PROCEEDINGS OF SCIENCE

Selection techniques of neutrino-induced cascades in the Baikal-GVD neutrino telescope

V.A. Allakhverdyan,[a] A.D. Avrorin,[b] A.V. Avrorin,[b] V.M. Aynutdinov,[b] Z. Bardačová,[d,*] I.A. Belolaptikov,[a] I.V. Borina,[a] N.M. Budnev,[e] V.Y. Dik,[a,j] G.V. Domogatsky,[b] A.A. Doroshenko,[b] R. Dvornický,[d] A.N. Dyachok,[e] Zh.-A.M. Dzhilkibaev,[b] E. Eckerová,[d] T.V. Elzhov,[a] L. Fajt,[f] A.R. Gafarov,[e] K.V. Golubkov,[b] N.S. Gorshkov,[a] T.I. Gress,[e] K.G. Kebkal,[c] O.G. Kebkal,[c] A. Khatun,[d] E.V. Khramov,[a] M.M. Kolbin,[a] K.V. Konischev,[a] A.V. Korobchenko,[a] A.P. Koshechkin,[b] V.A. Kozhin,[h] M.V. Kruglov,[a] V.F. Kulepov,[g] Y.M. Malyshkin,[a] M.B. Milenin,[b] R.R. Mirgazov,[e] D.V. Naumov,[a] V. Nazari,[a] D.P. Petukhov,[b] E.N. Pliskovsky,[a] M.I. Rozanov,[i] V.D. Rushay,[a] E.V. Ryabov,[e] G.B. Safronov,[b] D. Seitova,[a,j] B.A. Shaybonov,[a] M.D. Shelepov,[b] F. Šimkovic,[d,f] A.E. Sirenko,[a] A.V. Skurikhin,[h] A.G. Solovjev,[a] M.N. Sorokovikov,[a] I. Štekl,[f] A.P. Stromakov,[b] O.V. Suvorova,[b] V.A. Tabolenko,[e] Y.V. Yablokova[a] and D.N. Zaborov[b]

[a] *Joint Institute for Nuclear Research, Dubna, Russia*
[b] *Institute for Nuclear Research, Russian Academy of Sciences, Moscow, Russia*
[c] *EvoLogics GmbH, Berlin, Germany*
[d] *Comenius University, Bratislava, Slovakia*
[e] *Irkutsk State University, Irkutsk, Russia*
[f] *Czech Technical University in Prague, Prague, Czech Republic*
[g] *Nizhny Novgorod State Technical University, Nizhny Novgorod, Russia*
[h] *Skobeltsyn Institute of Nuclear Physics, Moscow State University, Moscow, Russia*
[i] *St. Petersburg State Marine Technical University, St.Petersburg, Russia*
[j] *Institute for Nuclear Physics ME RK, Almaty, 050032 Kazakhstan*

E-mail: bardacova4@uniba.sk

The neutrino telescope Baikal-GVD (Gigaton Volume Detector) has been designed to search for high-energy neutrino cosmic sources. It is located in pure water of Lake Baikal at a depth of 1366 m. Currently (year 2022) Baikal-GVD comprises 2880 optical modules divided to 10 independently operating clusters. Optical modules detect flashes of Cherenkov light from secondary charged particles induced in interactions of neutrinos with matter. Some charged and neutral current neutrino interactions lead to hadronic or electromagnetic cascade events. Apart from the neutrino cascades, the cascade-like light topologies can be also induced along the muon tracks. These event signatures, referred to as background cascades arise from the discrete stochastic energy losses of the muon. The cascades produced along the atmospheric muon bundles constitute the main background in neutrino cascade channel. In this paper, a developed, optimized, and tested algorithm for suppression of the background cascades is presented.

