Investigation and simulation of cascade-like neutrino events in ANTARES

Bachelorarbeit aus der Physik

vorgelegt von
Michael Moser

30. November 2014

Erlangen Centre for Astroparticle Physics
Physikalisches Institut
Friedrich-Alexander-Universität Erlangen-Nürnberg

1. Gutachter: Dr. Thomas Eberl
2. Gutachter: Prof. Dr. Gisela Anton

Abstract

In this work 59 shower neutrino events found in 6 years of ANTARES data are investigated. The investigation is done by means of a dedicated simulation of each individual event using the neutrino parameters extracted from measured data as an input to the simulation. The conclusion is that out of the 59 events there are several events with "mirror" solutions in the Dusj reconstruction algorithm and that the algorithm underestimates the energy of events by about 30%. Several methods such as a hit pattern approach are developed to distinguish between a "well" and a "badly" reconstructed event. Finally, it is demonstrated that a categorization of shower events in terms of "goodness" is now possible by employing the different methods that were designed.

Contents

1. Observation of a cosmic neutrino flux

At all times humans have been striving to investigate the cosmos by detecting incident light. This was only possible because photons, as a carrier of the electromagnetic force, can be detected easily.
However, this is also a major drawback as photons are likely to be absorbed in e.g. dust clouds before they can reach Earth. That's also the reason why we can only observe the outer layers of our sun because any light produced in the core will interact with other particles, e.g. in the outer layer, so that it may take over tens of thousands of years for a photon to escape from the sun [1].
To avoid this kind of problem one can try to detect neutrinos because they only interact by the weak force, but on the other hand this also makes detection a lot harder. Nevertheless, the ICECUBE collaboration succeeded in measuring a cosmic neutrino flux of $E^2 \cdot \Phi_{90\%} = 0.95 \pm 0.3 \cdot 10^{-8} \mathrm{GeV\,cm^{-2} \cdot sr^{-1} \cdot s^{-1}}$ per flavor at the 5.7 σ level in the 100 TeV – PeV range with their detector located in Antarctica [2]. The energy spectrum of the corresponding high-energy cosmic neutrino events is shown in figure 1.

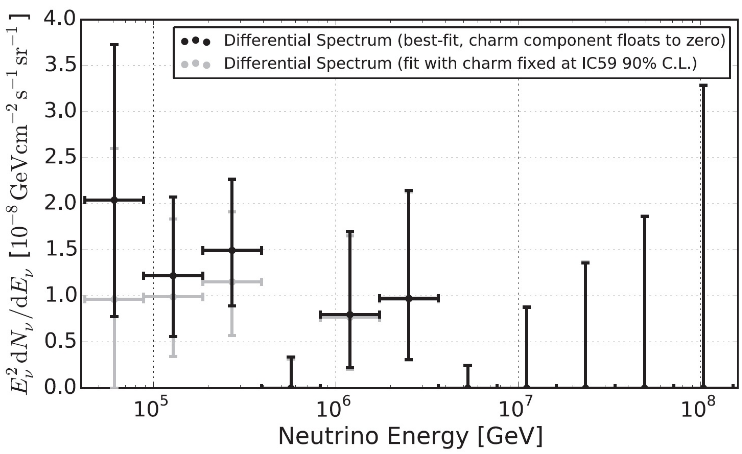


Figure 1: Extraterrestrial neutrino flux as a function of energy as detected by IceCube [2]

If the results are correct one should also be able to measure this flux in other neutrino observatories like ANTARES. In this context, Florian Folger conducted a search for a diffuse cosmic neutrino flux in his doctoral thesis [3] using the ANTARES data by developing a reconstruction algorithm for shower events.
Before continuing with the results of this work let's clarify that there are 2 kinds of events: track-like and shower-like events. Track-like events occur in charged current muon neutrino interactions with nucleons in which a hadronic shower and a muon is produced that can be detected by the resulting Cherenkov radiation in water [3]. Shower-like events can occur in both neutral current and charged current interactions. In neutral current events electron, myon or tau neutrinos produce a hadronic shower whereas in

charged current events electron or tau neutrinos produce a hadronic and an electromagnetic shower. As the sensitivity of the ANTARES detector hasn't yet been high enough to observe a significant signal contribution with shower events, a 90 % confidence upper limit on the cosmic neutrino flux was set that yields a value per neutrino flavour of $E^2 \cdot \Phi_{90\%} = 4.9 \cdot 10^{-8}\mathrm{GeV\,cm^{-2}} \cdot \mathrm{sr}^{-1} \cdot \mathrm{s}^{-1}$ [3].
Nevertheless, eight neutrino candidates with an energy above 10TeV were found [3] which is in line with an expected value of five atmospheric and three cosmic neutrinos considering the cosmic neutrino flux measured by IceCube. Unfortunately, the significance of these three events is not yet high enough, as the significance of measured excess over background is only 1.5σ, which lies well within systematic and statistical errors.
As a result of this ANTARES measurement there remain a few open questions. For example it is not clear whether some of the reconstructed neutrino events aren't just misreconstructed atmospheric muons because in some circumstances track-like atmospheric muons could be interpreted as showers by the reconstruction algorithm. Another question that needs to be answered is if the estimated energy and angular resolution is correct.
As a consequence of these issues, we investigate the quality of the so-called "Dusj" reconstruction algorithm developed by Florian Folger. The aim of this work is to review the detected neutrino events by taking the characteristics of these events and to use them as input parameters for the ANTARES detector simulation so that one can evaluate the reconstruction quality.

2. The ANTARES neutrino telescope

The detector whose data is used is the ANTARES neutrino telescope located off the French Mediterranean coast. The ANTARES detector consists of 12 strings with each carrying 25 "storeys" equipped with three optical modules, an electronics container and calibration devices where necessary [4]. A visualization of the layout can be seen in figure 2.

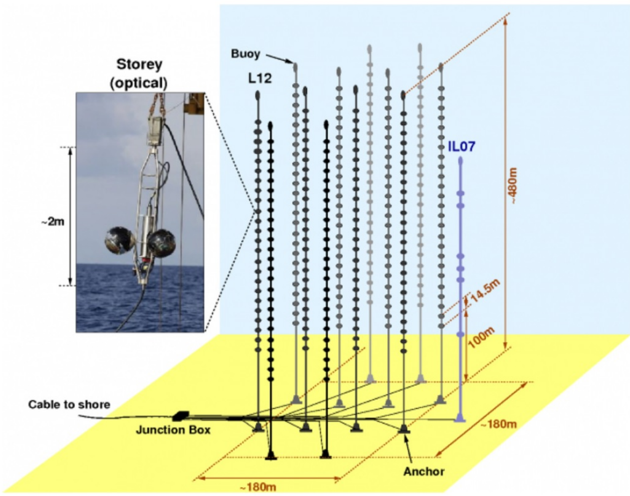


Figure 2: Schematic of the ANTARES neutrino telescope [4]

With this detector layout one can measure both track-like and shower-like events.
However, as there are a lot of atmospheric background muons entering the detector from the top [3] that could be misidentified as neutrino track events, the photomultipliers (PMTs) inside the optical modules (OMs) look down at an angle of 45° [4]. By using this geometry the Earth itself can be used as a background shield because neutrinos at the interesting energies can travel through all layers of our Earth whereas background like atmospheric muons cannot.
Each of the optical modules records a signal when a charge threshold of typically about 0.3 photo-electrons (pe) measured by a PMT, is surpassed [3]. Subsequently an analogue ring sampler (ARS) collects the charge caused by all incident photons within an integration time of 25ns [3]. Now the amplitude of the signal hit can be defined as the accumulated charge and the signal time of the hit as the time of the threshold crossing [3]. After the integration time window writing and sending the hit information results in a dead time of about 250ns [3]. However, there is a second ARS integrated in every OM that can take over the job of the other ARS at around 40ns after the integration time of the first ARS began [3].
Each of the photomultipliers has a time resolution of about 1.3ns [4] so that the arrival times of (Cherenkov-) photons can be precisely measured.
To synchronize the signal time measurements at the photomultipliers several calibration

techniques like pulsed light emission are used achieving an accuracy of about 0.5ns [4]. As the optical modules are moving with a varying sea current the position of each optical module needs to be thoroughly verified. For the exact calibration of position and orientation an "[...] acoustic system measuring running times of acoustic pulses between transmitters on the sea floor and receivers (hydrophones) on 5 storeys per string" [4] is used achieving a resolution of about 10cm which translates into a timing uncertainty of 0.5ns [4].
With the knowledge of time and position one can now e.g. reconstruct the muon trajectories of track-like neutrino events by measuring the arrival time of the resulting Cherenkov photons with the individual photomultipliers, because the muon is moving faster than light in water.

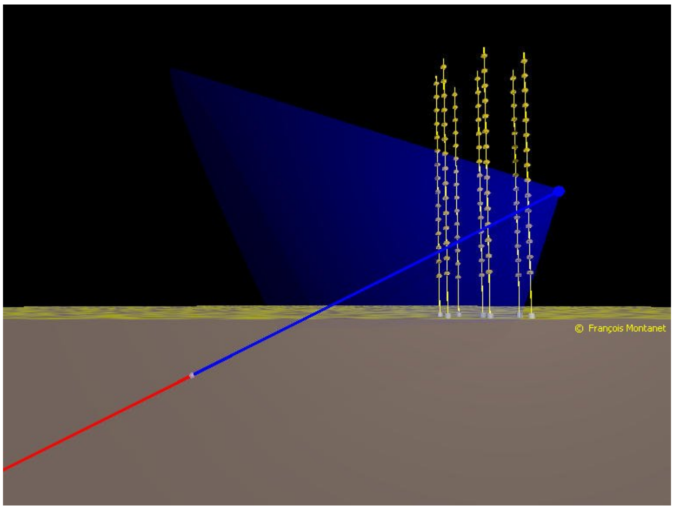


Figure 3: ANTARES detector principle: red: neutrino, blue: muon, blue area: Cherenkov cone [5]

Of course shower events can be reconstructed in the same way, as used in this work.
One drawback of using the Mediterranean Sea as a detector medium compared to ice as in IceCube is the strong background light caused by e.g. ^{40}K decays and bioluminescence. The decay of ^{40}K results in a rate of about 30kHz per photomultiplier [4] and bioluminescence accounts for a varying baseline rate of roughly a few 10kHz to a few 100kHz [4]. However there are also short bursts of bioluminescence in the MHz region [4], probably caused by bioluminescent animals.

3. The Dusj-Shower-Reconstruction

The Dusj-Reconstruction algorithm developed by Florian Folger consists of several modules integrated in the collaboration internal SeaTray framework.
At the beginning of every reconstruction hits from direct photons, excluding photons scattered in water, are selected followed by a module to distinguish between signal hits from the shower and background hits.
The essential parameters of neutrino events are then derived by employing a maximum likelihood estimation for the individual particle variables like energy or neutrino direction. At first the vertex and interaction time is computed through use of a Monte-Carlo based probability density function (PDF) table. Following, the shower energy and neutrino direction is reconstructed by using the reconstructed vertex as an input in another module.
Additionally another module "collects and evaluates potential cut parameters that can later be used to remove badly reconstructed events and to suppress muon track events" [3]. At last an event classification using a Random Decision Forest (RDF) was intended to be carried out to separate hadronic showers from muon tracks. However, this approach has been given up, the exact reasons can be found in [3]. In the end a cut on a vertex fit quality parameter was made instead. The entire reconstruction process can be seen in figure 4.

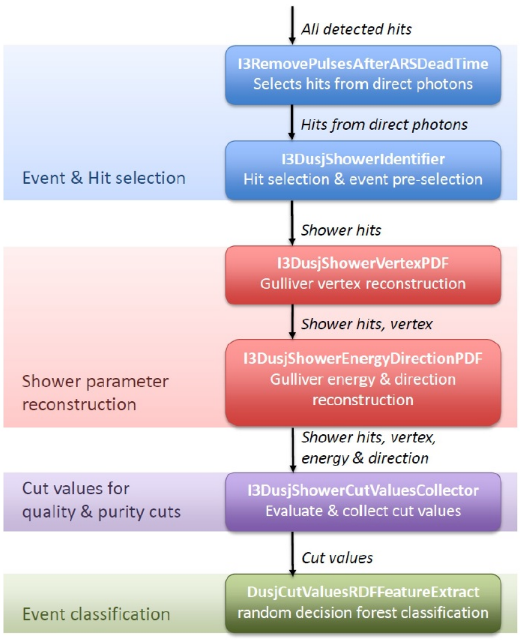


Figure 4: Work flow scheme of a Dusj-shower-reconstruction [3]

The result of this reconstruction characterizes a neutrino event in vertex position (coordinates of the initial neutrino interaction), direction (zenith, azimuth), energy and time

accompanied by several other quality parameters.
At the end of Florian Folger's investigation 59 reconstructed neutrino shower events were found.

All of the information of this chapter was written using [3], for more information on details in the reconstruction algorithm please consider [3].

4. Simulating the shower events individually

Now the final parameters of the Dusj-Reconstruction can be used as an input for the ANTARES detector simulation.
The simulation itself consists of two major modules: KM3 and TriggerEfficiency. The first component in the simulation chain is named KM3 and is responsible for simulating the light creation and propagation. TriggerEfficiency, as a second step, simulates the photon detection and the readout electronics inside the individual OMs. In addition, it also generates background light like e.g. from ^{40}K and bioluminescence.
For this work, we can now use the output of TriggerEfficiency as the input of the Dusj-Reconstruction to reconstruct the simulated neutrino event. By comparing the original reconstructed neutrino events from [3] to the reconstructed events with the simulated neutrinos as an input we can now analyze the properties of both the Dusj-Reconstruction and the detector simulation.
However, to use the reconstructed neutrinos from [3] a few modifications need to be done to the simulation chain.
First of all, the reconstructed vertex from a neutrino event does not represent the original vertex of the neutrino interaction because the reconstruction algorithm only finds the brightest measured point in space as the reconstructed vertex. In [6] it is shown that the longitudinal distance from the interaction point to the brightest point in space is energy dependent as can be seen in figure 5.

