Thesis submitted for the degree of
Master of Philosophy (Mixed)

Application of Importance Sampling for Point Source Analysis with the IceCube Neutrino Observatory

Ryan Thomas Burley

February 2020

Supervisors:
Assoc. Prof. Gary Hill
Prof. Gavin Rowell

Faculty of Sciences
School of Physical Sciences
Department of Physics
The University of Adelaide

Principal Superviso: Assoc. Prof. Gary Hill
Co-Supervisor: Prof. Gavin Rowell
Date of Submission: 7th February 2020

Abstract

The IceCube Neutrino Observatory observes astrophysical neutrinos produced by the most energetic processes in the Universe. To date, the exact sources of these neutrinos, particles with no electric charge and almost negligible mass, are still a mystery. In an attempt to identify the sources of the highest energy neutrinos, the IceCube Collaboration uses likelihood analysis to search for clustering of neutrino events in the sky. An important part of this analysis is knowing how often neutrinos randomly cluster on the sky to replicate what an astrophysical neutrino source would look like. However, numerous simulations are required to properly understand this, and hence so are excessive computational resources.

In this thesis, importance sampling is used to force rare clusters of neutrinos to occur on simulated skies. Two methods of importance sampling have been created to force these clusters to occur, a Gaussian weighting method and a binomial weighting method. Once these events are clustered, an appropriate weight can be applied to the sky the cluster is created on, and a likelihood analysis can be performed. We demonstrate how these methods can be used to identify the frequency at which rare clusterings of neutrinos occur, without the requirement of exhaustive computational time. We find that these rare clusters can be forced to occur on a sky with importance sampling, as can appropriate weights indicating the frequency the cluster would appear at a fixed point in space. However, further investigation is required to understand how to correctly apply sampling weights to the results when we perform the likelihood analysis over a full sky.

The result of using importance sampling to identify rare clusters of neutrinos is used to investigate the effectiveness of a new test statistic for hypothesis testing in point source analysis. The most powerful test statistic for this analysis is the maximum likelihood, $\hat{\mathcal{L}}$. This is obtained by maximising a likelihood function relative to the maximum number of signal events, $\hat{n}_s$, from some position on the sky. We construct a new statistic using a combination of the $\hat{\mathcal{L}}$ and $\hat{n}_s$ values, which has been suggested to be a more powerful test statistic than $\hat{\mathcal{L}}$ on its own. Using distributions obtained with importance sampling, we find that there is no evidence to indicate that a test statistic constructed using $\hat{\mathcal{L}}$ and $\hat{n}_s$ is more powerful than $\hat{\mathcal{L}}$ on its own. Furthermore, we find that it simply replicates the results of $\hat{\mathcal{L}}$ by itself, due to the strong correlation between the $\hat{\mathcal{L}}$ and $\hat{n}_s$ combinations in the null and alternate hypotheses tested.

Declaration of Originality

I, Ryan Thomas Burley, certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Date Signature

Acknowledgments

The last year two years have been immensely rewarding, and I am so grateful for everything I've learnt and all the fantastic people I've been able to meet. Firstly, I want to thank my supervisor Gary Hill. The knowledge and skills you've given me these past two years have made this experience all the more worthwhile, and the multiple conversations about Lost have been great. I also want to thank my co-supervisor Gavin Rowell for your support and wisdom throughout this degree, as well as Bruce Dawson for the valuable feedback you have given on my project. I also thank Sabrina Einecke for providing the template to write this thesis. I am grateful to have had the pleasure of meeting everyone in the High Energy Astrophysics Group, and I look forward to what the coming years have in store.

Thank you to Ella, whose editing and comments on the contents of this thesis were extremely helpful, and for all the laughs we've shared over the last year. It's hard to imagine when we met at the beginning of our undergrad that we'd be sitting across from each other every day talking about neutrinos, but here we are and I couldn't be happier.

There's no way I would have made it through the last few months without coming home every day to my ever cheerful house mates. Morgan and Mick, your support has meant more to me than you know. I also want to thank the rest of my friends who have supported me throughout this degree, especially Charlie, Elena, and Michael. You have all been such solid and treasured friends throughout my entire university experience and I am so lucky to have you all in my life.

I would like to thank Jo Fisher, Rick Fisher, Steve Hall, and Jeff Johnson, who ignited my interest in physics and maths. Without your past guidance, I would never have ended up on the path I am now.

Finally, I owe my most important thank you to my parents, Wendy and Bill, for your financial support for the entirety of this degree, and for your constant support and encouragement. This of course extends to the rest of my family who have constantly checked in on me. I genuinely look forward to you all being able to see this piece of work.

Contents

1 Neutrinos

Neutrinos, commonly denoted by the Greek letter ν, are subatomic particles in the Standard Model of particle physics. They are almost massless and have no electric charge, making them an ideal cosmic messenger to study the Universe at extragalactic distances. However, due to their unique properties, they have extremely small interaction rates and require large volumes for detection. Of particular interest are the mysterious high energy astrophysical neutrinos, produced in some of the most energetic processes in the Universe.

1.1 History

In 1896, Henri Becquerel discovered that uranium emitted rays similar to recently discovered x-rays. This phenomenon, later named "radioactivity" by Marie Curie in 1897, became an important focus of research in physics over the following decades. Within this research, alpha decay and gamma decay were discovered to release particles with discrete energies. Beta (β) decay however, released electrons and positrons with a continuous spectrum of energy. This was an issue, as it was well established that there should be a conservation of energy and angular momentum in these radioactive decays. In 1930, Wolfgang Pauli postulated the existence of the neutrino in response to this dilemma. [1] The imbalance between the total energy and angular momentum of the initial nucleus and the particles it emitted prompted Pauli to suggest another neutral particle that carried away these unaccounted for quantities. Pauli called this invisible particle a "neutron". This idea was not widely accepted until Enrico Fermi included the particle in his theory of β-decay in 1934. [2]. By this point, the name "neutron" referred to the neutron we know today when James Chadwick proved its existence in 1932. [3] As such, Fermi gave Pauli's invisible particle the name "neutrino", meaning "little neutral one" in Italian. Fermi illustrated his theory of β-decay by considering the decay of the neutron to a proton, electron, and antineutrino.

$$n \rightarrow p + e^- + \bar{\nu}_e \tag{1.1}$$

Unfortunately, due to the unique properties of the neutrino and the difficulty involved in detecting it, it was not experimentally observed until appropriate technological developments were made years later. In the 1950s, Clyde L. Cowan Jr and Frederick Reines realised that if the theory of β-decay is correct, then atomic weapon tests should produce massive neutrino fluxes. [4] As it would be logistically very difficult to observe the neutrino flux from one of these tests, they opted to use a nuclear reactor to observe the flux instead. The electron anti-neutrinos produced in the reactor could interact with protons in some detector to produce a neutron and a positron.

$$\bar{\nu}_e + p \rightarrow n + e^+ \tag{1.2}$$

The positron would then undergo an annihilation when it collides with an electron to produce gamma-rays, which could then be detected using scintillators and photomultipliers. Cowan and Reines understood that this would not be enough to solidify the neutrino's existence, so they decided the neutron in this reaction would have to be observed too. This could be achieved by placing cadmium in the detector, which when hit by a neutron, enters an excited state and decays through the emission of a photon. This process in known as "neutron capture".

$$n + {}^{108}\mathrm{Cd} \rightarrow {}^{109}\mathrm{Cd}^* \rightarrow {}^{109}\mathrm{Cd} + \gamma \tag{1.3}$$

By measuring the photon from the cadmium decay 5μs after measuring the characteristic photons from the electron-positron annihilation, definitive proof for the existence of the neutrino could be obtained. In 1956, Cowan and Reines successfully performed this experiment by detecting an average of 3 neutrino events per hour over 1371 hours. [5] This produced the first conclusive proof of the neutrino, for which Reines was awarded a Nobel Prize in 1995. Unfortunately, Cowan was ineligible for the award due to his death in 1974.

1.2 The Standard Model

Neutrinos are just one of many particles in the Standard Model of particle physics (Figure 1.1). The Universe consists of fundamental particles which are all moderated by four fundamental forces (gravity, electromagnetic, weak, and strong). The Standard Model has been carefully constructed throughout the 20th century, and describes the relationship between each of these particles and all of the forces, excluding gravity.

All the matter in the Universe is made of elementary particles from one of two groups, quarks or leptons, both of which are fermions (they have half integer spin and follow Fermi-Dirac statistics). Within these groups there are three generations; the most stable particles belong in the first generation, with particle instability increasing as the generation increases. For quarks, the first generation includes the up and down quark, the second has the charm and strange quark, and the third holds the top and bottom quark. Each quark has a color charge which can either be red, green, or blue. For leptons, the first generation consists of the electron (e) and electron neutrino (ν_e), the second has the muon (μ) and the muon neutrino (ν_μ), and the third holds the tau (τ) and the tau neutrino (ν_τ). Each of these particles have a corresponding antiparticle. These are particles with the same mass but opposite physical charge (electric or color). This results in a total of 6 types, or 3 "flavours", of neutrino.

The electromagnetic, strong, and weak forces are mediated by force-carrier particles called bosons (particles with integer spin that follow Bose-Einstein statistics). Interactions between fermions occur through the exchange of discrete amounts of energy transferred

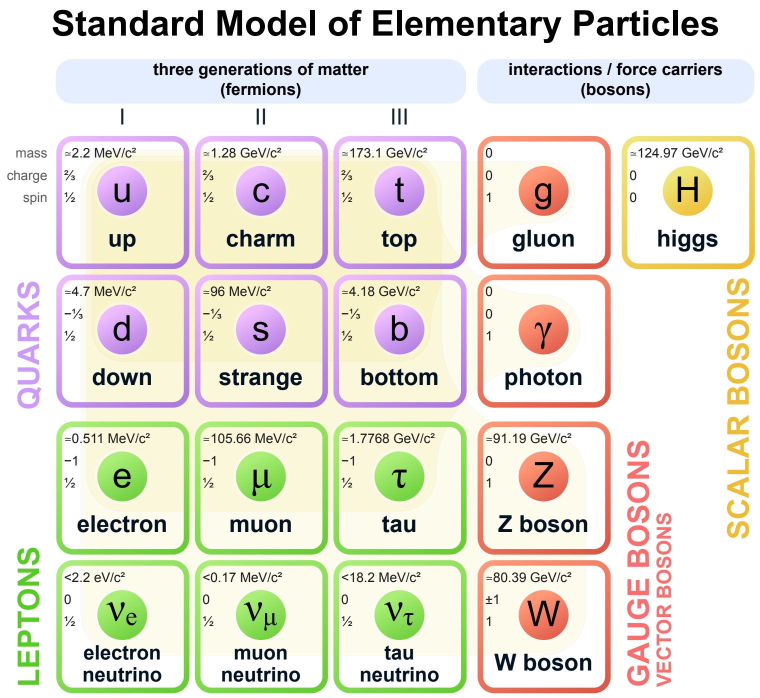

Figure 1.1: Visual depiction of the Standard Model of particle physics.

by these bosons. The electromagnetic force has infinite range and occurs through the exchange of photons (γ). The strong force is only effective at subatomic levels, and is mediated by gluons. The weak force is carried by W and Z bosons, and is also only effective at subatomic distances. Each of these bosons are their own antiparticle. There is also the Higgs boson, which accompanies the Higgs mechanism, a model which describes how the W and Z bosons have mass. This boson's existence was confirmed in 2012 by CERN, in which the particle was discovered to have a mass around 126 GeV$/c^2$ [6]. The force of gravity is notably absent here, as it is still not understood how this force fits into the Standard Model. [7] This arises from the complication of combining aspects of the quantum world with the general theory of relativity. However, as the force of gravity is almost negligible at microscopic scales, the Standard Model performs adequately without its inclusion.

In addition to the previously mentioned lack of gravity in the Standard Model, it also fails to include many other physical phenomena and theoretical predictions. Notably, the Standard Model predicts that neutrinos have zero mass, although neutrino oscillations (more on this in the next section) indicate that they are in fact massive particles. The model can be adjusted for this fact, but it brings further problems to the model, particularly how their mass arises compared to the other particles. Additionally, the current model fails to include dark matter [8] and dark energy, nor the imbalance between matter and antimatter in the Universe [9].

1.3 Neutrino Physics

1.3.1 Neutrino Interactions

As neutrinos have no electric charge they do not interact through the electromagnetic force, and as they have no color charge, they cannot interact through the strong force either. Additionally, as they have almost no mass, the force of gravity is negligible where neutrinos are concerned. This leaves the weak force as the only way for neutrinos to interact, which they achieve through the exchange of $W^{\pm}$ bosons in charged current interactions and Z^0 bosons in neutral current interactions, as shown in Figure 1.2.

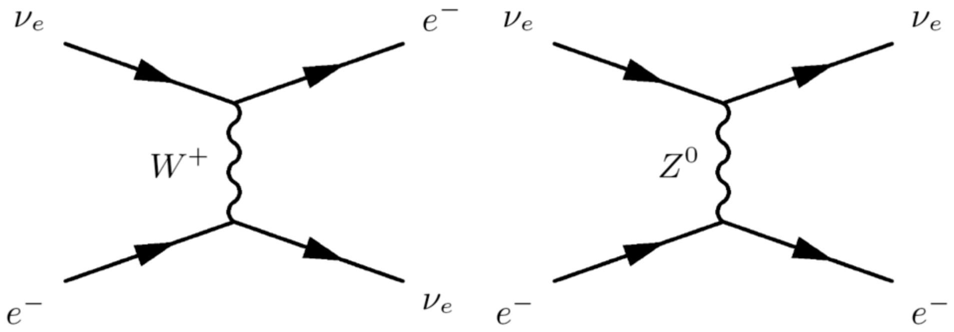

Figure 1.2: Feynman diagrams for an electron neutrino undergoing a charged current interaction (left) and a neutral current interaction (right). [10]

When neutrinos undergo charged current (CC) interactions, the $W^{\pm}$ boson induces the absorption of a neutrino which leaves the neutrino's corresponding lepton from its fermion generation ($e \rightarrow \nu_e$, $\mu \rightarrow \nu_\mu$, $\tau \rightarrow \nu_\tau$). The $W^{\pm}$ boson that mediates the interaction is the opposite charge of the charged lepton involved. For example, a W^+ boson would be exchanged in an interaction with an electron, while a W^- boson would be exchanged during a positron interaction. Overall, the CC interaction involves a transfer of charge, momentum, spin, and energy. Neutrino neutral current (NC) interactions involve the transfer of a Z^0 boson to a target particle. Here, the boson transfers energy, momentum, and spin, while the neutrino continues with the same flavour. [11]

1.3.2 Neutrino Oscillation

As previously mentioned, neutrinos come in one of three flavours: electron, muon, or tau. While neutrinos propagate through a medium, they are able to change flavour in a process referred to as "neutrino oscillation". The phenomenon of neutrino oscillation was first predicted by Bruno Pontecorvo in 1957 [12], and has been confirmed in multiple experiments since, most notably by the Super-Kamiokande Observatory in 1998 [13]. This discovery is especially important as it confirms that neutrinos have a finite mass.

Neutrino flavours are understood to be the result of a superposition of mass eigenstates, ν_1, ν_2, and ν_3, which all travel at different frequencies. As the neutrino propagates, the phase of the combined wave changes and a different neutrino flavour can be observed. The probability at which a neutrino with some flavour will be observed with another flavour depends on the degree of mixing between the mass eigenstates, the mass difference of the eigenstates, the energy of the neutrino, and the length it travelled. [14] An example of how this probability changes is shown in Figure 1.3.

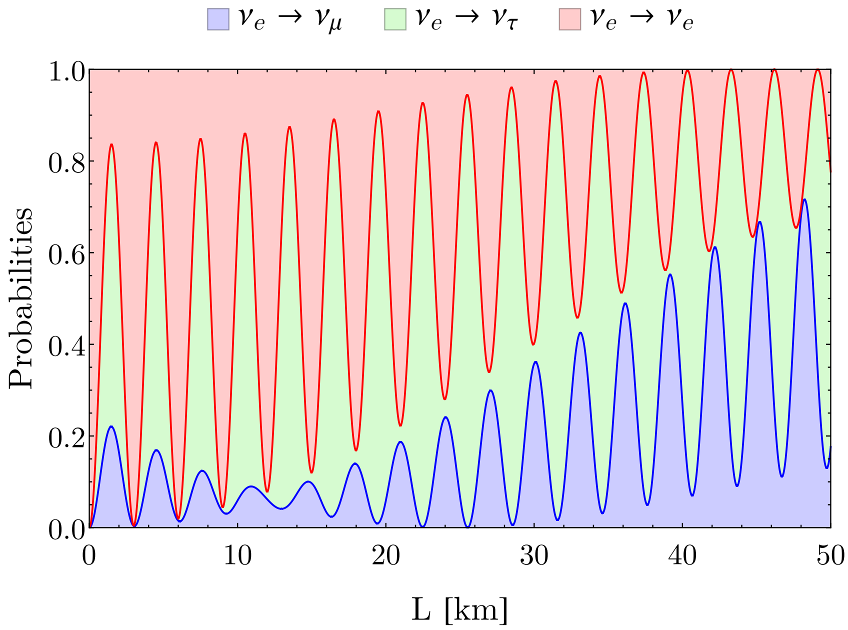

Figure 1.3: The probability of observing a 3 MeV electron neutrino as either an electron, muon, or tau neutrino after a given length. Blue indicates the electron neutrino is observed as a muon neutrino, green indicates it is observed as a tau neutrino, and red indicates it is observed with its original flavour. Note that the sum of probabilities at any given length is always one. [14]

1.4 Neutrino Astronomy

An incredible amount of neutrinos arrive at Earth every second, and the energy that each of these neutrinos have can vary over magnitudes of electron volts (eV). Unsurprisingly, this is due to a multitude of methods responsible for the production of these particles. Some of these neutrinos are nearly as old as the Universe, while others are produced only fractions of a second before hitting the Earth's surface.

1.4.1 Relic Neutrinos

Initially the Universe consisted of neutrinos, electrons, positrons, and photons, all in thermal equilibrium. In its first few seconds of existence, the Universe's temperature dropped below 2.5 MeV, resulting in the decoupling of neutrinos from the rest of the baryonic matter. These relic neutrinos remained at the same temperature as the photons while the rest of the Universe expanded, resulting in a cosmic neutrino background (CνB) with an approximate energy range between 0.1 meV and 1 μeV. When the temperature of the Universe dropped below the temperature of the mass of the electron, electron-positron annihilation occurred which caused an increase in the energy of the photons, resulting in the cosmic microwave background (CMB) that we observe today. [15] As the ratio of temperature of the CνB (T$_\nu$) and the CMB (T$_\gamma$) have remained the same since, we can consider entropy to find that these temperatures are related by:

$$T_\nu = \left(\frac{4}{11}\right)^{\frac{1}{3}} T_\gamma \tag{1.4}$$

The recorded value of T$_\gamma$ is 2.725 $\pm$ 0.001 K, resulting in an approximate value of T$_\nu$ $\sim$ 1.945 K [16]. Due to the intense complexities of detecting neutrinos at this energy, no relic neutrinos have been detected, or may ever be directly observed. [17] Regardless, the detection of these neutrinos would provide a unique view into the beginning stages of the Universe.

1.4.2 Solar Neutrinos

The Sun is essentially an enormous nuclear reactor, housing trillions of particles constantly undergoing nuclear fusion. In turn, this produces the bulk of the neutrinos received at Earth. The beginning of the Sun's nuclear fusion cycle starts with the burning of hydrogen to helium, which is possible due to the intense pressure and heat inside the star. First, proton-proton reactions produce deuterium (d), a positron, and an electron neutrino. This process produces approximately 86% of the solar neutrinos we receive at Earth.

$$p + p \rightarrow d + e^+ + \nu_e \tag{1.5}$$

This deuterium reacts with another proton to produce helium-3 and a photon.

$$d + p \rightarrow {}^3\mathrm{He} + \gamma \tag{1.6}$$

Next, helium-3 nuclei can react with each other to yield helium-4 and protons.

$$^{3}\mathrm{He} + {}^{3}\mathrm{He} \rightarrow {}^{4}\mathrm{He} + 2p \tag{1.7}$$

Then, this helium-4 can react with helium-3 to produce beryllium and a photon.

$$^{3}\mathrm{He} + {}^{4}\mathrm{He} \rightarrow {}^{7}\mathrm{Be} + \gamma \tag{1.8}$$

This beryllium can undergo electron capture, absorbing an electron which transforms one of its protons into a neutron. This reaction produces a lithium atom and releases an electron neutrino in the process. Neutrinos released through this process make up approximately 14% of the solar neutrino flux at Earth.

$$^{7}\mathrm{Be} + e^{-} \rightarrow {}^{7}\mathrm{Li} + \nu_{e} \tag{1.9}$$

The lithium atom can interact with a proton to produce two helium-4 nuclei, which can continue on according to Equation 1.8.

$$^{7}\mathrm{Li} + p \rightarrow {}^{4}\mathrm{He} + {}^{4}\mathrm{He} \tag{1.10}$$

Otherwise, the beryllium nuclei can interact with a proton to create a boron atom, while releasing a photon.

$$^{7}\mathrm{Be} + p \rightarrow {}^{8}\mathrm{B} + \gamma \tag{1.11}$$

Here, the boron will undergo β^{+} decay, resulting in a beryllium-8 nuclei, positron, and electron neutrino. The solar neutrino flux at Earth comprises of about 0.02% of these neutrinos. [18]

$$^{8}\mathrm{B} \rightarrow {}^{8}\mathrm{Be} + e^{+} + \nu_{e} \tag{1.12}$$

This beryllium-8 nuclei then splits into two helium-4 nuclei. The resultant nuclei from many of these reactions can participate in many of the other reactions listed, continuing the cycle. The majority of the neutrinos produced leave the Sun and travel through the solar system, resulting in a solar neutrino flux at Earth of around 7×10^{10} $\mathrm{cm^{-2}s^{-1}}$ with energies between 100 keV and 18 MeV [19]. Notably, each of the neutrinos produced in the above reactions are electron neutrinos. This was important in providing evidence for neutrino oscillations, which involved measuring solar neutrino flavour compositions at Earth and comparing them to the expected amount of neutrinos produced in the Sun.

1.4.3 Supernova Neutrinos

A supernova is an explosion which occurs at the end of the life of particular massive stars. During the lifetime of a star, it continually burns through its nucleic fuel, gradually increasing the atomic number of the nuclei, until only iron remains. At this point, if the stellar core has a mass larger than $1.44\mathrm{M}_{\odot}$, it will collapse under its own gravitational force. As the star collapses, deleptonization occurs in which the electrons in the star are forced into protons, producing a burst of electron neutrinos and neutrons.

$$e^- + p \rightarrow n + \nu_e \tag{1.13}$$

Additional neutrinos are produced due to thermal photons producing electron-positron pairs, which are immediately absorbed due to the intense density of matter during the collapse. As this occurs, virtual Z bosons are exchanged which can decay into neutrino and antineutrinos of any flavour.

$$e^- + e^+ \rightarrow Z \rightarrow \nu_\alpha + \bar{\nu}_\alpha \tag{1.14}$$

An immense amount of neutrinos produced in these reactions are released during the supernova. A notable example involves supernova SN 1987a, which was observed on February 24^{th}, 1987. On this day, the Kamiokande II detector, the Irvine-Michigan-Brookhaven (IMB) detector, and the Baksan Neutrino Observatory (BNO) detected 25 antineutrino events between them, all within 13 seconds [20, 21, 22]. Following this detection, SN 1987a was observed in visible light. The amount of neutrinos from this event matched the theoretical prediction that this supernova would have produced 10^{58} neutrinos with a total energy of 10^{46} J [23].

Similar to the cosmic neutrino background, the diffuse supernova neutrino background (DSNB) is a theorised flux of neutrinos and antineutrinos caused by all past supernovae events across the Universe. This flux has not been directly observed, but does have upper limits based on experimental results. Many different factors and models have been used to get an estimate for this flux, which currently has an upper limit of 1.9 $\text{cm}^{-2}\text{s}^{-1}$ [24].

1.4.4 Atmospheric and Astrophysical Neutrinos

Atmospheric and astrophysical neutrinos are produced as a result of some of the highest energy processes in the Universe. Astrophysical neutrinos are produced at the sources of these high energy processes and propagate astronomical distances to reach Earth. Cosmic rays (charged particles) are also produced at these sources. When cosmic rays reach the Earth, they interact in the atmosphere and produce a shower of particles which travel down to the Earth's surface. Neutrinos produced in these showers are called atmospheric neutrinos. Both of these concepts are covered in detail in Chapter 2.

1.4.5 Cosmogenic Neutrinos

Cosmogenic neutrinos are produced due to the interactions of cosmic rays with energies greater than 100 EeV. These ultra-high energy cosmic rays (UHECR) are limited in their travel by the Greisen-Zartsepin-Kuzmin (GZK) cut-off, where UHECR with energy greater than 60 EeV are prone to interactions with photons from the cosmic microwave background (CMB). [25] In this reaction, cosmic ray protons interact with CMB photons to produce pions, π, through a delta resonance, Δ^+, along with a proton or neutron which continues travelling with a lower energy. A similar process occurs for UHECR nuclei above this energy.

$$p + \gamma_{CMB} \rightarrow \Delta^{+} \rightarrow p + \pi^{0} \tag{1.15}$$

$$p + \gamma_{CMB} \rightarrow \Delta^{+} \rightarrow n + \pi^{+} \tag{1.16}$$

The π^{+} pions then go on to decay into cosmogenic neutrinos, with energies in the range of 100 PeV to 10 EeV.

$$\pi^{+} \rightarrow \mu^{+} + \nu_{\mu} \rightarrow e^{+} + \nu_{e} + \bar{\nu}_{\mu} + \nu_{\mu} \tag{1.17}$$

To date, no neutrinos within this range have been detected. Observing these particles will be useful for UHECR studies, as the GZK limit constrains the travel of these high energy cosmic rays to approximately 200 Mpc before they are below the GZK energy. Therefore, cosmogenic neutrinos are important to better understand the highest energy cosmic ray sources within 200 Mpc.

1.5 Summary

Neutrinos have a rich but relatively recent history, with many mysteries and unanswered questions. Their unique neutral charge and negligible mass combination make them the ideal cosmic messengers, which can assist in understanding a multitude of processes throughout the Universe. However, their low interaction probability make them hard to detect, with large volumes required to see them at the highest energies. The current advancements in observatories and particle detectors make this an exciting time for neutrino physics, and specifically, neutrino astronomy.

2 High Energy Astrophysics

2.1 Cosmic Rays

Charged cosmic rays are energetic protons and nuclei produced in a number of processes throughout the Universe. These particles propagate the interstellar medium at almost the speed of light. Many of the cosmic rays we receive at Earth are due to solar flares and coronal mass ejections from the Sun. The origin of the highest energy cosmic rays, however, is shrouded in mystery, although there are a number of proposed acceleration methods and sources for these particles.

2.1.1 Production of High Energy Cosmic Rays

There are two primary models to explain the production of high energy cosmic rays: Fermi's original theory and diffusive shock acceleration (DSA). Fermi's original theory of cosmic ray acceleration describes the energy change of cosmic rays within magnetic clouds. The magnetic field of these molecular gas clouds arises from the individual magnetic contributions of the ions it is comprised of. In Fermi's original theory, also referred to as the Fermi mechanism of the 2$^{\text{nd}}$ order, cosmic rays undergo collisionless scattering within these molecular clouds. That is, they don't collide with the atoms or ions in the cloud. In doing so, the cosmic rays exit the cloud with an energy which on average is greater than the energy it entered with. The average fractional energy gain of the cosmic rays depends on the angle at which the particle enters the cloud relative to the direction of propagation of the cloud, as well as the angle it exits the cloud. As a result, the average fractional energy gain, $\Delta E/E$, of the particle is proportional to the square of the ratio of the speed of the cloud to the speed of light, β, such that:

$$\frac{\Delta E}{E} = \frac{4}{3}\beta^2_{\text{cloud}} \tag{2.1}$$

The resulting energy gain of particles through this acceleration mechanism is relatively small due to there almost being as much energy lost in these collisions as there is energy gained. To explain high energy cosmic rays, Fermi's original theory was adapted to describe cosmic ray acceleration in supernova shocks. This process is known as both diffusive shock acceleration and the Fermi mechanism of the 1$^{\text{st}}$ order. Diffusive shock acceleration describes how cosmic rays can be accelerated to high energies due to passing back and forth across the shock wave of a supernova. As the particles cross the shock wave, they can interact with magnetised plasma much like the interactions within the clouds of Fermi's original theory. Following this, they can pass back through the shock front and interact in the same way. As the particle makes this passage multiple times, it can increase its energy exponentially. The average fractional energy gain, $\Delta E/E$, of this mechanism is:

$$\frac{\Delta E}{E} = \frac{V_S}{c} \tag{2.2}$$

where V_S is the speed of the shock front. As the particle could escape this cycle at any given crossing, cosmic rays produced in this process span multiple orders of magnitude in energy. Cosmic rays can also be produced in the intense magnetic fields of pulsars [26], and in binary star systems as matter is transferred from one star to the accretion disk of the other [27]. Additionally, it is theorised that the highest energy cosmic rays are produced in the relativistic jets through a diffusive shock acceleration process powered by super-massive black holes at the centre of active galactic nuclei (AGN) [28].

2.1.2 Energy Spectrum

The cosmic ray energy spectrum spans multiple orders of magnitude in electron volts and has an interesting but not well understood structure. Multiple experiments currently observe the highest energy cosmic rays, and have already documented much of their extensive energy range. A graph showing recent high energy cosmic ray flux results from assorted experiments is show in Figure 2.1.

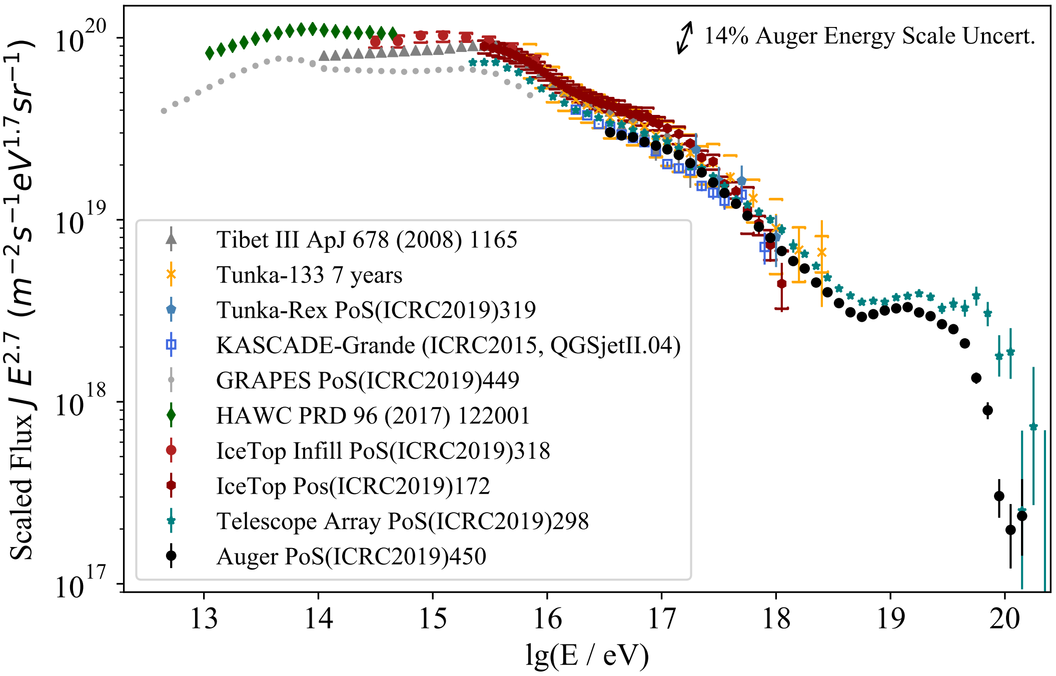

Figure 2.1: Energy spectra of high-energy cosmic rays from multiple experiments around the world. All measurements were obtained from cosmic ray air shower observations. [29]

It can be seen from Figure 2.1 that there are notable changes in the slope at energies of approximately 4 × 10^{15} eV, 1 × 10^{17} eV, and 1 × 10^{19} eV. These regions are referred to as the “first knee”, the “second knee”, and the “ankle”, respectively. The presence of the first knee has been known for many years and is thought to occur due to the maximum energy that supernova remnants, within our galaxy, can accelerate cosmic rays to. [30] The energy of the cosmic rays either side of this region can then be explained by considering magnetised winds, or similar extra acceleration processes, influencing cosmic rays when they exit supernova remnants. The second knee was identified much more recently and there are no clear theories on the cause of this feature. [31] The ankle has also been known about for years, and is proposed to be due to a purely extragalactic component, meaning that any cosmic ray above this energy is produced outside of our galaxy. [32] The spectrum then cuts off due to the GZK limit involving cosmic ray interactions with the cosmic microwave background, as described in Section 1.4.5.

2.2 Neutrinos

2.2.1 Astrophysical Neutrinos

The origin of the highest energy cosmic rays can be quite difficult to trace due to extragalactic magnetic fields altering their paths. However, neutrinos produced at cosmic ray production sites can assist in understanding the mysterious processes these particles are created in. When these cosmic rays are released, they can interact with other cosmic rays, atomic nuclei, or photons, and result in the production of neutrinos. There are two main reactions for neutrino production, hadronic collisions and pion photo-production. In hadronic collisions, protons collide with other protons to create charged or neutral pions. Each of these pions are produced with equal probability.

$$p + p \rightarrow \pi^{\pm} \tag{2.3}$$

$$p + p \rightarrow \pi^{0} \tag{2.4}$$

These charged pions then decay to a charged muon and neutrino, with the muon decaying to its respective electron or positron and two neutrinos.

$$\pi^{+} \rightarrow \mu^{+} + \nu_{\mu} \rightarrow e^{+} + \nu_{e} + \bar{\nu}_{\mu} + \nu_{\mu} \tag{2.5}$$

$$\pi^{-} \rightarrow \mu^{-} + \bar{\nu}_{\mu} \rightarrow e^{-} + \bar{\nu}_{e} + \nu_{\mu} + \bar{\nu}_{\mu} \tag{2.6}$$

The neutral pion decays into two photons, each with approximately the same energy.

$$\pi^{0} \rightarrow \gamma + \gamma \tag{2.7}$$

In pion photo-production, a proton collides with a photon to produce either a proton and neutral pion, or a neutron and positively charged pion.

$$p + \gamma \rightarrow p + \pi^{0} \tag{2.8}$$

$$p + \gamma \rightarrow n + \pi^{+} \tag{2.9}$$

The pions then decay along the same chain as described for hadronic collisions. [33] Together, the hadronic and pion photo-production collisions produce multiple neutrinos that can be detected at the IceCube Neutrino Observatory, which is discussed in great detail in Chapter 3.

