Measurement of differential production cross section for boosted top quarks in the all hadronic channel

NTUA

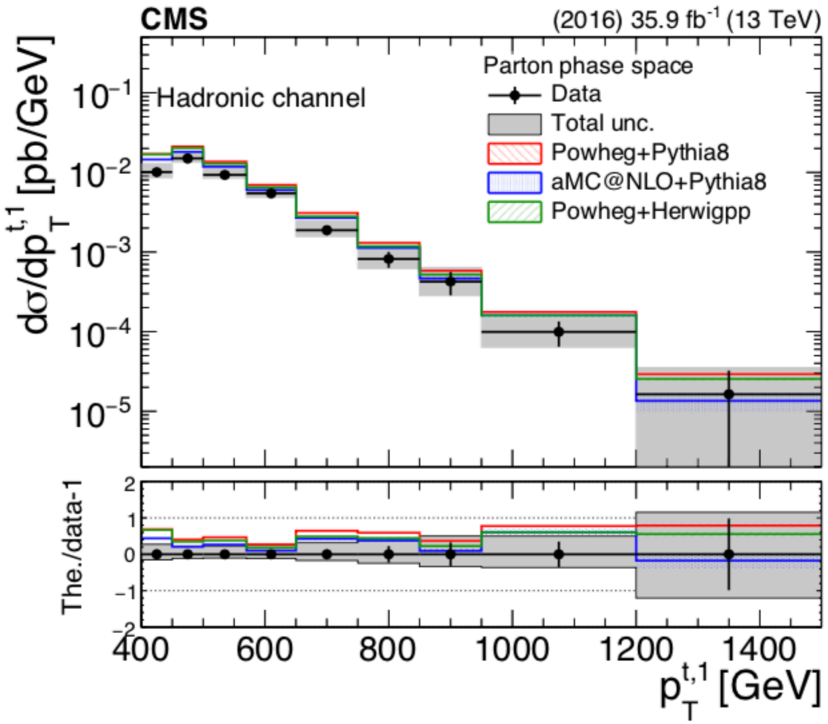
23/9/2020

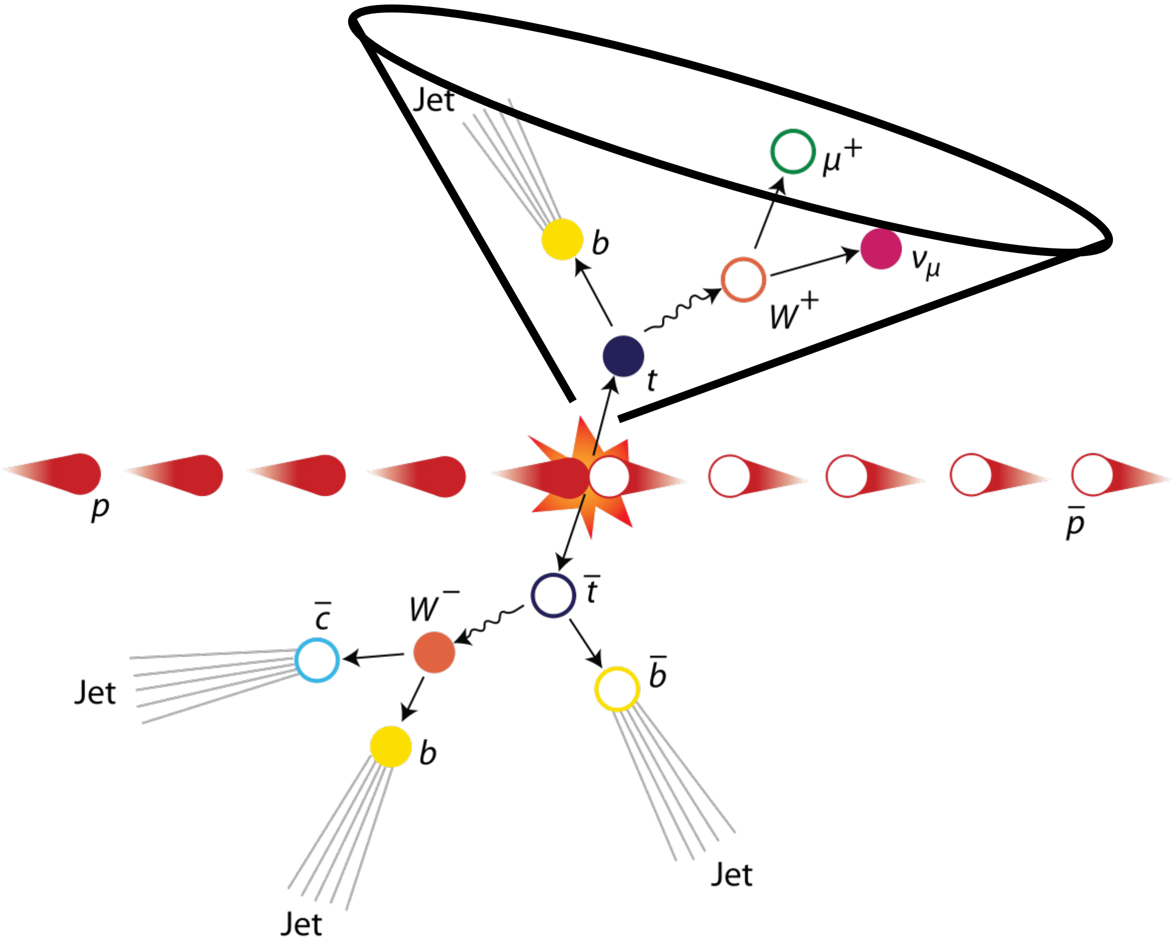
G. Bakas, K. Kousouris, I. Papakrivopoulos, G. Tsipolitis

Motivation

Top, anti-top production in the fully hadronic final state.

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





Overview

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 + $m_{SD}^{jet1,2} \in (50,300)GeV + 2btags$
Extended CR (CR_A) (QCD fit region)	Baseline + topTagger + $m_{SD}^{jet1,2} \in (50,300)GeV + 0btags$

- Variables of interest:
  - ttbar mass, pt, rapidity
  - Leading and Subleading jetPt and |jetY|
  - Nominal ttbar MC samples
  - Mtt samples (700-1000, 1000-Inf) (only for 2016)
- Baseline Parton cuts:
  - Jet Matching
  - partonPt[0],[1] > 400
  - |partonEta[0],[1]| < 2.4
  - mTTbarParton > 1000 GeV
- 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
- Btagging selection:
  - bTagging (medium WP **deepCSV**) (2016: 0.6321, 2017: 0.4941, 2018: 0.4184)
- Top Tagger WP:
  - **New top Tagger:** (2016: 0.2, 2017:0.0, 2018: 0.1)

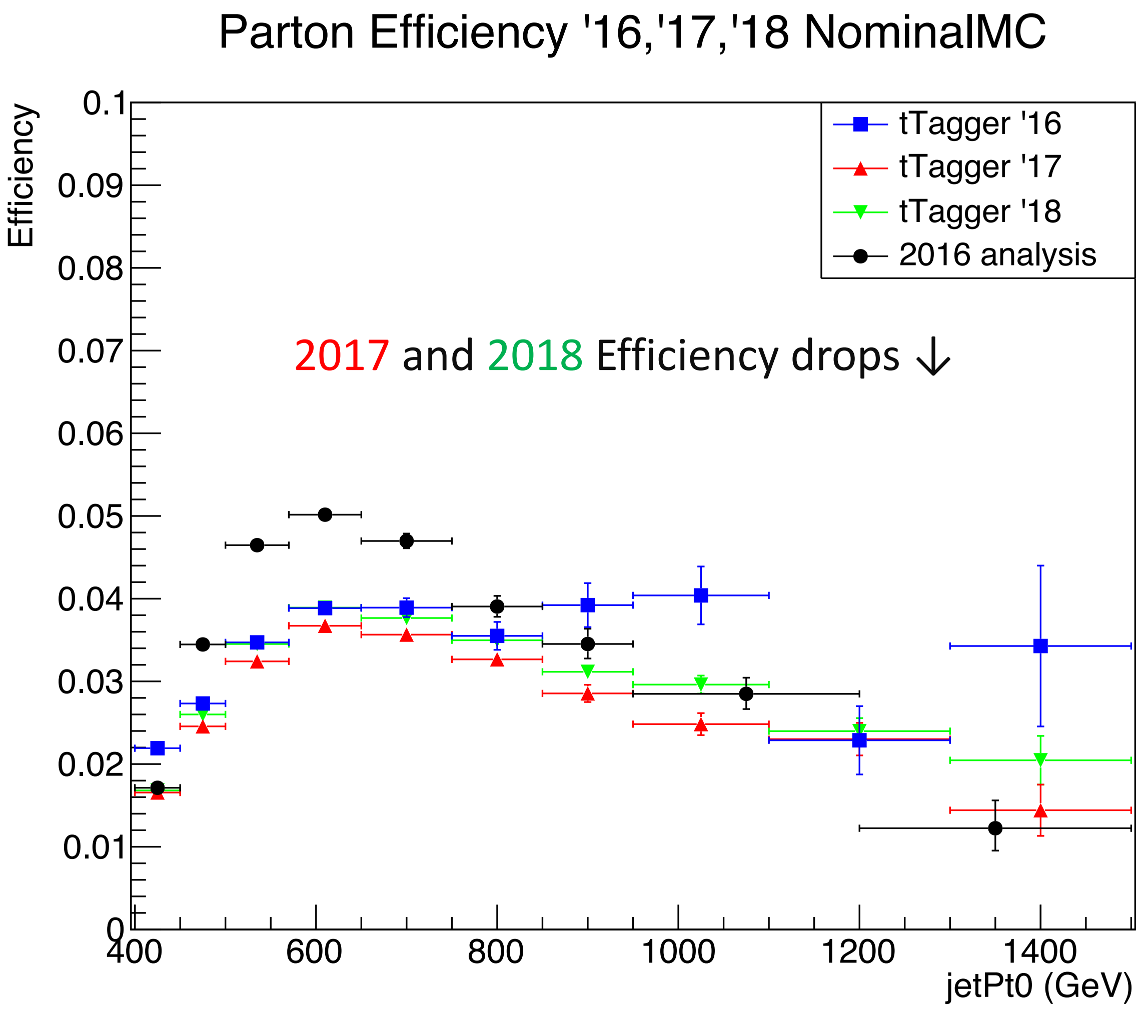
Goal is to Unfold to the Parton And Particle Levels
- Closure Tests with Nominal MC's
- Unfolding and extrapolation with Data

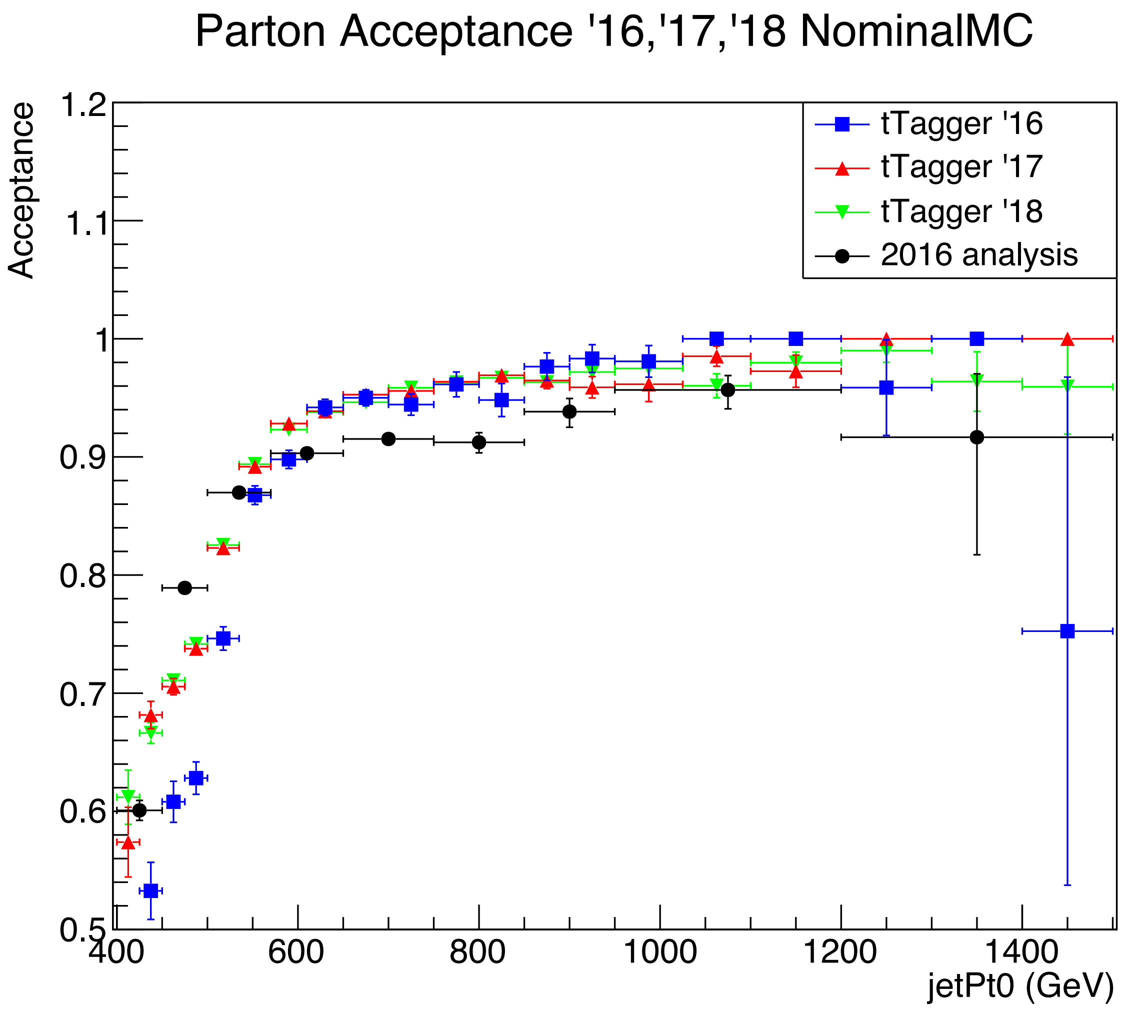
Discovered that our Control Region is contaminated from ttbar and Subdominant bkg:
- Extract this contribution from the Data CR distribution → pure QCD

Differences with TOP-18-013:
- mJJ and mTTbarParton cut at 1000 GeV instead of 800 GeV
- New top tagger, tagging jets and not events. The goal is to have higher efficiency in the far end of the spectrum

