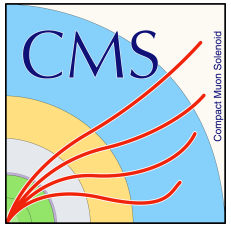
Status Report
Mass Fit and bTagging Efficiency (2016 and 2017)

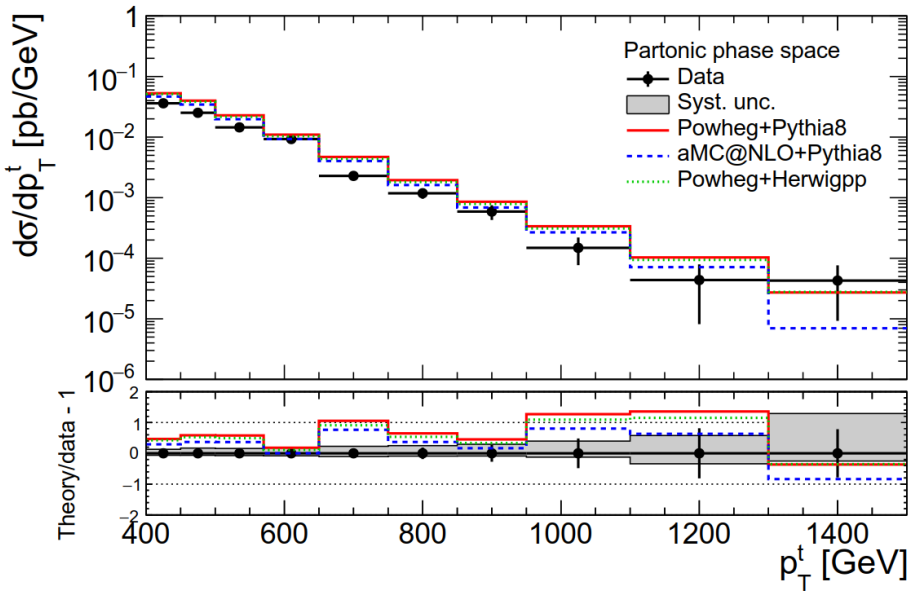
George Bakas, K.Kousouris, Ioannis Papakrivopoulos

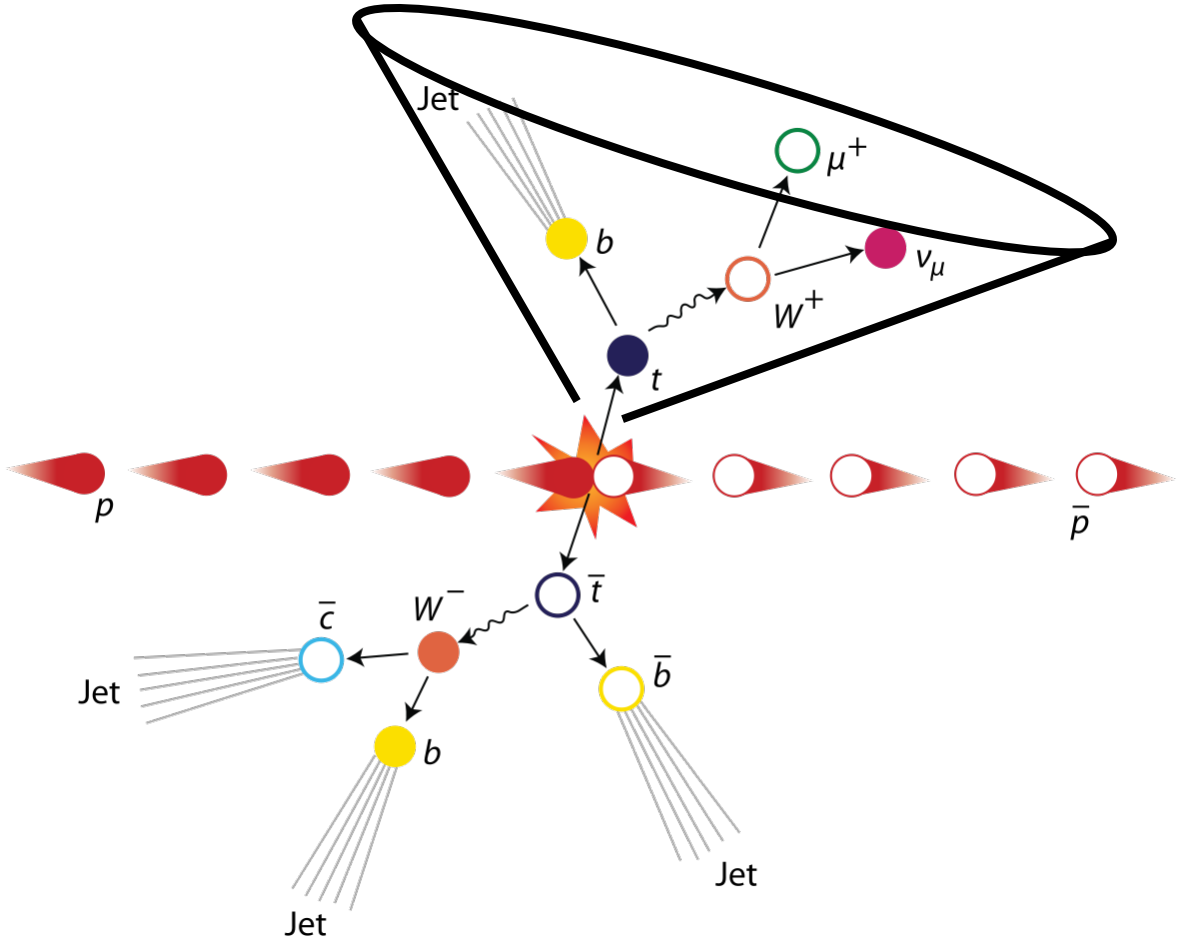


Motivation

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.





Top Angular Distributions

- We employ the dijet angular variable χ from the rapidities of the two leading jets
- Why χ?
  - 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

- We measure the variable χ using the difference of the rapidities of the two leading jets such as the corresponding rapidity in the ZMF is:

$$y^* = \frac{1}{2}(y_1 - y_2)$$

- χ is defined as $\chi = e^{|2y^*|} = e^{|y_1 - y_2|}$

- $|\cos(\theta^*)|$ distribution of the leading jet

- Selection:
  - Parton: partonPt > 400, |partonEta|< 2.4, mTTbarParton > 1000
  - Reco: jetPt>400, |jetEta| < 2.4, nLeptons = 0
  - Btagging Medium working point
  - Top tagger mva > 0.1
  - Jet mass soft Drop (120, 220)GeV

Overview of Signal Region, Control Region and SR_A region

- Signal Region (SR): Baseline Selection + topTagger Selection + mass Selection + 2btagged subjets
- Control Region (CR): Baseline Selection + topTagger Selection + mass Selection + 0btagged subjets

- Extension of Signal Region → SR_A = SR – Mass Selection cuts
- Selection:
  - Jet Matching

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

- 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 |η|<2.4
  - Trigger

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

- Top Tagger:
  - Tagger cut (**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.

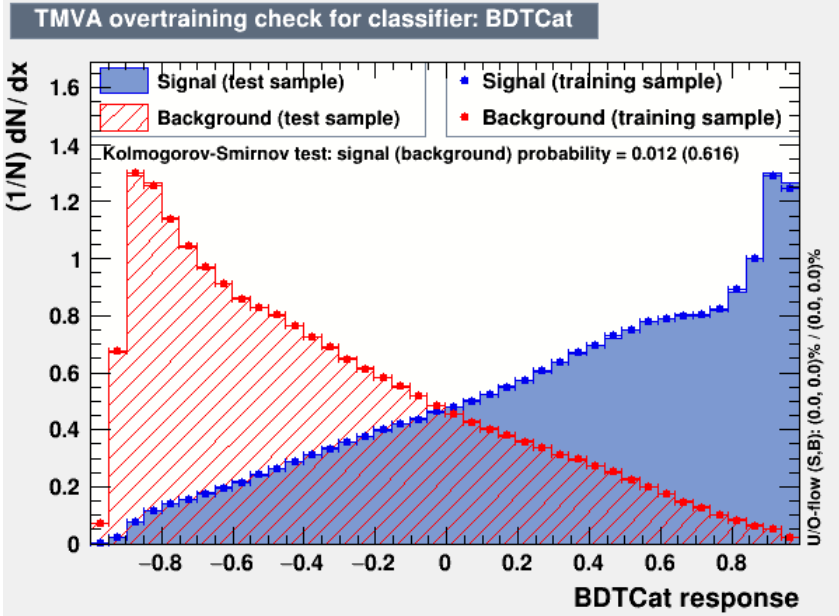
- Bins:
  - [400, 600] GeV
  - [600, 800] GeV
  - [800, 1200] GeV
  - [1200, inf) GeV
- BDT, used variables:
  - Leading and Sub-leading subjet mass
  - N-Subjetiness variables (tau1, tau2, tau3)
  - fraction of the jetPt over the total pt sum of the event.
  - Energy correlation functions (ecfB1N2,ecfB1N3, 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

$$Efficiency = \frac{\#events\ passing\ reco\ and\ parton\ selection}{\#events\ passing\ parton\ sel.\ from\ EventCounter}\ (vs\ Parton)$$

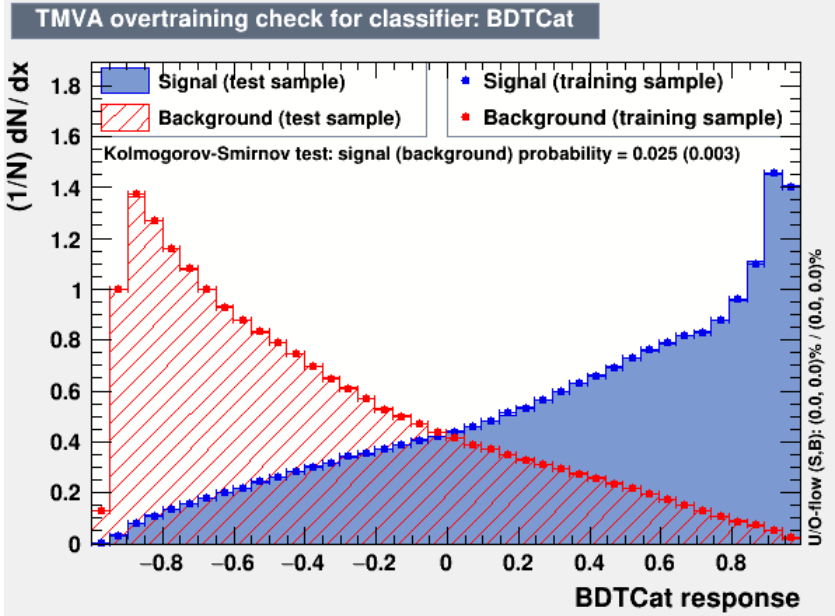
$$Acceptance = \frac{\#events\ passing\ reco\ and\ parton\ selection}{\#eventsing\ pass\ reco\ selection}\ (vs\ Reco)$$

Training Outputs

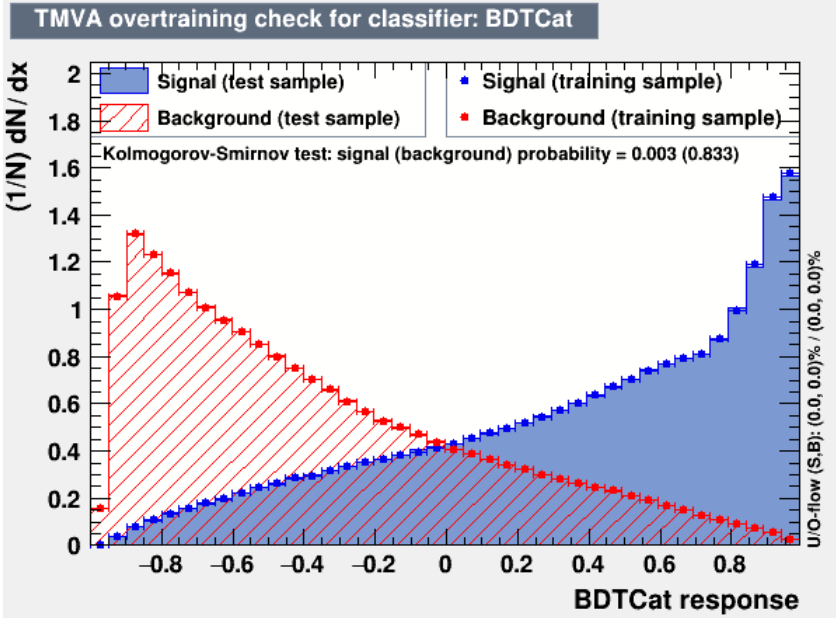
2016



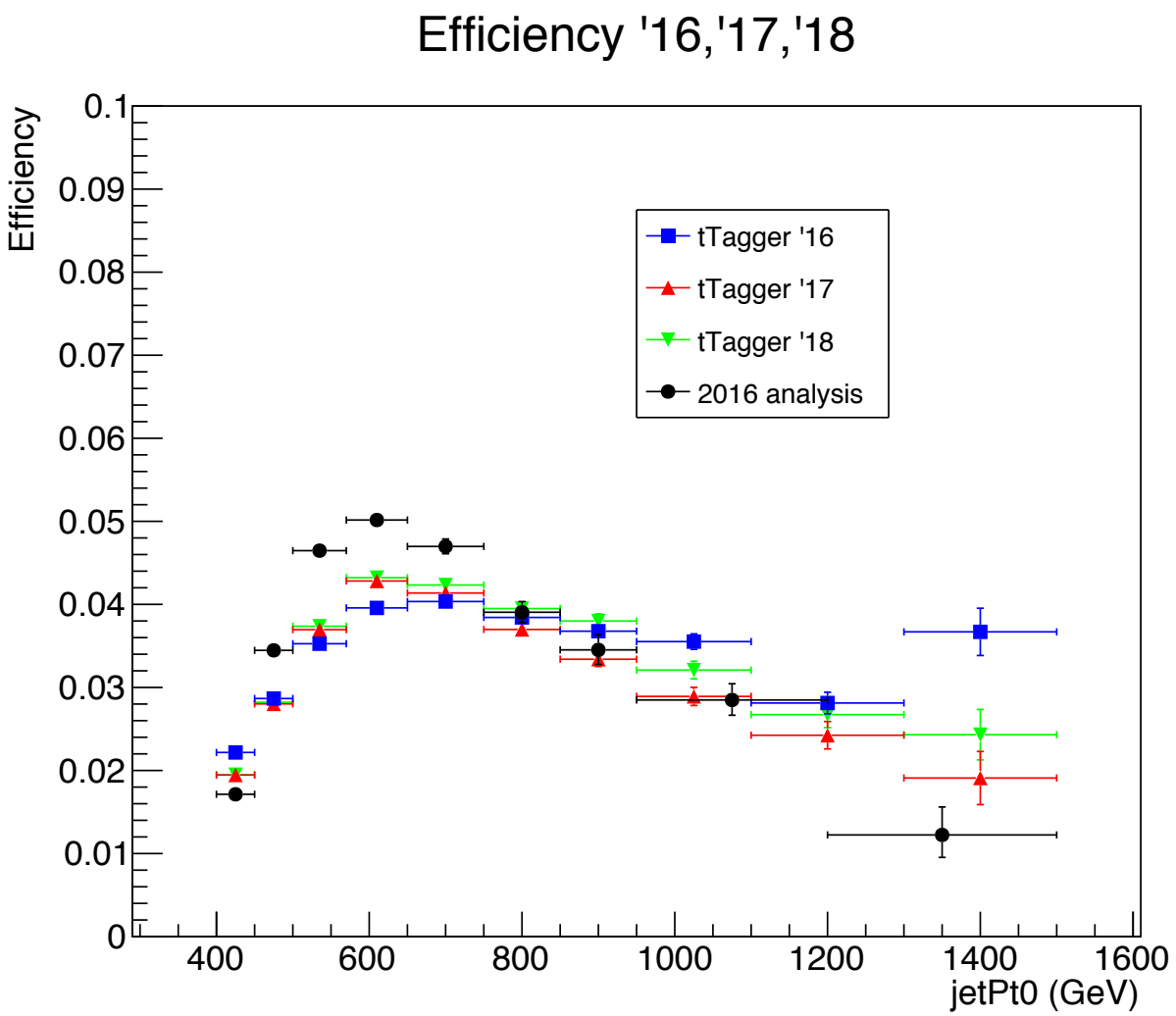
2017

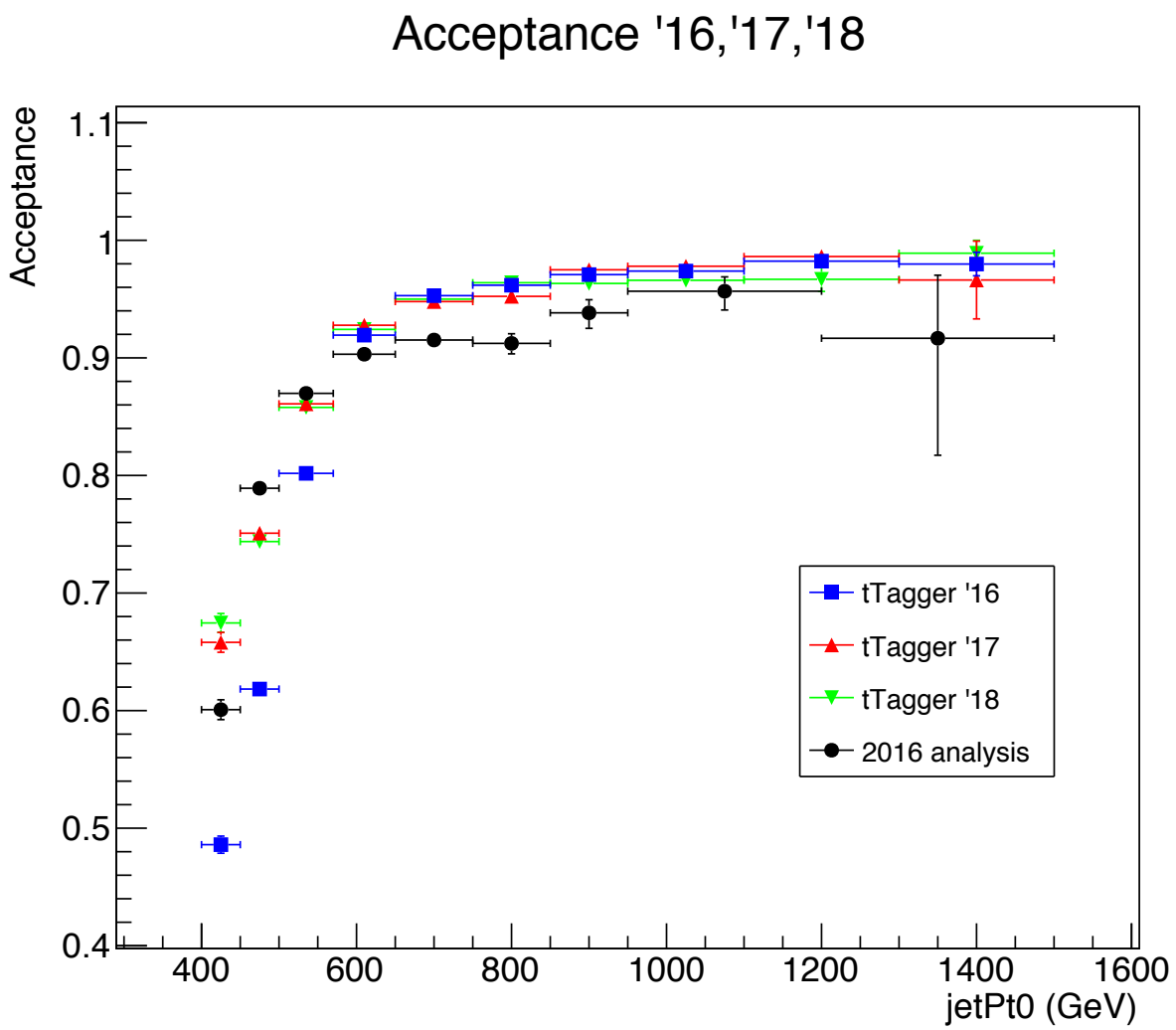


2018



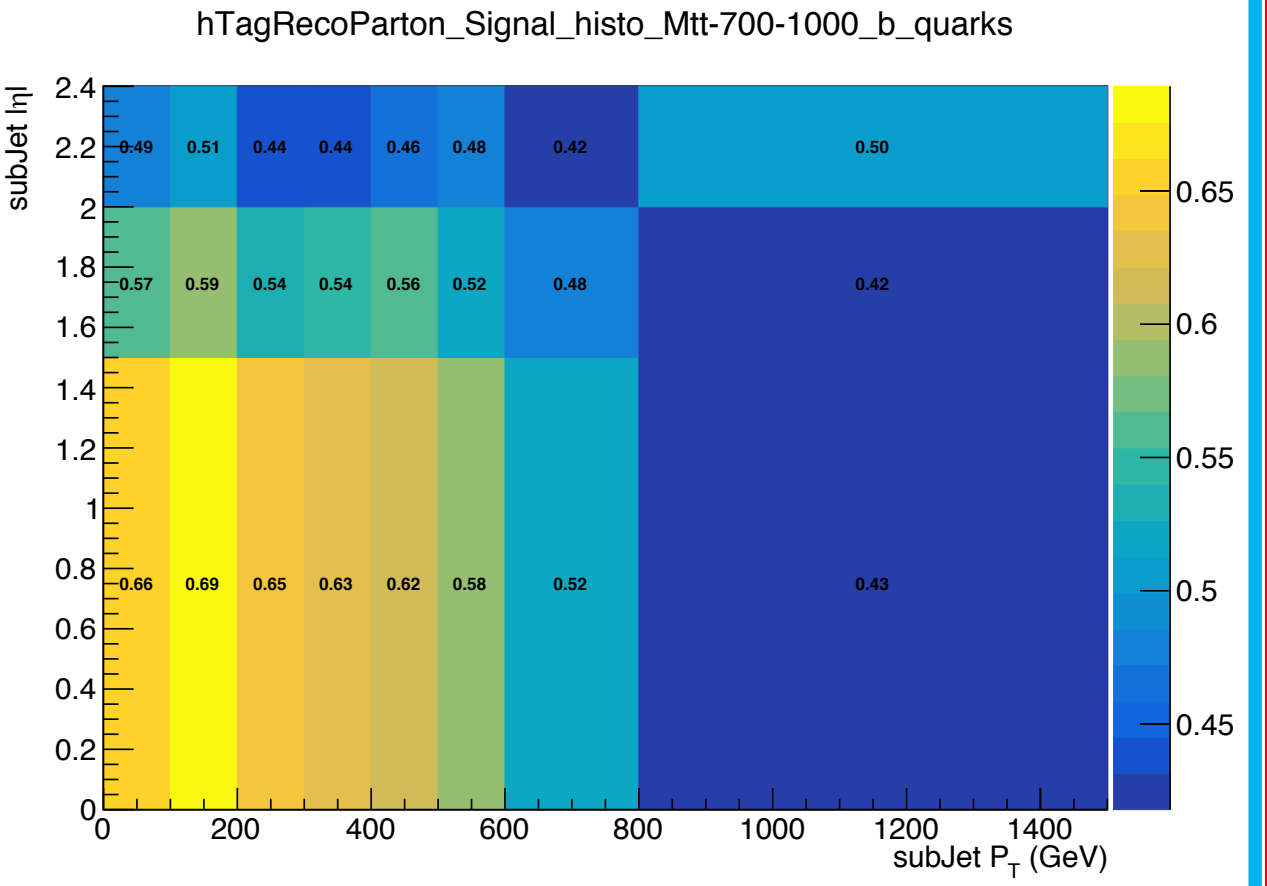
Efficiency and Acceptance for 2016, 2017 and 2018 and previous 2016 analysis