2.2.2 Atmospheric Neutrinos

When a cosmic ray penetrates the Earth's atmosphere, it can interact with an air molecule to produce a shower of particles. This interaction typically occurs around 15 kilometres above the Earth's surface, and results in the production of numerous particles which can be broken into three components: muonic, electromagnetic, and hadronic, as illustrated in Figure 2.2. This shower of particles is referred to as an extensive air shower (EAS), and follows a conical shape as particles are gradually produced and travel towards the Earth's surface. [34]

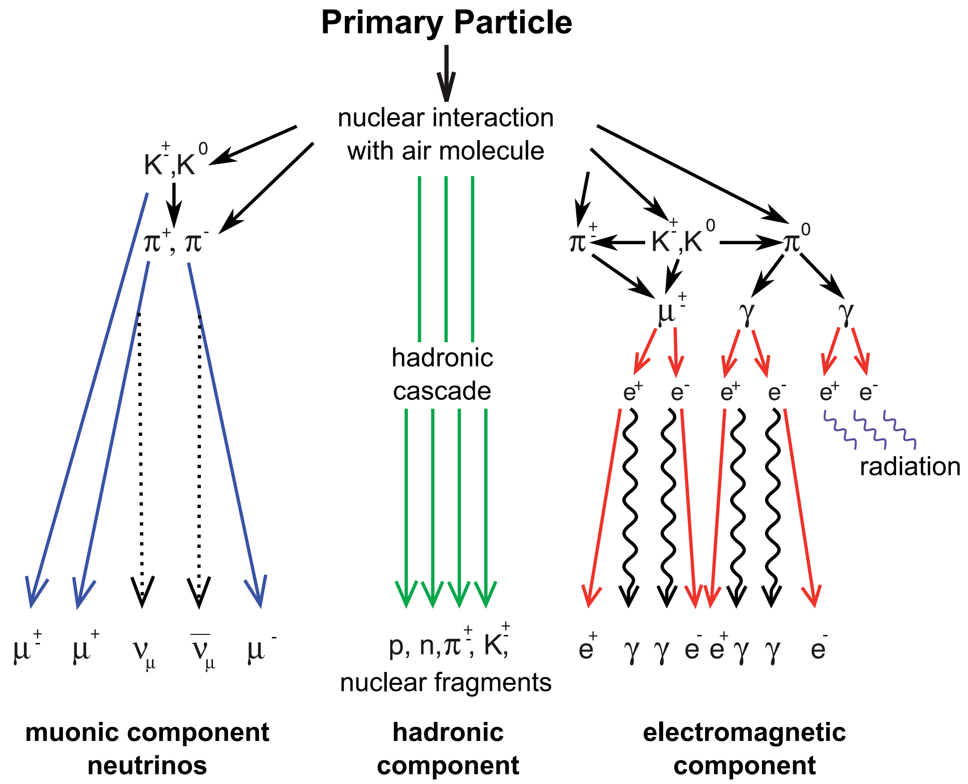

Figure 2.2: Diagram showing the three components (muonic, electromagnetic, hadronic) of secondary particles produced due to the interaction of a cosmic ray in the Earth's atmosphere. [35]

The muonic component consists of neutrinos and muons that are produced by the decay of charged kaons and pions created by the initial collision of the cosmic ray. Charged kaons can either decay to their corresponding charged muon and a muon neutrino, or to a corresponding charged pion and a neutral pion.

$$K^+ \rightarrow \mu^+ + \nu_\mu \tag{2.10}$$

$$K^- \rightarrow \mu^- + \bar{\nu}_\mu \tag{2.11}$$

$$K^\pm \rightarrow \pi^\pm + \pi^0 \tag{2.12}$$

These pions and muons can then decay in the same processes as described in Equations 2.5 and 2.6. The neutrinos produced here are the atmospheric neutrinos, and are important to understand for experiments attempting to detect astrophysical neutrinos. This concept will be expanded on further in Chapter 3.

The electromagnetic component of these air showers is due to the interactions of electrons, positrons, and photons. The photons, produced by the decay of neutral pions, can interact with atomic matter through pair production, where the photon's energy is converted into an electron-positron pair. However, this only occurs if the energy of the photon is greater than the rest mass energy of an electron/positron, 1.022 MeV.

$$\gamma \rightarrow e^- + e^+ \tag{2.13}$$

These electrons and positrons, as well as the electrons and positrons produced in the initial muon decays, can collide with their corresponding antiparticle to cause an electron-positron annihilation where two photons with equal energy are produced.

$$e^- + e^+ \rightarrow \gamma + \gamma \tag{2.14}$$

The resulting particles, in both pair production and electron-positron annihilation, go on to continuously repeat these processes until the relevant particles are below the production threshold energy.

The hadronic component is due to the primary cosmic ray producing other hadrons, such as protons and neutrons, with lower energy. These hadrons can produce their own particle showers and electromagnetic cascades, continuing on the cycle of the initial cosmic ray. Any of the particles produced due to the initial cosmic ray's interaction are referred to as secondary particles.

2.3 Gamma Rays

Gamma rays (γ-rays) are another particle which can help us understand the highest energy processes in the Universe. As explained in Section 2.2.1, high energy photons are produced in hadronic and pion photo-production collisions of cosmic rays at their sources. The photons produced in these reactions are called hadronic γ-rays. There is another category for these particles, called leptonic γ-rays, which are produced when relativistic electrons undergo leptonic processes. Namely, these processes are inverse Compton scattering, synchrotron radiation, and bremsstrahlung radiation.

Inverse Compton Scattering

Inverse Compton scattering (ICS) is the process in which relativistic electrons scatter off low energy photons, boosting the photon's energy and reducing the electron's energy. The rate of electron energy loss can be found by considering a Lorentz transformation for the scattering of the photon off the electron in both the laboratory frame and the electron's rest frame. The resulting relation for the electron energy loss rate is:

$$\frac{dE_e}{dt} \sim -\frac{4}{3} U_{rad} \sigma_T c \gamma^2 \tag{2.15}$$

Here, U_{rad} is the radiation energy density of the photons, σ_T is the Thomson cross section (the cross section for the elastic scattering of radiation from a free electron), and γ is the Lorentz factor.

Synchrotron Radiation

Synchrotron radiation is produced by the spiralling of relativistic electrons as they propagate through magnetic fields while under the effects of relativistic beaming. To obtain a relation for the rate at which electrons lose energy due to this process, both the relativistic beaming and a Lorentz transformation between the laboratory frame and the rest frame of the electron must be taken into account. This results in an electron energy loss rate of:

$$\frac{dE_e}{dt} \sim -\frac{4}{3} \sigma_T c U_B \beta^2 \gamma^2 \tag{2.16}$$

Where U_B denotes the energy density of the magnetic field and β is the velocity of the electron with respect to the speed of light. The synchrotron radiation emitted can range in energy over the entire electromagnetic spectrum, and its detection helps us understand the magnetic fields in different interstellar environments.

Bremsstrahlung Radiation

Bremsstrahlung radiation is the radiation produced when a charged particle is accelerated by passing through the electric and magnetic fields surrounding a nucleus. Cosmic ray electrons can produce relativistic bremsstrahlung radiation in the form of γ-rays as they undergo collisions with the constituent particles of interstellar gas. The relationship for energy loss rate of an electron per unit distance travelled due to this process is:

$$\frac{dE_e}{dl} = -E_e \frac{n_z Z(Z+1.3)e^6}{16\pi^3 \hbar \epsilon_0^3 c^5 m_e^2} \left[ln \left(\frac{183}{Z^{\frac{1}{3}}} \right) + \frac{1}{8} \right] \tag{2.17}$$

Here, Z denotes the atomic number of the scattering atom considered, while n_z is that same atom's number density and m_e is the mass of an electron. Measuring bremsstrahlung, while considering the effect of different molecules on its production, helps us to understand the composition of the elements within our galaxy.

2.4 Detection Methods

The detection of the aforementioned particles can be complex and quite often requires large areas or volumes to view them at their highest energies. Currently, multiple experiments around the globe utilise different methods to observe cosmic rays, neutrinos, and γ-rays. Three types of detection methods frequently used to detect these particles are Cherenkov detectors, imaging atmospheric Cherenkov telescopes, and scintillation detectors.

2.4.1 Cherenkov Detectors

Cherenkov detectors rely on a specific type of electromagnetic radiation produced by charged particles as they propagate through the detector, and are the primary type of instrument used for neutrino detection. This radiation is Cherenkov radiation, and is caused when a charged particle moves faster than the phase velocity of light in the medium it is traversing. As the charged particle passes through the medium, it excites the atoms and molecules surrounding it. These atoms and molecules then instantly return to their ground state through the emission of a photon. These photons are the Cherenkov radiation, which are in the visible light range of the electromagnetic spectrum and peak at a characteristic blue colour. The angle at which Cherenkov radiation is emitted due to a moving charge can be easily found by considering the geometry of this situation. This angle is given by:

$$\cos\theta = \frac{1}{n\beta} \tag{2.18}$$

Here, θ is the emission angle, n is the refractive index of the medium, and β is the ratio of the speed of the particle to the speed of light. This is also shown graphically in Figure 2.3. These detectors are typically volumetric and consist of a material which allows minimal absorption of the visible light that is Cherenkov radiation. The detector also consists of, or is surrounded by, photomultipliers that receive the radiation. The information from the photomultipliers can be pieced together to find the direction and energy of the original particle. This type of detector is typically used for neutrino detection. Notable examples of neutrino Cherenkov detectors include the IceCube Neutrino Observatory (discussed in detail in Chapter 3), Kamiokande II [37], and the Sudbury Neutrino Observatory [38].

2.4.2 Imaging Atmospheric Cherenkov Telescopes

Imaging atmospheric Cherenkov telescopes (IACTs) also use Cherenkov radiation to detect high energy particles, and are primarily used for high energy γ-ray detection. When high energy γ-rays interact in the Earth's atmosphere, they undergo pair production resulting in an electron and positron, which begins an electromagnetic cascade as described in Section 2.2.2. The electrons and positrons in the cascade travel at almost the speed of light and produce Cherenkov radiation. These resulting photons can be seen by the IACTs on the ground, which typically consist of mirrors with large surface areas that reflect the light onto a photomultiplier. As the base of the air shower for these high energy γ-rays can stretch over quite a large area, extending up to one kilometre in

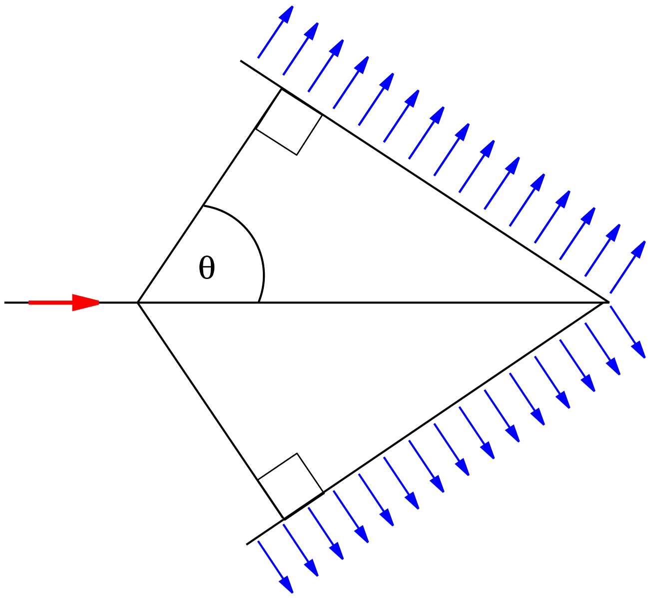

Figure 2.3: Graphical representation of the angle of Cherenkov emission, θ, as a particle (red arrow) moves faster than the phase velocity of light in the medium it is traversing. [36]

radius, multiple separate IACTs can be used to view the same air shower. Each individual IACT views some of the air shower, and together the data can be used to reconstruct the original energy and direction of the γ-ray. Many IACTs are currently used for high energy γ-ray astronomy, including H.E.S.S. [39], MAGIC [40], and VERITAS [41], while the next generation γ-ray detector, the Cherenkov Telescope Array (CTA), is currently under construction [42].

2.4.3 Scintillation and Fluorescence Detectors

For high energy cosmic ray observatories, a combination of scintillation detectors and fluorescence detectors are commonly used. Both of these detectors work concurrently to measure an extensive air shower and determine the nature of the original cosmic ray. These air showers also spread over a large area on the Earth's surface, so multiple scintillation detectors can be spread out to detect the secondary particles produced. A scintillator contains a luminescent material, which when struck by an incoming particle, can absorb its energy, and then return to ground state by emitting this energy as a photon. The scintillation detector also contains a photomultiplier, which detects the resulting light. Multiple scintillation detectors like this can be spread over some area to gather as much information about the air shower as possible. [43] This set-up is often accompanied by fluorescence detectors. As the electrons and positrons produced in these air showers propagate, they undergo inelastic collisions with nitrogen molecules in the air, causing their excitation. As a result, these molecules emit photons in the range between 290 and 430 nm to return to ground state, producing ultraviolet fluorescent light. [44] As such, telescopes that watch for this characteristic light can track the evolution of the air shower as it moves through the atmosphere. Additionally, having more than one of these fluorescence telescopes observing the same air shower can greatly improve the reconstruction of the cosmic ray's arrival direction. Current experiments that implement these methods to detect ultra high energy cosmic rays include the Pierre Auger Observatory [45] and the Telescope Array Project [46].

2.5 Summary

The Earth is continuously bombarded by high energy cosmic rays, neutrinos, and γ-rays created in assorted processes throughout the Universe. Although clear sources for the majority of these particles are not currently known, understanding the processes that create them and the relationships between the different particles can help us narrow down their origins. Additionally, the use of multiple telescopes and experiments around the world to detect these particles can assist in identifying their sources. Observing these particles is especially vital to find the mysterious ultra high energy cosmic ray acceleration sites in the Universe.

3 The IceCube Neutrino Observatory

The IceCube Neutrino Observatory is a cubic kilometre Cherenkov telescope located within the ice at the Admunsen-Scott South Pole Station, Antarctica. The primary goal of the IceCube Neutrino Observatory, colloquially referred to as IceCube, is to observe astrophysical neutrinos produced in the most energetic processes in the Universe. Currently, 52 institutions over 12 countries make up the IceCube Collaboration [47], all of which work on the scientific goals, design, and construction of the detector.

3.1 Design

The forefather for the IceCube Neutrino Observatory was the Antarctic Muon and Neutrino Detector Array (AMANDA), which was located at the same location. This detector collected data between 1997 and 2007, and provided the important result that the glacial ice of Antarctica could facilitate a Cherenkov detector suitable for neutrino detection. [48] IceCube instruments a cubic kilometre of this ice for Cherenkov detection. Construction on the detector began in 2004, and ran over seven austral summers until completed in 2010. The major component of this construction involved deployment of the "strings" which hold the digital optical modules (DOMs). Each DOM contains a ten-inch photomultiplier tube for Cherenkov radiation collection, which views this light with 2 nanosecond time resolution. They also contain other related electronics and calibration devices, while the string itself sends the information from each DOM to the IceCube laboratory located on top of the detector. To deploy a string, which holds 60 DOMs each separated by 17 metres, a hot water drill was used to melt a hole in the ice down to a depth of 2450 metres. The string is then quickly deployed down the borehole such that the DOMs are positioned between a depth of 1450 and 2450 metres, and the water is left to freeze over the string. This process was performed for all 86 strings over a hexagonal grid such that each string was separated by 125 metres. The finished product results in the cubic kilometre detector that is IceCube, as shown in Figure 3.1. The reason for the depth of the IceCube detector is to significantly reduce the background signal. Specifically, at this depth, only the interactions of neutrinos and muons are seen, as other charged particles cannot penetrate this distance from any direction around IceCube. The majority of the muons observed are due to cosmic ray interactions in the atmosphere. [49]

IceTop

IceTop is a cosmic ray detector array which spreads across a square kilometre over the top of the IceCube Neutrino Observatory. This surface array consists of 80 stations separated by a mean distance of 125 metres. Each of these stations contains two tanks filled with clear ice, separated by 10 metres, and both contain two regular DOMs. These DOMs

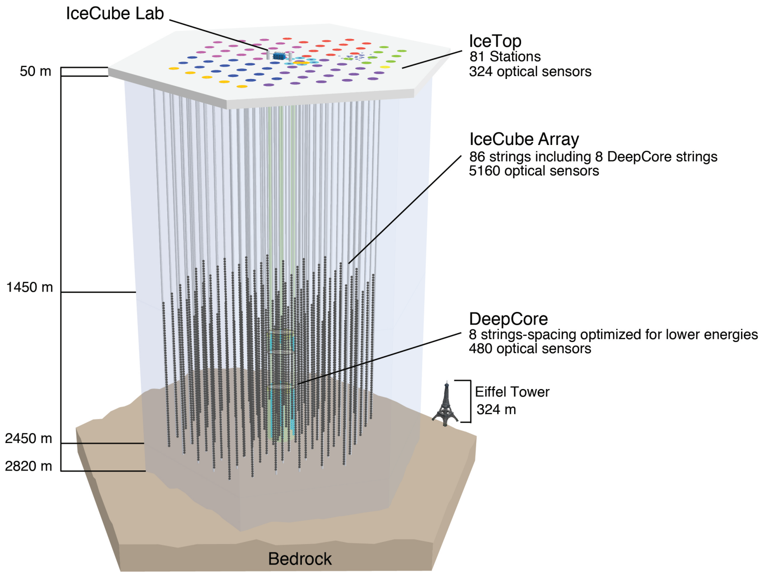

Figure 3.1: A scale diagram of the IceCube Neutrino Observatory. The locations of the IceCube Laboratory, IceTop, and the DeepCore array are also shown. The coloured string locations at the surface of the detetor indicate the summer season in which the strings were deployed. [49]

detect the Cherenkov radiation produced by the secondary particles in a cosmic ray's extensive air shower. Measuring these cosmic ray air showers above the detector can also act as a veto system for IceCube's operations. If an air shower is detected coincident with a neutrino detection, it can be identified as an atmospheric neutrino rather than an astrophysical neutrino. IceTop allows detection of cosmic rays between the energies of 10^{14} and 10^{17} eV, which covers both the first and second knee of the cosmic ray energy spectrum. [50]

DeepCore

DeepCore is a dense region of instrumentation at the centre of the IceCube Neutrino Observatory designed to detect neutrinos at energies as low as 10 GeV, an order of magnitude less than the detectable energy of the rest of the detector. An additional 8 strings with multiple DOMs are used in this configuration. Of these strings, six are positioned around a central IceCube string and use DOMs containing photomultiplier tubes with a higher quantum efficiency than that of the regular IceCube DOMs. The six strings have 50 of these high quantum efficiency DOMs (HQE-DOMs) separated vertically by 7 metres between a depth of 2100 and 2450 metres. The reason for this depth is a highly absorbing and scattering dust layer located between 2000 and 2100 metres under the Antarctic ice sheet. Directly above this dust layer, each string holds another 10 of these HQE-DOMs, separated by a spacing of 10 metres. This region above the dust layer acts as an extra veto region for the DeepCore instrumentation below it. The remaining two strings are comprised of a mix of both DOMs and HQE-DOMs, which are similarly positioned above and below the dust layer. These strings, named the DeepCore Infill strings, are horizontally separated by 42 metres. DeepCore also utilises 7 of the regular IceCube strings, totalling 15 strings. The DeepCore region of the detector is primarily used to measure atmospheric neutrino oscillations and is used to search for sterile neutrinos, a hypothesised type of neutrino which only interacts via the gravitational force. [51]

IceCube Upgrade and IceCube-Gen2

The IceCube Upgrade is the planned addition of 7 strings to the IceCube Neutrino Observatory over the 2022/2023 Antarctic Summer season. There will be seven hundred different optical sensors spread over these strings, which will be concentrated within the DeepCore array with a horizontal string spacing of 20 metres and vertical sensor spacing of 3 metres, as shown in Figure 3.2. Two new types of DOMs will be used in this region, the Multi-PMT Digital Optical Module (mDOM) and the Dual optical sensors in an Ellipsoid Glass for Gen2 (D-Egg). Both of these optical modules have improved photon detection efficiencies and advanced calibration capabilities. Also on these strings will be a number of devices used for updated calibration of the Antarctic ice. This region of the detector will primarily be used for advanced neutrino oscillation studies and investigating the neutrino mass hierarchy. [52]

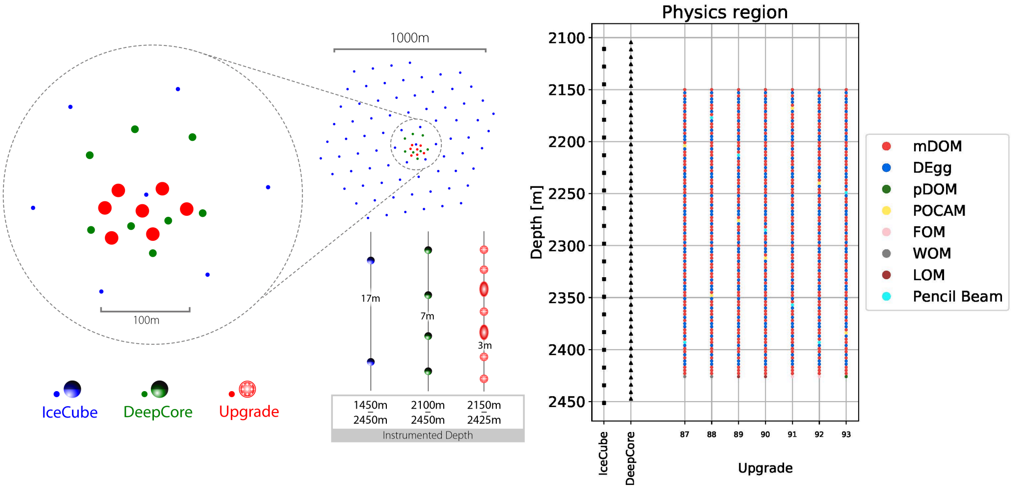

Figure 3.2: The left diagram shows the geometry of the IceCube Upgrade's position within the DeepCore array and the IceCube detector itself. The right diagram shows the position of the multiple assorted types of sensors being included in the IceCube Upgrade strings. [52]

The IceCube Upgrade, and hence the deployment of these prototype DOMs, is the first step towards the planned IceCube-Gen2 upgrade. This upgrade involves increasing the volume of the detector from one cubic kilometre to eight cubic kilometres. New strings for this configuration will be deployed with a horizontal spacing of 250 metres, rather than 125 metres, to reduce costs for the upgrade. This large increase in instrumented volume will lead to a magnitude increase in neutrino detection rates. [53]

3.2 Detection

As mentioned in Chapter 2.4, Cherenkov detectors such as the IceCube Neutrino Observatory do not detect neutrinos directly, but rather the Cherenkov radiation produced by the particles the neutrinos create when they interact. A detection of a neutrino in IceCube is classified as an "event", and the properties of this event depend on the flavour of the neutrino and how it interacts. There are three classifications of events that IceCube observes. The first is the "track" event, which occurs if a muon neutrino interacts via a charged current interaction inside or just before entering the detector. The charged current interaction produces a muon which travels through the detector, creating a track-like path, as shown in Figure 3.7. Events of this signature have approximately 1° of angular resolution. The second type of event is a "cascade" event. This occurs when an electron neutrino undergoes a charged current interaction, or if any flavour of neutrino undergoes a neutral current interaction. The result is a shower of particles with each particle producing its own Cherenkov radiation, also in Figure 3.7. As the cascade event looks the same for both current interactions, it is not possible to resolve the flavour of neutrino which caused the signal. Additionally, these events have a much worse angular resolution, between 10° and 15°, than the track events due to their geometry. The final type of event is a "double-bang" event, also depicted in Figure 3.7, which occurs when a tau neutrino undergoes a charged current interaction within the detector. Here, a tau particle is produced along with a hadronic shower. The tau particle has an extremely short lifetime of 2.9×10^{-13} s, and travels a short distance before decaying and creating a second hadronic shower, hence the double-bang. Due to the proximity of the two showers, this type of event is almost indistinguishable from a cascade event, and so far no double-bang signatures have definitively been seen in IceCube. [54]

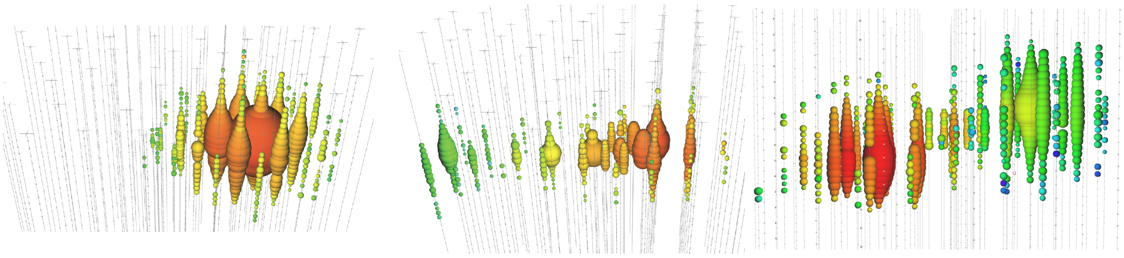

Figure 3.3: Simulations for the three types of event signatures seen in IceCube: cascade (left), track (middle), and double-bang (right). The dot in each figure represent a DOM in the detector, while the coloured sphere sizes are proportional to the amount of Cherenkov radiation received by the respective DOM. The colour indicates the arrival time of the radiation, red for an earlier time through to blue for a later time. [55]

3.3 Results

Since IceCube began taking data in 2005 with the deployment of the first string, there has been a variety of science performed across high energy astrophysics and neutrino physics. The first noteworthy example is the observation of an astrophysical neutrino flux, announced three years into IceCube's full-volume operations. Since then, a through-going astrophysical muon neutrino flux has also been identified. Another important result is the recent localisation of high energy neutrinos to blazar TXS 056+0506, making it the first known likely source of ultra high energy cosmic rays.

3.3.1 Evidence for High Energy Extraterrestrial Neutrinos

In 2013, the IceCube Collaboration reported the detection of two neutrinos with PeV energies [56]. Following this announcement, an improved sensitivity full sky search for high energy neutrinos was performed using the data from IceCube's first two years of operations. This analysis identified a further 26 high energy neutrino events, accounting for more neutrinos than expected from a purely atmospheric neutrino background [57].

In this analysis, the atmospheric neutrino background-only flux was constructed by considering a model of atmospheric neutrinos produced relative to the best estimates of the cosmic ray energy spectrum at the PeV energy range, as well as neutrinos produced due to a charm component. This charm component refers to the neutrinos produced due to the "prompt" decay of charmed mesons created in cosmic ray air showers. The flux for this type of neutrino production is much less steep than that due to regular atmospheric neutrino production, sometimes referred to as the "conventional" flux, but can become the dominant source of atmospheric neutrino flux depending on the energy being considered [58]. In total over these first two years of operations, there was an expected atmospheric neutrino and muon contribution of $10.6^{+5.0}_{-3.6}$ events between the energies of 30 and 1200 TeV. The 28 high energy neutrino events identified during this time had a significance of 3.6σ and 4.5σ, relative to two different models for the atmospheric neutrino flux due to the charm neutrino production. Under both models, more neutrinos were identified than expected due to a purely atmospheric background, and as such some of these must have an astrophysical origin. A likelihood analysis (more on likelihood analysis in Section 4.1) was performed in an attempt to identify the sources of these neutrinos, however no significant clustering hence no sources were found. The positions of these 28 events in the sky are shown in Figure 3.4.

These 28 events are also the first sample of high energy starting events (HESE) in IceCube. For an event to be classified in the HESE selection it must obey two requirements. First, the event must begin within the fiducial volume of the detector, which is surrounded by a veto region of DOMs, as shown in Figure 3.5. Specifically, no more than 2 of the first 250 observed photoelectrons (the electrons produced when the Cherenkov radiation hits a photomultiplier) can appear in the veto region. Secondly, there must be a large amount of light deposited within the detector to combat the background due to muons. The amount of photoelectrons produced must be more than 6000. The HESE selection

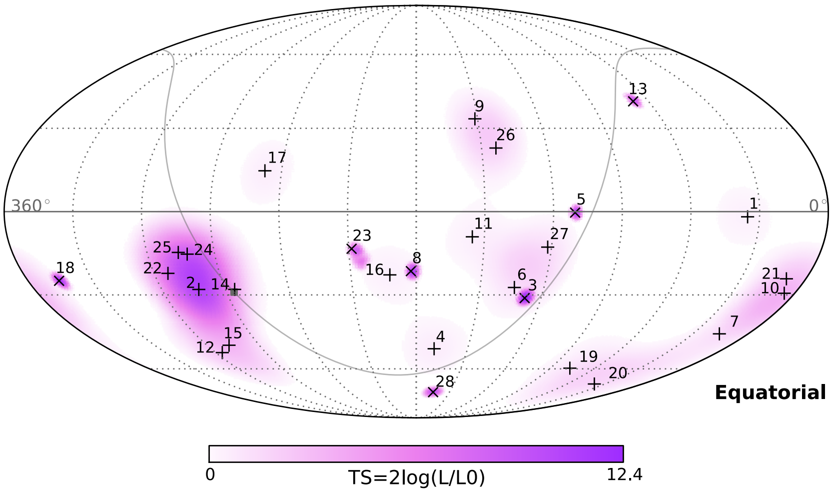

Figure 3.4: The 28 high energy neutrino events identified in the first two years of IceCube data. Their positions are shown on a sky map in equatorial coordinates, where the curved line is the galactic centre. Track events are indicated with a × while a + indicates a shower event. A colour map showing the test statistic from the likelihood analysis performed to identify clustering is also shown. [57]

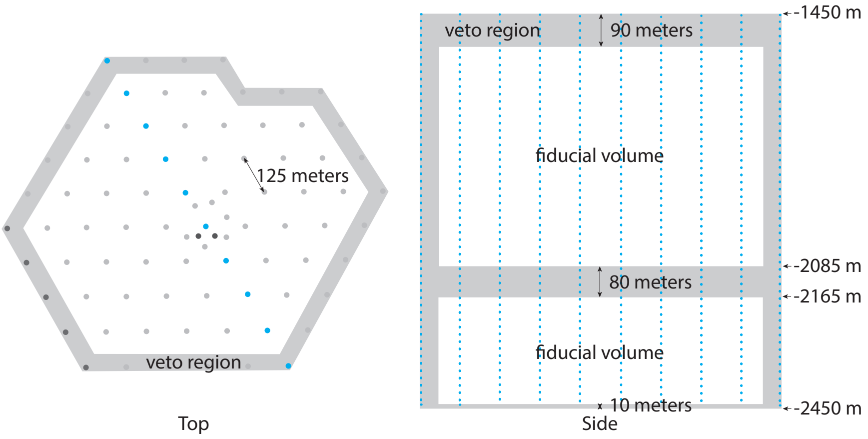

Figure 3.5: The layout of the IceCube detector including the fiducial volume and the veto region used for HESE selection. Note that the previously mentioned dust layer, where there is an increased level of light scattering and absorption, is part of the veto region. [57]

criteria is an efficient way to identify promising astrophysical neutrino events while saving computational time by immediately excluding events caused by down-going muons.

3.3.2 Through-going Astrophysical Muon Neutrino Flux

Measurements of the diffuse astrophysical neutrino flux is another important IceCube result. This flux is similar to the HESE selection in that it has particular constraints which suppress the effect of atmospheric neutrinos when searching for astrophysical neutrinos. The particular restriction here is that the diffuse flux only considers through-going muons events which come from the Northern Hemisphere, such that the Earth can absorb muons produced by atmospheric neutrinos. Although the majority of these through-going muons were due to the atmospheric neutrinos, the highest energy neutrinos in the first diffuse flux analysis were found to be inconsistent with a purely terrestrial origin at 3.7σ significance, providing further strong evidence for an astrophysical neutrino flux [59]. This diffuse flux estimate is constantly being updated as the runtime of IceCube increases, as so does the amount of events seen.

The current diffuse astrophysical neutrino flux is calculated according to the approximate 650,000 events fitting the criteria described above detected between May 2009 and December 2018. This 10 year analysis assumes a single power-law in energy and provides an updated diffuse flux of astrophysical muon neutrinos and anti-neutrinos, $d\phi_{\nu+\bar{\nu}}/dE$, of:

$$\frac{d\phi_{\nu+\bar{\nu}}}{dE} = (1.44^{+0.25}_{-0.24}) \left(\frac{E}{100\text{TeV}}\right)^{-2.28^{+0.08}_{-0.09}} \cdot 10^{-18}\text{GeV}^{-1}\text{cm}^{-2}\text{s}^{-1}\text{sr}^{-1} \tag{3.1}$$

The contributions taken into account for this estimate are conventional atmospheric neutrinos, prompt atmospheric neutrinos, and an isotropic flux of astrophysical neutrinos. [60]

3.3.3 Correlation of Neutrinos with Gamma Rays from Blazar TXS 0506+056

On September 22$^{\text{nd}}$, 2017, IceCube detected a track-like neutrino event with an approximate reconstructed energy of 290 TeV and a best fit arrival direction with a right ascension of $77.43^{+0.95}_{-0.65}$ degrees and declination of $+5.72^{+0.50}_{-0.30}$ degrees. The event, consequently called IceCube-170922A, resulted in a real-time alert being sent out to other multimessenger experiments to observe the arrival direction of the neutrino. The event was found to be coincident with a known γ-ray blazar, TXS 0506+056, by about 0.1°, shown in Figure 3.6. The follow up observations by the γ-ray instruments, the *Fermi* Large Area Telescope (*Fermi*-LAT) and the Major Atmospheric Gamma Imaging Cherenkov (MAGIC) telescopes, observed TXS 0506+056 to be in a flaring state of enhanced γ-ray emission at the time. [61]

To determine whether this neutrino was connected to the flaring blazar, a chance coincidence probability was calculated. This probability indicates whether the IceCube-170922A neutrino was correlated by chance with TXS 0506+056 while it was in a flaring γ-ray

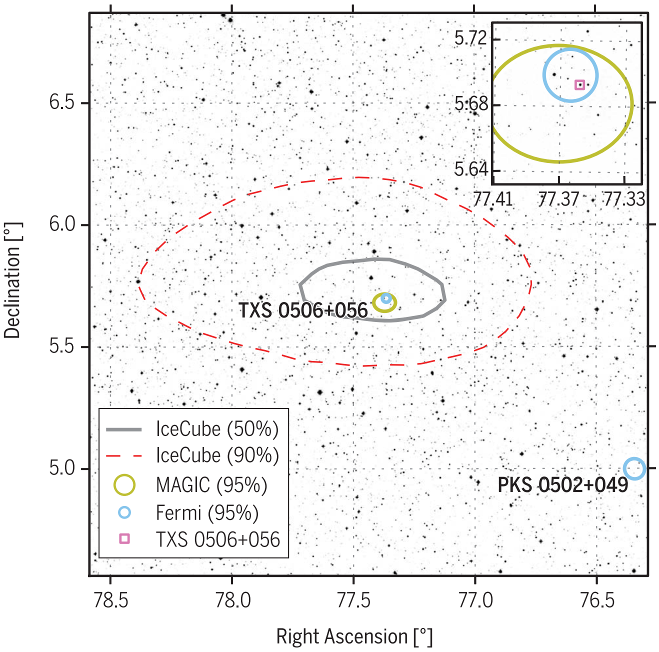

Figure 3.6: The position of the blazar TXS 0506+056 in the sky. Additionally shown are the containment regions for IceCube-170922A with 50% and 90% confidence, as well as the positional uncertainty to 95% for the γ-observations from *Fermi*-LAT and MAGIC. [61]

state. To study this, different models comparing the neutrino event to the assorted γ-ray production methods and observed fluxes were taken into account. All of these models in some way assumed that the hadronic interactions at the blazar that may produce the high energy neutrino could also produce the observed γ-ray flux. A likelihood ratio test was then performed with these models to identify the chance coincidence probability. This resulted in a chance coincidence that the neutrino was incorrectly associated with the flaring blazar of 3σ, under the requirement that the original neutrino production can be related to γ-ray production at the source or with the observed γ-ray flux. Although this was not irrefutable evidence that these observations were correlated, it was enough to prompt a study which looked at IceCube's archived data in the direction of the blazar.

