Progress Update

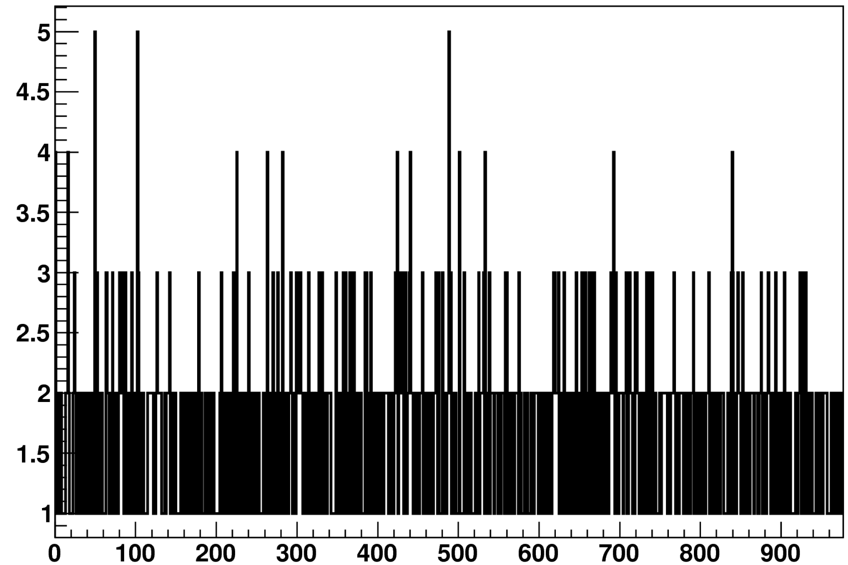
Trevor Towstego
UofT Neutrino/DM Meeting
July 13, 2018

Using New MC Files

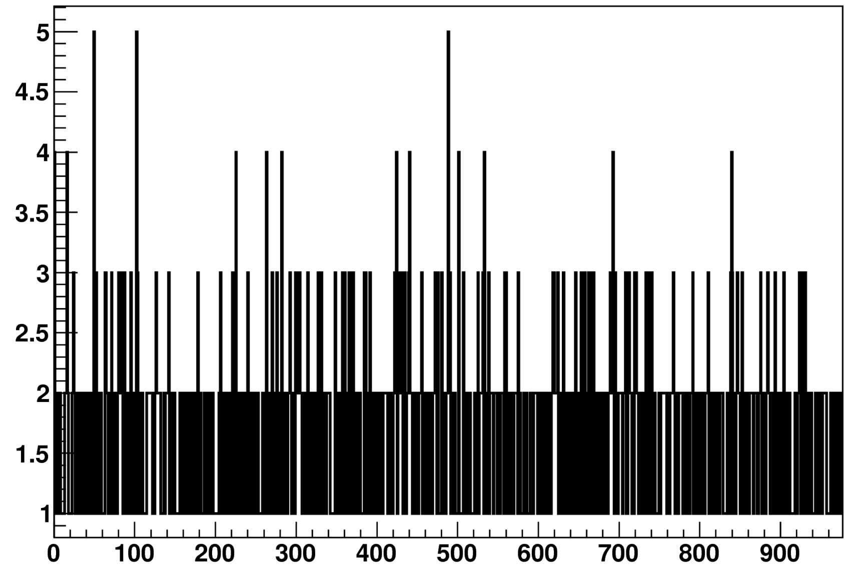
- Since last time, I’ve combined the previous T2K-SK MC files with the new MC files containing additional fiTQun information
  - Took MR information from new files, and everything else I need from old files
  - Used nev variable and went file-by-file to ensure events matched
- Made some code to verify results event-by-event
- Also made some plots to compare which MR fits are available before/after

MC Combination Verification (1 MC file only)

