Performance Study of a Time-Clustering Algorithm for a Neutrino Point Source Search with the IceCube Detector

Alexei Dikalyuk
work done in collaboration with
Elisa Bernardini
Robert Franke
DESY Zeuthen
dial_up@list.ru

September 11, 2008

In the following work we carried out a performance study of a time-clustering algorithm in the search for flares of neutrinos using the IceCube detector (configuration with 40 strings). For a fixed search bin size and different declinations (5, 40, 80 degree) we derived plots of the signal detection probability against source strength. We examined the behavior of these curves in dependence of the time-scale, in which the source is active. Two different methods of background estimation were used in this analysis. A comparison of these methods will also be presented. For this work approximately two months of IceCube-40 data with randomized right ascension were used ("blind analysis").

1 Introduction

Different observations of the candidates for neutrino sources tell us that their electromagnetic emission is very variable and often shows a flare behavior [1], [2]. According to several models one can expect that the neutrino emission from those sources has a similar character. The standard point source search may not be sensitive to such events due to the weak expected signals compared to the irreducible background. That is why it is important to develop new types of algorithms that are taking into account the time information of events.

The time-clustering algorithm presented in this work is one of such approaches. The algorithm looks for time clusters of neutrino events which come from a certain direction. The big advantage of this algorithm, is that it makes the neutrino flare search independent of any a priori assumption concerning the time structure of the signal.

Of course, if we want to show that this algorithm is more efficient than time integrated searches, we have to do a performance study at first. This will give

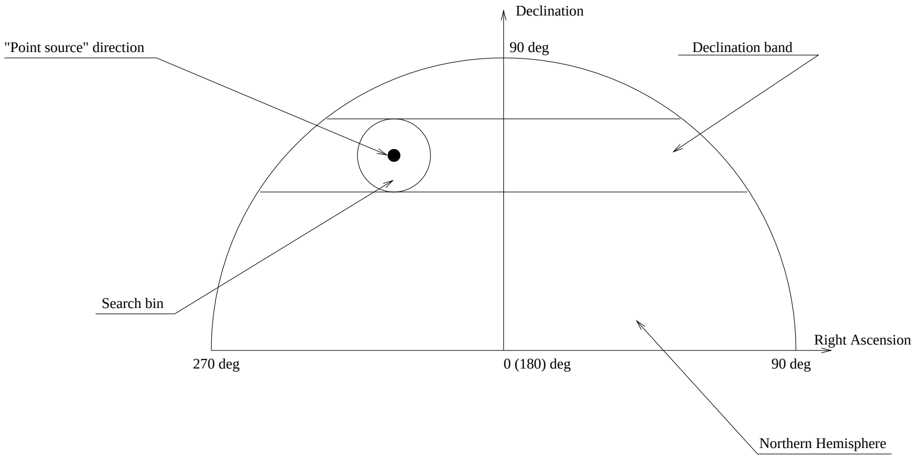


Figure 1: "Point source" direction, search bin, declination band.

us information about its range of application, i.e. in which cases and situations it is better to use it instead of standard ones. Also such analysis can lead us to new improvements, which are good to include into the algorithm or even give us ideas on new approaches.

2 Description of the Basic Algorithm

For the further clarification of our actions it is necessary to explain here the basic concepts of the time-clustering algorithm with a fixed bin size:

1. The data for these studies were produced by reconstruction level 2 using so called "soft cuts": we required a declination of more than 0 degrees (down-going tracks only with respect to the detector position), a number of direct hits (C-type) of more than 5 and an angular resolution (estimated from the so called "paraboloid fit") of less than 6 degrees. This procedure led to a decrease number of analyzed events.

2. We fix the declination and the right ascension of the "point source" and define the size of the search bin. This provides us also with the declination band, which corresponds to parameters that we described above (Figure 1).

