

Weekly Report

Shuiting Xin

July.21.2020

- Overview

- Part 1

- Fake Lepton composition

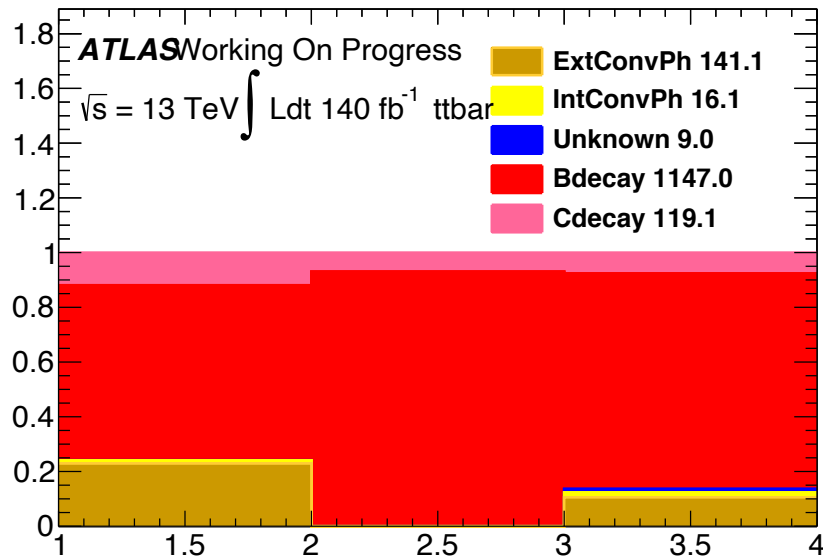
- Part 2

- Real and Fake Efficiency

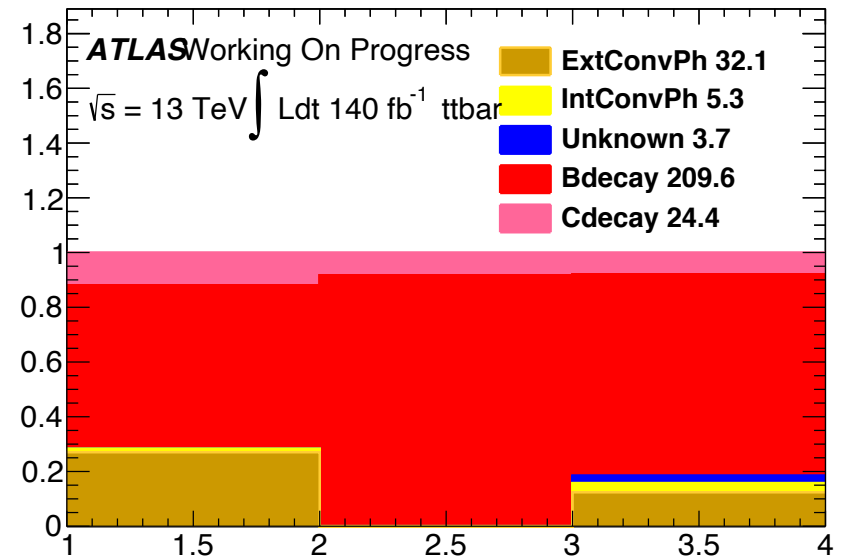
Part1 Fake Lepton composition

- Use ttHML variables, truth lepton type and origin for classification
- Five categories are considered
 - External Conversion
 - Internal Conversion
 - B decay(truth origin = 26,29,33)
 - C/Other decay(truth type =0,17)
 - Rest unknown
- Two CR choices for measurement
 - 1 || 2 jet, 0 bjet
 - 1 || 2 jet, 1 bjet

Leading lepton



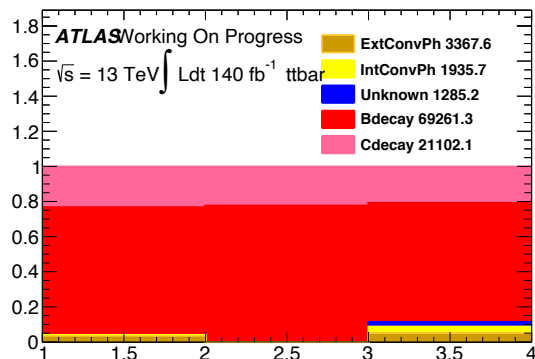
CR 1||2 jet



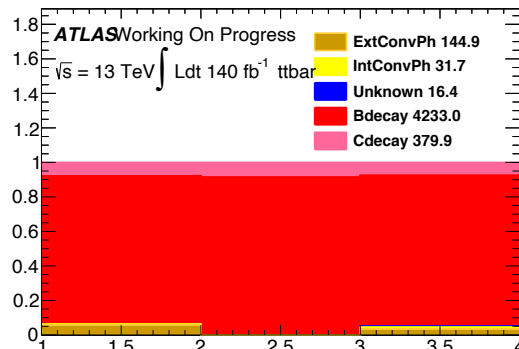
CR ≥ 3 jet

- Fake composition looks similar for CR and SR
- QMisID==0 required for both leptons
- Contains all loose lepton

