33RD INTERNATIONAL COSMIC RAY CONFERENCE, RIO DE JANEIRO 2013
THE ASTROPARTICLE PHYSICS CONFERENCE

ICRC 2013

Event reconstruction in IceCube based on direct event re-simulation

THE ICECUBE COLLABORATION[1]

[1]*See special section in these proceedings*

dima@icecube.wisc.edu

Abstract: We present a new method of event reconstruction in IceCube that searches through an extensive set of re-simulated events to find the best match to data. Several search methods, including the stochastic approximation and localized random search, were tried. We demonstrate the capability of the new method to reconstruct cascade-like as well as track-like events. While significantly slower than the parametrized probability distribution function (PDF) -based reconstructions, the new method is able to take into account all of the nuances present in the simulation that are sometimes omitted from the PDF-based description of data.

Corresponding author: Dmitry Chirkin[1]
[1] *Dept. of Physics and WIPAC, University of Wisconsin, Madison, WI 53706, USA*

Keywords: cascade reconstruction, track reconstruction, direct re-simulation.

1 Introduction

IceCube is a cubic-kilometer neutrino detector installed in the ice at the geographic South Pole [1] between depths of 1450 m and 2450 m. Detector construction started in 2005 and finished in 2010. Event reconstruction relies on the optical detection of Cherenkov radiation emitted by secondary particles produced in neutrino interactions in the surrounding ice or the nearby bedrock.

The development of the method described in this paper followed after we recognized that ice surrounding the detector does not have the azimuthal symmetry previously ascribed to it (see [2], [3]) and still used in our PDF-based reconstructions. While it is relatively easy to add a description of almost any conceivable detail to the detector simulation, adding even one new dimension to the existing probability tables is prohibitive, from the perspective of the computing time necessary to create the PDF tables, and storage of and access to the stored tables (which becomes slow if the tables cannot be fitted into the active computer memory).

In section 2 we discuss reconstruction of cascades, i.e., relatively short particle showers, which happen when neutrinos interact via the neutral current process, or when ν_e, $\overline{\nu}_e$ and ν_τ, $\overline{\nu}_\tau$ interact via the charged current process. In these cases all visible particles produced in the interaction are short-lived, either decaying or losing energy very quickly, producing an initial light pattern that can be often approximated as emanating from a point, when compared to the large scale of the IceCube detector.

In contrast, the charged current interactions of ν_μ, $\overline{\nu}_\mu$ create a muon (in addition to the initial cascade), which can propagate a great distance away from the interaction point, leaving behind a characteristic "track"-like pattern of light. We call these events "tracks" and discuss the application of our method to the track-like events in section 3.

2 Cascade reconstruction

We start by describing the cascade reconstruction, which purpose is to ascertain the parameters of the cascade-like event observed in data: energy (E), time and position (t_0, x, y, z), and direction (θ, ϕ). The algorithm works by simulating many cascades with various sets of parameters and choosing the one (or a few) that look most like data.

The reconstruction is started from a first guess with coordinates of the center of gravity of hits in the detector, a random direction, the energy of $E = 10^5$ GeV, and the start time of $t_0 = 0$ ns. In the localized random search version of the algorithm we specify the widths of the proposal distributions for space coordinates (Δr=10 m) and angular coordinates ($\Delta\Psi$=30 degrees). The proposal distributions are the 3-d Gaussian and 2-d Gaussian on a sphere (von-Mieses-Fisher distribution).

We compare the time-binned photo-electron charges registered by optical sensors between data and simulation with a likelihood function described in [4], which takes into account statistical fluctuations in both data and simulation, as well as a possible systematic disagreement between data and simulation, which is modeled with a log-normal Gaussian (describing the "model error" of the ice model, as explained in [2] and reduced to 20% in [3]). At each iteration 25 possible sets of space and angular coordinates (x, y, z, θ, ϕ) are sampled from the proposal distribution around the best cascade from the previous iteration (or the initial guess, if this is the first iteration). For each such set the photon simulation of [6] is run with the best values of E and t_0 from the previous iteration, producing Monte-Carlo hits. Then the best values of energy E and start time t_0 that maximize the likelihood function are found. This is done by varying the (possibly fractional) "effective" number of events represented by the simulation (n_s in [4]) and shifting the simulation around in time. Finally, each set is re-simulated with the optimized values of E and t_0. The final likelihood values for all 25 sets are computed and the event with the best value is selected. This best value and corresponding cascade parameters are saved and used as a starting point for the next iteration.

