Measurement of differential production cross section for high-pT top quarks in the all hadronic channel

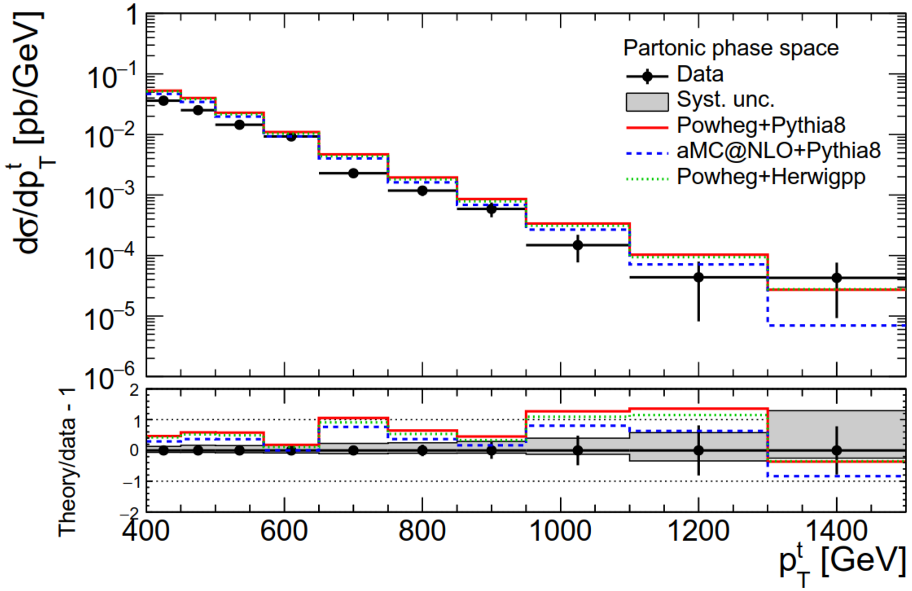
Top ttX meeting

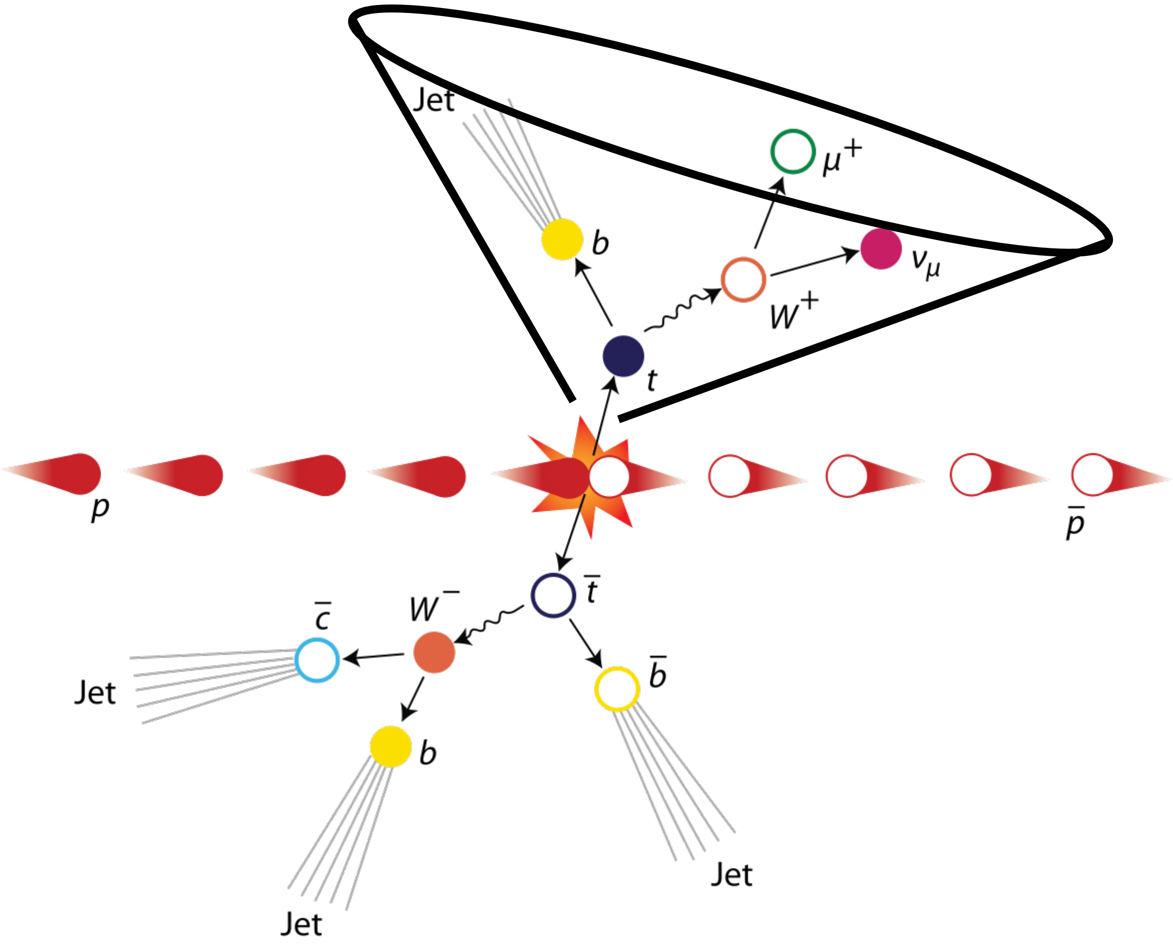
G. Bakas, K. Kousouris, I. Papakrivopoulos, G. Tsipolitis

Motivation

Top, anti-top production with fully hadronic final state.

Trying to identify two big jets that contain the products of the top/anti-top decay.





Overview

- Variables of interest:
  - ttbar mass, pt, rapidity
  - Leading and Subleading jetPt and |jetY|
  - Mtt samples (700-1000, 1000-Inf)

Region	Requirements
Signal Region (SR)	Baseline + topTagger + $m_{SD}^{jet1,2} \in (120,220) GeV + 2btags$
Control Region (CR)	Baseline + topTagger + $m_{SD}^{jet1,2} \in (120,220) GeV + 0btags$
Extended SR (SR_A) (QCD fit region)	Baseline + topTagger + $m_{SD}^{jet1,2} \in (50,300) GeV + 2btags$

- Top Angular distributions
  - $|\cos(\theta^*)|$ of the leading jet
  - $\chi = e^{|2y^*|} = e^{|y_1 - y_2|}$, where $y^* = \frac{1}{2}(y_1 - y_2)$
  - The distributions associated with the final states produced via QCD interactions are relatively flat in comparison with the distributions of the BSM models or new particles, which typically peak at low values of χ
  - Mtt sample (1000-Inf)

- Baseline Parton cuts:
  - Jet Matching
  - partonPt[0],[1] > 400
  - |partonEta[0],[1]| < 2.4
  - mTTbarParton > 1000

- Btagging selection:
  - bTagging (medium WP **deepCSV**) (2016: 0.6321, 2017: 0.4941, 2018: 0.4184)

- Baseline Reconstructed level cuts:
  - nJets > 1
  - nLeptons = 0
  - Dijet mass (mJJ) > 1000
  - Leading and Subleading jet p_T > 400
  - Leading and Subleading absolute jet eta $|\eta|<2.4$
  - Trigger

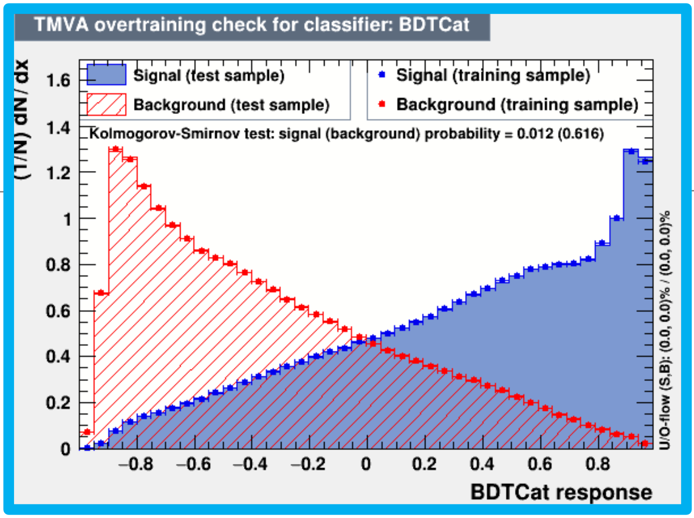
- Top Tagger WP:
  - **New top Tagger:** (2016: 0.2, 2017:0.0, 2018: 0.1)

Overview: Discriminator, Efficiency and Acceptance

The discriminator is a BDT trained individually for 2016, 2017 and 2018

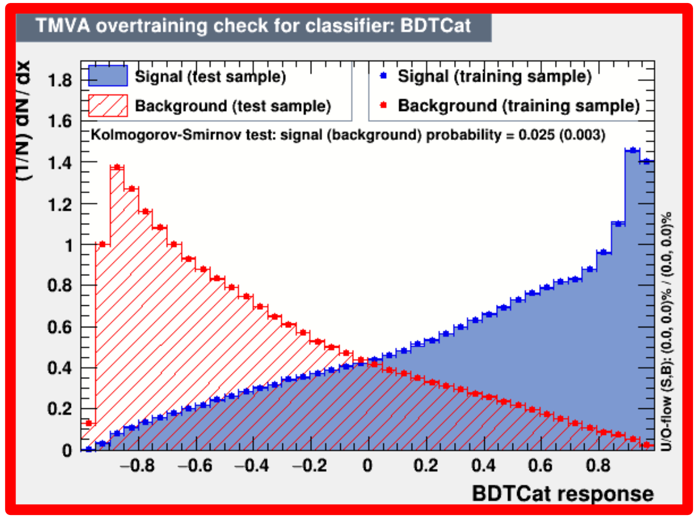
Category training: split the sample in categories based on Pt

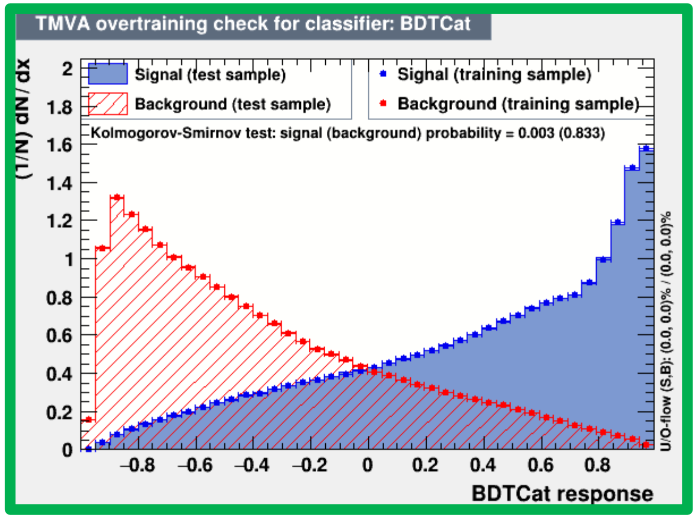
- Bins:
  - [400, 600] GeV
  - [600, 800] GeV
  - [800, 1200] GeV
  - [1200, inf) GeV
- BDT, used variables:
  - Leading and Sub-leading subjet mass
  - N-Subjetiness variables (tau1, tau2, tau3)
  - fraction of the jetPt over the total pt sum of the event.
  - Energy correlation functions (ecfB1N2,ecfB1N3, ecfB2N2, ecfB2N3)
- BDT Output consistency for the 3 years
- Calculation of Efficiency and acceptance for each year
  - We choose the WP's for each year so that the leading jet p_T efficiency is similar for all years



2016

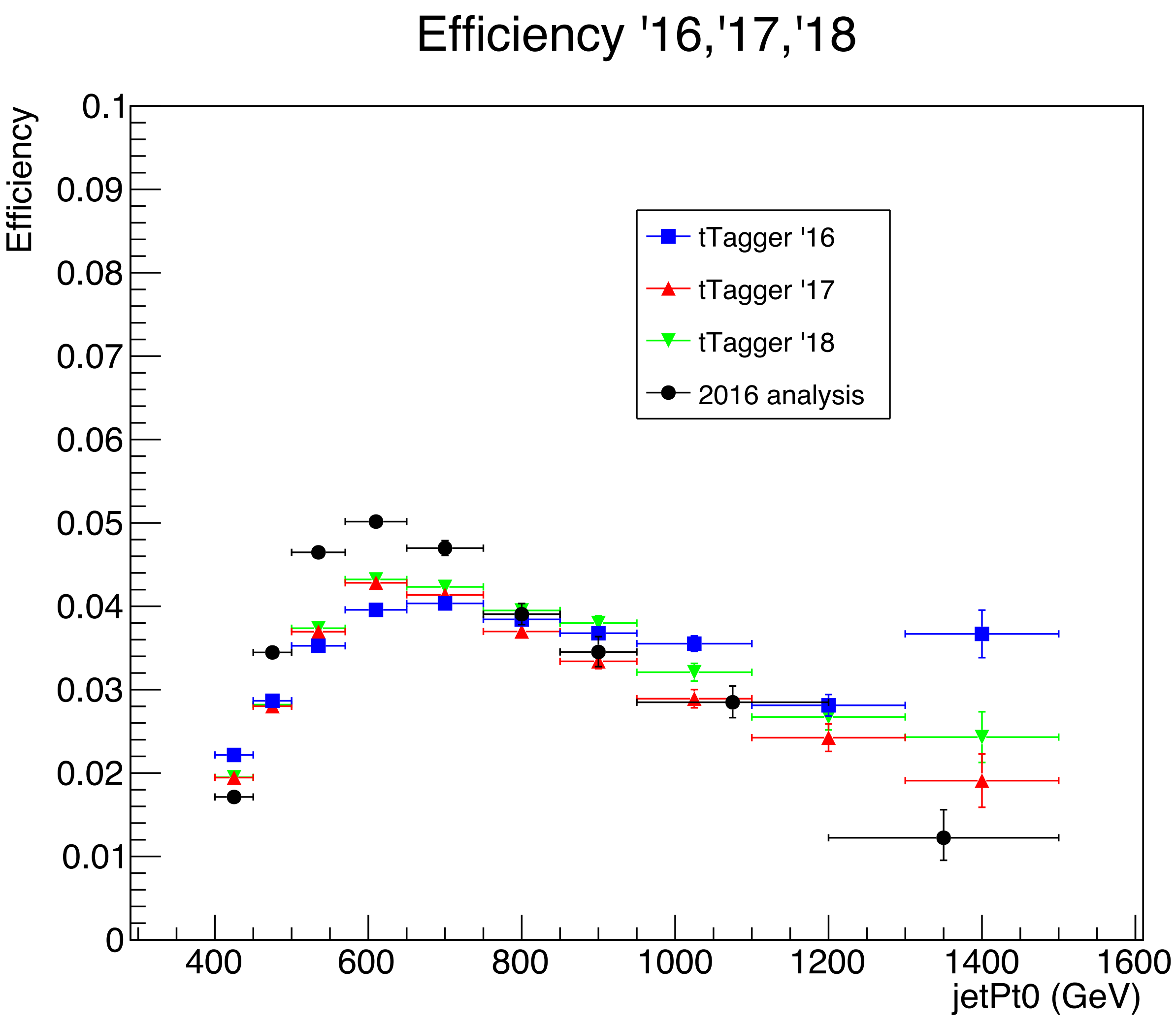
2017

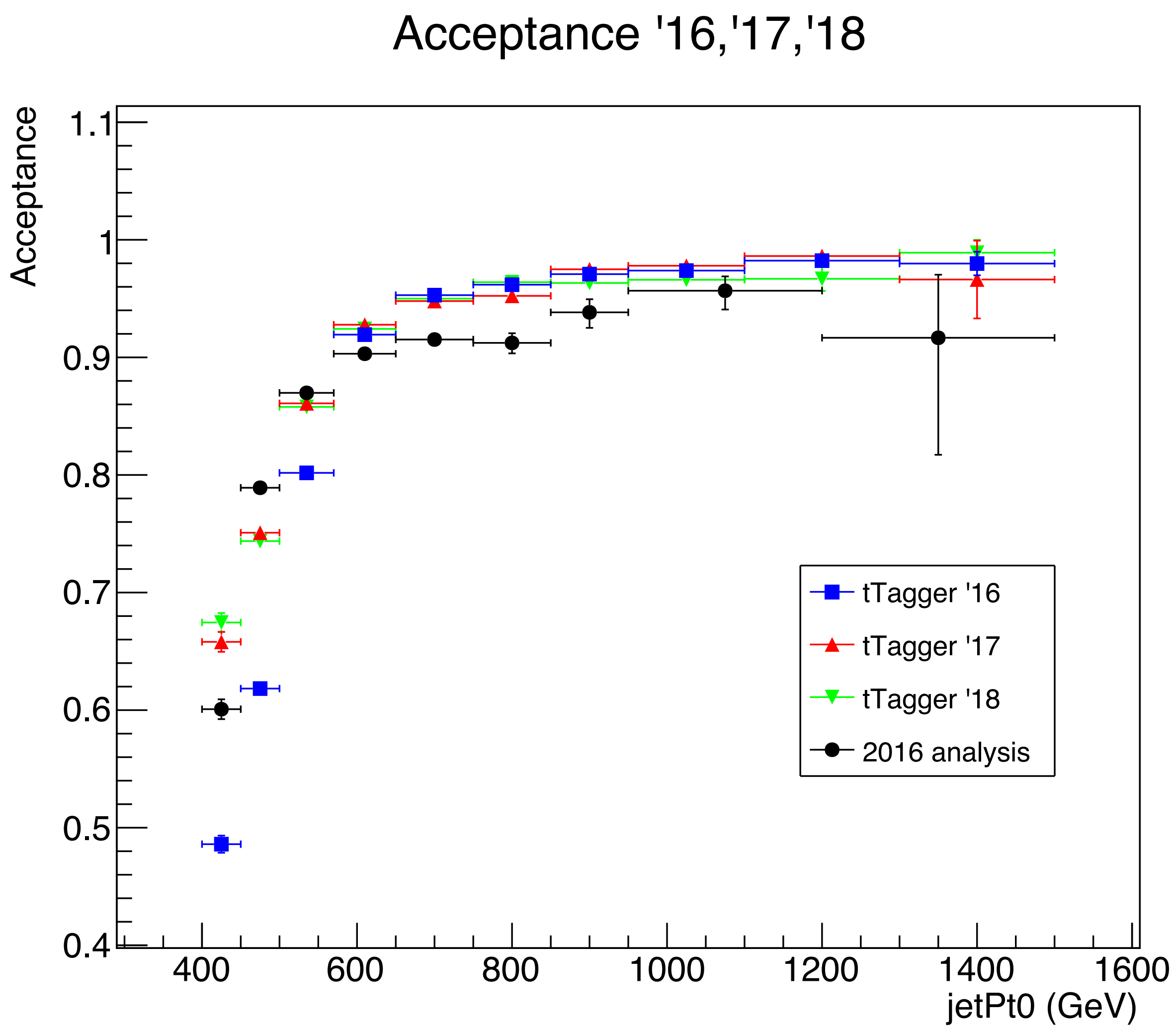




2018

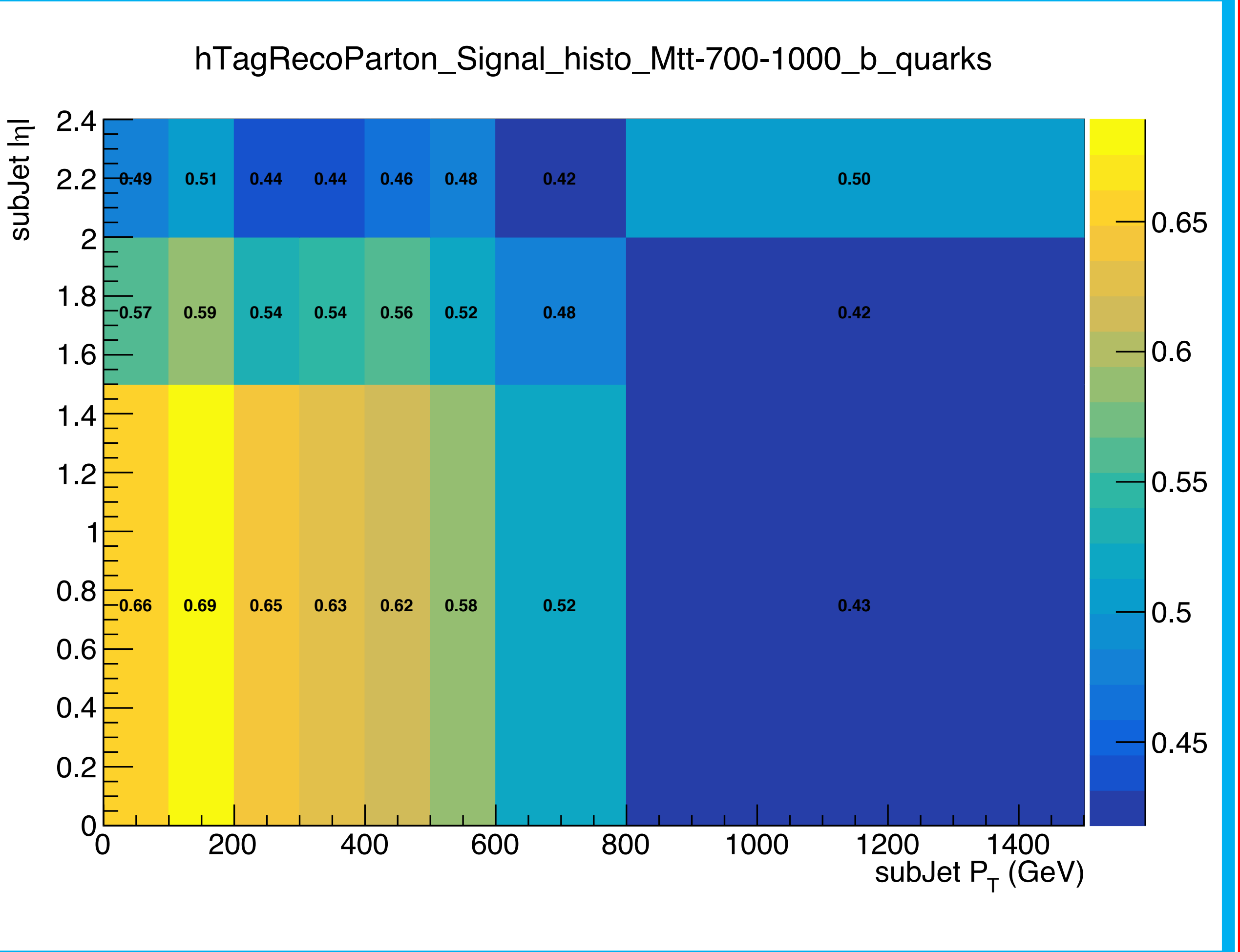
Efficiency and Acceptance for 2016, 2017 and 2018 and previous 2016 analysis



