Lancaster University

ν_μ selection update

Dominic Brailsford

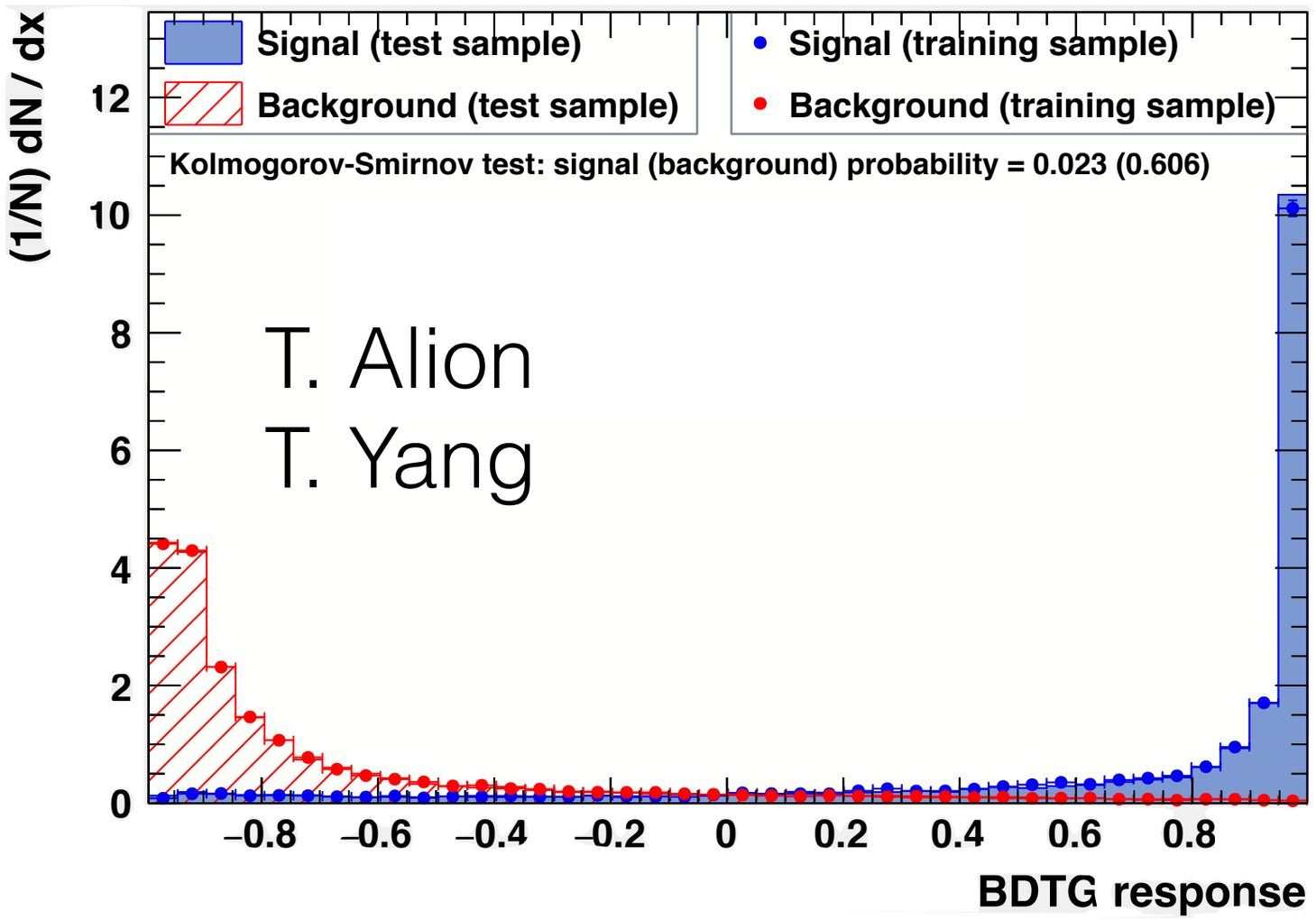
LBL meeting

27/02/17

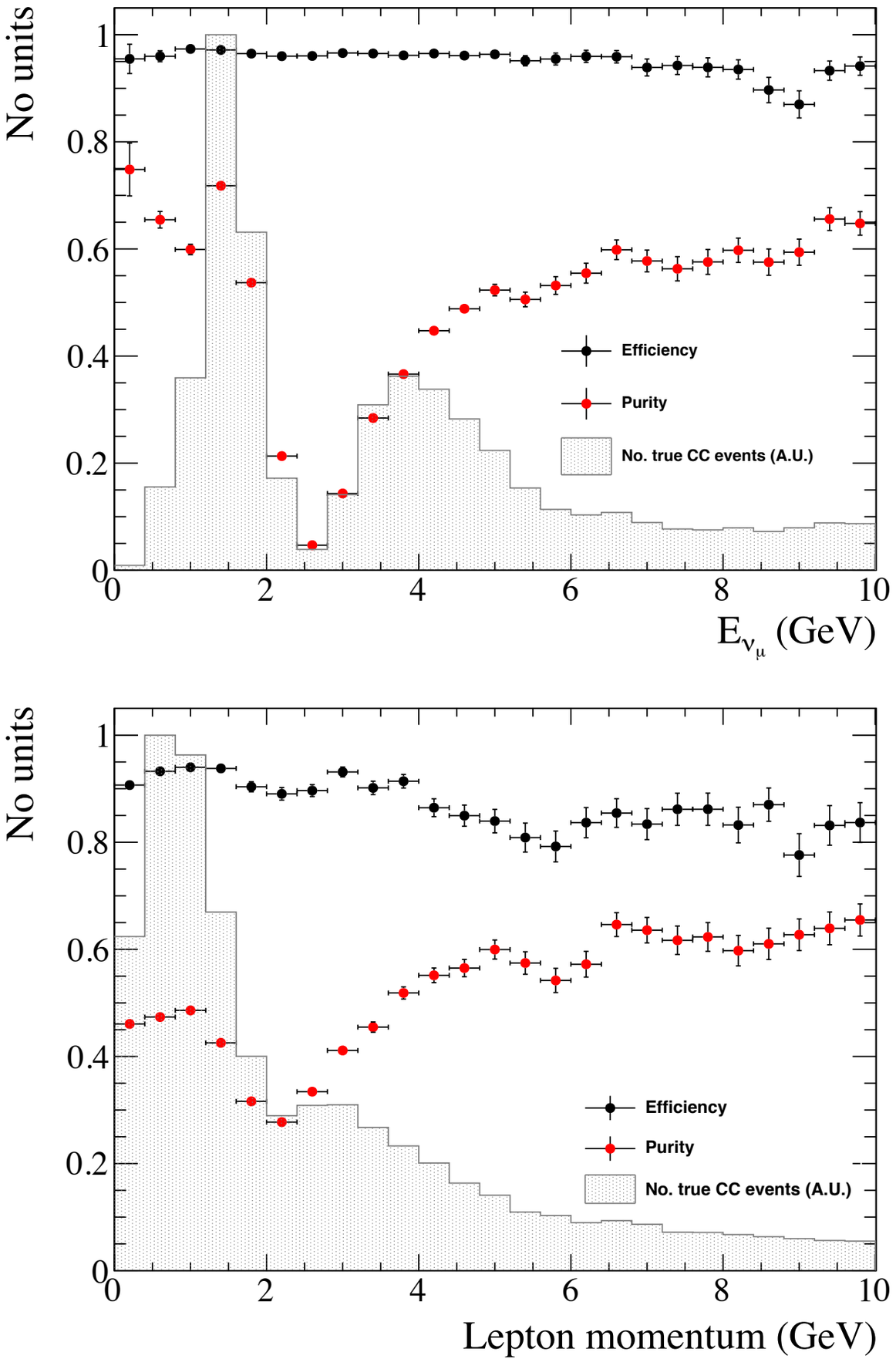
Thanks to R. Patterson and D. Cherdack for the tuning discussion

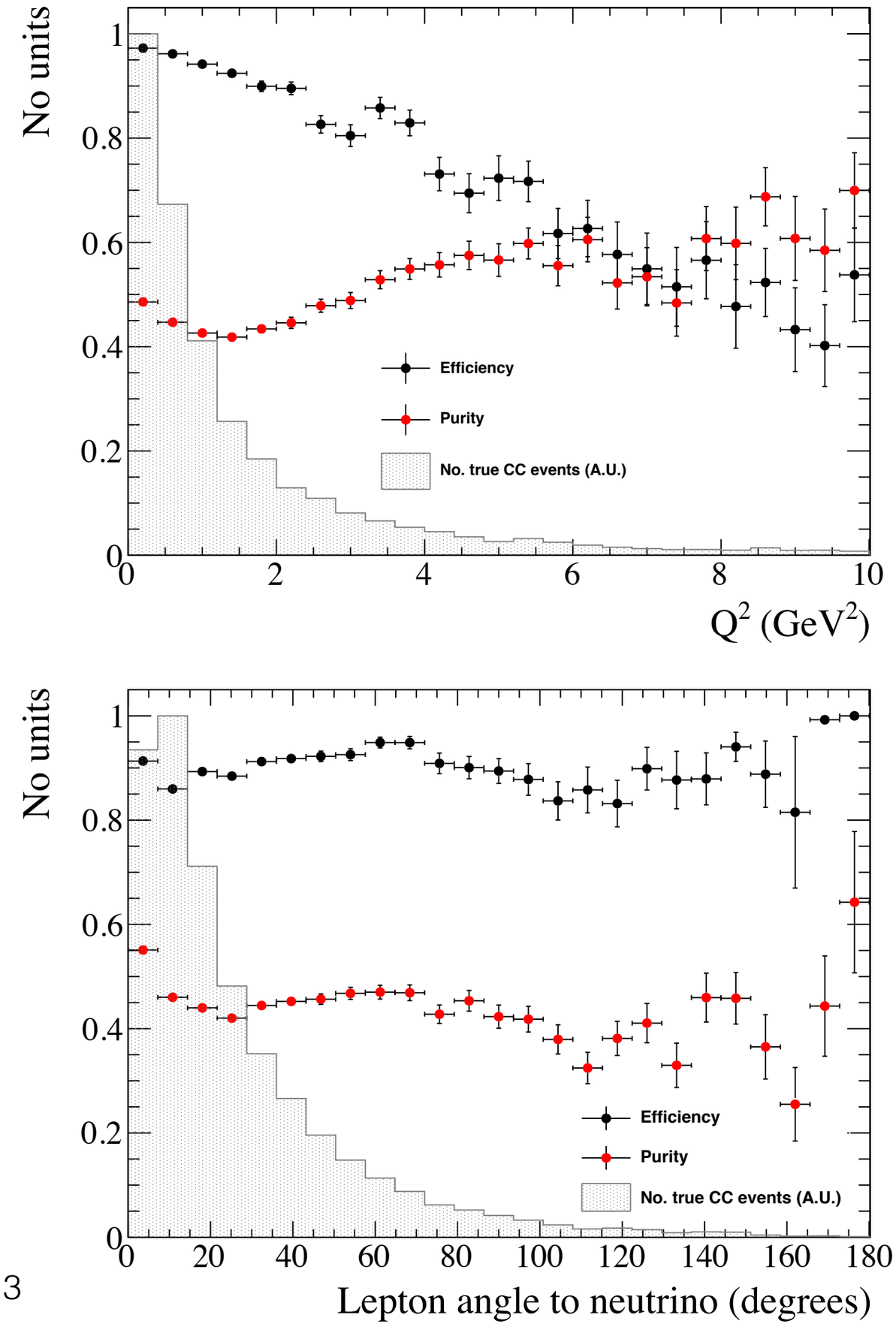
ν_μ Far Detector Event Selection

- Selection developed by T. Alion and T. Yang
- Uses a BDT to select CC ν_μ events. Assesses
  - Event topology
  - Event shape
  - Event charge
- Currently tuned to maximise eff. x pur.
  - Prefers MVA > 0

