technische universität
dortmund

Arbeit zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften
(Dr. rer. nat.)

Observation of High-Energy Neutrinos from the Milky Way

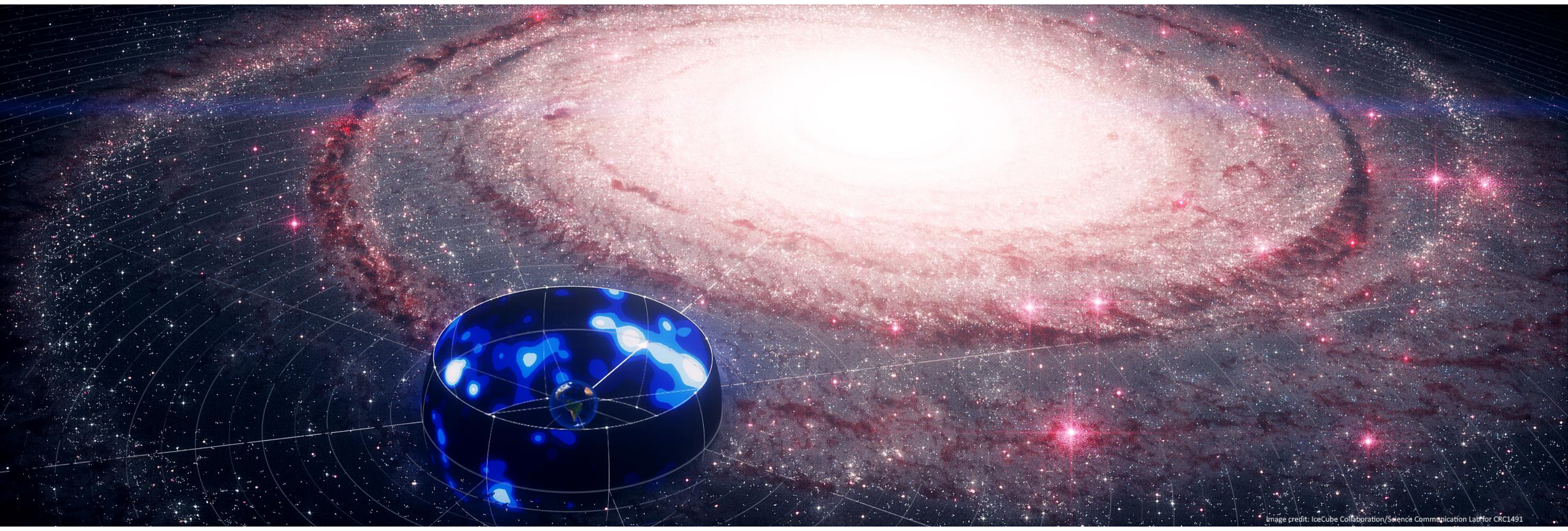


Image credit: IceCube Collaboration/Science Communication Lab for CRC1491

Mirco Hünnefeld
2023

Astroteilchenphysik
Fakultät Physik
Technische Universität Dortmund

Erstgutachter:	Prof. Dr. Dr. Wolfgang Rhode
Zweitgutachter:	Prof. Dr. Julia Tjus
Abgabedatum:	20. Juni 2023

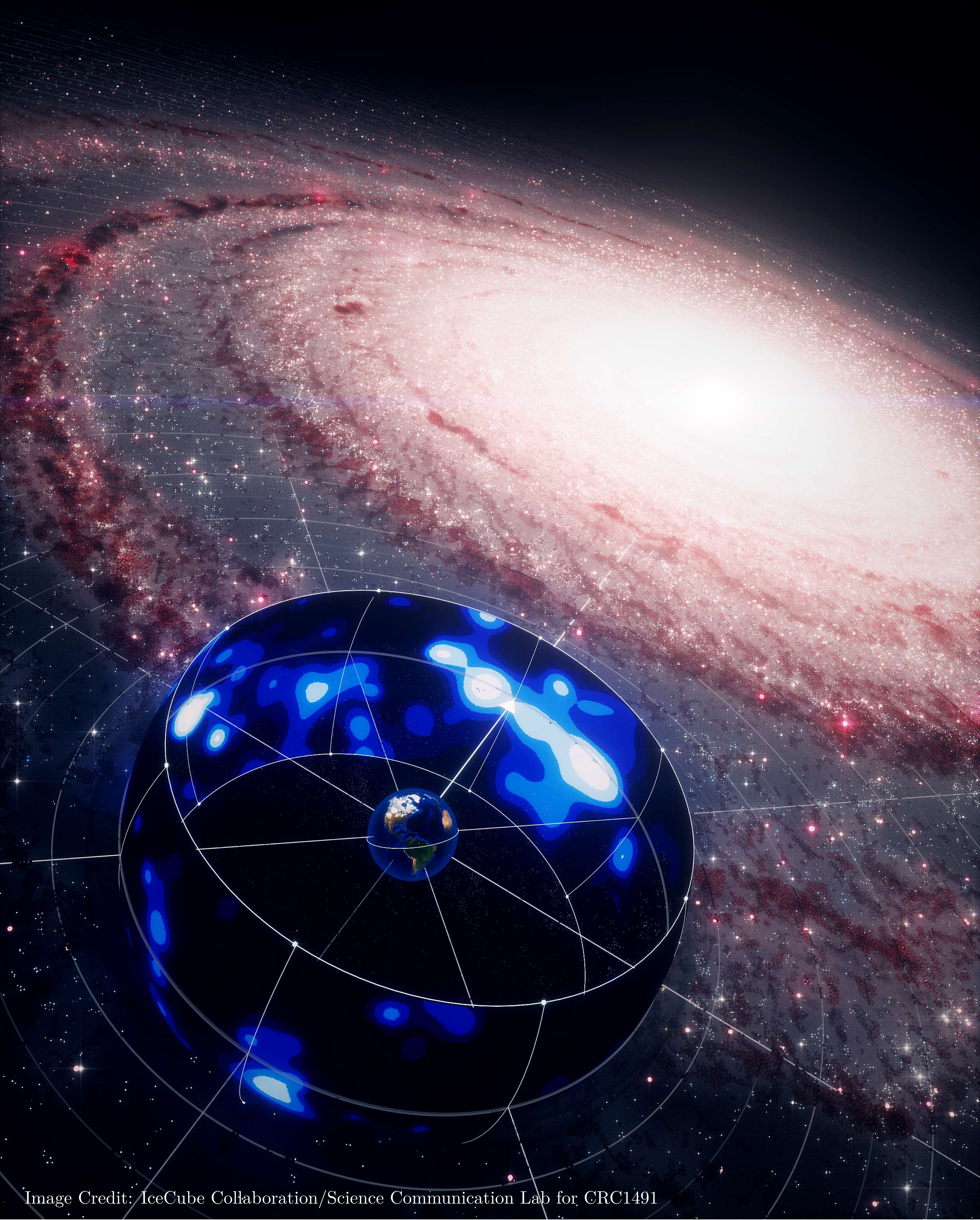


Image Credit: IceCube Collaboration/Science Communication Lab for CRC1491

Abstract

With the discovery of the astrophysical neutrino flux by IceCube in 2013, the foundation for neutrino astronomy was established. In subsequent years, the first point-like neutrino source candidates, a flaring blazar known as TXS 0506+056 and the active galaxy NGC 1068, emerged. Now, in this work, a new milestone in the rising field of neutrino astronomy is presented: the first observation of high-energy neutrinos from our own Galaxy — the Milky Way. A search for Galactic neutrino emission is performed on 10 years of IceCube data, rejecting the background-only hypothesis at the 4.5σ significance level. The observed Galactic neutrino flux, believed to originate from diffuse interactions of cosmic rays, possibly in addition to contributions from unresolved point-like sources, may explain up to 10 % of the astrophysical neutrino flux previously measured by IceCube. This observation is enabled by novel tools based on deep learning, developed in this dissertation. In comparison to prior IceCube analyses, the sensitivity is improved by a factor up to four, due to improved event reconstructions and an increased effective area by over an order of magnitude. These tools not only lead to the most sensitive neutrino dataset to date in the Southern Sky, but they also enable a wide variety of future applications and analyses that were previously unattainable.

Kurzfassung

Mit der Entdeckung des astrophysikalischen Neutrinoflusses durch IceCube im Jahr 2013 wurde der Grundstein für die Neutrinoastronomie gelegt. In den folgenden Jahren wurden die ersten Kandidaten für punktförmige Neutrinoquellen beobachtet, ein aufflackernder Blazar mit der Bezeichnung TXS 0506+056 und die aktive Galaxie NGC 1068. In dieser Arbeit wird nun ein neuer Meilenstein auf dem aufstrebenden Gebiet der Neutrinoastronomie vorgestellt: die erste Beobachtung von hochenergetischen Neutrinos aus unserer eigenen Galaxie — der Milchstraße. Mit der durchgeführten Analyse über 10 Jahre IceCube-Daten wird die Hintergrundhypothese mit einer Signifikanz von 4.5σ abgelehnt. Der beobachtete galaktische Neutrinofluss, vermutlich aus diffusen Wechselwirkungen der kosmischen Strahlung stammend, möglicherweise mit zusätzlichen Beiträgen aus unaufgelösten Punktquellen, könnte bis zu 10 % des zuvor von IceCube gemessenen astrophysikalischen Neutrinoflusses erklären. Diese Beobachtung wird durch neuartige, auf Deep Learning basierende Werkzeuge ermöglicht, die in dieser Dissertation entwickelt wurden. Im Vergleich zu früheren IceCube-Analysen wurde die Sensitivität um bis zu einem Faktor vier verbessert, was auf verbesserte Winkelrekonstruktionen und eine effektive Fläche, die um mehr als eine Größenordnung gesteigert wurde, zurückzuführen ist. Diese Werkzeuge führen nicht nur zum bisher sensitivsten Neutrino-Datensatz am Südhimmel, sondern sie ermöglichen auch eine Vielzahl von zukünftigen Anwendungen und Analysen, die bisher nicht möglich waren.

Contents

1 Introduction

Event reconstruction and classification, along with detector calibration, constitute core tasks vital to any physics experiment. Although often undervalued, advancements in these tasks may improve an experiment's sensitivity on similar scales as expensive detector upgrades or even the construction of entirely new detectors. This dissertation is largely dedicated to the development and improvement of such core components in the IceCube analysis chain. As such, the presented work demonstrates the potential that lies in method development, ultimately leading to the first observation of high-energy neutrinos from the Milky Way. In particular, this dissertation introduces deep learning-based reconstruction frameworks to the toolkit in IceCube. These are used as standalone tools in event reconstruction and classification tasks, but also in combination with traditional methods based on maximum-likelihood estimation.

In Part I, the concept of domain knowledge and the importance of its utilization is outlined. Two software frameworks are introduced that build upon this concept. A general reconstruction framework based on convolutional neural networks is developed in addition to a hybrid approach that combines maximum-likelihood estimation with deep learning. These methods allow for more accurate event reconstruction and for the creation of more efficient event samples as demonstrated in Part II. The improved methods are applied to searches for neutrino emission in IceCube's cascade channel, culminating in the most sensitive search for Galactic neutrino emission to date, detailed in Part III.

Parts of this dissertation have been published in peer-reviewed articles, books, and conference proceedings authored or co-authored by the author of this dissertation [1–7]. These parts, which may follow the published work, are indicated by gray "declaration" boxes at the beginning of the relevant section. Similarly, work carried out jointly with collaborators in the IceCube collaboration is identified in these declarations. The stated collaborator affiliation corresponds to the affiliation at the time at which the collaboration was carried out.

Part I

Redefining Event Reconstruction in IceCube

Every IceCube analysis relies on reconstruction methods to translate detected photons, measured as electric currents over time, to abstract event properties such as topology, particle direction, energy, and type. These reconstruction methods therefore constitute core components to any analysis performed in IceCube. As such, their performance is imminent and crucial to the outcome of the final analysis result. Event reconstruction methods thus play a central role to the success of IceCube's physics program.

In this dissertation, novel methods based on machine and deep learning are developed and introduced to the IceCube Neutrino Observatory, a neutrino telescope detailed in Chapter 3. These methods expand IceCube's toolset and capabilities in key areas, such as event classification and reconstruction tasks, but also for analysis level methods. In particular, a general software framework based on convolutional neural networks is introduced in Chapter 4. Chapter 5 highlights a hybrid reconstruction method that combines maximum-likelihood estimation with deep learning. Core concepts to these frameworks, built upon machine learning, is the utilization of available symmetries and domain knowledge, as discussed in the following Chapter 2.

2 Event Reconstruction and Domain Knowledge

From an epistemological perspective, testing of scientific hypotheses through dedicated experiments is a crucial step in the process of knowledge discovery in physics [8–10]. These hypotheses are tested by comparing data measured by detectors to the predictions made from theory. In particle physics, such theories can, for instance, predict the amount of particles with certain properties that arrive at the detector. These particles may undergo interactions in the detector volume, leading to measurable quantities such as electric currents and timestamps in photomultiplier tubes. A collection of these currents and timestamps over a certain time window is also referred to as an event.

In order to test the theory hypothesis, a connection must be established between the measured currents and timestamps in an event and the particle that caused these. This process of translating measured quantities to particle properties is referred to as event reconstruction. Event reconstruction relies on a model that connects the measured data with the event or particle properties. Part I is dedicated to various ways on how this model may be defined. Crucial to the success of these models is how accurate and complete the underlying generating process, which connects the measured data to the event properties, is described and how well its patterns and symmetries are utilized.

In principle, there are two main approaches in how this model is constructed. The first of which is a data-driven approach via machine learning. In this approach, only a limited amount of underlying symmetries and domain knowledge are used explicitly, while the majority is learned implicitly from the training data. The other approach, predominately via maximum-likelihood estimation (MLE), aims to explicitly define and exploit the generating process. However, this is often intractable in practice and therefore MLE commonly resorts to simplified models and imposed assumptions.

In addition to these two approaches, one can further distinguish between the direction in how the relation of measurable quantities and generating process is defined. In MLE, a model is built in “forward” direction: starting from a given hypothesis, the MLE model predicts the measurable quantities, which can then be compared to the measured data to compute the likelihood of the data under the given hypothesis. Typical machine learning models, on the other hand, go the reverse direction: starting from the measured data, these models attempt to directly estimate the posterior distribution of the physics quantities of interest. Both directions are viable solutions. However, as shown in the following sections, the construction of a model in “forward” direction facilitates the utilization of domain knowledge such as symmetries and physics laws. The reason for this is that the "forward" direction is analogous to how Monte Carlo (MC) simulations are set up. Any information applied in the simulation chain (symmetries, laws, detector effects, etc.) may also be used in the same manner in the construction of the model. This is not trivial in the "backward" direction, as many effects are convolved and often lead to degeneracies in the measured data.

In this dissertation, a data-driven approach via deep neural networks is introduced in addition to a hybrid approach that is designed to combine the strengths of both MLE and deep learning. Much of the discourse in this dissertation focuses on the utilization of explicit domain knowledge in these data-driven models. To emphasize the importance of explicit utilization of domain knowledge, some hypothetical examples are described in the following Section 2.1.

2.1 The Importance of Domain Knowledge

Declaration

This section follows the argumentation in Ref. [5], while adding explicit examples to demonstrate the importance of utilizing domain knowledge. The first paragraphs of this section are taken over verbatim as indicated by the quotation marks. Ref. [5] was written by the author of this dissertation and reviewed by the IceCube collaboration.

> “The reconstruction of neutrino events in the IceCube detector has traditionally relied on maximum-likelihood-based methods [11, 12]. More recently, deep learning architectures such as convolutional neural networks (CNN) [13] have surpassed traditional reconstruction methods in certain areas including high-energy cascade reconstruction [6]. Further applications in IceCube, also utilizing recurrent and graph neural networks, are illustrated in Refs. [3, 14–19].
>
> These applications based on deep learning illustrate a paradigm shift from explicit to implicit use of available information. In likelihood-based methods, domain knowledge, such as translational and rotational invariance, detector characteristics and physics laws, are implemented directly into the likelihood prescription. The aforementioned deep learning applications, however, must learn this information implicitly through the training data. CNNs are able to explicitly utilize translational invariance to a certain degree, but they lack the ability to directly include other information.
>
> The strength of deep learning lies in the universality of its methods. Typical deep

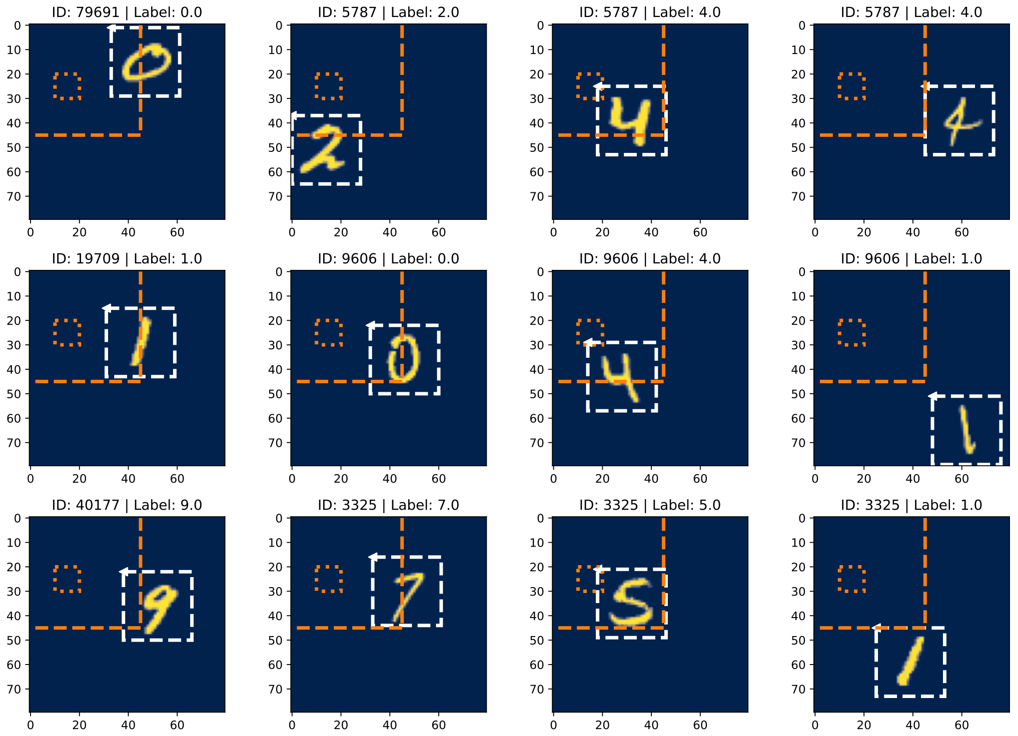
> learning architectures were developed for a generalized usage in a wide field of applications. While this enables the application across many different domains, it neglects potential benefits from explicit exploitation of domain knowledge. One such application is the field of image recognition. Input data in this field consist of an array of pixel values. Due to the broad range of this domain, the underlying generating process for these pixel values may not be known. However, certain information, such as scale, rotational and translational invariance, may still be shared across applications. Amongst other reasons, CNNs led to a breakthrough [20] in this field by exploiting common domain knowledge such as the importance of local pixels and translational invariance.
>
> In contrast to image data, the underlying generating processes for data in particle physics experiments are well understood. These experiments often employ extensive simulations, which implies that the physical processes and detector response are known to great detail. Maximum-likelihood-based methods aim to utilize the full extent of this information. However, due to computational limitations, these methods are often forced to apply simplifications and approximations. Standard deep learning architectures perform well for these tasks, but they lack the ability to fully exploit available domain knowledge. Similarly to CNNs for image recognition, the explicit utilization of domain knowledge in these architectures may help to advance the field of event reconstruction in particle physics experiments. ”[5]

In the following sub-sections, some hypothetical examples are constructed that demonstrate the importance of utilizing domain knowledge. Only by exploiting symmetries and patterns, can a model meaningfully inter- and extrapolate beyond the phase space covered by training data.

2.1.1 Example 1: Translational Invariance

The Modified National Institute of Standards and Technology (MNIST) [21] database of handwritten digits is used to demonstrate the importance of exploiting translational invariance. Images of the MNIST dataset are sampled and embedded at random locations in a larger image as demonstrated in Figure 2.1. Two exclusion regions are defined in which no MNIST images are placed for the training or test datasets. A standard, fully-connected neural network (NN) and a CNN are trained to classify the generated images (details on these architectures are provided in Section 3.2 [Machine and Deep Learning]). The trained models are then evaluated on the training and test dataset, as well as on images of each of the exclusion regions.

As shown in Figure 2.2, the fully-connected NN is not able to achieve the same performance as the CNN, despite having two orders of magnitude more trainable



(a) Training. (b) Test. (c) Interpolation. (d) Extrapolation.

Figure 2.1: Example images are shown for the training and test dataset as well as for the interpolation and extrapolation regions. These exclusion regions are indicated by dotted (interpolation) and dashed (extrapolation) orange lines. MNIST images (dashed, white lines) are inserted at random positions, under the constraint that the upper left corner (white triangle) is outside the specified exclusion regions for training and test data.

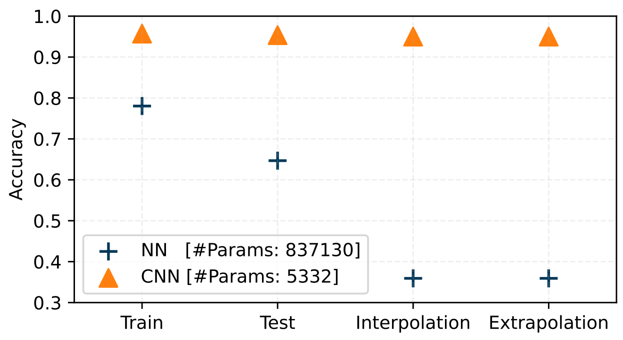


Figure 2.2: The prediction accuracy is shown for the fully-connected (NN) and convolutional neural network (CNN) for each of the datasets and exclusion regions. The CNN is capable of extrapolating to unseen data by exploiting symmetries.

parameters. Over-training is visible on the test dataset and there is poor performance for images sampled in the exclusion regions. In contrast, the CNN demonstrates a stable performance even when extrapolating to images in the exclusion regions. This robust extrapolation is enabled by utilizing translational invariance within the CNN architecture.

2.1.2 Example 2: Periodicity

In this example, the goal is to fit a one-dimensional, periodic function with a known period length as shown in the dark, solid line in Figure 2.3. Two scenarios are investigated in which 100 and 1000 training samples are generated for training. Training points are not generated for the exclusion regions indicated by shaded areas. A standard fully-connected NN is trained on the generated training set. In order to utilize the known period length T, a second fully-connected NN is trained on a modified input $x' = x \mod T$. Predictions for both models are shown in Figure 2.3. In contrast to the standard NN, the periodic NN is able to model the true underlying function well, despite the small number of training samples. This is achieved by effectively mapping training samples onto a smaller phase space by utilizing the symmetry in time. Similarly, the periodic NN is able to extrapolate into the exclusion regions due to the explicit exploitation of this domain knowledge. The standard NN is unable to find and exploit this pattern.

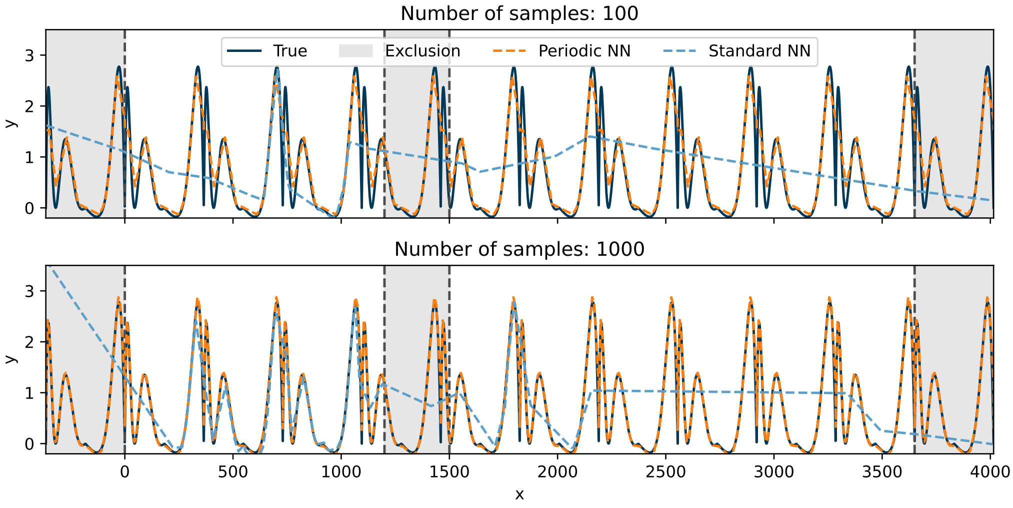


Figure 2.3: The predicted curves for the standard and periodic NN are overlaid on the true curve for each of the two training cases including 100 and 1000 data points in the training sample. Shaded regions are excluded from training.

2.1.3 Example 3: Known Approximate Solution

Similar to the previous example, a one-dimensional function $f(x)$ is modeled by two neural networks for varying number of sampled training points. In this case, an approximate solution, defined as $f_0(x)$, is known a priori. A standard NN is trained on the sampled points in addition to a second NN that employs a modification to utilize the approximate solution. This "modified" network is defined as $g(x) = f_0(x) \cdot \mathrm{NN}(x)$, where NN is a neural network with an identical architecture to the standard NN. In contrast to the standard NN, the modified NN therefore only has to learn a variable correction factor that is applied to the known approximate solution $f_0(x)$. Analogously to the previous examples, the exploitation of available domain knowledge — in this case the known approximate solution — allows for a more accurate modeling with less required training samples, while also enabling extrapolation to unseen data along the utilized symmetries.

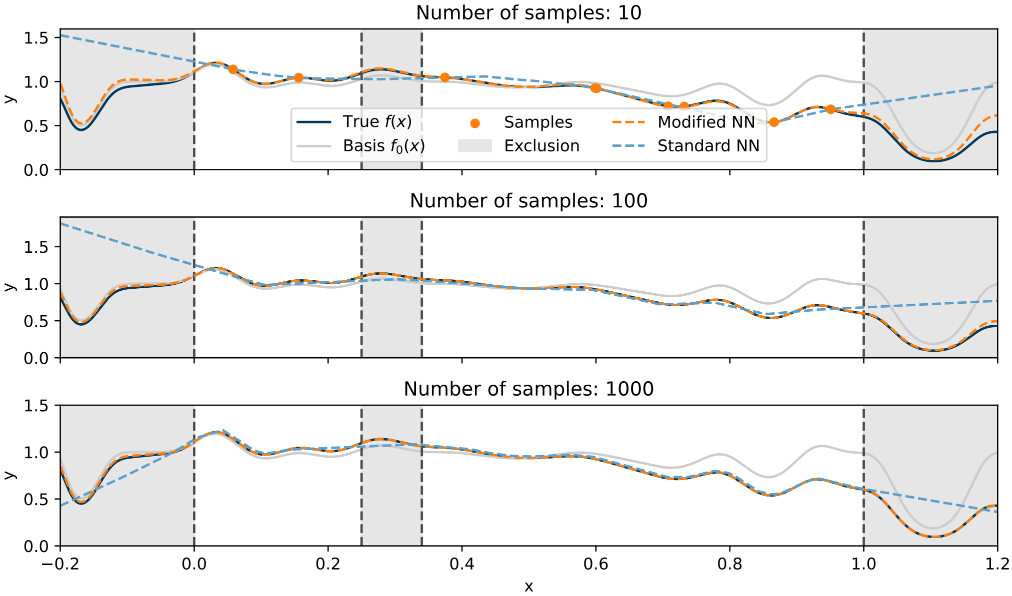


Figure 2.4: The predicted curves for the standard and modified NN are overlaid on the true curve for each of the three training cases including 10, 100 and 1000 data points in the training sample. Sampled training points for the first case are shown in the top panel as orange dots. Shaded regions are excluded from training.

3 Event Reconstruction in IceCube

Declaration

Parts of this chapter closely follow the description in Ref. [6], written by the author of this dissertation.

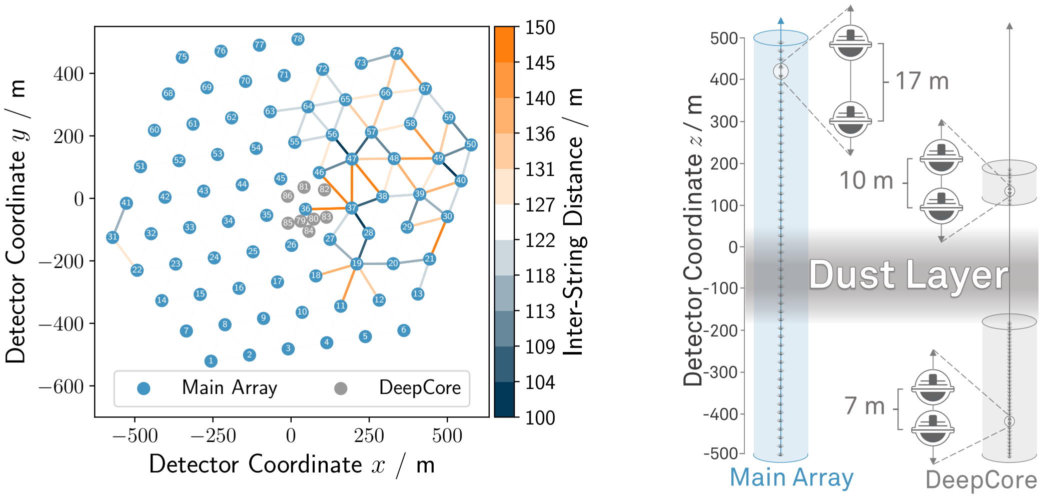
Reconstruction methods in IceCube have traditionally relied on maximum-likelihood estimation. While theoretically optimal, these methods pose challenges in practical applications. In this section, the IceCube detector is introduced. Afterwards, the basics on maximum-likelihood estimation and machine and deep learning are laid out, as these constitute core concepts to the developed reconstruction methods in this dissertation.

3.1 The IceCube detector

Declaration

The introduction to IceCube is taken verbatim from Ref. [6], written by the author of this dissertation, as indicated by the quotation marks.

> “The IceCube Neutrino Observatory is a neutrino detector located at the geographic South Pole, instrumenting a cubic kilometer of glacial ice. The detector consists of 5160 digital optical modules (DOMs) with a downward-facing 10 inch-diameter photomultipler tube (PMT) [22] (Figure 3.4a) installed on 86 vertical strings at depths between 1450 m and 2450 m. The origin of the IceCube coordinate system is placed in the detector center at a depth of 1948 m. The z-axis is chosen to point upwards towards the ice surface. PMT signals are digitized and buffered on the DOM mainboard with a timing resolution of about 2 ns. Upon readout request, these digitized waveforms are sent to computers on the surface of the detector array.
>
> The 86 strings of the detector may be grouped into three detector parts. While the main IceCube array, consisting of the first 78 strings, is arranged on an approximately triangular grid, the remaining 8 strings do not follow this symmetry. They are installed in a denser configuration called DeepCore [23] with variable distances to neighboring strings. Each of the 86 strings holds 60 DOMs. These are evenly distributed along the z-axis for strings of the main array. DOMs located on DeepCore strings are further grouped into an array above the dust layer (10



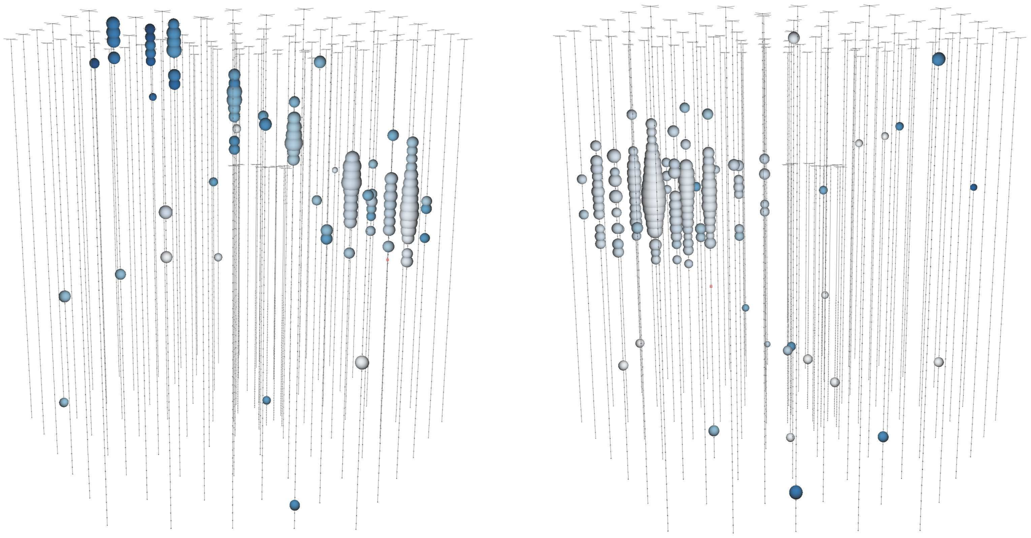
(a) Top view of IceCube

(b) Side view of Strings

Figure 3.1: A top view of the IceCube detector is shown on the left. The in blue depicted 78 strings are on an approximately triangular grid, while the DeepCore strings, shown in gray, are installed in a denser configuration. The color scale indicates the inter-string distance for the main IceCube array, which can substantially deviate from the usual spacing of 125 m. On the right, the DOM layout along the z-axis of the strings is illustrated. In contrast to the strings of the main array, the DOMs on the DeepCore strings (gray) are divided into two groups, one above the dust layer and one below. Figure and caption adopted from Ref. [6].

DOMs on each string) and an array below the dust layer (50 DOMs per string), as illustrated in Fig. 3.1b. The dust layer is a layer in the glacial ice with increased dust impurities that was produced about 60000 to 70000 years ago [24, 25]. It ranges from depths of 1950 m to 2100 m and results in increased scattering and absorption coefficients. The strings of the main IceCube array have an inter-string spacing of about 125 m. However, there are deviations as shown in Fig. 3.1a.

At trigger level, IceCube data consist of recorded waveforms where the amplitude of the waveform corresponds to the charge recorded by the DOM. An IceCube “event” is a set of these waveforms for many DOMs. Each DOM can measure multiple waveforms with variable starting times in a single event. While a typical event has a read out window of approximately 15 000 ns, this can vary depending on the trigger [26, pp.56-59]. In a subsequent step, a series of pulses, represented as a series of times and amplitudes of light observed in a DOM, are extracted from the recorded waveforms. The number of extracted pulses can vary by orders of magnitude and it is different for each DOM. A detailed description of the IceCube detector and its data acquisition system can be found in Ref. [26].



(a) Track Event.

(b) Cascade Event.

Figure 3.2: Example event views of simulated data are shown for an up-going muon entering the detector (left) and for the resulting particle cascade of a charged-current (CC) ν_e interaction inside the detector volume (right). Each DOM is represented by a sphere. The size of the sphere corresponds to the amount of collected light and the color indicates the arrival time of the photons (darker colors correspond to later times). Figure and caption taken from Ref. [6].

In order to detect neutrinos, IceCube measures Cherenkov photons produced by charged secondary particles resulting from neutrino interactions. The two primary detection channels consist of so called track-like events (tracks), muons induced by charged-current (CC) ν_μ interactions, as well as cascade-like events (cascades), which result from CC ν_e and ν_τ interactions in addition to neutral current interactions of all neutrino types [27, p. 6]. Muons deposit energy along their trajectory through the detector resulting in tracks as shown in Fig. 3.2a. Cascade-like events, on the other hand, produce a shower of secondary particles at the neutrino interaction vertex, which is generally not resolvable by IceCube, given its inter-string spacing of about 125 m. As a result, IceCube detects a spherical energy deposition as shown in Fig. 3.2b.

In contrast to track-like events, the missing lever arm and the spherical energy deposition make the angular reconstruction of cascades a challenging task in IceCube. Furthermore, cascade reconstructions are more susceptible to local ice properties than track reconstructions due to the local energy deposition of the cascade. For

> tracks, the energy depositions are distributed over large parts of the detector, which helps to average out local fluctuations. It is therefore crucial to understand the effect that systematic uncertainties in the description of the local ice properties may have on the reconstruction.”[6]

Commonly considered systematic uncertainties in IceCube consist of the bulk scattering and absorption coefficients of the glacial ice that affect the propagation of photons [24, 25], the DOM quantum efficiency at which incident photons are detected, the acceptance parameterization of the refrozen ice column surrounding the DOMs [28], and the ice anisotropy [29, 30], which leads to direction dependent scattering and absorption of photons as well as curved light trajectories due to birefringence. A set of systematic parameters describing the glacial ice in IceCube is also referred to as an ice model. Ice models are typically optimized in a high-dimensional fitting procedure to model calibration data runs (flasher runs) [31]. Despite these modeling challenges, cascade events are an important detection channel that expand IceCube's ability to probe the Southern Neutrino Sky, which is otherwise dominated by down-going atmospheric muons.

3.1.1 Detector Extensions: IceCube Upgrade and IceCube-Gen2

Extensions to the IceCube detector are envisioned to further improve the sensitivity of the detector. The IceCube Upgrade [32–34] will extend IceCube's energy threshold down to a few GeV by adding 7 strings with approximately 700 optical sensors within the lower DeepCore region in a dense spacing of 20 m (horizontal) and 3 m (vertical). A future high-energy extension, IceCube-Gen2 [35], will consist of a radio array in addition to an extended optical array. The module spacing for the optical array is planned for 240 m (horizontal) and 17 m (vertical), while the radio array employs an even larger spacing, as illustrated in Figure 3.3.

In contrast to the optical modules deployed in the current IceCube detector (Gen1 DOM), both of the extensions will employ multi-PMT modules, shown in Figure 3.4. The IceCube Upgrade will primarily rely on the multi-PMT digital optical module (mDOM) [37] and the D-Egg (Dual optical sensors in an Ellipsoid Glass for Gen2) [32]. Utilization of multiple PMTs in one module allows for an increased photon collection area, while also providing directional information on the detected photons.

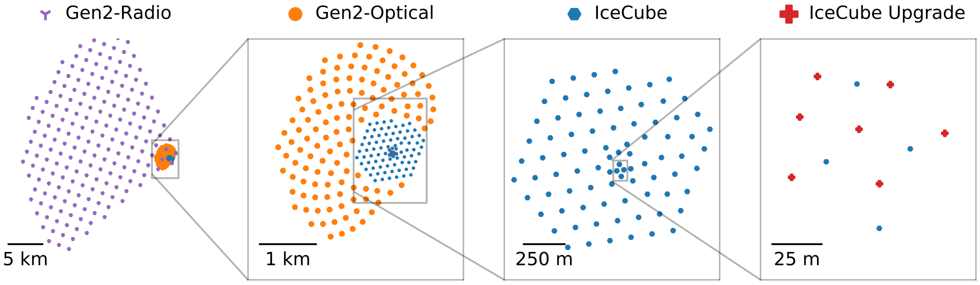


Figure 3.3: A top view is shown for the various components of the IceCube-Gen2 facility, which includes the current IceCube detector (blue) and the low-energy extension IceCube Upgrade (red). Figure taken from Ref. [36].

(a) Gen1 DOM (b) mDOM (c) D-Egg (d) mEgg (e) LOM−18

Figure 3.4: A schematic view of optical modules is shown. Gen1 DOMs [26] are used for the current IceCube detector, while mDOM [37] and D-Eggs [32] will be deployed in the IceCube-Upgrade; mEgg and LOM−18 are two of various potential designs for IceCube-Gen2 [38, 39]. Figure taken from Ref. [38].

3.2 Machine and Deep Learning

Machine learning (ML) is a subfield of artificial intelligence (AI) describing computer systems that are able to recognize and adapt patterns in existing data to make inferences, without following explicit instructions. The general goal of these systems is to model a function dependent on trainable parameters (θ):

$$f_\theta : X \mapsto \hat{Y} \tag{3.1}$$

that maps some input data (X) to a set of desired quantities ($Y \approx \hat{Y}$). Deep learning is a class of machine learning algorithms that commonly utilize artificial neural networks to model the relation defined in Eq. (3.1). Neural networks are particularly interesting for this task due to their property as universal approximators [40].

Artificial neural networks have been a field of research for many decades. Initial ideas on training multilayer neural networks via stochastic gradient descent in combination with backpropagation were described by different authors in the 1970s and 1980s [41–44]. However, due to lacking success in the training of these networks, the community had mostly abandoned this approach by the late 1990s. Research on neural networks regained momentum in the early 2000s with major breakthroughs following soon after in speech [45,46] and image recognition tasks [47], demonstrating the capabilities of deep neural networks. These breakthroughs were enabled by hardware innovations [48], increasing sizes of datasets, and new developments in the construction and training of neural network architectures. A new regularization technique, referred to as dropout [49], ReLU activiation functions [50], and data augmentation were a central element in the success story of deep neural networks [51]. Further details on deep learning are provided in Refs. [51, 52].

3.2.1 Artificial Neuron (Perceptron)

Neural networks are constructed from basic components, referred to as artificial neurons or perceptrons. The combination of such neurons in multiple layers, as illustrated in Figure 3.5, is called a multilayer perceptron (MLP).

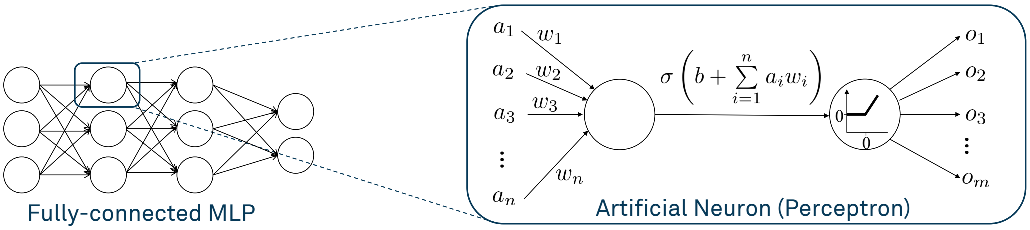


Figure 3.5: Neural network architectures, such as multilayer perceptrons (MLP), are comprised of multiple layers of interconnected artificial neurons (perceptrons).

In a feedforward network, information flows in one direction. An artificial neuron accumulates information over the inputs a_i. These inputs are either the outputs o_j of previous neurons, or the data provided to the network, if the neuron is part of the first layer in the network (input layer). A weighted sum over the inputs and weights (w_i) is performed. An offset via the bias term (b) is added, prior to applying a non-linear activation function $\sigma(x)$. Non-linear activation functions are required to allow the model to parameterize non-linear relations. A common choice for activation functions are rectified linear units (ReLU) [50]. The weights and bias term are free parameters that are tuned to the data during training (Section 3.2.4).

3.2.2 Fully-Connected Layers

A specific configuration of layers of neurons and choice of activation functions is referred to as the *architecture* of a neural network. On the left of Figure 3.5, a fully-connected MLP with four layers is shown. The first layer is utilized to provide data to the network and it is thus referred to as the input layer. The last layer defines the output of the network, and intermediate layers are referred to as hidden layers. In fully-connected (dense) layers, every neuron is connected to every neuron of the previous layer, resulting in a total of $(n + 1) \cdot k$ trainable parameters for n and k neurons in the previous and current layer.

3.2.3 Convolutional Layers

Convolutional neural networks (CNNs) [13] utilize convolutional layers that exploit translational invariance and locality to reduce the number of trainable parameters. Instead of connections to every neuron in the previous layer as in fully-connected (dense) layers, only connections to nearby neurons are established, as illustrated in Figure 3.6. The output of a neuron is computed analogously to Figure 3.5 as a weighted sum over the inputs utilizing a set of weights equal in length to the amount of connected neurons. This set of weights is shared across neurons in the same layer and referred to as the convolution kernel or filter. This re-use of weights leads to a reduction of trainable parameters and the exploitation of translational invariance: a feature in one part of the input layer will produce the same response as the same feature in another part of the input layer.

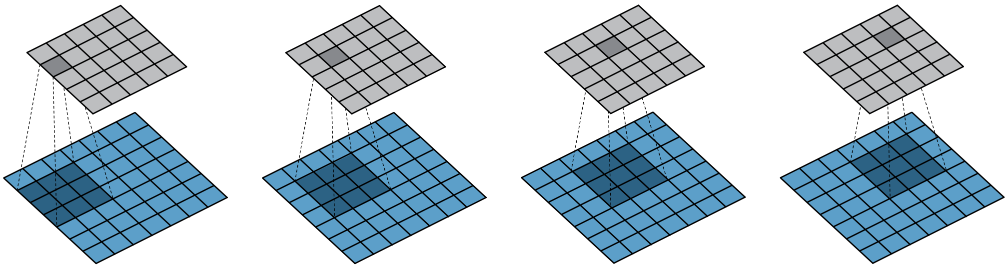


Figure 3.6: In convolutional layers, a sliding dot product is performed by shifting a set of weights (dark blue), referred to as a kernel or filter, over the input space (light blue). In this case, a two-dimensional convolution with a single 3×3 kernel is demonstrated. The result of the dot product is further modified by a bias term and subsequent non-linear activation function, similar to the computation at an artificial neuron illustrated on the right of Figure 3.5. When applied at every position in the input space, the output of the convolutional layer (gray) is obtained.

In Figure 3.6, the convolution is illustrated by a single kernel that is applied across the input layer (blue) to compute the output (gray). However, a convolutional layer may consist of many such kernels, which each create an output referred to as a feature map or channel. Similarly, a convolutional layer can have more than one input channels.

3.2.4 Training of Neural Networks

The trainable parameters of neural networks are tuned by optimizing an objective function ($\mathcal{L}$). This objective function, also referred to as the loss function, quantifies the performance of the neural network. In a supervised learning task, in which the true values for Y are known, the loss function is typically defined as a distance measure between the network's prediction $\widehat{Y} = f_{\theta}(X)$ and the true values Y.

During training of the neural network, batches of n examples are drawn from the training data. The data is passed through the network to evaluate the loss function $\mathcal{L}(\widehat{Y}, Y)$. The loss function is then iteratively minimized by adjusting the parameters of the model via stochastic gradient descent [53]. The gradients of the loss function in regard to the free parameters $\frac{\partial L}{\partial \theta}$ are calculated for the given batch of data via backpropagation [54], an application of the chain rule. Initially, the output of the network $\widehat{Y} = f_{\theta}(X)$ will be random, due to the random initialization of the free parameters θ of the model. These steps are repeated many times, until the loss converges in a minimum [51, 54].

3.3 Maximum-Likelihood Estimation

In contrast to methods based on machine learning, maximum-likelihood estimation requires detailed knowledge of the underlying statistical model. Maximum-likelihood estimation (MLE) is a commonly utilized method for parameter estimation due to its desirable properties for large sample sizes such as consistency, efficiency, and asymptotic normality. Consistency states that the parameter estimates converge to the true values with an increasing sample size. An efficient estimator achieves the smallest possible variance among all consistent estimators, and asymptotic normality indicates that the distribution of maximum-likelihood estimates becomes normal in the limit of infinite sample size.

A parametric model with unknown parameters θ is defined that describes the data distribution. The parameters of this assumed model are then estimated by maximizing the likelihood function, which quantifies the probability of observing the data given the set of parameter values. MLE is provably optimal in the limit of infinite sample sizes and in the case that the assumed parametric model correctly describes the data distribution. For small sample sizes these properties may be violated, and incorrect assumptions in the chosen model may lead to biases in the MLE estimate and to the construction of confidence intervals with improper coverage. An overview on maximum-likelihood estimation and hypothesis testing is provided in Ref. [55].

3.4 Traditional Reconstruction Methods in IceCube

The task of reconstruction methods in IceCube is to translate measured pulses to event properties such as the deposited energy, event direction, interaction vertex, and particle type. Most traditional and widely used methods in IceCube are based on maximum-likelihood estimation (Section 3.3). These methods aim to determine the event properties that best describe the data, by maximizing the likelihood of the observed light pattern under a given event hypothesis.

To first order, the various MLE-based reconstruction methods in IceCube are differentiated in the way the underlying probability density functions (PDFs) are modeled. These PDFs describe the relation between the event hypothesis and the expected light deposition pattern in the detector. Multiple variations on how the PDFs are constructed exist, including analytical descriptions [56] and computing-intensive re-simulations [57–60]. In theory, direct re-simulations provide the most accurate solution if sufficient re-simulations are performed to suppress statistical noise. The most common approach, however, is an efficient approximation via pre-computed and tabularized Monte Carlo simulations. Extensive simulations are performed on a grid of source hypotheses to generate a table describing the expected light yield in the detector. In order to obtain smooth interpolations and to generate gradients, spline functions are used to approximate the discrete tables. This procedure reduces the necessary compute time in comparison to direct re-simulations, but it requires the introduction of approximations to reduce the dimensionality of the MC tables. Further differentiation of MLE-based methods in IceCube are possible in the definition of the event hypothesis and utilized likelihood, as described in the following sub-sections.

3.4.1 Monopod: Cascade Reconstruction

Declaration

The description of Monopod is taken verbatim from Ref. [6], written by the author of this dissertation, as indicated by the quotation marks.

> “The current standard cascade reconstruction method is a maximum-likelihood estimation [called MONOPOD] that fits a point-like cascade template to the measured

> light deposition pattern in the detector [11]. The cascade template describes the expected light yield at each DOM for a given cascade hypothesis consisting of the interaction vertex and time, the deposited energy, and the direction of the primary neutrino. A binned Poisson likelihood is employed to find the cascade hypothesis that best describes the measured light deposition pattern in the detector. The cascade template is obtained from MC simulation and tabulated for various cascade-DOM-configurations. In order to reduce the dimensionality of the lookup tables, rotational and translational invariance in the x-y-plane are assumed. Second order corrections can then be applied to account for the inhomogeneities in the detector medium and photon propagation. Due to these simplifications, the standard reconstruction method maximizes an approximation to the true underlying likelihood. More advanced reconstruction methods exist within IceCube [57, 60], however, due to their computational complexity they can only be applied on single events. Additional information on [MONOPOD] is provided in Ref. [11].

”[6]

3.4.2 Taupede: Double-Cascade Reconstruction

The TAUPEDE reconstruction is an extension of the single cascade hypothesis utilized in MONOPOD. Two causally connected point-like energy depositions are assumed with the same direction, where the second energy deposition is placed on the trajectory of an hypothetical outgoing particle. This event hypothesis thus adds two free parameters to the single-cascade hypothesis consisting of the distance to the second energy deposition and its energy. TAUPEDE is primarily utilized to reconstruct double-cascade signatures expected from charged-current ν_τ interactions, in which the out-going τ decays after a short propagation distance.

3.4.3 Millipede: Segmented Track Reconstruction

Analogous to TAUPEDE, MILLIPEDE [11] extends the single cascade hypothesis to a series of energy depositions along an hypothesized track. These energy depositions may be placed in a pre-defined interval, typically 5 m to 15 m, and are all defined to point in the same direction. An additional continuous energy deposition along the track, modeling a minimum ionizing muon, may be added. The MILLIPEDE hypothesis thus adds N free parameters consisting of the energy of each of the N energy depositions.

3.4.4 SplineMPE: Track Rreconstruction

An alternative track reconstruction method to Millipede, requiring less parameters, is given by SplineMPE, which maximizes a multi-photo-electron (MPE) likelihood [61,62] given by:

$$\mathcal{L}_{\mathrm{MPE}}\left(\vec{x}|\vec{\theta}\right) = \prod_{i}^{\mathrm{1sthits}} n_i \cdot p(t_{\mathrm{res},i}|\vec{\theta}) \cdot \left(1 - P(t_{\mathrm{res},i}|\vec{\theta})\right)^{n_i - 1} \tag{3.2}$$

$$P(t_{\mathrm{res},i}|\vec{\theta}) = \int_{-\infty}^{tres} p(t|\vec{\theta})dt. \tag{3.3}$$

The probability density for the expected light arrival time $p(t|\vec{\theta})$ for a given track hypothesis $\vec{\theta}$ is obtained from spline tables, similarly to Millipede. An infinitely long track of a minimum ionizing muon is assumed for the hypothesis $\vec{\theta}$, omitting discrete energy depositions along the track as utilized in Millipede. The likelihood utilizes the first pulse measured at the i-th DOM as well as the total number of pulses n_i at that DOM. The first photon at a DOM is typically least scattered and therefore carries the most information content on its origin. Further details on the SplineMPE reconstruction are provided in Ref. [62].

4 DNN-reco: A deep learning framework for IceCube

Declaration

A large fraction of the results and figures presented in this chapter have been published in Ref. [6]. This chapter summarizes the findings of that publication and provides additional details on the generality of the framework, software architecture and additional data input formats not covered in the paper.

The DNN-reco framework is a software framework aimed to facilitate the training and usage of deep learning models in IceCube. It is intended as a modularized framework that allows the application of various deep learning architectures to IceCube data. As of writing this, DNN-reco supports any neural network (NN) architecture that utilizes a constant number of input features per DOM. This currently mainly includes convolutional neural networks (CNNs), but the modular structure allows to expand to other architectures and input data types as discussed in the following sections.

The framework is a generalization of the more specific work on CNNs for muon-neutrino reconstruction presented in Ref. [3]. An example of utilizing DNN-reco for the reconstruction of cascade events in IceCube is published in Ref. [6]. The main findings of that publication are summarized in Sections 4.4 and 4.5. In Part II, further examples are provided that illustrate the performance of DNN-reco for classification tasks within IceCube. These enable the creation of efficient and pure event selections. The DNN-reco framework has also been implemented for real-time alerts of cascade-like events, as detailed in Appendix Chapter A [Cascade Real-Time Alert Stream].

4.1 Notes on Software Framework

The DNN-reco project is a software framework for deep neural network applications in IceCube. This covers a wide range of supervised learning tasks such as regression (e.g. reconstruction of direction, energy and interaction vertex) and event classification (e.g. background rejection, flavor and topology identification). The software framework not only covers the development and training of neural network models, but also enables their application and distribution.

An often voiced concern in the context of machine learning is reproducibility. Trained models are very powerful, but can be susceptible to incorrect usage and mis-configured settings. Especially when working in a collaborative environment, where such models are shared between users, there have to be mechanisms in place to ensure their correct usage. DNN-reco aims to simplify and secure the process of model creation and distribution between users.

Reproducibility is therefore chosen as one of the guiding design concepts of DNN-reco. The software framework and dependent packages cover the complete production line from training data creation, over training of the neural networks and application of the trained models to new data. Throughout this process, used settings in the intermediate steps and especially in the training procedure are automatically saved.

Once a neural network model is trained, it can be exported. The exported models contain all necessary configuration and log files as well as software versions needed to reproduce the training procedure. In addition, all required settings are exported such that other users may apply these models with the provided `I3Module`[1]. Note, however, that the training process employs multiprocessing for the data input pipeline. It can therefore not be guaranteed that the training batches are created in the same order. This may result in small, but mostly minor differences if a model is re-trained from scratch. The DNN-reco framework ensures that an exported model is used with the correct settings and input parameters.

The code is available on GitHub[2] and Zenodo [64], together with additional documentation and examples. In order to allow for future extensions, the software framework is designed in a modularized structure. Individual modules are dynamically loaded depending on the central configuration file, which steers model training, export and application.

The driving classes of DNN-reco are:

- **SetupManager:** reads the configuration file, adds default values and provides meta information such as the current software versions.
- **NNModel:** sets up NN architecture and provides the minimizer and training routine. Configurable and dynamically loaded modules include:
  - **Model:** defines the architecture of the neural network. This can be any architecture that maps the input features to the expected number of outputs, which are defined by the labels loaded in the DataHandler class.

---

[1] IceCube data processing relies on individual modules, so called `I3Modules`, that are applied consecutively to a stream of data within the `IceTray` [63] framework.

[2] `https://github.com/icecube/dnn_reco`

  - **Event weight:** defines if and how event-weights are computed. These may be used to weight the per-event loss.
  - **Loss:** defines the loss function to use. This can be any function that maps the network's outputs or parameters to a scalar value, which is used to optimize the trainable parameters of the network.
  - **Learning rate scheduler:** optionally a more advanced learning rate schedule can be defined. This allows for flexible modifications of the learning rate as a function of the training iteration step.
  - **Evaluation:** if provided, user defined evaluation modules may be run on the validation set after every N training iterations. This can be used to compute additional performance metrics or to create evaluation plots.

- **DataTransformer:** normalizes the input data and labels. This is an essential component that facilities and speeds up the training procedure. Further details on the normalization and its importance are provided in Ref. [6].
- **DataHandler:** loads the input data, weights, labels and additional miscellaneous data. Can optionally also apply a filter or biasing strategy to the loaded events to limit or emphasize training to a certain subset of data. This filter may be based on any of the miscellaneous data, the labels or even on the current network predictions. Configurable modules include:
  - **Labels:** defines the labels to use. These can be loaded from the training files or computed on-the-fly.
  - **Misc:** defines any additional miscellaneous data that needs to be loaded in addition to the input features and labels. This typically consists of data needed for weighting or the application of the event filters.
  - **Filter:** Defines the filter to apply to loaded events.

These classes dynamically load the modules as specified in the configuration file. Each of the modules are specified by the file name and the function or class name within the specified file. As such, new user-defined modules may be added without interfering with existing code. Apart from the dynamic modules, each of the driving classes has a number of settings that are specified in the central configuration file.

As of writing this, the main focus of DNN-RECO is on CNNs with constant length, DOM-based inputs. Although currently not implemented, a future iteration of DNN-RECO may also generalize the input data pipeline of the DataHandler class via dynamically loaded modules. This would allow for more flexibility in the definition of the input features and thus support variable input sizes. See Section 4.3 [DOM-based Input Features] for further details on the input features to the NNs.

4.2 Neural Network Architecture

The DNN-reco framework currently supports any neural network architecture

$$f\colon \vec{X} \mapsto \hat{\vec{y}} \tag{4.1}$$

that maps a DOM-based feature matrix $\vec{X}$ of constant length to the desired output quantities $\hat{\vec{y}}$. This is particularly suited to convolutional neural networks (CNNs), but also graph neural networks working on fixed, DOM-based input vectors are appropriate architectures for this input format. As mentioned in the previous Section 4.1, variable sized input features may be supported in a future iteration of DNN-reco. This would allow for pulse-level input data and increased flexibility in the architecture choice. Details and further examples on the choice of input features $\vec{X}$ are provided in Section 4.3 [DOM-based Input Features].

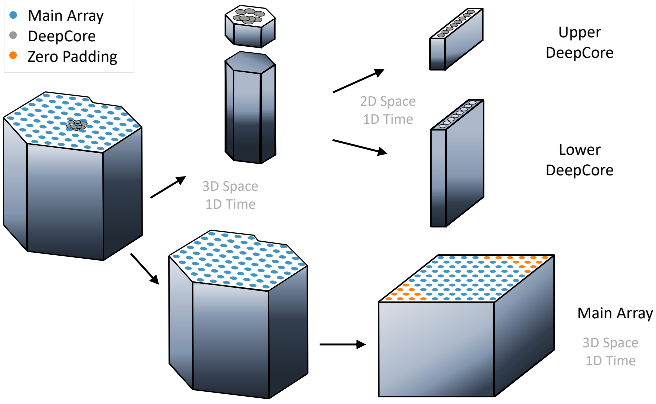


Figure 4.1: The main IceCube array and DeepCore strings are handled separately due to their differing geometry. Hexagonally shaped data of the main array can be transformed from an axial coordinate system into an orthogonal grid by padding with zeros (orange dots) and aligning the rows, which results in a 10×10 grid in the x-y-plane. Figure and caption are taken from Ref. [6].

The NN architecture is defined in a separate, dynamically loaded module (Section 4.1) and is thus fully user-defined. The current default architecture is illustrated in Figure 4.2. IceCube data is split up into three parts, the main IceCube array as well as lower and upper DeepCore, according to its geometry as shown in Figure 4.1. The resulting arrays are processed independently by three convolutional neural networks. The outputs of the CNNs are combined in a flattened layer and passed on to two small fully-connected sub-networks, which predict the quantities of interest (first sub-network) as well as the estimated uncertainty on each of these predictions (second sub-network). The convolutions over the main IceCube array utilize hexagonal convolution kernels in order to match the detector geometry. Further details on the architecture, design choices, training procedure, and loss functions are provided in Ref. [6].

The number of layers noted in Figure 4.2 are for the specific architecture that was used in Ref. [6]. Amongst other settings, the number of layers, convolution kernel sizes, activation functions and loss function are configurable and will generally depend on the task at hand. A neural network that is intended as a first data reduction step will generally have fewer layers and smaller kernel sizes in order to reduce computation time. Likewise, while a Gaussian Likelihood may be an appropriate loss function for regression tasks, event classification usually employs a loss function based on cross-entropy. The DNN-RECO-based neural networks presented in the context of this dissertation utilize a similar architecture to the one described above, while mainly modifying the network size, loss functions and input data format.

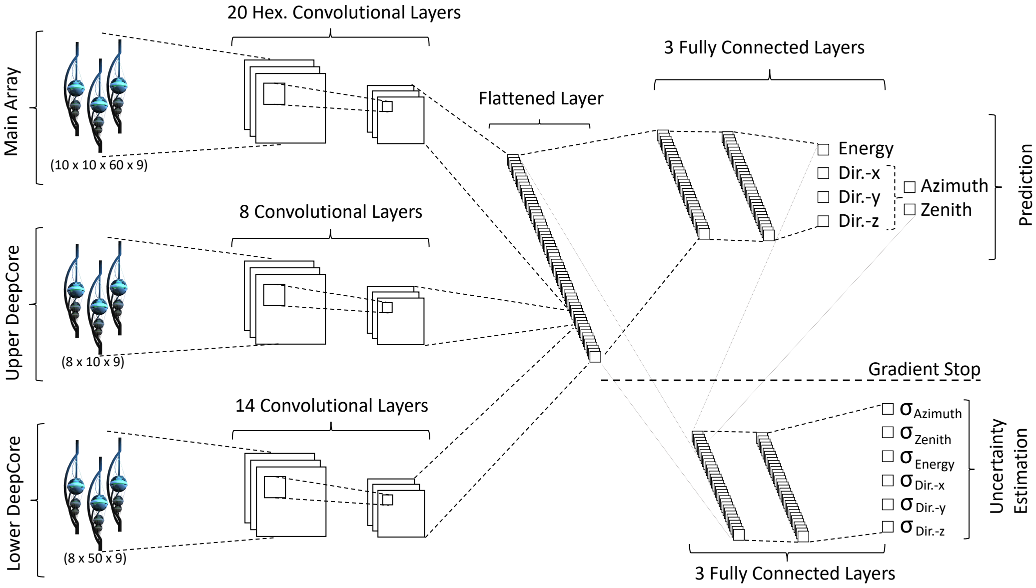


Figure 4.2: A sketch of the neural network architecture is shown. Data from the three sub-arrays are sequenced into convolutional layers. The result is flattened, combined, and passed on to two fully-connected sub-networks which perform the reconstruction and uncertainty estimation. The uncertainty-estimating sub-network also obtains the prediction output as an additional input. Figure and caption are taken from Ref. [6].

4.3 DOM-based Input Features

An IceCube event consists of a series of extracted pulses at each DOM (Section 3.1 [The IceCube detector]). The pulse series are of variable length that can fluctuate substantially between DOMs of an event as well as between events. Neural network architectures such as CNNs require a fixed input size for each input node. For the default architecture in the DNN-RECO framework (Section 4.2), the input nodes to the NN are represented by IceCube DOMs. For the application in CNNs, the variable sized pulse series at a given DOM must therefore be converted to a feature vector of fixed length. This section illustrates examples of how this can be achieved.

In general, the input features should be chosen based on the task at hand. The features need to contain all necessary information and ideally emphasize the relevant details through the choice of a suitable format. Explicit domain knowledge can be embedded in the construction of the input features. Moreover, the set of features can be constructed such that it is robust against mis-modeling in the MC simulation and against known systematic uncertainties (see also Section 4.5 [Robustness and Data/Simulation Agreement]).

The default input features, also utilized in the cascade reconstruction example from Ref. [6] (results summarized in Sections 4.4 and 4.5), employs a selection of nine input features. These nine input features are illustrated in Figure 4.3 and consist of:

1. c_{total}: the total charge at the given DOM.
2. $c_{500\mathrm{ns}}$: the charge collected within 500 ns after the first pulse.
3. $c_{100\mathrm{ns}}$: the charge collected within 100 ns after the first pulse.
4. t_{first}: the (relative) time of the first pulse.
5. t_{last}: the (relative) time of the last pulse.
6. $t_{20\%}$: the (relative) time at which 20% of the DOM charge is collected
7. $t_{50\%}$: the (relative) time at which 50% of the DOM charge is collected
8. t_{mean}: the average (relative) time of all pulses at the given DOM
9. t_{std}: the standard deviation of pulse times at the given DOM

The time variables are computed relative to a global offset, which is defined as the start of a 6000 ns long time window that maximizes the contained charge of each event. Utilization of relative timing allows for the exploitation of translational invariance of the data in time. This selection of input features is a general description of the pulse series and as such applicable to a wide variety of tasks.

In some cases, a faster selection and computation of input features is desired. To accomplish this, the previous selection may be reduced to the most important features ($c_{\mathrm{total}}, t_{\mathrm{first}}, t_{\mathrm{std}}$). The total charge at a given DOM c_{total} correlates to the energy deposited in the nearby detector medium. It is therefore crucial for any task relying on energy estimation. In combination with the time of the first pulse t_{first}, this provides valuable information on the origin of the measured photons. The first photons arriving at a given DOM have the highest chance of not having undergone scattering. A scattered photon will require a longer time to reach the PMT of a given DOM than an unscattered photon originating from the

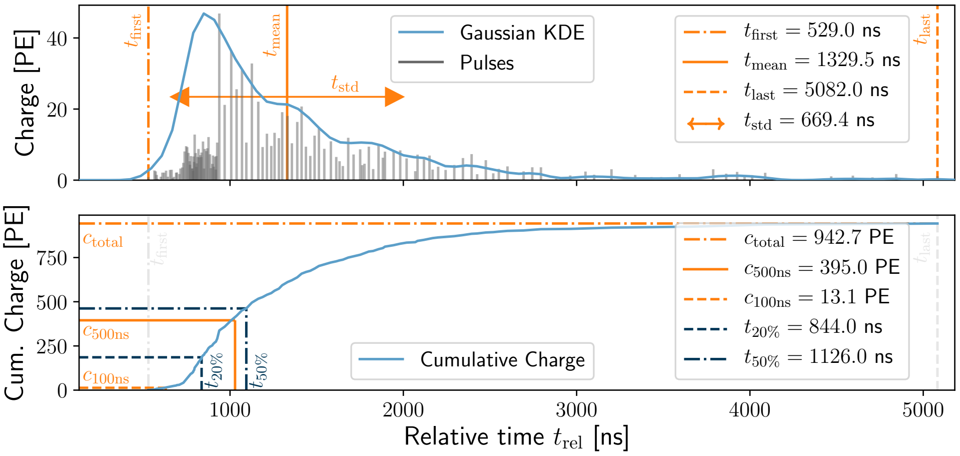


Figure 4.3: An example pulse series and corresponding input data for a single DOM is shown. The measured pulse series cannot directly be utilized by a CNN due to its varying length. The pulses are therefore reduced to nine input parameters (c_{total}, c_{500ns}, c_{100ns}, t_{first}, t_{last}, $t_{20\%}$, $t_{50\%}$, t_{mean}, t_{std}), which aim to summarize the pulse distribution. Figure and caption are taken from Ref. [6].

same source. The probability that the first photon has not undergone scattering also rises with an increasing number of detected photons, which is proportional to the total charge c_{total}. Unscattered photons give direct insight on the possible distance of a source to the given DOM. With the combination of multiple DOMs, the location of the source can then be identified. The standard deviation of the pulse times t_{std} correlates with the width of the pulse distribution and thus provides additional information on the distance and orientation of the source as well as a hint on additional nearby sources. The narrower the distribution, the closer the source is and the less scattered the photons are. This is typically the case for nearby sources pointing towards the DOMs. Very wide distributions can indicate significant scattering or photons from multiple sources. In contrast to the previous selection, the reduced selection uses a simpler definition of the global time offset that is faster to compute. In this case, the time offset is chosen as the average time over all pulses measured in an event. The reduced selection is implemented in c++ and it enables the implementation of small NNs with a per-event compute time of less than 1 ms.

While the previous feature selections are applicable to a wide range of tasks, specific applications may benefit from a dedicated set of input features. An example for the detection of incoming veto muons is illustrated in the following discussion. Event selections, focusing on astrophysical neutrinos that start in the detector, make use of the so called “self-veto” [65, 66] to further reduce the atmospheric neutrino background. High-energy atmospheric neutrinos are often accompanied by muons created in the same air shower. If detected, these accompanying muons can be used to veto the neutrino events as neutrinos of atmospheric origin.

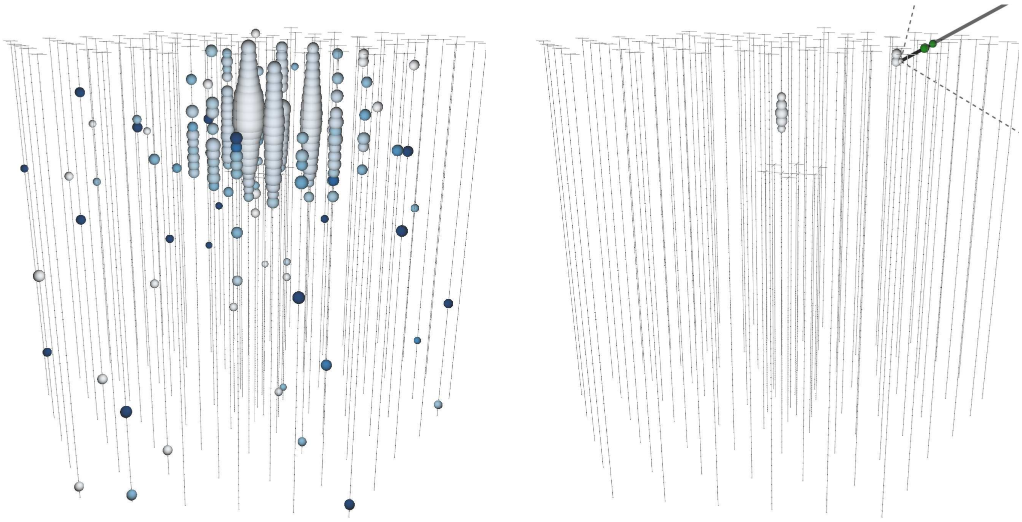


Figure 4.4: An event view of a cascade interaction in the upper detector is shown. The left panel illustrates the information a NN would obtain if provided with the total charge at a DOM (size of blob) together with the time of the first pulse (color of blob). On the right, dedicated input features are able to identify an incoming muon that is not directly visible in the data representation on the left. The dedicated features bin obtained charge in time bins relative to the assumed cascade interaction vertex. Shown is the charge collected in three bins that correspond to the time window in which one would expect to see a contribution of minimally scattered light from an accompanying muon. In addition, the true trajectory of the incoming muon is plotted in the right panel, which clearly links the charge in the two nearby DOMs to the accompanying muon.

In principle, the general feature selections, previously described, contain most of the necessary information for the detection of such muons. However, the signal of these faint muons can be hard to distinguish from the photons of the neutrino interaction and it may rely on precise timing information. A dedicated data representation can emphasize the contribution of the muons as illustrated in Figure 4.4. This greatly facilitates the training of neural networks and also improves their performance. Domain knowledge on the assumed vertex position and timing of accompanying muons is utilized to construct the data representation on the right of Figure 4.4. This domain knowledge is therefore explicitly included and does not have to be learned implicitly by the network from the training data.

Based on an assumed cascade interaction vertex, time bins are constructed at which photons of an in- or out-going muon are expected at a given DOM. For each DOM, the positive time difference Δt_d is computed that is necessary for a particle with a velocity of c to reach the d-th DOM from the assumed cascade interaction vertex. The bins at the d-th DOM are constructed based on (signed and fractional) multiples of this relative DOM-dependent time difference plus an optional absolute time delta. The bin sizes are configurable by the user. For the detection of incoming veto muons, bin edges $\vec{b}_d$ around the expected time of a muon directly passing by the d-th DOM such as:

$$\begin{aligned}\vec{b}_d = [&-\Delta t_d - 1000\,\mathrm{ns}, -\Delta t_d - 100\,\mathrm{ns},\\ &-\Delta t_d + 0\,\mathrm{ns}, -\Delta t_d + 100\,\mathrm{ns},\\ &-\Delta t_d + 300\,\mathrm{ns}, -\Delta t_d + 500\,\mathrm{ns},\\ &-\Delta t_d + 1000\,\mathrm{ns}, -\Delta t_d + 3000\,\mathrm{ns},\\ &-\Delta t_d + 5000\,\mathrm{ns}, -\Delta t_d + 10\,000\,\mathrm{ns}]\end{aligned} \tag{4.2}$$

may be useful. The pulse charges at the d-th DOM are then accumulated in the specified bins while also taking into account uncertainties on the vertex reconstruction. The overall time uncertainty

$$\sigma = \sqrt{\frac{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}{c^2} + \sigma_t^2} \tag{4.3}$$

is computed by adding the reconstruction uncertainties in quadrature. The charge in the j-th bin at the d-th DOM is then given by

$$C_{d,j} = \sum_i \left(q_i \cdot \int_{b_j}^{b_{j+1}} \mathcal{N}(x|t_i, \sigma)\,\mathrm{d}x \right) \tag{4.4}$$

where q_i and t_i are the charge and time of the i-th pulse, b_k is the k-th element of the specified bin edges $\vec{b}_d$ and $\mathcal{N}(x|t_i, \sigma)$ is a Gaussian distribution with mean t_i and standard deviation σ. Intuitively, this divides the pulse charge among the time bins based on the probability that the given pulse truly belongs to the specified time bin.

On the right of Figure 4.4, the charge from bins three to five, corresponding to the bin edges $[\Delta t_d + 0\,\mathrm{ns}, \Delta t_d + 100\,\mathrm{ns}, \Delta t_d + 300\,\mathrm{ns}, \Delta t_d + 500\,\mathrm{ns}]$, are shown. These are the bins in which one would expect to see photons from an accompanying muon. Figure 4.4 clearly shows that there is a second cluster of pulses that does not stand in causal relation to the shower of the neutrino interaction. The plotted Monte Carlo truth of the muon trajectory verifies this assertion. This data representation utilizes the knowledge, that muons must come in coincidence with the neutrino, in order to emphasize a potential signal from accompanying muons. For this, the reconstructed interaction vertex position and time is used, including their uncertainties. This data representation is one example for a specific task. It can be extended by including an assumed neutrino direction.

DNN-reco currently supports constant length, DOM-based input features. A selection of input features must therefore be performed. While the feature selections described in this section contain most of the available information, they are subject to information loss due to the necessary reduction of the pulse series. A future iteration of the framework may support variable sized, pulse-level inputs, as discussed in Section 4.1 [Notes on Software Framework]. This would allow for the inclusion of all available information, but also require architectures that can handle this input format. Graph and recurrent neural networks can generally handle variable input sizes and are therefore viable candidates.

4.4 Cascade Reconstruction Performance

Declaration

The performance plots illustrated here are published in Ref. [6] for the application of cascade reconstruction.

DNN-reco is a versatile tool, applicable to a wide range of tasks. Here, the performance for cascade event reconstruction is illustrated for a typical cascade event selection in IceCube, described in Ref. [67]. The event selection is designed to select high-energy events at TeV energies and above induced by charged-current ν_e and $\bar{\nu}_e$ as well as neutral-current interactions of all neutrino types. Applications and performance for classification tasks are detailed in Part II.

In the context of neutrino astronomy, the most relevant event properties are the interaction time, energy and direction of the incoming neutrino. While the event trigger provides sufficient timing resolution, inference of the energy and, in particular, direction require sophisticated reconstruction methods. These methods must be sensitive to subtle features in the light deposition pattern in the detector in order to accurately reconstruct the direction (Section 3.4 [Traditional Reconstruction Methods in IceCube]). In the following discussion, the performance of the default CNN architecture, defined in the DNN-reco framework, is compared against the current standard cascade reconstruction method in IceCube[3]. Details on the standard, maximum-likelihood-based reconstruction method (Monopod) are provided in Section 3.4.1 [Monopod: Cascade Reconstruction].

---

[3] Note that reconstruction methods in IceCube are subject to ongoing work and improvements. Performance comparisons shown in this dissertation are therefore not final.

Angular Resolution

From a neutrino astronomy perspective, the energy range above ∼ 10 TeV is of particular interest as this region is where the astrophysical neutrino contribution begins to dominate the cascade energy spectrum. This is also the energy region in which the CNN begins to outperform the angular resolution of the standard method. As shown in Figure 4.5, the CNN is able to improve the angular resolution by almost a factor of two at highest energies. This is a significant improvement that directly translates to a large boost in sensitivity towards searches for neutrino sources as discussed in Part III (Chapters 12 [Analysis Performance] and 13 [MESC 7yr Analysis with DNN-reco]).

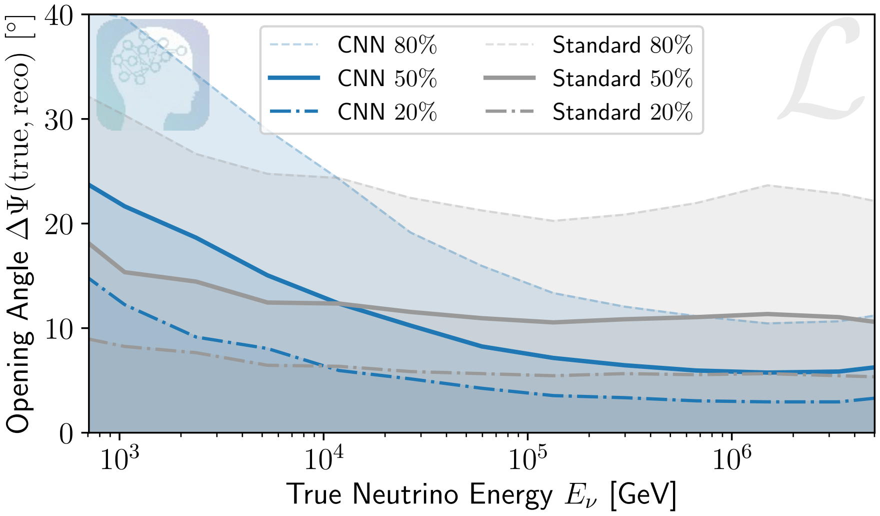


Figure 4.5: The angular resolution is shown as a function of neutrino energy for the standard reconstruction method and the newly developed CNN-based method. The shaded area and lines denote the 20%, 50%, and 80% quantiles. At higher energies, the resolution can be improved by up to 50 %. Systematic uncertainties are not included. Figure and caption are taken from Ref. [6].

Energy Resolution

The collected charge in an event can be used as a simple proxy for the deposited energy in the detector volume. More advanced algorithms also account for the orientation and location of the emitter as well as differing local ice properties that affect the amount of detected photons. This allows for the accurate reconstruction of energy that was deposited in electro-magnetic shower components within the detector. An inherent and unavoidable uncertainty in the energy reconstruction of the incoming neutrino stems from energy carried away by neutral particles or outgoing neutrinos as it is the case in neutral current interactions, for example. IceCube does not have the resolution to distinguish hadronic from electro-magnetic showers, apart from identifying outgoing, long-ranged leptons. Therefore, the performance of cascade energy reconstruction methods in IceCube is typically quantified on the deposited, electro-magnetic equivalent energy instead of the incoming neutrino energy. As shown in Figure 4.6, the standard and CNN-based reconstruction method provide comparable energy resolution. There is a tail of mis-reconstructed events at higher energies for the standard reconstruction method that are better reconstructed with the CNN.

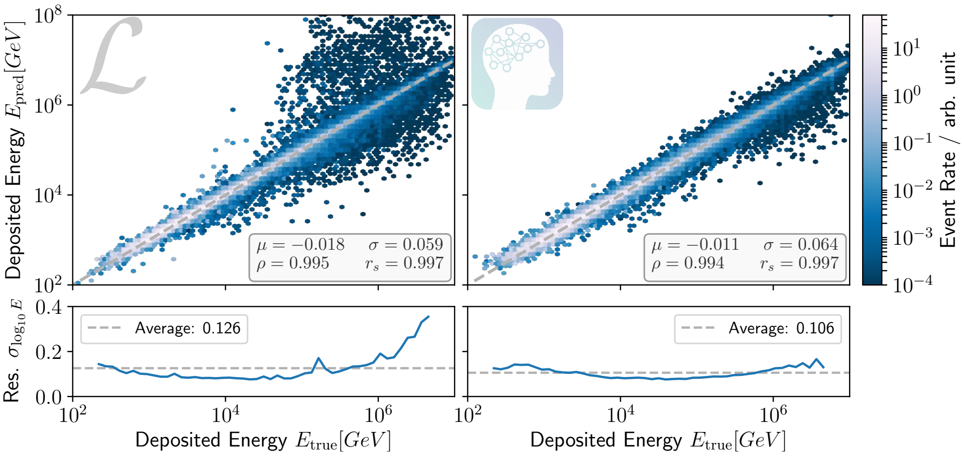


Figure 4.6: Correlation plots between the true and reconstructed deposited energy are shown for the standard reconstruction (left) and CNN (right). Mean μ and standard deviation σ of the residuals $y_{\text{pred}} - y_{\text{true}}$ are determined in $\log_{10}$-space as well as the Pearson ρ and Spearman r_s correlation coefficients. The energy resolution in the bottom panel is calculated according to Ref. [11]. Figure and caption are taken from Ref. [6].