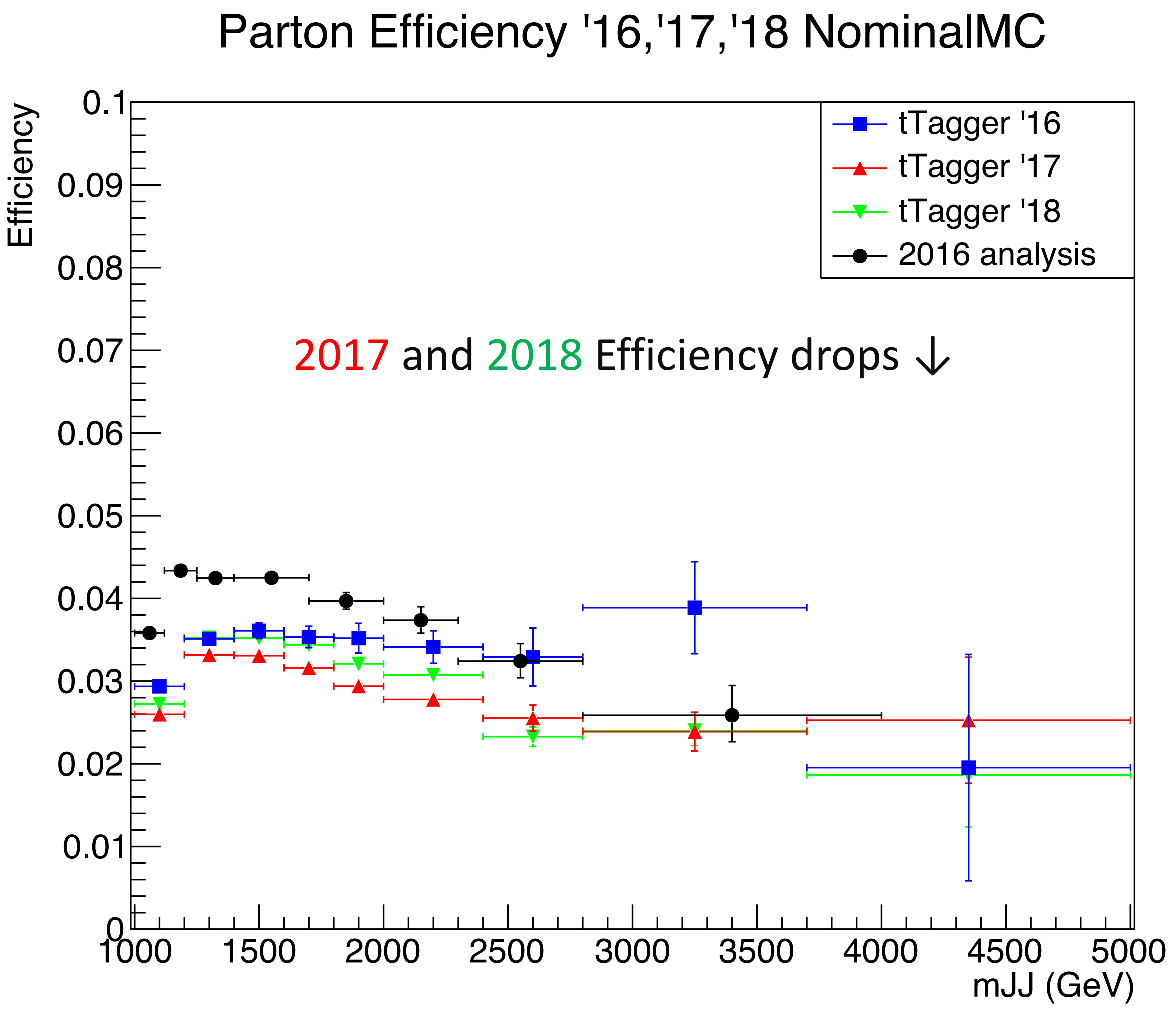
B tagging SF's are applied

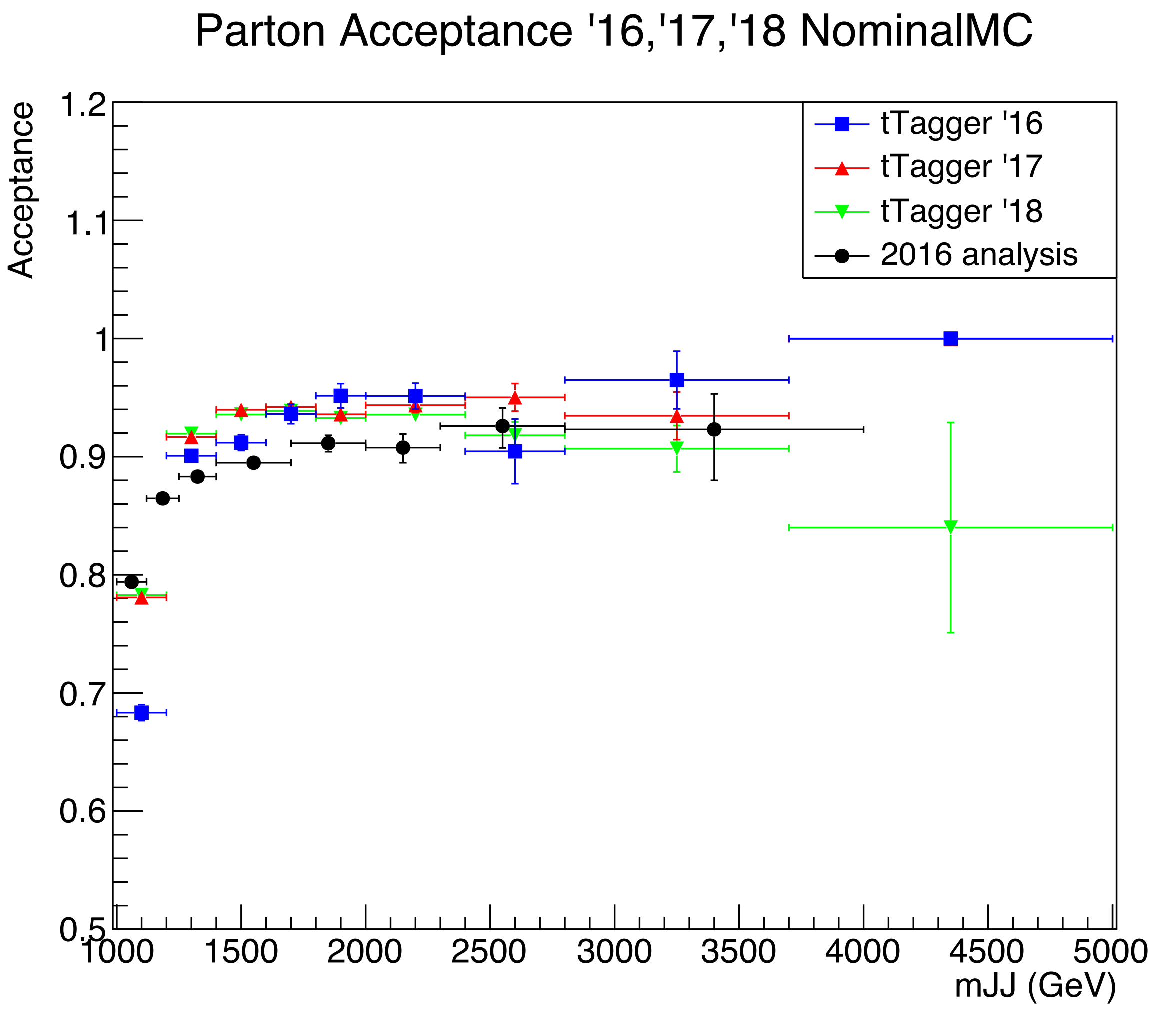
Efficiency and Acceptance Plots





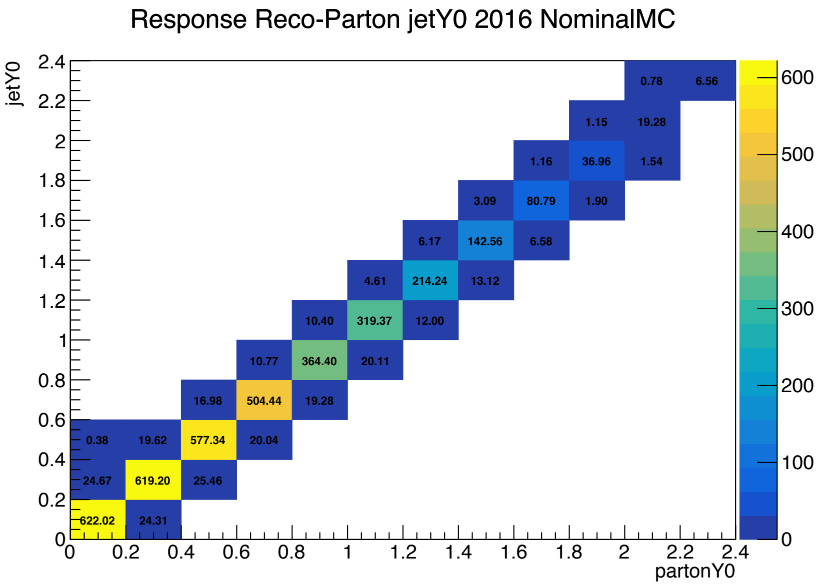
Efficiency and Acceptance Plots



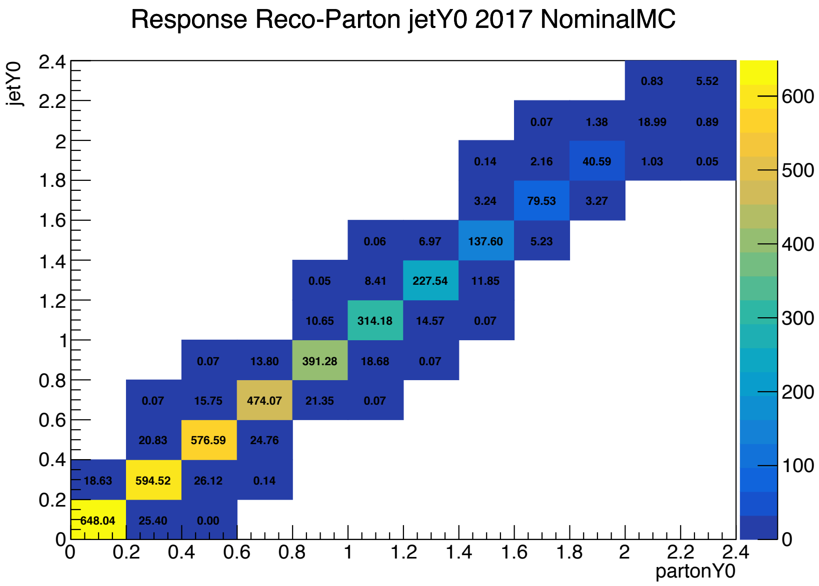


Response Matrices

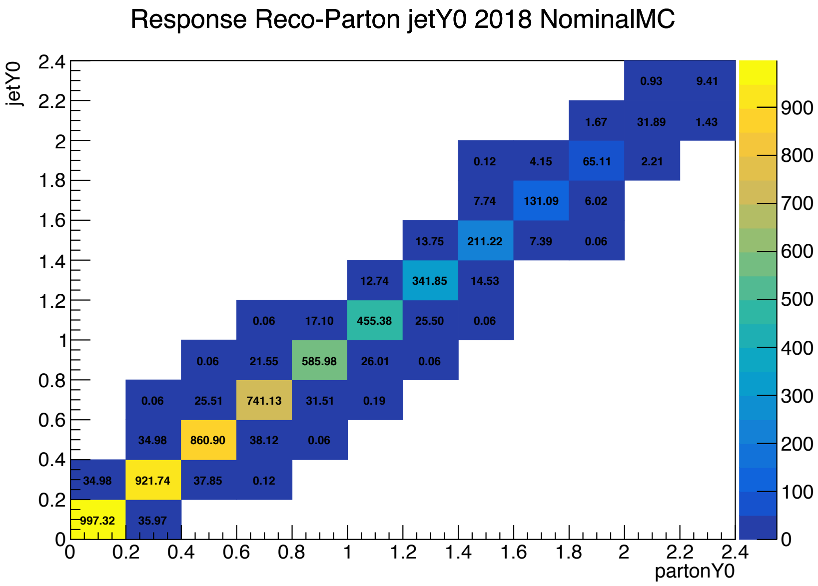
(2016)



(2017)



(2018)

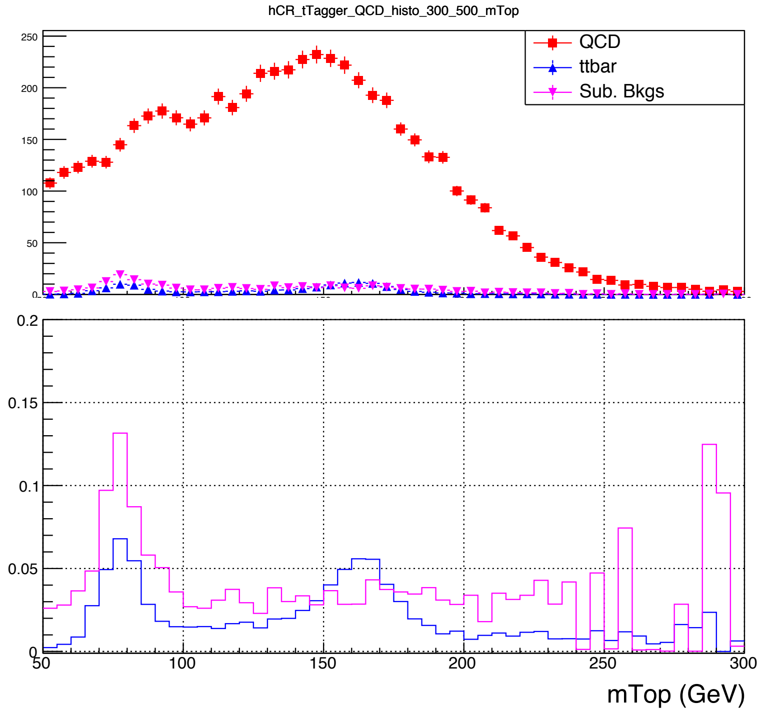


Assuming that response matrices are compatible to each other

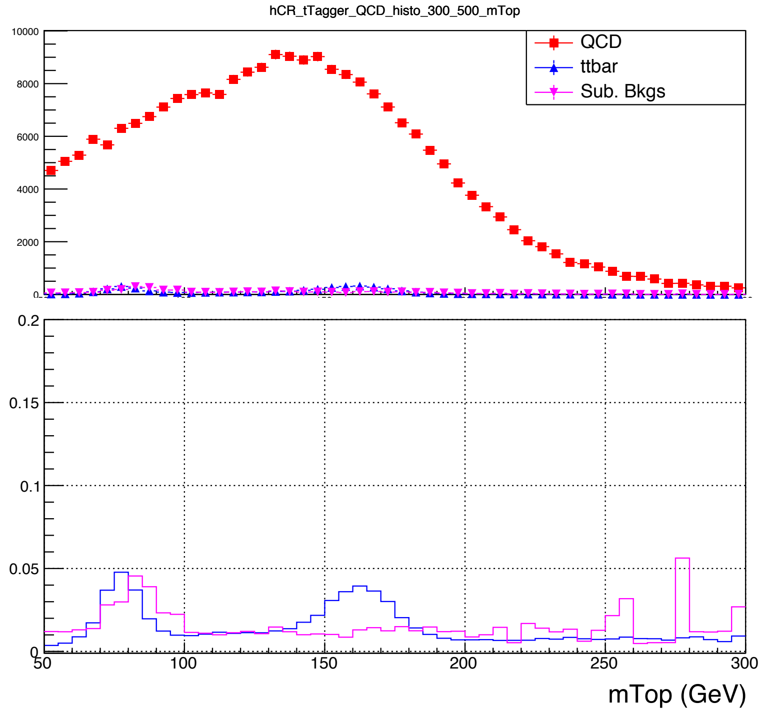
1. Unfold each year and combine results?
2. Combine Fiducial Measurements and then unfold