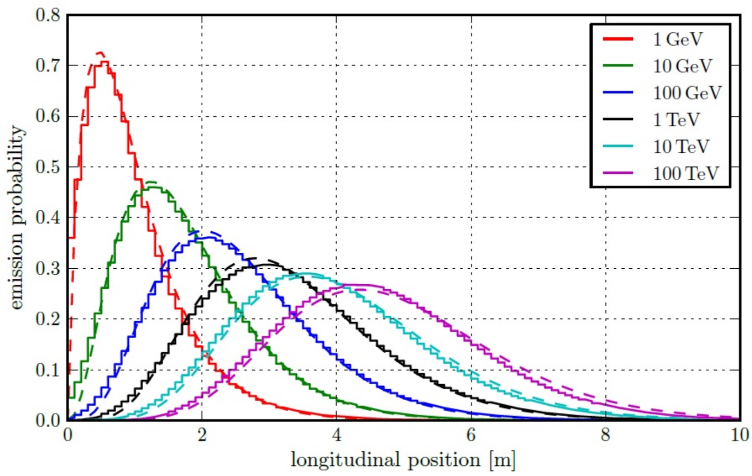


Figure 5: Longitudinal light distribution for cascades with different energies [6]

Hence, before we can set the 59 original reconstructed neutrino events as an input for the "Run by Run" simulation we have to manually shift the reconstructed vertex [6]:

$$\vec{r}_{\nu\text{-Interaction}} = \vec{r}_{\text{brightest}} - \mathrm{d}\,\vec{r} \cdot \text{VtxShift}(E) \tag{1}$$

$\vec{r}_{\nu\text{-Interaction}}$ is the point in space where the neutrino interaction happens, $\vec{r}_{\text{brightest}}$ is the brightest shower point in space, $\mathrm{d}\,\vec{r}$ is the neutrino direction and VtxShift(E) is the longitudinal shift to the brightest shower point in space:

$$\mathrm{VtxShift}(E) = 1.85 + 0.62 \cdot \ln \frac{E}{1\mathrm{GeV}} \tag{2}$$

Furthermore, the geometry of the ANTARES detector for the point in time when the neutrino event happened has to be extracted so that the simulation chain uses the correct detector geometry when simulating the event.
What has not been modified yet, but should be in the future, is that the background rates used as an input for the simulation are averaged over a segment of ANTARES data called "RunXXXXX" (where XXXXX represents numbers) that typically contains three hours of data and not only the single neutrino event. So instead of simulating the neutrino event with the baseline rate at the time of interaction the neutrino was simulated with a baseline rate averaged over the whole run. Even more important, bioluminescence bursts could be taken into account exactly as they occur in the event.
By means of the customized simulation chain each one of the 59 neutrino events was now simulated 1000 times and afterwards those 1000 simulations for every event were stored for further analysis. 1000 was chosen as a compromise between high enough statistics and limited computing time.
As this work investigates the 59 events found in [3], a table, which can be found in the appendix, shows the most important characteristics for each event.

5. Dusj reconstruction analysis of the simulated neutrinos

For the sake of simplicity from now on the original neutrinos with their reconstructed properties from [3] will be called "data neutrino" and the simulated and then reconstructed neutrinos will be called "reconstructed sim neutrino".
Most of the time we just compare the input of the simulation ("Monte Carlo (MC) neutrino") to the reconstructed sim neutrino properties. In this case the MC neutrino, as it is just the reconstructed data neutrino set as an input for the simulation, is always equal to the data neutrino.
As mentioned before, the ANTARES data is split into "Runs" containing data from a certain time period (normally about 3 hours). Since there isn't more than one neutrino event in one single run we can just refer to the RunID if we want to talk about a certain neutrino event.

In this chapter we create a series of first plots to get an idea how the basic properties of a run look like.

5.1. Vertex resolution

The first plot that can be created is a visualization of the projected longitudinal and perpendicular components of a vector from the MC vertex to the reconstructed sim vertex. Now it is possible to investigate how far and how often (1000 simulations) the reconstructed sim neutrino vertex is off the data neutrino vertex. An example can be seen in figure 6.

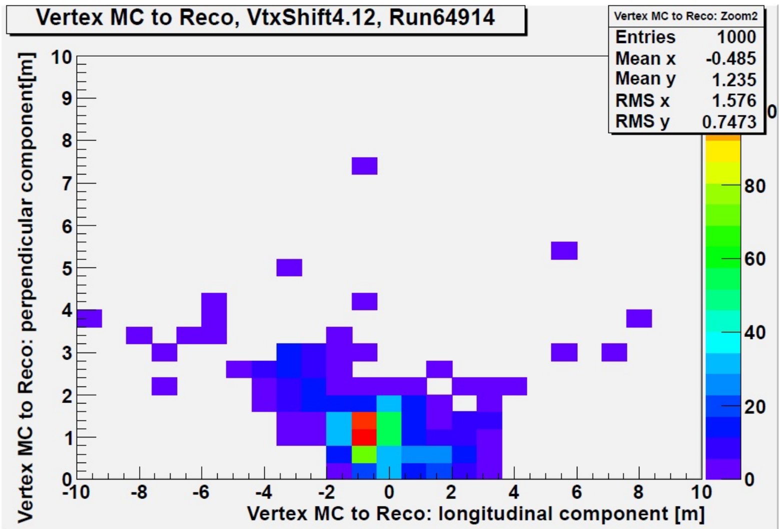


Figure 6: Projected longitudinal and perpendicular components of a vector from the data vertex to the reconstructed sim vertex for Run 64914

In this plot we can see that the most common spot of the simulation ("hotspot") is around

0.5m off in the longitudinal and 1.3m in the perpendicular component. One important aspect is that the characteristic spread or variance and the hotspot in the reconstructed sim vertex differs for each Run (= each neutrino event). For that reason this can be used as a certain quality parameter for the single neutrino events. An interesting phenomenon that occurs in four (run 26397/46852/60704/65434) out of 59 neutrino events is the emergence of a mirror solution as seen in figure 7.

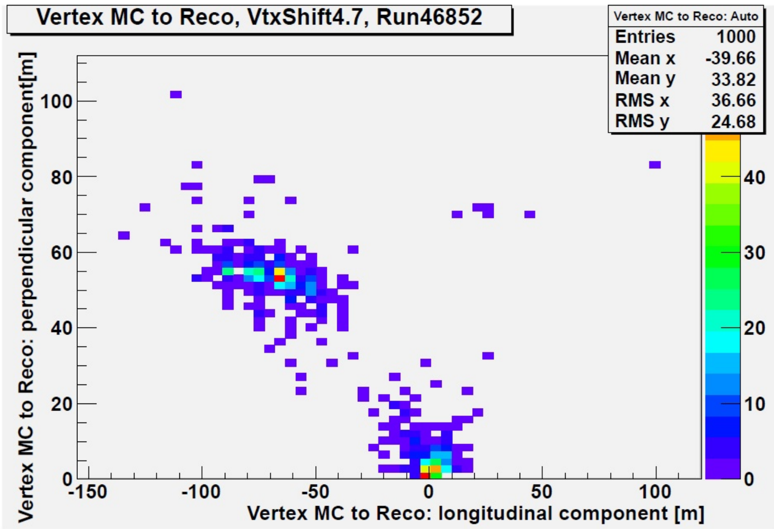


Figure 7: Projected longitudinal and perpendicular components of a vector from the data vertex to the reconstructed sim vertex for Run 46852. Characteristic property: mirror solution

For this event, the mirror solution is about 70 m off of the standard solution in the longitudinal component and about 50 m off in the perpendicular component.

Interestingly, run 26397 and run 46852 are both rather high energetic neutrino events (40TeV) that are both marked with a bad quality parameter of "1" in [3] (cf. appendix). To sum up, we can deduce that the Dusj reconstruction sometimes bounces between two equal solutions and the question arises if the reconstructed data event was the correct solution or if it was just a falsely reconstructed mirror solution. This is under the assumption that the simulation chain for the ANTARES detector is a decent representation of the real detector. To solve this issue we will try to set the mirror solution found with this plot as an input for the simulation chain later on.

For a more detailed investigation of the perpendicular or longitudinal components one can also plot them in 1D like it can be seen in figure 8.

It is also useful to calculate the distance between the MC vertex and the reconstructed sim vertex as it is shown in figure 9

In figure 7 one can clearly see the strong (in terms of frequency) mirror solution in the

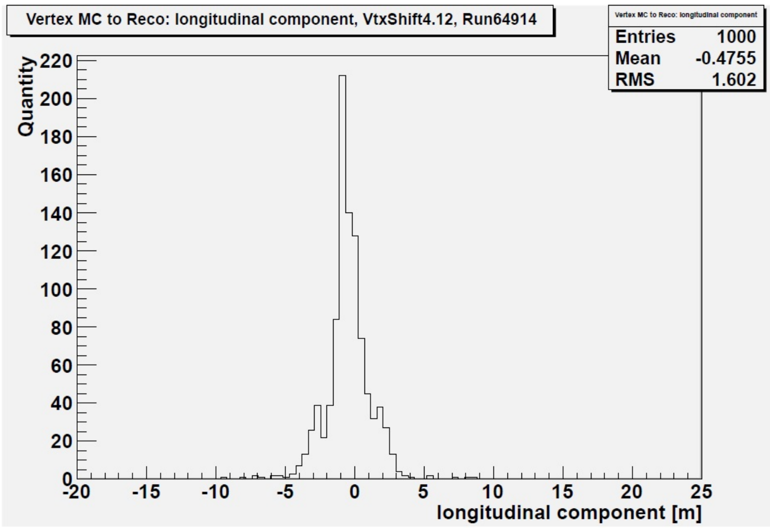


Figure 8: Projected longitudinal component of a vector from the data vertex to the reconstructed sim vertex for Run 64914

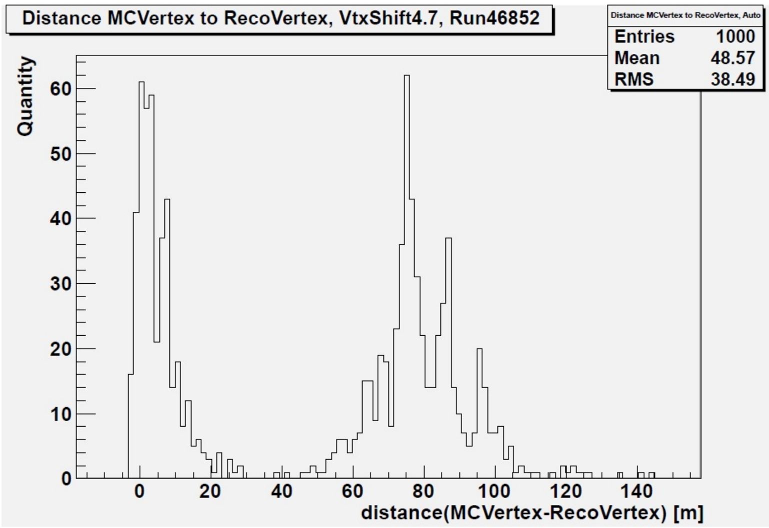


Figure 9: Distance from the MC vertex to the reconstructed sim vertex for Run 46852

vertex position for Run 46852 which is about 80m off the MC (= data) vertex.

5.2. Angular resolution

Another very important aspect for performing astronomy is the spread induced by the reconstruction algorithm in terms of angular resolution as the aim of a neutrino observatory eventually should be the creation of a celestial map. For this reason, a so called 2D "pointspread" plot was created that shows the difference in the direction (zenith, azimuth) of the data and the reconstructed sim neutrino. For the zenith axis, we can just take the difference in data and simulated zenith whereas we need to multiply the difference in azimuth with the sine of the data neutrino direction zenith. One exemplary plot can be found in figure 10.

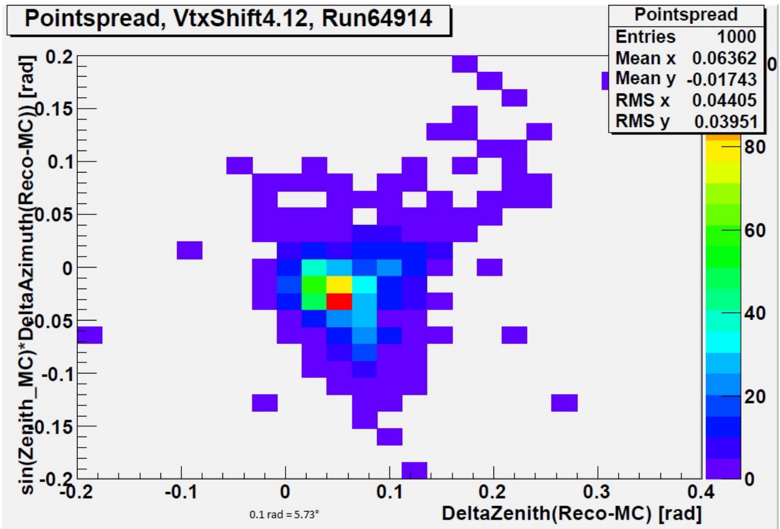


Figure 10: Pointspread for Run 64914: delta zenith to delta azimuth. 0.1rad = 5.73°

With the aid of this plot, one can quantify how good the angular resolution of the reconstruction algorithm roughly is. To get a quantitative overview of how much the hotspot offset spreads from zero and how much the spread roughly differs for each run, a table was created that can be found in the appendix as table 1.
The hotspots were chosen as the spots with the highest quantity while the hotspot spread in azimuth and zenith was measured from 100 % of the maximum (hotspot) down to 10%. Ideally, the pointspread should form a circle like one would expect it from a conventional photon telescope, however this does not happen as a lot of times the spread looks like an ellipse. This is to be expected, as the resolution of the ANTARES detector is not the same for zenith and azimuth. On that account to calculate the spread of the hotspot for each axis the cumulative spread (so both in positive and negative direction) was measured and then divided by 2. This is of course just a rough solution but in this case it will do the job.
In some of the 59 neutrino events a mirror solution is noticeable, this will be discussed below. In this case the hotspot and the hotspot spread was measured off the "regular" solution and not the mirror solution.
The smallest axis - ranges that were used to measure these values were -0.2 to 0.4 rad in zenith and -0.2 to 0.2 rad in azimuth with each of the two having 25 histogram bins. As a consequence, the minimum error in zenith is 0.024 and in azimuth it is 0.016. The error that is given in the table is always the minimum error as a result of bin sizes. Of course this method is not as precise as one would wish it to be, but the the sheer amount of

different looks for every event didn't allow for a more precise method without spending an excessive amount of time on the table.
Sometimes it wasn't possible to properly quantify hotspot offset and hotspot spread because the distribution didn't allow it. In this case only a "-" is given.
Additionally, in the case of multiple hotspots it was tried to estimate a main hotspot so that one could do the calculations from there on.
However, these values should still be interpreted as rough estimates as this table was only made to distinguish between "good" and "bad" and not "good" and "somehow good" or "medium" angular resolutions. If a more detailed investigation is needed one can just look it up in the Pointspread plots for each Run.

As mentioned before, similar to the vertex mirror solutions there are also mirror solutions in direction. One of these mirror solutions can be seen in figure 11.