Motivated by this correlation between IceCube-170922A and TXS 0506+056, the IceCube Collaboration looked back through archival data to see if there was any notable activity from the direction of this blazar in the past. A time-dependent likelihood analysis was performed over the 9.5 years of IceCube data which identified an increased period of neutrino activity from this direction between September 2014 and March 2015. This method utilised two different time windows to identify clustering, a Gaussian shape and a box-shape, which both independently identified the same significant period of neutrino activity. Specifically, 13 ± 5 events were found in excess during this period, compared to the expected atmospheric background rate. The significance of each of these events were dependent on their energies, their proximity to TXS 0506+056, and their clustering in time.

These events and the importance of these factors are shown according to their weights in Figure 3.6. To identify the significance of this time clustered selection of neutrinos, a time-dependent analysis was also performed on random datasets at the coordinates of TXS 0506+056. This resulted in a significance of 3.5σ that the neutrino clustering would incorrectly be associated with the blazar [62]. This result, accompanied by the detection of a high energy neutrino while the source was in a flaring state, suggests that blazars, such as TXS 0506+056, are likely sources of astrophysical neutrinos, and hence localised the first possible source of high energy cosmic ray acceleration.

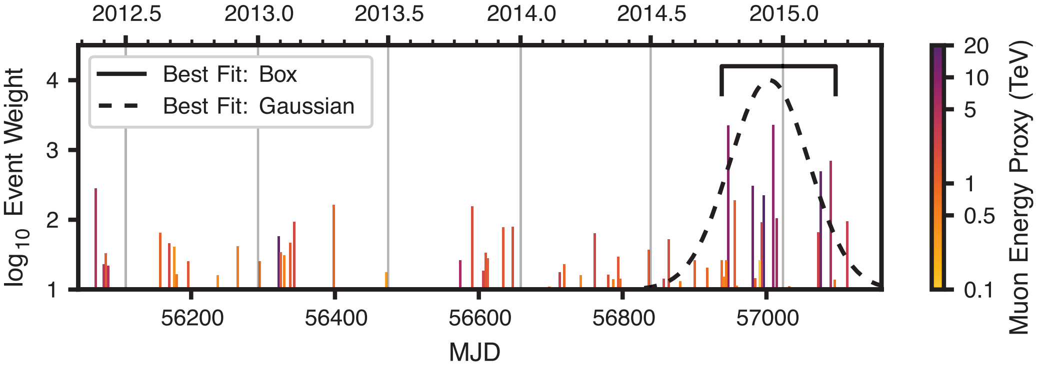

Figure 3.7: The time-independent weights of events during the IC86b period (IceCube data taken from 2012 to 2015). The weights are given according to the neutrino's energy, its proximity to the coordinates of TXS 0506+056, and its clustering in time relative to the other neutrinos. In this study, 13 ± 5 events were identified in excess to atmospheric backgrounds by a time-dependent likelihood analysis which utilised both a Gaussian time window and a box-shaped time window. [62]

3.4 Summary

The IceCube Neutrino Observatory is a Cherenkov radiation detector which instruments a cubic kilometre of ice beneath the Admunsen-Scott South Pole Station in Antarctica. It implements 5160 digital optical modules (DOMs), each containing photomultipliers and associated electronics, for detection of Cherenkov radiation produced by secondary particles resulting from neutrino interactions. Careful consideration is also implemented to reduce the background signal caused by atmospheric neutrinos produced in cosmic ray air showers. The success of IceCube over the last decade has lead to the first observed flux of astrophysical neutrinos and observation of a through-going astrophysical muon neutrino flux. Additionally, recent studies have identified the first known likely source of such a neutrino, blazar TXS 0506+056, and hence provided strong evidence for the first likely source of high-energy cosmic ray acceleration.

4 Likelihood Analysis and Importance Sampling

Likelihood analysis is a statistical method used by the IceCube Collaboration. After neutrino events are seen by the IceCube detector, likelihood analysis can be used to identify whether these events are clustered in a way that would indicate a single source produced them. In this chapter, we deconstruct the likelihood function and demonstrate how it can be used alongside neutrino sky simulations to search for astrophysical neutrino sources. An important piece of this analysis is understanding how often neutrinos randomly cluster together and imitate what neutrinos from a real source would look like. This involves numerous simulations where neutrino event data is scrambled and observing how often event clustering occurs as a result. The amount of simulations that have to be performed to view how often large numbers of these events cluster randomly increases dramatically, and as such the correct frequency for the random occurrence of large neutrino clusters cannot be obtained within any reasonable computing time. In this chapter, we introduce the concept of importance sampling, and how it can be implemented into these sorts of analyses to obtain the frequencies for these sought after rare clusters without the need of excessive computational resources.

4.1 Likelihood Analysis

Likelihood analysis involves estimating parameter values for some probability distribution using already gathered data. This can be achieved by maximising a likelihood function, which returns the parameter or parameters of a statistical model that best describe some data. In order to describe the likelihood function, first consider some random variable X, which can be obtained from a continuous probability density function, f_θ, defined by some parameter θ. The likelihood function, $\mathcal{L}(\theta \,|\, x)$, is then a function of the parameter θ given some outcome where $X = x$, as drawn from f_θ.

$$\mathcal{L}(\theta \,|\, x) = f_\theta(x) \tag{4.1}$$

Although $\mathcal{L}(\theta \,|\, x)$ is equal to $f_\theta(x)$ for x when the parameter θ is true, these two functions overall are not equivalent. A probability density function uses fixed parameters with some already known probability in order to predict any possible outcome of some model. The likelihood function uses one set of fixed data in order to find the set of parameters which would best describe the data. For example, consider a set of data, which could be obtained with two different parameter values, θ_1 and θ_2. If we find that $\mathcal{L}(\theta_2 \,|\, x) > \mathcal{L}(\theta_1 \,|\, x)$, we would conclude that the data would be more likely to have occurred if $\theta = \theta_2$. If we were to find the likelihood at each parameter θ, there would be some θ where the likelihood reaches a maximum value. This value is then the best fit likelihood, $\hat{\mathcal{L}}$, which is given by the best fit parameter value, $\hat{\theta}$, the parameter value which best describes the observed data. For convenience, the negative logarithm of the likelihood function,

$\log\mathcal{L}(\theta\,|\,x)$, can be taken [63]. Minimising the negative logarithm of the likelihood is equivalent to maximising the original likelihood and can greatly assist for computational purposes.

The likelihood function can then be used for hypothesis testing. This is performed by testing a null hypothesis, H_0, against an alternate hypothesis, H_1. The null hypothesis, H_0, says that there is nothing statistically significant about the relationship between some variables, although some varying degree of real effects can be expected. This hypothesis is assumed to be true until proven otherwise. For a typical IceCube point source case, the null hypothesis would describe that there is no significant relationship between some neutrinos and a source in the sky. The alternate hypothesis, H_1, suggests that the observed data of the experiment may indicate a relationship between some phenomena, and as such arises due to some other factor. For example, in IceCube, that neutrinos can be associated to some point source. Once these hypotheses are constructed, some data can be tested against them by constructing a test statistic, a singular value which characterises the dataset. In this case, consider that the null hypothesis and the alternate hypothesis are described by the parameter θ_0 and θ_1 respectively. A test statistic, λ, can be calculated by taking twice the negative logarithm of the ratio of the likelihood functions for these cases, which is primarily used in likelihood ratio tests. This test statistic is used for nested models, where the $\mathcal{L}(\theta_0\,|\,x)$ in question is in a simpler form compared to $\mathcal{L}(\theta_1\,|\,x)$. Specifically, θ_0 is a set value which is also a possible value for θ_1 in $\mathcal{L}(\theta_1\,|\,x)$.

$$\lambda = -2\log\frac{\mathcal{L}(\theta_0\,|\,x)}{\mathcal{L}(\theta_1\,|\,x)} \tag{4.2}$$

The test statistic can then be used to identify how significant a result is based on the difference of the ratio from, in this case, a value of zero, where $\mathcal{L}(\theta_0\,|\,x) = \mathcal{L}(\theta_1\,|\,x)$. A useful result of the test statistic constructed according to Equation 4.2 is that it may, under certain conditions, follow Wilks' theorem. This theorem states that if the null hypothesis, H_0, is true for some amount n of data, then as $n \to \infty$, the distribution of λ will follow a χ^2 distribution with a number of degrees of freedom equal to the difference between the dimensionality of free parameters in H_1 and H_0. This theorem applies under the assumption that the estimated parameters are within the possible parameter space of the likelihood functions. [64] This is a useful relationship, as if all appropriate conditions are met, a statistical significance for the hypothesis can then be calculated by considering the χ^2 distribution. Unfortunately, it is unknown whether Wilks' theorem can be directly applied to the likelihood distributions we investigate in this project. As such, we cannot simply extend a χ^2 distribution and hence must use importance sampling to forcefully obtain desired likelihood values.

4.2 Likelihood Fitting for Neutrino Data

The previously described likelihood analysis is frequently used by the IceCube Collaboration. Particularly relevant to this project is how the analysis is used to identify "hotspots" on the sky. That is, searching for points on the sky which correlate to increased amounts of neutrino emission, which may relate to a high energy neutrino source. The likelihood function typically used by IceCube in this type of analysis is given by:

$$\mathcal{L}(\theta \,|\, x) = \mathcal{L}(n_s \,|\, \vec{x}_s) = \prod_{i=1}^{N} \left[\frac{n_s}{N} S_i(\vec{x}_s) + (1 - \frac{n_s}{N}) B(\vec{x}_s)\right] \tag{4.3}$$

In this function, the value to be fitted is n_s, the number of signal neutrino events which may have originated from a point source at position, $\vec{x}_s$, on the sky. The number of events in the dataset is N, and each of these events are described by their probability density function, $S_i(x)$, which gives the expected distribution around a point in space according to the resolution of the IceCube detector. The likelihood function then accounts for the contribution of each event at the position in consideration, $\vec{x}_s$. Finally, $B(x)$ is the background probability density of the neutrinos, which details the chance that the neutrino event could have randomly occurred at any point across some space. In a typical IceCube analysis, the background probability density is estimated from experimental data [65]. Now, taking the negative logarithm of this function, we obtain:

$$-\log\mathcal{L}(\theta \,|\, x) = -\log\mathcal{L}(n_s \,|\, \vec{x}_s) = -\sum_{i=1}^{N} \log \left[\frac{n_s}{N} S_i(\vec{x}_s) + (1 - \frac{n_s}{N}) B(\vec{x}_s)\right] \tag{4.4}$$

The hypothesis testing in this situation usually goes as follows: the null hypothesis predicts that the neutrinos in some data are purely a result of the background probability density ($n_s = 0$), and the alternate hypothesis indicates the neutrinos in that same data are a combination of both signal and background neutrinos.

The signal probability density function of an event is often represented by a two-dimensional Gaussian function with the form:

$$S_i(\vec{x}_s, \vec{x}_i, \sigma_i) = \frac{1}{2\pi\sigma_i^2} \exp\left[\frac{-|\vec{x}_s - \vec{x}_i|^2}{2\sigma_i^2}\right] \tag{4.5}$$

Data taken by the IceCube detector gives the reconstructed position, $\vec{x}_i$, of the event with coordinates of right ascension and declination. The uncertainty, σ_i, indicates the resolution error in the reconstruction. Using this information, we can create neutrino skies by distributing events over some region with the properties described by their relative probability density function. After the events are distributed, a likelihood fit can be performed across the region to identify the best number of neutrino events, $\hat{n}_s$, at each position on the sky. The speed at which this likelihood fit is performed can be greatly increased by incorporating a minimising package such as MINUIT, which was incorporated into all likelihood analyses in this project. This is used rather than calculating likelihood values for all possible values of n_s to identify which value returns the maximum likelihood, which requires much more computational resources.

MINUIT

MINUIT is a package which performs the minimisation of a specified multi-parameter function. In doing so, it can take in a likelihood function and return the parameter which minimises the function, in our case, $\hat{n}_s$. To do this, MINUIT receives an input from the user of upper and lower value limits, an initial guess of the best fit parameter value, and a step size with which the minimiser tests parameter values. Here, we utilise the MIGRAD minimisation subroutine to compute the minimum of the function. In this subroutine, MIGRAD finds the numerical first derivative of the likelihood function at some point as calculated by MINUIT to determine the gradient of the function and hence which direction to calculate its next point. The subroutine continues to do this until a function minimum is reached, and the corresponding parameters are returned [66].

Performing the Likelihood Fitting

The process of performing a likelihood fit on neutrino data is shown henceforth on a Cartesian grid for simplicity. Over this space 2000 neutrino events have been randomly positioned, as shown in Figure 4.1.

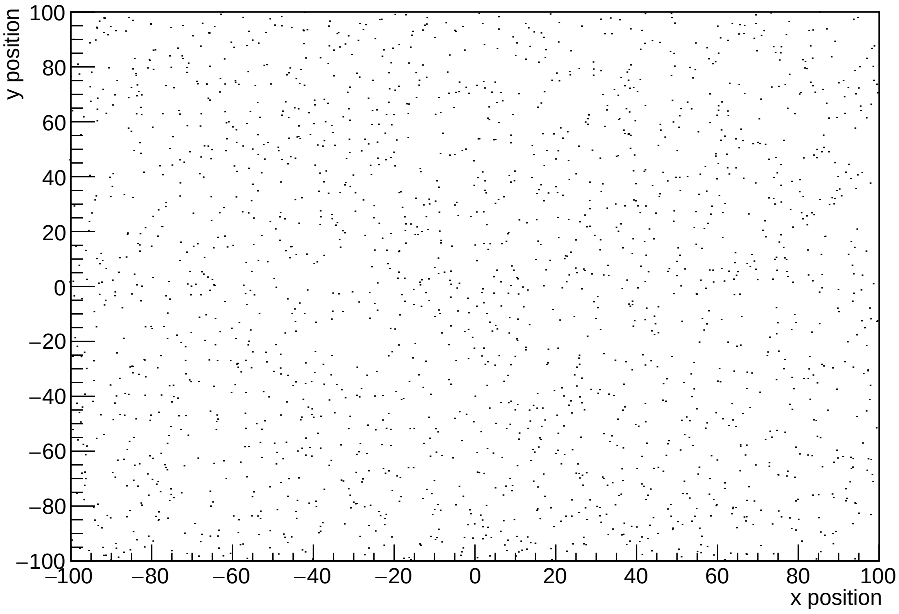

Figure 4.1: Skymap for 2000 events, each with an uncertainty of 1 unit and indicated by a black dot, randomly positioned on a 200 × 200 unit Cartesian grid.

The aforementioned likelihood fit can be performed to analyse the sky in Figure 4.1 with each position indicating its relative maximum likelihood value, $\hat{\mathcal{L}}$, or corresponding best fit number of signal events, $\hat{n}_s$, which we also refer to as the maximum signal events. As the likelihood values calculated within the work in the remainder of the thesis are extremely small values, the values are given as the logarithm of the likelihood. These results are illustrated in Figure 4.2 and 4.3.

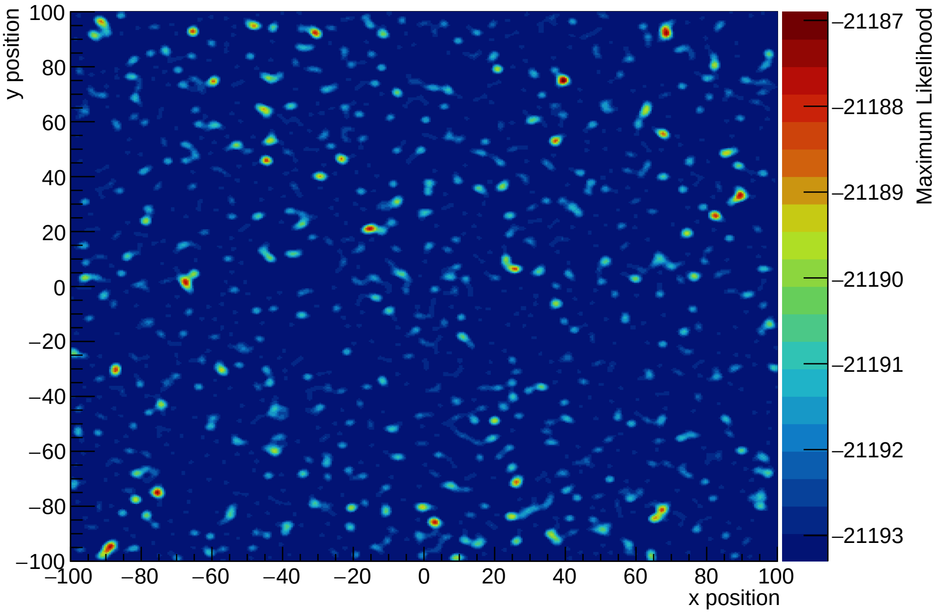

Figure 4.2: Maximum likelihood values using the 2000 events from Figure 4.1, where the colour plot shows the maximum likelihood $\hat{\mathcal{L}}$ value at each 0.5 unit interval on the sky. The hottest spot on this sky occurs at (39.5,75), where $\hat{\mathcal{L}}$ = -21186.9 and $\hat{n}_s$ = 4.63.

After this fit is performed, the maximum of the maximum likelihood values corresponds to the hotspot on the sky. The above process can be repeated multiple times on randomly generated skies to view how often clusters of neutrinos will form. Specifically, a new sky is repeatedly created, the same events are randomly placed on the sky, and the likelihood fit is performed. For each sky, the maximum of all the best fit likelihood values on the sky, $\hat{\mathcal{L}}$, and its corresponding $\hat{n}_s$ is kept. This is referred to as a full sky likelihood analysis, which we perform in Chapter 6 after we build the following machinery. This is an important aspect of neutrino point source analysis, as to calculate the significance for a real observation of neutrino clustering, we need to know how often we would expect to see the same cluster of neutrinos if a background only hypothesis is true.

First, we look at how often neutrino events will randomly cluster by considering some point on the map as the position of a potential astrophysical neutrino source. The above process is then performed, and we calculate $\hat{\mathcal{L}}$ relative to this source position. Here, we do this on the same Cartesian grid with a test source position at the origin, (0, 0). This procedure has been performed 10^5 times, such that the 2000 events have been randomly positioned each time. The maximum $\hat{\mathcal{L}}$ and $\hat{n}_s$ values at the origin are kept each time, and the distribution of these values are presented in Figure 4.4 and 4.5 respectively. In

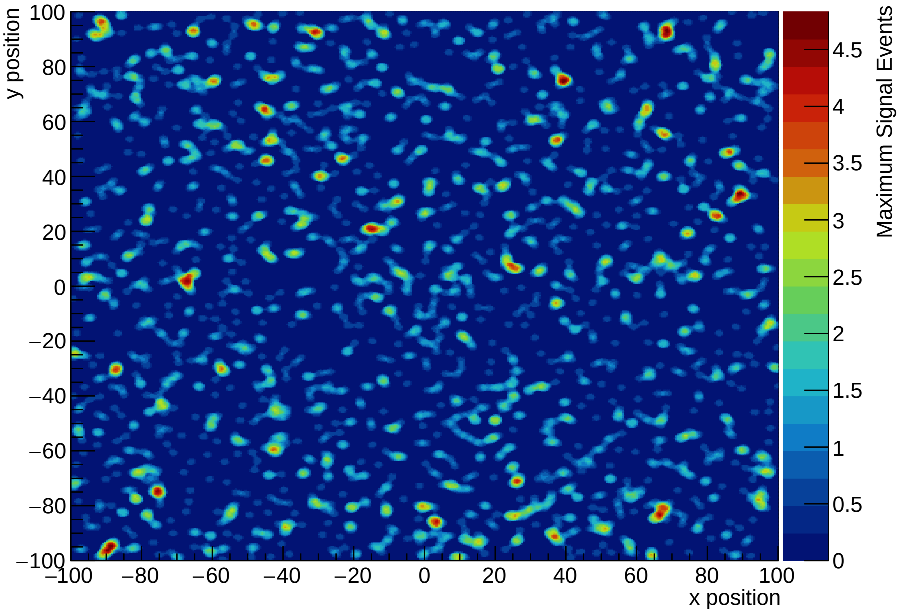

Figure 4.3: Maximum signal event values for the 2000 events from Figure 4.1, where the colour plot now shows the maximum signal events $\hat{n}_s$ at each 0.5 unit interval on the sky.

addition, Figure 4.6 shows the two-dimensional distribution of each $\hat{\mathcal{L}}$ value and its corresponding $\hat{n}_s$ from each simulation. In these figures, and for all the further figures in Chapter 4, "counts" indicates the number of skies with the indicated best fit parameter values.

From Figure 4.4 and 4.5, as expected, the majority of simulations provide us with the minimum possible $\hat{\mathcal{L}}$ for this scenario and an $\hat{n}_s$ value of zero, implying that in most of the simulations where events are randomly positioned on the sky, no event overlaps significantly with the test source position. The minimum $\hat{\mathcal{L}}$ which can be obtained for this scenario is found by simply setting n_s to zero in Equation 4.4. Here, the minimum $\hat{\mathcal{L}}$ is approximately -21193.3. As we scan through the distributions, it is clear that as the number of best fit neutrinos that randomly cluster increases, their frequency decreases. In Figure 4.6, which contains the combination of $\hat{\mathcal{L}}$ and $\hat{n}_s$ from each trial, it is also apparent that there are many combinations of the best fit parameters not seen during the 10^5 trials. For example, the largest $\hat{n}_s$ value obtained out of these trials was 7, so there is no indication as to how often we would expect a number of signal events greater than 7 to occur in this simulation. It is also important to notice that the $\hat{\mathcal{L}}$ and $\hat{n}_s$ values do not have a linear relationship. Specifically, $\hat{\mathcal{L}}$ can correlate to a number of different $\hat{n}_s$ values, and vice versa. This relationship will be explored more in Chapter 5.

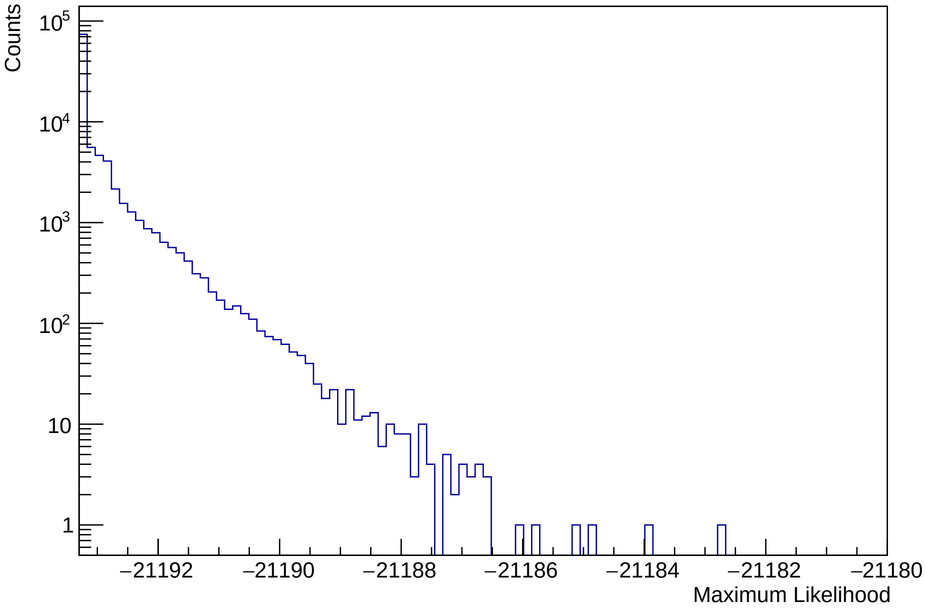

Figure 4.4: Distribution of the $\hat{\mathcal{L}}$ values found from a likelihood fit at the test source position (0, 0) for 10^5 skies with 2000 randomly distributed events.

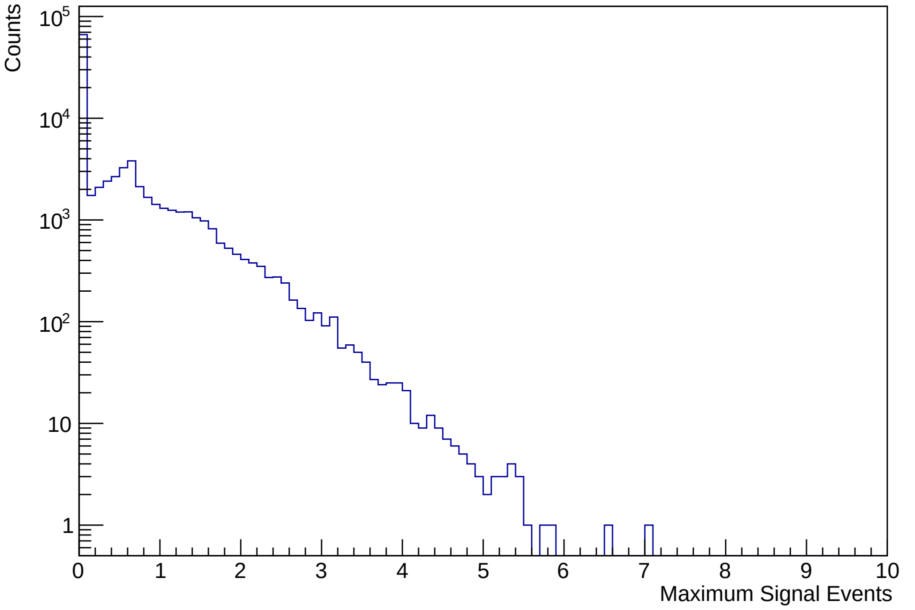

Figure 4.5: Distribution for the $\hat{n}_s$ values found from a likelihood fit at the test source position (0, 0) for 10^5 skies with 2000 randomly distributed events.

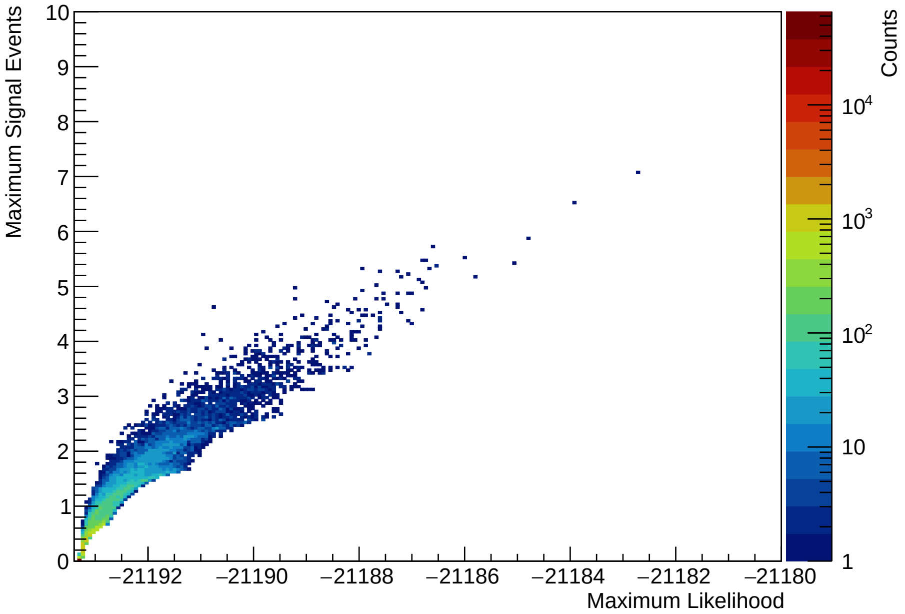

Figure 4.6: A colour plot for the combinations of $\hat{\mathcal{L}}$ and $\hat{n}_s$ from each simulation depicted in Figure 4.2 and Figure 4.3, where a likelihood fit is performed at the origin for 10^5 trials where 2000 events are placed randomly on a 200 $\times$ 200 Cartesian grid.

The parameters, $\hat{\mathcal{L}}$ and $\hat{n}_s$, obtained from each of the trials have a count of one, such that when we enter their information from each of the 10^5 trials into the histograms in Figures 4.4, 4.5, and 4.6, the integral for each of the distributions will equal 10^5. These distributions then provide an indication for how often we expect some result when the simulation is performed. For example, the total number of counts related to an approximate $\hat{n}_s$ value of zero in Figure 4.5 is around 68000. This value is also the "total count" of this bin. This implies that if this simulation is performed for 10^5 trials, an estimated 68000 of these trials will have a best fit parameter $\hat{n}_s$ of zero. By normalising the distribution, we can also say that if the simulation is performed just once, there is an approximately 68% chance we will obtain an $\hat{n}_s$ of zero. The issue with this process is that we are not obtaining large $\hat{n}_s$ values as the clustering which produces them is rare. The number of simulations that are run can be increased to identify these rare clusters, but vast amounts of extra simulations can be required simply to obtain a frequency for a given number of best fit neutrino events. This can be extremely computationally exhaustive. In addition, for $\hat{\mathcal{L}}$ and $\hat{n}_s$ values which are getting very small frequencies, we are also getting large error values as this process follows Poisson statistics (also referred to as particle counting statistics). Here, the error, σ, for some histogram bin is the ratio of the total count of the bin, X, to the square root of number of counts in the bin, N.

$$\sigma = \frac{X}{\sqrt{N}} \tag{4.6}$$

Here, the total count of the bin is equal to the number of counts in the bin, so the error is simply the square root of the total count of the bin, $\sqrt{X}$. As we are considering Poisson statistics, we ideally want large numbers of entries in some bin to obtain minimal error. As large signal events are unable to be observed in a limited number of the trials, and without an appropriate number of counts to return a small error, we can implement the method of importance sampling to obtain the desired frequencies for rare clustering in significantly less time.

4.3 Importance Sampling

As previously mentioned, it can be computationally exhaustive to understand how often rare clusters of neutrinos will appear on the sky. The approach taken in this project to combat this problem relies heavily on the concept of importance sampling. Importance sampling involves estimating values of some region in a distribution which would otherwise be difficult to obtain. This is done by creating some other distribution or distributions to sample values from, and applying an appropriate sampling weight, which indicates how often the value would be observed in the original distribution to the sampling distribution.

Suppose that the distribution of best fit likelihood values that was produced in Figure 4.4 could be represented by a probability density function $f(x)$. This is justified as we previously explained how the trials provide us with an idea of the chance that a given type of clustering would occur. The procedure previously used to obtain $f(x)$ is unfortunately limited by its inability to produce rare outcomes of neutrino clustering. In order to combat this issue, we can utilise importance sampling. We do this by constructing another sampling distribution, $p(x)$, which would reproduce the desired distribution, $f(x)$, after an infinite amount of trials. The sampling distribution $p(x)$ focuses on a particularly important region of interest of $f(x)$. In our case, $p(x)$ would be used to return likelihood values in the tail of $f(x)$, where we originally obtained no information due to a lack of statistics. As we do not naturally acquire this information, we must assign any values gathered with this method an appropriate sampling weight, W. This weight is simply the ratio of the desired distribution to the sampling distribution, evaluated at the value x:

$$W = \frac{f(x)}{p(x)} \tag{4.7}$$

Using this relationship, we now aim to construct appropriate sampling distributions in order to extend the tails of the distributions seen in Figures 4.4 and 4.6. This can be done by creating a sampling distribution, $f(x)$, which describe how neutrino events would be uniformly distributed over some space. Then, another sampling distribution, $p(x)$, is constructed such that it concentrates neutrino events around a desired point on that same space. Once the positions of the events are determined for some sky, the sampling weight, W, is then calculated and allocated to the sky. We subsequently demonstrate this process with two types of sampling distributions. The first, the Gaussian weighting method, uses a sampling distribution to concentrate the events around a set position. The second, the binomial weighting method, involves distributing events uniformly within a desired region according to a binomial probability.

4.4 Gaussian Weighting Method

The first method we have constructed for the importance sampling is a Gaussian weighting method. This process involves obtaining an event's position on the sky from a sampling function that concentrates events artificially about a desired point, and allocating each event an appropriate weight. This individual event weight is related to the sampling function used and provides an indication of how biased this forced clustering is. As the weighting method name suggests, we use a Gaussian function as the sampling function. The total weight, W, of the sky is then the multiplicative of each individual event weight. Any property of this sky will then have a weight of W. For example, the maximum best fit likelihood $\hat{\mathcal{L}}$ and best fit number of events $\hat{n}_s$ would have weight W. This weight is obtained through importance sampling and will indicate how often we expect some result in the original simulation if were able to run it for an infinite amount of time. To illustrate this method, we will apply it to the same 2000 events on a Cartesian grid sky used to create the skymaps in Figures 4.1, 4.2, and 4.3, in order to extend the distributions in Figures 4.4 and 4.6.

We begin by considering a flat normalised function, $h(x, y)$, across the grid, such that $h(x, y)$ is the same value at any (x, y). Then, we construct a normalised two-dimensional Gaussian function, $g(x, y)$, which extends over the space.

$$g(x,y) = \frac{1}{2\pi\sigma^2}\exp\left[-\left(\frac{(x-x_o)^2+(y-y_o)^2}{2\sigma^2}\right)\right] \tag{4.8}$$

We could sample the positions for all of the events according to the Gaussian, however this would concentrate all of the events around the position in consideration. This would not replicate the situation we want to observe, where some amount of events are distributed around a set position and the remaining events are distributed uniformly across the space. Due to this, we construct a sampling function, $p(x)$, which is a combination of both the flat function, $h(x,y)$, and the Gaussian function, $g(x,y)$. As such, the next step is to consider how many of the total events N, we bias to come from the test source position. This amount of events, n, sets a limit to whether an event's position is sampled according to $h(x,y)$ or $g(x,y)$. For each i^{th} event, a random number on $(0,1)$, ϕ, is generated. If $\phi < n/N$, then we sample its position from $g(x,y)$. The sampled position of an event according to the Gaussian function can be obtained using the Box-Muller transform, which is described in Appendix A. Otherwise, if $\phi > n/N$, we sample from $h(x,y)$. Either way, the event's position, (x_i, y_i), is returned. After the random number has been generated for an individual event and its position has been sampled from the respective distribution, we allocate the event a weight, w_i. As we are using importance sampling, the weight is the ratio of the desired function, $f(x)$, to the sampling function, $p(x)$:

$$w_i = \frac{f(x)}{p(x)} = \frac{h(x_i,y_i)}{(1-\frac{n}{N})h(x_i,y_i) + \frac{n}{N}g(x_i,y_i)} \tag{4.9}$$

This process is performed for all N events on the sky, with each event allocated its own weight w_i. The total weight of the sky, W, is then the multiplicative of each of the events' weights:

$$W = \prod_{i=1}^{N} w_i \tag{4.10}$$

This weight W is now given to any property of the sky, such as $\hat{\mathcal{L}}$ or $\hat{n}_s$. There are two important parameters that can be changed throughout this process in order to see rarer clustering of events. The first is the number of events we bias to come from the source position, n. By increasing this value, more events on average will be sampled from $g(x,y)$ in their relative simulation than from $h(x,y)$. This occurs as the larger the biased number of events, the greater the random number limit ϕ. The second is the value of σ in $g(x,y)$. Decreasing σ in this function creates a "tighter" Gaussian spread, so a position sampled from it will be closer to the function's centre. As w_i is inversely proportional to the combination of $h(x,y)$ and $g(x,y)$, if σ is small, $g(x,y)$ values near its centre will be large and the resulting w_i will be small. Consequently, W can then become very small, representing a clustering of events that would be near impossible due to random clustering. On the contrary, if σ is sufficiently large, $g(x,y)$ will essentially be flat over the space. This will result in an equal probability for events to be positioned anywhere on the space, regardless if their position is drawn from $h(x,y)$ or $g(x,y)$. As such, the resulting w_i values will all be approximately 1, resulting in a total weight W of approximately 1, essentially reproducing the case that all the events are randomly placed without any sampling method applied. This is important as the resulting distribution from the large σ case can be used to check that an appropriate distribution is being created.