3. We calculate the number of events, that are within the declination band (N_{band}) over the whole time period of observation (in our case approximately two months). Using this number we can estimate (this is a very rough estimation, we will discuss later a better one and provide a comparison between them) background rate in two steps:

    - using the expression:

$$N_{bin}^{bg} = N_{band} \cdot \frac{S_{bin}}{S_{band}}, \tag{1}$$

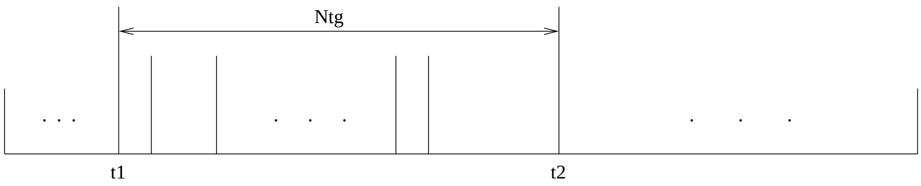


Figure 2: Idea of time-clustering procedure.

where: N_{bin}^{bg} - mean number of expected background events in the search bin; S_{bin} - area of the search bin; S_{band} - area of the declination band.

- using N_{bin}^{bg} we get the mean rate of expected background events ($\dot{N}_{bin}^{bg}$) within this search bin with the following equation:

$$\dot{N}_{bin}^{bg} = \frac{N_{bin}^{bg}}{t_{last} - t_{first}}, \quad (2)$$

where: t_{last} - arrival time of the very last event in our two month collection; t_{first} - arrival time of the very first event in our two month collection.

4. With the assumption that all events within the search bin are produced by background we built a histogram of the excess parameter (it will be defined a little bit later) distribution. We can do it using the following Monte-Carlo procedure. Each time we generate, using a Poisson distribution, a number of events with a mean of N_{bin}^{bg}. Using the time information from the whole period of observations (in our case it was two months: from 21 April till 30 June 2008), we draw randomly for each of these events an arrival time. When all preparations are finished, we can apply to this newly created data set the following algorithm:

- We examine all possible pairs of times within our data set (Figure 2).
- Using the background rate, we estimate mean number of expected background events within this time gap:

$$\mu = \dot{N}_{bin}^{bg} \cdot (t_2 - t_1). \quad (3)$$

- Using information from the data set, we count the number of events that are within the time gap (N_{tg}).
- For each of these pairs of times we calculate the Poisson probability (Pr), that this number of events is induced by background. Then we calculate the excess parameter (ξ) according to the expression:

$$\xi = -log_{10}(Pr). \quad (4)$$

- The last step is to find out the pair of times with the biggest excess parameter (ξ_{best}).

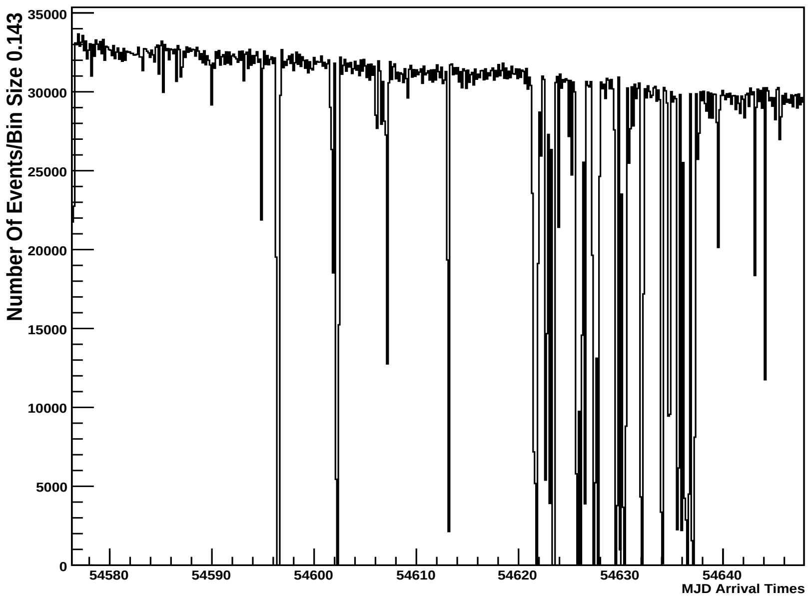


Figure 3: Arrival time histogram in data set, that we use in this study.

5. When this procedure is complete we can define, using the excess parameter distribution (it is a function of the corresponding declination band), its threshold value (ξ_{tresh}) for a certain declination band. We set the confidence level to 99.73% to define this threshold value.

6. Now we can choose events, which are within the search bin and apply to them the procedure, that is described in point 3. We assume that the signal is found, if ξ_{best} is bigger then ξ_{tresh}.