Uncertainty Estimation

As previously mentioned in Section 4.2 [Neural Network Architecture] and illustrated in Figure 4.2, the default CNN architecture also outputs an estimate for the uncertainty on the reconstructed quantities. Hence, the NN not only outputs a point estimate, but an approximation of the posterior distribution under the assumption that it is Gaussian distributed. This assumption often holds to a reasonable degree for the residuals $\Delta y = y_{\text{pred}} - y_{\text{true}}$ between the true and predicted values, y_{true} and y_{pred}, respectively. The uncertainty estimate σ_{pred} is trained together with the point

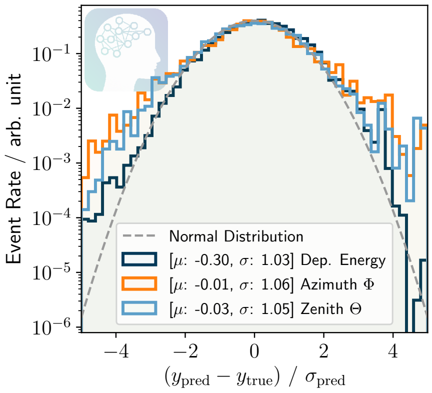


Figure 4.7: The pull distributions are shown for the reconstructed quantities in comparison to the normal distribution which indicates a good uncertainty estimator. These match well, apart from slight deviations in the tails of the distribution. Figure and caption are taken from Ref. [6].

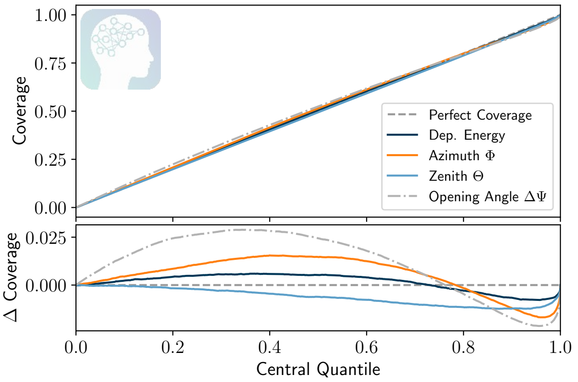


Figure 4.8: The coverage, i.e. the number of events that fall in a certain quantile, is shown as a function of the computed central quantile based on the uncertainty estimate. Perfect coverage is obtained on the dashed diagonal. The bottom panel shows the difference in coverage of each reconstructed quantity compared to the perfect one-to-one relation. Figure and caption are taken from Ref. [6].

estimate y_{pred} via an employed Gaussian likelihood

$$\mathcal{L} = \ln\left(\sigma_{\text{pred}}\right) + 0.5 \cdot \left(\frac{y_{\text{true}} - y_{\text{pred}}}{\sigma_{\text{pred}}}\right)^2 \tag{4.5}$$

as the loss function (constant terms omitted). This loss function is applied individually to each reconstructed quantity while neglecting potential correlations. These can, however, be included by utilizing a multivariate likelihood. Further details are provided in Ref. [6].

The assumption of the Gaussian distributed residuals can be tested by plotting the pull distribution, which is a histogram of the residuals normalized by the estimated uncertainty: $\Delta y_{\text{pred}}/\sigma_{\text{pred}}$. In case the residuals are truly distributed as Gaussians with the estimated width σ_{pred}, the pull distribution should resemble a normal distribution. The pull distributions for the three reconstructed quantities (azimuth, zenith, energy) are illustrated in Figure 4.7. These are well described by a normal distribution apart from deviations in the tails.

Alternatively, the uncertainty estimate of the neural network can be verified by computing the coverage. Based on the assumption of a Gaussian distribution with the per-event width of σ_{pred} as estimated by the neural network, the number of events can be calculated that should lie within a certain quantile around the prediction. This number can then be compared to how many events actually fall inside. The coverage is computed and plotted for varying quantiles in Figure 4.8. For all estimated quantities, the computed coverage follows the diagonal, which indicates an accurate uncertainty measure.

In addition to the previously mentioned event properties (azimuth, zenith, energy), the coverage on the overall opening angle $\Delta\Psi$ between true and reconstructed direction is computed. A simplified approach assumes a "circularized" error, i.e. a circular contour on the sphere, which is computed as

$$\sigma_{\Delta\Psi} = \sqrt{\frac{\sigma_{\Theta}^2 + (\sigma_{\Phi} \cdot \sin\Theta_{\text{CNN}})^2}{2}} \tag{4.6}$$

where Θ_{CNN} is the reconstructed zenith angle by the CNN and σ_{Θ} and σ_{Φ} the estimated uncertainty on the zenith and azimuth reconstruction, respectively. The coverage for the opening angle in Figure 4.8 is computed based on a Rayleigh distribution. A more suitable approach to define the per-event uncertainties on the sphere is discussed in supplement Chapter A [Cascade Real-Time Alert Stream].

Per-Event Runtime

In addition to an improved accuracy and reliable uncertainty estimate, the CNN is able to reduce the necessary computation time by almost three orders of magnitude in comparison to the standard reconstruction method. The CNN also reduces the spread in per-event runtime, indicated by the shaded region in Figure 4.9. The main contributor to the variability of the runtime is the computation of the DOM input parameters, which depends on the number of pulses in an event. A reduced feature selection and network size can further decrease the runtime to about 1 ms per event (Section 4.3 [DOM-based Input Features]). If the CNN is applied on a single CPU instead of a GPU, the per-event runtime increases by roughly a factor of 60. However, this factor will depend on the network size and the time necessary for data preprocessing versus the pass through the network.

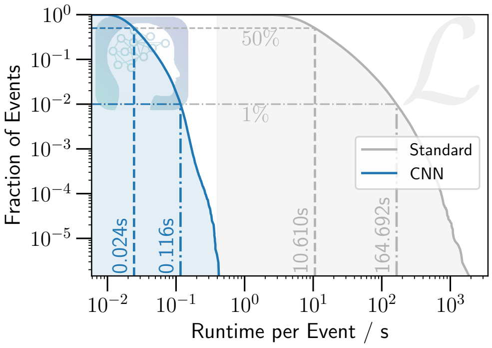


Figure 4.9: Per-event runtimes are shown as a survival function of the fraction of events exceeding a specified runtime. In contrast to the standard reconstruction, the CNN-based method is able to run on multiple cores as well as on a GPU. Runtimes shown for the CNN are for application on a GPU (NVIDIA GTX 980). The CNN outperforms the standard method by 2-3 orders of magnitude. Figure and caption are taken from Ref. [6].

4.5 Robustness and Data/Simulation Agreement

Declaration

Section 4.5 and its sub-sections are adopted verbatim from Ref. [6], as indicated by the quotation marks, with minor modifications applied. Ref. [6] is authored by the author of this dissertation, while the IceCube collaboration provided feedback and edits in the internal review process of the paper publication procedure.

> “In the previous section, the performance of the CNN-based method was shown to offer comparable, and for some tasks, even improved performance relative to the standard cascade reconstruction method in IceCube. However, the performance plots shown are for the MC baseline dataset. Limits in our knowledge of the detector and the underlying physics result in systematic uncertainties in the MC simulation. Some of the sources of systematics are known, such as the scattering and absorption coefficients of the glacial ice. The effect of these systematic uncertainties can be studied in dedicated simulations for which these parameters are varied. More challenging are unknown systematics that cannot be simulated and explicitly tested. A key quality parameter of a reconstruction method is therefore its data/MC agreement and the robustness of the method towards possible uncertainties in the MC simulation. In the context of this dissertation, robustness of a reconstruction method is defined as its insensitivity towards changes in the input data. Note that a robust method does not necessarily have to be an accurate one. In contrast, often a trade-off between reconstruction accuracy and robustness exists. In the following sections, the data/MC agreement of the presented reconstruction method will be investigated and its robustness tested against potential mis-modeling in the simulation.” [6]

4.5.1 Agreement for Baseline Simulation

> “A correct MC simulation will not contain differences to experimental data, although in practice, the simplified model in the simulation will not perfectly describe the data. While a simulation cannot accurately describe all low level physics, it is important that the resulting high-level variables, from which physics results are derived, are well described. Reconstruction methods produce such high-level variables and must therefore be robust towards low-level disagreements in the simulation. The data/MC agreement of a given reconstruction method can be

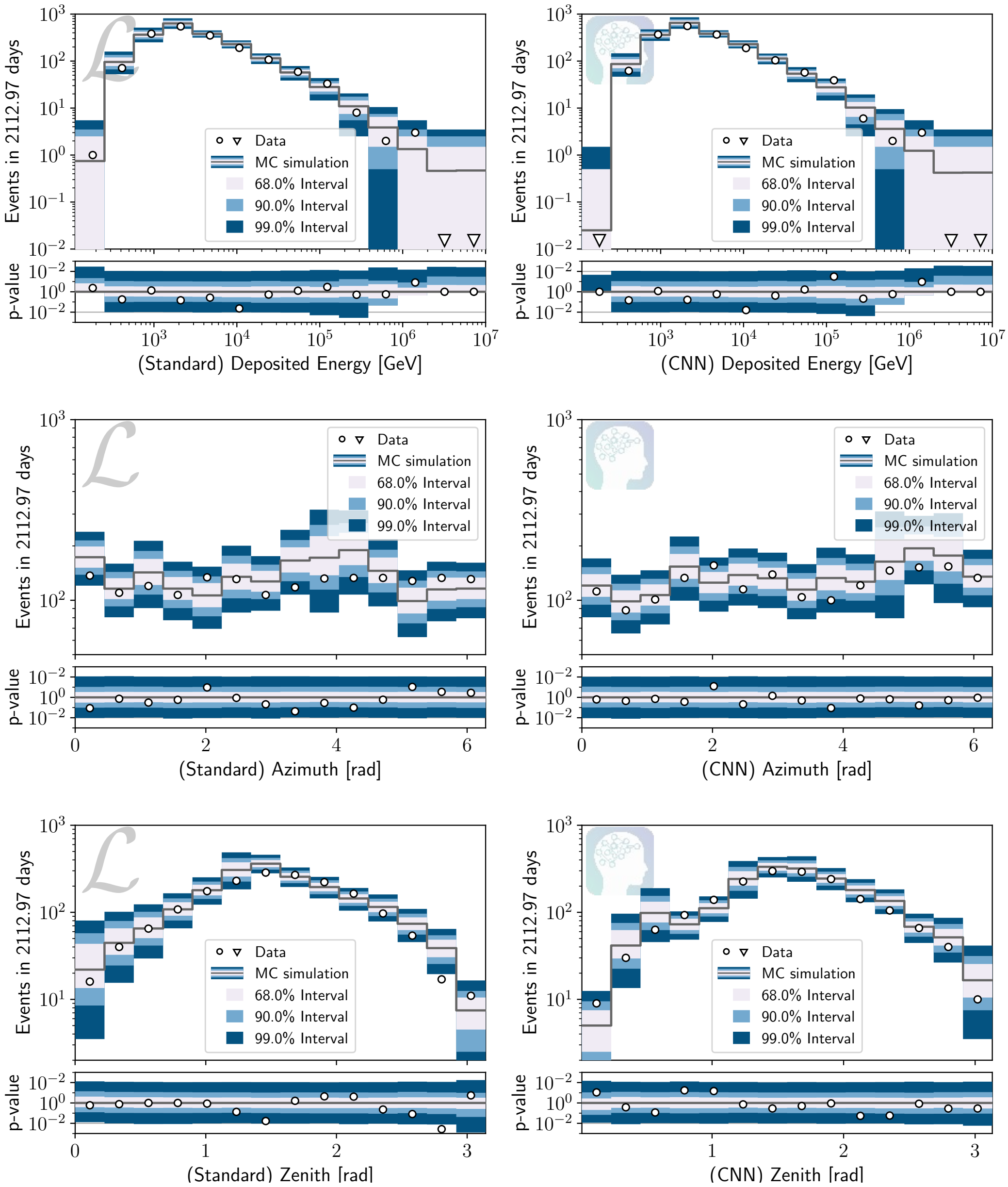


Figure 4.10: The simulated and measured distributions for the reconstructed quantities are shown. The left side shows the standard reconstruction and the right side shows the CNN-based method. Each plot is divided into two panels showing the number of events (top panel) and the significance of the fluctuations (bottom panel) in each bin. The significance calculation is based on the assumption of Poisson-distributed data and does not include systematic uncertainties. Figure and caption are taken from Ref. [6].

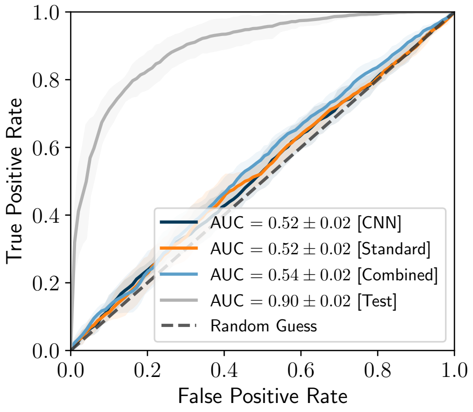


Figure 4.11: A random forest (RF) is trained in a 5-fold cross-validation to distinguish data events from simulated ones on the basis of different sets of features. The receiver-operating curves are shown for each set of features. The area under this curve (AUC) is a metric for the classifier performance. A value of 1.0 denotes a perfect classification while randomly assigning classes will result in an AUC of 0.5. Events are not distinguishable on the basis of the reconstructed values for either of the reconstruction methods. Figure and caption are taken from Ref. [6].

quantified by comparing distributions of analysis level parameters.

The agreement between simulated and experimental data for the distributions of the reconstructed zenith, azimuth, and deposited energy is investigated in Figure 4.10. On the left side, the distributions are compared for the standard reconstruction method and on the right for the CNN. The distributions and their agreement is comparable between the two reconstruction methods. Some smaller fluctuations exist, but overall, the simulation and the reconstructed quantities well describe the experimental data.

The comparisons of the one-dimensional distributions in Figure 4.10 are blind to some classes of multivariate disagreements. In principle, higher-dimensional distributions can be compared, but the data will become too sparse once the dimensionality reaches a certain level. An alternative approach is to employ a multivariate classifier to distinguish data from simulation [68, 69]. A random forest (RF) [70] classifier from the scikit-learn [71] package is trained in a 5-fold cross-validation. To quantify the found mis-match, the area under curve (AUC) metric of the receiver-operating curve is used. Ideally, the classifier should not be able to tell the difference between simulated and experimental data, i.e. the AUC should be around 0.5. If it is capable of separating the classes (AUC > 0.5), the feature importance vector of the random forest can be used to understand the origin of the disagreement. The feature importances hint at variables that are important for the separation task and might therefore have a higher mis-match.

To test the applicability of this method, a RF is trained on the output variables of both reconstruction methods in addition to artificially created variables that contain a certain mis-match. Variables $\mathrm{Var}_{x,c}$ and $\mathrm{Var}_{y,c}$ are added which are drawn from a bivariate normal distribution with a covariance of $+c$ for data events and $-c$ for

Table 4.1: Feature importances obtained from the RF, trained to distinguish simulated from experimental data, are shown for each of the four sets of features. The test case in which artificial features $\mathrm{Var}_{x/y,c}$ with varying degrees of mis-matches are added is shown on the left. An AUC score of 0.90 ± 0.02 is achieved. The importances of the CNN-only ($\mathrm{AUC} = 0.52 \pm 0.02$), standard reconstruction-only ($\mathrm{AUC} = 0.52 \pm 0.02$), and combined feature set including outputs from both reconstructions ($\mathrm{AUC} = 0.54 \pm 0.02$) are shown on the right from top to bottom, respectively. Table and caption are taken from Ref. [6].

Test

Feature Name	Importance
$\mathrm{Var}_{x,c=0.9}$	0.09 ± 0.07
$\mathrm{Var}_{y,c=0.9}$	0.08 ± 0.07
$\mathrm{Var}_{x,c=0.5}$	0.06 ± 0.05
$\mathrm{Var}_{y,c=0.5}$	0.05 ± 0.04
$\mathrm{Var}_{y,c=0.0}$	0.03 ± 0.02
CNN - Zenith	0.03 ± 0.02
Standard - Zenith	0.03 ± 0.02
CNN - Energy Unc.	0.03 ± 0.02
Standard - Energy	0.02 ± 0.02
CNN - Azimuth Unc.	0.02 ± 0.02
CNN - Energy	0.02 ± 0.02
$\mathrm{Var}_{x,c=0.0}$	0.02 ± 0.02
CNN - Azimuth	0.02 ± 0.02
CNN - Zenith Unc.	0.02 ± 0.02
Standard - Azimuth	0.02 ± 0.02

Real

Feature Name	Importance
CNN - Zenith	0.18 ± 0.17
CNN - Energy	0.17 ± 0.16
CNN - Azimuth	0.17 ± 0.16
Standard - Zenith	0.17 ± 0.17
Standard - Azimuth	0.17 ± 0.16
Standard - Energy	0.16 ± 0.16
Standard - Zenith	0.07 ± 0.06
CNN - Energy Unc.	0.06 ± 0.06
CNN - Zenith	0.06 ± 0.06
CNN - Azimuth	0.06 ± 0.05
Standard - Azimuth	0.06 ± 0.05
CNN - Azimuth Unc.	0.06 ± 0.05
CNN - Zenith Unc.	0.06 ± 0.05
Standard - Energy	0.05 ± 0.05
CNN - Energy	0.05 ± 0.05

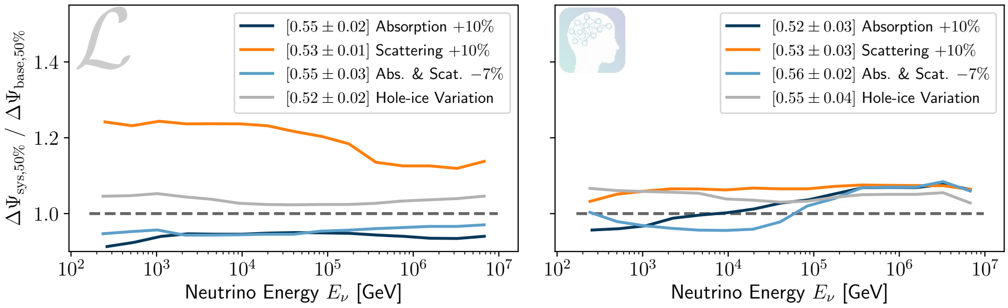
simulated events. The one-dimensional distributions of these variables for data and simulated events are indistinguishable by construction. Only a multivariate approach is able to find the disagreement. The resulting receiver-operating curve is shown in Figure 4.11 labeled as ”Test”. As expected, the RF is able to distinguish the events based on the added variables which contain mis-matches in correlated variables. The resulting feature importances also correctly order the artificial features according to their mis-match as shown on the left of Tab. 4.1.

Output features of each reconstruction method are tested independently as well as in a combined set. The resulting receiver-operating curves and feature importances are shown in Figure 4.11 and on the right of Tab. 4.1. Apart from the method test case, the RF is not capable of significantly distinguishing simulation from experi-

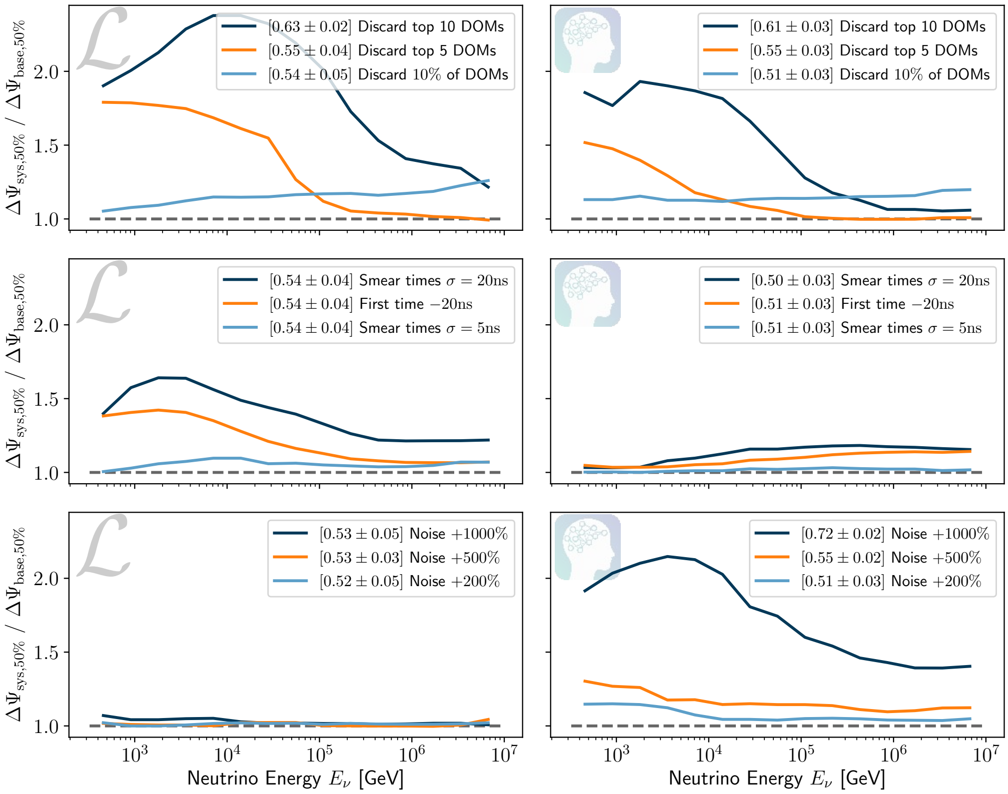
> mental data, boosting confidence in the baseline simulation and the reconstruction methods. Both reconstruction methods provide reasonable data/MC agreement, which indicates a certain level of robustness towards potential mis-modeling in the simulation.
>
> While the agreement on high-level variables is most important, the data/MC test can also be employed on low-level inputs. In this case, the RF isn’t applicable due to the high-dimensionality of the input data and a NN, preferably the same architecture, may be used instead. ”[6]

4.5.2 Ice Systematics

> “The accurate description of the detector medium is a challenging task for simulations in IceCube [31]. Ice properties affect how photons scatter and are absorbed in the detector medium and are therefore a major contributor to systematic uncertainties in the simulation. In order to find astrophysical neutrino sources, the impact of these uncertainties on the angular reconstruction must be well understood. Parameters studied in this section include the scattering and absorption coefficients of the ice as well as the parameterization of the hole-ice. These are the main sources of systematic uncertainty in IceCube. The hole-ice is the refrozen column of ice in which the DOM strings are embedded. After deployment of the detector strings, the melted water in the drill holes refreezes. The refrozen hole-ice has different optical properties with increased absorption and scattering coefficients due to enclosed bubbles and dust particles [28]. The accurate description of the hole-ice in the simulation is one of the key challenges. Four systematic datasets are obtained by varying the values of the baseline simulation.
>
> In Figure 4.12a, the robustness of the angular resolution towards these systematic variations is investigated. The ratio of angular resolution for the baseline versus the systematic variation is plotted as a function of neutrino energy. A ratio of 1.0 indicates that the systematic variation has no effect on the angular resolution. Ratios above 1.0 mean that the systematic variation has a negative effect on the directional reconstruction increasing the opening angle in comparison to the baseline. In contrast, systematic variations with ratios less than 1.0 improve the angular resolution with respect to the baseline. The spread in ratios across the different variations is a measure for the robustness of the reconstruction method. Since the exact values of these parameters are unknown, an ideal reconstruction method should be insensitive to these variations, i.e. produce the same results and therefore have a ratio close to 1.0.



(a) Variation of Ice Properties



(b) Pulse Modifications

Figure 4.12: The median angular resolution is compared between the baseline simulation $\Delta\Psi_{\text{base, }50\%}$ and the systematic variation $\Delta\Psi_{\text{sys, }50\%}$ for the standard reconstruction method (left) and the CNN-based method (right). A ratio close to 1.0 indicates a robust reconstruction method that is insensitive to the varied systematic parameter. For each systematic variation, the data/MC test via the RF (Section 4.5.1) is performed. The resulting AUC for each variation is shown in brackets in the legend. Figure and caption are taken from Ref. [6].

> Increasing the scattering coefficients will result in more diffuse photons arriving at the photo-multipliers and hence complicate the reconstruction. As a result, the angular resolution decreases for both reconstruction methods in comparison to the baseline simulation as shown in Figure 4.12a. In contrast, reducing the scattering coefficients or increasing the absorption results in an improved angular resolution for the standard reconstruction method. Overall, the spread across different ice systematics is larger for the standard method than for the CNN-based reconstruction. With a relative resolution change up to about 5% for the CNN in comparison to 24% for the standard reconstruction method, the CNN seems to be less sensitive to the varied ice-parameters. ”[6]

4.5.3 Pulse Modifications

> “In addition to the varied ice properties in Section 4.5.2, further unknown sources of systematics may exist. While the effect of these unknown systematics cannot be directly studied, the simulation can be altered in various ways to investigate how a reconstruction method reacts to certain modifications. In this study, several modifications are performed to the measured pulses to understand the reconstruction's dependence on single DOMs, its sensitivity to timing uncertainties, and the effect of an increased noise rate.
>
> The reconstruction's dependence on single DOMs is investigated by discarding these prior to the reconstruction. Entries in the input data tensor to the NN, which correspond to discarded DOMs, are filled with zeros. Three different variations are tested: discarding the top 5 and top 10 DOMs with the most collected charge and randomly discarding 10% of all DOMs. The top panel of Figure 4.12b shows the results of this test. The more light a DOM captures, the more information it carries for the reconstruction task. Thus, high-charge DOMs (DOMs with the most collected charge in an event) are crucial to the reconstruction. However, if the reconstruction relies solely, or to a big extent, on a single high-charge DOM, it is more susceptible to mis-modelling in the simulation. It is challenging to accurately describe local ice properties around a DOM, whereas averages over many DOMs are much more robust. As shown in the top panel in Figure 4.12b, both reconstruction methods have a similar energy-dependent behaviour. Discarding the highest-charge DOMs negatively impacts the angular resolution at lower energies, while it becomes less important at higher energies. This is due to the fact that saturated DOMs and overly bright DOMs are automatically excluded from the reconstruction. This applies to both reconstruction methods which are therefore, by construction, fairly robust to this systematic at higher energies. For lower energetic events, the number

of hit and saturated DOMs reduces and hence the exclusion of the n highest-charge DOMs becomes more and more relevant as the fraction of removed hit DOMs increases. Randomly discarding DOMs (light blue curve in top panel) results in a slightly worse angular resolution over the complete energy range with a minimal energy-dependence. While both reconstruction methods are similarly affected by the exclusion of DOMs, the standard method is more sensitive to single high-charge DOMs, which is driven by the underlying Poisson likelihood.

Unknown systematic uncertainties in the simulation might also affect the arrival times of photons at the photomultipliers. To investigate this, pulse times are smeared by a Gaussian with a width of 5 ns and 20 ns. In comparison, IceCube's timing resolution is on the order of a few ns [22,26]. Moreover, the time of the first pulse at a DOM is particularly important for directional reconstruction, since it is likely to come from a photon that has scattered the least, i.e. carries a lot of information on its origin. By shifting it by 20 ns, it can be tested how much the reconstruction relies on this time. The results are plotted in the middle panel of Figure 4.12b. Smearing pulse times by 5 ns, which is larger than IceCube's resolution, barely has an effect on the angular resolution. Interestingly, the standard method seems to rely more on timing information at lower energies than the CNN-based method. At energies around a few TeV and lower, the variations to the pulse times have almost no effect on the CNN angular reconstruction. This indicates that the CNN-based method is not using the timing information to the full extent possible. Timing information is more important for the reconstruction of low-energetic cascades as opposed to high-energetic cascades for which the charge asymmetry is a good measure of the cascade direction. This might explain why the standard reconstruction method achieves significantly better resolution at lower energies.

Neural networks can be manipulated to produce wrong results by adding subtle noise to images through evolutionary algorithms or gradient ascent [72,73]. This vulnerability is of great concern in security crucial applications such as autonomous driving. While explicit attacks are not a concern for the application in IceCube, potential mis-modeling in the noise simulation could impose a vulnerability. Instead of explicitly attempting to manipulate the CNN, the effect of an increased overall noise rate in the detector is studied. For this study, the noise rate is increased by 200 %, 500 %, and 1000 %. Note that the values studied here are purely hypothetical and much larger than actual uncertainties on the noise rate, which are on the order of only a few percent. The noise rate in IceCube is well monitored and stable [26]. As previously mentioned, the standard method internally relies on a Poisson likelihood which is driven by high-charge DOMs. A uniformly increased noise rate therefore barely has any effect. The standard reconstruction method is robust towards the increased noise rate. In contrast, the CNN is sensitive to this modification. In agreement with the previous findings, it seems that the CNN-based method is

> less focused on single DOMs, but more so on the overall distribution. Increasing the noise rate to the hypothetical case of an additional +1000% therefore has an increased impact worsening the angular resolution by a factor up to 2.3. Realistic variations of the noise rate on the order of a few percent do not have an impact on the angular resolution.
>
> The variations tested here cannot by any means cover all potential systematics in the simulation. Neural networks can interpolate well on the training data, but should be used with caution when extrapolating. It is hard to predict how potential mis-modeling in the simulation will affect the reconstruction method. While this might seem problematic, there are multiple ways to ensure the correctness of approaches based on deep learning. The neural network can be trained to be robust towards certain variations by including these in the training process, similar to the systematic datasets for the ice properties. The difficulty lies in compiling an exhaustive list of reasonable variations. Apart from this, the data/MC test from Section 4.5.1 is an additional safety check. The data/MC test is performed for every systematic variation of the simulation data and the resulting AUC scores are shown in brackets in the legends of Figure 4.12. Whenever the systematic variation gets large enough to significantly affect the angular resolution, the AUC also increases. For example, increasing the noise rate by 1000 % results in an AUC of 0.72 ± 0.02 . Hence, the RF classifier is clearly able to distinguish between data and simulated events. This increases the confidence that unaccounted systematics will be detected by the data/MC test, if they have significant impact on the reconstruction values. ”[6]

4.6 Conclusions

In the previous sections, the performance of a CNN within the DNN-RECO framework is illustrated for cascade event reconstruction. The CNN provides a robust approach, improving upon standard methods, with reliable per-event uncertainty estimates. While this showcases one specific example, the framework allows for a rich variety of applications. It thus enables applications and physics analyses that were previously infeasible. This includes the creation of novel and improved event selections, as described in Part II, as well as the generation of fast and robust seeds to more advanced reconstruction methods.

CNNs enable the exploitation of translational invariance — an approximate symmetry in IceCube data. More suitable architectures, such as graph neural networks, exist for IceCube's detector geometry [14, 16, 19]. However, further symmetries and domain knowledge are available, which these and other standard neural network architectures are unable to directly exploit. The inferior performance of the CNN at lower energies in comparison to the MLE-based approach, shown in Figure 4.5, suggests that available information is not optimally utilized by the CNN. Robustness tests in Section 4.5.3 [Pulse Modifications] indicate that this could be due to lacking sensitivity to time-based variables. A more adequate approach, that is capable of fully utilizing available information, is thus introduced in the next chapter in order to alleviate the discussed deficiencies.

5 Event-Generator: Combining Maximum-Likelihood and Deep Learning

Declaration

Individual parts of the content discussed in this section have been presented in Refs. [4, 5], written by the author of this dissertation.

In the previous Chapter 4, a deep learning based reconstruction method, DNN-reco, is introduced. This novel reconstruction method is able to expand IceCube's toolkit with a general, all-purpose method applicable to a broad range of regression and classification tasks. Although quite successful, DNN-reco has some significant drawbacks and limitations as detailed in Section 4.6 [Conclusions].

One of the major limitations of DNN-reco is its inability to include prior knowledge and symmetries in a unified and generalized manner. This is a drawback that applies to most standard deep learning architectures. These architectures and methods are predominantly designed to function in a generalized setting. Prior information and symmetries are implicitly learned from the training data. While this works well in many cases as also illustrated by the performance of DNN-reco, dedicated methods that utilize domain knowledge may outperform their generalized counterpart.

The recognition of patterns and symmetries is crucial for meaningful extrapolation to unseen data. In a standard neural network architecture, these patterns and symmetries have to be learned implicitly from the training data. Detailed studies are required to investigate if the trained models simply memorize the training data, or if they are capable of extracting the underlying patterns and symmetries. It can be argued that a standard dense network only has the capability to utilize the continuity of the input phase space to interpolate, but that it cannot actually extract any higher level patterns that are necessary for accurate extrapolation beyond the training data. This is a more pessimistic view that contradicts the commonly communicated abilities of neural networks to implicitly learn and exploit patterns.

For reliable extrapolation, a model must be able to resort to underlying patterns, laws and symmetries (Section 2.1 [The Importance of Domain Knowledge]). Network architectures such as convolutional neural networks utilize the assumption of translational invariance in the input data. Due to this, they are capable of applying learned patterns from one part of the phase space to another. This is one of the reasons for the breakthrough in image recognition [47]. The translational equivariance, which made this breakthrough possible, is explicitly built into the architecture, but it is not learned. Explicit inclusion of a priori known symmetries and patterns into the network architecture may be the key to success. The inclusion of such domain knowledge facilitates the training process as this knowledge does not have to be learned implicitly. In addition, the model may utilize these built-in symmetries to reliably extrapolate beyond the training data.

In contrast to DNN-reco, the maximum-likelihood-based reconstruction methods, described in Section 3.4 [Traditional Reconstruction Methods in IceCube], aim to utilize all available domain knowledge. Prior knowledge and symmetries are directly built into the the likelihood and PDF prescription. Due to this construction, they are more sensitive than the implicit models such as DNN-reco. However, in practice it is often infeasible to compute the PDF and the formulation of the correct likelihood may also be intractable. Typical MLE-based methods in IceCube therefore utilize simplifications as discussed in Section 3.4.

To combat the deficiencies of DNN-reco and the traditional MLE methods in IceCube, a novel reconstruction method, Event-Generator, is introduced in this chapter. Event-Generator is a reconstruction suite that combines the strenghts of maximum-likelihood estimation with those of deep learning.

5.1 Event Hypothesis

In a maximum-likelihood-based event reconstruction (Section 3.3), an event hypothesis, a set of parameters for an assumed parametric model, must be defined. In IceCube, the event hypothesis may be defined as a linear superposition of individual energy depositions in the detector, as illustrated in Figure 5.1. To a good approximation, photons arriving at DOMs are independent of each other as long as the DOM is not saturated and its response is in the linear regime. The process of pulse extraction from the measured cumulative waveform may add a slight dependence on other photon hits, but this can usually be neglected.

Any IceCube event may therefore be fully described by a superposition of energy depositions. In the following, these energy depositions will be referred to as a collection of sources S_i. A source S_i is defined by a set of hypothesis parameters $\vec{\xi}_i$ and a function

$$G_i(\vec{\xi}_i) = \left\{\vec{\lambda}_i, \vec{P}_i(t)\right\} \tag{5.1}$$

that translates the hypothesis $\vec{\xi}_i$ to an expected signal at each DOM. The function G is approximated by a neural network, as discussed in Section 5.4 [Neural Network Architecture]. It is therefore also referred to as a generative model or Event-Generator in the context of this dissertation. The signal at the d-th DOM from the i-th source is fully defined by the expected charge $\lambda_{(d,i)} = \left(\vec{\lambda}_i\right)_d$ and the arrival

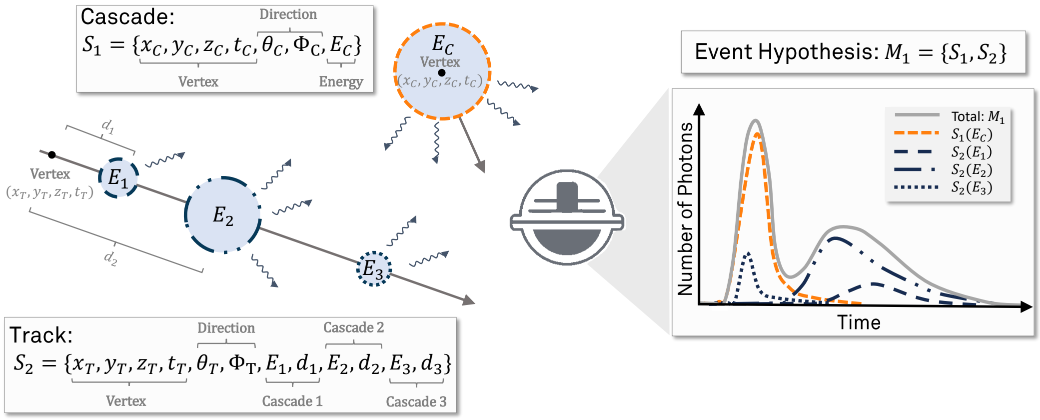


Figure 5.1: An example event hypothesis (M_1) is constructed of one cascade (S_1), consisting of 7 parameters, and a track (S_2), parameterized as three causally connected energy depositions along the particle trajectory. Each of these energy depositions is defined by the deposited energy (E_i) and the distance (d_i) to a vertex on the trajectory, resulting in a total of 12 parameters for the chosen track parameterization. On the right, a sketch is shown, demonstrating the superposition of the photons collected in a particular DOM from each of these energy depositions.

time PDF $P_{(d,i)}(t) = \left(\vec{P}_i(t)\right)_d$. The expected signal at the d-th DOM from a collection of N sources is therefore given by:

$$\lambda_d = \sum_{i=1}^{N} \lambda_{(d,i)}, \tag{5.2a}$$

$$P_d(t) = \frac{\sum_{i=1}^{N} \lambda_d^i \cdot P_d^i(t)}{\sum_{i=1}^{N} \lambda_d^i}, \tag{5.2b}$$

where the index i runs over the individual sources. The combined arrival time PDF $P_d(t)$ of the event hypothesis can be interpreted as a charge-weighted mixture density over the the individual time PDFs of each source.

In IceCube, MLE-based reconstructions typically use two fundamental sources, which are defined as track segments and cascades. A cascade is commonly implemented as a point-like energy deposition in the detector, while a track segment defines a continuous energy deposition along a straight line of a specified length. Event hypotheses are built as a superposition of these fundamental sources. A constructed event hypothesis may then be applied in a maximum-likelihood setting as discussed in Section 3.3.

As detailed in Section 3.4 [Traditional Reconstruction Methods in IceCube], the most accurate way to compute the expected light yield of a given source is via

MC simulation. However, this is often infeasible due to computational complexity. Most maximum-likelihood-based reconstruction methods in IceCube therefore resort to table-based or spline-based MC simulations. These allow to pre-compute the expected light yield on a grid of source hypotheses to which a high-dimensional spline may then be fit. This procedure reduces the necessary compute time for reconstruction, but comes at a cost. The dimensionality of the MC tables must be reduced by introducing simplifications. Further details are outlined in Section 3.4.

The Event-Generator follows a similar strategy as the spline-tables in the sense that a model is fit to MC simulations to describe the light yield as a function of the source hypothesis. In contrast to the spline tables, this is performed via deep neural networks. As a result, the model may be more complex and additional simplifications are not necessary. For example, the Event-Generator does not require to model cascades as point-like and symmetric in detector azimuth. The shower extension of the cascade as well as ice anisotropy (Section 3.1 [The IceCube detector]) can be directly modeled by the underlying neural network.

Within the Event-Generator framework, it is possible to define arbitrary sources S_i. The individual sources may also be grouped together as multi-sources:

$$M_j = \{S_1, ..., S_k\}. \tag{5.3}$$

A multi-source may also consist of a collection of multi-sources or of a combination of sources and multi-sources:

$$M_l = \{S_1, ..., S_k, M_1, ..., M_j\}, \tag{5.4}$$

which enables a recursive event hypothesis definition. An event hypothesis may consist of an arbitrary number of sources and (possibly nested) multi-sources. Mathematically, the multi-sources M_j are equivalent to an individual source and therefore the expected light-yield over a collection of multi-sources is computed recursively and analogously to Eq. (5.2).

The focus in this dissertation lies on the reconstruction of cascade-like neutrino events. These events are well described by a single or multi-cascade source. A cascade source may be parameterized by seven parameters:

$$\vec{\xi}_{\texttt{cascade}} = (x, y, z, \theta, \varphi, E, t) \tag{5.5}$$

with the interaction vertex position (x, y, z), the direction of the cascade (θ, φ), the deposited electromagnetic equivalent energy E and the time of the interaction t.

5.1.1 Nuisance Parameters and Systematic Uncertainties

Apart from the physics parameters of interest, the light yield of a given source may also depend on systematic uncertainties in the simulation. A major source of systematic uncertainty in IceCube is the accurate description of the glacial ice that is being used as the detector medium. Scattering and absorption coefficients of the detector medium have a direct influence on the expected light yield measurable at a DOM. Additional systematic uncertainties include the quantum efficiency of the DOMs (referred to as DOM efficiency), the parameterization of the refrozen hole-ice column surrounding the detector strings and the ice anisotropy (Section 3.1).

These systematic uncertainties have an effect on the expected light yield and should therefore be considered in the modeling of the source. The Event-Generator framework allows to include the systematic parameters in the cascade hypothesis as nuisance parameters

$$\vec{\xi}_{\mathrm{cascade,sys}} = (x, y, z, \theta, \varphi, E, t, \vec{\Xi}), \tag{5.6}$$

where $\vec{\Xi}$ are the above mentioned systematic parameters.

An alternative approach is to marginalize over the systematic uncertainties. This may be achieved by sampling training data according to the assumed priors on the systematic parameters via the SnowStorm method [74]. During training of the source model, the neural network is confronted with events simulated from different realizations of systematic parameters. As a result, the neural network learns to model the marginalized PDFs. Note that in this approach the likelihood prescription must be modified to employ a negative binomial instead of a Poisson distribution due to the over-dispersion (Section 5.3 [Likelihood Prescription]).

5.2 Time PDF Parameterization

Declaration

This section expands on the motivation for the choice of time parameterization laid out in Ref. [5], written by the author of this dissertation.

The difficulty in modeling a source S_i lies in the accurate description of the pulse arrival time PDF $\vec{P}_i(t)$ at each of the DOMs. Typical PDFs $\vec{P}_i(t)$ for single energy depositions at varying distances and angles are shown in Figure 5.5. IceCube has a timing resolution of a few nano seconds, while photons from a neutrino interaction can arrive many thousands of nano seconds after the time of interaction. The pulse arrival time PDFs must therefore be described over a large time window, but retain fine enough resolution.

In a binned approach, this would result in hundreds to thousands of necessary time bins. In addition, the combination of individual sources to multi-sources is only possible if the bin edges align. Apart from these complications, time invariance can only be utilized in multiples of the bin width, which does not allow for smooth gradients in time that are necessary in the event reconstruction step.

It is therefore desirable to utilize a continuous description of the time PDF. To accomplish this, a parameterization must be found that is normalized over its domain to ensure a proper PDF. The parameterization must be flexible enough to accurately describe the pulse arrival time PDF, while ideally keeping the number of free parameters low.

A viable choice for such a parameterization is a mixture model of asymmetric Gaussians (AGs) [75]. The pulse arrival time PDF $P_d(t)$ at the d-th DOM may be parameterized by a mixture model

$$P_d(t) = \sum_j^K w_j \cdot \mathrm{AG}(t \,|\, \mu_{(d,j)}, \sigma_{(d,j)}, r_{(d,j)}) \tag{5.7}$$

of K asymmetric Gaussians:

$$\mathrm{AG}(x \,|\, \mu, \sigma, r) = N \cdot \begin{cases} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), & x \leq \mu \\ \exp\left(-\frac{(x-\mu)^2}{2(\sigma r)^2}\right), & \text{otherwise} \end{cases} \tag{5.8}$$

$$N = \frac{2}{\sqrt{2\pi} \cdot \sigma(r+1)} \tag{5.9}$$

where the asymmetry is parameterized by r. The asymmetry of the mixture model components allows for a proper description of the PDF, while maintaining a low number of free parameters as illustrated in Figures 5.5 and 5.6. [5]

The continuous description of the time PDF $\vec{P}_i(t)$ enables the exploitation of time invariance of the neutrino interaction by shifting the argument from t to $t' = t + \mathit{\Delta} t$. It also facilitates the composition of multiple sources and therefore constitutes an essential element of the recursive structure of the EVENT-GENERATOR framework.

5.3 Likelihood Prescription

Declaration

This section summarizes the prescription laid out in Ref. [5], which is authored by the author of this dissertation.

In order to train a model for a given event hypothesis $\vec{\xi}$, an objective needs to be defined that will be optimized in the training procedure of the neural network.

While the Event-Generator framework allows for arbitrary objectives, it is usually advisable to use the same likelihood prescription that is also used during reconstruction in the MLE setup (see Section 3.3 for details on MLE).

In IceCube, the likelihood is typically chosen as a binned or unbinned extended Poisson likelihood [76] over the pulses measured at the DOMs. As described in Section 5.2, the source models in the Event-Generator framework utilize a mixture model to obtain a continuous description of the time PDFs $P_d(t \,|\, \vec{\xi})$. Therefore, the per-event likelihood is chosen as an unbinned Poisson likelihood over the measured pulses of the form:

$$\mathcal{L}_{\text{event}}\left(\vec{x} = \{\vec{c}, \vec{t}\} \,|\, \vec{\xi}\right) = \prod_{d}^{D} \mathsf{Poisson}\left(\sum_{i}^{N_d} c_{d,i} \,|\, \lambda_d\left(\vec{\xi}\right)\right) \cdot \prod_{i}^{N_d} P_d(t_{d,i} \,|\, \vec{\xi})^{c_{d,i}}, \quad (5.10)$$

with $D = 5160$ as the total number of DOMs, N_d as the number of pulses at the d-th DOM, and $c_{d,i}$ and $t_{d,i}$ are as charge and time of the i-th pulse at the d-th DOM. [5]

The expected charge at a DOM for a given event hypothesis and set of systematic parameters $\vec{\Xi}$ is Poisson distributed. However, some care must be taken if the model does not explicitly state the systematic parameters as nuissance parameters in the event hypothesis $\vec{\xi}$, but instead marginalizes over these. In this case, the Poisson expectation value itself is drawn from a random distribution via the sampled set of systematic parameters. As a result, higher fluctuations of the measured charge at a DOM are obtained than expected from a Poisson distribution. A Gamma-Poisson mixture distribution may be used to model this over-dispersion as shown in Figure 5.2. The Gamma-Poisson mixture distribution is a real-valued pendant to the negative binomial distribution. The parameterization defined in Ref. [77] is utilized:

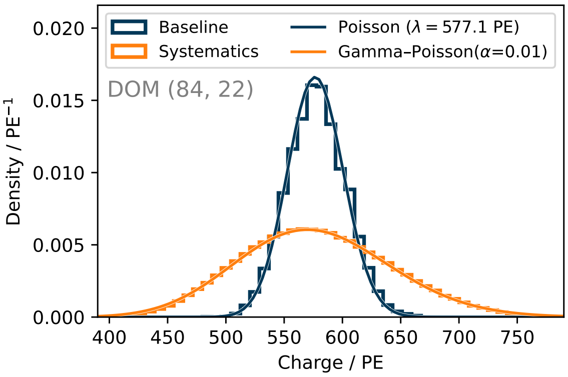


Figure 5.2: The distribution of total observed charge at DOM $(84, 22)$ is shown for re-simulations of one injected cascade event. In contrast to the baseline simulation with fixed values for systematic parameters (blue), marginalizing over different realizations of systematic parameters (orange) results in a charge distribution with increased variance. This over-dispersion may be modeled by a Gamma-Poisson mixture distribution.

$$\mathsf{GP}\left(z \,|\, \lambda, \alpha\right) = \frac{\Gamma(z + \frac{1}{\alpha})}{\Gamma(z+1)\Gamma(\frac{1}{\alpha})} \left(\frac{1}{1+\alpha\lambda}\right)^{\frac{1}{\alpha}} \left(\frac{\alpha\lambda}{1+\alpha\lambda}\right)^{z} \quad (5.11)$$

which utilizes a shape parameter α that leads to over-dispersion when $\alpha > 0$. As a result, a source S_i must also output the shape parameter $\vec{\alpha}$ for each DOM, which modifies Eq. (5.1) to:

$$G_i(\vec{\xi}_i) = \left\{\vec{\lambda}_i, \vec{\alpha}_i, \vec{P}_i(t)\right\} \tag{5.12}$$

and the likelihood from Eq. (5.10) is modified to:

$$\mathcal{L}_{\text{event}}\left(\vec{x} = \{\vec{c}, \vec{t}\} \,|\, \vec{\xi}\right) = \prod_d^D \mathsf{GP}\left(\sum_i^{N_d} c_{d,i} \,|\, \lambda_d(\vec{\xi}), \alpha_d(\vec{\xi})\right) \cdot \prod_i^{N_d} P_d(t_{d,i} \,|\, \vec{\xi})^{c_{d,i}} \tag{5.13}$$

to account for these changes. [5]

5.3.1 DOM and Time Window Exclusions

In experimental data runs, data of individual DOMs or certain time windows of a DOM may be excluded for various reasons such as hardware malfunction, calibration issues or PMT saturation. These exclusions must be accounted for by the reconstruction algorithm.

Exclusion of an entire DOM is straightforward: the expected light yield λ_d for the given DOM is set to zero and all measured pulses from that DOM are removed. The partial exclusion of one or multiple time windows at a specific DOM requires additional steps. Measured pulses within excluded time windows are removed from the event. Excluding these time windows modifies the pulse arrival time PDF and the expected light yield at the given DOM. The excluded fraction ϵ_d of the cumulative time distribution at the d-th DOM is computed via

$$\epsilon_d(\vec{\xi}) = \sum_{i=0}^{T} \int_{t_i^{\text{lower}}}^{t_i^{\text{upper}}} P_d(t \,|\, \vec{\xi}) \, \mathrm{d}t \tag{5.14}$$

where i iterates over T non-overlapping exclusion time windows, which are each defined by the tuple $\{t^{\text{lower}}, t^{\text{upper}}\}$ that define the beginning and end of the excluded time window. The pulse arrival time PDF is then renormalized:

$$P'_d(t \,|\, \vec{\xi}) = \frac{P_d(t \,|\, \vec{\xi})}{1 - \epsilon_d(\vec{\xi})} \tag{5.15}$$

and the expected light yield is scaled accordingly:

$$\lambda'_d(\vec{\xi}) = \lambda_d(\vec{\xi}) \cdot \left(1 - \epsilon_d(\vec{\xi})\right). \tag{5.16}$$

5.3.2 Priors and Constraints

Within the Event-Generator framework, a likelihood prescription may be composed of multiple terms. The per-event likelihood $\mathcal{L}_{\text{event}}$ from Eq. (5.10) or Eq. (5.13) constitutes an example of such a term. These individual likelihood terms are then combined to the overall likelihood via

$$\mathcal{L} = \prod_i \mathcal{L}_i \tag{5.17a}$$

$$\Longleftrightarrow \ln \mathcal{L} = \sum_i \ln \mathcal{L}_i . \tag{5.17b}$$

This implementation enables the addition of priors and constraints. Hypothesis parameter constraints are of particular importance for the event reconstruction step. They may be used to ensure that the optimizer stays within bounds of the phase space that is covered by the training dataset.

One of the commonly used priors is a uniform one within a specified boundary. Internally, this uniform bounded prior is implemented as a “pseudo” prior, where the boundary edges are smoothed out with a sharply increasing function defined as

$$-\ln \tilde{p}_{\text{uniform}}(x \,|\, x_{\text{lower}}, x_{\text{upper}}) = \begin{cases} h\left(\frac{x_{\text{lower}} - x}{x_{\text{upper}} - x_{\text{lower}}}\right), & x < x_{\text{lower}} \\ h\left(\frac{x - x_{\text{upper}}}{x_{\text{upper}} - x_{\text{lower}}}\right), & x > x_{\text{upper}} \\ 0, & \text{else} \end{cases} \tag{5.18}$$

with

$$h(z) = \exp\left(\zeta \cdot (z + 1)\right) - \exp\left(\zeta\right) \tag{5.19}$$

where the parameter ζ (typically set to a value of 10) defines how sharply the function increases outside of the specified boundaries.

Due to the smooth boundary of the prior, gradients are available, even if the minimizer exceeds the boundaries. This is referred to as a “pseudo” prior, because the normalization is not correct. For the reconstruction task, only likelihood differences and gradients are required. Therefore, a constant offset to the PDF within the bounds does not alter the result. Setting the negative log-likelihood within the bounds to zero allows for computational optimization. The contribution of this “pseudo” prior only needs to be calculated if the values are outside of the boundaries.

5.4 Neural Network Architecture

Declaration

This section follows the description from Ref. [5], written by the author of this dissertation. Parts of this section are adopted verbatim as indicated by quotation marks.

A challenge for maximum-likelihood based methods in IceCube lies in the high dimensionality of the task. In the standard methods, described in Section 3.4 [Traditional Reconstruction Methods in IceCube], tabulated MC simulations are utilized to describe the pulse arrival time PDF and expected charge at each DOM. However, this approach does not scale well to a large number of dimensions while maintaining a fine granularity, resulting in the necessity of simplifications via approximate symmetries to reduce dimensionality. On the other hand, neural networks are highly efficient at mapping high-dimensional relations in data. To exploit this property and to overcome shortcomings of traditional methods in IceCube, the generative model G from Eq. (5.1) or (5.12):

$$G(\vec{\xi}) = \begin{cases} \left\{\vec{\lambda}, \vec{\alpha}, \vec{P}(t)\right\}, & \text{marginalized} \\ \left\{\vec{\lambda}, \vec{P}(t)\right\}, & \text{else} \end{cases} \tag{5.20}$$

is modeled by a generative neural network. At the same time, it is necessary to make use of domain knowledge such as translational and rotational invariance of the neutrino interaction, physics laws and detector characteristics. These are incorporated into the neural network architecture to facilitate the training procedure and to enable reliable and meaningful inter- and extrapolation beyond the phase space covered in the training data (see Section 2.1 for details on the importance of utilizing domain knowledge in neural networks).

Relative displacement vectors and angles between the energy deposition and DOMs are computed and thus utilized as input to the neural network (NN), as shown in Figure 5.3. This allows for the utilization of the exact detector geometry as well as rotational and translational invariance of the neutrino interaction before detector effects are accounted for. The depicted per-DOM input parameters are suitable for sources that describe individual energy depositions or neutrino interactions within the detector. Other source definitions, such as through-going tracks, might require a modified selection of input variables in order to efficiently utilize available symmetries.

The neural network is required to compute the expected charge $\vec{\lambda}$ and pulse arrival time PDF $\vec{P}(t)$ at each DOM. As mentioned in Section 5.2, the time PDF is parameterized by a mixture model of asymmetric Gaussians. Hence, the architecture of the

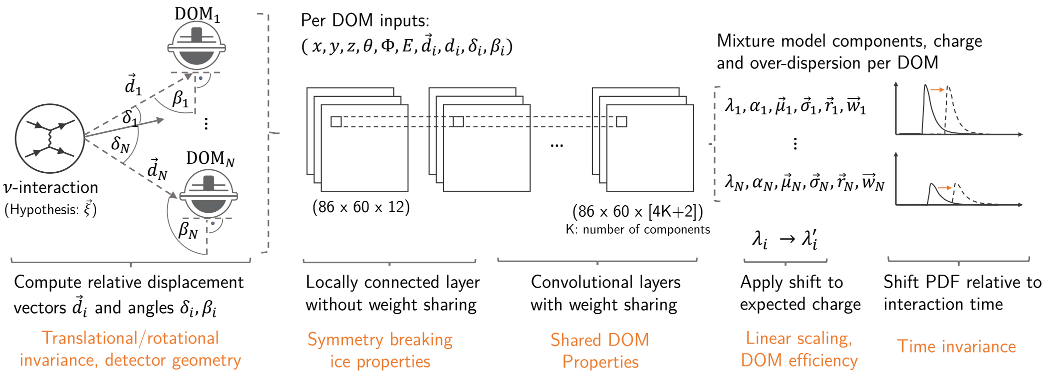


Figure 5.3: A sketch of the generative neural network architecture for cascade hypotheses is shown. Due to the construction in "forward direction", similar to the Monte Carlo simulation, domain knowledge (examples indicated in orange) can be explicitly included into the architecture. Figure and caption taken from Ref. [5].

generator NN is designed to output the mixture model parameters $\{\vec{\mu}_d, \vec{\sigma}_d, \vec{r}_d, \vec{w}_d\}$ and expected charge λ_d at each DOM. If the model marginalizes over systematic parameters, it must also output the over-dispersion parameter α_d. A series of convolutional layers with 1×1-kernels is utilized to exploit the independence and similar response of each individual DOM. Due to existing and efficient implementations in tensorflow [78], this is internally achieved via two dimensional convolutions[1]. Initial layers are locally connected, but these do not apply weight sharing across DOMs.

> This allows the NN to model the position and direction dependent, symmetry breaking optical properties of the ice. Subsequent layers utilize standard convolution operations with weight sharing. Therefore, after the initial locally connected layer, every DOM is treated equally. Additional domain knowledge, such as the linear scaling of collected charge to cascade energy or the differing quantum efficiency q_d of the DOMs, is directly incorporated into the architecture, by scaling the expected charge output:
>
> $$\lambda'_d = \lambda_d \cdot \frac{E}{10\,\mathrm{TeV}} \cdot q_d. \tag{5.21}$$
>
> In general, the architecture may be configured analogously to the MC simulation, while computationally expensive parts are replaced by a neural network approximation. Any domain knowledge that goes into the MC simulation, may therefore also be utilized in the generator NN. This is possible, in contrast to standard deep learning architectures, because the generator NN is defined in the same "forward"

---

[1] Note that the convolutional layers are used for their weight sharing property, while neighbor relations between input nodes are discarded due to the 1×1-kernels. Alternative implementations without the use of convolutional layers are possible.

> direction as the simulation. Standard deep learning applications, such as the CNN-based method, attempt to infer the posterior distribution of the quantities of interest from measured data, i.e. in “backward” direction compared to the simulation.

”[5]

The architecture described above is utilized for the cascade source models trained in the context of this dissertation. In contrast to the CNN-based approach (Chapter 4 [DNN-reco: A deep learning framework for IceCube]), it allows to exploit the exact detector geometry as well as translational and rotational invariance of the neutrino interaction rather than the approximate invariance in measured data, which is convolved with detector effects. In principle, more domain knowledge exists about the optical properties of the ice that could be included. The approach via the initial locally connected layers allow the NN to model these dependencies in a data-driven approach during the training procedure. However, it is possible to include more direct knowledge on the scattering and absorption of the ice. In a future iteration, this could be done by computing an effective absorption and scattering length between the neutrino interaction and receiver DOM. Further details on the trained models are provided in Appendix B.2.

5.5 Model Evaluation and Interpretability

In contrast to standard deep learning architectures, such as CNNs, the hybrid reconstruction approach applied in the Event-Generator framework is more interpretable. Individual components of the model can be visualized and cross-checked. Moreover, due to the modular structure, the contribution of individual DOMs to the overall likelihood can be computed. Selected DOMs may also be excluded from the reconstruction to investigate their impact on the reconstruction outcome. This results in an advanced toolset that can be used to interpret and evaluate results of the model. Standard deep learning architectures do not provide this functionality and are therefore sometimes negatively referred to as “black boxes”. In the following sections, the trained Event-Generator models (supplement Section B.2) are visualized, cross-checks against MC simulation are conducted, and sanity checks are performed. While these visualizations and checks are by no means exhaustive, they illustrate the potential of the available tools and demonstrate how results of the model can be studied in greater detail.

5.5.1 Comparison to Monte Carlo Simulation

In order to assess the validity of the trained models, the estimated charge deposition pattern in the detector can be compared to the true one obtained from MC simulations. Since the models are trained to estimate the average charge and time PDF at each DOM, a large amount of simulations must be performed such that statistical fluctuations are reduced. Due to the computational cost, the checks in this section are performed exemplary for one specific cascade event. A cascade of 10 TeV is injected below the dust layer in the central part of the detector. The same cascade event is re-simulated half a million times to obtain the average light yield at each DOM. An overview of the cascade simulation and investigated DOMs is provided in Figure 5.4.

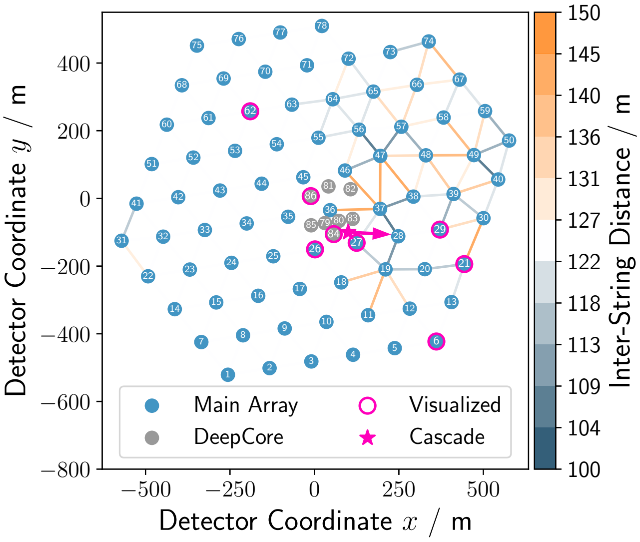


Figure 5.4: A 10 TeV cascade is injected at the coordinates (100 m, −100 m, −300 m) with a zenith and azimuth angle of 150° and 177°, respectively. The event is simulated half a million times and resulting pulses are combined to obtain the average light yield at each DOM. The cascade position and direction in the x-y-plane is plotted in addition to the surrounding IceCube strings. Strings containing DOMs that are investigated further in this section are highlighted in pink.

Pulse arrival time PDFs of 24 DOMs on 8 different strings are compared in Figure 5.5. The selected DOMs cover a wide range of distances and observation angles. In addition to the distribution of simulated pulses, the estimated distributions for each of the three cascade models (described in supplement Section B.2) are shown. Overall, the estimated densities match the true underlying ones well[2]. There are some minor deviations as visible in the peaks of the densities of DOMs $(84, 22)$ and $(84, 31)$, or slightly mis-aligned shapes of DOMs $(26, 35)$, $(6, 59)$ and $(62, 58)$. Interestingly, if present, these deviations tend to exist in each of the three independent generator models. This may indicate that common properties of the generator architecture or the shared training data may have led to the observed biases.

Deviations in low-charge DOMs such as $(26, 35)$, $(6, 59)$ and $(62, 58)$ are expected due to fluctuations present in the training data. Note that the event-generator models are trained on individual events, while the comparison in Figure 5.5 utilizes half a million of simulations of the same event. In fact, it's remarkable that the

---

[2] The utilized models are trained to reconstruct the maximum point of the cascade shower, which is computed by applying an energy-dependent shift along the shower axis. In order to compare against the interaction vertex, the reconstructed vertex is shifted back by this amount.

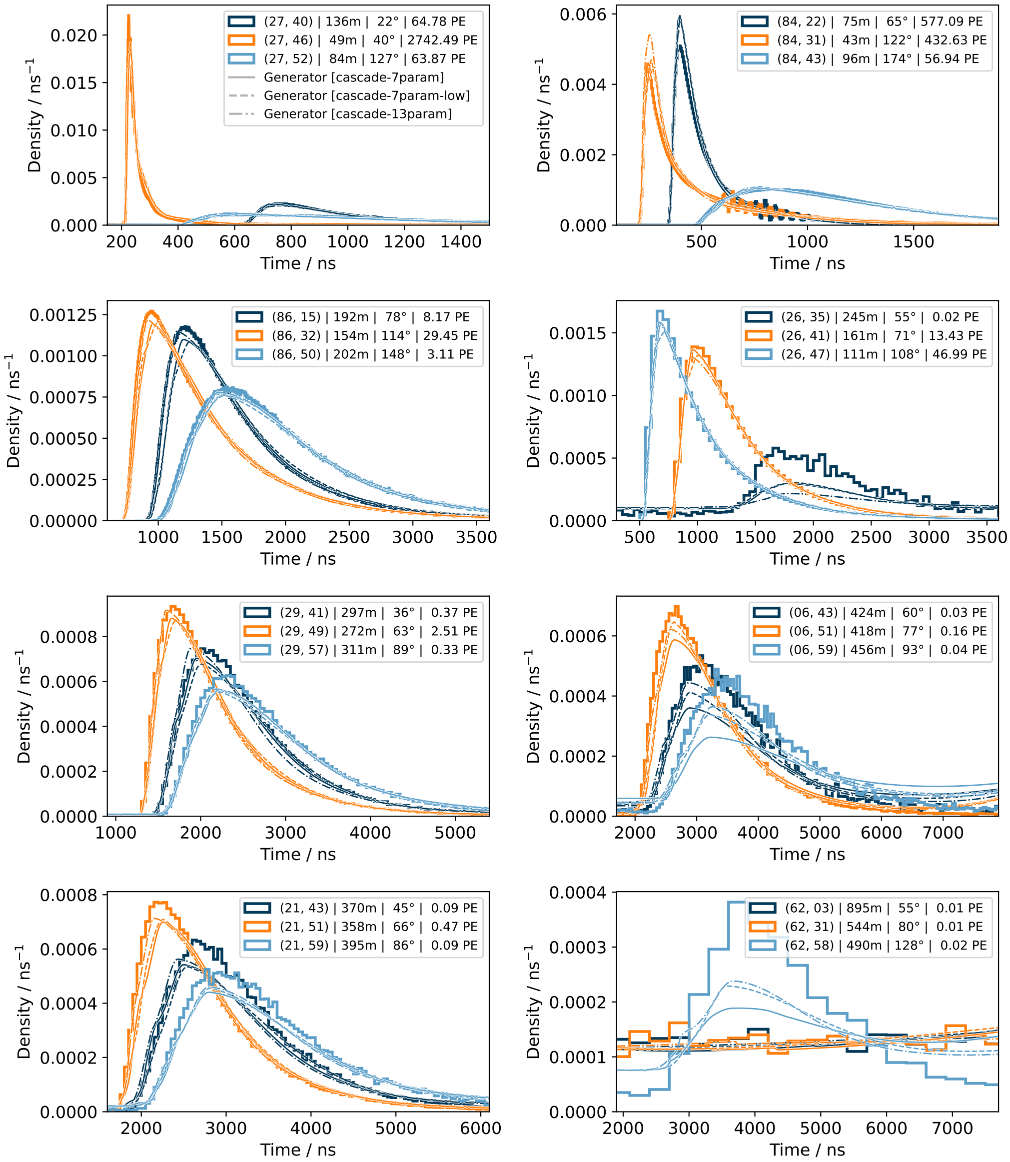


Figure 5.5: The pulse arrival time PDFs as estimated by the generator models are compared against the true underlying PDFs obtained from MC simulation. Each panel contains the distribution of simulated pulses for three different DOMs at a given string. The labels denote the string and DOM id, the distance to the cascade vertex, the opening angle with respect to the cascade direction and the average per-event charge. Apart from minor deviations, all three cascade models are capable of modeling the pulse distribution.

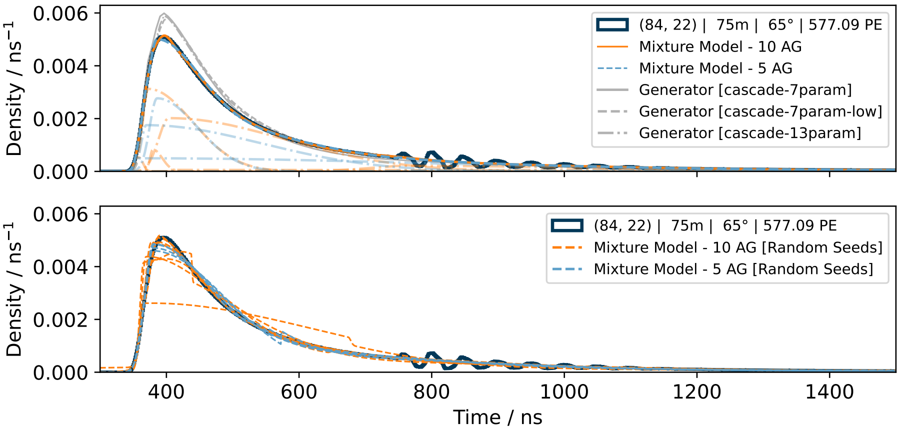


Figure 5.6: Both of the mixture models containing 5 and 10 AGs are able to model the underlying densities as demonstrated in the top panel for the time PDF of DOM $(84, 22)$. The set of chosen components (illustrated by dashed-dotted lines in upper panel) that lead to a good description of the PDF are not unique. The solution is degenerate and the success of the fitting algorithm depends on the chosen seeds as demonstrated in the lower panel. Ten random seeds for the positions of the components are drawn for each of the mixture models. The components are then fit via maximum-likelihood estimation and the obtained solutions are plotted as dashed lines.

models are able to recover the shape of these distributions despite only observing a photon at the given DOM in about every hundredth training event. During the training process, the Event-Generator model is forced to infer the average distributions by comparing across “similiar” events in the training data. Due to limited amounts of available training data, the phase space may not be covered as densely as required, resulting in a lack of “similiar” events and thus leading to reconstruction biases in certain areas. As demonstrated in Section 2.1 [The Importance of Domain Knowledge], symmetries and domain knowledge can be utilized to mitigate this effect by allowing to compare events at distant points in phase space along the given symmetries. The explicit exploitation of such domain knowledge is a contributing factor to the model’s ability to model even these faint DOMs.

While the mis-alignment for low-charged DOMs may be explained by statistical fluctuations, the deviations in the distribution peaks as visible in DOMs $(84, 22)$ and $(84, 31)$ must have a different underlying reason. With an average expected charge of about 500 PE, the modeling of these events should not depend strongly on statistical fluctuations except for a potential lack of “similar” events in the training data. As demonstrated in Figure 5.6, another difficulty lies in finding an optimal solution for the values of the mixture model components. The mixture models containing 5 and 10 AGs are capable of modeling the distribution peak as shown in the upper panel. However, the solution is degenerate and the success of the gradient-based minimizer[3] highly depends on the selected seeds. The bottom panel illustrates the results obtained from fits with different seeds for the position of the AGs. The mixture model becomes more expressive with an increasing amount of components, but it is also increasingly difficult to find a good solution. While the 5-component mixture model fits tend to achieve a decent overall solution, the 10-component model has outliers that fail to appropriately model the density. It also appears that the variation around the peak of the distribution is largest, indicating that this region is the most difficult to fit and the most sensitive to an inappropriate seed. Although the BFGS-minimizer in this example is not directly comparable to the stochastic gradient descent in the training of the EVENT-GENERATOR models, this may explain why the deviations seen for the EVENT-GENERATOR models mainly appear around these peaks. The fact that all three independent EVENT-GENERATOR models have similar deviations might further suggest that the found solution is a compromise for the shared training data, which averages over a multitude of “similar” time PDFs. In addition, while the architectures and hypotheses differ between the models, the chosen seeds for the mixture model components are identical. This will further contribute to finding a similar solution for each of the models.

Around 400 ns after the starting flank of DOMs $(84, 22)$ and $(84, 31)$, there are some oscillations visible in Figures 5.5 and 5.6. These are caused by the transition of the readout chip from the Analog Transient Waveform Digitizer (ATWD) to the fast analog-to-digital converter (fADC). The ATWD has a readout window of 427 ns with a sampling rate of 300 Msps while the fADC saves an interval of 6.4 µs with a sampling rate of 40 Msps. The oscillations are caused by the lower sampling rate. Further details on the readout electronics are provided in Ref. [26].

The 5 or 10 AG mixture models are not capable of modeling these oscillations. More components or perhaps a periodic basis function are required to achieve this. However, these oscillations are only relevant for high-charge DOMs, which have not saturated. As the position of these oscillations depends on the time of the first measured pulse, the signature will be washed out for fainter DOMs. Furthermore, the oscillations only happen in the tail of the distributions. It can therefore be assumed that the average description by the current mixture models is sufficient. This also demonstrates that the chosen parameterization via the mixture model is robust against small-scale fluctuations and that it is therefore a suitable candidate to obtain smooth PDFs. The number of components can be interpreted as a tuning parameter that enables regularization of the modeled densities. A lower number of components leads to less expressiveness and thus to a reduced capability to overfit.

---

[3] In this example, the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm is used within the SCIPY framework (`https://docs.scipy.org/doc/scipy/reference/optimize.minimize-bfgs.html`).

On the other hand, too few components will result in an over-simplification of the model and the inability to adequately model the underlying density.

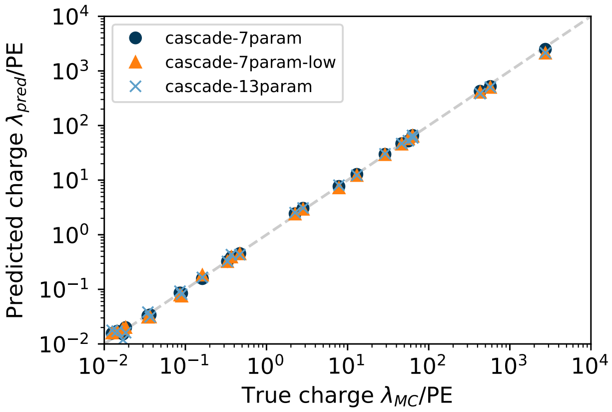


Figure 5.7: The estimated charge for each of the investigated DOMs is compared against the true average charge obtained from half a million re-simulations. As demonstrated by the alignment on the diagonal, each of the Event-Generator models is capable of correctly estimating the average charge over many orders of magnitude.

In addition to the photon arrival time PDFs, the Event-Generator models must also predict the collected charge at each DOM. Overall, the models are able to estimate the observed charge over many orders of magnitude as illustrated in Figure 5.7. However, a closer investigation of the charge distributions at individual DOMs in Figure 5.8 indicates potential deficiencies of the current models. While systematic uncertainties lead to a widening of the charge distribution that may be described by a gamma-Poisson mixture (Section 5.3), additional calibration and readout effects result in further features that are not well modeled by these distributions. In particular, a collection of events exist that exhibit a low number of charge for otherwise high-charge DOMs as shown in the lower panel of Figure 5.8. This collection of outliers cannot be adequately modeled by the gamma-Poisson mixture parameterization. As a result, the generator models are forced to find a compromise that leads to an overall bias in the average estimated charge. This mis-modeling of the charge PDF may degrade reconstruction performance and contribute to challenges in obtaining proper uncertainty contours (Section 5.6.2). The low-charge outlier events are likely results of calibration and readout effects. Investigations show that a large fraction of these events are marked accordingly and are thus appropriately handled (Section 5.3.1) in the training procedure of the models. However, this does not seem to be the case for all events and the models are still sensitive to some of these outliers. This will require further investigation.

If systematic parameters are included in the model hypothesis (Section 5.1), the impact and correct modeling of these may also be verified via MC simulations. Of the three trained Event-Generator models, the `cascade-13param` model includes six systematic parameters as nuisance parameters. The influence of these six systematic parameters on the pulse arrival time PDF is demonstrated in Figure 5.9. As shown, the DOM efficiency has no impact on the shape of the time PDF, while

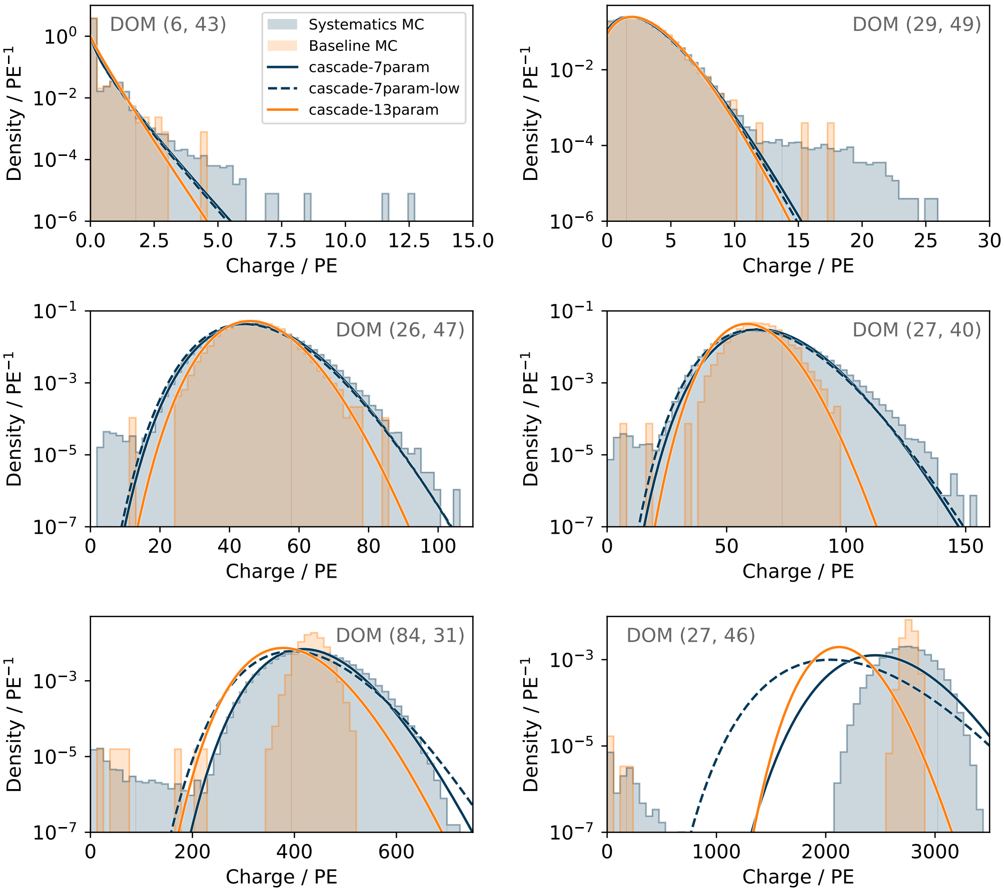


Figure 5.8: The distribution of observed charge at six different DOMs, obtained from event re-simulations, is compared against the estimated distributions predicted by three Event-Generator models. The inclusion of systematic variations (Systematic MC) results in a wider distribution than the simulations that utilize the baseline settings for systematic parameters (Baseline MC). In addition, there are further outliers, particularly at low charge, possibly due to calibration and readout effects.

the impact of other systematics generally depends on the orientation and distance between the emitting cascade and receiving DOM. Figure 5.9 illustrates the changes of the time PDF relative to the baseline exemplary for three different DOMs. In each case, the model accurately captures the influence of each of the systematic parameters. Thus, the Event-Generator model is sensitive to and capable of describing these subtle effects caused by systematic uncertainties. While this means that the impact of systematics on the event reconstruction results must be carefully investigated, it also provides an opportunity in which this approach could be used

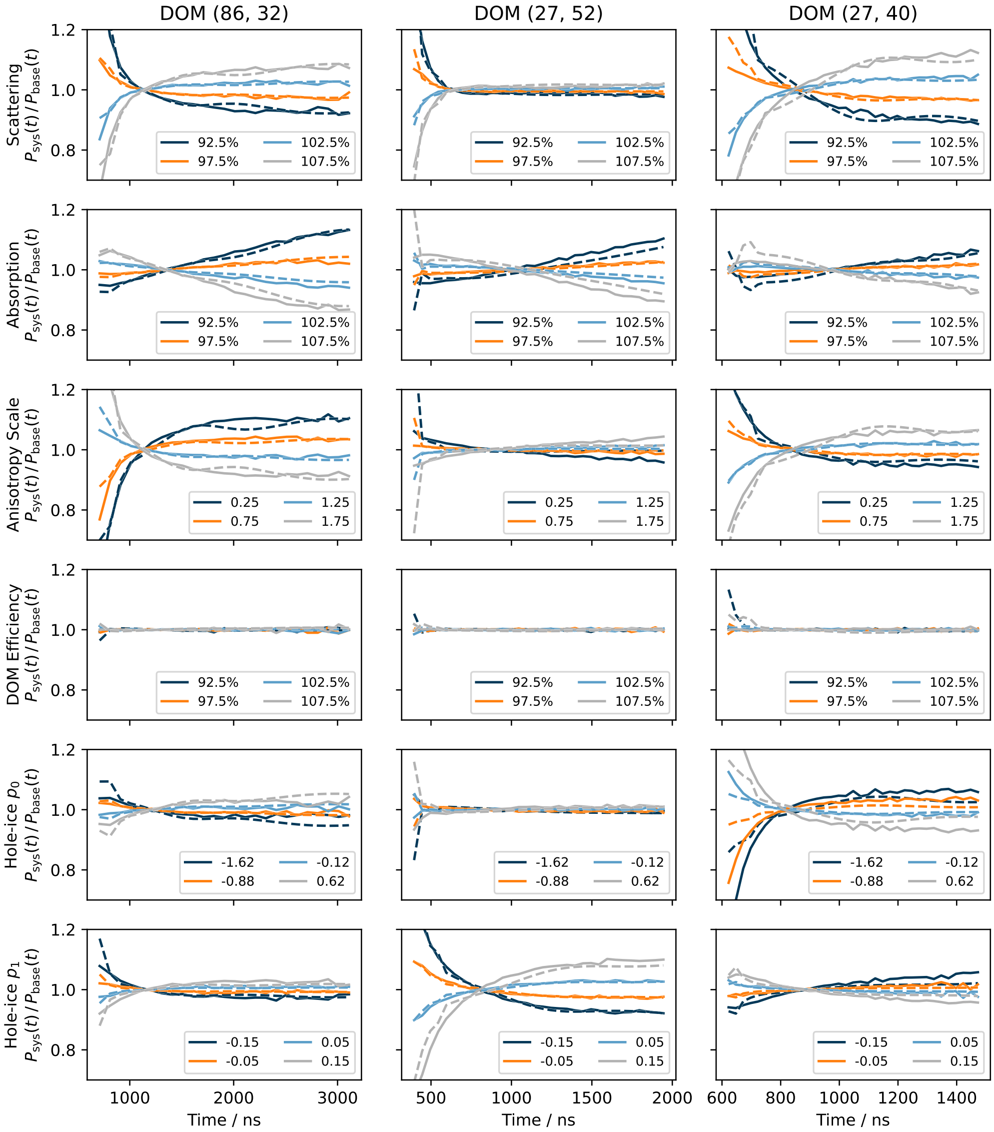


Figure 5.9: The impact of systematic parameters on the photon arrival time PDFs of three exemplary DOMs is shown relative to the baseline for the `cascade-13param` model predictions and the true underlying distribution as obtained from MC simulations. Ratios close to one indicate no or only minor impacts due to the specific systematic parameter (see DOM efficiency for example), while values further away from one demonstrate a larger influence on the arrival time PDF. The model predictions (dashed lines) generally follow the true distributions (solid lines) well.

to measure these parameters with the help of calibration data (Section 5.8 [Detector Calibration with EVENT-GENERATOR]).

