

# Statistical analysis on the origin of the IceCube neutrino events

Bachelorarbeit am Lehrstuhl für Astronomie  
von  
Juliane Wiehl



Fakultät für Physik und Astronomie  
Julius-Maximilians-Universität Würzburg

Würzburg, den 4. Juli 2016

*Verantwortlicher Hochschullehrer:*

Prof. Dr. Karl Mannheim



# Zusammenfassung

Im Rahmen dieser Arbeit wurde ein Algorithmus zur Untersuchung der Korrelation von detektierten IceCube Neutrino Events mit astronomischen Quellen entwickelt. Der IceCube Detektor am Südpol ist der derzeit größte Neutrinodetektor mit einem effektiven Volumen von etwa  $1 \text{ km}^3$ . Die Liste an detektierten Neutrino Events, die 2014 veröffentlicht wurde, enthält 37 hochenergetische Neutrino Events, die gegen mögliche Quellen getestet werden können. Ein Kandidat sind aktive Galaxienkerne, die kompakte innere Region von Galaxien, die eine außergewöhnlich hohe Leuchtkraft über das gesamte elektromagnetische Spektrum aufweisen. Besonders eine Untergruppe namens Blazare sind für die Emission von Neutrinos geeignet.

Ein Analyse der Korrelation möglicher Quellen wurde im Jahr 2014 schon von Glüsenkamp et al. durchgeführt. Davon ausgehend wurde der Algorithmus, der sich die sog. *unbinned maximum likelihood* Methode zu Nutze macht, aufbereitet und erweitert. Bei der Entwicklung wurden zur Berechnung zum Einen die Koordinaten der Events und der Quellen, zum Anderen die rekonstruierte Energie eines Events und der *Fermi*-Fluss der Quellen berücksichtigt.

Im Anschluss wurden mithilfe des Algorithmus die IceCube HESE-Events gegen die BL Lac, FSRQ und Uncertain Quellen aus dem *5th edition of the Roma-BZCAT* Katalog mit 1425, 1909 beziehungsweise 227 Elementen getestet. Der Parameter  $n_s$  entspricht der höchsten Übereinstimmung zwischen Neutrino Events und den Quellen und beträgt 17 für Uncertain, 6 für BL Lac und 7.5 für FSRQ. Das bedeutet, bei den Uncertain Quellen passt die Anzahl von 17 aus den 37 detektierten Neutrino Events am Besten zum untersuchten Emissionsmodell.

Der Algorithmus, der im Rahmen dieser Arbeit entwickelt wurde, bietet eine gute Grundlage für Erweiterungen, wie z.B. die Berücksichtigung der Detektionszeit eines Events. Außerdem wäre ein größerer Datensatz an Events hilfreich, um eine bessere statistische Aussage über die Herkunft der Neutrinos treffen zu können.

---

# Abstract

In this thesis an algorithm for analysing the correlation between IceCube neutrino events and astronomical sources was developed. IceCube is located at the South Pole and currently the neutrino detector with the largest integrated volume of about  $1 \text{ km}^3$ . The sample of detected events published in 2014 contains 37 high-energy neutrino events that can be tested against possible sources. A presumable candidate is given by active galactic nuclei, the compact inner regions of galaxies that show an exceptionally high luminosity throughout the entire electromagnetic spectrum. In particular, a subgroup named blazars is eligible for neutrino production.

A correlation test with possible sources has been performed previously by Glüsenskamp et al. in 2014. Based on that, the algorithm using the unbinned maximum likelihood method was re-evaluated and enhanced. Over the development process, the algorithm considers the event and source coordinates, the reconstructed energy of an event and the *Fermi*-flux of the sources.

Consecutively, the algorithm was used to test the IceCube HESE events against the BL Lac, FSRQ and Uncertain sample from the 5th edition of the Roma-BZCAT catalogue with 1425, 1909 and 227 members respectively. The parameter  $n_s$  corresponding to the highest agreement between neutrino events and the source sample is 17 for the Uncertain sample, 6 for BL Lac and 7.5 for FSRQ respectively. This means, out of the total number of 37 events tested, in the Uncertain sample a count of 17 events is most likely to match the emission model.

The algorithm developed within the framework of this thesis provides a good starting point for further extensions, for instance considering the event observation time. In order to improve the statement about the origin of the IceCube neutrino event a larger set of events would be of benefit for the statistical analysis.



# Contents

<b>1</b>	<b>Motivation</b>	<b>3</b>
<b>2</b>	<b>Active galactic nuclei and neutrinos</b>	<b>5</b>
2.1	Active galactic nuclei . . . . .	5
2.1.1	Structure and emission . . . . .	5
2.1.2	Classification . . . . .	6
2.1.3	Unification model . . . . .	8
2.2	Neutrino production and detection . . . . .	9
<b>3</b>	<b>The IceCube detector</b>	<b>13</b>
3.1	Construction . . . . .	13
3.2	Detection principle . . . . .	14
3.3	Effective area . . . . .	14
3.4	First results . . . . .	16
<b>4</b>	<b>Development and application of the algorithm</b>	<b>19</b>
4.1	Unbinned maximum likelihood method . . . . .	19
4.2	Developing the algorithm . . . . .	21
4.2.1	Model of the algorithm . . . . .	21
4.2.2	Signal PDF $S_i$ . . . . .	21
4.2.3	Background PDF $B_i$ . . . . .	24
4.3	Testing . . . . .	27
4.4	Error consideration . . . . .	29
<b>5</b>	<b>Conclusion and outlook</b>	<b>31</b>
	<b>Appendix</b>	<b>33</b>
	<b>Bibliography</b>	<b>39</b>
	<b>Acknowledgements</b>	<b>41</b>



# 1 Motivation

Already long ago, people started to look up to the nocturnal sky and were fascinated by the stars and whatever else appeared. However, the range of observation was limited by the capabilities of the human eye. As science progressed, so did the observation techniques and the gain in knowledge about the universe.

An important milestone in astronomy is the development of detectors and telescopes to measure radiation beyond the optical waveband. For measurements using particles other than photons, one has to bear in mind the deflection and absorption mechanisms occurring all along their trajectory. Here, considering neutrinos, their lack of charge comes in handy. As the deflection has no influence whatsoever, neutrinos provide a unique opportunity to directly indicate sources. But the apparent advantage turns into an inconvenience when it comes to detection. Neutrino detectors require a much more sophisticated approach than conventional photon detectors such as CCDs. Popular representatives might be Super-Kamiokande built in 1996 in Japan, the Borexino experiment situated near the Gran Sasso mountain in Italy or the ANTARES detector built in 2008 off the Mediterranean coast of France. The presently largest neutrino observatory is named IceCube and is located at the South Pole.

Considering astrophysical neutrinos, promising source candidates are blazars, a subtype of radio-loud active galactic nuclei. Since they can be observed along the entire electromagnetic spectrum, their coordinates and other properties are known, allowing for an attempt to match detected neutrinos with the source candidates. This has already been done among others by Bernhard 2014 [3], Ström 2015 [23] and Glüsenskamp et al. 2015 [7].

In particular, Glüsenskamp et al. provide an approach for correlation analysis using an unbinned maximum likelihood method. In this work, this algorithm is being re-evaluated and improved in terms of the assumptions made over the development in order to perform an enhanced version of the analysis relevant to the model. Desirably, it is possible to make a more precise statement about the suitability of presumed sources for the neutrino emission model as a result.



## 2 Active galactic nuclei and neutrinos

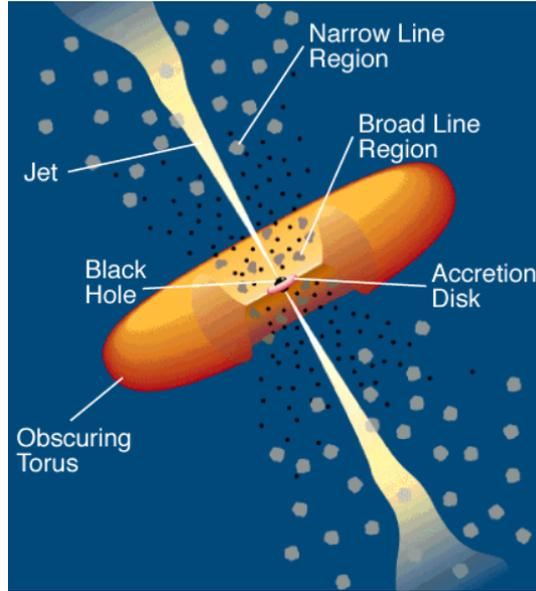
As described already in chapter 1, the astronomical objects that are of relevance here are *Active Galactic Nuclei* (AGN) in respect of neutrino emission and detection. The following section about AGN is based on Beckmann 2012 [2] and Krolik 1998 [16]. If no additional references are given, the section concerning neutrino production is built on Povh 1999 [19].

### 2.1 Active galactic nuclei

The expression *Active Galactic Nucleus* (AGN) refers to the inner region of a galaxy emitting enormous amounts of radiation compared to other galaxies. AGN are among the most luminous astronomical objects and thereby especially interesting for astronomers. It is necessary to distinguish an active galaxy (i.e. a galaxy containing an AGN in its center) from an ordinary galaxy. Active galaxies show a luminosity that can be several times higher compared to common galaxies. Whereas the size of the emitting region is comparable to that of our solar system and the responsible radiation is of non-stellar origin. Due to their size, AGN were believed to be stellar objects for quite some time because the host galaxy could not be resolved properly.

#### 2.1.1 Structure and emission

The schematic structure of an AGN can be seen in Fig. 2.1. The central engine consists of a massive black hole with a mass of up to  $10^{10}M_{\odot}$ . There is an accretion disk made of dust, gas and ionized gas respectively surrounding the black hole. Matter falling onto the accretion disk is moving in an orbit around the central body. Due to friction and turbulences, differential velocities of the orbiting matter occur. This matter then is transported inwards while angular momentum is carried outwards. In this process, potential energy in the gravitational field of the black hole is converted into thermal and kinetic energy, raising the temperature of the disk. The resulting radiation from the disk is mostly located in the optical or UV waveband. The spectrum of the accretion disk can be described by a black body spectrum where the temperature is dependent on the distance to the disk center. Inverse-Compton scattering can also cause X-ray emission from the disk. In a distance of about 0.1 to 1 pc from the central body,



**Figure 2.1:** Schematic visualization of an AGN showing the black hole as center body surrounded by an accretion disk. The sketch also features the two emission regions BLR and NLR as well as the obscuring dust torus and the relativistic outflows that are limited to radio-loud AGN. Credit: Urry and Padovani 1995 [24]

the *broad line region* (BLR) can be found above/below the accretion disk. Consisting of heavy, ionized clouds moving at velocities of 1000 to 10 000  $\frac{\text{km}}{\text{s}}$  this region causes broad line emission due to the turbulence of its components. Furthermore, a dust torus surrounding the accretion disk is located in a distance of 1 to 10 pc from the central black hole. In contrast to the BLR, the *narrow line region* (NLR) is found at a distance of about 100 pc from the accretion disk. Supposedly, its clouds are colder, less dense and moving much slower than in the BLR, causing the emission to be of narrow line type. Perpendicular to the accretion disk, relativistic outflows called jets can emerge. A jet consists of particles accelerated to almost the speed of light and magnetic field lines that are spirally arranged. Multi-wavelength observations show emission of AGN properties throughout the entire electromagnetic spectrum, from radio up to gamma rays with energies of several PeV.

### 2.1.2 Classification

Multi wavelength observations yield several types of AGN which are characterized by different emission properties. As a first classification, AGN can be divided into two classes depending on their radio emission properties. AGN that show an occurrence of jets are classified as radio-loud while representatives without this feature are known as radio-quiet. In the following a brief description of the most common types is given.

## Seyfert galaxies

Most of the observed AGN are radio-quiet, whereof two representatives are named *Seyfert 1* and *Seyfert 2* galaxies after they were discovered by Carl Seyfert in 1943 [22]. Observations showed that both types are lacking relativistic outflows in contrast to their radio-loud relatives with the exception of some Narrow Line Seyfert 1 galaxies (Schulz et al. 2016 [20]). In general, they are (primarily spiral) galaxies with a highly luminous core. The difference between Seyfert 1 and Seyfert 2 is mainly found in the spectrum. Seyfert 1 show strong broad emission lines along with narrow lines, whereas the other spectrum has but narrow lines.

