



Dissertation for the Degree  
Master in Physics and Astronomy

# COMPARISON OF STATISTICAL METHODS TO EVALUATE THE ICECUBE DISCOVERY POTENTIAL FOR STEADY POINT SOURCES

**PABLO CORREA**  
2016-2017

Promotor: Prof. Dr. Nick van Eijndhoven  
Co-promotor: Dr. Krijn D. de Vries  
Sciences & Bio-Engineering Sciences



*To Life*

*Op het Leven*

*A la Vida*

*A la Vie*



# Acknowledgments

First and foremost I would like to thank my family, especially my mother, father, grandmother, grandfather, and uncle, who have always supported me throughout the endeavor that is my life. They have given me the opportunity to become the person that I am today and to pursue any major or minor goals that I would want to achieve during my time on this Earth. One of these was to become a physicist, and I could never have succeeded in doing so if I did not have these amazing people standing behind me. Mamá, papá, abuela, abuelo, tito: ¡gracias por todo, os quiero mucho!

Next, I want to thank my friends, particularly those who are closest to me, the other members of The Uff Marones: Filip, Grisha, Laurens, Edo, Alessio, Florian, and Alexandre. Every moment in time that I share with these great lads, from our long intriguing discussions about humanity to the most ridiculous nights at Bar Deco, I cherish with all my heart. They have lead me to develop a mentality to enjoy every possible aspect of life, and this mentality has made the difficult moments in the making of this work much more bearable. Makers, bedankt voor alles, en op nog vele toekomstige avonturen!

Furthermore, I would like to acknowledge the fact that without the help of my classmates Jonas, Paul, Marine, Matthias, and Jonathan, who have been there since the beginning of these five years, I would probably not be writing this right now. Hopefully all of us get the chance to develop a successful career, whether or not it will be one with a lot of physics. Want we hebben het toch wel gedaan hè!

I also want to express a sincere thank you to my promoters, Nick and Krijn, who have given me more and better guidance in becoming a scientist than I could have ever hoped for. Since my first experience with research during my bachelor project two years ago, they have been the persons who were always available, not only for interesting discussions about physics, but also for any help I would need throughout the less and more difficult periods of my studies. Moreover, their support during the making of this work has been incredible, and on top of that, they have rigorously revised this manuscript multiple times. Nick, Krijn, bedankt—ik kijk er naar uit om nog met jullie samen te werken!

Not to forget in these acknowledgments are all the members of the IIHE, and in particular the ICECUBE team. Their constant availability to aid anyone with any problem gives the IIHE a very warm atmosphere, and this makes the department a really pleasant workspace. I also want to thank the IT team for allowing me to use that extra computation power which allowed me to finish this work.

Finally, I would like to express my gratitude towards Jorgen D'Hondt and Stijn Buitink, who have accepted to be a part of the reading committee for this thesis.



# Contents

<b>Acknowledgments</b>	<b>v</b>
<b>Contents</b>	<b>vii</b>
<b>Introduction</b>	<b>1</b>
<b>1 Neutrino Astronomy</b>	<b>3</b>
1.1 High-Energy Cosmic Rays and Cosmic Neutrinos . . . . .	3
1.1.1 The Unknown Origin of Cosmic Rays at the Highest Energies . . . . .	3
1.1.2 Production of High-Energy Cosmic Neutrinos . . . . .	5
1.2 The ICECUBE Neutrino Observatory . . . . .	6
1.2.1 Detection Principle and Event Topologies . . . . .	6
1.2.2 DEEPCORE and ICETOP . . . . .	9
1.2.3 Determination of a Neutrino Flux from an Event Rate . . . . .	10
1.3 The Birth of Neutrino Astronomy . . . . .	11
1.3.1 First Observations of High-Energy Cosmic Neutrinos . . . . .	11
1.3.2 Status of Steady Point Source Searches with ICECUBE . . . . .	12
<b>2 Statistics in Neutrino Astronomy</b>	<b>17</b>
2.1 An Overview of Statistics . . . . .	17
2.1.1 Probability Distribution Functions . . . . .	18
2.1.2 Likelihoods . . . . .	18
2.1.3 Classical Confidence Intervals and Confidence Limits . . . . .	19
2.1.4 Bayesian Credibility Intervals . . . . .	24
2.2 Techniques for a Statistical Analysis with ICECUBE . . . . .	25
2.2.1 Test Statistics and Pseudo-Experiments . . . . .	25
2.2.2 Discovery Potential . . . . .	28
2.2.3 Sensitivity . . . . .	28
2.2.4 Analysis of Experimental Data . . . . .	30
2.3 Additional Remarks . . . . .	31
<b>3 Toy Model of Steady Point Sources</b>	<b>33</b>
3.1 Toy Model . . . . .	33
3.1.1 Background Generation . . . . .	34
3.1.2 Signal Generation . . . . .	34
3.2 Analysis in a Simplified Counting Experiment . . . . .	35

<b>4</b>	<b>Statistical Methods</b>	<b>39</b>
4.1	Li-Ma Method . . . . .	40
4.2	Likelihood Method . . . . .	41
4.2.1	Expected Angular Distribution . . . . .	41
4.2.2	Maximization of the Likelihood . . . . .	42
4.2.3	Construction of the Likelihood Test Statistic . . . . .	43
4.3	PLT Method . . . . .	44
4.3.1	Signal as a Perturbation of the Background . . . . .	45
4.3.2	Maximization of the Score . . . . .	45
4.4	$\psi$ Method . . . . .	48
4.4.1	Assessment of Significance in a Bayesian Approach . . . . .	48
4.4.2	Binning of the Expected Angular Background Distribution . . . . .	49
<b>5</b>	<b>Results of the Toy Model</b>	<b>51</b>
5.1	Comparison of the Statistical Methods . . . . .	51
5.1.1	Background-Only Distributions . . . . .	51
5.1.2	Discovery Potentials and Sensitivities . . . . .	52
5.1.3	Probabilities for Discovery . . . . .	53
5.2	Implication for Steady Point Source Studies . . . . .	55
	<b>Conclusions and Outlook</b>	<b>61</b>
	<b>Summary</b>	<b>63</b>
	<b>Samenvatting in het Nederlands</b>	<b>65</b>
<b>A</b>	<b>Simulation Code of Pseudo-Experiments with the Toy Model</b>	<b>67</b>
<b>B</b>	<b>Distributions for Discovery Potentials and Sensitivities</b>	<b>75</b>
	<b>Bibliography</b>	<b>81</b>

# Introduction

More than a century ago, Victor F. Hess discovered the extraterrestrial particles that are now known as *Cosmic Rays* (CRs) [1]. Typically, these are considered to be charged atomic nuclei, as will be done implicitly henceforth. Even though they have been observed up to the highest energies ever detected in a particle, of the order of  $10^{20}$  eV [2], their origin remains one of the unresolved problems in the field of astroparticle physics. In particular, there is currently no evidence for the source of the CRs with energies above  $10^{18}$  eV, also called *Ultra-High-Energy Cosmic Rays* (UHECRs).

Because CRs are charged, they are deflected by magnetic fields. However, UHECRs cannot be contained by the magnetic field of the Milky Way [3], and they are therefore expected to have an extragalactic origin. Due to its charge, an (UHE)CR will not point back to its source when detected at Earth. On its path, it will have traversed unmapped (inter)galactic magnetic fields that deflect them in their trajectory. Thus, observations of UHECRs cannot be used to determine their origin.

Fortunately, hadronic interactions of UHECRs within the high-energy environment of their source are expected to produce *high-energy cosmic neutrinos* [4]. These cosmic neutrinos have no charge and a low interaction cross section, so they can reach Earth undeflected and quasi unattenuated. Therefore, they serve as ideal cosmic messengers of the high-energy environment in which they were produced. The determination of their source(s) will most likely resolve the problem of the UHECR origin.

The existence of high-energy cosmic neutrinos was not an established fact until their discovery with the ICECUBE neutrino telescope in 2013 [5]. However, due to the limited statistics at the highest energies, up to this date the origin of the observed high-energy cosmic neutrinos has not been determined yet. In this thesis, the focus lies on steady point sources of (high-energy) electromagnetic radiation as candidates to account for the observed high-energy cosmic neutrino flux.

These steady point sources mostly comprise different subclasses of Active Galactic Nuclei (AGN) [6]. The sources that have been investigated so far have relatively low population densities but high electromagnetic luminosities, which could imply a large flux of high-energy cosmic neutrinos coming from each of these sources. Yet, these objects have been discarded as the sole origin of the observed high-energy cosmic neutrinos [7], and therefore the interest of steady point source searches is shifting more towards objects with lower electromagnetic luminosities but higher population densities, or transient sources.

Thus, if these latter candidate sources are the sole origin of high-energy cosmic neutrinos, each of them is expected to produce a relatively low cosmic neutrino flux compared with the objects that have already been investigated. Therefore, relatively

less neutrino events in ICECUBE are expected for each candidate steady point source that corresponds with a high-density population. However, in the case of such a low signal per source, it might be that other statistical methods give a higher potential for a  $5\sigma$  discovery compared to the standard statistical method applied in point source analyses with ICECUBE, which is a maximum likelihood method. This is the main motivation for the work presented in this thesis, in which four different statistical methods, including the likelihood method, will be compared.

In the following, first a more detailed description of (UHE)CRs, high-energy cosmic neutrinos, ICECUBE, and candidate steady point sources is given in Chapter 1. Subsequently, Chapter 2 covers the frequentist techniques required to perform a statistical analysis with ICECUBE. Then, a toy model of steady point sources is constructed in Chapter 3, which will be used to compare the statistical methods that will be discussed in Chapter 4. Finally, in Chapter 5 the results are presented of this comparison using the toy model.

# Neutrino Astronomy

## Introduction

In this first Chapter, a brief overview of neutrino astronomy is presented. To begin, the link between Ultra-High-Energy Cosmic Rays and high-energy cosmic neutrinos is made in Section 1.1. Subsequently, Section 1.2 covers the main aspects of the ICECUBE neutrino observatory, such as the detection principle and the different signatures of high-energy neutrinos in the detector. Lastly, the first observations of high-energy astrophysical neutrinos, which mark the start of neutrino astronomy, are discussed in Section 1.3, where also the current status of searches for steady point sources as the possible origin of these observed high-energy cosmic neutrinos is presented.

## 1.1 High-Energy Cosmic Rays and Cosmic Neutrinos

### 1.1.1 The Unknown Origin of Cosmic Rays at the Highest Energies

The full Cosmic Ray spectrum above 10 TeV, including all detected compositions, is shown in Fig. 1.1, which has been obtained through observations of the CR flux  $F(E)$  over several orders of magnitude in energy  $E$  by multiple experiments [8–25]. It can be seen that the spectrum clearly follows a broken power law,  $F(E) \propto E^\Gamma$ , where the spectral index  $\Gamma$  depends on the energy range. Before the “knee” at  $4 \times 10^{15}$  eV,  $\Gamma \approx -2.7$ , after which it softens to  $\Gamma \approx -3.1$ . Around  $2 \times 10^{17}$  eV, the “second knee” marks a second steepening, so that  $\Gamma \approx -3.3$  up to the “ankle” at  $4 \times 10^{18}$  eV, which benchmarks a hardening of the spectrum up to  $\Gamma \approx -3$ . This last spectral index is observed up to approximately  $2 \times 10^{20}$  eV, after which there are no more statistics.

Because of the fact that their spectrum follows a (broken) power law, it can be deduced that CRs must be accelerated non-thermally [3, 26]. Up to the knee, it is expected that the CRs have a *galactic* origin, and in particular, that their energy is gained from *diffuse shock acceleration* in supernova remnants [27–31], based on an

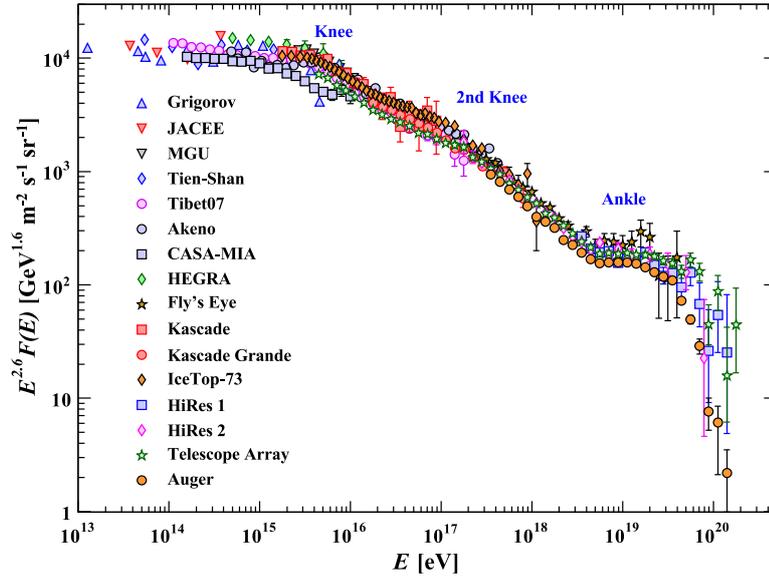


Figure 1.1: The Cosmic Ray spectrum following a broken power law, where the observed CR flux  $F(E)$  is scaled with  $E^{2.6}$  in order to clearly visualize the transitions in spectral index. In particular, it can be seen that at the ankle the spectrum hardens, possibly indicating that different, more powerful acceleration mechanisms start to dominate up to the highest energies. The different experiments that contributed to this plot are also indicated [32].

original proposal by Fermi [33]. The knee could then indicate the maximal achievable energy in the acceleration process for protons. In the case of heavier elements, this maximal energy would be proportional with their atomic number, so that the saturation of iron acceleration could explain the presence of the second knee.

Furthermore, the steepening of the CR spectrum at the ankle is a strong indication that other acceleration processes give rise to Ultra-High-Energy Cosmic Rays. Moreover, at these energies, the charged UHECRs cannot be confined by the galactic magnetic field, as the Larmor radius becomes larger than the average scale of the Milky Way. Therefore, the extremely energetic environments in which the UHECRs are accelerated most likely have an *extragalactic* origin [3, 26].

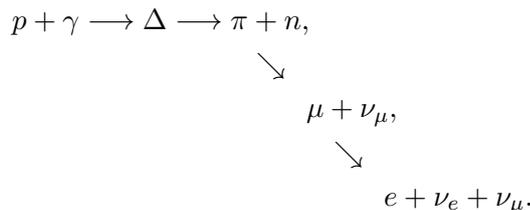
The end of the CR spectrum could be due the saturation of these extragalactic acceleration mechanisms. However, the fact that no UHECRs above this threshold are observed could also be explained by the *Greisen-Zatsepin-Kuzmin (GZK) effect* [34, 35], which states that at these energies the UHECRs interact with the Cosmic Microwave Background (CMB) through the  $\Delta$ -resonance, thereby reducing the mean free path of a proton in the Universe to a handful of megaparsecs. Moreover, heavier nuclei are expected to photo-disintegrate at these energies. Note that this is based on the current knowledge (at relatively low-energies) of the center-of-mass proton-photon interaction cross sections. The existence of the GZK limit could be disputed if more energetic UHECRs are observed, but this is unfeasible in current experiments as the UHECR flux would become too low for regular detections at these energies ( $\lesssim 1$  particle per

square kilometer per century) [3].

In any case, the deflection of (UHE)CRs by (inter)galactic magnetic fields implies that they do not point back to their source when detected at Earth. Since these magnetic fields are unknown, the origin of (UHE)CRs cannot be determined through direct observations. However, UHECRs are expected to produce non-charged high-energy cosmic neutrinos, as explained in more detail in the following Section, which will *not* be deflected. Moreover, since they interact rarely, high-energy cosmic neutrinos could be the ideal messengers of the extreme environments in which UHECRs are produced. Thus, determining the source of these high-energy cosmic neutrinos could resolve the problem of the UHECR origin.

### 1.1.2 Production of High-Energy Cosmic Neutrinos

Through hadronic interactions of the (UHE)CRs with the ambient medium at their production site, a large abundance of a.o. (charged) pions and kaons is expected to be produced, which will always give rise to neutrinos in their final decay products<sup>1</sup> [32]. Considering protons only for simplicity in this discussion—which can be generalized directly for higher elements—, one of the important interactions will be that of an energetic CR with ambient photons at the  $\Delta$ -resonance [4, 36],



Here, the most likely decay chain after the  $p\gamma$ -interaction is indicated as well; the neutron will also be able to interact hadronically or it will decay, in both cases producing an additional neutrino component.

For example, in case the ambient photon is in the UV regime, i.e. if it has an energy between 1 eV and 10 eV, then the threshold proton energy for the  $\Delta$ -resonance is about  $10^{16}$  eV (in the restframe of the ambient medium). Moreover, the high-energy cosmic muon neutrino originating directly from the pion decay will obtain about 5% of the proton energy, which would correspond with about 500 TeV in this case.

The higher the CR energy, the less the photon energy has to be in order to reach the  $\Delta$ -resonance. In particular, protons with energies above about  $10^{20}$  eV would start to be sensitive for  $p\gamma$ -interactions with photons of the CMB, assuming that currently known physics can still be applied at these energies. This is exactly the GZK effect discussed previously, which would give rise to so-called *cosmogenic neutrinos*. Note that cosmic neutrinos with energies above 100 PeV are also referred to as *Extremely-High-Energy (EHE) cosmic neutrinos*.

It should be remarked that the decay  $\Delta \rightarrow p + \pi^0$  followed by  $\pi^0 \rightarrow \gamma + \gamma$  will also occur, where two high-energy  $\gamma$ -rays result from the neutral pion decay. Consequently,  $\gamma$ -ray astronomy can also be used to search for indications of hadronic acceleration.

---

<sup>1</sup>Note that in the text no distinction will be made between particles and antiparticles, as this is not of direct relevance here and should also be clear from the context.

However,  $\gamma$ -rays are easily absorbed by (inter)galactic dust in contrast with high-energy cosmic neutrinos. Nevertheless,  $\gamma$ -ray astronomy is currently used in combination with neutrino astronomy in so-called *multi-messenger* studies.

Another important interaction is that of the CRs with nuclei from ambient gas or dust [4]. Again considering only protons, the multiplicity of cosmic neutrinos originating from  $pp$ -interactions is expected to be higher than in the case of  $p\gamma$ -interactions [37], especially at relatively low energies, as indicated in Fig. 1.2. This can be explained by the fact that a large amount of  $\pi$ - and  $K$ -mesons will be produced in the  $pp$ -interactions, where each of them will then obtain a relatively small fraction of the original CR energy.

After their production, high-energy cosmic neutrinos can travel directly towards Earth, where they can then be observed. Note that because of neutrino oscillations, high-energy cosmic neutrinos are expected to reach Earth with equal abundances in the three known neutrino flavours [38]. However, detecting high-energy cosmic neutrinos is a non-trivial task, as there is a relatively low probability for a high-energy cosmic neutrino to interact within a detector due to their low cross sections and low flux. Therefore, in order to observe high-energy cosmic neutrinos on a regular basis, such a detector must have a large enough volume, as is the case for the 1 km<sup>3</sup> ICECUBE neutrino telescope.

## 1.2 The ICECUBE Neutrino Observatory

### 1.2.1 Detection Principle and Event Topologies

Located deep in the Antarctic ice at the geographic South Pole, the ICECUBE detector [39], depicted in Fig. 1.3, is an expansion of the former AMANDA-II experiment [40]. It consists of 86 hexagonally spaced vertical strings that reach depths of about 2.5 km, each of which contain 60 *Digital Optical Modules* (DOMs), providing for a total of 5160 sensors.

These DOMs are essentially photon multipliers, which are designed to capture singular photons originating from Čerenkov radiation. This is emitted by highly energetic charged particles that traverse the ice within the detector. The Čerenkov photons will be observed in optical and UV frequencies, for which the South Pole ice is very transparent [41]—hence the choice for the location of ICECUBE. The depth of the DOMs ensures that background light from e.g. the sun is strongly suppressed, as well as secondary particles from atmospheric air showers induced by Cosmic Rays. The detection of a clear signature by the DOMs (see below), which cannot be produced by background light, is called an *event*.

When a high-energy cosmic neutrino  $\nu_\ell$  interacts weakly with a nucleus  $N$  of the ice near the detector, it will be through the exchange of a  $W$ -boson or a  $Z$ -boson,

$$\begin{aligned}\nu_\ell + N &\xrightarrow{W} \ell + X, \\ \nu_\ell + N &\xrightarrow{Z} \nu_\ell + X,\end{aligned}\tag{1.1}$$

where  $\ell$  is the corresponding lepton and  $X$  is a fragmenting hadronic state. These are called Charged Current (CC) interactions and Neutral Current (NC) interactions,

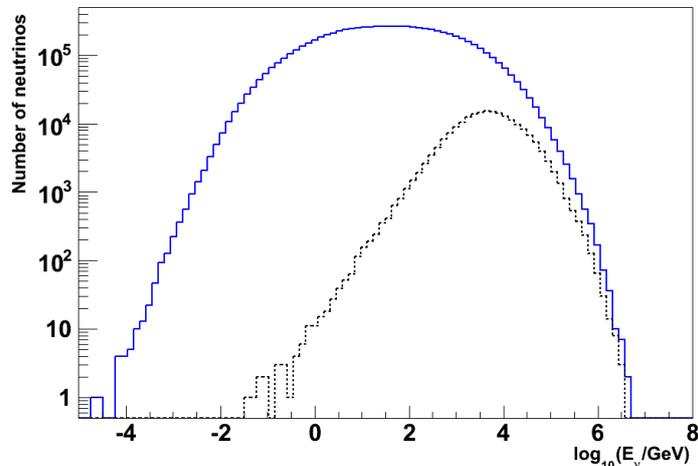


Figure 1.2: The simulated production of cosmic neutrinos in  $pp$ -interactions, indicated by the full blue line, and in  $p\gamma$ -interactions, denoted by the black dashed line, under the assumption that 80% of the CR interactions occur in the  $pp$ -channel, and 20% in the  $p\gamma$  channel. It is clear that more neutrinos are produced in  $pp$ -interactions, especially at energies below 1 TeV [37].

respectively. In both scenarios, the Čerenkov light emitted by the residual energetic charged particles when they pass through the detector is what will be observed by the DOMs.

The fragmenting hadronic state  $X$  will develop itself as a *hadronic shower* [32] in the detector. The corresponding *topology* is called a *cascade*, which is shown in Fig. 1.4a. If this cascade is fully contained within the detector, then a good estimation can be performed of the energy of  $X$ . Note that this will be a fraction of the original  $\nu_\ell$  energy, which remains unknown. Yet, the arrival direction of the  $\nu_\ell$  is not well-determined because of the “blob” structure of the cascade; the achievable angular resolution is about  $10^\circ$  for  $\nu_\ell$  energies above 100 TeV, and worse for lower energies.

On the other hand, if a CC interaction occurs, the lepton flavour will determine how the event manifests itself in the detector; the contribution of  $X$  in the topology remains the same. For the scenario that  $\ell = e$ , the electron will induce an *electromagnetic shower* [32] on top of the hadronic shower<sup>2</sup>. Consequently, the topology will also be a cascade. Although for this particular CC interaction the full energy of the original  $\nu_e$  is deposited in the detector, it can not be distinguished topologically from any  $\nu_\ell$  interacting via the NC.

Furthermore, if  $\ell = \mu$  the produced muon will lose its energy mostly due to stochastic losses and ionization; no second electromagnetic shower is induced. Moreover, because of its relatively long lifetime, it will be able to traverse the whole detector before decaying, leaving behind a *track* of Čerenkov light, as depicted in Fig. 1.4b. The energy

<sup>2</sup>Note that the hadronic shower will produce its own electromagnetic component through the production of neutral pions.

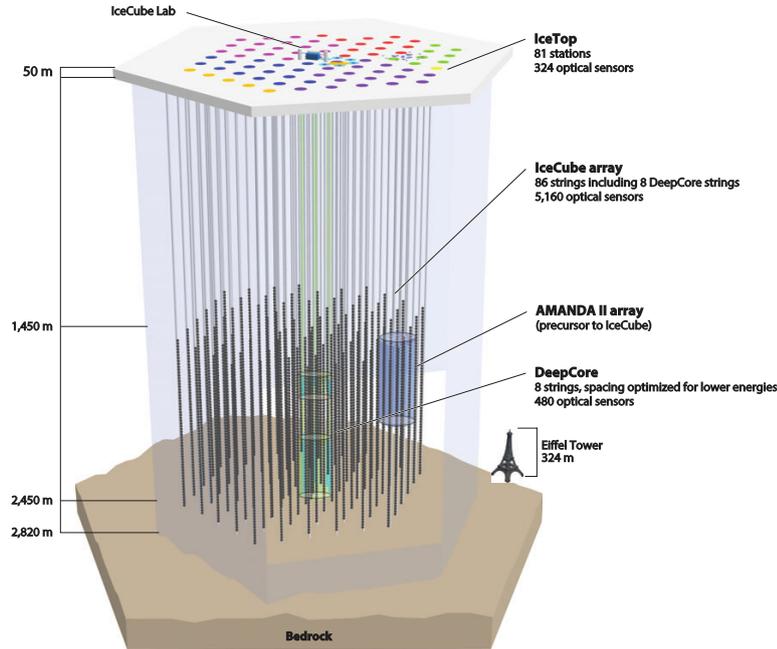


Figure 1.3: A schematic of the hexagonal ICECUBE neutrino observatory, visualizing the scale of the 86 strings, of which the DOMs are also shown, w.r.t. the Eiffel Tower. The DEEPCORE component of ICECUBE and the ICETOP array are indicated, as well as the AMANDA-II detector which was the precursor of ICECUBE. The ICECUBE lab is where the online triggering and data acquisition takes place [42].

of the  $\nu_\mu$  is poorly determined because of the fact that the muon is not contained in ICECUBE. However, due to this topology, for  $\nu_\mu$  energies above about 200 GeV the arrival direction of the  $\mu$ , and thus of the  $\nu_\mu$ , can be determined with a resolution  $\lesssim 1^\circ$ . As a consequence, muon neutrinos play a vital role in the search for the origin of high-energy cosmic neutrinos.

Lastly, the topology for the case  $\ell = \tau$  would depend on the energy of the  $\nu_\tau$  due to the short lifetime of the  $\tau$ . For relatively low energies, only a cascade would be observed, making the  $\nu_\tau$  indistinguishable from a  $\nu_e$  interacting via the CC and any  $\nu_\ell$  interacting via the NC. At energies above approximately 6 PeV, the  $\tau$  would be able to travel a short yet detectable distance in the ice, thereby producing a short track just as in the muon case. Afterwards, the  $\tau$  could decay leptonically, producing a  $\mu$  which would create a second, brighter track, or it could decay hadronically, inducing a second cascade. Thus, a  $\nu_\tau$  with a sufficient energy would also result in a good angular resolution and a distinctive topology. However, no (astrophysical) tau-flavoured neutrinos have been identified with ICECUBE so far [44].

Note that it can occur that the CC interaction for  $\ell = \mu$  does not occur inside the detector, but that the muon track is still observed as it propagates through it. Such an event becomes harder to distinguish from the main background observed with ICECUBE, which are atmospheric muons originating from Cosmic Ray air showers detected at a rate of about 2.8 kHz [45]. Therefore, typically a veto layer is constructed

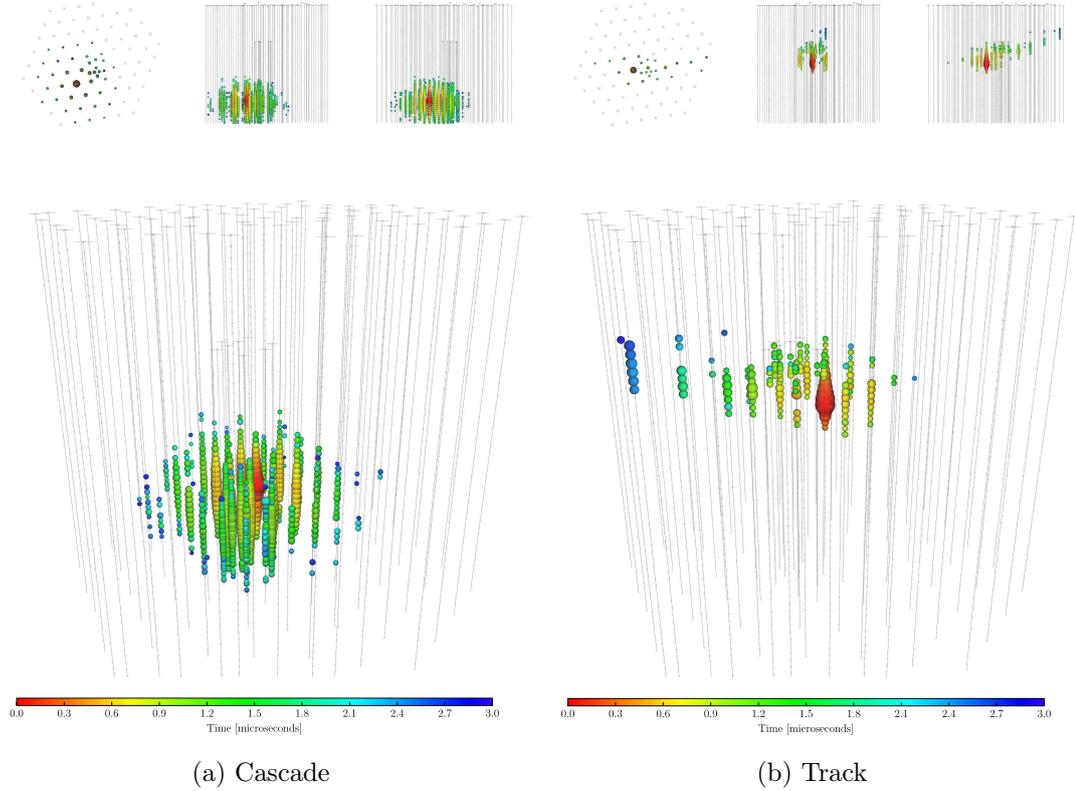


Figure 1.4: The two main topologies observed in ICECUBE, showing the temporal development of the Čerenkov radiation detected by the DOMs. First, a cascade is shown in Fig. 1.4a, where it can be seen that it is almost fully contained in the detector, meaning that the deposited energy can be estimated relatively well. Second, a track induced by a muon neutrino is plotted in Fig. 1.4b, clearly indicating that its arrival direction can be determined with a relatively high resolution; the cascade originating from the fragmenting hadronic state produced in the CC interaction can also be distinguished here. In both cases the depicted event was most likely produced by a high-energy cosmic neutrino [43].

from the outer strings in order to only analyze events (both tracks and cascades, the latter for energy confinement) that start inside the detector, as shown in Fig. 1.4.

### 1.2.2 DEEPCORE and ICETOP

With the whole ICECUBE detector, good track reconstructions can be performed for energies above about 200 GeV, as mentioned previously. However, a lower threshold is required for studies concerning a.o. dark matter, solar flares, and neutrino oscillations. For that purpose, six of the eight most central strings of ICECUBE have both a smaller horizontal spacing and a smaller vertical DOM spacing, so that the track reconstruction can be performed down to energies of the order of 10 GeV. These eight central string comprise the so-called DEEPCORE component of ICECUBE [46], shown in Fig. 1.3.

Additionally, Fig. 1.3 also indicates the 1 km<sup>2</sup> ICETOP array on the surface of ICECUBE [47], which consists of 162 Čerenkov ice tanks for the study of Cosmic Ray air showers. These studies have contributed to the determination of the CR spectrum shown in Fig. 1.1 [21]. Because it detects the atmospheric muons originating from the air showers, ICETOP is also used as a veto to suppress the atmospheric muon background in ICECUBE analyses.