5.5.2 Model Investigation

Cross-checks against MC simulation, as performed in the previous section, are most meaningful for the evaluation of the correctness of the generator model. However, these checks require large amounts of computational resources for the simulation and are therefore limited here to only one simulated cascade hypothesis. An alternative method to verify the validity of the fitted model is to investigate how the predicted distributions vary with respect to the event hypothesis. Once the models are trained, these estimated distributions can be obtained by a simple forward pass through the NN, which only requires a few tens of milliseconds on a CPU.

Crucial to a successful minimization process for the event reconstruction (Section 5.6) are smooth and continuous transitions between the distributions corresponding to varying event hypotheses. Figure 5.10 illustrates how the pulse arrival time PDF changes as a function of the zenith direction and vertex position of the cascade when varied around the nominal injection vertex $(100\,\mathrm{m}, -100\,\mathrm{m}, -300\,\mathrm{m})$ and direction of $(\theta = 150^\circ, \phi = 177^\circ)$. The underlying PDFs are expected to depend continuously on these parameters, and this is indeed the case for the model predictions. Apart from smooth transitions, the predicted distribution shapes match expectations based on the geometry of the emitting cascade and receiving DOM. The PDFs are more peaked and earlier in time, the closer the cascade is and the more direct the Cherenkov light is pointed towards the DOM. This is well visible in the top, left panel of Figure 5.10 when the injected cascade is closest to the position of DOM (27, 46) at a z-coordinate of about $-270\,\mathrm{m}$. Moving the injected cascade below the receiving DOM results in more diffuse light as the emitted photons are required to scatter upward in order to reach the DOM. The same trends can be seen for the distributions at DOMs (84, 22) and (86, 50) with coordinates $(57\,\mathrm{m}, -106\,\mathrm{m}, -239\,\mathrm{m})$ and $(-11\,\mathrm{m}, 7\,\mathrm{m}, -430\,\mathrm{m})$, respectively, and for varied zenith angles (right panel). These checks indicate that the trained models adequately model the pulse arrival time distributions.

IceCube events exhibit approximate symmetries including translational invariance in space and time as well as rotational invariance in the x-y-plane. These symmetries are directly incorporated into the EVENT-GENERATOR architecture in addition to other domain knowledge such as the linear energy scaling (Section 5.4). As a result, the EVENT-GENERATOR models are capable of exploiting this information and its predictions reflect upon these symmetries, as illustrated in Figures 5.11 and 5.12. Figure 5.11 highlights the expected linear relationship between injected cascade energy and estimated charge. This relationship is valid over many orders of magnitude and even extends into regions that are not covered by training data. A

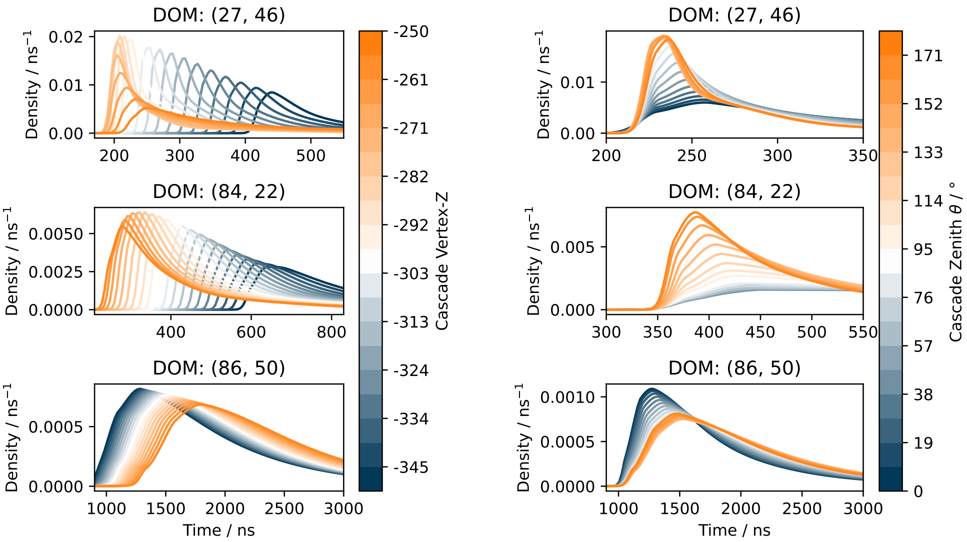


Figure 5.10: The pulse arrival time PDF, approximated by the `cascade-7param-low` model (similar for other models), is shown for three different DOMs of the same event. The left panel shows the effect of modifying the z-coordinate of the cascade interaction vertex, while the right panel illustrates the change due to the varying zenith angle. Figure and caption adopted from [5].

lower charge floor of less than 0.1 PE stems from the contribution of noise pulses, which are independent of the injected cascade energy.

Figure 5.12 demonstrates the approximate translational and rotational symmetries. In the left panel, translational symmetries are investigated by shifting the injected cascade to another position while maintaining the relative distance and orientation to the illustrated DOM. The legend shows the predicted number of photoelectrons (PE) at the given DOM, the string and DOM number, and the position of the injected cascade. In a perfectly homogeneous medium, the time arrival PDFs in Figure 5.12 should be identical for each of the configurations shown. However, translational invariance in IceCube is violated by ice layers with differing optical properties as well as an overall position-dependent tilt in the layers due to the structure of the underlying bedrock [24, 25, 30]. As a result, the curves corresponding to injected cascade positions $(-338\,\text{m}, 269\,\text{m}, -501\,\text{m})$ and $(46\,\text{m}, -19\,\text{m}, 387\,\text{m})$ differ in shape and predicted charge. The dotted blue curve corresponds to a cascade in the clear ice at the bottom of the detector. In this area, the ice is characterized by large scattering and absorption lengths resulting in a more peaked distribution and higher estimated charge. In contrast, the ice at a z-coordinate of 387 m is heavily influenced by dust impurities resulting in more scattered and absorbed light leading to less detected photons and a flatter and less distinct peak. The remaining curves

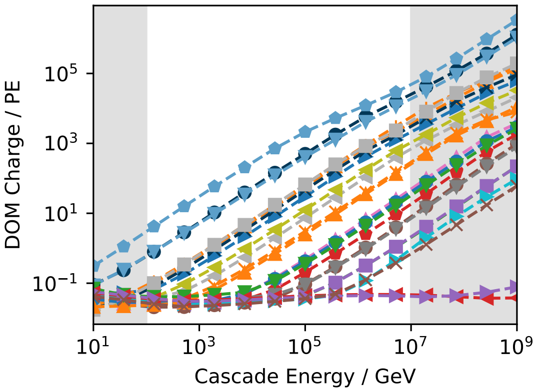


Figure 5.11: The estimated charge is shown as a function of the injected cascade energy for the `cascade-7param-low` model (similar for other models). Each of the lines corresponds to one of the investigated DOMs. Although energies below 100 GeV and above 10 PeV (shaded regions) are not covered by the training data, the expected linear relationship between predicted charge and cascade energy extends to these energies.

correspond to regions in the detector with more aligned ice properties resulting in near identical distributions, thus more strongly demonstrating the translational symmetry.

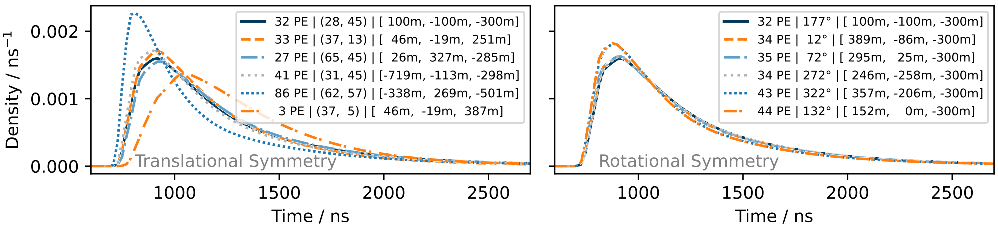


Figure 5.12: The estimated pulse arrival time PDFs for various shifted and rotated cascade-DOM configurations are shown to demonstrate translational and rotational symmetries in IceCube and deviations thereof. For a hypothetical homogeneous detector medium, all of the shown distributions would be identical due to the underlying symmetries of the neutrino interaction.

In the right panel of Figure 5.12, rotational symmetry is demonstrated by the rotation of the injected cascade in the x-y-plane around DOM (28, 45). In addition to a modified cascade vertex, the orientation of the injected cascade is also rotated in order to maintain the relative opening angle to the observing DOM. The resulting cascade azimuth angle and vertex is provided in the legend in addition to the estimated charge at DOM (28, 45). Similarly to the left panel, the resulting curves for each of these configurations should align in an homogeneous medium. Given that all of these cascades are injected in a similar area of the detector, the minor differences visible do not stem from differing dust concentrations in the ice, but rather from anisotropic light propagation. Due to the crystal structure of the ice, photons are subject to differing scattering and absorption lengths depending on their orientation [29, 30]. Absorption is reduced along the flow direction of the ice ($\sim 130^\circ$), resulting in larger estimated charge for the shown azimuth angles of 132° and 322°. By construction, the architecture of the Event-Generator models enables the utilization of these symmetries, while still providing enough freedom through the initial locally connected layers (Section 5.4) to model the aforementioned second order deviations.

5.5.3 Model Extrapolation

As outlined in Section 2.1 [The Importance of Domain Knowledge], utilization of available symmetries and domain knowledge is beneficial to a model's convergence and overall performance, and crucial to meaningful inter- and extrapolation to phase space not — or not sufficiently — covered by training data. The Event-Generator architecture aims to combat this by directly incorporating available information into the network architecture. As shown in Figures 5.11 and 5.12, due to the direct inclusion into the NN architecture, the models are able to extrapolate along the given symmetries. The training data for the Event-Generator models do not contain events below 100 GeV, only a handful events between 10 PeV and 100 PeV and no events above that. Yet, the models are able to extrapolate into this region by utilizing the linear energy scaling. The same applies for translational invariance by utilizing the relative displacement vectors as input features for each DOM (Figure 5.3). This allows the model to accurately model the pulse arrival time PDFs even for cascades injected far outside of the detector where training data is very sparse and insufficient, as demonstrated in the grey, dotted line on the left of Figure 5.12 corresponding to a cascade vertex of ($-719\,\mathrm{m}, -113\,\mathrm{m}, -298\,\mathrm{m}$). Time invariance is trivially incorporated by employing a continuous parameterization of the time arrival PDFs via mixture models, which allows for a simple offset to the function argument. This essentially removes the time dimension from the NN model and enables perfect extrapolation to any arbitrary cascade interaction time.

5.6 Event Reconstruction

The Event-Generator framework allows for the training of generative models that predict the pulse arrival pattern in the detector for a specified event hypothesis. Once such a model is trained, it may be used in a maximum-likelihood setup to reconstruct events as illustrated in Figure 5.13. Given an initial event-hypothesis, the predicted distribution of pulses in the detector is compared against the observed distribution of pulses by computing the likelihood (Section 5.3 [Likelihood Prescription]) of the observed pulses for this particular event-hypothesis. In a subsequent step, the event-hypothesis is adjusted such that the likelihood is maximized. This process is repeated until an optimum is found.

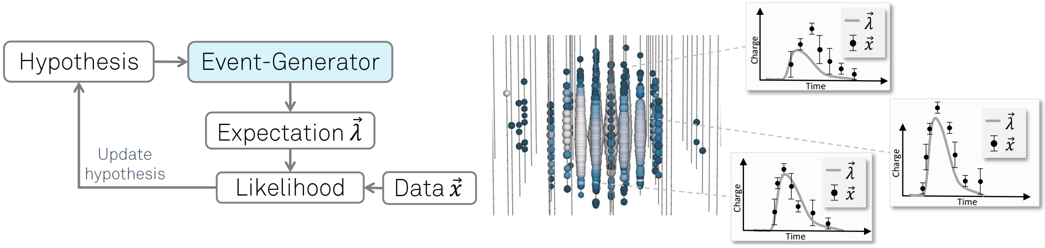


Figure 5.13: A trained EVENT-GENERATOR model may be used to reconstruct events via maximum-likelihood estimation (Section 3.3). Expected pulse arrival distributions $\vec{\lambda}$ are obtained from the EVENT-GENERATOR model and compared to observed data, as illustrated on the right. The likelihood of the observed pulses ($\vec{x}$) is optimized in an iterative fashion by updating the event hypothesis parameters.

In contrast to MC simulations, the EVENT-GENERATOR is differentiable with respect to the event-hypothesis. From a deep learning perspective, this corresponds to computing gradients with respect to the neural network inputs rather than the network's parameters. Although this differs from the typical utilization of gradients in backward-propagation during training of the model, the computation is analogous and possible because the neural network itself is differentiable. Available gradients with respect to the event-hypothesis not only enable an efficient optimization process during event reconstruction, but they also provide a cost-effective way to inspect the likelihood landscape and thus to estimate uncertainties. Due to the model architecture and its training, the EVENT-GENERATOR models mitigate noise and fluctuations in the simulation of individual events, leading to smooth descriptions of the underlying PDFs (Figure 5.10). This smoothness in addition to the continuous transitions between the pulse arrival time PDFs, corresponding to varying event hypotheses, are crucial for a minimizer to robustly and effectively converge in a proper minimum. These are reasons why the EVENT-GENERATOR can — in practice — outperform a maximum-likelihood-based reconstruction method [57–59] that directly utilizes MC simulations to obtain the photon arrival time PDFs, although these methods are theoretically optimal.

The EVENT-GENERATOR software framework utilizes a modular structure that allows the user to choose which minimizer and minimization strategy to utilize. Among more standard minimizers, the user may also choose to run a Markov-Chain-Monte-Carlo. Per default, a gradient-based minimizer, the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, is used within the SCIPY framework[4]. Figure 5.14 illustrates the final iteration of the optimization process for a given example event. The top six DOMs that most contribute to the overall log-likelihood value are shown. The observed pulse arrival distributions at each DOM are compared against the model predictions shown in blue. This particular event with a true deposited energy

---

[4] `https://docs.scipy.org/doc/scipy/reference/optimize.minimize-bfgs.html`

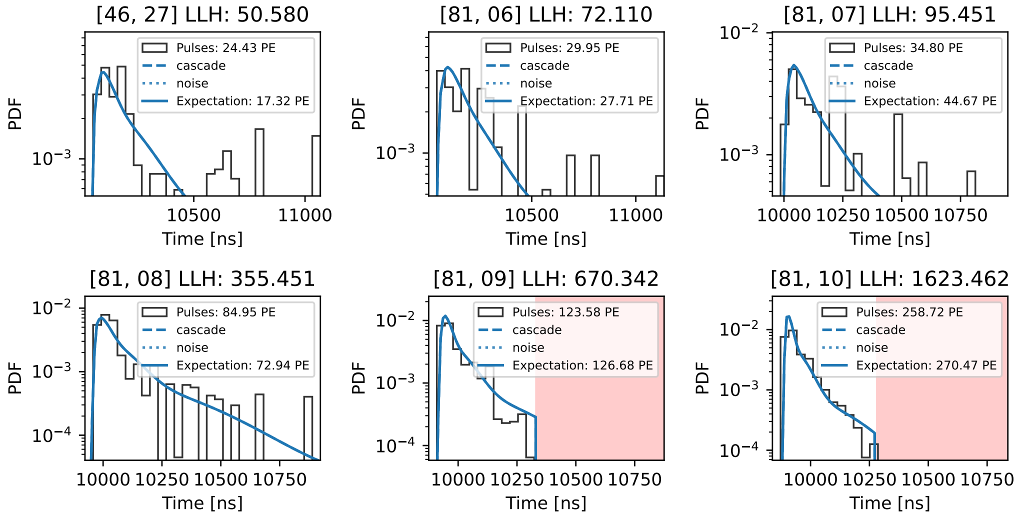


Figure 5.14: The observed distribution of pulses for a given example event is compared against `cascade-7param-low` model predictions for the six DOMs that most contribute to the overall log-likelihood value. Contributions of individual model source components are shown in dashed (cascade) and dotted (noise) lines while the combined model expectation is given in the solid blue line. The contribution of the noise component is sub-dominant here and therefore not visible. The title of each panel indicates the string and DOM number and its contribution to the log-likelihood. Regions shaded in red indicate time exclusion windows (5.3.1).

of 4192 GeV (reconstructed to 4089 GeV) is well reconstructed as indicated by the agreement between observed pulses and estimated distribution. The true vertex lies within 2.2 m and 2 ns of the reconstructed vertex and the direction is correctly reconstructed to within 4°.

5.6.1 Reconstruction Performance

In order to quantify the performance of the Event-Generator as a reconstruction method, comparisons against standard likelihood-based reconstruction methods in IceCube are provided in this section. Although the main focus in this dissertation is on the reconstruction of cascade-like events, the versatility and potential of the Event-Generator framework is demonstrated by also comparing against double-cascade (Taupede) and track reconstructions (SplineMPE, Millipede). More information on these reconstruction methods is provided in Section 3.4.

The performance for cascade-like events is evaluated on simulated ν_e events that pass the event sample utilized in Ref. [67], which is also referred to as the “Medium

Energy Starting Cascade" (MESC) sample. A few thousand of these ν_e events are re-simulated with various different ice models and realizations of systematic parameters via the SnowStorm method [74] to account for systematic uncertainties. This ensemble of events is then reconstructed with the trained EVENT-GENERATOR models as well as MONOPOD[5]. The resulting resolution curves as a function of deposited energy are illustrated in Figure 5.15.

As shown in Figure 5.15, the `cascade-7param` and `cascade-7param-low` models provide nearly identical reconstruction resolution. These models outperform the CNN, built within the DNN-RECO framework (Chapter 4), and the standard reconstruction method MONOPOD over the entire energy range. The improvement in angular resolution in comparison to MONOPOD ranges from a few tens of percent at TeV energies to almost a factor of two at a few hundred TeV. Systematic uncertainties limit resolution at higher energies. In an idealized scenario without any systematic uncertainties, the resolution may be further reduced down to 2–3° at 100 TeV. At lower energies, where the CNN experiences difficulties in exploiting timing information (Section 4.5.3 [Pulse Modifications]), the EVENT-GENERATOR models are capable of improving by over a factor of two in resolution. Improvements in angular resolution are most important for neutrino source searches for which they can have a considerable impact on the sensitivity of an analysis (Chapter 12 [Analysis Performance]).

In addition to the origin of a neutrino, the deposited energy is an important proxy variable utilized to distinguish signal from background events. The 7-parameter EVENT-GENERATOR models are able to improve the energy resolution on this ensemble of events by roughly 30%, as seen in the lower panel of Figure 5.15, resulting in a median resolution of about 0.03 in $\log_{10}$-space corresponding to a relative accuracy of about 7% of the true deposited energy. However, systematic uncertainties dominate the energy resolution and, if conservative ranges are assumed, the differences in resolution between the depicted reconstructions will converge. For instance, the uncertainty on the DOM efficiency linearly influences the observed light yield and therefore also the reconstructed energy. Although likely too large, an assumed uncertainty on the DOM efficiency of 10%, which would directly translate to an uncertainty on the energy reconstruction of at least the same amount, is not ruled out by detector calibration at this point.

The vertex position (second panel), while not directly used in the analysis, helps to identify the location and topology of an event. This information is beneficial to the event selection step (Chapter 7 [DNNCASCADE: A novel Cascade Dataset]). A median resolution between 2 m and 6 m is achieved depending on the reconstruction method and energy range. Timing resolution of the event trigger $\mathcal{O}\left(\mu s\right)$ is already sufficient for any analysis level requirements. However, investigation of the timing

---

[5] Here, MONOPOD is run with the BFRv2 spline tables including tilt and effective distance tables, which provides the best results compared to other configurations.

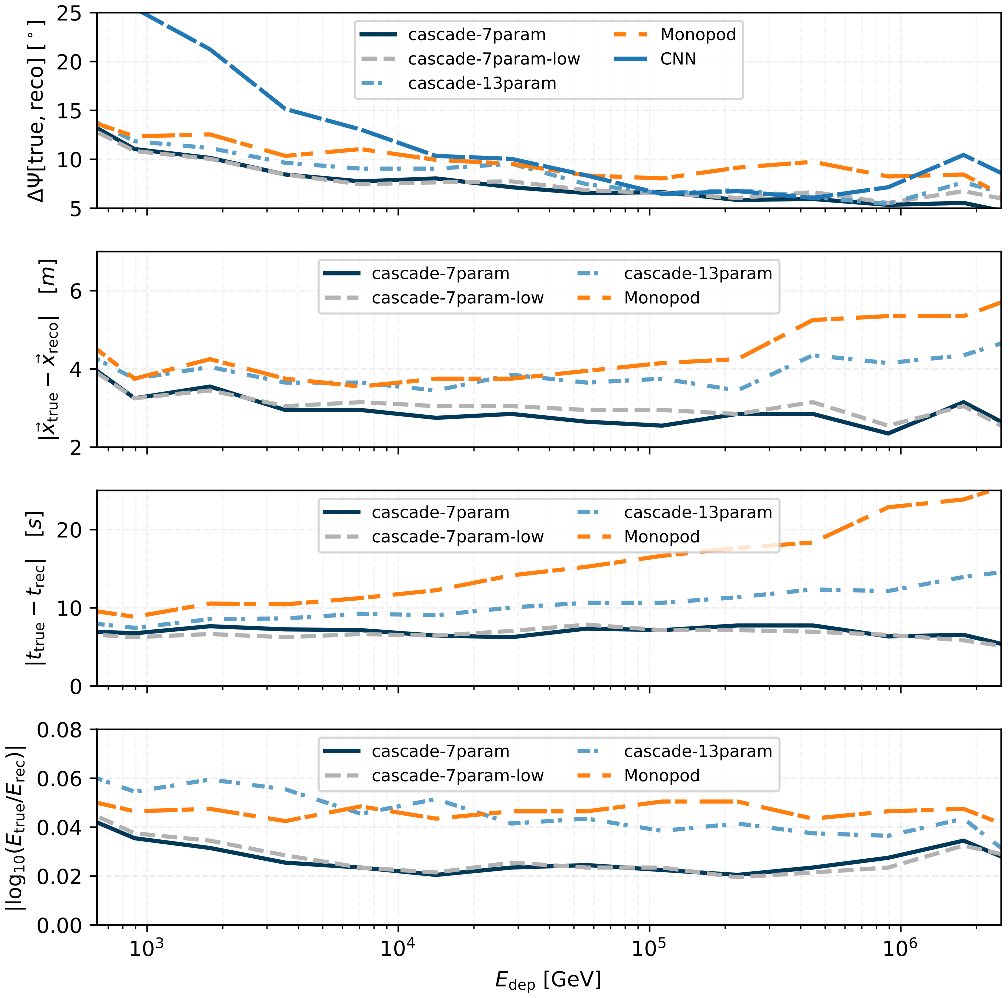


Figure 5.15: The resolution as a function of energy is shown for the cascade direction, vertex position, vertex time and deposited energy, from top to bottom panel, respectively. The resolution curves indicate the 50% quantile of the quantity as defined in the y-axis label at a given true deposited energy. Both 7-parameter Event-Generator cascade models provide the best resolution for all reconstructed quantities and energy ranges.

resolution can help to identify if the reconstruction works as intended. Plus, for more complex event signatures and separation thereof, accurate timing resolution of individual components is crucial. The `cascade-7param` and `cascade-7param-low` models achieve an energy independent resolution of about 8 ns, while the resolution for MONOPOD ranges from about 10 ns at TeV energies up to 25 ns at 1 PeV. The correlation of MONOPOD's time and vertex resolution curves to the deposited energy likely stems from the simplified assumption of a point-like energy deposition, which neglects an extension of the cascade shower. Note that the reconstructed vertex of MONOPOD is shifted back by the average, energy-dependent cascade extension in these comparisons. This significantly corrects the interaction vertex and time resolution, but cannot completely negate the impact of the simplified assumption.

Overall, the 7-parameter EVENT-GENERATOR models are able to improve the reconstruction accuracy for every quantity and energy range in comparison to the standard reconstruction method MONOPOD and the CNN within the DNN-RECO framework. This enhancement is possible because the EVENT-GENERATOR utilizes symmetries and domain knowledge, which the CNN is incapable of, while avoiding simplifications that are necessary for the spline table approach used in MONOPOD (further discussed in Section 3.4 [Traditional Reconstruction Methods in IceCube]). Thus, the EVENT-GENERATOR is capable of combining the strengths of MONOPOD and the CNN. A downside of the EVENT-GENERATOR in comparison to the CNN, however, is the required time to reconstruct an event, which is on the order of seconds, comparable to the runtime required by MONOPOD, in contrast to milliseconds as needed by the CNN. The reason for the increased runtime lies in the necessary optimization loop in the maximum-likelihood setup, while the CNN only requires a single forward pass through the network.

The `cascade-13param` model leads to inferior performance compared to the seven parameter models. This is interesting, because — in theory — the 13-parameter model should be more sensitive as it can fit the systematic parameters as nuisance parameters in the reconstruction step. In contrast, the 7-parameter models are forced to model the average PDFs, marginalized over the distribution of systematic parameters in the training data. This likely leads to more robust results at the cost of reconstruction accuracy. These preliminary results indicate that there may be additional challenges to overcome when expanding event hypotheses to higher dimensions to include systematic nuisance parameters. In principle, there are two main areas in which deficiencies may exist. Either the reconstruction is more challenging in high dimensions leading to bad local minima, or the model itself needs further optimization. It's possible that the training of the `cascade-13param` model did not properly converge and/or that the provided training data isn't sufficient to cover the 13-parameter phase space. Insufficient training data for a more detailed event hypothesis may lead to confusion rather than improved sensitivity since the patterns may not be evident enough given the noise and fluctuations. This will need further study and potentially a modified network architecture. Assuming

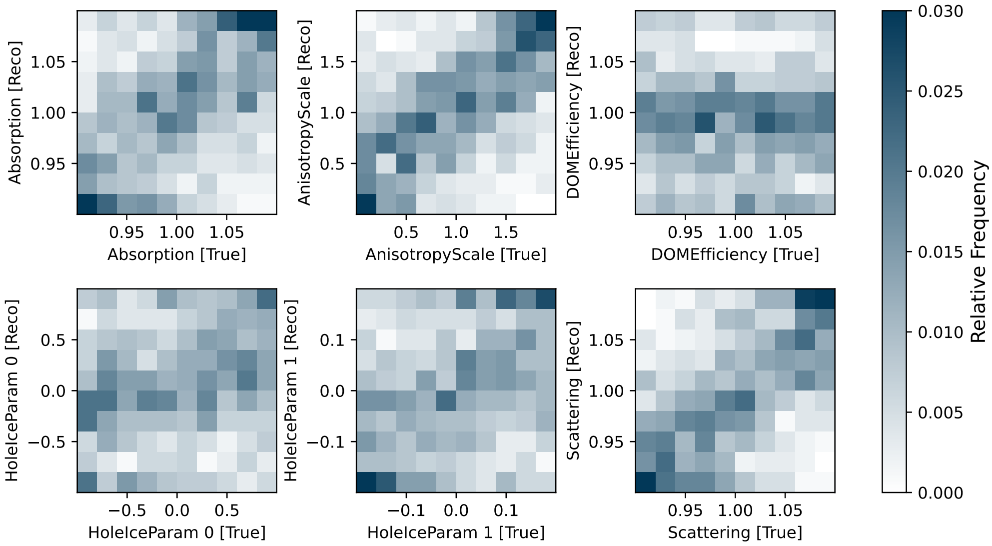


Figure 5.16: Correlation plots between true (x-axis) and reconstructed (y-axis) values of the `cascade-13param` model are shown for each of the six systematic parameters. Among these variables, the correlation for the strength of the ice anisotropy (AnisotropyScale) in the top, middle panel is strongest as indicated by the accumulation of data points on the diagonal.

that the relative changes in systematic parameters affect the pulse arrival time PDFs uniformly and independent of a specific DOM, then the network components pertaining to the systematics can be factored out. This would facilitate the training of the neural network as the dimensionality is effectively reduced from 13 parameters down to two separate problems of 7 cascade parameters (already solved) and 6 parameters for the modification of an arrival time PDF for a given set of systematic parameters.

Nevertheless, the results for the `cascade-13param` model show that it is in principle possible to include further nuisance parameters to the event hypothesis. In the previous section, Figure 5.9 demonstrates that the `cascade-13param` model is sensitive to the systematic parameters, apart from the DOM efficiency, which does not affect the shape of the underlying pulse arrival time PDF. As a result, the model should — to a certain degree — be capable of reconstructing the injected systematic parameters. As shown in Figure 5.16, there is indeed a weak correlation visible between reconstructed and true parameters. The correlation is strongest for the systematic parameter that regulates the strength of the ice anisotropy [29, 30], but it is also evident in the ice scattering and absorption parameters. However, the correlation is fairly weak and the parameters are mis-reconstructed for a large number of events. In Section 5.8, the potential of the Event-Generator framework is investigated for the task of detector calibration. For the presented study, the same cascade is injected in the detector for varying sets of systematic parameters. An Event-Generator model is then trained to estimate the underlying PDFs as a function of the six systematic parameters. This essentially reduces the dimensionality of the `cascade-13param` model down to six parameters for a fixed cascade hypothesis and the study can be interpreted as a test to investigate whether a more densely populated phase space leads to more sensitivity towards the systematic parameters. As shown in Figure 5.22, the trained model is able to accurately recover the injected systematic parameters. This generally indicates potential for a future 13-parameter model that may succeed the performance of the `cascade-13param` model discussed here.

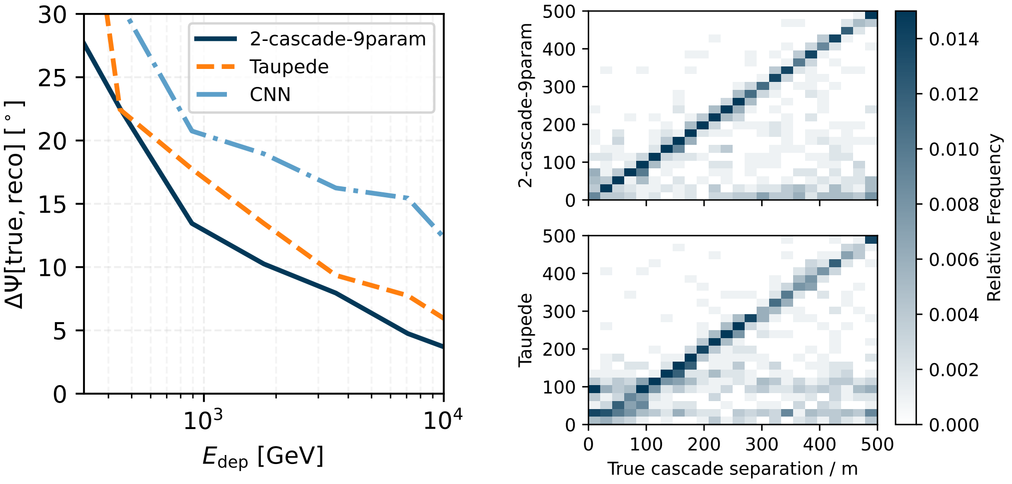


Figure 5.17: The reconstruction performance on the artificial double-cascade sample is demonstrated. On the left, the angular resolution is compared as a function of total deposited energy. On the right, the correlation between true and reconstructed cascade separation distance is shown for the Event-Generator reconstruction (top panel) and Taupede (bottom panel).

Although the emphasis in this dissertation is on the development and optimization of cascade reconstruction methods, the DNN-reco and Event-Generator frameworks are designed such that they are applicable to arbitrary event topologies. As discussed in Section 5.1 [Event Hypothesis], the Event-Generator framework can be utilized to construct multi-source objects consisting of individual sources. Here, a double-cascade event hypothesis is constructed based on the `cascade-7param-low` model, which is referred to as `2-cascade-9param` model in the following discussion. In addition to the original seven cascade parameters, the energy and distance to a second cascade interaction along the direction of flight are added. The `2-cascade-`

`9param` model is not retrained, but it simply acts as a wrapper around the underlying `cascade-7param-low` model, which is applied twice to obtain the light yield for the double cascade event.

A sample of artificial double-cascade events is simulated by injecting two correlated cascades with a shared direction and energies between 100 GeV and 10 TeV each. The second cascade is injected along the direction of flight of the first cascade with uniformly sampled separation distances between 0 m and 500 m. Note that these characteristics are not a realistic scenario for ν_τ induced double-cascade events, but they allow to better investigate the performance of the reconstruction methods. In addition to the `2-cascade-9param` model, the TAUPEDE reconstruction method, which is the equivalent standard reconstruction method for double-cascade events, and the CNN are applied on this ensemble and their performance is compared in Figure 5.17. For this ensemble of events, the EVENT-GENERATOR reconstruction is able to provide the most accurate angular reconstruction as indicated on the left of Figure 5.17. In particular at small separation distances, the `2-cascade-9param` model appears to be performing better than TAUPEDE, which experiences a larger fraction of mis-reconstructed events as visible in the bottom correlation plot on the right of Figure 5.17.

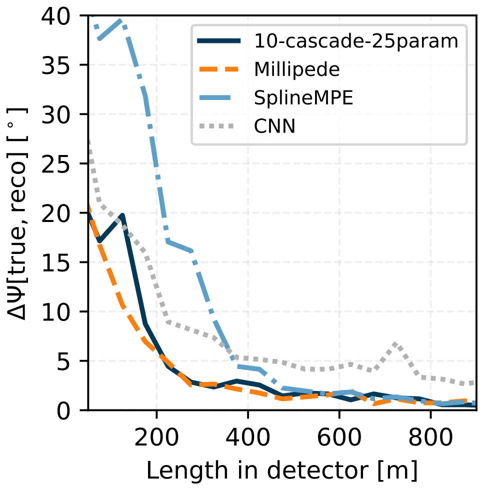


Figure 5.18: The median angular resolution on the ensemble of artificial track events, consisting of ten correlated cascades, is shown as a function of track length in the detector. The CNN is not specifically trained on this event topology. A dedicated CNN will provide an improved performance, but not surpass the other methods. The preliminary `10-cascade-25param` model provides comparable results to the current state-of-the-art reconstruction methods MILLIPEDE and SPLINEMPE, indicating potential for future EVENT-GENERATOR developments targeting track events.

Analogously to the `2-cascade-9param` model, a multi-source of ten correlated cascades, `10-cascade-25param`, can be defined on the basis of the underlying `cascade-7param-low` model. Similarly to the previous setup, artificial track events consisting of ten correlated cascades are simulated. The location of the cascades are sampled along the direction of flight with varying distances between 0 m and 1000 m of the first cascade. The resulting sample is then reconstructed with the `10-cascade-25param` model, the standard reconstruction method equivalent MILLIPEDE, the CNN, and the most widely used track reconstruction method

SplineMPE. Figure 5.18 shows the median angular resolution as a function of track length in the detector. The resolution of SplineMPe, Millipede and the Event-Generator converge for track lengths above 500 m. At smaller track lengths, Millipede and Event-Generator dominate with a slight preference for Millipede. In contrast to the `10-cascade-25param` event hypothesis, which fits for the location and energy of ten cascades along the track, Millipede only fits for the energies of cascades at fixed steps of 10 m along the hypothesized track. An equivalent event hypothesis with the Event-Generator framework is possible, but in its current implementation the Event-Generator framework does not scale well to multi-sources consisting of multiple hundred individual sources, as required for the Millipede hypothesis. This would require a forward and backward pass through the underlying NN for each of the cascades along the track. An alternative and possibly better solution is to directly train a track hypothesis model with some flexibilty to fit for stochastic energy losses along the track. The preliminary results shown here for the `10-cascade-25param` model demonstrate the potential of the Event-Generator for track reconstructions.

5.6.2 Uncertainty Estimation

Best-fit positions of reconstructed quantities (point estimates) are meaningless without appropriate measures on their uncertainty. As demonstrated in Section 4.4 [Cascade Reconstruction Performance] for the DNN-reco framework, trained parametric models for the uncertainty estimation of a class of similar events is feasible and such an approach can provide adequate coverage on the ensemble of events. This approach is also applicable for the Event-Generator framework (Section 7.3). Nevertheless, a maximum-likelihood-based reconstruction method allows another approach to estimate the uncertainty on a reconstruction on a per-event basis. The shape of the likelihood landscape can be utilized to derive confidence intervals based on the delta log-likelihood values. Different approaches utilizing the inverse Hessian matrix at the best-fit location, marginalized sky-scans, and Markov-Chain-Monte-Carlo (MCMC) are discussed in this section.

Hessian and Sandwich Estimator

Under certain conditions, the inverse of the observed Fisher information matrix (negative of the Hessian matrix), evaluated at the maximum-likelihood estimate, is an estimator of the asymptotic covariance matrix [55, 79]. However, this is only valid if the maximum-likelihood model is correctly specified. In the case of misspecification, the outer product of gradients (OPG) estimate [80] can be used in connection with the inverse Hessian matrix in order to derive a robust estimate of the covariance matrix, referred to as the *sandwich* estimator [79]. The OPG estimate is also relevant for extended maximum-likelihood estimation [76] for which the inverse asymptotic variance is not approximated by the Hessian (second order derivatives) as in standard MLE, but by a product of first oder derivatives.

The Hessian and sandwich estimators are resource efficient methods that utilize the local curvature of the likelihood space at the maximum-likelihood estimate to provide confidence intervals. These estimators require the calculation of first and second order derivatives, which may be computed via back-propagation in the Event-Generator framework.

Markov-Chain-Monte-Carlo

An alternative approach to the aforementioned estimators is a sampling technique of the posterior distribution via Markov-Chain-Monte-Carlo (MCMC). The Event-Generator allows the usage of multiple software frameworks and sampling strategies. In the study presented in this section, the emcee [81] software package is employed. Samples of the posterior distribution are generated via an ensemble of 14 walkers and 2000 sampled points each. These settings are a compromise between acceptable computation requirements and adequate modeling of the posterior distribution. In order to evaluate the coverage of the angular uncertainty estimates, the zenith and azimuth values of the generated samples are parameterized via two different distributions on the unit sphere. The von Mises-Fisher (vMF) [82] distribution, a 1D Gaussian equivalent on the sphere, with “circularized” angular uncertainties, symmetrical in azimuth and zenith, is employed in addition to the Fisher-Bingham distribution (FB_5)[6] [84], an analogue to the general bivariate normal distribution capable of modeling “elliptical” contours on the sphere.

Marginalized Sky Scans

The marginalized posterior distribution in reconstructed zenith and azimuth angles may also be investigated by performing a profile likelihood scan over the sky. In this approach, the sky is divided into equal area bins using the Hierarchical Equal Area isoLatitude Pixelisation (HEALPix) [85]. Optionally, a finer binning may be chosen around regions of interest to allow for more details of the scan while reducing computational requirements. At each pixel location, a maximum-likelihood estimation is performed while keeping the zenith and azimuth angles constant to the values at the current point in the sky. Confidence intervals may be constructed under consideration of Wilk’s theorem [86], or directly by interpreting the normalized profile likelihood scan as an approximation to the posterior distribution. In this

---

[6] The python package fb8 [83] available at `https://github.com/tianluyuan/sphere` is utilized.

section, the latter option is chosen with an additional parameterization via an FB_5 distribution, similar to the MCMC approach, as this provided the best results.

Evaluation of Estimators

Estimated contours at 68 % confidence level are compared for each of the described methods in Figure 5.19. In contrast to the vMF distribution, the FB_5 distribution is able to model elliptical contours, leading to a better description of the underlying distribution. This is particularly visible in the reconstruction illustrated in the bottom right panel, showing a strong correlation between zenith and azimuth angles. For this example, the best-fit vMF distributions for the sky scan and MCMC estimates are offset on either side of the reconstructed maximum-likelihood estimate (white triangle). This offset likely stems from an attempt to model the elongated PDF with a circular distribution. Nevertheless, in many cases, the circularized description via the vMF distribution results in adequate contours, achieving a similar coverage on the ensemble as the FB_5 equivalent.

The coverage for the described uncertainty estimators is investigated in Figure 5.20. The MCMC sampling provides nearly perfect coverage for low-energy events (top, left). However, with increasing energy (left to right), the estimator begins to under-estimate uncertainties. This trend is similar for the other methods[7] with the exception of the sandwich estimator. As discussed in Section 5.5.1 [Comparison to Monte Carlo Simulation], the current Event-Generator models exhibit difficulties in modeling the charge distribution due to rare outliers that are not described by a Poisson or negative binomial distribution (Figure 5.8). The impact of this modeling inability becomes more pronounced at higher energies. Thus, the likelihood models are mis-specified in this energy regime, resulting in potential reconstruction bias and unreliable uncertainty estimates. The sandwich estimator constitutes the only exception as it is able to account for model mis-specification.

To further verify the hypothesized impact of model mis-specification, round-trip simulations are generated. In these simulations, pulses at each of the DOMs are sampled directly from the Event-Generator PDFs (Section 5.7) that are also used in the event reconstruction. This round-trip eliminates any model mis-specification at higher energies. At lower energies, some discrepancies may remain due to discretization issues in number and charge of sampled pulses. Results are illustrated in the bottom row of Figure 5.20, indicating a much improved coverage and thus supporting the claimed impact of model mis-specification. Note also that the Hessian and sandwich estimates provide nearly identical coverage in the absence of mis-specification. Remaining under- and over-coverage of the Hessian and sandwich estimates is likely due to the underlying vMF distribution. The sky

---

[7] This trend is also visible for standard MLE reconstruction methods in IceCube, indicating an intrinsic modeling issue.

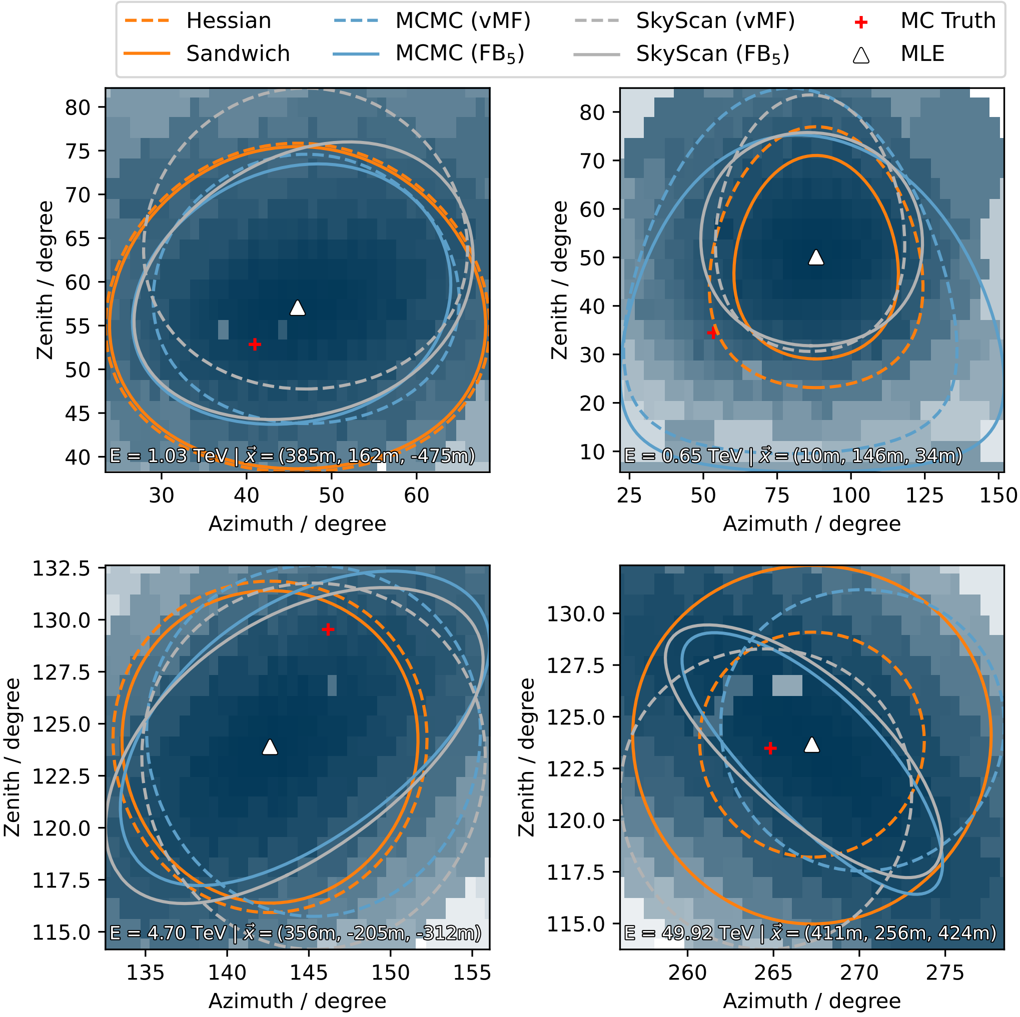


Figure 5.19: Contours (68 % C.L.) as estimated by each of the described methods are shown on top of the marginalized sky scan map for four example events. Darker colors correspond to locations of larger log-likelihood values. The true and reconstructed direction (MLE) are indicated by a red cross and white triangle, respectively. Note that the vMF contours do not appear circular in this visualization due to the aspect ratio and planar projection.

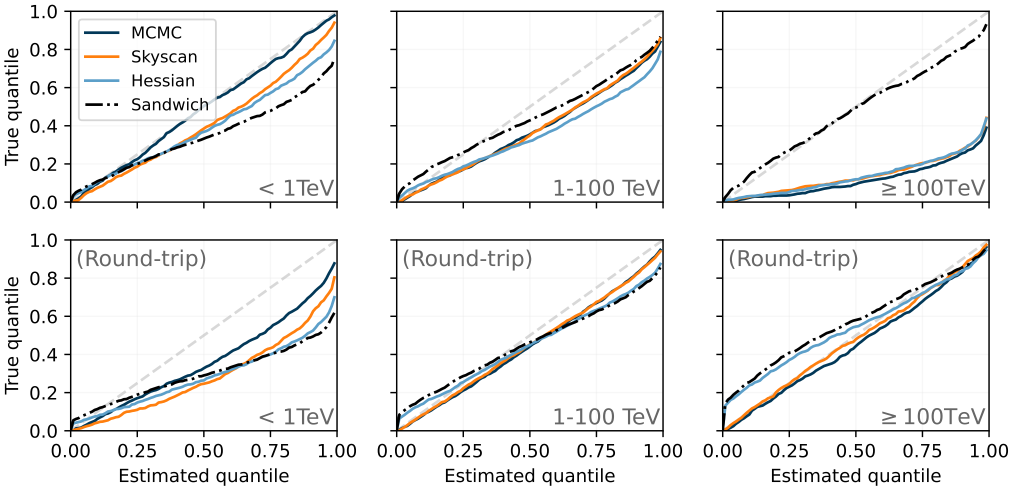


Figure 5.20: The coverage of the angular uncertainty estimate for each of the described methods is evaluated for different energies (columns) on the standard simulation (top) as well as round-trip simulations (bottom). Round-trip simulations re-simulate events by sampling pulses directly from the Event-Generator models (Section 5.7). Perfect coverage is obtained on the diagonal (dashed, light gray).

scan and MCMC results shown in Figure 5.20 are based on the more general FB_5 distribution.

In summary, neither of the presented uncertainty estimators are currently applicable to all energy ranges without further calibration. Thus, for the analysis described in this dissertation (Part III), a semi-parametric estimator is constructed based on these per-event estimators, leading to adequate coverage on the ensemble of events as demonstrated in Figure 7.9. Details are provided in Section 7.3 [Event Reconstruction and Uncertainty Estimation].

5.7 Event-Generator as a Simulation Tool

The Event-Generator is a versatile tool that can be utilized for multiple applications. In principle, the employed generative neural network is an approximation to the Monte Carlo (MC) simulation. As such it can be used to generate new events. This may be helpful to avoid computationally expensive MC simulations, especially at higher energies.

The generative model is trained to predict the average light yield for a given event hypothesis $\vec{\xi}$. In order to add event-by-event variations, pulses must be sampled

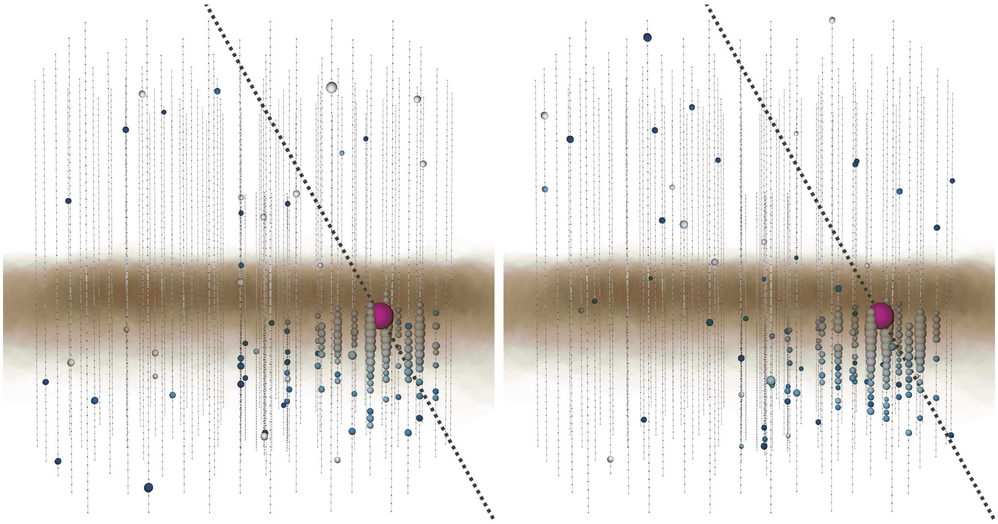


Figure 5.21: An event view of a cascade interaction just below the dust layer is shown for the Monte Carlo simulation on the left and the generative model on the right. Both events match well within the expected statistical fluctuations. The generative model is able to capture the attenuating effect of the dust layer. Figure and caption taken from Ref. [5].

from these expectation values. This is achieved by first sampling the total charge at each DOM. On average, a photon hitting the PMT will generate a measurable charge of about 1 Photoelectrons (PE), but there are deviations. To account for this, the sampled total charge at the d-th DOM is distributed over N_d pulses such that the distribution of pulse charges roughly corresponds to the single photon charge distribution as obtained from MC simulation.

Afterwards, the time of each pulse is sampled in two steps. For each pulse, two random numbers q_1 and q_2 are drawn between 0 and 1. The first of which is used to sample the corresponding mixture model component. As described in Section 5.2, the pulse arrival time PDF $P_d(t)$ from Eq. (5.7) for a source S_i is parameterized by a mixture model of asymmetric Gaussians (AGs). This still holds true for a multi-source M_j, which is just a linear combination of individual sources. Each AG can therefore be attributed a weight that quantifies the contribution to the overall time PDF $P_d(t)$. A cumulative sum

$$\left(\vec{\Sigma}_{\mathrm{cum}}\right)_k = \begin{cases} 0, & k = 0 \\ \sum_{i=1}^{k} w_i, & k > 0 \end{cases} \tag{5.22}$$

is then computed from all weights, where k goes from 0 to the number of mixture components K. Based on the first random number q_1, the k-th mixture component is chosen such that

$$\left(\vec{\Sigma}_{\text{cum}}\right)_{k-1} < q_1 \leq \left(\vec{\Sigma}_{\text{cum}}\right)_k \tag{5.23}$$

is valid. The second random number q_2 is used to sample the time of the i-th pulse

$$t_k^i = \text{CDF}_{\text{AG}}^{-1}(q_2|\mu_k, \sigma_k, r_k) \tag{5.24}$$

from the inverse CDF of the k-th asymmetric Gaussian mixture component given by:

$$\text{CDF}_{\text{AG}}^{-1}(q|\mu,\sigma,r) = \begin{cases} \mu + \sqrt{2}\cdot\sigma\cdot\text{erf}^{-1}(q\cdot(r+1)-1), & q < \frac{1}{r+1} \\ \\ \mu + r\cdot\sqrt{2}\cdot\sigma\cdot\text{erf}^{-1}\left(\frac{q\cdot(r+1)-1}{r}\right), & q \geq \frac{1}{r+1} \end{cases} \tag{5.25}$$

The resulting pulses are analogous to the pulses obtained from IceCube's full simulation chain including photon propagation, trigger and electronics simulation as well as pulse extraction. An example for a cascade event is illustrated in Figure 5.21. In the left panel, an event view is shown for the MC simulation of a cascade just below the dust layer. The same cascade as generated by the Event-Generator is shown on the right.

5.8 Detector Calibration with Event-Generator

The strength of the Event-Generator framework lies in its ability to model high-dimensional PDFs in a flexible and versatile fashion. Due to this property, the framework is well suited for calibration tasks such as the optimization of detector parameters that describe the properties of photon propagation in the glacial ice (Section 3.1 [The IceCube detector]). Calibration of the detector medium is a challenging and computationally expensive task. The ice model consists of many hundreds of free parameters [25, 28, 30], which complicates the search for a global optimum. In addition, the evaluation of a given set of parameters requires large amounts of simulations, rendering a global search nearly impossible. Current calibration approaches in IceCube [31] therefore aim to isolate individual components of the ice model, which are then optimized by performing grid searches in the parameter sub-space.

The Event-Generator framework may be used to reduce the necessary computation time and to facilitate this minimization procedure, potentially enabling a global search. Analogously to the arguments outlined in Section 5.6 [Event Reconstruction], the approach via the Event-Generator may be superior to the direct re-simulation due to the smoothing and interpolation capabilities of the underlying neural network. Individual parts of the parameter phase space may not matter for certain detector responses or their impact may be factored out and described by simple functional relations. The neural networks utilized in the EVENT-GENERATOR framework are able to exploit data properties such as these, which is more challenging to achieve in manual grid scans of the parameter space via direct re-simulations. As a result, the EVENT-GENERATOR requires less simulated data points, and also less photon statistics for each of these points, to appropriately cover the parameter phase space.

In order to demonstrate the potential of the EVENT-GENERATOR for detector calibration, a simplified study is performed here. A cascade with a deposited energy of 10 TeV is injected half a million times in the lower central part of the detector below the dust layer[8]. The parameters of the injected cascade are held constant for each injection, while systematic parameters pertaining to the ice model and detection efficiency of the DOMs are varied. These simulations emulate dedicated flasher runs in which calibrated LEDs on the DOMs are used to inject light in the detector [31]. The recorded detector response of such flasher runs is used to calibrate detector properties. In contrast to the cascade simulations used in this study, dedicated flasher simulations directly inject light at the position of the flashing DOM. While the differing injection locations should not alter the conclusion of this study, fixing the deposited energy to 10 TeV is a simplification that will result in overly optimistic results as the number of emitted photons per LED flash varies and is not well calibrated.

An EVENT-GENERATOR model equivalent to the `cascade-13param` model introduced in Section 5.5 is trained to predict the pulse arrival time PDFs as a function of the systematic parameters. Once trained, this model is employed to estimate the values of the systematic parameters on unseen events that were held out during training. In this approach, the nuisance parameters in the `cascade-13param` model are now used as the parameters of interest. Figure 5.22 demonstrates the capability of the trained model to recover the injected systematic parameters. A strong correlation between true and reconstructed values is visible. The least pronounced correlation is seen for the parameter `HoleIceParam 0`, which — together with `HoleIceParam 1` — varies the photon acceptance of the PMTs as a function of the incident photon angle. Although the varied set of systematic parameters is only a small subset of the tunable ice model parameters, these results are promising and they provide a proof of concept that the EVENT-GENERATOR framework is sensitive enough to accurately model and reconstruct the impact of detector properties.

Aside from facilitating optimization of the current generation of ice models, the EVENT-GENERATOR method is an interesting candidate for future non-parametric or semi-parametric ice models. One of the challenges in defining an accurate model for photon propagation and detector effects is that the current generation of ice

---

[8] This re-uses the simulations performed for the validation of the EVENT-GENERATOR models in Section 5.5.1 [Comparison to Monte Carlo Simulation].

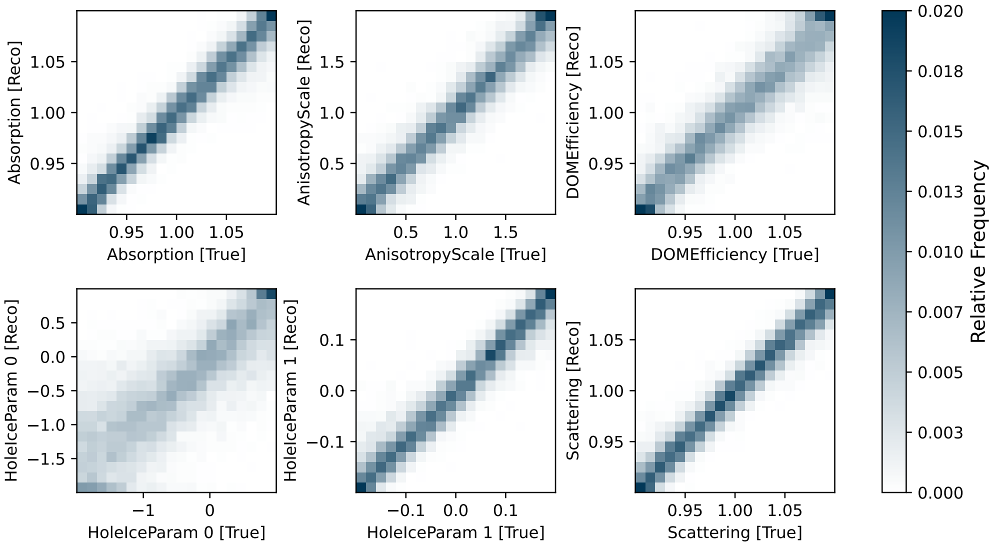


Figure 5.22: Correlation plots between true (x-axis) and reconstructed (y-axis) values are shown for each of the six varied systematic parameters. The model is able to recover the injected parameters as indicated by the strong correlation.

models requires a choice of parameterization. The chosen parameterization must be flexible enough to appropriately model the data, but it should not contain too many free parameters in order to prevent overfitting and to avoid bad convergence properties. A pre-defined parameterization is helpful to ensure interpretability of fitted parameters and to provide insight to the underlying physics such as absorption and scattering coefficients of individual layers of the ice. However, not all effects are required to be explicitly simulated and they are instead treated with effective parameterizations, disregarding the details of the underlying processes. Examples include noise simulation and certain electronic effects. Neural networks provide a powerful non-parametric tool that can be combined with explicit parametric models as demonstrated in the EVENT-GENERATOR approach. These models may be trained directly on experimental data, potentially in combination with simulations.

5.9 Software Framework and Reproducibility

The EVENT-GENERATOR framework follows similar design concepts as those applied in the DNN-RECO software package, outlined in Section 4.1 [Notes on Software Framework]. Among these are the modular software structure and the focus on reproducibility. The software is designed in such a way that settings defining individual sources S_i (Section 5.1 [Event Hypothesis]) are automatically logged and saved. Analogously to the DNN-reco framework, trained Event-Generator models can be exported and disseminated in a secure fashion that ensures their correct usage. Provided `I3Modules` allow users to apply exported models in the IceCube toolchain. Supported applications include event reconstruction, computation of covariance matrices, Markov-Chain-Monte-Carlos, sky scans, visualizations, and model-based simulations.

A core element of the Event-Generator framework is its recursive structure pertaining to the definition of nested event hypotheses via multi-sources M_j, as described in Section 5.1 [Event Hypothesis]. The software is designed to accommodate the construction of multi-sources, thus allowing to combine trained models to more complex event hypotheses without the necessity of re-training. Further details and documentation on the software framework are provided on GitHub[9] [87].

5.10 Applicability to IceCube Upgrade and Gen2

In contrast to the CNN-based approach within the DNN-reco framework (Chapter 4), the Event-Generator approach is easily extensible to future IceCube extensions such as the IceCube Upgrade and IceCube-Gen2, detailed in Section 3.1.1. In order to run the Event-Generator on the IceCube extensions, the generative model G from Eqs. (5.1) and (5.12) must be extended to also output the expected light yield at each of the PMTs of the upgrade modules. The remainder of the Event-Generator framework laid out in the previous sections remains valid.

The generator architecture can be modified by adding separate NNs for each module type (Gen1 DOM, mDOM, D-Egg). This allows to treat modules of the same type in a similar fashion and thus reduce the necessary number of parameters of the NN model. For the modules with multiple PMTs, additional input variables providing information on the PMT angle may be provided. The Event-Generator framework can handle arbitrary detector geometries and module types. It is therefore a suitable reconstruction method for future IceCube extensions.

5.11 Conclusions

Event-Generator is a software framework for event reconstruction that aims to exploit available domain knowledge in IceCube data. Symmetries are embedded in an explicit network architecture, allowing for the reduction of required training data and phase space coverage, as well as enabling extrapolation along these instilled

---

[9] `https://github.com/icecube/event-generator`

symmetries. The Event-Generator is a maximum-likelihood-based method that utilizes concepts from deep learning to model the underlying, high-dimensional PDFs. This construction enables the Event-Generator models to exploit available information and symmetries, while avoiding necessary simplifications required in the standard MLE-based methods in IceCube (Section 3.4). In contrast to the CNN, the Event-Generator model is more sensitive to timing-based information at the cost of increased computational requirements, leading to an improved angular resolution, particularly at lower energies. Event-Generator provides state-of-the-art reconstruction performance and thus constitutes an essential element in the first observation of Galactic neutrinos (Part III).

While the focus in this dissertation is on the reconstruction of cascade-like events, the Event-Generator software is a general likelihood-based framework that allows for the construction of arbitrary event hypotheses. Cascades — single energy depositions in the detector — are elementary components to any combined event hypothesis. Thus, their accurate modeling forms the basis for solving more complex topologies. Due to the work carried out in this dissertation, cascades are well understood within the Event-Generator framework, enabling future work on more complex hypotheses including track events with stochastic energy losses.

Part II

Leveraging Deep Learning for Event Sample Construction

In Part I, the DNN-reco and Event-Generator software frameworks are introduced. These frameworks extend IceCube's analysis toolkit with a rich set of reconstruction and classification methods based on deep learning. In particular, the versatility, speed, and ease of use of the DNN-reco framework opens up unprecedented possibilities for the construction of event samples that were previously unattainable. This enables unique analyses that require the selection of non-typical or challenging event topologies, but it also provides the opportunity to modernize and completely overhaul existing event selections.

In the following chapters, the versatility of the DNN-reco framework is demonstrated for the selection of muons stopping in the detector volume and for a selection of neutrino induced cascade events. Together with the Event-Generator reconstruction, this leads to a state-of-the-art sample of neutrino cascades in Chapter 7 [DNNCascade: A novel Cascade Dataset]. The developed cascade sample, achieved by leveraging the newly developed tools, enables for the first time the observation of high-energy neutrinos from the Galactic plane in Part III.

6 Selection of Stopping Muons

The DNN-RECO framework, introduced in Chapter 4, is a versatile tool that allows for the construction of event samples with arbitrary target event topologies. In combination with the fast inference speed of the employed CNNs, more sophisticated event selections can be applied in earlier stages of IceCube's data analysis pipeline. In this section, this property is demonstrated for the creation of a dataset of atmospheric muons stopping in the detector volume. This dataset can, for instance, be used to measure the depth-dependence of atmospheric muons or to infer properties of the originating air shower. The dataset is also an interesting candidate for detector calibration studies due to the abundance of muons in the minimal ionizing regime, whose properties are well described by existing theoretical predictions.

6.1 Event Selection Chain

The selection of stopping muons in itself is not particularly challenging because they are a common event topology with an event rate of more than 100 Hz. Event classification tools are therefore not necessarily required to have a large efficiency, but they must be computationally inexpensive to run due to the high data rate. Thus, the DNN-RECO framework is utilized with several small CNN architectures.

Three CNNs are trained to perform the classification task as well as the reconstruction of the stopping vertex depth and the zenith angle of the incoming muon. These quantities are essential for the analysis of the muon depth-intensity. The neural networks have the same underlying architecture that only differ in the output layer and applied loss functions (binary cross-entropy for the classification versus a Gaussian likelihood for the reconstruction). The number of layers, convolution kernels and kernel sizes are chosen such that the application requires less than 1 ms per event. Details on the network architecture and training procedure are provided in the supplement Section B.1.1. The reduced data input format consisting of $(c_{\mathrm{total}}, t_{\mathrm{first}}, t_{\mathrm{std}})$ as described in Section 4.3 [DOM-based Input Features] is utilized to enable such a fast inference speed. Training, distribution and deployment of the CNNs is facilitated by the DNN-RECO framework. Due to the existing framework, the neural networks can be setup, trained and deployed within a few days.

6.2 Event Selection Performance

The distribution of classification scores is shown on the left of Figure 6.1 for the MC simulation and experimental data. On the right, the CNN capability is demonstrated to select stopping muons with a high purity. A classification score requirement of ≥ 0.95 results in a sample purity of 98.6% and a selection efficiency of about 60% which corresponds to a rate of 63 events per second.

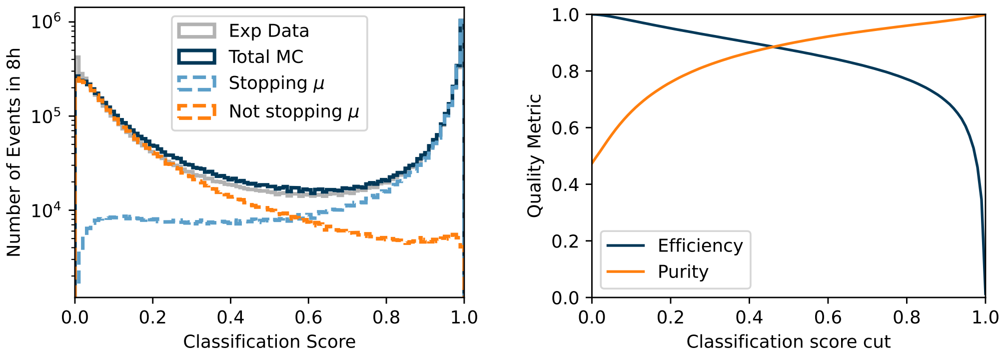


Figure 6.1: The score distribution of the classification CNN is shown on the left. On the right, the sample efficiency and purity is shown as a function of the applied classification score cut.

The reconstruction resolution is shown in the correlation plots in Figure 6.2. Due to the large data rate, strict quality cuts may be imposed by applying cuts on the uncertainty outputs of the CNN. These cuts may be tuned to optimize analysis performance. The shown foreground distribution is obtained by requiring an estimated uncertainty output of less than 20 m on the z-coordinate and less than 2° on the zenith angle. Due to these cuts, the standard deviation of the residuals between true and reconstructed values $\sigma(y_{\mathrm{true}} - y_{\mathrm{reco}})$ is reduced from 78.9 m down to 25.2 m and from 6.5° down to 2.7° for the stopping depth and zenith angle, respectively. The sample purity with these additional quality cuts is 97.7% and the efficiency and corresponding event rate drop down to 0.009% and 9 mHz, respectively. Although this event rate is much lower, the sample will still contain over 300 000 events per year, which is sufficient statistics to perform the analysis.

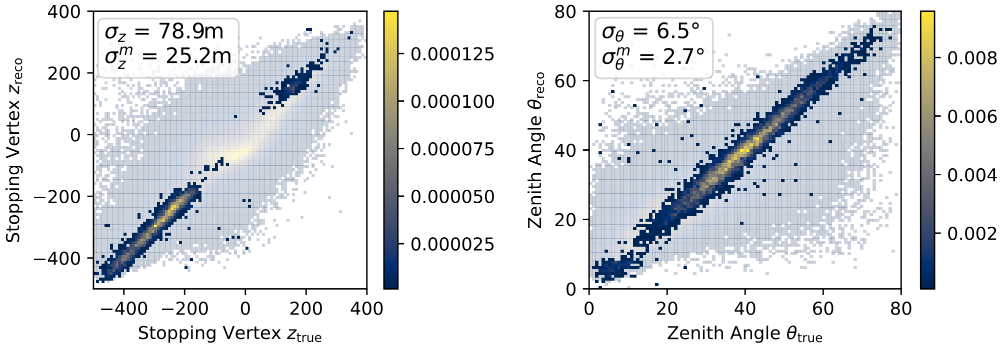


Figure 6.2: Correlation plots are shown for the z-coordinate of the stopping point (left) and the muon zenith angle θ (right). The light distribution in the background shows the correlation for the sample without any further quality cuts. The darker foreground distribution illustrates how the correlation can be improved by utilizing the network's uncertainty estimates. The standard deviation of the residuals between true and reconstructed values are shown in the upper left box.

6.3 Conclusions

Despite providing promising results, the presented dataset of stopping muons is merely a demonstration of how the DNN-reco framework can be utilized to facilitate the construction of new event samples. Additional verification and optimization of the dataset may be necessary in order to employ it for physics analyses. A complete example of how the developed frameworks can be utilized as building blocks in an event selection pipeline are outlined in the Chapter 7.

7 DNNCascade: A novel Cascade Dataset

Declaration

Parts of the dataset description in this Chapter are published in Ref. [1], written by the author of this dissertation and edited by the IceCube collaboration during the paper publication process.

The development of the DNN-reco and Event-Generator frameworks in Part I provide powerful tools that allow to modernize event selections in IceCube. In particular, the DNN-reco reduces the necessity for simplified data reduction cuts in early stages of the event selection pipeline and the Event-Generator method provides unprecedented reconstruction accuracy. Combined, these tools constitute core building blocks in the development of a novel neutrino-induced cascade event selection, referred to as DNNCascade. The DNNCascade sample is aimed at selecting astrophysical neutrinos to enable searches for neutrino emission in the Southern Sky, primarily of the Galactic plane as described in Part III.

7.1 Prospects and Challenges of Cascade Samples

IceCube has two main detection channels for neutrinos consisting of track and cascade events as outlined in Section 3.1 [The IceCube detector]. Track events can be further divided into starting tracks, events with neutrino interaction vertices inside the instrumented volume, and through-going tracks, events in which the neutrino-induced muon is created before entering the detector. Through-going tracks provide superior angular resolution and are thus typically preferred for neutrino source searches [88, 89]. However, event selections of these three event topologies are affected differently by existing backgrounds.

With a data rate of about 2.7 kHz at trigger level [26], the dominant background for neutrino source searches is composed of muons created in Earth's atmosphere by interactions of incoming cosmic rays. The same interactions also produce atmospheric neutrinos [90], which constitute a secondary background to astrophysical neutrinos (neutrinos with extraterrestrial origin). At this stage, background events, atmospheric muons and neutrinos, outnumber signal events, astrophysical neutrinos, by a ratio of roughly 10^8:1. In order to enable neutrino source searches, an event selection is required that enriches the signal purity in the selected subset of events.

Through-going muon samples are most sensitive in the Northern Sky as they use the shielding effect of Earth to reduce the atmospheric muon background. In the Southern Sky, these samples are unable to distinguish atmospheric muons from neutrino induced muons on a per-event basis[1]. Hence, starting events and in particular cascades are required to combat the dominating background of atmospheric muons.

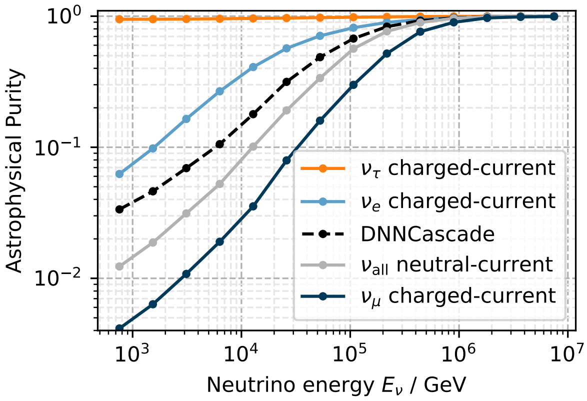


Figure 7.1: The fraction of astrophysical neutrinos is plotted as a function of neutrino energy for different neutrino flavors and interaction types as well as for the DNNCascade sample. Although ν_τ events are well suited for neutrino source searches due to their high astrophysical purity, these events are mostly indistinguishable below several hundred TeV from neutral-current events and charged-current ν_e interactions.

While atmospheric muons can be removed by either using the Earth as a shield or by rejecting events with light from incoming muons, atmospheric neutrinos constitute a mostly irreducible background that grows towards smaller energies, thus inducing a natural lower energy threshold for analyses. Nevertheless, a fraction of the atmospheric neutrinos from the Southern Hemisphere (down-going in IceCube) can be vetoed via the detection of accompanying muons that originate from the same cosmic-ray air-shower that also produced the neutrino [65, 66]. The remaining atmospheric neutrino background is dominated by muon neutrinos, which are largely detected as tracks in IceCube. As shown in Figure 7.1, cascade events, which are produced by neutral-current interactions of all neutrino flavors as well as charged-current interactions of electron and tau neutrinos, are more likely to be of astrophysical origin than tracks produced by charged-current muon neutrino interactions. In comparison to tracks, cascade samples therefore have a reduced contamination of atmospheric neutrinos by about an order of magnitude at TeV energies. Consequently, a lower energy threshold for neutrino source searches of about 1 TeV is achieved.

Despite the above outlined benefits of cascade samples, the selection and reconstruction of these events is challenging. Atmospheric muons can enter the first layers of the detector undetected until eventually depositing light inside the instrumented volume and thus imitate starting events. In particular, if the muon produces a catastrophic stochastic energy loss, these events will be challenging to distinguish from neutrino induced showers. The angular reconstruction of cascades is difficult due to the almost point-like energy depositions in comparison to IceCube's detector

---

[1] Probabilistic statements on the atmospheric or astrophysical origin can be made dependent on the energy and zenith angle of the incoming muon.

spacing. In addition, cascade reconstructions heavily rely on scattered photons and are thus more susceptible to systematic uncertainties of the detector medium.

Nevertheless, the increased astrophysical purity in the Southern Sky and the lower energy threshold of cascades is particularly important for searches of neutrino emission from the Galactic plane. For these searches of large scale structures, the impact of the inferior angular resolution in comparison to tracks is outweighed by the superior energy resolution and purity, as detailed in Chapter 9 [Leveraging Cascade Events]. Cascade events thus provide an important complimentary channel for neutrino astronomy.

7.2 Design and Setup of the Selection Pipeline

Declaration

Training and optimization of the muon and cascade classification boosted decision trees applied in selection step three, as well as tuning of the final analysis level cuts were performed by Stephen Sclafani (Drexel University, PA, USA).

Previous cascade event selections in IceCube rely on simple selection criteria based on event quantities such as the total measured charge in order to reduce the initial data rate [91–93]. In subsequent selection steps, more sophisticated and time consuming selection strategies are performed such as the definition of veto regions within the detector that are used to reject events identified as incoming muons [92, 93].

The developed DNN-reco framework (Chapter 4), built upon convolutional neural networks (CNNs) [13], allows for more efficient event selections. In particular, due to the fast inference speed of the neural networks (NNs), the necessity for simplified, initial data reduction steps is eliminated, thus enabling more complex filtering schemes at earlier stages of the data analysis chain. This allows for a lower energy threshold of the event selection and for the inclusion of more challenging cascade events, which are harder to reconstruct and distinguish from background due to their location at the boundaries of the instrumented volume or in regions of the ice with degraded optical clarity due to higher concentrations of impurities. [1]

The DNNCascade selection heavily relies on CNNs implemented within the DNN-reco framework in order to reduce the background of atmospheric muons and neutrinos by eight orders of magnitude. A staged selection approach, visualized in Figure 7.2, with subsequently more complex models is utilized, allowing for the early rejection of the majority of the background events and hence for the reduction of required computational resources. Multiple CNNs are applied at individual

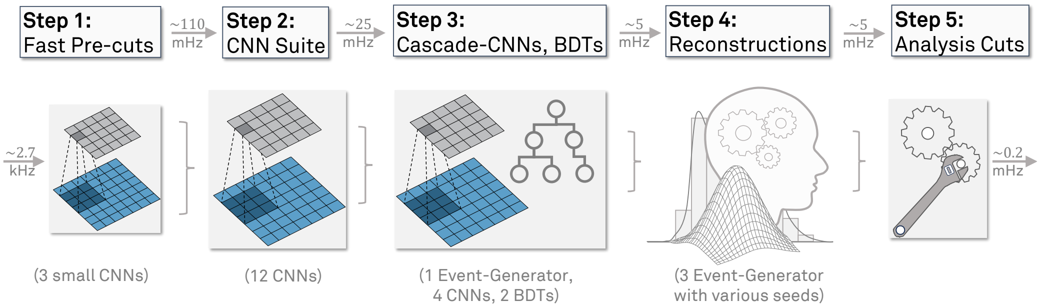


Figure 7.2: An overview of the DNNCascade selection is provided. Starting at all triggered events with an event rate of $\sim 2.7\,\mathrm{kHz}$, a layered selection is performed with the help of CNNs and boosted decision trees (BDTs). Final event reconstructions are performed with the Event-Generator framework (Chapter 5).

selections steps for improved performance. The individual models are trained on slightly different signal and background definitions and subsets of the training data, forcing each of the networks to focus on separate aspects of the over-arching selection goal. In addition, models of subsequent selection steps are trained on the subset of training data that pass the prior selection cuts, which enables the construction of dedicated models for the distinction of more subtle differences in signal and background event topologies that would otherwise remain unnoticed in the overwhelming abundance of less challenging background events.

The target label for the DNNCascade selection, denoted as cascades†, is chosen to align with event topology properties that are detectable given IceCube's resolution. Any neutrino interaction within 150 m of the convex hull around IceCube and a total secondary shower length of less than 100 m is considered as a cascade†. Hence, the target label (cascade†), used during training of the majority of models, differs from the physical definition of cascades. This definition is employed to facilitate the training of models by avoiding unnecessary confusion imposed by an impossible classification task. Targeting an experimental event topology definition such as cascades† also allows to utilize a single class of reconstruction methods. The DNNCascade event selection is designed to maximize sensitivity for neutrino source searches in a disjunct channel from existing track samples [88, 89, 94, 95]; neutrino flavor purity is not the primary goal.

As shown in Figure 7.3, the DNNCascade sample is able to retain a large fraction of cascades† and over 80% of all contained ν_e showers above 20 TeV at the final selection level. The corresponding background reduction for each selection step is demonstrated in Figures 7.4 and 7.5. A summary of the data rates, selection efficiency and purity at each selection step is provided in Tables 7.1, 7.2, and 7.3. More detailed information is provided in Appendix Tables C.2, C.3, and

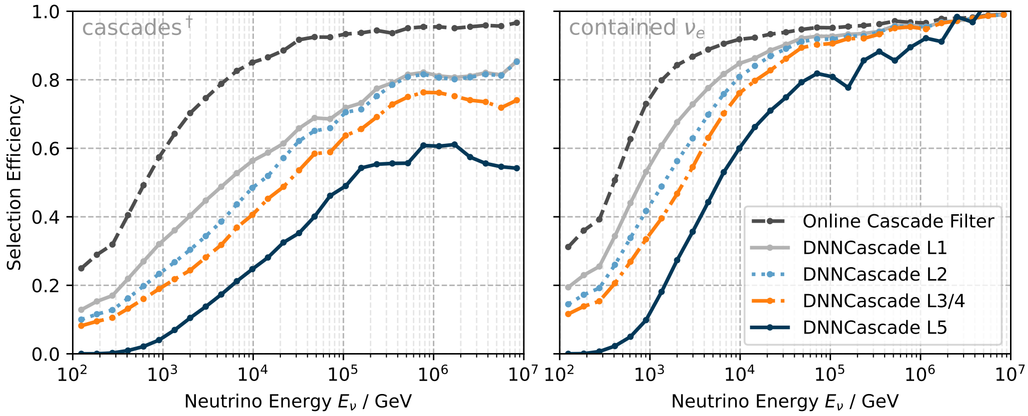


Figure 7.3: The selection efficiency of the DNNCascade sample in retaining cascades† and contained ν_e interactions is shown for each of the selection steps. The efficiency is calculated relative to all triggered events that passed any one of the online filters [26].

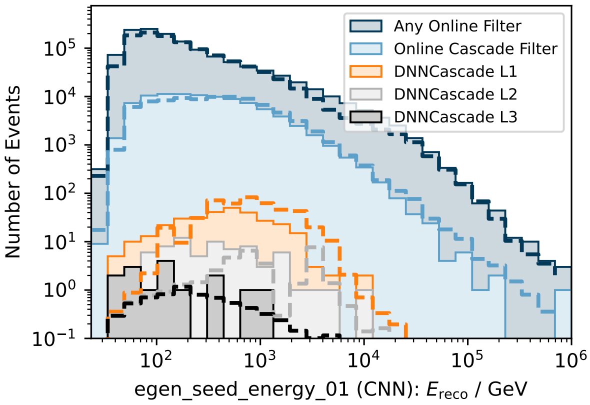


Figure 7.4: The energy distribution of experimental data is compared for different selection steps and a livetime of 47.85 minutes. From 1 274 738 triggered events that passed any of the online filters, only 314, 74, and 15 events remain at selection step one through three, respectively. Expectations from MC simulation are indicated by dashed lines.

C.1. Note that a more conservative background estimate is utilized by combining MuonGun [96] and CORSIKA [97, 98] simulation without accounting for overlap of single muons. In total, the atmospheric background is reduced by almost eight orders of magnitude, while the selection efficiency of neutrino-induced cascades is improved by a factor of 20 in comparison to the precursor event sample [67], as discussed in Section 7.4 [Selection Performance and Dataset Properties]. The computational requirements are provided in Table 7.4. The majority of the time is spent for the final event reconstructions in selection step four. This step may be sped up by reducing the number of seeds utilized in the minimization procedure. Further details on the individual selection steps and applied models are outlined below.

