Search for cosmic neutrinos with ANTARES

Bogazzi, C.

6. Event selection and data/MC comparison

I have done something very bad today by proposing a particle that cannot be detected; it is something no theorist should ever do.

Wolfgang Pauli

We now present the sample of neutrino candidates used for the search for point-like sources. The agreement between data and Monte Carlo is also discussed. First, we discuss the run selection and data taking conditions for the period of data considered

6.1. Selection of the runs

The data used for the analysis presented in this thesis were collected during the period January 2007 to December 2010. The first run considered started at 12:57 on January 28th 2007. Five strings were operational at that time. The data taking continued with this detector configuration until December 7^{th} 2007 when other five lines were deployed. During the period from the 3rd of March 2008 to the 25th of May, line 4 suffered a problem which prevented any communication with the junction box. Data taking in that period continued with 9 strings. During a three day sea campaign at the end of May 2008, line 4 was re-connected together with line 11 and line 12. On the 30th of May 2008 data taking with the completed detector started. Several detector maintenance operations occured during 2009 due to problems with Line 12, 9 and 6. Line 6 was recovered on the 27th of October 2009 and re-connected only one year after. The last run used in this analysis was taken on the 31st of December 2010. The run numbers, the corresponding data taking periods and the number of detector strings for each period are summarised in Table 6.1.

The run selection consists of a set of "sanity checks" of the data. This prevents inclusion of runs with synchronisation problems in the DAQ system and with non-physical event rates. Experimental runs which are used e.g. to test new calibrations or H.V. settings are also excluded. The livetime of the excluded runs is roughly 150 days.

The livetime of the selected 7419 runs is 813 days (of which 183 days correspond to the five line period as summarised in Table 6.2). The duty cycle is around 60% though it is increasing every year. Loss of efficiency is mainly due to periods of high bioluminescence and sea operations.

Run number	Data taking period	Number of lines
25700 - 30460	January - December 2007	5
30508 - 32491	December 2007 - March 2008	10
32529 - 34417	March - May 2008	9
34419 - 41671	May 2008 - October 2009	12
41673 - 52896	October 2009 - November 2010	11
52896 - 54250	May 2008 - December 2010	12

Table 6.1.: Data taking periods and the corresponding run numbers. The number of detector strings which were operational during each period is also shown.

Period/Configuration	Number of runs	Livetime [days]	Duty cycle
All	7419	813	58%
5 lines	1396	183	57%
12 lines	6023	630	60%
2007	1469	192	53%
2008	1987	181	50%
2009	1644	208	57%
2010	2319	232	64%

Table 6.2.: Livetime and number of runs for different data taking periods. The last column shows the data collection efficiency.

6.1.1. Sparking runs

A small number of runs contain events with an exceptionally high hit multiplicity. As an example, Figure 6.1 (left) shows the distribution of the number of hits for a run where this happens. It is believed that this is due to "sparking" OMs, i.e. OMs where the PMT suffers an high voltage surge. Although these high multiplicity events are not reconstructed with large values of Λ, meaning that they would not be selected as neutrino candidates, we veto these runs. Twenty-five runs were identified as "sparking", corresponding to roughly two days of livetime.

6.1.2. High baseline runs

In Figure 3.15 (top) we showed the distribution of the baseline for the period of data considered. Two periods of high baseline rate are clearly distinguishable. They correspond to the period of data taking May - August 2009 and August - September 2010. The Λ-distribution for events in these periods is compared to the rest of the data in Figure 6.2. It exhibits a contribution of events (around $\Lambda \simeq -7$) which is thought to consist of

