Status and performance of TauDiscriminant package

Marcin Wolter, Andrzej Zemła
IFJ PAN, Krakow, Poland

ATLAS Tau Workshop 2009
April 15 - 17
Niels Bohr Institute, Copenhagen

Status of the TauDiscriminant package

Set of ID tools running on ESD/AOD and on RDO after tauRec:

- **Tau1p3pDiscriCut –** baseline human-optimized cuts,
- **Tau1p3pDiscriCutTMVA –** cuts optimized using TMVA,
- **Tau1p3pDiscriLL –** projected likelihood ratio (TMVA),
- **Tau1p3pDiscriPDRS –** PDE_RS algorithm *(off in production),*
- **Tau1p3pDiscriNN –** Neural Network,
- **Tau1p3pEfficNN –** compensated NN (flat efficiency in E_T),
- **TauCommonLikelihood2008** – likelihood version using common variables,
- **TauLikelihood2007** – older likelihood version,
- **TauBDTAnalysis –** Boosted Decision Tree,
- **TauBDT** – Boosted Decision Tree, same variables as TauCommonLikelihood2008 *(off in production).*

Status of the package

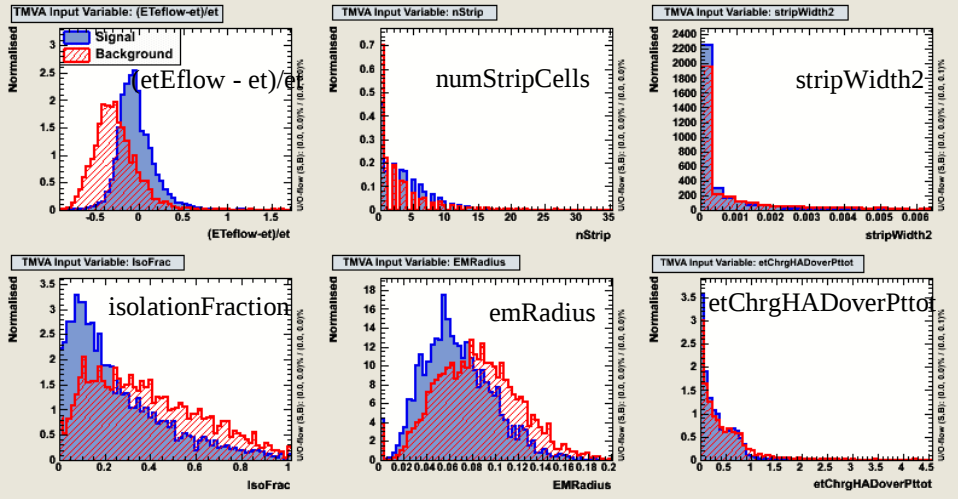
- **Tau1p3pDiscriPDRS** and **TauBDT** are switched off in production due to high CPU/memory consumption.
- New adaptive PDRS implemented – the size of the box adapts automatically to the density of training points.
- **Versions in CVS:**
  - *offline/PhysicsAnalysis/TauID/TauDiscriminant-00-00-29* – works with mc08 production
  - *offline/PhysicsAnalysis/TauID/TauDiscriminant-00-00-34* – (since ATHENA rel. 15.0.0) all the TauDiscriminant tools migrated to the new EDM. The new code is validated giving the same results.
- **New DiscriNN, EfficNN and CommonPDERS discirminants in preparation.**

New “track seed” Neural Network

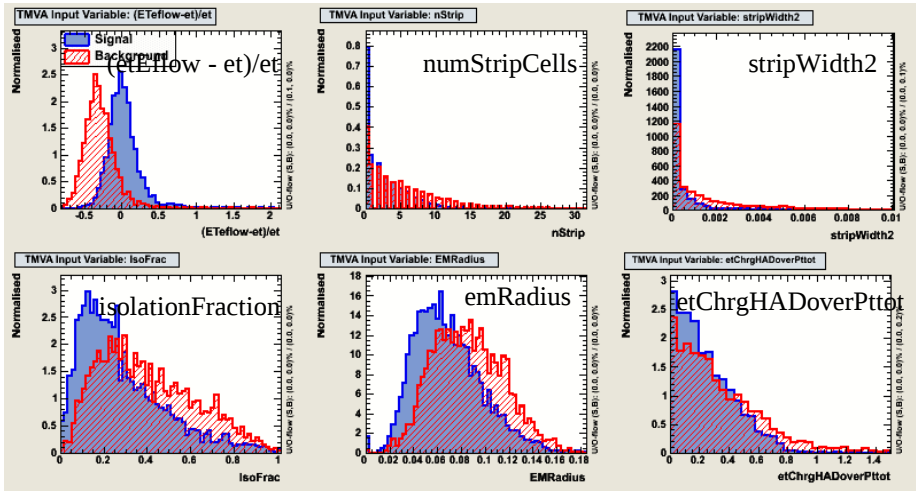
New neural network using track seed variables (tau1p3p)

- **Tau1p3pDiscriNN**
  - Trained on mc08 data
  - Three separate discriminants (instead of eight networks in the old version):
    1. 1 prong, no Pi0
    2. 1 prong, >1 Pi0
    3. Multi-prong, trained on 3-prong tau candidates
  - E_T no longer used as discriminating variable, (E_T -EtEflow)/E_T used instead
- **Tau1p3pEfficNN**
  - Flat id. efficiency in ETEflow (for W->had tau nu channel)
  - Flat discriminant distribution for signal (W had taunu)
  - **Not biasing the ntrack spectrum**