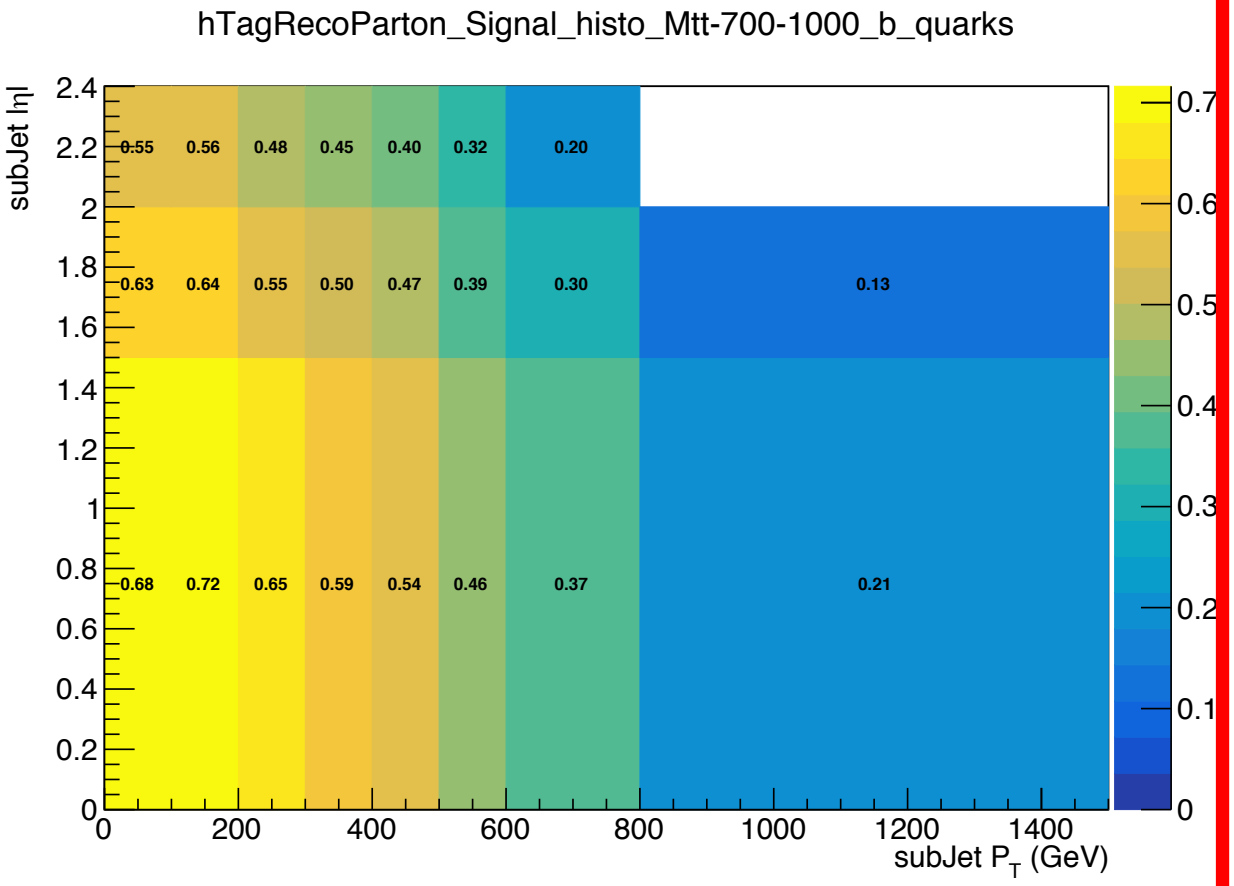


Btagging efficiency in eta, pT_{subJet} phase space

2016

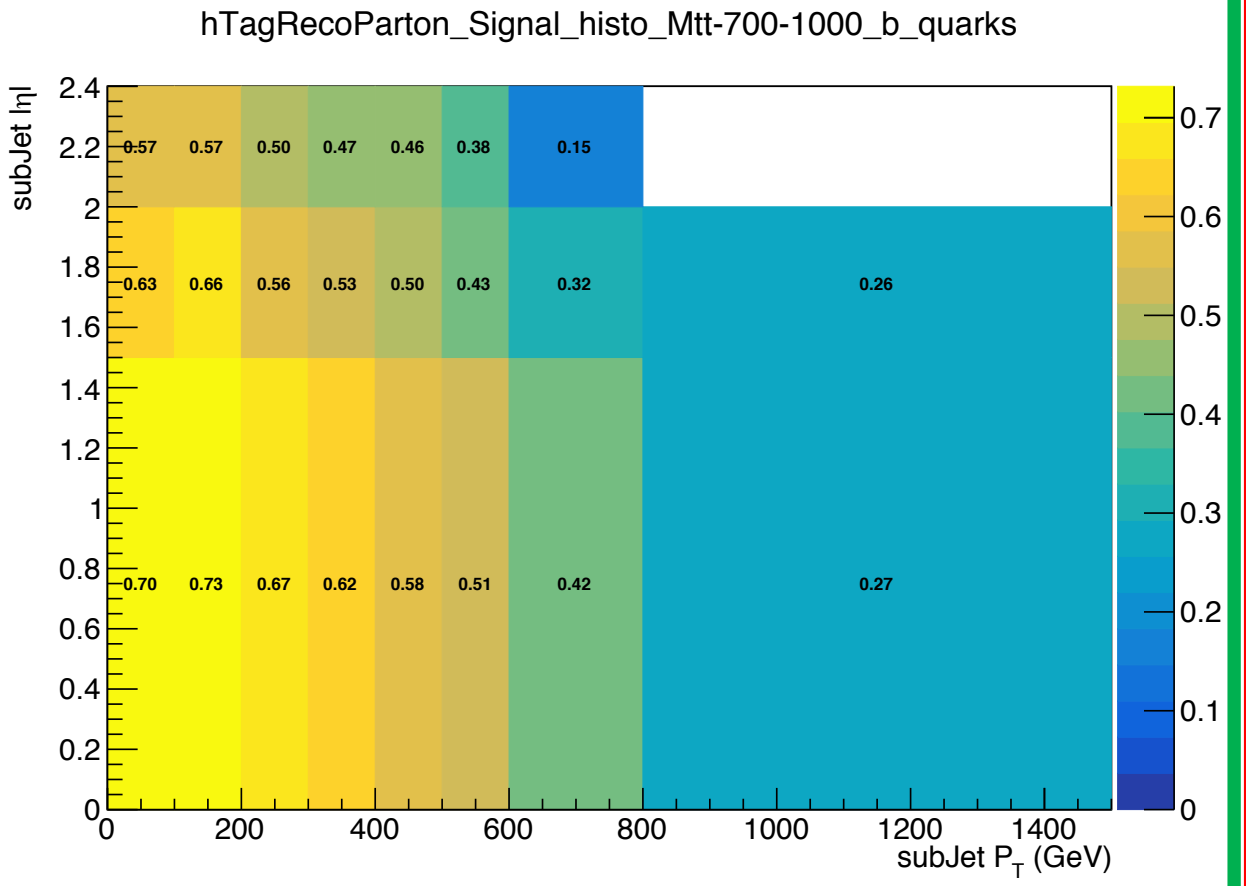


2017

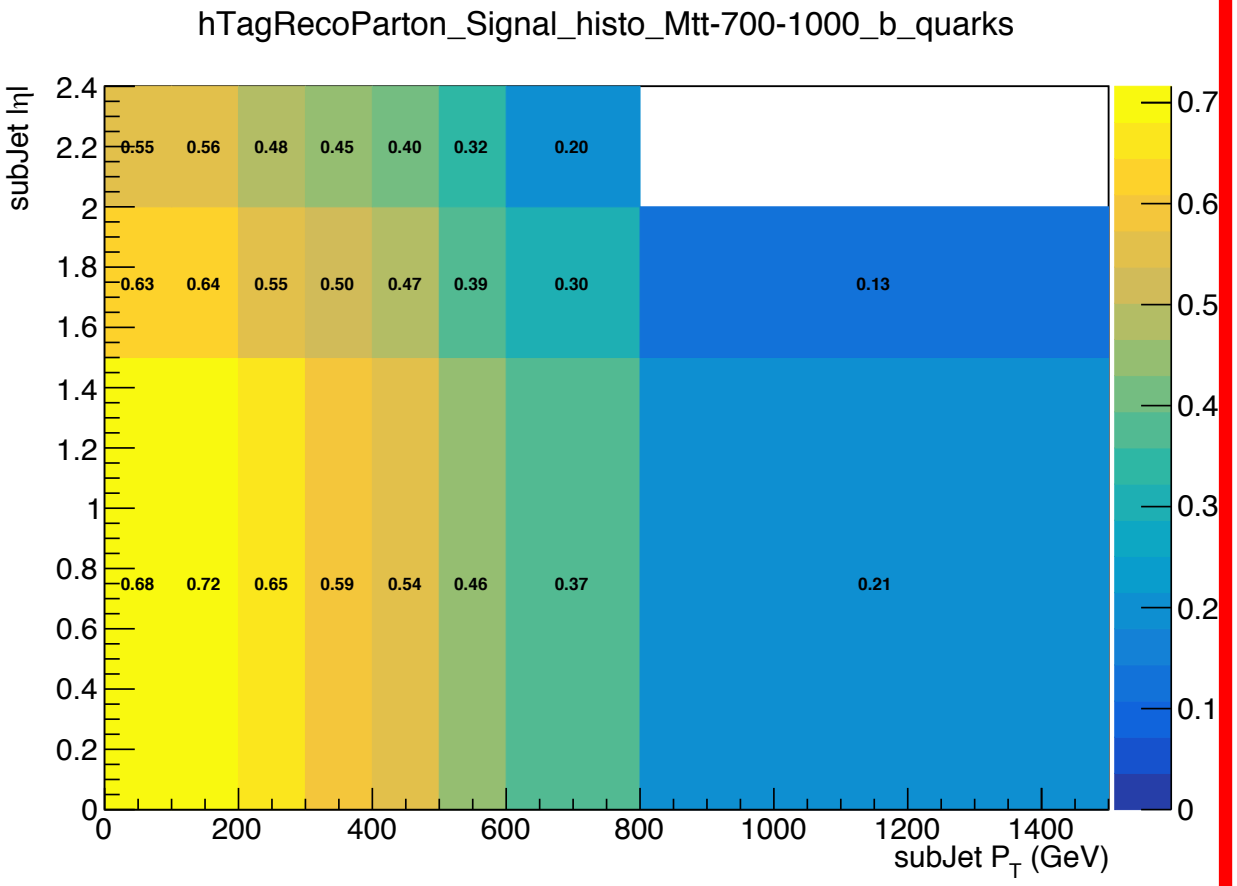


Btagging efficiency in eta, pT_{subJet} phase space

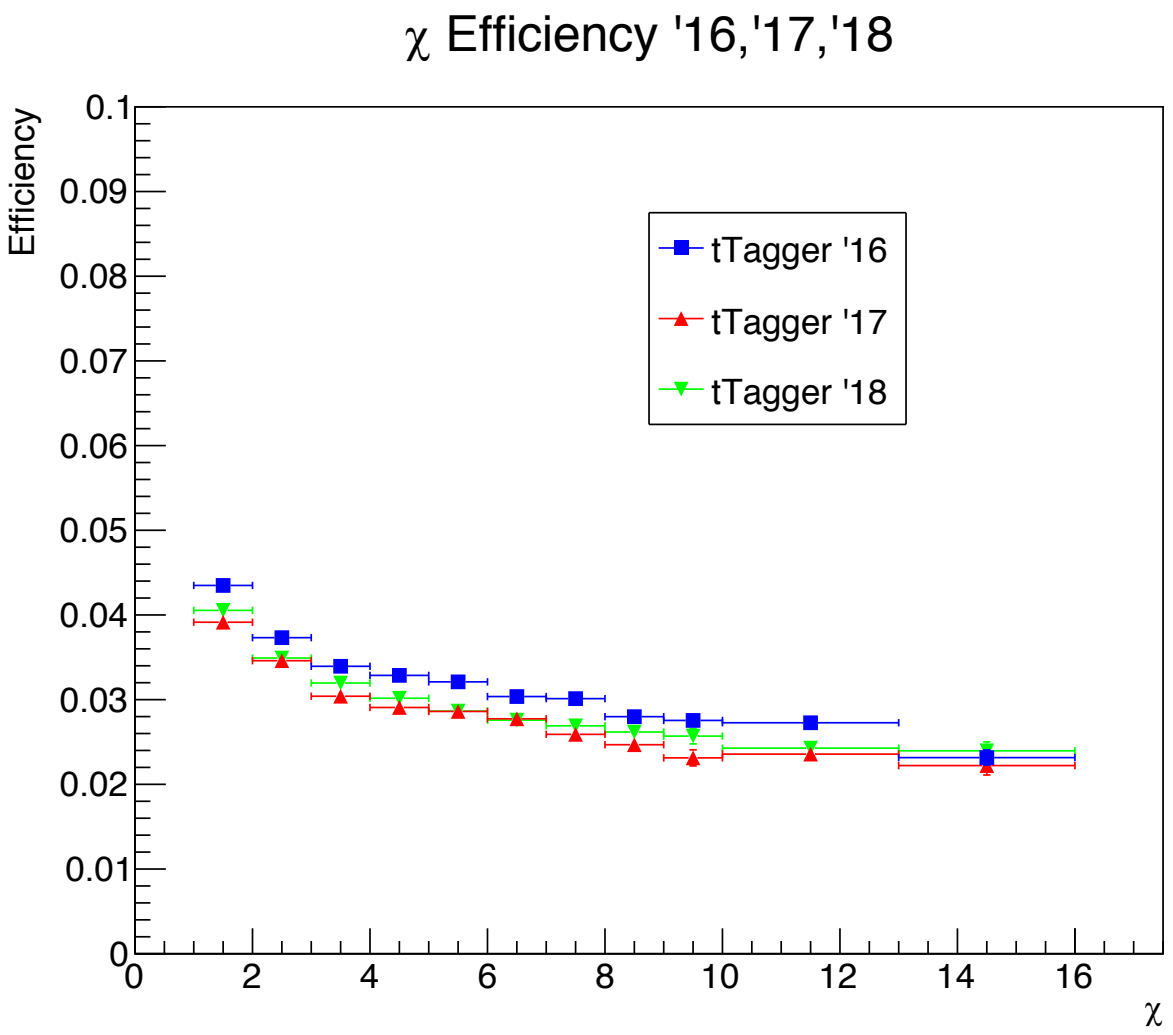
2018

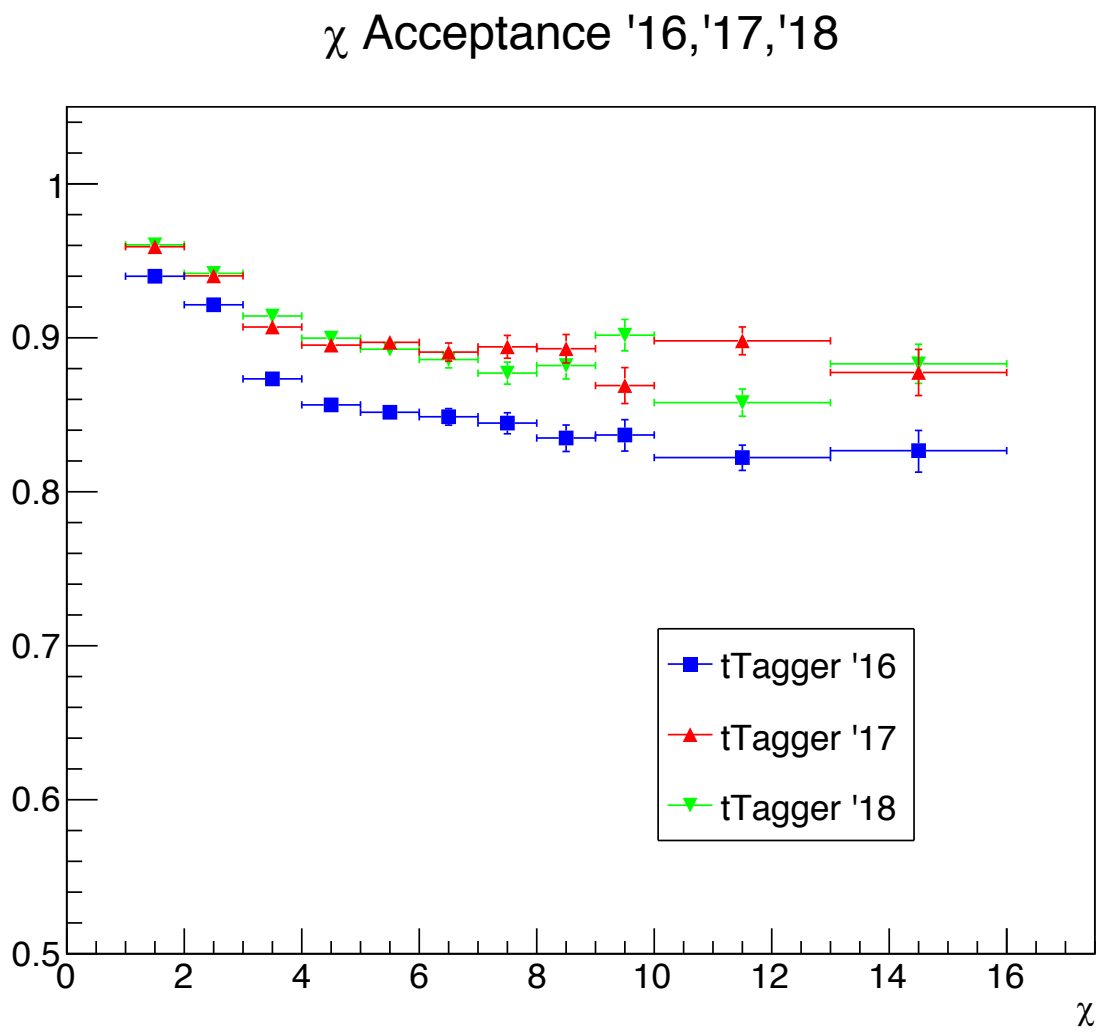


2016

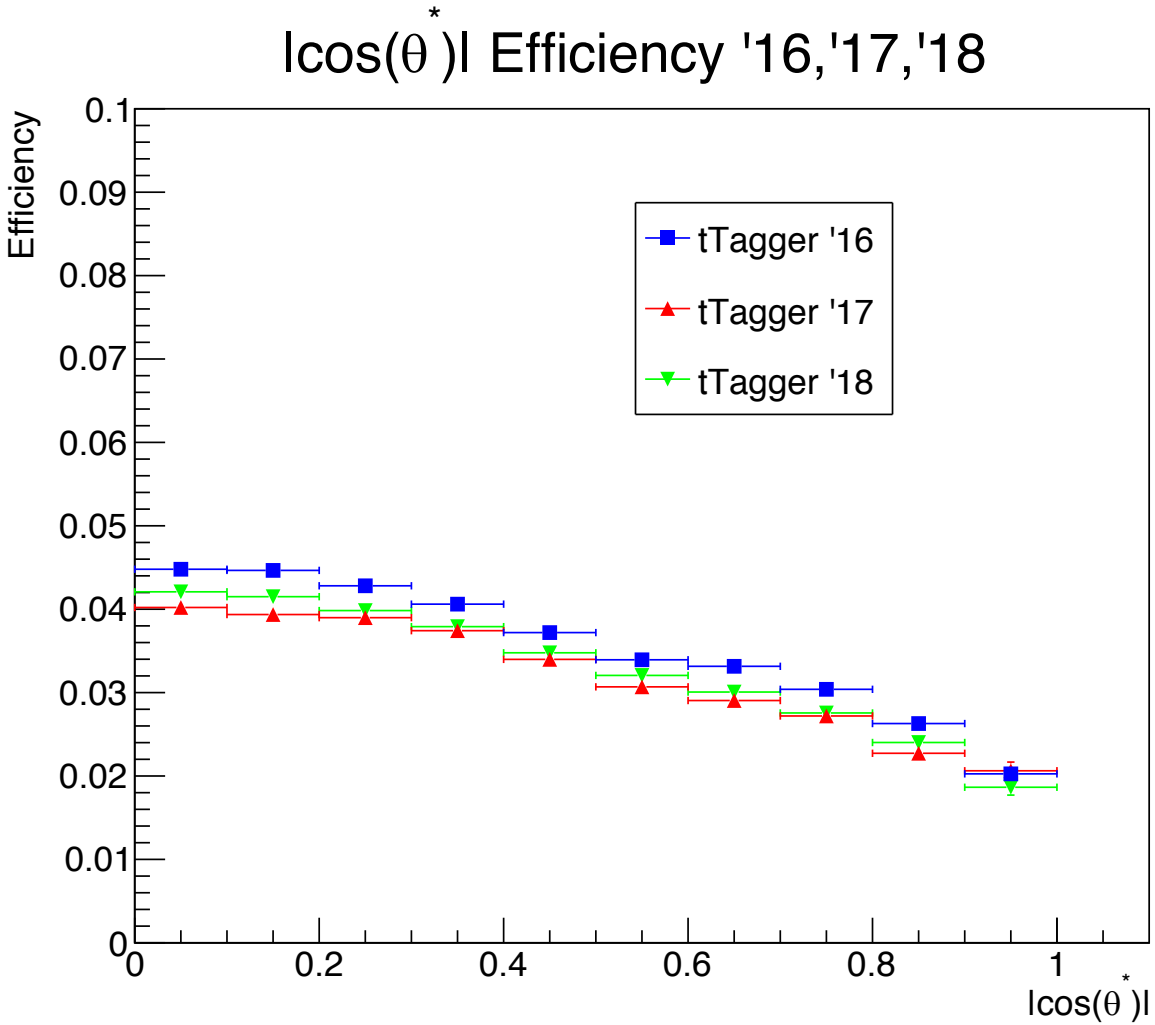


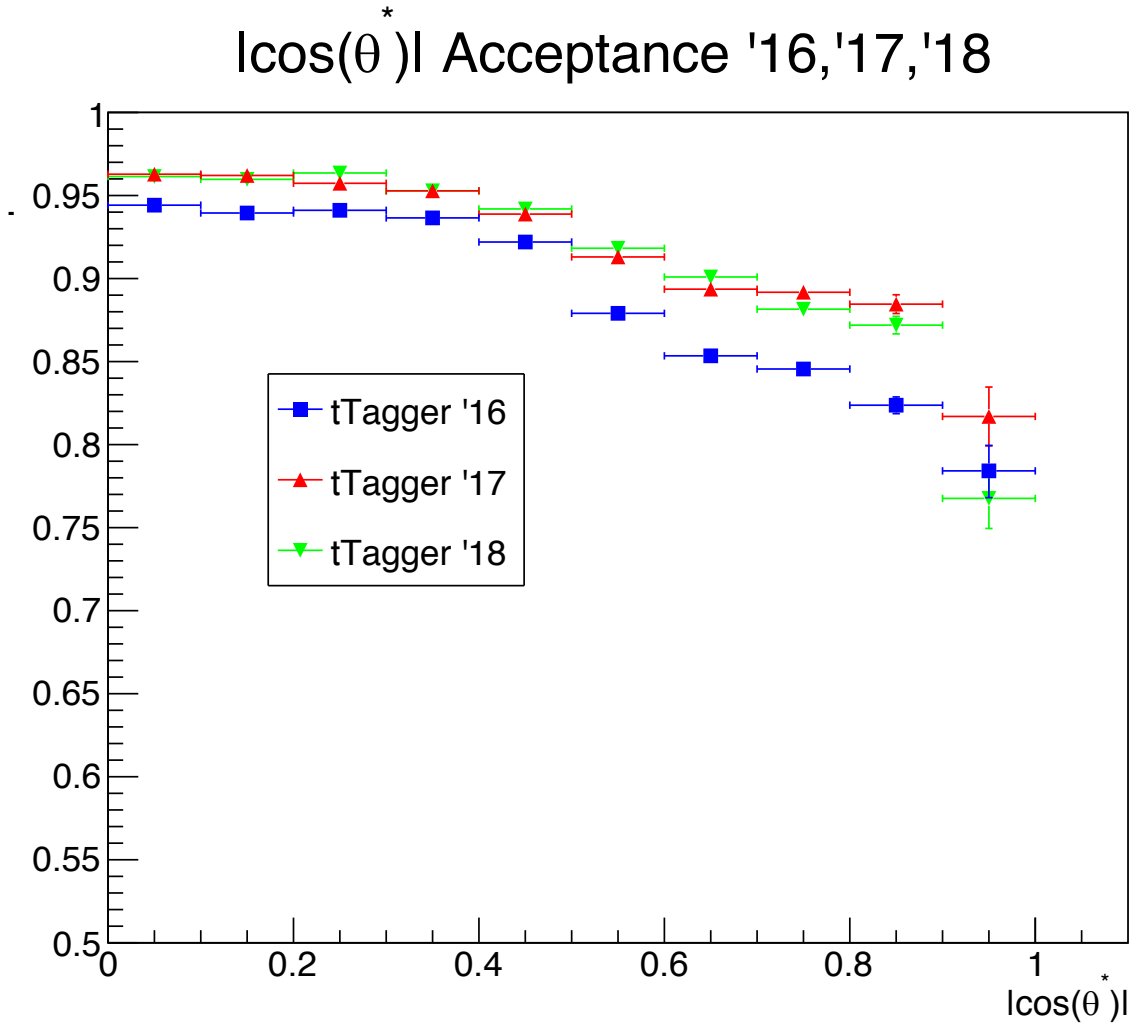
Efficiency, Acceptance for χ





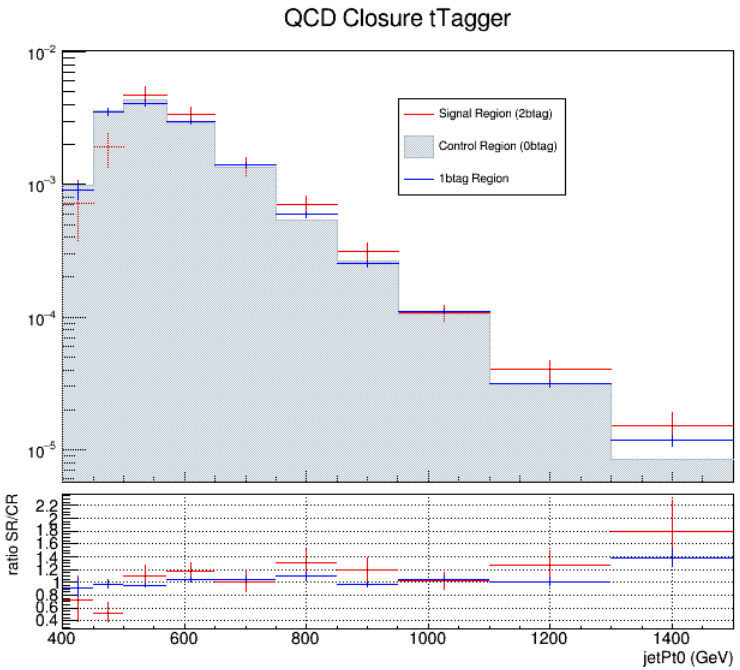
Efficiency, Acceptance for |cos(θ*)|



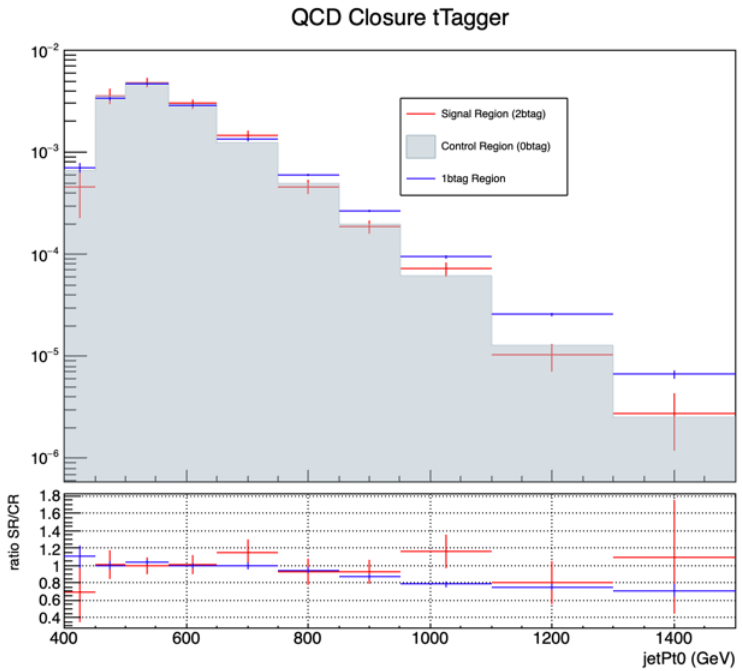


QCD Closure Tests ’16, ‘17, ’18 jetPt0

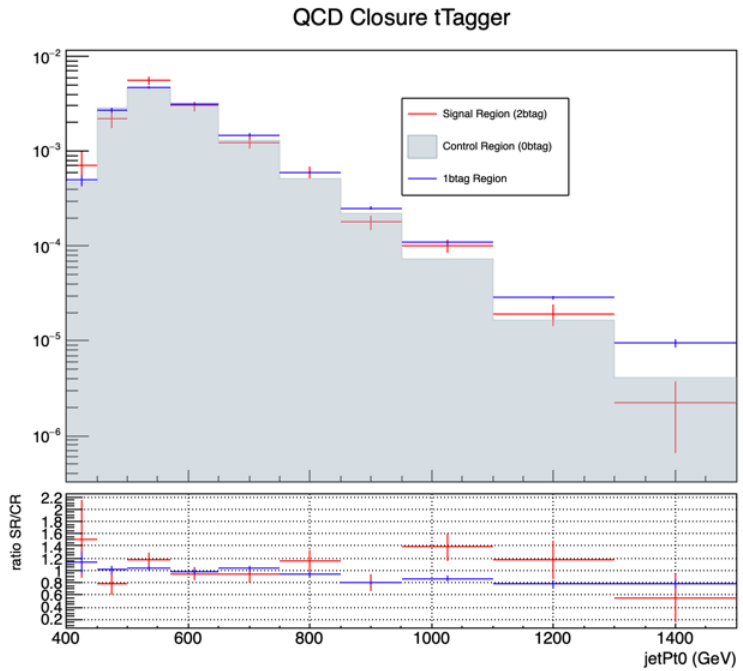
2016



2017



2018

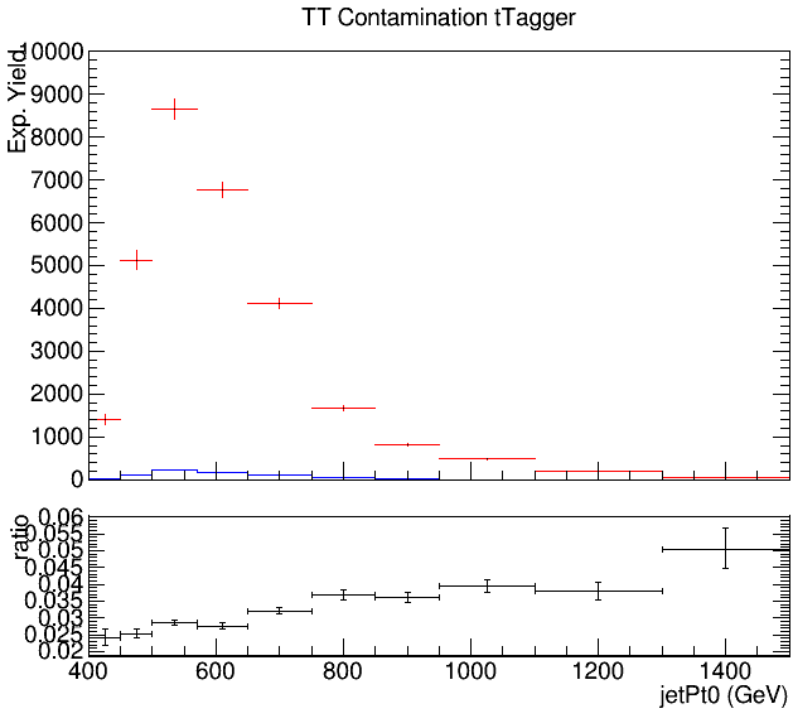


CR Contamination ’16,’17,’18 jetPt0

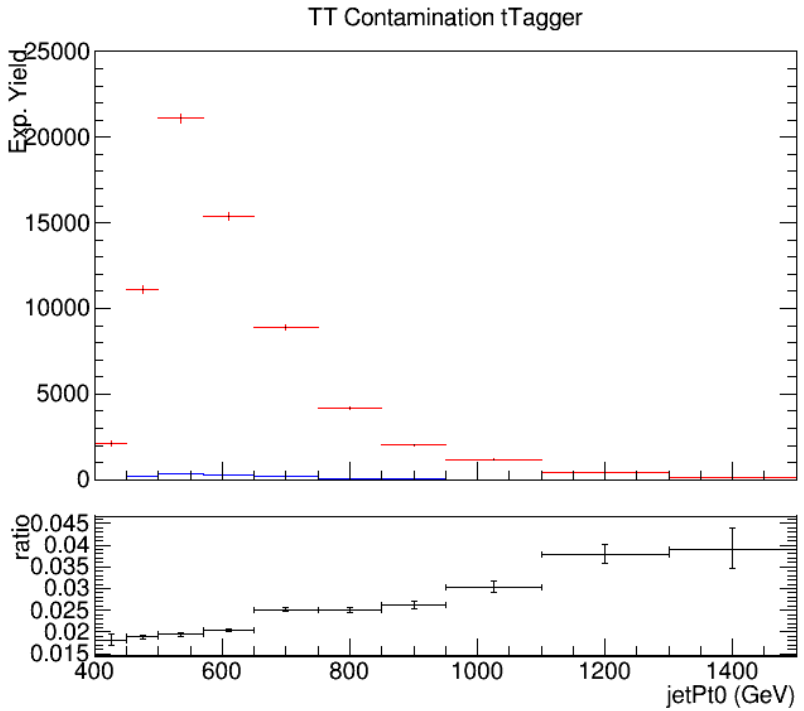
Control Region QCD sample

Control Region TT sample

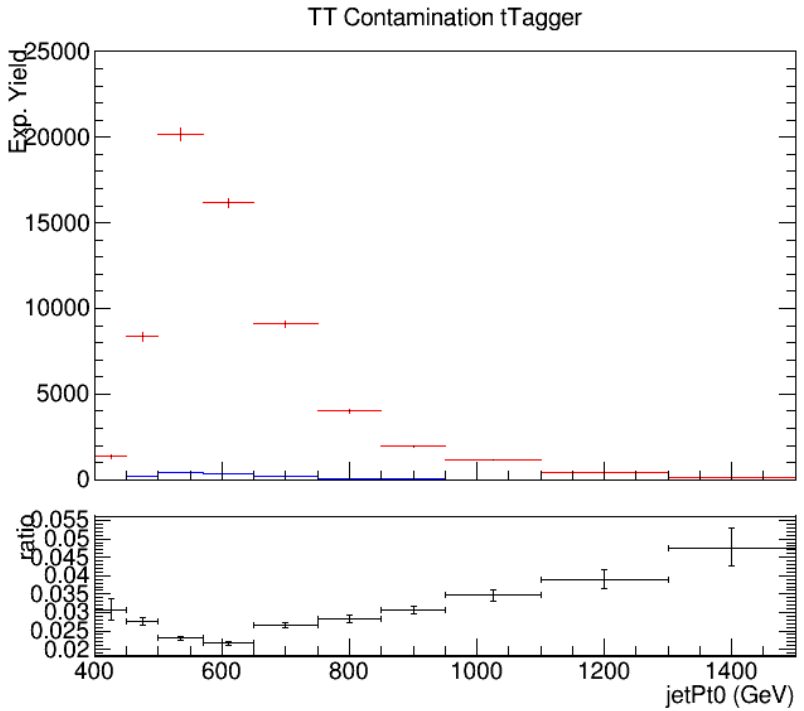
2016



2017

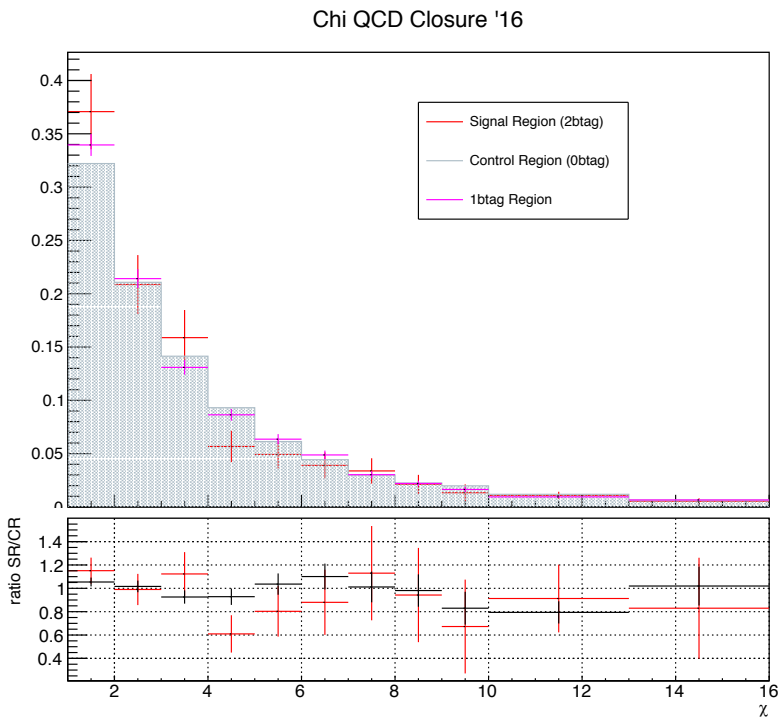


2018

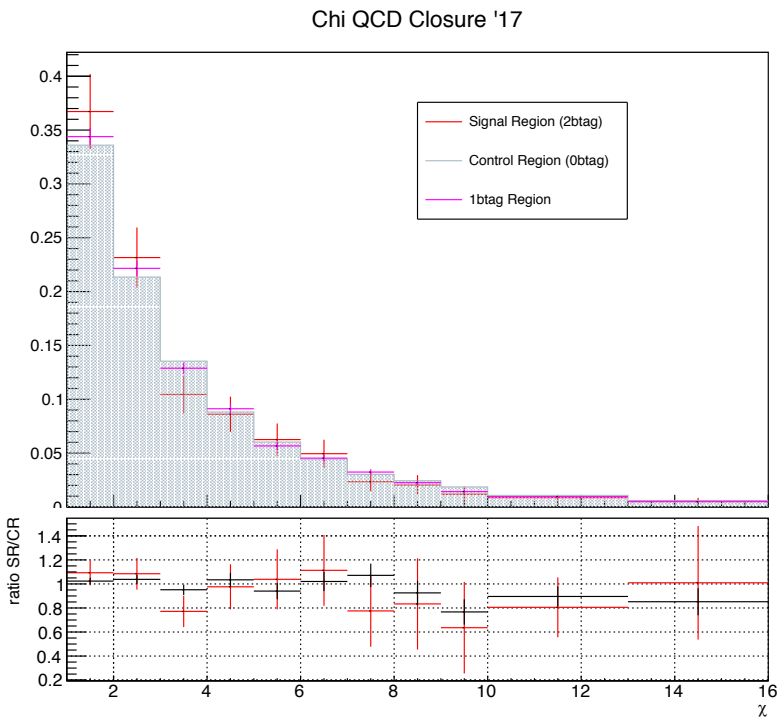


QCD Closure Tests ’16, ‘17, ’18 χ

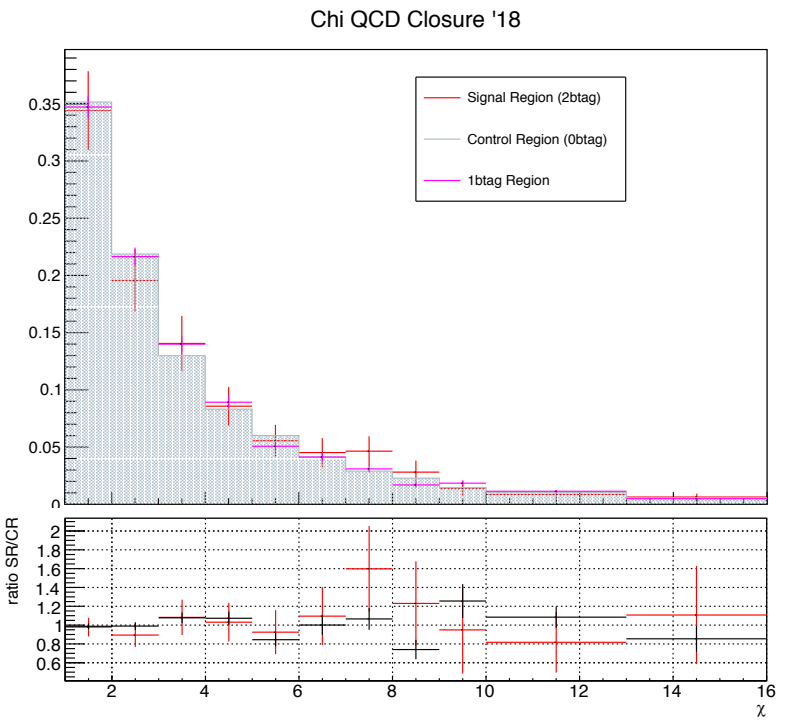