### 1.2.3 Determination of a Neutrino Flux from an Event Rate

The conversion of a recorded neutrino<sup>3</sup> event rate  $\dot{N}_{\text{det}}$ , representing an actual measurement or a constraint such as an upper limit, into a (differential) flux  $\Phi_\nu$  is a non-trivial task, since a variety of factors have to be taken into account. These not only include aspects of the ICECUBE detector, such as detection and reconstruction efficiencies, but also inherent physical quantities, such as the cross sections of the  $\nu_\ell N$ -interactions given in Eqs. 1.1. All this information is contained in the so-called *effective area*  $A_{\text{eff}}$  of ICECUBE.

The effective area  $A_{\text{eff}}(E, \theta)$  depends on both the neutrino energy  $E$  and the ICECUBE zenith angle  $\theta$ ; it has no azimuthal dependency due to the geometry of ICECUBE (see Fig. 1.3). As a consequence, the rate  $\dot{N}_{\text{det}}$  [s<sup>-1</sup>] is the convolution of  $\Phi_\nu(E, \Omega)$  [GeV<sup>-1</sup> cm<sup>-2</sup> sr<sup>-1</sup> s<sup>-1</sup>] with  $A_{\text{eff}}(E, \theta)$  [cm<sup>2</sup>], where  $\Omega$  denotes the solid angle dependency,

$$\dot{N}_{\text{det}} = \int A_{\text{eff}}(E, \theta) \Phi_\nu(E, \Omega) dE d\Omega. \quad (1.2)$$

Thus, in order to determine  $\Phi_\nu$  from a certain  $\dot{N}_{\text{det}}$ , a deconvolution of Eq. 1.2 has to be performed. The details of this process and of the determination of  $A_{\text{eff}}$ , done via Monte Carlo simulations, lie beyond the scope of this thesis, but more information can be found in [48–50].

Of particular interest is the energy-dependency of  $\Phi_\nu(E)$ , i.e. the neutrino spectrum, which can be inferred directly from an energy-binning of  $\dot{N}_{\text{det}}$  and taking into account  $A_{\text{eff}}(E)$ , averaged over zenith angle  $\theta$  for each energy bin. An example of this dependency is shown in Fig. 1.5, where  $A_{\text{eff}}(E)$  was determined assuming an equal flux of (anti)neutrinos for all flavours. The dependency is dominated by the  $\nu_\ell N$ -interaction cross sections, which are rising functions of energy [32]. As a consequence, the recorded event rate spectrum will be harder than the actual spectrum of the neutrino flux.

Different models of the spectrum can be found in the literature, see e.g. [51], but the generic choice is a power law,  $\Phi_\nu(E) \propto E^\gamma$ . In case the rate represents actual measurements, one fits the spectrum to the data, where the spectral index  $\gamma$  can be chosen as a free parameter. Contrastingly, if the rate is a constraint, the parameters of the spectrum have to be assumed a fortiori.

---

<sup>3</sup>The discussion in this Section is valid for all (anti)neutrinos, including atmospheric neutrinos and high-energy cosmic neutrinos.

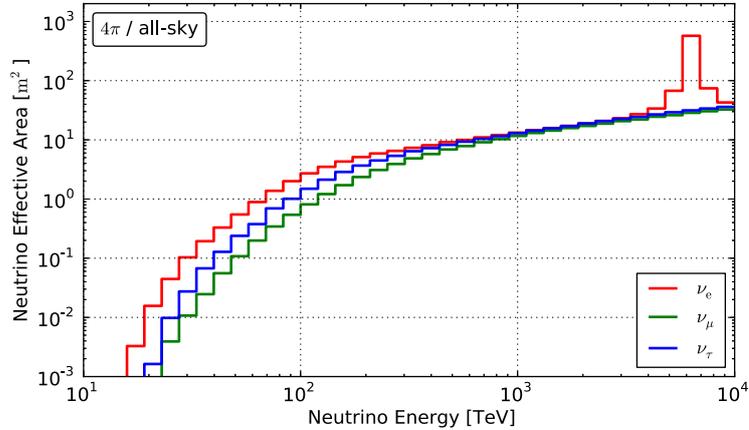


Figure 1.5: An example of the effective area  $A_{\text{eff}}$  averaged over zenith angle plotted as function of (anti)neutrino energy, for the three different flavours. The determination was performed under the assumption that the total neutrino flux has a (1:1:1) flavour abundance. It is clear that the effective area increases with energy. The peak around 6.3 PeV for the electron flavour is due to the Glashow resonance [52], resulting in an excess in  $W$ -boson production from interactions of the  $\bar{\nu}_e$  with atomic electrons [5].

## 1.3 The Birth of Neutrino Astronomy

### 1.3.1 First Observations of High-Energy Cosmic Neutrinos

The existence of high-energy cosmic neutrinos was unambiguously proven when they were discovered by the ICECUBE Collaboration in 2013 [5]. For a detector livetime of 988 days, the background-only hypothesis was rejected with a significance of  $5.7\sigma$  [43]—a detailed description of the assessment of significance is given in Chapter 2. In these three years, a total of 37 candidate high-energy cosmic neutrino events were observed with energies between 30 TeV and 2 PeV, two of which are shown in Fig. 1.4.

Generally, in high-energy cosmic neutrino studies with ICECUBE, the main background component is due to the aforementioned atmospheric muons. Furthermore, atmospheric neutrinos that also originate from Cosmic Ray air showers provide a second important contribution to the background [53].

For the analysis discussed here of a three year detector livetime, the measured data is shown in Fig. 1.6 together with these two contributions from the background. It is clear that for deposited event energies above 100 TeV the measurements cannot be described by the background alone. Two spectral power law fits were performed on the data, one with the spectral index fixed as  $\gamma = -2$ , and another where the spectral index was also determined from the fit, with a best-fit value  $\gamma = -2.3 \pm 0.3$ . For the  $E^{-2}$  spectrum, the high-energy cosmic neutrino flux per flavour (including antineutrinos) resulting from this analysis is given by

$$E^2\Phi_\nu = (0.95 \pm 0.3) \times 10^{-8} \text{ GeV cm}^{-2} \text{ sr}^{-1} \text{ s}^{-1}. \quad (1.3)$$

It should be noted that this is a *diffuse flux*, as the data of this analysis were obtained

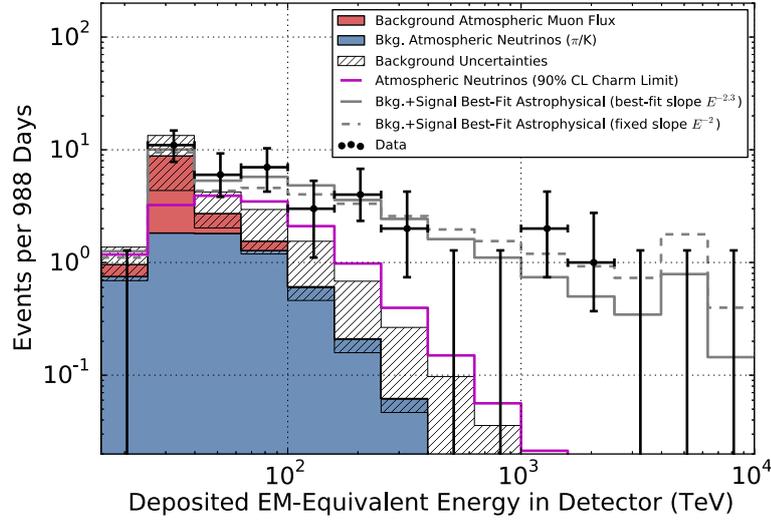


Figure 1.6: The high-energy astrophysical neutrino spectrum observed with ICECUBE after three years of performance. Above 100 TeV, the data clearly do not correspond with a pure background contribution of atmospheric muons and atmospheric neutrinos. Two power law fits of the spectrum are also shown, one with a fixed spectral index ( $\gamma = 2$ ) and one with a best-fit spectral index ( $\gamma = -2.3$ ). Note that the slopes of the fits in the plot are steepened w.r.t. their true spectral index because of the ICECUBE effective area contribution to the detected event rate (see Fig. 1.5). For further information on the composition of this plot, the reader is referred to [43].

over the whole sky without any preferred direction, in contrast with source investigations.

The discovery of high-energy cosmic neutrinos has laid the foundation for the field of neutrino astronomy. Apart from ICECUBE, there are other active neutrino observatories located in the Northern Hemisphere, such as the ANTARES and Baikal detectors [54, 55], although these have smaller detection volumes. Nevertheless, the northern 1 km<sup>3</sup> counterpart of ICECUBE, called KM3NeT [56], is currently being commissioned in the Mediterranean Sea. Furthermore, studies of (cosmogenic) EHE neutrinos with energies above several EeV are performed with the ARA, ARIANNA, and ANITA experiments [57–59]. Note that up to this date ICECUBE is the only detector with which high-energy cosmic neutrinos have been observed. It should also be mentioned that there is a proposal for an ICECUBE expansion with a volume of 10 km<sup>3</sup> [60, 61], in order to obtain high-energy cosmic neutrino statistics up to energies exceeding 10 PeV.

### 1.3.2 Status of Steady Point Source Searches with ICECUBE

Ever since the first measurement of the diffuse high-energy cosmic neutrino flux given in Eq. 1.3, several candidate sources, including extended sources, time-dependent point sources such as Gamma Ray Bursts (GRBs), and steady point sources such as Active Galactic Nuclei, have been studied with the purpose to find the origin of this flux. However, so far none of these investigations have resulted in a source determination.

As steady point sources are the objects of interest in this thesis, only the current status of these particular searches will be shortly discussed here. This overview is mainly based on [62], where the results of steady point sources searches with seven years of ICECUBE data are presented in detail. For more information about the analyses of the other source types, the reader is referred to [63–66]. Also, it is mentioned here that in the following, several statistical concepts will be encountered that are rigorously defined in Chapters 2 and 4.

The analysis with seven years of ICECUBE data focused exclusively on high-energy astrophysical muon neutrinos, because of the  $\lesssim 1^\circ$  resolution obtained on their arrival direction based on their track reconstruction. The Northern and Southern hemisphere were treated separately, because atmospheric muons from the Northern Hemisphere are strongly suppressed by the Earth, resulting in a large reduction of the background.

Apart from individual candidate source analyses, that lead to non-significant results, a scan of the full sky was performed using an unbinned maximum likelihood method with an isotropic background as the null hypothesis. The sky was gridded isotropically in patches smaller than the angular resolution, and for each grid segment, a value of the likelihood test statistic was determined from the data with a corresponding pre-trial  $p$ -value, which resulted in the all-sky map given in Fig. 1.7.

In the analysis, the lowest pre-trial  $p$ -values in the all-sky map, also called hotspots, were found to have values  $p = 1.82 \times 10^{-6}$  in the Northern Hemisphere, and  $p = 0.93 \times 10^{-6}$  in the Southern Hemisphere. However, these pre-trial  $p$ -values do not represent a final assessment of significance since one has to correct for the fact that multiple trials ( $1.9 \times 10^5$  per hemisphere) of the likelihood maximization were performed—one for each segment of the grid. Taking this into account resulted in post-trial values  $p = 29\%$  and  $p = 17\%$  for the Northern and Southern hemisphere hotspots, respectively. Conclusively, no significant excesses over the background-only hypothesis were observed in the all-sky search.

Furthermore, an analogous analysis was made of 74 individual galactic and extragalactic steady point source candidates of different types. Also here the results corresponded with a background-only expectation, as the lowest post-trial  $p$ -values were 37% and 9.3% in the Southern hemisphere<sup>4</sup>, and 54% in the Northern hemisphere.

Since no significant excesses were found over the expected background in both analyses, constraints have been put on the high-energy cosmic neutrino flux per source, denoted here as  $d\phi/dE_\nu$  with  $E_\nu$  the neutrino energy, and where an  $E_\nu^{-2}$  spectrum is assumed<sup>5</sup>. These constraints were constructed as confidence upper limits at 90% Confidence Level, which are shown in Fig. 1.8 for all declinations  $\delta$ .

The upper limits were obtained for each of the candidate steady point sources, and also for the two hotspots encountered in the all-sky search. Moreover, for the latter analysis, at each declination  $\delta$  the upper limit of the flux was determined such that if a source at that  $\delta$  would correspond with a flux higher than this limit, then in 90% of

---

<sup>4</sup>The reason why two southern  $p$ -values are mentioned here is because in the analysis the 74 sources were split into two categories, where one contained only southern sources, and the other included objects from both hemispheres. Hence, per category the trial correction was performed for each hemisphere.

<sup>5</sup>For constraints on other models, see the original paper [62].

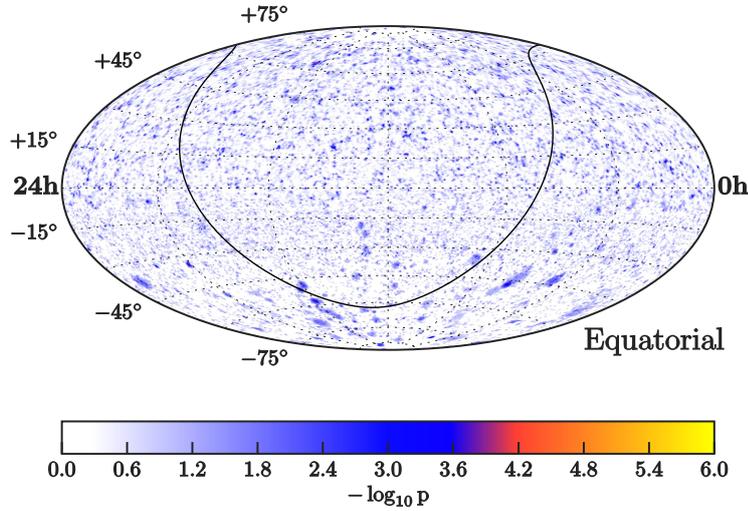


Figure 1.7: The map of the all-sky steady point source search obtained with seven years of ICECUBE data, depicted in equatorial coordinates (the equinox corresponds with a J2000 epoch). The pre-trial  $p$ -value is shown as  $-\log_{10} p$  for each segment of the gridded sky, from which it can be seen that no significant excesses were found w.r.t. the background-only hypothesis. The full black line in the plot indicates the plane of the Milky Way [62].

the experiments it would be detected with a significance larger than the one from the hotspot in the corresponding hemisphere.

In summary, no steady point sources have been found so far that are responsible for the diffuse high-energy cosmic neutrino flux measured with ICECUBE. Using a simple argumentation given in [7], one can obtain an estimate of the steady point source populations that could still account for the observed diffuse flux.

Let  $\rho$  be the population density of a specific source type, and  $L$  its characteristic (high-energy) electromagnetic luminosity. Then  $L\rho$  roughly represents the total emission per volume of this source population, which, in order to be a candidate source, should trespass the emission density corresponding with the measured diffuse flux. If one takes  $d = (4\pi\rho)^{-1/3}$  as a characteristic distance for the nearest source of the population, then one can estimate the neutrino flux  $\phi_{\nu}^{\text{ps}} = L/d^2$  of this nearest source according to  $\phi_{\nu}^{\text{ps}} \propto \rho^{-1/3}$ . Here it is assumed that a direct correlation is present of high-energy electromagnetic emission and high-energy neutrino production.

Further details behind this estimation lie beyond the scope of this thesis, but they can be found in [7]. The results of this analysis are shown in Fig. 1.9, where also transient point sources (GRBs and different supernova types) are included, but these will not be discussed here. Note that this plot was made with limits obtained using four years of ICECUBE data, although the difference with the seven year upper limits are not large enough to cause a significant influence on the shown plot.

There are six (extragalactic) steady point source populations considered in this work, i.e. Flat Spectrum Radio Quasars (FSRQs), BL Lacertae (BL Lac), Fanaroff-

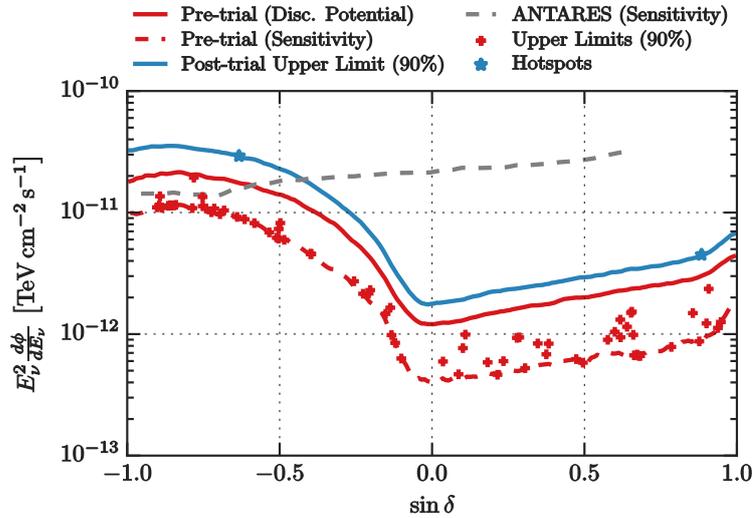


Figure 1.8: The 90% CL upper limits on the flux  $d\phi/dE_\nu \propto E_\nu^{-2}$  per source determined from the analyses of the full sky and of the 74 individual candidate steady point sources using seven years of ICECUBE data. The constraints are given for all candidate sources, and also of the all-sky hotspots, from which the all-sky upper limits at other declinations  $\delta$  (full blue line) are determined as described in the text. Also shown are the  $5\sigma$  Discovery Potential and sensitivity determined for all declinations before the unblinding of the data; as a comparison, the sensitivity of the ANTARES experiment is indicated as well. It can be observed that stronger constraints are obtained in the Northern Hemisphere, which is due to the fact that ICECUBE suffers less from atmospheric muons produced in the northern sky. This effect is also seen in the ICECUBE Discovery Potential and sensitivity [62].

Riley galaxies of types I and II (FRI resp. FRII), galaxy clusters, and starburst galaxies. The first four sources are different types of AGN [6]. What can be seen from Fig. 1.9 is that objects with relatively low population densities and high characteristic luminosities are disfavoured to solely account for the observed diffuse neutrino flux. However, FRI galaxies, starburst galaxies, and galaxy clusters, which have relatively low luminosities and high population densities, are not constrained by this estimation.

It should be noted that e.g. starburst galaxies could have significant electromagnetic obscuration at the highest energies due to the gas clouds in which the star formation occurs. This could be an indication that neutrino production is significantly enhanced through  $pp$ -interactions [37] (see Section 1.1.2). Moreover, hints for such obscured sources as the possible origins of high-energy cosmic neutrinos also arise from  $\gamma$ -ray observations with the Fermi-LAT satellite [67]. Therefore, the high-energy cosmic neutrino flux corresponding with such an obscured source might be larger than the estimated  $\phi_\nu^{\text{ps}}$  given here.

Finally, it can be concluded that if steady point sources are accountable for the diffuse high-energy cosmic neutrino flux measured in ICECUBE, they likely form part of a population with a high source density but low neutrino emission rate. As a conse-

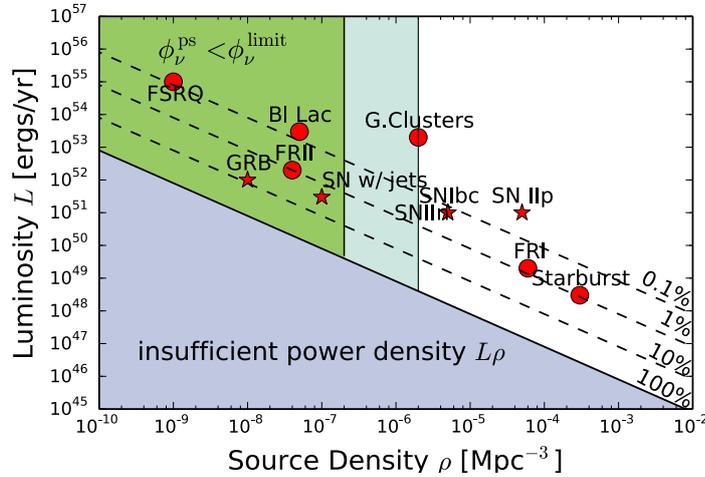


Figure 1.9: An estimative depiction of several transient and steady point source populations, based on their population density and characteristic electromagnetic luminosity, that have sufficient power densities to possibly be responsible for the diffuse high-energy astrophysical neutrino flux observed with ICECUBE. The diagonal percentages indicate the fraction of the corresponding power density that would be required to explain the diffuse flux. The green region illustrates which sources are disfavoured by the flux upper limit  $\phi_{\nu}^{\text{limit}}$  determined with four years of ICECUBE data, based on an estimate of the flux  $\phi_{\nu}^{\text{ps}}$  for a characteristic distance  $d = (4\pi\rho)^{-1/3}$  to the nearest source of each population. The blue region is the analogous limit for transient point sources; for more information on their interpretation, see [7].

quence, maximum likelihood analyses, such as the one discussed previously, of individual sources which are expected to have signatures that are hardly distinguished from background, might not be the most appropriate.

It is thus possible that other statistical methods perform better than the likelihood for such low signal event rates. This is the main motivation for the work presented in this thesis. In the following Chapters, the necessary statistics—of which several concepts have been used in this Section—will be introduced in order to perform a comparison of four statistical methods (including the likelihood) using a toy model of steady point sources.

# Statistics in Neutrino Astronomy

## Introduction

In experimental physics, measured data is analyzed through statistical methods in order to claim discoveries or set constraints on theoretical models. However, such a dataset will in general not exclusively contain information that is useful for the purpose of the experiment. The data will be a combination of *signal*, i.e. the information that is relevant to the tested model, and *background*, which are the remaining measurements. In a source analysis with ICECUBE for example, the signal would correspond with the neutrinos originating from the source (if they are present), while the background would mainly consist of atmospheric muons and atmospheric neutrinos from cosmic ray air showers.

Since the background typically dominates the dataset, it is a non-trivial task to search for a possible signal signature. In terms of statistics, the distribution of the full dataset, which contains signal and background, has to be distinguished from the background-only distribution, where the contribution of signal to the former is relatively small. This Chapter will cover the techniques to analyze such a dataset in the context of neutrino astronomy. First, a short overview of some relevant statistics is given in Section 2.1. Then, the statistical concepts needed for an ICECUBE analysis are defined in Section 2.2, followed by some final remarks in Section 2.3.

## 2.1 An Overview of Statistics

To start, a short overview of some statistical concepts is given which are relevant for this thesis. For a more detailed discussion of the most general topics, the reader is referred to [68–70]. It should be noted that here the emphasis lies on the physical interpretation of statistics rather than on the fundamental approach of mathematics.

### 2.1.1 Probability Distribution Functions

A stochastic variable, or in short a *statistic*  $X$ , is defined as a physical quantity that, when measured multiple times, will take on a different value for each measurement. In other words, the value of a measurement of  $X$  will be random, although there is a probability  $\mathcal{P}$  that can be assigned for it to result in a value  $x$ , i.e.  $\mathcal{P}(X = x)$ . In case one could perform an infinite amount of measurements, then one would know the whole set  $\Omega$ —the sample space, which can be continuous or discrete—of possible values for  $X$ , with their probability of occurrence.

This information is contained in the *Probability Density Function* (PDF)  $f_X(x|\boldsymbol{\theta})$ , which gives the probability density that a measurement of  $X$  will result in a value<sup>1</sup>  $x \in \Omega$ , given some physics parameters  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_m)$ ;  $\Omega$  can also depend on these parameters. Thus, the probability that  $X \in A$ , with  $A \subseteq \Omega$ , is given by

$$\mathcal{P}(X \in A) = \int_A f_X(x'|\boldsymbol{\theta}) dx',$$

where the integral should be appropriately replaced by a summation in the discrete case, which will be implicitly assumed henceforth. Furthermore, since by definition  $0 \leq \mathcal{P} \leq 1$ , the PDF should be normalized to one,

$$\int_{\Omega} f_X(x'|\boldsymbol{\theta}) dx' = \mathcal{P}(X \in \Omega) = 1.$$

One can generalize this by defining PDFs of multiple statistics, but these will not be of relevance here.

The most important thing to remember is that given the true value  $\boldsymbol{\theta}_t$  of the parameters  $\boldsymbol{\theta}$ , the PDF  $f_X(x|\boldsymbol{\theta}_t)$  can be used to determine the probability that a measurement of  $X$  will result in a value  $x$ . Since physics parameters are always determined from experiment, the exact value of  $\boldsymbol{\theta}_t$  is never known exactly due to systematic and statistical uncertainties. The true PDF  $f_X(x|\boldsymbol{\theta}_t)$  is therefore a theoretical concept; in practice a set of *estimators*  $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m)$  of  $\boldsymbol{\theta}_t$  are required. One method to obtain such estimators is explained in the next Section.

### 2.1.2 Likelihoods

Let  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  be a set of independent measurements of a certain statistic  $X$  which follows a PDF  $f_X(x|\boldsymbol{\theta})$ . The *likelihood*  $\mathcal{L}(\mathbf{x}|\boldsymbol{\theta})$  is defined as

$$\mathcal{L}(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^n f_X(x_i|\boldsymbol{\theta}),$$

which could be interpreted as the plausibility that this set of measurements would be the result of an experiment, given  $\boldsymbol{\theta}$ . However, the reasoning can be inverted: given  $\mathbf{x}$ ,  $\mathcal{L}(\boldsymbol{\theta}|\mathbf{x})$  is the likeliness that  $\boldsymbol{\theta}$  corresponds with these measurements. This is the interpretation that is mostly used in practice. In the following, the distinction between the interpretations will be implied by the used notation, i.e.  $\mathcal{L}(\boldsymbol{\theta}|\mathbf{x})$  or  $\mathcal{L}(\mathbf{x}|\boldsymbol{\theta})$ .

<sup>1</sup>To avoid initial confusions, in this Section a statistic will be denoted by a capital letter, e.g.  $X$ . The corresponding variable will then be labeled by the same, uncapitalized letter, e.g.  $x$ .

As mentioned previously, in experimental physics one can only estimate  $\theta_t$ . Given a set of measurements  $\mathbf{x}$ , one can e.g. obtain the estimators  $\hat{\theta}$  that best fit the observations. In other words, those estimators  $\hat{\theta}$  are to be found which maximize the likelihood, i.e. the Maximum Likelihood (ML) estimators  $\hat{\theta}_{\text{ML}}$ . In practice, the natural logarithm of the likelihood is used,

$$\ln \mathcal{L}(\boldsymbol{\theta}|\mathbf{x}) = \sum_{i=1}^n \ln f_X(x_i|\boldsymbol{\theta}),$$

and since the logarithm is a monotonic function,

$$\ln \mathcal{L}(\hat{\boldsymbol{\theta}}_{\text{ML}}|\mathbf{x}) = \max_{\boldsymbol{\theta}} \{\ln \mathcal{L}(\boldsymbol{\theta}|\mathbf{x})\}.$$

### 2.1.3 Classical Confidence Intervals and Confidence Limits

Since  $\hat{\theta}_j = g_j(\mathbf{x})$  with  $j \in \{1, 2, \dots, m\}$ , where each  $g_j(\mathbf{x})$  is a function of the measurements  $\mathbf{x}$ , the estimators are statistics themselves, and this gives rise to their statistical uncertainty. As a consequence, one needs to set up a degree of confidence or *Confidence Level* (CL)  $\alpha$  with which  $\theta_t$  is estimated, where  $0 \leq \alpha \leq 1$  is usually expressed as a percentage. Here, the procedure of Neyman<sup>2</sup> [71] will be followed to construct so-called *confidence intervals*, which follows a *frequentist* or *classical* approach. For simplicity, let  $m = 1$  such that the index  $j$  can be omitted; the upcoming discussion can be generalized directly for  $m > 1$ .

Let  $Y$  denote the estimator statistic following a PDF  $f_Y(y|\theta)$  which is assumed to be known. In the frequentist approach, one can, for a value of the theoretical parameter  $\theta$ , obtain a statement about the probability  $\mathcal{P}(Y|\theta)$  of  $Y$  being in a specified interval through the integration of  $f_Y(y|\theta)$ . In particular, for the true value  $\theta_t$  and a value  $y_0 = \hat{\theta} = g(\mathbf{x})$  determined from a set of measurements  $\mathbf{x}$ , one could determine what the probability  $\mathcal{P}(y_0|\theta_t)$  is that the measured  $y_0$  corresponds with the true value  $\theta_t$  (for a certain interval of  $y$ ). However, since the PDF of the statistic  $\Theta$  is not known, one can *not* obtain a statement about the probability  $\mathcal{P}(\theta_t|y_0)$  of  $\Theta = \theta_t$  lying in a certain interval of  $\theta$ , given the measured  $y_0$ .

The Neyman method works around this problem by finding an *acceptance interval*  $[y_1(\theta), y_2(\theta)]$ , typically simply connected, for each  $\theta$  such that

$$\mathcal{P}(y_1 \leq Y \leq y_2|\theta) = \int_{y_1}^{y_2} f_Y(y'|\theta) dy' = \alpha, \quad (2.1)$$

with the complementary conditions

$$\begin{aligned} \mathcal{P}(Y < y_1|\theta) &= \int_{-\infty}^{y_1} f_Y(y'|\theta) dy' = \beta, \\ \mathcal{P}(Y > y_2|\theta) &= \int_{y_2}^{+\infty} f_Y(y'|\theta) dy' = 1 - \alpha - \beta. \end{aligned} \quad (2.2)$$

---

<sup>2</sup>In the literature, the CL can alternatively be defined as  $1 - \alpha$ , following the original notation by Neyman.

In practice, it is commonly chosen to set  $\beta = (1 - \alpha)/2$ . Note that these conditions are arbitrary and in principle free to choose, as long as they are not influenced by the measured  $y_0$ , which will be explained below in more detail. For a fixed  $\alpha$ ,  $y_1(\theta)$  and  $y_2(\theta)$  form a so-called *confidence belt*  $\{[y_1(\theta), y_2(\theta)]\}$ , as depicted in Fig. 2.1a. Each horizontal line thus represents the containment of a probability  $\alpha$  within the belt, for the corresponding value of  $\theta$ .

In an experiment, one obtains a value  $y = y_0$ , which is drawn as a vertical line in Fig. 2.1a. The confidence interval  $I$  is now obtained by taking the union of all  $\theta$ -values which correspond with acceptance intervals intersected by the vertical line. In Fig. 2.1a the confidence interval is simply connected,  $I = [\theta_1, \theta_2]$ , although this is not required. In general,  $I$  is a *two-sided confidence interval*, and in the case that  $\beta = (1 - \alpha)/2$ , it is also called a *central confidence interval*.

It is imperative to realize that the construction of a confidence interval  $I$  does *not* imply that  $\mathcal{P}(\theta_t \in I | y_0) = \alpha$ . This would require the PDF of  $\Theta$  which is not known. In fact, if  $I$  is obtained at a CL of  $\alpha$ , it means that if the experiment is repeated multiple times, the resulting  $I$  will contain  $\theta_t$  in a fraction  $\alpha$  of all experiments, on average. Furthermore, it is important to note that the only reason why this claim can be made is because of the fact that each acceptance interval contains a probability  $\alpha$ .

Thus, if  $\alpha = 90\%$  and the experiment is repeated 100 times, then on average 90 of the 100 obtained confidence intervals will contain the true value  $\theta_t$ . That is what is meant in the case that only one  $I$  is e.g. published (typically there is only one experiment), and one states that this  $I$  was obtained at 90% CL.

Besides two-sided confidence intervals, one can also define *confidence limits*, i.e. *one-sided confidence intervals*. An upper limit with CL  $\alpha$  can be obtained by setting  $y_2 = +\infty$  for all  $\theta$  in Eq. 2.1,

$$\mathcal{P}(Y \geq y_1 | \theta) = \int_{y_1}^{+\infty} f_Y(y' | \theta) dy' = \alpha,$$

with the reduction of Eqs. 2.2 to the single condition

$$\mathcal{P}(Y < y_1 | \theta) = \int_{-\infty}^{y_1} f_Y(y' | \theta) dy' = 1 - \alpha.$$

For a  $y_0$  from an experiment, one again takes the union of the  $\theta$ -values that correspond with acceptance intervals  $[y_1(\theta), +\infty)$ , as shown in Fig. 2.1b. This union is the one-sided confidence interval  $I = (-\infty, \theta_2]$  such that the confidence upper limit is  $\theta_2$ —in the simply connected case, although otherwise one can still find an upper  $\theta_2$ .