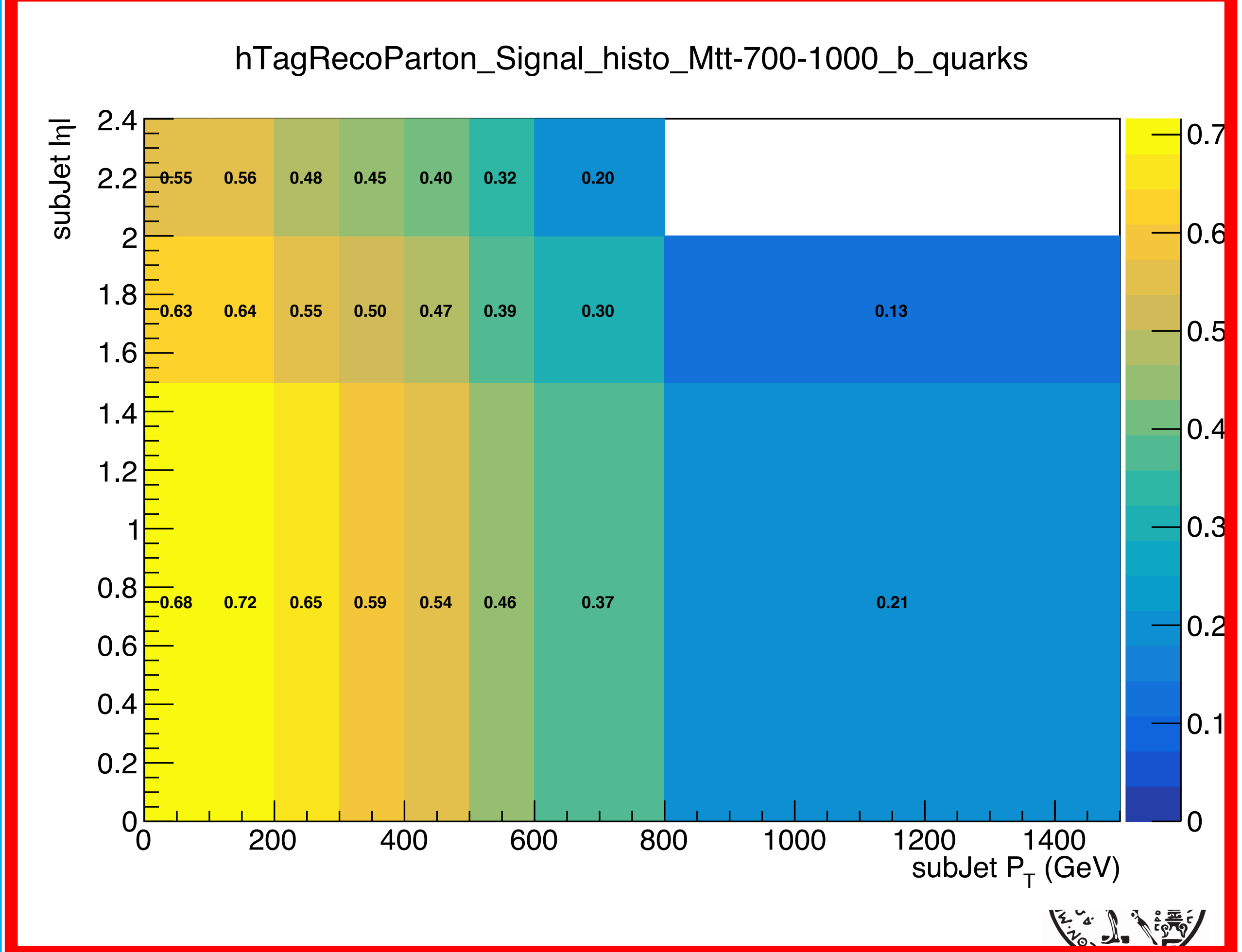


Btagging efficiency in eta, pT_{subJet} phase space

2016

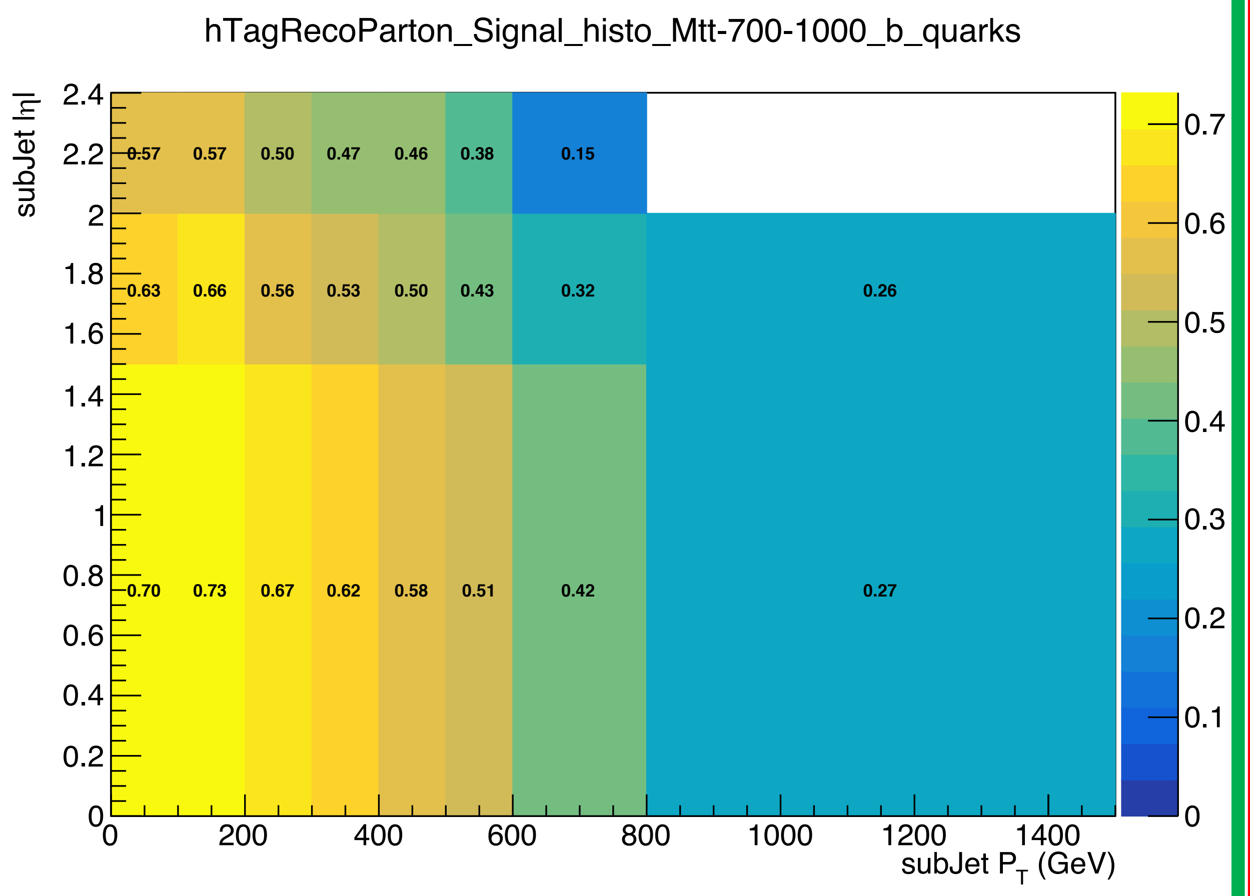


2017

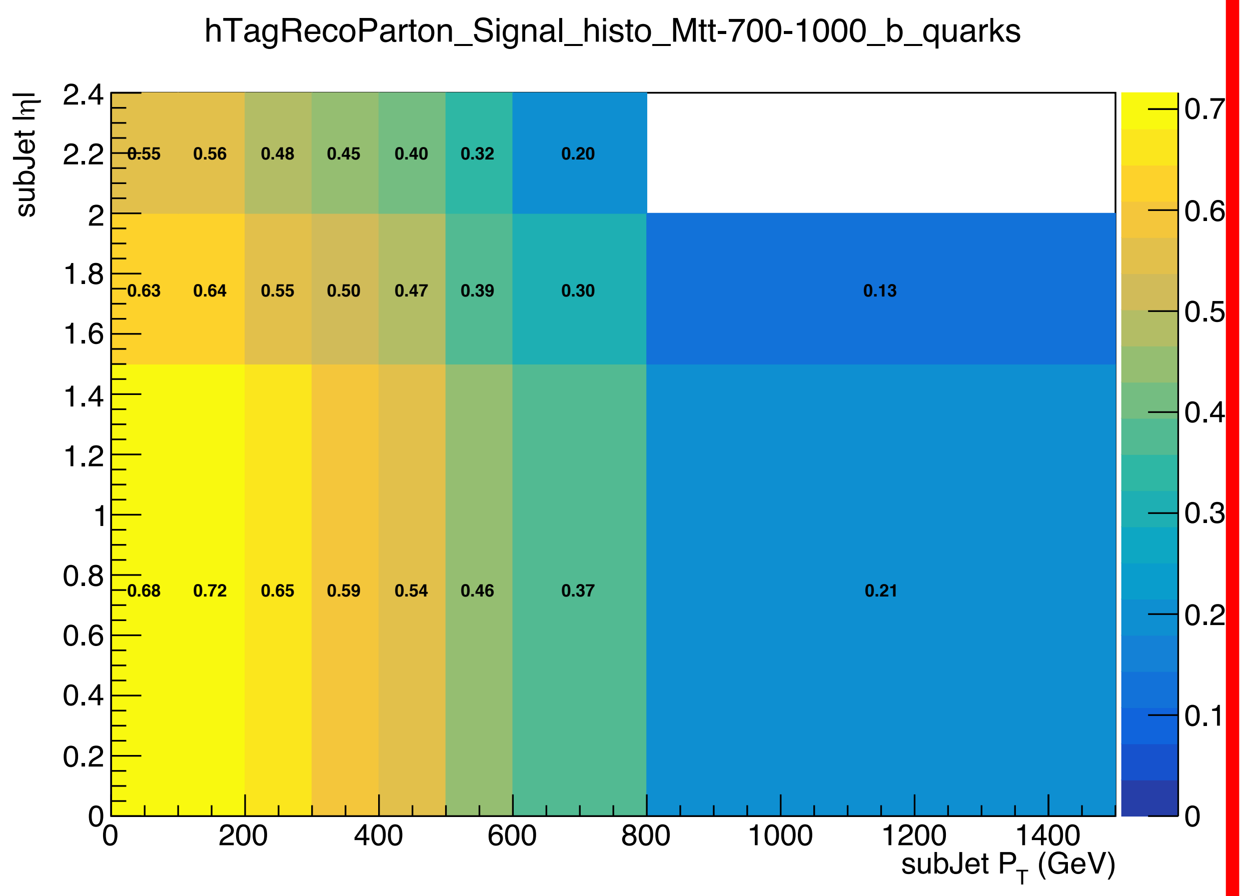


Btagging efficiency in eta, pT_{subJet} phase space

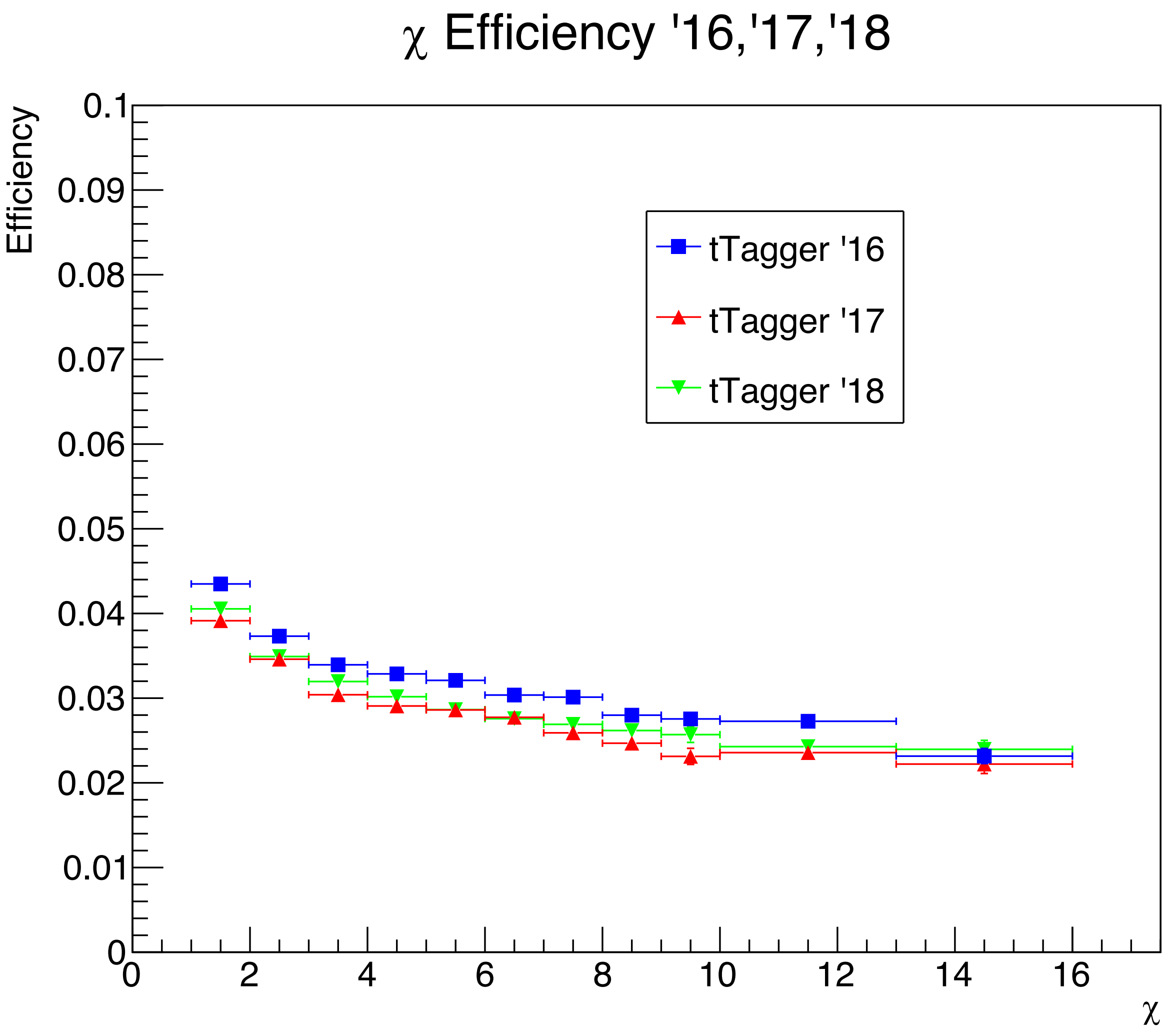
2018

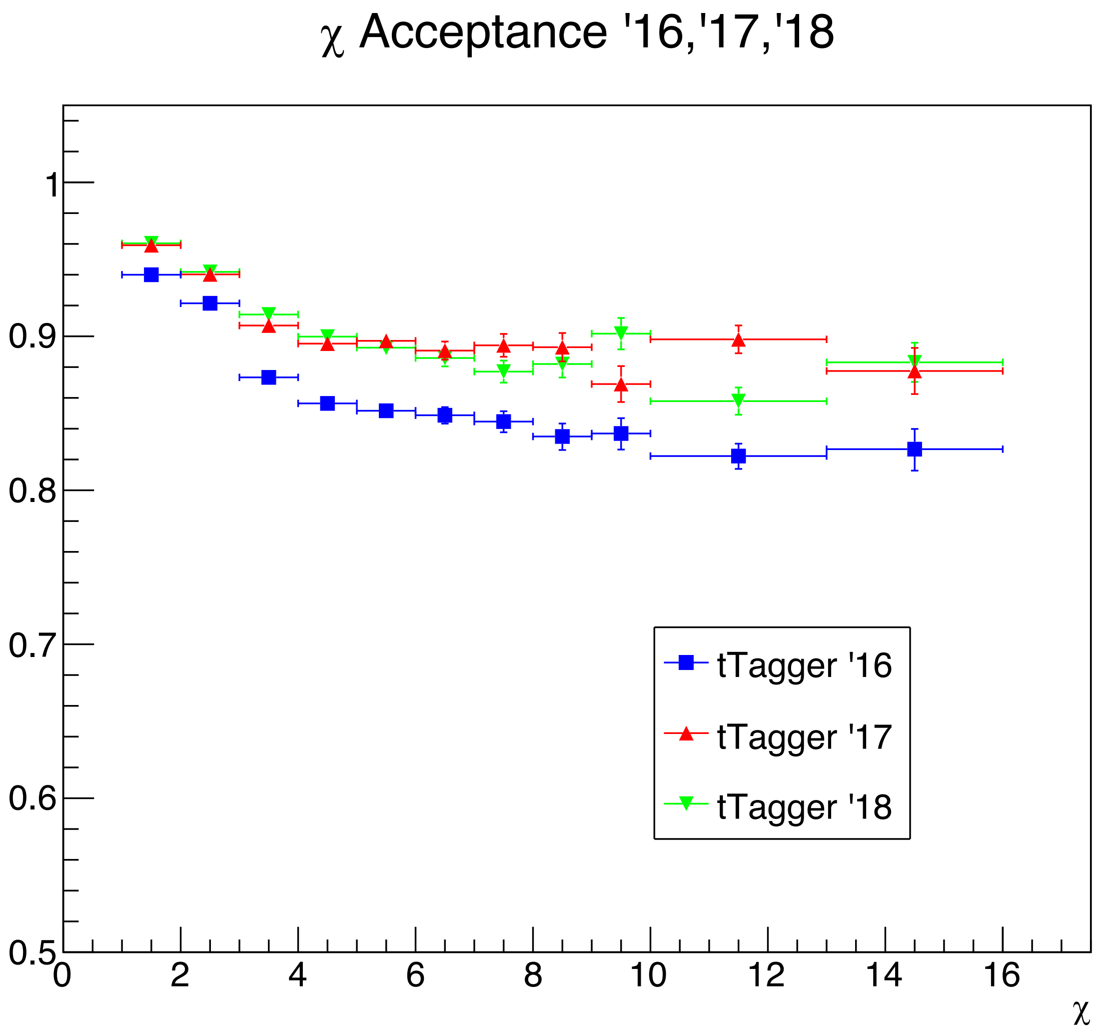


2017

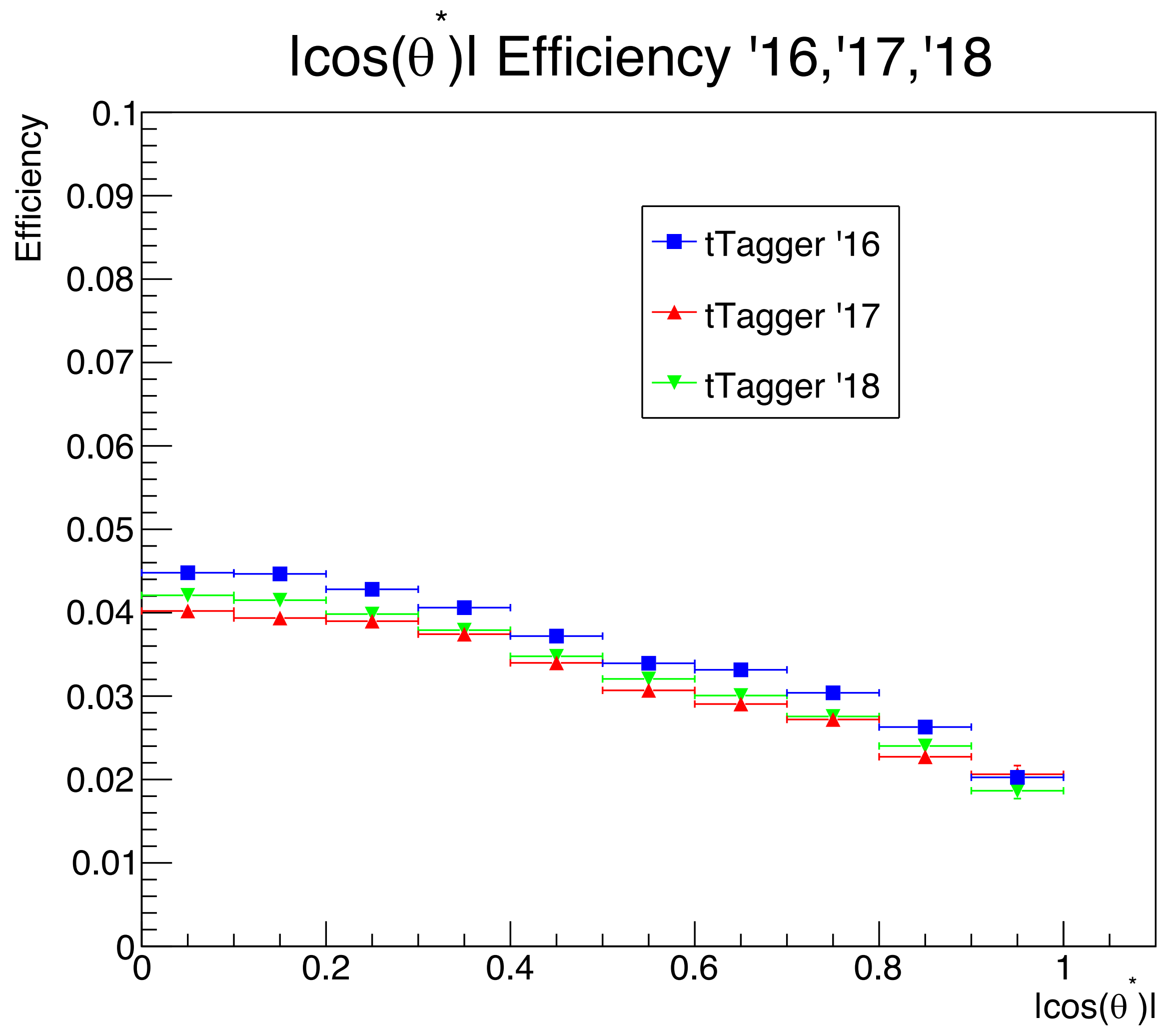


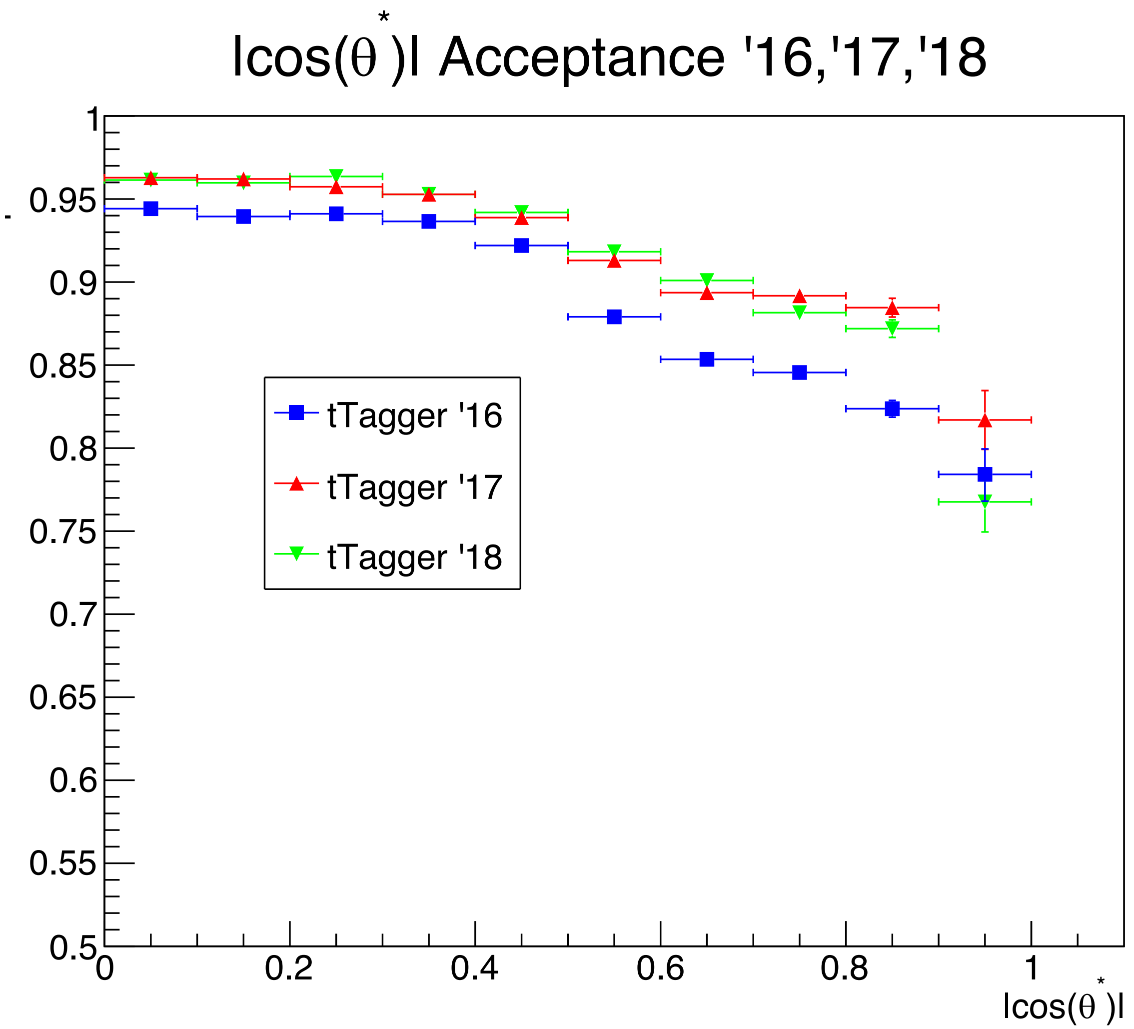
Efficiency, Acceptance for χ





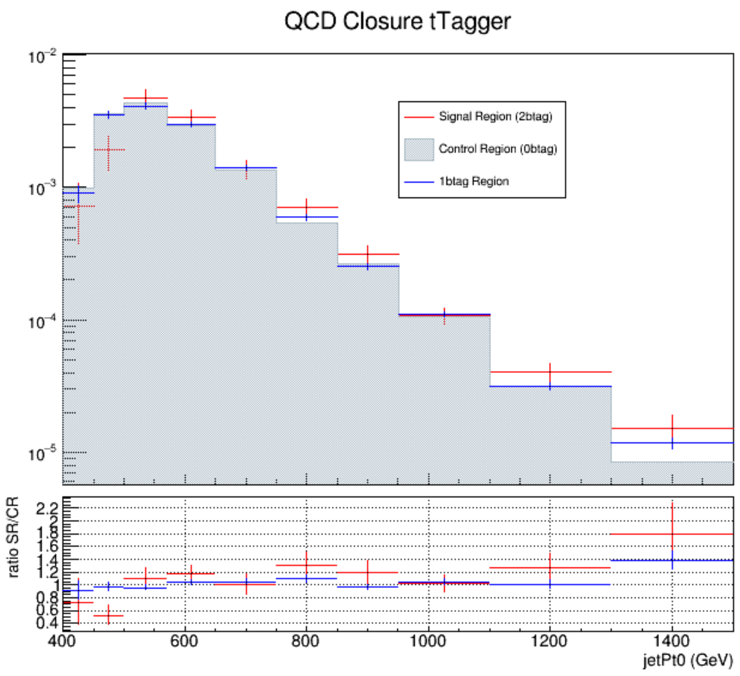
Efficiency, Acceptance for |cos(θ*)|



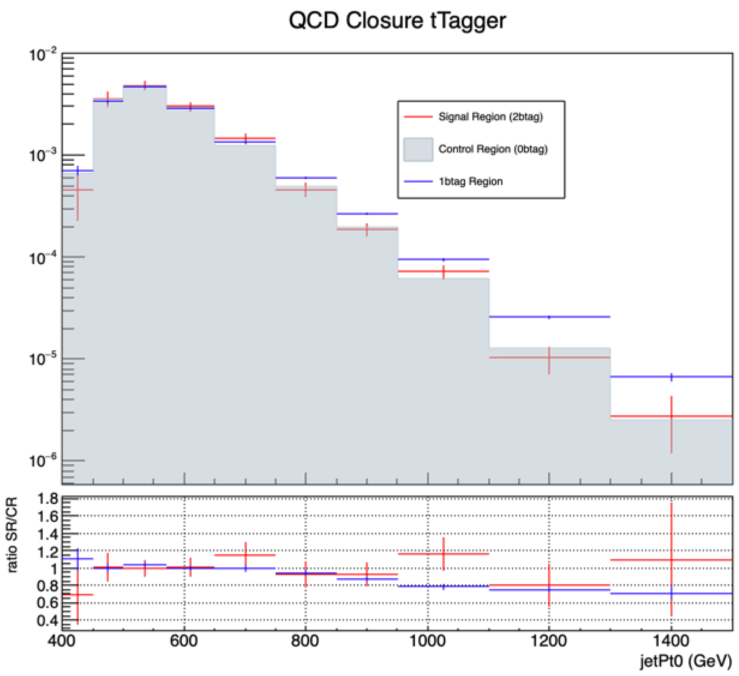


QCD Closure Tests ’16, ‘17, ’18 jetPt0

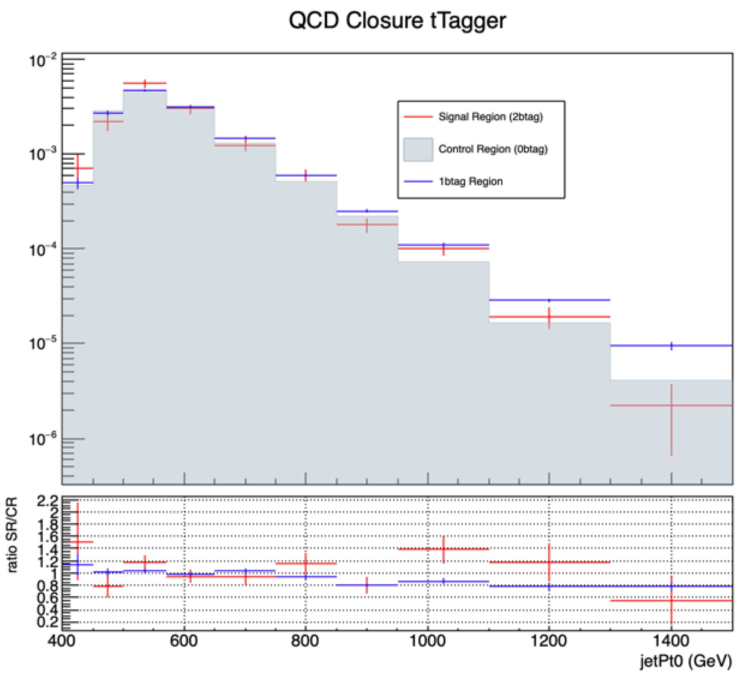
2016



2017

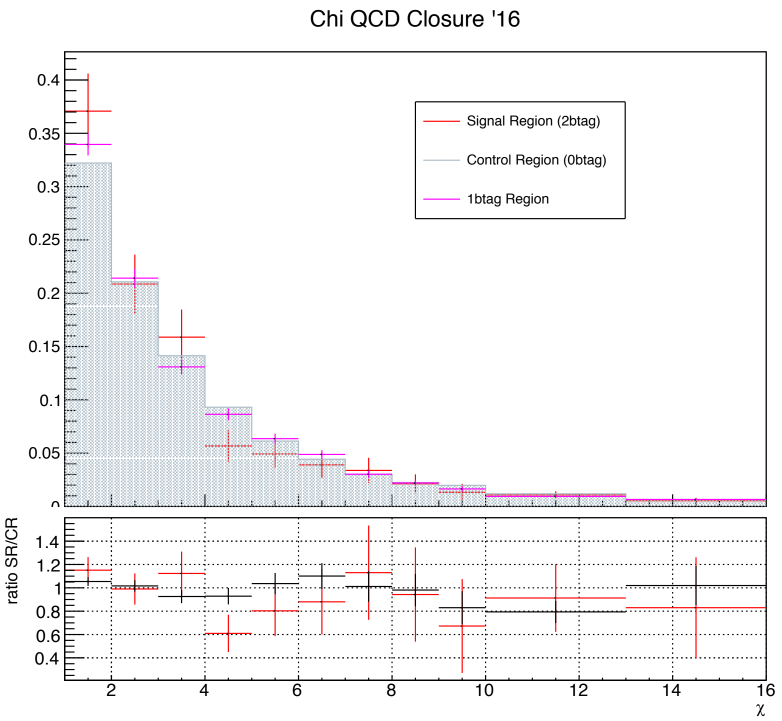


2018

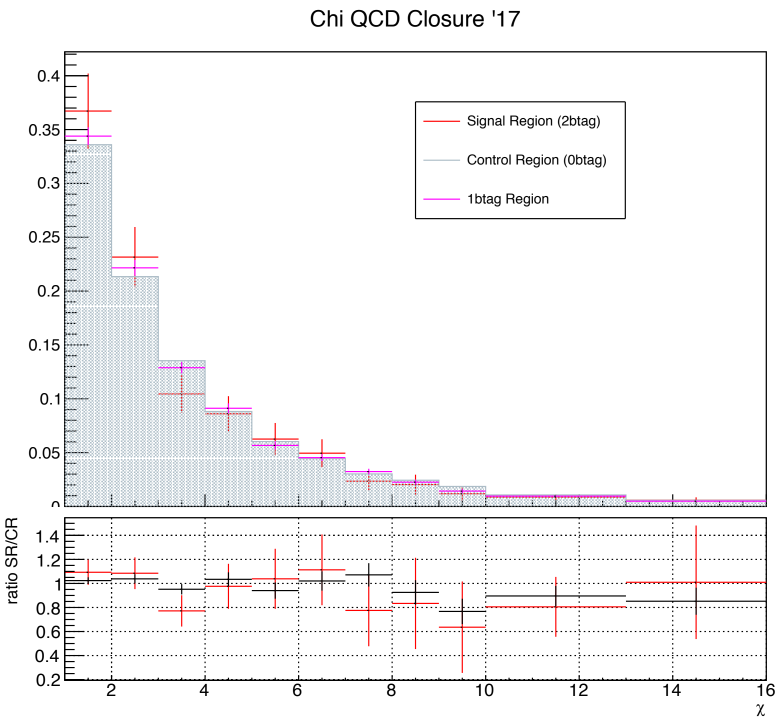


QCD Closure Tests ’16, ‘17, ’18 χ

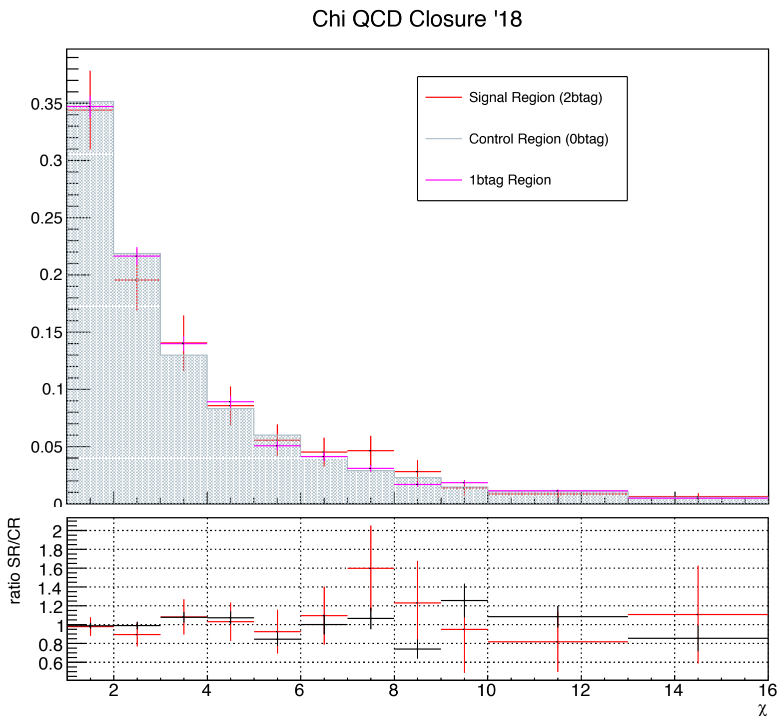
2016



2017

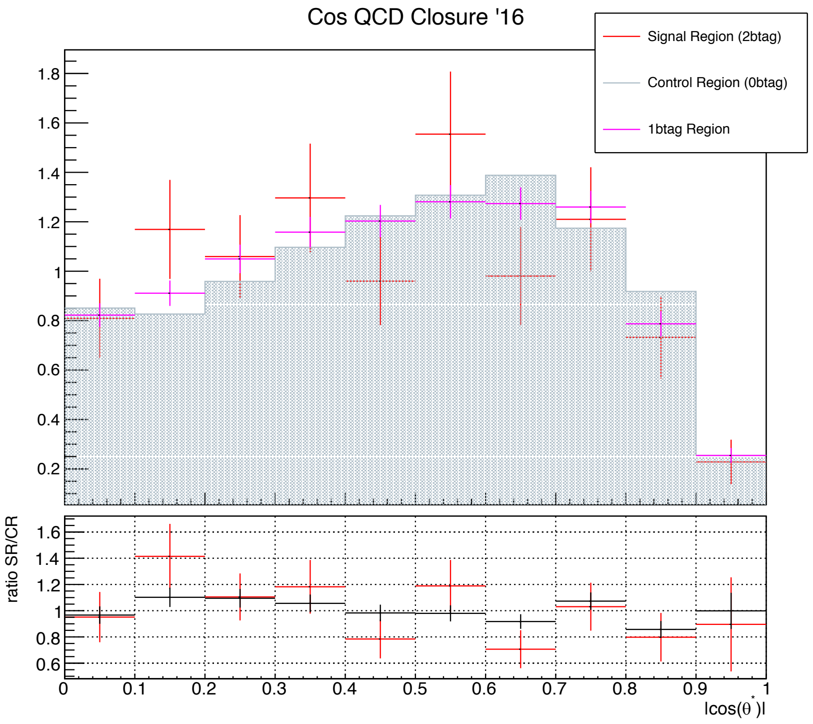


2018

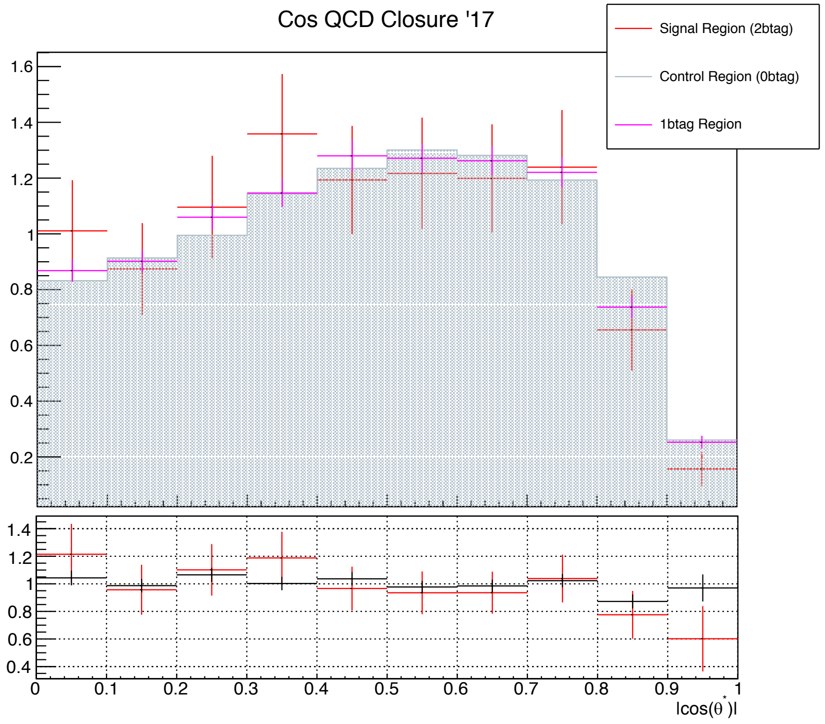


QCD Closure Tests '16, '17, '18 |cos(θ*)|

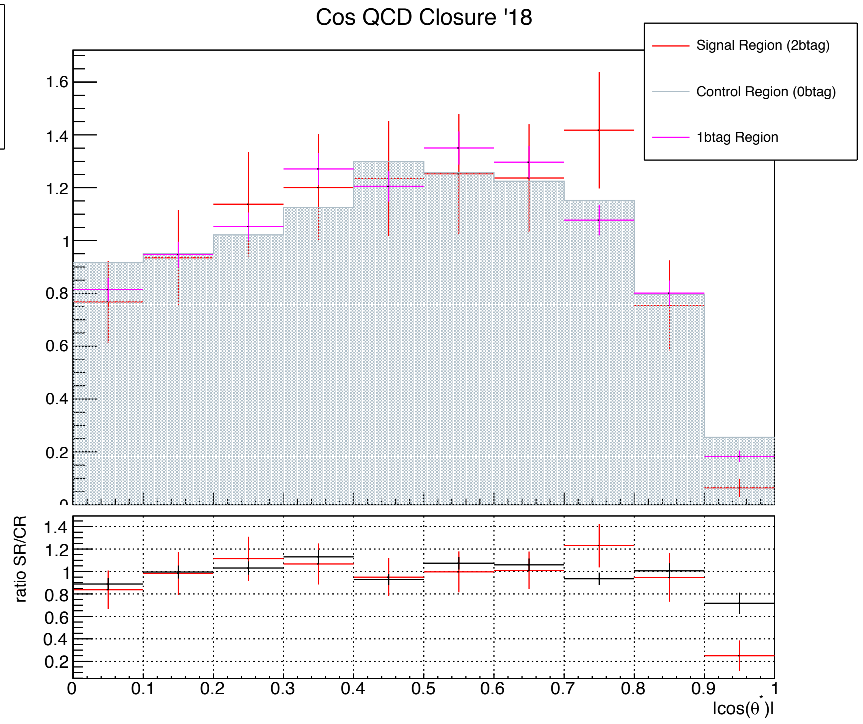
2016



2017



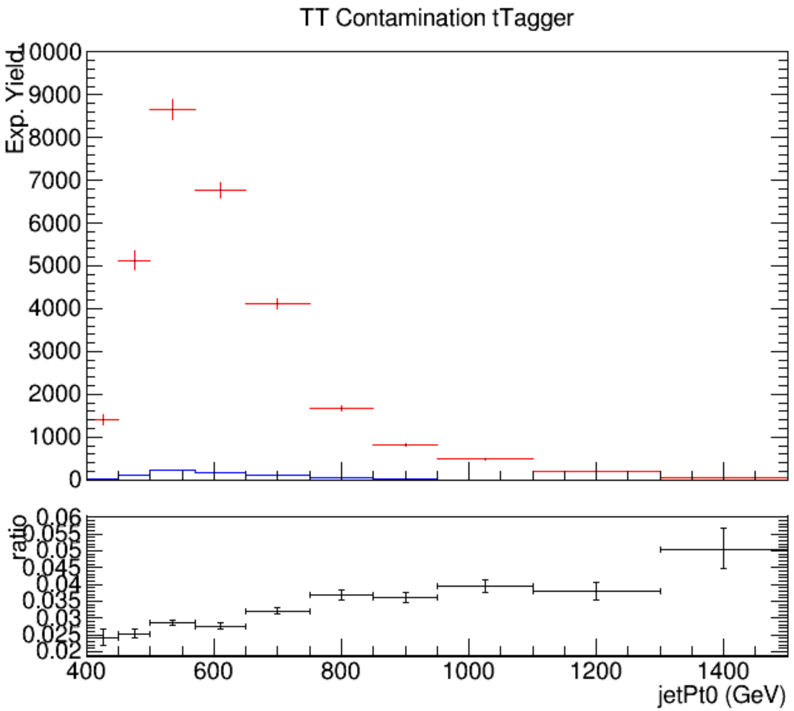
2018



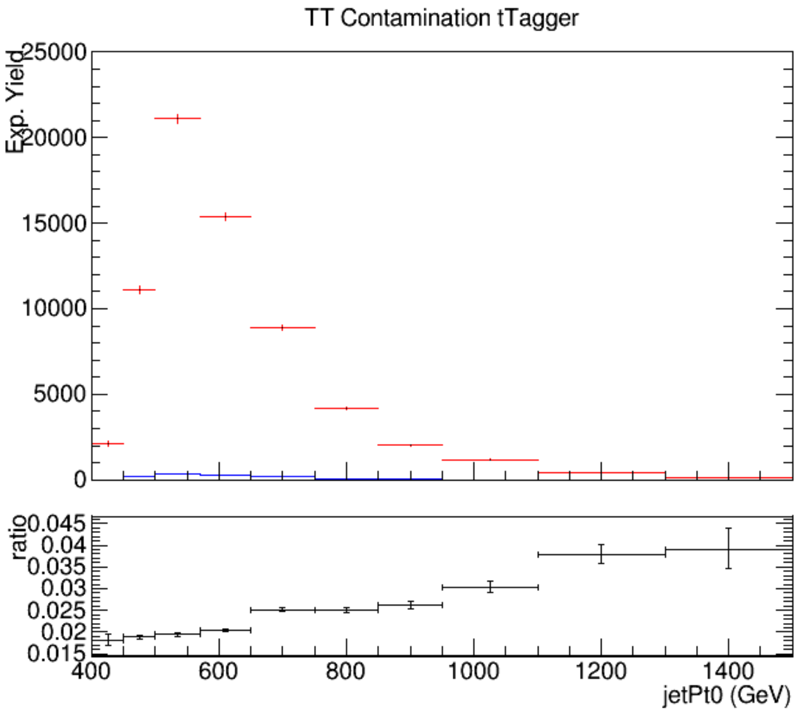
CR Contamination ’16,’17,’18 jetPt0

Control Region QCD sample
Control Region TT sample

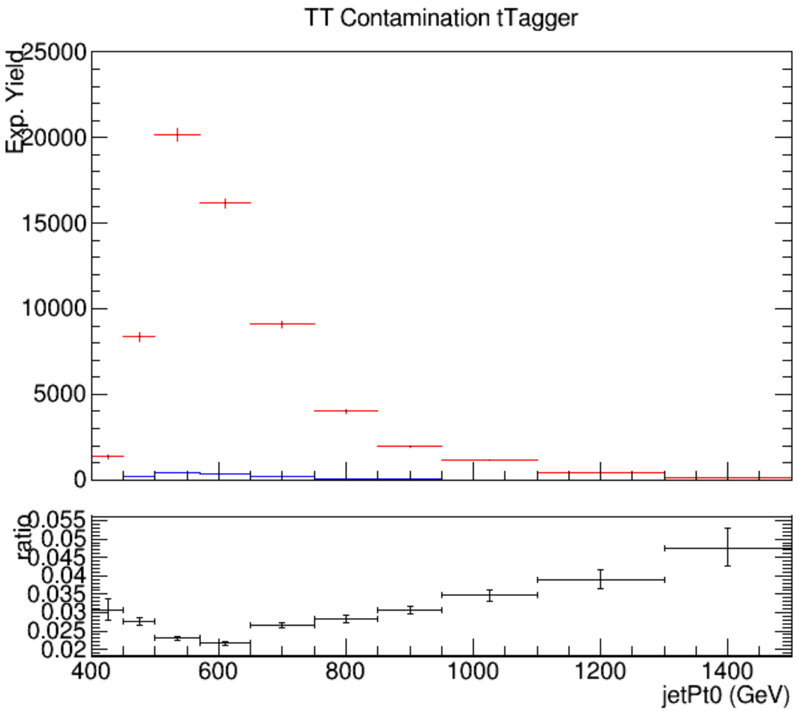