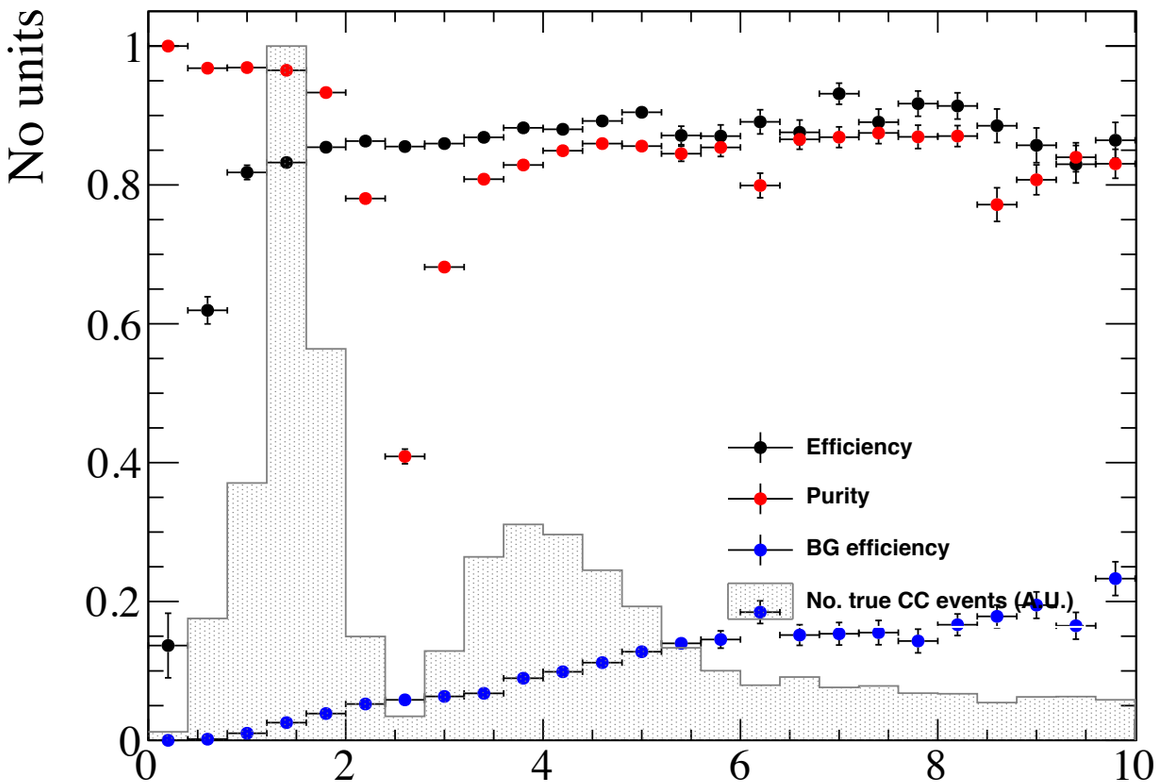


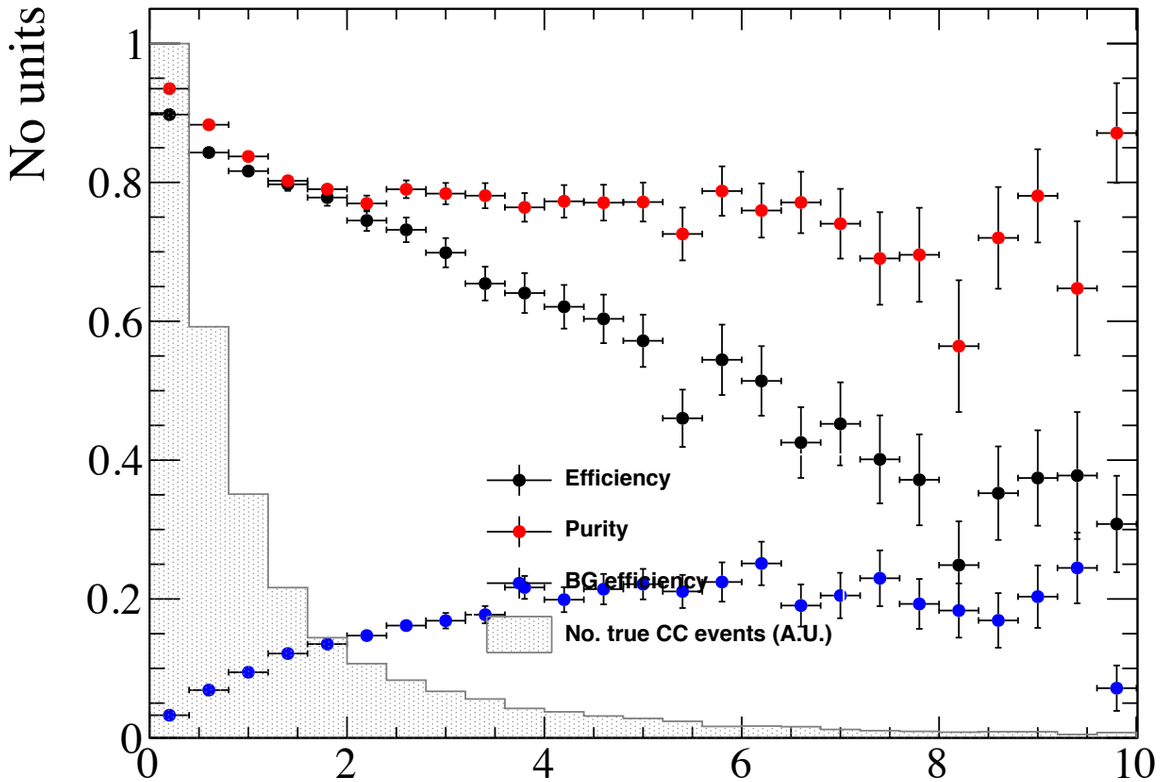
Efficiency and purity (before MVA cut)

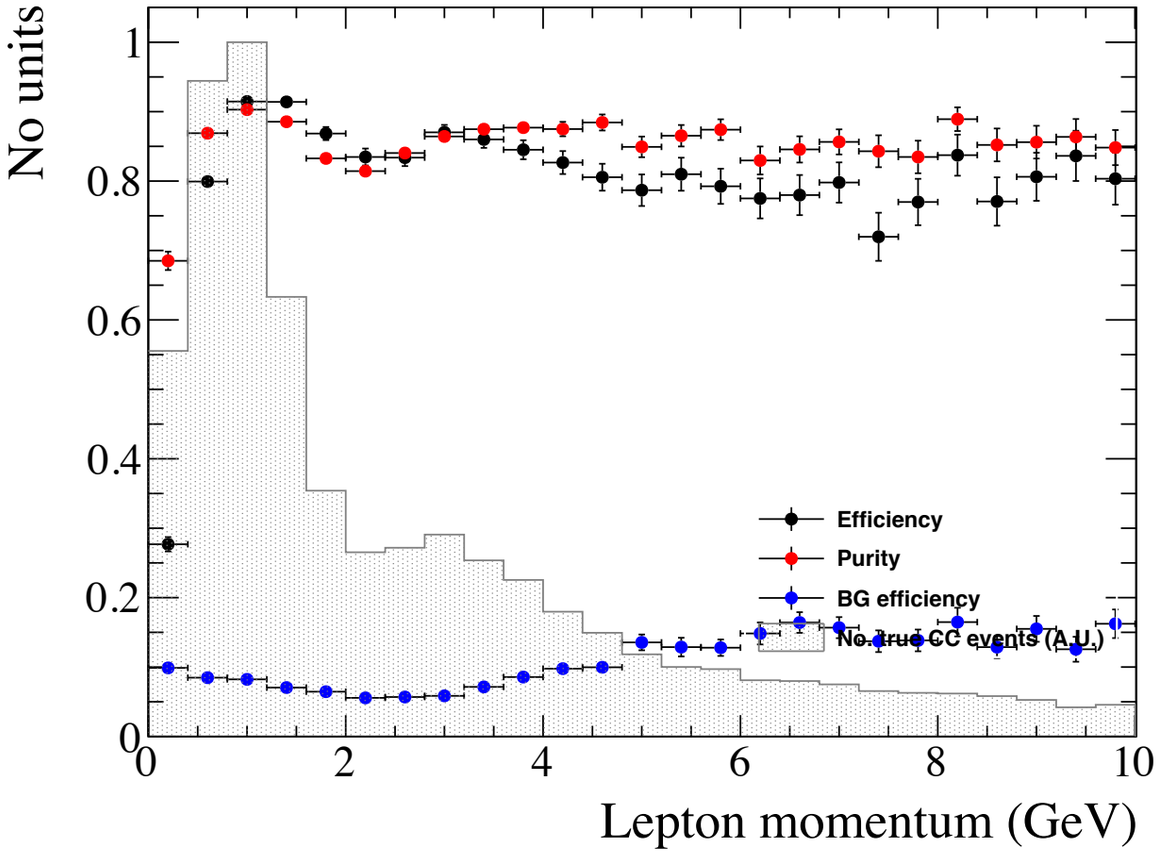


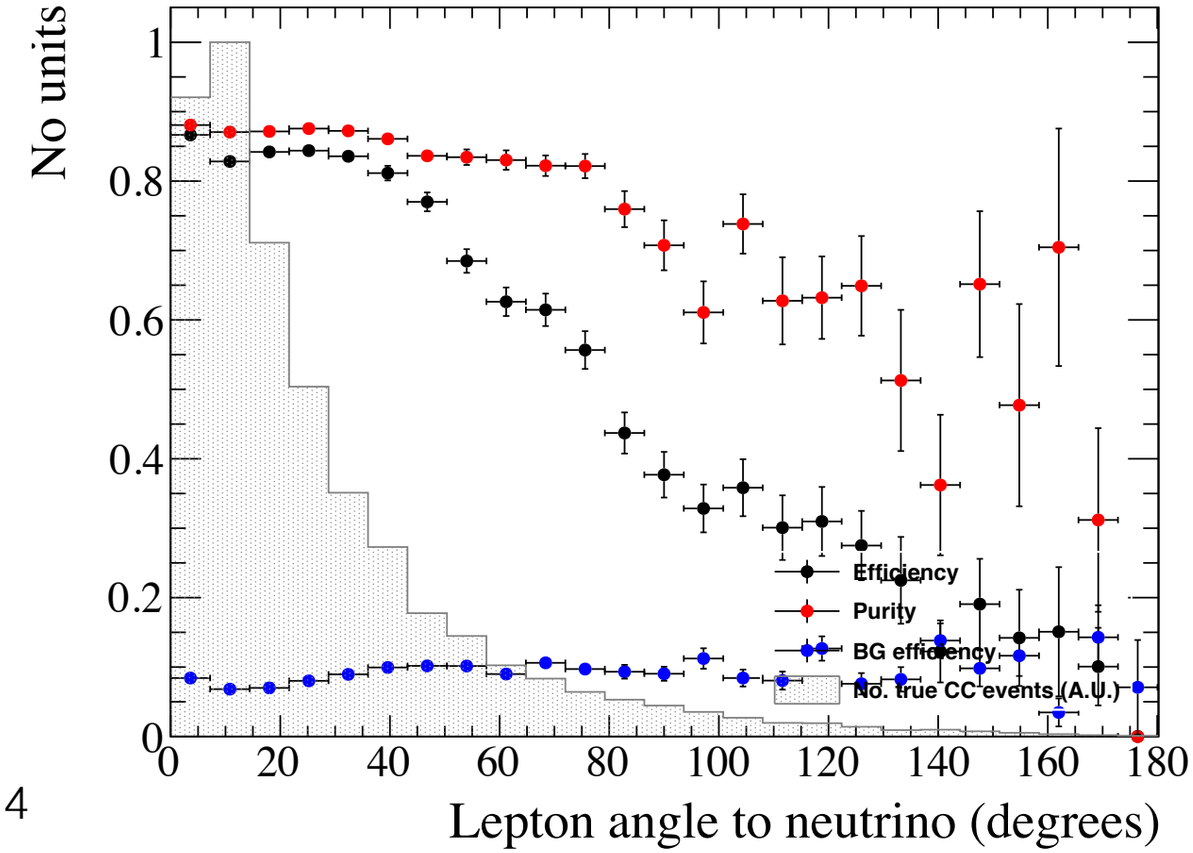


Efficiency and purity (MVA > 0)









Efficiency bias

- We see several cases of bias
- An efficiency drop off at high Q^2 before applying the MVA
- An efficiency drop off for low neutrino energy, low lepton momentum and high lepton angle after applying the MVA
- The bias in Q^2 is unchanged by the MVA cut meaning there are different underlying causes for this and the other biases

MVA with pandora reconstruction

- I've retrained the BDT and reprocessed the sample using pandora, rather than pmtrack

