UNIVERSITE DE MONS-HAINAUT

Physique Générale et Physique des Particules Elémentaires

Point source search in AMANDA using a maximum likelihood method ('97 data)

PPEl/UMH 2002 12-360

Thierry Castermans
Fernand Grard

Point source search in AMANDA using a maximum likelihood method ('97 data)

Introduction :

The analysis of the 1997 AMANDA data in the prospect of finding evidence for point sources has been currently conducted by binning the sky and comparing in each bin the number of observed events with an expected average background. This analysis has revealed no significant point source signal.

In this report, we will refer to the results of a similar analysis which have been presented by one of us (TC) in a DEA thesis[1]:

- After the appropriate cuts to get the best signal to background ratio, the total number of muon tracks which have been retained for the point source search in the Northern hemisphere is 2557. Fig. 1 shows the sky plot of these events versus declination (δ) and right ascension (α). No obvious accumulation of events can be seen.
- In order to evaluate the statistical significance of possible accumulation of events, the sky has been subdivided in bins ($\Delta\delta$, $\Delta\alpha$) taking into account the angular resolution: $(HWHM)_\delta = 2.2°$; $(HWHM)_\alpha = 0.25h$ as determined from MC simulation. In short, the sky has been subdivided in 10 equally wide declination slices ($\Delta\delta = 9°$), the first slice at the horizon has been subdivided in 24 bins with $\Delta\alpha = 1h$ (15°) and the others such that in each resulting bin, the subtended solid angle is about the same. The total number of bins is 154.
- The event density happens to be dependant on the declination angle. In each of the δ declination slices, the average number of background events per bin (b) has been determined by dividing the total number of observed events in that slice by the corresponding number of bins.
- In each bin, a statistical estimator ξ or "significance" is computed from the observed number of events N and the average background b, as:

$$\xi = -\log_{10} \Pi(N,b)$$

$\Pi(N,b)$ being the "Poissonian probability of excess":

$$\Pi(N,b) = \sum_{n=N+1}^{\infty} P(n,b)$$

- From the definition of ξ, it appears that an anormalously high N would lead to a high value of ξ. Fig. 2 shows the bin distribution as a function of ξ. The bins of interest for a point source search are of course those which are characterized by a high value of ξ, in particular the three bins with $\xi > 1.4$.
- In order to evaluate the probability that these bins may exist in the absence of point source signals, the observed ξ distribution is compared to the ξ frequency distribution which has been obtained by randomization of the data sample, i.e. by associating to each measured declination δ one of the measured right ascension angle chosen at random in the sample. The final ξ frequency distribution has been obtained by averaging the frequency distributions as a function of ξ from 100 randomized samples. This ξ frequency distribution is also shown on Fig. 2 for comparison with the experimental data.
- It can be seen that the two distributions are quite compatible for values of ξ less

[1] Thierry Castermans : « Recherche de sources ponctuelles de neutrinos d'origine astrophysique avec le détecteur AMANDA » Université de Liège, 2002.

than 1.4. Above that value, there are three bins present in the data: 2 bins in the interval $1.8 < \xi < 2.0$ where 0.5 bins are expected from the background; 1 bin in the interval $2.6 < \xi < 2.8$ to be compared with 0.067 bins from background. Taken at face value, these numbers lead to 7.6 % and 6.2 % probability that the observed bins are consistent with the background.

- The following remarks should be made:

- For $1.4 < \xi < 1.8$, no bin has been observed and the expected average number of bins from background in this interval is 2.59. The probability of occurrence of this downward fluctuation is 7.5%, similar to the probability for the above upward fluctuations. For $2.0 < \xi < 2.6$, no bin has been observed with an average number from background of 0.41, a situation which can occur with a probability of 66.3%. This is quite acceptable as it is beyond $\xi > 2.8$
- The histograms of Fig. 1 have been plotted with a step $\Delta\xi = 0.2$, which is about equal to the variation of ξ in the region of interest when one changes the number of observed events by one unit in a bin (whose average number is about 18 as it is in the present search). However in reality, the values of ξ evaluated from the different possible average number of background events and number of observed events in the bins covering the sky are spreading with a finer granularity over the whole range of ξ. The choice of the $\Delta\xi$ step has therefore a direct influence on the statistical significance of the observed effect as evaluated above. For example, if one makes two equal intervals instead of the one at 2.6 – 2.8, the observed bin will be located in one of the two smaller intervals in which the average expected number of bins would be about 0.067/2, i.e. about half the above quoted value. This would lead to a reduced probability that the observed bin is consistent with background. Choosing a larger step would instead lead to the reversed conclusion.

- Taking these remarks into consideration, one would rather evaluate the probability of occurrence of the three bins above $\xi = 1.4$ (below this limit, the statistics is comfortable and there is good agreement between data and background expectation) to be compared to a cumulated background of 3.6 bins. This probability is equal to 22.1%, well high enough to conclude that there is no significant evidence for a point source signal.

- As a possible point source signal may be spread over two adjacent bins, the above procedure has been repeated by shifting the bins by half a bin width in δ and/or α and has lead to the same conclusion.

In the following, we will concentrate on the "hot" bin characterized by $\xi = 2.61$ and centered at $\delta = 22.5°$; $\alpha = 14.73h$ with $\Delta\delta = 9°$ and $\Delta\alpha = 1.091h$. The average background in that bin is 18.36 events and the observed number of events is 32. Fig. 3 shows the bin configuration around this "hot" bin, with corresponding values of b, N and ξ.

A maximum likelihood procedure for determining the best point source location and its relative intensity will be applied to the data in an attempt to evaluate the statistical significance of this signal candidate.

The maximum likelihood procedure:

- The strategy at the basis of this procedure consists in searching for a point source signal in a restricted region of the sky. A "box" with adequate dimensions $\Delta\delta$, $\Delta\alpha$ has therefore to be defined.
- For each event (i) in the box, the likelihood amplitude is written as:

$$f_i = s.S(\delta_i, \alpha_i; \delta_0, \alpha_0) + (1-s)B(\delta_i, \alpha_i)$$

- s = relative intensity of the signal ; (1-s) = same for the background
- S, B = signal and background amplitudes, normalized to 1 in the interval (Δδ, Δα)
  - S, B represent the probability to find an event as signal or background event at a given point in the interval (Δδ, Δα)
  - the likelihood amplitude is so normalized to 1 :

$$\int_{\Delta_\delta,\Delta\alpha} f_i . d\delta . d\alpha = 1$$

  - as a first approximation, S has been written as the product of two uncorrelated gaussians centered on δ_0 and α_0 with fixed standard deviations σ_δ and σ_α:

$$S_{unnorm} = \frac{1}{\sigma_\delta \sqrt{2\pi}} e^{-\frac{(\delta-\delta_0)^2}{2\sigma_\delta^2}} \times \frac{1}{\sigma_\alpha \sqrt{2\pi}} e^{-\frac{(\alpha-\alpha_0)^2}{2\sigma_\alpha^2}}$$

which have to be normalized to 1 :

$$S = \frac{S_{unnorm}}{\int_{\Delta\delta,\Delta\alpha} S_{unnorm} . d\delta . d\alpha}$$

  - The background amplitude B(δ, α) has been assumed to be uniform in α and has been determined by least square fitting the event distribution as a function of δ to a suitable analytical function (B_{unnorm}), which should then also be normalized to unity in (Δδ, Δα) :

$$B(\delta,\alpha) = \frac{B_{unnorm}}{\int_{\Delta\delta,\Delta\alpha} B_{unnorm} . d\delta . d\alpha} = \frac{B_{unnorm}}{\Delta\alpha \int_{\Delta\delta} B_{unnorm} . d\delta}$$

- From the events located in the box (Δδ, Δα), the following likelihood function is constructed:

$$L = \prod_i f_i \equiv L(s, \delta_0, \alpha_0)$$

The most likely values of s, δ_0, α_0 consistent with the data and background are those which make L maximum. As at the limit of large statistics L becomes a gaussian as a function of s, δ_0, α_0, the standard errors on these quantities are evaluated considering that a shift of plus or minus one standard error with respect to the best values found would lead to a reduction of the maximum of lnL by 0.5.