Contamination Plots Medium WP (CR) 2016, 2017, 2018

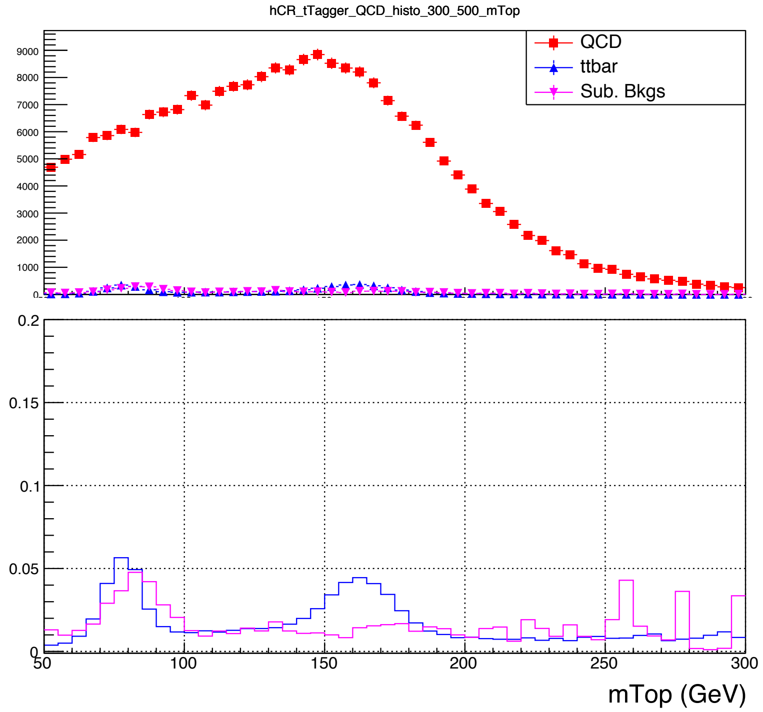
(2016)



(2017)



(2018)



Contamination Plots Medium WP (CR) 2016, 2017, 2018

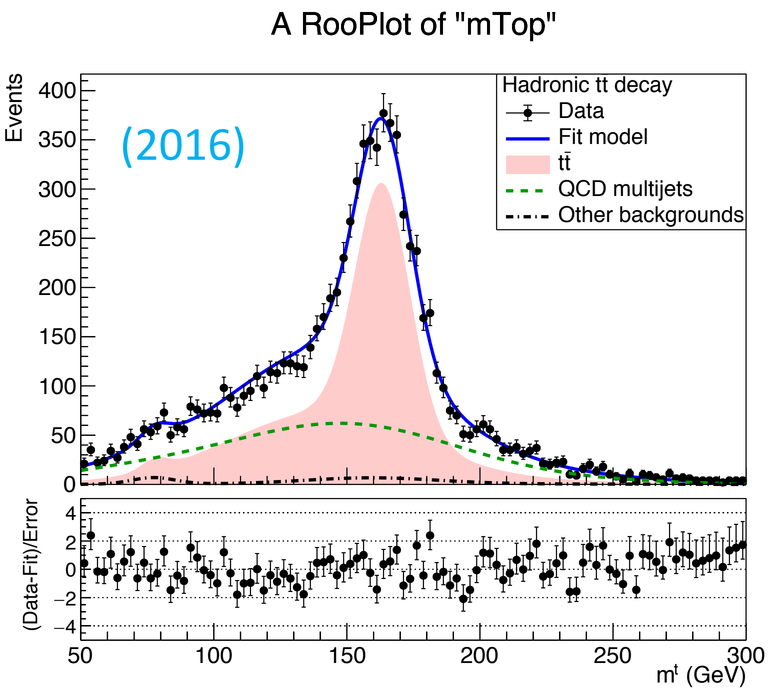
Comment

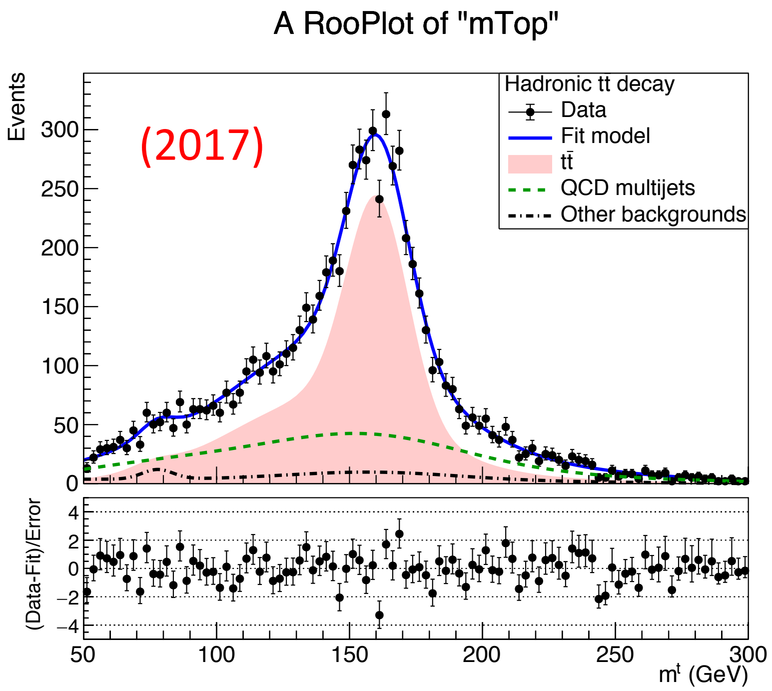
Shapes and in general contamination is NOT affected by btagging by Scale factors

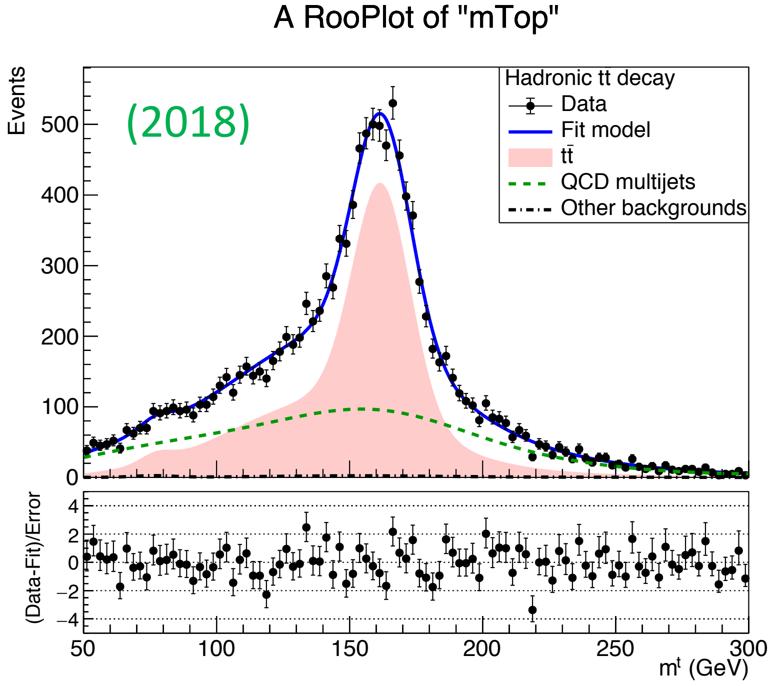
Mass Fit in Extended SR (SR_A)

$$QCD_0(m^t) = D_0(m^t) - T_0(m^t) - Sub_0(m^t)$$

- Both SR and Control Region use the Medium btag WP.
- Intuition is to remove the ttbar and subdominant bkg contribution from the data Control Region







Floating Parameter	FinalValue +/-	Error
kMassResol	9.2251e-01 +/-	2.73e-02
kMassScale	9.9891e-01 +/-	2.01e-03
kQCD_2b	6.9753e-02 +/-	5.26e-02
nFitBkg_2b	2.4472e+02 +/-	1.47e+02
⚠ nFitQCD_2b	2.9890e+03 +/-	1.74e+02
nFitSig2b	5.2763e+03 +/-	1.67e+02

Signal strength: r = 0.686668 ± 0.0263103 (new)

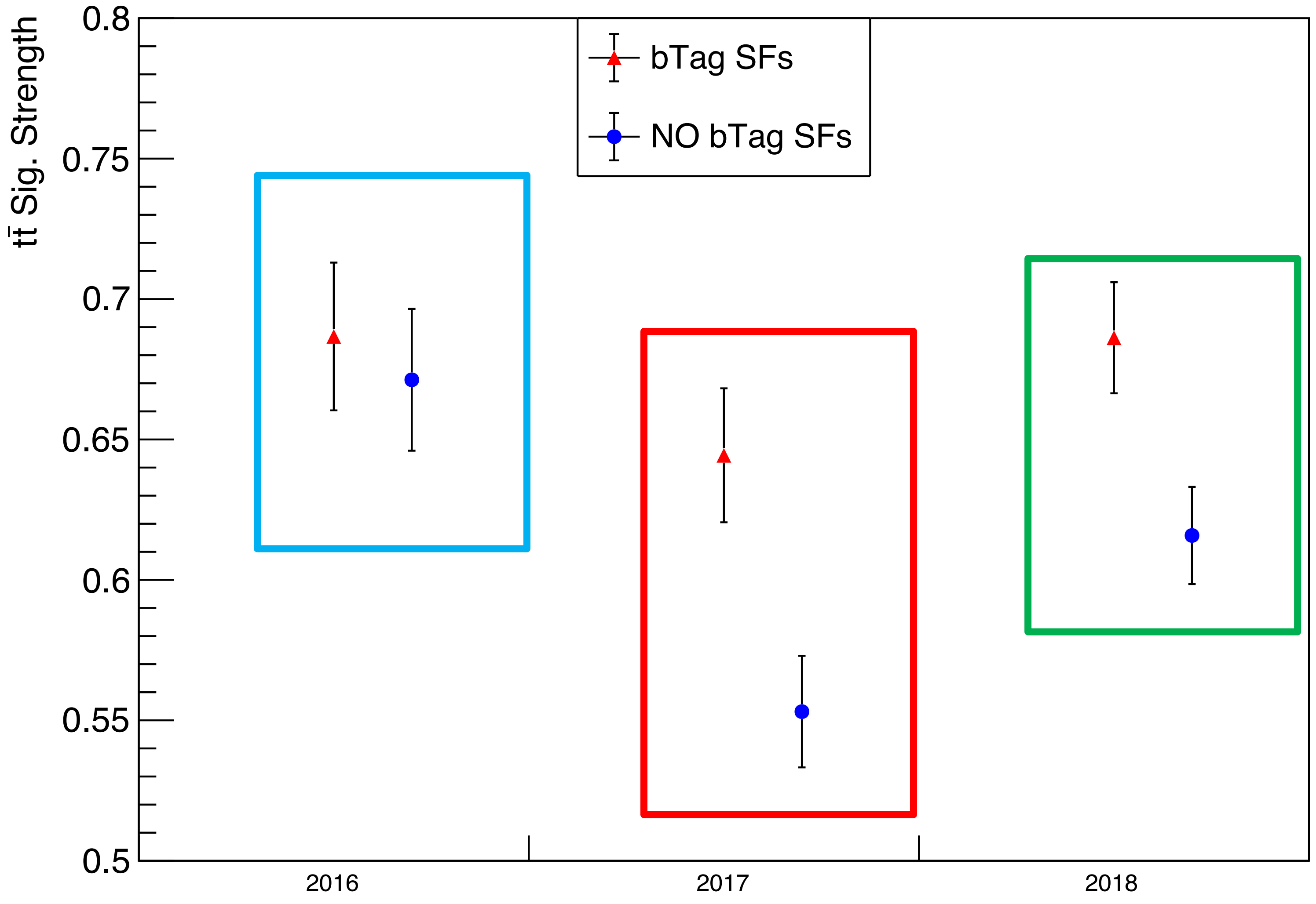
Floating Parameter	FinalValue +/-	Error
kMassResol	1.0998e+00 +/-	4.02e-02
kMassScale	9.8340e-01 +/-	2.66e-03
kQCD_2b	1.6593e-02 +/-	7.44e-03
nFitBkg_2b	4.9791e+02 +/-	2.68e+02
⚠ nFitQCD_2b	2.1662e+03 +/-	3.11e+02
nFitSig2b	4.8059e+03 +/-	1.50e+02

Signal strength: r = 0.644361 ± 0.023851 (new)

Floating Parameter	FinalValue +/-	Error
kMassResol	1.0071e+00 +/-	2.84e-02
kMassScale	9.9006e-01 +/-	1.92e-03
kQCD_2b	1.2862e-02 +/-	2.71e-03
nFitBkg_2b	1.0737e+02 +/-	2.97e+02
⚠ nFitQCD_2b	5.0444e+03 +/-	2.83e+0[illegible]
nFitSig2b	7.6593e+03 +/-	1.86e+0[illegible]

Signal strength: r = 0.683382 ± 0.0200866

Signal Strength Results



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)

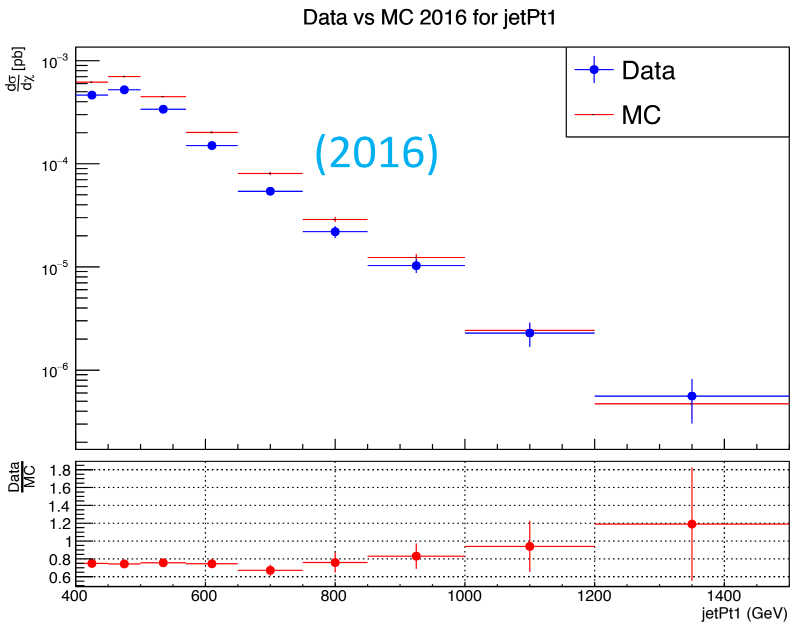
- Where x_{reco} is the respected variable of interest (ttbar mass, pt, rapidity, leading and subleading jetPt and |jetY|)