2016



2017



2018

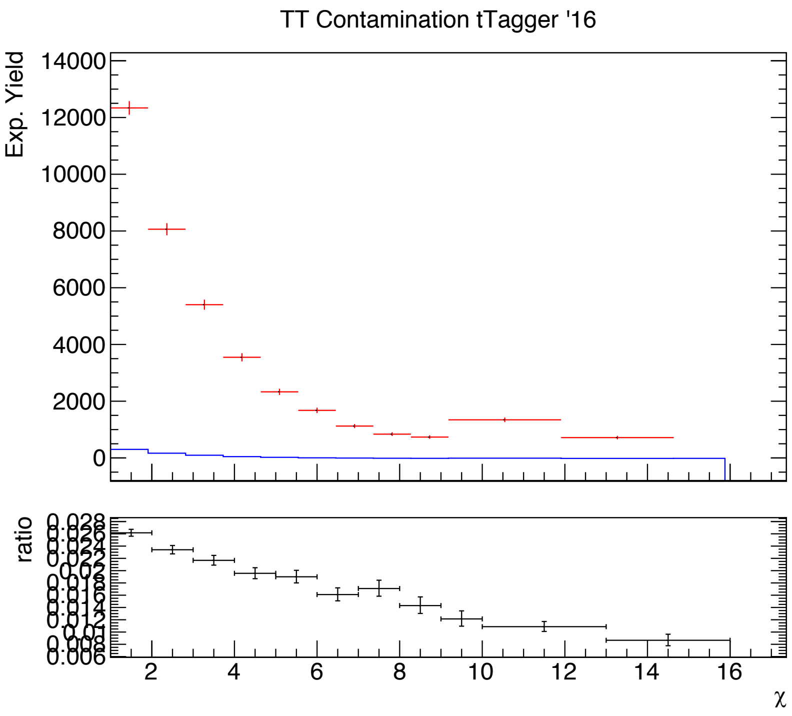


CR Contamination ’16,’17,’18 χ

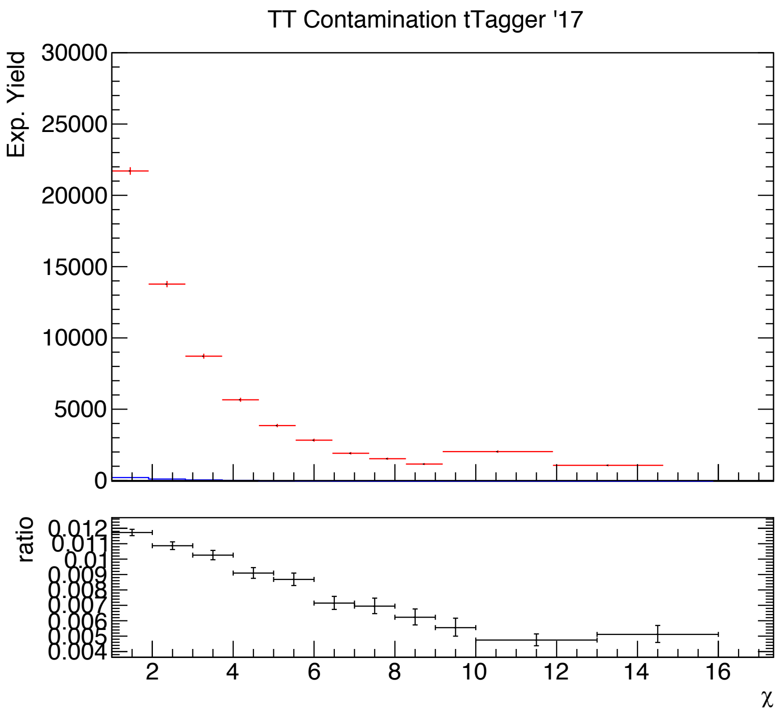
Control Region QCD sample

Control Region TT sample

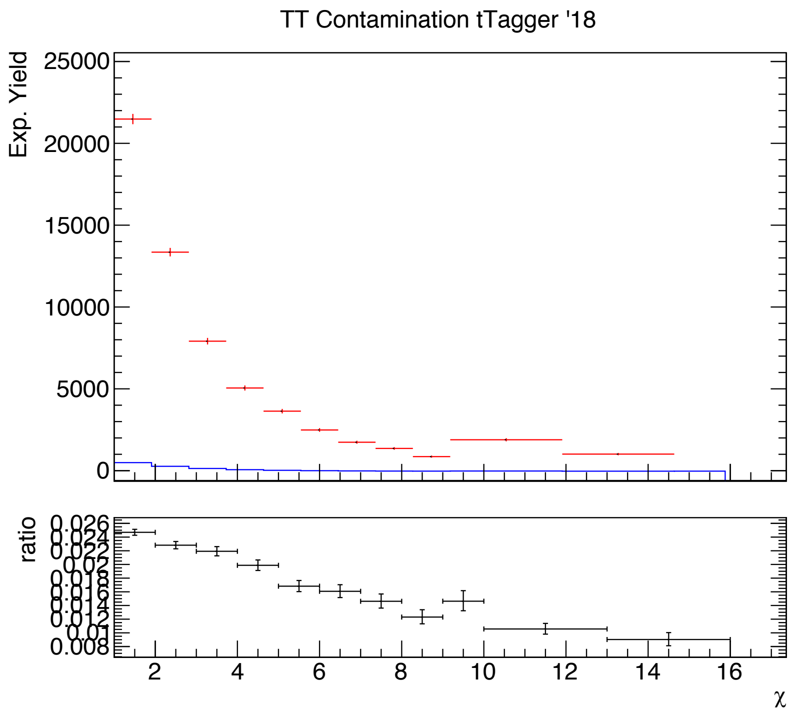
2016



2017



2018

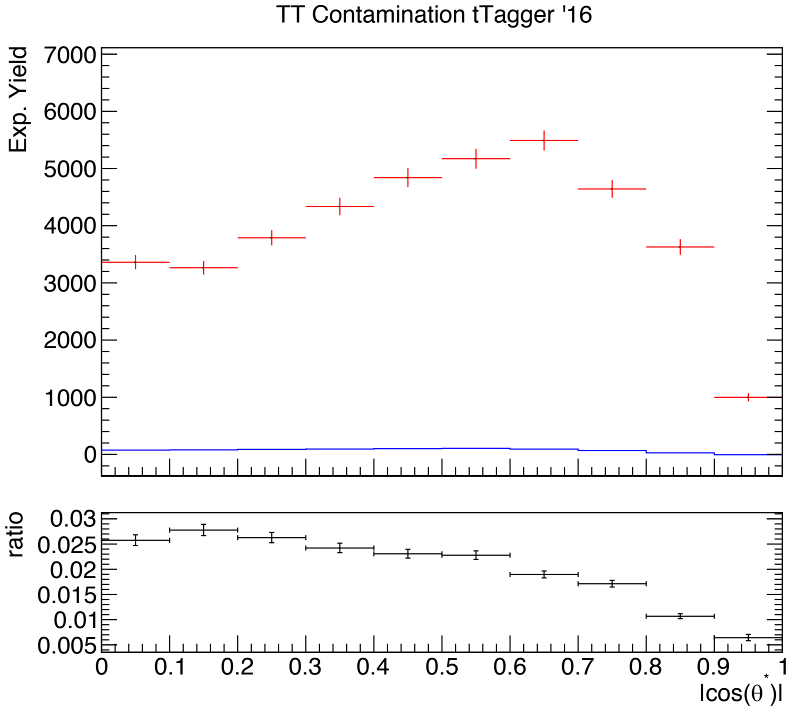


CR Contamination ’16,’17,’18 |cos(θ*)|

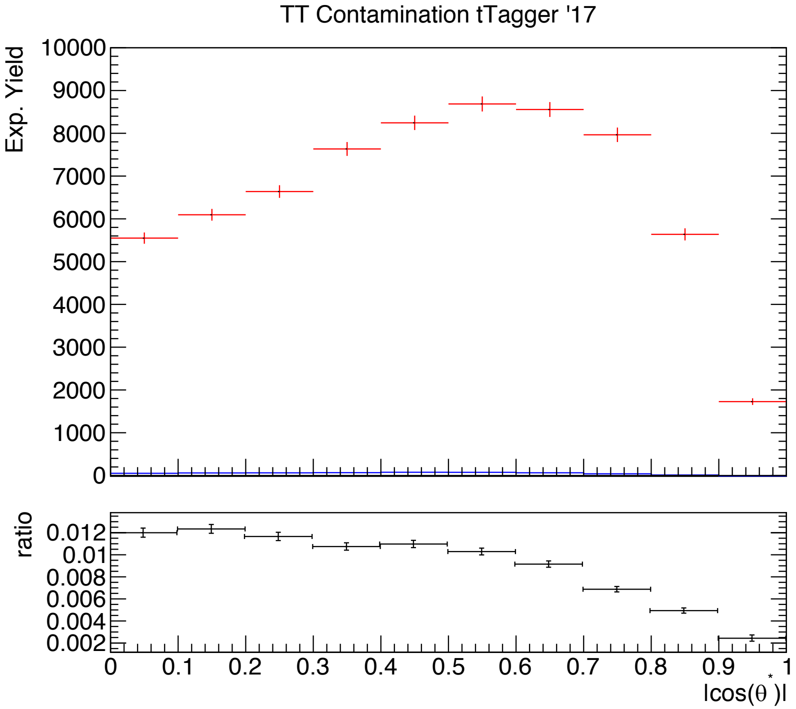
Control Region QCD sample

Control Region TT sample

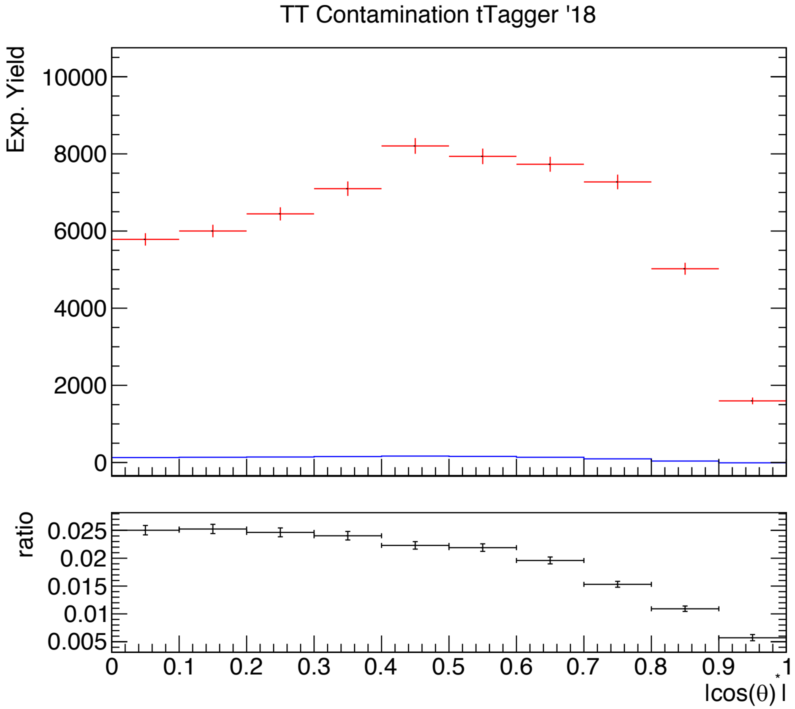
2016



2017



2018



Signal Extraction

Measured dist from data

Fitted number of QCD events in SR_A

QCD shape taken from Data (CR)

$$S(x_{reco}) = D(x_{reco}) - C_{bkg}^{yield} N_{QCD}^{fit} C_{QCD}^{shape}(x_{reco}) Q(x_{reco}) - B(x_{reco})$$

Fiducial Yield

Transfer factor from SR_A to SR

QCD shape correction factor

Subdominant bkg shape and contribution (MC)

- Where x_{reco} is the respected variables of interest (ttbar mass,pt, rapidity, leading and subleading jetPt and |jetY|)

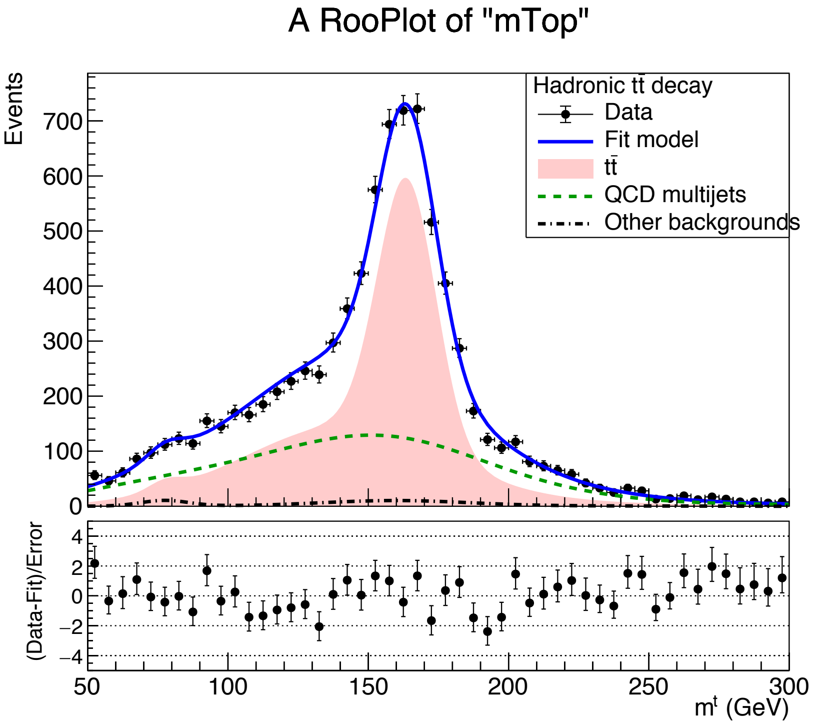
- We deploy a simultaneous fit in 3 regions (0,1,2) btag because we do not have a pure Control Region.
  - Our data CR is contaminated

$$D(m^t)^{(i)} = N_{tt}^{(i)} T^{(i)}(m^t, k_{MassScale}, k_{MassResolution}) + N_{bkg}^{(i)} B(m^t)(1 + k_1 x) + N_{sub}^{(i)} O^{(i)}(m^t)$$