27th European Cosmic Ray Symposium - ECRS
25-29 July 2022
Nijmegen, the Netherlands

[*] Speaker

PoS(ECRS)098

1. The Baikal-GVD Neutrino Experiment

The Baikal Gigaton Volume Detector is an underwater cubic kilometer scale neutrino telescope deployed in the southern part of Lake Baikal. It is located 3.6 km offshore at a depth of 1366 m [1]. The experiment aims to detect high-energy astrophysical neutrinos using a 3D array of 13-inch optical modules (OMs) housing 10-inch Hamamatsu photo-multiplier tube (PMT). These detector elements record the Cherenkov photons from secondary particles produced in neutrino interactions. The OMs are arranged along vertical strings anchored on the lakebed and pulled up by a buoy. The strings are in turn sub-arranged into functionally independent clusters (see Fig. 1). A single cluster consists of 8 strings with 36 OMs each. Currently (year 2022) the Baikal-GVD is composed of 10 clusters. The cluster effective volume for neutrino cascade events with energy above 100 TeV is ≈ 0.05 km^3 [2]. As of April 2022, a total Baikal-GVD effective volume is ≈ 0.5 km^3, hence it is currently the largest neutrino telescope in the Northern Hemisphere.

There are two main channels of neutrino detection. One consists of cascade events induced by charged current (CC) electron neutrino (ν_e) and tau neutrino (ν_τ) interactions, as well as neutral current interactions (NC) of all neutrino flavors. The other one consists of track events, i.e. muons, which result from CC muon neutrino (ν_μ) interactions. The cascade events are of great interest as most of the high-energy cosmic neutrino events identified by the IceCube experiment were cascades [3]. However, the overwhelming background from downgoing muon bundles, produced in air showers in the atmosphere above the detector dominates in both detection channels. The so-called background cascades originate from discrete stochastic energy losses along muon tracks (bremsstrahlung, photonuclear processes, direct electron-positron pair production). The illustration of background cascades along one muon track is shown in Fig. 2 (left).

In this paper, developed selection techniques for distinction between neutrino-induced cascades and background cascades are presented. The techniques were developed and optimized using Monte Carlo (MC) simulations for the early part of the 2019 season. The individual selection methods result in the output variables, which were in turn used for training a machine learning algorithm - Boosted Decision Trees (BDTs) [4]. The presented analysis is performed only with single-cluster data.

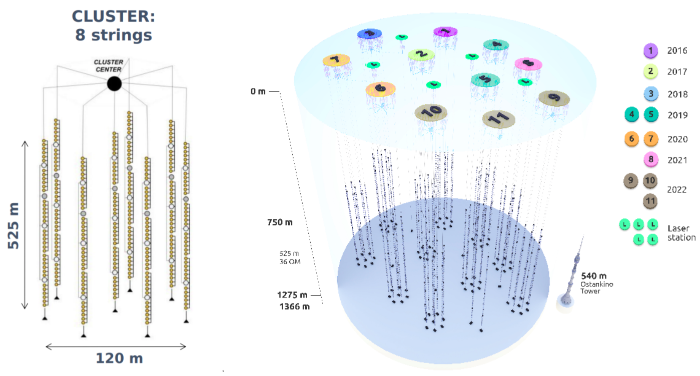

Figure 1: Illustration of the Baikal-GVD configuration. An individual cluster with 8 strings is depicted on the left side. The right side shows the detector in 2022 consisting of 10 clusters and 2 strings of the 11th cluster. The detector deployment annual progression is shown in the legend.

2. Monte Carlo - Data comparison and Suppression Techniques

The separation between cascades induced in neutrino interactions and background cascades from atmospheric muons is an important task. Standard rejection strategy for the atmospheric muon bundles is to select only upward-going events. However, the background cascade can be misreconstructed as an upgoing event. For that reason, it is necessary to develop a technique, which enables us to efficiently suppress the background cascades.

