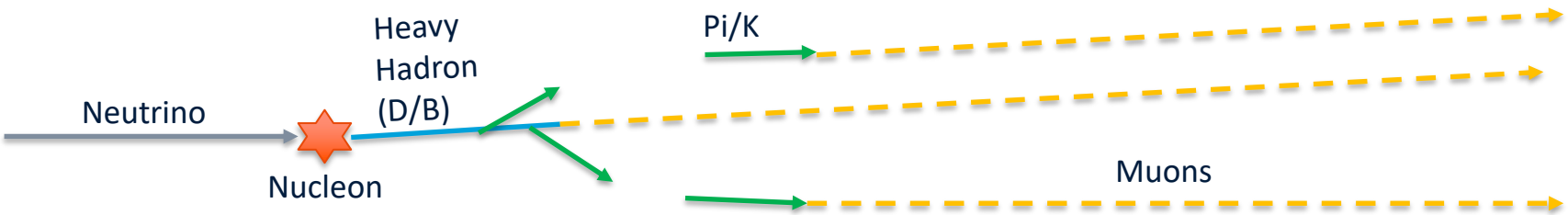
Using boosted decision trees to find muons from neutrino-induced hadronic cascades

Pepijn Kruidenier
p.kruidenier@student.maastrichtuniversity.nl
Maastricht Science Programme BTR
External Supervisor: Aart Heijboer
Internal Supervisor: Lorenzo Reverbi

Maastricht University

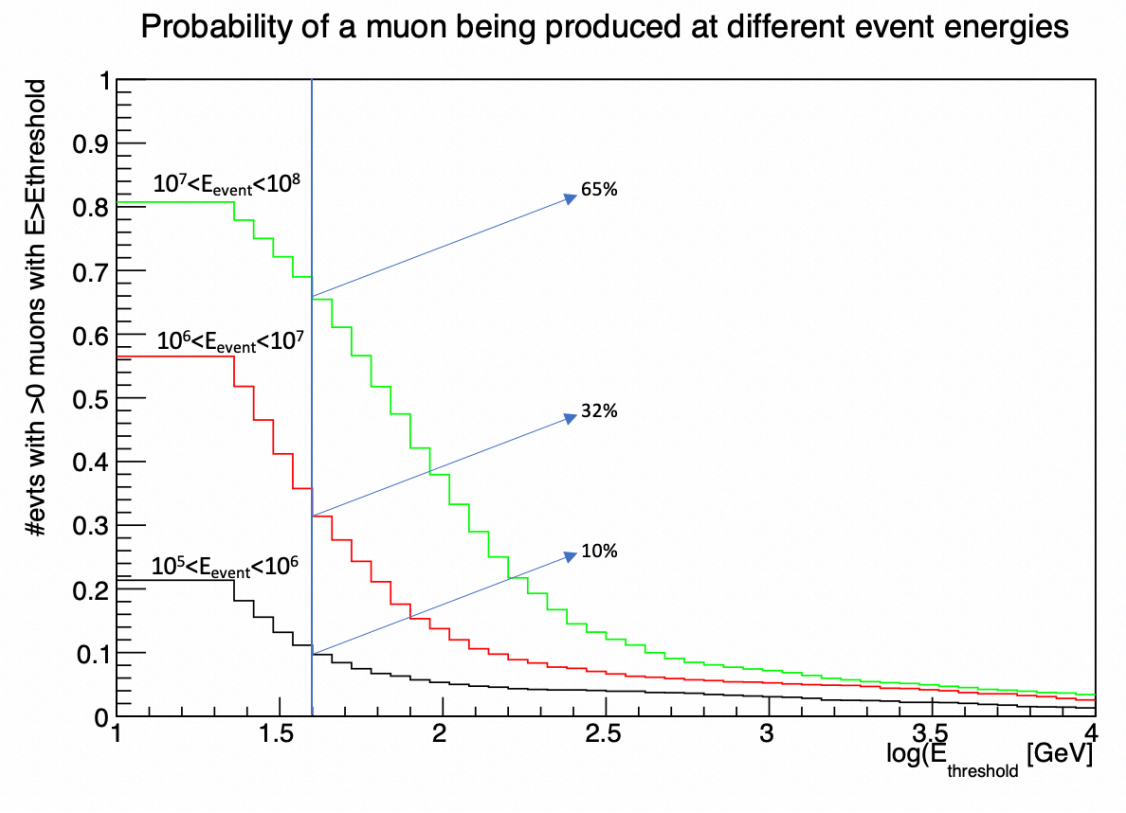
Muon production

- Why are they important
- Track hypothesis
- Shower Hypothesis
  - Pions 99% branching rate
  - Semileptonic decay (10% branching rate)



