

Summer Student Project

for NOvA collaboration, Fermilab

Looking for charm production in NOvA

Author: Roman Lavička¹

Supervisor: Keith Matera²

¹FNSPE CTU in Prague, ²NOvA Fermilab

1 Introduction

One of the most abundant particles in the universe are neutrinos. The neutrinos are very important in many processes in the nature. The beta decay is a good example for all of them since it is essential for a thermonuclear fusion, an interaction which powers stars. There would not be a life in a form we know today on Earth without the energy provided by the Sun. Therefore it seems to be a good idea to learn more about the neutrinos and processes connected with them. Unfortunately, due to its negligible mass and ability to interact weakly only it is difficult to create, catch or measure them. Scientists all around the world throughout last decades grouped together to build several impressive facilities which focus on a neutrino research. One of them is located near Chicago, Illinois, at Fermi National Accelerator Laboratory (Fermilab).

There are several neutrino experiments at Fermilab. The NuMI Off-axis ν_μ Appearance (NOvA) is an experiment which is looking for neutrino oscillations. Its primary goals are to look for an oscillation of muon neutrino to electron neutrino (appearance) and an oscillation of muon neutrino to muon neutrino (disappearance). Measuring these probabilities one can put new limits on some physics quantities, such as ordering of neutrino masses or symmetry between matter and antimatter. Since the NOvA has become operational in 2014, the experiment NOvA gathered a lot of physics data. Recent results can be found in [1] and [2].

During the summer 2016 we investigated a possibility of a measurement of events, where neutrino interacts with a nucleon and creates a charm quark and muon. This later results in a single or dimuon event visible in a detector. A motivation of this effort is to provide a complementary measurement of the collaboration NOMAD in an energy range from 0 to 20 GeV [3]. Next pages describe and discuss current results.

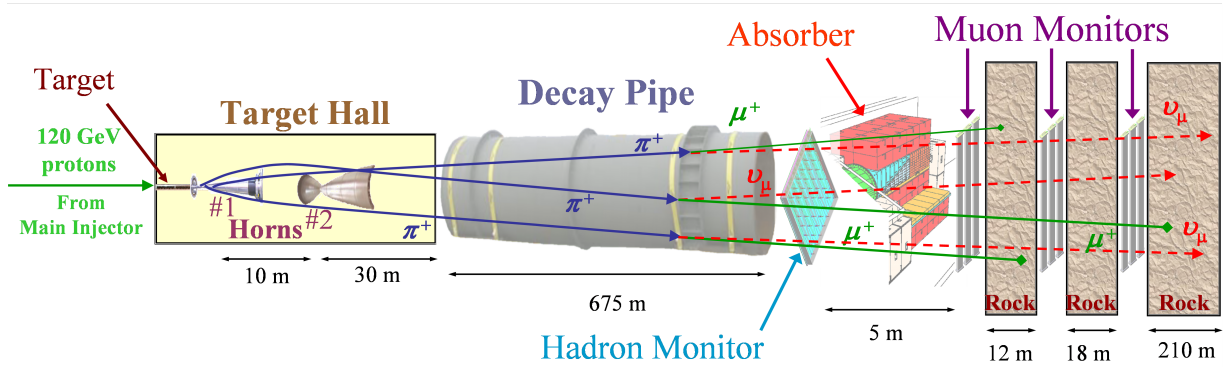


Figure 1: A scheme of the beamline in Fermilab.

2 Experiment

Because the collaboration NOvA's physics tasks are motivated with neutrino physics, it needs an intense beam of neutrinos called NuMI. A creation of this beam is sketched in Fig. 1. Everything starts with protons accelerated to an energy of 120 GeV in the Main Injector. These protons are delivered to a graphite target, where they interact and create primarily a lot of charged pions, mainly in a forward and semi-forward direction. The pions are then focused in two magnetic horns and sent to an almost 700 metres long decay pipe. When the pions travel through this pipe, they have enough time to decay into a muon and muon neutrino. Next the beam pass an hadronic absorber, which deflects all remaining protons and mesons made in the initiating collision. Than the beam propagates through a 230 metres of rock, where the rest of the muons is absorbed and only a beam of muon neutrinos with an energy ranging from 1 - 20 GeV remains.

While an axis of the NuMI beam points towards the Soudan Mine in Minnesota (hitting detectors of experiments MINOS and MINERvA), the experiment NOvA is constructed 14.6 miliradians off. This yields in a peak in an energy spectrum of neutrinos around 2 GeV and a long energetic tail, which is a fact we want to exploit in our analysis as it will be described in Sec. 3.

The experiment consists of two detectors - the near detector (ND) and the far detector (FD). As depicted in Fig. 2 the ND is located at Fermilab and the FD is 810 kilometres away in Ash River in Minnesota. In Fig. 3 comparisons of sizes of both detectors to a human are shown. Clearly, the FD is larger so it would be a better idea to use this one for our research. Unfortunately, due to a long distance a ratio of cross sections of the beam and the detector is much smaller than for the ND. Even if we catch hundreds of neutrinos in the ND per a second, we have only 1 neutrino in the FD per a minute. Therefore we used the near detector in our analysis.