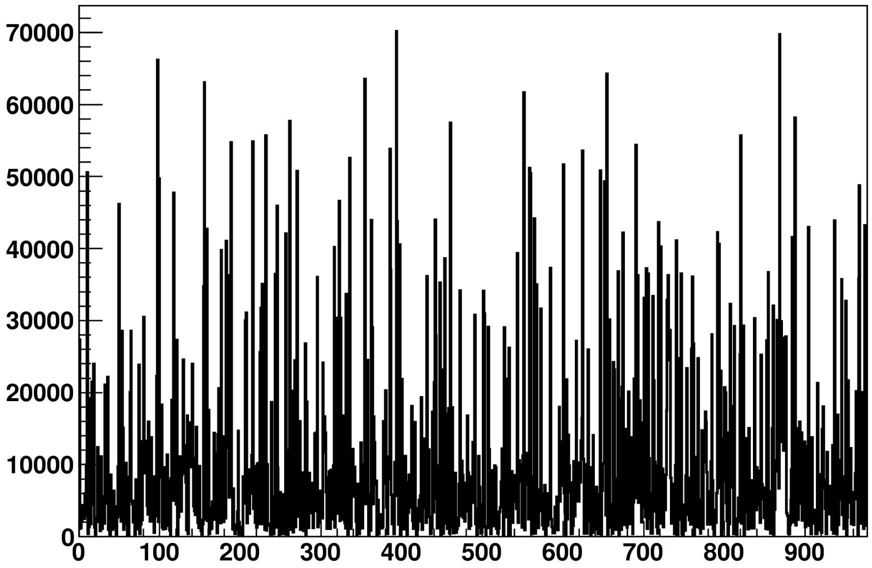
fqnse: fQ-only file



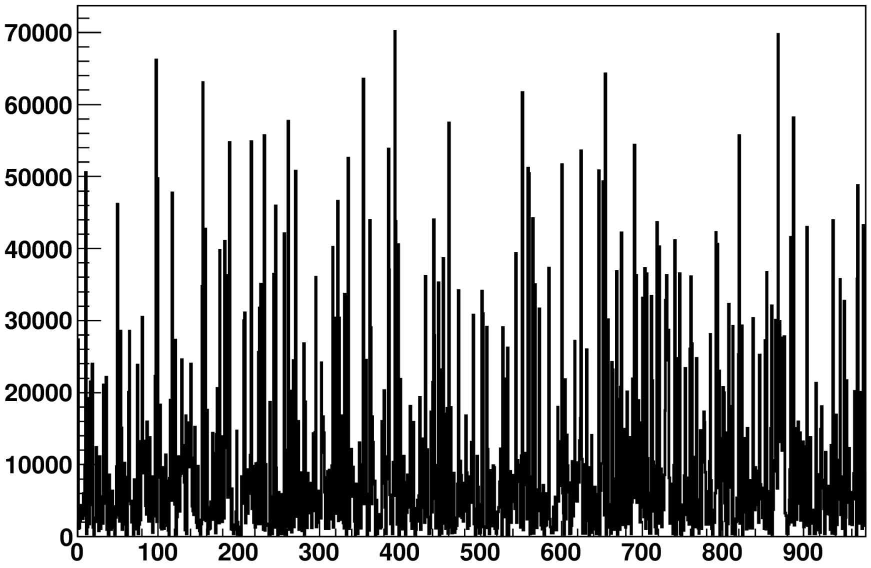
fqnse: merged file



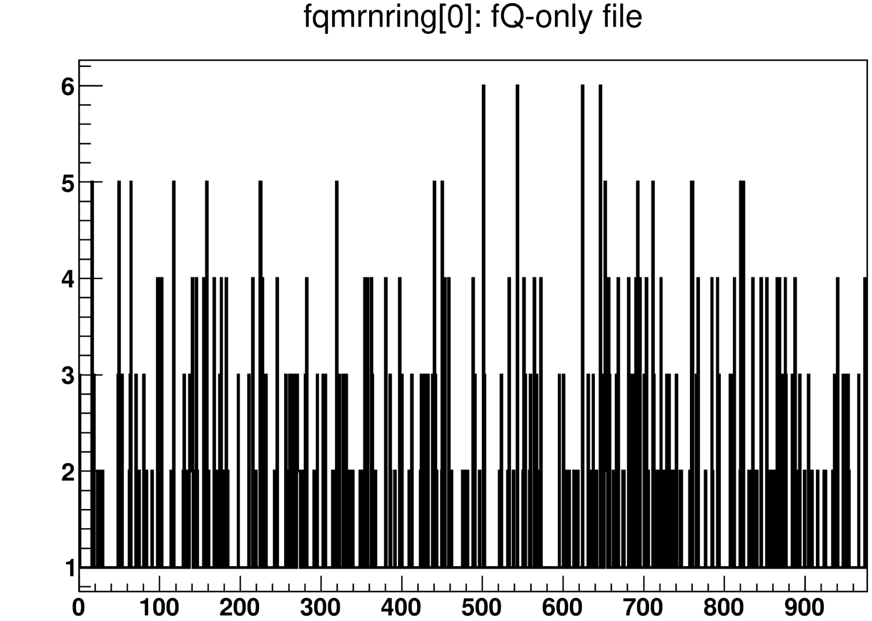
fq1rnll[0][1]: fQ-only file

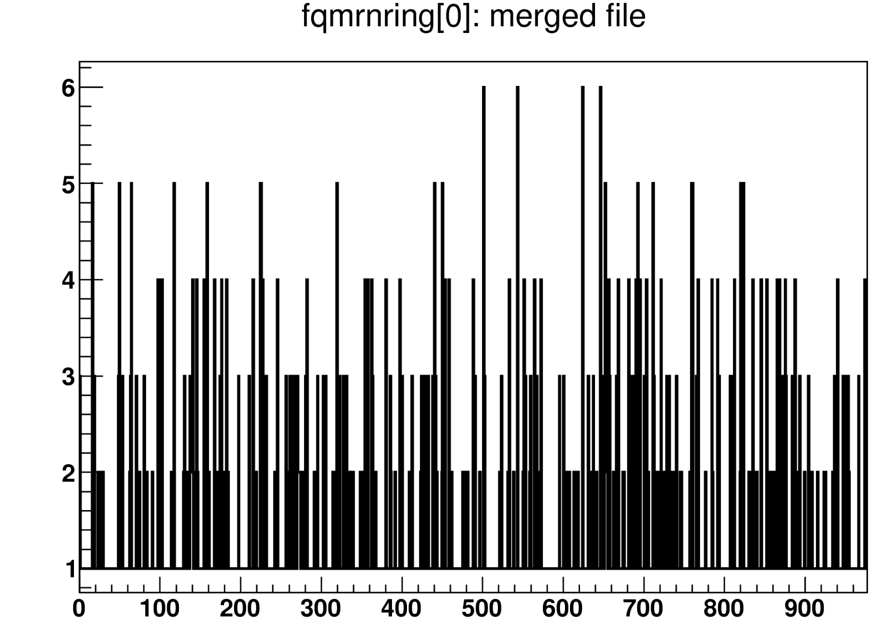


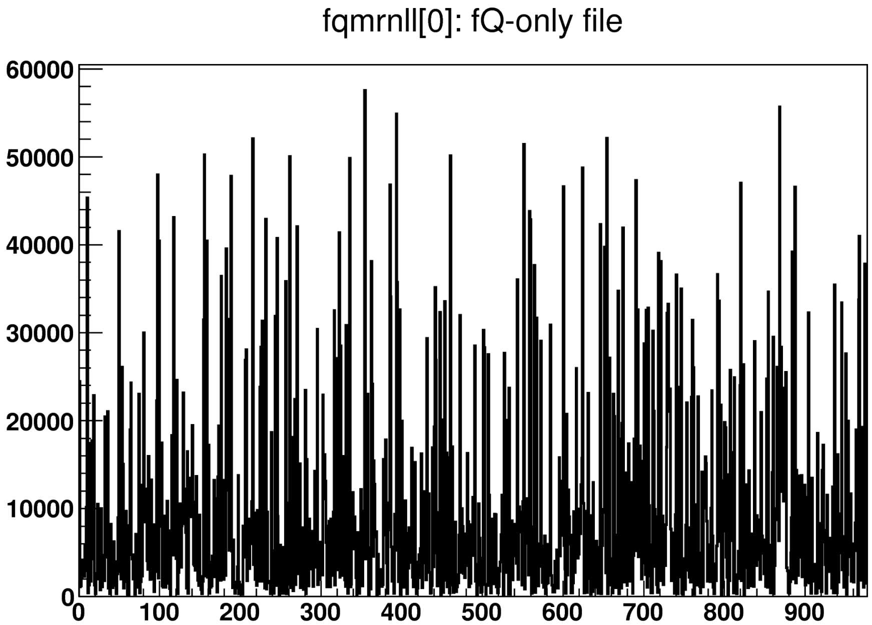
fq1rnll[0][1]: merged file

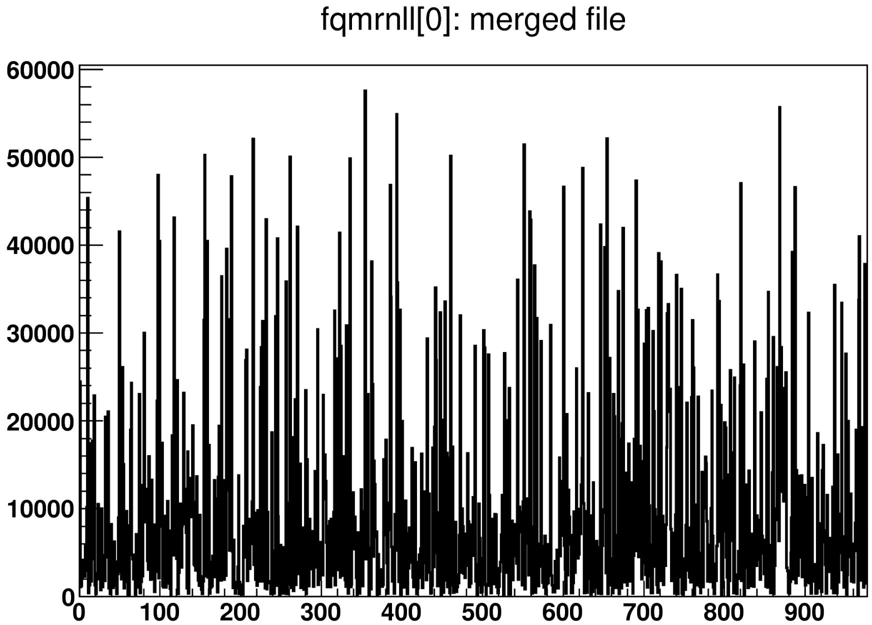


MC Combination Verification (1 MC file only)