Muon flux

- 15,000 electron neutrino
- 10% 100TeV
- 32% 1PeV
- 60% 10PeV



BDT

- TMVA
- kBDT
  - 850 Trees
  - Maximum depth of 3 nodes
  - AdaBoost
  - Gini index
- Input several variables

Input Dataset

- 2 data sets which include 20,000 ν_e events in ARCA115.
  - 1 CC and 1 NC
- Cuts
  - No hits
  - has_muon
  - Detector range
  - First hits
- 8400 events left, 5×10^6 mc hits
  - 10% used for training and 90% for evaluation

Variables

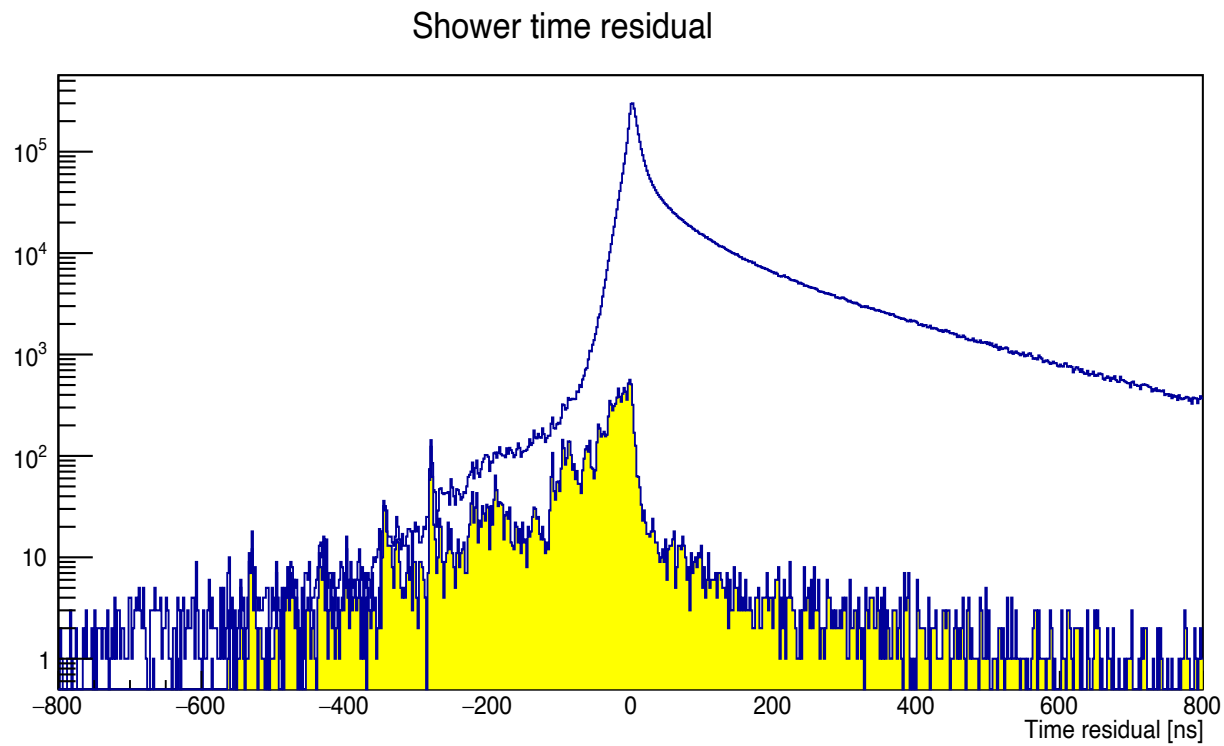
1. Time residual computed with shower-hypothesis
2. Time residual computed with track-hypothesis
3. Distance from (hypothetical) muon track to the hit
4. Angle between between hit direction and shower direction
5. Shower direction reconstructed shower energy
6. Track Energy
7. Muon yes/no

Using best_reco_trk (evt_util.hh) and mc_hits
Help BDT with splitting muons and other

Time residual computed with shower-hypothesis

$$residual_shower(trk, hit) = hit.t - (trk.t - \frac{hit.pos - trk.pos}{v_light})$$

