Lancaster University

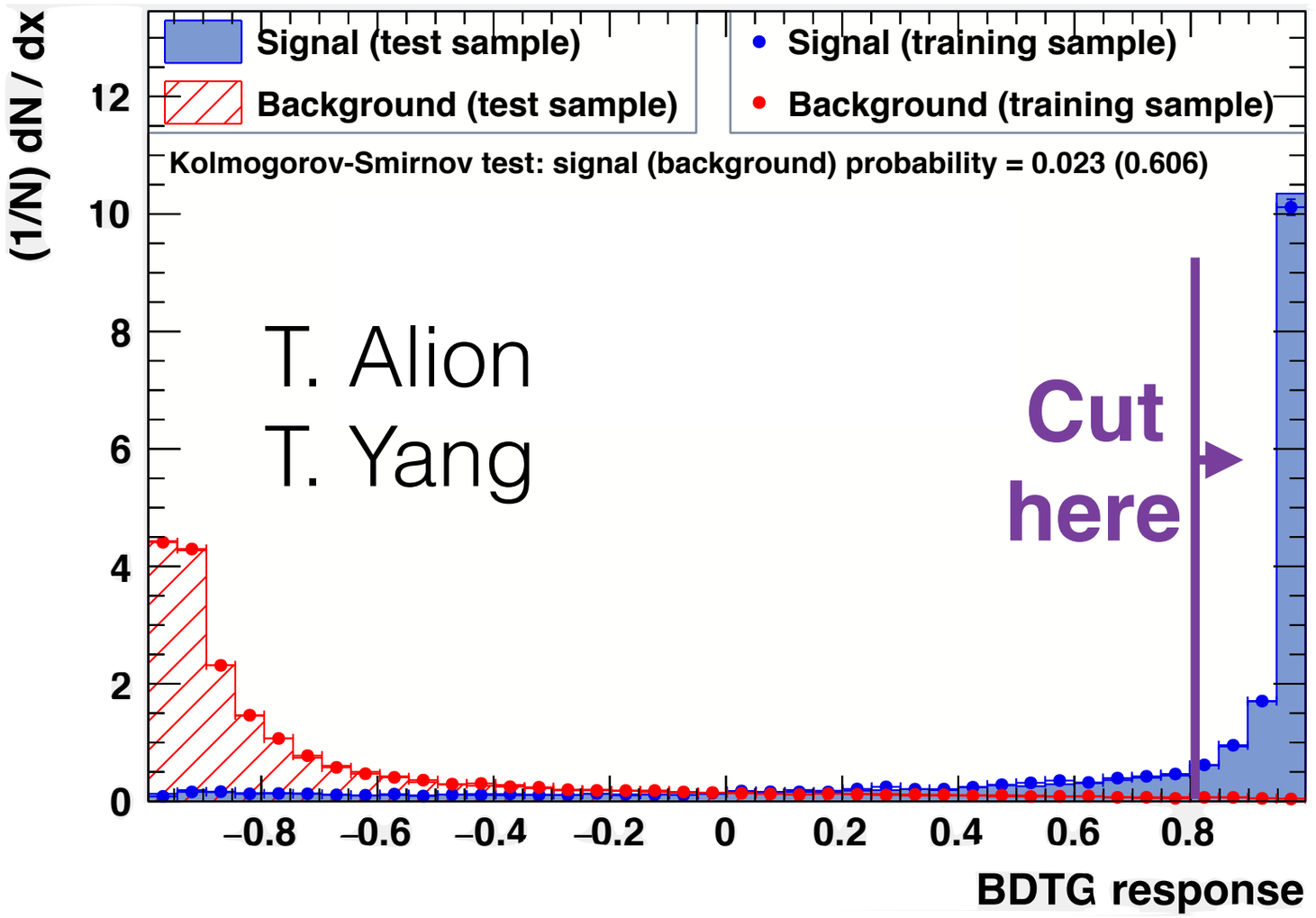
ν_μ Far Detector Event Selection

Dom Brailsford
Tyler Alion
Tingjun Yang

Long Baseline Physics Meeting
19/12/16

ν_μ Far Detector Event Selection

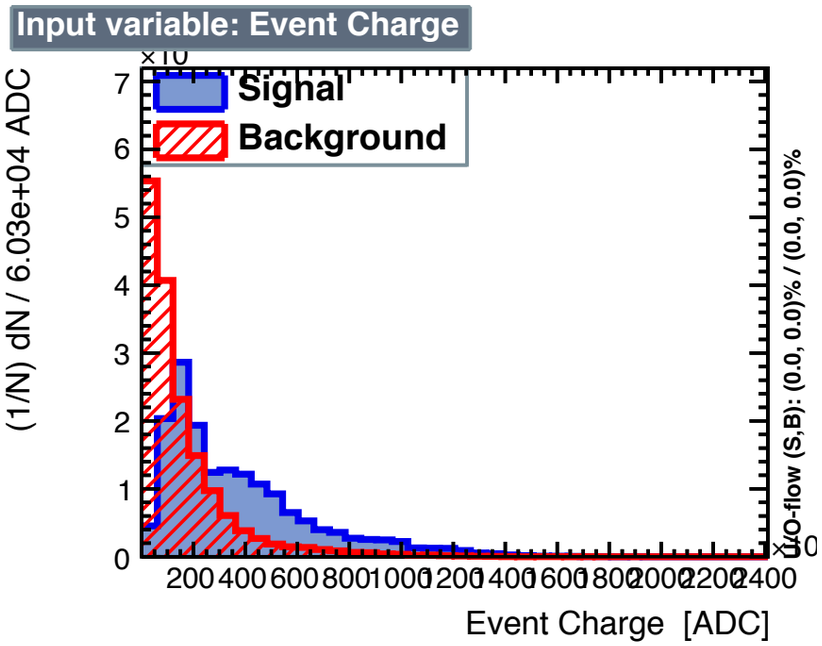
- Selection developed by T. Alion and T. Yang
- Uses a BDT to select CC ν_μ events. Assesses
  - Event topology
  - Event shape
  - Event charge
- Events selected with MVA response greater than 0.8
  - This value is not tuned