2016



2017



2018

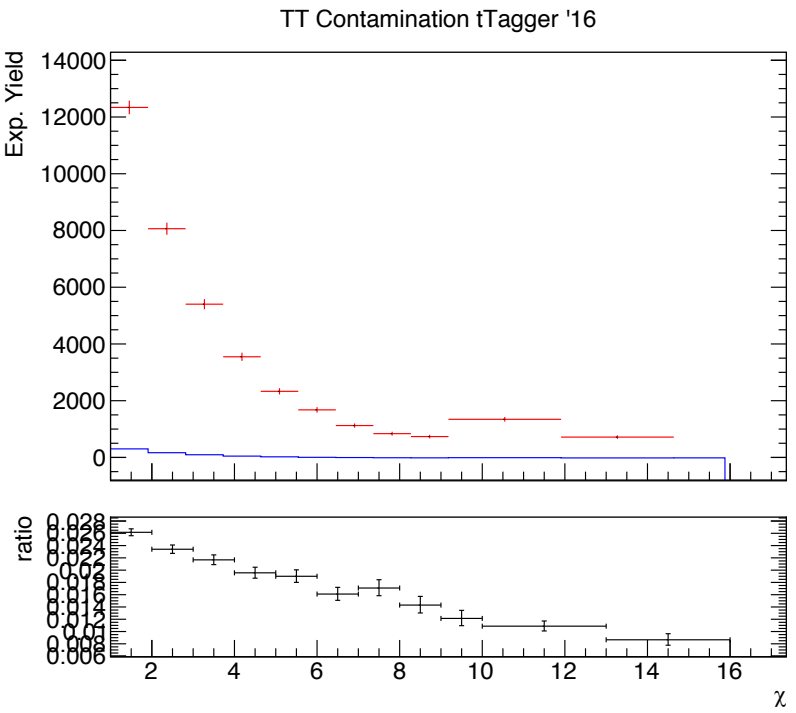


CR Contamination ’16,’17,’18 χ

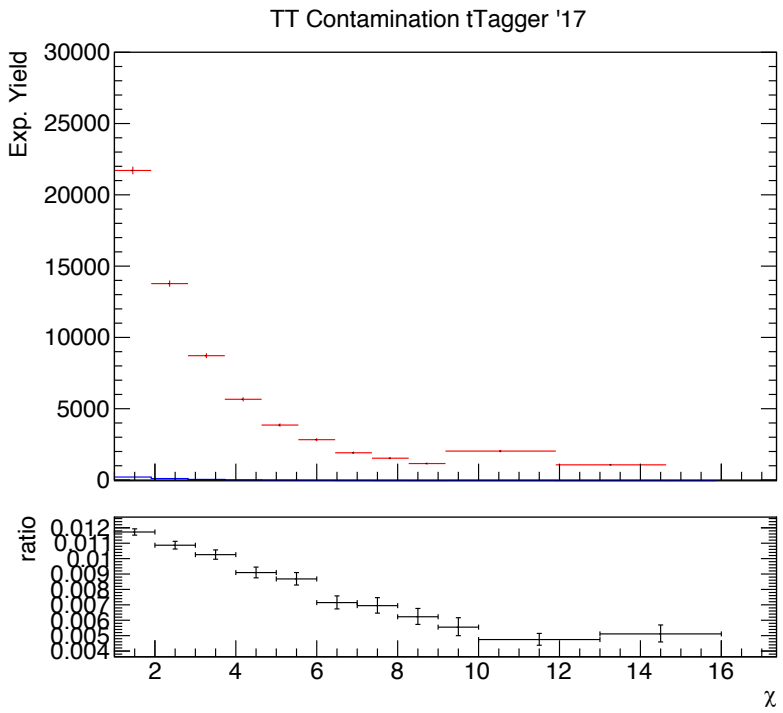
Control Region QCD sample

Control Region TT sample

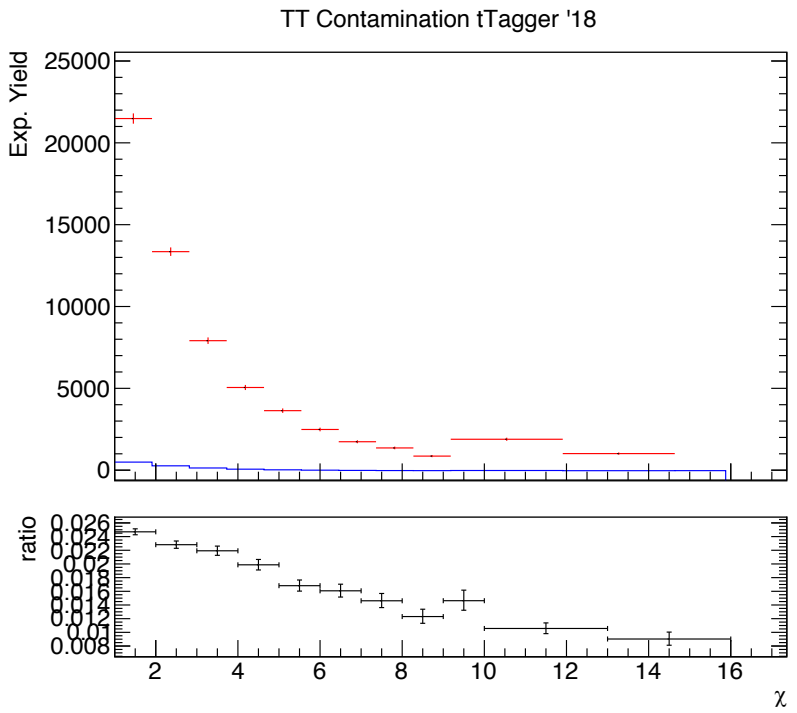
2016



2017

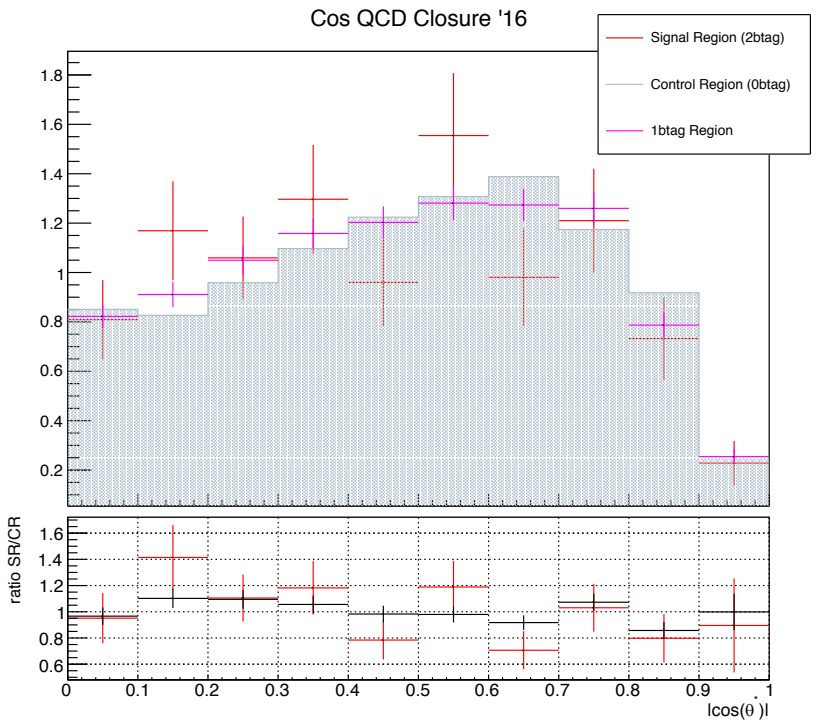


2018

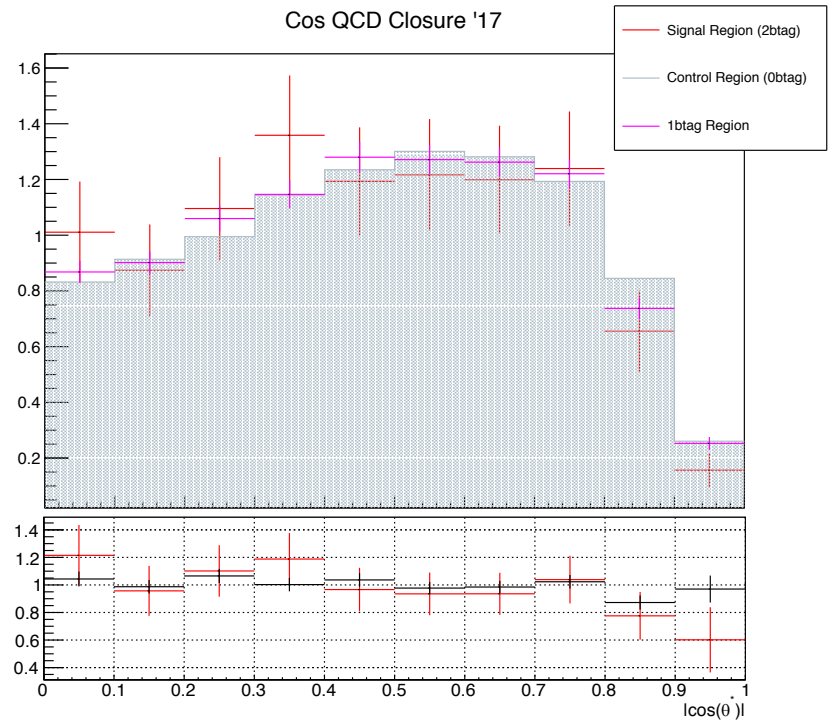


QCD Closure Tests ’16, ‘17, ’18 |cos(θ*)|

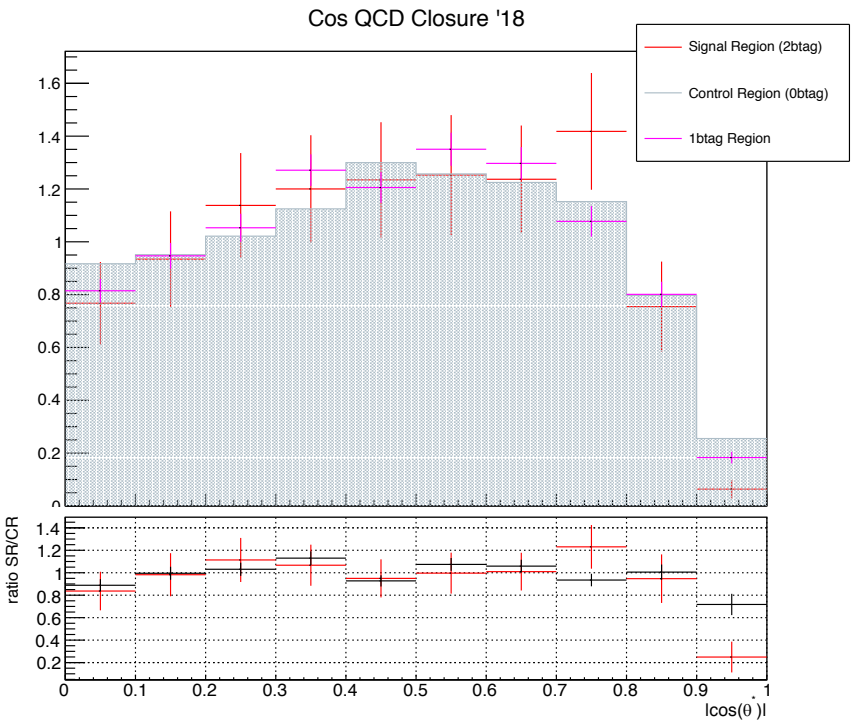
2016



2017



2018

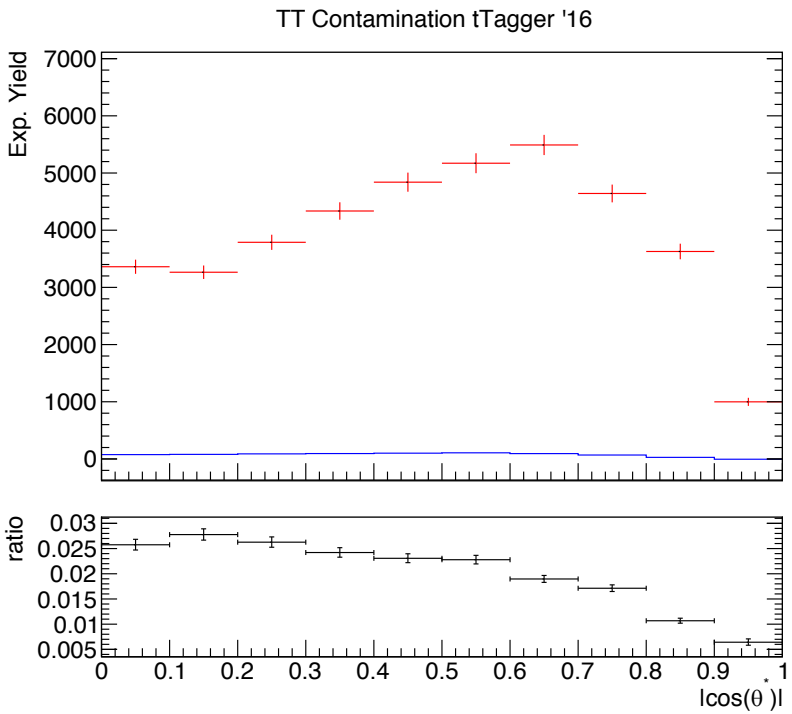


CR Contamination ’16,’17,’18 |cos(θ*)|

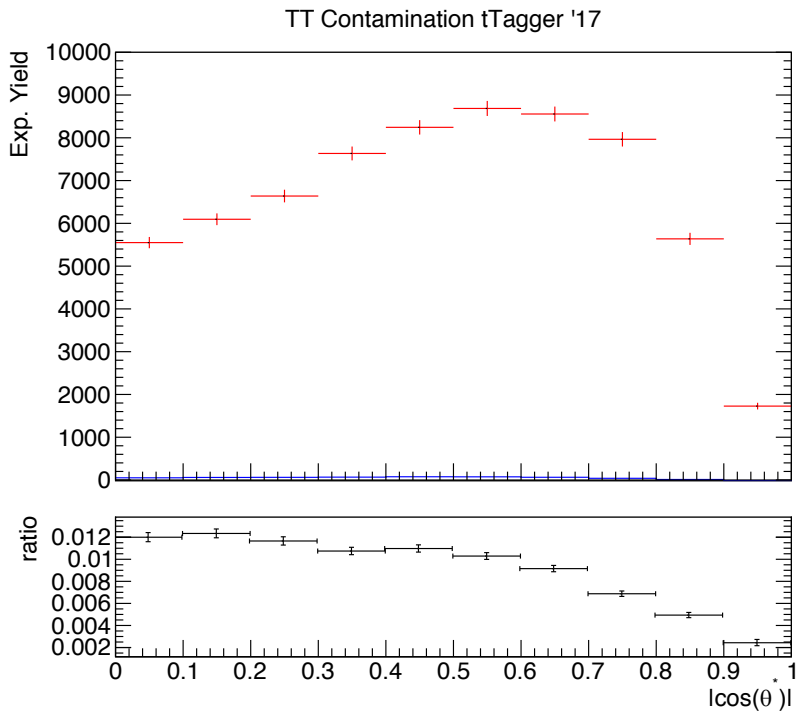
Control Region QCD sample

Control Region TT sample

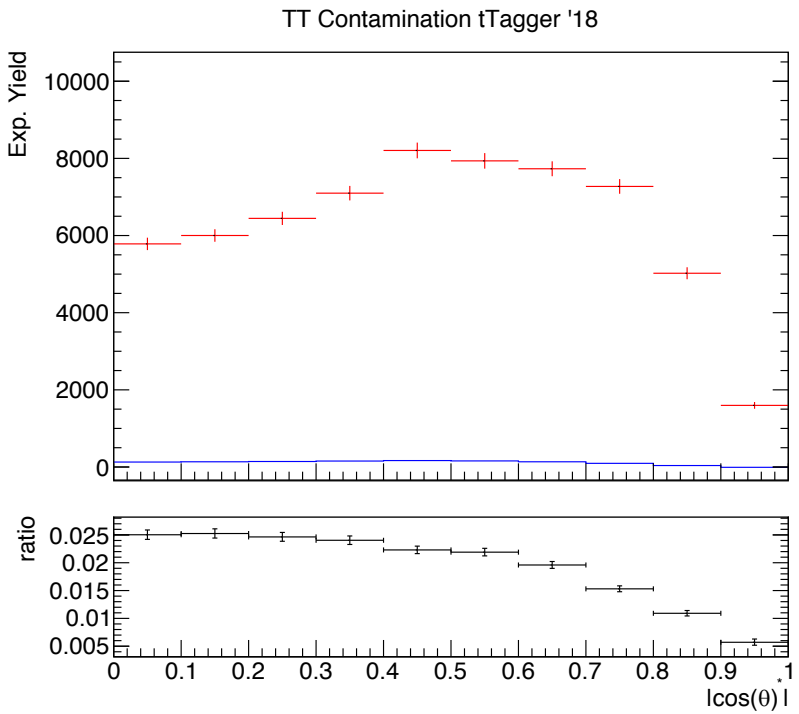
2016



2017



2018



Simultaneous Fit in 3 regions

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

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

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

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

  - 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 $\mathrm{e_b}$ is the b tagging efficiency and $\mathrm{N_{tt}}$ is the total ttbar yield.