- Muons highlighted in yellow
- Expect only muons with negative time residual

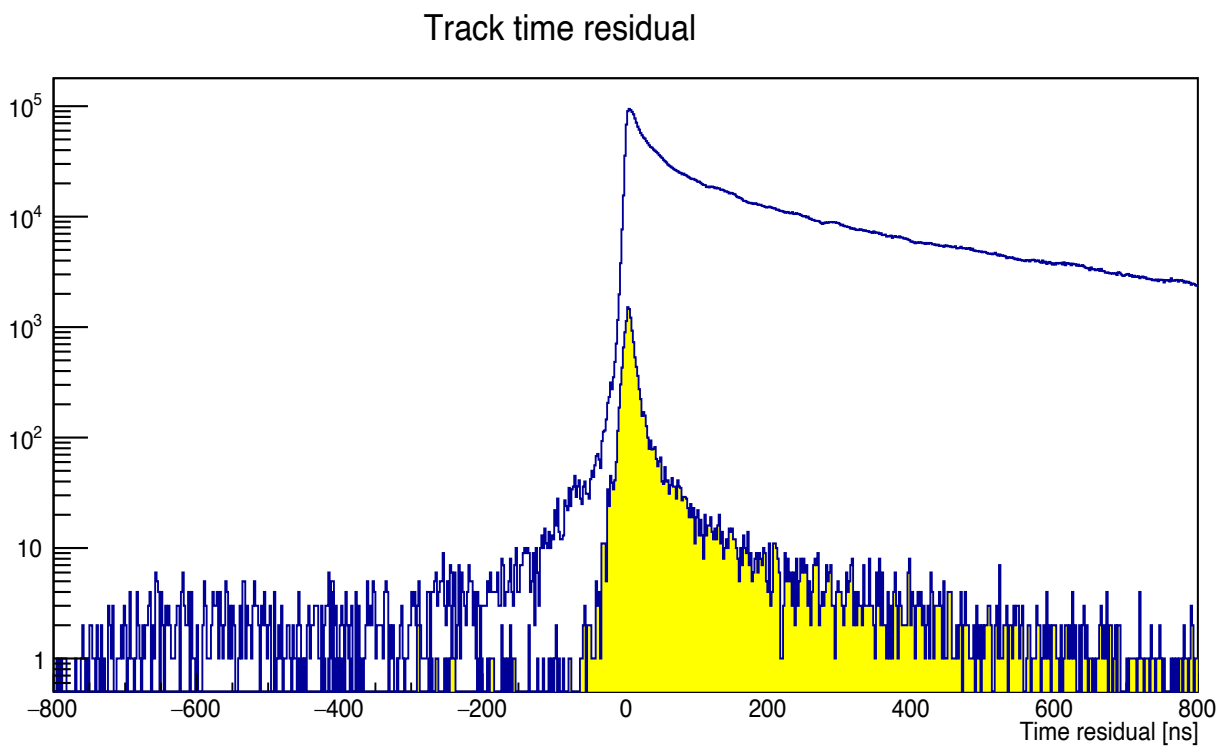


Time residual computed with track hypothesis

$$residual_track(trk, hit) = hit.t - ROOT.time_at(trk, hit)$$