The ND is a compact detector 350 feet underground with dimensions 2.9 x 4.2 x 14.3 m³ [4]. The body is assembled from modules of extruded Poly Vinyl Chloride loaded with titanium dioxide. The modules are divided into 32 cells with proportions 3.8 x 5.9

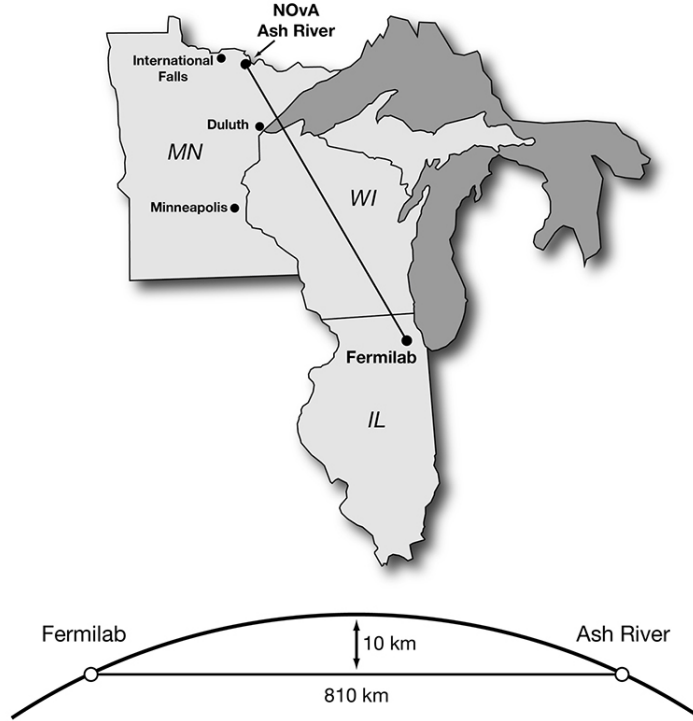


Figure 2: A scheme of the beamline in a macroscopic scale.

centimetres. Each cell is filled with liquid scintillator and read-out with a wave-length shifting fiber which has an avalanche photo-diode on both ends.

3 Theory

Today we classify neutrinos according to their flavour or mass into three types. We can have electron, muon or tauon neutrino or we have 3 different mass states of neutrinos. Each flavour state neutrino consists of an currently unknown mixture of 3 mass state neutrinos. Because of the different masses, each mass state neutrino travels with different velocity and their superposition is changing in time. Therefore the mixture is fluctuating and we observe a flavour state transforming into another flavour state. This is called the neutrino oscillation and this is a primary object of investigation of NOvA collaboration.

But the collected data can be also used for another analysis. In our case it is a usage of the near detector to study a production of charm quark. A general diagram of such a process is shown in Fig. 4. The charm quark then hadronized into a baryon (mostly lambda) or a meson (mostly D-meson). These hadrons have very short lifetime and decay almost immediately through several decays, where some of them produce high energetic

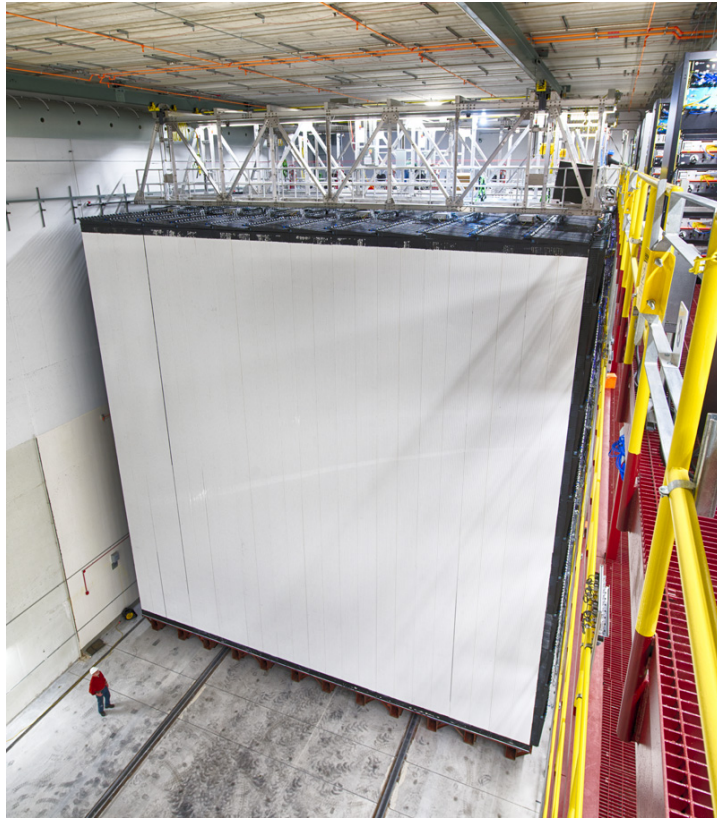
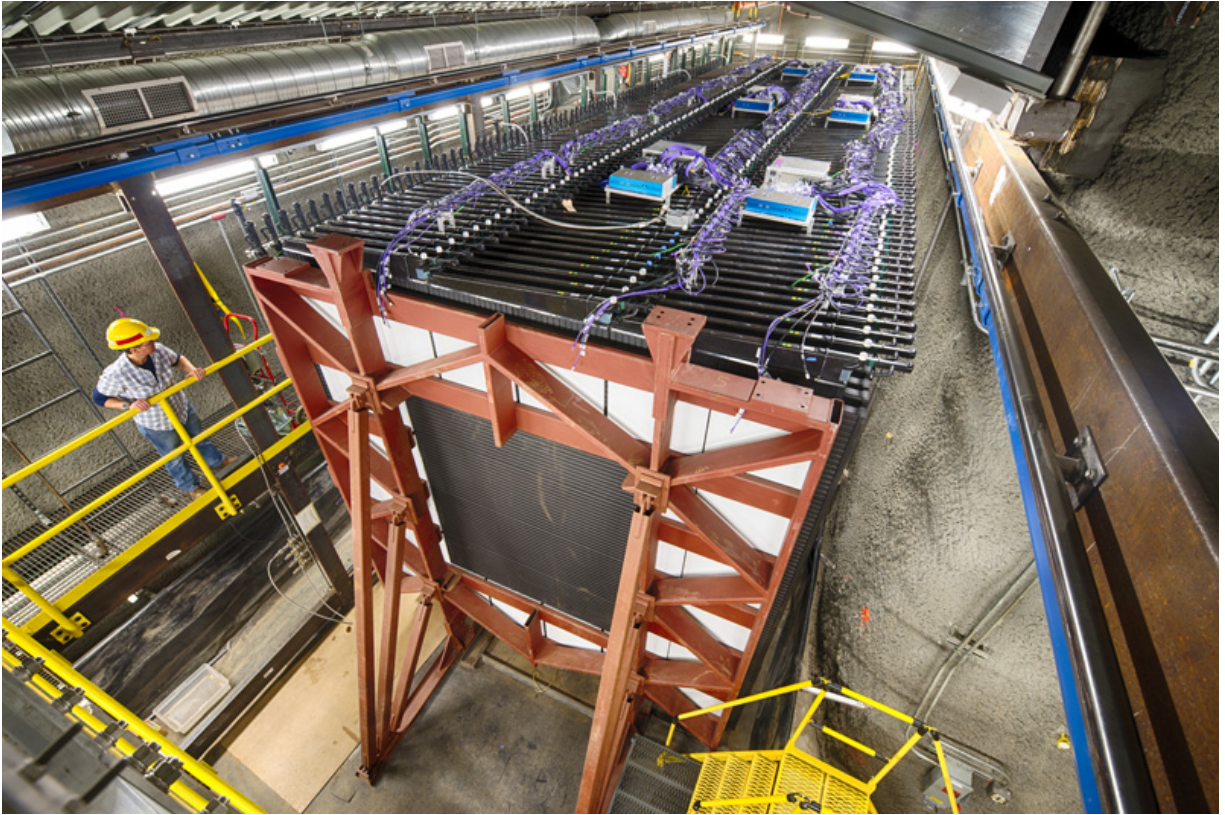


