Search for tWZ production with Run 2 and Run 3 data

Alberto Belvedere, Dennis Schwarz, Katerina Lipka, Roman Kogler

DESY top meeting
21 January 2025

Introduction

- CADI: TOP-24-009
- TWZ analysis twiki
- Analysis note: AN-2024/266
- Paper draft
- Previous presentation:
  - tX meeting Feb. 21st - CADI talk Oct. 9th
  - tX meeting May 29th - tX meeting Nov. 29th
- Object review:

Category	Status	Link to PubTalk	Comments
TRIG	Signed off	Link	
EGM		Link	Waiting for approval of lepton SFs
MUO	Signed off	Link	
JME	Signed off	Link	
BTV	Signed-off	Link	
STAT	To be submitted		
Combine	Signed-off	Link	

DRAFT
CMS Paper

The content of this note is intended for CMS internal use and distribution only

2025/01/14
Archive Hash: d005273-D
Archive Date: 2025/01/14

Measurement of tWZ production using Run 2 and 3 data

The CMS Collaboration

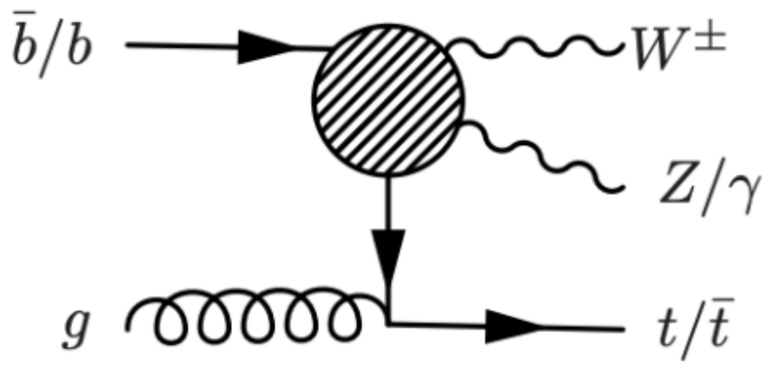
Abstract

This article presents a measurement of single top quark production in association with a W and a Z boson in proton-proton collisions. Using data corresponding to an integrated luminosity of 138 fb^{-1} recorded with the CMS detector at the CERN LHC at a cener-of-mass energy of 13 TeV together with 13.6 TeV data corresponding to an integrated luminosity of 61.9 fb^{-1}, a detailed analysis is performed to identify and characterize tWZ events. Events are selected if they contain three or four charged leptons, which can be electrons or muons. State-of-the-art machine learning algorithms and sophisticated reconstruction methods result in an unprecedented sensitivity to tWZ production. The measurement results in a statistical significance of XX standard deviations, with 3.5 expected.

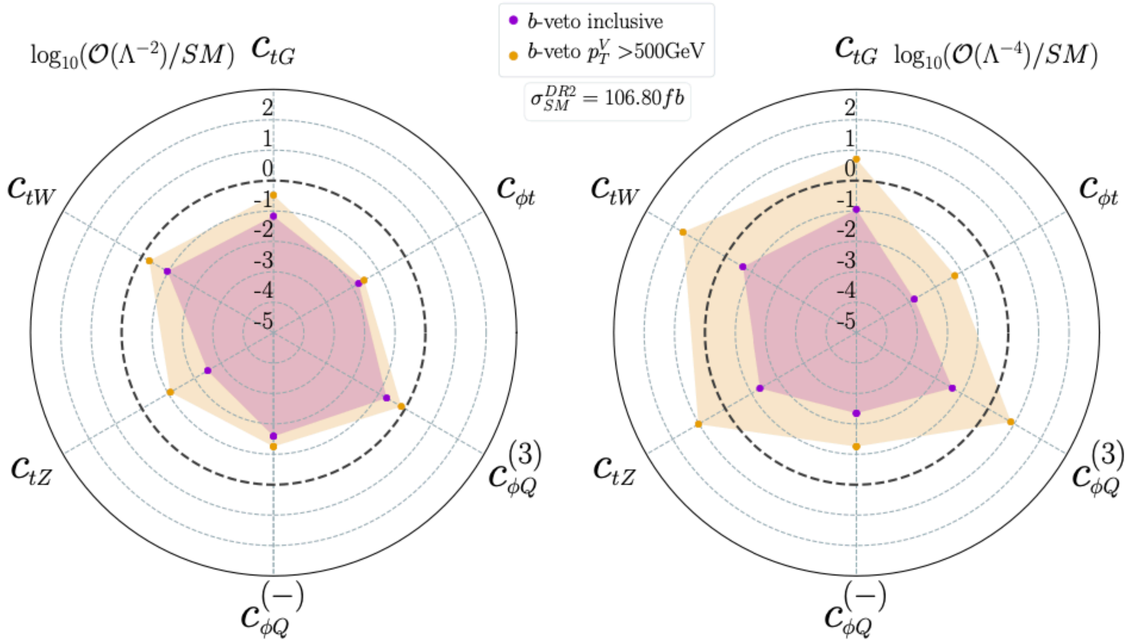
tWZ in multi-lepton final state analysis

arXiv:1904.05637 arXiv:2111.03080

- tWZ has **never been observed**, only the evidence has been published (TOP-22-008).
- Its discovery would allow the analysis of SMEFT operators that are only very loosely constrained.
- New physics could arise from **bW → tZ** vertices.
- This would lead in **energy growth** for some specific operators.



Analysis on Run2 + 2022 and 2023 data



Analysis strategy

The tWZ process has a very **small cross-section** but it's possible to exploit **multi-lepton final states** and **b-tag information** to increase signal to background ratio.

Most important **background** processes are ttZ, WZ, and TZQ:

- **Similar final state** but much **larger cross-section**.
- Large **normalisation uncertainties**.

→ **Cutting-edge ML algorithms** are employed to separate signal from background.
→ **Dedicated CRs** to constrain the normalisation of such backgrounds.

Other sources of background are DY and TT plus an additional **fake lepton**:

- Fake leptons are **not perfectly modelled** by the MC

→ **Dedicated CRs to fit MC to data** and improve the fake lepton description.

2022 and 2023 data-taking years are analysed together with **Run 2** to exploit $\approx 200\, fb^{-1}$ and compensate for the small tWZ cross-section.

Object selection and correction

Run 2

Leptons

- p_T > 25,15,10,(10) GeV
- $|\eta|$ < 2.5 (2.4)
- ID: TopLeptonMVA (e tight, μ medium)

Jets

- p_T > 25 GeV
- $|\eta|$ < 2.5
- ID: Tight
- B-tag: DJ medium

Run 3

Leptons

- p_T > 25,15,15,(15) GeV
- $|\eta|$ < 2.5 (2.4)
- ID: PromptMVA (e tight, μ medium)

Jets

- p_T > 25 GeV
- $|\eta|$ < 2.5
- ID: Tight
- B-tag: ParT medium

Corrections

- Pile-up reweighting
- L1 pre-firing (only in 2016 and 2017)
- Leptons ID SFs
- Muon Rochester corrections (only in Run 2)
- Electron Scale&Smearing (not in 2023)
- Electron reconstruction SFs
- B-tag SFs
- JEC
- JER
- Jet veto maps (only in Run 3)
- WZ njet reweighting

Regions definition

Signal regions

Semi-leptonic tWZ

- At least 2 jets
- At least 1 b-jet
- Exactly 3 leptons
- $|m(ll)-m(Z)| < 15$ GeV

Fully-leptonic tWZ

- At least 1 jet
- At least 1 b-jet
- Exactly 4 leptons
- $|m(ll)-m(Z)| < 15$ GeV

Control regions

WZ CR

- Exactly 3 leptons
- 0 b-jet
- $|m(ll)-m(Z)| < 15$ GeV
- MET > 40 GeV

DY NP CR

- Exactly 3 leptons
- 0 b-jet
- $|m(ll)-m(Z)| < 15$ GeV
- MET < 40 GeV

ZZ CR

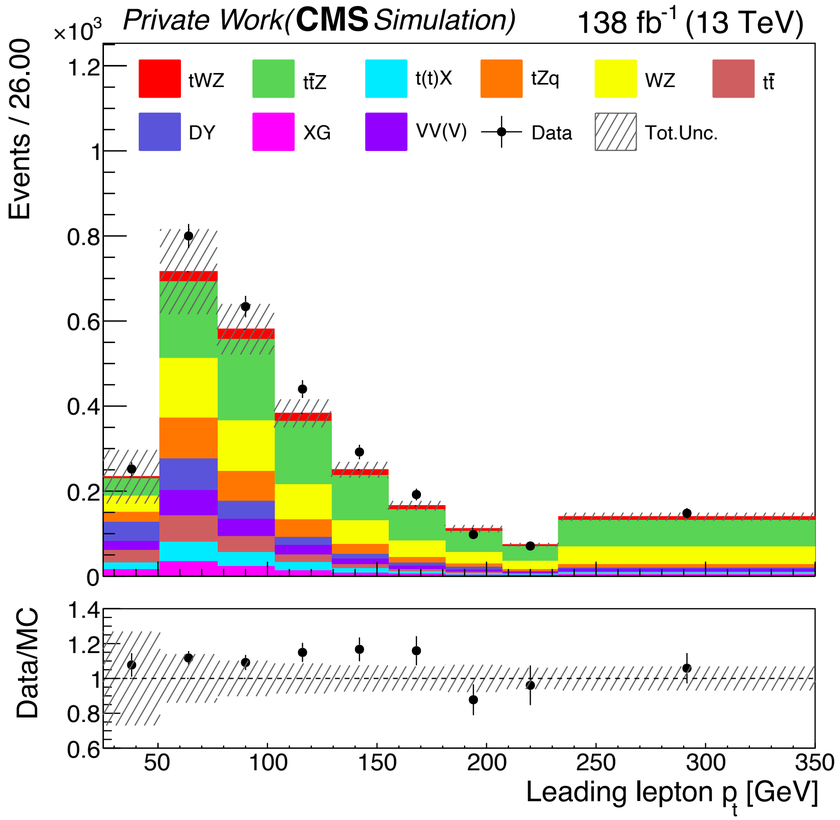
- At least 1 jet
- At least 1 b-jet
- Exactly 4 leptons
- $2x|m(ll)-m(Z)| < 15$ GeV

ttX NP CR

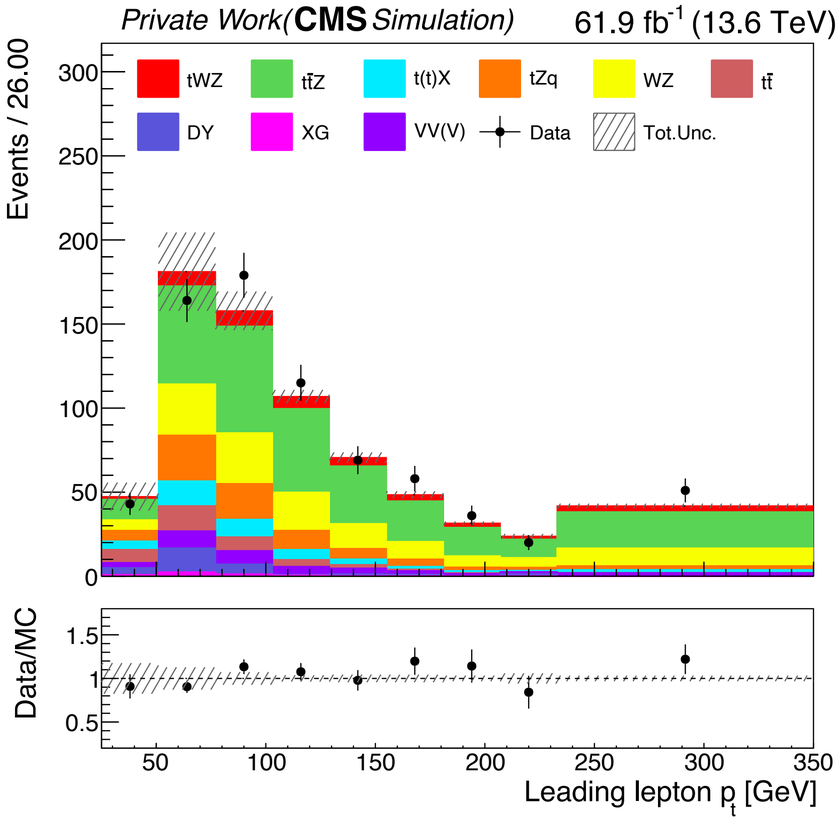
- At least 2 jet
- At least 1 b-jet
- Exactly 3 leptons
- $|m(ll)-m(Z)| > 15$ GeV

Semi-leptonic signal region distributions

Run 2



Run 3

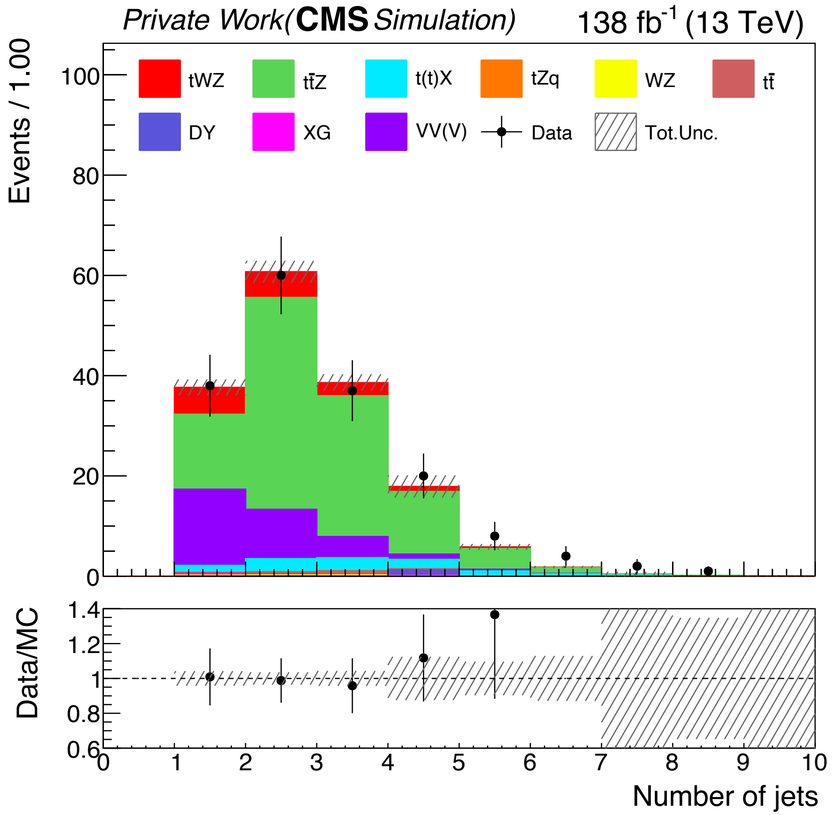


Good agreement in both data-taking periods, with a small excess of data visible in Run 2.

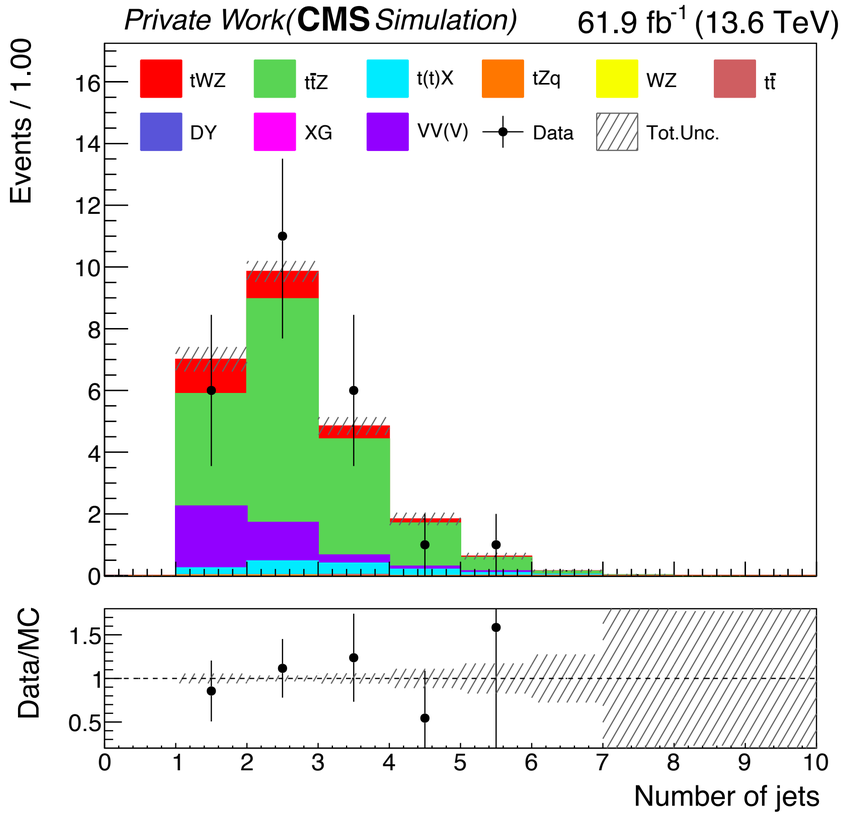
The background is overwhelming, with the main contribution being $\mathbf{t\bar{t}Z}$.

Fully leptonic signal region distributions

Run 2



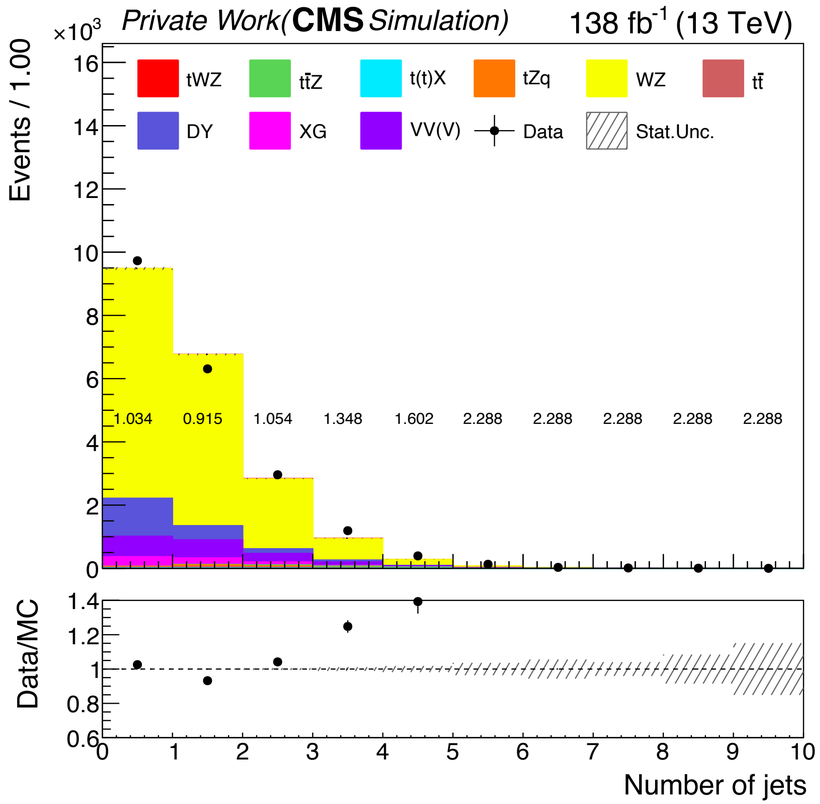
Run 3

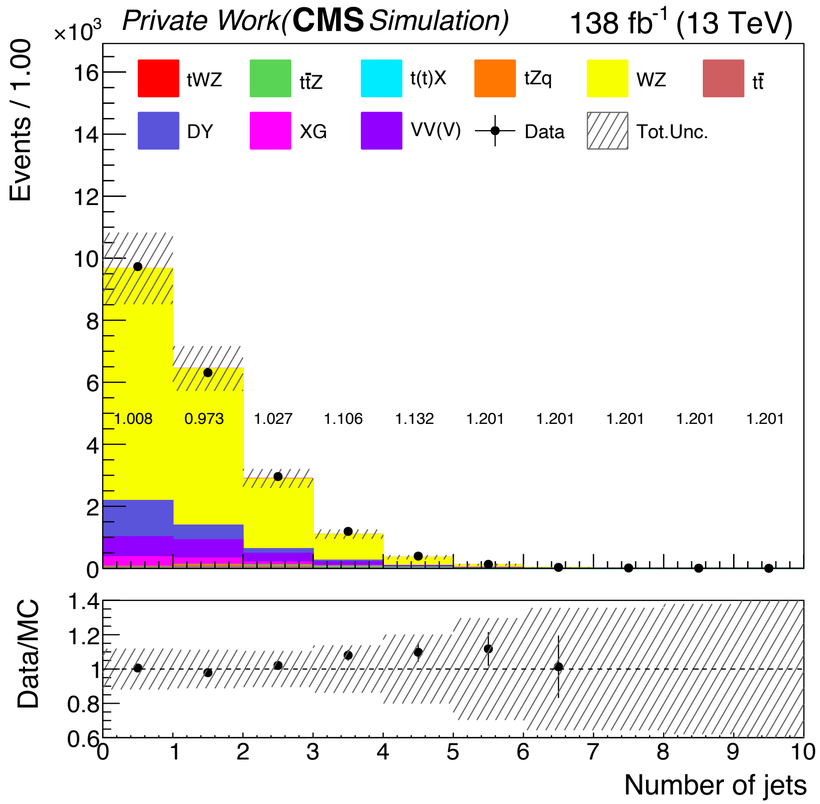


Good agreement in both data-taking periods.

