Measurement of CP-violating Wtb couplings using EFT in single top t-channel production with CMS Run 2 data

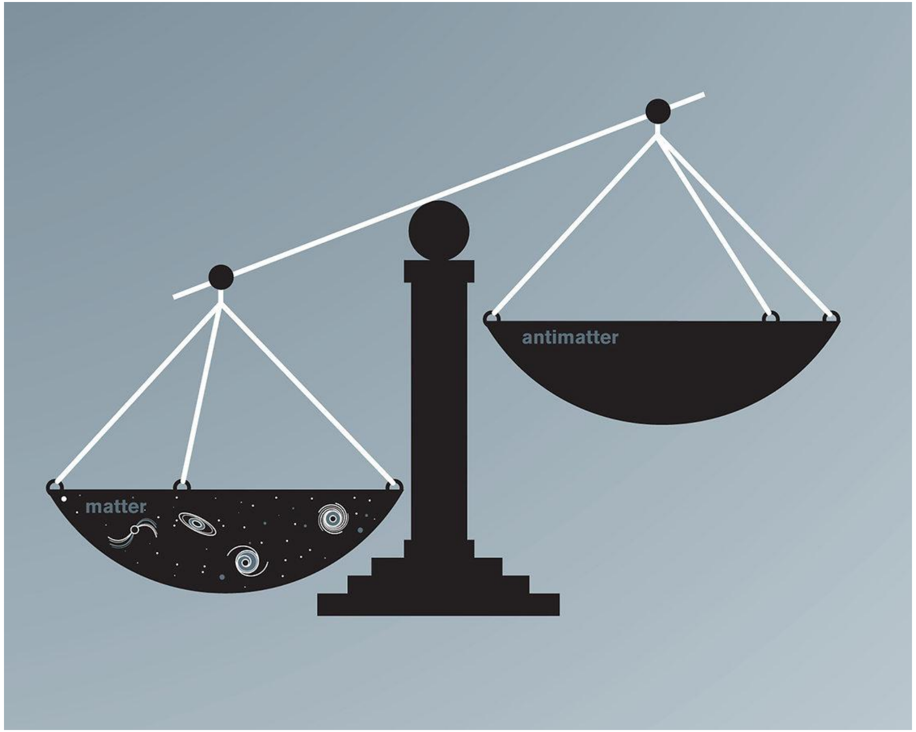
Particle group meeting
27/05/2025

Nicolas Chanon (Staff), Arnab Purohit (Postdoc), Enzo Fillaudeau (PhD)

CMS group at IP2I Lyon

Context : Matter and antimatter asymmetry

The universe is baryon-number asymmetric

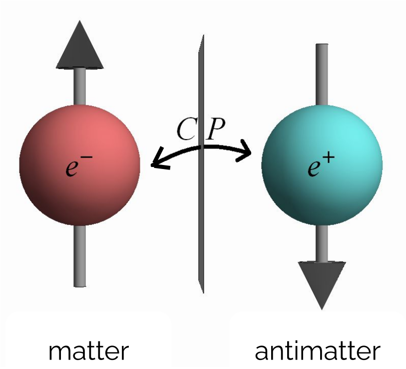


SM prediction: $$\eta_{SM} = \frac{n_B - n_{\bar{B}}}{n_\gamma} \propto 10^{-27}$$

Observation: $$\eta_{obs} = \frac{n_B - n_{\bar{B}}}{n_\gamma} \propto 10^{-10}$$

$$\Rightarrow \frac{\eta_{SM}}{\eta_{obs}} \propto 10^{-17}$$

Discrepancy between the SM prediction and observations

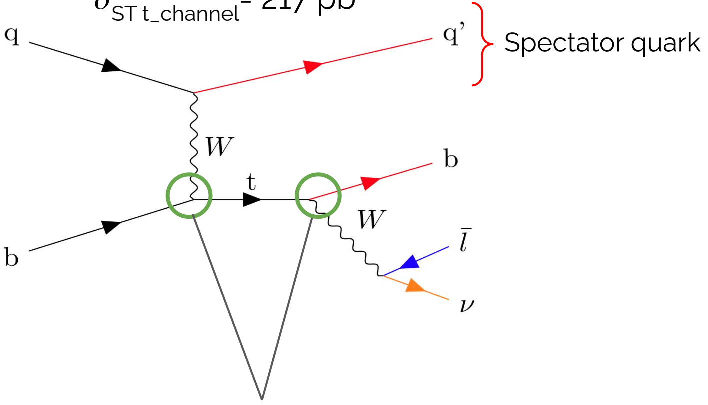


Looking for new CP violation sources involving top quarks Beyond the Standard Model (BSM).

Goal of the analysis

Measure CP violation through the t-channel production of single-top quarks and their subsequent decay to Wb. *This process has the advantage of involving twice the Wtb vertex*

$\sigma_{\text{ST t\_channel}}$ = 217 pb



Analysis performed in the context of Effective Field Theory

$$\mathcal{L}_{eff}^{(6)} = \mathcal{L}_{SM} + \sum_i \frac{C_i^{(6)}}{\Lambda_i^2} O_i^{(6)} + h.c.$$

Non-Hermitian Operators involving two quarks and a boson	Wilson Coefficients
$O_{uW}^{(6)} = (\bar{q}\sigma^{\mu\nu}\tau^I t)\tilde{\varphi}W_{\mu\nu}^I$	$c_{tW} + ic_{tW}^I$
$O_{dW}^{(6)} = (\bar{q}\sigma^{\mu\nu}\tau^I b)\varphi W_{\mu\nu}^I$	$c_{bW} + ic_{bW}^I$
$O_{\varphi ud}^{(6)} = (\tilde{\varphi}^\dagger iD_\mu\varphi)(\bar{u}\gamma^\mu d)$	$c_{\varphi tb} + ic_{\varphi tb}^I$

CP violation if

$c^i_{tW} \neq 0$, $c^i_{bW} \neq 0$ or $c^i_{\varphi tb} \neq 0$

➔ EFT impacts both the production and decay of top quark
➔ This vertex can be modified by CP violation
➔ The effect is canceled in ttbar process

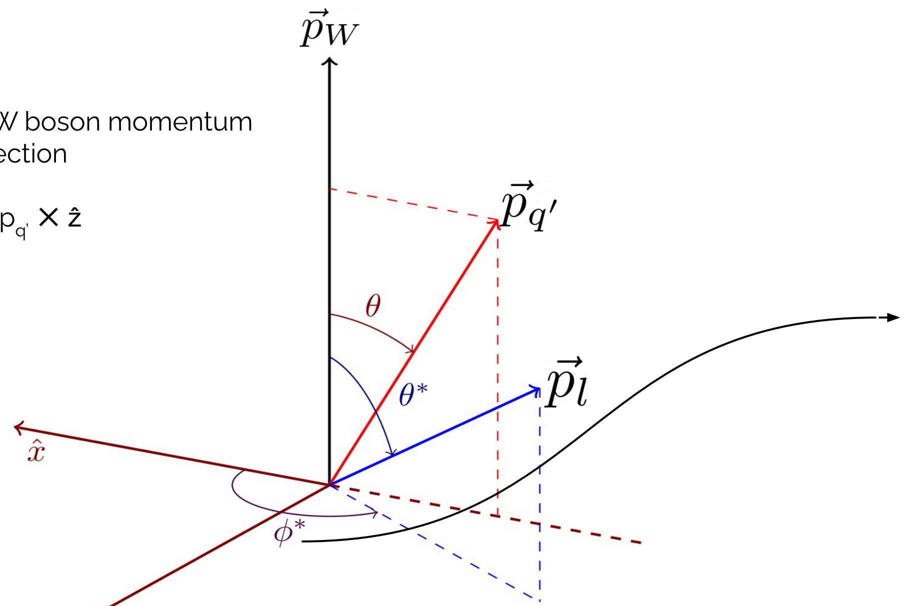
Measuring CP violation : EFT impact on angular variables

Top quark rest frame

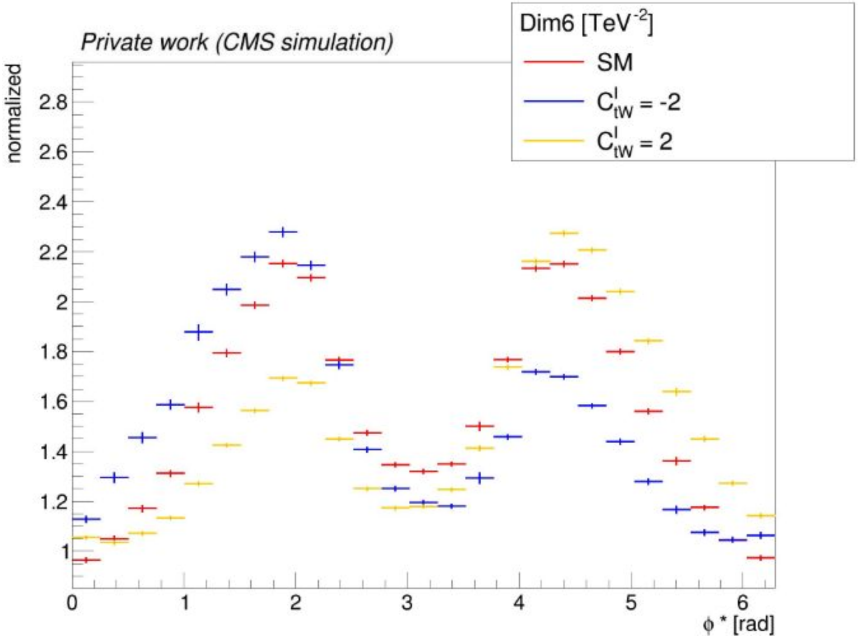
Reference frame used in ATLAS 8 TeV [arXiv:1707.05393]

$\hat{\mathbf{z}}$: W boson momentum direction

$\hat{y} = p_{q'} \times \hat{z}$



ϕ^*= angle between x-axis and projection of lepton momentum on the xy plane

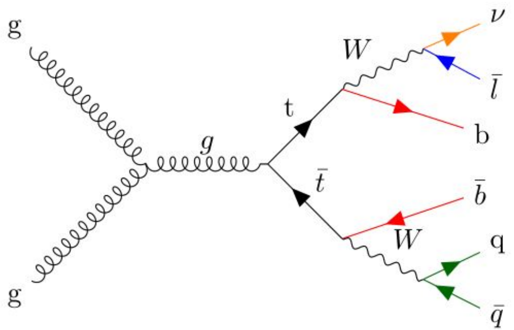


The shape of the distribution varies depending on the value of the EFT coefficient

The amount of CP violation can be extracted using such angular distributions

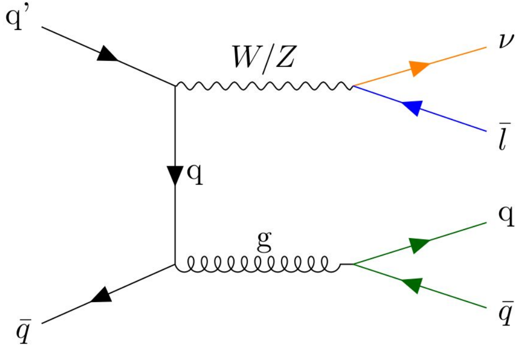
Main background processes :

ttbar



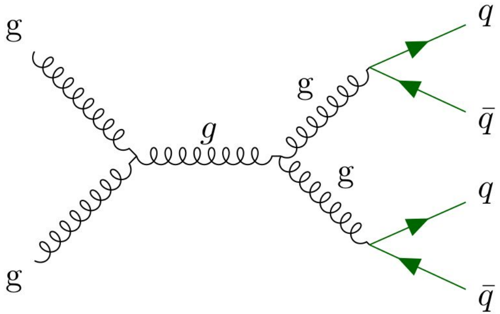
σ_{ttbar}= 452 pb

W/Z + jets (with Drell-Yan)