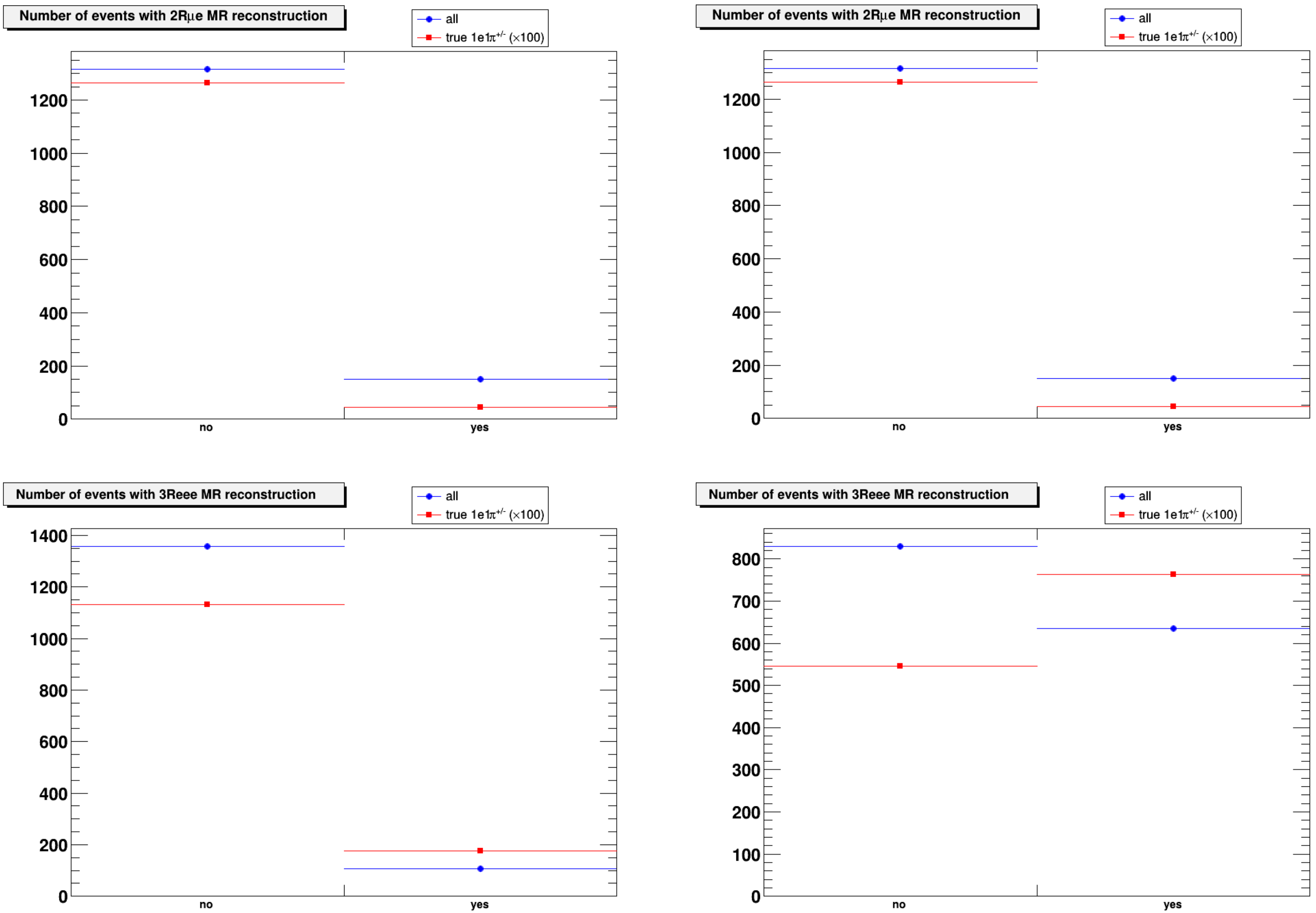




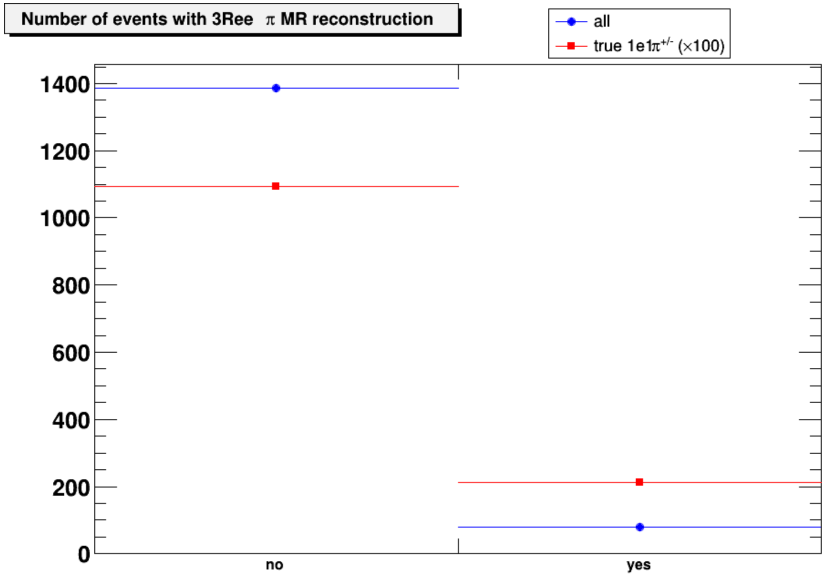


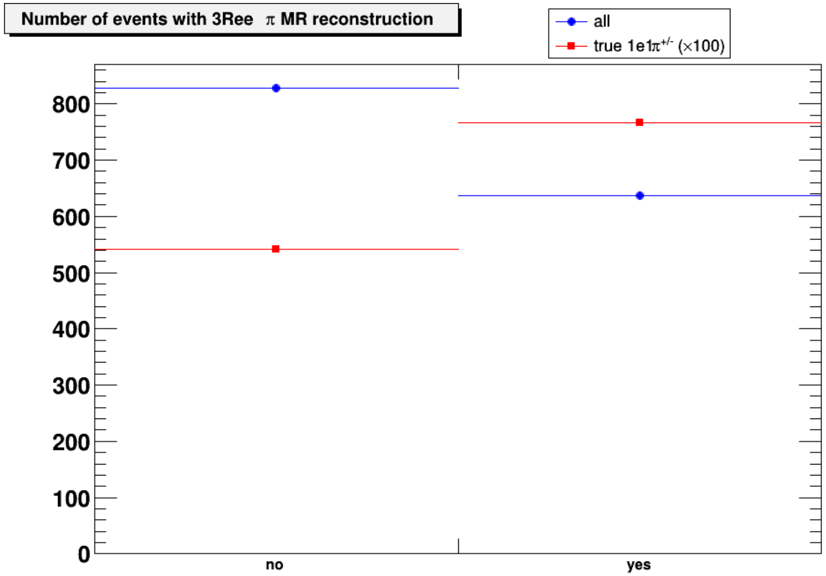


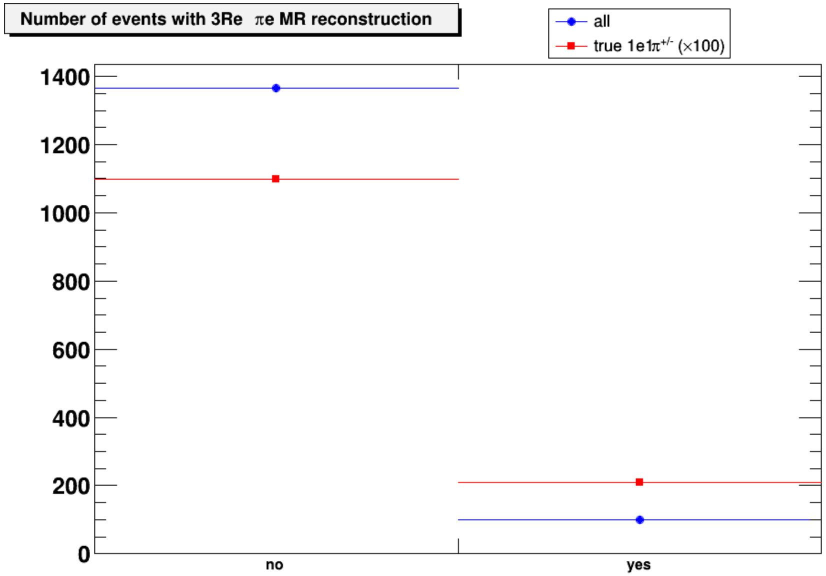
Available MR Fits: Before & After

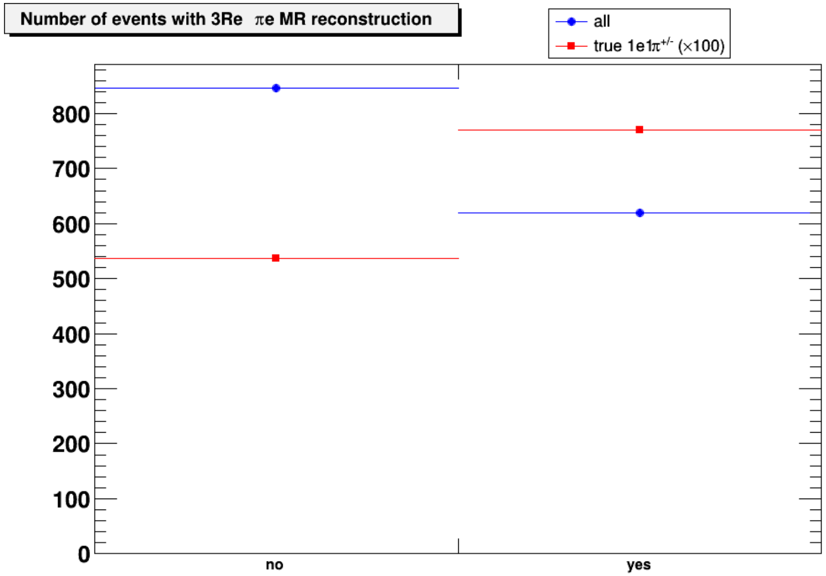


Available MR Fits: Before & After

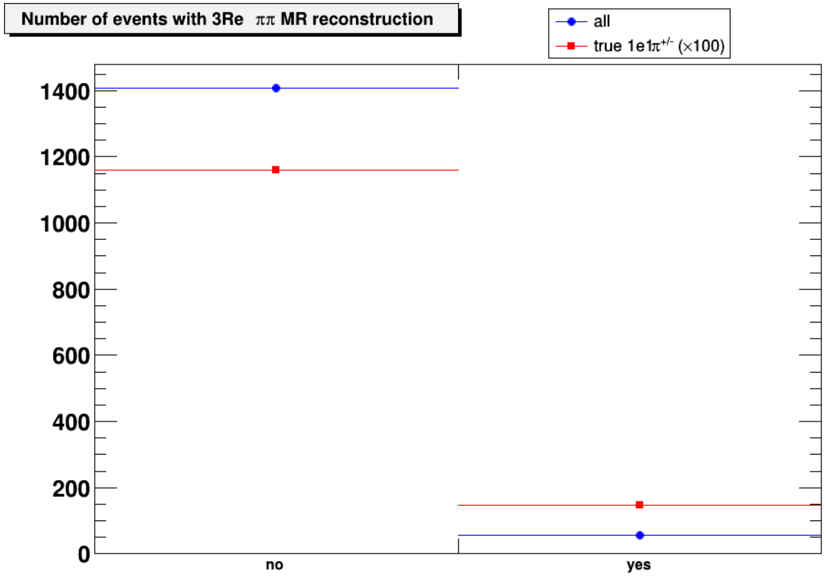


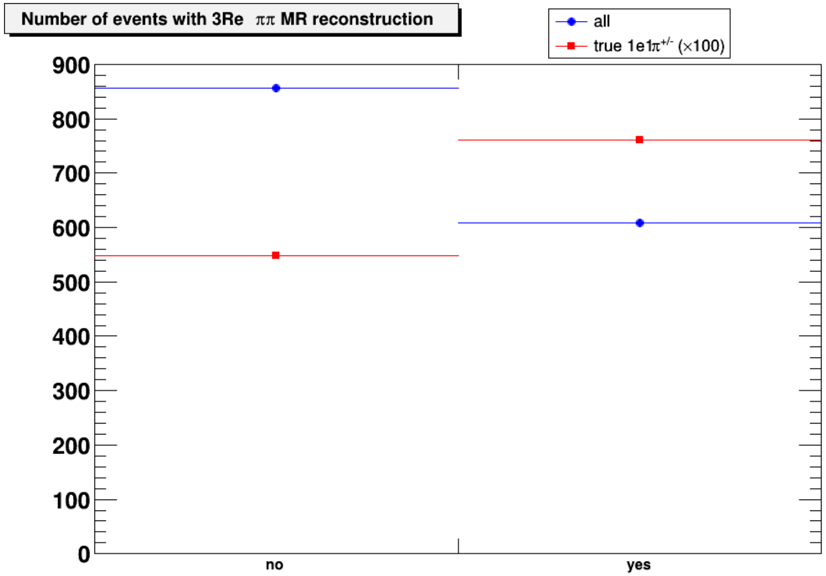






Available MR Fits: Before & After





- Although not all events have 2Rμe and 3R fits, a majority of true $1e1\pi^{+/-}$ events have the 3R fits of interest

BDTs

- Preliminary cuts:
  - FCFV
  - possible 2Repi
    - 0 de: i2repi==0 || i2rpie==0 || i3repipi==0
    - 1 de: (i1re==0 && !Is1re && !Is1re1de) || i2ree==0 || i2repi==0 || i2rpie==0 || i2rmue==0 || i3repipi==0
  - 1/2 sub-events
    - separate samples
  - $E_{rec}(1e,1\pi) < 1.5$ GeV
- Since last time, looked at benefit of adding 3R likelihood information (with padding)

BDTs: Trial 1

- BDT variables:
  - nll1Re - nll1Rμ