Table 7.1: The data rate of experimental data and MC simulation is shown for each of the DNNCascade selection steps and for multiple subsets of simulation. Triggered events that pass any one of the online filters are denoted by ‘T’; ‘L0’ refers to the Online Cascade Filter.

	Data Rate / mHz					
Data Type	**T**	**L0**	**L1**	**L2**	**L3/4**	**L5**
Atmospheric μ	384633	30977	158	9.356	1.512	0.020
CORSIKA	327786	24873	113	5.238	1.210	0.011
MuonGun	56847	6104	45.0	4.117	0.302	0.009
Neutrinos ν	19.6	6.920	3.235	2.631	2.287	0.167
MC simulation	384653	30984	162	12.0	3.798	0.187
Experimental data	444008	32754	109	25.8	5.225	0.196

Table 7.2: The selection efficiency of experimental data and MC simulation with respect to triggered events that passed any of the online filters (T) is shown for each of the DNNCascade selection steps and for multiple subsets of simulation.

	Efficiency / %					
Data Type	**T**	**L0**	**L1**	**L2**	**L3/4**	**L5**
Atmospheric μ	100	8.1	0.04	2.43e-03	3.93e-04	5.26e-06
CORSIKA	100	7.6	0.03	1.60e-03	3.69e-04	3.37e-06
MuonGun	100	10.7	0.08	7.24e-03	5.31e-04	1.61e-05
Neutrinos ν	100	35.2	16.47	13.40	11.64	0.85
MC simulation	100	8.1	0.04	3.12e-03	9.87e-04	4.86e-05
Experimental data	100	7.4	0.02	5.81e-03	1.18e-03	4.41e-05

Table 7.3: The selection purity based on MC simulation is shown for each of the DNNCascade selection steps and for multiple subsets of simulation.

	Purity / %					
Data Type	**T**	**L0**	**L1**	**L2**	**L3/4**	**L5**
Atmospheric μ	100	100	98	78	40	11
CORSIKA	85	80	70	44	32	5.9
MuonGun	15	20	28	34	7.9	4.9
Neutrinos ν	5.1e-03	2.2e-02	2.0	22	60	89

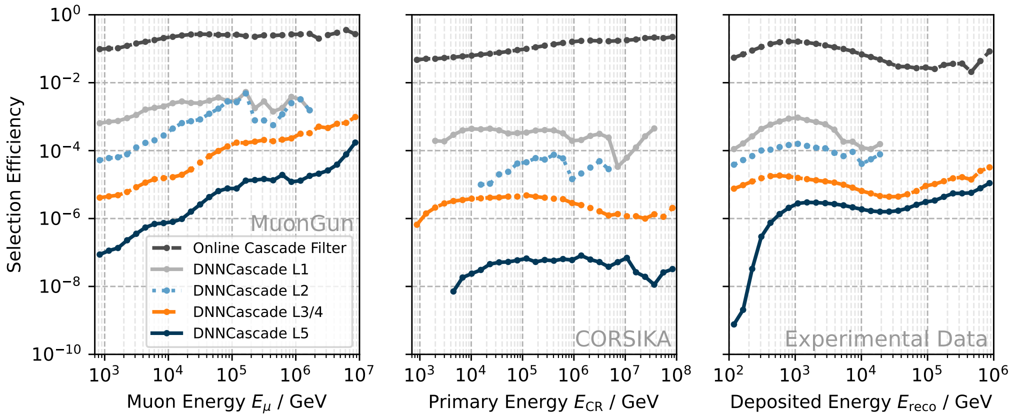


Figure 7.5: The background rejection of atmospheric muon simulation via MuonGun [96] (left) and CORSIKA [97, 98] (middle) is shown as well the reduction in experimental data (right), which is dominated by muons until selection step three.

Table 7.4: The computational requirements of the DNNCascade sample are shown for the individual selection steps if run on GPUs on the IceCube computing cluster (NVIDIA GeForce GTX 980/1080). For the per-event runtime, the 10%, 50%, 90% quantiles are provided in addition to the average value. The total runtime is scaled to the livetime of the full DNNCascade sample of almost ten years. Note that these are only approximate values with large variations possible depending on the employed hardware.

	Per-event Runtime / s				
Level	**10%**	**50%**	**90%**	**Average**	**Total / days**
L1	2.1e-03	2.5e-03	3.5e-03	8.5e-03	4.2
L2	9.2e-02	0.1	0.2	0.1	12
L3	0.1	0.4	1.3	0.6	22
L4	71	123	217	146	101

7.2.1 Step 1: Fast pre-cut Neural Networks

The starting point of the event selection is events passing the ONLINE CASCADE FILTER [26, 91] with an event rate of about 30 Hz. In principle, the DNNCascade selection can be applied on all triggered events, without negatively impacting the selection outcome. In fact, without the required passing condition of the ONLINE CASCADE FILTER, more lower energy events are retained, while the background level remains the same at the final selection level. However, since the passing condition is already computed and because the additional lower energy events are mostly irrelevant for time integrated source searches due to the large atmospheric neutrino background, the ONLINE CASCADE FILTER is a sufficient starting point for the cascade selection that allows to reduce the computational footprint.

After the initial cascade filter cut, background events dominate the sample with about one million events for every astrophysical neutrino event. Three fast and simple CNN architectures are applied to discard the majority of background events and thus to further reduce the data rate for subsequent selection steps. The properties of these networks and the applied score cut are provided in Table B.2. Additional information on the neural network setup and training configuration are provided on GitHub and Zenodo [64] (supplement Section B.1.2). These CNNs require a per-event runtime of about 1 ms on a GPU (NVIDIA GeForce GTX 980/1080) and they are able to further reduce the atmospheric background by about 99.7% while retaining more than half of all signal events above 500 GeV, compared to the ONLINE CASCADE FILTER as detailed in Tables C.2, C.3, and C.1. This reduction of background events by almost three orders of magnitude allows for larger and more complex CNN architectures to be applied in subsequent steps.

7.2.2 Step 2: DNN-reco Reconstruction Suite

After the data rate is reduced down to about 100 mHz in selection step one, a suite of twelve CNNs, detailed in Table B.1, are run for various classification and regression tasks. These models have similar architectures that mainly differ in their training objectives [64]. They are more complex than the precut NNs, requiring a per-event runtime of around ten milliseconds. The CNN classification model EVENT_SELECTION_DNN_CSCD_L3B_CUT2_STARTING_EVENTS_300M_FAST_MEDIUM_01 is utilized to further reduce the atmospheric muon background. Events with a classification score of < 0.90 are rejected, arriving at a data rate of about 20 mHz and a neutrino purity of 20%. The outputs of additional classification networks are utlized as input variables to the gradient boosted decision trees (BDTs) defined in selection step three. Further CNNs are run to obtain initial reconstructions for event properties such as direction, interaction vertex, and deposited energy. These reconstruction results will be used as seed values for the EVENT-GENERATOR methods in subsequent steps as well as input variables to the BDTs.

7.2.3 Step 3: Cascade-based CNNs and Boosted Decision Trees

The third selection step applies a fast EVENT-GENERATOR reconstruction based on the CASCADE-7PARAM model by choosing a high-tolerance for the minimizer. The initial seed hypothesis for the minimizer is constructed from the outputs of the EGEN_SEED_* CNN models run in the prior selection step. This reconstruction takes about 1 s per event depending on the amount of observed light and it is used to

Table 7.5: The utilized input features and their importance are shown for the μ-BDT. The name prefixes event__selection__egen__[1] and event__selection__dnn__cscd__[2] are omitted for brevity.

Feature	**Importance**
vertex__starting__events__300m__fast__medium__01[1]	0.5626
vertex__starting__nue__300M__fast__medium__01[1]	0.3325
l3c__track__numu__cc__vs__nue__cc__01[2]	0.0186
l3b__cut2__starting__events__300m__fast__medium__01[2]	0.0144
(fast) cascade-7param energy	0.0143
vertex__track__numu__cc__vs__starting__01[1]	0.0132
vertex__track__numu__cc__vs__nue__cc__01[1]	0.0113
l3a__starting__cascades__150m__red__summary__stats__fast__02[2]	0.0096
l3c__cut2__starting__events__300m__fast__medium__01[2]	0.0094
l3c__track__numu__cc__vs__starting__01[2]	0.0088
(fast) cascade-7param zenith	0.0052

obtain an improved estimate for the cascade vertex, which is employed by cascade-based NNs applied afterwards. These networks, detailed in Table B.3, utilize the estimated cascade vertex in order to compute input variables that are more sensitive to incoming and outgoing particles as described in Section 4.3 [DOM-based Input Features]. Once the cascade-based NNs have been applied, two gradient boosted decision trees (BDTs) [99] are run. The first BDT, denoted as μ-BDT, classifies and scores events as atmospheric muons or neutrinos. The resulting "muon-score" is a representation of the muon-ness of an event. The second BDT, referred to as cascade-BDT, is trained to classify between charged-current muon-neutrinos and neutrinos of all other flavors and interaction types. This "cascade-score" corresponds to the likeliness of an event being a cascade.

The BDTs obtain the outputs of the previously run CNNs and Event-Generator reconstructions as input features. Tables 7.5 and 7.6 provide information on the chosen features and their importance. For both BDTs, the vertex-based models are most important to the classification task, demonstrating that a careful and adequate choice of input variables to the NNs is essential to maximize performance. A loose selection criteria of μ-BDT < 0.1 (< 0.01) is chosen to further reduce the atmospheric muon background by over 80% (98%). Stricter cuts, optimized to a specific analysis, are applied later in the final selection step five.

Table 7.6: The utilized input features and their importance are shown for the cascade-BDT. The name prefixes EVENT_SELECTION_EGEN_[1] and EVENT_SELECTION_DNN_CSCD_[2] are omitted for brevity.

Feature	**Importance**
VERTEX_TRACK_NUMU_CC_VS_STARTING_01[1]	0.7494
VERTEX_TRACK_NUMU_CC_VS_NUE_CC_01[1]	0.1255
L3C_TRACK_NUMU_CC_VS_STARTING_01[2]	0.0535
L3C_TRACK_NUMU_CC_VS_NUE_CC_01[2]	0.0199
(fast) CASCADE-7PARAM energy	0.0162
VERTEX_STARTING_NUE_300M_FAST_MEDIUM_01[1]	0.0078
(fast) CASCADE-7PARAM zenith	0.0075
L3B_CUT2_STARTING_EVENTS_300M_FAST_MEDIUM_01[2]	0.0059
L3C_CUT2_STARTING_EVENTS_300M_FAST_MEDIUM_01[2]	0.0050
L3A_STARTING_CASCADES_150M_RED_SUMMARY_STATS_FAST_02[2]	0.0049
VERTEX_STARTING_EVENTS_300M_FAST_MEDIUM_01[1]	0.0045

7.2.4 Step 4: Final Event Reconstructions

At this point, with a data rate of 5 mHz, more computing intensive EventGenerator reconstructions with multiple seeds are run[2]. Three different models, defined in Sections 5.5 and 5.6, for a single cascade, two correlated cascades, and two independent cascades are applied. In total, these reconstructions require a per-event runtime of about two minutes due to the chosen number of minimizer seeds. Afterwards, three NNs are run to obtain an estimate on the angular uncertainty for each of the reconstructed event hypotheses. Based on the estimated uncertainties and the likelihood of each of the event hypotheses, the best reconstruction result is chosen with the corresponding uncertainty estimate, referred to as $\hat{\sigma}$-NN. Further details on the event reconstruction and uncertainty estimation are provided in Section 7.3. No additional cuts are applied in this selection step. Events remaining at this step, together with the applied reconstructions, are referred to as DNNCascade Level 4. This level is intended as a general starting point for dedicated selection criteria that optimize specific analyses in selection step five.

7.2.5 Step 5: Analysis Optimized Selection Criteria

In contrast to prior selection steps, selection step five is intended to be specific to a given analysis. Available BDT classification scores and event reconstructions in

---

[2] The final data processing performed for the DNNCascade sample located on the IceCube clusters at /data/ana/PointSource/DNNCascade/ applied a stricter cut of μ-BDT < 0.01 prior to running DNNCascade L4 in order to reduce required disk space and computational requirements by 50%.

DNNCascade Level 4 enable analyzers to fine-tune high-level selection criteria to maximize performance of a given analysis. The selection step five, outlined in this disseration and referred to as DNNCascade Level 5, is therefore specific to the analysis described in Part III, aimed to maximize sensitivity and discovery potential to neutrino emission from the Galactic plane.

For neutrino source searches, an event sample must provide excellent event reconstruction performance. Higher backgrounds and atmospheric muon contamination can be tolerated in comparison to diffuse flux measurements [91, 92, 95, 100, 101], leading to more lenient selection criteria as outlined in Table 7.7. The applied BDT score cuts result in an atmospheric muon contamination of about 5 to 10%, while the astrophysical neutrino contribution is estimated to around 7%, assuming the flux given in Ref [91].

Table 7.7: The selection criteria to arrive at the Galactic plane optimized DNNCascade Level 5 are presented.

Passing Criteria	**Details**
μ-BDT < 0.005	Reject atmospheric muons
cascade-BDT > 0.1	Reject starting tracks
$E_{\mathrm{dep}} > 500\,\mathrm{GeV}$	Remove low energy events
$\hat{\sigma}$-NN $< 40^\circ$	Require minimal reconstruction quality
$-700\,\mathrm{m} < z < 600\,\mathrm{m}$	Remove mis-reconstrcuted events
$r = \sqrt{x^2 + y^2} < 700\,\mathrm{m}$	Remove mis-reconstrcuted events

The cuts on the estimated angular resolution $\hat{\sigma}$-NN and the vertex position aim to maintain a minimal reconstruction quality by limiting the distance a cascade is allowed to be from the instrumented volume. Events reconstructed with vertices far outside of the detector boundaries may also be an indication of failed reconstructions with bad local minima. Such events are therefore removed from the data sample. A minimal energy is imposed in order to reduce atmospheric backgrounds and a cut on the cascade-BDT is applied to reduce overlap to existing track samples. Finally, an additional correction, described in Section 7.3, is applied to the estimated angular uncertainties to account for potential detector systematics. At this stage, the atmospheric muon background is reduced by almost eight orders of magnitude and well reconstructed neutrinos now dominate the sample, enabling the search for neutrino emission.

7.2.6 Notes on Software Availability

The DNNCascade selection software is available on GitHub[3]. Selection steps one through three are grouped together in a single `I3Module` and referred to as DNNCascade Level 3. DNNCascade Level 4 and Level 5 align with the selection steps four and five and are provided in separete `I3Modules`. DNNCascade Level 4 is intended as a basis for analyzers to fine-tune the DNNCascade selection to their specific needs. Hence, the analysis level cuts applied in DNNCascade Level 5 will be unique to a certain analysis for which these cuts are optimized.

7.3 Event Reconstruction and Uncertainty Estimation

Declaration

This section follows the dataset description from Ref. [1], written by the author of this dissertation. Parts of this section are adopted verbatim as indicated by quotation marks.

> “Due to the DOM’s largely linear response to light intensity, any event hypothesis in IceCube can be deconstructed as a superposition of individual energy losses. The Event-Generator framework makes use of this fact: only a single neural network needs to be trained to perform this elementary mapping. Complex event hypotheses may then be constructed from the superposition of multiple inference steps of this elementary NN, as detailed in Chapter 5 [Event-Generator: Combining Maximum-Likelihood and Deep Learning].
>
> Since this analysis is focused on cascade events, most events are well modeled by a single energy deposition. However, fluctuations in energy depositions along the particle shower as well as incoming/outgoing particles and coincident events may require a more complex event hypothesis. To this purpose, three event hypotheses, defined in Sections 5.5 and 5.6, are applied: a single energy deposition, two causally connected energy depositions, and two independent energy depositions. Of these three reconstructed hypotheses, the reconstruction result from the model hypothesis with the lowest estimated angular uncertainty and highest likelihood is chosen on a per-event basis.”[1]

The resulting angular resolution at DNNCascade Level 5 is demonstrated in Figure 7.6. An improvement in angular resolution over the entire energy range is

---

[3] `https://github.com/icecube/dnn_cascade_selection` (Access rights to IceCube GitHub organization are required)

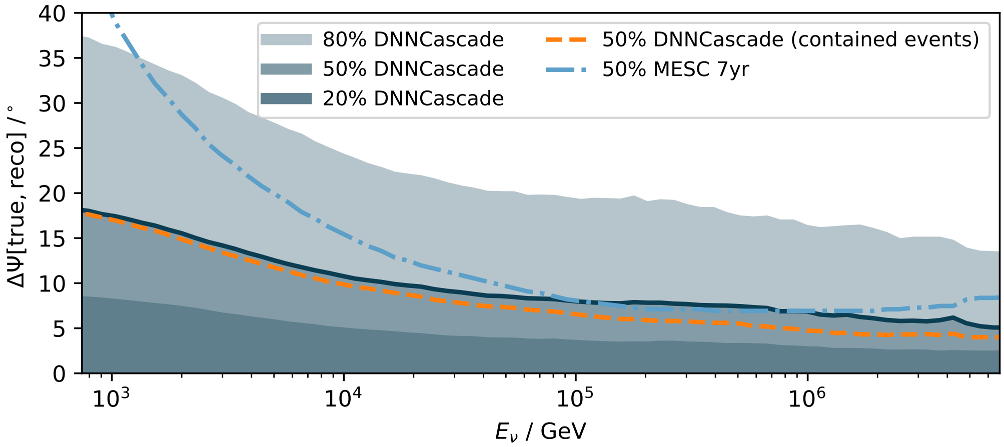


Figure 7.6: The angular resolution as a function of true neutrino energy is shown on simulated events for the DNNCascade sample and the previous cascade selection, MESC 7yr [67]. An additional curve in orange demonstrates the angular resolution of contained events. Systematic uncertainties are not included in these curves. Figure and caption adopted from Ref [1].

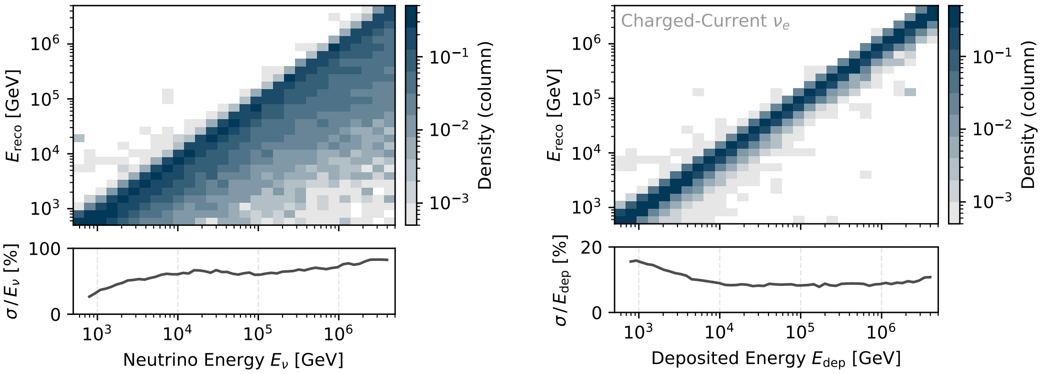


Figure 7.7: The energy resolution on simulated events is shown for the neutrino energy (left) and for the deposited energy of charged-current electron neutrinos (right). The reconstruction method aims to infer the true deposited energy (E_{dep}) as shown on the right, which is a lower bound to the true neutrino energy (E_ν) [11]. The top panel illustrates the correlation between reconstructed and true quantities, and the bottom one shows the relative resolution, defined here as the 68% quantile of the absolute values of the relative residuals $|(E_{\mathrm{true}} - E_{\mathrm{reco}})/E_{\mathrm{true}}|$. Figure and caption adopted from Ref [1].

observed for the DNNCascade sample in comparison to the precursor cascade selection, despite a higher fraction of more challenging events. This is accomplished by the inclusion of domain knowledge via the novel hybrid reconstruction framework Event-Generator [5] introduced in Chapter 5, which results in a better exploitation of information than the DNN-reco-based CNN [3, 6] used in the previous cascade selection. The most significant gains are achieved at energies below a few tens of TeV, reaching up to more than a factor of two at 1 TeV. As discussed in Part III, the majority of the Galactic plane signal is observed in this energy region, emphasizing the value of the improved methods. The energy resolution of the sample is comparable to previous cascade selections. It is illustrated in Figure 7.7 as a function of neutrino energy and deposited energy. A vertex resolution of 6 m is achieved at 1 TeV, which improves to 4 m at 10 TeV and then flattens out to about 3.5 m at higher energies. [1]

> “The estimated per-event angular uncertainty for each of these reconstructed event hypotheses is obtained by three independent, fully-connected NNs. These NNs are trained via a von Mises-Fisher likelihood [82] to estimate the circularized angular uncertainty on each of these three reconstructions using high-level features such as the reconstructed event properties, the delta log-likelihood values between the hypotheses, and the second derivatives of the likelihood. As discussed in Section 5.6.2 [Uncertainty Estimation], this method provides better coverage than an uncertainty estimator based on the Hessian of the likelihood evaluated at the best fit position. More sophisticated scans or sampling of the high-dimensional likelihood landscape were not performed due to computational constraints. A circularized angular uncertainty estimate is used, symmetrical in right ascension and declination. [This assumption simplifies the analysis tools used for source searches in Part III, but it is not entirely correct as cascades, in particular, can exhibit asymmetric uncertainty contours, often elongated in right ascension due to the detector layout. See also supplement Chapter A [Cascade Real-Time Alert Stream] and Section 5.6.2 for further details. A more complete description of this second order effect in future work may further improve the sensitivity and discovery potential of the analyses presented in Part III.]” [1]

> “Systematic uncertainties on detector properties have an energy-dependent impact on the angular resolution, further degrading the resolution shown in Figure 7.6 by about 5% at 1 TeV up to about 25% at PeV energies. These systematic uncertainties encompass properties of the detector medium including absorption and scattering coefficients, ice anisotropy, and the acceptance parameterization of the refrozen ice column surrounding the DOMs, which has increased scattering and absorption properties due to enclosed air bubbles and dust impurities, as well as the DOM quantum efficiency (Section 3.1). To account for these systematic effects, the

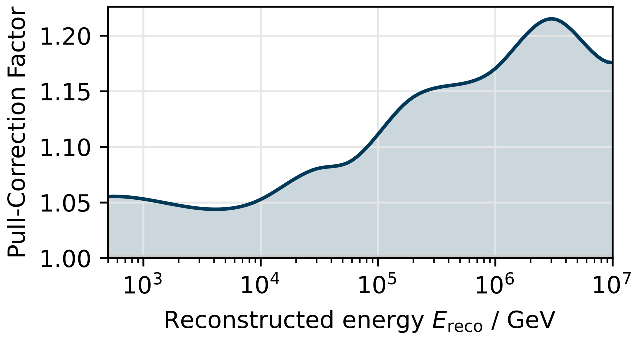


Figure 7.8: An additional correction factor, shown here as a function of reconstructed energy, is applied to the estimated angular uncertainty, $\hat{\sigma}$-NN, to account for systematic uncertainties. The resulting, corrected uncertainty estimate is used in the final analysis and it is referred to as σ.

> estimated angular uncertainty is further corrected by an energy-dependent scaling factor, as shown in Figure 7.8, that is determined from a systematic dataset that continuously samples from the estimated systematic uncertainties [74]. The coverage of the resulting uncertainty estimate is compared in Figure 7.9 for the baseline MC simulation and the simulation with varied systematic parameters. On the baseline simulation, the uncertainty estimate is conservatively over-covering by about 15% due to the applied scaling factor. This over-coverage is reduced when introducing systematic uncertainties on the detector properties as shown on the right of Figure 7.9. In both cases, the coverage does not strongly depend on the neutrino energy. ” [1]

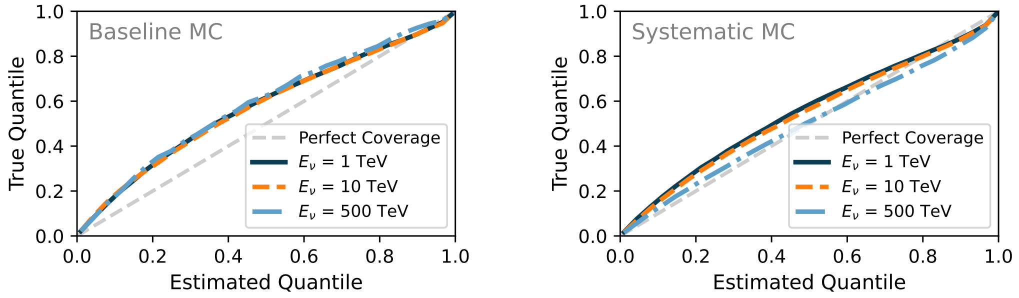


Figure 7.9: The coverage of the angular uncertainty estimator is evaluated for different neutrino energies on the baseline MC simulation (left) and on a simulation with varied systematic parameters (right). For an accurate uncertainty estimate, the estimated fraction of events for which the true direction lies within a certain quantile (x-axis) should match the true frequency at which the true direction falls into that quantile (y-axis). Perfect coverage is obtained along the diagonal indicated by the gray, dashed line. Curves that lie above this diagonal over-cover the true uncertainty, i.e. the estimated uncertainty is larger than the true uncertainty of the reconstruction method. Figure and caption taken from Ref [1].

7.4 Selection Performance and Dataset Properties

The newly developed DNNCascade sample provides a much improved selection efficiency as demonstrated in Figure 7.10. In comparison to the precursor cascade event selection [67], the effective area is improved by an order of magnitude, resulting in about ten times as many selected astrophysical neutrinos assuming an isotropic astrophysical flux according to the cascade measurement in Ref. [91]. In particular in the Southern Sky (down-going in IceCube), the efficiency of the DNNCascade sample is unmatched and therefore crucial to the success of studies on neutrinos from the Galactic plane. This improvement arises from the deployment of novel selection techniques with improved performance and reduced computational requirements, enabling the application of more advanced selection strategies on larger data rates, as discussed in Section 7.2 [Design and Setup of the Selection Pipeline]. As a result, the energy threshold of the event selection is reduced to about 500 GeV compared to several TeV in earlier work.

> A significant contribution to the increase in efficiency relative to the previous selection [67] is the inclusion of events with interaction vertices near the boundary or outside of the instrumented volume as illustrated in Figure 7.21. These "partially contained" and "uncontained" events are more difficult to distinguish from background events and their reconstruction is challenging due to only observing a fraction of the deposited energy. Thus, these events were removed in most previous IceCube analyses. The application of deep learning based tools in this analysis allows it to retain a significant fraction of these and other more challenging events, while ensuring an acceptable reconstruction quality and level of background contamination. An example for an "uncontained" and "contained" cascade event is shown in Figure 7.20. About 17.5% of all events in the sample have a reconstructed interaction vertex outside of the instrumented volume and are thus considered as uncontained events. An additional 2.5% of events are located in a region with increased dust impurities. Events such as these are often excluded from analyses due to the poor reconstruction quality. [1]

As shown by the median opening angle in the first three panels of Figure 7.11, the angular resolution does degrade for these events. The impact of the dust layer on the angular resolution is particulary visible in Panels 7.11b and 7.11c, while Panel 7.11a demonstrates the degraded resolution for events far outside of the instrumented volume. However, the estimated angular uncertainty σ is also larger for these events and thus accurately models the degraded resolution. This is visible in the last three panels of Figure 7.11, which illustrate the median opening angle as a function of estimated uncertainty versus energy, distance to the convex hull around IceCube as well as the z-coordinate of the vertex position. In each of these cases, the median opening angle $\Delta\Psi_{50\%}$ is mostly homogeneous along the y-axis and thus well

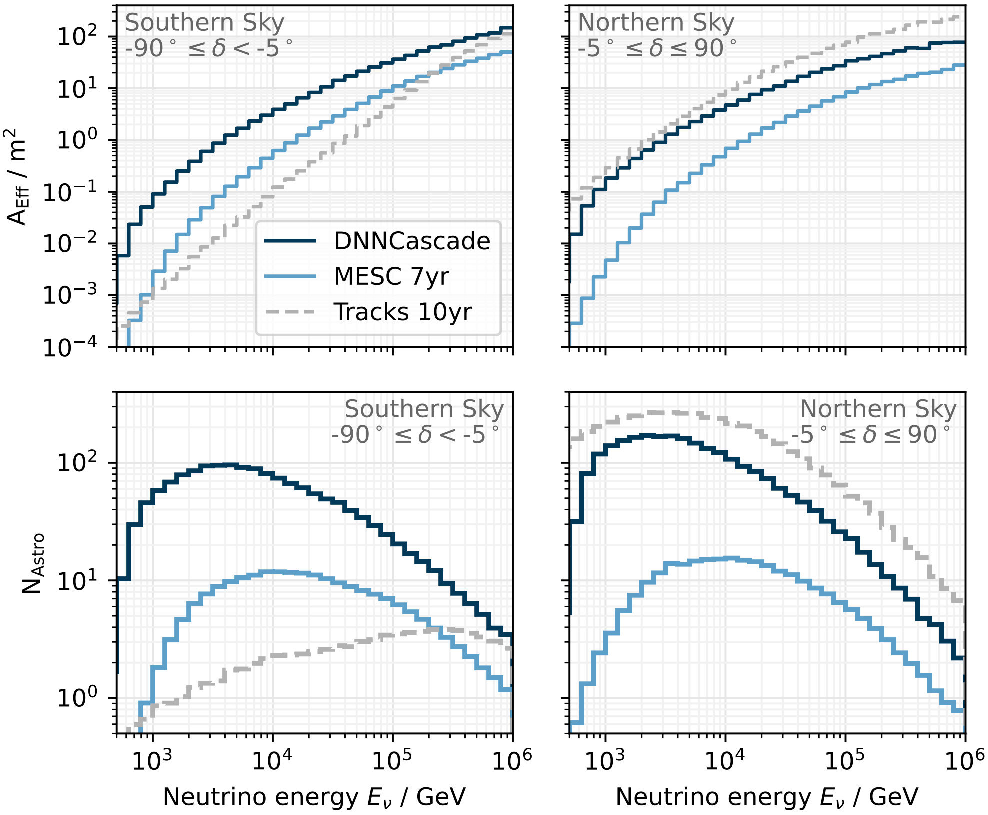
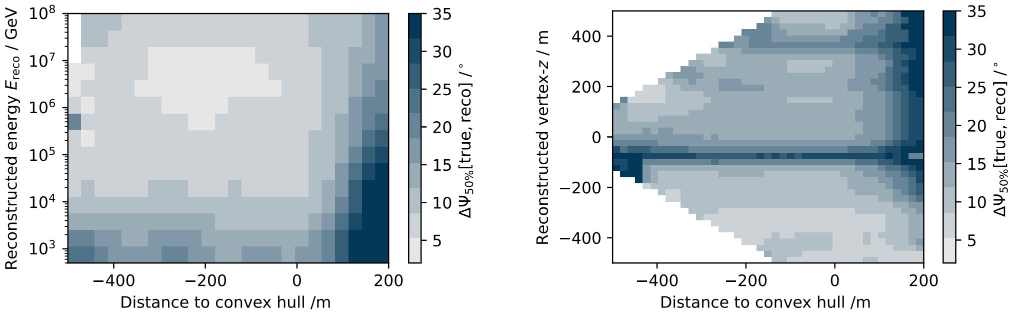
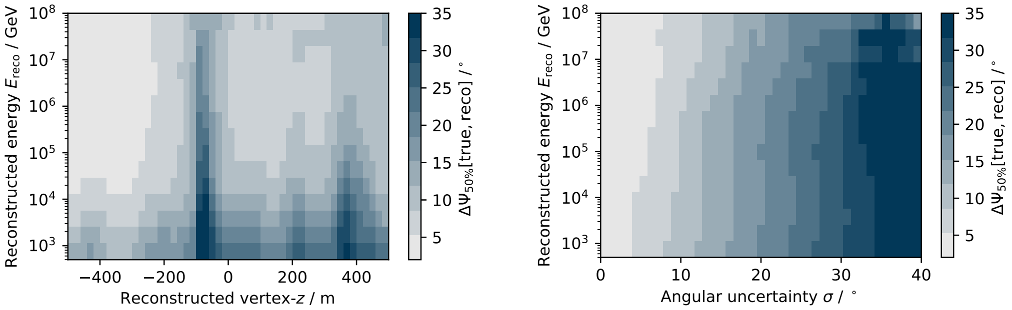


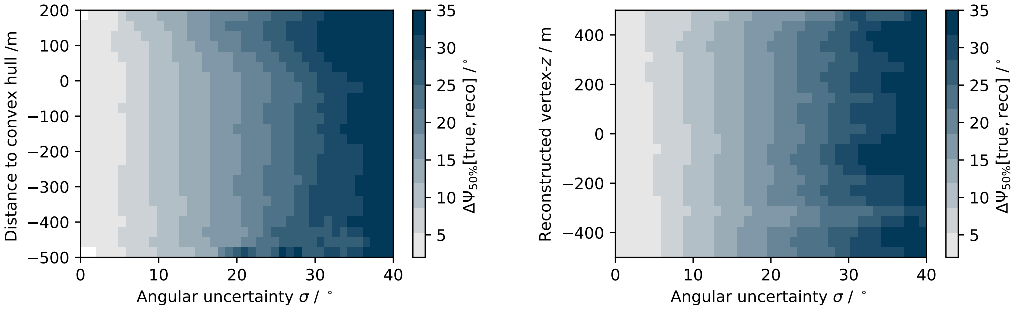
Figure 7.10: The all-flavor neutrino effective area of the DNNCascade sample is compared to the precursor cascade event selection (MESC 7yr) [67] and the track event selection (Tracks 10yr) [88] in the upper panel. The effective area is averaged over solid angle in the declination range between −90° and −5° (left) and between −5° and 90° (right) for the Southern and Northern Sky, respectively. The bottom panel shows the number of signal events in the Southern (left) and Northern (right) Sky per energy bin for each event selection scaled to the livetime of the DNNCascade sample (9.64 years). An isotropic astrophysical flux according to the cascade measurement in Ref. [91] is assumed in addition to equal contributions of each flavor at Earth due to neutrino oscillations.



(a) Distance vs. energy. **(b)** Distance vs. vertex-z.



(c) Vertex-z vs. energy. **(d)** Uncertainty vs. energy.



(e) Uncertainty vs. distance. **(f)** Uncertainty vs. vertex-z.

Figure 7.11: The median opening angle, $\Delta\Psi_{50\%}$, between true and reconstructed direction is shown for different two-dimensional projections. Darker colors correspond to smaller opening angles and thus better reconstructions.

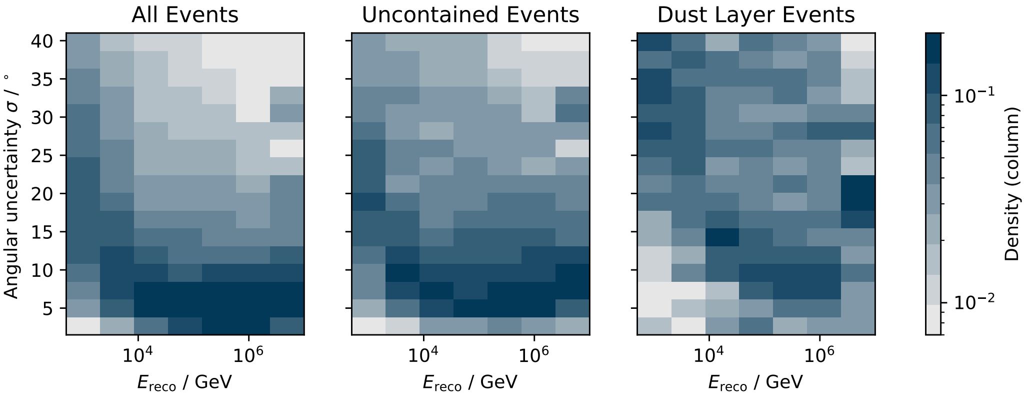


Figure 7.12: The column-normalized distribution of estimated angular uncertainty σ is shown as a function of reconstructed energy. These are shown for three categories of events: entire sample (left), only "uncontained" events (middle), and only events in the dust layer (right).

described by the estimated uncertainty. Further evidence for the correct modeling is demonstrated by the distribution of estimated angular uncertainty for different classes of events in Figure 7.12. As expected, the estimated uncertainty increases for dust layer and uncontained events. Overall, the estimated angular uncertainty provides good coverage on the ensemble of events as demonstrated in Figure 7.9. Thus, not only is the sample able to retain these more challenging events, but it is also able to appropriately model their degraded reconstruction resolution.

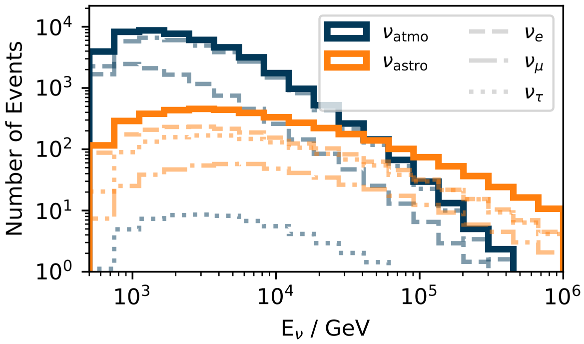
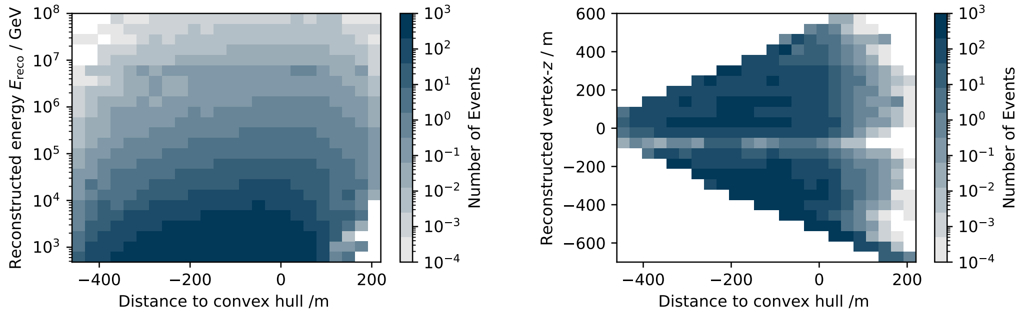


Figure 7.13: The neutrino flavor composition of the DNNCascade sample based on simulation is illustrated as a function of neutrino energy. Contributions to each neutrino flavor are shown in addition to the total number of atmospheric (solid, dark blue) and astrophysical (solid, orange) events.

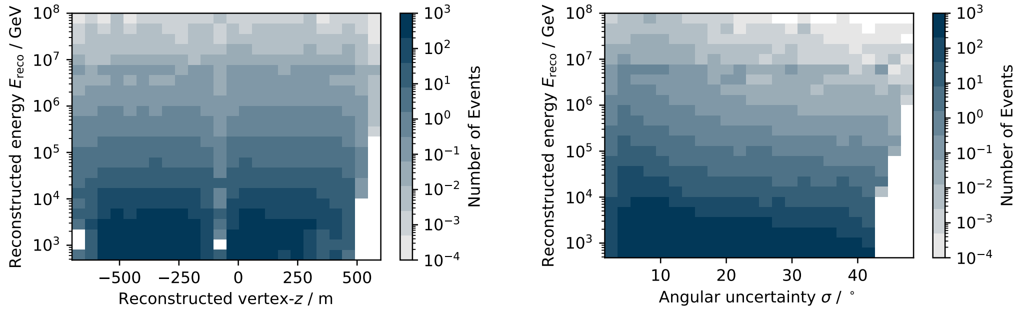
The overall dataset composition separated out into the three neutrino flavors is depicted in Figure 7.13. Although the cascade selection discards a large amount of neutrino-induced muons from charged-current ν_μ interactions, the atmospheric contribution to the cascade sample is still dominated by ν_μ. This arises from the disproportional amount of ν_μ produced in Earth's atmosphere compared to ν_e and ν_τ, see also Section 7.1 [Prospects and Challenges of Cascade Samples] for more details. Neutrinos of extraterrestrial origin become dominant above several tens of TeV, depending on the assumed astrophysical spectrum. These astrophysical neutrinos consist mostly of ν_e and ν_τ.

The distribution of selected events is studied for various two dimensional projections in Figure 7.14. Panel 7.14a shows that with increasing energy, more events are selected further outside of the instrument volume. Lower energy events do not inject enough light into the detector for accurate event reconstruction and background rejection and are thus discarded from the sample. Projections including the z-coordinate of the interaction vertex highlight the effect of the dust layer as a large dip in number of selected events at coordinates from approximately $-150\,\mathrm{m}$ to $0\,\mathrm{m}$. The majority of events are selected in the deep ice, which has improved optical clarity and thus optimal properties for event reconstruction.



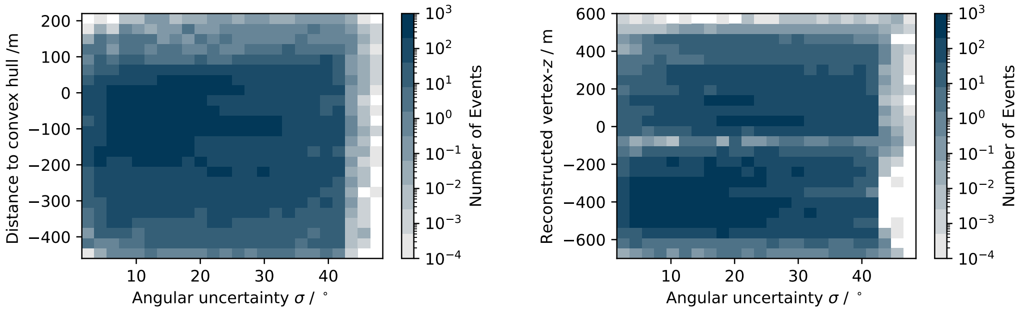
(a) Distance vs. energy.

(b) Distance vs. vertex-z.



(c) Vertex-z vs. energy.

(d) Uncertainty vs. energy.



(e) Uncertainty vs. distance.

(f) Uncertainty vs. vertex-z.

Figure 7.14: The expected number of events in the DNNCascade sample is shown for bins in different two-dimensional projections.

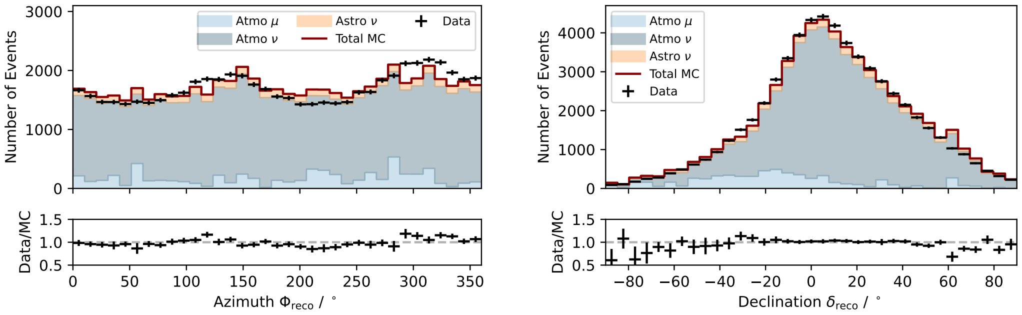
7.5 Data/MC Agreement

Previous cascade event selections are primarily designed for the study of the diffuse neutrino spectrum. For the success of these types of analyses, it is crucial that the experimental data is well described by the MC simulation as background estimates and analysis results are derived from and directly depend on the simulation. While the signal region is typically well modeled by simulation, accurate description of the atmospheric muon background is inherently difficult. This arises from the lack of sufficient statistics of simulated air showers at final event selection level, but also from large systematic uncertainties on the muon flux that contribute to the so-called "muon puzzle" [102]. Previous analyses therefore conservatively discard events from regions in which atmospheric muons may play a larger role.

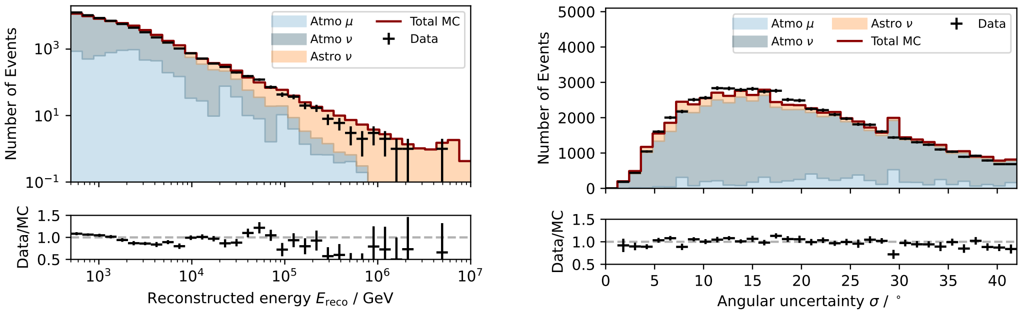
In contrast, the DNNCascade sample is designed and optimized for searches of neutrino emission. These searches typically employ data-driven methods in which the background is modeled from scrambled experimental data as described in Chapter 10 [Analysis Methods]. Simulation of backgrounds and their accurate modeling is therefore not strictly required. However, incorrect modeling will result in loss of analysis sensitivity due to inclusion of unmodelled and potentially avoidable backgrounds. Disagreements between MC simulation and experimental data may also be indications of systematic effects that could invalidate reconstruction results and their corresponding uncertainty estimates, which are both based on simulation. Due to these reasons, the agreement of experimental data with expectations obtained from MC simulation is explored in this section.

In the following comparison plots, astrophysical neutrino contributions assume an isotropic astrophysical flux according to Ref. [91] and an equal flavor ratio at Earth. The contributions of atmospheric neutrinos are obtained from the cascade-equation solver MCEq [103], the Hillas model (H3a) for cosmic rays [104, 105] and the SIBYLL 2.3c hadronic interaction model [106]. Self-Veto [65, 66] passing fractions are calculated with NuVeto [107]. Atmospheric muon background is simulated by the simulation packages CORSIKA [97, 98] and MuonGun [96]. These are combined without taking overlap into considerations. As MuonGun is only used to simulate single muons at detector entry, the naïve combination of these two simulation sets will lead to a conservative double counting of these type of background events. Systematic uncertainties on the assumed fluxes are not taken into account and only the baseline values are utilized.

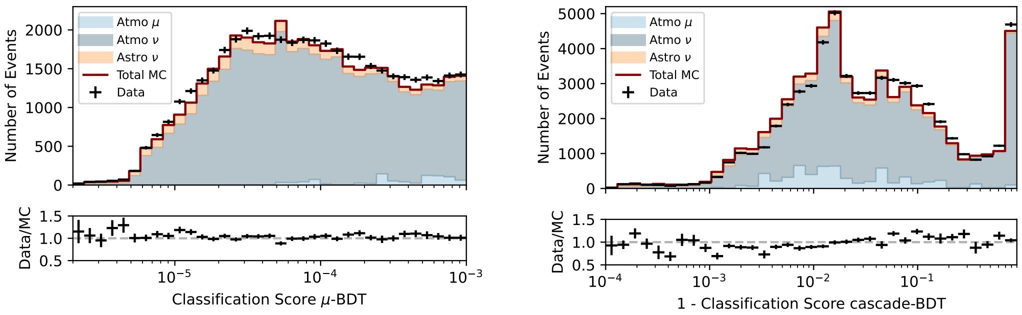
Overall, data agrees well with predictions from MC simulation for key analysis variables of the DNNCascade selection, as demonstrated in Figures 7.15 and 7.16 as well as in the supplement Section C.2 [Data/MC Agreement]. Good agreement is also obtained for the fraction of uncontained events as illustrated in Figure 7.22. Some discrepancies are visible in the energy spectrum (Panel 7.15c), azimuth (Panel 7.15a) and vertex-z (Panel 7.16c) distributions. Differences in the energy



(a) Azimuth. (b) Declination.

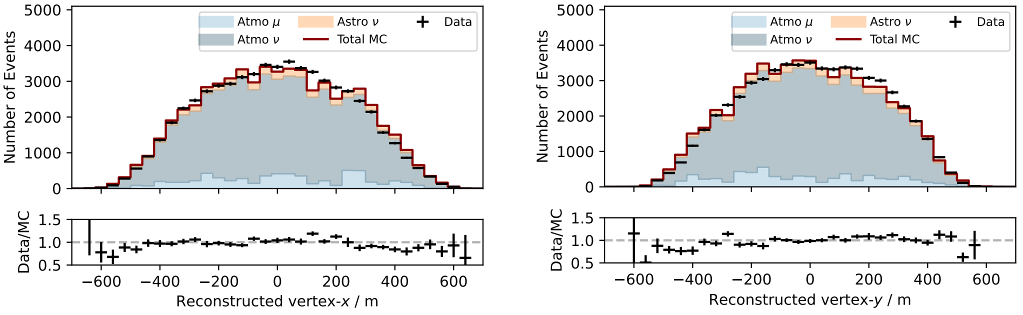


(c) Energy. (d) Uncertainty.



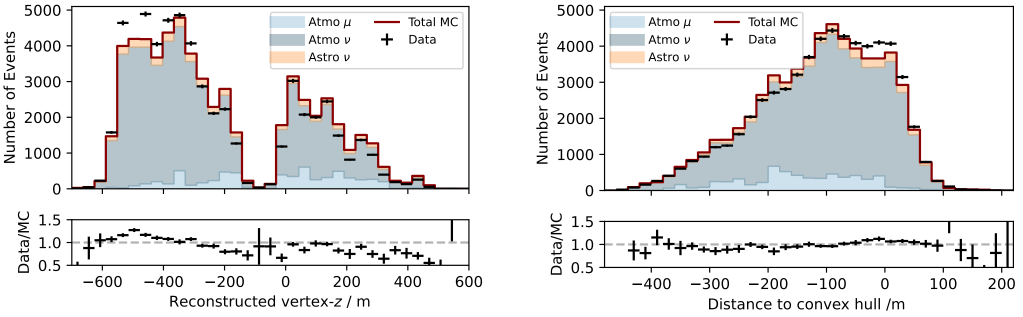
(e) μ-BDT. (f) cascade-BDT.

Figure 7.15: The experimental data distributions in key analysis observables are compared to expectations obtained from MC simulations. Systematic uncertainties are not included.



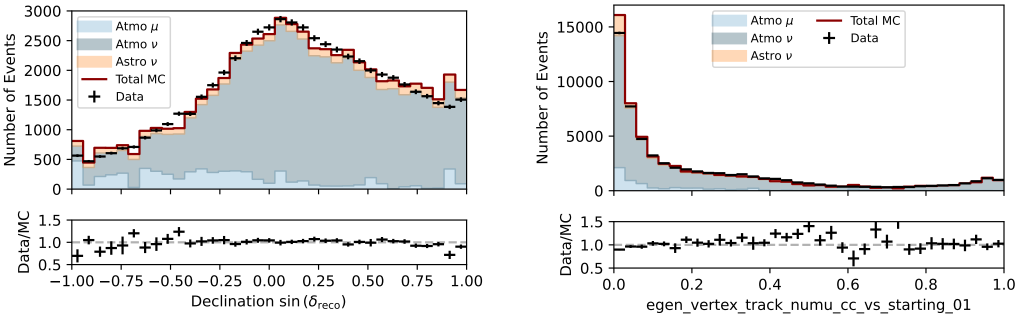
(a) Vertex-x.

(b) Vertex-y.



(c) Vertex-z.

(d) Distance.



(e) Sine Declination.

(f) Vertex-based CNN.

Figure 7.16: Experimental data distributions are compared to expectations obtained from MC simulations. Systematic uncertainties are not included.

spectrum are expected due to large uncertainties on the assumed fluxes, which are disregarded here. The ratio plot in Panel 7.15c indicates that the slope of the atmospheric neutrino flux below reconstructed energies of 10 TeV may be softer than assumed in the Hillas model (H3a). The distribution of reconstructed energies also hints at interesting features in the spectrum that will require further studies in dedicated analyses. Disagreements in the reconstructed azimuth and vertex-z distributions could indicate a contribution of an unmodeled class of events or the impact of systematic effects of the detector and glacial ice.

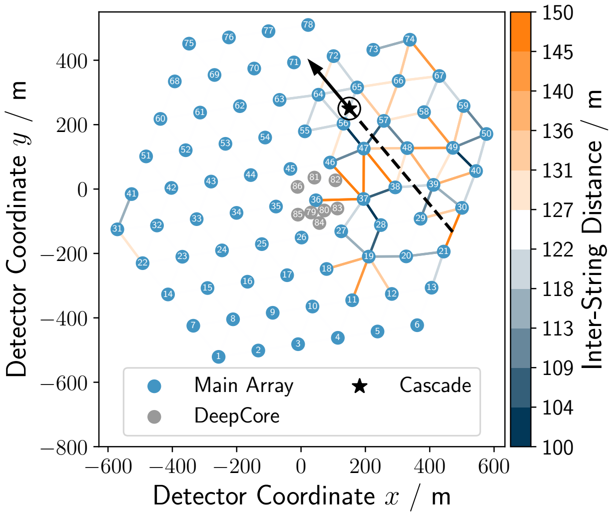
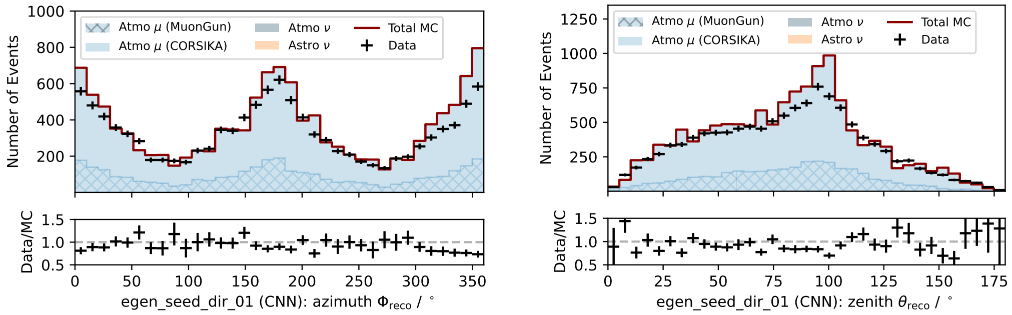


Figure 7.17: Inter-string distances of neighbouring strings are highlighted. The vertex location with x-y coordinates of (150 m, 250 m) is indicated by a black star and the azimuth angle of 310° is demonstrated by a black arrow. The dashed line shows that the trajectory of such a particle would enter the detector between strings 21 and 30 with an inter-string spacing of 147 m.

The excess of data events around azimuth angles of 310° could be linked to the layout of the detector that exhibits lanes through the instrumented volume, parallel to the axes of the triangular grid on which the detector strings are deployed, see Figure 7.17 for a visualization of the detector layout. Muons traveling on these lanes in between strings could enter the detector undetected and thus imitate starting events. This effect could be amplified by gaps and irregularities in the string layout. For instance, the distance between Strings 21 and 30 at the outer edge of the detector is 147 m, which is substantially larger than the average distance of 125 m. Stricter selection criteria would be necessary to reject muons entering in between these strings. In principle, these effects should be visible in MC simulation and therefore accounted for by the classifiers and the event selection pipeline. However, these events constitute a very narrow class of background muons that may not be simulated in required abundance. Due to the large initial atmospheric muon flux, single MC events can obtain large weights resulting in large flux uncertainties comparable to the neutrino flux levels. Although not conclusive, Panels 7.16a and 7.16b could indicate a light excess of events from the regions around reconstructed vertex x and y coordinates of (150 m, 250 m). These coordinates, combined with an azimuth angle of around 310°, would be roughly consistent with a particle trajectory passing through the gap between Strings 21 and 30, as illustrated by the dashed black line in Figure 7.17. It is also possible that differences in active DOMs in experimental data runs versus simulation may cause muon leaks. Further investigation and dedicated simulations are required to substantiate or disprove these hypotheses.

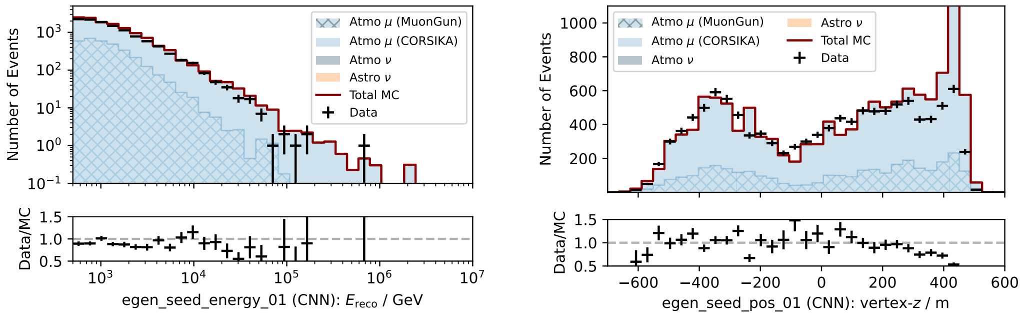
Alternative explanations include the direction dependence of photon propagation in the glacial ice. This ice anisotropy has been a longstanding calibration issue [29] known to cause discrepancies in the photon counts originating from photons traveling parallel versus perpendicular to the flow axis of the glacial ice. Recent breakthroughs were achieved by the introduction of birefringence in the modeling of the ice [30]. The neutrino simulation sets (21535, 21536, and 21537) used for the DNNCascade sample and analysis described in Part III utilize the first generation of the birefringence ice models (BFRv1) with an incomplete description of the ice layer tilt. Newer ice model versions have since been released that improve the agreement to experimental calibration data. In particular, the improved ice layer tilt description in newer models may reduce the current disagreement in the distribution of reconstructed z-vertices shown in Panel 7.16c. The distribution of z-vertices shows one of the largest discrepancies between MC prediction and experimental data. Unmodeled muon background may be a cause for this, but it also seems likely that systematic differences in the ice model description could be an underlying reason.

In any case, the overall description of atmospheric muon background seems adequately described by CORSIKA and MuonGun as demonstrated in Figure 7.18, which investigates the agreement in a region dominated by atmospheric muon background. This region is defined by looser selection criteria consisting of an energy cut of $E > 500\,\mathrm{GeV}$ and a maximum μ-BDT score value of 0.98. In order to explain the discrepancies in Figures 7.15 and 7.16 by a global scaling factor applied to the expected atmospheric muon signal, the normalization would have to be increased by a factor of two to ten. This would be in stark contradiction to the experimental data distributions for the background region shown in Figure 7.18. The overall normalization of CORSIKA and MuonGun simulation therefore seems correct, but it is possible that certain phase space is not properly covered or not simulated densely enough, resulting in mis-modeling of tails of distributions that become relevant at final selection levels. Additional comparison plots for various variables at DNNCascade Level 5 are provided in the supplement in Section C.2.



(a) Azimuth (CNN).

(b) Zenith (CNN).



(c) Energy (CNN).

(d) Vertex-z (CNN).

Figure 7.18: Experimental data distributions of the DeepLearningReco_event_selection_egen_seed* CNNs are compared to expectations obtained from MC simulations in a background dominated region. For the applied cuts ($E_{\mathrm{reco}} > 500\,\mathrm{GeV}$, μ-BDT < 0.98 and passed Online Cascade Filter), less than four neutrinos are expected for a livetime of about 48 minutes over a background of about 12 000 atmospheric muons. Systematic uncertainties are not included.

7.6 Dataset Unblinding

The DNNCascade selection is applied to ten years of IceCube data ranging from May 2011 to May 2021, corresponding to a livetime of about 9.64 years. The data rate in each experimental data year is stable as shown in Figure 7.19 with an average rate of 0.196 mHz.

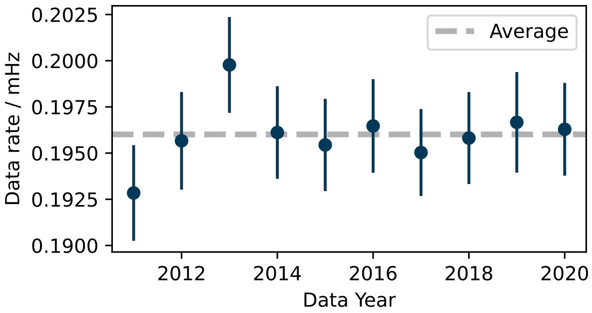
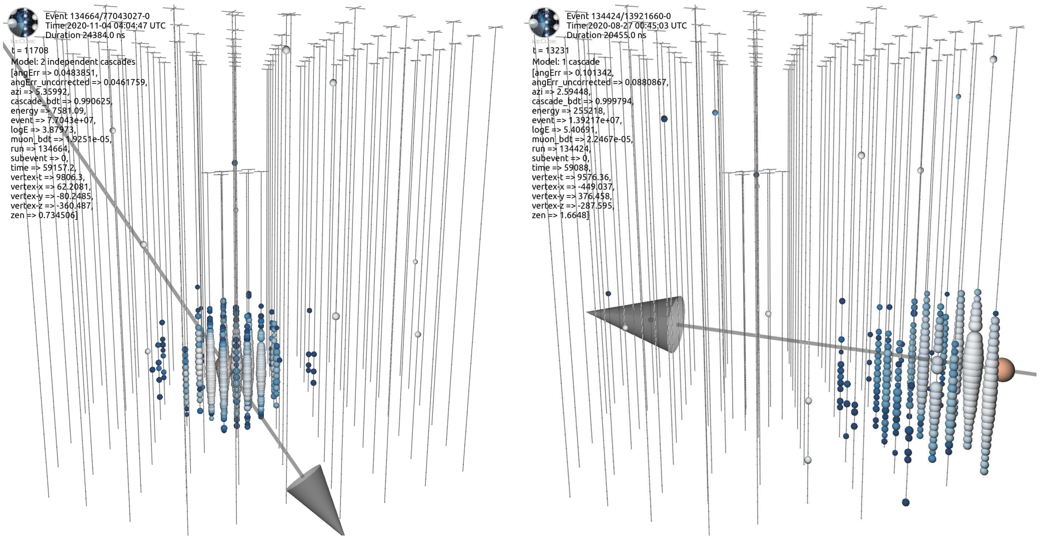


Figure 7.19: Data rates for each year of data taking are shown. Individual yearly rates are well consistent with the overall average rate of the sample.

A total of 59592 events are selected. Compared to the precursor event selection [67], referred to as the "Medium Energy Starting Cascade" (MESC 7yr) sample, with 1980 events over seven years, this corresponds to an increase of 20 times as many events per livetime. A significant fraction of the additional events are obtained by the lower energy threshold and the inclusion of partially contained and uncontained events, enabled by novel selection techniques as described in Section 7.2 [Design and Setup of the Selection Pipeline]. Example event views of a contained and uncontained cascade event are shown in Figure 7.20. These example events are among the top 20 events that contribute most to the signal from the Galactic plane emission search for the π^0 model as described in Part III. Figure 7.21 demonstrates that the reconstructed interaction vertices of the DNNCascade selection extend beyond the convex hull around the detector in contrast to the MESC 7yr sample. The inclusion of such events enable the selection of additional PeV events, including the most energetic cascade event detected to date with a reconstructed deposited energy of about 5.4 PeV, illustrated in Figure 7.24.

As discussed in Section 7.5 [Data/MC Agreement], the inclusion of partially contained and uncontained events may lead to a stronger dependence on atmospheric muon simulation to achieve an accurate modeling. In order to asses potential impacts, Figure 7.22 compares the fraction of events with vertices reconstructed outside of the detector volume as a function of reconstructed energy for experimental data and MC simulation. The fraction of uncontained events in experimental data well match expectations based on simulation, implying a correct description of the atmospheric muon background.



```
Event 134664/77043027-0
Time 2020-11-04 04:04:47 UTC
Duration 24384.0 ns
t = 11708
Model: 2 independent cascades
[angErr => 0.0483851,
angErr_uncorrected => 0.0461759,
azi => 5.35992,
cascade_bdt => 0.990625,
energy => 7581.09,
event => 7.7043e+07,
logE => 3.87973,
muon_bdt => 1.9251e-05,
run => 134664,
subevent => 0,
time => 59157.2,
vertex-t => 9806.3,
vertex-x => 62.2081,
vertex-y => -80.2485,
vertex-z => -360.487,
zen => 0.734506]
```

```
Event 134424/13921660-0
Time 2020-08-27 00:45:03 UTC
Duration 20455.0 ns
t = 13231
Model: 1 cascade
[angErr => 0.101342,
angErr_uncorrected => 0.0880867,
azi => 2.59448,
cascade_bdt => 0.999794,
energy => 255218,
event => 1.39217e+07,
logE => 5.40691,
muon_bdt => 2.2467e-05,
run => 134424,
subevent => 0,
time => 59088,
vertex-t => 9576.36,
vertex-x => -449.037,
vertex-y => 376.458,
vertex-z => -287.595,
zen => 1.6648]
```

Figure 7.20: Event views are shown for a "contained" (left) and an "uncontained" (right) cascade event. The colored blobs indicate DOMs that registered light, where the size of these blobs corresponds to the amount of observed light and the color indicates the time of the first registered light from early (white) to late (dark blue) times. The brown sphere and grey arrow show the reconstructed vertex and direction of the neutrino interaction. Figure and caption adopted from Ref. [1]

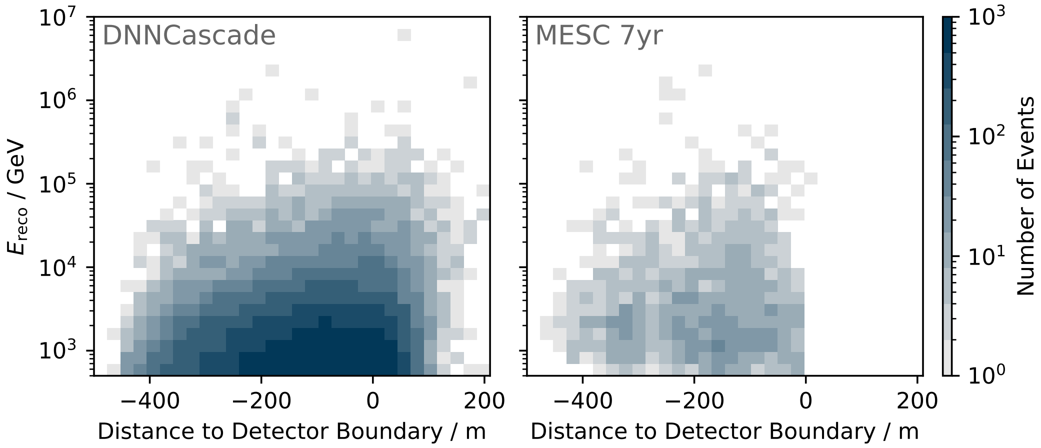


Figure 7.21: The distribution of experimental data events in reconstructed energy and distance to the detector boundary are compared for this work and the previous cascade selection, MESC 7YR [67]. Negative distance values indicate that the interaction vertex is reconstructed inside the instrumented volume and thus these events are referred to as contained events. Figure and caption adopted from Ref. [1].

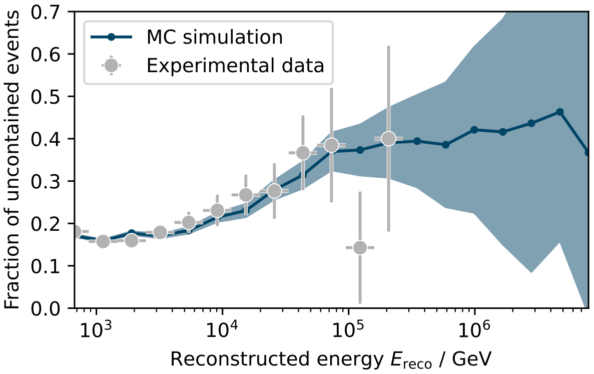


Figure 7.22: The fraction of "uncontained" events is shown as a function of reconstructed energy. With increasing energy, more events are "uncontained", up to a fraction of about 40% above 100 TeV. Experimental data shows good agreement to expectations obtained from MC simulation.

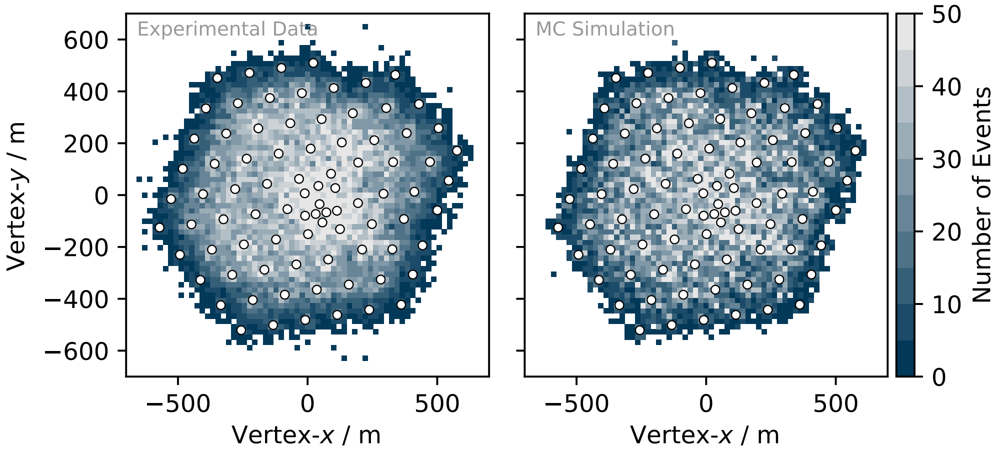


Figure 7.23: The distribution of reconstructed vertex positions in the x-y-plane are shown for the experimental data (left) and for the MC simulation (right). Locations of strings are indicated by white dots.

An overview of the reconstructed vertex positions in the x-y-plane is provided in Figure 7.23. Distributions for experimental data match those expected from MC simulation. The seemingly grainier structure on the right of Figure 7.23 for the simulation stems from limited statistics of atmospheric muon simulation at final event selection level. These simulated muon events obtain large weights and thus result in large uncertainties in each of the shown bins. With more simulation data, the simulation distribution will further smoothen out. Note that larger bins were not chosen here in order to visualize fine structures such as the dip in number of events in between strings 21 and 30; see also Figure 7.17 for an overview of the string layout.

As part of the dataset unblinding, the top ten events with the highest reconstructed energies are individually unblinded and inspected in detector coordinates. An overview of these events is provided in Table 7.8. Events also contained in the MESC 7YR sample are shown in Table 7.9. An event view of the most energetic event with a deposited energy of 5.4 PeV is provided in Figure 7.24. This event was also detected in the the "Extremely High-Energy" (EHE) [108] and "PeV energy partially contained event sample" (PEPE) [109] event selections. There are indications that this event is a Glashow Resonance event [110, 111]. Within the studies reported in Ref. [111], a dedicated and computing intensive event reconstruction [111, 112] was applied. The results obtained are in good agreement with the reconstructed values computed by the EVENT-GENERATOR within the DNNCASCADE sample. Additional event views are provided in the supplement in Section C.3 [Event Views].

Table 7.8: An overview of the top ten events with the highest reconstructed energies is shown. Half of the events are also contained in the MESC 7YR sample.

Run	**Event**	**Year**	E / **PeV**	θ	Φ	σ	**MESC**	**Name**
128897	2121632-0	2016	5.35	71.9°	358.8°	5.4°	no	Hydrangea
121240	72944671-0	2012	1.96	47.2°	333.1°	8.6°	yes	Big Bird
133609	17773654-0	2020	1.62	70.5°	297.8°	42.0°	no	–
133946	2834525-0	2020	1.25	79.6°	140.3°	7.4°	no	–
119316	36556705-0	2012	1.08	28.7°	296.1°	5.9°	yes	Ernie
132433	44612142-0	2019	0.89	105.9°	206.5°	13.1°	no	–
118545	63733662-0	2011	0.87	65.0°	324.6°	4.2°	yes	Bert
122918	19795769-0	2013	0.79	95.4°	126.1°	7.1°	no	–
128224	10435404-0	2016	0.65	69.2°	252.5°	8.5°	yes	–
122819	44356016-0	2013	0.60	54.1°	339.0°	24.7°	yes	–

Table 7.9: Event properties as used in the MESC 7YR sample are shown for the overlapping high-energy events contained in both samples.

Run	**Event**	**Year**	E_{MESC} / **PeV**	θ_{MESC}	Φ_{MESC}	σ_{MESC}	**Name**
121240	72944671-0	2012	1.99	48.1°	326.5°	7.0°	Big Bird
119316	36556705-0	2012	1.25	21.1°	244.3°	5.7°	Ernie
118545	63733662-0	2011	1.03	70.3°	328.0°	6.3°	Bert
128224	10435404-0	2016	0.64	62.2°	241.1°	8.4°	–
122819	44356016-0	2013	1.87	106.8°	18.7°	52.6°	–

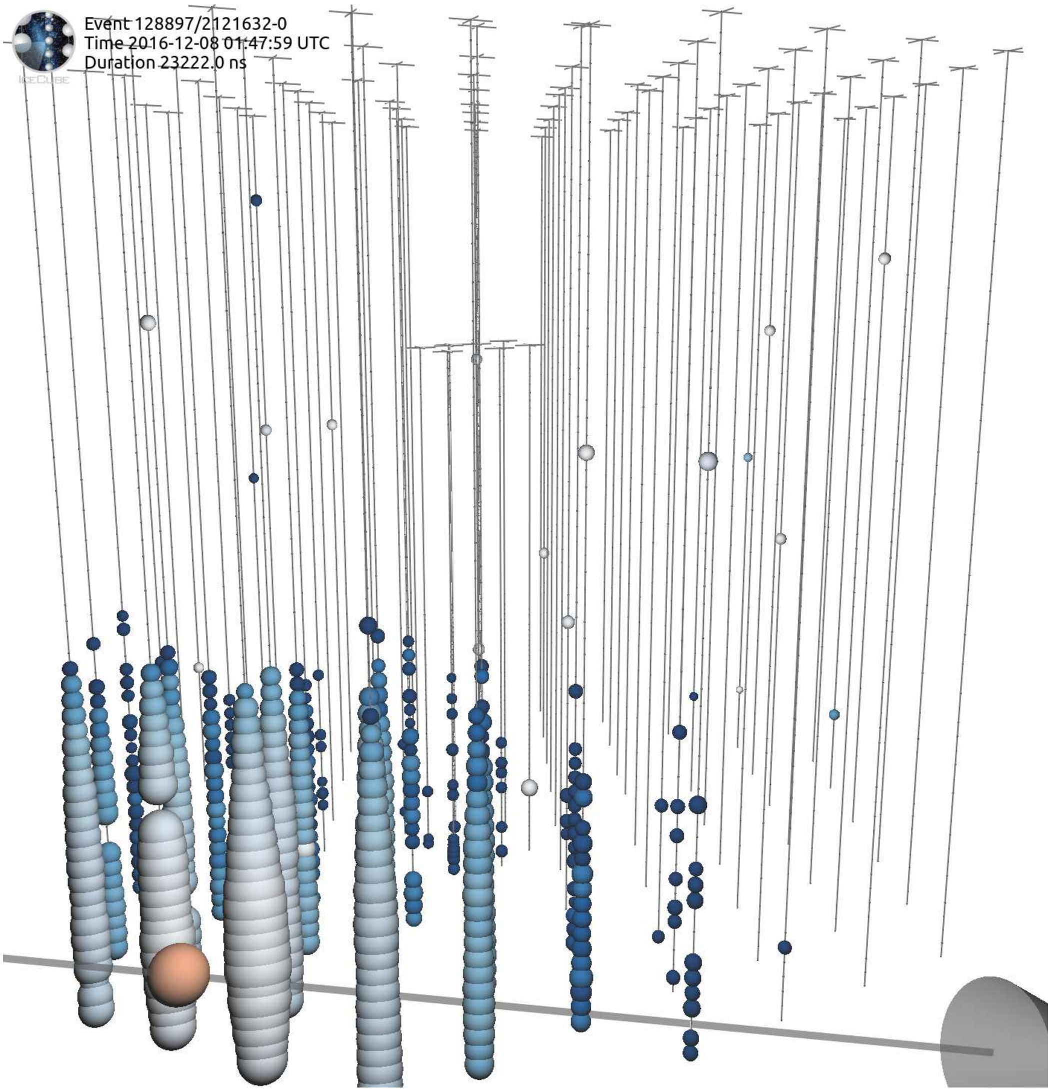


Figure 7.24: The event view of the most energetic event in the DNNCascade sample is shown with a reconstructed deposited energy of 5.4 PeV. The colored blobs indicate DOMs that registered light, where the size of these blobs corresponds to the amount of observed light and the color indicates the time of the first registered light from early (white) to late (dark blue) times. The brown sphere and grey arrow show the reconstructed vertex and direction of the neutrino interaction.

7.7 Conclusions

Novel methods, developed in this dissertation, are applied in the selection of cascade-like, neutrino-induced events on 10 years of IceCube data. These methods enable the most sensitive neutrino sample to date in the Southern Sky, with an increase in selection efficiency by over an order of magnitude in comparison to prior work by IceCube [67]. This improvement in selection efficiency, paired with the updated reconstruction methods, results in an enhanced sensitivity to the Galactic plane by a factor of three and thus ultimately to the first observation of high-energy neutrinos from the Milky Way, as discussed in Part III.

While the specific analysis level cuts in selection step 5 (Section 7.2.5) are optimized for the search of neutrino emission from the Galactic plane, the developed DNNCascade selection provides a rich portfolio of future applications, including measurements of the diffuse neutrino spectrum. Event views provided in supplement Section C.3 and cross-checks performed in the previous sections indicate that the methods perform as advertised. Overall, the generated sample demonstrates the potential of updated methods, capable of leading to equivalent improvements as expected from expensive detector upgrades (Section 3.1.1 [Detector Extensions: IceCube Upgrade and IceCube-Gen2]).

Part III

Neutrino Emission from the Galactic Plane

Declaration

Parts of the following chapters are published in Ref. [1].

IceCube has discovered a flux of astrophysical neutrinos in 2013 [93] and presented evidence for the first point-like neutrino sources, a flaring blazar known as TXS 0506+056 [113, 114] and the active galaxy NGC 1068 [88, 89]. However, the sources responsible for the majority of the astrophysical neutrino flux remain elusive. A considerable fraction of the observed astrophysical neutrino flux is theorized to originate in our own Galaxy by interactions of cosmic rays at their acceleration sites and during propagation through the interstellar medium. Methods developed in Part I, combined with the novel cascade event selection introduced in Part II, are leveraged to search for neutrino emission with cascade events. Cascade events constitute a complimentary channel to track events that is particularly sensitive to the Southern Neutrino Sky and thus to neutrino emission from the Galactic plane (GP). These searches are presented here, leading to the first observation of our Galaxy in high-energy neutrinos and thus identifying the origin of up to 10% of the astrophysical neutrino flux.

8 The Milky Way as a Source of Neutrinos

Declaration

This Chapter closely follows the introduction laid out in Ref. [1], co-authored by the author of this dissertation, with minor modifications applied.

The Milky Way, shown in Figure 8.1, is an emitter across the electromagnetic spectrum, observed from radio to gamma rays. Observations in different wavelengths of light allow for insight into the structure of our Galaxy as well as identification of sources of the highest energy photons. Above 1 GeV, the Milky Way is the most prominent feature of the sky. A majority of this observed gamma-ray flux, illustrated in the second panel of Figure 8.1, is believed to consist of photons from the decays of neutral pions (π^0) produced by cosmic rays colliding with the interstellar medium within our Galaxy [117].

However, photons can also be produced in interactions of energetic electrons or absorbed by ambient matter, and so additional messengers are required to probe the cosmic-ray interactions and hadronic sources in the Galaxy. Along with neutral pions (π^0), cosmic-ray interactions will create charged pions ($\pi^\pm$), and among their decay products are neutrinos. In contrast to photons, these neutrinos rarely interact on their way to Earth and directly trace the interactions of cosmic rays, thus providing a complementary probe of the Galaxy. A diffuse neutrino flux concentrated along the Galactic plane has been predicted [118–121], but has remained elusive until now. The expected TeV-energy neutrino flux based on an extrapolation of the GeV-energy *Fermi*-LAT π^0 measurement [117] is shown in the third panel of Figure 8.1 and further described in Section 11.5 [Diffuse Galactic Plane Searches]. Apart from the diffuse emission, our Galaxy is densely populated with numerous high-energy gamma-ray sources, several classes of which are considered potential cosmic-ray accelerators and therefore neutrino sources [122–126]. The combination of these potential neutrino fluxes makes the Galactic plane a long-hypothesized source of neutrinos and frequent target for neutrino source searches.