Track seed discriminating variables

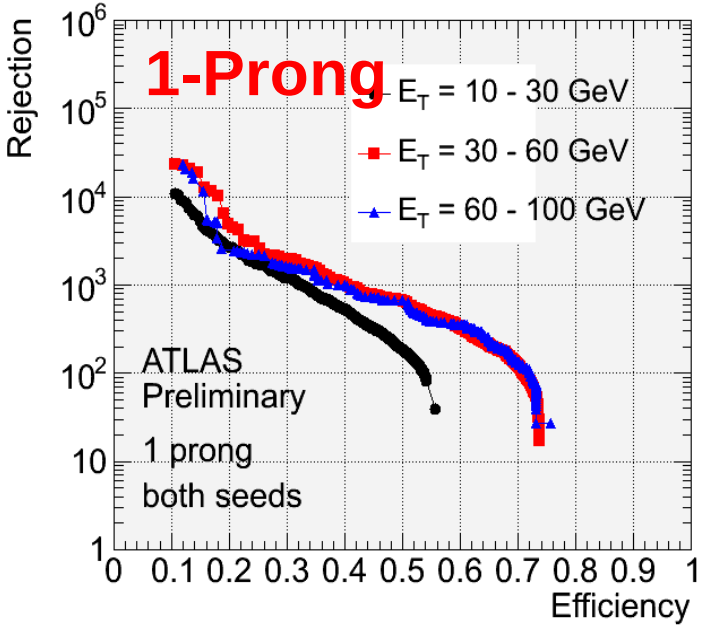


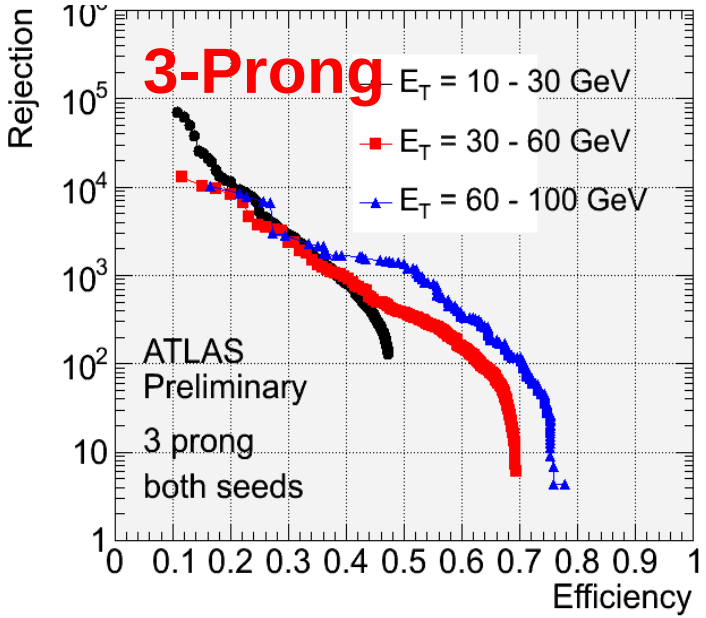
1-Prong



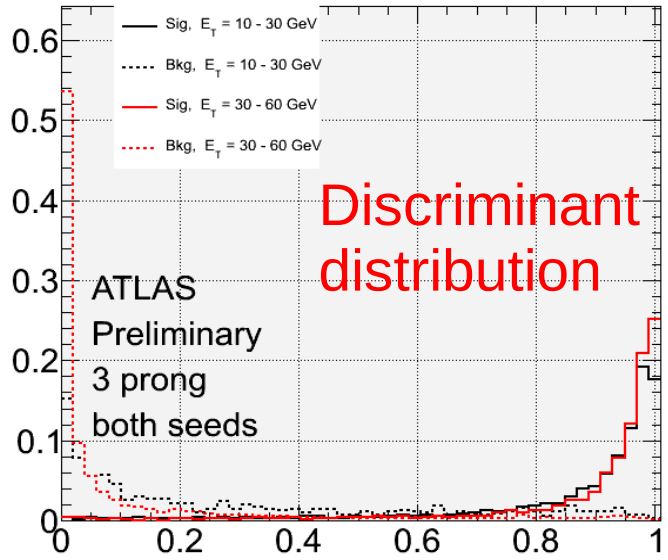
3-Prong

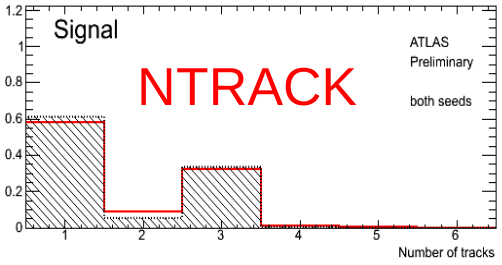
Performance of new Tau1p3pDiscriNN

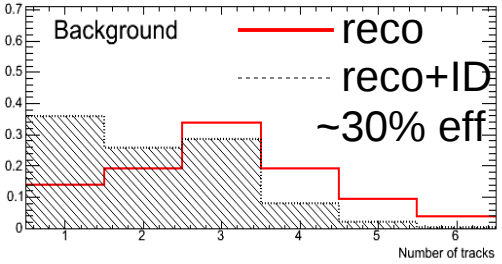




		10-30 Gev eff=30%	30-60 Gev eff=30%
Old **Tau1p3pDiscriNN**	1 prong	1900 ± 140	1700 ± 200
	3 prong	2800 ± 240	1700 ± 190
New **Tau1p3pDiscriNN**	1 prong	1200 ± 100	2100 ± 420
	3 prong	2600 ± 290	2500 ± 520



