Fermilab

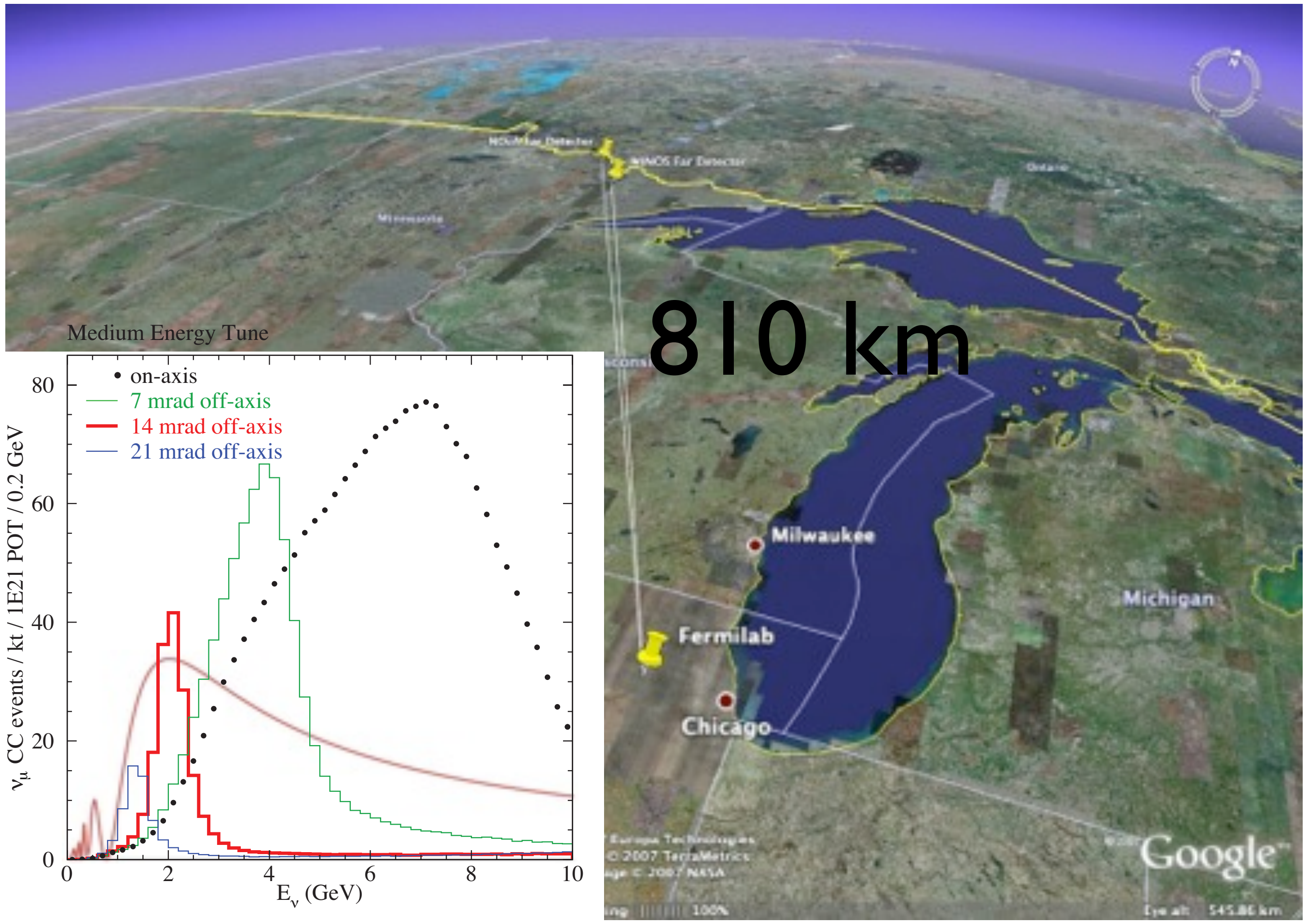
ν_e identification in the NOvA Near Detector events

Ciro Riccio

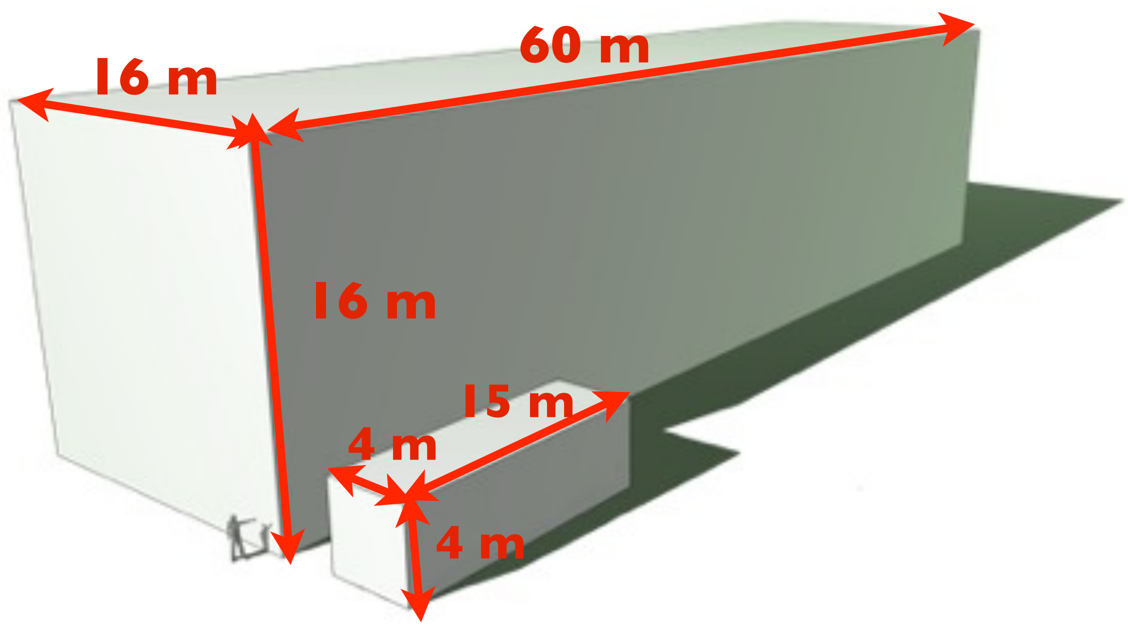
Supervisors: Xuebing Bu and Pat Lukens

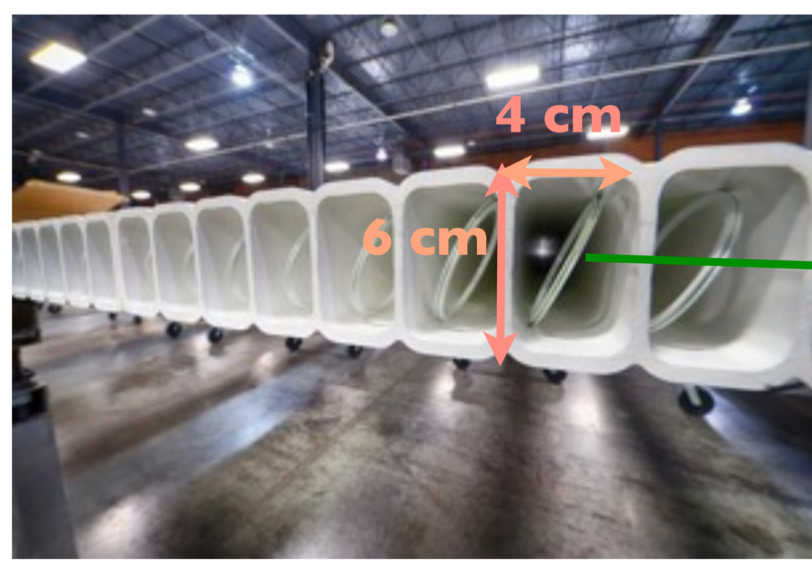
September 25th, 2014

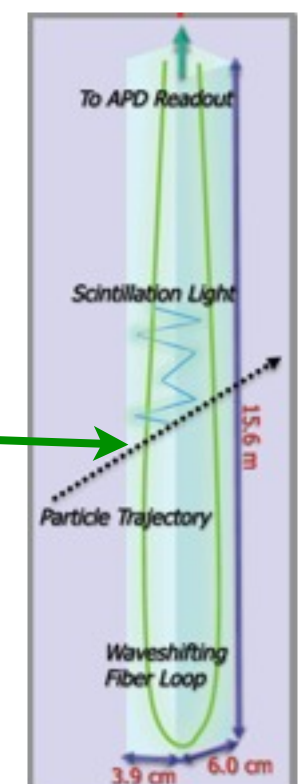
The NOvA experiment



- NOvA NuMI Off-Axis νe Appearance is optimized for the detection of $\nu_\mu \rightarrow \nu_e$ and $\overline{\nu}_\mu \rightarrow \overline{\nu}_e$ oscillations
- NOvA includes:
  - Main Injector now @ 360 kW used to produce the beam
  - A 14 kt "totally active" tracking liquid scintillator calorimeter sited 14.6 mrad off the NuMI beam axis at a distance of 810 km (Far Detector, FD)
  - A 300 ton Near Detector (ND) identical to the far detector sited 14.6 mrad off the NuMI beam axis at a distance of 1 km and 105 m underground. It is used to study the background compositions and contributions for oscillation analysis