## QSOs/Quasars

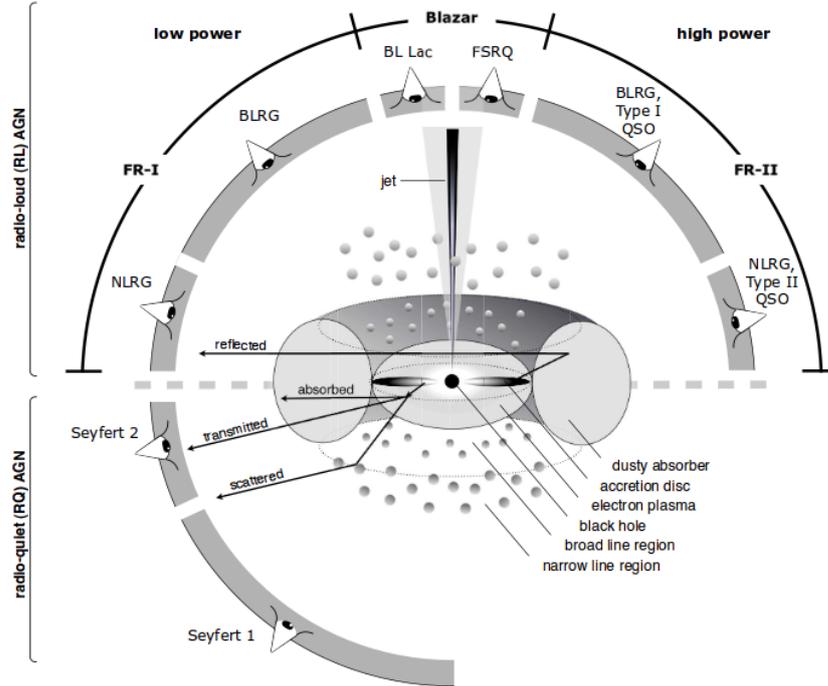
Another radio-quiet type are *quasi-stellar objects* (QSOs). Historically, the name was formed because the core completely outshined the host galaxy in the observed waveband and therefore these objects appeared point-like, resembling stars. Basically, QSOs show broad emission lines in the optical waveband like Seyfert 1 galaxies, but with higher luminosity. A radio-quiet relative to QSOs is called Quasar which stands for quasi-stellar radio source. In contrast to QSOs, quasars feature jets, while both types are commonly spotted at high redshifts.

## Blazars

The objects called *Blazars* are characterized by their variability over short time periods ranging from just hours up to months and polarized radiation. Basically, blazars show emission throughout the entire spectrum. However, depending on the spectrum they can be further divided into *BL Lac* and *FSRQ* objects, where the former is short for *BL Lacertae object*, named after its prototype BL Lacertae. The latter is an abbreviation of *Flat Spectrum Radio Quasar*. BL Lac have a spectrum presenting strongly polarized, weak narrow lines and a non-thermal continuum, whereas FSRQ show strong, broad lines and a flat radio spectrum. The variability of the radiation implies that the emitting region is compact. This can be explained by a relativistic outflow emitting along the line of sight. The spectrum energy distribution features two significant peaks, where the low energy peak is induced by synchrotron radiation. For the high energy peak, inverse Compton scattering is suspected to be the cause. The precise underlying model has not been substantiated yet, however, another attempt to explain the distribution is given by hadronic models.

## Radio galaxies

Another representative of the radio-loud active galaxies are *radio galaxies*. They manifest a high emission in the radio band. A special feature is given by lobes. The relativistically accelerated particles within a jet hit the interstellar medium and cause a rise in



**Figure 2.2:** Visualization of the unification model: Different types of AGN can be reduced to one basic type. The variations in observational properties occur due to differing viewing angles. Credit: Beckmann 2012 [2]

temperature of interstellar gas. This results in regions of excessive synchrotron emission which are very strong in the radio waveband. Further distinction according to Fanaroff and Riley 1974 [5] yield the subclasses *FR-I* and *FR-II*, first of which is characterized by a decrease in luminosity along the jets when distance to the core increases. In contrary, the latter shows a rather weakly emitting core, whereas the jets produce more and more radiation with increasing distance.

### 2.1.3 Unification model

In 1995, Urry and Padovani [24] suggested that the observational differences regarding radio-loud AGN types are based on only one structure type dependent on the viewing angle and the luminosity. The dust torus plays an important role here since it is responsible for obscuring certain emitting regions (namely the accretion disc or the BLR) towards an observer on earth. Another factor is whether an observer is located within the solid angle of the jet. Therefore, the appearance of galaxies of type Seyfert 2 is conditioned by the BLR being shadowed by the torus, resulting in a mere narrow line spectrum. Contrarily, for the Seyfert 2 type both the BLR and NLR are observable

since no dust torus is located along the line of sight, manifesting in the spectrum. A similar consideration can be done for radio-loud objects. Here, jets provide an additional feature, which results in a high variability of the source when observed along the jet axis as presented in blazar spectra.

## 2.2 Neutrino production and detection

First postulated by Pauli in 1930 trying to explain the plausability of the  $\beta$ -decay, neutrinos have a spin of  $\frac{1}{2}$  and are therefore considered fermions. They can be divided into three subclasses according to their flavor, that is, electron, muon and tau neutrinos ( $\nu_e, \nu_\mu, \nu_\tau$ ) including their corresponding antineutrinos ( $\bar{\nu}_e, \bar{\nu}_\mu, \bar{\nu}_\tau$ ). In addition, neutrinos show no charge and are very light-weight. Their only possible interaction is through gravitation or weak interaction (e.g. the  $\beta$ -decay). Regarding weak interaction, neutrinos are able to couple to  $W^\pm$  and  $Z^0$  bosons, depending on the change of charge during the interaction. These interactions can be distinguished as charged-current (CC) or neutral-current (NC) [25]. In particular, aside from beta-decays neutrinos can also be produced in other processes. Considering blazars, there are two different models being able to explain the measured radiation, hadronic and leptonic models. The first one is in line with cosmic radiation. Given a high energetic particle interacting with matter, neutrinos can be obtained from either nucleon interactions as

$$\begin{aligned} p + p &\rightarrow \pi^0 + 2p \\ p + p &\rightarrow \pi^+ + n + p \end{aligned} \quad (2.1)$$

or photonuclear interactions according to

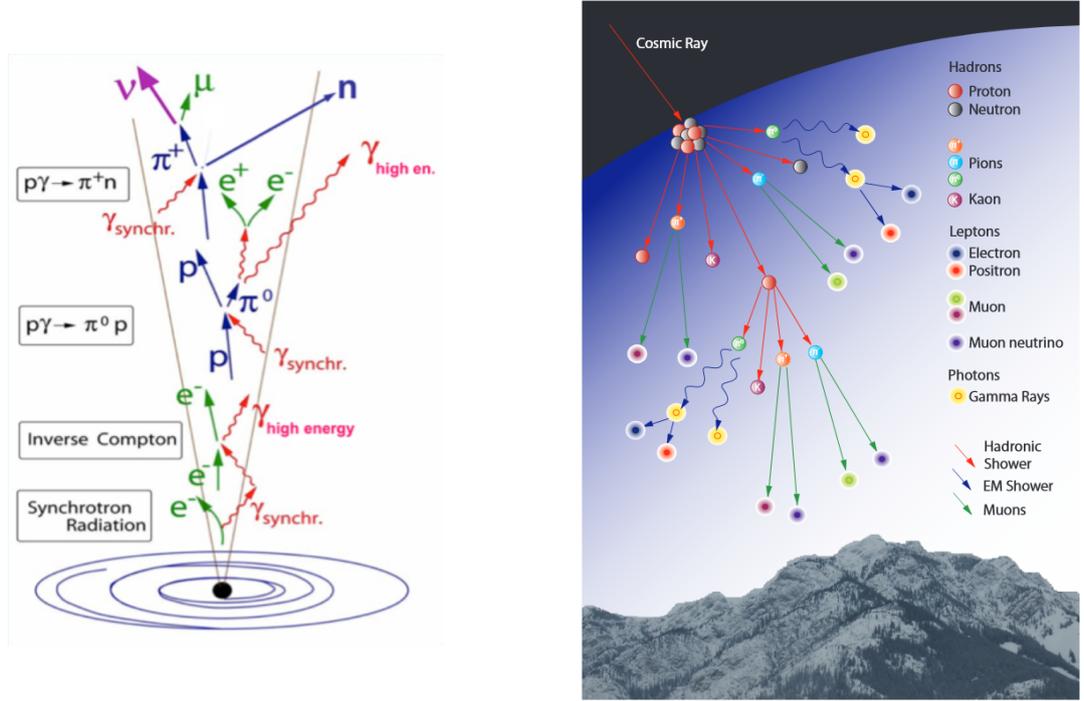
$$\begin{aligned} p + \gamma &\rightarrow \pi^+ + n \\ p + \gamma &\rightarrow \pi^0 + p \end{aligned} \quad (2.2)$$

producing pions in the process either way, which in turn are able to create neutrinos through their interaction:

$$\begin{aligned} \pi^0 &\rightarrow \gamma + \gamma \\ \pi^+ &\rightarrow \mu^+ + \nu_\mu \\ (\pi^- &\rightarrow \mu^- + \bar{\nu}_\mu). \end{aligned} \quad (2.3)$$

The muon produced in the second interaction can itself decay further as

$$\begin{aligned} \mu^+ &\rightarrow e^+ + \nu_e + \bar{\nu}_\mu \\ (\mu^- &\rightarrow e^- + \bar{\nu}_e + \nu_\mu) \end{aligned} \quad (2.4)$$



**Figure 2.3:** Left: Sketch of the hadronic interaction chain occurring in relativistic jets of AGN resulting in neutrino and gamma ray production. Credit: Katz and Spiering 2012 [15]  
 Right: Sketch of the interaction chain producing atmospheric showers. Credit: Bernhard 2014 [3]

creating even more neutrinos. The pions  $\pi^0$  and  $\pi^+$  are expected to be produced equally, which means that neutrino production is likely to be proportional to photon production due to photo-pion-production (Kadler et al. 2016 [14]). At first glance the neutrino flavor distribution ratio  $\nu_e : \nu_\mu : \nu_\tau$  might not be even but taking into account the fact that the probability of a neutrino to change its flavor while going from one point to another point in space is not neglectable, the ratios change as well, resulting in approximately 1 : 1 : 1 after propagation of a long distance.

Compared to other particles in respect to the survey of astronomical objects, neutrinos show a clear advantage. Charged particles for example can be deflected through a magnetic field while high energy photons might be absorbed [12]. On the contrary, due to the lack of charge, the trajectory of neutrinos through space is nearly undisturbed. That means the direction of a neutrino measured on earth matches the direction to the origin and the reconstructed path directly points to a possible source [3, 25]. Nevertheless, one has to be careful as neutrinos measured on earth can also originate from the earth's atmosphere (called atmospheric neutrinos). High energetic, hadronic cosmic particles can induce a chain of interactions when hitting molecules in the atmosphere

(see Fig. 2.3). The resulting shower has an electromagnetic and hadronic branch, while the second is of importance for neutrino production. The hadronic shower part produces pions along with kaons, which can also interact according to

$$\begin{aligned} K^+ &\rightarrow \mu^+ + \nu_\mu \\ K^- &\rightarrow \mu^- + \bar{\nu}_\mu \\ K^+ &\rightarrow \pi^0 + e^+ + \nu_e \end{aligned} \quad (2.5)$$

where the last process is much less likely compared to the other two. The spectrum of cosmic rays  $\frac{dN}{dE} \propto E^{-2.7}$  is not fully applicable for atmospheric neutrinos. Due to the fact that pions and kaons possibly collide and transfer some of their energy into atmospheric neutrinos before their decay (which is caused by their rather long lifetime in comparison with other elementary particles), the occurrence of high energetic neutrinos changes. Therefore, for atmospheric neutrinos the spectrum results in being steeper as reported by [10]:

$$\left(\frac{dN}{dE}\right)_{n,atm} \propto E^{-3.7}. \quad (2.6)$$

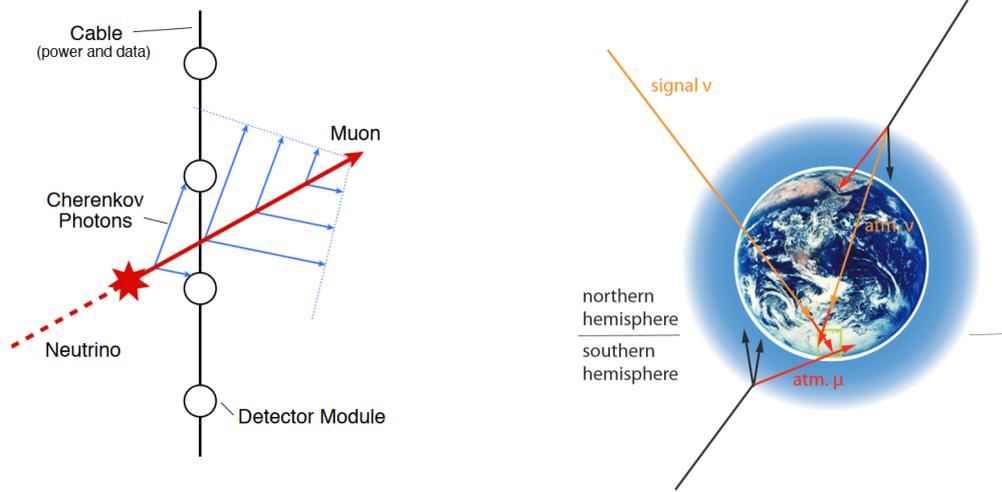
When neutrinos interact with charged particles, the excitation or motion of the latter results in observable electromagnetic radiation which allows for neutrino detection. Both CC and NC interactions are eligible, creating further particles over the recoil. Plus, in a CC interaction a charged lepton is generated. If the charged particles produced over the interaction have a sufficiently large velocity on their way through the transparent medium, Cherenkov light is being emitted. Cherenkov radiation is the result of charged particles going through a medium at a velocity greater than the phase velocity of the medium and can be measured (see Fig. 2.4). Coherent photons are emitted by the electrons interacting with the particle over conservation of momentum and energy. The radiation produced in the process is traveling in a particular angle  $\theta$  from the trajectory direction of the particle:

$$\theta = \arccos\left(\frac{1}{n\beta}\right) \quad (2.7)$$

where  $n$  is the refraction index of the medium and  $\beta = \frac{v}{c}$  is the particle velocity. The total emitted energy  $dE$  can be described by the Frank-Tamm-formula

$$dE = \frac{e^2}{4\pi} \mu(\omega) \omega \left(1 - \frac{1}{n^2(\omega)\beta^2}\right) dx d\omega \quad (2.8)$$

for the radiation frequency  $\omega$  and a particle that has travelled the distance  $x$ .  $\mu(\omega)$  is the permeability and  $n(\omega)$  the refraction index dependent on the frequency. In order to measure Cherenkov light, one has to distinguish two possible kinds of events, track and



