

Development of the Online Filter for Selection of Cascade-like Events in IceCube

K. Husband^{1,2} and Supervisor: E. Middell¹

¹*DESY, Plantanallee 6, D-15738 Zeuthen, Germany*

²*Magdalene College, Cambridge University, CB3 0AG, UK*

(Dated: September 3, 2010)

Improvements in the IceCube online filter are investigated with the aim of improving the signal passing efficiency and reducing the CPU needed. Firstly it was found that the LineFit velocity cannot be replaced by the time variance. Secondly two new cut variables were found from the cumulative charge distribution on an event-wise basis and the charge-energy ratio decay over distance. Finally these variables along with the tensor of inertia eigenvalue ratio were optimized and a possible new filter was found, passing 73.0% $E^{-2} \nu_e$ signal for a background of 1%.

I. INTRODUCTION

IceCube is a 1 km³ Cerenkov neutrino telescope situated within the ice at the South Pole. On completion in 2011 it will consist of 86 strings of photomultiplier tubes (PMT's) contained within Digital Optical Modules (DOMs) which detect the amplitude and arrival time of photons. A neutrino traversing the detector can undergo weak interactions with nuclei in the ice creating a deluge of charged particles and consequently Cerenkov radiation. On the other hand, the background of atmospheric muons leave long tracks of Cerenkov radiation as they move close to the speed of light through the detector.

The background rate is over a million times greater than the signal, so effective filters are necessary to select which events are to be kept for further analysis. This study will look at the existing online filter which must

have an event selection rate within the bandwidth of the satellite transmitting the data from pole. There is also an additional constraint on the amount of CPU power at the pole due to the inaccessibility of the location and so any reduction in the amount of power needed for the filter would also be beneficial.

The purpose of this study is to look for new variables as well as for improvements in the existing variables to improve both filter efficiency and reduce its CPU usage. The signals detected are described in the next section followed by the detector details in Section 3. The event signatures are outlined in Section 4 before explaining the variables in Section 5. The optimization of the new and old cut variables is discussed in Section 6, before conclusions are drawn in Section 7.

II. SIGNAL

The majority of neutrinos arriving at the Earth are produced inside the Sun by nuclear fusion reactions with an energy of a few MeV. Neutrinos are also produced in the upper atmosphere by the interaction of primary cosmic rays, mainly protons, to produce muons and secondary cosmic rays. These secondary cosmic rays include pions which decay as follows [7]:

$$\pi^+ \longrightarrow \mu^+ + \nu_\mu \longrightarrow e^+ + \nu_e + \bar{\nu}_\mu + \nu_\mu \quad (1)$$

The neutrinos produced in this interaction are known as atmospheric neutrinos. This reaction also produces muons which are the main background signal for IceCube. In addition there are astrophysical neutrinos, with energies above 1 TeV, from processes such as supernovae or those that generate high energy cosmic rays. Finally there are also low energy relic neutrinos left over from the Big Bang which also stream through the Earth [7]. IceCube is designed to look for high energy astrophysical neutrinos and so has an energy threshold of around 100 GeV with the deep core.

This filter is designed to select cascade-like events which arise from all flavours of neutrino.

III. ICECUBE DETECTOR

IceCube consists of 5160 DOMs spaced at 17 m intervals on 86 separate strings, spac-

ing 125 m, sunk to a depth of 2.45-1.45 km[1] (Fig. 1). The bottom half of each DOM contains a photomultiplier tube (PMT) encased in mu-metal to protect it from the Earth's magnetic field. The signal this detects is then digitalized within the DOM to avoid losses in the long wires to the surface. Each DOM also contains a LED panel for testing purposes. To avoid spurious signals, for example from radioactive decays, the DOMs will only trigger if the nearest or next-nearest DOM on a string also fires. The ice at this depth is

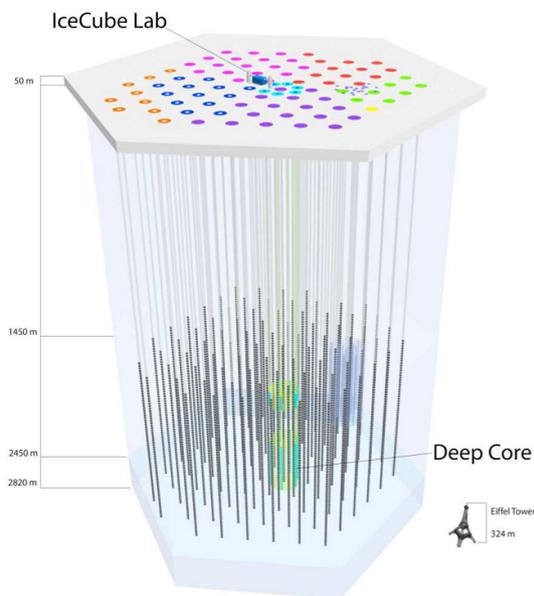


FIG. 1: Schematic of the IceCube detector showing the strings of DOMs in grey. The colours indicate the year the strings were installed: red 06/07, pink 07/08, purple 08/09, blue 09/10 and orange this coming winter.[2]