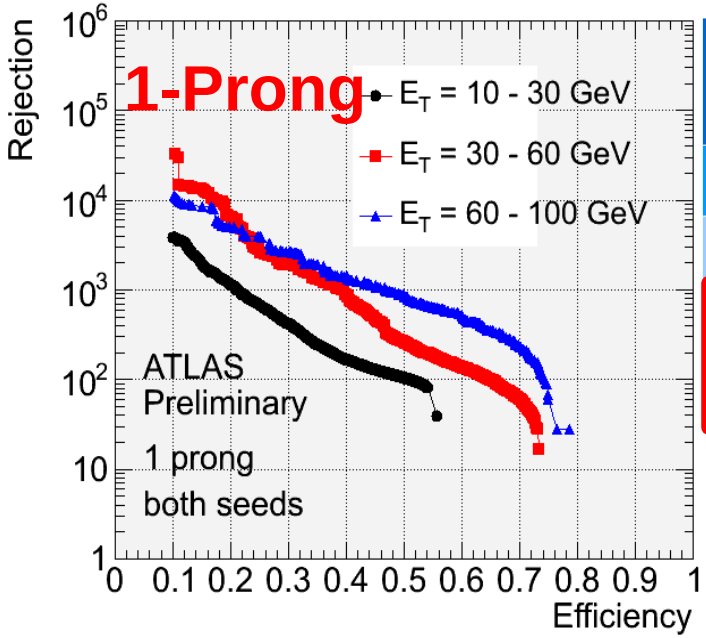


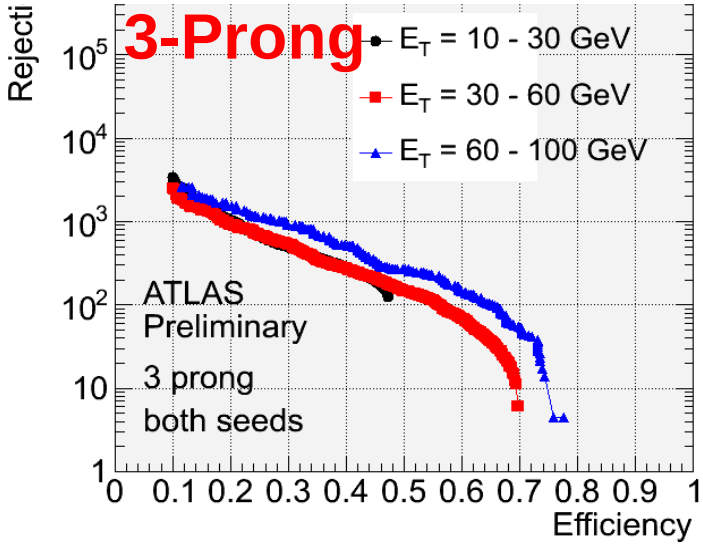


The discriminant is biasing the ntrack spectrum – to avoid this we should remove some discriminating variables for 3-prong candidates.

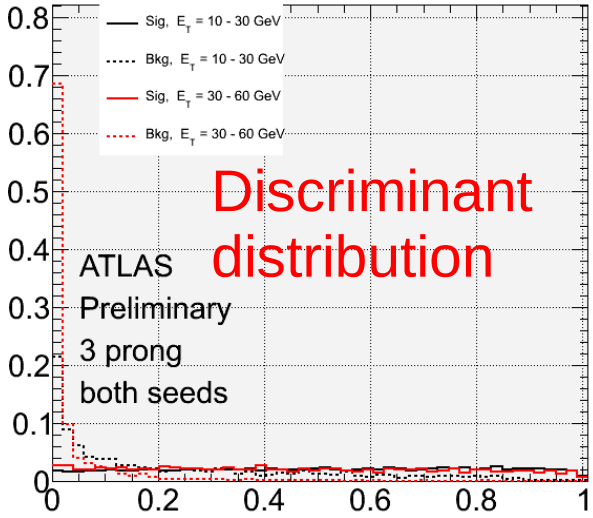
Signal: W had tau nu, Background: J0+J1+J2+J3

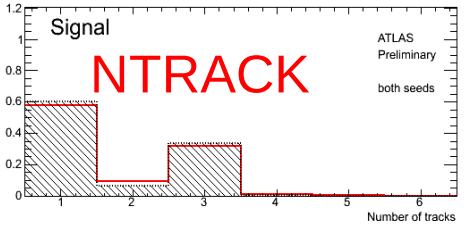
Performance of new Tau1p3pEfficNN

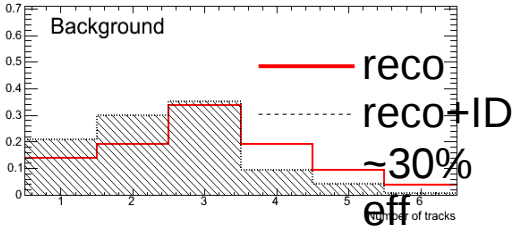




		10-30 Gev eff=30%	30-60 Gev eff=30%
Old **Tau1p3pEfficNN**	1 prong	420 ± 20	1000 ± 90
	3 prong	850 ± 50	1100 ± 100
New **Tau1p3pEfficNN**	1 prong	420 ± 20	1900 ± 370
	3 prong	500 ± 30	530 ± 50







The discriminant is NOT so much biasing the ntrack spectrum (in contrast to the old Tau1p3pEfficNN) – rejection for 3-prong candidates is reduced by removing ***rwidth2Trk3P*** and ***trFlightPathSig***.