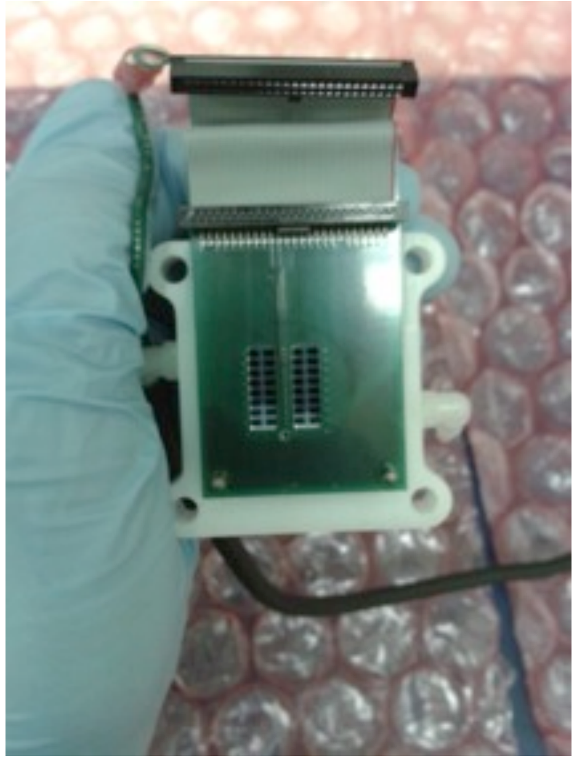






APDs Quality Assurance Test

Visual test



Pressure and flow test

Electrical test

ν_e identification in the ND

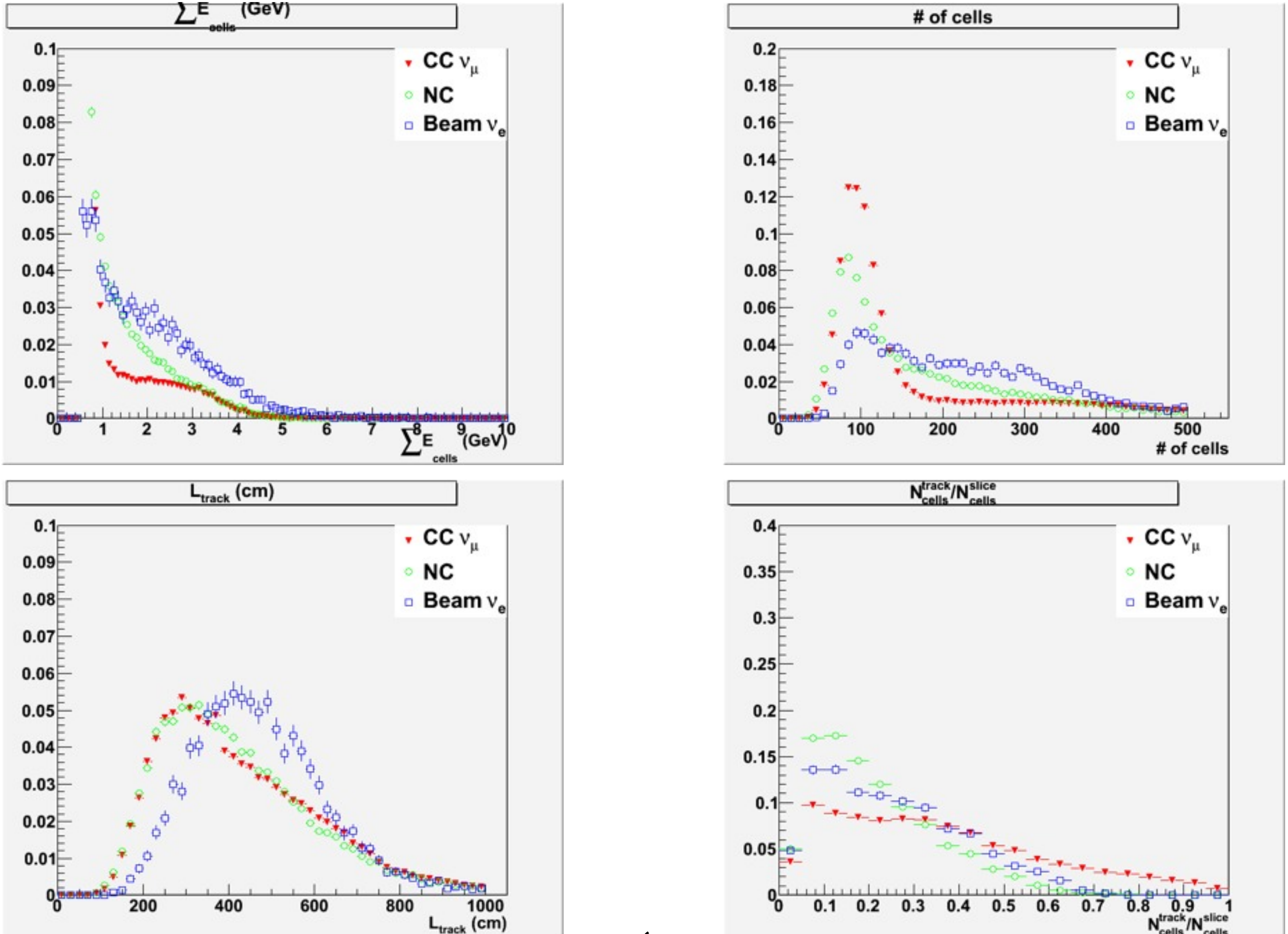
In order to identify ν_e events I used Boosted Decision Trees (BDT):

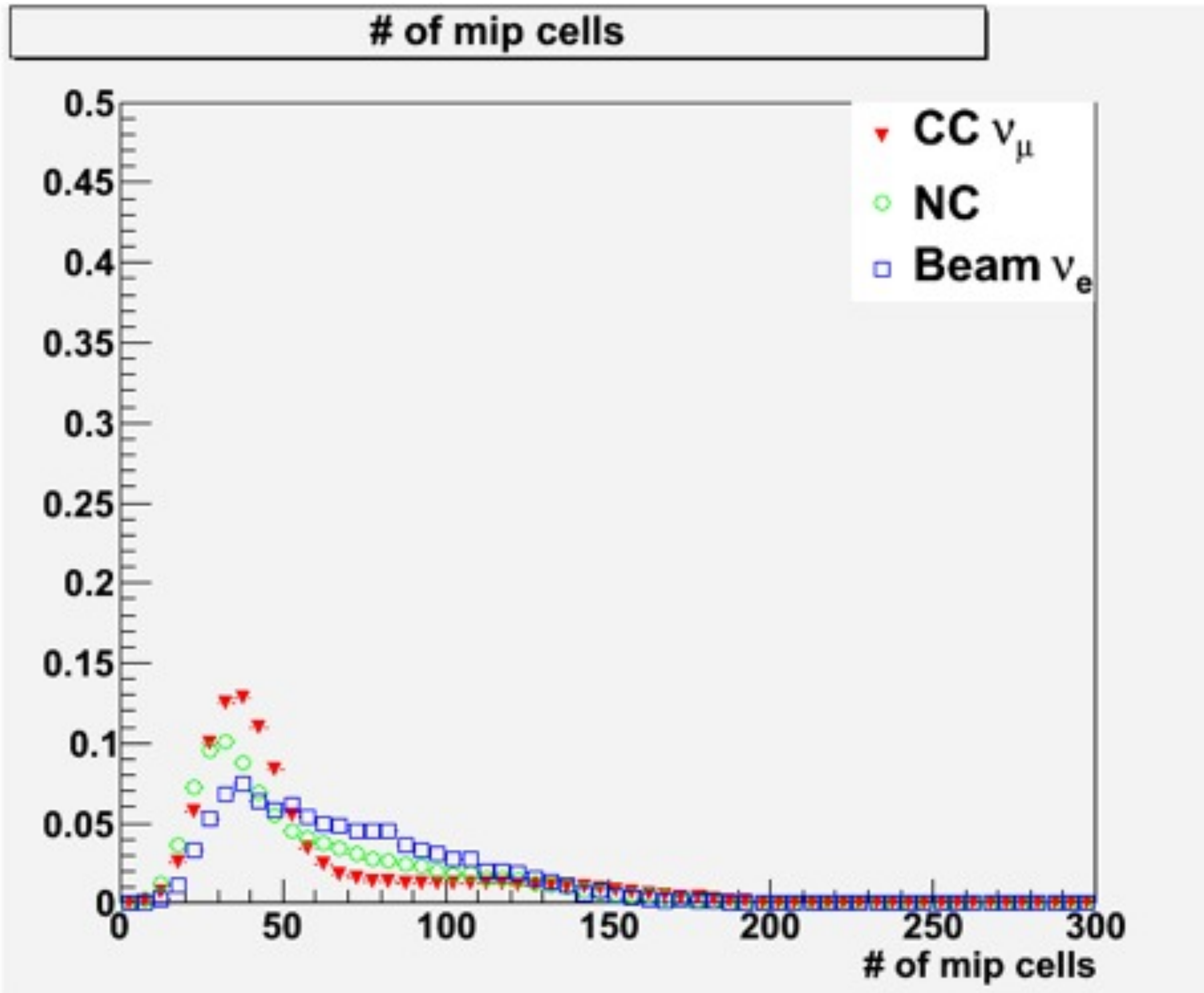
- BDT is a classifier implemented in TMVA;
- The BDT was trained and tested using well known signal and background samples;
- The BDT was applied to 1779 MC files for a total of 8.9 x 10^{19} POT to identify ν_e events in ND

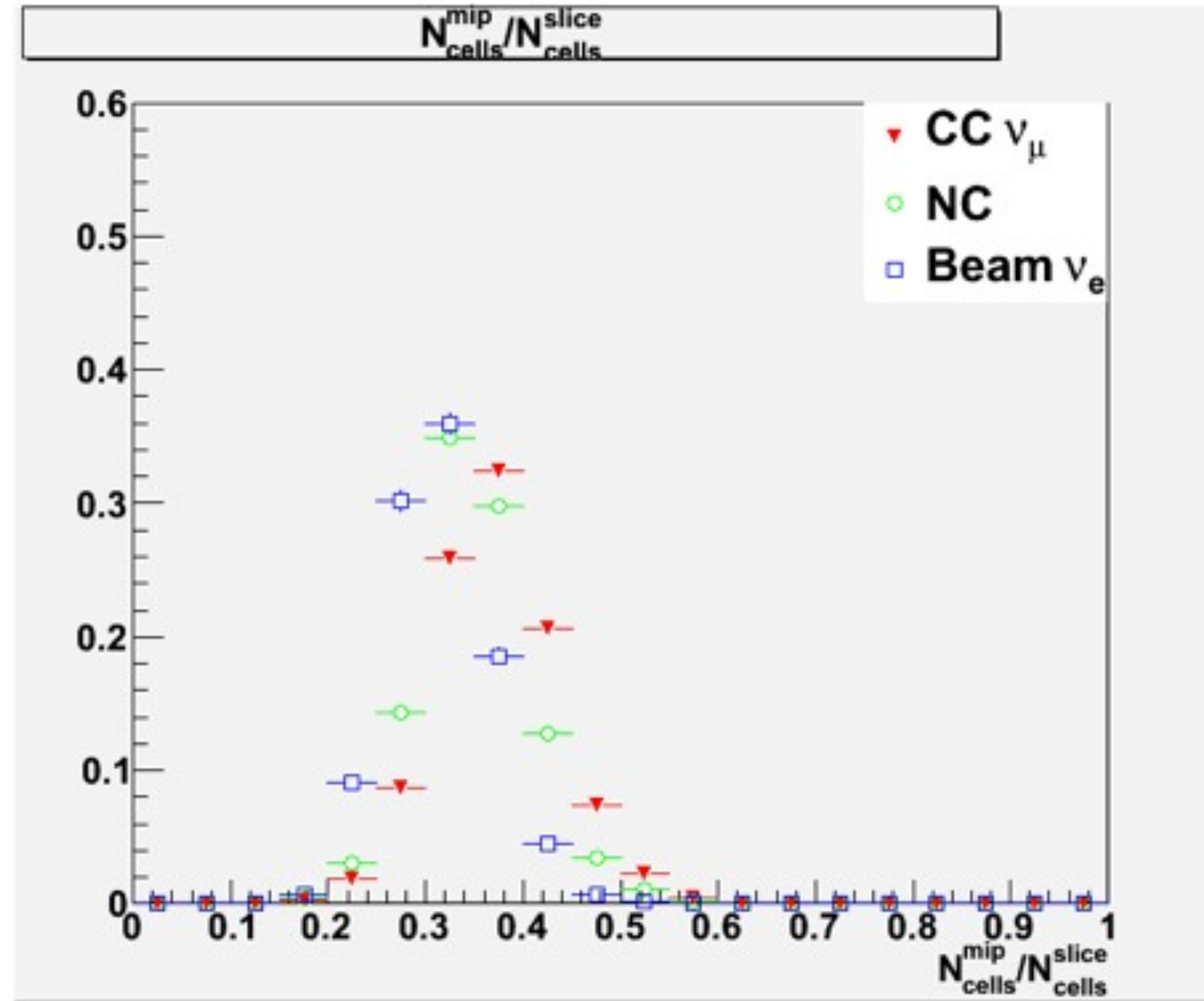
List of variables used to train and test BDT and for PID

