Progress Update

Trevor Towstego
UofT Neutrino/DM Meeting
June 28, 2018

T2K-SK MC Issues

- Looking towards adding more likelihood ratios (2Rμe, 3R, etc.)
  - Yoshida-san has recently re-processed T2K-SK MC
    - new ROOT files contain all fiTQun information, but don't contain other information to make file size more manageable
  - Trying combine relevant fiTQun information and other information I need from older T2K-MC files
    - In doing so, I found some discrepancies with the T2K-MC files I had been using previously, which Yoshida-san and Cris have been helping me with
- I'm using Sophie's copy of the T2K-SK 14c MC on the NEUT cluster at TRIUMF

```
Processing MC 1: added 201 files.
391706 antiNue events.
Processing MC 2: added 200 files.
200000 antiNue events.

Processing MC 1: added 201 files.
391934 signalNueb events.
Processing MC 2: added 200 files.
200000 signalNueb events.

Processing MC 1: added 201 files.
391948 signalNue events.
Processing MC 2: added 200 files.
200000 signalNue events.

Processing MC 1: added 201 files.
391812 bkgNue events.
Processing MC 2: added 200 files.
200000 bkgNue events.

Processing MC 1: added 1000 files.
979171 antiNumu events.
Processing MC 2: added 1000 files.
1000000 antiNumu events.

Processing MC 1: added 1000 files.
979404 numu events.
Processing MC 2: added 1000 files.
1000000 numu events.
```

MC 1 = previous MC files
MC 2 = new MC files with fiTQun information

Recall from January...

Previous cutflow result:

Sample	cut	numu/numub CC	intrinsic nue/nueb CC	osc nue/nueb CC	numu/numub NC	intrinsic nue/nueb NC	Signal	Bkgd	Purity	FOM
2Repi	baseline	1.48	1.00	0.88	3.17	0.11	0.88	5.76	0.13	0.342
	Erec < 1.5 GeV	0.28	0.41	0.79	2.45	0.08	0.79	3.22	0.20	0.392

New cutflow result:

Sample	cut	numu/numub CC	intrinsic nue/nueb CC	osc nue/nueb CC	numu/numub NC	intrinsic nue/nueb NC	Signal	Bkgd	Purity	FOM
2Repi	baseline	1.48	2.00	1.76	3.17	0.21	1.76	6.87	0.20	0.601
	Erec < 1.5 GeV	0.28	0.82	1.57	2.45	0.16	1.57	3.72	0.30	0.683

- It appeared that ALL nue event rates are higher by a factor of 2 after the change (not just oscillated, but intrinsic as well)
- I changed the script back to how it was before, re-ran it, and got the **same result**
- Note that the root file that this script uses has not changed since December (to my knowledge) and is stored locally on my laptop

Recall from January...

Previous cutflow result:

Sample	cut	numu/numub CC	intrinsic nue/nueb CC	osc nue/nueb CC	numu/numub NC	intrinsic nue/nueb NC	Signal	Bkgd	Purity	FOM
2Repi	baseline	1.48	1.00	0.88	3.17	0.11	0.88	5.76	0.13	0.342
	Erec < 1.5 GeV	0.28	0.41	0.79	2.45	0.08	0.79	3.22	0.20	0.392

New cutflow result:

Sample	cut	numu/numub CC	intrinsic nue/nueb CC	osc nue/nueb CC	numu/numub NC	intrinsic nue/nueb NC	Signal	Bkgd	Purity	FOM
2Repi	baseline	1.48	2.00	1.76	3.17	0.21	1.76	6.87	0.20	0.601
	Erec < 1.5 GeV	0.28	0.82	1.57	2.45	0.16	1.57	3.72	0.30	0.683

- It appeared that ALL nue event rates are higher by a factor of 2 after the change (not just oscillated, but intrinsic as well)
- I changed the script back to how it was before, re-ran it, and got the **same result**
- Note that the root file that this script uses has not changed since December (to my knowledge) and is stored locally on my laptop