To perform data-MC comparison, experimental data from the early part of the season 2019 (between April 1 and June 30) and corresponding MC were used. Total effective livetime of the experimental data sample is 353 days (for all 5 clusters together). A cascade reconstruction software which utilizes parallel tool [5] has been applied for processing of both MC and experimental data. Fig. 2 (right) shows the comparison between the rate of reconstructed cascade events for experimental data and MC datasets as a function of the reconstructed zenith angle. Cascade-like events with horizontal distance of the reconstructed cascade position from the cluster centre $\rho < 100$ m are displayed. No quality criteria are applied to the datasets. Note that most of the reconstructed events from the experimental data are downgoing ($\theta > 90°$), where the background atmospheric muons (μ_{atm}) dominates. On the contrary, neutrino cascade events due to atmospheric and astrophysical fluxes [6] (ν^{atm}, ν^{astro}, respectively) are several orders lower than the experimental data. In Table 1, the event rates (per year per cluster) of the reconstructed cascades with $\rho < 100$ m for the experimental data and MC μ_{atm} are given. The data/MC ratio is defined as a ratio between number of the reconstructed events for data and MC μ_{atm}. The data-MC comparison shows that majority of the reconstructed cascades from the experimental data comes from MC μ_{atm}. For that reason, successful analysis for the selection of neutrino cascades is required. Table 2 displays the expected ν_μ^{atm}, ν_e^{atm}, and ν^{astro} cascade event rates for the configuration of one fully active cluster. In the next sections performance of several developed and optimized methods is presented. Their effect has been studied for MC μ_{atm}, ν^{atm} and ν^{astro} datasets. In general, these techniques are based on the compact topology of the neutrino cascade events or finding causally connected hits from the muon track in the event.

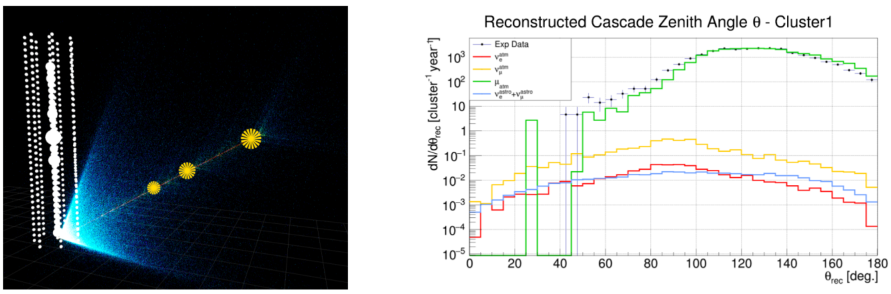

Figure 2: Left: Illustration of the background cascades (yellow blobs) produced along one muon track (blue track). Right: Distribution of the reconstructed zenith angle for cascade-like events. $\theta > 90°$ corresponds to downgoing events and $\theta < 90°$ to upgoing. The experimental data from season 2019, Cluster1 are displayed by black dots. Green histogram shows the MC simulation of μ_{atm}. ν_e^{atm} and ν_μ^{atm} are shown in red and yellow, respectively. Astrophysical ν_e^{astro} and ν_μ^{astro} neutrinos are combined in one histogram displayed in blue.

Table 1: Cascade event rates for experimental data and MC background cascades from μ_{atm}

$\rho < 100$ m [yr^{-1} cluster^{-1}]	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5
Experimental Data	23 680	23 535	31 239	34 783	32 544
MC μ_{atm} Expectation	23 539	21 740	27 491	26 328	23 784
Data/MC ratio	1.01	1.08	1.14	1.32	1.37

Table 2: Expected event rates for MC neutrino cascades

Neutrino Cascades	$\rho < 100$ m [yr^{-1} cluster^{-1}]
ν_μ^{atm}	6.34
ν_e^{atm}	0.64
$\nu_e^{\text{astro}} + \nu_\mu^{\text{astro}}$	0.8

nTrackHits Filter

The nTrackHits method discriminates between background and signal cascades exploiting the presence of muon track in the atmospheric muon bundle and its absence in neutrino cascade events. This method is designed to find as many track hits as possible. Hits are tagged as track hits based on the following hit time condition:

$$t_1 < t_i - T_i^{\text{track}} < t_2, \tag{1}$$

where t_i is the detected hit time on the OM, $t_1 = -100$ ns, $t_2 = 25$ ns, and T_i is the expected time of the detected hit on the OM from the muon track, calculated according to the equation:

$$T_i^{\text{track}} = t_{\text{recoCascade}} + (\text{sLong} - \text{lLong}) \cdot \frac{1}{\text{c}} + \sqrt{\text{sPerp}^2 + \text{lLong}^2} \cdot \frac{1}{\text{c}_\text{w}}, \tag{2}$$

where $t_{\text{recoCascade}}$ is the reconstructed cascade time, c_w is the speed of light in water, and c is the speed of light in vacuum. Other components used in Eq. 2 are depicted in Fig. 3 (left). Fig. 3 (right) shows the detected hit time dependence t_i as a function of the OM z coordinate for the MC μ_{atm} event. The displayed hits come from the muon track, random noise, and background cascades. Yellow line indicates the expected time for track hits determined by the Eq. 2. The green lines delimit the expected time window for hits originating from the cascade. Especially, the hits that have a time residual t_{res}: $-20 < t_{res}/ns < -20$ with respect to the reconstructed cascade. Note that, both cascade and track hits follow the theoretically expected times. In order to increase the efficiency of the nTrackHits method, number of track hits (denoted as *nTrackHits*) is determined in each track direction by scanning over zenith and azimuth angle in a pre-defined cone with an apex angle $\approx 40°$ around the reconstructed cascade direction (see Fig. 4 (left)). Reconstructed position and time of the cascade fix the track in a single point.

The distribution of *nTrackHits* for the MC μ_{atm} background cascades (red), neutrino-induced cascades (together $\nu_{e,\mu}^{\text{atm}}$ and $\nu_{e,\mu}^{\text{astro}}$ in blue), and experimental data (black points) is shown in Fig. 4 (right). As expected, the largest number of *nTrackHits* is obtained for the background cascades produced along the muon track. In case of neutrino cascades, non-zero values of *nTrackHits* are due to falsely identified noise or cascade hits. Majority of the experimental data are in a good agreement with the MC μ_{atm} cascades.

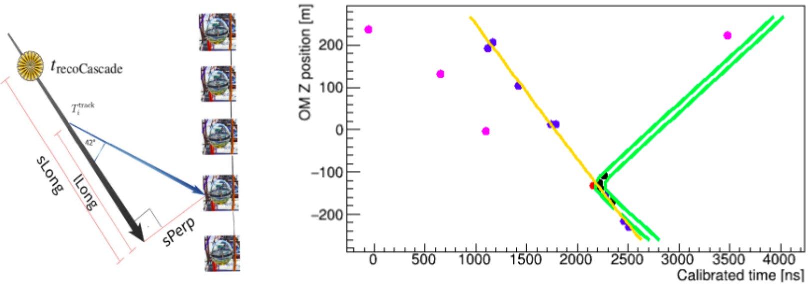

Figure 3: Left: Scheme of the geometry used for the calculation of predicted time of the detected photon originating from the track T_i^{track}, defined in Eq. 2. Right: The dependence of the detected hit time on the OM's z coordinate along one string for MC simulated μ_{atm} event. $t_{\text{recoCascade}}$ is shown by red dot, track hits by blue, cascade hits are shown in black and noise hits by pink. Yellow line displays the expected time T_i^{track} for track hits. Green band corresponds to the expected time interval, when the cascade photons should arrive at the OMs.

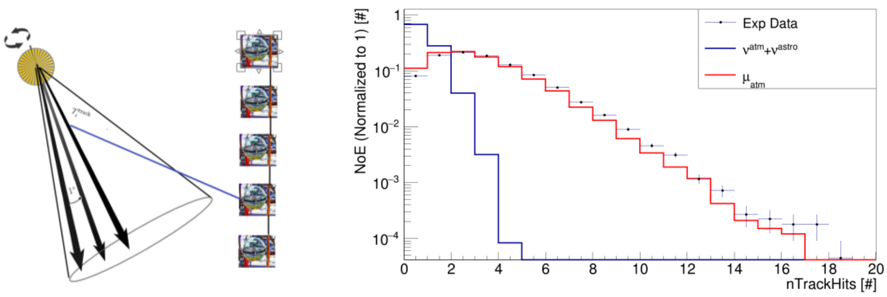

Figure 4: Left: Illustration of the searching for the muon track direction in which the highest nTrackHits value is obtained. Right: Distribution of found nTrackHits for signal cascades ($\nu_{e,\mu}^{\text{atm}}$ and $\nu_{e,\mu}^{\text{atro}}$) and background cascades (μ_{atm}) shown in blue and red, respectively. Experimental data are shown by black points.

