

NTUA Top Tagger

Tag & Probe methodology



Overview

- BDT Input and Output in the SR_B Region
 - SR_B : Baseline selection + tight Mass Cut (120,220) GeV, **no TopTagger** Selection
 - Leading + subleading in different p_T regions: [400,600], [600,800], [800,1200]
 - **Find** Data vs MC Input and Output for UL our Analysis [here](#)
- Top Tagger Scale Factors
 - Data is subtracted QCD and Subdominant bkgs (MC) so that the data sample is pure

$$efficiency = \frac{Tight \ \& \ SR}{Tight \ \& \ Probe} = \frac{\# (1 \ jet \ pass \ baseline + Tight \ TopTagger \ Cut \ AND \ 1 \ jet \ pass \ SR)}{\# (1 \ jet \ pass \ baseline + Tight \ TopTagger \ Cut \ AND \ 1 \ jet \ pass \ only \ baseline)}$$

- Implemented Randomization (check random jet) to fill histogram to avoid p_T bias
- Divide the phase space into p_T regions: [400-600]GeV, [600-800]GeV, [800-Inf]GeV
- For the QCD estimation, we perform a fit in both regions (Tight & Probe, Tight & SR):
 - Shape of QCD is estimated from Data while inverting btagging requirement
 - # QCD events in each region is calculated from fit using the Leading JetMassSoftDrop variable
 - To scale the ttbar → fit the Leading JetMassSoftDrop in each region and get the signal strength
 - For the evaluation of Signal distribution from data, we do the following:

$$\forall \ region: S(x) = D(x) - N_{QCD}d_0(x) - Sub.Bkg(x)$$

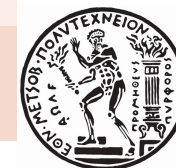


Signal Selection

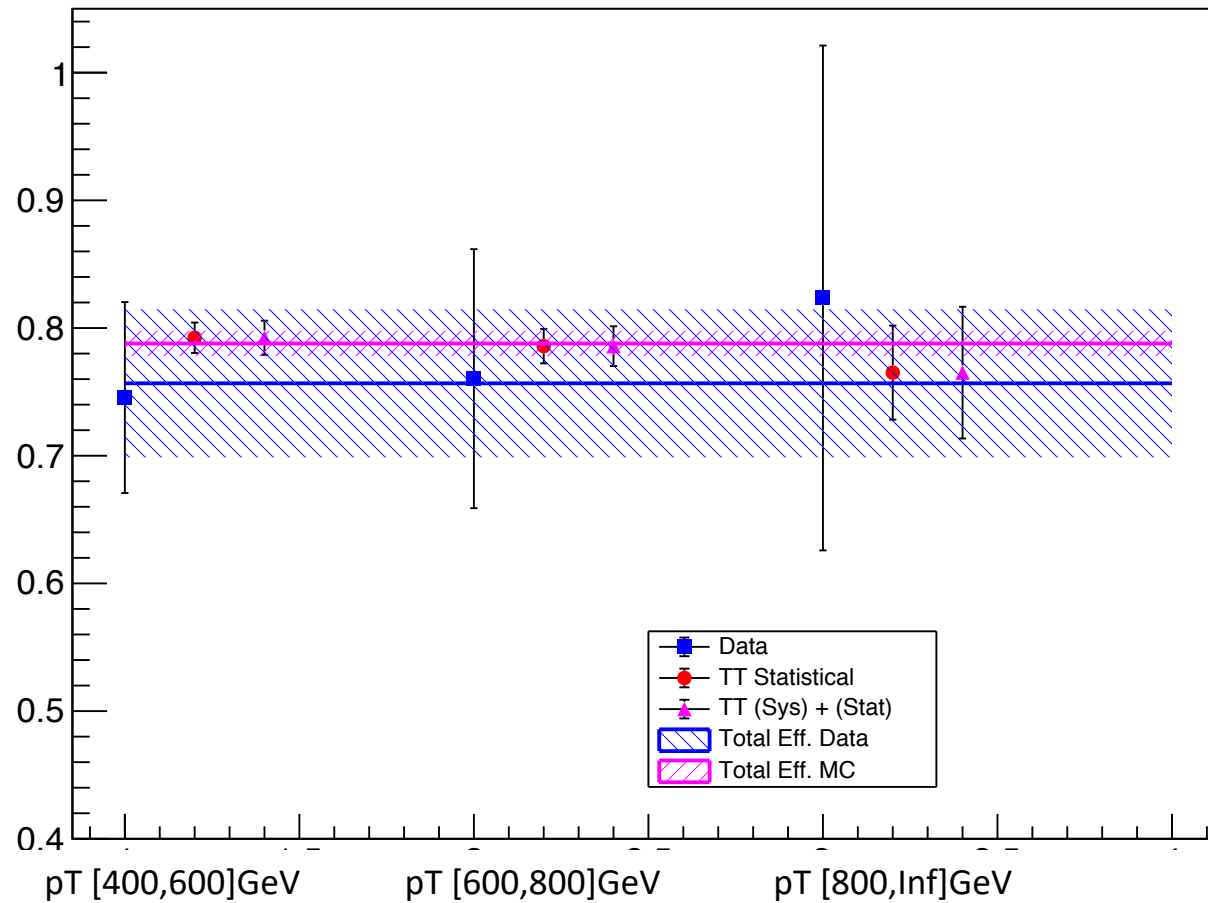
Variables	Selected Cut
pT leading jets	> 450 GeV
pT 2 nd leading jets	> 400 GeV
Njets	> 1
N leptons	= 0
eta (both leading jets)	< 2.4
mJJ	> 1000 GeV
jetMassSoftDrop (only for fit)	(50,300) GeV
Top Tagger	> 0.2
B tagging (2 btagged jets)	> Medium WP
Signal Trigger	