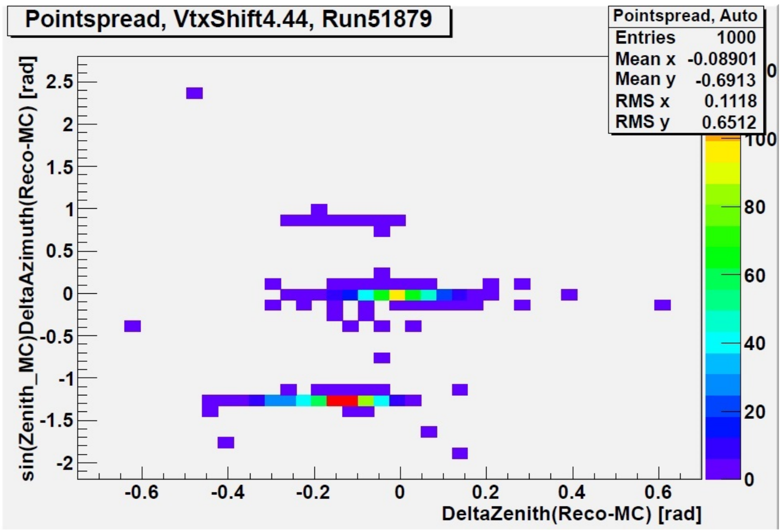


Figure 11: Pointspread for Run 51879: delta zenith to delta azimuth. 0.1rad = 5.73°

In this particular case, the mirror solution (the one in the bottom left) even got more entries than the "normal" solution and the main offset from the "normal" solution seems to be in the azimuthal direction by about 1.3rad = 75°. Akin to the mirror solutions in the vertex position one can now set this mirror solution as an input for the simulation to possibly figure out if it originally was the correct solution (so that the reconstructed data neutrino direction was just a minor false mirror solution).

As a result of the discovery of a mirror solution in both the vertex and the direction the analysis suggests that the Dusj reconstruction sometimes gets stuck in a minor maximum.

If just a local celestial map is desired to see what the direction of the neutrino event was one can look at figure 12 which shows the reconstructed zenith to the reconstructed azimuth.

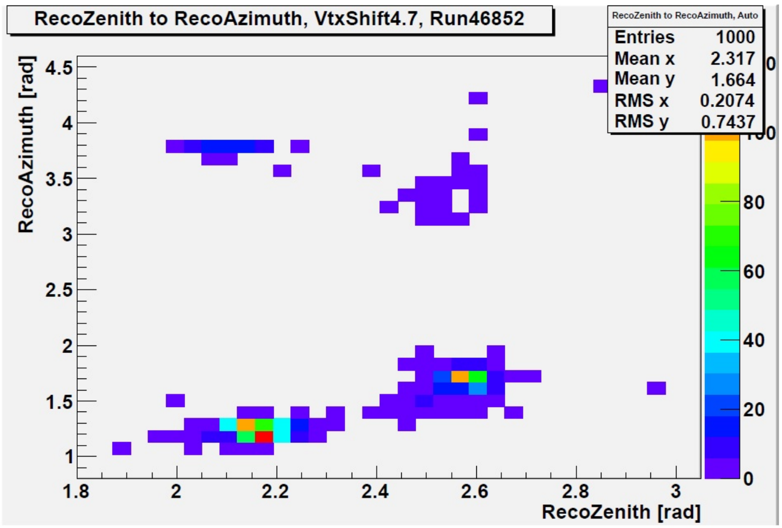


Figure 12: Reconstructed zenith to reconstructed azimuth for Run 46852: The mirror solution is easily recognized

In the case of Run 46852 the mirror solution is about 0.4rad = 23° in zenith and about 0.5rad = 29° in azimuth off of the data solution.

To determine the angle between the data neutrino direction and the reconstructed sim neutrino a 1D plot was made as shown in figure 13.

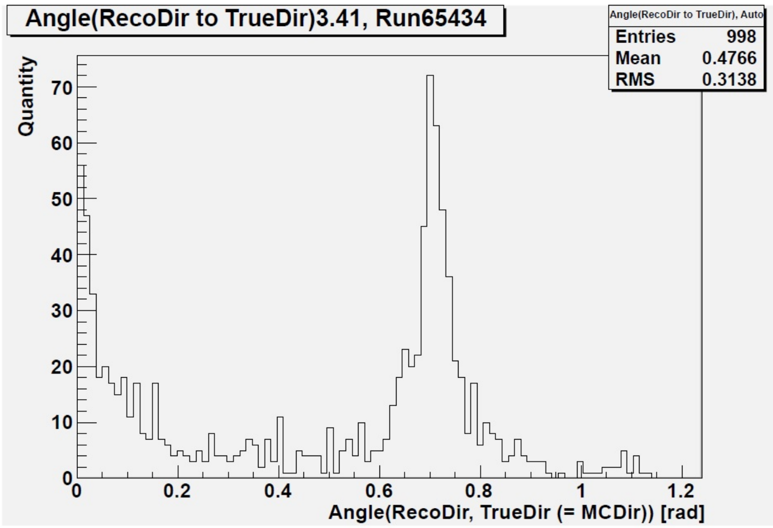


Figure 13: Angle between the data neutrino direction and the reconstructed sim neutrino direction for Run 65434: The mirror solution is easily recognized

In this example one can easily see the "normal" and the mirror solution.

For some runs there are both vertex and direction mirror solutions. To ultimately determine to which vertex a mirror solution in the pointspread and vice versa is associated a plot was made that shows the distance between the MC vertex and the reconstructed sim vertex versus the angle between the data direction and the reconstructed sim direction. One example for this is Run 60704 as shown in figure 14.

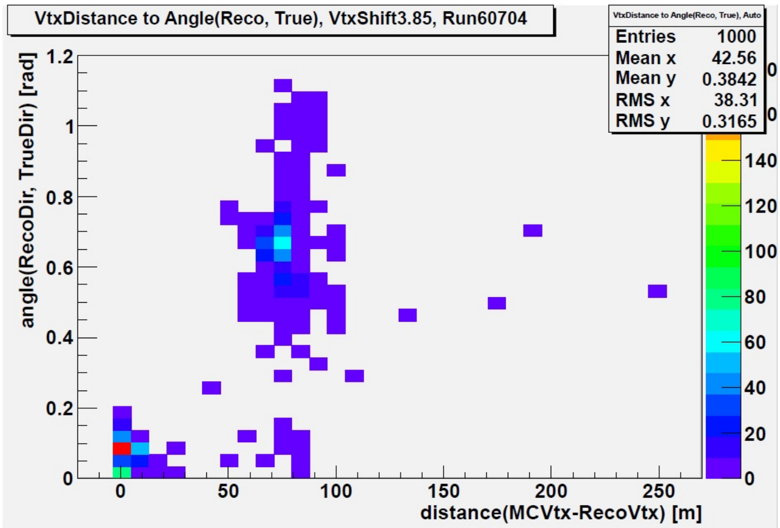


Figure 14: Distance between MC vertex and reconstructed sim vertex versus the angle between the data neutrino direction and the reconstructed sim neutrino direction for Run 60704

If one now compares this plot with the longitudinal & perpendicular plot as well as with the plot for the angle between the data neutrino direction and the reconstructed sim neutrino one can figure out that for this run the mirror solution in the vertex also belongs to the mirror solution in the direction.

Another question is if the mirror solution for the vertex and the direction is somehow correlated to the error in the energy reconstruction. To investigate this issue, it is possible to plot the reconstructed sim energy minus the MC energy versus the angle between the data direction and the reconstructed sim direction. A typical example can be seen in figure 15 for Run 47679 which has a weak mirror solution in the direction.
In this case the mirror solution has worse reconstructed energy, however this isn't the typical case for all runs. Sometimes the mirror solution also has a better reconstructed energy and there are also Runs where both the "normal" and the mirror solution have the same reconstructed energy. As a consequence, it seems like there is another effect that is causing these variabilities. Ultimately, there is no general relation between the goodness of the energy reconstruction and the appearance of a mirror solution direction

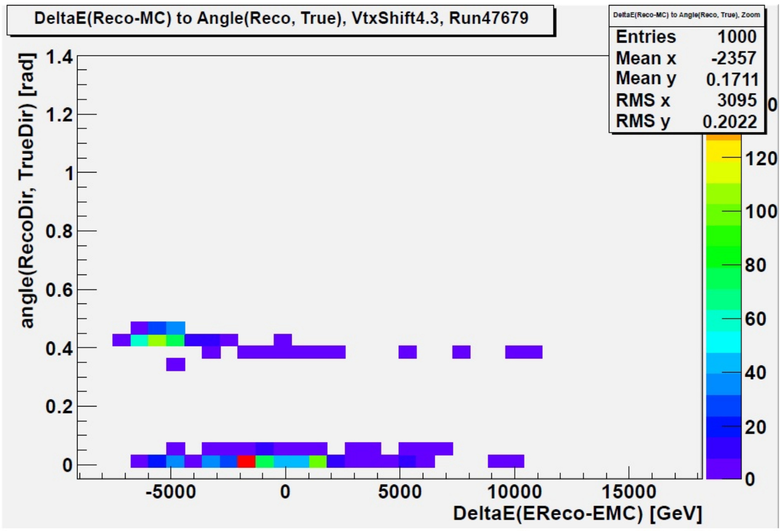


Figure 15: Energy error versus the angle between the data direction and the reconstructed sim direction for Run 47679

wise.

5.3. Vertex reconstruction quality

To further investigate the Dusj reconstruction itself, especially to see if a Run has a large spread in the quality of the vertex fit reconstruction, the "DusjShowerRecoVertexFitReducedLogLikelihood", which is a quality parameter of the Dusj reconstruction, was investigated for every vertex fit.
As such it can be interpreted as a quality parameter for the vertex reconstruction of the Dusj algorithm. In [3] a cut was set on this parameter at a value of 7.9 so that investigated events with a value greater than 7.9 were discarded. Therefore, we shouldn't see values much above 7.9 though some outliers are expected as a result of the spread in vertex position. One of these distributions can be seen in figure 16.
For this run, the distribution looks like expected, the vertex fit quality is Gaussian-like distributed around a value of about 7.9 that is equal to the cut value that was used in [3] for atmospheric muon suppression. Similarly, in most cases the vertex fit quality has a high value (significantly above 7.9) when the vertex reconstruction didn't properly succeed like for example in Run 65434. In this case the vertex position was misreconstructed for about 50m in both the longitudinal and perpendicular component because the mirror solution has almost all entries (most reconstructed sim neutrinos have the mirror solution vertex). Thus, the mean of the "DusjShowerRecoVertexFitReducedLogLikelihood" distribution is about 10 which suggests that the reconstructed vertex in [3] probably was a falsely reconstructed vertex as the cut value of 7.9 is only achieved at the very left end of the distribution seen in figure 17.
However, in some cases there are runs with a mirror solution in the vertex that has as

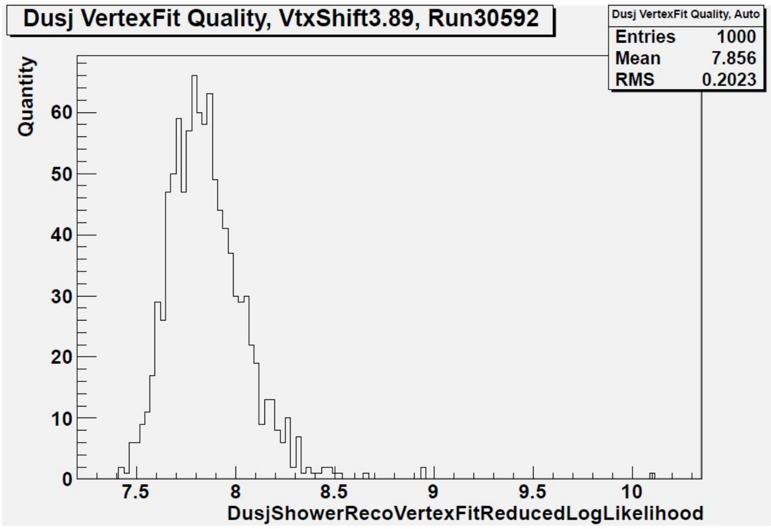


Figure 16: Vertex reconstruction quality for Run 30592

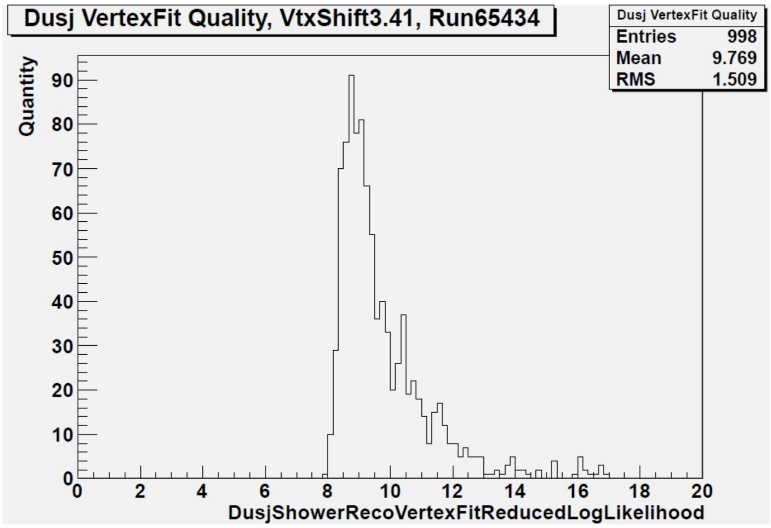


Figure 17: Vertex reconstruction quality for Run 65434: the distribution is far above the original cut value of 7.9

many entries out of 1000 simulations as the "normal" solution so that both are equiprobable. One of these events is the data neutrino event in Run 46852.
Now it stands to reason if the vertex fit quality of both of them is equal or if one could be discarded by looking at the "DusjShowerRecoVertexFitReducedLogLikelihood" value. For this purpose, a 2D plot was made that shows the distance of the MC vertex to the reconstructed sim vertex in comparison to the vertex fit quality.
The result can be seen in figure 18.

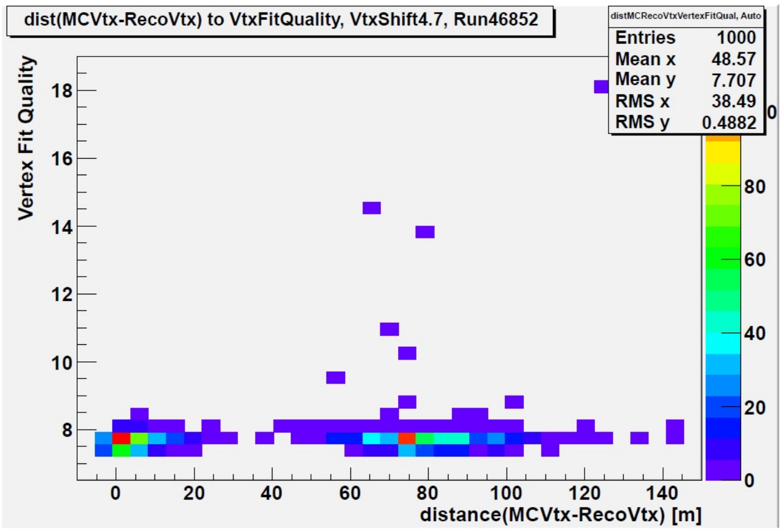


Figure 18: Vertex reconstruction quality for Run 46852: the vertex fit quality of the normal and the mirror solution are equal

Figure 18 shows that in this case the Dusj vertex reconstruction fit quality is equal for both solutions so that it doesn't allow for a differentiation between a "correct" and a "false" vertex.