σ_{Wjets}= 84 847 pb

QCD processes



Other backgrounds :

- ➔ Diboson
- ➔ ttX
- ➔ Single top s-channel and tW process

Event samples

- **Single Top/AntiTop t-channel (SM)** : ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8
- **ttbar semileptonic** : TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8
- **W + Jets** : WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8
- **QCD** : pT binned (mu or EM enriched) QCD_Pt-50To80_MuEnrichedPt5_TuneCP5_13TeV-pythia8
- **tW top and anti-top** : ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8

EFT samples :

- ➢ Generated using the dim6Top LO UFO model from Madgraph5
- ➢ A reweighting technique is used to get all the EFT combination

Object definition

The **single isolated muon** trigger with p_T > 24 GeV (2016, 2018) and p_T > 27 GeV (2017) are used in this analysis

Jet selection

	Good jets
|η|	< 4.7
p_T 2016 (GeV)	> 40
p_T 2017 and 2018 (GeV)	> 40 (|η| < 2.4) > 60 (2.4 < |η| < 4.7)
Overlap	Removed overlap between jets and leptons in a ΔR < 0.4 cone
Jet Id	**Tight** (discriminate real jets from fake lepton, pile up and detector noise)

b-jets :

- ➢ Must be good jets
- ➢ \|η\| < 2.5 (2017 & 2018)
- ➢ \|η\| <2.4 (2016)
- ➢ Tight/medium working point of DeepJet tagger used

Muon selection

	Isolated muon	Loose muon
|η|	< 2.4	< 2.4
pT 2017 (GeV)	> 30	> 10
pT 2016 and 2018 (GeV)	> 26	> 10
Relative isolation	< 15%	< 25%
Id	tight	loose

A reversed isolated muon is also defined by reverting the isolation : > 40%

➔ All recommended CMS corrections are applied (including pileup and b-tag weights, muon Rochester momentum correction, JEC, MET phi modulation)

Analysis regions

Loose preselection applied to reduce fake lepton contribution :

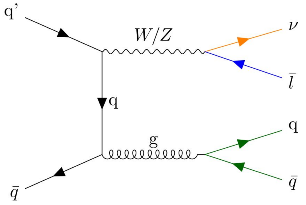
- ➢ **MET** > 20 GeV
- ➢ **Transverse mass of W boson with fixed lepton p_T at 45 GeV ($M_T^{W,fix}$**) > 25 GeV

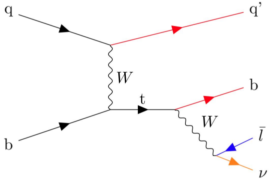
Lepton selection :

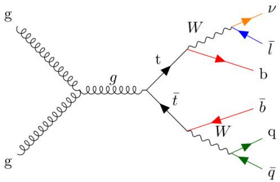
- Exactly one isolated tight muon
- Veto events with additional loose muons
- Veto events with veto electrons

Event categorization based on the number of Jets and b-tagged jets : Signal region (SR) , W/Z-Jets control region (CR) and ttbar CR

	= 0 tight and 1 medium b-jet	= 1 tight b-jet	>= 2 tight b-jet
= 2 jets	W/Z-Jets Control Region.	Signal Region	Not Used
≧ 3 jets	Not Used	Not Used	TTbar Control Region

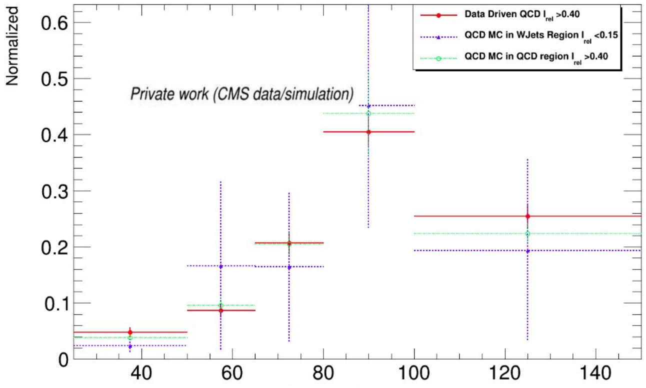






+ QCD measurement region (sideband region) :

- ❖ defined for each signal and control region
- ❖ same selection but reverted muon isolation (>0.4)

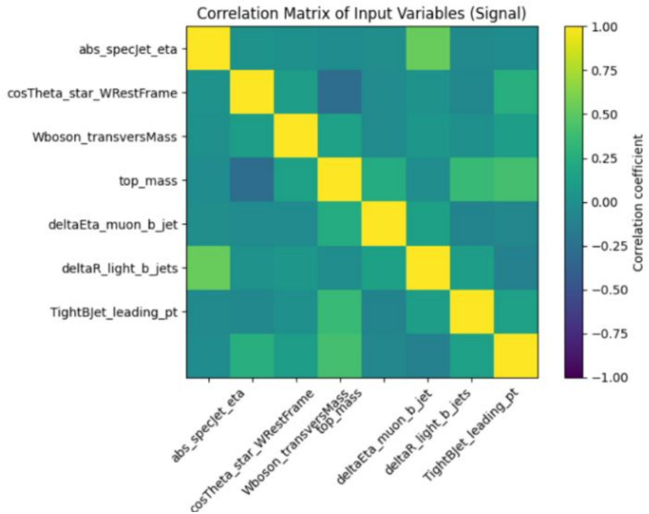


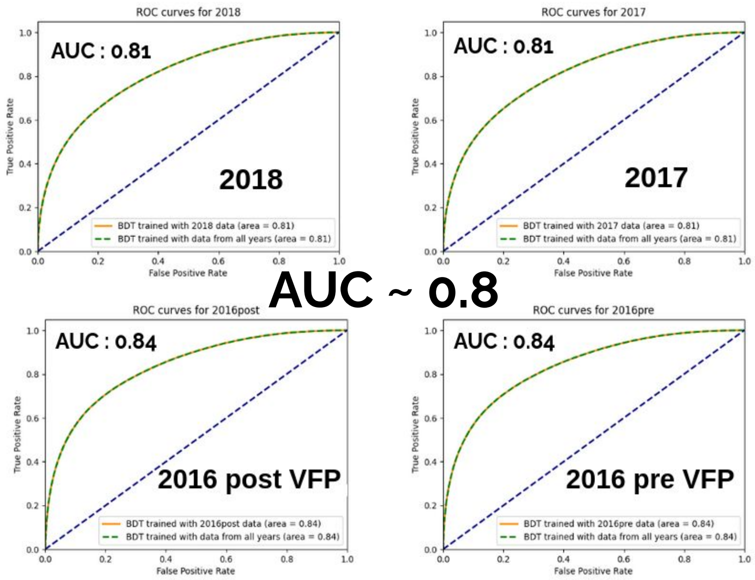
Comparison of the distribution of $M_T^{W,fix}$ for different QCD templates

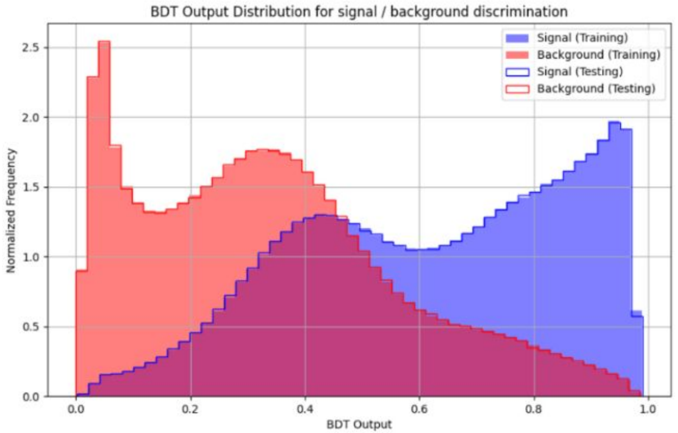
Signal Region

In the SR we require : **1** isolated tight muon, **2** jets and **1** tight b-jet

- **BDT trained in the SR using XGBoost to discriminate single top (ST) t-channel from other SM background process**
- *Input variables for the BDT* : cos(θ^*), η of the spectator jet, $M_T^{W,fix}$, M_{top}, $\Delta\eta(\mu$, b-jet), ΔR(spec jet, b-jet), b-jet p_T, lepton p_T



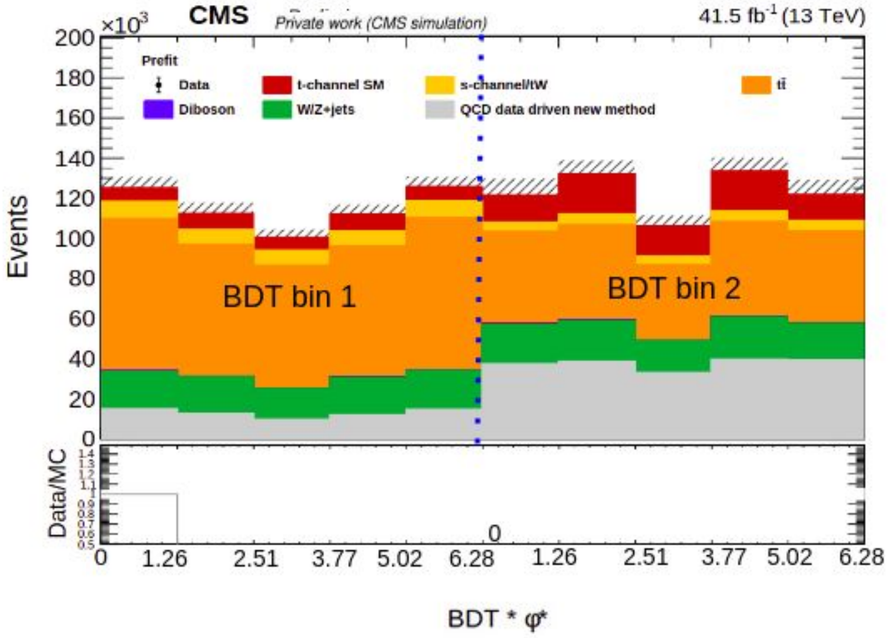




Signal Region

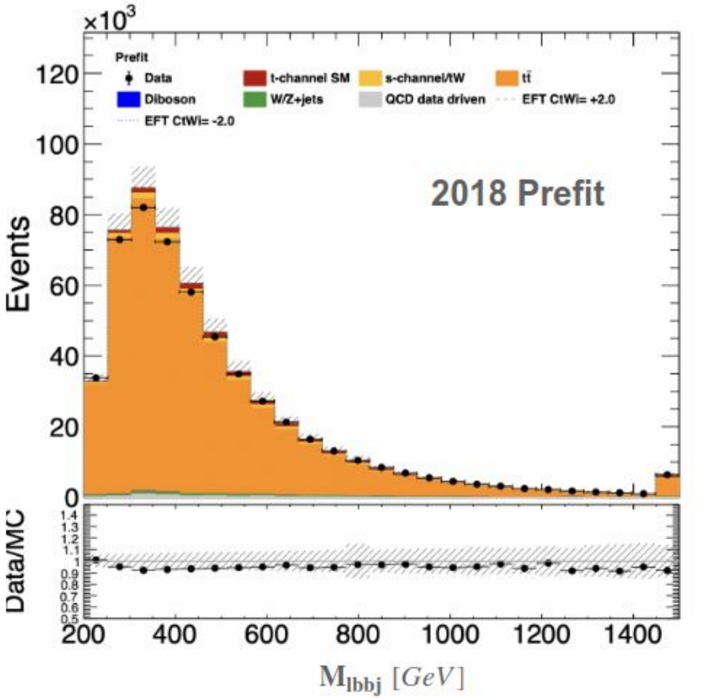
As the measurement observable we are using the ϕ^* variable in bins of the BDT

- ➔ Constructed by unrolling the 2D distribution to get a 1D distribution
- ➔ ϕ^* : Measure the CP violation (c_{tW}^{i} Wilson Coefficient)
- ➔ BDT : Increase the sensitivity of the analysis



TTbar Control Region

In the ttbar control region, we are using **invariant mass of the lepton and three jets** (2 tight b- tagged jets and a third jet with the highest p_T) system.