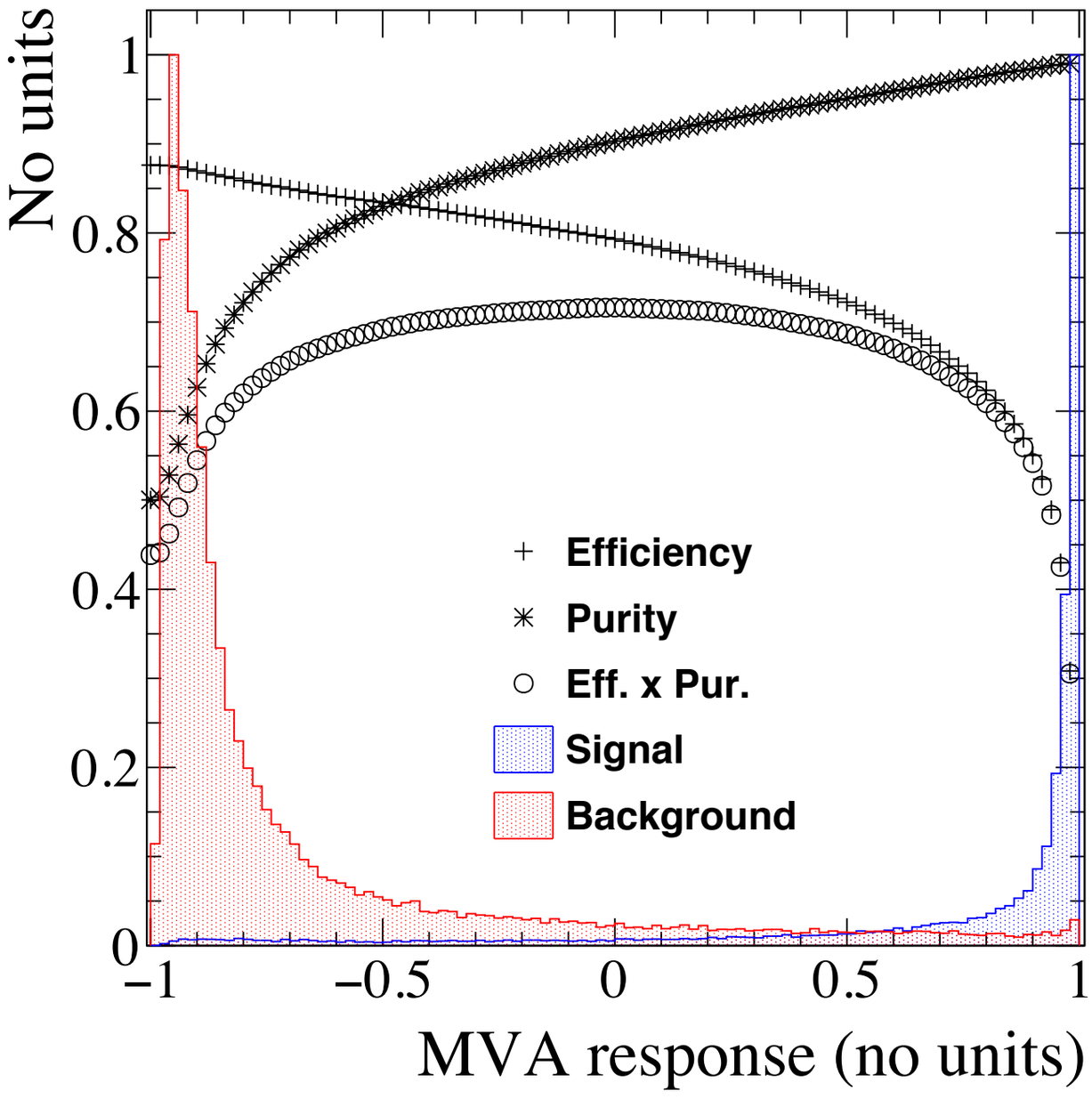
- The input sample is the same one used to make the plots on the previous slides

- In the following set of plots, the ONLY difference to what you have already seen is that a different reconstruction algorithm (and BDT training) has been used

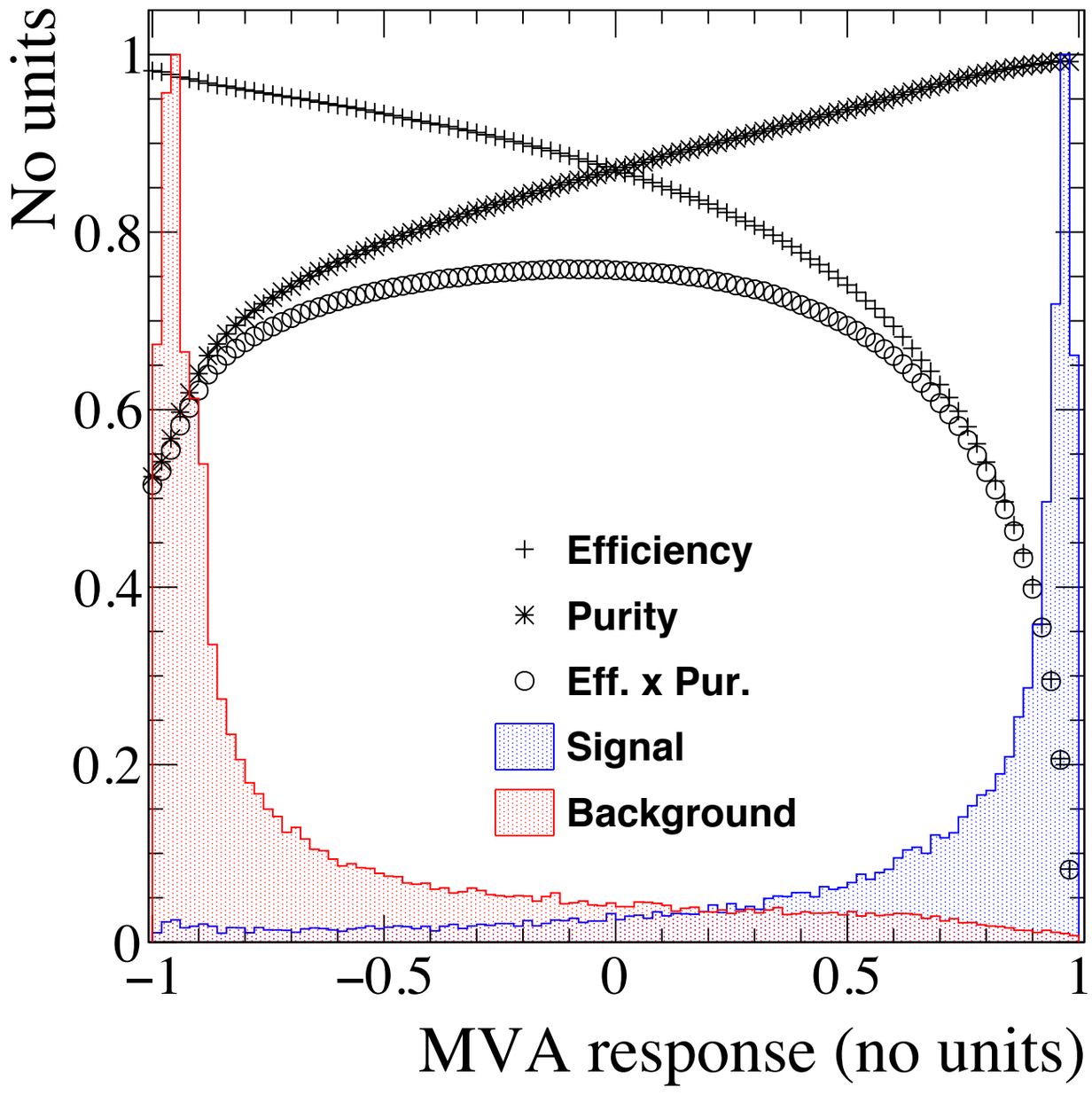
- It should also be noted that pandora has been run out of the box i.e. there has been no tuning for the far detector setup

