Measurement of the atmospheric neutrino energy spectrum

Palioselitis, D.

Citation for published version (APA):
EVENT SELECTION

The present chapter contains information about data selection, the background rejection of atmospheric muons and the event selection criteria applied to define the final data sample. The vast majority of events detected in ANTARES are due to downward-going atmospheric muons that dominate over the incoming neutrino flux by many orders of magnitude. As we will see in the following sections, rejecting the events arriving above the horizon is not enough to purify the final sample and additional selection criteria need to be introduced. Section 5.1 contains the description of the selection of runs for this analysis, taking into account various characteristics of the data taking conditions. In section 5.2 we discuss the atmospheric muon background and define the cuts used in this analysis. Sections 5.3 and 5.4 contain a series of comparisons between data and Monte Carlo which allow for an evaluation of our understanding of the detector and the physical processes that take place.

5.1 Data selection

The data used for this analysis were collected during the period May 2008 to December 2010 with a 12-line detector configuration. The run numbers and the corresponding data taking periods are shown in table 5.1.

There are various quantities that characterize the quality of the data taking conditions, such as the percentage of active optical modules during a run or the hit rates. The baseline of a run is defined as the mean of a gaussian fit on the L0 rate of each PMT, averaged over all PMTs. The percentage of TimeSlices where the rate was higher than the baseline plus 20% is averaged over all PMTs to give the burst fraction. The mean rate is the average rate over all ARS’s that measured more than 10 kHz. Quality Basic (QB) is a flag that characterizes the quality of the data taking conditions, based on optical background and detector state information. It is defined as follows:

- QB = 1 : Basic selection of runs available for physics analyses.
- QB = 2 : At least 80% of the OMs that are expected to work are active.
5. Event selection

<table>
<thead>
<tr>
<th>Run number</th>
<th>Data taking period</th>
</tr>
</thead>
<tbody>
<tr>
<td>034419 - 035878</td>
<td>May - September 2008</td>
</tr>
<tr>
<td>035880 - 038230</td>
<td>October - December 2008</td>
</tr>
<tr>
<td>038233 - 039775</td>
<td>January - March 2009</td>
</tr>
<tr>
<td>040098 - 041626</td>
<td>April - June 2009</td>
</tr>
<tr>
<td>041628 - 043534</td>
<td>July - September 2009</td>
</tr>
<tr>
<td>043540 - 045536</td>
<td>October - December 2009</td>
</tr>
<tr>
<td>045540 - 047694</td>
<td>January - March 2010</td>
</tr>
<tr>
<td>047706 - 049766</td>
<td>April - June 2010</td>
</tr>
<tr>
<td>049770 - 052305</td>
<td>July - September 2010</td>
</tr>
<tr>
<td>052307 - 054250</td>
<td>October - December 2010</td>
</tr>
</tbody>
</table>

Table 5.1: Data taking periods in groups of three months and the corresponding run numbers. The physics runs are not consecutive since calibration and acoustic data taking runs are excluded.

- QB = 3 : Baseline $\leq$ 120 kHz and burst fraction $\leq$ 0.4.

- QB = 4 : Baseline $\leq$ 120 kHz and burst fraction $\leq$ 0.2.

These run sets are cumulative, i.e. runs that satisfy QB=3, satisfy QB=2 as well. All runs from the database with QB $\geq$ 1 are chosen for the final data sample. Additionally, calibration runs and runs that were used for tests or high voltage tunings, as well as runs where both T3 and 3N triggers were absent are excluded from the analysis. In rare occasions electrical breakdowns can occur in the PMT’s causing them to spark, producing a signal similar to a hadronic or electromagnetic shower. Runs where such “sparking events” were identified are also excluded. Only runs that are longer than 5 minutes and contain more than one hundred events are kept. The final data sample consists of 5037 runs corresponding to 517.8 days of live time. After the application of all quality cuts, described in the following sections, 831 neutrino events were found. For the energy reconstruction of each event, only working optical modules were considered, while OM’s with unusually high or low rates were excluded from the fit.

Distributions of the baseline, burst fraction, mean number of active OM’s and mean rates for the final run selection are plotted against run number in figure 5.1. The trend observed on the mean number of active optical modules is due to the fact that some lines were not operational for certain periods of time as discussed in section 2.9. The hit rates remain stable and at a low level (less than 80-90 kHz for the baseline) for most of the used runs, with some short periods of increased optical background activity. Around 33% of the runs used have a burst fraction higher than 0.4, 81% of which have less than 80% of the expected active OMs actually operational.
Figure 5.1: The mean number of active OMs, the mean rates and the baseline rate in kHz, and the burst fraction for all the runs of the period May 2008 to December 2010 that are used in this analysis. On top of the figure the periods where certain lines were not operational are indicated. These are responsible for the drops and rises on the number of active OMs.
5.2 Atmospheric muon background rejection

The goal of the event selection is to obtain a neutrino candidate sample with the following characteristics:

- well reconstructed tracks
- high purity sample of neutrino events i.e. low atmospheric muon contamination and
- as large an efficiency as possible in order to keep statistics high.