QEarly and BranchRatio Filter

The next discrimination method, called QEarly was inspired by the ANTARES collaboration [7], however, some variations were introduced. QEarly is supposed to compare the total charge of track hits $Q_{trackHits}$ (identified by the nTrackHits method) to the summed charge of hits, which fulfill the cascade hit selection criteria - $QcascadeHits$ (green band in Fig. 3 (right)). It is defined by the following equation: $QEarly = \log_{10}\left(\frac{(\mathrm{Q_{trackHits}}+\mathrm{a})}{\mathrm{Q_{cascadeHits}}}\right)$, where $a = 1$ is an additional constant, which prevents the logarithm from becoming infinite in case of *nTrackHits* equals zero. Fig. 5 (left) shows the performance of the QEarly separation variable. Background cascades with the muon track have greater values of QEarly compared to the signal cascades.

The main purpose of separation variable, called BranchRatio filter, is to suppress downgoing cascades from μ_{atm} events. It compares a number of hit OMs that are located above the reconstructed z coordinate of the cascade $\mathrm{nOMs_{upper}}$ and the number of the OMs below it $\mathrm{nOMs_{lower}}$, defined as: $BranchRatio = \frac{\mathrm{nOMs_{upper}}}{\mathrm{nOMs_{lower}}}$. As can be seen in Fig. 5 (right) the BranchRatio reaches lower values for background cascades, since they come from the downgoing muons and the OMs that are located below the cascade are more likely to be hit. Neutrino cascades obtain higher values of the

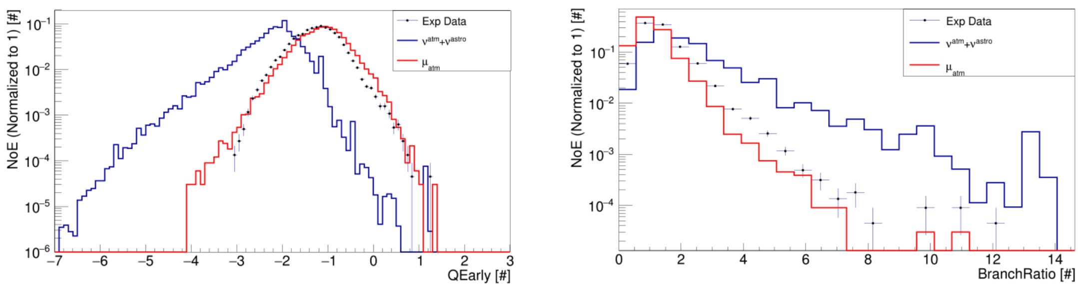

Figure 5: Left: The distributon of QEarly variable for the neutrino cascades (ν^{atm} and ν_{atro}) and background cascades (μ_{atm}) shown in blue and red, respectively. Experimental data are displayed by black points. Right: The performance of a BranchRatio discriminating variable.

BranchRatio, because in this detection channel upgoing events are also present.

3. Boosted Decision Trees and Results

In order to take the most of the discrimination potential of the established variables and to reduce a dimensionality of separation parameters, a multivariate event classifier Boosted Decision Trees was applied (from ROOT TMVA package [4]). The BDTs were trained using the 5 variables: nTrackHits, QEarly, BranchRatio, χ^2 - value of the Chi-Square after final cascade position reconstruction, θ - reconstructed zenith angle. In order to find a potential neutrino cascade candidate, only events with reconstructed cascade position inside the instrumented cluster volume ($\rho < 60$ m) and reconstructed as upgoing ($\theta < 90°$) were used in this multivariate analysis. As an input for the BDTs, upgoing neutrino cascades from MC $\nu_{e,\mu}^{\text{atm}}$ and $\nu_{e,\mu}^{\text{astro}}$ were used as a signal sample and misreconstructed upgoing background cascades from MC μ_{atm} as a background sample. Approximately 80 % of the MC datasets were used for the BDT training and the remaining events were used for testing. The distribution of the BDT output value is shown in Fig. 6 (left). Blue dots correspond to the training events of signal cascades and red dots to the background. Solid histograms with filled areas correspond to the testing events. After training and testing, TMVA evaluates the relative importance of the individual discriminating variables. The most important input to the BDTs is QEarly. The variable nTrackHits also exhibits high separation power.