MVA distribution

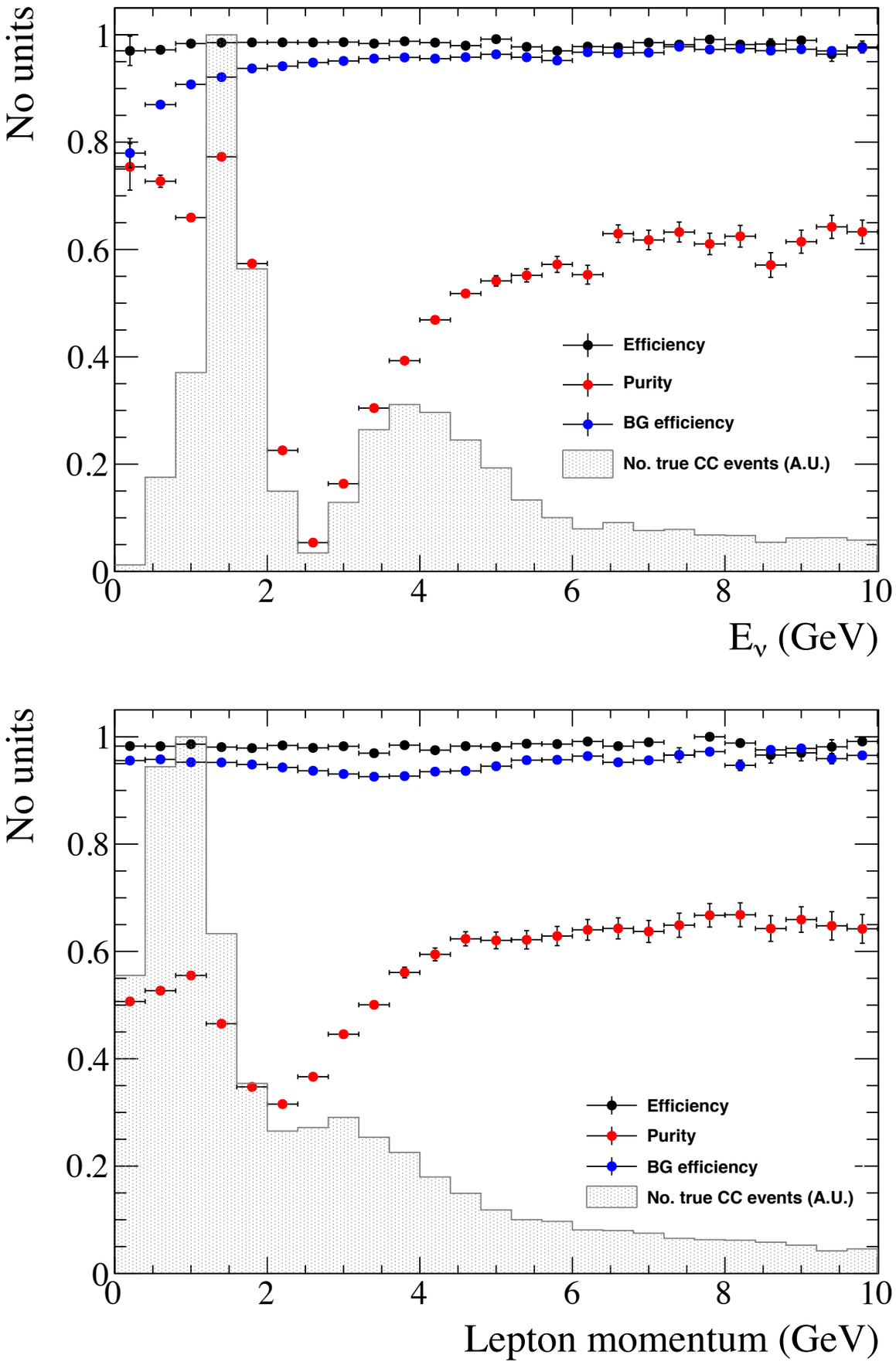
PMA track reconstruction

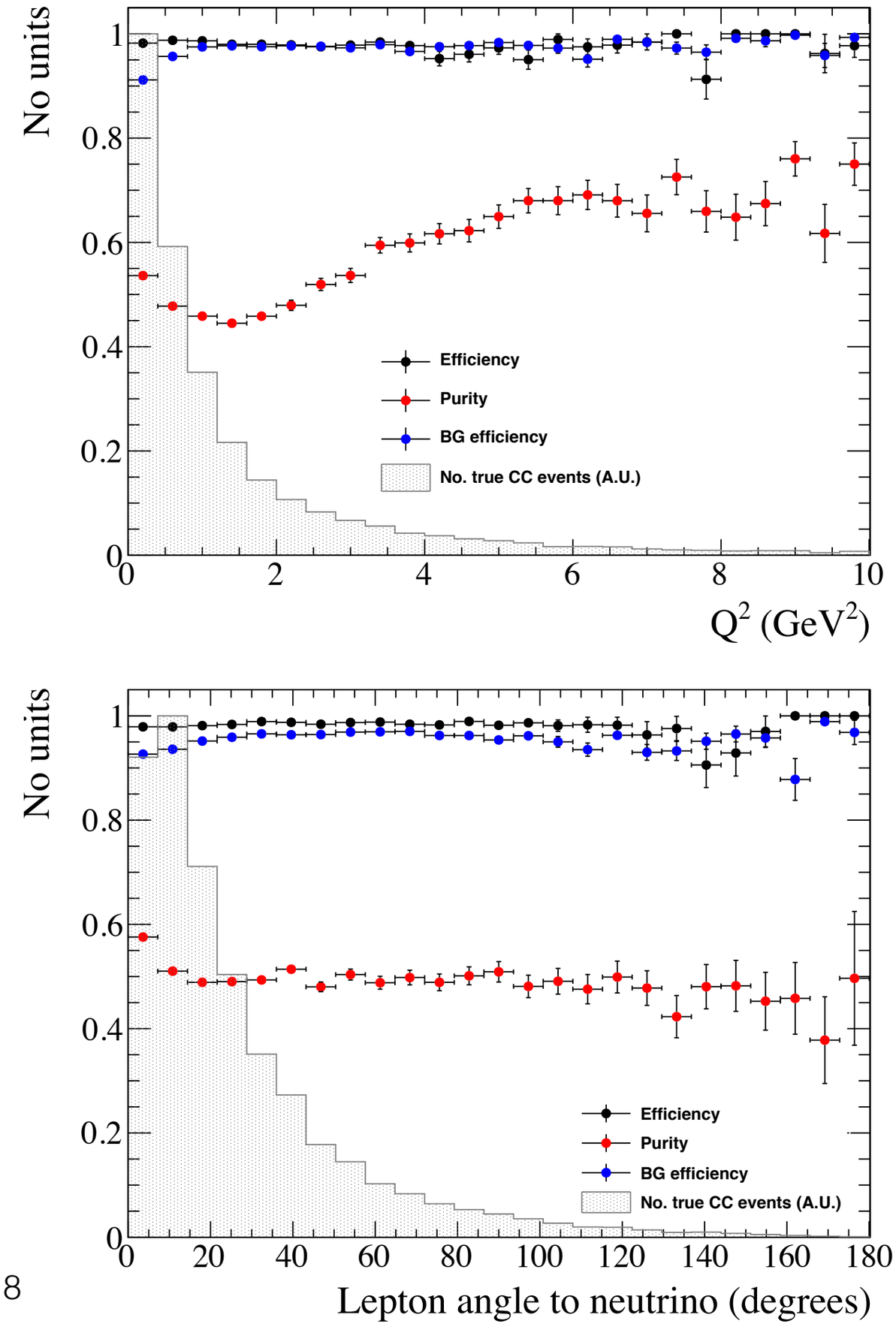


Pandora reconstruction

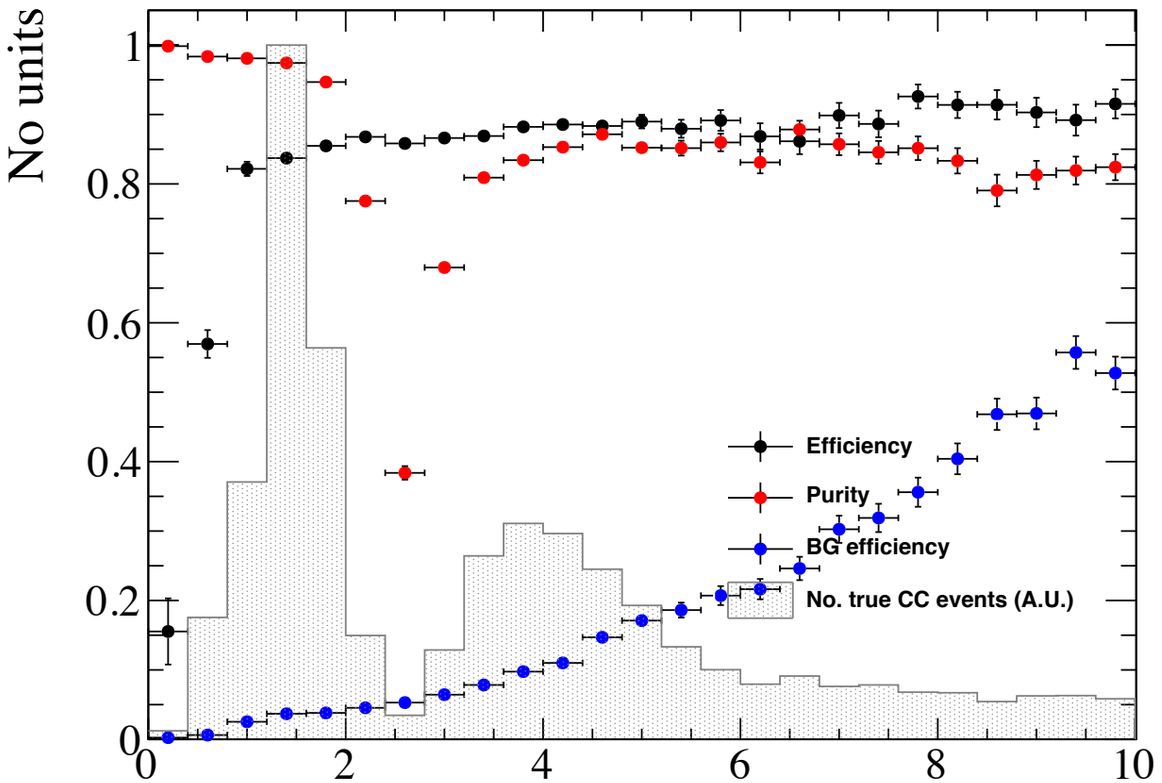


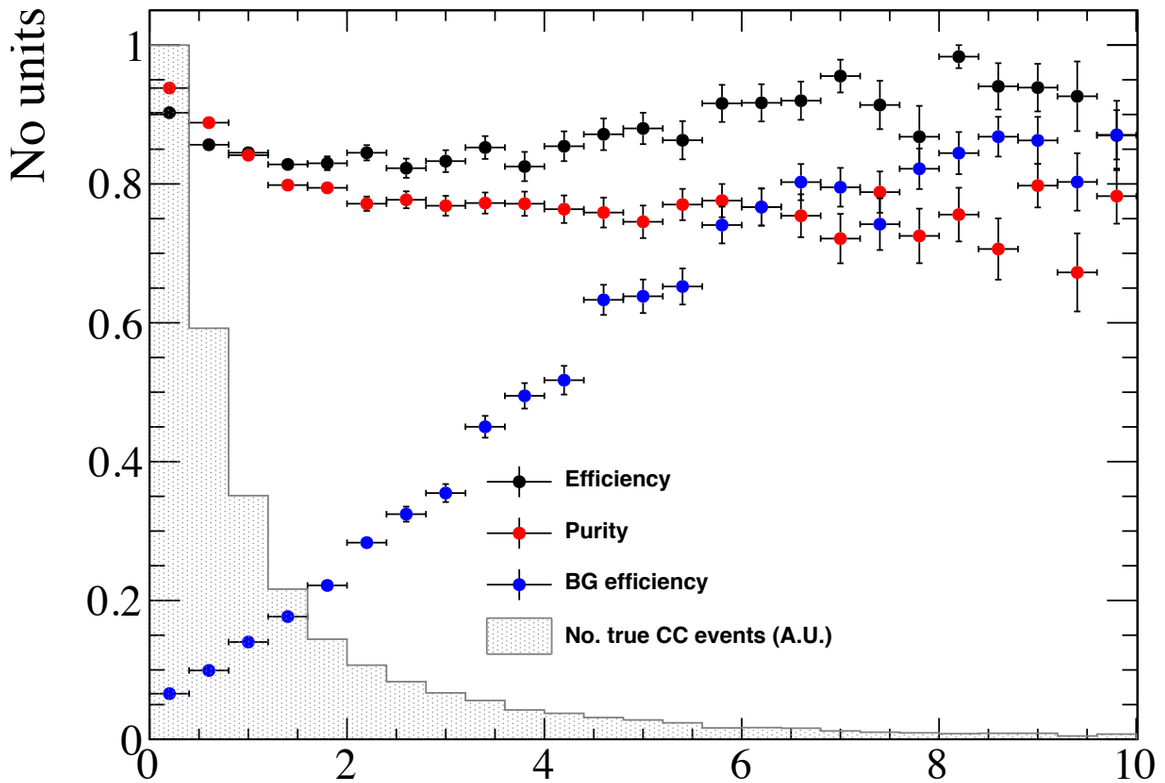
Pandora before selection





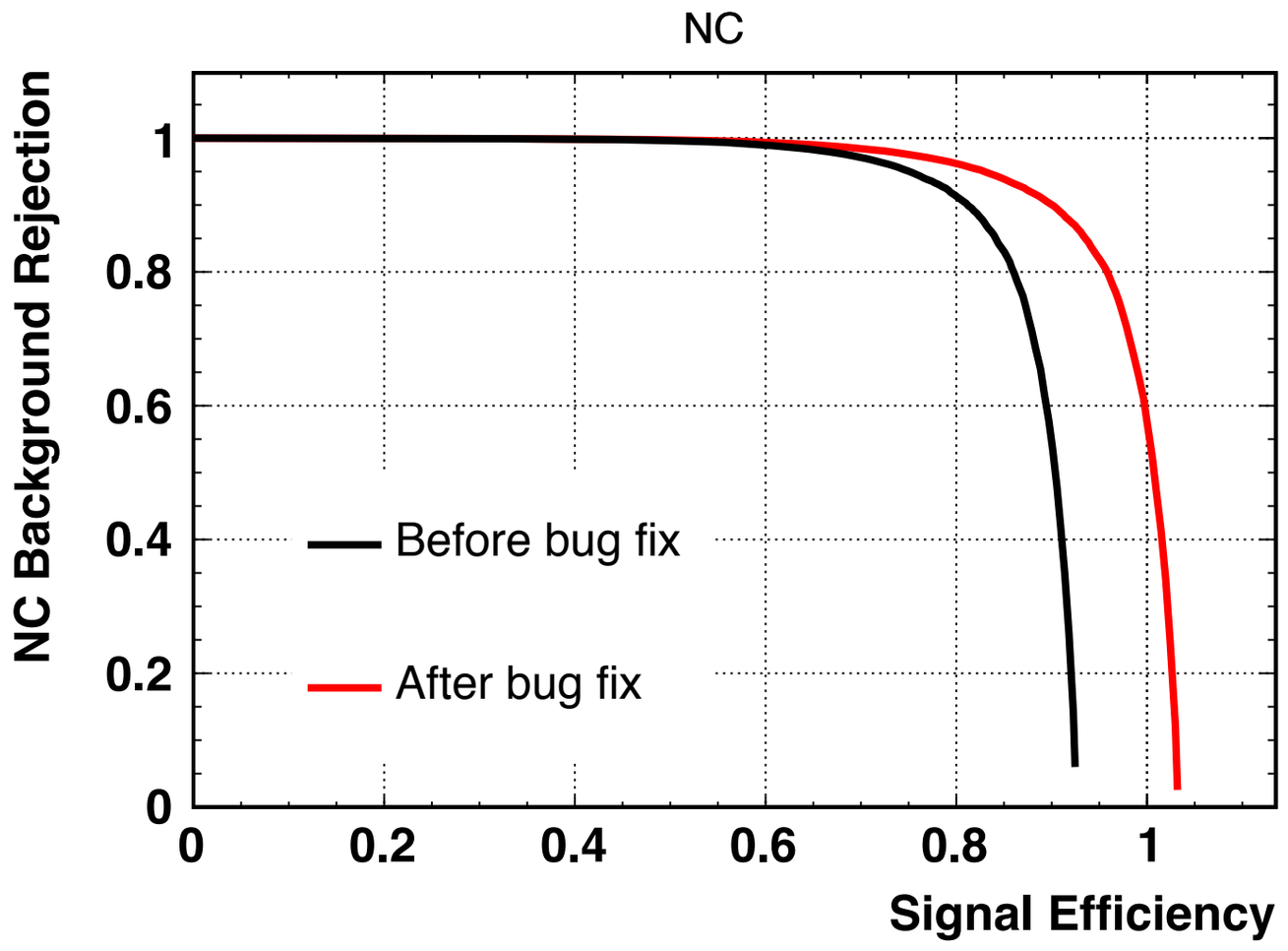
Pandora after selection (MVA > 0)