Also here the confidence upper limit at a CL  $\alpha$  does *not* mean that there is a probability  $\alpha$  for  $\theta_t$  to be smaller than  $\theta_2$ . The correct statement is that for multiple repetitions of the experiment, the resulting upper limit  $\theta_2$  will be larger than the true  $\theta_t$  in a fraction  $\alpha$  of the experiments, on average. Note that the construction of a confidence lower limit is completely analogous, by setting  $y_1 = -\infty$  in Eq. 2.1.

Terminologically, the constructed  $I$  has the correct *coverage* if

$$\mathcal{P}(Y \in [y_1(\theta), y_2(\theta)]) = \alpha, \quad \forall \theta \in I.$$

In other words, each  $\theta$  covers its corresponding acceptance interval  $[y_1(\theta), y_2(\theta)]$  at the stated CL. In case

$$\exists \theta \in I : \mathcal{P}(Y \in [y_1(\theta), y_2(\theta)]) < \alpha \text{ or } \mathcal{P}(Y \in [y_1(\theta), y_2(\theta)]) > \alpha,$$

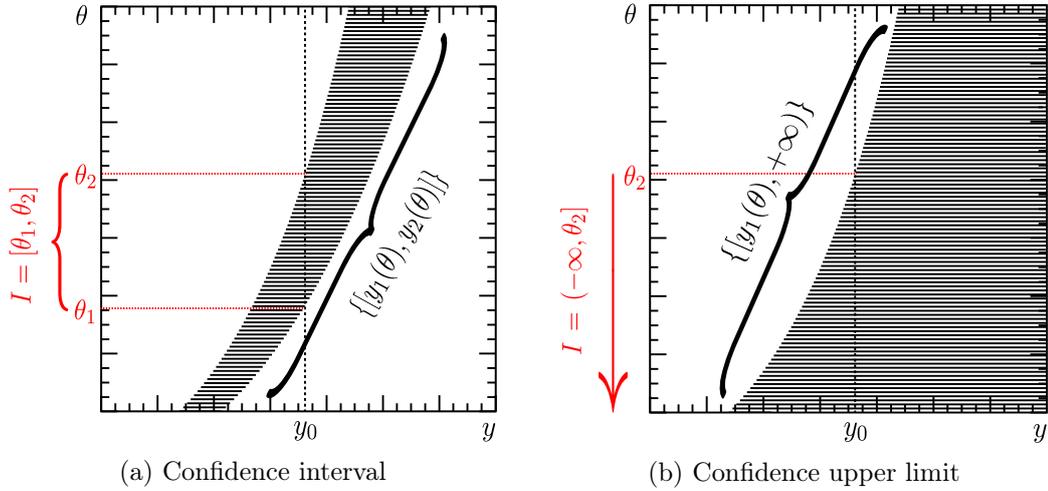


Figure 2.1: An illustrative template for the construction of a central confidence interval, shown in Fig. 2.1a, and a confidence upper limit, depicted in Fig. 2.1b, following the description in the text. The respective confidence belts are also indicated; typically only the functions  $y_1(\theta)$  and/or  $y_2(\theta)$  are shown without depicting each acceptance interval, which will also be done here subsequently. Adapted from [72].

one speaks of an *undercoverage* or *overcoverage*, respectively. This means that at least one  $\theta$  covers its corresponding acceptance interval with a probability less or greater than the stated CL. The former scenario has to be avoided, since then one claims to state a confidence interval with a CL of  $\alpha$ , whilst the actual CL is less than  $\alpha$ . In other words, one claims that on average a fraction  $\alpha$  of repetitions of the experiment will result in an  $I$  which contains  $\theta_t$ , whilst the average is actually less than  $\alpha$ .

In the latter case one also speaks of *conservatism*. This is unavoidable in the construction of confidence intervals of a discrete quantity: an acceptance interval for a certain  $\theta$  will generally not exactly contain a probability  $\alpha$  due to the discreteness. Since undercoverage is to be avoided, one chooses the first interval that contains a probability which trespasses  $\alpha$  as the acceptance interval. The resulting confidence interval will therefore lead to overcoverage.

Consider, for example, one of the scenarios illustrated by Feldman and Cousins<sup>3</sup> [72], where one is trying to measure a small mass  $\theta = \mu$  (in arbitrary units), and that the estimator  $Y$  of the mass follows a Gaussian distribution with a standard deviation  $\sigma = 1$ ,

$$f_Y(y|\mu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\mu - y)^2}{2}\right).$$

The confidence belts for the construction of a central confidence interval and confidence upper limit, both at 90% CL, are respectively shown in Figs. 2.2a and 2.2b. Since the true mass  $\mu_t \geq 0$ , one limits the range of  $\mu$  to positive values only, as done in the plots. This limitation does not cause an issue; if e.g. the one-sided case is considered, then

<sup>3</sup>Unfortunately, Feldman and Cousins use the confusing notation  $\mathcal{P}(\mu \in [\mu_1, \mu_2]) = \alpha$  to state that on average a fraction  $\alpha$  of the confidence intervals  $[\mu_1, \mu_2]$  contain the value  $\mu$ .

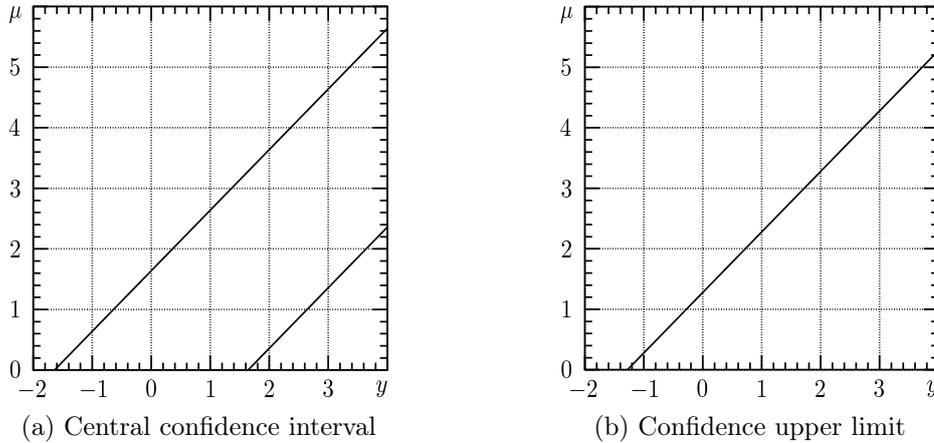


Figure 2.2: The confidence belts at 90% CL used to construct a central confidence interval, shown in Fig. 2.1a, and a confidence upper limit, depicted in Fig. 2.2b, for an estimator  $y$  which follows a Gaussian distribution with an expectation value  $\mu \geq 0$  and a standard deviation of unity. This  $\mu$  represents a mass of which the true value is estimated by  $y$ . In both cases there is a correct coverage [72].

on average 90 of the 100 determined confidence intervals  $[0, \mu_2]$  (instead of the usual  $(-\infty, \mu_2]$ ) with upper limit  $\mu_2$  will contain  $\mu_t$ , *because* of the known fact that  $\mu_t \geq 0$ .

However, it is perfectly possible that a measurement  $y_0 < 0$  is obtained. In the scenario of the upper limit, one can see from Fig. 2.2b that if  $y_0 = -1.8$  for example, the one-sided confidence interval would be empty. This does *not* mean that there is a 90% probability for  $\mu_t$  to be “nothing”, which is furthermore nonsensical. The true interpretation is that, if the experiment had been repeated 100 times, the result from the particular experiment with  $y_0 = -1.8$  means that this experiment lies in the average 10% which obtains a one-sided confidence interval that does *not* contain  $\mu_t$ .

As an attempt to resolve this, one might opt to modify the correct confidence belt of Fig. 2.1b by stating for each  $y < 0$  the upper limit one would obtain is the one corresponding with  $y = 0$ . This modification is shown in Fig. 2.3a. Nevertheless, this will unavoidably introduce conservatism. This can be seen by taking any  $\mu \in [0, 1.24]$ , for which the acceptance interval then contains the whole range of  $y$ , so the coverage of this  $\mu$  is 100% of the  $y$ -values instead of the stated 90%. Since at least a part of the  $\mu \in [0, 1.24]$  will be in the one-sided confidence interval, the stated upper limit will always lead to overcoverage.

Another “self-made” confidence belt is constructed when the following “flip-flop” reasoning is applied. If an experiments measures a  $y_0$  with a significance of at least  $3\sigma$  (see Section 2.2.1), i.e.  $y_0 \geq 3$  in the Gaussian example considered here, then a central confidence interval is obtained. On the other hand, if the significance is less than  $3\sigma$ ,  $y_0 < 3$ , then an confidence upper limit is acquired.

The resulting confidence belt can then be visualized by taking the belt of Fig. 2.2b up to  $y_0 < 3$ , and “pasting” it with the belt of Fig. 2.2a for  $y_0 \geq 3$ . This is shown in Fig. 2.3b. Now take e.g.  $\mu = 2$ . For that  $\mu$ , the acceptance interval is  $[y_1(\mu = 2), y_2(\mu = 2)] = [0.72, 3.64]$ , for which  $\mathcal{P}(Y \in [0.72, 3.64] | \mu = 2) = 85\% < 90\%$ .

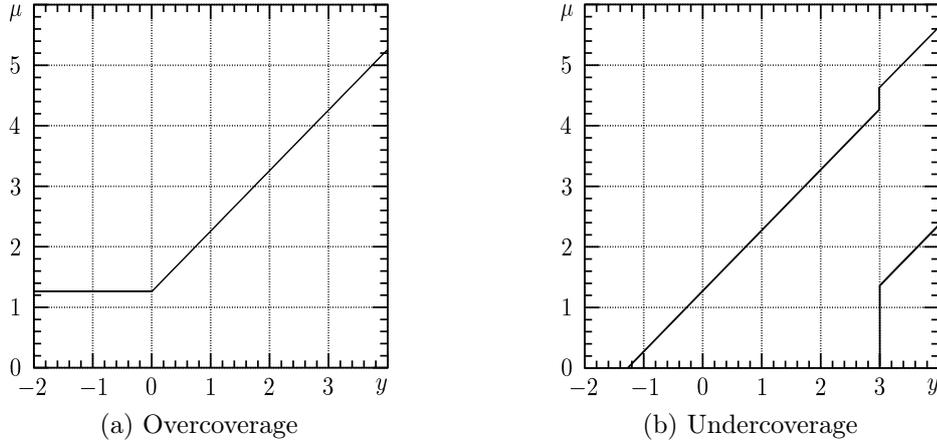


Figure 2.3: Two alterations of the correct confidence intervals that were shown in Fig. 2.2. The first, shown in Fig. 2.3a, occurs when Fig. 2.2b is truncated at a value  $y = 0$  in order to avoid empty one-sided confidence intervals. This always leads to overcoverage. Secondly, the confidence belt corresponding with a “flip-flop” reasoning is shown in Fig. 2.3b, which can be seen as the “pasting” of Fig. 2.2a with Fig. 2.2b at  $y = 3$ . Here, any confidence limit which includes a  $\mu \in [1.36, 4.28]$  will lead to undercoverage. Adapted from [72].

Thus,  $\mu = 2$  undercovers its acceptance interval. Moreover, each  $\mu \in [1.36, 4.28]$  undercovers, with a coverage of 85% instead of the stated 90%. Each stated confidence upper limit or confidence interval which contains a subset of  $[1.36, 4.28]$  will therefore lead to undercoverage.

The reason why this “flip-flopping” goes wrong, is because the conditions of Eq. 2.2 were influenced by the (anticipation of the) measurement  $y_0$ . Depending on the value of  $y_0$ , the conditions for a central confidence limit or a confidence upper limit were applied. It is exactly this altering of the conditions which introduces acceptance intervals that contain less than the stated probability of 90%. One must *first* choose a set of conditions, i.e. decide if one  $\mu$  wants to determine a central interval *or* an upper limit, *then* construct a confidence belt based on these, respectively leading to Fig. 2.2a or Fig. 2.2b, and *afterwards* look at the measurement  $y_0$  and thereby state the correct one-sided or two-sided confidence interval.

In summary, the problems in the two examples given above are that empty confidence intervals can be obtained, and that one does not know when to state a central interval or an upper limit.

Feldman and Cousins [72] resolved this through the design of an ordering principle serving as the measurement-independent conditions with which to construct acceptance intervals. If  $\theta$  is used again to generally denote a certain parameter, one first calculates for each fixed value  $y$  of the estimator the *physically allowed* value  $\theta = \theta_{\text{best}}$  which maximizes the likelihood  $\mathcal{L}(\theta|y)$ , i.e. which corresponds best with  $y$ . Subsequently, for each fixed value of  $\theta$  one calculates a likelihood ratio  $R(y|\theta)$  for all  $y$  by determining the likelihood  $\mathcal{L}(y|\theta)$ , and normalizing it with the maximized likelihood  $\mathcal{L}(\theta_{\text{best}}|y)$  at

that  $y$ ,

$$R(y|\theta) = \frac{\mathcal{L}(y|\theta)}{\mathcal{L}(\theta_{\text{best}}|y)} = \frac{f_Y(y|\theta)}{f_Y(y|\theta_{\text{best}})}.$$

Values of  $y$  are then added to the acceptance interval by starting with the largest  $R(y|\theta)$ , and continuing towards lower  $R(y|\theta)$  until the acceptance interval contains a probability  $\alpha$ . This ordering ensures that the acceptance interval for a  $\theta$  contains the values of  $y$  which are most likely to occur, given that specific  $\theta$ .

When applying the ordering principle to the Gaussian example, one first realizes that  $\theta_{\text{best}} = \mu_{\text{best}} \geq 0$ , since physically  $\mu$  represents a mass. Therefore,  $\mu_{\text{best}} = 0$  for all  $y < 0$ . The resulting confidence belt at a CL  $\alpha = 90\%$  is depicted in Fig. 2.4. For the largest  $y$ -values, this belt corresponds with the standard two-sided belt of Fig. 2.2a. Going towards smaller values of  $y$ , the lower edge of the belt smoothly crosses  $\mu = 0$ , and from that point, one would quote an upper limit or a one-sided confidence interval with a lower bound equal to zero. Furthermore, for  $y < 0$ , the upper edge of the belt never goes below zero. Thus, the ordering method does indeed resolve the two problems encountered before.

Finally, it should be mentioned that the construction of a confidence belt does not require  $Y$  to be a direct estimator of  $\theta = \theta_t$ ; this assumption was made for simplicity in the discussion. If there is any statistic  $Y$  which can be calculated as a function of the measurements  $\mathbf{x}$  and which follows a known PDF  $f_Y(y|\theta)$ , then the techniques given here can be applied to find the confidence interval of  $\theta$  for a certain measured  $Y = y_0$ .

### 2.1.4 Bayesian Credibility Intervals

As an alternative to the frequentist approach, one can use *Bayesian* statistics to obtain a statement about a certain theoretical quantity  $\theta_t$ . The discussion here will be limited, since in the rest of this thesis exclusively frequentist principles will be applied, with one exception in the description of a statistical method in Chapter 4. More information on the topic of Bayesian statistics can be found in [73–76].

Consider two propositions  $A$  and  $B$ , and some priorly known information  $I$ . The Bayesian principle finds its origin in the theorem of Bayes, which states that

$$\mathcal{P}(B|AI) = \mathcal{P}(B|I) \frac{\mathcal{P}(A|BI)}{\mathcal{P}(A|I)}, \quad (2.3)$$

where  $\mathcal{P}(B|AI)$  is the probability that  $B$  is true under the condition that  $A$  and  $I$  are true, etc. In particular, using the same notations as in the previous Section, if  $y_0$  is a measured value of the estimator of the theoretical quantity  $\theta$  with a true value  $\theta_t$ , then one can set  $A = y_0$  and  $B = \theta_t$  in Eq. 2.3,

$$\mathcal{P}(\theta_t|y_0I) = \mathcal{P}(\theta_t|I) \frac{\mathcal{P}(y_0|\theta_tI)}{\mathcal{P}(y_0|I)}.$$

The important implication of the theorem of Bayes is that if the PDF  $f_Y(y|\theta)$  is known, as is generally the case, one can determine the PDF  $f_{\Theta}(\theta|y)$  of  $\Theta$  using  $f_Y(y|\theta)$  and some prior information  $I$ . *This* is what distinguishes Bayesian statistics from classical statistics. *Credibility intervals* or *Bayesian confidence intervals* with a

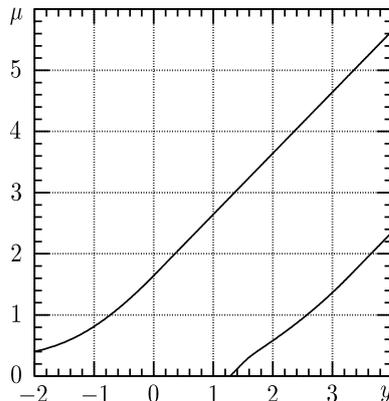


Figure 2.4: The confidence belt at 90% CL that is obtained through the usage of the ordering principle developed by Feldman and Cousins, for the Gaussian example. The two problems of empty confidence intervals and choosing between two-sided and one-sided intervals, which were unsuccessfully tackled in Fig. 2.3, are now righteously resolved [72].

certain CL  $\alpha$ —one-sided or two-sided—can then be obtained from a measured  $y_0$  by constructing a range  $[\theta_1, \theta_2]$  for which

$$\mathcal{P}(\Theta \in [\theta_1, \theta_2]) = \int_{\theta_1}^{\theta_2} f_{\Theta}(\theta' | y_0) d\theta' = \alpha.$$

Therefore, in contrast with a frequentist confidence interval, the Bayesian confidence interval *does* correspond with a range  $[\theta_1, \theta_2]$  which has a probability  $\alpha$  of containing the true  $\theta_t$ .

## 2.2 Techniques for a Statistical Analysis with ICECUBE

### 2.2.1 Test Statistics and Pseudo-Experiments

Given a set of measurements  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  of a certain stochastic physical observable  $X$ , one can define a so-called *test statistic*  $T = T(\mathbf{x})$  which can in principle be any real function of the measurements. In ICECUBE, the physical observable is an event in the detector where a.o. the energy, arrival direction, and time of detection of the neutrino are used to determine  $T$  (although the latter is of less relevance in steady sources). Since per experiment one dataset  $D$  is obtained, which contains both signal and background, only one value<sup>4</sup>  $T = T_D$  will be acquired. The purpose of this  $T_D$  is to quantify the degree of belief in the hypothesis  $\mathcal{H}_0$  that  $D$  exclusively contains background<sup>5</sup>.

First, the PDF of  $T$  has to be considered under  $\mathcal{H}_0$ , i.e. the background-only PDF  $\mathcal{T}_0(t|\boldsymbol{\theta}_0)$ , with  $\boldsymbol{\theta}_0$  a set of physical parameters (see below). The *p-value* of a certain  $T_p$

<sup>4</sup>From this point onwards, the convention from Footnote 1 will usually be omitted to avoid an excess in notations.

<sup>5</sup>Note that the discussion given here follows a frequentist approach.

w.r.t.  $\mathcal{T}_0$  is then defined as

$$p = \mathcal{P}(T \geq T_p | \mathcal{H}_0) = \int_{T_p}^{+\infty} \mathcal{T}_0(t' | \boldsymbol{\theta}_0) dt'.$$

In particular, if  $T_D = T_p$ , the probability under  $\mathcal{H}_0$  that a new experiment will result in a value of  $T$  that is larger than the original  $T_D$  is thus  $p$ . Consequently, the smaller the value of  $p$ , the less the degree of belief in  $\mathcal{H}_0$ , or the higher the *significance* of  $T_D$ . In order to claim a discovery in (astro)particle physics, i.e. to have a significance that is high enough to state that  $D$  cannot solely consist of background measurements, a value  $p = 5.73 \times 10^{-7}$  is required. This is called a  $5\sigma$  significance since it corresponds with the integral of a one-sided Gaussian PDF from  $5\sigma$  up to infinity, where  $\sigma$  is the standard deviation. In general, a certain significance is expressed<sup>6</sup> in units of  $\sigma$ .

It should be noted that here a one-sided test statistic has been considered. One could also define a two-sided (central) test statistic, and in this case the two values  $(T_1)_p$  and  $(T_2)_p$  are required to define the  $p$ -value as the probability for which

$$\begin{aligned} \frac{p}{2} &= \mathcal{P}(T \leq (T_1)_p | \mathcal{H}_0) = \int_{-\infty}^{(T_1)_p} \mathcal{T}_0(t' | \boldsymbol{\theta}_0) dt', \\ \frac{p}{2} &= \mathcal{P}(T \geq (T_2)_p | \mathcal{H}_0) = \int_{(T_2)_p}^{+\infty} \mathcal{T}_0(t' | \boldsymbol{\theta}_0) dt'. \end{aligned}$$

These are not of relevance for this thesis, but the discussion given here can directly be generalized for two-sided test statistics.

Practically, there are two manners in which the  $\mathcal{T}_0$ -PDF can be obtained through the usage of *pseudo-experiments*, which replicate the actual experiment. The first option is to simulate a set of background events in a pseudo-experiment utilizing Monte Carlo techniques, from which a value of  $T$  can be computed. One then performs many of these background-only pseudo-experiments to obtain the full PDF.

The most challenging part of this method is that in order to obtain  $\mathcal{T}_0$  up to a significance of  $5\sigma$  with enough statistical precision, one needs to perform  $\mathcal{O}(10^9)$  pseudo-experiments. Therefore, the computation time will be large. This can in first instance be remedied by performing e.g.  $\mathcal{O}(10^6)$  pseudo-experiments and fitting an expected functionality of  $\mathcal{T}_0$  through the simulated measurements. As an example, Figs. 2.5a and 2.5b show the results of the same test statistic  $T = \lambda$  (see Chapter 4 for its definition) for two different analyses, one with ICECUBE [77] and one with AMANDA-II [78]—the precursor of ICECUBE—, using  $10^9$  pseudo-experiments and the combination of  $5 \times 10^6$  pseudo-experiments with a fit, respectively.

On the other hand, one can opt to perform pseudo-experiments using experimental data of which it is known to be background, although this will depend on the analysis in question. In an ICECUBE steady point source analysis for example, one takes a certain patch on the sky around the candidate source. Typically a declination band containing this patch is then constructed. Subsequently, all (or a part) of the ICECUBE data

---

<sup>6</sup>For clarity, in the text the distinction will be made to express a  $p$ -value as a probability and a significance in units of  $\sigma$ , although they are conceptually identical. Therefore, when stating the expression “a significance of  $5\sigma$  has been obtained” for example, it is equivalent to stating  $p = 5.73 \times 10^{-7}$ . As a consequence, if  $p_1$  corresponds with  $3\sigma$  and  $p_2$  corresponds with  $5\sigma$ , then  $p_1 > p_2$ .

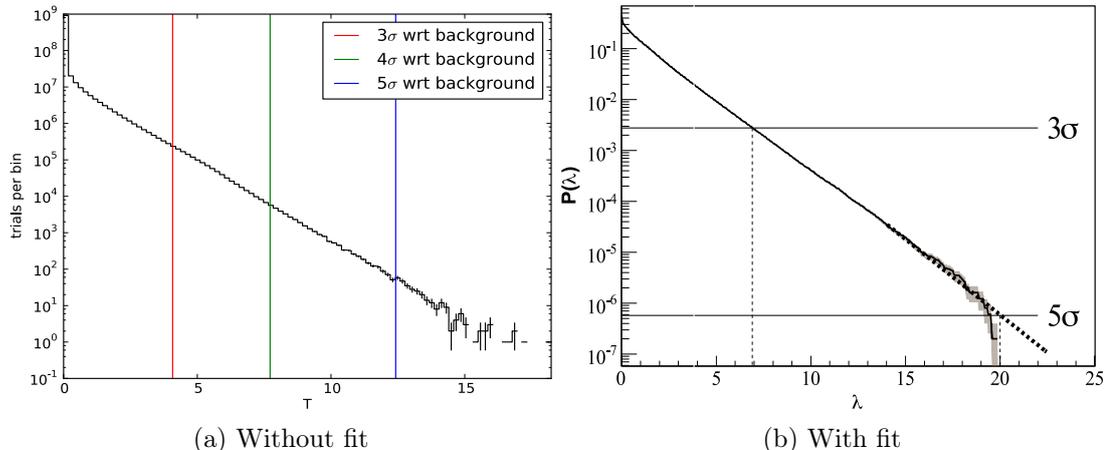


Figure 2.5: Examples of the determination of the background-only PDF  $\mathcal{T}_0$  for the test statistic  $T = \lambda$  in two different analyses, indicating some significances and corresponding  $T_p$ -values. Fig. 2.5a depicts the trial histogram of  $10^9$  pseudo-experiments, which relates directly to  $\mathcal{T}_0$  if normalized [77]. Alternatively, Fig. 2.5b shows the histogram of the integrated  $\mathcal{T}_0$ , i.e. the  $p$ -value as function of  $\lambda$ . Here, the histogram was obtained by performing  $5 \times 10^6$  pseudo-experiments, and a fit was made through it, indicated by the dashed line [78]. Note that the two plots should not be compared directly, as they correspond with values of  $T = \lambda$  for two different analyses.

within this band are scrambled, that is, the azimuthal component of an event is given a random value from a uniform distribution over  $[0, 2\pi]$  whilst keeping the declination fixed. This is plausible due to the azimuthal symmetry of ICECUBE<sup>7</sup>. The scrambling ensures that a possible signal signature in the source patch is washed away in the declination band. Therefore,  $\mathcal{T}_0$  can be obtained by scrambling multiple times—these are thus the pseudo-experiments—and determining the value of  $T$  in this patch for each scramble.

Once  $\mathcal{T}_0$  is known, one wants to determine how sensitive the test statistic  $T$  is for a signal detection and what its potential for a discovery is. To achieve this, the behavior of  $T$  in the presence of signal has to be known. Therefore, the background *and* signal PDFs  $\mathcal{T}_s(t|\theta_s)$  of  $T$  are required for different values of the signal  $s > 0$ , where the injection of signal a fortiori requires the usage of simulated pseudo-experiments (only for the signal; experimental data can still be used for the background). Fortunately, in these cases the necessary amount of simulations will be limited to  $\mathcal{O}(10^4)$ . This will become clear in the following Sections, where the concepts of Discovery Potential and sensitivity are quantified.

Note that  $\mathcal{T}_s(t|\theta_s)$  should be interpreted as  $f_T(t|s)$  in the notation of Section 2.1 for a PDF, that is, the amount of signal events  $s$  is the parameter which fully determines the PDF of the statistic  $T$ , and it has an unknown true value. The other physical

<sup>7</sup>It is a common misunderstanding to state that the location of ICECUBE at the geographic South Pole is the reason behind the azimuthal symmetry of the detector. In fact, it is because of the rotation of the Earth (along the azimuth) that this symmetry is present.

parameters  $\theta_s$  are completely defined for a certain value of  $s \geq 0$ ; their inclusion in the notation  $\mathcal{T}_s(t|\theta_s)$  is useful if the analytic expression of the PDF is known, which is generally *not* the case. In practice it is always through the performance of pseudo-experiments at a fixed  $s$  that the corresponding PDF is determined. If known a priori, the functional form of the PDF serves as a natural choice to perform the fit with for the case  $s = 0 \Rightarrow \theta_s = \theta_0$ , in order to reach a  $5\sigma$  significance (see Chapter 3 for an example).

### 2.2.2 Discovery Potential

As mentioned previously, in order to obtain a discovery a significance of at least  $5\sigma$  w.r.t. the background-only PDF  $\mathcal{T}_0$  is required. By determining the signal-plus-background PDF  $\mathcal{T}_s$  for different signal injections  $s > 0$  through the usage of pseudo-experiments, the probability to obtain a value  $T$  from  $\mathcal{T}_s$  with a significance of at least  $5\sigma$  can be determined, which will be larger the more signal is injected. For the purpose of calculating this percentage, the sufficient amount of pseudo-experiments per value  $s$  is  $\mathcal{O}(10^4)$ .

The amount of injected signal for which in  $\delta\%$  of the trials<sup>8</sup> a value  $T$  drawn from  $\mathcal{T}_s$  will have at least a  $p$ -value corresponding with  $k\sigma$ , is defined as the  $\delta\%$  *Discovery Potential* (DP) at  $k\sigma$  significance. The 50% DP at  $5\sigma$  significance, or in short, the DP<sup>9</sup>, is of most relevance in the literature, and it is shown schematically in Fig. 2.6 for the exemplary case  $T = \lambda$ . The DP is the amount of signal required so that the probability that a  $5\sigma$  discovery can be claimed from the experiment is 50%.

Preferably, the DP is as low as possible, since this indicates that less signal is required in the data  $D$  in order to claim a discovery. Therefore, an important part of an analysis is its optimization to find the minimal DP. This involves a.o. the usage of multiple test statistics and the optimization of cuts on the data resulting from Boosted Decision Tree trainings.

Since for an injected signal corresponding with the DP, in 50% of the trials one would not be able to claim a discovery, it might also be useful to determine the 90% DP at  $5\sigma$  significance. This is called the *Least Detectable Signal* (LDS). However, in practice it often occurs that a discovery is preceded by a value  $T_D$  with a lower significance in the early stages of an experiment. Thus, for a first indication of a strong excess in signal over background one might be interested in the 50% DP at  $3\sigma$  significance. As an example, the  $3\sigma$  DP,  $5\sigma$  DP, and LDS for the analysis of [78] ( $T = \lambda$ ) are depicted in Fig. 2.7.

### 2.2.3 Sensitivity

Additional to the DP, one wants to know how much signal is required such that it can be distinguished from a background fluctuation. This *sensitivity* is defined as the amount of injected signal for which in 90% of the trials a value  $T$  drawn from  $\mathcal{T}_s$  will have at most a  $p$ -value of 50%. Rephrased, for a signal injection corresponding with the

<sup>8</sup>In principle, the amount of pseudo-experiments required should be scaled with  $\delta$ , but  $\mathcal{O}(10^4)$  will suffice for the values of  $\delta$  considered in this work.

<sup>9</sup>If the  $\delta$ - and  $k$ -values are not mentioned explicitly, they are always assumed to be the ones corresponding with the 50% DP at  $5\sigma$  significance unless clear from the context.

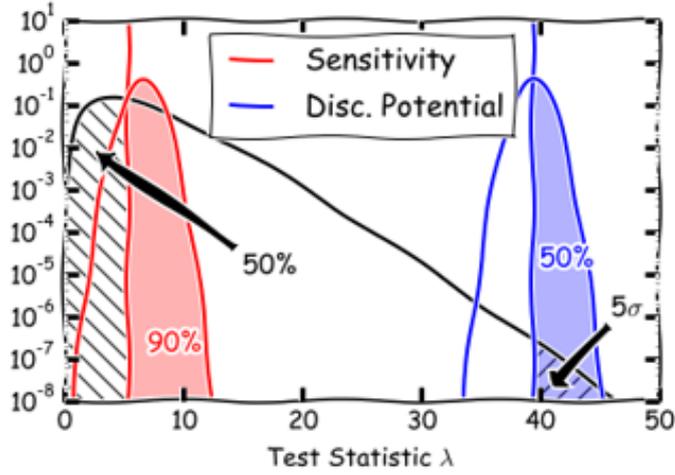


Figure 2.6: A schematic illustration used to define the concepts of Discovery Potential and sensitivity for the example  $T = \lambda$ . The background-only PDF  $\mathcal{T}_0$  is shown in black, and the signal-plus-background PDFs  $\mathcal{T}_s$  corresponding with the DP and sensitivity are colored in blue and red, respectively [50].

