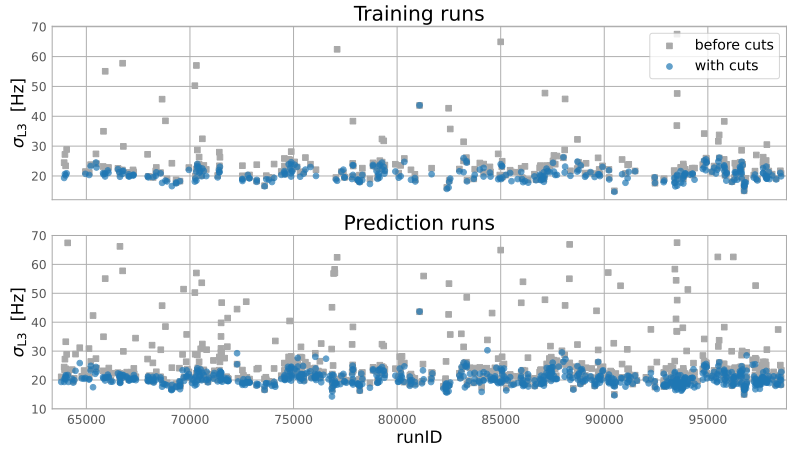


Figure 6.9: Standard deviation of the L3 rate σ_{L3} for the training and prediction datasets. Here only runs with timecuts are illustrated. The grey squares shows the raw σ_{L3} of the complete runs and the blue circle after excluding periods by the time cuts.

The fundamental objective of the data assessment is to identify and remove periods in which the rate of γ-like events r_γ are reduced by the bad data quality. We use the reconstructed event rate r_{reco} as a representative for r_γ. It is the rate of reconstructed events before the background suppression. Thus it is dominated by hadronic events. Both depend on the same observational conditions. However, r_{reco} has the advantage of a significantly larger statistic which makes it easy to study. We use it to investigate the performance of the time cuts.

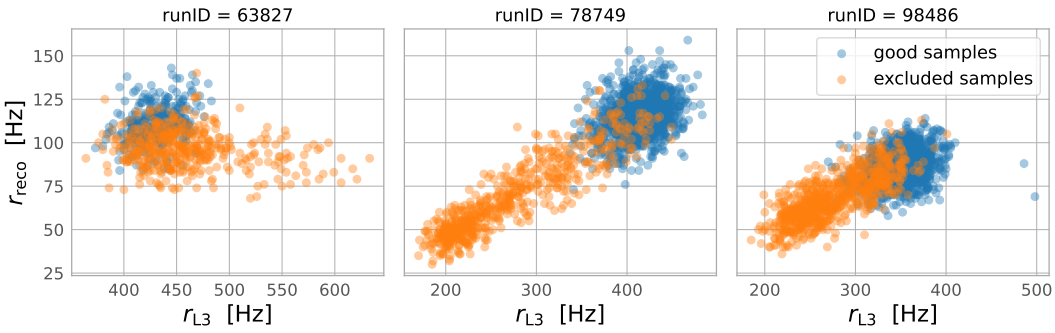


Figure 6.10: Rate of reconstructed events as a function of the L3 rate. The blue data points are samples that are included in the analysis. The orange points in the *left* graph are excluded due to a detected Moonrise. In the *middle* and *right* time cuts were specified to remove cloudy periods.

In Figure 6.10 we show three examples for the distribution of r_{reco} and r_{L3} with a 1 second sampling for individual runs. The blue points are samples that are included in the analysis, and the orange points lie within the determined time cuts. In these examples, the automatic DQM was

able to identify the periods that have varying r_{reco}. The changes in the plot left are due to a detected Moonrise and the other both show effects due to clouds. Compared to the clouds, the effect of the Moonrise on the reconstructed rate is less significant. The standard reconstruction is rather robust against changes in the NSB level. As the rates still tend to be a bit lower, we exclude the detected Moonrises nonetheless.
In the next step, we select all runs from the prediction dataset for which the automatic DQM sets a time cut. This results in a list of 700 runs. We calculate the rate of $r_{\text{reco},20}$ with 20 second sampling. This corresponds to the time binning used later in this analysis (see Chapter 8). A high standard deviation of these samples within the run is a sign of uncontrolled influences in the data. Thus we calculate the standard deviation $\sigma_{\text{reco},20}$ of these samples. Figure 6.11 shows the distribution for the selected runs. Due to the automatic time cuts, periods with significant deviations from the stable conditions are excluded for the blue histogram. Thus, the majority of runs have a lower $\sigma_{\text{reco},20}$ which indicates stable observing conditions.

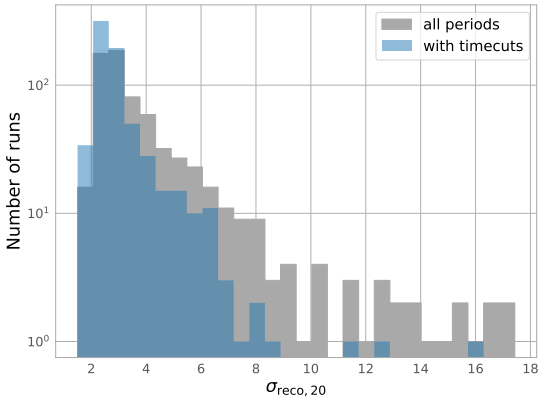


Figure 6.11: Standard deviation of the reconstructedd event rate in 20 seconds binning $\sigma_{\text{reco},20}$ for the 700 runs with time-cuts. The grey distribution includes all periods in these runs, and the blue considers only periods with good data quality. With the time cuts, periods with unstable conditions are removed, and thus $\sigma_{\text{reco},20}$ is lower.

Simulation of PBH Evaporation Signals

7

In previous chapters we discussed the concepts of the novel deep-learning based transient detection. In this thesis, we implement this method to search for evaporating primordial black holes with VERITAS. The Hawking radiation which constitutes this burst was reviewed in section 2.2. In section 7.1 we introduce a description of the expected γ-ray signals of these events. Using this parametrization, we simulate the time-dependent signals using the package `gammapy`. This is covered in section 7.2. These signals are superimposed onto a background dataset generated from observations as part of the data preparation described in Chapter 8. Thus, background events are not included into these simulations.

7.1 γ-ray Signals from Evaporating PBHs

In section 2.2 we introduced the production mechanisms of γ-rays by PBHs. We base our simulations on the theoretical work and parametrization derived in [27]. The most important process is the fragmentation and hadronization of quarks or gluons. The γ-ray production is dominated by the decay $\pi^0 \longrightarrow 2\gamma$. Due to the Lorentz boost, the angles between the γ-rays is very small. Nonetheless, over the distance of $\mathcal{O}$(pc), the split between the photons is too large to be detected by the same detector. This justifies the assumption that only one of the photons could be detected.[1] In this thesis, we use the pion fragmentation model assuming that the total fragmentation can be approximated by the production of pions. Using this assumption, a reasonable parametrization for the γ-ray contribution at energies $E_\gamma \gtrsim 1\,\mathrm{GeV}$ was derived as a function of the dimensionless parameter

[27]: Ukwatta et al. (2016), 'Primordial Black Holes: Observational characteristics of the final evaporation'

1: The expected maximum detectable distance d_{max} is about $\mathcal{O}$(pc) for VERITAS. For nearby PBHs, $d \ll 1\,\mathrm{pc}$, both photons might be detectable, making this a conservative assumption. However, the expected γ-ray signals in these cases are very strong and typically well detectable in any case.

$$x_\gamma = \frac{E_\gamma}{k_\mathrm{B} T_\mathrm{BH}(\tau)} \approx 1.287 \times 10^{-4} \left(\frac{E_\gamma}{1\,\mathrm{GeV}}\right) \left(\frac{\tau}{1\,\mathrm{s}}\right)^{1/3}, \tag{7.1}$$

where τ is the remaining lifetime of the BH, k_B is the Boltzman constant and T_BH is the BH temperature. The fragmentation contribution to the instantaneous spectrum is described by the empirical function [27]

[27]: Ukwatta et al. (2016), 'Primordial Black Holes: Observational characteristics of the final evaporation'

$$\begin{aligned}\left(\frac{\mathrm{d}^2 N_\gamma}{\mathrm{d}E_\gamma \mathrm{d}t}\right)_{\mathrm{frag.}} &= A x_\gamma^{-3/2}[1 - \Theta_S(x_\gamma - 0.3)] \\ &\quad + B \exp(-x_\gamma)[x_\gamma(x_\gamma + 1)]^{-1}\Theta_S(x_\gamma - 0.3)\end{aligned} \tag{7.2}$$

where $A = 6.339 \times 10^{23}\,\mathrm{GeV^{-1}s^{-1}}$, $B = 1.1367 \times 10^{24}\,\mathrm{GeV^{-1}s^{-1}}$, and the function $\Theta_S(x) = 0.5(1 + \tanh(10x))$.
A Monte Carlo (MC) simulation of the fragmentation can be performed with the Herwig event generator [137]. Compared to the MC simulation, the parametrization is accurate to ±15% in the range of x_γ between 0.1 and 10 At smaller and larger x_γ the precision is ±3%.
The direct Hawking emission is parametrized by [27]

[137]: Bellm et al. (2013), 'Herwig++ 2.7 Release Note'

[27]: Ukwatta et al. (2016), 'Primordial Black Holes: Observational characteristics of the final evaporation'

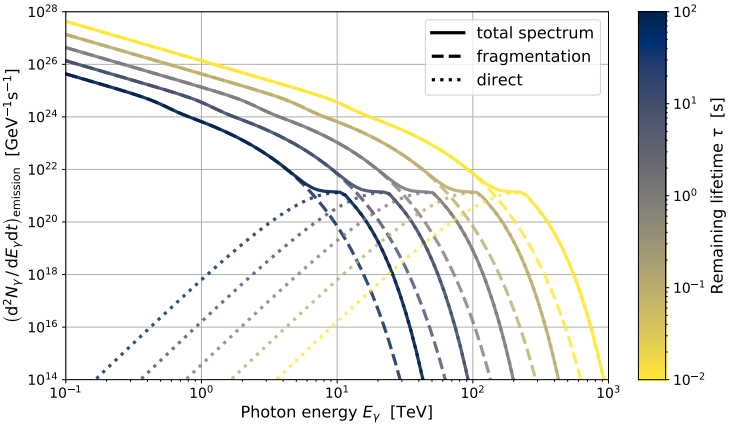


Figure 7.1: Instantaneous emission of γ-rays during the final 100 s of the BH evaporation. The colors note distinct values of the remaining lifetime $\tau \in \{100, 10, 1, 0.1, 0.01\}$ s. The total spectra (solid lines) are composed by contributions of the pion fragmentation (dashed lines) and direct Hawking emission of γ-rays (dotted lines).

$$\left(\frac{\mathrm{d}^2 N_\gamma}{\mathrm{d}E_\gamma \mathrm{d}t}\right)_{\text{direct}} = \frac{1.13 \times 10^{19}\,\text{GeV}^{-1}\text{s}^{-1} x_\gamma^6}{\exp(x_\gamma) - 1} F(x_\gamma), \tag{7.3}$$

where F is

$$F(x_\gamma) = \begin{cases} 1.0 & , x_\gamma \leq 2 \\ \exp\{[-0.0962 - 1.982(\ln x_\gamma - 1.908)] & \\ \quad \times [1 + \tanh(20(\ln x_\gamma - 1.908))]\} & , x_\gamma > 2\,. \end{cases}$$

Instead of this parametrization, many works also use the description introduced in [41]. It is derived by averaging the energy of the peak quark flux over the last τ seconds of the BH life time $Q \approx 4 \times 10^4 (\tau/1\,\text{s})^{-1/3}$ GeV. It inaccurately describes the behavior of the spectrum at E_γ slightly smaller than Q. Also the exponential fall off at the highest energies is overestimated. Equation 7.3 provides a more suited description of the direct γ-ray emission.

[41]: Linton et al. (2006), 'A new search for primordial black hole evaporations using the Whipple gamma-ray telescope'

Finally, also the decay of fundamental particles produces γ-rays. This component is small compared to the production via direct emission and pion fragmentation. Thus, it is neglected in this consideration. The total γ-ray spectrum at time τ is described by

$$\left(\frac{\mathrm{d}^2 N_\gamma}{\mathrm{d}E_\gamma \mathrm{d}t}\right)_{\text{emission}} = \left(\frac{\mathrm{d}^2 N_\gamma}{\mathrm{d}E_\gamma \mathrm{d}t}\right)_{\text{frag.}} + \left(\frac{\mathrm{d}^2 N_\gamma}{\mathrm{d}E_\gamma \mathrm{d}t}\right)_{\text{direct}}. \tag{7.4}$$

We show the instantaneous emitted spectrum for five different remaining lifetimes τ between 100 s and 0.01 s in Figure 7.1. The dashed and dotted lines show emission from fragmentation and direct emission, respectively. As τ approaches the final evaporation, the contribution of directly Hawking-emitted γ-rays shifts to higher energies, and the emission from the pion decay is scaled up.

7.2 Simulation of Evaporation Signals

This thesis aims to constrain the rate of PBH evaporations. A critical parameter for this work is the detection efficiency of the bursts. We derive this by applying the deep-learning-based transient detection to simulated signals. For this work, we use `gammapy` [138] in version `v0.20.1`.

[138]: Deil et al. (2017), 'Gammapy - A prototype for the CTA science tools'

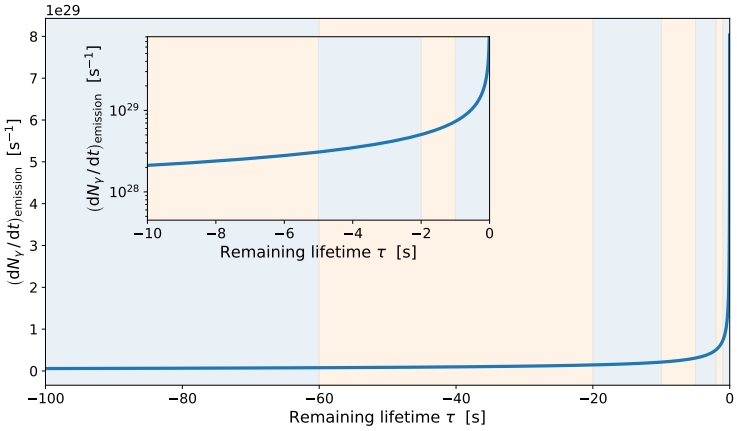


Figure 7.2: Emitted light curve of γ-rays in the energy range of 0.1 TeV and 100 TeV during the final 100 s of the BH evaporation. The smaller subplot inside focuses on the last 10 s and uses a logarithmic y-scale. The blue and orange shaded colors illustrate the time-binning used for the simulations.

We simulate the last 100 s of the BH evaporation corresponding to the most significant expected excess of events. Figure 7.2 shows the integrated light curve in the energy range of 0.1 TeV and 100 TeV. The main plot shows the full simulated time, and the smaller sub-graph shows more details during the last 10 s. As the BH is closer to the end of the lifetime, $\tau \to 0\,\text{s}$, the emission approaches infinity, $\left(\mathrm{d}N_\gamma / \mathrm{d}t\right)_{\text{emission}} \to \infty$.
As shown in Figure 7.1, not only the normalization changes in this process but also the spectral shape. Currently, it is not possible to directly simulate events coming from such changing spectra in `gammapy`. Thus, we separate the total time interval in seven shorter bins. The eight edges $[\tau_0, \ldots, \tau_7]$ are $[-100, -60, -20, -10, -5, -2, -1, -10^{35}]$ s. They are illustrated by the blue and orange shaded regions in Figure 7.2. The shorter intervals closer to the final burst reflect the fast changing spectrum at this time. The last value -10^{35} s is selected for numerical reasons to avoid a flux of ∞. For the calculation of the average spectrum, any sufficiently small final value ($\tau_7 \lesssim 10^{-20}$) yields stable results.
For the simulations, we calculate the average spectra within each of the seven time bins. The average γ-ray spectrum in the time interval τ_i and τ_{i+1} ($\tau_i < \tau_{i+1}$) until the final evaporation is

$$\left[\frac{\mathrm{d}^2 N_\gamma}{\mathrm{d}E_\gamma \mathrm{d}t}\right]_{\tau_i}^{\tau_{i+1}} = \int_{\tau_i}^{\tau_{i+1}} \frac{\mathrm{d}^2 N_\gamma}{\mathrm{d}E_\gamma \mathrm{d}t} \mathrm{d}t. \tag{7.5}$$

The average spectra in the seven time bins are shown in Figure 7.3. The shape of the average spectrum during final 1 s significantly differs from the other spectra. Especially, it roughly follows a broken power law with a break around $E_{\text{break}} \approx 10\,\text{TeV}$. For $E_\gamma < E_{\text{break}}$, the spectrum has an power law index of -1.5. The reason for this is the dominant fragmentation contribution at $E_\gamma \lesssim kT_{\text{BH}}$. At higher energies, $E_\gamma > E_{\text{break}}$, the direct Hawking-emitted γ-rays dominate. The index of -3 in this range comes from the fact that $\tau \propto T_{\text{BH}}^{-3}$ [27].
Above spectra describe the emitted photons. Ultimately, we are interested in photons that might be detectable. The Hawking radiation is homogeneous over the full solid angle. This leads to a scaling factor incorporating the distance of the BH to the observer d

$$\frac{\mathrm{d}^2 N_\gamma}{\mathrm{d}E_\gamma \mathrm{d}t} = \frac{1}{4\pi d^2} \left(\frac{\mathrm{d}^2 N_\gamma}{\mathrm{d}E_\gamma \mathrm{d}t}\right)_{\text{emission}}. \tag{7.6}$$

[27]: Ukwatta et al. (2016), 'Primordial Black Holes: Observational characteristics of the final evaporation'

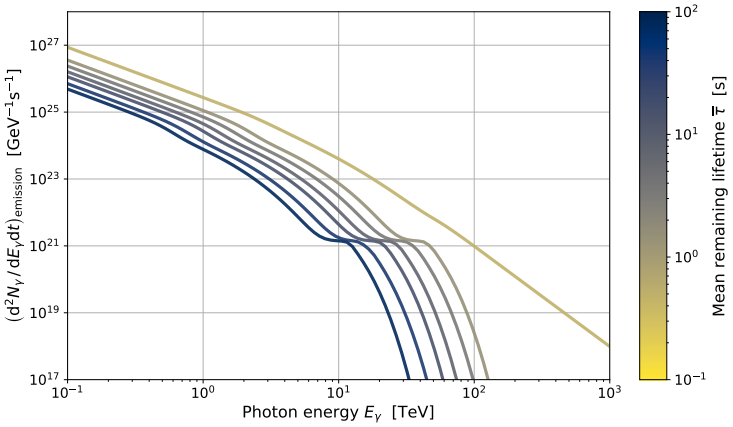


Figure 7.3: Average simulated emission of γ-rays calculated using Equation 7.5. The colors denote average time $\overline{\tau}_i = (\tau_i + \tau_{i+1})/2$ within each time bin. The average spectral shape of the last time bin differs significantly from the previous spectra. At $E_\gamma \gtrsim 10\,\text{TeV}$, it is dominated by the direct Hawking-emitted photons where the index represents the fact that $\tau \propto T_{\text{BH}}^{-3}$.

The BH distance d is the only source-dependent parameter in the detectable spectrum.
For the simulations also the detector response is crucial and has to be included. It is described by the effective area $A(E_\gamma, \hat{\Theta})$, where $\hat{\Theta}$ are all observational parameters which influence the instrument performance. For VERITAS this includes the pointing elevation and azimuth, NSB level, the offset angle ζ, observation season, and instrument configuration. ζ is defined by the position of the source relative to the pointing position of the instrument.
We load the full enclosure IRFs corresponding to a selected run from the DL3 files. We introduced these above in section 6.2. In section 5.3, we introduced the concept of meta bins grouping the runs by the most important differences. As discussed below in subsection 9.3.2, the number of relevant meta bins in this work is 24. By randomly selecting 10 runs from each meta bin, we ensure that all observing conditions are covered by the simulations. If not enough runs are available in one meta bin, we use the maximum available. The minimum number of runs in a single meta bin is eight.

Table 7.1: Summary of simulated PBH parameters.

Parameter	N	Values
τ	7	edges: $-[100, 60, 20, 10, 5, 2, 1, 10^{-35}]$ s
d	25	logarithmically between 0.15 pc and 2.5 pc
ζ	11	linearly between 0° and 2°

For all selected runs, we run the simulations with varying source parameters. We use 25 distances d logarithmically spaced between 0.15 pc and 2.5 pc. Further, we use eleven offset angles ζ linearly distributed between 0° and 2°. These parameters are summarized in Table 7.1. As part of the data preparation described in Chapter 8, the seven time bins are merged to one 100 s single evaporation event. This leaves a total of 275 combinations of the source parameters d and ζ that are simulated. For each combination and sampled observing run, 20 simulated evaporations are produced. For most meta bins, this yields a total of 200 simulated events per parameter combination. In total, $\sim 1.3 \times 10^6$ BH evaporation bursts are simulated covering the full parameter space.

Data Preparation Pipeline 8

This work is the first application of this novel deep-learning-based transient detection for actual observational IACT data. As it is not a standard analysis method, we need to develop a pipeline to prepare the VERITAS data for this method. Previously we developed a pipeline to assess the data quality and determine intervals suited for this analysis. We follow the approaches we introduced in section 5.3 for this analysis. This chapter discusses how we implement these strategies to prepare VERITAS data into `trans_finder`.

The data preparation is crucial for the performance of the transient detection method. Its purpose is to read the selected VERITAS data and generate the inputs for the RNN. The main features of the input data are the counts of γ-like events divided into ROIs, energy bins, and time bins. Figure 8.1 shows a schematic illustration of the data preparation pipeline. It provides an interface for the results of the automatic DQM with `vts_datacheck`. We read them from the `hdf5` files and perform a preselection of the observation. Details of this procedure are described in subsection 6.3.5. The DQM results also contain the periods p with stable observing conditions identified by the DQM. We search for transient signals within these intervals.

Table 8.1: Columns of the event-list stored in the DL3 files.

Name	Unit	Description
`TIME`	s	arrival time
`ENERGY`	TeV	reconstructed energy
`IS_GAMMA`		boolean value indicating if event is γ-like
`BDT_SCORE`		Score used in γ-hadron separation. Values closer to one indicate γ-like events
`EVENT_TYPE`		Number of images for this event (*multiplicity*)
`RA`	deg	reconstructed origin in ICRS
`DEC`	deg	
`ALT`	deg	reconstructed origin in horizontal system
`AZ`	deg	
`Xoff`	deg	reconstructed origin relative to center of FoV
`Yoff`	deg	

The data preparation is built on the VERITAS standard event reconstruction performed with EventDisplay. We read the list of reconstructed γ-like events from the DL3 files. Table 8.1 summarizes the information stored in these files. Primarily these are the time of arrivals, the reconstructed energy, and the reconstructed origin in the *International Celestial Reference System* (ICRS) [139] for all events. They also store metadata, such as the score used for the γ-hadron separation and whether it is a γ-like event. The `EVENT_TYPE` refers to the number of telescopes that have recorded an image of the events. In the following, we refer to this parameter as the *multiplicity* of the events.

Furthermore, the time series data stored by `vts_datacheck` is also loaded. These are the elevation, azimuth, and L3 rates during the investigated periods. We assign the observational information to each event using linear interpolation in time.

[139]: Arias et al. (1995), 'The extragalactic reference system of the International Earth Rotation Service, ICRS.'

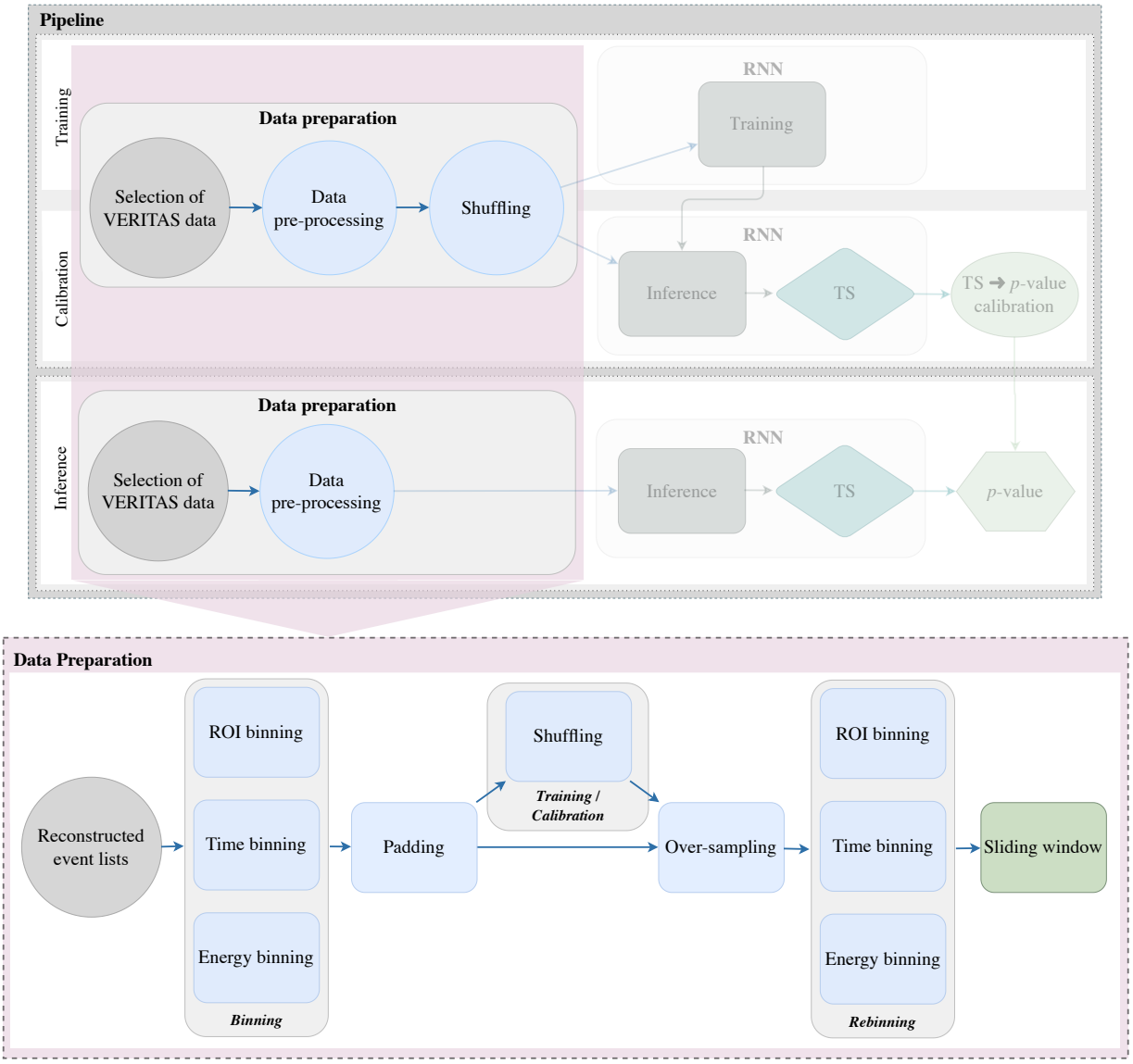


Figure 8.1: Schematic illustration of the data preparation pipeline. It generates the data required for this transient detection method. The data selection module provides an interface to apply the automatic time-cuts discussed in the previous chapter. The reconstructed γ-like event lists are read from the DL3 files. First, the events of each period p are binned in energy, time, and ROIs, which yields the time series to analyze $\mathcal{S}_\mathrm{p}$. We apply padding with background data at the beginning and end for each ROI and energy bin. For the training and calibration phases, the method requires a background dataset. We implement a shuffling algorithm that removes the contamination of possible transient signals from the data. Furthermore, we apply an oversampling of the events. It adds each event multiple times and varies its position and energy according to the expected uncertainty. This new list of events is rebinned to generate the series $\mathcal{S}_{\mathrm{p}*}$. Finally, using a sliding window approach, we extract the time series to be passed to the RNN $\mathcal{S}_\mathrm{RNN}$.

First, we calculate the time series $\mathcal{S}_p$ which covers the full length of one observing period. Each time step has a duration of $T_\mathrm{step} = 20\,\mathrm{s}$. The total number of steps depends on the duration of the observation. We assign the counts of γ-like events passing the γ-hadron selection to each time bin. These time series are calculated for three energy bins and 37 regions of interest filling the FoV. The ROIs are introduced in section 8.1 and time and energy bins are defined in section 8.2.

The required input to the RNN are time series $\mathcal{S}_\mathrm{RNN}$ of length $\tau_\mathrm{RNN} = \tau_\mathrm{enc} + \tau_\mathrm{dec}$, where τ_enc and τ_dec are the steps of the encoder and decoder, respectively. In this analysis we use $\tau_\mathrm{enc} = 20$ and $\tau_\mathrm{dec} = 5$ which results in a total length of the time series $\tau_\mathrm{RNN} = 25$. Only the decoder steps are tested for transient signals. In order to probe the full series $\mathcal{S}_\mathrm{p}$ for

transient signals, we implement a sliding window approach. It moves $\mathcal{S}_{\text{RNN}}$ through the period. By padding the beginning and end of the periods with background data, we generate the auxiliary sereis $\mathcal{S}_{\text{p}*}$. With the combination of padding and sliding window, the transient signals can be searched within the full series $\mathcal{S}_{\text{p}}$. We expand on these approaches in section 8.3.
The training and calibration phases require background datasets. The actual data could include transient signals and thus is not directly suited. We use a shuffling algorithm to smear out these contaminations. In section 8.4 we describe its implementation into the data preparation pipeline. Further, we implement an oversampling that creates an event cloud for each γ-like event. It involves the uncertainties of the direction and energy reconstruction. The oversampling that we discuss in section 8.5, helps to deal with the gaps between the ROIs. The data preparation pipeline changes event features (i.e., during shuffling) and artificially adds new events (i.e., during padding and oversampling). We rebin the updated event list using the same binning approach as before. This yields a new series that can be prepared with the sliding window.

8.1 Region of Interests

The ROIs are circular fields in the FoV with a radius of r_{ROI}. The radius should be large enough to include the majority of the source events. However, the larger the ROIs are, the higher the background counts. Thus, it is a trade-off between including a larger part of the signal and increasing the background. We select r_{ROI} such that a significant fraction of the reconstructed γ-rays are within one single ROI in case of perfect alignment between ROI and transient point source.

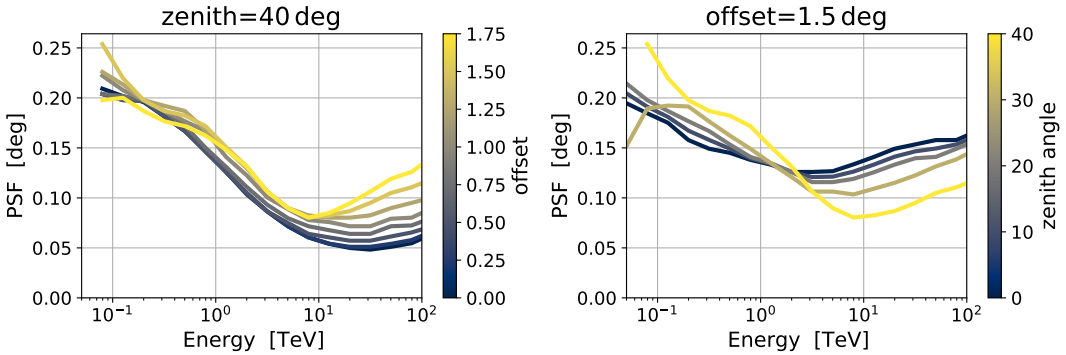


Figure 8.2: Dependency of the PSF on the zenith and offset angles. The azimuth and pedestal variance have secondary effects and thus are not shown here. On the left, we use a fixed zenith angle of $\theta = 40°$ and vary the offset angle. In the right plot, the offset is fixed at 1.5° and the θ is varied. The fixed parameters roughly correspond to the values with the largest expected PSF in this analysis.

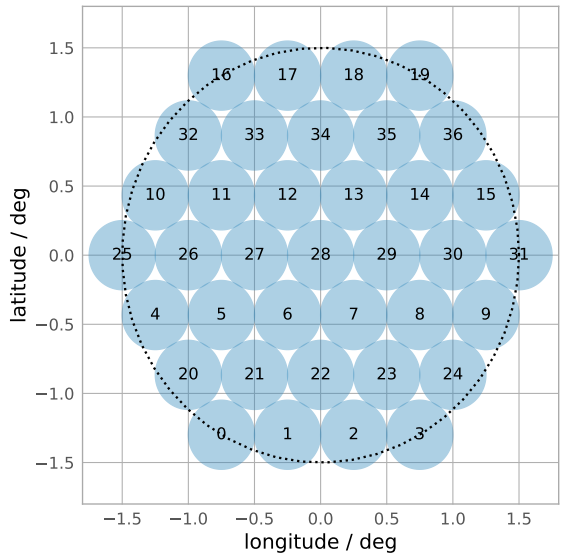
Figure 8.2 shows the distributions of the PSF for the most critical parameters, the zenith and offset angles. VERITAS has the largest PSF values at high offset angles to the center of the FoV, large zenith angles, and small energies. The other parameters, such as the azimuth angle and pedestal variance, have secondary effects on the PSF. This analysis includes observations with zenith angles $\theta \leq 44°$. The lowest energies considered are 100 GeV and the most distant ROI is 1.5° away from the center of the FoV. The PSF for observations in this analysis usually is

$\leq 0.25°$. Thus, we select $r_{\text{ROI}} = 0.25°$.

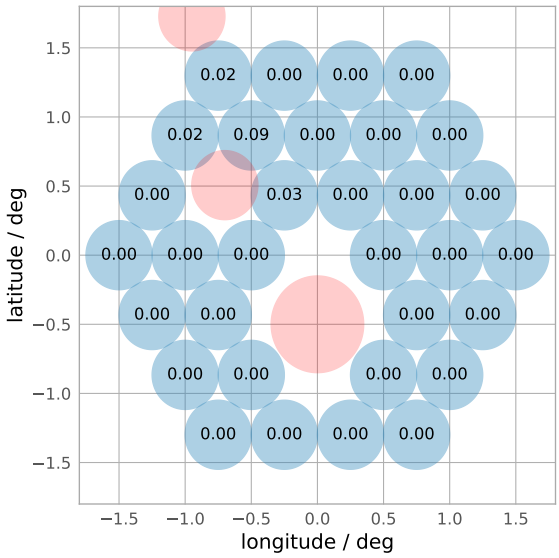
We define the ROI positions, in a `SkyOffsetFrame` implemented in `astropy` [140, 141]. These are spherical coordinates relative to the origin. The axes in this frame are the longitude and latitude angles. We use the instrument's pointing position as the frame's origin. Thus, the `SkyOffsetFrame` is centered around the middle of the FoV. We position ROIs in the FoV to have a maximum distance from the origin of 1.5°. We start adding ROIs in rows of fixed latitude. Each following row is shifted by $\Delta\text{lon} = r_{\text{ROI}}$ and $\Delta\text{lat} = \sqrt{3} r_{\text{ROI}}$. This way, the ROIs can fill the FoV without overlap and the least possible gaps. We define a total of 37 ROIs. Figure 8.3a illustrates their locations relative to the pointing position.

[140]: Astropy Collaboration et al. (2013), 'Astropy: A community Python package for astronomy'

[141]: Astropy Collaboration et al. (2018), 'The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package'



(a) In total, we define 37 ROIs that cover the inner 1.5° of the camera without overlap. The numbers inside the ROIs are the unique identifiers of the ROIs.



(b) We remove ROIs with more than 10% overlap for this analysis. The numbers in the ROIs show their corresponding overlap with the exclusion regions.

Figure 8.3: The blue circles show the ROIs in the `SkyOffsetFrame` relative to the center of the FoV. The red circles indicate the excluded regions within the FoV due to bright stars and known VHE γ-ray sources. We use the fractional overlap as a correction for the remaining ROIs.

We exclude positions with known VHE γ-ray sources and bright stars. In their presence, the background rate of γ-like events diverges from the customarily expected background rates. The treatment of these specific cases is beyond the scope of this thesis. We define circular exclusion regions around stars based on their brightness. 0.25°, 0.3° and 0.35° for stars brighter than magnitudes 6, 4, and 2, respectively. We use the Hipparcos 2 catalog for the star parameters [142].

To identify relevant γ-ray sources for VERITAS, we use the VTSCat [143]. It contains a list of sources detected or observed by VERITAS. We exclude sources with circular regions of a radius of 0.35°. As discussed in Chapter 6, we conservatively select the data for the training and calibration phases. Thus, we exclude regions around every observed source in the VTSCat. It prevents contamination by weak sources at these run phases. We slightly loosen the exclusion for the inference phase by including observed but not detected sources.

We calculate the overlapping area A_{ex} of each ROI with the excluded regions. The fraction to the total area A_{tot} that is excluded is $(A_{\text{ex}}/A_{\text{tot}})_{\text{ROI}}$. If $(A_{\text{ex}}/A_{\text{tot}})_{\text{ROI}} > 0.1$, the ROI is removed for this observing run. Other-

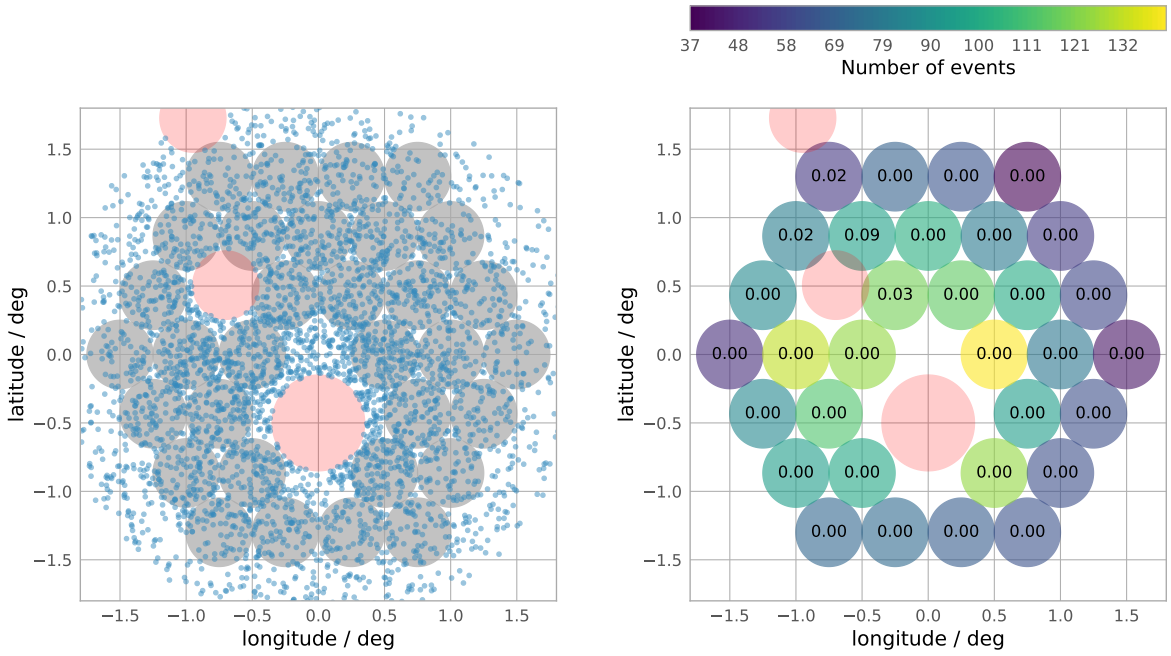
[142]: van Leeuwen (2007), 'Validation of the new Hipparcos reduction'

[143]: Patel et al. (2021), 'VTSCat: The VERITAS Catalog of Gamma-Ray Observations'

wise, we apply a correction for the expected fraction of lost events. For this correction, a homogeneous distribution of events within each ROI is assumed. The corrected number of event counts is

$$n_{\gamma,\,\mathrm{corr}}(\mathrm{ROI}) = n_{\gamma}(\mathrm{ROI})\frac{1}{1-\left(\frac{A_{\mathrm{ex}}}{A_{\mathrm{tot}}}\right)_{\mathrm{ROI}}}. \tag{8.1}$$

We show an example of the exclusion regions in Figure 8.3b. In this illustration, we remove ROIs excluded from the analysis. The numbers in each remaining ROI are the fractional overlaps with the excluded areas $(A_{\mathrm{ex}}/A_{\mathrm{tot}})_{\mathrm{ROI}}$. The counts of γ rays within all ROIs with $(A_{\mathrm{ex}}/A_{\mathrm{tot}})_{\mathrm{ROI}} > 0.0$ are corrected according to Equation 8.1.