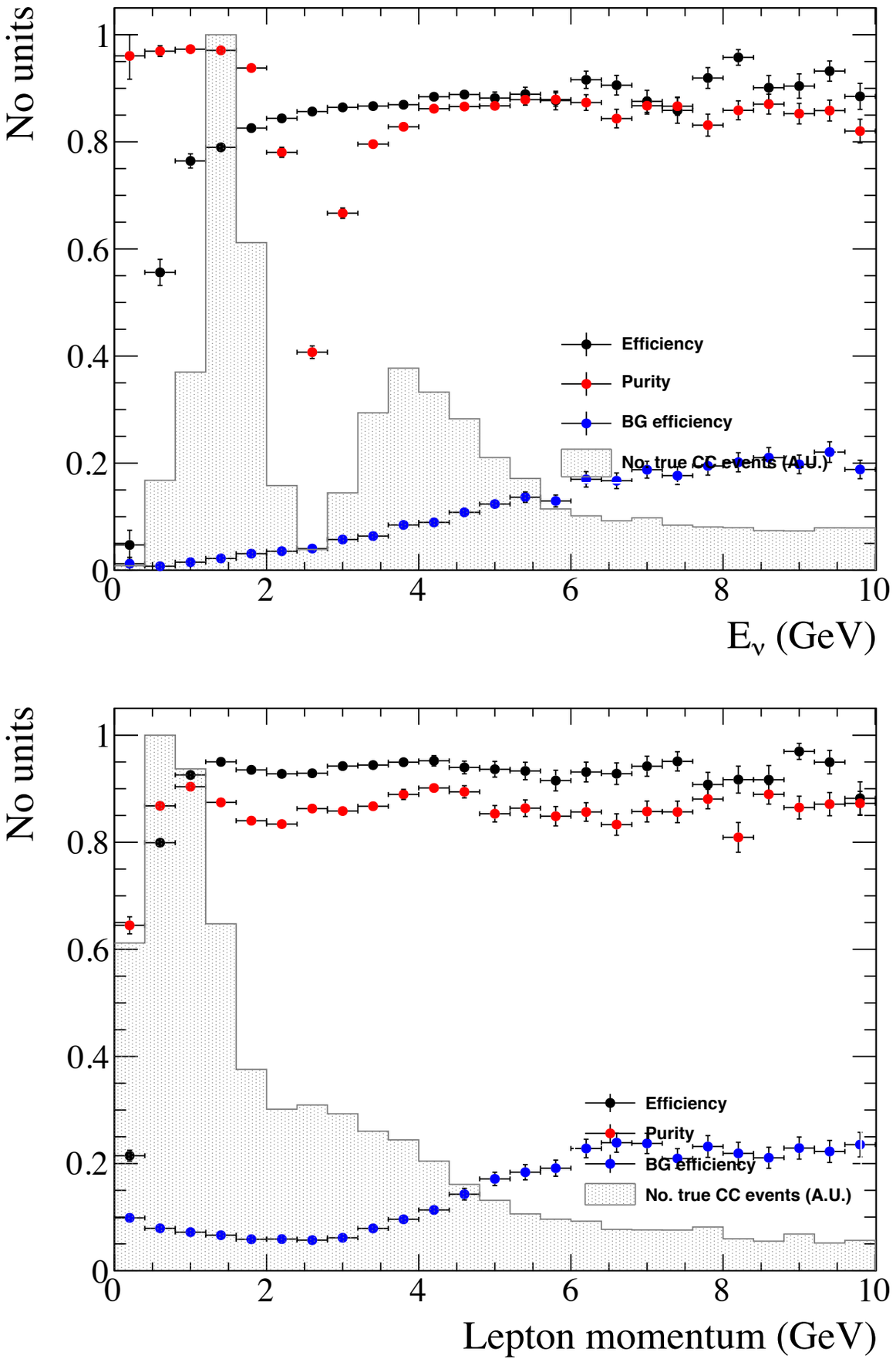


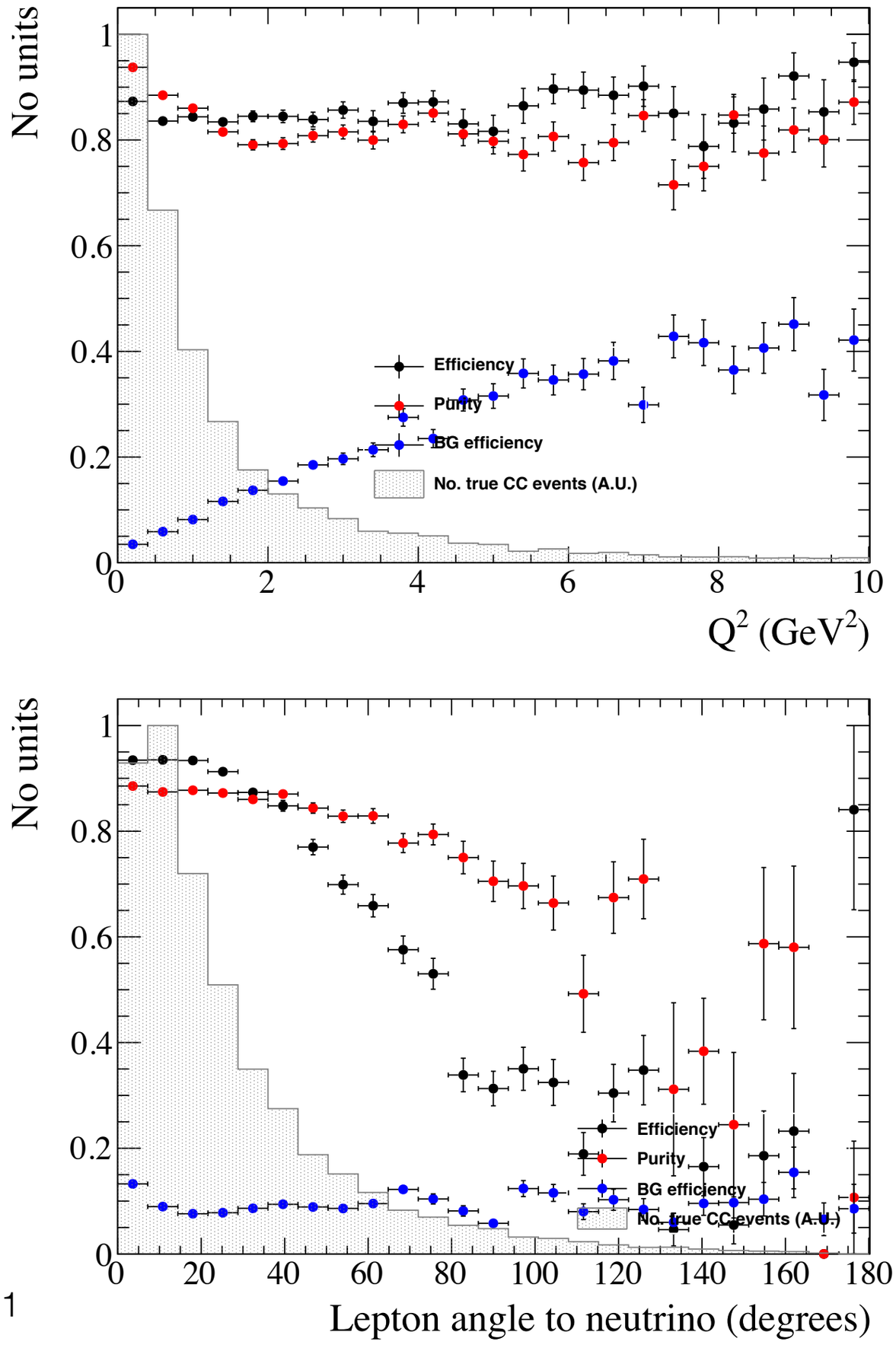
A bug has been found

- MVASelect codes uses finite-size arrays for holding reco. vertex info in memory
  - Max array size == 100
- PMA track likes to make a lot of tracks -> makes a lot of vertices
- In a lot of cases, the signal vertex was being dropped -> the event flagged as out of fiducial volume



Bug fixed eff. and pur. (MVA > 0)





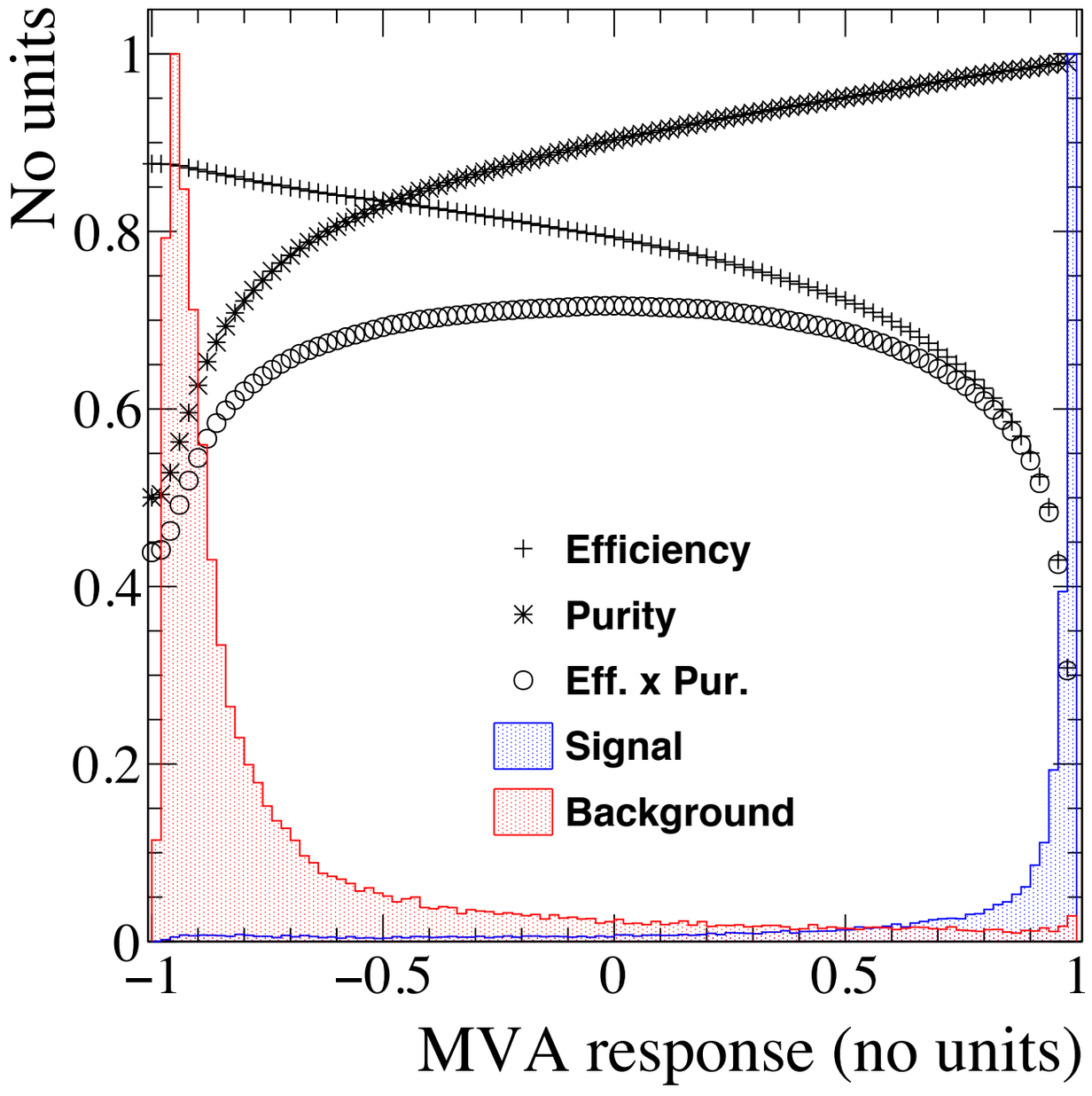
Observations and future work in this area

- The bug fix does fix the Q^2 bias
- The bug fix does not fix any biases from cutting on the MVA

- Test an alternative simple reconstruction using pandora PID
  - Unfortunately the pandora PID is not stored in the reco files by default. More processing will be needed before I can access the PID
- Fully reprocess the PMTrack and Pandora samples with the bug fix
- Press on with finding removing the selection bias

Tuning the selection

- The core of the selection is one cut on the BDT
  - This cut requires tuning
- So far I've tuned on eff. x pur.
  - Suggests MVA > 0
- This tuning metric does not address what we are trying to measure