This process is repeated 400 times. Every 40 iterations the proposal distribution width parameters Δr and $\Delta\Psi$ are updated to the RMS values of space and angular coordinates calculated from the best 20 of the previous 40

iterations. Finally, the solution is calculated by averaging the best 160 iteration results. Some measure of statistical uncertainties is given by the spread of parameters of the best 160 cascades, although a strict statistical meaning of this spread is unclear. The specific numbers listed above (25 sets per iteration, 400 iterations, initial parameters of the proposal distribution) resulted in a robust, yet relatively fast reconstruction.

In addition to the "localized random search" method described above we have tried two more approaches, which yielded somewhat less robust performance. The first of these is the simultaneous perturbation stochastic approximation [5] and its variations, which is a derivative-based maximization search method, where the derivative is calculated "stochastically" from consecutive iterations, but the step size is reduced in a controlled manner. The second method was inspired by the Markov chain Monte-Carlo method; here each proposed step is either taken or rejected based on whether the calculated value of the likelihood at the new coordinates is better or worse than at the current point. Since the likelihood value at given coordinates is calculated from a simulation performed anew for each new evaluation, it is a "stochastic" quantity, forming a distribution, rather than assuming a single value. So, it is often possible to jump from a point with parameters providing a (on average) better description of data to a point with parameters providing a (on average) worse description of the data, just like in a typical Markov chain Monte-Carlo. Although the iterations of the method do, in fact, form a Markov chain, this Markov chain is not reversible, with no obvious probabilistic interpretation of the resulting sequence of iterations.

3 Track reconstruction

The algorithm described in the previous section also works for track reconstruction with a following modification. For each proposed set of space and angular coordinates, defining a track (x, y, z, θ, ϕ), small cascades of equal energy are simulated every 7.5 meters along the part of the track that is inside the detector with a fixed combined energy of 10^6 GeV. Each such cascade (numbered with index j) produces a pattern of hits in the detector, creating a column in a matrix A_{ij}, where i represents a time bin in a DOM (A_{ij} being the photo-electron charge in that bin left by the cascade j).

We now have the unfolding problem similar to that discussed in [7], where the energy of each cascade j needs to be scaled by "weights" w_j so that the sum $\sum_j A_{ij} w_j$ best matches the hit counts in data d_i. This problem is solved taking into account the Poisson nature of both data and simulation sets of each cascade (as described in [4]), resulting in the set of unfolded simulation weights w_i. The first guess is obtained by solving the $\sum_j A_{ij} w_j = d_i$ with a non-negative least squares (NNLS) algorithm, and the solution is then refined for the full likelihood description with a derivative method (BFGS2 of GSL, [8]).

The unfolding procedure is repeated for various values of t_0, and the best value of t_0 is chosen to maximize the likelihood comparing the data with a superposition of cascades weighted with the unfolded weights w_j determined for that t_0. The unfolded pattern of weights corresponding to the best value of t_0 is applied to the energies of cascades along the track, and the event is re-simulated.

The procedure described above is optionally repeated one more time starting with small cascades of energies proportional to w_j (but still with a total sum of 10^6 GeV), adding simulation to the pool created during the first pass (i.e., simulation sets from both passes are used for the energy loss unfolding part of the calculation). The final pattern of cascade energies is re-simulated one more time and the final likelihood value for this iteration is computed.

The entire calculation described in this section is performed for each of the 25 trial sets of each of the 400 iterations of section 2. Fig. 4 demonstrates the convergence behavior of the method for a track event.

4 Estimating uncertainties

Statistical uncertainties can be estimated with an Approximate Bayesian Computation (ABC) method as in [9], [10].

The idea is to substitute the probability of data D given event parameters θ, $p(D|\theta)$, in the Bayes expression for the conditional probability density

$$p(\theta|D) = \frac{p(D|\theta) \cdot p(\theta)}{p(D)},$$

with a probability to obtain data events that are sufficiently similar to the event under investigation. The similarity is defined with a "distance" between the other possible data events $\hat{D}$ and our data event D

$$\rho(\hat{D}, D) \leq \varepsilon,$$

which we substitute with the minus log likelihood expression of [4], comparing the time-binned photo-electron charges between data and simulation.