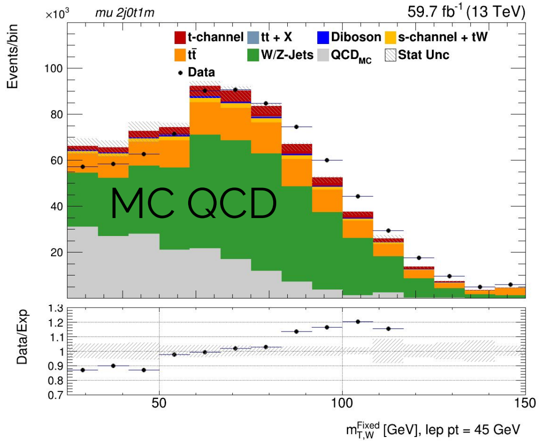


- ➔ Variable chosen for its power to discriminate between ttbar and the rest
- ➔ Good data/MC agreement observed

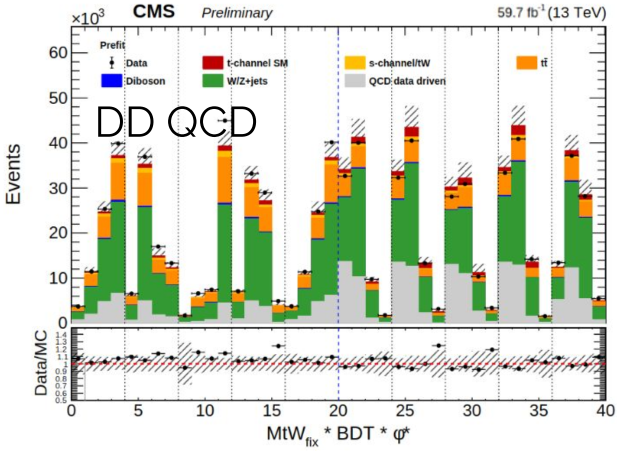
W/Z + jets Control Region

In the W-jets CR we use the **transverse mass of the W boson with fixed lepton p_T at 45 GeV ($M_T^W{}_{,fix}$)** in bins of our main observable (3D distribution unrolled in 1D)

➔ Variable chosen for its power to discriminate between W+jets and the rest



Bad data/MC agreement using MC QCD



We still get unsatisfactory data/MC agreement due to bad QCD modelling : **We need a new method to estimate the QCD**

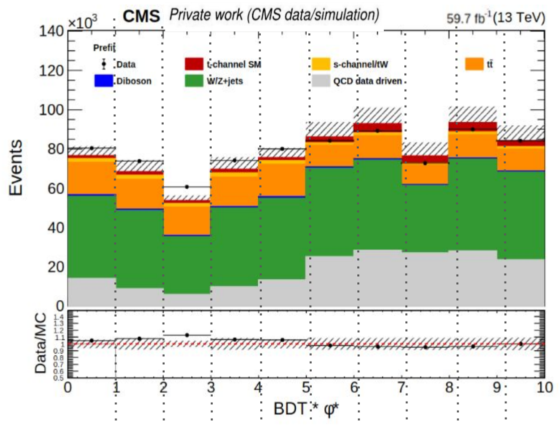
QCD estimation strategy

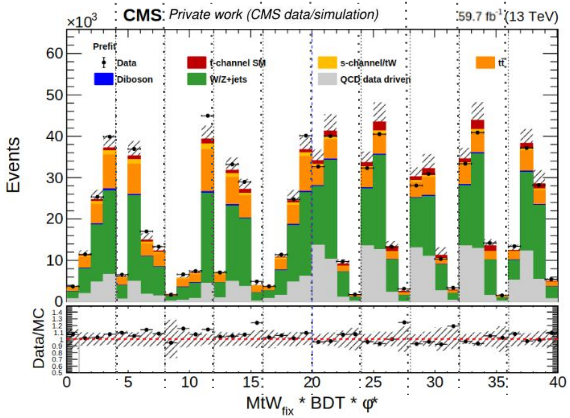
- *There is no standard method to estimate QCD (under discussion within TOP PAG)*
- *We present here our QCD estimation strategy*

Let's take any variable

For each bin of this variable :

1. **Build the $M_T^{W,fix}$ template** (transverse mass of the W with lepton p_T fixed at 45 GeV) for every process after optimizing the binning of the variable and $M_T^{W,fix}$
2. Replace the QCD MC template by the **data driven QCD template** (obtained from the QCD sideband region) and set its normalization to that of QCD MC from SR.
3. Perform a **simultaneous fit** over all bins of the chosen variable with **n Parameters Of Interests (POIs) for the QCD normalization** (n bins of the variable)

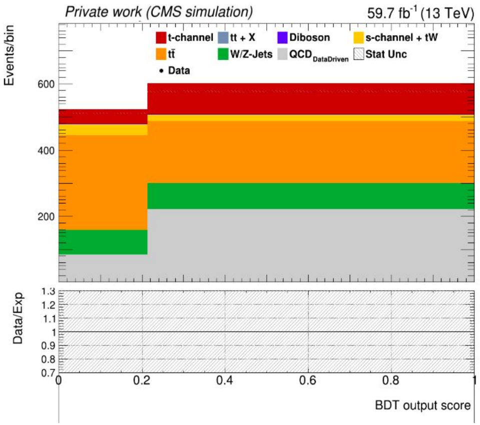




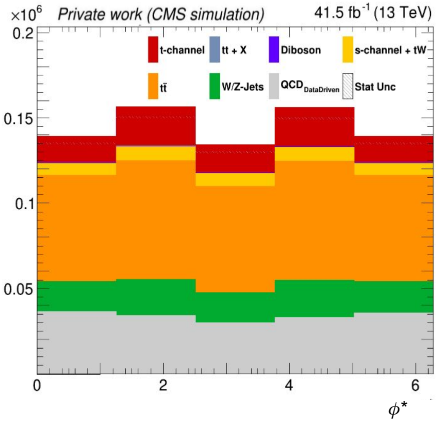
SM t-channel rate parameter is fixed to 1

Binning optimization

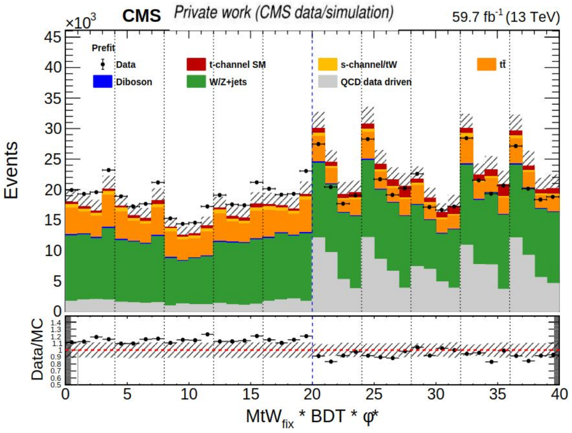
- **BDT binning** : We are using **2 bins of BDT with equal number of W+jets events** in the SR in each of them to reduce the impact of the lack of statistics for this process



- **ϕ^* binning** : Binning optimized in the SR by **maximizing the significance of ST t-channel with EFT over SM** (5 bins of ϕ^*)



- **$M_T^{W,fix}$** : Tried optimizing wrt significance of QCD over the rest but found better results using a binning with **equal number of W+jets events** in each bin. We are using 4 bins of $M_T^{W,fix}$



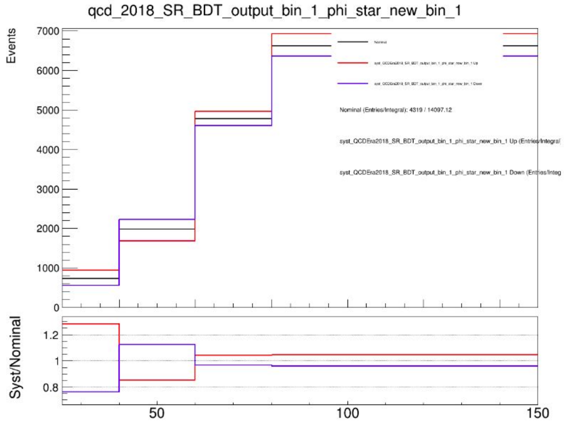
Systematic uncertainties

NAMES	USED	CORRELATION
Muon Reco	Y	All channel
Muon HLT	Y	All channel
Muon ID	Y	All channel
Muon Iso	Y	All channel
B Tagging SF	Y	All channel
Prefiring	Y	All channel
Pile Up	Y	All channel
Top pt	Y	Only TTbar
Lumi	Y	All channel
JES/JER	Y	As per Recom
ISR	Y	Not correlated
FSR	Y	Not correlated
PDF + alpha_s	Y	All channels
QCD Scale	Y	All channels
Diff btw NLO and LO	Y	Only EFT+ST

We introduce 2 uncertainties in this analysis :

- Data Driven QCD shape/era uncertainty (constructed by choosing different data taking era)
- Difference between NLO (POWHEG samples) and LO (Madgraph samples)

All uncertainties are correlated over bins of the main observable except **QCD era uncertainty**.



