Data-based evaluation of direction reconstruction for IceCube cascade events by utilizing starting tracks

The IceCube Collaboration
(a complete list of authors can be found at the end of the proceedings)

E-mail: yosuke.ashida@icecube.wisc.edu

The IceCube Neutrino Observatory instruments a cubic-kilometer of glacial ice and has been the first experiment to identify high-energy astrophysical neutrinos. There are two main morphologies of IceCube events: tracks and cascades. Tracks result from muons, while cascades result from particle showers induced by in-ice interactions. The directional reconstruction of cascades is less precise than that of tracks, which limits the sensitivity of astrophysical neutrino analyses with cascade events. In order to improve the directional reconstruction of cascade events, accurate ice modeling is essential. However, potential biases might exist in data stemming from unconstrained systematic uncertainties. In this work, feasibility studies to better understand the ice using a data-driven approach are performed, where photons that are likely to have been induced by the hadronic cascade part of muon neutrino charged-current interactions are categorized using probability density functions in time, distance and angle, and the reconstructed direction with this pseudo-cascade is compared with the track direction. In this proceedings, methodology and results are detailed and a path towards better understanding of the ice is discussed.

Corresponding authors: Yosuke Ashida\textsuperscript{1*}
\textsuperscript{1} Department of Physics and Astronomy, University of Utah

* Presenter
1. Introduction

The IceCube Neutrino Observatory provides important science results, e.g., the discovery of astrophysical neutrinos \cite{1}. It instruments a cubic-kilometer of glacial ice at the geographic South Pole and consists of 5,160 digital optical modules (DOMs) in 86 vertical drill holes to detect Cherenkov light emitted from relativistic charged particles traveling inside ice. Observed neutrino events in IceCube can be categorized into two main morphologies: tracks from high-energy muons produced in $\nu_\mu$ charged-current (CC) interactions and cascades from neutral-current (NC) interactions or $\nu_e/\nu_\tau$ CC interactions. Due to their large extension, track directions are inferred much more precisely than cascades. In IceCube, the directional pointing accuracy is usually deduced from Monte-Carlo (MC) simulation as no reference point sources are available. Cosmic ray absorption by the Moon leads to a localized deficit of muons in the sky, but this is only useful for track events and is also limited by the fact that the Moon is not a point source \cite{2}.

In this work, a data-based method is employed by comparing cascade directions to tracks on a per-event basis in order to probe ice model effects (see Ref. \cite{3} for the preceding work that motivated the present study). Here, a selection comprised of predominantly $\nu_\mu$ CC interactions is utilized. Such events are deep inelastic and composed of an initial hadronic cascade in addition to an outgoing muon track. On average, the track and cascade directions are expected to be aligned at high energies ($E_\nu \gtrsim 1$ TeV). For each event, the detected photons are split into those likely to have been produced by the hadronic cascade and those likely emitted by the outgoing muon. The hadronic part of the event is reconstructed with a cascade reconstruction algorithm and the resulting direction is compared to the reconstructed track direction. A newly developed sample based on the starting track selection (ESTES) \cite{4–7}, specifically the one tailored for the neutrino source analysis, is used. The benefit of using this sample is that a high purity of $\nu_\mu$ CC events is achieved ($\sim 97\%$ for the entire sample) with most events having sufficient track length for accurate reconstruction. The performance of track reconstruction in this sample is high; the median standard deviation evaluated by MC is as good as 0.5 degrees in both zenith and azimuth directions, hence the track direction can be used well as a reference for the cascade direction. Based on MC, the median neutrino energy of the sample is around 2 TeV, thus the median hadronic shower energy on which this result is based will be a fraction of that.