The background is overwhelming, with the main contribution being $t\bar{t}Z$.

WZ CR and njet reweighting in Run 2

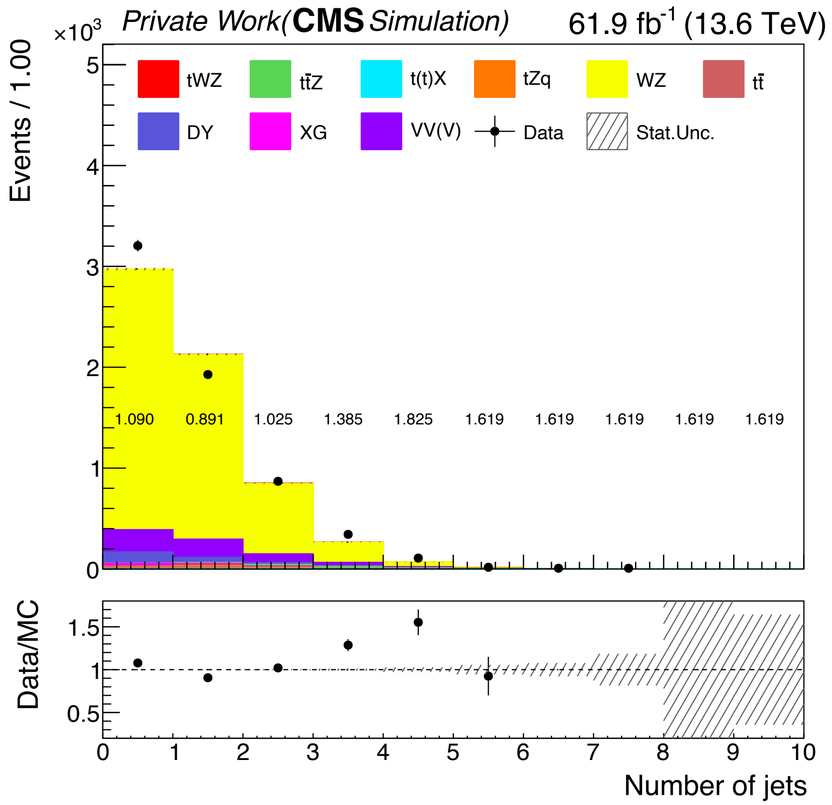


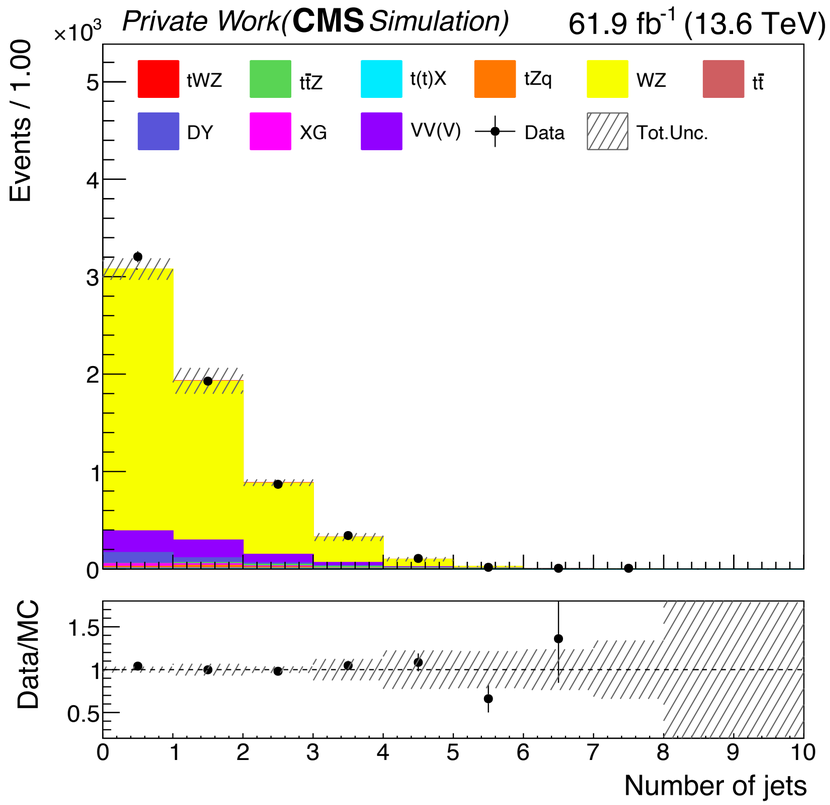


Effect of the WZ njet reweighting in Run 2

A **systematic uncertainty** is applied in the following way:
-**down**: the distribution we would have if we didn't apply the SFs
-**up**: the distribution we would have if we applied twice the SFs

WZ CR and njet reweighting in Run 3

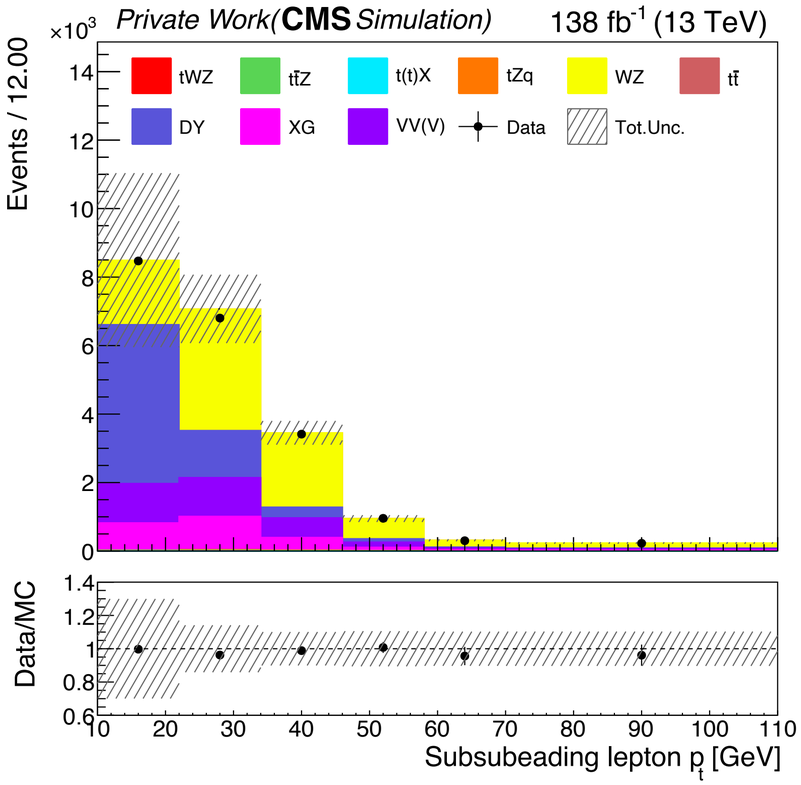




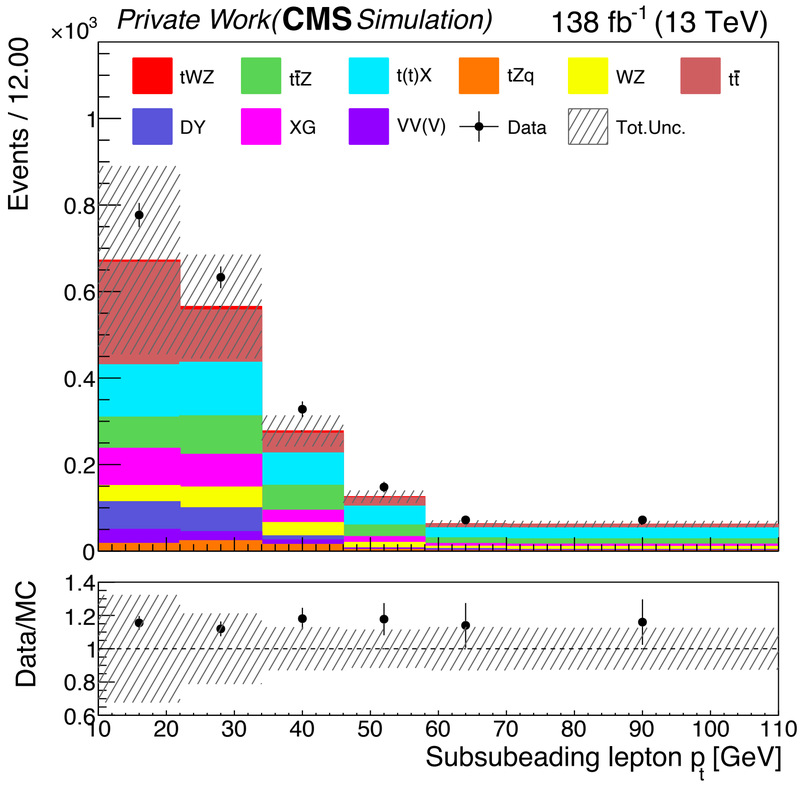
Slightly smaller mis-model in Run3,
but similar agreement after the application of the SFs.

DY, ttX, and ZZ CRs in Run 2

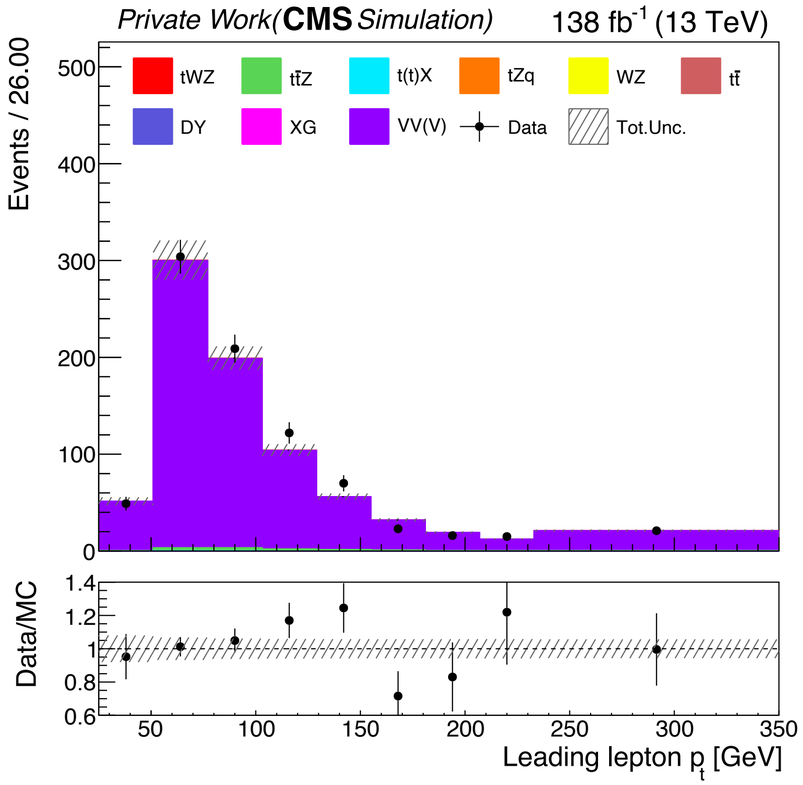
DY CR



ttX CR



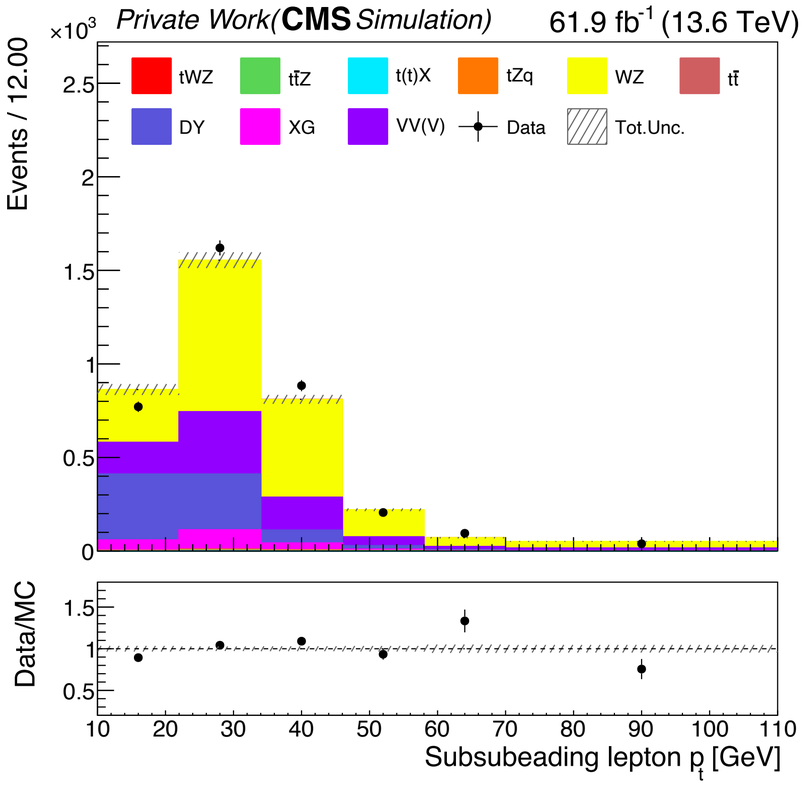
ZZ CR



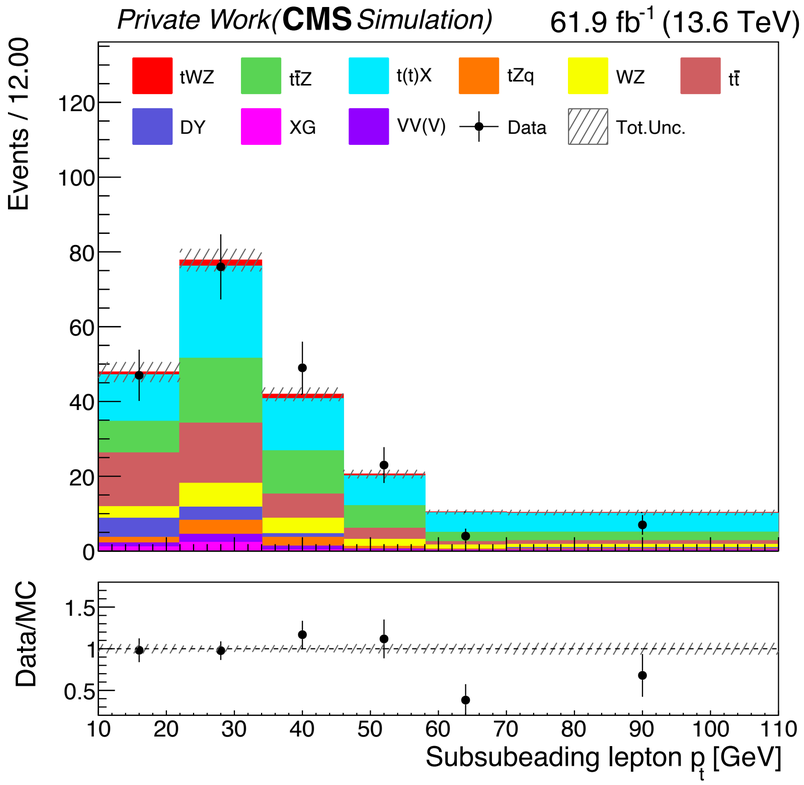
Good data-MC agreement in the all the regions.
Flat 15% excess of data in the ttX CR fixed with non-prompt estimation.

DY, ttX, and ZZ CRs in Run 3

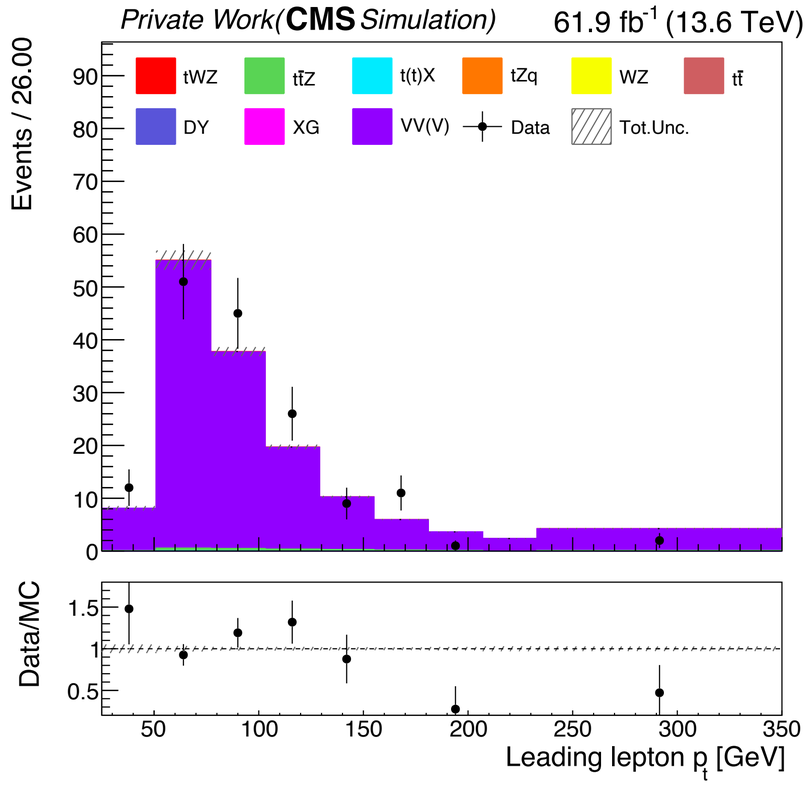
DY CR



ttX CR



ZZ CR



Good data-MC agreement in the all the regions.

The excess of data observed in Run 2 in the ttX CR is not present in Run 3, probably due to smaller contribution from nonprompt lepton backgrounds.

Run 2 ML classifier performance

Using the **Particle Transformer** algorithm to distinguish between signal and background.

