Robust Statistics in Muon Track Reconstruction

M. G. Aartsen[2], R. Abbasi[27], Y. Abdou[22], M. Ackermann[41], J. Adams[15], J. A. Aguilar[21], M. Ahlers[27], D. Altmann[9], J. Auffenberg[27], X. Bai[31,1], M. Baker[27], S. W. Barwick[23], V. Baum[28], R. Bay[7], J. J. Beatty[17,18], S. Bechet[12], J. Becker Tjus[10], K.-H. Becker[40], M. Bell[38], M. L. Benabderrahmane[41], S. BenZvi[27], J. Berdermann[41], P. Berghaus[41], D. Berley[16], E. Bernardini[41], A. Bernhard[30], D. Bertrand[12], D. Z. Besson[25], G. Binder[8,7], D. Bindig[40], M. Bissok[1], E. Blaufuss[16], J. Blumenthal[1], D. J. Boersma[39], S. Bohaichuk[20], C. Bohm[34], D. Bose[13], S. Böser[11], O. Botner[39], L. Brayeur[13], H.-P. Bretz[41], A. M. Brown[15], R. Bruijn[24], J. Brunner[41], M. Carson[22], J. Casey[5], M. Casier[13], D. Chirkin[27], A. Christov[21], B. Christy[16], K. Clark[38], F. Clevermann[19], S. Coenders[1], S. Cohen[24], D. F. Cowen[38,37], A. H. Cruz Silva[41], M. Danninger[34], J. Daughhetee[5], J. C. Davis[17], C. De Clercq[13], S. De Ridder[22], P. Desiati[27], M. de With[9], T. DeYoung[38], J. C. Díaz-Vélez[27], M. Dunkman[38], R. Eagan[38], B. Eberhardt[28], J. Eisch[27], R. W. Ellsworth[16], S. Euler[1], P. A. Evenson[31], O. Fadiran[27], A. R. Fazely[6], A. Fedynitch[10], J. Feintzeig[27], T. Feusels[22], K. Filimonov[7], C. Finley[34], T. Fischer-Wasels[40], S. Flis[34], A. Franckowiak[11], R. Franke[41], K. Frantzen[19], T. Fuchs[19], T. K. Gaisser[31], J. Gallagher[26], L. Gerhardt[8,7], L. Gladstone[27], T. Glüsenkamp[41], A. Goldschmidt[8], G. Golup[13], J. A. Goodman[16], D. Góra[41], D. Grant[20], A. Groß[30], M. Gurtner[40], C. Ha[8,7], A. Haj Ismail[22], P. Hallen[1], A. Hallgren[39], F. Halzen[27], K. Hanson[12], D. Heereman[12], P. Heimann[1], D. Heinen[1], K. Helbing[40], R. Hellauer[16], S. Hickford[15], G. C. Hill[2], K. D. Hoffman[16], R. Hoffmann[40], A. Homeier[11], K. Hoshina[27], W. Huelsnitz[16,2], P. O. Hulth[34], K. Hultqvist[34], S. Hussain[31], A. Ishihara[14], E. Jacobi[41], J. Jacobsen[27], K. Jagielski[1], G. S. Japaridze[4], K. Jero[27], O. Jlelati[22], B. Kaminsky[41], A. Kappes[9], T. Karg[41], A. Karle[27], J. L. Kelley[27], J. Kiryluk[35], F. Kislat[41], J. Kläs[40], S. R. Klein[8,7], J.-H. Köhne[19], G. Kohnen[29], H. Kolanoski[9], L. Köpke[28], C. Kopper[27], S. Kopper[40], D. J. Koskinen[38], M. Kowalski[11], M. Krasberg[27], K. Krings[1], G. Kroll[28], J. Kunnen[13], N. Kurahashi[27], T. Kuwabara[31], M. Labare[13], H. Landsman[27], M. J. Larson[36], M. Lesiak-Bzdak[35], M. Leuermann[1], J. Leute[30], J. Lünemann[28], J. Madsen[33], R. Maruyama[27], K. Mase[14], H. S. Matis[8], F. McNally[27], K. Meagher[16], M. Merck[27], P. Mészáros[37,38], T. Meures[12], S. Miarecki[8,7], E. Middell[41], N. Milke[19], J. Miller[13], L. Mohrmann[41], T. Montaruli[21,3], R. Morse[27],

*Corresponding author. Email: wellons@icecube.wisc.edu, Phone: 304-542-4464, Address: Wisconsin Institutes for Discovery, 330 N. Orchard St., Madison, WI 53715

[1]Physics Department, South Dakota School of Mines and Technology, Rapid City, SD 57701, USA