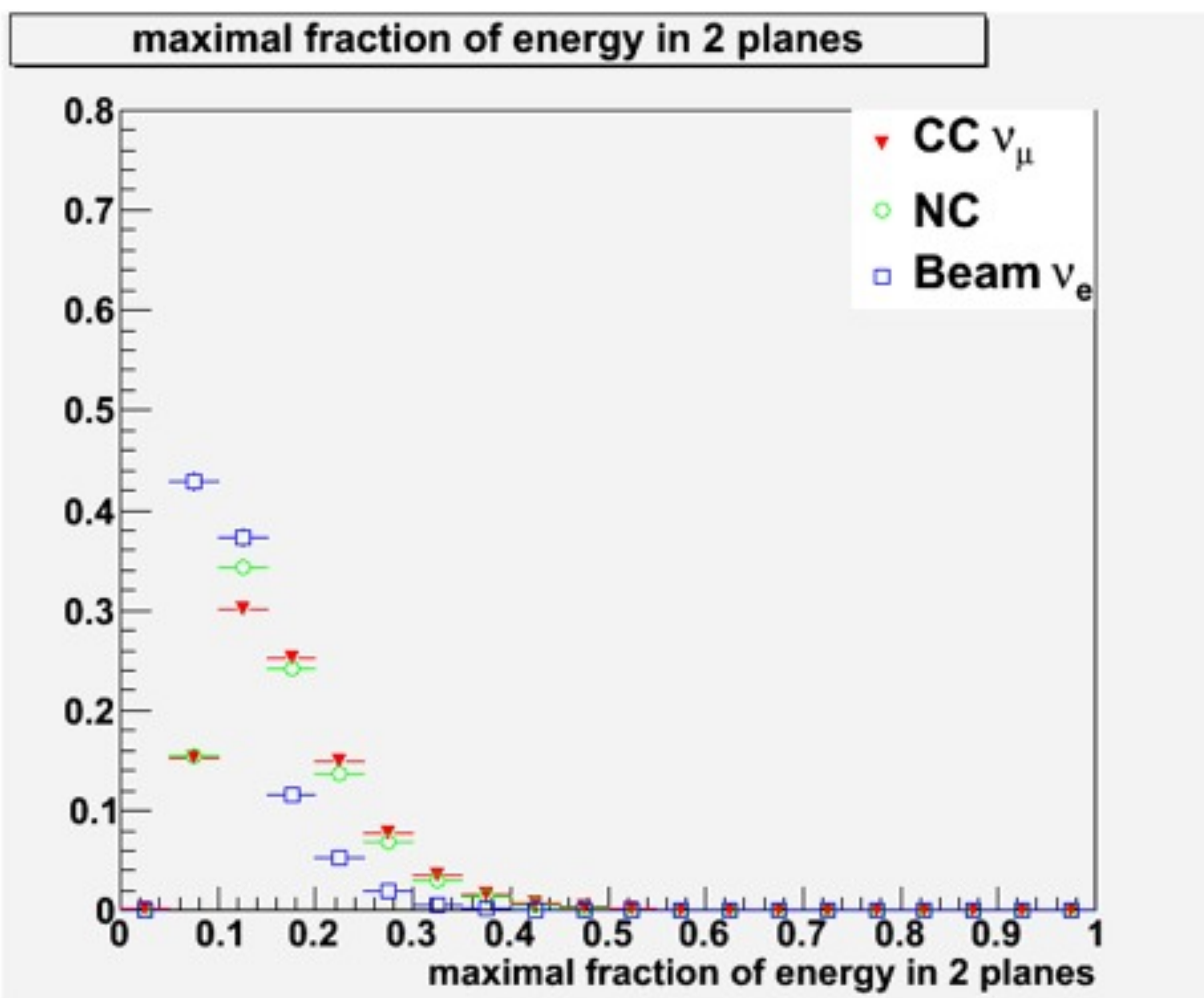
- ΣE_{cells} is the summed energy of all cells associated to the slice with the maximum number of associated cells;
- N_{cells} is the number of cells associated to the slice with the maximum number of cells;
- L_{track} is the lenght of the track;
- The ratio of number of cells associated to the longest track over N_{cells};
- Number of MIP cells (Nmip defined requiring 100 < PECorr < 245, PECorr is corrected photo-electrons);
- The ratio N_{cells} over N_{mip};
- Fraction of energy in first 20 planes;
- Maximal fraction of energy in 2 planes. Reflects the condensity of the longitudinal shower;
- Maximal fraction of energy in 6 planes;
- Fraction of energy in 2σ (σ = 2 cm) road. The ν_e should have relatively narrower transverse shower than the π^0 ;
- Fraction of energy out 3σ road;
- Number of 2D prongs;
- Number of 3D prongs;
- Energy balance between 2 most energetic 2D prongs;
- Energy balance between 2 most energetic 3D prongs.