(a) The grey and red circles are the ROIs and exclusion regions respectively. The blue dots illustrate the origin of each γ-like event within this specific observing run.

(b) The colors illustrate the total number of events within each ROI. If an ROI overlaps with an exclusion region, we correct the number of events using Equation 8.1.

Figure 8.4: Assignment of events to the ROIs. On the left, we show the reconstructed position of the individual γ-like events. Based on the spatial coincidence, they are assigned to the ROIs. We show the number of events for each ROI after the assignment on the right.

In Figure 8.4a we show the reconstructed origins of the recorded γ-like events by the blue dots. They are superimposed on the grey ROIs that fill the inner 1.5° of the FoV. We show the same FoV with exclusion regions displayed in Figure 8.3b. Events falling into any of the exclusion regions are removed for this analysis. We assign events to each ROI based on spatial coincidence with the reconstructed origin. It leads to a spatial binning of the events. The events falling in one of the gaps are not associated with any ROI. Consequently, only the events within the ROIs contribute to the search for transient signals. In order to also include events falling into the gaps, we implement an oversampling of the events. We describe this oversampling approach in section 8.5.

8.2 Time and Energy Bins

To determine the periods that are analyzed, we use the automatic DQM developed in Chapter 6. All events with a temporal coincidence with the time cuts are removed from the analysis. The distinct periods p with good observing conditions have start times $T_{\text{start, p}}$ and end times $T_{\text{end p}}$. We define a time series $\mathcal{S}_\text{p}$ for each period that spans the whole duration. Each time steps of the series has a duration of $T_\text{step} = 20\,\text{s}$. The total number of steps τ_p of each series $\mathcal{S}_\text{p}$ is given by the duration of the period

$$\tau_\text{p} = \left\lceil \frac{T_{\text{end p}} - T_{\text{start, p}}}{T_\text{step}} \right\rceil. \tag{8.2}$$

We note that $\mathcal{S}_\text{p}$ can not be used as input to the RNN directly. It has a variable length of $\tau_\text{p} \neq \tau_\text{RNN}$ steps. We transform it to the input sequences $\mathcal{S}_\text{RNN}$ utilizing a sliding window approach introduced in section 8.3. Furthermore, this method is only sensitive to transient signals within the decoder steps. Directly using $\mathcal{S}_\text{p}$, it could only investigate parts of the total period p.

We calculate the number of γ-like events in each step τ_i, where $i \in \{1, \ldots, p\}$. Due to edge effects, the duration $T(\tau_\text{p})$ of the last time bin τ_p can be shorter than 20 s. If its duration is less than 90% of T_step, we remove this time bin from the analysis. Otherwise, we correct the event counts for the fraction of lost time assuming a constant event rate during the time interval

$$n_{\gamma,\,\text{corr}}(\tau_\text{p}) = n_\gamma(\tau_\text{p}) \frac{T_\text{step}}{T(\tau_\text{p})}. \tag{8.3}$$

Including information about the spectrum of the γ-like counts can improve the detection probability of transient signals. We split the complete list of events into three energy bins. The first energy bin covers the low end of VERITAS sensitivity range from 100 GeV to 330 GeV. The middle bin covers events from 330 GeV to 1 TeV. The last bin covers the highest energy interval of the VERITAS sensitivity. Its range is from 1 TeV up to 100 TeV.

Figure 8.5 shows an example of the applied time and energy binning. Displayed is the same observing run shown in Figure 8.4b. The ROI number 29 we show here has the highest integrated number of events. The maximum number of events for this example is three counts in one of the energy and time bins. For this period, the last time bin τ_p is shorter than 90% of the time step duration T_step. Thus it is removed from the analysis.

Under the background hypothesis, the expected number of γ-like events $> 1\,\text{TeV}$ is low. As shown in section 7.1, the emitted γ-rays from PBH evaporations can reach these energies. This low-background bin can be very powerful for detecting these signals. However, during training and calibration counts of mostly 0 can be difficult to handle. Thus, we manually increase the background counts by adding Poisson noise with $\lambda = 2$. This is applied during all analysis stages and thus treated consistently. While it increases the threshold for detection, it also has an stabilizing effect on the analysis.

The number of expected γ-like events in first energy bin, $0.1 < E/\text{TeV} <$

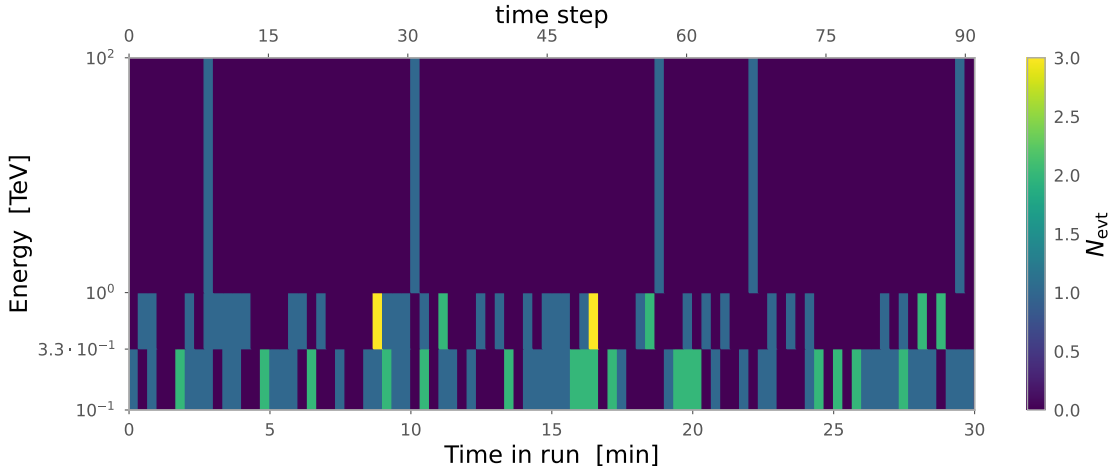


Figure 8.5: Series $\mathcal{S}_p$ of counts in ROI 29 of Figure 8.4b. All events are binned in energy and time bins as described above. The individual counts in each bin reach from 0 to a maximum of 3 counts.

0.33, depend energy threshold. It varies between the runs due to the observing conditions.[1] Within the data which we consider in this thesis, it can rise to above 330 GeV. For these runs, no events in the first energy bin are expected. Thus, we also add Poissonian noise to this energy bin. This additional noise is added just before passing the series to the RNN. It is the latest data preparation stage. In particular, it happens after the events are oversampled. We describe the details of the oversampling approach in section 8.5. The following graphs do not include this additional noise terms.

Above, we discussed the definition of the ROIs, energy, and time bins. We calculate the time series $\mathcal{S}_p$ with the counts of γ-like events for each bin. The counts are the main features of the analysis used to calculate the test statistic. The results from the different ROIs and energy bins are combined later as part of the calibration pipeline. The calibration is described in section 9.3.

1: We study influences to determine auxiliary parameters which describe these differences in section 9.1.

8.3 Padding and Sliding Window

We perform a blind search for transient signals. Thus, we do not know a priori at which of the τ_p steps of the series $\mathcal{S}_p$ we expect a transient signal. As a consequence, it is impossible to fix the position of the time series $\mathcal{S}_{RNN}$ such that the decoder position coincides with the transient signal. Instead, it can occur at any time, and the detection method has to scan through the entire period $\mathcal{S}_p$. We implement a sliding window algorithm to move $\mathcal{S}_{RNN}$ with the τ_{RNN} time steps through the initial series $\mathcal{S}_p$. At each iteration of the sliding window, the time series is moved by one time step. This way, a potential signal is guaranteed to also 'slide' through the decoder phase.

This sliding window approach is illustrated in Figure 8.6. Each horizontal grey box represents one iteration n_{slide} of the sliding window. The blue circles represent the time steps of $\mathcal{S}_p$ that are investigated. During the sliding window, each of the τ_p steps are filled once into each of the τ_{dec} decoder steps. The green rectangles represent the decoder. Each time step of $\mathcal{S}_p$ is filled τ_{dec} times into different positions in the decoder

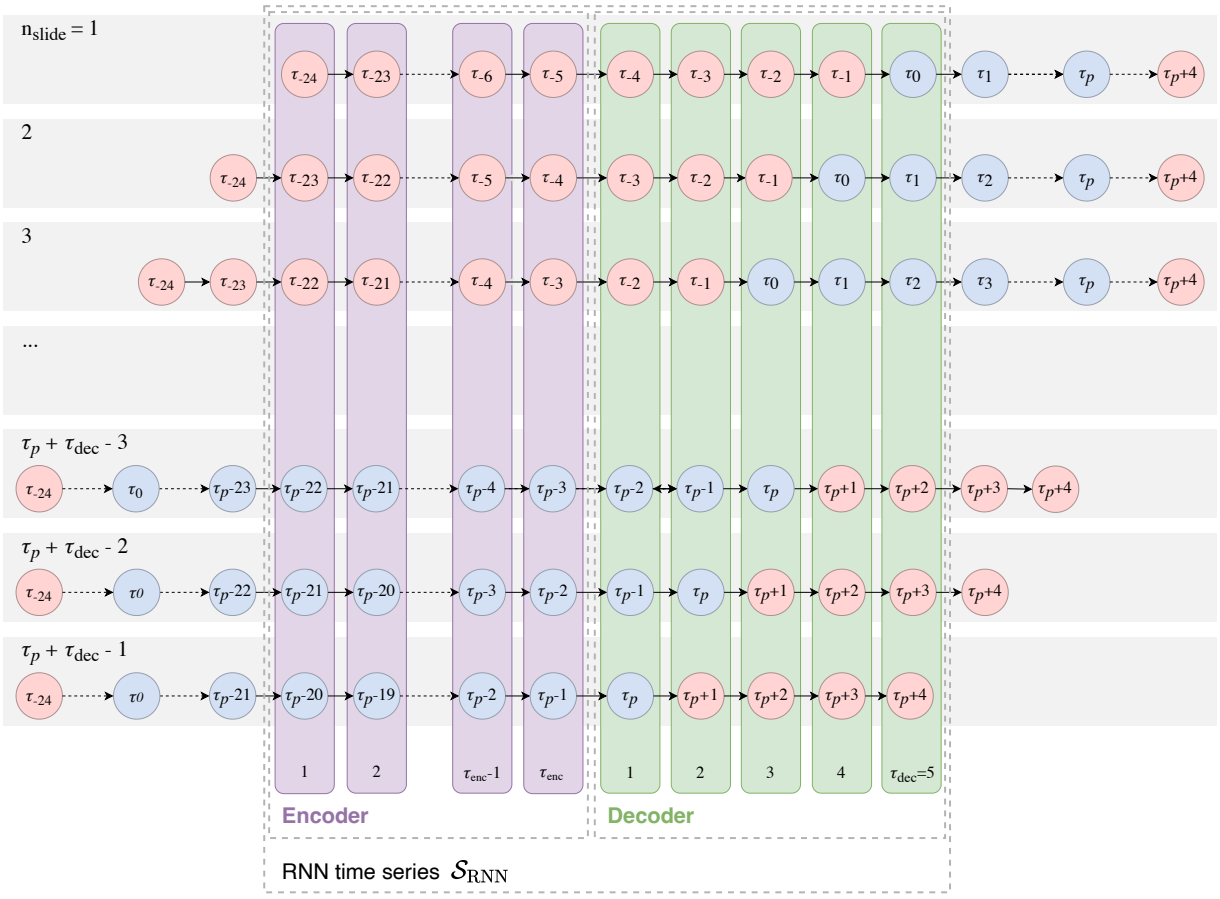


Figure 8.6: Illustration of the sliding window and padding. The blue circles represent the τ_p time steps of the investigated series $\mathcal{S}_p$. The grey boxes in the background illustrate the different iterations n_{slide} of the sliding window. The sliding window extracts parts of the time series to fill the time series $\mathcal{S}_{RNN}$. They consist of the encoder and decoder steps. The method only searches for transient signals in the decoder steps. Thus, we deploy a sliding window approach that fills each step of $\mathcal{S}_p$ into all decoder steps. We need padding data to scan through the period's beginning and end. The red circles illustrate these. The padded time series $\mathcal{S}_{p*}$ are longer than the initial series by a total of $\tau_{enc} + 2\tau_{dec} - 2$ time steps. During the first iteration, the first time step τ_0 is filled into the last decoder step, and the padding steps fill the rest. At each iteration n_{slide}, the sliding window is moved by one time step. In total $\tau_p + \tau_{dec} - 1$ iterations are performed. In the final realization, the last step τ_p is filled into the first decoder step.

phase. Thus, each time step is tested multiple times for a transient signal, increasing the trial factors. The sliding window is also applied to generate the background data for the calibration phase. Thus, the background test statistic distribution also includes the statistical fluctuations multiple times. This way, these trial factors are automatically corrected by the basic calibration approach of this analysis.[2]

To search for transient signals within the full range of $\mathcal{S}_p$, we need to ensure that each step is filled to the decoder. We generate a new time series $\mathcal{S}_{p*}$ with additional padding steps at the beginning and end of the period. In Figure 8.6, $\mathcal{S}_{p*}$ is the joint of blue and red circles. The red circles represent the padding time steps added to $\mathcal{S}_p$. This padding data corresponds to background counts under identical observing conditions. Their primary use is to fill the missing steps of $\mathcal{S}_{RNN}$. The padding at the beginning fills the encoder during the first iterations of n_{slide}. It allows starting the transient search at the beginning of the investigated period $\mathcal{S}_p$. The padding at the end allows continuing the sliding window until the last time step of $\mathcal{S}_p$.

For the padding, we randomly select events from the period $\mathcal{S}_p$. With this approach, the padding data follows the same energy spectrum as the

2: Nonetheless, we need to correct for other trial factors in this analysis. See more details on their treatment during calibration in section 9.3.

initial data. The period potentially can include transient signals which could conflict with the background hypothesis. Thus, to generate the padding data, we scramble the timestamps of all events within the period. In this thesis, we search for PBHs, for which the expected signals are in the duration of few seconds. The minimum allowed duration of the period is 5 min. Thus, these potential signals is smeared out during the shuffling. After the shuffling, we group the event into time bins of length T_{step} and append these to the $\mathcal{S}_{\text{p}}$.
We add $\tau_{\text{enc}} + \tau_{\text{dec}} - 1$ time steps at the beginning of each time period. For the first iteration $n_{\text{slide}} = 1$ of the sliding window, the first time step of $\mathcal{S}_{\text{p}}$ coincidences with the last decoder step. All other $\tau_{\text{enc}} + \tau_{\text{dec}} - 1$ steps of $\mathcal{S}_{\text{RNN}}$ are filled with the padding steps. During further iterations, $\mathcal{S}_{\text{RNN}}$ covers different parts of the series. In the last sliding window, the last step of $\mathcal{S}_{\text{p}}$ is filled into the first decoder step. Thus, we pad the time series with $\tau_{\text{dec}} - 1$ time steps at the end. The padded series $\mathcal{S}_{\text{p*}}$ has more steps than the initial series $\mathcal{S}_{\text{p}}$

$$\tau_{\text{p*}} = \tau_{\text{p}} + \tau_{\text{enc}} + 2\tau_{\text{dec}} - 2. \tag{8.4}$$

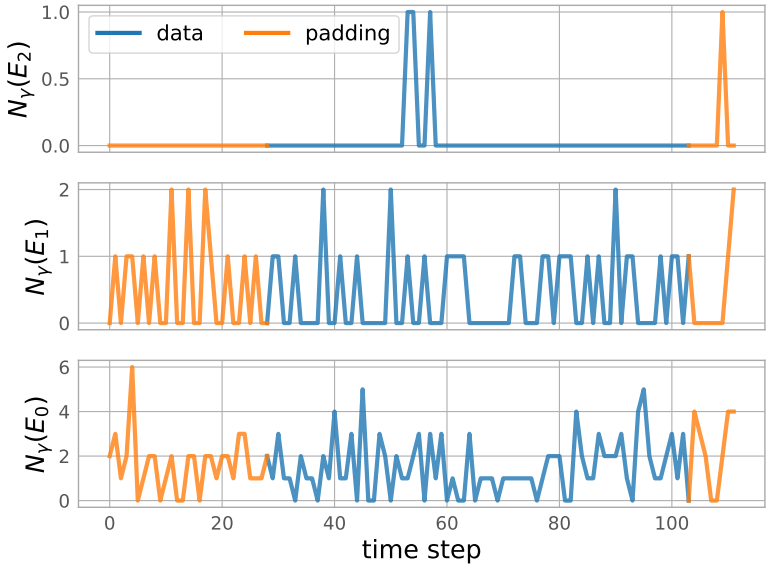


Figure 8.7: An example of the padded series $\mathcal{S}_{\text{p*}}$ in the three energy bins within one ROI. From bottom to top, the energy bin increases. The blue graphs illustrate the period with recorded events. The transient signals are searched within all of this period. The orange time steps are added as padding data. With this, the decoder steps of sliding window series $\mathcal{S}_{\text{RNN}}$ covers all recorded time steps.

We show examples of the padded series $\mathcal{S}_{\text{p*}}$ in Figure 8.7. Illustrated are the number of event counts in the three energy bins for one selected ROI. The blue curves show the recorded time steps, and the orange highlights the padding data. The event counts during the padding steps follow a Poisson distribution with the mean rate λ of $\mathcal{S}_{\text{p}}$

$$\lambda(\eta) = \frac{1}{\tau_{\text{p}}} \sum_{n=1}^{\tau_{\text{period}}} N_{\gamma,n}(\eta), \tag{8.5}$$

where $N_\gamma(\eta)$ are the counts of γ-like events for each energy bin η. We have selected the data such that the intra-run change of the rate of γ-like are expected to be modest. Thus, ignoring possible transient signals, the counts in $\mathcal{S}_{\text{p}}$ follow roughly the same distributions.

8.4 Shuffling

As mentioned in section 5.3, the training and calibration run phases require a dataset without transient signals. In this work, we use actual observational data for these phases. It enables the input data generation without modeling the instrument response function. This approach does not have the risk of a data-simulation mismatch. However, the real data can contain transient signals, which contaminate the training data. A transient signal causes a temporal increase in the rate of detected γ-like events. Scrambling the timestamps of the events can flatten out these signals. We use this concept to generate the background dataset.
The background rates are also affected by numerous effects, which can change the expected rates on $\mathcal{O}(\text{min})$. The shuffling approach needs to maintain these temporal features. Otherwise, it would create a mismatch between the shuffled and unshuflled data. Particularly, the calibration phase would be affected by differences in the distributions.
We use an approach similar to the sliding window to keep the time-dependent features throughout the shuffling. We define a shuffling interval covering τ_{shuffle} time steps. In this thesis, we set $\tau_{\text{shuffle}} = 3\tau_{\text{dec}} = 15$. This corresponds to a window duration of 300 s. For the science case of PBHs, the most significant signal is expected only in the last few seconds to the final burst. We investigate the efficiency of removing transient signals in Figure 8.8.
The shuffling window is sliding through the padded series $\mathcal{S}_{\text{p}*}$. For each iteration $i \in \{0, \ldots, \tau_{\text{p}*}\}$, we select the events within the time steps τ_i and $\tau_i + \tau_{\text{shuffle}}$. We shuffle the timestamps for these events using a uniform distribution within the shuffle interval. A rebinning in time assigns them to their new time steps. The counts in this auxiliary time series $\mathcal{S}_{\text{shuffle},i}$ follow a Poisson distribution with the mean rate λ_i of the shuffle interval.

$$\lambda_i = \frac{1}{\tau_{\text{shuffle}}} \sum_{\text{n}=\tau_i}^{\tau_i+\tau_{\text{shuffle}}} N_{\gamma,\text{n}} \ , \tag{8.6}$$

where $N_{\gamma,\text{n}}$ are the number of counts at time step n. We randomly select one of the time steps of $\mathcal{S}_{\text{shuffle}}$. The events of this interval are chosen as the new events of time step τ_i.
The shuffling of each time step i we require a total of τ_{shuffle} steps. To be able to shuffle all timesteps, an additional $\tau_{\text{shuffle}} - 1$ padding steps are required. These are temporarely added to $\mathcal{S}_{\text{p}*}$. After shuffling, the initial length of $\mathcal{S}_{\text{p}*}$ is restored.
The window size τ_{shuffle} defines how effectively transient signals are removed from the input data. The larger the window, the more a potential transient signal is smeared. However, also more of the background-specific features are removed. Thus, the window should be selected as small as possible to remove a significant contribution of the transient signal. For PBHs, the bulk of the evaporation signal is expected within a few seconds. Having a 300 s long shuffle window, is able to smear these short signals. Further, the $\mathcal{O}(\text{min})$ changes of the background rates are largely unfazed.
In Figure 8.8, we investigate the efficiency of the shuffle algorithm in removing transient signals. We perform a toy simulation assuming constant Poisson distributions for the background and signal counts

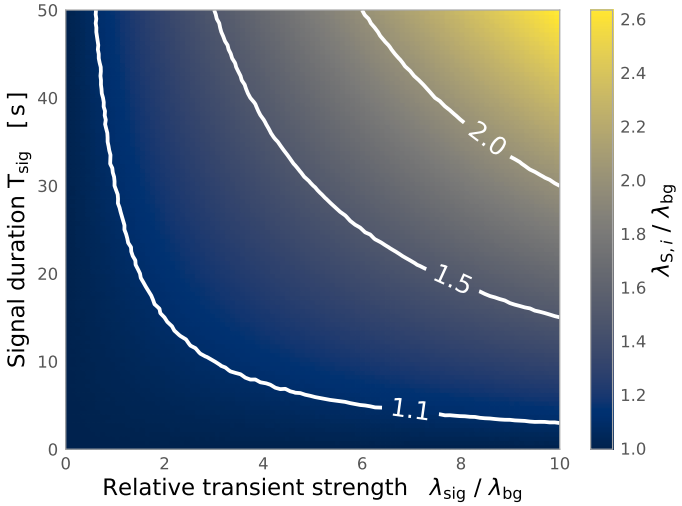


Figure 8.8: The efficiency of the shuffling algorithm for toy simulation. We simulate constant Poisson distributions for the background and the source with different signal durations and strengths. The colors give the relative increase of the shuffled counts to the background counts $\lambda_{S,i}/\lambda_{bg}$.

with means λ_{sig} and λ_{bg}, respectively. The x-axis shows the strength of the transient signal relative to the background $\lambda_{sig}/\lambda_{bg}$. We also alter the duration of the signal T_{sig} relative to the duration of the shuffling window $T_{shuffle} = 300\,s$. The duration of the signal is shown on the y-axis. The average counts after the shuffling are

$$\lambda_{S,i} = \lambda_{bg} + \frac{T_{sig}}{T_{shuffle}} \frac{\lambda_{sig}}{\lambda_{bg}} \tag{8.7}$$

The contribution of physical sources depends on the source's time profile, energy spectrum, and the instrument's response. Simulations of the transient source are required to estimate it precisely. A detailed study of this effect are beyond the scope of this thesis.

We show the results of the shuffling algorithm in Figure 8.9. The initial data is the identical with graph in Figure 8.7. The dashed grey lines display this unscrambled data. As the shuffling window moves through the time series, the padding and actual data are shuffled together. Again, we illustrate the padded parts in orange and the observation data in blue. The shuffling is repeated 20 times. We determine the median number of counts and illustrate it using solid lines. The shaded area corresponds to the 1σ containment. As shown in Figure 8.8, fluctuations with times $< \tau_{dec}$ are in average smeared out. However, this shuffling approach keeps trends in the data on time scales of $\tau_{shuffle}$ and slower. In this example, this is most evident in the lowest energy bin. Starting at time step 60, the initial counts are mostly 0 and 1 for roughly 15 time steps. The shuffling only uses events in the interval of $\tau_{shuffle}$. Thus, the scrambled counts match closely with the actual data. The average counts of the initial series rise to $\sim$ 2 counts at time step $\sim$ 90. The shuffled dataset shows an identical increase.

The training and calibration phases require enough shuffled data covering the observations' entire phase space. In order to generate additional data, we can repeat the shuffling multiple times. This generates independent realizations of the background dataset according to the expected statistical uncertainties following the Poisson distribution. By reshuffling, we ensure a sufficient amount of background data in the less populated areas of the phase space.

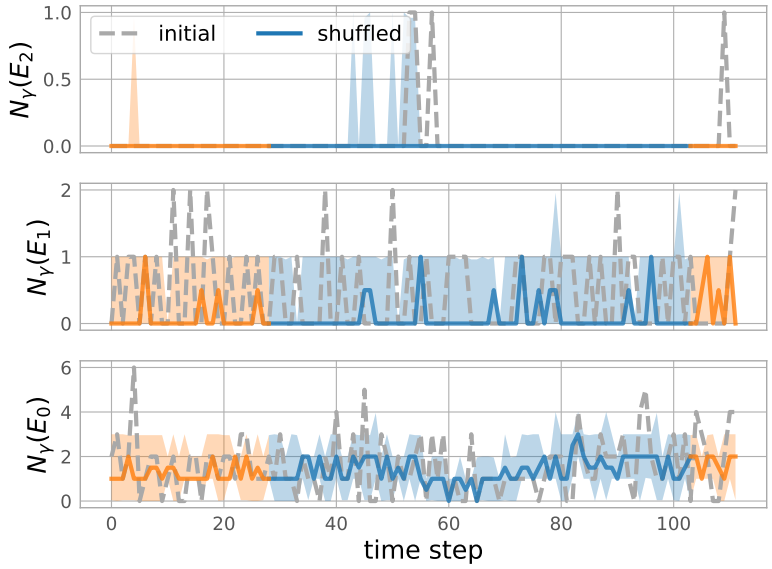


Figure 8.9: Shuffled series of event counts in one ROI for the different energy bins. Shown is the same observation run as in Figure 8.7. The dashed grey curves display the initial counts. For this graph, we repeat the shuffling 20 times. We determine the median value at each time step shown by the solid line. The shaded area shows the 1σ containment. The blue area corresponds to the recorded data, and the orange curve illustrates the padding steps.

8.5 Oversampling

We mentioned above the gaps between the circular ROIs. Figure 8.3b shows that the gaps can be even more prominent due to exclusion regions. In order to not entirely lose the events which fall in these gaps, we add an oversampling of the γ-like events. This adds every detected event multiple times converting it into a cloud of events. We add each γ-like event 20 times to the analysis. Including the oversampling for each run phase, yields consistent results.
The events have uncertainties in the reconstructed positions and energies. As discussed in Chapter 4, the position uncertainty is given by the *point spread function*. As discussed above, the energy reconstruction is associated with an uncertainty and potential bias. This is described by the *energy dispersion*. An example of these distributions is shown in Figure 4.2. They are estimated from simulated events and are stored as part of the IRFs in the DL3 files.
During the padding and shuffling phases, the γ-events are assigned following a Poisson distribution. The relative spread of the Poisson distribution decreases for larger mean values λ

$$\frac{\sigma}{\lambda} = \frac{1}{\sqrt{\lambda}}, \tag{8.8}$$

where σ is the standard deviation of the distribution. Performing the padding and shuffling after the oversampling would decrease the expected spread of the distribution of counts. Thus, the oversampling is the last step of the data preparation. All shuffled and original datasets have the same relative spread from the Poisson distribution.
During the oversampling, we vary the reconstructed event positions and energies according to the known uncertainty for each event γ. We interpolate the graphs from Figure 4.2b to calculate the radius $r_{\mathrm{PSF},\gamma}$. This radius includes 68% of events with the same energy and location as the initial event. We vary the position of the oversampled events and assume a radial symmetric probability. The shift in radius $r_{\mathrm{shift},\gamma}$ for each

event is drawn from a normal distribution

$$r_{\text{shift},\gamma} = |\mathcal{N}(0, r_{\text{PSF},\gamma})|. \tag{8.9}$$

The polar angles ϕ_γ are sampled with a uniform distribution between 0 and 2π. The sampled values are converted to the shift in the latitude and longitude of each event

$$\Delta\text{lat} = r_{\text{shift},\gamma} \cos(\phi_\gamma) \tag{8.10}$$
$$\Delta\text{lon} = r_{\text{shift},\gamma} \sin(\phi_\gamma) \tag{8.11}$$

We update the initial reconstructed position with these shifts to obtain the new positions of events. Calculating the shift in camera coordinates defined by the `SkyOffsetFrame` avoids significant uncertainties due to projection effects. From this frame, the new event positions are converted back to the origin in the ICRS.

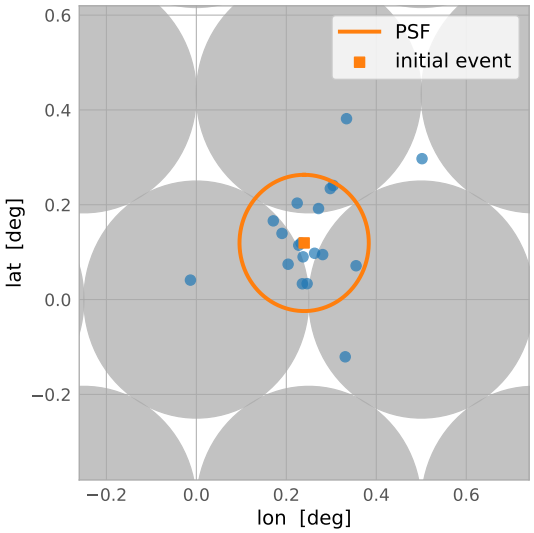


Figure 8.10: Example of the position of the oversampled events. The orange circle displays the initial event. It was recorded during an observation at 59° elevation and 286° azimuth angle. It has an energy of ~ 185 GeV and an offset of ~ 0.2678°. The 1σ uncertainty region of the reconstructed position is illustrated by the orange circle with radius ~ 0.1437°. The blue points give the position of the 19 oversampled events. The grey circles illustrate the ROIs that fill the FoV.

We show an example of the oversampled event positions in Figure 8.10. It displays a subset of the complete FoV. The grey circles show the ROIs. The orange square illustrates the initial event position. The 1σ uncertainty of the reconstructed origin is the orange circle. We oversample this event 19 times. The blue points illustrate the updated positions of these events. This oversampling approach converts each event into a cloud of events. While the initial event falls within a gap, a fraction of the event cloud is contained in the ROIs. Thus, the initial event will also partly contribute to the analysis. The probabilities of contributing to a certain ROI are distributed following the uncertainties of the reconstruction.
In order to validate the new event positions, we repeated the oversampling 10^5 times for the event shown in Figure 8.10. Its PSF is $r_{\text{PSF},\gamma} \sim 0.1437°$. We calculate the distance of the new event positions to the initial event. The blue distribution in Figure 8.11 is the distribution of distances. We expected it to follow a normal distribution with $\sigma = r_{\text{PSF},\gamma}$. The orange dashed line indicates the 1σ quantile of this distribution. As expected, it matches the value of the PSF, which is given by the solid black line.
To include the uncertainty of the energy reconstruction, we sample the energy values according to the energy dispersion of the events. Figure 4.2a shows an example of the energy dispersion. We assume that

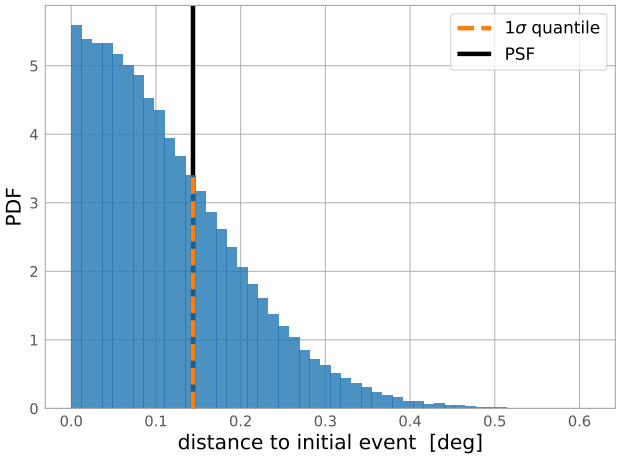


Figure 8.11: PDF of the oversampled distances from the initial position. For this plot, the oversampling for the event shown in Figure 8.10 is done by a factor 10^5. We calculate the distance of the new events to the initial one to determine the blue distribution. The orange dashed line highlights the 1σ quantile. It is equivalent to the expected value of $r_{\text{psf}} \sim 0.1437°$. The solid black line indicates the expected value from the PSF.

the reconstructed energy E_{rec} of the initial event is the actual energy E_{true}. We draw a random value $(E_{\text{rec}}/E_{\text{true}})_\gamma$ from the energy distribution for each added event γ. We evaluate it at the offset relative to the center of the FoV d_{off} and the energy $E = E_{\text{rec}}$ of the initial event. The orange curve in Figure 8.12 shows an example of $E_{\text{rec}}/E_{\text{true}}$ for an individual event. The initial energy of this event is $E_{\text{rec}} = 0.08\,\text{TeV}$ and its offset is $d_{\text{off}} = 0°$. The blue histogram shows the values sampled from this distribution.

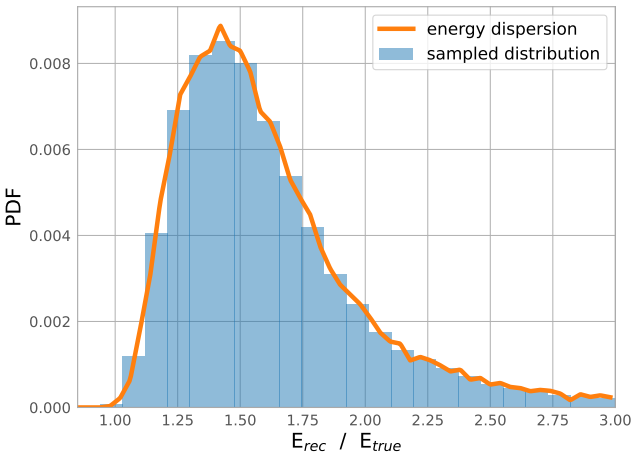


Figure 8.12: The orange curve shows the energy dispersion at $E_{\text{true}} = 8\,\text{GeV}$ for a run with 79° elevation and 146° azimuth. It is the vertical slice of the interpolated values of Figure 4.2a at the given energy. We sample this distribution 10^5 times for validation. The blue histogram shows the distribution of the sampled $(E_{\text{rec}}/E_{\text{true}})_\gamma$ values.

For the example event of 8 GeV, the reconstruction has a significant energy bias b_γ. Directly using the sampled distribution with bias would shift the new energy. The bias must be corrected to keep the scales of the initial and new energy values comparable. We calculate b_γ at the energy of the given event. In the calculation of the new energy of the event $E_{\text{rec, new}}$, the bias in the denominator

$$E_{\text{rec,new}} = E_{\text{rec}} \frac{(E_{\text{rec}}/E_{\text{true}})_\gamma}{b_\gamma}. \tag{8.12}$$

This formula shifts the energies according to the energy dispersion. However, the contribution by the bias is factored out. We show the distribution of $E_{\text{rec, new}}$ for the example event with $E_{\text{rec}} = 0.08\,\text{TeV}$ in Figure 8.13. As expected the values of $E_{\text{rec, new}}$ vary around E_{rec}. With the bias-correction, the median of the new energies is equivalent with the initial energy.

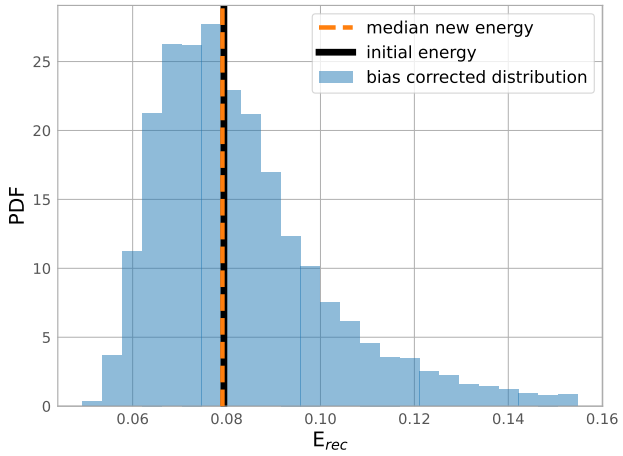


Figure 8.13: For the example of Figure 8.12, we calculate the new energies using Equation 8.12. The median of the distribution of $E_{\text{rec, new}}$ (orange dashed line) matches the energy of the initial energy (solid black line).

Statistically, the event clouds are equivalent to working with event densities. Each event represents a probability density to belong to a certain ROI and energy bin. This approach can also help to increase the statistics and thus has a stabilizing effect on the analysis.

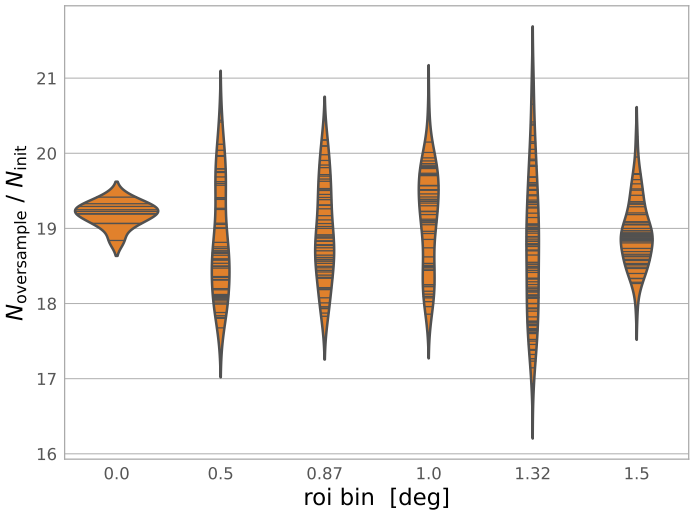


Figure 8.14: Distribution of oversampling factor $N_{\text{oversample}}/N_{\text{init}}$ in each ROI. We group the ROIs by their distance to the camera center. Each violin plot in the horizontal axis represents one group of ROIs. Each horizontal grey line illustrates the ratio obtained from one iteration of oversampling. The underlying distributions of each group are estimated using a kernel density estimate.

The oversampling increases the number of total events by a factor of 19. On average, we expect the counts of events in each ROI to be increased by this factor. In Figure 8.14, we validate the oversampling for a run without exclusion regions. The ROIs are defined as illustrated in Figure 8.3a. We group the ROIs based on their distance to the camera center. In total, we have six groups illustrated independently in the graph. We repeat the oversampling ten times to increase the statistic. For each ROI, we calculate the ratio of the number of oversampled to initial events $N_{\text{oversample}}/N_{\text{init}}$. As the positions are varied randomly according to the PSF, each iteration yields an independent ratio. The horizontal grey lines in the distributions of each group show the obtained ratios. We estimate the underlying distribution using a kernel density estimate. To do this, we determine the bandwith from Scotts rule [3].

The distributions vary around the expected ratio of 19 for each group. The variances, though, increase further away from the center. The reason for this is in the approximately radial symmetric acceptance. Only in

3: see also Equation 6.3.2

the camera center rings around the ROI have the same acceptance. For other ROIs, it is higher in the direction of the center and smaller on the opposite side. Thus, the event rates around the ROIs vary more significantly. The size of the ring from which events can fall into each ROI during oversampling depends on the PSF size. Typically, the PSF increases further away from the center. Thus, the ring size also increases. Eventually, this leads to the increased variances of these distributions.