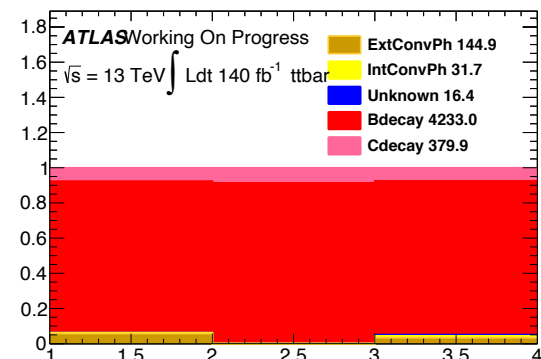
Subleading lepton



CR 1 || 2 jet ,1 b



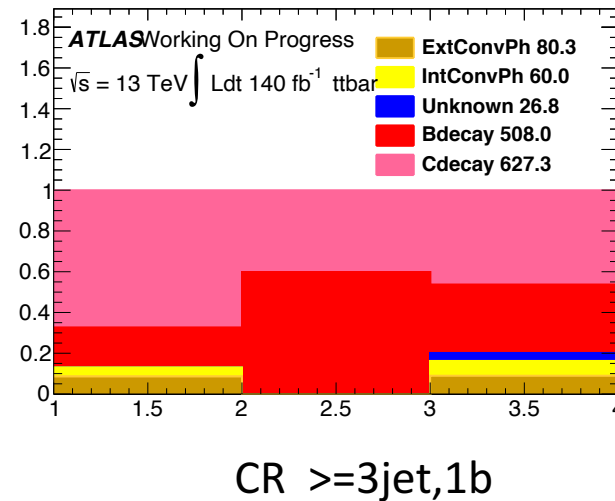
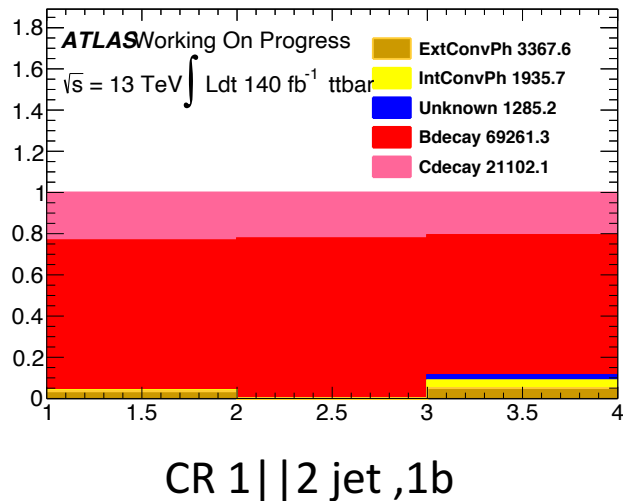
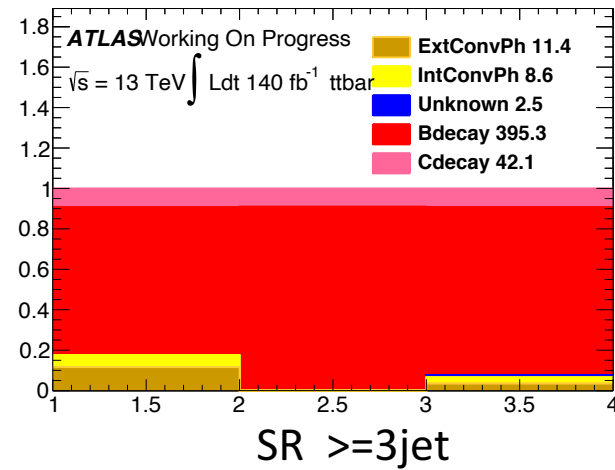
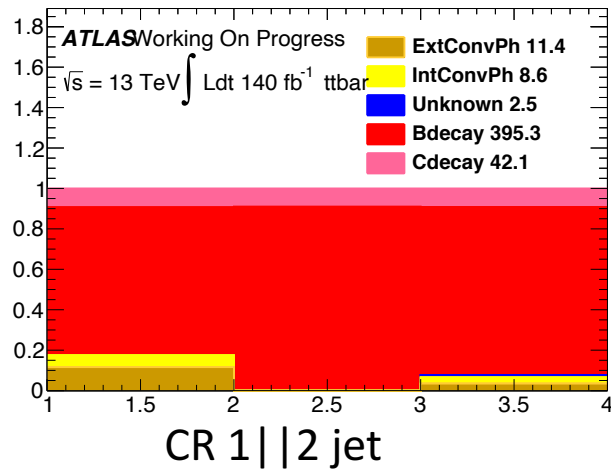
CR 1 || 2 jet