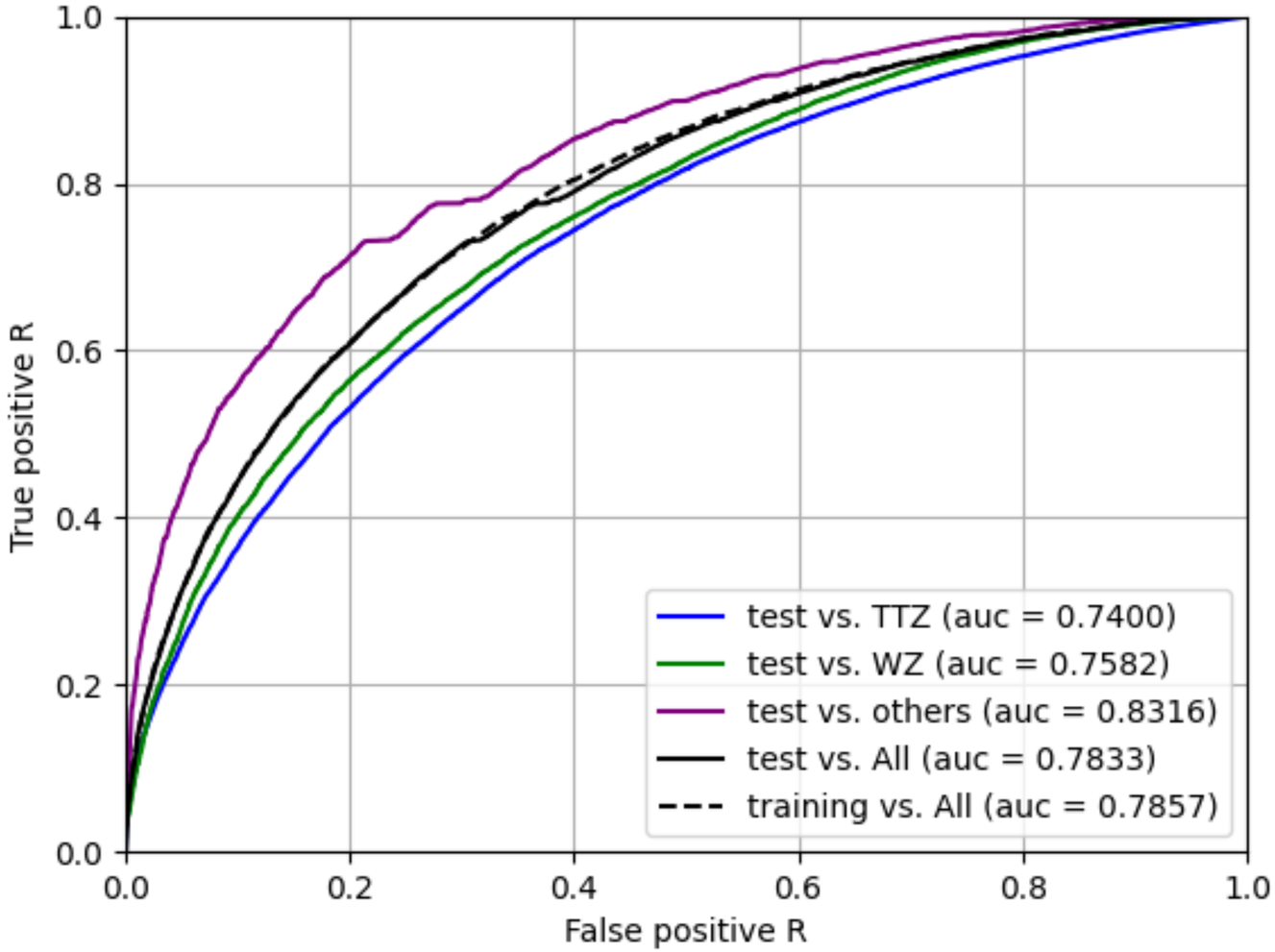
Variables used in the algorithm:

- 4-momentum of the first **6 jets**.
- 4-momentum of the **4/5 leptons**.
- log(p_T), log(E), η, **b-tag**, and pdgID.

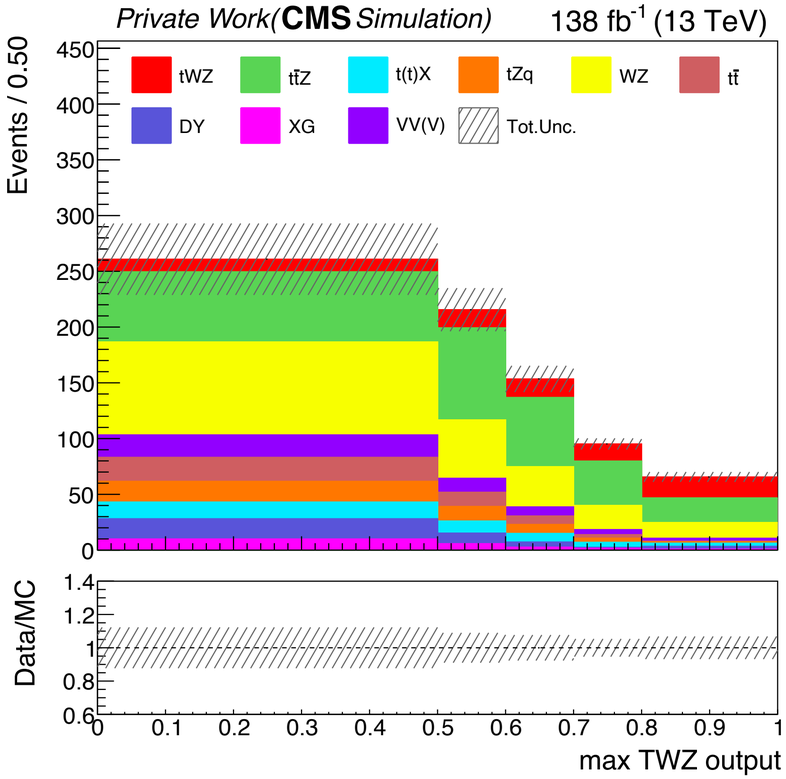
In the b-tag information inputs the different b-tag WPs are exploited.

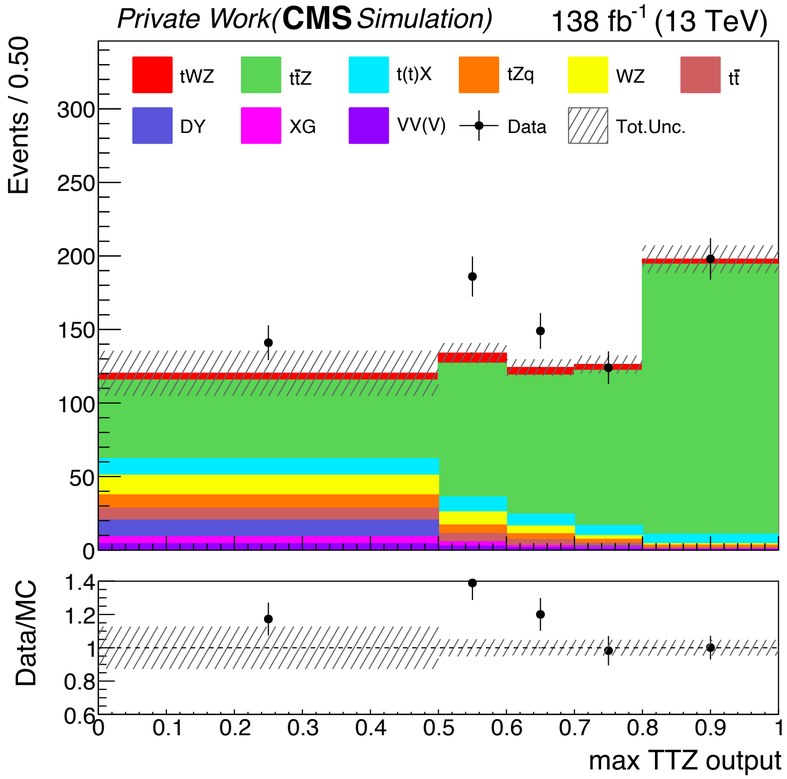
4 different output nodes:

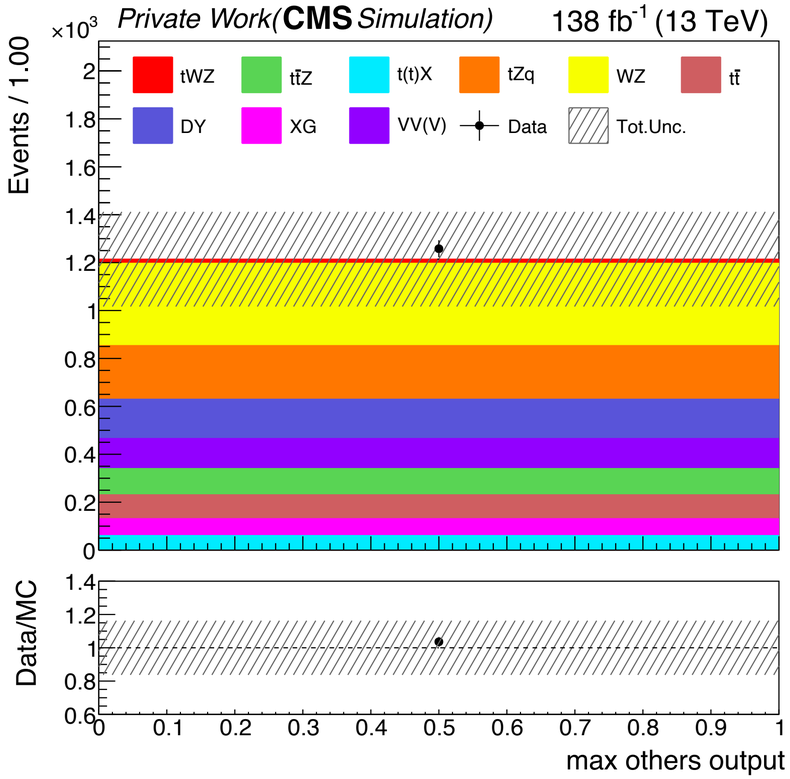
- TWZ
- TTZ
- WZ
- Other backgrounds



ML output distributions for the semi-leptonic SR







The three distributions are created by dividing the events according to the maximum value among the output scores and merging the other and the WZ node.

Run 3 ML classifier performance

The performance of the ML algorithm on **Run 3** data is very poor due to the **poor statistics.**

Asked for an extension of the current samples to reach the following statistics on Run 3:

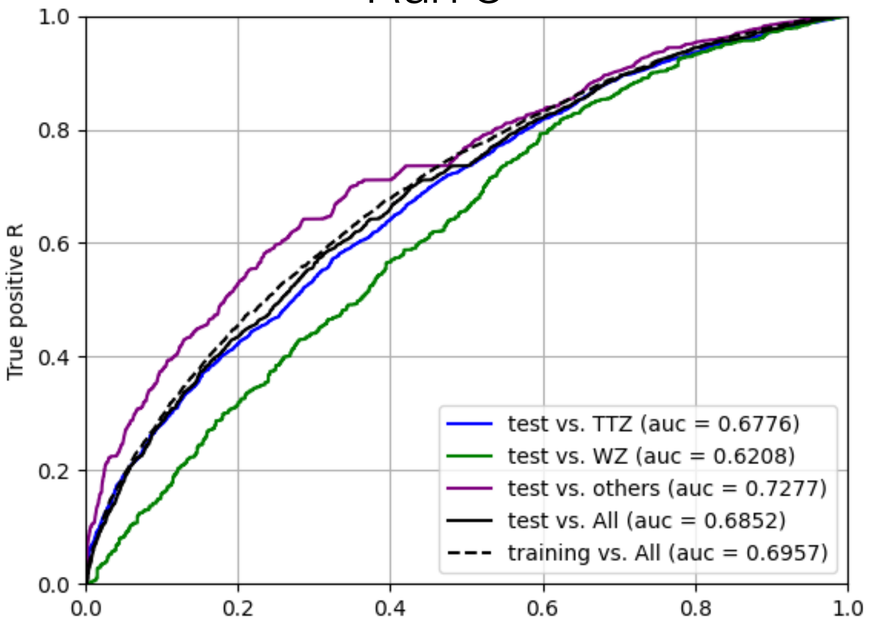
TWZ: ~10 M → ~35 M

TTZ: ~4 M → ~35 M

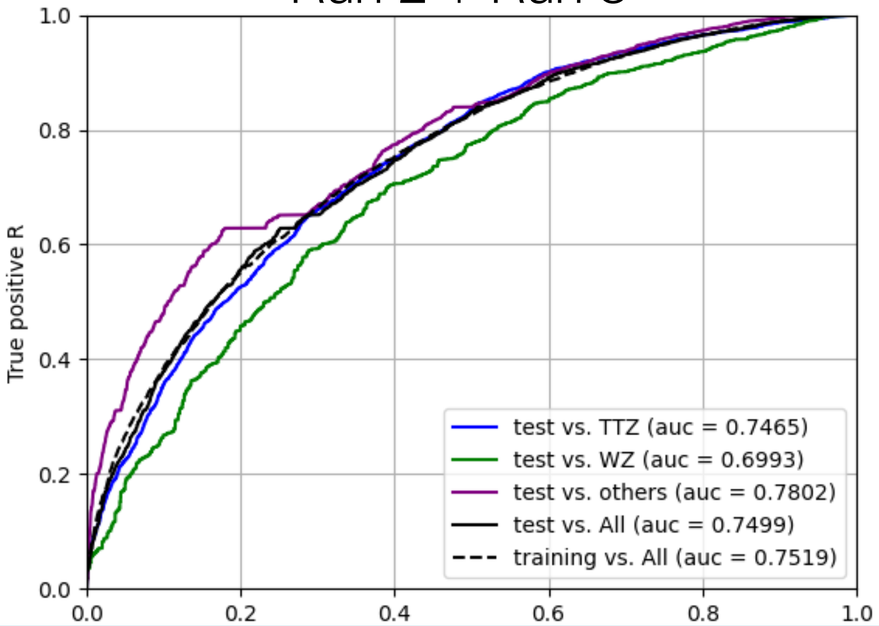
WZ: ~20 → ~95 M

For the moment running **inclusively on Run 2 and Run 3**, adding a variable to discriminate between the two data-taking periods.

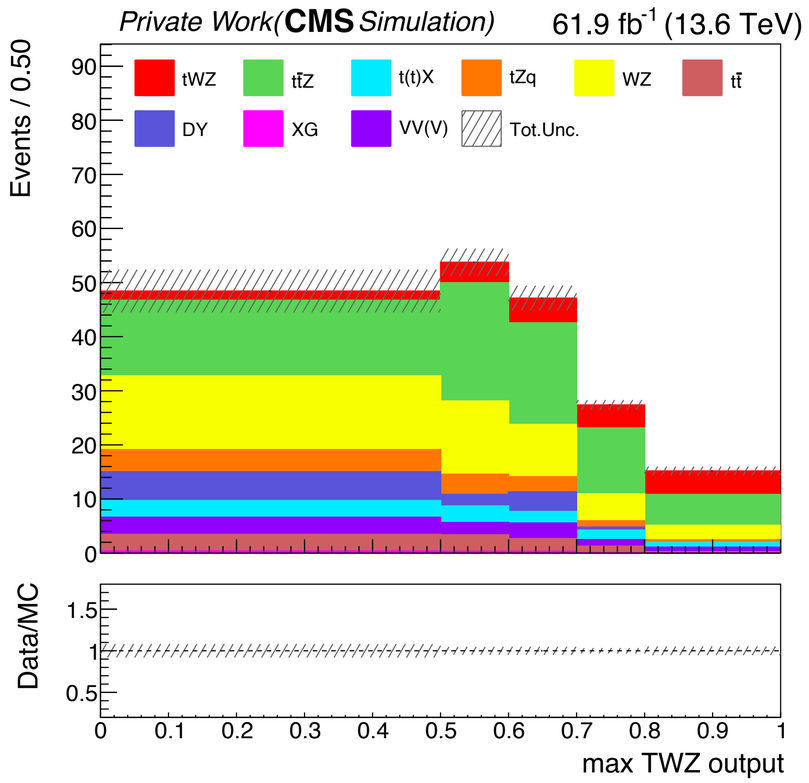
Run 3

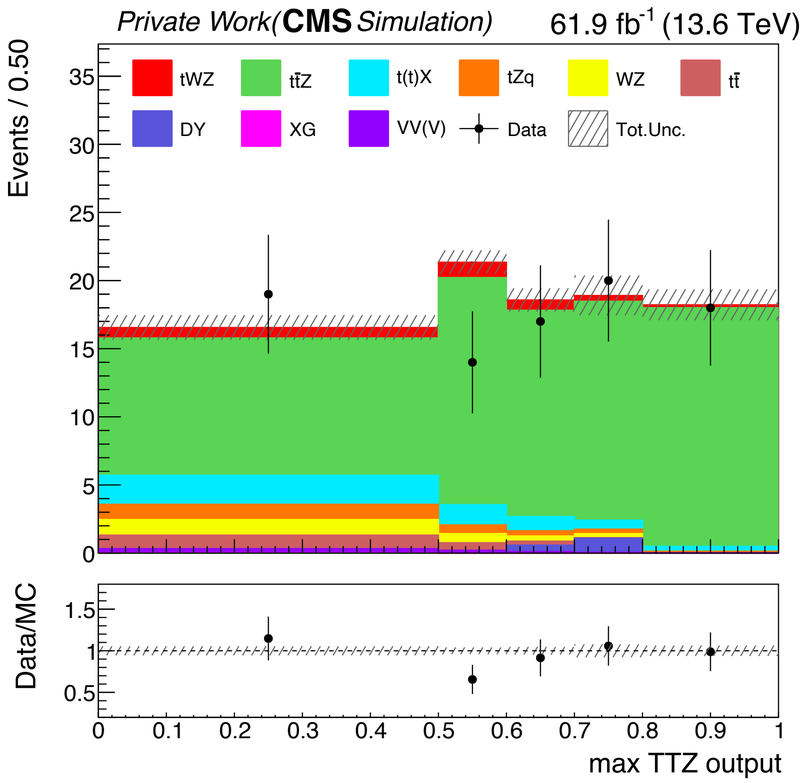


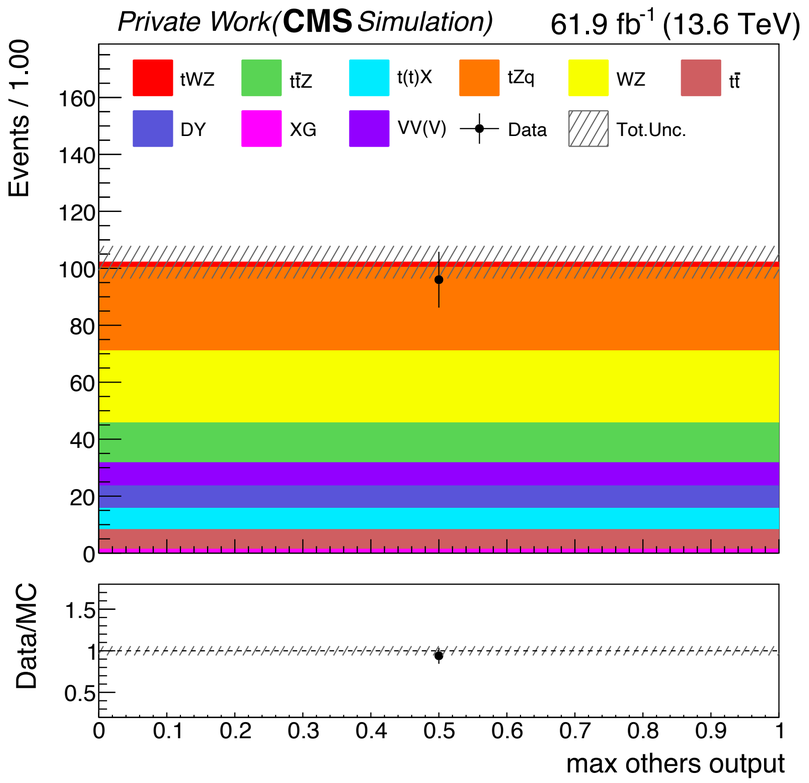
Run 2 + Run 3



ML output distributions for the semi-leptonic SR





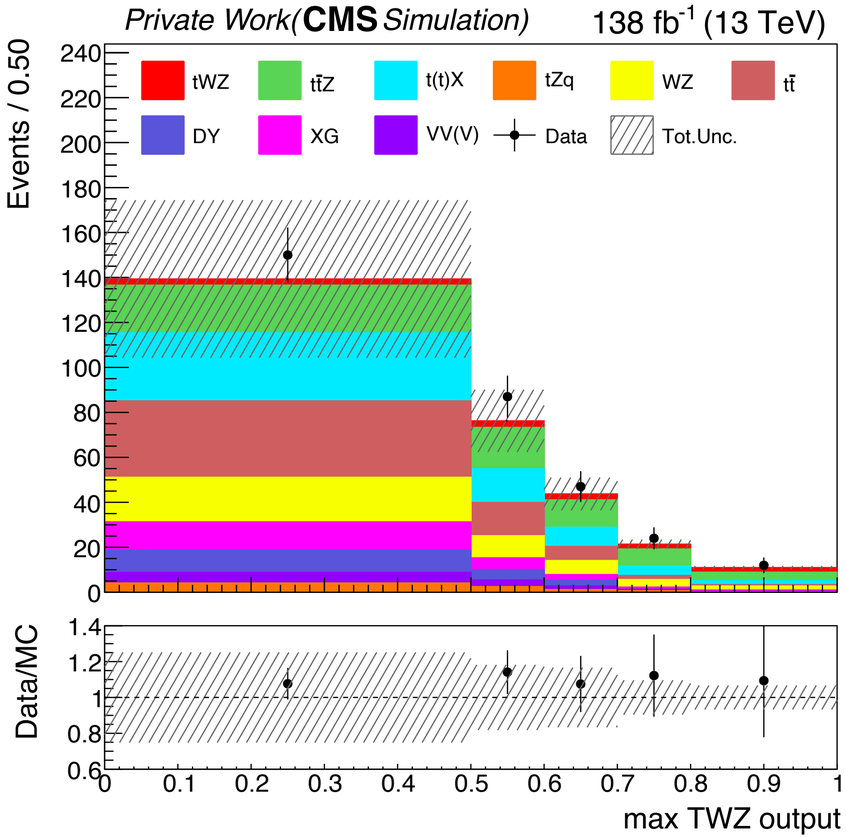


The three distributions are created by dividing the events according to the maximum value among the output scores and merging the other and the WZ node.