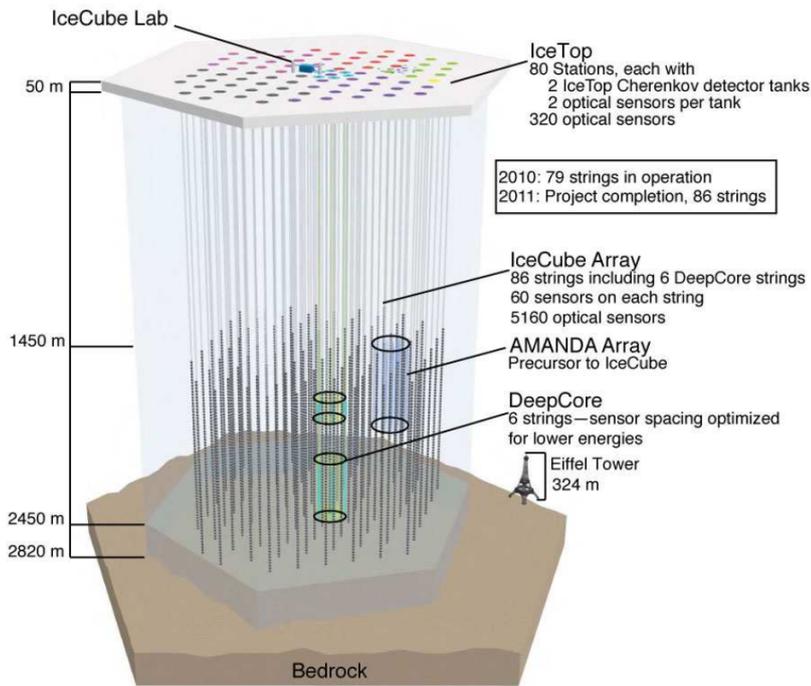
**Figure 2.4:** Left: Sketch of the neutrino detection principle using a line of sensors; the incoming neutrino (dashed red) interacts creating a muon (solid red) which itself produces Cherenkov photons (blue). Credit: Weaver 2015 [25]  
 Right: Visualization of particles detected by IceCube. Atmospheric particles from the northern hemisphere are shielded towards the detector, so mostly cosmic neutrinos get through. From the southern hemisphere both cosmic as well as atmospheric particles are measured as they can enter the detector from above. Credit: Bernhard 2014 [3]

shower. Considering a track type event, the emitting (secondary) particle may travel through the medium while hardly interacting, producing Cherenkov light all along the way. They are able to scatter stochastically which can lead to daughter particles causing showers. A shower event refers to a particle which is not able to travel quite as long without interacting, creating other secondary particles. Altogether, the process results in diffuse radiation emission. Therefore, the calculation of the trajectory of the primary particle for a shower like event is way worse than for a track. For instance, in IceCube the medium angular error is typically  $< 2^\circ$  for tracks and reaches up to  $> 40^\circ$  for showers (see Tab. 3.1). In return, the energy resolution for showers is better than for tracks. The reason is that, for showers, the total energy is likely deposited within the detector volume. This is not applicable to tracks though as they are able to pass the detector eventually [25].

# 3 The IceCube detector

Located at the south pole, built under a thick layer of ice, the IceCube detector is meant to find high energetic neutrinos originating from astronomical objects. The detector measures the Cherenkov light which is produced when neutrinos interact with charged particles within the ice using highly sensitive photo multipliers [12].

## 3.1 Construction



**Figure 3.1:** Structure of the IceCube detector consisting of IceTop and the DOM array in a depth of 1450 to 2450m with subarrays AMANDA and DeepCore. Credit: IceCube Collaboration 2011 [10]

The IceCube detector consists of 86 strings containing 60 digital optical modules (short DOM) each, resulting in a total number of 5160 DOMs. The structure can be seen in Fig. 3.1. The DOMs are connected through a cable for power supply and communication. There is a sub array with smaller spacing between the DOMs, namely

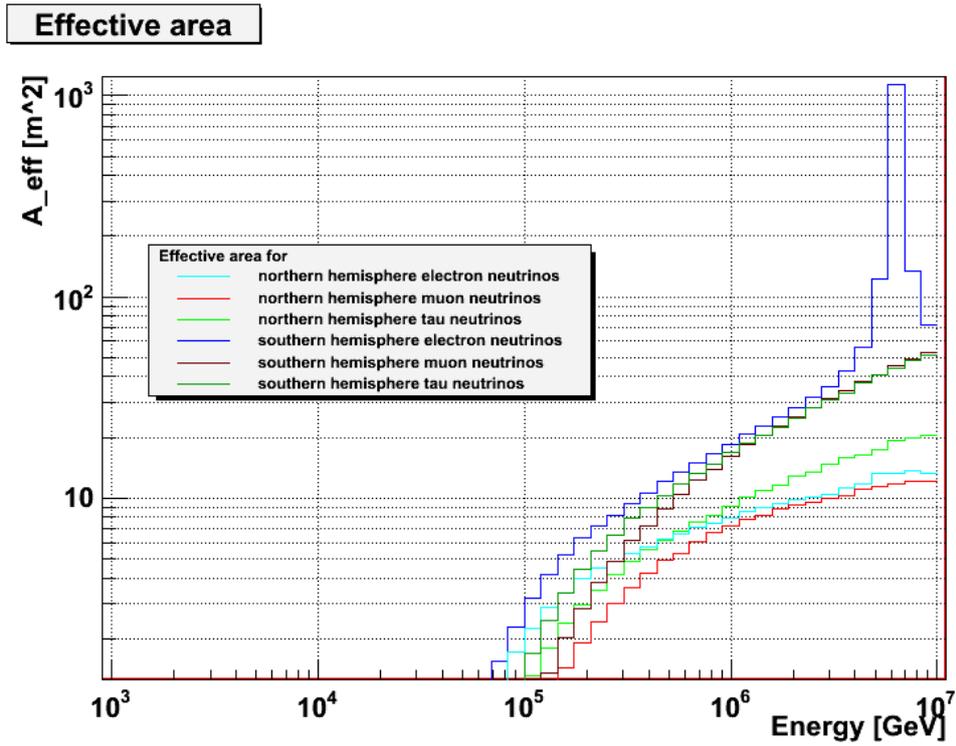
DeepCore which is meant to target neutrino energies of  $\sim 10$  GeV. Another sub array named AMANDA was the former experiment to the current detector. On the surface there are 324 optical sensors in an array to measure cosmic ray air-showers which plays an important role in background suppression. The integrated volume is  $1 \text{ km}^3$ . A DOM contains a photomultiplier tube and the electronics for Cherenkov radiation measurement [12].

## 3.2 Detection principle

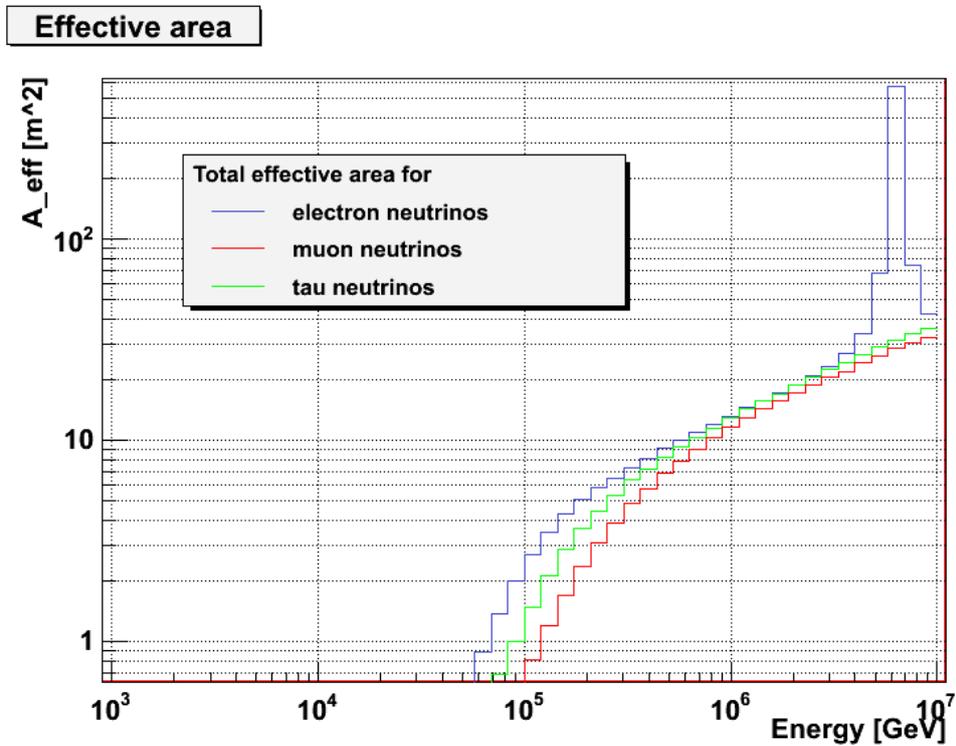
As shown in section 2.2, the neutrino interaction can be divided into charged (CC) and neutral current (NC) interactions. In a CC interaction, the charged lepton created has an energy of about 50% up to 80% of the initial neutrino energy, whereas the percentage itself depends on the energy as well [9]. The energy difference is transferred to a hadronic shower. Possible CC interactions feature muon neutrinos as well as electron and taon neutrinos, however, the properties vary significantly. When a muon is produced it can be measured as a track type event with an angular resolution of  $< 1^\circ$  [3]. As the particle passes the detector volume the energy resolution is rather bad compared to showers. For electron neutrinos, an electron is created. Due to bremsstrahlung, this particle ends up in both a hadronic cascade as well as an electromagnetic shower, whereas the latter gets about 0.5 to 0.8% of the particle energy. The interaction completely takes place within the detector, therefore the energy resolution is much better than for tracks. Unfortunately, the shower is rather assymmetric, which makes it more difficult to determine the electron trajectory. Due to an elongation of the shower in the electron direction it can be done nevertheless [8]. Tau neutrinos produce taons with a lifetime of order  $10^{-13}$ . The decay has two possible outcomes, whereas for low energies one cannot tell the difference to a shower caused by an electron neutrino. Aside from that, the characteristics of a high energetic tau neutrino have not yet been observed in IceCube [3].

## 3.3 Effective area

The efficiency of a detector can be described using the so called effective area. This quantity corresponds to the area of a hypothetical detector with an efficiency of 100% to detect neutrinos of given energy  $E$  in an angle  $\theta$  with respect to the zenith. In addition,  $A_{eff}$  depends on the neutrino flavor. The  $A_{eff}$  data is distinguished by the northern and southern hemisphere according to the background topology. Particles from the northern hemisphere have to pass the earth before reaching the IceCube detector at the South Pole. As a result, background muons are shielded towards the detector, however, background neutrinos are not affected. The background event rate is  $\sim 10^5$  events per year. For the southern hemisphere, the shielding property of the earth has no effect and



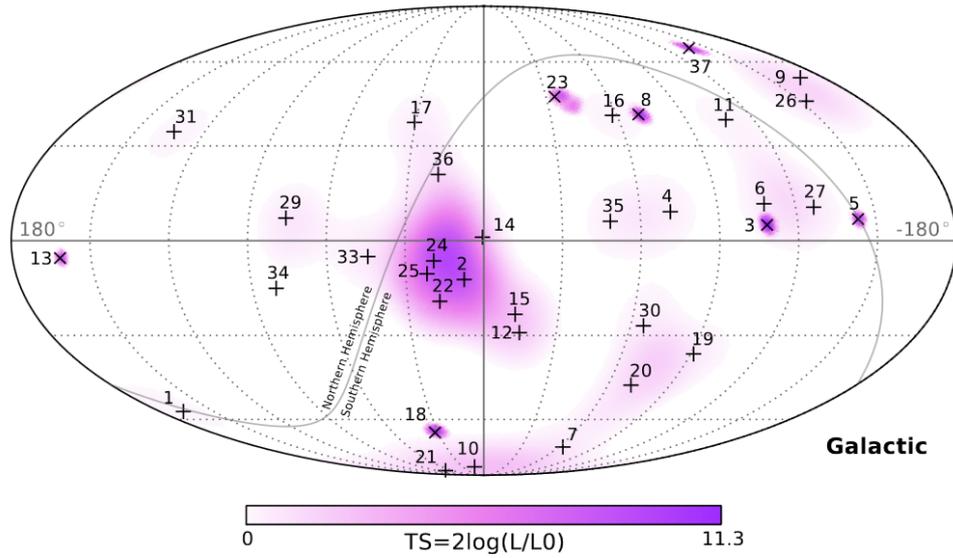
**Figure 3.2:** Effective area  $A_{eff}$  as a function of energy for the individual neutrino flavors distinguished by hemisphere. Data credit: IceCube Collaboration 2013 [11]



**Figure 3.3:** All sky effective area  $A_{eff}$  as a function of energy for the individual neutrino flavors. Data credit: IceCube Collaboration 2013 [11]

the background event rate is immense ( $\sim 10^{11}$  events per year) [3]. The effective area is shown individually per flavor and hemisphere in Fig. 3.2. An all sky average per flavor can be seen in Fig. 3.3. Remarkably, the southern hemisphere  $\nu_e$  effective area features a peak at  $\approx 6.3$  PeV which is caused by the Glashow resonance (see Glashow 1960 [6]).