Signal: W had tau nu, Background: J0+J1+J2+J3

New adaptive CommonPDERS

- Uses variables from both: track seed and calo seed collection (same as used by TauCommonLikelihood)
- Currently a discriminant for multiprong candidates ready (15 variables), 1-prong in preparation

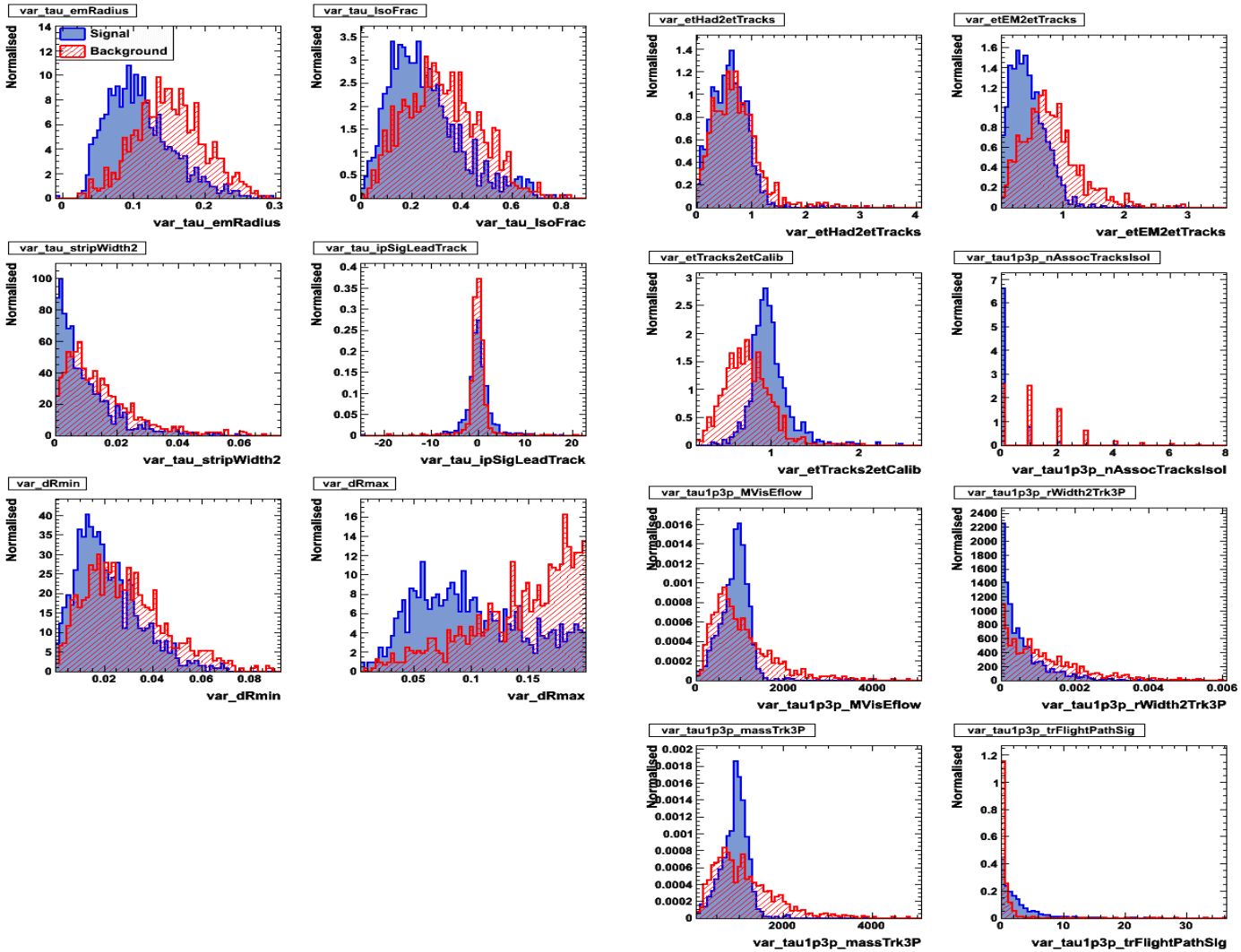
Discriminating variables

Calo seed

- EmRadius
- IsoFrac
- StripWidth2
- NumStripCells
- IpSigLeadTrack
- EtHad2etTracks
- etEM2etTracks
- EtTracks2etCalib

Track seed

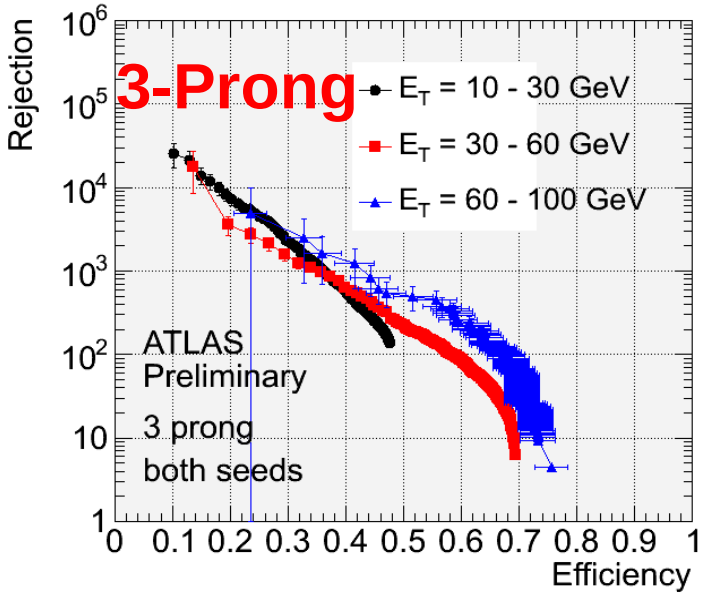
- nAssocTracksIsol
- MvisEflow
- MassTrk3P
- trFlightPathSig
- dRmin
- dRmax