[2]Los Alamos National Laboratory, Los Alamos, NM 87545, USA

[3]also Sezione INFN, Dipartimento di Fisica, I-70126, Bari, Italy

[4]NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA

R. Nahnhauer[41], U. Naumann[40], H. Niederhausen[35], S. C. Nowicki[20], D. R. Nygren[8], A. Obertacke[40], S. Odrowski[30], A. Olivas[16], M. Olivo[10], A. O'Murchadha[12], L. Paul[1], J. A. Pepper[36], C. Pérez de los Heros[39], C. Pfendner[17], D. Pieloth[19], N. Pirk[41], J. Posselt[40], P. B. Price[7], G. T. Przybylski[8], L. Rädel[1], K. Rawlins[3], P. Redl[16], R. Reimann[1], E. Resconi[30], W. Rhode[19], M. Ribordy[24], M. Richman[16], B. Riedel[27], J. P. Rodrigues[27], C. Rott[17], T. Ruhe[19], B. Ruzybayev[31], D. Ryckbosch[22], S. M. Saba[10], T. Salameh[38], H.-G. Sander[28], M. Santander[27], S. Sarkar[32], K. Schatto[28], M. Scheel[1], F. Scheriau[19], T. Schmidt[16], M. Schmitz[19], S. Schoenen[1], S. Schöneberg[10], L. Schönherr[1], A. Schönwald[41], A. Schukraft[1], L. Schulte[11], O. Schulz[30], D. Seckel[31], S. H. Seo[34], Y. Sestayo[30], S. Seunarine[33], C. Sheremata[20], M. W. E. Smith[38], M. Soiron[1], D. Soldin[40], G. M. Spiczak[33], C. Spiering[41], M. Stamatikos[17,4], T. Stanev[31], A. Stasik[11], T. Stezelberger[8], R. G. Stokstad[8], A. Stößl[41], E. A. Strahler[13], R. Ström[39], G. W. Sullivan[16], H. Taavola[39], I. Taboada[5], A. Tamburro[31], S. Ter-Antonyan[6], S. Tilav[31], P. A. Toale[36], S. Toscano[27], M. Usner[11], D. van der Drift[8,7], N. van Eijndhoven[13], A. Van Overloop[22], J. van Santen[27], M. Vehring[1], M. Voge[11], M. Vraeghe[22], C. Walck[34], T. Waldenmaier[9], M. Wallraff[1], R. Wasserman[38], Ch. Weaver[27], M. Wellons[27,*], C. Wendt[27], S. Westerhoff[27], N. Whitehorn[27], K. Wiebe[28], C. H. Wiebusch[1], D. R. Williams[36], H. Wissing[16], M. Wolf[34], T. R. Wood[20], K. Woschnagg[7], C. Xu[31], D. L. Xu[36], X. W. Xu[6], J. P. Yanez[41], G. Yodh[23], S. Yoshida[14], P. Zarzhitsky[36], J. Ziemann[19], S. Zierke[1], A. Zilles[1], M. Zoll[34], B. Recht[42], C. Ré[42]

[1]*III. Physikalisches Institut, RWTH Aachen University, D-52056 Aachen, Germany*
[2]*School of Chemistry & Physics, University of Adelaide, Adelaide SA, 5005 Australia*
[3]*Dept. of Physics and Astronomy, University of Alaska Anchorage, 3211 Providence Dr., Anchorage, AK 99508, USA*
[4]*CTSPS, Clark-Atlanta University, Atlanta, GA 30314, USA*
[5]*School of Physics and Center for Relativistic Astrophysics, Georgia Institute of Technology, Atlanta, GA 30332, USA*
[6]*Dept. of Physics, Southern University, Baton Rouge, LA 70813, USA*
[7]*Dept. of Physics, University of California, Berkeley, CA 94720, USA*
[8]*Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[9]*Institut für Physik, Humboldt-Universität zu Berlin, D-12489 Berlin, Germany*
[10]*Fakultät für Physik & Astronomie, Ruhr-Universität Bochum, D-44780 Bochum, Germany*
[11]*Physikalisches Institut, Universität Bonn, Nussallee 12, D-53115 Bonn, Germany*
[12]*Université Libre de Bruxelles, Science Faculty CP230, B-1050 Brussels, Belgium*
[13]*Vrije Universiteit Brussel, Dienst ELEM, B-1050 Brussels, Belgium*
[14]*Dept. of Physics, Chiba University, Chiba 263-8522, Japan*
[15]*Dept. of Physics and Astronomy, University of Canterbury, Private Bag 4800, Christchurch, New Zealand*
[16]*Dept. of Physics, University of Maryland, College Park, MD 20742, USA*
[17]*Dept. of Physics and Center for Cosmology and Astro-Particle Physics, Ohio State University, Columbus, OH 43210, USA*
[18]*Dept. of Astronomy, Ohio State University, Columbus, OH 43210, USA*
[19]*Dept. of Physics, TU Dortmund University, D-44221 Dortmund, Germany*
[20]*Dept. of Physics, University of Alberta, Edmonton, Alberta, Canada T6G 2E1*