Figure 3: NOvA near detector (left) and NOvA far detector (right). Taken from [4].

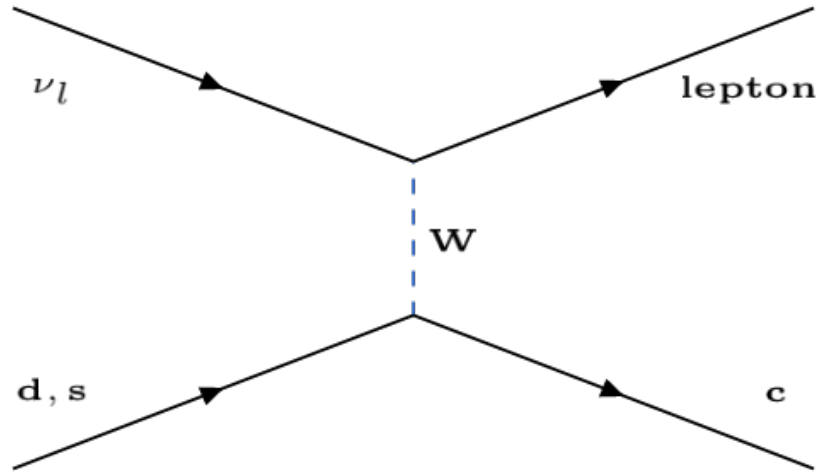


Figure 4: A diagram of a charm interaction. ν_l stands for a neutrino of a specific lepton, W is an intermediate boson and d, s and c are down, strange and charm quarks respectively.

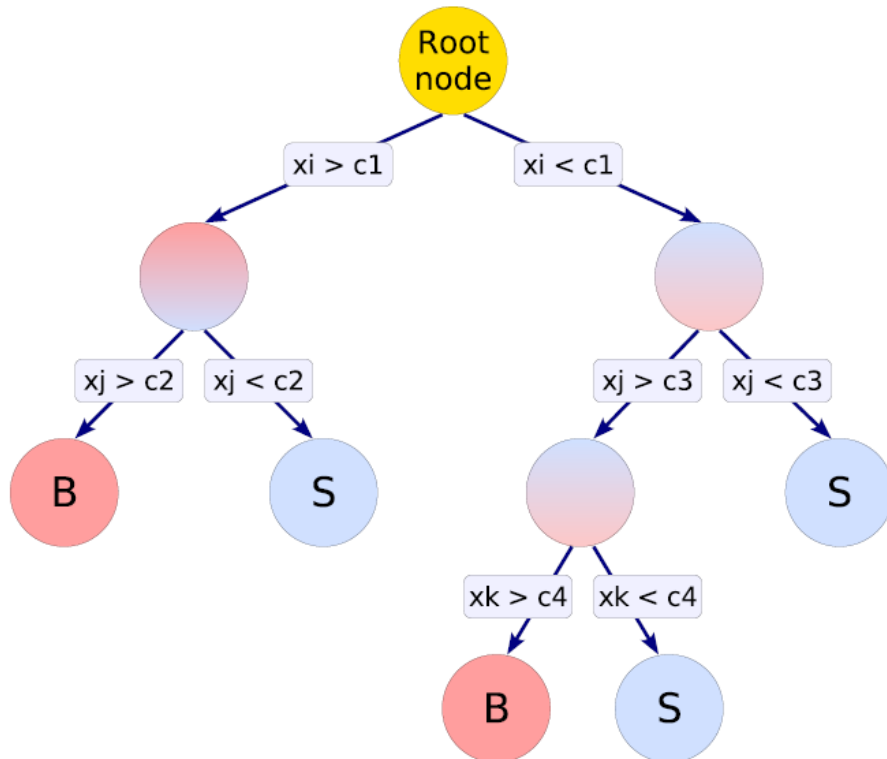


Figure 5: A scheme of a boosted decision tree. Each level represents one decision criterion. B is background, S is signal.

muons. In the end we can see one or two high energetic muons in the detector what can be a sign of the charm production.