### 3.4 First results



**Figure 3.4:** Galactic plot of the 37 (36) HESE events detected by IceCube in 988 days of measurement. Accumulation in direction of the Galactic Center can be observed. Credit: IceCube Collaboration 2014 [12]

In mid 2014, the IceCube Collaboration [12] published the results obtained from 988 days of measurements, dating from 2010 to 2013. The outcome comprises 37 events containing an estimated number of  $8.4 \pm 4.2$  cosmic ray muons and  $6.6^{+5.9}_{-1.6}$  atmospheric neutrino background events. The event sample referred to as *High Energy Starting Events* (HESE) is shown in Fig. 3.4. An accumulation towards the Galactic Center can be observed. The event properties can be found in Tab. 3.1. For events 32 (HESE-32) and 28 (HESE-28) constraints have to be made. As HESE-32 is caused by two coinciding cosmic muons, it is lacking coordinates and an energy value, leaving only an observation time. HESE-28 is rated presumable as background because it triggered hits in the IceTop cosmic air shower array.

The IceCube Collaboration [13] published a list containing 17 additional events in 2015. The resulting sample with a total number of 54 events is obtained from 1347 days ranging from 2010 to 2014.

ID	Dep. Energy (TeV)	Observation Time (MJD)	Decl. (deg.)	R.A. (deg.)	Med. Angular Error (deg.)	Event Topology
1	$47.6^{+6.5}_{-5.4}$	55351.3222143	-1.8	35.2	16.3	Shower
2	$117^{+15}_{-15}$	55351.4659661	-28.0	282.6	25.4	Shower
3	$78.7^{+10.8}_{-8.7}$	55451.0707482	-31.2	127.9	$\lesssim 1.4$	Track
4	$165^{+20}_{-15}$	55477.3930984	-51.2	169.5	7.1	Shower
5	$71.4^{+9.0}_{-9.0}$	55512.5516311	-0.4	110.6	$\lesssim 1.2$	Track
6	$28.4^{+2.7}_{-2.5}$	55567.6388127	-27.2	133.9	9.8	Shower
7	$34.3^{+3.5}_{-4.3}$	55571.2585362	-45.1	15.6	24.1	Shower
8	$32.6^{+10.3}_{-11.1}$	55608.8201315	-21.2	182.4	$\lesssim 1.3$	Track
9	$63.2^{+7.1}_{-8.0}$	55685.6629713	33.6	151.3	16.5	Shower
10	$97.2^{+10.4}_{-12.4}$	55695.2730461	-29.4	5.0	8.1	Shower
11	$88.4^{+12.5}_{-10.7}$	55714.5909345	-8.9	155.3	16.7	Shower
12	$104^{+13}_{-13}$	55739.4411232	-52.8	296.1	9.8	Shower
13	$253^{+26}_{-22}$	55756.1129844	40.3	67.9	$\lesssim 1.2$	Track
14	$1041^{+132}_{-144}$	55782.5161911	-27.9	265.6	13.2	Shower
15	$57.5^{+8.3}_{-7.8}$	55783.1854223	-49.7	287.3	19.7	Shower
16	$30.6^{+3.6}_{-3.5}$	55798.6271285	-22.6	192.1	19.4	Shower
17	$200^{+27}_{-27}$	55800.3755483	14.5	247.4	11.6	Shower
18	$31.5^{+4.6}_{-3.3}$	55923.5318204	-24.8	345.6	$\lesssim 1.3$	Track
19	$71.5^{+7.0}_{-7.2}$	55925.7958619	-59.7	76.9	9.7	Shower
20	$1141^{+143}_{-133}$	55929.3986279	-67.2	38.3	10.7	Shower
21	$30.2^{+3.5}_{-3.3}$	55936.5416484	-24.0	9.0	20.9	Shower
22	$220^{+21}_{-24}$	55941.9757813	-22.1	293.7	12.1	Shower
23	$82.2^{+8.6}_{-8.4}$	55949.5693228	-13.2	208.7	$\lesssim 1.9$	Track
24	$30.5^{+3.2}_{-2.6}$	55950.8474912	-15.1	282.2	15.5	Shower
25	$33.5^{+4.9}_{-5.0}$	55966.7422488	-14.5	286.0	46.3	Shower
26	$210^{+29}_{-26}$	55979.2551750	22.7	143.4	11.8	Shower
27	$60.2^{+5.6}_{-5.6}$	56008.6845644	-12.6	121.7	6.6	Shower
28	$46.1^{+5.7}_{-4.4}$	56048.5704209	-71.5	164.8	$\lesssim 1.3$	Track
29	$32.7^{+3.2}_{-2.9}$	56108.2572046	41.0	298.1	7.4	Shower
30	$129^{+14}_{-12}$	56115.7283574	-82.7	103.2	8.0	Shower
31	$42.5^{+5.4}_{-5.7}$	56176.3914143	78.3	146.1	26.0	Shower
32	-	56211.7401231	-	-	-	Coincident
33	$385^{+46}_{-49}$	56221.3424023	7.8	292.5	13.5	Shower
34	$42.1^{+6.5}_{-6.3}$	56228.6055226	31.3	323.4	42.7	Shower
35	$2004^{+236}_{-262}$	56265.1338677	-55.8	208.4	15.9	Shower
36	$29.9^{+3.0}_{-2.6}$	56308.1642740	-3.0	257.7	11.7	Shower
37	$30.8^{+3.3}_{-3.5}$	56390.1887627	20.7	167.3	$\lesssim 1.2$	Track

**Table 3.1:** HESE events observed by IceCube in 988 days of measurements. Event 32 is caused by two coinciding cosmic muons, therefore only the observation time could be determined. Aside from event 32, event 28 shows hits in the IceTop cosmic air shower array as well, which means they are most likely background muons. Credit: IceCube Collaboration 2014 [12]



# 4 Development and application of the algorithm

Hereafter, an introduction into the concept of the maximum likelihood method is given. Subsequently, the development of the algorithm applying the unbinned maximum likelihood method is described. Using the algorithm, the correlation between IceCube neutrino events and AGN is analyzed afterwards.

## 4.1 Unbinned maximum likelihood method

A common approach to determine a certain parameter of a statistical model is the so called *maximum likelihood method*. The following description is based on Blobel 1998 [4] and Schwarze 1997 [21]. Shortly, the method selects the parameter which yields maximum agreement of a model with a given set of data. When the probability function  $\rho$  for one observational result  $x$  is

$$\begin{aligned} \rho : \Omega &\rightarrow [0; 1] \\ x &\mapsto \rho(x|\theta), \end{aligned} \tag{4.1}$$

where  $\Omega$  is the sample space, the corresponding likelihood function is given by

$$\begin{aligned} L : \Theta &\rightarrow [0; 1] \\ \theta &\mapsto \rho(x|\theta) \end{aligned} \tag{4.2}$$

with  $\Theta$  being the space of all possible parameter values. The value of  $L$  can be interpreted as the probability that one observes  $x$  given the parameter  $\theta$ . The maximum likelihood estimator determines the value of  $\theta$  which maximizes the likelihood and therefore the probability to measure  $x$ . Suppose we have a variable  $X$  whose probability density  $f$  is a function of parameter  $q$ . Given a random sample, the function  $f$  can be factorized according to

$$f(x_1, x_2, \dots, x_n; q) = \prod_{i=1}^n f_{X_i}(x_i; q) \tag{4.3}$$

where  $n$  is the number of independently and uniformly distributed realizations of the random variable. Now it is possible to determine the density for fixed (observed) values  $x_i$  as a function of  $q$ . This results in the following likelihood function

$$L(q) = \prod_{i=1}^n f_{X_i}(x_i; q). \quad (4.4)$$

Maximizing this function with respect to  $q$  will most likely be difficult since  $L$  is not necessarily differentiable. This is the reason why in many cases one takes the log-likelihood function whose advantage is obvious. By taking the logarithm of the whole function, the product is being transformed into a sum, and the function can be computed much easier. Conveniently, the monotonicity of the logarithm does not change the position of the maximum, so the calculation of the maximum can be simplified.

$$\log(L) = \log\left(\prod_{i=1}^n f_{X_i}(x_i; q)\right) = \sum_{i=1}^n \log(f_{X_i}(x_i; q)) \quad (4.5)$$

Generally, the maximum likelihood estimator selects the set of values that yields the best possible agreement with an underlying statistical model. Especially, one uses a slightly different method in physics, the *unbinned maximum likelihood method*. In this case, the (log-)likelihood function is given by

$$L(n_s) = \prod_{i=1}^N \left(\frac{n_s}{N} \cdot S_i + \left(1 - \frac{n_s}{N}\right) \cdot B_i\right) \quad (4.6)$$

$$\log[L(n_s)] = \sum_{i=1}^N \log\left(\frac{n_s}{N} \cdot S_i + \left(1 - \frac{n_s}{N}\right) \cdot B_i\right) \quad (4.7)$$

where one computes the (log-)likelihood function  $L$  dependent on the total number of events  $N$ , the signal probability distribution function  $S_i$  and the background probability distribution function  $B_i$  for each event  $i$ . The function  $L$  is being maximized in respect to the parameter  $n_s$ . When considering a set of events, it is generally unknown whether the detected particles originate from the sources or belong to background. Although it is intended to eliminate background events from consideration as much as possible, a total correction is clearly not guaranteed. Thus, the set of events represents a composition of 'interesting' source like neutrinos and background particles. The parameter  $n_s$  corresponds to the number of events that are most likely to originate from sources within the tested sample, i.e. the number yielding the maximum agreement with the emission model, while  $(N - n_s)$  events are expected to be caused by background particles.

## 4.2 Developing the algorithm

### 4.2.1 Model of the algorithm

The following statistical analysis is based on the model presented in Glüsenskamp et al. 2015 [7]. Starting from Eq. 4.7, we will have a closer look at all individual parts of the likelihood for a better understanding. The probability distribution function (PDF) for an event  $i$  to be of signal type is given as

$$S_i = \frac{\sum_j^{N_{src}} w_{j,tot} \cdot S_j(x_i; \sigma_i) \cdot \epsilon_j(E_i)}{\sum_j^{N_{src}} w_{j,tot}}, \quad (4.8)$$

where  $S_j$  is the point spread function of source  $j$  at a certain celestial coordinate  $x_i$  (i.e. the reconstructed event origin coordinate).  $S_j$  is taking into account the angular error of the event  $\sigma_i$  which is supposed to be shaped like a 2D-Gaussian.  $w$  is a weighting factor ( $w_{tot} = w_{src} \cdot w_{dec}$ ), consisting of  $w_{src}$  and  $w_{dec}$ . First,  $w_{src}$  is adapted depending on the weighting model one uses for the sources. It is either  $w_{src} = 1$  for an equal weighting scheme or  $w_{src} \propto F_\gamma$  allowing for the association between neutrinos and the gamma flux  $F_\gamma$  of an individual source caused by neutral pion decays (see section 2.2). The second factor,  $w_{dec}$ , is motivated by the properties of the detector, a declination dependent detector response weighting. Furthermore, there is the energy PDF  $\epsilon_j$ , which is obtained through Monte Carlo simulations. The background PDF  $B_i$  is in that case computed from the data distribution and the reconstructed energy.

During the development process, the algorithm is not tested on a source sample but rather a fictional sample, where the coordinates are set equal the event coordinates. The source flux therefore is not considered in this stage and an equal weighting scheme is applied. Regarding the coordinates, i.e. declination and right ascension along with the angular error, this method is supposed to yield the best possible agreement. Furthermore, only the smaller sample of HESE events (i.e. 36 events, see section 3.4) was used for testing.

