29th International Cosmic Ray Conference Pune (2005) **00**, 101–106

Algorithms to search for point-like neutrinos sources with the ANTARES telescope

J.A. Aguilar[a] on behalf of the ANTARES collaboration
(a) IFIC – Insituto de Física Cospuscular, CSIC – Universitat de València, apdo. 22085, E-46071 Valencia, Spain
Presenter: J.A. Aguilar (aguilars@ific.uv.es), spa-aguilar-sanchez-JA-abs2-og25-poster

The detection of astrophysical sources of high energy neutrinos is one of the main motivations for building a neutrino telescope. The ANTARES telescope is being deployed in the Mediterranean Sea. The detector will consist in a tridimensional array of photomultipliers (PMTs) that will detect the Cherenkov light induced by the muons produced in neutrino interactions. Since the neutrino fluxes from point-like sources are expected to be small, it is of the utmost importance to take advantage of the ANTARES accurate pointing power (angular resolution less than 0.3 degrees for E_ν greater than 10 TeV) to disentangle a possible signal from the unavoidable atmospheric neutrino background. In order to distinguish an excess of neutrino events from the background, several search algorithms have been developed within the ANTARES collaboration. Some of them perform binned searches of clusters, some others use likelihood maximization techniques and recently a method based on the well known Expectation-Maximization clustering algorithm has been devised. In this contribution, these different methods are reviewed and their discovery potential using the ANTARES neutrino telescope are presented.

1. Introduction

The ANTARES collaboration [1] has started the construction of an underwater neutrino telescope in the Mediterranean Sea at a depth of 2.4 km. The first line will be deployed during 2005, and the construction of the whole detector (12 strings) will be completed by 2007. Each line is equipped with 75 10" photomultiplier tubes (PMTs) joined in groups of three PMTs that make 25 floors along the line. The lines are kept vertical by means of a buoy located at their upper end. The mean distance among lines is about 65 m and the instrumented length starts at 100 m over the seabed and covers about 350 m.

The muon track reconstruction is performed using the arrival times of the Cherenkov photons on the PMTs and the signal amplitudes. This time information is essential for the muon track reconstruction and therefore an accuracy of about 0.5 ns in the relative time calibration among PMTs is desirable to perform a good reconstruction procedure despite optical background such as decaying ^{40}K and bioluminescence. The resulting angular resolution for the ANTARES telescope at high energies (above 10^5 GeV) is less than 0.2°. At lower energies, the angular resolution is dominated by the angle between the muon track and the original neutrino direction. The angular resolution together with the effective area determine the performances of a neutrino telescope. The good pointing accuracy and large effective area at high energies (0.06 km^2 at 100 TeV) allows the ANTARES telescope to set restrictive upper limits to neutrino fluxes from the point-source candidates. Consequently, different algorithms and techniques for searching point-like sources have been devised within the collaboration. In the absence of real data, the methods described in the following sections have been applied to Monte Carlo simulated data.

2. Binned methods

Binned methods [2] look for a cluster of events around any direction in the sky or in bins of grids in which the sky is divided. The significance of clusters is estimated in comparison with the distribution of atmospheric neutrino events uniform in right ascension. These methods rely on the knowledge of the point spread function of neutrino induced events around a point-like source. Nevertheless no hypothesis is assumed about the neutrino source beyond the neutrino emission according to a power law with spectral index of 2. We studied and compared two different approaches:

- The **grid method** in which the sky is divided in a grid in declination and right ascension. Given a bin, the probability that it includes the observed or a higher number of events is computed in the hypothesis that all events are due only to background using the mean value of the corresponding declination band.
- The **cluster method** looks for events in a cone around each of the measured events. A significant excess of events in a cone would indicate the presence of a point-like source. In this case, the relative significance is obtained by computing the probability for the background to produce the observed or higher number of events in the cone.

Both methods require the selection of the initial size or shape of the bins or cones. This choice is performed through an optimization of the bin size comparing signal to noise. Figure 1 shows the discovery potential (for a confidence level of 5 sigmas) for both the cluster and the grid methods for different declinations as a function of the mean number of events emitted by the source.