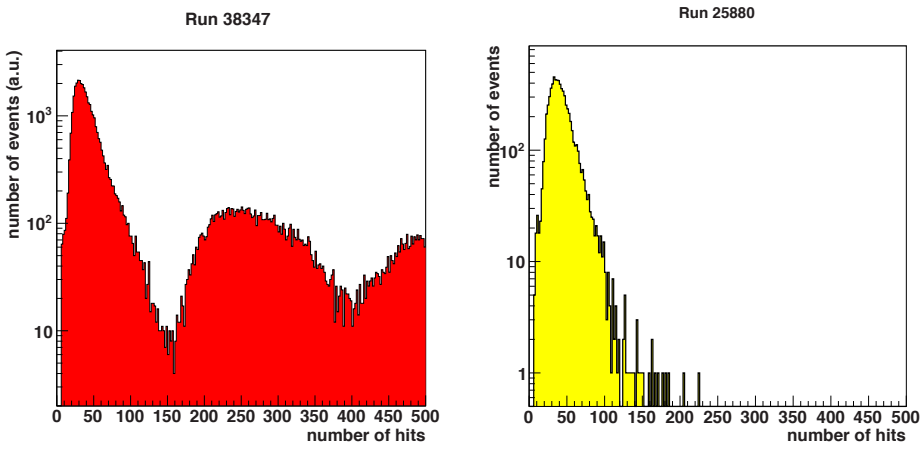

Figure 6.1.: Left: number of hits for the run 38447 showing an excess of high multiplicity. These excess is attributed to sparking OMs. Right: number of hits for run 25880 with no excess of high multiplicity.

only optical background, (the events at higher values of lambda are due to atmospheric muons). However, the selection of the sample relies on a cut on $\Lambda > -5.2$ which rejects these events.

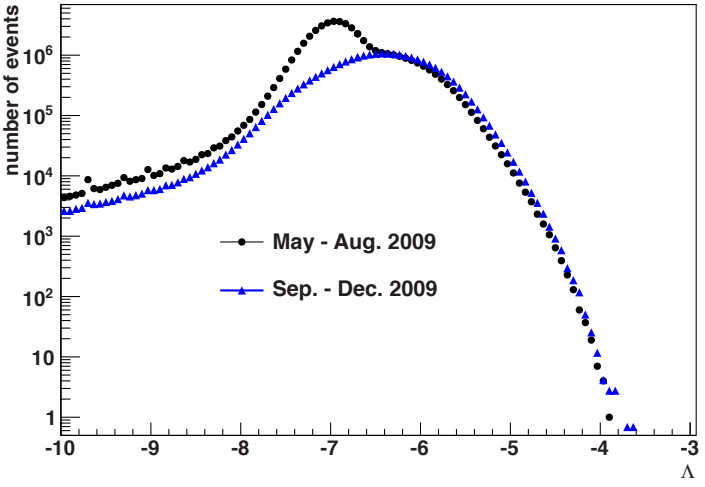

Figure 6.2.: Λ distribution for two periods of data taking. The peak around $\Lambda \simeq -7$ is explained by high optical background.

6.2. Event selection

The cuts applied to select neutrino candidates serve to reject the background due to cosmic rays. They are:

- $\theta < 90^\circ$. Neutrino candidates events are first selected requiring tracks reconstructed as upgoing. With this cut, and without applying any other, it has been estimated that 90% of the atmospheric muons and 20% of the atmospheric neutrinos are rejected.
- $\beta < 1^\circ$. The estimated angular uncertainty on the muon track direction, β, defined in Section 4.3, is required to be smaller than 1 degree. The error estimate on the direction of the reconstructed muon track is an important variable for the discrimination of the mis-reconstructed atmospheric muons. Figure 6.3 shows the distribution of β for both signal neutrinos reconstructed within 2 degrees of the true neutrino direction and for upgoing atmospheric muons. Most of the signal events have values of β smaller than 1 degree. The cut $\beta < 1^\circ$ rejects 47% of the mis-reconstructed atmospheric muons.

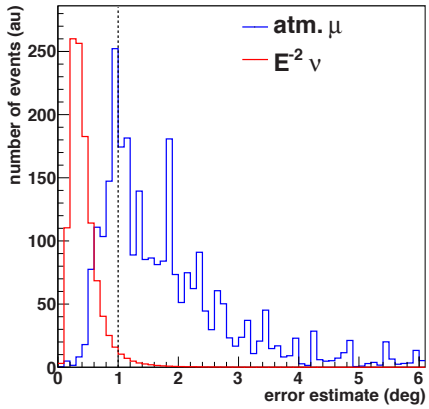

Figure 6.3.: Distribution of the estimated angular uncertainty, β, for E_ν^{-2} signal neutrinos (red) and for upgoing mis-reconstructed atmospheric muons (blue). The vertical dashed line indicates where the selection cut is applied ($\beta < 1^\circ$).