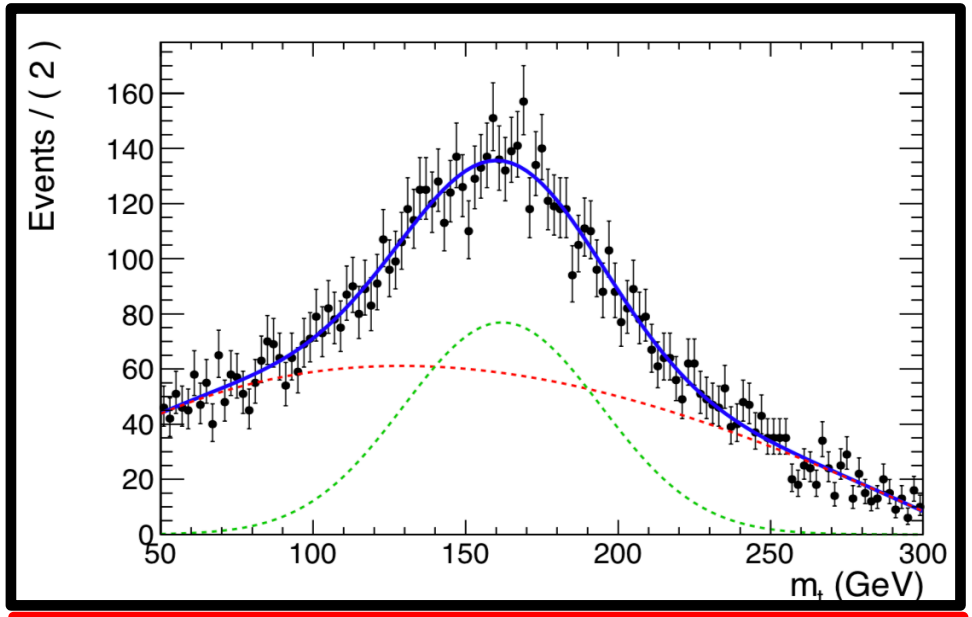
  We can either leave $\mathrm{e_b}$ and $\mathrm{N_{tt}}$ as free parameters in the fit or $N_{tt}^{(0)}, N_{tt}^{(1)}, N_{tt}^{(2)}$

- 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:

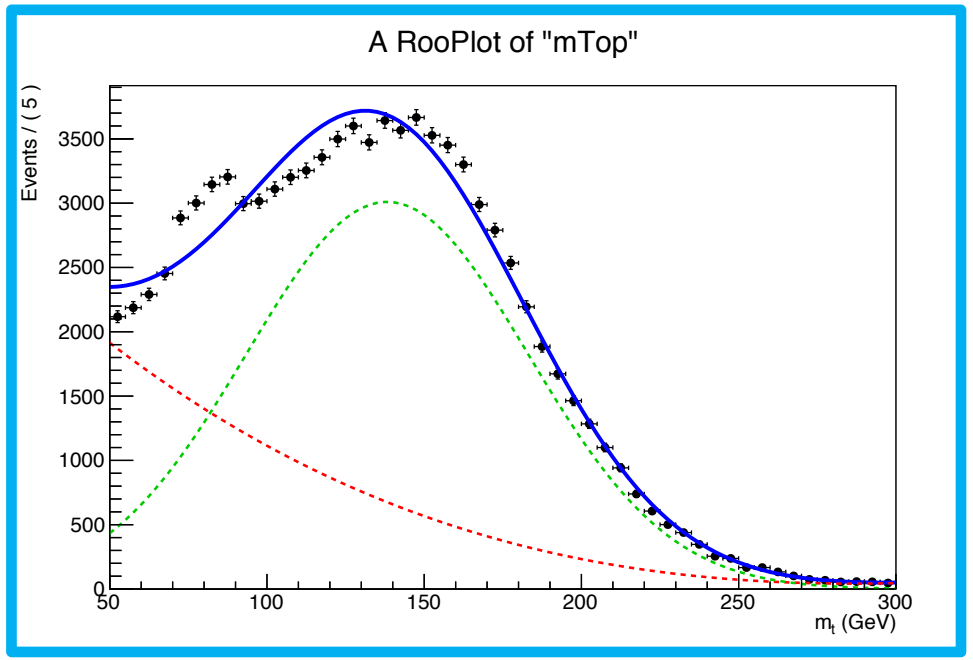
$e_b = \frac{\#subjets\ with\ flavour\ id\ requirement + deepCSV\ btagged}{\#subjets\ with\ flavour\ id\ requirement\ (b)}$, where all selected events pass baseline + parton selection

Template Fit results in CR

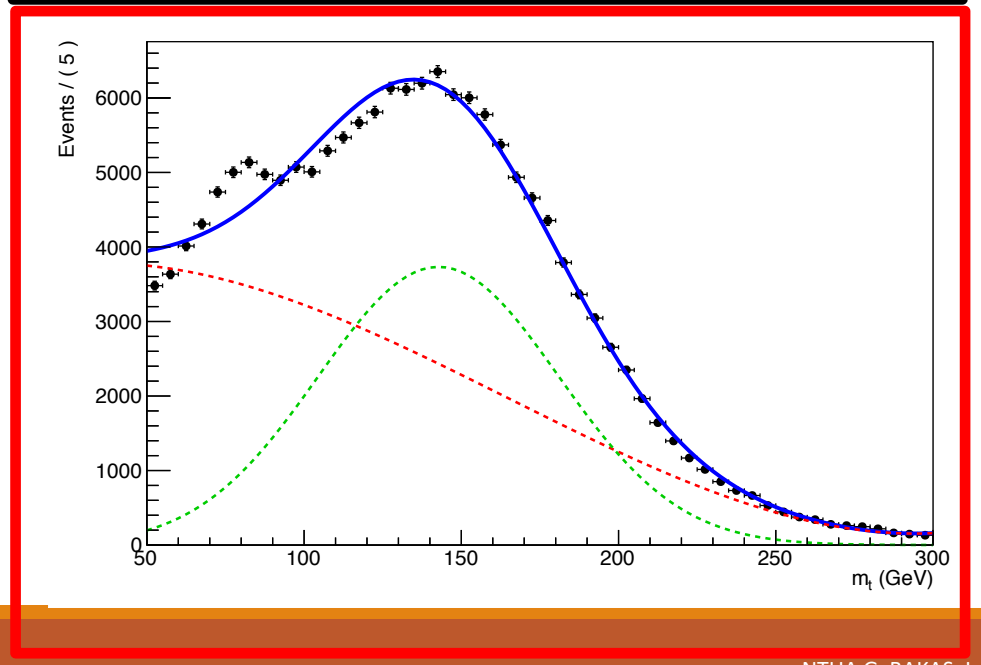
2016



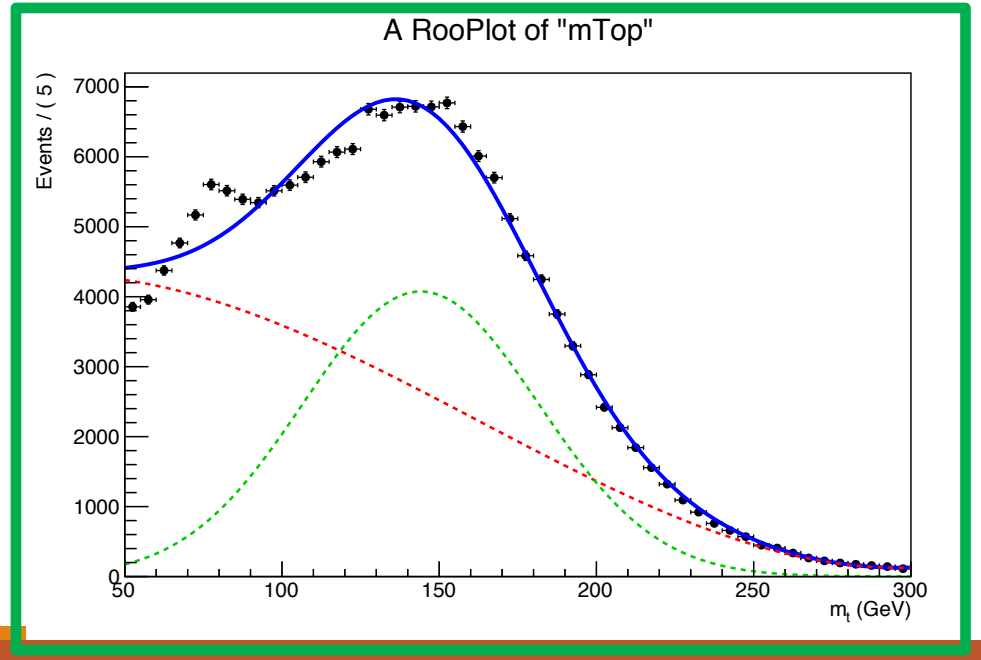
2016



2017

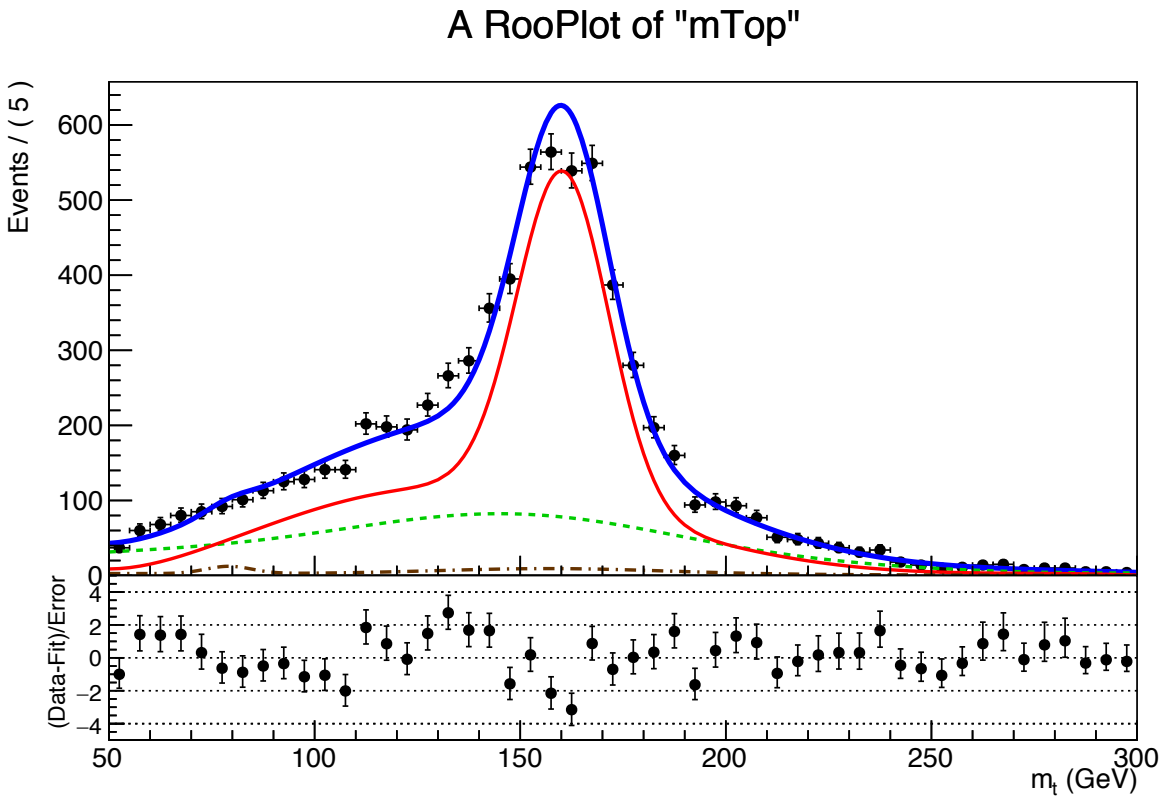


2018

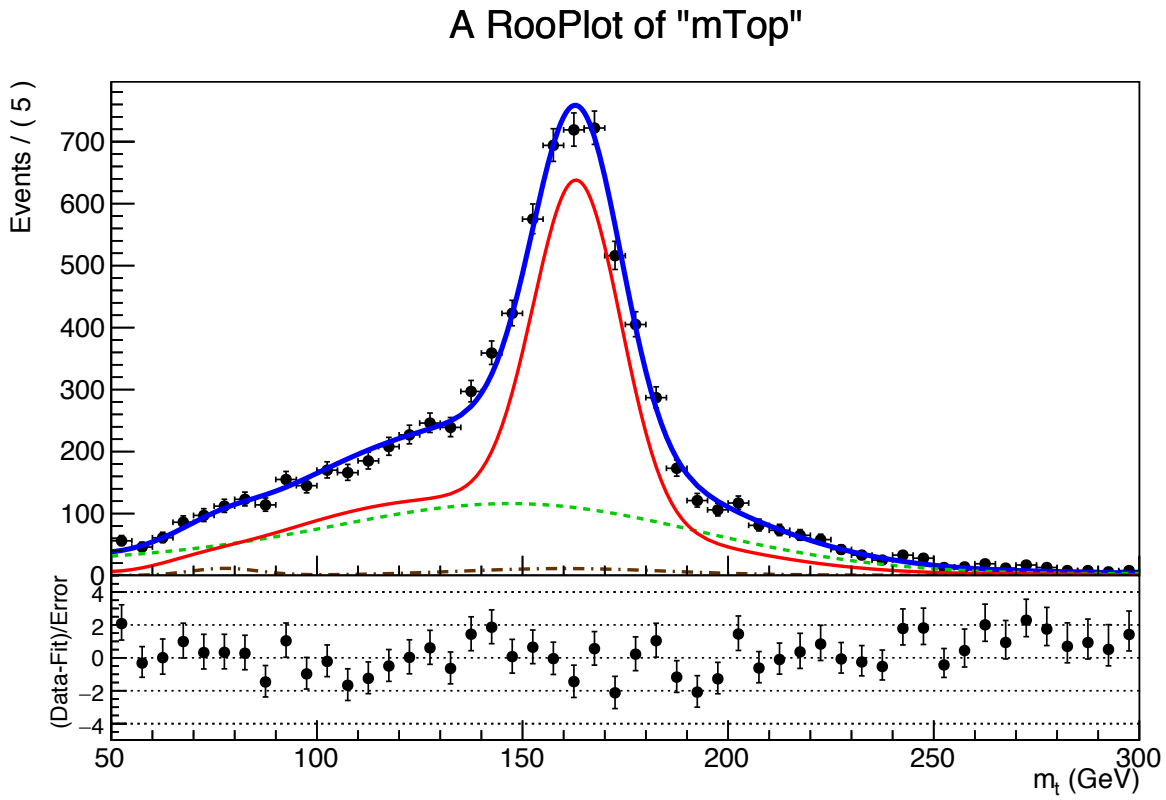


Simultaneous Fit in 3 regions for 2017 and 2016 when eb is fixed

Signal Region (2btag) (2017)



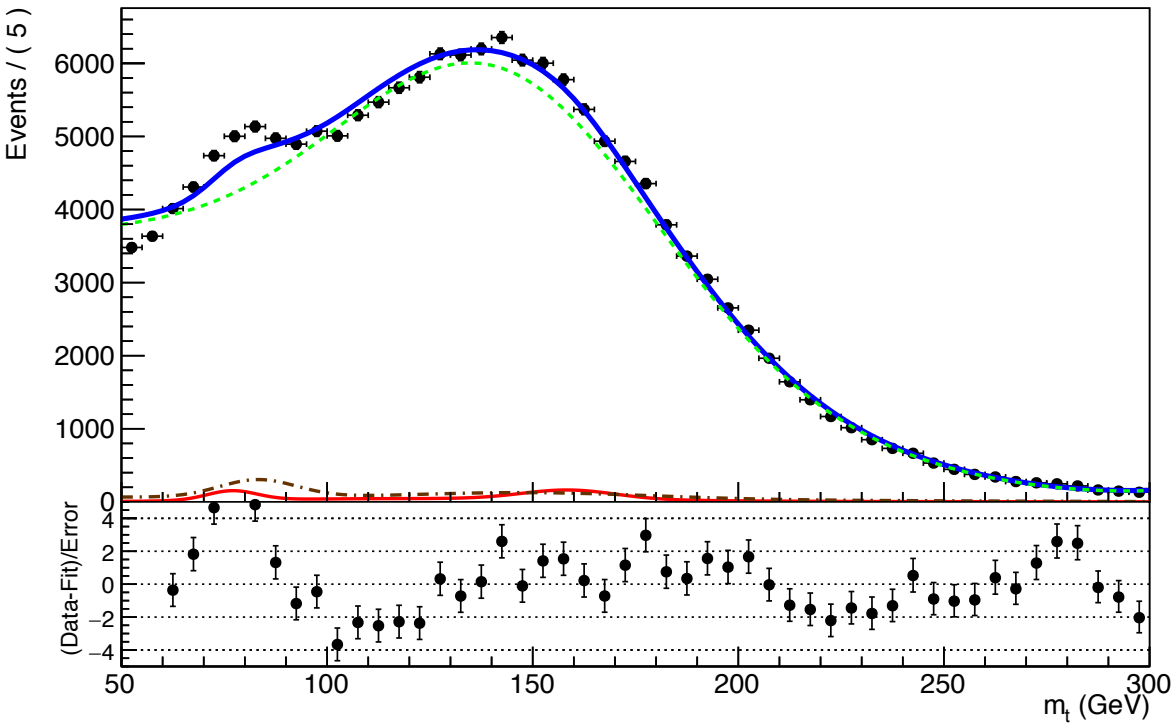
Signal Region (2btag) (2016)



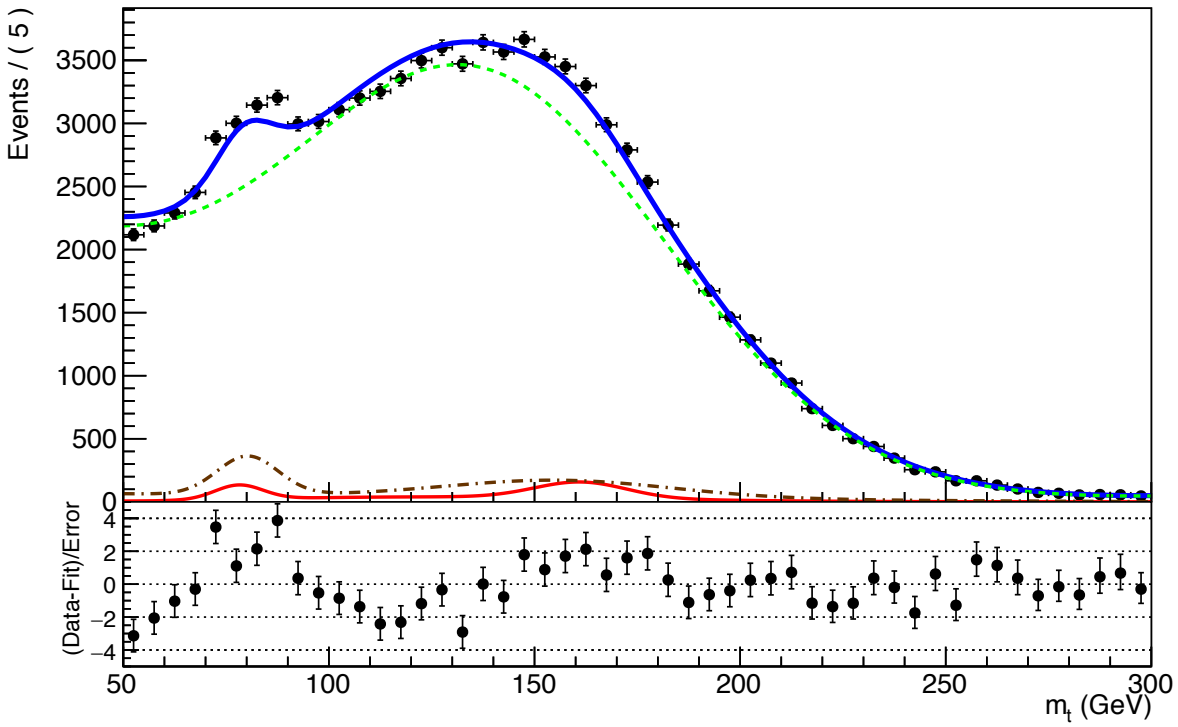
Result of the template fit on data in SR. The red line shows the ttbar contribution, the green line shows the QCD, and the brown line shows the subdominant backgrounds

Simultaneous Fit in 3 regions for 2017 and 2016 when eb is fixed

Control Region (0btag) (2017)



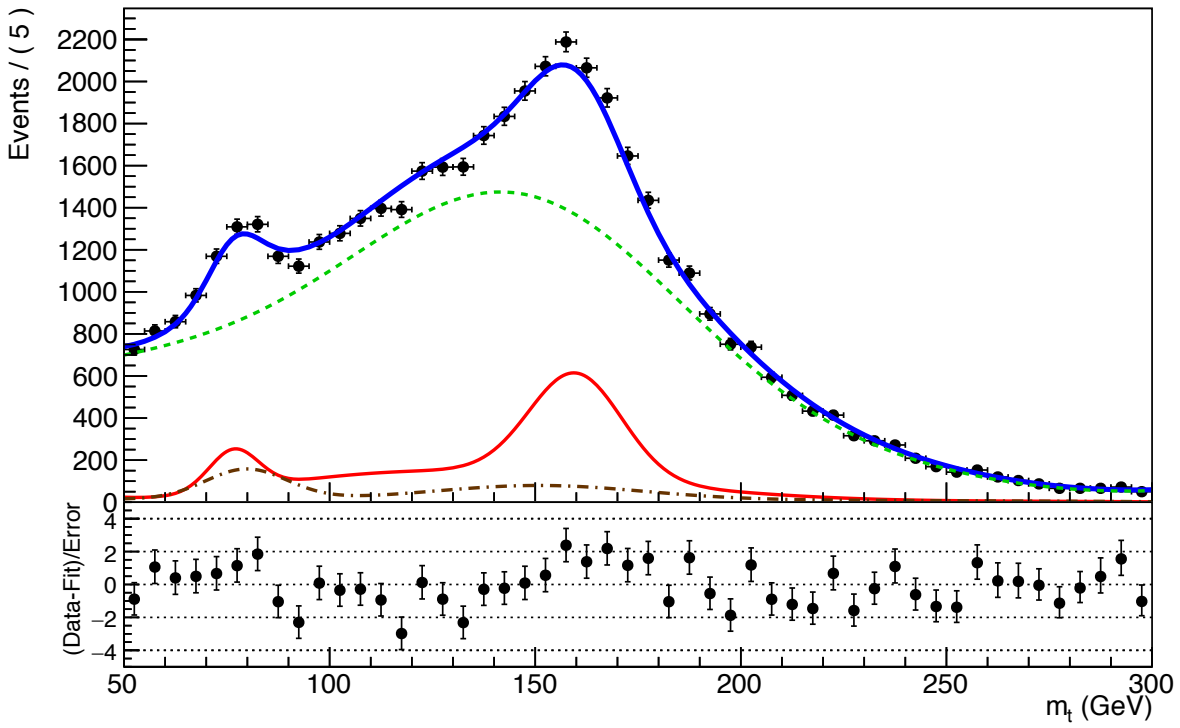
Control Region (0btag) (2016)