5.2.1 Atmospheric muon background

The atmospheric muon flux is the dominant background in the search for neutrino candidate events as shown in figure 5.2. The rates that correspond to atmospheric muons

![Figure 5.2: Simulated atmospheric muon ($\mu^\pm$) and neutrino ($\nu_\mu + \bar{\nu}_\mu$) induced muon events per year livetime at the detector level. No cuts are applied. The atmospheric muon background dominates the neutrino flux, being around four orders of magnitude higher. The energy range shown here is from $10^2$ GeV to $10^{5.5}$ GeV, divided in 10 bins.](image)

lie $\sim$ 4 orders of magnitude above the neutrino rates. The first step towards the reduction of this background is to reject all events that are reconstructed as downward-going. Atmospheric muons that are created in the atmosphere above the horizon will reach the detector, leading to a downward-going event signature. On the other hand, atmospheric muons created below the horizon, that should give an upward-going event signature, will never be able to traverse the Earth and reach the detector. This directional selection decreases the rates of atmospheric muons by an order of magnitude, as shown in
5.2 Atmospheric muon background rejection

Figure 5.3: Simulated atmospheric muon ($\mu^+$) and neutrino ($\nu_\mu + \bar{\nu}_\mu$) induced muon events per year livetime at the detector level. Events reconstructed as downward-going are rejected. No additional quality cuts are performed. The muon background is still higher than the muon neutrino rates due to mis-reconstructed atmospheric muon contamination. The energy range shown here is from $10^2$ GeV to $10^{5.5}$ GeV, divided in 10 bins.

Figure 5.3 where only upward-going events are shown. This step is clearly not sufficient for the complete suppression of the muon background, since the muon rates are still a few orders of magnitude above the neutrino flux. The reason for this is that many atmospheric muon events are mis-reconstructed as upward-going, thus faking neutrino events. In order to reject these muon events, we need to examine the quality of the track fit in more detail.

The left and right plots of figure 5.4 show the relation between the reconstructed zenith angle and the track reconstruction quality parameter $\Lambda$ for atmospheric muons and neutrinos respectively. Most of the atmospheric muon events belong to the upper half of the left plot in figure 5.4, corresponding to downward-going events. The ones that have been reconstructed as upward-going (lower half) tend to have lower values of $\Lambda$ as well. This is expected since we have seen that lower values of $\Lambda$ correspond to less well-reconstructed tracks. Upward-going neutrinos on the other hand, can take higher $\Lambda$ values, indicating well-reconstructed tracks. Figure 5.5 illustrates what this situation looks like for real data. Both muon and neutrino events populate this plot. A cut rejecting all downward-going events suppresses a large fraction of the atmospheric muon flux. An additional $\Lambda$ cut rejects the rest of the atmospheric muon contamination, leaving the well-reconstructed upward-going neutrinos still in the sample. Finally, a cut on the estimated angular error from the track reconstruction algorithm is applied, in order to keep well-reconstructed events with small uncertainty in their direction. This
is relevant for the reconstruction of the energy, where well-reconstructed tracks provide the best results.

The cumulative distribution of the \( \Lambda \) value for simulated muon and neutrino events reconstructed as upward-going is shown in figure 5.6. Due to low statistics on the simulation of atmospheric muons, there are very few entries for high \( \Lambda \) values (see section 3.2). In order to circumvent this problem and determine the most appropriate cut at \( \Lambda \) successfully, we fit the tail of the \( \Lambda \) cumulative distribution for muons with an exponential function, indicated in the figure with a dashed line. A cut at \( \Lambda = -4.8 \) suppresses the atmospheric muon background contamination to less than one percent.

5.2.2 Selection criteria

We summarize here the cuts used, in this analysis, to reject the atmospheric muon background and select the most well-reconstructed events. The selection criteria chosen are the following:

- The event must be reconstructed as upward-going, i.e. \( \cos \theta_{\text{rec}} < 0^{\circ} \).
- The value of \( \Lambda \) must be larger than \(-4.8\).
- The angular error estimate \( \hat{\beta} \) from the track fit must be smaller than one degree.