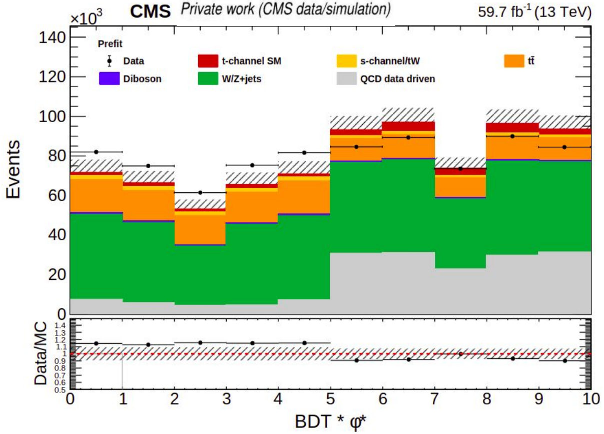
Down : 2018 A, B, C era
Up : 2018 D era
Nominal

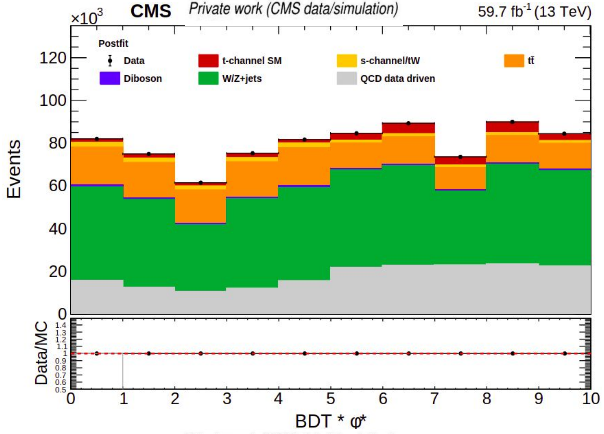
$M_T^{W,fix}$ distribution in BDT bin 1 and ϕ^* bin 1 with Up and Down variations of QCD era uncertainty

QCD Background only fit with data unblinded in the W-Jets region is performed

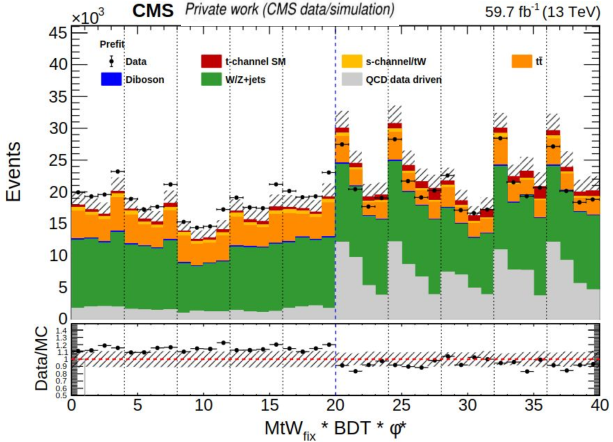
Prefit **Postfit**

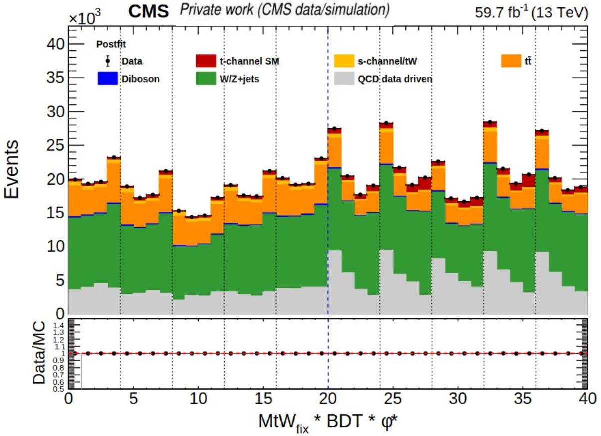
Integrated over $M_T^{W,fix}$





$M_T^{W,fix}$ in bins of BDT x ϕ^* (used in the QCD fit)

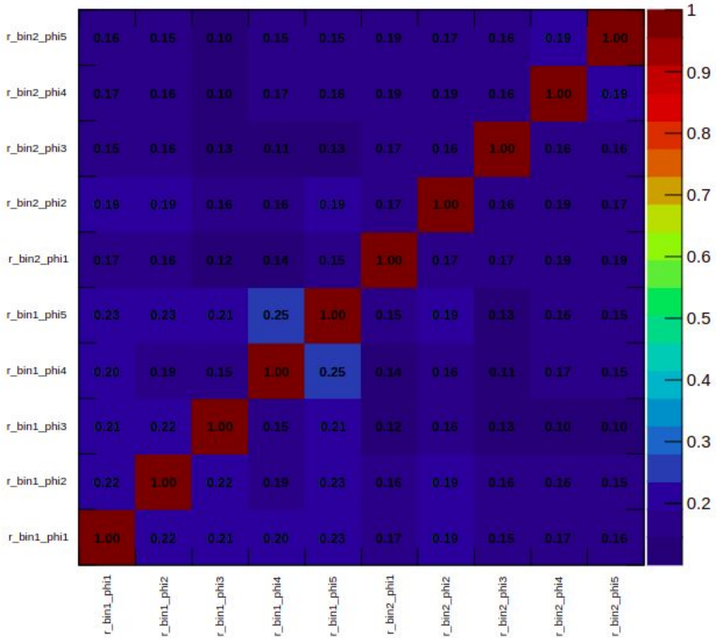




QCD background fit results   2018 results

Uncertainties include both statistical and systematic

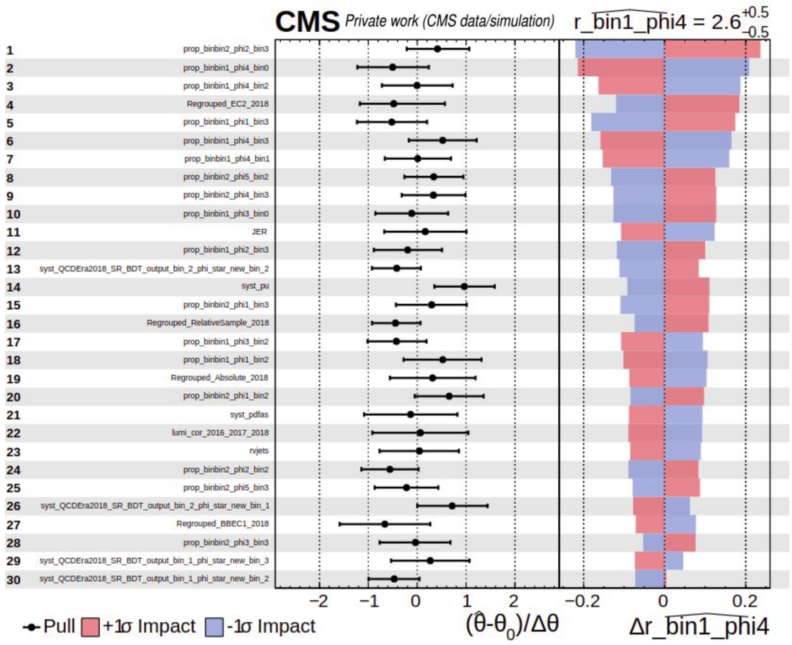
POI Name	Best Fit Value	Uncertainty
r_bin1_phi1	2.130	+/- 0.136
r_bin1_phi2	2.144	+/- 0.169
r_bin1_phi3	2.397	+/- 0.212
r_bin1_phi4	2.581	+/- 0.215
r_bin1_phi5	2.115	+/- 0.145
r_bin2_phi1	0.714	+/- 0.033
r_bin2_phi2	0.733	+/- 0.033
r_bin2_phi3	1.002	+/- 0.041
r_bin2_phi4	0.791	+/- 0.035
r_bin2_phi5	0.717	+/- 0.032

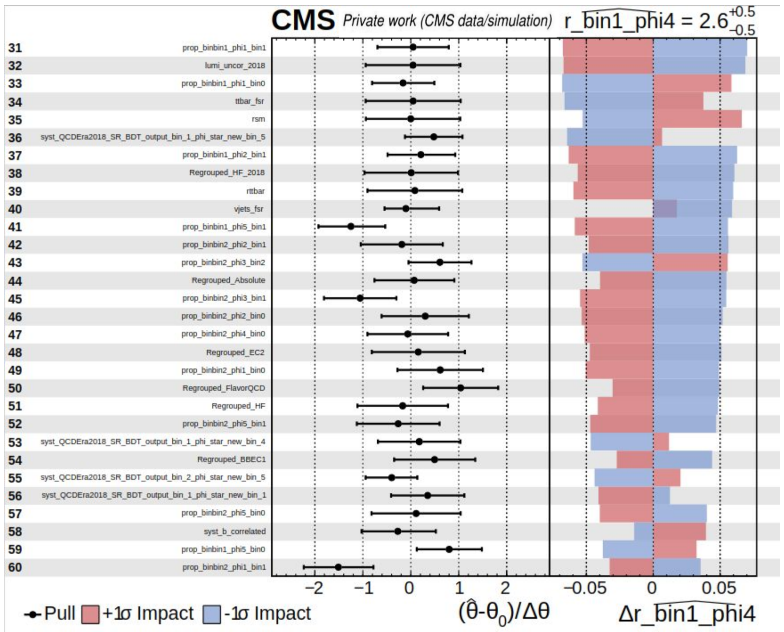


➢ Low correlations (<0.25) between the POIs

Fit diagnostics

2018 results





➔ The impact of the nuisances on the POI of BDT bin 1 and phi star bin 4 is presented

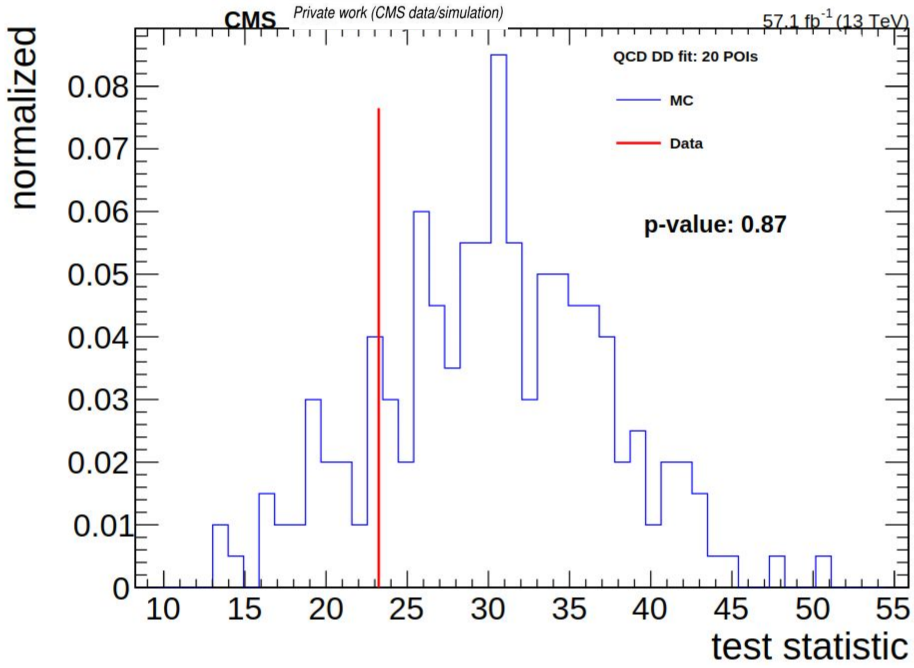
- Pulls of the nuisances are not very large
- MC statistical uncertainties have a large impact on the POIs

Fit diagnostics

2018 results

- A goodness of fit (GoF) test with the saturated model where likelihood ratio is used as test statistic, and the alternate hypothesis is considered to be the data itself, is performed.