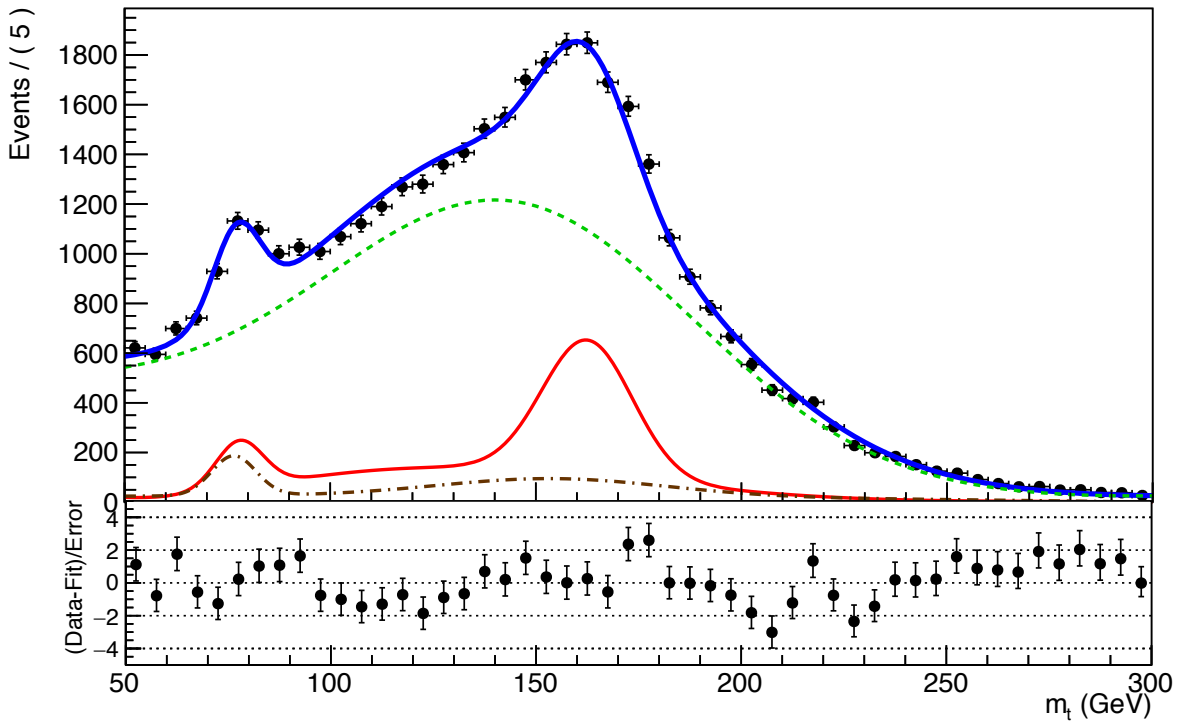
Result of the template fit on data in CR. The red line shows the tt contribution, the green line shows the QCD, and the brown line shows the subdominant backgrounds

Simultaneous Fit in 3 regions for 2017 and 2016 (1btag Region)

2017

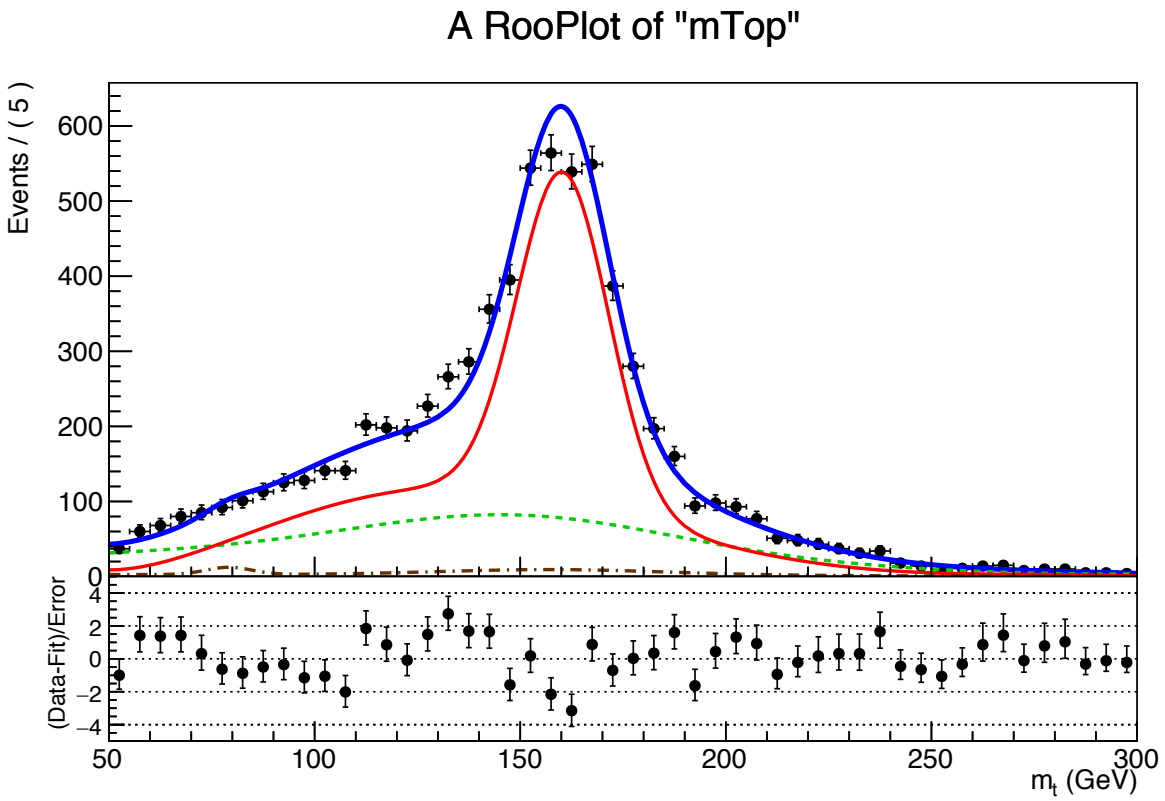


2016

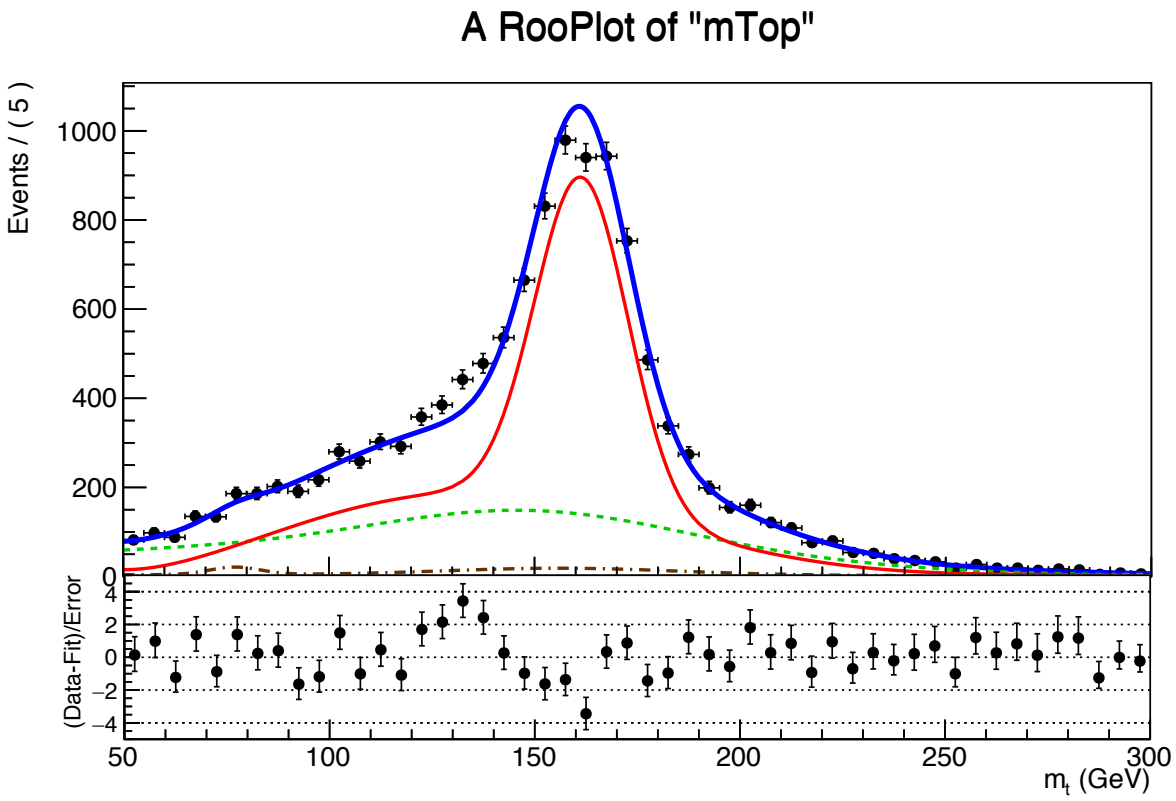


Simultaneous Fit in 3 regions for 2017 and 2018 when eb is fixed

Signal Region (2btag) (2017)



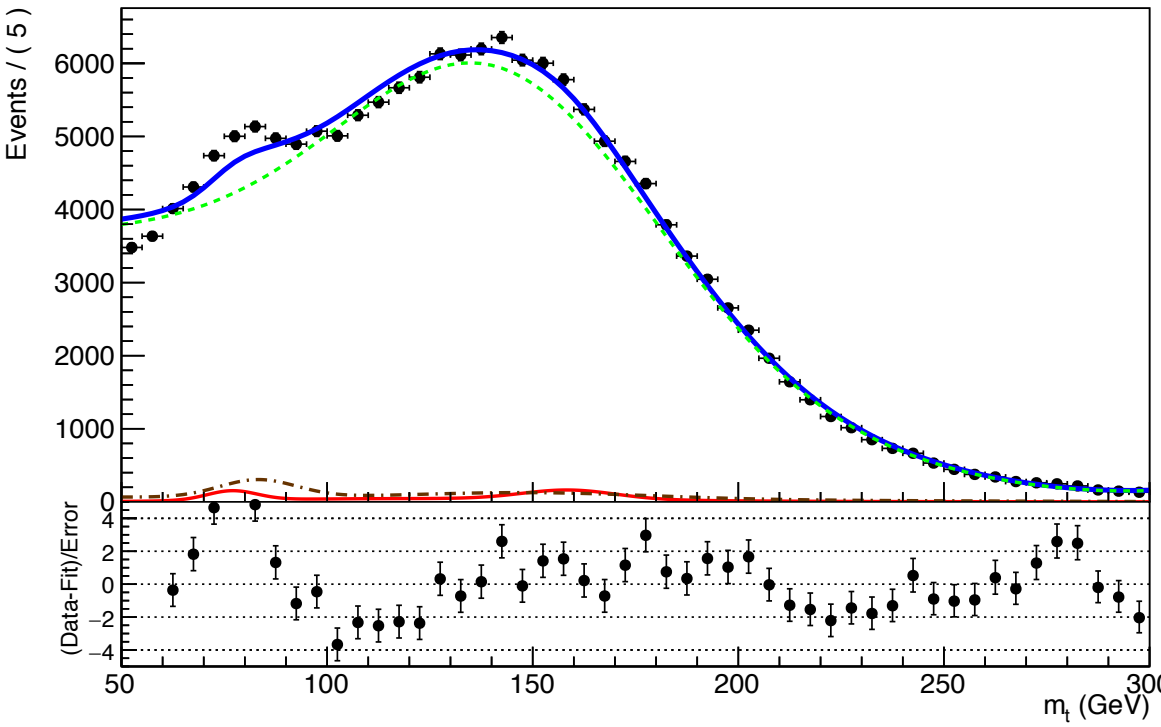
Signal Region (2btag) (2018)



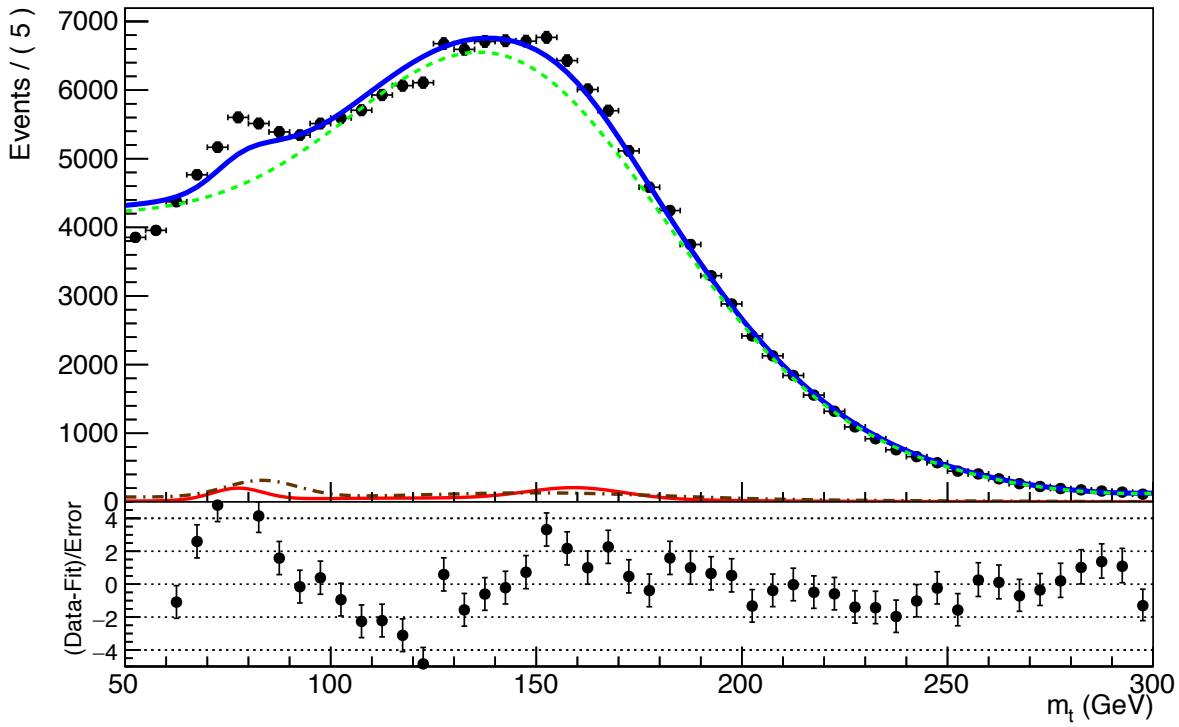
Result of the template fit on data in SR. The red line shows the ttbar contribution, the green line shows the QCD, and the brown line shows the subdominant backgrounds

Simultaneous Fit in 3 regions for 2017 and 2018 when eb is fixed

Control Region (0btag) (2017)



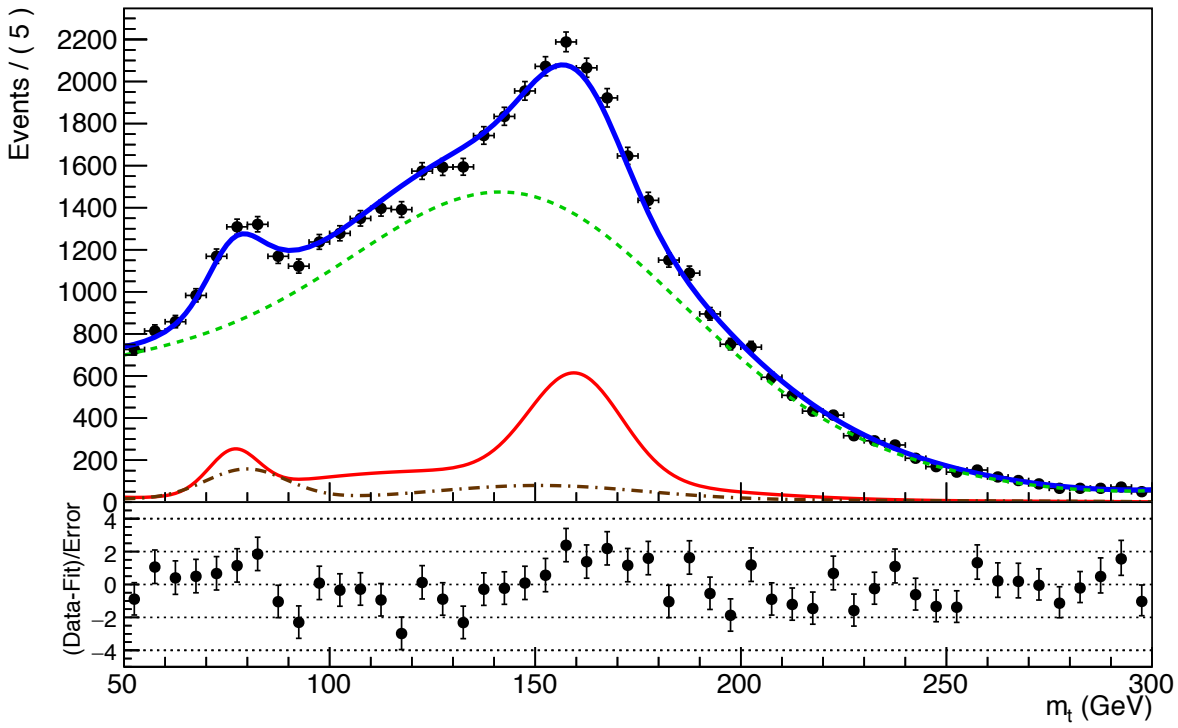
Control Region (0btag) (2018)



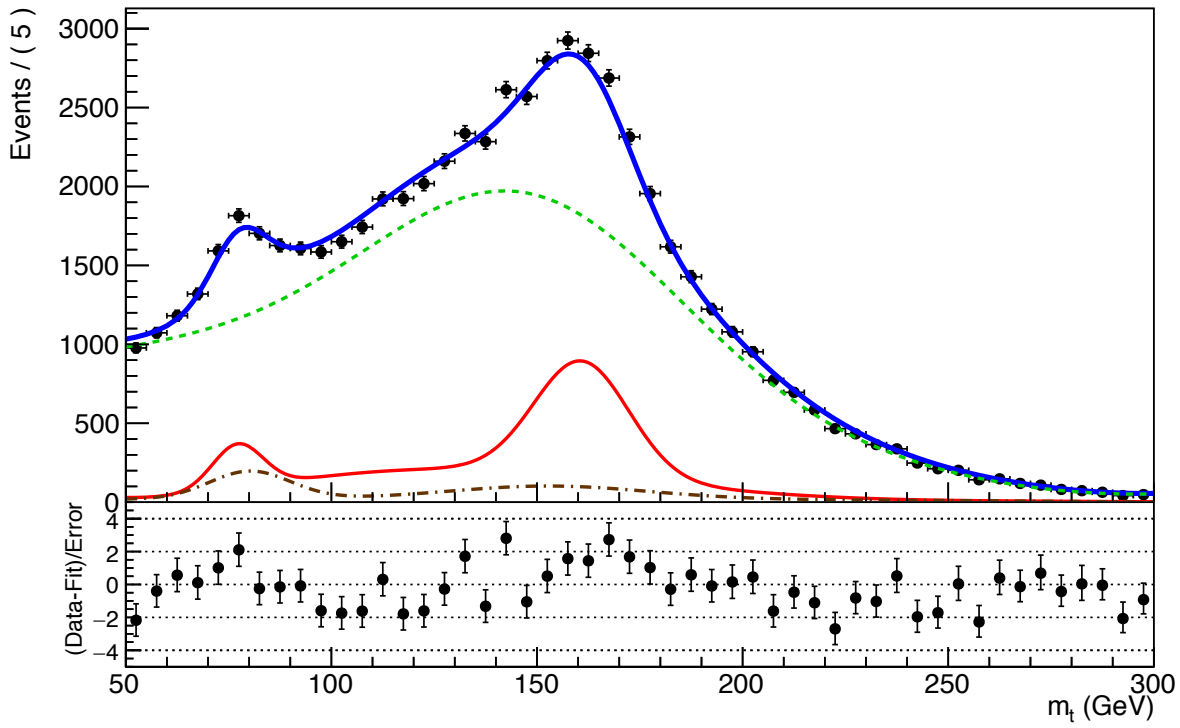
Result of the template fit on data in CR. The red line shows the tt contribution, the green line shows the QCD, and the brown line shows the subdominant backgrounds

Simultaneous Fit in 3 regions for 2017 and 2018 (1btag Region)

2017



2018



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

2016

Floating Parameter	FinalValue +/-	Error
kMassResol	9.2150e-01 +/-	2.07e-02
kMassScale	1.0023e+00 +/-	1.60e-03
kQCD_1b	6.3680e-03 +/-	4.58e-04
kQCD_2b	5.9385e-02 +/-	3.48e-02
nFitBkg_0b	4.5269e+03 +/-	4.25e+01
nFitBkg_1b	2.3356e+03 +/-	2.73e+02
nFitBkg_2b	2.0703e+02 +/-	2.32e+01
nFitQCD_0b	8.8323e+04 +/-	3.13e+02
nFitQCD_1b	3.0542e+04 +/-	2.62e+02
nFitQCD_2b	2.8400e+03 +/-	1.55e+02
nFitSig	1.3955e+04 +/-	3.69e+02

Ntt expected = 16351.2
Ntt observed = 13955.3
r = 0.85347

2017

Floating Parameter	FinalValue +/-	Error
kMassResol	9.8167e-01 +/-	2.54e-02
kMassScale	9.8694e-01 +/-	1.88e-03
kQCD_1b	4.6427e-03 +/-	3.05e-04
kQCD_2b	1.1491e-02 +/-	3.97e-03
nFitBkg_0b	4.0852e+03 +/-	4.44e+01
nFitBkg_1b	2.0768e+03 +/-	6.62e+01
nFitBkg_2b	2.1331e+02 +/-	4.46e+01
nFitQCD_0b	1.5591e+05 +/-	4.08e+02
nFitQCD_1b	3.7752e+04 +/-	2.83e+02
nFitQCD_2b	2.1182e+03 +/-	1.33e+02
nFitSig	1.3738e+04 +/-	3.84e+02

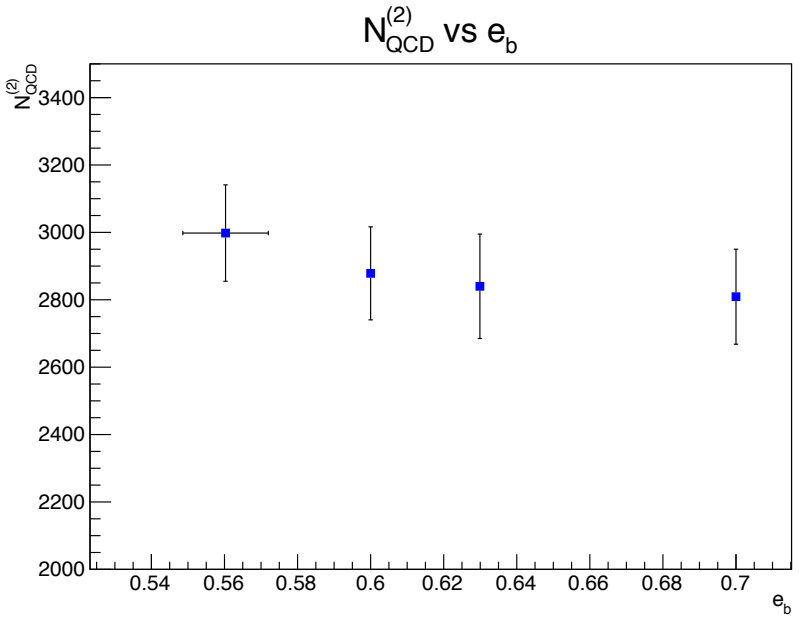
Ntt expected = 23720.7
Ntt observed = 13737.8
r = 0.579147

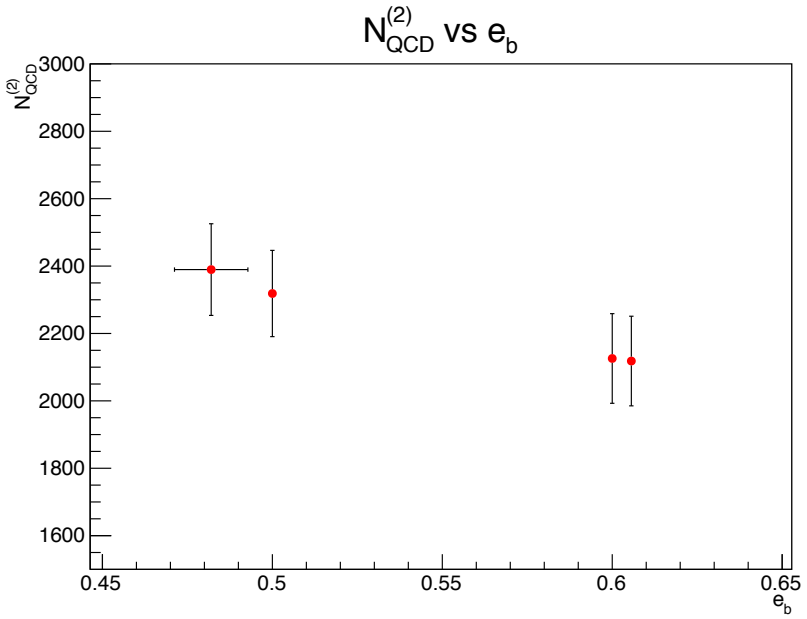
2018

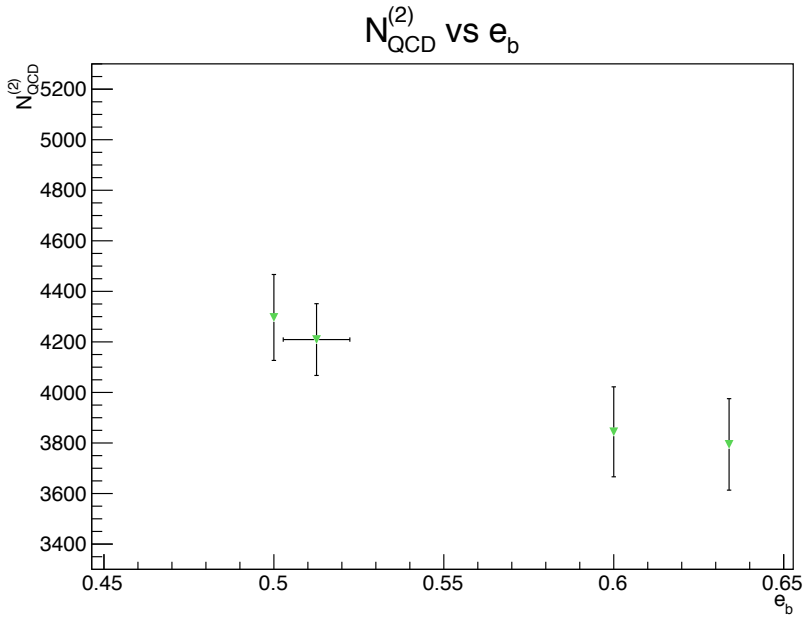
Floating Parameter	FinalValue +/-	Error
kMassResol	9.8988e-01 +/-	2.10e-02
kMassScale	9.9003e-01 +/-	1.48e-03
kQCD_1b	4.0033e-03 +/-	2.41e-04
kQCD_2b	1.0290e-02 +/-	2.65e-03
nFitBkg_0b	4.1842e+03 +/-	4.02e+01
nFitBkg_1b	2.6492e+03 +/-	6.75e+01
nFitBkg_2b	4.1749e+02 +/-	3.89e+01
nFitQCD_0b	1.6973e+05 +/-	4.26e+02
nFitQCD_1b	5.0264e+04 +/-	3.25e+02
nFitQCD_2b	3.7944e+03 +/-	1.81e+02
nFitSig	2.0663e+04 +/-	4.54e+02