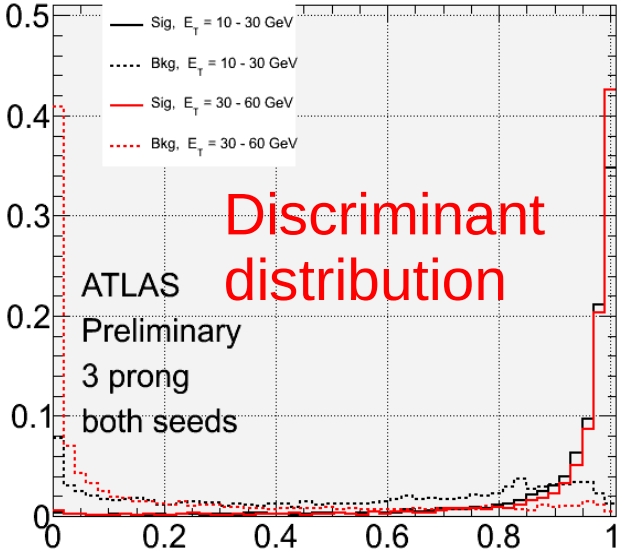


Performance of new TauCommonPDRS

Preliminary results

		10-30 Gev eff=30%	30-60 Gev eff=30%
TauCommonPDERS	3 prong	2300 ± 100	1700 ±250
TauCommonLikelihood	3 prong	2100 ± 160	2700 ±390





Signal: W had tau nu, Background: J0,J1,J2,J3

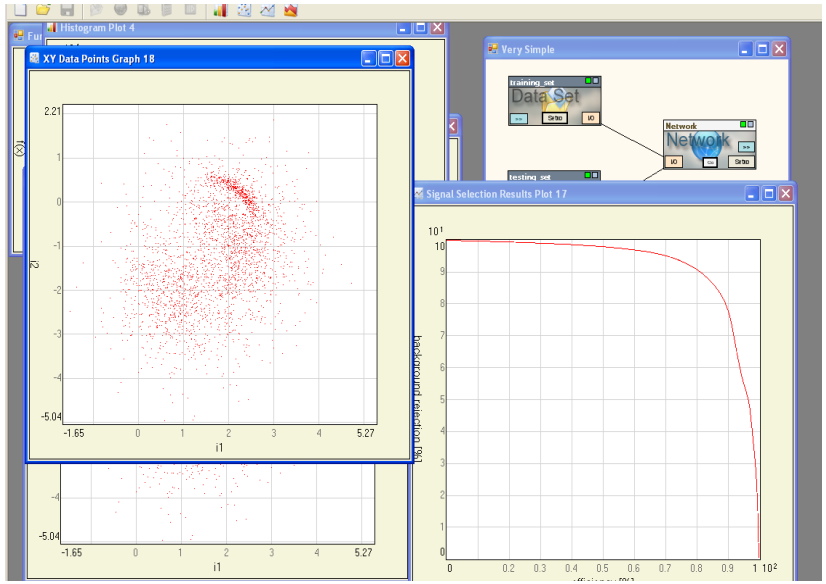
- *3-prong performs well, but for higher p_T still worse than TauCommonLikelihood*
- *Problems with 1-prong candidates - not enough reference events*
- *Still plenty of work...*

Future plans

- Finishing CommonPDERS optimization
- Common Neural Network – might be worthwhile to include, since PDERS must be off during production due to high CPU and memory consumption.
- Use a new tool for that – NetMaker is building a NN with number of hidden layers/nodes fitted to the complexity of the problem:

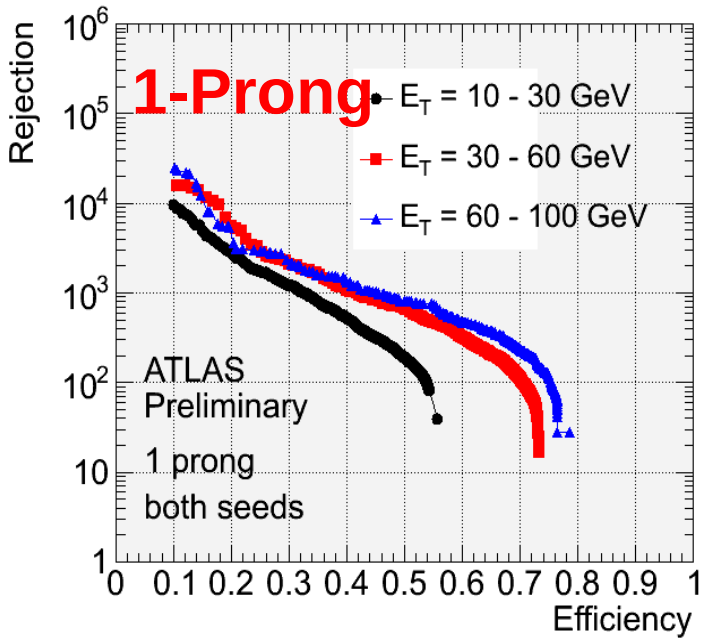
R. Sulej:
http://www.ire.pw.edu.pl/~rsulej/NetMaker/

