A new method for identifying neutrino events in IceCube data

Dmitry Chirkin* **for the IceCube collaboration**

University of Wisconsin, Madison, U.S.A.

Abstract**. A novel approach for selecting high-quality muon neutrino events in IceCube data is presented. The rate of air shower events mis-reconstructed as interesting signal is first reduced via the use of the geometrical (software) trigger. The final event selection is performed with a machine-learning method, designed specifically for IceCube data. It takes into account some generic properties of IceCube events, e.g., the fact that separation of signal from background is more difficult (i.e., the cuts on the quality parameters must be tightened) for horizontal rather than vertically up-going tracks). The method is very robust and cannot be over-trained by construction. Its performance is compared to other techniques.**

Keywords**: neutrino search, machine learning, event selection**

I. INTRODUCTION

Analysis of the AMANDA-II data resulted in limits placed on the extraterrestrial fluxes of neutrinos and exotic particles. A variety of search methods were developed and resulted in a number of techniques that are now being applied anew to the analysis of the first few years of the IceCube data.

A common first step of these efforts is achieving a high purity atmospheric (plus possible extraterrestrial) neutrino event sample, with only a small contaminating fraction of mis-reconstructed atmospheric muon events. Only neutrino events could have directions up, having crossed much of Earth overburden; however, selecting events with reconstructed direction up leaves many of the mis-reconstructed atmospheric muon events, improving the ratio of neutrino to contaminating muon events (initially at $\sim 10^{-6}$) by only a factor of ~ 100.

The problem is further exacerbated by a highly uneven contamination of the mis-reconstructed muons in several of the analysis variables, most importantly the zenith angle. This contamination is smaller for up-going directions and increases for more horizontal tracks, growing fast near and above the horizon. It is therefore difficult to arrive at an event selection method that provides cut surfaces simultaneously for all zenith angles, leading several authors to split optimization of cuts into several disjointed problems in different zenith angle bins. This approach leads to fluctuations of the cut parameters from one zenith angle bin to the next that are perceived as unphysical.

Additionally, in situations with limited simulated data splitting it in several zenith angle bins was undesirable.

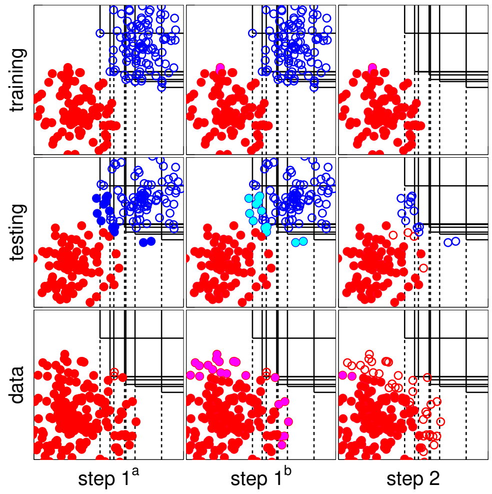

Fig. 1. Simple example: lower values on both x and y mean better quality. Two event populations are shown: red is signal for the training and testing simulated event sets, and all data in the data set. Blue points (located up and to the right from red points) describe background events. In steps 1^a and 1^b empty circles show background events removed by the *skeleton cuts*, aqua and pink points show background and signal (or data) events respectively that are removed by the machine quality parameter cut set at 1.5. At step 2 events removed by step 2 are shown as empty circles and events additionally removed by the quality parameter cut are shown in pink, same as before. The black lines show the *skeleton cuts* and go through the out-most background events of the training set.

This author has also performed the SVM-based event selection optimization and found that the difficulty with training the SVM increases as angles just above the horizon are included into the consideration eventually leading to some drop in performance for wider zenith angle ranges.

The above considerations led to development of a new framework for selecting and applying cuts on quality parameters, that in the following is called "Subset Browsing Method", or *SBM* for short.

The quality parameters used with the event selection method of this paper include and build upon those discussed previously [1].