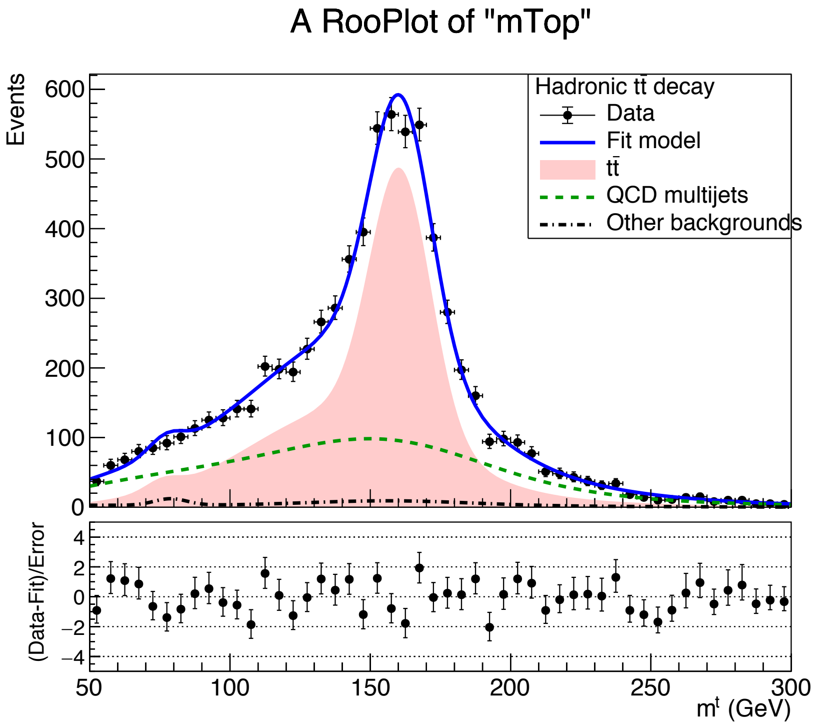
- We assume that $N_{tt}^{(0)} = (1 - \mathrm{e_b})^2 \mathrm{N_{tt}}, \mathrm{N_{tt}^{(2)}} = \mathrm{e_b^2 N_{tt}}$ and $\mathrm{N_{tt}^{(1)}} = 2(1 - e_b) e_b N_{tt}$ where e_b is the b tagging efficiency and N_{tt} is the total ttbar yield.

Simultaneous Fit in 3 regions for 2016, 2017, 2018 when eb is free (SR)

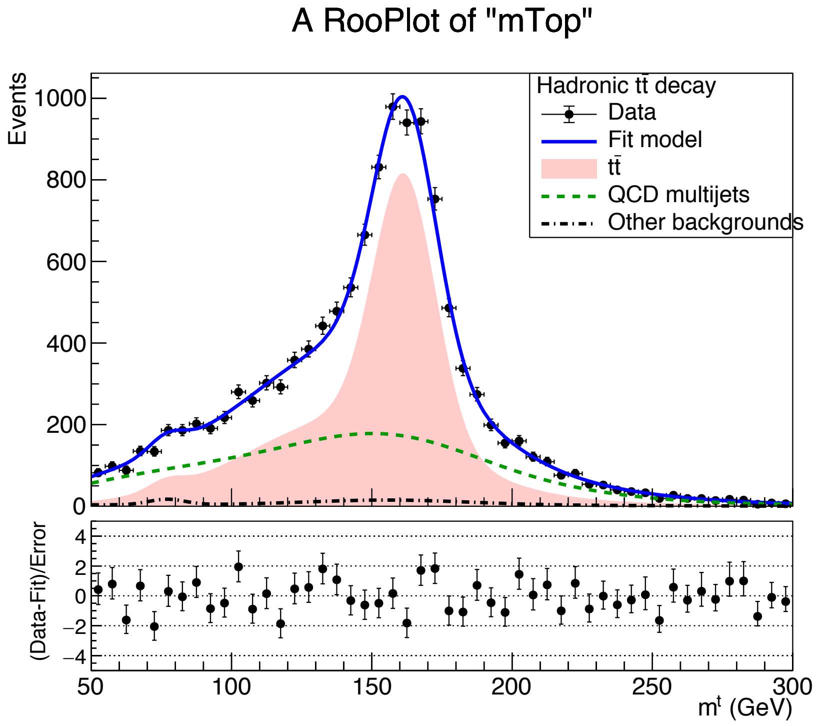
Signal Region (2btag) (2016)



Signal Region (2btag) (2017)



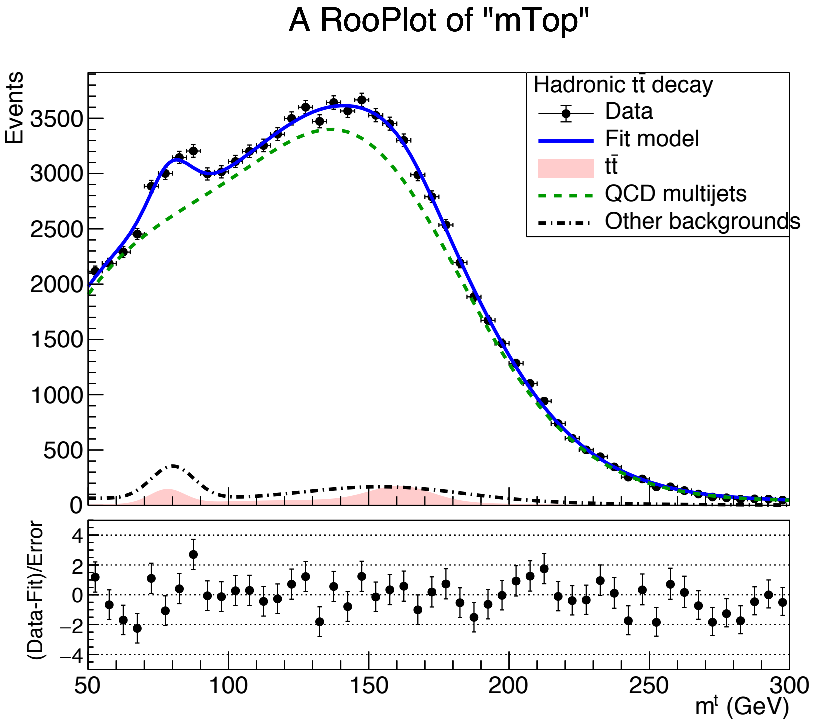
Signal Region (2btag) (2018)



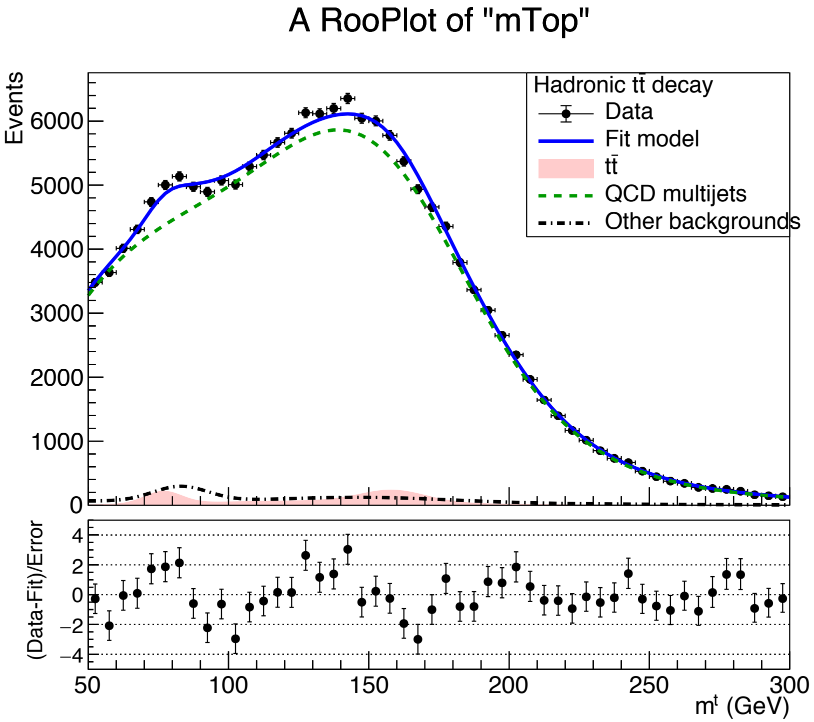
Result of the simultaneous fit on data in SR. The red line shows the ttbar contribution, the green line shows the QCD, and the black line shows the subdominant backgrounds

Simultaneous Fit in 3 regions for 2016, 2017, 2018 when eb is free (CR)

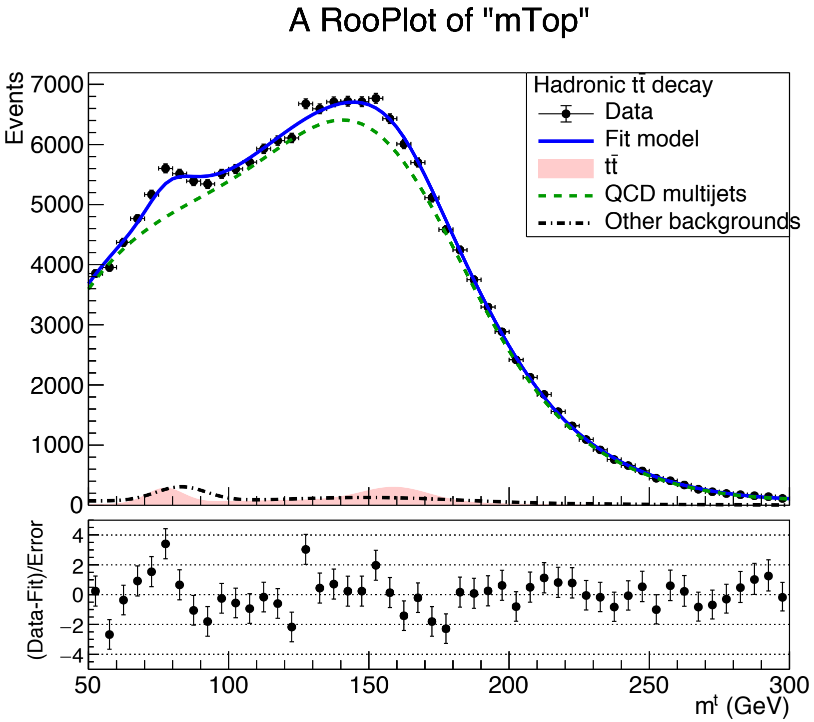
Control Region (0btag) (2016)



Control Region (0btag) (2017)



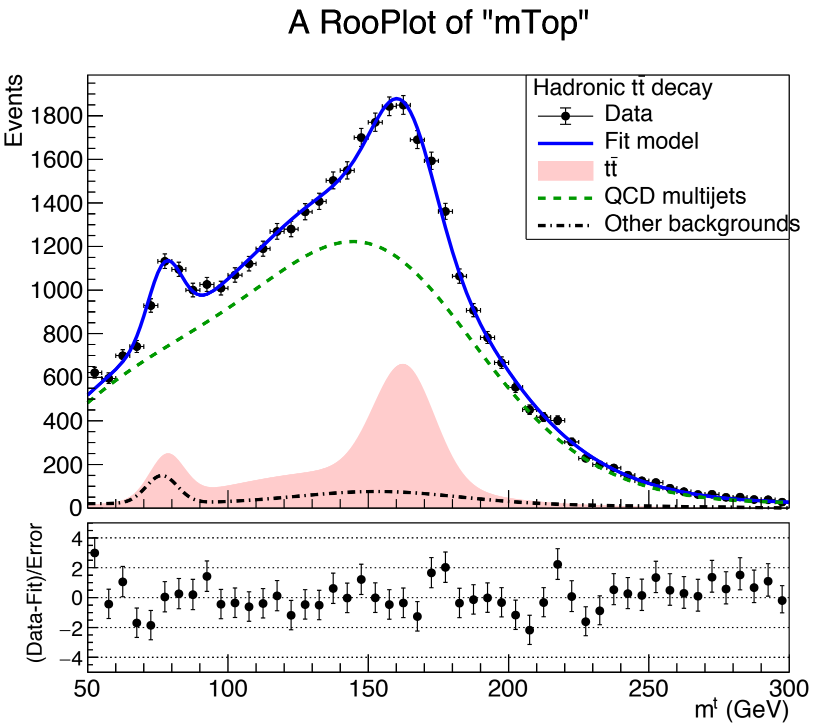
Control Region (0btag) (2018)



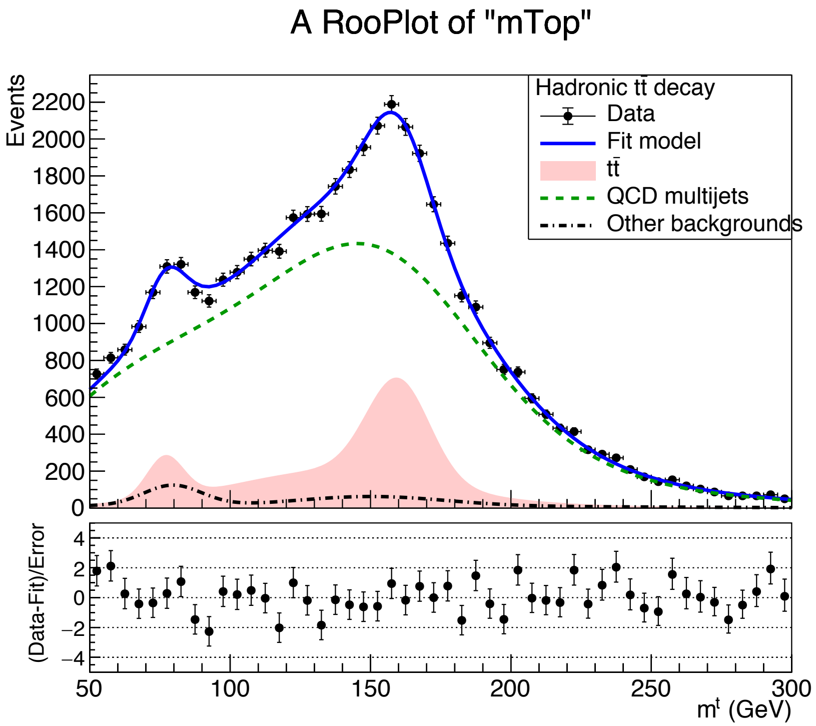
Result of the simultaneous fit on data in CR. The red line shows the ttbar contribution, the green line shows the QCD, and the black line shows the subdominant backgrounds