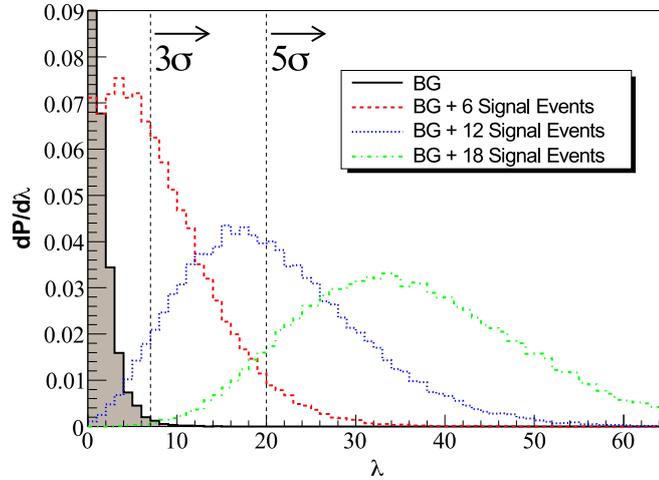


Figure 2.7: The determination of the  $3\sigma$  DP,  $5\sigma$  DP, and LDS w.r.t. to the pure background (BG) distribution  $\mathcal{T}_0$  (cf. Fig. 2.5b), depicted in black. The PDFs  $\mathcal{T}_s$  corresponding with these quantities are shown respectively in red, blue, and green, and the related amounts of injected signal events  $s$  are also indicated [78].

sensitivity, a trial  $T$  will be above the median of  $\mathcal{T}_0$  at a CL of 90%. This is illustrated in Fig. 2.6. As in the case of the DP determination,  $\mathcal{T}_s$  is determined for different signal injections  $s > 0$ , each requiring  $\mathcal{O}(10^4)$  pseudo-experiments, until the sensitivity is found.

An alternative and compatible definition of the sensitivity found in the literature is

the average confidence upper limit at 90% CL, weighted with  $\mathcal{T}_0$ . That is, one uses the ordering principle of Feldman and Cousins, discussed in Section 2.1.3, to construct for each  $s \geq 0$  acceptance intervals that contain of 90% of the probability determined from  $\mathcal{T}_s$ . Hence, a confidence belt is obtained at 90% CL, from which one can determine the upper limit  $s_2(T)$  for each  $T$ . The sensitivity  $s_{\text{sens}}$  is now defined as

$$s_{\text{sens}} = \int_{-\infty}^{+\infty} s_2(T') \mathcal{T}_0(T') dT'.$$

However, in this thesis solely the first definition of the sensitivity will be considered, which determines the 90% confidence upper bound w.r.t. to the median rather than calculating the weighted average.

In analogy with the DP, the optimization of the sensitivity is an important part of the analysis, which is achieved through the same techniques as in the DP case. Also here the optimal sensitivity is the lowest amount of injected signal needed to be distinguishable from the background. Note that typically an analysis is optimized for the DP or the sensitivity, but not both simultaneously.

Lastly, the sensitivity is typically not optimized in source analyses with ICECUBE, since here the sources of the observed cosmic neutrino flux are yet to be discovered, implying that an optimization of the DP is more relevant. Contrastingly, in analyses concerning dark matter searches with ICECUBE, the sensitivity might be optimized rather than the DP, since currently no evidence has been provided for (relatively low-energy) neutrinos originating from dark matter interactions.

## 2.2.4 Analysis of Experimental Data

In order to prevent biases in the analysis, the DP and sensitivity are quantities determined from pseudo-experiments that are independent of the measurements. Note that these quantities are merely used to optimize the analysis. After the optimization is complete, the actually observed dataset  $D$  of the experiment becomes available<sup>10</sup>, from which a single value  $T_D$  of each considered test statistic is determined. This  $T_D$  is then compared with the background-only PDF  $\mathcal{T}_0$  in order to obtain its  $p$ -value  $p_D$ .

Depending on the value of  $p_D$ , a different type of result will be given from an analysis of the ICECUBE data. First, consider the scenario where  $p_D$  corresponds with a significance that is less than  $5\sigma$ . In that case, the confidence belt at 90% CL is constructed using the ordering principle of Feldman and Cousins. For the specific value of  $T_D$ , one thus obtains a confidence interval for  $s$  at 90% CL.

In contrast, if  $p_D$  corresponds with a significance of at least  $5\sigma$ , one wants to state an observed value  $s_{\text{obs}}$  and its  $\pm 1\sigma$  statistical uncertainties (note that the latter  $\sigma$  does not refer to a significance of  $T_D$ ). This is done by constructing the confidence belt at 68.27% CL using the ordering principle of Feldman and Cousins. For the determined  $T_D$ , one then sets  $s_{\text{obs}} = s_{\text{best}}$ , i.e. the value of  $s$  which best fits the likelihood in the ordering principle. The upper limit and lower limit of the corresponding confidence interval then comprise the  $+1\sigma$  and  $-1\sigma$  uncertainties, respectively.

<sup>10</sup>This is called data unblinding in ICECUBE analyses; the optimizations of the DP and sensitivity form part of the blind analysis, where there is no knowledge of  $D$  in order to prevent biases.

Yet, this choosing of the CL as function of the value of  $p_D$  obtained from the data is another form of the “flip-flop” reasoning discussed in Section 2.1.3. The confidence belt “thins” from 90% CL to 68.78% CL after trespassing the  $5\sigma$  threshold, which leads to undercoverage w.r.t. the 90% CL region, and overcoverage w.r.t. the 68.78% CL region. Hence the  $\pm 1\sigma$  uncertainties on an observation with at least a  $5\sigma$  significance will overcover, but this conservatism is not really an issue if a signal is discovered. On the other hand, undercoverage is avoided for the lowest significances where the most important constraints are performed, since the corresponding  $T_D$ -values lead to confidence intervals that do not contain any values of  $s$  that undercover.

However, undercoverage for more intermediate significances—which could be between  $3\sigma$  and  $5\sigma$  for example—does occur. One could argue that if an experiment obtains a result with such an intermediate significance, it will definitely be performed again with more data. This will probably lower the significance of the new result w.r.t. the original result if the latter was a background fluctuation, or probably increase it if there is an actual signal present. Therefore, this new result has a relatively large likeliness of avoiding undercoverage. Unfortunately, this argument is not a justification of the fact that undercoverage does occur in practice when a confidence interval at 90% CL is given at these intermediate significances.

The discussion given above is based on the determination of one value  $p_D$  from the data  $D$ . Yet, it is common that in an analysis several values  $(p_D)_i$  with  $i \in \{1, 2, \dots, N\}$  are determined from  $N$  trials. In a steady point source analysis for example, such as the one discussed in Chapter 1, for each of the  $N$  considered candidate sources a  $(p_D)_i$  is determined, which are called pre-trial  $p$ -values.

Such *pre-trial  $p$ -values* do not correspond with a correct assessment of significance—if e.g.  $N = 10^8$  trials would be performed and no signal would be present, still several  $(p_D)_i$  would be smaller than  $5.73 \times 10^{-7}$ . As a consequence, one needs to correct for the fact that  $N$  trials were performed. For that, one typically finds the smallest  $(p_D)_{\min} = \min_i \{(p_D)_i\}$ , and then one calculates the probability  $P$  to obtain a value  $p < (p_D)_{\min}$  in  $N$  trials. This is given by a binomial distribution,

$$\begin{aligned} P &= \frac{N!}{(N-1)!} (p_D)_{\min} [1 - (p_D)_{\min}]^{(N-1)} \\ &= N (p_D)_{\min} [1 - (p_D)_{\min}]^{(N-1)}. \end{aligned}$$

The probability  $P$ , that would then correspond with the single  $p_D$  mentioned above, is therefore called the *post-trial* or *trial-corrected  $p$ -value*, of which a  $5\sigma$  significance is required to claim a discovery.

## 2.3 Additional Remarks

Firstly, it should be emphasized that the result of an analysis is to be interpreted cautiously. Recall that a (frequentist) confidence interval at a CL  $\alpha$  does not have of probability  $\alpha$  to contain the true value of the signal. The correct statement is that in a fraction  $\alpha$  of the experiments, the resulting confidence interval will contain the true value. However, one should bear in mind that both overcoverage and undercoverage do occur in practice.

Secondly, when comparing different test statistics, there is in principle no unique definition of “the best statistical method”. Considering the DP scenario—the following reasoning can in principle also be applied to sensitivities, although in this case it is typically not relevant in practice—, it might be argued that the optimal test statistic is the one for which the lowest value of the DP is obtained, i.e. the value for which  $\delta\% = 50\%$  if the trials will trespass the  $k\sigma = 5\sigma$  threshold. However, it can occur that a method gets a 10% probability to exceed  $5\sigma$  when injecting a small number of signal events, such that it becomes the optimal statistical method for a very low signal. Thus, it might be useful to consider different values of  $\delta$  for a fixed  $k$  and vice versa to determine the best test statistic in each scenario. This will be taken into consideration in Chapter 5.

Furthermore, from the experimental point of view, the DP and sensitivity are expressed as signal rates rather than number of signal events, which will also be done in this thesis. Yet, in references of ICECUBE source analyses such as the one discussed in Chapter 1, these quantities are typically denoted as cosmic neutrino fluxes. The conversion from an event rate to a flux was also mentioned in Chapter 1, where one has to take into account the effective area of the ICECUBE detector.

Also to be noted is that one has to be careful when using experimental data to determine the background-only PDF  $\mathcal{T}_0$ . It is imperative to ensure that the optimization of the statistical methods should not be biased towards the data  $D$  that has to be analyzed. In the scrambling example given above, this is avoided since the data  $D$  is mixed with other experimental data in the declination band.

Finally, it should be remarked that the frequentist techniques outlined in Section 2.2 are not exclusive in particle physics (where also Bayesian principles are used). They are developed with the purpose to discover a signal signature in the data rather than to obtain precision measurements, as is typically done in the determination of Standard Model quantities in experiments at e.g. the Large Hadron Collider. In the latter analyses, much more attention goes into uncertainties (statistical and systematic) of parameters, to obtain results that are as exact as possible. This is, however, not of immediate relevance in neutrino astronomy as this field is currently in a phase of claiming discoveries rather than making precision measurements.

# Toy Model of Steady Point Sources

## Introduction

In order to investigate different statistical methods for steady point source analyses with ICECUBE, a framework is needed with which to perform pseudo-experiments. For this purpose, a toy model is constructed in Section 3.1. Furthermore, to illustrate the concepts of Discovery Potential and sensitivity developed in the previous Chapter, the analysis of a simple counting experiment in the context of the toy model is given in Section 3.2.

## 3.1 Toy Model

The toy model described here is based on the one given in [79], which was used to generate events for a GRB analysis, i.e. it simulated both the time and angular position of an event. Since the interest of this work lies in *steady* point sources, the event generation needs to cover an extended time range, which was chosen to be one year. Therefore, only the event position on the sky is of relevance here; the toy model does not take the energy of an event into consideration.

Consider a steady point source (PS) with a location vector  $\mathbf{r}_{\text{PS}}$  on the sky. The opening angle  $\alpha$  of a certain point on the sky, given by a location vector  $\mathbf{r}$ , w.r.t. the PS can then be defined through the scalar product of these two vectors. That is,

$$\alpha = \arccos\left(\frac{\mathbf{r}_{\text{PS}} \cdot \mathbf{r}}{|\mathbf{r}_{\text{PS}}||\mathbf{r}|}\right),$$

where  $0 \leq \alpha \leq \pi$ . Subsequently, a patch of size  $\alpha_{\text{max}}$  around the PS can be defined as the collection of points on the sky for which  $\alpha \leq \alpha_{\text{max}}$ . For the choice of  $\alpha_{\text{max}}$ , such a patch will be used to define the region of the sky which is scanned for a signal coming from the PS.

The following Sections describe how  $\alpha$  is generated for both background and signal events. Note that the toy model can directly be generalized for multiple point sources, but this will not be of relevance in this thesis.

### 3.1.1 Background Generation

The average background event rate  $\dot{N}_b$  of atmospheric neutrinos detected by ICECUBE, which is the only background component considered in the toy model, is roughly 10 events per hour from the Northern Hemisphere [53]. Note that for the purpose of the toy model, the distinction between the background rates from the two different hemispheres is not relevant. Therefore this rate  $\dot{N}_b = 2.8$  mHz is taken, from which the average number of background events  $\Lambda_b$  in one year can be determined. The probability to obtain a certain number of background events  $N_b$  in one year is given by a Poisson PDF,

$$\mathcal{P}(N_b \text{ events}) = \frac{(\Lambda_b)^{N_b} e^{-\Lambda_b}}{N_b!}. \quad (3.1)$$

Thus, for each pseudo-experiment the number of background events to be simulated is drawn randomly<sup>1</sup> from this distribution.

In the toy model it is assumed that the background events are distributed isotropically. Therefore, for each background event,  $\cos \alpha$  is chosen randomly from a uniform distribution, which takes into account the solid angle effect. Explicitly, the background PDF  $B_\alpha$  of  $\alpha$  (not  $\cos \alpha$ ) is given by

$$B_\alpha(\alpha) = \frac{\sin \alpha}{2}.$$

### 3.1.2 Signal Generation

The rate of injected signal events  $\dot{N}_s$ , or equivalently, the number of injected signal events<sup>2</sup>  $N_s$  corresponding with a time span of one year, is a free parameter in the toy model. The generation of such an event takes into account the ICECUBE resolution  $\sigma_\alpha$  on the opening angle  $\alpha$  (e.g.  $\sim 1^\circ$  for muon neutrinos as mentioned in Chapter 1). Therefore, the signal PDF  $S_\alpha$  of  $\alpha$  is chosen to be a single-sided Gaussian PDF,

$$S_\alpha(\alpha) = \sqrt{\frac{2}{\pi\sigma_\alpha^2}} \exp\left(-\frac{(\mu_\alpha - \alpha)^2}{2\sigma_\alpha^2}\right), \quad (3.2)$$

where the expectation value  $\mu_\alpha = 0^\circ$  is the opening angle corresponding with the considered point source PS. For each simulated signal event, a random value is drawn from this  $S_\alpha$ .

It should be noted that  $S_\alpha$  is not exactly normalized over the physical range  $[0^\circ, 180^\circ]$  of  $\alpha$ , but rather over all positive values  $[0^\circ, \infty)$ , i.e.

$$\int_{0^\circ}^{\infty} S_\alpha(\alpha') d\alpha' = 1 \quad \implies \quad \int_{0^\circ}^{180^\circ} S_\alpha(\alpha') d\alpha' < 1.$$

<sup>1</sup>Such a drawing is in practice always pseudo-random, but this detail will be omitted in the text.

<sup>2</sup>Note that  $N_s$  corresponds with the number  $s$  of the previous Chapter; from this point, “ $s$ ” is simply used as a label.

However, for an angular resolution  $\sigma_\alpha = 1^\circ$  used in this work, the discrepancy is negligible, and therefore Eq. 3.2 was used for the generation of signal events.

## 3.2 Analysis in a Simplified Counting Experiment

Take a patch of size  $\alpha_{\max}$  centered around the PS, and let  $N = N_b + N_s$  be the total amount of generated events on the full sky corresponding with a timespan of one year. The conceptually simplest statistical method is to count which fraction of  $N$  is contained within this patch. This number  $n = n_b + n_s$  is therefore the test statistic considered here,  $T = n$ , where  $n_b$  and  $n_s$  are the amount of background and signal events in the patch, respectively<sup>3</sup>.

The goal of this counting experiment is to reproduce Fig. 2.6 for the choice  $T = n$ , so that the determination of the Discovery Potential and sensitivity is illustrated in a practical application. The patch size was chosen to be  $\alpha_{\max} = 5^\circ$  such that quasi all simulated signal events would fall within the patch for the choice of a muon neutrino resolution  $\sigma_\alpha = 1^\circ$ . For simplicity and due to the illustrative purposes of this Section, it was also chosen to let  $N_b = \Lambda_b$  fixed for all pseudo-experiments instead of drawing this value from Eq. 3.1. Therefore, the PDF  $\mathcal{T}_{N_s}$  of  $T = n$  for a certain signal injection  $N_s \geq 0$  is a Poisson distribution,

$$\mathcal{T}_{N_s}(n|\lambda) = \mathcal{P}(n = n_b + n_s \text{ events in patch}) = \frac{\lambda^n e^{-\lambda}}{n!}, \quad (3.3)$$

where  $\lambda = \lambda(N_s) = \lambda_b + \lambda_s(N_s)$  is the average number of events in this patch<sup>4</sup>, composed of the average number of background and signal events ( $\lambda_b$  and  $\lambda_s$ , respectively), which depends on  $N_s$ . Recall that in general, the analytic form of  $\mathcal{T}_{N_s}$  is not known a priori.

Firstly, the background-only PDF  $\mathcal{T}_0$  of the test statistic has to be determined, of which the analytic formulation can be found by setting  $N_s = 0$  and thus also  $n_s = 0$  and  $\lambda = \lambda(N_s = 0) = \lambda_b$  in Eq. 3.3,

$$\mathcal{T}_0(n|\lambda_b) = \mathcal{P}(n = n_b \text{ events in patch}) = \frac{\lambda_b^n e^{-\lambda_b}}{n!}. \quad (3.4)$$

Since the functional form of  $\mathcal{T}_0$  is known,  $10^5$  pseudo-experiments were performed and, in order to reach a  $5\sigma$  significance, a fit  $f(x)$  of the resulting histogram was made according to Eq. 3.4,

$$f(x) = \frac{(\lambda_b^{\text{fit}})^x e^{\lambda_b^{\text{fit}}}}{x!},$$

with  $\lambda_b^{\text{fit}}$  the fit parameter. The histogram and fit corresponding with the background-only PDF are shown in Fig. 3.1, and the result of the fit<sup>5</sup> was a value  $\lambda_b^{\text{fit}} = 167 \pm 14$  which corresponds well with the mean  $\lambda_0^{\text{hist}}$  of the histogram,  $\lambda_b^{\text{hist}} = 167$ . The uncertainty on  $\lambda_b^{\text{fit}}$  will be omitted in the remainder of this illustrative discussion.

<sup>3</sup>To avoid confusion, it is mentioned here that an upper-case  $N$  stands for the number of events over the full sky, whilst a lower-case  $n$  stands for the number of events within a patch of size  $\alpha_{\max}$ . In particular,  $n_s \leq N_s$ .

<sup>4</sup>This  $\lambda$  corresponds with  $\theta_s$  in the notation of Chapter 2.

<sup>5</sup>Also because of this illustrative example, uncertainties on the amount of bin entries were not taken into account.

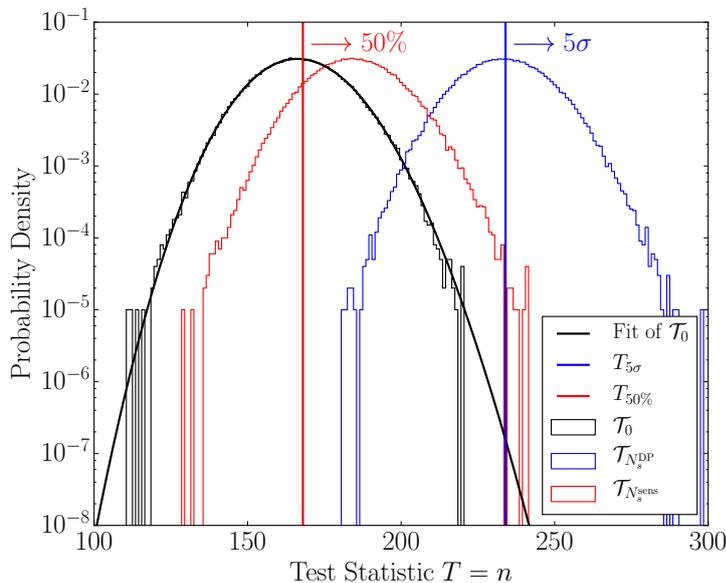


Figure 3.1: The reproduction of Fig. 2.6 in the context of the simplified counting experiment. In black, the distribution of the background-only PDF  $\mathcal{T}_0$  is shown together with the fit through this histogram. The vertical lines represent the values  $T_p = T_{50\%}$  (red) and  $T_p = T_{5\sigma}$  (blue) determined from  $\mathcal{T}_0$  which are required for the determination of the sensitivity and the DP, respectively. The distributions of the background plus signal PDFs  $\mathcal{T}_{N_s}$  for  $N_s = N_s^{\text{sens}}$ , the amount of injected signal events corresponding with the sensitivity, and  $N_s = N_s^{\text{DP}}$ , the amount of injected signal events corresponding with the DP, are respectively depicted in red and blue.

Before continuing, a cross-check can be performed of the determined values of  $\lambda_b^{\text{fit}}$  and  $\lambda_b^{\text{hist}}$ . Let  $\lambda_b^{\text{exp}}$  denote the expected average number of background events in the patch. Then, if  $\Omega$  and  $\Omega_{\alpha_{\text{max}}}$  respectively denote the solid angles of the full sky and the patch of size  $\alpha_{\text{max}} = 5^\circ$ , one can find that, recalling that  $N_b = \Lambda_b$  is fixed,

$$\lambda_b^{\text{exp}} = N_b \frac{\Omega_{\alpha_{\text{max}}}}{\Omega} = N_b \frac{2\pi(1 - \cos \alpha_{\text{max}})}{4\pi} = 167,$$

which is in good agreement with  $\lambda_b^{\text{fit}}$  and  $\lambda_b^{\text{hist}}$ .

Subsequently, the values  $T_p = n_p$  that mark the  $p$ -value of 50% and the  $5\sigma$  significance (noted as  $T_{5\sigma}$  instead of  $T_p = T_{5.73 \times 10^{-7}}$ ) w.r.t.  $\mathcal{T}_0$ , needed for the determination of respectively the sensitivity and the DP, were obtained. These are also indicated in Fig. 3.1. Their values are  $T_{50\%} = n_{50\%} = 167$  and  $T_{5\sigma} = n_{5\sigma} = 234$ . Note that the former value was obtained using the histogram whilst the latter value was found using the performed fit.

Finally, the distributions of  $\mathcal{T}_{N_s}$  for different values of  $N_s$  were determined using  $10^5$  pseudo-experiments for each value. According to the definitions of Chapter 2, the number of signal injections  $N_s^{\text{sens}}$  corresponding with the sensitivity was determined as the smallest  $N_s$  which exceeded the requirement that in 90% of the trials of  $\mathcal{T}_{N_s}$ , an obtained value of  $T = n$  would be larger than  $T_{50\%} = n_{50\%}$ . Analogously, the number

of signal injections  $N_s^{\text{DP}}$  corresponding with the DP was determined as the smallest  $N_s$  which exceeded the requirement that in 50% of the trials of  $\mathcal{T}_{N_s}$ , an obtained value of  $T = n$  would be larger than  $T_{5\sigma} = n_{5\sigma}$ .

The distributions of  $\mathcal{T}_{N_s}$  corresponding with  $N_s^{\text{sens}} = 18$  and  $N_s^{\text{DP}} = 67$  are shown in Fig. 3.1, which can now be directly compared with Fig. 2.6. For completeness, the resulting sensitivity  $\dot{N}_s^{\text{sens}}$  and DP  $\dot{N}_s^{\text{DP}}$ , denoted here as rates, are thus given by

$$\begin{aligned}\dot{N}_s^{\text{sens}} &= 0.57 \mu\text{Hz}, \\ \dot{N}_s^{\text{DP}} &= 2.12 \mu\text{Hz}.\end{aligned}$$



# Statistical Methods

## Introduction

This Chapter covers the different statistical methods that will later be compared using the toy model. Since the interest of this thesis lies in steady point sources, event distributions over time are irrelevant, as mentioned previously. Moreover, because energies are not considered in the toy model, they are not taken into account in the developed test statistics<sup>1</sup>. Here, the methods are developed for angular observation windows of ICECUBE centered around a certain source<sup>2</sup>. In other words, a patch of size  $\alpha_{\max}$  is taken centered around the source, and the test statistics will be constructed from a set of  $n$  measured opening angles  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)$  of the events within the patch<sup>3</sup>.

The considered statistical methods use different aspects of the (expected) angular distribution  $\mathcal{F}_\alpha$  of  $\alpha$  on the patch, which will contain both background and signal components. The latter will depend on the angular resolution  $\sigma_\alpha$  that causes the signal events to be smeared on the sky around the point source.

First, in Section 4.1 a link is made with the simplified counting experiment of the previous Chapter by considering a test statistic based on a counting method, which is compatible with the integral  $\int \mathcal{F}_\alpha d\alpha$  over the patch. Therefore, it will be most sensitive for a relatively strong signal-to-noise event ratio, which is favoured by patch sizes that are relatively small w.r.t. the angular resolution.

Subsequently, the construction of a likelihood test statistic using  $\mathcal{F}_\alpha$  is discussed in Section 4.2, which is the method mostly used in ICECUBE analyses. The method

---

<sup>1</sup>It should be mentioned that the inclusion of the energy distributions would significantly influence the test statistics, with the exception of the Li-Ma method (see below) in which the distributions are essentially integrated over.

<sup>2</sup>To be exact, it should be called a “candidate source”, but this detail will be omitted in the text since the toy model provides a fictitious source.

<sup>3</sup>Recall that here no *real* measurements will be used—pseudo-experiments are performed to determine  $\boldsymbol{\alpha}$ —, but the evaluation of the test statistics for actual measurements is completely analogous.

can be shown to work best for a reasonable amount of signal events and a patch of reasonable size, such that the signal contribution to  $\mathcal{F}_\alpha$  is prominent in the likelihood.

Furthermore, Section 4.3 focuses on a test statistic that is based on the variations of the previous likelihood w.r.t. the signal strength  $\eta$ , which is essentially using  $\partial\mathcal{F}_\alpha/\partial\eta$ . This “needle in a haystack” method is therefore sensitive for a small amount of signal events that will alter the “shape” of  $\mathcal{F}_\alpha$ , which will also occur for relatively small patch sizes.

Finally, as an alternative to the three anticipated methods which are used in unbinned analyses, a statistical method in which the background component of  $\mathcal{F}_\alpha$  is binned will be considered in Section 4.4. The signal contribution of the total angular distribution is not considered in this case. Thus, a reasonable amount of signal combined with a relatively large patch size is required for the method to become optimal, since then the contribution of the signal events will be most distinguishable w.r.t. the overall “shape” of the expected background distribution.

## 4.1 Li-Ma Method

Just as in the simplified counting experiment of Chapter 3, the method developed by Li and Ma [80] is based on counting the amount of events  $n$  in the patch of size  $\alpha_{\max}$  around the source, and also the amount of events  $n_{\text{off}}$  in a patch of the same size  $\alpha_{\max}$  that does *not* contain the source. Both quantities follow a Poisson PDF, as in the example of Chapter 3.

Li and Ma use a likelihood ratio to determine the significance  $\Sigma$  of the observed  $n$  and  $n_{\text{off}}$  w.r.t. a background-only hypothesis. Since the reasoning behind this likelihood ratio method is analogous to the discussion given in Section 4.2, the details of the determination of  $\Sigma$ , which can be found in [80], are omitted here. The result is given by

$$\Sigma = \sqrt{2} \left( n \ln \left[ \left( \frac{1+\beta}{\beta} \right) \left( \frac{n}{n+n_{\text{off}}} \right) \right] + n_{\text{off}} \ln \left[ (1+\beta) \left( \frac{n_{\text{off}}}{n+n_{\text{off}}} \right) \right] \right)^{1/2}, \quad (4.1)$$

where  $\beta = t_n/t_{n_{\text{off}}}$  with  $t_n$  and  $t_{n_{\text{off}}}$  the time used for the observations of  $n$  and  $n_{\text{off}}$ , respectively.

The factor  $\beta$  is of relevance in e.g.  $\gamma$ -ray astronomy, for which this method was originally developed, because a  $\gamma$ -ray satellite such as the Fermi Large Area Telescope [81] can only be focused on a limited region of the sky determined by its field of view. Therefore, in  $\gamma$ -ray astronomy,  $n$  and  $n_{\text{off}}$  are typically determined separately in time. However, in neutrino astronomy with ICECUBE, data is taken continuously over the whole sky, such that  $n$  and  $n_{\text{off}}$  can be determined simultaneously. Thus,  $\beta = 1$  can be chosen for the purposes of this work, and then Eq. 4.1 reduces to

$$\Sigma = \sqrt{2} \left( n \ln \left[ 2 \left( \frac{n}{n+n_{\text{off}}} \right) \right] + n_{\text{off}} \ln \left[ 2 \left( \frac{n_{\text{off}}}{n+n_{\text{off}}} \right) \right] \right)^{1/2}.$$

This  $\Sigma$  now denotes the Li-Ma test statistic. Note that from the definition of  $\Sigma$ , a  $k\sigma$  significance will correspond with a value  $\Sigma = k$ , where  $k \geq 0$ .

## 4.2 Likelihood Method

Consider a set of  $n$  measurements  $\alpha$  within a patch of size  $\alpha_{\max}$  centered around a steady point source. The goal is to estimate from these measurements the number of signal events  $n_s$  that are present in the patch. As mentioned in Chapter 2, a possibility is to determine the ML estimator  $\hat{n}_s = (\hat{n}_s)_{\text{ML}}$  of  $n_s$  through the maximization of the likelihood  $\mathcal{L}(n_s|\alpha)$ .

One option is to determine this likelihood by omitting the angular information, and simply count the amount of events  $n$  in the patch and determine their probability of occurrence, given by a Poisson distribution. This is the reasoning behind the Li-Ma method of Section 4.1. On the other hand, the likelihood can be obtained from the expected angular PDF  $\mathcal{F}_\alpha$  of  $\alpha$  within the patch. The latter will be the topic of this Section and it will be used to construct the prominent test statistic in ICECUBE analyses; this discussion is based on [77–79].

### 4.2.1 Expected Angular Distribution

Since the measurements  $\alpha$  consist of both background and signal, the angular PDF  $\mathcal{F}_\alpha$  is a linear combination of the expected background and signal distributions on the patch of size  $\alpha_{\max}$ , denoted by  $\mathcal{B}_\alpha$  and  $\mathcal{S}_\alpha$ , respectively. This requires  $\mathcal{B}_\alpha$  and  $\mathcal{S}_\alpha$  to be *normalized on the patch*. If there are  $n_s$  signal events and  $n - n_s$  background events in the patch, then

$$\mathcal{F}_\alpha(\alpha|n_s) = \binom{n_s}{n} \mathcal{S}_\alpha(\alpha) + \binom{n - n_s}{n} \mathcal{B}_\alpha(\alpha), \quad (4.2)$$

which is also normalized on the patch by construction.

The expected atmospheric neutrino background in ICECUBE, which is the only background component considered, is isotropic, and therefore  $\mathcal{B}_\alpha$  is given by

$$\mathcal{B}_\alpha(\alpha) = \frac{\sin \alpha}{1 - \cos \alpha_{\max}}, \quad (4.3)$$

where the solid angle effect was taken into account. Furthermore, the signal originating from a source is expected to be smeared because of the angular resolution  $\sigma_\alpha$  of ICECUBE ( $\sigma_\alpha = 1^\circ$  for muon neutrinos). Thus,  $\mathcal{S}_\alpha$  is taken as a one-sided Gaussian,

$$\mathcal{S}_\alpha(\alpha) = \mathcal{N} \exp\left(-\frac{(\mu_\alpha - \alpha)^2}{2\sigma_\alpha^2}\right), \quad (4.4)$$

where by definition the expectation value  $\mu_\alpha = 0^\circ$ , and where

$$\mathcal{N}^{-1} = \int_0^{\alpha_{\max}} \exp\left(-\frac{\alpha'^2}{2\sigma_\alpha^2}\right) d\alpha',$$

which ensures a normalization of  $\mathcal{S}_\alpha$  on the patch. For completeness, both PDFs are shown in Fig. 4.1 for the case  $\alpha_{\max} = 5^\circ$  and  $\sigma_\alpha = 1^\circ$ .