However, this algorithm has one major disadvantage: the current background estimation is very rough. Indeed, we use the whole period of observations, but actually the number of events within each sub period can vary drastically as can be seen in Figure 3. This means that in some cases we overestimate the background, in some - underestimate it. To deal with this, we can use a different approach. We can estimate the background using information only from that sub region of time, that we are going to analyze, i.e. the time range of the time cluster currently under investigation.

Another thing that still can be improved is the size of the search bin. Actually it is not necessary to fix it a priori. We can perform a very simple mental experiment (Figure 4). It is obvious, that the background and signal distribution vary in space, but signal events are only distributed near the source position, while the background is uniformly distributed. We know that the number of background events is bigger than the number of signal events. So what if we will look not into the search bin, but into the whole sky? What will we see? Nothing

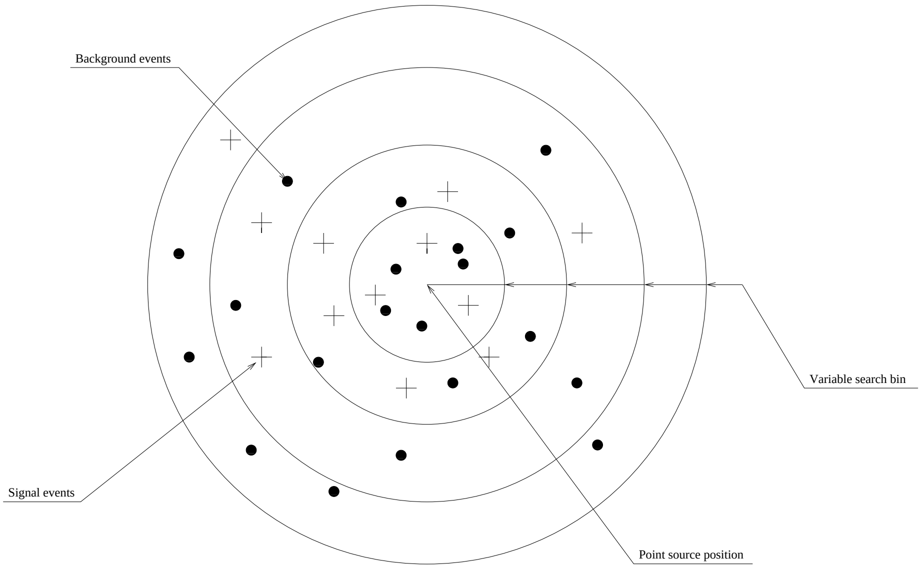


Figure 4: Illustration of the mental experiment, concerning a variable bin size. One can see that we cut a lot of background when shrinking the bin size. It increases the ratio of signal to background and increases the probability to detect the signal.

but background. Let us shrink the search bin to the point source position. By this procedure we will cut away big amount of background, but the signal events are still near the source location. Till some moment we can expect, that the probability of signal detection will increase. But if we continue to shrink the size of the search bin to the point where the source is located, it will contain less and less events and the probability will decrease one more time. So one can conclude that an optimal size of the search bin exists, that provide us the best probability to find the signal.

3 Performed Tests and Discussion of Results

3.1 Estimation of the Signal Point Spread Function

To carry out a performance study of the algorithm properly the first thing that we have to find out (because we use such concepts as fixed bin and variable bin size, which means that we work with space distributed events) is how our signal events are distributed in space. Actually, we don't care about their real angular distribution, the only thing that we have to know is how far away from the source location these events can emerge.

To construct this signal point spread function we use the following technique. Using information about the shape of the IceCube 40 detector one can do a full Monte Carlo simulations of the process of muons track detection within it and then implement all algorithms that the IceCube community uses for track reconstruction. The point is that in this case we know real tracks of muons, so