Simultaneous Fit in 3 regions for 2016, 2017, 2018 when eb is free (1btag region)

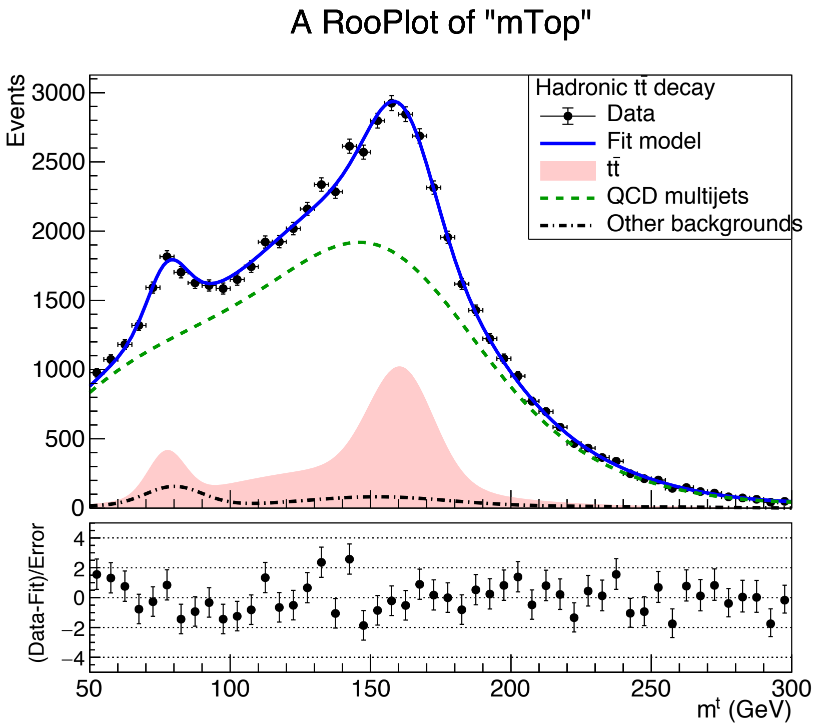
1btag region (2016)



1btag region (2017)

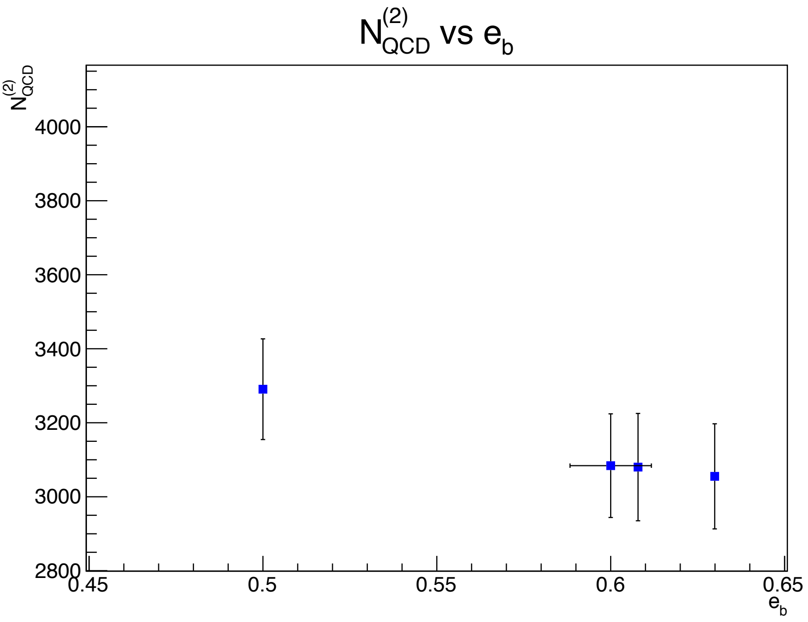


1btag region (2018)

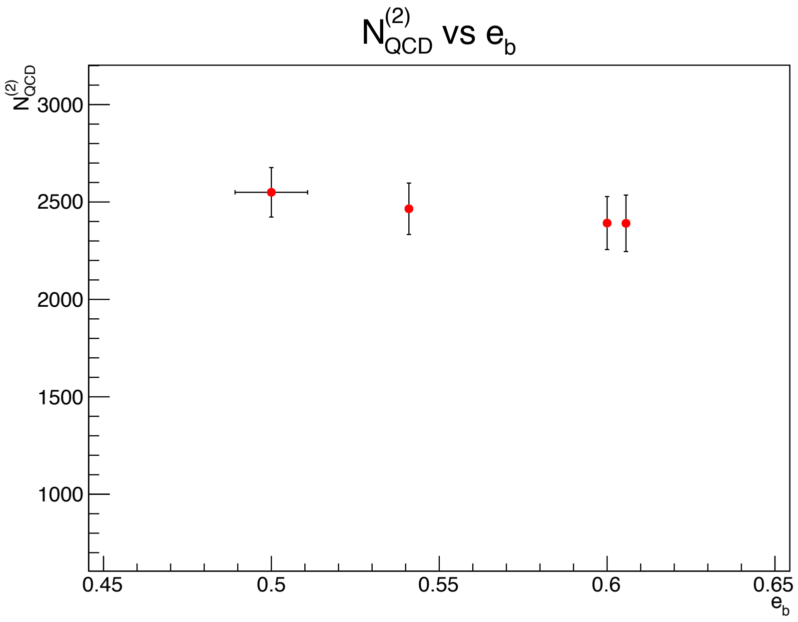


- We are checking for different values of e_b , the output of the $N^{(2)}_{QCD}$ for 2016 , 2017, 2018
- Calculated btagging efficiency for all years
  - btagging efficiency when the parameter is set as a free nuisance in the simultaneous fit
  - 2016: eb (fit) ≈ 0.61 and eb (calculated) ≈0.63
  - 2017: eb(fit) ≈ 0.55 and eb (calculated) ≈ 0.61
  - 2018: eb(fit) ≈ 0.57 and eb (calculated) ≈ 0.63

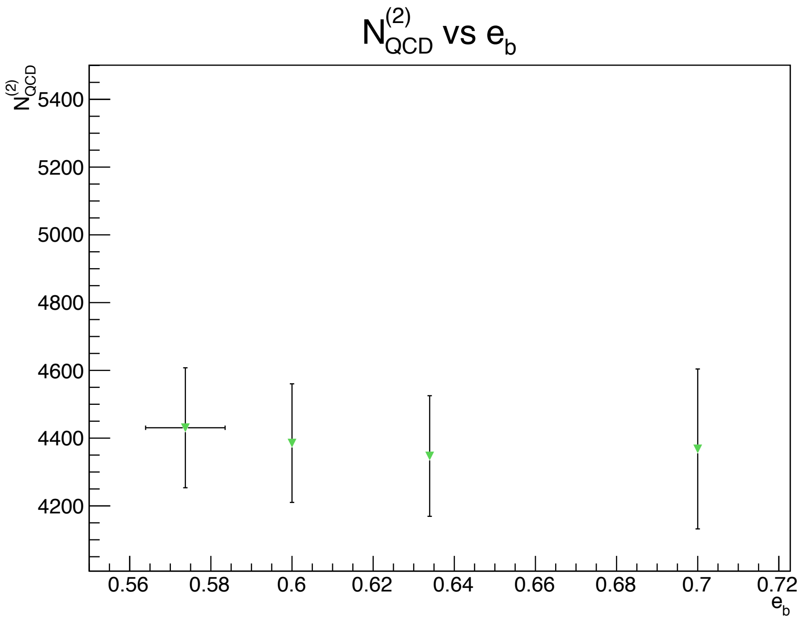
2016



2017



2018



Summary

- Delivered a new top Tagger that discriminates top jets from QCD jets

- Efficiencies and Acceptances for measured variables
  - New variables: top angular
  - btagging efficiency in the |eta|, pT subjet phase space

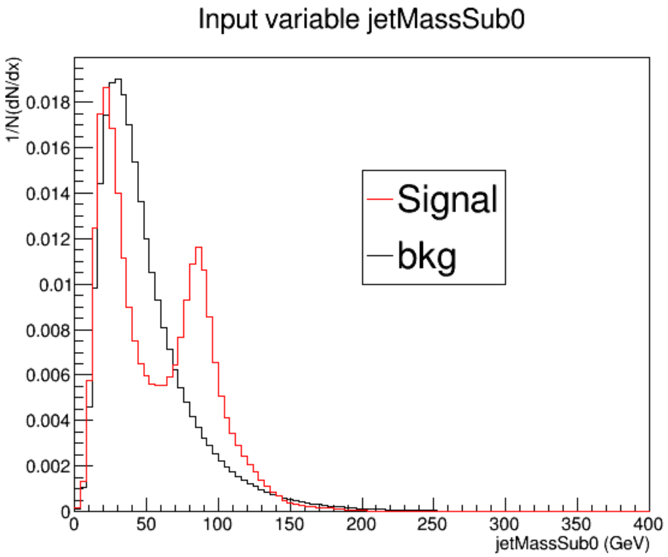
- QCD Closure tests to ensure that the shape of the QCD in the CR and in SR are consistent

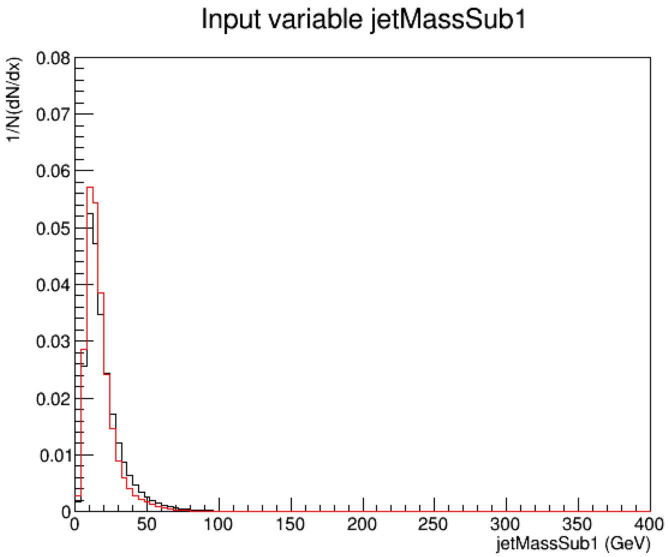
- ttbar contamination in the CR for measured variables

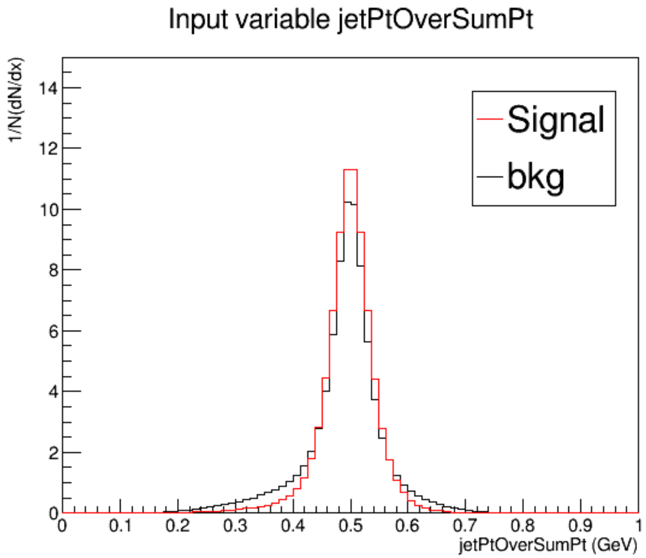
- Signal Extraction
  - Fit to extract the N_{QCD} in the extended Signal Region A (SR_A)
  - Simultaneous fit in 3 regions to suppress the ttbar contribution in the CR
  - N_{QCD} not affected from b-tagging efficiency

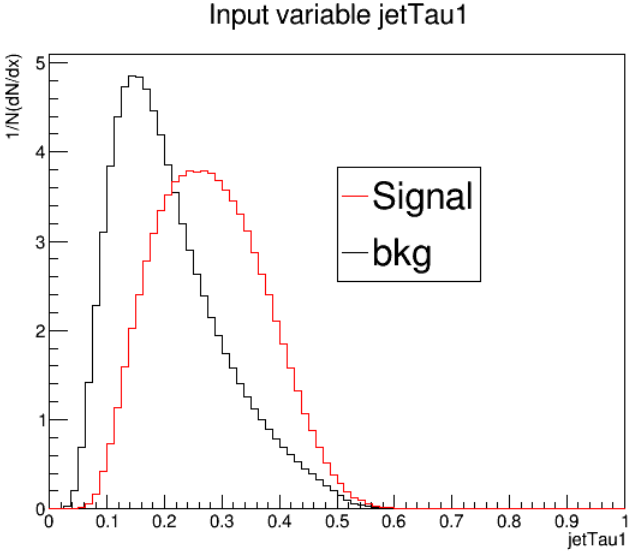
BACKUP SLIDES

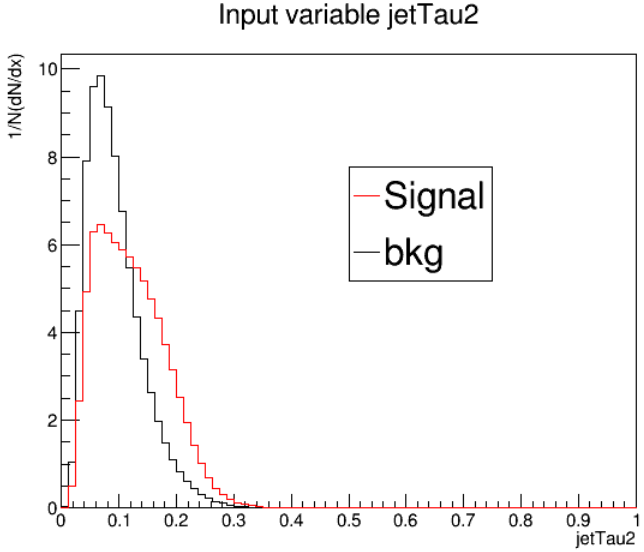
Training variables 2017

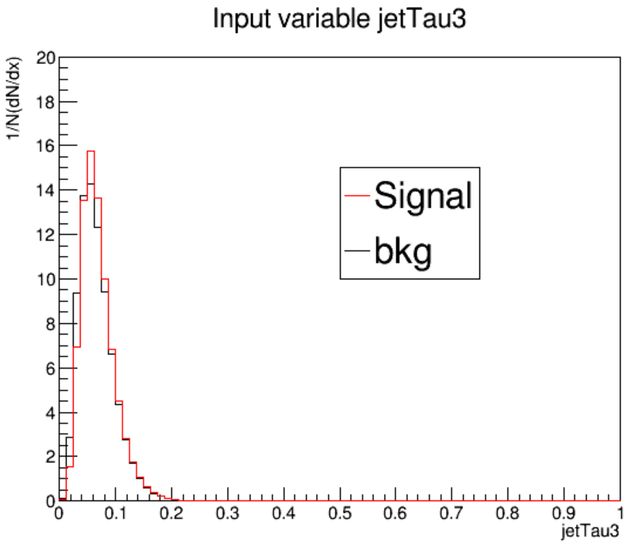




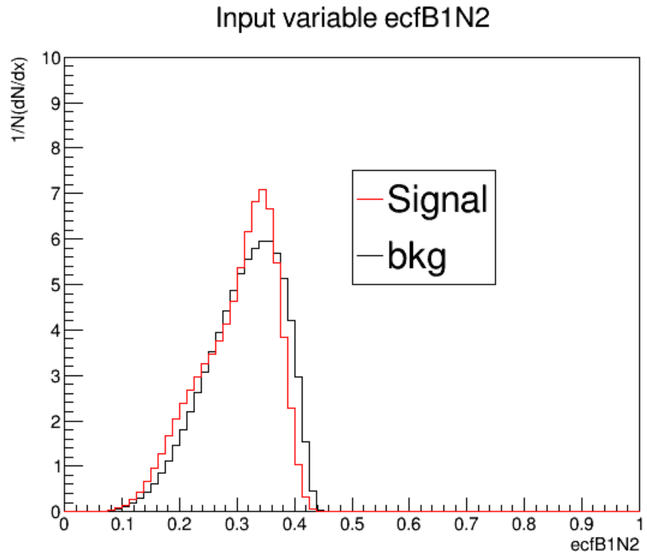


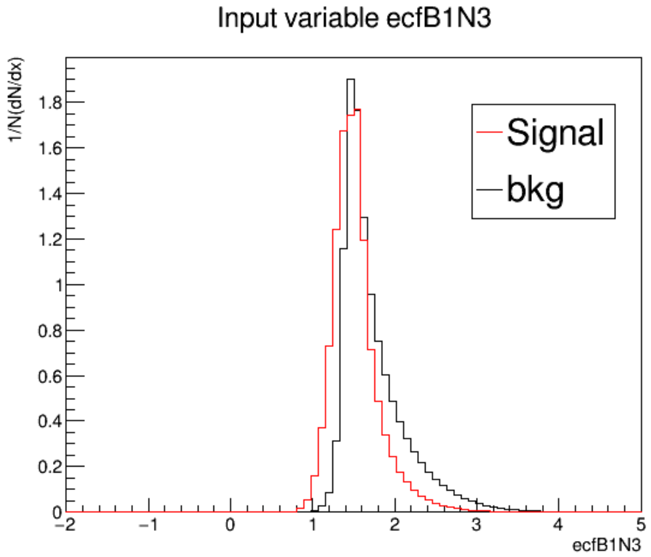


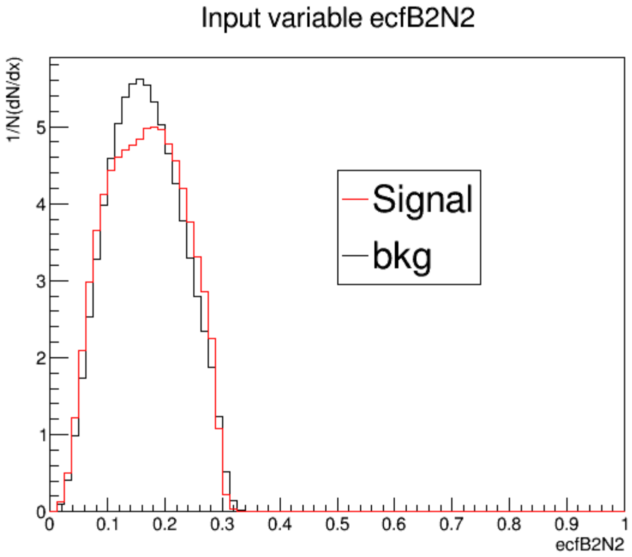


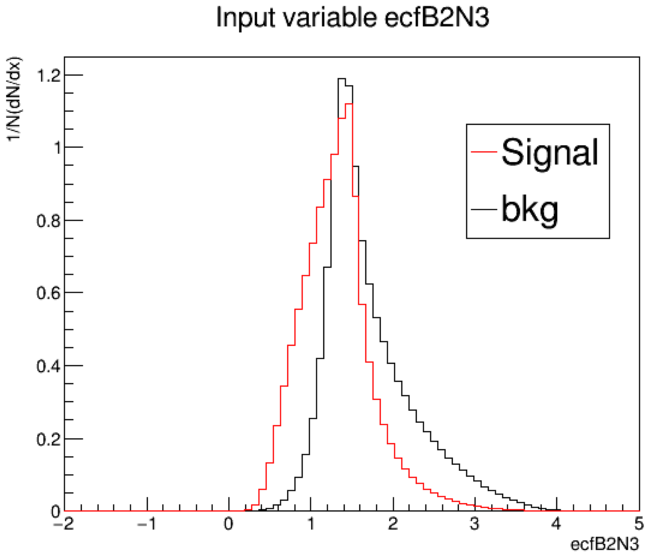


Training variables 2017









Signal Extraction

- Dominant Background is the QCD multijet production → mimic the topological substructure of a top decay jet
- We suppress the QCD using b-tagging and our newly developed topTagger
- Remaining contribution: data driven technique based on the assumption that if the b-tagging requirement is reverted:
  - Pure QCD sample
  - Jet kinematic properties are not affected
- To extract the signal:

$$S(x) = D(x) - R_{yield} N_{QCD} Q(x) - B(x)$$

- S(x) shape of the signal, D(x) shape of the data Q(x) is the QCD shape taken from the CR data and B(x) is the shape and contribution of the subdominant bkgs taken from simulation.
- R_{yield} is a transfer factor needed to get the QCD normalization from the signal region A to the signal region: $R_{yield} = \frac{N^{SR}}{N^{SR_A}}$
- N_{QCD} is the QCD absolute normalization taken from a fit on the leading jet top mass candidate variable → this is the reason why we perform a fit in the Signal Region A (SR_A)

$$D(m^t)^{(i)} = N_{tt}^{(i)} T^{(i)}(m^t, k_{MassScale}, k_{MassResolution}) + N_{bkg}^{(i)} B(m^t)(1 + k_1 x) + N_{sub}^{(i)} O^{(1)}(m^t)$$

where i is the region of interest (0, 1, 2) btag

- We perform a simultaneous fit in 3 regions (0, 1, 2) btag because we do not have a pure Control Region.
  - Our CR from data is contaminated because of the new topTagger
- $k_{MassScale}$, $k_{MassResolution}$: account for any differences between data and simulation in scale and resolution of the m^t
- $1+k_{slope}m^t$: linear modification factor to account for any difference observed in the Closure test of the QCD for the m^t

Simultaneous Fit in 3 regions

- Simultaneous fit in 3 regions (2btag, 1btag and 0btag)

$$D(x)^{(0)} = N_{tt}^{(0)} T^{(0)}(x, k_{MassScale}, k_{MassResolution}) + N_{bkg}^{(0)} B(x, \vec{p}) + N_{sub}^{(0)} O^{(0)}(x)$$

$$D(x)^{(2)} = N_{tt}^{(2)} T^{(1)}(x, k_{MassScale}, k_{MassResolution}) + N_{bkg}^{(2)} B(x, \vec{p})(1 + k_1 x) + N_{sub}^{(2)} O^{(1)}(x)$$

$$D(x)^{(1)} = N_{tt}^{(1)} T^{(2)}(x, k_{MassScale}, k_{MassResolution}) + N_{bkg}^{(1)} B(x, \vec{p})(1 + k_2 x) + N_{sub}^{(1)} O^{(2)}(x)$$

  - We do a simultaneous fit because we do not have a pure Control Region.
    - Our CR from data is contaminated because of the new topTagger
  - $N_{sub}^{(0)}$ is limited in $0.9 N_{sub,MC}^{(0)}$ *up to* $1.1 N_{sub,MC}^{(0)}$
  - We assume that $N_{tt}^{(0)} = (1 - \mathrm{e_b})^2 \mathrm{N_{tt}}, \mathrm{N_{tt}^{(2)}} = \mathrm{e_b^2 N_{tt}}$ and $\mathrm{N_{tt}^{(1)}} = 2(1 - e_b) e_b N_{tt}$ where e_b is the b tagging efficiency and N_{tt} is the total ttbar yield.