After applying these selection criteria to the data sample and comparing it to the Monte Carlo distribution we find an overall normalization difference of 22%. Figure 5.7 shows the distribution of the cosine of the reconstructed zenith angle for data and Monte Carlo for the whole period of 517.8 days. The neutrino simulation is lower by 22% as indicated by the linear fit on the ratio of data over Monte Carlo shown in figure 5.8. This is within the theoretical normalization uncertainties of the level of 30%.

![Figure 5.4: Zenith angle of the reconstructed track as a function of \( \Lambda \). The z-axis stands for the number of events per bin. The left plot corresponds to the atmospheric muon background and the right one to the atmospheric neutrinos. The entries are normalized to one year livetime.](image-url)
5.2 Atmospheric muon background rejection

Figure 5.5: Zenith angle of the reconstructed track as a function of $\Lambda$ for one year of data. The $z$-axis stands for the number of events per bin. This plot contains atmospheric muon as well as neutrino events. Events with $\cos \theta_{\text{zenith}} < 0$ and $\Lambda > -4.8$ are kept.

Figure 5.6: Cumulative distribution of the track fit quality parameter $\Lambda$ for atmospheric muons ($\mu^\pm$) and neutrinos ($\nu_\mu + \bar{\nu}_\mu$). The dashed line indicates the exponential fit used to estimate the muon contribution at higher $\Lambda$ values (see text) and the straight vertical line denotes the value of the cut.

the remainder of this work, the neutrino Monte Carlo is scaled up to match the data. Such an overall normalization will not affect the unfolded result, as discussed in section 4.4.

Let us now consider the efficiency of the aforementioned selection criteria and the purity of the selected sample. The efficiency of a selection process is defined as the fraction of the size of the selected sample corresponding to the signal after the cuts, over the size of the initial signal sample before any cuts:

$$\text{Efficiency} = \frac{N_{\text{signal}}^{\text{cut}}}{N_{\text{signal}}^{\text{all}}}.$$  \hspace{1cm} (5.2.1)
For the purposes of this analysis, all neutrinos are considered to be our signal while atmospheric muons constitute our background. The purity of a selected sample is defined as the ratio of the selected sample size corresponding to the signal over the size of the selected sample, counting both signal and background:

$$\text{Purity} = \frac{N_{\text{signal}}^{\text{cut}}}{N_{\text{cut}}^{\text{signal}} + N_{\text{cut}}^{\text{background}}}.$$  (5.2.2)

The atmospheric muon and neutrino simulation rates per year, as well as the corresponding rate calculated from data are shown in table 5.2. The last two columns give the efficiency and purity levels. Sample I refers to the initial sample of all reconstructed tracks. Sample II is the resulting sample after rejecting the downward-going tracks, suppressing the atmospheric muon background by around an order of magnitude. After the $\Lambda$ and $\beta$ cuts, we obtain sample III, where the atmospheric muon flux is almost completely suppressed, leading to a pure neutrino sample. The last sample contains only those events whose energy was reconstructed. This final selection has practically no effect at such an advanced stage of the selection process. This was anticipated already due to the high efficiency of the energy reconstruction algorithm (see figure 3.29).

### 5.3 Data - Monte Carlo comparisons

In order to verify that the detector and the physics that govern the experiment are well understood it is necessary to perform a series of comparisons between data and Monte Carlo. In what follows, the atmospheric neutrino flux by Honda and Martin et. al is used to provide the Monte Carlo distributions normalized for the data livetime of 517.8 days and multiplied by a scale factor of 1.22 as discussed in section 5.2.2.
5.3 Data - Monte Carlo comparisons

We begin the examination of the data and Monte Carlo distributions by looking at the angular information provided by the track reconstruction. The zenith and azimuthal distributions before the selection process and after the application of all quality cuts are shown in figures 5.9 and 5.10 respectively.

The applied cuts completely suppress the atmospheric muon flux, visible in both figures on the right, where only atmospheric neutrino events remain. The flatness of the azimuthal distribution reflects the symmetry of the neutrino flux for this angle.

Figure 5.8: Distribution of the ratio of data over Monte Carlo for the cosine of the reconstructed zenith angle. A linear fit indicates an overall normalization difference of 22%.

Figure 5.9: Zenith angle of the reconstructed track for data and Monte Carlo. A cosine of -1 corresponds to tracks traveling vertically upward. The left plot includes all events. Events after selection are shown on the right plot.
Table 5.2: Atmospheric muon and neutrino rates per year, for different samples. The efficiency and purity of the selection is also shown in the last columns. Samples correspond to all reconstructed tracks (sample I), tracks reconstructed as upward-going (sample II), events after the Λ and ˆβ cuts (sample III) and events where the energy reconstruction algorithm succeeded in finding a most probable energy (sample IV).