MC Issue Now Fixed

- In December of 2017, someone used the “hadd” command to put events of type
  nue_x_nue,
  nuebar_x_nuebar,
  numu_x_nue, and
  numubar_x_nuebar
  into a single ROOT file
  - Unfortunately, this file was left in the same directory as the other files, so ever since then my code has been double-counting all ν_e events
- FOMs will be worse, but work done towards improving it is still viable!

```
Processing MC 1: added 200 files.
195853 antiNue events.
Processing MC 2: added 200 files.
200000 antiNue events.

Processing MC 1: added 200 files.
195967 signalNueb events.
Processing MC 2: added 200 files.
200000 signalNueb events.

Processing MC 1: added 200 files.
195974 signalNue events.
Processing MC 2: added 200 files.
200000 signalNue events.

Processing MC 1: added 200 files.
195906 bkgNue events.
Processing MC 2: added 200 files.
200000 bkgNue events.

Processing MC 1: added 1000 files.
979171 antiNumu events.
Processing MC 2: added 1000 files.
1000000 antiNumu events.

Processing MC 1: added 1000 files.
979404 numu events.
Processing MC 2: added 1000 files.
1000000 numu events.
```

Some Remaining MC Discrepancies

- Cris suggested the differing event numbers could be due to events with unphysical true vertices being removed at the fillnt stage
- With nev variable, should be able to combine MC files into new ones that contain all the information I need to further investigate BDTs
  - Currently working on code to do this
  - Some concerns about not all multi-ring information
    - Cris suggested all 200 MR fits should be there, but after briefly looking at MC files in TBrowser I'm not so sure

```
Processing MC 1: added 200 files.
195853 antiNue events.
Processing MC 2: added 200 files.
200000 antiNue events.

Processing MC 1: added 200 files.
195967 signalNueb events.
Processing MC 2: added 200 files.
200000 signalNueb events.

Processing MC 1: added 200 files.
195974 signalNue events.
Processing MC 2: added 200 files.
200000 signalNue events.

Processing MC 1: added 200 files.
195906 bkgNue events.
Processing MC 2: added 200 files.
200000 bkgNue events.

Processing MC 1: added 1000 files.
979171 antiNumu events.
Processing MC 2: added 1000 files.
1000000 antiNumu events.

Processing MC 1: added 1000 files.
979404 numu events.
Processing MC 2: added 1000 files.
1000000 numu events.
```

Fixed MC: Improved eπ-like Cut

old baseline

	osc. ν_e CC FOM	true 1e1π	other	purity	eff.	1e1π FOM	net purity	net eff.
0de	0.370	0.71	3.10	18.5%	27.5%	0.362	31.92%	32.39%
1de	0.697	1.75	2.15	45.0%	34.9%	0.888		

new baseline: 2Reπ + 2Rπe + 1Re + 2Ree + 3Reππ + 2Rμe

	osc. ν_e CC FOM	true 1e1π	other	purity	eff.	1e1π FOM	net purity	net eff.
0de	0.381	0.70	2.32	23.3%	27.4%	0.405	32.60%	44.89%
1de	0.994	2.71	4.73	36.4%	53.8%	0.992		

Efficiency: 32.4% → 44.9%
Purity: 31.9% → 32.6%

Note: I did attempt to combine the 2Rπe and 2Rμe samples into a single sample, but couldn't get performance to be better than it would be without the 2Rμe sample

Using BDTs: Attempt 1

- 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
- BDT variables:
  - nll1re-nll1rmu
  - nll1re-nll2repi
  - nll1re-nll2rpie
  - nll1re-nll2ree
  - nll1rmu-nll2repi
  - nll1rmu-nll2rpie
  - nll1rmu-nll2ree
  - nll2repi-nll2rpie
  - nll2repi-nll2ree
  - nll2rpie-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