Previous searches performed by IceCube and ANTARES [67, 127–130] could not establish the Galactic plane as a neutrino source (p-values ≥ 0.02). The advent of deep learning in data science has enabled the development of new tools [5, 6, 131], discussed in Part I, that allow the identification of a larger number of neutrino interactions in detector data, with improved angular resolution, as demonstrated in Part II. These methods have been used in this dissertation to perform the most sensitive searches to date for neutrinos from the Galactic plane, resulting in the

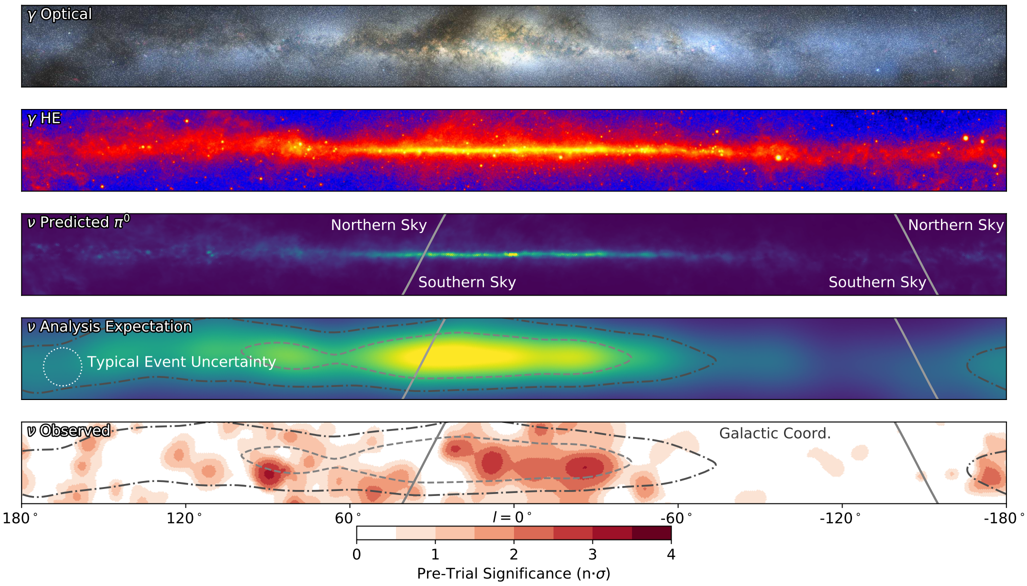


Figure 8.1: A multi-messenger view of the Milky Way galaxy is shown, centered on the Galactic center and viewed in galactic coordinates. Each panel shows the entire Galactic plane in a band of ±15° in galactic latitude, with each panel having a unique color scale. The top panel shows the view in the optical range as seen by A. Mellinger [115], which is partly obscured by clouds of gas and dust that absorb optical photons. The second panel shows the integrated flux in gamma rays as seen by the *Fermi*-LAT 12 year survey [116]. The third panel is the emission template for the expected neutrino flux (Section 11.5), taken to match the π^0 template from *Fermi*-LAT measurements of the diffuse gamma-ray emission [117]. The fourth panel shows the emission template from panel 3 convolved with the IceCube detector acceptance for the DNNCascade sample and the angular uncertainty of a typical signal event (7°). For orientation purposes, panel 4 also includes the central regions that contain 20% and 50% of the overall expected diffuse neutrino emission signal. The final panel shows the pre-trial significance of the all-sky seach for point-like sources, described in Sections 11.1 and 14.2, using the cascade neutrino event sample in the same band of the Galactic plane, with the expected neutrino emission regions as in panel 4. Grey lines in the bottom three panels indicate the Northern-Southern Sky horizon line at the IceCube detector. Figure and caption adopted from Ref. [1].

> observation of neutrinos from the Milky Way galaxy with a significance of 4.5σ using 10 years of IceCube data.
>
> Establishing the Galactic plane as a source of neutrinos complements the all-sky extra-galactic diffuse neutrino signal observed by IceCube [91,95,132]. The detection also confirms our understanding of the signal from interactions of cosmic rays in the Galaxy and enables the pursuit of a wide spectrum of scientific questions. These include identifying the Galactic sources of high-energy cosmic rays and understanding their propagation, mapping the distribution of sources and matter in our Galaxy, and probing physics in extreme astrophysical environments [133]. ”[1]

9 Leveraging Cascade Events

Searches for neutrino sources must find a way to manage the overwhelming background of atmospheric muons and neutrinos. When searching for an excess of signal events over background, the sensitivity to such a signal can be improved by either reducing the level of background in the data set, or by reducing the area searched in the sky. Assuming that the signal is a point-like source, both of these methods will increase the signal to noise ratio and therefore facilitate detection of the source.

Analyses based on tracks [88, 89] aim to utilize the superior angular resolution of through-going events in order to reduce the required search area around the position of an assumed source. This works well for point-like sources for which the search area can be reduced according to the angular uncertainty of the signal events. However, for extended sources such as the Galactic plane, there is a limit up to which an improved angular resolution can reduce thre required search area, as this is bounded by the size of the emission region of the source. Thus, alternative options that allow to reduce the overall background levels are required to increase the signal to noise ratio.

As discussed in Section 7.1 [Prospects and Challenges of Cascade Samples], although cascade events have an inferior angular resolution compared to tracks, their signal to noise ratio is about an order of magnitude better at TeV energies. On top of the improved astrophysical purity, cascades also provide a much improved energy resolution than through-going events, which can be used as a further mechanism to increase the signal to noise ratio in the analysis. Cascade events are therefore not only a viable, but also an important complimentary channel for neutrino source searches.

In comparison to tracks, cascades are also beneficial when the spatial emission profile of the assumed source is poorly defined or subject to large uncertainties. As demonstrated in Figure 9.1, cascade-based searches are more robust to spatial uncertainties on the hypothesized emission regions. For this test, the sensitivity to the π^0 emission template (Section 11.5 [Diffuse Galactic Plane Searches]) is investigated with the DNNCascade sample and a starting track sample [94]. Pseudo experiments are generated with signal injected for the nominal case and for spatial templates that are rotated by various angles. While the track-based search is highly susceptible to such a mis-modeled spatial template, the cascade-based search remains stable up to rotations of about 5°. Even at rotations of 10°, a significance of around 4σ is expected for the DNNCascade analysis. The robustness of the DNNCascade GP analysis is beneficial for the detection of a poorly defined signal, but it complicates further interpretation of the spatial components, as discussed in Section 15.7 [Model Confusion and Unresolved Sources].

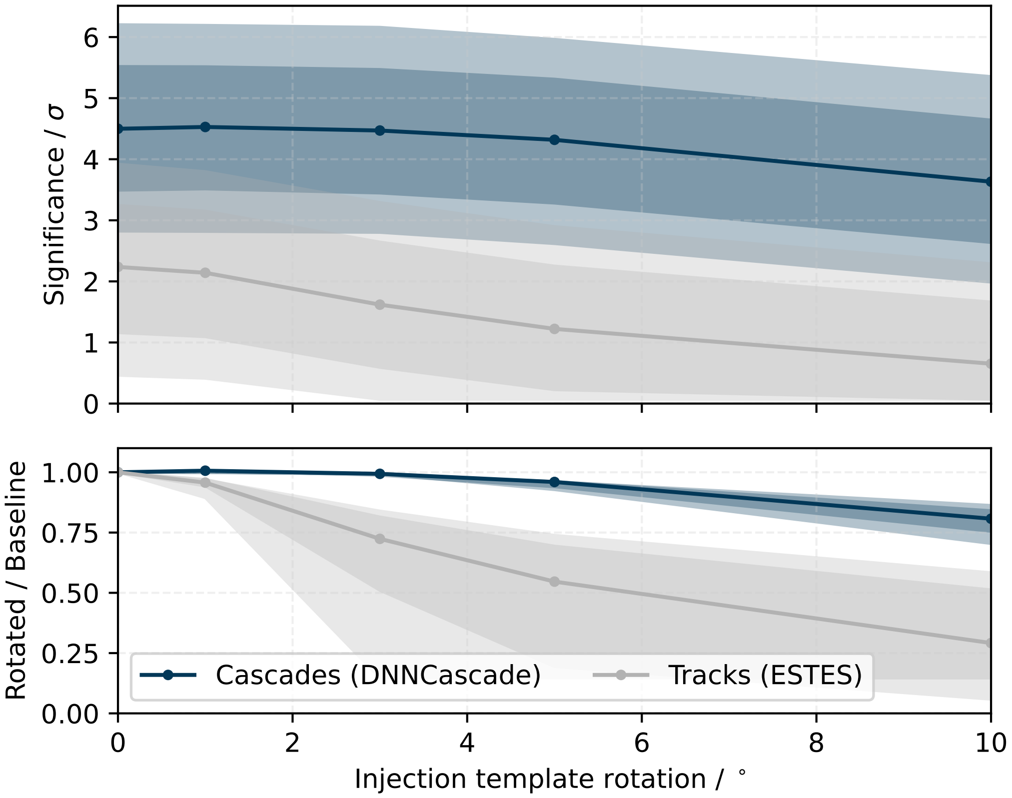


Figure 9.1: The robustness towards spatial mis-modeling of the π^0 Galactic plane template (Section 11.5) is tested for the DNNCascade sample and the "Enhanced Starting Track Event Selection" (ESTES) [94]. Signal events are injected in pseudo experiments based on the measured π^0 flux (Table 14.1), but for modified spatial templates. The spatial templates are rotated by various angles to emulate mis-modeling of the spatial PDF. The significance is then evaluated for these trials assuming the baseline template and it is illustrated in the top panel. The lower panel shows the relative change in significance due to this mis-modeling.

10 Analysis Methods

Declaration

The description of the analysis methods is adopted from Ref. [1], co-authored by the author of this dissertation, with some additional information provided here.

The analysis methods and performed searches are described in this chapter. A data-driven search method is employed to mitigate the impact of systematic uncertainties.

10.1 Maximum-Likelihood-Based Search Method

Declaration

This section closely follows the description laid out in Ref. [1] with some minor modifications applied.

> The search for neutrino sources, both point sources and extended, is based on the maximum-likelihood technique, largely using the standard IceCube methods from Ref. [134]. A mixture PDF is utilized to describe contributions from signal (S_i) and background (B_i) terms:
>
> $$\mathcal{L}(n_s, \vec{\theta}_s) = \prod_{i=1}^{N} \frac{n_s}{N} S_i(\vec{x}_i, \vec{\theta}_s) + (1 - \frac{n_s}{N}) B_i(\vec{x}_i) \tag{10.1}$$
>
> where n_s is the number of signal events, N is the total number of events, and the event observables
>
> $$\vec{x}_i = (\alpha_i, \delta_i, \sigma_i, E_i)\,. \tag{10.2}$$
>
> are given by the right ascension (α_i), declination (δ_i), event angular uncertainty (σ_i), and reconstructed energy (E_i). The signal hypothesis depends on the parameters $\vec{\theta}_s$. A common signal hypothesis consists of neutrino emission from a particular source, following a
>
> $$\Phi(\mathrm{E}) = \Phi_0 \cdot \left(\frac{\mathrm{E}}{100\,\mathrm{TeV}}\right)^{-\gamma_s} \tag{10.3}$$

power-law spectrum. In this case, the signal hypothesis parameters

$$\vec{\theta}_s = (\alpha_s, \delta_s, \gamma_s)\,. \tag{10.4}$$

are defined by the location of the source (α_s, δ_s) and the spectral index (γ_s).

In order to account for the signal contribution in the data-driven background estimate, detailed in Section 10.4 [Data-driven Search Method and Systematics], a modification of the likelihood defined in Eq. (10.1) is used, first introduced in [127]. The data-driven PDF, $\tilde{D}_i$, does not correspond to pure background B_i in Eq. (10.1), but includes the expected contribution from the signal PDF averaged over right ascension, $\tilde{S}_i$, in proportion to the strength of the signal that is fit:

$$\tilde{D}_i = (1 - \frac{n_s}{N})B_i + \frac{n_s}{N}\tilde{S}_i. \tag{10.5}$$

The likelihood equation is then rearranged to arrive at the "signal-subtracted" likelihood function Eq. (10.6):

$$\mathcal{L}(n_s, \vec{\theta}_s) = \prod_{i=1}^{N} \frac{n_s}{N} S_i(\vec{x}_i, \vec{\theta}_s) + \tilde{D}_i(\vec{x}_i) - \frac{n_s}{N}\tilde{S}_i(\vec{x}_i, \vec{\theta}_s). \tag{10.6}$$

The test statistic Λ is defined by the ratio of the maximized likelihood and the likelihood for the null hypothesis, shown in Eq. (10.7):

$$\Lambda = -2\ln\frac{\mathcal{L}(n_s = 0)}{\mathcal{L}(\hat{n}_s, \vec{\theta}_s)} \tag{10.7}$$

where $\hat{n}_s$ is the best-fit number of signal events. As described in Section 10.4, this test statistic is calculated for experimental data and it is converted to a p-value by comparing the experimental result to the test statistic distribution obtained from background pseudo-experiments generated from scrambled experimental data. ”[1]

10.2 Analysis Tool Simplifications

The analysis is performed with the CSKY software[1], which is a general software package designed for neutrino source searches. In order to reduce computational requirements and to facilitate the analysis, simplifications are introduced to the signal and background terms utilized in the likelihood defined in Eq. (10.6).

---

[1] `https://github.com/icecube/csky` (IceCube GitHub access required)

The signal term S_i, as utilized in Eq. (10.6), requires a high-dimensional PDF over the four event observables $\vec{x}_i$ plus the signal hypothesis parameters $\vec{\theta}_s$. While the data-derived background term $\tilde{D}_i(\vec{x}_i)$ does not depend on the assumed signal hypothesis, it still requires the construction of a four-dimensional PDF. The construction of such a PDF is not feasible given the limited number of available data events (59592). Instead, the dimensionality is reduced by decomposing the signal and background terms into a product of two separate PDFs pertaining to the spatial and energy components:

$$S_i(\vec{x}, \vec{\theta}) \approx S_i^{\text{space}}(\alpha_i, \delta_i \,|\, \alpha_s, \delta_s, \sigma_s) \cdot S_i^{\text{energy}}(E_i \,|\, \delta_i, \gamma_s) \tag{10.8}$$

and

$$\tilde{D}_i(\vec{x}_i) \approx \tilde{D}_i^{\text{space}}(\alpha_i, \delta_i) \cdot \tilde{D}_i^{\text{energy}}(E_i \,|\, \delta_i). \tag{10.9}$$

The spatial term $\tilde{D}_i^{\text{space}}$ from Eq. (10.9) can be further simplified by exploiting that background is uniformly distributed in right ascension due to the rotation of Earth and IceCube's location at the South Pole, arriving at:

$$\tilde{D}_i(\vec{x}_i) \approx \frac{1}{2\pi} \cdot \tilde{D}_i^{\text{space}}(\sin \delta_i) \cdot \tilde{D}_i^{\text{energy}}(E_i \,|\, \delta_i), \tag{10.10}$$

which is the commonly used likelihood definition in IceCube source searches [134]. This dimensionality reduction is possible by dropping weak conditional dependencies in each of the terms. Terms such as the conditional PDF over the event angular uncertainty σ can be omitted under the assumption that the PDFs are similar for background and signal events and thus cancel in the likelihood ratio defined in Eq. (10.7).

The spatial signal term S_i^{space} further uses a von Mises-Fisher [82] distribution on the sphere with "circularized" angular uncertainties σ, symmetrical in detector azimuth and zenith. This is a simplification that greatly facilitates and speeds up the analysis software, but it is not entirely correct as cascades do not necessarily have symmetrical uncertainties due to the differing detector spacing along the z-axis versus x-y-plane. See also Section 7.3 [Event Reconstruction and Uncertainty Estimation] and supplement Chapter A [Cascade Real-Time Alert Stream] for further discussion. Similar limitations apply to the template searches introduced in Section 10.3, which employ a symmetrical Gaussian smearing to modify templates to account for event uncertainties.

While the discussed simplifications enable the employed analysis tools, they may result in minor biases and analysis performance degradation.

10.3 Template Analysis

In addition to searches for point-like sources, the CSKY software also supports searches based on pre-defined templates. The full sky templates are binned in equal-solid-angle bins, convolved with the detector acceptance, and smeared with a Gaussian to generate the spatial PDF term S_i^{space} in Eq. (10.8). Example templates with detector acceptance and a smearing of 7° and 15° applied are provided in Figure 11.3. Note that the spatial template is independent of the energy; templates with energy-dependent spatial components are first marginalized over the energy spectrum of the DNNCascade sample to remove the conditional dependence. The templates are smeared with Gaussian distributions for a fixed set of angular uncertainties. For each event, the closest pre-computed template is then chosen to evaluate the likelihood in Eq. (10.6) based on the per-event angular uncertainty.

10.4 Data-driven Search Method and Systematics

Declaration

This section closely follows the description laid out in Ref. [1] with some minor modifications applied.

The source searches performed in this dissertation utilize a data-driven method to model backgrounds and to evaluate the analysis significance. This procedure allows to construct robust and conservative p-values that are correct even in the presence of detector systematics as these are — by construction — already accounted for.

> “Due to IceCube’s location at the South Pole and the rotation of Earth, backgrounds to this analysis (atmospheric events and an assumed-isotropic astrophysical neutrino background) are uniformly distributed in right ascension. Similarly, the impact of any terrestrial sources of systematic uncertainty will also be distributed evenly in right ascension. The data-driven search method makes use of this, generating background pseudo-experiments by randomizing the right ascension values of the observed experimental data. This randomization will remove signal from sources by evening out anisotropic structure on the sky (in given declination bands), while maintaining the influence of systematic uncertainties and the statistical properties of the background events in experimental data. [Note that for searches of point-like sources an additional local randomization in declination is performed, as discussed in Section 12.1 [Point-like Source Searches]]. This data-driven procedure prevents the risk of false discovery of anisotropic structures on the sky.” [1]

The test statistic Λ from Eq. (10.7) is evaluated for each of these background pseudo-experiments, resulting in a background distribution of Λ_B values. These values describe the distribution of test statistic values that are generated under the null hypothesis by random chance. The analysis p-value is then obtained by comparing the test statistic Λ for the (non-scrambled) experimental data to the fraction of generated pseudo-experiments Λ_B that have a value greater or equal to the experimental data one.

Although the computed p-value of the analysis correctly accounts for systematic uncertainties and mis-modeling in the MC simulation, improved modeling will improve analysis performance and thus generally result in more significant results if signal is present. Projected analysis sensitivities and effective areas, used to convert the fitted number of events n_s to an observed best-fit flux normalization, are based on simulations of the detector. Flux conversions are thus dependent on correct modeling, requiring adequate accounting of systematic uncertainties, as described in Section 15.2 [Flux Measurement]. [1]

> Systematic uncertainties pertaining to ice-modeling and detector properties (Section 3.1) are accounted for in multiple places. The CNNs in the event selection pipeline, discussed in Part II, are trained on a variety of different simulation datasets with different systematic properties. This ensures that the CNNs are more robust towards these known sources of systematic uncertainties. Quantitative comparisons on the impact of these systematics on the CNNs are presented in Section 4.5 [Robustness and Data/Simulation Agreement]. The neural network models utilized in the hybrid method, Event-Generator, is also trained on simulation sets that sample different sets of systematic parameters for each subset of events [74]. Furthermore, for the loss function, the model utilizes Eq. (5.13) that is able to model the over-dispersion resulting from the marginalization of systematic parameters. As discussed in Section 7.3 [Event Reconstruction and Uncertainty Estimation], an additional scaling factor is applied to the per-event angular uncertainty estimates to account for the aforementioned sources of systematic uncertainties. Finally, the confidence interval construction of the best-fit flux normalization of the Galactic plane models, reported in Table 14.1 and Figures 15.3, account for these systematic uncertainties. As described in Section 15.2, the confidence intervals are constructed by inversion of the likelihood ratio test via the approach by Feldman and Cousins [135] while also sampling a concrete realization of systematic parameters for every trial. As such, the obtained intervals not only include statistical uncertainties, but also the aforementioned known sources of systematic uncertainties.
>
> Overall, this set of systematic uncertainties impact the sensitivity by up to 20% and the effective area by about 10%. The largest impact is on the angular resolution of high-energy events, which are most affected by systematic uncertainties in the

> modeling of the glacial ice. Nevertheless, all of these (and other unidentified) effects are — by construction — accounted for in the p-value calculation. Estimates of the best-fit flux normalizations are susceptible to systematic uncertainties, but the analysis itself is robust due to the data-driven search method. ”[1]

11 Performed Searches

Declaration

The description of the individual searches is adopted from Ref. [1], co-authored by the author of this dissertation.

A suite of searches for different signal hypotheses are performed as described in the following sections. These are common searches performed on new datasets, chosen to align with those searches performed in the precursor analysis [67]. Minor modifications were applied in order to tune the analysis to the properties of the DNNCascade sample.

11.1 All-sky Search

The “all-sky search” is an unbiased search for an excess of neutrino emission over background. A test for a point-like source is performed over the full sky on a grid of points of equal solid angle bins spaced 0.45° apart. At each location in the sky, the number of signal events n_s and the power-law spectral index γ_s is determined that maximize the test-statistic from Eq. (10.7). The method matches previous searches [67, 88]. Although individual points are highly correlated with their neighbors due to the large angular uncertainty of cascades, this test still entails a large (∼500) trial factor as a result of searching the entire sky. The brightest spot in each hemisphere is reported. [1]

11.2 Source List

In order to reduce the large trial factor of the unbiased all-sky search in Section 11.1, assumptions can be made on the signal hypothesis. Utilization of prior information restricts the search space and is therefore more sensitive to the assumed signal hypothesis. However, if the priors are incorrectly chosen this will result in an inability to detect neutrino emission. For this search, the assumption is made that a correlation exists between the neutrino flux and the GeV gamma-ray flux reported in the 4FGL catalog [136]. A list of 109 sources is defined *a priori* with the same methods as utilized in Ref. [88], but optimized for the declination dependent sensitivity of the DNNCascade event selection. For each of the locations in this source list, a test for a point-like source is performed similar to the all-sky scan. As illustrated in Figure 11.1, the majority of the selected sources are extragalactic objects.

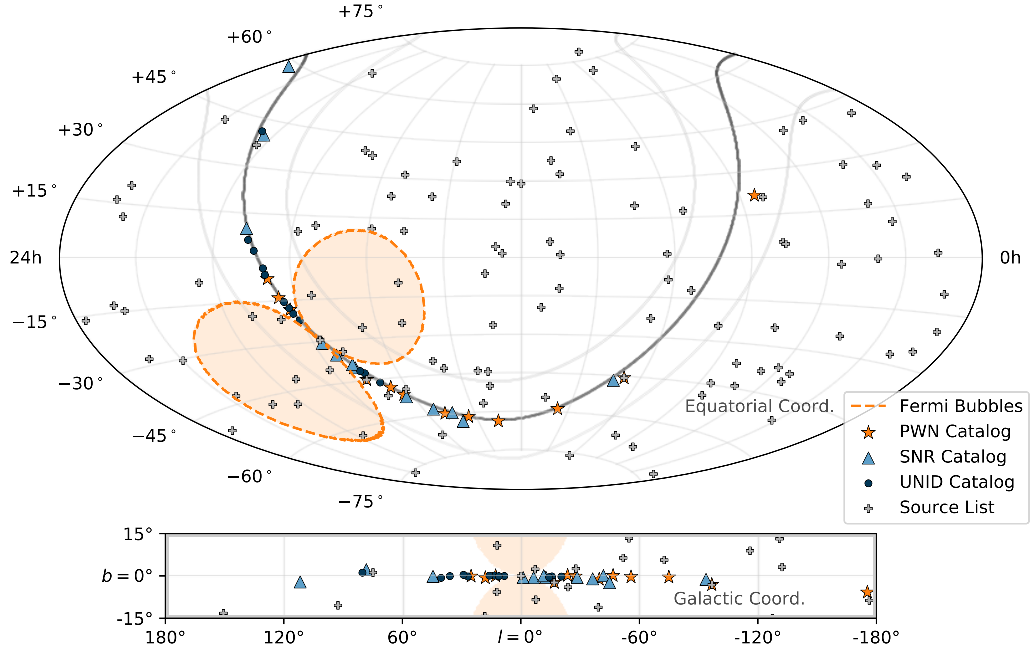


Figure 11.1: The sources in each of the catalogs and source list are shown as well as the *Fermi* Bubbles in an Aitoff projection in equatorial coordinates. An insert is added highlighting the region around the Galactic plane in galactic coordinates.

11.3 Catalog Stacking

In addition to introducing priors, many weak sources can be “stacked” in a single analysis to improve the sensitivity compared to many individual searches for each of the sources [137]. In contrast to random background, the neutrino flux will accumulate coherently. Hence, a stacking analysis will be more sensitive in the case that many of the stacked sources emit neutrinos.

Fermi-LAT measurements [117] indicate that the total gamma-ray signal from the Galactic plane includes the contribution of strong gamma-ray sources. This motivates searches for correlated neutrino emission from three distinct catalogs of Galactic sources classified as Supernova Remnants (SNR), Pulsar Wind Nebulae (PWN), and other unidentified Galactic sources (UNID) based on observations in TeV gamma rays [138, 139]. For each of these three catalogs, the 12 sources with the largest expected neutrino flux are selected based on the observed gamma-ray emission above 100 GeV and the detector sensitivity, following methods in [88]. The selected catalogs are listed in the supplement Table D.1 and visualized in Figure 11.1. A stacking search is performed for each of these catalogs, assuming that each source in the catalog contributes equally to the total neutrino flux as a point-like source with a shared power-law spectrum. The analysis fits for the spectral index (γ_s) and total number of signal events (n_s) of the catalog.

11.4 *Fermi* Bubbles

A template search is performed for neutrino emission from the *Fermi* Bubbles [140], a pair of gamma-ray emission regions that extend up to $\sim 55^\circ$ above and below the Galactic center. The template is constructed by adding two circular lobes with a radius of 25° tangent to each other at the Galactic center, as shown in Figure 11.1. Positions within these lobes are assigned equal probability for the emission of neutrinos following a power-law spectrum with a spectral index of $\gamma_s = 2$ and an optional spectral cutoff. Multiple spectral cutoffs including 50 TeV, 100 TeV, 500 TeV are tested in addition to an unbroken power-law.

11.5 Diffuse Galactic Plane Searches

As in previous analyses [67, 127], searches for diffuse neutrino emission from the Galactic plane (GP) are performed based on three different template models, referred to as π^0, KRA_γ^5 and KRA_γ^{50}. Each of these Galactic plane templates is derived from the same underlying gamma-ray observations [117]. The model predictions depend on the assumed distribution and emission spectrum of Cosmic Ray (CR) sources in our Galaxy, the properties of CR diffusion in the Interstellar Medium (ISM) as well as the spatial distribution of target gas.

The π^0 spatial template is derived from the π^0 gamma-ray component of the `GALPROP` [141] output files available from the supplemental material for the $^{S}\mathtt{S}^{Z}\mathtt{4}^{R}\mathtt{20}^{T}\mathtt{150}^{C}\mathtt{5}$ model [117][1]. The integrated π^0 gamma-ray component is normalized over the sky to serve as a spatial probability template only, identical to previous IceCube analyses [67, 127]. An unbroken power-law spectrum according to Eq. (10.3) with a spectral index of $\gamma_s = 2.7$ is assumed, based on the assumption that the MeV-to-GeV π^0 component, measured by the *Fermi*-LAT gamma-ray telescope, follows a single $\mathrm{E}^{-2.7}$ power-law to TeV energies with the same spatial emission profile. Note that this is a simplified assumption that disregards CR injection and propagation properties. The π^0 model tested for in this analysis should therefore be understood as a well motivated spatial template, designed for searches of neutrinos from the Galactic plane, and less as an internally consistent model of diffuse neutrino emission from pion decay.

---

[1] https://galprop.stanford.edu/PaperIISuppMaterial/

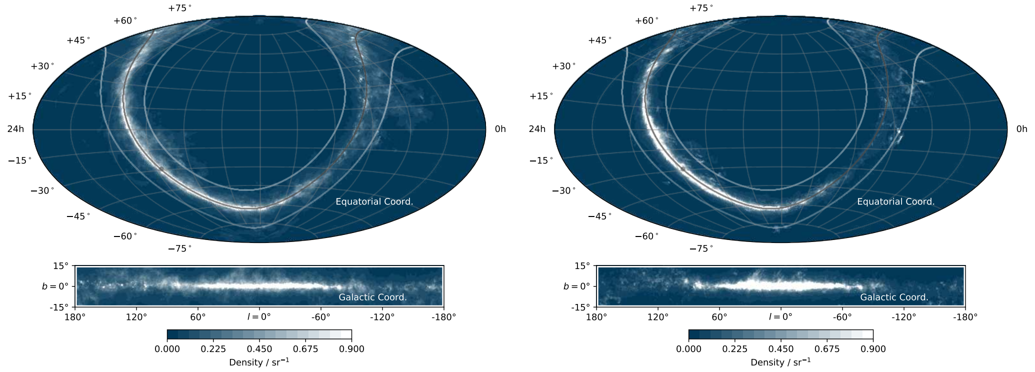


Figure 11.2: An Aitoff projection of the spatial templates for the π^0 (left) and KRA_γ^5 at 10 TeV(right) are shown in equatorial coordinates. The lower panel highlights the region around the Galactic plane in galactic coordinates.

As described in detail in Ref. [142], the KRA_γ models are based on the same underlying gas measurements as the π^0 template model, but the authors explain an apparent under-prediction of the observed gamma-ray flux at higher energies by modeling a radially dependent diffusion coefficient for the cosmic-ray propagation, using the cosmic-ray transport software package `DRAGON` [143] and a diffuse gamma-ray simulation code `Gamma-Sky` [144]. As shown in Figure 11.2, this results in more concentrated neutrino emission from the Galactic center region, whereas the π^0 model predicts events more dispersed along the Galactic plane. The spatial distribution of the KRA_γ^{50} model is similar to the KRA_γ^5 distribution shown in Figure 11.2. In contrast to the π^0 model, the KRA_γ models show a variable spectrum in different spatial regions, obtain an on-average harder neutrino spectrum, and include a spectral cutoff at the highest energies [142]. The neutrino flux predicted by the KRA_γ models, binned in equal solid angle and logarithmic energy bins, is available on Zenodo [145].

In this analysis, the KRA_γ models are tested with a template that uses a constant, model-averaged spectrum over the sky, as described in Section 10.3 [Template Analysis], roughly corresponding to an $\mathrm{E}^{-2.5}$ power law, with either a 5 PeV or a 50 PeV cosmic-ray energy cutoff for the KRA_γ^5 and KRA_γ^{50} models respectively. The corresponding cutoff for the neutrino spectrum in these models is at about 20 times lower energies. As detailed in Section 10.3 [Template Analysis] and illustrated in Figure 11.3, each neutrino emission model is converted to a spatial template and is then convolved with the detector acceptance and the event's estimated angular uncertainty to create an event-specific spatial PDF. For all three models, the spectrum is held constant and only the total number of signal events (n_s) is maximized, corresponding to the best-fit flux normalization.

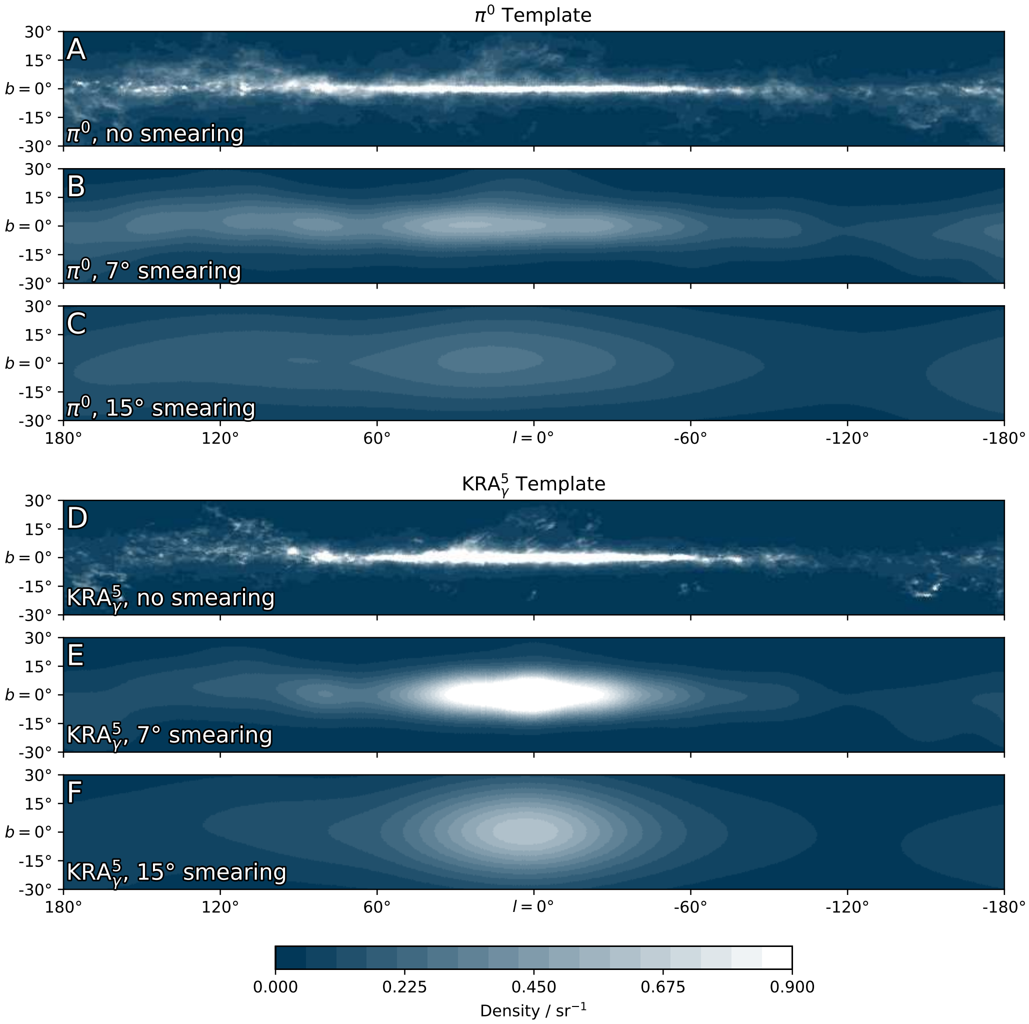


Figure 11.3: The spatial templates for the π^0 (A-C) and KRA$_\gamma^5$ (D-F) models of diffuse Galactic neutrino emission are shown. Each panel shows the Galactic plane in a band of $\pm 30°$ in latitude (b) and $\pm 180°$ longitude (l) in Galactic coordinates. The spatial template for the KRA$_\gamma^5$ model is obtained by marginalizing over the energy distribution of the DNNCascade sample. For the analysis, the models are first convolved with the IceCube detector acceptance (A, D) and then smeared with a Gaussian corresponding to the event uncertainty. Here, two example analysis templates are illustrated for a smearing of 7° (B, E) and 15° (C, F). The spatial distribution of the KRA$_\gamma^{50}$ model is very similar to the KRA$_\gamma^5$ one shown here.

12 Analysis Performance

Methods developed in Part I, combined with the novel cascade event selection introduced in Part II, are leveraged in this disseration to maximize sensitivity for searches of neutrino emission with cascade events. Over the course of this dissertation, the sensitivity and discovery potential of such searches has been improved by up to 400 %, as demonstrated for the π^0 and KRA$_\gamma$ GP analyses in Figure 12.1, or for searches of point-like sources, shown in Figure 12.2.

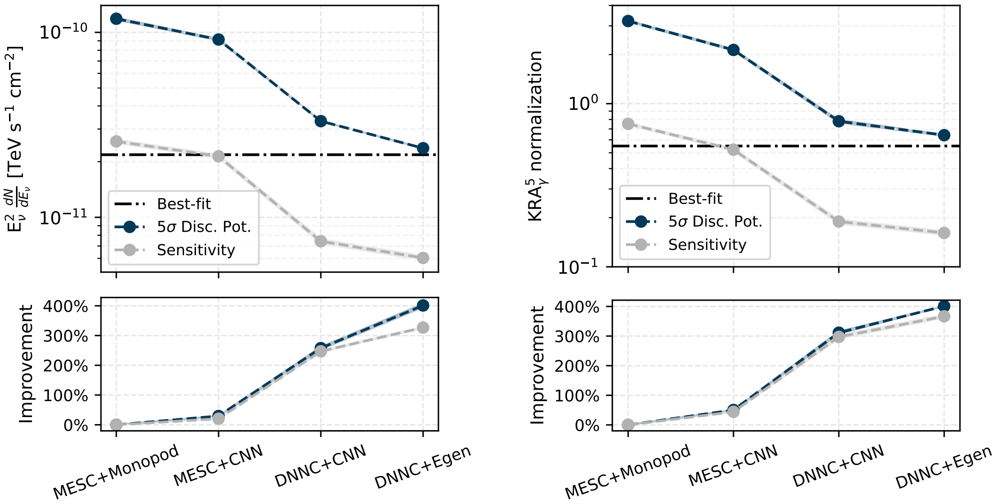


Figure 12.1: The improvement in sensitivity and 5σ discovery potential due to methods and datasets developed in this dissertation are shown for the π^0 (left) and KRA$_\gamma^5$ (right) GP analyses. Prior to this dissertation, the MESC 7yr [67] event sample with the Monopod reconstruction (MESC+Monopod), described in Section 3.4.1, was the state-of-the-art cascade GP analysis. Monopod was then replaced with the newly developed CNN, described in Part I, and eventually published in the MESC 7yr analysis [67] ((MESC+CNN)). The GP analysis conducted in this dissertation utilizes the newly developed DNNCascade event sample together with the Event-Generator framework (DNNC+Egen), resulting in an overall improvement up to 400 %.

Apart from method development, searches for neutrino emission will also gain sensitivity from an increased amount of collected experimental data. The sensitivity and discovery potential for these types of searches roughly scale with the square root of the experimental livetime. Thus, in order to gain the same level of improvements as achieved with the developed methods in this dissertation, the original analysis based on the MESC sample and Monopod reconstruction (MESC+Monopod in Figure 12.1) would have to be run for a total livetime of 175 years. The MESC 7yr analysis with the DNN-reco-based CNN (MESC+CNN in Figure 12.1) would require an extension in livetime of about 75 years. Assuming annual operation costs of about \$3 million[1], the developed methods result in savings upwards of \$225 million on top of enabling results on a reasonable timeline. The observation of the Galactic plane in high-energy neutrinos therefore relies on the newly developed tools and cannot be achieved by an increased livetime alone.

The following sections highlight the performance of the individual analyses, also in comparison to prior searches. Properties of the analyses such as biases in the fitted source parameters are investigated.

12.1 Point-like Source Searches

The sensitivity and discovery potential of the DNNCascade analysis to point-like sources are illustrated in Figure 12.2 as a function of declination. The sensitivity is defined as the flux at which 90% of signal trials, trials injected with signal events corresponding to a given flux, result in a test-statistic value greater than the median test-statistic value of background trials. Similarly, the discovery potential at $n \cdot \sigma$ is defined as the flux at which 50% of signal trials result in a significance of at least $n \cdot \sigma$. As demonstrated in Figure 12.2, a considerable improvement over the precursor analysis, MESC 7yr [67], is achieved leading to reductions in sensitivity of a factor of four. To date, the DNNCascade sample is the most sensitive neutrino dataset of any of the existing neutrino telescopes for point-like sources in the Southern Hemisphere; in the Northern Hemisphere, track-based samples provide superior resolution.

For the all-sky scan and source list search, defined in Sections 11.1 and 11.2, the number of signal events n_s and spectral index γ_s of an assumed point-like source are fitted. In an ideal scenario, the analysis should be able to — on average — recover the injected spectrum and amount of signal events. If this is not the case, this can be an indication of mis-modeling of the underlying PDFs in the likelihood from Eq. (10.6), which may lead to biases and degradation in analysis performance. In particular, biases need to be accounted for when converting the fitted number of events to a neutrino flux, as detailed in Section 15.2 [Flux Measurement]. In Figure 12.3, the recovery of the injected source parameters is tested at five different declinations. There is some bias visible at harder spectra as shown in the upper left panel of Figure 12.3, potentially caused by the simplifications introduced in the likelihood, described in Section 10.2 [Analysis Tool Simplifications]. For this declination, the background test-statistic distribution in Figure 12.4 also shows a bump and a larger tail towards high test-statistic values, which indicates the presence

---

[1] This is a crude approximation based on the NSF funding report from 2019: `https://www.nsf.gov/about/budget/fy2019/pdf/40f_fy2019.pdf`.

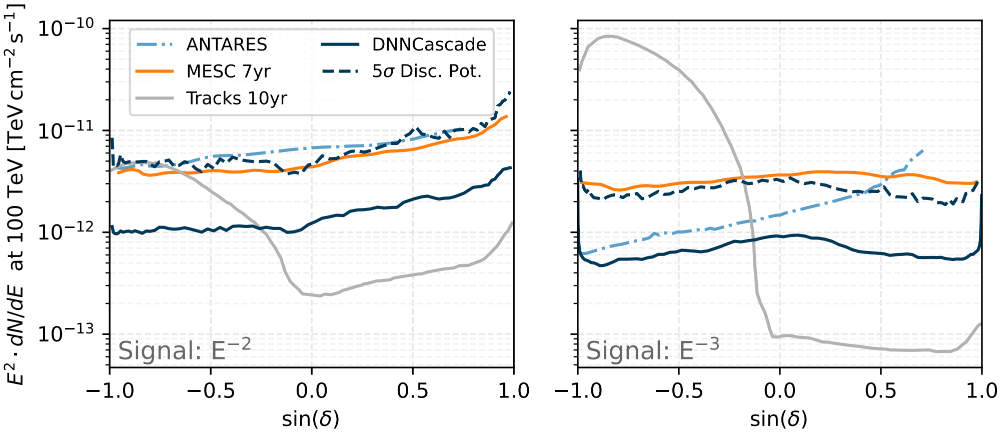


Figure 12.2: The declination dependent sensitivity to point-like sources with an assumed spectral index of $\gamma_s = 2$ (left) and $\gamma_s = 3$ (right) of the DNNCascade sample are compared to the precursor cascade analysis (MESC 7yr) [67], to IceCube tracks (Tracks 10yr) [88] and to recent results from ANTARES [146]. Sensitivities shown for the ANTARES experiment are from Ref. [147] for the E^{-2} and Ref. [148] for the E^{-3} signal injection spectrum. The 5σ discovery potential for the DNNCascade analysis is also shown.

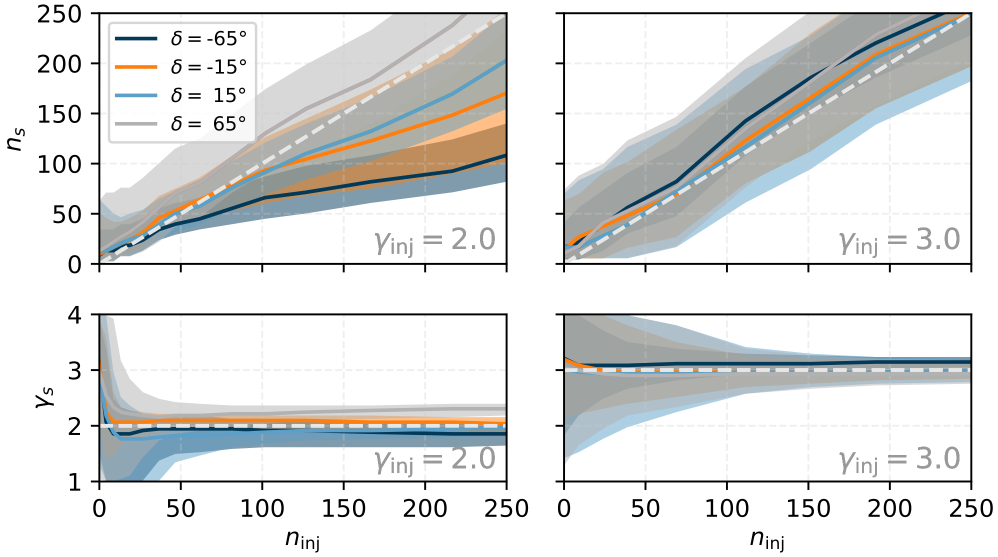


Figure 12.3: The recovery of injected source parameters is tested for four different declinations and a spectral index of $\gamma_{\mathrm{inj}} = 2$ (left column) and $\gamma_{\mathrm{inj}} = 3$ (right column). The upper panel shows the fitted number of signal events (n_s) and the lower panel the fitted spectral index (γ_s) versus the true number of injected events (n_{inj}).

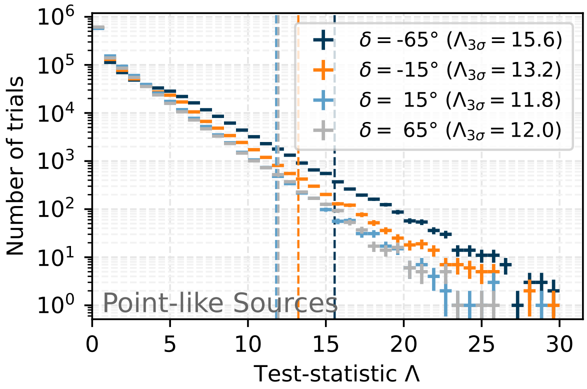


Figure 12.4: The distribution of test-statistic values is shown for background trials at four different declinations. Dashed vertical lines indicate the test-statistic value required to achieve a significance of 3σ. Differences between the distributions can arise from scrambling of individual signal-like events in the data-driven search method described in Section 10.4.

of individual signal-like events. These events are scrambled in right ascension to obtain background trials, as described in Section 10.4 [Data-driven Search Method and Systematics]. Thus, the presence of such events in a given declination band can lead to trials with larger test-statistic values if the signal-like event happens to be scrambled on top of the assumed source location. These events are typically high-energetic and therefore likely of astrophysical origin. However, by construction, the data-driven search method includes these events in the background modeling and trial generation, leading to larger background test-statistic values and thus to more conservative results. In contrast to background modeling based on MC simulation, the data-driven method is less sensitive to such events. In order to mitigate this effect, a small Gaussian smearing of 3° is applied in declination (excluding events within 8° of the poles) during scrambling. This is motivated by the assumption that background rates are expected to be stable on the order of a few degrees in declination.

Apart from degraded sensitivity, these bumps are not further problematic for the analysis as they are accounted for in the evaluation of the p-value. However, due to these bumps, the background test-statistic distribution has to be generated by brute force over millions of background scrambles; they cannot be approximated by a X^2 distribution. The distribution of correlated background trials for evaluation of the post-trial p-value of the all-sky scan is given in Figure 14.3. An example all-sky scan for a given background trial is provided in Figure 12.5.

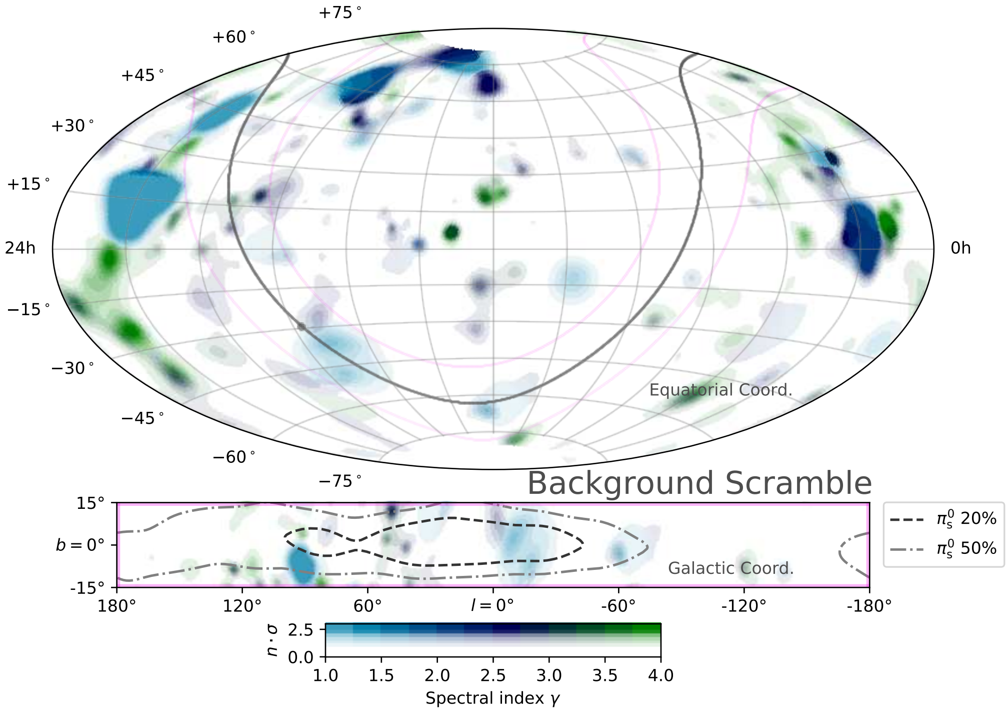


Figure 12.5: The best-fit spectral index is shown, weighted by significance as a function of direction, in equatorial coordinates (J2000) and Aitoff projection, for a background scramble of the all-sky scan. The pixel opacity is scaled by the pre-trial significance so that more opaque locations are more significant. The Galactic plane is indicated by a grey curve with a magenta band, and the region between ±15° in galactic latitude is highlighted in the lower panel. Also shown here are contours corresponding to the central 20% and 50% region of the π^0 model convolved with detector acceptance and smeared with a Gaussian corresponding to the uncertainty of a typical signal event (7°).

12.2 Catalog Stacking

The stacking of point-like sources in pre-defined catalogs, described in Section 11.3, follows similar methods to those used for the point-like source searches. The recovery of injected catalog parameters is illustrated in Figure 12.6. As demonstrated, the stacking analysis is able to recover these values accurately with only some minor bias visible for the PWN catalog for harder inejction spectra and a large number of injected signal events. Note, however, that these tests assume that the signal hypothesis is correct, i.e. that all sources in the catalog emit with the same strength and spectral index. Bias may be introduced if this assumption is violated. The sensitivity in terms of $\mathrm{E}^2 \frac{dN}{dE}$ at 100 TeV is provided in Table 14.4. The background test-statistic distribution for the stacking analyses, shown on the left of Figure 12.7, are less dependent on individual events in the background trial generation. Therefore, these distributions exhibit smaller bumps than those of the point-like source searches shown in Figure 12.4. The test-statistic distributions roughly follow X^2 distributions between 1.2 and 1.4 degrees of freedom with larger deviations at large test-statistic values, as demonstrated by the dashed lines in the left panel of Figure 12.7.

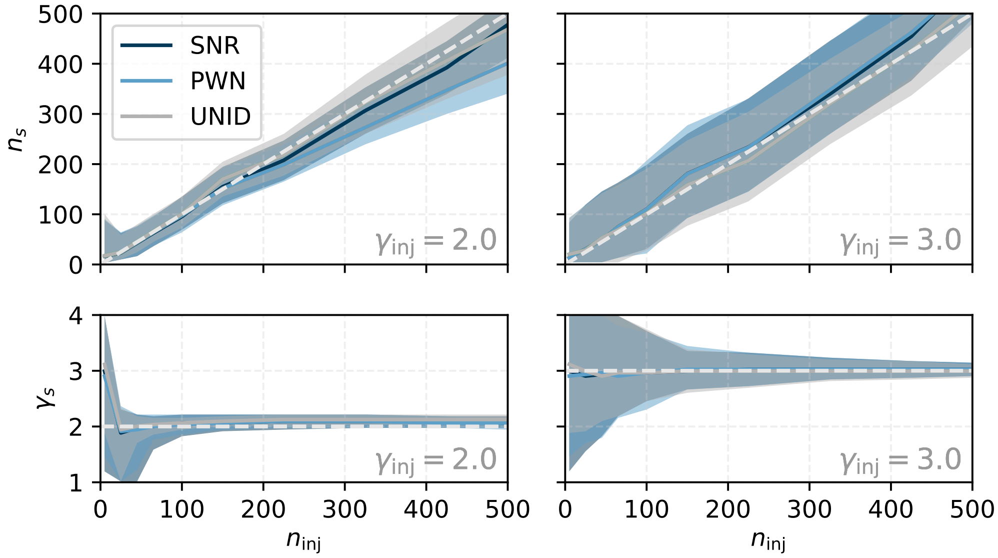


Figure 12.6: The recovery of injected source parameters is tested for the stacking analyses of each of the three source catalogs and a spectral index of $\gamma_{\mathrm{inj}} = 2$ (left column) and $\gamma_{\mathrm{inj}} = 3$ (right column). The upper panel shows the fitted number of signal events (n_s) and the lower panel the fitted spectral index (γ_s) versus the true number of injected events (n_{inj}).

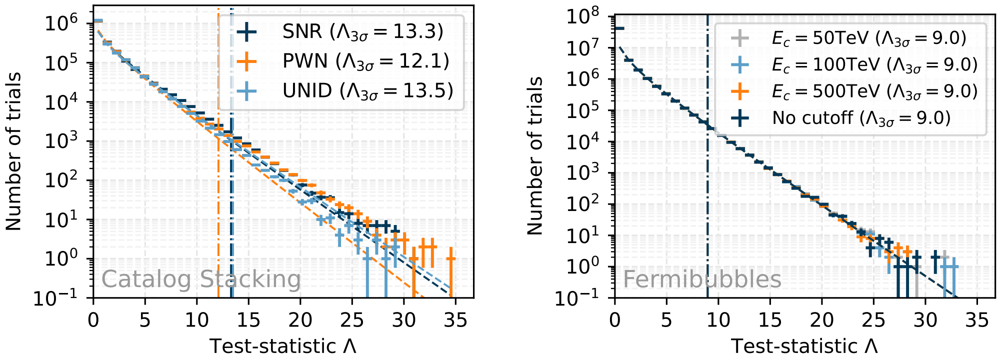


Figure 12.7: The background test-statistic distributions are shown for the catalog stacking searches (left) and the Fermibubbles analyses (right). Required test-statistic values to achieve a pre-trial significance of 3σ are indicated by vertical, dashed-dotted lines. The test-statistic distributions for the Fermibubbles are well described by a X^2 distribution with one degree of freedom as demonstrated by the dashed lines.

12.3 Template Searches

The Fermibubbles and Galactic plane template searches only fit for the number of signal events while keeping the spectrum fixed to the model prediction. Thus, these searches only have one degree of freedom, which is reflected in the background test-statistic distributions. As demonstrated on the right of Figure 12.7, the Fermibubbles test-statistic distributions for each of the spectral cutoffs are well described by a X^2 distribution with one degree of freedom. The same applies to the distributions for the Galactic plane as shown in the supplement Figure D.4. The distribution of correlated background trials for evaluation of the post-trial p-value are provided in Figure 14.2 and 14.5 for the Galactic plane searches and Fermibubbles, respectively.

Overall, the template searches are able to recover the injected number of signal events well, as illustrated in Figure 12.8. There is some larger bias visible for the Fermibubbles template corresponding to the flux without a spectral cutoff. Minor biases in the Galactic plane searches are accounted for when converting the fitted number of events to a neutrino flux, as detailed in Section 15.2 [Flux Measurement]. A comparison of the sensitivity and 5σ discovery potential between the work described in this dissertation and the precursor analysis [67] (Chapter 13) is provided in Tables 12.1 and 12.2 for the diffuse Galactic plane and Fermibubbles searches, respectively. The DNNCascade analysis is able to improve the sensitivity and discovery potential flux for each of the searches by a factor of three to four.

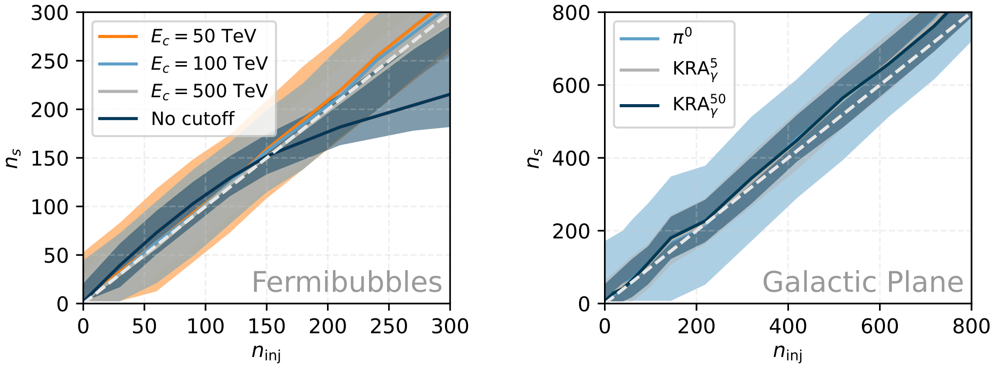


Figure 12.8: The bias in recovered number of signal events is investigated for the Fermibubbles (left) and Galactic plane (right) templates.

Table 12.1: The sensitivity (Sens.) and 5σ discovery potential (5σ DP) flux is shown for the DNNCascade analysis and the precusor selection (MESC 7yr) [67] for the three tested Galactic plane models. Values for the KRA_γ models are provided in percentages of the predicted model flux. The values for the π^0 model are the per-flavor neutrino flux given as $E^2 \frac{dN}{dE}$ at 100 TeV in units of 10^{-11} TeV cm^{-2} s^{-1}.

	DNNCascade		**MESC 7yr**	
GP Model	**Sens.**	**5σ DP**	**Sens.**	**5σ DP**
π^0	0.6	2.4	2.1	9.1
KRA_γ^5	16 %	64 %	52 %	214 %
KRA_γ^{50}	12 %	47 %	34 %	142 %

Table 12.2: The sensitivity (Sens.) and 5σ discovery potential (5σ DP) flux for the Fermibubbles search are compared between the DNNCascade anaylsis and the precusor selection (MESC 7yr) [67]. The per-flavor neutrino flux is provided as $E^2 \frac{dN}{dE}$ at 1 TeV in units of 10^{-11} TeV cm^{-2} s^{-1}.

	DNNCascade		**MESC 7yr**	
Cutoff	**Sens.**	**5σ DP**	**Sens.**	**5σ DP**
50 TeV	2.1	8.2	6.8	28
100 TeV	1.4	5.8	4.7	19
500 TeV	0.7	3.1	2.2	9.2
No cutoff	0.3	1.4	1.0	4.5

13 MESC 7yr Analysis with DNN-reco

Declaration

The MESC 7YR analysis [67] described in this chapter was performed by Michael Richman (Drexel University, PA, USA). The author of this dissertation developed the employed cascade event reconstruction, investigated impacts of systematic uncertainties, and validated the reconstruction's performance during analysis review.

The "Medium Energy Starting Cascade Sample" (MESC), precursor to the newly developed DNNCascade selection, is applied on seven years of IceCube data, resulting in 1980 selected cascade-like events. Similarly to the DNNCascade analysis, the MESC 7YR analysis [67] performs multiple searches for neutrino emission with particular emphasis on the Southern Sky and the Galactic plane. For this analysis, the CNN-based cascade event reconstruction [6], described in Part I, was developed to improve upon the angular resolution of astrophysical cascades.

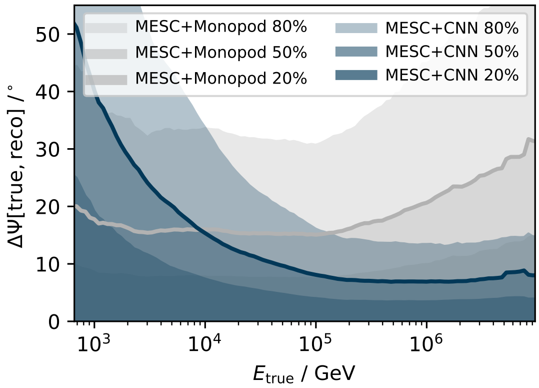


Figure 13.1: The angular resolution as a function of true neutrino energy is shown on simulated events for the MESC sample utilizing Monopod and the CNN-based method. The performance of Monopod at the time of the MESC 7YR analysis [67] is shown; the reconstruction has since been considerably improved, almost achieving the same resolution at higher energies than the CNN.

As shown in Figure 13.1, the angular resolution is improved by over a factor of two at higher energies. This energy region is dominated by astrophysical neutrinos and therefore particular important for searches of neutrino sources with harder spectra. With the help of the CNN-based reconstruction method, the sensitivity to point-like sources with a spectral index of $\gamma_s = 2$ could be reduced by about a factor of two. The improvement for sources with softer emission spectra and spectral cutoffs is less pronounced due to the reconstruction challenges the CNN faces at lower energies, as discussed in Part I. For the Galactic plane analyses, the implementation of the CNN results in an improvement of over 50%, as demonstrated in Figure 12.1.

The results of the Galactic plane analyses are summarized in Table 13.1. A minor excess of data events is observed from the Galactic plane with a significance of around 2σ. These fluctuations are consistent with the background-only hypothesis. In Figure 13.2, an *a posteriori* likelihood scan of the integrated, per-flavor Galactic flux of the diffuse π^0 template is shown. The obtained results are consistent with those presented in Ref. [127]. Note, however, that these contours are only based on an excess of signal events at the level of less than 2σ. An Aitoff projection in Figure 13.3 visualizes the results of the unbiased all-sky search. The most significant source candidate is located at $(\alpha, \delta) = (271.23^\circ, 7.78^\circ)$ with a pre-trial p-value of 1.8×10^{-3} (2.9σ), corresponding to a post-trial p-value of 0.69 [67].

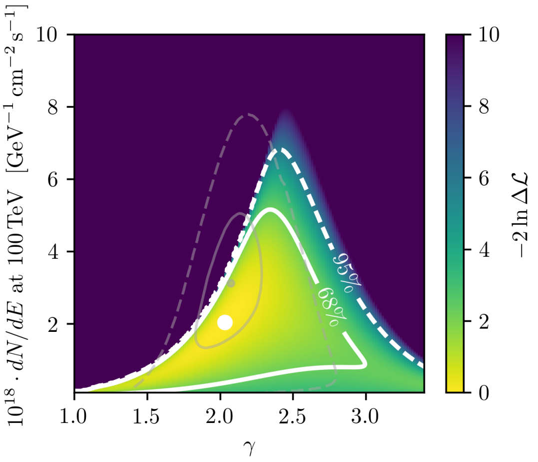


Figure 13.2: An *a posteriori* likelihood scan of the spatially-integrated, per-flavor Galactic flux is shown as a function of normalization and spectral index. Solid (dashed) contours indicate 68% (95%) confidence regions. Grey contours show the result of past IceCube work using tracks from the Northern Sky [127], for comparison. Figure and caption taken from Ref. [67].

Table 13.1: Results for the diffuse Galactic template searches performed in the MESC 7yr analysis [67] are shown. Sensitivity, fitted flux, and upper limits are given as multiples of the model prediction for KRA$_\gamma$ models, and as $E^2\, dN\,/\,dE$ at 100 TeV in units of $10^{-11} \cdot \mathrm{TeV cm^{-2} s^{-1}}$ for *Fermi*-LAT π^0 decay.

Template	p-value	$N\sigma$	Fitted Flux	Upper Limit
KRA$_\gamma^5$	0.021	2.03	0.85	1.7
KRA$_\gamma^{50}$	0.022	2.01	0.65	0.97
Fermi-LAT π^0	0.030	1.88	3.3	6.6

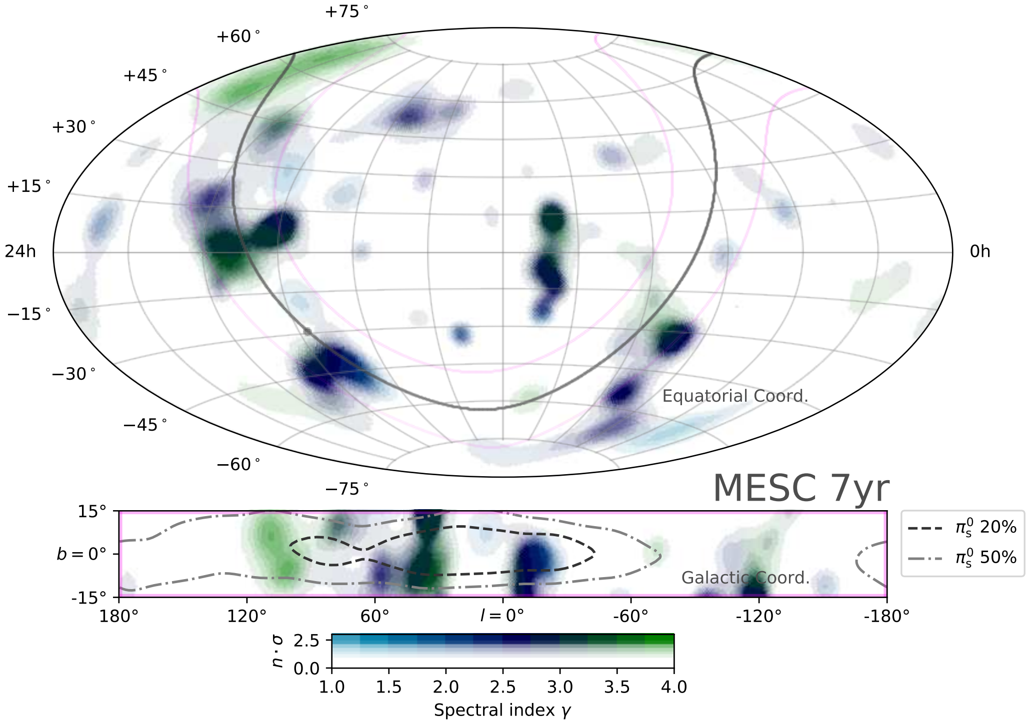


Figure 13.3: The best-fit spectral index is shown, weighted by significance as a function of direction, in equatorial coordinates (J2000) and Aitoff projection, for a background scramble of the all-sky scan. The pixel opacity is scaled by the pre-trial significance so that more opaque locations are more significant. The Galactic plane is indicated by a grey curve with a magenta band, and the region between ±15° in galactic latitude is highlighted in the lower panel. Also shown here are contours corresponding to the central 20% and 50% region of the π^0 model convolved with detector acceptance and smeared with a Gaussian corresponding to the uncertainty of a typical signal event (7°).

14

DNNCascade Analysis Results

An overview of the DNNCascade analysis results is provided in Table 14.1. Further results on the individual analyses, described in Chapter 11 [Performed Searches], are presented in the following sections. Interpretation and further context is provided in Chapter 15 [Implications of Galactic Neutrinos].

Diffuse GP Analyses	**Sensitivity Φ**	**p-value**	**Best-Fit Φ**
π^0	5.98	$1.26{\times}10^{-6}$ (4.71σ)	$21.8\,^{+5.3}_{-4.9}$
KRA_γ^5	$0.16{\times}$MF	$6.13{\times}10^{-6}$ (4.37σ)	$0.55^{+0.18}_{-0.15}{\times}$MF
KRA_γ^{50}	$0.11{\times}$MF	$3.72{\times}10^{-5}$ (3.96σ)	$0.37^{+0.13}_{-0.11}{\times}$MF
Stacking Analyses		**p-value**	
SNR		$5.90{\times}10^{-4}$ (3.24σ)*	
PWN		$5.93{\times}10^{-4}$ (3.24σ)*	
UNID		$3.39{\times}10^{-4}$ (3.40σ)*	
Summary of Other Analyses		**p-value**	
Fermi Bubbles		0.06 (1.52σ)	
Source List		0.22 (0.77σ)	
Hotspot (North)		0.28 (0.58σ)	
Hotspot (South)		0.46 (0.10σ)	

Table 14.1: An overview of the DNNCascade analysis results are presented. Sensitivity and best-fit flux normalization is given in units of model flux (MF) for KRA_γ templates and per neutrino flavor as $\mathrm{E}^2\,\frac{dN}{dE}$ at 100 TeV in units of 10^{-12} TeV cm^{-2} s^{-1} for the π^0 analyses. Pre-trial p-values for each individual result are shown for the three diffuse Galactic plane analyses and three stacking analyses, and post-trial p-values are shown for the summaries of other analyses. Due to the spatial overlap of the stacking catalogs with the diffuse Galactic plane templates, strong correlations between these searches are expected. Significance values for the stacking catalog results, denoted with an "*", indicate that these results are consistent with the diffuse Galactic plane template search results, as discussed in Section 15.7 [Model Confusion and Unresolved Sources]. Table and caption are adopted from Ref. [1].

14.1 Galactic Plane Searches

Of the performed searches, the searches for diffuse neutrino emission from the Galactic plane result in the most significant rejection of the background-only hypothesis at 4.71σ, 4.37σ, and 3.96σ, corresponding to 748, 276, and 211 signal events (n_s) for the π^0, KRA_γ^5, and KRA_γ^{50} templates, respectively. The contribution to the overall test-statistic value for each of the Galactic plane searches is visualized in Figure 14.1 and compared to contributions of a typical background trial.

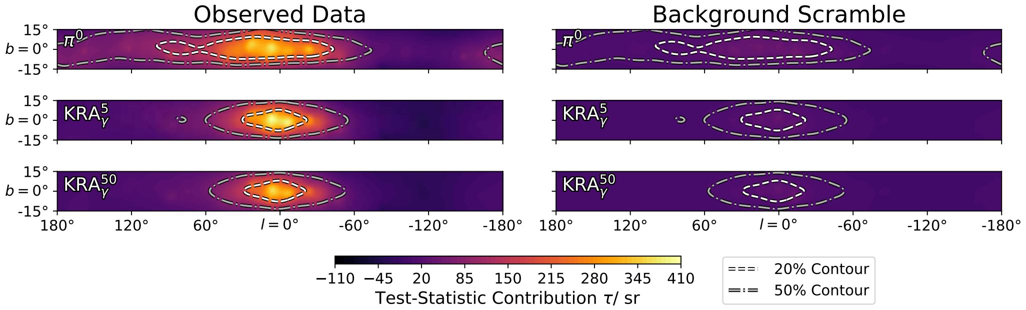


Figure 14.1: The contribution to the test-statistic Λ is illustrated in galactic coordinates for each of the three tested Galactic plane models. The overall test-statistic value can be obtained by integration over the sky. The contribution for the observed data (left) is compared to the contribution for a typical background scramble (right). For illustrative purposes, overlaid are the central model contours that contain 20% and 50% of the expected flux. The 50% contours contain about 1.64 sr, 0.70 sr and 0.65 sr for the π^0, KRA_γ^5 and KRA_γ^{50} models respectively. Figure and caption are taken from Ref. [1].

A conservative estimate of the post-trial significance for the most significant result, accounting for the three GP tests performed, results in a significance of 4.48σ. However, this assumes that these tests are independent of each other. When accounting for correlations between the analyses, as shown in Figure 14.2, a trial-corrected p-value of 2.81×10^{-6} is obtained, corresponding to 4.54σ. For this calculation, background trials are performed and evaluated for each of the three Galactic plane templates. The most significant of the three pre-trial results (Λ_{pre}) is chosen for each of the background trials to construct the background distribution illustrated in Figure 14.2. The trial-corrected p-value is obtained by evaluating the fraction of background trials that are as least as significant as the analysis result, illustrated by a solid, vertical line.

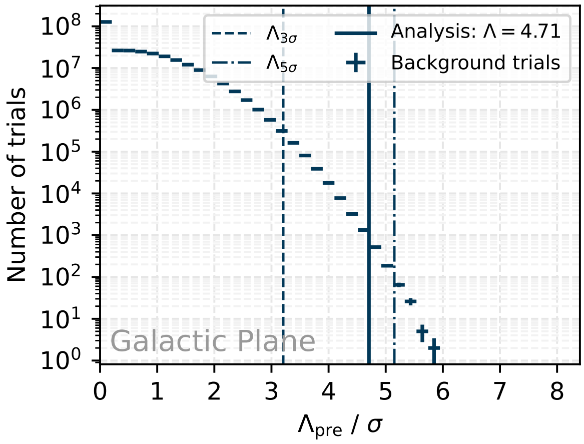


Figure 14.2: The most significant pre-trial significance (Λ_{pre}) is chosen as a test-statistic to generate background trials for the Galactic plane searches. A dashed (dashed-dotted) vertical line indicate the 3σ (5σ) level. The analysis result is illustrated by the solid, vertical line.

14.2 All-sky Search

The results of the unbiased all-sky search are summarized in Table 14.2 and Figure 14.4. No significant point-like source is found after accounting for trials; individual warm spots are consistent with background fluctuations. However, there is an accumulation of over-fluctuations with consistent spectral index visible along the Galactic plane as shown in Figure 14.4 and further discussed in Section 15.6 [Interpretation of All-sky Maps]. The background distribution utilized to evaluate the post-trial significance is provided in Figure 14.3. Further plots, including visualizations in Galactic coordinates, are provided in Section 15.6 and the supplement Section D.2.

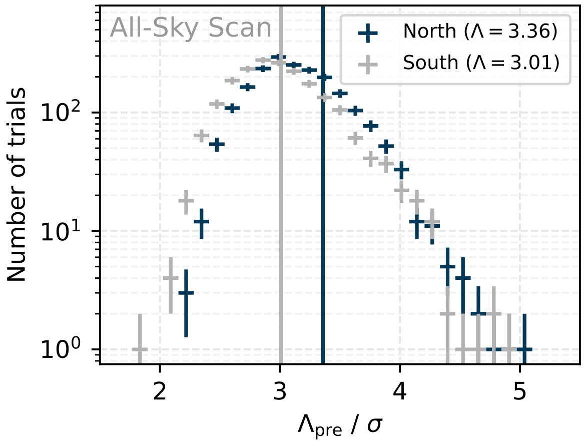


Figure 14.3: The most significant pre-trial significance (Λ_{pre}) is chosen as a test-statistic to generate background trials for the each of the hemispheres in the all-sky search. A dashed (dashed-dotted) vertical line indicate the 3σ (5σ) level. The analysis result is illustrated by the solid, vertical line.

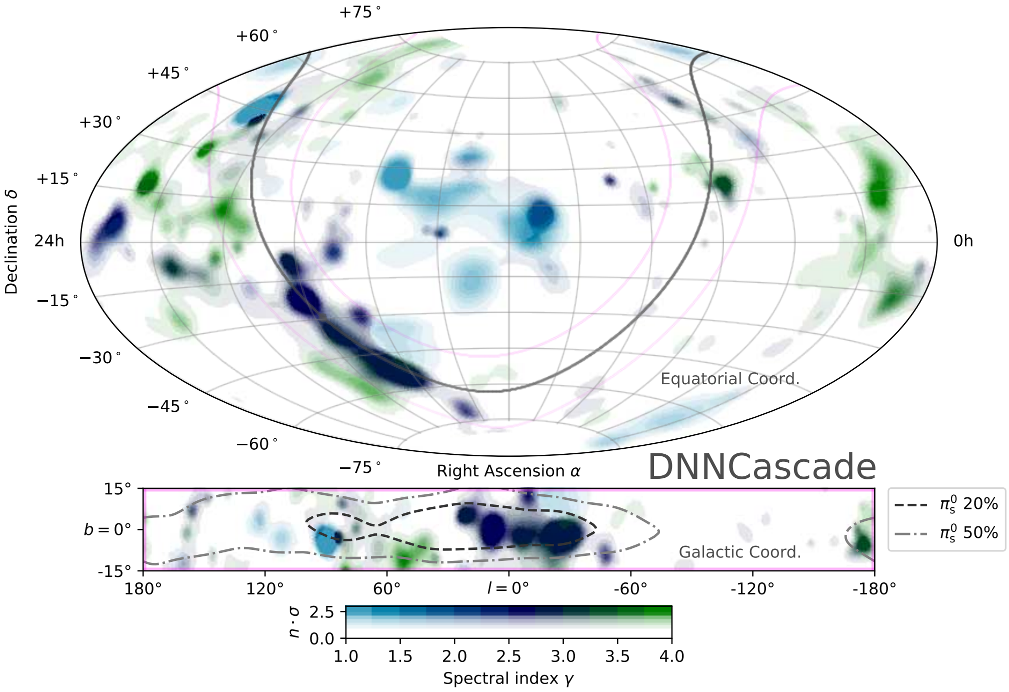


Figure 14.4: The best-fit spectral index is shown, weighted by significance as a function of direction, in equatorial coordinates (J2000) and Aitoff projection, for the all-sky search. The pixel opacity is scaled by the pre-trial significance so that more opaque locations are more significant. The Galactic plane is indicated by a grey curve with a magenta band, and the region between ±15° in galactic latitude is highlighted in the lower panel. Also shown here are contours corresponding to the central 20% and 50% region of the π^0 model convolved with detector acceptance and smeared with a Gaussian corresponding to the uncertainty of a typical signal event (7°). Figure and caption adopted from Ref. [1].

Table 14.2: Results of the all-sky search for the most significant point in the Northern and Southern hemisphere. [1]

Analysis	$\alpha[^\circ]$	$\delta[^\circ]$	n_s	γ	p-value$_{\text{pre}}$	p-value$_{\text{post}}$
Hotspot (North)	337.9	17.6	213.7	3.6	3.92×10^{-4}	0.28
Hotspot (South)	248.1	-50.9	90.2	2.9	1.31×10^{-3}	0.46

14.3 Fermibubbles

Results for the Fermibubbles searches are shown in Table 14.3. No significant excess is found. After trial correction, utilizing the background test-statistic distribution in Figure 14.5, a post-trial p-value of 0.06 is found, corresponding to 1.52σ.

Table 14.3: The best fit number of signal events (n_s), pre-trial significance, and 90% Upper Limits (UL) are shown for the Fermi Bubbles model with various exponential cutoffs. The per-flavor neutrino flux upper limits are given as $E^2 \frac{dN}{dE}$ at 1 TeV in units of 10^{-12} TeV cm^{-2} s^{-1}. [1]

Cutoff	n_s	Pre-trial p-value	UL Φ
50 TeV	96.0	0.03 (1.88σ)	<52.5
100 TeV	70.9	0.05 (1.65σ)	<34.1
500 TeV	34.4	0.12 (1.17σ)	<15.0
No Cutoff	23.7	0.14 (1.06σ)	<5.9

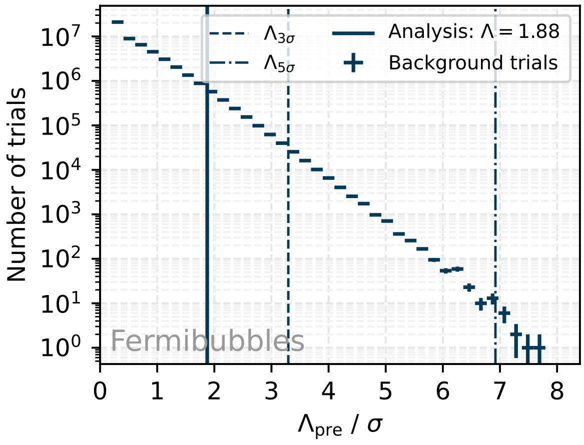


Figure 14.5: The most significant pre-trial significance (Λ_{pre}) is chosen as a test-statistic to generate background trials for the Fermibubbles searches. A dashed (dashed-dotted) vertical line indicate the 3σ (5σ) level. The analysis result is illustrated by the solid, vertical line.

14.4 Stacking Catalogs

Table 14.4 shows an overview of the stacking searches for each of the three tested Galactic catalogs. Each of the catalogs is observed with a significance above 3σ. However, as discussed in Section 15.7 [Model Confusion and Unresolved Sources], these analyses are correlated to the diffuse GP searches due to the large angular uncertainty of cascade events. The results obtained here are consistent with expectations derived from an assumed diffuse π^0 emission scenario.

Table 14.4: Sensitivity, number of signal events (n_s), pre-trial significance, best-fit spectrum (γ), best-fit flux (Φ), and 90% flux upper limits (UL Φ) for the Galactic stacking catalog analyses are provided. Upper limits are evaluated with respect to a source emitting following an E^{-2} spectrum. The per-flavor neutrino flux sensitivity, best-fit flux, and flux upper limits are given as $E^2 \frac{dN}{dE}$ at 100 TeV in units of 10^{-12} TeV cm^{-2} s^{-1} for the entire catalog of sources. Best-fit spectral index γ and flux Φ have been corrected for bias. The 68% confidence intervals, profiling along each dimension, are quoted as described in Section 15.2 [Flux Measurement]. [1]

Catalog	Sens. Φ	n_s	Pre-trial p-value	Best-Fit γ	Best-Fit Φ	UL Φ
SNR	2.24	218.6	5.90×10^{-4} (3.24σ)	$2.7^{+0.3}_{-0.3}$	$6.2^{+5.1}_{-3.8}$	<9.0
PWN	2.25	279.6	5.93×10^{-4} (3.24σ)	$3.0^{+0.3}_{-0.3}$	$3.8^{+4.3}_{-2.5}$	<9.5
UNID	1.89	238.4	3.39×10^{-4} (3.40σ)	$2.8^{+0.4}_{-0.3}$	$5.0^{+4.3}_{-3.5}$	<7.8

14.5 Source List

No significant emission from any of the sources in the pre-defined source list is found. The largest significance is seen from PMN J1650-5044 located at $(\alpha, \delta) = (252.6^\circ, -50.8^\circ)$ with a pre-trial p-value of 0.002 (2.9σ), corresponding to a post-trial p-value of 0.06 (1.52σ). The candidate neutrino sources identified by IceCube, the flaring blazar TXS 0506+056 [113, 114] and the active galaxy NGC 1068 [88, 89], are included in the tested source list. Neither of these sources are observed significantly with the DNNCascade analysis. The number of signal events is found to be $n_s = 0$ for TXS 0506+056, leading to upper limits for an E^{-2}-powerlaw of $\Phi_2 \leq 0.61 \cdot 10^{-12}$TeVcm^{-2}s^{-1} in terms of $E^2 \frac{dN}{dE}$ at 100 TeV. For NGC 1068, the best-fit results in $n_s = 87.1$ and $\gamma_s = 4.0$ with a p-value of 0.25. An upper limit for an E^{-3}-powerlaw of $\Phi_3 \leq 1.36 \cdot 10^{-12}$TeVcm^{-2}s^{-1} is set in terms of $E^2 \frac{dN}{dE}$ at 100 TeV. The cascade analysis is less sensitive to point-like sources in the Northern Hemisphere than track-based analyses [88, 89], as visualized in the declination-dependent sensitivity in Figure 12.2. The DNNCascade analysis results for these sources align with expectations derived from injection trials. The set upper limits are well above the reported fluxes [89, 113] and thus consistent with the prior measurements. A detailed list of results for each of the sources is provided in the supplement Table D.2.

15 Implications of Galactic Neutrinos

The results of the DNNCascade analysis, presented in Chapter 14, demonstrate strong evidence for the observation of high-energy neutrinos from the direction of the Galactic plane. The background-only hypothesis is rejected at 4.5σ after accounting for trials. Observing such an accumulation of neutrinos along the spatial π^0 template by random chance is therefore extremely unlikely with a p-value of 2.81×10^{-6}. Despite this strong neutrino signal from the Galactic plane, the underlying sources cannot be clearly identified with this analysis, as further discussed in Section 15.7 [Model Confusion and Unresolved Sources]. The observed signal is consistent with diffuse emission from the Galactic plane, potentially in combination with emission by a population of unresolved sources.