Subsequently, a cut on the BDT response value was optimized. The performance of the BDT classifier can be evaluated according to Significance $S/\sqrt{S+B}$ for a certain number of signal and background events. S and B represent the numbers of signal and background events after applying a specified cut on the BDT value. The maximal value of the Significance is considered to evaluate the BDT cut value most precisely. However, one should keep in mind the signal efficiency and background rejection factor, which corresponds to the optimal BDT cut value. For that purpose, the event rates of neutrino induced cascades are scaled down according to realistic neutrino fluxes (atmospheric and astrophysical) in order to estimate the cut on the BDT value. Fig. 6 (right) shows distribution of the BDT output value for signal cascades, background and experimental data for Cluster1, season 2019. The event rate is calculated per one year. All together, according to maximal significance we applied a cut of 0.48 on the BDT value. After applying this cut, signal efficiency of

$\approx$ 49% is reached and the background is reduced to less than or of the order of 1 event per cluster per year. However, we can observe that the statistics in case of the MC background events is lower and thus the fluctuations can cause an instability in the Boosted Decision Trees.

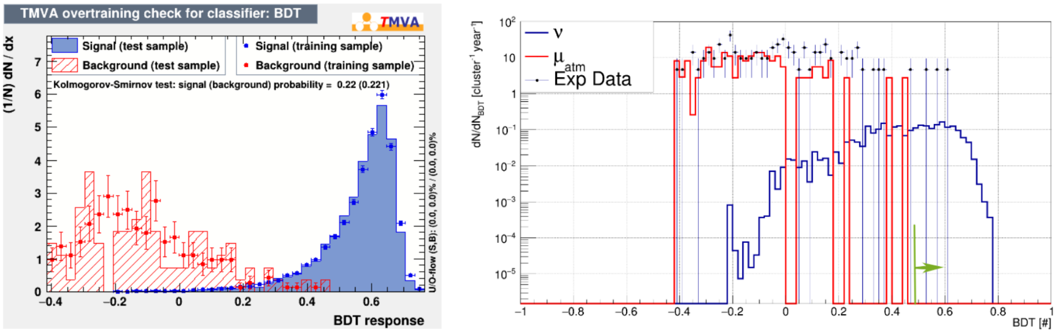

Figure 6: Left: The BDT response value distribution for MC signal (blue) and background (red) cascades. Right: The distribution of the BDT output value for MC signal (blue) and background (red) cascades and experimental data from Cluster1. Green line indicates cut applied on the BDT value.

To find some interesting high energy events in the experimental data we applied a cut on the reconstructed energy (E_{reco} > 50 TeV) in addition to the cut on the BDT value. As Table 3 shows, only one event satisfied these conditions. This event was reconstructed in the Cluster1 as upgoing (θ = 70.9 deg) with an energy E_{reco} = 83.3 TeV. Its BDT response value was found to be of 0.6.