generally very clear, although it is inhomogeneous due to a few large dust layers which absorb photons (especially around depths of 1.95-2.15km). The ice also contains small air bubbles and dust grains, which results in significant scattering of photons.

This experimental setup was verified in IceCube's predecessor AMANDA. In addition

there is a 'deep core' section where the DOMs are more closely spaced to enable measurements of lower energy events.

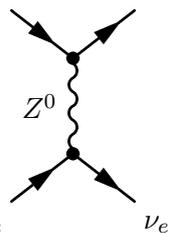
IV. EVENT SIGNATURES

There are two main types of event signature: cascades and tracks.

Cascades arise from hadronic showers produced when a neutrino interact with an atomic nuclei within the ice. As neutrinos can only interact via the weak force (gravity can be neglected due to their small mass) they undergo two types of interaction: charged-current (CC) and neutral-current (NC)[1].

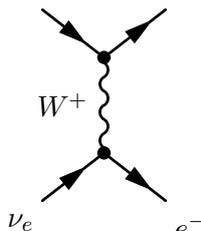
NC Example:

$Nuclei \Rightarrow d$ $d \Rightarrow HadronicCas.$



CC Example:

$Nuclei \Rightarrow d$ $u \Rightarrow HadronicCas.$



The resulting hadronic cascade is scattered through the ice resulting in a nearly spherical photon distribution. The intensity of photons in a cascade and hence charge, Q , decays rapidly with distance, d . This decay can be fitted by Eq. 2 [1, 3] where E is the energy and the fit parameters $A=3.3 \text{ GeV}^{-1}\text{m}$ and $B=\frac{1}{29} \text{ m}$.

$$\frac{Q}{E} = \frac{A}{d} \exp(-Bd) \quad (2)$$

All flavours of neutrino create nearly point-like cascades, although there are certain differences between them [5]. Electron neutrinos and tau neutrinos produce a corresponding lepton which does not produce a long track as it either loses its energy rapidly through bremsstrahlung or decays to other leptons in the case of the tau. The decay of the tau can create a distinctive 'double-bang' signature if its energy is sufficient to separate the two cascades; around 1PeV. On the other hand charge-current interactions of muon neutrinos produce a long track from the muon in addition to the cascade enabling good angular reconstruction, although there is a significant background from atmospheric muons.

Muons are traveling close to the speed of light and so will emit Cerenkov radiation along their track without significantly reducing their energy. From the Cerenkov formula the angle of Cerenkov radiation in ice is $\sim 41^\circ$ [3]. Also in contrast to the weakly interacting neutrinos, the muons cannot traverse the Earth and so will only be seen traveling downwards in the detector. Finally the muon can occasionally knock an electron out of a nuclei in the ice, which will lead to an electromagnetic cascade through bremsstrahlung resulting in the creation of electron-positron pairs and hence more bremsstrahlung [3]. This will also be detected as a cascade-like event in the detector.

These differences, in shape, charge distribution etc., will be exploited to create a filter to separate the signal from the background.

V. RECONSTRUCTION

If triggered the PMT inside each DOM records a continuous waveform from which the number of pulses, corresponding to photons, must be determined. Either we use the 'FeatureExtractor' to fit pulses of different times, and amplitudes into the waveform, known as multi-peak extraction¹ or alternatively the number of photons is calculated from the total amplitude of the waveform. In this case, known as single-peak extraction², all the photons are given the same start time. The charge, time and position relative to the detector centre of each pulse in the multi-peak case or of the combined pulse in the single-peak are referred to as c_i , t_i and r_i in later sections. As the multi-peak method requires a more complicated algorithm it is more CPU intensive than the single-peak.