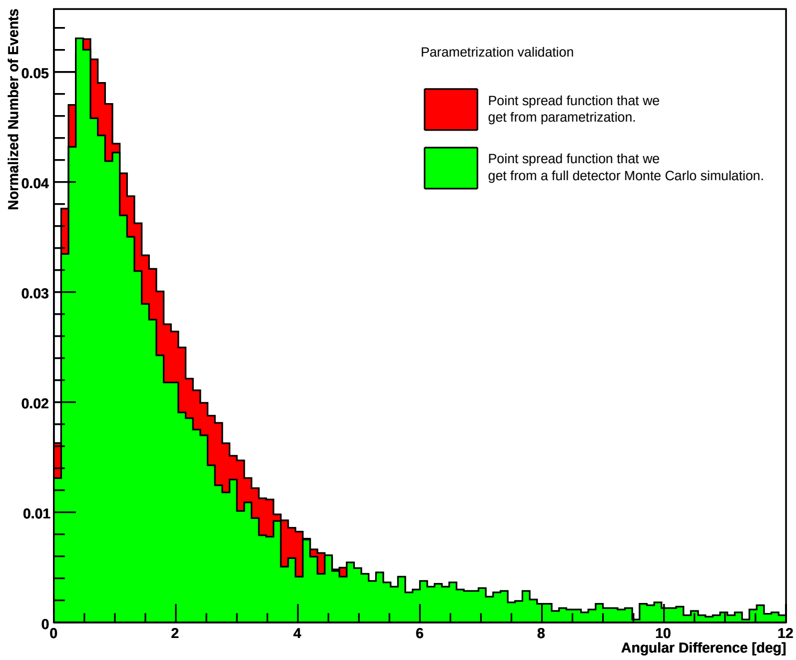


Figure 5: Comparison between the point spread function given by parametrization and obtained from a full detector Monte Carlo simulations.

one can calculate the angular difference between the reconstructed track and the real one. Having this information, we can find the proper parametrization for the angular difference distribution and use it in our simulations.

This technique was implemented in our studies and a parametrization was found. To get a better agreement with the "real" angular difference distribution some free parameters were introduced. The chosen parametrization is:

$$\omega_1 = 0.29,\ \omega_2 = 1.2,\ \sigma_1 = 1.054 + 0.0053 \cdot \sin\delta,$$
$$\sigma_2 = 0.5097 + 2.7786 \cdot \sin\delta, \tag{5}$$
$$x = gaus(0, (1-\omega_1)\cdot\sigma_1 + \omega_1\cdot\sigma_2),\ y = gaus(0, (1-\omega_1)\cdot\sigma_2 + \omega_1\cdot\sigma_1),$$
$$\Delta\psi = \omega_2 \cdot \sqrt{x^2+y^2},$$

where ω_1, ω_2 - free parameters, gaus(μ,σ) - Gaussian distribution with a mean value of μ and a variance of σ, $\Delta\psi$ - angular difference that we are interested in.

Figure 5 compares this parametrization to the distribution from a full detector Monte Carlo simulations.

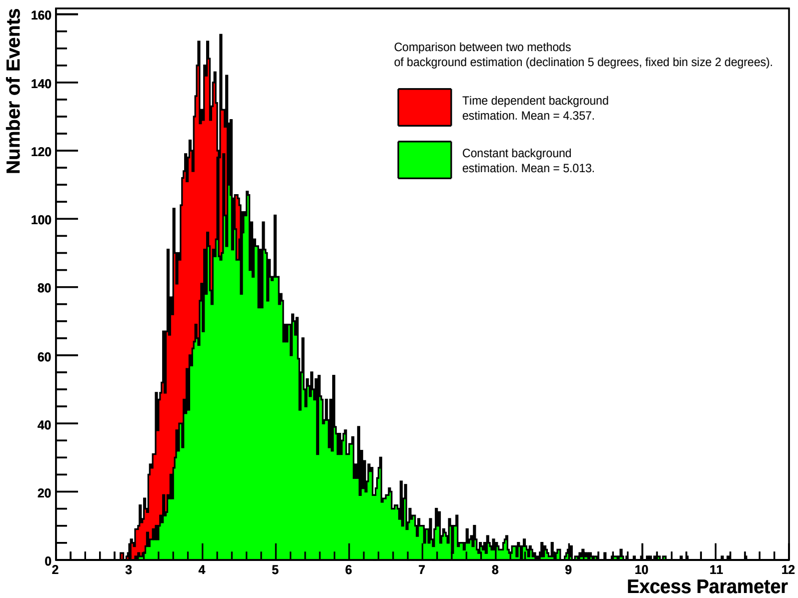


Figure 6: Example of excess parameter distribution for a source declination of 5 deg.