[21]Département de physique nucléaire et corpusculaire, Université de Genève, CH-1211 Genève, Switzerland
[22]Dept. of Physics and Astronomy, University of Gent, B-9000 Gent, Belgium
[23]Dept. of Physics and Astronomy, University of California, Irvine, CA 92697, USA
[24]Laboratory for High Energy Physics, École Polytechnique Fédérale, CH-1015 Lausanne, Switzerland
[25]Dept. of Physics and Astronomy, University of Kansas, Lawrence, KS 66045, USA
[26]Dept. of Astronomy, University of Wisconsin, Madison, WI 53706, USA
[27]Dept. of Physics and Wisconsin IceCube Particle Astrophysics Center, University of Wisconsin, Madison, WI 53706, USA
[28]Institute of Physics, University of Mainz, Staudinger Weg 7, D-55099 Mainz, Germany
[29]Université de Mons, 7000 Mons, Belgium
[30]T.U. Munich, D-85748 Garching, Germany
[31]Bartol Research Institute and Department of Physics and Astronomy, University of Delaware, Newark, DE 19716, USA
[32]Dept. of Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK
[33]Dept. of Physics, University of Wisconsin, River Falls, WI 54022, USA
[34]Oskar Klein Centre and Dept. of Physics, Stockholm University, SE-10691 Stockholm, Sweden
[35]Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794-3800, USA
[36]Dept. of Physics and Astronomy, University of Alabama, Tuscaloosa, AL 35487, USA
[37]Dept. of Astronomy and Astrophysics, Pennsylvania State University, University Park, PA 16802, USA
[38]Dept. of Physics, Pennsylvania State University, University Park, PA 16802, USA
[39]Dept. of Physics and Astronomy, Uppsala University, Box 516, S-75120 Uppsala, Sweden
[40]Dept. of Physics, University of Wuppertal, D-42119 Wuppertal, Germany
[41]DESY, D-15735 Zeuthen, Germany
[42]Dept. of Computer Sciences, University of Wisconsin, Madison, WI 53706, USA

Abstract

The IceCube detector is a high-energy neutrino telescope located at the geographic South Pole. Neutrinos cannot be directly observed and must be inferred from their interactions with other particles. These interactions sometimes generate a muon, which in turn emits observable light. At the energies the IceCube detector is sensitive to, the neutrino and generated muon have almost parallel paths, so the neutrino path can be inferred from a reconstruction of the muon path. However, reconstructing the muon path from the observed light is challenging due to noise, outliers in the data, and the possibility of simultaneously multiple muons inside the detector.

This manuscript describes our work on two problems: (1) the *path reconstruction* problem, in which, given a set of observations, our goal is to recover the path of a muon, and (2) the *coincident event* problem, which is to determine how many muons are active in the detector during a time window. Rather than solving these problems by developing more complex physical models, our approach is to augment the detector's current models with simple filters and robust statistical techniques. Using the metric of median angular resolution, a standard metric for path reconstruction, our solution improves the accuracy

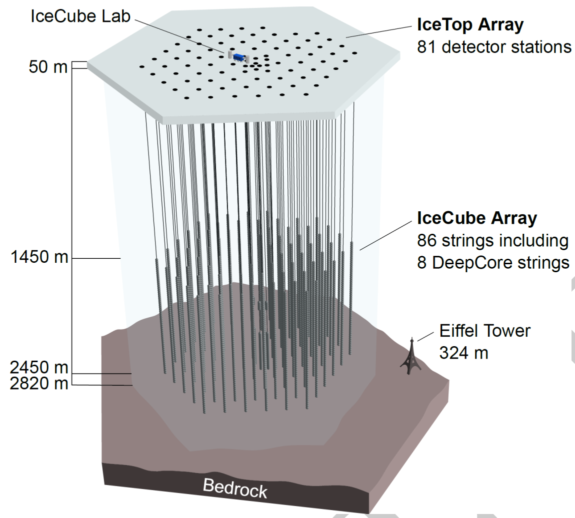

Figure 1: The IceCube neutrino detector in the Antarctic ice. A picture of the Eiffel Tower is shown for scale.

in the reconstruction direction by 13%. Our solution for the coincident-event problem accurately determines the number of muons 98% of the time, which is an improvement of 86% over the software previously used in IceCube.