time_at returns the expected time that the Cherenkov light should reach the hit

Expect time residuals above zero



Distance between track and hit

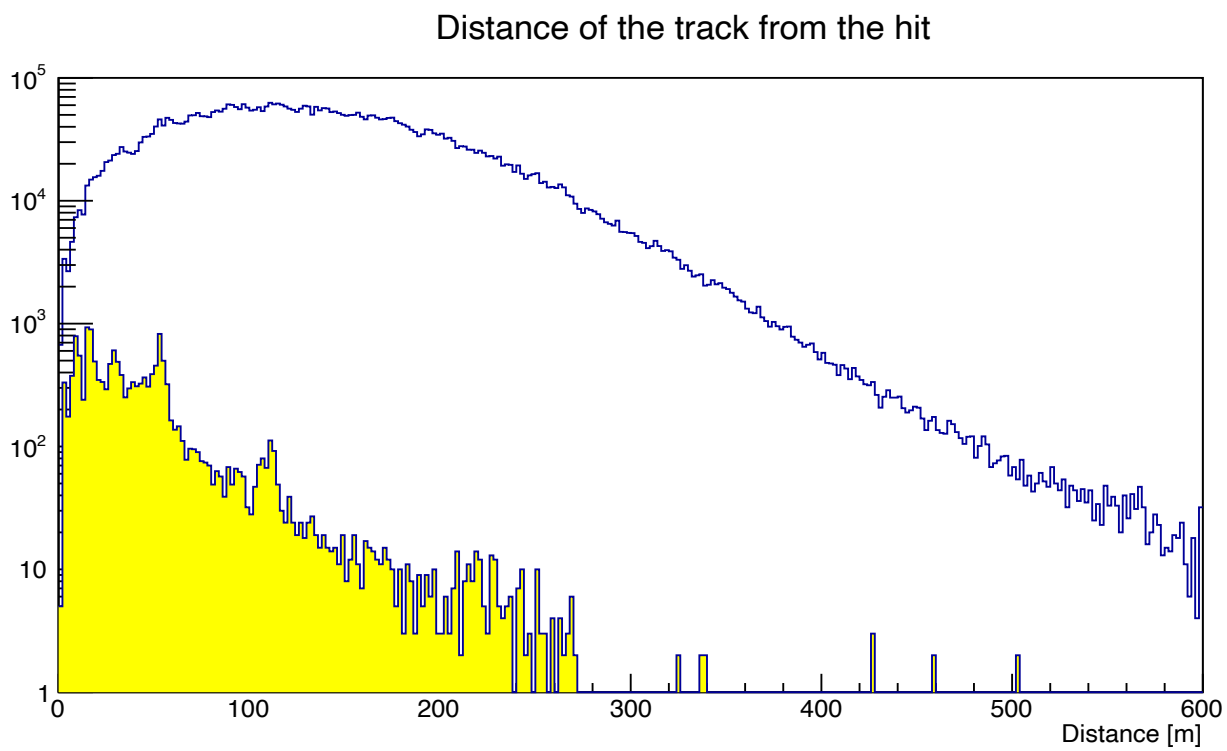
$v = hit.pos - trk.pos$