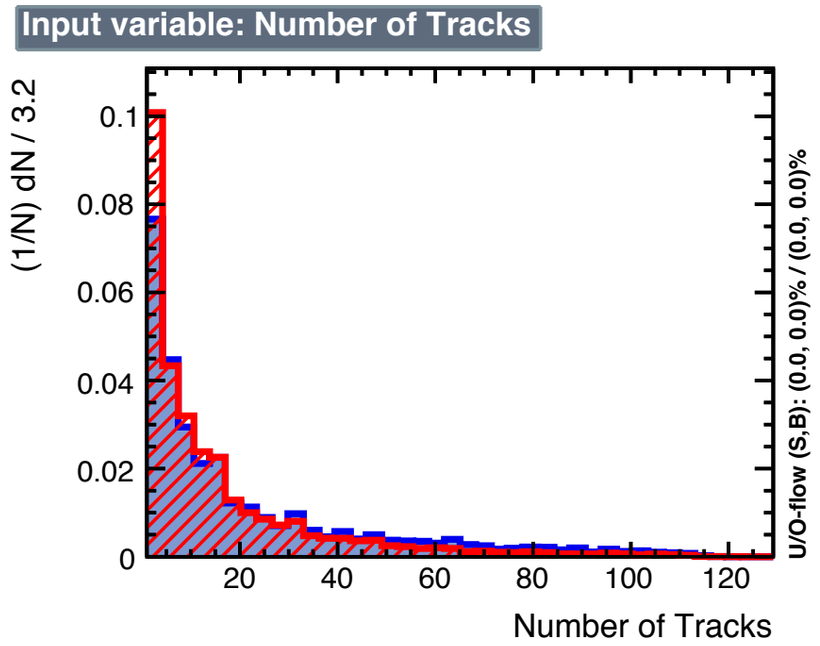


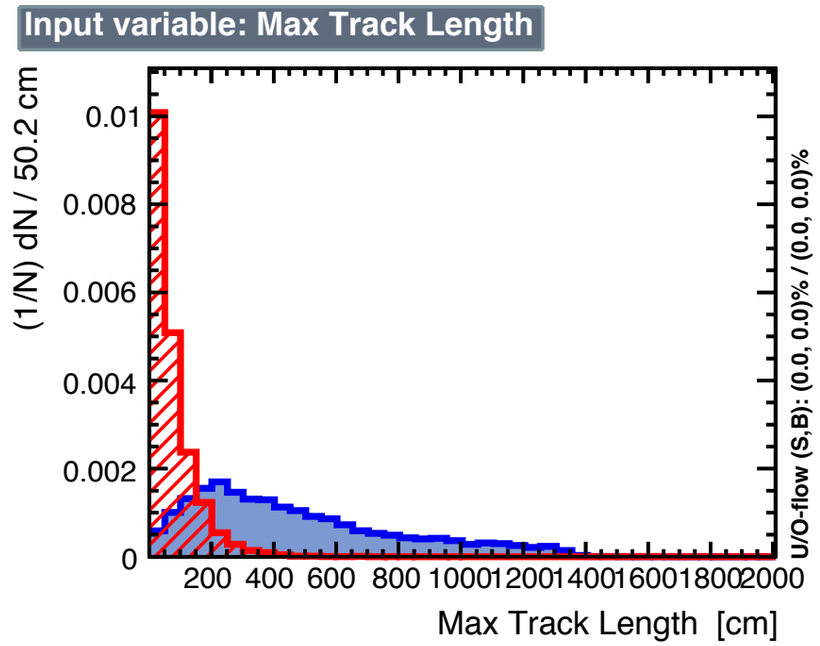
BDT Inputs

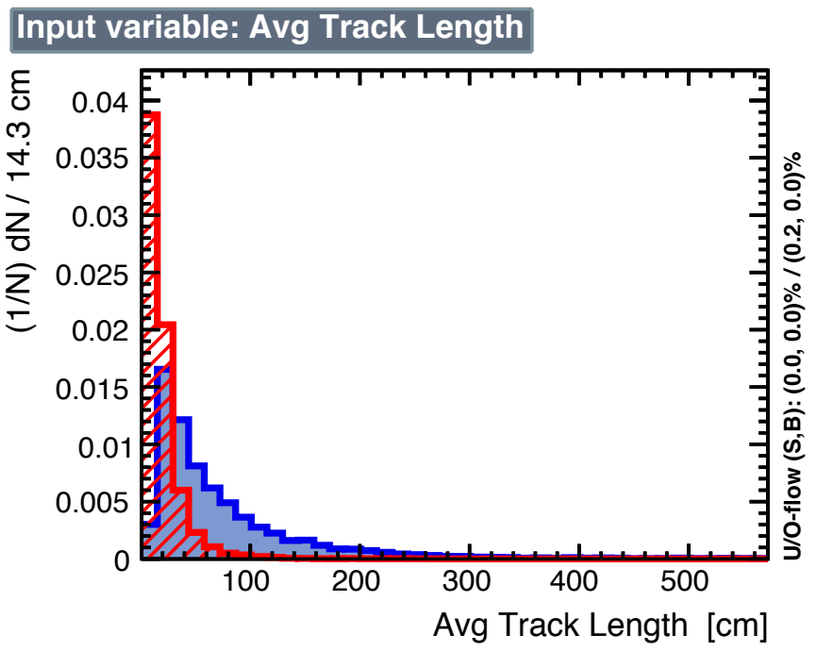
- Total collection plane hit charge
- Number of tracks
- Maximum track length
- Average track length
- Longest track (d)E/(d)x
- Signal fluctuation
  - Q1/Q2 where Q1 (Q2) is the sum of the top (bottom) 50% of wire charge
- Transverse track profile
  - Fraction of charge within 200 ticks of longest track
  - Fraction of charge on longest track
  - Longest track PIDA
  - Maximum fraction of charge in 5, 10, 50 and 100 wires
  - Direction cosines of longest track
  - Fractional transverse energy

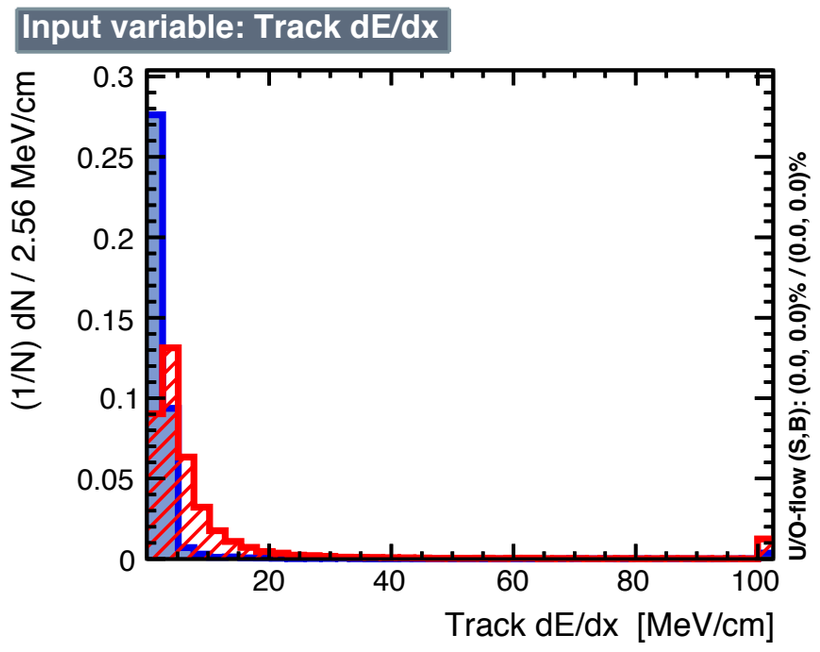
BDT Inputs

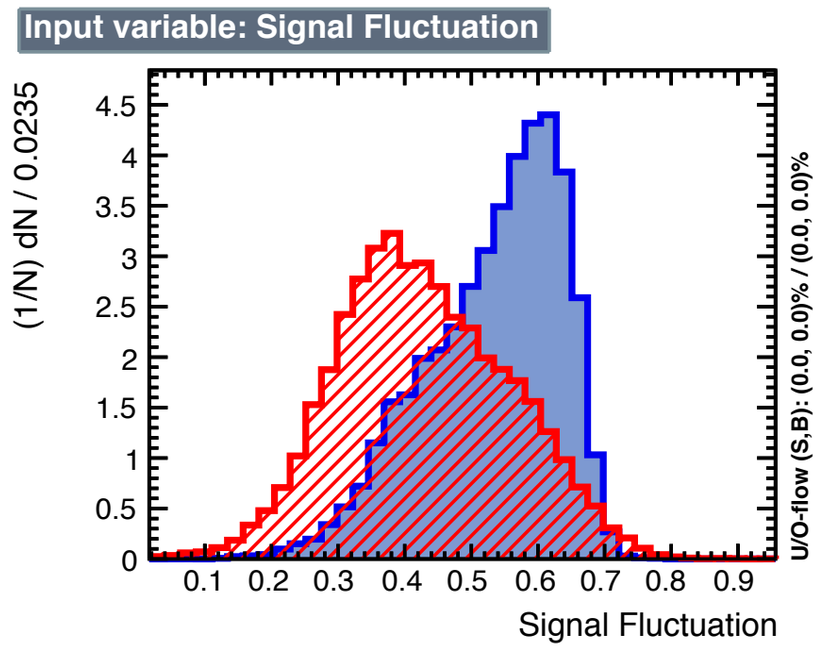

Input variable: Event Charge
Signal
Background
(1/N) dN / 6.03e+04 ADC
Event Charge [ADC]
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%


Input variable: Number of Tracks
(1/N) dN / 3.2
Number of Tracks
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%

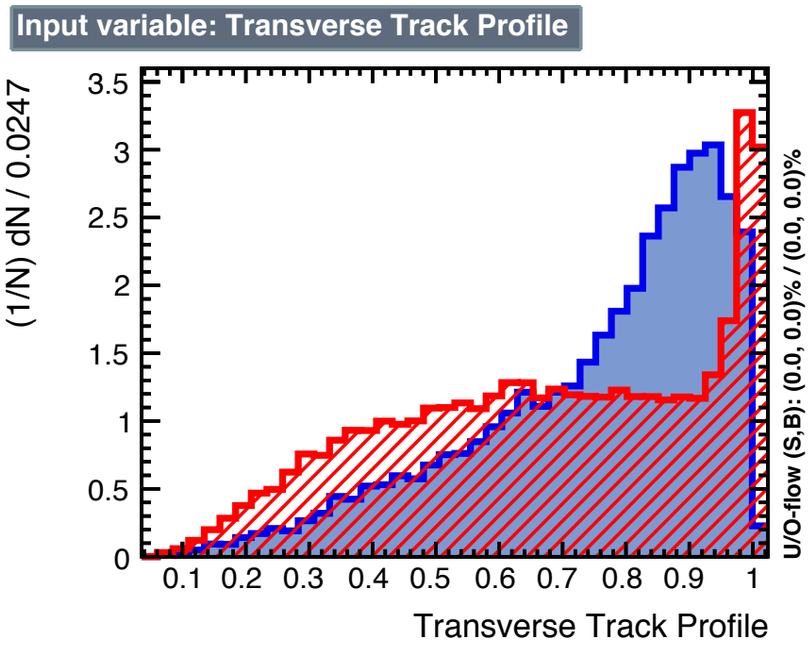

Input variable: Max Track Length
(1/N) dN / 50.2 cm
Max Track Length [cm]
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%

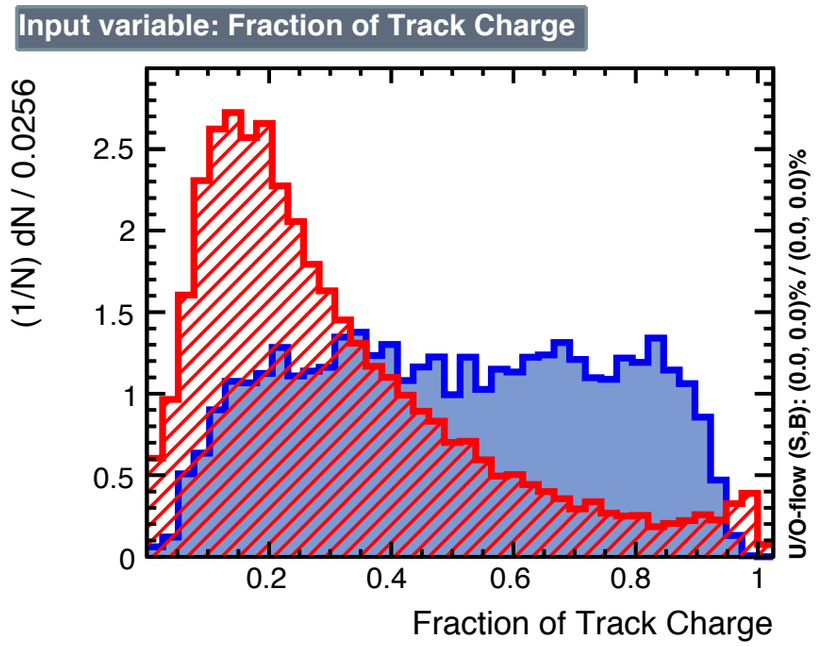

Input variable: Avg Track Length
(1/N) dN / 14.3 cm
Avg Track Length [cm]
U/O-flow (S,B): (0.0, 0.0)% / (0.2, 0.0)%