The following analysis uses simulated MonteCarlo data based on IceCube with 79 strings (IC79). The background is simulated from primary cosmic rays producing showers in the atmosphere³ and the signal simulates electron neutrinos⁴. A weighting scheme is applied so that the signal events have the E^{-2} energy spectrum of astrophysical neutrinos and that the weights sum to the detection rate in Hz. Astrophysical neutrinos are expected

to come from environments, such as supernova remnants, where Fermi acceleration can occur, giving the E^{-2} power law. Fermi acceleration occurs when charged particles bounce back and forth across 'magnetic mirrors' created from moving magnetic inhomogeneities, for example across shock waves (see [4]).

A. Tensor of Inertia

We can reconstruct the position of an event by assuming each DOM has a mass proportional to the charge it detects and then finding the centre of gravity of this distribution. This gives a reasonable result for cascades, although for tracks it is likely to focus on any bright electromagnetic cascades along the track produced from bremsstrahlung of a displaced electron. In addition the tensor of inertia of this mass distribution can be calculated. For a spherical, cascade-like distribution the three eigenvalues will be similar whilst for a 2D track one eigenvalue is likely to be close to zero. Hence the ratio of the smallest eigenvalue to the sum of all the eigenvalue (ToI ratio) provides a good cut variable (Fig. 2).

B. LineFit Velocity

The LineFit reconstruction fits a straight line through the triggered DOMs by minimizing Eq. 3 where the centre of gravity, r , and the LineFit velocity, v , are free parameters

¹ As this is commonly used by the cascade group it has also been referred to in legends as 'Cascade' or 'Cas'.

² As this is commonly used by the muon group this has also been referred to in legends as 'Muon'.

³ This is also referred to as 'Corsika' or 'Cor', after the program which simulated it.

⁴ This is referred to as 'Nue' in legends.

[8].

$$r_i = r + v \times t_i \quad (3)$$

The analytic solution gives the LineFit velocity as:

$$|\vec{v}| = \frac{\langle r_i t_i \rangle - \langle r_i \rangle \langle t_i \rangle}{\langle t_i^2 \rangle - \langle t_i \rangle^2} = \frac{A - B}{C} \quad (4)$$

where angle brackets denote average over all i [6]. The velocity of cascades should be close to zero whilst the background tracks should have a velocity close to the speed of light (0.3 m(ns)^{-1}). Hence this can be used as a cut variable (Fig. 3). Note that the electron neutrino data with the single-peak extraction gives a higher LineFit velocity than with the multi-peak extraction, whilst they are the same for background.

If the variance of time in the denominator of this equation (term C) is the signifi-

cant factor, assuming that the variance of the arrival times is bigger for cascades than for tracks, then we could reduce the LineFit velocity variable to just term C . Instead, here we find that neither term A , B or C was the dominating one.

Next the formula was rewritten as:

$$|\vec{v}| = \frac{\langle (r_i - \langle r_i \rangle) t_i \rangle}{\langle t_i^2 \rangle - \langle t_i \rangle^2} \quad (5)$$

It has been shown that the magnitude of the position of the DOM with reference to the centre of gravity ($r_i - \langle r_i \rangle$) is the term that differs between single and multi-peak extraction (Fig. 4). Finally charge weighting the velocity (Eq. 6) does not enhance the separation between signal and background in either the single or multi-peak mode, so this is not an effective cut variable (Fig. 5). Again there is a difference between single and multi-peak

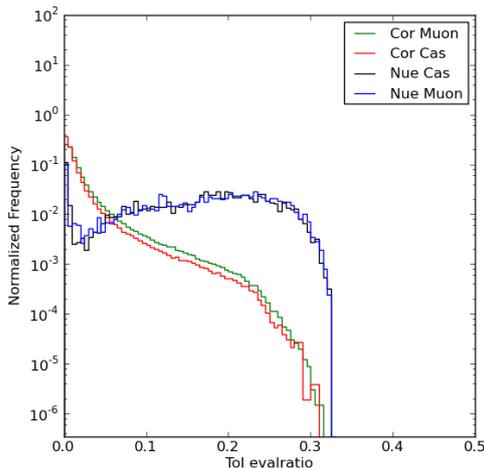


FIG. 2: A normalized histogram showing the tensor of inertia eigenvalue ratio (ToI) for signal and background using both single and multi-peak extraction. This shows that selecting events with $\text{ToI} > \sim 0.1$ would remove significant background whilst keeping the majority of the signal.