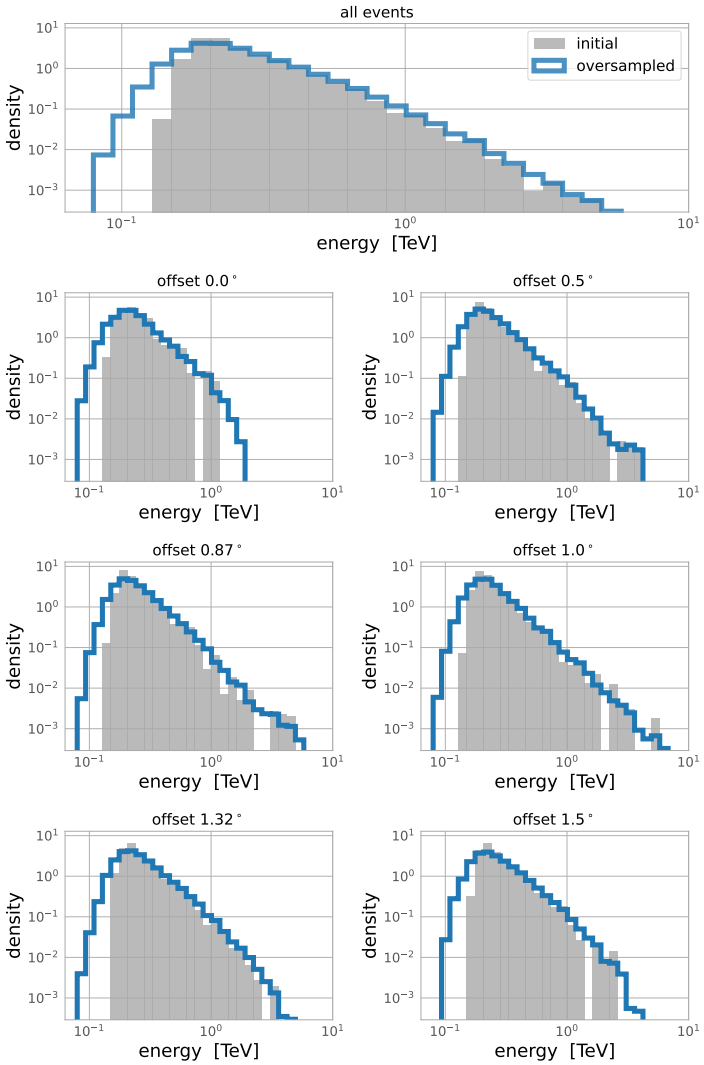


Figure 8.15: Energy spectrum of initial and oversampled events. The first row includes all events within the FoV. The other subplots show the distribution of the events within the ROIs at a given offset to the camera center.

In Figure 8.15 we compare the normalized energy spectra of the oversampled (blue) and initial (grey) events. Following the energy dispersion, the energies of the oversampled events can be higher or lower. Thus, the overall range of the spectra increases by including these uncertainties. Above the peak of the distribution, the spectra follow roughly a power-law spectrum. We expect more events from the most populated energy bins falling in the less populated bins than in the opposite direction. Consequently, we expect a harder power-law spectrum due to the oversampling. In the top row of Figure 8.15 we show the spectra for all events within the FoV. It proves that the overall shift of the power-law index is a minor effect. It also reveals the effect of the increased energy range at low energies.[4]
The other graphs show the spectra for individual ROIs at different offsets

4: As the oversampling affects every stage of the analysis, it is self-consistent. The oversampling is only applied for identifying significant transient signals. In case of a detection, the physical properties are investigated with traditional data analysis tools. These, use the original data without oversampling.

to the camera center. We notice the effects of the increased statistic with the oversampling. With the included energy dispersion, the spectrum does not contain any gaps.

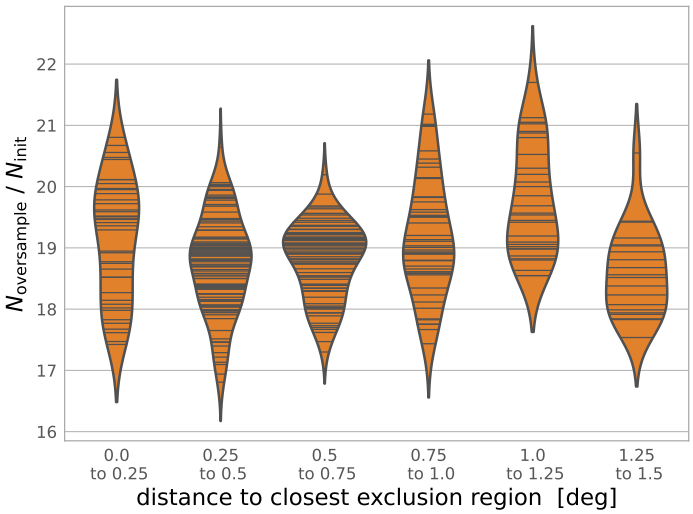


Figure 8.16: The number of oversampled events in ROIs as a function of the distance to the exclusion region. We group ROIs based on their distances to the closest exclusion region.

The smearing, according to the uncertainties, can lead to edge effects. We discussed the widening of the energy spectra in the previous paragraph. Another edge case is the proximity of exclusion regions. Events within these regions are removed before the oversampling takes place. Thus, no events from these areas can end up in one of the ROIs. However, the oversampled position of events can fall within an exclusion region when the initial location is close to it. We remove these oversampled events from the analysis. This edge effect can cause a decreased oversampled number of counts for ROIs in proximity to an exclusion region. We study this effect in Figure 8.16. In contrast to the previous discussion, we investigate the FoV with exclusion region, which we have displayed in Figure 8.3b. The ROIs are grouped based on the distance to the closest exclusion region. We do not find a significant decrease in the oversampled ratio close to the exclusion regions. Thus, this oversampling approach is also stable for FoVs with different exclusion regions.

8.6 Preparation of Simulations

Chapter 7 introduced the simulation of the PBH evaporation signals. We can superimpose it to a background dataset to generate the expected observable signal. The background is generated from the actual observations using the shuffling from section 8.4. The injection of the signal is done in two phases. First, the simulated signal is prepared into series $\mathcal{S}_{\text{sim}}$ in the phase *prep_sim*. Later, these $\mathcal{S}_{\text{sim}}$ are superimposed to background series during *inject_sim*.
In the first phase, only deal with the simulated signals. We combine the events from the seven individually simulated time windows. It yields the simulated γ-rays within the duration $T_{\text{sim}} = 100\,\text{s}$ until the final evaporation of PBHs. Each burst is shifted by a random time between 0 s and $T_{\text{step}} = 20\,\text{s}$. It accounts for the serendipitous times of the burst relative to the defined timesteps. We also rotate all individual γ-rays

around the center of the FoV by an angle $\phi \in [0, 2\pi)$ assuming a radial symmetric sensitivity of VERITAS.[5] This represents that the burst can also have a random location relative to the ROIs. We then use the same methods introduced above to group the events in energy, time, and ROIs. In this phase, where we only deal with simulated signals, we consider all 37 ROIs without exclusion regions. As we look at 100 s long signals, the resulting series $\mathcal{S}_{\text{sim}}$ have $\tau_{\text{sim}} = 5$ steps.

In subsection 9.3.2, we group the observation based on similar data characteristics. We select a random observing run from the same meta bin for signal injection for each $\mathcal{S}_{\text{sim}}$. In this way, the observing conditions if the simulated signal and background are matched. Further, we randomly select a timestep from the background data to which the signal is injected. We then add the binned counts of $\mathcal{S}_{\text{sim}}$ to the selected background time. The background data contains the nominal definition of exclusion regions. ROIs, which are excluded in the background, are also ignored during the injection. Thus, the configuration used for the injected data represents the actual dataset.

5: For technical reasons, the simulations at each offset ζ are performed with the same polar angle $\phi = 0$. The VERITAS IRFs are radial symmetric as well. In contrast to simulating with a random ϕ, the only difference is the location of the γ-rays. After the rotation during the data preparation, there is no systematic difference.

9 Training and Calibration

VERITAS detects an inevitable rate of background events passing the nominal background selection. These are mainly γ-like cosmic-ray showers, electrons, and positrons. In the direction of the galactic plane, we expect an additional component of diffuse γ-rays. The background rate depends on the observing conditions which impact the instrument response. These effects can lead to varying rates between individual runs but also change the rates within a single observation. To model these influences, we add auxiliary parameters to the inputs to the RNN. These describe the underlying differences between observations.
In section 9.1, we study the effects of different observables to the background rates. This work informs us which auxiliary parameters provide crucial inputs. We describe the training of the RNN in section 9.2. With this results, we are able to estimate the expected background counts. The calibration approach discussed in section 9.3 allows interpretation of the results with a statistical significance.

9.1 Auxiliary Parameters

In traditional analyses, the instrument characteristics are described by the instrument response functions, including effective areas and radial acceptances. These allow the interpretation of observations in terms of physical parameters, such as fluxes and energy spectra. This deep-learning-based transient detection approach has the advantage that it does not rely on the modelling of instrument response. With suited auxiliary parameters, the RNN can learn the instrument response under different conditions from the observational data itself. These parameters describe the effects which influence the detected event rates. During the training, the RNN generates an abstract description of the inputs.
In subsection 9.1.1, we discuss the calculation of the auxiliary parameters as part of the data preparation. Subsequently, we investigate which auxiliary parameters provide crucial information for the RNN. We subdivide this study into two parts. In subsection 9.1.2, we investigate the runwise background event rates. Due to the large differences of observing conditions, we expect these to reveal the most important effects. Subsequently, we investigate the changes of rates within the individual runs in subsection 9.1.3. This informs the decision on the auxiliary parameters which are required to model the intra-run rate changes.

9.1.1 Calculation of Auxiliary Parameters

We deal with different types of auxiliary parameters. We have the information that is constant for the complete sequence δ_{RNN} and data that changes for each of the τ_{RNN} steps in the sequence. An instance of the former is the offset of the ROI to the camera center. The series δ_{RNN} are assigned to one ROI and thus this value is constant. An example of

the latter is the zenith angle of observation that changes in the course of a single observing run. As discussed above, the same long short term memory cell is representing all time steps. Thus, they also require the identical inputs. Even a constant parameter is internally passed to each step.
When dealing with variable data within $\mathcal{S}_{\text{RNN}}$, we discriminate between two types of data. The first is information that is assigned to a specific point in time. For instance, this are the time series stored in the data base such as the zenith and azimuthal angle of observation, or the L3 trigger rate. As the steps τ of $\mathcal{S}_{\text{RNN}}$ represent one specific 20 s long interval in time, we can assign each step with the corresponding mean value. However, for the padded time steps these data do either not exist or are not meaningful. Thus, we use a linear extrapolation of the existing time steps to assign the values to the padded data. As these parameters directly describe a point in time, they are not affected by the shuffling.
The second data type is assigned to each individual recorded γ-like event. An example for this type is the multiplicity of the events. These parameters are assigned to a given time through the timestamps of the events. The auxiliary parameter $p(\tau)$ at step τ is calculated as the average over the values p_i for all events that are assigned to τ. During padding and shuffling, the γ-like events are reassigned to different time bins. Thus, this procedure works identically in these cases.

9.1.2 Runwise Event Rates

For this study we select all 5725 training runs. We also apply the identical time cuts and exclusion regions as used during the analysis. The remaining events of this dataset are dominated by the background. We calculate the average background rate for each run, correcting for the excluded fractions as described by Equation 8.1 and Equation 8.3. The rates are normalized to the integrated solid angles. Unless stated differently, we refer to the average rate within the inner 1.75° of the FoV.

Table 9.1: Overview of the potential auxiliary parameters. More details to these variables are given in the corresponding paragraphs below.

Parameter	Description
$\sec(\theta)$	Secant of zenith angle of observation
$\lvert\alpha\rvert$	Azimuthal angle observation
ΔT_{ref}	Years after reference point 2012.08.01
l, b	Galactic coordinates
R_{L3}	L3 trigger rate
`dead_time`	Fraction of time with busy data acquisition
`ped_var`	Average pedestal variance for complete run
$M(\text{E}_{\text{total}})$	Average multiplicity in total energy range
$M(\text{E}_0)$, $M(\text{E}_1)$, $M(\text{E}_2)$	Average multiplicity per energy bin
ζ	Offset angle relative to center of FoV

Below we study the dependency of the runwise background rates on potential auxiliary parameters. Table 9.1 shows the candidates we investigate. The goal is to identify variables with a strong relationship to the target background rates. The Spearman's rank r_s coefficient is a measure of how well two variables are described by a monotonic function [134]. A value of 1 corresponds to a perfect correlation, −1 to anti-correlation, and 0 to no correlation between the parameters.
We also aim to identify linearly correlated auxiliary parameters. Their

[134]: Spearman (1904), 'The Proof and Measurement of Association Between Two Things'

information are redundant to the RNN. However, they increase the total dimension of the in- and outputs and thus the complexity of the model. This possibly can harm the robustness of the DNN [144]. We use the Pearson correlation coefficient r_p as measure of the linear correlation between two variables [145].

[144]: Ortiz Jimenez et al. (2020), 'Redundant features can hurt robustness to distributions shift'
[145]: Pearson (1895), 'Note on Regression and Inheritance in the Case of Two Parents'

Zenith Angle or Observation

The zenith angle of observation θ is among the best known parameter to change the expected background rate. For a pointing closer to the horizon, the propagation path through the atmosphere increases. This length is described by

$$\sec(\theta) = \frac{1}{\cos(\theta)}, \tag{9.1}$$

where θ is the zenith angle of observation.

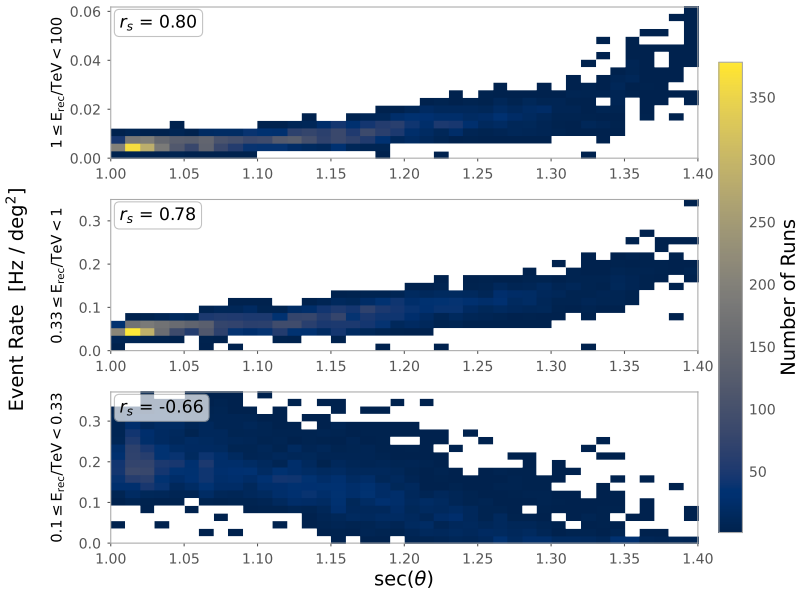


Figure 9.1: Correlation of γ-like event rates with $\sec(\theta)$ for all considered runs.

We show the dependency between the background event rate and the $\sec(\theta)$ in Figure 9.1. Each row represents one of the energy bins considered in this analysis. In the lowest energy bin, $E_{\text{rec}} < 0.33$ TeV, the event rate decreases for larger $\sec(\theta)$. The Spearman's rank coefficient of $r_s = -0.66$ confirms this. The relatively weak signals of faint air showers are less probable to be detected with the larger propagation paths increasing the lower energy threshold of the instrument. For the observations with $\sec(\theta) \gtrsim 1.35$, the event rate typically drops to ~ 0. For the two higher energy bins, we find the opposite trend with $r_s = 0.78$ and $r_s = 0.8$, respectively. The rates here increase by up to a factor ~ 6. For events at these energies, the Cherenkov signals are stronger and able to reach the ground level. In this case, projection effects increase the collection area causing higher rates. Due to the significant changes, $\sec(\theta)$ is an important auxiliary parameter.
Below we are particularly interested in the influences by other observable parameters. It is difficult to investigate less significant effects for the complete dataset. Thus, we group all observations into three bins $\sec(\theta) \leq$

1.05, $1.05 < \sec(\theta) \leq 1.15$, and $1.15 < \sec(\theta) \leq 1.4$. For the 1807 runs in the first bin, the rate differences due to the zenith angle of observation are only marginal. This allows identifying subdominant influences by other parameters. However, all systematic effects need to be included into this study. So if necessary, we also investigate the groups of 1937 and 1981 runs for the second and third $\sec(\theta)$ bins, respectively.

Azimuthal Angle of Observation

The magnetic field of the earth impacts the lateral development of atmospheric air showers. Images of IACTs are sensitive to these distortions [146]. Especially differences between pointing north ($\alpha = 0°$) and south ($\alpha = 180°$) are expected. The differences between the pointing east ($\alpha = 90°$) or west ($\alpha = 270°$) are negligible. For simplification, we transform the azimuth by first wrapping the angle at 180°. In this notation the azimuth is mapped from −180° to 180°. We take the absolute value of this angle which yields $|\alpha|$. It describes the distance from pointing north independent of the rotation direction. We show a mapping of $|\alpha|$ to the initial azimuth in Table 9.2. Contrary to α, we expect a monotonic relationship of $|\alpha|$ to the background rates

[146]: Chadwick et al. (1999), 'Geomagnetic effects on atmospheric Cherenkov images'

Table 9.2: Mapping of $|\alpha|$ to the pointing direction.

$\lvert\alpha\rvert$	α	direction
0°	0°	*north*
90°	90°, 270°	*east*, *west*
180°	180°	*south*

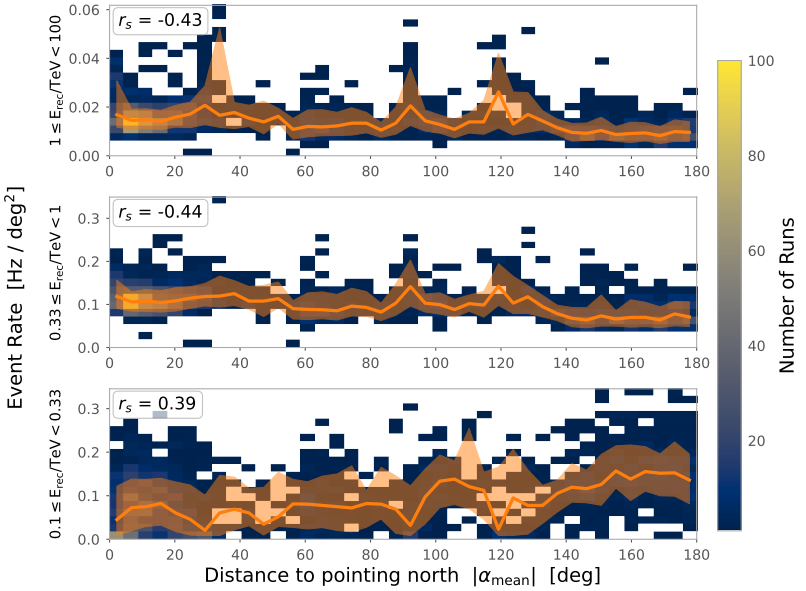


Figure 9.2: Correlation of γ-like event rates with the mean azimuth angle for runs with $1.15 < \sec(\theta) \leq 1.4$.

The effects of the azimuth angle are more significant for observations at higher θ. For runs with $\sec(\theta) \leq 1.05$, the maximum correlation with the background rates in any energy bin is $|r_s| \leq 0.19$. In Figure 9.2, we show the background rate as a function of $|\alpha|$ for the group of runs at $1.15 < \sec(\theta) \leq 1.4$. We also calculate the medians and 1σ-areas and illustrate these in orange. For this group, we find $|r_s| \sim 0.4$ for all three energy bins. This is a hint for a possible dependency. It also confirms the expectation that the largest differences are evident between pointing north $\alpha \sim 0°$ and south $\alpha \sim 180°$. The changes in the event rates are roughly a factor ~ 2. In the lowest energy bin, the rates increase when pointing south. However, in the two higher energy bins, we observe the inverse trend. This can be explained by the variation of the primary energy depending on the pointing direction [147]. Due to the identified effects, we use the $|\alpha|$ as auxiliary parameter to the RNN.

[147]: Ivanov (1999), 'Azimuthal effect on extensive air showers of cosmic rays'

Time

We discussed the influences of aging effects on the VERITAS sensitivity in subsection 4.2.6. Especially, the loss of faint events is inevitable as the quantum efficiency of PMTs and the mirror reflectivity decreases [80]. The impact is strongest for low energy events. We show the distribution of average rates for the group of runs with $\sec(\theta) \leq 1.05$ in Figure 9.3. As reference time, we use 01/08/2012 which is before the first observation considered in this work. We name the difference of the time of observation and this reference ΔT_{ref}.

[80]: Adams et al. (2022), 'The throughput calibration of the VERITAS telescopes'

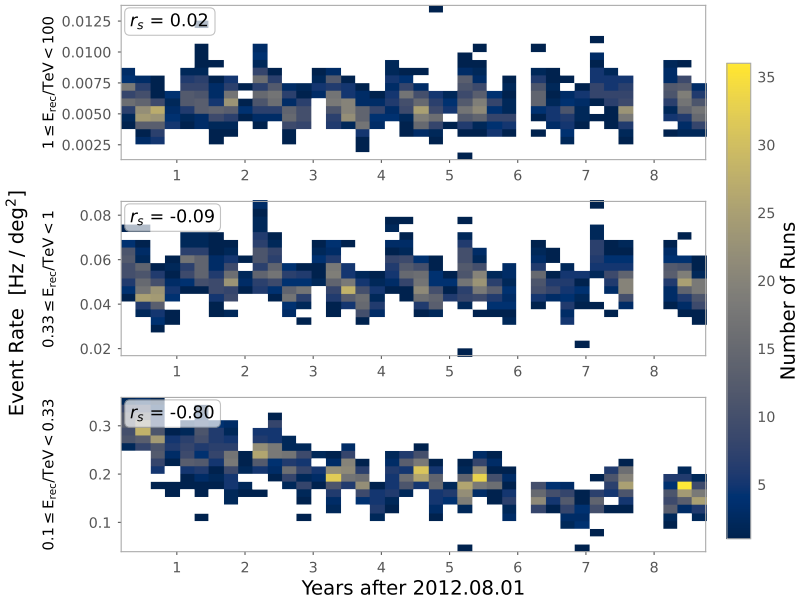


Figure 9.3: Correlation of γ-like event rate with the reference time for runs with $\sec(\theta) \leq 1.05$.

A correlation of $r_s = -0.8$ in the lowest energy bin confirms the dominating effect of the long-term degeneration of the instrument. Additionally, we find periodic structures affecting all energies with a period of one year. The summer in Arizona is monsoon time. During this period no observations are taken. We define one season as one year of observations starting in August. At the beginning of the seasons, the amplifier gains for the flash analog-to-digital converter are tuned.[1] This also affects the raw rate of detected γ-like events. During each season, the atmospheric conditions change.[2] The intra-season effect on the rate is at the level of up to $\sim$ 33% for each season. Using ΔT_{ref} as a floating point number allows the RNN to also extract these periodic features. However, this dependency is not trivial. Below, we identify the L3 rate as an additional parameter that can help describing the intra-season changes. We study this in more depth in Figure 9.8.

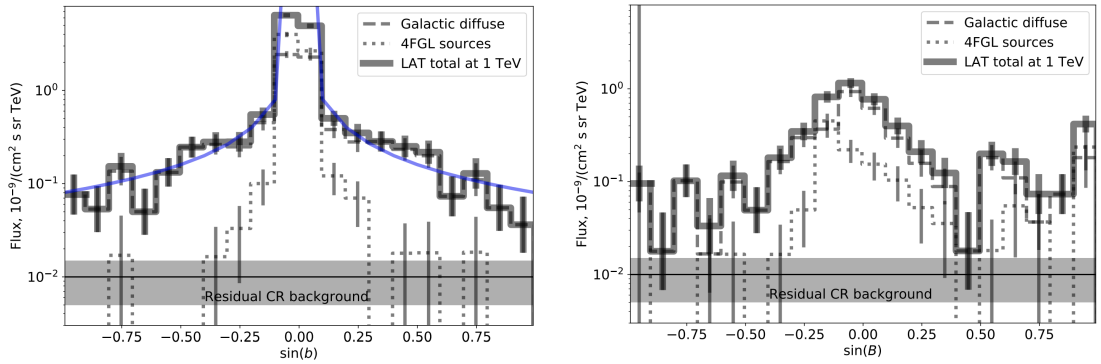
1: See subsection 4.1.3.

2: In the standard VERITAS data analysis, this effect is considered by using independent sets of IRFs for summer and winter.

Galactic Coordinates

Diffuse γ-rays from the Milky Way contribute to the total background. Up to 300 GeV, the spectral shape of the diffuse γ-ray at the Galactic ridge ($|l| < 30°$) differs from the outer Galactic plane ($150 < l < 210°$) [149]. In the TeV range, however, the spectral slopes are similar and only their scale varies. Figure 9.4 shows the flux profiles at 1 TeV for the inner and outer regions of the Galactic plane. The strongest diffuse component

[149]: Neronov et al. (2020), 'Galactic diffuse gamma-ray emission at TeV energy'



(a) Flux profile of the inner Galactic plane $-60° < l < 60°$. The blue curves is the expected emission from a disc with a constant thickness.

(b) Flux profile of the outer Galactic plane $120° < l < 240°$.

Figure 9.4: Galactic latitude flux profiles at 1 TeV of the inner and outer Galactic plane. The Dashed and dotted lines illustrate the galactic diffuse and the resolved 4FGL catalog [148] source flux components. The solid lines are the total flux observed flux. The grey band shows of the residual cosmic ray background level. [149].

comes from Galactic plane when pointing at $|l| < 60°$ and $|b| < 5°$ [150]. From all processed VERITAS runs, we find 91 runs within the Galactic plane that match all other selection criteria for training. The reason for the relatively small number of runs is the location of VERITAS in the northern hemisphere. Galactic plane observations from this position, usually have large zenith angle of observation. Thus, they typically are not considered in this thesis. The diffuse emission provides a challenging source of systematic uncertainty. As mentioned in subsection 6.3.5, we decide to exclude all Galactic plane observations from the datasets.

In future works, Galactic observations might be included which would allow the detection of Galactic transient sources such as novae [151]. Thus, we also investigate possible effects in the background rates in the VERITAS data. The smallest Galactic longitude among the 91 runs is at $|l| \sim 30°$ with the majority being $|l| > 40°$. To study the impacts of the Galactic plane, we select 14 runs from the galactic plane observed with $\Delta T_{\text{ref}} < 1\,\text{yr}$, $|\alpha| > 140°$ and $57° < \epsilon < 63°$. In Figure 9.5, we compare their event rates to 33 training runs taken outside of the galactic plane but with the identical range of observing conditions. The underlying distributions of rates at all energy bins are estimated using a kernel density estimate. The blue and orange distributions represent the runs outside and inside the Galactic plane, respectively. The dashed lines highlight the median and dotted lines show 1σ quantiles.

[150]: Mohrmann, L. et al. (2019), 'Validation of open-source science tools and background model construction in astronomy'

[151]: null et al. (2022), 'Time-resolved hadronic particle acceleration in the recurrent nova RS Ophiuchi'

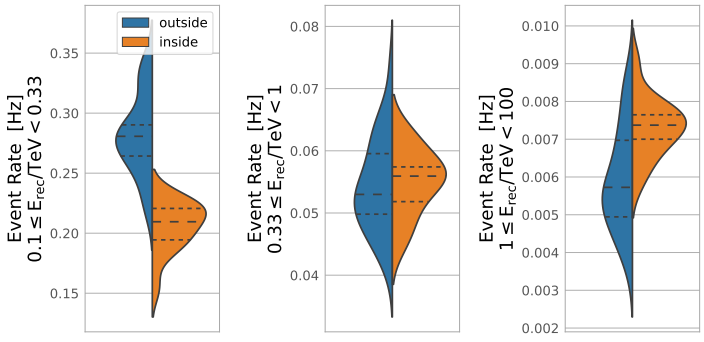


Figure 9.5: Background counts for observations in and outside of the Galactic plane.

We find a $\sim$ 20% decrease of the low energy γ-ray rate in the Galactic

plane. In the highest energy bin, we find an increase of the same order. This indicates, that the background spectra for the selected runs at the Galactic plane are harder. We note, that all selected runs from Galactic plane have the identical FoV. This might be a source of systematic difference that is not covered here. As number of VERITAS observations with sec $\theta \leq 1.4$ is small in the Galactic plane is limits. Thus, studying this dependencies in the scope of this thesis is challenging. Including these Galactic runs into this analysis would require further investigation of the systematic effects. This future work should also include observations at smaller zenith angles of observation to increase the coverage of the Galactic plane.

L3 Trigger Rate and Dead Time

The L3 trigger rate R_{L3} corresponds to the total rate of events written to disk. Whenever a trigger occurs, the data acquisition system is occupied by writing out the data. During this time, it can not record new events. For nominal VERITAS observations this so-called *dead time* is in the range of $\sim 10\%$ to $\sim 18\%$. We show the distribution of the run-wise L3 rates and dead times in Figure 9.6. The Pearson coefficient for these two variables is $r_p = -0.92$. Due to their almost perfect linear anti-correlation, adding both quantities is redundant. In the following, we investigate the L3 trigger rate as potential auxiliary parameter.

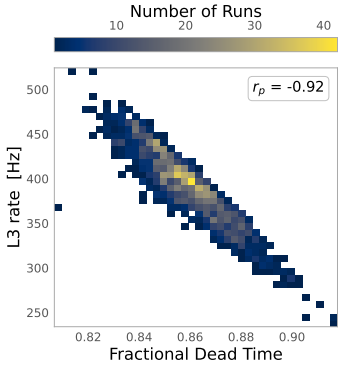


Figure 9.6: Correlation of L3 trigger rate and dead time at sec(θ) ≤ 1.05.

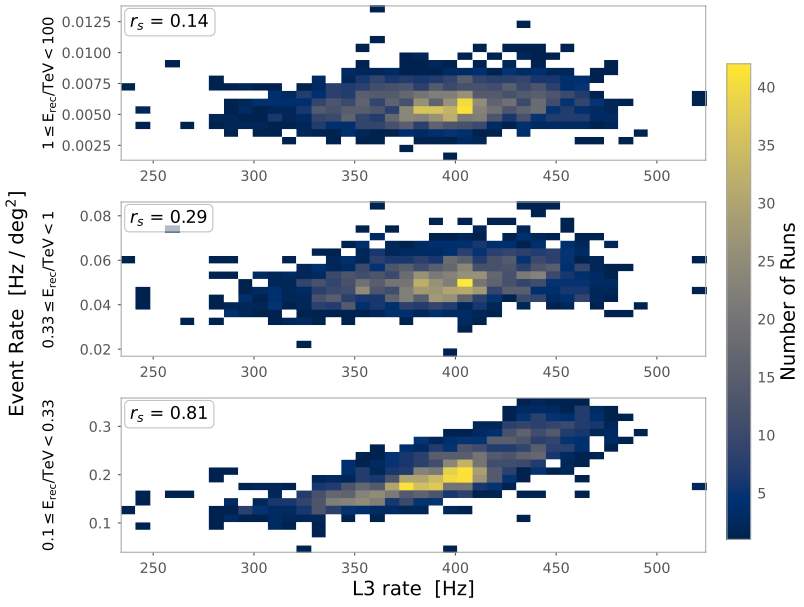


Figure 9.7: Correlation of event rates with mean L3 trigger rate for sec(θ) $\leq$ 1.05.

In subsection 4.1.3, we discussed how the VERITAS trigger system is built to suppress unwanted background. Nonetheless, an inevitable rate of background events is recorded. The L3 trigger rate is dominated by cosmic-rays and fluctuations of the NSB. These are also the most dominant contributions to the background rates. This suggests that the R_{L3} might be a suited auxiliary parameter.
Figure 9.7 shows the background event rates as a function of the L3 rate for runs with sec $\theta \leq 1.05$. A rising L3 rate is a proxy for an increased background event rate. This correlation is most significant at the lowest energy bin with $r_s = 0.81$. Above, we found similar correlations between

the event rates and ΔT_{ref}. Thus, we investigate whether L3 is redundant or provides relevant information. We focus this investigation on the lowest energy bin in which we found the strongest correlations between the rates and these two parameters.

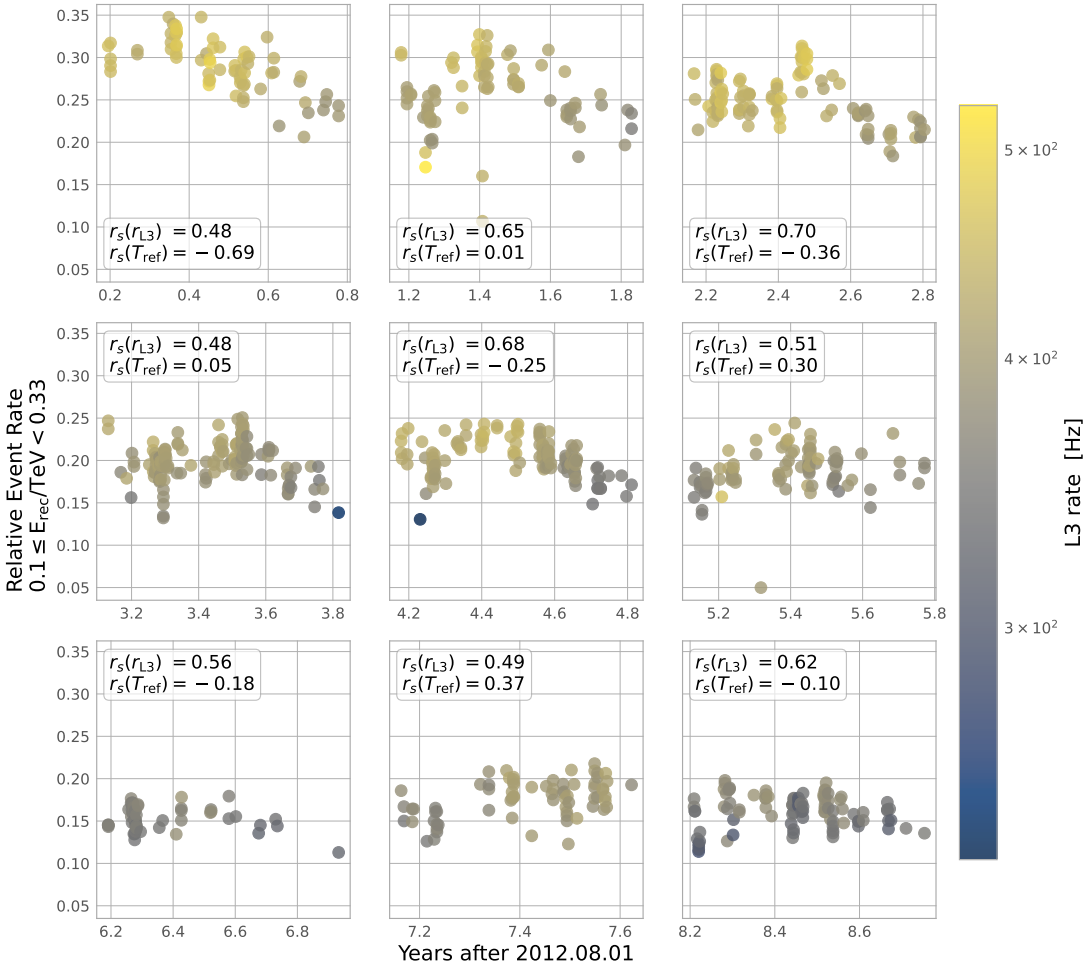


Figure 9.8: Run-wise background event rate in the first energy bin $0.1 \leq E_{rec}/TeV < 0.33$ as a function of the reference time ΔT_{ref}. Each point corresponds to an individual run. The colors denote the average L3 rate R_{L3} during each run. This plot only shows runs with $\sec(\theta) \leq 1.05$ and pointing south. Each subgraph shows one of the nine seasons that are used in this work. For each season, we calculate the correlation between the background rates with R_{L3} and ΔT_{ref} using Spearman's correlation coefficient.

Figure 9.8 shows the background rates in the first energy bin covering the range $0.1 \leq E_{rec}/TeV < 0.33$. Here we select only runs with $\sec(\theta) \leq 1.05$ and pointing south. This selection is based on the previous discussion to have a more homogeneous set of observations. Each VERITAS season is represented independently in one of the nine sub-graphs. The point locations show the background rates as a function of ΔT_{ref} for all selected runs. The colors represent the average L3 rates R_{L3} during these runs.
Particularly, we are interested whether the R_{L3} provides a refining information compared to the reference time ΔT_{ref}. For each individual season, we show values of r_s between the background rates and both variables in the corresponding subplots. The correlation with R_{L3} is in the range from $|r_s(R_{L3})| = 0.48$ up to $|r_s(R_{L3})| = 0.7$. This indicates an important correlation between R_{L3} and the target background rate. For ΔT_{ref}, the correlations are from $|r_s(\Delta T_{ref})| = 0.01$ to $|r_s(\Delta T_{ref})| = 0.69$. We find that $|r_s(R_{L3})| > |r_s(\Delta T_{ref})|$ for all but the first season. The L3 rate is a sensitive parameter to describe the intra-season changes in the

background rates in the first energy bin.
The figures for the two higher energy bins are in Figure A.1 and Figure A.2. For these energy ranges, the reference time $|r_s(\Delta T_{\text{ref}})|$ typically is higher compared to the first energy bin. Thus, the intra-season effects are better described by the time. Indeed, the correlations $|r_s(R_{\text{L3}})|$ are typically weaker than $|r_s(\Delta T_{\text{ref}})|$. ΔT_{ref} describes accurately the long-term trends for all energy bins. In the two higher energy bins it is sufficient to describe the intra-season effects. At lower energies, however, the L3 rate more precisely describes the changes of the energy threshold within the seasons. Both, ΔT_{ref} and R_{L3}, are important variables to describe the target background event rates.

Pedestal Variance

In subsection 4.2.1, we introduced the pedestal variance as a proxy parameter for the brightness of the night sky background. Figure 9.9 shows the run-wise event rate of the low zenith angle group as function of the pedestal variance. The orange lines and bands correspond to the column-wise median and 1σ-intervals, respectively. For the two high energy bins we find the expected correlation between the rates and the pedestal variance. The correlations are described with $r_s = 0.49$ and $r_s = 0.44$, respectively. Within this group of runs, the median rates increase by $\sim 50\%$ as the NSB is brighter. The larger noise component due to the brightness causes a larger component of background events. Thus, considering this effect during the training of the RNN is important.

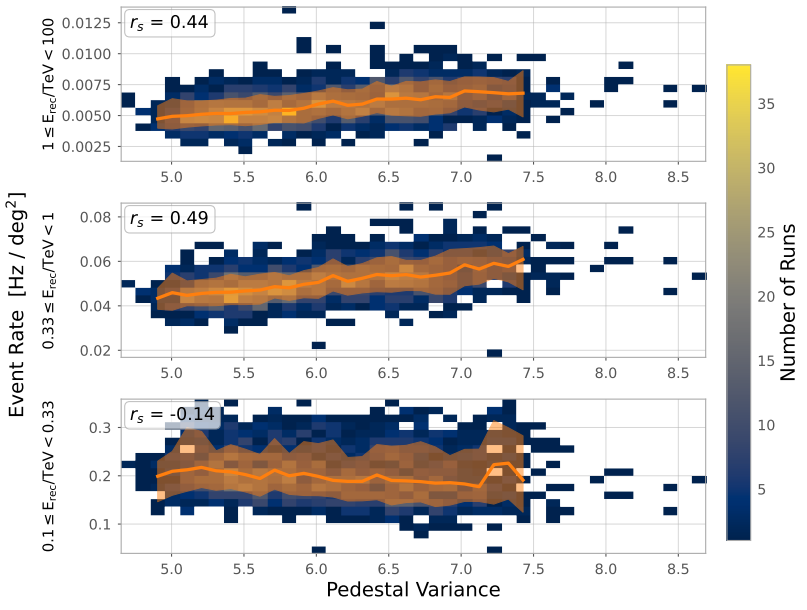


Figure 9.9: Correlation of event rates with pedestal variance for $\sec(\theta) \leq 1.05$. The orange line shows the median of each pedestal variance bin. The shaded area corresponds to the 1σ containment area.

With $r_s = -0.14$ we do not find a clear correlation in the lowest energy bin. The thresholds during the image cleaning are adjusted to the measured pedestal variance. For a bright NSB, the signals need to be significantly higher to pass the cleaning. Typically, images of faint air showers correspond to low energy events. The competing effects of overall more events due to the noise and the higher threshold overall cause a nearly stable background rate in this energy bin.

In the input DL3-files, the pedestal variance is stored as the average value for the complete run. However, this parameter can change on the timescales of a single run. Possible reasons include a bright moon rise or fall. For simplicity, we have excluded such effects for this analysis during the data selection[3]. However, data with minor influences of changes of the NSB still can be included. In future applications, the search for transient signals might also be extended to observations with clearly changing NSB. Thus, we seek ways to include this information as a function of the time.

Figure 9.10 shows the correlation of the run-wise average L3 rate and pedestal variance. For this group of runs we do not find a correlation between the average L3 rate and the pedestal variance. A possible reason for this is that we previously excluded strong effects during the DQM. The L3 rate is not sensitive enough for the more subtle differences remaining. Thus, it is not suited to replace the pedestal variance. Below, we investigate the average multiplicity of the events. For this parameter, we find a clear correlation to the pedestal variance. This parameter can be calculated for each time step and thus provides the sought-for time dependent description of the NSB level.