Keywords: IceCube, Track reconstruction, Neutrino telescope, Neutrino astrophysics, Robust Statistics

---

1. Introduction

The IceCube neutrino detector searches for neutrinos that are generated by the universe's most violent astrophysical events: exploding stars, gamma ray bursts, and cataclysmic phenomena involving black holes and neutron stars [1]. The detector, roughly a cubic kilometer in size, is located near the geographic South Pole and is buried to a depth of about 2.5 km in the Antarctic ice [2]. The detector is illustrated in Figure 1 and a more complete description is given in Section 2.

We examine two problems that arise in the IceCube detector's neutrino detection:

1. *Reconstruction*, in which the path of a muon is reconstructed from the observed light at different positions and times in the detector.
2. *Coincident Event Detection*, in which we detect the number of muons inside the detector, and assign observed photons to a muon.

The IceCube Collaboration has spent considerable effort on both of these problems over the last decade, as they are a critical step for data analysis. They have developed sophisticated domain models that take into account the interaction of near- and far-field effects of light, and have undertaken complex mapping efforts to understand the effects of photon propagation in the ice [3, 4]. Our solutions do not further refine the detailed modeling of these physical effects, but instead augment the models with off-the-shelf statistical techniques combined with some simple data filtering to remove outliers.

Related Work. Track reconstruction and coincident event detection challenges are ubiquitous in particle physics [5–7], both in particle accelerators and cosmic particle detectors. While the work described in this manuscript builds on the previous technique developed for the IceCube detector [3], our techniques are general purpose, and potentially have applications in detectors beyond IceCube.

Outline. We begin by describing the necessary background on the IceCube detector in Section 2. In Section 3, we describe the reconstruction pipeline including the prior IceCube software, then we discuss our work and its results. Section 4 describes our work on coincident event detection, and follows a parallel structure to Section 3. We describe how in this application, a simple heuristic approach is an improvement over the prior software. We close with a conclusion in Section 5.

2. Background

The IceCube detector is composed of 5160 optical detectors, each composed of a photomultiplier tube (PMT) and onboard digitizer[8]. The PMTs are spread over 86 vertical strings arranged in a hexagonal shape, with a total instrumented volume of approximately a cubic kilometer. The PMTs on a given string are separated vertically by 17 m, and the string-to-string separation is roughly 125 m.

When a neutrino enters the telescope, it sometimes interacts with the ice and generates a muon. As the muon travels though the detector, it radiates light[9], which is observed by the PMTs and broken down into discrete *hits*[10]. A collection of hits is called an *event*, and if the number of hits in an event is sufficiently large, the muon path reconstruction algorithm is triggered.

In addition to neutrinos, muons can also be generated by cosmic rays. IceCube analyses on neutrinos are not interested in cosmic ray muons, and the detector attempts to separate out the cosmic ray muons from the neutrino muons. The primary mechanism for this separation is reconstructing the muon path, and determining if the muon was traveling downwards into the Earth or upwards out of the Earth. Since neutrinos can penetrate the Earth but cosmic ray muons cannot, it follows that a muon traveling out of the Earth must have been caused by a neutrino. Thus, by selecting only the muons that are reconstructed as up-going, the cosmic ray muons can, in principle, be removed from the data.

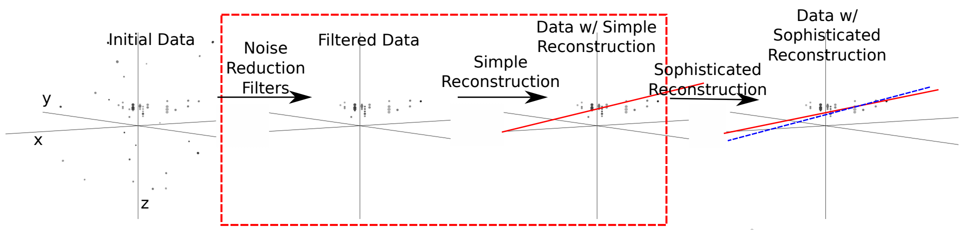

Figure 2: The reconstruction pipeline used to process data in the IceCube detector. After initial data is collected, it is then processed by some simple noise filters, which remove clear outliers. This cleaned data is processed by a simple reconstruction algorithm (red line), which is used as the seed for the more sophisticated reconstruction algorithm (dashed blue line). The sophisticated reconstruction is then evaluated as a potential neutrino. Our work in the reconstruction problem makes changes to the filtering and simple reconstruction step (indicated by the dashed red box).