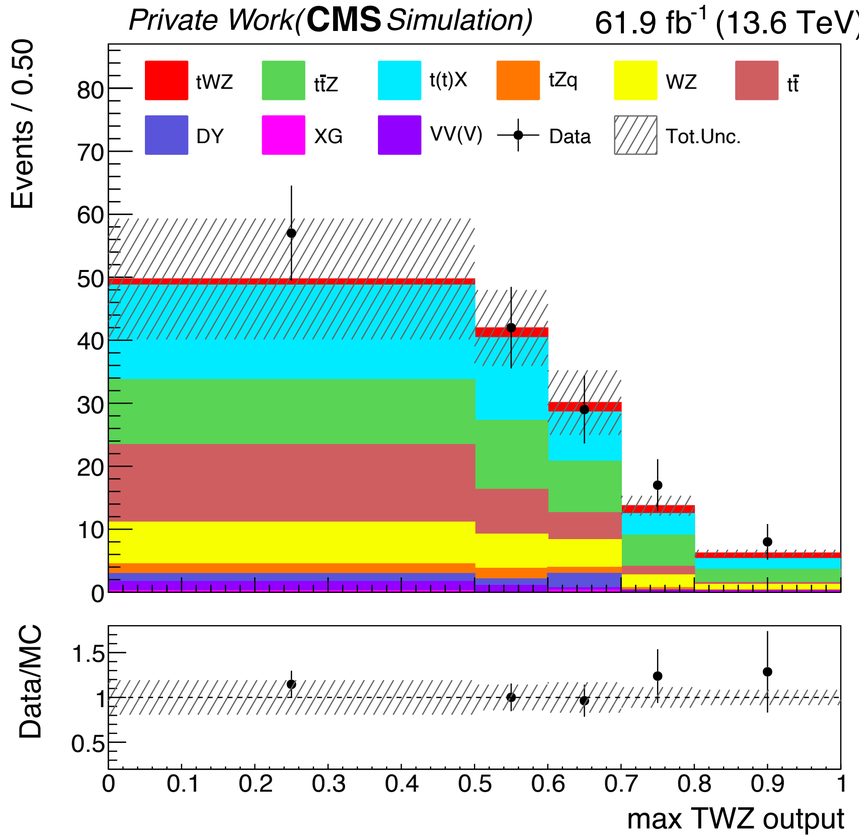
Good data-MC agreement in the TTZ and others output nodes.

Max tWZ output distributions in the ttX CR

Run 2



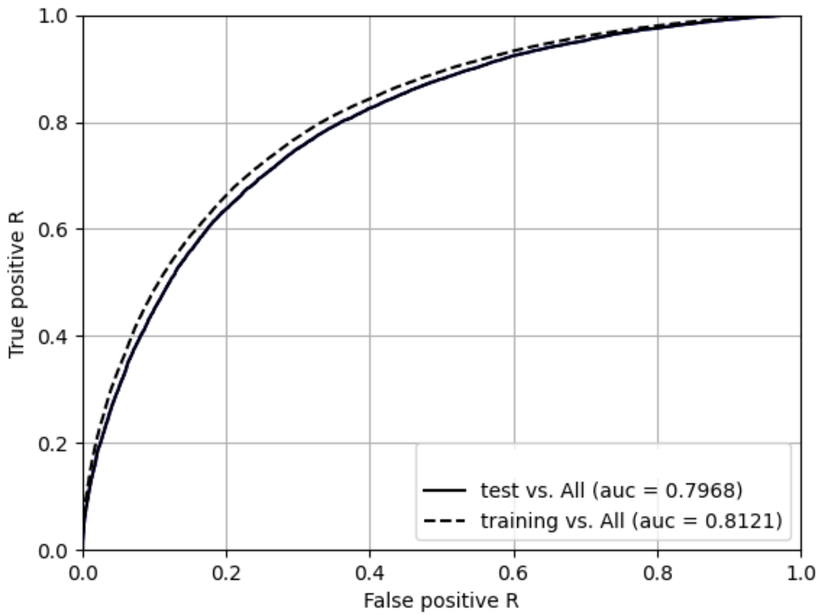
Run 3



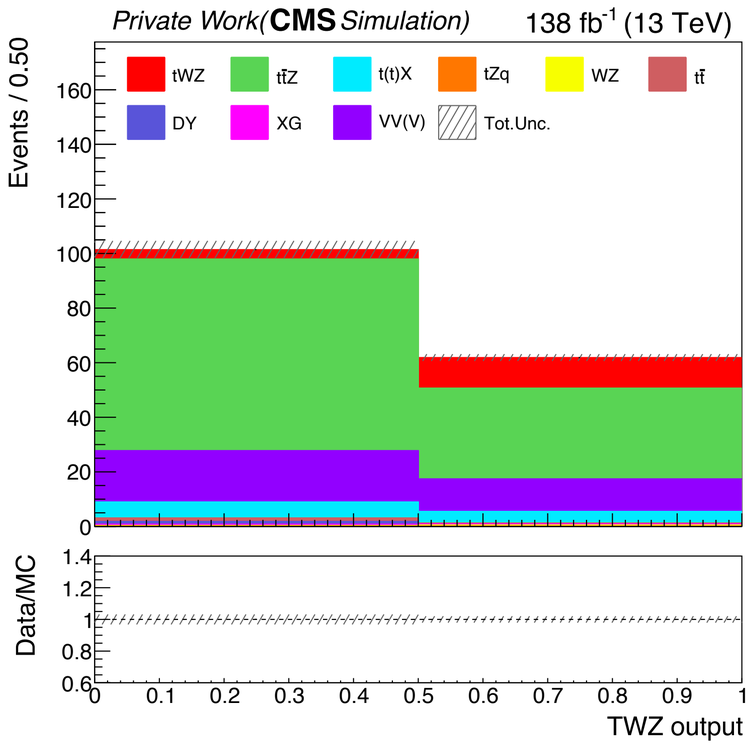
Good data-MC agreement for the max TWZ output distributions in the ttX CR for both data-taking years.

ML for fully leptonic signal region

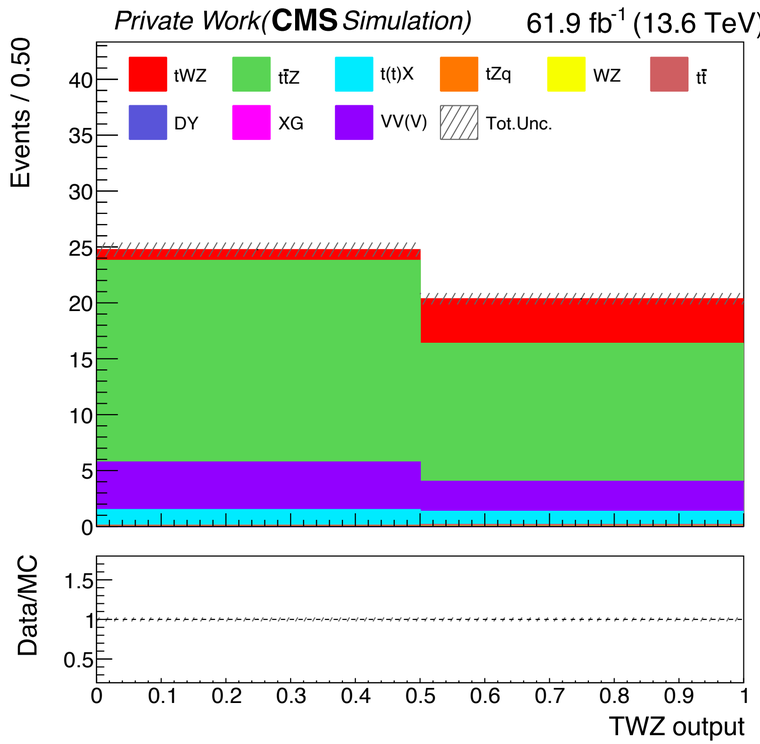
Performance



Run 2



Run 3



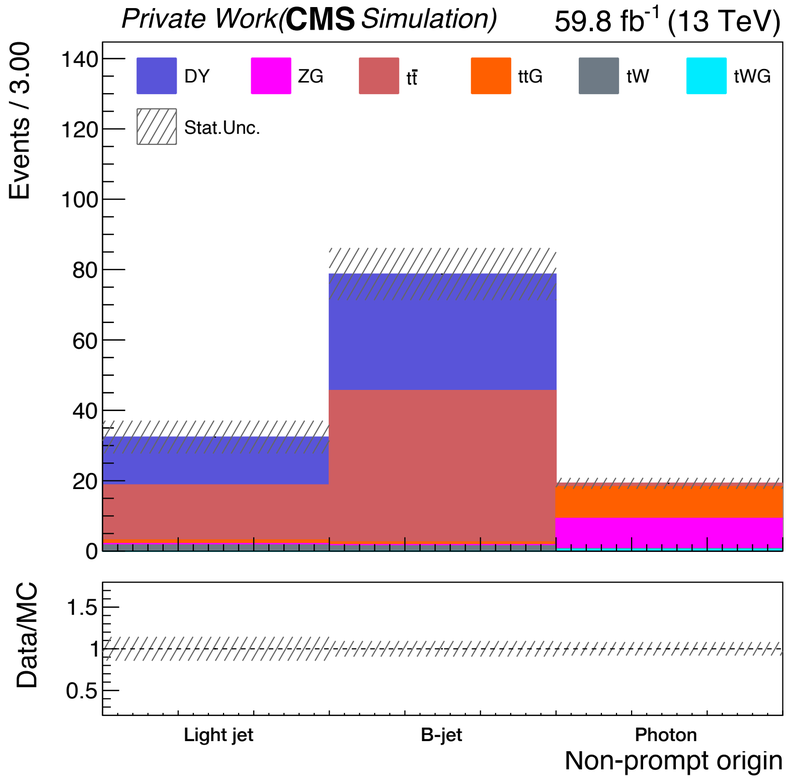
New ML algorithm trained on TWZ, TTZ, TTH, and ZZTo4L.

Fake-leptons background

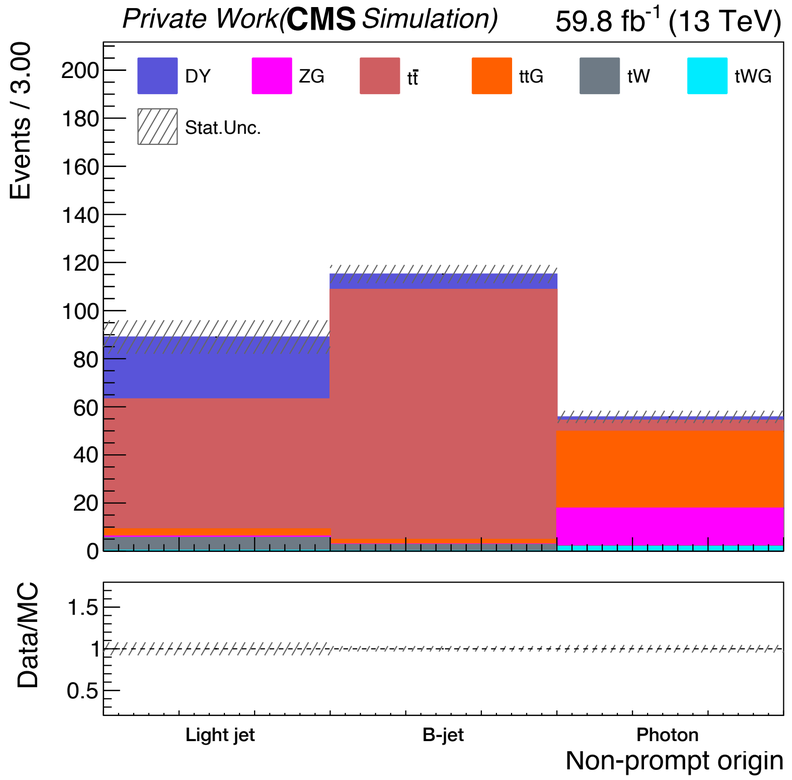
Origin of the fake lepton: light jet, b-jet, and photon.

This is obtained by matching the fake leptons at reco-level with the gen-level particles in the MC.

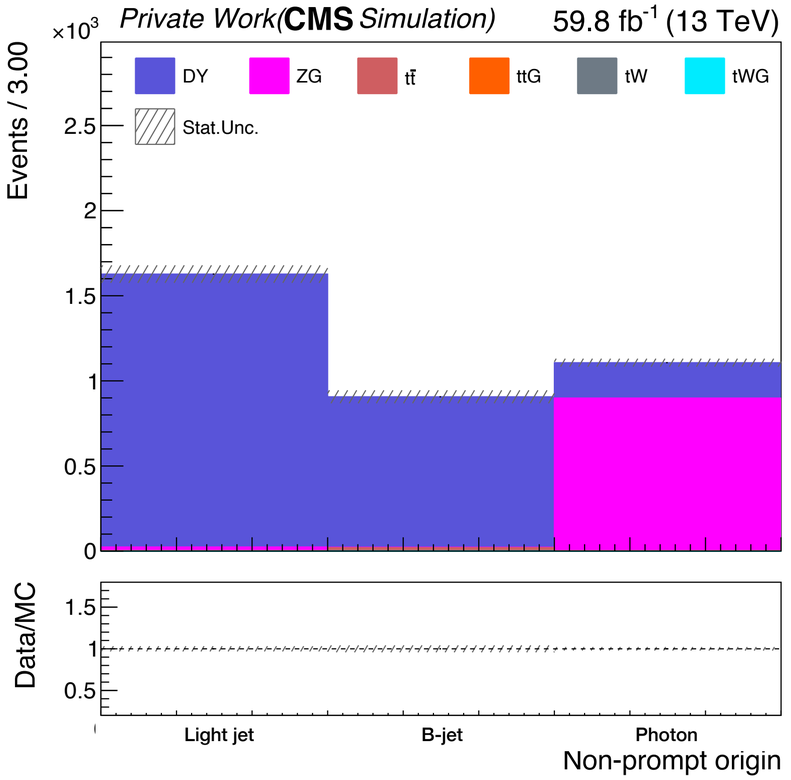
Semi-leptonic tWZ



ttX CR



DY CR



Fake-leptons background

Fake leptons estimation performed **directly in the fit** to estimate this contribution directly data.

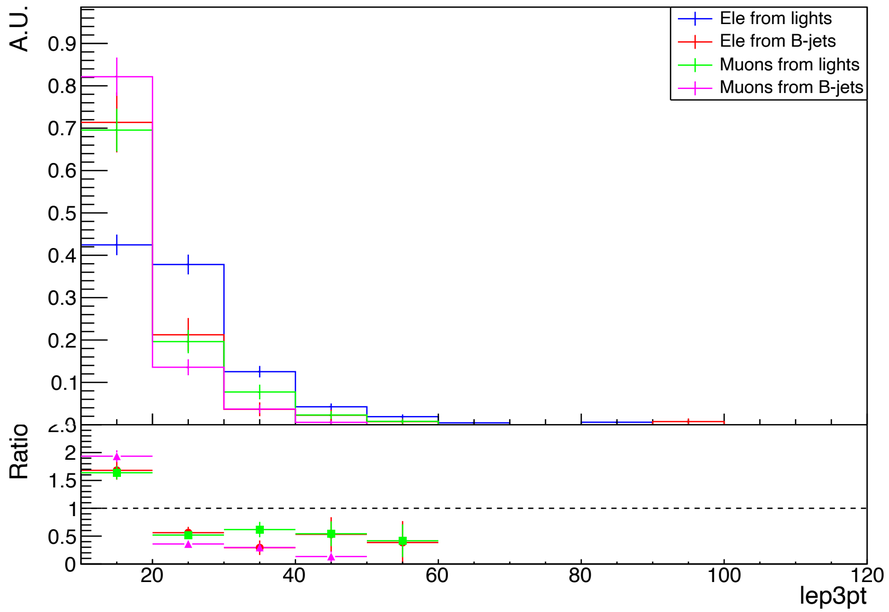
Three templates created using the **trailing lepton** distribution to correct shape and normalisation of the NP backgrounds:

- **bjetFake**: uncostrained
- **unknownFake**: uncostrained

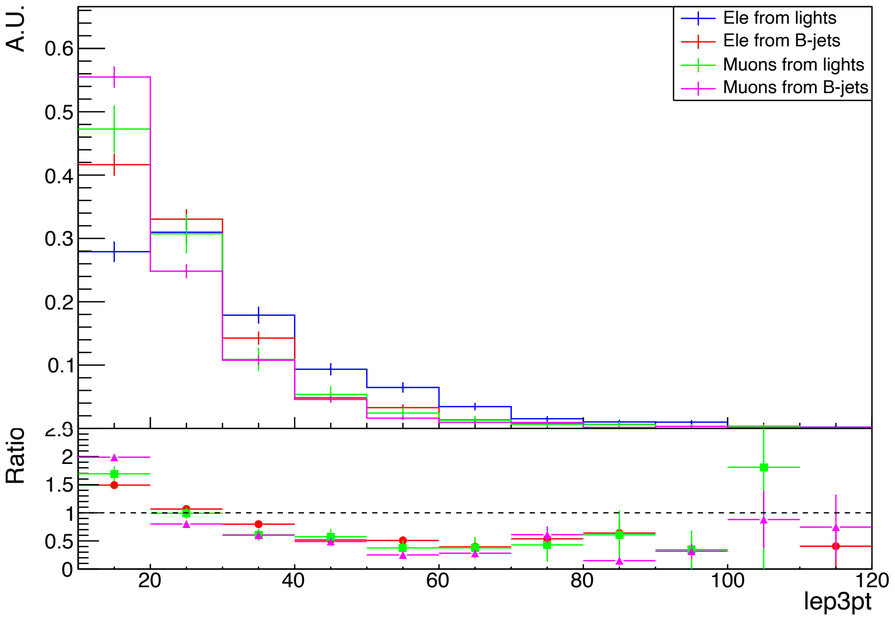
} up: 3 x nominal; down: 0 x nominal

- **photonFake**: gauss. constrain ± 0.5xnominal

DY CR



ttX CR



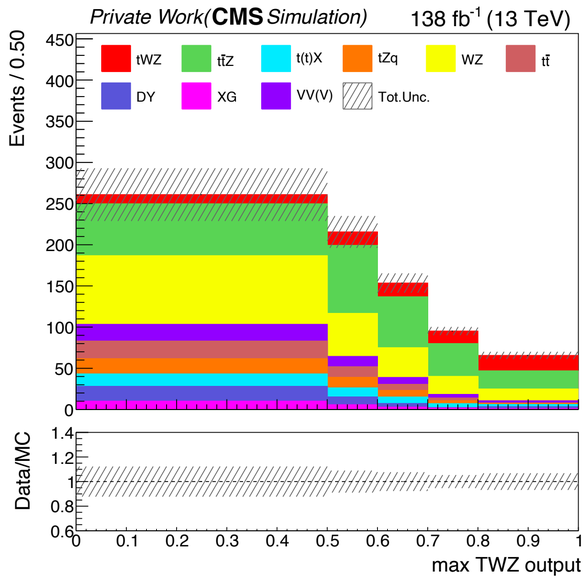
Systematic uncertainties (experimental)

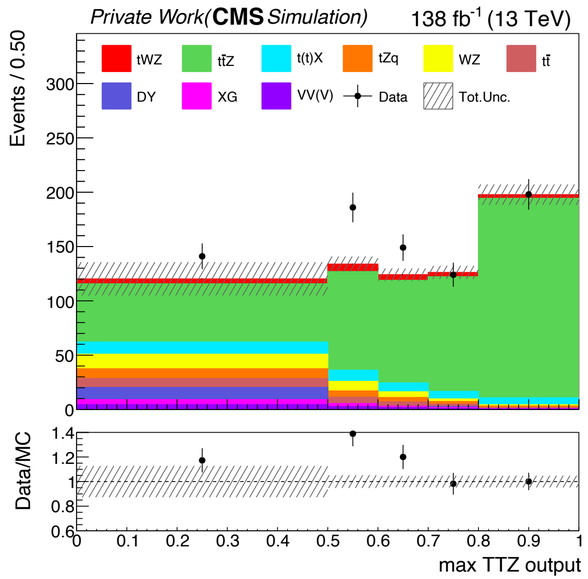
- Pileup, correlated
- B-tag ID SFs, (HF, LF)⊗(correlated, uncorrelated)
- Lepton ID SFs, correlated and uncorrelated parts
- Electron reco SFs
- Fake leptons, uncorrelated
- L1 pre-firing, correlated
- JEC (reduced splitting), correlated
- JER, uncorrelated
- Unclustered energy, uncorrelated
- Trigger, (2% flat) uncorrelated
- Luminosity, correlated and uncorrelated parts

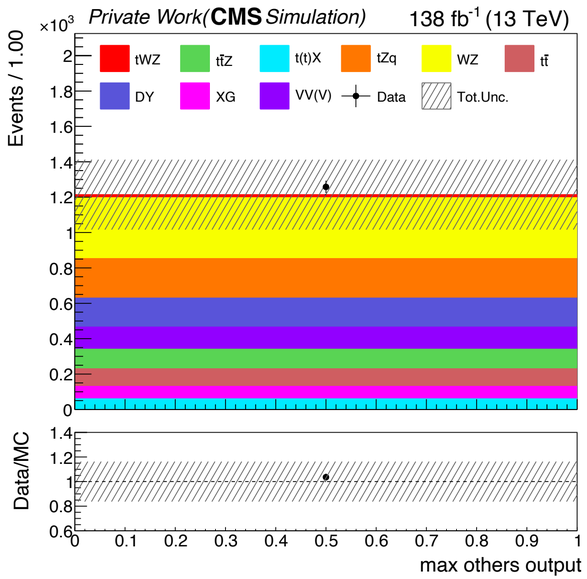
Systematic uncertainties (theoretical)

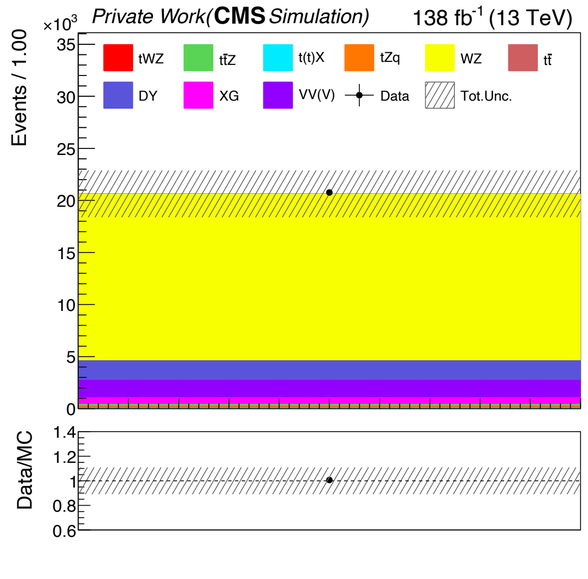
- DR modeling
- WZ njet reweighting
- QCD scales, uncorrelated across processes
- PDF, 100 variations
- ISR, uncorrelated across processes
- FSR, correlated across processes
- DY: 10%
- TTZ: 7%
- TTX: 10%
- VV(V): 10%
- TT: 5%
- TZQ: 11%
- WZ: 10%
- XG: 10%

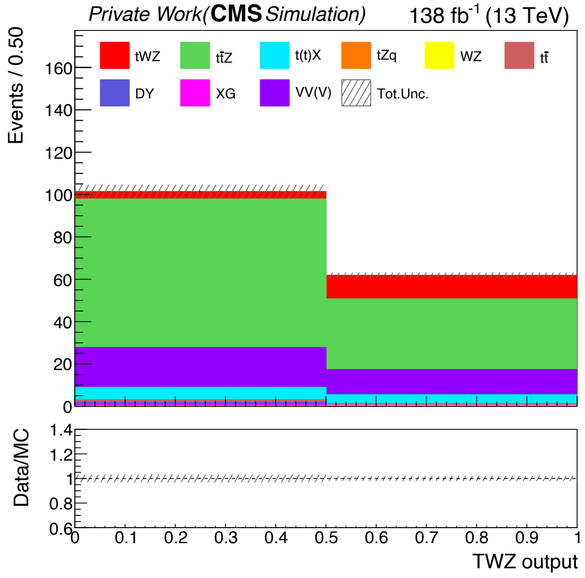
Input distributions to the fit for Run 2

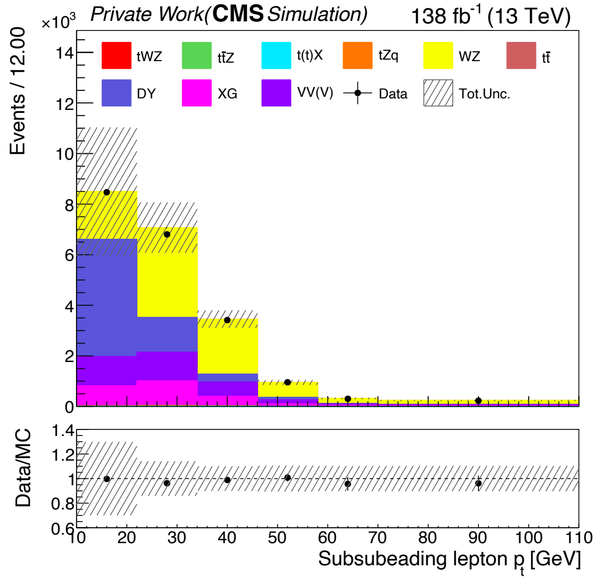


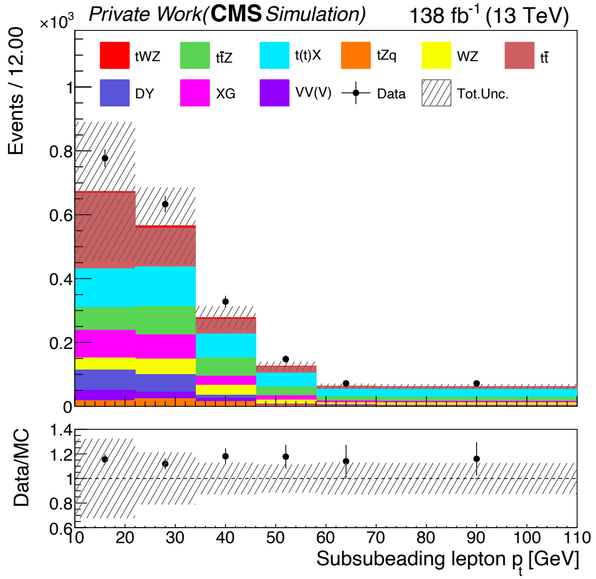


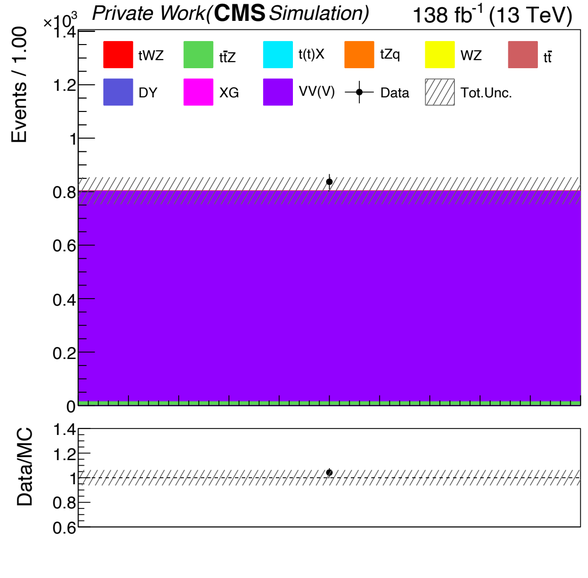




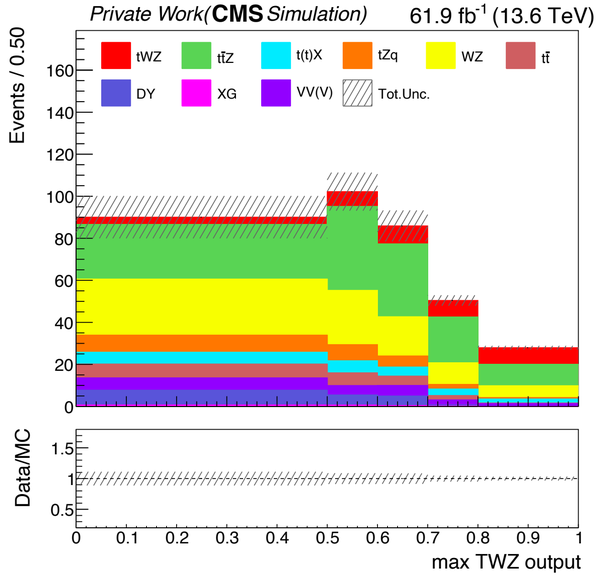


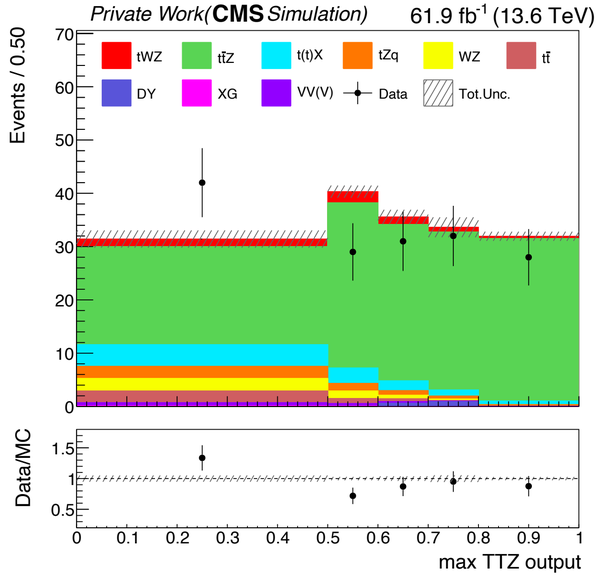


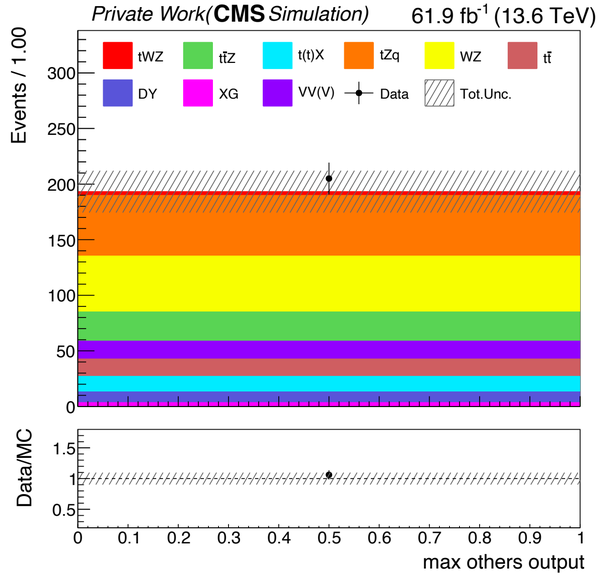


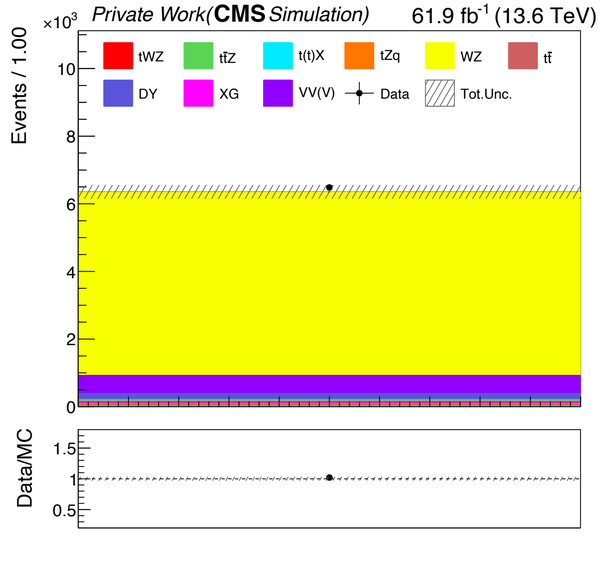


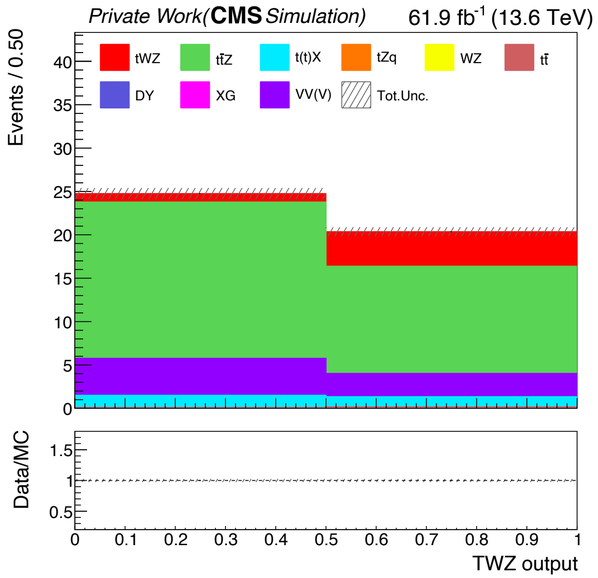
Input distributions to the fit for Run 3