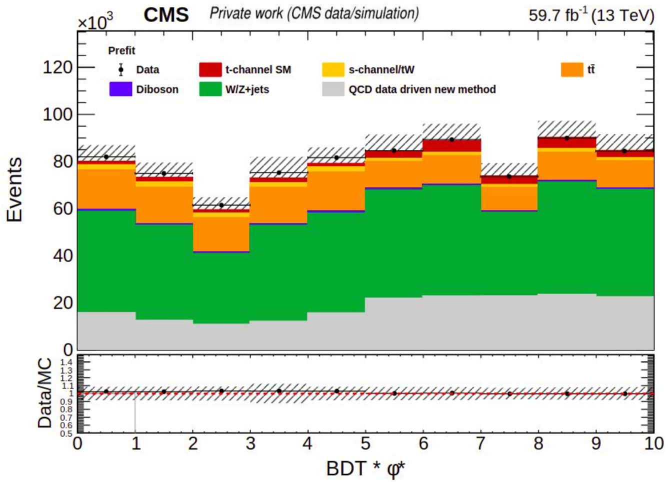
- The observed P-value shows, the model adequately fits the data

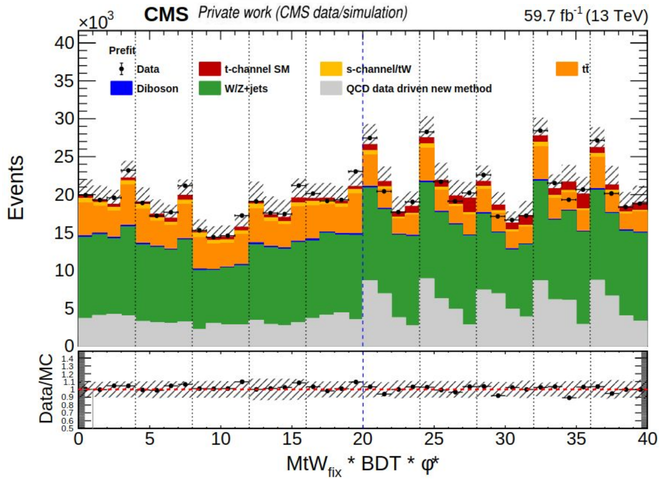


New template for the final EFT fit

2018 results

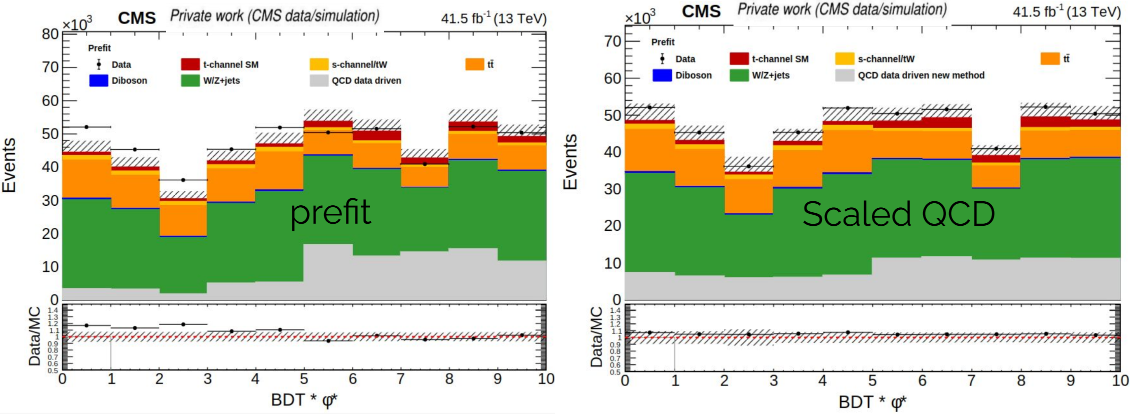
- A new QCD template (“aka” Scaled QCD) is constructed from the best fit values of the POIs
- The templates does **not** include pulled nuisances from background data fit
- Upper plot will be the input to the final EFT fit
- Clear improvement of the data/MC agreement compared to prefit (slide 12)





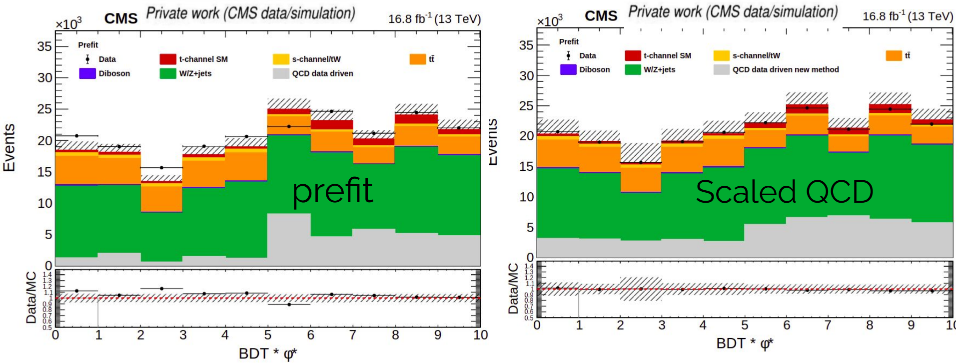
QCD background fit results for other years

2017

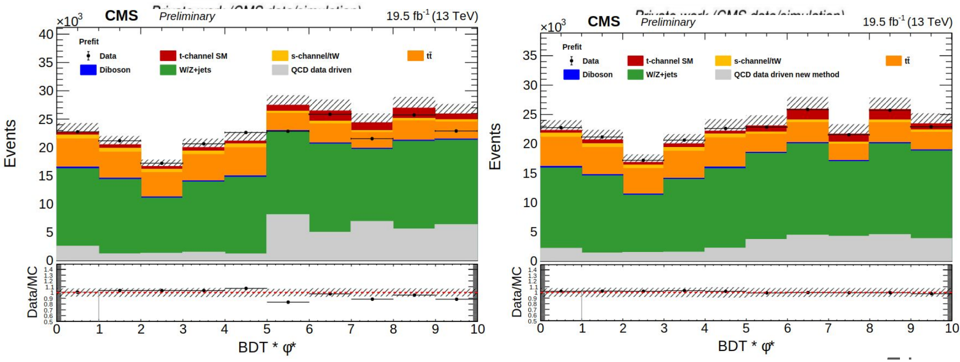


➔ Same method is used for other years showing same improvement in the data/MC agreement as in 2018.

2016 postVFP



2016 preVFP



Conclusion

What has been done :

- ➔ Produced private EFT samples
- ➔ Divided the analysis into signal and control regions
- ➔ Trained a BDT to improve the sensitivity of the analysis on EFT measurement
- ➔ Developed a promising QCD estimation strategy

Plans :

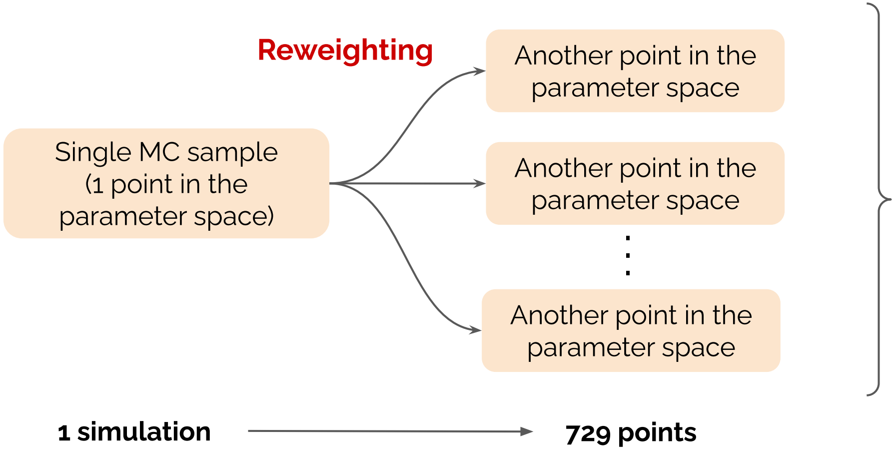
- ❑ Currently working on the final EFT fit for all years
- ❑ Write an analysis note

Thank you for your attention

Backup

Reweighting technique

- We produce a simulation sample for single top production including EFT coefficients at top production and decay
- <u>Reweighting method:</u> different regions of the parameter space to be probed with a single Monte Carlo (MC) sample



$\{ C_{tW} , C^{I}_{tW} \} = \{-2,0,2\}$

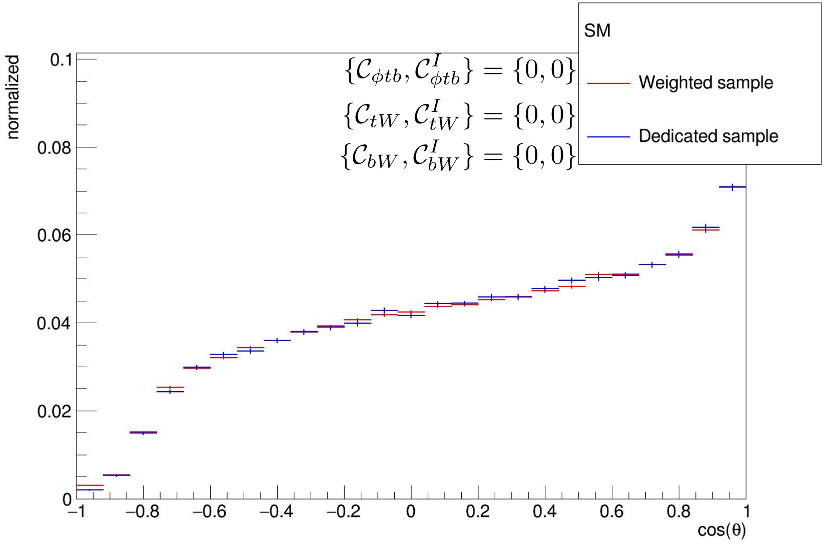
$\{ C_{bW} , C^{I}_{bW} \} = \{-2,0,2\}$

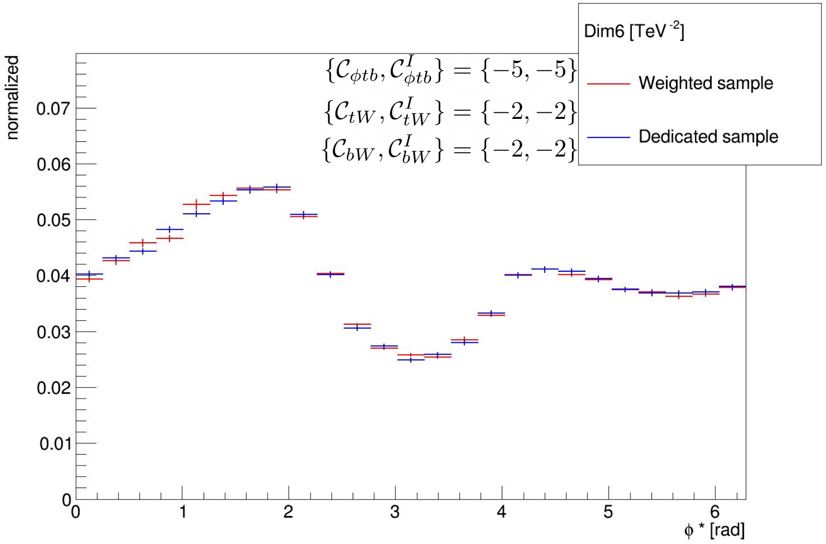
$\{ C_{\varphi tb} , C^{I}_{\varphi tb} \} = \{-5,0,5\}$