Tuning the selection

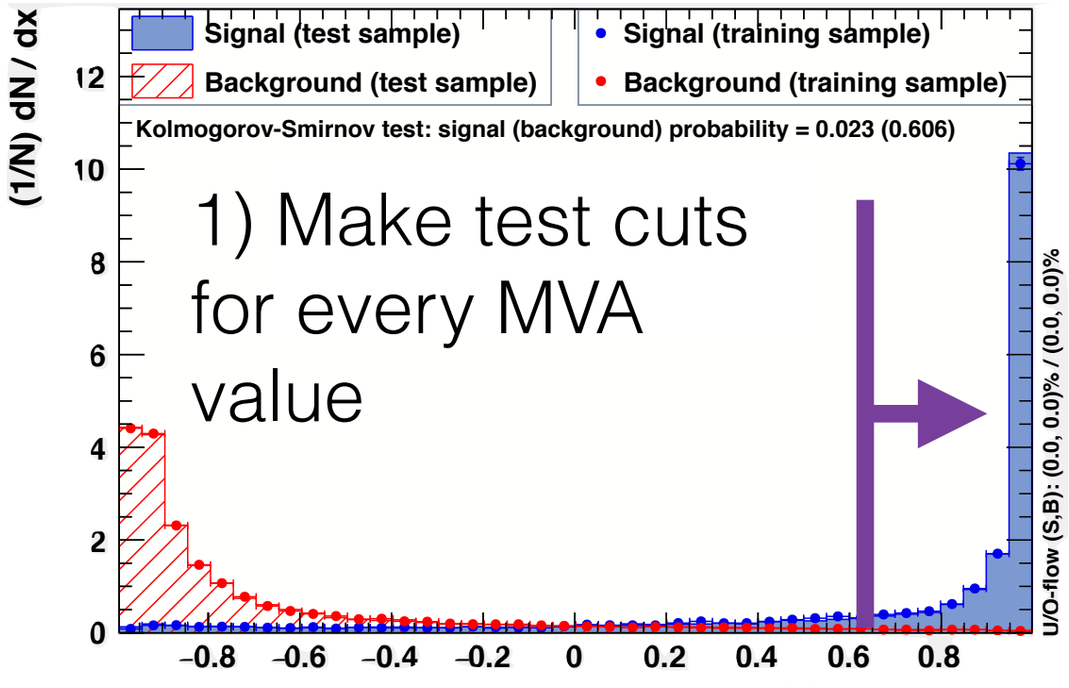
$$\mathcal{A}_{CP} \sim \frac{\cos\theta_{23}\sin 2\theta_{12}\sin\delta_{\mathrm{CP}}}{\sin\theta_{23}\sin\theta_{13}}\left(\frac{\Delta m^2_{21}L}{4E_\nu}\right) + \text{matter effects}$$

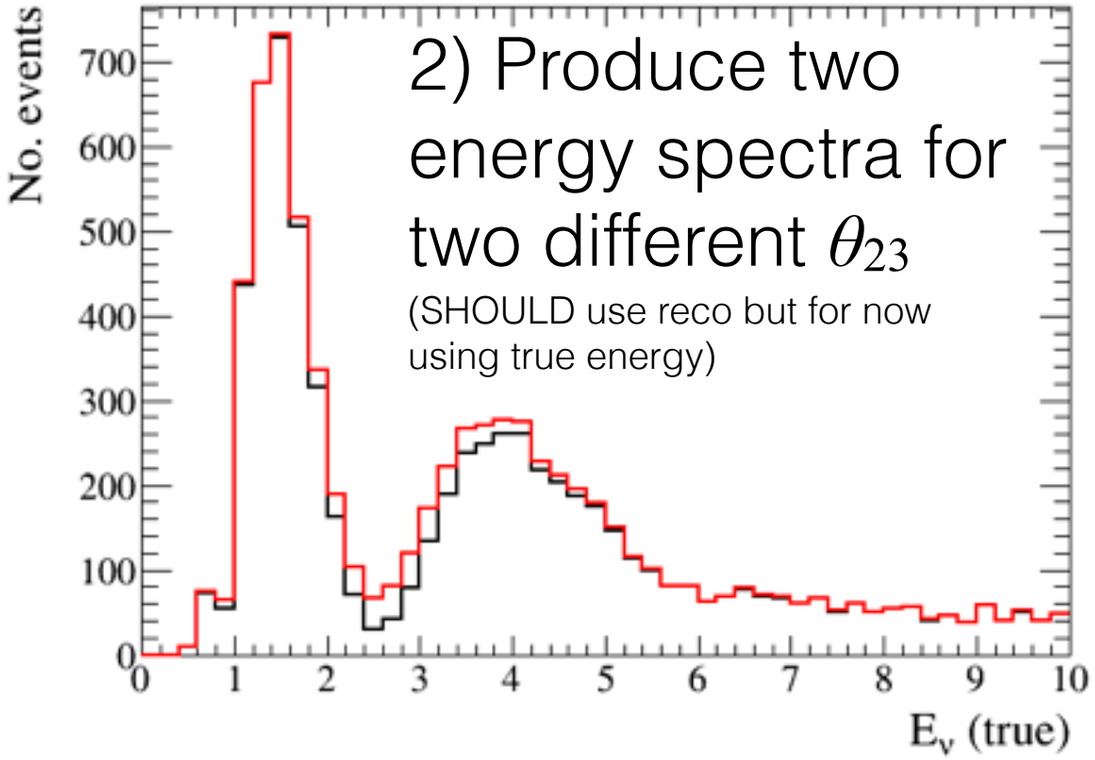
Taken from DUNE CDR

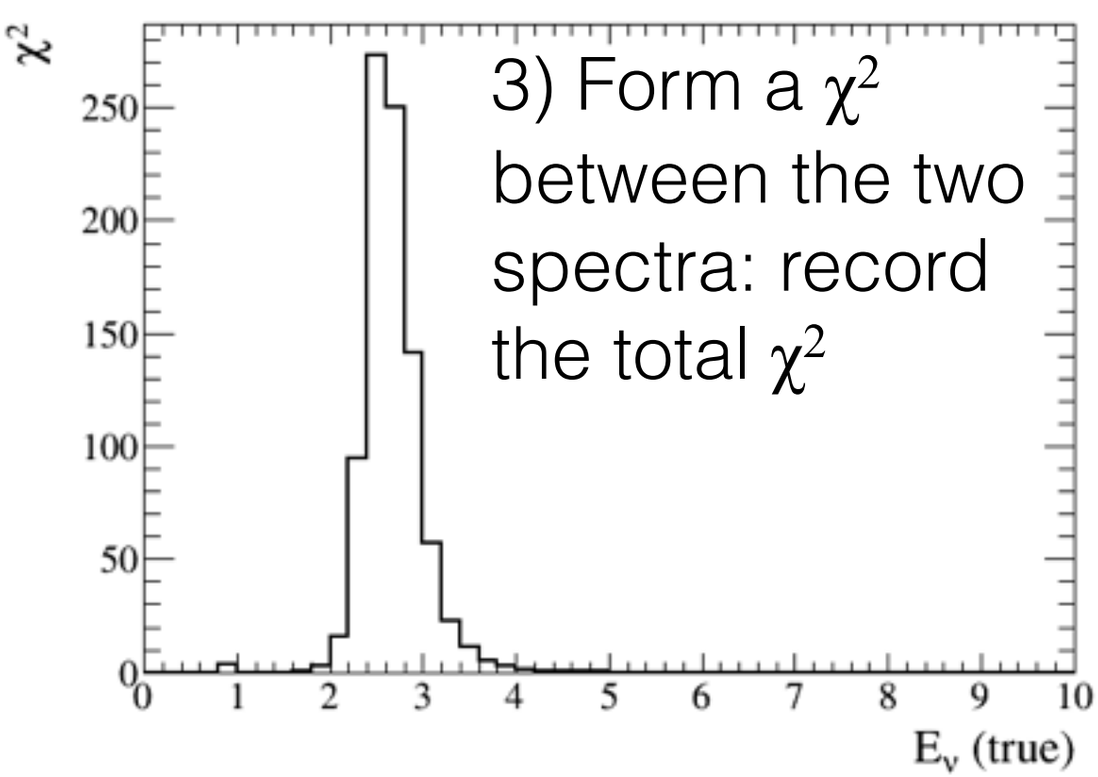
- We obviously want to measure δ_{CP} and we can get at that parameter by measuring the the CP asymmetry (defined above)

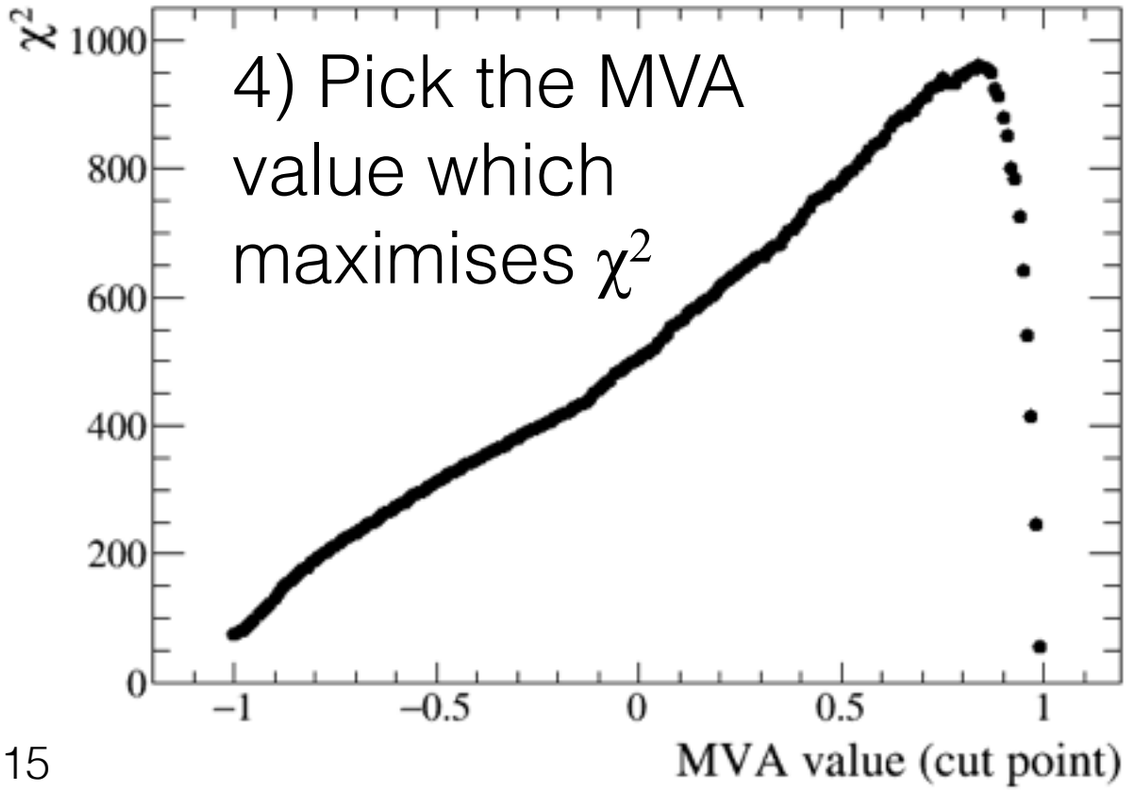
- ν_μ disappearance provides a handle on θ_{23} and so we should maximise the ν_μ selection's sensitivity to that parameter

How to tune


Signal (test sample)
Background (test sample)
Signal (training sample)
Background (training sample)
Kolmogorov-Smirnov test: signal (background) probability = 0.023 (0.606)
(1/N) dN / dx
U/O-flow (S,B): (0.0, 0.0)% / (0.0, 0.0)%
BDTG response
1) Make test cuts for every MVA value


No. events
E_ν (true)
2) Produce two energy spectra for two different θ_{23}
(SHOULD use reco but for now using true energy)