The multivariate analysis (MVA) or it's ROOT implementation Toolkit for Multi-Variate Analysis (TMVA)[6] is a perfect instrument to distinguish between charm events (signal) and the rest (background). This tool includes several packages to solve the problem. In this analysis we used so-called Boosted Decision Tree (BDT). This method as an input needs a set of variables which are further analysed. The method is depicted in Fig. 5. It loops over all possible cuts on each variable. As shown, it takes an event and asks for the first variable, whether it passed a cut or not. According to it, it is sent to one branch or another. In the next node, it is asked again on cut on the next variable (the selection criterion varies from branch to branch). This continues over all variables and in the end the decision whether it is a signal or background is made. This tree needs to be trained using data from Monte Carlo simulations. When the tree is trained, it returns a list of weights for all variables and a cut output value, which represents the best MVA value for the set of variables. The MVA value reflects the efficiency of the acceptance/rejection of data scaled to a general range.

4 Implementation

During 9 weeks of work on this project we had to get acquainted with NOvA computing procedures, to learn with CAFAna tool and a structure of storage of data on a reconstruction level, displaying events, to use GRID for time-eating computations and to apply a multivariate analysis to find the best selections for our data. All this needed to develop several macros, which are described in this section and are saved in a development folder at NOvA hard drives located in `/nova/app/users/rlavicka/charmRecoAnalysis-S16-03-04/Charmanalysis/`.

4.1 Event displays

Looking at the event displays is one of the less complicated tasks in our analysis from the programming point of view. First, we had to get a specific file containing our wanted event. The macro `find-events.pl` in `EVD` folder serves this purpose. One only need to write correct definition name of the data snapshot and a name of a `.txt` file containing information about event inside the macro. The `.txt` file should in each line (representing 1 wanted event) have the numbers related to the event in this order: run, subrun, event, slice, cycle and time of the file creation. An example is in a file `dimuon.txt`. Running this macro will download the specific file with the specific event from the tape (SAM) into current folder. One can use a visual framework to look at the event by writing

```
nova -c evd.fcl -s <samplename.root>
```

into terminal. The framework is usually set via `edit→Configure Drawing→RawDrawingOptions` where one mark *by time* in a line `Color` and adjust range in a line `TimeRange`. To list

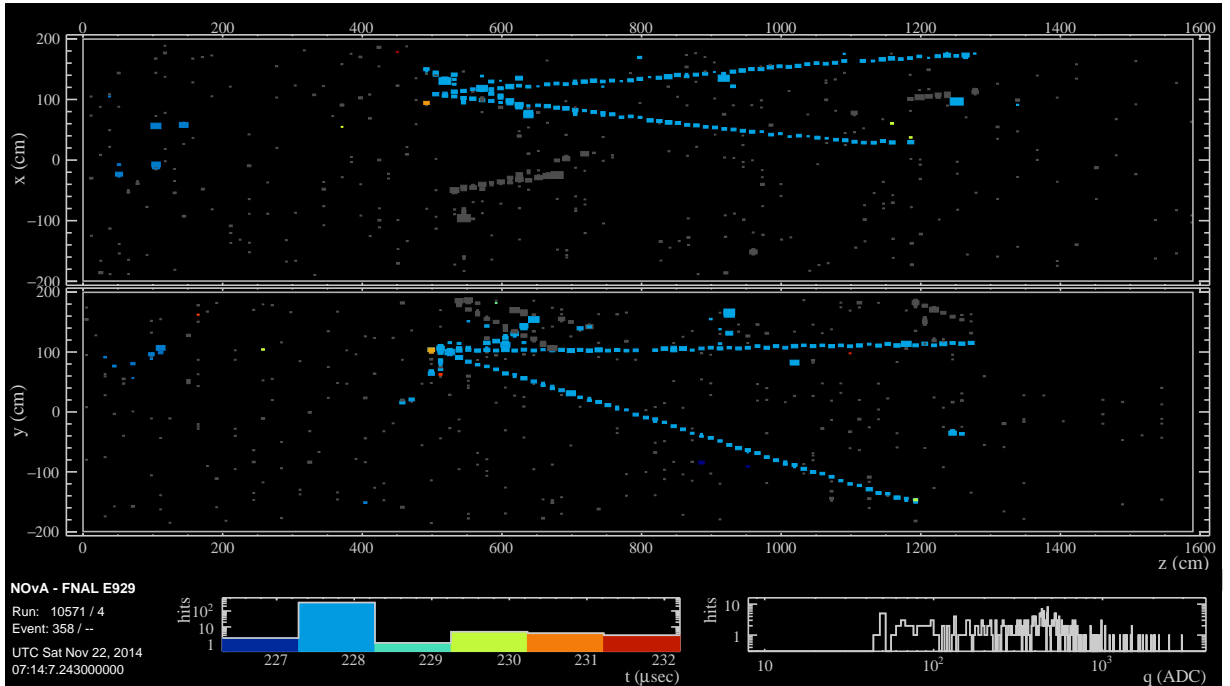


Figure 6: An example event display of an event containing 2 muons obtained from the Monte Carlo simulations.

through slices it is necessary to mark *slice* in a line *Zoom* in a bookmark *Geometry*. An example of event display with these options on is in Fig. 6. Widely useful option is to use information from the MCTruth module to display trajectories and/or interaction formula. In this case option *trajectories (short)* has to be marked in a line *Draw (Text)* in a bookmark *Simulation*.

4.2 Monte Carlo variables

To read data form a SAM file and save it in an easy-to-access format was a challenge. Two macros named *charmjob.fcl* and *CharmAnalysis_module.cc* were created and are located in a folder called *fcl*. The fickle file is the one which should be run using ART by typing

```
nova -c charmjob.fcl
```

to terminal. This command can be extended with a switch *-s <filename.txt>*, where the text file contains names of samples of the SAM snapshot which wanted to be processed. The second file is a module, which is called by the fickle file and contains a code, which does with the sample what we want to be done.

A structure of the fickle file is simple and is well commented. We only call attention on the list of input modules in a block called *physics*. Here it is crucial the names match names of modules in the SAM file. This can be easily check by eye using a predefined ART macro by calling

```
nova -c eventdump.fcl <samplename.root>
```

in terminal.