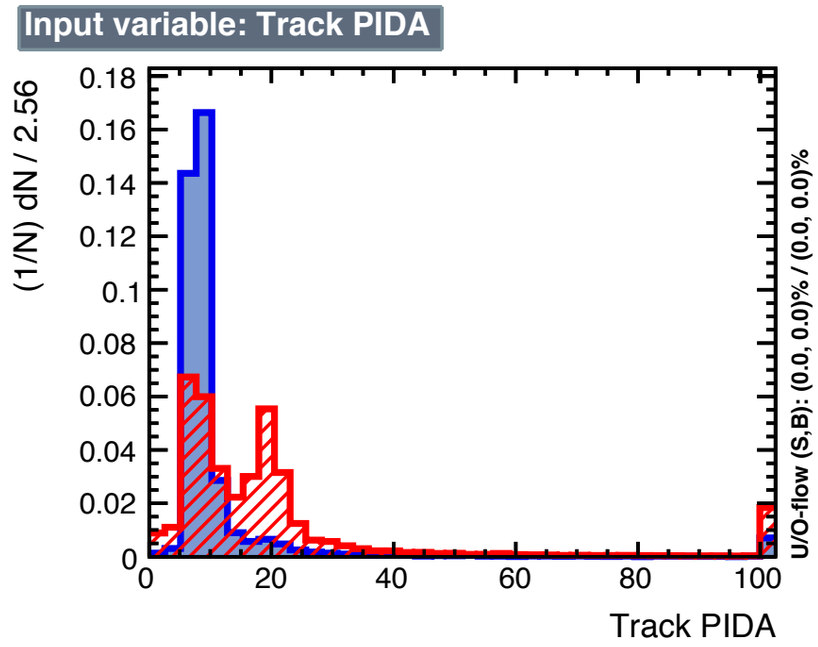

Input variable: Track dE/dx
(1/N) dN / 2.56 MeV/cm
Track dE/dx [MeV/cm]
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%

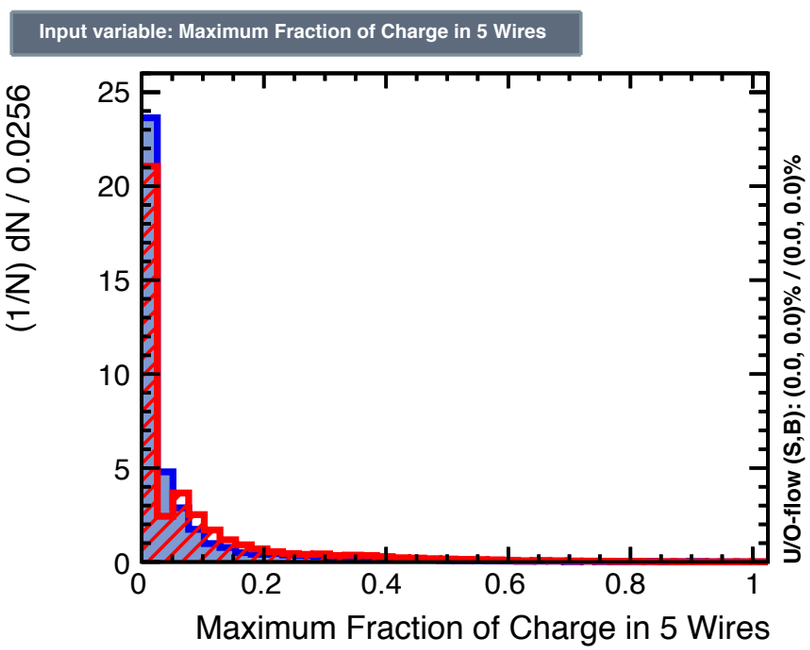

Input variable: Signal Fluctuation
(1/N) dN / 0.0235
Signal Fluctuation
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%

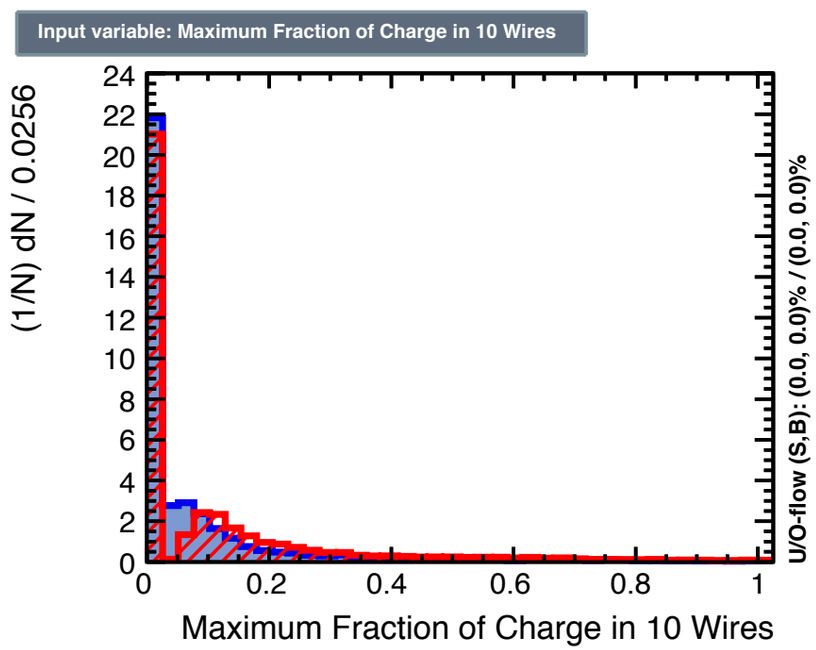
BDT Inputs

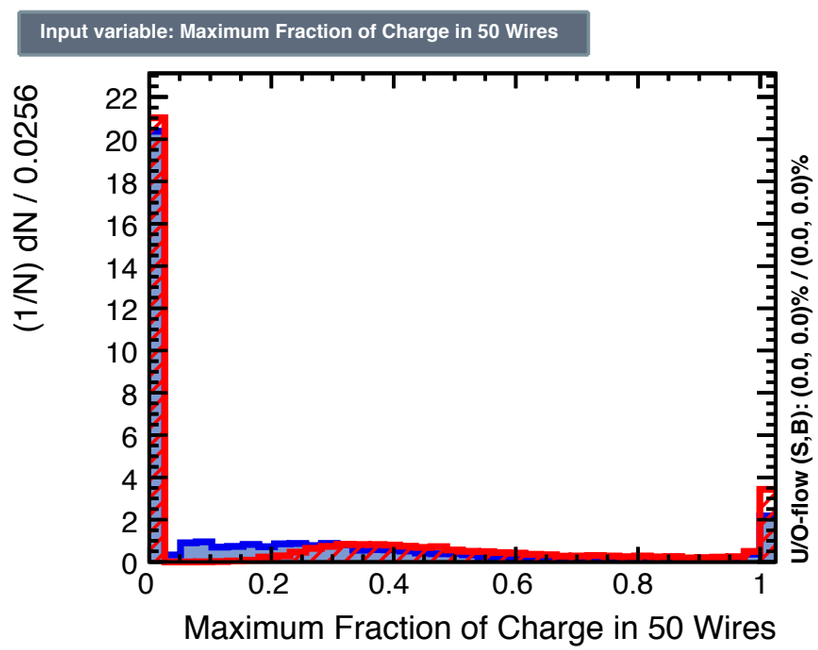

Input variable: Transverse Track Profile
(1/N) dN / 0.0247
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Transverse Track Profile


Input variable: Fraction of Track Charge
(1/N) dN / 0.0256
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Fraction of Track Charge

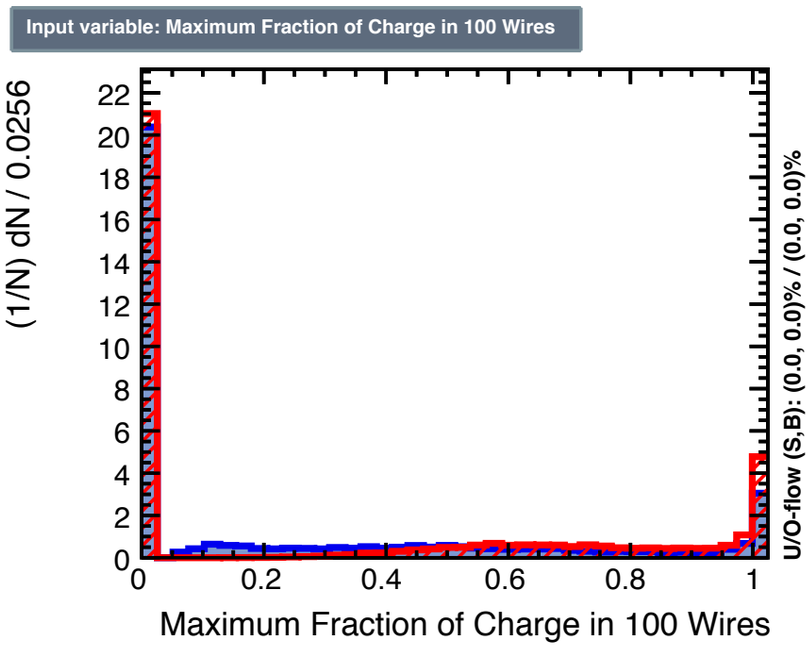

Input variable: Track PIDA
(1/N) dN / 2.56
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Track PIDA

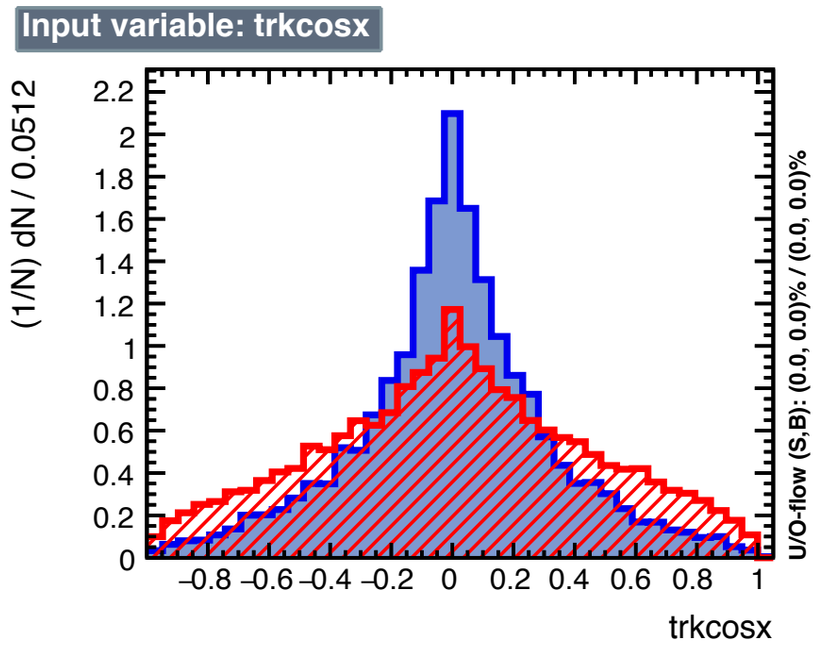

Input variable: Maximum Fraction of Charge in 5 Wires
(1/N) dN / 0.0256
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Maximum Fraction of Charge in 5 Wires