  - nll1Re - nll2Reπ
  - nll1Re - nll2Rπe
  - nll1Re - nll2Ree
  - nll1Rμ - nll2Reπ
  - nll1Rμ - nll2Rπe
  - nll1Rμ - nll2Ree
  - nll2Reπ - nll2Rπe
  - nll2Reπ - nll2Ree
  - nll2Rπe - nll2Ree

		Signal	Bkgd	Purity	Eff	FOM
2Reπ	New BL	0.70	2.32	23.3%	27.4%	0.405
	BDT 1	0.71	1.84	27.9%	27.8%	**0.446**
2Reπ1de	New BL	2.71	4.73	36.4%	53.8%	0.992
	BDT 1	2.52	2.64	48.8%	50.0%	**1.108**

- NTrees = 850
- MaxDepth = 3

Note: Signal = true 1e1$\pi^{+/-}$ events

BDTs: Trial 2

- BDT variables:
  - nll1Re - nll1Rμ
  - nll1Re - nll2Reπ
  - nll1Re - nll2Rπe
  - nll1Re - nll2Ree
  - nll1Rμ - nll2Reπ
  - nll1Rμ - nll2Rπe
  - nll1Rμ - nll2Ree
  - nll2Reπ - nll2Rπe
  - nll2Reπ - nll2Ree
  - nll2Rπe - nll2Ree
  - p_e 1Re
  - p_μ 1Rμ
  - p_e 2Reπ
  - p_π 2Reπ
  - p_e 2Rπe
  - p_π 2Rπe
  - p_{e1} 2Ree
  - p_{e2} 2Ree

		Signal	Bkgd	Purity	Eff	FOM
2Reπ	New BL	0.70	2.32	23.3%	27.4%	0.405
	BDT 1	0.71	1.84	27.9%	27.8%	0.446
	BDT 2	0.61	0.98	38.4%	23.7%	**0.484**
2Reπ1de	New BL	2.71	4.73	36.4%	53.8%	0.992