$l = v \cdot trk.dir$

$k2 = v \cdot v - l * l$

$d = \sqrt{k2}$ if $k2 > 0$

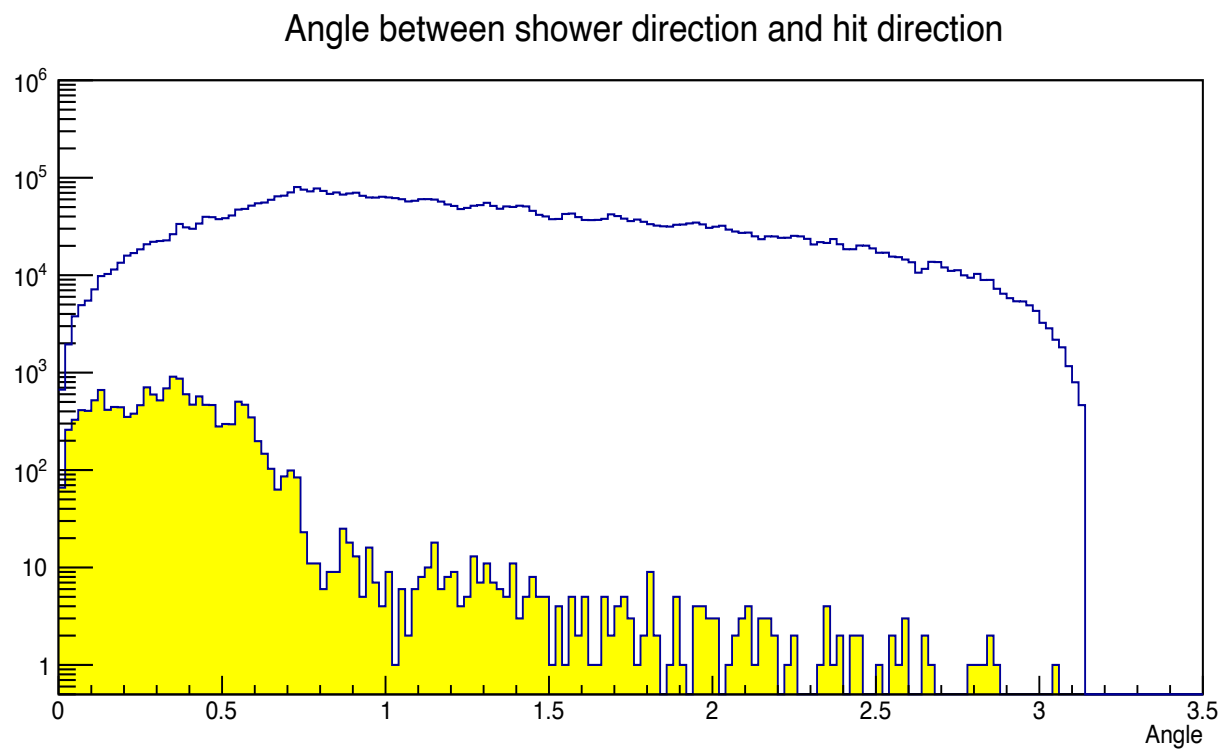
Else $d = 0$



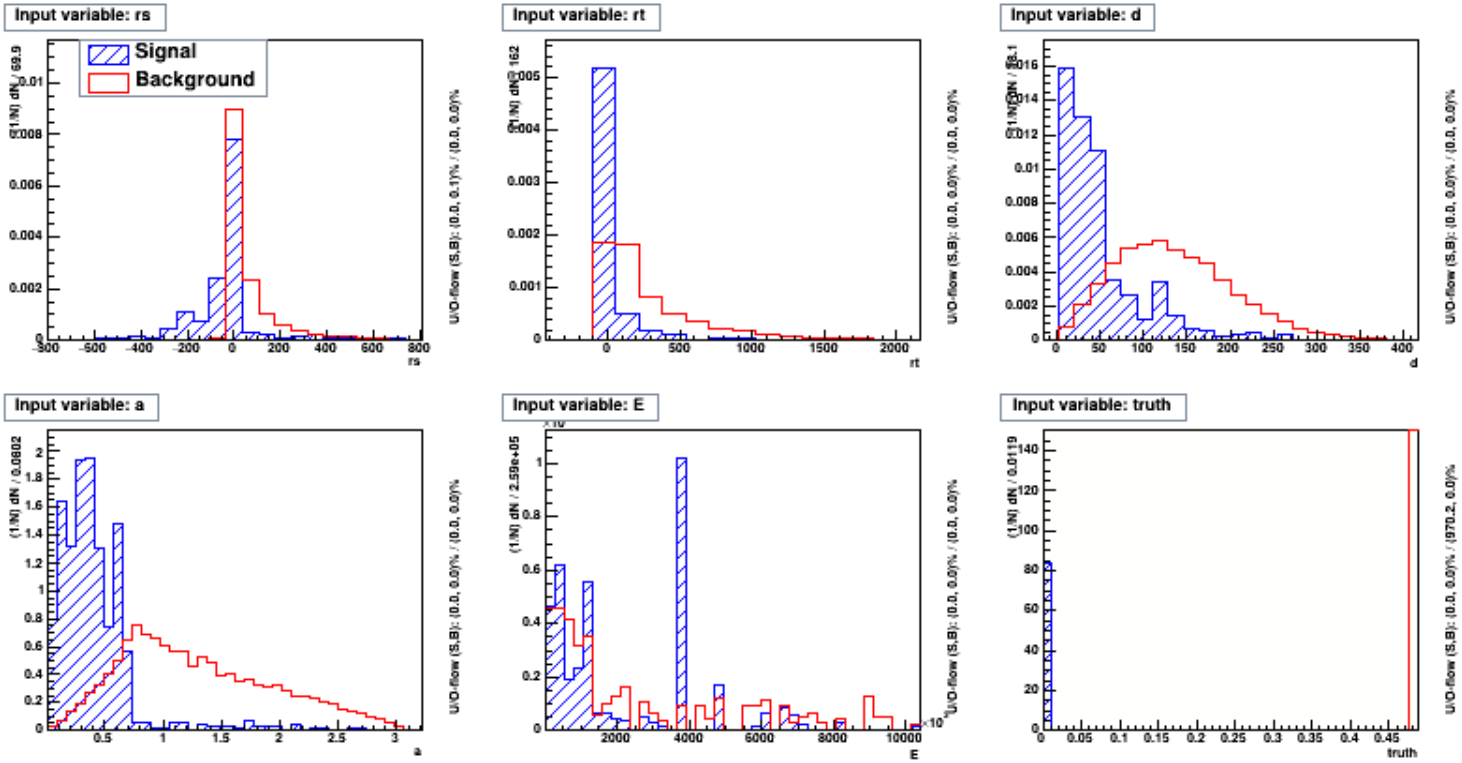
Angle between shower direction and hit direction