To be mentioned is that  $\mathcal{B}_\alpha$  and  $\mathcal{S}_\alpha$  should not be confused with  $B_\alpha$  and  $S_\alpha$  of Chapter 3, as they have completely different purposes. The latter are used to generate background and signal events in the toy model, whilst the former denote the expected distributions of the background and signal events within a patch of size  $\alpha_{\max}$ . Note however that if a resolution  $\sigma_\alpha$  is chosen for the generation of signal with  $S_\alpha$ , it is the same  $\sigma_\alpha$  used in  $\mathcal{S}_\alpha$ .

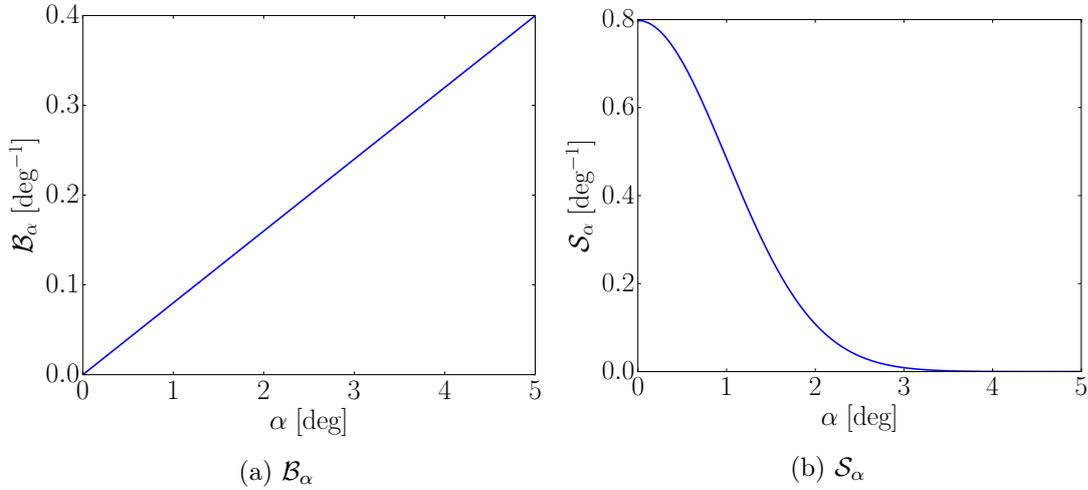


Figure 4.1: The expected background and signal PDFs  $\mathcal{B}_\alpha$  and  $\mathcal{S}_\alpha$  used for the determination of the likelihood test statistic, respectively shown in Figs. 4.1a and 4.1b, for a patch size  $\alpha_{\max} = 5^\circ$  and an ICECUBE resolution  $\sigma_\alpha = 1^\circ$ .

## 4.2.2 Maximization of the Likelihood

As mentioned in Chapter 2, typically the logarithm of the likelihood<sup>4</sup> is used in practice, i.e.  $\ln \mathcal{L}(n_s | \boldsymbol{\alpha})$  for a set of measurements  $\boldsymbol{\alpha}$ . To calculate the likelihood, Eq. 4.2 has to be evaluated for each  $\alpha_i$  with  $i \in \{1, 2, \dots, n\}$ ,

$$\begin{aligned} \ln \mathcal{L}(n_s | \boldsymbol{\alpha}) &= \sum_{i=1}^n \ln \mathcal{F}_\alpha(\alpha_i | n_s) \\ &= \sum_{i=1}^n \ln \left[ \left( \frac{n_s}{n} \right) \mathcal{S}_\alpha(\alpha_i) + \left( 1 - \frac{n_s}{n} \right) \mathcal{B}_\alpha(\alpha_i) \right]. \end{aligned} \quad (4.5)$$

Recall that  $n_s$  is an unknown parameter, which has to be determined from the observed  $\boldsymbol{\alpha}$ .

The estimator  $\hat{n}_s$  that best fits the measurements  $\boldsymbol{\alpha}$  is the value of  $n_s$  which maximizes the likelihood in Eq. 4.5. However, a lower bound on  $n_s/n$  has to be introduced in the maximization in order to prevent  $\ln \mathcal{L}(n_s | \boldsymbol{\alpha})$  to become infinite. This bound is taken from [78], and is given by

$$\frac{n_s}{n} > \begin{cases} \max_i \{ \mathcal{B}_\alpha(\alpha_i) / [\mathcal{B}_\alpha(\alpha_i) - \mathcal{S}_\alpha(\alpha_i)] \} & \text{if } \exists \alpha_i : \mathcal{S}_\alpha(\alpha_i) > \mathcal{B}_\alpha(\alpha_i), \\ -1 & \text{else.} \end{cases}$$

The reason why the unphysical values  $n_s < 0$  are allowed, is because it is perfectly possible that, in case the amount of signal events is negligible in comparison with the amount background events, a statistical fluctuation in the background results in a ML estimator  $\hat{n}_s < 0$ . This will be taken into account when developing the test statistic in

<sup>4</sup>For brevity, the “log likelihood” will just be called the likelihood from this point onwards.

Section 4.2.3. For the same reason, in case of a negligible amount background events w.r.t. the amount of signal events, the physical bound  $n_s/n \leq 1$  could also be exceeded by  $\hat{n}_s/n$ , although this scenario is very unlikely to occur.

To illustrate the maximization in context of the toy model, Figs. 4.2a and 4.2b show the likelihood and the corresponding values of  $\hat{n}_s/n$  for the cases of 0 and 20 injected signal events, respectively. The patch was chosen to have a size  $\alpha_{\max} = 5^\circ$  and the ICECUBE resolution was taken as  $\sigma_\alpha = 1^\circ$ . In the example of Fig. 4.2a it can be seen that the background fluctuated in such a way that  $\hat{n}_s < 0$ .

### 4.2.3 Construction of the Likelihood Test Statistic

In order to construct a valuable test statistic for a certain set of measurements  $\alpha$ , one cannot simply use  $\ln \mathcal{L}(\hat{n}_s|\alpha)$ , i.e. the likelihood evaluated at  $n_s = \hat{n}_s$  in Eq. 4.5. Consider a scenario where there are two sets  $\alpha_1$  and  $\alpha_2$  of which the former contains many more events (both background and signal) than the latter, but which both give rise to the same ML estimator  $\hat{n}_s$ . Purely because of the fact that there are more events in  $\alpha_1$ , one would have that  $\ln \mathcal{L}(\hat{n}_s|\alpha_1) < \ln \mathcal{L}(\hat{n}_s|\alpha_2)$ . This could then be interpreted as “there is a smaller likeliness that there is signal present in set  $\alpha_1$  than in set  $\alpha_2$ ”, which is erroneous since the same  $\hat{n}_s$  was obtained.

This issue is remedied by constructing a likelihood ratio  $\ln R_\lambda(\hat{n}_s|\alpha)$  for a set  $\alpha$  which serves as a normalization of  $\mathcal{L}(\hat{n}_s|\alpha)$ ,

$$\ln R_\lambda(\hat{n}_s|\alpha) = \ln \left( \frac{\mathcal{L}(\hat{n}_s|\alpha)}{\mathcal{L}_0(\alpha)} \right).$$

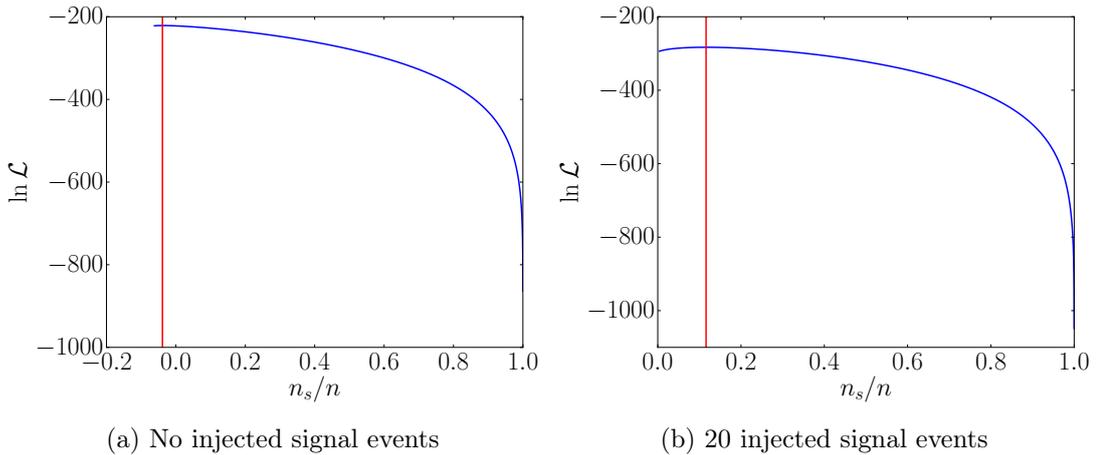


Figure 4.2: The likelihood functions for 0 injected signal events, depicted in Fig. 4.2a, and 20 injected signal events, shown in Fig. 4.2b, using the toy model, determined for a patch size  $\alpha_{\max} = 5^\circ$  and an ICECUBE resolution  $\sigma_\alpha = 1^\circ$ . For each case, the red line indicates the maximum likelihood estimator  $\hat{n}_s/n$ ; in the background-only scenario,  $\hat{n}_s < 0$  for the presented example.

Here,

$$\begin{aligned}\ln \mathcal{L}_0(\boldsymbol{\alpha}) &= \ln \mathcal{L}(n_s = 0 | \boldsymbol{\alpha}) \\ &= \sum_{i=1}^n \ln [\mathcal{B}_{\alpha}(\alpha_i)],\end{aligned}$$

which represents the likeliness that the same set  $\boldsymbol{\alpha}$  purely consists of background.

A common misinterpretation in how the normalization factor  $\ln \mathcal{L}_0(\boldsymbol{\alpha})$  of the likelihood ratio is determined occurs when one writes  $\boldsymbol{\alpha} = \boldsymbol{\alpha}_b + \boldsymbol{\alpha}_s$  as the sum of its background component  $\boldsymbol{\alpha}_b$  and its signal component  $\boldsymbol{\alpha}_s$ . One could then erroneously calculate the normalization factor as  $\ln \mathcal{L}_0(\boldsymbol{\alpha}_b) = \ln \mathcal{L}(n_s = 0 | \boldsymbol{\alpha}_b)$ , i.e. by only evaluating the background events  $\boldsymbol{\alpha}_b$  in the likelihood. This is *not* correct; one should use the full set  $\boldsymbol{\alpha}$  of events. Moreover, if  $\boldsymbol{\alpha}$  would consist of experimental data, one would not be able to distinguish  $\boldsymbol{\alpha}_b$  and  $\boldsymbol{\alpha}_s$  during the analysis, so that  $\ln \mathcal{L}_0(\boldsymbol{\alpha}_b)$  would be non-determinable.

Finally, the likelihood test statistic  $\lambda$  is defined as

$$\begin{aligned}\lambda &= 2 \operatorname{sign}(\hat{n}_s) \ln R_{\lambda}(\hat{n}_s | \boldsymbol{\alpha}) \\ &= 2 \operatorname{sign}(\hat{n}_s) [\ln \mathcal{L}(\hat{n}_s | \boldsymbol{\alpha}) - \ln \mathcal{L}_0(\boldsymbol{\alpha})].\end{aligned}$$

The  $\operatorname{sign}(\hat{n}_s)$  ensures that negative ML estimators  $\hat{n}_s < 0$  can be distinguished from the positive values  $\hat{n}_s > 0$  in the test statistic, since the ratio  $\ln R_{\lambda}(\hat{n}_s | \boldsymbol{\alpha})$  is not sensitive for the sign of  $\hat{n}_s$ . The extra factor 2 is included because in that case, for  $n \rightarrow \infty$  and if the background-only hypothesis is true,  $\lambda \sim \chi_{\nu}^2$ , i.e. the likelihood test statistic follows a  $\chi^2$  distribution with  $\nu$  degrees of freedom, which are the free parameters used in the likelihood maximization, assuming that these are stochastically independent [69, 78, 80]. For the  $\lambda$  constructed here, the only free parameter was  $n_s$ , such that  $\nu = 1$ .

Note that it is this  $\lambda$  which was previously encountered in Chapter 1 for the steady point source analysis [62] and in Chapter 2 for the analyses of [77, 78], although in those examples the respective energy distributions of the events were also included in the likelihood, as is standard for ICECUBE analyses. Then one typically performs a likelihood maximization with two degrees of freedom, i.e.  $n_s$  and the spectral index  $\gamma$  of the energy distribution. Thus, in those cases one would expect that  $\lambda \sim \chi_2^2$ . However,  $n_s$  and  $\gamma$  are typically correlated, such that  $\nu < 2$  in the  $\chi^2$  distribution. In practice, one typically fits a  $\chi_{\nu}^2$  through the obtained background-only distribution of  $\lambda$  to reach a  $5\sigma$  significance, where  $\nu$  is then taken as a free parameter. An example of such a fit was shown in Fig. 2.5b.

### 4.3 PLT Method

Pilla, Loader, and Taylor developed their self-named PLT method [82] with the purpose of finding “a needle in a haystack”, that is, a small signal component overwhelmed by a dominating sea of background. The method essentially considers the variations of the likelihood defined in the previous section, which might be a more promising approach if the amount of signal in the data is severely suppressed.

### 4.3.1 Signal as a Perturbation of the Background

First, the expected angular PDF  $\mathcal{F}_\alpha$  of Eq. 4.2 can be rewritten as

$$\mathcal{F}_\alpha(\alpha|\eta, \mu_\alpha) = \eta \mathcal{S}_\alpha(\alpha|\mu_\alpha) + (1 - \eta) \mathcal{B}_\alpha(\alpha),$$

with  $\eta = n_s/n$ , where  $\mathcal{B}_\alpha$  and  $\mathcal{S}_\alpha$  are given by Eqs. 4.3 and 4.4, respectively, although here  $\mu_\alpha$  is taken as a free parameter. In the PLT method, the signal is considered as a perturbation of the background with a strength  $\eta \in [0, 1]$  located at  $\mu_\alpha$ .

For a set of measurements  $\boldsymbol{\alpha}$ , a so-called *score*  $S(\mu_\alpha|\boldsymbol{\alpha})$  can be given to them by considering the variation of the likelihood  $\ln \mathcal{L}(\eta, \mu_\alpha|\boldsymbol{\alpha})$  (see Eq. 4.5) with the perturbation strength  $\eta$ . This derivative is then evaluated for a perturbation strength equal to zero,

$$\begin{aligned} S(\mu_\alpha|\boldsymbol{\alpha}) &= \left. \frac{\partial \ln \mathcal{L}(\eta, \mu_\alpha|\boldsymbol{\alpha})}{\partial \eta} \right|_{\eta=0} \\ &= \sum_{i=1}^n \left. \frac{\partial \ln [\eta \mathcal{S}_\alpha(\alpha_i|\mu_\alpha) + (1 - \eta) \mathcal{B}_\alpha(\alpha_i)]}{\partial \eta} \right|_{\eta=0} \\ &= \sum_{i=1}^n \frac{\mathcal{S}_\alpha(\alpha_i|\mu_\alpha) - \mathcal{B}_\alpha(\alpha_i)}{\mathcal{B}_\alpha(\alpha_i)}. \end{aligned} \quad (4.6)$$

The idea behind taking the derivative at  $\eta = 0$  is that then the expectation value of  $S(\mu_\alpha|\boldsymbol{\alpha})$  is zero for all  $\mu_\alpha$  if the background-only hypothesis is true, whilst it would peak at the true value of  $\mu_\alpha$  if signal would be present.

Note that for the development of the test statistic in [82], the score is normalized to take into account the dependence of its random variability on the value of  $\mu_\alpha$ . However, the exact value of  $S(\mu_\alpha|\boldsymbol{\alpha})$  is not of real interest in this work since the goal is to compare its background-only PDF with its signal-plus-background PDFs, for different signal injections with the toy model. Therefore, this normalization is not taken into consideration here.

### 4.3.2 Maximization of the Score

The PLT test statistic  $S$  is now defined as the score which is maximized for a value  $\mu_\alpha = (\mu_\alpha)_{\max} \in [0^\circ, \alpha_{\max}]$ ,

$$S = S((\mu_\alpha)_{\max}|\boldsymbol{\alpha}) = \max_{\mu_\alpha} S(\mu_\alpha|\boldsymbol{\alpha}).$$

To be completely correct in the interpretation of the PLT method, for each  $\mu_\alpha$  the expected signal PDF  $\mathcal{S}_\alpha$  of Eq. 4.4 should be normalized correctly in the calculation of  $S((\mu_\alpha)_{\max}|\boldsymbol{\alpha})$ ; the normalization constant  $\mathcal{N}$  depends on the value of  $\mu_\alpha$ ,

$$\mathcal{N}^{-1} = \mathcal{N}^{-1}(\mu_\alpha) = \int_0^{\alpha_{\max}} \exp\left(-\frac{(\mu_\alpha - \alpha')^2}{2\sigma_\alpha^2}\right) d\alpha'.$$

However, taking this into account can result in erroneous maximizations of the score if a few signal events are injected using the toy model. This is due to the fact

that in those cases local maxima are hard to distinguish computationally from the desired global maximum on the considered patch. Therefore, it was chosen to let  $\mathcal{N} = \mathcal{N}(\mu_\alpha = 0)$  be fixed for all values of  $\mu_\alpha$ . This alters the value of the test statistic  $S$ , but the exact value of  $S$  is not of relevance here, as mentioned previously. Figs. 4.3a and 4.3b show, choosing  $\alpha_{\max} = 5^\circ$  and  $\sigma_\alpha = 1^\circ$ , the respective maximizations of the score for 0 signal event injections and of the score for 20 signal event injections using the toy model.

Furthermore, it was checked that  $(\mu_\alpha)_{\max}$  still behaved as expected being interpreted as the location of the perturbation which maximizes the score. In fact, it should converge to the true location  $(\mu_\alpha)_{\max} = 0^\circ$  the more signal is injected with the toy model. To do so, for each number of injected signal events  $N_s \in \{0, 5, \dots, 100\}$  (recall from Chapter 3 that  $N_s$  is generally *not* equal to  $n_s$ ),  $(\mu_\alpha)_{\max}$  was determined 10 times from which the average  $\langle(\mu_\alpha)_{\max}\rangle$  was computed. The result is given in Fig. 4.4, where the expected convergence is present up to around  $N_s = 40$ , after which  $\langle(\mu_\alpha)_{\max}\rangle$  remains relatively constant around  $0.2^\circ$ . This observed bias is due to the fixation of  $\mathcal{N} = \mathcal{N}(\mu_\alpha = 0)$ . For a simulated set  $\alpha$  with 50 signal injections ( $\alpha_{\max} = 5^\circ$  and  $\sigma_\alpha = 1^\circ$ ), Fig. 4.5 shows the evaluation of  $\mathcal{S}_\alpha(0^\circ|\alpha_i)$  and  $\mathcal{S}_\alpha(0.2^\circ|\alpha_i)$  (without the correct normalization) for the different  $\alpha_i$  of  $\alpha$ . Since  $\sum_{i=1}^n \mathcal{S}_\alpha(0.2^\circ|\alpha_i) = 43.68 \text{ deg}^{-1} > \sum_{i=1}^n \mathcal{S}_\alpha(0^\circ|\alpha_i) = 37.94 \text{ deg}^{-1}$ , the former is favoured in the score maximization. Note that for this particular  $\alpha$ ,  $(\mu_\alpha)_{\max} = 0.28^\circ$ . The bias does not present an issue, however, since for a fixed  $\mu_\alpha$  and a growing sample  $\alpha$  with increasing  $N_s$ , the test statistic  $S$  will be larger on average due to the summation in Eq. 4.6. This still allows the distinction of a PDF of  $S$  for e.g. 50 and 70 injected signal events, which is of most relevance for this work.

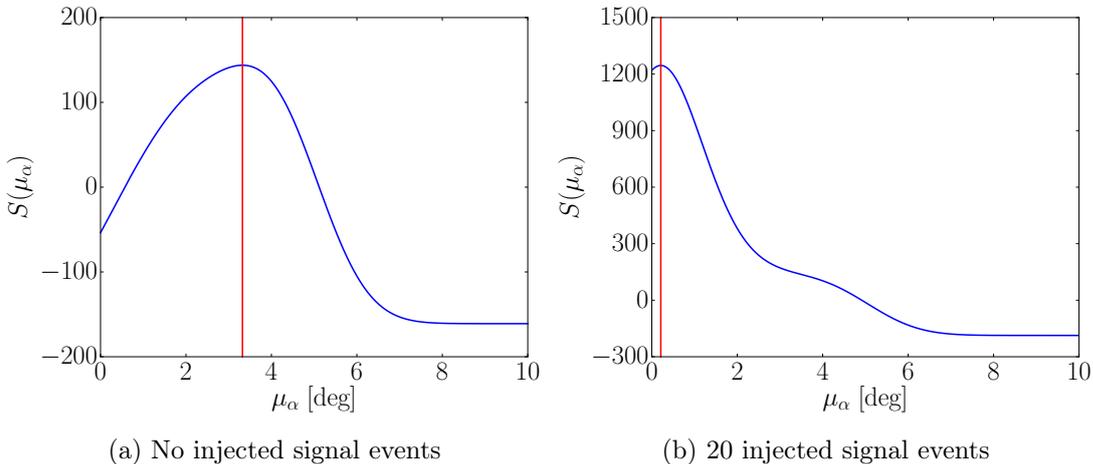


Figure 4.3: The score  $S(\mu_\alpha|\alpha)$  as function of the location  $\mu_\alpha$  of the signal perturbation for the choices  $\alpha_{\max} = 5^\circ$  and  $\sigma_\alpha = 1^\circ$ , depicted in Figs. 4.3a and 4.3b respectively for 0 and 20 signal injections with the toy model. The red line in each scenario denotes the score maximizer  $(\mu_\alpha)_{\max}$ , of which the corresponding score is the PLT test statistic  $S$ . Note that for the examples shown here, the value of  $S$  for 20 signal injections is about an order of magnitude larger than its value in the case of only background.

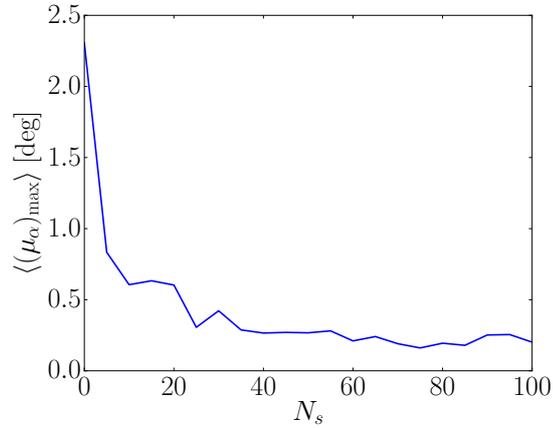


Figure 4.4: The behavior of the average score maximizer  $\langle(\mu_\alpha)_{\max}\rangle$  as a function of the number of signal injections  $N_s$  with the toy model, where for each probed  $N_s$  the average value was computed from 10 determinations of the score maximizer  $(\mu_\alpha)_{\max}$ . Also here,  $\alpha_{\max} = 5^\circ$  and  $\sigma_\alpha = 1^\circ$ . The expected trend towards zero is present up to about  $N_s = 40$ , after which  $\langle(\mu_\alpha)_{\max}\rangle$  remains relatively constant.

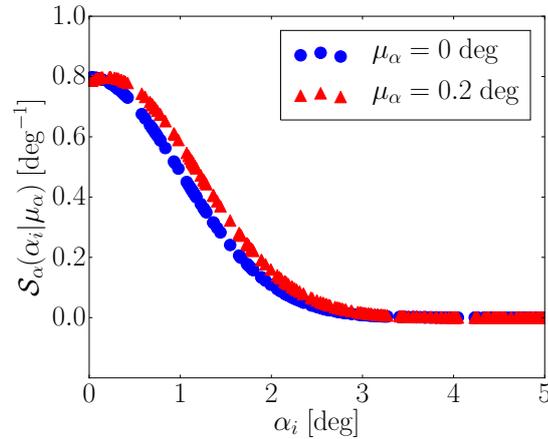


Figure 4.5: The evaluation of the signal distributions  $\mathcal{S}_\alpha(\alpha_i|0^\circ)$  and  $\mathcal{S}_\alpha(\alpha_i|0.2^\circ)$ , using a fixed normalization  $\mathcal{N} = \mathcal{N}(\mu_\alpha = 0)$ , for a simulated set of events  $\boldsymbol{\alpha} = \{\alpha_i\}$  with 50 signal injections and the choices  $\alpha_{\max} = 5^\circ$  and  $\sigma_\alpha = 1^\circ$ . Mostly, the distribution with  $\mu_\alpha = 0.2^\circ$  has larger values which are favourable for the score maximization. This causes the bias that is observed in Fig. 4.4.

Finally, an interesting remark of the PLT test statistic  $S$  can be made by examining Eq. 4.6, and in particular, the division by  $\mathcal{B}_\alpha(\alpha_i)$  for a certain  $\alpha_i$  of the measurements  $\boldsymbol{\alpha}$ . This can occasionally lead to large values of  $S$  if  $\mathcal{B}_\alpha(\alpha_i)$  is close to zero, which occurs if an event has a value  $\alpha_i$  close to zero. However, assuming that the background is isotropic—as done in the toy model—the probability that this occurs for  $\boldsymbol{\alpha}$  containing only background is rather low. On the other hand, since the detected signal is

expected to follow a Gaussian peaking at the true source location, a signal event has a relatively large probability to obtain an  $\alpha_i$  close to zero. Note that in order to avoid a possible division by zero in Eq. 4.6, a lower limit of  $10^{-8}$ , corresponding with numerical accuracies, was set for the values of  $\mathcal{B}_\alpha(\alpha)$  for the smallest  $\alpha$ .

Thus, signal events are expected to give these “outliers” in  $S$ -values much more frequently than background events. For example, in Fig. 4.3b the resulting value of  $S$  is, in this particular case, about an order of magnitude higher than the value of  $S$  in Fig. 4.3a. As a consequence, it will be interesting to investigate the development of the probability that a significance of  $5\sigma$  is obtained as function of the amount of signal events. In contrast, this probability is fixed at 50% for the determination of the Discovery Potential. Moreover, the convergence towards the DP is expected to be rather slow, since for more signal events the value of  $S$  does not increment much in Eq. 4.6, except if such an “outlier” is present. This will be taken into consideration when examining the PDFs of  $S$  (and the other test statistics) in Chapter 5.

## 4.4 $\psi$ Method

Up to this point, the discussed statistical methods do not require a binning of the data to obtain their test statistic. As an alternative, the binning-based  $\psi$  method, originally developed in [83], is considered here. Moreover, it is based on a Bayesian assessment of significance, which serves as a complementary distinction with the previously mentioned methods, where a frequentist reasoning was used to construct the test statistics.

### 4.4.1 Assessment of Significance in a Bayesian Approach

The  $\psi$  method uses the full power of the theorem of Bayes to construct its test statistic. Let  $\mathcal{H}_0$  be the hypothesis that the measurements  $D = \alpha$  only consist of background, and  $\mathcal{H}_1$  any alternative to  $\mathcal{H}_0$ . Furthermore, suppose some prior information  $I$  is known. Then, the evaluation of Eq. 2.3 leads to

$$\begin{aligned}\mathcal{P}(\mathcal{H}_0|DI) &= \mathcal{P}(\mathcal{H}_0|I) \frac{\mathcal{P}(D|\mathcal{H}_0I)}{\mathcal{P}(D|I)}, \\ \mathcal{P}(\mathcal{H}_1|DI) &= \mathcal{P}(\mathcal{H}_1|I) \frac{\mathcal{P}(D|\mathcal{H}_1I)}{\mathcal{P}(D|I)}.\end{aligned}\tag{4.7}$$

One can now define the *evidence*  $e(\mathcal{H}_1|DI)$  for the alternative  $\mathcal{H}_1$  relative to the background-only  $\mathcal{H}_0$ , given the data  $D$  and prior information  $I$ , by taking the ratio of Eqs. 4.7 on a decibel scale,

$$\begin{aligned}e(\mathcal{H}_1|DI) &= 10 \log_{10} \left[ \frac{\mathcal{P}(\mathcal{H}_1|DI)}{\mathcal{P}(\mathcal{H}_0|DI)} \right] \\ &= 10 \log_{10} \left[ \frac{\mathcal{P}(\mathcal{H}_1|I)}{\mathcal{P}(\mathcal{H}_0|I)} \right] + 10 \log_{10} [\mathcal{P}(D|\mathcal{H}_1I)] - 10 \log_{10} [\mathcal{P}(D|\mathcal{H}_0I)] \\ &\leq 10 \log_{10} \left[ \frac{\mathcal{P}(\mathcal{H}_1|I)}{\mathcal{P}(\mathcal{H}_0|I)} \right] - 10 \log_{10} [\mathcal{P}(D|\mathcal{H}_0I)] \\ &= e(\mathcal{H}_1|I) + \psi.\end{aligned}\tag{4.8}$$

Here, the inequality was obtained using that  $0 \leq \mathcal{P} \leq 1$ . Also,  $e(\mathcal{H}_1|I)$  was used to denote the evidence for  $\mathcal{H}_1$  relative to  $\mathcal{H}_0$  given  $I$ , and the Bayesian observable  $\psi$  was defined as

$$\psi = -10 \log_{10} [\mathcal{P}(D|\mathcal{H}_0I)]. \quad (4.9)$$

The result of Eq. 4.8 means that the evidence for  $\mathcal{H}_1$  relative to  $\mathcal{H}_0$  given  $D$  and  $I$  cannot exceed  $e(\mathcal{H}_1|I) + \psi$ . Since  $e(\mathcal{H}_1|I)$  does not depend on the data  $D$ , it is  $\psi$  which quantifies the degree of belief in  $\mathcal{H}_0$ . This can be seen by taking two datasets  $D_1$  and  $D_2$  with respective  $\psi$ -values  $\psi_1$  and  $\psi_2$ , and supposing that  $\psi_1 < \psi_2$ . In that case,  $e(\mathcal{H}_1|D_1I) < e(\mathcal{H}_1|D_2I)$ , because  $e(\mathcal{H}_1|I)$  remains the same for the two datasets. Thus, the evidence for  $\mathcal{H}_1$  relative to  $\mathcal{H}_0$  given  $D = D_1$  and  $I$  is less than in the case  $D = D_2$ . As a consequence, the degree of belief in  $\mathcal{H}_0$  is less if the value of  $\psi$  given the data  $D$  is larger. Therefore,  $\psi$  is the quantity used as a test statistic here. Note that although the construction of  $\psi$  is based on Bayesian principles, the evaluation of  $\psi$  as a test statistic will still follow the frequentist approach outlined in Chapter 2.

#### 4.4.2 Binning of the Expected Angular Background Distribution

Consider the expected background-only distribution  $\mathcal{B}_\alpha$  given in Eq. 4.3, and divide the measurements  $\alpha$  in  $m$  bins over the patch of size  $\alpha_{\max}$ , where the  $k^{\text{th}}$  bin has a width  $w_k = \alpha_k - \alpha_{k-1}$  with  $k \in \{1, 2, \dots, m\}$ . A certain  $\alpha_k$  indexed by  $k$  indicates one of the bin-edges, *not* a measurement  $\alpha_i$  of  $\alpha$  in the patch, indexed by  $i \in \{1, 2, \dots, n\}$ . Then the expected probability  $p_k$  of a background event  $\alpha$  to fall in bin  $k$  is given by the integration of  $\mathcal{B}_\alpha$  over the bin,

$$\begin{aligned} p_k &= \int_{\alpha_{k-1}}^{\alpha_k} \frac{\sin \alpha'}{1 - \cos \alpha_{\max}} d\alpha' \\ &= \frac{\cos w_{k-1} - \cos w_k}{1 - \cos \alpha_{\max}}. \end{aligned}$$

By definition,  $\sum_{k=1}^m p_k = 1$ .