5.4. Combined energy and angular fit reconstruction quality

Another Dusj reconstruction fit quality parameter that can be investigated is the "DusjShowerRecoFinalFitReducedLogLikelihood" that combines both the energy and the direction fit quality. This value is quite similar to the Dusj vertex reconstruction quality parameter however this time the mirror solution in the pointspread cannot be seen in distribution as easily because the quality parameter is a combined value of energy and direction.
Therefore, the two pointspread "peaks" are too close together to properly see them and this also suggests that the energy reconstruction is independent of the angular reconstruction which is as expected.
One example can be seen in figure 19 for Run 43639 which has a strong mirror solution in the pointspread.

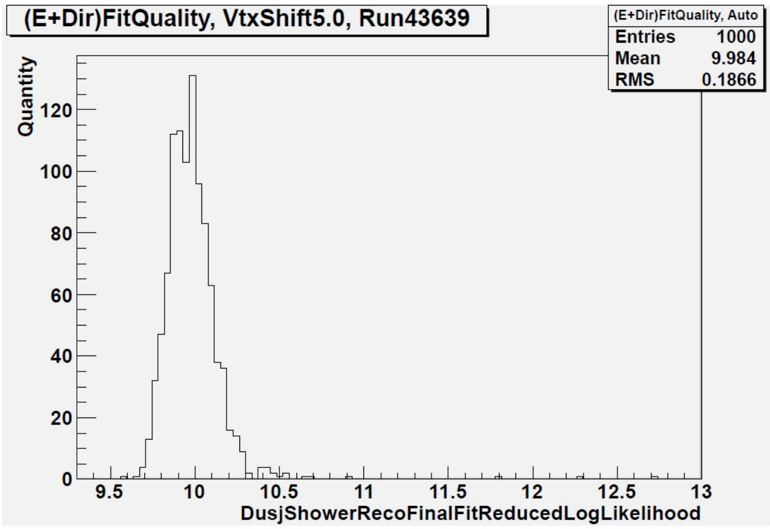


Figure 19: Combined energy and direction reconstruction quality for Run 43639 which has a mirror solution in the pointspread

6. Visualization of the background rates

Sometimes, when a potentially "bad" or falsely reconstructed data event is recognized the question arises if the background rate is responsible for this circumstance. In this particular case, it is possible to investigate the "onshore" or the "offshore" background rates in the ANTARES data.
Hereby, the background rates are measured with the ARS whereas the "offshore" background rates are measured directly at the photomultiplier voltage output. Even though the "offshore" background rates have a worse resolution than the standard background rates they are by all means needed because the standard background measurement with the ARS can only measure frequencies up to a certain point. So, if the maximal frequency that the ARS can measure is exceeded (typically in the MHz range) one can still take a look at the "offshore" background rates.
With this knowledge it is possible to plot both the standard background rates and the "offshore" background rate for each OM of the ANTARES detector.
These plots can be created for the data neutrino events and for the reconstructed sim neutrino events. For the sim neutrino events the arithmetic mean out of 1000 simulations was taken for the calculation of the background rates. On top of this, only the OMs marked as properly calibrated in the ANTARES calibration database were used for the calculation.
As mentioned in chapter 4, the background rates that serve as an input for the simulation are not the same rates as the measured background rates for every data neutrino event. By comparing the measured background rates from the neutrino event to the simulated background rates one can roughly compare the implication of this deficiency in the current simulation chain.
On top of this, as mentioned before, it is now possible to determine whether a high background rate was one of the reasons for a bad Dusj reconstruction.
An example for a simulated background rate plot can be seen in figure 20, the color indicates the background rate in kHz.
OMs and Lines that are not shown were not working at the time of the neutrino event.

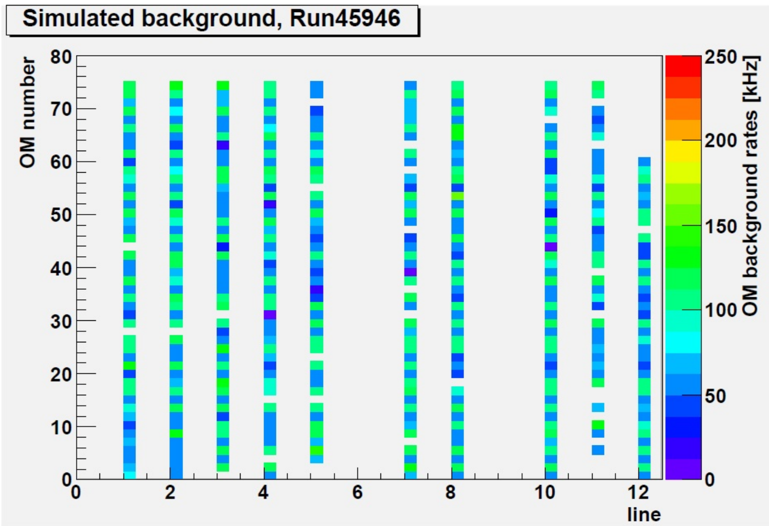


Figure 20: Simulated background rates for all 12 ANTARES lines for Run 45946. Colors: background rate in [kHz]

7. Investigation of the Dusj reconstruction in terms of energy

One aspect that hasn't been examined yet is the energy reconstruction of the Dusj algorithm. For this purpose, a 1D histogram was created that shows the difference between the data neutrino energy (= MC energy) and the reconstructed sim neutrino energy as it is shown in figure 21 for Run 26397.

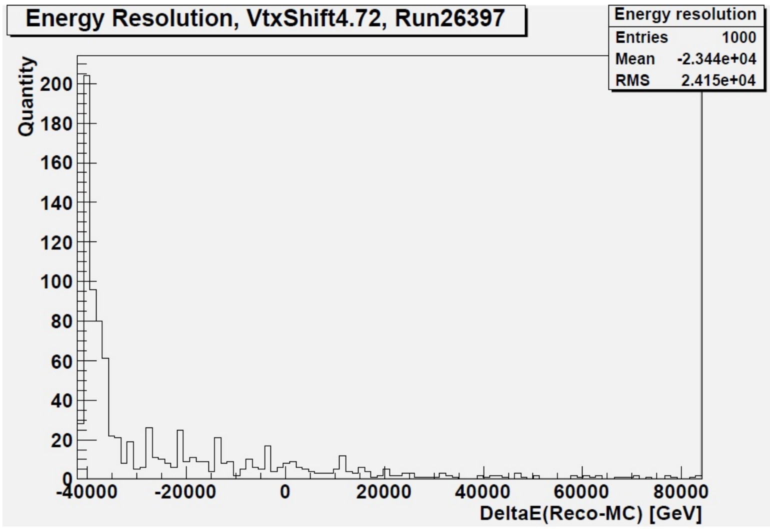


Figure 21: Reconstructed sim energy minus MC energy to quantity for Run 26397

What has to be kept in mind is that figure 21 only shows data points in between a range of about −40.000GeV and 80.000GeV. This matters because it was discovered that there are around a dozen reconstructed sim energies with a reconstructed energy of up to 25PeV. This occurrence has to be considered later on.

As one can see in figure 21 the reconstructed energy is a lot smaller than the MC energy. This tells us that the originally reconstructed energy for the data neutrino of Run 26397 in [3] probably was too small by a large margin and this applies to almost all other data neutrino events, too. For Run 26397 the reconstructed data energy is about 42 TeV whereas the reconstructed sim neutrino, if we take the maximum of the distribution which suffices for this demonstration, only yields about 2 TeV. If we would scale the data neutrino energy with this value we would get about 800 TeV instead of 42 TeV. However, 800 TeV is very unlikely if one considers the flux spectrum and the size of the ANTARES detector.

From this point of view one could now consider to rescale the originally reconstructed data neutrino energy if the distribution in these plots is sufficiently sharp.

To give an overview for all data neutrino events a graph was created that shows the

Run-Number on the x-axis and the percentage deviation off the data neutrino (= MC) events for the reconstructed sim (= Reco) neutrino events. This value is calculated in the following way:

$$\frac{E(\mathrm{MC}) - E(\mathrm{Reco})}{E(\mathrm{MC})} \tag{3}$$

For E(Reco) the arithmetic mean out of 1000 simulations is taken. However, as mentioned before, if one would include all of the massive outliers the validity of the whole value would be distorted. For this reason, another plot was made that shows the reconstructed sim energies for all runs in one plot as it can be seen in figure 22.

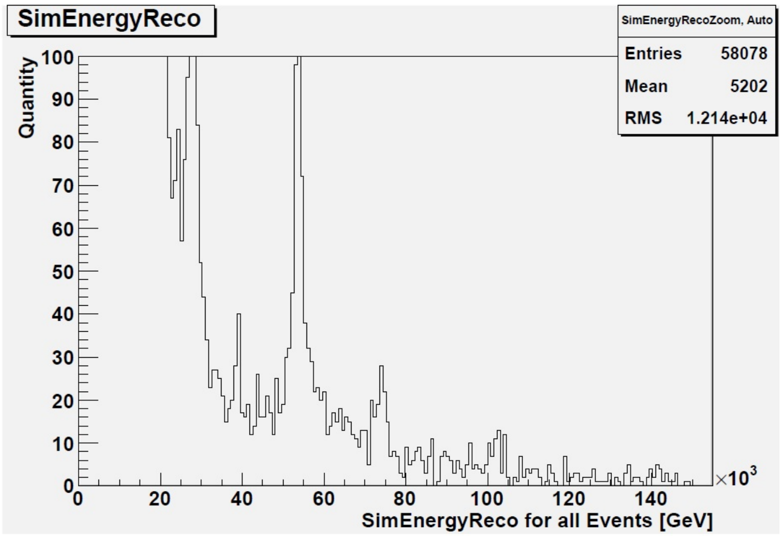


Figure 22: Reconstructed sim energies for all Runs

Now we can figure out that reconstructed energies above 150TeV are very rare. On that account a cut was taken at 150TeV so that no simulations with a reconstructed energy of above 150TeV would be included in the arithmetic mean for E(Reco).

Now equation 3 yields a maximum value of 1 if the reconstructed sim energy is 0, an ideal value of 0 if $E(\mathrm{MC}) = E(\mathrm{Reco})$ and a value < 0 if the reconstructed sim energy is larger than the MC (= data) energy. Additionally, one can define the RMS of the individual distributions as error bars in the graph which gives a rough idea of how broad the distributions are. The resulting graph can be seen in figure 23.
Now we can also calculate how large the deviation is on average by fitting the data points with a horizontal line. However as the Runs 47838, 52087, 60226 and 65434 are either quite off all of the other data points or the spread in their distributions is very large they were excluded from the fit shown in figure 24.
As a result we can conclude, under the assumption that the simulation itself does not affect the energy reconstruction by a large margin, that the Dusj reconstruction under-

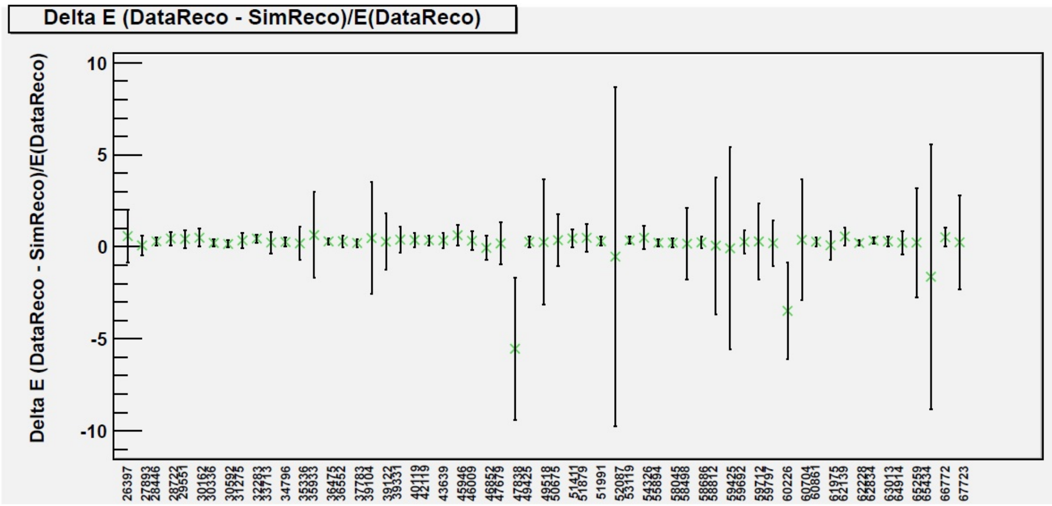


Figure 23: Percentage deviation off the data neutrino (= MC) events for the reconstructed sim (= Reco) neutrino events

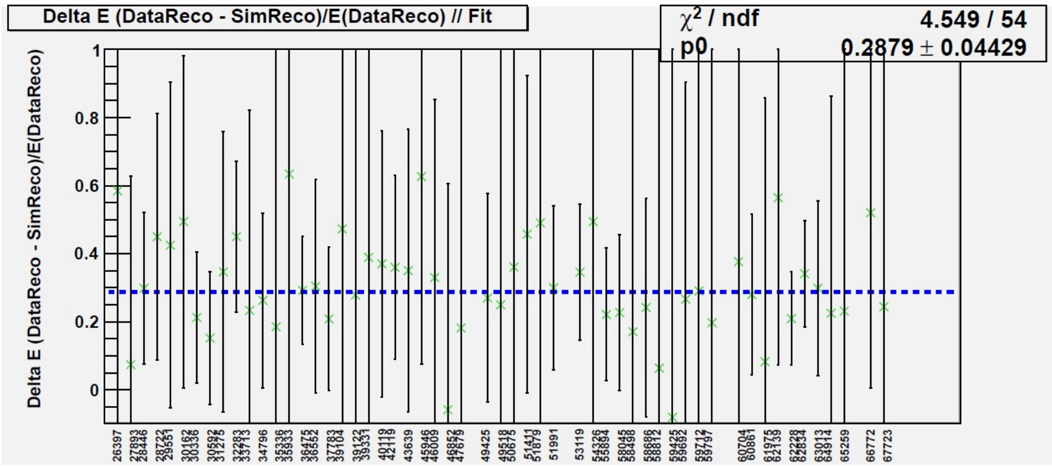
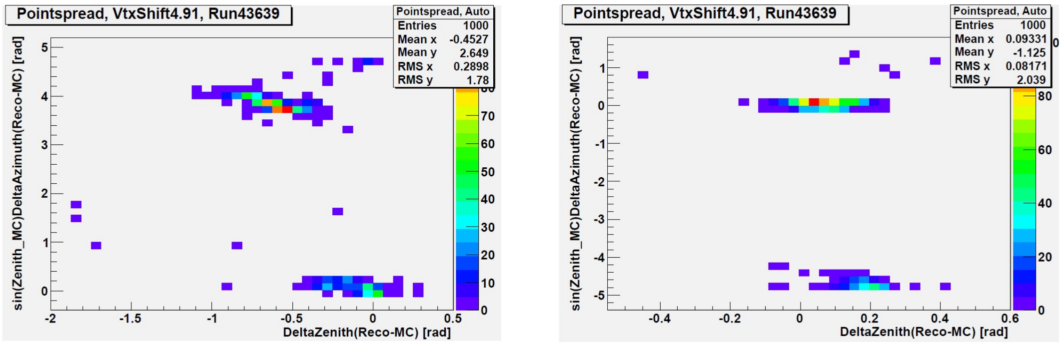


Figure 24: Percentage deviation off the data neutrino (= MC) events for the reconstructed sim (= Reco) neutrino events

estimates the true event energy by about $p0 = 28,8\% \pm 4.4\%$ on average. However, the small reduced chi square ($\chi^2_{\text{red}} \approx 0.1$) indicates that the hypothesis of a constant energy shift is probably wrong. In the future it should be investigated, if there is a hypothesis that better matches the underestimation in energy.