2.1. Challenges in Neutrino Detection

Recovering the muon path from the light measurements is the *reconstruction* problem. The reconstruction algorithms used in the detector have several challenges, which must be overcome. The underlying mechanics are stochastic and incompletely modeled, the data is noisy and contains outliers, and the computational abilities of the detector are limited.

Modeling Difficulties. The underlying physics of the system are nontrivial to model. The muon's light is scattered by the dust and air crystals in the ice medium. This scattering is both complex and stochastic, and the scattering properties of the ice vary with depth [11].

Noise. An unescapable challenge is the noise inherent in the data. The PMTs are so sensitive to light that they can record hits even in the absence of nearby muons. These hits can arise from photons generated either by radioactive decay inside the PMT or the triboluminescence [12] of the ice.

Computational Constraints. The reconstruction algorithms are also limited in complexity by the computing resources available at the South Pole. The path reconstruction algorithm has to process about 3000 muons per second, so algorithms with excessive computational demands are discouraged.

3. Reconstruction Problem

By augmenting the reconstruction algorithm with some simple filters and classical data analysis techniques, we show significant improvement in the reconstruction algorithm's accuracy.

3.1. Prior IceCube Software

The muon path reconstruction process (outlined in Figure 2) starts when the number of detected hits exceeds a preset threshold and the data collection step triggers. After the initial data is collected, it then passes though a series of simple filters to remove obvious outliers, described more in [13].

This is followed by a simple reconstruction algorithm *linefit*, which simply finds the track that minimizes the sum of the squares of the distances between the track and the hits. More formally, assume there are N hits, and denote the position and time of the ith hit as $\vec{x}_i$ and t_i respectively. Let the reconstructed muon path have a velocity of $\vec{v}$, and let the reconstructed path pass though point $\vec{x}_0$ at time t_0. Then linefit reconstruction solves the *least-squares* optimization problem

$$\min_{t_0, \vec{x}_0, \vec{v}} \sum_{i=1}^{N} \rho_i(t_0, \vec{x}_0, \vec{v})^2, \tag{1}$$

where

$$\rho_i(t_0, \vec{x}_0, \vec{v}) = \|\vec{v}(t_i - t_0) + \vec{x}_0 - \vec{x}_i\|_2 . \tag{2}$$

The linefit reconstruction is primarily used to generate an initial track or *seed* for a more sophisticated reconstruction.

The reconstruction algorithm used in the sophisticated reconstruction *SPE*, is described further in [3]. SPE takes as input the least-squares reconstruction and the event data, and uses a likelihood maximization algorithm to reconstruct the muon path.

3.2. Algorithm Improvement

SPE is dependent on the seed. Given a seed that is inaccurate by greater than or equal to 6°, SPE typically cannot recover, and also produces a reconstruction inaccurate by greater than or equal to 6°. In addition, the likelihood space for SPE can contain multiple local maxima, so improving the accuracy of a seed already near the true solution still improves the accuracy of SPE. Thus, we focused our work on improving the quality of the seed.

As indicated in Equation 1, a least-squares fit models the muon as a single point moving in a straight line, and hits are penalized quadratically in their distance from this line. Thus there is an implicit assumption in this model, which is that all the hits will be near the muon. There are several pitfalls in this assumption:

1. It ignores the scattering effects of the ice medium. Some of the photons can scatter for over a microsecond, which means that when they are recorded by a PMT, the muon will be over 300 m away.
2. While the noise reduction steps remove most of the outlier noise, the noise hits that survive can be far from the muon. Since these outliers are given quadratic weight, they exert a huge influence over the model.

The first pitfall is a case of the model being incomplete and not modeling the data, and the second amounts to the model not being robust to noise. Our solution was twofold: improve the model and increase the noise robustness by replacing least squares with robust statistical techniques.

3.2.1. Improving the Model

The least-squares model does not model the scattering, and thus hits generated by photons that scattered for a significant length of time are not useful predictors of the muon's position. We found that a simple filter could identify these scattered hits, and generate an accuracy improvement of almost a factor of two by removing them from the dataset.

More formally, for each hit h_i, the algorithm looks at all neighboring hits within a neighborhood of r, and if there exists a neighboring hit h_j with a time stamp that is t earlier than h_i, then h_i is considered a scattered hit, and is not used in the simple reconstruction algorithm. Optimal values of r and t were found to be 156 m and 778 ns by parameter search.

3.2.2. Adding Robustness to Noise

One of the fundamental problems with least squares is that outliers are given a quadratic influence, whereas an ideal model would give outliers zero influence. Such an ideal model does not exist, but classical statistics has developed models where outliers can be marginalized. We experimented replacing the least-squares model with a variety of more robust models: a deadzone-linear fit, a one-norm fit, and a Huber fit [14].