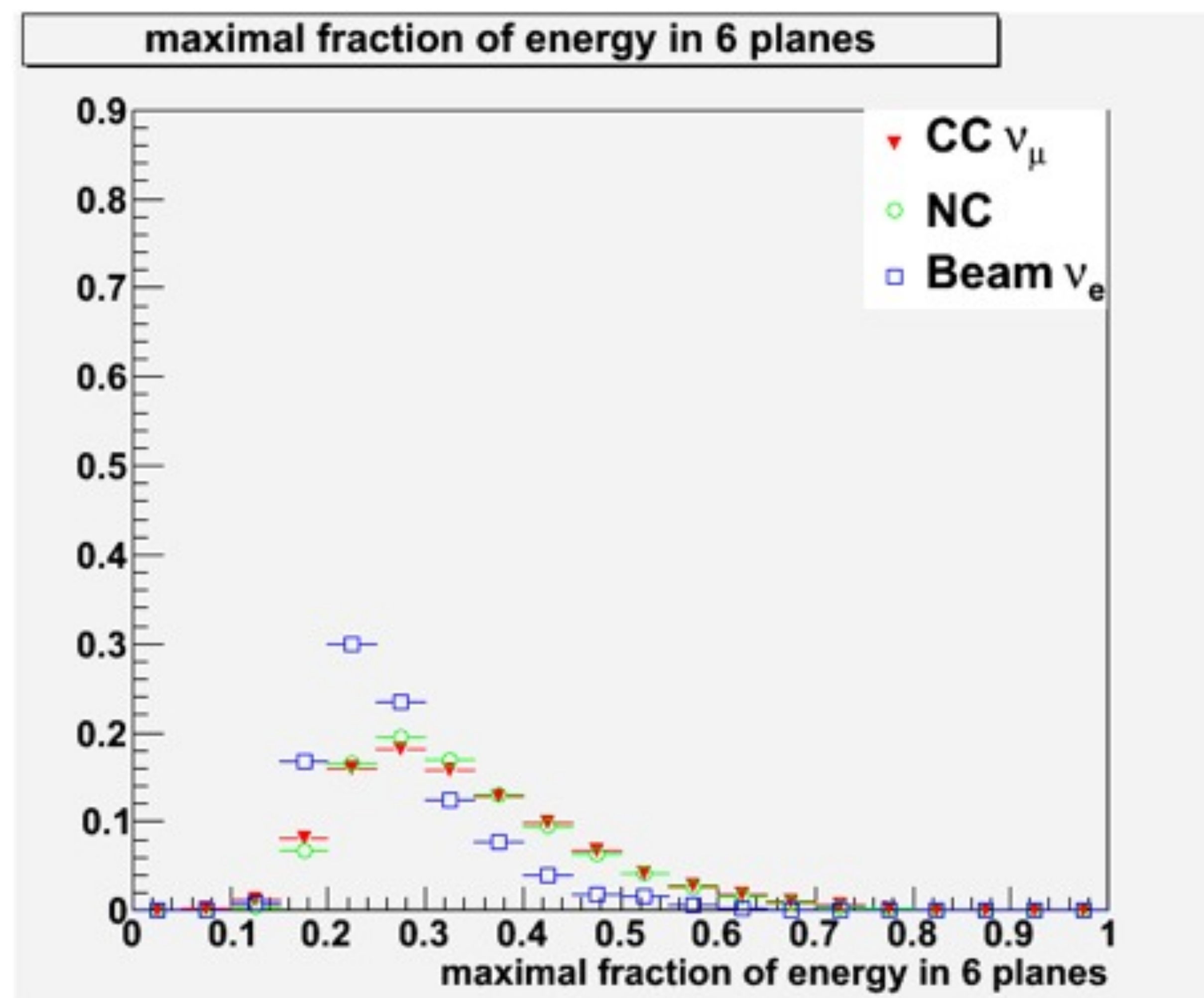
Input Variables

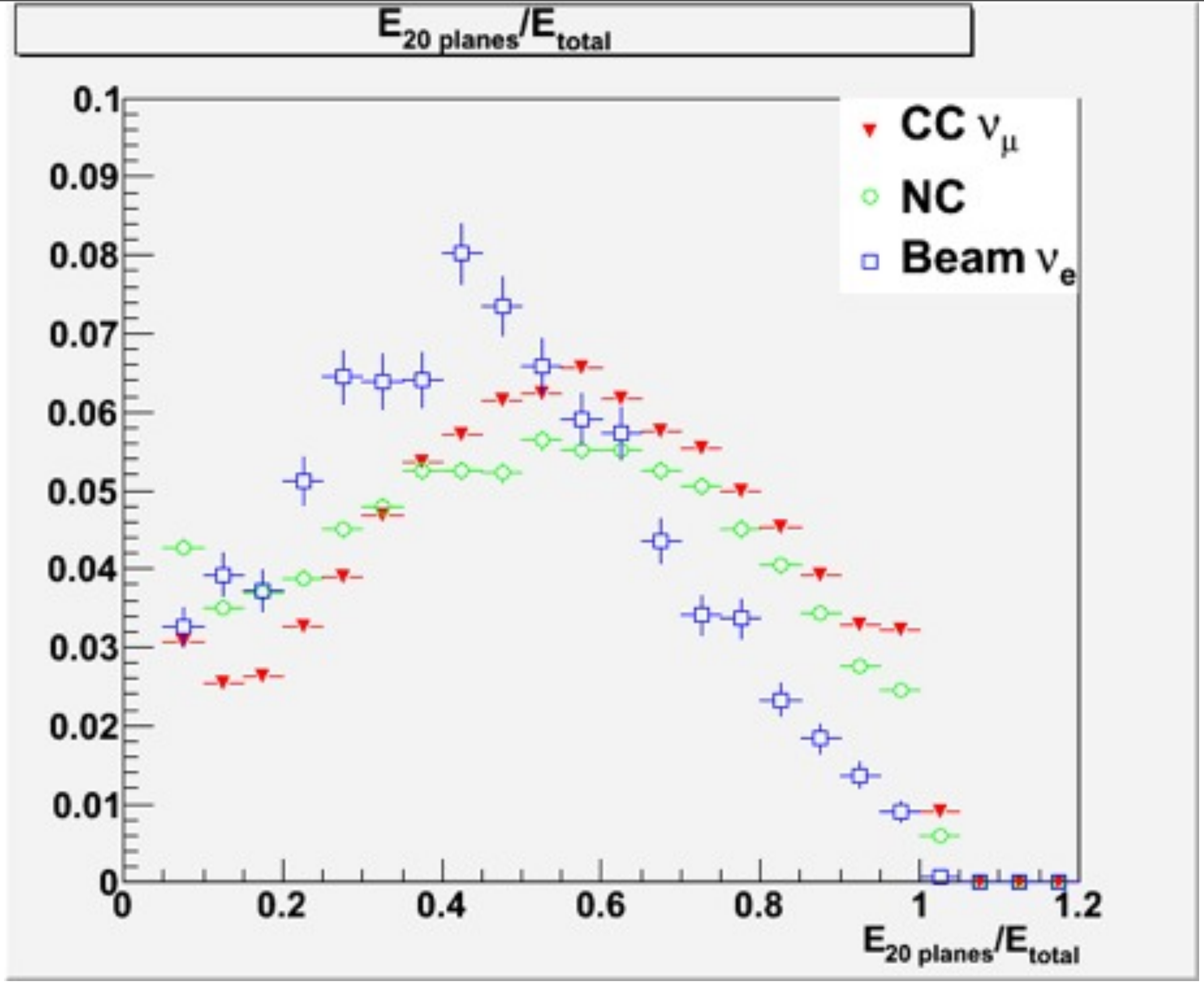


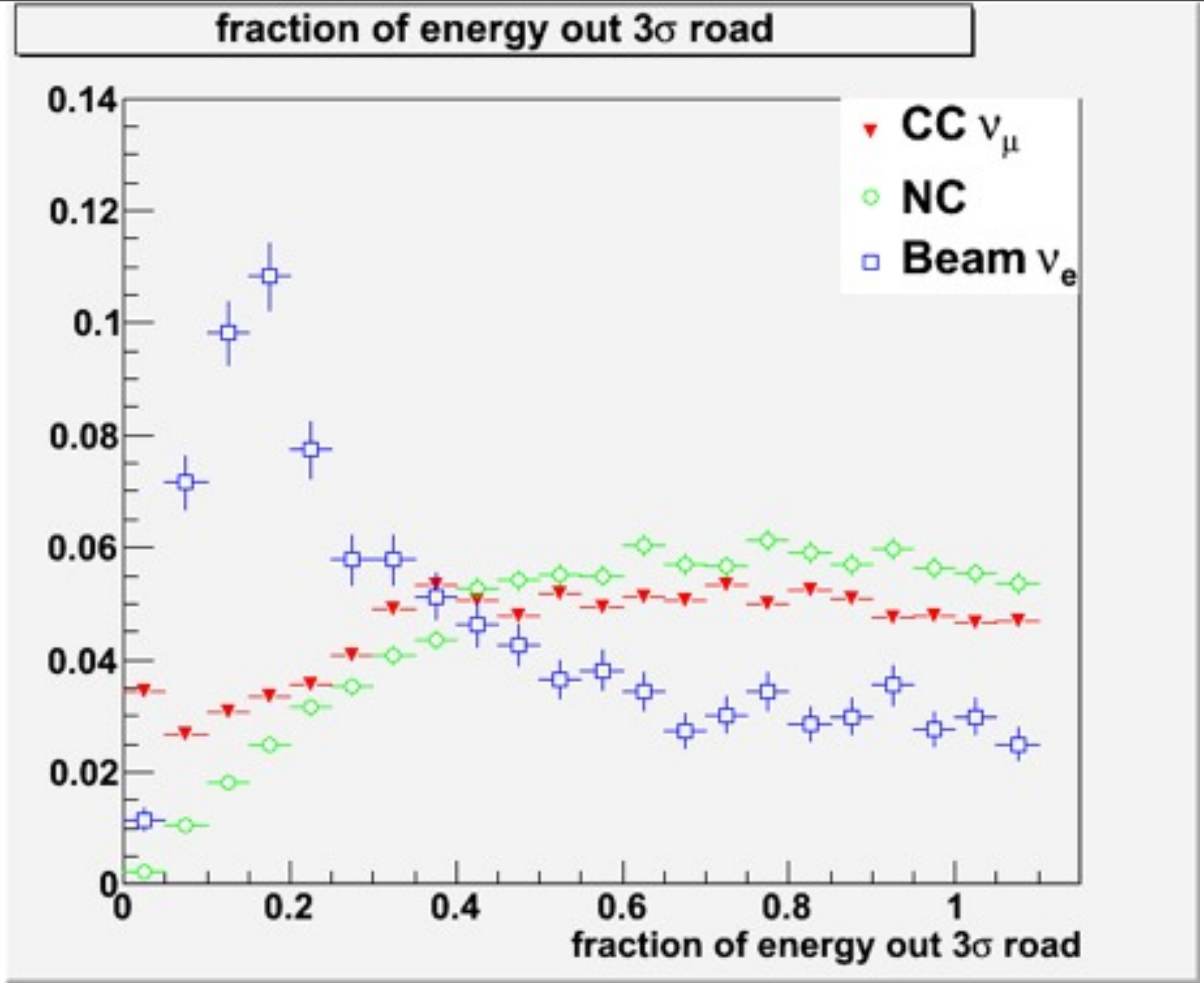

# of mip cells
CC ν_μ
NC
Beam ν_e
# of mip cells

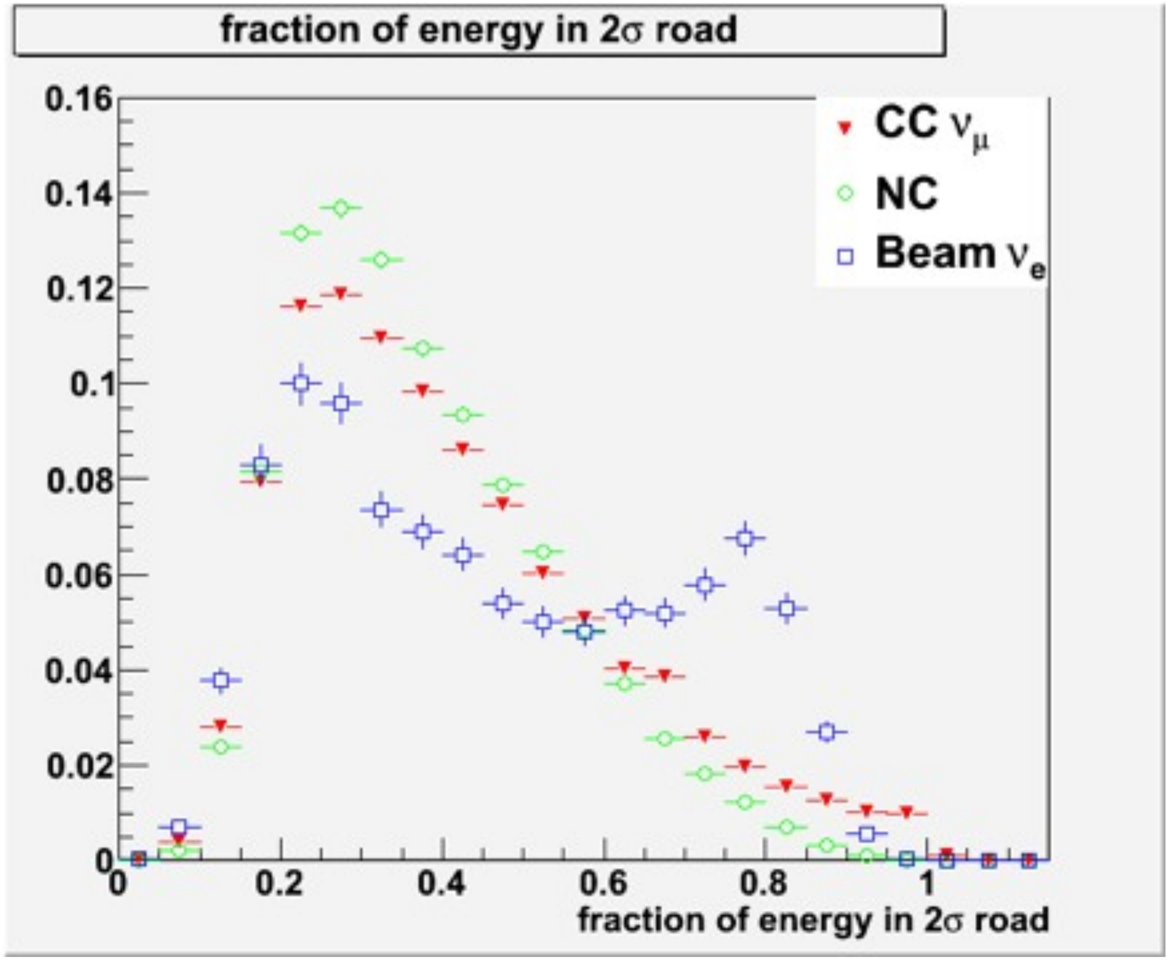

$N^{mip}_{cells}/N^{slice}_{cells}$
CC ν_μ
NC
Beam ν_e
$N^{mip}_{cells}/N^{slice}_{cells}$

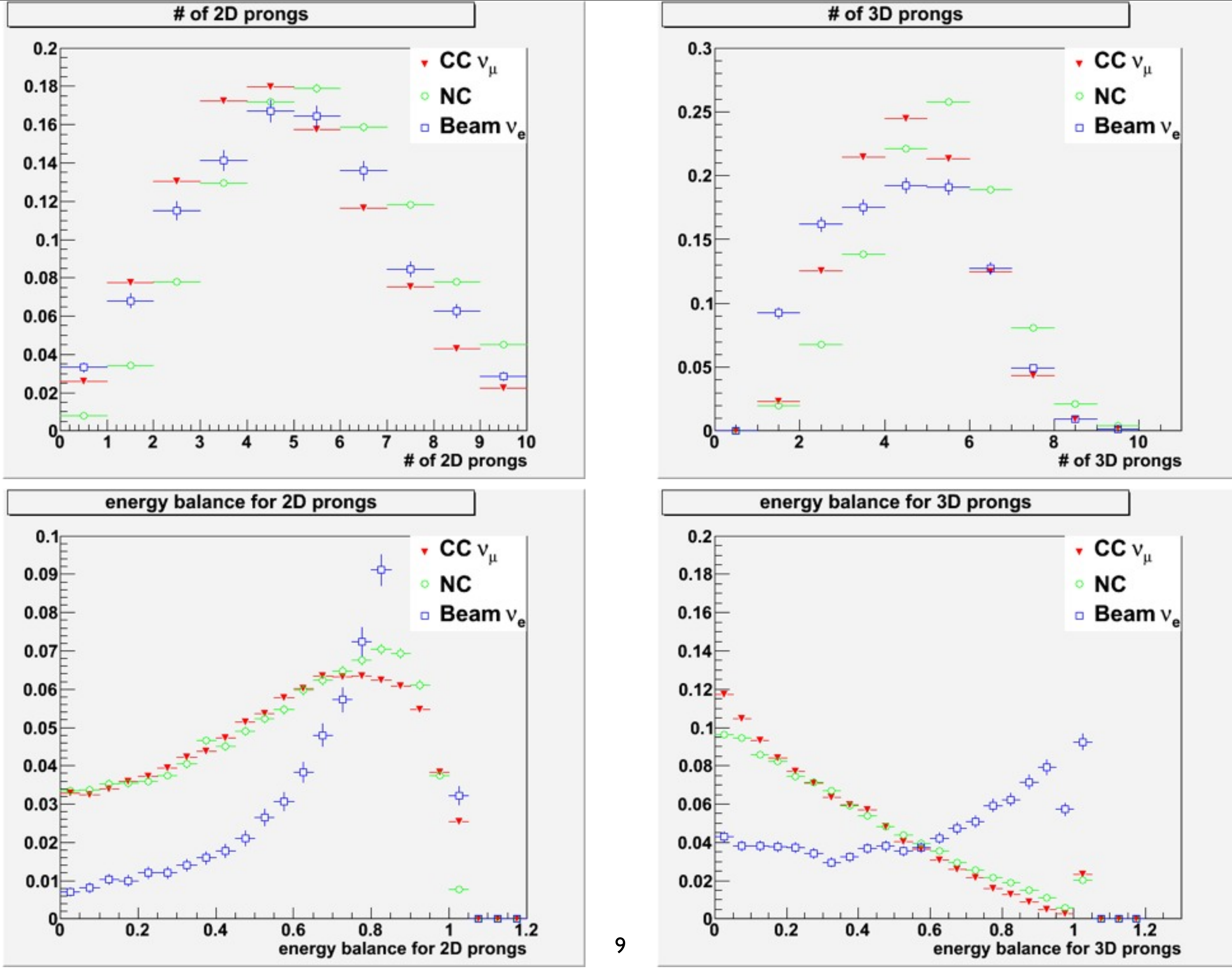

maximal fraction of energy in 2 planes
CC ν_μ
NC
Beam ν_e
maximal fraction of energy in 2 planes


maximal fraction of energy in 6 planes
CC ν_μ
NC
Beam ν_e
maximal fraction of energy in 6 planes