### 4.2.2 Signal PDF $S_i$

In this section a more detailed description of the individual ingredients of the signal PDF will be given. The basic equation is only considering the signal PDF, neglecting the  $B_i$  term. The signal PDF is again simplified in terms of its individual parts (regarding the weighting factor  $w_{j,tot}$ , the energy PDF  $\epsilon_j$  and the point spread function  $S_j$ , see above).

In the end, the function reads

$$\log [L] = \sum_{i=1}^N \log \left( \frac{n_s}{N} \cdot S_i \right), \quad (4.9)$$

$$S_i = \frac{\sum_j^{N_{src}} S_j(x_i; \sigma_i)}{N_{src}}. \quad (4.10)$$

Regarding Eq. 4.8, the signal PDF changes concerning the weight factor and the energy PDF (in particular  $w_{j,tot} = \epsilon_j(E_i) = 1$ ). The point spread function (PSF)  $S_j$  is at this point modelled as a step function where  $\alpha$  is the angular distance between the event coordinates  $x_i$  and the source  $j$ .  $\sigma_i$  is the angular error of event  $i$ . The PSF reads

$$S_j(x_i; \sigma_i) = \begin{cases} 1 & \alpha < \sigma_i \\ 0 & \alpha > \sigma_i. \end{cases} \quad (4.11)$$

The angular distance  $\alpha$  is obtained as

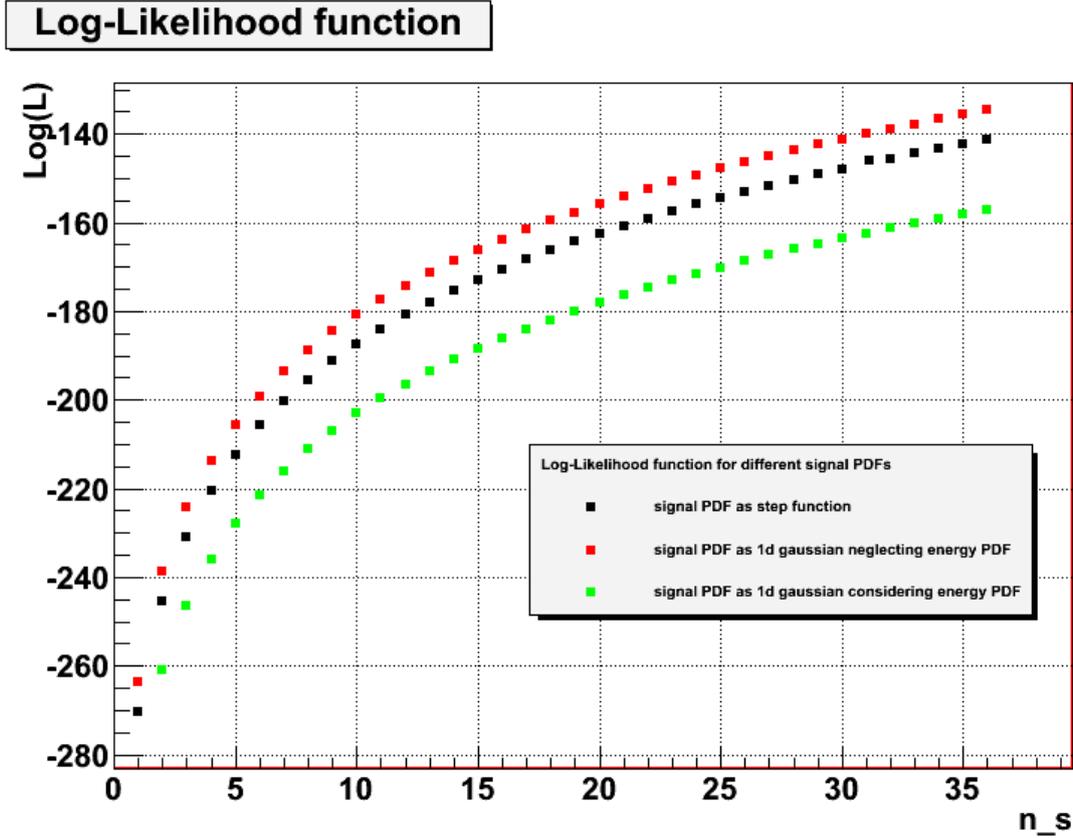
$$\alpha = \arccos \left( \frac{\cos(d) \cdot \cos(a) \cdot \cos(d_s) \cdot \cos(a_s) + \cos(d) \cdot \sin(a) \cdot \cos(d_s) \cdot \sin(a_s) + \sin(d) \cdot \sin(d_s)}{\sqrt{(\cos(d) \cdot \cos(a))^2 + (\cos(d) \cdot \sin(a))^2 + (\sin(d))^2} \cdot \sqrt{(\cos(d_s) \cdot \cos(a_s))^2 + (\cos(d_s) \cdot \sin(a_s))^2 + (\sin(d_s))^2}} \right) \quad (4.12)$$

with  $d$ ,  $a$  as declination and right ascension of the reconstructed origin of the event and  $d_s$ ,  $a_s$  similarly for the source. Note that the used PDFs are always normalized (i.e.  $\sum S_i = \sum B_i = \sum \epsilon_i = 1$ ). These simplifications yield a log-likelihood function of a form that can be seen in Fig. 4.1 (black). The shape of the graph is induced by momentarily neglecting the background PDF. For that reason, one obtains a linear function for the argument of the logarithm. Then, the logarithm creates the characteristic logarithmic shape. From that point, it is not possible to perform the maximum likelihood method. When the background PDF  $B_i$  is taken into account, it will change the shape of the graph significantly as the background provides a decrease opposing to the signal PDF which will result in the likelihood function to have a maximum.

The next step is to model the PSF  $S_j$  more accurately. This is done with a gaussian representation. The underlying model suggests a 2D-Gaussian at this point, but more conveniently one can also use a 1D-Gaussian combined with the angular distance instead of the 2D-Gaussian with spherical coordinates. Using Eq. 4.12 the resulting 1D-Gaussian is given by

$$S_j(\alpha_{i,j}; \sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \cdot \exp \left( -\frac{\alpha_{i,j}^2}{2\sigma_i^2} \right) \quad (4.13)$$

with  $\alpha_{i,j}$  now being the angular distance between event  $i$  and source  $j$ . The result can be seen in Fig. 4.1 (red). The shape remains the same, however, the range changes. The next extension is the energy PDF  $\epsilon_i$ . Due to the efficiency of the detector, the



**Figure 4.1:** Loglikelihood functions obtained by neglecting the background term  $B_i$ . The signal term  $S_i$  differs, as a first approach it uses a step function for the PSF  $S_j$  (Eq. 4.11) (black). The other two outcomes are created as the PSF is being modelled as a 1D-Gaussian and the energy PDF  $\epsilon_{S,j}$  is either neglected (red) or considered according to Eq. 4.14 (green).

energy PDF depends on the reconstructed energy of the event  $E_i$  and the direction of the source (in particular the zenith angle of the neutrino):

$$\epsilon_{S,j}(E_i) = E_i \cdot A_{eff}(E_i, x_{S,j}) \quad (4.14)$$

where  $A_{eff}$  is the effective area dependent on the neutrino energy  $E_i$  and direction  $x_j$ , which is described in section 3.3.

Taking the energy PDF into account, one obtains a log-likelihood function according to Fig. 4.1 (green) with a similar shape but a different range.

The last implementation of  $S_i$  is the weighting factor  $w_j$ . So far, the equal weighting scheme ( $w_j = 1$ ) is applied. The weighting factor depends on the individual source properties. As shown in section 2.2, the neutrino flux of a blazar is assumed to be

proportional to the gamma flux  $F_\nu \propto F_\gamma$ . Thus,  $w_j$  can be calculated as

$$w_j = \frac{F_j}{F_{max}} \quad (4.15)$$

where  $F_j$  is the individual gamma flux of source  $j$  and  $F_{max}$  is the maximum gamma flux of all sources considered in the current test run. A declination dependent treatment according to the model (see section 4.2.1) is however only done through the energy PDF.

### 4.2.3 Background PDF $B_i$

The background term which represents the probability for a detected neutrino to be of background origin is implemented in the likelihood function now. Since the background neutrino direction is expected to be isotropically distributed, the particles lack a favoured direction and their origin direction can be simulated by using random numbers. The first approach for the background PDF is similar to the first attempt for  $S_i$ , however only the event coordinates are of importance at the moment:

$$B_i = \frac{\sum_j^{N_{bg}} S_j(x_i; \sigma_i)}{N_{bg}} \quad (4.16)$$

where  $S_j$  is the PSF of event  $i$  with coordinates  $x_i$  and angular error  $\sigma_i$  evaluated at the coordinates of a background neutrino  $j$ . For every source considered in the computation a background counterpart is being generated, therefore  $N_{bg} = N_{src}$ .

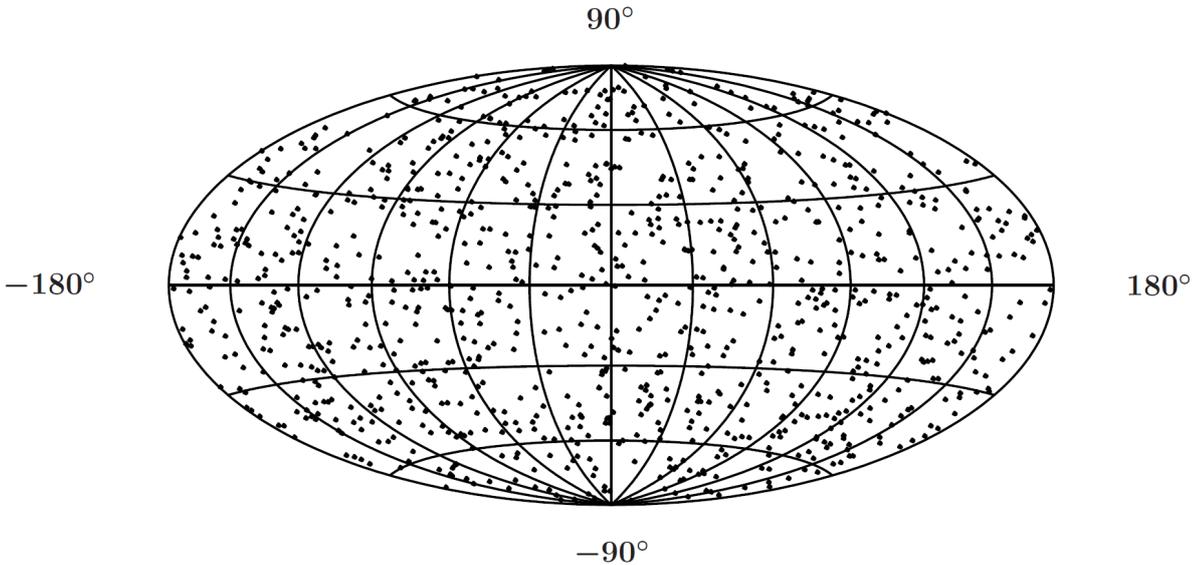
Usually random numbers are generated as a uniform distribution. In order to determine random coordinates, i.e. declination and right ascension, the random numbers have to be transformed. An approach of directly using the random numbers as declination  $\delta \in [-90^\circ; +90^\circ]$  and right ascension  $\alpha \in [0^\circ; 360^\circ]$  will result in accumulation points at both poles. A proper distribution of coordinates can be obtained by using

$$\delta = \arccos(2r_1 - 1) \quad (4.17)$$

$$\text{and } \alpha = 2\pi \cdot r_2 \quad (4.18)$$

for the transformation of two uniformly distributed numbers  $r_1, r_2 \in [0; 1]$  into desired coordinates. This transformation is tested in advance resulting in the distributions seen in Fig. 4.2, no accumulation points observed.

Using random numbers, it is not sufficient to just consider one single outcome of the algorithm since the results vary significantly. This step changes the log-likelihood function drastically according to Fig. 4.3 which shows different possible log-likelihood results. Clearly, the maximum of the function can vary. Therefore, from this point on the consideration of individual log-likelihood functions no longer makes sense. Instead, the count of maxima occurring per value of  $n_s$  in a large number of simulated runs will be considered.