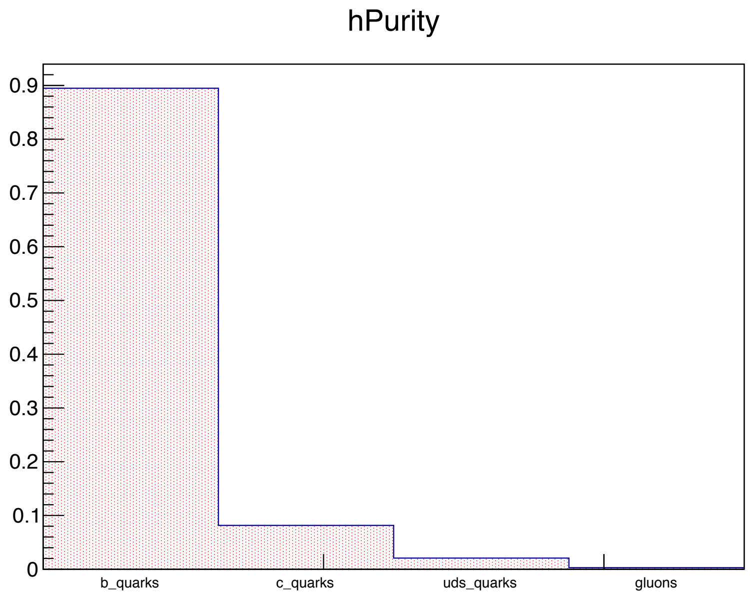
  We can either leave e_b and N_{tt} as free parameters in the fit or $N_{tt}^{(0)}, N_{tt}^{(1)}, N_{tt}^{(2)}$

- btagging efficiency and the Ntt yield are highly correlated.
  - We decided to try and fix the btagging parameter → calculated b-tagging
  - For the btagging efficiency calculation:

$$e_b = \frac{\#subjets\ with\ flavour\ id\ requirement + deepCSV\ btagged}{\#subjets\ with\ flavour\ id\ requirement\ (b)}$$, where all selected events pass baseline + parton selection

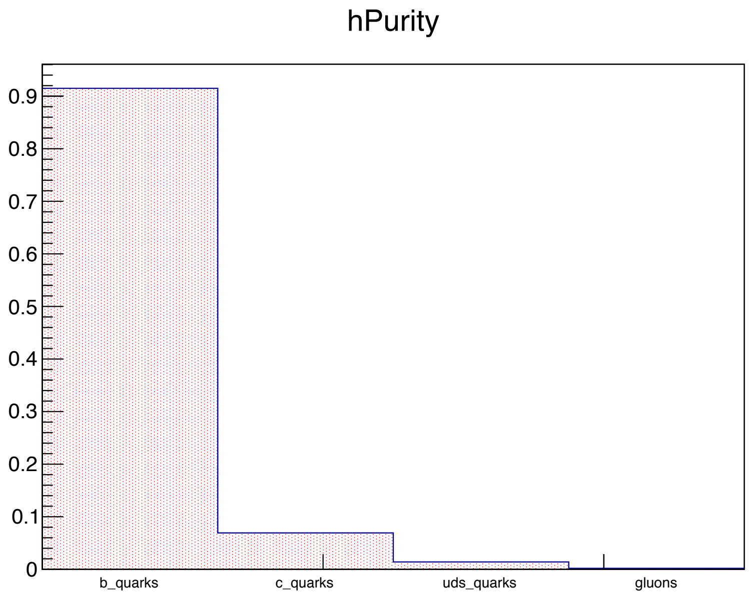
Btagging purity

2016



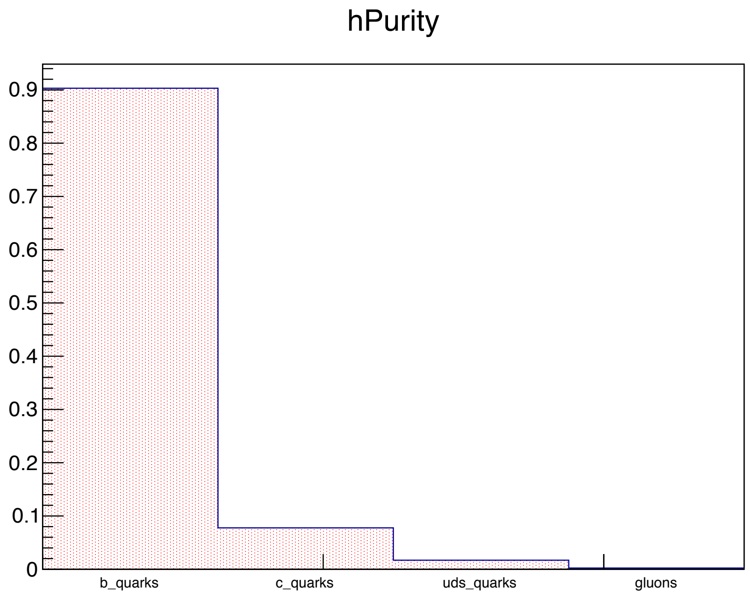
Purity ≈ 0. 894

2017



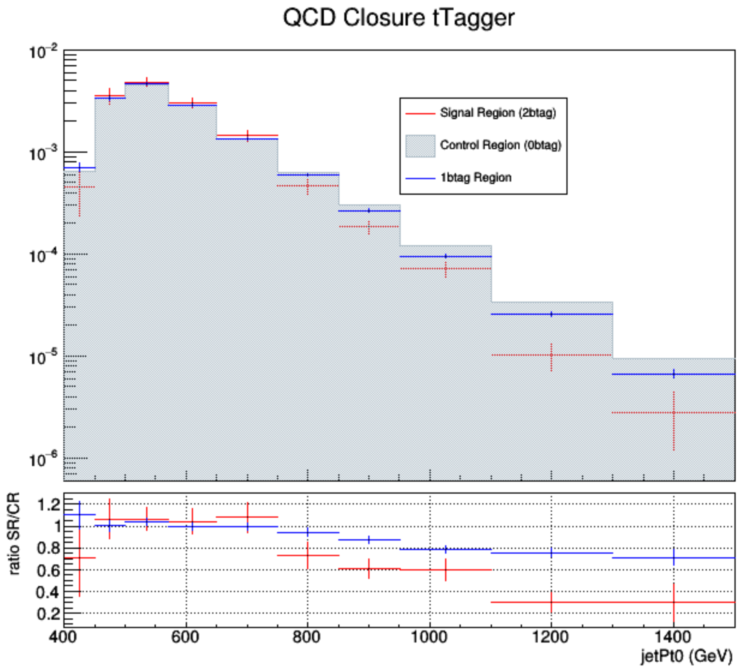
Purity ≈ 0.915

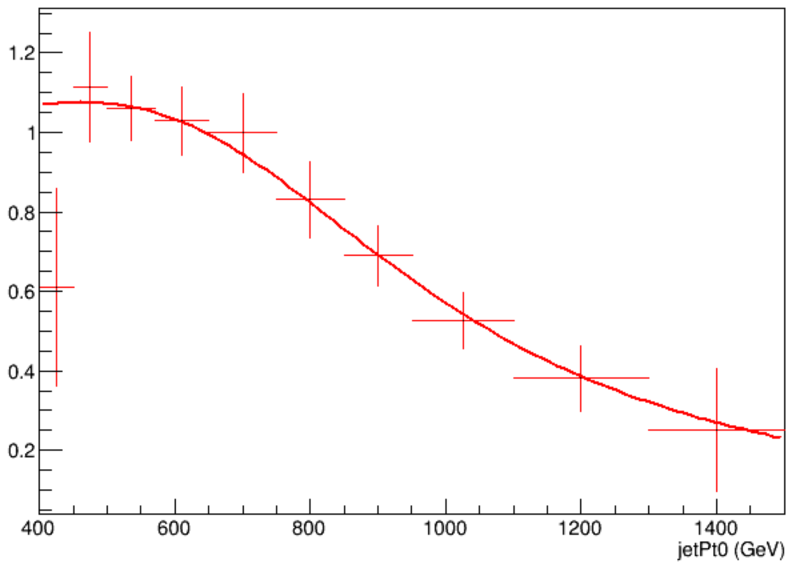
2018

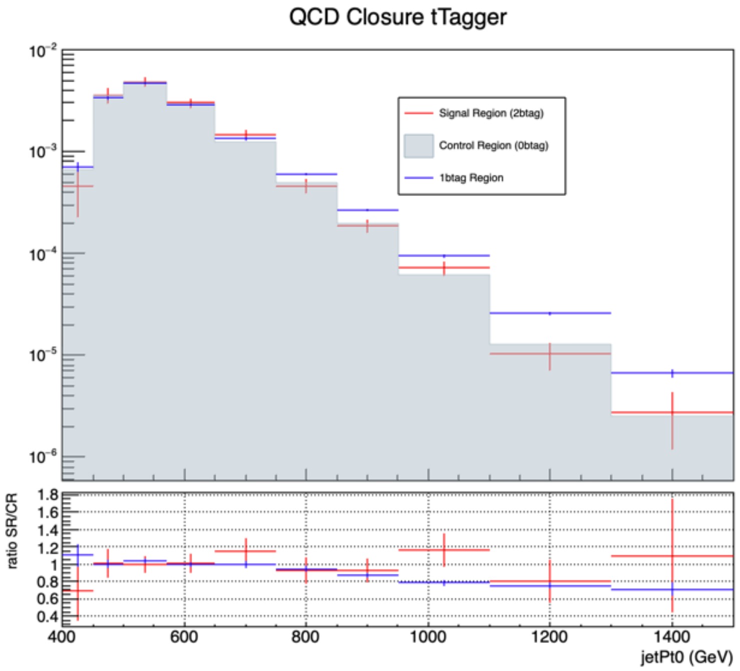


Purity ≈ 0.903

2017 QCD Closure and the fit ratio

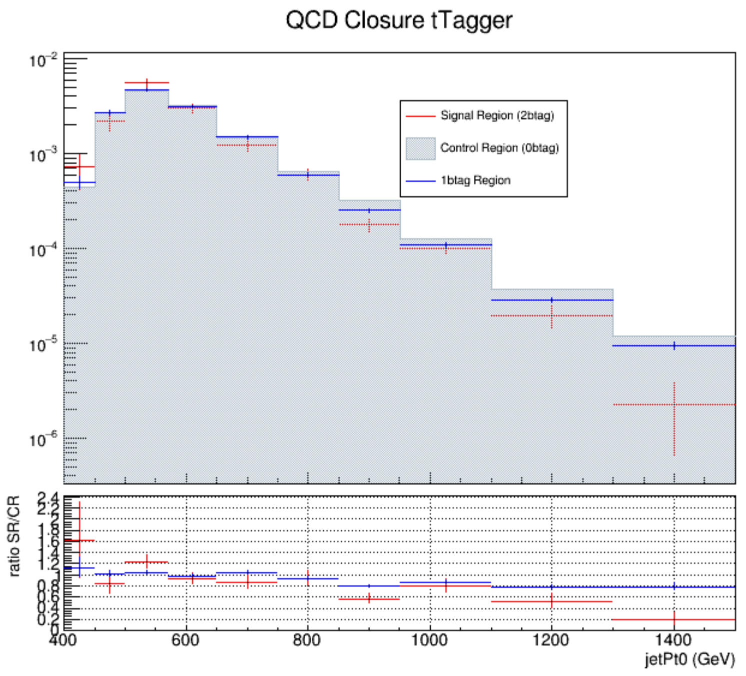


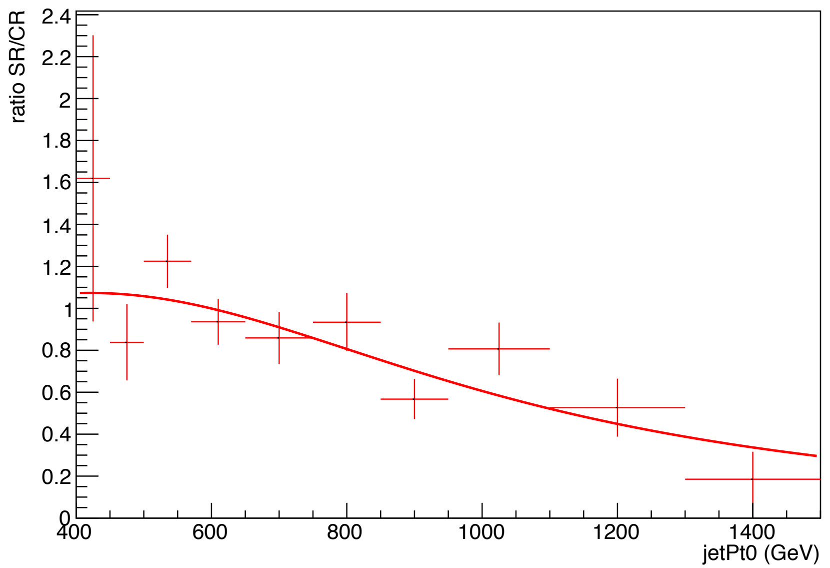


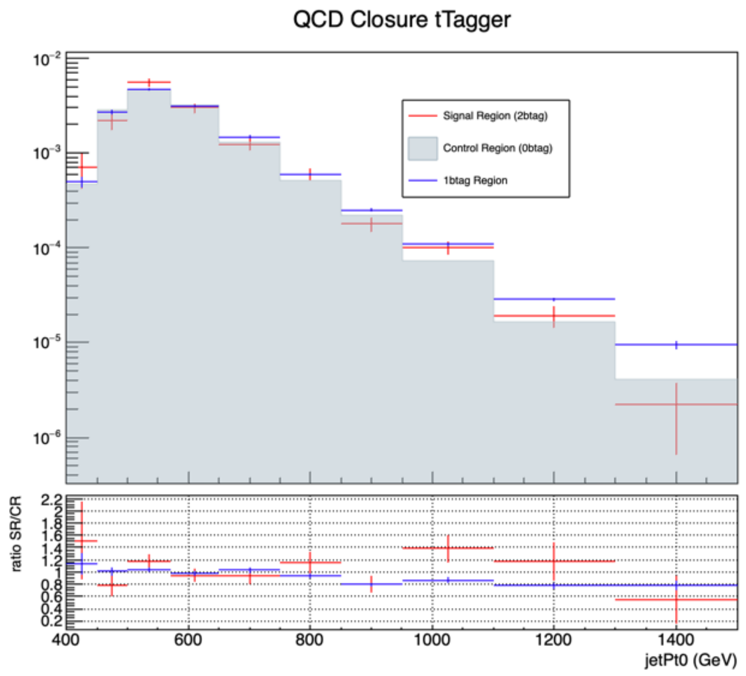


Discovered shape in the SR/CR → Fit the ratio with the function: → Output of the new QCD closure

2018 QCD Closure and the fit ratio



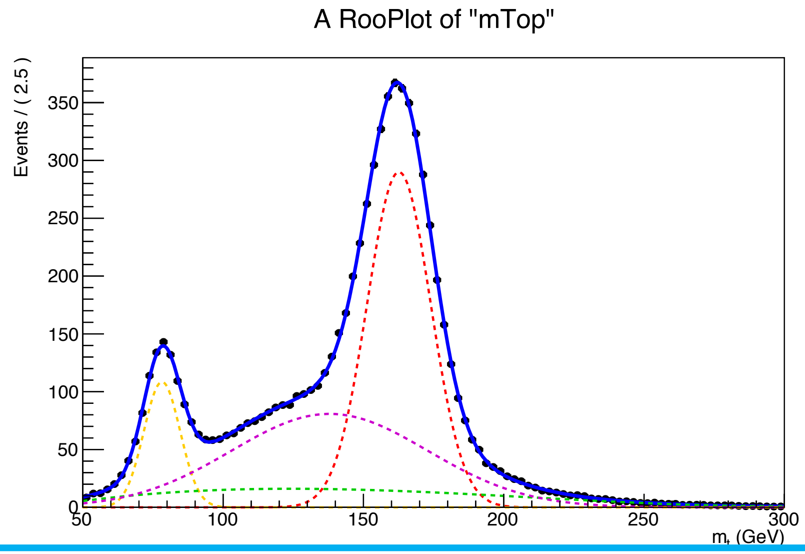


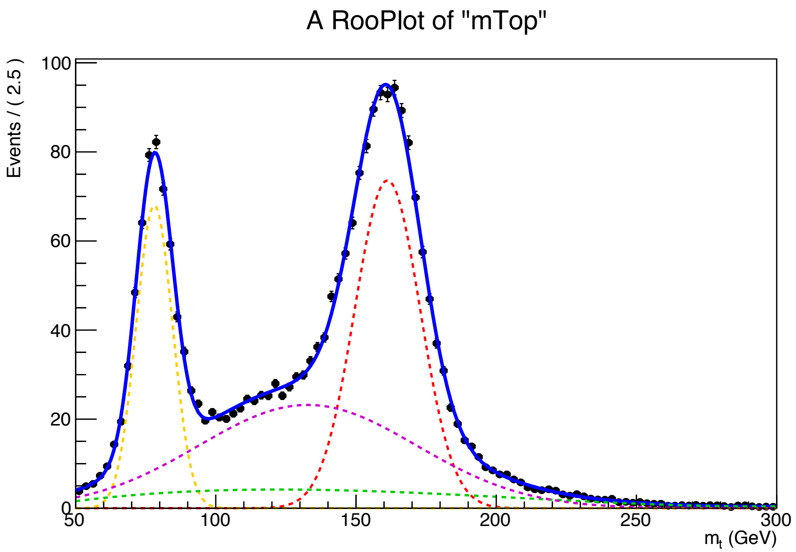


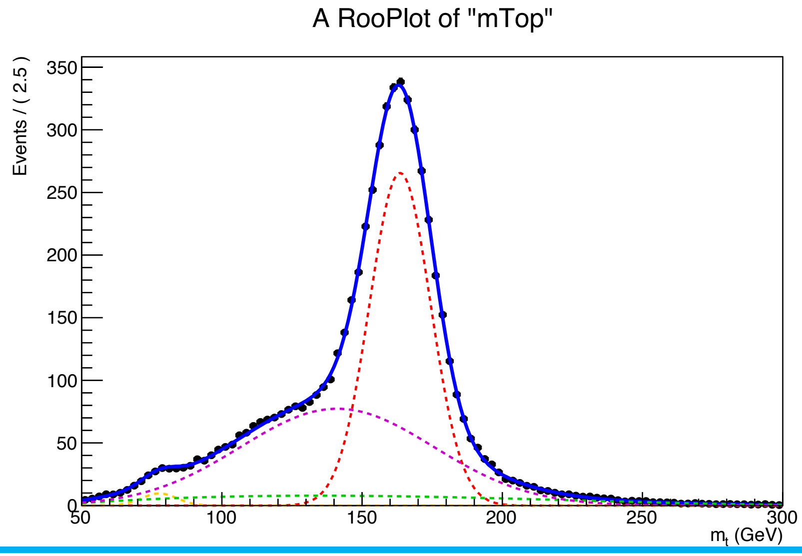
Discovered shape in the SR/CR → Fit the ratio with the function: → Output of the new QCD closure