Our 6 axes of the parameter space — Different combination values

The reweighting method allows to produce a single sample instead of 729

Reweighting technique

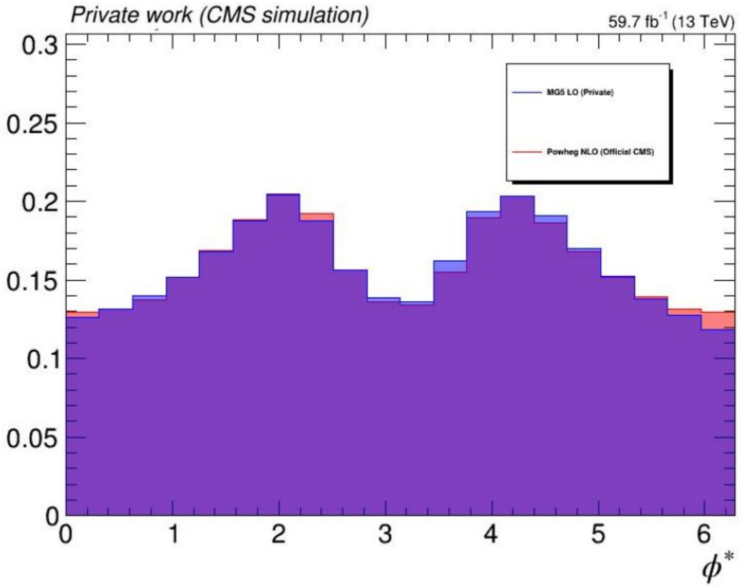


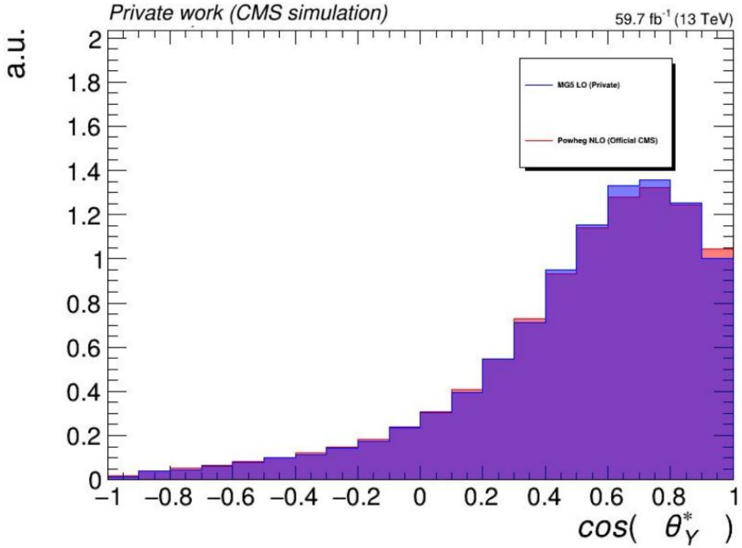


Comparing reweighted distributions of cos(θ) and ϕ^* to dedicated (non-reweighted) samples at two different distant points of the parameter space

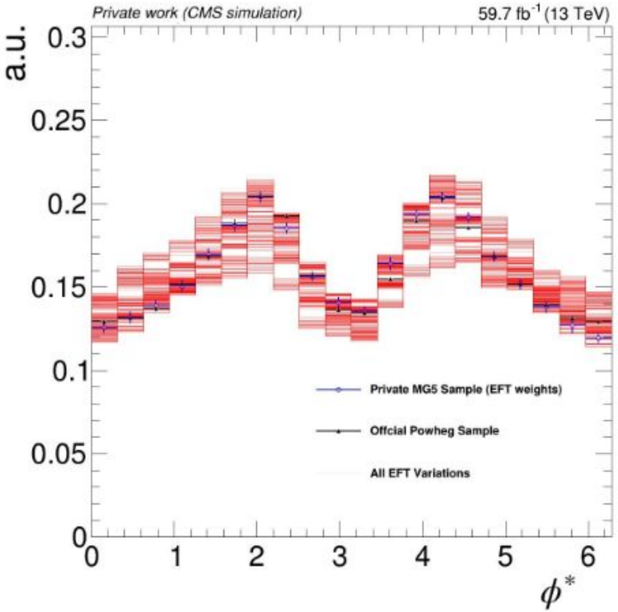
➔ **Reweighting is validated**

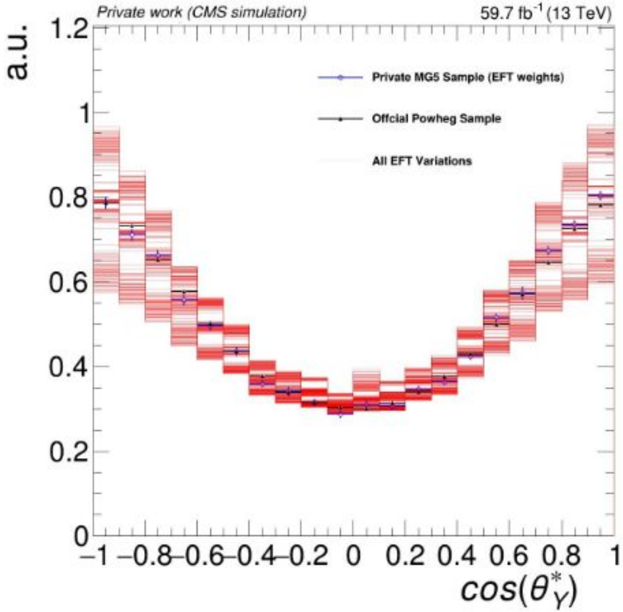
Reweighting technique



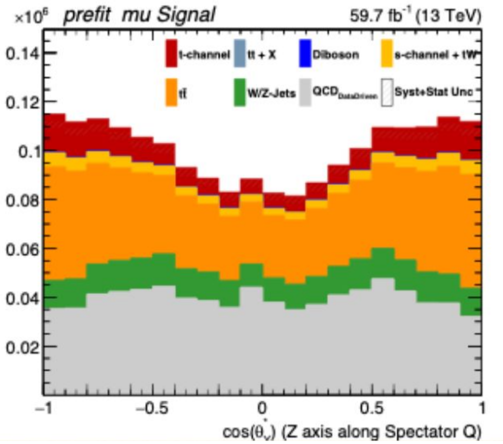


Reweighting technique

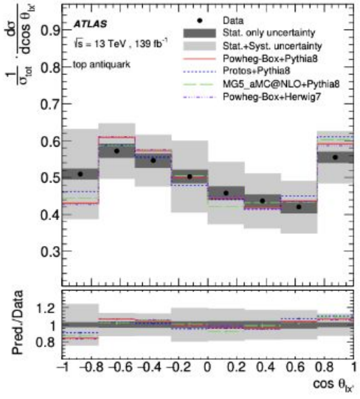




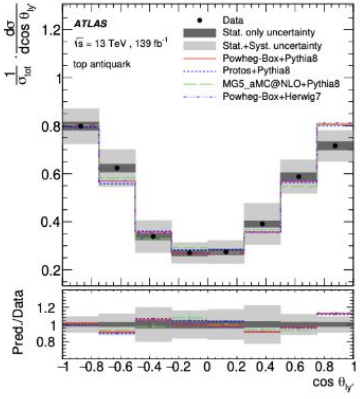
ATLAS results



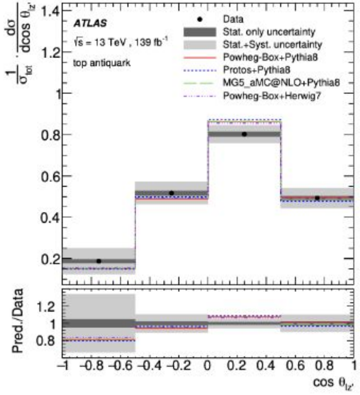
	C_{tW}		C_{itW}	
	68% CL	**95% CL**	**68% CL**	**95% CL**
All terms	$[-0.3, 0.8]$	$[-0.9, 1.4]$	$[-0.5, -0.1]$	$[-0.8, 0.2]$
Order $1/\Lambda^4$	$[-0.3, 0.8]$	$[-0.9, 1.4]$	$[-0.5, -0.1]$	$[-0.8, 0.2]$
Order $1/\Lambda^2$	$[-0.3, 0.8]$	$[-0.8, 1.5]$	$[-0.6, -0.1]$	$[-0.8, 0.2]$



(a)

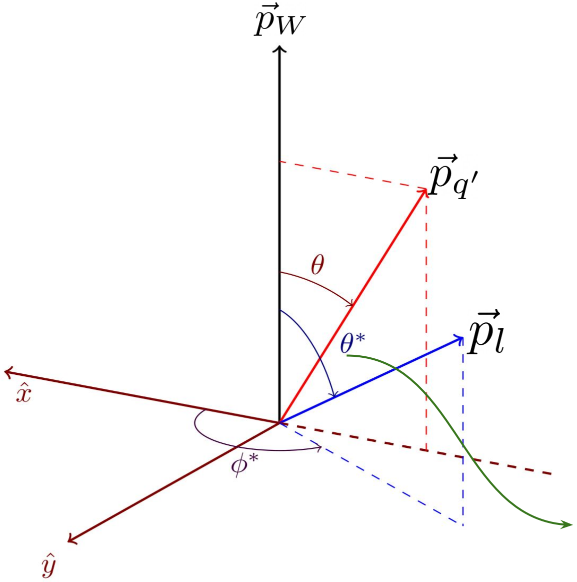


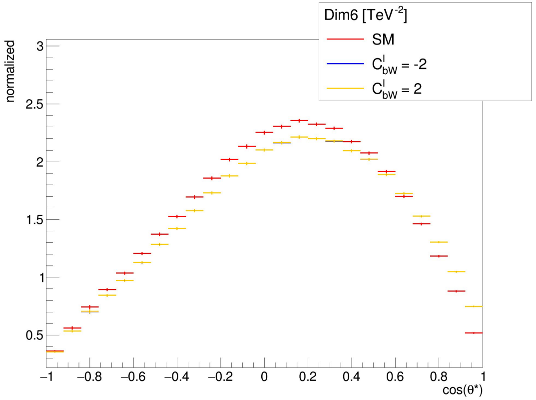
(b)



JHEP11(2022)040

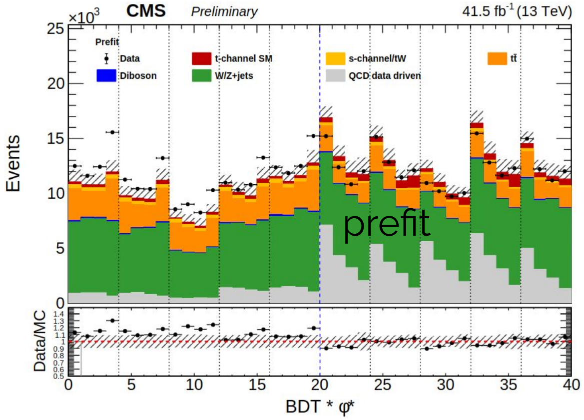
Impact of EFT on kinematic variable

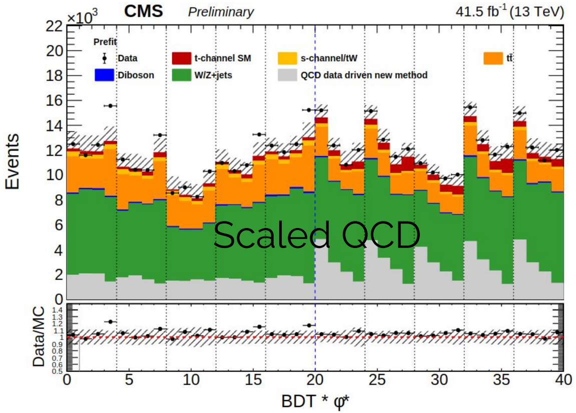




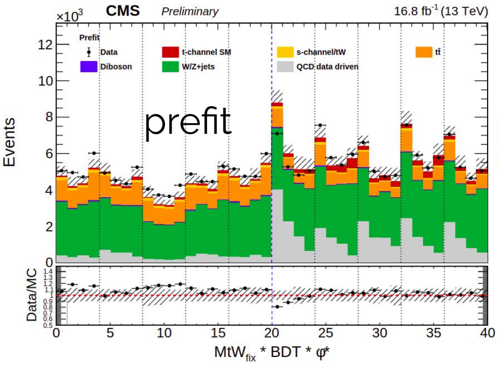
QCD background fit results for other years