$E_{20\ planes}/E_{total}$
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
0.2
0.4
0.6
0.8
1
1.2
CC ν_μ
NC
Beam ν_e
$E_{20\ planes}/E_{total}$


fraction of energy out 3σ road
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
0.2
0.4
0.6
0.8
1
CC ν_μ
NC
Beam ν_e
fraction of energy out 3σ road

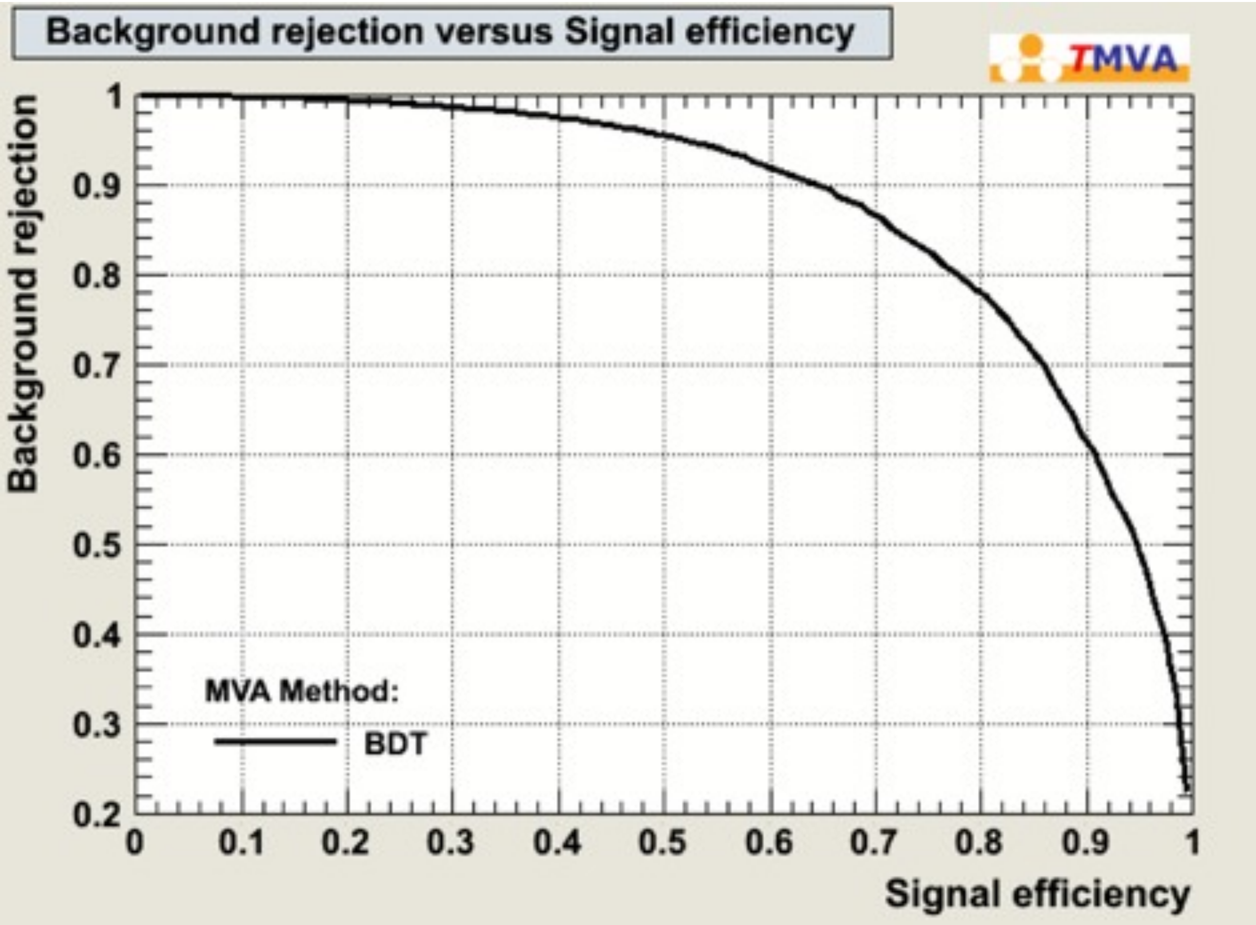

fraction of energy in 2σ road
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
0.2
0.4
0.6
0.8
1
CC ν_μ
NC
Beam ν_e
fraction of energy in 2σ road



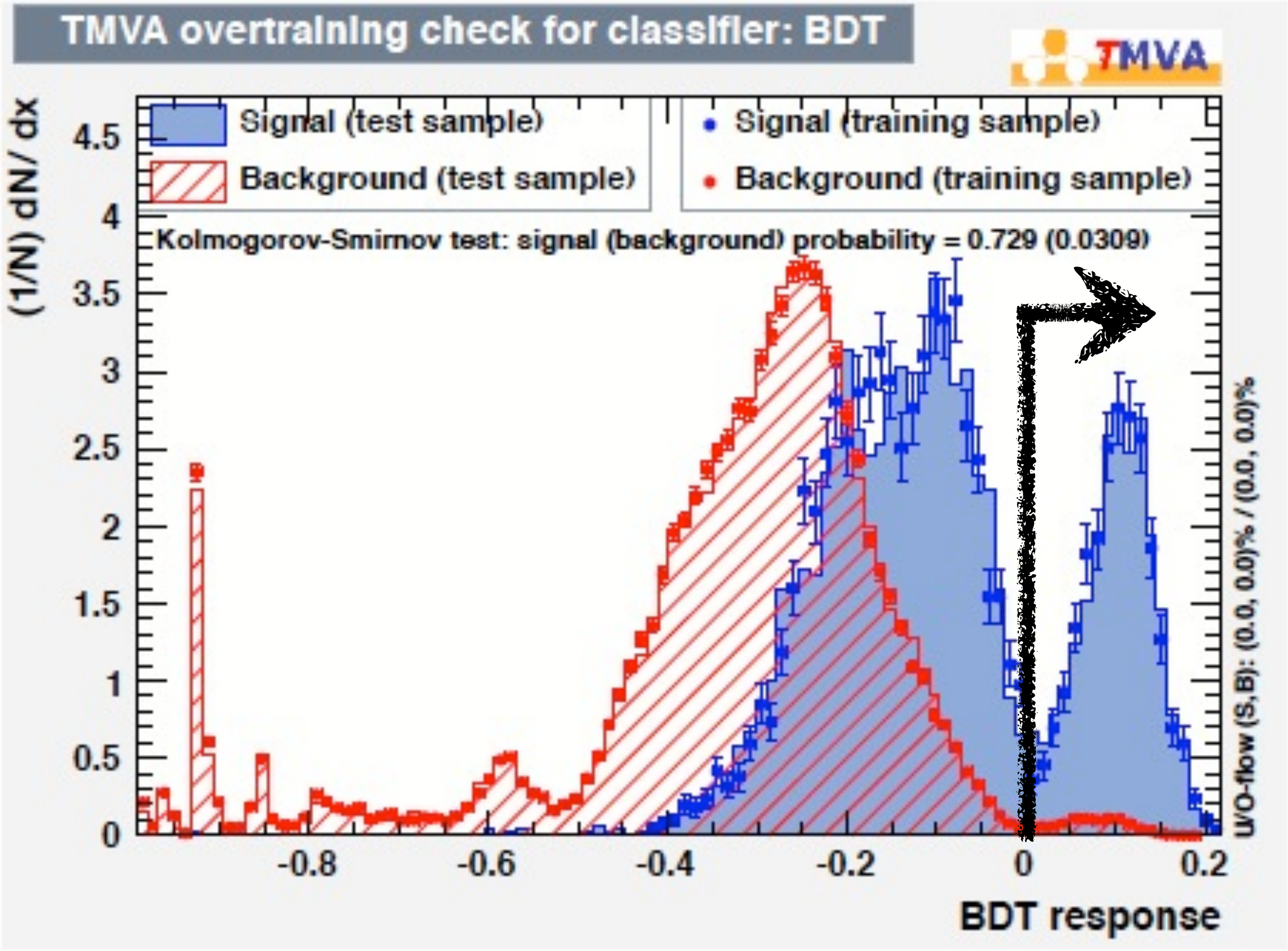
# of 2D prongs
CC ν_μ
NC
Beam ν_e
# of 2D prongs
# of 3D prongs
CC ν_μ
NC
Beam ν_e
# of 3D prongs
energy balance for 2D prongs
CC ν_μ
NC
Beam ν_e
energy balance for 2D prongs
energy balance for 3D prongs
CC ν_μ
NC
Beam ν_e
energy balance for 3D prongs

TMVA Output

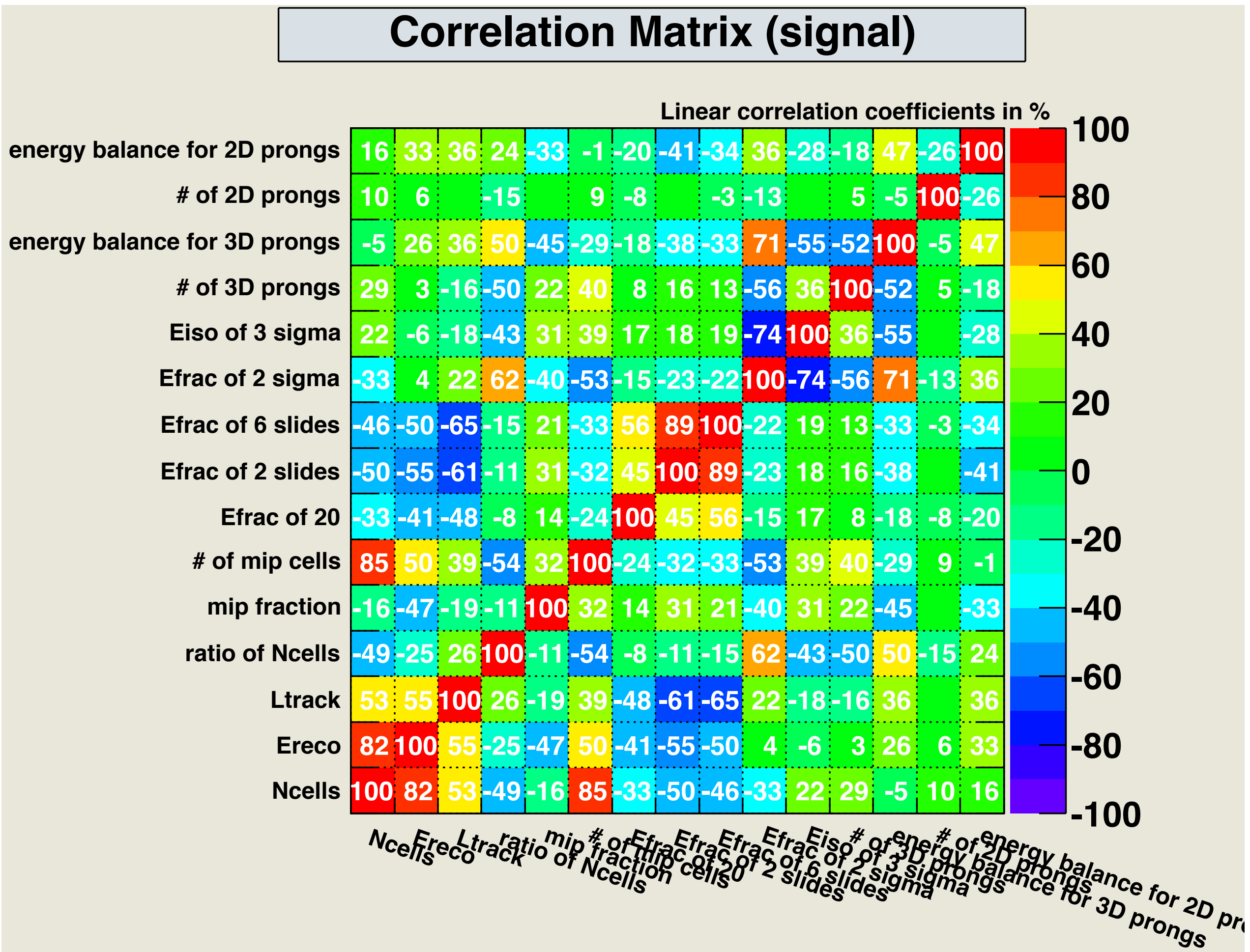
Background rejection versus Signal efficiency