Since the prior of the track parameters x, y, z, θ, ϕ can be often assumed to be flat in 3-d space of coordinates and on a 2-d sphere of directions (as neutrinos interact uniformly anywhere inside the detector), the sampling method of choice is a Markov Chain (rather than simple rejection), where steps are sampled from a proposal distribution and accepted when the "distance" condition is satisfied. For more details of this procedure (including a proof), see [9]. The parameters of the proposal distribution and ε in the "distance" expression are estimated from the best steps of the minimization procedure of sections 2, 3.

Estimating statistical uncertainties using this method takes approximately the same amount of time as the reconstruction (two days on a specially-configured computer). It is unclear if adding nuisance parameters to this procedure to additionally account for the systematic uncertainties is realistic as the step acceptance rate is lowered with each additional parameter (leading to a "curse of dimensionality").

5 Reconstruction Examples

The method of this report was applied to the 28 high energy events of [11]. Figure 1 compares the reconstructed energy and zenith with the main reconstruction method used in [11]. Out of the 28 events only one reconstructed as a track, with many reconstructed losses along its path (as shown in figure 2, also see figure 3). In the other 6 events that

have a muon in addition to the larger cascade the muon part of the event is largely ignored by this reconstruction. We believe this happens when the contribution to the likelihood from the few hits left by the muon is so small that it is largely "washed out" by the fluctuations of the much larger contribution from the cascade part of the event. Finally, figure 4 illustrates how the method converges to the solution.

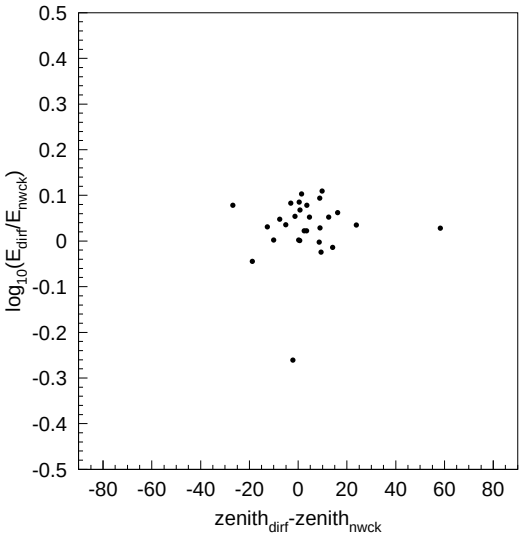


Fig. 1: Comparison between the results of the reconstruction method described in this report with the main reconstruction used in [11] for the 28 HE events of [11]. Most of the points are within the average estimated uncertainties of 10% in energy and 10 degrees in the zenith angle. The energy of the outlier is overestimated by the reconstruction method of [11], since it did not include the effect of the ice layer tilt (change in depth of a given ice layer at different x, y across the detector, see [2]) in the PDF used by the reconstruction, whereas the method of this report used our best description of the ice, including the tilt. This particular event was affected more than others because it lay at depth where the ice properties are quickly changing (just above the main dust layer, [2]), and far from the ice tabulated position in (x, y) at (0, 0), removed from it almost along the line of largest tilt (gradient direction of the ice layer elevation). The uncertainties for this event were estimated in [11] by reconstructing events similar to this one, simulated with the best knowledge of ice. The median of the energy uncertainty range was found to be a good agreement with the energy reconstructed with the method of this report. The outlier in zenith is for a cascade event with a particularly poor reconstructed angular resolution.

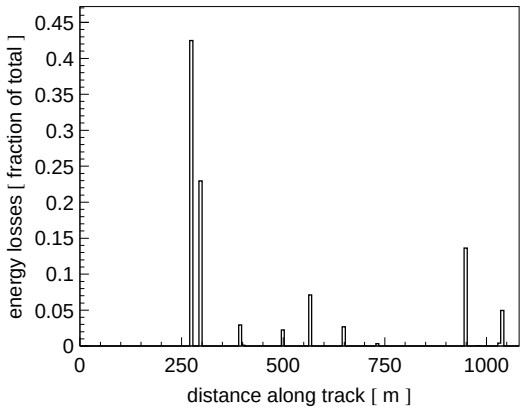


Fig. 2: Reconstructed energy losses along the best-fit track of a track-like event shown in figure 3 (one of the 28 high energy events of [11]). The losses are normalized to the total energy lost in the volume of the detector.