2. Photon Selection and Cascade Reconstruction

Cherenkov photons contributing to each detected event are categorized into those likely to have originated from the initial hadronic cascade and otherwise. This is achieved based on log-likelihood ratios (rlogL) calculated using probability density functions (PDFs). Two types of 2D PDFs are prepared: arrival time vs. distance, and opening angle vs. distance. Figure 1 shows photon arrival times as a function of distance from the reconstructed interaction vertex, for simulated $\nu_\mu$ CC starting events at ESTES final level, with the photon propagation being based on the ice model Spice 3.2.1. The emission of Cherenkov photons from the initial hadronic cascade only happens close to the interaction vertex (up to $\sim 10$ m) while the muon emits light continuously along its trajectory. The distributions look as expected and are normalized so as to be used as PDFs later. Secondly, as shown in Figure 2, the opening angle between the reconstructed muon direction and
the vector from the vertex to a detecting DOM is taken as a parameter and plotted for different distances from the interaction vertex. It is found that the hadronic cascade-originated photons are detected around the vertex and in a wide angle range because the initial hadron and its daughter particles undergo multiple scattering. In contrast, the muon-originated photons are mainly detected far away from the vertex, which distribute at small opening angles in the figure. The rlogL is then calculated for each detected photon using these PDFs as,

$$r\log L = \log \left( \frac{L_h}{L_\mu} \right), \quad L_h = \prod_i \text{PDF}^i_h, \quad L_\mu = \prod_i \text{PDF}^i_\mu,$$

where $i$ stands for the index of two types of PDFs and $h$ ($\mu$) for the hadron (muon) likeliness. Here, the two 2D PDFs are known to be correlated (and feasibility of using 3D PDFs is under study). The calculated rlogL distributions are shown for fractional events randomly chosen from the simulated $\nu_\mu$ CC events in Figure 3. It should be noted that a relative scale between the hadronic and muonic-photon distributions differs event-by-event depending on some factors such as energy split and track length, which may affect the resulting performance. Such impact on individual events will be checked in detail as a future work.

**Figure 1:** The correlated 2D PDFs of the arrival time of photons as a function of distance from the vertex to DOMs for the hadronic- (left) and muonic-photons (right) produced using the simulated $\nu_\mu$ CC events.

**Figure 2:** The correlated 2D PDFs of the opening angle between the muon direction and the vertex-to-DOM vector as a function of distance from the vertex to DOMs for the hadronic- (left) and muonic-photons (right) produced using the simulated $\nu_\mu$ CC events.
IceCube Pseudo-Cascade Study

Figure 3: Log-likelihood ratios for the hadronic and muonic photons from the simulated $\nu_\mu$ CC events. Note that for reconstruction, the direct correspondence between photons and pulses is lost and pulses are selected using the photon classification as an approximation. See text for more details.

Table 1: Summary of the pulse samples used in this study.

<table>
<thead>
<tr>
<th>Sample (pulse set)</th>
<th>Selection criteria</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudo-cascade</td>
<td>All pulses with $0 &lt; \text{rlogL} &lt; 6$</td>
<td>Data, MC</td>
</tr>
<tr>
<td>Near-cascade</td>
<td>All hadronic pulses</td>
<td>MC</td>
</tr>
<tr>
<td>Cascade with hadronic</td>
<td>Hadronic pulses with $0 &lt; \text{rlogL} &lt; 6$</td>
<td>MC</td>
</tr>
<tr>
<td>pulse loss</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cascade with muonic</td>
<td>All hadronic pulses and muonic pulses with $0 &lt; \text{rlogL} &lt; 6$</td>
<td>MC</td>
</tr>
<tr>
<td>pulse contamination</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In IceCube, detected photons are converted into pulses which may contain more photons than one in case those arrive at a DOM in a short time. This processing is performed in MC to make one-to-one comparison to observed data (Data) possible, and information on the parent particle is lost at the pulse level. The rlogL in Figure 3 is based on detected photons and some accuracy is lost when applied to pulses. In this study, multiple sets of pulses are prepared depending on the selection criteria as summarized in Table 1. First, the pulses with $0 < \text{rlogL} < 6$ are selected and the resulting set is referred to as a pseudo-cascade sample. This selection is equally applied to both Data and MC. Another set of pulses which are likely be emitted from the hadronic cascade is prepared based on the MC information. These pulses are selected as ones recorded within $\pm 5$ ns from the true hadronic photons (red in Figure 3) and are referred to as a near-cascade sample. It may be that this time-based selection also includes pulses generated by the muon or later stochastic losses, as the parent particle association is not preserved at pulse level. This set is used as a truth reference for the pseudo-cascade sample. In addition, two more sets of pulses are prepared from MC, so to be able to investigate the impact of the imperfect rlogL-based pulse selection. One of these sets contains only the pulses originating from the hadronic cascade which fall within the $0 < \text{rlogL} < 6$ selection window. This sample is called cascade with hadronic pulse loss. The other set is produced by adding the muon-induced pulses within $0 < \text{rlogL} < 6$ to the hadronic cascade-induced pulses and referred to as cascade with muonic pulse contamination. Note again that these hadronic or muonic
pulses are selected within a ±5 ns time window around the true hadronic or muonic photon arrivals, respectively. By comparing the results from these two sets with that from the near-cascade set, the selection impact can be estimated.