Control Region Selection

Variables	Selected Cut
pT leading jets	> 450 GeV
pT 2 nd leading jets	> 400 GeV
N leptons	= 0
eta (both leading jets)	< 2.4
mJJ	> 1000 GeV
jetMassSoftDrop (only for fit)	(50,300) GeV
Top Tagger	> 0.2
B tagging (0 btagged jets)	< Medium WP
Control Trigger	



TagAndProbe Efficiency per Pt region (2016 preVFP)



Efficiency--

eff data: 0.757 ± 0.058

eff ttbar: $0.788 \pm (\text{stat}) 0.009 \pm (\text{stat} + \text{systematic}) 0.01$

Efficiency per Pt region

eff data pT[400-600]: 0.746 ± 0.075

eff ttbar pT[400-600]: $0.792 \pm (\text{stat}) 0.012 \pm (\text{stat} + \text{systematic}) 0.014$

eff data pT[600-800]: 0.76 ± 0.101

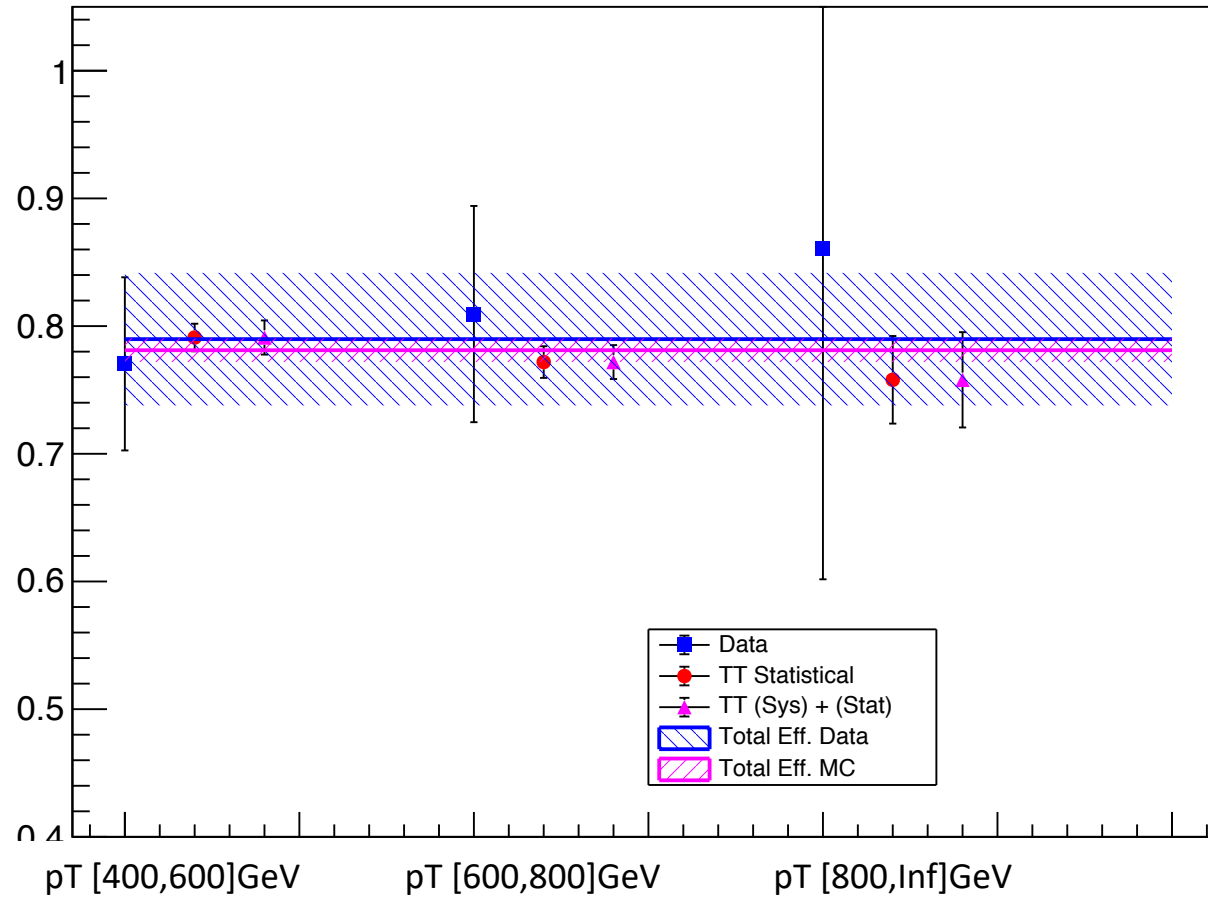
eff ttbar pT[600-800]: $0.786 \pm (\text{stat}) 0.014 \pm (\text{stat} + \text{systematic}) 0.016$

eff data pT[800-Inf]: 0.824 ± 0.197698

eff ttbar pT[800-Inf]: $0.765 \pm (\text{stat}) 0.037 \pm (\text{stat} + \text{systematic}) 0.052$



TagAndProbe Efficiency per Pt region (2016 postVFP)



ciency--

data: 0.79 ± 0.052

ttbar: $0.781 \pm (\text{stat}) 0.008 \pm (\text{stat} + \text{systematic}) 0.009$

ciency per Pt region

data pT[400-600]: 0.77 ± 0.068

ttbar pT[400-600]: $0.791 \pm (\text{stat}) 0.011 \pm (\text{stat} + \text{systematic}) 0.013$