II. SIMPLE EXAMPLE

First consider a simple example employing only two parameters describing events that are, in some sense, better, as the value of the quality parameter decreases. These can be, e.g., zenith angle (0 degrees for up-going to 90 for horizontal tracks) and the estimated angular resolution (e.g., describing the half-width of the likelihood function at the minimum corresponding to the reconstructed track direction). Both of these can be used

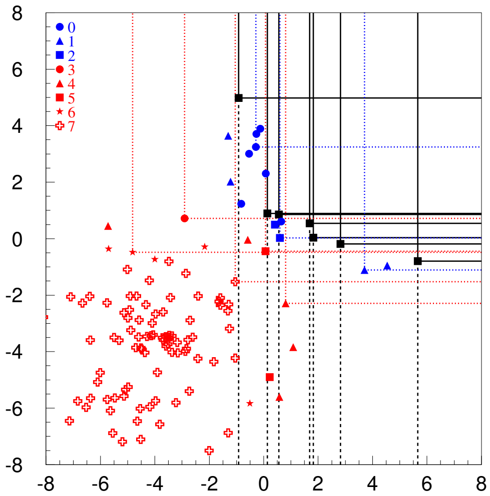

Fig. 2. Shown are events remaining in the testing data set after the application of step 1^a, signal in red, and background in blue. The legend indicates the value of the quality parameter for the shown events. For each value of the quality parameter the bad sides of a representative event are shown with straight lines. The quality parameter is simply the count of the "outlying" background events of the training data set that gave rise to the "angle cuts" of step 1^a (indicated with black squared and black solid lines.

to remove the background of mis-reconstructed events, one through the basic reconstruction property, and the other though our prior knowledge that the contamination is higher for tracks near the horizon.

The toy simulated events are divided into two groups (randomly): the training set that is used for the training of the machine, and the testing set, that is used to judge its performance. Both sets, while drawn from the same parent distribution, are statistically independent. The toy "data" events are also simulated and drawn from a somewhat wider signal distribution. The 3 steps of the machine application are the steps 1^a, 1^b, and 2 as shown on Figure 1.

The training of the machine in this simple example is achieved by identifying the "outlying" background events (on the signal side of the distribution), and creating the "angle cuts" (shown with black straight lines) by cutting away everything on the bad sides of such cuts (i.e., everything above and to the right of the background event, including that background event itself).

The cuts so identified will obviously remove all background events in the training data set. As seen from the second row of the Figure 1, these cuts do not remove all of the background events when applied to the testing data set, so a further step, here called 1^b is necessary. Using the angle cuts derived in step 1^a a quality parameter (*SBN*) is constructed, which is simply the count of "angle cuts" of step 1^a that fall into the bad quadrant (up and to the right) of a tested event, see Figure 2. This quality parameter could also be constructed as a weighted sum, as described in the following section, shown for comparison as *SBM* in Figure 4.

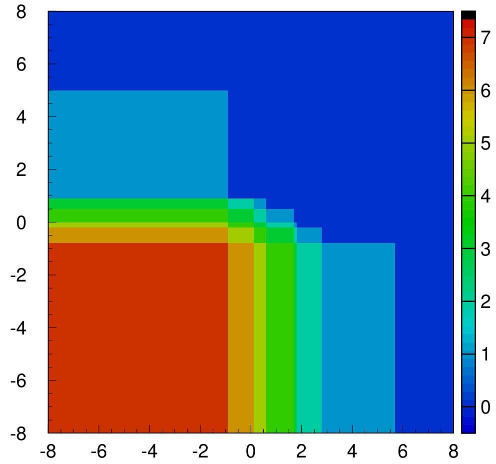

Fig. 3. Map of the quality parameter (*SBN*)

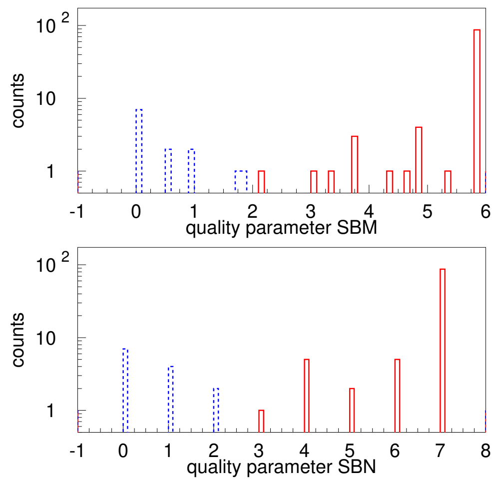

Fig. 4. quality parameters: simple sum over the "angle cuts" (*SBN*), and weighted sum (*SBM*). Red solid and blue dotted lines show the distribution of signal and background events, respectively.