- $\Lambda > -5.2$. To further reject mis-reconstructed atmospheric muons, events with the quality variable Λ larger than -5.2 are selected. This value was chosen to optimise the neutrino flux needed to have a 50% chance of discovering the signal at the 3σ (or 5σ) significance level assuming an E_ν^{-2} spectrum. Table 6.4 shows the values of the flux needed for discovery for a ν-source at three declinations [1]. Among the three Λ cuts, $\Lambda > -5.2$ gives the lowest, i.e. best, discovery potentials.

The effect of the selection cuts on data, expected background and signal efficiency are summarised in Table 6.3

The selected sample consists of 3058 neutrino candidate events out of a total of roughly 4×10^8 triggered events. From Monte Carlo simulations, it has been estimated that the

[1] The values presented here are for a candidate list search (see Chapter 7).

Cut	Data	Atm. μ	Atm. ν	$E_\nu^{-2}\nu$[%]	$E_\nu^{-2}\nu(\alpha < 1^o)$[%]
Triggered events	3.94×10^8	3.06×10^8	1.54×10^4	100	100
$\theta < 90^\circ$	6.08×10^7	2.98×10^7	1.24×10^4	61	57
$\theta < 90^\circ + \beta < 1^\circ$	3.90×10^7	1.57×10^7	8352	44	53
$\theta < 90^\circ + \beta < 1^\circ + \Lambda > -5.2$	3058	358	2408	23	44

Table 6.3.: Number of events before and after applying the selection cuts described in the text for data (second column) and Monte Carlo simulations. The fourth column shows the percentage of signal events assuming a neutrino flux proportional to an E_ν^{-2} spectrum. The last column the same but for signal events reconstructed within 1 degree from the true direction.

$\Lambda >$	δ [°]	$n_{5\sigma}$	$\phi_{5\sigma}$ [$\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$]
-5.0	-70	4.43	6.42×10^{-8}
-5.2	-70	5.08	6.09×10^{-8}
-5.4	-70	6.42	6.7×10^{-8}
-5.0	-30	4.33	9.19×10^{-8}
-5.2	-30	4.82	8.36×10^{-8}
-5.4	-30	5.83	8.74×10^{-8}
-5.0	10	3.88	1.17×10^{-7}
-5.2	10	4.29	1.04×10^{-7}
-5.4	10	6.38	1.33×10^{-7}

Table 6.4.: For three different declinations (second column) and cuts on Λ (first column) discovery potentials are computed: mean number of signal events (fourth column) and flux (last column) needed to claim a 5σ discovery. The cut $\Lambda > -5.2$ optimises it.

atmospheric muon contamination of this sample is around 14% (see Table 6.3). A Galactic sky map of these events is given in Figure 6.4.

6.3. Data - Monte Carlo comparison

The comparison between data and simulations is an important step for the kind of analysis. Good agreement between data and MC represents a hint of good understanding of the physics processes and detector response of the experiment.

The Monte Carlo used in this analysis corresponds to a run-by-run simulation [151]. Using the measured optical background rates, OM conditions and run duration, a realistic simulation of the physics and data taking process for each run is obtained.

Figure 6.5 shows the cumulative distribution of the Λ variable for upgoing tracks. The cut $\beta < 1^\circ$ is also applied. Overall, the agreement between data and the simulations is