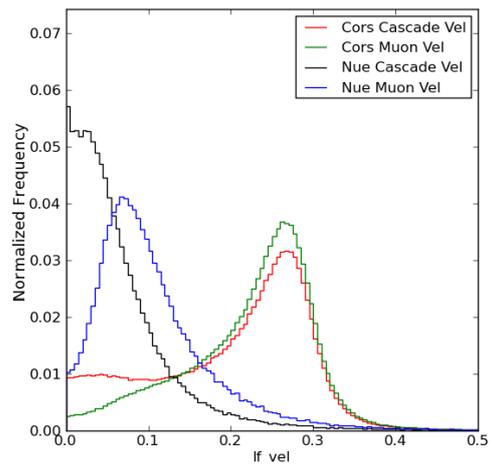


FIG. 3: A normalized histogram showing the LineFit velocity (m(ns)^{-1}) for signal and background using both single and multi-peak extraction. The muon background is around the speed of light as expected. Note the difference in single and multi-peak extraction for neutrinos.

extraction for only the neutrino data.

$$|\vec{v}| = \frac{\langle r_i c_i t_i \rangle - \langle r_i c_i \rangle \langle t_i \rangle}{\langle c_i \rangle (\langle t_i^2 \rangle - \langle t_i \rangle^2)} \quad (6)$$

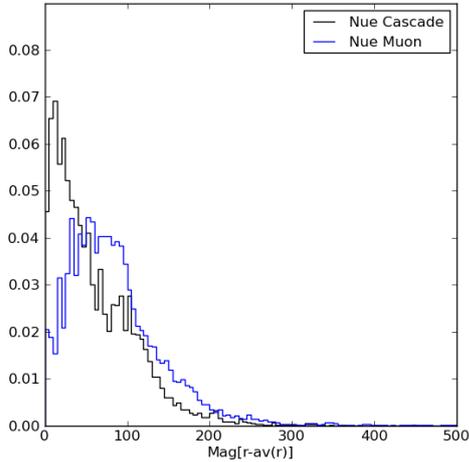


FIG. 4: The magnitude of a DOMs position with reference to the centre of gravity for ν 's using both single (blue) and multi-peak (black) extraction. Again the multi-peak has smaller values than the single-peak.

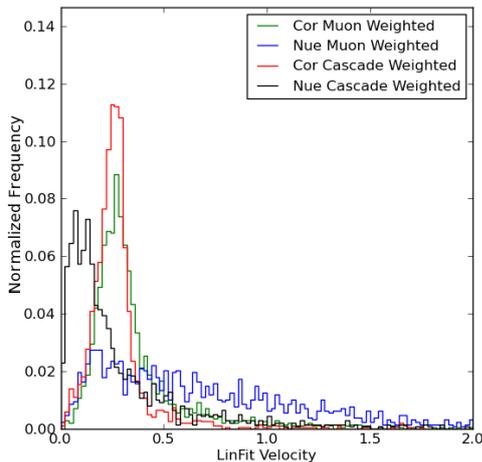


FIG. 5: A normalized histogram showing the LinFit velocity weighted with charge (Eq. 6) for signal and background data sets using both single and multi-peak extraction. Similarly to the unweighted case the multi-peak signal has smaller values than the single-peak for signal, whilst both extraction modes have similar values for background data. Note also that the single-peak data in both cases has a larger tail.

C. Cumulative Charge

Firstly, the cumulative distribution of charge over time may be calculated for a single event from summing the instantaneous charge over time, as depicted in Fig. 6. Taking the time (e.g. t_{50}) at a certain per-

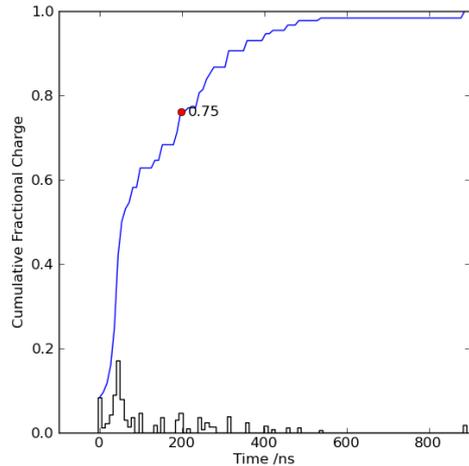


FIG. 6: The charge distribution of a single ν_e event (black) with the cumulative distribution shown in blue. The red dot marks t_{75} at 75% total charge.

cent of the total charge (e.g. 50%) will give an idea of the spread of the charge. t_{50} is defined in Eq. 7, where c_i refers to the charge at DOM- i .

$$t_{50} : \frac{\sum_{t_i \leq t_{50}} c_i}{\sum_i c_i} = 0.5 \quad (7)$$

Cascades are expected to collect more light close to the interaction vertex as the intensity dies rapidly (Eq. 2) whilst the muons will have some very long times due to their long tracks giving a wider spread of times. Hence this could be an effective cut variable, especially if used with the single-peak extraction.