CR >=3jet

- Fake composition looks similar for CR no bjet and SR

More plots in 1/0bjet CR requiring 2 tight lepton



- Clearly to see more statistics with bjet
- Find fake composition has a strong dependence on Nbjet

Real and Fake Efficiency update

- Previously select e/mu by its charge
- Switch to tag lepton pt

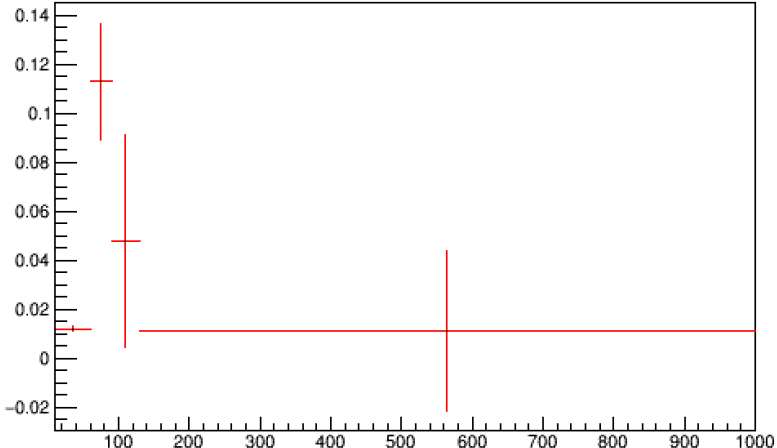
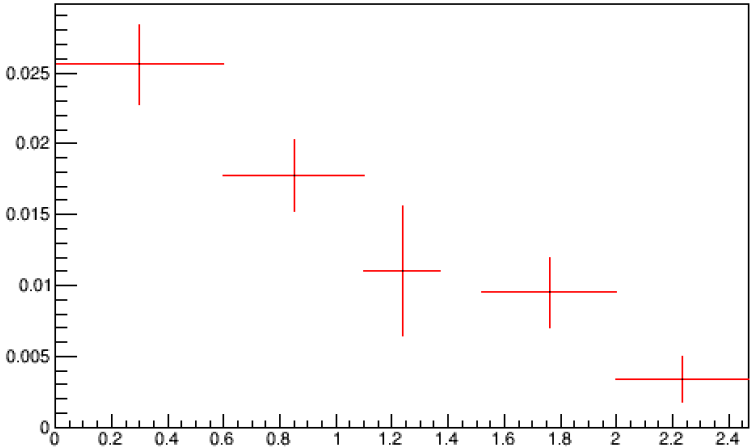
$e^{\pm}e^{\mp}$		$\mu^{\pm}\mu^{\mp}$		$e^{\pm}\mu^{\pm}, \mu^{\pm}e^{\pm}$		$\mu^{\pm}\mu^{\pm}$	
Tag	probe	Tag	Probe	Tag	Probe	Tag	Probe
Tight	Loose	Tight	Loose	Tight	Loose	Tight	Loose
Leading pt	Subleading pt	Leading pt	Subleading pt			Leading pt	Subleading pt