Of the models that we tested, the Huber penalty function gave the greatest increase in reconstruction accuracy. More formally, we replace Equation 1 with the optimization problem

$$\min_{t_0,\vec{x}_0,\vec{v}} \sum_{i=1}^{N} \phi(\rho_i(t_0,\vec{x}_0,\vec{v})), \tag{3}$$

where the Huber penalty function $\phi(\rho)$ is defined as

$$\phi(\rho) \equiv \left\{ \begin{array}{ll} \rho^2 & \text{if } \rho < \mu \\ \mu(2\rho - \mu) & \text{if } \rho \geq \mu \end{array} \right. . \tag{4}$$

Here, $\rho_i(t_0,\vec{x},\vec{v})$ is defined in Equation 2 and μ is a constant calibrated to the data (in our work, the optimal value of μ is 153 m).

The Huber penalty function has two regimes. In the near-hit regime ($\rho < \mu$) hits are assumed to be strongly correlated with the muon's path, and the Huber penalty function behaves like least squares, giving these hits quadratic weight. In the far-hit regime ($\rho \geq \mu$), hits are given linear weights as they are more likely to be noise.

In addition to its attractive robustness properties, the Huber fit's weight assignment also has the added benefit that it inherently labels points as outliers (those with $\rho \geq \mu$). Thus, once the Huber fit is computed, we can go one step

Table 1: Median angular resolution (degrees) for reconstruction improvements. The first line is the accuracy of the prior least-squares model, and the subsequent lines are the accuracy measurements from cumulatively adding improvements into the simple reconstruction algorithm.

Algorithm	θ_{med}
Linefit Reconstruction (Least-Squares)	9.917
With Addition of Logical Filter	5.205
With Addition of Huber Regression	4.672
With Addition of Outlier Removal	4.211

farther and simply remove the labeled outliers from the dataset. A better fit is then obtained by computing the least-squares fit on the data with the outliers removed.

3.3. Results

To measure the improvement generated by our changes, we use the metric of *median angular resolution* θ_{med}, which is a standard metric used in the collaboration. The angular resolution of a reconstruction is the arc-distance between the reconstruction and the true path. Removing the scattered hits and adding robustness to the model generates measurable a improvement to the model's accuracy, as shown in Table 1.

We can improve the median angular resolution of the simple reconstruction by 57.6%. Seeding SPE with the improved simple reconstruction generates an improvement in the angular resolution of 12.9%. These improvements in the reconstruction algorithm result in 10% fewer atmospheric muons erroneously reconstructed as up-going, and 1% more muons correctly reconstructed as up-going.

4. Coincident Event Problem

In our second experiment, we look at the problem of determining when more than one muon has entered the detector. In the most common case, a single muon will pass though the detector and generate an event before exiting. These events are processed by the pipeline described in Figure 2. However, for roughly 9% of the events collected by the data collection algorithm, more than one muon will be passing though the detector simultaneously, an occurrence known as a *coincident event*.

One of the primary sources of background noise in the scientific analyses of the IceCube Collaboration is coincident background muons that have been erroneously reconstructed as neutrinos. To see why this occurs, consider the coincident event shown in Figure 3. There are two clear groups of hits; however, the reconstruction algorithm treats them as a single group, resulting in a erroneous reconstruction. In the ideal case, the reconstruction algorithm would identify coincident events and split them, as in Figure 4.

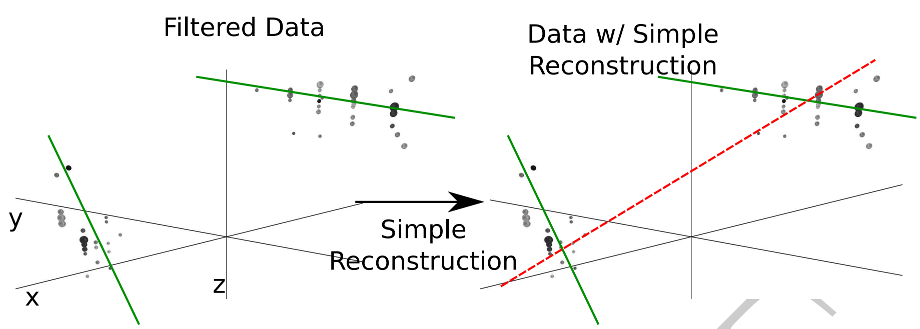

Figure 3: In this example, an event that is clearly composed of two muons (actual tracks shown as green thick lines) is treated as a single muon, and thus the reconstruction (shown in dashed red) is inaccurate.

The challenge in this example is determining the number of muons in an event. In our results, we find that a simple spatial clustering algorithm can solve this classification problem with less than 2% error.