3.2 Comparison between Different Approaches in Background Estimation

As we mentioned above, we used two different ways of background estimation (section 2). To examine them, we carried out a number of tests.

At first we constructed the distribution of the maximum excess parameter using each of this techniques. As an example, we show here this plot only for a declination of 5 degrees (Figure 6).

To estimate this effect numerically, we will plot the detection probability against source strength for fixed search bin (2 degrees) at declination 5 degrees and for two different time length of source activity (Figure 7). Now we can see, that the time-depending background estimation approach is better than the constant rate background estimation. Also we could roughly estimate the efficiency of the binned approach from the data that we had. The detection probability increases by about $\sim$10%.

As the time-dependent background estimation method provides us with a better detection probability than the constant rate estimation, in all further studies we will use only it.

3.3 Fixed Search Bin Size

In this chapter we will provide plots of the signal detection probability against source strength for three different declinations at 5, 40 and 80 degrees

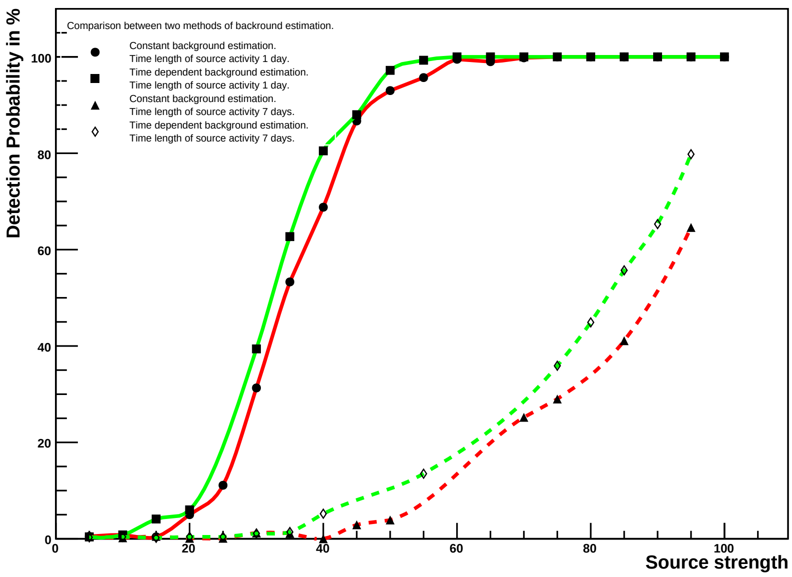
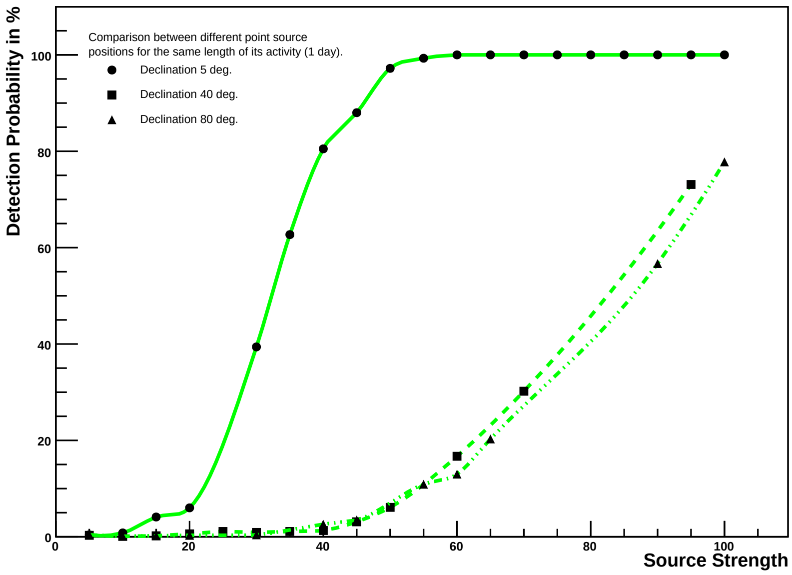


Figure 7: Detection probability against source strength for declination at 5 deg.

(Figure 8 and Figure 9). We examined different time lengths of source activity because we don't know how variable sources are.