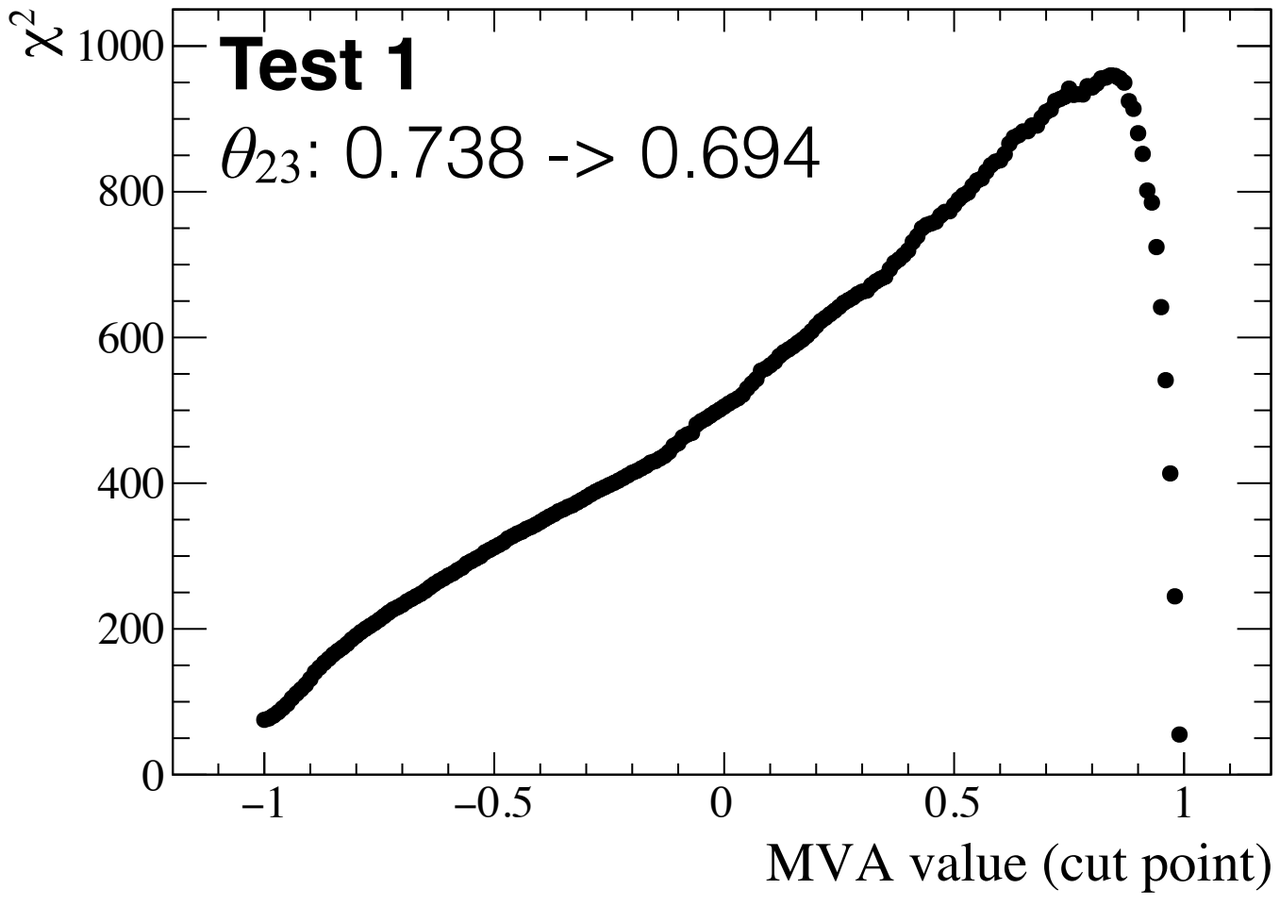

χ^2
E_ν (true)
3) Form a χ^2 between the two spectra: record the total χ^2

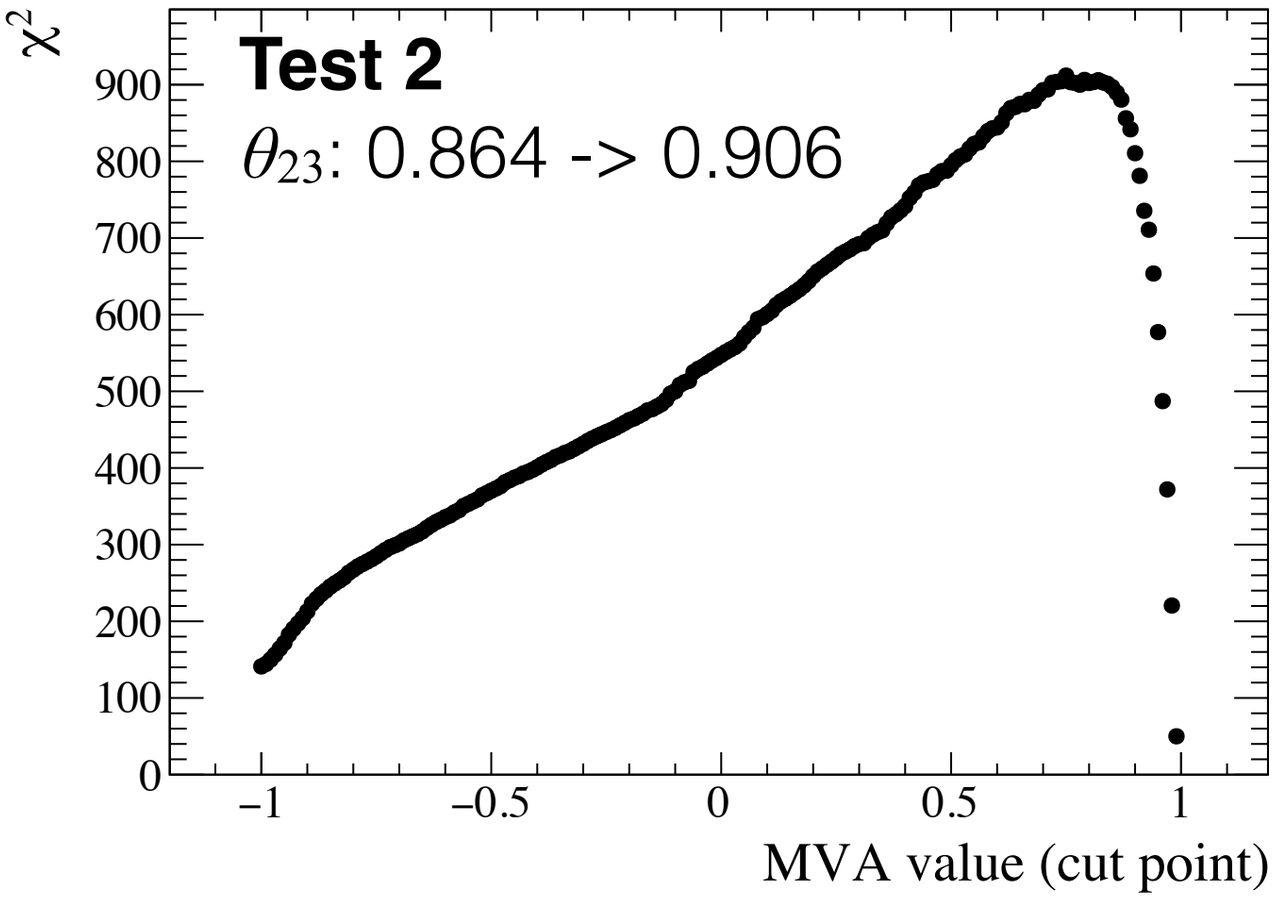

χ^2
MVA value (cut point)
4) Pick the MVA value which maximises χ^2

Which θ23 values?

Test no.	θ_{23} A	θ_{23} B	Why?
1	0.738	0.695	CDR NH - 1σ
2	0.864	0.906	CDR IH + 1σ
3	0.785	0.738	Max. mix. to CDR NH
4	0.785	0.864	Max. mix. to CDR IH
5	0.738	0.864	CDR NH -> CDR IH