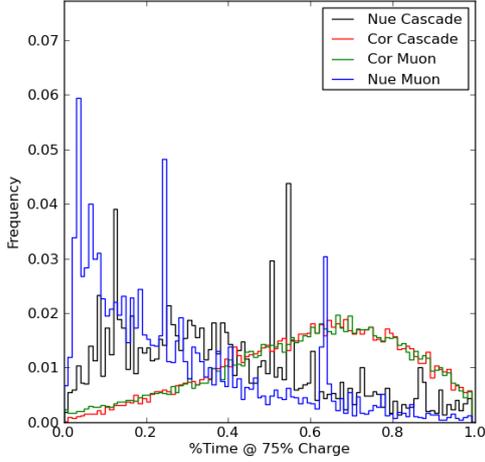


FIG. 7: The times, relative to the total event time, at 75% of the total charge (t_{75}) for an event. Hence selecting events with $t_{75} < \sim 0.5$ would reduce the background without too much loss of signal. Note the small number of statistics resulting in the spiky distribution.

Next the parameter percentage of total charge was optimized to find the value which gives the largest difference between the signal and background. The results are shown in Fig. 8. Hence taking the time around 60-70% of the total charge will give the best cut results. This can be further seen in Fig. 9.

Using the same variable on a DOM-wise basis with multi-peak extraction defined by Eq. 8, where c_{ij} is the charge of pulse- i at DOM- j and t_{ij} is the time of pulse- i relative to the time of the first pulse at DOM- j , did not show any marked differences between signal and background. Hence this was not investigated further.

$$t_{50} : \frac{\sum_j \sum_{t_{ij} \leq t_{50,j}} c_{ij}}{\sum_j \sum_i c_{ij}} = 0.5 \quad (8)$$

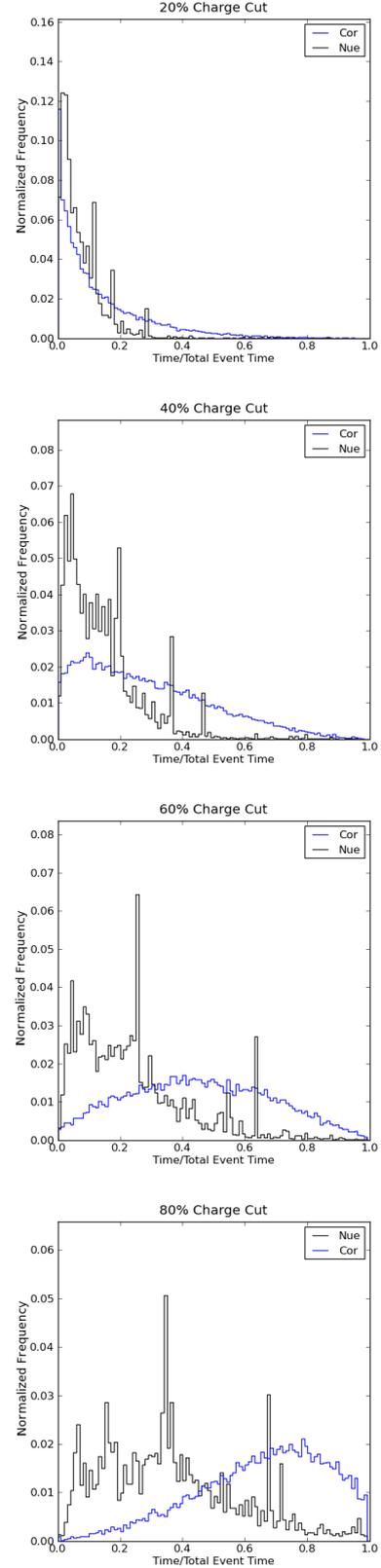


FIG. 8: Normalized histograms showing the difference between signal (black) and background (blue) for different percentage charge (multi-peak). Here an 80% charge cut gives the greatest separation. Note the small number of statistics resulting in the spiky distribution.