3: See also subsection 6.3.3.

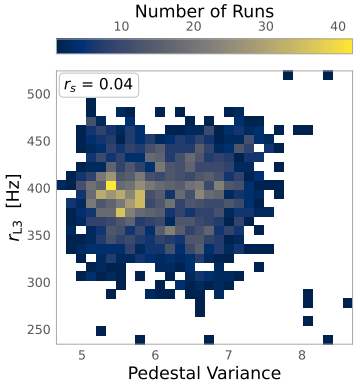


Figure 9.10: Correlation of run-wise average pedestal variance and average L3 rate for runs with $\sec(\theta) \leq 1.05$

Multiplicity

The multiplicity describes the number of images of each reconstructed event. The geometrical reconstruction of the origin requires at least two images. Thus, for VERITAS the multiplicity is in the range of two to four images. The expected multiplicity depends on the energy of the primary particle. Bright high-energetic air-showers produce strong Cherenkov-signals that can be observed with many telescopes simultanously. Also the NSB level can directly influence the multiplicity. The increased image cleaning thresholds lead to less surviving images.

The multiplicity is an attribute of the events. As discussed in subsection 9.1.1, we calculate the average value for each time step. Coming from the events themselves, we can calculate the multiplicity for each of the three energy bins, $M(\mathrm{E}_0)$, $M(\mathrm{E}_1)$, and $M(\mathrm{E}_2)$. We do not find a clear indication, that splitting the multiplicity into the energy bins provides significantly different information than the total over all energy bins $M(\mathrm{E}_{\mathrm{total}})$. Also, the statistical uncertainties for the total multiplicity is smaller. Thus, $M(\mathrm{E}_{\mathrm{total}})$ is preferred and investigated below.

The multiplicity might be a potential proxy parameter for the pedestal variance. Contrary to the pedestal variance, it can be calculated as time-dependent feature. Figure 9.11 shows the distribution between these two quantities. With $r_p = -0.9$ we confirm a linear anti-correlation. As expected, the multiplicity can be used as a sensitive parameter to describe changes of the pedestal variance. Its main advantage is the possible time-dependent calculation that can also describe intra-run changes of the NSB.

Figure 9.12 shows the γ-like rates for $\sec(\theta) \leq 1.05$ as function of $M(\mathrm{E}_{\mathrm{total}})$. Except from the inverse slopes, the correlation is similar to pedestal variance in Figure 9.9. It prooves, that the multiplicity is a suitable parameter to replace the pedestal variance which might also be helpful to describe intra-run changes. We investigate this possiblity further in subsection 9.1.3.

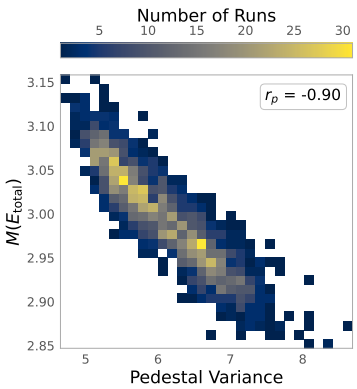


Figure 9.11: Correlation of pedestal variance and total multiplicity $M(\mathrm{E}_{\mathrm{total}})$ for runs with $\sec(\theta) \leq 1.05$.

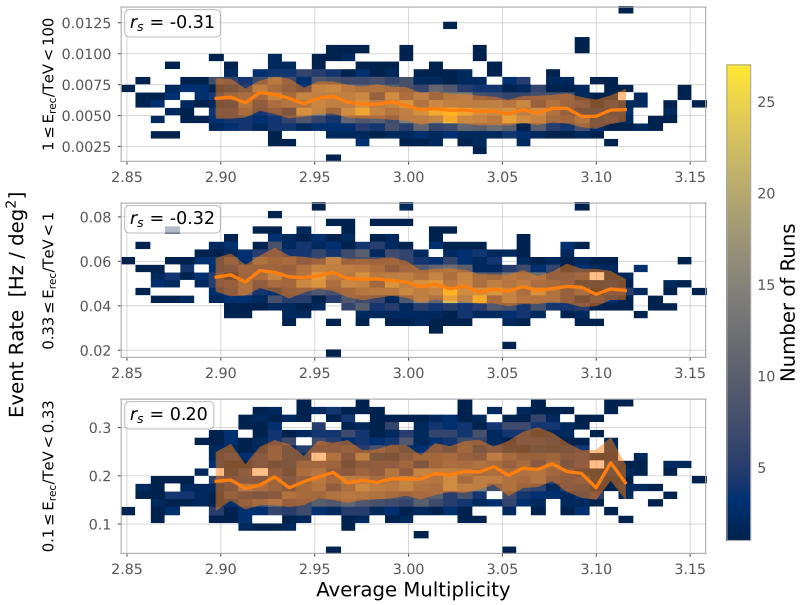
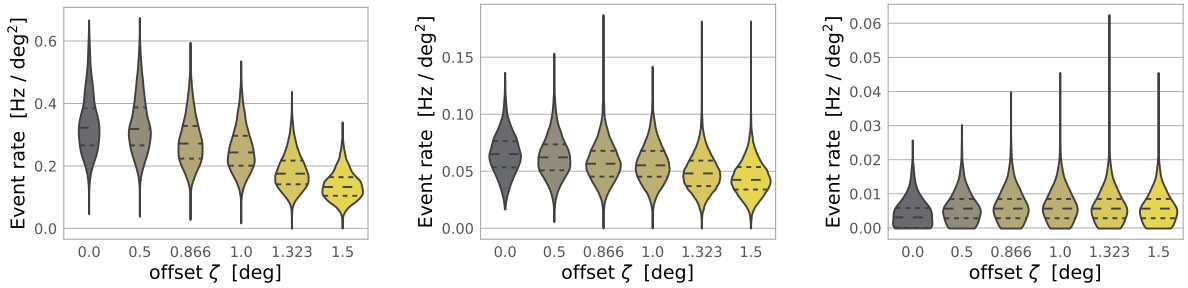


Figure 9.12: Correlation of event rates with multiplicity for $\sec(\theta) \leq 1.05$. The orange line shows the median of each multiplicity bin. The shaded area corresponds to the 1σ containment area.

Offset angle

The sensitivity of the instrument depends on the position of a ROI within the FoV. This sensitivity is roughly radial symmetric around the center of the FoV. In traditional analyses, the dependency is described by the radial acceptance. In this thesis, we directly use the offset angle ζ, instead. The LSTM extracts the critical differences between the individual ROIs from the data itself.



(a) Energy bin $0.1 \leq E_{rec}/TeV < 0.33$.

(b) Energy bin $0.33 \leq E_{rec}/TeV < 1$.

(c) Energy bin $1 \leq E_{rec}/TeV < 100$.

Figure 9.13: Dependency between ROI-wise event rate and offset angle ζ.

We validate this by splitting the FoV into the same ROIs used for the analysis [4]. In total we have ROIs with 6 distinct offset angles $\zeta \in \{0.0°, 0.5°, 0.866°, 1.0°, 1.323°, 1.5°\}$. ROIs with more than 10% of the area excluded are removed. For the remaining ROIs, we investigate the dependency of the background rates on the offset angle. For each group of offsets we estimate the distribution of rates with a kernel density estimate. The rates for all energy bins are illustrated in Figure 9.13. The dashed and dotted lines show the median and 1σ-containment.
For the two lower energy bins, we find that the rates decrease further away from the center of the FoV. Only the first two groups at 0° and 0.5° show overall similar distributions. This is consistent with the radial acceptance used in the standard VERITAS analysis. In the highest energy

4: See section 8.1 for more details about the ROI definition.

bin, the median rates are slightly higher for large ζ. However, the statistic at this energy bin for a single ROI is very limited. This study confirms the expected dependency of the background rates on the offset angle ζ. Thus, we include it as auxiliary parameter to describe the differences between the ROIs.

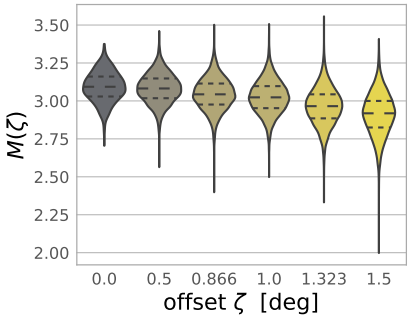


Figure 9.14: Dependency of ROI-wise multiplicity $M(\zeta)$ on offset angle ζ for runs with $\sec(\theta) \leq 1.05$.

Previously, we discussed that the average multiplicity is calculated directly from the events. During the data preparation, the events are assigned to individual ROIs. Thus, the average multiplicity can also be calculated independently for each ROI. We show the dependency of the ROI-wise multiplicity $M(\zeta)$ on the offset angle in Figure 9.14. The median multiplicity decreases more to the edge of the FoV. However, the calculation of this quantity is more challenging due to the small statistics in individual ROIs. As discussed above, M_{total} and ζ are sensitive auxiliary parameters. Adding $M(\zeta)$ is only meaningful if it adds relevant information not transported yet.

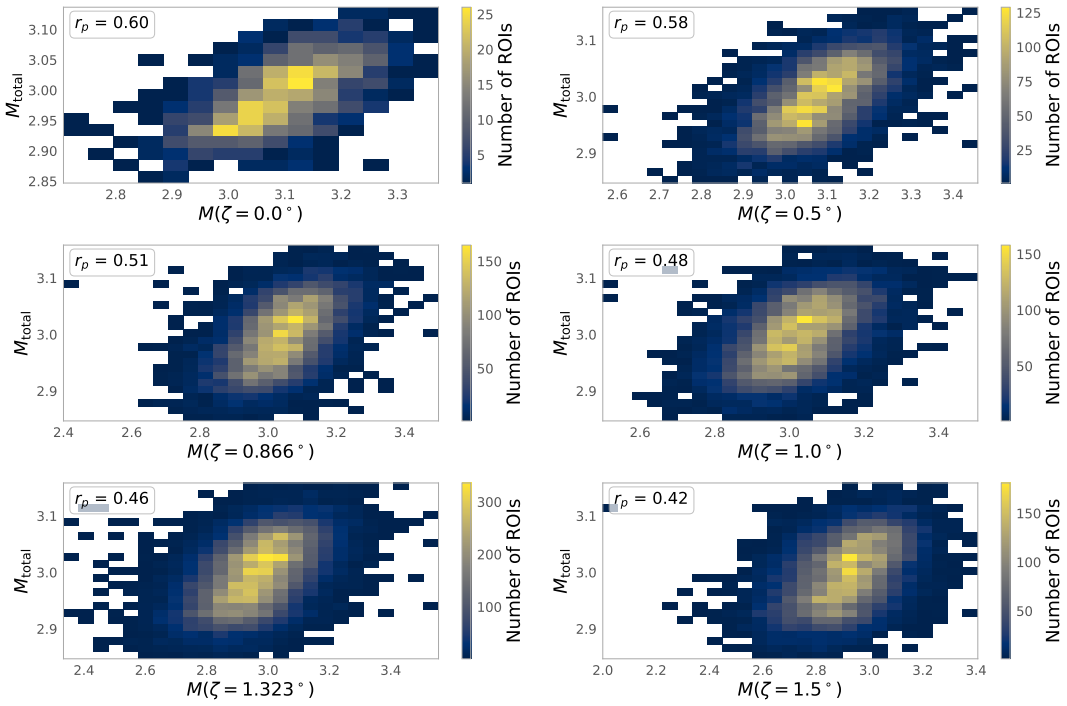


Figure 9.15: Comparison of ROI-wise $M(\zeta)$ and FoV-wise $M(\zeta_{\text{total}})$ multiplicity. Each sup-graph represents one ROI group at offset ζ with $\zeta \in \{0.0^\circ, 0.5^\circ, 0.866^\circ, 1.0^\circ, 1.323^\circ, 1.5^\circ\}$.

We investigate the correlations of $M(\zeta)$ and M_{total} in Figure 9.15. Each

subplot uses the multiplicity of one ζ-group in the x-axis. The y-axis shows M_{total} using all events in the complete FoV. Due to the lower statistic, the spread of the distributions of $M(\zeta)$ are significantly larger for all ζ. We find linear correlations in the range from $r_p = 0.42$ to $r_p = 0.6$. We do not find an indication, that $M(\zeta)$ adds crucial information not included in M_{total}. Adding parameters with larger larger noise can be counterproductive for the performance. Thus, the FoV-wise multiplicity M_{total} is the preferred input to the RNN.

Summary

Above we discussed each candidate for the auxiliary parameters for the RNN. For this we investigated the run-wise averages of these parameters in different zenith angle bins. We summarize the outcome of this study in Table 9.3. The column *intra-run* specifies whether the parameter potentially can be used to describe the intra-run changes. Below, we investigate which of these parameters are relevant to describe these short term effects.

Table 9.3: Summary of run-wise study of auxiliary parameters. The column *use* specifies whether each parameter is used in the RNN or not. Parameters highlighted in *intra-run dependency* can be used to describe the intra-run variations of the background. In a separate study below, we investigate if they provide relevant information. In the final column we summarize the most important points of this study for each parameter.

Parameter	Use	Intra-run	Comments
$\sec(\theta)$	✓	✓	Propagation length of atmospheric showers
$\lvert\alpha\rvert$	✓	✓	Effects on lateral shower development by geomagnetic field
ΔT_{ref}	✓	✓	Long-term degeneration of instrument
l, b	✗	✗	Observations in the galactic plane are excluded
R_{L3}	✓	✓	Intra-season changes of the instruments
`dead_time`	✗	✓	Redundant due to linear correlation to L3 rate
`ped_var`	✗	✗	Only accessible as average number for run
$M(\text{E}_{\text{total}})$	✓	✓	Description of time dependent NSB to replace `ped_var`
$M(\text{E}_0)$, $M(\text{E}_1)$, $M(\text{E}_2)$	✗	✓	Correlation to $M(\text{E}_{\text{total}})$, limited by low statistic
ζ	✓	✗	Radial acceptance for different ROIs

9.1.3 Intra-run Event Rates

VERITAS observing runs usually have a duration between 15 min and 30 min. During this time, the rate of detected γ-like events can be influenced by changes in the observing conditions. In this subsection, we are interested in finding the auxiliary parameters which track significant changes during the individual runs.
In subsection 9.1.2, we investigated which parameters are important to describe the background event rates. In total we identified six suited auxiliary parameters. Five of these, $\sec(\theta)$, α, ΔT_{ref}, R_{L3}, and $M(E_{\text{total}})$, can optionally be used as time-dependent input to $\mathcal{S}_{\text{RNN}}$. We aim to identify which of the five potential parameters describe the short term intra-run changes of rates. We use the full training set of 5725 runs for this study. In the previous subsection, we showed that the offset angle ζ is sufficient to describe the changes within the ROI. Thus, we integrate the event counts within the inner 1.75° of the camera. Also time cuts and exclusion regions are handled as described above.

To investigate the intra-run effects, we apply the time-binning in 20 sec that is also used during the analysis. The investigated time series correspond to the input series $\mathcal{S}_\mathrm{p}$ described in Chapter 8. It has a total of τ_p time steps N_η which contains the number of γ-like counts in each energy bin $\eta \in \{0, 1, 2\}$. We fit a linear curve to the series of each energy bin

$$f_\eta(\tau) = S(N_\eta)\tau + C(N_\eta), \tag{9.2}$$

where τ are the increasing integer numbers of the time bins. The slopes $S(N_\eta)$ characterize intra-run changes that span the complete run. For instance, a monotonic increase of N_η corresponds to a positive slope. Below, we use this slope to identify trends in the dataset. We also calculate the standard deviations $\sigma(N_0)$, $\sigma(N_1)$, and $\sigma(N_2)$ to characterize variations during the runs. From a pure Poisson distribution, we expect $\sigma(N) = \sqrt{\overline{N}}$, where $\overline{N}$ is the mean observed counts. In order to compare $\sigma(N_\eta)$ from all runs, we normalize these to the expectation

$$\hat{\sigma}(N_\eta) = \frac{\sigma(N_\eta)}{\sqrt{\overline{N_\eta}}}. \tag{9.3}$$

For a perfect Poisson distribution the expectation value 1 is independent of $\overline{N_\eta}$. This allows identifying deviations independent of the overall scale of the counts. Contrary to the slope, $\hat{\sigma}(N_\eta)$ is also sensitive to non-monotonic trends during one run. These can be dips or bumps in the event counts on shorter time scales.

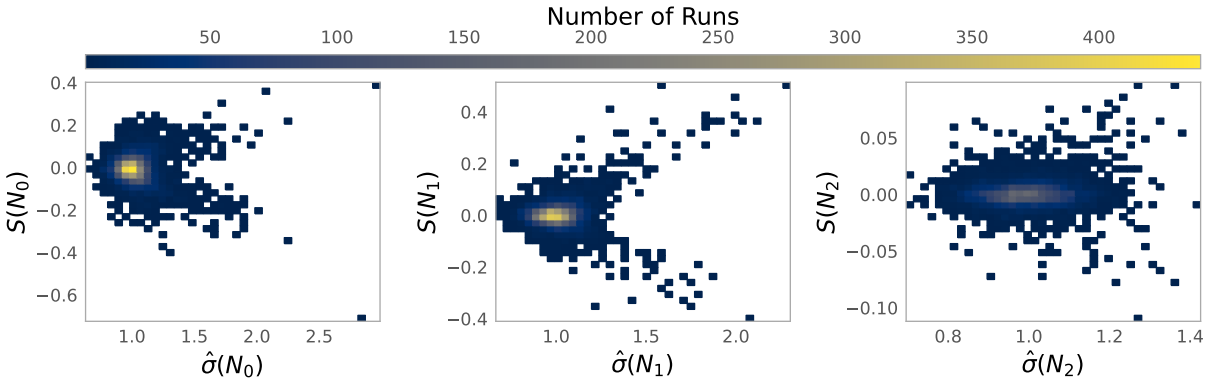


Figure 9.16: Dependency of the standard deviations σ and the slopes S for the counts in the three energy bins n_0, n_1, n_2.

Figure 9.16 shows the dependency of the $\hat{\sigma}(N_\eta)$ and $S(N_\eta)$ for each energy bin. Runs with $\hat{\sigma}(N_\eta) \sim 1$ and $S(N_\eta) \sim 0$ are considered to have stable rates. In this study, we are particularly interested in variable runs. For these the standard deviation is larger than expected $\hat{\sigma}(N_\eta) > 1$. In case of a monotonic rise or fall $|S(N_\eta)| > 0$. In the graph, we find two groups of outliers with either positive or negative slopes. To identify auxiliary parameters that cause these intra-run changes, we use the Spearman's rank correlation coefficient r_s. If for an individual run one parameter describes these change, we expect it to be correlated to N_η. Thus, we expect a significant $|r_s| \gg 0$ for runs with large $\hat{\sigma}(N_\eta)$ or $S(N_\eta)$. For runs without intra-run changes, we also do not expect a correlation. Below, we investigate each of the five potential time-dependent parameters. Due to its sensibility to monotonic and non-monotonic changes, we predominantly use $\hat{\sigma}(N_\eta)$ to identify variable runs. However, $S(N_\eta)$

provides important insights for understanding the results.

Time

For each of the runs we calculate the correlation of ΔT_{ref} to the counts in the energy bins $r_s(N_\eta, \Delta T_{\text{ref}})$. Figure 9.17 shows it as a function of $\hat{\sigma}(N_\eta)$. ΔT_{ref} increases linearly during the runs. Thus, a correlation with the time series of γ-like counts N_η only occurs for runs with a steady increase or decrease of the rates. This corresponds to $|S(N_\eta)| > 0$, and the sign of $S(N_\eta)$ defines whether r_s is positive or negative. Due to the fact that the correlations can be either positive or negative, it does not provide important information to describe the underlying effects. Thus, this '*c-shaped*' distribution is an clear indication for an irrelevant correlation. The same conclusion can be also drawn from the distribution of $r_s(N_\eta, \Delta T_{\text{ref}})$ and $S(N_\eta)$.

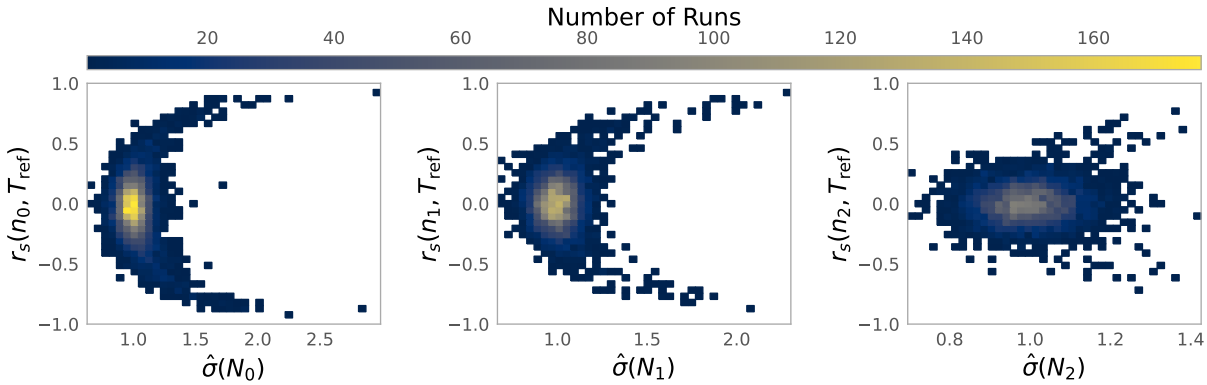


Figure 9.17: Intra-run correlations of the counts in the three energy bins n_0, n_1, n_2 with ΔT_{ref} as a fuction of $\hat{\sigma}(n)$.

While ΔT_{ref} is crucial to describe the aging of the telescope, it is negligible on the time-scales of an individual run. Thus, using a constant value throughout each run is less complex. In this thesis, we use the starting time of the observing run as auxiliary parameter.

Zenith Angle of Observation

In subsection 9.1.2 we discussed that the zenith angle of observation is among the most critical parameters to describe the sensitivity of the instrument. The auxiliary parameter $\sec(\theta)$ provides an accurate description of this effects. The changes of θ within each run are typically up to few degrees. The distribution of the maximum changes within the training runs are shown in Figure 9.18. For low zenith angles $\theta \lesssim 25°$ the sensitivity is mostly independent of these changes. However, at lower elevations they cause significant changes of the background rates for an individual run.

We show the distribution the correlations $r_s(N_\eta, \sec(\theta))$ with $\hat{\sigma}(n)$ for each energy bin in Figure 9.19. For the first energy bin, runs with large $\hat{\sigma}(n_0)$ typically have a significant anti-correlation $r_s(n_0, \sec(\theta)) \to -1$. The reason for this is the rise of the energy threshold within a single run as $\sec(\theta)$ increases. At the second energy bin we find a positive correlation $r_s(n_1, \sec(\theta)) \to 1$. As mentioned above, a possible explanation is due

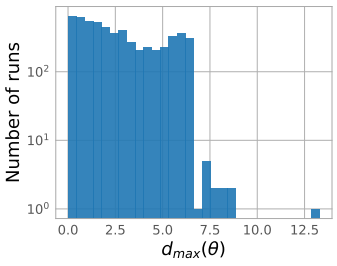


Figure 9.18: Histogram of the maximum intra-run changes of the zenith angle $d_{\text{max}}(\theta)$.

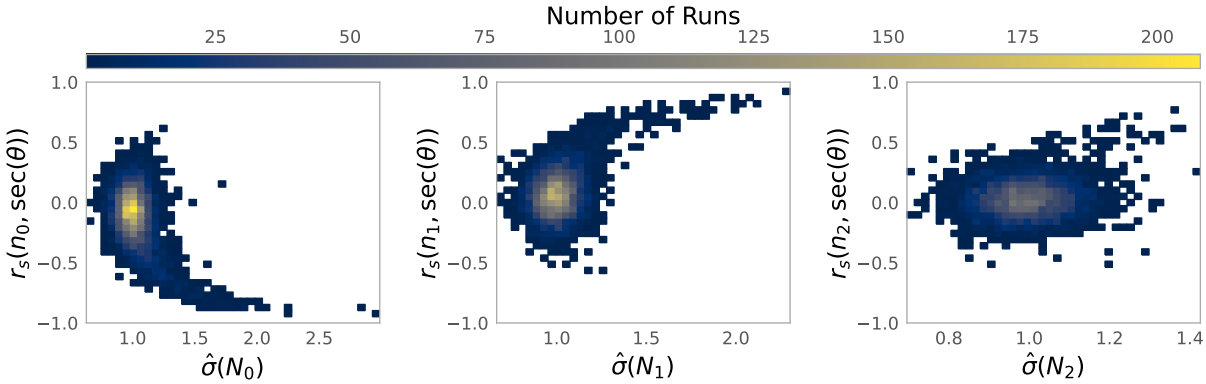


Figure 9.19: Intra-run correlations of the counts in the three energy bins n_0, n_1, n_2 with $\sec(\theta)$ as a fuction of $\hat{\sigma}(n)$.

to the projection effects. In the highest energy bin, we find a hint for a similar effect. However, the low statistic of the background counts makes a further investigation challenging. Nonetheless, the elevation is a relevant time-dependent input to describe the intra-run changes of the background rates.

Azimuth

The intra-run correlations with the absolute azimuth angle $r_s(n_1, |\alpha|)$ are given in Figure 9.20. Except from the sign of r_s, the graphs are similar to these discussed above for $\sec(\theta)$. This indicates, that $|\alpha|$ provides meaningful input to describe the intra-run changes of background rates. For the two low energy bins, the distributions reveal a smaller populations of outliers, at the inverse side as the bulk. The reason for this is that the elevation is a predominant effect. Thus, in cases that the intra-run change is mostly driven by the elevation, it can obscure the expected changes due to the azimuth. During training with the time-dependent auxiliary parameters azimuth and elevation, the RNN can also learn to estimate these effects.

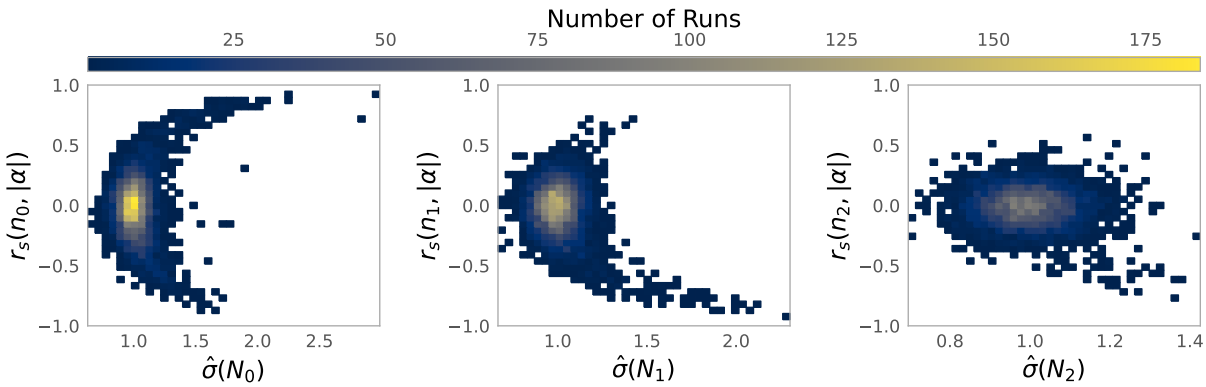


Figure 9.20: Intra-run correlations of the counts in the three energy bins n_0, n_1, n_2 with $|\alpha|$ as a function of $\hat{\sigma}(n)$.

L3 rate

In subsection 9.1.2, we discussed that the L3 rate is an important parameter to describe intra-season effects between runs. Thus, it is natural to assume that it also might help to describe effects on even shorter time scales. We also used it in Chapter 6 to identify periods with poor data quality.

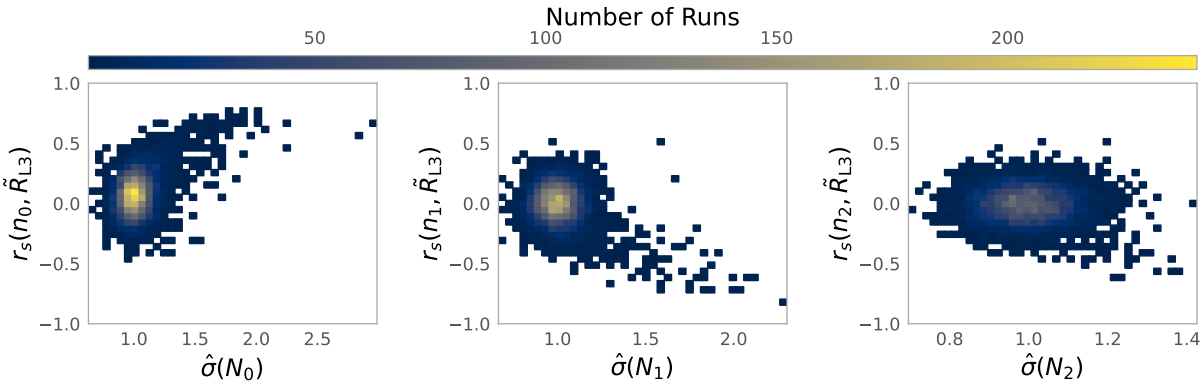


Figure 9.21: Intra-run correlations of the counts in the three energy bins n_0, n_1, n_2 with $\tilde{R}_{L3}$ as a fuction of $\hat{\sigma}(n)$.

To investigate the intra-run correlations, we calculate the median L3 rate of each time step $\tilde{R}_{L3}$. Figure 9.21 presents the distributions of the correlations with the event counts. Compared to previous correlations for $\sec(\theta)$ and $|\alpha|$, the total values $|r_s|$ are smaller. This is expected as we selected only data with modest L3 rate changes for this work. Nonetheless, we find the expected one-sided outliers in the distributions. Thus, $\tilde{R}_{L3}$ is a sensitive time-dependent auxiliary parameter.

Multiplicity

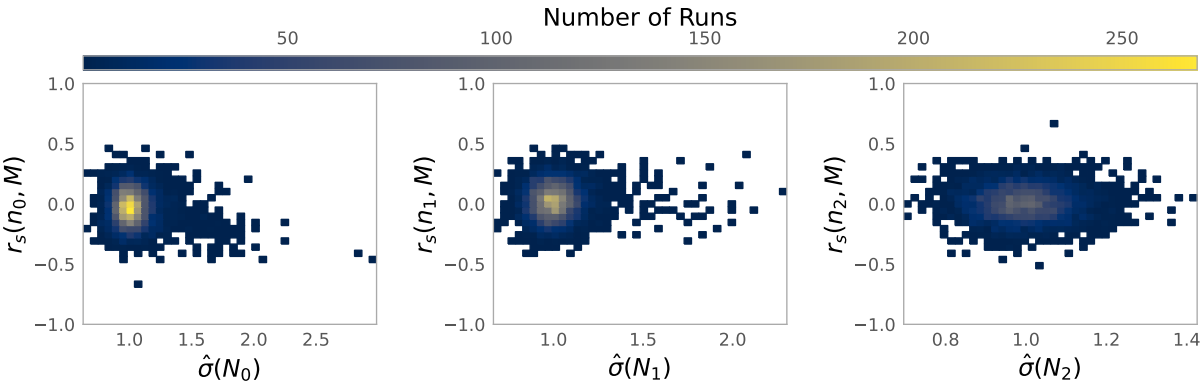


Figure 9.22: Intra-run correlations of the counts in the three energy bins n_0, n_1, n_2 with $M(E_{\text{total}})$ as a function of $\hat{\sigma}(n)$.

Previously, we identified that the multiplicity $M(E_{\text{total}})$ is a suited parameter to describe the pedestal variance. In the distributions in Figure 9.22, we do not find clear hints for a correlation with the intra-run rates. One reason for this might be the automatic data selection that specifically aims to identify periods with stable NSB conditions. Thus, periods in which the multiplicity actually might provide crucial information, such as a possible moon-rise or set, are excluded from the data-set. We conclude that for the selected data used in this thesis, the multiplicity does not

provide crucial intra-run information. Thus, we use the mean multiplicity spanning the complete observing run for this work. In future work, periods with changing NSB might be included and the multiplicity should be reevaluated as time-dependent auxiliary parameter.

Summary

We summarize the results of this study in Table 9.4. We found the expected correlations with the event rates for $\sec(\theta)$, $|\alpha|$, and R_{L3}. In the RNN we use these as time-dependent auxiliary parameters. The correlations found with ΔT_{ref} are not relevant to the network. The starting time of the observations is sufficient for describing the differences. For the dataset considered in this work, the multiplicity $M(\mathrm{E_{total}})$ is not correlated to the intra-run changes. In this thesis, we use the mean multiplicity spanning the complete run.

Table 9.4: Summary of the study of intra-run auxiliary parameters. The parameters $\sec(\theta)$, $|\alpha|$, and $\tilde{R}_{\mathrm{L3}}$ are passed to each time step. ΔT_{ref} and $M(\mathrm{E_{total}})$ are used as constant parameters for each run. The descriptions behind the arrows summarizes how the constant values are derived.

Parameter		Time Dependent Auxiliary Parameter
ΔT_{ref}	✗	→ Constant starting point of observing run
$\sec(\theta)$	✓	
$\lvert\alpha\rvert$	✓	
$\tilde{R}_{\mathrm{L3}}$	✓	
$M(\mathrm{E_{total}})$	✗	→ Mean multiplicity spanning the full observing run

9.2 Training and Validation

During the training the weights of the networks in the LSTM are updated. We discussed the architecture of the RNN in Chapter 5. We consider 5725 runs from nine years of observations for training. Loading all data at once requires several hundreds of GB of memory. This exceeds the capabilities of usual machines. Thus, we define a sequential loading scheme for training. Figure 9.23 shows a schematic illustration of this approach. We split the training of the LSTM into multiple periods. During each period, we randomly sample 500 runs from the total list of runs. Loading the prepared data yields a set of $\mathcal{S}_{\mathrm{RNN}}$ time series. Each has exactly τ_{RNN} steps and contains all target and auxiliary parameters. We split the total loaded data into 70% training and 30% validation data. During the periods, we repeat the training for ten training epochs. The training process is repeated for multiple periods to cover the complete training dataset.
By keeping the number training epochs during each period small, prevents an over optimization of the weights for the selected subset of runs. However, 500 runs are typically sufficient to get a good representation of the distribution of the observing conditions. By frequently exchanging the training training data, the total composition of data is varied. As we show below, these points of transition provides insights into the stability of the network.
As we discuss below, calculating the total cost function takes into account the absolute difference to the target values. Different scales can lead to

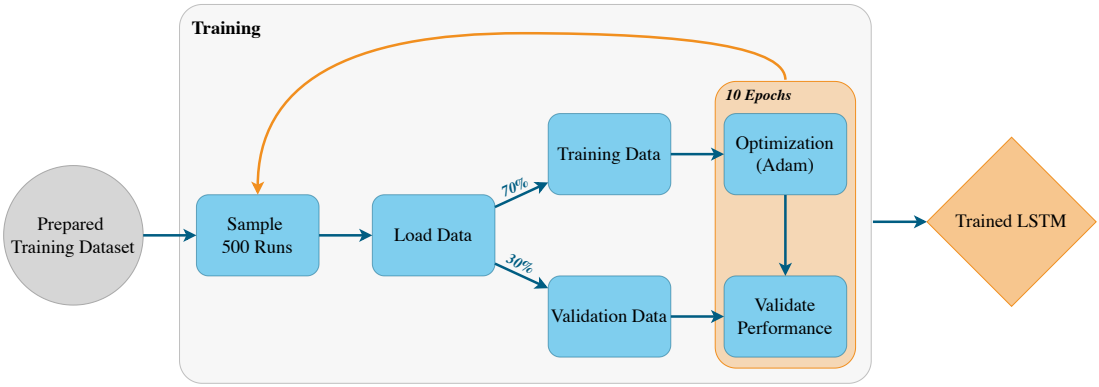


Figure 9.23: Illustration of the training process of the LSTM. We use the prepared training dataset. The amount of data is too large to load it simultaneously. Thus, we perform the training in multiple periods. The grey box illustrates one training period which is repeatedly until the training is concluded. First, we sample 500 runs from it and load this data to get the series δ_{RNN}. These are randomly split into the training and validation samples at a ratio of 70% to 30%. We use Adam optimizer to update the weights and biases of the LSTM for ten epochs. At each epoch, we use the validation data to monitor the current performance. We repeat the process of sampling runs and training until the cost function converges to a minimum.

certain parameters of them dominating the total cost. To avoid this, we scale each input feature x to the same range. The scaled quantity is

$$x_{\text{scaled}} = 2\frac{x - q_m}{q_h - q_l}, \tag{9.4}$$

where q_l, q_m, and q_h are the 0.0001, 0.5, and 0.9999 quantiles respectively. The new value x_{scaled} is roughly in the range of -1 to 1. Using the q_l and q_h instead of the extreme values makes this calculation less sensitive to outliers. Inside the RNN, only the scaled features are used. However, during the calculation of the test statistic, the scaling of the inputs and predictions is reversed. In this way, the calculation uses the actual counts of γ-like events.