- The determination of the best value of s with its standard error σ_s should in principle allows one to evaluate the statistical significance of a candidate signal.

Application of the above maximum likelihood procedure

- Signal amplitude:
  The following values have been chosen:

$$\sigma_\delta = 3°; \sigma_\alpha = 0.333h$$

- Dimensions and location of the search box:
  The box is centered on the "hot" bin n^r 61

$$\delta = 22.5°;\ \alpha = 14.73h$$

  Δδ and Δα have both been set equal to 6 standard deviations:

$$\Delta\delta = 18° ,\ \Delta\alpha = 2h$$

  The total number of data events in the box is 80.
  Taking into account that the average background in the hot bin is 18.36 and the number of observed events 32, a rough estimation of s would give:

$$s \approx \frac{32 - 18.36}{80} = 0.17$$

- <u>Background amplitude:</u>
  The background amplitude has been determined by least square fitting a fourth order polynomial to the whole event distribution as a function of δ. The following B_{unnorm} function has been obtained:

$$B_{unnorm} = 0.0002\delta^4 + 0.0009\delta^3 - 0.9452\delta^2 + 12.315\delta + 531.54$$

Results:

- With s having been set equal to 0.17 (it should be set ≠ 0!), the best values of δ_0 and α_0 have been determined by an iterative procedure until the minimum of (-lnL) has been found to a good approximation:
  - α_0 is given a value close to the center of the box and (-lnL) is plotted versus δ_0
  - δ_0 is then given the value corresponding to the minimum of (-lnL) and (-lnL) is plotted versus α_0
  - α_0 is given the value at the minimum of (-lnL) and (-lnL) is again plotted versus δ_0
  - and so on …

The best values and corresponding errors are:
- $\alpha_0 = 14.7125 \pm 0.1725 (= 2.6°)$
- $\delta_0 = 22.2 \pm 2.6°$

to be compared with the center of bin 61: α = 14.73h ; δ = 22.5°, which is quite satisfactory.

- Fig. 4 shows the plot of (-lnL) versus s, with α_0 and δ_0 set respectively to 14.7125 and 22.2 . (-lnL) has its minimum in the vicinity of s = 0 with an estimated standard deviation of 0.07, a result which could be taken as a confirmation that the excess of events in the "hot" bin among the statistical fluctuations in the surrounding bins does not lead to a significant signal

- In order to check the good operational behavior of this maximum likelihood procedure, events have been added in the center of the “hot” bin in order to get a really significant signal, i.e. a resulting value of s which is consistent with the expected one. The following Table summarizes the results:

Nr of events added	Expected s value	Best s value	σ_s	ξ
0	0.17	0.00	0.07	2.61
5	0.22	0.08	0.10	4.07
10	0.27	0.17	0.09	5.80
20	0.34	0.31	0.09	9.97
30	0.40	0.40	0.09	14.96
40	0.45	0.47	0.09	20.63

The corresponding values of the significance ξ are given in the last column.
Fig. 5 to 9 show the distributions of (-lnL) versus s.

- The maximum likelihood search has been repeated by choosing a box which is not centered on the “hot” bin but shifted by 1.5 standard deviations in both δ and α directions. The total number in the box is 73.

The efficiency with which the signal is localized and its intensity is determined is practically the same as above:

- Localization of the signal: $\delta = 22.15 \pm 2.6°$; $\alpha = 14.72 \pm 0.2$ h

- Results from the addition of events:

Nr of events added	Expected s value	Best s value	σ_s
0	0.19	0.00	0.08
5	0.24	0.05	0.10
10	0.29	0.15	0.12
20	0.36	0.30	0.10
30	0.42	0.40	0.10
40	0.47	0.50	0.10