This method can be applied to the original scenario presented in Section 4.2 in order to extend the distributions in Figures 4.4 and 4.6. In this case, as the grid is 200 × 200, $h(x, y) = 1/40000$. As we are considering the origin, $(0, 0)$, as the test source position, we centre the Gaussian at these coordinates and it takes the form:

$$g(x,y) = \frac{1}{2\pi\sigma^2 A}\exp\left[-\left(\frac{x^2+y^2}{2\sigma^2}\right)\right] \tag{4.11}$$

The new factor, A, in this function is required as depending on the σ chosen, the function may not be normalised on the Cartesian space under consideration. In this case, A is the double integral of $g(x, y)$ over the 200 × 200 grid, and ensures that the function is appropriately normalised. Additionally, when in combination with $h(x, y)$, the entire sampling distribution $p(x)$ is normalised across the space.

$$A = \int_{-100}^{100}\int_{-100}^{100} \frac{1}{2\pi\sigma^2}\exp\left[-\left(\frac{x^2+y^2}{2\sigma^2}\right)\right]\,dx\,dy \tag{4.12}$$

Furthermore, our total number of events, N, is 2000. To demonstrate the above method, we will first take the number of events we bias from the test source position, n, to be 4. For the Gaussian function, $g(x, y)$, a σ value of 500 is used to show that the "flat" Gaussian function reproduces a similar distribution as seen in Figures 4.4 and 4.6. The result of using these parameters is shown in Figures 4.7 and 4.8.

As seen, using this biasing method with the appropriate parameters returns the original distribution of $\hat{\mathcal{L}}$ and $\hat{n}_s$ values with no biasing applied. Next, the same simulation was performed 10^5 times for three different Gaussian σ values of 2, 1, and 0.5, all with a bias number of events from the source of 4. Histograms of the $\hat{\mathcal{L}}$ values from each of these simulations, including errors according to Equation 4.6, are shown in Figure 4.9, while the histograms of $\hat{\mathcal{L}}$ and $\hat{n}_s$ combinations for each simulation are in Appendix B.

We can see in Figure 4.9 that this method is returning extremely small counts, down to a minimum of 10^{-19}. As this likelihood value appeared with a count of 10^{-19} in 10^5 trials, this means we would expect this result in 1 in 10^{24} of the original trials with no sampling method applied. This is just an extreme example of the rare clustering that can be observed utilising this method. As each of the distributions in Figure 4.9 show different regions of the same distribution, we can combine them to create an overall distribution for likelihood values and their associated frequencies. This is done by constructing an average weighted mean of each bin for the overall distribution using the counts and associated errors from the relative bin of each individual bias simulation. For some number of simulations, i, which each produce a region of the overall distribution with j bins, each with a count value of x^j and error σ^j, the value of the j^{th} bin in the overall distribution, X^j, is:

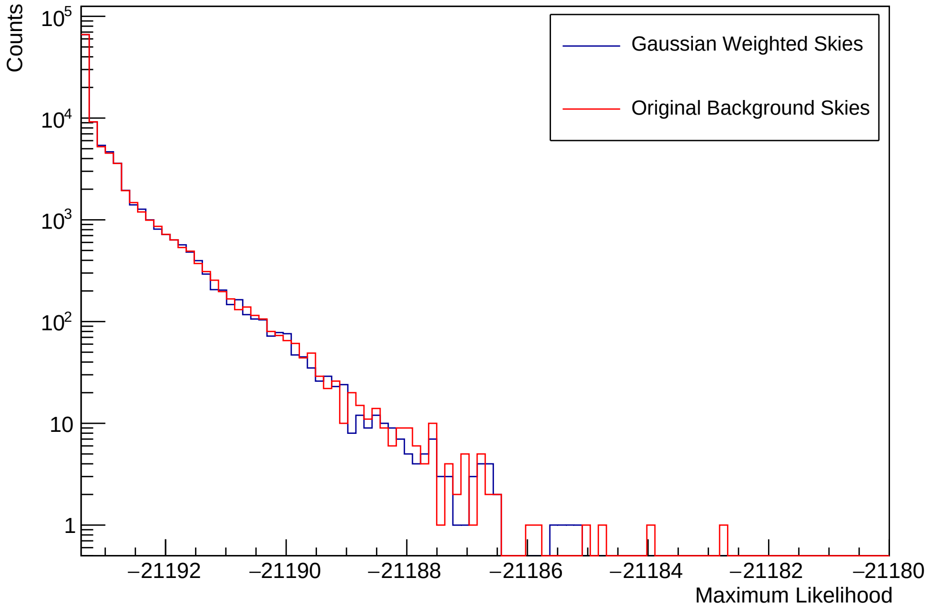

Figure 4.7: A comparison of the $\hat{\mathcal{L}}$ values from 10^5 simulations with the Gaussian weighting method applied for 4 biased events and a sampling σ of 500 against the $\hat{\mathcal{L}}$ from 10^5 of the same randomly generated skies with no biasing method applied. The two distributions are almost identical, demonstrating that the given Gaussian weighting parameters reproduce the original simulation.

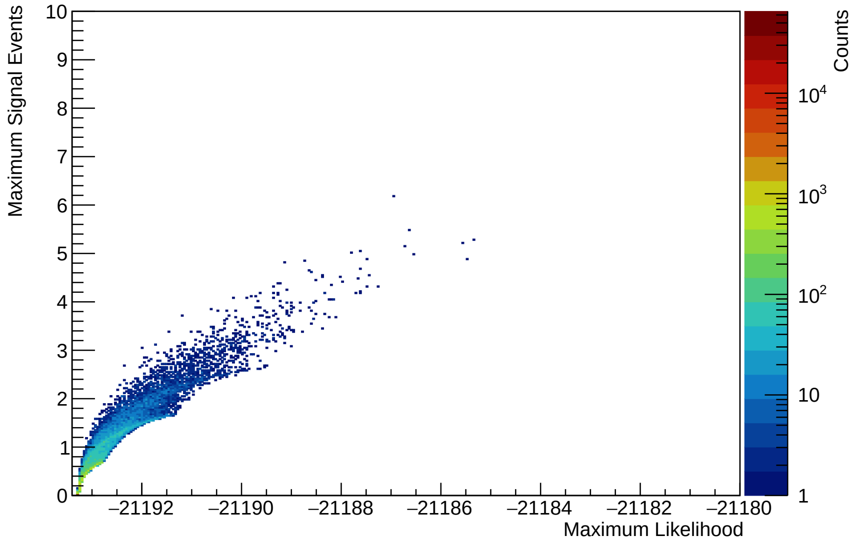

Figure 4.8: A colour plot for the combinations of $\hat{\mathcal{L}}$ and $\hat{n}_s$ from 10^5 simulations with the Gaussian weighting method applied for 4 biased events and a sampling σ of 500. This plot replicates the result of the original case in Figure 4.6 where no biasing is applied.

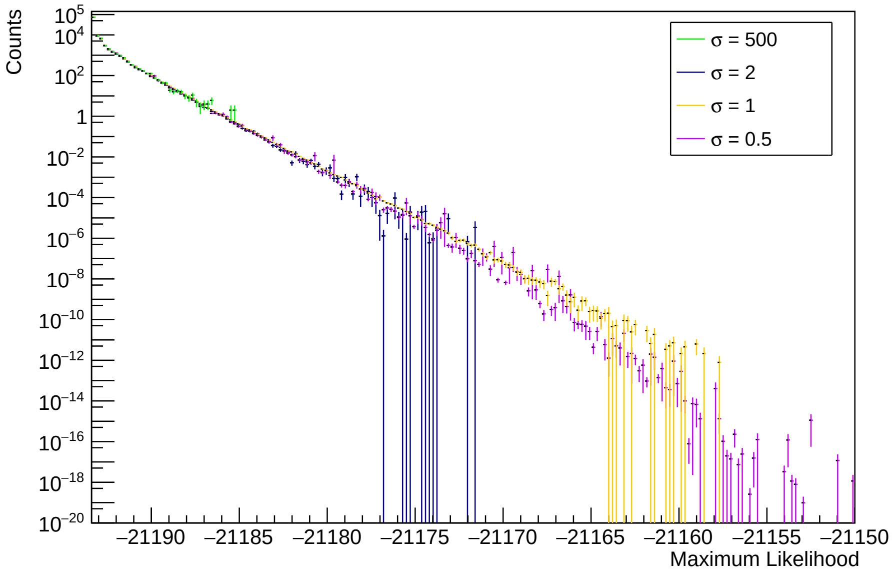

Figure 4.9: The resulting $\hat{\mathcal{L}}$ using the Gaussian weighting method with parameters of 4 biased events from a test source position at (0, 0) with a sampling σ of 500, 2, 1, and 0.5. Each individual σ value simulation was performed 10^5 times. The purple distribution is notably beneath the other distributions, due to how tight the Gaussian sampling function is.

$$X^j = \frac{\displaystyle\sum_i \frac{x_i^j}{{\sigma_i^j}^2}}{\displaystyle\sum_i \frac{1}{{\sigma_i^j}^2}} \tag{4.13}$$

This process has been adapted to the two-dimensional histograms produced with the Gaussian weighting, and has been applied to combine the simulations with the parameters of 4 biased events from the source position with assorted sampling σ values of 500, 2, 1, and 0.5, as shown in Figures 4.10 and 4.11.

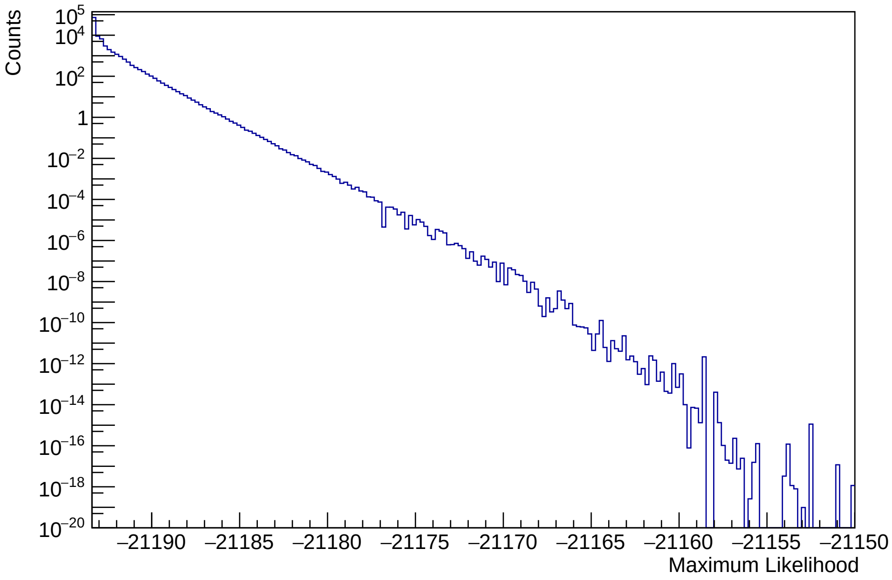

Figure 4.10: The combination of $\hat{\mathcal{L}}$ values from simulations run using the Gaussian weighting method with parameters of 4 biased events from a test source position at (0, 0) with a sampling σ of 500, 2, 1, and 0.5, each performed 10^5 times. These simulations were combined according to Equation 4.13.

So far, the Gaussian weighting method has allowed us to obtain rare clustering and assign appropriate weights for clusters of neutrino events for which simulations would have had to run for unjustifiable amounts of time to occur naturally. However, from inspection of Figure 4.10, we see that as we move along the tail of the likelihood distribution, the slope becomes steeper above an approximate $\hat{\mathcal{L}}$ value of -21170. The cause of this is explicitly seen in Figure 4.9, as the distribution due to a sampling σ of 0.5 is considerably below the distributions produced by the other σ values. This noticeably brings down the weights of the large likelihood values in the tail of Figure 4.10. A change of behaviour is also apparent in Figure 4.11, where the spread of the distribution becomes narrower as we increase in $\hat{\mathcal{L}}$ and $\hat{n}_s$. Both of these occurrences are caused by the "tight" clustering that

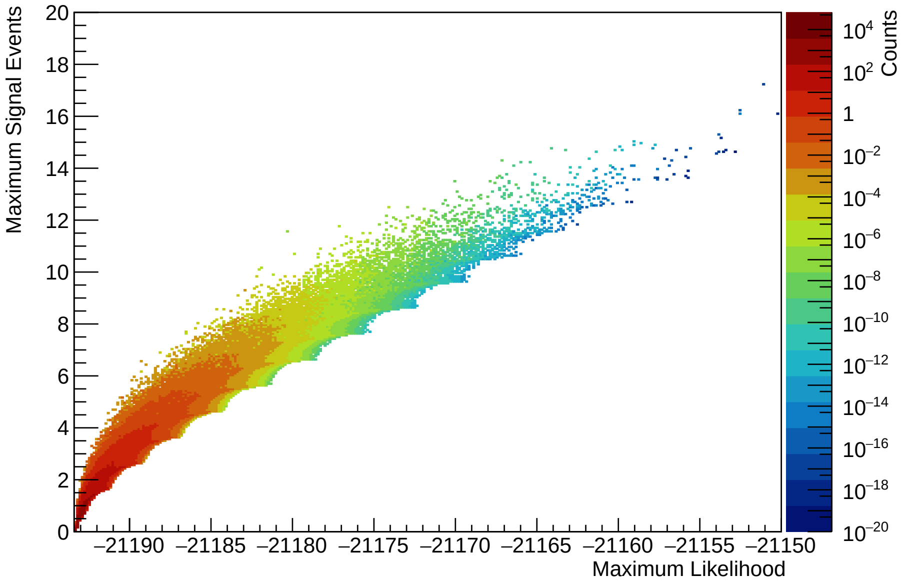

Figure 4.11: The combination of $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs from simulations run using the Gaussian weighting method with parameters of 4 biased events from a test source position at (0, 0) with a Gaussian σ of 500, 2, 1, and 0.5, each performed 10^5 times. These simulations were combined according to Equation 4.13.

occurs due to the sampling of event positions. As the Gaussian becomes narrower, the event positions we sample are forced close together around the test source position. This results in $\hat{\mathcal{L}}$ and $\hat{n}_s$ combinations which closely follow the "sawtooth" pattern along the bottom of possible values on the histogram in Figure 4.11, which indicates the maximum $\hat{\mathcal{L}}$ possible for a given $\hat{n}_s$. To obtain the currently absent lower values of $\hat{\mathcal{L}}$ for a given $\hat{n}_s$, we want to consider large amounts of events randomly clustered around some region, without being forced towards the centre of said region. To assess this, a similar weighting method was created which we refer to as the binomial weighting method.

4.5 Binomial Weighting Method

The purpose of the binomial weighting method is to obtain large signal event values for a given maximum likelihood not observed with the Gaussian weighting method. As the limitations of the Gaussian method resulted from the sampled events being tightly clustered towards the centre of some region, this method allows for uniform positioning of the events over a specified region and gives the entire sky an appropriate weight, W. To perform the binomial weighting method, we begin by considering some number of events, N. We then choose some closed region, U, on a space where the events are to be clustered. After this region is chosen, the base probability, p, is found, which indicates the probability that if an event were randomly placed over the entire space, its position would be inside U. This probability is simply equal to the ratio of the area U to the area of the entire space. Next, a bias probability, p', is chosen. This is used to determine how many of the N events are placed within the closed region. We choose the number of events in the region by sampling an integer from a binomial distribution, which is a discrete probability distribution of the form:

$$P(k \,|\, n, p) = \binom{n}{k} p^k (1-p)^{n-k} \tag{4.14}$$

where:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{4.15}$$

Here, k indicates the number of successes, whether a trial is "in" or "out", in a series of n independent trials where p is the probability that a trial will be successful. In our case, k indicates the number of events which will be placed inside the closed region. By inputting our number of events, N, and our bias probability p', we are returned some integer number which we use as the number of events placed in the closed region. The remaining events are then randomly placed outside of the closed region, and the previously used likelihood fit can be performed. The weight, W, given to this sky, and hence to any property of the sky, is a ratio of the value of the probability mass function such that we obtained k successes for N trials with the base probability p, to the value of the same probability mass function with the same values except that the probability is the bias probability p'. Once again, we can relate this to the importance sampling such that the base probability is the desired distribution $f(x)$, and the bias probability is the sampling distribution, $p(x)$:

$$W = \frac{f(x)}{p(x)} = \frac{P(k \,|\, N, p)}{P(k \,|\, N, p')} \tag{4.16}$$

Using this process and the allocated weight, we are now able to simulate the rare neutrino clustering of a large number of neutrinos equally distributed over some region of interest. By choosing a large p' (relative to p), the denominator of W will typically be larger than the numerator, as sampling from a binomial distribution with probability p' will on average return a larger integer than if sampled with p. In turn, by forcing an increased number of events into the closed region, it will be given an appropriately smaller weight. By drastically increasing p', we can force extremely rare clusters of events to occur. It is important to be able to generate these cases as they can identify assortments of clusters that the Gaussian weighting method cannot. In the binomial method, if a large number of events are forced to cluster, they are evenly distributed within the desired region. However, because of how the Gaussian weighting method works, if a large number of events are forced to cluster, their positions will be biased towards the centre of the region of interest due to how the event positions being sampled according to a Gaussian distribution. For some set-up, we can also check that this method returns the same outcome as would be expected with no biasing applied by simply setting $p' = p$. Here, events are naturally randomly distributed over the entire space and the weight of the sky would be one.

Once again, to demonstrate this method, it is applied to the same 2000 events each with a resolution error of 1 on a 200 × 200 unit Cartesian grid. This is performed to force rare clusters of neutrinos to occur, which correspond to $\hat{\mathcal{L}}$ values not seen in Figure 4.11, with appropriate weights applied. Here, we choose a circular closed region with an area of 1/2000 of the entire space's area such that the base probability, p, is 0.0005. Once again, we consider the origin (0, 0) as our test position. As such, events that are positioned within the closed region then have their coordinates (x, y) sampled according to the process described in Appendix B. This process returns a random position within a circle of radius R. As the closed region under consideration is $U = 1/2000$ of the entire 200 × 200 grid, we use an R of $\sqrt{20/\pi}$. We begin by setting our bias probability, p', to the base probability, p, to check that this method returns a similar distribution to those seen naturally in Figures 4.4, 4.5, and 4.6. The result of performing this simulation 10^5 times with $p' = p$ is shown in Figures 4.12 and 4.13.

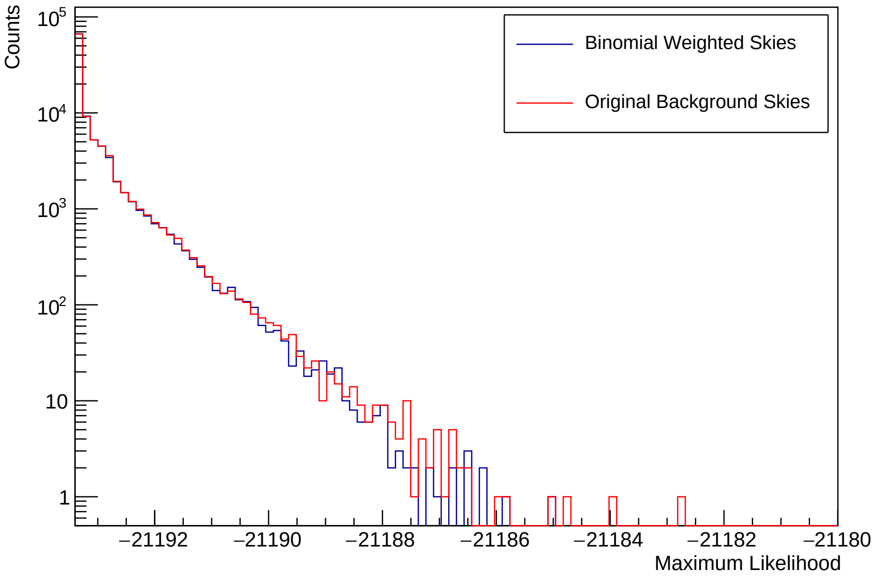

Figure 4.12: A comparison of the $\hat{\mathcal{L}}$ values from 10^5 simulations with the binomial weighting method applied for a base probability p and bias probability p' of 0.0005 against the $\hat{\mathcal{L}}$ from 10^5 of the same randomly generated skies with no biasing method applied. Both of these simulations produce similar $\hat{\mathcal{L}}$ distributions.

We see from Figure 4.12 that applying this biasing method under the correct conditions returns the natural distribution expected when no biasing is applied. Following the process of the Gaussian weighting method, the sampling parameters in the binomial weighting method are changed in order to extend the distributions depicted in Figures 4.4 and 4.6. To demonstrate this, the bias probability, p', is set to be 0.001, 0.0025, and 0.005. The results for the $\hat{\mathcal{L}}$ values from each of these simulations with appropriate errors

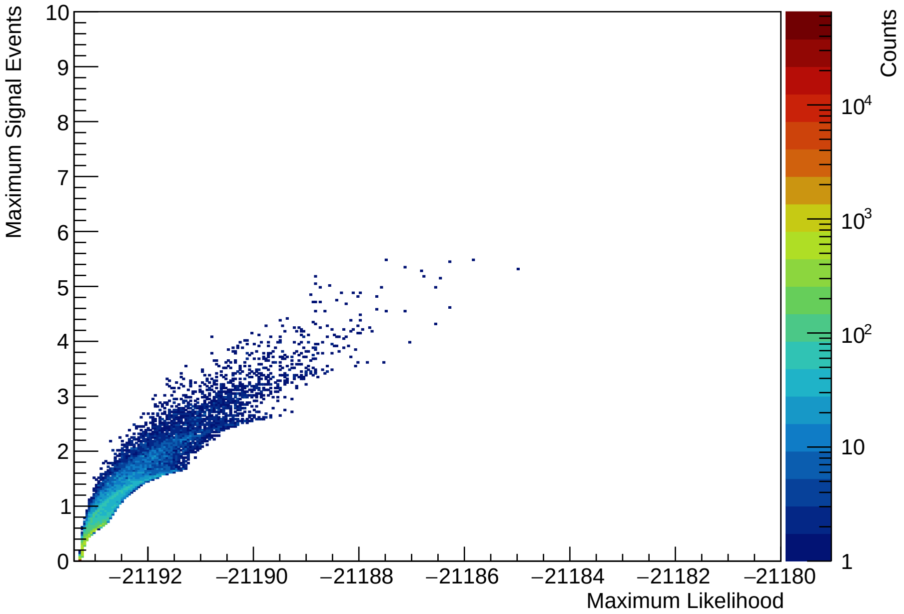

Figure 4.13: A colour plot for the combinations of $\hat{\mathcal{L}}$ and $\hat{n}_s$ from the 10^5 simulations with the binomial weighting method applied for a base probability p and bias probability p' of 0.0005. Again, this plot replicates the result of the original case in Figure 4.6 where no biasing is applied

using Equation 4.6 are shown in Figure 4.14, while the $\hat{\mathcal{L}}$ and $\hat{n}$ combination plots are in Appendix D.

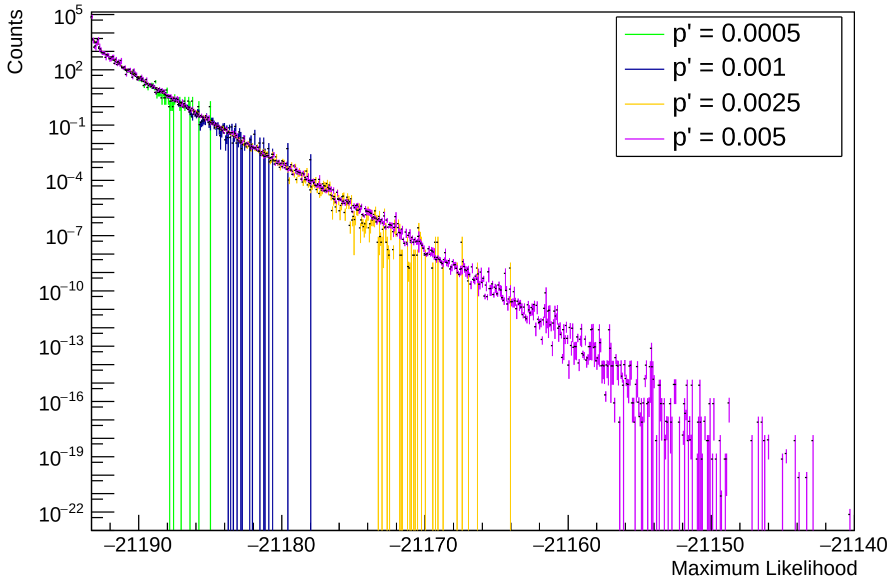

Figure 4.14: The resulting $\hat{\mathcal{L}}$ using the binomial weighting method with parameters of a base probability p of 0.0005 around a test source position at (0, 0) with a bias probability p′ of 0.0005, 0.001, 0.0025, and 0.005. Each individual p′ value simulation was performed 10^{5} times.

Again, the application of importance sampling has extended the original likelihood distribution, this time with the binomial weighting method. Now, we can again use Equation 4.13 to combine the distributions from each of these simulations. The result of combining the $\hat{\mathcal{L}}$ distributions is presented in Figure 4.15, while the combination of the $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs from each simulation is depicted in Figure 4.16.

It is clear from Figure 4.16 that the application of the binomial weighting method returns a different sample of the possible $\hat{\mathcal{L}}$ and $\hat{n}_s$ combinations than that seen with the Gaussian weighting method. However, it does have the inverse effect in that even though it has a wider spread as the distribution extends to include more of the possible combinations, it fails to obtain the values which created the clear “sawtooth” pattern previously seen in Figure 4.11. An explicit example of the difference between the regions mapped out by the Gaussian weighting method and the binomial weighted is shown in Figure 4.17.

This shows that while neither weighting method is able to fill out the entire space of possible $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs, a combination of the two methods should be able to map out the entire parameter space.

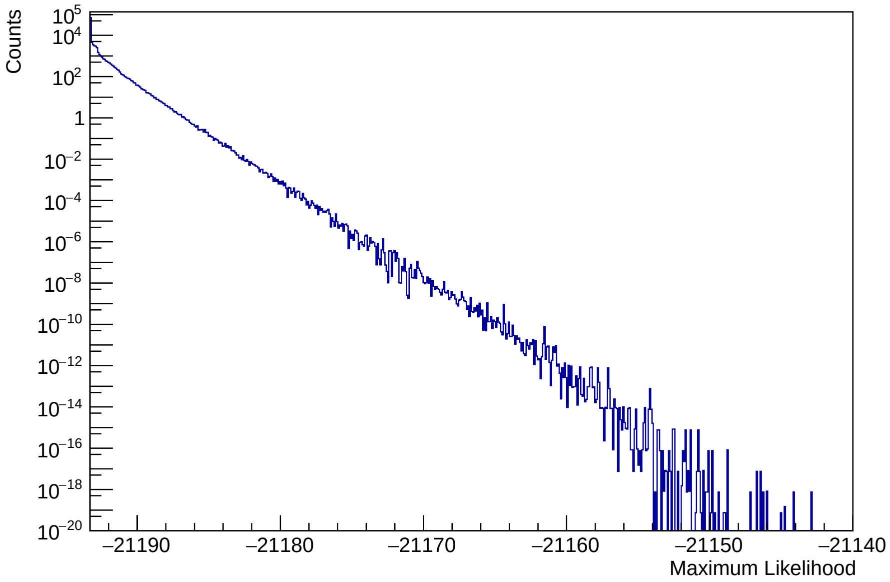

Figure 4.15: The combination of $\hat{\mathcal{L}}$ values from simulations run using the binomial weighting method with parameters of a base probability p of 0.0005 around a test source position at (0, 0) with a bias probability p′ of 0.0005, 0.001, 0.0025, and 0.005, each performed 10^5 times. These simulations were combined according to Equation 4.13.

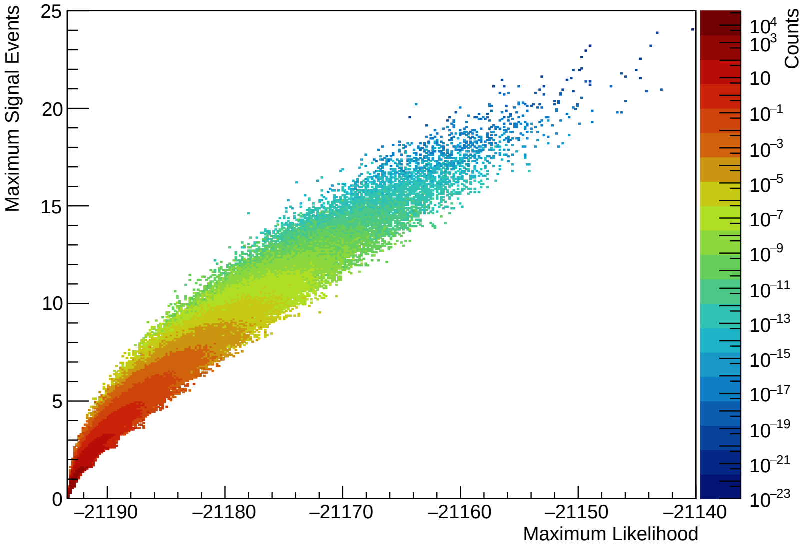

Figure 4.16: The combination of $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs from simulations run using the binomial weighting method with parameters of a base probability p of 0.0005 around a test source position at (0,0) with a bias probability p′ of 0.0005, 0.001, 0.0025, and 0.005, each performed 10^5 times. These simulations were combined according to Equation 4.13.

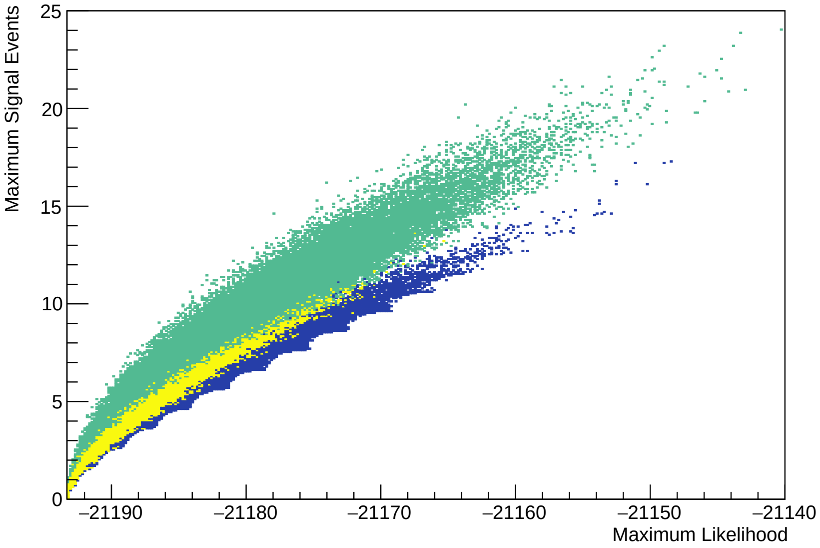

Figure 4.17: An explicit example of the different regions of the space of possible $\hat{\mathcal{L}}$ and $\hat{n}_s$ combinations obtained using the Gaussian and binomial weighting methods. The dark blue region shows combinations obtained using Gaussian weighting with 4 biased events and a sampling σ of 0.5, the green region shows combinations obtained using binomial weighting with a bias probability p' of 0.005, and the yellow region shows combinations obtained in both weighting methods.

4.6 Combining Weighting Methods

The Gaussian weighting method and the binomial weighting method have been constructed to force rare clusters of neutrinos to form, which then have appropriate weights applied to quantify the probability of observing that cluster naturally. However, as the results in each simulation are mapping out different regions of the possible $\hat{\mathcal{L}}$ and $\hat{n}_s$ combinations on the same parameter space, they need to be combined appropriately to construct full distribution that includes all possible parameter combinations and their relative weights that can be used for other analyses. Much like how different simulations with varying sampling parameters were combined within the Gaussian weighting and the binomial weighting, the results from the Gaussian weighting and the binomial weighting can also be combined for some sky. We are able to do this as we expect that for either the Gaussian or the binomial method, any set of sampling parameters will, after an infinite amount of time, return the same distribution. To demonstrate this we apply the previously described weighting methods to the same 2000 events with a 1 unit resolution error over a 200 × 200 region formerly used to illustrate the processes. A number of different parameters have been used with the weighting methods and are listed in Table 4.1. The combination of the $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs from these simulations is shown in Figure 4.18, while the combination of all $\hat{\mathcal{L}}$ values is depicted in Figure 4.19 due to a combination of the Gaussian and binomial methods against a χ^2 distribution with one degree of freedom.

Once again, Equation 4.13 was used to combine the distributions produced by each of these simulations. In addition, when the results from each simulation are combined, we choose not to include the results from bins which have less than three entries in them before the average weighted mean is calculated. This was done because bins with small numbers of entries carry large errors due to the inverse proportionality between the error and sample size as described in Equation 4.6. In turn, this can drastically affect the final result as calculated with Equation 4.13.

Table 4.1: Sampling parameters used to create Figure 4.18 and 4.19.

Weighting	Biased Events	Gaussian σ	Base Probability p	Bias Probability p'
Gaussian	4	500	-	-
Gaussian	4	2	-	-
Gaussian	4	1	-	-
Gaussian	4	0.5	-	-
Gaussian	10	2	-	-
Gaussian	10	1	-	-
Gaussian	20	1	-	-
Binomial	-	-	0.0005	0.0005
Binomial	-	-	0.0005	0.001
Binomial	-	-	0.0005	0.0015
Binomial	-	-	0.0005	0.002
Binomial	-	-	0.0005	0.0025
Binomial	-	-	0.0005	0.003
Binomial	-	-	0.0005	0.0035
Binomial	-	-	0.0005	0.004
Binomial	-	-	0.0005	0.0045
Binomial	-	-	0.0005	0.005
Binomial	-	-	0.0005	0.01

As seen on inspection of Figure 4.18, by combining the Gaussian and binomial weighting methods with assorted sampling parameters, we have effectively extended the distribution of possible $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs and their respective weights for the previously defined neutrino events on the simulated sky. This is in comparison to the combination of Gaussian method simulations in Figure 4.11 and the combination of binomial method simulations in Figure 4.16. We expect that much like the parameter pairs along the sawtooth edge of the distribution which indicate the minimum $\hat{n}_s$ for a given $\hat{\mathcal{L}}$, the cut-off of values along the "top" of the distribution indicate the maximum $\hat{n}_s$ for a given $\hat{\mathcal{L}}$. That is, a combination of parameter pairs above this region would not be seen in any simulation, as the likelihood function would never fit any of these values for this specific number of events on this specified region.