A map of the quality parameter (*SBN*) is shown in Figure 3. It is clear that through application of the quality parameter the initial "jaggedness" of the cut structure achieved in step 1^a (due to statistical fluctuations in the positions of the outlying background events) is much smoothed. Figure 4 shows that a value of *SBN*=2.5 completely separates signal from background in the testing data set of this simple example.

III. Unsimulated Events and Step 2

After the application of steps 1^a and 1^b to the real IceCube data it became obvious that, although much of the background events like those present in our simulation was removed, some "unsimulated" background remained in data. This affected agreement in variable distributions and was particularly evident upon visual inspection. One class of such events appeared to contain two or more coincident but independent muon hit patterns that happened near each other with much overlap in time and failed to be split by the topological trigger. These events will likely have been eventually simulated if more simulation was performed.

Another class of events appeared to contain a bright EM or hadron shower (*cascade*), with rate of occurrence much higher than that predicted by the simulation. While these events, when understood, could be very interesting, making a valuable contribution to the final event selection, it is unclear at this point whether they should be classified as signal or background. Moreover, since the are not simulated as either, the detector effective area to these events cannot be estimated, so they do not contribute to any of the physics results.

The authors electing to use other machine learning (and simple cut) methods deal with this by raising the quality parameter of their respective analysis past the point where all simulated background is removed and to the point where an agreement between data and signal simulation is achieved. This is possible because the quality parameter judges not only how far a given event is from the background region, but also how close it is to the signal region.

In the method presented here the quality parameter is constructed using only the information about the outlying background events that form the "angle cuts" (after some initial amount of signal events is carved out by the step 1^a). Thus, the quality parameter judges only the distance to the background region (contrary to other approaches). To achieve the "similarity with signal events" another step (called step 2) becomes necessary.

This event selection step is achieved (after the application of steps 1^a and 1^b) by removing all events that, were they used as "angle cuts" would not have affected the remaining signal events present in the training set. The effect of step 2 is demonstrated in the last column of Figure 1. While this author believes step 2 is necessary in the presence of the unsimulated background (as much as tightening of the quality parameter is necessary in other approaches), its application to the current IceCube data remains under much discussion at the time of writing of this proceeding.

IV. Multi-dimensional generalization

First we re-iterate that the SBM method relies on the important observation that most of the quality parameters used in the analysis of IceCube data have the following property: as the fits become less constrained at lower number of channels N_{ch} or strings N_{str} the cuts on the quality parameters necessary to reach a given signal purity need to be tightened. Alternatively the cuts applied to quality parameters of events with higher N_{ch} or N_{str} can be relaxed somewhat. A similar behavior of cuts on quality parameters can be argued for their dependence on the reconstructed zenith angle θ: at angles closer to the horizon the number of background events seeping through is higher than for tracks going up closer to the vertical, so to reach the same purity the cuts on the quality parameters need to be tighter for events with higher reconstructed zenith angle θ. To summarize, we introduce the following

Basic cut property: the cuts necessary to reach the same signal purity satisfy the following conditions:

$$c(\theta^\star, N_{ch}, N_{str}) \le c(\theta^0, N_{ch}, N_{str}) \text{ for } \theta^\star \ge \theta^0$$

$$c(\theta, N_{ch}^\star, N_{str}) \le c(\theta, N_{ch}^0, N_{str}) \text{ for } N_{ch}^\star \le N_{ch}^0$$

$$c(\theta, N_{ch}, N_{str}^\star) \le c(\theta, N_{ch}, N_{str}^0) \text{ for } N_{str}^\star \le N_{str}^0$$

This relies on the **assumption** that lower cut values imply tighter cuts. Some of the quality parameters may need to be taken with a minus sign or as one over their value to satisfy this assumption. Parameters θ, N_{ch}, and N_{str} that allow such a behavior of cuts are in the following called *basic cut variables*. The following discussion is simplified with a

Definition: a cut $c^\star$ defined for a set of events with $\theta^\star$, $N_{ch}^\star$, and $N_{str}^\star$ is said to be operating on a *subset* of events of a cut c^0 defined for a set of events with θ^0, N_{ch}^0, and N_{str}^0 if $\theta^\star \ge \theta^0$, $N_{ch}^\star \le N_{ch}^0$, and $N_{str}^\star \le N_{str}^0$.