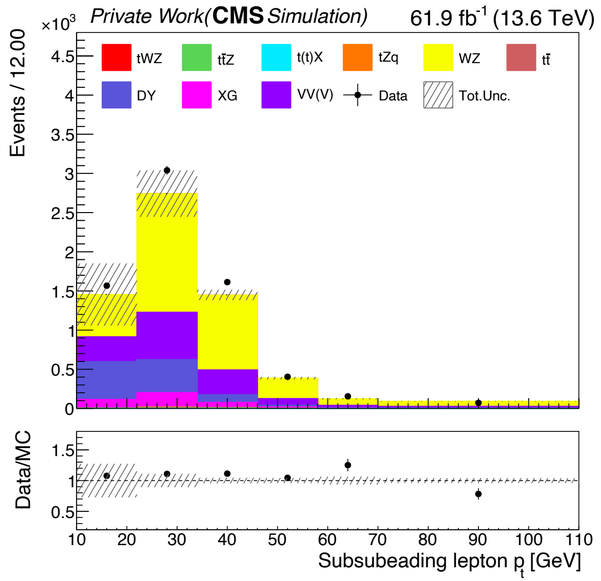


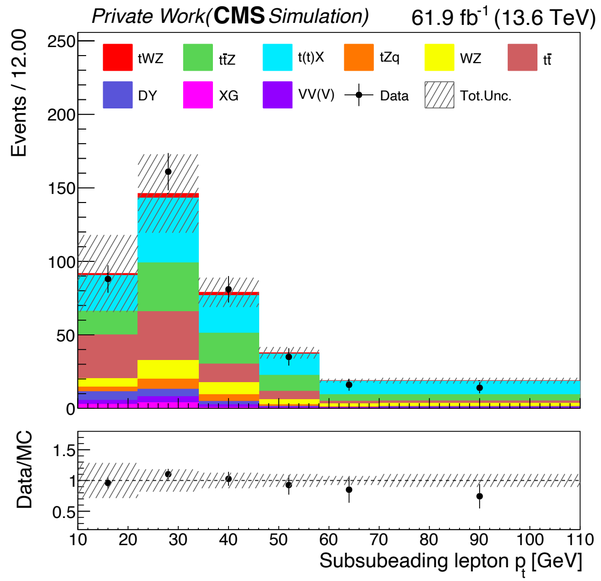


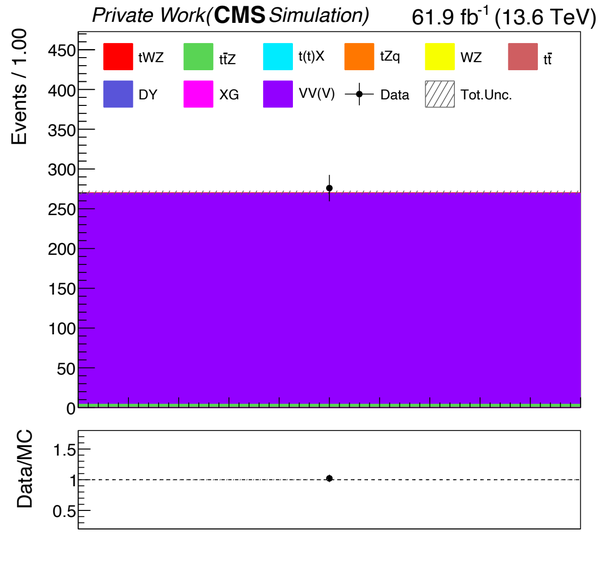






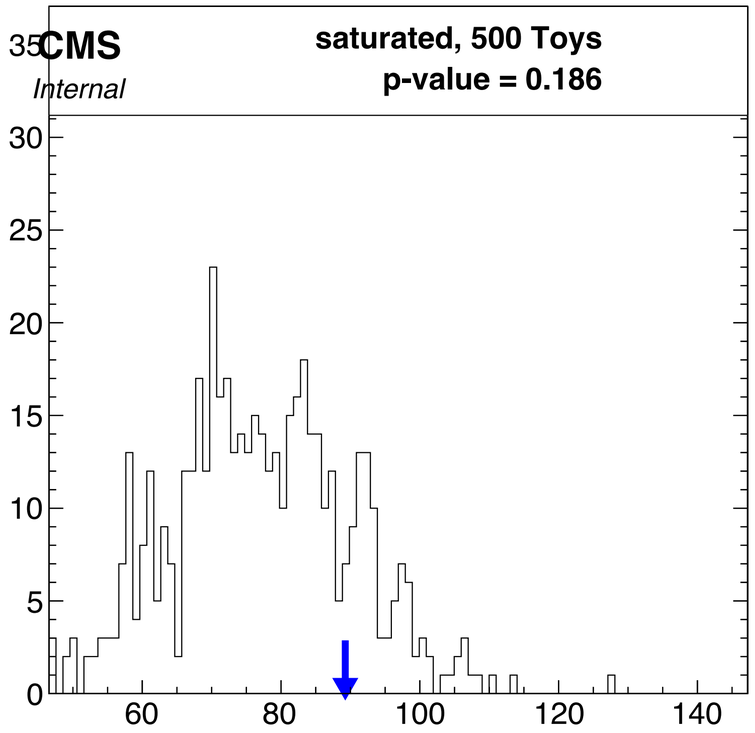


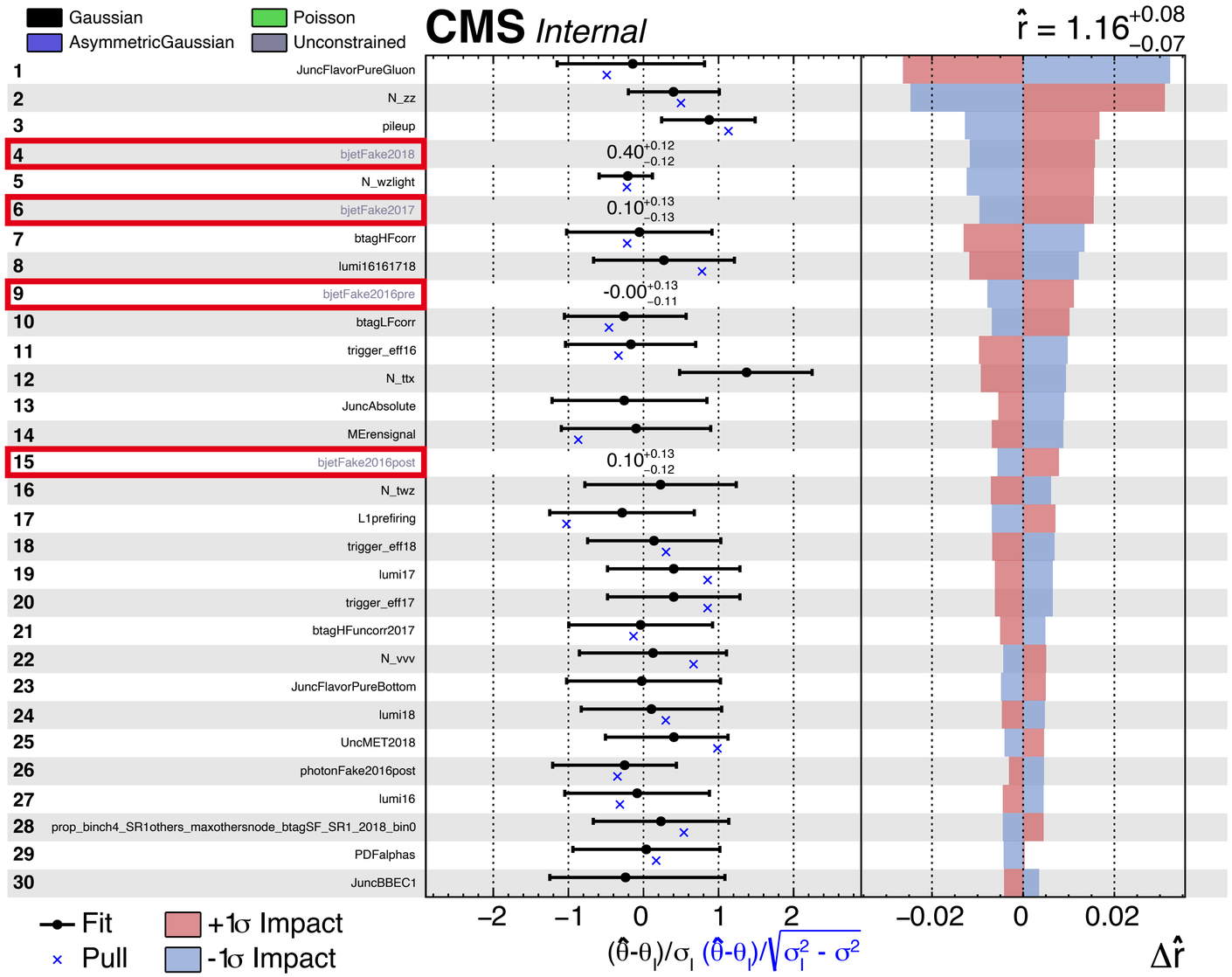




Run 2 partial unblinded fit

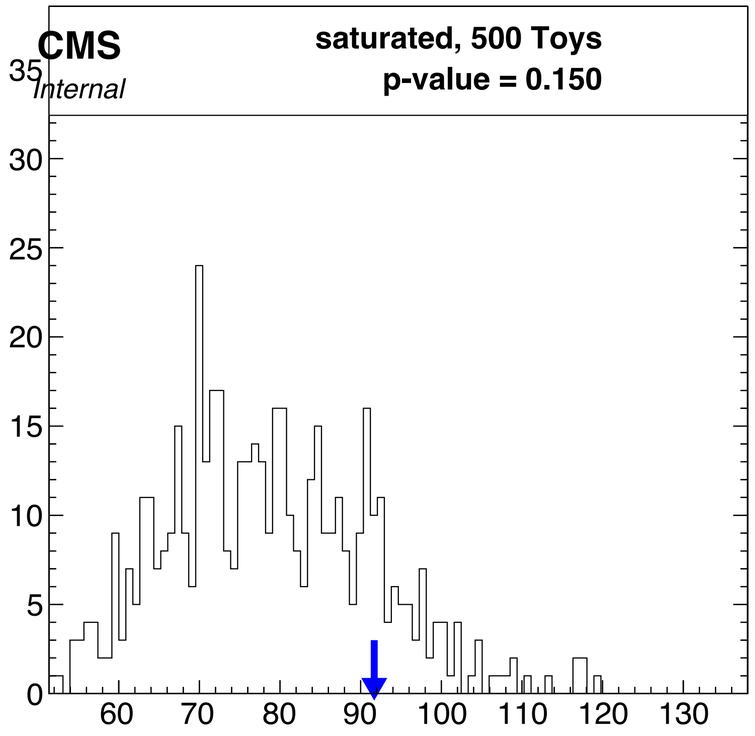
- ttZ used as POI.
- tWZ ML algorithm output blined.
- Simultaneous fakes normalisation measurement possible.
- Good p-value for the GOF test.

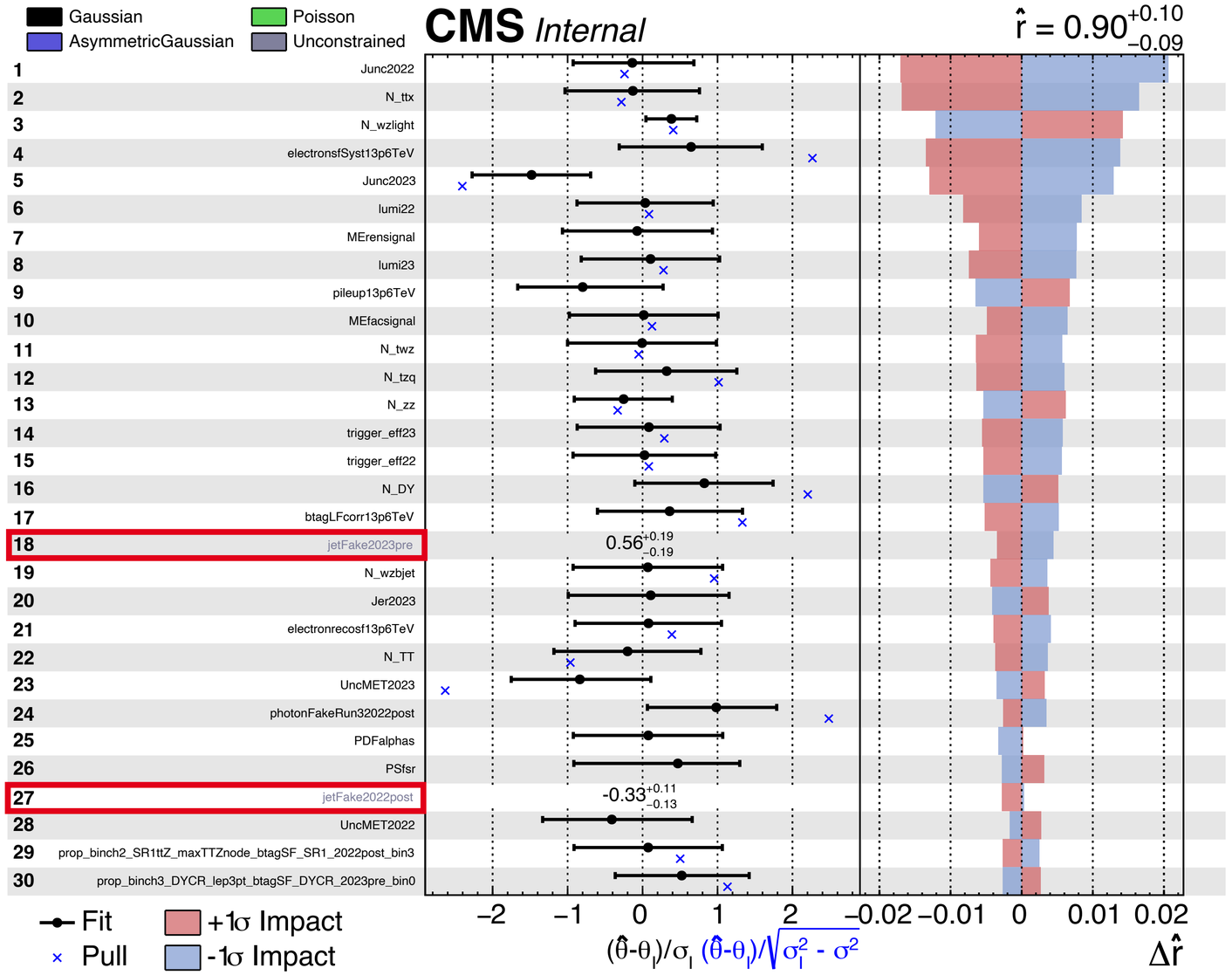




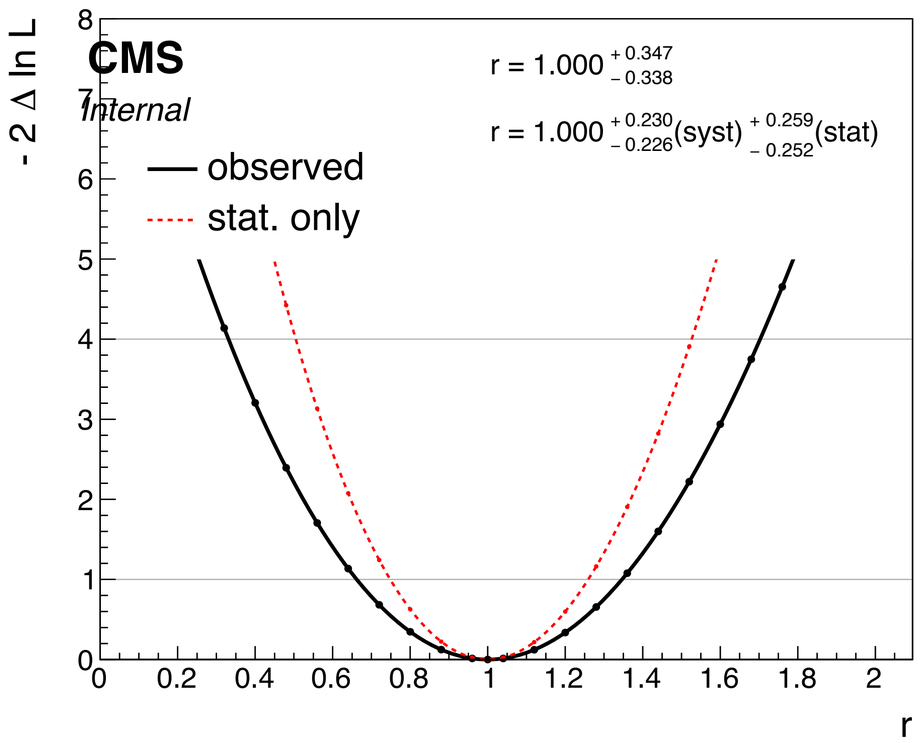
Run 3 partial unblinded fit

- ttZ used as POI.
- tWZ ML algorithm output blined.
- Simultaneous fakes normalisation measurement possible.
- Good p-value for the GOF test.