This distribution could also be extended further, notably by increasing the bias probability

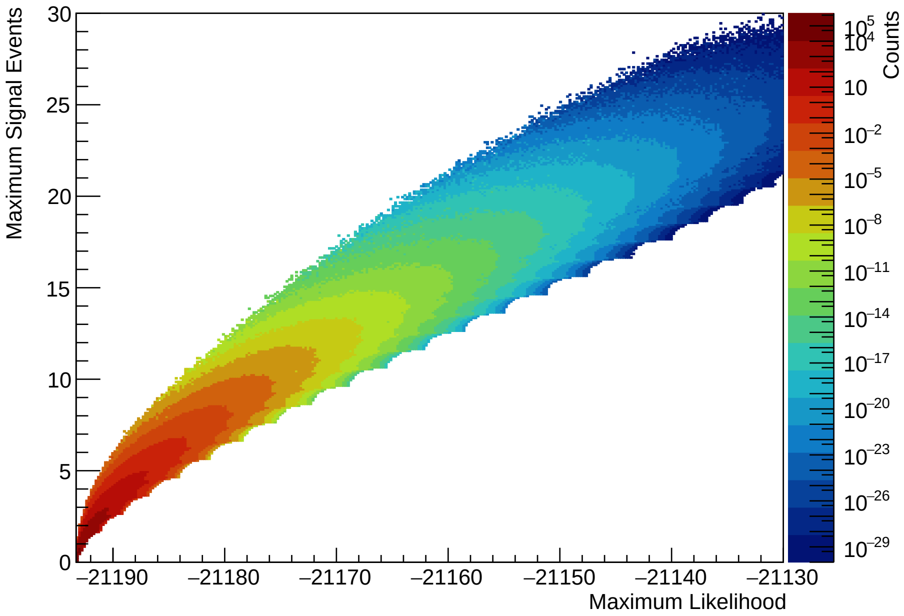

Figure 4.18: The average weighted mean of the $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs from the simulations described in Table 4.1. These simulations were combined according to Equation 4.13. This distribution has been drastically extended compared to the combination of Gaussian method simulations in Figure 4.11 and the combination of binomial method simulations in Figure 4.16.

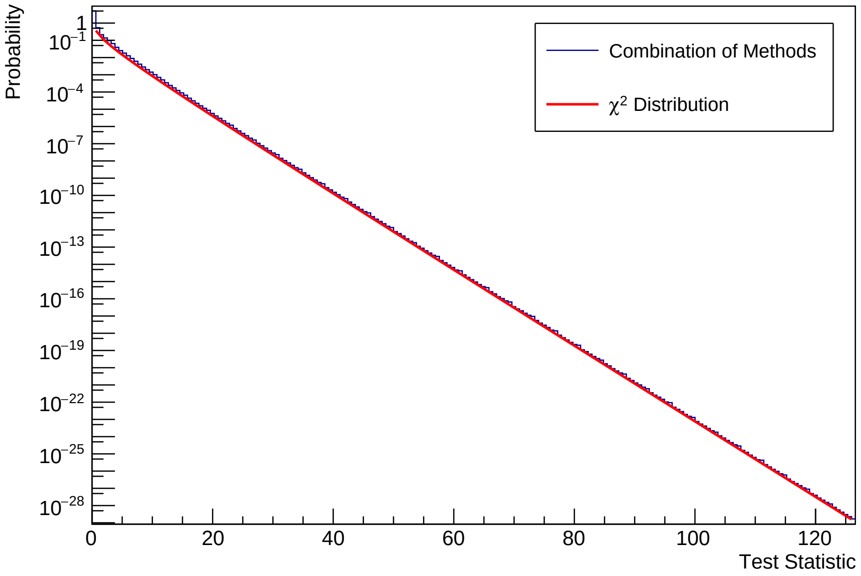

Figure 4.19: The average weighted mean of the $\hat{\mathcal{L}}$ values from the simulations described in Table 4.1 is shown in blue, plotted against a χ^2 distribution with one degree of freedom shown in red. These simulations were combined according to Equation 4.13.

for the binomial weighting or increasing the biased number of events for the Gaussian weighting. However, this may not be necessary depending on the analysis being performed. For example, clustering described by the parameters pairs indicated by the green bands in Figure 4.18 identify clustering with a probability between around 10^{-15} and 10^{-20} of occurring, which would be more than sufficient for most analyses.

It is also interesting to note that by taking the $\hat{\mathcal{L}}$ values and creating their respective test statistic, λ, as described by Equation 4.2, the distribution of test statistic values fits to a χ^2 distribution with one degree of freedom, shown in Figure 4.16. A degree of freedom of one is used as this is the difference in dimensions of parameters between a case of n_s set to zero, a fixed values, and a case of varying n_s, one free parameter. As previously described, Wilks' theorem says that if the null hypothesis is true, then the distribution of λ should follow a χ^2 distribution. This result may indicate that the combination of the weights created using importance sampling are returning the correct weights for how often these clusters would occur naturally, if we were to assume that Wilks' theorem describes the expected answer. Additionally, it can provide a comparable significance from a χ^2 test to the significance calculated from the test statistic of these results, which will be explored more in Chapter 5. This is only useful however if we were sure that our hypotheses can be considered under Wilks' theorem, which as previously mentioned, we are unsure whether it can.

4.7 Summary

Likelihood analysis is an extremely useful and frequently utilised method applied to many different analyses by the IceCube Collaboration. This method is heavily used for point source analysis, where we need to understand how often we would expect to see clustering of neutrino events due to random background to compare to real neutrino event clustering seen in IceCube. This process is limited by the amount of neutrino sky simulations required to find the frequency of this clustering and the computational resources required to perform them. Here, we have used the method of importance sampling to force rare clustering of neutrinos to appear, and allocated them weights which indicate how often they would appear in the background-only events simulation. Specifically, we have introduced two new methods for this importance sampling, a Gaussian weighting method and binomial weighting method, and demonstrated how a combination of the two can effectively obtain weights for any given clustering of a number of events on a specified region. Specifically, we were able to identify very rare clusters of neutrinos which have probabilities of less than 10^{-28} of occurring.

5 Application of Importance Sampling for Parameter Pair Constructed Test Statistics

We introduced the concept of test statistics in Chapter 4. Often denoted with λ, test statistics are a singular value which characterises a dataset. These values can then be used to identify how significant a result is, and are used in hypothesis testing. In this chapter, we investigate the effectiveness of combining the maximum likelihood, $\hat{\mathcal{L}}$, and its corresponding best fit maximum signal events, $\hat{n}_s$, into a new single test statistic. We do this by performing likelihood analyses on skies created according to a null hypothesis and an alternate hypothesis. Two-dimensional histograms containing the $\hat{\mathcal{L}}$ and its $\hat{n}_s$ are created, and a ratio of the frequency of each $\hat{\mathcal{L}}$ and $\hat{n}_s$ pair in both hypotheses can be taken as the test statistic. However, importance sampling is required to use the test statistic correctly, as we require knowledge of the frequency of the rare $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs, corresponding to rare clusterings of neutrinos. This new test statistic is then compared to $\hat{\mathcal{L}}$ using a power versus significance analysis, which according to the Neyman-Pearson Lemma, is the most powerful test statistic.

5.1 Test Statistics

5.1.1 Neyman-Pearson Lemma

The Neyman-Pearson Lemma describes the best test statistic that can be used in hypothesis testing. First, consider some hypothesis test, where the null hypothesis H_0 and the alternate hypothesis H_1 are described by the parameters θ_0 and θ_1 respectively. Now consider a test statistic, $\lambda(x)$, which is a likelihood ratio of H_0 to H_1, where H_0 is rejected by H_1 at some threshold $\lambda(x) \leq \xi$ with a significance of α. This significance is then:

$$\alpha = P(\lambda(x) \leq \xi \,|\, H_0) \tag{5.1}$$

with:

$$\lambda(x) = \frac{\mathcal{L}(\theta_0 \,|\, x)}{\mathcal{L}(\theta_1 \,|\, x)} \tag{5.2}$$

The Neyman-Pearson Lemma states that the test statistic, $\lambda(x)$, is then the most powerful test statistic at the significance level α [67], and applies to fixed hypotheses. This is similar the test statistic used for the likelihood ratio test described in Chapter 4, apart from applying twice the negative logarithm. In this chapter, we investigate a new test statistic, based on a combination of the maximised likelihood $\hat{\mathcal{L}}$ and best fit $\hat{n}_s$, and compare how powerful it is to only using the maximum likelihood $\hat{\mathcal{L}}$ as the test statistic.

5.1.2 Power and Significance

Statistical power and significance are important concepts used in hypothesis testing. Significance, α, indicates the probability that the null hypothesis is rejected when it is in fact true. This outcome is also referred to as a type I error. A type II error is then the probability of correctly rejecting the null hypothesis when the alternate hypothesis is true, and has probability β of occurring. As such, the statistical power is then $1 - \beta$, and describes the probability of accepting the alternate hypothesis when it is true. For some given distributions of test statistics constructed for a null hypothesis and an alternate hypothesis, we can calculate the statistical power and significance for any given test statistic in either distribution. We can then compare this to a test statistic from some experimental result to conveniently find its statistical power and significance. A graphical example of this is illustrated in Figure 5.1, where likelihood values are considered as the test statistic in some null and alternate hypothesis.

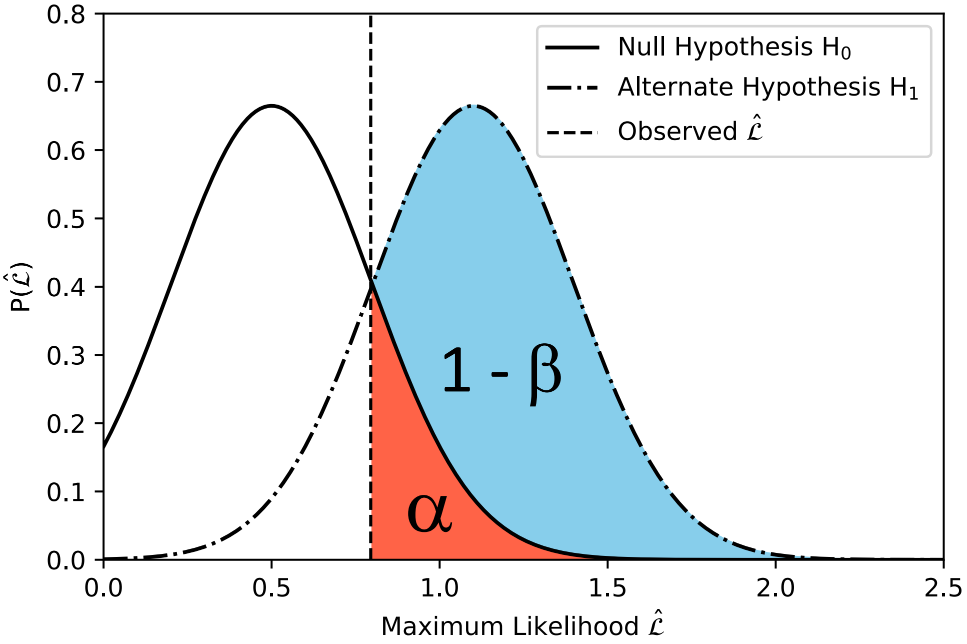

Figure 5.1: An example of distributions describing the probability of obtaining some likelihood value for a null hypothesis and alternate hypothesis. An observed likelihood value at the indicated point will have some statistical significance α and statistical power $1 - \beta$. These values are equal to the area under the null hypothesis distribution (red) and alternate hypothesis distribution (blue and red) for $\hat{\mathcal{L}}$ greater than the critical value indicated.

If the ratio between the $P(\hat{\mathcal{L}})$ for the null hypothesis to the $P(\hat{\mathcal{L}})$ for the alternate hypothesis is taken for any given $\hat{\mathcal{L}}$, the smallest likelihood ratio value indicates the likelihood value which, if observed, provides the strongest indication that the alternate hypothesis is better than the null hypothesis. This minimum likelihood ratio indicates the likelihood value at which we begin the rejection region, α, of rejecting the null hypothesis, which for the likelihood ratios, increases as the likelihood ratio value increases. For the distributions in Figure 5.1, the significance increases from right to left for a given likelihood value as the likelihood ratio is smallest at the larger likelihood values. As an example, consider the null and alternate hypothesis distributions of likelihood values in Figure 5.1, and an experiment that produced the likelihood $\hat{\mathcal{L}}$ at approximately 0.8. At this point, consider that the area under the null hypothesis distribution from this observed likelihood to any greater likelihood value is 0.1. This value indicates a significance, α, of 0.1, and implies that there is a 10% chance that the likelihood value would be observed if the null hypothesis were true. Now, at the same experimental likelihood, consider the area under the alternate hypothesis distribution between this likelihood value and any greater value is 0.9. This value is the statistical power, $1-\beta$, and indicates that there is a 90% chance that the likelihood value would be observed if the alternate hypothesis were true. Ideally, the distributions should be separated as much as possible to obtain a small significance with a large power, such that we can predominantly accept an experimental likelihood value according to the alternate hypothesis with little chance that the null hypothesis also explains it. This same procedure can be performed at every likelihood value to obtain a power and significance for any given likelihood that could result from the experiment. This information can then be plotted to create a power versus significance plot. An example of applying this process to the distributions in Figure 5.1 is shown in Figure 5.2. Each point on Figure 5.2 shows some possible likelihood value observable in either hypothesis, and the power and significance for that value if it was observed. Ideally, the power vs. significance would be pushed up into the top left corner, such that there is a large probability of accepting any result if the alternate hypothesis is true, and a small chance it would be observed if the null hypothesis is true.

Graphing the power and significance in this manner due to different null and alternate hypothesis combinations can be extremely useful to understand the best test statistic to test some hypothesis, as it conveniently shows how the significance changes with power for a tested hypothesis. Ideally, we desire the highest attainable power for a required level of significance.

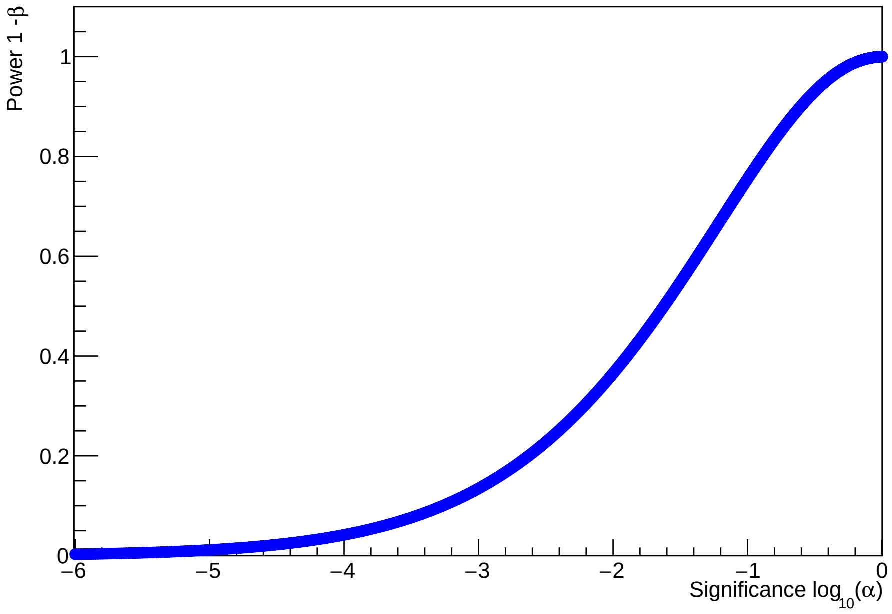

Figure 5.2: A power versus significance plot created by integrating the null and alternate likelihood value distributions from Figure 5.1. Each point indicates the power and significance of a given likelihood which may be observed due to either hypothesis.

5.2 Creating Test Statistics and Comparing Power and Significance

In this section we investigate the effectiveness of using a test statistic which is a combination of $\hat{\mathcal{L}}$ and $\hat{n}_s$. This is created by producing a two-dimensional plot of $\hat{\mathcal{L}}$ and its corresponding $\hat{n}_s$ for a number of trials representing both a null hypothesis and an alternate hypothesis. The counts in the two-dimensional plot indicate the frequency that a given $\hat{\mathcal{L}}$ and $\hat{n}_s$ pair occurs in the relative hypothesis. We can then take a ratio of the frequency due to both hypotheses, and use it as our new test statistic, which we refer to as the “parameter pair ratio” test statistic or “ratio” for short. To accurately investigate this, we require the use of the Gaussian and binomial importance sampling methods to obtain rare clusters and hence represent all relevant $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs. This test statistic is then compared, via a power vs. significance analysis, to the maximum likelihood, $\hat{\mathcal{L}}$, on its own.

5.2.1 Maximum Likelihood and Maximum Signal Events as a Test Statistic

In order to explore this concept, we once again consider the distribution of 2000 neutrino events with resolution error of $\sigma = 1$ across a 200 × 200 Cartesian space. For this, we create test statistics from a null hypothesis, H_0, where we expect zero signal events from a possible source position, and from an alternate hypothesis, H_1, where we expect three signal events from the source position. We set the origin of the Cartesian grid, (0, 0), as our test source position and perform a likelihood fit at this point, to find the best fit parameters $\hat{\mathcal{L}}$ and $\hat{n}_s$. To investigate the null hypothesis, we create a neutrino sky by randomly positioning the 2000 events across the space. For the alternate hypothesis, where we expect three signal events at the test source position, we must distribute three of the total number of events around this point. We treat the positions of these source events as normally distributed around the origin, and as such their coordinates can be obtained using the Box-Muller transform described in Appendix A. The remaining events are then placed randomly on the sky. Here we have performed both the null hypothesis and the alternate hypothesis simulation for 10^6 trials. Specifically, random neutrino skies have been simulated and a likelihood fit has been performed at the origin for both the null hypothesis and alternate hypothesis cases to obtain the best fit parameters $\hat{\mathcal{L}}$ and $\hat{n}_s$. The distribution of $\hat{\mathcal{L}}$ values from both hypotheses are plotted in Figure 5.3, while the $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs from each hypothesis are shown in Figures 5.4 and 5.5. Again, the "counts" in each of the figures in Chapter 5 indicate the number of skies with the indicated best fit parameter values, or indicated ratio values.

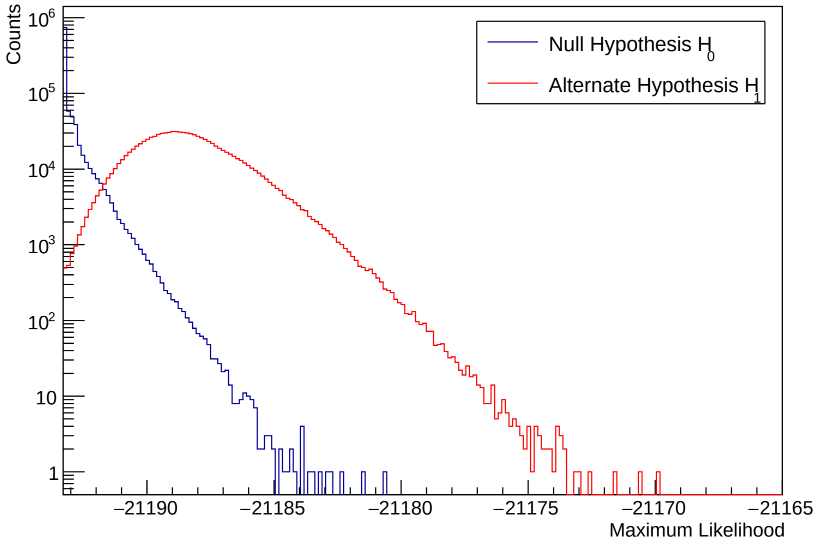

Figure 5.3: The maximum likelihood $\hat{\mathcal{L}}$ values from 10^6 trials for both the null hypothesis (blue) and the alternate hypothesis (red). As expected, the majority of null hypothesis trials fit to the minimum possible likelihood of -21193.3, while the majority of alternate hypothesis trials fit to a greater likelihood, indicating the presence of at least one neutrino event near the origin.

Upon inspection of Figure 5.3, both hypotheses clearly produce different $\hat{\mathcal{L}}$ distributions. Furthermore, the alternate hypothesis plot for $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs in Figure 5.5 shows the expected peak of $\hat{n}_s$ values at three, as this is the amount we forced to be distributed around the origin. Any values that are larger than this are due to some number of the remaining background events on the sky overlapping with this position. Any values that

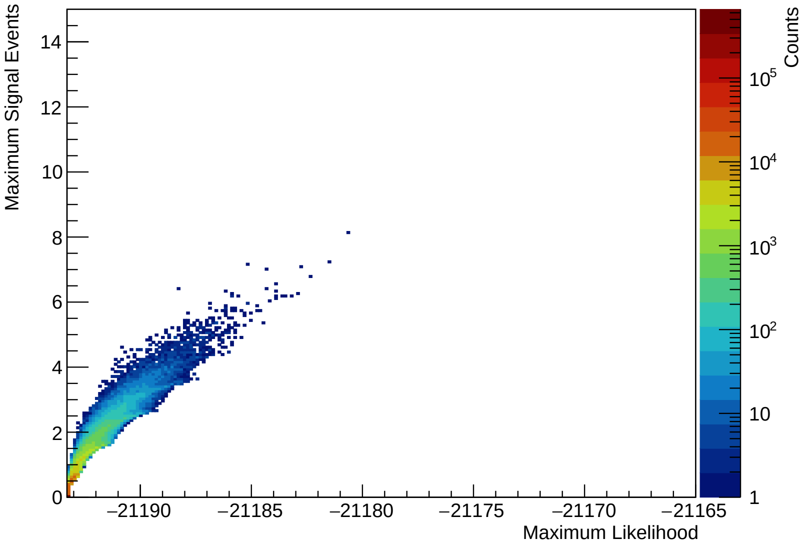

Figure 5.4: A colour plot for the $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs from each of the 10^6 null hypothesis trials where 2000 events are randomly positioned on a 200 × 200 Cartesian grid and a likelihood fit is performed at (0, 0). The majority of trials fit to an $\hat{n}_s$ of 0, and there is a notable lack of counts above an approximate $\hat{n}_s$ of 3, which is expected due to no biasing of events.

are less this occur due to the Box-Muller transform distributing the three events away from one another relative to the origin, resulting in a smaller fitted $\hat{n}_s$. The $\hat{\mathcal{L}}$ distributions can now be used to produce a power vs. significance plot which can be compared to a given test statistic from experimental data.

5.2.2 Parameter Pair Ratios as a Test Statistic

The idea of combining $\hat{\mathcal{L}}$, and its best fit signal events, $\hat{n}_s$, into a single test statistic has previously been suggested to provide a more powerful test statistic than $\hat{\mathcal{L}}$ on its own [68]. This may seem counter-intuitive, as we have already detailed the Neyman-Pearson Lemma and its consequences, but it sounds plausible that if a test statistic parametrises a trial to both the $\hat{\mathcal{L}}$ and $\hat{n}_s$ value it fits to, we may be able to construct a test statistic more useful than using the $\hat{\mathcal{L}}$ value alone. As mentioned in Chapter 4, the parameter pairs do not follow a one to one relationship, such that $\hat{\mathcal{L}}$ can correspond to multiple assorted $\hat{n}_s$ values and vice versa. This relationship can possibly used to our advantage to create a new test statistic. This test statistic is the ratio of how often $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs are observed for the alternate hypothesis to how often the same $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs are observed for the null hypothesis. We can describe the histograms in Figure 5.4, the null hypothesis, and Figure 5.5, the alternate hypothesis, as having bins with total counts $x^{jk}_{H_0}$ and $x^{jk}_{H_1}$ respectively where each histogram has j rows and k columns. The j rows then separate

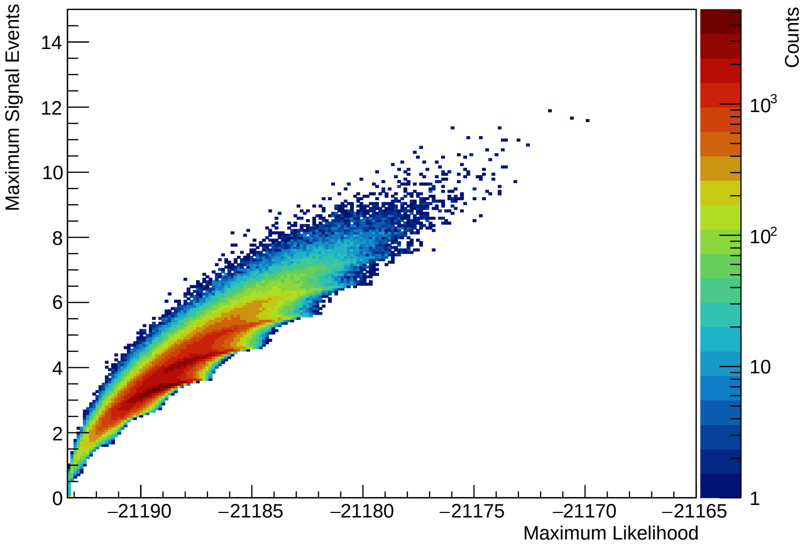

Figure 5.5: A colour plot for the $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs from each of the 10^6 alternate hypothesis trials where 3 events are distributed according to a Box-Muller transform around (0, 0) and the remaining 1997 events are randomly positioned on a 200 × 200 Cartesian grid. The likelihood fit is then performed at (0, 0). As expected, the $\hat{n}_s$ counts peak at 3 due to how the alternate hypothesis was constructed. There is also a slightly smaller peak at 4, indicating that in a numerous trials, one of the background neutrinos has also been positioned close to the origin.

the $\hat{n}_s$ values into a discrete range, while the k columns do the same for the $\hat{\mathcal{L}}$ values. The ratio value for each respective bin, X^{jk}, is then calculated with:

$$X^{jk} = \frac{x^{jk}_{H_1}}{x^{jk}_{H_0}} \tag{5.3}$$

The result of these calculations is shown in Figure 5.6. There are however a number of $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs in the alternate hypothesis plot (Figure 5.5) which do not have a corresponding value in the null hypothesis plot (Figure 5.4), so the respective bins in the ratio plot, Figure 5.6, have been left empty, as we cannot divide the entry by zero.

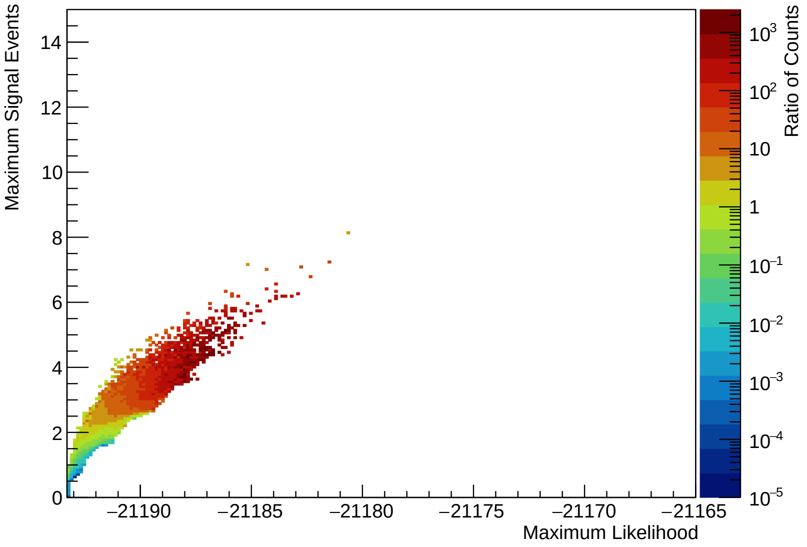

Figure 5.6: A colour plot indicating the ratio value of the frequency of an alternate hypothesis $\hat{\mathcal{L}}$ and $\hat{n}_s$ pair to the frequency of a null hypothesis $\hat{\mathcal{L}}$ and $\hat{n}_s$ pair. All bins which correlate to a $\hat{\mathcal{L}}$ and $\hat{n}_s$ pair of zero for the null hypothesis but have a value for the alternate hypothesis have been given a value of zero.

The next step is to allocate each of the hypothesis trials its corresponding ratio value, to create a new distribution of null and alternate hypothesis test statistics. We do this by allocating each of the trials with a $\hat{\mathcal{L}}$ and $\hat{n}_s$ pair in Figures 5.4 and 5.5 to its corresponding ratio value in the Figure 5.6 plot. As there are numerous alternate hypothesis trials without a null hypothesis trial value to divide by, we put the trial skies in this situation into an "overflow bin", such that they are still represented and are taken into account when considering the power and significance. As the ratio values vary over orders of magnitude, their distribution according to the null and alternate hypothesis is shown as a logarithmic plot in Figure 5.7, with the overflow bin set at a value of 6. Out of the 10^6 trials performed for the test hypothesis, 204190 of these have no corresponding ratio due to a lack of $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs in the null hypothesis trials, which as over 20% of trials, is a particularly large portion.

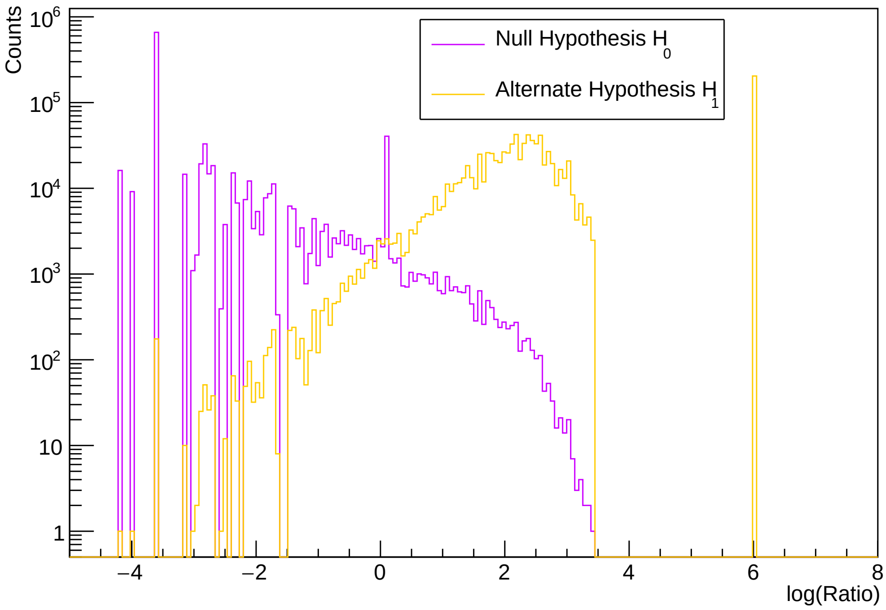

Figure 5.7: The ratio distributions where each trial with some $\hat{\mathcal{L}}$ and $\hat{n}_s$ pair from the null and alternate hypothesis has been allocated a ratio value equal to the ratio of the frequency that the parameter pair was observed in the test hypothesis trials to the null hypothesis trials. The overflow bin at 6 consists of the trials from the alternate hypothesis whose parameter pairs were not observed in the null hypothesis, hence a ratio could not be taken.

We can now create a power and significance plot for the ratio values due to the ratio distributions and compare it to the power and significance due to the maximum likelihood $\hat{\mathcal{L}}$, the best test statistic according to the Neyman-Pearson Lemma. This is shown in Figure 5.8. Both distributions stop at an expected significance of -6 as this indicates the one null hypothesis trial of the 10^6 trials with the greatest $\hat{\mathcal{L}}$ value in Figure 5.3 and greatest ratio value in Figure 5.7.

As seen in Figure 5.8, a useful power vs. significance graph for the distributions cannot be constructed as the ratio power only reaches a value of 0.2 due to the presence of the overflow bin. The reason for this cut-off is the approximately 20% of trials which were placed in the overflow bin. As the power vs. significance plot is created by integrating the null and alternate hypothesis distributions bin by bin, the point at a power of 0.2 represents 20% of the alternate hypothesis trials and none of the null hypothesis trials. This currently suggests that the ratio test statistic is an improvement over the maximum likelihood on its own, as for significance values lower than where $\log_{10}\alpha$ is approximately -3, the power $(1 - \beta)$ is higher than that for the maximum likelihood. Although, we cannot say that this is definitively any better without testing this case where no overflow bin

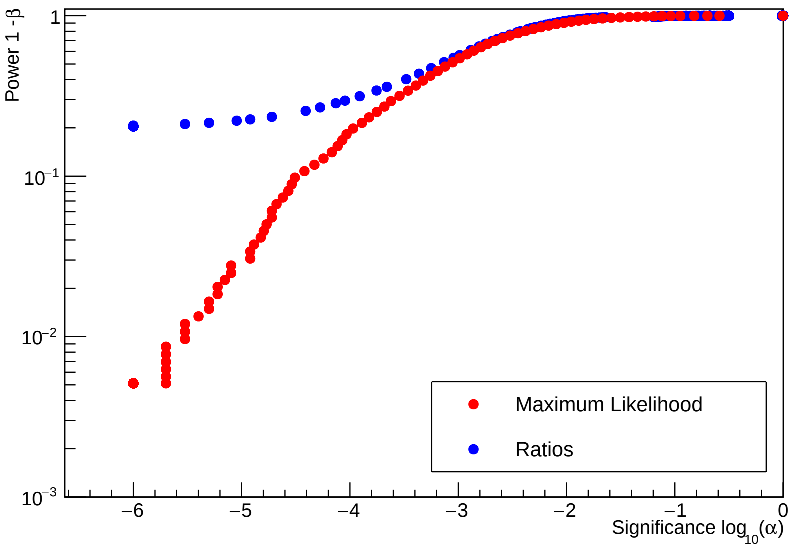

Figure 5.8: A power and significance for the distributions of test statistics for the null hypothesis and alternate hypothesis due to the maximum likelihood $\hat{\mathcal{L}}$ (red) and due to a ratio of the frequency of their parameter pairs (blue), created by integrating the distributions from Figure 5.3 and 5.7 respectively. It appears that using the parameter pair ratio is a better test statistic, however this could be an artefact of using the overflow bin.

required. Fortunately, we can utilise importance sampling to force rare clustering in the null hypothesis to occur, enabling us to obtain correct ratios for every $\hat{\mathcal{L}}$ and $\hat{n}_s$ pair. A distribution for the null hypothesis parameter pairs of interest is shown in Figure 5.9, which is constructed using the same Gaussian sampling and binomial sampling simulations listed in Table 4.1, as this was created for the same number of events on the same size Cartesian grid.

Seen clearly in Figure 5.9, using importance sampling has provided frequencies for all parameter pairs which are seen for the test hypothesis, and we can now create appropriate test statistics from the ratio values without requiring an overflow bin. The result of taking the ratio of the alternate hypothesis with importance sampling frequencies to the null hypothesis is again calculated with Equation 5.3, with these values shown in Figure 5.10. To be exact, Figure 5.10 is then just the ratio of counts in each bin in Figure 5.5 to the corresponding bin in Figure 5.9.