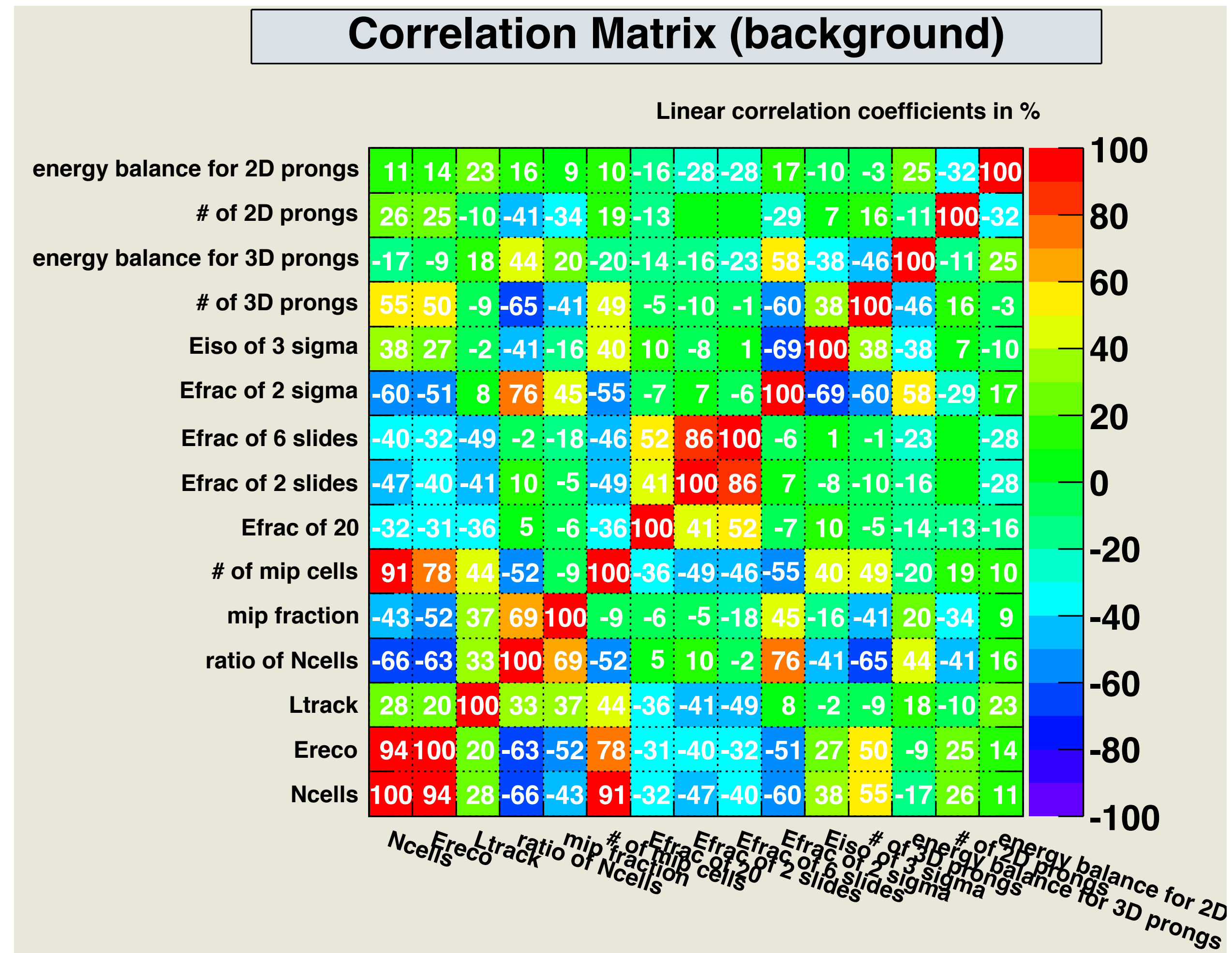


Overtraining check plot



Correlation Matrices for signal and background

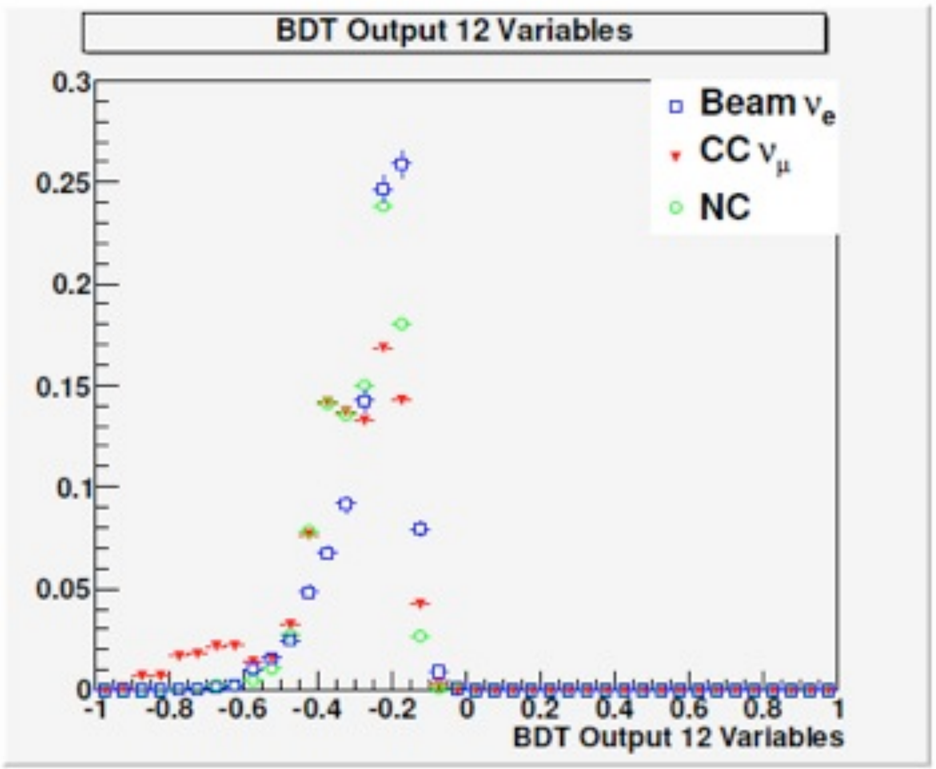


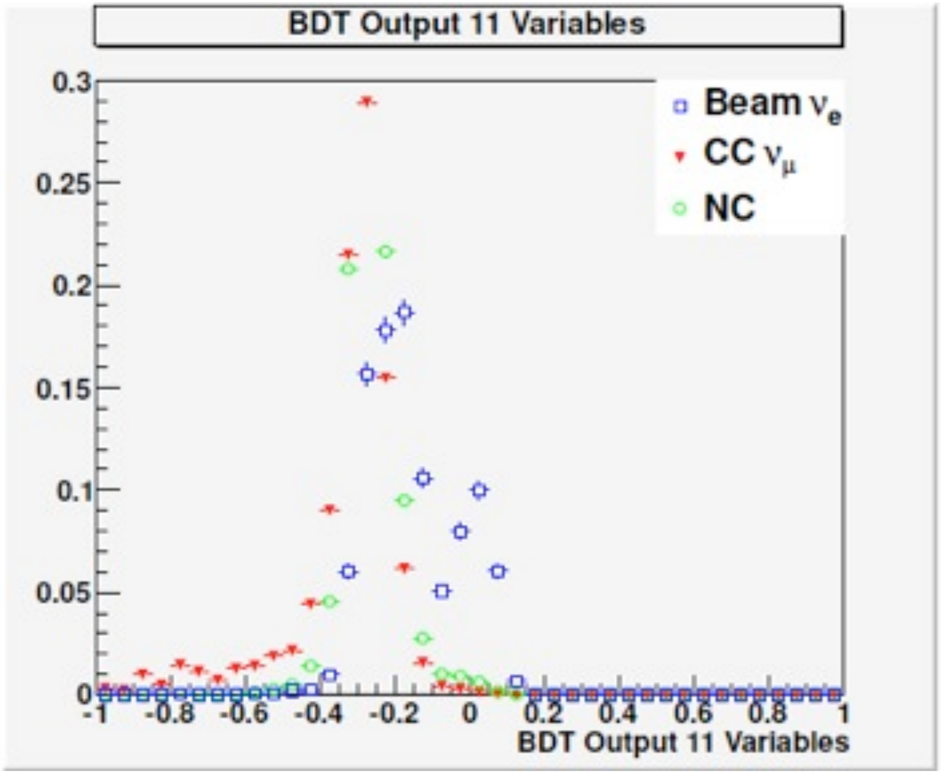


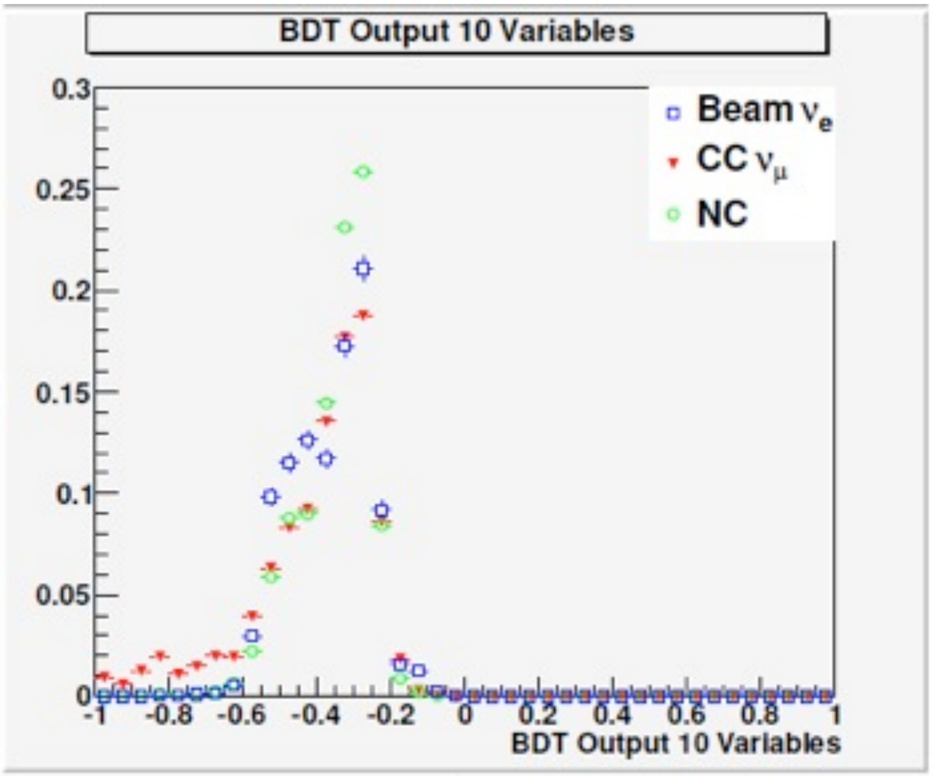
Some variables are correlated

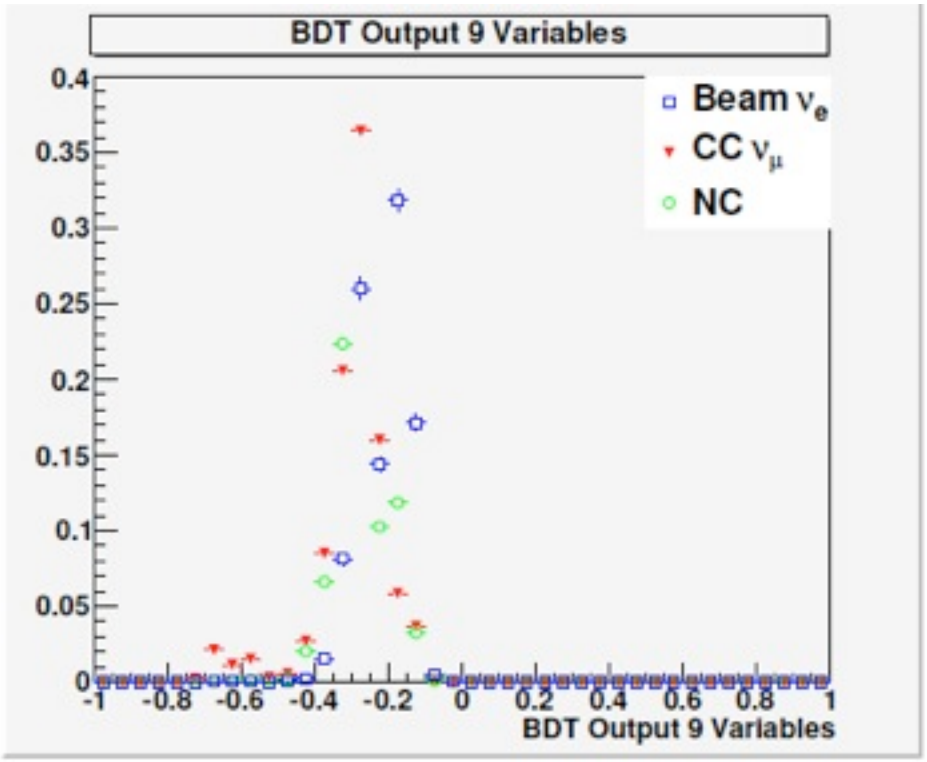
Some variable are correlated

BDT Output

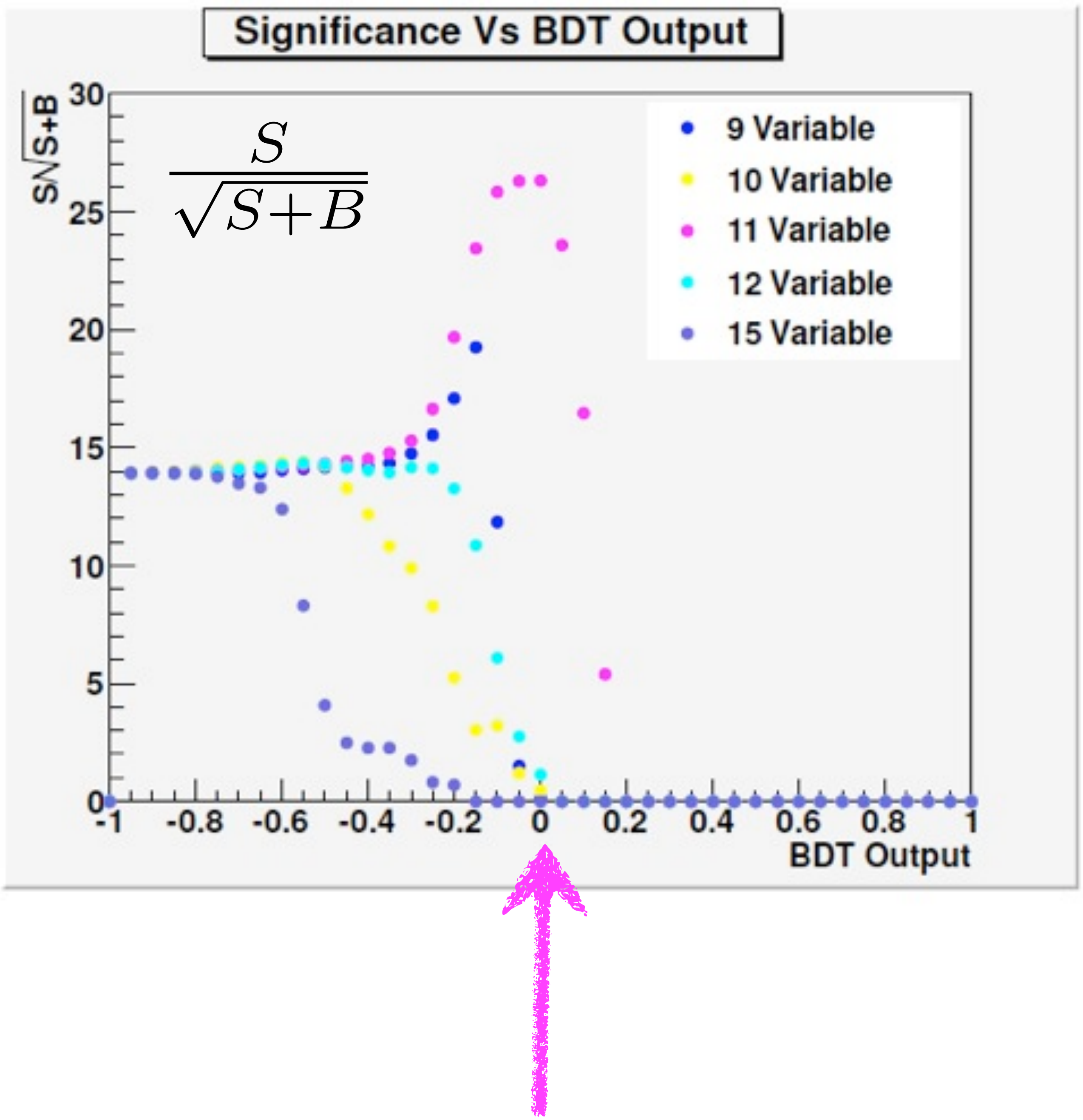

BDT Output 12 Variables
Beam ν_e
CC ν_μ
NC
BDT Output 12 Variables

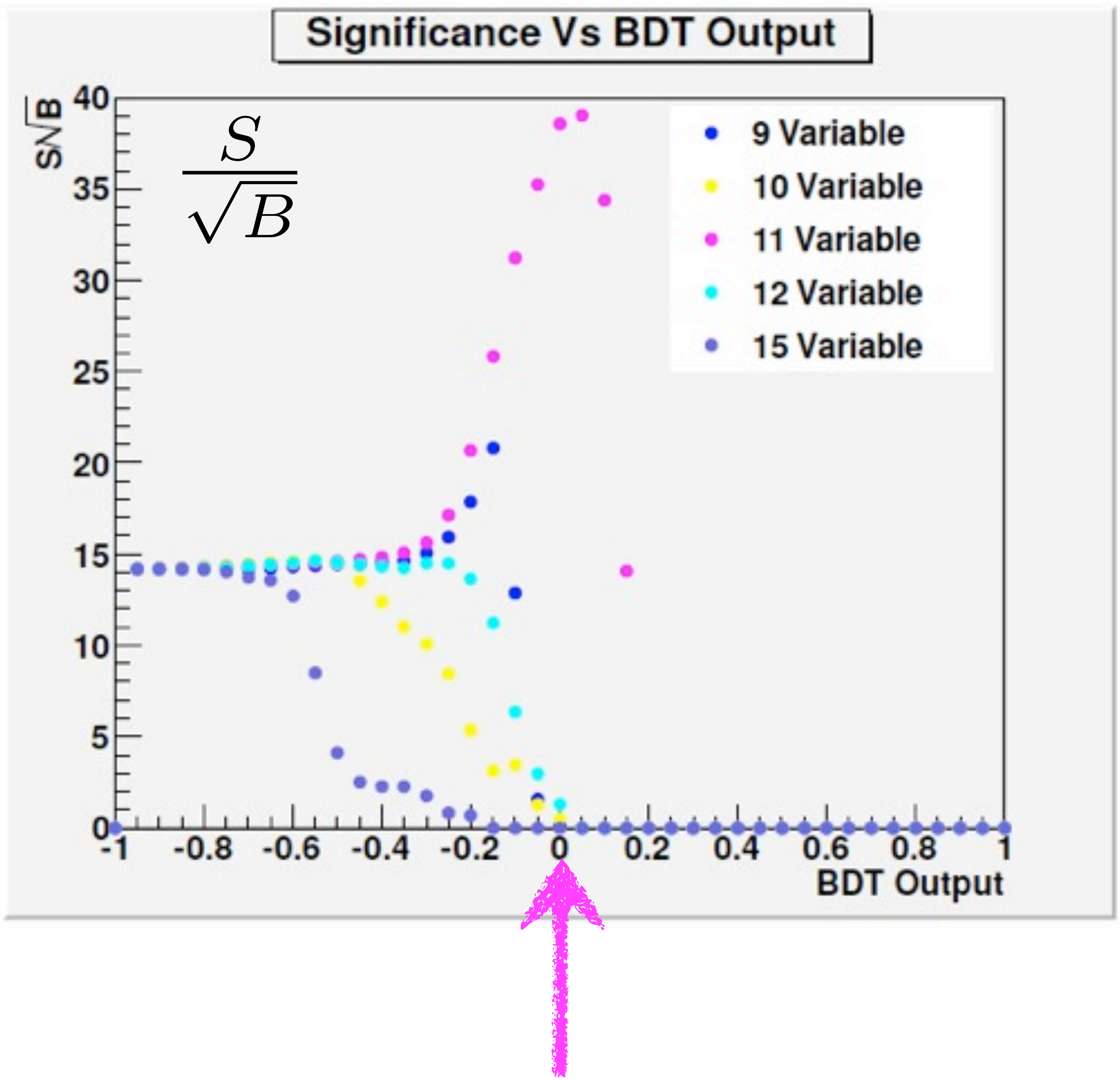

BDT Output 11 Variables
Beam ν_e
CC ν_μ
NC
BDT Output 11 Variables


BDT Output 10 Variables
Beam ν_e
CC ν_μ
NC
BDT Output 10 Variables