In section 5.3, we introduced the general concept of the network. The final output of the probabilistic layer for each feature x are normal distributions $\mathcal{N}(\mu_x, \sigma_x^2)$. Contrary to a deterministic output, the goal of the training is to maximize the probability of the distributions describing the target values. For an individual feature x the log-likelihood is given by

$$l(\theta_x) := \log p(x; \theta_x). \tag{9.5}$$

Thus, we can define the cost function as the sum of all features

$$C = -\sum_x l(\theta_x). \tag{9.6}$$

During training the weights in the LSTM are optimized to yield the minimal C. We also use a L2-regularization [152]. The regularization adds a term

[152]: Ng (2004), 'Feature Selection, L1 vs. L2 Regularization, and Rotational Invariance'

$$C_{\text{train}} = C + C_{\text{L2}}, \tag{9.7}$$

where the regularization is given by

$$C_{\text{L2}} = \lambda_{\text{L2}} \sum_i w_i^2. \tag{9.8}$$

Here w_i are all weights in the RNN and λ_{L2} is the scale. This regularization term is added during training, however not during validation. It provides a penalty for large weights and helps to prevent overtraining in which the DNN learns random fluctuations of the training dataset. We study the impact of the hyper-parameter λ_{L2} on the training process below.
We apply a recurrent dropout of $d_{\text{Rec}} = 0.2$ during the training phase [153, 154]. For the LSTM, this randomly suppresses 20% of the candidate vectors $Tilde\mathbf{c}_t$ added to the cell state

[153]: Semeniuta et al. (2016), 'Recurrent Dropout without Memory Loss'
[154]: Hinton et al. (2012), 'Improving neural networks by preventing co-adaptation of feature detectors'

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i_t} \odot d_{\text{Rec}}(\tilde{\mathbf{c}}_t). \tag{9.9}$$

Except for the dropout, it is equivalent to Equation 5.25. Further, we add a dropout of $d_{\text{DecIn}} = 0.2$ of the inputs vectors of the decoder phase. In the chosen architecture, these are the repeated hidden states calculated by the encoder. With these dropouts, the LSTM deals with larger variations during training. As a consequence, the neurons are expected to be less dependent on the other hidden units [155]. With the dropout more robust networks can be trained. Furthermore, it is an additional method to prevent overtraining.
The training data is used to optimize the weights using the Adam optimizer with learning rate ϵ. During each period p, we start with an initial learning rate ϵ_p. It is kept constant for the first three training epochs of the period before decreasing it by 5% each subsequent epoch. From one epoch to the next, the total learning rate is decreased by 10%

[155]: Srivastava et al. (2014), 'Dropout: a simple way to prevent neural networks from overfitting'

$$\epsilon_{p+1} = 0.9 \cdot \epsilon_p. \tag{9.10}$$

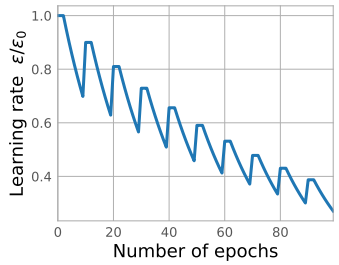


Figure 9.24: Learning rate ϵ for training of the LSTM relative to the initial learning rate ϵ_0

This learning rate schedule gives more importance every time new data is loaded. Decreasing ϵ during the periods, contributes to not over-optimizing to each subset of runs. Figure 9.24 illustrates this learning rate schedule. The impact of the initial learning rate ϵ_0 at period $p = 0$ is investigated below.
For selecting the size of the neural network, we train the network with different number of layers L and neurons. The objective is to provide enough flexibility to achieve a good description of the input data. However, keeping the total size modest prevents overtraining and helps building simpler model. We find that an LSTM with three layers with 32, 64, and 128 neurons, respectively, fulfills these criteria. In the following, we use this LSTM size. We also note, that a further optimization of this hyper-parameters might influence the performance.
During training, the model parameters are optimized to minimize C_{train}. However, the performance is validated on the independent validation dataset C_{valid}. Naively, it is expected that the $C_{\text{train}} < C_{\text{valid}}$. However, C_{train} has the additional L2-regularization term and is influenced by the dropouts. Both of these typically increase the cost functions. In this case, it is expected that $C_{\text{train}} > C_{\text{valid}}$.
The initial learning rate ϵ_0 and the scale of the L2-regularization λ_{L2} are among the most important hyper-parameters during training [156]. We scan a range of parameters to find suitable values for these parameters. For each iteration, we train the LSTM for three periods. We monitor C_{train} and C_{valid} to identify which values provide a stable training. We use a fixed random seed for the selection of runs and splitting the data for training and validation. In this way, each training process utilizes the identical datasets. Nonetheless, the training is subject to further

[156]: Bengio (2012), 'Practical Recommendations for Gradient-Based Training of Deep Architectures'

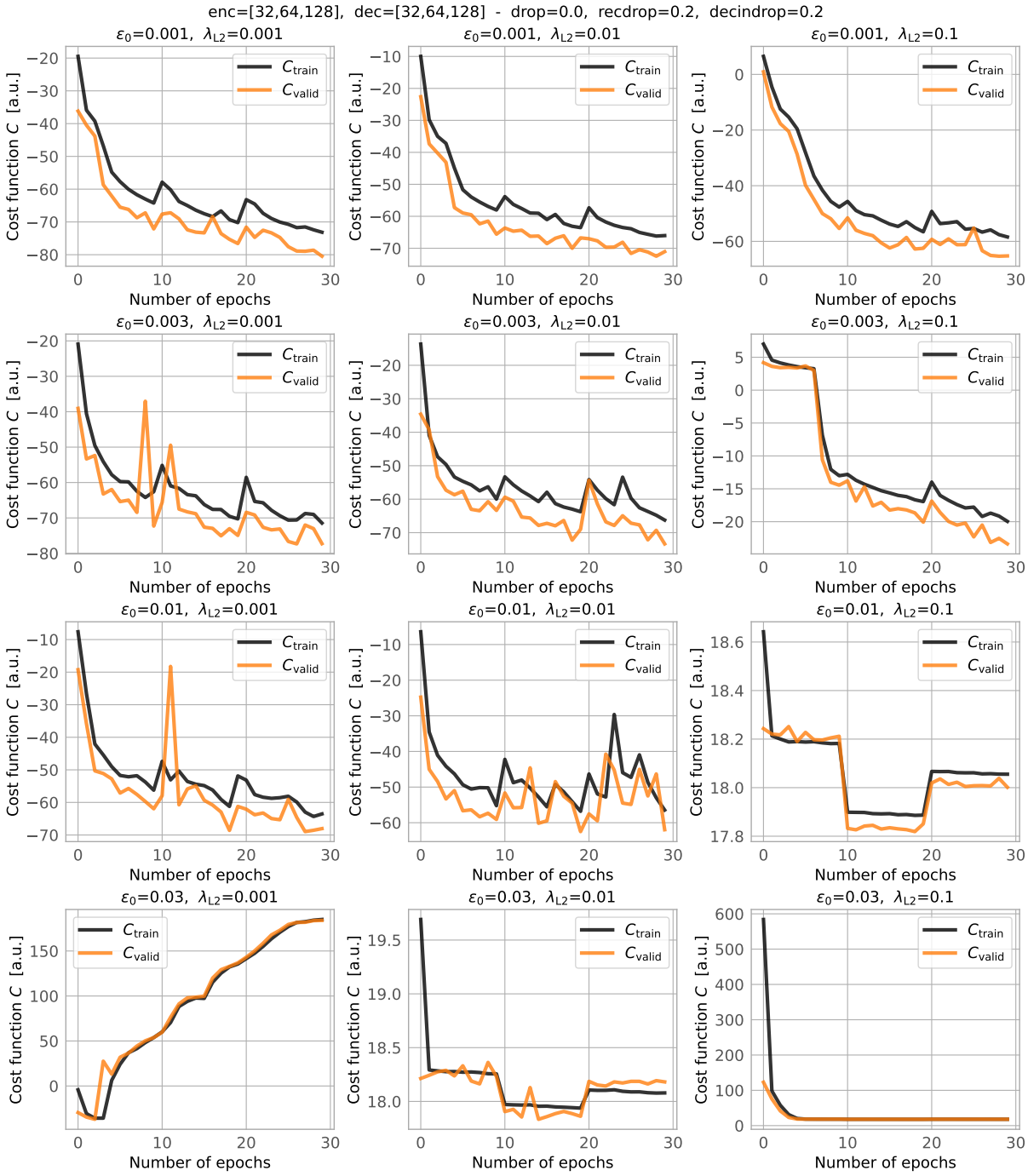


Figure 9.25: Cost functions during the first 3 training periods of the LSTM with different hyper-parameters. The rows corresponds to different initial learning rates ϵ_0 and the columns to the scale of the normalization λ_{L2}. Black curves show the training cost C_{train} and orange the cost for the validation data C_{valid}.

randomization. Among these are the selection of the mini-batches and the initialization of the weights of the LSTM. In this thesis, we use a batch size of $B = 128$.

We show the results for the optimization of the hyper parameters in Figure 9.25. Each row shows the results fixed learning between 0.001 up to 0.03. The columns correspond to the scale of L2-regularization between 0.001 and 0.1. Overall, we find that the largest values, $\epsilon_0 = 0.03$ and $\lambda_{L2} = 0.1$, do not yield stable results. These correspond to the last row

Table 9.5: Summary of training hyper-parameters.

Hyper-parameter	Value	Comment
L	3	Number of layers
n_{neurons}	[32, 64, 128]	Number of neurons per layer
d_{rec}	0.2	Recurrent dropout
d_{DecIn}	0.2	Decoder input dropout
B	128	mini-batch size
p	10	Number of epochs per periods
n_{runs}	500	Loaded runs per period
λ_{L2}	0.003	L3-regularization scale
ϵ_0	0.001	Initial training rate
f_p	0.1	Decrease of ϵ per period
f_e	0.05	Decrease of ϵ per epoch

and last column, respectively. Among the observed effects are steadily increasing cost functions, inability to learn the trends, and relatively slow and unstable learning process. Further, we also find instances in which the training fails all together, $(\epsilon_0, \lambda_{\text{L2}}) \in [(0.03, 0.01), (0.01, 0.1)]$. The changes in C in these cases are only due to the exchange of the training and validation data. Reasons for these effects might be the random initialization of the weights. Repeating this study several times, showed a high probability of getting such results for these large values of ϵ_0 and λ_{L2}.

The remaining six graphs ($\lambda_{\text{L2}} \in \{0.001, 0.01\}$, $\epsilon_0 \in \{0.001, 0.003, 0.01\}$) all show the expected decrease of the cost functions. We note that the evolution C is smoother for smaller ϵ_0. Also the best values of C after three training periods are reached for $\epsilon_0 = 0.001$. We select this value for the training of the LSTM. The lowest values of the cost functions are reached for $\lambda_{\text{L2}} = 0.001$. However, it is a trade-off with smoother graphs that are reached using a larger regularization scale. We decide to select an intermediate value of $\lambda_{\text{L2}} = 0.003$. We summarize all hyper-parameters used during training in Table 9.5. With these, we initially run the training for a total of 20 periods. With randomly selecting 500 runs at each period, in average all training runs are loaded more than once.

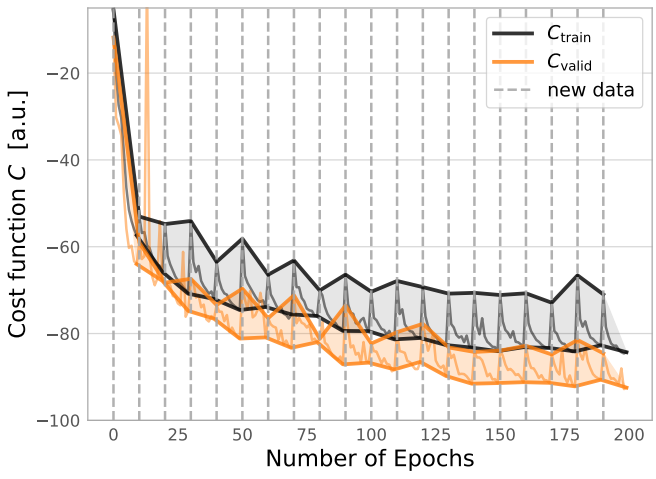


Figure 9.26: Cost function during training of LSTM with envelope. The shaded area illustrates the range of the cost function reached at the beginning and the end of each period.

We monitor closely the cost functions to investigate the performance of the LSTM. Their evolution during the training is illustrated in Figure 9.26.

The start of the periods are illustrated by the vertical dashed grey lines. As they are associated with a replacement of the the data, C_{train} and C_{valid} show the characteristic jumps. Within each period, the RNN optimizes the weights for the selected data. This leads to an overall decrease of C. The shaded areas illustrate the enveloping values at the beginning and the end of the periods. The starting values are particularly interesting as they represent how robust the performance is for different selections of data. For a stable performance, we expect this value to be rather constant. The values in the end of the periods show the performance after optimizing the network for this certain data. We find that the enveloping cost functions are roughly constant starting at period 13.
To decide if the training should be continued further, we investigate these periods. We show a zoomed graph of C in Figure 9.27. The dashed lines show fits to the starting and end values of C_{valid}. The slopes over these seven periods are 0.002 and -0.008, respectively. Thus, the overall trending changes of the cost function over these periods is negligible. Also it is sign of a robust network that performance is stable with different slices of the input data. Thus, we conclude, that the network is successfully trained and finish after the total of 20 training periods.

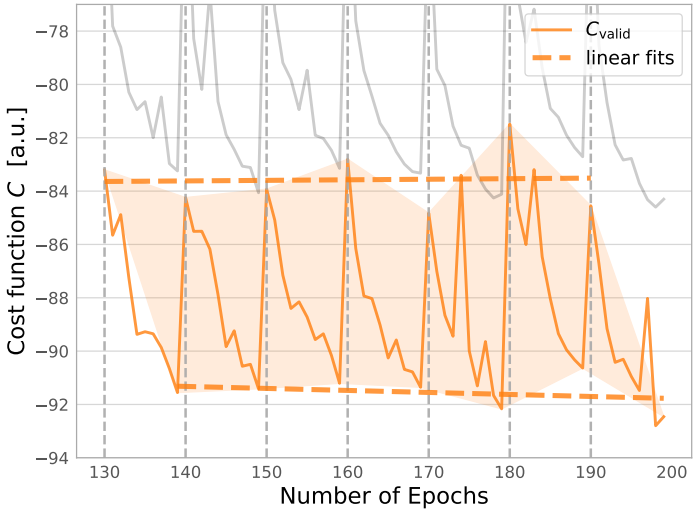


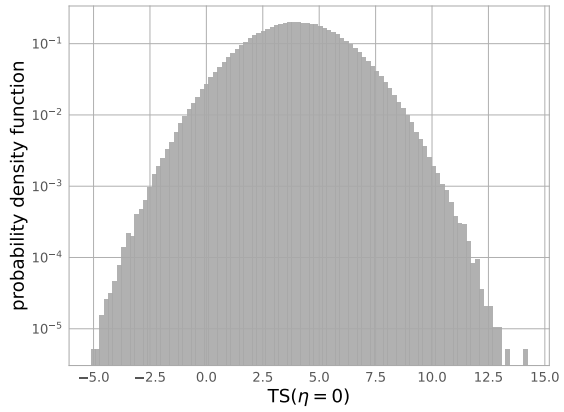
Figure 9.27: Zoom in to the cost function during training of LSTM. The grey line shows the cost function of the training dataset C_{train}

9.3 Calibration Pipeline

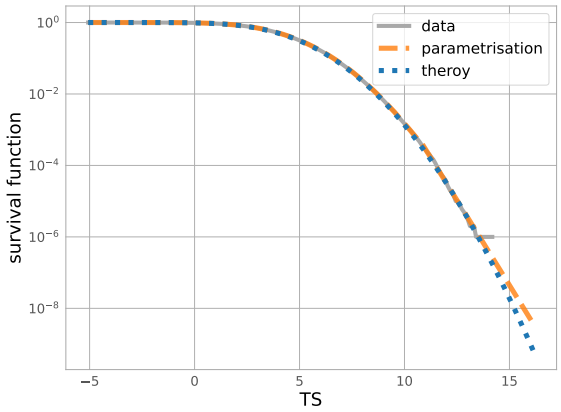
As discussed above, the calibration allows to interpret a recorded signal with a statistical significance. The general concept of turning any test statistic to a probability based on background samples was briefly introduced before. The only requirement to the TSs is that they increase for a more significant detection, i.e. stronger excess of γ-rays.
In order to store and evaluate the TS distributions we parameterize the background results. We give details on this parametrisation in subsection 9.3.1. This calibration depends on the overall background rate and the accuracy of the RNN. For an accurate description, we split the total data-set into meta bins. The calibration is done independently in each of these meta bins. We describe the selection of these bins in subsection 9.3.2. The calibration is implemented in three individual stages. The TS for a

single series and energy bin was introduced previously in Equation 5.29. The following stages combine the TS values of each previous stage to obtain a new TS. The individual steps of the calibration pipeline are discussed in subsection 9.3.3.

9.3.1 Parametrisation of the Test Statistic



(a) Probability density function of the normal distribution.



(b) Survival function of the normal distribution. The orange dashed line shows the parametrisation of the survival function that yields the conversion to a p-value. The theoretical value is illustrated by the dotted blue curve.

Figure 9.28: Probability density function and survival function of a normal distribution with $\mu = 4$ and $\sigma = 2$ reconstructed constructed from $n_s = 10^6$ samples.

The TS values for the background distributions are roughly distributed following a normal distribution. Its probability density function (PDF) is given by

$$\mathrm{PDF}(\mathrm{TS}) = \frac{1}{2\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\mathrm{TS}-\mu}{\sigma}\right)^2\right], \tag{9.11}$$

where μ is the mean and σ^2 is the variance of the distribution. An example of this distribution is shown in Figure 9.28a for $n_s = 10^6$ samples. Both parameters are determined by the quality of the background predictions from the RNN. We will discuss this in more depth in section 9.3.3. The fraction of values above a certain value is described by the survival function (SF)

$$\mathrm{SF}(\mathrm{TS}) = 1 - \mathrm{cdf}(\mathrm{TS}) = 1 - \int_{-\infty}^{\mathrm{TS}} \mathrm{PDF}(x)\,\mathrm{d}(x). \tag{9.12}$$

Here CDF is the cumulative density function integrating the PDF up to a certain TS value. It describes the frequency that the value is smaller than TS. Figure 9.28b gives the survival function of the distribution. Evaluated on the background dataset, the survival function corresponds to the probability of obtaining a certain value of $x \geq \mathrm{TS}$. Thus, it directly describes the sought-for mapping of the TS as the statistical probability. We refer to this value as the p-value. A small p-value corresponds to a small probability of an event being caused by a statistical fluctuation of the background. This corresponds to a higher significance for the

detection. The minimum p-value that can be inferred directly depends on the number of background events

$$\min(p) = \frac{1}{n_s}. \tag{9.13}$$

Creating a sufficient amount of background data to extend to required range of very small p-values is not feasible. However, a mapping from the TS p-value is crucial to determine the significance of rare events. Thus, we use a parametrisation of the survival function to extend to higher significances. We fit the final part of the survival function by a log-linear curve

$$f(TS) = 10^{\mathrm{TS} \times a + b}, \tag{9.14}$$

where a describes the steepness of SF and b is a normalisation. For this fit, we use the range covering the last two orders of magnitudes in the p-values. We use this fit to extrapolate the graph to smaller p-values.
The complete parametrisation is the combination of the direct interpolation for sufficiently small TS values and the log-linear fit at higher TS. In Figure 9.28, this parametrisation is shown by the orange curve. We note that for the normal distribution the analytical value of the survival function is given by

$$\mathrm{sf}(\mathrm{TS}) = \frac{1}{2}\left[1 - \mathrm{erf}\left(\frac{\mathrm{TS} - \mu}{\sigma\sqrt{2}}\right)\right], \tag{9.15}$$

where erf is the error function. As illustrated by the blue dotted curve in Figure 9.28b, it is concave in the log-linear space. Approximating it by Equation 9.14 yields higher p-values making this parametrisation more conservative.

9.3.2 Calibration Meta Bins

The interpretation of the test statistic depends on the absolute background level. Also the accuracy of the background estimation directly influences the expected distributions. A bias in the prediction leads to a shift μ of the PDF and SF. Further, a large spread of the predictions corresponds to larger error σ corresponding to a more gentle fall of the SF. To consider the differences in the observed data, we group these into meta bins. These contain data with similar conditions which are represented in small changes of the background rates. We base this selection on the previous investigation for the auxiliary parameters in section 9.1. Selecting more meta bins improves the agreement of data in the bins. However, it also decreases the statistic which leads to a less precise estimate of the survival function. Thus, it is important to consider the most important differences while keeping the total number of bins low. Previously we identified $\sec(\theta)$, $|\alpha|$, ζ, and ΔT_{ref} as the most critical parameters to describe the background event rates.
Figure 9.1 shows the smooth transition of the event rate as a function of $\sec(\theta)$. We define a total of four linear meta bins with width 0.1 in this parameter spanning the total range from 1.0 to 1.4. The main difference in the azimuth are due to the pointing towards north or south. This was illustrated in Figure 9.2. Thus, we define two meta bins with $|\alpha| \leq 90$

and $|\alpha| > 90$.
The most significant differences due to the time are the long term changes. For the calibration, we study the differences between each of the nine epochs. We show the median event rate and the 1σ-containment for each energy bin in Figure 9.29. Due to the changes in the energy threshold, the most significant differences are at the lowest energy bin. We define three meta bins in T_{eff}, $[0, 3)$, $[3, 6)$ and $[6, 9)$, which group together epochs with similar rates.

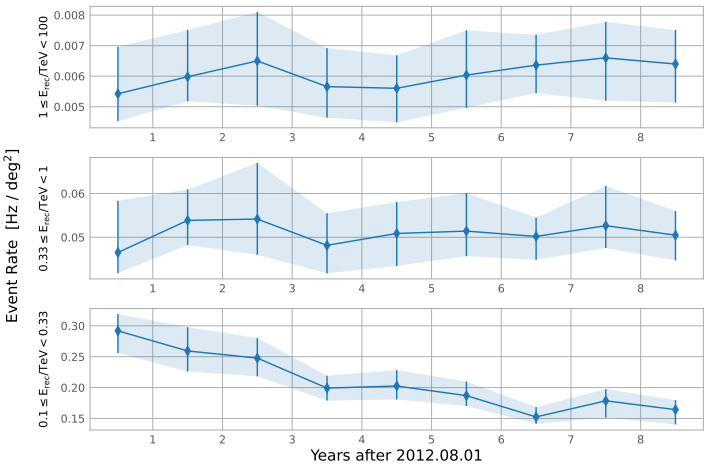


Figure 9.29: The markers show the season-wise median rate per VERITAS run for the three energy bins. The error bars and bands show the 1σ containment.

The correlation of the background rate and the offset angle is investigated in Figure 9.13a. In this work, we have six discrete values of ζ corresponding to the locations of the ROIs. The ROIs at 0° and 0.5° show almost identical distributions. This is in alignment with the radial acceptance which is roughly constant in this range. For the calibration, we do not need to differentiate between these two. As all other offset angles have revealed a decreasing sensitivity these are grouped in independent meta bins. In total, we define five offset meta bins.
Table 9.6 shows the summary of meta bins. The values of $\sec(\theta)$, $|\alpha|$, and T_{eff} do not change substantially within a single observing run. We use the run-wise mean values to group each observations into on of the meta bins. Thus, we refer to these as run-wise meta bins. In contrast, the offset angle ζ is a ROI-wise meta bin. Each observing run encompasses several ROIs. Thus, the data from a single run is subdivided into these five meta bins.

Table 9.6: Summary of meta bins. In total we define 24 run-wise meta bins in the three parameters $\sec(\theta)$, $|\alpha|$, and T_{eff}. In combination with the five ROI-wise bins in ζ, the total number of meta bins is 120.

Parameter	N	Range	Scope
$\sec(\theta)$	4	[1.0, 1.1, 1.2, 1.3, 1.4]	run
$\|\alpha\|$	2	[0, 90, 180] deg	run
T_{eff}	3	[0, 3, 6, 9] yr	run
ζ	5	$(0, 0.5)$, 0.866, 1.0, 1.323, 1.5	ROI

9.3.3 Calibration Stages

The most elemental TS is calculated for an individual energy bin in one ROI. However, the information of a single transient signal might be spread over several energy bins and ROIs. Thus, combining these

information can improve to overall detection efficiency. In this part, we introduce the novel approach of calibration stages. They allow sequential combination of results from the previous stages to a new test static. This TS itself may be calibrated to be interpreted as a p-value using the parametrisation in subsection 9.3.1.
In this thesis, we deploy two individual stages. The first stage deals with the single time series per energy bins and ROIs. During the second calibration stage, the energy bins are combined. These stages, deal with ROI specific results. Thus, the relevant meta bins include the run- and ROI-wise meta bins. We summarize these in Table 9.7.

Stage	Description	Meta Bins
1	single energy and ROI	$\sec(\theta)$, $\lvert\alpha\rvert$, `epoch`, ζ
2	combine energy bins	$\sec(\theta)$, $\lvert\alpha\rvert$, `epoch`, ζ
3	combine neigboring ROIs	$\sec(\theta)$, $\lvert\alpha\rvert$, `epoch`, ζ
4	combine full FoV	$\sec(\theta)$, $\lvert\alpha\rvert$, `epoch`

Table 9.7: Overview of calibration stages. The first two stages deal with single energy bins and ROIs. The energy bins are combined in the second calibration stage. In the future, the calibration stages might be extended by a combination of neighboring ROIs and the full FoV.

In future work, the calibration might be expanded to also combine the results from neighboring ROIs. This might improve the sensitivity to events which are not well aligned with the ROIs. Further, also the combination to a full FoV TS might be possible. For this stage, only the run-wise meta bins would be relevant. Including these two calibration stages is beyond the scope of this work.
In case of the detection of a transient signal in a combined TS, the signal can be traced back to the earlier calibration stages. This is possible due to the sequential construction of the TS. It gives direct insight to the locations and energy bins which are important for the detection.
The calibration interpret the single windows $\mathcal{S}_{\text{RNN}}$ of the sliding window. They do not account for the total number of series which are probed. For the final calculation of the significance, the trial factors need to be considered. This converts the p-value from a local significance to the global significance. We discuss the trial factors subsequent to the three calibration stages.

Single Energy Bin Calibration

The first calibration stage calculates a TS for each energy bin η and each ROI r. We define this most fundamental test statistic in Equation 9.16:

$$\text{TS}_{\text{single}}(\eta, r) = \sum_{\tau=1}^{\tau_{\text{dec}}} \frac{S(\tau, \eta, r) - B(\tau, \eta)}{\sqrt{B(\tau, \eta, r) + 1}}. \tag{9.16}$$

It incorporates the difference between the measured signal $S(\tau, \eta)$ and predicted background $B(\tau, \eta)$ summed over all τ_{dec} decoder steps of $\mathcal{S}_{\text{RNN}}$. An increasing S yields a higher TS fulfilling the critical condition. Further, we apply a filter on the minimum number of events. For each energy bin, we check if it contains at least two γ-like counts in two of the τ_{dec} steps. If this conditions is not matched for at least one energy bin, the TS for all energy bins is set to -1 which corresponds to a downward fluctuation. This prevents a possible detection with events in only few time bins which has a stabilizing effect on the analysis.
For the calibration we evaluate Equation 9.16 for the prepared background dataset. The complete data is grouped into the 24 run-wise and 5 ROI meta

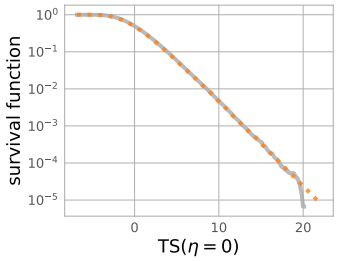


Figure 9.30: Survival function of the TS distribution in the first energy bin.

bins. In each meta bin, we calculate the survival function. Figure 9.30 shows an example for the first energy bin $\eta = 0$ of a single meta bin. These are parametrized using the the procedure described in subsection 9.3.1. Due to the filter on the number of γ-rays, an increased number of events might occur at TS $= -1$. This effect is in the range of the direct parametrisation of the TS. Thus, the potential step in the survival function is handled correctly without any influence on the log-linear fit. Evaluating the survival function yields the probability $p_{\text{single}}(\eta, r)$ at the first stage.

Combine Energy Bins

The goal of this second calibration stage is the combination of the three energy bins η. The intention behind this stage is that the transient signal might cause an excess in all three energy bins. Also the new test statistic has to increase for a stronger signal. As p_{single} decreases logarithmically, we can define a new statistic

$$\text{TS}_{\text{e}}(r) = -\sum_{\eta} \log_{10}\, p_{\text{single}}(\eta, r). \tag{9.17}$$

This test statistic is calculated in the same 120 meta bins as the first stage. Again, the survival function for the background data is parametrized to get the mapping to $p_{\text{e}}(r)$.

Trial Factors

Considering the trial factors is inevitable for calculating an accurate statistical probability. The calibration pipeline introduced above intrinsically takes care of much of this. The sliding window approach to convert $\mathcal{S}_{\text{p}}$ to $\mathcal{S}_{\text{RNN}}$ leads to the same time bin being probed τ_{dec} times. This sliding window is equally applied to the inference and the background dataset for calibration and training. As every statistical fluctuation is sliding through the complete decoder for the calibration, the distribution of TS values also includes the effects of multiple occurrence of each value. Hence, the mapped p-values are corrected for this effect. Equally, the combination of multiple TS values is done identical for both datasets. Thus, these possible trial factors are handled correctly with the mapping to the p-values.
The only effect not considered is the number of total time positions probed by the analysis. Below, we assume that the sought-for detectable signals have a duration of τ_{sig} steps. For the simulated signals, we fixed the total signal length to $\tau_{\text{sig}} = 5$ steps. In Figure 9.31, we highlight the position of the five signal steps by the orange circles.
Not every single time step is probed independently but the sum of all $\tau_{\text{dec}} = 5$ steps[5]. The four steps before and after this signal are illustrated by the blue circles. Every row represents one selected iteration of the sliding window. The purple area shows the position of the decoder window at each iteration. We label the shown iteration with $n = 1$ meaning the signal just enters the decoder, $n = \tau_{\text{dec}} = 5$ where the signal overlaps completely with the decoder, and $n = \tau_{\text{dec}} + \tau_{\text{sig}} - 1 = 9$ corresponding to the last sliding window with signal contributing to the decoder. We expect the highest significance of detection for one of the iterations that

5: See the fundamental test statistic defined in Equation 9.16.

covers the complete signal with the decoder. In this graph this corresponds to $n = 5$. This graph shows that a signal of length τ_{sig} is tested a total of $n_{\text{test}} = \tau_{\text{dec}} + \tau_{\text{sig}} - 1$ times.

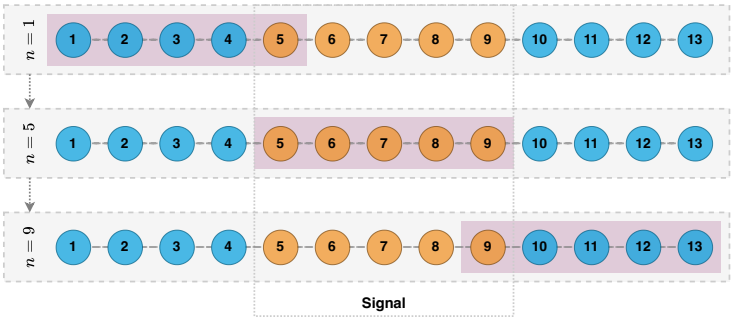


Figure 9.31: Location of the decoder window $\tau_{\text{dec}} = 5$ relative to signal with duration $\tau_{\text{sig}} = 5$ during different iterations of the sliding window n.

Above discussion assumes that the relevant duration of the signal is known a priori. In the case of PBHs, the expected strength of the signal is scaled with the distance of the PBH $N \propto d^{-2}$. For evaporations with distances $d \ll 1\,\text{pc}$, also detectable signals of more than five steps are expected[6]. However as the signals get fainter, only the last time bins might be relevant. In a consequence τ_{sig} can be flexible. In this work, we investigate the results manually to identify which candidate events are originated by the same or discrete events. A generalization of this approach is beyond the scope of this thesis.
We use the most conservative value for n_{test} to correct for the trial factors. This is done by assuming the shortest possible signal duration $\tau_{\text{sig}} \geq 1$. This leads a minimum value of $n_{\text{test}} \geq 5$. In a dataset with n_{total} probed series, the upper limit for the number of independently detectable signals is

$$n_{\text{trial}} = \frac{n_{\text{total}}}{n_{\text{test}}} \leq \frac{n_{\text{total}}}{5}. \tag{9.18}$$

The value of n_{total} incorporates the total observing time, the number of probed ROIs, and the duration of the timesteps.
In this work we define the detection of a transient source to be an outlier of $> 5\sigma$ from the background. This corresponds to a confidence interval of $\alpha_{\text{det.}} \sim 5.73 \times 10^{-7}$. We modify this confidence interval according to the Šidák correction [157]

$$\alpha_{\text{corr.}} = 1 - (1 - \alpha_{\text{det.}})^{1/n_{\text{trial}}} \tag{9.19}$$

in order to arrive at a global significance.

6: The strongest signals are close to the final evaporation of the PBHs. Thus, the highest detection power is during the last time steps. As only detection or non-detection is relevant in this work, simulating the last five steps is sufficient.

[157]: Šidák (1967), 'Rectangular Confidence Regions for the Means of Multivariate Normal Distributions'

10 Search for PBH Evaporation

In the previous chapters, we introduced all the necessary preparation to search for transient sources in archival VERITAS data using the deep learning transient detection. In this section, we apply this method to search for the evaporation bursts of PBHs. It is the first time in the curse of this thesis that we use the unshuffled dataset. Thus, in section 10.1, we validate its consistency to the scrambled data used for training and calibration. Using the simulated signals, we then calculate the effective search volume V_{eff} for PBH evaporation events. This V_{eff} and the observation time T_{obs} define the constraining power of this analysis. We discuss this estimation in section 10.2. After estimating the detection power, we apply the DL transient detection to the entire selected VERITAS dataset. We investigate these results in section 10.3. Finally, we present the conclusions and prospects of this study in section 10.4.

10.1 Validation of Shuffling

In section 8.4, we developed the approach for generating the required background dataset from real data. This background data is used during the training and calibration phases of the method. The objective of shuffling the γ-like events in time is to remove the contamination of potential transient sources. Detected transient sources correspond to a short excess of γ-like events. By scrambling the events, these are distributed over multiple time bins. Outliers with a large number of events are smeared. We introduced this concept in section 8.4. The key feature of our approach is a sliding shuffling window that uses a 300 s long range of the data. As discussed above, it should be able to flatten strong transient signals while still keeping the $\mathcal{O}$(min) long features of the data.

Figure 10.1 shows the distributions of event counts per energy bin η for the shuffled training data (blue) and inference data (orange). In this graph, we only consider periods used during both phases. Thus, the data are comparable. The peaks in the distribution at every multiple of 20 are due to the oversampling. If the initial event is in the middle of an ROI, all oversampled events likely end in the same ROI. Closer to the edge of the ROIs, the probability that fractions of the oversampled events end in different ROIs increases. The distributions are in good agreement, up to high counts in all three energy bins. The slight difference at the end of the distributions is the expected behavior.

Also, validating the data of individual runs does not show any differences. As we discuss in section 9.1, also additional parameters are used in the RNN. These are essential to describe the different observing conditions and influence the predictions of background counts. For completeness, we also investigate these for systematic differences. The used auxiliary parameters are the secant of zenith angle of observation $\sec(\theta)$, the absolute azimuth angle of observation $|\alpha|$, the reference time ΔT_{ref}, the L3-rate R_{L3}, the offset angle of the ROIs ζ, and the total multiplicity

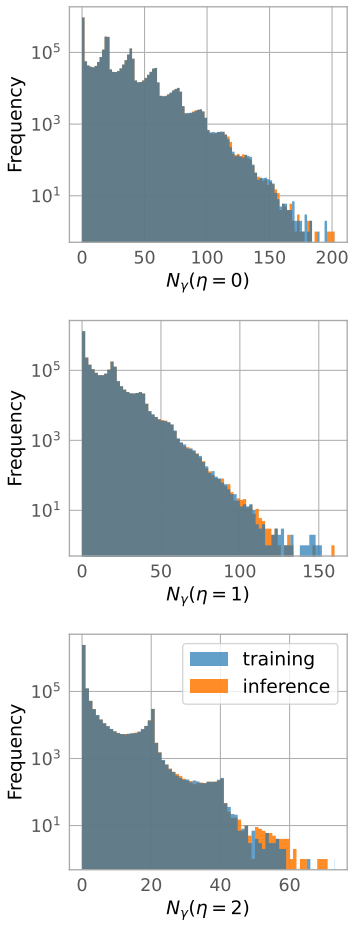


Figure 10.1: Distribution of γ-like event counts in the training (*blue*) and inference data (*orange*). Each graph represents one of the three energy bins η.

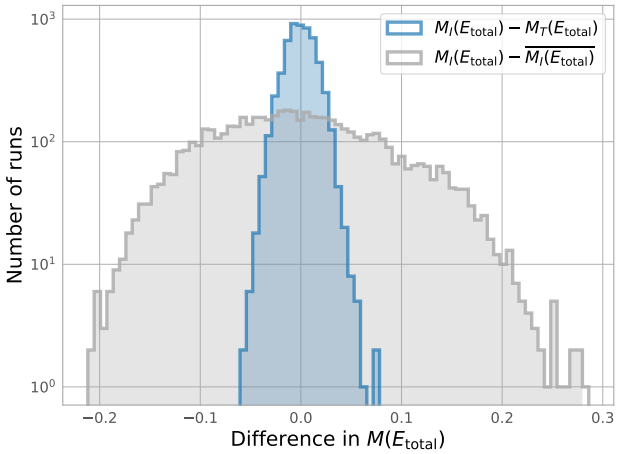


Figure 10.2: Difference in $M(E_{\text{total}})$ between training and inference data. The blue graph shows the distribution of the run-wise differences of the multiplicity between the inference data M_I and training data M_T. To put this difference into context, we also show the distribution of multiplicities in the unshuffled inference data in grey. For comparison, we subtract the mean value $\overline{M_I}$ from this data.

$M(E_{\text{total}})$. From these $\sec(\theta)$, $|\alpha|$, and R_{L3} are used as time dependent parameters.
Only the multiplicity $M(E_{\text{total}})$ is calculated directly from the events. As the composition of the data is slightly changed during the shuffling, the average multiplicity is also affected. We show this difference for all runs in Figure 10.2. The blue distribution shows the absolute run-wise difference of the multiplicities $M_I(E_{\text{total}})$ in the inference data to $M_T(E_{\text{total}})$ in the training data. To put these differences into context, the variations from one run to the other are relevant. Thus, the grey distribution shows the total spread of the multiplicities in the inference data set. For this comparison, we subtract the mean value over all observing runs in $\overline{M_I(E_{\text{total}})}$. The observed differences between M_I and M_T are due to statistical fluctuations in the shuffling. Compared to the spread between individual runs, this effect is usually small. The other five auxiliary parameters are independent of the γ-like events. Thus, the shuffling does not affect these. This study shows that the training and inference datasets are in good agreement.

10.2 Detection Efficiency and Effective Volume

The goal of this thesis is to constrain the abundance of PBHs. This section applies our implemented DL-based transient detection method to the simulated dataset. The goal is to determine the effective search volume V_{eff}, which is probed by this analysis. En route, we calculate the expected detection efficiency for PBHs.
Previously, we introduced the simulation of PBH evaporation signals for VERITAS. These are superimposed to a shuffled background dataset. We discussed this approach above in section 8.6. In this section, we apply the transient detection method to probe whether we can detect these signals. Contrary to the real analysis, we know that a source is present. For the simulations, we are only interested in whether a burst is detected or not. We calculate the significance of an event from the p_e-value with

$$\sigma_e = \text{isf}_{\text{norm}}\left(\frac{p_e}{2}\right), \tag{10.1}$$

where isf$_{\text{norm}}$ is the inverse survival function of the normal distribution, and p_e is the statistical probability assigned by the calibration after combining the individual energy bins. We consider a simulated PBH detected if we obtain a 5σ significance for any timestep and ROI of the prepared simulation data. Counting the fraction of simulated bursts that are detected yields the detection efficiency.

In section 7.2, we discussed that we perform a set of simulations for each meta bin. This way, the full range of observing conditions we consider in this analysis is represented. The background data into which we inject the signals are selected from the same meta bin. It is prepared using identical data selection, ROI settings, and exclusion regions that are also applied for the inference data. Thus, the superimposed dataset represents all properties of the actual data. Applying the procedure to each meta bin gives an accurate average of the detection efficiency for the given observing conditions of this bin. The PBH distance d and location ζ are known for the simulations. We calculate the fraction of detected PBH bursts $f_m(d, \zeta)$ as a function of the source parameters and the meta bin m. We represent the 24 run-wise meta bins by a tuple $m = (m(|\alpha|), m(\Delta T_{\text{ref}}), m(\sec(\theta)))$. These bins are summarized in Table 9.6. The range of each bin is $m(|\alpha|) \in (0, 1)$, $m(\Delta T_{\text{ref}}) \in (0, 1, 2)$, and $m(\sec(\theta)) \in (0, 1, 2, 3)$.

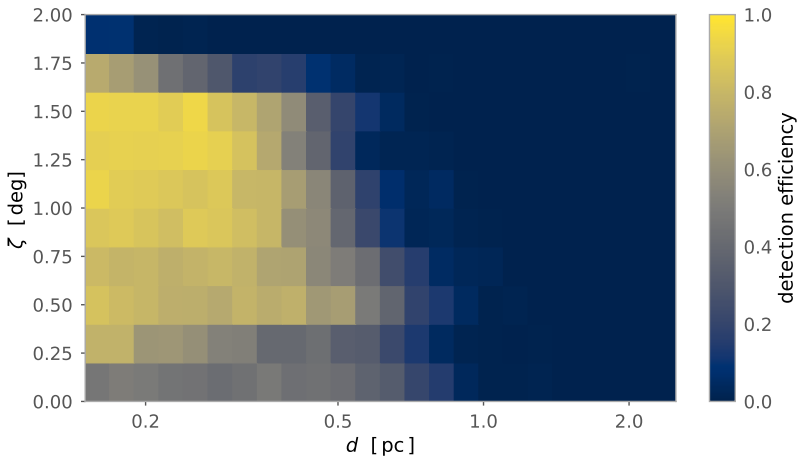


Figure 10.3: Detection efficiency $f_{(0,0,0)}(d, \zeta)$ for meta bin $m = (0, 0, 0)$ as a function of the PBH distance d and offset angle ζ.

Figure 10.3 shows the calculated detection efficiency $f_{(0,0,0)}(d, \zeta)$ in meta bin $m = (0, 0, 0)$. We obtain a better detection efficiency for events at small distances d. This is expected as the detectable signal scales with d^{-2} Further, we find that the detection efficiency sharply decreases at a given distance $d_{\text{max}}(\zeta)$. In general, observations closer to the camera center, small ζ, yield a larger d_{max}. This reflects the higher sensitivity closer to the camera center, which is crucial for detecting faint signals.