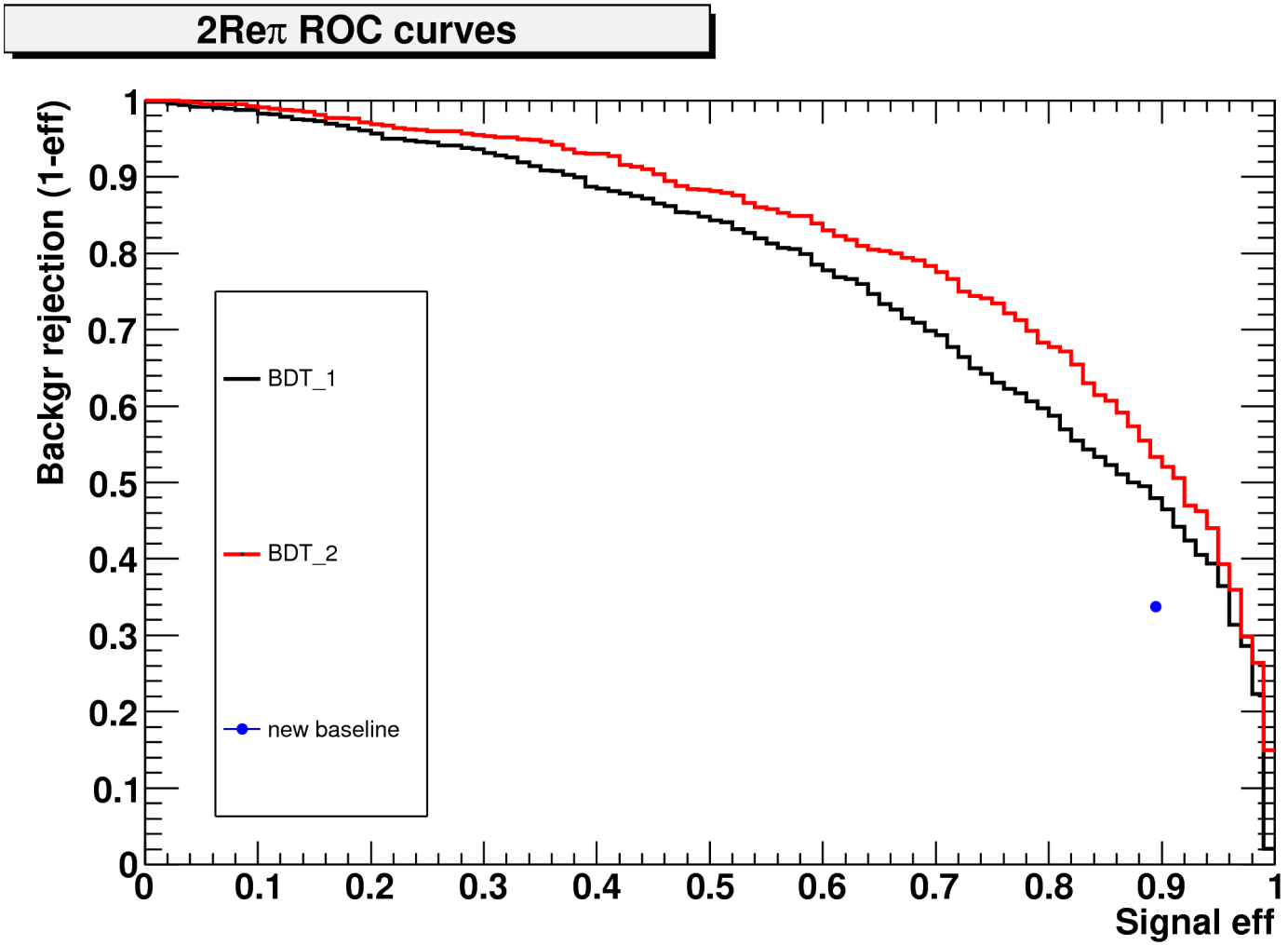
Using BDTs: Attempt 2

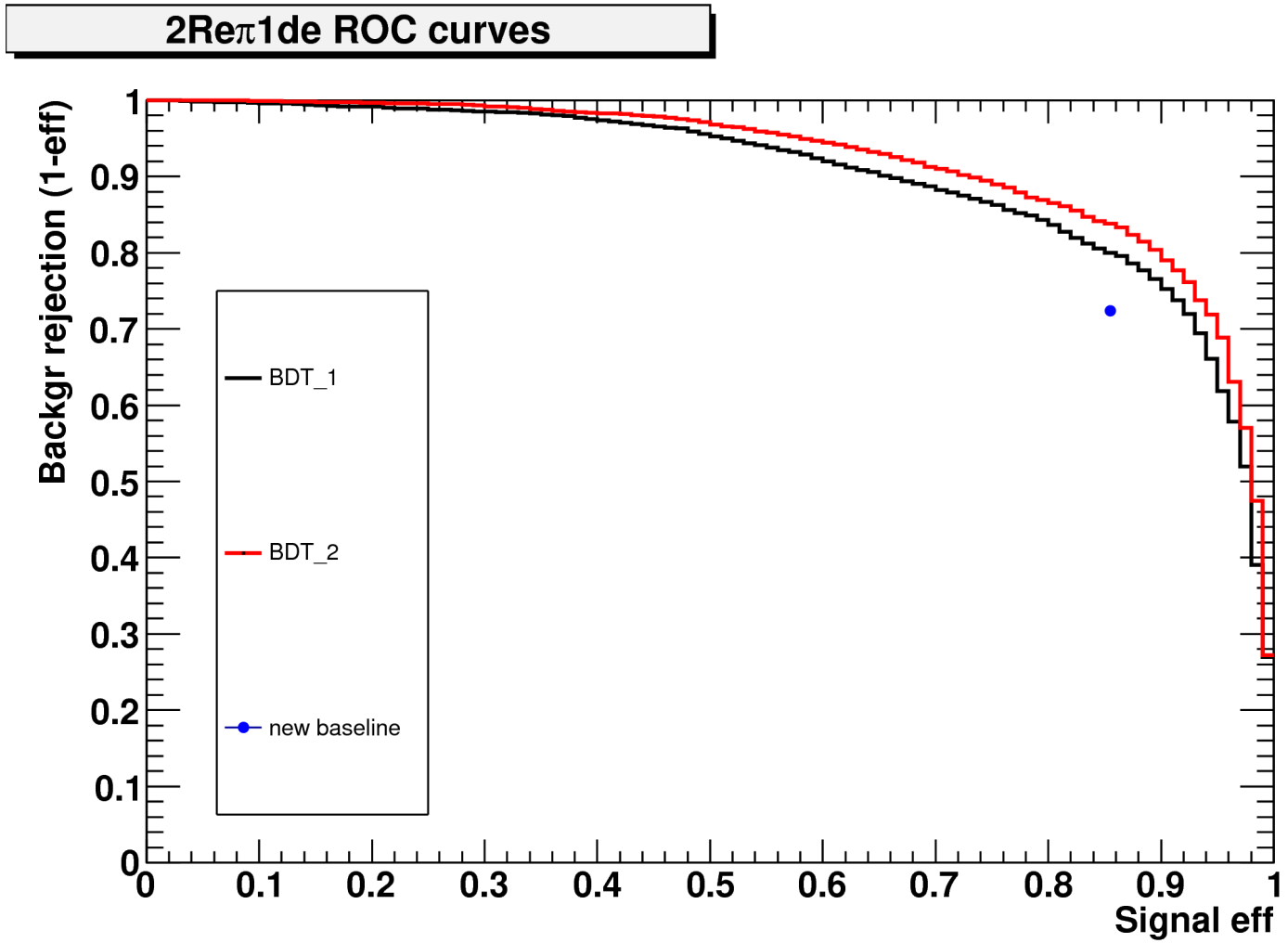
- Preliminary cuts:
  - FCFV
  - possible 2Repi
    - 0 de: i2repi==0 || i2rpie==0 || i3repipi==0
    - 1 de: (i1re==0 && !Is1re && !Is1re1de) || i2ree==0 || i2repi==0 || i2rpie==0 || i2rmue==0 || i3repipi==0
  - 1/2 sub-events
    - separate samples
  - $E_{rec}(1e,1\pi) < 1.5$ GeV
- BDT variables:
  - nll1re-nll1rmu
  - nll1re-nll2repi
  - nll1re-nll2rpie
  - nll1re-nll2ree
  - nll1rmu-nll2repi
  - nll1rmu-nll2rpie
  - nll1rmu-nll2ree
  - nll2repi-nll2rpie
  - nll2repi-nll2ree
  - nll2rpie-nll2ree
  - pe_1re
  - pmu_1rmu
  - pe_2repi
  - ppi_2repi
  - pe_2rpie
  - ppi_2rpie
  - pe1_2ree
  - pe2_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


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


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

Comments from T2K-SK Meeting

Roger:

- Should start to consider how to deal with systematics if going forward with a BDT-based analysis
  - In general, BDT seems like the right way to go, but not much precedent for dealing with BDT systematics in T2K
  - Ask some ATLAS people around the office
- Before moving fully towards BDT-based analysis:
  - Get an idea of what additional events will be present when using BDTs vs. cuts-based analysis
    - i.e. are these events actually good 2Reπ/2Rπe fits?
      - Look at E_{rec}, p_e, p_π, and vertex resolution
    - Since we are adding 3-ring (and possibly 4-ring) events, make sure I understand what fiTQun is reconstructing as additional rings
      - Look at secondaries stack