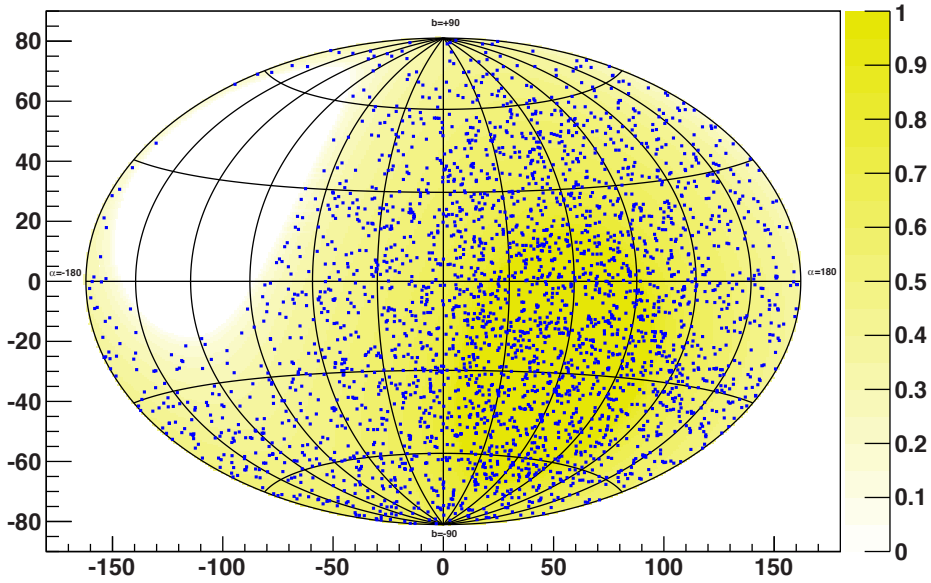

Figure 6.4.: Galactic skymap of the 3058 neutrino candidate events. The different shades of yellow indicate different visibility at the ANTARES site.

good, well within the systematic uncertainties. The excess of data compared to the Monte Carlo at the lowest value of Λ is due to the non-simulated contribution of events consisting of solely optical background as already discussed in Section 6.1.2.

Figure 6.6 shows the distribution of the selected run numbers. In general, the predictions underestimate the data and the ratio data/MC is around 1.4 for most of the period considered. The two periods of high bioluminescence responsible for the peak at $\Lambda \simeq -7$ in Figure 6.2 are clearly distinguishable from the large discrepancy between data and Monte Carlo. The zenith angle distribution with (bottom) and without quality cuts (top) is shown in Figure 6.7. The agreement between data and MC improves after requiring well reconstructed tracks. The azimuth angle distribution is shown in Figure 6.8 again for triggered events only (top) and event passing the quality cuts $\Lambda > -5.2$ and $\beta < 1^\circ$ (bottom).

Finally, in Figure 6.9 and Figure 6.10 we show the distribution of the angular error and the number of hits respectively. For the first plot, upgoing tracks with $\Lambda > -5.2$ are selected. For the distribution of the number of hits, all the final cuts are applied. Five events with large values of N_{hits} are not modelled by the simulations. These five events have high values of Λ, meaning that it is unlikely that they are mis-reconstructed upgoing muons. We cannot exclude that these events are due to an unknown sparking OM although this is unlikely since sparking runs have lower values of Λ. Table 6.5 shows the run number, the frame index, the values of N_{hits} and Λ, and the number of lines in the reconstruction for these events.

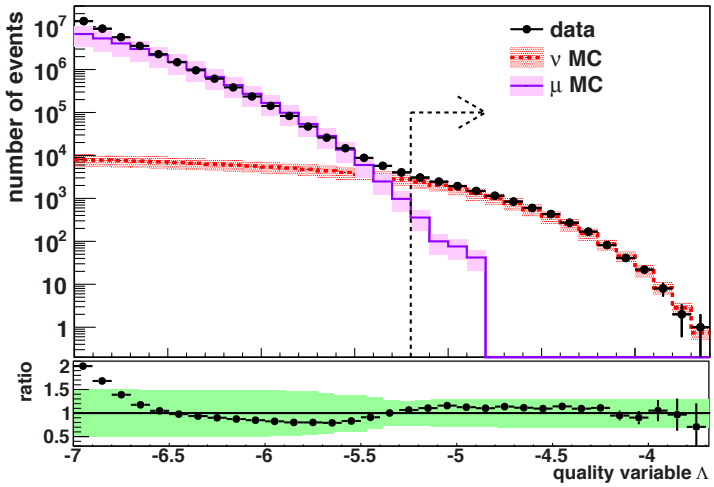

Figure 6.5.: Cumulative distribution of Λ for upgoing tracks. The cut on the angular error estimate $\beta < 1^\circ$ is also applied. The arrow shows where the selection cut is applied ($\Lambda > -5.2$). The magenta histogram represents atmospheric muons simulated with the MUPAGE package and the red histogram is for atmospheric neutrinos (upgoing and downgoing) weighted with the Bartol flux. The ratio of the data over the Monte Carlo is also shown. This is also valid for the following plots.