	BDT 1	2.52	2.64	48.8%	50.0%	1.108
	BDT 2	2.58	2.35	52.4%	51.3%	**1.162**

- NTrees = 850
- MaxDepth = 3

Note: Signal = true 1e1$\pi^{+/-}$ events

BDTs: Trial 3

- BDT variables:
  - nll1Re - nll1Rμ
  - nll1Re - nll2Reπ
  - nll1Re - nll2Rπe
  - nll1Re - nll2Ree
  - nll1Rμ - nll2Reπ
  - nll1Rμ - nll2Rπe
  - nll1Rμ - nll2Ree
  - nll2Reπ - nll2Rπe
  - nll2Reπ - nll2Ree
  - nll2Rπe - nll2Ree
  - nll2Reπ - nll3Reee
  - nll2Reπ - nll3Reeπ
  - nll2Reπ - nll3Reπe
  - nll2Reπ - nll3Reππ
  - nll2Rπe - nll3Reee
  - nll2Rπe - nll3Reeπ
  - nll2Rπe - nll3Reπe
  - nll2Rπe - nll3Reππ
  - nll2Ree - nll3Reee
  - nll2Ree - nll3Reeπ
  - nll2Ree - nll3Reπe
  - nll2Ree - nll3Reππ

		Signal	Bkgd	Purity	Eff	FOM
2Reπ	BDT 2	0.61	0.98	38.4%	23.7%	0.484
	BDT 3*	0.59	0.62	48.7%	22.9%	**0.536**
	BDT 33**	0.54	0.60	47.3%	21.1%	***0.506***
2Reπ1de	BDT 2	2.58	2.35	52.4%	51.3%	1.162
	BDT 3*	2.62	2.74	48.8%	52.0%	**1.131**
	BDT 33**	2.20	1.60	57.7%	43.7%	***1.127***