Main cut property: a cut operating on a given set of events also operates on all its subsets.

To rephrase, a cut c^0 defined for a set of events with θ^0, N_{ch}^0, and N_{str}^0 also applies to any set of events with $\theta^\star$, $N_{ch}^\star$, and $N_{str}^\star$ (that has its own cut $c^\star$), if $\theta^\star \ge \theta^0$, $N_{ch}^\star \le N_{ch}^0$, and $N_{str}^\star \le N_{str}^0$. To prove we need to show that $c^0 \ge c^\star$. Using 2 intermediate sets of events, and the basic cut property introduced above

$$c^0 = c(\theta^0, N_{ch}^0, N_{str}^0) \ge c(\theta^\star, N_{ch}^0, N_{str}^0) \ge$$
$$c(\theta^\star, N_{ch}^\star, N_{str}^0) \ge c(\theta^\star, N_{ch}^\star, N_{str}^\star) = c^\star$$

This property allows us to consider all cuts as operating not only on events with $\theta = \theta^0$, $N_{ch} = N_{ch}^0$, and $N_{str} = N_{str}^0$, but rather on all events with $\theta \ge \theta^0$, $N_{ch} \le N_{ch}^0$, and $N_{str} \le N_{str}^0$.

Definition: The cut c^0 associated with θ^0, N_{ch}^0, and N_{str}^0 is considered *redundant* if there exists another cut $c^\star$ associated with some other $\theta^\star$, $N_{ch}^\star$, and $N^\star$ such that

$$c^\star \le c^0,\ \theta^\star \le \theta^0,\ N_{ch}^\star \ge N_{ch}^0, \text{ and } N_{str}^\star \ge N_{str}^0.$$

This is because the new cut $c^\star$ clearly implies c^0 by the main cut property.

For each background event i_b in the training simulated data set its quality parameters are used to create n cuts associated with θ^{i_b}, $N_{ch}^{i_b}$, and $N_{str}^{i_b}$ of the event. The signal purity $p^{i_b} = s^{i_b}/(s^{i_b} + b^{i_b})$ of the events with $\theta \ge \theta^{i_b}$, $N_{ch} \le N_{ch}^{i_b}$, and $N_{str} \le N_{str}^{i_b}$ is then calculated and used to find the i_b that defines cuts in a region with the worst purity. Out of the n cuts associated with i_b the cut that results in a smallest loss of signal events is then chosen and applied to the whole subset on which this cut operates. To accelerate this process if a cut in encountered that removes no signal events it is immediately used without taking into account the purity of the subset of events on which this cut operates.

This procedure is then repeated until the background events in the training simulated data set are exhausted. At that point all cuts of all background events are cycled through once again, and those that result in no further loss of remaining signal events (which are said to form a *core* or signal events) are saved into the *trained cut set* of the machine. One can further *reduce* this set by removing the *redundant* cuts from it, thus resulting in an *irreducible trained cut set*, which is the result of this machine training procedure.

One can obviously remove all background events (i.e., reach a 100% signal purity) by applying all cuts from the *irreducible trained cut set* to the training simulated dataset. However, when the same is applied to the separately generated *testing* simulated dataset a number of background events seep through and the signal purity never reaches 100% (or even $10-20\%$ in our tests).

This may happen, e.g., if we encounter a background event with, say, $N^\star_{ch}$ higher than $N^{i_b}_{ch}$ of every background event in the training simulated data set, thus there are no cuts available that would remove such an event from the testing data set.

