Event identification for KM3NeT/ARCA

Thomas Heid, on behalf of the KM3NeT Collaboration
Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen Centre for Astroparticle Physics
E-mail: thomas.heid@fau.de

Abstract. KM3NeT is a large research infrastructure consisting of a network of deep-sea neutrino telescopes. KM3NeT/ARCA will be the instrument detecting high-energy neutrinos with energies above 100 TeV. This instrument gives a new opportunity to observe the neutrino sky with very high angular resolution to be able to detect neutrino point sources. Furthermore it will be possible to probe the flavour composition of neutrino fluxes, and hence production mechanisms, with so-far unreached precision. Neutrinos produce different event topologies in the detector according to their flavour, interaction channel and deposited energy. Machine-learning algorithms are able to learn features of topologies to discriminate them. In previous analyses only two event types were regarded, namely the shower and track topology. With good timing resolution and precise reconstruction algorithms it is possible to separate into more event types, for example the double bang topology produced by tau neutrinos. The final goal is to distinguish all three neutrino flavors as much as possible. To resolve this issue the KM3NeT collaboration uses deep neural networks trained with Monte Carlo events of all neutrino types. This contribution shows the ability of KM3NeT/ARCA to classify events in more than two neutrino event topologies. Furthermore, the borders between detectable classes are shown, such as the minimum distance the tau has to travel before decaying into a tau neutrino to be detected as double bang event.

1. From light patterns in KM3NeT/ARCA neutrino telescope to event types
The main objective of the KM3NeT/ARCA neutrino telescope is to explore the sources of high-energy astrophysical neutrinos [1], where ARCA stands for Astroparticle Research with Cosmics in the Abyss. Evidence of a diffuse astrophysical neutrino flux has been discovered by IceCube [2, 3]. A key observable is the flavour composition, which holds information on the production mechanisms, non-standard oscillations and other new physics. Solving these issues requires new methods with improved neutrino flavour separation power.

In this work advanced topological features, calculated with simulated events, are used to train a neural network for the separation into topologically inspired classes. This work will be a key input to identify the flavour composition in future analyses.

2. Implementation and performance
Event identification in KM3NeT/ARCA is trained on features, which are characterising values calculated from light patterns in the detector. These enable the algorithms to separate several event classes. A feature can also be an output value of reconstruction algorithms such as reconstructed energy or quality parameters, as described in the KM3NeT Letter of Intent [1].

The standard simulation of KM3NeT/ARCA [1] is used. For training only well-reconstructed neutrinos are selected. Atmospheric muons complete the sample for evaluation. The statistics
used for training and evaluation are on the order of $10^4$ events per target class in the range from $10^3$ GeV to $10^8$ GeV.

Five classes were defined for training. Track events are events with a high-energy muon. These are mainly produced in charged current (CC) interactions of muon neutrinos. Track events are divided into three sub-classes. The sub-classes are: up-going and down-going, through-going tracks and the starting tracks. Up and down-going events are chosen to keep the possibility to reject atmospheric muons. The fifth class, cascade-like events, originate from neutral current (NC) interactions and CC electron neutrino interactions. All these have in common the topology of one single cascade. Double bang events are defined as tau neutrinos with a decaying tau with a path length larger than 20 m. Taus with shorter path length are assumed to be indistinguishable from a single cascade and are therefore classified as such.

The complete learning and application chain consists of several steps, which are explained in the following. In the first step each feature is standardised by removing the mean and scaling to unit variance (otherwise some features can dominate the learning process over others). Here, the scikit-learn package [4] is used. From around 100 initial features, a principle component analysis is used to reduce the number to 57.

To avoid imbalanced data-sets in each energy range the sample is homogenised in logarithmic energy for each class.

The python theanets package [5] is used to build up a three-layer neural network. The input layer has as many neurons as input features. The output layers consists of as many neurons as output classes. The hidden layer is optimised to 40 neurons. The final goal of optimising the hyper-parameters (the parameters set before the actual training) of a neural network is to get high sensitivity in some physics analysis. In the present case, which is not devoted to a specific analysis, we have focused on the energy range above $10^5$ GeV where the IceCube findings indicate a dominance of the astrophysical neutrino flux [2]. This analysis tries to achieve a mostly uniform behaviour of performance in this region.

Several sets of hyper-parameters were tested. For each set an independent neural network is trained. To get an estimate of the stability of the procedure cross-validation is applied. The complete sample is split into ten sub-samples. An independent training is performed on one sub-sample and then applied on the remaining sub-sets. This is repeated for all then combinations.

The figures 1 and 2 show the performance of the final network for each of the target classes. Error bars include both the statistical uncertainties and the uncertainties of the trained classifier. Further studies have shown that the statistical uncertainties are the most dominating part in this case. This leads to the conclusion that the classifier is very stable and that the variance is small enough so that one can use this for further analyses.

The left panel of figure 1 shows that the fraction of muon neutrino CC events correctly classified (as through-going in the upwards direction and starting events) is above 90 percent. Main contributions to other classes are to the double bang events in the range from $10^6$ GeV to $10^7$ GeV. Very high-energy events also show the tendency to be reconstructed as up-going. The correlation between through-going and starting events can also be seen in the unclassified distribution. That is an artefact resulting from the event selection and reconstruction.

The right panel of figure 1 shows the classification in the case of CC anti-electron neutrino events. The fraction of events correctly classified as cascades is around 90%. For energies around $10^{6.8}$ GeV (Glashow resonance) a larger fraction of events tends to be classified as down-going. For energies above $10^7$ GeV the fraction of events classified as double-bang increases.

Figure 2 shows the distribution of target classes resolved in the tau path length. The threshold where more tau double bang events can be correctly classified is at around 30 meters, which is slightly above the defined threshold of 20 m. For path lengths exceeding 1000 m one of the showers will necessarily occur outside the detector volume and thus won’t be detected.
Figure 1. Distribution of charged current anti-electron neutrino events (right) and up-going charged current muon neutrinos (left) to target classes. Error bars include both the statistical errors and the uncertainty in the classifier.

Figure 2. Distribution of charged current tau neutrino event, where the tau lepton decay to a further shower, to target classes. Error bars show composition of statistical errors and uncertainty in the classifier.

3. Event Identification and its implications to further studies

The machine learning algorithm developed for KM3NeT/ARCA has the capability of correctly classifying cascade and track events at a 90% level. For this first study specialised reconstructions developed specifically to identify event classes (e.g. an in-development tau double-bang reconstruction) was not yet used. With this and other methods, one expects further improvements. With this event identification first studies towards a flavor composition analysis are under way.

References
[1] KM3NeT-Collaboration 2016 *Journal of Physics G: Nuclear and Particle Physics* **43** 084001
[2] IceCube-Collaboration (IceCube) 2013 *Science* **342** 1242856
[3] IceCube-Collaboration (IceCube) 2015 *Phys. Rev. Lett.* **114** 171102
[4] Pedregosa F and et al 2011 *Journal of Machine Learning Research* **12** 2825–2830
[5] Johnson L and et al 2016 theanets: Version 0.7.3 https://github.com/lmjohns3/theanets