$$angle(v1, v2) = cos^{-1}\left(\frac{v1 \cdot v2}{\|v1\|\|v2\|}\right)$$

Where $v1 = h.pos - trk.pos$
and $v2 = trk.dir$

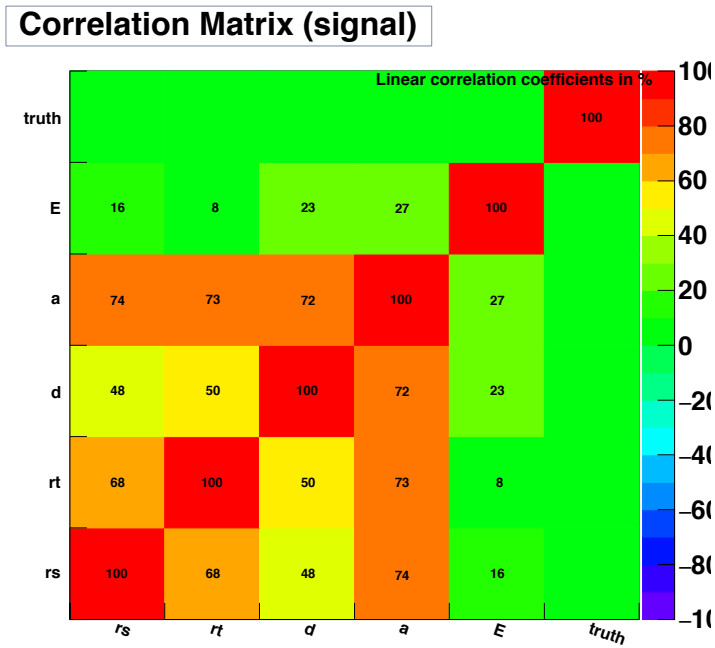


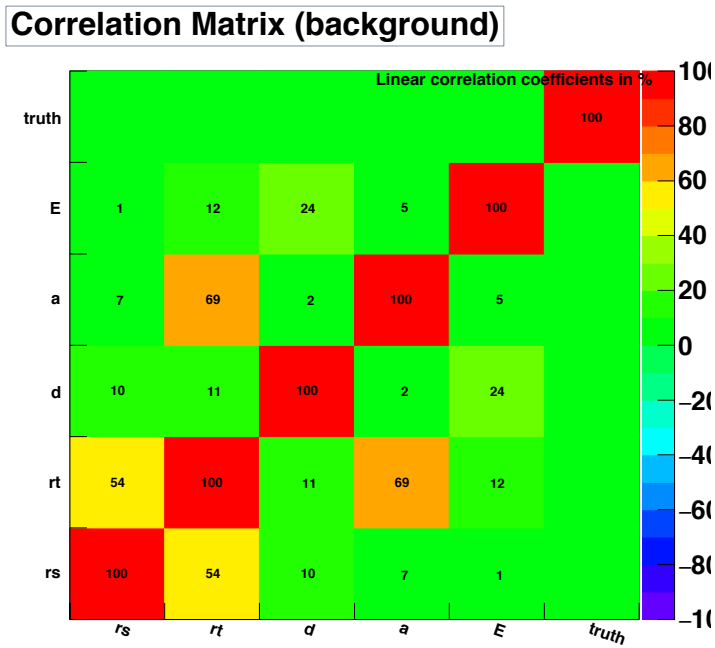
Variables


Input variable: rs
Signal
Background
Input variable: rt
Input variable: d
Input variable: a
Input variable: E
Input variable: truth