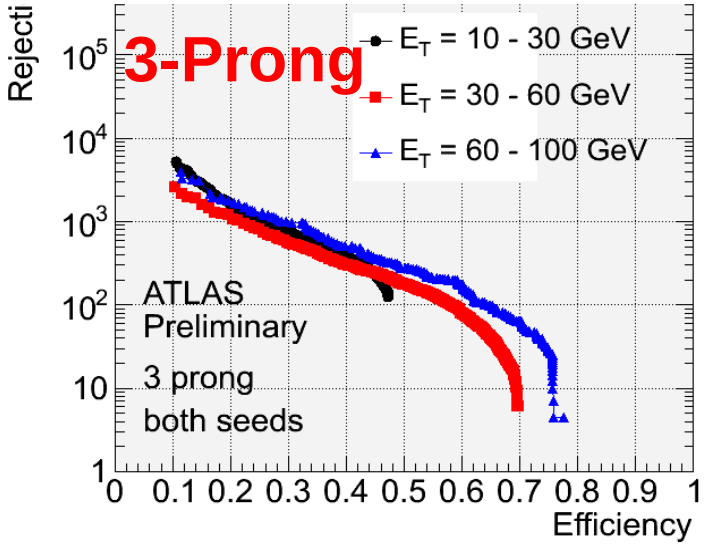
- We plan to switch from *Cone4H1TopoJets* to “antiKt” jet definition. Since all the discriminants were trained against TopoJets, this requires verification and probably retraining.



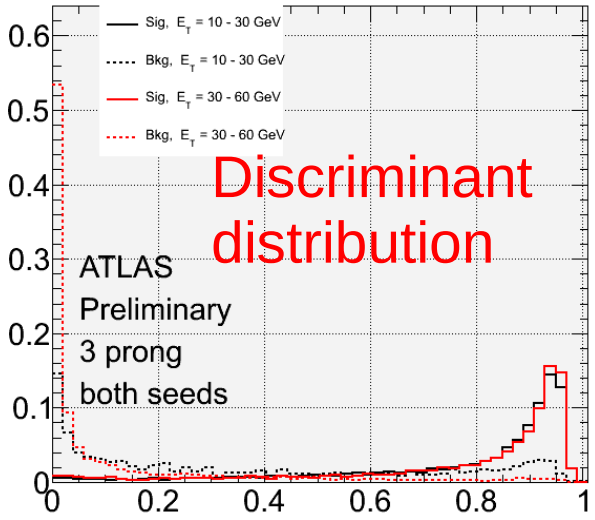
Backup slides

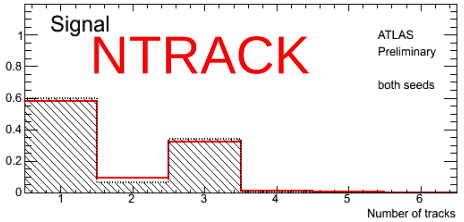
Performance of new Tau1p3pDiscriNN ***rwidth2Trk3P*** and ***trFlightPathSig*** removed.

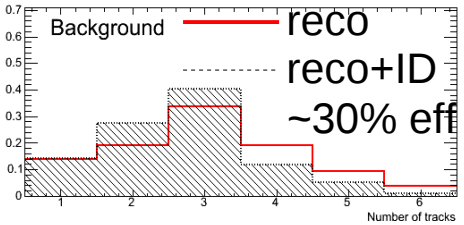




		10-30 Gev eff=30%	30-60 Gev eff=30%
Tau1p3pDiscriNN	1 prong	1200 ± 110	2000 ± 440
(all variables)	3 prong	3000 ± 370	2400 ± 610
Tau1p3pDiscriNN	1 prong	1200 ± 100	2100 ± 420
(rwidth2Trk3P and trFlightPathSig removed)	3 prong	770 ± 50	570 ± 60