D. Charge Propagation

The charge-energy ratio for a cascades has been shown to decay rapidly over distance (Eq. 2) whilst the ratio for a track should be roughly constant. Hence providing another cut variable. In the following figures (Figs. 10 and 11) this can be seen for the single-peak extraction in both the idealized case, using the Monte Carlo position and energy, and in the 'real' case, estimating the energy with the total charge of an event and position with the centre of gravity. It can be seen that this variable relies heavily on the interaction vertex which is not accurately represented by the centre of gravity, especially for the background muons where it is likely to select an electromagnetic cascade along the track (see section 5.A). This will always be inside the detector whilst the actual interaction vertex is likely to be outside the detector, hence giving the much smaller range of lengths for the

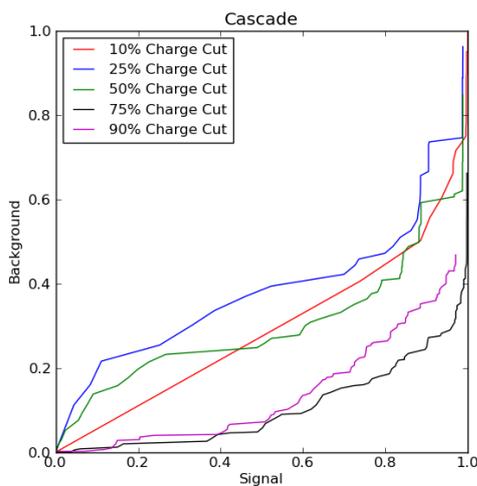


FIG. 9: The background and signal efficiencies for different percentage charge cuts. Hence for most of the signal efficiency range 75% is the most effective cut.

background muons.

In order to separate the signal and background the histogram was projected onto another axis. This was done by counting the number of points which lie between lines

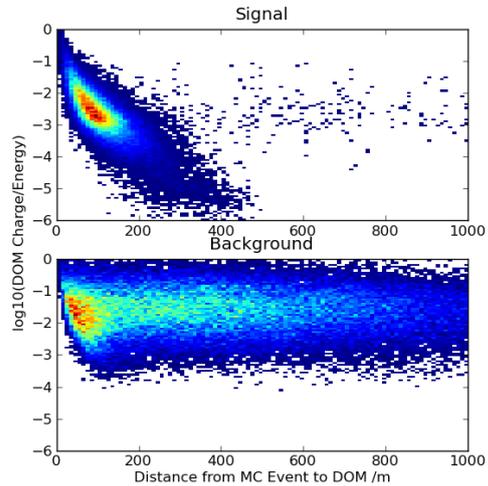


FIG. 10: The distribution of charge/event energy over distance away from the MonteCarlo event centre for each DOM in the idealized case. The background charge/total charge is shown to be roughly constant over distance, whilst the signal charge/total charge falls rapidly. Hence this could be used to separate the signal and background.

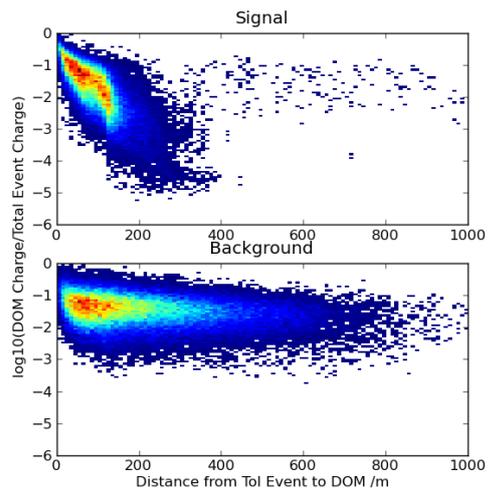


FIG. 11: The distribution of charge/total event charge over distance away from the ToI event centre for each DOM in the 'real' case. The same pattern is seen for signal and background, hence this could be used as a possible cut variable.

$\frac{Q}{Q_{TOTAL}} = \exp(-\frac{x}{50} + 20 - \frac{C}{3})$ where C varies between 0 and 100 and histogramming this with C as the x-variable. A large separation between signal and background was found for the weighted mean (M_{CD}) of these points for a single event. Here each point was weighted by its distance (in m) divided by 100 m to give more importance to distant DOMs (Figs. 12 and 13).

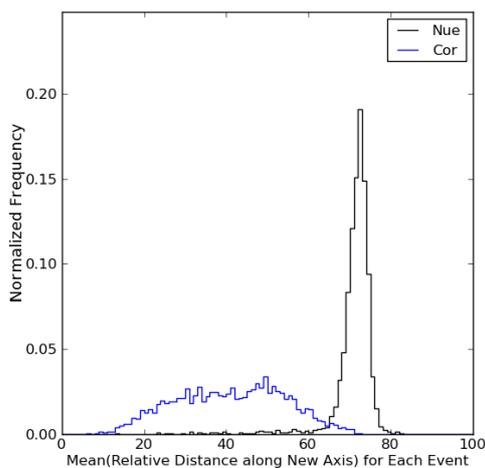


FIG. 12: A normalized histogram showing the weighted mean of an event's charge decay (M_{CD}) projected onto the new axis in the idealized case, using the MonteCarlo position and energy. Hence selecting events with a mean $> \sim 60$ would keep the majority of signal whilst removing most of the background.