data pT[600-800]: 0.809 ± 0.085

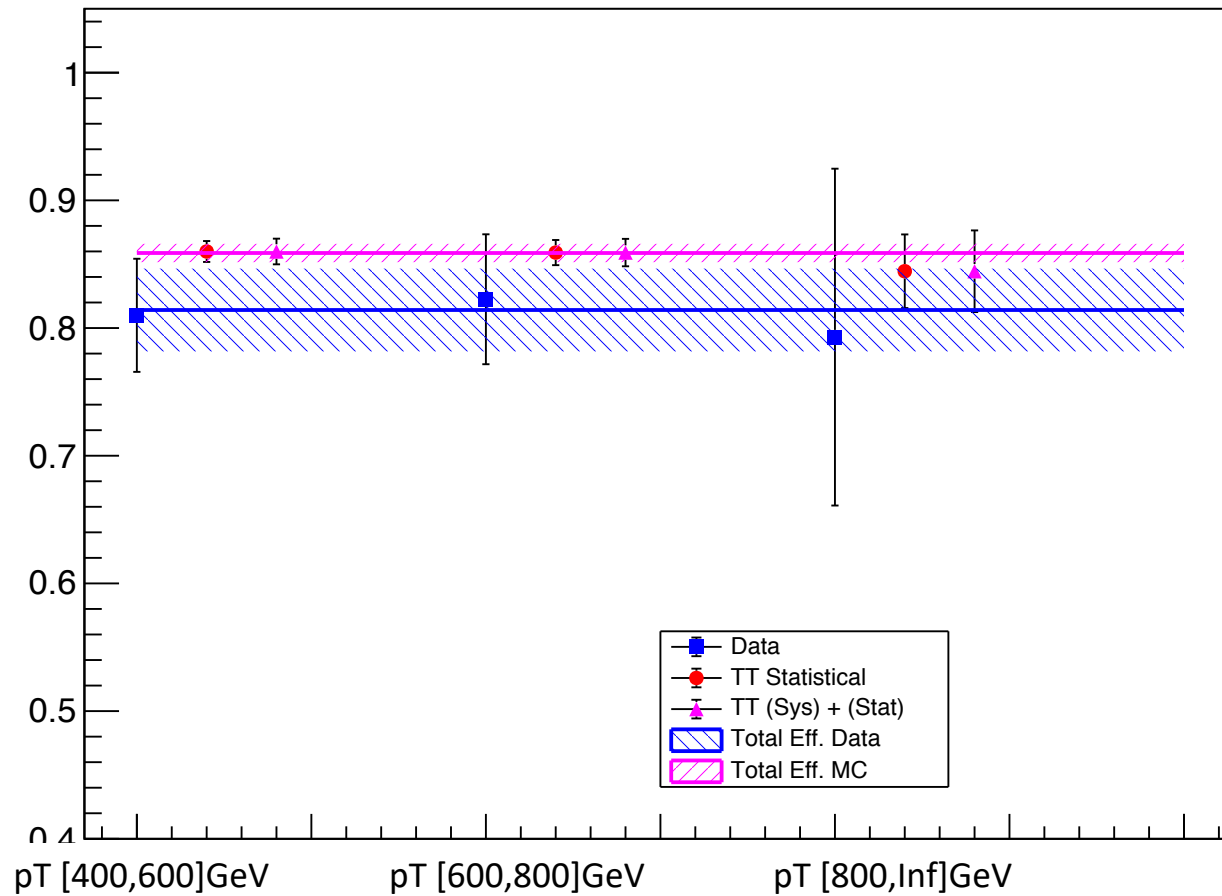
ttbar pT[600-800]: $0.772 \pm (\text{stat}) 0.012 \pm (\text{stat} + \text{systematic}) 0.013$

data pT[800-Inf]: 0.861 ± 0.259

ttbar pT[800-Inf]: $0.758 \pm (\text{stat}) 0.034 \pm (\text{stat} + \text{systematic}) 0.037$



TagAndProbe Efficiency per Pt region(2017)



icity--

a: 0.814 ± 0.033

ar: $0.859 \pm (\text{stat}) 0.006 \pm (\text{stat} + \text{systematic}) 0.007$

--

icity per Pt region

a pT[400-600]: 0.81 ± 0.044

ar pT[400-600]: $0.86 \pm (\text{stat}) 0.008 \pm (\text{stat} + \text{systematic}) 0.01$

--

a pT[600-800]: 0.823 ± 0.051

ar pT[600-800]: $0.859 \pm (\text{stat}) 0.01 \pm (\text{stat} + \text{systematic}) 0.011$