A way around this is to find at least one cut c with $\theta^c \geq \theta^\star$, $N^c_{ch} \leq N^\star_{ch}$, and $N^c_{str} \leq N^\star_{str}$ such that $c^\star = q^\star \leq c$. By the *main cut property* the cut that would achieve the same purity p_c on the subset defined by $\theta^\star$, $N^\star_{ch}$, and $N^\star_{str}$ as the cut c on the subset defined by θ^c, N^c_{ch}, and N^c_{str} is necessarily no more tight as the cut c. That is, applying the cut c on the subset defined by $\theta^\star$, $N^\star_{ch}$, and $N^\star_{str}$ achieves at least the same or higher level of purity as p_c. Now, if an event defined by its quality parameters $c^\star = q^\star$ but at the *basic cut variables* θ^c, N^c_{ch}, and N^c_{str} of the cut c is passed by the *trained cut set*, cut c is called the *purity cut* defined for the original event (defined by its own $c^\star$, $\theta^\star$, $N^\star_{ch}$, and $N^\star_{str}$).

Existence of at least one such cut c for each of the testing data set events passed by the machine guarantees that the purity in the regions with extrapolated θ, N_{ch}, and N_{str} is at least as good or better than in the regions for which background events existed in the training simulated data set. This is an important advantage of the discussed method compared to the other machine learning techniques.

Counting all *purity cuts* available for a given testing data set event provides one with an important machine quality parameter which value is higher for events that are more likely to be signal and lower for events that are more likely to be background. It appears that a cut on this parameter improves the purity in all subsets by equal amount. In order to improve the purity in all subsets to the same final value one may weight the terms in the quality parameter sum with the initial purity of the training simulated data set on the subsets of the cuts used in the sum.

We call the machine learning method described here the *subset browsing method* because of the technique in which one has to *browse* through the *subsets* on which the cuts of the *trained cut set* are defined to calculate the quality parameter separating signal from background. The quality parameter itself is called the *SBM quality parameter*: sbm.

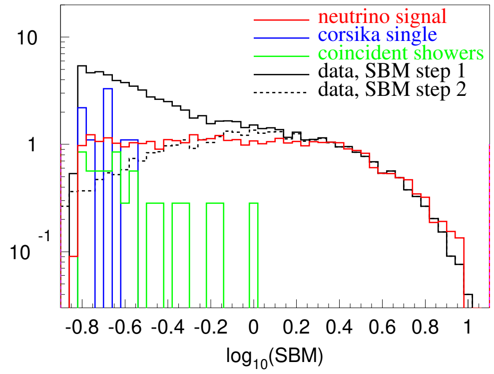

Fig. 5. Distribution of the quality parameter after step 1^a (for 1 year of IC-22 data). The 95% estimated purity is achieved at SBM=0.36.

The *irreducible trained cut set* forms a “skeleton” of cuts that are applied to the testing data set achieving the *initial SBM cut level*: $sbm = 0$. The sbm quality parameter is usually normalized so that the highest value of sbm of a background event in a testing simulated data set is 1 (e.g., in the plot of the sbm in Figure 5).

The SBM performance in separating signal and background in IceCube data appears to be substantially better than that of other methods if the number of training simulated data set events is low. The SBM extrapolation behavior is very robust and the performance of extrapolation improves as the testing simulated data set statistics increases, since more *purity cuts* become available to testing data set events in regions less populated by *trained cut set* cuts. The machine cannot be over-trained (a condition in which other methods, when improperly used, may elect to go around individual background events of the training data set) because the cuts, by construction, are monotonous functions of the basic parameters. Additionally, the learning method itself is very simple and has virtually no parameters to set (such as type and parameters of the kernel function of SVM). Most of the machine is implemented with only an application of the $\leq$ operator (and a simple summation for estimating the quality parameter).

V. Comparison with other approaches

The following table compares the method performance with best results achieved in other studies. The number of signal events per day in simulation at 95% or higher estimated purity level or where agreement with data is reached, whichever is higher, is chosen here as a gauge of performance.

IceCube year	SBM	other
IC-22 2007	29.0	SVM: 26.6
IC-40 2008	83.1	BDT: 71.0

References

[1] D. Chirkin, et al., *Effect of the improved data acquisition system of IceCube on its neutrino-detection capabilities*, 30th ICRC, Merida, Mexico