For example, to calculate muon fake eff

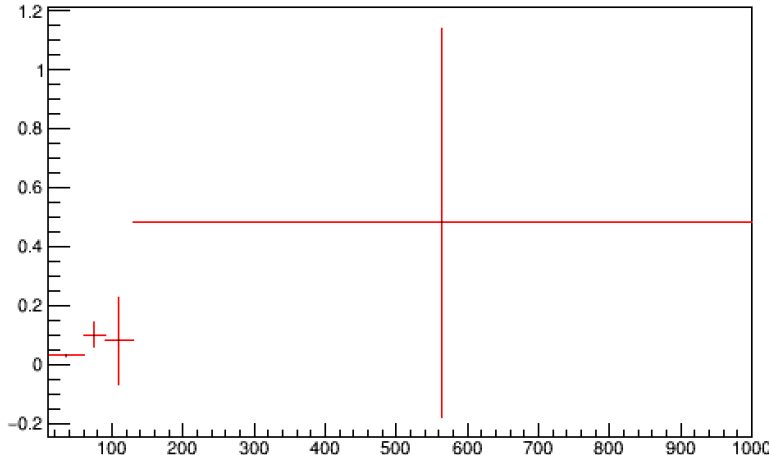
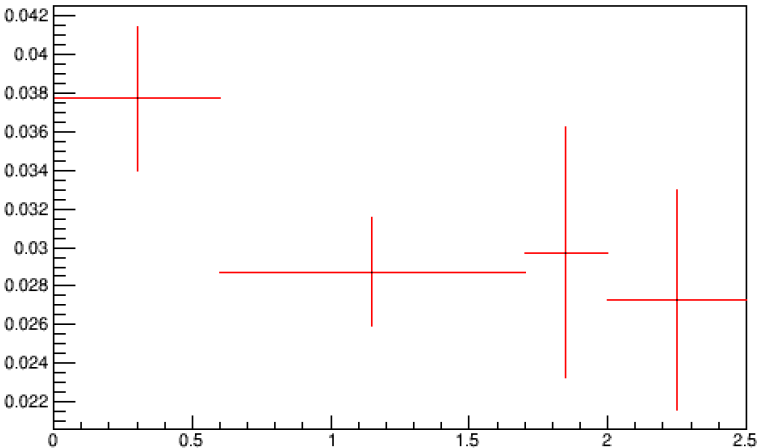
$$eff = \frac{N_{subleading,Tight}^{Data} - N^{MC}}{N_{subleading,Loose}^{Data} - N^{MC}}$$

Fake Efficiency 0b jet CR

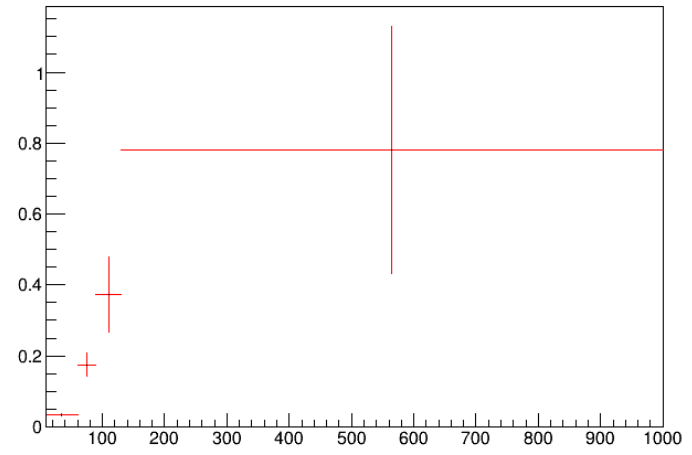
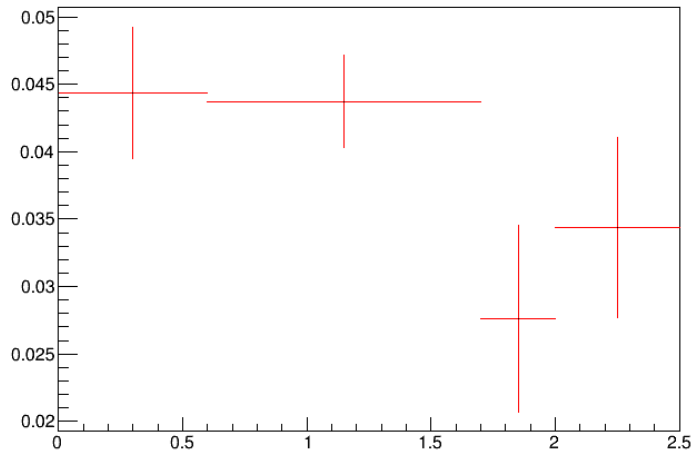
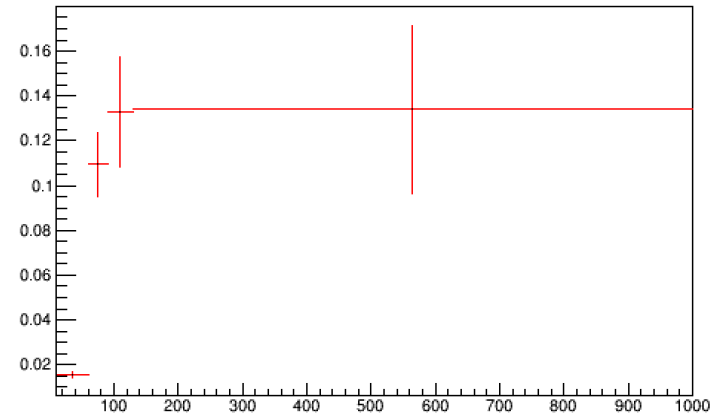
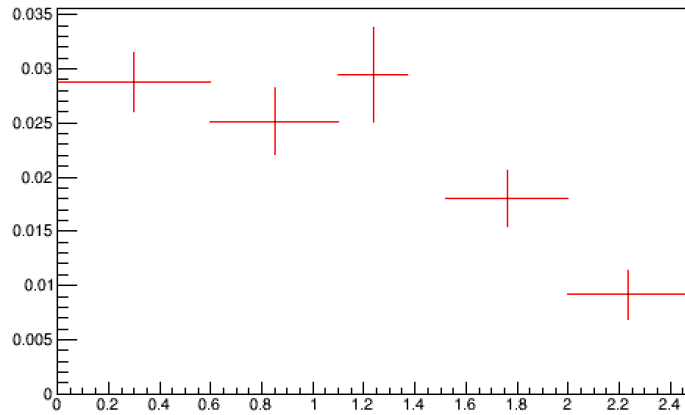
Electron