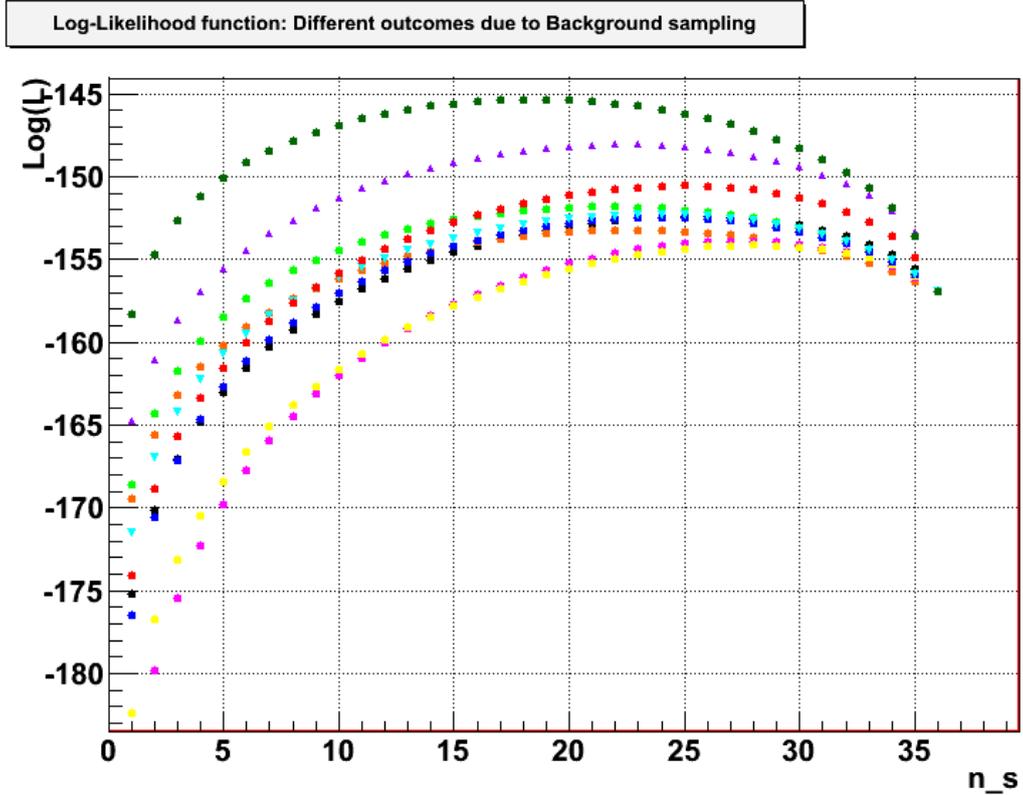
**Figure 4.2:** Result of the transformation of random numbers applied to celestial coordinates with  $N_{rnd} = 1000$  runs. As can be seen, no accumulation points are being created. Similar plots using  $N_{rnd} = 100$  and  $10000$  can be found in the Appendix (Fig. A.1) for comparison.

As the individual value of the log-likelihood is not of importance, the values of  $n_s$  at which a maximum occurs can be saved in a histogram. The histograms obtained using different numbers of Monte Carlo runs are given in the Appendix (Fig. A.2 to A.4). It also becomes clear that increasing the number of runs does not change the range of the maximum but makes the corresponding histogram smoother.

The next step in the development of the background term is the consideration of the background neutrino energy  $E$ . Background neutrinos are expected to follow a spectrum  $\frac{dN}{dE} \propto E^{-3.7}$  (see Eq. 2.6) [10]. By generating the background sample through random numbers, each neutrino gets a set of coordinates and an energy value. The transformation is done using the relation

$$P(E_{rnd}) = \left( \frac{1-r}{E_{min}^{2.7}} + \frac{r}{E_{max}^{2.7}} \right)^{-\frac{10}{27}} \quad (4.19)$$

where  $E_{rnd}$  is the obtained energy,  $r \in [0; 1]$  is a random number and  $E_{min(max)}$  the lower (upper) limit for  $E_{rnd}$ . In this case,  $E_{min(max)}$  is the minimum (maximum) energy of the events. The spectrum has been simulated using random numbers according to Eq. 4.19. The resulting histogram is shown in the Appendix (Fig. A.8). The background PDF



**Figure 4.3:** Different results for the loglikelihood function using the first approximation for the background PDF  $B_i$  considering but the event coordinates. The plot shows that the position of the maximum can vary due to randomly generated background neutrino samples.

now reads

$$B_i = \frac{\sum_j^{N_{bgr}} S_j(x_i; \sigma_i) \cdot \epsilon_{B,j}(E_i)}{N_{bgr}} \quad (4.20)$$

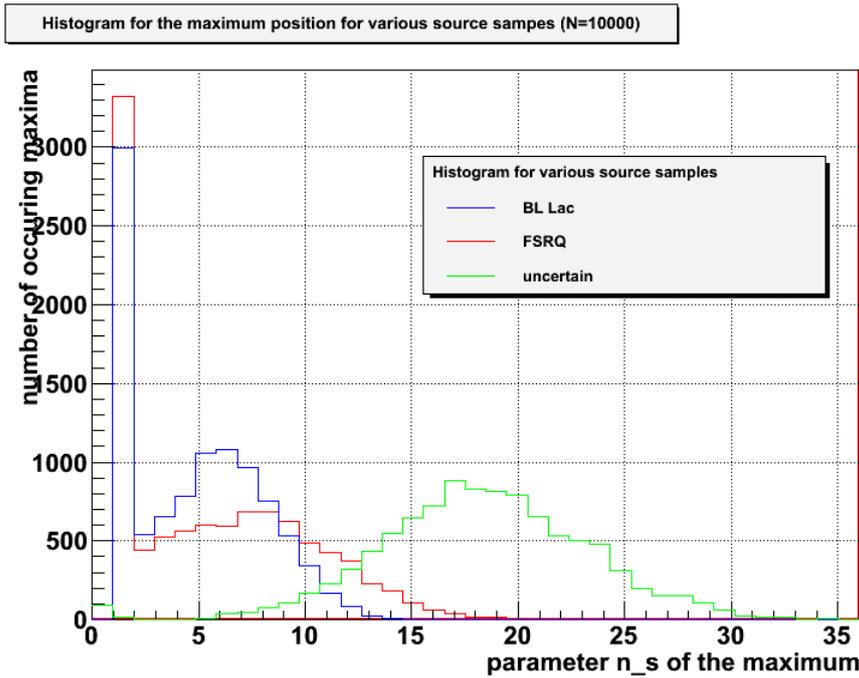
where the agreement of the coordinates is computed in the same way as in the signal term using the Gaussian  $S_j$  with the coordinates of background neutrino  $j$ . The number of background neutrinos is equal to the number of sources tested ( $N_{bgr} = N_{src}$ ). The energy PDF used for  $B_i$ ,  $\epsilon_{B,j}$  is however slightly different towards its signal counterpart:

$$\epsilon_{B,j}(E_i) = \begin{cases} 0 & |E_i - E_{rnd,j}| > \sigma_i \\ E_i \cdot A_{eff}(E_i, x_{B,j}) & |E_i - E_{rnd,j}| \leq \sigma_i \end{cases} \quad (4.21)$$

where  $E_i$  is the reconstructed energy of event  $i$  with error  $\sigma_i$  and  $x_{B,j}$  the origin coordinates of simulated background neutrino  $j$ . This results in neglecting a simulated

neutrino relevant to the background PDF when its energy is not consistent with the energy of event  $i$  within the error range. At this point, the resulting function has again been evaluated using different numbers of Monte Carlo runs. The histograms can be seen in the Appendix (Fig. A.5 to A.7).

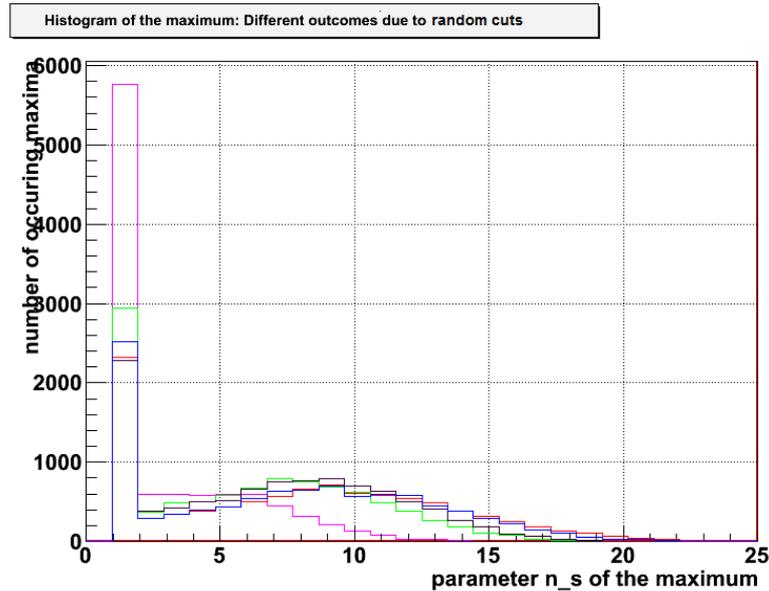
### 4.3 Testing



**Figure 4.4:** Resulting histograms evaluating the log-likelihood functions for different source samples with  $N = 10000$  runs. The peaks at  $n_s = 0$  occur due to the fact that the tail  $< 0$  of the histogram is being cumulated, therefore they cannot be considered maxima.

At this stage in algorithm development, actual testing can be done. As described in chapter 2, the sources which the IceCube events are tested against are blazars. In particular, the corresponding data is obtained from the catalogue published by Massaro et al. 2015 [17]. The set is divided into three groups, namely *BL Lac*, *FSRQ* and *Uncertain* with 1425, 1909 and 227 members respectively. For the computation, the algorithm required only the coordinates, i.e. right ascension and declination, as well as the *Fermi*-flux at an energy of 1 – 100 GeV.

Using  $N_{MC} = 10000$  simulation runs, one obtains a histogram according to Fig. 4.4. In order to make sure the discrepancy between the BL Lac and FSRQ sample is not caused by the different sample sizes, one can randomly choose FSRQ sources to be



**Figure 4.5:** Tests performed on the FSRQ sample randomly cut down to the size of the BL Lac sample. The outcome is highly dependent on the randomly selected sources that are cut out.

cut out, resulting in the sample having the same size as BL Lac. Considering random numbers, this has to be done several times. In fact, the result varies significantly as can be seen in Fig. 4.5. Therefore it makes no sense to adjust the sizes this way. Another possibility to determine the influence of the sample size is to consider only the ( $\gamma$ -)brightest sources of every sample. However, the difference in the sizes is being considered in the number of randomly created background events, resulting in every source having a virtual background 'source' counterpart.

Regarding Fig. 4.4, a maximum of the histogram at a larger value of  $n_s$  indicates a better agreement with the model for the sample. The values and ratios are shown in Tab. 4.1. Thus, the Uncertain sample apparently provides a much higher agreement than BL Lac and FSRQ. However, as stated in Massaro et al. 2015 [18], the Uncertain sample is composed of sources with peculiar properties that also show blazar characteristics. For example, there are galaxies containing a blazar nucleus with a low luminosity or crossover objects between BL Lac and radio galaxies. As a consequence the sample is difficult to rate. Besides, FSRQ has a maximum at a slightly larger value of  $n_s$  than BL Lac. This implies that the FSRQ sample yields a higher agreement with the model than BL Lac.

sample	$n_s$ of the maximum	ratio
BL Lac	6	16.6%
FSRQ	7.5	20.8%
Uncertain	17	47.2%

**Table 4.1:** Results of a test using  $N = 10000$  Monte Carlo simulated individual likelihood functions. The ratio implies the percentage of neutrino events that yields maximum agreement with the model of emission by the corresponding source sample.

## 4.4 Error consideration

In view of the histograms and plots presented earlier, one eventually notices the lack of error bars. Clearly, when working with measured data, error consideration is crucial. The IceCube events provide an error in coordinates (i.e. the medium angular error) as well as an error of the reconstructed energy. The first error mentioned comes into play in the PSF  $S_j$ . The second one is considered in the effective area and when the energy agreement with the background spectrum is determined. Altogether, it is possible to determine error bars for the (log-)likelihood function. Ensuing from that, the important part of the (log-)likelihood function is the maximum ( $n_s | \log(L(n_s))$ ), which obtains an error in both parameters. The error in  $n_s$  is essential as it belongs to the item of interest. However, the algorithm evaluates a large number of (log-)likelihood functions ( $N_L$ ), hence it is analogous with a series of measurements. The evaluation ensues statistically and the individual errors are dropped. Usually, a statistical evaluation yields an expected value along with the standard deviation, in sum being suitable for an error consideration. In this case, the distribution is asymmetrical, thus the measure of interest is not the expected value but rather the mode ( $M$ ). This measure, the value with the highest appearance in a set of data, is unfortunately lacking a standard deviation equivalent. In order to determine its error, one would have to perform another large number of simulations ( $N_m$ ). This would result in a set of mode values of size  $N_m$  which can be evaluated statistically by determining the expected value and standard deviation of the mode  $M$ .

So far the error consideration does not pose a major problem. The limiting factor is the computation time. Having two levels of simulation, the number of background sampling runs multiplies ( $N_{tot} = N_L \cdot N_m$ ). Nevertheless the simulation numbers should not be reduced without flinching as a reduction corrupts the suitability for statistical evaluation. It is the principle 'the more the better' for statistical treatment. The challenge is finding numbers  $N_L, N_m$  that are large enough to justify a statistical evaluation while being small enough to maintain a reasonable computation time.



## 5 Conclusion and outlook

Based on the model introduced by Glüsenkamp et al. 2015 [7], an algorithm has been developed to determine the correlation between IceCube neutrino events and their presumed sources, namely blazars. The algorithm allows for the most important criteria regarding a source correlation test, that is, coordinates and energy consideration. However, one can say that it is difficult to clearly prove the origin of the neutrinos detected by IceCube. Solely a comparing statement is possible. Nevertheless, a basic algorithm has been built, which can be further enhanced easily given a larger time span.