Ntt expected = 30676.2
Ntt observed = 20662.5
r = 0.673567

- We are checking for different values of e_b , the output of the $N^{(2)}_{QCD}$ for 2016 and 2017
- Points of interest are from 0.4 - 0.8 but especially 0.5-0.7 for 2016 and 0.45-0.6 5 for 2017
  - Calculated btagging efficiency for both years
  - btagging efficiency when the parameter is set as a free nuisance in the simultaneous fit
  - 2016: eb (fit) ≈ 0.56 and eb (calculated) ≈0.63
  - 2017: eb(fit) ≈ 0.48 and eb (calculated) ≈ 0.61
  - 2018: eb(fit) ≈ 0.52 and eb (calculated) ≈ 0.63



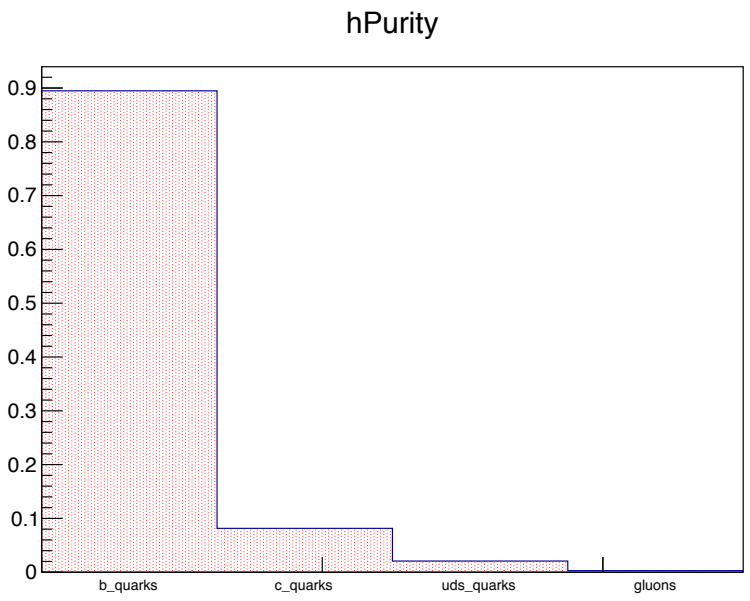




BACKUP SLIDES

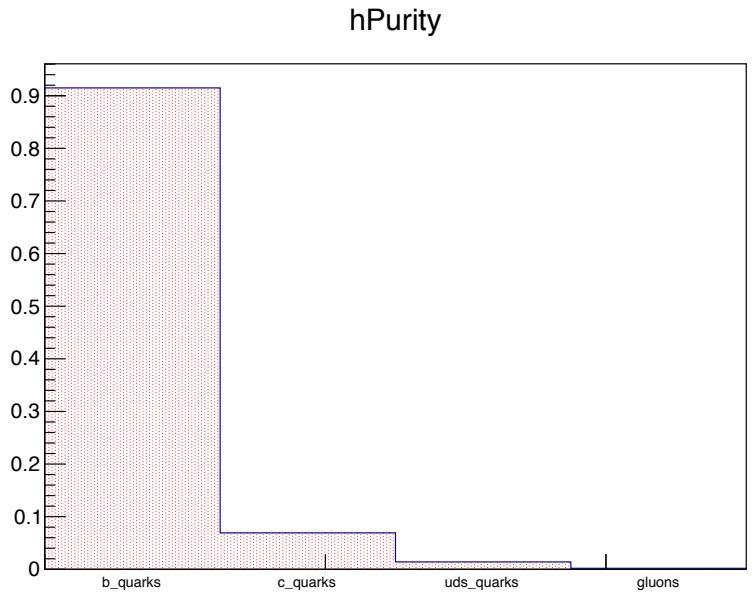
Btagging purity

2016



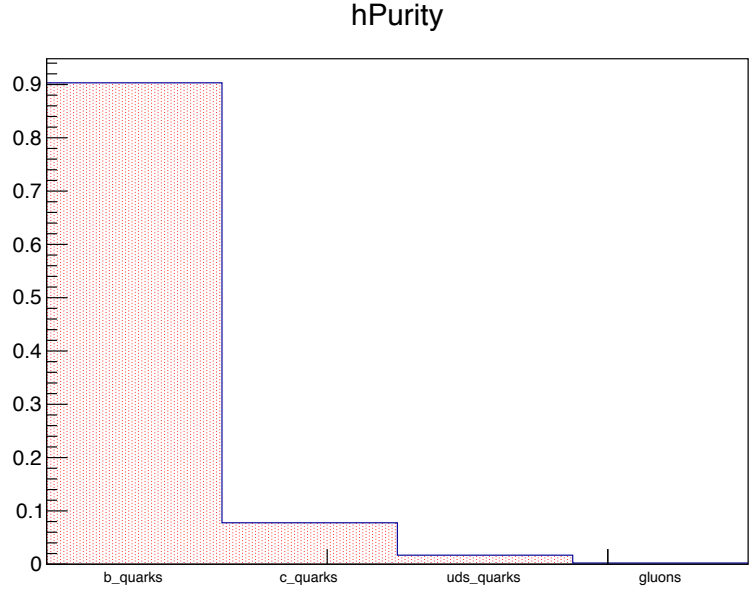
Purity ≈ 0. 894

2017

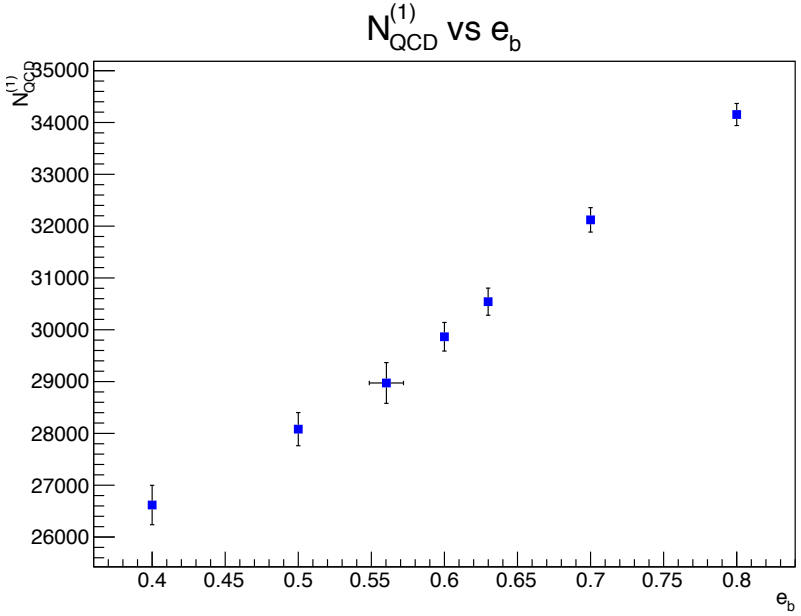


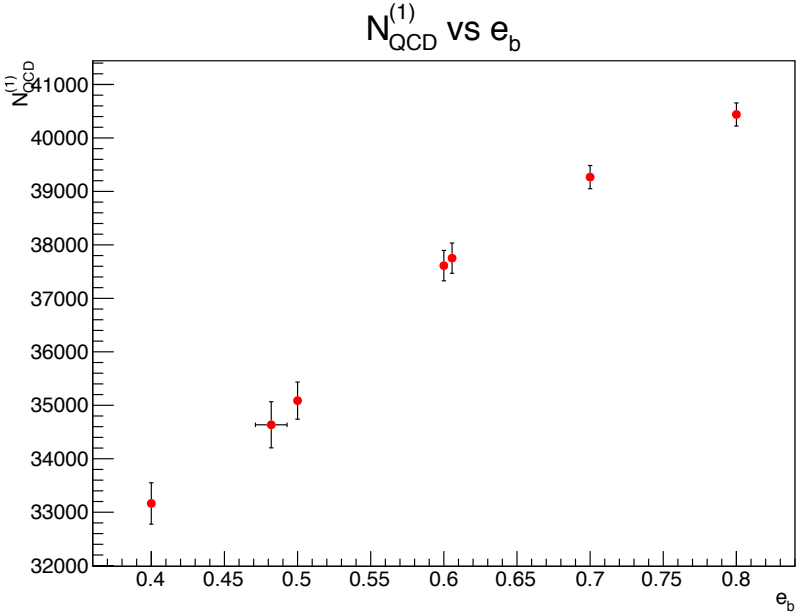
Purity ≈ 0.915

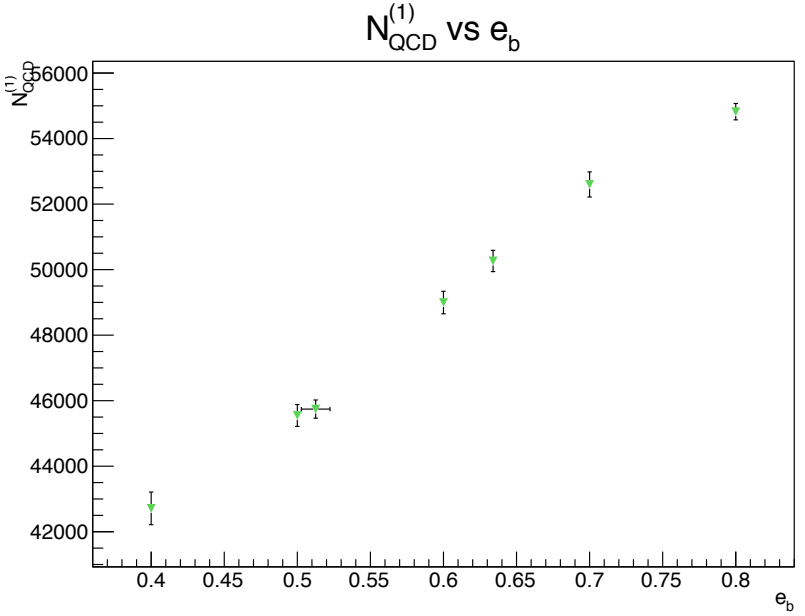
2018



Purity ≈ 0.903



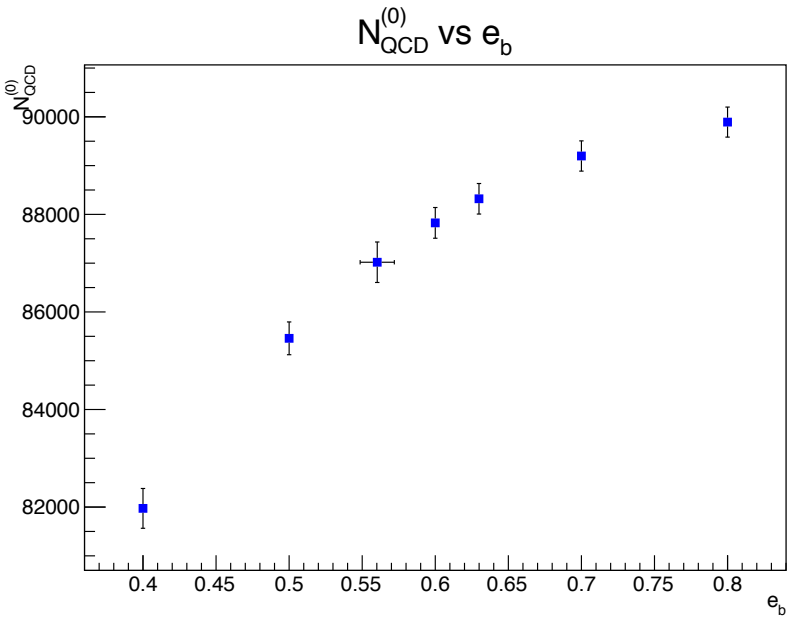


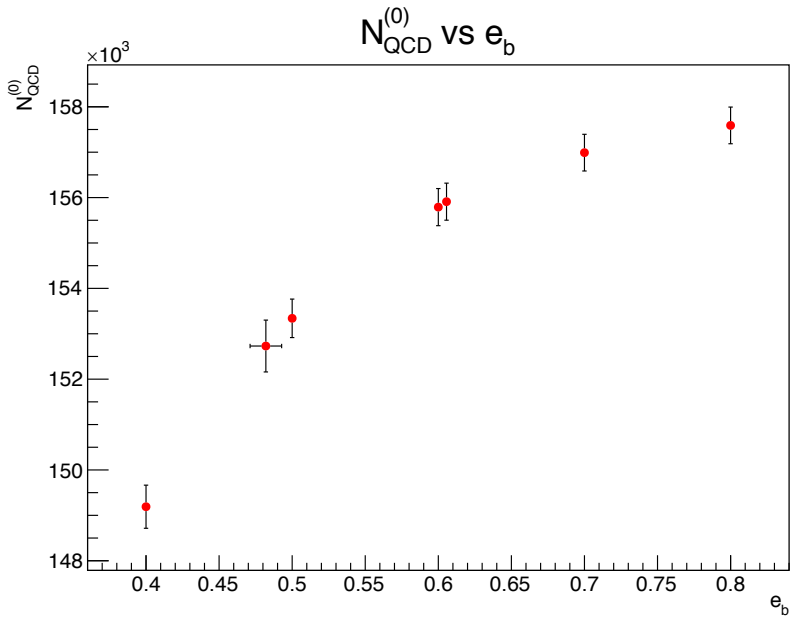


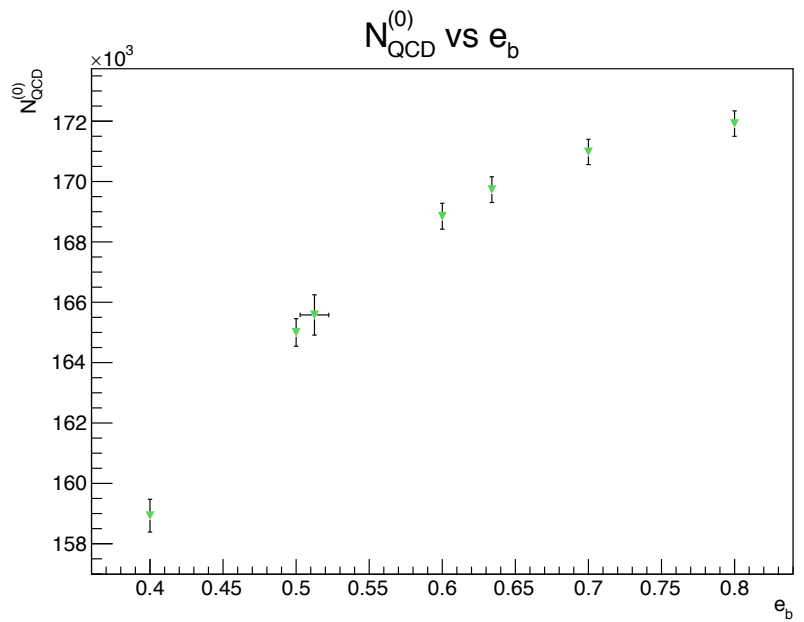
2016

2017

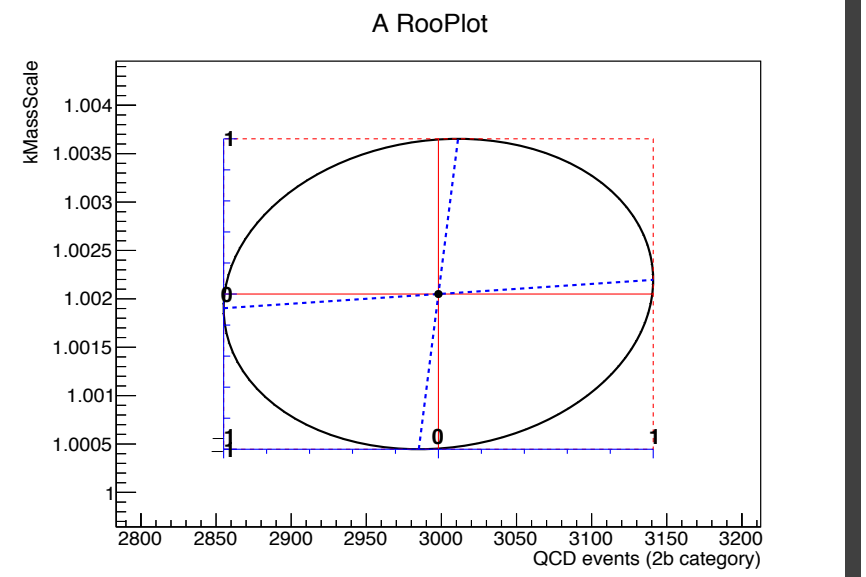
2018

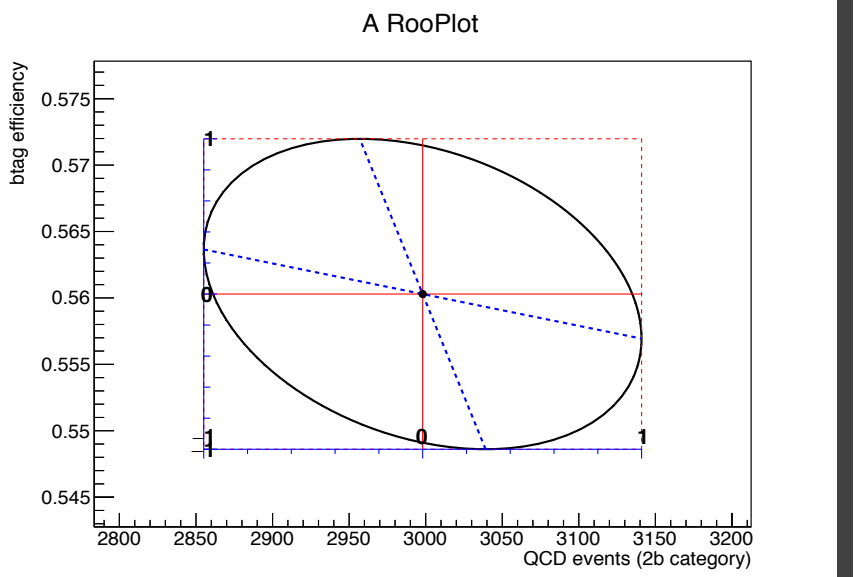


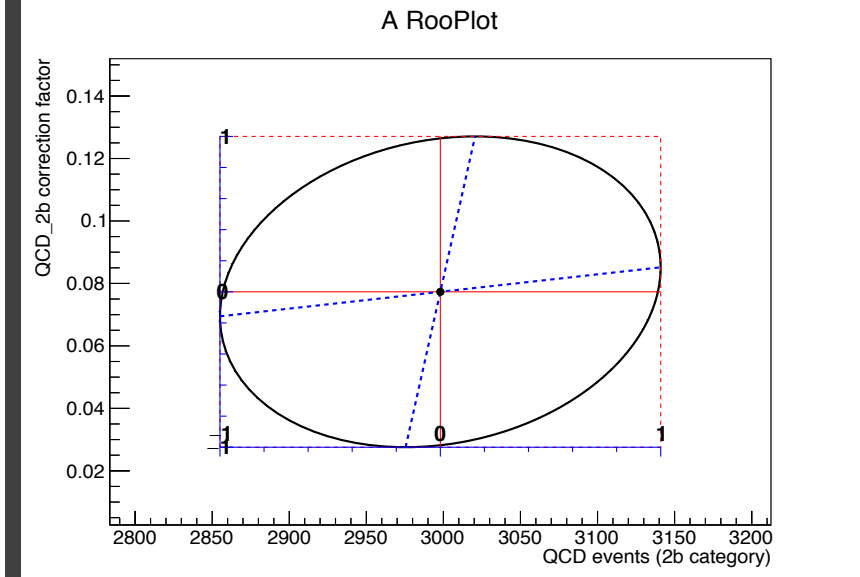


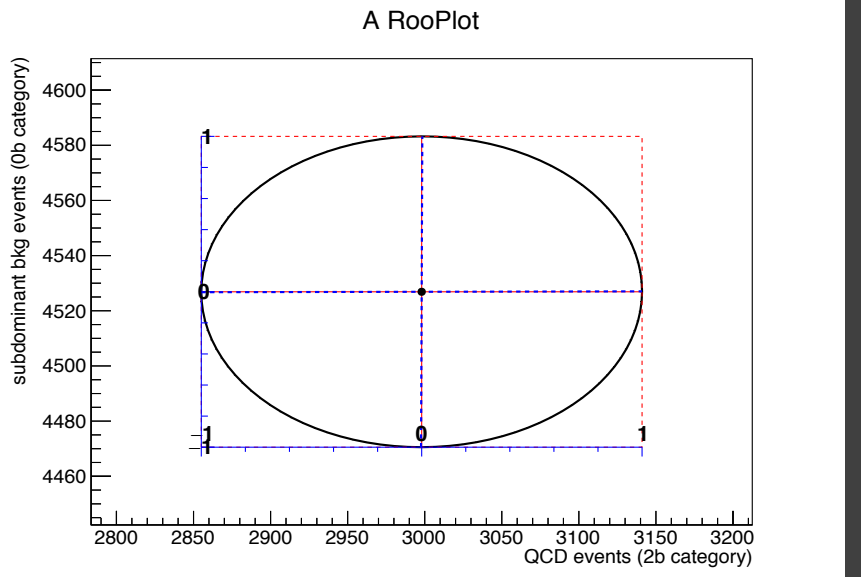


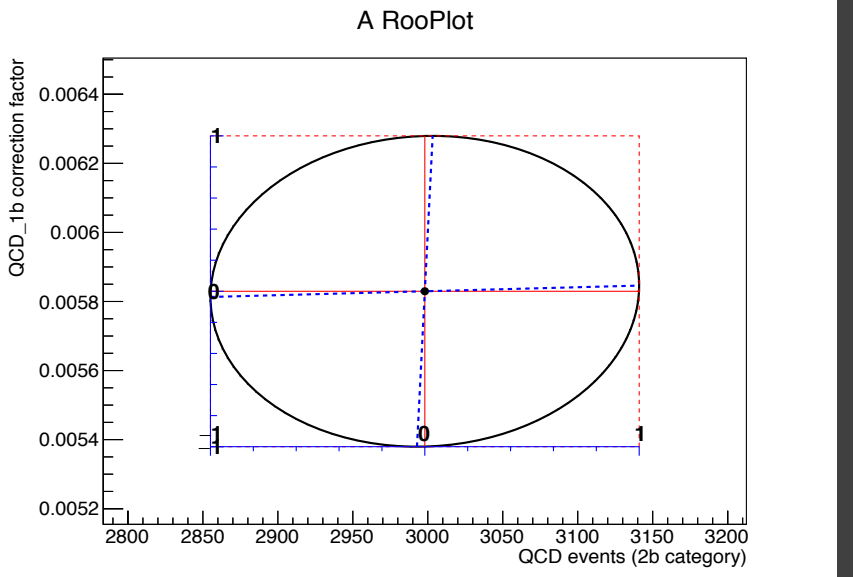
Correlation plots $N_{QCD(2)}$ vs all nuisances from fit when eb runs free for 2016