Evt ID	Run number	Frame index	$N_{\rm hits}$	Λ	$N_{\rm lines}$
1	35473	7183	152	-4.7	11
2	35583	58091	157	-4.8	8
3	46018	14072	150	-4.5	7
4	49420	33730	148	-4.7	9
5	28702	77100	147	-4.8	5

Table 6.5.: For the five events with large values of number of hits shown in Figure 6.10, we report in this table the run number (second column), the frame index (third column), the value of $N_{\rm hits}$ (fourth column), the value of Λ (fifth column) and the number of lines hitted by Cherenkov photons emitted by these events (last column).

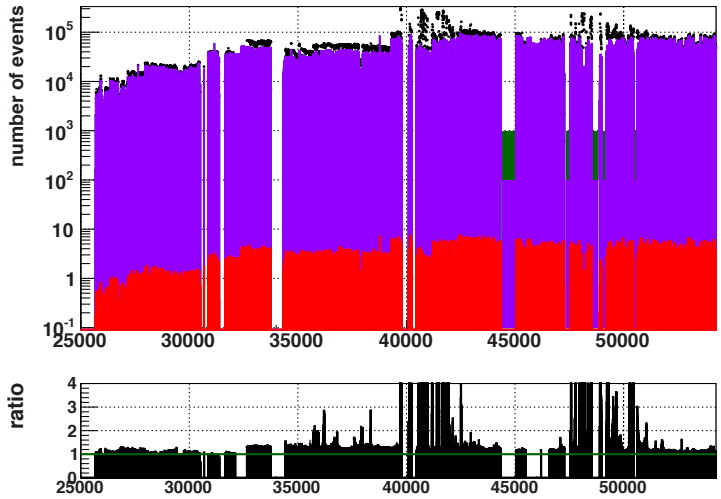

Figure 6.6.: Run number distribution of all reconstructed events passing either the 2T3 or the 3N trigger. This plot illustrated the run-by-run MC scheme as all runs in the data are simulated. Data are in general underestimated by the MC predictions. The two period of high baseline discussed in Section 6.1.2 can be identified by the large data - MC discrepancy.

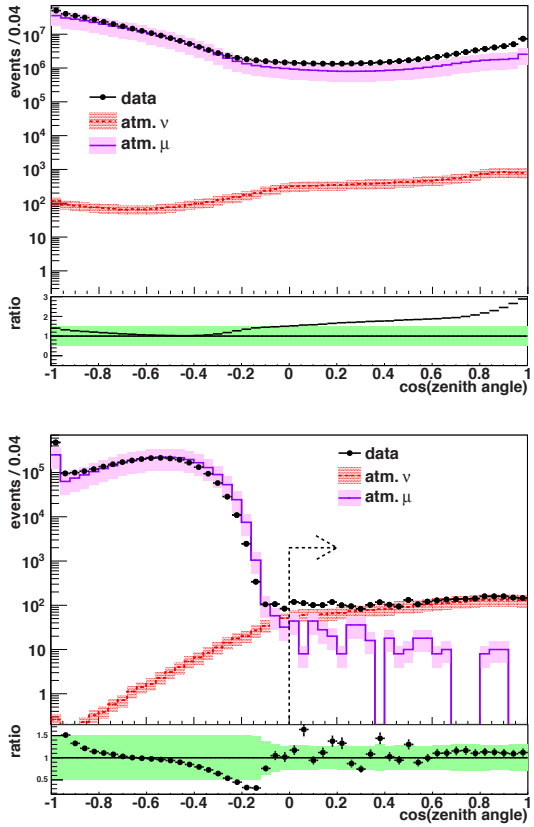

Figure 6.7.: Distribution of the cosine of the zenith angle for all triggered events (top) and for events passing the quality criteria $\Lambda > -5.2$ and $\beta < 1°$ (bottom). The dashed line shows the upgoing the events. The magenta and red bands show the systematic uncertainties on the simulation that were discussed in Chapter 5. In both plots, the bottom panel shows the ratio between data and MC. The green band the total systematic uncertainty.