A series of the IceCube cascade reconstruction tools are applied to these pulse sets (see Ref. [8] for the reconstruction updates). For the reconstructions in this study, the ice model Spice BFR-v2 which takes into account the birefringent micro-structure of the ice [9, 10] is used along with the h2-50cm hole ice model. Note that the MC production of the starting track sample is performed based on a previous model Spice 3.2.1.

3. Results

The resulting zenith and azimuth directions from the pseudo-cascade sample and three MC-truth-based samples are compared with the track directions as shown in Figures 4 and 5, respectively. In zenith, a negative bias is observed in Data and it is larger than that of the MC pseudo-cascade sample, while their widths are similar (see the left panel of the figure). In azimuth, no sizable bias is seen in both Data and MC and the widths are comparable.

From comparison between the MC pseudo-cascade and MC near-cascade samples, it is found that the pseudo-cascade bias in zenith is shifted mainly by contamination of the muonic pulses and its resolution in azimuth is deteriorated mainly by loss of the hadronic pulses. The angular resolution, which is derived from the space angle between the reconstructed direction and the true direction, involves both zenith and azimuth. Currently, the angular resolution for pseudo-cascades is worse than that for near-cascades and this is a subject of further studies; here, only the separate distributions in zenith and azimuth are focused on. With the MC studies performed, it is shown that these distributions can potentially provide hints on the ice modeling.

![Figure 4](image_url)

**Figure 4:** Reconstructed zenith angle difference between cascade and track parts from the pseudo-cascade samples for Data and the nominal MC (left) and from the MC-truth-based samples (right). MC distributions are weighted with the best-fit parameters from the ESTES diffuse analysis [4] and all distributions are area normalized. In the left panel, the median and 25%/75% percentiles of the distribution are $-6.3^{+9.8}_{-11.2}$ degrees in Data and $-2.7^{+9.5}_{-10.2}$ degrees in MC. The corresponding ones in the right panel are $-0.9^{+8.4}_{-8.5}$ degrees for black, $-1.0^{+9.1}_{-9.1}$ degrees for red, and $-2.5^{+9.4}_{-9.4}$ degrees for blue distributions.

In the following part, the zenith bias is discussed in more details. Since the bias obtained from the Data distribution in the left panel of Figure 4 is for pseudo-cascades, it is corrected for the
Figure 5: Same as Figure 4 but for azimuth. In the left panel, the median and 25%/75% percentiles of the distribution are $0.4^{+23.3}_{-23.5}$ degrees in Data and $0.0^{+23.9}_{-23.3}$ degrees in MC. The corresponding ones in the right panel are $0.1^{+17.3}_{-16.9}$ degrees for black, $0.1^{+24.8}_{-23.6}$ degrees for red, and $0.1^{+18.2}_{-17.8}$ degrees for blue distributions.