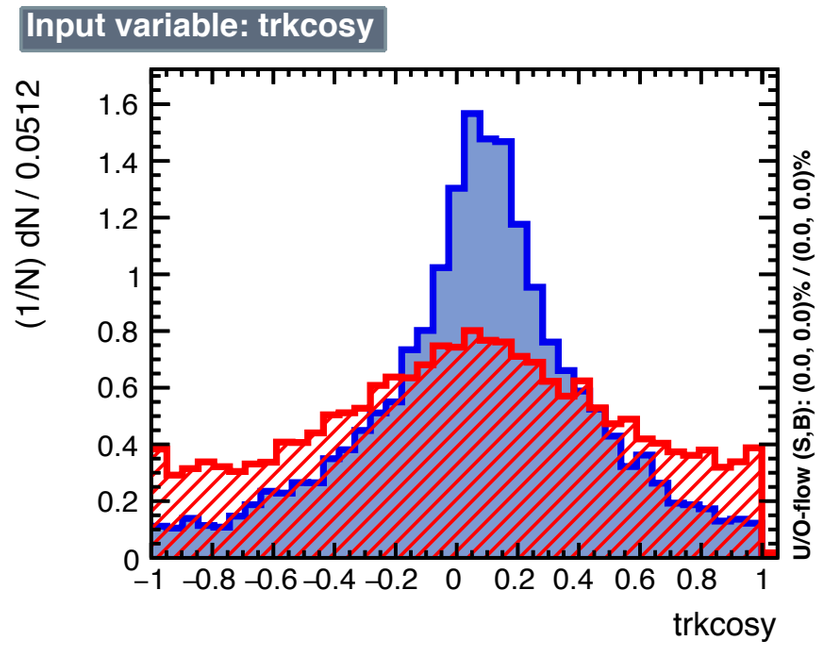

Input variable: Maximum Fraction of Charge in 10 Wires
(1/N) dN / 0.0256
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Maximum Fraction of Charge in 10 Wires

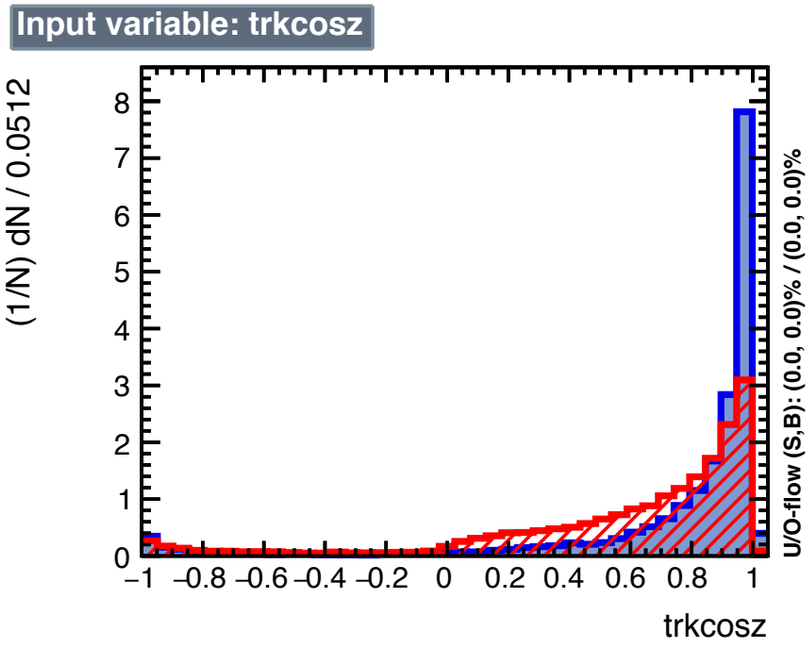

Input variable: Maximum Fraction of Charge in 50 Wires
(1/N) dN / 0.0256
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Maximum Fraction of Charge in 50 Wires

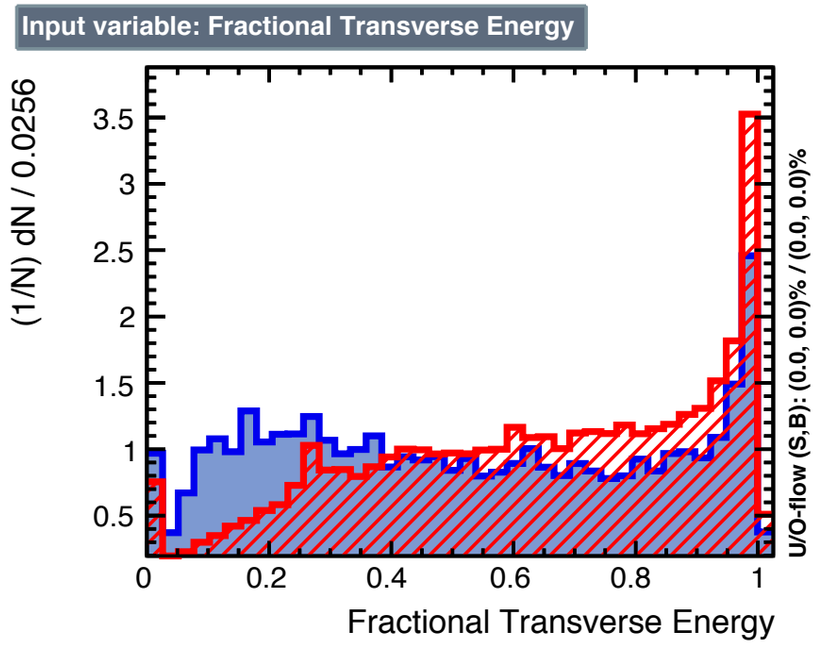
BDT Inputs


Input variable: Maximum Fraction of Charge in 100 Wires
(1/N) dN / 0.0256
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Maximum Fraction of Charge in 100 Wires


Input variable: trkcosx
(1/N) dN / 0.0512
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
trkcosx


Input variable: trkcosy
(1/N) dN / 0.0512
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
trkcosy

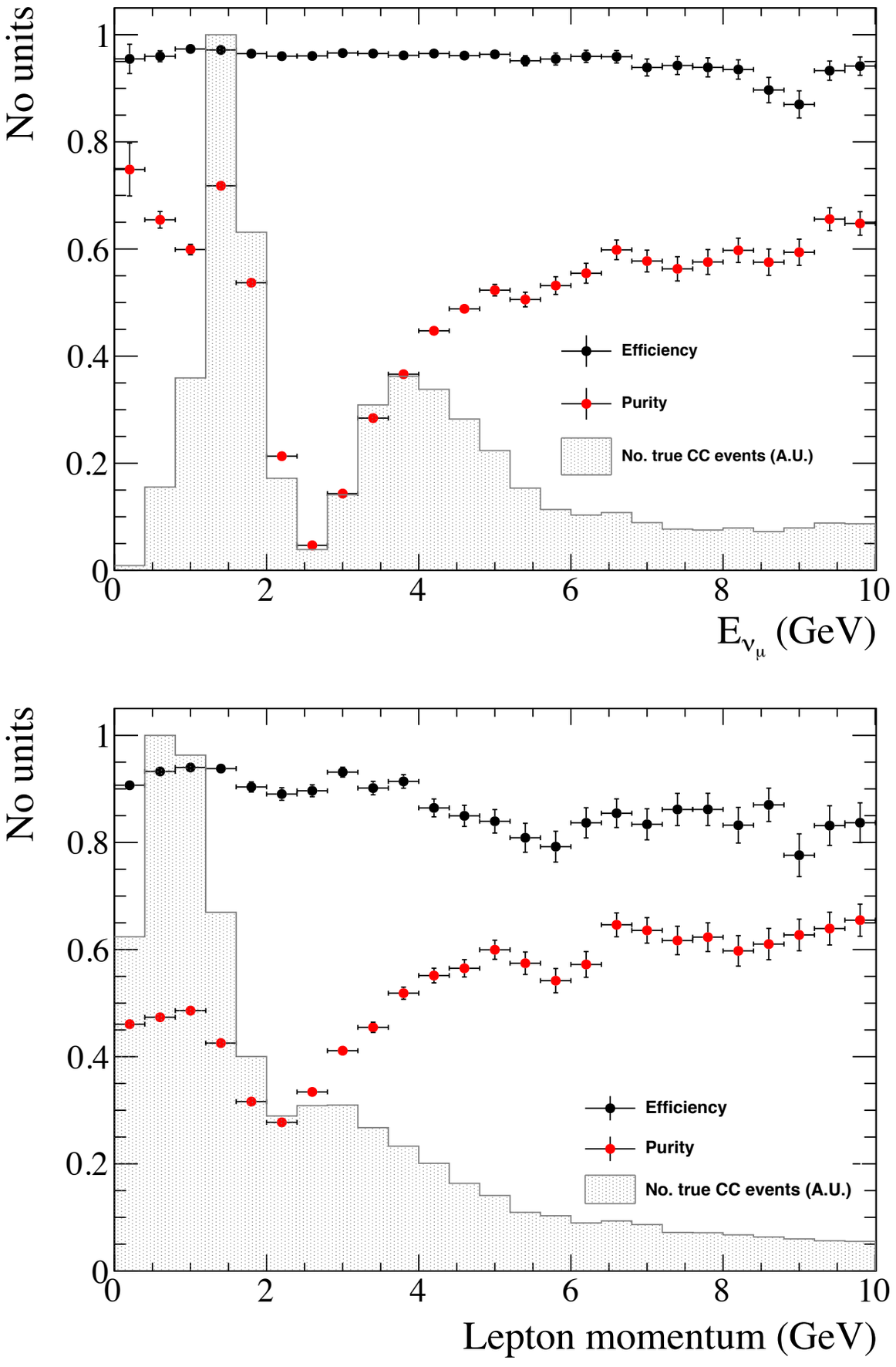

Input variable: trkcosz
(1/N) dN / 0.0512
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
trkcosz