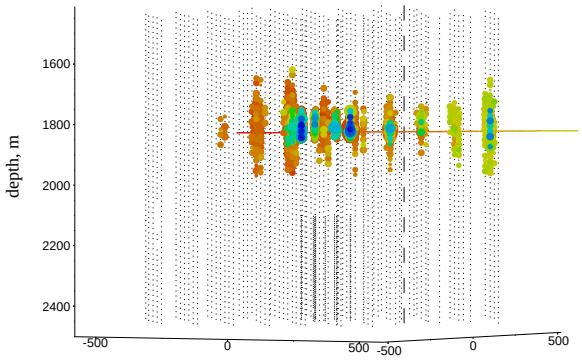


Fig. 3: Illustration of a track-like event, one of the 28 high energy events of [11].

References

[1] A. Achterberg et al., Astropart. Phys. 26, 155 (2006).
[2] M. Aartsen et al., NIM-A (2013) p. 73, arXiv:1301.5361.
[3] D. Chirkin et al., these proceedings.
[4] D. Chirkin, arXiv:1304.0735 (submitted to NIM-A).
[5] http://en.wikipedia.org/wiki/Simultaneous_perturbation_stochastic_approximation
[6] D. Chirkin, DOI: 10.1016/j.nima.2012.11.170.
[7] N. Whitehorn, awesome millipede reference.
[8] GNU Scientific Library, http://www.gnu.org.
[9] P. Marjoram et al., PNAS (2003) vol. 100 no. 26, p. 15324.
[10] E. Cameron, A. Pettitt, MNRAS, vol. 425, issue 1, p. 44.
[11] N. Whitehorn, C. Kopper, these proceedings.

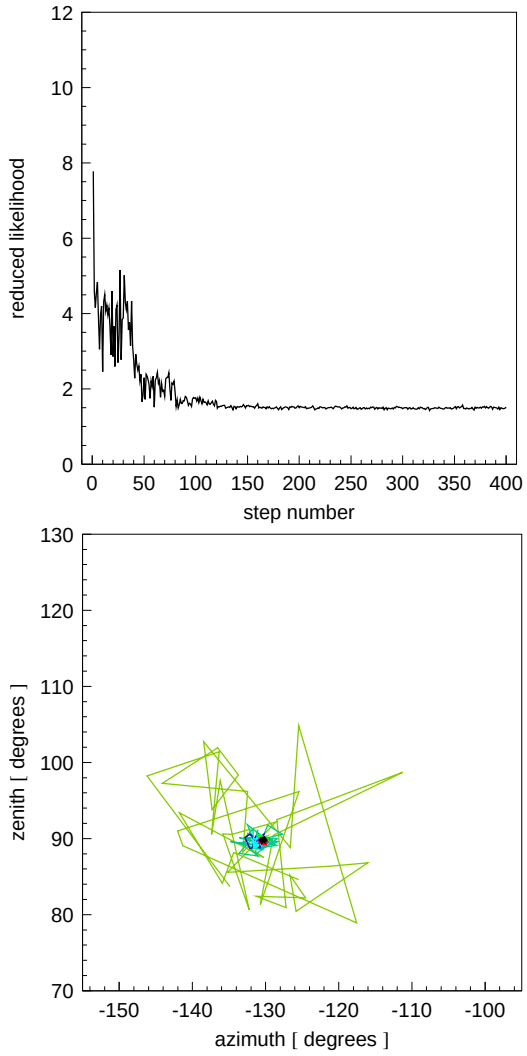


Fig. 4: Top: Reduced log likelihood (log likelihood divided by the number of degrees of freedom) after each of the 400 iterations. Some stepping structure evident in the plot is related to the fact that the parameters of the proposal distribution are updated every 40 steps. Bottom: Convergence of the best fit direction: results of successive iterations are connected by line segments for visual clarity. This is a reconstruction of a well-defined track, and results in the method wandering around in a very narrow range of zenith and azimuth. This reconstruction took about 2 days on a computer with 5 GPUs. Most of the calculation time was spent on the GPUs, which were used to run the photon propagation simulation.