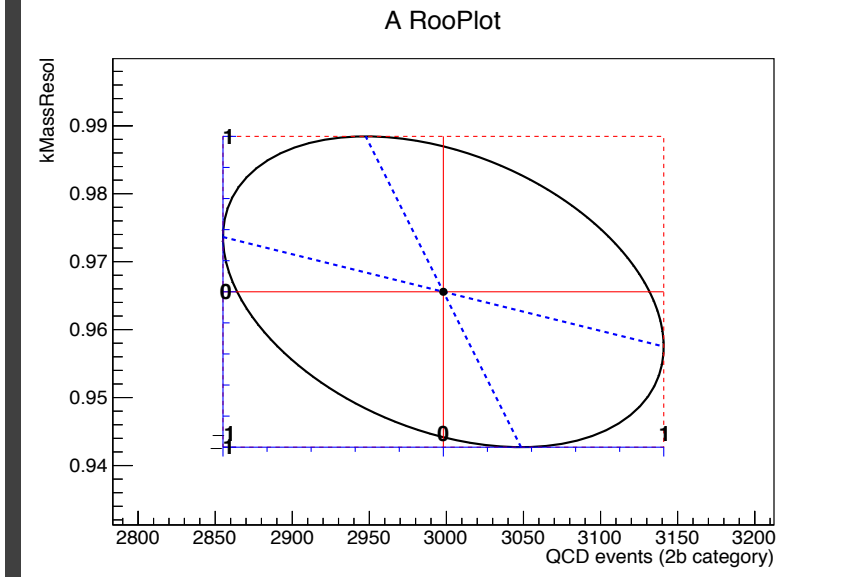




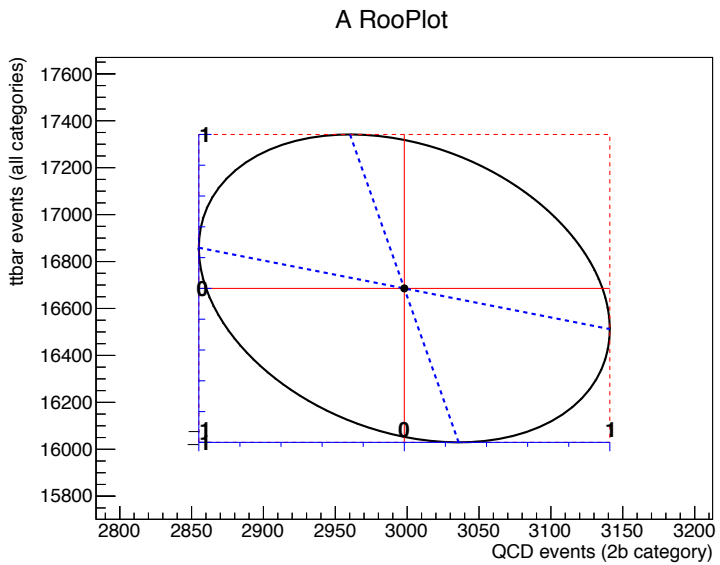


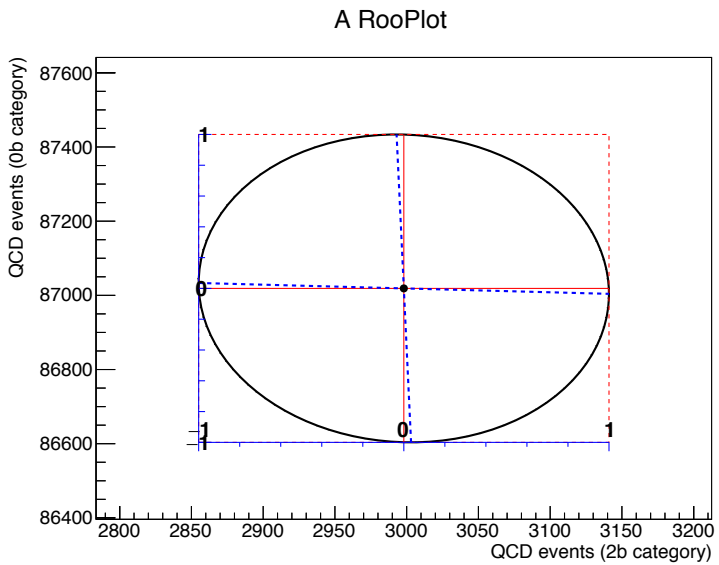


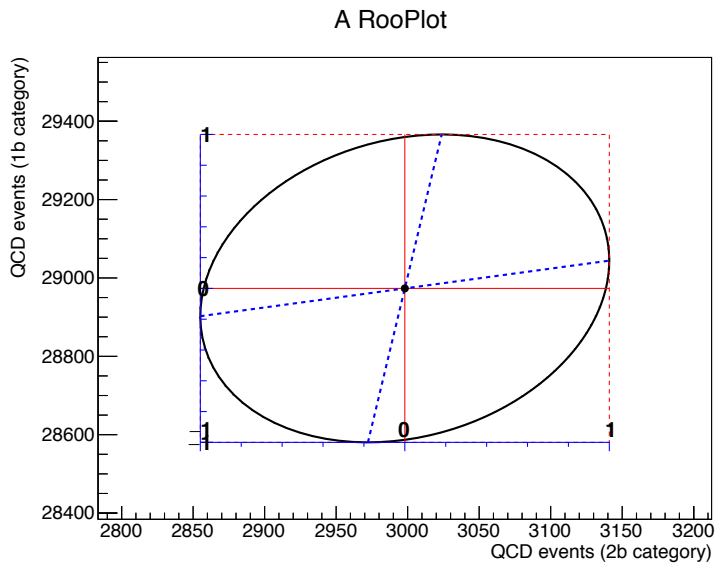


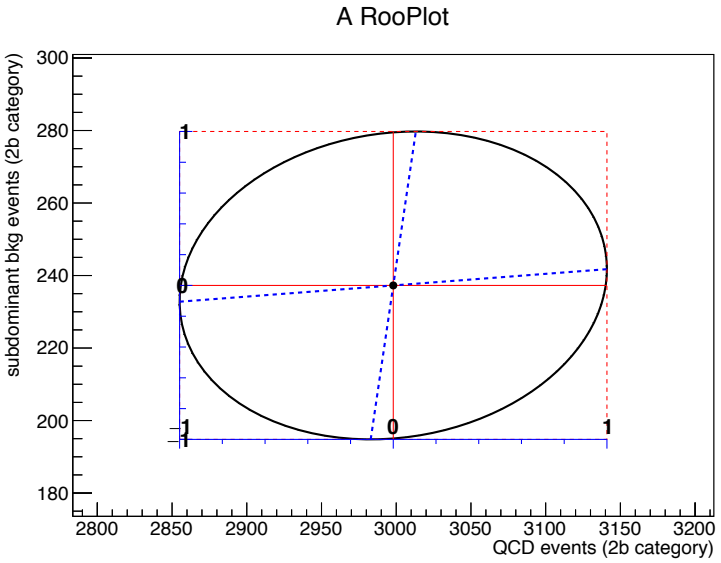


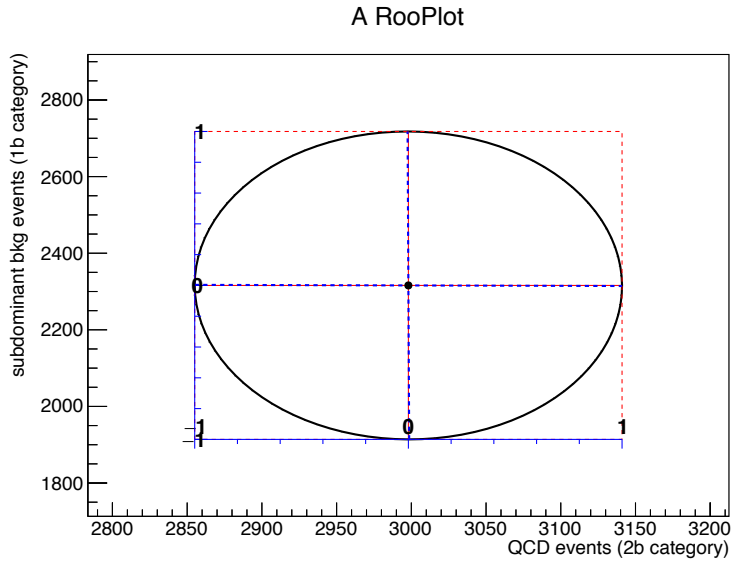
Correlation plots $N_{QCD(2)}$ vs all nuisances from fit when eb runs free for 2016



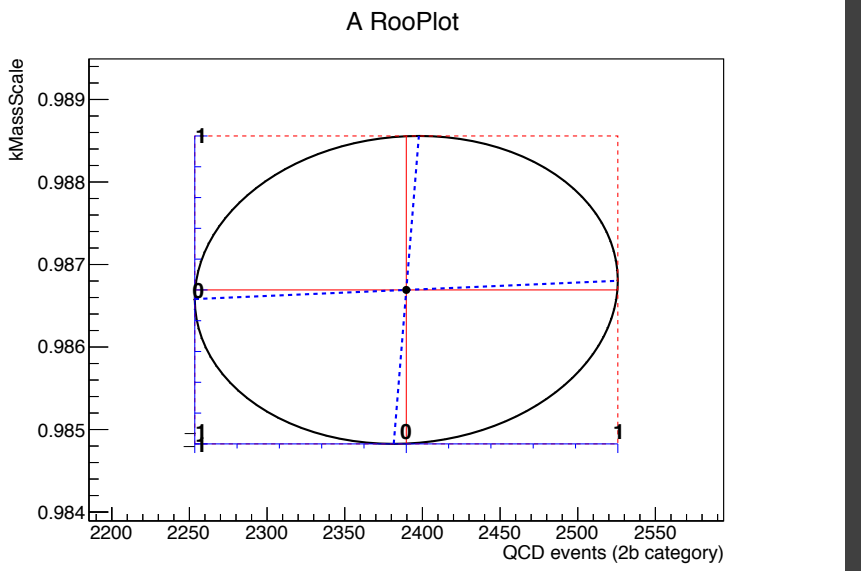


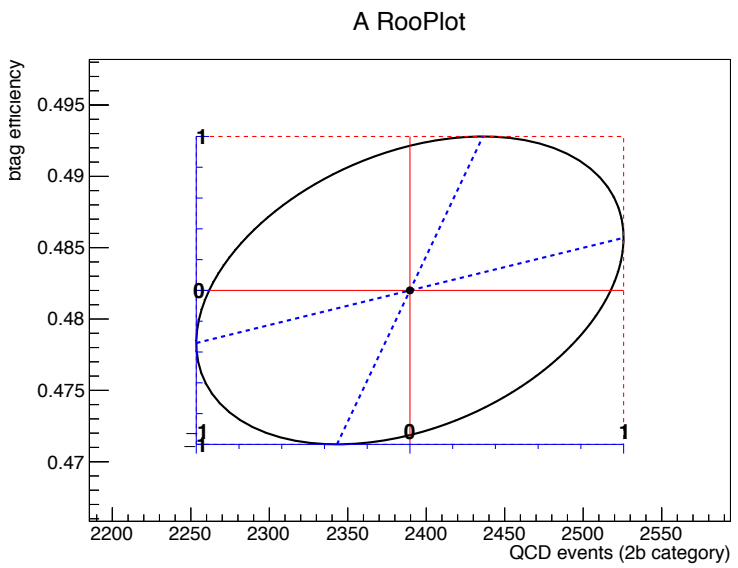


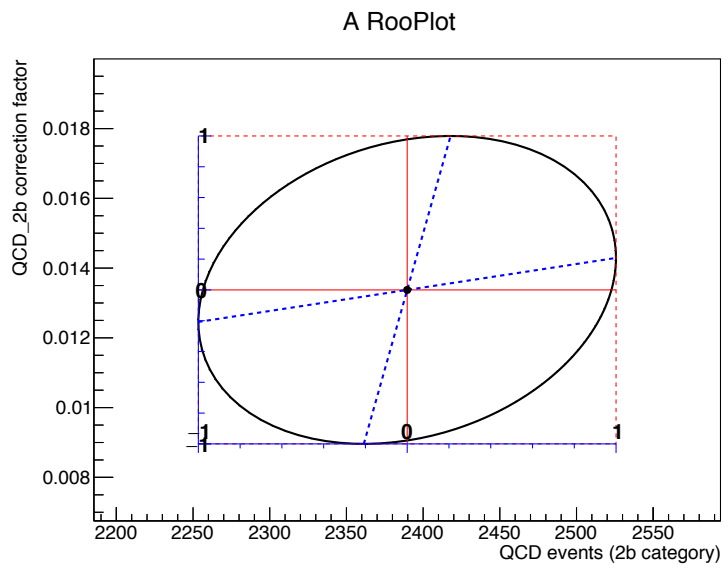


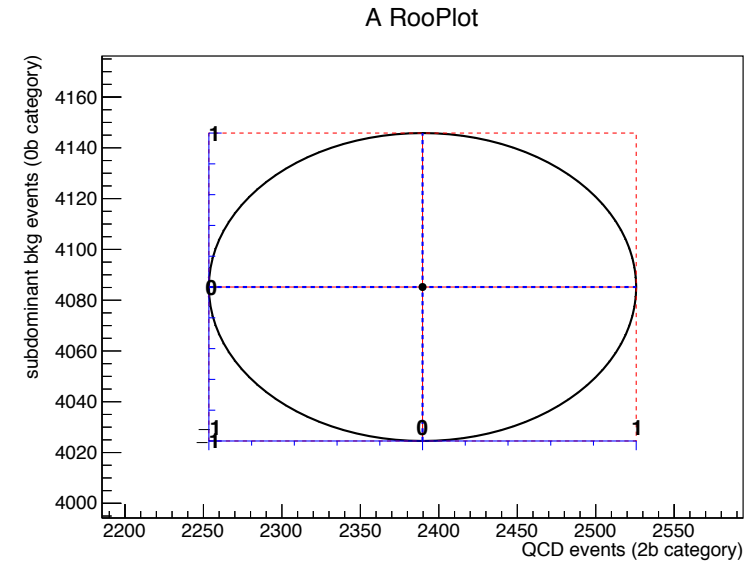


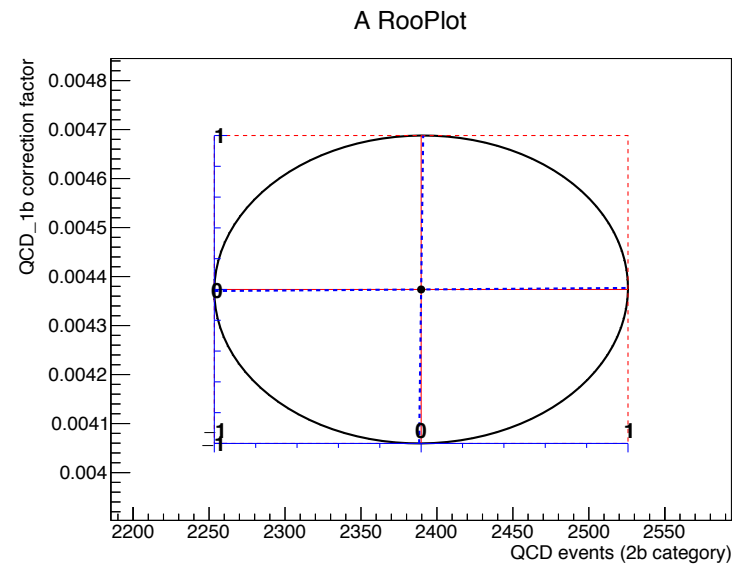
Correlation plots $N_{QCD(2)}$ vs all nuisances from fit when eb runs free for 2017

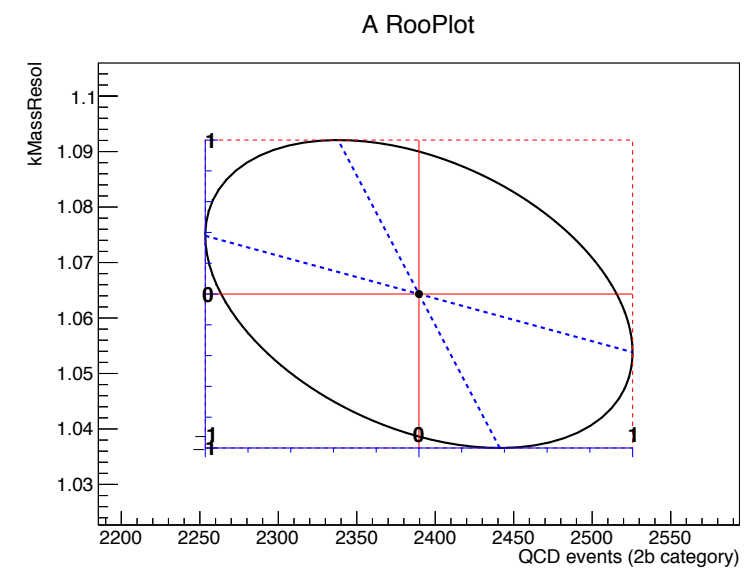




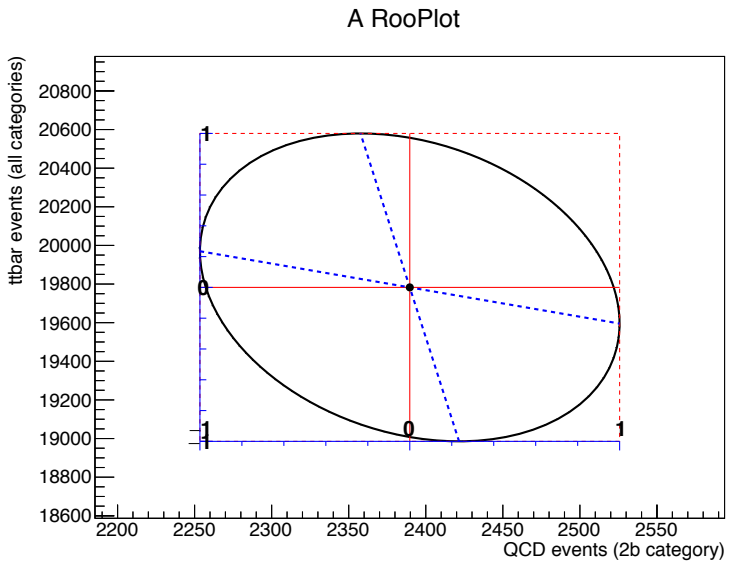


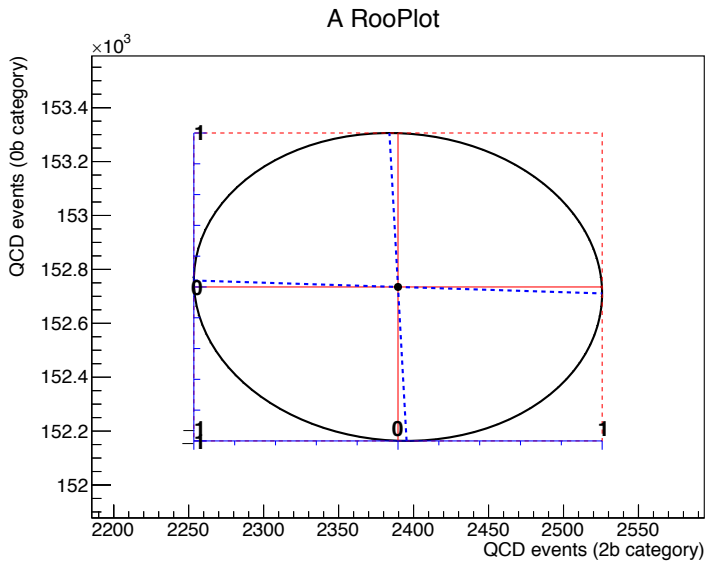


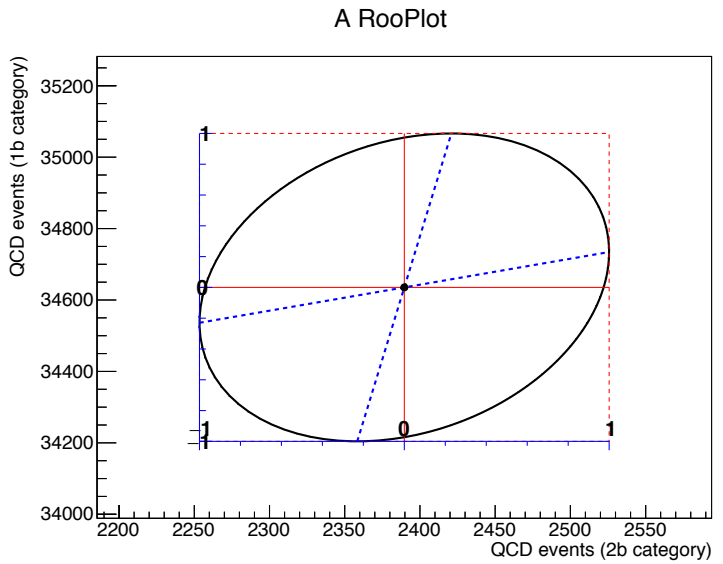


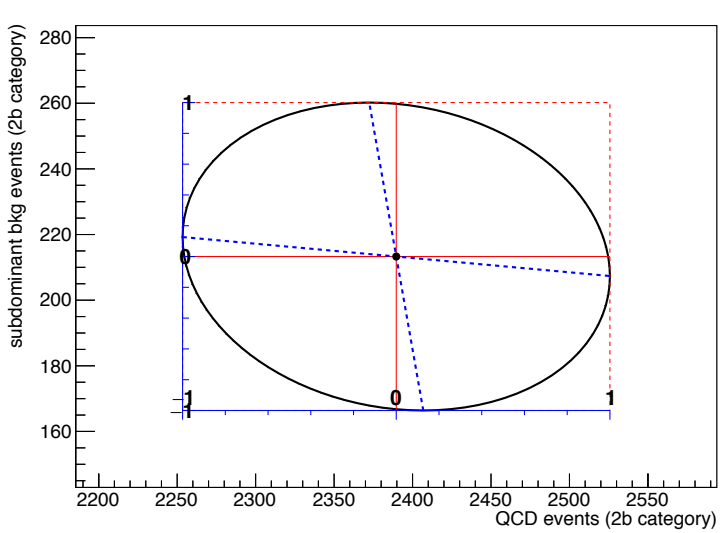


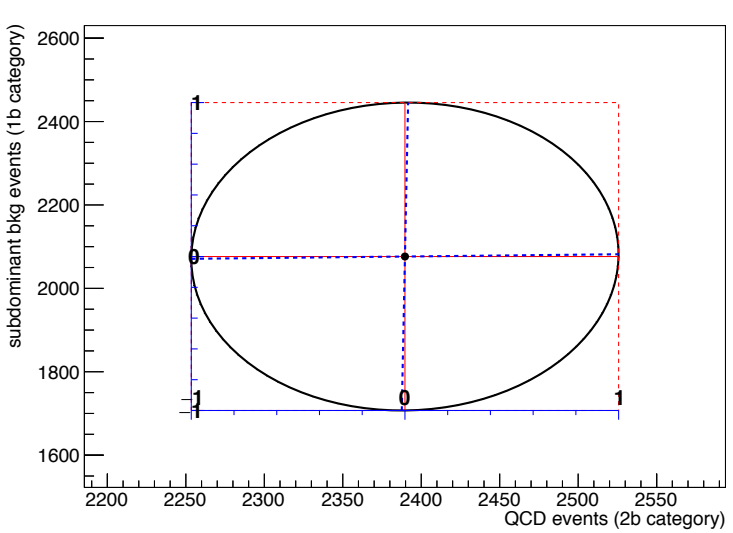
Correlation plots $N_{QCD(2)}$ vs all nuisances from fit when eb runs free for 2017



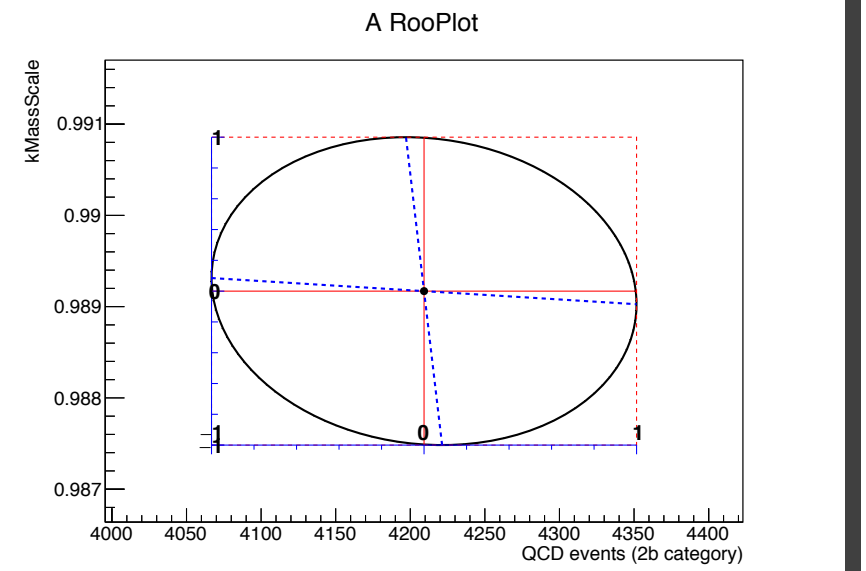


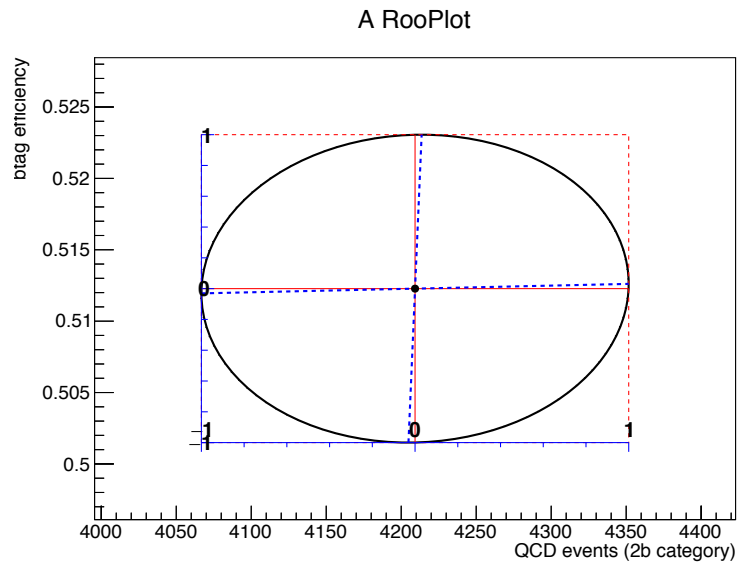


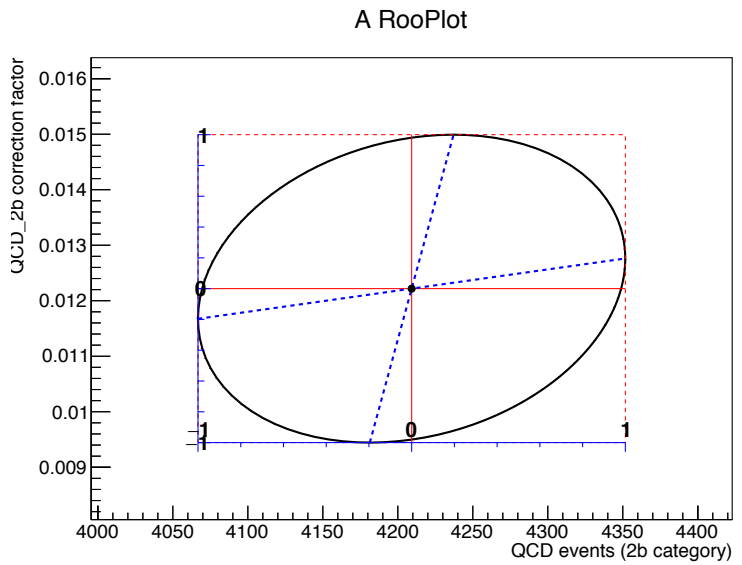


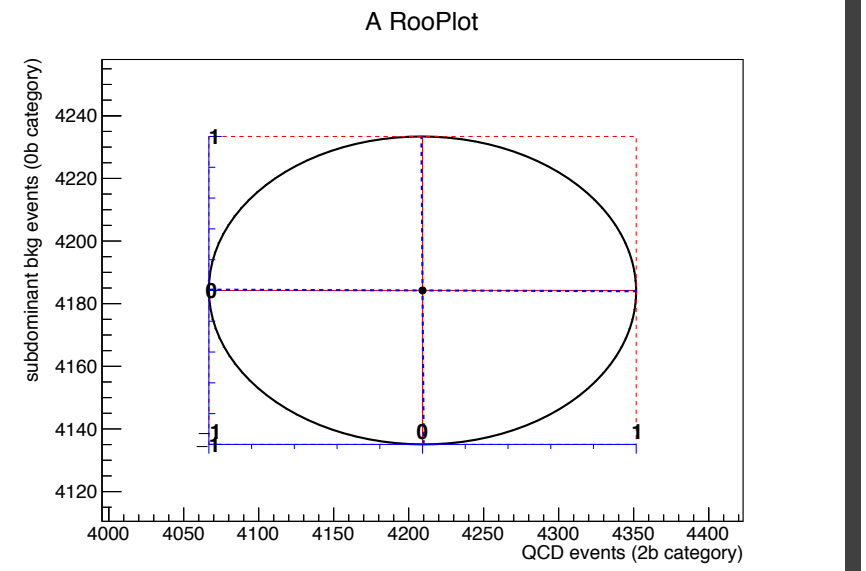


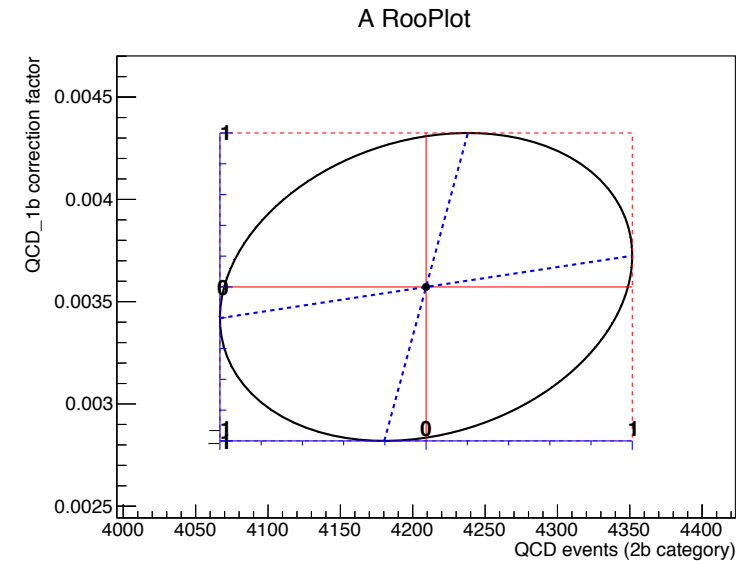
Correlation plots $N_{QCD(2)}$ vs all nuisances from fit when eb runs free for 2018