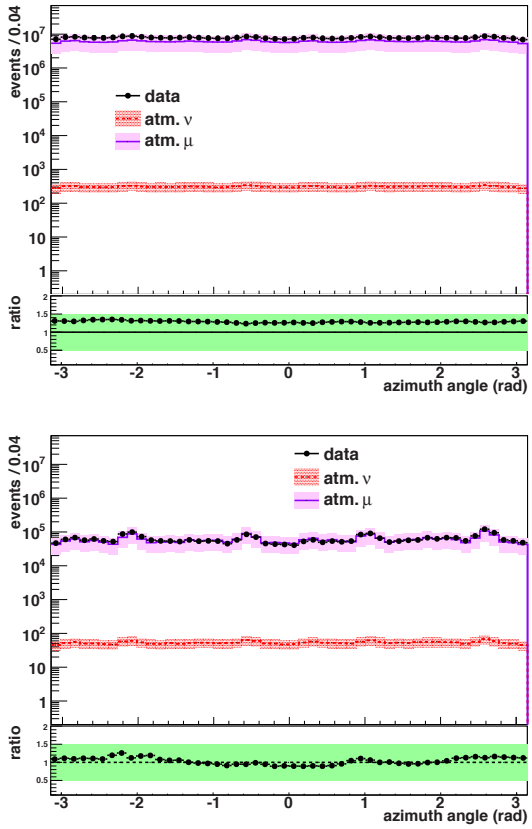

Figure 6.8.: Distribution of the azimuth angle for all triggered events (top) and for events passing the quality criteria $\Lambda > -5.2$ and $\beta < 1^\circ$ (bottom). The magenta and red bands show the systematic uncertainties on the simulation. The bottom panel in both plot shows the ratio of data over the MC. The green band indicates the total systematic uncertainty associated.

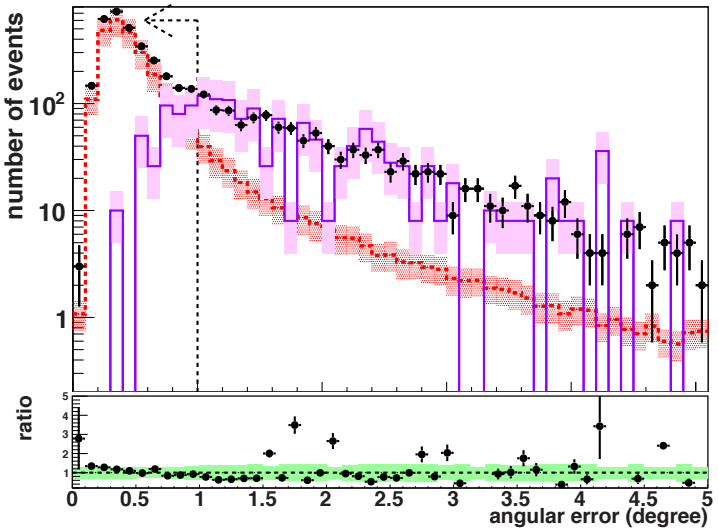

Figure 6.9.: Distribution of the angular uncertainty for reconstructed upgoing tracks after applying a quality cut $\Lambda > -5.2$. The arrow shows where the selection cut is applied.

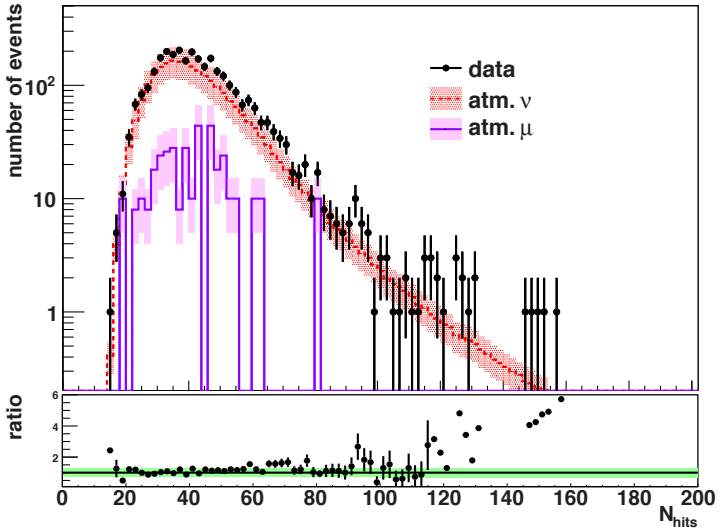

Figure 6.10.: N_{hits} distribution for upgoing selected events after applying the quality selection described in the text. Five events with large values of N_{hits} are visible at the tail of the distribution.