We note, that at $d \lesssim 0.5\,\text{pc}$ the detection efficiency decreases for $\zeta \to 0°$. This might seem non-intuitive as the sensitivity increases closer to the center of the FoV. The origin of this effect is the definition of exclusion regions. Two reasons need to be considered for this interpretation. The first is a geometrical effect. When an exclusion region and PBH evaporation occur at a small ζ, the probability of overlap is considerable. In the case of $\zeta = 0°$, it reaches 100%. However, further away from the center, the probability of spatial coincidence decreases. The second reason is the definition of exclusion regions around sources. Typically, VERITAS observes sources at an offset angle of $\zeta = 0.5°$. We use a circular exclusion

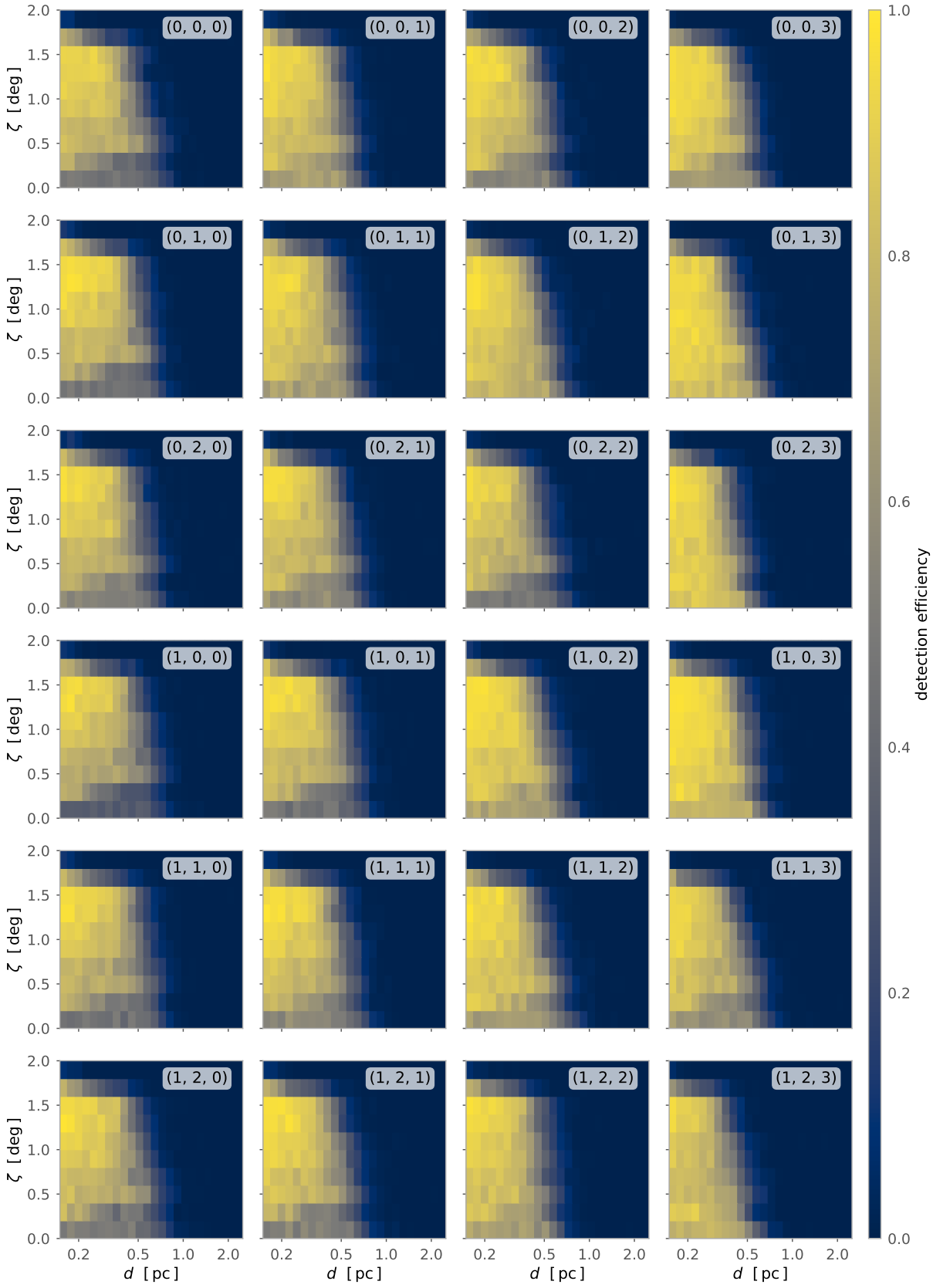


Figure 10.4: Detection efficiency with exclusion regions for the 24 run-wise meta bins as a function of the PBH distance d and offset angle ζ. The meta bins are indicated by the numbers in the upper right corner of each graph. This tuple indicates the meta bin in the azimuthal observation angle $m(|\alpha|)$, the reference time $m(\Delta T_{\text{ref}})$, and the secant of the zenith angle of observation $m(\sec(\theta))$.

region with a radius of 0.35° around VHE γ-ray sources detected with VERITAS. This is sufficient to overlap with more than 10% to multiple ROIs. All of these ROIs are excluded from the analysis. This also affects the only ROI positioned at center $\zeta = 0°$.[1] Having this ROI excluded for many observations limits the detection efficiency close to the camera center.

1: E.g. see the example in Figure 8.3b.

We show the detection probabilities f_m for all meta bins in Figure 10.4. As expected, the detection efficiency depends on the observing conditions represented by the meta bins. Each of these graphs can be interpreted as a representative of the mean performance for the corresponding meta bin. In general, the efficiency to detect faint events at lager d decreases for lower zenith angles $m(\sec(\theta)) \rightarrow 3$. This is due to the increase in the energy threshold, which causes a decrease in low-energy events. Observations taken in more recent data, where the throughput is lower, $m(\Delta T_{\text{ref}}) \rightarrow 2$, increase this effect. Also, a relative decrease of f_m due to the exclusion regions changes from one meta bin to the other. E.g. for $m = (0, 2, 3)$ we barely observe this effect. This is because the fraction of runs with detected sources is small in the selected runlist.

Having determined $f_m(d, \zeta)$, we can calculate the effective search volume $V_{\text{eff}}(m)$. It is the integral of the detection efficiency in spherical coordinates

$$V_{\text{eff}}(m) = \int_0^{2\pi} \int_0^{\infty} \int_0^{\infty} f_m(d, \zeta) d^2 \sin(\zeta) \, \mathrm{d}d \, \mathrm{d}\zeta \, \mathrm{d}\phi \tag{10.2}$$

$$= 2\pi \int_0^{\zeta_{\text{max}}} \int_0^{d_{\text{max}}} f_m(d, \zeta) d^2 \sin(\zeta) \, \mathrm{d}d \, \mathrm{d}\zeta, \tag{10.3}$$

where ϕ is the polar angle around the center of the FoV. In our case, the limits ζ_{max} and d_{max} correspond to the maximum values we simulated. Based on expectations from Figure 10.3, we choose $\zeta_{\text{max}} = 2°$ and $d_{\text{max}} = 2.5\,\text{pc}$. Figure 10.4 shows that at both of these maximum values, no further PBH bursts are detected. Thus, increasing the maximum values would be redundant.

We perform this integration numerically using the detection probabilities shown in Figure 10.4 The distance d of the simulated PBHs is in 25 logarithmic steps from 0.15 pc to 2.5 pc. We perform the integration of Equation 10.3 over d in logarithmic scale. The offset angle ζ has 11 linear values between 0° and 2°. We use a two-dimensional interpolation of the detection efficiency to obtain arbitrary values for the integration. We also extrapolate the graphs to distances lower than 0.15 pc. For this we use the constant values of $f_m(\zeta, d = 0.15\,\text{pc})$. It is a conservative assumption, as the number of γ-rays reaching the detector increases with d^{-2}. On the other hand, the exclusion regions remain a limiting factor independent of d. However, when solving the integration of a cone, the exact values of f_m at small distances have a minor impact on the total results.

We show the results for the effective volumes in for all 24 run-wise meta bins m in Figure 10.5. Typically, the volume is larger for small zenith angles of observations ($m(\sec(\theta)) \rightarrow 0$) and pointing south $m(|(|\alpha)) = 1$. This is an expected trend, as these observations typically correspond to a higher sensitivity of VERITAS. The differences between the T_{ref}-meta bins are most evident for observations at large zenith angles. Among the meta bins, the values of effective volume vary by a factor up to ~ 2.3.

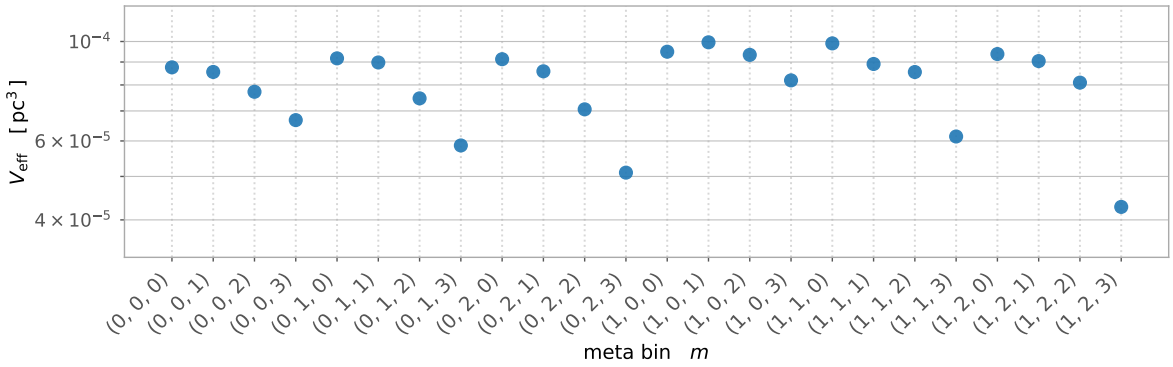


Figure 10.5: Comparison of the effective search volume V_{eff}^m for all meta bins m. The tuples m consist of the bins in the azimuthal observation angle $m(|\alpha|)$, the reference time $m(\Delta T_{\text{ref}})$, and the secant of the zenith angle of observation $m(\sec(\theta))$.

10.3 Constraining the Rate of PBH Bursts

In this section, we run the inference stage of the DL-based transient detection. It is performed on the unscrambled dataset we prepared in Chapter 8. We consider observations taken between 2012 and 2021 for this thesis. The data is selected using the automatic data quality assessment developed in Chapter 6. We summarize the selection of the data in subsection 6.3.5.

During this work, we have set one of the 9496 runs aside to run some validation. The selected run is located within meta bin $m = (0, 2, 0)$. In the following, it is removed from the dataset. The total remaining observing is 4222.36 hours. For a fixed effective search volume, the expected number N of detecable bursts scales linearly with the observation time T_{obs}. The total observation time is distributed over all 24 run-wise meta bins. Figure 10.6 shows the observing time T_{obs}^m per meta bin m with orange squared markers. We also show V_{eff}^m which we introduced before in Figure 10.5 in blue.

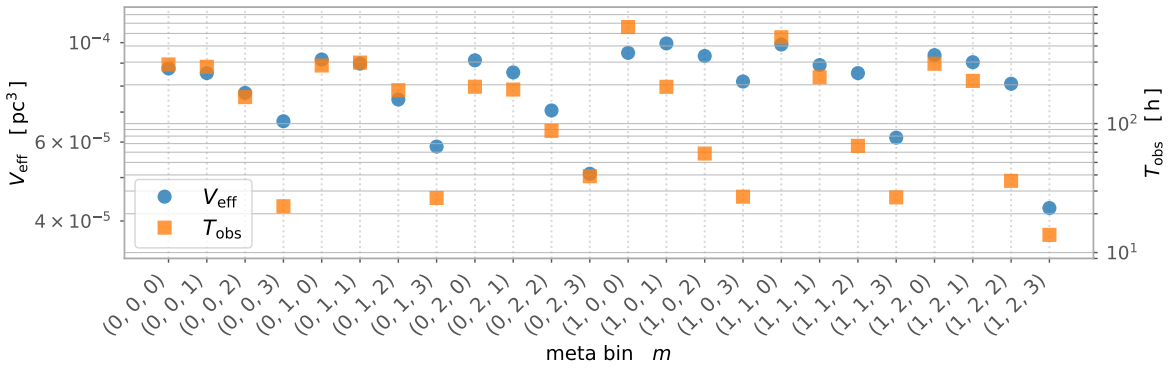


Figure 10.6: The orange squared markers show the observation time T_{obs}^m per meta bin m. The blue points illustrate V_{eff}^m which we introduced in Figure 10.5. The tuples m consist of the bins in the azimuthal observation angle $m(|\alpha|)$, the reference time $m(\Delta T_{\text{ref}})$, and the secant of the zenith angle of observation $m(\sec(\theta))$.

10.3.1 Calculation of Upper Limits

Having the effective search volume and the observing time, the number of expected events in one meta bin is

$$N^m = r_{\text{burst}} V_{\text{eff}}^m T_{\text{obs}}^m, \tag{10.4}$$

where r_{burst} is the local rate of PBH evaporations per volume. For the full dataset, the expected number is

$$N = \sum_m N^m = r_{\text{burst}} \sum_m V_{\text{eff}}^m T_{\text{obs}}^m. \tag{10.5}$$

We aim to constrain the rate r_{burst} which we can calculate from the number of observed events

$$r_{\text{burst}} = \frac{N}{\sum_m V_{\text{eff}}^m T_{\text{obs}}^m}. \tag{10.6}$$

The observed number of counts follows a Poisson distribution. The probability of obtaining N detections is denoted by $P(N|\lambda)$, where λ is the mean value. To calculate an upper limit, we set this probability to $1 - c$, where c is the confidence level

$$P(N|\lambda) = \frac{\lambda^N e^{-\lambda}}{N!} \equiv 1 - c. \tag{10.7}$$

Empirically solving this equation for λ yields the upper limit on the actual number of counts $\text{UL}_c(N)$. We calculate upper limits at the 0.99 and 0.95 confidence intervals in this work. We show the value of $\text{UL}_{0.99}(N)$ as a function of the detected number of events N in Figure 10.7. In case no events are observed the limits are $\text{UL}_{0.99}(0) \approx 4.6$ and $\text{UL}_{0.95}(0) \approx 3.0$. The upper limit of the sought-for rate is

$$r_{\text{burst}} < \frac{\text{UL}_c(N)}{\sum_m V_{\text{eff}}^m T_{\text{obs}}^m}. \tag{10.8}$$

In the next subsection, we investigate the results of the DL-transient search to determine the number of detected candidates N.

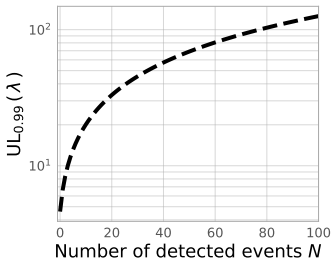


Figure 10.7: The upper limit on the number of detected counts λ as a function of the number of observed events N at a confidence interval $c = 0.99$.

10.3.2 Search for PBH Bursts

In this section, we investigate the results for the full dataset. We show the distribution of σ^2 Figure 10.8. The first three graphs show the results of the single energy bin calibrations $\sigma^2_{\text{single}}(\eta)$. The graph on the right shows the calibrated results of the combination of the three energy bins σ^2_{e}.
For a dataset with pure background, the σ^2 distribution should follow a χ^2-distribution with one degree of freedom. The grey dashed line shows this expected curve. The blue distribution presents the shuffled training dataset. It follows the expected χ^2-distribution. Having these distribution matching confirms the validity of the calibration approach. However, the slopes of all four σ^2 distributions for the unscrambled inference data are smaller than for the training data. We expect the inference data also to be dominated by background. Thus, except for outliers due to transient sources, it should also follow a χ^2-distribution. The difference indicates a systematic difference between the two datasets created during the

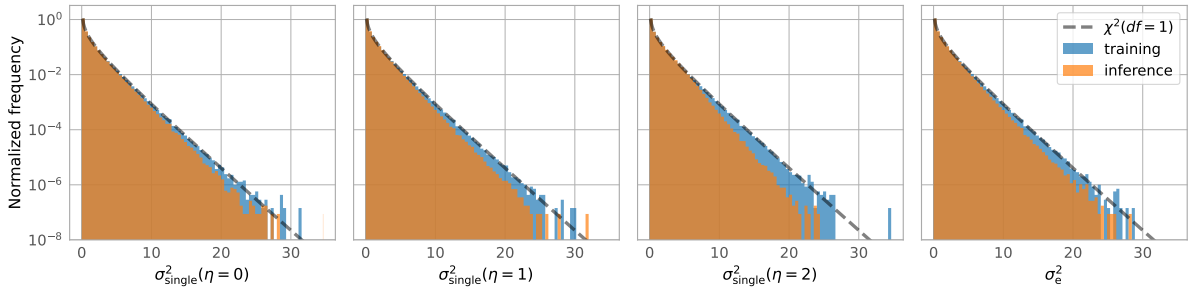


Figure 10.8: Distribution of σ^2 for all calibration stages. The three graphs on the left show the single energy bin calibration $\sigma^2_{\text{single}}(\eta)$. The graph on the right presents the combination of these three energy bins, σ^2_{e}. For a background dataset, this distribution should follow a χ^2-distribution with one degree of freedom. The grey dashed line shows this reference. The blue distribution shows the significance distribution for the shuffled training data set, which matches the χ^2-distribution. The results for the inference data are shown in orange. We expect that the inference data is also dominated by background and, except for outliers, follows the χ^2-distribution. However, the slope is significantly lower, indicating that the significances are systematically too low.

shuffling approach. Having systematically lower significances σ for the inference means that the search is too conservative.

When comparing the three energy bins $\eta \in \{0, 1, 2\}$, the mismatch is more evident in the high-energy bins. The main difference between these energy bins is the total number of events, which decreases for higher energies. This indicates that the uncertainty might be related to the scrambling approach in sparsely populated data. While validating the data preparation in section 10.1, we did not find a noticeable difference between the training and prediction data. All auxiliary parameters are identical and not modified during the training. The only minor difference we could identify is that outliers with many counts are reduced during the shuffling. However, as the objective of the shuffling is to remove potential transient signals, this is the sought-for behavior. The level of this effect can not explain a systematic difference on the scale we observe here. Further, by reducing the counts, we would naively expect smaller significances for the training data compared to the inference data. However, Figure 10.8 shows that the results during the inference phase have lower significances.

The results at the first three energy bins $\sigma^2_{\text{single}}(\eta)$ are directly given by the differences between the predicted counts of γ-like events by the long-short-term memory and the input counts. The LSTM considers the sequence of counts in the time series. Thus, a subtle difference between the datasets' temporal composition might influence the predictions. For this thesis, we consider this as a source of systematic uncertainty. Future work should investigate this difference further.

For this thesis, we proceed with the results that include this systematic uncertainty. To estimate its effects on this analysis, we investigate the results of the last stage of the calibration pipeline σ_e. We show the survival function of the significance distribution in Figure 10.9. The orange curve represents the significance distribution of the inference data. Assuming it is mostly background-dominated, it should follow the SF of a folded normal distribution with $\mu = 0$ and $\sigma = 1$.[2] We show the expectation with the dashed grey line. Further, we fit the SF of a folded normal distribution to the inference data and obtain values of $\mu = 0.43$ and $\sigma = 0.88$. The dotted black curve shows this parametrization. The difference in μ affects mostly low significance values and thus is less relevant for detecting transient sources. The distribution spread of < 1

2: Using a folded normal distribution corresponds to the fact that our analysis only assigns positive significance values.

is due to the underestimated significance. This difference dominates at high-significance values.

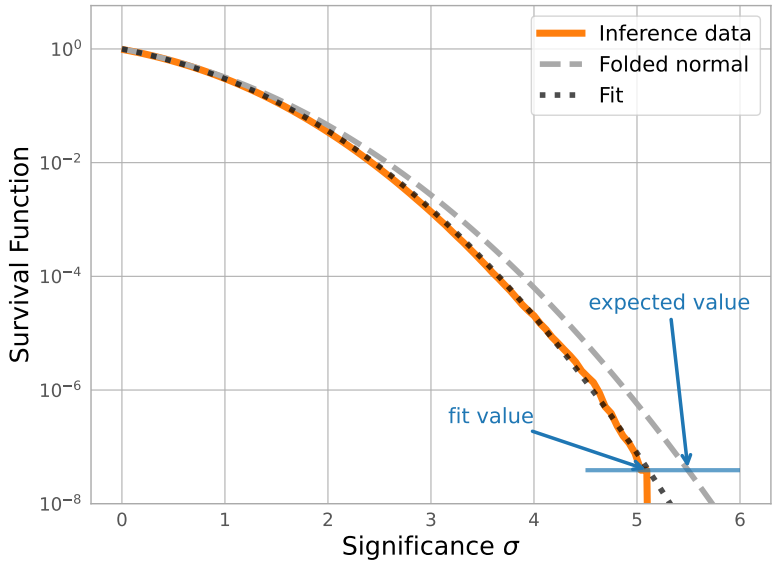


Figure 10.9: Survival function of the significance distribution. The orange curve survival function of all significances obtained in the inference dataset. We parameterize it with a folded normal distribution which is shown by the black dotted curve. The expected distribution is a folded normal distribution with $\mu = 0$ and $\sigma = 1$. The dashed grey line shows this expectation. The blue arrows indicate the positions of the quantiles $q = 1/n_{\text{total}}$ for both folded normal distributions. The difference of the significances is $\Delta\sigma_e \approx 0.38$.

In this thesis, we aim to detect transient signals. For this purpose, we are particularly interested in the systematic uncertainty for the most significant events. The total number of probed series is $n_{\text{total}} = 25712614$. Thus, the lowest value in the survival function is $q = 1/n_{\text{total}}$. We evaluate the inverse survival function (ISF) for the parametrized end expected distribution at ISF(q). For the fitted curve we find $\sigma_{\text{fit}} \approx 5.11$ and the expectation is $\sigma_{\text{exp}} \approx 5.50$. We indicate these two values by the blue arrows in Figure 10.9. The difference is $\Delta\sigma_e \approx 0.38$. By using the smallest possible quantile q, $\Delta\sigma_e$ corresponds to the largest difference in the significance. In the following, we use it as an estimate for the maximum effect this uncertainty can have on the analysis results.

Before trial factor correction, we have two significant detections with $\sigma_e > 5$. The first occurs during VERITAS run 84091 with significance $\sigma_e = 5.18$. The second pre-trial detection is during run 88737 with $\sigma_e = 5.05$. After correcting for the systematic uncertainty $\Delta\sigma_e$, these two signals correspond to 5.56σ and 5.44σ, respectively. As discussed in the next paragraph, after correcting for trial factors in this blind search, these detections are not relevant in terms of PBHs. Nonetheless, these events might be interesting for further investigations.

We discussed the calculation of the global significance in section 9.3.3. Starting from the total number of probed series, $n_{\text{total}} = 25712614$, we use Equation 9.18 to estimate the upper limit on the number of trial factors $n_{\text{trial}} = 5142523$. We use the inverse of Equation 9.19 to calculate the p_{pt}-values post-trial factor correction.

$$p_{\text{pt}} = 1 - (1 - p_e)^{n_{\text{trial}}}. \tag{10.9}$$

Using p_{pt} in Equation 10.1 yields the global significance of the events. The largest post-trial significance is at 0.41σ. Including the correction $\Delta\sigma_e$ for the systematic uncertainty, its global significance is 1.52σ. We conclude, that after correcting for the trial factors, no evaporating PBH was identified in the data set. The scientific interpretation of this results

is independent of correcting for the systematically lower significances. In the following we use $N = 0$ to constrain the local rate of PBH evaporations. Using the observing time and effective search volumes, which we show in Figure 10.6, the upper limit at the 99% confidence interval is

$$r_{\text{burst}}(99\%\,\text{CL}) < 1.07 \times 10^5\,\text{pc}^{-3}\text{yr}^{-1}. \tag{10.10}$$

For comparison to existing limits, we also calculate the 95% confidence upper limit, which yields

$$r_{\text{burst}}(95\%\,\text{CL}) < 6.97 \times 10^4\,\text{pc}^{-3}\text{yr}^{-1}. \tag{10.11}$$

We show a comparison of this constraint with previously existing limits in Figure 10.10.

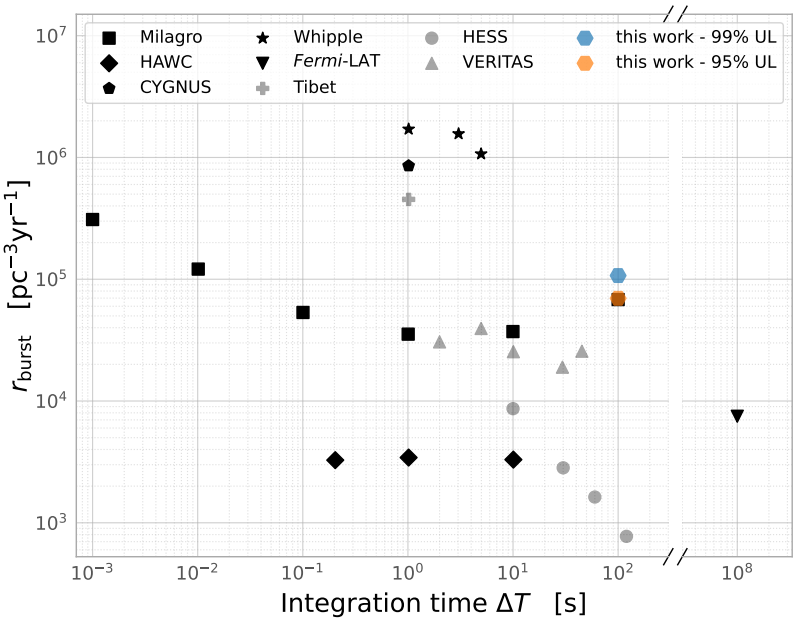


Figure 10.10: Upper limits on the local rate of PBH evaporation bursts. The black and grey markers, respectively, show peer and not peer reviewed published limits. The results of HAWC [39], CYGNUS [40], *Fermi*-LAT [42], Tibet air shower array [43], H.E.S.S. [44], VERITAS [45] show the upper limits at the 99% confidence interval. The limits from Milagro [38] and Whipple [41] show the 95% confidence interval. The blue and orange markers show the constraints we derive with this analysis at both confidence limits.

The traditional approach of existing studies uses a fixed window with duration ΔT to search for bursts of events. The distributions of the number of counts within this time window are compared to a shuffled background dataset. If PBH evaporations are present, it might cause a detectable excess of counts in this distribution. Our analysis uses an alternative approach in which we do not directly compare the number of counts in a fixed window. For comparison, we illustrate our results at $\Delta T = 100\,\text{s}$. This corresponds to the duration of the search window, which is given by the number of decoder steps $\tau_{\text{dec}} = 5$ and the duration of the time bins $T_{\text{step}} = 20\,\text{s}$. However, a significant excess in one of the decoder steps might also be sufficient to cause a detection. Thus, the duration of the signals can be flexible.

10.4 Summary and Outlook

The results we present in subsection 10.3.2 are the first results for real data obtained with this DL-based transient detection method. We have defined many novel concepts for this approach which are tested by this

first analysis. One strength of our analysis is the small number of required assumptions. Due to the data-based nature, it does not depend on modeling the effective area and the radial acceptance. The characteristics of the instrument response are directly learned from the data. Usually, generating effective areas and radial acceptances is computationally heavy and requires careful validation. Removing the need for these makes changing the cuts of the γ/hadron separation light. Thus, these can be optimized easily for the sought for science case.
In this thesis, we implemented an anomaly detection. Its goal is to detect a transient signal as the divergence from the background. Thus, its application is not limited to PBHs. Only a few hyperparameters of this search, such as the length of the decoder τ_{dec} and the duration of the shuffling window, are directly motivated for the specific science case of PBHs. However, the transient search itself does not require assumptions about the source's temporal or spectral model. Thus, this DL based transient detection method might also be used to search for other transient sources, such as γ-ray bursts or flaring blazars, in the future.
The limits we set on the rate of PBH evaporations are in the same order of magnitude as existing limits for VERITAS and Milagro. Compared to existing VERITAS limits, the dataset used in this analysis is considerably larger. Previous work used 747 hours of observations. Our dataset includes a total of 4222.36 hours. However, this work presents the first application of this method and leaves room for future improvements beyond the scope of this thesis.
The DL-based transient detection method provides an independent analysis method to search for PBH bursts. Contrary to the existing approach to searching for the PBH bursts, we can assign an event-by-event significance. It allows following up on individual events which might be detected during the blind search. This might provide a deeper understanding of the nature of the transient signals. In the context of PBHs, it might allow probing the theory of Hawking radiation for individual bursts.
For future work, our analysis method has significant room for improvement. In the following, we will discuss future points that might be considered.

- First, the differences in the significance distributions we showed in Figure 10.8 require further attention. In the current state, the significance of the inference is systematically underestimated. As discussed above, it hints at a subtle difference between the shuffled training and inference data. For this thesis, we have validated that the scientific interpretation is independent of this uncertainty. However, in future work the source of systematic uncertainty should be investigated more carefully. Solving the mismatch will lead to a more accurate, higher significance during the inference.

- In the future, the hyper-parameters of this analysis might be optimized. Among these is the number of decoder steps τ_{dec}. In this thesis, we used a decoder with duration 100 s. The strongest signal from PBH bursts has a duration of $\mathcal{O}(s)$. Thus, it might be interesting also to test shorter decoder windows. Optimizing the search window might improve the signal-to-noise ratio. It roughly corresponds to varying ΔT during the traditional PBH searches.

- A further improvement for future analyses might be more sophisticated calibration stages. Currently, every ROI is treated independently. However, signals from one PBH burst might be distributed into multiple neighboring ROIs. We mentioned the possibility of combining neighboring ROIs in subsection 9.3.3. In the blind search, the source location can be anywhere in the FoV. There are three probable cases of the location relative to the ROIs. If most of the signal is in one ROI, using a group of the single ROI ($g = 1$) should be optimal. If the source is located between two ROIs, the best results are expected by combining these two ($g = 2$). Finally, it might also be located between three ROIs in which $g = 3$ is preferred. A priori, it is unclear which group g yields the best performance. However, all three groups can be probed for every individual ROI r using

$$\mathrm{TS}_g(r) = \max_{c \in c_g(r)} \left(-\frac{1}{g} \sum_{r' \in c} \log_{10} p_{\mathrm{e}}(r') \right). \tag{10.12}$$

  Here $c_g(r)$ is the set of combination which include ROI r and $g - 1$ of its not excluded neighbors. For $g = 1$, the group only consists of r itself. The typical calibration approach can map these test statistics to probabilities $p_g(r)$. Eventually, the value which yields the best results is chosen

$$\mathrm{TS}_b(r) = \max_{g \in \{1,2,3\}} \left(-\log_{10} p_g(r) \right). \tag{10.13}$$

  By mapping this test statistic to $p_b(r)$, we obtain a probability that incorporates the three likely cases for the relative location of the signals. It might significantly improve the sensitivity to transients that are in the gap between ROIs. We especially expect an improvement for signals which are currently not detectable when split into multiple ROIs. This might improve the detection efficiency for faint PBH bursts close to $d_{\max}(\zeta)$. Due to the d^2-term in the integrand of Equation 10.3, improvements of f_m at large distances d have a particularly large influence to the effective volume V_{eff}.

- In section 8.1, we define the regions of interest as circular regions. With this configuration, not the entire FoV is covered by the ROIs. Due to the oversampling approach introduced in section 8.5, γ-like events falling in gaps can also partly contribute to the event counts. However, it might be beneficial to consider hexagonal ROIs instead. These would fill the FoV without gaps. Thus, no γ-like events are lost in the first place.

- Figure 10.3 showes that the exclusion regions significantly affect the detection efficiency for PBHs. For this thesis, we conservatively choose the size of the exclusion regions. It might be possible to improve the detection efficiency by selecting these more carefully in the future. Especially not all detected sources need to be removed, but only those with a significant excess on short time scales. Also, the exclusion around stars might be optimized to leave a larger fraction of the ROIs. For this consideration, the brightness of the γ-ray sources and stars is critical. The effects of reducing or removing exclusion regions might be accounted for by additional auxiliary

parameters, such as the distance to a source or star.

This thesis presents the first search for transient signals using this DL based transient detection method on actual observation data. It defines the critical concepts of how to conduct the search and implements the entire data preparation pipeline for VERITAS. This allows for running the different stages of the analysis and also testing possible improvements efficiently. Thus, we anticipate faster development circles moving forward.

Optical Microlensing by Primordial Black Holes with IACTs

11

In section 2.3, we discussed the possibility of PBHs in the sublunar range ($10^{17} \lesssim M/\mathrm{g} \lesssim 10^{23}$) contributing a substantial fraction of the DM. At the upper end, this window is constrained by microlensing studies. These works search for events in which the BH passes the line of sight (LoS) to a background star. During this process, the PBH acts as gravitational lens, causing a temporal enhancement of the measurable signal. The optical microlensing currently constrains over eleven orders of magnitude in the BH mass. We show a simplified version of the PBH microlensing limits by the blue regions in Figure 11.1. The MACHO and EROS experiments both monitored stars in the Large Magellanic Cloud with roughly 24 h cadence constraining the mass range of $[10^{-7}, 10]\ M_\odot$ [158, 159]. The first two years of Kepler data, monitoring Galactic stars with a sampling speed of 30 min, probe PBH masses down to $10^{-8}\ M_\odot$ [160]. The Subaru Hyper Suprime-Cam (HSC) constrains the abundance to $10^{-11}\ M_\odot$ with observations of Andromeda with a cadence of 2 min [161]. The original limits are shown by the dashed line. However, later work pointed out that with an updated distribution of the star sizes, these results are only valid to masses $> 10^{-10}\ M_\odot$ [162]. The updated results are illustrated by the solid blue line.

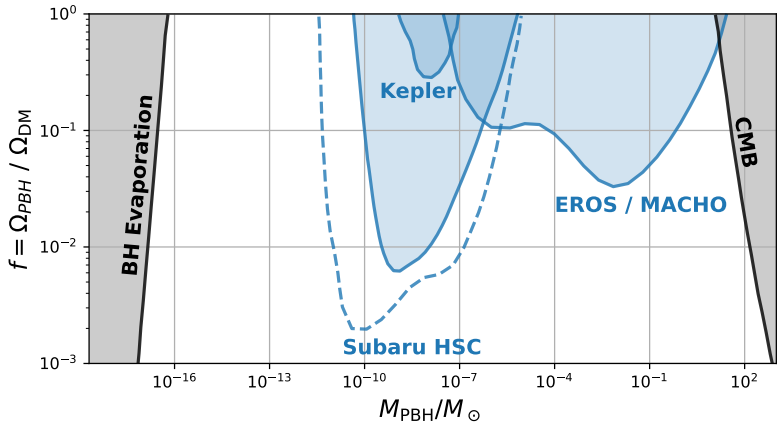


Figure 11.1: Clipped overview of the abundance of non-evaporation PBHs. Microlensing constraints are shown in blue. [163]

At their low mass range, above works are usually limited by the sampling speed, reaching from 2 min to several hours. As discussed in section 3.2, IACTs are optical telescopes optimized for ultra fast $\mathcal{O}$(ns) signals in the near-UV band. While the optical quality is modest compared to traditional optical instruments, the fast sampling speed sets them apart for fast optical astronomy. As proved in previous works [164, 165], they provide excellent capabilities to monitor fast brightness changes of astrophysical sources. In this chapter, we investigate the capabilities of IACTs to constrain low mass PBH $< 10^{-10}\ M_\odot$. In section 11.1, we review the critical formalism of optical microlensing. We study the possibilities of IACTs constraining the currently sublunar mass range in section 11.2.

[158]: Alcock et al. (2000), 'The MACHO Project: Microlensing Results from 5.7 Years of Large Magellanic Cloud Observations'

[159]: Tisserand et al. (2007), 'Limits on the Macho Content of the Galactic Halo from the EROS-2 Survey of the Magellanic Clouds'

[160]: Griest et al. (2014), 'Experimental limits on Primordial Black Hole Dark Matter From the first 2 yr of Kepler Data'

[161]: Niikura et al. (2019), 'Microlensing constraints on primordial black holes with Subaru/HSC Andromeda observations'

[162]: Smyth et al. (2020), 'Updated constraints on asteroid-mass primordial black holes as dark matter'

[164]: Hassan et al. (2019), 'Proving the outstanding capabilities of Imaging Atmospheric Cherenkov Telescopes in high time resolution optical astronomy'

[165]: W. et al. (2019), 'Direct measurement of stellar angular diameters by the VERITAS Cherenkov telescopes'

11.1 Optical Microlensing

Optical microlensing occurs when a small massive object is passing through close to the LoS between the observer and a background star. In this process, the massive object, i.e. the PBH, acts as a gravitational lens. In case of perfect alignment of star and lens, the Einstein ring can be observed. Its size is described by the Einstein radius r_E which is the characteristic distance for gravitational lensing. It depends on the relative distances between the lensing PBH and the background star to the observer $x = d_{PBH}/d_*$ and the mass of the lens M_{PBH},

$$r_E(x) = \sqrt{\frac{4G}{c^2} M_{PBH} d_* x(1-x)}. \tag{11.1}$$

It is appropriate to measure distances in units of r_E. Accordingly, the projected star radius is given by

$$U_* = \frac{Rx}{r_E(x)}, \tag{11.2}$$

where R is the radius of the star. The instantaneous distance of the PBH to the LoS at time t is described by the impact parameter

$$u(t) = \sqrt{u_{min}^2 + \left(\frac{2(t-t_0)}{t_E}\right)^2} \tag{11.3}$$

where u_{min} is the closest approach of the lens to the LoS at time t_0, and $t_E = 2r_E/v_T$ is the time for a lens to cross the Einstein ring with transverse velocity v_T.

The gravitational lensing creates multiple images of the same source. Below we assume circular sources with a constant surface brightness. Depending on the values of u relative to U_*, the shape of images is either in a ring, two arcs or a slightly deformed ellipse. Example of the first two cases are shown in the small graphs in Figure 11.2. The shape of the lensed images are described by equation (3) of [166]. These are shown by the blue shapes. The non-lensed images of the source with radius $U_* = 0.1$ are shown by the black shapes below. The shown images correspond to the time of the closest approach u_{min} for three different values 0.02, 0.09, and 0.2. Due to the small extent of the lens, the shapes are not individually resolvable. However, the observed solid angle of the sources increases. This effect leads to an apparent amplification of the source that is the quotient between the area of the blue lensed image and the initial star size.

[166]: Witt et al. (1994), 'Can Lensed Stars Be Regarded as Pointlike for Microlensing by MACHOs?'

The amplification A_{FS} for a finite source is described by equations (9)-(11) of [166]. We show the light curves with the blue lines in Figure 11.2. As comparison we also show the expected curves A_{PS} for the point source limit by the grey dashed lines. For small impact parameters, $u_t \ll U_*$, the point source causes a larger amplification. At $u_t \approx U_*$, the finite source results are slightly larger. Finally, at $u_t \gg U_*$ the differences are only marginal. The detectable duration of such events also depends on the sensitivity A_{thresh} of the instrument to detect an amplification. This value directly yields the observable duration T_{obs} at which $A > A_{thresh}$. For the examples, we show T_{obs} by the dotted horizontal lines assuming that an amplification by a factor 4 would be detectable for the given source

[166]: Witt et al. (1994), 'Can Lensed Stars Be Regarded as Pointlike for Microlensing by MACHOs?'

by an arbitrary experiment. A higher sensitivity of the instrument also increases the expected event duration.

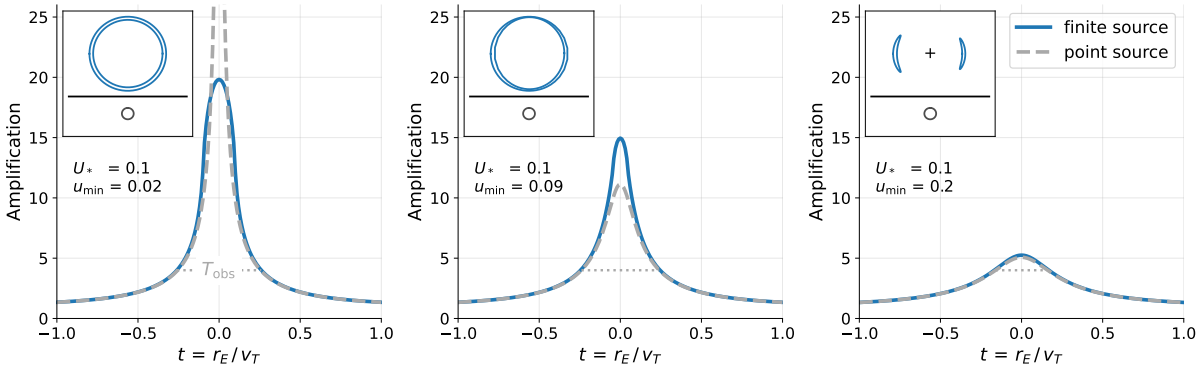


Figure 11.2: Apparent light curves during microlensing events with different minimum impact parameter $u_{\min}$. Blue shows the finite size light curves for a source with $U_* = 0.1$. The grey dashed curve shows the light curves for the assumption of a point source. The boxes in the upper left of each graph show the lensed (blue) and initial (black) images of the target star at the point of closest approach $u_{\min}$.