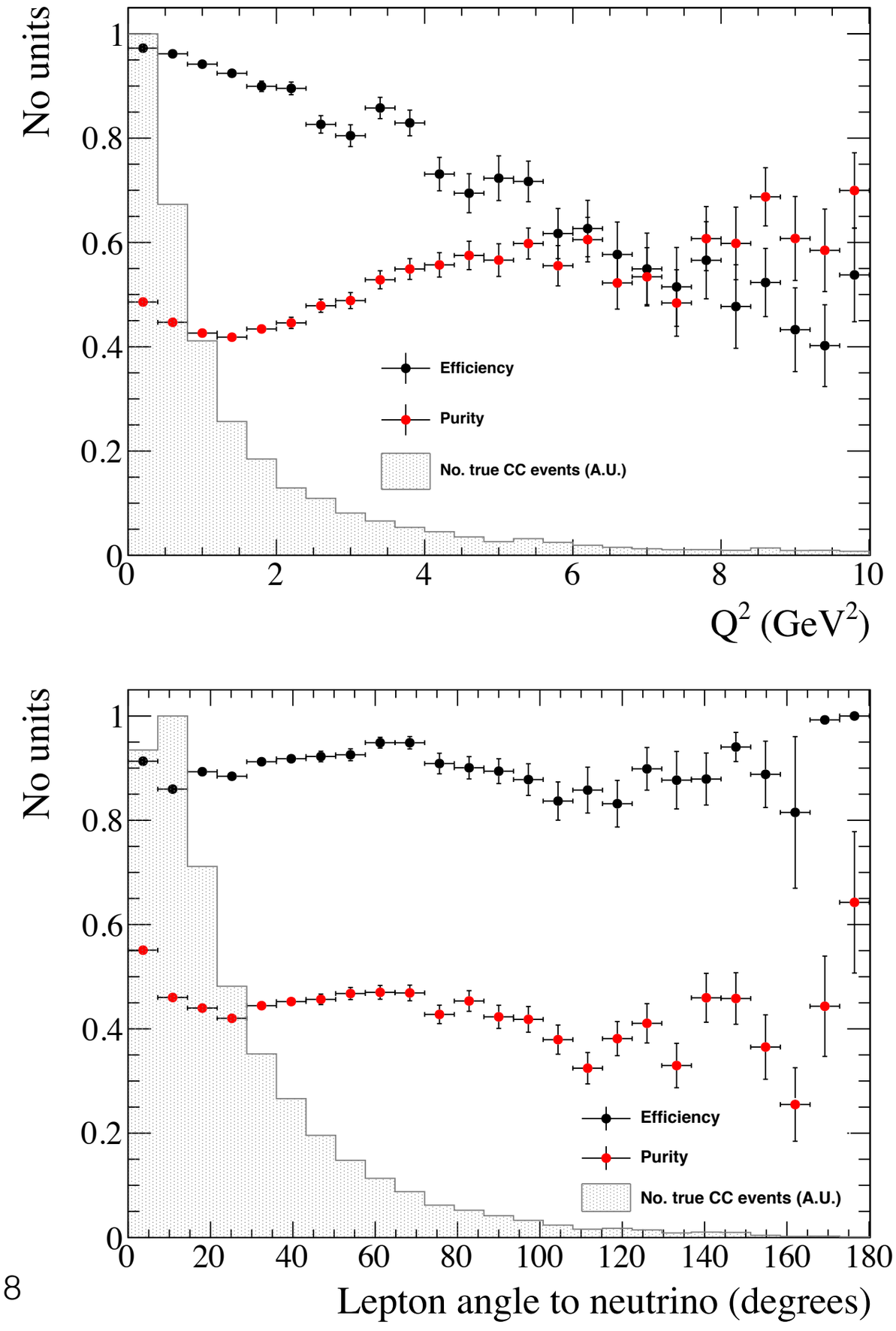

Input variable: Fractional Transverse Energy
(1/N) dN / 0.0256
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
Fractional Transverse Energy

What's been going on

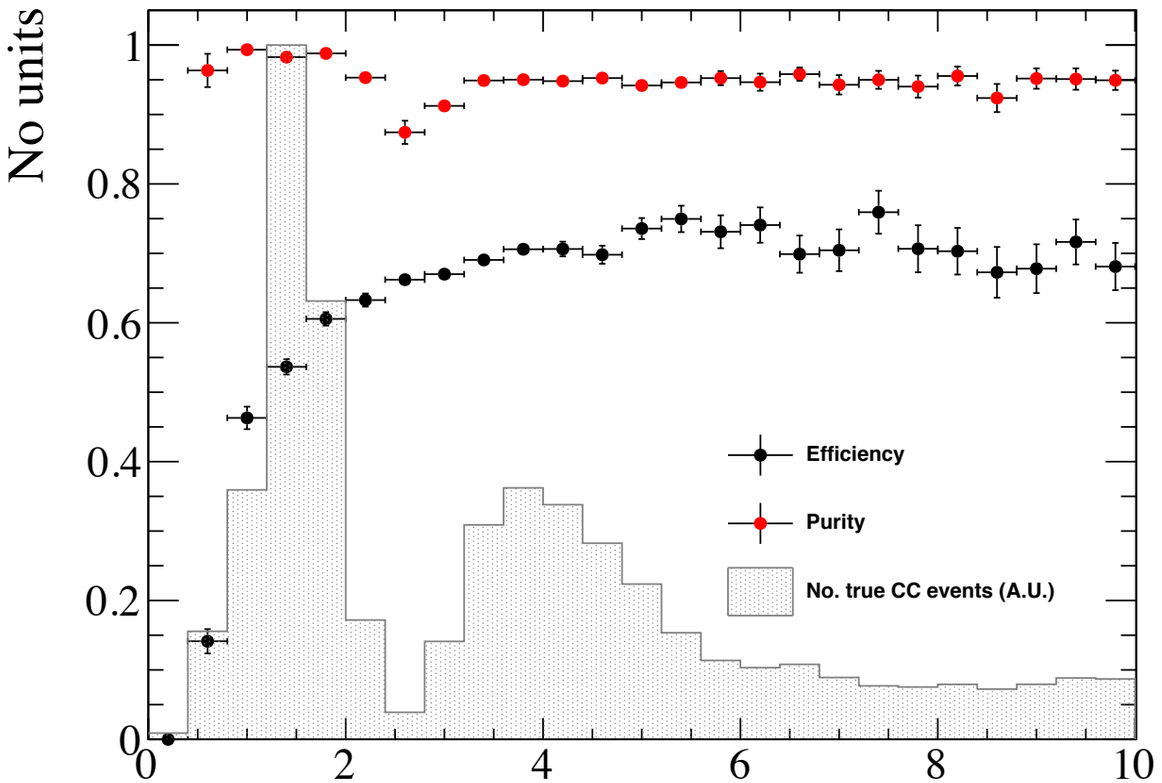
- Moved to dunetpc v06_11_00 (generated my own samples for testing)
- Under the hood changes
- Automated POT accounting
- **Characterising the current selection**
  - Efficiencies, purities, tunings etc.

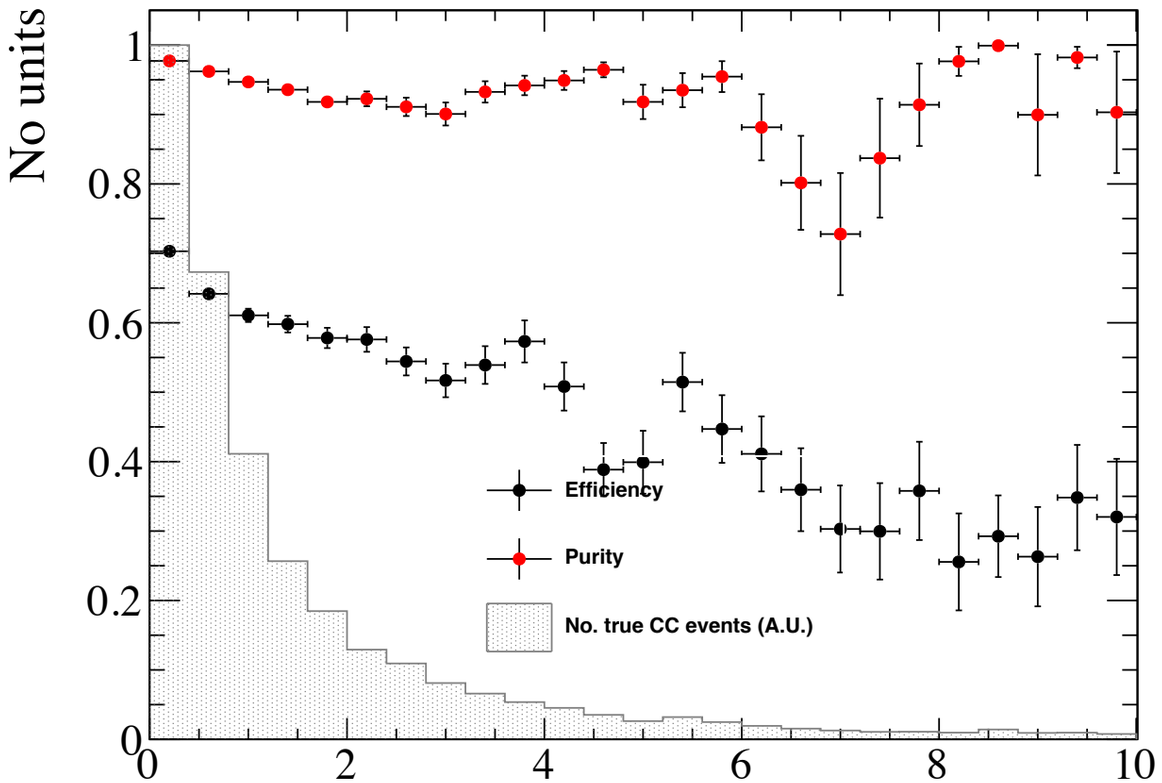
Efficiency and purity (before MVA cut)





Efficiency and purity (MVA > 0.8)