The CharmAnalysis module is a little bit more complicated. We only pinpoint important parts of the code in this paragraph. Each time the fickle file loads a data sample, the CharmAnalysis class is constructed, which sets classes (i.e. BackTracker or TFileService) ART needs for a communication. For each sample *beginJob*, *respondToOpenInputFile*, *reconfigure*, *analyze* and *endJob* functions are called respectively. The fickle file loops over all events the sample contains (usually 2000 events) in the *analyze* function. All histograms or trees needed to store our wanted variables are set in the function *beginJob* and are saved in the function *endJob*. The *reconfigure* function serves to reconnect the input modules with our analyser module and the last function is there to obtain meta data informations about the opened sample.

The most comprehensive part is the analyser module. Here we recover variables we want for our analysis from the SAM files. In the first part we set everything what is needed for the communication with the input modules. The second part is a loop over slices (sub-events). In this part we check the slice and set some switches. Then we obtain and save informations from the sample to local variables. This is done directly in the code when it is necessary or via some functions defined in the header of the class. We should mention the filling of true variables which keep the information about the generated particles (not reconstructed). They are usually stored in an object called MCTruth and each slice contains more of them. Several conditions applied on them should pick only one object per slice. After that we save informations about the presence of the charm quark in the interact, if it comes from the sea quark and the number of muons generated according to the back tracker. The first one is an important variable for us to distinguish between signal and background and the other gives an complementary view helping us to understand what is going on.

At the end of the loop we decide, whether the sub-event was completely reconstructed in the near detector and if it was signal or not. Variables are stored according to it.

One more thing is important to mention about current version of this module. The output can be stored in two different ways. The first way is to store every variable to its own branch in a ROOT object called TTree [6]. To store data in this object is necessary for the multivariate analysis, which is described in Sec. 4.4. The second approach is to fill prepared histograms with the variables. A raw histograms are saved in a *.root* file. A macro called *GetRatio.C* located in a directory *ratio* is written to load this file, set all histograms, make ratios of signal and background for most of the variables and save it in a *.pdf* format.

Because of a pre-defined form of the module, we cannot easily use both methods of data storage simultaneously and one have to switch from one type to another manually commenting and un-commenting some parts of the code. This solution is not optimal and it originated from a development of the code. The next step would be to use a TTree object only to save variables which means the macro *GetRatio.C* needs to be rewritten. Unfortunately we didn't consider this to be a crucial problem, therefore we used our limited

time on some other urgent issues.

4.3 GRID

Analysing large amount of events demands a high computational power which would take more than a reasonable amount of time for a laptop. Therefore collaboration NOvA dispose with an array of processors which users can use for their calculations. Once user sets his working environment properly, he can create a project on GRID to run his analysis running a predefined macro called *submit_nova_art.py*. This can be set with options which are listed in [8]. To simplify the act the options can be written in a configuration file which can be append to the macro with a switch *-f*. The configuration file we used is located in a folder *grid* and is named *runcharmjob.cfg*.

A user should set carefully several thing in the file, such as name of the project, where the fickle file is located or which snapshot should be taken. All options are well commented. We will only stress out that for usage of TFileService class one need to set an option *--histTier* with a name which will be appended to a name of an output file.

Each snapshot is usually divided into smaller files. These files contain some number of events. If you want to analyze more events than there are in a single file, you have to express it in the configuration file. An option *--filer_per_job* serves to tell to grid, how many files each job will look at. A number of jobs to be processed is set with *--njobs* option. The power of grid is that it can work on more jobs at the same moment (up to hundreds). An only restriction on users is to set jobs to last between 1 to 6 hours. Therefore the configuration file should contain a proper ratio of number of jobs and number of files per job.

4.4 TMVA

As mentioned before we used ROOT's toolkit for multivariate analysis to train the boosted decision tree method with our Monte Carlo data and to retrieve a selection criterion. Original and adjusted macros from the original ROOT package are located in a folder *TMVA*. There are 3 macros which should be touch.

The macro *classification.C* is the one used to train the tree. It is very well commented and before every use a user needs to take care of several lines. First, names of input and output files had to be set properly in TString objects called *fname* and *outfileName* respectively. Then all used variables has to be assign to the TMVA's object of class Factory via a function *AddVariable*. These have to match with names of branches of TTree objects in the input *.root* file. How to handle this is described in more detail in [7].

Running this macro results in several files. One is the output *.root* file containing histograms for instance with correlations among variables, cut efficiencies or over-training checks. These can be easily accessed with running a macro *TMVAGui.C* and are automatically plotted into a directory *TMVA/plot*. Other outputs are tables with weights for each variable found by BDT and are stored in *.xml* format in a folder *TMVA/weights*. These

Variable	Pre-selection
Length of the longest track	>152 cm
Length of the 2. longest track	>16 cm
Ratio of hits of the longest track	<0.88
Ratio of hits of the 2. longest track	<0.64

Table 1: Pre-selection criteria as suggested by BDT method.

are used by a macro *application.C*. This one serves to read the weights using TMVA's object of class Reader and applying them on our data, returning the MVA output value for each event. This can be compared to the cut criterion from the BDT method and accept (or reject) the event. In addition, this macro stores a histogram with numbers of accepted and rejected events into *.root* file made by TFile object called *target*.

5 Results

In this section we are going to present and discuss our results. As it emerges from the previous chapters, our output is mostly graphical. As our input we used Monte Carlo generated data which for the NOvA collaboration are known under a definition name

*prod_pidpart_R16-03-03-
prod2reco.a_nd_genie_nonswap_nogenierw_fhc_nova_v08_epoch1-3c.v1_prod2-
snapshot*

and contain around 100 millions of event of neutrino interactions with NOvA's near detector. How we choose between signal and background and what events we store is told in Sec. 4.2.