8. Setting the mirror solutions as an input for the simulation

As already mentioned in chapter 5.1 it is still an unsolved issue whether some of the observed mirror solutions in either the vertex or direction could be the "true" solution. However, if we set the mirror solutions as an input for the simulation chain and compare those new reconstructed vertices / directions to the ones with the data neutrino input from [3] one could maybe figure this out.
For this purpose, five runs with either a mirror solution in the vertex or direction (Run 43639 (direction), Run 45946 (direction), Run 46852 (vertex), Run 47679 (direction), Run 51879 (direction)) were set as a simulation input. After this procedure the standard plots that are described in the previous chapters were created for the simulated mirror solutions. At this point, one can now compare the reconstruction of the data neutrino input and of the mirror solution neutrino input.

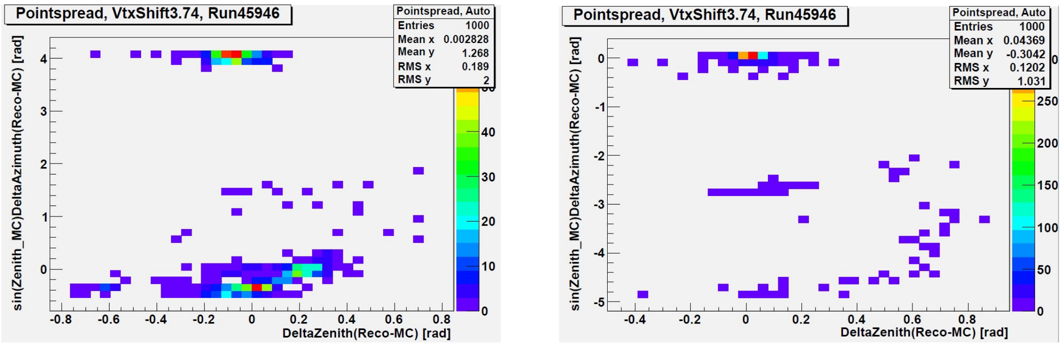


(a) Standard data neutrino sim input

(b) Mirror solution (direction) neutrino sim input

Figure 25: Comparison of the Pointspread for Run 43639 with either the original data neutrino or with the mirror solution (direction) neutrino: Both solutions just change places. $0.1\text{rad} = 5.73°$

Figure 25 shows a comparison of the pointspread plot for both inputs. It can be recognized that both solutions just change places if you set the mirror solution in the direction as an input for the simulation chain.
However, run 45946 shows a different behaviour as it can be seen in figure 26.
In this case the initial mirror solution doesn't differ by much after the reconstruction and on top of this the standard data solution disappears. This means that if the mirror solution was the "true" solution, a significant second solution wouldn't have been produced in the Dusj reconstructions. However this effect could also be happening due to a systematic error in the simulation framework. As a result the disappearance of the standard solution is a hint that the standard solution might be the "true" solution, however this is not completely sure. That's why another method is needed to better distinguish between a standard solution and a mirror solution. Therefore, later on a hit pattern approach is executed. By combining both methods one can try to derive the "true" solution.

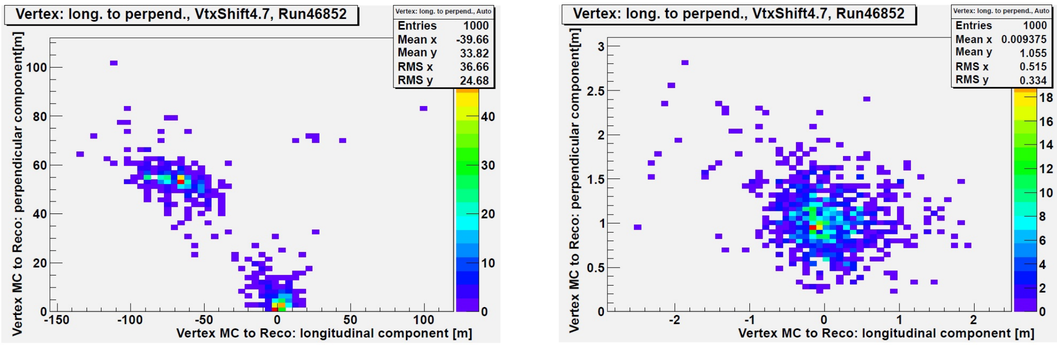


(a) Standard data neutrino sim input

(b) Mirror solution (direction) neutrino sim input

Figure 26: Comparison of the Pointspread for Run 45946 with either the original data neutrino or with the mirror solution (direction) neutrino: the initial mirror solution is spread around zero and the standard data solution disappears. $0.1\text{rad} = 5.73°$

The same effect happens for Run 46852 even though it has a mirror solution in the vertex as seen in figure 27.



(a) Standard data neutrino sim input

(b) Mirror solution (vertex) neutrino sim input

Figure 27: Comparison of the vertex components for Run 46852 with either the original data neutrino or with the mirror solution (vertex) neutrino: the initial mirror solution is spread around zero and the standard data solution disappears

Before, if one sets the mirror solution in the vertex as an input of the simulation chain the standard data neutrino vertex solution disappears and the mirror solution is the new "good" solution.

At last, for run 47679 the mirror solution (direction) just changes place with the standard solution whereas for run 51879 the standard solution disappears and the mirror solution (direction) is again the new "good" solution.
As a summary for the direction mirror solutions, the standard solution disappears 2 out

of 4 times and for the mirror solution in the vertex the standard solution disappears too. However, this still doesn't answer the question why sometimes the standard data solution disappears and why sometimes it does not. If one would want to create a celestial map out of e.g. these shower events for each possible neutrino event it would need to be investigated if the reconstructed solution is maybe only a mirror solution. Unfortunately, the question if an event is the "true" solution or only the mirror solution can't be answered in all cases just by simulating the reconstructed event.

9. Time residual analysis

For further analysis in the next chapter time residual cuts need to be made to separate the background hits from the neutrino event hits. For this purpose, we can compare the point in time of a hit or also called "calibrated pulse" of a photomultiplier to the point in time when we would expect a signal at that certain moment because we know the time of the MC neutrino event. Calibrated pulses are those signals that fulfill the conditions mentioned in chapter 2.

Therefore, the time residual needs to be calculated in the following way:

$$\text{TimeResidual} = t_{\text{hit}} - t_{\text{expected}} \tag{4}$$

$$\text{and} \quad t_{\text{expected}} = t_{\text{MCEvent}} + t_{\text{travel time}} \tag{5}$$

t_{hit} is the point in time where a hit at a photomultiplier is recored, t_{expected} is the point in time when we would expect the hit at this photomultiplier, t_{MCEvent} is the time of the initial neutrino event set as an input for the simulation chain and $t_{\text{travel time}}$ is the time that photons would need from the neutrino interaction vertex to the PMT.
If we now plot this for every photomultiplier and for 1000 simulations we get the distribution in figure 28. The result is a small stable background from -4000ns to 3000ns that we want to cut out and signal hits in the range of -20ns to 120ns as seen in figure 28.

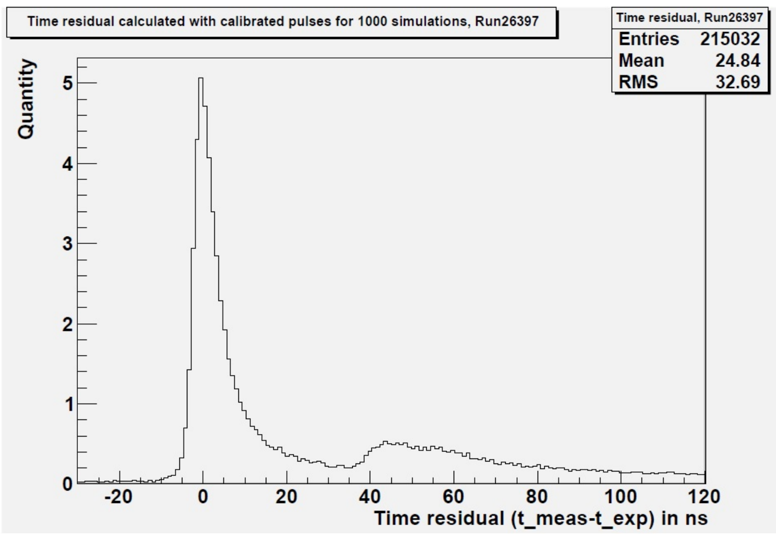


Figure 28: Time residual for all PMTs with recorded hits for 1000 simulations

As expected we can identify a significant peak located at around 0ns. A second small peak can be seen at approximately 45ns which is the result of the functional principle of the two ARS as described in chapter 2.

An important outcome is that we can now cut all hits that lie outside a certain time window to suppress background light. In this work a cut was made at -20ns and 30ns.

10. Hit pattern analysis

Now that we have developed a cut criterion we can try to identify "bad" events (Dusj reconstruction failing in certain aspects) by exerting a hit pattern analysis.
At first one can compare the total charge of a PMT for a data neutrino event to the distribution that 1000 simulations for the reconstructed sim neutrinos yield. However, as we don't want that "bad" PMTs distort our analysis we exclude them from further analysis and plotting. "Bad" PMTs are PMTs that are not marked with an "OK" status flag, e.g. PMTs that are somehow damaged, in the calibration database. On top of this, only the hits are taken that satisfy the time residual specified in chapter 9. Hits in a simulation or in data that are "OK" but either fail in the time residual or they just don't register a hit are inserted with a value of 0.
At this point, it is possible to create a charge distribution (1000 simulations) for each PMT and the measured total charge in the data event marked with a red vertical line as seen in figure 29.

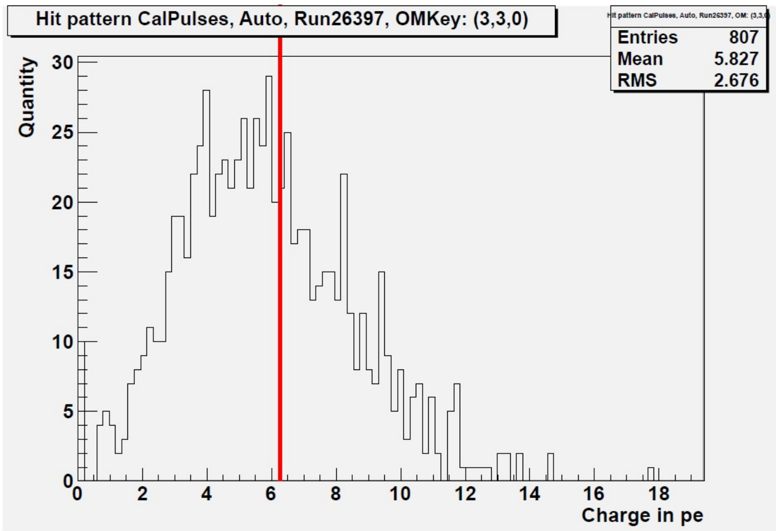


Figure 29: Distribution of the measured charge of a specific PMT for 1000 simulations. Vertical red line: data event. The "OMkey" that is specified in the title is a key in the data that links to a certain PMT

If the reconstructed sim charge of the PMT is correct, the data event (vertical red line) should reside in the simulated distribution. Obviously it is not feasible to look at every plot for every PMT because there are so many PMTs. For this reason, a so called "Q-Value" was created. For the data events it is calculated as follows:

$$Q_{\text{All}} = \sum_{\text{PMT: } i=1} \frac{\text{charge}_{\text{data},i} - \text{MeanCharge}_{\text{sim},i}}{\sqrt{\text{MeanCharge}_{\text{sim},i}}} \quad , \tag{6}$$

where $\text{charge}_{\text{data},i}$ is the measured total charge for a certain PMT in the data, $\text{MeanCharge}_{\text{sim}}$

is the mean of the distribution for the total charge of this PMT out of 1000 simulations and the sum runs over all PMTs. Hence, we can split this sum into two individual sums, Q_{Hit} and $Q_{\text{No Hit}}$:

$$Q_{\text{Hit}} = \frac{1}{N_{\text{PMTs}}} \sum_{\text{PMT: } i=1} \frac{\text{charge}_{\text{data},i} - \text{MeanCharge}_{\text{sim},i}}{\sqrt{\text{MeanCharge}_{\text{sim},i}}} \text{ and charge}_{\text{data},i} \neq 0 \quad , \tag{7}$$

$$Q_{\text{No Hit}} = \frac{1}{N_{\text{PMTs}}} \sum_{\text{PMT: } i=1} \frac{0 - \text{MeanCharge}_{\text{sim},i}}{\sqrt{\text{MeanCharge}_{\text{sim},i}}} \text{ and charge}_{\text{data},i} = 0 \tag{8}$$

Q_{Hit} only sums over those PMTs that have a hit (= charge) in the original data whereas $Q_{\text{No Hit}}$ only sums over those PMTs that do not have a hit in the data. At last to get a normed mean value for all PMTs we just need to divide these Qs by the amount of PMTs summed over (N_{PMTs}). As a result we now have the deviation of the mean of the simulated charge distribution for every PMT from the measured charges in the data neutrino events.
However, this brings up the question if these values also lie in the distribution of the deviations for the 1000 simulations. For a good event in terms of reconstruction one would expect that the mean deviation of the data event charge from the MeanCharge out of the 1000 simulations should not be larger than any charge deviation of a single simulation from the MeanCharge out of 1000 simulations. Therefore, we need to calculate the deviation of single simulations from the mean over 1000 simulations to illustrate the variance that the sim chain produces. These values are calculated as follows:

$$Q_{\text{All, sim}} = \frac{1}{N_{\text{PMTs}}} \sum_{\text{PMT: } i=1} \frac{\text{charge}_{\text{sim},i} - \text{MeanCharge}_{\text{sim},i}}{\sqrt{\text{MeanCharge}_{\text{sim},i}}} \tag{9}$$

$$Q_{\text{Hit, sim}} = \frac{1}{N_{\text{PMTs}}} \sum_{\text{PMT: } i=1} \frac{\text{charge}_{\text{sim},i} - \text{MeanCharge}_{\text{sim},i}}{\sqrt{\text{MeanCharge}_{\text{sim},i}}} \text{ and charge}_{\text{sim},i} \neq 0 \tag{10}$$

$$Q_{\text{No Hit, sim}} = \frac{1}{N_{\text{PMTs}}} \sum_{\text{PMT: } i=1} \frac{0 - \text{MeanCharge}_{\text{sim},i}}{\sqrt{\text{MeanCharge}_{\text{sim},i}}} \text{ and charge}_{\text{sim},i} = 0 \tag{11}$$

Hence, the only difference in the calculation is that one takes the charge of a PMT in one of the 1000 simulations instead of the charge measured in the original data. If we now calculate these values for all 1000 simulations one can put each of them in a 1D histogram together with the corresponding Q-values for the measured data. In terms of a better comprehensibility we also add the index "data" to the Q-values calculated for the data events: $Q_{\text{All, data}}$, $Q_{\text{Hit, data}}$ and $Q_{\text{No Hit, data}}$.
After all of this preliminary work it is now possible to plot the Q-values for both the data and the simulation into one single histogram. One example can be seen for Q_{Hit} in figure 30.