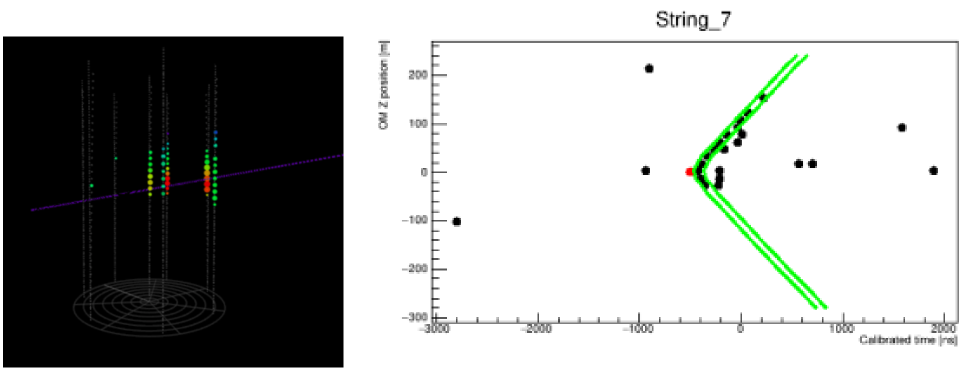

Figure 7: Left: Side view of the neutrino cascade candidate (see Table 3). Right: Dependence of the detected hit times (horizontal axis) on the OM z coordinate (vertical axis) for string 7 in the Cluster1 for the neutrino cascade candidate.

Table 3: Reconstructed parameters of the most energetic event found in data 2019 for Cluster1: Cluster, Energy, Zenith angle, Azimuth angle, Horizontal distance, Likelihood, Total charge, Total number of hits, Number of hits used in the reconstruction, and Number of track hits.

Cl	E_{rec} [TeV]	θ [°]	ϕ [°]	ρ [m]	L	Q [p.e.]	nHits	nRecoHits	nTrackHits
1	83.3	70.9	4.96	47.65	1.01	1665.01	106	44	1

4. Conclusion

We have developed and optimized several selection methods for the neutrino cascades. They were optimized with Monte Carlo simulations for season 2019. Subsequently, experimental data from season 2019 were processed and analyzed. The Boosted Decision Trees that provide their result in one discriminating output value have been used. The BDTs were trained and tested with the contained upgoing signal and background cascades. In search for the neutrino cascade candidate with higher energy in experimental data we applied cuts on the BDT output value and the reconstructed energy. One contained upgoing event was reconstructed with 83.3 TeV energy.

5. Acknowledgements

This work is supported in the framework of the State project "Science" by the Ministry of Science and Higher Education of the Russian Federation under the contract 075-15-2020-778. We acknowledge support by the VEGA Grant Agency of the Slovak Republic under Contract No. 1/0607/20 and by the Ministry of Education, Youth and Sports of the Czech Republic under the INAFYM Grant No. CZ.02.1.01/0.0/0.0/16_019/0000766. We also acknowledge the technical support of JINR staff for the computing facilities (JINR cloud).

References

[1] Baikal-GVD Collaboration: A.D. Avrorin *et al.*, Status of the Baikal-GVD Neutrino Telescope, EPJ Web of Conf. **207** (2019) 01003, doi:10.1051/epjconf/201920701003

[2] Baikal-GVD Collaboration: A.D. Avrorin et al., Search for cascade events with BaikalGVD, PoS-ICRC2019-0873, arXiv:1908.05430, doi:10.22323/1.358.0873

[3] IceCube Collaboration: M. G. Aartsen et al., Atmospheric and astrophysical neutrinos above 1 TeV interacting in IceCube. Physical Review D 91, 022001 (2015), doi: 10.1103/PhysRevD.91.022001

[4] A. Hoecker et al., TMVA 4 - Toolkit for Multivariate Data Analysis with ROOT, Users Guide (2017). arXiv:physics/0703039

[5] O. Tange, GNU Parallel - The Command-Line Power Tool, The USENIX Magazine (2011), doi: 10.5281/zenodo.16303

[6] IceCube Collaboration, M. G. Aartsen et al., A Combined Fit of the Diffuse Neutrino Spectrum using IceCube Muon Tracks and Cascades, PoS(ICRC2021)1129, arXiv:2107.10003v1 [astro-ph.HE], doi:10.22323/1.395.1129

[7] ANTARES Collaboration: A. Albert et al., An algorithm for the reconstruction of neutrino-induced showers in the ANTARES neutrino telescope, arXiv:1708.03649 [astro-ph.IM], doi:10.3847/1538-3881/aa9709