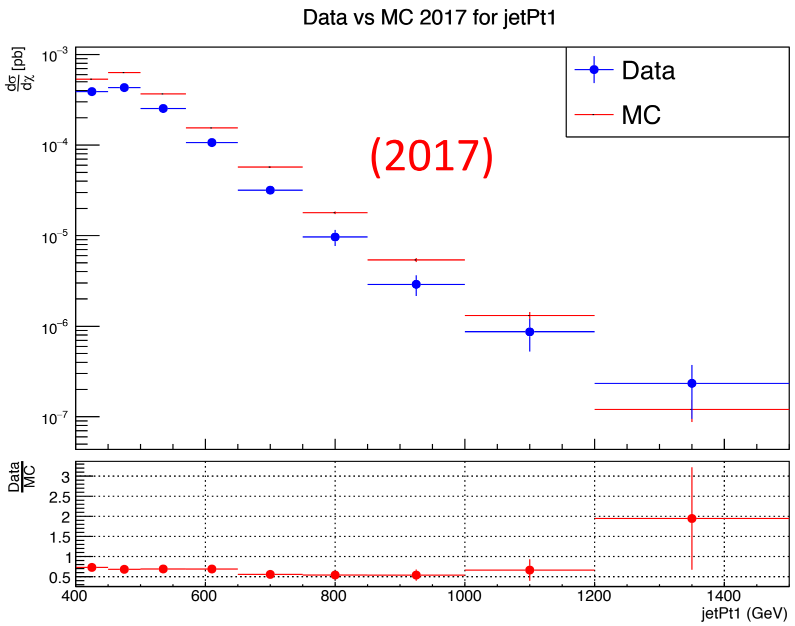
- We deploy a fit in the Signal Region (2btag) to extract the N_{QCD}^{fit}

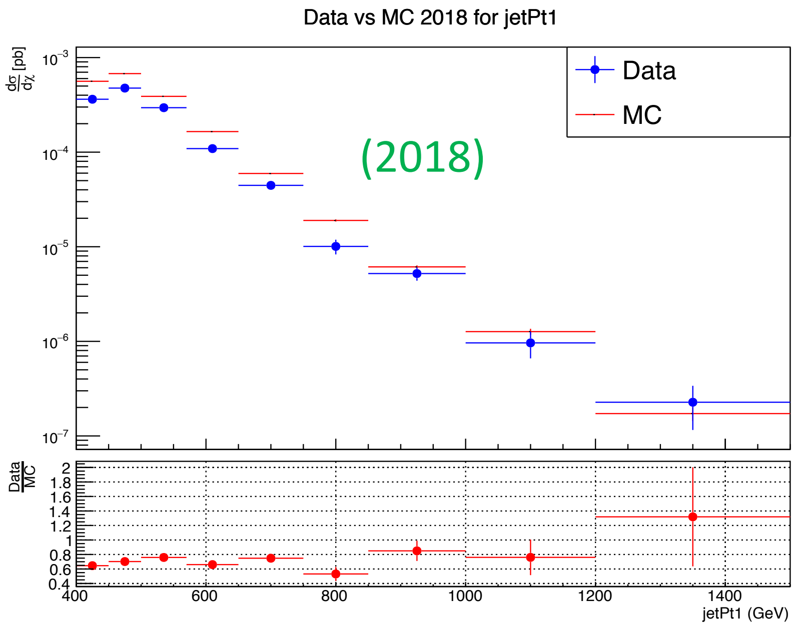
$$D(m^t)^{(i)} = N_{tt}^{(i)} T^{(i)}(m^t, k_{MassScale}, k_{MassResolution}) + N_{bkg}^{(i)} B(m^t)(1 + k_1 x) + N_{sub}^{(i)} O^{(i)}(m^t)$$

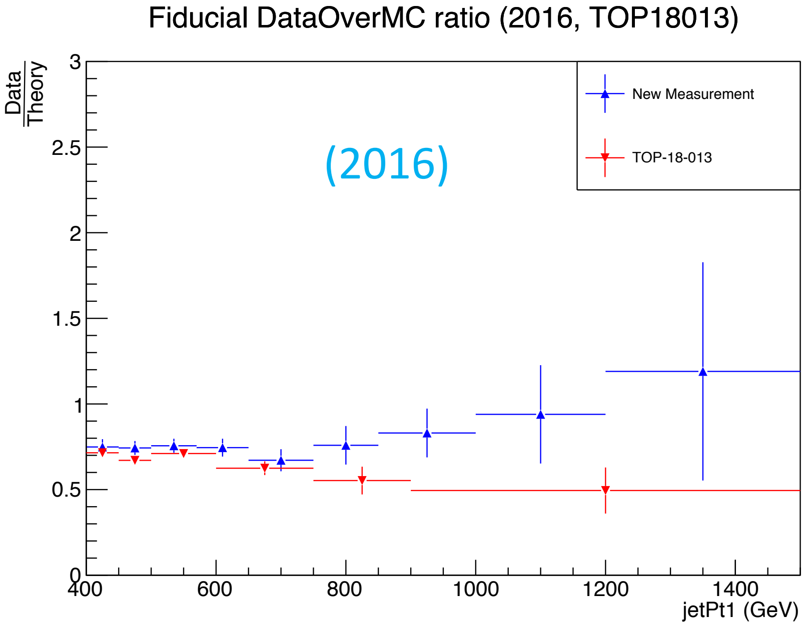
- Our data CR is contaminated from ttbar and subdominant bkgs which has to be dealt with.

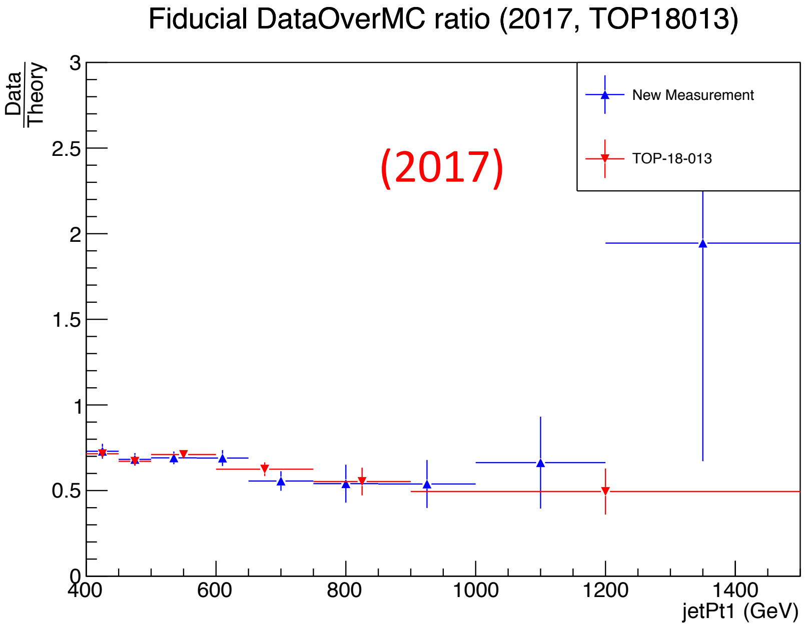
Fiducial Differential Cross Section

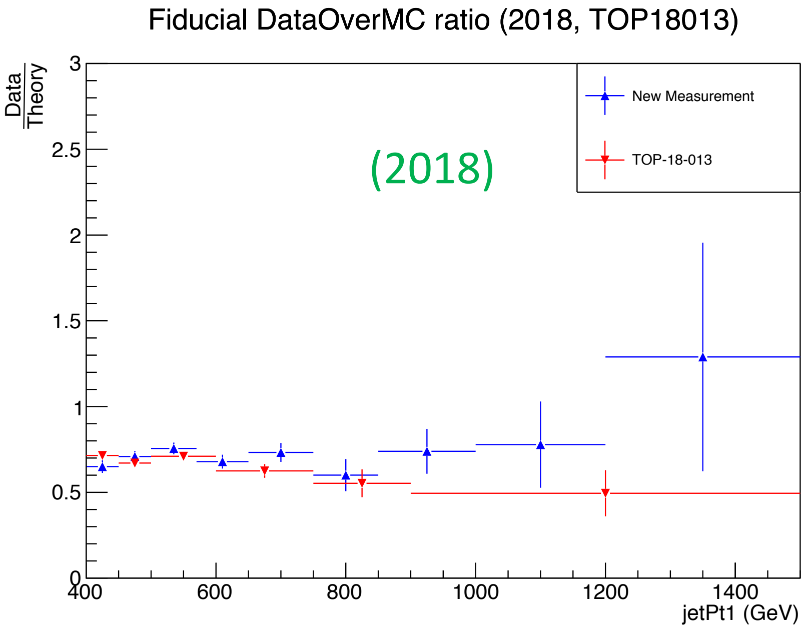




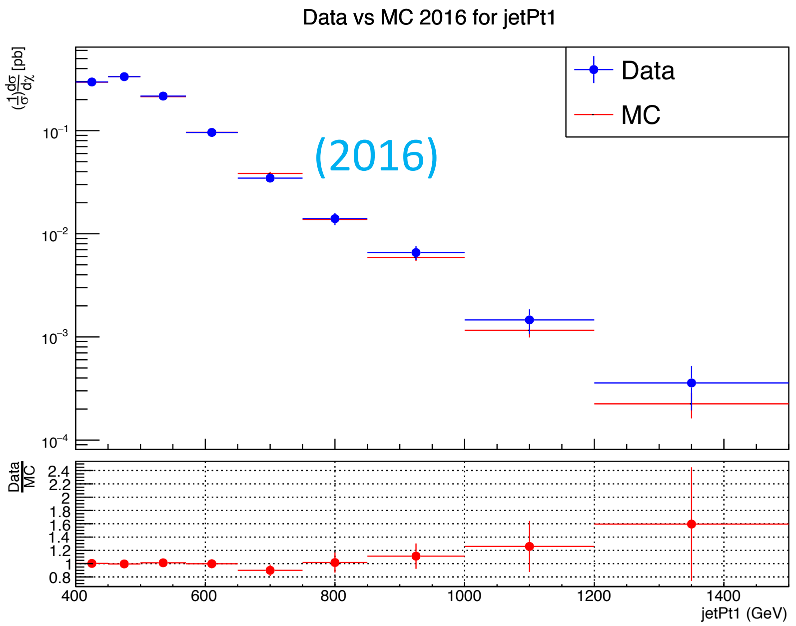


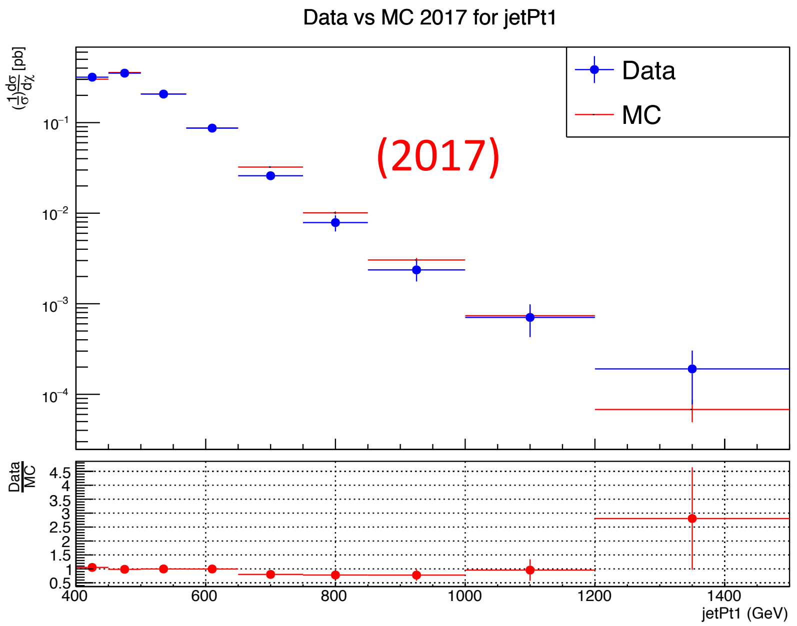


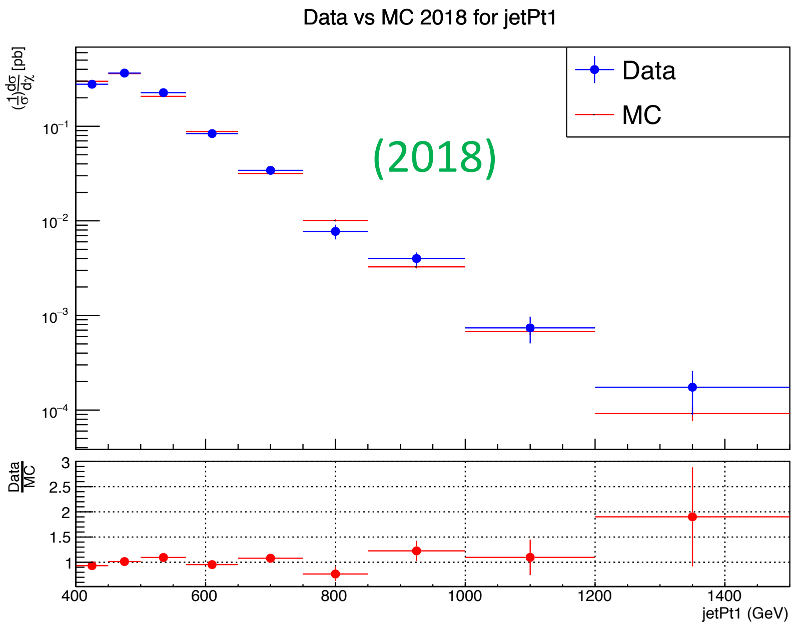


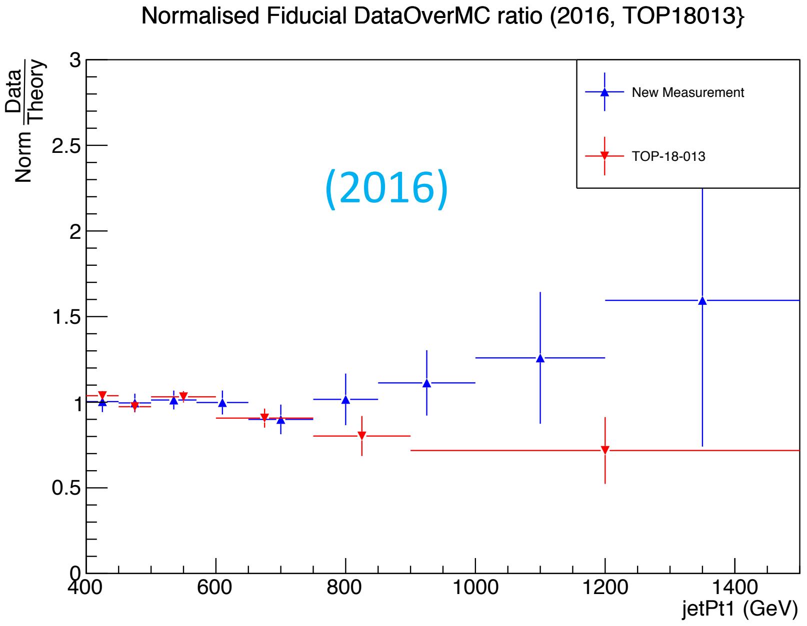


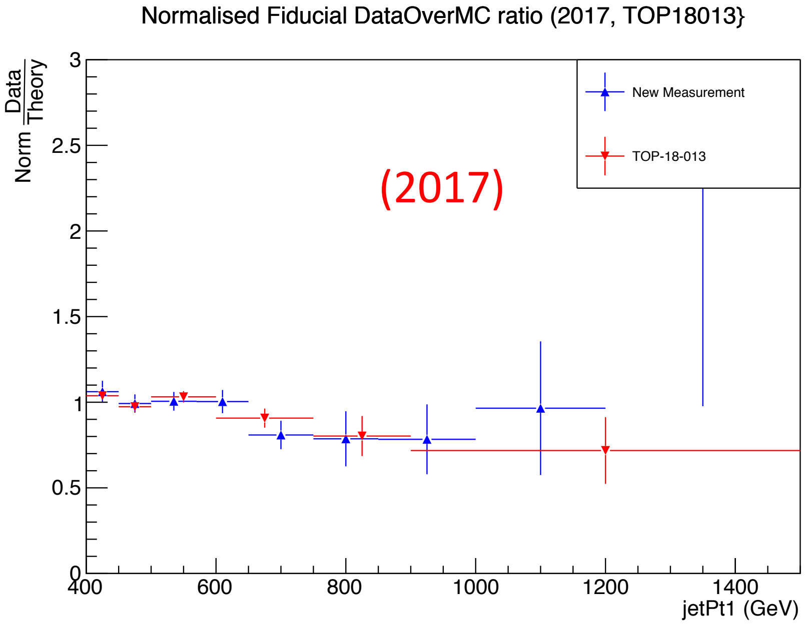
Fiducial Differential Cross Section (Normalized)