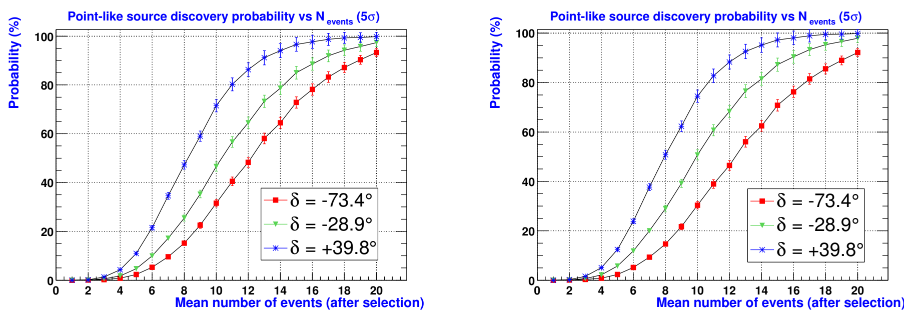

Figure 1. Results obtained with the grid method (left) and the cluster method (right) after one year of data taking for different declination bands as a function of the average number of events emitted by the source.

3. Likelihood ratio search method

This method was the first developed in ANTARES that does not rely on binning [3]. The likelihood ratio (LR) method operates by testing the data compatibility with two hypotheses: one, often called H_0, is the *null* hypothesis and the other one called H_1. The *null* hypothesis is assumed to be the presence of background only, and H_1 considers the existence of a neutrino point source in addition to the atmospheric background. This compatibility is accomplished by computing the value of the so-called test statistic:

$$\lambda = \log\left[\frac{P(data|H_1)}{P(data|H_0)}\right] \quad (1)$$

Expressions for $P(data|H_{1,0})$ rely on the parametrization of the (energy dependent) effective area and angular resolution of the detector and of the probability of finding a muon with a given reconstructed energy as a function of the neutrino energy. Once the ingredients for the construction of the likelihood for both background and signal are available, maximization is performed numerically. The distributions of the test statistics are computed for the case of background-only samples. A discovery is made if the test statistic exceeds a critical value λ_c that determines different confidence levels (CL). The discovery potential or probability of discovering a source is the probability of the source plus background distribution to exceed the value λ_c. Figure 2 shows the discovery potential as a function of the expectation value of the number of observed signal events. The neutrino flux needed for a discovery of 50% is also shown.

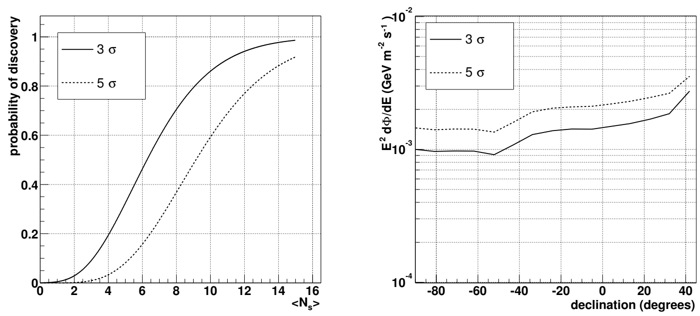

Figure 2. Left: Discovery potential for two confidence levels for a declination of -81°. Right: Neutrino flux from a point source needed to yield a probability of 50% of discovery in a one-year full sky search as a function of the declination.