Impacts from the logL-based pulse selection to be compared to the near-cascade result in a fair way. This is performed by the formula: $\Delta_{\text{corrected}} = \Delta_{\text{pseudocasc}} - (\Delta_{\text{hadloss}} + \Delta_{\text{casc}}) - (\Delta_{\text{mucontami}} + \Delta_{\text{casc}})$. Here, $\Delta_{\text{pseudocasc}}$ is the median of the pseudo-cascade distributions (left panel of Figure 4) and considered for Data and MC as $\Delta_{\text{pseudocasc, Data}}$ and $\Delta_{\text{pseudocasc, MC}}$, respectively. The correction part is composed of three medians, $\Delta_{\text{casc}}$, $\Delta_{\text{hadloss}}$, and $\Delta_{\text{mucontami}}$, from the near-cascade, cascade with hadronic pulse loss, and cascade with muonic pulse contamination distributions in the right panel of Figure 4. $\Delta_{\text{corrected, MC}}$ is compared with the near-cascade result ($\Delta_{\text{casc}}$) on purpose of validating the correction formula above because its form is essentially a guess. The median of each distribution is obtained for $\nu_\mu$ CC events and this is achieved for Data with a subtraction of the non-$\nu_\mu$ CC contribution using MC with considering the relative total number of events between Data and MC. It should be mentioned that the correction effect by this formula is not large in the present study, but still it is applied for the later results.

In the correction formula, $\Delta_{\text{casc}}$, $\Delta_{\text{hadloss}}$, and $\Delta_{\text{mucontami}}$ are affected by systematic uncertainties regarding physics and detector modeling. A major source of systematic uncertainty is the parametrization of the so-called hole ice, the refrozen column of ice in which DOMs are installed, where the photon propagation differs from that in the bulk ice (see Ref. [11] for more details about the hole ice). Figure 6 shows the zenith bias resulting from MC simulation sets with different assumed hole ice parameters, $\epsilon_{\text{angular, p0}}$ and $\epsilon_{\text{angular, p1}}$. It is found that the zenith bias is largely sensitive to both of the hole ice parameters. The h2-50cm model is used in the reconstruction and corresponds roughly to $(\epsilon_{\text{angular, p0}}, \epsilon_{\text{angular, p1}}) = (-0.4, 0.075)$ in the new parametrization scheme. The impacts of all considered systematic uncertainties are shown in Figure 7. The effect of uncertainties relating to the physics modeling is small in comparison to detector systematic uncertainties. The largest impact comes from hole ice parameters; however, their effect on the converted results from Data is not sizable enough to fully compensate the observed bias from zero. Consistent values of the converted results from the MC pseudo-cascade and the MC near-cascade results ensure that the assumption about the correction formula in this study is reasonable.

The results for different zenith regions, $-1.0 < \cos \theta_{\text{zenith}} < -0.5$, $-0.5 < \cos \theta_{\text{zenith}} < 0.0$, and $0.0 < \cos \theta_{\text{zenith}} < 0.5$, are shown in Figure 8. As a reference, the MC near-cascade results with
IceCube Pseudo-Cascade Study

Figure 6: Zenith bias from the MC production with different choices of hole ice parameters, $\epsilon_{\text{angular}, p0}$ (left) and $\epsilon_{\text{angular}, p1}$ (right), in comparison with that from Data. In the left and right panels, $\epsilon_{\text{angular}, p1}$ and $\epsilon_{\text{angular}, p0}$ are set to 0.0, respectively. Dashed lines represent the median of each MC distribution.