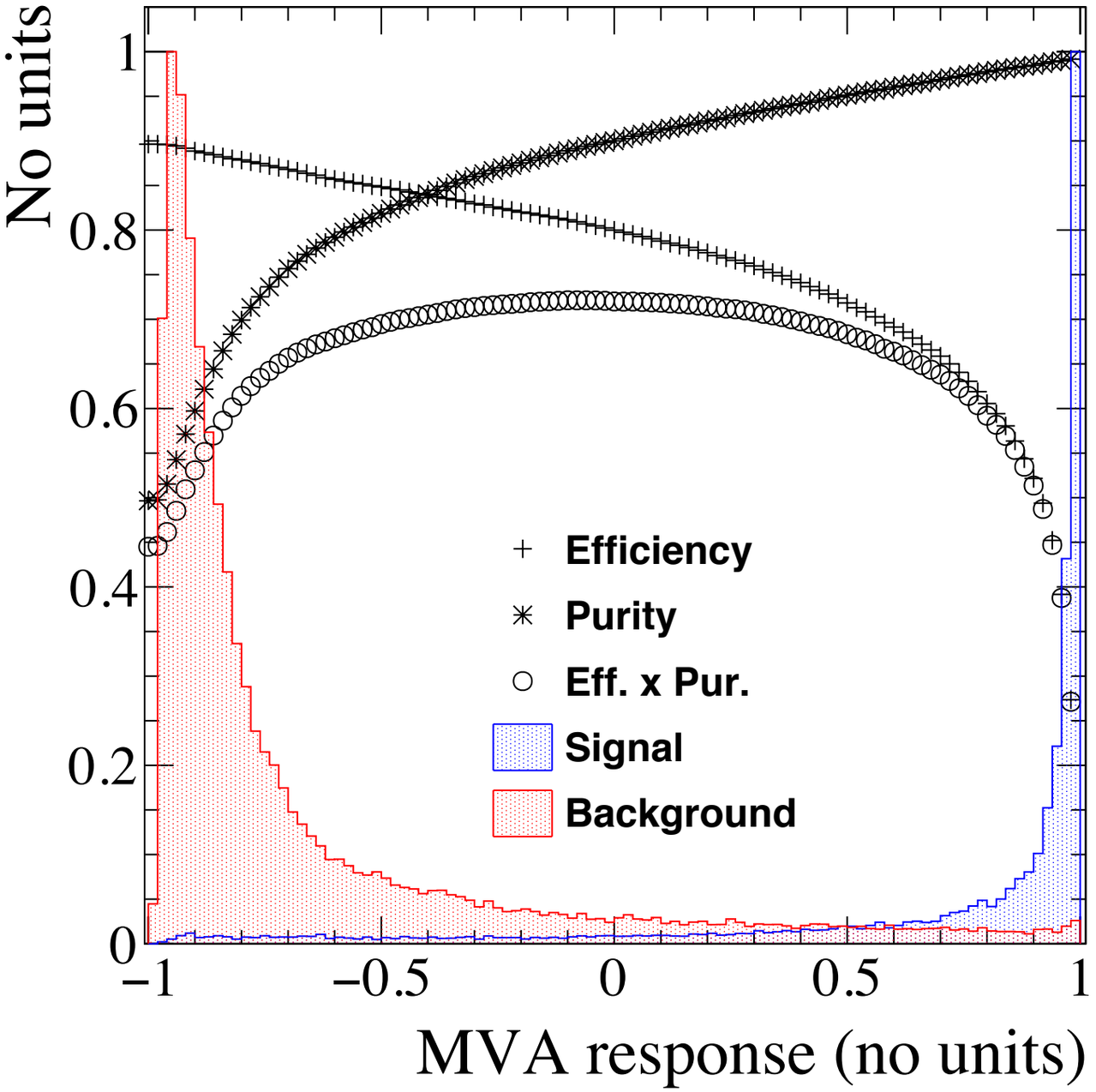


Efficiency and purity

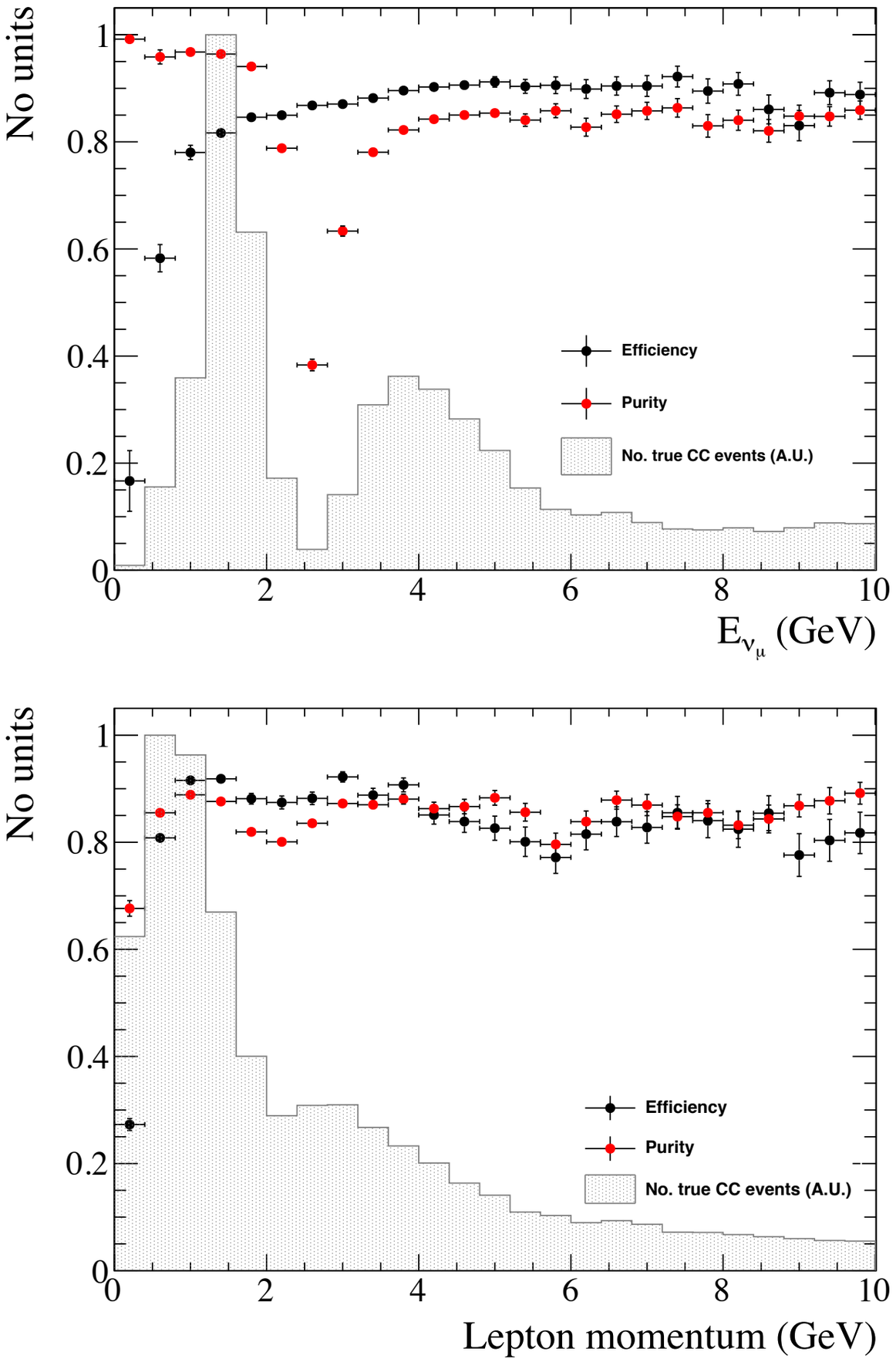
- The selection is quite biased in the four key distributions
- There is also a smaller bias in Q^2 before the selection takes place
    - Only requirement is that the reconstructed vertex is in the fiducial volume
- Two approaches to combatting the biases:
    - Tune the MVA cut
    - Remove variables from the MVA