The probability  $\mathcal{P}(D|\mathcal{H}_0I)$  that the measurements  $D = \alpha$  is supported by the background-only hypothesis  $\mathcal{H}_0$  is now given by a multinomial distribution [73],

$$\mathcal{P}(D|\mathcal{H}_0I) = \frac{n!}{n_1! n_2! \dots n_m!} p_1^{n_1} p_2^{n_2} \dots p_m^{n_m},$$

where  $n_k$  is the number of entries in bin  $k$ , such that  $\sum_{k=1}^m n_k = n$ . Direct substitution in Eq. 4.9 then leads to

$$\psi = -10 \left[ \log_{10} n! + \sum_{k=1}^m (n_k \log_{10} p_k - \log_{10} n_k!) \right],$$

which is how the  $\psi$  test statistic is determined in this work.

However, the binning should be approached with care, since it can influence the determined  $p$ -value for an actually observed dataset [84]. To avoid this, the bin widths  $w_k$  should be chosen small enough such that in most of the cases (preferably all),  $n_k \leq 1$ . For a patch size  $\alpha_{\max}$  of some degrees, an appropriate way to take this into account is

by setting the number of bins  $m = 10^4$ , each with a constant width  $w_k = w = \alpha_{\max}/m$ . As an example, Figs. 4.6a and 4.6b show the applied binning after choosing  $\alpha_{\max} = 5^\circ$ , respectively for 0 and 20 signal injections with the toy model. The resulting “bar code plots” respectively contain no and one bin with  $n_k > 1$ , for these particular examples. The maximal amount of times that  $n_k > 1$  per background-only pseudo-experiment in a set of  $10^3$  simulations was determined to be 8, which is reasonable for  $m = 10^4$ .

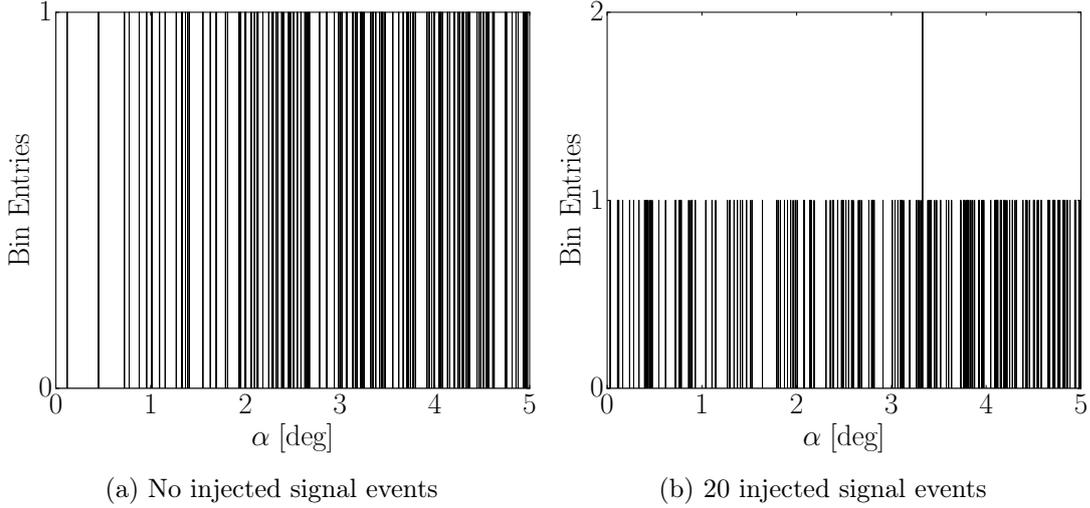


Figure 4.6: The “bar code plots” illustrating the binning used in the determination of the  $\psi$  test statistic with  $\alpha_{\max} = 5^\circ$ , for 0 and 20 injected signal events using the toy model, respectively depicted in Figs. 4.6a and 4.6b. The requirement that the majority of bins should contain at most one (simulated) measurement is clearly fulfilled.

# Results of the Toy Model

## Introduction

In this Chapter, the four different statistical methods are compared using pseudo-experiments performed with the toy model. The full simulation code used for this purpose can be found in Appendix A. For that, the angular resolution was chosen as  $\sigma_\alpha = 1^\circ$ , according to the ICECUBE angular resolution for muon neutrinos. Furthermore, the size of the patch scanned for a signal signature was taken as  $\alpha_{\max} = 5\sigma_\alpha = 5^\circ$ , to ensure that almost all of the signal events are contained within the patch, as is typically done in ICECUBE point source analyses.

The comparison of the four methods is presented in Section 5.1, where different arguments will be considered to determine the optimal method(s). After that, the implications of the obtained results on the current status of steady point source searches, which was discussed in Chapter 1, are given in Section 5.2.

## 5.1 Comparison of the Statistical Methods

### 5.1.1 Background-Only Distributions

Firstly, the background-only PDFs of the different test statistics have been determined by performing  $10^9$  pseudo-experiments without signal injection. This ensured that a  $5\sigma$  significance, required for the determination of Discovery Potentials, was obtained with enough statistical precision, as argued in Chapter 2.

The resulting histograms of the four distributions are plotted in Fig. 5.1, where the probability densities have been calculated as the normalized bin entries. Note that negative values of the likelihood test statistic  $\lambda$  are included in the first positive bin, as is commonly done in ICECUBE analyses. Furthermore, it can be remarked that the scale of the PLT test statistic  $S$  is logarithmic in contrast with the other test statistics, which is expected from its construction in Chapter 4.

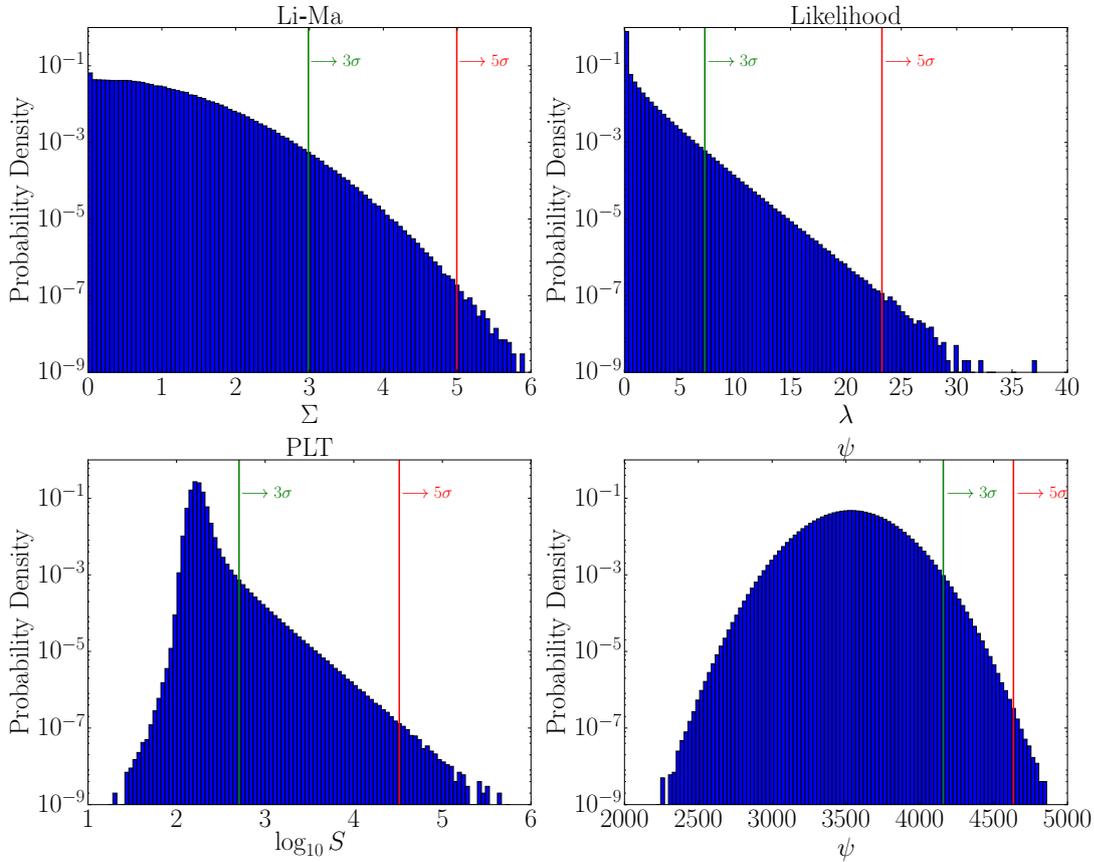


Figure 5.1: The background-only PDFs of the four different test statistics considered in this work. The test statistic values corresponding with both  $3\sigma$  and  $5\sigma$  significances are also shown. It can be seen that in all cases a good statistical precision is achieved up to a  $5\sigma$  significance.

For each test statistic, it can be seen from Fig. 5.1 that up to a  $5\sigma$  significance the development of the distributions are relatively well-determined, as required. Above  $5\sigma$  the distributions become less stable; the probability densities per bin fluctuate since the bin entries are low. Consequently, above  $5\sigma$  there is a relatively large uncertainty on the determined probability density. The reason why these uncertainties are not indicated is because they will not influence the results of the presented study, since at most a  $5\sigma$  significance ( $p = 5.73 \times 10^{-7}$ ) is needed in what follows, up to which the uncertainties can be neglected. Note that this is due to the fact that  $10^9$  background-only pseudo-experiments were performed.

### 5.1.2 Discovery Potentials and Sensitivities

In order to compare the statistical methods, for each signal injection  $N_s \in [1, 200]$ , representing the simulated amount of signal events in one year, the four signal-plus-background PDFs were determined. These were obtained through the performance of

$10^5$  pseudo-experiments per value of  $N_s$ . Some of the signal-plus-background PDFs are depicted in Appendix B.

First, the 50% DPs at  $3\sigma$  and  $5\sigma$ , and the 90% DP at  $5\sigma$  were determined for each test statistic. Figs. 5.2a and 5.2b respectively show the development of the significance exceeded by 50% and 90% of the trials in function of  $N_s$ , indicating the different thresholds ( $3\sigma$  and  $5\sigma$ ) corresponding with each DP. It can be seen that the likelihood method performs best in all cases, since for each DP it requires the least amount of signal  $N_s$  in order to reach the DP threshold. Furthermore, the  $5\sigma$  DPs are not reached by the PLT method for  $N_s \leq 200$ , which at small  $N_s$ -values is comparable with the likelihood. Because of the relatively poor performance of the PLT method for the  $5\sigma$  DPs, which was also anticipated in Chapter 4, their values (corresponding with values  $N_s > 200$ ) were not determined here.

It should be mentioned that in Fig. 5.2 one can observe that the significances do not always rise smoothly when incrementing  $N_s$ —see the PLT method, for example. This is due to the fact that statistical fluctuations have an influence in the comparison of the significances for e.g.  $N_s = 50$  and  $N_s = 51$ . In other words, there is a statistical uncertainty on the determined significances. However, they do not influence the conclusions drawn here that the likelihood performs best. Yet, in practice one would typically perform e.g. five sets of  $10^5$  pseudo-experiments for this example, and then take the average of the five significances obtained for each  $N_s$ , per test statistic. Note that this remark will also be valid for upcoming plots, although it will not be reemphasized.

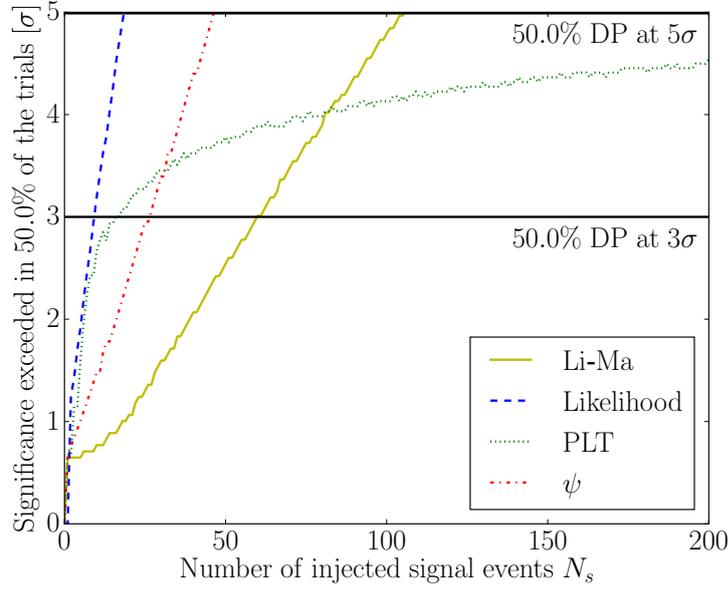
Subsequently, the sensitivities were determined for each of the test statistics. Instead of significance, Fig. 5.3 plots the upper  $p$ -value for 90% of the trials as function of  $N_s$ , thereby also showing the sensitivity threshold (50%). In this case, both the likelihood and the PLT methods perform the best, since they trespass the sensitivity threshold at the lowest value of  $N_s$  compared with the other statistical methods.

To put everything into perspective, Fig. 5.4 shows the sensitivities and all DPs obtained for all statistical methods through the conversion of  $N_s$  to an injected signal rate  $\dot{N}_s = N_s/(1 \text{ yr})$ . It is clear that the likelihood method performs the best overall, since it has the lowest values for  $\dot{N}_s$  corresponding with the sensitivity and the DPs. The PLT method does well for the sensitivity and  $3\sigma$  DP, but the worst for the  $5\sigma$  DPs. Moreover, it can be seen that overall the  $\psi$  method performs better than the Li-Ma method. A comparison per test statistic of the signal-plus-background PDFs with the background-only PDFs for all the  $N_s$ -values shown in Fig. 5.4 can be found in Appendix B, where these  $N_s$ -values and corresponding  $\dot{N}_s$ -values are also tabulated.

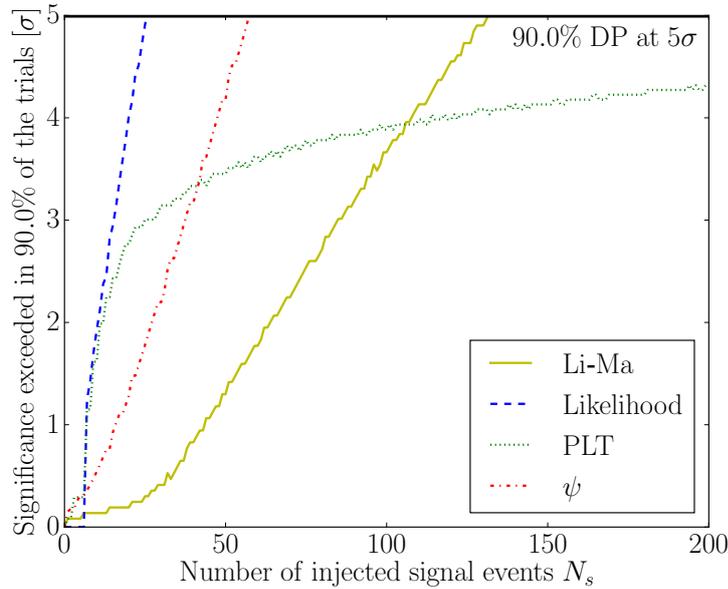
### 5.1.3 Probabilities for Discovery

In Chapter 2 it was mentioned that the best statistical method is not solely defined by the one with the smallest DP (sensitivities will not be considered in the remainder of this discussion), where 50% or 90% of the trials are required to exceed a  $3\sigma$  or  $5\sigma$  significance. It is also useful to investigate which method performs best for an exceeding percentage of e.g. 10%, that is, the method for which the least injected signal is required such that the probability for a  $3\sigma$  or  $5\sigma$  significance would be 10%.

Thus, in contrast with the development of significance for a fixed exceeding percentage shown in Fig. 5.2, the exceeding percentage was plotted for a fixed significance



(a) 50% exceeding probability



(b) 90% exceeding probability

Figure 5.2: The plots of the significance exceeded in 50% of the trials, illustrated in Fig. 5.2a, and in 90% of the trials, depicted in Fig. 5.2b, as function of the amount of injected signal events  $N_s$ . The thresholds corresponding with the different DPs are also shown (horizontal black lines). It can be seen that the likelihood method behaves the best, since it requires the least amount of signal injections in order to obtain all different DPs.

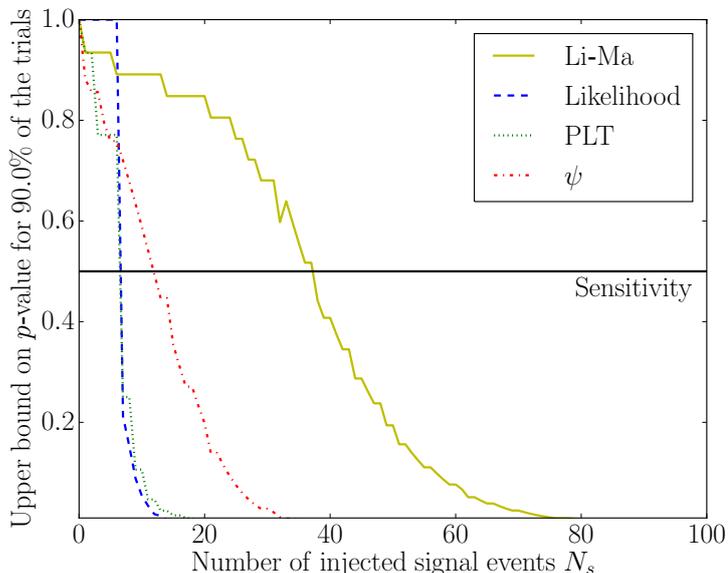


Figure 5.3: The  $p$ -value exceeded in 90% of the trials as function of the amount of injected signal events  $N_s$ . The threshold corresponding with the sensitivity is also indicated. One can observe that both the likelihood and PLT methods perform the best, by trespassing the threshold for the sensitivity at the smallest value of  $N_s$  compared with the Li-Ma and  $\psi$  methods.

as a function of  $N_s$ . Figs. 5.5a and 5.5b respectively show these dependencies for fixed significances of  $3\sigma$  or  $5\sigma$ , thereby also indicating the different DP thresholds (50% and 90%). For the  $3\sigma$  case, it can be seen that the PLT and likelihood essentially coincide up to an exceeding probability of about 10%, after which the likelihood method dominates. The Li-Ma and  $\psi$  methods perform worse than the likelihood and the PLT methods.

Yet, the more relevant scenario is that of a  $5\sigma$  significance, as this is the significance required for a discovery. From Fig. 5.5b it can be observed that for the major part, the likelihood method performs best. However, for the lowest values of  $N_s$ —a zoomed version of Fig. 5.5b for the smallest  $N_s$ -values is given in the next Section—the PLT method has a higher probability of discovery than the other methods, up to approximately 0.1% after which the likelihood method starts to dominate. This behavior of the PLT method was anticipated in Chapter 4. For the larger  $N_s$ -values, the PLT method also becomes worse compared with the Li-Ma and  $\psi$  methods.

## 5.2 Implication for Steady Point Source Studies

With the use of Fig. 1.9, in Chapter 1 an estimate was made of the steady point source types that could solely be responsible for the diffuse high-energy cosmic neutrino flux measured with ICECUBE. Recall that to do so, the flux of the nearest candidate steady point source of a certain population was determined assuming a strong correlation

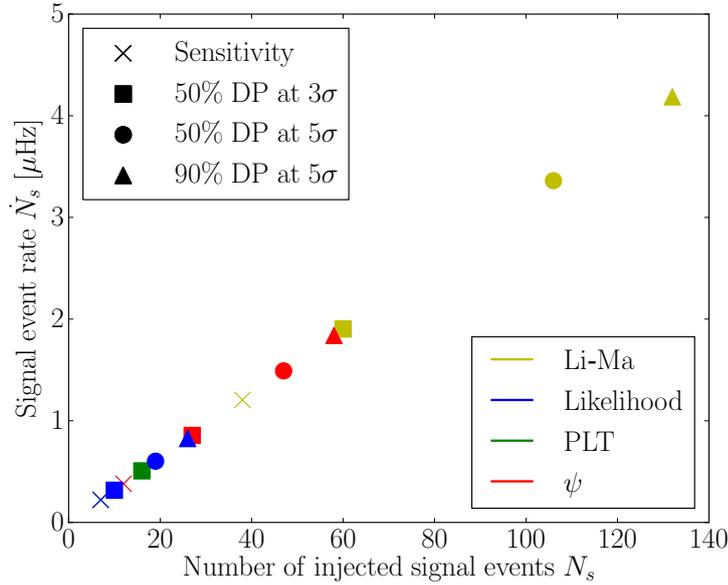


Figure 5.4: A plot showing the sensitivities and all DPs for the four statistical methods, with the exception of the two  $5\sigma$  DPs for the PLT method, which were not determined. Also note that the likelihood and PLT sensitivities coincide. It can be concluded from this comparison of the different methods that the likelihood performs best overall.

between high-energy electromagnetic and high-energy neutrino emission. Also, the electromagnetic emission per volume from the entire population was required to at least equalize the emission per volume corresponding with the diffuse flux.

The result of this estimation was that if there would be such a steady point source population, it would most likely be one with a relatively high population density  $\rho$ , but a rather low neutrino emission rate per source. Therefore, the event rate  $\dot{N}_s$  that could be detected by ICECUBE for such a candidate source would be relatively small. The goal is now to obtain a rough estimate of  $\dot{N}_s$  per source for the six different steady point source populations of Fig. 1.9, and see where they fall in the context of Fig. 5.5b. Then it can be determined, based on the latter plot, which statistical method would give the highest probability for a discovery per singular candidate source of each type.

First, for each population, the number of sources  $N_{\text{sources}}$  within an arbitrarily chosen distance  $R = 100$  Mpc was estimated as

$$N_{\text{sources}} = \rho \frac{4\pi R^3}{3},$$

where  $\rho$  was taken directly from Fig. 1.9. Next, it was assumed that all the sources within 100 Mpc would be responsible for the observed diffuse flux, which has a rate  $\dot{N}_{\text{diffuse}} = 37/(988 \text{ days})$ , as given in [43]. Therefore, the event rate per source was roughly approximated as  $\dot{N}_s = \dot{N}_{\text{diffuse}}/N_{\text{sources}}$ , without taking into account differences in neutrino flavours.

The resulting values of  $\dot{N}_s$  for the six different considered candidate sources are given in Tab. 5.1. Here, it is furthermore indicated which sources have been excluded

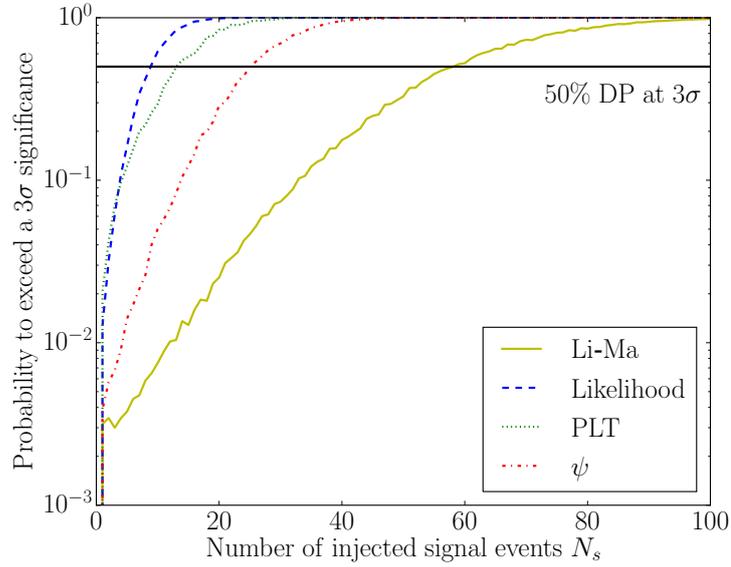
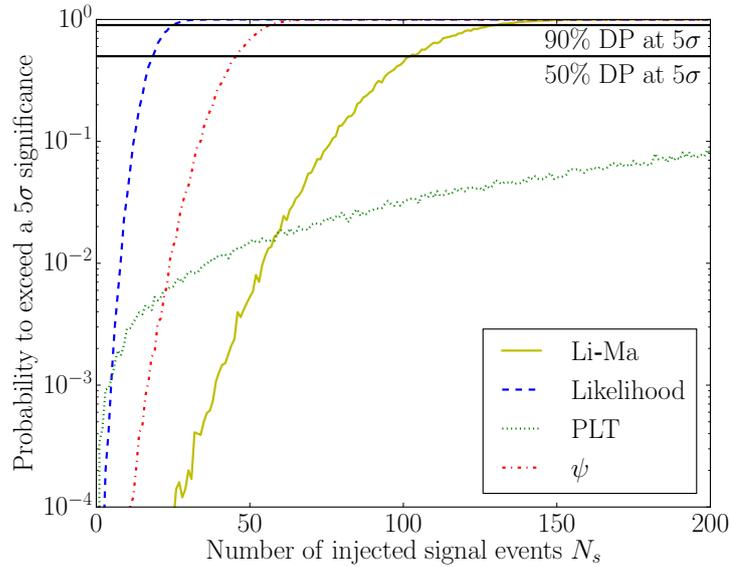
(a)  $3\sigma$  significance(b)  $5\sigma$  significance

Figure 5.5: The plots depicting the probability to exceed  $3\sigma$  and  $5\sigma$  significances, respectively shown in Figs. 5.5a and 5.5b. Also here the threshold for the different DPs are indicated. Mostly, the likelihood reaches a certain exceeding probability the fastest as function of  $N_s$ . However, in the  $5\sigma$  case, it can be seen that for exceeding probabilities below 0.1% the PLT method performs best. A zoomed version of Fig. 5.5b can be found in Fig. 5.6.

as the sole origin of the observed diffuse flux based on Fig. 1.9. For the different  $\dot{N}_s$ -values, one can determine which statistical method would give the highest probability for a discovery through a direct comparison with Fig. 5.6, which is a zoomed version of Fig. 5.5b for injected signal rates. The results are also included in Tab. 5.1.

One can conclude that for the three candidate steady point source types with the highest population densities, and thus lowest expected neutrino emission rate per source, the PLT method might give a higher probability for a  $5\sigma$  discovery when analyzing a single source of that type. Note that these three populations, i.e. galaxy clusters, starburst galaxies, and FRI galaxies, are those which were not excluded by the argumentation given in Chapter 1.

Thus, the PLT method might be considered for future steady point source studies with ICECUBE, even though the probability of detecting such low high-energy cosmic neutrino fluxes per source is small for a detection volume of  $1 \text{ km}^3$ . Larger detectors will be needed to significantly enhance this detection probability if no discoveries of the high-energy cosmic neutrino origin are made in the near future.

Furthermore, it should be remarked that one can also apply a *stacking* of the data, which is typically done in ICECUBE analyses if low signal rates are expected. This means that for each candidate steady point source of a population one takes a patch with size  $\alpha_{\text{max}}$  centered on the source, and then “stacks” all the different patches on top of each other. One would therefore essentially scan one patch in which the data corresponding with all candidate sources are combined.

Stacking would thus increase the amount of signal events on the patch, which would make the “shape” of a possible signal more prominent w.r.t. the background, although the signal-to-noise ratio would remain the same. Consequently, the likelihood and  $\psi$  methods are expected to improve from stacking (see for instance [83]), in contrast with the Li-Ma method, which is not sensitive to this “shape”. Moreover, the probability to

Table 5.1: A summary of parameters and conclusions drawn for six different populations of candidate steady point sources. For each source type these include the population density  $\rho$ ; if the population is excluded as the sole population for the observed diffuse neutrino flux; the estimated event rate  $\dot{N}_s$  that would be observed per source; and the statistical method which would give the highest probability for a discovery. This tabular can be compared directly with Fig. 5.6.

Population	$\rho$ [ $\text{Mpc}^{-3}$ ]	Excluded	$\dot{N}_s$ [ $\mu\text{Hz}$ ]	Method
FSRQ	$1 \times 10^{-9}$	Yes	$1.0 \times 10^2$	Likelihood
BL Lac	$3 \times 10^{-8}$	Yes	3.4	Likelihood
FRII	$5 \times 10^{-8}$	Yes	2.1	Likelihood
G. Clusters	$2 \times 10^{-6}$	No	$5.2 \times 10^{-2}$	PLT
Starburst	$6 \times 10^{-5}$	No	$1.7 \times 10^{-2}$	PLT
FRI	$3 \times 10^{-4}$	No	$3.4 \times 10^{-4}$	PLT

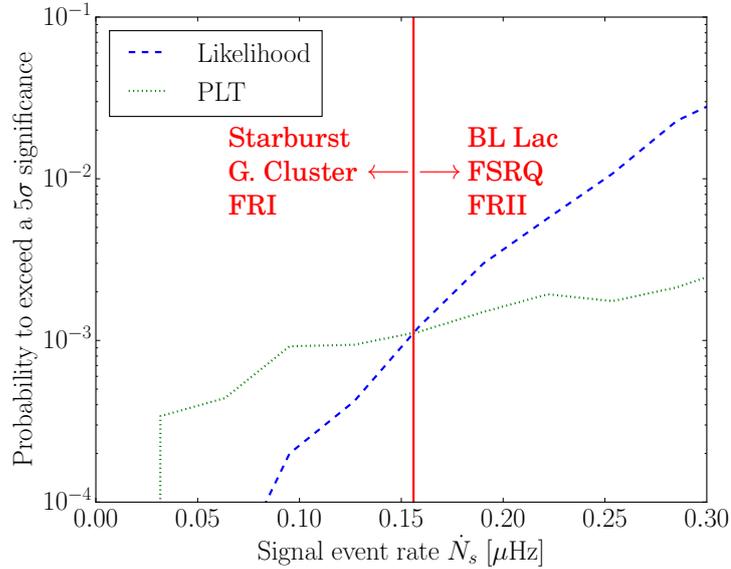


Figure 5.6: The plot given in Fig. 5.5b, for signal rates  $\dot{N}_s$ , zoomed around the region where the comparison is made between the likelihood and PLT methods. For the six different candidate steady point source populations, it is indicated with the red vertical line which method could give the highest probability for a  $5\sigma$  discovery in a single-source analysis.

obtain a  $5\sigma$  significance with the PLT method might increase slightly. Nevertheless, a future study (using the toy model) should be performed to quantitatively investigate the effect of stacking on the statistical methods.

Finally, it should be mentioned that a threshold of 0.1% for the discovery probability is rather low in order to make a decisive statement on the PLT method being favourable in future analyses. However, the choices of  $\alpha_{\max} = 5^\circ$  and  $\sigma_\alpha = 1^\circ$ , which are typical in ICECUBE analyses since these are optimal values for the standardly used likelihood method, are expected to play a role in the comparison given here. Therefore, altering  $\alpha_{\max}$  and  $\sigma_\alpha$  might give different results; the Li-Ma and  $\psi$  methods could then also play an important role in the comparison. Unfortunately, such an investigation has not been performed here, as it would require a significant amount of additional computation time that does not fall within the time scope of this thesis.



## Conclusions and Outlook

In this work, a toy model of steady point sources was constructed with which to compare four different statistical methods through pseudo-experiments. Only the angular distributions of events were taken into consideration. Background events were generated isotropically, whilst signal events were smeared w.r.t. the location of a fictitious steady point source by taking into account the angular resolution  $\sigma_\alpha$  of the detector. The test statistics corresponding with the different methods were then determined by scanning a certain patch with size  $\alpha_{\max}$  centered on the fictitious source. For consistency with standard ICECUBE analyses, it was chosen to set  $\alpha_{\max} = 5\sigma_\alpha = 5^\circ$ .

Of the considered methods, the Li-Ma method only took into account the amount of events within the patch, without applying any angular information. Comparatively, the likelihood method, which is the standard statistical method applied in (steady) point source analyses with ICECUBE, did consider the angular distribution of the events. This was also done with the PLT method, in which essentially the derivative of the likelihood was considered. Finally, to provide some contrast with these three unbinned methods, the  $\psi$  method was based on a binning principle of the angular distributions.