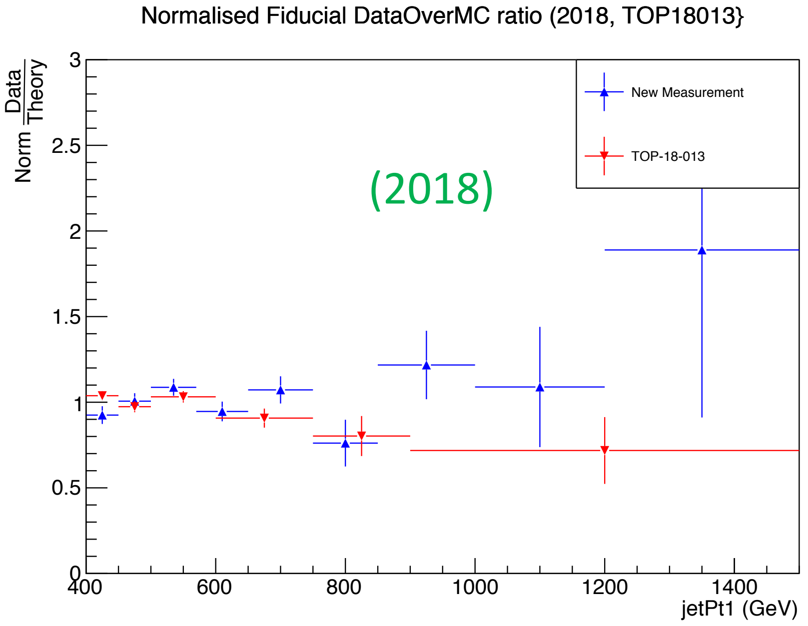








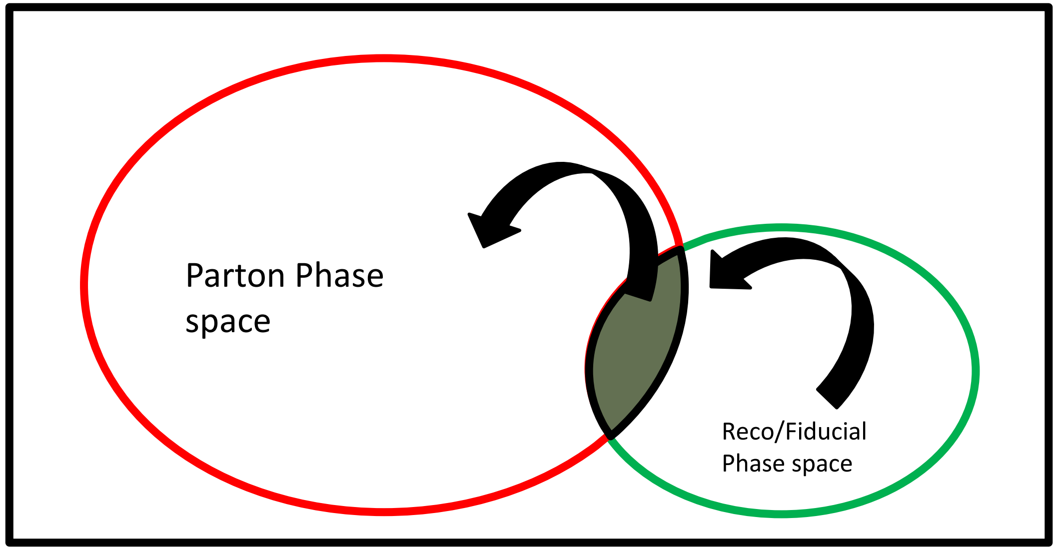




Parton & Particle levels

Parton

Observable	Requirement
$p_T^{t,\bar{t}}$	> 400 GeV
$\lvert\eta^{t,\bar{t}}\rvert$	< 2.4
$m_{t\bar{t}}$	> 1000 GeV



Particle level Top Candidates

Observable	Requirement
N_{jets}	>1
$p_T^{jet1,2}$	> 400 GeV
$\lvert\eta^{jet1,2}\rvert$	< 2.4
$m_{SD}^{jet1,2}$	(120, 220) GeV
m_{jj}	> 1000 GeV

$$\frac{d\sigma_i^{\text{unf}}}{dx} = \frac{1}{\mathcal{L}\cdot\Delta x_i}\cdot\frac{1}{f_{2,i}}\cdot\sum_j\left(R_{ij}^{-1}\cdot f_{1,j}\cdot S_j\right)$$

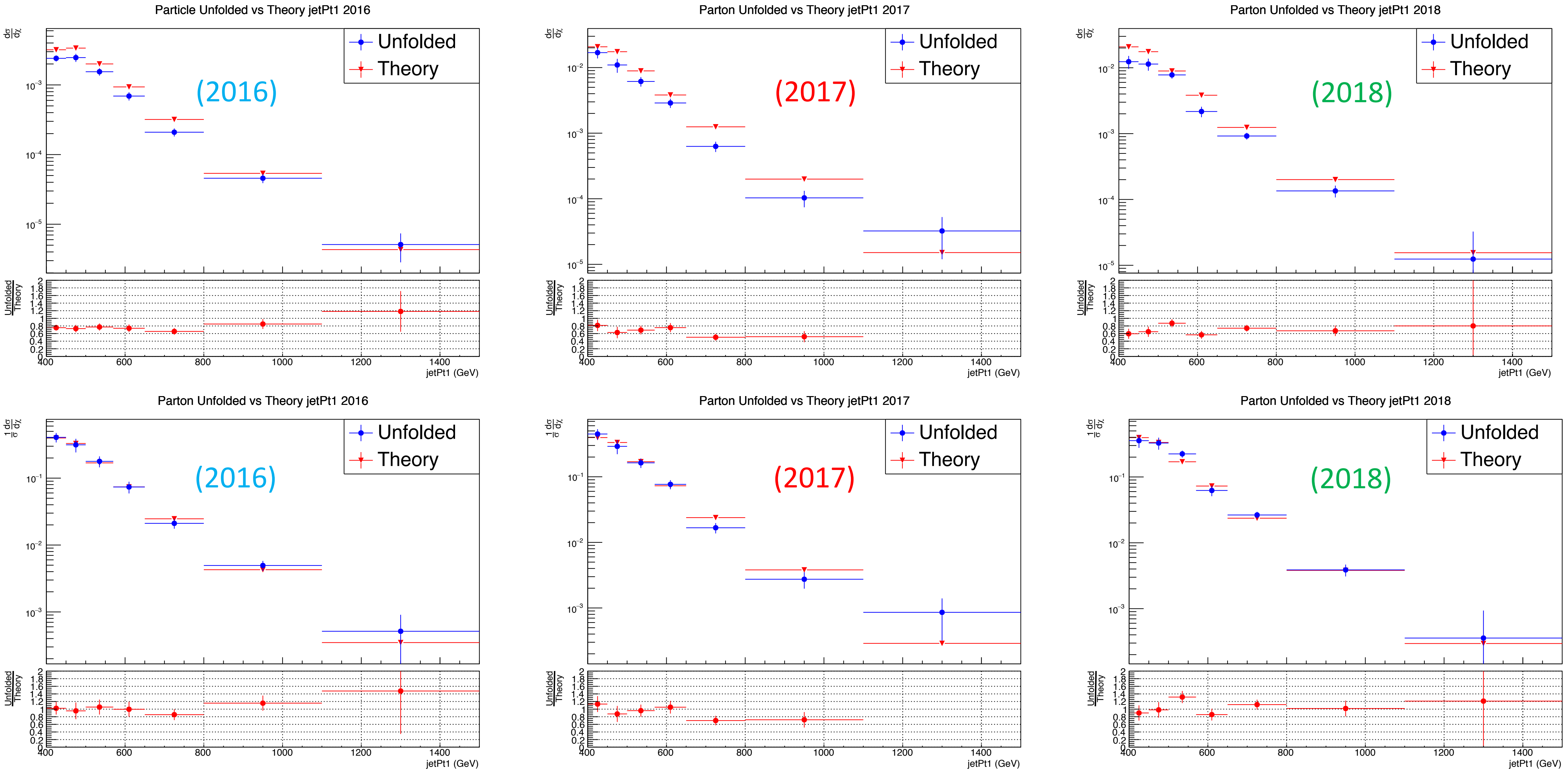
efficiency of the reco+true selection

migration matrix

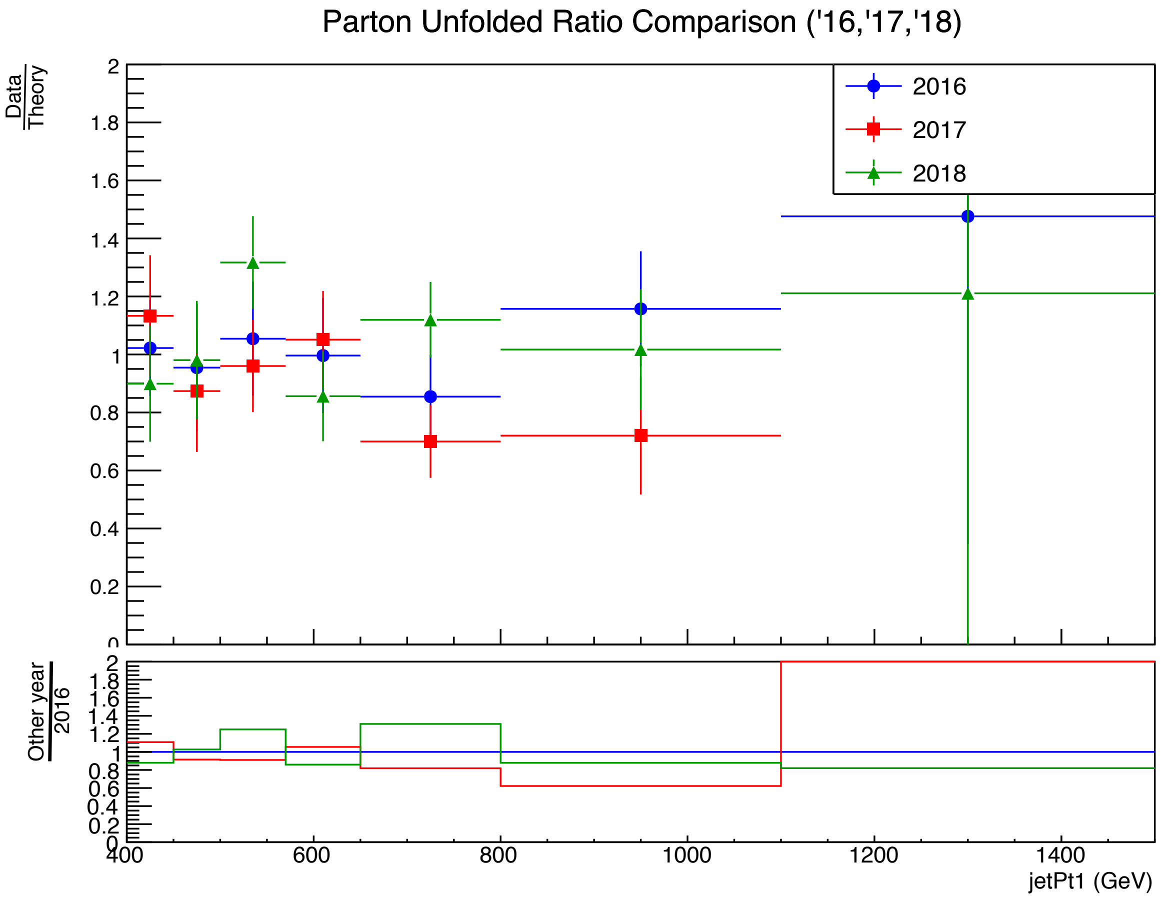
reco efficiency of the reco+true selection

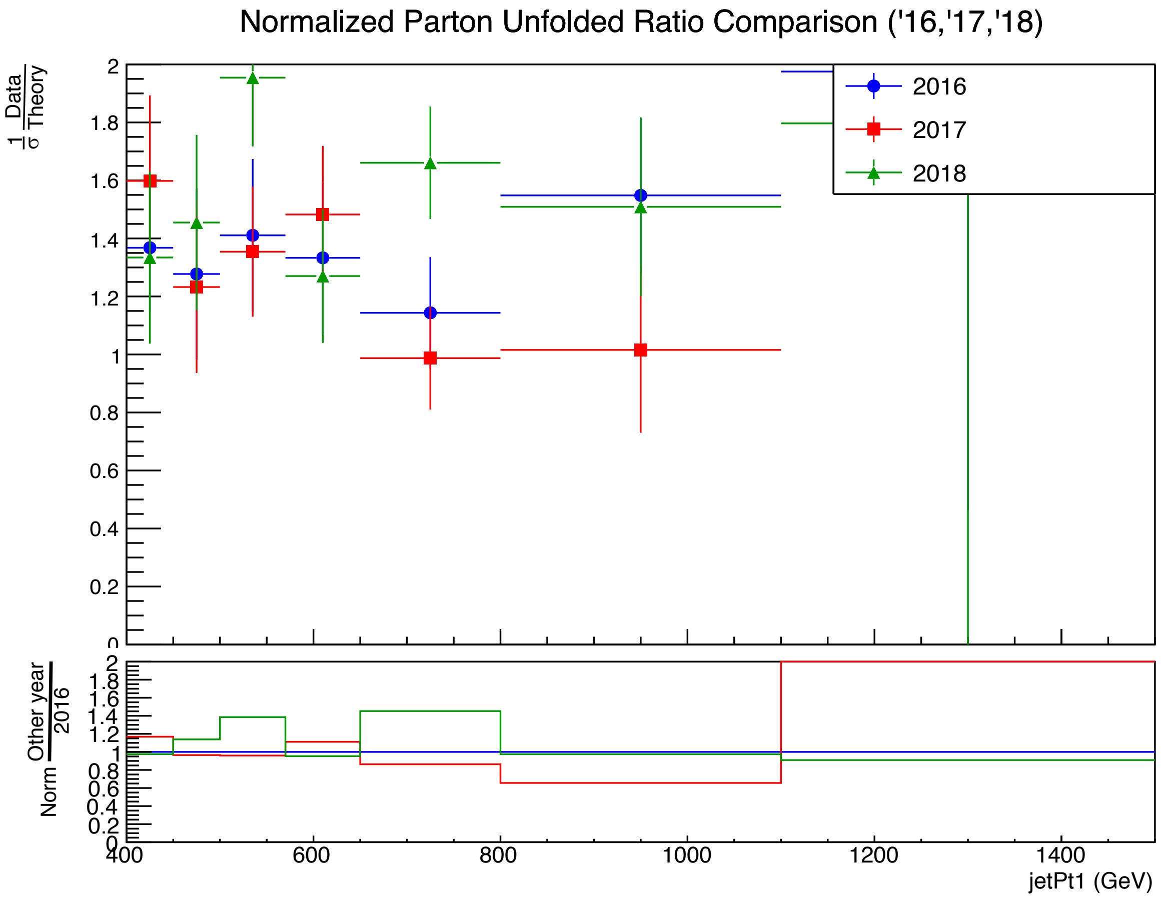
Unfolding: simple response matrix inversion w/o regularisation