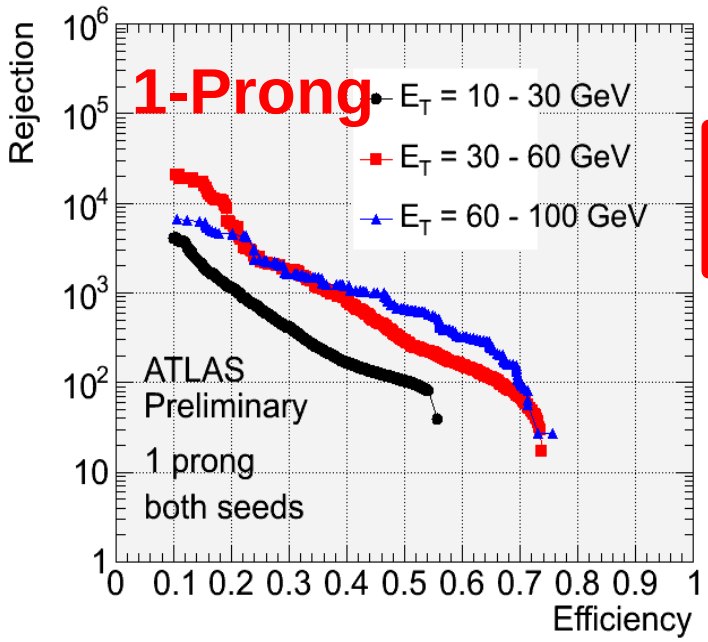


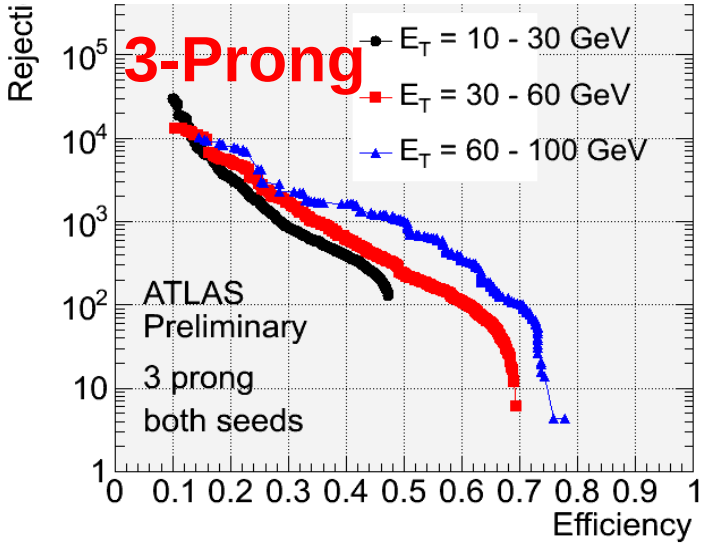


The discriminant is NOT significantly biasing the ntrack spectrum – rejection for 3-prong candidates is reduced by removing ***rwidth2Trk3P*** and ***trFlightPathSig***.

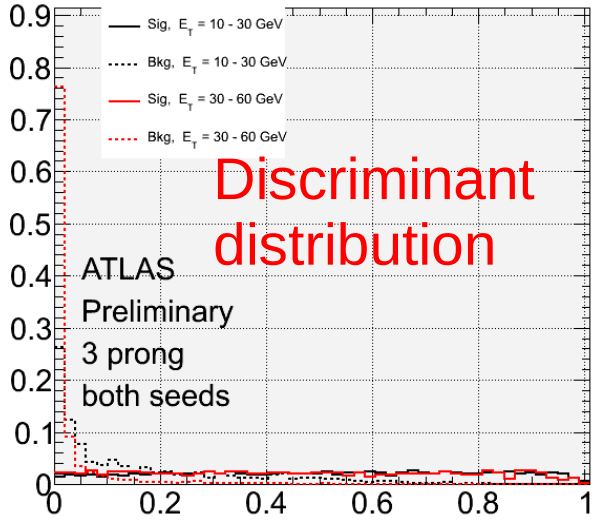
Signal: W had tau nu, Background: J0+J1+J2+J3

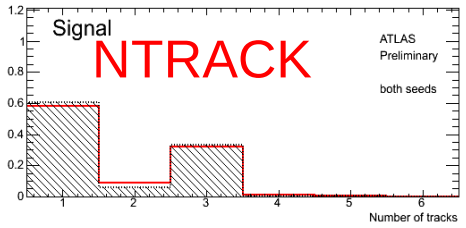
Performance of new Tau1p3pEfficNN
all variables used

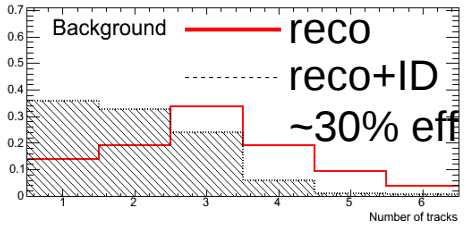




		10-30 Gev eff=30%	30-60 Gev eff=30%
Tau1p3pEfficNN	1 prong	420 ± 20	1900 ± 370
(all variables)	3 prong	820 ± 60	1700 ± 310
Tau1p3pEfficiNN	1 prong	420 ± 20	1900 ± 370
(rwidth2Trk3P and trFlightPathSig removed)	3 prong	500 ± 30	530 ± 50