* 3R nlls padded with zeros
** 3R nlls padded with 50,000

- NTrees = 850
- MaxDepth = 3

Note: Signal = true 1e1$\pi^{+/-}$ events

BDTs: Trial 4

- BDT variables:
  - nll1Re - nll1Rμ
  - nll1Re - nll2Reπ
  - nll1Re - nll2Rπe
  - nll1Re - nll2Ree
  - nll1Rμ - nll2Reπ
  - nll1Rμ - nll2Rπe
  - nll1Rμ - nll2Ree
  - nll2Reπ - nll2Rπe
  - nll2Reπ - nll2Ree
  - nll2Rπe - nll2Ree
  - nll3Reee - nll3Reeπ
  - nll3Reee - nll3Reπe
  - nll3Reee - nll3Reππ
  - nll3Reeπ - nll3Reπe
  - nll3Reeπ - nll3Reππ
  - nll3Reπe - nll3Reππ
  - nll2Reπ - nll3Reee
  - nll2Reπ - nll3Reeπ
  - nll2Reπ - nll3Reπe
  - nll2Reπ - nll3Reππ
  - nll2Rπe - nll3Reee
  - nll2Rπe - nll3Reeπ
  - nll2Rπe - nll3Reπe
  - nll2Rπe - nll3Reππ
  - nll2Ree - nll3Reee
  - nll2Ree - nll3Reeπ
  - nll2Ree - nll3Reπe
  - nll2Ree - nll3Reππ

		Signal	Bkgd	Purity	Eff	FOM
2Reπ	BDT 3*	0.59	0.62	48.7%	22.9%	0.536
	BDT 4*	0.55	0.51	52.1%	21.5%	**0.537**
	BDT 44**	0.50	0.48	50.9%	19.5%	***0.505***
2Reπ1de	BDT 2	2.58	2.35	52.4%	51.3%	1.162
	BDT 4*	2.58	2.52	50.6%	51.4%	**1.144**
	BDT 44**	2.47	2.30	51.8%	49.2%	***1.132***

* 3R nlls padded with zeros
** 3R nlls padded with 50,000

- NTrees = 850
- MaxDepth = 3

Note: Signal = true 1e1$\pi^{+/-}$ events

BDTs: Trial 5

- BDT variables:
  - nll1Re - nll1Rμ
  - nll1Re - nll2Reπ
  - nll1Re - nll2Rπe
  - nll1Re - nll2Ree
  - nll1Rμ - nll2Reπ
  - nll1Rμ - nll2Rπe
  - nll1Rμ - nll2Ree
  - nll2Reπ - nll2Rπe
  - nll2Reπ - nll2Ree
  - nll2Rπe - nll2Ree
  - nll3Reee - nll3Reeπ
  - nll3Reee - nll3Reπe
  - nll3Reee - nll3Reππ
  - nll3Reeπ - nll3Reπe
  - nll3Reeπ - nll3Reππ
  - nll3Reπe - nll3Reππ

		Signal	Bkgd	Purity	Eff	FOM
2Reπ	BDT 4*	0.55	0.51	52.1%	21.5%	0.537
	BDT 5*	0.51	0.33	61.2%	20.0%	**0.561**
	BDT 55*	0.48	0.29	62.1%	18.8%	***0.547***
2Reπ1de	BDT 2	2.58	2.35	52.4%	51.3%	1.162
	BDT 5*	2.52	1.89	57.1%	50.0%	**1.199**
	BDT 55*	2.23	1.33	62.6%	44.3%	***1.181***

- nll2Reπ - nll3Reee
- nll2Reπ - nll3Reeπ
- nll2Reπ - nll3Reπe
- nll2Reπ - nll3Reππ
- nll2Rπe - nll3Reee
- nll2Rπe - nll3Reeπ
- nll2Rπe - nll3Reπe
- nll2Rπe - nll3Reππ
- nll2Ree - nll3Reee
- nll2Ree - nll3Reeπ
- nll2Ree - nll3Reπe
- nll2Ree - nll3Reππ

* 3R nlls padded with zeros
** 3R nlls padded with 50,000

- p_e 1Re
- p_μ 1Rμ
- p_e 2Reπ
- p_π 2Reπ
- p_e 2Rπe
- p_π 2Rπe
- p_{e1} 2Ree
- p_{e2} 2Ree