Now that we have ratio values for each of the $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs observed in the alternate hypothesis, we can allocate each trial from each hypothesis its relative ratio value and use this as our new test statistic. In this case, we now have multiple trials which contributed to the null hypothesis that have no corresponding weights for parameters pairs in the alternate hypothesis. However, as we divide by the frequency of the null hypothesis, the ratios allocated to these trials are simply zero. Additionally, as previously mentioned,

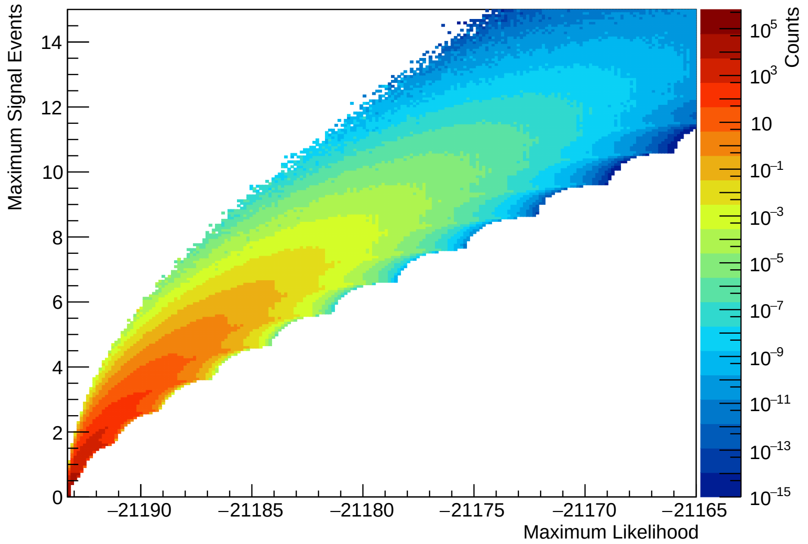

Figure 5.9: A colour plot for the $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs for the null hypothesis trials using the importance sampling methods from Chapter 4. Again, 2000 events are randomly positioned on a 200 × 200 Cartesian grid and a likelihood fit is performed at (0, 0). Both the Gaussian and binomial sampling methods have been used to extend the distribution.

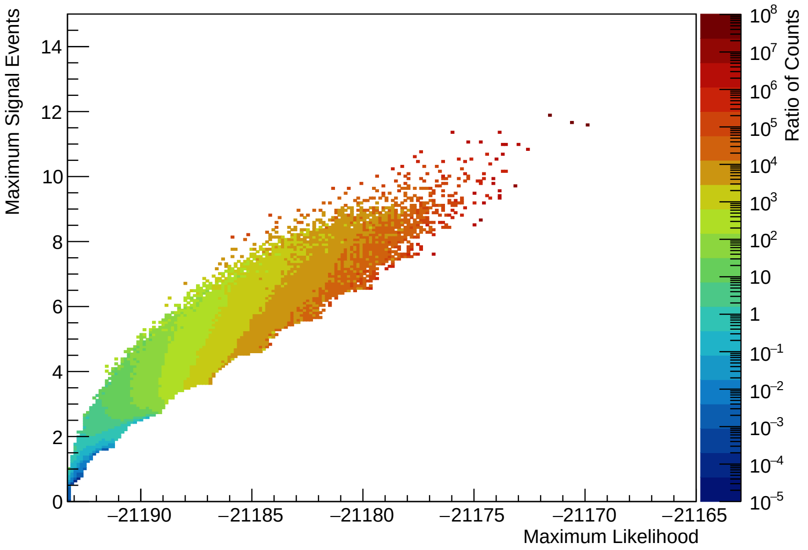

Figure 5.10: A colour plot indicating the ratio value of the frequency of an alternate hypothesis $\hat{\mathcal{L}}$ and $\hat{n}_s$ pair to the frequency of a null hypothesis $\hat{\mathcal{L}}$ and $\hat{n}_s$ pair. All bins correlating to a filled bin for the alternate hypothesis now have a respective ratio value, due to the use of importance sampling in the null hypothesis simulation.

when we use importance sampling to obtain the weight of a sky, the weight applies to any property of the sky. This includes the ratio statistics created here, such that when creating the ratio value distribution for the null hypothesis, the weight for a given trial is also given to the respective parameter pair ratio value. The resulting distributions from allocating each of the trials from the null and alternate hypothesis is shown in Figure 5.11, where each ratio value expectedly has contributions from both hypotheses' trials. The logarithm of the ratio values is shown due to the range in orders of magnitude of possible values.

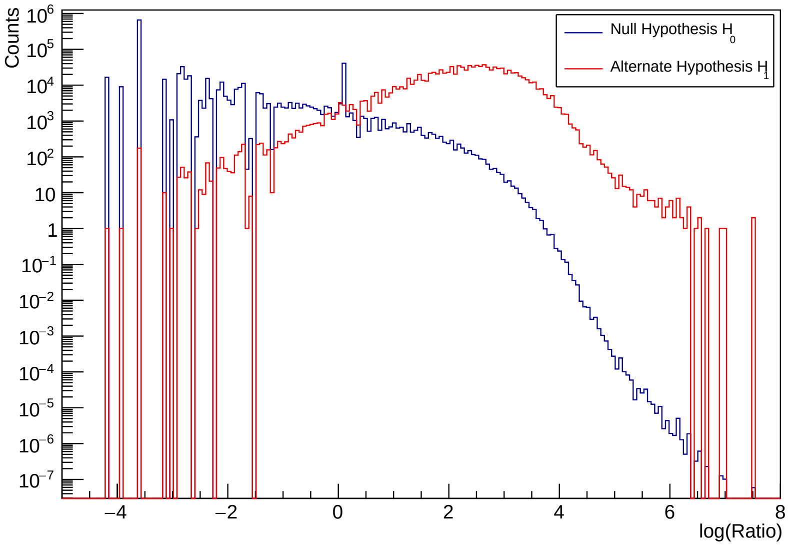

Figure 5.11: The ratio distributions where each trial with a calculated $\hat{\mathcal{L}}$ and $\hat{n}_s$ pair from the null and alternate hypothesis has been allocated a ratio value equal to the ratio of the frequency that the parameter pair was observed in the alternate hypothesis trials to the null hypothesis trials. Each trial from both the null and alternate hypothesis now has a ratio value.

The effect of the overflow bin, and how it has been mitigated, is shown in Figure 5.12. Here, the distribution of parameter pair ratios with the overflow bin from Figure 5.7 is plotted against the results with the use of importance sampling from Figure 5.11. In addition to these plots, the green distribution indicates where the trials that originally went in the overflow bin are put when they now have a corresponding ratio value. The green distribution shows that the overflow bin trials were distributed throughout likelihood value bins covered by both the hypothesis distributions when no importance sampling was applied. This clearly indicates that the trials which originally had no ratio value cannot be placed in the overflow bin located at a greater test statistic than any of the others obtained in the simulations. Additionally, this indicates that the power vs. significance plot in Figure 5.8 does not accurately portray the performance of the test statistic we have constructed here.

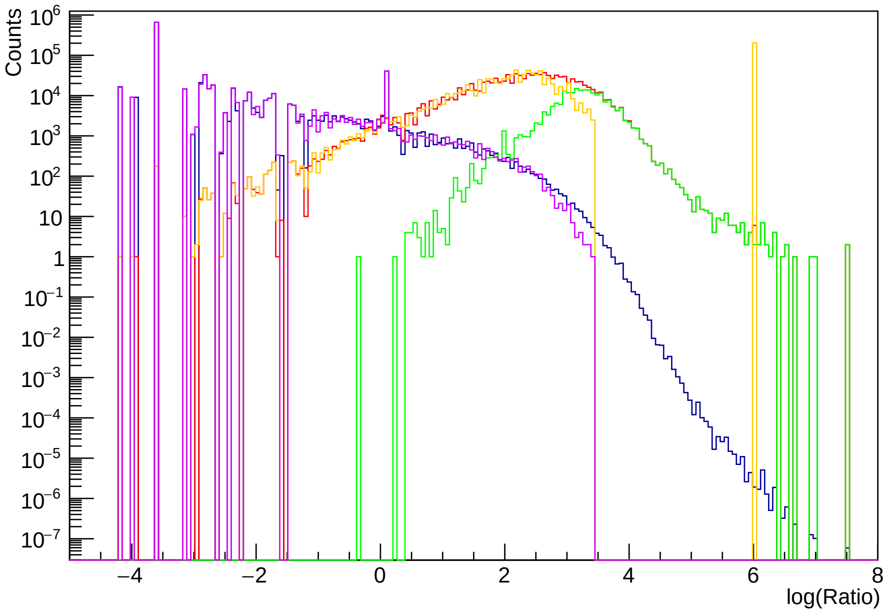

Figure 5.12: The null (red) and alternate (blue) hypothesis for parameter pair ratio distributions using importance sampling from Figure 5.11 against the null (purple) and alternate (orange) hypothesis for the parameter pair ratio distributions using an overflow bin from Figure 5.11. The green distribution indicates where the trials which were originally in the overflow bin were placed once importance sampling was used to provide them with ratio values. These overflow bin trials have been placed before and after the cut off of value covered by the original null hypothesis distribution without importance sampling.

Now, we create a power vs. significance plot using the null and alternate hypothesis parameter pair ratio value distributions after having used importance sampling to extend the null hypothesis distribution. As we compare this to the maximum likelihood $\hat{\mathcal{L}}$ power and significance, we must also construct the $\hat{\mathcal{L}}$ power vs. significance plot using the null hypothesis likelihood values with importance sampling weights applied. Additionally, the use of importance sampling now results in a much lower significance than -6 for the maximum likelihood, the value at which it ended in Figure 5.8. As a result, the power and significance for the test and null hypothesis due to the maximum likelihood $\hat{\mathcal{L}}$ and due to a ratio of the frequency of the parameter pairs is shown in Figure 5.13, with a closer look in Figure 5.14.

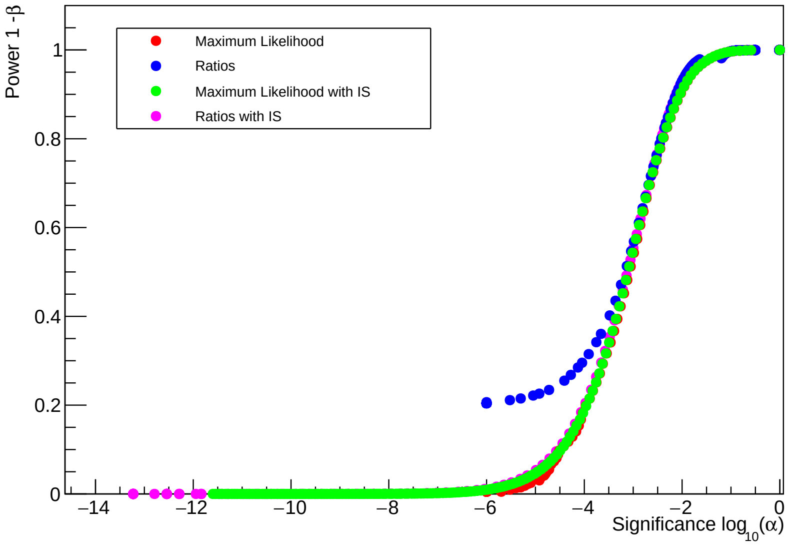

Figure 5.13: A power vs. significance plot for the distributions of test statistics for the null hypothesis and alternate hypothesis using the maximum likelihood $\hat{\mathcal{L}}$ (green) and using a ratio of the frequency of their parameter pairs this time using importance sampling (pink). Again, these plots are created by integrating over the null and alternate hypothesis distribution to values higher than a given test statistic. These are compared to the power vs. significance in Figure 5.8 due to $\hat{\mathcal{L}}$ (red) and the ratio values (blue), both without importance sampling.

The first primary result to observe from Figure 5.14 is that using importance sampling greatly increases both the power and significance levels that these test statistics can be tested at, compared to the power vs. significance plots produced in Figure 5.8. However, both Figure 5.13 and 5.14 also indicate that there is no overwhelming difference in the power and significance due to a test statistic constructed from the maximum likelihood compared to one constructed from a ratio of the frequency of parameter pairs. Figure 5.14 shows that the deviation between these points occurs below an approximate $\log_{10}\alpha$ value -6. This difference is not relatively large however, and could be some artefact of using the importance sampling method in one of the hypotheses and not the other. Explicitly,

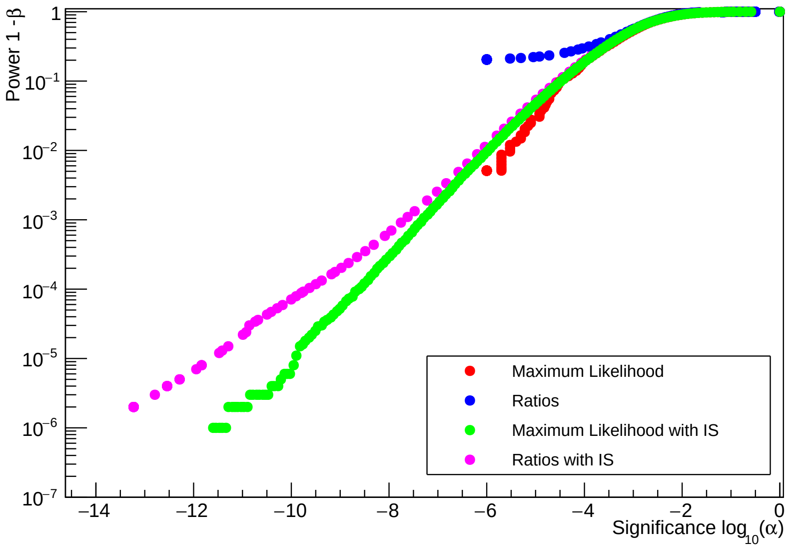

Figure 5.14: A closer look at the region of interest in Figure 5.13, where the power vs. significance plots separate. The plots notably move away from each other below a significance of -6. The power for the ratio test statistic here is slightly larger, although at this small significance this not a useful result.

not extending the alternate hypothesis distribution with importance sampling. Even if this difference is real, for this case at least, it only provides the slightest improvement. To clarify further, consider the significance at a $\log_{10}\alpha$ value of -10. At this level, the power only improves from 6×10^{-6} to 7×10^{-5}, which is only an improvement in that the relevant test statistic would occur in 0.0007% of trials if the alternate hypothesis were true, rather than in 0.00006% of trials. Altogether, this is not a particularly useful improvement on the significance. Additionally, this adds unnecessary and time-consuming steps to the procedure in order to obtain the power and significance for a given experimental test statistic. Either way, this phenomenon would have to be investigated further to obtain a definitive answer for the difference in these test statistics.

We can look at this effect in more detail by considering how the $\hat{n}_s$ values change for a set of fixed $\hat{\mathcal{L}}$ values in both the null and alternate distributions. This can be checked by taking "slices" along fixed $\hat{\mathcal{L}}$ bins of the two-dimensional $\hat{\mathcal{L}}$ and $\hat{n}_s$ plots for the null hypothesis, Figure 5.9, and the alternate hypothesis, 5.5. A set of these slices taken from Figures 5.5 and 5.9 is shown, normalised, in Figure 5.15.

Each of the slice distributions have been normalised so that the shape between the hypotheses can be compared. We see that the shapes of the distributions are similar, particularly for the distributions representing the slice of a lower likelihood value. If these shapes are the same, this is an indication that constructing a ratio test statistic, based off the parameter pair, will not be useful as it will not provide new information with which the

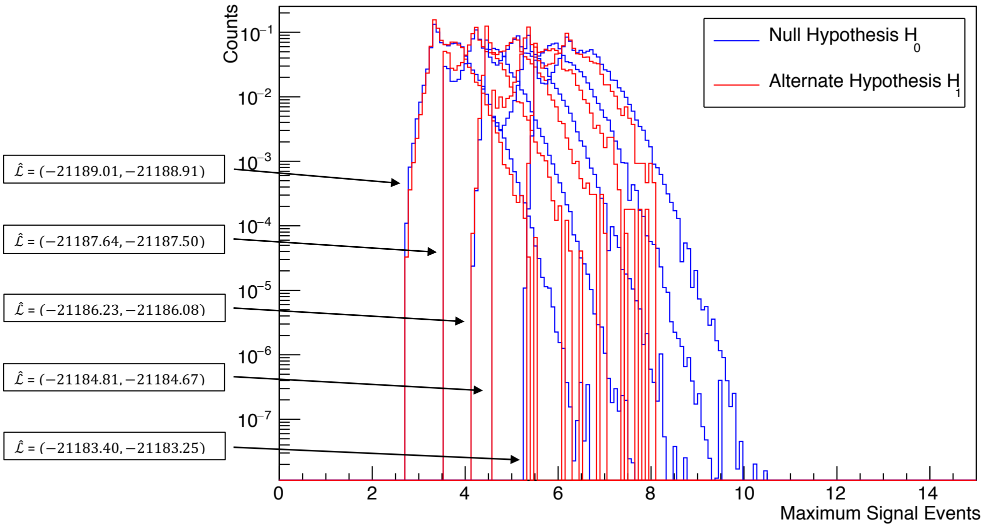

Figure 5.15: Assorted distributions showing how the maximum signal events $\hat{n}_s$ changes for a given discrete range of $\hat{\mathcal{L}}$ values. The range of $\hat{\mathcal{L}}$ for each distribution is stated in the respective boxes. The shape of these distributions is very similar for the lower $\hat{\mathcal{L}}$ values, while the shapes seem to become different between the hypothesis as $\hat{\mathcal{L}}$ increases.

hypotheses can be compared. At the lower $\hat{\mathcal{L}}$ values, there are numerous counts for the relative slice in both hypotheses. At the larger $\hat{\mathcal{L}}$ values, the alternate hypothesis lacks counts and hence has a notably different shape to the corresponding null hypothesis slice. In this case, a beneficial exercise may be to apply the importance sampling methods to the alternate hypothesis, such that we force rare clustering to occur and obtain the weights these parameters values in this hypothesis would be observed with.

Another way to look at the lack of difference between using $\hat{\mathcal{L}}$ and the $(\hat{\mathcal{L}}, \hat{n}_s)$ ratio value as the test statistic is by considering the correlation between these test statistics for each hypothesis. This is compared by creating a two-dimensional histogram for each hypothesis where a trial is placed according to its $\hat{\mathcal{L}}$ value and its $(\hat{\mathcal{L}}, \hat{n}_s)$ ratio value. The result of this process is shown for the null hypothesis in Figure 5.16 and for the alternate hypothesis in Figure 5.17. Note that there are millions of counts in Figure 5.16, as this includes all trials used to extend the $\hat{\mathcal{L}}$ distribution with importance sampling.

Both of these two-dimensional plots between the $\hat{\mathcal{L}}$ value and the ratio value for the null and alternate hypotheses indicate that there is a high level of correlation between the values. Furthermore, this correlation implies there is no benefit to constructing a test statistic according to the ratio of the frequency of parameter pairs, as the ratio value is just be representing its respective likelihood value. If the new test statistic constructed in this chapter is in fact no more powerful than $\hat{\mathcal{L}}$ on its own, this may just be further indication that the Neyman-Pearson Lemma is true.

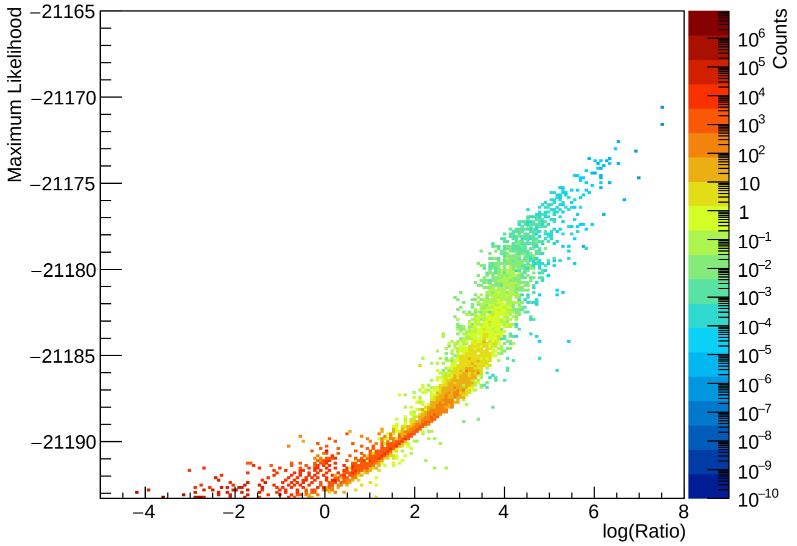

Figure 5.16: A two-dimensional plot for the null hypothesis where each count indicates a trial and its corresponding $\hat{\mathcal{L}}$ and ratio value. As the null hypothesis was extended with importance sampling, the histogram contains all trials used to extend the $\hat{\mathcal{L}}$ distribution. The resulting distribution seems to indicate there is a high degree of correlation between the $\hat{\mathcal{L}}$ and ratio values.

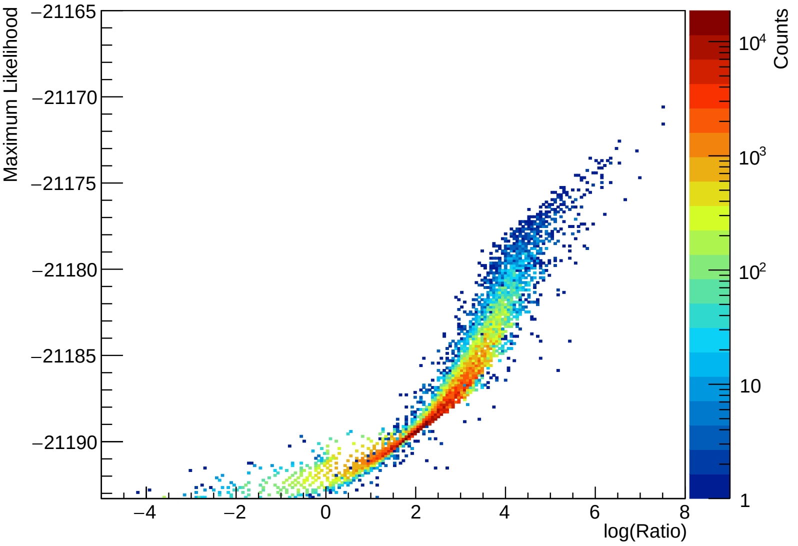

Figure 5.17: A two-dimensional plot for the alternate hypothesis where each count indicates a trial and its corresponding $\hat{\mathcal{L}}$ and ratio value. The resulting distribution also indicates a high level of correlation between the $\hat{\mathcal{L}}$ and ratio values.

5.3 Summary

Test statistics are an important part of hypothesis testing and are vital to obtain the significance of potential neutrino clustering observed with IceCube. Here, we have described test statistics and how they can be used in hypothesis testing, and in particular detailed the Neyman-Pearson Lemma, which says that using the maximum likelihood in likelihood analysis provides the most powerful test statistic. This theory has been implemented to construct our own test statistics to test a null hypothesis and an alternate hypothesis of simulated neutrino events on some space. We also utilised the ratio of frequency of best fit parameter combinations from both hypotheses to approach this problem differently, which required the assistance of the importance sampling method to generate the rare backgrounds. We were unable to find any significant difference between using the maximum likelihood as a test statistic compared to using one constructed by taking a ratio of frequency of best fit parameter pairs of $\hat{\mathcal{L}}$ and $\hat{n}_s$. Further investigation showed that there is a high correlation between the ratio values and the $\hat{\mathcal{L}}$ values. We also found that the distributions of $\hat{n}_s$ values for a fixed $\hat{\mathcal{L}}$ look very similar in both the null and alternate hypothesis. In conclusion, the ratio test statistic seems to just replicate the use of the $\hat{\mathcal{L}}$ test statistic, and as such indicates that as the Neyman-Pearson Lemma suggests, $\hat{\mathcal{L}}$ is the most powerful test statistic.

6 Application of Importance Sampling for Rare Cluster Generation in a Full Sky Likelihood Analysis

The ideal application for the importance sampling methods detailed in Chapter 4 is to use them in the real likelihood analyses performed within the IceCube Collaboration. In this chapter we show how we can translate the Gaussian and binomial sampling methods to force clusters of simulated neutrino events to occur on actual skies that resemble the full sky seen by IceCube. Again, this is demonstrated by extending the distribution of maximum likelihood values and their corresponding weights for clustering of neutrinos which would otherwise rarely occur if a background only hypothesis is true, and the clusters and are naturally sampled.

6.1 Likelihood Fitting on the Full Sky

To emphasise the usefulness of the importance sampling methods for likelihood analyses, we need to demonstrate how they work with the actual sky observed by IceCube. This means that the space we are considering is one that resembles the Universe around us, and any events on it have positions described in right ascension and declination coordinates. It is important to note that likelihood analyses in IceCube are also often maximised according to the energy of the neutrino, however here we will only focus on how the importance sampling methods are used to manipulate the spatial positions of events to create rare background clusterings. The IceCube Collaboration uses both binned [69] and unbinned [70] likelihood analyses to search for astrophysical neutrino sources. Unbinned likelihood analyses follow the same process already used to obtain the best fit likelihood values in this thesis. Specifically, we find the best fit likelihood by considering the continuous point spread function due to each neutrino event on the space. By maximising the likelihood using these point spread functions, the neutrino information incorporated into the likelihood functions represents each neutrino as accurately as possible, hence being unbinned. Binned likelihood analyses on the other hand involve fitting the likelihood with the neutrino information binned. For example, the maximisation of the likelihood at each point on the space may on the amount of neutrinos which have a significant contribution at that point of space. A histogram of information obtained using a binned likelihood analyses approaches the same result as an unbinned likelihood analysis as the bin size is reduced. As such, using an unbinned likelihood analysis produces a much more accurate result.

Again, the previously used logarithmic likelihood function, Equation 4.4, applies for the full sky likelihood analysis, albeit with some slight changes.

$$-\log\mathcal{L}(n_s\,|\,\vec{x}_s) = -\sum_{i=1}^{N}\log\left[\frac{n_s}{N}S_i(\vec{x}_s) + (1-\frac{n_s}{N})B(\vec{x}_s)\right] \tag{6.1}$$

The first difference is that we now consider our background probability density, $B(\vec{x}_s)$, as one that describes the entire 4π steradians of the sky that IceCube observes. In previous chapters, that background probability density depended on the Cartesian grid under consideration. For our full sky case, we use:

$$B(\vec{x}_s) = \frac{1}{4\pi} \tag{6.2}$$

Another difference is that when we consider the spatial coordinates of an event, or any position $\vec{x}_s$ on the sky, they are now in celestial coordinates instead of Cartesian coordinates which were used previously. Celestial coordinates consist of right ascension, α, and declination, δ. Right ascension corresponds to the longitude of the Earth, and runs between 0 and 2π radians or 0 and 360°. Declination corresponds to the latitude of the Earth. It begins at the zenith of the North Pole as $\frac{\pi}{2}$ radians or 90°, and goes to $-\frac{\pi}{2}$ radians or -90° at the zenith of the South Pole.

Following the same procedure of Chapter 4, we aim to create randomly generated skies of neutrino events and find how often the background events naturally form neutrino clusters. Previously, a likelihood fit was performed at a singular point on the sky to obtain the best fit parameters. This time, to replicate a typical process performed by IceCube, the likelihood fit is performed over the entire sky. This means that the likelihood fit on any given randomly generated sky will locate the position with the greatest likelihood value, which is then the "hotspot" on the sky, indicating the most likely position where neutrino emission occurs along the line of sight to these coordinates. Creating numerous random skies and performing this process then gives us an indication for how often clustering of neutrino events on the real sky occurs due to random chance. To randomly distribute such events on a sky, we must sample their positions from an evenly distributed region. This can be achieved by sampling right ascension, α, and declination, δ, according to:

$$\alpha = 2\pi\eta_1 \tag{6.3}$$

$$\delta = \cos^{-1}[2\eta_2 - 1] - \frac{\pi}{2} \tag{6.4}$$

where η_1 and η_2 are uniformly distributed random numbers between 0 and 1. The signal probability density function, $S_i(\vec{x}_s)$, for each i^{th} event is still described by:

$$S_i(\vec{x}_s, \vec{x}_i, \sigma_i) = \frac{1}{2\pi\sigma_i^2}\exp\left[\frac{-|\vec{x}_s - \vec{x}_i|^2}{2\sigma_i^2}\right] \tag{6.5}$$

The error resolution, σ_i, must be in the same units that are used for the coordinates. The angular separation in radians between $\vec{x}_s$ and $\vec{x}_i$, with coordinates (α_s, δ_s) and (α_i, δ_i) respectively, can be obtained with:

$$|\vec{x}_s - \vec{x}_i| = \cos^{-1}[\sin(\delta_s)\sin(\delta_i) + \cos(\delta_s)\cos(\delta_i)\cos(\alpha_s - \alpha_i)] \tag{6.6}$$

where $\alpha \in [0, 2\pi]$ and $\delta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. All of this information can now be used to create random neutrino skies. As we aim to obtain the hotspot on the sky (the largest likelihood value) we must scan across the sky and maximise the likelihood at uniformly separated positions across the space. The largest likelihood value will then correspond to such a hotspot. For the following demonstrations and examples, we will distribute 2000 events over the artificial sky. Each of these events have a resolution error of 0.01 radians, which is approximately 0.573°. An example of the likelihood fit performed uniformly across the sky for 2000 events is depicted in Figure 6.1.

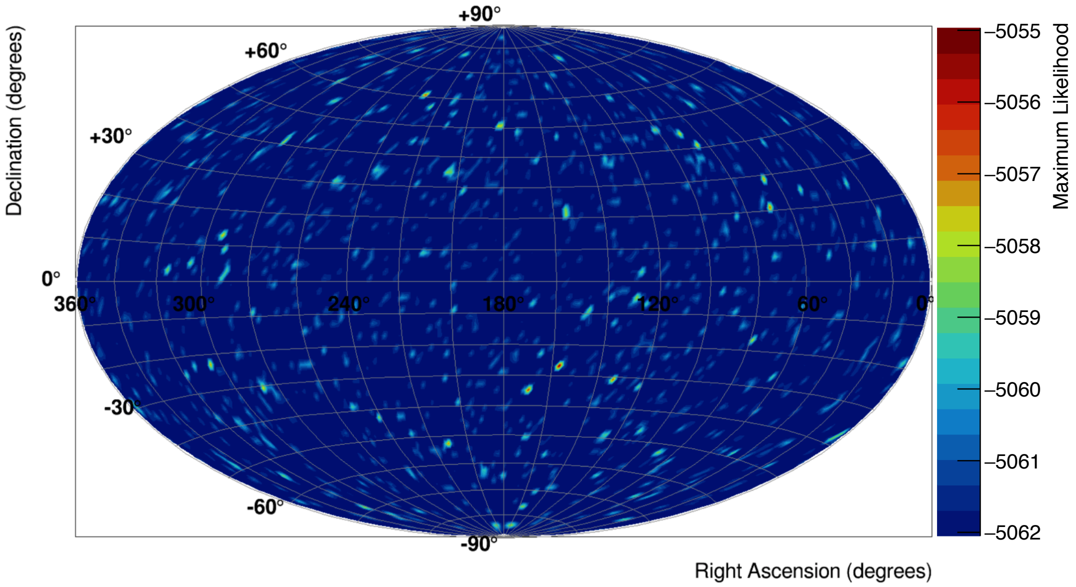

Figure 6.1: An example skymap showing the sky created for each trial, containing 2000 events each with an error resolution of 0.01 radians. The maximum likelihood $\hat{\mathcal{L}}$ is calculated at each one degree interval in both right ascension and declination. The hotspot, indicated by the absolute maximum $\hat{\mathcal{L}}$ value, occurs at (158°,-62°) with an approximate $\hat{\mathcal{L}}$ value of -5054.9.

In this case, the maximum likelihood is calculated at each 1 degree interval in both right ascension and declination, such that there is a total of 64800 likelihood maximisations that occur per sky, of which the greatest value signifies the hotspot on the sky. This is already a significant increase in the work required to obtain a test statistic for the sky, compared to one maximum likelihood fit per sky in Chapter 4 and 5. A distribution of the greatest $\hat{\mathcal{L}}$ value from 10^5 of these trials is shown in Figure 6.2, while a plot showing each greatest $\hat{\mathcal{L}}$ value and its corresponding $\hat{n}_s$ value is shown in Figure 6.3. Again, "counts" in these figures and the rest of the figures in Chapter 6 indicate the number of skies with the indicated best fit parameter values.

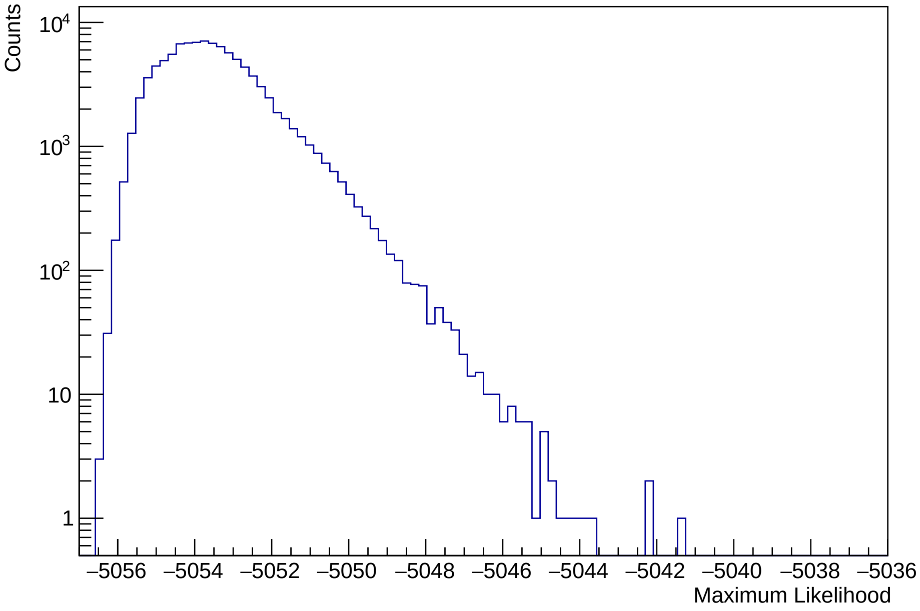

Figure 6.2: Distribution of the maximum $\hat{\mathcal{L}}$ values found from a likelihood fit over the full sky for 10^5 trials with 2000 randomly distributed events.