On Figure 8 one can see a high sensitivity of our algorithm to the short signals on different declinations. On Figure 9 one can also observe that we can get the signal events on the 1 week scale (actually we have to think not about absolute time lengths, but about relative ones to the whole period of observations, because background has uniform distribution in time) but only at a declination of 5 degrees. We believe, that this fact represents the dependency of the point spread function against declination, which is connected with the shape of IceCube-40. IceCube-40 provides better resolution to tracks which are closer to the equator. So we can see this effect on the plot. This explanation is also in good agreement with the Figure 8, because in this case we have two different factors which have influence on the curve position. The first is an expected number of background events and the second is the distribution of the signal events within the bin. As our tracks are closer to the north pole they are resolved worse, but also from the data distribution we know, that on the pole we get extremely less number of events than in the middle of the sky. From one hand we have a lot of background in the middle, but also we have more signal events than on the north pole. The situation is opposite on the pole. That is why curves which correspond to declinations of 40 and 80 degrees are close to each other on Figure 8.

The main conclusion, that we can make according to this plots, is that the fixed bin size approach is quite effective on relatively small ($\sim 0.01 - 0.02$) time scales, compared to the whole time period that is analyzed.



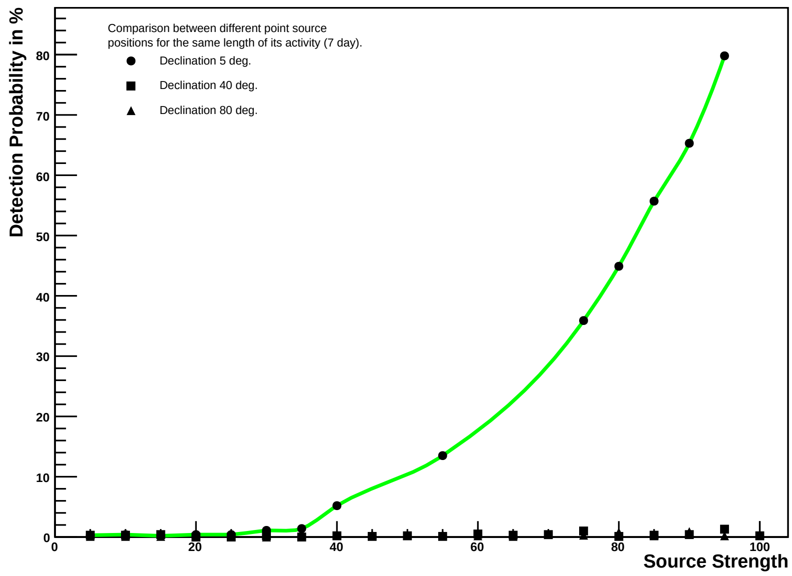


Figure 8: Detection probability against source strength for declination at 5, 40, 80 deg for 1 and 7 days signal length.

4 Conclusion

The performance study of the time-clustering algorithm was performed. For this purposes a parametrization of point spread function of signal events was found (Expression 5). We discover that the point spread function strongly dependents on declination. While at 5 degree declination all of signal events were within the search bin of 2 deg, at 40 - there was only half of them there and at 80 - one fifth.

A comparison between two background estimation approaches was performed. We showed that time dependent background estimation makes this algorithm more sensitive (∼10%) then the constant background estimation approach.

For a fixed bin size approach we showed that this algorithm is strongly dependent on the point spread function on different declinations but also on the time period when source was active. The last dependence is stronger the smaller this period is compared to the whole period of observation.

Still one more extension of this algorithm can be made. We can examine its behavior on different bin sizes. One can study for predefined signal strength and declination what will happen if we will change size of the bin. This study can provide us with a "best size" for the search bin, which increases sensitivity of the algorithm.

Also one more thing can be done here. Actually one can make a comparison between this algorithm and the time integrated point source search.

Acknowledgments

I want to thank every person here in DESY Zeuthen who made this program possible and our stay here comfortable. I am grateful to all of AT group people who was very sympathetic and friendly. The very special credit I want to make to Elisa Bernardini for her advices and problem discussion and Robert Franke for his help in code developing and result discussion.

References

[1] K. Satalecka Time clustering search for neutrino flares with AMANDA-II

[2] M. Ackermann, Search for signals from cosmic point-like sources of high energy neutrinos in 5 years of AMANDA-II data, PhD-Thesis, 2007