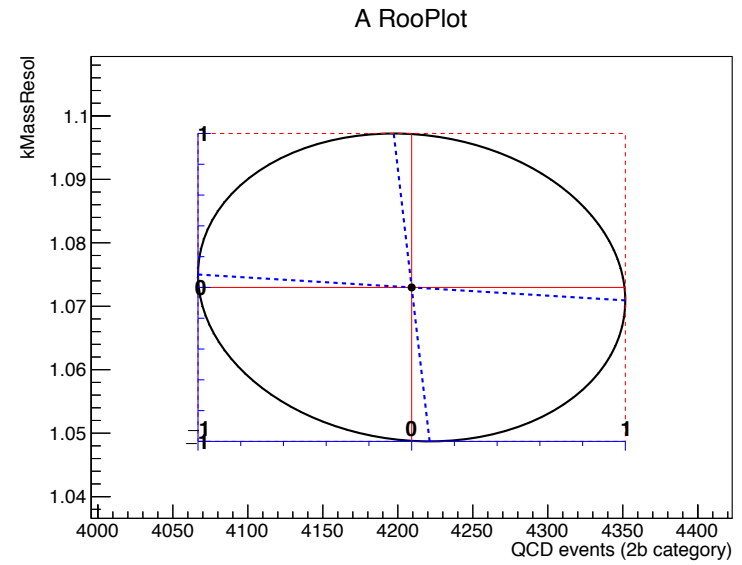




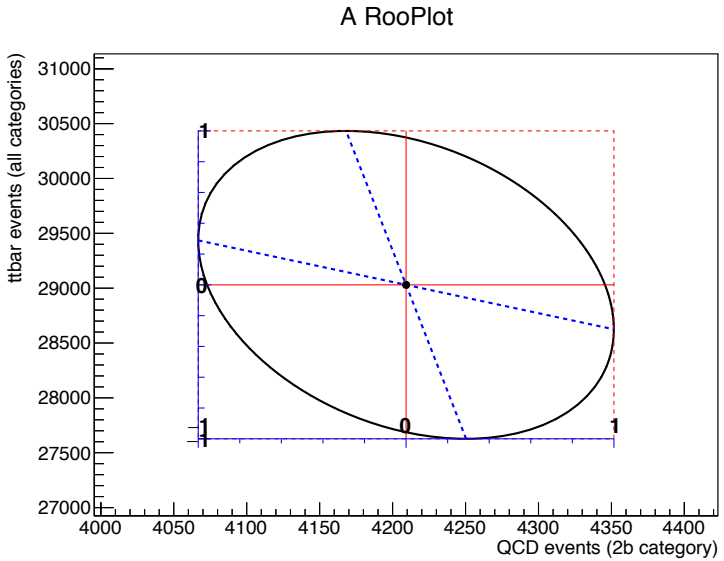


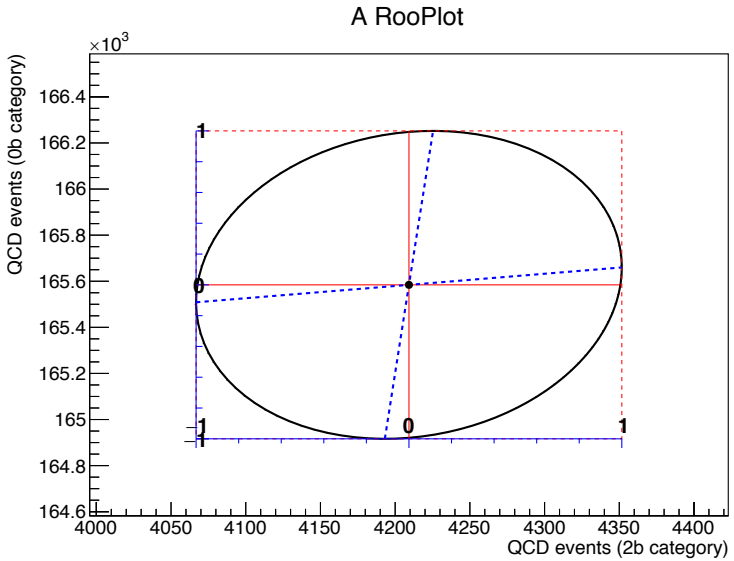


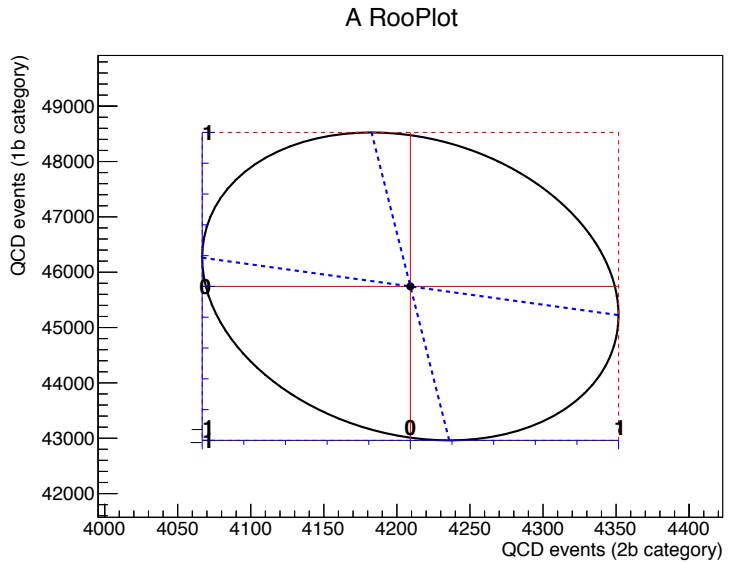


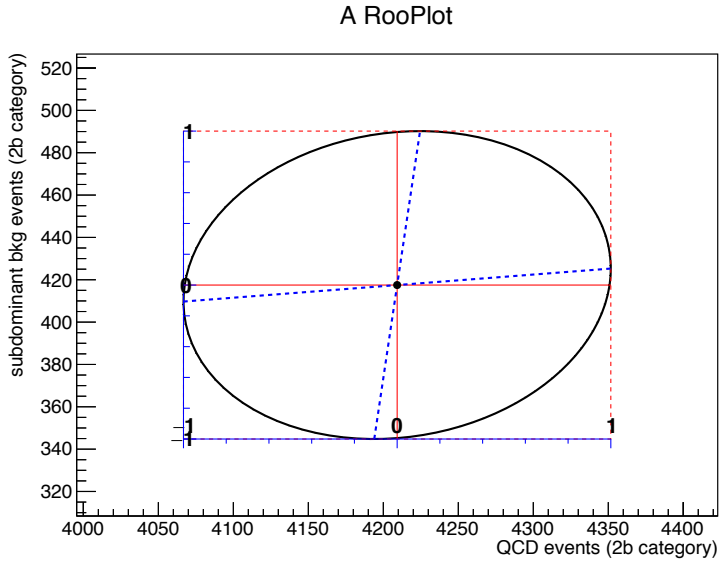


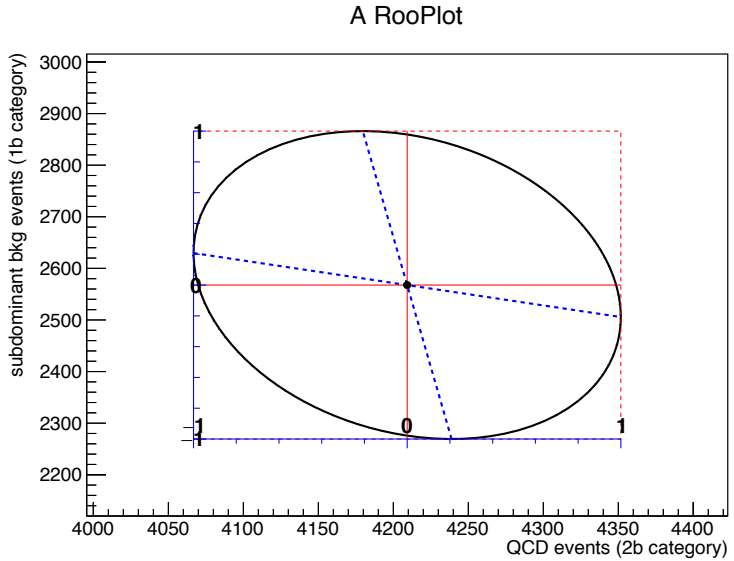
Correlation plots $N_{QCD(2)}$ vs all nuisances from fit when eb runs free for 2018



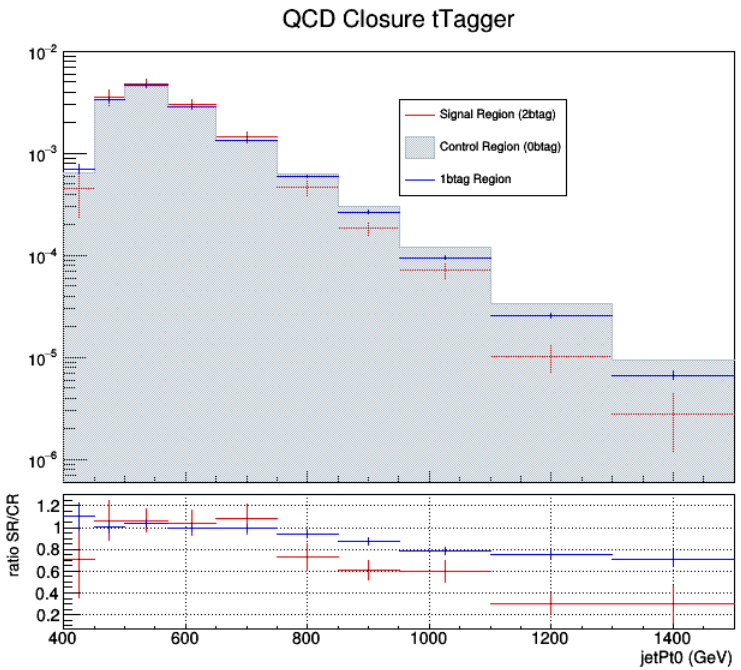


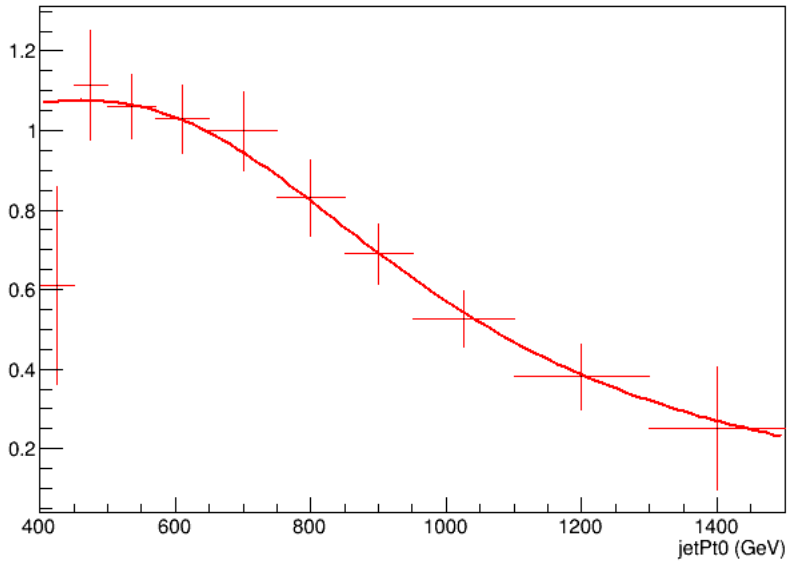


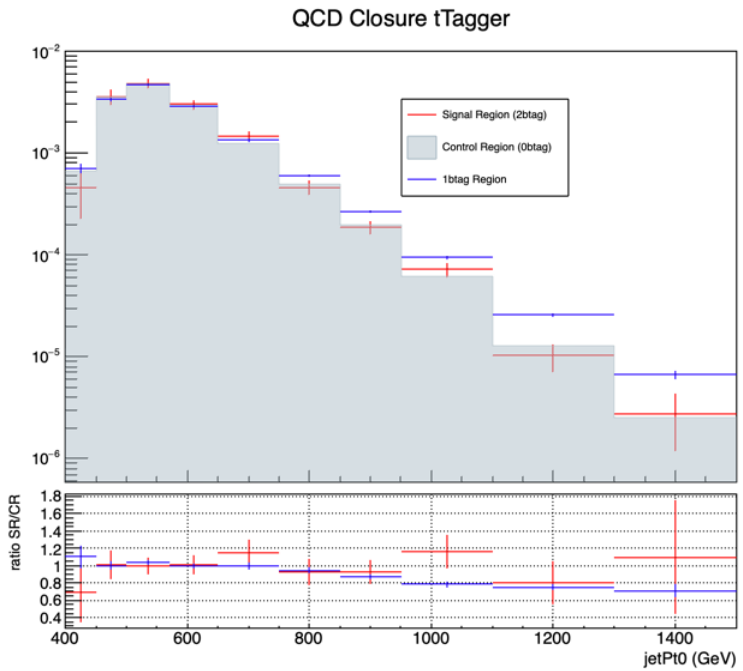




2017 QCD Closure and the fit ratio

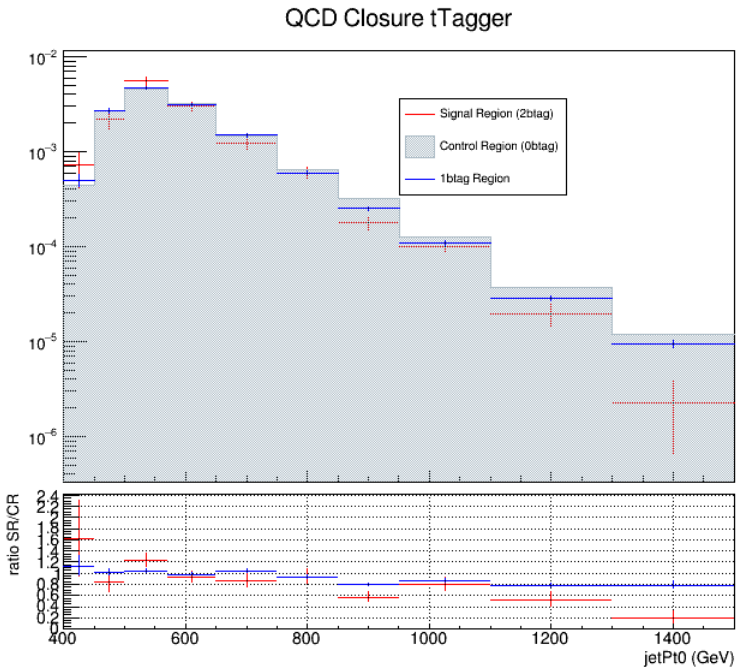


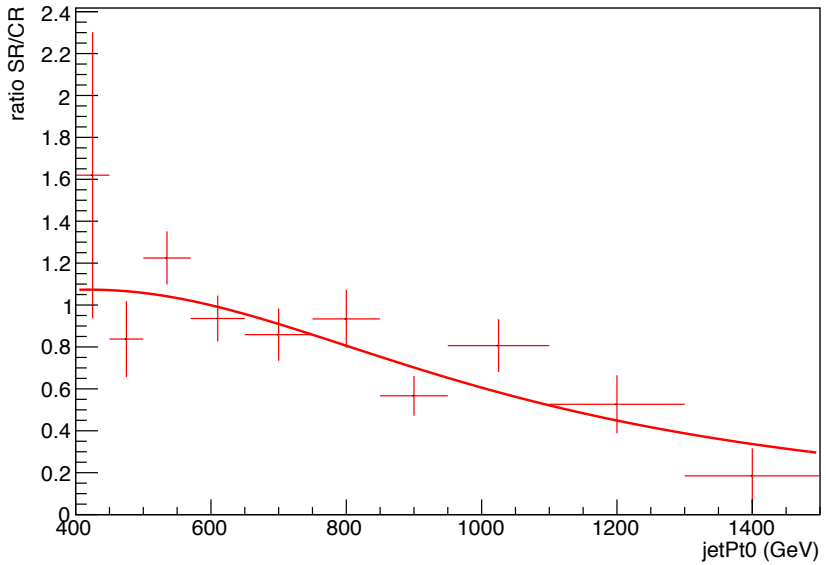


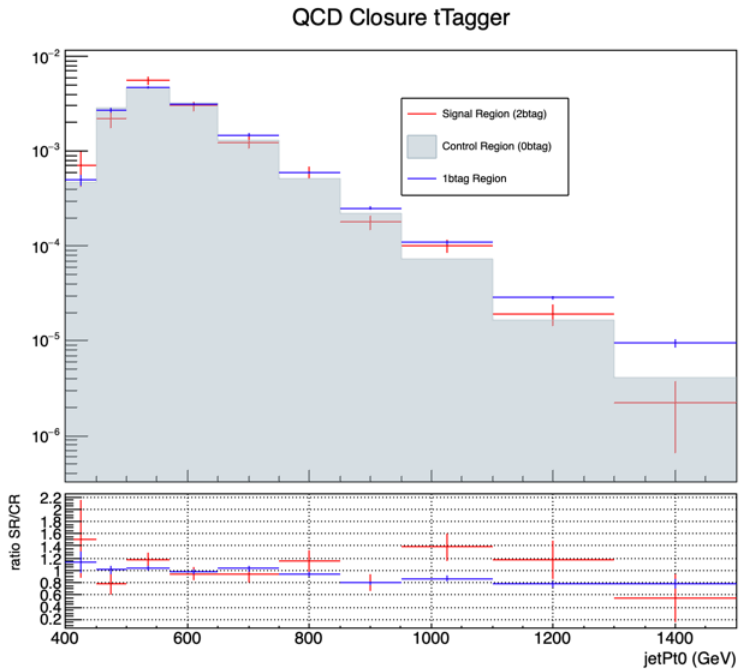


Discovered shape in the SR/CR → Fit the ratio with the function: → Output of the new QCD closure

2018 QCD Closure and the fit ratio

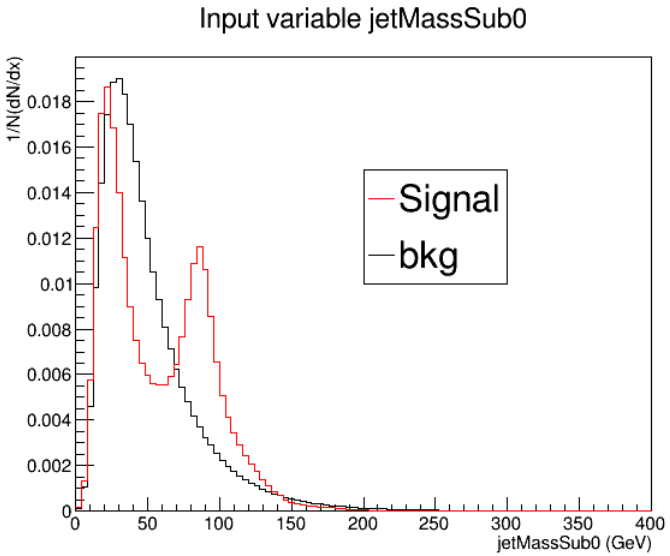


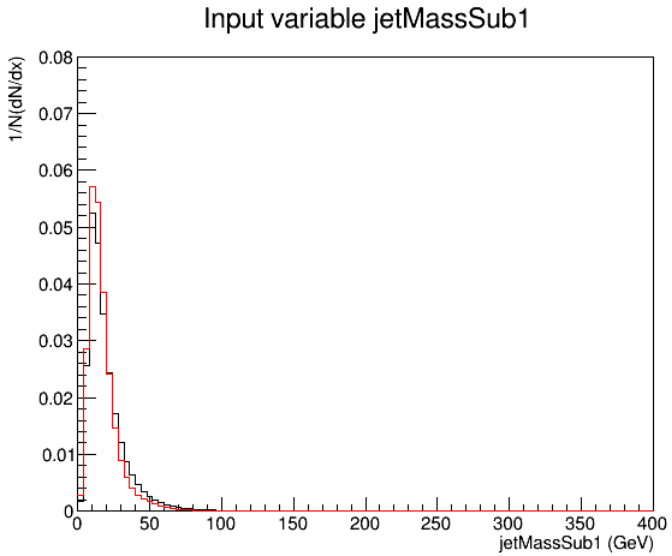


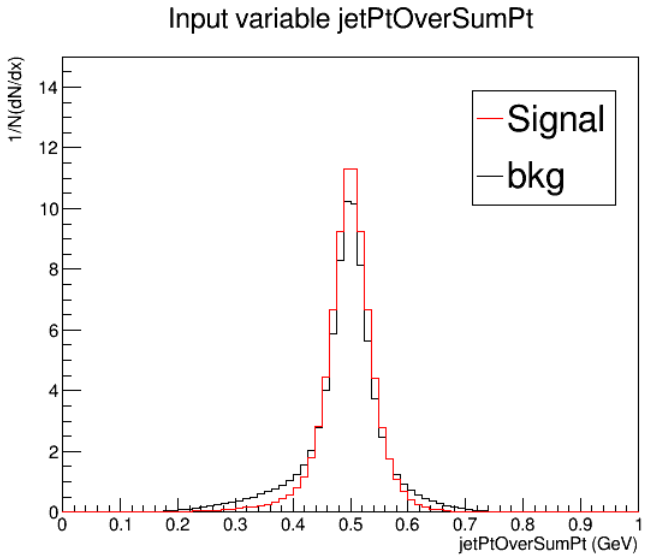


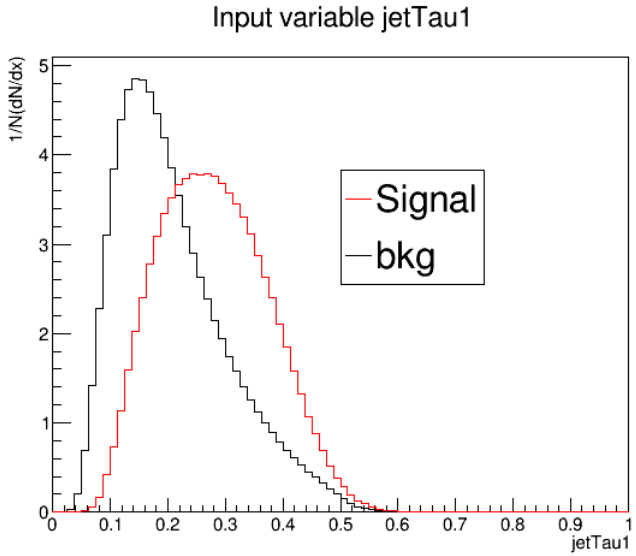
Discovered shape in the SR/CR → Fit the ratio with the function: → Output of the new QCD closure

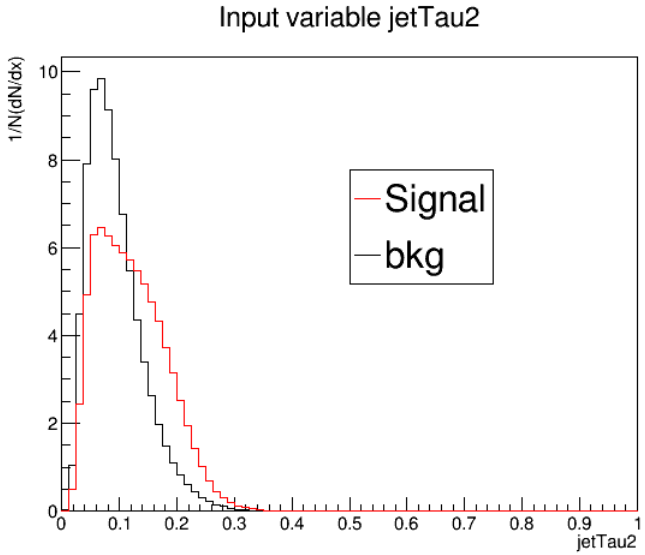
Training variables 2017

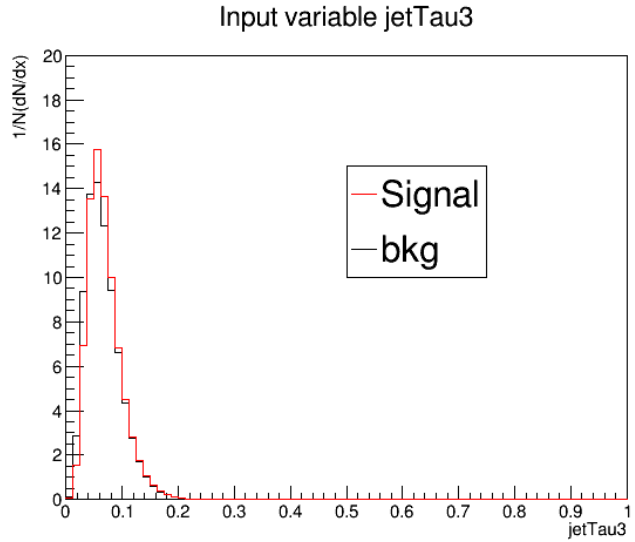












Training variables 2017