Parton Differential Cross Section



Parton Differential Cross Section Comparison





Tag And Probe

- Top Tagger Scale Factors
  - Validation method to ensure that no SF’s are needed
  - From data we subtract QCD and Subdominant bkgs (MC) so that the data sample is pure

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

  - Randomization: Randomly select leading/subleading jet to use as tag or probe to avoid pT bias
  - Divide the phase space into pT regions based on the topTagger categories: [400-600] GeV, [600-800] GeV, [800-Inf] GeV

Tag And Probe Calculations

(2016)

eff data: 0.781 ± 0.038
eff ttbar: 0.772 ± 0.014

Efficiency per Pt region
eff data pT[400-600]: 0.761 ± 0.042
eff ttbar pT[400-600]: 0.778 ± 0.016
----
eff data pT[600-800]: 0.851 ± 0.100
eff ttbar pT[600-800]: 0.748 ± 0.031
----
eff data pT[800-Inf]: 0.886 ± 0.160
eff ttbar pT[800-Inf]: 0.775 ± 0.063

(2017)

eff data: 0.857 ± 0.040
eff ttbar: 0.875 ± 0.0072

Efficiency per Pt region
eff data pT[400-600]: 0.872 ± 0.047
eff ttbar pT[400-600]: 0.874 ± 0.008
----
eff data pT[600-800]: 0.795 ± 0.088
eff ttbar pT[600-800]: 0.876 ± 0.018
----
eff data pT[800-Inf]: 0.797 ± 0.186
eff ttbar pT[800-Inf]: 0.899 ± 0.045

(2018)

eff data: 0.798 ± 0.034
eff ttbar: 0.839 ± 0.005

Efficiency per Pt region
eff data pT[400-600]: 0.793 ± 0.04
eff ttbar pT[400-600]: 0.836 ± 0.006
----
eff data pT[600-800]: 0.829 ± 0.066
eff ttbar pT[600-800]: 0.851 ± 0.013
----
eff data pT[800-Inf]: 0.752 ± 0.13
eff ttbar pT[800-Inf]: 0.865 ± 0.032

Summary

- Applied the b-tagging SF's
  - Significant effect on the ttbar signal strength on 2017 and 2018

- We have extracted the first results of the cross section in the Fiducial and the parton/particle levels

- Start investigating ttbar Systematic Uncertainties

- 2016 Nominal MC:

Tune CUETP8M2T4	TuneCP5
/TT_TuneCUETP8M2T4_13TeV-powheg-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM	/TTToHadronic_TuneCP5_PSweights_13TeV-powheg-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM (TTToSemiLeptonic, TTTo2L2Nu)

- Investigate on how to combine the measurements between the three years
  - Combine them in the fiducial level and extract the cross section?
  - Extract the cross sections individually and combine the measurements in the unfolded level?

BACKUP

Signal Selection

Variables	Selected Cut
pT (both 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, 0, 0.1
B tagging (2 btagged jets)	> Medium WP
Signal Trigger	

Control Region Selection

Variables	Selected Cut
pT (both 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, 0, 0.1
B tagging (0 btagged jets)	< Medium WP
Control Trigger	

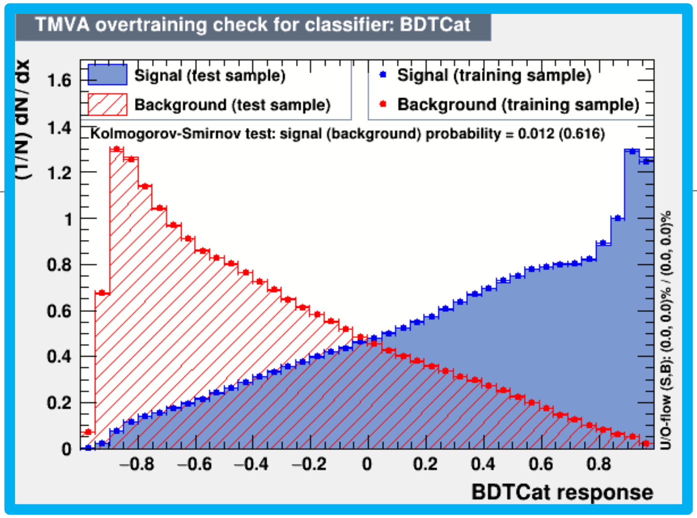
Year	Type of File	DAS
2016	TT Mtt 700-1000	/TT_Mtt-700to1000_TuneCUETP8M2T4_13TeV-powheg-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM
	TT Mtt 1000-Inf	/TT_Mtt-1000toInf_TuneCUETP8M2T4_13TeV-powheg-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v2/MINIAODSIM
	TT Nominal	/TT_TuneCUETP8M2T4_13TeV-powheg-pythia8/RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3-v1/MINIAODSIM
2017	TT Mtt 700-1000	--
	TT Mtt 1000-Inf	--
	TT Nominal Hadronic	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v1/MINIAODSIM
	TT Nominal Semilepton	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v2/MINIAODSIM
	TT Nominal Dilepton	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14-v2/MINIAODSIM
2018	TT Mtt 700-1000	--
	TT Mtt 1000-Inf	--
	TT Nominal Hadronic	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	TT Nominal Semilepton	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM
	TT Nominal Dilepton	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15-v1/MINIAODSIM

Overview: Discriminator, Efficiency and Acceptance

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

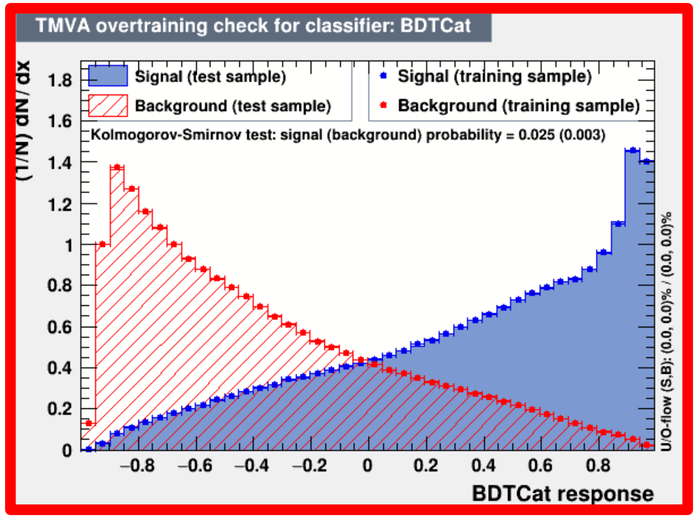
Category training: split the sample in categories based on Pt

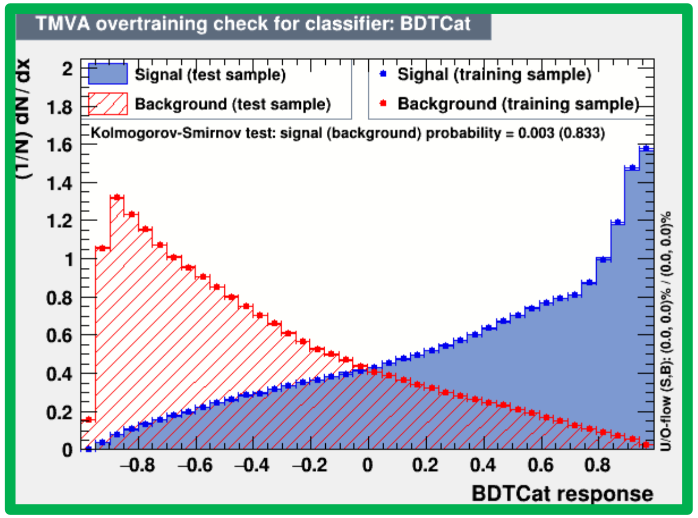
- Bins:
  - [400, 600] GeV
  - [600, 800] GeV
  - [800, 1200] GeV
  - [1200, inf) GeV
- BDT, used variables:
  - Leading and Sub-leading subjet mass
  - N-Subjetiness variables (tau1, tau2, tau3)
  - fraction of the jetPt over the total pt sum of the event.
  - Energy correlation functions (ecfB1N2,ecfB1N3,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