Template fit results Signal Region A

2016



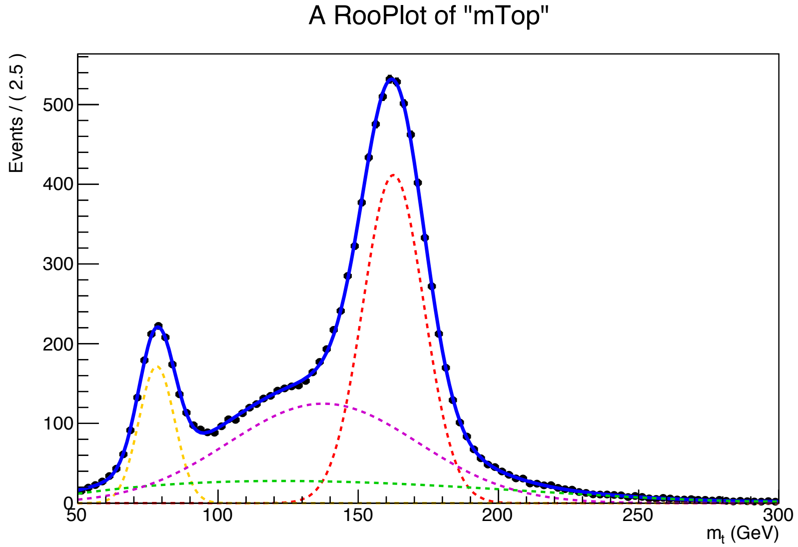


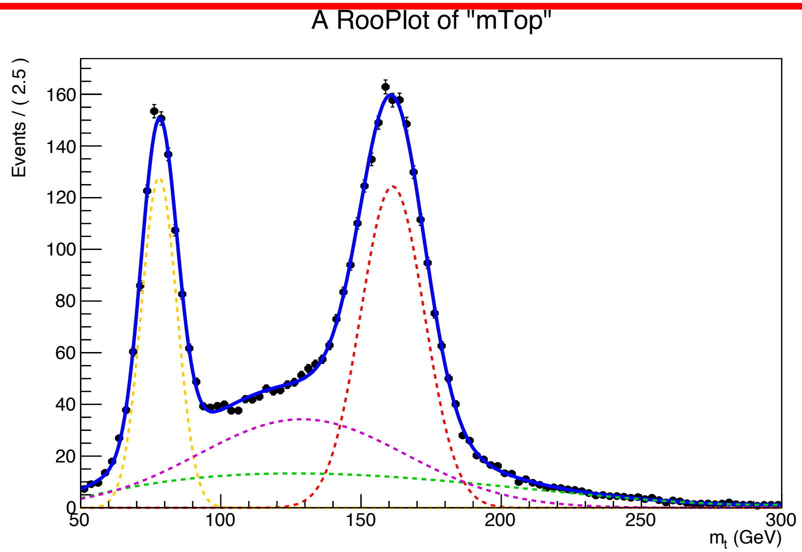


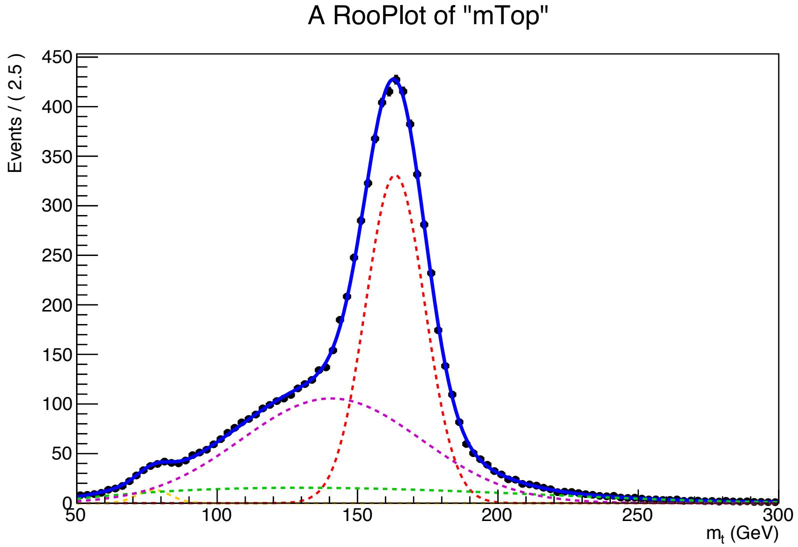
1btag region

0btag region

2btag region



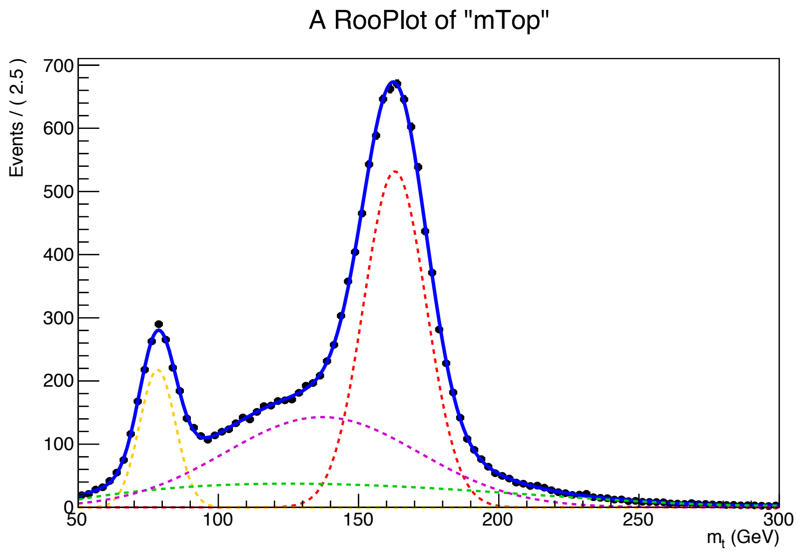




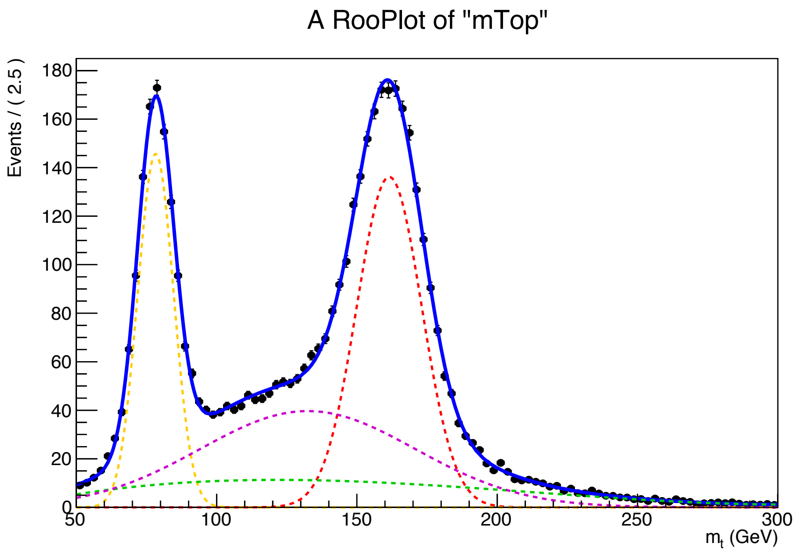
2017

Template fit results Signal Region A

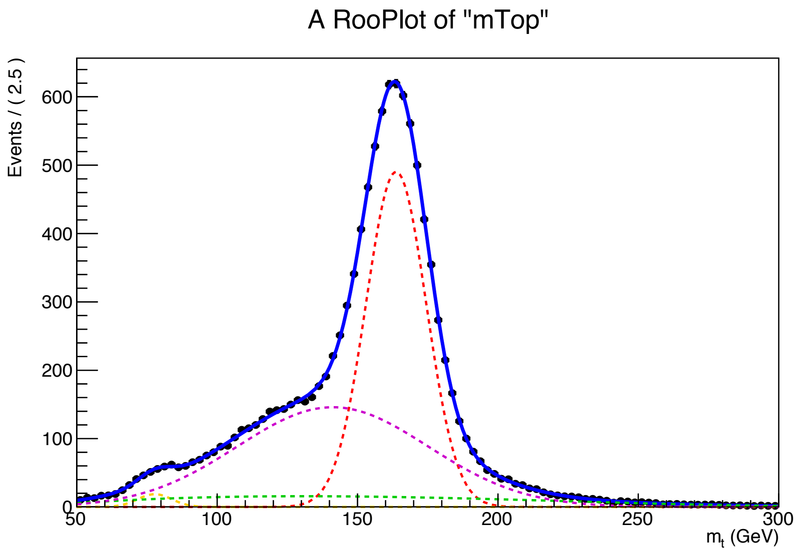
2018



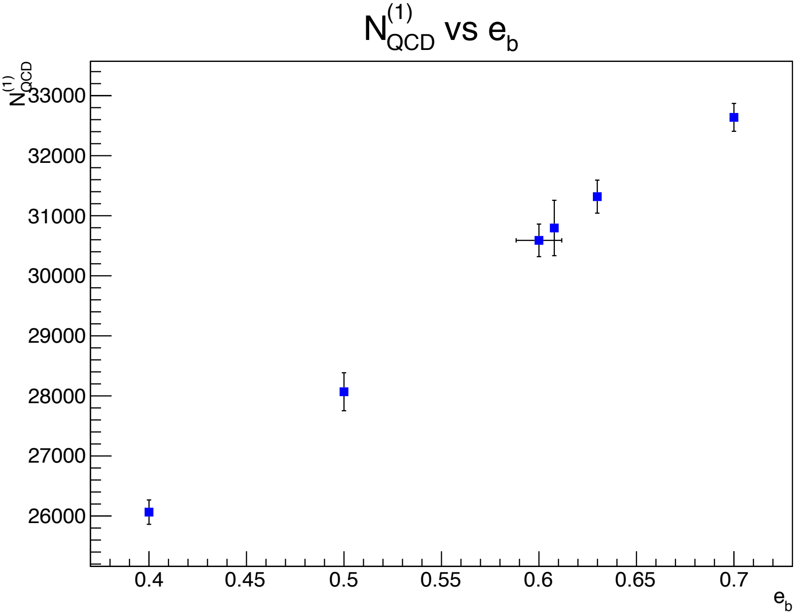
1btag region

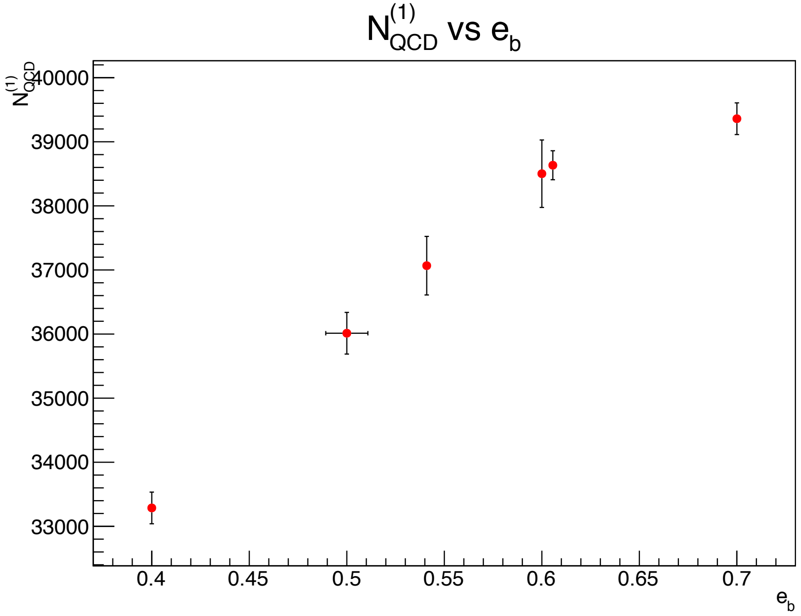


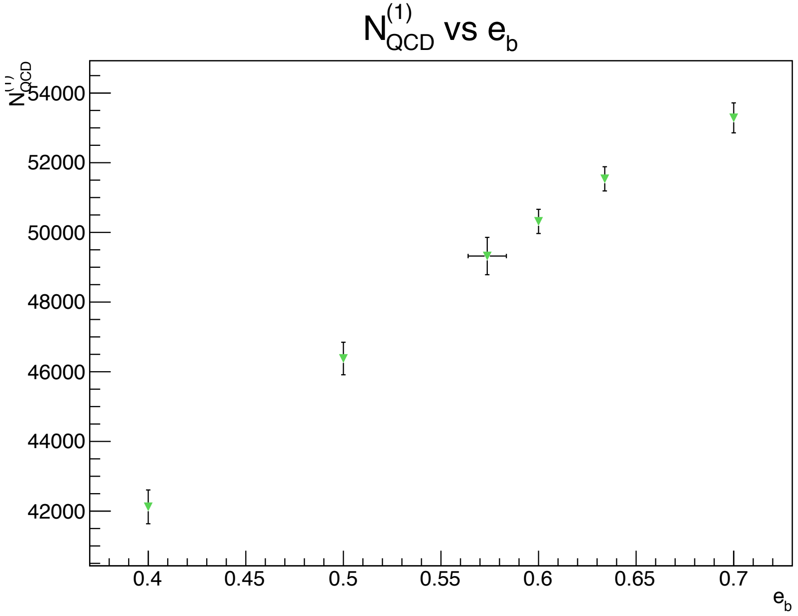
0btag region



2btag region



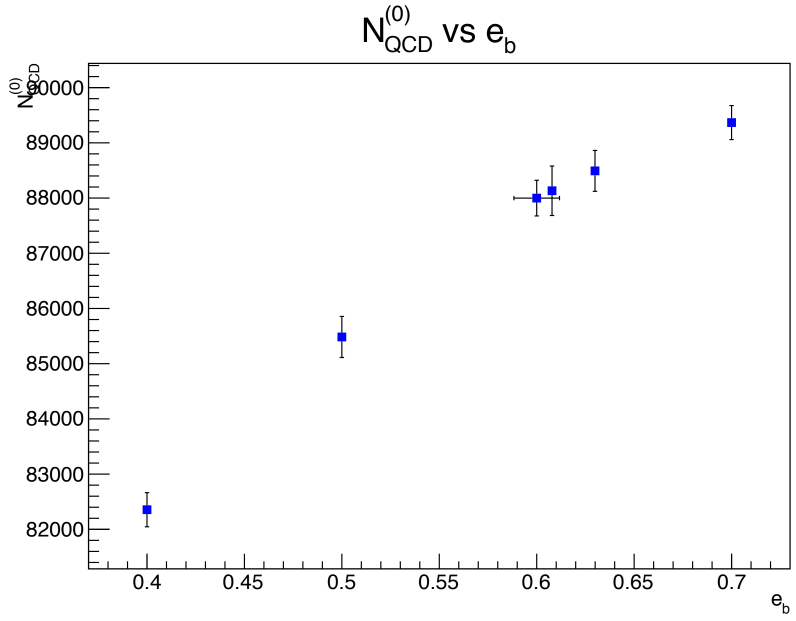


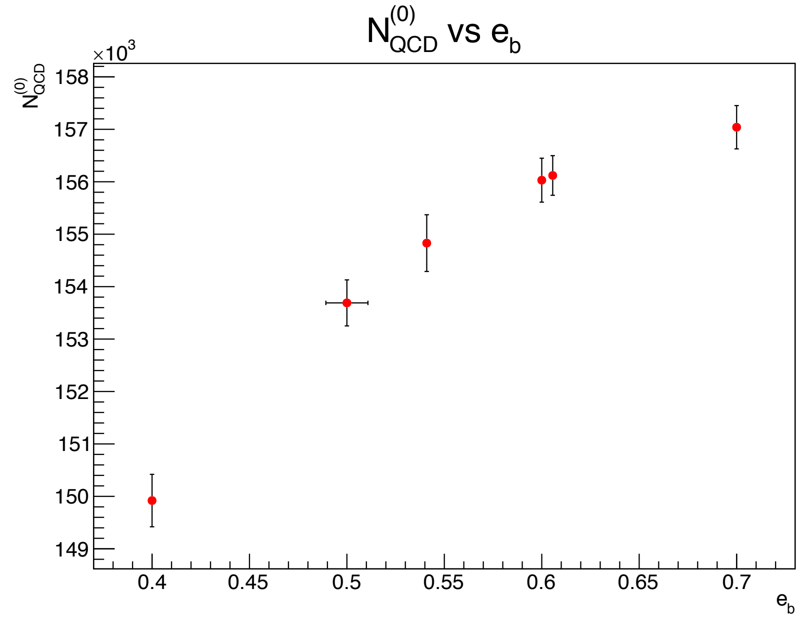


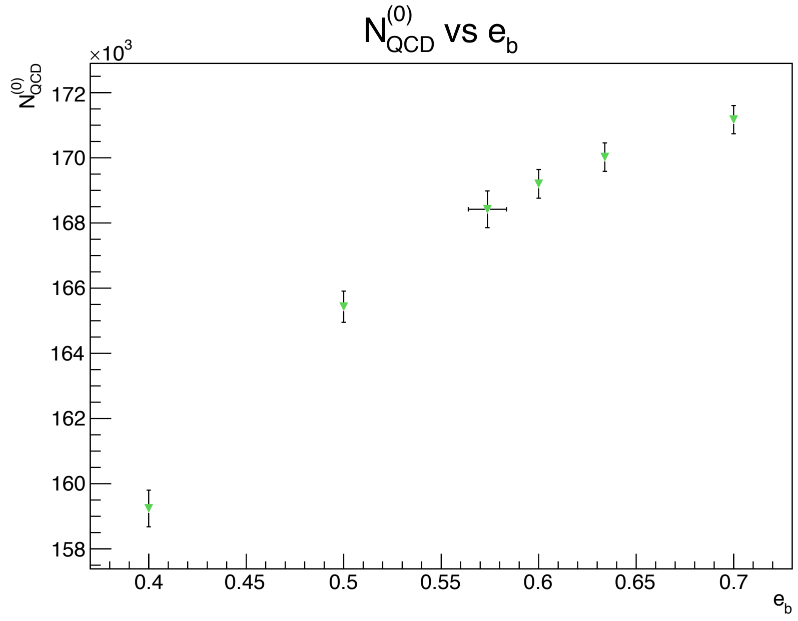
2016

2017

2018







Simultaneous Fit in 3 regions for 2016 ,2017 and 2018 (nuisances) with free eb

2016

Floating Parameter	FinalValue +/- Error
btagEff	6.0786e-01 +/- 1.53e-02
kMassResol	9.5079e-01 +/- 2.42e-02
kMassScale	1.0009e+00 +/- 1.76e-03
kQCD_1b	6.0406e-03 +/- 5.05e-04
kQCD_2b	9.6764e-02 +/- 6.83e-02
nFitBkg_0b	4.5268e+03 +/- 6.84e+02
nFitBkg_1b	1.9110e+03 +/- 3.13e+02
nFitBkg_2b	1.9415e+02 +/- 2.48e+01
nFitQCD_0b	8.8132e+04 +/- 4.48e+02
nFitQCD_1b	3.0796e+04 +/- 4.61e+02
nFitQCD_2b	3.0802e+03 +/- 1.45e+02
nFitSig	1.4153e+04 +/- 7.33e+02
qcd_b0	6.6661e-01 +/- 5.74e-01
qcd_b1	1.4002e+00 +/- 1.62e+00
qcd_b2	3.3898e-02 +/- 5.46e-02
qcd_b3	3.3916e-02 +/- 3.37e-02
qcd_b4	1.6260e-02 +/- 1.52e-02
qcd_f1	6.9097e-01 +/- 2.57e-02
qcd_mean	1.5055e+02 +/- 1.05e+00
qcd_sigma	3.3739e+01 +/- 1.30e+00

Ntt expected: 16351
Ntt observed: 14153
Signal strength r: 0.865584

2017

Floating Parameter	FinalValue +/- Error
btagEff	5.4097e-01 +/- 1.39e-02
kMassResol	1.0289e+00 +/- 3.03e-02
kMassScale	9.8332e-01 +/- 2.06e-03
kQCD_1b	4.2097e-03 +/- 3.15e-04
kQCD_2b	1.5460e-02 +/- 5.16e-03
nFitBkg_0b	4.0852e+03 +/- 5.03e+02
nFitBkg_1b	1.6992e+03 +/- 2.84e+02
nFitBkg_2b	2.1330e+02 +/- 4.44e+01
nFitQCD_0b	1.5483e+05 +/- 5.41e+02
nFitQCD_1b	3.7067e+04 +/- 4.57e+02
nFitQCD_2b	2.4652e+03 +/- 1.32e+02
nFitSig	1.5594e+04 +/- 7.87e+02
qcd_b0	5.2866e-01 +/- 8.49e-02
qcd_b1	1.1003e+00 +/- 1.63e-01
qcd_b2	8.4662e-02 +/- 3.49e-02
qcd_b3	5.5285e-02 +/- 1.06e-02
qcd_b4	1.9483e-02 +/- 3.70e-03
qcd_f1	7.1106e-01 +/- 1.24e-02
qcd_mean	1.5184e+02 +/- 6.37e-01
qcd_sigma	3.2820e+01 +/- 6.76e-01

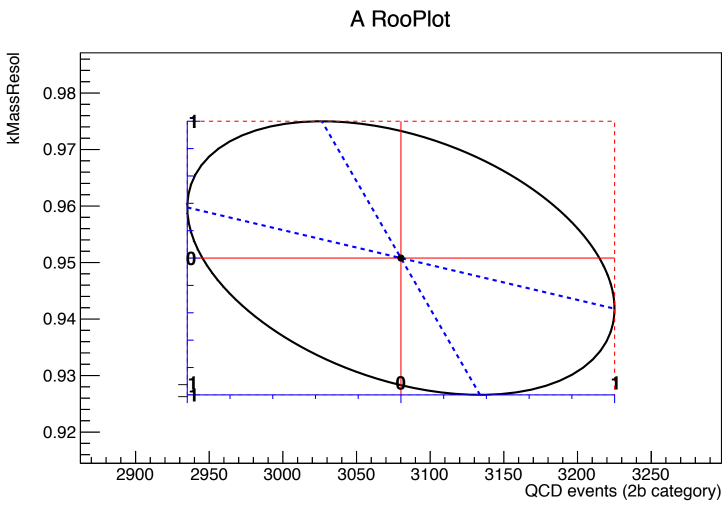
Ntt expected: 23721
Ntt observed: 15594
Signal strength r: 0.657402

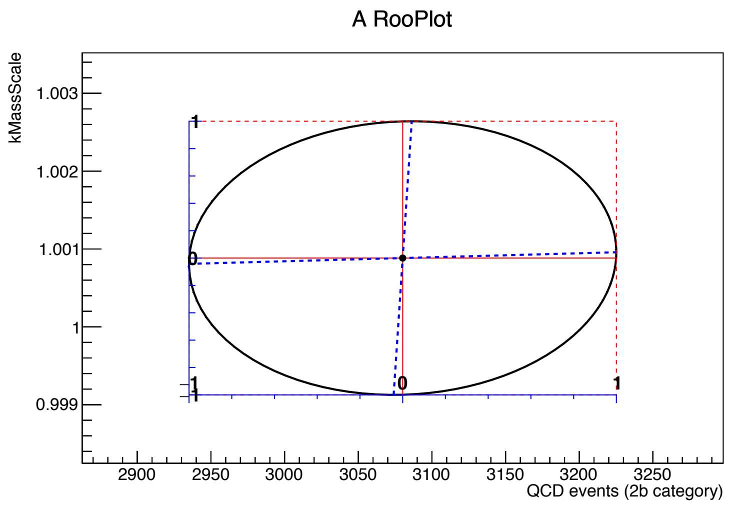
2018

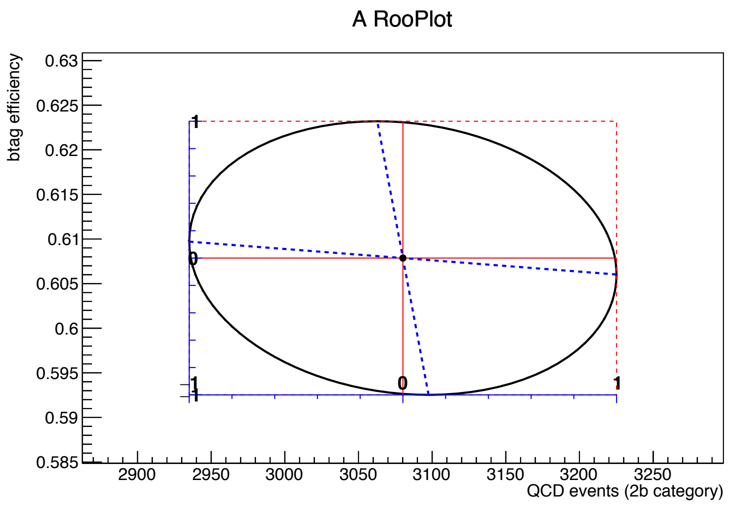
Floating Parameter	FinalValue +/- Error
btagEff	5.7371e-01 +/- 1.12e-02
kMassResol	1.0251e+00 +/- 2.43e-02
kMassScale	9.8728e-01 +/- 1.61e-03
kQCD_1b	3.6804e-03 +/- 2.60e-04
kQCD_2b	1.3533e-02 +/- 3.24e-03
nFitBkg_0b	4.1842e+03 +/- 5.05e+02
nFitBkg_1b	2.1675e+03 +/- 3.15e+02
nFitBkg_2b	3.4159e+02 +/- 7.52e+01
nFitQCD_0b	1.6842e+05 +/- 5.64e+02
nFitQCD_1b	4.9321e+04 +/- 5.36e+02
nFitQCD_2b	4.4306e+03 +/- 1.77e+02
nFitSig	2.2906e+04 +/- 8.92e+02
qcd_b0	4.0935e-01 +/- 6.23e-02
qcd_b1	8.5950e-01 +/- 1.25e-01
qcd_b2	1.0234e-01 +/- 2.32e-02
qcd_b3	2.4955e-02 +/- 6.63e-03
qcd_b4	1.1844e-02 +/- 2.02e-03
qcd_f1	7.3124e-01 +/- 1.01e-02
qcd_mean	1.5274e+02 +/- 5.33e-01
qcd_sigma	3.1322e+01 +/- 6.11e-01

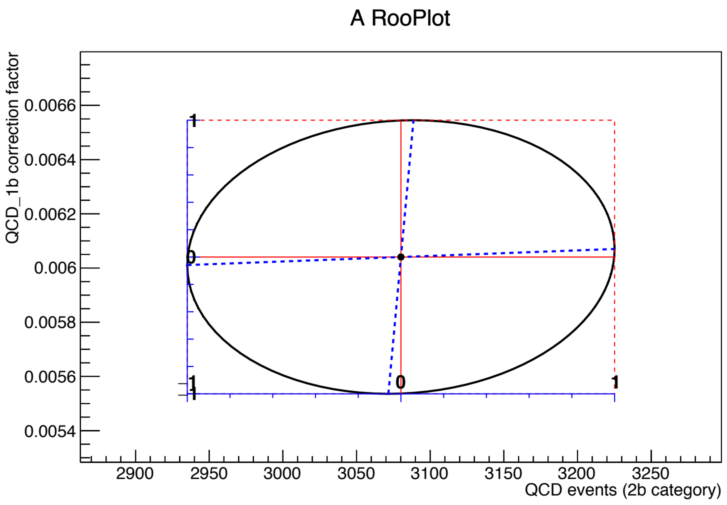
Ntt expected: 30676
Ntt observed: 22906
Signal strength r: 0.746688

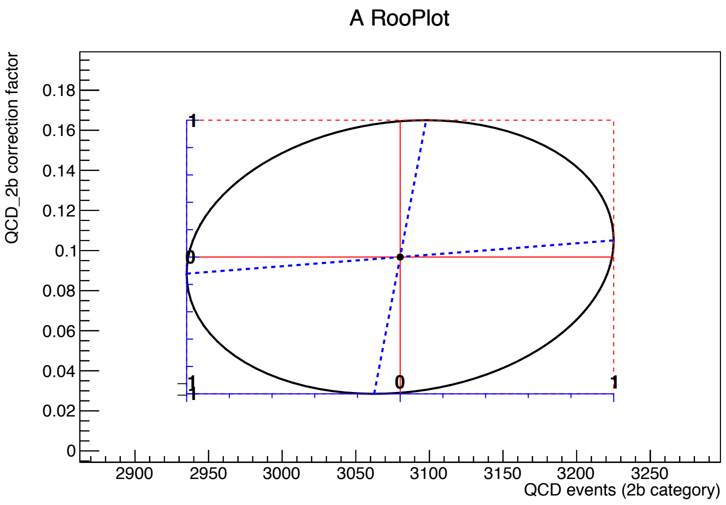
Correlation plots $N_{QCD(2)}$ vs all nuisances from fit when eb runs free for 2016