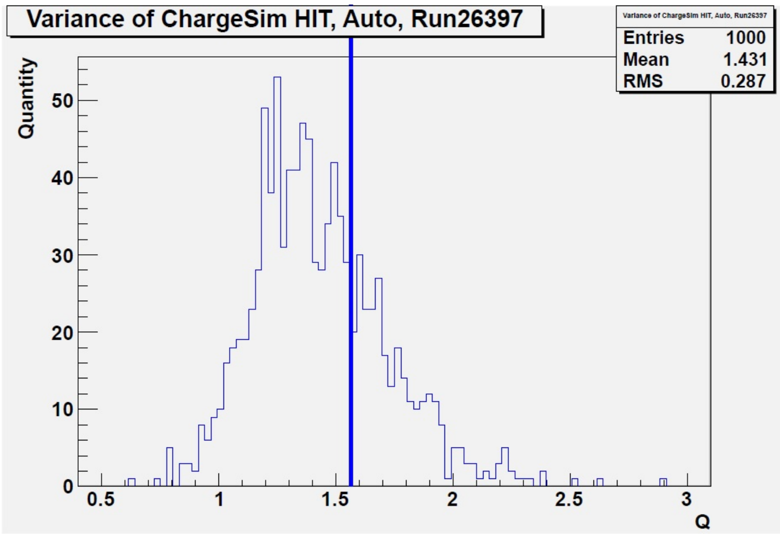


Figure 30: Distribution of the simulated Q_{Hit} values for Run 26397. Vertical blue line: Q_{Hit} value calculated for the data event

One aspect that has to be remembered though is that for the reconstruction algorithm only the PMTs with a significant charge are of interest. So to deduce whether an event has a good Dusj reconstruction or not only the PMTs with a high enough charge should be taken.
In this work, four approaches to select the most interesting PMTs are tested. One selection is to only take the PMTs with a MeanCharge (calculated through 1000 simulations) larger than 1pe (pe = photoelectronvolt). Another approach is to only select a certain number of PMTs with the largest charge. In this work values of 10, 30 and 50 were used. With that in mind we can derive 4 new Q values for either data or sim:

$$Q_{\text{All, MeanCharge>1, data/sim}} \tag{12}$$

$$Q_{\text{All, 10PMTs with highest charge, data/sim}} \tag{13}$$

$$Q_{\text{All, 30PMTs with highest charge, data/sim}} \tag{14}$$

$$Q_{\text{All, 50PMTs with highest charge, data/sim}} \tag{15}$$

The three different approaches for the "highest charge" selection were taken because the energy of a neutrino event determines how many PMTs have a significant charge.
At last we can plot all of these values into a single plot as it can be seen in figure 31. "Hit" values are painted in blue, "No Hit" in red, "MeanCharge $>$ 1" in black, "10 PMTS with largest charge" in magenta, "30 PMTS with largest charge" in green and "50 PMTS with largest charge" in orange.
As one can see in figure 31 for Run 28446 the calculated Q values for the data event are in agreement with the simulated distributions which indicates that the Dusj reconstruction

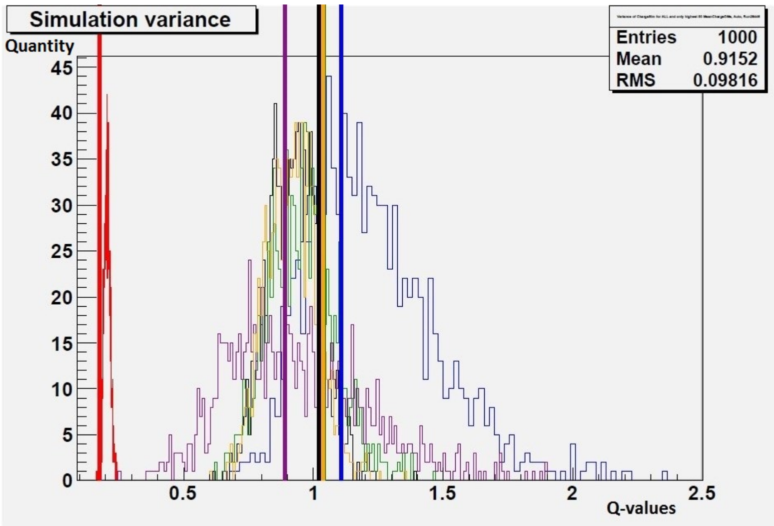


Figure 31: Distribution of the simulated Q values for Run 28446. Vertical lines: Q values calculated for the data event. The green vertical line is concealed by the orange and black lines. "Hit": blue, "No Hit": red, "MeanCharge > 1": black, "10 PMTs with largest charge": magenta, "30 PMTs with largest charge": green and "50 PMTs with largest charge": orange.

of this neutrino event should be quite good. It also can be seen that the Q value distribution for the "10 PMTs with largest charge" selection is rather broad. This, however, is to be expected as only a total of 10 PMTs is used for the calculations compared to the 30 or 50 distributions.

If we now want to compare the Q-data values with the Q-sim values for all Runs simultaneously one can plot the mean of the Q-sim values with the RMS as error bars and the Q-data values into one plot as seen in 32. Circles indicate the Q-data values and the Q-sim values are marked as squares.

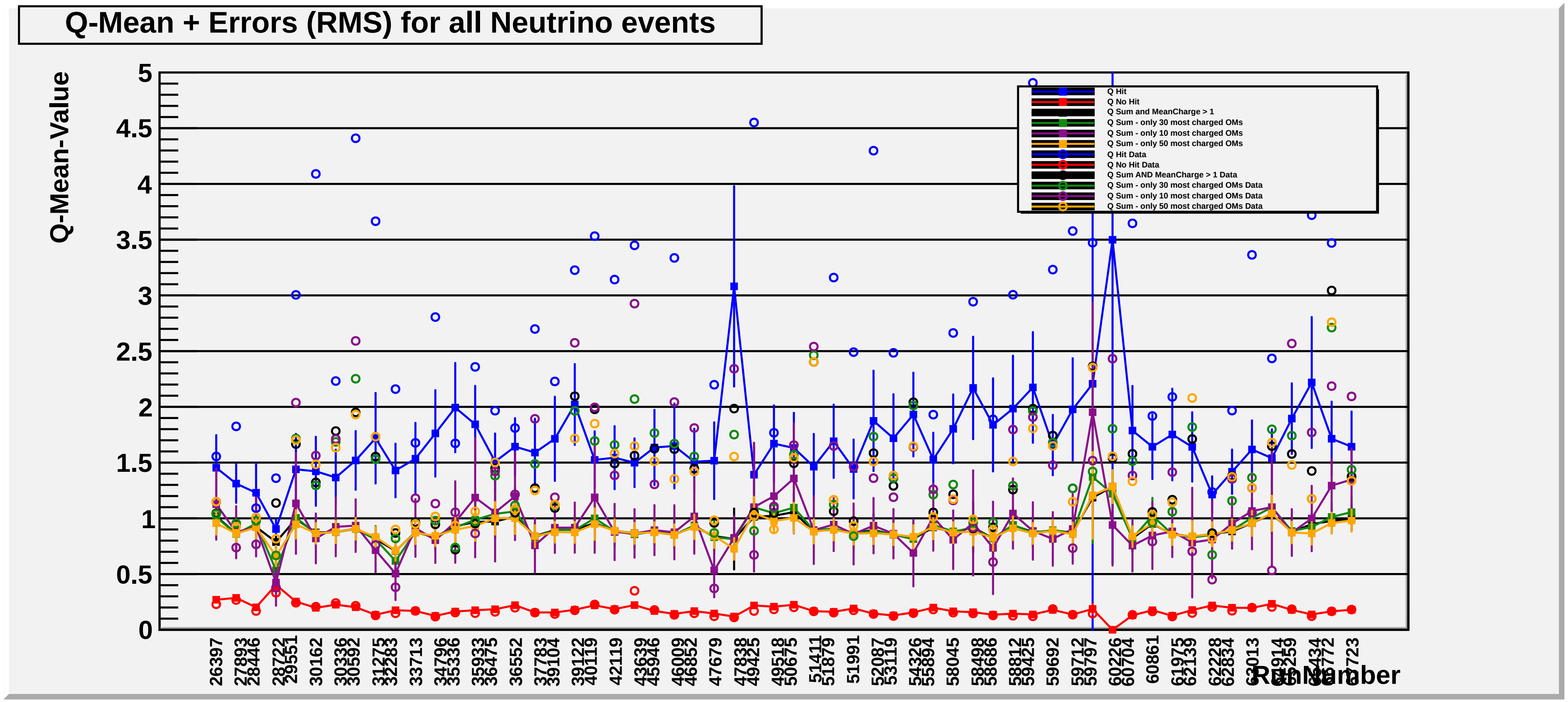


Figure 32: Q-sim mean values and RMS (squares and error bars) and Q-data values (circles) for each Run. "Hit": blue, "No Hit": red, "MeanCharge > 1": black, "10 PMTs with largest charge": magenta, "30 PMTs with largest charge": green and "50 PMTs with largest charge": orange.

Now one can see that the results for most Runs aren't as perfect as for Run 28446. Some times the Q-data values lie in between the Q-sim RMS but most of the time they are not. Actually, as a matter of fact, at times the Q-data values are far off the Q-sim values which might indicate a badly Dusj reconstructed event.

10.1. Hit pattern values and the underestimation of energy by the Dusj reconstruction

However, small deviations are expected because as we know now the Dusj reconstruction underestimates the energy of events by about 30%. Since the emitted light in an event is proportional to its energy and therefore also to the collected charge of the PMTs one would expect that the Q-data values would approach the Q-sim values (= smaller Q-data values) if the energy of the data event is corrected previously.
Therefore, in the following an energy correction of a single run will be presented to demonstrate this effect on figure 32.
As the underestimation in energy can be fluctuating for each run one has to investigate the reconstructed energies of the single run for the 1000 simulations at first, confer figure 33.

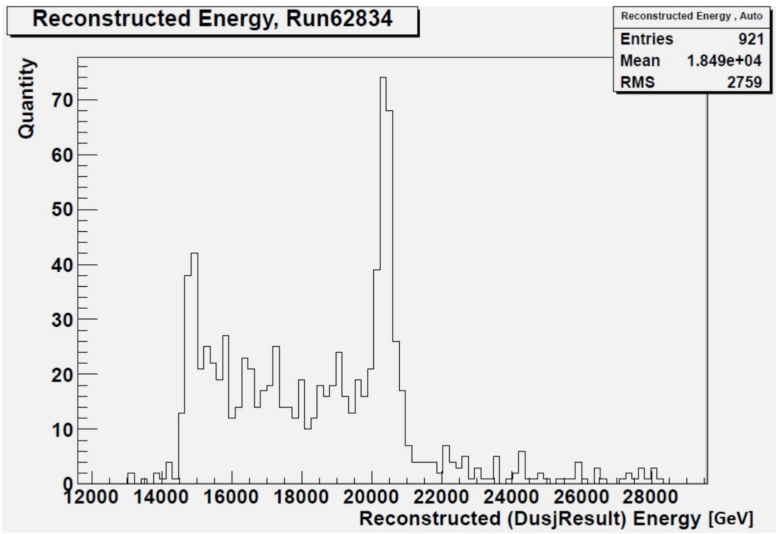
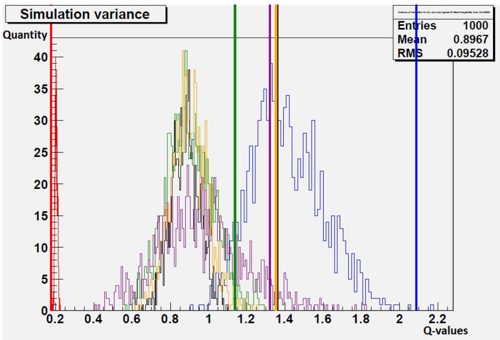


Figure 33: Distribution of the reconstructed energy for run 62834

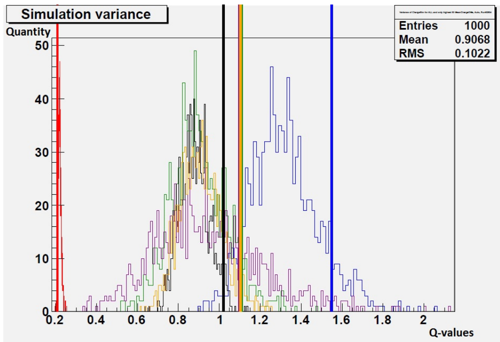
The originally reconstructed energy of this event in [3] is 28.1TeV. As one can clearly see in figure 33 the reconstructed energies are below this value by about 34 %. Since the distribution in figure 33 is broad, the mean value of it is taken to rescale the originally reconstructed energy. Of course, one could argue to take the position of the highest peak or other values but as this approach is just a basic demonstration the mean value is chosen.
With that in mind it is now possible to rescale the reconstructed energy of [3]:

$$\frac{28.1\text{TeV}}{18,49\text{TeV}} \cdot 28.1\text{TeV} = 42.7\text{TeV} \tag{16}$$

At this point, we can set the reconstructed event in Run 62834 with a corrected energy value as an input for the sim chain to investigate the effect of the energy shift on the Q-data values. To compare both the event with the original energy and the event with the new corrected energy one can now plot the Q-values of the sim chain for the corrected event in a histogram as seen in figure 34.