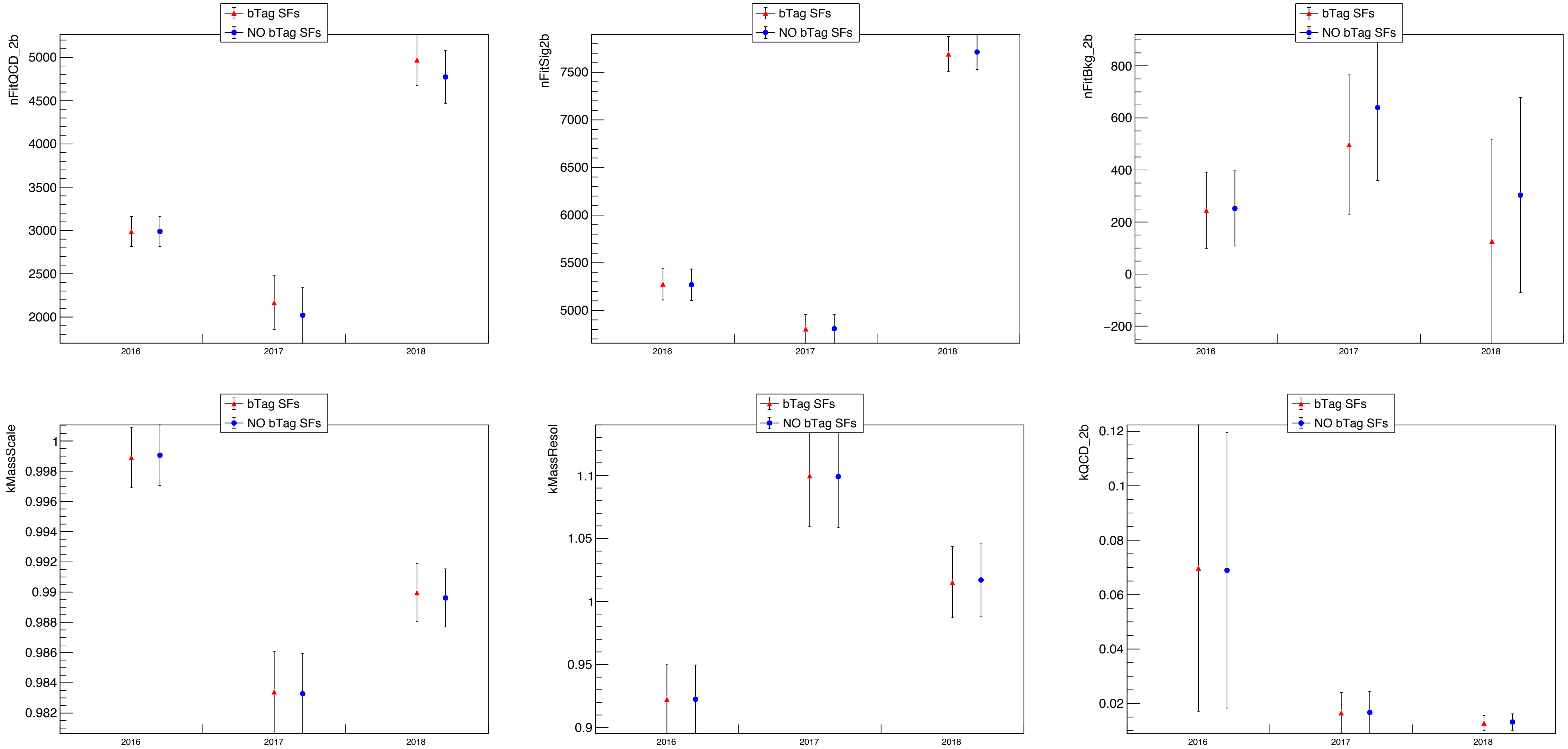
2017





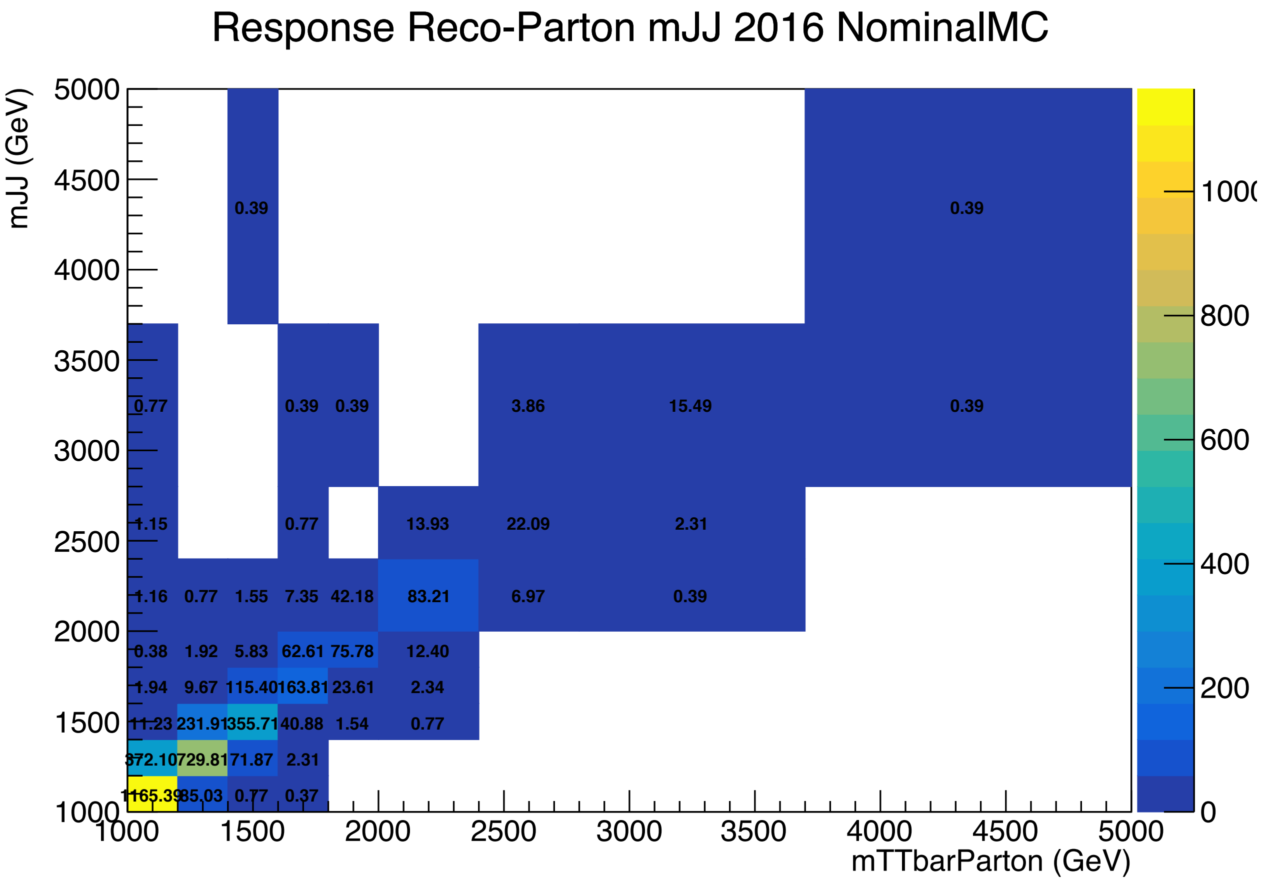
2018

Fit Params Results Comparison

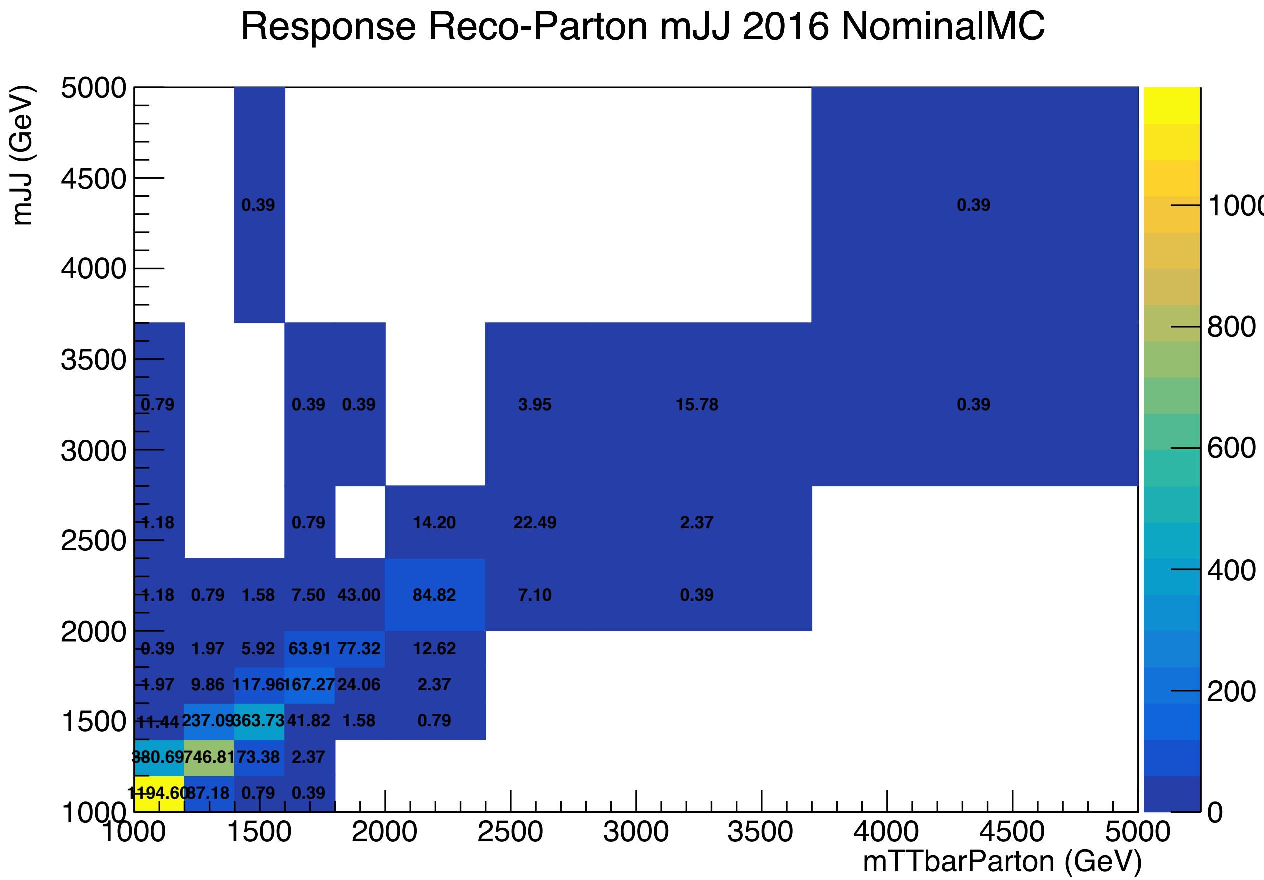


Response Matrices 2016

b tagging SF's

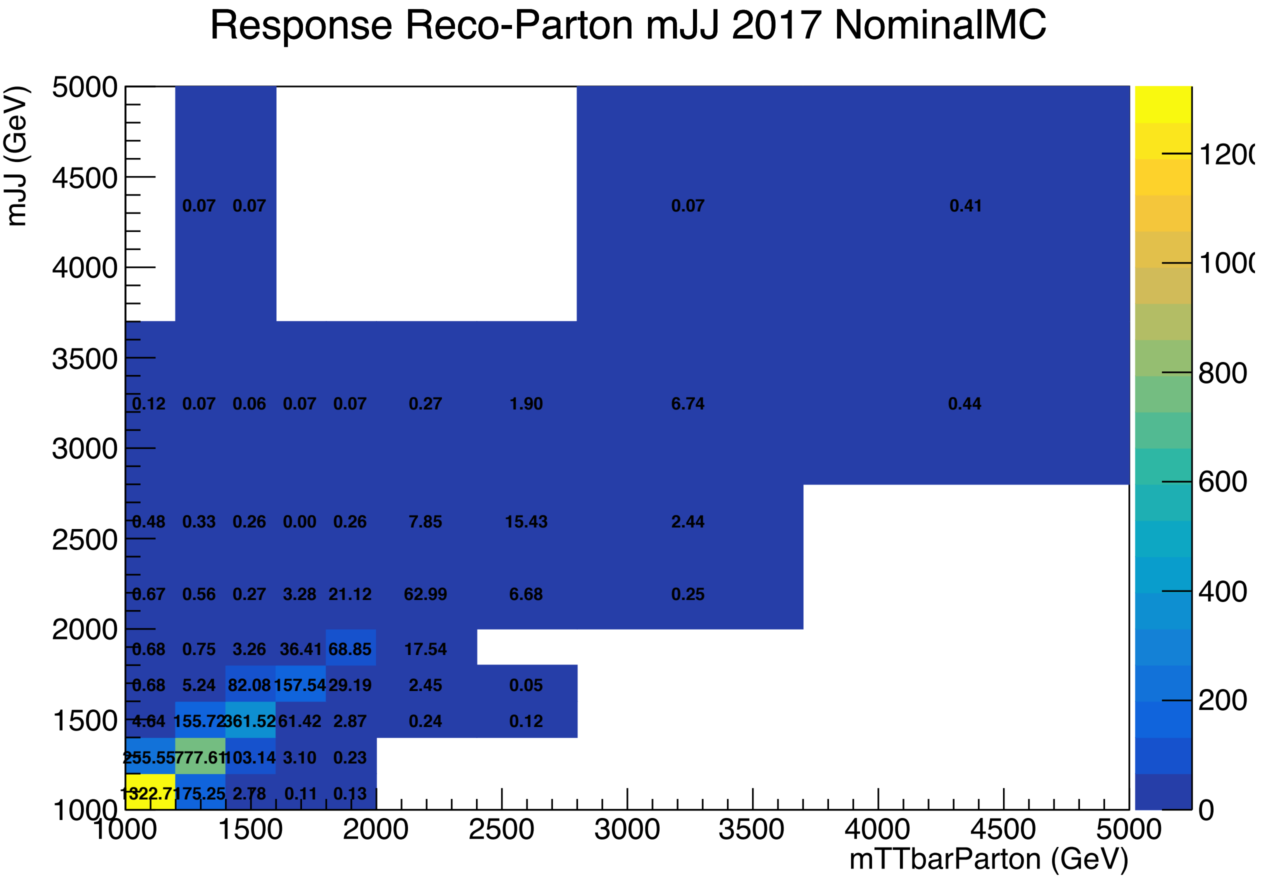


without b tagging SF's

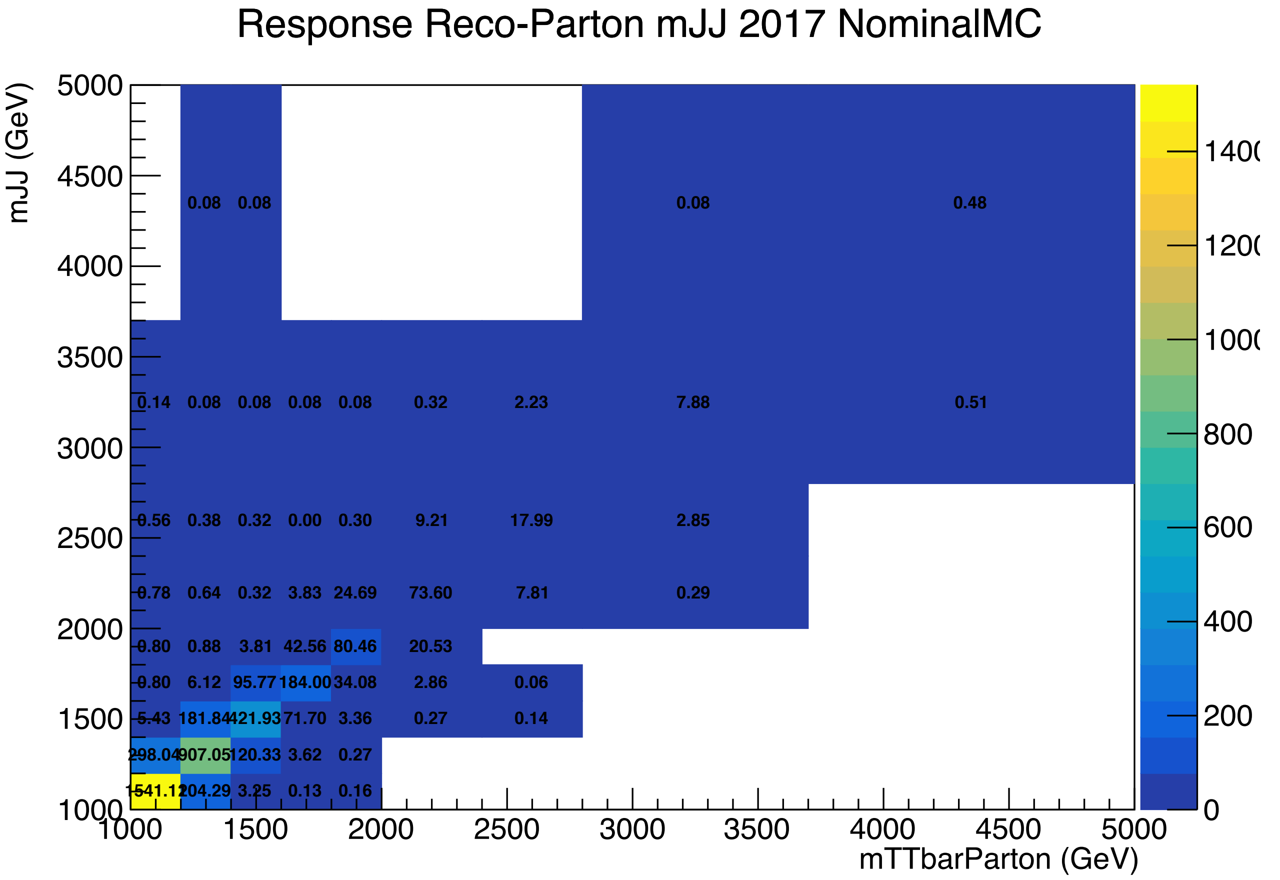


Response Matrices 2017

b tagging SF's

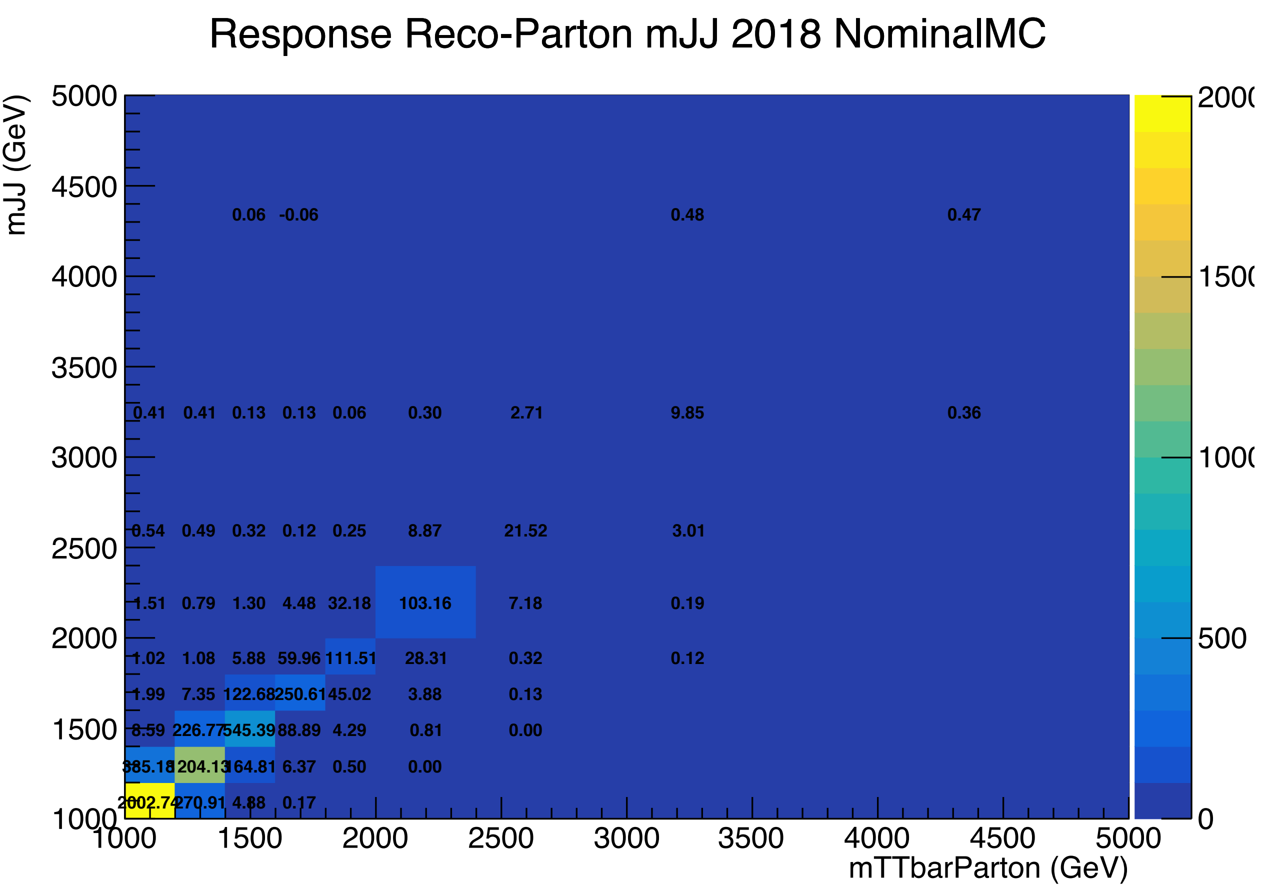


without b tagging SF's

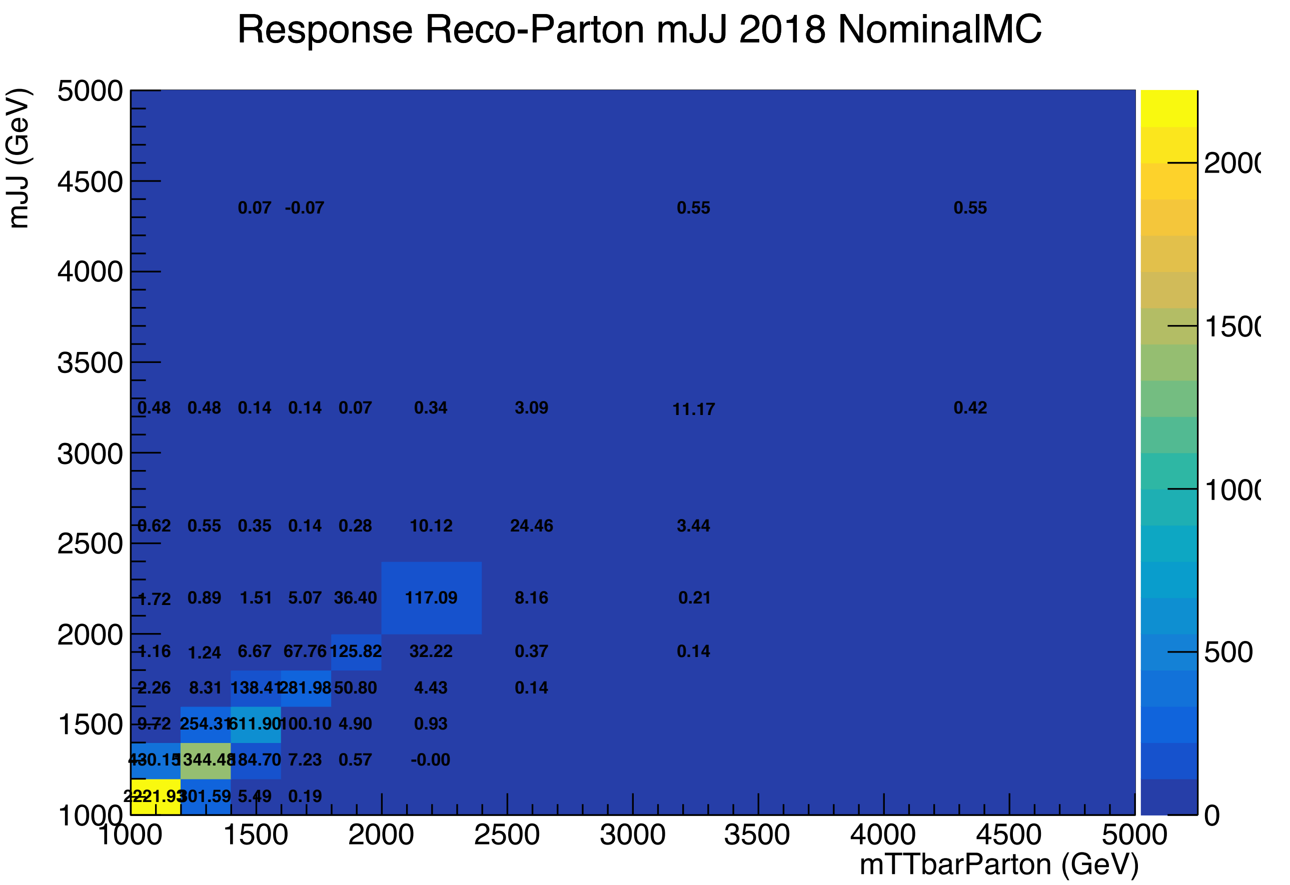


Response Matrices 2018

b tagging SF’s

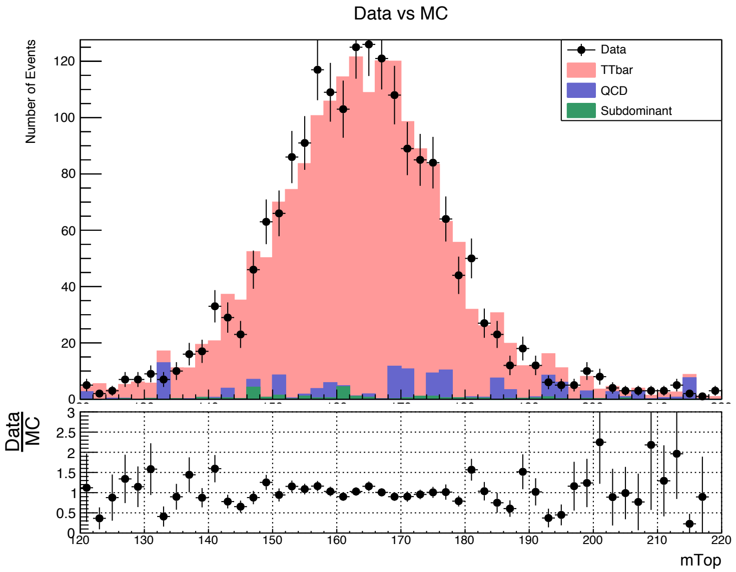


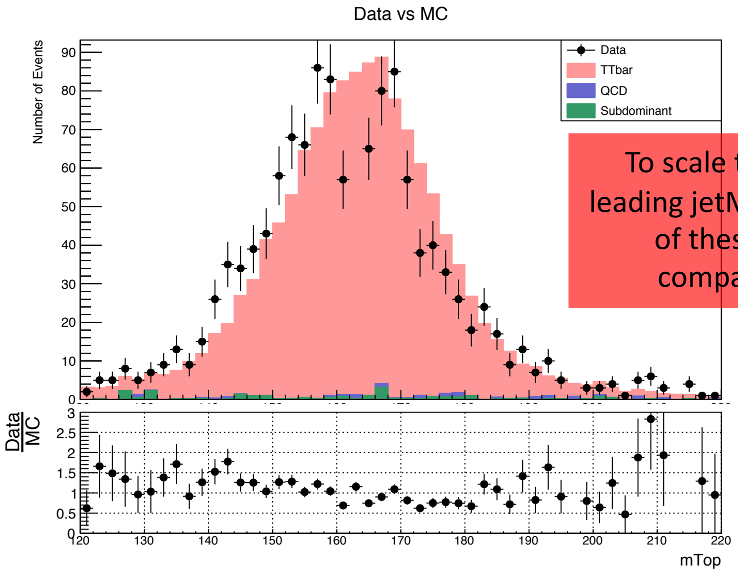
without b tagging SF’s

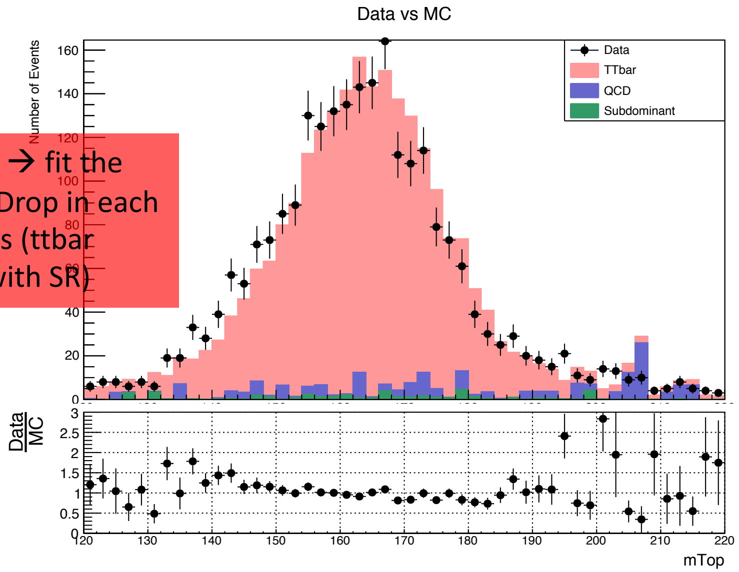


TagAndProbe Efficiency Plots

Tight TopTagger + SR TopTagger

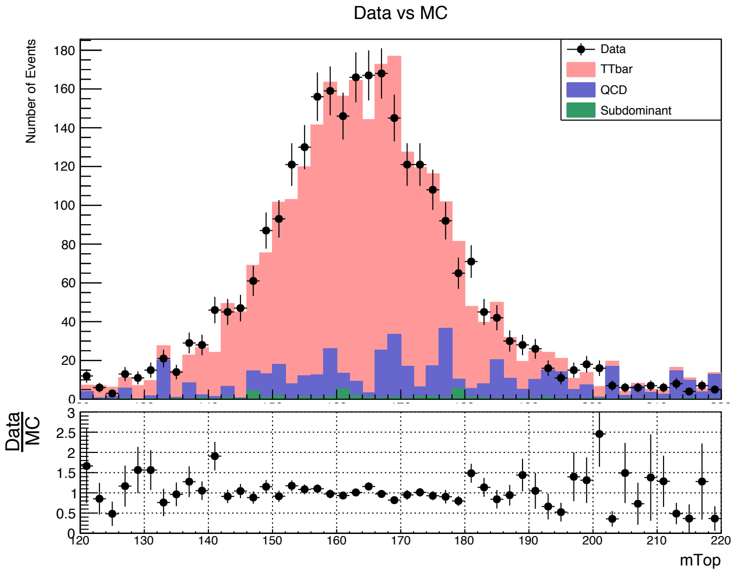


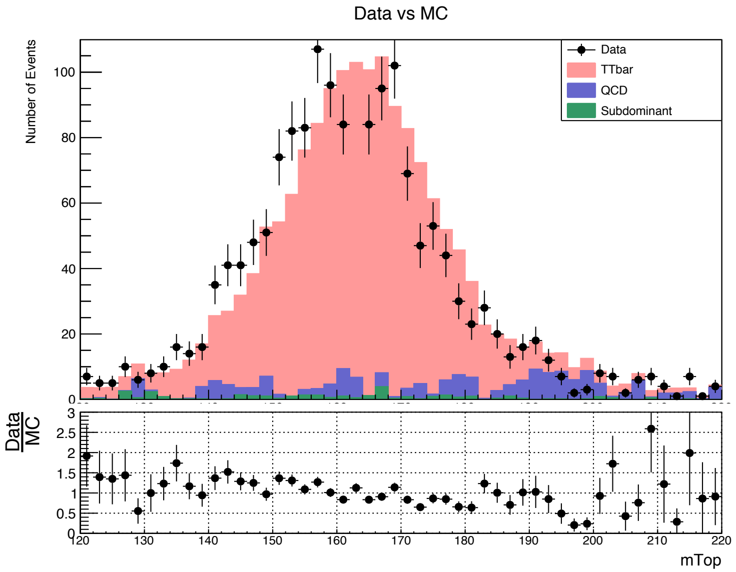


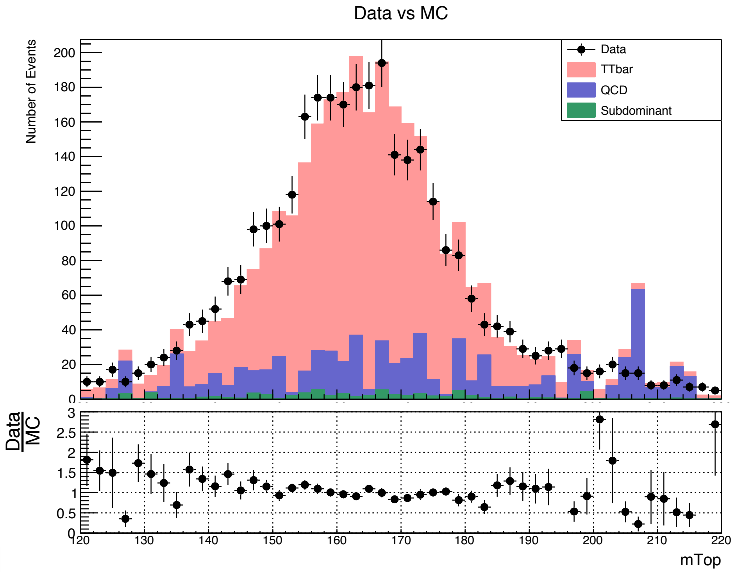


To scale the ttbar → fit the leading jetMassSoftDrop in each of these regions (ttbar compatible ~ with SR)

Tight TopTagger + Probe







Tag And Probe Calculations 2016

b tagging SF’s

Efficiency--with btagging SF's
eff data: 0.781 ± 0.038
eff ttbar: 0.772 ± 0.014

Efficiency per Pt region
eff data pT[400-600]: 0.761 ± 0.042
eff ttbar pT[400-600]: 0.778 ± 0.016
----
eff data pT[600-800]: 0.851 ± 0.100
eff ttbar pT[600-800]: 0.748 ± 0.031
----
eff data pT[800-Inf]: 0.886 ± 0.160
eff ttbar pT[800-Inf]: 0.775 ± 0.063

without b tagging SF’s

Efficiency--without btagging SF's
eff data: 0.782 ± 0.039
eff ttbar: 0.772 ± 0.014

Efficiency per Pt region
eff data pT[400-600]: 0.762 ± 0.043
eff ttbar pT[400-600]: 0.778 ± 0.016
----
eff data pT[600-800]: 0.854 ± 0.103