The situation is different for the zenith distribution which drops as we approach the horizontal direction. This is due to the different path length neutrinos have to travel as they traverse the Earth. Vertically upward neutrinos have a higher probability of interacting in rock and producing a detectable muon, whereas for horizontal neutrinos this probability is lower. The distribution of the track fit quality parameter Λ as well as its cumulative distribution for data and Monte Carlo are shown in figure 5.11. Only tracks reconstructed as upward-going are included in these plots. It is evident that an additional Λ cut at this stage is sufficient to eliminate the atmospheric muon contamination in the sample.
5.3 Data - Monte Carlo comparisons

Figure 5.11: Left plot: \( \Lambda \) distribution for data and atmospheric muon and neutrino simulations. Right plot: \( \Lambda \) cumulative distribution for data and simulations. Only tracks reconstructed as upward-going are shown here. No additional quality cuts are applied.

Figure 5.12: Distribution of the angular error estimate \( \hat{\beta} \) from the track fit for data and Monte Carlo. The left plot shows all events while the right plot shows all the events that passed the selection criteria.

Figure 5.12 shows the distribution of the angular error estimate \( \hat{\beta} \) from the track fit. The left plot shows the data and Monte Carlo distributions for all track reconstructed events while on the left only the final selection is shown. Finally, the distributions for the number of lines and the number of hits used in the final stage of the track fit, for both data and Monte Carlo, are shown in figures 5.13 and 5.14 respectively. In both figures, the right plot consists only of neutrino events, since atmospheric muons are completely suppressed after the directional and \( \Lambda \) cuts. There is an overall good agreement between data and Monte Carlo, with only a normalization difference of 22%.
5. Event selection

Figure 5.13: Distribution of the number of lines that were used in the final fit stage of the track reconstruction. On the left plot all events are shown. The right plot includes only events that passed the selection cuts.

Figure 5.14: Distribution of the number of hits used in the final fit stage of the track reconstruction. The left plot includes all events. Events reconstructed as upward-going that passed the quality cuts are shown on the right.

5.4 Distribution of the energy observable

The most important quantity for the unfolding of the atmospheric neutrino spectrum is the reconstructed energy, i.e. our energy observable. This section contains the comparisons between data and simulations regarding the output of the energy reconstruction algorithm described in detail in chapter 3. Figure 5.15 shows the distribution of the energy reconstruction variable for upward-going events with an angular error estimate of less than one degree and different Λ cuts. The sample is dominated by atmospheric muons especially at higher energies when Λ > −5.6. The reconstructed energy distribution in data follows the simulations as more strict Λ cuts are applied and the atmospheric muon background is increasingly suppressed. For Λ > −5 there is still a small atmospheric muon component present. The sensitivity of the unfolded result for different choices of Λ in the vicinity of Λ = −4.8 will be examined in the next chapter, in section 6.2.

After applying the event selection cuts described in the previous sections, the at-
Figure 5.15: Distribution of the reconstructed energy variable for data and simulations. All events reconstructed as upward going and with $\hat{\beta} < 1^\circ$ are shown for the different $\Lambda$ indicated in each figure. The atmospheric muon Monte Carlo sample corresponds to 1/10th of the livetime and each event is weighted accordingly (see section 3.2).

Atmospheric muon flux is completely suppressed. This is shown in figure 5.16. In this plot we see the distribution of the energy reconstruction variable in data, following the Monte Carlo prediction for the atmospheric neutrino energy distribution. The agreement between the reconstructed energy distributions for data and neutrino simulations is satisfactory. This energy spectrum is the distribution that is used as the measured input vector in the unfolding process described in chapter 4.

### 5.5 Summary

In order to unfold the spectrum of atmospheric neutrinos we need to have a pure neutrino sample. The aim of the event selection process introduced in this chapter was to eliminate as much as possible the presence of atmospheric muon contamination in the final sample. Additionally, well-reconstructed tracks have a higher probability of
5. Event selection

Figure 5.16: Distribution of the reconstructed energy for data and Monte Carlo after the application of the selection criteria described in section 5.2.2. No atmospheric muon events are present in the sample.

having their energy reconstructed accurately. The selection criteria that satisfy these conditions consist of selecting upward-going tracks, with Λ and $\hat{\beta}$ values that indicate a high quality track fit. The purity of the final sample is almost 100% with an efficiency over the total number of upward-going events of $\sim$8%. Finally, the data - Monte Carlo comparisons show a satisfactory level of agreement, indicating a good understanding of the detector and the physical processes taking place.