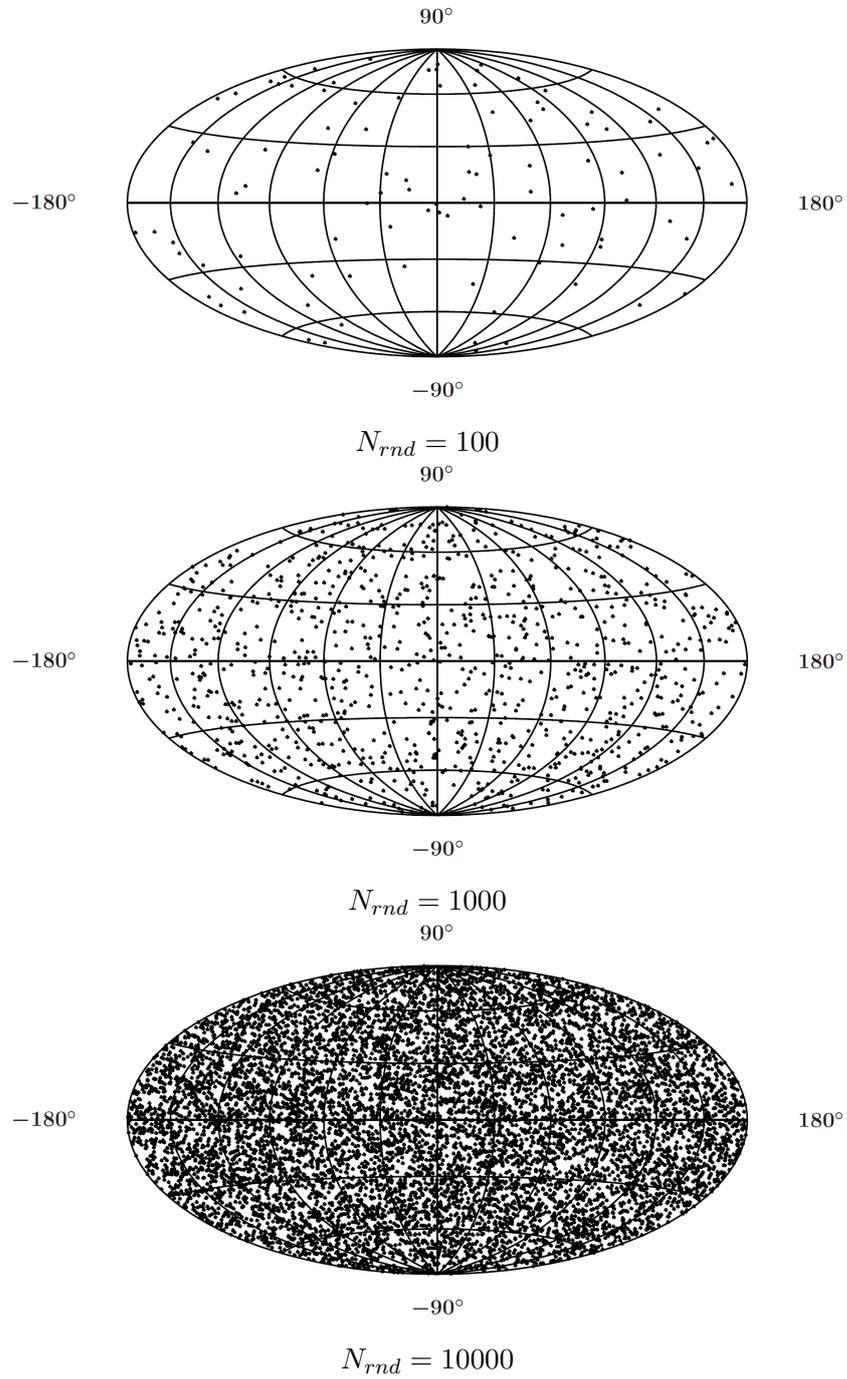
Based on the assumptions made during the development, it is possible to make a comparing statement of the samples tested. So, it is shown that the FSRQ sample has maximum agreement at a number of 7.5 neutrino events, whereas with BL Lac only 6 neutrino events are much likely to be suitable. The Uncertain sample provides a number of 17 events, therefore it is the sample with the highest agreement out of the three tested samples. It has been ruled out that this discrepancy compared to FSRQ and BL Lac is caused by a shift in the flux weighting due to an exceptionally bright source in the Uncertain sample. Although the dependency of the sample size has been investigated regarding BL Lac (1425 members) and FSRQ (1909 members), it might be responsible for this result considering the small size of the Uncertain sample (227 members). Therefore it would be important to further evaluate the influence of the sample sizes. This can be done by cutting the BL Lac and FSRQ samples down to the size of the Uncertain sample, whereas either random sources or the least luminous ones are dropped. Besides, it would be interesting to show a histogram obtained from a test against a purely randomly simulated source sample and whether its maximum is located at a higher or smaller value of  $n_s$  compared to the tested samples.

More precisely, other properties can be considered in the computation of the likelihood, for instance the time component of an event (see Tab. 3.1). It can be matched against the variability of a source. The algorithm can also be enhanced regarding the energy. In the current status, it only considers the background spectrum whereas the source neutrino spectrum (assumed  $\frac{dN}{dE} \propto E^{-2.3}$ ) has not been taken into account.

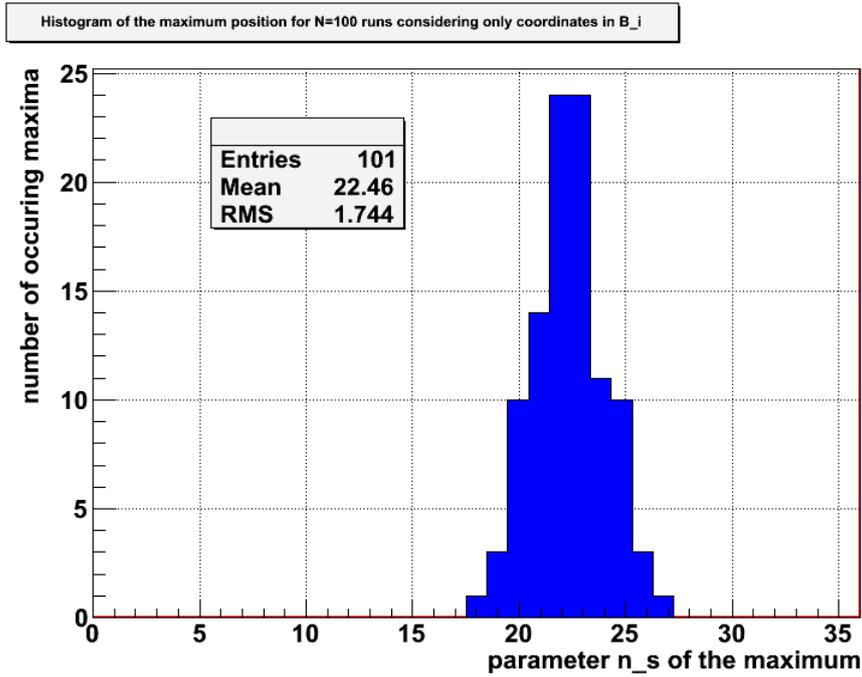
Another property which is hopefully being enhanced in the future is the event sample size. For statistical treatment, a data count of 37 is barely enough to consider at all. Due to the fact that the IceCube detector has just been up and running for a couple of years now, one can be confident that there are a lot more neutrino events to be detected. Having a larger data set, it will be much easier to make a reliable statement about the origin of extraterrestrial high-energy neutrino events.



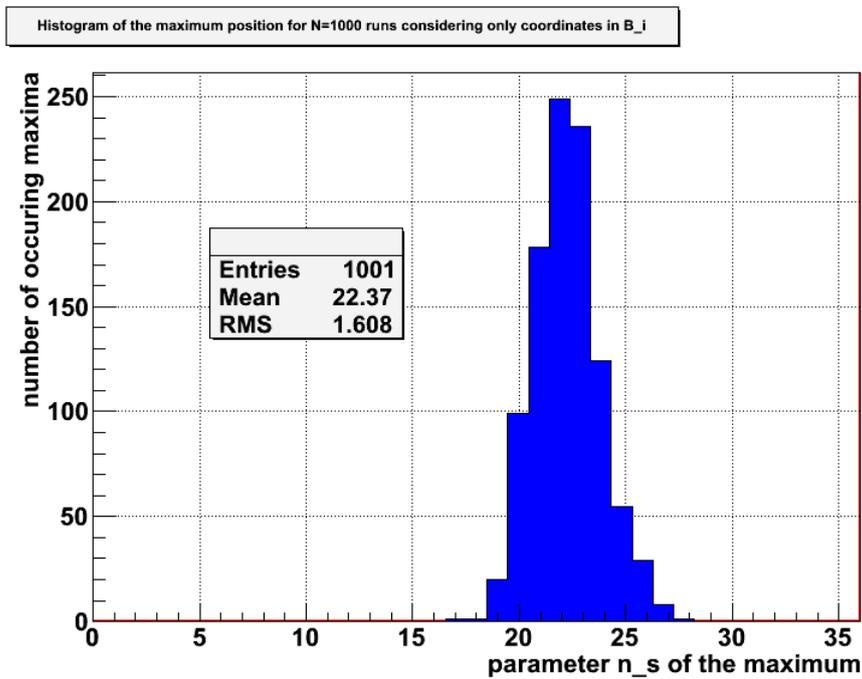
# Appendix



**Figure A.1:** Testing the transformation of random numbers applied to celestial coordinates with different numbers of runs  $N_{rnd}$ . As can be seen, no accumulation points are being created.



**Figure A.2:** Histogram of the maximum position when  $B_i$  is only considering the event coordinates for  $N_{rnd} = 100$  runs.



**Figure A.3:** Histogram of the maximum position when  $B_i$  is only considering the event coordinates for  $N_{rnd} = 1000$  runs.

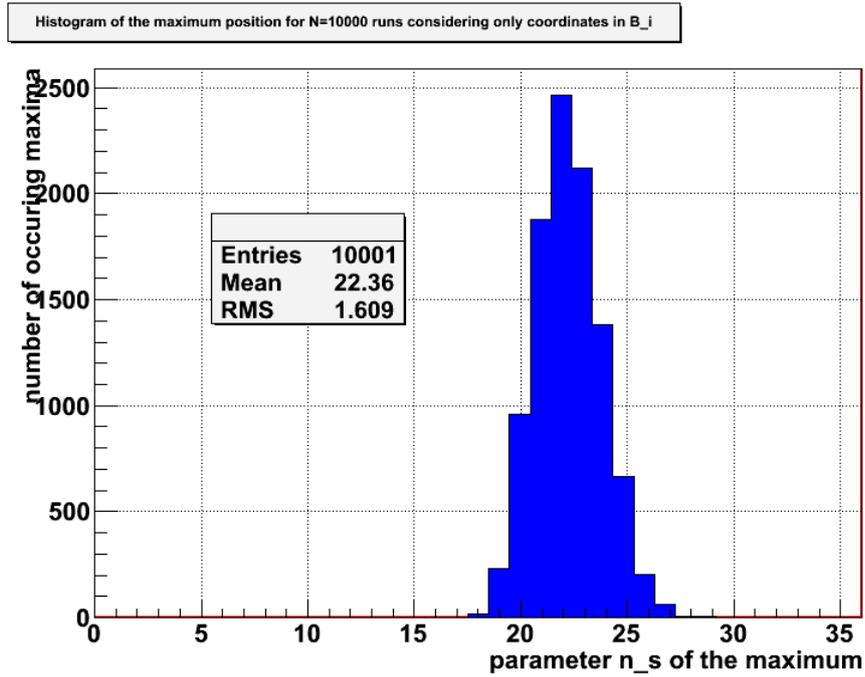


Figure A.4: Histogram of the maximum position when  $B_i$  is only considering the event coordinates for  $N_{rnd} = 10000$  runs.