kBDT analysis of data

- angle

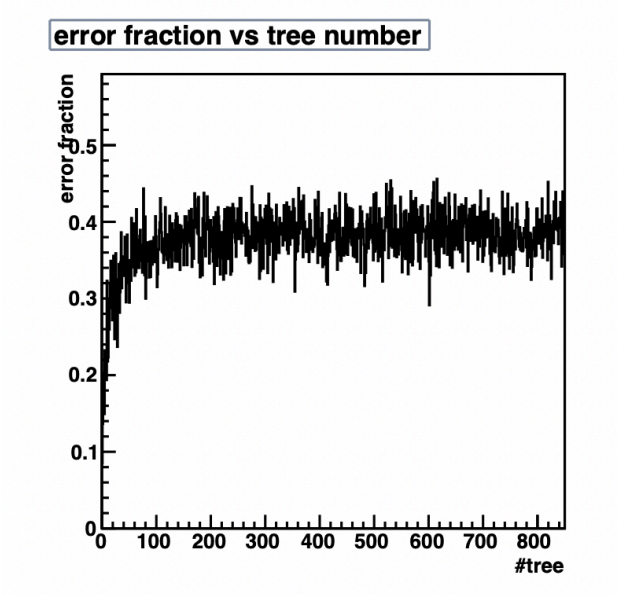


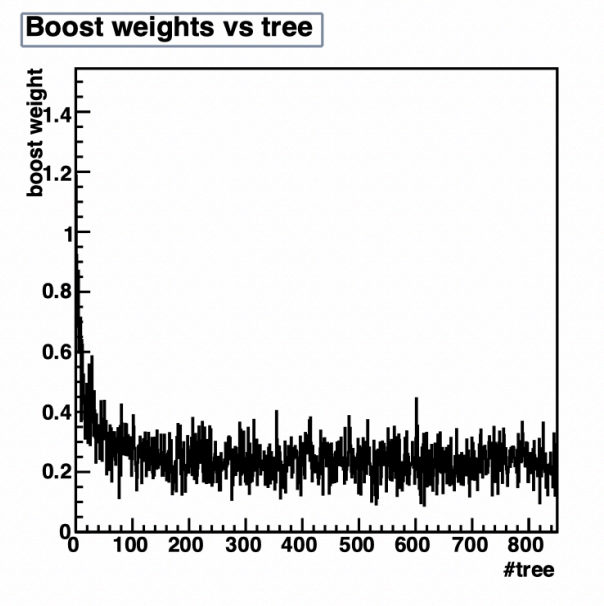


Variable ranking (method specific)

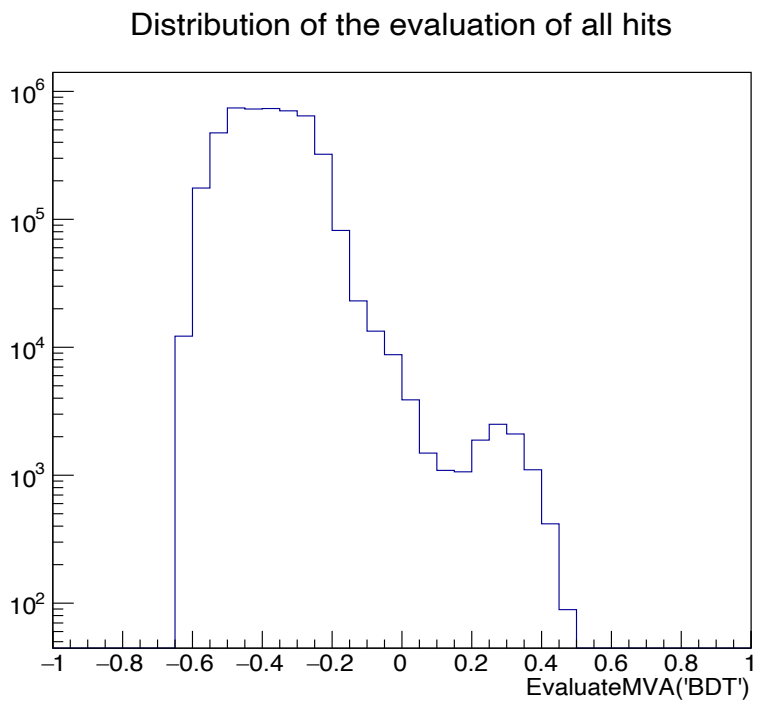
```
: ---------------------------------------
: Rank : Variable  : Variable Importance
: ---------------------------------------
:    1 : E         : 2.653e-01
:    2 : a         : 2.490e-01
:    3 : rs        : 1.763e-01
:    4 : d         : 1.732e-01
:    5 : rt        : 1.362e-01
:    6 : truth      : 0.000e+00
: ---------------------------------------
```