4.1. Prior IceCube Software

Coincident events have been a concern in the IceCube analysis [15] for years, and some software has been developed to handle coincident events. As a baseline of comparison, we use the *TTrigger* software, which is described in [16].

4.2. Algorithm Improvement

Our solution to this problem is a proximal clustering algorithm. The intuition in proximal clustering is that points local in space and time are probably from the same muon. The proximal clustering algorithm iterates though each pair of hits (i, j) and builds an adjacency matrix $\mathbf{A}$ as

$$\mathbf{A}_{ij} = \begin{cases} 1 & \text{if } \|\Delta x^2 + \Delta y^2 + \Delta z^2 + (c\Delta t)^2\|_2 \leq \alpha, \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where $\Delta x, \Delta y, \Delta z$ and Δt are the space and time differences between the pair of hits, and α is tuned to the data. The clustering can be recovered by extracting the connected components of the graph defined by $\mathbf{A}$.

4.2.1. Improving the Model

When implemented naively, proximal clustering succeeded for the majority of the events, but failed if there was a gap in the muon track, which can occur when the muon travels though dusty ice. If there is a significantly large gap, the algorithm erroneously separates the hits into two clusters.

To get around this, an additional heuristic is added, *track connecting*. After the data segmentation is finished, track connecting determines if separate clusters should be combined. It computes the mean position and time of each

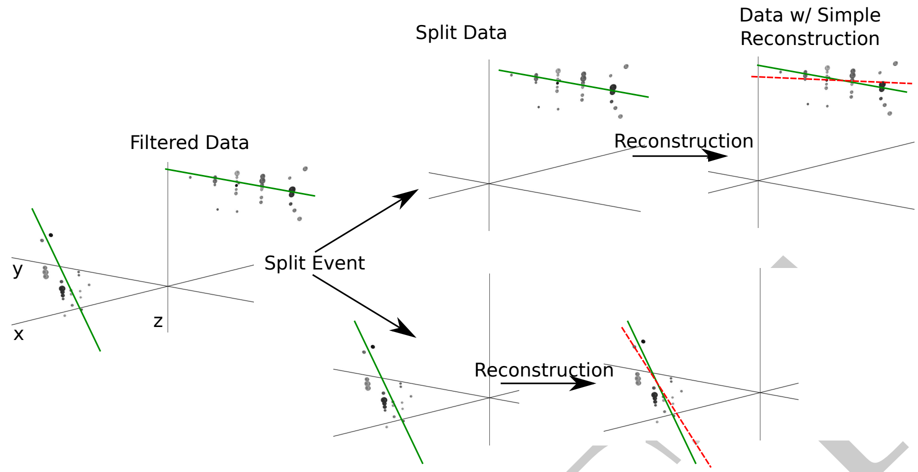

Figure 4: Ideally, the detector would split coincident events before computing the reconstruction. Splitting the event results in more accurate reconstructions (reconstructions shown in red, true muon tracks shown in green). Note the difference in the reconstructions compared with Figure 3.

cluster, and connects a hypothetical muon track t between each pair of subspaces.

It checks if the speed s of the hypothetical path is within 25% of the speed of light c, and it checks that the mean distance between hits and t in both clusters is less than 60 m. If t passes both checks, the clusters are combined.

4.2.2. Adding Robustness to Noise

Proximal clustering is susceptible to noise. Noise hits close to two disjoint tracks will be considered adjacent to both tracks, connecting the two tracks in the adjacency matrix.

One heuristic that worked well at mitigating this problem was to use all the hits in building the adjacency matrix. During data collection, some hits are marked as coincident, which indicates that both they and a neighboring PMT reported a hit. These hits have a high probability of not being noise hits, and thus exclusively using them to build the adjacency matrix mitigates the problem of erroneously connecting two tracks.

After the proximal clustering algorithm has extracted the tracks from the adjacency matrix, the hits not used in the construction of the adjacency matrix are simply assigned to the closest reconstructed track.

4.3. Results

There were two competing goals for coincident event detection algorithms: the algorithm should be conservative enough that events containing single paths are not erroneously split, and aggressive enough that a useful fraction of coincident events are split correctly. Erroneously discarding events containing neutrinos is worse than erroneously allowing additional noise into the data pool,

Table 2: Error Rates for Classification Algorithms

Algorithm	E_{Single} %	E_{Multiple}%	E_{tot} %
Trivial	0.0	100.0	8.3
TTrigger	11.5	31.8	13.2
Proximal clustering	0.2	18.9	1.8

as noise can be eliminated by future filtering of the data pool. Our algorithm is tuned to keep almost all of the single events correctly unsplit, while still correctly splitting 80% of the coincident events.