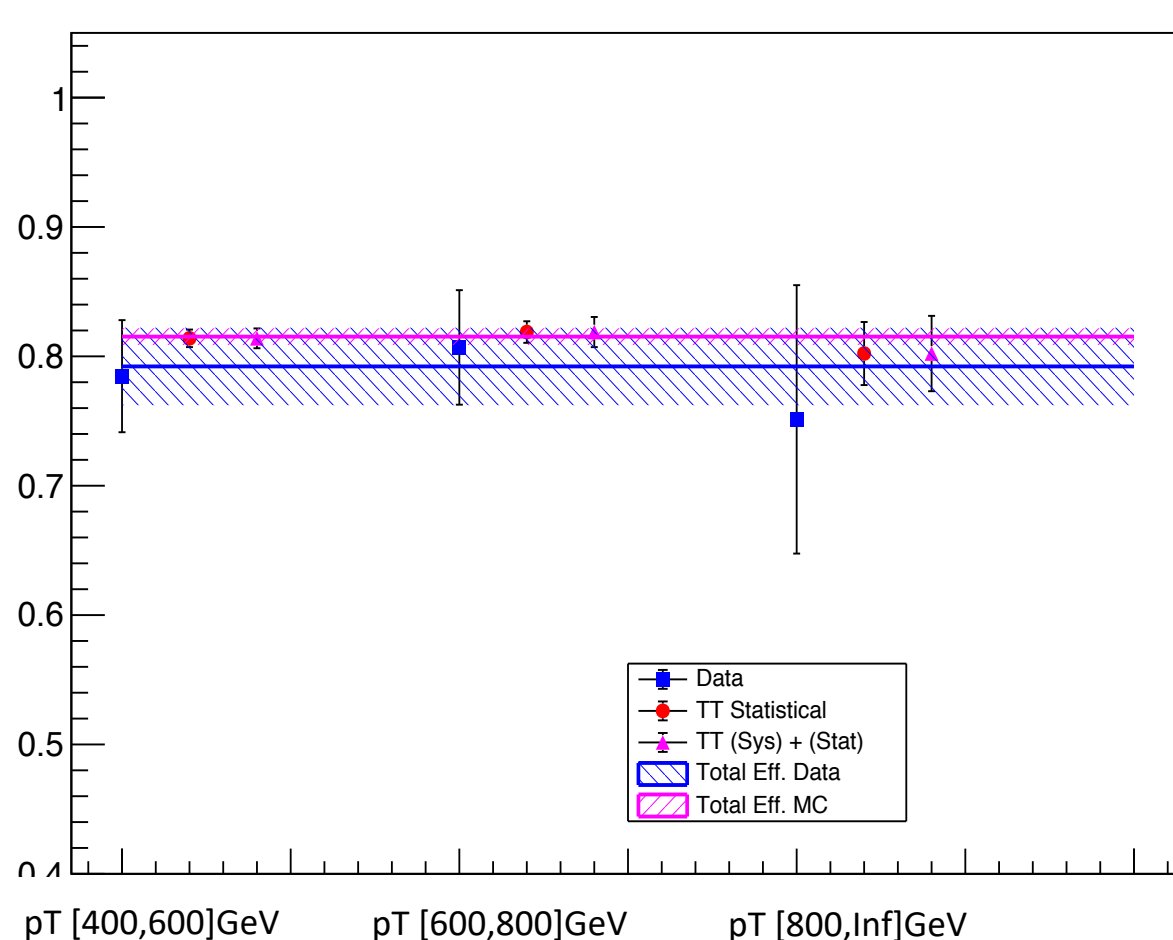
--

a pT[800-Inf]: 0.793 ± 0.132

ar pT[800-Inf]: $0.845 \pm (\text{stat}) 0.029 \pm (\text{stat} + \text{systematic}) 0.032$



TagAndProbe Efficiency per Pt region(2018)



Efficiency--

eff data: 0.792 ± 0.03

eff ttbar: $0.815 \pm (\text{stat}) 0.005 \pm (\text{stat} + \text{systematic}) 0.006$

Efficiency per Pt region

eff data pT[400-600]: 0.785 ± 0.043

eff ttbar pT[400-600]: $0.814 \pm (\text{stat}) 0.007 \pm (\text{stat} + \text{systematic}) 0.008$

eff data pT[600-800]: 0.81 ± 0.044

eff ttbar pT[600-800]: $0.819 \pm (\text{stat}) 0.008 \pm (\text{stat} + \text{systematic}) 0.012$

eff data pT[800-Inf]: 0.753 ± 0.11

eff ttbar pT[800-Inf]: $0.802 \pm (\text{stat}) 0.024 \pm (\text{stat} + \text{systematic}) 0.029$