First, we were able to plot ratio of signal and background for different quantities. These we used to look for some significant differences, which would help us to distinguish between signal and background by a simple cut on a specific variable. In Fig. 7 we have plots of lengths of the longest and the second longest tracks in each event as an instance. Here we can see that charmed events have more likely longer tracks of muons. This holds with the fact that charmed events should produce more energetic muons than the other. Therefore we believe that a pre-selection cut on track length or muon energy would help.

Except the track length we identified other variables, which could be helpful. Among these belongs a ratio of hits of the track to the other hits (for both tracks). The BDT method can also suggest, which pre-selection criteria use and these are written in Tab. 1. Although we could be more strict according to our plots, we have decided to apply cuts from BDT for consistency.

From the plots we have picked a set of variables. Their correlations are in Fig. 8. As we can see, most of them are not correlated, which is good. But some (like Trk0CalE and

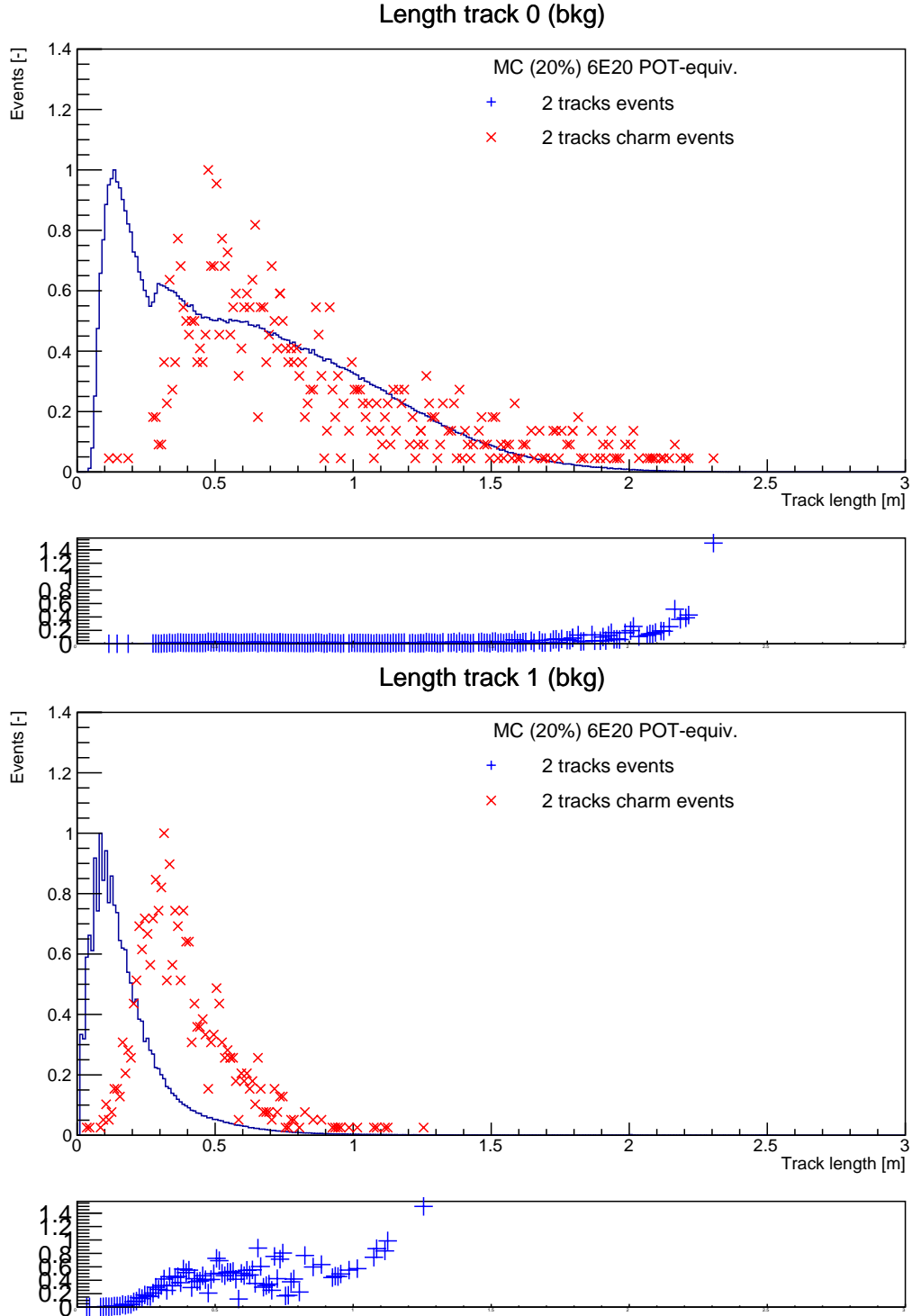


Figure 7: Plots of the lengths of the longest (top) and the second longest (bottom) tracks in an event. Each plot is divided into 2 pads, where the top one contains a histogram for a signal and a background and the bottom one their ratio.