Figure 7: Zenith bias with the correction formula based on the nominal MC and the different systematic uncertainties. The left and right panels correspond to the cases with physics and detector systematic uncertainties, respectively. In each panel, the converted results from the Data pseudo-cascade (red), the converted results from the MC pseudo-cascade (blue), and the MC near-cascade results (grey) are shown. Different hole ice parameters are also plotted (the converted results from the MC pseudo-cascade are consistent with the MC near-cascade results as seen in Figure 7). Note that the $\nu_\mu$ CC purity differs bin-by-bin and so does the subtraction effect. The observed zenith bias based on Data is significantly negative in every bin and its size is larger for northern sky events. The zenith bias dependence on deposited energy of the hadronic part of $\nu_\mu$ CC events is also investigated and no strong dependence is found for the region of $10^2$–$10^4$ GeV. These dependence could be a key to understanding the ice features and further improvement of the IceCube reconstruction tool.

4. Conclusion and Outlook

In this study, a data-based method is employed on the ESTES neutrino source search sample to compare per-event reconstructions of cascade-like pulses against that of the full event. Two types
IceCube Pseudo-Cascade Study

Figure 8: Zenith bias converted from the Data pseudo-cascade with systematic uncertainties for different cosine zenith bins in comparison with the MC near-cascade results with different hole ice parameters. Note that $\epsilon_{\text{angular},p1}$ is set to 0.0 for the case of $\epsilon_{\text{angular},p0} = \pm 1.0$ and $\epsilon_{\text{angular},p0}$ is set to 0.0 for $\epsilon_{\text{angular},p1} = \pm 0.2$

of 2D PDFs are employed to calculate log-likelihood ratios for selecting Cherenkov photons from the initial hadronic cascade in $\nu_\mu$ CC events. Noting that the correspondence between photons and pulses is not one-to-one, different sets of pulses are selected based on these log-likelihood ratios and passed to angular reconstruction. The resulting zenith and azimuth directions are then compared to the reconstructed track zenith and azimuth. In particular, the zenith bias derived from the pseudo-cascade sample is converted to the bias for cascades using MC-truth-based samples. Here, systematic uncertainties regarding physics and detector modeling are considered. The results imply that a modification of the hole ice can describe observed data better, but the effect is not completely understood and further studies are needed.

References

The authors gratefully acknowledge the support from the following agencies and institutions: USA – U.S. National Science Foundation-Office of Polar Programs, U.S. National Science Foundation-Physics Division, U.S. National Science Foundation-EPSCoR, Wisconsin Alumni Research Foundation, Center for High Throughput Computing (CHTC) at the University of Wisconsin–Madison, Open Science
IceCube Pseudo-Cascade Study

Grid (OSG), Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS), Frontera computing project at the Texas Advanced Computing Center, U.S. Department of Energy-National Energy Research Scientific Computing Center, Particle astrophysics research computing center at the University of Maryland, Institute for Cyber-Enabled Research at Michigan State University, and Astroparticle physics computational facility at Marquette University; Belgium – Funds for Scientific Research (FRS-FNRS and FWO), FWO Odysseus and Big Science programmes, and Belgian Federal Science Policy Office (Belspo); Germany – Bundesministerium für Bildung und Forschung (BMBF), Deutsche Forschungsgemeinschaft (DFG), Helmholtz Alliance for Astroparticle Physics (HAP), Initiative and Networking Fund of the Helmholtz Association, Deutsches Elektronen Synchrotron (DESY), and High Performance Computing cluster of the RWTH Aachen; Sweden – Swedish Research Council, Swedish Polar Research Secretariat, Swedish National Infrastructure for Computing (SNIC), and Knut and Alice Wallenberg Foundation; European Union – EGI Advanced Computing for research; Australia – Australian Research Council; Canada – Natural Sciences and Engineering Research Council of Canada, Calcul Québec, Compute Ontario, Canada Foundation for Innovation, WestGrid, and Compute Canada; Denmark – Villum Fonden, Carlsberg Foundation, and European Commission; New Zealand – Marsden Fund; Japan – Japan Society for Promotion of Science (JSPS) and Institute for Global Prominent Research (IGPR) of Chiba University; Korea – National Research Foundation of Korea (NRF); Switzerland – Swiss National Science Foundation (SNSF); United Kingdom – Department of Physics, University of Oxford.