On inspection of Figure 6.2 and 6.3, it is clear that the best fit parameter observed most frequently is not the minimum possible parameter value, as it was for the simulations in Chapter 4 ($\hat{\mathcal{L}} \sim$ -21193.3, $\hat{n}_s = 0$). A most frequent $\hat{n}_s$ of greater than zero is expected as the minimum $\hat{n}_s$ value for a hotspot on the full sky is larger than the minimum $\hat{n}_s$ value observed when only calculating the likelihood at a fixed point. Thus, if the events were evenly distributed across the space, with no overlap of events, we know that the minimum possible $\hat{n}_s$ value in any of these trials is slightly less than one. This is because the likelihood function in Equation 6.1 will never fit to exactly one, even if only one event is being considered. This is due to the background probability density component always adding some contribution to the likelihood value. From Figure 6.3, the minimum $\hat{n}_s$ observed in the 10^5 trials is approximately 2.7, which indicates that in each of the trials there was some level of clustering occurring, with a peak of approximately 3.6 best fit signal events. The maximum $\hat{n}_s$ value observed is around 8.2, although trials with $\hat{n}_s$ values between this value and 6 are generally sparse and provide little information about how often clustering with these numbers of signal events would occur randomly. Specifically, the 10^5 trials performed here mean we can only consider any of the results to a significance of 10^{-5}. Desirably, we want to understand the rare clusters to a much lower significance than this, for which the increase in trials needed would require an exponential increase in computational time. Once again, to combat this dilemma, we

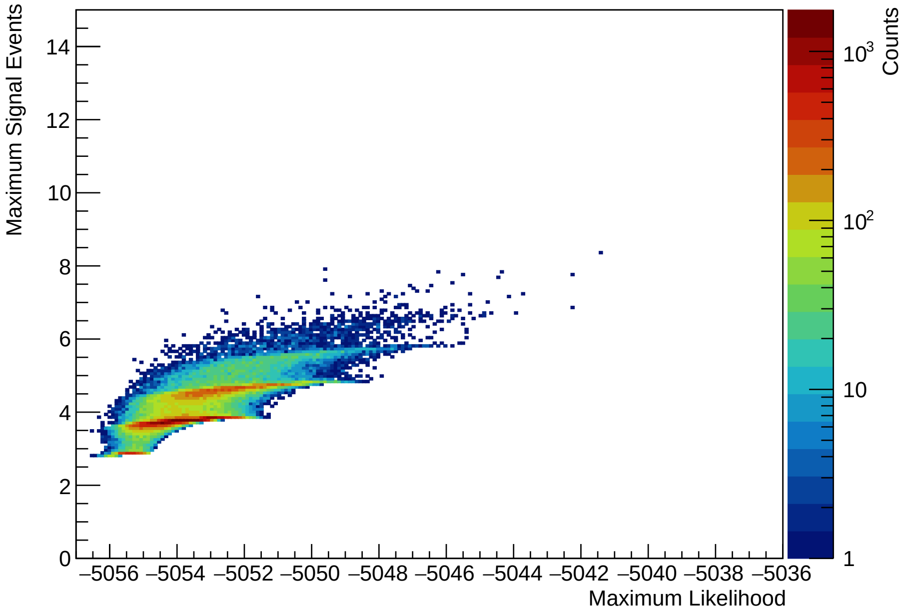

Figure 6.3: Distribution of the maximum $\hat{\mathcal{L}}$ values and their corresponding $\hat{n}_s$ found from a full sky likelihood fit for the 10^5 trials in Figure 6.2.

apply the previously constructed importance sampling methods to now observe biased clustering on the full sky likelihood analysis.

6.2 Application of Importance Sampling

6.2.1 Gaussian Weighting Method for Full Sky

To reiterate, the Gaussian weighting method relies on concentrating event positions around some point and allocating each event a weight. This weight value is a combination of the two possible sampling distributions that the position, in right ascension and declination, could be sampled from. The first distribution, $h(\alpha, \delta)$, describes the probability that an event has an equal chance of being placed anyway on the sky. As the total area of the sky is 4π steradians, this distribution is simply:

$$h(\alpha, \delta) = \frac{1}{4\pi} \tag{6.7}$$

The second distribution, $g(\alpha, \delta)$, concentrates events around some position (α_0, δ_0), for which a Gaussian function is used. The spread of the Gaussian, σ, can be altered to "tighten" the function, causing the events to be concentrated closer together. This distribution then takes the form:

$$g(\alpha,\delta) = \frac{1}{2\pi\sigma^2}\exp\left[-\left(\frac{(\alpha-\alpha_o)^2+(\delta-\delta_o)^2}{2\sigma^2}\right)\right] \tag{6.8}$$

As we did in Chapter 4, we require a combination of the flat and Gaussian distributions to create our new sampling distribution. If all event positions were selected according to the Gaussian function, they would all be artificially placed around the desired point. However, we only want a select amount of events around the desired position. Again, we also choose a biased number of events, n, which have their positions chosen according to $g(\alpha,\delta)$ by using the Box-Muller transform described in Appendix A. For each of the N total events, we choose a random number ϕ on (0,1). If $\phi < n/N$, the event position is sampled from $g(\alpha,\delta)$. Otherwise, if $\phi > n/N$, we sample from $h(\alpha,\delta)$, such that the right ascension and declination are found with Equation 6.4. This method produces some "flat" distribution of events across the sky, while there will be some bump (abundance of events) due to the Gaussian component. The weight of each i^{th} event is then obtained by dividing the natural probability density function by the biased probability density function:

$$w_i = \frac{h(\alpha_i,\delta_i)}{(1-\frac{n}{N})h(\alpha_i,\delta_i)+\frac{n}{N}g(\alpha_i,\delta_i)} \tag{6.9}$$

The total weight for the entire sky is then the multiplicative of each individual event's weight.

$$W = \prod_{i=1}^{N} w_i \tag{6.10}$$

Aside from the transformation of this process from Cartesian coordinates to celestial coordinates, when adapting this process from Chapter 4, we also change the position, (α_0,δ_0), that the Gaussian function, $g(\alpha,\delta)$, is centred on for each individual sky trial. This must be done as we are now performing a maximum likelihood search across the entire sky, and so we must construct the simulation such that the cluster of events could occur at any point on the space. Furthermore, the biased hotspot can then occur anywhere on the sky.

6.2.2 Binomial Weighting Method for Full Sky

The binomial weighting method must also be adjusted for celestial coordinates and a full sky likelihood analysis. However, this change is much simpler than that required for the Gaussian weighting method. Again, the binomial weighting method focuses on distributing some number of the total events, N, within some closed circular region, U, on the sky, which is just some portion of the total 4π steradians of the sky. The base probability, p, again describes the probability that if an event was placed randomly on the sky, it would be within U. This probability is then found by taking the ratio of the area of the closed region to the area of the entire sky:

$$p = \frac{U}{4\pi} \tag{6.11}$$

A bias probability, p', is then selected to determine how many of the N events are placed within U with a random position. An integer, k, is then sampled from a binomial distribution according to the chosen bias probability p', which has the form:

$$P(k \,|\, N, p') = \binom{N}{k} {p'}^{k} (1 - p')^{N-k} \tag{6.12}$$

where:

$$\binom{N}{k} = \frac{N!}{k!(N-k)!} \tag{6.13}$$

Here, the integer k, obtained by sampling this distribution, indicates the number of events which are in U. The weight, W, for this sky is then the ratio of the probability k was obtained from the probability mass function with N events and base probability p, to the probability k was obtained from the same probability mass function with N and bias probability p'.

$$W = \frac{P(k \,|\, N, p)}{P(k \,|\, N, p')} \tag{6.14}$$

Again, this method is now applied to a full sky likelihood analysis, and so the position at which U is centred at for each trial is sampled uniformly in $\cos(\delta)$. The effect of observing a biased cluster of events due to the binomial weighting method at any point on the sky can then be investigated. Once we have our number of events k, which are placed within U, we must randomly distribute them with this region. For each of the k events, we provide them with a random position on a spherical surface according to Equations 6.3 and 6.4, and if this position is within U, we allocate the event these coordinates. If the coordinates are not within U, we can simply sample a new position until it is in the closed region. The remaining $N - k$ background events then also have their coordinates chosen according to Equations 6.3 and 6.4. The position for each of these remaining events is accepted provided that they are not within U. Using all of this information, we can now force rare clusters of neutrinos to appear on the sky, according to both the binomial and Gaussian weighting methods.

6.2.3 Results

The Gaussian and binomial importance sampling weighting methods are used with the parameters described in Table 6.1 to demonstrate forced neutrino clustering in the full sky maximum likelihood analysis. Each of these simulations are performed with 10^5 trials. As both the Gaussian and binomial weighting methods involve choosing a position at which the biased neutrino events are clustered around, we perform a finer likelihood search around this biased position in addition to the full sky likelihood search. Once this position is identified, the finer search scans over approximately $\pm 5.73^{\circ}$ in both right ascension and declination of the hotspot position in 0.0573° intervals, and adds an extra 40000 likelihood maximisations per sky. We do this to obtain a more accurate value of the absolute maximum likelihood value, as due to the clustering, we expect the hotspot will occur around this position.

Table 6.1: Sampling parameters used to create Figures 6.4 and 6.5.

Weighting	Biased Events	Gaussian σ	Base Probability p	Bias Probability p'
Gaussian	4	0.01	-	-
Gaussian	4	0.005	-	-
Binomial	-	-	0.0005	0.001
Binomial	-	-	0.0005	0.0025
Binomial	-	-	0.0005	0.005

Figure 6.4 shows the absolute maximum likelihood, $\hat{\mathcal{L}}$, distributions acquired from each trial in each simulation. Each of the bins for a discrete range of likelihood values from each simulation is combined as an average weighted mean (Equation 4.13), with the results of this combination shown in Figure 6.5. More trials and a greater assortment of sampling parameters are desirable to construct each of these plots, however the number of likelihood maximisations required per sky impose a severe limit on the number of trials which can be performed within some computational time.

It is immediately obvious that the shape of the distributions in Figure 6.4 show some unusual behaviour. In particular, the distributions from each of the simulations seem to have their counts reduced by several orders of magnitudes around an approximate $\hat{\mathcal{L}}$ of -5047. This is clarified by comparing the average weighted mean of these distributions to the original distribution from Figure 6.2 where no importance sampling has been used, which is shown in Figure 6.5. We see here that the shape of the $\hat{\mathcal{L}}$ distribution creates a significantly different distribution to what we expect to obtain, that is, the distribution seen in Figure 6.2. This behaviour of the one-dimensional distributions in Figure 6.4 implies there is some other unaccounted for factor impacting these biasing methods. Upon closer inspection of the data which describes how the neutrino events are clustered, it turns out that this method does possibly have a flaw. Specifically, we have constructed these simulations such that we expect the absolute maximum likelihood to occur in the immediate vicinity of the position where the biased events are clustered. What is actually occurring in many of trials is that the full sky likelihood search identifies an absolute maximum likelihood value correlated to a hotspot position not near where the neutrino events are forced to cluster. To be explicit, the background neutrino events are randomly clustering together elsewhere on the sky and creating the hotspot. Note that the counts of likelihood values in Figure 6.4 below a likelihood of -5050 have approximately the same value as the counts for the same likelihood in the natural background trial distribution. This region indicates where the full sky likelihood analysis identifies a hotspot not due to biased events. As these clusters form due to randomly placed background events on the sky, this region of the distribution replicates the background-only event distribution. The remaining distribution then predominantly describes likelihood values acquired due to

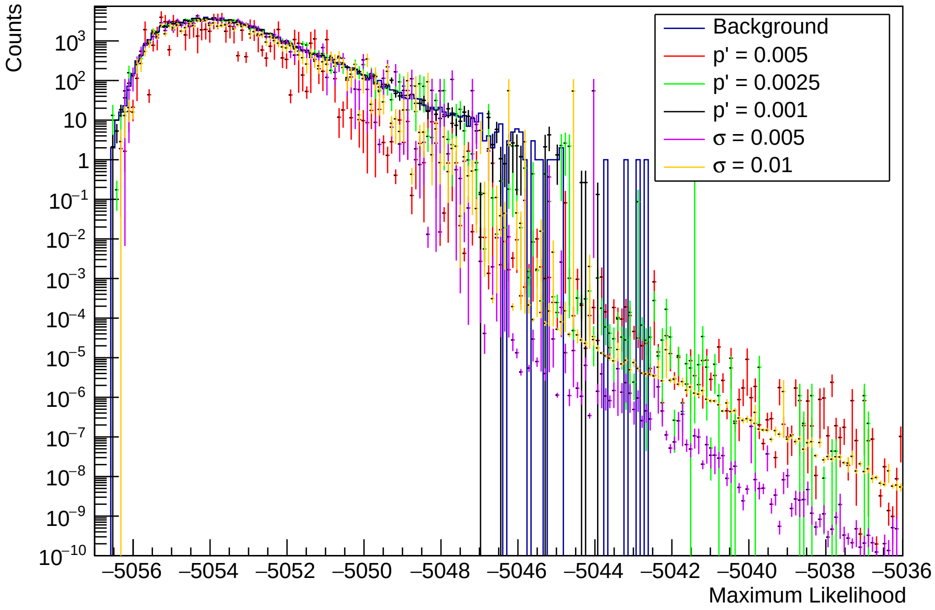

Figure 6.4: The distribution of $\hat{\mathcal{L}}$ values from each of the simulations performed with the sampling parameters described in Table 6.1 against the background-only simulation from Figure 6.2. Each simulation was performed with 10^5 trials. There is a notable change in slope of the distributions at an approximate $\hat{\mathcal{L}}$ of -5047.

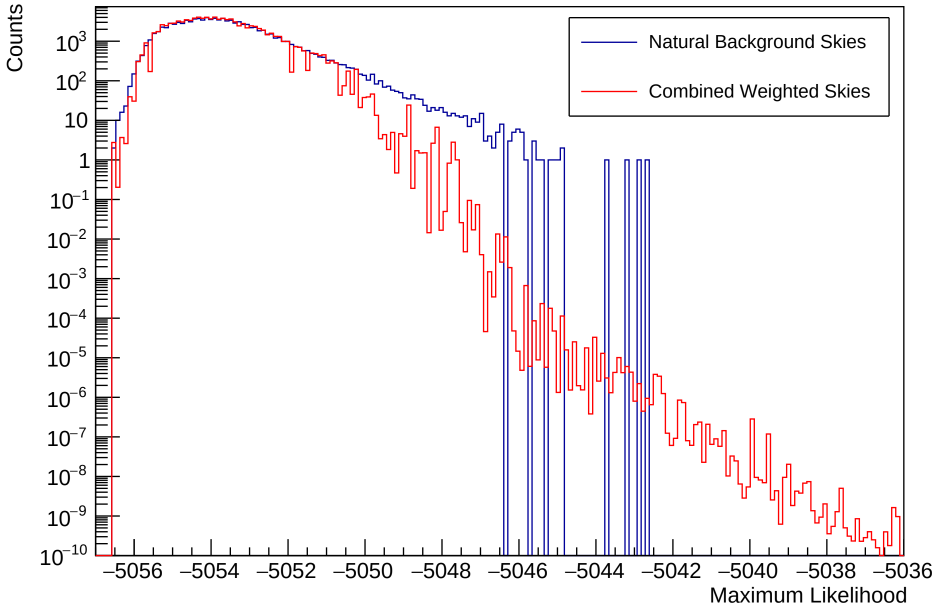

Figure 6.5: The resulting distribution due to creating an average weighted mean for discrete ranges of $\hat{\mathcal{L}}$ values from Figure 6.4. This is shown against the background-only simulation from Figure 6.2. Once again, there is a change in slope seen at an approximate $\hat{\mathcal{L}}$ of -5047, and the combination of the distributions notably does not follow the distribution seen due to the background-only trials.

the biased clusters. As such, this technique will produce two different distributions, one for clusters which are caused due to the biased events, and the other which is simply a background event clustering distribution.

To further investigate the difference between these regions of the distributions, we look at the regions separately. For this, we separate out the trials where we believe the absolute maximum $\hat{\mathcal{L}}$ value on the sky is caused due to the importance sampling biasing. To determine this limit, we calculate the angular separation between the position on the sky the biased events were forced to cluster about and the position which the likelihood search identifies as the hotspot. This calculation is made using Equation 6.6. We now determine the maximum distance from the biasing position such that if the hotspot occurred within this distance, it was due to the biased events. This distance is quite subjective, as it depends on the sampling parameters used for a given distribution, and how one determines whether a cluster was caused due to biased events. For example, a hotspot in a trial could occur near the edge of the region where the biased events are placed, and hence the largest $\hat{\mathcal{L}}$ may occur due to a combination of biased events and background events. Because of this, we set the maximum distance the hotspot can occur relative to the clustering position as approximately 3° or 0.05 radians. With this requirement set, we can look at the separate components of one of the distributions from Figure 6.4. Figure 6.6 shows the distribution of likelihood values for the binomial weighting simulation where $p' = 0.005$. The "on hotspot" distribution indicate the trials where the hotspot occurred within 3° of the point biased events are positioned about, while the "off hotspot" distribution consists of trials where the cluster occurs outside of this limit. These are plotted against the original background trials from Figure 6.2.

Figure 6.6 shows that the using importance sampling in the full sky likelihood does indeed create two separate likelihood distributions according to whether a hotspot occurs due to the biased events or not. The largest likelihood values in the "off hotspot" distribution for the hotspots away from the biasing point, have counts which would suggest they are due to biased events. This agrees with the previous statement that the angular separation chosen which determines whether the hotspot is due to the biasing of events or not is a subjective choice. Although the counts for the distribution of likelihoods for hotspots due to the biasing is orders of magnitude below the the natural background distribution, it does seem to have the same shape. This may be an indication that the sampling parameters we have used to create the importance sampling simulations have caused events to be too compactly clustered such that they are not identified in a full sky likelihood analysis. The sampling parameters here were chosen as their relative size on the full sky case is similar to those we used in Chapter 4, where the likelihood was calculated always at a fixed point. The scenario of clustering events according to sampling parameters on the same scale between a likelihood analysis at a fixed point and over a full sky may have been an incorrect step to take. As a result, our simulations have produced distributions where the biased clusters we are looking for have been severely under-sampled, and overall the counts for each of these likelihoods have been heavily reduced. Ideally, we would run these simulations with sampling parameters that make progressively small changes to how we would expect events to be placed naturally, and find whether there

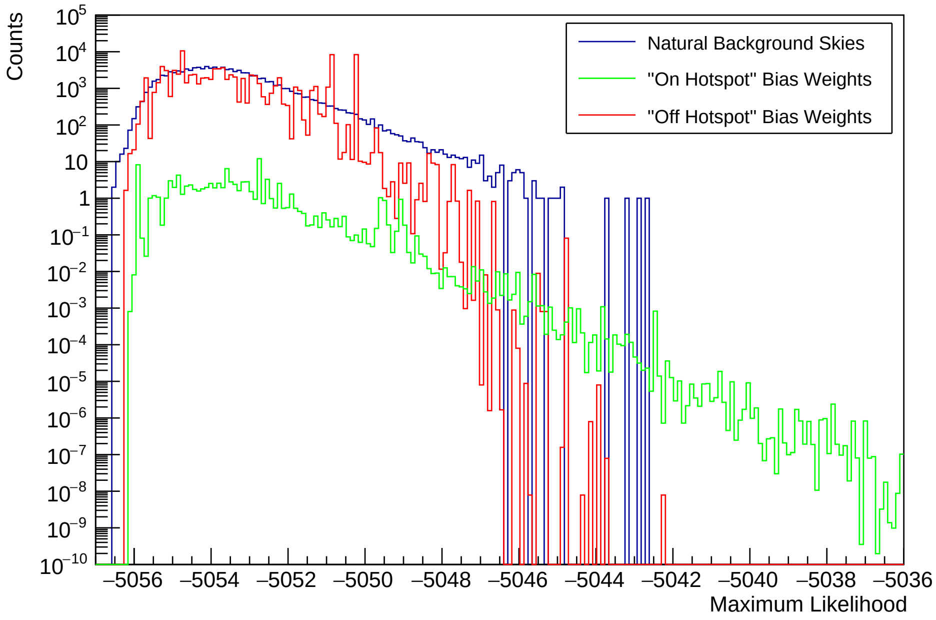

Figure 6.6: The distribution of trials from the binomial $p' = 0.005$ simulation where trials are considered to be "on hotspot" if the hotspot occurred within 3° of the biasing position (green), or "off hotspot" if the hotspot was greater than 3° away (red), both with weights applied. Also plotted is the natural background skies distribution from Figure 6.2 (blue), which at lower $\hat{\mathcal{L}}$ values has a similar shape to the "off hotspot" distribution.

is no split in the distribution between the regions for hotspots due to the biased events and those that are not. Unfortunately, time limitations of this project prevent this further investigation.

Although we have found that the distributions have not produced the expected results, we saw that the distribution in Figure 6.6 considering clusters that are due to the biased events follows a similar shape to the background trials distribution. This same behaviour also occurs in each of the other simulations when we separate the trials according to whether the hotspot occurs within 3° of the the biasing position. As such, we consider the trials which fit this criteria from each simulation. With this requirement set, we once again create a distribution of each of the $\hat{\mathcal{L}}$ values from each simulation in Table 6.1. As each simulation loses a different number of trials due to maximum distance of the hotspot restriction, the distribution for each simulation has been normalised so that they are comparable. Table 6.2 is a reduced version describing the sampling parameters used in each simulation from Table 6.1, which now also contains the number of trials for which the hotspot distance was within 3° of the biasing position. The individual normalised distribution of $\hat{\mathcal{L}}$ values from each simulation is shown in Figure 6.7 against a normalised version of the background-only events $\hat{\mathcal{L}}$ distribution from Figure 6.2. Each of these simulations are then combined as an average weighted mean, again with Equation 4.6. The resulting distribution is shown in Figure 6.8 against a normalised distribution of the background trials from Figure 6.2, as well as a normalised unweighted distribution of the trials which were not within 3° of the biasing position.

Table 6.2: Sampling parameters used to create Figures 6.7 and 6.8. The final column indicates how many trials produced a hotspot within the maximum allowed angular separation of 3° of the biasing position.

Weighting	Gaussian σ	Bias Probability p'	Number of Trials
Gaussian	0.01	-	56543
Gaussian	0.005	-	65653
Binomial	-	0.001	549
Binomial	-	0.0025	21270
Binomial	-	0.005	89424

The individual $\hat{\mathcal{L}}$ distributions in Figure 6.7 all now have a very similar slope at an approximate $\hat{\mathcal{L}}$ value of -5052 and higher. Although there is no expectation that these distributions should decrease at the same rate, it already indicates the combination of the simulations will produce a closer distribution to the background trials than the resultant distribution in Figure 6.5. The result of taking an average weighted mean of each of the simulations in Figure 6.7 returns the total distribution of $\hat{\mathcal{L}}$ values expected if the background-only simulation could be run for a much larger amount of trials than the trials ran here. This is shown in Figure 6.8. Although this distribution is much more similar in shape and slope to the original distribution, there is still a lot of inconsistencies between the bin values of the average weighted mean and the original trials. Assuming

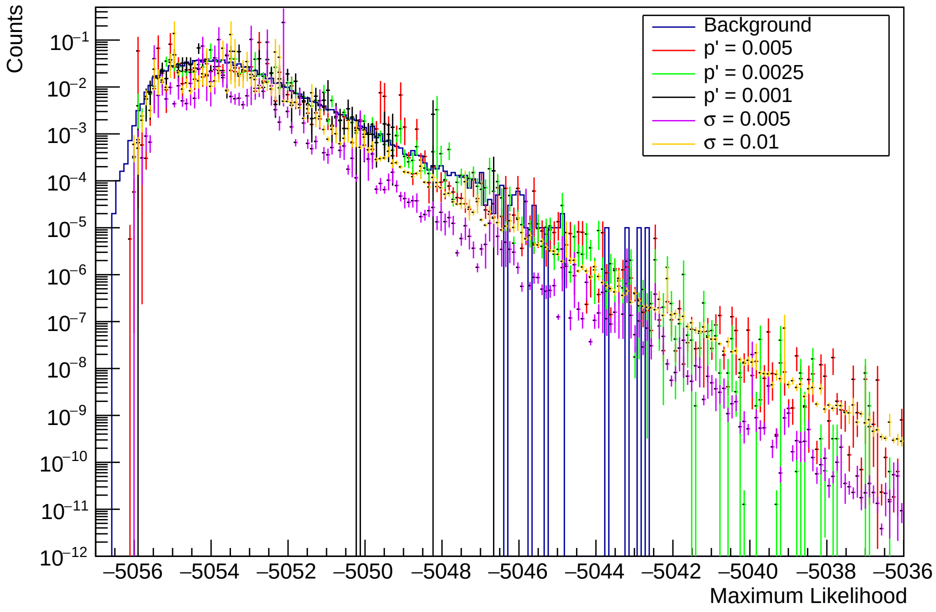

Figure 6.7: The normalised distribution of $\hat{\mathcal{L}}$ values from each of the simulations performed with sampling parameters described in Table 6.2 against the normalised background-only simulation from Figure 6.2. Each distribution seemingly follows a shape similar to that of the background-only simulation, however at different scales.

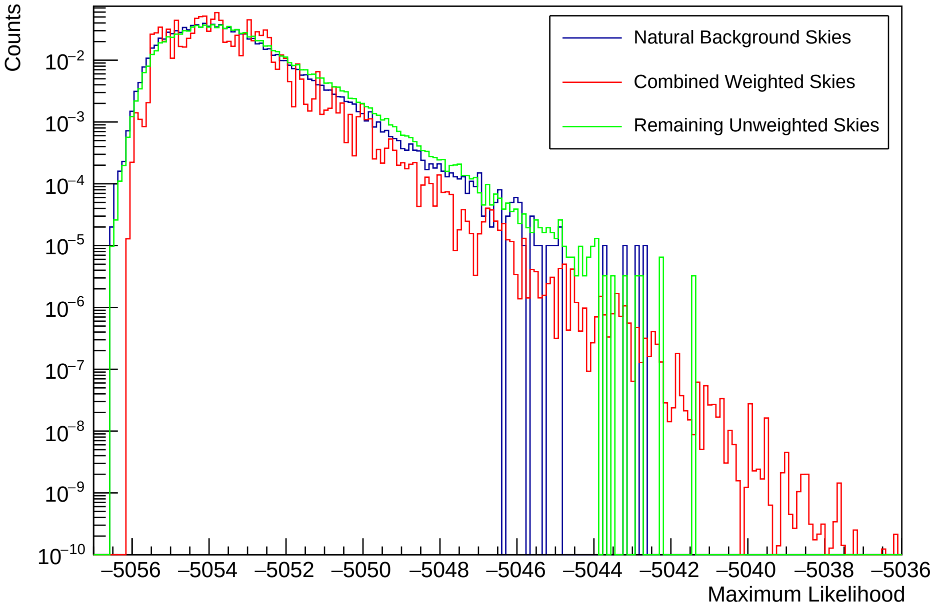

Figure 6.8: The resulting distribution due to making an average weighted mean for discrete ranges of $\hat{\mathcal{L}}$ values from Figure 6.7. This is shown against the normalised background-only simulation from Figure 6.2. The result of combining the simulation which fit the 3° angular separation criteria now looks a lot similar to the background only trials distributions, however there is still notable disagreement between the bin values of each distribution. Also plotted is the normalised distribution of the remaining trials which did not fit the 3° angular separation cut, which closely follow the natural background skies distributions.

that the selection criteria placed on the trials was the correct step to take, there is a notable lack of trials overall that are used to create the distribution, as described in Table 6.2. It is possible that if these simulations were performed for a greatly increased number of trials, and as such a lot more trials fit the criteria to be included, the resulting final distribution would “smooth out” out due to the increased amount of statistics. We can also observe that there are a number of bins around the peak which hold a lot of the total counts of the distribution. Again, if more trials were included, these peaks may be reduced with some of their contributing counts being distributed throughout the rest of the distribution, increasing the overall height of the rest of the distribution and perhaps fitting a lot closer to the original background distribution. Of course, it would also be ideal to perform multiple more simulations with different sampling parameters to those described in Table 6.1. Unfortunately, due to the time taken to run these simulations, this was not able to be investigated.

Another check we can consider is observing the distribution of the trials which did not fit the 3° angular separation criteria. These trials are shown without the weights obtained from importance sampling, and normalised for comparison, in Figure 6.8. We show them in this way to understand what likelihood values are obtained from hotspots not due to biased events. The unweighted distribution of the trials which indicate the hotspot occurring away from the biased events looks quite similar to the distribution produced by the original background-only events trials, except that it seems to be shifted slightly towards larger $\hat{\mathcal{L}}$ values. Specifically, when considering the peak of the natural background skies distribution at an approximate $\hat{\mathcal{L}}$ of -5054, the respective counts in the remaining unweighted skies distribution are smaller for $\hat{\mathcal{L}}$ less than -5054, but have larger counts values for greater $\hat{\mathcal{L}}$ values. We expect that the natural background and remaining unweighted skies distributions should be identical, as the remaining unweighted trials should essentially also just be identifying hotspots due to purely background event clustering. A potential reason for this dissimilarity is that some of the clusters which create the hotspot in these trials are sometimes still being created due to the influence of biased events. This may imply that the 3° angular separation cut on where we consider clusters to be due to the biased events or only background events may not be appropriate, overall increasing the $\hat{\mathcal{L}}$ values obtained in these trials. Another issue may be due to the region around the biasing position we perform a finer likelihood search over, in addition to the full sky likelihood analysis. As there is a difference in the size of this region and that where we consider our cut on whether clusters are formed due to biased events, the hotspot which forms above this cut but due to background only events may occur in the finer search region. In this case, the $\hat{\mathcal{L}}$ values obtained due to background events can on average have a greater $\hat{\mathcal{L}}$ value than the original background only events trials which did not have an as fine incremental likelihood search. To investigate this further, we would have to recreate the background sky trials with a likelihood search with searches at the same detail across the entire sky. Again, due to computational limits, this check is unable to be performed within the constraints of this investigation. Another likely reason for an increase in the average $\hat{\mathcal{L}}$ of the remaining unweighted skies distribution compared to the natural background skies $\hat{\mathcal{L}}$ values is because the remaining unweighted sky trials represent hotspots which occur away from the biasing position. We expect that as we bias events close to each other, there should be some clustering with a larger likelihood associated with this orientation, although a hotspot due to the remaining background events can occur with a larger likelihood, producing the maximum likelihood $\hat{\mathcal{L}}$. Overall, the remaining unweighted skies distribution then has a greater average $\hat{\mathcal{L}}$ as the trials which are in it relate only to trials which had a $\hat{\mathcal{L}}$ for background hotspots large enough to exclude the biased only cluster distribution. Altogether, the results do seem to indicate that the importance sampling methods can be used to force rare clusters in the full sky likelihood analysis. However, further checks and simulations are required to ensure that correct weights due to importance sampling are being applied. Finally, it may be appropriate to trial the full sky likelihood analysis in Cartesian coordinates much like the simulations in Chapter 4. In this way, it would most likely be easier to identify any issues with using importance sampling in the full sky likelihood analysis.

6.3 Summary

The Gaussian and binomial weighting methods described in Chapter 4 have been adapted to force rare clusterings of neutrinos to occur on a full sky replicating the coordinate system that the IceCube Neutrino Observatory observes. Demonstrations of likelihood analyses over the full sky have then been performed to obtain the maximum likelihood value, $\hat{\mathcal{L}}$, which correspond to these rare neutrino clusters, for which appropriate weights are then applied. Obtaining this information can then extend the distributions of the $\hat{\mathcal{L}}$ values that would be seen if background-only events simulations were performed for a number of trials which would otherwise be computationally exhaustive. This investigation found that there are some unpredicted flaws to this current method, as a full sky likelihood analysis does not always identify the absolute maximum likelihood of a sky at position where the biasing of neutrino events occurs. An attempt to mitigate this effect has been trialled by only considering hotspots due to biased events. This identified rare clusters which would naturally occur with a 10−10 probability. However, additional improvements to the analysis have been suggested. Further consideration of this analysis and more statistics and computational resources are required to properly identify and resolve any current issues, which is unfortunately not possible due to the time restrictions of this project.

7 Conclusion

7.1 Results

In this work, we investigated two concepts closely related to the likelihood analysis performed for astrophysical neutrino point source analysis by the IceCube Collaboration. The first of these involved using importance sampling to extend the level of significance which experimental results can be compared at. Specifically, when a likelihood analysis identifies a cluster of neutrinos on the sky which replicates a neutrino source, we need to understand how often neutrinos would randomly be positioned together to mimic this same orientation. Typically to understand this, real neutrino data is scrambled over the sky multiple times to see how often it randomly clusters together. This process does not identify how often the rarest clusters of neutrinos occur naturally. To combat this, we have used importance sampling to force these rare clusters of neutrinos to happen on a sky. A sampling weight is then calculated for this sky, which indicates how often the rare cluster would occur naturally. In order to identify all types of clusters that would happen, we constructed two sampling methods for this: the Gaussian weighting method and the binomial weighting method. We have shown that these methods correctly force different types of neutrino clusters to occur on a coordinate system replicating that observed by IceCube. However, although likelihood values were obtained for rare clustering of events on the real sky coordinate system, the weights calculated for these values did not follow the expected likelihood value distribution when compared to our background only trials simulation. The cause of this issue may have been due to the sampling parameters chosen for the importance sampling simulations, which could have caused events to be too tightly clustered and as such were not identified in the full sky likelihood search.

The second major component of this work was investigating the usefulness of a new test statistic for hypothesis testing in point source analysis. It is known from the Neyman-Pearson Lemma that the maximum likelihood, $\hat{\mathcal{L}}$, is the most powerful test statistic in hypothesis testing. However, it has been previously suggested that a test statistic constructed as a combination of $\hat{\mathcal{L}}$ and the maximum signal events, $\hat{n}_s$, it maximises to could be even more powerful [68]. We investigated this by running multiple trials of null and alternate hypothesis, and constructed the test statistic as the ratio of the frequency that a $\hat{\mathcal{L}}$ and $\hat{n}_s$ combination occurs in the alternate hypothesis to that same pair occurring the null hypothesis. To fully understand this, we required the previously created importance sampling methods to extend the distribution of the null hypothesis $\hat{\mathcal{L}}$ and $\hat{n}_s$ values. By considering a power versus significance analysis between using $\hat{\mathcal{L}}$ on its own compared to the ratio value due to $\hat{\mathcal{L}}$ and $\hat{n}_s$, we found no improvement in the ratio test statistic. Through further investigation, we found that $\hat{\mathcal{L}}$ and $\hat{n}_s$ ratio values simply correlate to the $\hat{\mathcal{L}}$ values, and as such the ratio test statistic reproduces the power and significance of the $\hat{\mathcal{L}}$ test statistic.

7.2 Future Work

The obvious next step for the work is to completely understand why forcing neutrino clustering through importance sampling on the full sky did not return the expected maximum likelihood, $\hat{\mathcal{L}}$, distributions. Ideally, this issue would already be resolved, however this was not possible due to time limitations of this work. Once this issue is resolved, it would also be extremely beneficial to investigate further improvements to reduce the computational time of these simulations. This would be particularly useful as we would preferably be able to calculate the likelihood over an entire sky at a much finer scale than done in this work. This would also drastically increase the time taken for these simulations, so a solution to this may involve additional minimisation techniques in overall method.

Of course, if the previously listed issues are resolved, we would like to test the importance sampling methods on real neutrino data obtained with IceCube. By applying these methods on real data, we can compare the test statistics we obtain with importance sampling to those obtained with natural simulations of scrambling event data in other research. If there is an agreement between these two distributions of test statistics, this is an ideal test to check whether the importance sampling methods do indeed work with real neutrino data. This would further indicate that the importance sampling methods constructed in this work correctly identify the frequency of rare neutrino clusters which would be seen if natural simulations were run for long enough. As the typical likelihood analysis used in point source analysis incorporates the energy of the neutrino as well, it would also be useful to investigate whether importance sampling can be applied to the energy too. Overall, there is a lot more promising work to be done with the concept of importance sampling for rare cluster generation in point source analysis.