(a) Standard data neutrino sim input with an energy of 28.1TeV



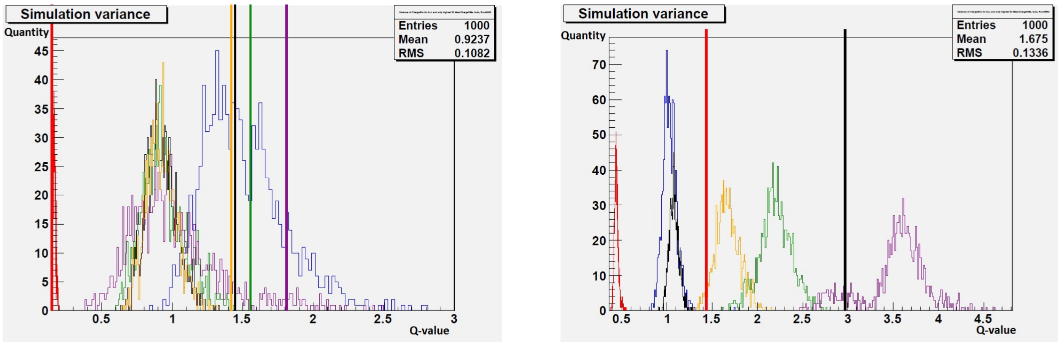
(b) Data neutrino input with a corrected energy of 42.7TeV

Figure 34: Comparison of the hit pattern for Run 62834 with either the original data neutrino or with the data neutrino with corrected energy: the Q-data values approach the Q-sim distribution with an event that is corrected in energy. Vertical lines: Q values calculated for the data event, distributions: Q-sim. "Hit": blue, "No Hit": red, "MeanCharge > 1": black, "10 PMTs with largest charge": magenta, "30 PMTs with largest charge": green and "50 PMTs with largest charge": orange.

As expected the Q-data values match the Q-sim distributions better than without a corrected energy. Hence, especially for investigations with the hit pattern approach it is advisable to correct all of the originally reconstructed data neutrinos from [3] with an individually upscaled energy.

10.2. Investigation of mirror solutions with the hit pattern approach

As the aim of the hit pattern method is to separate "bad" events from "good" events it could also help to identify which solution (normal vs. mirror) is correct.
For this reason, the hit patterns of the events from chapter 8 were investigated for both the standard data neutrino sim chain input and the mirror solution neutrino sim chain input. In these instances the energy was not corrected as it isn't necessary for the differentiation.
One example of comparing the hit patterns of these two simulation chain inputs can be seen in figure 35.



(a) Standard data neutrino sim input

(b) Mirror solution (vertex) neutrino sim input

Figure 35: Comparison of the hit pattern for Run 46852 with either the original data neutrino or with the mirror solution (vertex) neutrino: the hit pattern of the mirror solution input is significantly worse that the standard input. Vertical lines: Q values calculated for the data event, distributions: Q-sim. "Hit": blue, "No Hit": red, "MeanCharge $>$ 1": black, "10 PMTs with largest charge": magenta, "30 PMTs with largest charge": green and "50 PMTs with largest charge": orange.

In this case for the hit pattern approach the mirror solution input is far worse than the standard neutrino input. This is consistent with chapter 8 where it was seen that the original data vertex disappears if we set the mirror solution vertex as an input for the simulation chain.

Similar to Run 46852 the mirror solution hit patterns for Run 43639, 45946 and 51879 are worse, although not as bad as for Run 46852, than the original hit patterns. The only difference is Run 47679 (direction) where both the standard solution input and the mirror solution input are as good as the other. As of chapter 8 for this Run the standard and the mirror solution only change places in the pointspread if the mirror solution is set as an input for the simulation chain.

As a conclusion one can argue that a distinction between a "true" solution and a mirror solution with the hit pattern approach is not always possible. On top of this, the statistics are just too low for a general assertion as there are only 5 events out of the 59 events whose mirror solutions are separated good enough to be set as an input for the simulation chain.

11. Summary

Finally, after having developed all those methods one can now try to combine them to get an idea if a certain event is "good" (in terms of reconstruction quality) or "bad".
E.g. for Run 60226 the hit pattern (chapter 10) as well as the energy reconstruction (chapter 7) and the "DusjShowerRecoVertexFitReducedLogLikelihood" (chapter 5) is bad. Even though the pointspread looks good one should generally classify this event as a badly reconstructed event.

In the following, the same procedure shall be applied for the eight most energetic neutrino events.
The most energetic event that was found in [3] is run 43639 with an energy of 87.5 TeV. This event has a mirror solution in the direction which has about twice as many hits as the standard solution. However, there is no mirror solution in the vertex. Hence, as expected the vertex fit quality is good, as the mean of the vertex fit quality value is 7.8 with an RMS of 0.44. The average background rate of the PMTs is about 100 to 150 kHz which is acceptable. On top of this, the underestimation in energy is about 35 % which is not far from the average value for all runs of 28.8 %. If one sets the mirror solution as an input for the simulation chain, the standard and the mirror solution only change places and the hit pattern of the mirror solution input is worse than with the standard input. Hence, once cannot tell if the mirror or the standard solution is the correct one so that this event reconstruction should be considered as bad.

The next event that will be investigated is run 26397 with an energy of 42.1 TeV. This event has both a mirror solution in the vertex and in the direction. The mirror solution in the vertex occurs more often as the standard solution whereas for the solution in the direction it is the other way round. Thus, the vertex fit quality is worse than it is normally the case (Mean: 8.2, RMS: 0.5). The same applies for the combined energy and direction fit quality. Background rates are within reason, however, the underestimation in energy is about 60 % which is a huge difference from the average value for all runs. As such it seems that the reconstruction of energy is not very good for this run. As this event has mirror solutions in both the vertex and the direction it wasn't tried to set them as an input for the sim chain. Nevertheless this should be reconsidered in the future for a better classification of this event. Oddly enough the hit pattern for this event is actually very good, all of the Q-data values lie within the error bars of the Q-sim values. However using all of this information, one should still classify this event as a bad event because it is known that there are mirror solutions in both the vertex and the direction and on top of this the reconstruction of energy is rather bad.

Following, is run 46852 with an energy of 39.3 TeV. Similar to run 26397 this event has a mirror solution in both the vertex and the direction. This time the mirror solution in the vertex has as many hits as the standard solution in contrast to the mirror solution in the direction which occurs more often than the standard solution. Surprisingly the vertex fit quality is very good with a mean of 7.7 with an RMS of 0.5. The same applies for the combined fit in energy and direction. Background rates are good and the underestimation in energy on average is very good with about 0 %, however the RMS is also very large with a value of 60 %. This also means that this time the energy could be overestimated as well. If one sets the mirror solution in the vertex as an input for the simulation chain the standard vertex solution disappears, however as expected the situation for the direction is unaltered. At last the hit pattern is quite good and if one sets the mirror solution in the vertex as an input for the simulation chain the hit pattern gets very bad. This is another hint that the mirror solution in the vertex might indeed be a bad solution. In the future the same should also be tried for a mirror solution in the direction. In the end it could also be tried to set both the mirror solution in the vertex and the direction as an input simultaneously. As a result of the appearance of mirror solutions and the large RMS of the reconstruction in energy this event should also be considered as bad.

The next event is run 28722 with an energy of 39.1 TeV. This event has a very weak, in terms of hits out of the 1000 simulations, mirror solution in both the direction and the vertex. Interestingly, one can see two peaks in the distribution of the vertex fit quality: a large (mean of about 7.5) and a small one (mean of about 10.8). This indicates that the minor mirror solutions are indeed bad ones. The same effect can be seen in the combined energy and direction fit. Background rates are normal and the underestimation of energy is about 45 % with an RMS of about 35 %. Setting the mirror solutions as an input for the simulation wasn't tried in this work because the mirror solutions are so weak. The hit pattern looks good, however, the Q-sim distributions are significantly smaller than the average Q-sim values for a run. As it can't be rouled out that the reconstructed data neutrino isn't a reconstructed mirror solution this event should still be classified as bad.

Coming up is run 62834 with an energy of 28.1 TeV. This event has no mirror solution at all. As expected both the vertex fit quality and the combined energy and direction fit quality is very good. The background light is ok and the underestimation of energy is about 20 % with a an RMS of 10 % that is smaller than on average for all runs. On top of this, the hit pattern is very good for this event. Hence, this event should be considered as a well reconstructed event.

The next event is run 49425 with an energy of 21.4 TeV. Similar to run 62834 it has no mirror solution and both fit qualities are good. The background rates of this event are quite high, however as these are not the exact background rates at the event time one can't figure out if they were equally bad at the point in time of the neutrino interaction for this event. The underestimation in energy is about 28 % with an RMS of about 30 %. Lastly, the hit pattern is also quite good. As a result one should classify this event as a good event even though it seems that its reconstruction is not as good as it is for run 62834.

Another highly energetic event is run 27893 with an energy of 16.3 TeV. This event has no mirror solution at all. One could argue that it has a very, very weak mirror solution in the direction but that would just be definition dependent. The vertex fit quality is very good whereas the combined fit quality is only mediocre. Background rates are normal and the underestimation of energy is about 8 % with an RMS of about 55 %. However, the hit pattern is very good. As a consequence, it looks like this event should be considered as a "good" event.

The last investigated neutrino candidate is run 51879 with an energy of 15.6 TeV. This event has a strong mirror solution in the direction which occurs more often than the standard solution, however, there's no mirror solution in the vertex. The vertex fit quality is good as one would expect it to be and interestingly the combined fit quality is also good. Background rates are ok and the underestimation in energy is about 50 % with a large RMS of about 200 %. Contrarily the hit pattern is very good. As a result the reconstruction of this event seems to be good in terms of vertex position and neutrino direction, however the RMS of the reconstructed energy is very large. As such one should consider this event as a bad event.

In general the standard practice to be taken should be to investigate runs with the several methods developed in this work (hit pattern, vertex resolution, angular resolution, energy reconstruction, ...) individually so that one can get an idea how good an event was reconstructed regarding certain fit values (energy, direction, vertex, ...). After having done that for all events the next step would be to create a table where the outcome of every method (e.g. "good" or "bad") is specified. Hereby it could be figured out which events are mostly "bad". However, on that note it should also be determined which reconstruction parameters are the most important ones.
E.g. a misreconstructed vertex alone wouldn't be a severe issue because the reconstructed vertex position isn't necessarily a crucial parameter if one wants to conduct neutrino astronomy.
Another subject that should be undertaken in the future is to rescale all of the 59 events individually to a higher energy as the Dusj reconstruction algorithm underestimates it. As a result the hit pattern analysis should be easier because the Q-data offsets from the Q-sim distributions should be alleviated.

On top of this, as the final aim of a neutrino telescope should be to create a skymap of cosmic neutrino sources, mirror solutions especially in the direction should be further investigated. With the applied methods in this work one could learn a lot about those mirror solutions, however, it was not possible to properly distinguish between "true" solutions and mirror solutions.

As a closing remark it can be said that still a lot can be learned about Dusj reconstructed events by investigating single possible neutrino events as it was demonstrated in this work.

References

[1] R. Mitalas and K. R. Sills. "On the photon diffusion time scale for the sun." In: *Astrophysical Journal* 401 (Dec. 1992), p. 759. DOI: 10.1086/172103.

[2] M. G. Aartsen et al. "Observation of High-Energy Astrophysical Neutrinos in Three Years of IceCube Data." In: *Physical Review Letters* 113.10, 101101 (Sept. 2014), p. 101101. DOI: 10.1103/PhysRevLett.113.101101. arXiv: 1405.5303 [astro-ph.HE].

[3] Florian Folger. "Search for a diffuse cosmic neutrino flux using shower events in the ANTARES neutrino telescope." doctoral thesis. Friedrich-Alexander Universität Erlangen-Nürnberg, 2014.

[4] U. F. Katz and C. Spiering. "High-energy neutrino astrophysics: Status and perspectives." In: *Progress in Particle and Nuclear Physics* 67 (July 2012), pp. 651–704. DOI: 10.1016/j.ppnp.2011.12.001. arXiv: 1111.0507 [astro-ph.HE].

[5] URL: http://antares.in2p3.fr/Gallery/3D/princip1.jpeg.

[6] Claudio Kopper. "Performance Studies for the KM3NeT Neutrino Telescope." doctoral thesis. Friedrich-Alexander Universität Erlangen-Nürnberg, 2010.

List of Figures

Appendix A Important parameters for the 59 investigated events

The following tables were taken from [3]. Hereby one can derive the most crucial aspects of a certain event.

Run ID	Event ID	Trigger counter	Detector configuration	Date [UTC]	Fitted energy [TeV]	Fitted zenith [°]	# hits / strings	Total charge [pe]	Vertex position	Run burst fraction	Run mean rate [kHz]	Quality basic
26397	**129099**	**4046**	**5Line**	**2007-03-02 06:37**	**42.1**	**126**	**42/3**	**168**	**29 *h*, 45 *v****	**0.41**	**91**	**1**
27893	**74913**	**21742**	**5Line**	**2007-05-26 19:43**	**16.3**	**98**	**75/3**	**321**	**61 *h****	**0.05**	**63**	**4**
28446	20087	96972	5Line	2007-06-22 11:45	2.6	177	64/4	383	Detector	0.06	65	4
28722	**45202**	**426**	**5Line**	**2007-07-13 03:29**	**39.1**	**106**	**186/5**	**1373**	**22 *h****	**0.04**	**63**	**4**
29551	113754	1678	5Line	2007-09-19 19:35	2.1	99	47/4	194	25 *h**	0.57	125	1
30162	59159	703	5Line	2007-11-02 07:27	1.6	115	51/3	197	20 *h**	0.11	60	1
30336	124385	1501	5Line	2007-11-17 14:47	1.1	111	42/3	193	Center	0.07	55	4
30592	10231	13420	10Line	2007-12-10 13:39	1.9	109	64/5	264	Detector	0.16	60	4
31275	90189	1336	10Line	2008-01-12 18:18	1.6	108	56/5	319	Detector	0.07	52	4
32283	70238	18906	10Line	2008-02-24 13:33	3.0	134	85/4	568	Detector	0.02	54	4
33713	100118	9127	10Line	2008-04-23 03:48	1.5	159	47/4	210	Detector	0.14	63	4
34796	33243	6663	12Line	2008-06-12 18:08	1.8	108	55/3	217	Detector	0.04	63	4
35336	42797	10550	12Line	2008-09-13 11:48	1.0	101	44/4	146	Center	0.22	148	3
35933	30513	74729	12Line	2008-10-02 21:20	1.1	143	70/6	282	Detector	0.28	194	3
36475	31623	9522	12Line	2008-10-18 02:45	4.7	122	67/4	648	14 *v**	0.22	145	3
36552	7188	4464	12Line	2008-10-21 23:33	1.6	150	56/4	280	Center	0.48	344	2
37783	25390	13777	12Line	2008-12-12 08:40	4.7	112	70/5	280	Detector	0.15	92	4
39104	57792	37941	12Line	2009-02-14 17:29	2.8	117	50/5	168	7 *v**	0.33	191	3
39122	63278	81877	12Line	2009-02-15 18:21	8.4	117	48/4	167	47 *h**	0.10	73	4
39331	63304	50873	12Line	2009-02-25 13:39	7.4	105	51/5	151	19 *h*, 44 *v**	0.14	96	4
40119	6	46	12Line	2009-04-13 08:02	2.2	104	52/3	452	Detector	0.31	243	3
42119	148518	280014	12Line	2009-07-25 19:47	6.0	101	73/4	429	Detector	0.09	118	4
43639	**66178**	**479501**	**12Line**	**2009-10-05 06:08**	**87.5**	**130**	**36/3**	**74**	**84 *h*, 1 *v****	**0.09**	**90**	**4**
45946	71285	2061417	12Line	2010-01-20 08:21	1.1	134	50/4	197	Detector	0.11	73	4