As $U_* \propto M_{\mathrm{PBH}}^{-0.5}$, the finite source effects are most critical for small PBH masses. This study targets the currently unconstrained range $M_{\mathrm{PBH}} < 10^{-10} M_\odot$. Thus, we only consider this more realistic finite source case. In this case, the maximum reached amplification $A_{\mathrm{max,FS}}$ with the lens directly at the LoS ($u = 0$) depends on the projected star radius

$$A_{\mathrm{max,FS}} = \frac{4 + U_*^2}{U_*}. \tag{11.4}$$

As the PBH is closer to the star, $x \to 1$, the maximum amplification decreases, $A_{\mathrm{max,FS}} \to 1$. This yields a maximum distance x_{max} at which the instrument is just sensitive enough to detect this event. For heavy PBHs, $U_* \to 0$ and thus $x_{\mathrm{max}} \approx 1$. In this case, the microlensing event would be detectable at any arbitrary position along the LoS. However, for lighter PBHs the maximum value is proportional to the mass $x_{\mathrm{max}} \propto M_{\mathrm{PBH}}$. The position of the transition between the constant and mass dependent x_{max} depends on the star parameters R, d_*, and magnitude m as well as the instrument's sensitivity A_{thresh}. For the best selected target star and the sensitivity of VERITAS at a 50 Hz sampling, the transition is around $M_{\mathrm{PBH}} \approx 10^{-8}\, M_\odot$[1]. The optical depth for the finite source limit is calculated by [167]

$$\tau_{\mathrm{FS}} \approx \rho_0 \frac{\pi}{3} \frac{d_* R^2 x_{\mathrm{max}}^3}{M_{\mathrm{PBH}}}, \tag{11.5}$$

where $\rho_0 \approx 7.9 \times 10^{-3} M_\odot/\mathrm{pc}^3$ is the local dark matter density. Here we assume a constant DM density along the LoS. This is appropriate as the relevant target stars for IACTs are close-by ($d_* < 1\,\mathrm{kpc}$). The corresponding total detectable event rate is

$$\Gamma_{\mathrm{FS}} \approx 2 \frac{\tau_{\mathrm{FS}}\, v_c}{\pi x_{\mathrm{max}} R}, \tag{11.6}$$

where $v_c \approx 220\,\mathrm{km/s}$ is the halo circular velocity. In case of large M_{PBH} with $x_{\mathrm{max}} \approx 1$, we find $\Gamma_{\mathrm{FS}} \propto M_{\mathrm{PBH}}^{-1}$. This is due to the assumption of a monochromatic mass function in which the total number of PBHs in the DM halo depends on the mass $n_{\mathrm{PBH}} \propto M_{\mathrm{PBH}}^{-1}$. However, for small PBH

1: See the discussion about the star selection and sensitivity of IACTs below.

[167]: Griest et al. (2011), 'Microlensing of Kepler Stars as a Method of Detecting Primordial Black Hole Dark Matter'

masses $x_{\max} \propto M_{\mathrm{PBH}}$ and the detectable rate is $\Gamma_{\mathrm{FS}} \propto M_{\mathrm{PBH}}$. The average detectable duration of the microlensing events is

$$\langle t_e \rangle = \frac{\tau_{\mathrm{FS}}}{\Gamma_{\mathrm{FS}}}. \tag{11.7}$$

For small masses $\tau_{\mathrm{FS}} \propto M_{\mathrm{PBH}}^2$ and thus $\langle t_e \rangle \propto M_{\mathrm{PBH}}$. The average duration of events decreases together with the mass. This proves, that the sensitivity to fast optical transient events is critical for detecting microlensing events with light PBHs.

11.2 Microlensing Observations with IACTs

Existing microlensing searches use traditional optical instruments. Contrary to IACTs, they provide a good optical precision which allows monitoring a large number of stars simultaneously. For this, they use CCD cameras with high pixel densities. The fastest sampling speed among these studies was 2 min.
Contrary to these instruments, IACTs are particularly optimized to detect the ~ ns flashes of Cherenkov light produced in atmospheric air showers. Their optical performance is discussed in more details in [168]. In addition to their optimized cameras, the most critical attribute are the large optical reflectors. They help minimizing atmospheric scintillation noise. However, the precision in measuring fluxes is at best modest compared to the traditional instruments. Below we refer to VERITAS as representative for the current generation of IACTs. At a sampling speed of 2400 Hz VERITAS is sensitive to measure the flux of a ~ 10.2 mag object with a uncertainty of [165]

[168]: Daniel (2019), 'Non-Gamma-ray Applications of TeV Telescopes'

[165]: W. et al. (2019), 'Direct measurement of stellar angular diameters by the VERITAS Cherenkov telescopes'

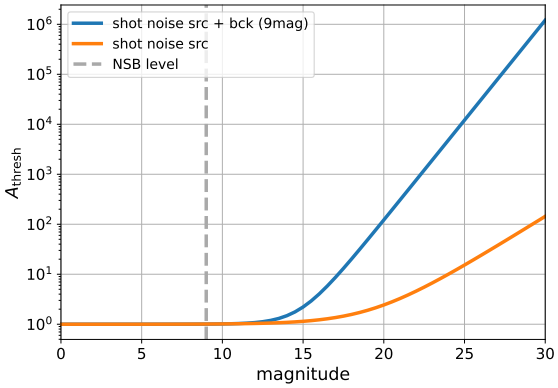
$$\frac{\Delta F}{F}\left(2400\,\mathrm{Hz}, 10.2\,\mathrm{mag}\right) = 0.1. \tag{11.8}$$

Each of the 499 pixels cover a patch of the sky with ~ 0.16° diameter. Thus, they integrate a large amount of background light compared to the much denser CCD cameras. Extragalactic targets, such as stars within as the Large Magellanic Cloud or the Andromeda nebular, have a narrow angular separation. Thus, the individual stars are not resolvable with IACT cameras. Only bright galactic stars which are capable to outshine the total background can be a possible targets. Below we conservatively assume that only one bright foreground star could be detected at once by an IACT.

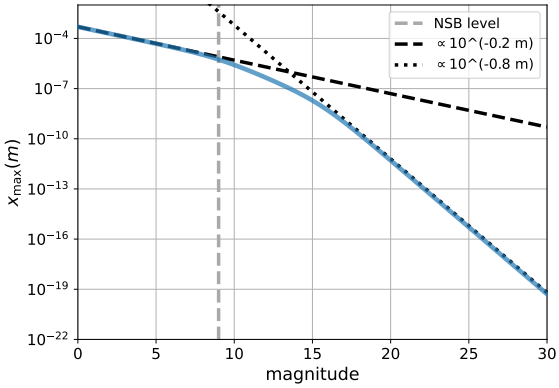
11.2.1 Target Selection

As mentioned above, the stellar properties are crucial for describing the expected microlensing events. They directly affect the optical depth and thus the expected detectable event rate. We investigate the optimal target stars for constraining the event rate for PBH masses below $10^{-10}\,M_{\odot}$. For this work, we only consider shot noise as a source of uncertainty. The shot noise can be modeled by a Poisson process. The relative uncertainty of flux measurements decreases as $\Delta F/F \propto N^{-0.5}$, where N is the total number of detected photons This number is composed by contributions of the source N_{src} and the night sky background N_{bck}. We use the baseline sensitivity

stated in Equation 11.8 to scale the relative uncertainty to arbitrary magnitudes and sampling speeds. As discussed below, we use 50 Hz in this work. Further, we assume a constant NSB level of magnitude $m = 9$. The sensitivity to detect flux changes A_{thresh} directly depends on the relative uncertainty. We show the dependency of A_{thresh} on the magnitude of the star in Figure 11.3a. For this graph, we assume $M_{\text{PBH}} = 10^{-10} M_{\odot}$, a star distance $d_* = 50\,\text{pc}$, and a stellar radius $R = 10\,R_{\odot}$. The orange graph only considers the shot noise by the source. In the blue curve, also the NSB is considered. For very bright stars the value is roughly constant $A_{\text{thresh}} \sim 1$. This threshold increases exponentially for fainter stars. The calculation with the NSB leads to a larger slope. This reflects the fact that the total amplification needs to be sufficient to overcome the additional constant background. The transition happens around $\sim 13\,\text{mag}$.



(a) Detection threshold A_{thresh} as function of the star magnitude. Contrary to the blue curve, the NSB is ignored for the orange graph.



(b) Maximum relative distance x_{max} between the observer and PBH to observer and source star the as a function of the star magnitude.

Figure 11.3: Required amplification and maximum distance as function of the magnitude of the star. For this graphs, we use values of $M_{\text{PBH}} = 10^{-10}\,M_{\odot}$, $d_* = 50\,\text{pc}$, and $R = 10\,R_{\odot}$. The dashed grey vertical line illustrates the NSB level of 9 mag.

Using above result for A_{thresh}, we calculate x_{max} for the graph with the NSB contribution. This result of this empirical function is shown in Figure 11.3b. Below the transition point, we find $x_{\text{max}} \propto 10^{-0.2m}$ and above $x_{\text{max}} \propto 10^{-0.8m}$. These are shown by the black dashed and dotted lines, respectively. We apply an interpolation $f(m)$ describing the of x_{max} for the reference parameters mentioned above. It can be generalized to arbitrary source values and PBH masses using

$$x_{\text{max}}(m) \approx f(m) \left(\frac{d_*}{50\,\text{pc}}\right) \left(\frac{R}{10\,R_{\odot}}\right)^{-2} \left(\frac{M_{\text{PBH}}}{10^{-10}\,M_{\odot}}\right). \tag{11.9}$$

Inserting Equation 11.5 and Equation 11.9 into Equation 11.6 yields the dependency of the event rate on the source and PBH parameters

$$\Gamma_{\text{FS}}(m) \propto \frac{d_*^3\, f(m)^2}{R^3} M_{\text{PBH}}. \tag{11.10}$$

We define the ideal target star to maximize the expected detectable event rate. It is a balance between the star parameters. A large source distance d_* corresponds to a large amount of DM along the LoS. A small star radius and small magnitude are critical to obtain a large absolute amplification during the microlensing. We investigate the objects of the JSDC catalog

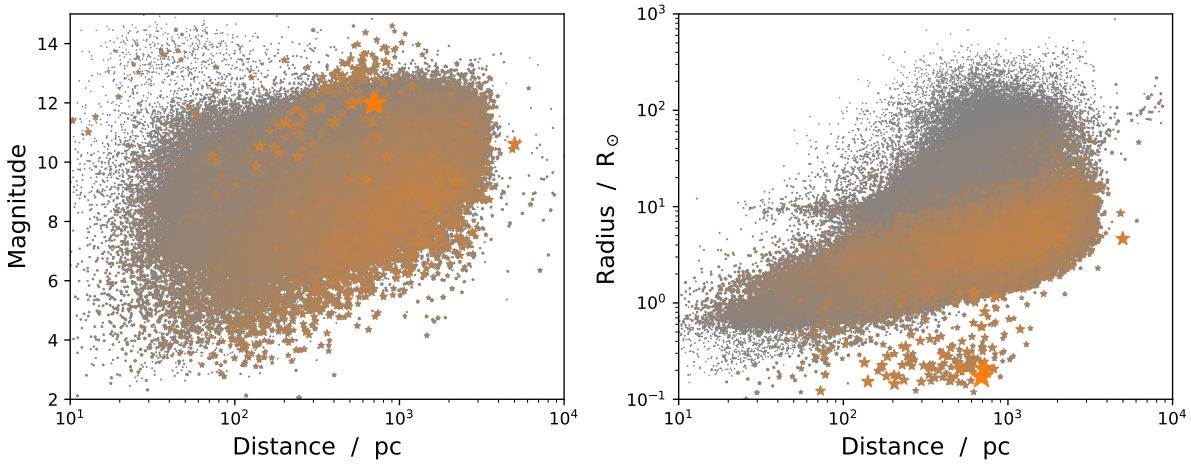


Figure 11.4: Distribution of star parameters relevant for microlensing. The larger markers and orange colors highlight target stars with a high expected rate of microlensing events in the mass range $M_{\mathrm{PBH}} < 10^{-10} M_{\odot}$. [163].

[169]: Chelli et al. (2016), 'Pseudomagnitudes and differential surface brightness: Application to the apparent diameter of stars'

[170]: Wenger et al. (2000), 'The SIMBAD astronomical database. The CDS reference database for astronomical objects'

[171]: Geier, S. (2020), 'The population of hot subdwarf stars studied with Gaia - III. Catalogue of known hot subdwarf stars: Data Release 2'

[172]: Heber (1986), 'The atmosphere of subluminous B stars. II. Analysis of 10 helium poor subdwarfs and the birthrate of sdB stars.'

[169] and query the SIMBAD database [170] to obtain the distances of these objects. In total we have 433,378 candidates with all required information. We show the distribution of the critical star parameters in Figure 11.4. For each star, we estimate the relative expected event rate using Equation 11.10. Stars with high expected rates are highlighted by larger orange markers.

The majority of the best targets are B-type stars. Especially, hot subdwarf stars (sdO/Bs) prove to be excellent targets. These are situated in the extreme horizontal branch [171, 172]. Particularly, their very small radii make them excellent targets for this microlensing study. The optimal target is identified to be the sdO/B star PG 0240+046 with a V magnitude of 11.98, distance $d_* = 692\,\mathrm{pc}$ and radius $R = 0.174\,R_{\odot}$.

In this selection, we ignore the potential variability of stars. These are not suited as targets for IACTs. Further, we also don't consider the possible saturation of pixels. It can limit the maximum brightness of stars that can be monitored and thus exclude the stars with very low magnitudes. However, the identified best target with a magnitude of 11.98 is well observable with IACTs. We also note that only a small fraction of O-type stars are included in the JSDC catalog. These stars also might be attractive targets for microlensing due to their large ratio of brightness to radius.

11.2.2 Event rates

At their corresponding low mass end, existing contraints are typically limited by the minimum detectable event duration. Thus, the calculation of the expected event rate is performed by integrating the differential rate $d\Gamma/dt_e$ over the relevant range of event durations t_e. For examples of this approach see [167, 173, 174]. The limiting factor for IACTs, however, is not the sampling speed but the sensitivity to flux changes and the small number of simultaneously monitored stars. These factors are directly considered in the event rate given by Equation 11.6. Thus, this integration is not required for this study.

To calculate the event rate, we assume a monochromatic mass profile of the PBHs in the DM halo. Further, we assume that at any time during the operations one star with attributes such as PG 0240+046 is observed.

[167]: Griest et al. (2011), 'Microlensing of Kepler Stars as a Method of Detecting Primordial Black Hole Dark Matter'

[173]: Griest (1991), 'Galactic Microlensing as a Method of Detecting Massive Compact Halo Objects'

[174]: Griest et al. (2013), 'New Limits on Primordial Black Hole Dark Matter from an Analysis of Kepler Source Microlensing Data'

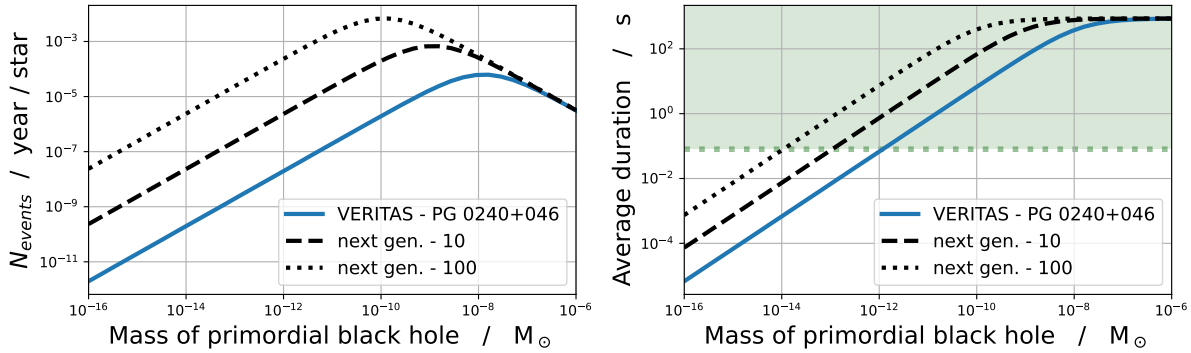


Figure 11.5: *Left*: Expected event rate for the optimal target PG 0240+046 as function of the PBH mass calculated using Equation 11.6. The blue curve shows the expectation using the sensitivity of VERITAS. The dashed and dotted black lines correspond to projects to next generation IACTs assuming 10 or 100 times smaller relative uncertainties, respectively. *Right*: Average duration of the detectable events calculated using Equation 11.7 for VERITAS and next generation instruments.

Due to this optimistic assumption, the calculated event rates should be considered as upper limits. This is sufficient for the feasibility study we present in this thesis. For a more precise analysis, all monitored stars during the observing schedule should be analysed and considered accordingly. This addition is beyond the scope of this work.
The total number of detected photons N scale linearly with the sampling duration t_s. Accordingly, the relative error caused by shot noise decreases with $t_s^{-0.5}$. In principle, the sampling speed for optical observations can be adjusted to the required science case. To reduce this relative uncertainty, the sampling speed is selected as large as possible so that the microlensing events can still be detected. In this work, we assume 50 Hz. Compared to the reference of 2,400 Hz discussed above, it yields a $\sim 1/7$ times lower relative error. We identify a microlensing event by at least four consecutive sample being enhanced by more than 3σ. This corresponds to less than one expected fake positive event per year of observing time. It results in a sensitivity to events with $t_e > 0.08$ s.
The blue solid line in the left panel of Figure 11.5 presents the expected event rate for VERITAS as a function of M_{PBH}. The transition between constant $x_{\max}$ to the mass dependent regime is around $10^{-8}\,M_\odot$. This corresponds to the peak of the expected event rate. In the relevant range of masses $\leq 10^{-10}\,M_\odot$, the rate decreases linearly with M_{PBH}. The reason for this is the decreasing $x_{\max}$. On the right side we show the average detectable event duration calculated with Equation 11.7. Again, the blue curve shows the results expected for a study with VERITAS. The detectable range of average event duration is highlighted by the green area. It proves, that the fast sampling could allow a detection down to $\sim 10^{-12}\,M_\odot$ for VERITAS. In the mass range we aim to constrain, the highest rate is obtained for $10^{-10}\,M_\odot$. However, even at this mass, an event is only expected every $\sim 10^6$ years of observations.
Given the low expected event rate for VERITAS, we investigate the possibilities for next generation instruments as well. The largest improvements are expected due to larger reflecting areas and better quantum efficiencies of the photodetectors. These directly improve the quantity $A_{\mathrm{thresh}} - 1$ that describes the detectable relative brightening of the star. We investigate the results assuming changes by the future instrument by a factor of 1/10 and 1/100 compared to VERITAS. These two cases are illustrated by the black dashed and dotted lines in Figure 11.5, respectively. We find

that below the peak in detectable events the rate increase by a factor $(A_{\text{thesh}} - 1)^{-2}$. Further, the maximum is shifted to lower masses. The event duration increases with $(A_{\text{thesh}} - 1)^{-1}$ in this range. Even with the improved sensitivity by a factor 100, we only expect an event after ~ 150 years of observations with the given sampling speed.

As mentioned above, the relative errors depend on the sampling duration t_s and thus also the critical sensitivity changes $(A_{\text{thresh}} - 1) \propto t_s^{-0.5}$. At masses below the peak of detectable events, we find $\Gamma_{\text{FS}} \propto t_s$ and $\langle t_e \rangle \propto \sqrt{t_s}$. Above this point, the event rate and duration are independent of the changes. As the transition moves to lower masses, it allows tuning the sampling speed to have the slowest possible sampling that still is sufficient to detect these events. E.g., for the 100 times improved next generation event, a 1 Hz sampling yields a peak around $10^{-11}\, M_\odot$. At this peak the expected event rate is ~ 0.05 events per year and the average duration is still well detectable.

We note that for a more detailed study also diffraction effects have to be considered [175, 176]. These are relevant as the Schwarzschild radius of the lensing PBH in this mass range is similar to the observed wavelength $R_{\text{sch}} \approx \lambda$. In contrast to the assumption above, this effect introduces a time dependent oscillation of the apparent amplification. In this case, the maximum reached brightening is decreased. Considering this effect in the study above, would further decrease the total rate of detectable events.

IACTs are also limited to observing during the night. Also a promising target star is not within the FoV during all observations. Considering these aspects, decreases the rate of detectable microlensing events further. Even when ignoring the diffractive microlensing and other effects the detectable PBH microlensing event rates are modest for the current and next generation of IACTs. We conclude that the optical capabilities of IACTs are not sufficient to constrain the PBH abundance in the sublunar mass window using microlensing observations.

[175]: Jaroszynski et al. (1995), 'Diffraction Effects in Microlensing of Q2237+0305'

[176]: Naderi et al. (2018), 'Primordial black hole detection through diffractive microlensing'

12 Conclusions and Outlook

Primordial black holes have been involved in explaining many cosmological observations. Yet, to date their existence has neither been confirmed nor entirely excluded. The largest unconstrained mass range is the sublunar window which reaches from about 10^{17} g to about 10^{23} g. The mass of PBHs below this window is small enough that the Hawking radiation might be relevant. Those created at about 10^{15} g would evaporate today. PBHs with masses in the sublunar window and above are unaffected by Hawking radiation. However, they might contribute to dark matter, which comprises about 26% of the Universe's critical density. This thesis contributes to the ongoing search for these objects by developing new methods to detect PBHs with VERITAS. In the first part, we aimed to detect the burst of γ-rays caused by evaporating PBHs. For this, we presented the implementation of a deep-learning-based transient detection method.

We considered a data set of about 9 years of observations with VERITAS between 2012 and 2021. Manually validating the quality of this data is not feasible and would be a potential source of inconsistencies. Thus, we developed a sophisticated automatic data quality assessment. It is based on the VERITAS internal database. The data selection can be performed without downloading and analyzing the data. This approach saves time and unnecessary computational effort. Further, assessing the quality based on low-level auxiliary parameters prevents possible selection biases on the high-level results. Besides simple filtering of the general configurations and system status, it also performs a time-series analysis of the L3-trigger rate and FIR temperature to identify temporary effects. The three most common features in the data are short spikes or drops in the rates, clouds in the FoV, and changes in the night sky brightness. They can influence the instrument's sensitivity and thus might not be suited for scientific analysis. We developed algorithms that specifically aim to identify these effects. They make it possible to mask these parts with time-cuts rather than excluding the entire observing run. This method is applicable beyond the scope of this thesis. This automatic selection of data can create consistent datasets for scientific analyses. For this thesis, we selected about 4222 hours of observations after time-cuts which are probed for bursts by evaporating PBHs.

We implemented a new approach for VERITAS to probe this dataset. The deep-learning-based transient detection method allows for conducting a blind search. This thesis presented the first application of this method to actual observational IACT data. It is data-driven, which reduces the need for external modeling of the instrument response function. The method can be split into two core elements. The first part is a long short-term memory recurrent neural network with an encoder-decoder architecture. It works with time series of γ-like counts. These sequences represent periods within the full dataset. The encoder is the background interval. Based on this input, the LSTM predicts the expected background counts of γ-like events during the decoder time steps. A higher event count in the observed data compared to the prediction indicates a transient signal. In the second part, the results of the LSTM are used in a calibration

pipeline to assign a statistical significance to possible transient signals. The training of the LSTM and preparation of the calibration require background datasets without transient signals. As this method has not been used for a search with real IACT data before, this thesis first defined the required strategies. Some of the key questions were:

- How can we create the required background datasets from the observation data?
- To conduct a blind search, the decoder needs to cover the entire probed period. However, the DL transient detection method always requires full time series of encoder plus decoder. How can we probe the full dataset consistently?
- Which data features are critical to describing the instrument response?

The newly developed data preparation pipeline reflects the answers to these critical questions. Among the methods are approaches for shuffling the data to remove transient signals, the definition of region of interests and exclusion regions, preparing and padding time series, and a sliding window approach. Further, we investigated which features describe the changes in the data. We found that the secant of the zenith angle of observation $\sec(\theta)$, the azimuthal angle of observation $|\alpha|$, the age of the instrument ΔT_{ref}, the L3 trigger rate R_{L3}, the multiplicity of events M, and the offset angle ζ describe the effects on the rate of γ-like events. These are used as auxiliary parameters for the LSTM. Based on this study, we also defined meta bins that correspond to the most significant systematic changes in the γ-like event rate. The calibration is done independently for each of these meta bins. This technical work allows searching for transient signals in the archival VERITAS data. In the implemented anomaly detection, only the expected duration of the transient signals is directly used. Thus, it might easily be adjusted for different transient sources of VHE γ-rays, such as γ-ray bursts or blazar flares from objects not yet detected in the TeV-range.

In this thesis, we probed the data for the $\mathcal{O}$(s) signals from evaporating PBHs. We conducted simulations of the signals which might be detectable with VERITAS. These cover the full range of observing conditions in the dataset. The simulations are superimposed on a shuffled dataset created from actual data. This avoids possible systematic uncertainties in simulating the background. We used this dataset to determine the detection efficiency for evaporating PBHs as a function of the observing conditions, the PBH distance, and the position in the FoV. With this efficiency, we calculate the effective search volume in which this analysis can detect PBH bursts. Depending on the observing conditions, it is up to $V_{\text{eff}} = 9.96 \times 10^{-5}\,\text{pc}^3$. This calculation takes into account the exact setup used for the actual search.

With the estimate of the probed volume, we applied this method to the full unshuffled dataset. After correcting for trial factors, no burst caused by an evaporating PBH was detected. Thus, we used the total observation time and V_{eff} to constrain the local rate of PBH evaporations to $r_{\text{burst}} < 1.07 \times 10^5\,\text{pc}^{-3}\text{s}^{-1}$ at the 99% confidence interval. This limit is in the range of existing constraints.

Future work can build on the implemented analysis and improve the power of this new analysis technique. We discussed some of the critical aspects which might be investigated. One point is a systematic uncer-

tainty on the statistical significance. Currently, these are estimated too conservatively. The most probable source is the shuffling of sparsely populated data. In this thesis, we estimate the effect of this uncertainty. The scientific results on the rate of PBH evaporations are not affected by this uncertainty.
Further improvements might include additional calibration stages combining signals from neighboring ROIs. This could improve the detection of weak transient signals close to the border of an ROI. Also, the ROIs could be optimized to use a hexagonal instead of a circular shape. It would allow covering the entire FoV without gaps. This analysis also showed that exclusion regions significantly impact the detection efficiency of transient signals. For the scope of this thesis, we selected these conservatively. Carefully revising these might allow us to include more of the field of view in the search. Finally, the hyper-parameters of this analysis might be further optimized for detecting PBHs. Notably, the duration of the sensitive window, given by the decoder length, might be adjusted. This might improve the signal-to-noise ratio.
The second part of this thesis investigates whether IACTs might constrain PBH masses within the sublunar mass window. For this, we probed optical microlensing as an alternative approach for IACTs. Sampling speeds on the order of minutes to hours of traditional optical telescopes are typically a constraining factor in extending PBH limits to lower masses. IACTs can take optical observations with up to $\mathcal{O}$(ns) sampling. In this study, we used VERITAS as a representative of the current generation of IACTs. For this work, we reviewed microlensing and the optical performance of IACTs and investigated which stars might be suitable targets for this study. We found that a bright star with a small radius at a large distance is preferred. Our investigation shows that the small radii of hot subdwarf stars make them excellent targets. We used the ideal target star PG 0240+046 to calculate the expected rate of PBH microlensing events. For this, we determined an appropriate sampling speed that IACTs might use. It needs to be fast enough to detect these transient optical events. However, a slower sampling reduces the relative uncertainties of the Poissonian noise. For VERITAS, as representative of the current generation of IACTs, we find an expected rate of detectable events of $\sim 10^{-6}\,\mathrm{yr}^{-1}$ in the sublunar mass range with 50 Hz sampling speed. We also investigated the changes for possible next-generation instruments. Due to the higher sensitivity to optical transient, the optimal sampling speed decreases to about 1 Hz. With this configuration, the expected detectable rate is $\sim 0.05\,\mathrm{yr}^{-1}$. We conclude that the modest precision of IACTs to measure optical fluxes is the most constraining aspect. Hardware advances increasing the optical flux sensitivity, however, would improve the expected rate. Paired with the fast sampling speed, it might also allow investigating the regime of diffractive microlensing for light PBHs.

Appendix

A

Background Rate Dependencies

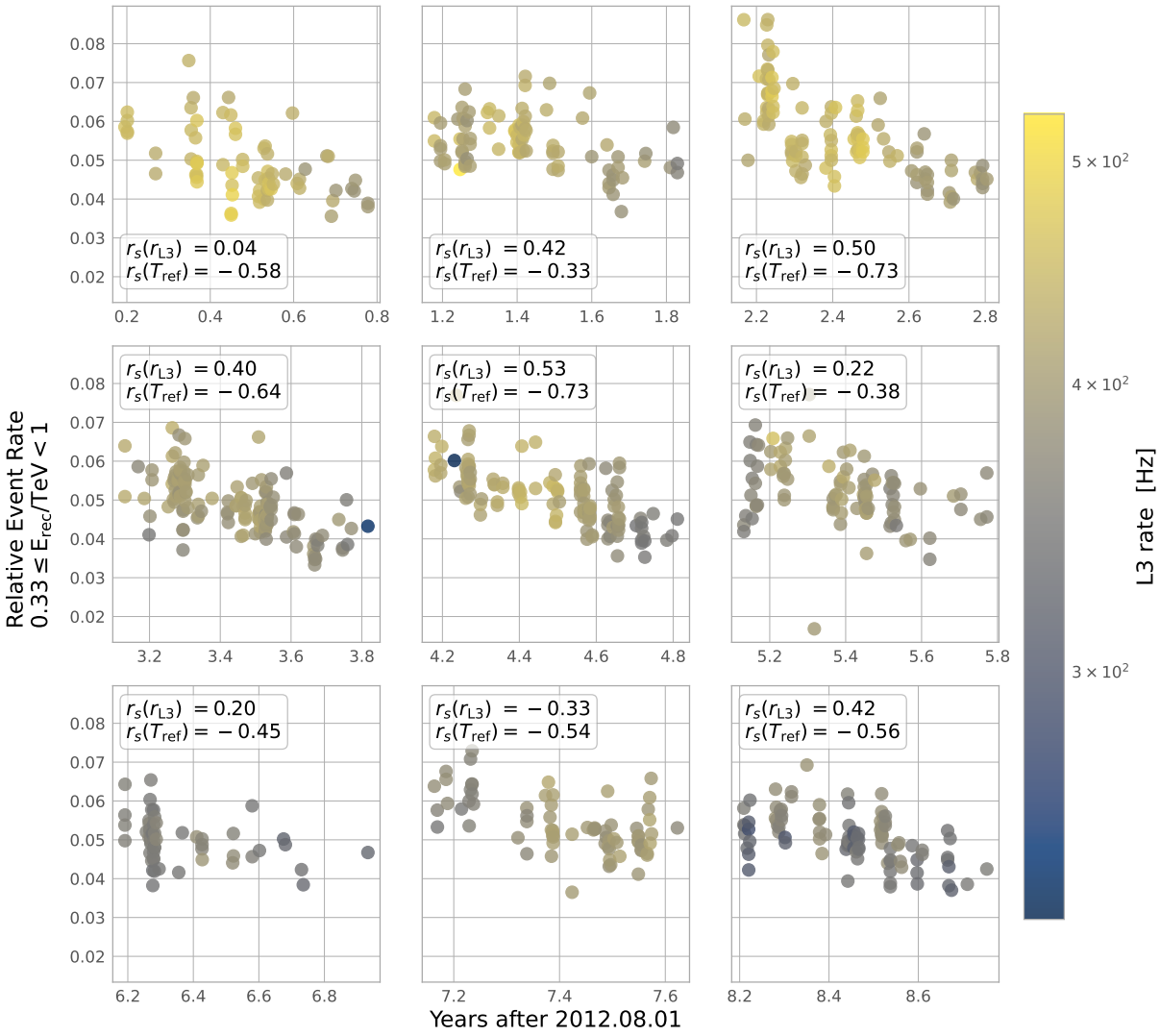


Figure A.1: Run-wise background event rate in the second energy bin R_1 ($0.33 \leq E_{rec}/TeV < 1$) in dependency of the reference time T_{ref}. Each point corresponds to an individual run. The colors denote the average L3 rate r_{L3} during each run. This plot only shows runs with $\sec(\theta) < 1.05$, pointing south, and in an extragalactic field. Each subgraph shows one of the nine seasons that are used in this work. For each season, we calculate the correlation between the background rates R_1 with r_{L3} and T_{ref} using Spearman's correlation coefficient.

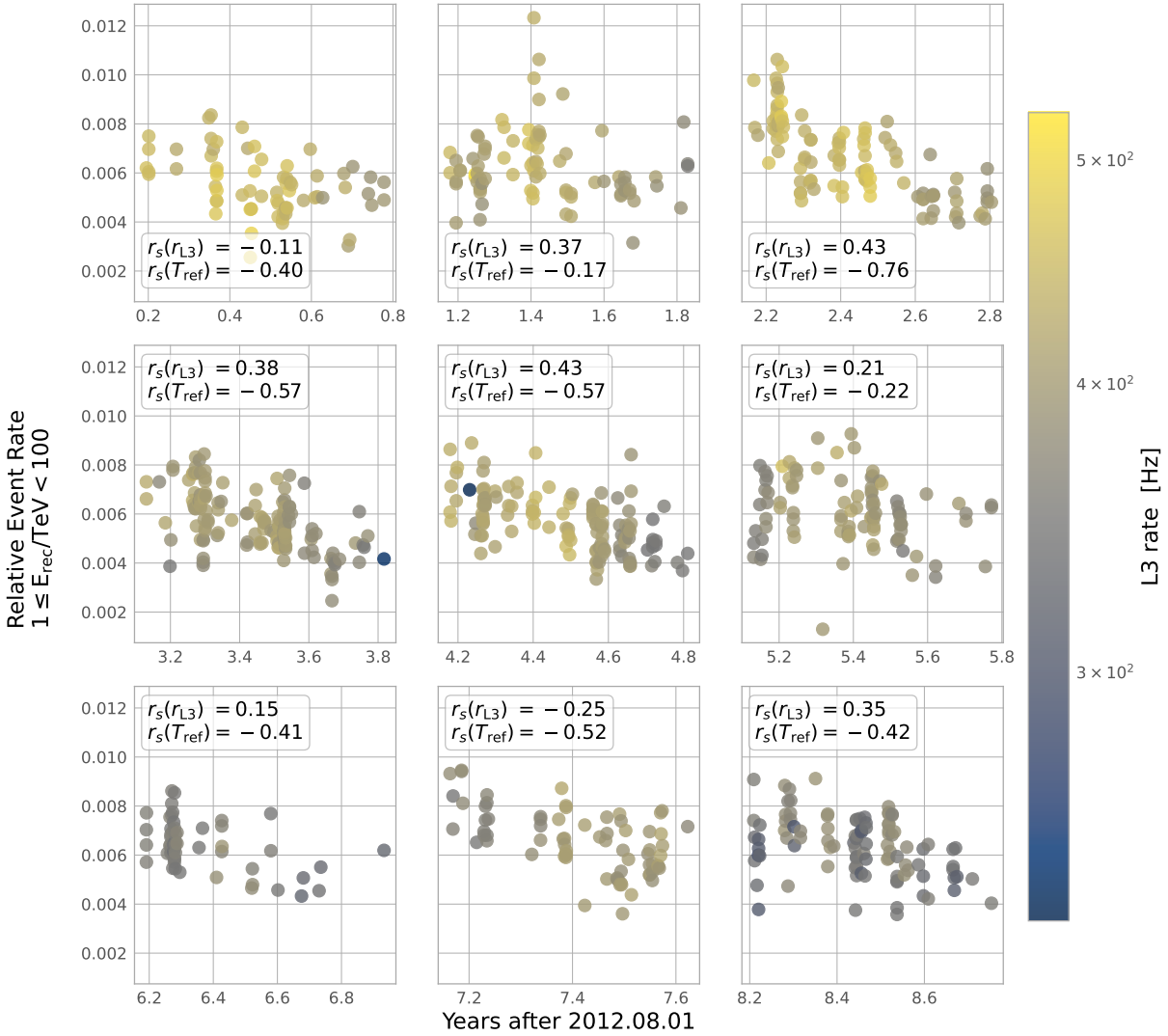


Figure A.2: Run-wise background event rate in the third energy bin R_2 ($1 \leq E_{rec}/TeV < 100$) in dependency of the reference time T_{ref}. Each point corresponds to an individual run. The colors denote the average L3 rate r_{L3} during each run. This plot only shows runs with $\sec(\theta) < 1.05$, pointing south, and in an extragalactic field. Each subgraph shows one of the nine seasons that are used in this work. For each season, we calculate the correlation between the background rates R_1 with r_{L3} and T_{ref} using Spearman's correlation coefficient.

Bibliography

[1] Bernard Carr et al. 'Constraints on primordial black holes'. In: *Reports on Progress in Physics* 84.11 (2021), p. 116902. DOI: 10.1088/1361-6633/ac1e31 (cited on pages 1, 6, 10, 11).

[2] S. W. Hawking. 'Black hole explosions?' In: Nature 248.5443 (Mar. 1974), pp. 30–31. DOI: 10.1038/248030a0 (cited on pages 1, 5).

[3] Nabila Aghanim et al. 'Planck 2018 results-VI. Cosmological parameters'. In: *Astronomy & Astrophysics* 641.A6 (2020) (cited on pages 1, 5, 9).

[4] Kim Griest, Agnieszka M. Cieplak, and Matthew J. Lehner. 'New Limits on Primordial Black Hole Dark Matter from an Analysis of Kepler Source Microlensing Data'. In: *Phys. Rev. Lett.* 111 (18 Oct. 2013), p. 181302. DOI: 10.1103/PhysRevLett.111.181302 (cited on page 1).

[5] Ya. B. Zel'dovich and I. D. Novikov. 'The Hypothesis of Cores Retarded during Expansion and the Hot Cosmological Model'. In: Soviet Ast. 10 (Feb. 1967), p. 602 (cited on page 5).

[6] Stephen Hawking. 'Gravitationally Collapsed Objects of Very Low Mass'. In: *Monthly Notices of the Royal Astronomical Society* 152.1 (Apr. 1971), pp. 75–78. DOI: 10.1093/mnras/152.1.75 (cited on page 5).

[7] D. N. Page and S. W. Hawking. 'Gamma rays from primordial black holes.' In: ApJ 206 (May 1976), pp. 1–7. DOI: 10.1086/154350 (cited on page 5).

[8] Roland Lehoucq et al. 'New constraints on the primordial black hole number density from Galactic γ-ray astronomy'. In: *Astronomy & Astrophysics* 502.1 (2009), pp. 37–43 (cited on page 5).

[9] K.M. Belotsky and A.A. Kirillov. 'Primordial black holes with mass $10^{16} - 10^{17}$ g and reionization of the Universe'. In: *Journal of Cosmology and Astroparticle Physics* 2015.01 (2015), pp. 041–041. DOI: 10.1088/1475-7516/2015/01/041 (cited on page 5).

[10] PN Okele and MJ Rees. 'Observational consequences of positron production by evaporating black holes'. In: *Astronomy and Astrophysics* 81 (1980), p. 263 (cited on page 5).

[11] C. G. Lacey and J. P. Ostriker. 'Massive black holes in galactic halos ?' In: ApJ 299 (Dec. 1985), pp. 633–652. DOI: 10.1086/163729 (cited on page 5).

[12] Rachel Bean and João Magueijo. 'Could supermassive black holes be quintessential primordial black holes?' In: *Phys. Rev. D* 66 (6 2002), p. 063505. DOI: 10.1103/PhysRevD.66.063505 (cited on page 5).

[13] N. Afshordi, P. McDonald, and D. N. Spergel. 'Primordial Black Holes as Dark Matter: The Power Spectrum and Evaporation of Early Structures'. In: *The Astrophysical Journal* 594.2 (2003), pp. L71–L74. DOI: 10.1086/378763 (cited on page 5).