15.1 Energy Range of Galactic Plane Measurement

The Galactic plane template analyses keep the energy spectrum fixed to each of the models' predictions, while only the normalization is fit for. As long as the assumed energy spectrum matches reality, this procedure produces correct estimates on the flux normalization. However, if the true energy spectrum differs from the spectrum tested for, biases are introduced and the resulting best-fit flux normalization is not easily interpreted.

For example, the π^0 template assumes an energy spectrum that follows a single, unbroken power-law, described in Eq. 10.3, with a spectral index of $\gamma_s = 2.7$. However, a spectral cutoff is expected to occur in the Galactic CR spectrum, as more energetic particles cannot be confined and thus accelerated to such energies in the Galactic plane. Depending on the energy at which this spectral cutoff occurs, the π^0 best-fit flux normalization must strike a compromise between explaining the excess observed at lower energies and the lack of expected events at highest energies. The reported spectra for the GP analyses must therefore never be understood as a valid flux measurement at a given energy, but only as a measurement of an assumed spectrum over an entire energy range. Thus, when interpreting the reported best-fit flux normalization and in particular comparing to other flux measurements, these caveats need to be considered.

There are different ways in how such energy ranges can be defined, each with unique properties. In any case, the maximal range is bounded by the scope of the DNNCascade dataset, which spans from approximately 500 GeV to multiple PeV. The "sensitivity range" is defined as the central energy range at which the analysis sensitivity drops by a certain, pre-defined percentage when limiting signal injection to this specified range. The "sensitivity range" therefore provides information on what energies the analysis is sensitive to. Such an energy range is thus only dependent on MC simulation. An alternative definition that is informed by experimental data and that aims to answer the question: "What energy range supports the measurement?" is referred to as the "TS-based energy range". For this energy range, the distribution of experimental data events is investigated, weighted by their contribution to the overall test-statistic (TS or Λ) value of the analysis. A weighted PDF in true neutrino energy is constructed by accumulating the individual energy PDFs of each data event, $P(E_\nu|E_{\mathrm{rec}}, \delta_{\mathrm{rec}})$, which describe the likely true neutrino energies as a function of reconstructed energy and declination. Here, only events with a positive test-statistic contribution are included. The same energy range can also be constructed on pseudo data, generated by simulation. Comparison of the TS-based ranges on experimental and simulated data allows for some insight on the true energy spectrum.

Table 15.1: Quantiles of the energy ranges based on the sensitivity and contribution to the test-statistic are provided for the GP analyses. Quantiles are given in units of TeV.

Model	**Method**	**5%**	**16%**	**50%**	**84%**	**95%**
π^0	Sens.	1.5	3.6	—	99	282
	TS	0.80	1.30	4.26	21.15	68.87
	TS (MC)	0.796 ± 0.004	1.30 ± 0.01	4.4 ± 0.1	24.1 ± 1.8	82.8 ± 8.5
KRA$_\gamma^5$	Sens.	5.2	10.1	—	81	145
	TS	0.97	1.89	7.42	37.14	99.71
	TS (MC)	0.98 ± 0.01	1.94 ± 0.03	8.2 ± 0.5	46.5 ± 6.5	120.0 ± 16.8
KRA$_\gamma^{50}$	Sens.	6.9	18.2	—	162	357
	TS	1.02	2.08	8.88	50.58	145.38
	TS (MC)	1.03 ± 0.01	2.14 ± 0.05	10.4 ± 0.9	66.5 ± 13.5	230.4 ± 26.8

Results for each of the mentioned energy ranges are summarized in Table 15.1. Comparisons between the TS-based ranges obtained from experimental data and simulation show good agreement for the lower boundaries of the central 68% and 90% ranges. However, the median and upper quantiles of the energy ranges show a shift towards higher energies in simulation. This suggests that less high-energy events are observed from the Galactic plane than expected in each of the three tested models. Potential explanations for this include — but are not limited to — a spectral break or cutoff in the Galactic CR population at lower energies than expected, or a contribution of unresolved, Galactic neutrino sources that emit at lower energies. Dedicated analyses are required to obtain a meaningful and reliable spectral measurement that is less dependent on spectral assumptions and potential contamination of unresolved sources.

15.2 Flux Measurement

The best-fit number of signal events (n_s) can be converted to a flux by accounting for the effective area of the DNNCascade sample (Figure 7.10). Contrary to the significance calculation via the data-driven search method (Section 10.4 [Data-driven Search Method and Systematics]), the flux conversion depends on MC simulation and it is therefore affected by systematic uncertainties. A confidence interval is constructed by inversion of the likelihood ratio test via the approach by Feldman and Cousins [135]. In order to account for systematics, a concrete realisation of systematic parameters is sampled for every generated trial via the SnowStorm method [74].

For every true flux normalization (and spectral index for the stacking analyses) a likelihood ratio test is performed to answer the question if the specified flux can be excluded at a given confidence level. The following test-statistic is used:

$$\tau = -2\ln \frac{L_0\left(n_s = n_{\mathrm{inj}}^{\dagger}\,[, \gamma_s = \gamma_{\mathrm{inj}}^{\dagger}]\right)}{L_1\left(n_s = n_{\mathrm{fit}}\,[, \gamma_s = \gamma_{\mathrm{fit}}]\right)}, \tag{15.1}$$

where $n_{\mathrm{inj}}^{\dagger}$ and $\gamma_{\mathrm{inj}}^{\dagger}$ are the bias-corrected number of injected signal events and injected spectral index. The bias-correction is applied such that the true number of injected signal events n_{inj} is corrected to provide the value the analysis would fit for this injected flux, and vice versa for the spectral index. L_0 and L_1 are the "signal-subtracted" likelihood function from Eq. (10.6) for the null-hypothesis (the flux is equal to the injected one) and the alternative hypothesis (the flux is different from the injected one). The parameters n_{fit} and γ_{fit} are the best-fit parameters that maximize the likelihood for each generated trial.

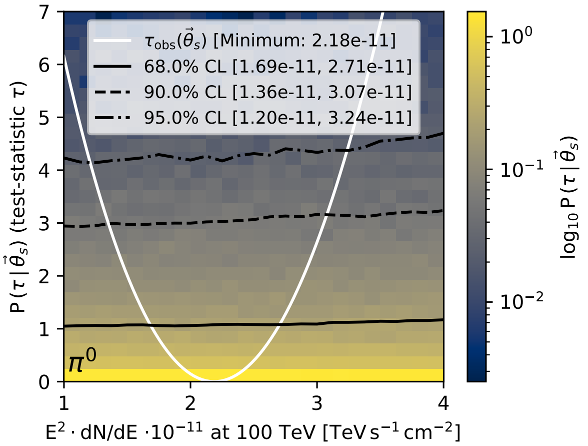


Figure 15.1: The confidence interval construction for the flux measurement of the π^0 model is illustrated. The distribution of test-statistic values (τ) from Eq. (15.1) is shown (y-axis and color map) for each true injected flux normalization (x-axis). Critical values are shown as black lines for the 68%, 90%, and 95% confidence levels. The observed test-statistic (τ_{obs}), evaluated on experimental data, is provided in white.

The test-statistic τ from Eq. (15.1) is evaluated for many generated trials to construct a distribution of test-statistic values for every true injected flux normalization as shown in Figure 15.1. Trials are generated by randomizing the experimental data (Section 10.4) and then injecting signal events from MC simulation corresponding to the assumed true flux given by $\vec{\theta}_s = (n_{\mathrm{inj}}, \gamma_{\mathrm{inj}})$. The generated distribution of test-statistic values allows for the determination of critical values for a given confidence level as a function of true flux normalization (shown as black lines). The critical values indicate the quantiles of the test-statistic distribution, i.e. the 90% critical value is the test-statistic value $\tau_{90\%}$ at which 90% of trials have a τ value less or equal to $\tau_{90\%}$. These critical values are then compared to the test-statistic observed in experimental data (τ_{obs}). The observed test-statistic as a function of flux, $\tau_{\mathrm{obs}}(\vec{\theta}_s)$, is computed by evaluating Eq. (15.1) on the experimental data without applying randomization and signal injection as performed in the aforementioned trials. The final confidence interval is constructed by selecting the interval of assumed true flux values that result in an observed test-statistic τ_{obs} lower than the corresponding critical values. The resulting interval describes the collection of assumed flux values that cannot be rejected at the given confidence level.

An analogous procedure is performed for the construction of the two-dimensional confidence intervals for the stacking analyses, as illustrated in the left column of supplement Figure D.6. These distributions are adequately described by X^2 distributions following Wilk's theorem [86], which allows to draw smooth contours as demonstrated in Figure 15.2 and on the right column of the supplement Figure D.6. In lack of applicability of Wilk's theorem, reduction of statistical fluctuations in the contours is only possible by utilization of excessive amounts of computational resources due to the large number of required trials in higher dimensions. Figures highlighting the construction of confidence intervals for the Galactic plane models and stacking catalogs are provided in the supplement Section D.3.

Utilizing the computed confidence intervals, the sky-integrated best-fit flux for each of the Galactic plane models is illustrated in Figure 15.3 and compared to the predicted energy spectra. The highlighted range for each observation corresponds to the central 90% neutrino energy range of observed events according to the "TS-based energy range", detailed in Section 15.1. While these ranges provide insight into the neutrino energies that support the measurement, the fit is performed on the entire sample of events, with neutrino energies up to a few PeV. Model-to-model flux comparisons can vary depending on the regions of the sky considered. The KRA_γ best-fit model normalizations are lower than the model expectations, and could be an indication of a spectral cutoff that is inconsistent with the 5 PeV and 50 PeV values assumed. The simple extrapolation of the π^0 model from GeV energies to 100 TeV (Section 11.5) predicts a neutrino flux that is a factor of ~5 below the observed best-fit flux, indicating limitations of the simplified model. Nevertheless, as discussed in Section 15.5, the π^0 best-fit flux appears consistent with recent

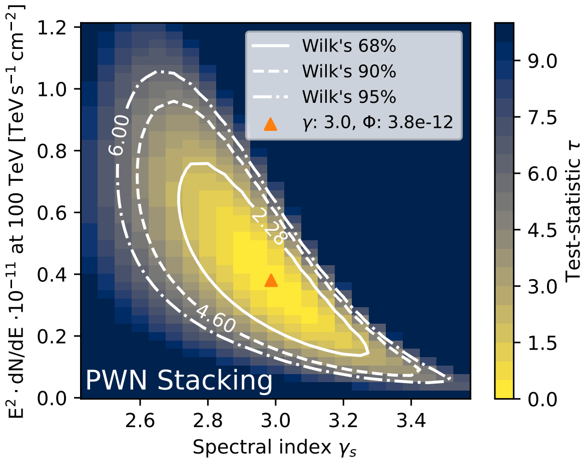


Figure 15.2: The confidence interval construction for the flux measurement of the PWN catalog stacking is illustrated. The distribution of observed test-statistic values (τ_{obs}) from Eq. (15.1) is shown as a function of spectral index γ_s (x-axis) and flux normalization (y-axis). Contours for 68%, 90%, and 95% confidence levels are drawn in white according to Wilk's theorem by utilizing a X^2 distribution of two degrees of freedom.

observations of 100 TeV gamma rays from the Tibet Air Shower Array [149]. The π^0 model mismatch could arise from propagation and spectral differences for cosmic rays in the Galactic center region or from contributions from unresolved neutrino sources. Further details are discussed in Section 15.3. The results of this analysis confirm the presence of Galactic plane neutrino emission for every model tested.

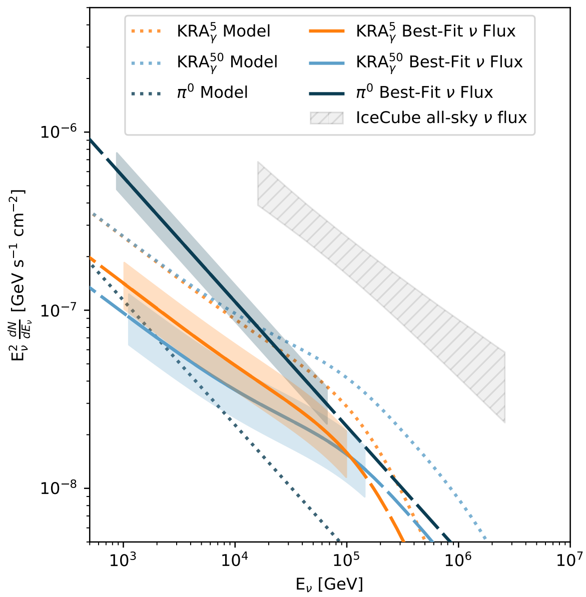


Figure 15.3: The energy scaled, sky-integrated, per-flavor neutrino flux is shown versus energy for each of the Galactic plane models and corresponding best-fit values in the region of the central 90% energy range that contribute to the observed significance for each model. These results are based on the all-sky (4π sr) template and presented as an all-sky flux. 1σ uncertainty on the measured flux normalization is shown for each observation. Also shown is the measured flux from IceCube [91] scaled to all an all-sky flux by multiplying by 4π, with corresponding 1σ uncertainty. Figure and caption adopted from Ref. [1].

The neutrinos observed from the Galactic plane contribute to the all-sky astrophysical diffuse flux previously observed by IceCube [91, 95, 132]. Shown in Figure 15.3 is a recent all-sky astrophysical flux measurement utilizing cascade-like events [91].

The measurement from each of the Galactic template models corresponds to roughly 10% of the astrophysical flux at 30 TeV[1]. For these comparisons, it should be noted that the observed Galactic flux is integrated over the entire sky, but local flux contributions along the central region of the Galactic plane will be higher.

15.3 Extrapolation of the *Fermi*-LAT π^0 Measurement

As detailed in Section 11.5 [Diffuse Galactic Plane Searches], the π^0 model utilized in this analysis is an extrapolation of *Fermi*-LAT measurements at GeV gamma-ray energies. The π^0 model is therefore to be considered as a well motivated spatial template, but not necessarily as a self-consistent model of diffuse Galactic plane emission. In particular, energy- and sptially-dependent properties of cosmic ray injection and propagation are disregarded in this extrapolation. Figure 15.4 visualizes the extrapolation from GeV gamma rays up to 100 TeV (orange, dotted). The underlying *Fermi*-LAT data (black points) are utilized to fit the $^{S}\mathtt{S}^{Z}\mathtt{4}^{R}\mathtt{20}^{T}\mathtt{150}^{C}\mathtt{5}$ model [117] via the cosmic ray transport package `GALPROP` [141]. The total MC prediction of the $^{S}\mathtt{S}^{Z}\mathtt{4}^{R}\mathtt{20}^{T}\mathtt{150}^{C}\mathtt{5}$ model (solid, black) and the component corresponding to the π^0 gamma-ray component (dashed, black) are also shown. While the model agrees well with the measurements up to a few GeV, an excess of data events over the MC prediction is observed at higher energies.

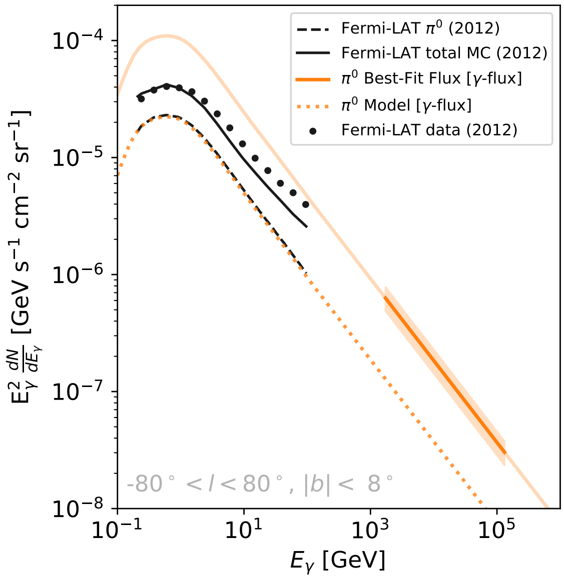


Figure 15.4: The energy scaled gamma ray flux, averaged over the Galactic plane region from $|l| < 80°$ and $|b| < 8°$, is shown versus energy for the π^0 model utilized in this analysis (orange, dotted lines), as well as for the underlying *Fermi*-LAT measurements. The total expected gamma ray flux of the $^{S}\mathtt{S}^{Z}\mathtt{4}^{R}\mathtt{20}^{T}\mathtt{150}^{C}\mathtt{5}$ model is illustrated in the solid, dark line and the component arising from π^0 decay is shown in the dashed, dark line. The best-fit values from this work (solid, orange) is shown in terms of gamma ray flux together with the "TS-based energy range" (shaded region).

---

[1] The aforementioned caveats on the spectral assumption and energy range apply for this comparison.

The authors of Ref. [117] state that this discrepancy may arise from the contribution by point sources such as pulsars, supernova remnants, and pulsar wind nebulae, as only a fraction of these sources are detected by *Fermi*-LAT. Alternate explanations include the existence of young cosmic ray sources in the inner Milky Way that inject with a harder spectrum relative to the steady-state cosmic ray population; as well as uncertainties in the correction of solar modulation effects, which directly impact the prediction of the π^0 component [117]. Furthermore, there are indications that this mis-modeling may be correlated to the cosmic ray composition itself (Figure 30 in Ref. [117]). While proton spectra are well described up to 100 TeV, the helium spectrum begins to deviate above a few hundred GeV. In any case, the observed discrepancy seems to increase with energy, indicating that extrapolation to even higher energies must be taken with care. These factors could contribute to the apparent over-shooting of the diffuse neutrino measurement via the π^0 template with respect to the $^{S}\texttt{S}^{Z}\texttt{4}^{R}\texttt{20}^{T}\texttt{150}^{C}\texttt{5}$ model prediction.

15.4 Prior Galactic Neutrino Searches

Prior IceCube analyses have searched for neutrino emission from the Galactic plane. The precursor cascade analysis (MESC 7yr) [67], described in Chapter 13, observed an excess of neutrino events from the Galactic plane over background at the level of 2σ significance. Another analysis utilizing seven years of IceCube tracks from the Northern Sky [127] rejects the background-only hypothesis at 1.5σ. With the result of the DNNCascade analysis, these mild over-fluctuations can be interpreted as early signs of an emerging Galactic neutrino flux. Although the observed signal is weak in these prior analyses, and the fitted flux therefore not necessarily robust, comparisons between these analyses may shed further light on the properties of the observed Galactic neutrino flux.

The *a posteriori* likelihood scans of the π^0 model in the DNNCascade and MESC 7yr analysis are compared in Figure 15.5. Contours are created in a simplified approach via Wilk's theorem [86] and without consideration of systematic uncertainties. Thus, caveats apply to the following comparisons, as systematic uncertainties and adequate handling of biases in the analyses, as detailed in Section 15.2 [Flux Measurement], may further impact the illustrated contours. Overall, the provided contours are consistent to each other given the large uncertainties on the prior measurement. A minor shift ($\sim 1\sigma$) towards softer spectra is observed in the DNNCascade measurement. A larger tension on the level of $\sim 2\sigma$ is seen for the prior track analysis (contours shown in Figure 13.2), which favors harder spectra. Note that the spectral measurement is based on the assumption of an unbroken power-law. A cosmic ray cutoff at lower energies will affect and bias the measured spectrum. The fitted flux normalization at 100 TeV reference energy is consistent between all analyses.

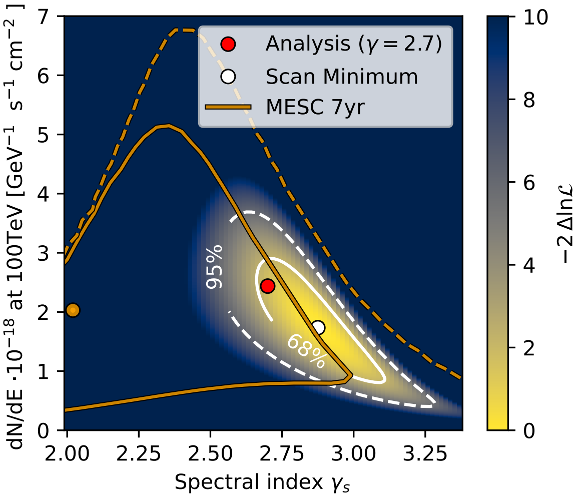


Figure 15.5: An *a posteriori* likelihood scan of the spatially-integrated, per-flavor Galactic flux for the π^0 model is shown as a function of normalization and spectral index. Contours (white lines) are constructed via Wilk's theorem [86], without an additional bias-correction. This is a simplifed procedure from the one described in Section 15.2. Systematic uncertainties are not included. The results of the precursor analysis (MESC 7yr) [67] from Figure 13.2 are shown in orange.

Both of the MESC 7yr [67] and DNNCascade analyses exploit cascade topologies and are therefore expected to be correlated due to a larger overlap of events. However, the lower energy threshold of the DNNCascade sample may result in differences in the spectral measurement.

In addition to a higher energy threshold, track analyses primarily probe the Northern Sky, whereas cascade analyses are most sensitive to the Galactic center region in the Southern Sky. On top of the different hemisphere, track-based analyses will also depend more on the spatial emission profile. Recently, an analysis based on ten years of the starting track sample, ESTES [94], was performed. Conclusions of this analysis are still in progress, but preliminary results indicate that the analysis may begin to differentiate between the three tested Galactic plane models, favoring the DNNCascade π^0 result over the KRAγ ones. At this point, it is unclear whether or not this is due to spectral or spatial differences in the emission models, or due to an additional contribution of unresolved sources. If the tension between track- and cascade-based analyses persist in upcoming measurements, further conclusions about the the spatial and spectral properties of the Galactic neutrino flux may be drawn.

Apart from IceCube analyses, a search for neutrino emission from the Galactic ridge [130], defined as the central Galactic plane region within 30° in longitude and 2° in latitude, has been performed. This search also finds an excess over background at the level of $\sim 2\sigma$. The reported flux for this sky region is compared in Figure 15.6 to the corresponding flux measured in this work. Note that the ANTARES measurement is a dedicated measurement from this sky region, while the DNNCascade result is an alternate representation of the sky-integrated best-fit values, scaled to this region of the sky according to the relative template flux. The

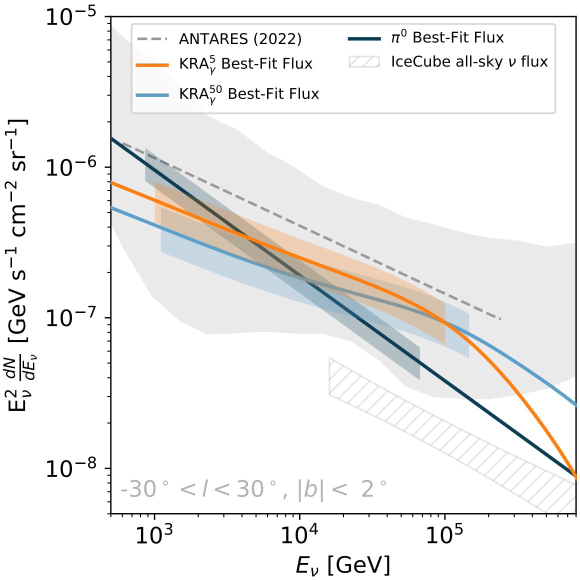


Figure 15.6: The energy scaled, per-flavor neutrino flux, averaged over the Galactic center region ($|l| < 30°$ and $|b| < 2°$), is shown versus energy for each of the Galactic plane best-fit values and corresponding 1σ uncertainty bands. These fluxes are not independent measurements in these parts of the sky, but rather an alternative presentation of the sky-integrated best-fit values. A dedicated measurement for this sky region from ANTARES [130] is shown in grey. Also shown is the measured isotropic, astrophysical flux from IceCube [91] with corresponding 1σ uncertainty.

68% confidence level of the ANTARES measurement (shaded region), albeit subject to large uncertainties, is well consistent with the observed flux in this work.

15.5 Diffuse hadronic emission?

The observed neutrino flux from the Galactic plane provides strong evidence for the existence of hadronic interaction scenarios in the Milky Way. In combination with gamma-ray measurements, the neutrinos may provide insight to the understanding of the processes responsible for the acceleration and propagation of cosmic rays.

Diffuse Galactic emission is believed to be dominated by hadronic interactions, due to the larger abundance of cosmic ray nuclei in comparison to leptons and due to the larger Inverse Compton and synchrotron losses that prevent energetic leptons of reaching far beyond their acceleration sites [119, 121, 133, 150]. Diffuse gamma-ray measurements up to multiple hundred TeV [149, 151] from the Tibet ASγ and LHAASO collaborations therefore imply the presence of a corresponding neutrino flux. Under the assumption of a hadronuclear (pp) scenario, the measured gamma-ray flux is converted to an expected neutrino flux, following the relation outlined in Refs. [152–154], while neglecting gamma-ray attenuation. The resulting, neutrino equivalent flux is shown in Figure 15.7 for three different sky regions.

As in Figure 15.6, the presented best-fit flux of the π^0 and KRA$_\gamma$ models is obtained by scaling the sky-integrated best-fit values with the relative template contribution from each of the defined sky regions. The flux measurement of the π^0 model is in good agreement with (preliminary) diffuse LHAASO-KM2A measurements [151] at tens of TeV and with gamma ray observations from the Tibet Air Shower Array [149] at 100 TeV, as demonstrated in the upper and middle panels in Figure 15.7. The lower panel includes a prediction for the diffuse Galactic neutrino flux [155], based on the gamma-ray observations from the Tibet Air Shower Array [149]. Each of the measured fluxes for the three tested models is in good agreement with this prediction.

The neutrino flux prediction is based on the assumption that the highest energy bin of the Tibet measurement includes a contribution ($\sim 40\%$) from unresolved sources originating from the Cygnus Cocoon region [156, 157]. The diffuse emission model is therefore only built based on 60% of the stated Tibet flux above 400 TeV. The DNNCascade π^0 measurement only reaches up to ~ 100 TeV and can therefore not constrain the highest energy data point by Tibet. Given that the lower data points align with the observed neutrino flux, it is plausible that the observed gamma-ray flux in this energy range is indeed of hadronic origin. However, at this point, the DNNCascade analysis cannot exclude the contribution of point-like sources to the observed Galactic plane flux (Section 15.7 [Model Confusion and Unresolved Sources]). Therefore, the true diffuse neutrino flux is likely to be lower as derived in this measurement, which limits the conclusions that can be drawn in regard to the nature of the observed gamma-ray flux.

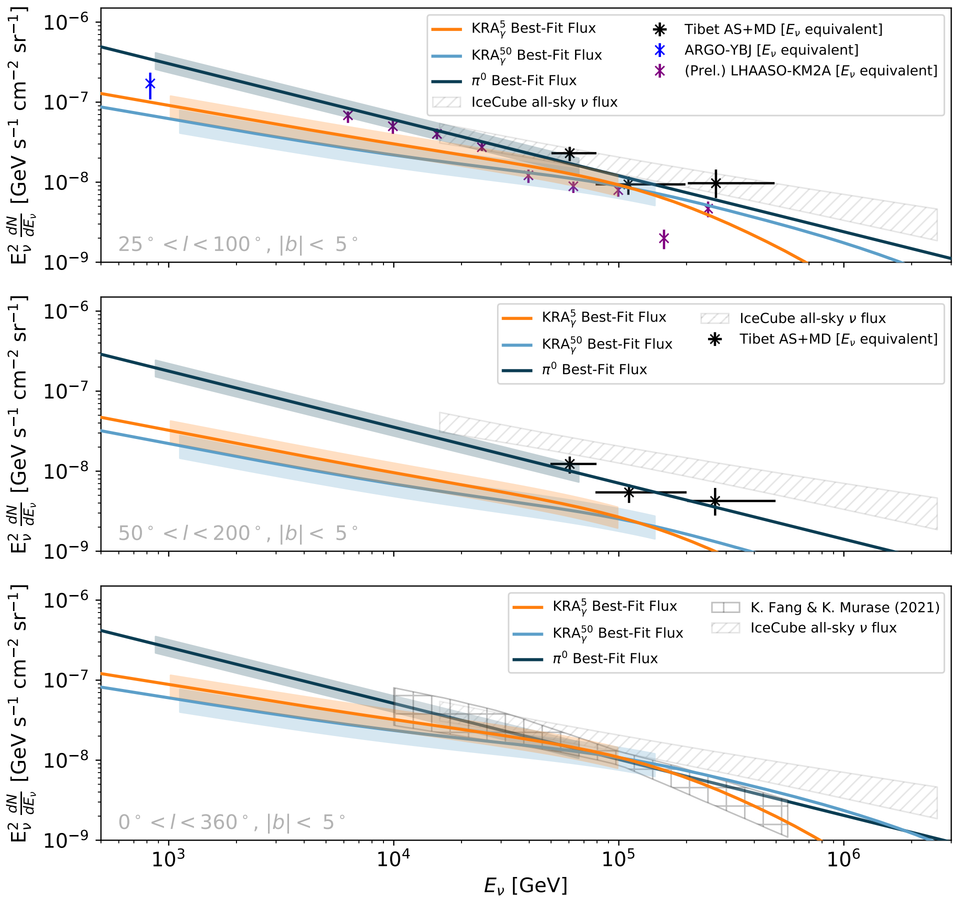


Figure 15.7: The energy scaled, per-flavor neutrino flux averaged over different areas of the sky is shown versus energy for each of the Galactic plane best-fit values and corresponding 1σ uncertainty bands. The shaded regions indicate the central 90% TS-based energy range (Section 15.1). The average flux values are obtained by multiplying the total, sky-integrated neutrino flux from Table 14.1 and Figure 15.3 with the relative template contribution from each of the three depicted sky regions, as indicated on the lower left of each panel. These fluxes are therefore not independent measurements in these parts of the sky, but an alternative presentation of the sky-integrated best-fit values. The top panels include gamma-ray measurements from the Tibet Air Shower Array [149], ARGO-YBJ [158], and preliminary results from LHAASO-KM2A [151], converted to a neutrino flux assuming a hadronuclear (pp) scenario, following the relation outlined in Refs. [152–154] while neglecting gamma-ray attenuation. The bottom panel includes a prediction for the diffuse Galactic neutrino flux (K. Fang & K. Murase (2021) [155]), derived from recent gamma-ray measurements. Also shown is the measured isotropic, astrophysical flux from IceCube [91] with corresponding 1σ uncertainty (hatched). Figure and caption are modified from Ref. [1].

15.6 Interpretation of All-sky Maps

As presented in Section 14.2 [All-sky Search], no significant point-like emission is found in the unbiased all-sky scan or source list search after accounting for trials. Nevertheless, the all-sky maps do exhibit interesting properties. While no individual warm spot is significant by itself, there is an accumulation of signal over background events along the Galactic plane and in particular in the Galactic center region, as shown in Figures 14.4, 15.8, and 15.9.

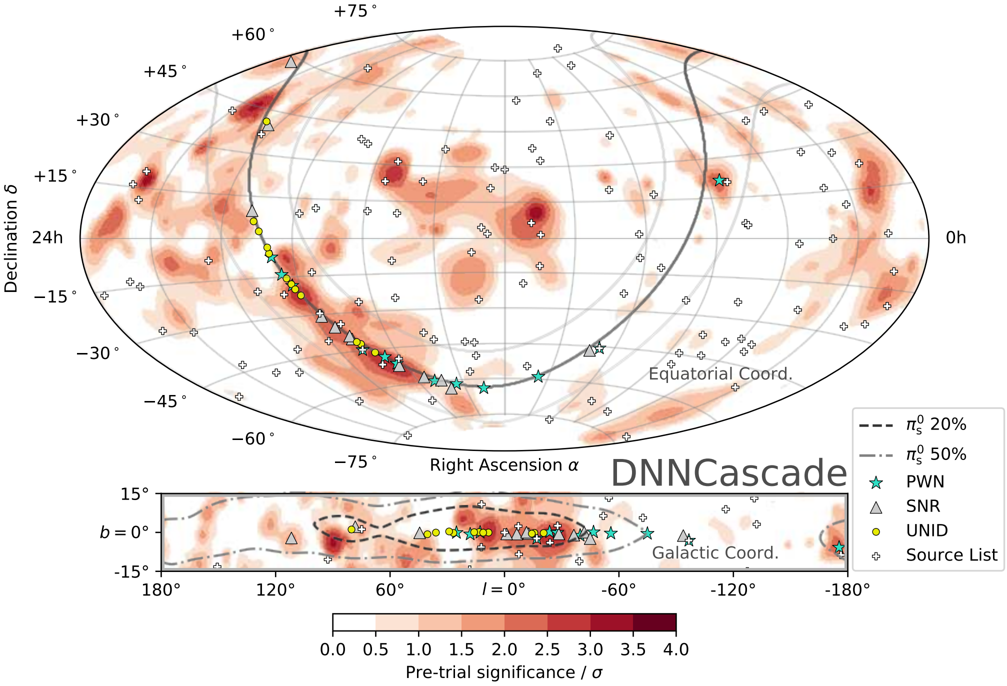


Figure 15.8: The pre-trial significance is shown as a function of direction, in equatorial coordinates (J2000) and Aitoff projection, for the all-sky search. The Galactic plane is indicated by a grey curve, and the region between ±15° in galactic latitude is highlighted in the lower panel. Overlaid are the sources from the stacking catalog and source list searches. Also shown here are contours corresponding to the central 20% and 50% region of the π^0 model convolved with detector acceptance and smeared with a Gaussian corresponding to the uncertainty of a typical signal event (7°).

These over-fluctuations are further interesting because the best-fit spectral index (Figure 14.4) at these locations are not only consistent to each other, but also consistent to expectations from diffuse Galactic neutrino emission models. If driven by statistical background fluctuations, these warm spots are likely to exhibit a larger variance in the fitted spectral index. Thus, the all-sky maps visually support the hypothesis of neutrino emission from the Galactic plane. The non-detection in the point-like source searches also implies that the emission that is present in the Galactic template analyses is not due to a single point-like source (further discussion on the contribution of unresolved sources in Section 15.7).

As illustrated in Figures 15.8 and 15.9, some locations in the sky have excess emission over the background expectations, including some in spatial coincidence with or nearby known gamma-ray emitters, such as the Crab Nebula, the Cygnus Cocoon [156, 157], and the blazar 3C 454.3. However, after accounting for trials factors, no single point in the map is statistically significant (Table 14.2).

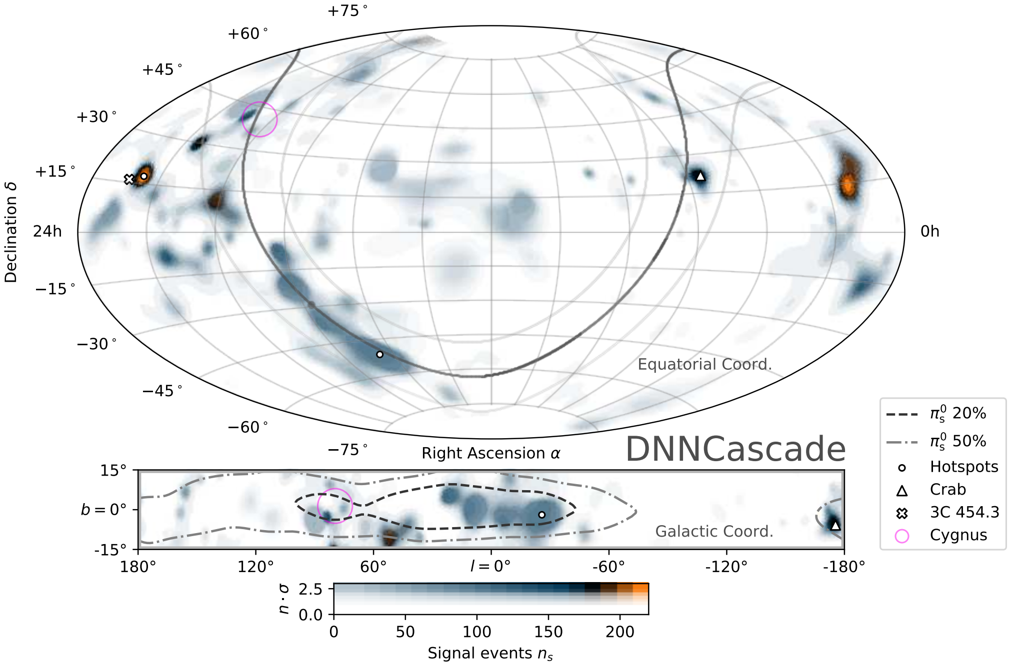


Figure 15.9: Similar to Figure 15.8, but showing the best-fit number of signal events, weighted by significance as a function of direction, in equatorial coordinates (J2000) and Aitoff projection, for the all-sky search. The pixel opacity is scaled by the pre-trial significance so that more opaque locations are more significant. Overlaid are the most-significant points of the all-sky search in either hemisphere (Hotspots), and the location of the Crab Nebula, Cygnus Cocoon, and the blazar 3C 454.3.

The ability to spatially resolve a neutrino source in the all-sky scan depends on the expected number of signal events and their energy distribution. Due to the typical 5° to 20° angular uncertainty of individual cascade events, the combination of many such events is required to improve the ability to resolve a source in the sky. Assuming the source is a point-like source with parameter values as measured at these warm spots, a resolution of a few degree is expected. Note that the current analysis utilizes symmetrical uncertainty contours in zenith and azimuth, which may underestimate uncertainties in right ascension, as discussed in Section 10.2 [Analysis Tool Simplifications].

The locations of the most significant spots in each hemisphere of the the all-sky search are shown in Figure 15.9. The hottest spot in the Southern Sky is located on the Galactic plane near the Galactic center region, consistent with expectations given the observed Galactic neutrino flux. Although not significant, the hotspot in the Northern Sky is interesting due to its measured properties and its vicinity to the blazar 3C 454.3 ($\varDelta\delta \approx 1.5^\circ, \varDelta\alpha \approx 5.8^\circ$), one of the brightest blazars in the *Fermi*-LAT gamma-ray sky. The best-fit number of signal events is particularly large with a value of $n_s = 213.7$, shown in Figure 15.9, weighted by the pre-trial significance. A corresponding soft spectral index with $\gamma_s = 3.6$ is fitted, which could also be an indication for a spectral cutoff at lower energies, if this emission were from an assumed source. Although not taken advantage of in the analysis, the background trial distributions depend on the fitted spectral index; the distributions are shifted towards smaller test-statistic values for a larger (softer) fitted spectral index. Hence, the all-sky scan may be slightly biased towards selecting hotspots with a harder spectral index. The significance of the Northern hotspot could possibly be slightly underestimated. However, accounting for this bias is not likely to have a large impact on the result.

15.7 Model Confusion and Unresolved Sources

The DNNCascade analysis finds strong evidence for the emission of neutrinos from the Galactic plane. Presented searches for diffuse emission models and catalog stacking searches of Galactic sources reject the background-only hypothesis at the 4.5σ and $> 3\sigma$ significance level, respectively (Table 14.1). Multiple emission scenarios are possible to explain the observed neutrino flux.

Source ambiguity can be investigated using model injection tests, where the best-fit neutrino signal from one source search is simulated, and the expected results in all other analyses are examined. These injection tests may be evaluated by comparing the median significance obtained in the simulated signal trials against the observed significance in the performed search (Figure 15.10), and by performing a two-tailed hypothesis test to infer if the observed significance is consistent with expectations for each injected model (Figure 15.11).

Injecting a signal with the observed best-fit signal strength observed in the π^0 model analysis produces a median significance that is consistent with the best fit values for all other tested hypotheses within the expected statistical fluctuations (second row in Figure 15.10). This includes the 3σ excess observed in Galactic source catalog

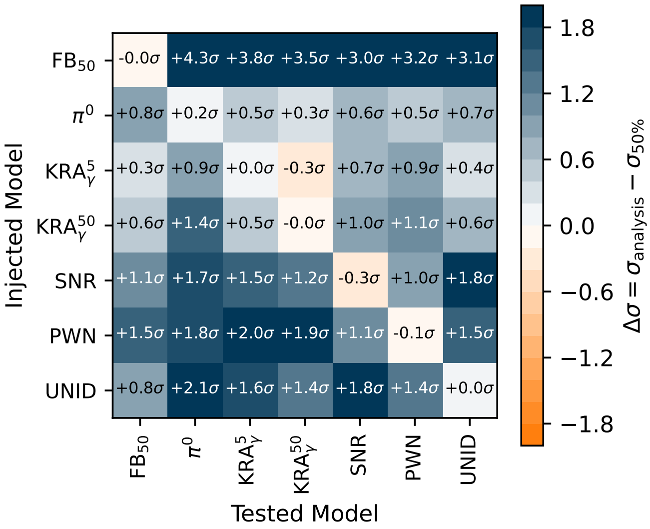


Figure 15.10: Model injection tests are performed, where the best-fit flux from one search is simulated, and the expected results in all other analyses are examined. The difference between the observed significance from the analysis (σ_{analysis}) and the median significance over all simulated signal trials ($\sigma_{50\%}$) is shown as a matrix. Large deviations indicate that the injected model (row) is not capable of explaining the results seen in the applied source search (column).

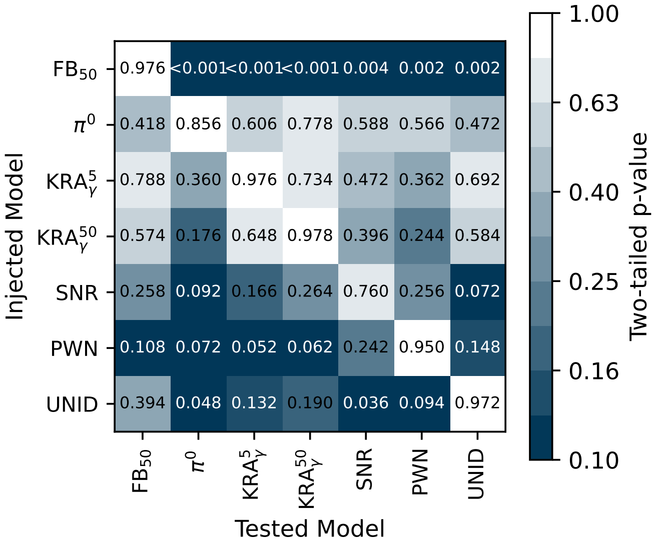


Figure 15.11: Model injection tests are evaluated for various combinations of injected (rows) and tested (columns) models, similar to Figure 15.10, but by utilizing a two-tailed hypothesis test as a comparison metric. The null hypothesis states that the observed analysis results are explained by the injected model. Thus, small p-values reject the injected model (row) of being a likely explanation for the results seen by the applied source search (column).

searches. Similarly, injecting the best-fit π^0 flux results in a two-sided p-value of 0.418 in the most pessimistic case for the Fermibubbles results.

Individually injecting the best-fit excess of any one of the tested Galactic source catalogs at the flux level observed does not recover the observed π^0 or KRA_γ model results with p-values below 0.1 for the two-tailed test. However, the angular resolution of the sample and the small number of equally weighted sources included in these catalogs does not constrain emissions from these broad source populations. It is plausible that many independently contributing sources from this source-dense region could show a comparable result to diffuse emission from interactions in the interstellar medium.

While the observed significances seem to favor the π^0 over the KRA_γ models in Figures 15.10 and 15.11, the differences are not sufficient to make conclusive statements on the underlying properties of the observed Galactic plane flux. The model injection tests only test for the simple case that only one model may be correct. It is probable that the observed neutrino flux originates from a combination of diffuse emission and a contribution of unresolved sources.

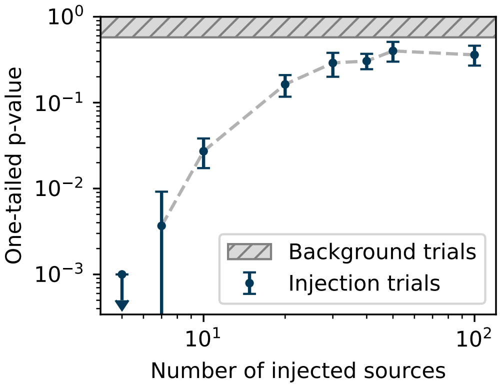


Figure 15.12: A simulation study is performed to infer the minimum number of sources required to explain the π^0 model results, while not excluded by the observed all-sky search analysis, which does not find a contribution of strong point-like sources. A one-tailed hypothesis test is performed to reject the null hypothesis of k sources being responsible for the observed Galactic plane neutrino flux.

This raises the question if any constraints can be placed on the contribution of unresolved sources. To examine this, a simulation study is performed in which signal events are injected for a defined number of sources randomly sampled according to the spatial PDF of the π^0 model. Each sampled source location is assumed to emit the same flux, and the sum over all sampled sources is set to match the best-fit π^0 flux. A one-tailed hypothesis test is evaluated for every k injected sources with

- H_0^k: the observed hottest all-sky pre-trial p-value is distributed according to k sources injected via the spatial π^0 template.
- H_1^k: the observed hottest all-sky pre-trial p-value is smaller (more significant) than expected for the k injected sources.

Hence, the simulation study aims to infer the number of sources that the observed flux must be minimally distributed over, given the observed all-sky scan results. If the total flux is distributed over less sources, the all-sky scan is likely to have seen more significant results than observed, and the hypothesis H_0^k of k sources can be rejected. As demonstrated in Figure 15.12, values of $k \leq 10$ are rejected at the 95% confidence level. Hence, the observed flux must at least be distributed over $\gtrsim 10$ sources (95% confidence level). More stringent constraints on the role of unresolved sources may be obtained by the inclusion of track-based samples due to their superior pointing resolution.

Supplement Figures D.7, D.8, and D.9 show example all-sky search maps for simulated injection trials for each of the Galactic plane and Galactic stacking catalog searches. In each injection trial, the best-fit flux is injected on top of the scrambled background events. These maps illustrate the expected scale and distribution of warm spots. Even for diffuse emission models (left panels), the sky-scans are expected to observe a number of distinct warm spots. This is due to the limited statistics of the data sample leading to a relatively small number of signal events distributed over the sky.

16 Outlook and Future Measurements

The observed excess of neutrinos from the Galactic plane provides strong evidence for the Milky Way as a source of high-energy neutrinos. This observation thus complements IceCube's measurement of the diffuse extra-galactic neutrino flux and it also confirms our present understanding of the interactions of cosmic rays within our Galaxy. Further characterization of this Galactic component of the astrophysical neutrino flux will require identification of sources and their spectra, and an evaluation of the strength of the diffuse Galactic emission in this matter-dense and source-dense region of the sky.

From Evidence to Detection

In the field of particle and astro-particle physics, it is common to refer to observations at the $\geq 3\sigma$ significance level as "evidence", while the phrasing of "detection" is typically reserved for significances above 5σ, corresponding to p-values of less than 3×10^{-7}. A natural question arises when the DNNCascade analysis will reach this detection threshold of 5σ significance.

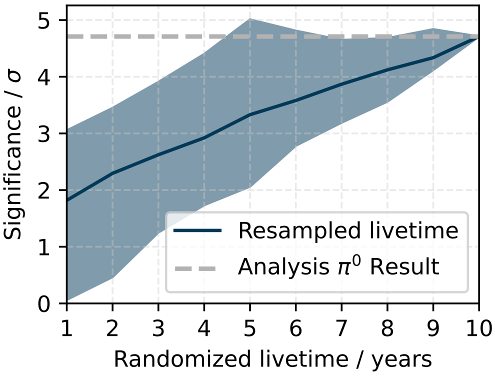


Figure 16.1: The evolution of the π^0 analysis significance over time is evaluated by applying the analysis to randomized subsamples of the DNNCascade dataset. The shaded region indicates the 90% confidence interval. Note that the uncertainty vanishes for the full livetime, as only one realization — the 10 year DNNCascade dataset — can be drawn due to the sampling without replacement. Extrapolations will be subject to large uncertainties.

In order to assess the time development of the significance of the π^0 Galactic plane analysis, randomized subsets of the data sample are selected, corresponding to livetimes between one and ten years. This randomization allows to retain blindness for time dependent analyses in the Galactic plane region, while mimicking the assumed steady diffuse Galactic signal. The π^0 template analysis is then performed on each of these subsamples, arriving at the significances shown in Figure 16.1. Similary, the sensitivity and discovery potential is investigated in Figure 16.2 for dataset subsamples corresponding to livetimes of three to ten years.

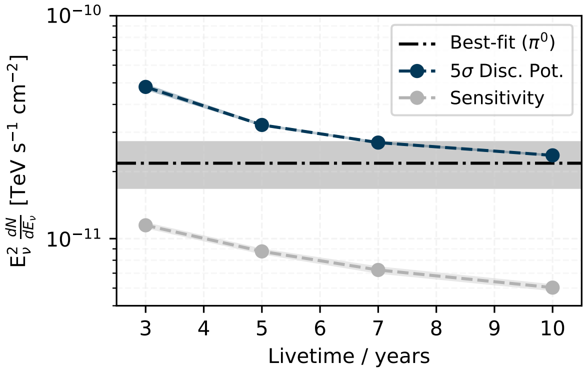


Figure 16.2: The sensitivity and discovery potential are shown for the Galactic plane π^0 analysis as a function of dataset livetime. Also shown is the best-fit flux and corresponding 1σ uncertainty band. Sensitivity and discovery potential roughly scale with the inverse square root of the dataset livetime.

The analysis presented in this work was performed on ten years of IceCube data, collected between May 2011 and May 2021. Assuming that the functional time dependence in Figures 16.1 and 16.2 may be extrapolated, two to three additional years of dataset livetime will greatly increase the probability of surpassing the detection threshold with this analysis. In addition, newer developments in event reconstruction and uncertainty estimation, modeling in the analysis itself, as well as detector and ice calibration, may provide further analysis improvements beyond the increased livetime.

Spectral and Spatial Measurements

Deepening our understanding of the observed Galactic neutrino signal will require the dissection into constituent components and spectral measurements thereof. Identifying the contribution of currently unresolved sources will be crucial to this process. The observed neutrino excess is strong enough to enable initial spectral and spatial measurements. Despite the relatively large angular uncertainties of cascades, the DNNCascade sample can already be utilized to constrain the broad regions from which the Galactic neutrino excess originates from. For such an analysis, the Galactic plane may be divided in multiple sections — preferably aligned to regions studied in existing gamma-ray measurements. This will allow for more appropriate comparisons than shown in Figures 15.6 and 15.7. If sensitivity allows, additional separation into energy intervals will further advance the understanding and modeling of cosmic rays and resulting hadronic interactions in our Galaxy.

Beyond the DNNCascade Sample

Cascades are one of two main detection channels in IceCube. Combined analyses utilizing the DNNCascade sample as well as starting and through-going track-based event samples are therefore a natural path forward. Tracks will provide an independent verification channel of the observed signal, but will — in particular — also allow to further constrain the spatial emission profile and contribution of unresolved sources. Previous track-based analyses [127] with seven years of IceCube data observed a minor excess of neutrinos from the Galactic plane at the $\sim 2\sigma$ significance level. Updates to these analyses are currently ongoing. In addition to IceCube, contributions from other neutrino telescopes with advantageous coverage of the Southern Sky [159–161] will help to understand the properties of the observed Galactic neutrino signal.

Part IV

Conclusions and Outlook

17

A new Milestone of Neutrino Astronomy

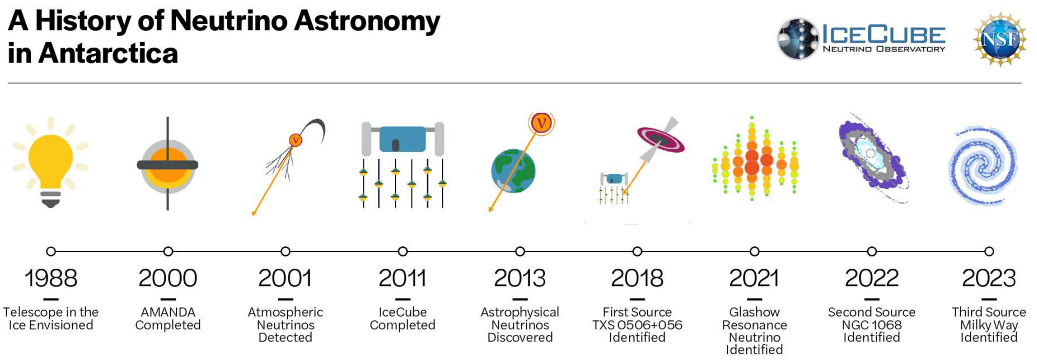


Figure 17.1: An overview of the history of neutrino astronomy in Antarctica is illustrated. Image credit: IceCube/NSF.

From the perception of the first neutrino telescopes in the late 1980s, the construction of AMANDA [162] in 2000 and the completion of ANTARES [146] and IceCube [26] in 2008 and 2011, to recent discoveries: the rising field of neutrino astronomy is evolving with an increasing pace. In the last decade, major milestones have been achieved (Figure 17.1) with the detection of the astrophysical flux in 2013 [93], the first evidence for point-like neutrino sources in 2018 [113, 114] and 2022 [88, 89], and the identification of a Glashow resonance event in 2021 [111] as an independent confirmation for the existence of astrophysical neutrinos. The observation of the Galactic plane in high-energy neutrinos, presented in this dissertation, constitutes yet another achievement on this scientific quest, complementing prior measurements and confirming our current understanding of cosmic rays in the Milky Way. This observation enables a multi-messenger perspective on our Galaxy, previously confined to measurements in the electro-magnetic spectrum. As such, a new era of Galactic neutrino astronomy is established, providing new insight to the pursuit of scientific questions regarding the sources of high-energy cosmic rays and their propagation, the structure and morphology of matter in the Milky Way, physics in extreme environments, and the properties of dark matter [133]. These advancements are enabled by the tools developed in this dissertation.

This dissertation introduces two novel software frameworks based on deep learning to advance key tasks in IceCube involving event classification and reconstruction. Based on these frameworks, an improved selection of neutrino-induced cascade events is constructed, increasing selection efficiency by over an order of magnitude, resulting in the most sensitive cascade neutrino sample to date. The developed sample is leveraged in a search for neutrino emission from the Galactic plane, leading to the first observation of high-energy neutrinos from our own galaxy — the Milky

Way. With this analysis, the most-significant high-energy neutrino source to date is found, rejecting the null hypothesis at the 4.5σ significance level.

Although much of the discourse in this dissertation is focused on the observed Galactic neutrino flux, the developed methods enable a wide range of applications, reaching beyond searches for astrophysical sources with cascade events. The tools enable the construction of dedicated event samples and the reconstruction of arbitrary event topologies and desired observables. Thus, novel analyses may be pursued that were previously unattainable. The DNNCascade analysis on Galactic neutrino emission is one such analysis that demonstrates the impact these tools may have: savings of 75 years of detector livetime and multiple $100 million, and the establishment of the Milky Way as a source of high-energy neutrinos.

18 Future Developments

Developed methods in this dissertation set the foundation for various analyses that were previously not sensitive enough or simply infeasible to conduct. The DNNCascade GP analysis is one example for which the developed methods made a considerable impact. The observation of the Galactic neutrino flux will prompt further measurements to dissect the observed GP signal, as discussed in Chapter 16 [Outlook and Future Measurements]. However, these are only a certain class of analyses that have become possible.

Direct Measurement of the Prompt Atmospheric Neutrino Flux

An intrinsic background to the astrophysical flux component in the diffuse neutrino flux measured by IceCube [91, 95, 100] is given by neutrinos produced in Earth's atmosphere as a result of air showers induced by cosmic ray interactions [90]. These atmospheric neutrinos may further be separated into two components referred to as conventional and prompt, where the prompt component is distinguished by a harder energy spectrum extending to higher energies. Interpretation of the astrophysical flux in this energy regime requires an understanding of the prompt atmospheric neutrino flux. However, large uncertainties pertain to the prompt component and thus far, it has eluded detection due to the dominant contribution of the conventional component at lower energies and the astrophysical flux at higher energies. Given the developed reconstruction methods, a unique opportunity arises to measure the prompt neutrino flux. A search for neutrino interactions in the detector volume, accompanied by lower energy muons coincident in direction and time of the neutrino, may be conducted. The neutrino in events such as these is likely to have been created in the same air shower from which the accompanying muons originate. A selection of such events would therefore reduce the contribution of astrophysical neutrinos and allow for the direct detection of the prompt neutrino flux. Essential to the success of such an analysis is a high selection efficiency as these event topologies are rare. This type of event selection was deemed infeasible with traditional reconstruction methods. However, initial tests indicate that sufficient signal events may be retained with the novel tools developed in this dissertation.

DNNTrack and DNNNeutrino Samples

The main focus in this dissertation is on the selection and reconstruction of cascade-like events. However, the developed frameworks allow for general applications — including track topologies. Initial studies detailed in Section 5.6, show the potential of the Event-Generator for the reconstruction of track events. Dedicated models and optimization of the event hypothesis parameterization may lead to a better reconstruction resolution than currently achievable via available track reconstruction methods in IceCube. Similarly, the construction of track-based event selections may be improved analogously to how the DNNCascade sample improved upon prior cascade samples. Construction of a future DNNTrack event selection, or even a general neutrino sample, DNNNeutrino, is a natural extension of the work presented in this dissertation.

Analysis Method Development

In addition to the construction of new event samples, improvements to existing analysis methods as well as the development of novel analysis techniques are closely connected to the work described in this dissertation. A more accurate modeling and parameterization of reconstruction uncertainties will enable advanced utilization in analyses and real-time applications (supplement Chapter A [Cascade Real-Time Alert Stream]). As discussed in Section 10.2 [Analysis Tool Simplifications], current analyses rely on approximations to model high-dimensional dependencies. Similarly to the modeling of high-dimensional PDFs with neural networks in the Event-Generator framework, deep learning may be utilized to model the underlying analysis PDFs more precisely. Finally, with neutrino signals arising in searches for large scale emission regions, such as the Galactic plane, novel analysis techniques must be developed to bridge the gap between the utilized template analyses and searches for point-like emission. Initial studies indicate that this may be accomplished via the construction of a mixture model of spatial templates of varying granularity. Such a mixture model could be used, for instance, to measure the emission regions along the Galactic plane, and thus to better understand the observed Galactic neutrino signal. In addition to the unbiased all-sky search for point-like emission, a new analysis technique to perform an unbiased search for large scale structures is required. This is arguably challenging to define and parameterize with traditional methods, however, a hybrid approach, leaning on traditional likelihood methods for a sound statistical interpretation and deep learning for the complex modeling, may lead to the desired results. Such an approach would have analogies to the Event-Generator methodology.

Bibliography

[1] IceCube collaboration, R. Abbasi et al., *Observation of high-energy neutrinos from the Galactic plane*, *Science* (2023) .

[2] M. Hünnefeld, *Deep learning for IceCube*, Discovery in Physics, pp. 245–258 (2022), www.scopus.com.

[3] IceCube collaboration, M. Huennefeld, *Deep Learning in Physics exemplified by the Reconstruction of Muon-Neutrino Events in IceCube*, *PoS* **ICRC2017** (2018) 1057.

[4] IceCube collaboration, M. Huennefeld, *Reconstruction Techniques in IceCube using Convolutional and Generative Neural Networks*, *EPJ Web Conf.* **207** (2019) 05005.

[5] IceCube collaboration, M. Huennefeld, *Combining Maximum-Likelihood with Deep Learning for Event Reconstruction in IceCube*, *PoS* **ICRC2021** (2021) 1065 [`2107.12110`].

[6] IceCube collaboration, R. Abbasi et al., *A convolutional neural network based cascade reconstruction for the IceCube Neutrino Observatory*, *JINST* **16** (2021) P07041 [`2101.11589`].

[7] IceCube collaboration, S. Sclafani et al., *A Search for Neutrino Sources with Cascade Events in IceCube*, *PoS* **ICRC2021** (2021) 1150 [`2107.09103`].

[8] K. Morik and W. Rhode, eds., *Volume 2 Machine Learning under Resource Constraints - Discovery in Physics*, De Gruyter, Berlin, Boston (2023), doi:10.1515/9783110785968.

[9] W. Rhode, *On probabilistic rationalism*, Tech. Rep. 1, TU Dortmund, Experimentelle Physik 5b (12, 2020).

[10] K.R. Popper, *The Logic of Scientific Discovery*, Hutchinson, London (1934).

[11] IceCube collaboration, M. Aartsen et al., *Energy Reconstruction Methods in the IceCube Neutrino Telescope*, *JINST* **9** (2014) P03009 [`1311.4767`].

[12] AMANDA collaboration, J. Ahrens et al., *Muon track reconstruction and data selection techniques in AMANDA*, *Nucl. Instrum. Meth. A* **524** (2004) 169 [`astro-ph/0407044`].

[13] Y. LeCun et al., *Handwritten Digit Recognition with a Back-Propagation Network*, in *Advances in Neural Information Processing Systems 2*, pp. 396–404, Morgan-Kaufmann (1990), http://papers.nips.cc/paper/293-handwritten-digit-recognition-with-a-back-propagation-network.pdf.

[14] IceCube collaboration, N. Choma et al., *Graph Neural Networks for IceCube Signal Classification*, `1809.06166`.

[15] IceCube collaboration, M. Kronmueller and T. Glauch, *Application of Deep Neural Networks to Event Type Classification in IceCube*, *PoS* **ICRC2019** (2020) 937 [`1908.08763`].

[16] IceCube collaboration, R. Abbasi et al., *Reconstruction of Neutrino Events in IceCube using Graph Neural Networks*, *PoS* **ICRC2021** (2021) 1044 [`2107.12187`].

[17] IceCube collaboration, R. Abbasi et al., *Reconstructing Neutrino Energy using CNNs for GeV Scale IceCube Events*, *PoS* **ICRC2021** (2021) 1053 [`2107.11446`].

[18] IceCube collaboration, S. Yu et al., *Direction Reconstruction using a CNN for GeV-Scale Neutrinos in IceCube*, *PoS* **ICRC2021** (2021) 1054 [`2107.02122`].

[19] IceCube collaboration, R. Abbasi et al., *Graph Neural Networks for low-energy event classification & reconstruction in IceCube*, *JINST* **17** (2022) P11003 [`2209.03042`].

[20] A. Krizhevsky, I. Sutskever and G.E. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*, *Commun. ACM* **60** (2017) 84.

[21] L. Deng, *The mnist database of handwritten digit images for machine learning research*, *IEEE Signal Processing Magazine* **29** (2012) 141.

[22] IceCube collaboration, R. Abbasi et al., *Calibration and Characterization of the IceCube Photomultiplier Tube*, *Nucl. Instrum. Meth. A* **618** (2010) 139 [`1002.2442`].

[23] IceCube collaboration, R. Abbasi et al., *The Design and Performance of IceCube DeepCore*, *Astropart. Phys.* **35** (2012) 615 [`1109.6096`].

[24] IceCube collaboration, M. Aartsen et al., *South Pole glacial climate reconstruction from multi-borehole laser particulate stratigraphy*, *J. Glaciol.* **59** (2013) 1117.

[25] IceCube collaboration, M. Ackermann et al., *Optical properties of deep glacial ice at the South Pole*, *J. Geophys. Res.* **111** (2006) D13203.

[26] IceCube collaboration, M. Aartsen et al., *The IceCube Neutrino Observatory: Instrumentation and Online Systems*, *JINST* **12** (2017) P03012 [`1612.05093`].

[27] IceCube collaboration, M. Aartsen et al., *The IceCube Realtime Alert System*, *Astropart. Phys.* **92** (2017) 30 [`1612.06028`].

[28] IceCube collaboration, M. Rongen, *Measuring the optical properties of IceCube drill holes*, *EPJ Web Conf.* **116** (2016) 06011.

[29] ICECUBE collaboration, D. Chirkin, *Evidence of optical anisotropy of the South Pole ice*, in *33rd International Cosmic Ray Conference*, p. 0580, 2013.

[30] IceCube collaboration, R. Abbasi et al., *In-situ estimation of ice crystal properties at the south pole using led calibration data from the icecube neutrino observatory*, *The Cryosphere Discussions* **2022** (2022) 1.

[31] IceCube collaboration, M. Aartsen et al., *Measurement of South Pole ice transparency with the IceCube LED calibration system*, *Nucl. Instrum. Meth. A* **711** (2013) 73 [`1301.5361`].

[32] IceCube collaboration, R. Abbasi et al., *D-Egg: a dual PMT optical module for IceCube*, *JINST* **18** (2023) P04014 [`2212.14526`].

[33] IceCube collaboration, A. Ishihara, *The IceCube Upgrade - Design and Science Goals*, *PoS* **ICRC2019** (2021) 1031 [`1908.09441`].

[34] IceCube collaboration, W.Y. Ma, *Physics Potential of the IceCube Upgrade*, *J. Phys. Conf. Ser.* **1468** (2020) 012169.

[35] IceCube-Gen2 collaboration, M.G. Aartsen et al., *IceCube-Gen2: the window to the extreme Universe*, *J. Phys. G* **48** (2021) 060501 [`2008.04323`].

[36] IceCube-Gen2 collaboration, B. Clark, *The IceCube-Gen2 Neutrino Observatory*, *JINST* **16** (2021) C10007 [`2108.05292`].

[37] IceCube collaboration, L. Classen, C. Dorn, A. Kappes, T. Karg, M. Kossatz, A. Kretzschmann et al., *A multi-PMT Optical Module for the IceCube Upgrade*, *PoS* **ICRC2019** (2020) 855 [`1908.10802`].

[38] IceCube-Gen2 collaboration, N. Shimizu et al., *Performance studies for a next-generation optical sensor for IceCube-Gen2*, *PoS* **ICRC2021** (2021) 1041 [`2108.05548`].

[39] IceCube-Gen2 collaboration, V. Basu et al., *A next-generation optical sensor for IceCube-Gen2*, *PoS* **ICRC2021** (2021) 1062 [`2107.08837`].

[40] K. Hornik, M. Stinchcombe and H. White, *Multilayer feedforward networks are universal approximators*, *Neural Networks* **2** (1989) 359.

[41] P. Werbos, *Beyond regression : new tools for prediction and analysis in the behavioral sciences*, Ph.D. thesis, Harvard University, 1974.

[42] D. Parker, M.I. of Technology and S.S. of Management, *Learning Logic: Casting the Cortex of the Human Brain in Silicon*, Technical report: Center for Computational Research in Economics and Management Science, Massachusetts Institute of Technology, Center for Computational Research in Economics and Management Science (1985).

[43] Y. Lecun, *Une procédure d'apprentissage pour réseau à seuil asymétrique*, *Proceedings of Cognitiva 85, Paris* (1985) 599.

[44] D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Learning Representations by Back Propagating Errors*, *Nature* **323** (1986) 533.

[45] A. r. Mohamed, G.E. Dahl and G. Hinton, *Acoustic modeling using deep belief networks*, *IEEE Transactions on Audio, Speech, and Language Processing* **20** (2012) 14.

[46] G.E. Dahl, D. Yu, L. Deng and A. Acero, *Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition*, *IEEE Transactions on Audio, Speech, and Language Processing* **20** (2012) 30.

[47] A. Krizhevsky, I. Sutskever and G.E. Hinton, *ImageNet Classification with Deep Convolutional Neural Networks*, in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, pp. 1097–1105 (2012), https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[48] R. Raina, A. Madhavan and A.Y. Ng, *Large-scale deep unsupervised learning using graphics processors*, in *Proceedings of the 26th International Conference on Machine Learning (ICML-09)*, A.P. Danyluk, L. Bottou and M.L. Littman, eds., p. 110, 2009, 10.1145/1553374.1553486.

[49] N. Srivastava et al., *Dropout: A simple way to prevent neural networks from overfitting*, vol. 15, pp. 1929–1958, 2014, http://jmlr.org/papers/v15/srivastava14a.html.

[50] X. Glorot, A. Bordes and Y. Bengio, *Deep sparse rectifier neural networks*, in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, vol. 15 of *Proceedings of Machine Learning Research*, pp. 315–323, PMLR, 11–13 Apr, 2011, http://proceedings.mlr.press/v15/glorot11a.html.

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

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

[53] D.P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, *CoRR* **abs/1412.6980** (2014) [1412.6980].

[54] Y.A. LeCun, L. Bottou, G.B. Orr and K.-R. Müller, *Efficient BackProp*, in *Lecture Notes in Computer Science*, pp. 9–48, Springer Berlin Heidelberg (2012),10.1007/978-3-642-35289-8_3.

[55] R.J. Rossi, *Mathematical statistics: An Introduction to Likelihood Based Inference*, John Wiley & Sons, Nashville, TN (Dec., 2018).

[56] D. Pandel, *Bestimmung von Wasser- und Detektorparametern und Rekonstruktion von Myonen bis 100 TeV mit dem Baikal-Neutrinoteleskop NT-72*, Ph.D. thesis, Humboldt-Universität zu Berlin, 1996.

[57] IceCube collaboration, D. Chirkin, *Event reconstruction in IceCube based on direct event re-simulation*, in *33rd International Cosmic Ray Conference*, p. 0581, 2013 [1309.7010].

[58] IceCube collaboration, S. Nowicki, *Improving reconstruction of GeV-scale neutrinos in icecube-deepcore by direct event simulation*, *Physics in Canada* **74** (2018) 125.

[59] S. Nowicki, *Improving reconstruction of GeV-scale neutrinos in IceCube-DeepCore by direct event simulation*, in *XXVIII International Conference on Neutrino Physics and Astrophysics*, p. 209, June, 2018,10.5281/zenodo.1300644.

[60] C. Haack, L. Lu and T. Yuan, *Improving the directional reconstruction of PeV hadronic cascades in IceCube*, *EPJ Web Conf.* **207** (2019) 05003.

[61] AMANDA collaboration, J. Ahrens et al., *Muon track reconstruction and data selection techniques in AMANDA*, *Nucl. Instrum. Meth. A* **524** (2004) 169 [astro-ph/0407044].

[62] K. Schatto, *Stacked searches for high-energy neutrinos from blazars with IceCube*, Ph.D. thesis, Mainz U., 2014-06-02.

[63] T. DeYoung, *IceTray: A software framework for IceCube*, in *Computing in high energy physics and nuclear physics. Proceedings, Conference, CHEP'04, Interlaken, Switzerland, September 27-October 1, 2004*, pp. 463–466, 2005, http://doc.cern.ch/yellowrep/2005/2005-002/p463.pdf.

[64] M. Huennefeld, *icecube/dnn_reco: Version 1.0.1*, June, 2023. 10.5281/zenodo.8003189.

[65] S. Schonert, T.K. Gaisser, E. Resconi and O. Schulz, *Vetoing atmospheric neutrinos in a high energy neutrino telescope*, *Phys. Rev. D* **79** (2009) 043009 [0812.4308].

[66] T.K. Gaisser, K. Jero, A. Karle and J. van Santen, *Generalized self-veto probability for atmospheric neutrinos*, *Phys. Rev. D* **90** (2014) 023009 [`1405.0525`].

[67] IceCube collaboration, M. Aartsen et al., *Search for Sources of Astrophysical Neutrinos Using Seven Years of IceCube Cascade Events*, *Astrophys. J.* **886** (2019) 12 [`1907.06714`].

[68] M. Börner et al., *Measurement/Simulation Mismatches and Multivariate Data Discretization in the Machine Learning Era*, in *Astronomical Data Analysis Software and Systems XXVII*, vol. 522, p. 431, 2019.

[69] D. Martschei et al., *Advanced event reweighting using multivariate analysis*, *J. Phys. Conf. Ser.* **368** (2012) 012028.

[70] L. Breiman, *Random Forests*, *Machine Learning* **45** (2001) 5.

[71] F. Pedregosa et al., *Scikit-learn: Machine learning in Python*, vol. 12, pp. 2825–2830, 2011, https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf.

[72] C. Szegedy et al., *Intriguing properties of neural networks*, *CoRR* **abs/1312.6199** (2014) [`1312.6199`].

[73] A. Nguyen, J. Yosinski and J. Clune, *Deep neural networks are easily fooled: High confidence predictions for unrecognizable images*, in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, June, 2015, `10.1109/cvpr.2015.7298640`.

[74] IceCube collaboration, M. Aartsen et al., *Efficient propagation of systematic uncertainties from calibration to analysis with the SnowStorm method in IceCube*, *JCAP* **10** (2019) 048 [`1909.01530`].

[75] T. Kato, S. Omachi and H. Aso, *Asymmetric Gaussian and Its Application to Pattern Recognition*, in *Lecture Notes in Computer Science*, pp. 405–413, Springer Berlin Heidelberg (2002), `10.1007/3-540-70659-3_42`.

[76] R.J. Barlow, *Extended maximum likelihood*, *Nucl. Instrum. Meth. A* **297** (1990) 496.

[77] D. Salinas et al., *DeepAR: Probabilistic forecasting with autoregressive recurrent networks*, *International Journal of Forecasting* **36** (2020) 1181.

[78] M. Abadi et al., *Tensorflow: Large-scale machine learning on heterogeneous distributed systems*, *CoRR* **abs/1603.04467** (2016) [`1603.04467`].

[79] M. Taboga, “Covariance matrix of the maximum likelihood estimator.” Lectures on probability theory and mathematical statistics. Kindle Direct Publishing. Online appendix. https://www.statlect.com/fundamentals-of-statistics/maximum-likelihood-covariance-matrix-estimation (2021).

[80] E.R. Berndt, B. Hall, R. Hall and J. Hausman, *Estimation and Inference in Nonlinear Structural Models*, in *Annals of Economic and Social Measurement, Volume 3, number 4*, pp. 653–665, National Bureau of Economic Research, Inc (1974), https://EconPapers.repec.org/RePEc:nbr:nberch:10206.

[81] D. Foreman-Mackey, D.W. Hogg, D. Lang and J. Goodman, *emcee: The MCMC Hammer*, *Publ. Astron. Soc. Pac.* **125** (2013) 306 [1202.3665].

[82] R.A. Fisher, *Dispersion on a sphere*, *Proc. R. Soc. Lond. A.* **217** (1953) 295.

[83] T. Yuan, *The 8-parameter Fisher-Bingham distribution on the sphere*, 1906.08247.

[84] J.T. Kent, *The Fisher-Bingham Distribution on the Sphere*, *Journal of the Royal Statistical Society. Series B (Methodological)* **44** (1982) 71.

[85] K.M. Górski and E. Hivon, “HEALPix: Hierarchical Equal Area isoLatitude Pixelization of a sphere.” Astrophysics Source Code Library, record ascl:1107.018, July, 2011.

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

[87] M. Huennefeld, *icecube/event-generator: Version 1.0.3*, June, 2023. 10.5281/zenodo.8003172.

[88] IceCube collaboration, M. Aartsen et al., *Time-Integrated Neutrino Source Searches with 10 Years of IceCube Data*, *Phys. Rev. Lett.* **124** (2020) 051103 [1910.08488].

[89] IceCube collaboration, R. Abbasi et al., *Evidence for neutrino emission from the nearby active galaxy NGC 1068*, *Science* **378** (2022) 538 [2211.09972].

[90] T.K. Gaisser and M. Honda, *Flux of atmospheric neutrinos*, *Ann. Rev. Nucl. Part. Sci.* **52** (2002) 153 [hep-ph/0203272].

[91] M. Aartsen et al., *Characteristics of the diffuse astrophysical electron and tau neutrino flux with six years of icecube high energy cascade data*, *Phys. Rev. Lett.* **125** (2020) 121104.

[92] IceCube Collaboration collaboration, M.G. Aartsen et al., *Atmospheric and astrophysical neutrinos above 1 TeV interacting in IceCube*, *Phys. Rev. D* **91** (2015) 022001.

[93] IceCube collaboration, M. Aartsen et al., *Evidence for High-Energy Extraterrestrial Neutrinos at the IceCube Detector*, *Science* **342** (2013) 1242856 [1311.5238].

[94] IceCube collaboration, R. Abbasi et al., *Searches for and Characterization of Astrophysical Neutrinos using Starting Track Events in IceCube*, *PoS* **ICRC2021** (2021) 1130 [2107.09811].

[95] IceCube collaboration, R. Abbasi et al., *Improved Characterization of the Astrophysical Muon–neutrino Flux with 9.5 Years of IceCube Data*, *Astrophys. J.* **928** (2022) 50 [2111.10299].

[96] J. van Santen, *Neutrino Interactions in IceCube above 1 TeV: Constraints on Atmospheric Charmed-Meson Production and Investigation of the Astrophysical Neutrino Flux with 2 Years of IceCube Data taken 2010–2012*, Ph.D. thesis, Wisconsin U., Madison, 11, 2014.

[97] J.N. Capdevielle, P. Gabriel, H.J. Gils, P. Grieder, D. Heck, J. Knapp et al., *The karlsruhe extensive air shower simulation code corsika*, Tech. Rep. (1992), 10.5445/IR/270033168.

[98] D. Heck, J. Knapp, J.N. Capdevielle, G. Schatz and T. Thouw, *Corsika: A monte carlo code to simulate extensive air showers*, Tech. Rep. (1998), 10.5445/IR/270043064.

[99] T. Chen and C. Guestrin, *XGBoost: A Scalable Tree Boosting System*, in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, KDD '16, (New York, NY, USA), p. 785, ACM, 2016, 10.1145/2939672.2939785.

[100] IceCube collaboration, M.G. Aartsen et al., *Measurement of the ν_μ energy spectrum with IceCube-79*, *Eur. Phys. J. C* **77** (2017) 692 [1705.07780].

[101] IceCube collaboration, C. Kopper, *Observation of Astrophysical Neutrinos in Six Years of IceCube Data*, *PoS* **ICRC2017** (2018) 981.

[102] J. Albrecht et al., *The Muon Puzzle in cosmic-ray induced air showers and its connection to the Large Hadron Collider*, *Astrophys. Space Sci.* **367** (2022) 27 [2105.06148].

[103] A. Fedynitch, R. Engel, T.K. Gaisser, F. Riehn and T. Stanev, *Calculation of conventional and prompt lepton fluxes at very high energy*, *EPJ Web Conf.* **99** (2015) 08001.

[104] T.K. Gaisser, *Spectrum of cosmic-ray nucleons, kaon production, and the atmospheric muon charge ratio*, *Astropart. Phys.* **35** (2012) 801 [1111.6675].

[105] T.K. Gaisser, T. Stanev and S. Tilav, *Cosmic Ray Energy Spectrum from Measurements of Air Showers*, *Front. Phys. (Beijing)* **8** (2013) 748 [1303.3565].

[106] A. Fedynitch, F. Riehn, R. Engel, T.K. Gaisser and T. Stanev, *Hadronic interaction model sibyll* 2.3c *and inclusive lepton fluxes*, *Phys. Rev. D* **100** (2019) 103018.

[107] C.A. Argüelles, S. Palomares-Ruiz, A. Schneider, L. Wille and T. Yuan, *Unified atmospheric neutrino passing fractions for large-scale neutrino telescopes*, *JCAP* **07** (2018) 047 [`1805.11003`].

[108] IceCube collaboration, M.G. Aartsen et al., *Differential limit on the extremely-high-energy cosmic neutrino flux in the presence of astrophysical background from nine years of IceCube data*, *Phys. Rev. D* **98** (2018) 062003 [`1807.01820`].

[109] IceCube collaboration, L. Lu, *Multi-flavour PeV neutrino search with IceCube*, *PoS* **ICRC2017** (2018) 1002.

[110] S.L. Glashow, *Resonant Scattering of Antineutrinos*, *Phys. Rev.* **118** (1960) 316.

[111] IceCube collaboration, M.G. Aartsen et al., *Detection of a particle shower at the Glashow resonance with IceCube*, *Nature* **591** (2021) 220 [`2110.15051`].

[112] C. Haack, L. Lu and T. Yuan, *Improving the directional reconstruction of PeV hadronic cascades in IceCube*, *EPJ Web Conf.* **207** (2019) 05003.

[113] IceCube collaboration, M. Aartsen et al., *Neutrino emission from the direction of the blazar TXS 0506+056 prior to the IceCube-170922A alert*, *Science* **361** (2018) 147 [`1807.08794`].

[114] IceCube, Fermi-LAT, MAGIC, AGILE, ASAS-SN, HAWC, H.E.S.S., INTEGRAL, Kanata, Kiso, Kapteyn, Liverpool Telescope, Subaru, Swift NuSTAR, VERITAS, VLA/17B-403 collaboration, M. Aartsen et al., *Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A*, *Science* **361** (2018) eaat1378 [`1807.08816`].

[115] A. Mellinger, *A color all-sky panorama image of the milky way*, *Publ. Astron. Soc. Pac.* **121** (2009) 1180.

[116] https://svs.gsfc.nasa.gov/14090. NASA Goddard Space Flight Center.

[117] Fermi-LAT collaboration, M. Ackermann et al., *Fermi-LAT Observations of the Diffuse Gamma-Ray Emission: Implications for Cosmic Rays and the Interstellar Medium*, *Astrophys. J.* **750** (2012) 3 [`1202.4039`].

[118] F.W. Stecker, *Diffuse Fluxes of Cosmic High-Energy Neutrinos*, *Astrophys. J.* **228** (1979) 919.

[119] V.S. Berezinsky, T.K. Gaisser, F. Halzen and T. Stanev, *Diffuse radiation from cosmic ray interactions in the galaxy*, *Astropart. Phys.* **1** (1993) 281.

[120] G. Ingelman and M. Thunman, *Particle production in the interstellar medium*, `hep-ph/9604286`.

[121] C. Evoli, D. Grasso and L. Maccione, *Diffuse Neutrino and Gamma-ray Emissions of the Galaxy above the TeV*, *JCAP* **06** (2007) 003 [`astro-ph/0701856`].

[122] Fermi-LAT collaboration, M. Ackermann et al., *Detection of the Characteristic Pion-Decay Signature in Supernova Remnants*, *Science* **339** (2013) 807 [`1302.3307`].

[123] D. Guetta and E. Amato, *Neutrino flux predictions for galactic plerions*, *Astropart. Phys.* **19** (2003) 403 [`astro-ph/0209537`].

[124] W. Bednarek, *Neutrinos from the pulsar wind nebulae*, *Astron. Astrophys.* **407** (2003) 1 [`astro-ph/0305430`].

[125] M.C. Gonzalez-Garcia, F. Halzen and S. Mohapatra, *Identifying Galactic PeVatrons with Neutrinos*, *Astropart. Phys.* **31** (2009) 437 [`0902.1176`].