Appendices

A Box-Muller Transform

The Box-Muller transform returns a pair of independent, normally distributed random numbers, x and y [71]. This is done by obtaining a length, R, and an angle, θ, according to some uniformly distributed random numbers. The connection between these sets of numbers is shown in Figure A.1.

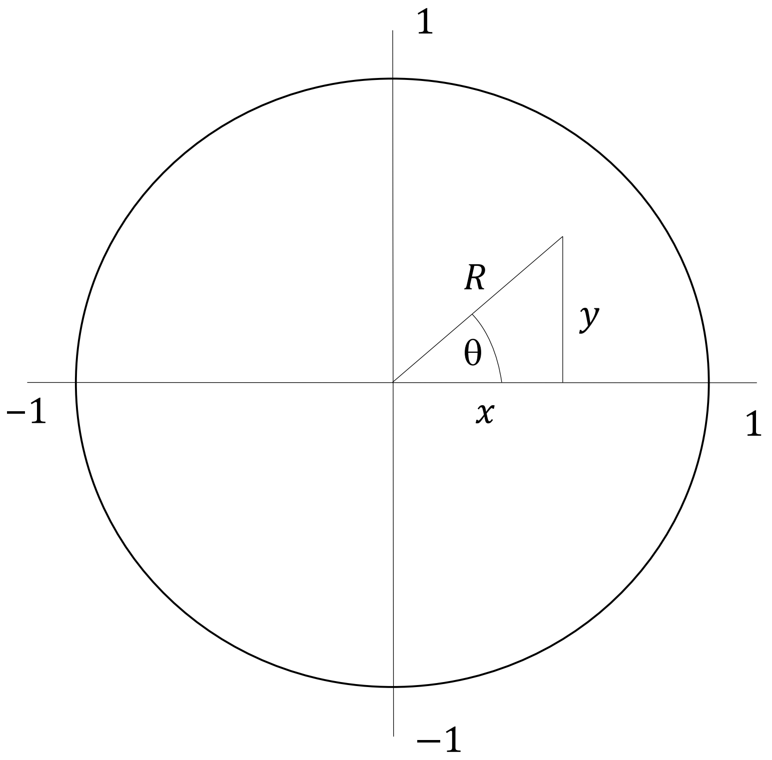

Figure A.1: Two normally distributed random numbers, x and y, which are obtained according to two uniformly distributed random numbers.

We obtain R and θ by taking a set of random numbers η_1 and η_2 on $(0, 1)$, such that:

$$R = \sqrt{-2\ln\eta_1} \tag{A.1}$$

$$\theta = 2\pi\eta_2 \tag{A.2}$$

The coordinates, (x, y), are then:

$$x = R \times \cos(\theta) \tag{A.3}$$
$$y = R \times \sin(\theta) \tag{A.4}$$

If we desire coordinates (x_1, y_1), such that they are distributed according to a Gaussian function which is centred at (x_μ, y_μ) and has a standard deviation σ, the coordinates obtained through the Box-Muller transform are then modified according to:

$$x_1 = (x \times \sigma) + x_\mu \tag{A.5}$$
$$y_1 = (y \times \sigma) + y_\mu \tag{A.6}$$

B Gaussian Weighting Plots

B.1 Gaussian $\hat{\mathcal{L}}$ and $\hat{n}_s$ Plots

The following graphs show the results used to create Figure 4.9. Each simulation contains 10^5 trials which biased 4 events of 2000 from a test source position at the origin, (0,0), on a 200×200 unit Cartesian grid. The sampling σ values used for each simulation are 500, 2, 1, and 0.5.

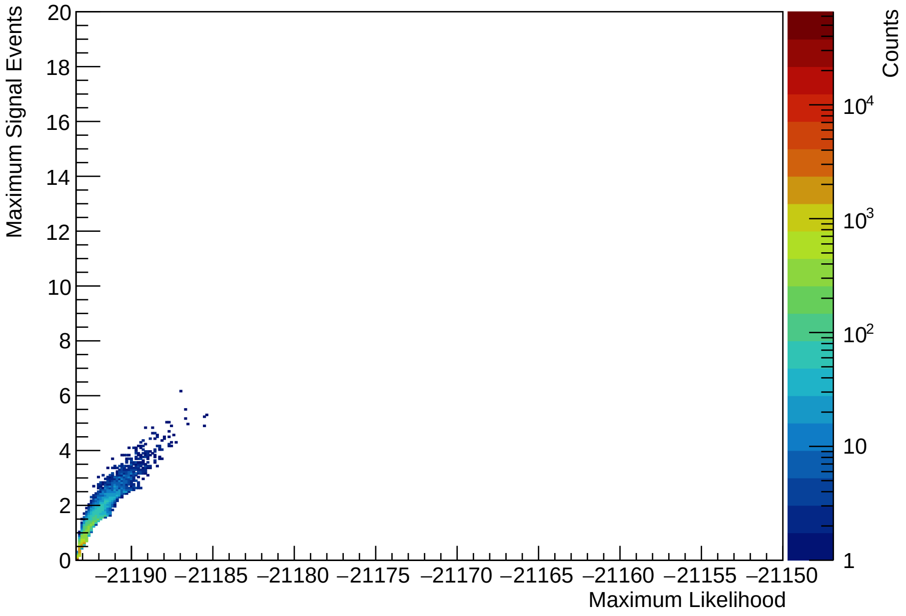

Figure B.1: The $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs with generated weights for the Gaussian weighting method with 4 biased events and a sampling σ of 500.

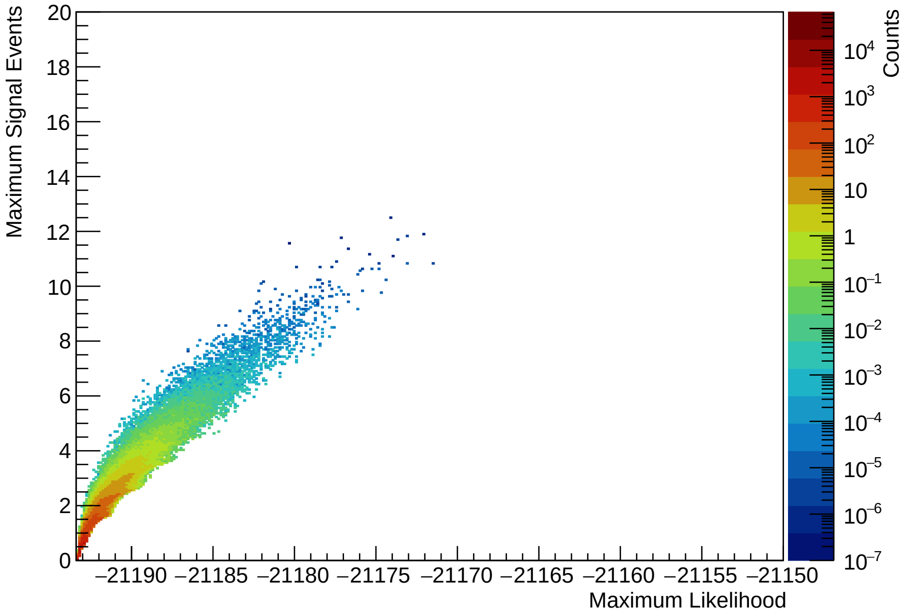

Figure B.2: The $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs with generated weights for the Gaussian weighting method with 4 biased events and a sampling σ of 2.

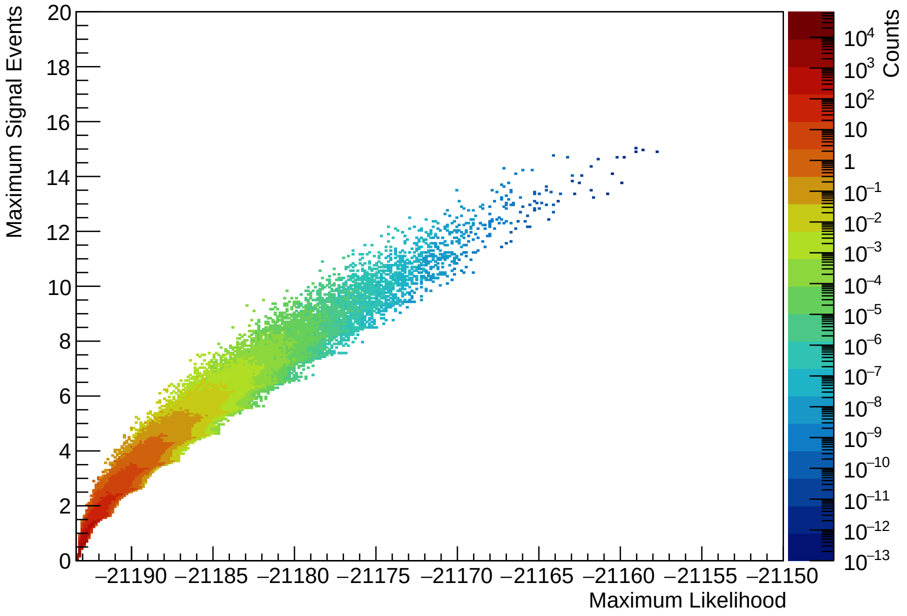

Figure B.3: The $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs with generated weights for the Gaussian weighting method with 4 biased events and a sampling σ of 1.

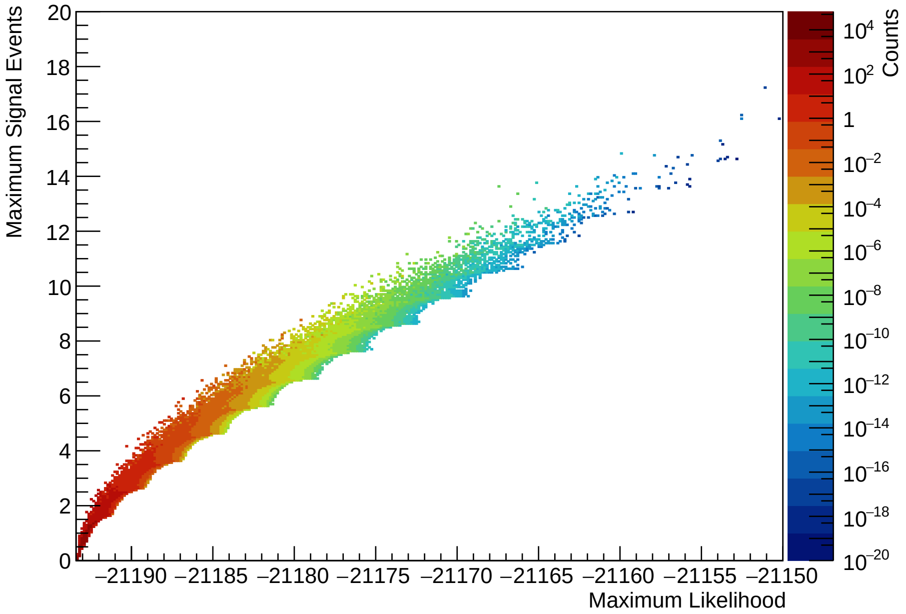

Figure B.4: The $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs with generated weights for the Gaussian weighting method with 4 biased events and a sampling σ of 0.5.

C Sampling Uniformly within a Circle

A pair of random numbers, x and y, sampled uniformly within a circle with radius R, can be found with the following procedure. The position is found by obtaining a length, r, and an angle, θ, according to some uniformly distributed random numbers. The connection between these sets of numbers is shown in Figure C.1.

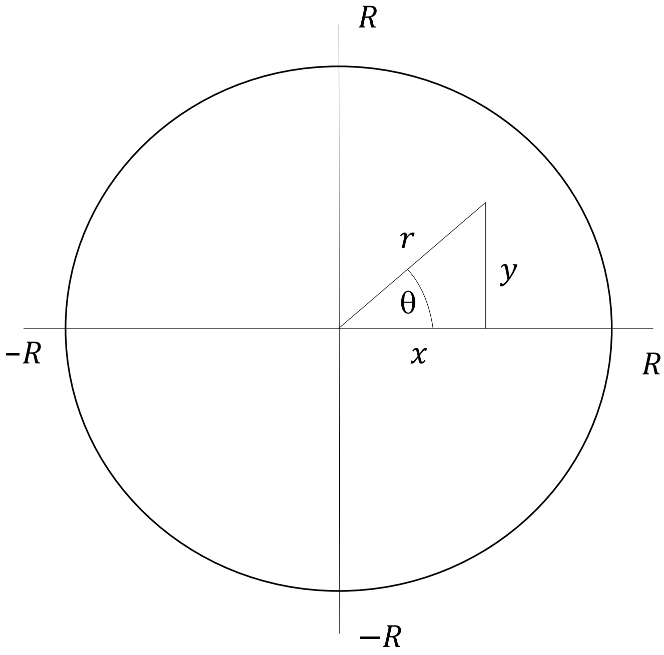

Figure C.1: Two random numbers, x and y, which are obtained according to two uniformly distributed random numbers.

We obtain r and θ by taking a set of random numbers η_1 and η_2 on $(0, 1)$, such that:

$$r = R \times \sqrt{\eta_1} \tag{C.1}$$
$$\theta = 2\pi\eta_2 \tag{C.2}$$

The coordinates, (x, y), are then:

$$x = r \times \cos(\theta) \tag{C.3}$$
$$y = r \times \sin(\theta) \tag{C.4}$$

D Binomial Weighting Plots

D.1 Binomial $\hat{\mathcal{L}}$ and $\hat{n}_s$ Plots

The following graphs show the results used to create Figure 4.14. Each simulation contains 10^5 trials using the binomial weighting method with a base probability p of 0.0005 that any of the 2000 events are randomly positioned around a test source position at the origin, (0,0), on a 200 × 200 unit Cartesian grid, within a radius of $\sqrt{20/\pi}$ units. The sampling bias probability values used for each simulation are 0.0005, 0.001, 0.0025, and 0.005.

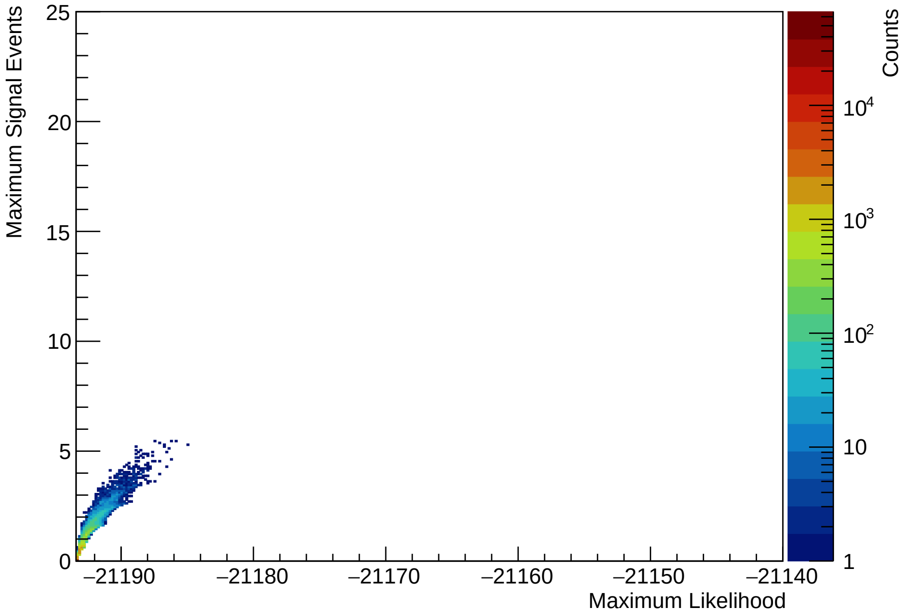

Figure D.1: The $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs with generated weights for the binomial weighting method with a sampling bias probability p' of 0.0005.

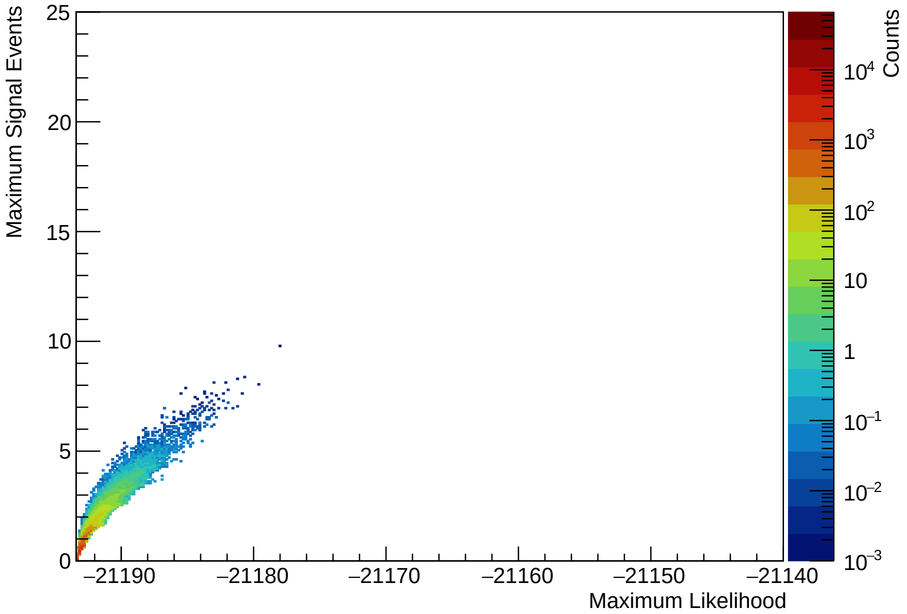

Figure D.2: The $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs with generated weights for the binomial weighting method with a sampling bias probability p' of 0.001.

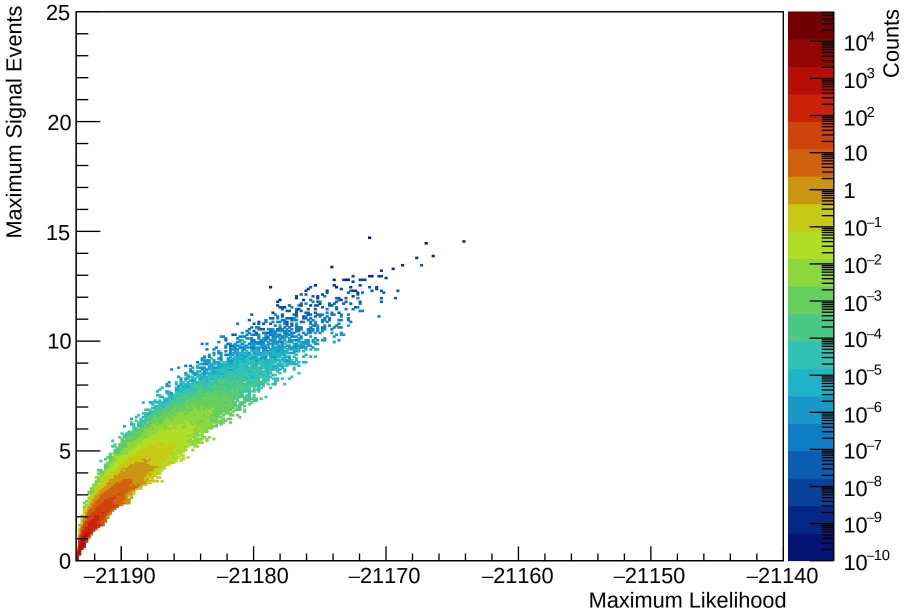

Figure D.3: The $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs with generated weights for the binomial weighting method with a sampling bias probability p' of 0.0025.

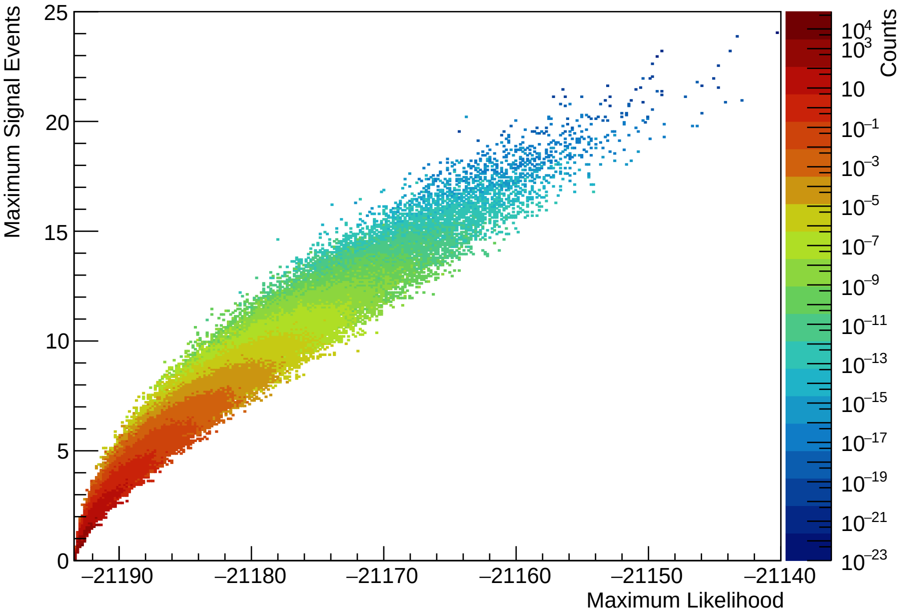

Figure D.4: The $\hat{\mathcal{L}}$ and $\hat{n}_s$ pairs with generated weights for the binomial weighting method with a sampling bias probability p' of 0.005.

Bibliography

[1] W. Pauli, *Open letter to the group of radioactive people at the gauverein meeting in tübingen.* 1930.

[2] E. Fermi, "Tentativo di una Teoria Dei Raggi β," *Il Nuovo Cimento*, vol. 11, no. 1, pp. 1–19, Jan. 1934.

[3] J. Chadwick, "The existence of a neutron," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 136, no. 830, pp. 692–708, 1932.

[4] F. Reines and C. L. Cowan, "A proposed experiment to detect the free neutrino," *Phys. Rev.*, vol. 90, pp. 492–493, 3 1953.

[5] C. L. Cowan, F. Reines, F. B. Harrison, H. W. Kruse, and A. D. McGuire, "Detection of the free neutrino: A confirmation," *Science*, vol. 124, no. 3212, pp. 103–104, 1956.

[6] G. Aad, T. Abajyan, B. Abbott, *et al.*, "Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc," *Physics Letters B*, vol. 716, no. 1, 1–29, 2012.

[7] L. N. Chang and C. Soo, "Standard model with gravity couplings," *Physical Review D*, vol. 53, no. 10, 5682–5691, 1996.

[8] C. Gross, A. Polosa, A. Strumia, A. Urbano, and W. Xue, "Dark matter in the standard model?" *Phys. Rev. D*, vol. 98, p. 063 005, 6 2018.

[9] J. A. S. Lima and D. Singleton, "Matter–antimatter asymmetry and other cosmological puzzles via running vacuum cosmologies," *International Journal of Modern Physics D*, vol. 27, no. 11, p. 1 843 016, 2018.

[10] J. A. Formaggio and G. P. Zeller, "From eV to EeV: Neutrino Cross Sections Across Energy Scales," *Rev. Mod. Phys.*, vol. 84, pp. 1307–1341, 2012.

[11] C. Giusti and M. V. Ivanov, "Neutral current neutrino-nucleus scattering: Theory," *Journal of Physics G: Nuclear and Particle Physics*, vol. 47, no. 2, p. 024 001, 2020.

[12] B. Pontecorvo, "Mesonium and anti-mesonium," *Sov. Phys. JETP*, vol. 6, p. 429, 1957, [Zh. Eksp. Teor. Fiz.33,549(1957)].

[13] Y. Fukuda, T. Hayakawa, E. Ichihara, *et al.*, "Evidence for oscillation of atmospheric neutrinos," *Physical Review Letters*, vol. 81, no. 8, 1562–1567, 1998.

[14] G. Fantini, A. Rosso, V. Zema, and F. Vissani, "Introduction to the formalism of neutrino oscillations: A tutorial for graduate students and young researchers," in. May 2018, pp. 37–119.

[15] D. Wands, O. F. Piattella, and L. Casarini, "Physics of the cosmic microwave background radiation," *Astrophysics and Space Science Proceedings*, 3–39, 2016.

[16] A. V. Ivanchik and V. Y. Yurchenko, "Relic neutrinos: Antineutrinos of primordial nucleosynthesis," *Physical Review D*, vol. 98, no. 8, 081301, p. 081 301, 2018.

[17] A. Ringwald, *How to detect big bang relic neutrinos?* XI International Workshop on Neutrino Telescopes, 2005.

[18] C. Grupen, *Astroparticle Physics*. Berlin: Springer, 2005, Translation of Astroteilchenphysik : Wiesbaden:Vieweg, 2000.

[19] A. Ianni, "Solar neutrinos and the solar model," *Physics of the Dark Universe*, vol. 4, pp. 44 –49, 2014, DARK TAUP2013.

[20] K. Hirata, T. Kajita, M. Koshiba, *et al.*, "Observation of a neutrino burst from the supernova sn1987a," *Phys. Rev. Lett.*, vol. 58, pp. 1490–1493, 14 1987.

[21] R. M. Bionta, G. Blewitt, C. B. Bratton, *et al.*, "Observation of a neutrino burst in coincidence with supernova 1987a in the large magellanic cloud," *Phys. Rev. Lett.*, vol. 58, pp. 1494–1496, 14 1987.

[22] E. N. Alexeyev, L. N. Alexeyeva, I. V. Krivosheina, and V. I. Volchenko, "Detection of the neutrino signal from SN 1987A in the LMC using the INR Baksan underground scintillation telescope," *Physics Letters B*, vol. 205, no. 2-3, pp. 209–214, 1988.

[23] A. Burrows and J. M. Lattimer, "Neutrinos from SN 1987A," *Astrophysical Journal Letters*, vol. 318, p. L63, Jul. 1987.

[24] K. Bays, T. Iida, K. Abe, *et al.*, "Supernova relic neutrino search at super-kamiokande," *Physical Review D*, vol. 85, no. 5, 2012.

[25] R. Aloisio, D. Boncioli, A. di Matteo, *et al.*, "Cosmogenic neutrinos and ultra-high energy cosmic ray models," *Journal of Cosmology and Astroparticle Physics*, vol. 2015, no. 10, 006–006, 2015.

[26] J. S. Heyl, R. Gill, and L. Hernquist, "Cosmic rays from pulsars and magnetars," *Monthly Notices of the Royal Astronomical Society: Letters*, 25–29, 2010.

[27] R. Walter and M. Balbo, "H carinae: Particle acceleration and multimessenger aspects," *Astronomische Nachrichten*, vol. 340, no. 4, 273–277, 2019.

[28] F. M. Rieger, V. Bosch-Ramon, and P. Duffy, *Fermi acceleration in astrophysical jets*. Dordrecht: Springer Netherlands, 2007, pp. 119–125.

[29] F. G. Schröder, *News from cosmic ray air showers (icrc 2019 – cosmic ray indirect rapport)*, 36th International Cosmic Ray Conference, 2019.

[30] J. n. Candia, S. Mollerach, and E. Roulet, "Cosmic ray spectrum and anisotropies from the knee to the second knee," *Journal of Cosmology and Astroparticle Physics*, vol. 2003, no. 05, 003–003, 2003.

[31] D. R. Bergman and J. W. Belz, "Cosmic rays: The second knee and beyond," *Journal of Physics G: Nuclear and Particle Physics*, vol. 34, no. 10, R359–R400, 2007.

[32] S. Thoudam, J. P. Rachen, A. van Vliet, *et al.*, "Cosmic-ray energy spectrum and composition up to the ankle: The case for a second galactic component," *Astronomy and Astrophysics*, vol. 595, A33, 2016.

[33] D. Biehl, D. Boncioli, A. Fedynitch, L. Morejon, and W. Winter, "Astrophysical neutrino production and impact of associated uncertainties in photo-hadronic interactions of uhecrs," *EPJ Web of Conferences*, vol. 208, B. Pattison, Y. Itow, T. Sako, and H. Menjo, Eds., p. 04 002, 2019.

[34] M. D. Domenico, M. Settimo, S. Riggi, and E. Bertin, "Reinterpreting the development of extensive air showers initiated by nuclei and photons," *Journal of Cosmology and Astroparticle Physics*, vol. 2013, no. 07, 050–050, 2013.

[35] A Haungs, J Blumer, B Fuchs, *et al.*, "Kcdc — the kascade cosmic-ray data centre," *Journal of Physics: Conference Series*, vol. 632, p. 012 011, 2015.

[36] A. Horvath, *Cherenkov*, `https://upload.wikimedia.org/wikipedia/commons/6/6b/Cherenkov.svg`, Accessed: 2019-10-2.

[37] C. W. Walter, "The super-kamiokande experiment," *Neutrino Oscillations*, 19–43, 2008.

[38] J. Boger *et al.*, "The Sudbury neutrino observatory," *Nucl. Instrum. Meth. A*, vol. 449, pp. 172–207, 2000.

[39] K BERNLOHR, "The optical system of the h.e.s.s. imaging atmospheric cherenkov telescopes. part i: Layout and components of the system," *Astroparticle Physics*, vol. 20, no. 2, 111–128, 2003.

[40] J. Aleksić, S. Ansoldi, L. Antonelli, *et al.*, "The major upgrade of the magic telescopes, part i: The hardware improvements and the commissioning of the system," *Astroparticle Physics*, vol. 72, 61–75, 2016.

[41] T. Weekes, H Badran, S. Biller, *et al.*, "Veritas: The very energetic radiation imaging telescope array system," *Astroparticle Physics*, vol. 17, pp. 221–243, May 2002.

[42] U. B. de Almeida, *Cherenkov telescope array science: Multi-wavelength and multi-messenger perspectives*, 36th International Cosmic Ray Conference, 2019.

[43] P Sommers and S Westerhoff, "Cosmic ray astronomy," *New Journal of Physics*, vol. 11, no. 5, p. 055 004, 2009.

[44] B. Keilhauer, M. Bohacova, M. Fraga, *et al.*, "Nitrogen fluorescence in air for observing extensive air showers," *EPJ Web of Conferences*, vol. 53, K.-H. Kampert, M. Fukushima, R. Engel, and B. Pattison, Eds., p. 01 010, 2013.

[45] The Pierre Auger Collaboration, "The pierre auger cosmic ray observatory," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 798, pp. 172 –213, 2015.

[46] M. Fukushima, "Telescope array project for extremely high energy cosmic rays," *Progress of Theoretical Physics Supplement - PROG THEOR PHYS SUPPL*, vol. 151, pp. 206–210, Jan. 2003.

[47] The IceCube Collaboration, *Collaborating organizations*, `https://icecube.wisc.edu/collaboration/institutions`, Accessed: 2019-10-28.

[48] F. Halzen, "Astroparticle physics with high energy neutrinos: From amanda to icecube," *The European Physical Journal C*, vol. 46, no. 3, 669–687, 2006.

[49] M. Aartsen, M. Ackermann, J. Adams, *et al.*, "The icecube neutrino observatory: Instrumentation and online systems," *Journal of Instrumentation*, vol. 12, no. 03, P03012–P03012, 2017.

[50] R. Abbasi, Y. Abdou, M. Ackermann, *et al.*, "Icetop: The surface component of icecube," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 700, 188–220, 2013.

[51] R. Abbasi, Y. Abdou, T. Abu-Zayyad, *et al.*, "The design and performance of icecube deepcore," *Astroparticle Physics*, vol. 35, no. 10, pp. 615 –624, 2012.

[52] A. Ishihara, *The icecube upgrade – design and science goals*, 36th International Cosmic Ray Conference, 2019.

[53] M. G. Aartsen, M. Ackermann, J. Adams, *et al.*, *Icecube-gen2: A vision for the future of neutrino astronomy in antarctica*, 2014.

[54] P. Coloma, P. A. Machado, I. Martinez-Soler, and I. M. Shoemaker, "Double-cascade events from new physics in icecube," *Physical Review Letters*, vol. 119, no. 20, 2017.

[55] The IceCube Collaboration, *The detection of neutrinos in icecube*, `https://masterclass.icecube.wisc.edu/en/learn/detecting-neutrinos`, Accessed: 2019-10-17.

[56] M. G. Aartsen, R. Abbasi, Y. Abdou, *et al.*, "First observation of pev-energy neutrinos with icecube," *Phys. Rev. Lett.*, vol. 111, p. 021 103, 2 2013.

[57] M. G. Aartsen, R. Abbasi, Y. Abdou, *et al.*, "Evidence for high-energy extraterrestrial neutrinos at the icecube detector," *Science*, vol. 342, no. 6161, 2013.

[58] R. Enberg, M. H. Reno, and I. Sarcevic, "Prompt neutrino fluxes from atmospheric charm," *Phys. Rev. D*, vol. 78, p. 043 005, 4 2008.

[59] M. Aartsen, K. Abraham, M. Ackermann, *et al.*, "Evidence for astrophysical muon neutrinos from the northern sky with icecube," *Physical Review Letters*, vol. 115, no. 8, 2015.

[60] J. Stettner, *Measurement of the diffuse astrophysical muon-neutrino spectrum with ten years of icecube data*, 36th International Cosmic Ray Conference, 2019.

[61] M. Aartsen, M. Ackermann, J. Adams, *et al.*, "Multimessenger observations of a flaring blazar coincident with high-energy neutrino icecube-170922a," *Science*, vol. 361, no. 6398, 2018.

[62] M. Aartsen, M. Ackermann, J. Adams, *et al.*, "Neutrino emission from the direction of the blazar txs 0506+056 prior to the icecube-170922a alert," *Science*, vol. 361, no. 6398, pp. 147–151, 2018.

[63] T. Severini, *Likelihood Methods in Statistics*, English. Oxford University Press, 2000.

[64] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *Ann. Math. Statist.*, vol. 9, no. 1, pp. 60–62, Mar. 1938.

[65] M. G. Aartsen, K. Abraham, M. Ackermann, *et al.*, "All-sky search for time-integrated neutrino emission from astrophysical sources with 7 yr of icecube data," *The Astrophysical Journal*, vol. 835, no. 2, p. 151, 2017.

[66] F. James and M. Roos, "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations," *Comput. Phys. Commun.*, vol. 10, pp. 343–367, 1975.

[67] J. Neyman and E. S. Pearson, “On the Problem of the Most Efficient Tests of Statistical Hypotheses,” *Philosophical Transactions of the Royal Society of London Series A*, vol. 231, pp. 289–337, 1933.

[68] G. C. Hill, “The muon problem: Finding cascade only point sources in a sky with background track events,” IceCube Spring Collaboration Meeting 2015, 2015.

[69] J. B. Alba, “Search for neutrino point sources with icecube 22-strings,” *Nuclear Physics B - Proceedings Supplements*, vol. 188, pp. 267 –269, 2009, Proceedings of the Neutrino Oscillation Workshop.

[70] J. Braun, J. Dumm, F. Palma, *et al.*, “Methods for point source analysis in high energy neutrino telescopes,” *Astroparticle Physics*, vol. 29, pp. 299–305, May 2008.

[71] G. E. P. Box and M. E. Muller, “A note on the generation of random normal deviates,” *Ann. Math. Statist.*, vol. 29, no. 2, pp. 610–611, Jun. 1958.