Overtraining Check



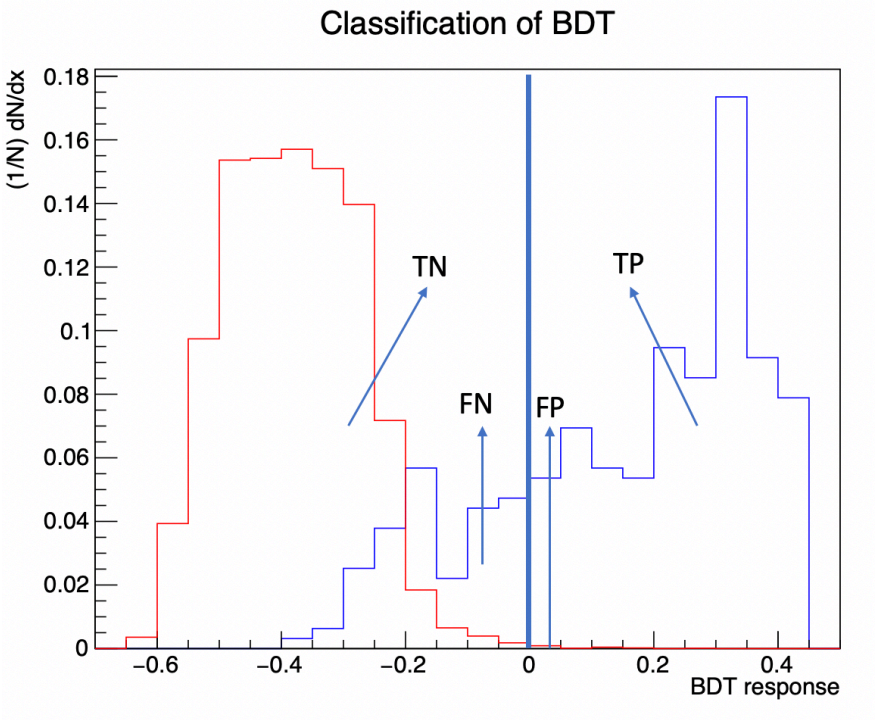


Performance of the BDT

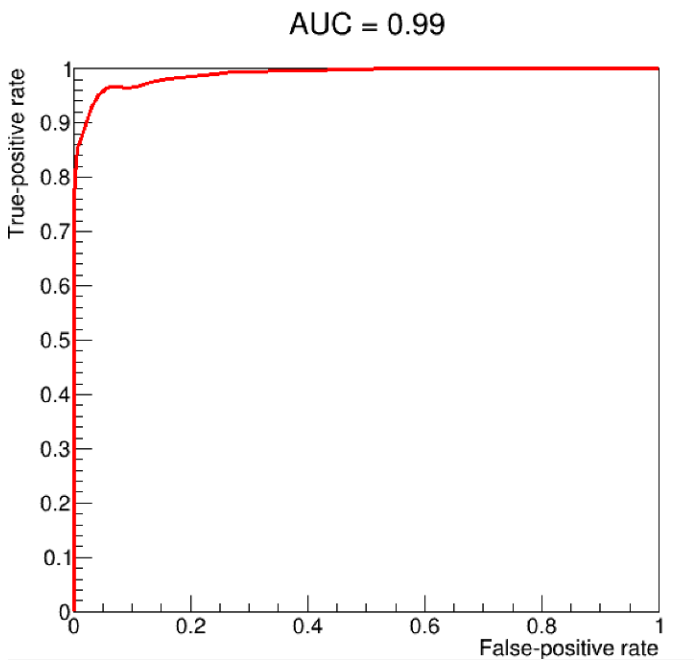


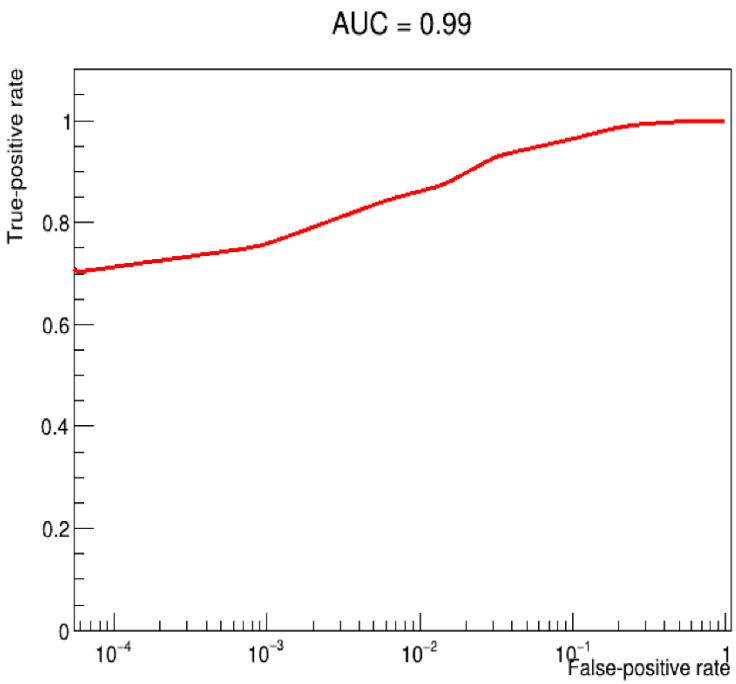
Performance of the BDT

	Classified Positive	Classified Negative
Labelled Positive	True Positive: 10459	False Negative: 3340
Labelled Negative	False Positive: 4757	True Negative: 4654157



Performance of the BDT





For one event

- Classification threshold 0.05
- 96% of signal correct

	Classified Positive	Classified Negative
Labelled Positive	True Positive: 3.69	False Negative: 0.99
Labelled Negative	False Positive: 0.15	True Negative: 1957.04

Issues and Improvements

- Background noise
- Threshold energy
- Same dataset
- Different BDT settings
- Different MVA
- Script that can identify muons in an event

Conclusion

Thank you!