[126] F. Halzen, A. Kheirandish and V. Niro, *Prospects for Detecting Galactic Sources of Cosmic Neutrinos with IceCube: An Update*, *Astropart. Phys.* **86** (2017) 46 [`1609.03072`].

[127] IceCube collaboration, M.G. Aartsen et al., *Constraints on Galactic Neutrino Emission with Seven Years of IceCube Data*, *Astrophys. J.* **849** (2017) 67 [`1707.03416`].

[128] ANTARES, IceCube collaboration, A. Albert et al., *Joint Constraints on Galactic Diffuse Neutrino Emission from the ANTARES and IceCube Neutrino Telescopes*, *Astrophys. J. Lett.* **868** (2018) L20 [`1808.03531`].

[129] ANTARES collaboration, A. Albert et al., *New constraints on all flavor Galactic diffuse neutrino emission with the ANTARES telescope*, *Phys. Rev. D* **96** (2017) 062001 [`1705.00497`].

[130] ANTARES collaboration, A. Albert et al., *Hint for a TeV neutrino emission from the Galactic Ridge with ANTARES*, `2212.11876`.

[131] M. Huennefeld, *icecube/event-generator: Version 1.0.2*, Dec., 2022. 10.5281/zenodo.7412035.

[132] IceCube collaboration, R. Abbasi et al., *The IceCube high-energy starting event sample: Description and flux characterization with 7.5 years of data*, *Phys. Rev. D* **104** (2021) 022002 [`2011.03545`].

[133] M. Ahlers, Y. Bai, V. Barger and R. Lu, *Galactic neutrinos in the TeV to PeV range*, *Phys. Rev. D* **93** (2016) 013009 [`1505.03156`].

[134] J. Braun, J. Dumm, F. De Palma, C. Finley, A. Karle and T. Montaruli, *Methods for point source analysis in high energy neutrino telescopes*, *Astropart. Phys.* **29** (2008) 299 [`0801.1604`].

[135] G.J. Feldman and R.D. Cousins, *A Unified approach to the classical statistical analysis of small signals*, *Phys. Rev. D* **57** (1998) 3873 [`physics/9711021`].

[136] FERMI-LAT collaboration, S. Abdollahi et al., *Fermi Large Area Telescope Fourth Source Catalog*, *Astrophys. J. Suppl.* **247** (2020) 33 [`1902.10045`].

[137] ICECUBE collaboration, R. Abbasi et al., *Time-Integrated Searches for Point-like Sources of Neutrinos with the 40-String IceCube Detector*, *Astrophys. J.* **732** (2011) 18 [`1012.2137`].

[138] http://tevcat2.uchicago.edu/. TeVCat 2.0.

[139] http://gamma sky.net/. Gamma-Cat.

[140] M. Su, T.R. Slatyer and D.P. Finkbeiner, *Giant Gamma-ray Bubbles from Fermi-LAT: AGN Activity or Bipolar Galactic Wind?*, *Astrophys. J.* **724** (2010) 1044 [`1005.5480`].

[141] A.W. Strong and I.V. Moskalenko, *Propagation of cosmic-ray nucleons in the galaxy*, *Astrophys. J.* **509** (1998) 212 [`astro-ph/9807150`].

[142] D. Gaggero, D. Grasso, A. Marinelli, A. Urbano and M. Valli, *The gamma-ray and neutrino sky: A consistent picture of Fermi-LAT, Milagro, and IceCube results*, *Astrophys. J. Lett.* **815** (2015) L25 [`1504.00227`].

[143] C. Evoli, D. Gaggero, D. Grasso and L. Maccione, *Cosmic-Ray Nuclei, Antiprotons and Gamma-rays in the Galaxy: a New Diffusion Model*, *JCAP* **10** (2008) 018 [`0807.4730`].

[144] C. Evoli, D. Gaggero, D. Grasso and L. Maccione, *A common solution to the cosmic ray anisotropy and gradient problems*, *Phys. Rev. Lett.* **108** (2012) 211102 [`1203.0570`].

[145] D.Gaggero, D. Grasso, A. Marinelli, A. Urbano and M. Valli, *Kragamma cosmic-ray diffusion models*, Sept., 2022. 10.5281/zenodo.7070823.

[146] ANTARES collaboration, M. Ageron et al., *ANTARES: the first undersea neutrino telescope*, *Nucl. Instrum. Meth. A* **656** (2011) 11 [`1104.1607`].

[147] ANTARES collaboration, G. Illuminati, *Searches for point-like sources of cosmic neutrinos with 13 years of ANTARES data*, *PoS* **ICRC2021** (2021) 1161.

[148] J. Aublin, *ANTARES search for point sources of neutrinos with 11 yr of data: a likelihood stacking analysis*, *PoS* **ICRC2019** (2019) 840.

[149] Tibet ASgamma collaboration, M. Amenomori et al., *First Detection of sub-PeV Diffuse Gamma Rays from the Galactic Disk: Evidence for Ubiquitous Galactic Cosmic Rays beyond PeV Energies*, *Phys. Rev. Lett.* **126** (2021) 141101 [`2104.05181`].

[150] P. De la Torre Luque, D. Gaggero, D. Grasso and A. Marinelli, *Prospects for detection of a galactic diffuse neutrino flux*, *Front. Astron. Space Sci.* **9** (2022) 1041838 [`2209.10011`].

[151] LHAASO collaboration, S. Zhao, R. Zhang, Y. Zhang and Q. Yuan, *Measurement of the diffuse gamma-ray emission from Galactic plane with LHAASO-KM2A*, *PoS* **ICRC2021** (2021) 859.

[152] K. Murase, M. Ahlers and B.C. Lacki, *Testing the Hadronuclear Origin of PeV Neutrinos Observed with IceCube*, *Phys. Rev. D* **88** (2013) 121301 [`1306.3417`].

[153] M. Ahlers and K. Murase, *Probing the Galactic Origin of the IceCube Excess with Gamma-Rays*, *Phys. Rev. D* **90** (2014) 023010 [`1309.4077`].

[154] M. Ahlers and F. Halzen, *Opening a New Window onto the Universe with IceCube*, *Prog. Part. Nucl. Phys.* **102** (2018) 73 [`1805.11112`].

[155] K. Fang and K. Murase, *Multimessenger Implications of Sub-PeV Diffuse Galactic Gamma-Ray Emission*, *Astrophys. J.* **919** (2021) 93 [`2104.09491`].

[156] A.U. Abeysekara et al., *HAWC observations of the acceleration of very-high-energy cosmic rays in the Cygnus Cocoon*, *Nature Astron.* **5** (2021) 465 [`2103.06820`].

[157] M. Ackermann et al., *A Cocoon of Freshly Accelerated Cosmic Rays Detected by Fermi in the Cygnus Superbubble*, *Science* **334** (2011) 1103.

[158] ARGO-YBJ collaboration, B. Bartoli et al., *Study of the Diffuse Gamma-ray Emission From the Galactic Plane With ARGO-YBJ*, *Astrophys. J.* **806** (2015) 20 [`1507.06758`].

[159] KM3NeT collaboration, S. Aiello et al., *Sensitivity of the KM3NeT/ARCA neutrino telescope to point-like neutrino sources*, *Astropart. Phys.* **111** (2019) 100 [`1810.08499`].

[160] Baikal-GVD collaboration, V.A. Allakhverdyan et al., *Observations of track-like neutrino events with Baikal-GVD*, *PoS* **ICRC2021** (2021) 1177 [`2108.04704`].

[161] P-ONE collaboration, M. Agostini et al., *The Pacific Ocean Neutrino Experiment*, *Nature Astron.* **4** (2020) 913 [`2005.09493`].

[162] E. Andres et al., *Observation of high-energy neutrinos using Cherenkov detectors embedded deep in Antarctic ice*, *Nature* **410** (2001) 441.

[163] GCN/AMON Notice, IceCubeCascade-221229a, `https://gcn.gsfc.nasa.gov/notices_amon_icecube_cascade/137489_30229466.amon`, IceCube (2022).

[164] GCN CIRCULAR 33128, `https://gcn.gsfc.nasa.gov/gcn3/33128.gcn3`, IceCube (2022).

Glossary

AG Asymmetric Gaussian. 50, 59, 60, 81

AI Artificial Intelligence. 14

ANTARES Astronomy with a Neutrino Telescope and Abyss environmental RESearch project. 144

BDT Boosted Decision Tree. 94, 98–101, 116

CDF Cumulative Probability Function. 82

CNN Convolutional Neural Network. v, 6–8, 16, 21, 23–25, 27, 30–32, 34–38, 40–44, 56, 85, 86, 88, 89, 93, 94, 98, 99, 104, 117, 135, 142, 143, 150, 201–203

CORSIKA COsmic Ray SImulations for KAscade — an Air Shower Simulation Program. 95, 97, 112, 116

CR Cosmic Ray. 139, 159, 160

DOM Digitial Optical Module. 10–13, 20, 21, 23–30, 46, 47, 49, 51, 52, 54–70, 73, 78, 81, 83, 116, 219

EHE Extremely High-Energy. 121

ESTES Enhanced Starting Track Event Selection. 130, 166

FB$_5$ Fisher-Bingham distribution. 77, 78, 80

GCN The Gamma-ray Coordinates Network. 200, 201

GeV Giga–electron volt. 13, 66, 67, 69, 75, 98, 101, 106, 116, 117, 126, 137–139, 159, 162, 164, 165

GP Galactic Plane. 125, 129, 139, 142, 154, 157, 159, 160, 182

HEALPix Hierarchical Equal Area isoLatitude Pixelisation. 77, 201

ISM Interstellar Medium. 139

MC Monte Carlo. 5, 18, 26, 35, 39, 40, 43, 48, 54, 56, 58, 61–64, 68, 80, 81, 89, 95, 96, 112–118, 120, 135, 145, 160–162, 164, 207–218

MCMC Markov-Chain-Monte-Carlo. 76–78, 80

MESC Medium Energy Starting Cascade. 103, 107, 118, 119, 121, 142–144, 150, 151, 165, 166

ML Machine Learning. 14

MLE Maximum-Likelihood Estimation. 4, 5, 17, 18, 43, 46, 47, 51, 77, 78, 86

MLP Multilayer Perceptron. 15, 16

MNIST Modified National Institute of Standards and Technology database. 6, 7

MPE Multi-Photo-Electron. 20

MuonGun IceCube toolkit for efficiently simulating the flux of atmospheric muons deep under the ice. 95, 97, 112, 116

NN Neural Network. 6–9, 21–25, 27, 28, 39, 54, 55, 64, 67, 85, 93, 98–101, 104, 105, 204–206, 215–218

OPG Outer Product of Gradients. 76

PDF Probability Density Function. 18, 46, 47, 49–51, 54, 57–68, 72–74, 78, 81–83, 86, 130–134, 140, 143, 160, 174, 183, 201, 202

PEPE PeV energy partially contained event sample. 121

PeV Peta–electron volt. 66, 67, 72, 104, 118, 121, 122, 140, 159, 162

PMT Photomultiplier Tube. 10, 13, 26, 52, 81, 83, 85

PWN Pulsar Wind Nebulae. 138, 147, 232, 237

ReLU Rectified linear units. 15

RF Random Forest. 37–40, 43

SNR Supernova Remnants. 138, 232, 237

TeV Tera–electron volt. 30, 31, 42, 55, 57, 70, 72, 75, 92, 94, 104, 106, 110, 115, 120, 126, 129, 131, 138–140, 147, 151, 158, 160, 162–165, 167, 168

UNID Unidentified Galactic sources. 138, 232, 237

vMF Von Mises-Fisher distribution. 77–79

Part V

Supplementary Material

A Cascade Real-Time Alert Stream

Declaration

The cascade real-time alert stream was developed in cooperation with Dr. Timothée Grégoire (Penn State U., PA, USA). The author of this dissertation contributed the event reconstruction and uncertainty estimation procedure, as well as the software tools to automate the reconstruction. Dr. Grégoire developed the event selection and setup the alert pipeline.

Despite having detected a flux of astrophysical neutrinos in 2013 [93], the sources for the majority of the astrophysical neutrino flux remain elusive. First evidence for a transient, point-like neutrino source, a blazar known as TXS 0506+056, was initially found via a multi-messenger campaign [113, 114]. An IceCube realtime alert for high-energy track events, sent out on 22 September 2017, allowed the community to perform followup observations with photons of different wavelengths [114]. These observations detected a gamma-ray flare from the blazar TXS 0506+056, consistent in direction and time with the neutrino alert, which prompted an additional archival search in IceCube data [113].

Motivated by this success, IceCube's realtime efforts are extended to also encompass cascade events. Although cascades provide decreased angular resolution in comparison to tracks, their superior signal purity and energy resolution make them an interesting candidate for followup observations. The DNN-reco framework (Chapter 4) is utilized to perform a rapid event reconstruction in real-time. The framework is extended to describe asymmetric uncertainty contours, discussed in Section A.1. Results are then distributed via the Gamma-ray Coordinates Network (GCN). A description of the cascade real-time alert stream is provided on the GCN website[1] and the software for the real-time selection is available on GitHub[2]. Contours for an example cascade alert from 2022 are illustrated in Figure A.1.

A.1 Asymmetric Uncertainty Contours

Applications in IceCube often utilize "circularized" angular uncertainties, as described in Eq. 4.6. The circularized uncertainty estimates are symmetrical in right ascension (detector azimuth) and declination (detector zenith) and thus simplify analysis tools (Section 10.2 [Analysis Tool Simplifications]). While this simplification is justified for the analysis of large event samples, it is not an accurate description

---

[1] `https://gcn.gsfc.nasa.gov/doc/High_Energy_Neutrino_Cascade_Alerts.pdf`

[2] `https://github.com/icecube/realtime`

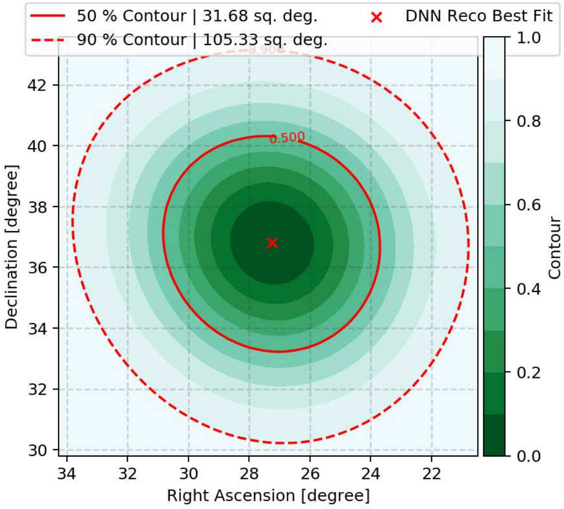


Figure A.1: Reconstructed contours of an example alert detected on December 29th, 2022 are shown. A deposited energy of $85.2^{+9.8}_{-8.8}$ TeV is reconstructed. The GCN notice for this alert is available in Ref. [163]. An internal follow-up analysis searching for additional track-like neutrino events in a two day time window did not find a significant neutrino excess with a p-value of 0.24 [164].

of the per-event uncertainty relevant to followup observations of individual realtime alerts, which are often asymmetric.

A simple expansion of the circularized uncertainties is to allow for an asymmetry between detector azimuth and zenith, leading to elliptical contours. Due to IceCube's geometry, uncertainties in azimuth are often larger than for the zenith angle, but this may depend on the topology and location of the neutrino interaction in the ice. Elliptical contours are an adequate description for a large fraction of IceCube events.

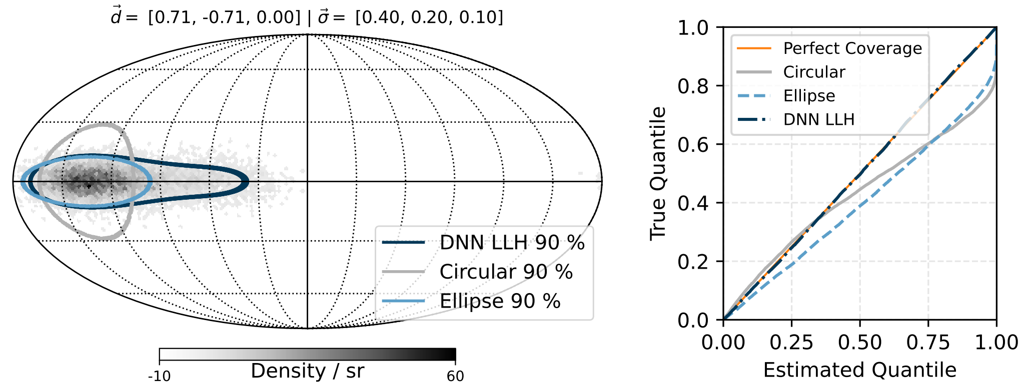
As described in Chapter 4, the default DNN-reco CNN architecture estimates the event direction by individually reconstructing the three components of the direction vector. A Gaussian uncertainty estimate is provided for each of these estimates. Internally, the CNN model architecture ensures the normalization of the direction vector, which adds a correlation between the estimated components that is not covered by the reported Gaussian description. In order to achieve a more accurate description of the uncertainty contours, the following heuristic procedure is performed:

1. The sky is divided into equal area bins using the Hierarchical Equal Area isoLatitude Pixelisation (HEALPix) [85]. For each of the resulting pixels, the product over the three Gaussian likelihood terms for each of the direction vector components is computed.

2. The resulting likelihood values are normalized over all pixels to correspond to a PDF over the sky.

3. Direction vectors are then sampled from this PDF and the standard deviations for each individual vector component are evaluated. These should correspond to the standard deviations estimated by the model.
4. If the standard deviations of the sampled components are not compatible with the ones provided by the model, the uncertainties are individually rescaled for each of the components. Steps 1–3 are then repeated using the updated values.
5. Utilize the resulting PDF to obtain contours over the sky.

This procedure allows for the description of asymmetric uncertainty contours as estimated by the utilized CNN. The improved ability to model asymmetric contours via this approach, labeled as DNN LLH, is demonstrated in Figure A.2. Future models may directly estimate the posterior on the sphere via a suitable parameterization, directly encompassing correlations of the direction vector components.



(a) Contours.

(b) Coverage.

Figure A.2: The DNN-RECO contour construction is demonstrated for points sampled from an asymmetric distribution defined by the reconstructed direction vector $\vec{d}$ and uncertainty estimates $\vec{\sigma}$. Sampled points and corresponding contours for each of the three contour constructions are shown on the left. The standard approach via circularized uncertainty estimates (Circular) leads to under-coverage in azimuth and over-coverage in zenith angles. Elliptical contours (Ellipse) can better capture this dependence, but are unable to describe further asymmetries. On the right, the coverage for each of the approaches is computed, showing that the heuristic approach (DNN LLH) provides an accurate description for this example.

B Neural Network Models

B.1 DNN-reco Models

The architecture and training procedure of the neural networks utilized in this dissertation are defined in central configuration files. These files are provided in the DNN-reco software framework [64] in the configs/ directory on GitHub[1] with file paths matching the model names.

B.1.1 Models for Selection of Stopping Muons

Models utilized for the selection of stopping muons include:

- **`Stopping muon classifier:`** A small CNN trained to select stopping muon events. Configuration file: stopping_muons_hoinka_no_filter.yaml.
- **`Stopping muon direction:`** A small CNN trained to reconstruct the incoming direction of stopping muon events. Configuration file: stopping_muons_hoinka_no_filter___direction.yaml.
- **`Stopping muon` r `and` z:** A small CNN trained to reconstruct the depth (z) and radius r in the x-y-plane of the stopping point of the muon. Configuration file: stopping_muons_hoinka_no_filter___stopping_rz.yaml.

B.1.2 Models for DNNCascade Selection

In addition to the information provided in the in the DNN-reco software framework [64], summary information is provided here on the models used within the DNNCascade event selection.

---

[1] `https://github.com/icecube/dnn_reco/tree/v1.0.1/configs/`

Table B.1: The training objective and restrictions on the data used during training of the models are shown for the NNs utilized in selection step two. The shower length, the longest length of any of the secondary particles, is defined as L.
*The shared model name prefix EVENT_SELECTION_ is omitted for brevity.
**The DNN_CSCD_L3A_STARTING_EVENTS_03 model was only used for training of models and is not used in event selection pipeline.

Model Name*	**Objective**	**Data restrictions**
CASCADE_DIR_01	Regression of cascade direction	Cascade L3, starting within 60 m, charged-current ν_e, $L < 30\,\mathrm{m}$
CASCADE_ENERGY_01	Regression of cascade energy	Cascade L2, starting within 60 m, charged-current ν_e, $L < 30\,\mathrm{m}$
CASCADE_POS_01	Regression of cascade position	Cascade L2, starting within 300 m, $L < 30\,\mathrm{m}$
CASCADE_TIME_01	Regression of cascade time	Cascade L2, starting within 300 m, $L < 30\,\mathrm{m}$
EGEN_SEED_DIR_01	Regression of starting event direction	μ-BDT $< 1 \times 10^{-2}$
EGEN_SEED_ENERGY_01	Regression of starting event energy	μ-BDT $< 1 \times 10^{-2}$
EGEN_SEED_POS_01	Regression of starting event position	μ-BDT $< 1 \times 10^{-2}$
EGEN_SEED_TIME_01	Regression of starting event time	μ-BDT $< 1 \times 10^{-2}$
DNN_CSCD_L3A_STARTING_EVENTS_03**	Classification of atmospheric muons versus starting neutrino events within 300 m	Cascade L2
DNN_CSCD_L3B_CUT2_STARTING_EVENTS_300M_FAST_MEDIUM_01	Classification of atmospheric muons versus starting neutrino events within 300 m	Cascade L2, for ν-simulation: starting within 300 m, DNN_CSCD_L3A_STARTING_EVENTS_03 ≥ 0.5
DNN_CSCD_L3C_CUT2_STARTING_EVENTS_300M_FAST_MEDIUM_01	Classification of atmospheric muons versus starting neutrino events within 300 m	selection step 1, for ν-simulation: starting events within 300 m
DNN_CSCD_L3C_TRACK_NUMU_CC_VS_NUE_CC_01	Classification of charged-current ν_μ versus charged-current ν_e	selection step 1, starting within 150 m, charged-current ν_μ and ν_e
DNN_CSCD_L3C_TRACK_NUMU_CC_VS_STARTING_01	Classification of charged-current ν_μ versus starting events within 150 m	selection step 1, starting within 150 m

Table B.2: The training objective and restrictions on the data used during training of the models is shown for the fast-precut NNs utilized in selection step 1. A classification score value of ≥ 0.95 is required for events to pass this initial selection step.

Model Name	**Objective**	**Data restrictions**
event__selection__dnn__cscd__L3a__starting__events__300m__red__summary__stats__ fast__02	Classification of atmospheric muons versus starting neutrino events within 300 m of the detector	All muon background; Only neutrinos starting within 300 m of the detector
event__selection__dnn__cscd__L3a__starting__events__150m__red__summary__stats__ fast__02	Classification of atmospheric muons versus starting neutrino events within 150 m of the detector	All muon background; Only neutrinos starting within 150 m of the detector
event__selection__dnn__cscd__L3a__starting__cascades__150m__red__summary__stats__ fast__02	Classification of atmospheric muons versus cascade† events within 150 m of the detector	All muon background; Only cascades† within 150 m of the detector

Table B.3: Details of the cascade-based models utilized in selection step three are shown. These NNs utilize inputs computed based on an assumed cascade vertex as described in Section 4.3 [DOM-based Input Features].
*The shared model name prefix event__selection__egen__vertex__ is omitted for brevity.

Model Name*	**Objective**	**Data restrictions**
starting__events__300m__fast__medium__01	Classification of atmospheric muons versus starting events within 300 m	For ν-simulation: starting within 300 m
starting__nue__300m__fast__medium__01	Classification of atmospheric muons versus starting cascades within 300 m	For ν-simulation: starting ν_e within 300 m
track__numu__cc__vs__nue__cc__01	Classification of charged-current ν_μ versus charged-current ν_e	Starting within 150 m, charged-current
track__numu__cc__vs__starting__01	Classification of charged-current ν_μ versus starting cascades	Starting within 150 m

B.2 Event-Generator Models

The architecture and training procedure of the neural networks utilized in this dissertation are defined in central configuration files. These files are provided in the Event-Generator software framework [87].

Models utilized in this dissertation include:

- **`cascade-7param`:** A single-cascade model with seven parameters. Configuration file: cascade_7param_noise_tw_BFRv1Spice321.yaml.
- **`cascade-7param-low`:** Same as `cascade-7param`, but with a smaller NN size to reduce memory consumption. Configuration file: cascade_7param_noise_tw_BFRv1Spice321_low_mem.yaml.
- **`cascade-13param`:** A single-cascade model including systematic uncertainties as nuisance parameters. Configuration file: cascade_13param_noise_tw_Spice321.yaml.
- **`2-cascade-9param`:** A two-cascade model with two causally connected energy depositions pointing in the same direction. Configuration file: starting_multi_cascade_7param_noise_tw_BFRv1Spice321_low_mem_n002.yaml.
- **`10-cascade-25param`:** A model consisting of ten causally connected energy depositions pointing in the same direction. Configuration file: starting_multi_cascade_7param_noise_tw_BFRv1Spice321_low_mem_n010.yaml.
- **`2-cascade-independent`:** A two-cascade model with two independent cascades. Configuration file: independent_multi_cascade_7param_noise_tw_BFRv1Spice321_low_mem_n002.yaml.

C DNNCascade Selection

This chapter contains additional Material supporting the DNNCascade event selection, introduced in Chapter 7 [DNNCascade: A novel Cascade Dataset]. In Section C.1 [Event Selection Performance], tables are provided detailing the purity, rate, and efficiency of each of the event selection steps. These values are computed for various classes of signal events. Section C.2 [Data/MC Agreement] provides additional data/MC comparison plots for a number of variables. In Section C.3 [Event Views], event views are shown for the top ten most energetic events in the DNNCascade sample.

As part of the analysis unblinding, the most important events of the Galactic plane searches were also unblinded in detector coordinates. These events are selected iteratively by choosing the event with the largest contribution to the test-statistic. Once chosen, the event is excluded from the analysis and the significance is re-evaluated on the remaining events. This procedure is repeated until the analysis significance of the remaining sample drops below 3σ. This test allows to investigate how dependent the analysis result is on an individual event. For the presented Galactic plane analyses, the significance drops continuously with an increased number of events removed; there is no strong dependence on individual events. The event views of these selected events that contribute most to the significance of the Galactic plane analyses are also shown in Section C.3 [Event Views]. The caption for each of the event views denotes the ordering at which the specific event was removed for the π^0, KRA_γ^5, and KRA_γ^{50} analysis. If no number is provided for a specific template search, the event is not part of the top most contributing events for that template.

C.1 Event Selection Performance

Tables describing the selection performance of the DNNCascade sample introduced in Chapter 7 [DNNCascade: A novel Cascade Dataset].

Table C.1: The selection purity based on MC simulation is shown for each of the DNNCascade selection steps and for multiple subsets of simulation.

	Purity / %					
Data Type	**T**	**L0**	**L1**	**L2**	**L3/4**	**L5**
Atmospheric μ	100	100	98	78	40	11
CORSIKA	85	80	70	44	32	5.9
MuonGun	15	20	28	34	7.9	4.9
Neutrinos ν	5.1e-03	2.2e-02	2.0	22	60	89
ν_e	3.4e-04	1.6e-03	0.2	2.0	5.1	23
ν_μ	4.8e-03	2.1e-02	1.8	20	55	64
ν_τ	1.3e-05	7.9e-05	9.5e-03	0.1	0.3	2.7
ν_μ (CC)	4.2e-03	1.8e-02	1.6	17	49	35
ν_μ (CC, contained, $>$ 500 GeV)	4.5e-04	3.6e-03	0.4	4.8	13	27
ν_μ (CC, contained, $>$ 1 TeV)	1.9e-04	1.7e-03	0.2	2.5	7.3	20
ν_μ (CC, contained, $>$ 20 TeV)	1.4e-06	1.5e-05	2.4e-03	3.2e-02	0.1	0.4
ν_e (contained)	2.3e-04	1.3e-03	0.2	1.7	4.4	20
ν_e (contained, $>$ 500 GeV)	4.3e-05	4.3e-04	6.4e-02	0.7	1.8	19
ν_e (contained, $>$ 1 TeV)	1.8e-05	1.9e-04	3.1e-02	0.4	1.0	14
ν_e (contained, $>$ 20 TeV)	3.9e-07	4.6e-06	8.6e-04	1.1e-02	3.6e-02	0.7
ν_μ (contained)	2.2e-03	1.2e-02	1.3	14	37	52
ν_μ (contained, $>$ 500 GeV)	5.5e-04	4.5e-03	0.5	6.2	17	52
ν_μ (contained, $>$ 1 TeV)	2.5e-04	2.3e-03	0.3	3.4	9.5	42
ν_μ (contained, $>$ 20 TeV)	1.9e-06	2.1e-05	3.4e-03	4.5e-02	0.1	1.1
ν_τ (contained)	6.5e-06	5.0e-05	7.2e-03	8.1e-02	0.2	2.2
ν_τ (contained, $>$ 500 GeV)	3.9e-06	3.8e-05	5.9e-03	6.8e-02	0.2	2.2
ν_τ (contained, $>$ 1 TeV)	2.7e-06	2.9e-05	4.6e-03	5.5e-02	0.2	2.1
ν_τ (contained, $>$ 20 TeV)	2.9e-07	3.3e-06	6.0e-04	8.0e-03	2.5e-02	0.4
cascades†	1.2e-03	6.1e-03	0.6	6.6	16	67
cascades† ($>$ 500 GeV)	3.5e-04	2.9e-03	0.3	3.3	8.3	66
cascades† ($>$ 1 TeV)	1.8e-04	1.7e-03	0.2	2.0	5.1	53
cascades† ($>$ 20 TeV)	2.9e-06	3.3e-05	4.7e-03	6.1e-02	0.2	2.6
ν (>500 GeV)	1.9e-03	1.1e-02	1.0	11	31	88
ν (>1 TeV)	1.1e-03	6.3e-03	0.6	6.6	19	70
ν (>20 TeV)	2.5e-05	1.8e-04	1.5e-02	0.2	0.6	3.0
Conventional	5.0e-03	2.2e-02	2.0	22	59	81
Prompt	1.2e-05	7.3e-05	7.9e-03	9.0e-02	0.2	1.6
Astrophysical	4.6e-05	2.8e-04	3.0e-02	0.3	0.9	7.0
Conventional (contained)	2.4e-03	1.3e-02	1.4	15	40	67
Prompt (contained)	5.1e-06	3.9e-05	5.3e-03	6.1e-02	0.2	1.3
Astrophysical (contained)	2.0e-05	1.5e-04	2.1e-02	0.2	0.6	5.7
Conventional (cascades†)	1.2e-03	5.9e-03	0.6	6.3	16	58
Prompt (cascades†)	5.1e-06	4.0e-05	5.1e-03	5.6e-02	0.1	1.5
Astrophysical (cascades†)	2.2e-05	1.7e-04	2.2e-02	0.2	0.6	6.7

Table C.2: The data rate of experimental data and MC simulation is shown for each of the DNNCascade selection steps and for multiple subsets of simulation. Triggered events that pass any one of the online filters are denoted by 'T'; 'L0' refers to the Online Cascade Filter.

	Data Rate / mHz					
Data Type	**T**	**L0**	**L1**	**L2**	**L3/4**	**L5**
Atmospheric μ	384633	30977	158	9.356	1.512	0.020
CORSIKA	327786	24873	113	5.238	1.210	0.011
MuonGun	56847	6104	45.0	4.117	0.302	0.009
Neutrinos ν	19.6	6.920	3.235	2.631	2.287	0.167
ν_e	1.297	0.508	0.309	0.238	0.192	0.043
ν_μ	18.3	6.387	2.911	2.381	2.084	0.119
ν_τ	0.050	0.024	0.015	0.013	0.011	0.005
ν_μ (CC)	16.0	5.708	2.528	2.084	1.849	0.066
ν_μ (CC, contained, $>$ 500 GeV)	1.712	1.103	0.674	0.575	0.505	0.051
ν_μ (CC, contained, $>$ 1 TeV)	0.737	0.533	0.347	0.304	0.277	0.038
ν_μ (CC, contained, $>$ 20 TeV)	0.005	0.005	0.004	0.004	0.004	0.001
ν_e (contained)	0.870	0.389	0.267	0.207	0.167	0.037
ν_e (contained, $>$ 500 GeV)	0.164	0.133	0.104	0.085	0.070	0.036
ν_e (contained, $>$ 1 TeV)	0.069	0.060	0.051	0.044	0.038	0.025
ν_e (contained, $>$ 20 TeV)	0.001	0.001	0.001	0.001	0.001	0.001
ν_μ (contained)	8.340	3.565	2.021	1.643	1.387	0.097
ν_μ (contained, $>$ 500 GeV)	2.134	1.388	0.881	0.742	0.636	0.096
ν_μ (contained, $>$ 1 TeV)	0.953	0.701	0.477	0.411	0.359	0.078
ν_μ (contained, $>$ 20 TeV)	0.007	0.007	0.006	0.005	0.005	0.002
ν_τ (contained)	0.025	0.015	0.012	0.010	0.008	0.004
ν_τ (contained, $>$ 500 GeV)	0.015	0.012	0.010	0.008	0.007	0.004
ν_τ (contained, $>$ 1 TeV)	0.010	0.009	0.007	0.007	0.006	0.004
ν_τ (contained, $>$ 20 TeV)	0.001	0.001	0.001	0.001	0.001	0.001
cascades†	4.727	1.886	1.041	0.790	0.623	0.124
cascades† ($>$ 500 GeV)	1.354	0.906	0.522	0.398	0.317	0.124
cascades† ($>$ 1 TeV)	0.684	0.514	0.310	0.242	0.193	0.100
cascades† ($>$ 20 TeV)	0.011	0.010	0.008	0.007	0.007	0.005
ν ($>$500 GeV)	7.312	3.326	1.580	1.320	1.171	0.165
ν ($>$1 TeV)	4.134	1.967	0.928	0.788	0.712	0.131
ν ($>$20 TeV)	0.095	0.056	0.025	0.023	0.022	0.006
Conventional	19.4	6.813	3.175	2.580	2.243	0.151
Prompt	0.046	0.023	0.013	0.011	0.009	0.003
Astrophysical	0.176	0.086	0.049	0.041	0.036	0.013
Conventional (contained)	9.139	3.912	2.258	1.825	1.531	0.125
Prompt (contained)	0.020	0.012	0.009	0.007	0.006	0.002
Astrophysical (contained)	0.078	0.047	0.034	0.029	0.025	0.011
Conventional (cascades†)	4.624	1.821	0.998	0.755	0.593	0.109
Prompt (cascades†)	0.020	0.013	0.008	0.007	0.006	0.003
Astrophysical (cascades†)	0.084	0.053	0.036	0.029	0.024	0.012
MC simulation	384653	30984	162	12.0	3.798	0.187
Experimental data	444008	32754	109	25.8	5.225	0.196

Table C.3: The selection efficiency of experimental data and MC simulation with respect to triggered events that passed any of the online filters (T) is shown for each of the DNNCascade selection steps and for multiple subsets of simulation.

	Efficiency / %					
Data Type	**T**	**L0**	**L1**	**L2**	**L3/4**	**L5**
Atmospheric μ	100	8.1	0.04	2.43e-03	3.93e-04	5.26e-06
CORSIKA	100	7.6	0.03	1.60e-03	3.69e-04	3.37e-06
MuonGun	100	10.7	0.08	7.24e-03	5.31e-04	1.61e-05
Neutrinos ν	100	35.2	16.47	13.40	11.64	0.85
ν_e	100	39.1	23.80	18.31	14.81	3.30
ν_μ	100	34.9	15.91	13.02	11.39	0.65
ν_τ	100	48.9	30.57	25.28	21.45	10.02
ν_μ (CC)	100	35.6	15.75	12.99	11.53	0.41
ν_μ (CC, contained, $>$ 500 GeV)	100	64.4	39.38	33.58	29.49	2.97
ν_μ (CC, contained, $>$ 1 TeV)	100	72.4	47.10	41.19	37.58	5.16
ν_μ (CC, contained, $>$ 20 TeV)	100	89.4	73.24	71.20	72.69	15.12
ν_e (contained)	100	44.7	30.68	23.85	19.16	4.20
ν_e (contained, $>$ 500 GeV)	100	81.1	63.08	51.75	42.51	21.99
ν_e (contained, $>$ 1 TeV)	100	86.7	72.86	63.20	54.71	36.49
ν_e (contained, $>$ 20 TeV)	100	95.3	92.75	91.96	92.42	82.09
ν_μ (contained)	100	42.7	24.23	19.70	16.63	1.16
ν_μ (contained, $>$ 500 GeV)	100	65.1	41.28	34.78	29.81	4.52
ν_μ (contained, $>$ 1 TeV)	100	73.6	50.05	43.09	37.68	8.21
ν_μ (contained, $>$ 20 TeV)	100	91.2	76.50	74.39	74.44	28.14
ν_τ (contained)	100	61.2	46.40	38.88	33.46	16.33
ν_τ (contained, $>$ 500 GeV)	100	79.5	63.57	54.74	47.57	27.32
ν_τ (contained, $>$ 1 TeV)	100	85.2	70.94	62.77	55.62	36.93
ν_τ (contained, $>$ 20 TeV)	100	92.5	88.15	86.69	84.60	69.87
cascades†	100	39.9	22.02	16.71	13.17	2.63
cascades† ($>$ 500 GeV)	100	66.9	38.55	29.38	23.42	9.13
cascades† ($>$ 1 TeV)	100	75.1	45.38	35.44	28.22	14.57
cascades† ($>$ 20 TeV)	100	93.2	69.22	66.36	59.95	43.51
ν ($>$500 GeV)	100	45.5	21.61	18.06	16.01	2.26
ν ($>$1 TeV)	100	47.6	22.46	19.06	17.23	3.18
ν ($>$20 TeV)	100	58.8	26.25	24.14	22.67	5.93
Conventional	100	35.1	16.35	13.29	11.55	0.78
Prompt	100	48.9	27.65	23.27	20.35	6.44
Astrophysical	100	48.6	27.99	23.49	20.26	7.48
Conventional (contained)	100	42.8	24.70	19.97	16.76	1.36
Prompt (contained)	100	61.1	43.86	37.05	31.89	12.42
Astrophysical (contained)	100	60.2	43.89	37.01	31.79	13.74
Conventional (cascades†)	100	39.4	21.58	16.32	12.82	2.36
Prompt (cascades†)	100	63.6	42.11	34.36	28.79	14.30
Astrophysical (cascades†)	100	63.2	42.18	34.59	28.95	14.77
MC simulation	100	8.1	0.04	3.12e-03	9.87e-04	4.86e-05
Experimental data	100	7.4	0.02	5.81e-03	1.18e-03	4.41e-05

C.2 Data/MC Agreement

Data/MC comparison plots for the DNNCascade event selection are shown for various variables.

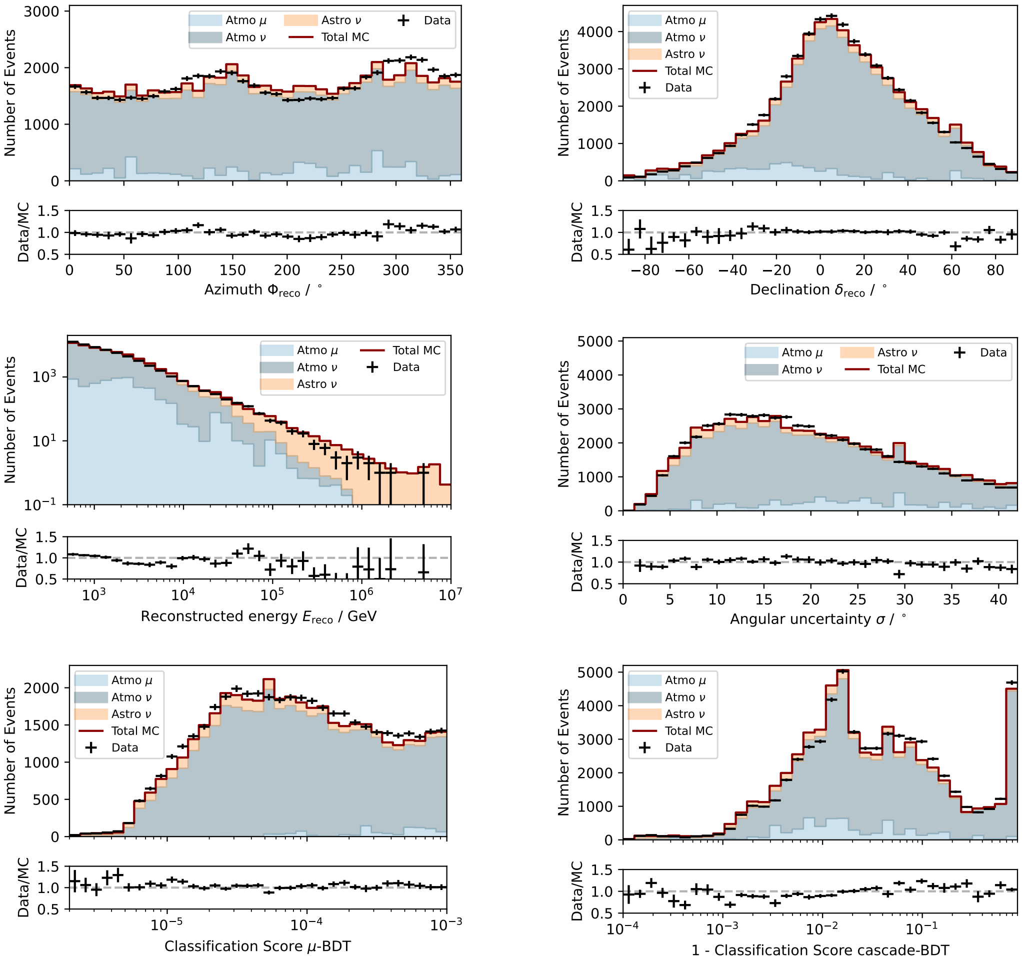


Figure C.1: The experimental data distributions in key analysis observables are compared to expectations obtained from MC simulations. Note that these are the same as in Figure 7.15, added here to facilitate comparison. Systematic uncertainties are not included.

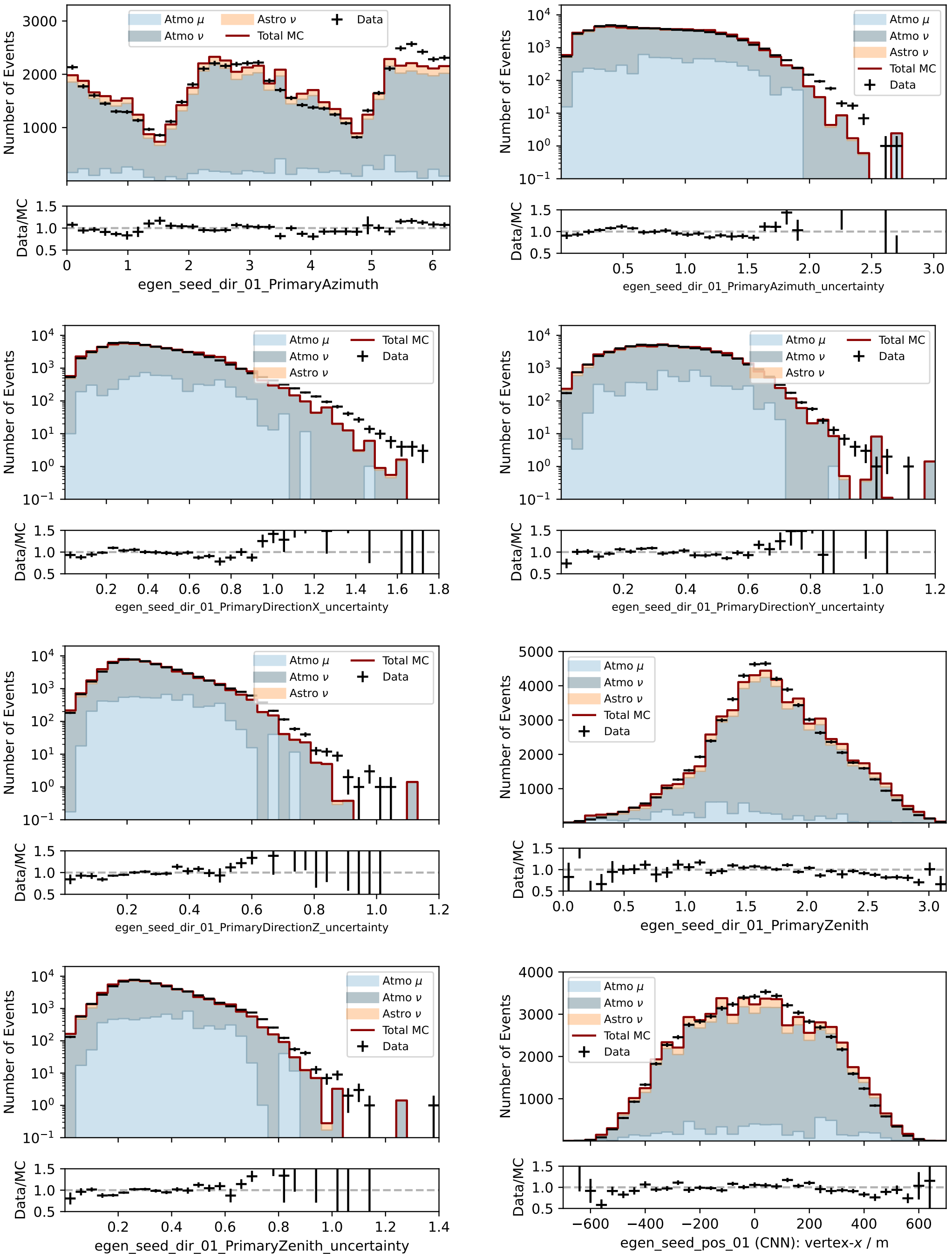


Figure C.2: Experimental data distributions of DNNCascade regression models are compared to expectations obtained from MC simulations. Systematic uncertainties are not included.

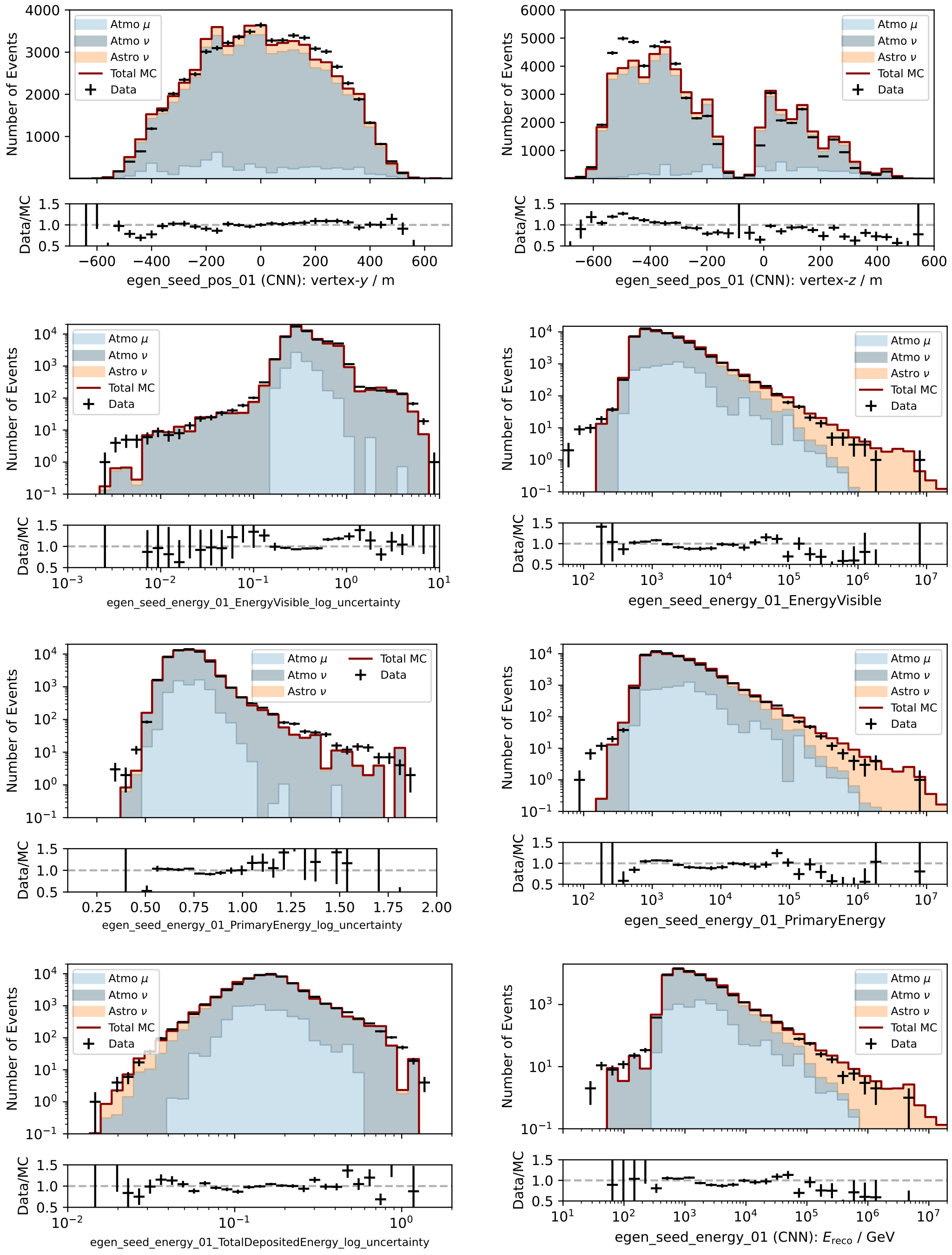


Figure C.3: Experimental data distributions of DNNCascade regression models are compared to expectations obtained from MC simulations. Systematic uncertainties are not included.

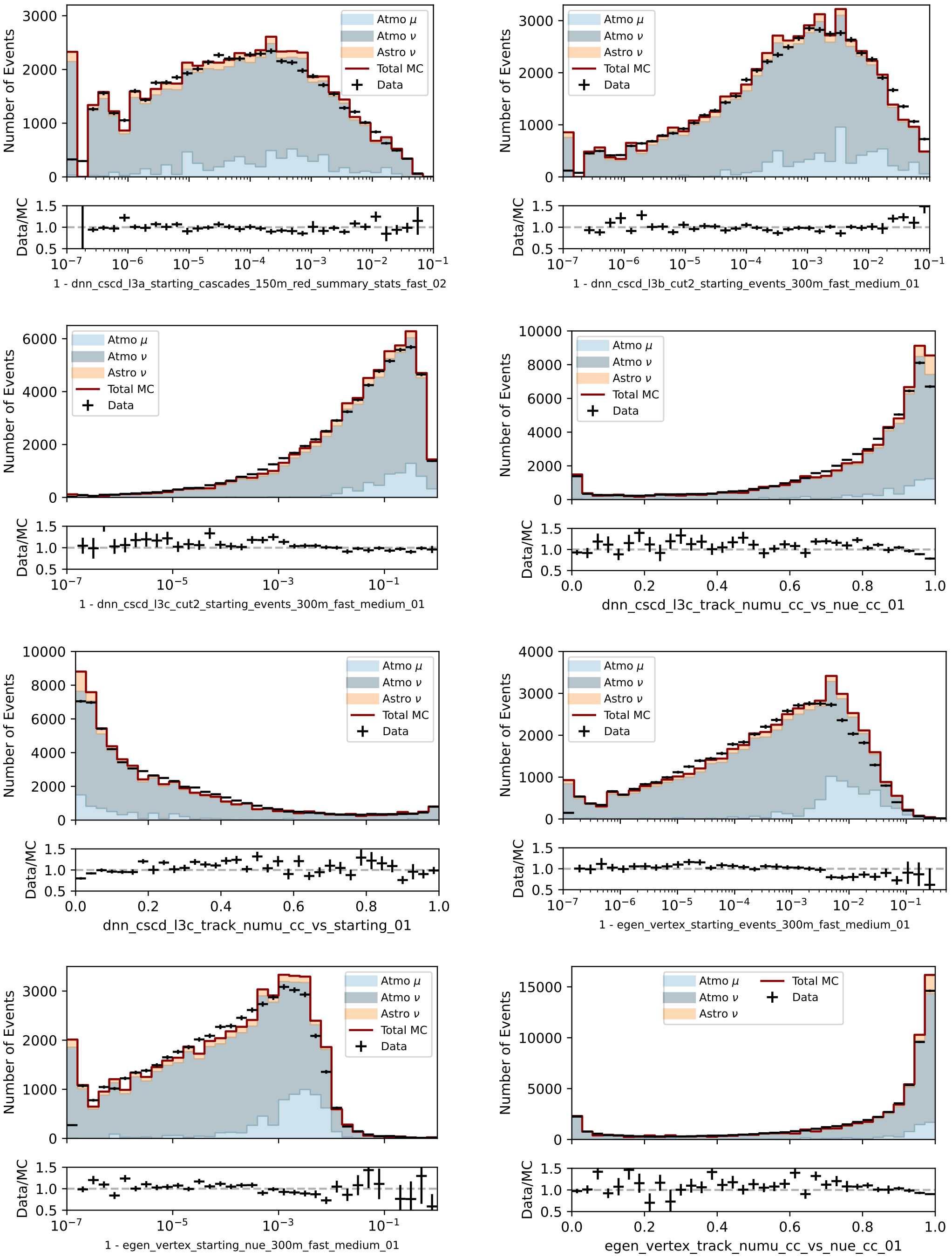


Figure C.4: Experimental data distributions of DNNCascade classification models are compared to expectations obtained from MC simulations. Systematic uncertainties are not included.

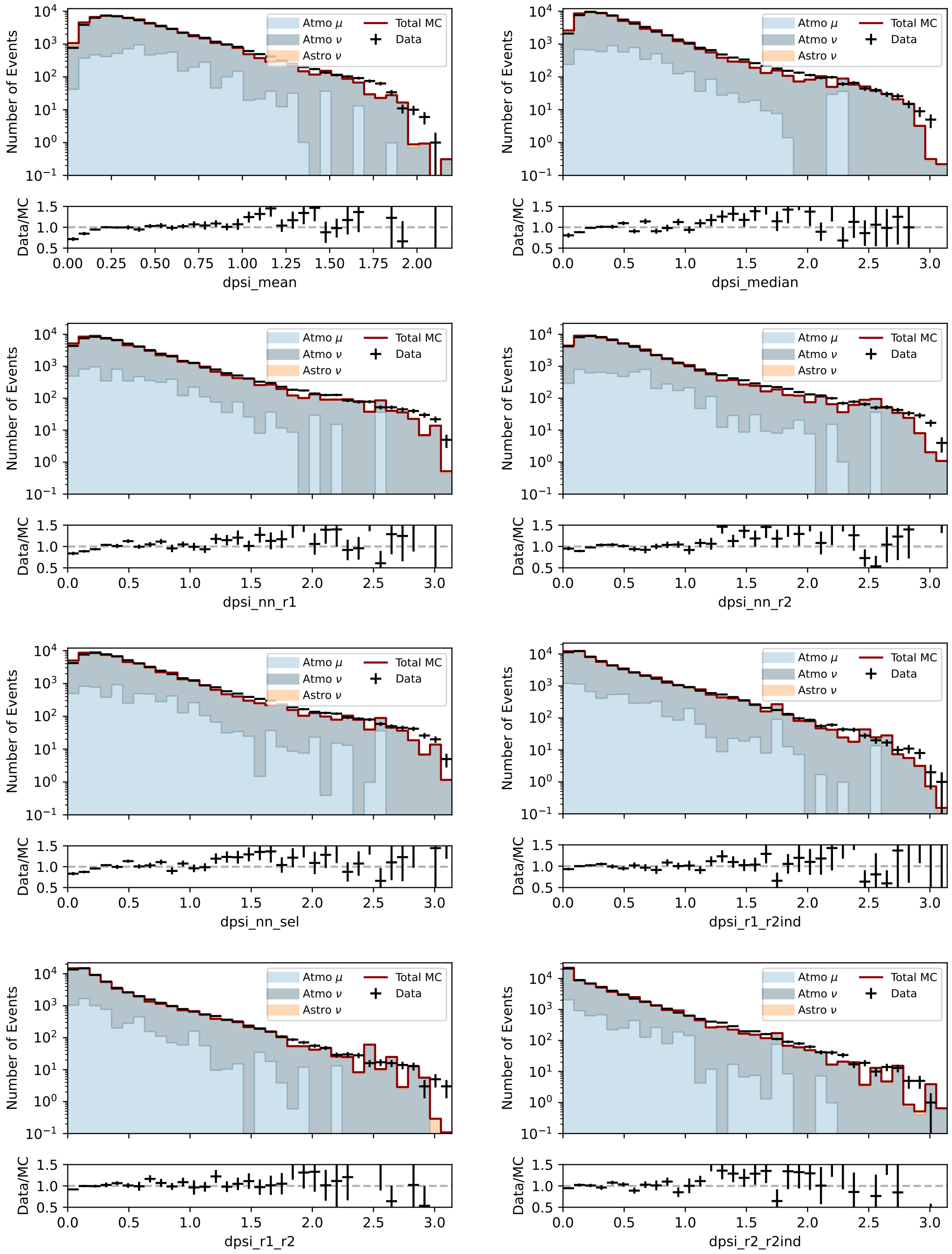


Figure C.5: Experimental data distributions of observables used in the σ-NNs are compared to expectations obtained from MC simulations. These are opening angles between different reconstruction methods of both DNN-reco and Event-Generator models. Systematic uncertainties are not included.

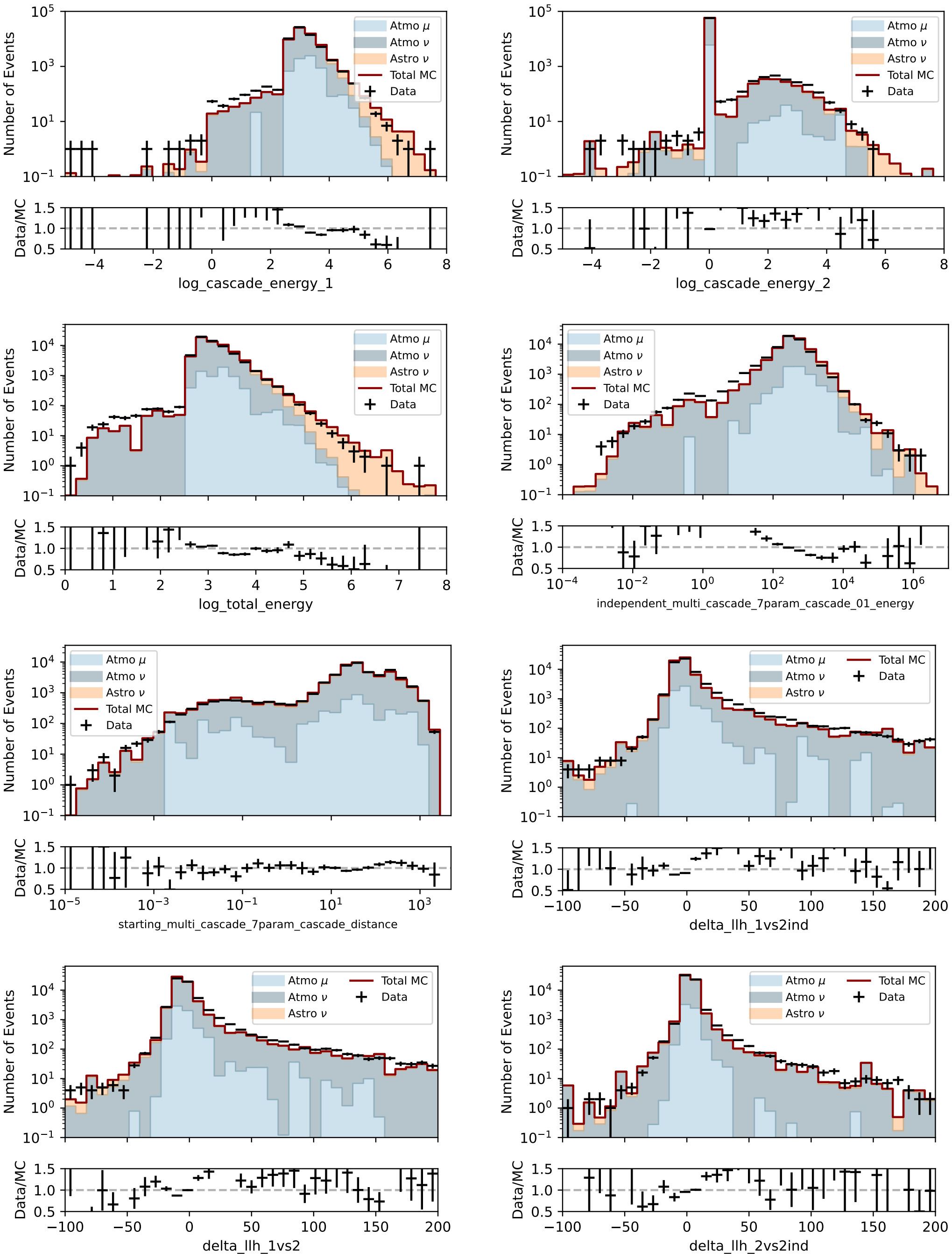


Figure C.6: Experimental data distributions of observables used in the σ-NNs are compared to expectations obtained from MC simulations. Systematic uncertainties are not included.

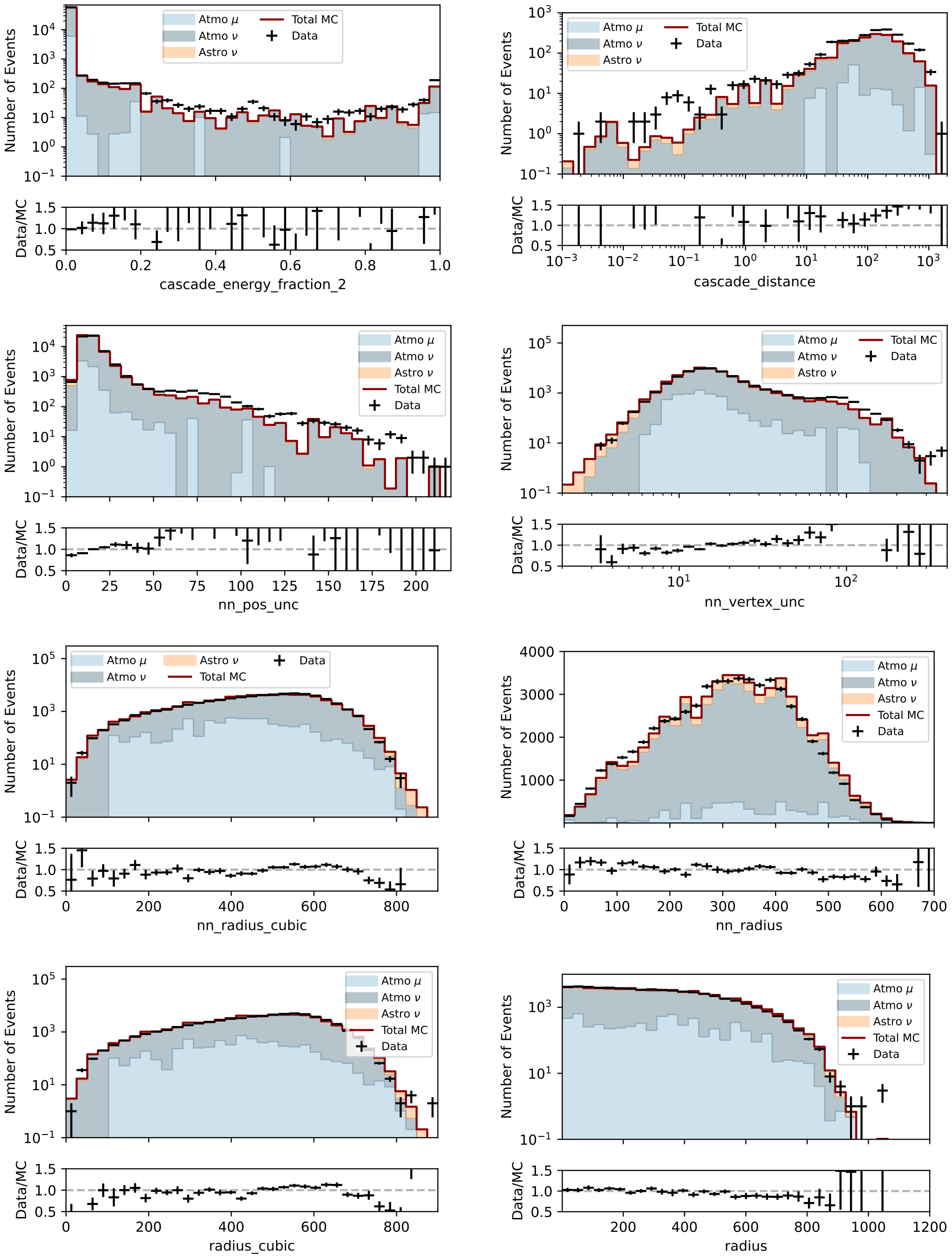


Figure C.7: Experimental data distributions of observables used in the σ-NNs are compared to expectations obtained from MC simulations. Systematic uncertainties are not included.

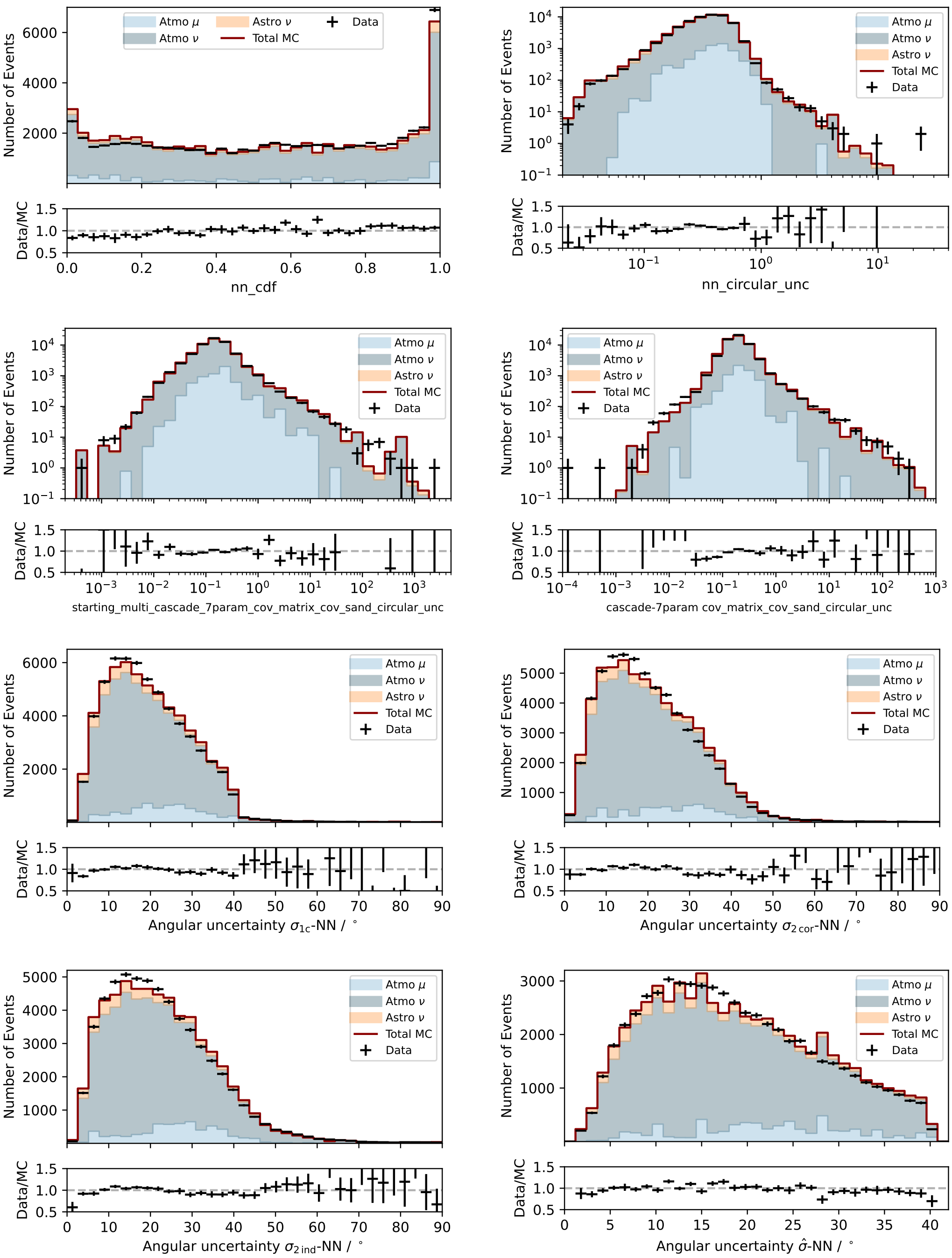
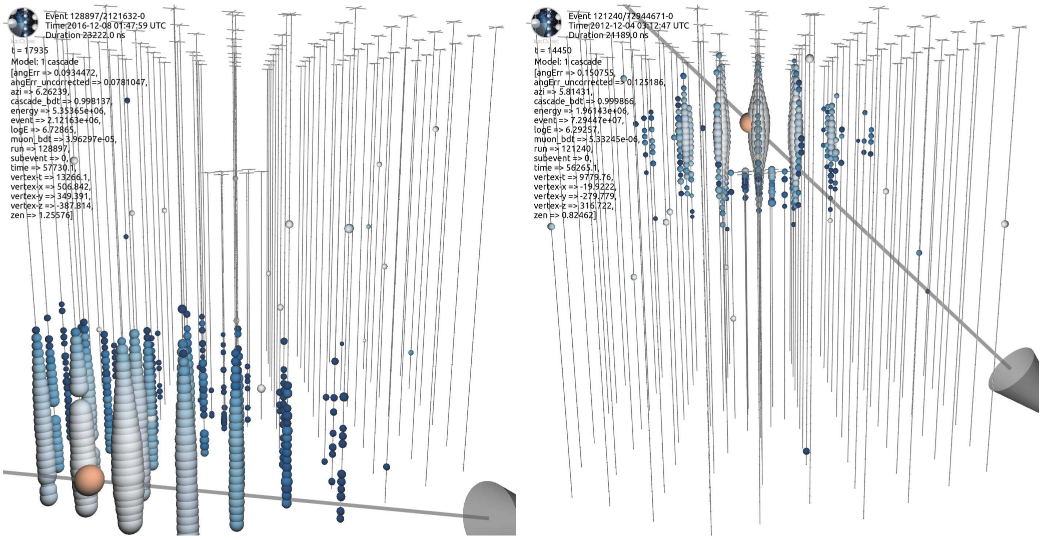


Figure C.8: Experimental data distributions of observables used in the σ-NNs as well as the σ-NNs themseleves are compared to expectations obtained from MC simulations. Systematic uncertainties are not included.

C.3 Event Views

Event views of the DNNCascade selection are shown for events with the highest reconstructed energy and for the events contributing most to the Galactic plane result. Colored blobs indicate light measured at DOMs in the IceCube detector, from early times (light) to late times (dark blue). The orange sphere and gray arrow illustrate the selected Event-Generator reconstruction result (Section 7.3 [Event Reconstruction and Uncertainty Estimation]). Detailed values of the reconstructed quantities are provided in the text on the top left of each event view. The first three lines provide the event header information. The following two lines show at which relative event time (t) the event view was generated and the name of the selected Event-Generator event hypothesis.

The unblinded events are high-quality events with cascade topologies as expected; no signs of background contamination are visible. This is demonstrative of the advertised performance of the DNNCascade event selection.



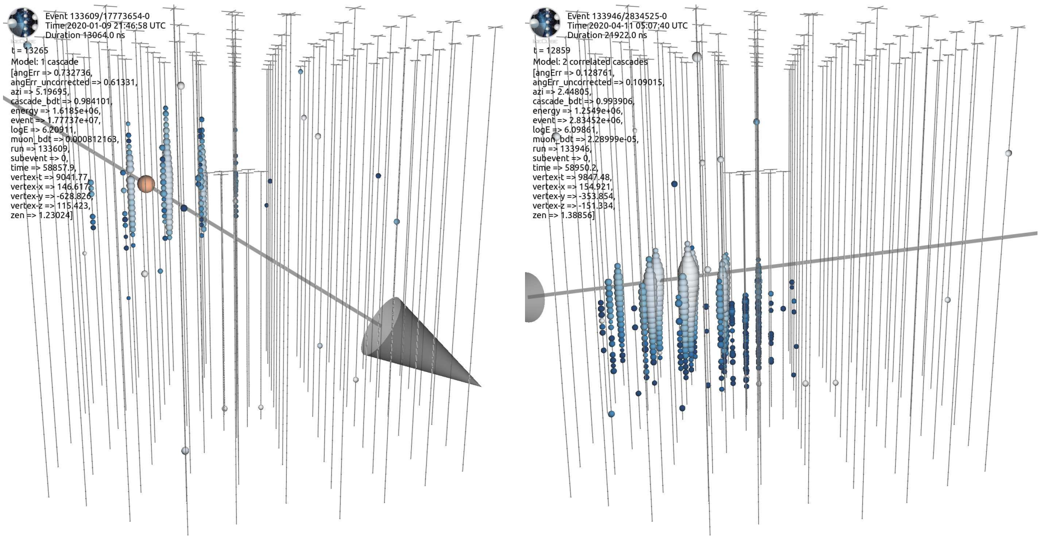
```
Event 128897/2121632-0
Time 2016-12-08 01:47:59 UTC
Duration 23222.0 ns
t = 17935
Model: 1 cascade
[angErr => 0.0934472,
angErr_uncorrected => 0.0781047,
azi => 6.26239,
cascade_bdt => 0.998137,
energy => 5.35365e+06,
event => 2.12163e+06,
logE => 6.72865,
muon_bdt => 3.96297e-05,
run => 128897,
subevent => 0,
time => 57730.1,
vertex-t => 13266.1,
vertex-x => 506.842,
vertex-y => 349.391,
vertex-z => -387.814,
zen => 1.25576]
```

(a) 5.4 PeV.

```
Event 121240/72944671-0
Time 2012-12-04 03:12:47 UTC
Duration 21189.0 ns
t = 14450
Model: 1 cascade
[angErr => 0.150755,
angErr_uncorrected => 0.125186,
azi => 5.81431,
cascade_bdt => 0.999866,
energy => 1.96143e+06,
event => 7.29447e+07,
logE => 6.29257,
muon_bdt => 5.33245e-06,
run => 121240,
subevent => 0,
time => 56265.1,
vertex-t => 9779.76,
vertex-x => -19.9222,
vertex-y => -279.779,
vertex-z => 316.722,
zen => 0.82462]
```

(b) 2.0 PeV.

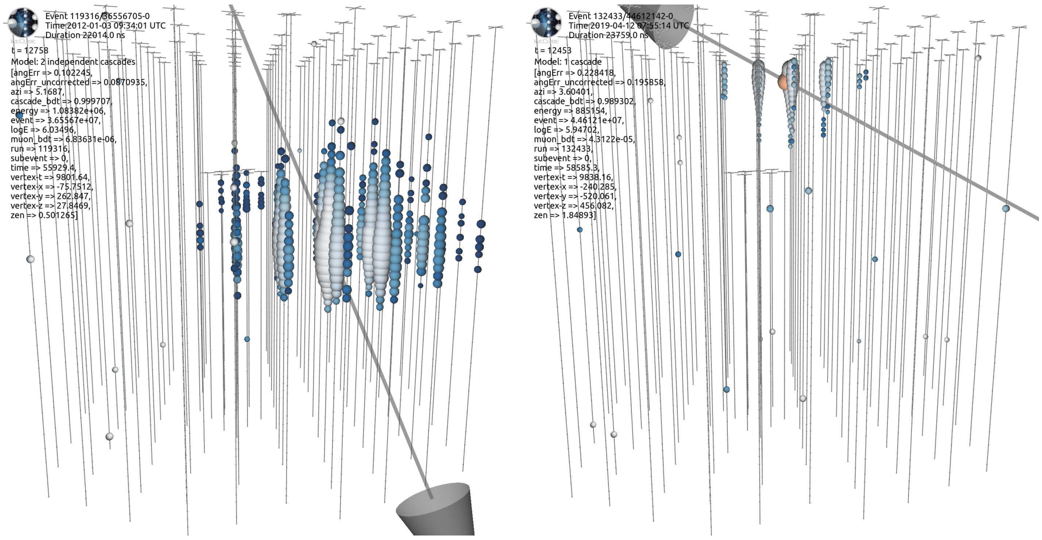
Figure C.9: Events 1–2 of the top ten events in the DNNCascade sample with the highest reconstructed deposited energies are shown.



```
Event 133609/17773654-0
Time 2020-01-09 21:46:58 UTC
Duration 13064.0 ns
t = 13265
Model: 1 cascade
[angErr => 0.732736,
angErr_uncorrected => 0.61331,
azi => 5.19695,
cascade_bdt => 0.984101,
energy => 1.6185e+06,
event => 1.77737e+07,
logE => 6.20911,
muon_bdt => 0.000812163,
run => 133609,
subevent => 0,
time => 58857.9,
vertex-t => 9041.77,
vertex-x => 146.617,
vertex-y => -628.826,
vertex-z => 115.423,
zen => 1.23024]
```

```
Event 133946/2834525-0
Time 2020-04-11 05:07:40 UTC
Duration 21922.0 ns
t = 12859
Model: 2 correlated cascades
[angErr => 0.128761,
angErr_uncorrected => 0.109015,
azi => 2.44805,
cascade_bdt => 0.993906,
energy => 1.2549e+06,
event => 2.83452e+06,
logE => 6.09861,
muon_bdt => 2.28999e-05,
run => 133946,
subevent => 0,
time => 58950.2,
vertex-t => 9847.48,
vertex-x => 154.921,
vertex-y => -353.854,
vertex-z => -151.334,
zen => 1.38856]
```

(a) 1.6 PeV. **(b)** 1.3 PeV.

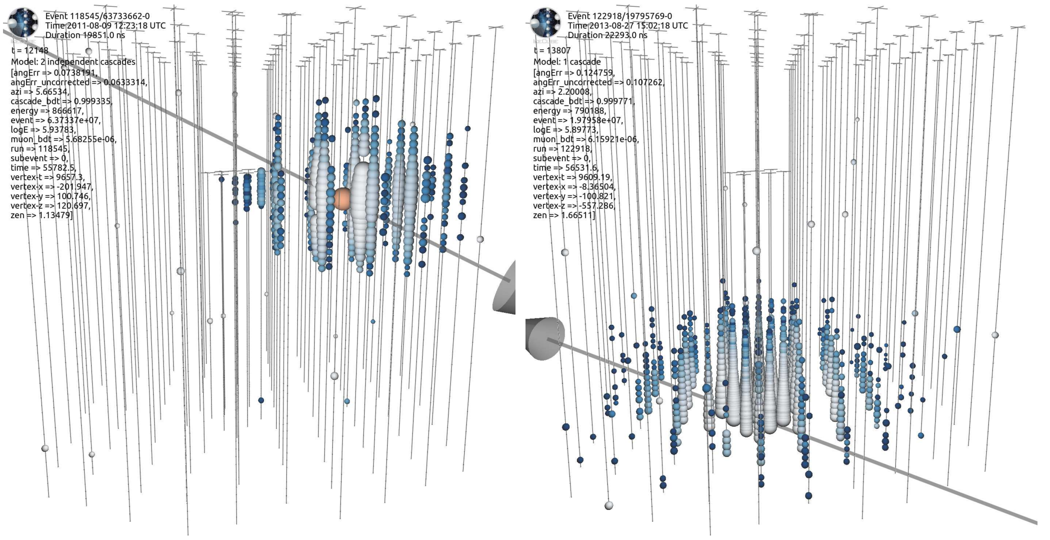


```
Event 119316/36556705-0
Time 2012-01-03 09:34:01 UTC
Duration 22014.0 ns
t = 12758
Model: 2 independent cascades
[angErr => 0.102245,
angErr_uncorrected => 0.0870935,
azi => 5.1687,
cascade_bdt => 0.999707,
energy => 1.08382e+06,
event => 3.65567e+07,
logE => 6.03496,
muon_bdt => 6.83631e-06,
run => 119316,
subevent => 0,
time => 55929.4,
vertex-t => 9801.64,
vertex-x => -75.7512,
vertex-y => 262.847,
vertex-z => 27.8469,
zen => 0.501265]
```

```
Event 132433/44612142-0
Time 2019-04-12 07:55:14 UTC
Duration 23759.0 ns
t = 12453
Model: 1 cascade
[angErr => 0.228418,
angErr_uncorrected => 0.195858,
azi => 3.60401,
cascade_bdt => 0.989302,
energy => 885154,
event => 4.46121e+07,
logE => 5.94702,
muon_bdt => 4.3122e-05,
run => 132433,
subevent => 0,
time => 58585.3,
vertex-t => 9838.16,
vertex-x => -240.285,
vertex-y => -520.061,
vertex-z => 456.082,
zen => 1.84893]
```

(c) 1.1 PeV. **(d)** 885 TeV.

Figure C.10: Events 3–6 of the top ten events in the DNNCascade sample with the highest reconstructed deposited energies are shown.