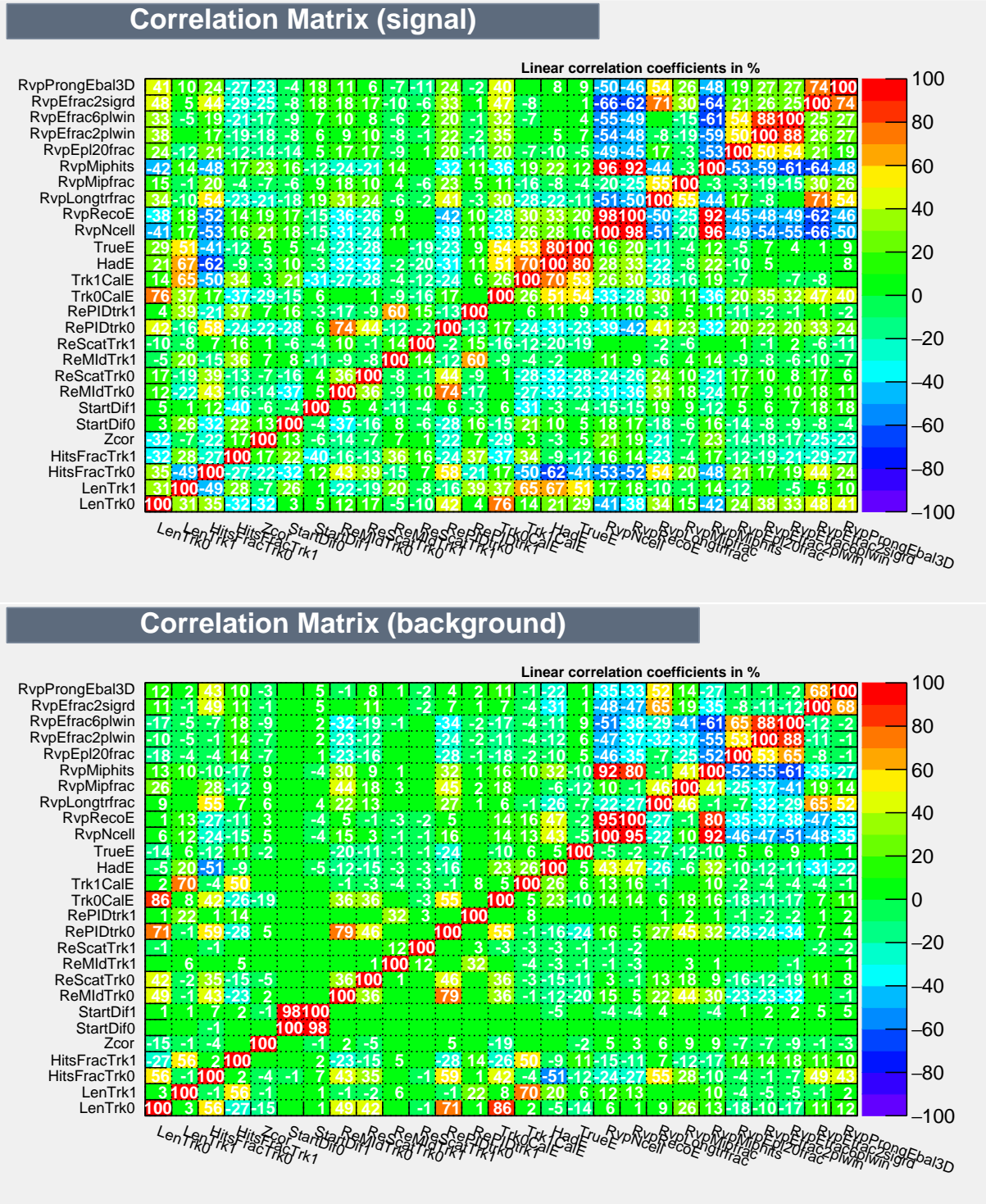


Figure 8: Correlations of variables used by BDT for signal (top) and background (bottom).

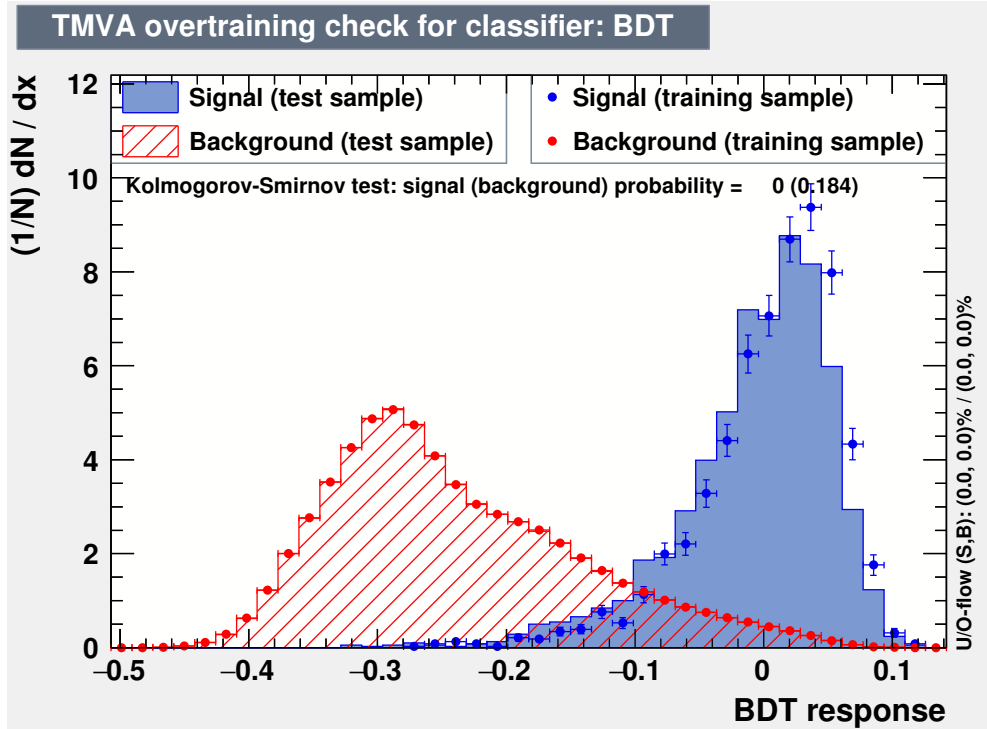
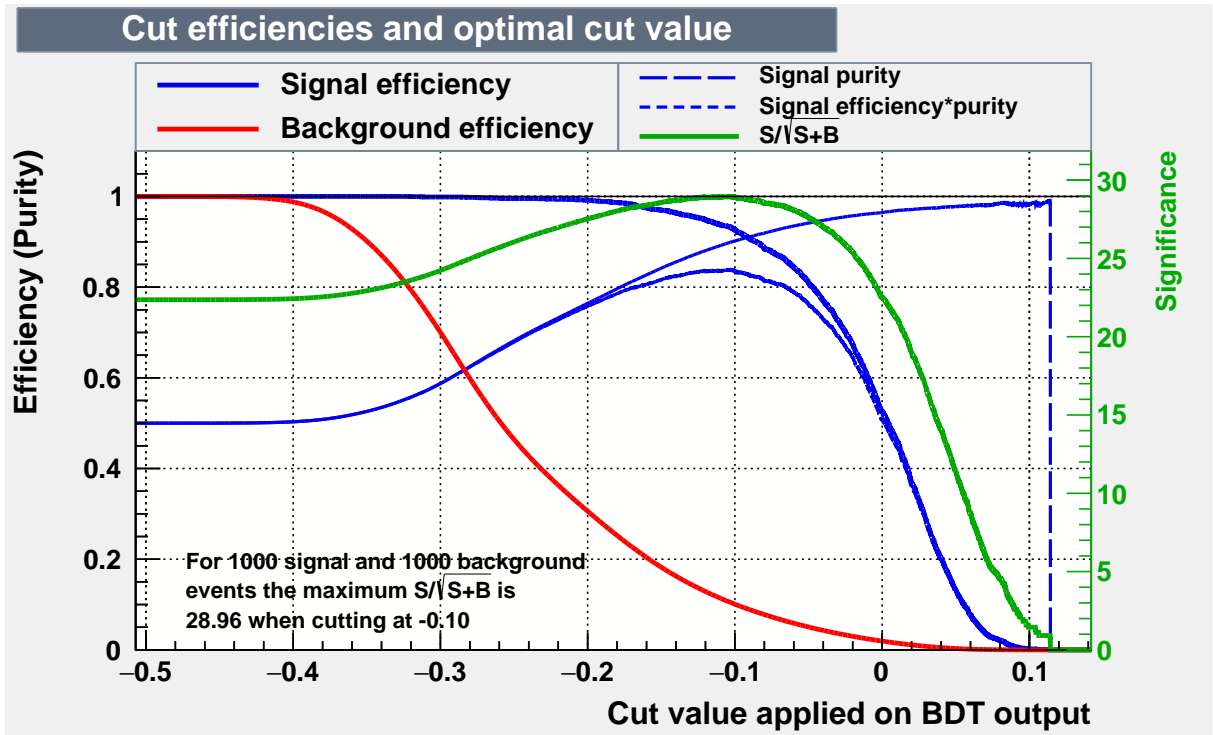