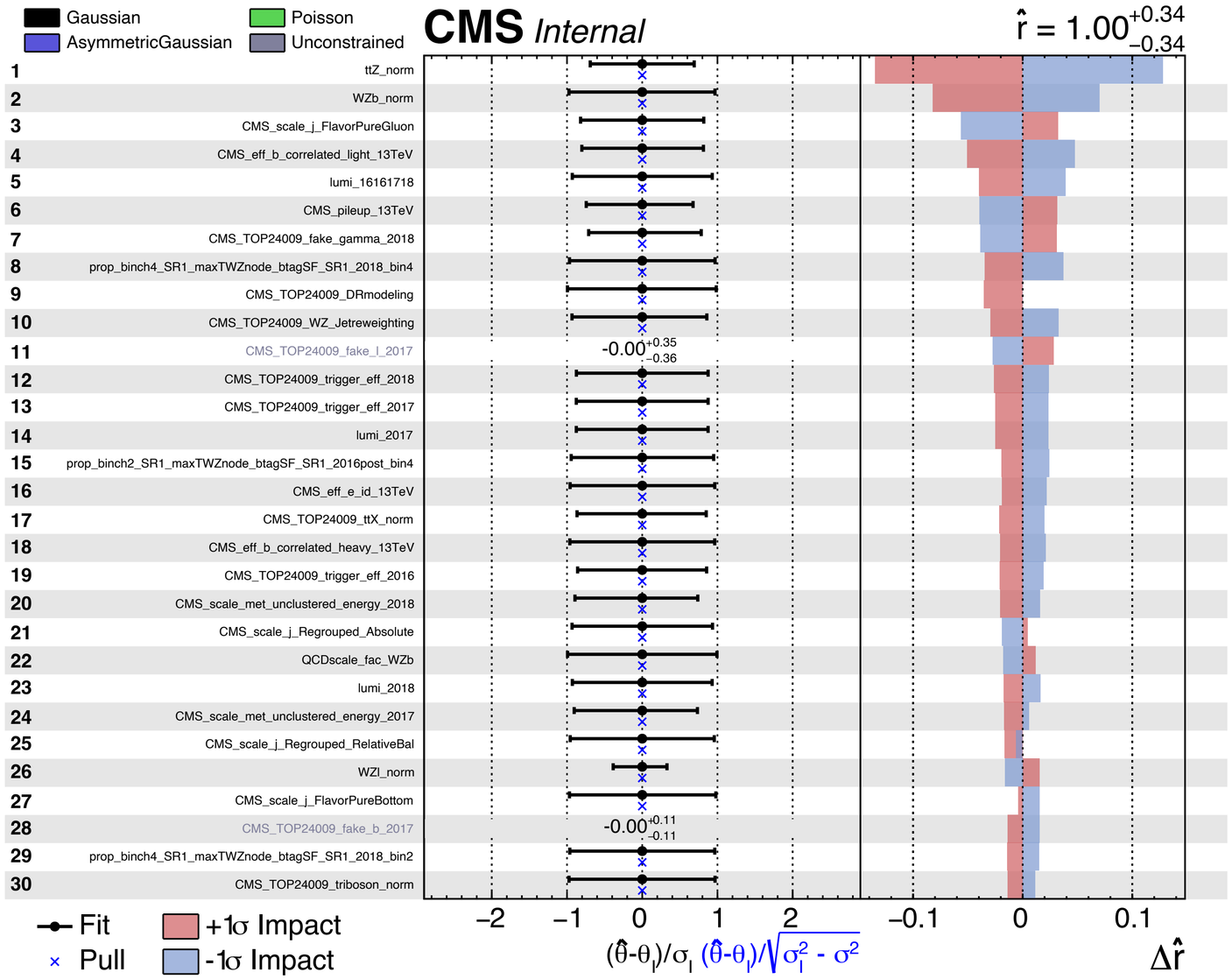




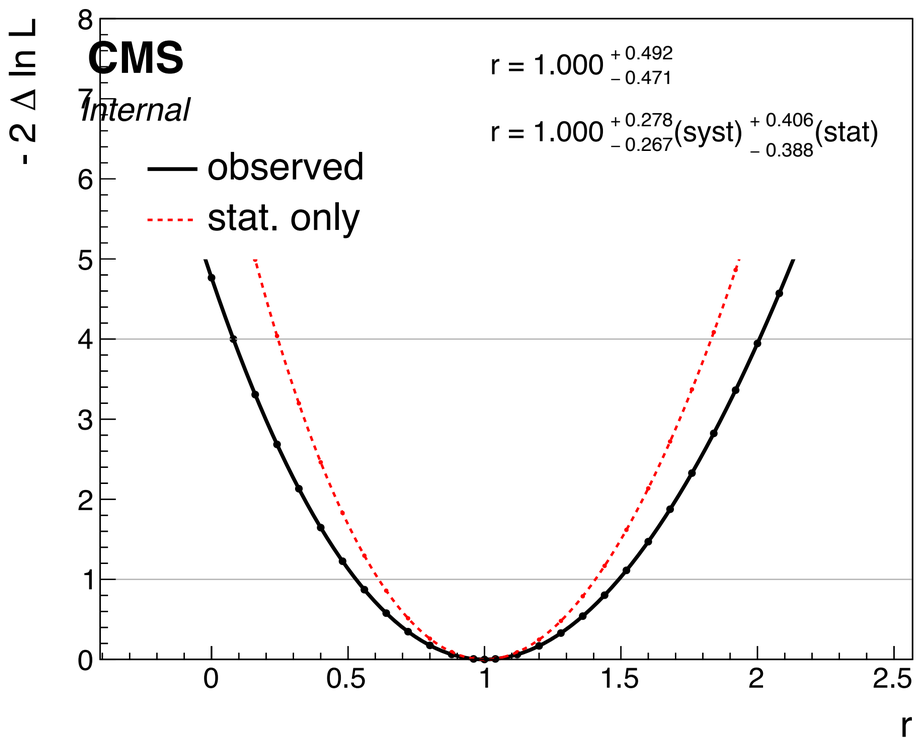
Run 2 fit on Asimov data

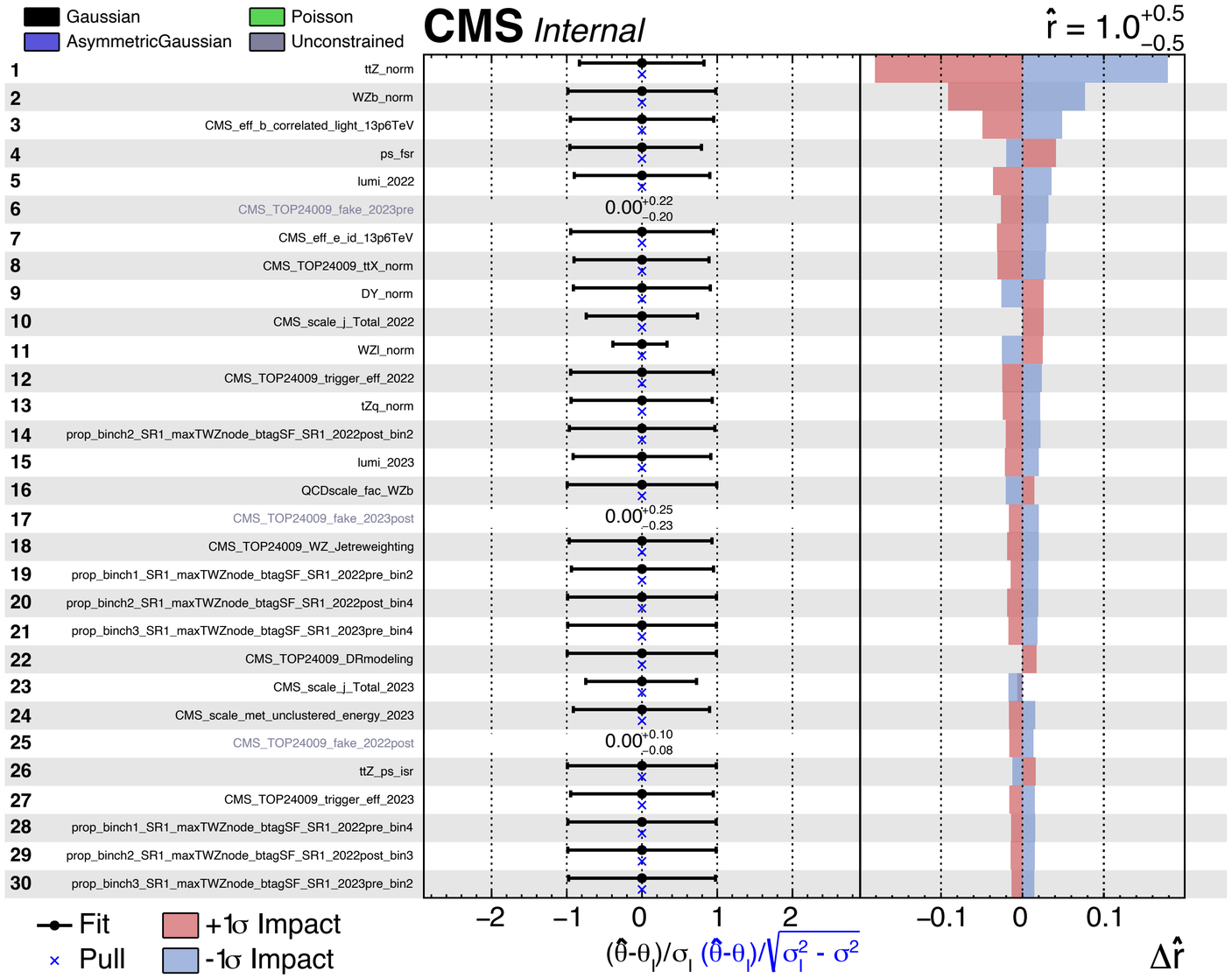


- Fit on Asimov data with full Run2 luminosity.
- Expected σ: 3.02 (1.36 for TOP-22-008)



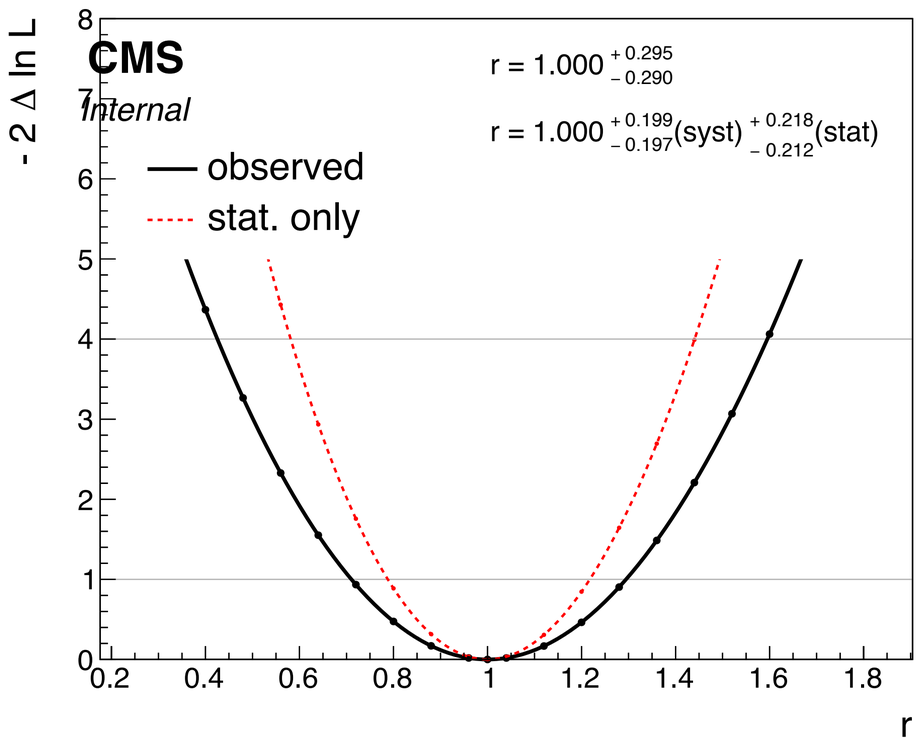
Run 3 fit on Asimov data



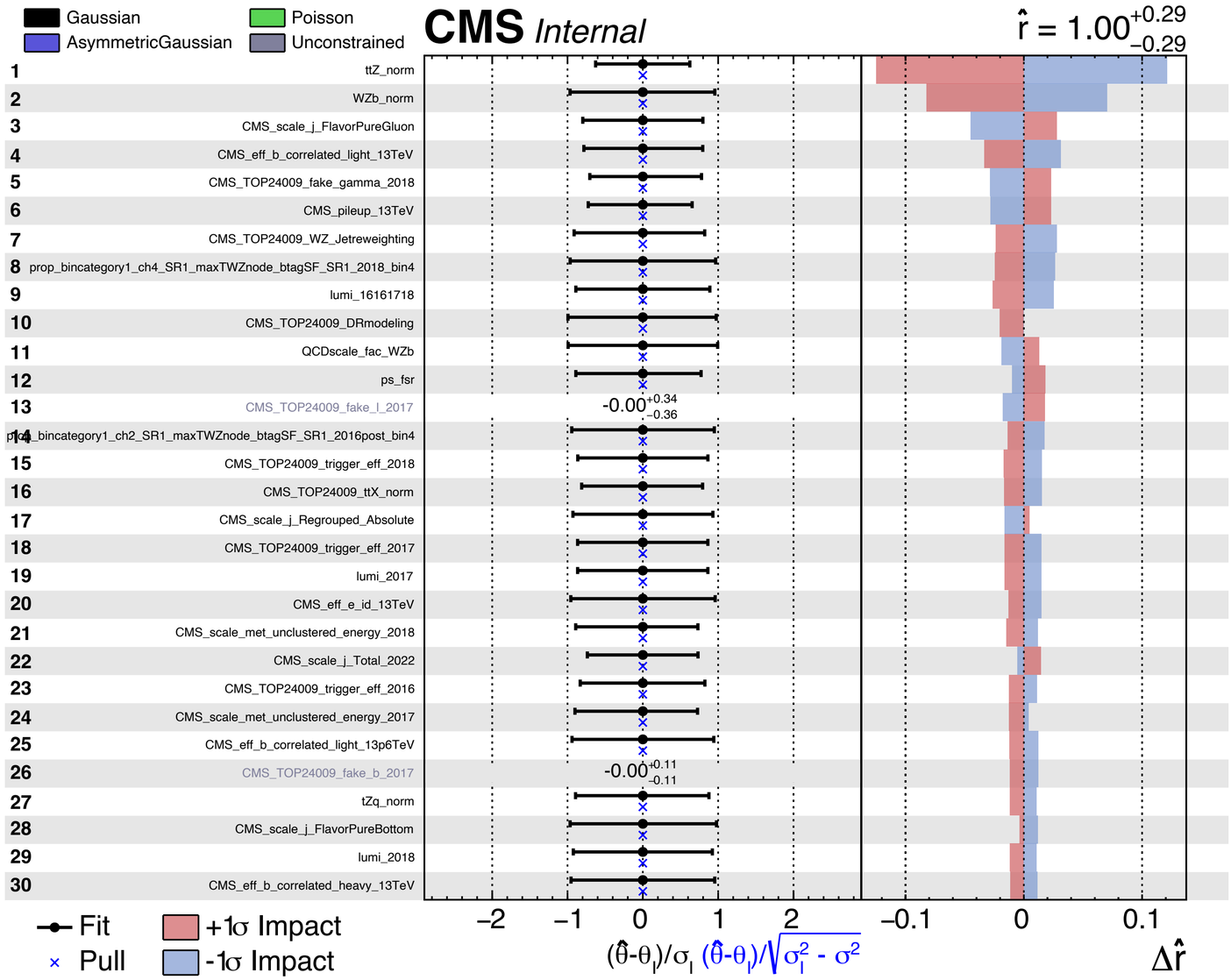


- Fit on Asimov data with full Run3 luminosity.
- Expected σ: 2.18

Simultaneous fit on Run 2 and Run 3



- Simultaneous fit on Run 2 and Run 3.
- Expected σ: 3.52



BACKUP

New TWZ cross-section and TWZ vs TWll comparison

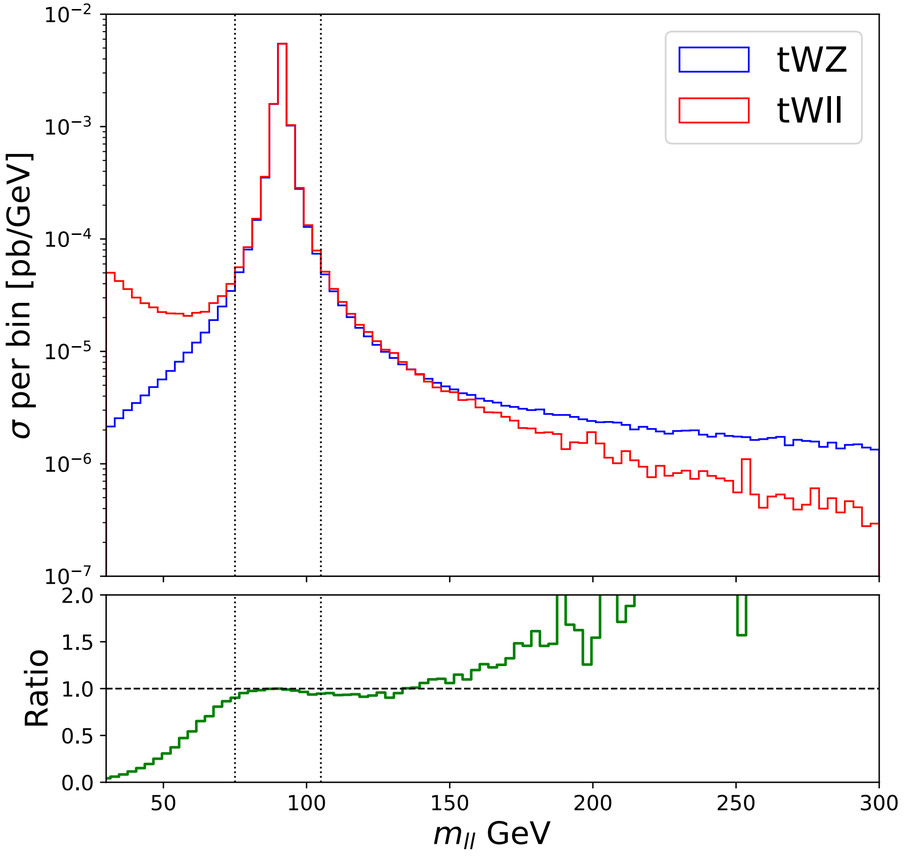
While checking possible differences between TWZ and TWll discovered a **bug** in the application of the branching ratios for the **TWZ semi-leptonic cross-section.**

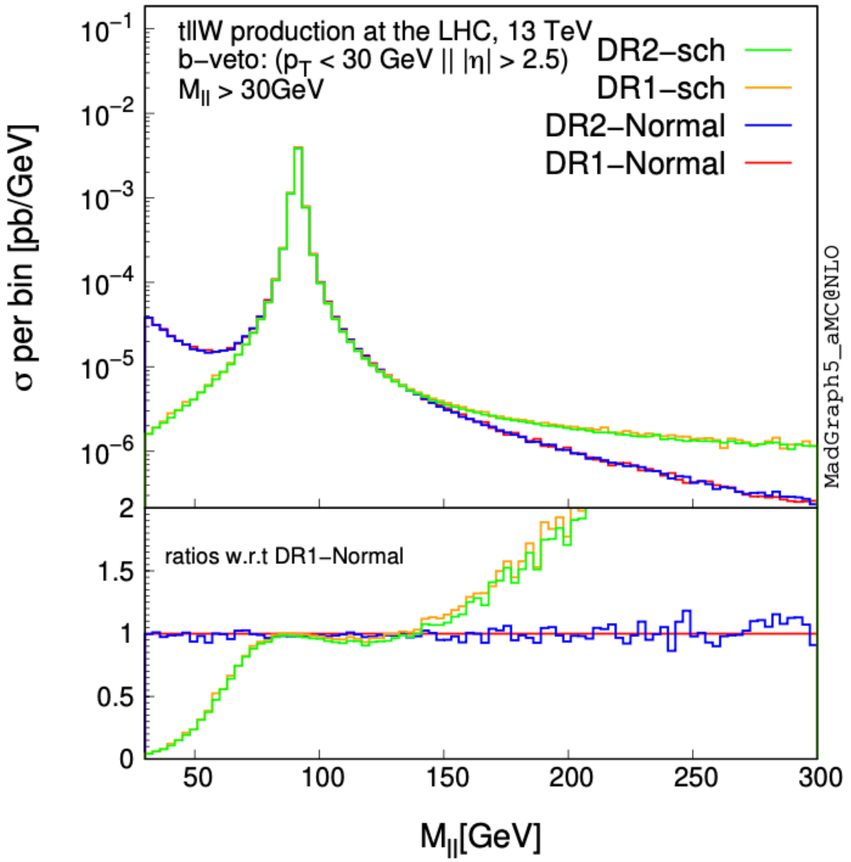
Change in cross-section: 4.1 fb → 3.0 fb

Contacted the author of arXiv:2111.03080 to get an histogram of tWll as function of m_{ll}.

Verified what we obtained with the results of their paper.

The **SF** in the the Z window **is very small.**
SF = **1.0058**





ML model

ParT paper

Github repository

Different ML models have been tested:

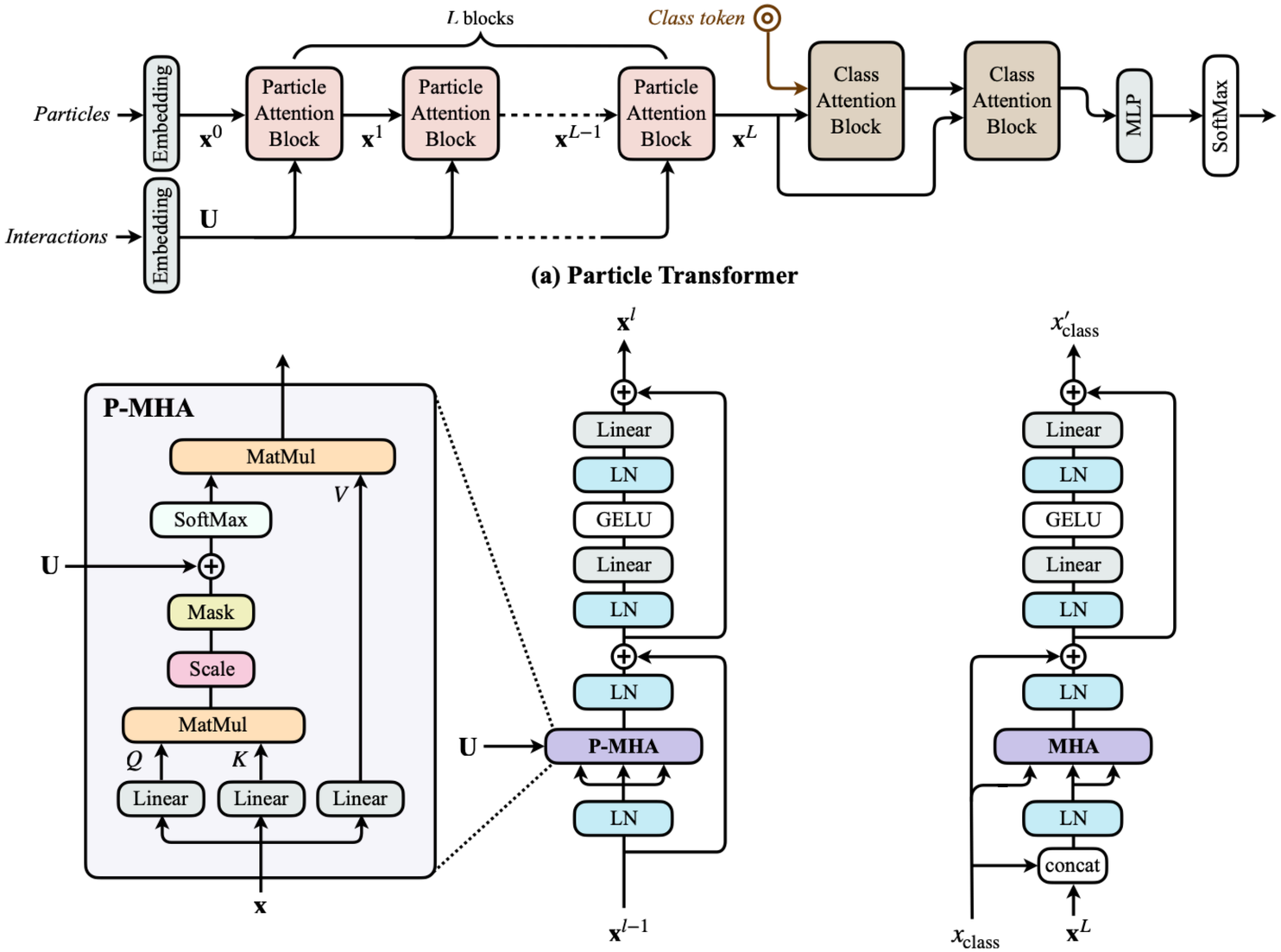
- Boosted decision trees.
- Fully connected feed forward NN.
- **Particle Transformer**.

Particles **variables** and **features**:

- 4-momentum of the first **6 jets**.
- 4-momentum of the **4/5 leptons**.
- log(p_T), log(E), η, b-tag, and pdgID.