BDT Output 9 Variables
Beam ν_e
CC ν_μ
NC
BDT Output 9 Variables

Significance Vs BDT Output



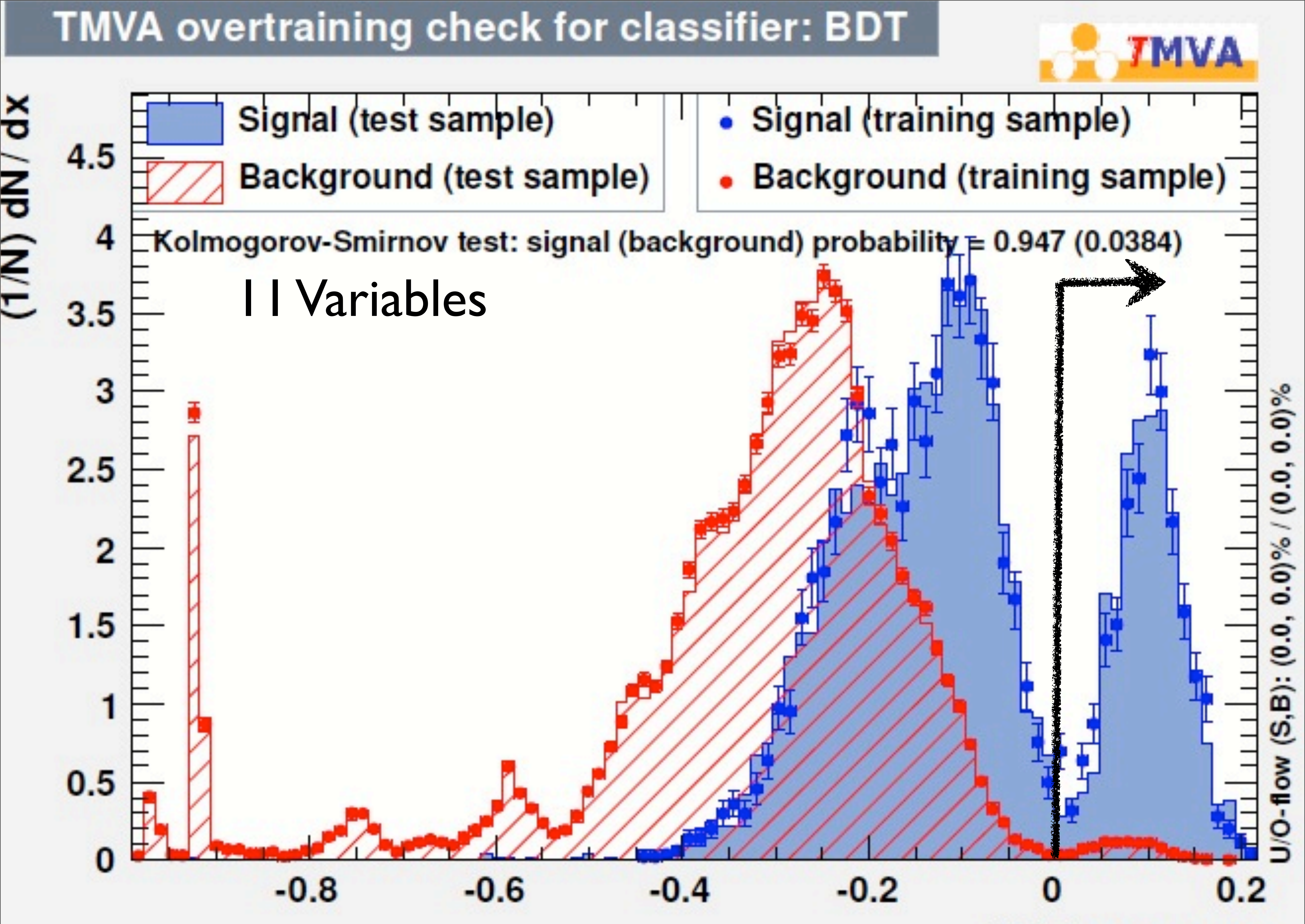


Requiring BDT Output largest than 0 and 11 variables

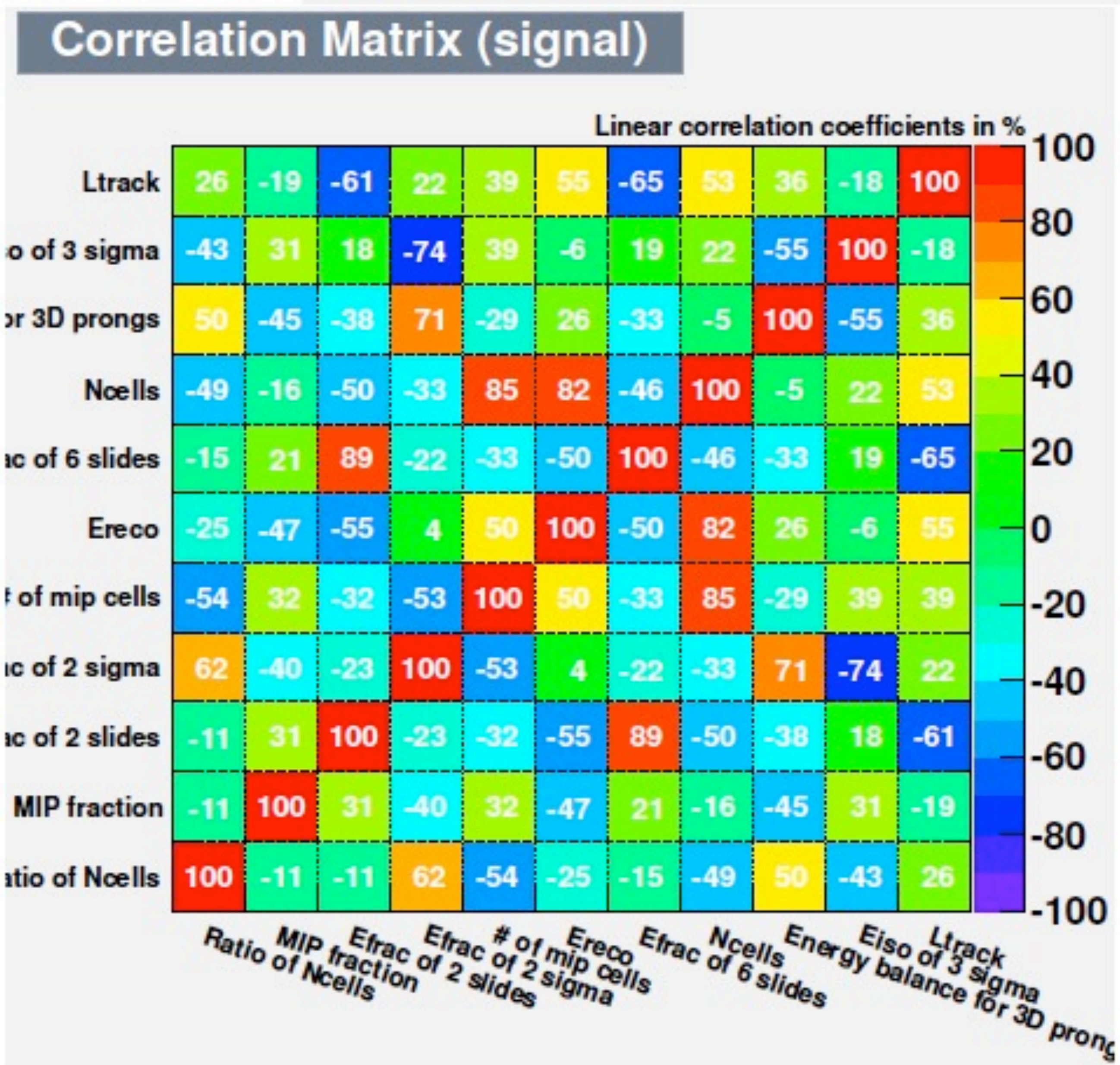
$$\frac{S}{\sqrt{S+B}} = 26\ \%$$

Requiring BDT Output largest than 0 and 11 variables

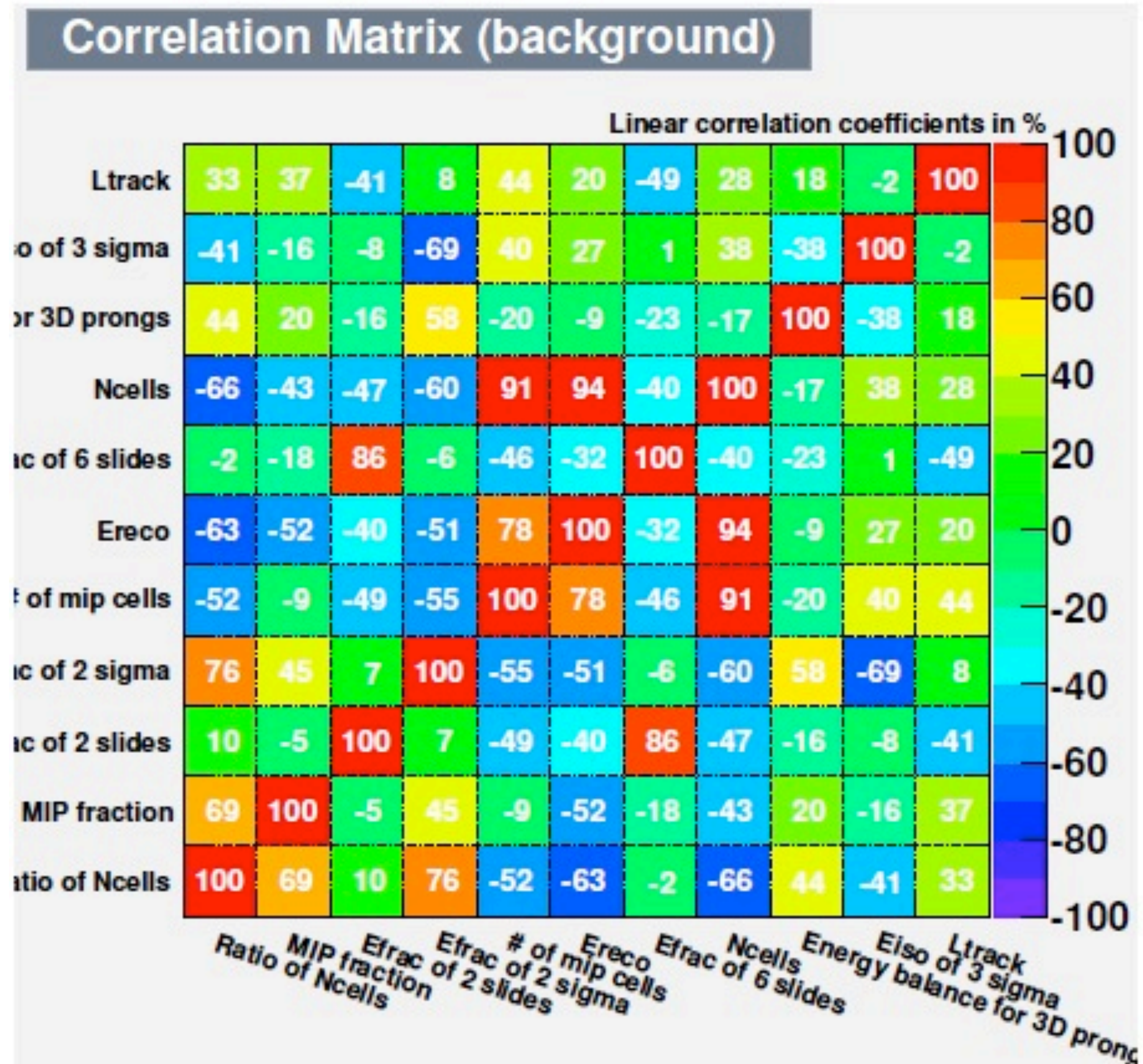
$$\frac{S}{\sqrt{B}} = 39\ \%$$


TMVA overtraining check for classifier: BDT
TMVA
Signal (test sample)
Background (test sample)
Signal (training sample)
Background (training sample)
Kolmogorov-Smirnov test: signal (background) probability = 0.947 (0.0384)
11 Variables
(1/N) dN / dx
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
BDT response
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
-0.8
-0.6
-0.4
-0.2
0
0.2

Correlation Matrix



11 Variables



Reducing the number of correlated variables we can reduce sources of systematic errors

Conclusions

- BDT was been trained, tested and then it are applied to MC files using 15 variables;
- The number of variables are reduced;
- $\frac{S}{\sqrt{S+B}}$ and $\frac{S}{\sqrt{B}}$ are evaluated varying the BDT Output between -1 and 1;
- Requiring BDT output > 0 and using 11 variable

$\frac{S}{\sqrt{S+B}}$ = 26 % $\frac{S}{\sqrt{B}}$ = 39 %

Thank you for your attention!