The first main conclusion of the statistical method comparison was that the likelihood method performs best overall in the context of Discovery Potentials (DPs) and sensitivities. It is the method for which the 50% DPs at  $3\sigma$  and  $5\sigma$ , and the 90% DP at  $5\sigma$  were determined to be the lowest. Moreover, it has the lowest sensitivity, although this value is the same as obtained with the PLT method, which performs well for small signal injections, but poorly for larger signal injections. At these large values, the  $\psi$  and Li-Ma methods are better than the PLT. Moreover, the  $\psi$  method performs better overall than the Li-Ma method.

Secondly, the likelihood method also performs well in the context of probabilities for a  $5\sigma$  discovery. However, for the smallest injected signals, up to a discovery probability of about 0.1%, the PLT method performs best. This corresponds well with its interpretation as a “needle in a haystack” method. For the largest injected signals, the PLT method is the worst one, and overall the  $\psi$  method performs better than the Li-Ma method.

As a consequence, it was investigated if there are populations of candidate steady point sources for which the PLT method might be preferred for an analysis. To do so, for six populations a rough estimation was made of the signal event rate required per source within 100 Mpc in order to be fully accountable for the diffuse high-energy cosmic neutrino flux measured with ICECUBE.

It was found that for three populations, i.e. galaxy clusters, starburst galaxies, and FRI galaxies, the PLT method could be considered in future analyses of these candi-

date sources. Moreover, these source types have relatively large population densities but small expected high-energy neutrino emission rates, which were determined to be the candidate steady point sources that could still likely be solely responsible for the observed diffuse flux. This is based on the fact that up to this date, no steady point sources have been found that could be the origin of this diffuse flux.

However, it can be argued that less than an 0.1% probability for a discovery might not be enough in order to consider the PLT method in future analyses. Yet, the standard choices  $\alpha_{\max} = 5^\circ$  and  $\sigma_\alpha = 1^\circ$  are typically made in ICECUBE analyses since these are optimal values with which to perform a likelihood analysis. Therefore, a first outlook for future work is to consider different combinations of  $\alpha_{\max}$  and  $\sigma_\alpha$ —a value  $\sigma_\alpha = 10^\circ$  would e.g. correspond with the angular resolution of a cascade in ICECUBE—in order to determine which values of these parameters optimize each method.

In particular, for a small  $\alpha_{\max}/\sigma_\alpha$  ratio, the amount of signal that will fall in the scanned patch will be relatively small. As such, the PLT method might give higher probabilities for a discovery, since its strength lies in very low signals compared with a dominating background. This will become interesting for future ICECUBE analyses if these probabilities are found to be in the 1% to 10% regime. Note however that the signal event rates for which the PLT method becomes prominent might be too low to be detectable with ICECUBE in the near future, meaning that a larger detector might be required to discover a signal from the corresponding candidate steady point sources.

Furthermore, for low expected signal rates one would typically stack the data of different candidate steady point sources in ICECUBE analyses. Consequently, the effect of this stacking on the four statistical methods should also be investigated in the future using the toy model. This is expected to be most beneficial for the likelihood and  $\psi$  methods. In addition, it might also be useful to include energy distributions in the toy model, as this is typically done in ICECUBE (steady) point source analyses, and which is expected to significantly influence the different test statistics.

In short, it can be concluded that for  $\alpha_{\max} = 5\sigma_\alpha = 5^\circ$  the likelihood method has the best DPs, and also the best sensitivity together with the PLT method. Furthermore, for the lowest signal rates, the PLT is expected to give the highest probability for a  $5\sigma$  discovery, which makes the PLT method a possible candidate statistical method for future steady point source analyses with ICECUBE and larger detectors that are yet to be built.

## Summary

Since the discovery of high-energy cosmic neutrinos with the ICECUBE experiment, no sources have been found that are responsible for the detected diffuse high-energy cosmic neutrino flux. In particular, constraints have been put on the steady point sources that could be the sole origin of these high-energy cosmic neutrinos, favoring the sources belonging to a population with a high number density and a low expected high-energy neutrino emission rate per source.

Therefore, in this thesis it has been investigated if different statistical methods, i.e. the Li-Ma, PLT, and  $\psi$  methods, perform better than the standard maximum likelihood method used in ICECUBE analyses, with a particular interest for the best method at low signal rates. For that, a toy model of steady point sources has been developed with which to perform simulated pseudo-experiments based exclusively on the angular distributions of signal and background.

Overall, the likelihood method has been found to perform best for a scan over a patch of  $5^\circ$  on the sky around the source and an angular resolution of  $1^\circ$ . It has the lowest Discovery Potential and lowest sensitivity compared to the other methods, although the latter was found to be shared with the PLT method, which performs well for small signal event rates but poorly for large signal event rates. On the other hand, the  $\psi$  method globally performs better than the Li-Ma method. Moreover, it was determined that for the lowest signal event rates, the PLT method has the largest probability for a  $5\sigma$  discovery, up to about 0.1%, after which the likelihood method dominates.

Furthermore, it has been studied if the PLT method could be used in future steady point source analyses with ICECUBE. An estimate has been made of the expected high-energy neutrino rate per source for six different steady point source populations that could be responsible for the observed diffuse high-energy cosmic neutrino flux. It has been found that for FRI galaxies, starburst galaxies, and galaxy clusters the estimated rates are low enough for the PLT method to be considered as an alternative to the standard likelihood method. These three source types also correspond with candidate steady point source populations that have high population densities.



# Samenvatting in het Nederlands

Het bestaan van hoog-energetische kosmische neutrino's werd ontegensprekelijk bevestigd door hun ontdekking met het ICECUBE experiment. Nochtans zijn de bronnen van de gedetecteerde diffuse hoog-energetische kosmische neutrinoflux momenteel nog niet geïdentificeerd. De tot nu onsuccesvolle zoektochten naar de oorsprong van de hoog-energetische kosmische neutrino's hebben geleid tot beperkingen voor de mogelijke bronnen die deze diffuse flux zouden kunnen verklaren.

In het bijzonder zijn er limieten opgesteld voor de mogelijke populaties van vaste (punt)bronnen die verantwoordelijk zouden kunnen zijn voor de diffuse flux. De types van vaste puntbronnen die niet onder deze restricties vallen, en dus van bijzonder belang zijn voor toekomstige zoektochten naar vaste puntbronnen, komen overeen met objecten die een hoge populatiedichtheid hebben, en waarvan er per bron een lage neutrino-emissie wordt verwacht.

Voor deze lage emissies, die zouden corresponderen met een zwak signaal in de ICECUBE detector, is het mogelijk dat andere statistische methoden betere resultaten opleveren dan de likelihood methode die standaard in ICECUBE-analyses wordt gebruikt. Bijgevolg is er in deze thesis een vergelijking van statistische methoden gemaakt met een focus op de optimale methode voor een zwak signaal. Hiervoor werden, de Li-Ma, PLT, en  $\psi$  methoden gebruikt als mogelijke alternatieven voor de likelihood methode. Bovendien werd er een startmodel geconstrueerd voor de simulaties van pseudo-experimenten, waarbij er exclusief gebruik werd gemaakt van de hoekdistributies van het signaal en de achtergrond.

Een eerste resultaat van dit onderzoek is dat, voor data genomen in een lgebied op de hemel met een grootte van  $5^\circ$  gecentreerd op de bron en een hoekresolutie van  $1^\circ$ , de likelihood methode over het algemeen de beste resultaten oplevert. De likelihood methode heeft allereerst het hoogste potentieel om te leiden tot een ontdekking. Bovendien heeft de likelihood methode de beste gevoeligheid voor een signaaldetectie, hoewel dit optimum wordt gedeeld met de PLT methode, dewelke goed is voor zwakke signalen, maar het slechtst voor sterke signalen. De  $\psi$  methode is voor alle signaalsterktes beter dan de Li-Ma methode.

Een tweede resultaat is dat voor de zwakste signalen, de hoogste kans op een ontdekking wordt gegeven door de PLT methode, tot ongeveer een kans van 0.1%. Voor hogere kansen corresponderend met hogere signaalsterktes levert de likelihood methode de beste resultaten op.

Tenslotte werd er onderzocht of men de PLT methode zou kunnen gebruiken in toekomstige studies omtrent vaste puntbronanalyses met ICECUBE. Hiervoor werden er zes populaties van vaste puntbronnen beschouwd die mogelijk de diffuse hoog-

energetische kosmische neutrinoflux zouden kunnen verklaren. Voor elke populatie is er een afschatting gemaakt van de verwachte emissiefrequentie van hoog-energetische neutrino's per bron. Er is geconcludeerd geweest dat FRI sterrenstelsels, starburststelsels, en clusters van sterrenstelsels zouden kunnen corresponderen met een signaal dat laag genoeg is opdat men de PLT methode zou kunnen gebruiken als een alternatief voor de standaard likelihood methode. Deze bronnen komen bovendien overeen met vaste puntbronpopulaties die een hoge populatiedichtheid hebben.



## Simulation Code of Pseudo-Experiments with the Toy Model

This Appendix contains the Python script used for the Monte Carlo simulations of pseudo-experiments in the context of the toy model. For each performed pseudo-experiment, the four different test statistics considered in this work are calculated. After the completion of all pseudo-experiments, the simulated test statistic values are stored in an external file.

```
1 import numpy as np
2 import argparse
3 import os
4 import time
5 from scipy import integrate
6 from scipy.optimize import minimize
7 from scipy.special import factorial
8
9 #####
10 # TOY MODEL OF STATISTICAL METHODS #
11 #####
12
13 # This code is used to simulate pseudo-experiments via Monte Carlo techniques.
14 # A pseudo-experiment generates one year of IceCube atmospheric neutrino events
15 # as background. The signal is injected through a Gaussian distribution.
16 # Different test statistics are calculated for each pseudo-experiment, i.e.
17 # different statistical methods are applied. These are then written onto a file
18 # so that they can be analyzed in other scripts.
19
20
21 ##### ABBREVIATIONS #####
22
23 # TS = Test Statistic
24 # sim = simulation
25 # bkg = background
26 # sign = signal
27 # norm = normalization constant
28 # LLH = likelihood
29 # MC = Monte Carlo
30
31
32 ##### DEFINITIONS #####
33
```

```

34 # Simulate signal and background events and return the opening angles alpha of the events
    within a certain patch around a fictitious source
35 # Also return the amount of events (without alpha info, because only needed for the Li-Ma
    method) for a patch of the same size on the opposite side of the sky
36 def SimulateEvents(N_bkg,N_sign,alpha_max,mean_sign,sigma_sign):
37     # N_bkg = Number of background events
38     # N_sign = Number of injected signal events
39     # alpha_max [deg] = Opening angle of the considered patch = maximal allowed opening
    angle to accept
40     # mean_sign [deg] = Expectation value for the Gaussian signal PDF, i.e. the source
    location
41     # sigma_sign [deg] = Standard deviation for the Gaussian signal PDF
42
43     # Without loss of generality, consider source at spherical coordinates (theta,phi) =
    (0,0)
44     # In that case, alpha = theta
45     # phi will not be needed, since the TSs used only require theta
46
47     # List of opening angles
48     alphas = []
49
50     # Number of events in the patch that does not contain the source
51     N_off = 0
52
53     # The background is simulated isotropically on the sky, i.e. cos(alpha) is uniform
54     for i in range(0,N_bkg):
55         alpha = np.rad2deg( np.arccos( np.random.uniform(-1.,1.) ) ) # [deg]
56         if alpha < alpha_max: # Only accept the events within the specified patch that
            contains the source
57             alphas = np.append(alphas,alpha)
58         if alpha > 180. - alpha_max: # Only accept the events within the specified patch
            that does not contain the source
59             N_off += 1
60
61     # The signal is simulated according to a gaussian distribution centered on the source (
    phi = 0) located at mean_sign with a standard deviation sigma_sign
62     for i in range(0,N_sign):
63         alpha = np.abs( np.random.normal(mean_sign,sigma_sign) ) # [deg]
64         if alpha < alpha_max: # Only accept the events within the specified patch that
            contains the source
65             alphas = np.append(alphas,alpha)
66         if alpha > 180. - alpha_max: # Only accept the events within the specified patch
            that does not contain the source
67             N_off += 1
68
69     return alphas, N_off # [deg]
70
71
72 # Returns the background PDF value of a certain opening within a patch of size alpha_max
73 def BkgDistribution(alpha,alpha_max):
74     # alpha [deg] = Opening angle
75     # alpha_max [deg] = Opening angle of the considered patch
76
77     # Isotropic PDF, definition can be found in L. Brayer, Ph.D. thesis, 2015
78     return ( np.sin( np.deg2rad(alpha) )/(1. - np.cos( np.deg2rad(alpha_max) ) ) )/np.rad2
    deg(1) # [deg^-1]
79
80
81 # Returns the normalization factor for the signal PDF within a patch of size alpha_max
82 def NormalizeSignDistribution(alpha_max,mean_sign,sigma_sign):
83     # alpha_max [deg] = Opening angle of the considered patch
84     # mean_sign [deg] = Expectation value for the Gaussian signal PDF, i.e. the source
    location
85     # sigma_sign [deg] = Standard deviation for the Gaussian signal PDF
86
87     # Gaussian PDF centered around alpha = mean_sign, i.e. the source location
88     # Integrate Gaussian PDF over whole patch
89     return integrate.quad(lambda x: np.exp( ( (mean_sign - x)**2 )/(-2.*sigma_sign**2) ), 0
    , alpha_max)[0] # [deg]; "[0]" to get the value of the integral
90
91
92 # Returns the background PDF value of a certain opening angle alpha within a patch of size
    alpha_max with the correct normalization
93 def SignDistribution(alpha,alpha_max,mean_sign,sigma_sign,norm_sign):
94     # alpha [deg] = Opening angle
95     # alpha_max [deg] = Opening angle of the considered patch
96     # mean_sign [deg] = Expectation value for the Gaussian signal PDF, i.e. the source
    location

```

```

97 # sigma_sign [deg] = Standard deviation for the Gaussian signal PDF
98 # norm_sign [deg] = Normalization factor for the Gaussian signal PDF
99
100 # Gaussian PDF centered around alpha = mean_sign, i.e. the source location, definition
    can be found in L. Brayeur, Ph.D. thesis, 2015
101 return np.exp( ( (mean_sign - alpha)**2 )/(-2.*sigma_sign**2) )/norm_sign # [deg^-1]
102
103
104 # ----- LIKELIHOOD METHOD -----
105
106 # Returns the minus log likelihood of a set of simulated opening angles
107 # This function is coded such that scipy.optimize.minimize can work with it
108 def minuslogLLH(x, alphas, alpha_max, mean_sign, sigma_sign, norm_sign):
109     # x = Fraction of the number of signal events in the patch over the total (signal +
        background) number of events in the patch
110     # alphas [deg] = Array of simulated opening angles
111     # alpha_max [deg] = Opening angle of the considered patch
112     # mean_sign [deg] = Expectation value for the Gaussian signal PDF, i.e. the source
        location
113     # sigma_sign [deg] = Standard deviation for the Gaussian signal PDF
114     # norm_sign [deg] = Normalization factor for the Gaussian signal PDF
115
116     # Definition can be found in J. R. Braun, Ph.D. thesis, 2009
117     return (-1.)*sum( np.log( (1.-x[0])*BkgDistribution(alphas[:,alpha_max] + x[0]*
        SignDistribution(alphas[:,alpha_max,mean_sign,sigma_sign,norm_sign) ) ) )
118
119
120 # Returns the an array of the log likelihoods of a set of simulated opening angles for
    different x-values
121 # This function is used for plotting the log likelihood as function of x
122 def logLLH_ForPlotting(x_array, alphas, alpha_max, mean_sign, sigma_sign, norm_sign):
123     # x_array = Array of fractions of the number of signal events in the patch over the
        total (signal + background) number of events in the patch
124     # alphas [deg] = Array of simulated opening angles
125     # alpha_max [deg] = Opening angle of the considered patch
126     # mean_sign [deg] = Expectation value for the Gaussian signal PDF, i.e. the source
        location
127     # sigma_sign [deg] = Standard deviation for the Gaussian signal PDF
128     # norm_sign [deg] = Normalization factor for the Gaussian signal PDF
129
130     # Array of log likelihoods
131     logLLHs = []
132
133     for x in x_array:
134         logLLHs = np.append( logLLHs, -1.*minuslogLLH([x], alphas, alpha_max, mean_sign,
            sigma_sign, norm_sign) )
135         # The "[" are used because x[0] in minuslogLLH doesn't exist if x is a float
136
137     # Returns an array corresponding with the given x-values dedicated for plotting
138     return logLLHs
139
140
141 # Returns the lower log likelihood bound of x given a set of simulated opening angles
142 # Ensures that the log doesn't become infinite in scipy.optimize.minimize
143 def GetMinimizerLowerBound(alphas, alpha_max, mean_sign, sigma_sign, norm_sign):
144     # alphas [deg] = Array of simulated opening angles
145     # alpha_max [deg] = Opening angle of the considered patch
146     # mean_sign [deg] = Expectation value for the Gaussian signal PDF, i.e. the source
        location
147     # sigma_sign [deg] = Standard deviation for the Gaussian signal PDF
148     # norm_sign [deg] = Normalization factor for the Gaussian signal PDF
149
150     # Definition of the bound can be found in J. R. Braun, Ph.D. thesis, 2009
151
152     # List of lower bound candidates
153     bound_candidates = []
154
155     for alpha in alphas:
156         B = BkgDistribution(alpha, alpha_max)
157         S = SignDistribution(alpha, alpha_max, mean_sign, sigma_sign, norm_sign)
158         if S > B: # Condition to add to bound_candidates
159             bound_candidates.append( B/(B-S) )
160
161     # The lower bound is strict, but scipy.optimize.minimize uses closed intervals
162     # Add 0.01 (arbitrary choice) to the bound to take this into account, then the log will
        never become infinite in scipy.optimize.minimize
163     if len(bound_candidates) == 0:
164         return -1 + 0.01

```

```

165     else:
166         return max(bound_candidates) + 0.01
167
168
169 # Returns the Maximum Likelihood (ML) estimator for a set of simulated opening angles
170 def GetMLEstimator(alphas, alpha_max, mean_sign, sigma_sign, norm_sign):
171     # alphas [deg] = Array of simulated opening angles
172     # alpha_max [deg] = Opening angle of the considered patch
173     # mean_sign [deg] = Expectation value for the Gaussian signal PDF, i.e. the source
174         location
175     # sigma_sign [deg] = Standard deviation for the Gaussian signal PDF
176     # norm_sign [deg] = Normalization factor for the Gaussian signal PDF
177
178     # Get the lower bound for the minimizer
179     # This is not just 0 because statistic fluctuations in the background PDF can lead to
180     # negative values
181     lower_bound = GetMinimizerLowerBound(alphas, alpha_max, mean_sign, sigma_sign, norm_sign)
182
183     # The log likelihood is to be maximized, or, the minus log likelihood is to be minimized
184     # The upper bound of x is 1 by definition
185     minimum = minimize(minuslogLLH,(0.5,),args=(alphas, alpha_max, mean_sign, sigma_sign,
186         norm_sign), bounds=((lower_bound, 1.)),)
187
188     return minimum.x # ".x" to obtain the value of the minimum from the minimizer
189
190
191 # Returns the likelihood TS for a set of simulated opening angles
192 def GetLLH(alphas, alpha_max, mean_sign, sigma_sign, N_sign, norm_sign):
193     # alphas [deg] = Array of simulated opening angles
194     # alpha_max [deg] = Opening angle of the considered patch
195     # mean_sign [deg] = Expectation value for the Gaussian signal PDF, i.e. the source
196         location
197     # sigma_sign [deg] = Standard deviation for the Gaussian signal PDF
198     # N_sign = Number of injected signal events
199     # norm_sign [deg] = Normalization factor for the Gaussian signal PDF
200
201     # Get the correct normalization for the signal PDF
202     #norm_sign = NormalizeSignDistribution(alphas, alpha_max, mean_sign, sigma_sign)
203
204     # Obtain the ML estimator
205     MLEstimator = GetMLEstimator(alphas, alpha_max, mean_sign, sigma_sign, norm_sign)
206
207     # For the definition of the TS, see e.g. J. R. Braun, Ph.D. thesis, 2009
208     # We define -logLLH -> take into account in formula
209     return 2.*np.sign(MLEstimator)*(minuslogLLH([0.], alphas, alpha_max, mean_sign, sigma_sign,
210         norm_sign) - \
211         minuslogLLH([MLEstimator], alphas, alpha_max, mean_sign,
212         sigma_sign, norm_sign))
213
214
215 # ----- PLT METHOD -----
216
217 # Returns the minus score of a set of simulated opening angles
218 # This function is coded such that scipy.optimize.minimize can work with it
219 def minusScore(x, alphas, alpha_max, sigma_sign, norm_sign):
220     # x [deg] = Expectation value for the Gaussian signal PDF, i.e. the source location
221     # alphas [deg] = Array of simulated opening angles
222     # alpha_max [deg] = Opening angle of the considered patch
223     # sigma_sign [deg] = Standard deviation for the Gaussian signal PDF
224     # norm_sign [deg] = Normalization factor for the Gaussian signal PDF
225
226     # Calculate the values of the bkg PDF for the set of simulated opening angles
227     BkgValues = BkgDistribution(alphas[:, alpha_max])
228
229     # Set a lower bound on the bkg distribution values to not divide by zero in the PLT TS
230     for value in BkgValues:
231         if value < 1e-8:
232             value = 1e-8
233
234     # Definition can be found in R. S. Pilla, C. Loader, and C. Taylor, 2005
235     return (-1.)*sum( ( SignDistribution(alphas[:, alpha_max], x, sigma_sign, norm_sign) -
236         BkgValues ) / BkgValues )
237
238
239 # Returns the maximizer of the score for a set of simulated opening angles
240 def GetScoreMaximizer(alphas, alpha_max, sigma_sign, norm_sign):
241     # alphas [deg] = Array of simulated opening angles
242     # alpha_max [deg] = Opening angle of the considered patch

```

```

236 # sigma_sign [deg] = Standard deviation for the Gaussian signal PDF
237 # norm_sign [deg] = Normalization factor for the Gaussian signal PDF
238 # MLEstimator = Maximum Likelihood estimator calculated from the simulated opening
      angles
239
240 # The score is to be maximized, or, the minus score is to be minimized
241 minimum = minimize(minusScore,(0.5,),args=(alphas,alpha_max,sigma_sign,norm_sign),
      bounds=((0.,10.)),)
242
243 return minimum.x # ".x" to obtain the value of the minimum from the minimizer
244
245
246 # Returns the PLT TS for a set of simulated opening angles
247 def GetPLT(alphas,alpha_max,sigma_sign,N_sign,norm_sign):
248 # alphas [deg] = Array of simulated opening angles
249 # alpha_max [deg] = Opening angle of the considered patch
250 # sigma_sign [deg] = Standard deviation for the Gaussian signal PDF
251 # N_sign = Number of injected signal events
252 # norm_sign [deg] = Normalization factor for the Gaussian signal PDF
253
254 # Get the correct normalization for the signal PDF
255 #norm_sign = NormalizeSignDistribution(alphas,alpha_max,mean_sign,sigma_sign)
256
257 # Obtain the score maximizer
258 ScoreMaximizer = GetScoreMaximizer(alphas,alpha_max,sigma_sign,norm_sign)
259
260 # For the definition of the TS, see R. S. Pilla, C. Loader, and C. Taylor, 2005
261 # We define -Score -> take into account in formula
262 return (-1.)*minusScore(ScoreMaximizer,alphas,alpha_max,sigma_sign,norm_sign)
263
264
265 # ----- PSI METHOD -----
266
267 # Returns the probability that an angle falls in a certain bin based on the uniform
      background PDF
268 def GetBinProbability(N_bin,bin_width,alpha_max):
269 # N_bins = Number of bins for the histogram of opening angles
270 # bin_width [rad] = The width of the bins for the histogram of opening angles
271 # alpha_max [deg] = Opening angle of the considered patch
272
273 # The definition can be found in D. Bose, et al, 2014
274 return ( np.cos( (N_bin-1)*bin_width ) - np.cos( N_bin*bin_width ) )/( 1 - np.cos( np.
      deg2rad(alpha_max) ) )
275
276
277 # Returns the binning based psi TS for a set of simulated opening angles
278 def GetPsi(alphas,alpha_max,N_bins,N_sign):
279 # alphas [deg] = Array of simulated opening angles
280 # alpha_max [deg] = Opening angle of the considered patch
281 # N_bins = Number of bins for the histogram of opening angles
282 # N_sign = Number of injected signal events
283
284 # Make the histogram of opening angles
285 # Preferably you want N_bins to be large enough so that the histogram has 0 or 1 entry
      most of the time
286 hist, bin_edges = np.histogram(alphas,N_bins,(0,alpha_max))
287
288 # Calculate the bin width
289 bin_width = np.deg2rad(alpha_max/N_bins) # [rad]
290
291 # For the definition of the TS, see N. van Eijndhoven, 2008
292 psi = -10*sum( np.log10( range(1,len(alphas)+1) ) ) # [dB]
293
294 for k in range(0,N_bins):
295     bin_value = hist[k]
296     if bin_value != 0.:
297         psi += -10*( bin_value*np.log10( GetBinProbability(k+1,bin_width,alpha_max) ) -
            np.log10( factorial( bin_value ) ) )
298
299 return psi # [dB]
300
301
302 # ----- LI-MA METHOD -----
303
304 # Returns the Li-Ma TS for a set of simulated opening angles
305 def GetLiMa(alphas,N_off):
306 # alphas [deg] = Array of simulated opening angles

```

```

307     # N_off = Number of events that lie in a patch of size alpha_max that does not contain
           the source
308
309     # Calculate the total number events in the patches centered on and off the source
310     N_on = float(len(alphas))
311     N_off = float(N_off)
312
313     # For the definition of the TS, see Li and Ma, 1983
314     return np.sqrt( 2.*( N_on*np.log( 2.*N_on/(N_on + N_off) ) + N_off*np.log( 2.*N_off/(
           N_on + N_off) ) ) )
315
316
317     ##### APPLICATIONS #####
318
319     # Make a parser
320     parser = argparse.ArgumentParser()
321
322     # Argument (required) that allows to choose the amount of times the TSs have to be
           simulated
323     parser.add_argument("--sim",dest="sim",help="REQUIRED: give amount of times the TSs have to
           be simulated",required=True)
324
325     # Argument (required) that allows to choose the amount of signal you want to inject in your
           TS simulations
326     parser.add_argument("--sign",dest="sign",help="REQUIRED: give amount of signal that has to
           be injected",required=True)
327
328     # Argument (required) that specifies the set number of the file you are exporting
329     parser.add_argument("--set",dest="set",help="REQUIRED: give the set number of the export
           file",required=True)
330
331     # Argument (optional) that sets the angular size of the patch around the source
332     parser.add_argument("--patch",dest="patch",help="OPTIONAL: give the angular size of the
           patch around the source [deg]")
333
334     # Argument (optional) that sets the source location, i.e. the expectation value for the
           Gaussian signal PDF
335     parser.add_argument("--loc",dest="loc",help="OPTIONAL: give the source location (the polar
           angle) on the sky [deg]")
336
337     # Argument (optional) that sets the standard deviation for the Gaussian signal PDF
338     parser.add_argument("--sigma",dest="sigma",help="OPTIONAL: give the standard deviation for
           the Gaussian signal PDF [deg]")
339
340     arguments = parser.parse_args()
341
342     # Number of times you want to simulate
343     N_sim = int(arguments.sim)
344
345     # Number of signal injections you want to make
346     N_sign = int(arguments.sign)
347
348     # The set of the file you are exporting
349     file_set = str(arguments.set)
350
351     # Angular size of the patch around the source
352     if arguments.patch is None:
353         alpha_max = 5. # [deg]; DEFAULT
354     else:
355         alpha_max = float(arguments.patch) # [deg]
356
357     # Expectation value for the Gaussian signal PDF, i.e. the source location
358     if arguments.loc is None:
359         mean_sign = 0. # [deg]; DEFAULT
360     else:
361         mean_sign = float(arguments.loc) # [deg]
362
363     # Standard deviation for the Gaussian signal PDF
364     if arguments.sigma is None:
365         sigma_sign = 1. # [deg]; DEFAULT
366     else:
367         sigma_sign = float(arguments.sigma) # [deg]
368
369
370     # IceCube atmospheric neutrino background rate (roughly)
371     rate_bkg = 1./(6*60) # [Hz]; 10 per hour
372
373     # Average number of background events in 1 year

```

```

374 average_N_bkg = int(rate_bkg*365*24*3600)
375
376 # Set the seed for the random number generator
377 newseed = int( time.time()*1e6 - int(time.time()*1e-3)*1e9 ) # repeat period of the order
    of 10 years
378 np.random.seed(newseed)
379
380 # Number of bins, chosen large enough so that the histogram (for psi) has 0 or 1 entry most
    of the time
381 N_bins = 10000
382
383 # Normalization factor for the Gaussian signal PDF
384 norm_sign = NormalizeSignDistribution(alpha_max,mean_sign,sigma_sign) # [deg]
385
386 # Make an array of N_sim rows and 4 columns for each of the TSs
387 TS_array = np.zeros((N_sim,4))
388
389 # Get all TSs for a set of simulations
390 for i in range(0,N_sim):
391     # Obtain the number of bkg events in one year from a Poisson PDF with expectation value
    average_N_bkg
392     N_bkg = np.random.poisson(average_N_bkg)
393
394     # Simulate a set of opening angles
395     alphas, N_off = SimulateEvents(N_bkg,N_sign,alpha_max,mean_sign,sigma_sign)
396
397     # Fill in the TS array
398     TS_array[i][0] = GetLiMa(alphas,N_off)
399     TS_array[i][1] = GetLLH(alphas,alpha_max,mean_sign,sigma_sign,N_sign,norm_sign)
400     TS_array[i][2] = GetPLT(alphas,alpha_max,sigma_sign,N_sign,norm_sign)
401     TS_array[i][3] = GetPsi(alphas,alpha_max,N_bins,N_sign)
402
403
404 # Export the TS_array in an .npz file with the given set number
405
406 # Set the output directory where the .npz file will be exported, specified by the parser
    parameters
407 outdir = "/data/user/pcorcam/StatisticalMethods/MC_output_npz/signal_%i_patch_%f_mean_%
    f_sigma_%f/" % (N_sign,alpha_max,mean_sign,sigma_sign)
408
409 # Check if the output directory exists
410 if os.path.exists(outdir):
411     print "WARN: ** output dir already exists:", outdir
412     if not os.path.isdir(outdir):
413         raise RuntimeError("output directory exists and is not a directory")
414 else:
415     os.makedirs(outdir)
416
417 # Set the title of the .npz file
418 filetype = "MC_toymodel_signal_%i_patch_%f_mean_%f_sigma_%f_set_%s.npz" % (N_sign,
    alpha_max,mean_sign,sigma_sign,arguments.set)
419
420 # Save the TS_array in the .npz file, also include the parser parameters for easy
    identification
421 np.savez(outdir+filetype,[N_sign,alpha_max,mean_sign,sigma_sign],TS_array)

```



## Distributions for Discovery Potentials and Sensitivities

In this Appendix, some signal-plus-background distributions, determined with  $10^5$  pseudo-experiments, are plotted together with the respective background-only distributions, determined with  $10^9$  pseudo-experiments, for each test statistic. This is done for the values of the injected signal  $N_s$  corresponding with the sensitivities (Fig. B.1), the 50% DPs at  $3\sigma$  (Fig. B.2), the 50% DPs at  $5\sigma$  (Fig. B.3), and the 90% DPs at  $5\sigma$  (Fig. B.4). These  $N_s$ -values are given in Tab. B.1 together with the corresponding event rates  $\dot{N}_s = N_s/(1 \text{ yr})$ .