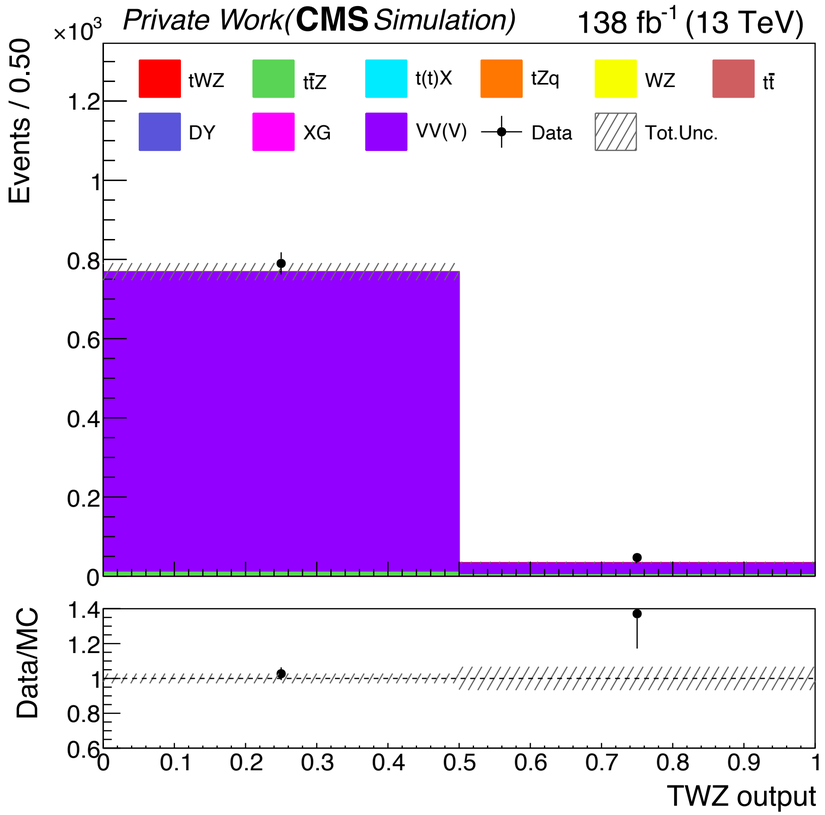
Interactions:

- $\Delta = \sqrt{(y_a - y_b)^2 + (\phi_a - \phi_b)^2}$
- $k_T = min(p_{T,a}, p_{T,b})\Delta$
- $z = min(p_{T,a}, p_{T,b})/(p_{T,a} + p_{T,b})$
- $m^2 = (E_a + E_b)^2 - ||\mathbf{p}_a + \mathbf{p}_b||^2$

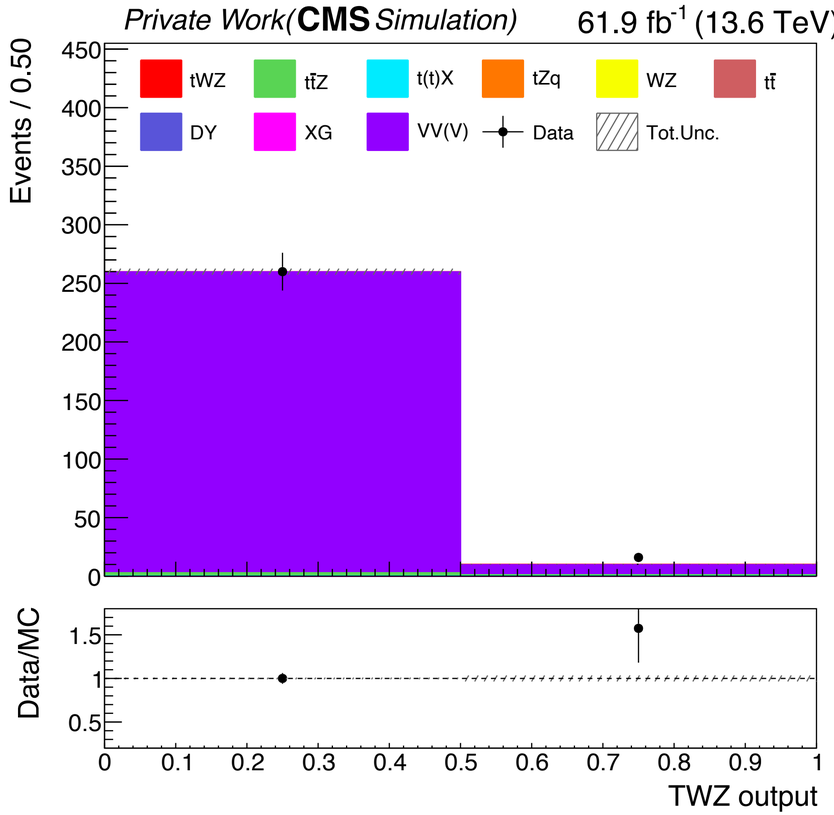


Fully leptonic SR algorithm in the ZZ CR

Run 2

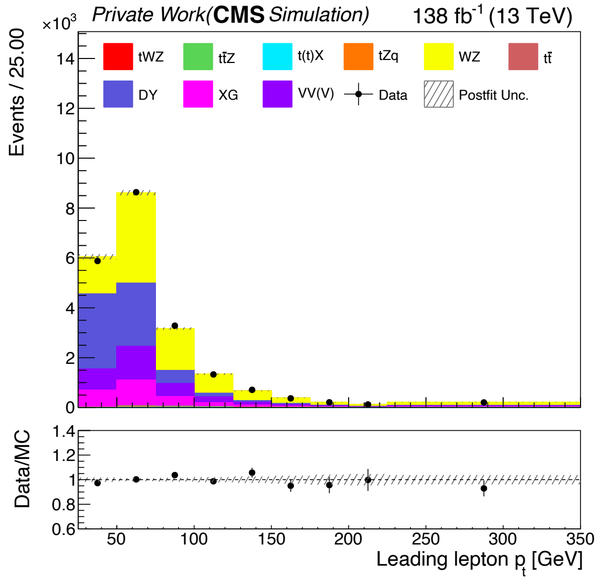


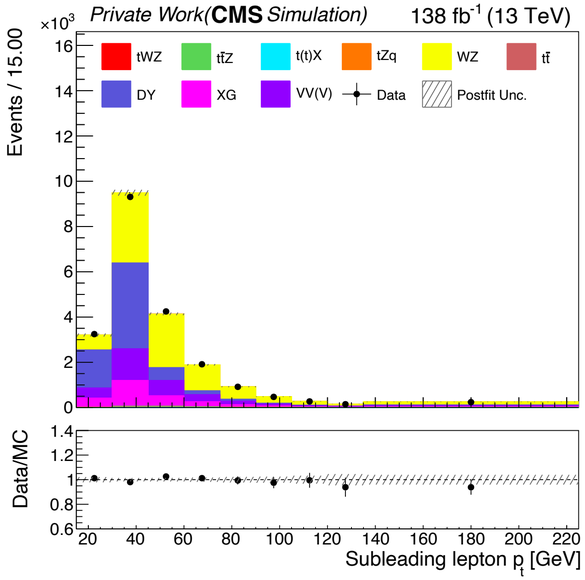
Run 3

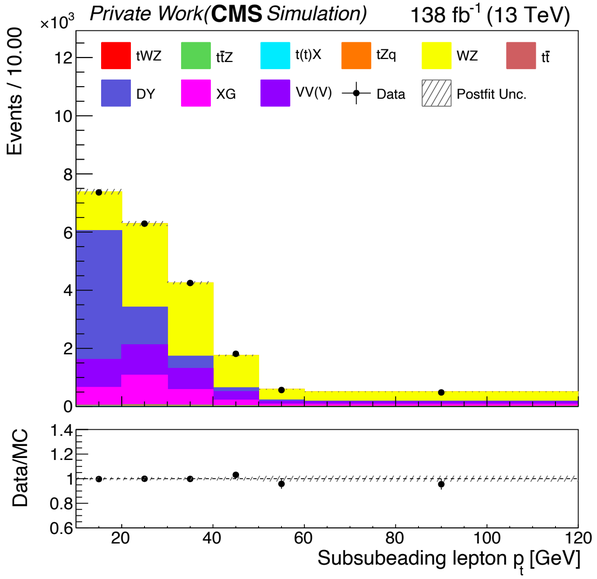


Pre vs Post fit distributions (DY CR)

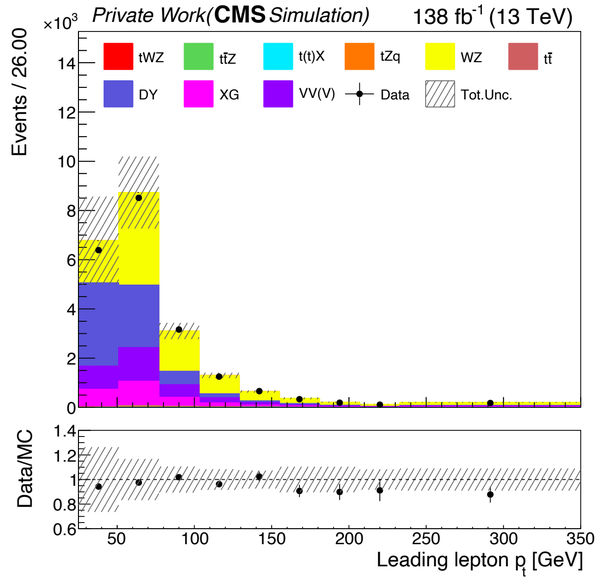
Post-fit

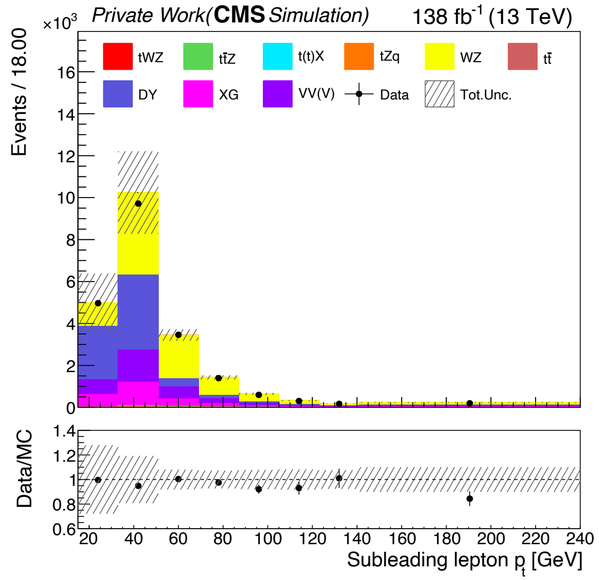


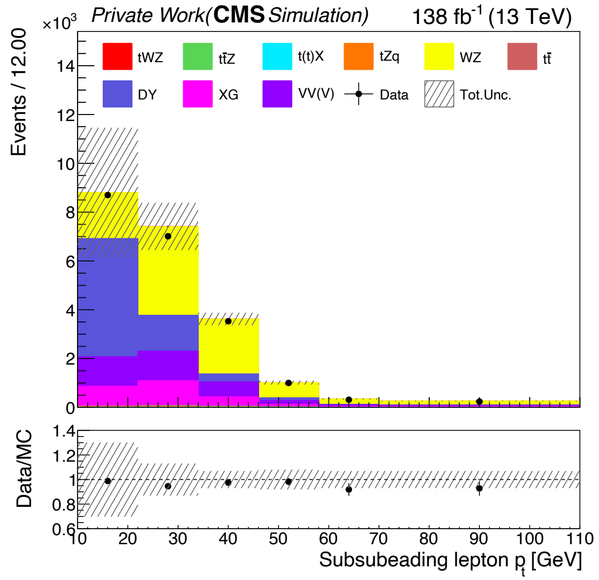




Pre-fit

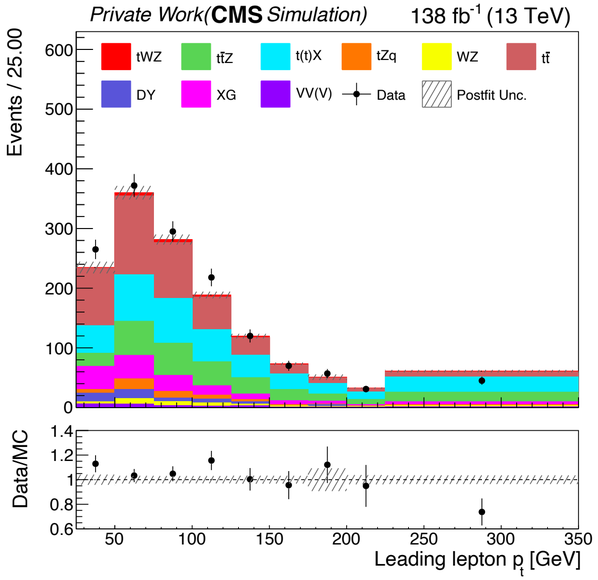


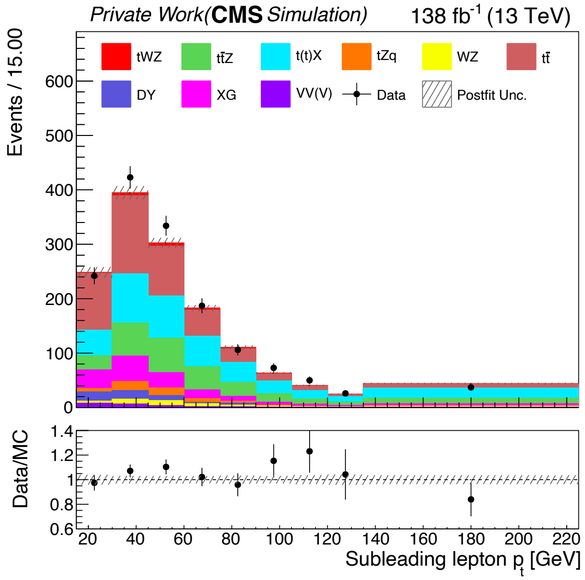


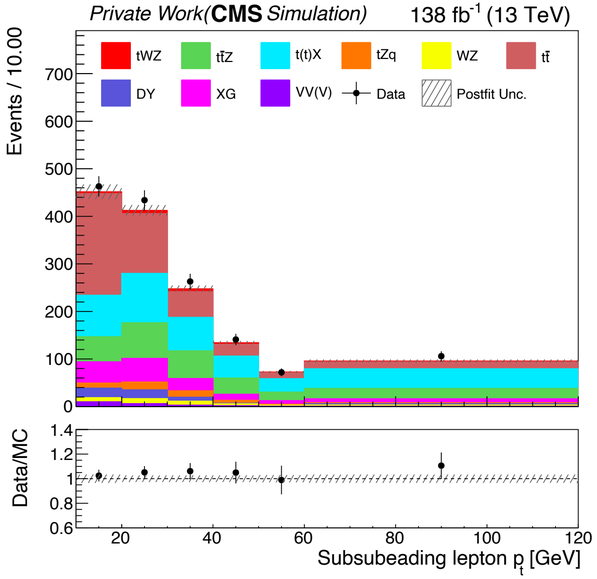


Pre vs Post fit distributions (ttX CR)

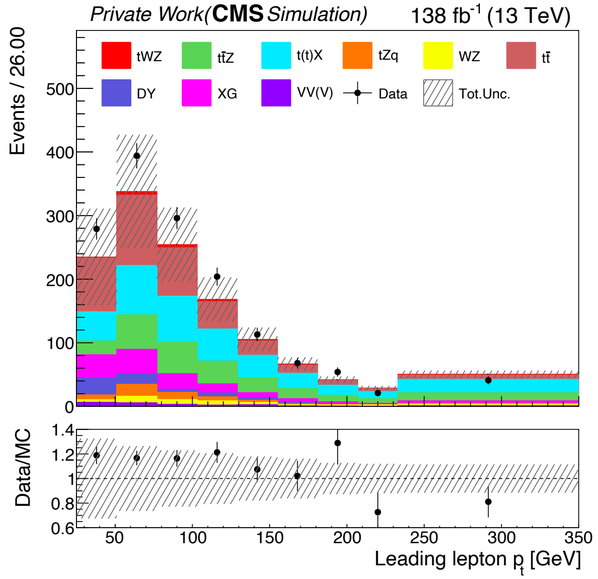
Post-fit

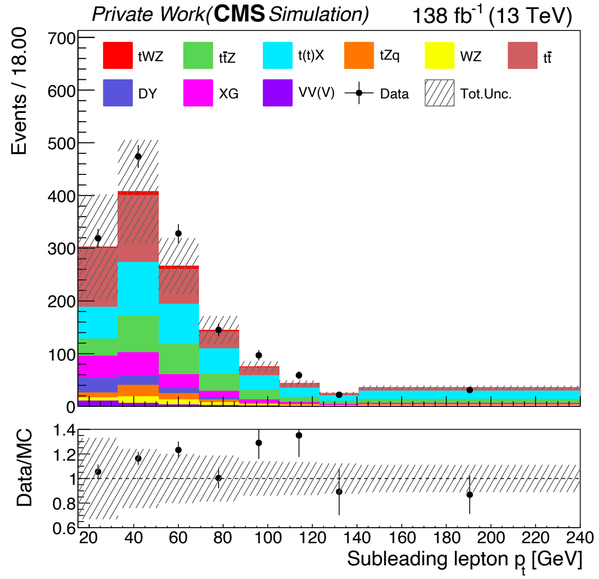


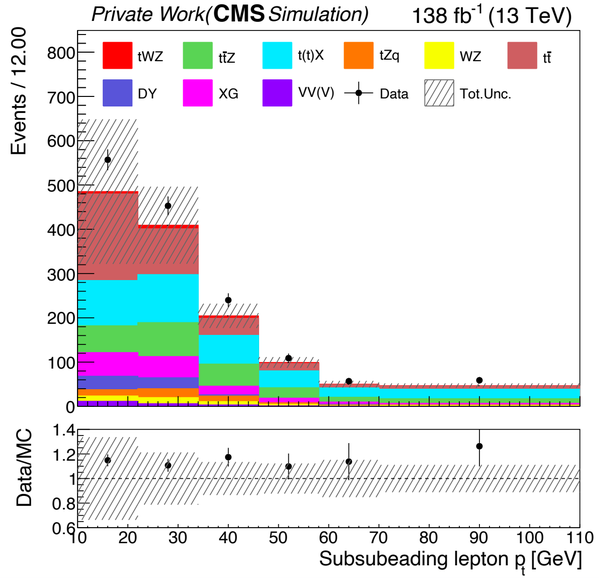




Pre-fit







Systematic uncertainties (differences with Run2)

- No split in stat and sys for lepton ID SFs
- B-tag ID SFs, (HF, LF)⊗(correlated) (no uncorrelated part)
- JEC (one single nuisance, reduced splitting not available yet)
- Merged light and bjet fake categories

ttZ normalisation uncertainty

Theory (2020) from arXiv:2001.03031

Process	μ_0	NLO[fb]	NLO+NLL[fb]	NLO+NLLwC[fb]	NLO + NNLL[fb]	K_{NNLL}
$t\bar{t}Z$	Q	$661^{+13.8\%}_{-12.5\%}$	$698^{+11.5\%}_{-10.1\%}$	$795^{+10.6\%}_{-9.7\%}$	$847^{+8.1\%}_{-8.2\%}$	1.28
	H_T	$694^{+13.6\%}_{-12.6\%}$	$723^{+11.0\%}_{-9.8\%}$	$805^{+10.0\%}_{-9.5\%}$	$848^{+7.9\%}_{-8.0\%}$	1.22
	$Q/2$	$752^{+12.5\%}_{-12.1\%}$	$770^{+10.6\%}_{-9.4\%}$	$824^{+8.8\%}_{-8.8\%}$	$854^{+7.1\%}_{-7.8\%}$	1.14
	$H_T/2$	$788^{+11.7\%}_{-11.9\%}$	$798^{+10.7\%}_{-9.5\%}$	$834^{+8.1\%}_{-8.4\%}$	$855^{+6.6\%}_{-7.7\%}$	1.09
	$M/2$	$841^{+9.4\%}_{-11.1\%}$	$848^{+11.2\%}_{-9.7\%}$	$858^{+7.1\%}_{-7.9\%}$	$874^{+6.7\%}_{-7.8\%}$	1.04

ttZ discovery paper from CMS (2019) (arXiv:1812.05900): 15%

tWZ+TTZ and tZq simultaneous measurement from CMS (2024) (arXiv:2410.23475): 6.1%

Latest TTZ measurement from ATLAS (2024) (arXiv:2312.04450): 6.6%

The normalisation uncertainty of ttZ has been **updated to 7%** in the analysis.