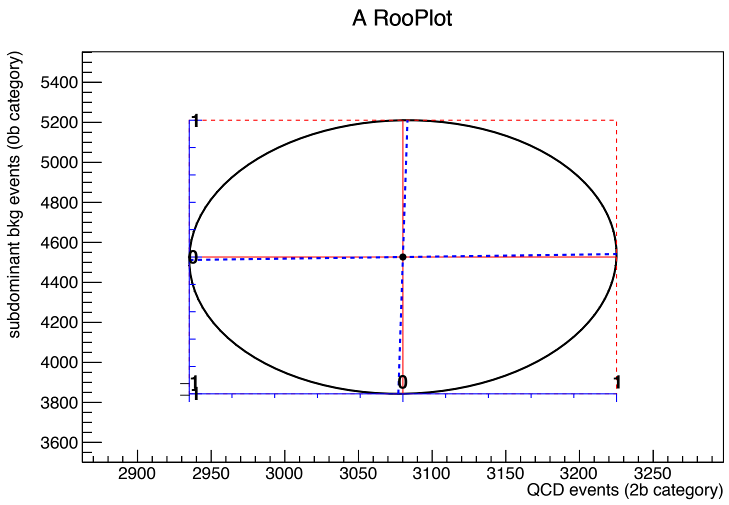




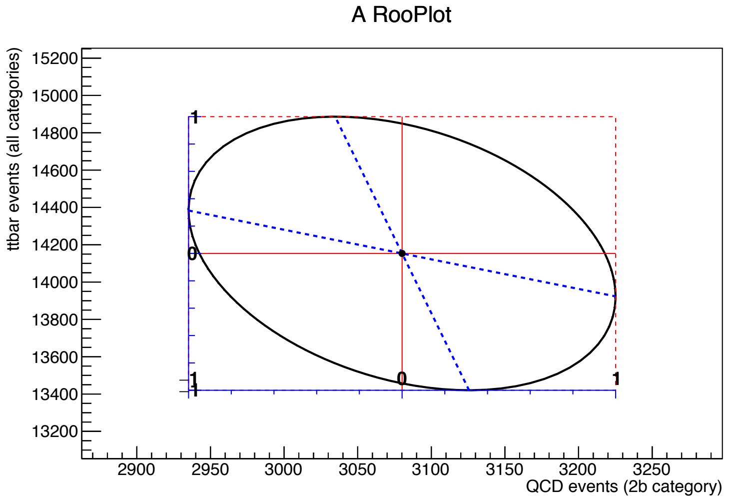


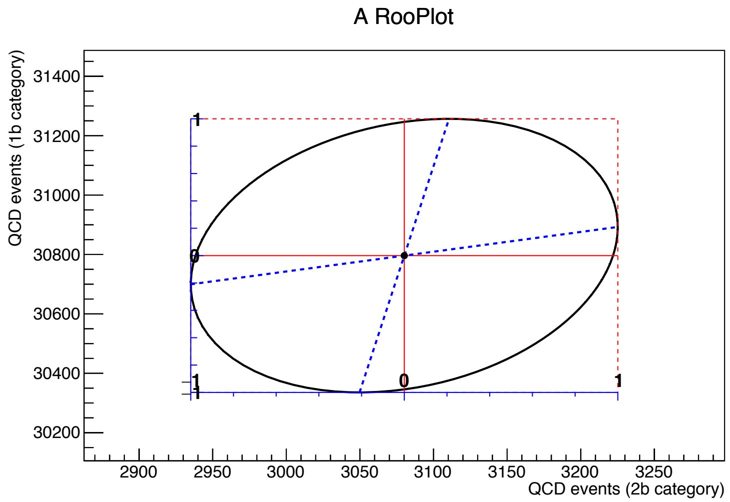


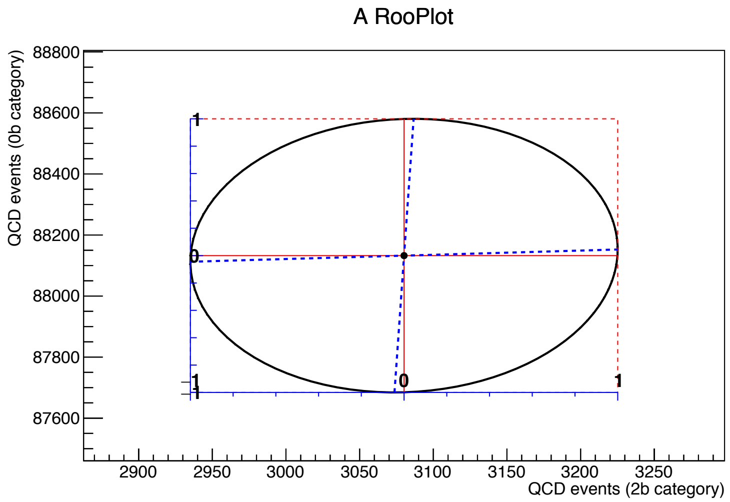


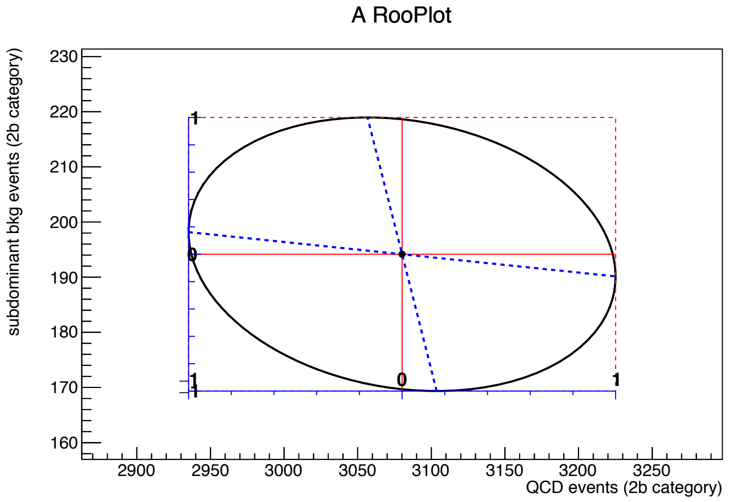


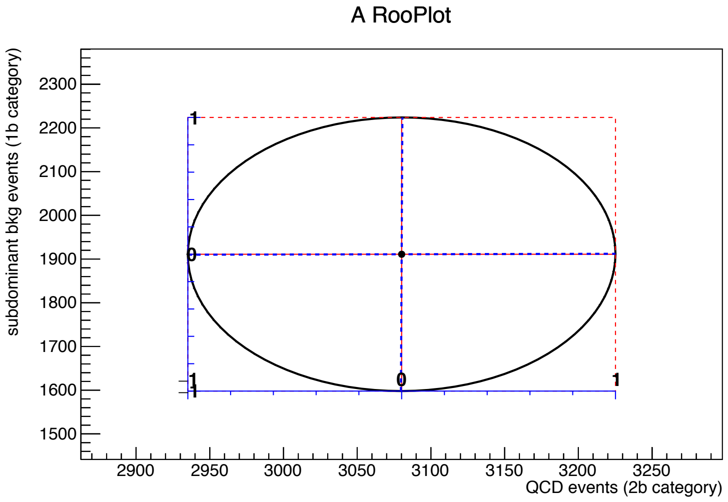
Correlation plots $N_{QCD(2)}$ vs all nuisances from fit when eb runs free for 2016



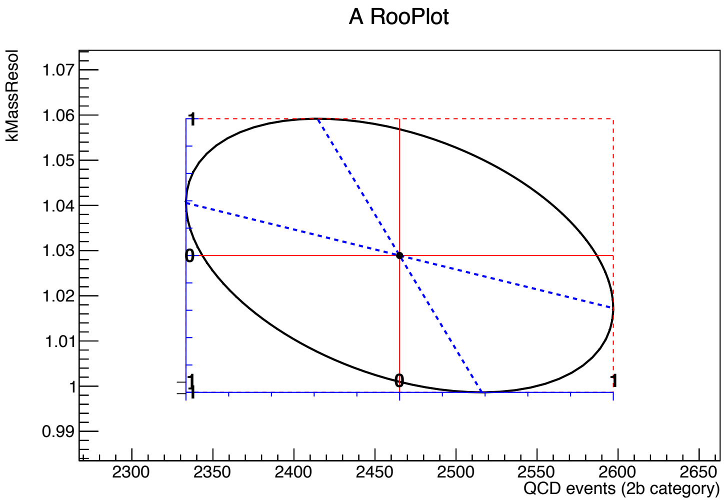


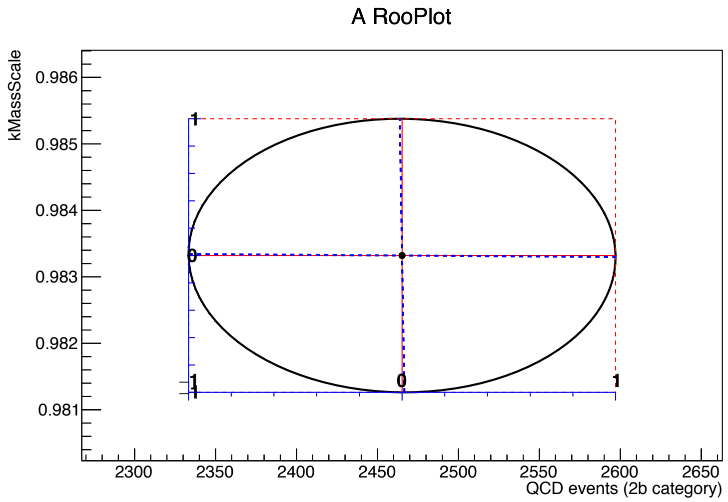


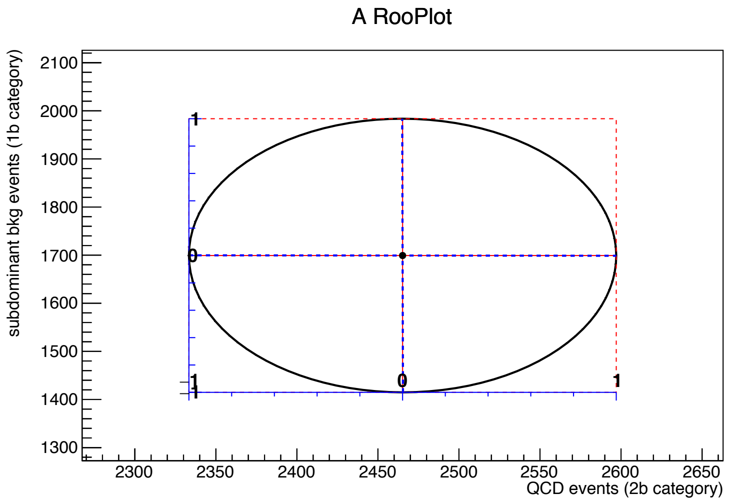


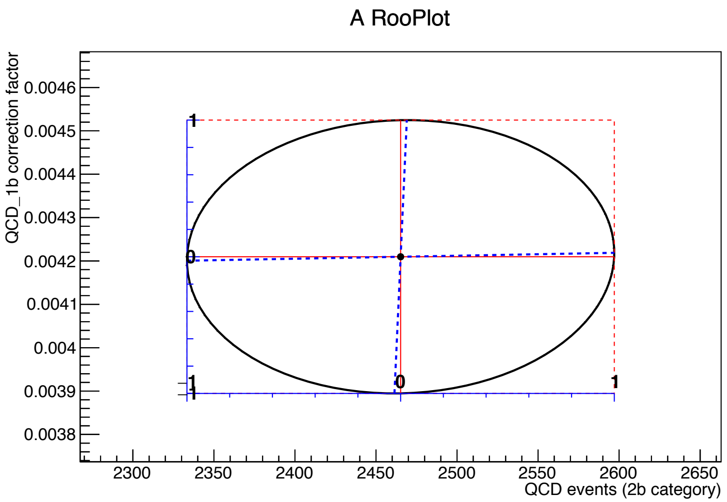


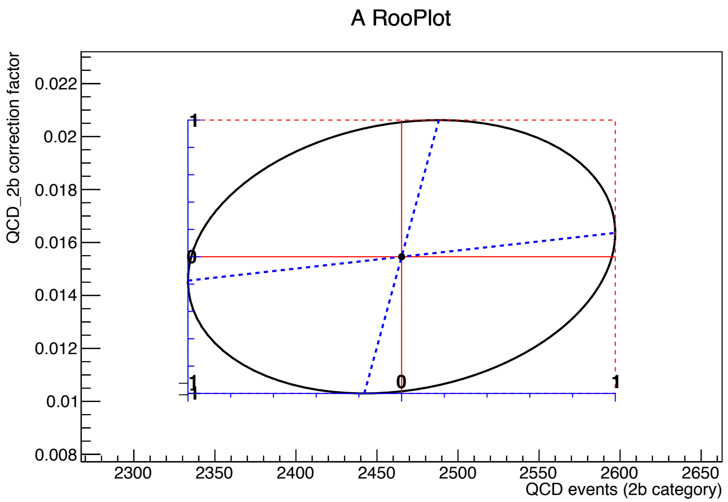
Correlation plots $N_{QCD(2)}$ vs all nuisances from fit when eb runs free for 2017

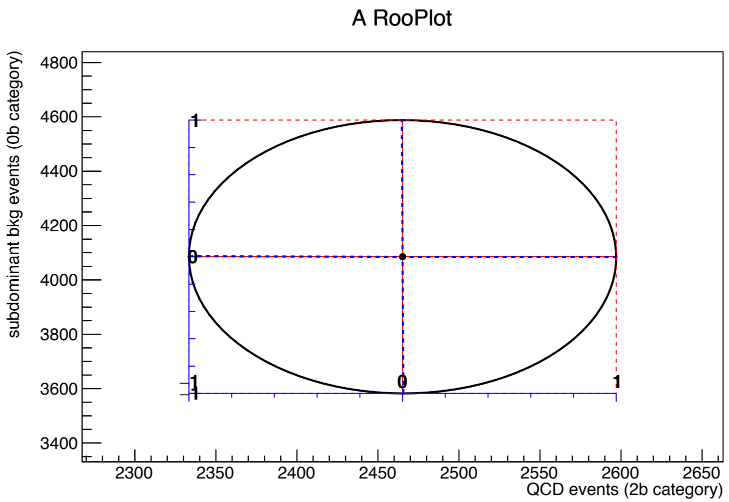




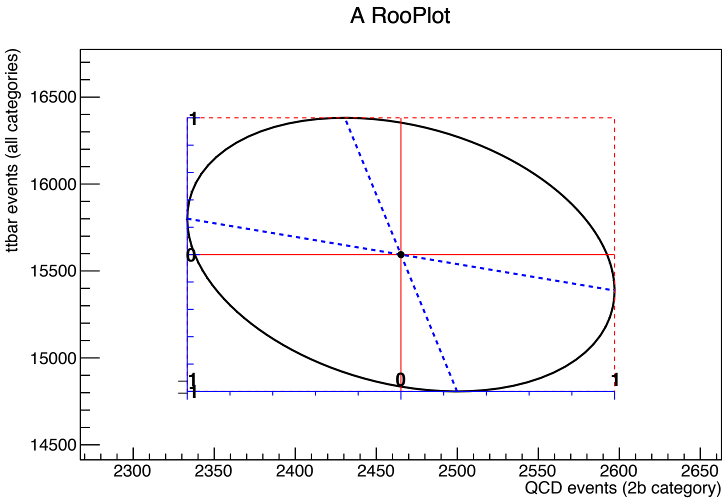


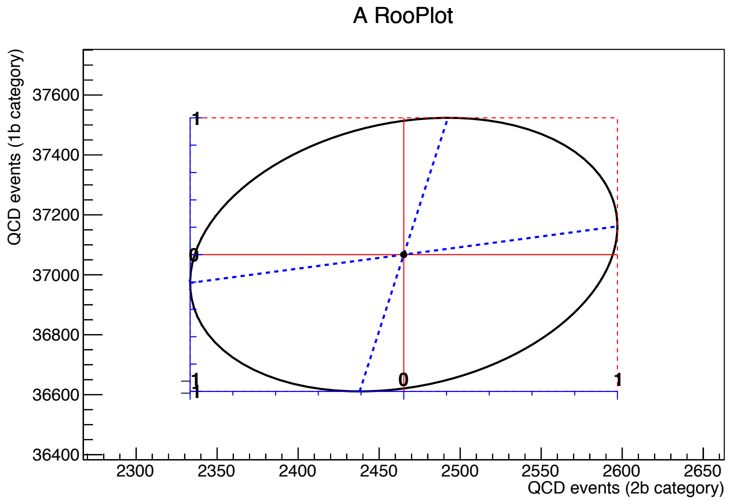


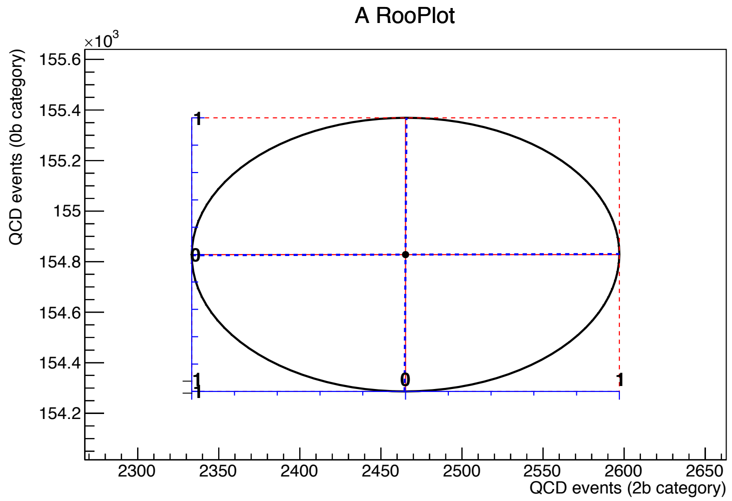


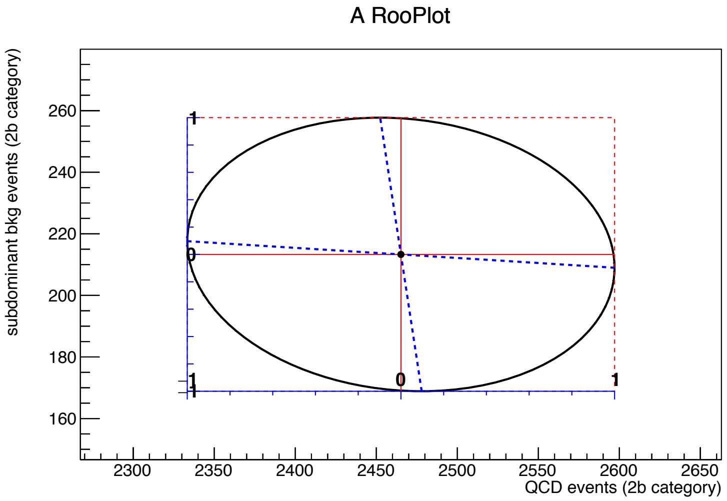


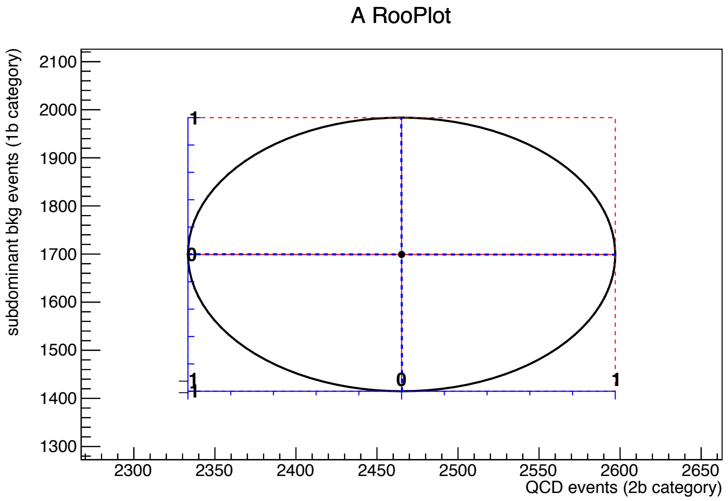
Correlation plots $N_{QCD(2)}$ vs all nuisances from fit when eb runs free for 2017



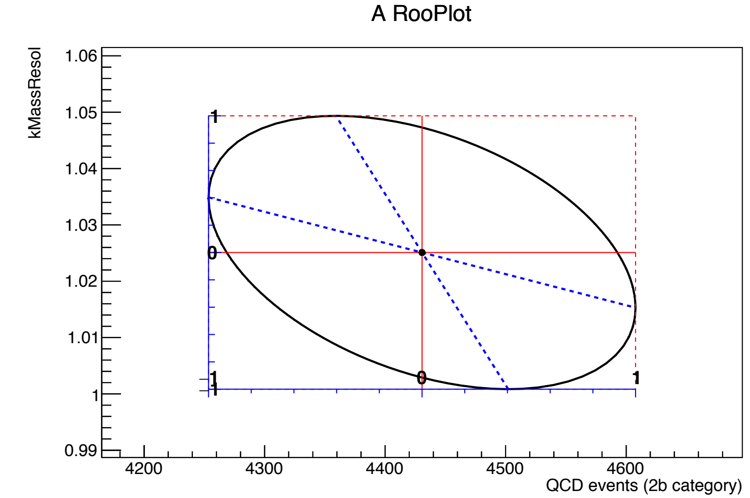


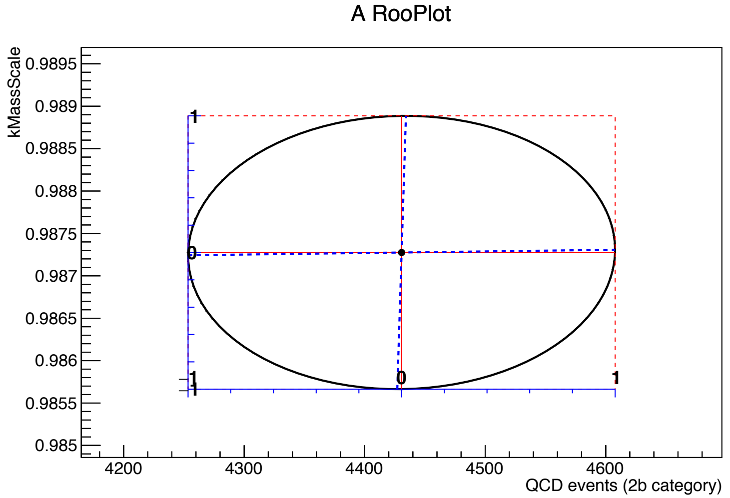


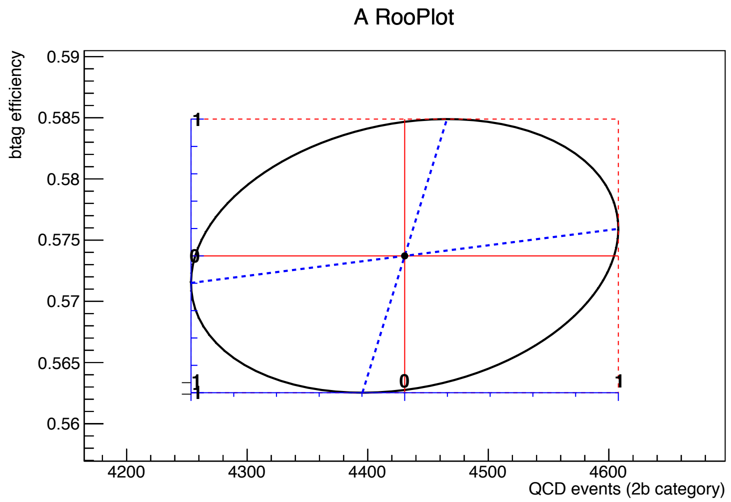


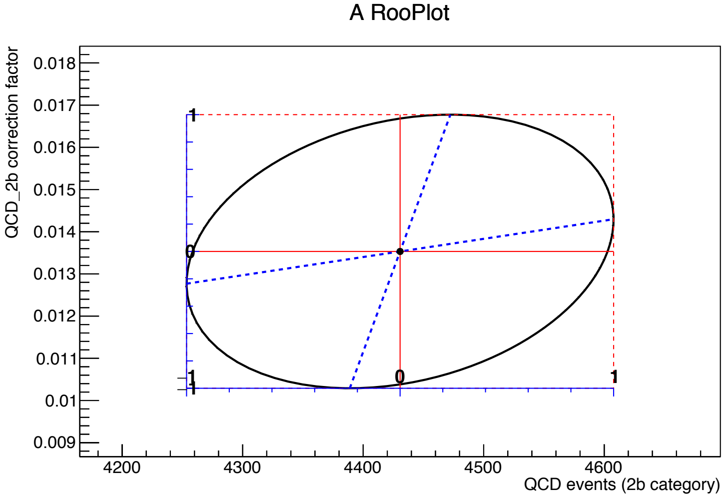


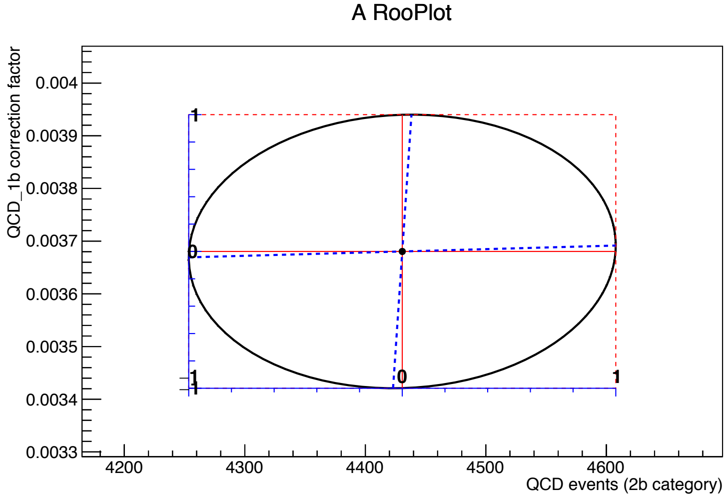
Correlation plots $N_{QCD(2)}$ vs all nuisances from fit when eb runs free for 2018

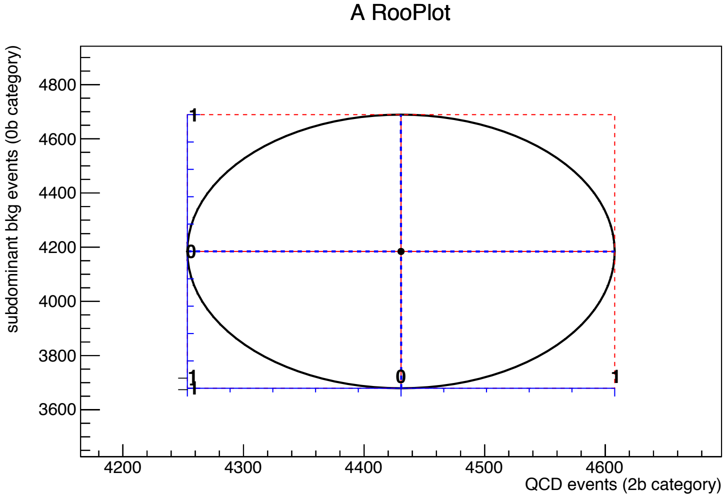












Correlation plots $N_{QCD(2)}$ vs all nuisances from fit when eb runs free for 2018