VI. OPTIMIZATION

The amount of signal and background passing a filter from these new variables in combination or in combination with existing single-peak variables is optimized under the boundary condition of a background of 1%, given by the satellite bandwidth. This is shown in the following graphs: Figs. 14 and 15. The existing filter using a combination

of LineFit velocity and tensor of inertia ratio with the multippeak extraction achieves an efficiency of 71.7% [6], although with a smaller detector. For the larger detector a similar result is obtained using the same single-peak variables. A slightly improved efficiency of 73.0% can be obtained using M_{CD} and the tensor of inertia eigenvalue ratio (Fig. 15). This filter is also slightly better at passing low energy events (Fig. 16).

VII. CONCLUSION

In conclusion several variables are investigated as possible cut variables, including LineFit velocity, cumulative charge with time and charge-energy ratio decay over distance. It is found that:

- The LineFit velocity cannot be re-

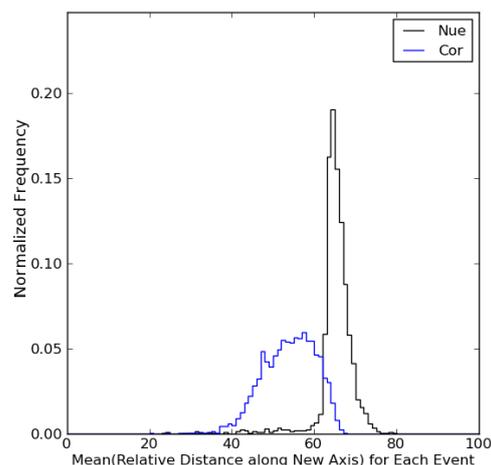


FIG. 13: A normalized histogram showing the weighted mean of an event's charge decay (M_{CD}) projected onto the new axis in the 'real' case, using the ToI position and the total charge of an event. Although the separation between signal and background is not as pronounced as in the idealized case, selecting events with a mean $> \sim 60$ would still be an effective cut.

placed by the variance of time.

- The cumulative charge distribution on an event-wise basis was shown to be a possible cut variable although not on a DOM-wise basis.
- The distance weighted mean along a new axis of the charge-energy ratio decay (M_{CD}) was also shown to be a possible cut variable.

- A possible new filter is suggested with cut values of: $M_{CD} > 61.4$, ToI ratio > 0.14 passing 73.0% signal for a background of 1%.

Future work should include optimization of the line used for the projection of the charge-energy ratio decay histogram and the weighting mechanism of the mean (M_{CD}). As the charge decay variable can be used with single-peak extraction there would be CPU power

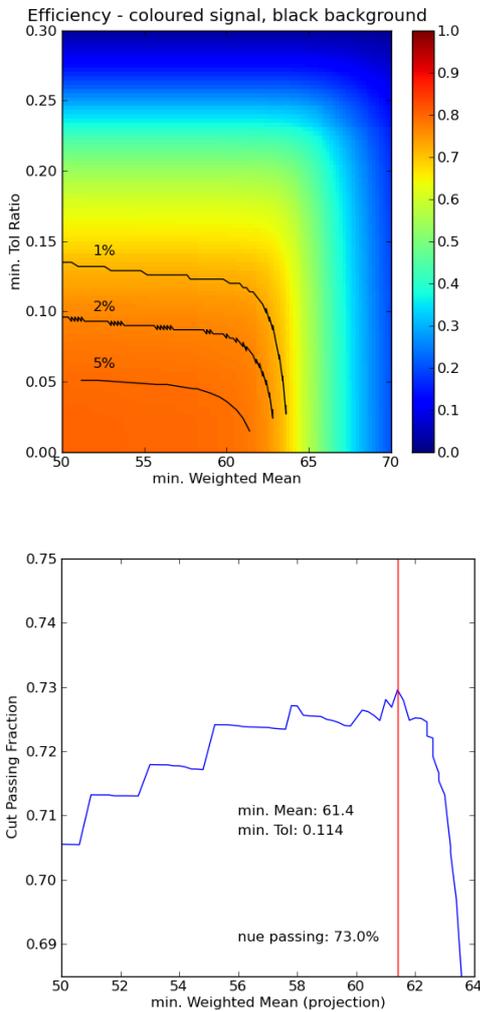


FIG. 14: TOP: The signal passing efficiency is shown in colour for the cut variables: weighted mean of charge decay (M_{CD}) and tensor of inertia eigenvalue ratio. Overlaid in black are the positions of possible cuts yielding a background passing rate of 1, 2 & 5%. BOTTOM: For 1% background the cut values are optimized for maximum signal.