eff ttbar pT[600-800]: 0.748 ± 0.031
----
eff data pT[800-Inf]: 0.888 ± 0.161
eff ttbar pT[800-Inf]: 0.775 ± 0.064

Tag And Probe Calculations 2017

b tagging SF’s

Efficiency-- with btagging SF's
eff data: 0.857 ± 0.040
eff ttbar: 0.875 ± 0.0072

Efficiency per Pt region
eff data pT[400-600]: 0.872 ± 0.047
eff ttbar pT[400-600]: 0.874 ± 0.008
----
eff data pT[600-800]: 0.795 ± 0.088
eff ttbar pT[600-800]: 0.876 ± 0.018
----
eff data pT[800-Inf]: 0.797 ± 0.186
eff ttbar pT[800-Inf]: 0.899 ± 0.045

without b tagging SF’s

Efficiency-- without btagging SF's
eff data: 0.864 ± 0.043
eff ttbar: 0.875 ± 0.007

Efficiency per Pt region
eff data pT[400-600]: 0.880 ± 0.049
eff ttbar pT[400-600]: 0.874 ± 0.008
----
eff data pT[600-800]: 0.8 ± 0.091
eff ttbar pT[600-800]: 0.876 ± 0.018
----
eff data pT[800-Inf]: 0.796 ± 0.2
eff ttbar pT[800-Inf]: 0.898 ± 0.045

Tag And Probe Calculations 2018

b tagging SF’s

Efficiency-- with tag SF's
eff data: 0.816 ± 0.032
eff ttbar: 0.839 ± 0.005

Efficiency per Pt region
eff data pT[400-600]: 0.8176 ± 0.038
eff ttbar pT[400-600]: 0.837 ± 0.006
----
eff data pT[600-800]: 0.809 ± 0.063
eff ttbar pT[600-800]: 0.847 ± 0.013
----
eff data pT[800-Inf]: 0.772 ± 0.132
eff ttbar pT[800-Inf]: 0.868 ± 0.032

without b tagging SF’s

Efficiency-- without tag sf's
eff data: 0.822 ± 0.034
eff ttbar: 0.839 ± 0.005

Efficiency per Pt region
eff data pT[400-600]: 0.824 ± 0.039
eff ttbar pT[400-600]: 0.837 ± 0.006
----
eff data pT[600-800]: 0.819 ± 0.066
eff ttbar pT[600-800]: 0.847 ± 0.013
----
eff data pT[800-Inf]: 0.789 ± 0.141
eff ttbar pT[800-Inf]: 0.868 ± 0.032