4. Expectation-Maximization

The Expectation-Maximization [4] algorithm is a general approach to maximum-likelihood estimations for finite mixture model problems. As in the case of LR, this method is subject to parametrization of the density distribution for both the signal and background neutrinos. Nevertheless, in this case sources are supposed to follow gaussian distributions. This turns out to be a reasonable assumption. The background distribution is inferred from the Monte Carlo data sample, but it could be obtained directly from the real data by scrambling the right ascension coordinate of the measured events. No energy information or further performances of the detector were used in this case. The parameters are determined by maximizing the likelihood using the EM algorithm. The EM assumes the existence of a *missing* or *hidden* information, in our case the knowledge of whether a neutrino event comes from a given source or it is produced by the atmospheric neutrino background. Hence, the real observed data can be understood as an *incomplete* sample, so that, adding a new vector **z** we can build the *complete* data sample, where **z** is just a class indicator vector that tells whether an event comes from a source or not. Maximization of likelihoods analytically intractable can be easily accomplished by means of this methodology. The EM general method follows an iterative procedure where each iteration has two steps:

1. E-step: The expected value of the *complete* data log-likelihood, conditional to the observed data, is computed for a given set of parameters $\{\mathbf{\Psi}^{(m)}\}$.
2. M-step: Find $\mathbf{\Psi} = \mathbf{\Psi}^{(m+1)}$ that maximizes the expected value. This maximization will lead to the maximization of the desirable log-likelihood of the *complete* data sample.

The model selection criteria is performed by means of the *bayesian information criterion* or BIC. This BIC value is an approximation of the integrated likelihood when the number of events, n, is high enough:

$$2\log[p(D|M_k)] \approx 2\log p(D|\hat{\boldsymbol{\theta}}_k, M_k) - \nu_k \log(n) = \mathrm{BIC}_k \tag{2}$$

where M_k stands for the different models or candidate clusters and ν_k is the number of the independent parameters to estimate. From the BIC distribution we can infer the discovery potential of a point-like source to be detected as can be seen in figure 3.

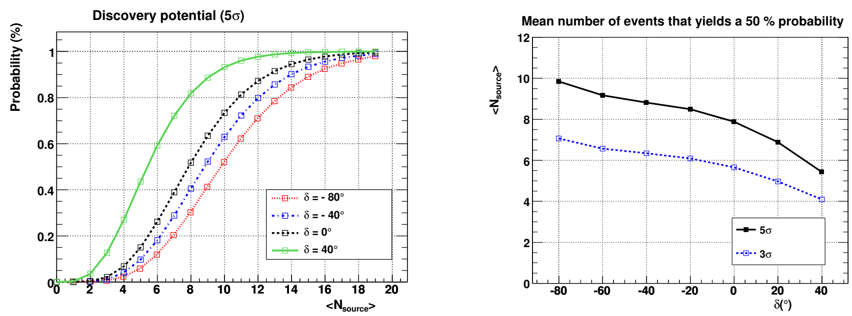

Figure 3. Left: Discovery potential as a function of the number of events emitted by the source for different declinations for a 5 σ confidence level. Right: number of events that yields a probability of 50 % as a function of the declination for two confidence levels.

5. Conclusions

The ANTARES location leads to a privileged sight of the most interesting areas of the sky like the Galactic Centre where neutrino source candidates are expected. Therefore, different searching algorithms have been developed within the collaboration. Each one of the presented techniques has different advantages as well as drawbacks. When comparing among methods it has to be taken into account that different event selection criteria after the reconstruction should lead to different number of detected signal events and therefore different discovery potential versus expected number of events. The binned methods and the Expectation-Maximization pattern recognition algorithms were applied making a minimum use of the detector performances as given by Monte Carlo simulation. They provide, therefore, very robust results. The Likelihood-Ratio method, on the other hand, shows a better sensitivity, but uncertainties in the detector's response can have a larger impact on its results. In this sense, these three techniques are broadly complementary and will allow ANTARES to look for neutrino sources in a very thorough and timely manner.

References

[1] ANTARES Collaboration. *A deep sea telescope of high energy neutrinos*, astro-ph/9907432, 1999.

[2] E. Carmona, *Study of the event reconstruction and expected performances for the point-like sources of the future ANTARES neutrino telescope*, PhD Thesis 2003, Universitat de València (Spain).

[3] A. Heijboer, *Track Reconstruction and Point Source Searches with ANTARES*, PhD Thesis 2004, Universiteit van Amsterdam, (The Netherlands).

[4] A.P. Dempster, N.M. Laird and D.B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society Series B, 1977. 39:1-38.