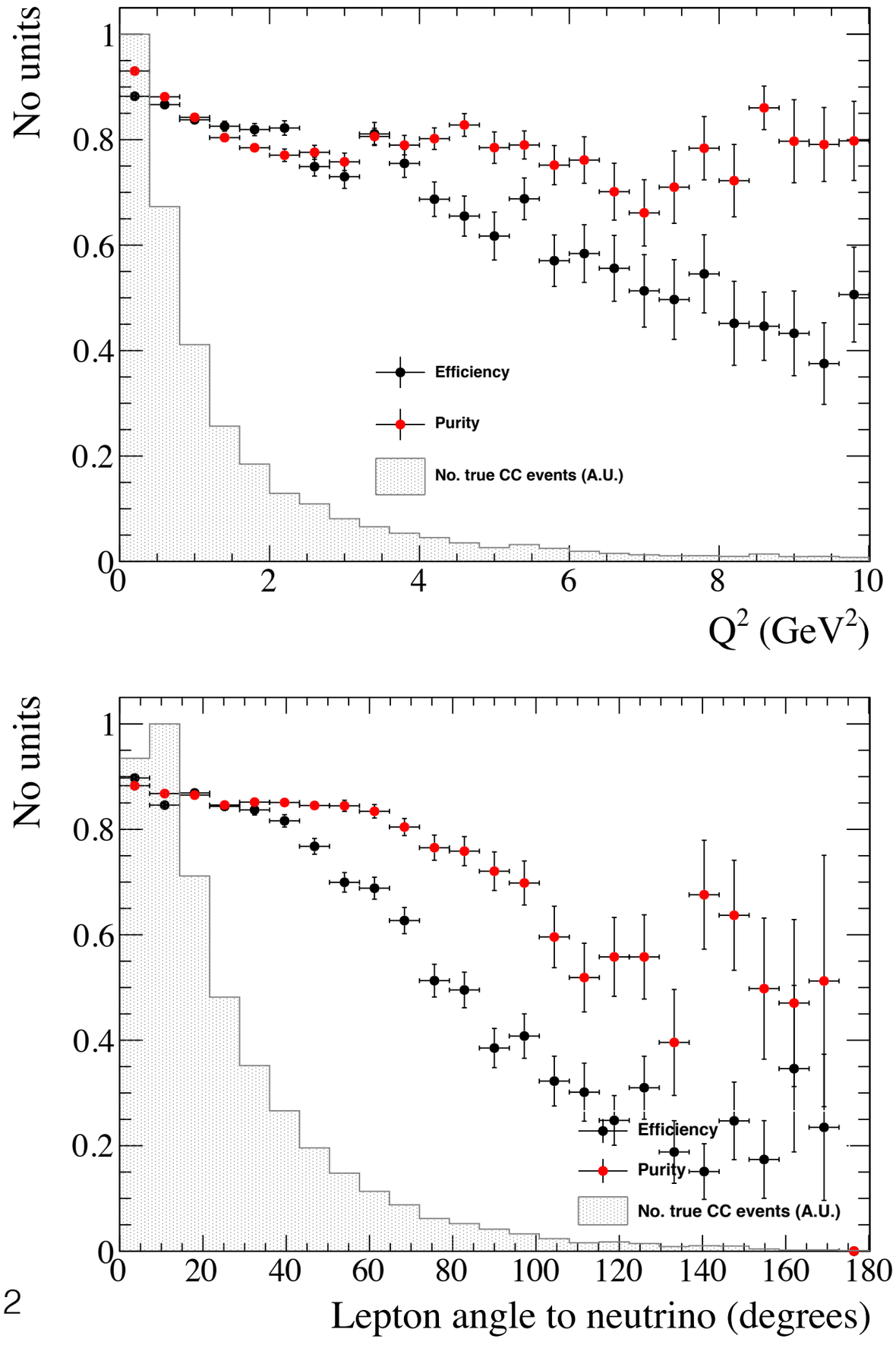
MVA cut tuning

- Opted for tuning eff. x pur.
- The eff. x pur. metric is shown on the right (open circles)
- Distribution peaks at mva==-0.1
  - Efficiency: 81.1%
  - Purity: 88.9%
  - Eff. x Pur: 0.72
- This cut value is a lot more forgiving than 0.8…



Efficiency and purity (MVA > -0.1)





Efficiency and purity after tuning the MVA cut

- Efficiency has improved from ~61% to 81% (+20%)
- The sample purity has degraded from ~95% to 89% (-6%)
- The bias remains but is much improved
- I am currently investigating the BDT input variables to pick out where the bias is coming from

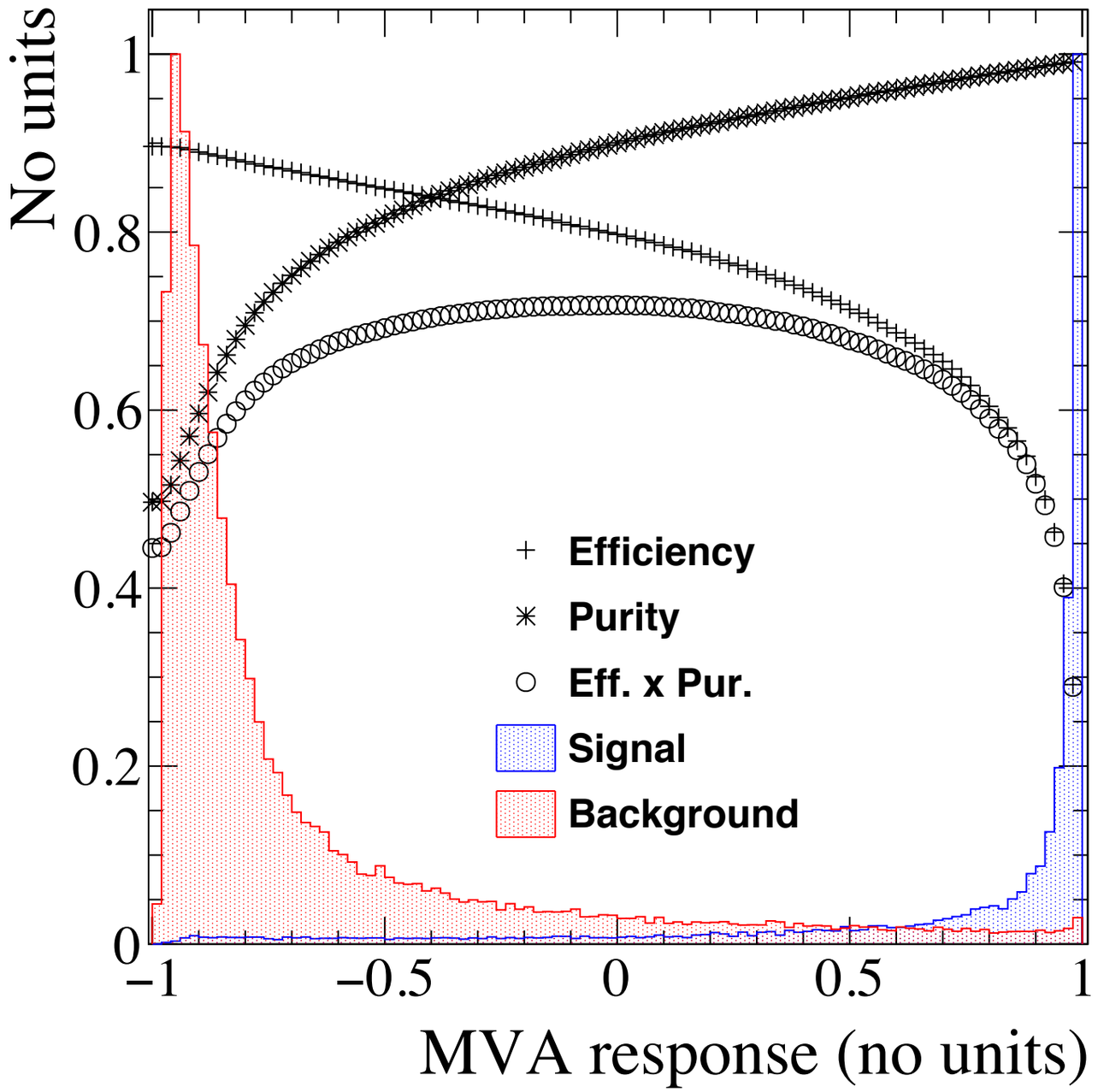
BDT Inputs (bias?)

- Total collection plane hit charge
- Number of tracks
- Maximum track length
- Average track length
- Longest track (d)E/(d)x
- Signal fluctuation
  - Q1/Q2 where Q1 (Q2) is the sum of the top (bottom) 50% of wire charge
- Transverse track profile
  - Fraction of charge within 200 ticks of longest track
  - Fraction of charge on longest track
  - Longest track PIDA
  - Maximum fraction of charge in 5, 10, 50 and 100 wires
  - Direction cosines of longest track
  - Fractional transverse energy