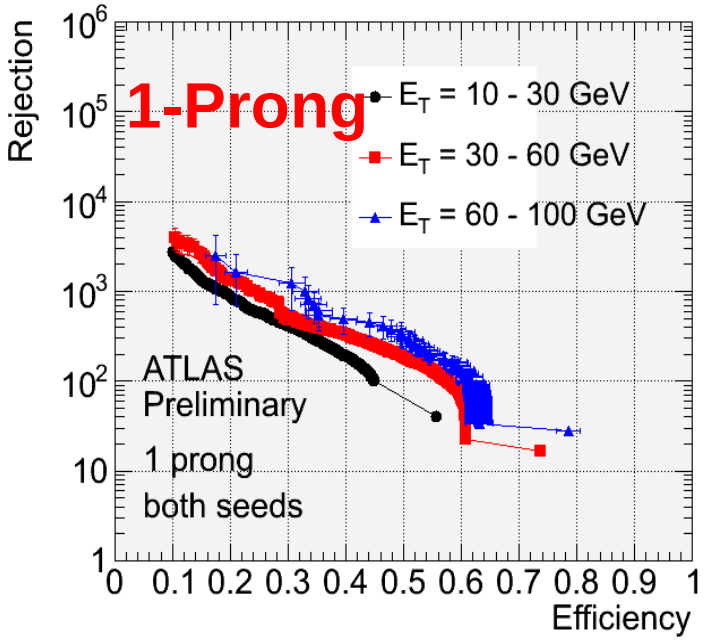


The discriminant is BIASING the ntrack spectrum

Signal: W had tau nu, Background: J0+J1+J2+J3

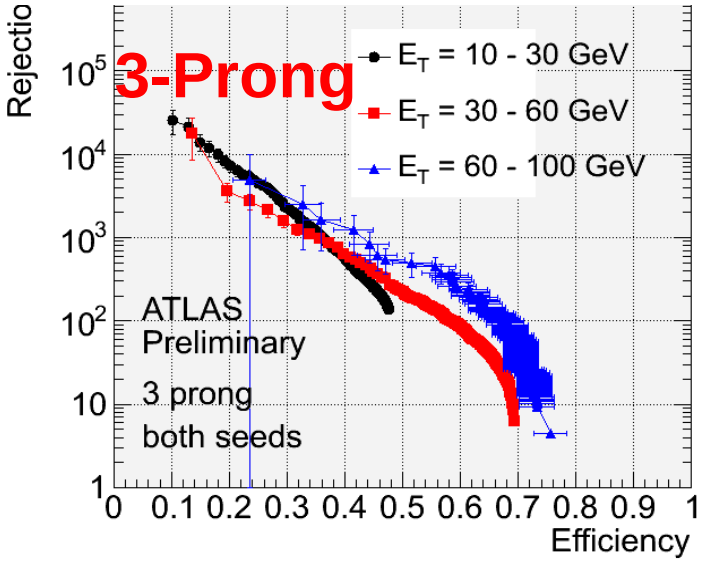
Performance of new TauCommonPDRS

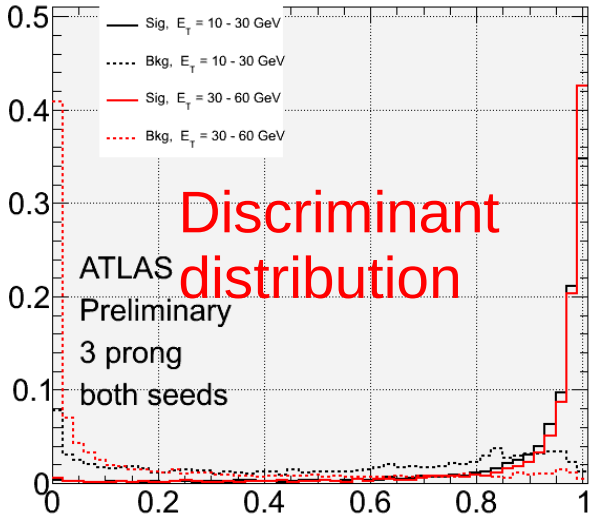
Preliminary results

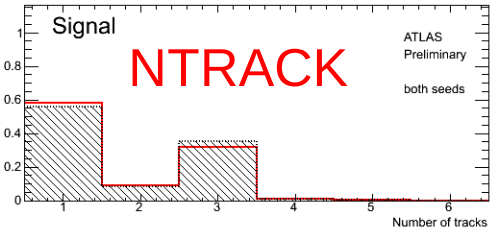


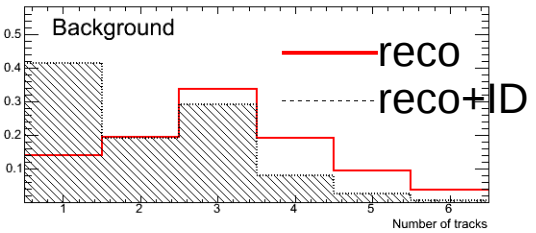
		10-30 Gev eff=30%	30-60 Gev eff=30%
TauCommonPDERS	1 prong	410 ± 20	530 ± 90
	3 prong	2200 ± 170	1600 ±250

- *1-prong performance is not satisfactory - not enough reference events*
- *3-prong performs better, but also not as expected*









Signal: W had tau nu, Background: J0,J1,J2,J3

Algorithm	Tracks	10 - 30 GeV		30 - 60 GeV		60 - 100 GeV	
		eff [%]	rej	eff [%]	rej	eff [%]	rej
Tau1p3pDiscriCut	1 prong	36	260 ± 8	52	120 ± 3	38	330 ± 20
	3 prong	35	470 ± 20	61	30 ± 1	55	40 ± 1
Tau1p3pDiscriCutTMVA	1 prong	27	1100 ± 60	59	150 ± 4	66	170 ± 9
	3 prong	39	460 ± 20	59	40 ± 1	60	60 ± 1
Tau1p3pDiscriLL	1 prong	30	920 ± 50	30	1200 ± 110	30	1200 ± 150
	3 prong	30	1200 ± 70	30	490 ± 30	30	480 ± 40
Tau1p3pDiscriPDRS	1 prong	30	1400 ± 90	30	1100 ± 100	30	630 ± 60
	3 prong	30	1500 ± 90	30	690 ± 50	30	450 ± 40
Tau1p3pDiscriNN	1 prong	30	1900 ± 140	30	1700 ± 200	30	1100 ± 140
	3 prong	30	2800 ± 240	30	1700 ± 190	30	1600 ± 260
Tau1p3pEfficNN	1 prong	30	420 ± 20	30	1000 ± 90	30	2200 ± 360
	3 prong	30	850 ± 50	30	1100 ± 100	30	5100 ± 1300
TauCommonLikelihood	1 prong	30	740 ± 40	30	1700 ± 170	30	3400 ± 720
	3 prong	30	2100 ± 160	30	2700 ± 390	30	14900 ± 4800
TauLikelihood	1 prong	30	360 ± 10	30	1200 ± 130	30	4300 ± 920
	3 prong	30	610 ± 30	30	340 ± 20	30	530 ± 40
TauBDTAnalysis	1 prong	30	870 ± 50	30	3000 ± 380	30	3700 ± 770
	3 prong	30	4200 ± 450	30	3100 ± 480	30	5400 ± 1400
TauBDT	1 prong	30	300 ± 10	30	630 ± 120	30	1700 ± 270
	3 prong	30	570 ± 20	30	370 ± 20	40	1000 ± 120

Signal: W had tau nu, Background: J0, J1,J2,J3, J4