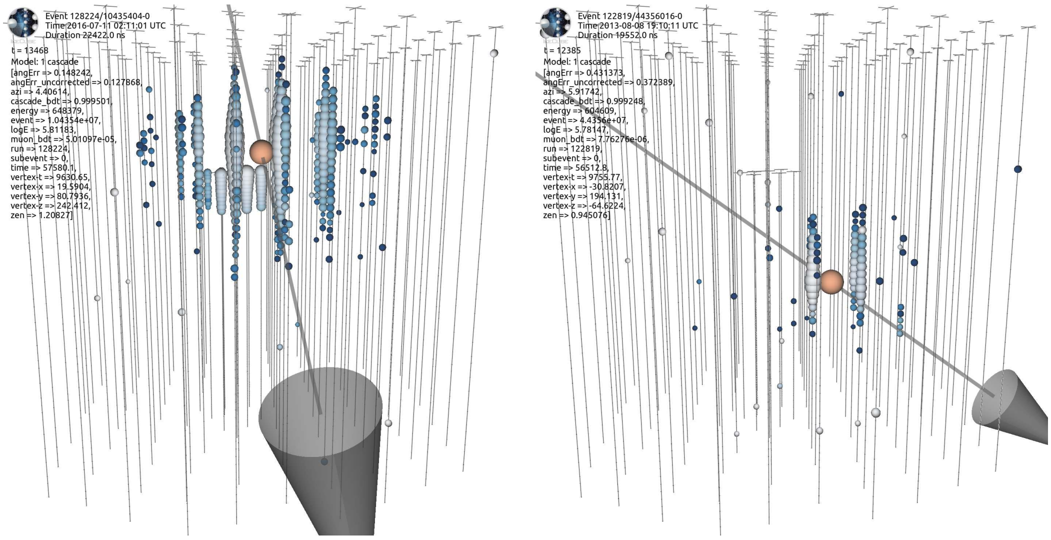


```
Event 118545/63733662-0
Time 2011-08-09 12:23:18 UTC
Duration 19851.0 ns
t = 12148
Model: 2 independent cascades
[angErr => 0.0738191,
angErr_uncorrected => 0.0633314,
azi => 5.66534,
cascade_bdt => 0.999335,
energy => 866617,
event => 6.37337e+07,
logE => 5.93783,
muon_bdt => 5.68255e-06,
run => 118545,
subevent => 0,
time => 55782.5,
vertex-t => 9657.3,
vertex-x => -201.947,
vertex-y => 100.746,
vertex-z => 120.697,
zen => 1.13479]
```

```
Event 122918/19795769-0
Time 2013-08-27 15:02:18 UTC
Duration 22293.0 ns
t = 13807
Model: 1 cascade
[angErr => 0.124759,
angErr_uncorrected => 0.107262,
azi => 2.20008,
cascade_bdt => 0.999771,
energy => 790188,
event => 1.97958e+07,
logE => 5.89773,
muon_bdt => 6.15921e-06,
run => 122918,
subevent => 0,
time => 56531.6,
vertex-t => 9609.19,
vertex-x => -8.36504,
vertex-y => -100.821,
vertex-z => -557.286,
zen => 1.66511]
```

(a) 867 TeV.

(b) 790 TeV.



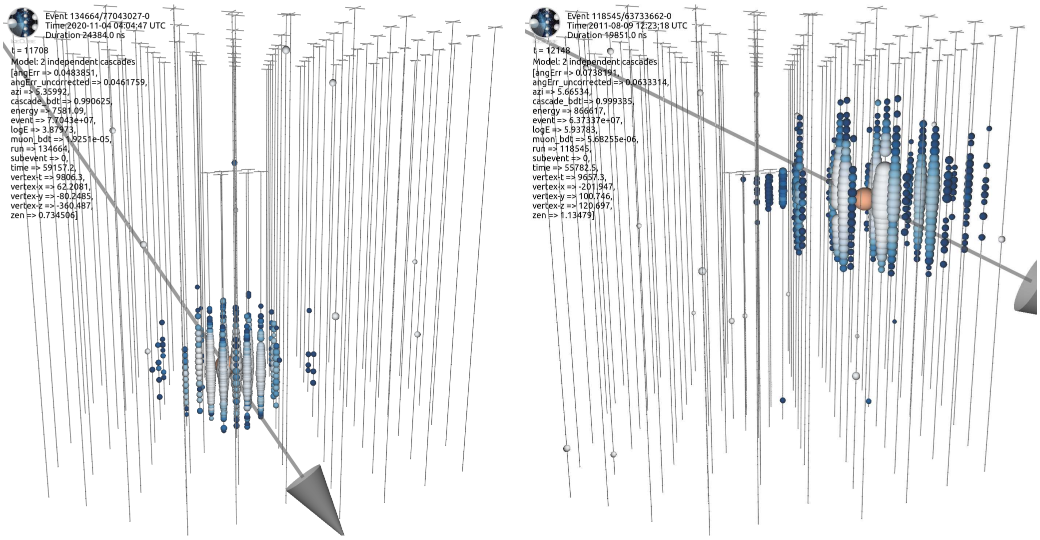
```
Event 128224/10435404-0
Time 2016-07-11 02:11:01 UTC
Duration 22422.0 ns
t = 13468
Model: 1 cascade
[angErr => 0.148242,
angErr_uncorrected => 0.127868,
azi => 4.40614,
cascade_bdt => 0.999501,
energy => 648379,
event => 1.04354e+07,
logE => 5.81183,
muon_bdt => 5.01097e-05,
run => 128224,
subevent => 0,
time => 57580.1,
vertex-t => 9630.65,
vertex-x => 19.5904,
vertex-y => 80.7936,
vertex-z => 242.412,
zen => 1.20827]
```

```
Event 122819/44356016-0
Time 2013-08-08 19:10:11 UTC
Duration 19552.0 ns
t = 12385
Model: 1 cascade
[angErr => 0.431373,
angErr_uncorrected => 0.372389,
azi => 5.91742,
cascade_bdt => 0.999248,
energy => 604609,
event => 4.4356e+07,
logE => 5.78147,
muon_bdt => 7.76276e-06,
run => 122819,
subevent => 0,
time => 56512.8,
vertex-t => 9755.77,
vertex-x => -30.8207,
vertex-y => 194.131,
vertex-z => -64.6224,
zen => 0.945076]
```

(c) 648 TeV.

(d) 605 TeV.

Figure C.11: Events 7–10 of the top ten events in the DNNCascade sample with the highest reconstructed deposited energies are shown.

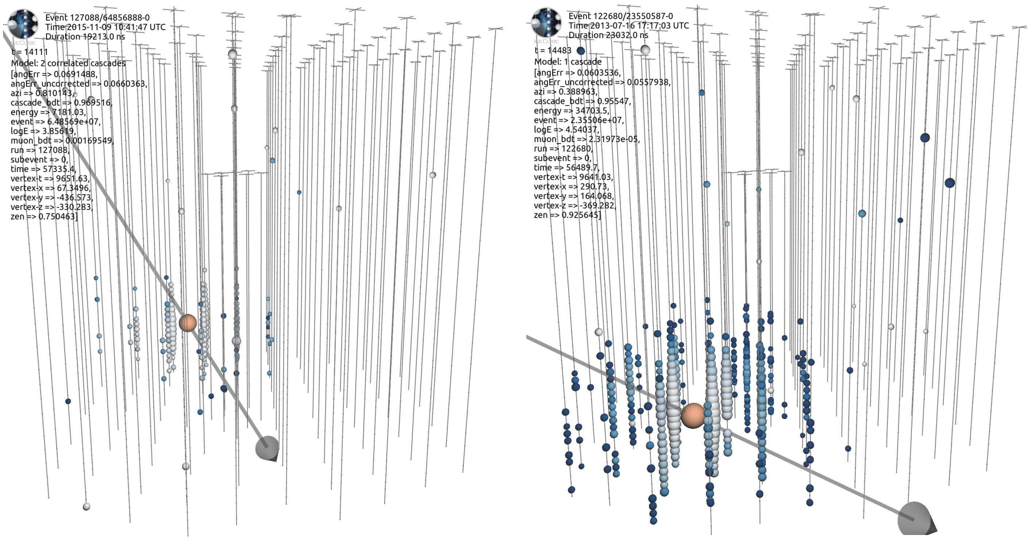


```
Event 134664/77043027-0
Time 2020-11-04 04:04:47 UTC
Duration 24384.0 ns
t = 11708
Model: 2 independent cascades
[angErr => 0.0483851,
angErr_uncorrected => 0.0461759,
azi => 5.35992,
cascade_bdt => 0.990625,
energy => 7581.09,
event => 7.7043e+07,
logE => 3.87973,
muon_bdt => 1.9251e-05,
run => 134664,
subevent => 0,
time => 59157.2,
vertex-t => 9806.3,
vertex-x => 62.2081,
vertex-y => -80.2485,
vertex-z => -360.487,
zen => 0.734506]
```

```
Event 118545/63733662-0
Time 2011-08-09 12:23:18 UTC
Duration 19851.0 ns
t = 12148
Model: 2 independent cascades
[angErr => 0.0738191,
angErr_uncorrected => 0.0633314,
azi => 5.66534,
cascade_bdt => 0.999335,
energy => 866617,
event => 6.37337e+07,
logE => 5.93783,
muon_bdt => 5.68255e-06,
run => 118545,
subevent => 0,
time => 55782.5,
vertex-t => 9657.3,
vertex-x => -201.947,
vertex-y => 100.746,
vertex-z => 120.697,
zen => 1.13479]
```

(a) π^0: #1; KRA_γ^5: #5; KRA_γ^{50}: #8.

(b) π^0: #2; KRA_γ^{50}: #4.



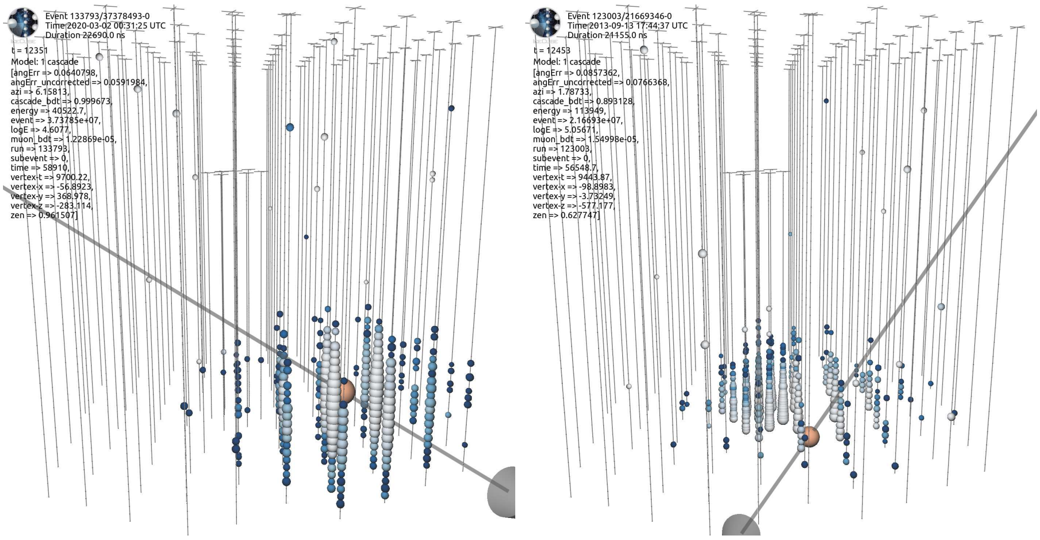
```
Event 127088/64856888-0
Time 2015-11-09 10:41:47 UTC
Duration 19213.0 ns
t = 14111
Model: 2 correlated cascades
[angErr => 0.0691488,
angErr_uncorrected => 0.0660363,
azi => 0.810143,
cascade_bdt => 0.969516,
energy => 7181.03,
event => 6.48569e+07,
logE => 3.85619,
muon_bdt => 0.00169549,
run => 127088,
subevent => 0,
time => 57335.4,
vertex-t => 9651.63,
vertex-x => 67.3496,
vertex-y => -436.573,
vertex-z => -330.283,
zen => 0.750463]
```

```
Event 122680/23550587-0
Time 2013-07-16 17:17:03 UTC
Duration 23032.0 ns
t = 14483
Model: 1 cascade
[angErr => 0.0603536,
angErr_uncorrected => 0.0557938,
azi => 0.388963,
cascade_bdt => 0.95547,
energy => 34703.5,
event => 2.35506e+07,
logE => 4.54037,
muon_bdt => 2.31973e-05,
run => 122680,
subevent => 0,
time => 56489.7,
vertex-t => 9641.03,
vertex-x => 290.73,
vertex-y => 164.068,
vertex-z => -369.282,
zen => 0.925645]
```

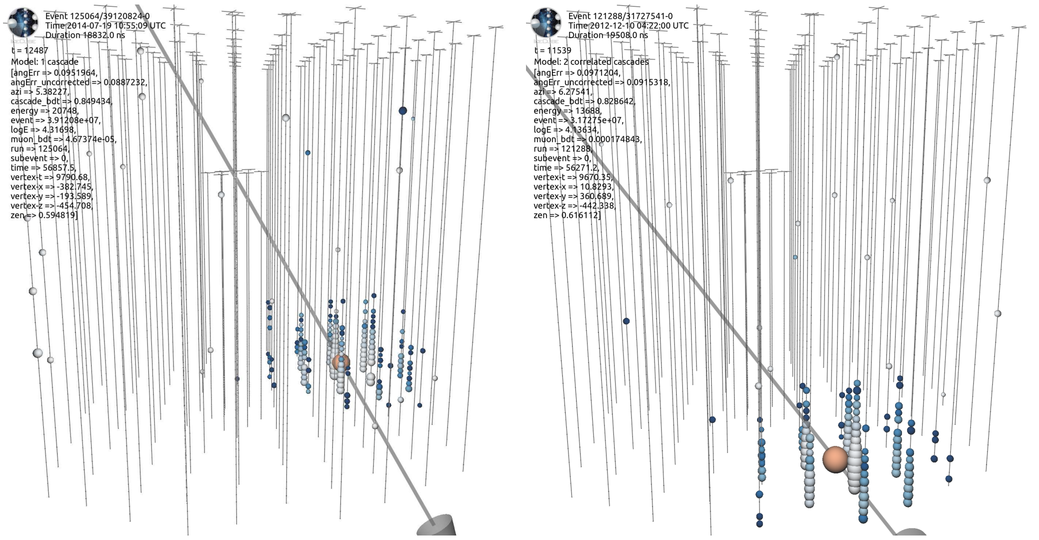
(c) π^0: #3.

(d) π^0: #4; KRA_γ^5: #3; KRA_γ^{50}: #3.

Figure C.12: Events 1–4 of the top events contributing most to the Galactic plane analyses. The ranking for each of the analysis templates is given in the captions.

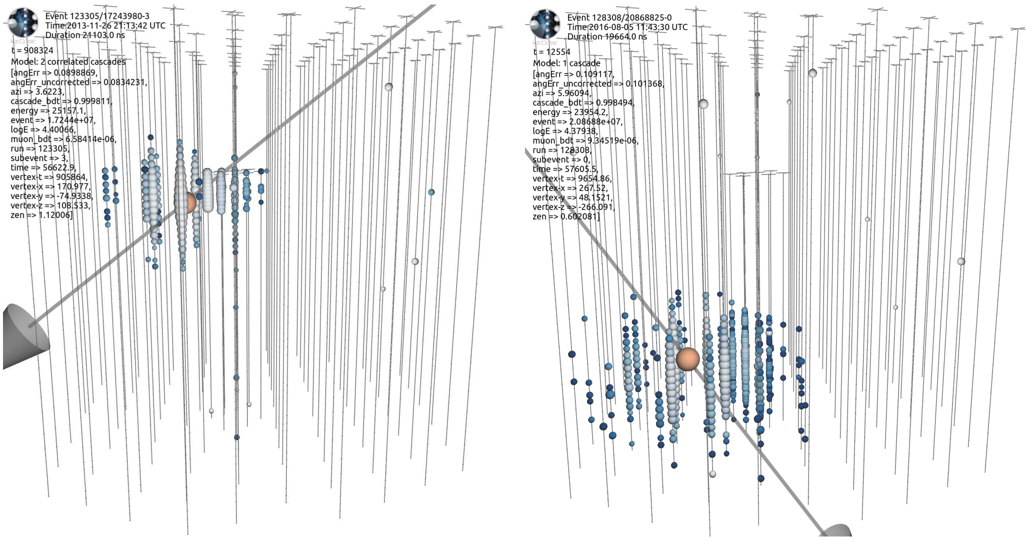


(a) π^0: #5; KRA_γ^5: #2; KRA_γ^{50}: #2. **(b)** π^0: #6; KRA_γ^5: #6; KRA_γ^{50}: #6.



(c) π^0: #7. **(d)** π^0: #8.

Figure C.13: Events 5–8 of the top events contributing most to the Galactic plane analyses. The ranking for each of the analysis templates is given in the captions.

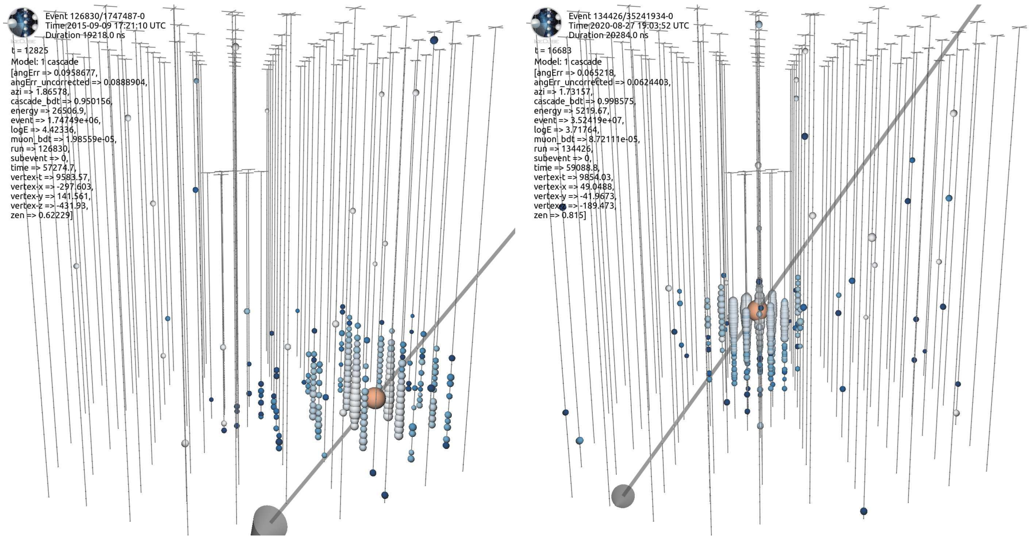


```
Event 123305/17243980-3
Time 2013-11-26 21:13:42 UTC
Duration 21103.0 ns
t = 908324
Model: 2 correlated cascades
[angErr => 0.0898869,
angErr_uncorrected => 0.0834231,
azi => 3.6223,
cascade_bdt => 0.999811,
energy => 25157.1,
event => 1.7244e+07,
logE => 4.40066,
muon_bdt => 6.58414e-06,
run => 123305,
subevent => 3,
time => 56622.9,
vertex-t => 905864,
vertex-x => 170.977,
vertex-y => -74.9338,
vertex-z => 108.533,
zen => 1.12006]
```

```
Event 128308/20868825-0
Time 2016-08-05 11:43:30 UTC
Duration 19664.0 ns
t = 12554
Model: 1 cascade
[angErr => 0.109117,
angErr_uncorrected => 0.101368,
azi => 5.96094,
cascade_bdt => 0.998494,
energy => 23954.2,
event => 2.08688e+07,
logE => 4.37938,
muon_bdt => 9.34519e-06,
run => 128308,
subevent => 0,
time => 57605.5,
vertex-t => 9654.86,
vertex-x => 267.52,
vertex-y => 48.1521,
vertex-z => -266.091,
zen => 0.602081]
```

(a) π^0: #9; KRA_γ^5: #4; KRA_γ^{50}: #5.

(b) π^0: #10.



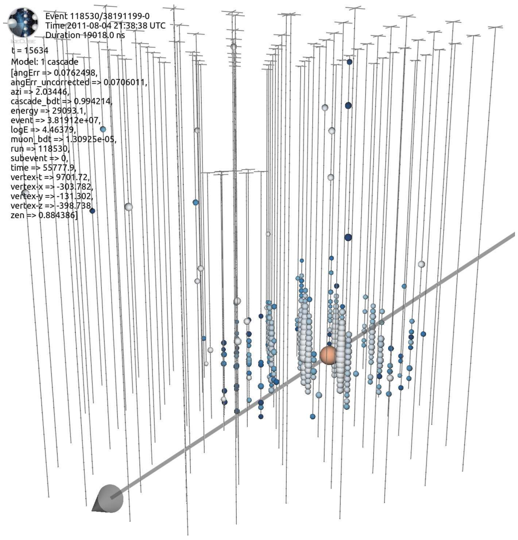
```
Event 126830/1747487-0
Time 2015-09-09 17:21:10 UTC
Duration 19218.0 ns
t = 12825
Model: 1 cascade
[angErr => 0.0958677,
angErr_uncorrected => 0.0888904,
azi => 1.86578,
cascade_bdt => 0.950156,
energy => 26506.9,
event => 1.74749e+06,
logE => 4.42336,
muon_bdt => 1.98559e-05,
run => 126830,
subevent => 0,
time => 57274.7,
vertex-t => 9583.57,
vertex-x => -297.603,
vertex-y => 141.561,
vertex-z => -431.93,
zen => 0.62229]
```

```
Event 134426/35241934-0
Time 2020-08-27 19:03:52 UTC
Duration 20284.0 ns
t = 16683
Model: 1 cascade
[angErr => 0.065218,
angErr_uncorrected => 0.0624403,
azi => 1.73157,
cascade_bdt => 0.998575,
energy => 5219.67,
event => 3.52419e+07,
logE => 3.71764,
muon_bdt => 8.72111e-05,
run => 134426,
subevent => 0,
time => 59088.8,
vertex-t => 9854.03,
vertex-x => 49.0488,
vertex-y => -41.9673,
vertex-z => -189.473,
zen => 0.815]
```

(c) π^0: #11.

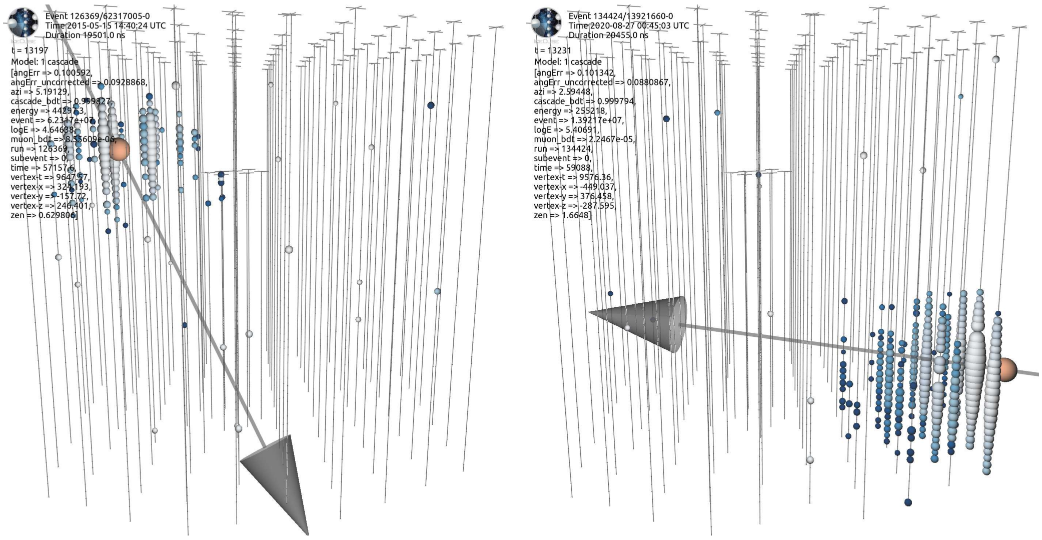
(d) π^0: #12.

Figure C.14: Events 9–12 of the top events contributing most to the Galactic plane analyses. The ranking for each of the analysis templates is given in the captions.



(a) π^0: #13; KRA$_\gamma^5$: #8.

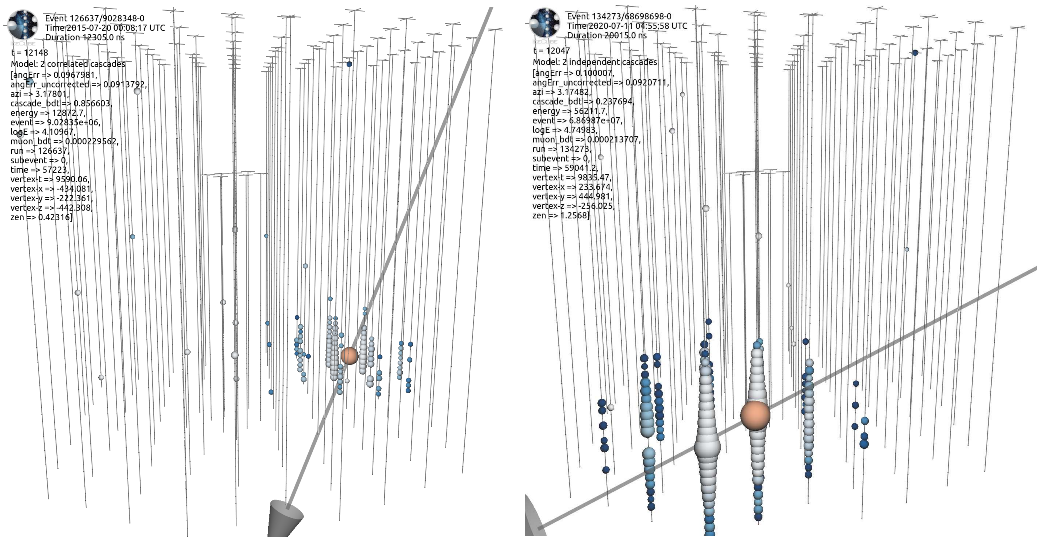
(b) π^0: #14.



(c) π^0: #15.

(d) π^0: #16.

Figure C.15: Events 13–16 of the top events contributing most to the Galactic plane analyses. The ranking for each of the analysis templates is given in the captions.

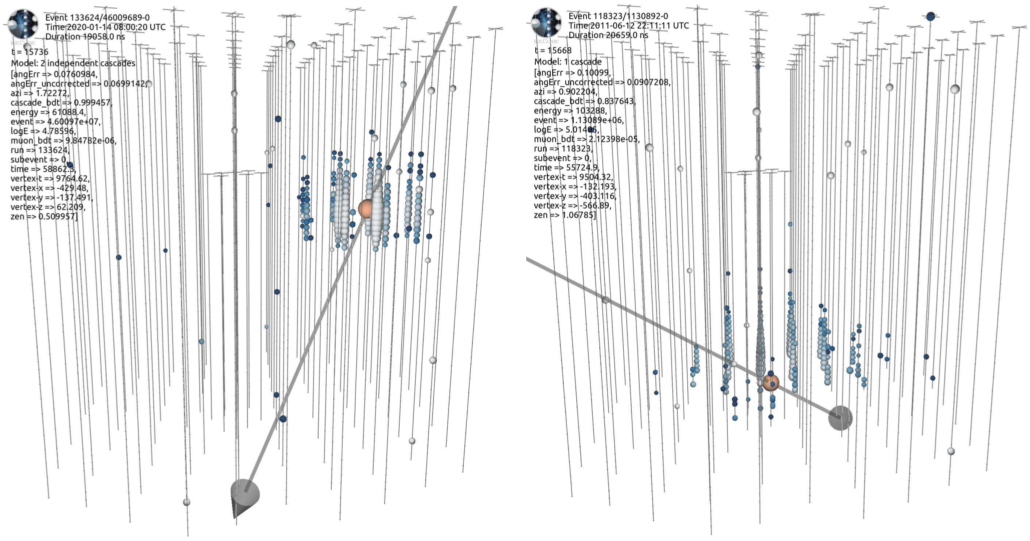


```
Event 126637/9028348-0
Time 2015-07-20 00:08:17 UTC
Duration 12305.0 ns
t = 12148
Model: 2 correlated cascades
[angErr => 0.0967981,
angErr_uncorrected => 0.0913792,
azi => 3.17801,
cascade_bdt => 0.856603,
energy => 12872.7,
event => 9.02835e+06,
logE => 4.10967,
muon_bdt => 0.000229562,
run => 126637,
subevent => 0,
time => 57223,
vertex-t => 9590.06,
vertex-x => -434.081,
vertex-y => -222.361,
vertex-z => -442.308,
zen => 0.42316]
```

```
Event 134273/68698698-0
Time 2020-07-11 04:55:58 UTC
Duration 20015.0 ns
t = 12047
Model: 2 independent cascades
[angErr => 0.100007,
angErr_uncorrected => 0.0920711,
azi => 3.17482,
cascade_bdt => 0.237694,
energy => 56211.7,
event => 6.86987e+07,
logE => 4.74983,
muon_bdt => 0.000213707,
run => 134273,
subevent => 0,
time => 59041.2,
vertex-t => 9835.47,
vertex-x => 233.674,
vertex-y => 444.981,
vertex-z => -256.025,
zen => 1.2568]
```

(a) π^0: #17.

(b) π^0: #18; KRA$_\gamma^5$: #10.



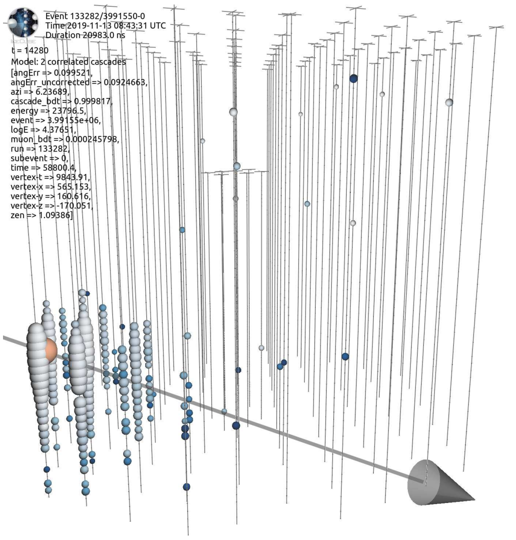
```
Event 133624/46009689-0
Time 2020-01-14 08:00:20 UTC
Duration 19058.0 ns
t = 15736
Model: 2 independent cascades
[angErr => 0.0760984,
angErr_uncorrected => 0.0699142,
azi => 1.72272,
cascade_bdt => 0.999457,
energy => 61088.4,
event => 4.60097e+07,
logE => 4.78596,
muon_bdt => 9.84782e-06,
run => 133624,
subevent => 0,
time => 58862.3,
vertex-t => 9764.62,
vertex-x => -429.48,
vertex-y => -137.491,
vertex-z => 62.209,
zen => 0.509957]
```

```
Event 118323/1130892-0
Time 2011-06-12 22:11:11 UTC
Duration 20659.0 ns
t = 15668
Model: 1 cascade
[angErr => 0.10099,
angErr_uncorrected => 0.0907208,
azi => 0.902204,
cascade_bdt => 0.837643,
energy => 103288,
event => 1.13089e+06,
logE => 5.01405,
muon_bdt => 2.12398e-05,
run => 118323,
subevent => 0,
time => 55724.9,
vertex-t => 9504.32,
vertex-x => -132.193,
vertex-y => -403.116,
vertex-z => -566.89,
zen => 1.06785]
```

(c) π^0: #19.

(d) π^0: #20; KRA$_\gamma^5$: #1; KRA$_\gamma^{50}$: #1.

Figure C.16: Events 17–20 of the top events contributing most to the Galactic plane analyses. The ranking for each of the analysis templates is given in the captions.

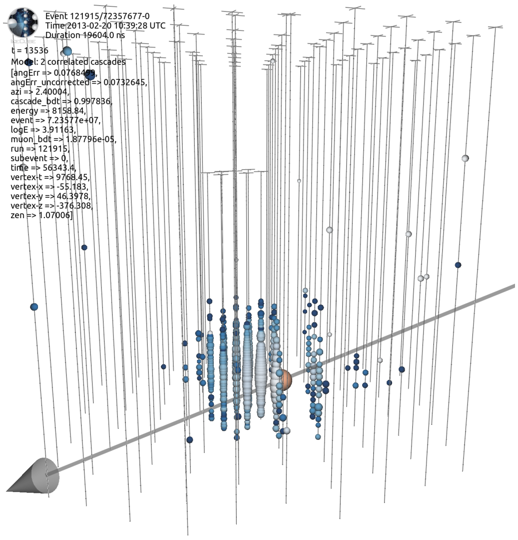


```
Event 133282/3991550-0
Time 2019-11-13 08:43:31 UTC
Duration 20983.0 ns
t = 14280
Model: 2 correlated cascades
[angErr => 0.099521,
angErr_uncorrected => 0.0924663,
azi => 6.23689,
cascade_bdt => 0.999817,
energy => 23796.5,
event => 3.99155e+06,
logE => 4.37651,
muon_bdt => 0.000245798,
run => 133282,
subevent => 0,
time => 58800.4,
vertex-t => 9843.91,
vertex-x => 565.153,
vertex-y => 160.616,
vertex-z => -170.051,
zen => 1.09386]
```

```
Event 130461/69540342-0
Time 2017-12-29 19:24:55 UTC
Duration 19726.0 ns
t = 12453
Model: 1 cascade
[angErr => 0.0830467,
angErr_uncorrected => 0.0786781,
azi => 3.75351,
cascade_bdt => 0.979482,
energy => 11193.1,
event => 6.95403e+07,
logE => 4.04895,
muon_bdt => 3.89664e-05,
run => 130461,
subevent => 0,
time => 58116.8,
vertex-t => 9797.22,
vertex-x => -227.479,
vertex-y => -185.133,
vertex-z => -353.215,
zen => 0.933624]
```

(a) KRA$_\gamma^5$: #7.

(b) KRA$_\gamma^5$: #9.



```
Event 121915/72357677-0
Time 2013-02-20 10:39:28 UTC
Duration 19604.0 ns
t = 13536
Model: 2 correlated cascades
[angErr => 0.0768499,
angErr_uncorrected => 0.0732645,
azi => 2.40004,
cascade_bdt => 0.997836,
energy => 8158.84,
event => 7.23577e+07,
logE => 3.91163,
muon_bdt => 1.87796e-05,
run => 121915,
subevent => 0,
time => 56343.4,
vertex-t => 9768.45,
vertex-x => -55.183,
vertex-y => 46.3978,
vertex-z => -376.308,
zen => 1.07006]
```

```
Event 125525/4818890-0
Time 2014-11-08 03:41:03 UTC
Duration 27545.0 ns
t = 19458
Model: 2 correlated cascades
[angErr => 0.180776,
angErr_uncorrected => 0.160347,
azi => 4.83743,
cascade_bdt => 0.936533,
energy => 133363,
event => 4.81889e+06,
logE => 5.12504,
muon_bdt => 0.000128664,
run => 125525,
subevent => 0,
time => 56969.2,
vertex-t => 16555.9,
vertex-x => 123.594,
vertex-y => -528.674,
vertex-z => -488.976,
zen => 0.943642]
```

(c) KRA$_\gamma^5$: #11.

(d) KRA$_\gamma^{50}$: #7.

Figure C.17: Events 21–24 of the top events contributing most to the Galactic plane analyses. The ranking for each of the analysis templates is given in the captions.

D DNNCascade Analysis

Supplemental material for the DNNCascade analysis presented in Part III is provided here.

D.1 Performed Analyses

Table D.1: The sources and coordinates [138, 139] of galactic sources used in the stacking catalogs analyses. [1]

Catalog	Source	α [°]	δ [°]
SNR	Vela Junior	133.0	-46.33
SNR	RX J1713.7-3946	258.36	-39.77
SNR	HESS J1614-518	243.56	-51.82
SNR	HESS J1457-593	223.7	-59.07
SNR	SNR G323.7-01.0	233.63	-57.2
SNR	HESS J1731-347	262.98	-34.71
SNR	Gamma Cygni	305.27	40.52
SNR	RCW 86	220.12	-62.65
SNR	HESS J1912+101	288.33	10.19
SNR	HESS J1745-303	266.3	-30.2
SNR	Cassiopeia A	350.85	58.81
SNR	CTB 37A	258.64	-38.54
PWN	Vela X	128.29	-45.19
PWN	Crab nebula	83.63	22.01
PWN	HESS J1708-443	257.0	-44.3
PWN	HESS J1825-137	276.55	-13.58
PWN	HESS J1632-478	248.01	-47.87
PWN	MSH 15-52	228.53	-59.16
PWN	HESS J1813-178	273.36	-17.86
PWN	HESS J1303-631	195.75	-63.2
PWN	HESS J1616-508	244.06	-50.91
PWN	HESS J1418-609	214.69	-60.98
PWN	HESS J1837-069	279.43	-6.93
PWN	HESS J1026-582	157.17	-58.29
UNID	MGRO J1908+06	286.91	6.32
UNID	Westerlund 1	251.5	-45.8
UNID	HESS J1702-420	255.68	-42.02
UNID	2HWC J1814-173	273.52	-17.31
UNID	HESS J1841-055	280.23	-5.55
UNID	2HWC J1819-150	274.83	-15.06
UNID	HESS J1804-216	271.12	-21.73
UNID	HESS J1809-193	272.63	-19.3
UNID	HESS J1843-033	280.75	-3.3
UNID	TeV J2032+4130	307.93	41.51
UNID	HESS J1708-410	257.10	-41.09
UNID	HESS J1857+026	284.30	2.67

D.2 Analysis Results

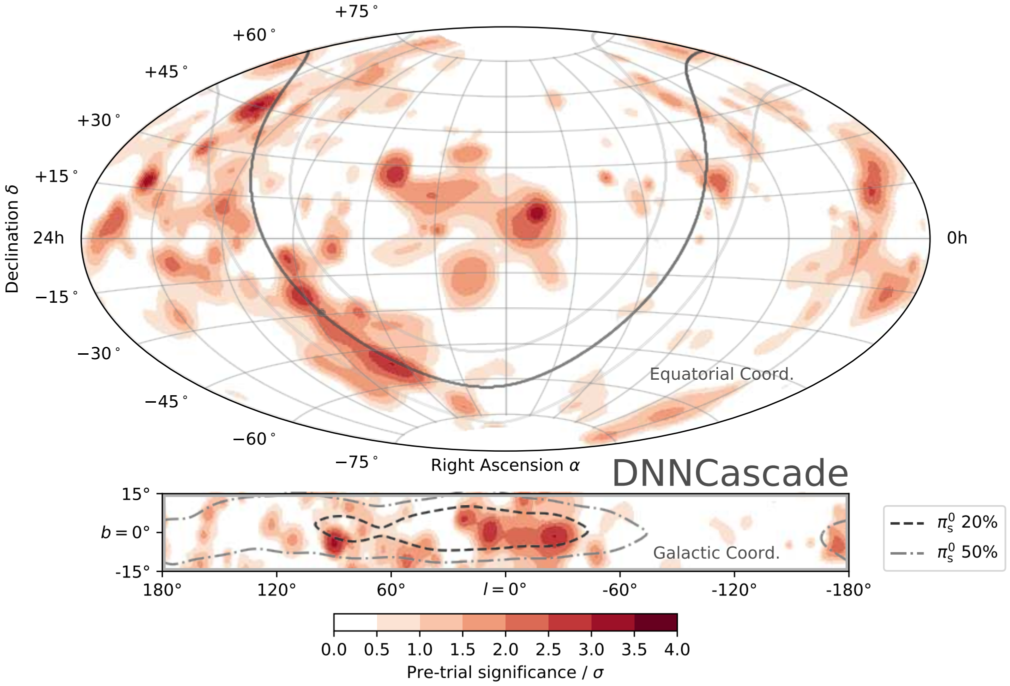


Figure D.1: The pre-trial significance is shown as a function of direction, in equatorial coordinates (J2000) and Aitoff projection, for the all-sky search. Same as Figure 15.8, but without overlaid sources.

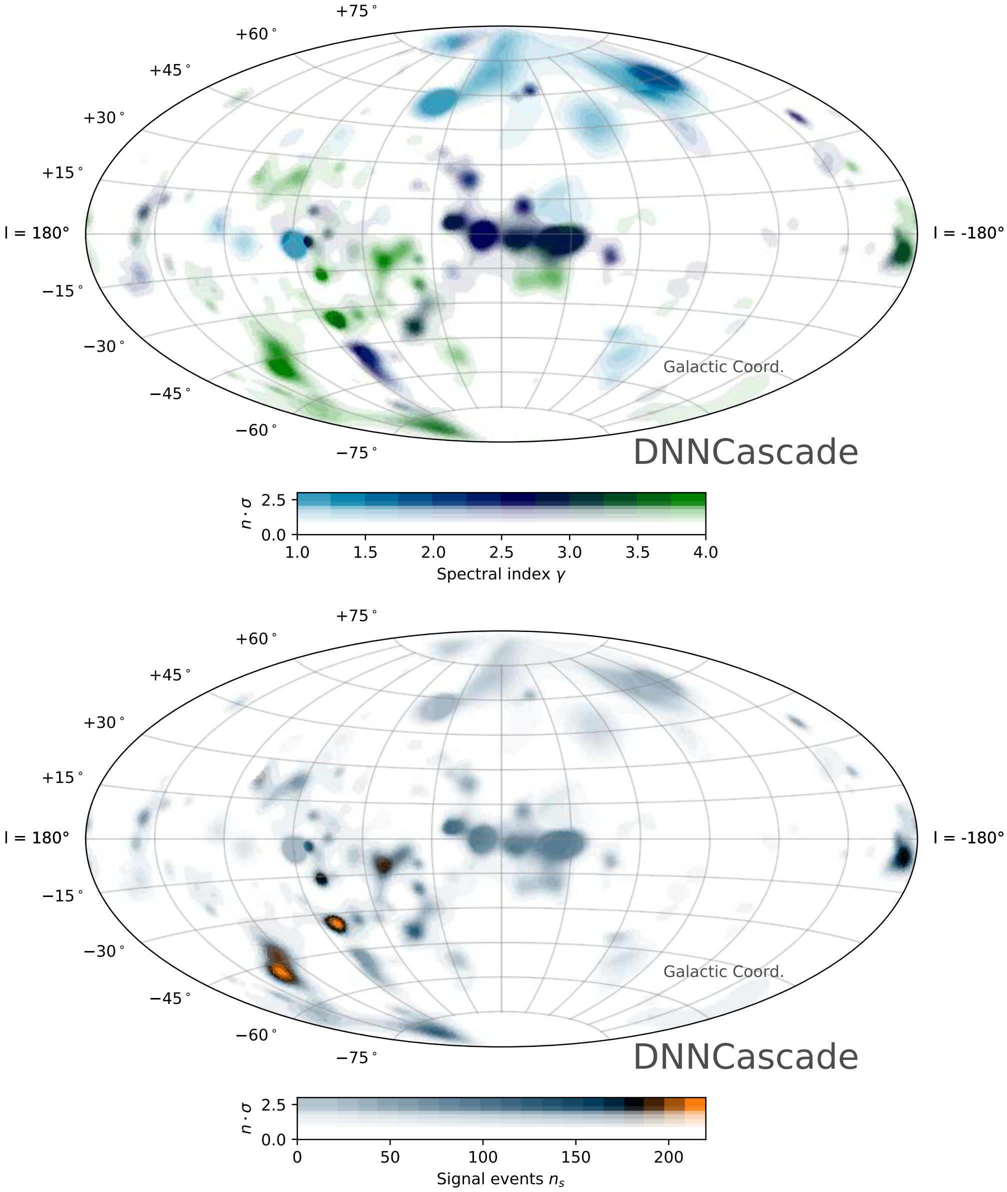


Figure D.2: The best-fit spectral index (top) and number of signal events (bottom) for the all-sky search are shown, weighted by significance as a function of direction. The pixel opacity is scaled by the pre-trial significance so that more opaque locations are more significant. Similar to Figures 14.4 and Figure 15.9, but utilizing galactic coordinates.

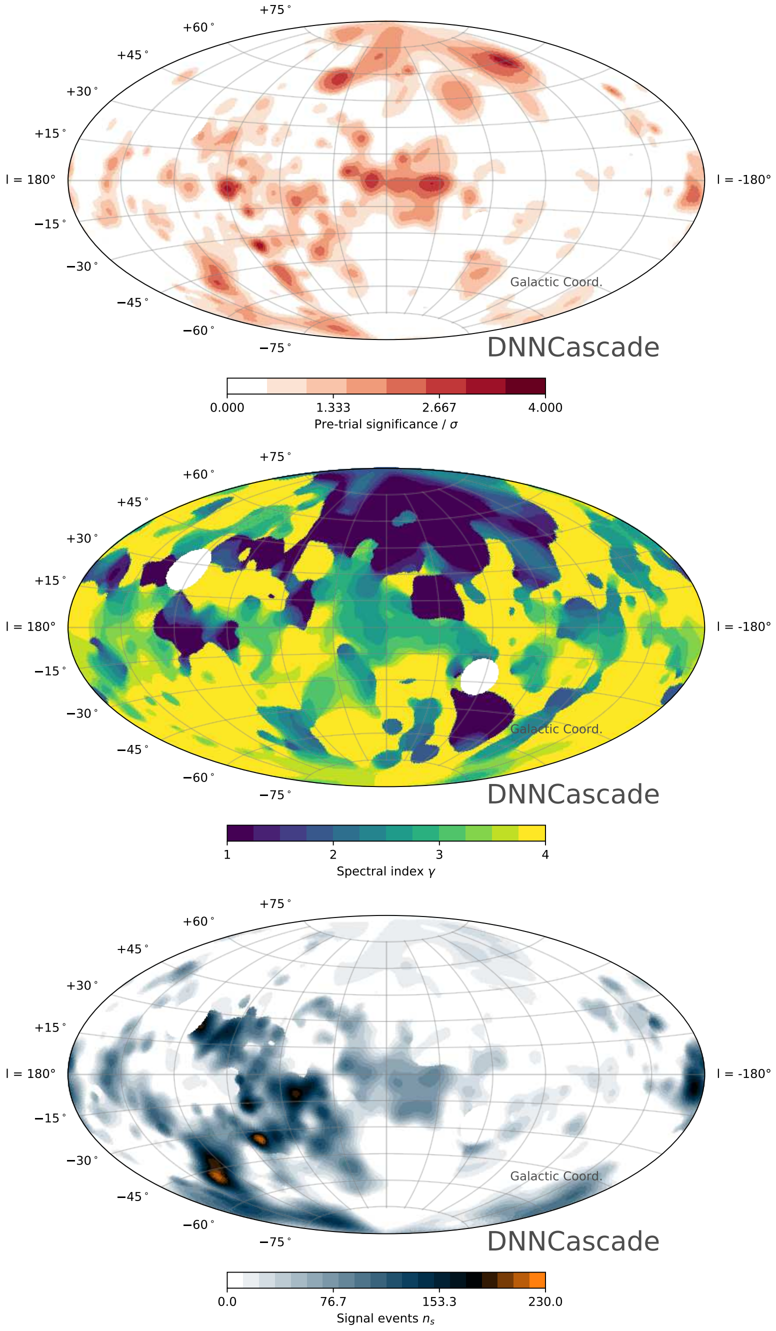
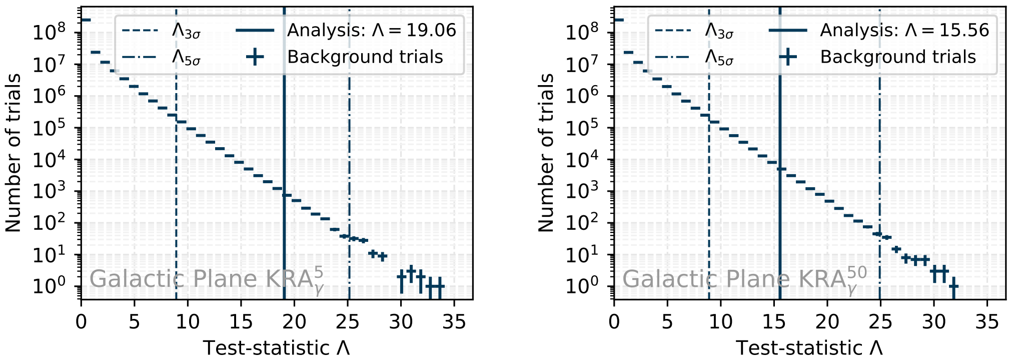
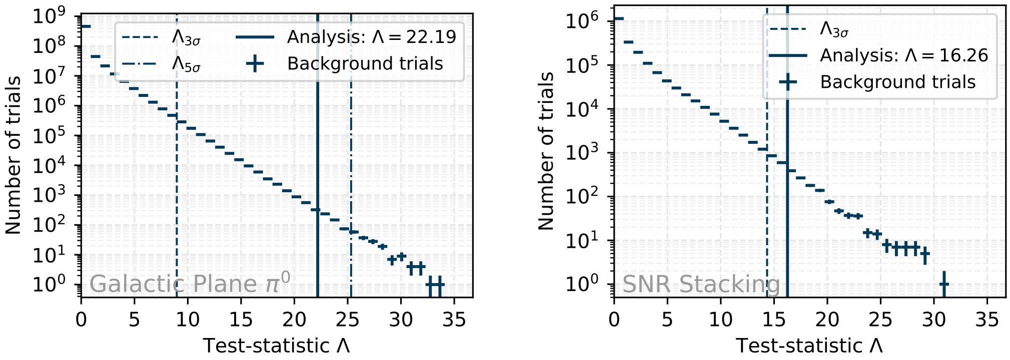


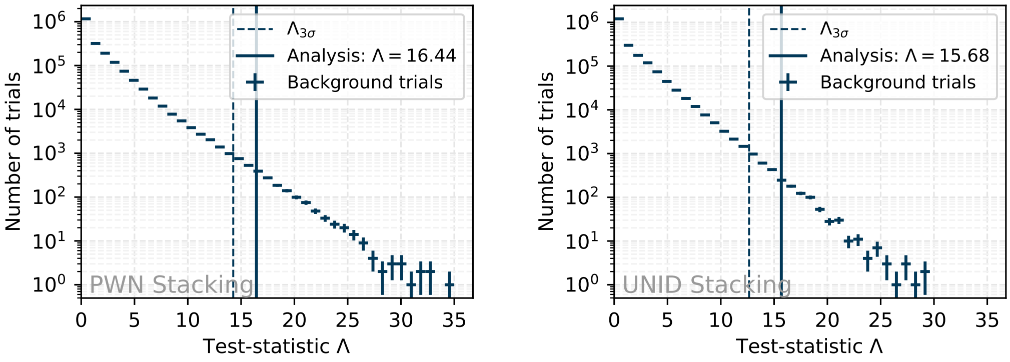
Figure D.3: The pre-trial significance (top), best-fit spectral index (middle), and best-fit number of signal events (bottom) are shown for the all-sky search in Galactic coordinates and Aitoff projection. Note that the best-fit parameters at non-significant positions in the sky depend on statistical fluctuations of the background.



(a) Galactic plane KRA$^5_\gamma$. **(b)** Galactic plane KRA$^{50}_\gamma$.



(c) Galactic plane π^0. **(d)** Stacking catalog SNR.



(e) Stacking catalog PWN. **(f)** Stacking catalog UNID.

Figure D.4: The background test-statistic distributions are shown for the Galactic plane template and for the catalog stacking searches. The required test-statistic value for a pre-trial significance of 3σ (5σ) is indicated by a dashed (dashed-dotted) line. Analysis results are shown as solid lines.

	Source Name	α [°]	δ [°]	n_s	γ	p-value	UL Φ_2	UL Φ_3
1	G343.1-2.3	257.0	-44.3	77.6	2.9	0.011	<3.34	<1.58
2	HESS J0835-455	128.3	-45.2	0.2	2.0	0.719	<0.59	<0.28
3	PKS 0426-380	67.2	-37.9	3.0	2.3	0.640	<0.81	<0.42
4	PKS 2155-304	329.7	-30.2	61.5	4.0	0.249	<1.52	<1.02
5	Mkn 421	166.1	38.2	0.0	–	–	<1.13	<0.32
6	PKS 0537-441	84.7	-44.1	1.1	1.8	0.576	<0.89	<0.45
7	PKS 0447-439	72.4	-43.8	0.7	2.1	0.698	<0.62	<0.31
8	BL Lac	330.7	42.3	7.6	1.6	0.340	<2.73	<0.77
9	PG 1553+113	238.9	11.2	0.0	–	–	<0.68	<0.35
10	TXS 0518+211	80.4	21.2	149.3	3.4	0.035	<4.24	<1.59
11	PKS 0235+164	39.7	16.6	0.0	–	–	<0.67	<0.31
12	PKS 1424+240	216.8	23.8	7.7	1.3	0.095	<3.55	<1.26
13	3C 66A	35.7	43.0	85.8	3.7	0.238	<3.44	<0.89
14	TXS 0506+056	77.4	5.7	0.0	–	–	<0.61	<0.39
15	AP Librae	229.4	-24.4	0.3	1.0	0.725	<0.61	<0.34
16	S5 0716+71	110.5	71.3	1.9	1.0	0.532	<3.94	<0.58
17	B2 1215+30	184.5	30.1	5.2	2.2	0.573	<1.80	<0.53
18	MH 2136-428	324.9	-42.6	0.0	–	–	<0.53	<0.27
19	PKS 2233-148	339.1	-14.6	25.8	4.0	0.584	<0.93	<0.65
20	Mkn 501	253.5	39.8	4.1	1.0	0.514	<2.09	<0.60
21	PMN J1603-4904	241.0	-49.1	84.5	3.1	0.010	<3.49	<1.59
22	S2 0109+22	18.0	22.8	147.8	4.0	0.077	<3.75	<1.31
23	PKS 0301-243	45.9	-24.1	44.3	3.6	0.405	<1.37	<0.76
24	4C +01.28	164.6	1.6	16.1	1.2	0.025	<3.27	<2.42
25	PKS 0700-661	105.1	-66.2	28.8	2.9	0.239	<1.72	<0.82
26	TXS 0628-240	97.7	-24.1	0.9	3.0	0.746	<0.55	<0.31
27	PKS 0823-223	126.5	-22.5	2.0	2.0	0.553	<1.01	<0.58
28	PKS 0735+17	114.5	17.7	63.6	4.0	0.376	<2.16	<0.91
29	PMN J1329-5608	202.3	-56.1	40.3	4.0	0.276	<1.48	<0.77
30	PMN J0531-4827	83.0	-48.5	0.0	–	–	<0.55	<0.26
31	MG1 J021114+1051	32.8	10.9	28.7	4.0	0.622	<1.22	<0.66
32	PKS 1440-389	221.0	-39.1	11.0	1.0	0.175	<2.07	<1.10
33	OT 081	267.9	9.6	35.2	2.8	0.356	<1.88	<1.10
34	OJ 287	133.7	20.1	28.6	2.7	0.473	<1.84	<0.74
35	PKS 1101-536	166.0	-54.0	0.0	–	–	<0.46	<0.23
36	TXS 0141+268	26.2	27.1	1.4	3.2	0.706	<1.17	<0.37
37	1H 1013+498	153.8	49.4	4.2	2.6	0.676	<1.65	<0.36
38	PKS 0048-09	12.7	-9.5	115.2	3.6	0.089	<2.35	<1.68
39	PMN J1650-5044	252.6	-50.8	88.1	2.9	0.002	<4.02	<1.93
40	PKS 0118-272	20.1	-27.0	44.8	3.9	0.378	<1.38	<0.82

	Source Name	α [°]	δ [°]	n_s	γ	p-value	UL Φ_2	UL Φ_3
41	1H 1914-194	289.4	-19.4	26.9	2.5	0.152	<2.30	<1.26
42	PKS 0332-403	53.6	-40.1	9.8	2.4	0.481	<1.13	<0.59
43	OJ 014	122.9	1.8	2.6	4.0	0.771	<0.57	<0.39
44	PMN J1918-4111	289.6	-41.2	54.6	3.7	0.256	<1.72	<0.89
45	PKS 1936-623	295.3	-62.2	26.6	4.0	0.420	<1.19	<0.55
46	1H 1720+117	261.3	11.9	6.8	2.7	0.712	<0.91	<0.49
47	PMN J1610-6649	242.7	-66.8	10.1	2.9	0.588	<0.84	<0.41
48	PMN J0334-3725	53.6	-37.4	10.7	2.4	0.456	<1.25	<0.65
49	TXS 1714-336	259.4	-33.7	67.7	2.9	0.028	<2.68	<1.54
50	PKS 2005-489	302.4	-48.8	0.0	–	–	<0.54	<0.26
51	PKS B1056-113	164.8	-11.6	19.6	2.1	0.096	<2.27	<1.66
52	RGB J2243+203	341.0	20.4	163.1	4.0	0.022	<4.53	<1.82
53	1ES 1959+650	300.0	65.1	150.0	4.0	0.071	<7.30	<1.14
54	S4 0814+42	124.6	42.4	15.4	4.0	0.568	<1.74	<0.49
55	KUV 00311-1938	8.4	-19.4	109.8	3.6	0.063	<2.91	<1.61
56	PMN J2250-2806	342.7	-28.1	59.1	4.0	0.244	<1.62	<1.05
57	1RXS J130421.2-435308	196.1	-43.9	7.1	2.3	0.473	<1.11	<0.57
58	PMN J0810-7530	122.8	-75.5	16.5	4.0	0.598	<0.79	<0.44
59	3C 454.3	343.5	16.2	106.2	4.0	0.188	<2.67	<1.25
60	PKS 1424-41	217.0	-42.1	10.1	1.0	0.280	<1.63	<0.83
61	3C 279	194.0	-5.8	10.1	1.0	0.120	<1.90	<1.63
62	CTA 102	338.2	11.7	129.6	4.0	0.102	<3.09	<1.60
63	PKS 1510-089	228.2	-9.1	4.8	1.0	0.212	<1.79	<1.34
64	PKS 0454-234	74.3	-23.4	0.0	–	–	<0.55	<0.30
65	PKS 1502+106	226.1	10.5	8.7	1.0	0.422	<1.79	<0.99
66	PKS 1830-211	278.4	-21.1	70.4	2.8	0.033	<3.07	<1.68
67	PKS 2326-502	352.3	-49.9	6.6	2.1	0.265	<1.66	<0.84
68	PKS 0727-11	112.6	-11.7	8.2	2.6	0.598	<0.83	<0.65
69	4C +21.35	186.2	21.4	11.4	1.5	0.162	<3.09	<1.15
70	PMN J2345-1555	356.3	-15.9	11.9	3.3	0.646	<0.78	<0.53
71	4C +01.02	17.2	1.6	74.8	3.8	0.336	<1.66	<1.18
72	PKS 2023-07	306.4	-7.6	122.8	3.5	0.060	<2.32	<1.76
73	Ton 599	179.9	29.2	14.5	2.4	0.405	<2.49	<0.74
74	4C +38.41	248.8	38.1	3.5	1.0	0.573	<1.84	<0.54
75	PKS 1244-255	191.7	-25.8	3.9	1.0	0.128	<2.04	<1.28
76	B2 1520+31	230.5	31.7	9.3	1.0	0.013	<6.59	<1.93
77	PKS 1124-186	171.8	-19.0	0.0	–	–	<0.56	<0.33
78	4C +28.07	39.5	28.8	0.0	–	–	<1.15	<0.36
79	PKS 1730-13	263.3	-13.1	51.2	2.7	0.128	<2.27	<1.59
80	PKS 0805-07	122.1	-7.9	2.4	2.3	0.676	<0.69	<0.53
81	PKS 0208-512	32.7	-51.0	4.9	1.8	0.144	<1.97	<1.03

	Source Name	α [°]	δ [°]	n_s	γ	p-value	UL Φ_2	UL Φ_3
82	PKS 0336-01	54.9	-1.8	11.0	4.0	0.729	<0.62	<0.47
83	PKS 0502+049	76.3	5.0	0.0	–	–	<0.61	<0.39
84	PKS 0402-362	61.0	-36.1	2.2	2.2	0.672	<0.71	<0.38
85	PMN J1802-3940	270.7	-39.7	76.7	3.1	0.043	<2.81	<1.39
86	PKS 1622-253	246.4	-25.5	36.6	2.6	0.089	<2.26	<1.36
87	PKS 2052-47	314.1	-47.2	0.0	–	–	<0.55	<0.26
88	3C 273	187.3	2.0	11.3	1.0	0.227	<2.01	<1.41
89	PKS 2142-75	326.8	-75.6	2.6	1.0	0.744	<0.49	<0.26
90	MG1 J123931+0443	189.9	4.7	9.3	1.0	0.217	<2.21	<1.42
91	MG2 J201534+3710	303.9	37.2	113.6	3.1	0.039	<5.29	<1.51
92	Galactic Centre	266.4	-29.0	58.7	2.7	0.022	<2.79	<1.67
93	PKS 2247-131	342.5	-12.8	7.3	4.0	0.709	<0.62	<0.46
94	NGC 1275	50.0	41.5	29.3	3.3	0.529	<2.01	<0.57
95	PKS 0521-36	80.7	-36.5	5.7	2.0	0.237	<1.75	<1.00
96	Cen A	201.0	-43.5	9.4	2.4	0.398	<1.30	<0.66
97	LMC	80.0	-68.8	31.5	4.0	0.327	<1.38	<0.71
98	SMC	14.5	-72.8	43.1	1.0	0.174	<1.77	<0.94
99	NGC 4945	196.4	-49.5	7.3	2.4	0.425	<1.28	<0.63
100	NGC 253	11.9	-25.3	51.4	3.8	0.328	<1.43	<0.88
101	NGC 1068	40.7	-0.0	87.1	4.0	0.252	<1.88	<1.36
102	M 82	148.9	69.7	22.9	2.9	0.564	<3.44	<0.53
103	Arp 220	233.7	23.5	15.2	1.0	0.009	<5.18	<2.03
104	M 31	10.8	41.2	18.2	2.9	0.487	<2.21	<0.62
105	NGC 3424	162.9	32.9	0.0	–	–	<1.19	<0.34
106	IC 678	168.6	6.6	19.2	1.8	0.005	<4.55	<2.85
107	NGC 5380	209.3	37.5	10.6	1.0	0.249	<3.30	<0.93
108	Arp 299	172.1	58.5	0.0	–	–	<1.25	<0.23
109	NGC 2146	94.5	78.3	4.2	1.0	0.445	<4.71	<0.73

Table D.2: Sources in the source-list analysis are provided together with pre-trial p-values. Per-flavor 90% flux upper limits are shown as $E^2 \frac{dN}{dE}$ at 100 TeV in units of 10^{-12} TeV cm^{-2} s^{-1} for sources emitting following an E^{-2} (UL Φ_2) and E^{-3} (UL Φ_3) spectrum. [1]

D.3 Interpretation of Results

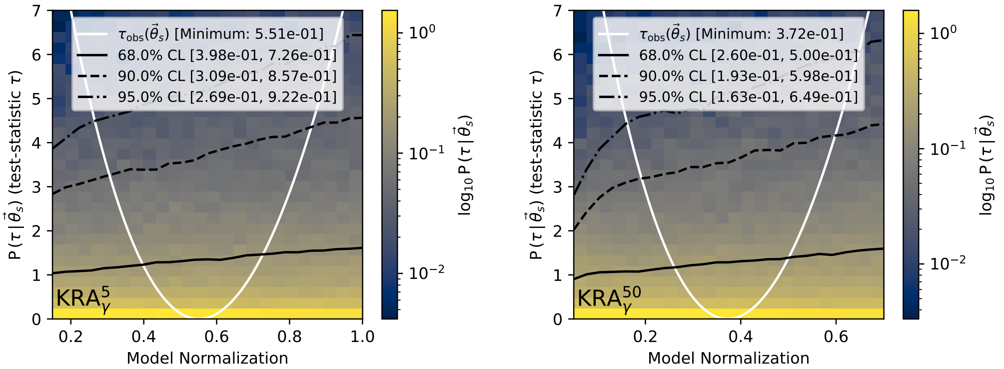
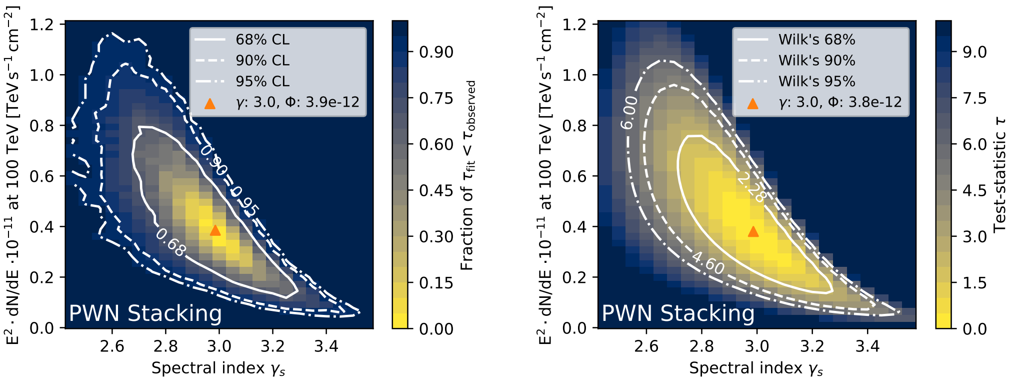
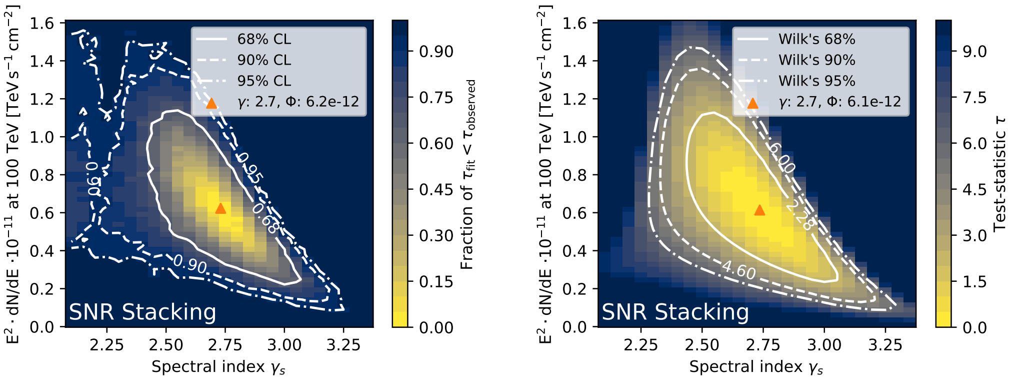


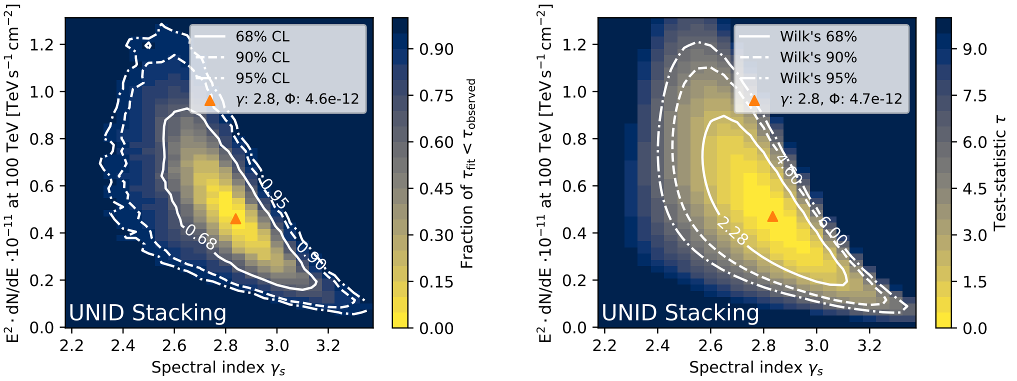
Figure D.5: The confidence interval construction for the flux measurement of the KRA_γ models is illustrated. The distribution of test-statistic values (τ) from Eq. (15.1) is shown (y-axis and color map) for each true injected flux normalization (x-axis). Critical values are shown as black lines for the 68%, 90%, and 95% confidence levels. The observed test-statistic ($\tau_{\rm obs}$), evaluated on experimental data, is provided in white.



(a) PWN. (b) PWN using Wilk's theorem.

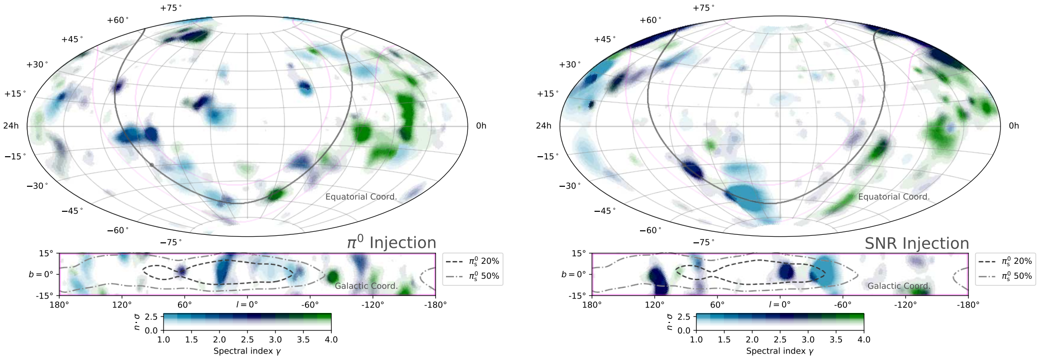


(c) SNR. (d) SNR using Wilk's theorem.

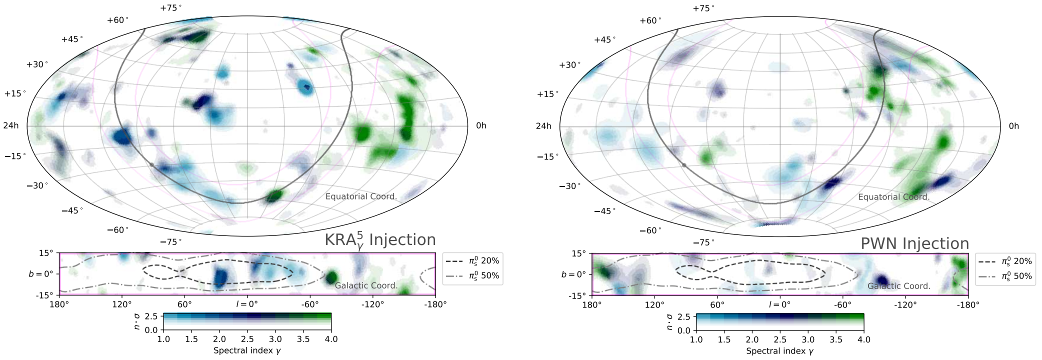


(e) UNID. (f) UNID using Wilk's theorem.

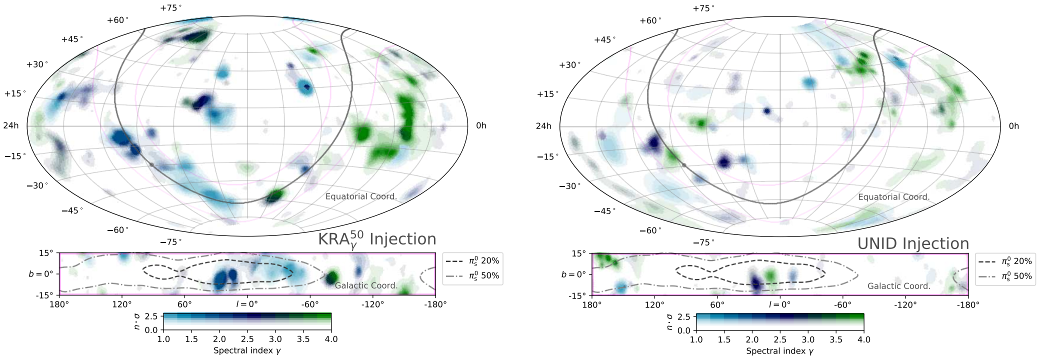
Figure D.6: The confidence interval construction for the flux measurement of the catalog stacking analyses are illustrated. Contours corresponding to 68%, 90%, and 95% confidence level are drawn in white based on trials (left column) and according to Wilk's theorem (right column).



(a) π^0. **(b)** SNR catalog.

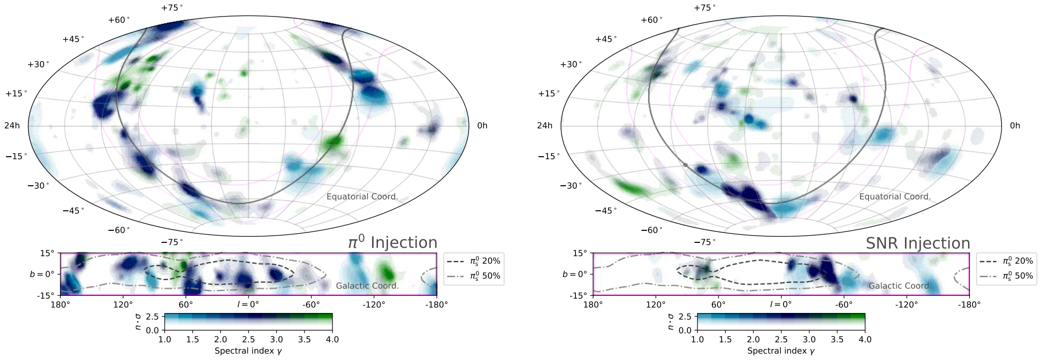


(c) KRA^5_γ. **(d)** PWN catalog.

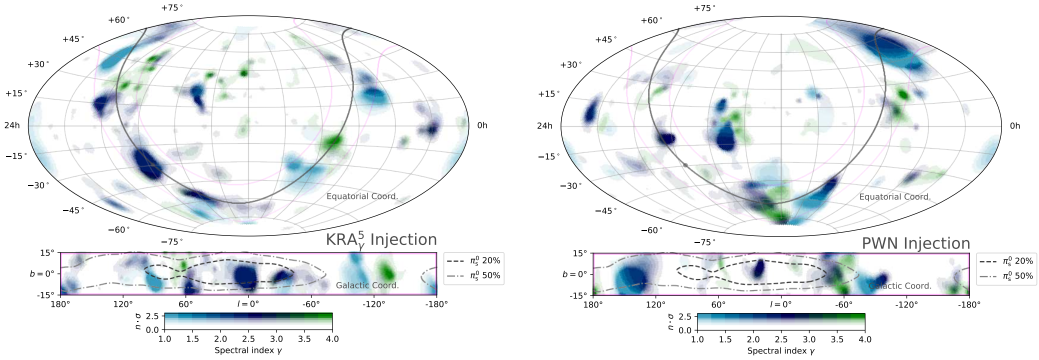


(e) KRA^{50}_γ. **(f)** UNID catalog.

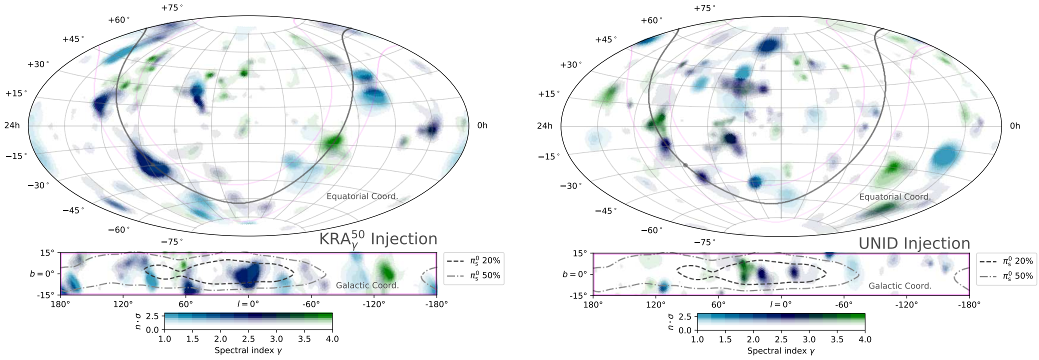
Figure D.7: The all-sky search result is shown for simulated injection trials where the best-fit flux is injected into background scrambles for each of the tested Galactic plane models and stacking catalogs. The Galactic plane template searches utilize the same underlying background scrambles in this test.



(a) π^0. **(b)** SNR catalog.

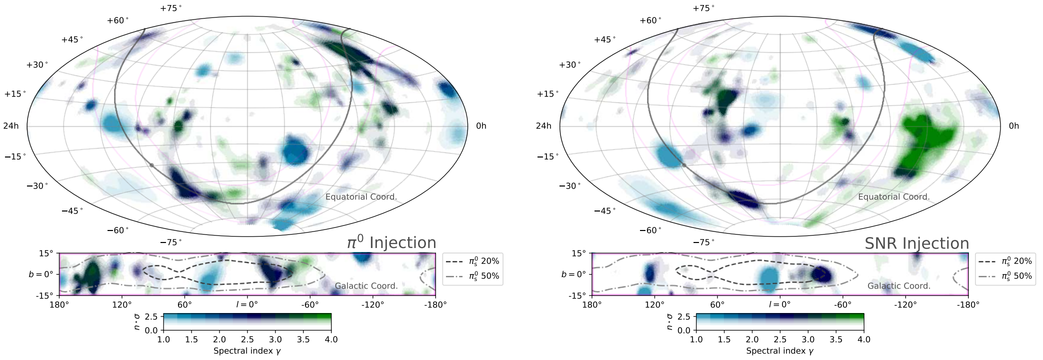


(c) KRA^5_γ. **(d)** PWN catalog.

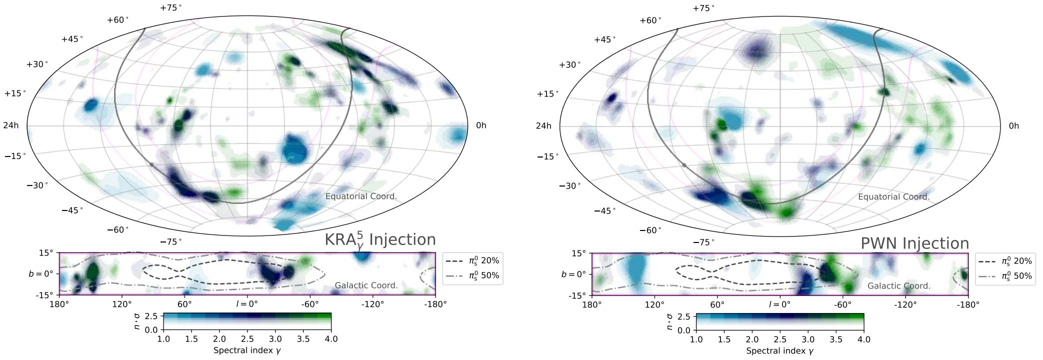


(e) KRA^{50}_γ. **(f)** UNID catalog.

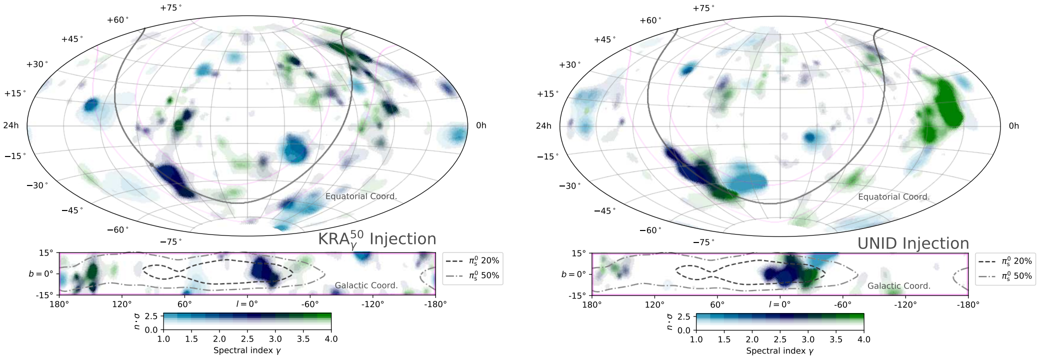
Figure D.8: Same as Figure D.7, but for a different random seed.



(a) π^0. **(b)** SNR catalog.



(c) KRA^5_γ. **(d)** PWN catalog.



(e) KRA^{50}_γ. **(f)** UNID catalog.

Figure D.9: Same as Figure D.7 and D.8, but for a different random seed.

Acknowledgments

I am deeply indebted to my advisor Prof. Dr. Dr. Wolfgang Rhode, who has supported me throughout the years, enabling many opportunities and never having turned down a request. Thank you for the trust placed in me and the freedom given to pursue my own path.

I would also like to thank the entire working group, and in particular my office colleagues. It has been a fantastic time with a great deal of laughter in- and outside of work. A special thanks goes out to Andrea Teichmann, who not only keeps the working group functioning, but also the people within.

Thank you to everyone that helped to proof-read this document, and to Prof. Dr. Julia Tjus who kindly agreed to act as second advisor to this dissertation.

Looking back, there have been a great deal of people, whom I have learned enormously from. Many of these great individuals are part of the IceCube family. In particular, I would like to thank Mathis Börner and Mike Richman, who have been extremely influential to me. Thank you, Mike, for that message many years ago, and the shared excitment as we realized we are on to something. My thanks also go out to Steve Sclafani and Prof. Naoko Kurahashi Neilson. It has been an honor to have worked together with you on this great discovery.

I would like to express my deepest appreciation to my friends and family, who have stood by me on this journey; to my parents, for their unwavering support and for enabling me to pursue my dreams; to my children, who have driven me nuts, but have always managed to warm my heart in the most challenging of times; and to my wonderful wife, for the love and patience in these stressful times. Words cannot express my gratitude to you, Marissa.

Last but not least, I am grateful for the financial support received from the following funding agencies and organizations, without which this endeavor would not have been possible: the collaborative research centers 876 and 1491 (DFG) as well as the BMBF, the Wilhelm and Else Heraeus Foundation and the DAAD for enabling the attendance to conferences, and Fulbright for funding a remarkable research stay with Spencer Klein at Lawrence Berkeley National Laboratory.

Eidesstattliche Versicherung

Ich versichere hiermit an Eides statt, dass ich die vorliegende Abschlussarbeit mit dem Titel „Observation of High-Energy Neutrinos from the Milky Way“ selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt, sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Ort, Datum

Unterschrift

Belehrung

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50 000 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden (§ 63 Abs. 5 Hochschulgesetz –HG–).

Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft.

Die Technische Universität Dortmund wird ggf. elektronische Vergleichswerkzeuge (wie z. B. die Software „turnitin“) zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen.

Ort, Datum

Unterschrift