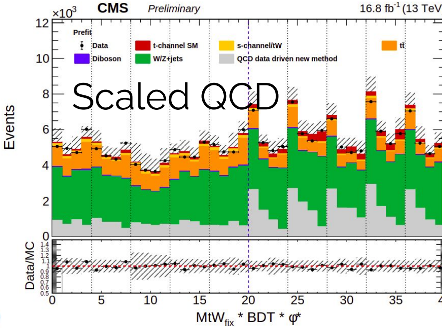
2017



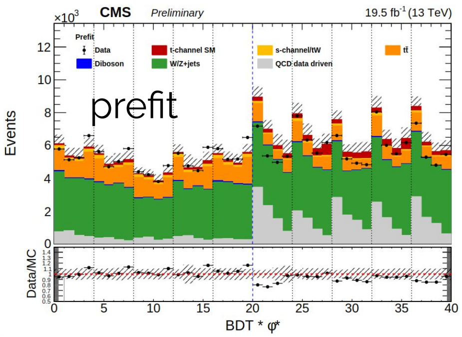


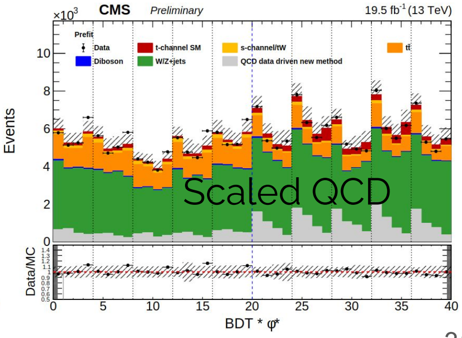
2016 postVFP





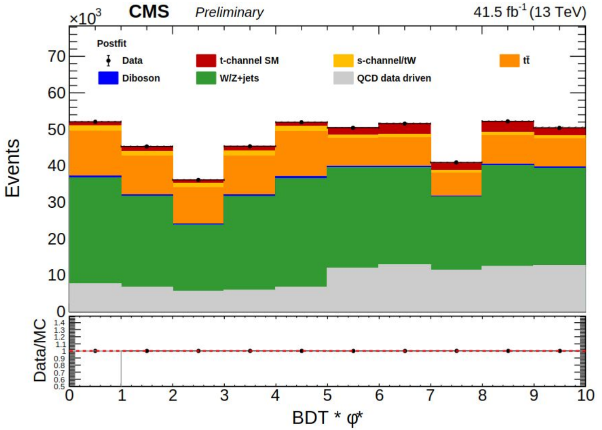
2016 preVFP

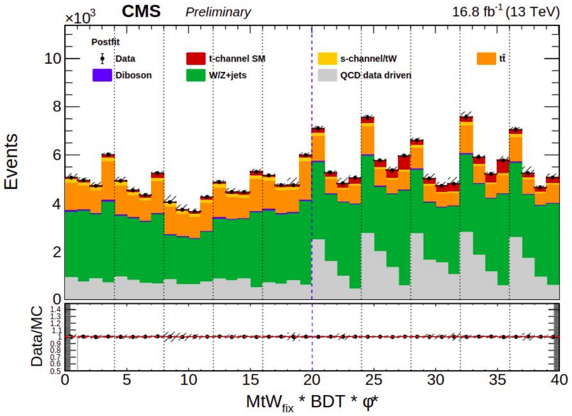




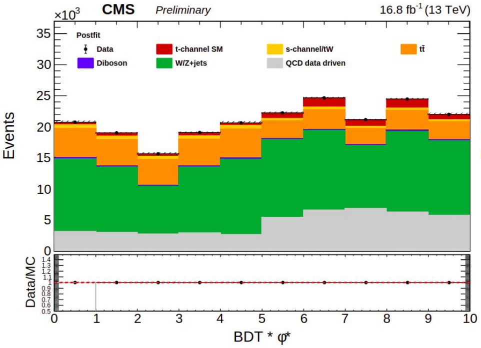
QCD background fit results for other years *postfit*