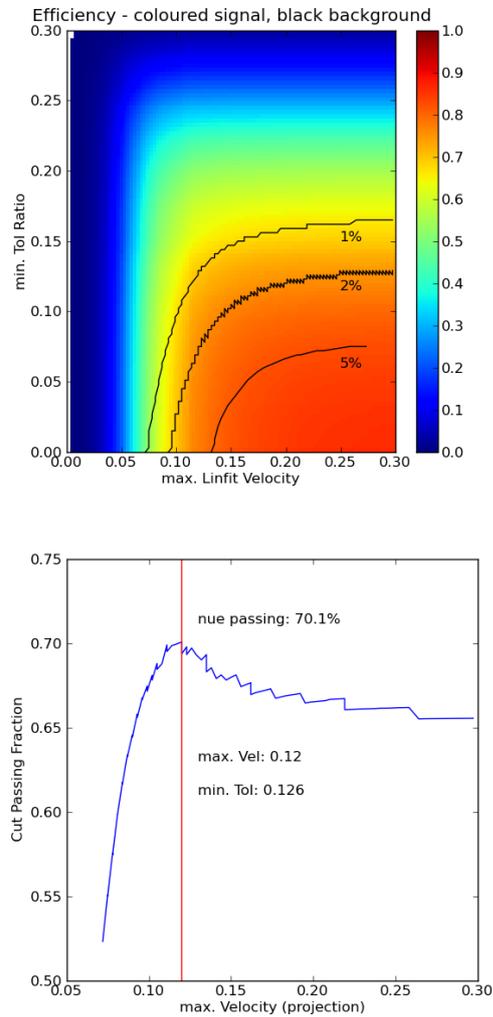


FIG. 15: TOP: The signal passing efficiency is shown in colour for the cut variables velocity and tensor of inertia ratio. Overlaid in black are the positions of possible cuts yielding a background passing rate of 1, 2 & 5%. BOTTOM: For 1% background the cut values are optimized for maximum signal.

left to run vertex reconstruction algorithms at the pole to replace the center of gravity, which could be investigated further. In addition other existing variables such as the zenith angle could be included into the filter. Finally a comparison of simulated data with the real data should be made.

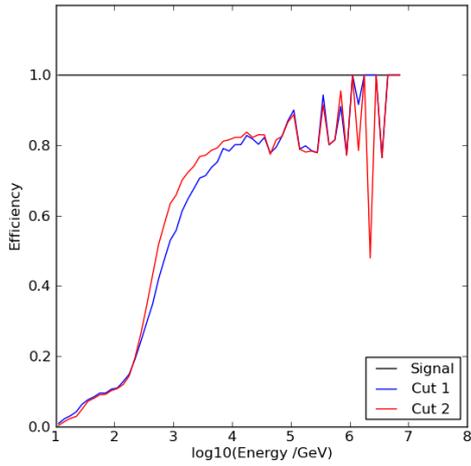


FIG. 16: The effect of these cuts (Cut1 is Velocity $< 0.12 \text{ m}(\text{ns})^{-1}$ & ToI > 0.126 , Cut2 is $M_{CD} > 61.4$ & ToI > 0.114) on the energy spectrum. Both cuts provide similar efficiencies across the energy range, in particular both have a poor efficiency for low energy events.

-
- [1] E. Middell. *2008 Reconstruction of Cascade-Like Events in IceCube*
- [2] P. Desiati. *2010 Neutrino Astrophysics and Galactic Cosmic Ray Anisotropy in IceCube*
- [3] M. Kowalski. *2003 Search for Neutrino Induced Cascades with AMANDA-II Detector*
- [4] M. S. Longair. *1981 Stars, the Galaxy and the Interstellar Medium, CUP*
- [5] E. Middell. *2009 Cascade and ν_c Reconstruction in IceCube, MANTS*
- [6] E. Middell & S. Panknin. *2009 TFT Proposal, Request for the Cascade Online Filter*
- [7] C. Giunti & C. W. Kim. *2007 Fundamentals of Neutrino Physics and Astrophysics, OUP*
- [8] H. Landsman. *2010 IceRec LineFit Module Documentation*