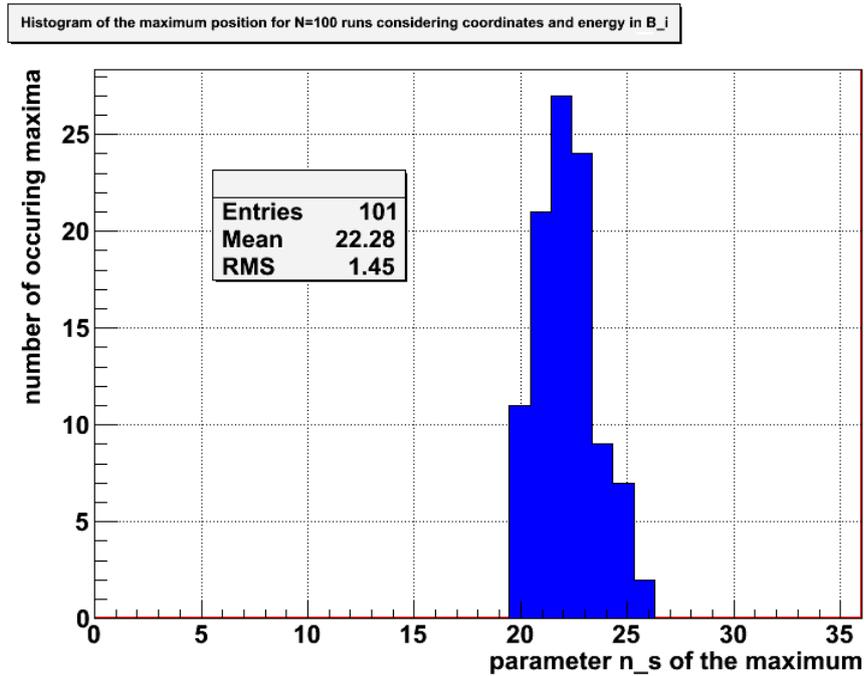
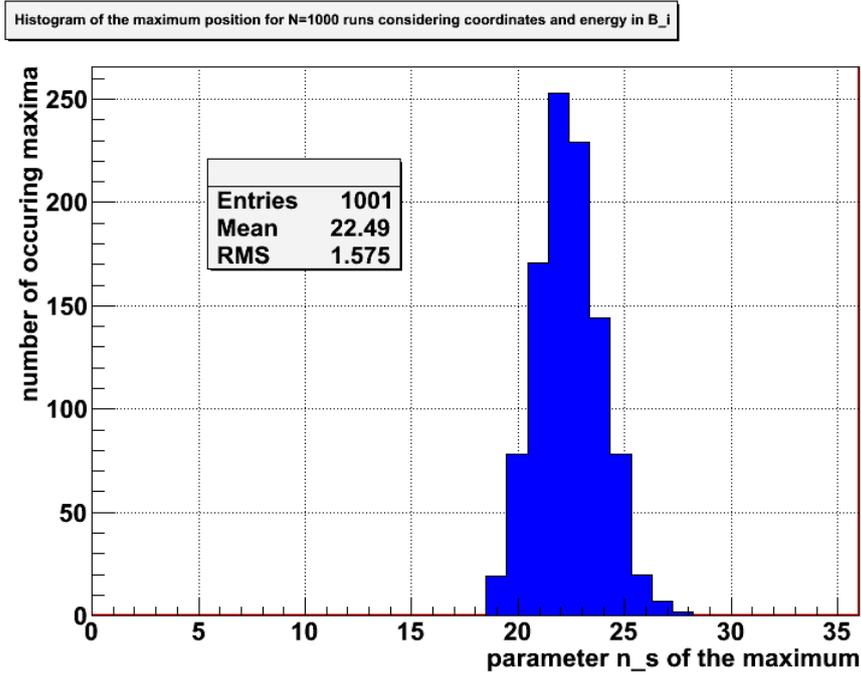
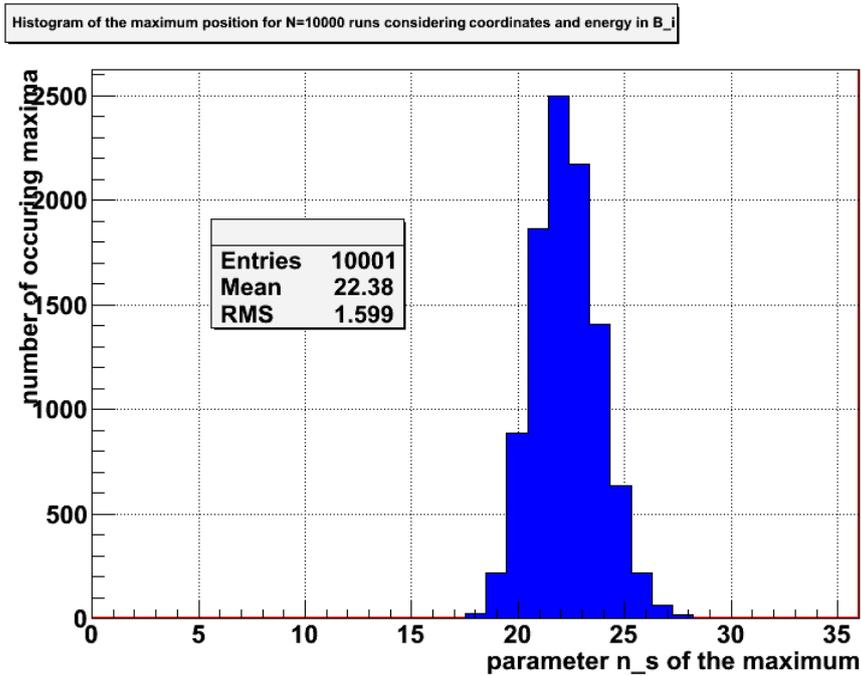


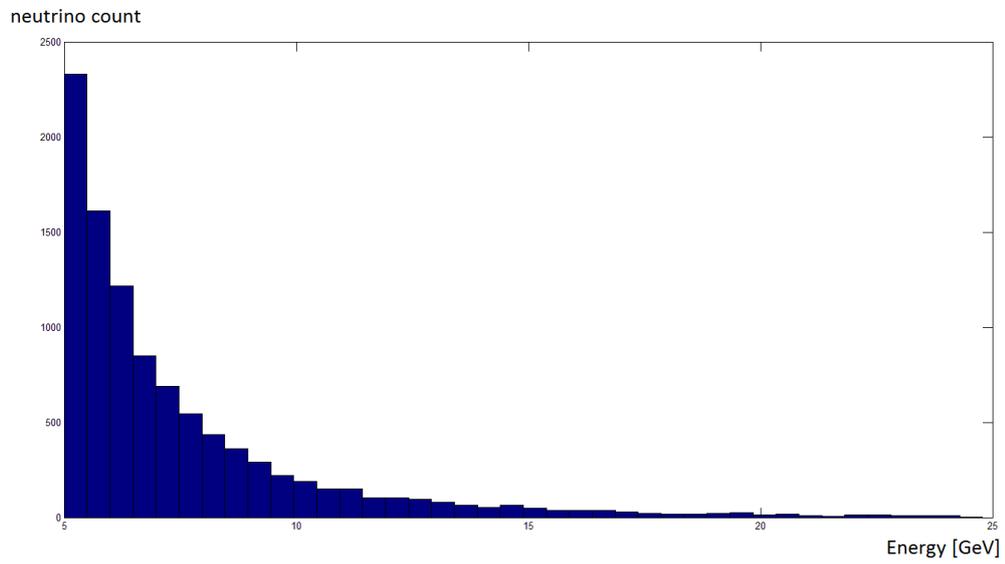
Figure A.5: Histogram of the maximum position when  $B_i$  is now considering event coordinates as well as the reconstructed energy for  $N_{rnd} = 100$  runs.



**Figure A.6:** Histogram of the maximum position when  $B_i$  is now considering event coordinates as well as the reconstructed energy for  $N_{rnd} = 1000$  runs.



**Figure A.7:** Histogram of the maximum position when  $B_i$  is now considering event coordinates as well as the reconstructed energy for  $N_{rnd} = 10000$  runs.



**Figure A.8:** Exemplary visualization of the distribution obtained from Eq. 4.19 using  $N_{rnd} = 10000$  random numbers and 40 classes with  $E_{min} = 5$  GeV and  $E_{max} = 25$  GeV.

# Bibliography

- [1] Aartsen M. G. et al. Energy Reconstruction Methods in the IceCube Neutrino Telescope. *JINST*, 9:P03009, 2014.
- [2] Volker Beckmann and Chris Shrader. *Active Galactic Nuclei*. Wiley-VCH, 2012.
- [3] Anna Bernhard. *Origin of IceCube's Astrophysical Neutrinos: Autocorrelation, Multi-Point-Source and Time-Structured Searches*. PhD thesis, Technische Universität München, 2014.
- [4] Volker Blobel and Erich Lohrmann. *Statistische und numerische Methoden der Datenanalyse*. Teubner Verlag, 1998.
- [5] Bernard L. Fanaroff and Julia M. Riley. The morphology of extragalactic radio sources of high and low luminosity. *Mon. Not. R. Astron. Soc.*, 167, 1974.
- [6] Sheldon L. Glashow. Resonant Scattering of Antineutrinos. *Phys. Rev.*, 118(316), 1960.
- [7] Thorsten Glüsenkamp. Analysis of the cumulative neutrino flux from Fermi-LAT blazar populations using 3 years of IceCube data. *Proceedings for the RICAP-14 conference*, 2015.
- [8] Francis Halzen. Astroparticle Physics with High Energy Neutrinos: from AMANDA to IceCube. *Eur. Phys. J.*, C46:669–687, 2006.
- [9] Francis Halzen and Spencer R. Klein. IceCube: An Instrument for Neutrino Astronomy. *Rev.Sci.Instrum.*, 81(081101), 2010.
- [10] IceCube Collaboration. Measurement of the atmospheric neutrino energy spectrum from 100 GeV to 400 TeV with IceCube. *Physical Review D*, 83.1(012001), 2011.
- [11] IceCube Collaboration. *Search for contained neutrino events at energies above 30 TeV in 2 years of data*, 2013. <http://icecube.wisc.edu/science/data/HE-nu-2010-2012> [Accessed: 2015-12-04].
- [12] IceCube Collaboration. Observation of High-Energy Astrophysical Neutrinos in Three Years of IceCube Data. *Phys. Rev. Lett.*, 113(101101), 2014.

- [13] IceCube Collaboration. The IceCube Neutrino Observatory - Contributions to ICRC 2015 Part II: Atmospheric and Astrophysical Diffuse Neutrino Searches of All Flavors. *Proceedings for the 34th International Cosmic Ray Conference*, 2015.
- [14] Kadler M. et al. Coincidence of a high-fluence blazar outburst with a PeV-energy neutrino event. *Nature Physics*, 2016.
- [15] Ulrich F. Katz and Christian Spiering. High-energy neutrino astrophysics: Status and perspectives. *Progress in Particle and Nuclear Physics*, 67(3):651–704, 2012.
- [16] Julian H. Krolik. *Active Galactic Nuclei: From the Central Black Hole to the Galactic Environment*. Princeton University Press, 1998.
- [17] E. Massaro, A. Maselli, C. Leto, P. Marchegiani, M. Perri, P. Giommi, and S. Piranomonte. *The Roma BZCAT 5th edition Multi-frequency Catalogue of Blazars*, 2015. [http://www.asdc.asi.it/bzcat/#start\\_page](http://www.asdc.asi.it/bzcat/#start_page) [Accessed: 2016-02-10].
- [18] Massaro E. et al. The 5th edition of the Roma-BZCAT. A short presentation. *Astrophys. Space Sci.*, 357, 2015.
- [19] Bogdan Povh and Klaus Rith. *Teilchen und Kerne: Eine Einführung in die physikalischen Konzepte*. Springer, 1999.
- [20] Schulz R. et al. The gamma-ray emitting radio-loud narrow-line Seyfert 1 galaxy PKS 2004-447. II. The radio view. *A&A*, 588:A146, 2016.
- [21] Jochen Schwarze. *Grundlagen der Statistik – Band 2: Wahrscheinlichkeitsrechnung und induktive Statistik*. Nwb Verlag, 1997.
- [22] Carl Seyfert. C. Nuclear Emission in Spiral Nebulae. *Astrophys. Journal*, 97(28), 1943.
- [23] Rickard Ström. *Exploring the Universe Using Neutrinos. A Search for Point Sources in the Southern Hemisphere Using the IceCube Neutrino Observatory*. PhD thesis, Uppsala University, 2015.
- [24] C. Megan Urry and Paolo Padovani. Unified Schemes for Radio-Loud Active Galactic Nuclei. *Publications of the Astronomical Society of the Pacific*, 107(715):803–845, 1995.
- [25] Christopher Weaver. *Evidence for Astrophysical Muon Neutrinos from the Northern Sky*. PhD thesis, University of Wisconsin-Madison, 2015.

# Acknowledgements

As a last point, I would like to express my gratitude to Prof. Karl Mannheim, Prof. Matthias Kadler and the chair of astronomy for giving me the opportunity to work on this topic. Especially I am thankful for the patience Prof. Mannheim and Prof. Kadler showed. Further, many thanks go out to Michael Kreter for guiding me and also Dr. Dorit Glawion for proof reading and helping out whenever. Moreover, I am thanking the work group for the warm welcome and the helpfulness.

Special thanks go out to my fellow students, namely G-Boys and Milena, who are responsible for the study days to become much less unbearable than I probably would have thought at the beginning of my studies. At that point I would also like to mention the boys from the bachelor office, thanks for all the inspiring discussions and the good time.

At last, my parents and my siblings Cornelia and Michael should not go unmentioned. Thank you for always being prepared to have my back. I am also deeply thankful to Denise and her flatmates for the incredible welcome and everything she did, and to Ko for making me who I am today.

## *Bibliography*

---

# Erklärung

Hiermit erkläre ich, dass die vorliegende Arbeit nach allgemeiner Studien- und Prüfungsordnung für die Bachelor- und Masterstudiengänge (ASPO) an der Julius-Maximilians-Universität Würzburg selbstständig und nur unter Benutzung der angegebenen Quellen und Hilfsmittel angefertigt wurde.

Des Weiteren wurde die Arbeit noch keiner anderen Prüfungsbehörde zur Erlangung eines akademischen Grades vorgelegt.

Würzburg, 04. Juli 2016

---

Juliane Wiehl