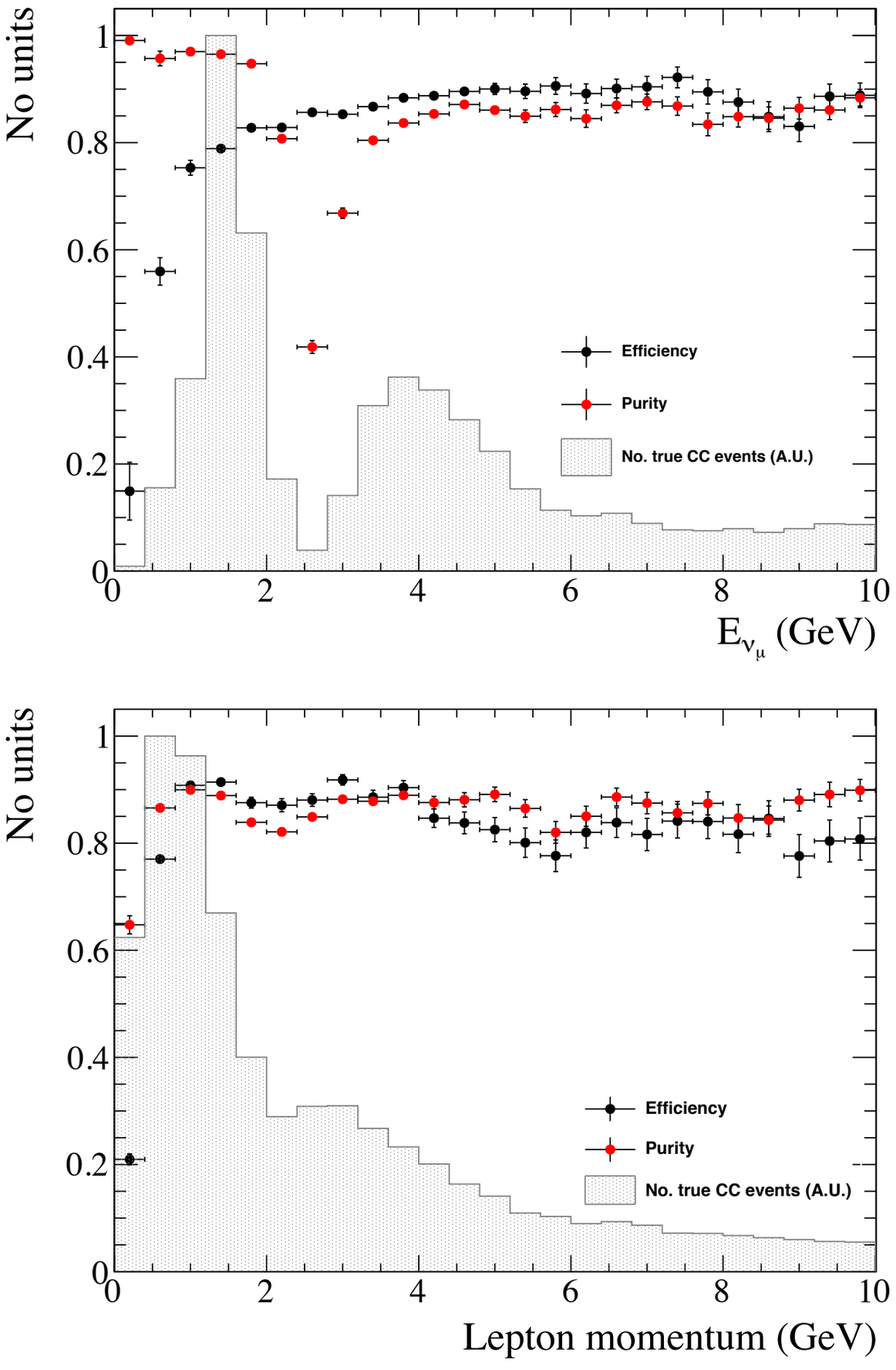
Remove, retrain, retune and rerun

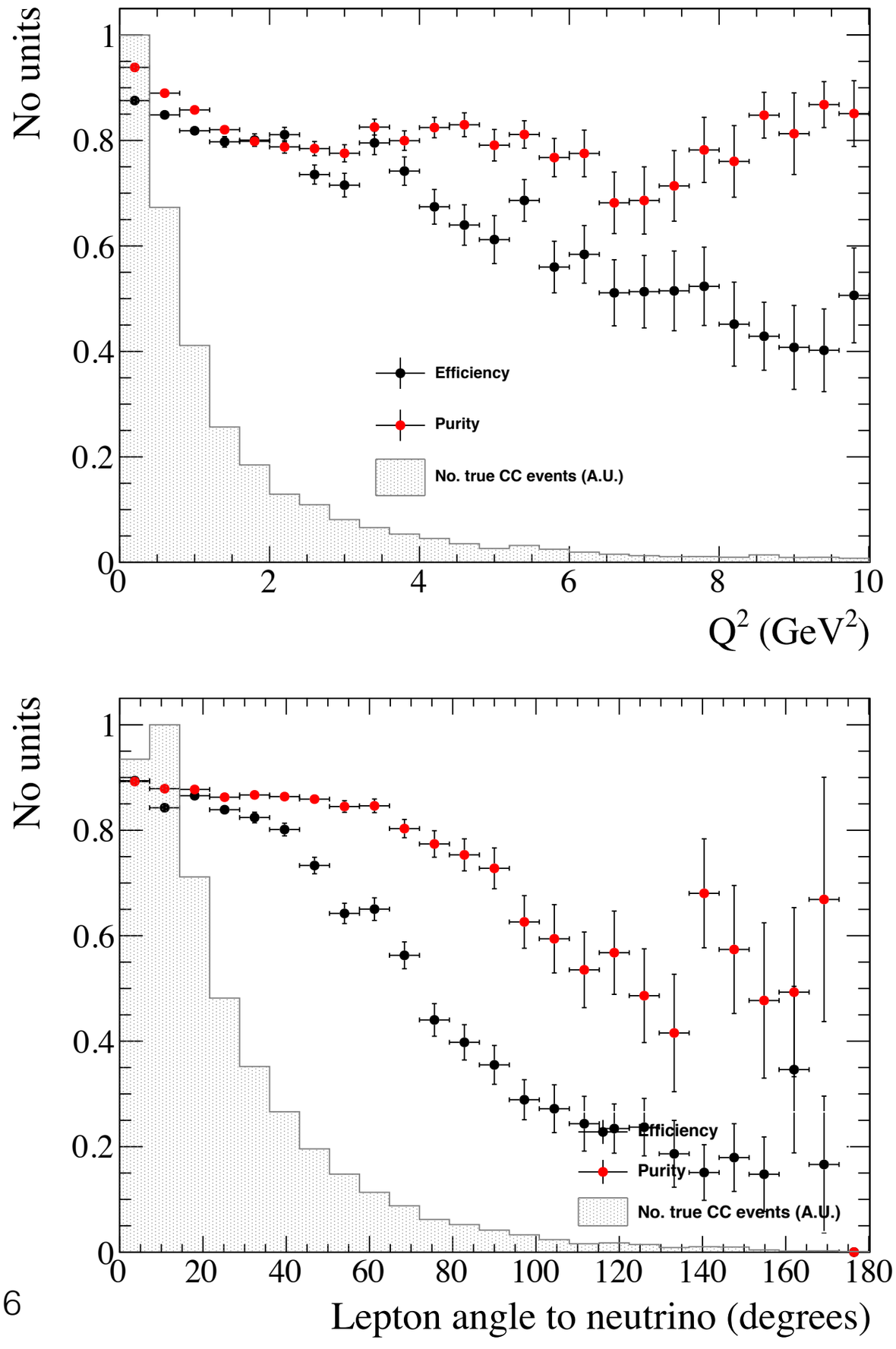
Retuning the MVA

- Distribution peaks at mva==0.0
  - Efficiency: 79.8%
  - Purity: 90.1%
  - Eff. x pur: 0.712
- Removing the three track angle variables has resulted in ~1% level changes



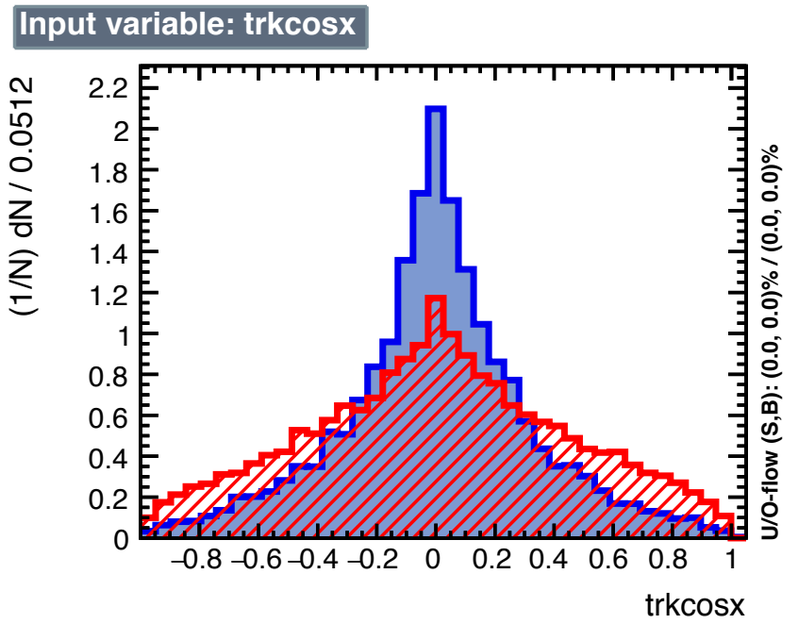
Efficiency and purity (MVA (no track cosine input) > 0)

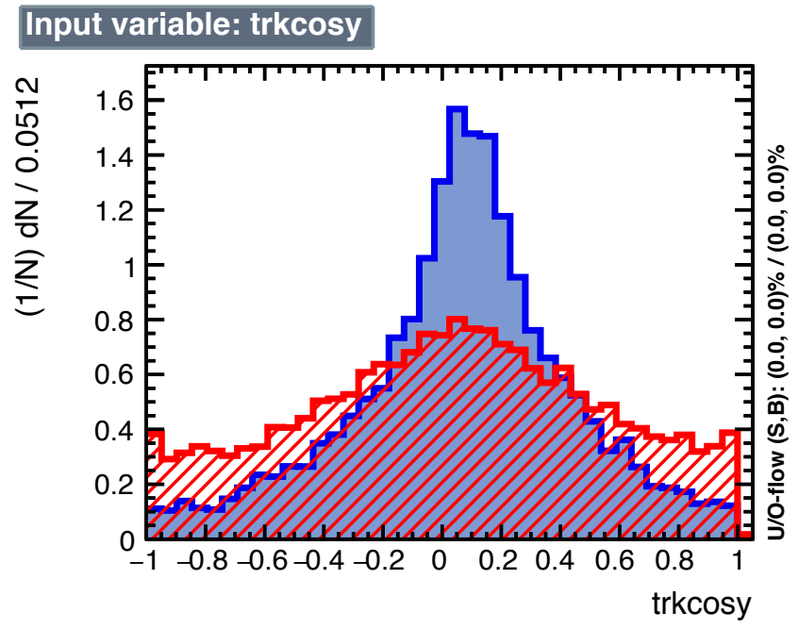


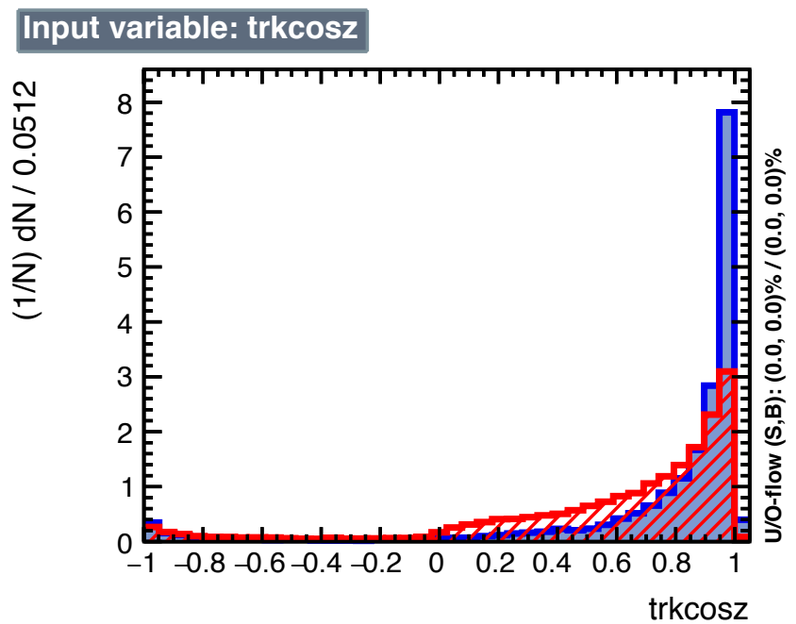


Removal of track angle cosines

- Essentially made no difference not only to the bias but also to the selection performance
- This actually is not that surprising when considering the variable separation (see right plots)
- Rather than guessing where the bias hides, I think a binary search-like removal of the variables is the most optimum way to find the offending variable(s)
- Given that the outgoing lepton angle is a physics observable, should we be cutting on it?







Summary

- I have taken over the vast amount of work and effort put in by Tyler and Tingjun

- The selection has now been characterised and it's evident there are some biases in key variables (notably neutrino energy)

- The selection has now been tuned, resulting in a 20% gain in efficiency and a small 6% loss in purity

  - Tuned mva cut: -0.1 (0.8 before)

  - Selection efficiency: 81.1% (~60% before)

  - Selection purity: 88.9% (~95% before)

- Work to find the biases is ongoing