- NTrees = 850
- MaxDepth = 3

Note: Signal = true 1e1$\pi^{+/-}$ events

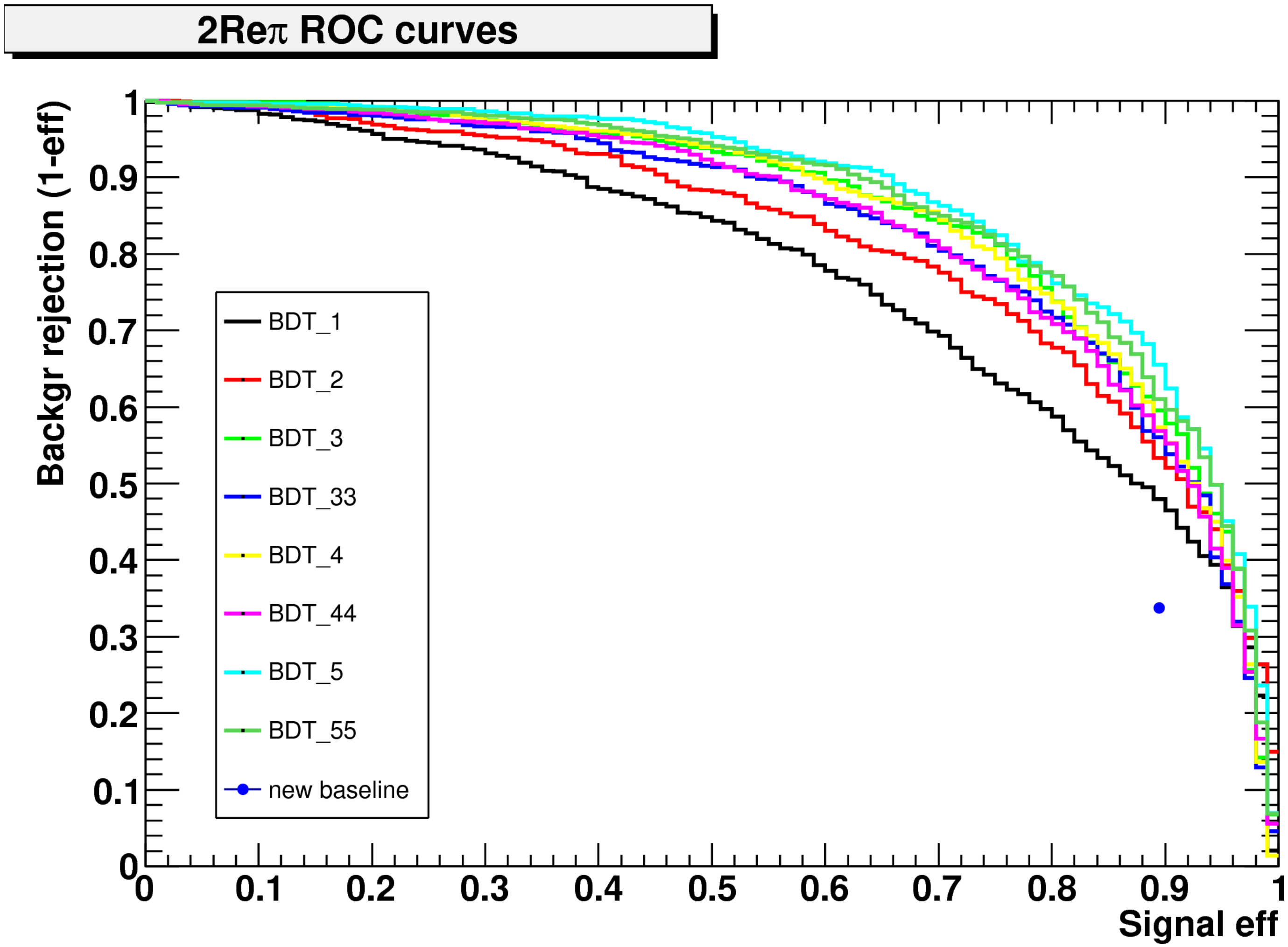
BDT Summary

		Signal	Bkgd	Purity	Eff	FOM
2Reπ	BDT 1	0.71	1.84	27.9%	27.8%	0.446
	BDT 2	0.61	0.98	38.4%	23.7%	0.484
	BDT 3*	0.59	0.62	48.7%	22.9%	0.536
	BDT 4*	0.55	0.51	52.1%	21.5%	0.537
	BDT 5*	0.51	0.33	61.2%	20.0%	0.561
2Reπ1de	BDT 1	2.52	2.64	48.8%	50.0%	1.108
	BDT 2	2.58	2.35	52.4%	51.3%	1.162
	BDT 3*	2.62	2.74	48.8%	52.0%	1.131
	BDT 4*	2.58	2.52	50.6%	51.4%	1.144
	BDT 5*	2.52	1.89	57.1%	50.0%	1.199

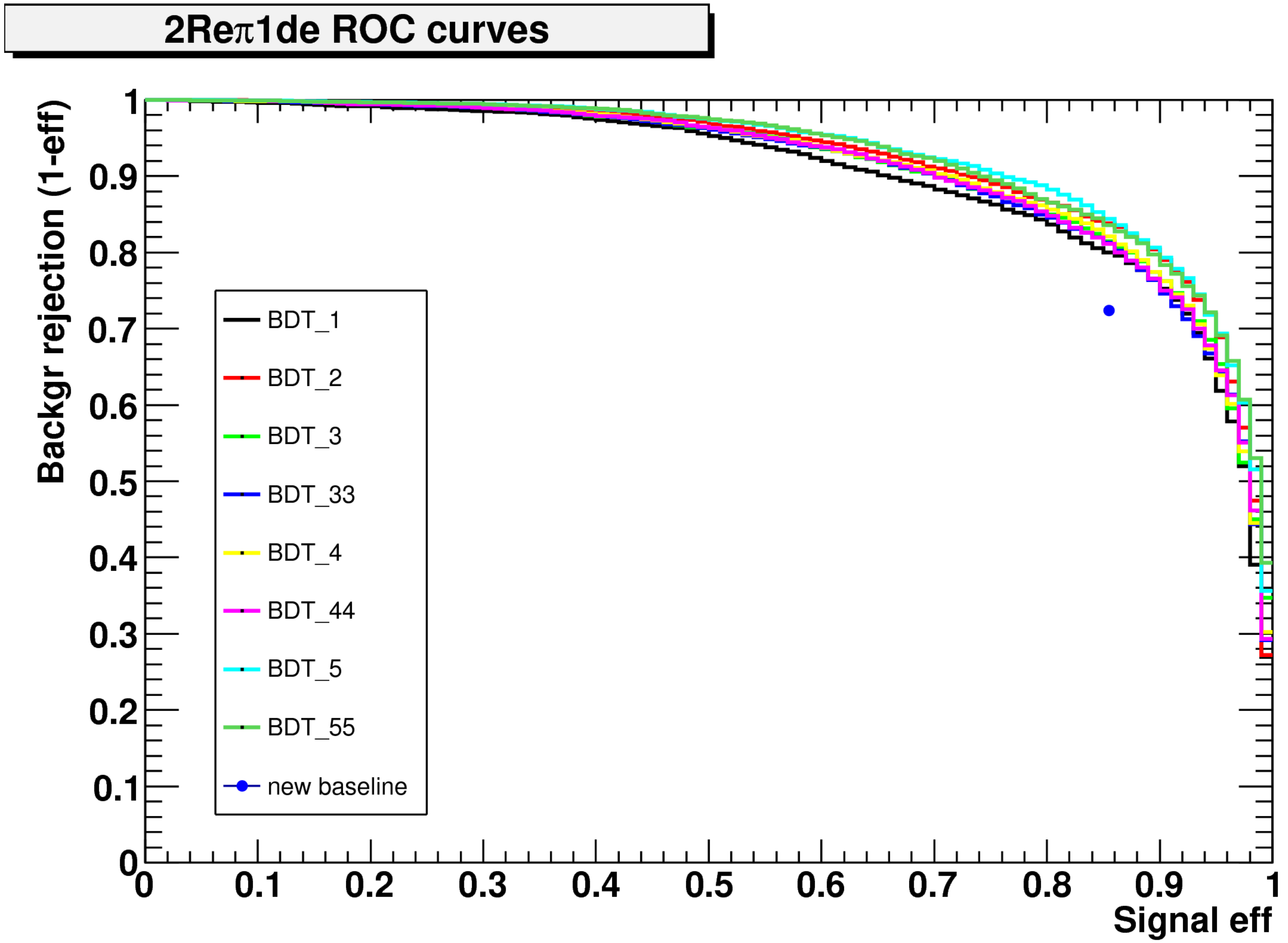
* 3R nlls padded with zeros

BDT 1: 1Rnll, 2Rnll
BDT 2: 1Rnll, 2Rnll, kinematics (1R & 2R)
BDT 3: 1Rnll, 2Rnll, 3Rnll vs 2Rnll
BDT 4: 1Rnll, 2Rnll, 3Rnll
BDT 5: 1Rnll, 2Rnll, 3Rnll, kinematics (1R & 2R)