Figure 9: Results of the boosted decision tree method of MVA. The top plot shows efficiencies of the cut on background and signal events. The bottom plot is check for BDT over training.

LenTrk0) embody large correlations and had to be scratched. In addition, we scratched a variable TrueE and HadE since they represent an information directly from the Monte Carlo generator and we could not use it for real data. On the other hand we kept all variables starting with Rvp. Unfortunately we didn't have time to investigate them more, so we hoped they don't ruin the result. As it will be mention later this is a place where we should pay more attention in future.

When we had prepared trees containing our chosen variables filled with informations from signal and background events, we could start using BDT. The method is described above and results are in Fig. 9. In the top plot the blue line represents an effect of the cut on the signal events, the red line tells us an efficiency of the cut on the background events and the green line shows

$$\frac{S}{\sqrt{S+B}} \quad (1)$$

where S stands for a number of signal events and B for a number of background events. The maximum of the green line gives us a BDT output value, which we use to cut on when looking on a data. The second plot in Fig. 9 tells us, if BDT method was not over trained. It seems it finished well.

Due to a lack of time, we were able to check the efficiency of the found cut only with Monte Carlo. Applying our trained tree on signal (background) and select only those events having the BDT output value >-0.1 , we got a percentage of events passing the selection which is 70% (87%). Taking into account that we have 4 705 signal events and 10 441 570 background events we have a ratio of signal to background 0.001, which means that after our selections, we will still have only 1 wanted event between 1000 unwanted events.

This is a horrible number and means that we have to repeat a whole process of selection carefully and find any mistake we could make. Appearance of the error is highly likely since we didn't have enough time to properly focus on the BDT method. If we will conclude that we did everything right or we will not be able to find a better selection even after all corrections, we should consider to use a different variables or maybe a different decay modes. If nothing helps, we should start asking, if we are able to measure charmed events with NOvA near detector.

References

- [1] P. Adamson *et al.* [NOvA Collaboration], “First measurement of electron neutrino appearance in NOvA,” *Phys. Rev. Lett.* **116** (2016) no.15, 151806 doi:10.1103/PhysRevLett.116.151806 [arXiv:1601.05022 [hep-ex]].
- [2] P. Adamson *et al.* [NOvA Collaboration], “First measurement of muon-neutrino disappearance in NOvA,” *Phys. Rev. D* **93** (2016) no.5, 051104 doi:10.1103/PhysRevD.93.051104

- [3] F. Vannucci, “The NOMAD Experiment at CERN,” *Adv. High Energy Phys.* **2014** (2014) 129694. doi:10.1155/2014/129694
- [4] D. S. Ayres *et al.* [NOvA Collaboration], “The NOvA Technical Design Report,” FERMILAB-DESIGN-2007-01.
- [5] K. Matera *et al.*, “Collaboration NOvA internal notes.”
- [6] I. Antcheva *et al.*, “ROOT: A C++ framework for petabyte data storage, statistical analysis and visualization,” *Comput. Phys. Commun.* **180** (2009) 2499 [arXiv:1508.07749 [physics.data-an]].
- [7] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, “TMVA: Toolkit for Multivariate Data Analysis,” *PoS A CAT 040* (2007) [physics/0703039].
- [8] NOvA collaboration, “The NOvA Offline Workbook,” [https://cdcvs.fnal.gov/redmine/projects/novaart/wiki/Submitting_NOvA_ART_Jobs cited on 7.10.2016].