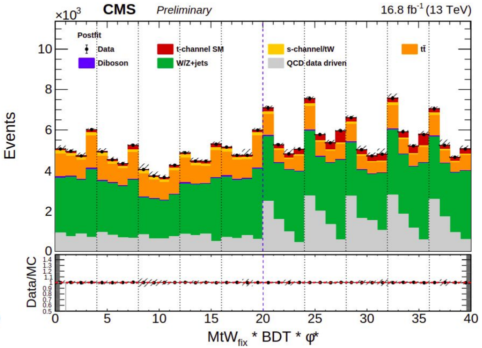
2017



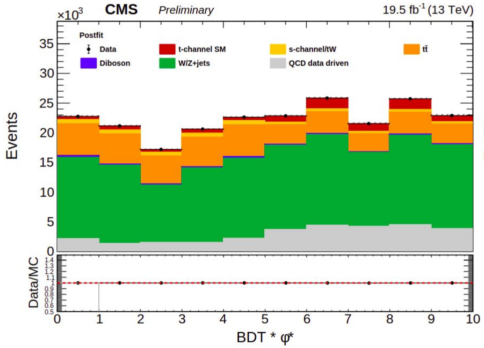


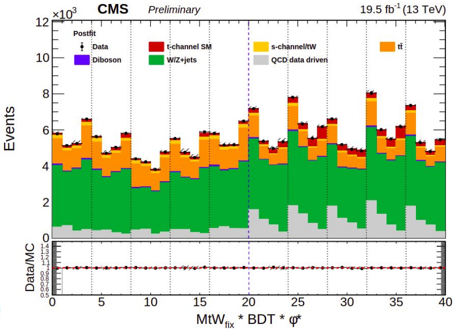
2016 postVFP





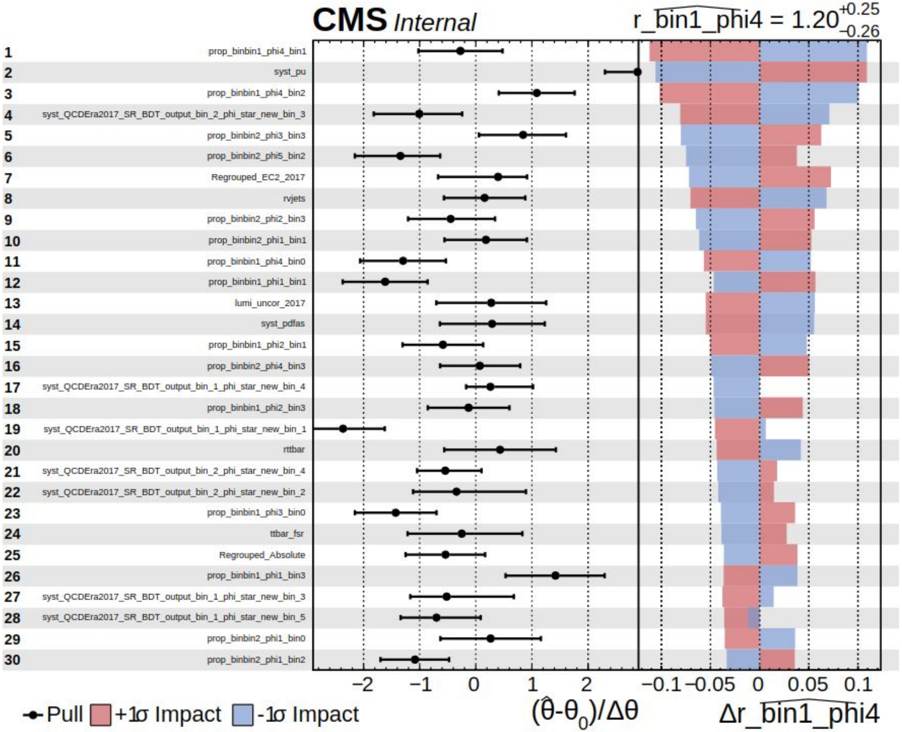
2016 preVFP

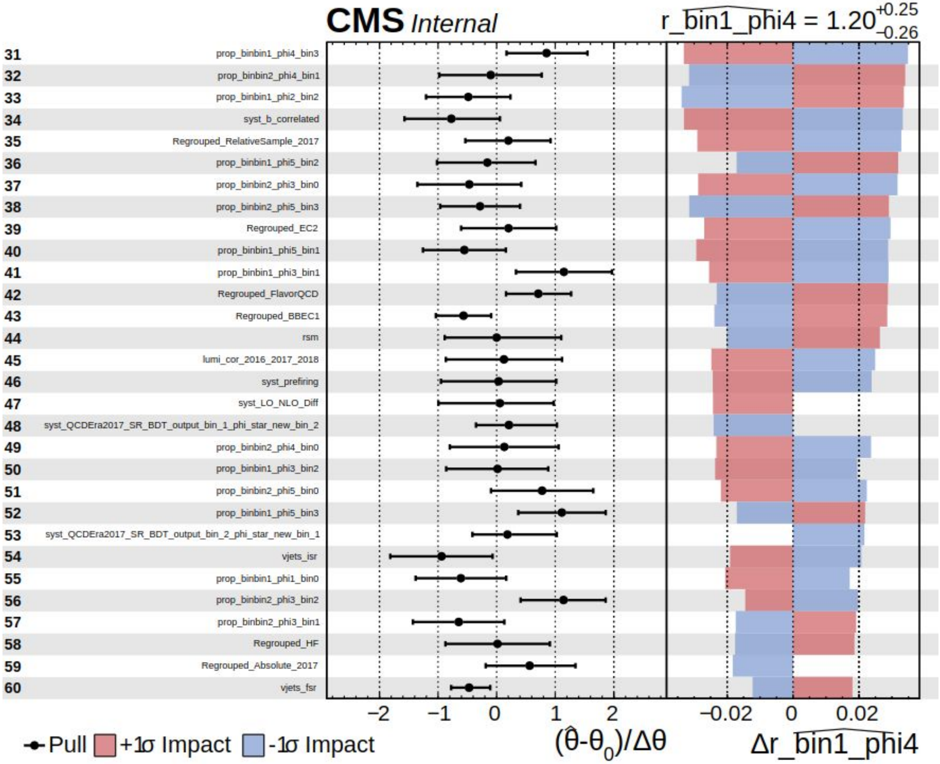




Fit diagnostics

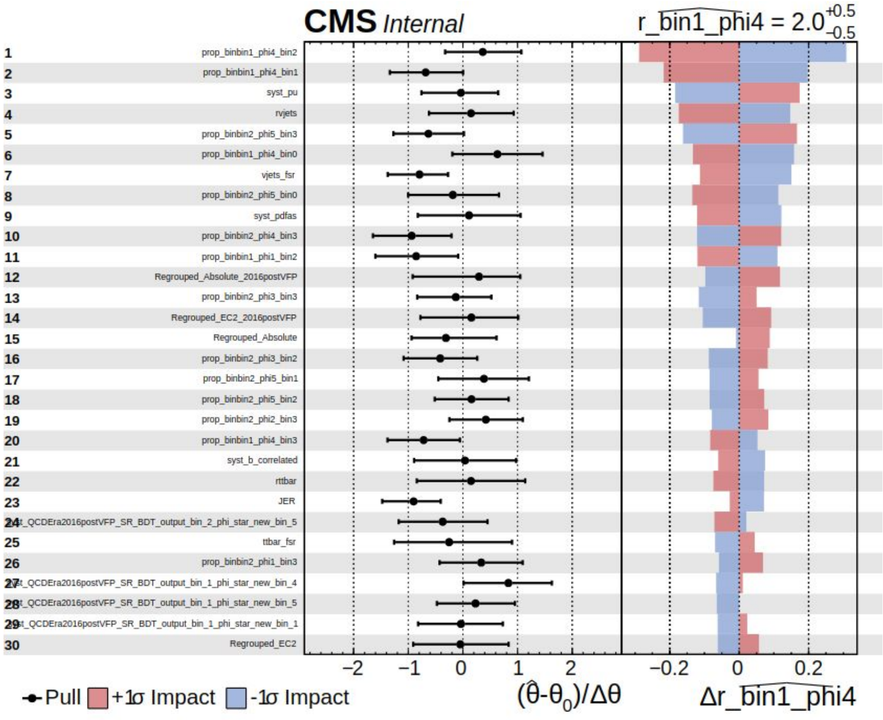
2017 results

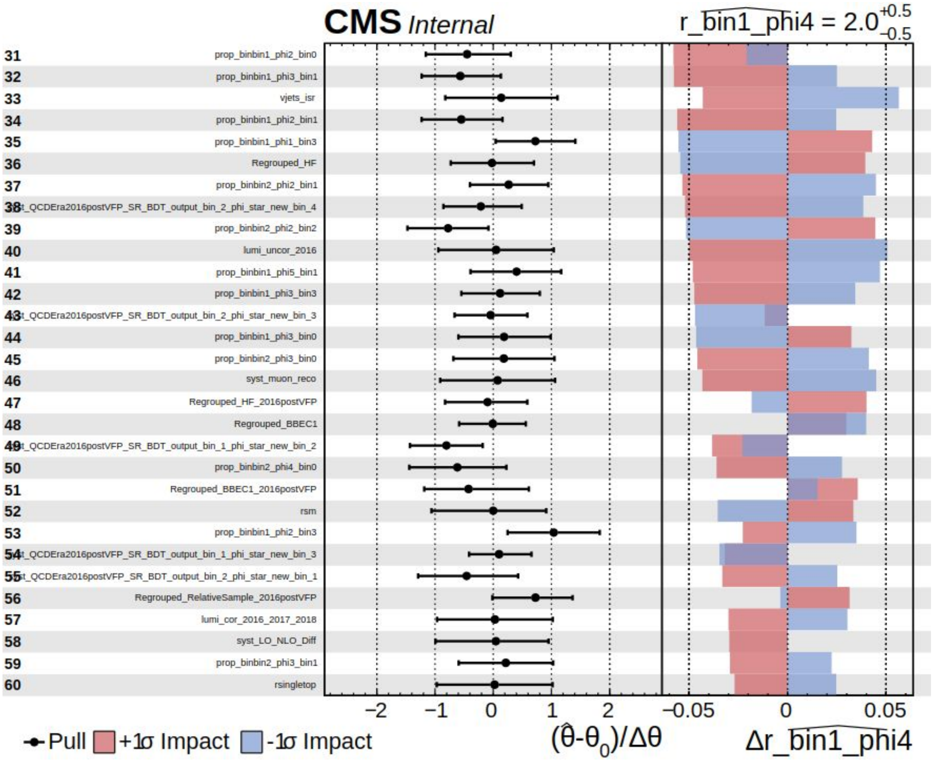




Fit diagnostics

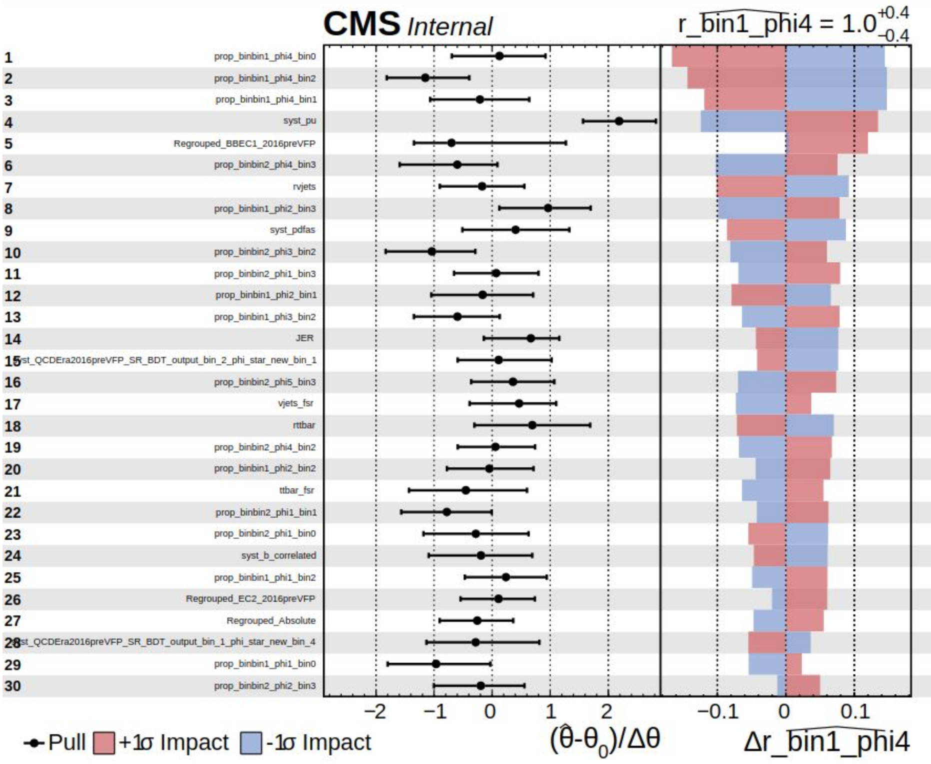
2016post results

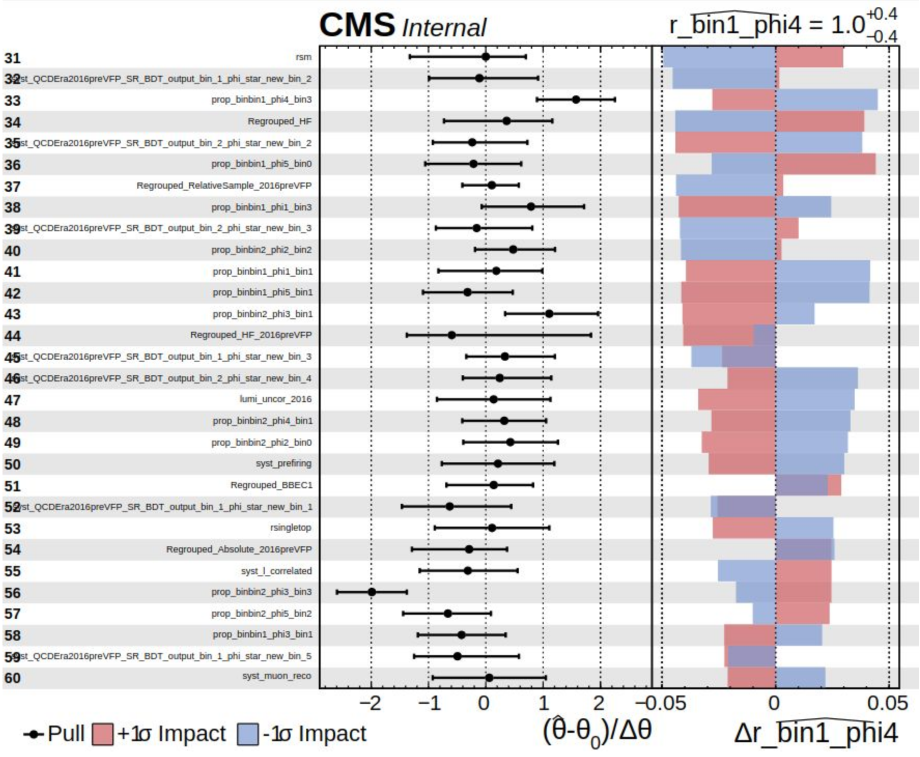




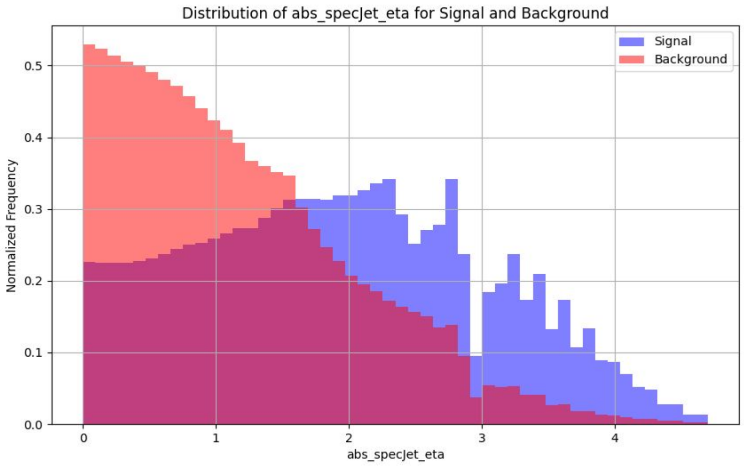
Fit diagnostics

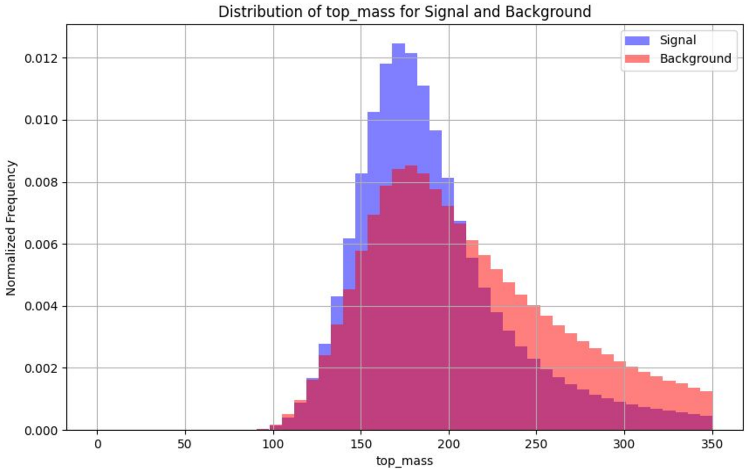
2016pre results

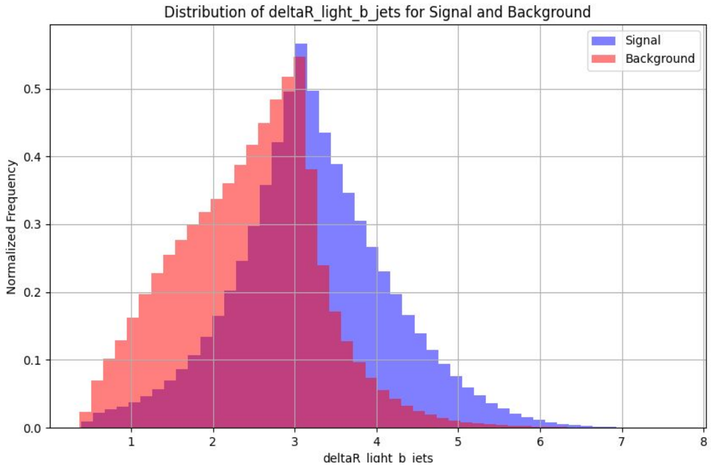




Distribution of input variables for the SR BDT







Binning optimization with significance

Formula used to evaluate significance (Poisson-Poisson model) taken from ATL-PHYS-PUB-2020-025 :

$$\sqrt{\sum_i \left(n_i \ln\left[\frac{n_i(b_i+\sigma_i^2)}{b_i^2+n_i\sigma_i^2}\right] - \frac{b_i^2}{\sigma_i^2}\ln\left[1+\frac{\sigma_i^2(n_i-b_i)}{b_i(b_i+\sigma_i^2)}\right]\right)}$$

n = QCD + non QCD

b = non QCD

$$\boldsymbol{\sigma} = \sqrt{\sum evWeight^2}$$

Optimization performed with the differential evolution algorithm from scipy.optimize (evolutionary algorithm)