muon



Fake efficiency 1b jet CR



ML experience from Bihuan

- Generally, Xgboost don't provide much improvement for most of the cases
- DNN can handle multi-classification task well, but for binary , it doesn't show advantages with respect to BDT. Someone claimed that applying DNN with low-level variables(lepton pt, eta, jet pt ...) has some improvements
- Suggestion
 - Avoid training with data-driven bkg
 - Train two/three classifier and combine them
 - The combine method is called maximum likelihood ratio

Plan

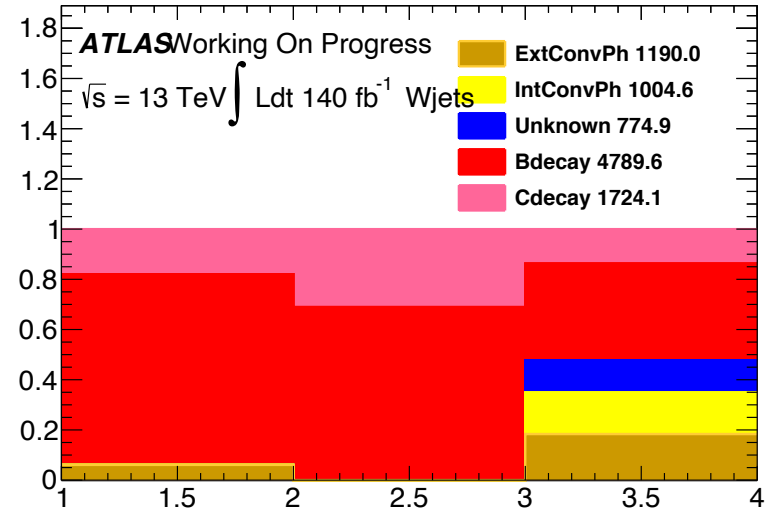
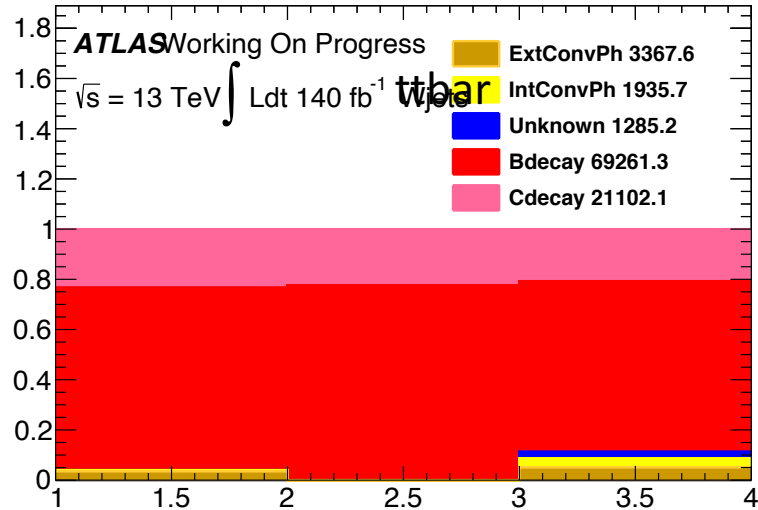
- How to validate efficiencies?
- CR checks and Optimazition.
- MM Closure test

Summary

- Fake estimation learning from ttW measurement
- Still much of the detailed work to be done
- Bruce tells me he starts to finalize the paper.

W+jets check

1 bjet CR



0 bjet CR