- Comparing BDT 2 to BDT 5 shows that 3R likelihood information significantly improves selection (much more so for 2Reπ selection)
- Kinematic information seems to be more useful in 2Repi1de selection



2Reπ ROC curves
Backgr rejection (1-eff)
Signal eff
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
BDT_1
BDT_2
BDT_3
BDT_33
BDT_4
BDT_44
BDT_5
BDT_55
new baseline


2Reπ1de ROC curves
Backgr rejection (1-eff)
Signal eff
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
BDT_1
BDT_2
BDT_3
BDT_33
BDT_4
BDT_44
BDT_5
BDT_55
new baseline

Current/Future Work

- Currently working on expanding input sample to TMVA
  - Memory issue comes primarily from the input sample being too large
    - Even with <10 float variables, ran into memory issues when expanding input sample liberally
  - Currently looking for loose cuts to purify input sample while maintaining large efficiency
    - So far, while I've been able to improve input efficiency, final performance isn't as good
      - Perhaps require larger tree depth (>3) or 4R likelihood information to deal with more efficient (but less pure) input?
- Also on the to-do list:
  - Investigate how to approach systematic uncertainties when using BDTs
  - Systematic method of removing BDT variables that don't significantly benefit selection performance
  - Evaluate performance of 2Reπ fit for true $1e1\pi^{+/-}$ events where fqmrpid[0][*] ≠ $1e1\pi^{+/-}$
    - Start looking at secondary stack (today)

BDT Variable Importance

2Reπ

```
Ranking input variables (method specific)...
Ranking result (top variable is best ranked)
--------------------------------------------
Rank : Variable   : Variable Importance
--------------------------------------------
   1 : pe2_2ree   : 4.152e-02
   2 : nll8       : 4.145e-02
   3 : nll13      : 3.702e-02
   4 : nll6       : 3.630e-02
   5 : nll20      : 3.549e-02
   6 : nll19      : 3.549e-02
   7 : pmu_1rmu   : 3.276e-02
   8 : nll4       : 3.276e-02
   9 : nll1       : 3.193e-02
  10 : nll18      : 3.110e-02
  11 : nll5       : 3.100e-02
  12 : nll9       : 3.092e-02
  13 : nll3       : 3.086e-02
  14 : nll14      : 3.049e-02
  15 : nll2       : 2.993e-02
  16 : pe_2repi   : 2.974e-02
  17 : nll22      : 2.970e-02
  18 : nll11      : 2.946e-02
  19 : nll10      : 2.882e-02
  20 : pe_2rpie   : 2.804e-02
  21 : nll21      : 2.679e-02
  22 : nll26      : 2.556e-02
  23 : pe1_2ree   : 2.383e-02
  24 : nll12      : 2.321e-02
  25 : pe_1re     : 2.308e-02
  26 : nll24      : 2.200e-02
  27 : nll15      : 2.176e-02
  28 : ppi_2repi  : 2.073e-02
  29 : nll16      : 2.066e-02
  30 : nll23      : 2.049e-02
  31 : nll7       : 2.044e-02
  32 : nll17      : 1.998e-02
  33 : nll28      : 1.968e-02
  34 : nll27      : 1.965e-02
  35 : nll25      : 1.950e-02
  36 : ppi_2rpie  : 1.782e-02
--------------------------------------------
```

2Reπ1de

```
Ranking input variables (method specific)...
Ranking result (top variable is best ranked)
--------------------------------------------
Rank : Variable   : Variable Importance
--------------------------------------------
   1 : nll7       : 4.334e-02
   2 : pe_1re     : 4.152e-02
   3 : nll1       : 4.061e-02
   4 : nll9       : 3.850e-02
   5 : nll19      : 3.630e-02
   6 : pe2_2ree   : 3.581e-02
   7 : nll8       : 3.556e-02
   8 : nll4       : 3.527e-02
   9 : nll6       : 3.526e-02
  10 : pe1_2ree   : 3.425e-02
  11 : nll5       : 3.364e-02
  12 : nll13      : 3.196e-02
  13 : pmu_1rmu   : 3.085e-02
  14 : nll18      : 3.076e-02
  15 : nll10      : 3.071e-02
  16 : nll20      : 3.027e-02
  17 : nll3       : 2.985e-02
  18 : nll14      : 2.812e-02
  19 : nll22      : 2.806e-02
  20 : nll15      : 2.724e-02
  21 : pe_2repi   : 2.684e-02
  22 : nll21      : 2.659e-02
  23 : nll12      : 2.538e-02
  24 : nll17      : 2.404e-02
  25 : nll16      : 2.378e-02
  26 : nll25      : 2.356e-02
  27 : nll24      : 2.341e-02
  28 : nll11      : 2.291e-02
  29 : nll2       : 2.224e-02
  30 : nll26      : 2.155e-02
  31 : nll28      : 1.861e-02
  32 : ppi_2rpie  : 1.791e-02
  33 : pe_2rpie   : 1.719e-02
  34 : nll23      : 1.474e-02
  35 : nll27      : 1.338e-02
  36 : ppi_2repi  : 0.000e+00
--------------------------------------------
```

```
"nll1 := nll1re-nll1rmu"
"nll2 := nll1re-nll2repi"
"nll3 := nll1re-nll2rpie"
"nll4 := nll1re-nll2ree"
"nll5 := nll1rmu-nll2repi"
"nll6 := nll1rmu-nll2rpie"
"nll7 := nll1rmu-nll2ree"
"nll8 := nll2repi-nll2rpie"
"nll9 := nll2repi-nll2ree"
"nll10 := nll2rpie-nll2ree"
"nll11 := nll2repi-nll3reee"
"nll12 := nll2repi-nll3reepi"
"nll13 := nll2repi-nll3repie"
"nll14 := nll2repi-nll3repipi"
"nll15 := nll2rpie-nll3reee"
"nll16 := nll2rpie-nll3reepi"
"nll17 := nll2rpie-nll3repie"
"nll18 := nll2rpie-nll3repipi"
"nll19 := nll2ree-nll3reee"
"nll20 := nll2ree-nll3reepi"
"nll21 := nll2ree-nll3repie"
"nll22 := nll2ree-nll3repipi"
"nll23 := nll3reee-nll3reepi"
"nll24 := nll3reee-nll3repie"
"nll25 := nll3reee-nll3repipi"
"nll26 := nll3reepi-nll3repie"
"nll27 := nll3reepi-nll3repipi"
"nll28 := nll3repie-nll3repipi"
```

E_{res} using eπ/πe reconstruction