* Distance in meters to detector edge, (h) horizontally, (v) vertically

Figure 36: Most important parameters for the 59 investigated events

Run ID	Event ID	Trigger counter	Detector configuration	Date [UTC]	Fitted energy [TeV]	Fitted zenith [°]	# hits / strings	Total charge [pe]	Vertex position	Run burst fraction	Run mean rate [kHz]	Quality basic
46009	9533	400972	12Line	2010-01-23 01:54	0.9	129	47/3	283	Detector	0.09	64	4
46852	**51708**	**917709**	**12Line**	**2010-02-25 19:44**	**39.3**	**143**	**91/6**	**603**	**84 v***	**0.13**	**87**	**1**
47679	12186	77946	12Line	2010-03-30 16:42	9.1	148	69/3	557	5 h, 20 v*	0.23	138	1
47838	65111	673623	12Line	2010-04-06 03:50	0.8	124	44/3	310	8 h*	0.08	67	4
49425	**32175**	**104853**	**12Line**	**2010-06-19 10:41**	**21.4**	**101**	**88/6**	**562**	**22 v***	**0.36**	**235**	**1**
49518	64893	1587	12Line	2010-06-22 11:44	1.0	105	50/5	143	Center	0.60	438	1
50675	10418	889360	12Line	2010-07-29 03:20	6.7	136	60/5	269	5 h*	0.53	396	1
51411	20931	590464	12Line	2010-08-23 00:46	2.5	114	65/5	420	Detector	0.26	163	3
51879	**9640**	**274664**	**12Line**	**2010-09-10 11:27**	**15.0**	**120**	**50/3**	**318**	**40 h***	**0.16**	**100**	**4**
51991	6122	153796	12Line	2010-09-15 06:43	5.4	94	71/4	380	Detector	0.18	103	4
52087	97516	465265	12Line	2010-09-20 06:03	0.3	160	37/3	169	Detector	0.25	145	3
53119	29228	609027	12Line	2010-11-14 20:33	1.2	145	55/3	265	Detector	0.06	56	4
54326	36824	4031901	12Line	2011-01-04 10:05	1.2	110	63/4	371	10 h*	0.16	85	4
55894	106093	204907	12Line	2011-03-11 01:47	2.9	114	82/6	387	Detector	0.27	136	1
58045	78338	186199	12Line	2011-06-14 04:55	1.1	157	53/3	273	Detector	0.11	111	4
58498	29935	388601	12Line	2011-07-07 06:08	1.4	106	45/4	250	27 h*	0.42	204	1
58686	30898	1716155	12Line	2011-07-16 15:35	0.8	107	49/4	332	Detector	0.11	83	4
58812	25119	67234	12Line	2011-07-24 03:54	1.1	101	46/4	126	Detector	0.41	232	1
58892	39915	17513	12Line	2011-08-04 15:44	0.5	154	47/4	181	Center	0.10	90	4
59425	99043	90861	12Line	2011-09-01 07:23	0.7	100	48/5	200	Detector	0.15	101	4
59692	54812	680785	12Line	2011-09-15 08:33	5.9	99	61/4	331	51 h*	0.16	95	4
59712	19763	1057429	12Line	2011-09-16 06:00	1.4	100	39/4	230	4 h*	0.19	104	4
59797	6532	2136	12Line	2011-09-19 01:20	4.8	94	77/7	452	Detector	0.19	109	4
60226	91161	1165187	12Line	2011-10-12 02:13	1.5	135	46/4	166	Detector	0.31	132	1

* Distance in meters to detector edge, (h) horizontally, (v) vertically

Figure 37: Most important parameters for the 59 investigated events

Run ID	**Event ID**	**Trigger counter**	**Detector configuration**	**Date [UTC]**	**Fitted energy [TeV]**	**Fitted zenith [°]**	**# hits / strings**	**Total charge [pe]**	**Vertex position**	**Run burst fraction**	**Run mean rate [kHz]**	**Quality basic**
60704	47177	5632054	12Line	2011-11-02 10:54	1.7	133	50/5	205	Detector	0.09	65	4
60861	33966	41994	12Line	2011-11-09 14:24	2.8	120	66/3	465	Detector	0.17	84	4
61975	85577	2908926	12Line	2012-01-04 17:15	1.5	97	59/4	153	Center	0.10	62	4
62139	7176	2287775	12Line	2012-01-13 04:49	2.7	102	55/3	427	15 h*	0.37	220	3
62228	38130	279600	12Line	2012-01-16 16:19	7.1	96	97/7	519	Detector	0.20	94	4
62834	**30474**	**384475**	**12Line**	**2012-02-19 17:11**	**28.1**	**118**	**99/7**	**456**	**69 v***	**0.16**	**66**	**4**
63013	66930	11840	12Line	2012-02-29 01:07	7.8	137	65/4	263	36 h*	0.38	214	3
64914	34922	24770	12Line	2012-06-19 05:00	4.5	99	62/5	517	Detector	0.39	307	2
65259	40468	89673	12Line	2012-07-04 21:42	1.6	96	59/4	227	Detector	0.13	106	4
65434	19098	251505	12Line	2012-07-12 16:09	0.3	115	40/3	133	Center	0.37	198	3
66772	36322	3313506	12Line	2012-09-19 02:56	1.8	126	48/4	313	Detector	0.12	67	1
67723	71283	449623	12Line	2012-11-06 07:31	2.3	175	55/7	220	34 v*	0.23	103	1

* Distance in meters to detector edge, (h) horizontally, (v) vertically

Figure 38: Most important parameters for the 59 investigated events

Appendix B Pointspread table

In this table the 8 most energetic neutrino events are marked in bold font.
The errors that are given are always the minimum error as it only indicates error caused by bin sizes. The errors for the hotspot spread are the same as for the hotspot offset and they were cut in the table for readability reasons.
The parentheses behind the "yes" in the mirror solution indicate if the Run has a strong (and at least equal frequency in comparison to the "normal" solution) or weak (else) mirror solution. The minimum amount for a weak mirror solution is defined as 100 entries in a localized cluster.
All values were rounded to the second decimal place.

RunID	hotspot offset [rad] (Zenith, Azimuth)	hotspot spread [rad] (Zenith, Azimuth)	Mirror solution?
26397	$(0 \pm 0.05, -0.4 \pm 0.15)$	(0.25, 0.15)	yes (strong)
27893	$(0.03 \pm 0.02, -0.06 \pm 0.03)$	(0.04, 0.04)	no
28446	$(-0.1 \pm 0.02, -0.03 \pm 0.02)$	(0.09, 0.02)	no
28722	$(-0.1 \pm 0.02, 0.05 \pm 0.02)$	(0.04, 0.4)	no
29551	$(0.04 \pm 0.02, -0.06 \pm 0.02)$	(0.06, 0.08)	yes (strong)
30162	(-)	(-)	no
30336	$(0 \pm 0.02, 0 \pm 0.02)$	(0.08, 0.06)	no
30592	$(-0.01 \pm 0.02, -0.08 \pm 0.02)$	(0.14, 0.07)	no
31275	$(0.13 \pm 0.02, 0.12 \pm 0.02)$	(0.1, 0.05)	no
32283	$(0.04 \pm 0.02, -0.06 \pm 0.02)$	(0.05, 0.04)	no
33713	$(0 \pm 0.02, -0.03 \pm 0.02)$	(0.09, 0.08)	no
34796	$(0 \pm 0.02, 0 \pm 0.02)$	(0.06, 0.04)	yes (strong)
35336	$(0.03 \pm 0.02, 0.05 \pm 0.02)$	(0.08, 0.06)	no
35933	$(-0.13 \pm 0.02, 0.19 \pm 0.02)$	(0.14, 0.04)	no
36475	$(0 \pm 0.02, 0 \pm 0.02)$	(0.05, 0.05)	no
36522	$(0 \pm 0.02, -0.03 \pm 0.02)$	(0.08, 0.08)	yes (weak)
37783	$(0.1 \pm 0.02, 0 \pm 0.02)$	(0.07, 0.05)	no
39104	$(-0.04 \pm 0.02, 0.1 \pm 0.02)$	(0.07, 0.1)	yes (strong)
39122	$(0 \pm 0.02, -0.05 \pm 0.02)$	(0.23, 0.12)	no
39331	$(0.05 \pm 0.09, 0 \pm 0.15)$	(0.3, 0.23)	no
40119	$(0.07 \pm 0.02, 0 \pm 0.02)$	(0.13, 0.07)	no
42119	$(0.04 \pm 0.02, 0.07 \pm 0.02)$	(0.13, 0.05)	no

Table 1: Offset and spread of the pointspread plots for each run

RunID	hotspot offset [rad] (Zenith, Azimuth)	hotspot spread [rad] (Zenith, Azimuth)	Mirror solution?
43639	(0.01 ± 0.02, 0 ± 0.02)	(0.18, 0.14)	yes (strong)
45946	(0.2 ± 0.02, 0 ± 0.02)	(0.28, 0.22)	yes (strong)
46009	(-0.01 ± 0.02, 0 ± 0.02)	(0.1, 0.03)	no
46852	(0.08 ± 0.02, -0.03 ± 0.02)	(0.05, 0.07)	yes (strong)
47679	(0.08 ± 0.02, -0.02 ± 0.02)	(0.06, 0.07)	yes (strong)
47838	(-)	(-)	no
49425	(0 ± 0.02, 0.1 ± 0.02)	(0.07, 0.12)	no
49518	(0.05 ± 0.02, 0.18 ± 0.02)	(0.2, 0.18)	no
50675	(-0.01 ± 0.08, 0 ± 0.02)	(0.38, 0.13)	no
51411	(0.05 ± 0.02, 0.2 ± 0.02)	(0.13, 0.1)	no
51879	(0 ± 0.02, 0.02 ± 0.02)	(0.25, 0.11)	yes (strong)
51991	(0.35 ± 0.04, 0.1 ± 0.18)	(0.26, 0.09)	no
52087	(0.01 ± 0.02, 0.03 ± 0.02)	(0.15, 0.1)	no
53119	(0.1 ± 0.02, 0.07 ± 0.02)	(0.11, 0.1)	no
54326	(0.05 ± 0.02, -0.05 ± 0.02)	(0.14, 0.13)	no
55894	(0 ± 0.02, 0.01 ± 0.02)	(0.08, 0.06)	no
58045	(0 ± 0.02, -0.02 ± 0.02)	(0.08, 0.1)	no
58498	(0.03 ± 0.02, -0.04 ± 0.02)	(0.06, 0.08)	no
58686	(0.07 ± 0.02, 0.03 ± 0.02)	(0.1, 0.04)	no
58812	(0.09 ± 0.02, 0.05 ± 0.02)	(0.25, 0.13)	no
59425	(0 ± 0.02, 0 ± 0.02)	(0.18, 0.15)	no
59692	(0 ± 0.02, -0.11 ± 0.02)	(0.08, 0.06)	no
59712	(0.11 ± 0.02, 0.08 ± 0.02)	(0.2, 0.15)	no
59797	(0.15 ± 0.02, -0.01 ± 0.02)	(0.12, 0.03)	no
60226	(0 ± 0.04, -0.1 ± 0.06)	(0.57, 0.18)	no
60704	(-0.01 ± 0.02, 0.06 ± 0.02)	(0.11, 0.11)	yes (strong)
60861	(-0.03 ± 0.02, -0.02 ± 0.02)	(0.11, 0.1)	no
61975	(0.2 ± 0.02, 0.12 ± 0.02)	(0.17, 0.1)	no
62139	(0.05 ± 0.02, -0.01 ± 0.02)	(0.2, 0.21)	no
62228	(0 ± 0.02, 0 ± 0.02)	(0.06, 0.04)	no
62834	(0 ± 0.02, -0.03 ± 0.02)	(0.02, 0.05)	no
63013	(0 ± 0.02, -0.06 ± 0.02)	(0.1, 0.1)	no
64914	(0.05 ± 0.02, -0.03 ± 0.02)	(0.06, 0.06)	no
65259	(0.05 ± 0.02, -0.03 ± 0.02)	(0.18, 0.19)	no
65434	(-0.1 ± 0.02, 0.03 ± 0.02)	(0.2, 0.14)	yes (strong)
66772	(-0.05 ± 0.06, -0.2 ± 0.13)	(0.09, 0.13)	no
67723	(-0.05 ± 0.02, -0.02 ± 0.02)	(0.08, 0.06)	no

Table 2: Offset and spread of the pointspread plots for each run

Statutory Declaration

I declare that I have developed and written the enclosed thesis entirely by myself and have not used sources or means without declaration in the text. Any thoughts or quotations which were inferred from these sources are clearly marked as such. This thesis was not submitted in the same or in a substantially similar version, not even partially, to any other authority to achieve an academic grading and was not published elsewhere.

Michael Moser Erlangen, 30.10.2014

Acknowledgements

Zuletzt möchte ich mich bei all denjenigen bedanken, die mich beim Anfertigen dieser Bachelorarbeit unterstützt haben. Herzlich bedanken möchte ich mich insbesondere bei

- meiner Familie, die mich während des Studiums und insbesondere während dieser Arbeit immer unterstützt haben
- meinem Betreuer Thomas Eberl für die Vergabe dieses sehr interessanten Themas und dafür, dass du immer ein offenes Ohr und Zeit für mich hattest. Ohne diese Unterstützung wäre diese Arbeit nicht möglich gewesen
- meinem Betreuer Jannik Hofestädt, der mir insbesondere bei den softwarespezifischen Problemstellungen dieser Arbeit stark weitergeholfen und damit diese Arbeit ermöglicht hat
- meinen Zimmerkollegen Thomas Heid und Steffen Hallmann für die freundliche Aufnahme und dafür, dass sie immer ein offenes Ohr für progammiertechnische Fragen hatten
- der ECAP Beachvolleyball Gruppe, die mir eine dringend nötige Abwechslung zur Arbeit geboten hat