Table B.1: The values for the injected signal  $N_s$  and signal rate  $\dot{N}_s$  corresponding with the different DPs and sensitivities for all four statistical methods. The only exceptions are the  $5\sigma$  DPs of the PLT method, which were not determined.

Sensitivities	$N_s$	$\dot{N}_s$ [ $\mu\text{Hz}$ ]	50% DPs at $3\sigma$	$N_s$	$\dot{N}_s$ [ $\mu\text{Hz}$ ]
Li-Ma	38	1.2	Li-Ma	60	1.9
Likelihood	7	0.2	Likelihood	10	0.3
PLT	7	0.2	PLT	16	0.5
$\psi$	12	0.4	$\psi$	27	0.9
50% DPs at $5\sigma$	$N_s$	$\dot{N}_s$ [ $\mu\text{Hz}$ ]	90% DPs at $5\sigma$	$N_s$	$\dot{N}_s$ [ $\mu\text{Hz}$ ]
Li-Ma	106	3.4	Li-Ma	132	4.2
Likelihood	19	0.6	Likelihood	26	0.8
PLT	/	/	PLT	/	/
$\psi$	47	1.5	$\psi$	58	1.8

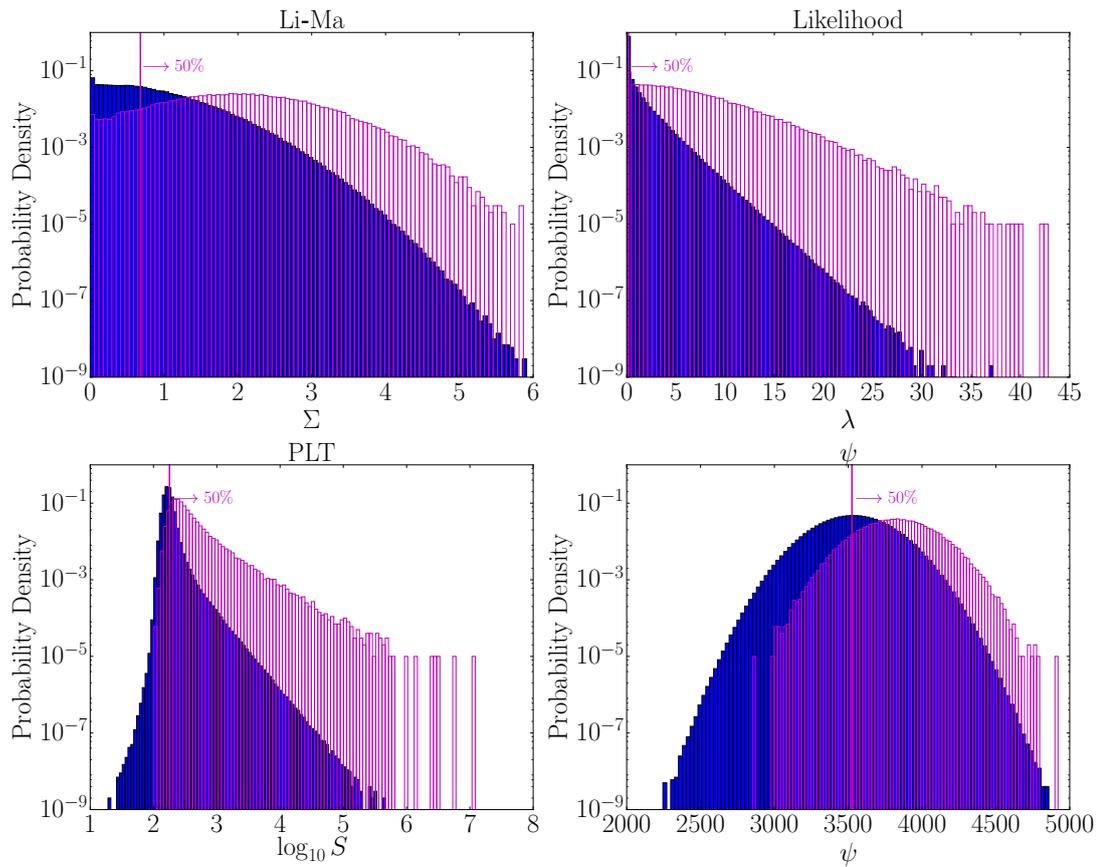
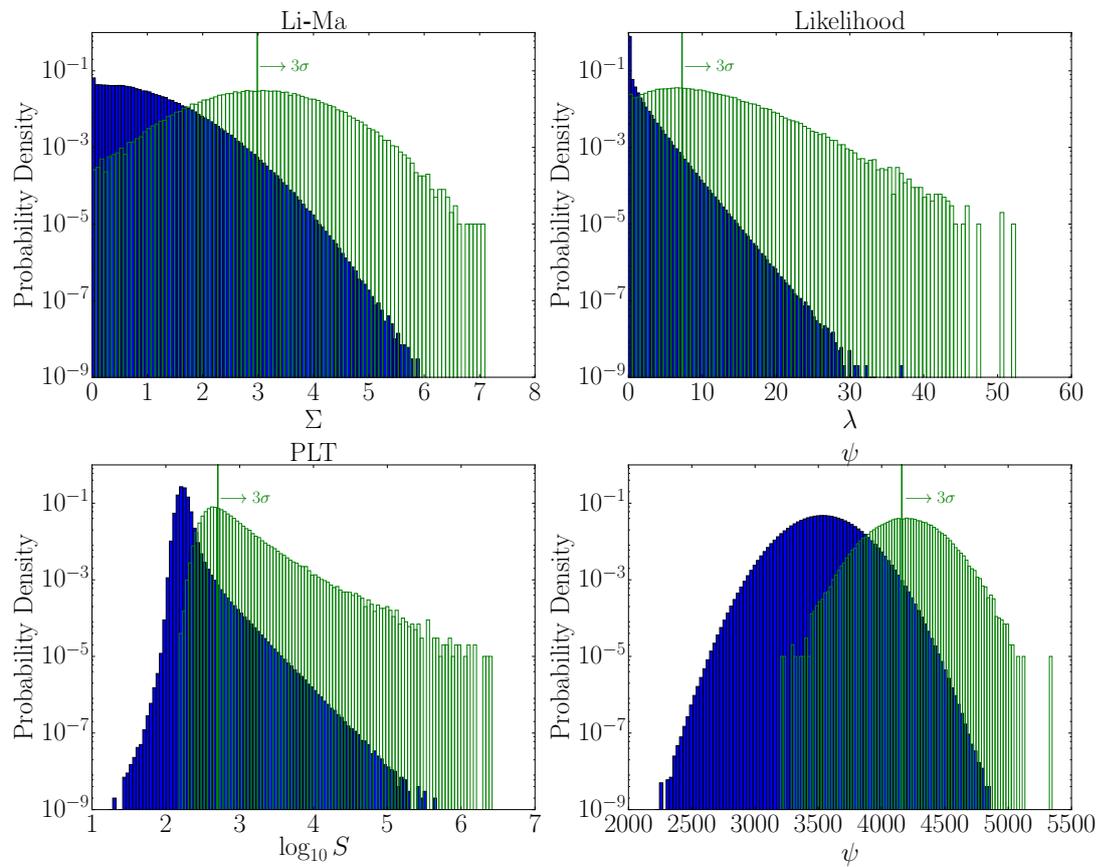


Figure B.1: The signal-plus-background distributions (empty histograms) corresponding with the sensitivities of the four statistical methods. The background-only distributions (filled histograms) are also shown for comparison.

Figure B.2: Same as Fig. B.1 for the 50% DPs at  $3\sigma$ .

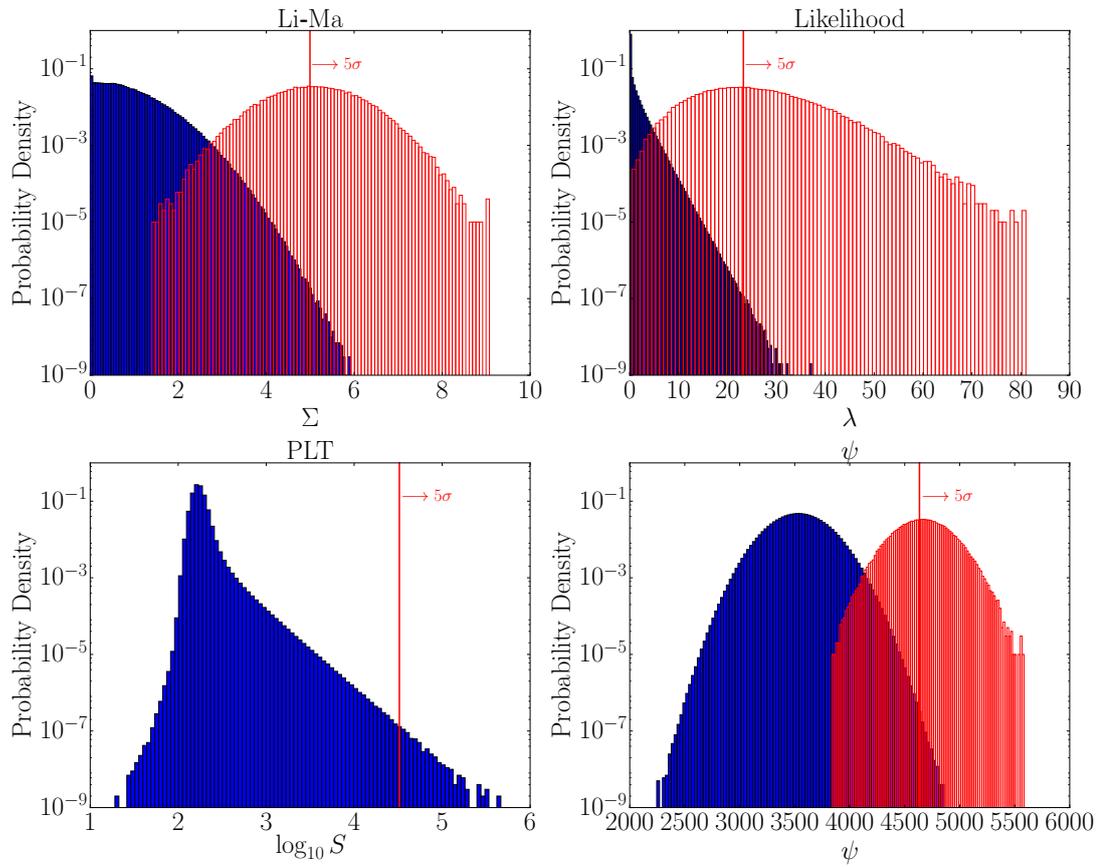


Figure B.3: Same as Fig. B.1 for the 50% DPs at  $5\sigma$ . Note that the DP corresponding with the PLT method was not determined.

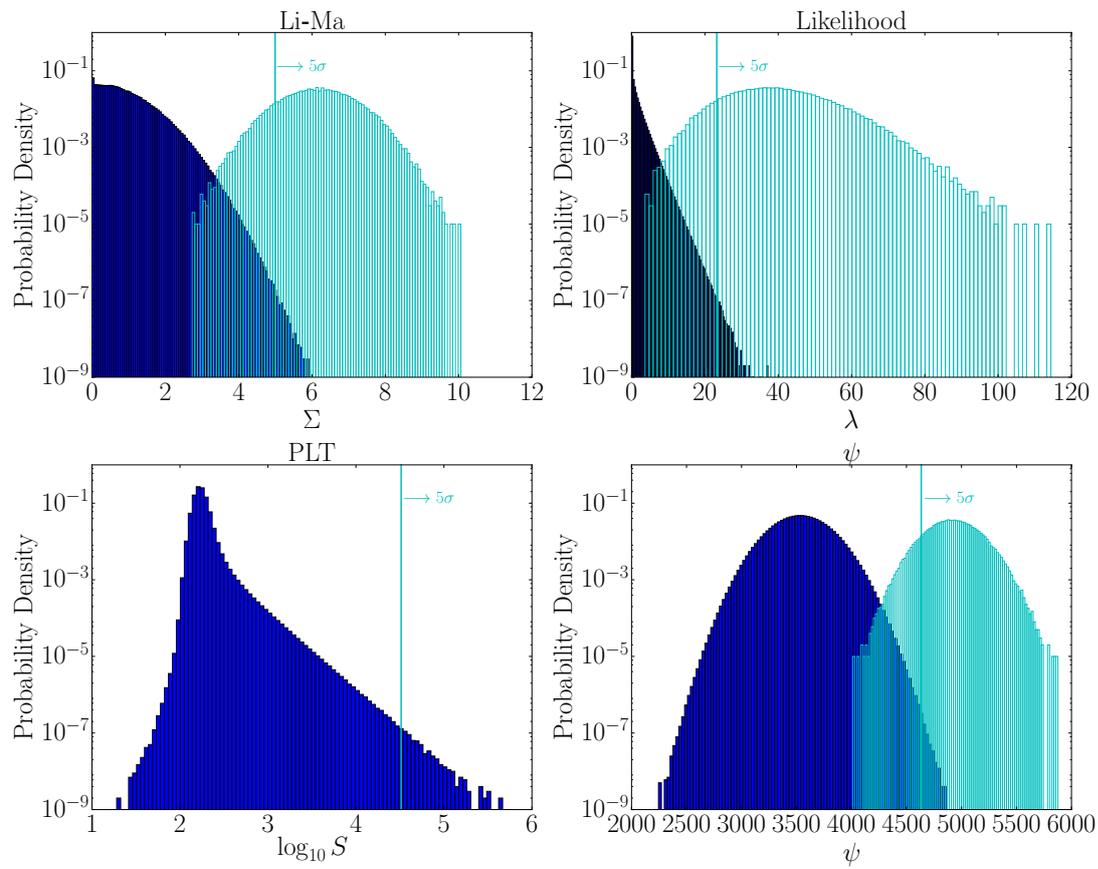


Figure B.4: Same as Fig. B.1 for the 90% DPs at  $5\sigma$ . Note that the DP corresponding with the PLT method was not determined.

## Bibliography

- [1] V. F. Hess, *Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten*. Physikalischen Zeitschrift **13**, 1912, pp. 1084–1091, title in English: “On the Observations of Low Level Radiation during Seven Free Balloon Flights”.
- [2] I. C. Mariş [Pierre Auger Collaboration], *Measurement of the energy spectrum of cosmic rays at the highest energies using data from Pierre Auger Observatory*. European Physical Journal Web of Conferences **53**, 2013, p. 04002, arXiv: 1307.0324.
- [3] J. Blümer, R. Engel, and J. R. Hörandel, *Cosmic rays from the knee to the highest energies*. Progress in Particle and Nuclear Physics **63**(2), 2009, pp. 293–338, arXiv: 0904.0725.
- [4] K. Murase, *On the origin of high-energy cosmic neutrinos*. American Institute of Physics Conference Proceedings **1666**(1), 2015, p. 040006, arXiv: 1410.3680.
- [5] M. G. Aartsen, et al. [ICECUBE collaboration], *Evidence for High-Energy Extraterrestrial Neutrinos at the IceCube Detector*. Science **342**(6161), 2013, arXiv: 1311.5238.
- [6] H. Netzer, *The Physics and Evolution of Active Galactic Nuclei*. Cambridge University Press: Cambridge, 2013.
- [7] M. Kowalski, *Status of High-Energy Neutrino Astronomy*. Journal of Physics: Conference Series **632**(1), 2015, p. 012039, arXiv: 1411.4385.
- [8] N. L. Grigorov, et al., *Investigation of energy spectrum of primary cosmic particles with high and superhigh energies of space stations \*proton\**. Yadernaya Fizika **11**, 1970, pp. 1058–1069.
- [9] K. Asakimori, et al. [JACEE Collaboration], *Energy spectra and composition of cosmic rays above 1 TeV per nucleon*. Proceedings of the 22<sup>nd</sup> International Cosmic Ray Conference **2**, 1991, pp. 57–60.
- [10] K. Asakimori, et al. [JACEE Collaboration], *Energy spectra of protons and helium nuclei above 5 TeV/nucleon*. Proceedings of the 22<sup>nd</sup> International Cosmic Ray Conference **2**, 1991, pp. 97–100.
- [11] K. Asakimori, et al. [JACEE Collaboration], *Cosmic Ray Composition and Spectra: (II) Helium and  $Z > 2$* . Proceedings of the 23<sup>rd</sup> International Cosmic Ray Conference **2**, 1993, pp. 25–29.

- [12] T. V. Danilova, et al., *The Energy Spectrum of the Primary Cosmic Rays in the Range  $10^{13}$  -  $10^{16}$  eV*. Proceedings of the 15<sup>th</sup> International Cosmic Ray Conference **8**, 1977, p. 129.
- [13] Y. A. Fomin, et al., *Energy Spectrum of Cosmic Rays at Energies of  $5 \times 10^{15}$  -  $5 \times 10^{17}$  eV*. Proceedings of the 22<sup>nd</sup> International Cosmic Ray Conference **2**, 1991, p. 85.
- [14] M. Amenomori, et al. [Tibet AS $\gamma$  Collaboration], *The Cosmic-Ray Energy Spectrum between  $10^{14.5}$  and  $10^{16.3}$  eV Covering the “Knee” Region*. Astrophysical Journal **461**, 1996, p. 408, doi: 10.1086/177069.
- [15] M. Nagano, et al., *Energy spectrum of primary cosmic rays between  $10^{14.5}$  and  $10^{18}$  eV*. Journal of Physics G: Nuclear Physics **10**(9), 1984, p. 1295, doi: 10.1088/0305-4616/10/9/016.
- [16] F. Arqueros, et al. [HEGRA Collaboration], *Energy spectrum and chemical composition of cosmic rays between 0.3 PeV and 10 PeV determined from the Cherenkov light and charged particle distributions in air showers*. Astronomy & Astrophysics **359**, 2000, pp. 682–694, arXiv: astro-ph/9908202.
- [17] M. A. K. Glasmacher et al., *The cosmic ray composition between  $10^{14}$  eV and  $10^{16}$  eV*. Astroparticle Physics **12**, 1999, pp. 1–17, doi: 10.1016/S0927-6505(99)00076-6.
- [18] T. Antoni et al. [KASCADE Collaboration], *KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems*. Astroparticle Physics **24**, 2005, pp. 1–25, arXiv: astro-ph/0505413.
- [19] M. Amenomori et al. [TIBET III Collaboration], *The All-particle spectrum of primary cosmic rays in the wide energy range from  $10^{14}$  eV to  $10^{17}$  eV observed with the Tibet-III air-shower array*. Astrophysical Journal **678**, 2008, pp. 1165–1179, arXiv: 0801.1803.
- [20] W. D. Apel et al. [KASCADE Grande Collaboration], *Kneelike structure in the spectrum of the heavy component of cosmic rays observed with KASCADE-Grande*. Physical Review Letters **107**, 2011, p. 171104, arXiv: 1107.5885.
- [21] M. G. Aartsen et al. [ICECUBE Collaboration], *Measurement of the cosmic ray energy spectrum with IceTop-73*. Physical Review **D88**(4), 2013, p. 042004, arXiv: 1307.3795.
- [22] R. U. Abbasi et al. [HiRes Collaboration], *First observation of the Greisen-Zatsepin-Kuzmin suppression*. Physical Review Letters **100**, 2008, p. 101101, arXiv: astro-ph/0703099.
- [23] J. Abraham et al. [Pierre Auger Collaboration], *Observation of the suppression of the flux of cosmic rays above  $4 \times 10^{19}$  eV*. Physical Review Letters **101**, 2008, p. 061101, arXiv: 0806.4302.

- [24] D. Ivanov [Telescope Array Collaboration], *TA Spectrum Summary*. Proceedings of Science **ICRC2015**, 2016, p. 349.
- [25] I. Valiño [Pierre Auger Collaboration], *The flux of ultra-high energy cosmic rays after ten years of operation of the Pierre Auger Observatory*. Proceedings of Science **ICRC2015**, 2016, p. 271.
- [26] P. Bhattacharjee and G. Sigl, *Origin and propagation of extremely high-energy cosmic rays*. Physics Reports **327**, 2000, pp. 109–247, arXiv: astro-ph/9811011.
- [27] J. G. Kirk and P. Duffy, *Particle acceleration and relativistic shocks*. Journal of Physics **G25**, 1999, pp. R163–R194, arXiv: astro-ph/9905069.
- [28] L. O. Drury, *An introduction to the theory of diffusive shock acceleration of energetic particles in tenuous plasmas*. Reports on Progress in Physics **46**, 1983, pp. 973–1027, doi: 10.1088/0034-4885/46/8/002.
- [29] R. Blandford and D. Eichler, *Particle Acceleration at Astrophysical Shocks: A Theory of Cosmic Ray Origin*. Physics Reports **154**, 1987, pp. 1–75, doi: 10.1016/0370-1573(87)90134-7.
- [30] F. C. Jones and D. C. Ellison, *The plasma physics of shock acceleration*. Space Science Reviews **58**, 1991, pp. 259–346, doi: 10.1007/BF01206003.
- [31] D. Caprioli, P. Blasi, and E. Amato, *Non-linear diffusive acceleration of heavy nuclei in supernova remnant shocks*. Astroparticle Physics **34**, 2011, pp. 447–456, arXiv: 1007.1925.
- [32] C. Patrignani, et al. [Particle Data Group], *Review of Particle Physics*. Chinese Physics C **40**(10), 2016, p. 100001, doi: 10.1088/1674-1137/40/10/100001.
- [33] E. Fermi, *On the Origin of the Cosmic Radiation*. Physical Review **75**, 1949, pp. 1169–1174, doi: 10.1103/PhysRev.75.1169.
- [34] K. Greisen, *End to the Cosmic-Ray Spectrum?* Physical Review Letters **16**(17), 1966, pp. 748–750, doi: 10.1103/PhysRevLett.16.748.
- [35] G. T. Zatsepin and V. A. Kuzmin, *Upper Limit on the Spectrum of Cosmic Rays*. Pis’ma v Zhurnal’ Experimental’noi i Teoreticheskoi Fiziki **4**(17), 1966, p. 144.
- [36] J. P. Rachen and P. Meszaros, *Photohadronic neutrinos from transients in astrophysical sources*. Physical Review **D58**, 1998, p. 123005, arXiv: astro-ph/9802280.
- [37] G. Maggi, et al., *Obscured flat spectrum radio active galactic nuclei as sources of high-energy neutrinos*. Physical Review **D94**(10), 2016, p. 103007, arXiv: 1608.00028.
- [38] J. G. Learned and S. Pakvasa, *Detecting tau-neutrino oscillations at PeV energies*. Astroparticle Physics **3**, 1995, pp. 267–274, arXiv: hep-ph/9405296.

- [39] M. G. Aartsen et al. [ICECUBE Collaboration], *The IceCube Neutrino Observatory: Instrumentation and Online Systems*. Journal of Instrumentation **12**(03), 2017, p. P03012, arXiv: 1612.05093.
- [40] J. Ahrens et al. [AMANDA Collaboration], *Physics and operation of the AMANDA-II High Energy Neutrino Telescope*. Proceedings of SPIE – The International Society for Optical Engineering **4858**, 2003, p. 79, arXiv: astro-ph/0211269.
- [41] M. Ackermann et al., *Optical properties of deep glacial ice at the South Pole*. Journal of Geophysical Research: Atmospheres **111**(D13), 2006, p. D13203, doi: 10.1029/2005JD006687.
- [42] T. Gaisser and F. Halzen, *IceCube*. Annual Review of Nuclear and Particle Science **64**, 2014, pp. 101–123, doi: 10.1146/annurev-nucl-102313-025321.
- [43] M. G. Aartsen et al. [ICECUBE Collaboration], *Observation of High-Energy Astrophysical Neutrinos in Three Years of IceCube Data*. Physical Review Letters **113**, 2014, p. 101101, arXiv: 1405.5303.
- [44] D. Xu [ICECUBE Collaboration], *Search for Astrophysical Tau Neutrinos with IceCube*. Proceedings of Science **ICHEP2016**, 2016, p. 452, arXiv: 1702.05238.
- [45] M. G. Aartsen et al. [ICECUBE Collaboration], *Characterization of the Atmospheric Muon Flux in IceCube*. Astroparticle Physics **78**, 2016, pp. 1–27, arXiv: 1506.07981.
- [46] R. Abbasi et al. [ICECUBE Collaboration], *The Design and Performance of IceCube DeepCore*. Astroparticle Physics **35**, 2012, pp. 615–624, arXiv: 1109.6096.
- [47] R. Abbasi et al. [ICECUBE Collaboration], *IceTop: The surface component of IceCube*. Nuclear Instruments and Methods in Physics Research **A700**, 2013, pp. 188–220, arXiv: 1207.6326.
- [48] J. Kunnen, *Direct Walk Reconstruction and Effective Area Determination in IceCube*. Internal Note of the ICECUBE Collaboration, 2011.
- [49] A. Ishihara and S. Yoshida, *Effective area and event rate using the event weighting technique*. Internal Note of the ICECUBE Collaboration, 2006.
- [50] Internal Documentation of the ICECUBE Collaboration.
- [51] M. G. Aartsen et al. [ICECUBE Collaboration], *A combined maximum-likelihood analysis of the high-energy astrophysical neutrino flux measured with IceCube*. Astrophysical Journal **809**(1), 2015, p. 98, arXiv: 1507.03991.
- [52] S. L. Glashow, *Resonant Scattering of Antineutrinos*. Physical Review **118**(1), 1960, pp. 316–317, doi: 10.1103/PhysRev.118.316.
- [53] R. Abbasi et al. [ICECUBE Collaboration], *Measurement of the atmospheric neutrino energy spectrum from 100 GeV to 400 TeV with IceCube*. Physical Review **D83**, 2011, p. 012001, arXiv: 1010.3980.

- [54] M. Ageron et al., *Studies of a full-scale mechanical prototype line for the ANTARES neutrino telescope and tests of a prototype instrument for deep-sea acoustic measurements*. Nuclear Instruments and Methods in Physics Research **A581**, 2007, pp. 695–708, doi: 10.1016/j.nima.2007.08.148.
- [55] A. Avrorin et al., *The Baikal neutrino experiment*. Nuclear Instruments and Methods in Physics Research **A626–627**, 2011, pp. S13–S18, doi: 10.1016/j.nima.2010.06.209.
- [56] A. Margiotta [KM3NeT Collaboration], *Perspectives of the KM3NeT project*. Nuclear and Particle Physics Proceedings **279–281**, 2016, pp. 182–189, doi: 10.1016/j.nuclphysbps.2016.10.026.
- [57] P. Allison et al., *Design and Initial Performance of the Askaryan Radio Array Prototype EeV Neutrino Detector at the South Pole*. Astroparticle Physics **35**, 2012, pp. 457–477, arXiv: 1105.2854.
- [58] L. Gerhardt, et al., *A prototype station for ARIANNA: a detector for cosmic neutrinos*. Nuclear Instruments and Methods in Physics Research **A624**, 2010, pp. 85–91, arXiv: 1005.5193.
- [59] J. T. Link et al., *Radio Detection of UHE Neutrinos with the Antarctic Impulsive Transient Antenna (ANITA) Experiment: Instrumentation*. Proceedings of the 29<sup>th</sup> International Cosmic Ray Conference
- [60] E. Blaufuss et al. [ICECUBE-Gen2 Collaboration], *The IceCube-Gen2 High Energy Array*. Proceedings of Science **ICRC2015**, 2016, p. 1146.
- [61] P. Coppin, *InIce veto studies for the future IceCube-Gen2 neutrino observatory*. Master’s thesis, Vrije Universiteit Brussel, 2017.
- [62] M. G. Aartsen et al. [ICECUBE Collaboration], *All-sky Search for Time-integrated Neutrino Emission from Astrophysical Sources with 7 yr of IceCube Data*. Astrophysical Journal **835**(2), 2017, p. 151, arXiv: 1609.04981.
- [63] M. G. Aartsen et al. [ICECUBE Collaboration], *Searches for Extended and Point-like Neutrino Sources with Four Years of IceCube Data*. Astrophysical Journal **796**(2), 2014, p. 109, arXiv: 1406.6757.
- [64] M. G. Aartsen et al. [ICECUBE Collaboration], *Searches for Time Dependent Neutrino Sources with IceCube Data from 2008 to 2012*. Astrophysical Journal **807**(1), 2015, p. 46, arXiv: 1503.00598.
- [65] M. G. Aartsen et al. [ICECUBE Collaboration], *Search for Prompt Neutrino Emission from Gamma-Ray Bursts with IceCube*. Astrophysical Journal **805**(1), 2015, p. L5, arXiv: 1412.6510.
- [66] M. G. Aartsen et al. [ICECUBE Collaboration], *Search for Transient Astrophysical Neutrino Emission with IceCube-DeepCore*. Astrophysical Journal **816**(2), 2016, p. 75, arXiv: 1509.05029.

- [67] K. Murase, D. Guetta, and M. Ahlers, *Hidden Cosmic-Ray Accelerators as an Origin of TeV-PeV Cosmic Neutrinos*. Physical Review Letters **116**(7), 2016, p. 071101, arXiv: 1509.00805.
- [68] M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics*. Griffin: London, 1977.
- [69] W. T. Eadie, et al., *Statistical Methods in Experimental Physics*. North-Holland: Amsterdam, 1971.
- [70] R. J. Barlow, *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*. Wiley: Hoboken, 2013.
- [71] J. Neyman, *Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability*. Philosophical Transactions of the Royal Society of London **A236**(767), 1937, pp. 333–380, doi: 10.1098/rsta.1937.0005.
- [72] G. J. Feldman and R. D. Cousins, *Unified approach to the classical statistical analysis of small signals*. Physical Review D **57**(7), 1998, pp. 3873–3889, arXiv: physics/9711021.
- [73] E. T. Jaynes, *Probability Theory*. Cambridge University Press: Cambridge, 2013.
- [74] E. D. Feigelson and G. J. Babu, *Statistical Challenges in Modern Astronomy III*. Springer: Heidelberg, 2002.
- [75] P. Gregory, *Bayesian Logical Data Analysis for the Physical Sciences*. Cambridge University Press: Cambridge, 2010.
- [76] R. D. Cousins, *Why isn't every physicist a Bayesian?* American Journal of Physics **63**(5), 1995, pp. 398–410, doi: 10.1119/1.17901.
- [77] M. Richman, *A Search for Muon Neutrinos Coincident with Northern Gamma-Ray Bursts Using IceCube*. Ph.D. thesis, University of Maryland, 2015.
- [78] J. R. Braun, *A Maximum-Likelihood Search for Neutrino Point Sources with the AMANDA-II Detector*. Ph.D. thesis, University of Wisconsin-Madison, 2009.
- [79] L. Brayeur, *Search for High-Energy Neutrinos Associated with Long Gamma Ray Bursts using the IceCube Detector*. Ph.D. thesis, Vrije Universiteit Brussel, 2015.
- [80] T.-P. Li and Y.-Q. Ma, *Analysis methods for results in gamma-ray astronomy*. Astrophysical Journal **272**, 1983, pp. 317–324, doi: 10.1086/161295.
- [81] W. B. Atwood et al., *The Large Area Telescope on the Fermi Gamma-ray Space Telescope Mission*. Astrophysical Journal **697**, 2009, pp. 1071–1102, arXiv: 0902.1089.
- [82] R. S. Pilla, C. Loader, and C. C. Taylor, *New Technique for Finding Needles in Haystacks: Geometric Approach to Distinguishing between a New Source and Random Fluctuations*. Physical Review Letters **95**(23), 2005, p. 230202, arXiv: physics/0505200.

- [83] N. van Eijndhoven, *On the observability of high-energy neutrinos from gamma ray bursts*. *Astroparticle Physics* **28**(6), 2008, pp. 540–546, arXiv: [astro-ph/0702029](#).
- [84] D. Bose, et al., *Bayesian approach for counting experiment statistics applied to a neutrino point source analysis*. *Astroparticle Physics* **50–52**, 2013, pp. 57–64, arXiv: [1212.2008](#).