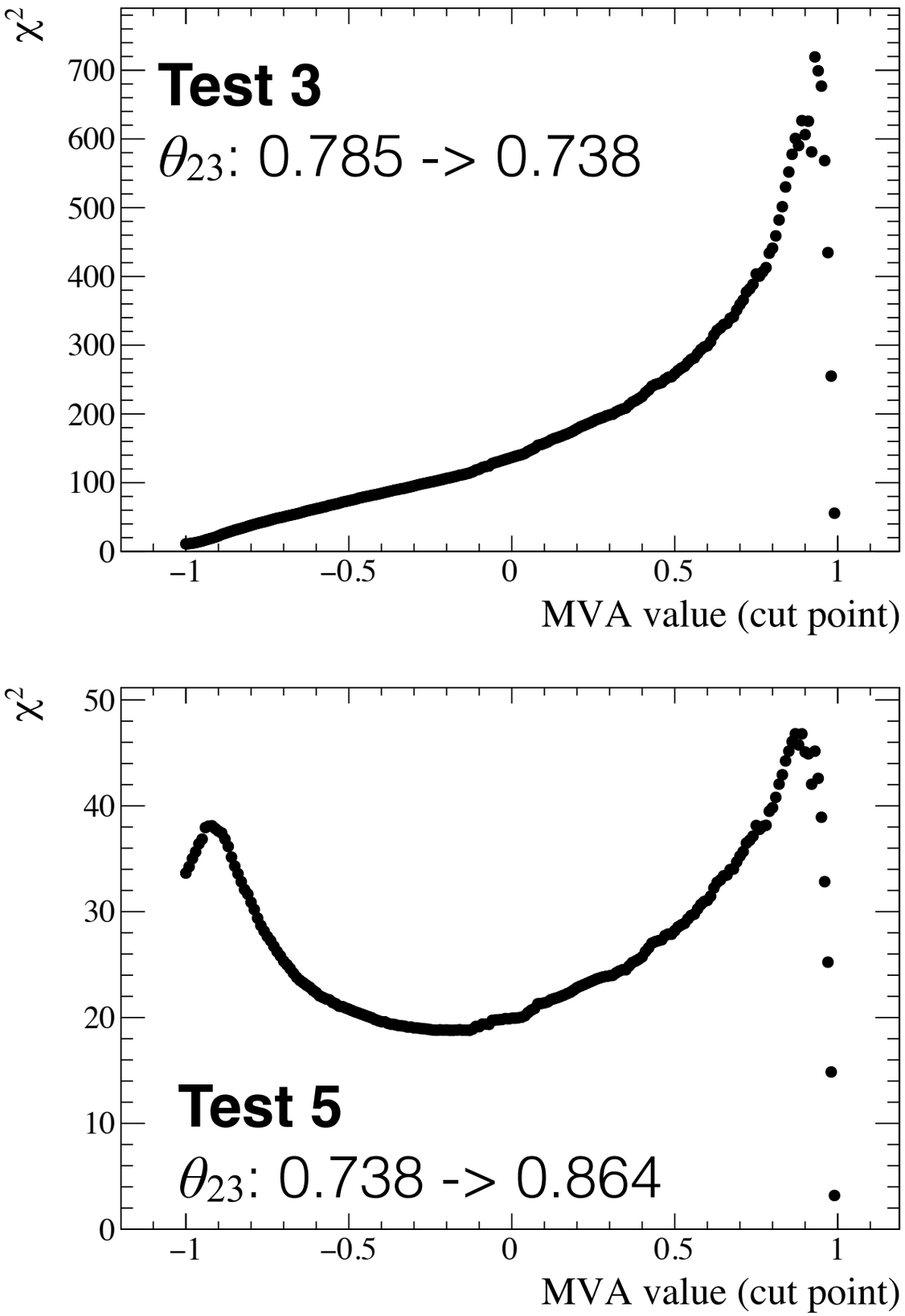
χ^2 distributions

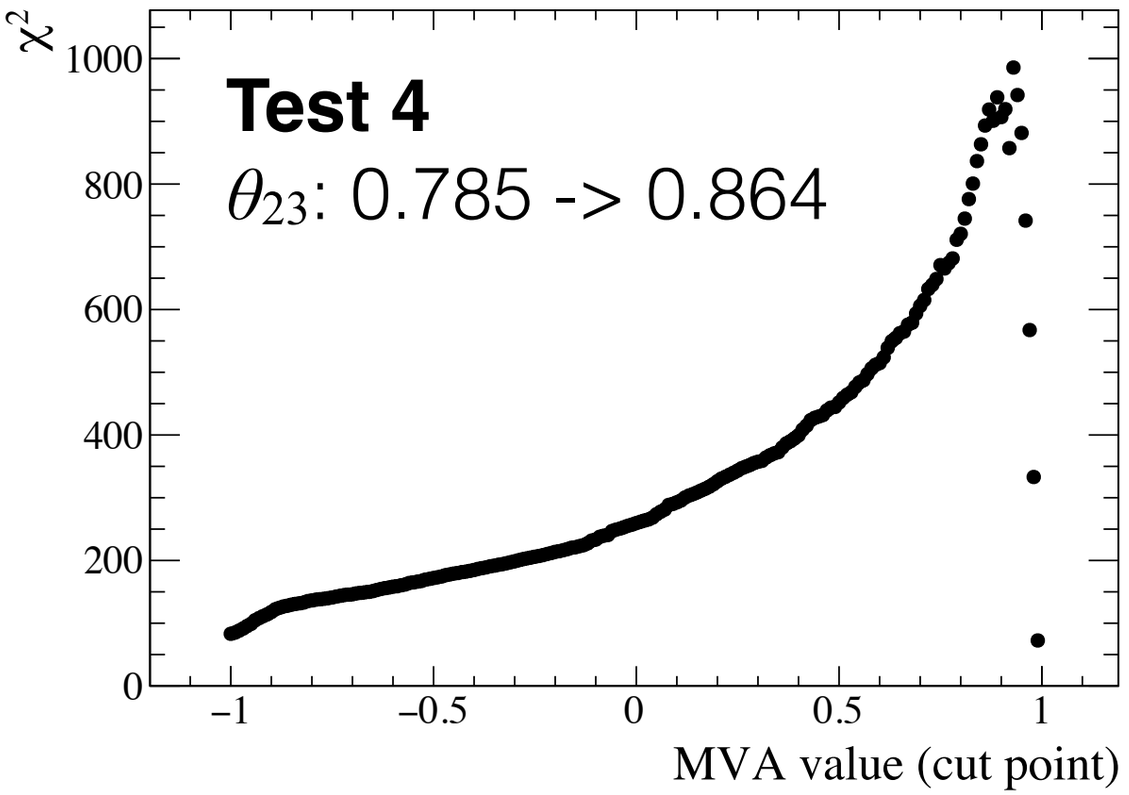




Distributions are similar

χ^2 distributions





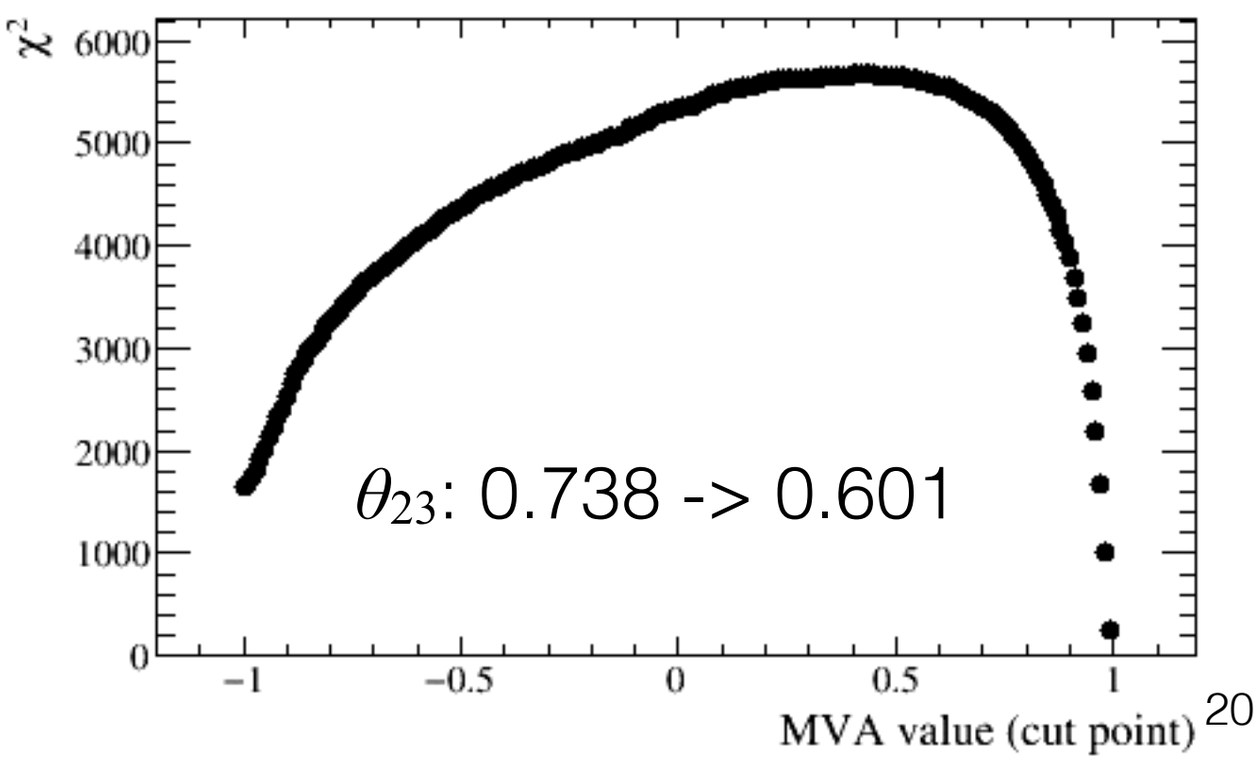
- Test 3 and 4 also look similar though there is a funky wobble around the peak
- Test 5 shows a double peak structure, both peaks with a very low χ^2

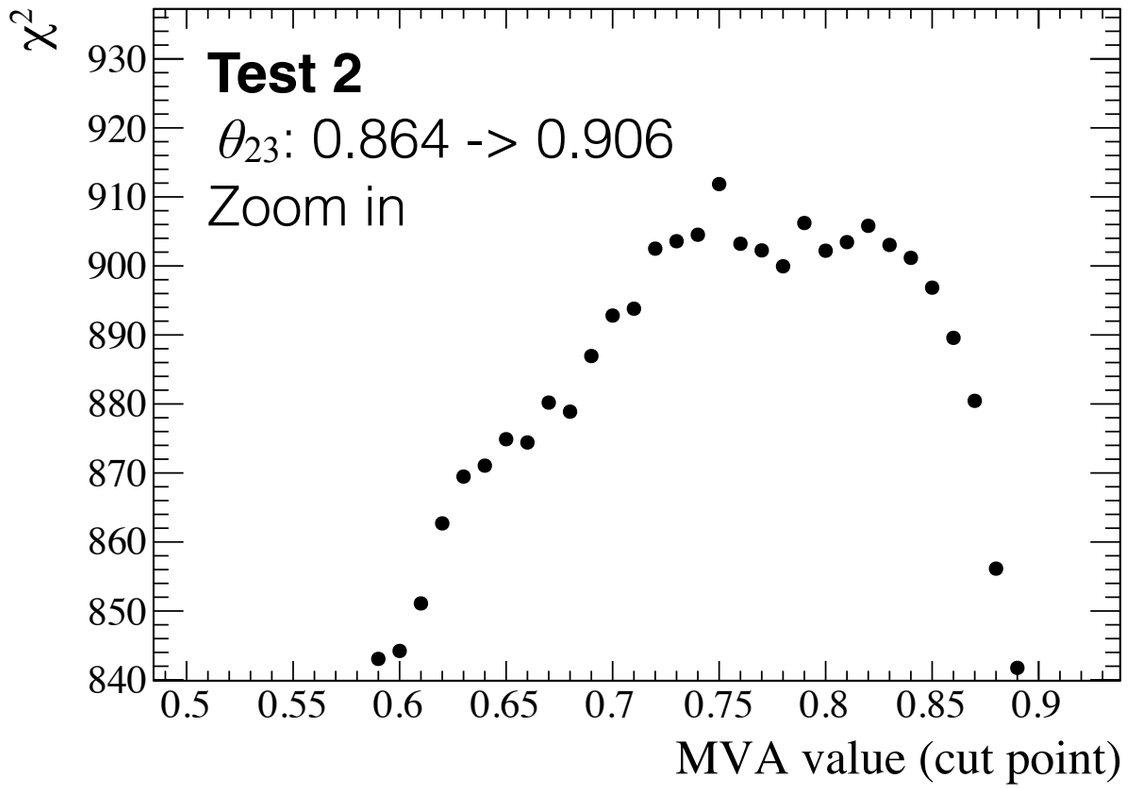
Preferred MVA values

Test no.	θ_{23} A	θ_{23} B	Max χ^2	Pref. MVA value
1	0.738	0.695	960	0.84
2	0.864	0.906	912	0.75
3	0.785	0.738	719	0.93
4	0.785	0.864	986	0.93
5	0.738	0.864	47	0.87

Discrepancies in preferred MVA values

- I think part of the issue is low statistics in areas where the signal/background ratio is changing very rapidly

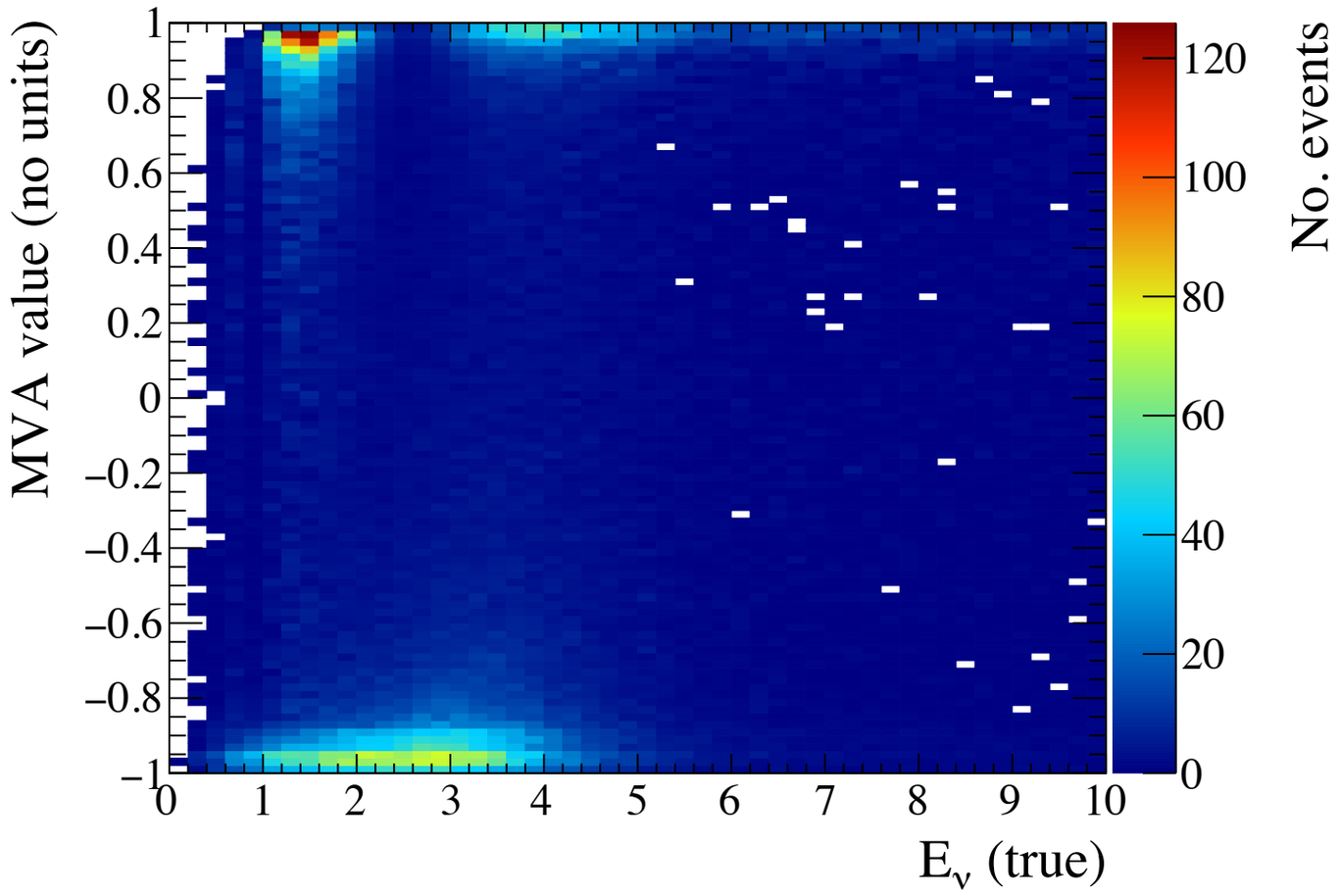




- However, picking more extreme θ_{23} comparisons does change the shape of the χ^2 distribution in a smooth manner

Using the MVA as a reco. variable in the fit

- The BDT is not perfect
  - We lose signal when we cut on it
  - We are potentially losing sensitivity
- Instead, why not try to incorporate the MVA into the fitting?
- Is this worth doing? We can test. Produce MVA vs E_ν distributions for two values of θ_{23} and measure the χ^2
  - If the χ^2 for the 2D dist. is larger than the 1D equivalent (what I've already shown) then there is more sensitivity in the 2D distribution



χ2 values

Test no.	θ_{23} A	θ_{23} B	Max χ^2 (1D)	χ^2 (2D)
1	0.738	0.695	960	1182
2	0.864	0.906	912	1125
3	0.785	0.738	719	908
4	0.785	0.864	986	1289
5	0.738	0.864	47	99

In all cases, the 2D χ^2 is at least 20% larger than the 1D equivalent

Some thought(s)

- In all test cases, the 2D distribution shows more sensitivity to θ_{23} than its 1D equivalent

- Does this mean we should proceed with this route?

  No

- Effects of systematics degrade sensitivity -> reduce the χ^2 contribution coming from each bin

- The 2D distribution contains a lot more bins so it is potentially a lot more sensitive to systematic effects

Summary

- The selection has been tested using pandora reconstruction
  - Performance is similar to PMA track
- T. Yang has found a bug in MVASelect which solves the Q^2 bias
- Biases after applying the MVA cut still remain
- How to tune the MVA cut is currently being studied. Two options have been presented, both of which try to maximise sensitivity to θ_{23}
  - Cutting directly on the MVA variable (1D)
  - Passing the MVA off as a reco. variable to the fitters (2D)
- At a first glance, the 2D method is more sensitive to θ_{23} effects

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