4.3.1. Measurements

We modified the reconstruction pipeline shown in Figure 2, in between the noise cleaning and the simple reconstruction, by adding a step for coincident event detection, as shown in Figure 4. This step takes cleaned data and attempts to classify the event as a single-track or multiple-track event.

We ran each algorithm on two datasets of simulated data. One dataset comprised single-muon events, and the other dataset comprised multiple-muon events. In each dataset, we measured the classification error E, which is the faction of events that were misclassified. To get a global measurement, we compute the *total error* E_{tot}, defined as

$$E_{tot} = w_{\text{Single}} E_{\text{Single}} + w_{\text{Multiple}} E_{\text{Multiple}}. \tag{6}$$

For computing E_{tot}, we use $w_{\text{Single}} = 0.917$ and $w_{\text{Multiple}} = 0.083$, which is the frequency in which single-muon and multiple-muon events appear in data simulating the distribution of events that trigger the reconstruction algorithm.

We present our results for the coincident event problem by measuring how well each algorithm performs at determining the number of subspaces in an event.

There are two natural comparisons for our work: the prior software TTrigger, as well as the trivial algorithm, which always classifies each event as a single-track event. Clearly, the latter will always get the single-track events correct, and always get the multiple-track events wrong. We provide a comparison of these techniques in Table 2. As shown, our software classifies the number of muons in the detector 86% better than TTrigger.

5. Conclusions

The problems in the IceCube detector are complex, and the data is noisy and contaminated with outliers. Despite the complexity of the problems, we found that we can achieve significant improvement via classical data analysis algorithms and simple models.

We looked at the problem of general reconstruction improvement, and found that by applying a simple filter to the data and adding some robustness to the fitting algorithm, we got superior reconstructions in the noisy environments of the IceCube data. Our reconstruction software runs on-site, and is included in all IceCube analysis.

We also looked at the problem of determining the number of muons in the detector. We found that proximal clustering, the simplest algorithm that we tried, was as good as or better than all other tested algorithms. Our proximal clustering algorithm was an 86% improvement over the current software.

References

[1] IceCube Collaboration, IceCube webpage, http://icecube.wisc.edu/.

[2] IceCube Collaboration, First year performance of the IceCube neutrino telescope, Astroparticle Physics 26 (3) (2006) 155–173.

[3] IceCube Collaboration, Muon Track Reconstruction and Data Selection Techniques in AMANDA, Nuclear Instruments and Methods in Physics Research Section A 524 (2004) 169–194.

[4] IceCube Collaboration, Measurement of South Pole ice transparency with the IceCube LED calibration system IceCube Collaboration, Nuclear Instruments and Methods in Physics Research Section A.

[5] ATLAS Collaboration, Tracking and vertexing with the ATLAS detector at the LHC, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 650 (1) (2011) 218–223.

[6] R. S. Chivukulaa, M. Goldena, E. H. Simmons, Multi-jet physics at hadron colliders, Nuclear Physics B 363 (1) (1991) 83–96.

[7] S. Ellis, J. Huston, K. Hatakeyama, P. Loch, M. Tönnesmann, Jets in hadron–hadron collisions, Progress in Particle and Nuclear Physics (60) (2008) 484–551.

[8] IceCube Collaboration, Calibration and characterization of the IceCube photomultiplier tube, Nuclear Instruments and Methods in Physics Research Section A 618 (2010) 139–152.

[9] IceCube Collaboration, An improved method for measuring muon energy using the truncated mean of dE/dx, Nuclear Instruments and Methods in Physics Research Section A.

[10] IceCube Collaboration, The icecube data acquisition system: Signal capture, digitization, and timestamping, Nuclear Instruments and Methods in Physics Research Section A 601 (3) (2009) 294–316.

[11] M. Wolf, E. Resconi, Verification of South Pole glacial ice simulations in IceCube and its relation to conventional and new, accelerated photon tracking techniques, Master's thesis, Max-Planck-Institut für Kernphysik Heidelberg (September 2010).

[12] IceCube Collaboration, IceCube sensitivity for low-energy neutrinos from nearby supernovae, Astronomy & Astrophysics 535 (A109) (2011) 18.

[13] M. Ackermann, Searches for signals from cosmic point-like sources of high energy neutrinos in 5 years of AMANDA-II data, Ph.D. thesis, Humboldt-Universität zu Berlin (2006).

[14] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2009.

[15] IceCube Collaboration, Measurement of the atmospheric neutrino energy spectrum from 100 GeV to 400 TeV with IceCube, Physical Review D 83 (1).

[16] D. Chirkin, Measurement of the atmospheric neutrino energy spectrum with IceCube, Proceedings of the 31st ICRC.