[14] B. J. Carr and S. W. Hawking. 'Black Holes in the Early Universe'. In: *Monthly Notices of the Royal Astronomical Society* 168.2 (Aug. 1974), pp. 399–415. DOI: 10.1093/mnras/168.2.399 (cited on page 6).

[15] Ilia Musco and John C. Miller. 'Primordial black hole formation in the early universe: critical behaviour and self-similarity'. In: *Classical and Quantum Gravity* 30.14, 145009 (July 2013), p. 145009. DOI: 10.1088/0264-9381/30/14/145009 (cited on page 6).

[16] Tomohiro Harada, Chul-Moon Yoo, and Kazunori Kohri. 'Erratum: Threshold of primordial black hole formation [Phys. Rev. D 88, 084051 (2013)]'. In: Phys. Rev. D 89.2, 029903 (Jan. 2014), p. 029903. DOI: 10.1103/PhysRevD.89.029903 (cited on page 6).

[17] Ilia Musco. 'Threshold for primordial black holes: Dependence on the shape of the cosmological perturbations'. In: *Phys. Rev. D* 100 (12 2019), p. 123524. DOI: 10.1103/PhysRevD.100.123524 (cited on page 6).

[18] Albert Escrivà, Cristiano Germani, and Ravi K. Sheth. 'Universal threshold for primordial black hole formation'. In: *Phys. Rev. D* 101 (4 2020), p. 044022. DOI: 10.1103/PhysRevD.101.044022 (cited on page 6).

[19] B. J. Carr. 'The primordial black hole mass spectrum.' In: ApJ 201 (Oct. 1975), pp. 1–19. DOI: 10.1086/153853 (cited on page 6).

[20] J. C. Niemeyer and K. Jedamzik. 'Dynamics of primordial black hole formation'. In: *Phys. Rev. D* 59 (12 1999), p. 124013. DOI: 10.1103/PhysRevD.59.124013 (cited on page 7).

[21] B. J. Carr et al. 'Constraints on primordial black holes from the Galactic gamma-ray background'. In: *Phys. Rev. D* 94 (4 2016), p. 044029. DOI: 10.1103/PhysRevD.94.044029 (cited on page 7).

[22] M. Yu. Khlopov, B. A. Malomed, and Ya. B. Zeldovich. 'Gravitational instability of scalar fields and formation of primordial black holes'. In: *Monthly Notices of the Royal Astronomical Society* 215.4 (Aug. 1985), pp. 575–589. DOI: 10.1093/mnras/215.4.575 (cited on page 7).

[23] S.W. Hawking. 'Black holes from cosmic strings'. In: *Physics Letters B* 231.3 (1989), pp. 237–239. DOI: https://doi.org/10.1016/0370-2693(89)90206-2 (cited on page 7).

[24] Bernard Carr and Florian Kühnel. 'Primordial Black Holes as Dark Matter: Recent Developments'. In: *Annual Review of Nuclear and Particle Science* 70.1 (2020), pp. 355–394. DOI: 10.1146/annurev-nucl-050520-125911 (cited on page 7).

[25] Andrew F. Heckler. 'Calculation of the Emergent Spectrum and Observation of Primordial Black Holes'. In: *Phys. Rev. Lett.* 78 (18 1997), pp. 3430–3433. DOI: 10.1103/PhysRevLett.78.3430 (cited on page 7).

[26] Stephen W Hawking. 'Particle creation by black holes'. In: *Euclidean quantum gravity*. World Scientific, 1975, pp. 167–188 (cited on page 7).

[27] T.N. Ukwatta et al. 'Primordial Black Holes: Observational characteristics of the final evaporation'. In: *Astroparticle Physics* 80 (2016), pp. 90–114. DOI: https://doi.org/10.1016/j.astropartphys.2016.03.007 (cited on pages 7, 8, 57, 59).

[28] Jane H. MacGibbon and B. R. Webber. 'Quark- and gluon-jet emission from primordial black holes: The instantaneous spectra'. In: *Phys. Rev. D* 41 (10 1990), pp. 3052–3079. DOI: 10.1103/PhysRevD.41.3052 (cited on page 8).

[29] Don N. Page. 'Particle emission rates from a black hole: Massless particles from an uncharged, nonrotating hole'. In: *Phys. Rev. D* 13 (2 1976), pp. 198–206. DOI: 10.1103/PhysRevD.13.198 (cited on page 8).

[30] Y Akrami et al., Planck Collaboration. 'Planck 2018 results. I. Overview and the cosmological legacy of Planck'. In: *Astronomy & Astrophysics* A1 (2020) (cited on pages 8, 9).

[31] William H. Kinney and Pierre Sikivie. 'Evidence for universal structure in galactic halos'. In: *Phys. Rev. D* 61 (8 2000), p. 087305. DOI: 10.1103/PhysRevD.61.087305 (cited on page 9).

[32] F. de Paolis, G. Ingrosso, and F. Strafella. 'Dark Matter in X-Ray-emitting Elliptical Galaxies'. In: ApJ 438 (Jan. 1995), p. 83. DOI: 10.1086/175056 (cited on page 9).

[33] Douglas Clowe, Anthony Gonzalez, and Maxim Markevitch. 'Weak-Lensing Mass Reconstruction of the Interacting Cluster 1E 0657-558: Direct Evidence for the Existence of Dark Matter'. In: *The Astrophysical Journal* 604.2 (2004), pp. 596–603. DOI: 10.1086/381970 (cited on page 9).

[34] P. J. E. Peebles. 'Large-scale background temperature and mass fluctuations due to scale-invariant primeval perturbations'. In: ApJ 263 (Dec. 1982), pp. L1–L5. DOI: 10.1086/183911 (cited on page 9).

[35] Richard H Cyburt, Brian D Fields, and Keith A Olive. 'Primordial nucleosynthesis in light of WMAP'. In: *Physics Letters B* 567.3 (2003), pp. 227–234. DOI: https://doi.org/10.1016/j.physletb.2003.06.026 (cited on page 10).

[36] B. J. Carr et al. 'New cosmological constraints on primordial black holes'. In: *Phys. Rev. D* 81 (10 2010), p. 104019. DOI: 10.1103/PhysRevD.81.104019 (cited on page 11).

[37] Sandeep Kumar Acharya and Rishi Khatri. 'CMB and BBN constraints on evaporating primordial black holes revisited'. In: *Journal of Cosmology and Astroparticle Physics* 2020.06 (2020), pp. 018–018. DOI: 10.1088/1475-7516/2020/06/018 (cited on page 11).

[38] Aous A. Abdo et al. 'Milagro Limits and HAWC Sensitivity for the Rate-Density of Evaporating Primordial Black Holes'. In: *Astroparticle Physics* 64 (2015), pp. 4–12 (cited on pages 12, 118).

[39] A. Albert et al. 'Constraining the local burst rate density of primordial black holes with HAWC'. In: *Journal of Cosmology and Astroparticle Physics* 2020.04 (2020), pp. 026–026. DOI: 10.1088/1475-7516/2020/04/026 (cited on pages 12, 118).

[40] D. E. Alexandreas et al. 'New limit on the rate-density of evaporating black holes'. In: *Phys. Rev. Lett.* 71 (16 1993), pp. 2524–2527. DOI: 10.1103/PhysRevLett.71.2524 (cited on pages 12, 118).

[41] E T Linton et al. 'A new search for primordial black hole evaporations using the Whipple gamma-ray telescope'. In: *Journal of Cosmology and Astroparticle Physics* 2006.01 (2006), pp. 013–013. DOI: 10.1088/1475-7516/2006/01/013 (cited on pages 12, 58, 118).

[42] M. Ackermann et al. 'Search for Gamma-Ray Emission from Local Primordial Black Holes with the *Fermi* Large Area Telescope'. In: *The Astrophysical Journal* 857.1 (2018), p. 49. DOI: 10.3847/1538-4357/aaac7b (cited on pages 12, 118).

[43] M. Amenomori et al. 'Search for 10 TeV Gamma Bursts from Evaporating Primordial Black Holes with the Tibet Air Shower Array'. In: *International Cosmic Ray Conference*. Vol. 2. International Cosmic Ray Conference. Jan. 1995, p. 112 (cited on pages 12, 118).

[44] Thomas Tavernier et al. 'Limits on primordial black hole evaporation from H.E.S.S. observations.' In: *Proceedings of 37th International Cosmic Ray Conference — PoS(ICRC2021)*. Vol. 395. 2021, p. 518. DOI: 10.22323/1.395.0518 (cited on pages 12, 118).

[45] Sajan Kumar. 'Constraining the evaporation rate of Primordial black holes using archival data from VERITAS'. In: 2019 (cited on pages 12, 118).

[46] M. Cassanyes. *MAGIC sensitivity to Primordial Black Hole bursts and modelization of BH chromospheres and gamma-ray emission spectra*. 2015 (cited on page 12).

[47] R. López-Coto et al. 'Prospects for the observation of Primordial Black Hole evaporation with the Southern Wide field of view Gamma-ray Observatory'. In: *Journal of Cosmology and Astroparticle Physics* 2021.08 (2021), p. 040. DOI: 10.1088/1475-7516/2021/08/040 (cited on page 12).

[48] Reshmi Mukherjee and Roberta Zanin. *Advances in Very High Energy Astrophysics*. WORLD SCIENTIFIC, 2023 (cited on page 12).

[49] A.U. Abeysekara et al. 'On the sensitivity of the HAWC observatory to gamma-ray bursts'. In: *Astroparticle Physics* 35.10 (2012), pp. 641–650. DOI: https://doi.org/10.1016/j.astropartphys.2012.02.001 (cited on page 13).

[50] Gaurang B. Yodh. 'The MILAGRO gamma ray observatory'. In: *Nuclear Physics B - Proceedings Supplements* 52.3 (1997), pp. 264–268. DOI: https://doi.org/10.1016/S0920-5632(96)00902-4 (cited on page 13).

[51] Michihiro Amenomori et al. 'Tibet air shower array: results and future plan'. In: *Journal of Physics: Conference Series* 120 (2008), p. 062024 (cited on page 13).

[52] J. Hinton and SWGO Collaboration. 'The Southern Wide-field Gamma-ray Observatory: Status and Prospects'. In: *37th International Cosmic Ray Conference. 12-23 July 2021. Berlin*. Mar. 2022, p. 23 (cited on page 13).

[53] J. Holder et al. 'The first VERITAS telescope'. In: *Astroparticle Physics* 25.6 (July 2006), pp. 391–401. DOI: 10.1016/j.astropartphys.2006.04.002 (cited on page 13).

[54] Felix Aharonian et al. 'Observations of the Crab nebula with HESS'. In: *Astronomy & Astrophysics* 457.3 (2006), pp. 899–915 (cited on page 13).

[55] J. Aleksić et al. 'The major upgrade of the MAGIC telescopes, Part I: The hardware improvements and the commissioning of the system'. In: *Astroparticle Physics* 72 (2016), pp. 61–75. DOI: https://doi.org/10.1016/j.astropartphys.2015.04.004 (cited on page 13).

[56] J. Aleksić et al. 'The major upgrade of the MAGIC telescopes, Part II: A performance study using observations of the Crab Nebula'. In: *Astroparticle Physics* 72 (2016), pp. 76–94. DOI: https://doi.org/10.1016/j.astropartphys.2015.02.005 (cited on page 13).

[57] M. Actis et al. 'Design concepts for the Cherenkov Telescope Array CTA: an advanced facility for ground-based high-energy gamma-ray astronomy'. In: *Experimental Astronomy* 32.3 (Dec. 2011), pp. 193–316. DOI: 10.1007/s10686-011-9247-0 (cited on page 13).

[58] Zhen Cao et al. 'Ultrahigh-energy photons up to 1.4 petaelectronvolts from 12 γ-ray Galactic sources'. In: *Nature* 594.7861 (2021), pp. 33–36 (cited on page 13).

[59] IE Tamm and IM Frank. 'Coherent radiation of fast electrons in a medium'. In: *Dokl. Akad. Nauk SSSR*. Vol. 14. 3. 1937, pp. 107–112 (cited on page 14).

[60] F. Schmidt and J Knapp. *CORSIKA Shower Images*. URL: https://www-zeuthen.desy.de/~jknapp/fs/showerimages.html (visited on 10/09/2022) (cited on pages 14, 15).

[61] Gernot Maier. 'The origin of cosmic rays and TeV gamma-ray astronomy'. In: *European Physical Journal Web of Conferences*. Vol. 52. European Physical Journal Web of Conferences. June 2013, p. 10001. DOI: 10.1051/epjconf/20125210001 (cited on page 15).

[62] Particle Data Group et al. 'Review of Particle Physics'. In: *Progress of Theoretical and Experimental Physics* 2020.8 (Aug. 2020). 083C01. DOI: 10.1093/ptep/ptaa104 (cited on page 15).

[63] 'Light Pulses from the Night Sky associated with Cosmic Rays'. In: *Nature* 171.4347 (1953), pp. 349–350. DOI: 10.1038/171349A0 (cited on page 16).

[64] Jamie Holder. 'Atmospheric Cherenkov Gamma-Ray Telescopes'. In: *The WSPC Handbook of Astronomical Instrumentation*. 2015. Chap. Chapter 6, pp. 117–136. DOI: 10.1142/9789811203817_0006 (cited on pages 16, 17).

[65] Daniel Mazin. 'A study of very high energy gamma-ray emission from AGNs and constraints on the extragalactic background light'. Dissertation. München: Technische Universität München, 2007 (cited on page 16).

[66] A. M. Hillas. 'Cerenkov Light Images of EAS Produced by Primary Gamma Rays and by Nuclei'. In: *19th International Cosmic Ray Conference (ICRC1985)*. Vol. 3. International Cosmic Ray Conference. 1985, p. 445 (cited on page 16).

[67] VERITAS collaboration. *VERITAS website*. URL: http://veritas.sao.arizona.edu/ (visited on 18/04/2022) (cited on page 19).

[68] J. Holder et al. 'Status of the VERITAS Observatory'. In: *AIP Conference Proceedings* 1085.1 (2008), pp. 657–660. DOI: 10.1063/1.3076760 (cited on page 19).

[69] Maria Krause, Elisa Pueschel, and Gernot Maier. 'Improved γ/hadron separation for the detection of faint γ-ray sources using boosted decision trees'. In: *Astroparticle Physics* 89 (2017), pp. 1–9. DOI: https://doi.org/10.1016/j.astropartphys.2017.01.004 (cited on pages 19, 23).

[70] A. McCann et al. 'A new mirror alignment system for the VERITAS telescopes'. In: *Astroparticle Physics* 32.6 (2010), pp. 325–329. DOI: https://doi.org/10.1016/j.astropartphys.2009.10.001 (cited on page 19).

[71] D. B. Kieda. 'Status of the VERITAS Upgrade'. In: *Proceedings of 32th International Cosmic Ray Conference — PoS(ICRC2011)*. Vol. 09. 2011, pp. 14–17. DOI: 10.7529/ICRC2011/V09/0343 (cited on page 19).

[72] John M. Davies and Eugene S. Cotton. 'Design of the quartermaster solar furnace'. In: *Solar Energy* 1.2-3 (Apr. 1957), pp. 16–22. DOI: 10.1016/0038-092X(57)90116-0 (cited on page 20).

[73] Roland Winston and Jay M. Enoch. 'Retinal Cone Receptor as an Ideal Light Collector'. In: *J. Opt. Soc. Am.* 61.8 (1971), pp. 1120–1121. DOI: 10.1364/JOSA.61.001120 (cited on page 20).

[74] J. Hall et al. 'Veritas CFDs'. In: *Proceedings of 28th International Cosmic Ray Conference — PoS(ICRC2003)*. Vol. 5. International Cosmic Ray Conference. July 2003, p. 2851 (cited on page 20).

[75] M. K. Daniel. 'The VERITAS standard data analysis'. In: *Proceedings of 30th International Cosmic Ray Conference — PoS(ICRC2007)*. Vol. 3. International Cosmic Ray Conference. Jan. 2008, pp. 1325–1328 (cited on page 21).

[76] Gernot Maier and Jamie Holder. 'Eventdisplay: An Analysis and Reconstruction Package for Ground-based Gamma-ray Astronomy'. In: *Proceedings of 35th International Cosmic Ray Conference — PoS(ICRC2017)*. Vol. ICRC2017. 2018, p. 747. DOI: 10.22323/1.301.0747 (cited on page 22).

[77] D. Heck et al. *CORSIKA: a Monte Carlo code to simulate extensive air showers.* 1998 (cited on page 22).

[78] Akira Okumura et al. 'Development of Non-sequential Ray-tracing Software for Cosmic-ray Telescopes'. In: *Proceedings of 32th International Cosmic Ray Conference — PoS(ICRC2011)*. Vol. 9. Oct. 2011, p. 211. DOI: 10.7529/ICRC2011/V09/1192 (cited on page 22).

[79] N. Otte. *CARE website*. URL: https://otte.gatech.edu/care/ (visited on 19/04/2022) (cited on page 22).

[80] C. B. Adams et al. 'The throughput calibration of the VERITAS telescopes'. In: A&A 658, A83 (Feb. 2022), A83. DOI: 10.1051/0004-6361/202142275 (cited on pages 24, 83).

[81] Oludare Isaac Abiodun et al. 'State-of-the-art in artificial neural network applications: A survey'. In: *Heliyon* 4.11 (2018), e00938. DOI: https://doi.org/10.1016/j.heliyon.2018.e00938 (cited on page 27).

[82] George Cybenko. 'Approximation by superpositions of a sigmoidal function'. In: *Mathematics of control, signals and systems* 2.4 (1989), pp. 303–314 (cited on page 27).

[83] Kurt Hornik. 'Approximation capabilities of multilayer feedforward networks'. In: *Neural Networks* 4.2 (1991), pp. 251–257. DOI: 10.1016/0893-6080(91)90009-T (cited on page 27).

[84] Y. Lecun et al. 'Gradient-based learning applied to document recognition'. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324. DOI: 10.1109/5.726791 (cited on page 27).

[85] Cosmin Anitescu et al. 'Artificial Neural Network Methods for the Solution of Second Order Boundary Value Problems'. In: *Computers, Materials & Continua* 59.1 (2019), pp. 345–359. DOI: 10.32604/cmc.2019.06641 (cited on page 28).

[86] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 'Deep learning'. In: *Nature* 521 (2015), p. 436 (cited on page 28).

[87] Vinod Nair and Geoffrey E. Hinton. 'Rectified Linear Units Improve Restricted Boltzmann Machines'. In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. ICML'10. Haifa, Israel: Omnipress, 2010, pp. 807–814 (cited on page 28).

[88] Ajay Shrestha and Ausif Mahmood. 'Review of Deep Learning Algorithms and Architectures'. In: *IEEE Access* 7 (2019), pp. 53040–53065. DOI: 10.1109/ACCESS.2019.2912200 (cited on page 28).

[89] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. 'Learning representations by back-propagating errors'. In: *nature* 323.6088 (1986), pp. 533–536 (cited on page 29).

[90] Léon Bottou et al. 'Stochastic gradient learning in neural networks'. In: *Proceedings of Neuro-Nımes* 91.8 (1991), p. 12 (cited on page 29).

[91] Diederik P. Kingma and Jimmy Ba. 'Adam: A Method for Stochastic Optimization'. In: *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*. Ed. by Yoshua Bengio and Yann LeCun. 2015 (cited on pages 30, 37).

[92] Sergey Ioffe and Christian Szegedy. 'Batch normalization: Accelerating deep network training by reducing internal covariate shift'. In: *International conference on machine learning*. PMLR. 2015, pp. 448–456 (cited on page 30).

[93] Kaiming He et al. 'Deep Residual Learning for Image Recognition'. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90 (cited on page 30).

[94] Alex Graves and Navdeep Jaitly. 'Towards End-to-End Speech Recognition with Recurrent Neural Networks'. In: *Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32*. ICML'14. Beijing, China: JMLR.org, 2014, II–1764–II–1772 (cited on page 30).

[95] Hyun-Je Song, A-Yeong Kim, and Seong-Bae Park. 'Translation of Natural Language Query Into Keyword Query Using a RNN Encoder-Decoder'. In: *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '17. Shinjuku, Tokyo, Japan: Association for Computing Machinery, 2017, pp. 965–968. DOI: 10.1145/3077136.3080691 (cited on page 30).

[96] Tom Young et al. 'Recent Trends in Deep Learning Based Natural Language Processing [Review Article]'. In: *IEEE Computational Intelligence Magazine* 13.3 (2018), pp. 55–75. DOI: 10.1109/MCI.2018.2840738 (cited on page 30).

[97] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 'Sequence to sequence learning with neural networks'. In: *Advances in neural information processing systems* 27 (2014) (cited on page 30).

[98] Hansika Hewamalage, Christoph Bergmeir, and Kasun Bandara. 'Recurrent Neural Networks for Time Series Forecasting: Current status and future directions'. In: *International Journal of Forecasting* 37.1 (2021), pp. 388–427. DOI: https://doi.org/10.1016/j.ijforecast.2020.06.008 (cited on page 30).

[99] S. Hochreiter. *Untersuchungen zu dynamischen neuronalen Netzen*. 1991 (cited on page 31).

[100] Yoshua Bengio, Patrice Simard, and Paolo Frasconi. 'Learning long-term dependencies with gradient descent is difficult'. In: *IEEE transactions on neural networks* 5.2 (1994), pp. 157–166 (cited on page 32).

[101] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. 'On the difficulty of training recurrent neural networks'. In: *Proceedings of the 30th International Conference on Machine Learning*. Ed. by Sanjoy Dasgupta and David McAllester. Vol. 28. Proceedings of Machine Learning Research 3. Atlanta, Georgia, USA: PMLR, 2013, pp. 1310–1318 (cited on page 32).

[102] Sarath Chandar et al. 'Towards non-saturating recurrent units for modelling long-term dependencies'. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 01. 2019, pp. 3280–3287 (cited on page 32).

[103] Devansh Arpit et al. 'h-detach: Modifying the LSTM gradient towards better optimization'. In: *Proceedings of the 7th International Conference for Learning Representations (ICLR)*. 2019 (cited on page 32).

[104] Sepp Hochreiter and Jürgen Schmidhuber. 'Long short-term memory'. In: *Neural computation* 9.8 (1997), pp. 1735–1780 (cited on page 32).

[105] Kyunghyun Cho et al. 'Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation'. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1724–1734. DOI: 10.3115/v1/D14-1179 (cited on page 32).

[106] Ming Liang and Xiaolin Hu. 'Recurrent convolutional neural network for object recognition'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 3367–3375 (cited on page 34).

[107] E B Postnikov et al. 'Gamma/Hadron Separation in Imaging Air Cherenkov Telescopes Using Deep Learning Libraries TensorFlow and PyTorch'. In: *Journal of Physics: Conference Series* 1181 (2019), p. 012048. DOI: 10.1088/1742-6596/1181/1/012048 (cited on page 34).

[108] Robert Daniel Parsons and S Ohm. 'Background rejection in atmospheric Cherenkov telescopes using recurrent convolutional neural networks'. In: *The European Physical Journal C* 80.5 (2020), pp. 1–11 (cited on page 34).

[109] D. Nieto Castaño et al. 'CTLearn: Deep Learning for Gamma-ray Astronomy'. In: *36th International Cosmic Ray Conference (ICRC2019)*. Vol. 36. International Cosmic Ray Conference. July 2019, p. 752 (cited on page 34).

[110] Samuel Spencer et al. 'Deep learning with photosensor timing information as a background rejection method for the Cherenkov Telescope Array'. In: *Astroparticle Physics* 129 (2021), p. 102579 (cited on page 34).

[111] Tjark Miener et al. 'IACT event analysis with the MAGIC telescopes using deep convolutional neural networks with CTLearn'. In: *arXiv* (2021). Cited by: 0 (cited on page 34).

[112] D. Nieto et al. 'Reconstruction of IACT events using deep learning techniques with CTLearn'. In: *30th Astronomical Data Analysis Software and Systems*. Jan. 2021 (cited on page 34).

[113] Samira Rezaei et al. 'DECORAS: detection and characterization of radio-astronomical sources using deep learning'. In: *Monthly Notices of the Royal Astronomical Society* 510.4 (2022), pp. 5891–5907 (cited on page 34).

[114] Amandin Chyba Rabeendran and Larry Denneau. 'A Two-stage Deep Learning Detection Classifier for the ATLAS Asteroid Survey'. In: *Publications of the Astronomical Society of the Pacific* 133.1021 (2021), p. 034501. DOI: 10.1088/1538-3873/abc900 (cited on page 34).

[115] Ignacio Becker et al. 'Scalable end-to-end recurrent neural network for variable star classification'. In: *Monthly Notices of the Royal Astronomical Society* 493.2 (2020), pp. 2981–2995 (cited on page 34).

[116] A Möller and T de Boissière. 'SuperNNova: an open-source framework for Bayesian, neural network-based supernova classification'. In: *Monthly Notices of the Royal Astronomical Society* 491.3 (Dec. 2019), pp. 4277–4293. DOI: 10.1093/mnras/stz3312 (cited on page 34).

[117] Thorben Finke, Michael Krämer, and Silvia Manconi. 'Classification of Fermi-LAT sources with deep learning using energy and time spectra'. In: *Monthly Notices of the Royal Astronomical Society* 507.3 (2021), pp. 4061–4073 (cited on page 34).

[118] Timothy D. Gebhard et al. 'Convolutional neural networks: A magic bullet for gravitational-wave detection?' In: *Phys. Rev. D* 100 (6 2019), p. 063015. DOI: 10.1103/PhysRevD.100.063015 (cited on page 34).

[119] Dmitry A Duev et al. 'Real-bogus classification for the zwicky transient facility using deep learning'. In: *Monthly Notices of the Royal Astronomical Society* 489.3 (2019), pp. 3582–3590 (cited on page 34).

[120] Tom Charnock and Adam Moss. 'Deep Recurrent Neural Networks for Supernovae Classification'. In: *The Astrophysical Journal* 837.2 (2017), p. L28. DOI: 10.3847/2041-8213/aa603d (cited on page 34).

[121] Trisha A. Hinners, Kevin Tat, and Rachel Thorp. 'Machine Learning Techniques for Stellar Light Curve Classification'. In: *The Astronomical Journal* 156.1 (2018), p. 7. DOI: 10.3847/1538-3881/aac16d (cited on page 34).

[122] Daniel Muthukrishna et al. 'RAPID: early classification of explosive transients using deep learning'. In: *Publications of the Astronomical Society of the Pacific* 131.1005 (2019), p. 118002 (cited on page 35).

[123] Minghu Zhang et al. 'Data-Driven Anomaly Detection Approach for Time-Series Streaming Data'. In: *Sensors* 20.19 (2020). DOI: 10.3390/s20195646 (cited on page 35).

[124] Marco A.F. Pimentel et al. 'A review of novelty detection'. In: *Signal Processing* 99 (2014), pp. 215–249. DOI: https://doi.org/10.1016/j.sigpro.2013.12.026 (cited on page 35).

[125] Iftach Sadeh. 'Data-driven Detection of Multimessenger Transients'. In: *The Astrophysical Journal* 894.2 (2020), p. L25. DOI: 10.3847/2041-8213/ab8b5f (cited on pages 35–37).

[126] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/ (cited on page 35).

[127] François Chollet et al. *Keras*. https://keras.io. 2015 (cited on page 35).

[128] Christoph Deil et al. 'Open high-level data formats and software for gamma-ray astronomy'. In: *6th International Symposium on High Energy Gamma-Ray Astronomy*. Vol. 1792. American Institute of Physics Conference Series. Jan. 2017, p. 070006. DOI: 10.1063/1.4969003 (cited on page 42).

[129] Harm Schoorlemmer, Jim Hinton, and Rubén López-Coto. 'Characteristics of extensive air showers around the energy threshold for ground-particle-based γ -ray observatories'. In: *European Physical Journal C* 79.5, 427 (May 2019), p. 427. DOI: 10.1140/epjc/s10052-019-6942-x (cited on page 47).

[130] Dorota Sobczyńska and Włodek Bednarek. 'Influence of clouds on the parameters of images measured by IACT at very high energies'. In: *Journal of Physics G: Nuclear and Particle Physics* 41.12 (2014), p. 125201. DOI: 10.1088/0954-3899/41/12/125201 (cited on page 47).

[131] Sobczyńska Dorota et al. 'An analysis method for data taken by Imaging Air Cherenkov Telescopes at very high energies under the presence of clouds'. In: *Astroparticle Physics* 120 (2020), p. 102450. DOI: https://doi.org/10.1016/j.astropartphys.2020.102450 (cited on page 47).

[132] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall, 1986 (cited on page 49).

[133] David W. Scott. *Multivariate density estimation : theory, practice, and visualization*. eng. Wiley series in probability and mathematical statistics. Applied probability and statistics. New York ; Wiley, 1992 (cited on page 49).

[134] C. Spearman. 'The Proof and Measurement of Association Between Two Things'. In: *American Journal of Psychology* 15 (1904), pp. 88–103 (cited on pages 49, 80).

[135] Pauli Virtanen et al. 'SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python'. In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2 (cited on page 50).

[136] G. E. P. Box. 'Non-Normality and Tests on Variances'. In: *Biometrika* 40.3-4 (Dec. 1953), pp. 318–335. DOI: 10.1093/biomet/40.3-4.318 (cited on page 51).

[137] J. Bellm et al. 'Herwig++ 2.7 Release Note'. In: *arXiv e-prints*, arXiv:1310.6877 (Oct. 2013), arXiv:1310.6877 (cited on page 57).

[138] C. Deil et al. 'Gammapy - A prototype for the CTA science tools'. In: *35th International Cosmic Ray Conference (ICRC2017)*. Vol. 301. International Cosmic Ray Conference. Jan. 2017, p. 766 (cited on page 58).

[139] E. F. Arias et al. 'The extragalactic reference system of the International Earth Rotation Service, ICRS.' In: A&A 303 (Nov. 1995), pp. 604–608 (cited on page 61).

[140] Astropy Collaboration et al. 'Astropy: A community Python package for astronomy'. In: A&A 558, A33 (Oct. 2013), A33. DOI: 10.1051/0004-6361/201322068 (cited on page 64).

[141] Astropy Collaboration et al. 'The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package'. In: AJ 156.3, 123 (Sept. 2018), p. 123. DOI: 10.3847/1538-3881/aabc4f (cited on page 64).

[142] F. van Leeuwen. 'Validation of the new Hipparcos reduction'. In: A&A 474.2 (Nov. 2007), pp. 653–664. DOI: 10.1051/0004-6361:20078357 (cited on page 64).

[143] Sameer Patel et al. 'VTSCat: The VERITAS Catalog of Gamma-Ray Observations'. In: *Proceedings of 37th International Cosmic Ray Conference — PoS(ICRC2021)*. Vol. 395. 2021, p. 812. DOI: 10.22323/1.395.0812 (cited on page 64).

[144] Guillermo Ortiz Jimenez et al. 'Redundant features can hurt robustness to distributions shift'. In: (2020) (cited on page 81).

[145] Karl Pearson. 'Note on Regression and Inheritance in the Case of Two Parents'. In: *Proceedings of the Royal Society of London Series I* 58 (Jan. 1895), pp. 240–242 (cited on page 81).

[146] P. M. Chadwick et al. 'Geomagnetic effects on atmospheric Cherenkov images'. In: *J. Phys. G* 25 (1999), pp. 1223–1233. DOI: 10.1088/0954-3899/25/6/312 (cited on page 82).

[147] A. A. Ivanov. 'Azimuthal effect on extensive air showers of cosmic rays'. In: *26th International Cosmic Ray Conference (ICRC26), Volume 1*. Vol. 1. International Cosmic Ray Conference. Jan. 1999, p. 403 (cited on page 82).

[148] S. Abdollahi et al. 'iFermi/i Large Area Telescope Fourth Source Catalog'. In: *The Astrophysical Journal Supplement Series* 247.1 (2020), p. 33. DOI: 10.3847/1538-4365/ab6bcb (cited on page 84).

[149] A. Neronov and D. Semikoz. 'Galactic diffuse gamma-ray emission at TeV energy'. In: A&A 633, A94 (Jan. 2020), A94. DOI: 10.1051/0004-6361/201936368 (cited on pages 83, 84).

[150] Mohrmann, L. et al. 'Validation of open-source science tools and background model construction in astronomy'. In: *A&A* 632 (2019), A72. DOI: 10.1051/0004-6361/201936452 (cited on page 84).

[151] null null et al. 'Time-resolved hadronic particle acceleration in the recurrent nova RS Ophiuchi'. In: *Science* 376.6588 (2022), pp. 77–80. DOI: 10.1126/science.abn0567 (cited on page 84).

[152] Andrew Y. Ng. 'Feature Selection, L1 vs. L2 Regularization, and Rotational Invariance'. In: *Proceedings of the Twenty-First International Conference on Machine Learning*. ICML '04. Banff, Alberta, Canada: Association for Computing Machinery, 2004, p. 78. DOI: 10.1145/1015330.1015435 (cited on page 97).

[153] Stanislau Semeniuta, Aliaksei Severyn, and Erhardt Barth. 'Recurrent Dropout without Memory Loss'. In: *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. Osaka, Japan: The COLING 2016 Organizing Committee, Dec. 2016, pp. 1757–1766 (cited on page 98).

[154] Geoffrey E. Hinton et al. 'Improving neural networks by preventing co-adaptation of feature detectors'. In: *CoRR* abs/1207.0580 (2012) (cited on page 98).

[155] Nitish Srivastava et al. 'Dropout: a simple way to prevent neural networks from overfitting'. In: *The journal of machine learning research* 15.1 (2014), pp. 1929–1958 (cited on page 98).

[156] Yoshua Bengio. 'Practical Recommendations for Gradient-Based Training of Deep Architectures'. In: *Neural Networks: Tricks of the Trade: Second Edition*. Ed. by Grégoire Montavon, Geneviève B. Orr, and Klaus-Robert Müller. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 437–478. DOI: 10.1007/978-3-642-35289-8_26 (cited on page 98).

[157] Zbyněk Šidák. 'Rectangular Confidence Regions for the Means of Multivariate Normal Distributions'. In: *Journal of the American Statistical Association* 62.318 (1967), pp. 626–633. DOI: 10.1080/01621459.1967.10482935 (cited on page 107).

[158] C. Alcock et al. 'The MACHO Project: Microlensing Results from 5.7 Years of Large Magellanic Cloud Observations'. In: *Astroph. J.* 542.1 (2000), pp. 281–307. DOI: 10.1086/309512 (cited on page 123).

[159] P. Tisserand et al. 'Limits on the Macho Content of the Galactic Halo from the EROS-2 Survey of the Magellanic Clouds'. In: *Astron. Astrophys.* 469 (2007), pp. 387–404. DOI: 10.1051/0004-6361:20066017 (cited on page 123).

[160] K. Griest et al. 'Experimental limits on Primordial Black Hole Dark Matter From the first 2 yr of Kepler Data'. In: *Astroph. J.* 786.2 (2014), p. 158. DOI: 10.1088/0004-637x/786/2/158 (cited on page 123).

[161] H. Niikura et al. 'Microlensing constraints on primordial black holes with Subaru/HSC Andromeda observations'. In: *Nat. Astron.* 3 (6 2019), pp. 524–534. DOI: 10.1038/s41550-019-0723-1 (cited on page 123).

[162] N. Smyth et al. 'Updated constraints on asteroid-mass primordial black holes as dark matter'. In: *Phys. Rev. D* 101 (6 2020), p. 063005. DOI: 10.1103/PhysRevD.101.063005 (cited on page 123).

[163] Konstantin Johannes Pfrang. 'Optical Microlensing by Primordial Black Holes with IACTs'. In: *PoS* ICRC2021 (2021), p. 495. DOI: 10.22323/1.395.0495 (cited on pages 123, 128).

[164] T. Hassan and M. Daniel. 'Proving the outstanding capabilities of Imaging Atmospheric Cherenkov Telescopes in high time resolution optical astronomy'. In: *PoS*. International Cosmic Ray Conference ICRC2019, 692 (2019), p. 692 (cited on page 123).

[165] Benbow W. et al. 'Direct measurement of stellar angular diameters by the VERITAS Cherenkov telescopes'. In: *Nat. Astron.* 3 (6 2019), pp. 511–516. DOI: 10.1038/s41550-019-0741-z (cited on pages 123, 126).

[166] Hans J. Witt and Shude Mao. 'Can Lensed Stars Be Regarded as Pointlike for Microlensing by MACHOs?' In: ApJ 430 (Aug. 1994), p. 505. DOI: 10.1086/174426 (cited on page 124).

[167] Kim Griest et al. 'Microlensing of Kepler Stars as a Method of Detecting Primordial Black Hole Dark Matter'. In: *Phys. Rev. Lett.* 107 (23 2011), p. 231101. DOI: 10.1103/PhysRevLett.107.231101 (cited on pages 125, 128).

[168] Michael Daniel. 'Non-Gamma-ray Applications of TeV Telescopes'. In: *PoS* ICRC2019 (2019), p. 007. DOI: 10.22323/1.358.0007 (cited on page 126).

[169] A. Chelli et al. 'Pseudomagnitudes and differential surface brightness: Application to the apparent diameter of stars'. In: *A&A* 589, A112 (2016), A112. DOI: 10.1051/0004-6361/201527484 (cited on page 128).

[170] M. Wenger et al. 'The SIMBAD astronomical database. The CDS reference database for astronomical objects'. In: *A&A Suppl. Ser.* 143 (2000), pp. 9–22. DOI: 10.1051/aas:2000332 (cited on page 128).

[171] Geier, S. 'The population of hot subdwarf stars studied with Gaia - III. Catalogue of known hot subdwarf stars: Data Release 2'. In: *A&A* 635 (2020), A193. DOI: 10.1051/0004-6361/202037526 (cited on page 128).

[172] U. Heber. 'The atmosphere of subluminous B stars. II. Analysis of 10 helium poor subdwarfs and the birthrate of sdB stars.' In: *A&A* 155 (1986), pp. 33–45 (cited on page 128).

[173] K. Griest. 'Galactic Microlensing as a Method of Detecting Massive Compact Halo Objects'. In: *Astrophysical Journal* 366 (1991), p. 412. DOI: 10.1086/169575 (cited on page 128).

[174] K. Griest et al. 'New Limits on Primordial Black Hole Dark Matter from an Analysis of Kepler Source Microlensing Data'. In: *Phys. Rev. Lett.* 111 (18 2013), p. 181302. DOI: 10.1103/PhysRevLett.111.181302 (cited on page 128).

[175] M. Jaroszynski and B. Paczynski. 'Diffraction Effects in Microlensing of Q2237+0305'. In: ApJ 455 (Dec. 1995), p. 443. DOI: 10.1086/176593 (cited on page 130).

[176] T. Naderi, A. Mehrabi, and S. Rahvar. 'Primordial black hole detection through diffractive microlensing'. In: *Phys. Rev. D* 97 (10 2018), p. 103507. DOI: 10.1103/PhysRevD.97.103507 (cited on page 130).

Acknowledgements

I feel blessed for everyone I met while preparing this thesis. Particularly, I want to thank

Dr. Elisa Pueschel, who gave me the chance to work on this thesis and found a balance between providing close supervision and yet leaving me the required freedom

Dr. Iftach Sadeh, for every discussion we had, the feedback you provided, and for always keeping an eye on the bigger picture

Dr. Tarek Hassan, who introduced me to scientific working during my first years and helped me to find my niche

Prof. Dr. Martin Pohl, for taking the responsibility to be my official supervisor

Dr. Gernot Maier, for always supporting me with information about VERITAS and its data analysis

Victor, for being a friend far beyond DESY's gate

Chiara, who was my office mate for most of my Ph.D. and also smiled with me through the tough times

Marianna, Mireia, Pedro, and Tobias, who have been great colleagues and friends on and off DESY during my PhD

My greatest gratitude is to **my Family**, Birgit, Michael, and Sebastian, for being the most significant source of support. This work would have never been possible without you.

Selbstständigkeitserklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig und nur mit den angegeben Quellen und Hilfsmitteln angefertigt habe. Alle Stellen der Arbeit, die ich aus diesen Quellen und Hilfsmitteln dem Wortlaut oder dem Sinne nach entnommen habe, sind kenntlich gemacht und im Literaturverzeichnis aufgeführt. Weiterhin versichere ich, dass ich die vorliegende Arbeit, weder in der vorliegenden, noch in einer mehr oder weniger abgewandelten Form, als Dissertation an einer anderen Hochschule eingereicht habe.

Potsdam, den 26. Oktober, 2022

Konstantin Johannes Pfrang