Search for the SM Higgs Boson in the $H \rightarrow \tau^{-}\tau^{+} \rightarrow \mu^{-}\mu^{+}\bar{\nu}_{\mu}\nu_{\mu}\nu_{\tau}\bar{\nu}_{\tau}$ Channel

Dissertation Endorsement

Vallary Bhopatkar

22 Nov 2017

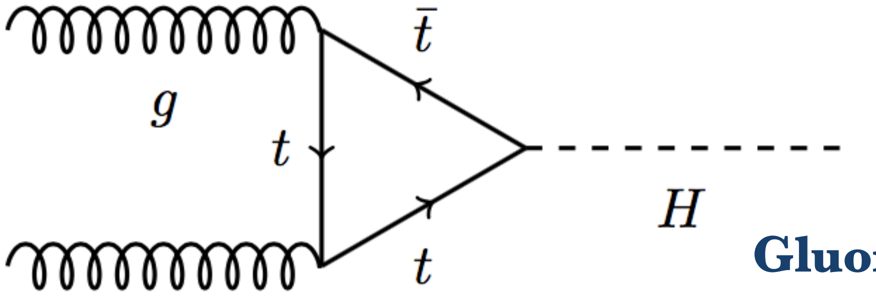
Outline

The standard model (SM) $\boldsymbol{H \rightarrow \tau\tau \rightarrow \mu\mu}$ analysis using full 2016 Run II dataset with $\sqrt{s} = 13\text{TeV}$, $\mathcal{L} = 35.9\ \text{fb}^{-1}$.

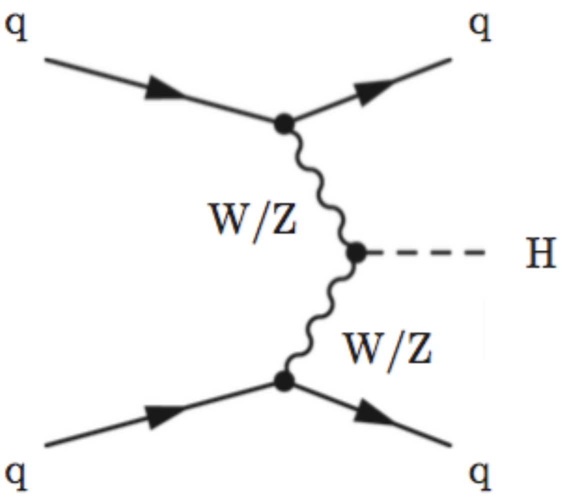
- SM Higgs: Overview
- Event Selection
- Event Categorization
- MC Corrections and Background Estimation Techniques
- Boosted Decision Trees (BDT) Multivariate Analysis for Background Suppression
- Statistical Analysis and Final Results

Higgs: Production and Decay Modes

Production Modes

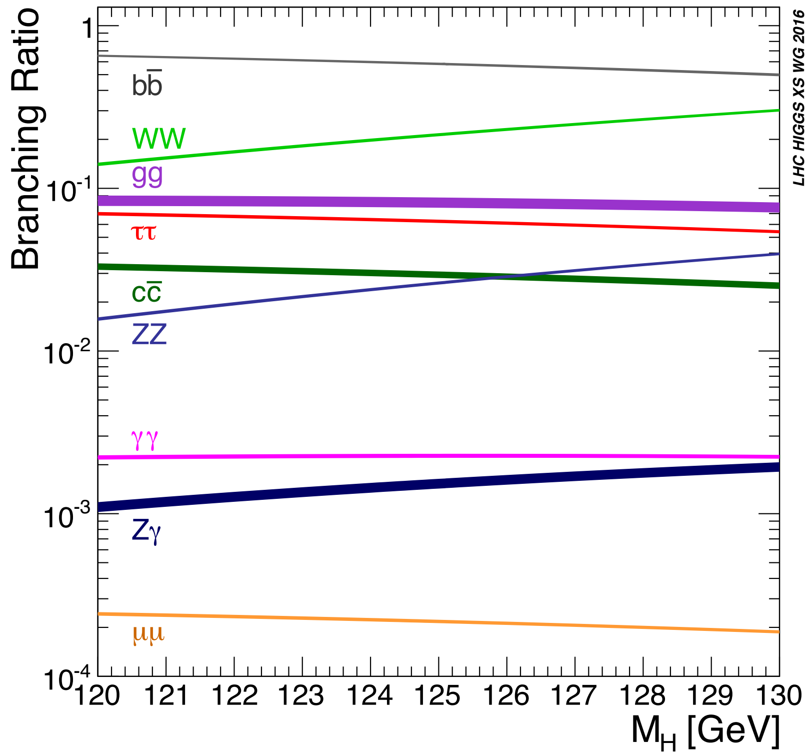


Gluon-Gluon Fusion (GGF, ggH)



Vector Boson Fusion (VBF, qqH)

Decay Modes



Will be focusing on the **ττ** decay mode

Overview: $\boldsymbol{H \rightarrow \tau\tau}$ Decay Mode

- τ leptons decay hadronically and leptonically
- Promising channel to study Yukawa couplings to leptons
  - Highest event yield among the leptonic channels
  - Lower background compared to $b\bar{b}$ decays
- This dissertation focuses on the leptonic channel where a pair of tau leptons decays into a pair of muons

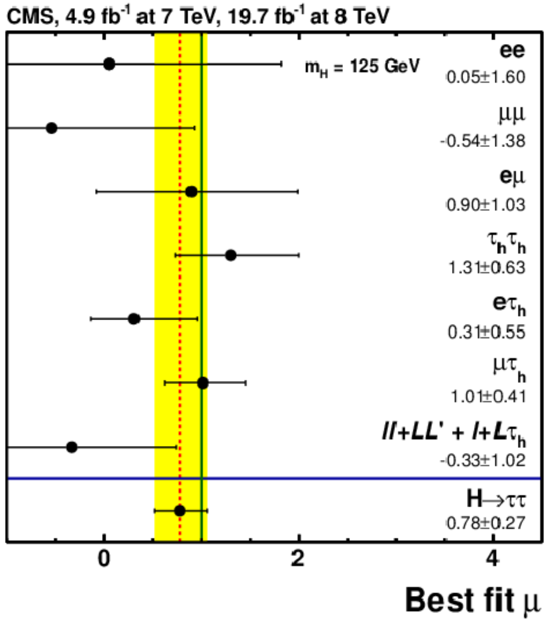
- Tau decay modes with BR

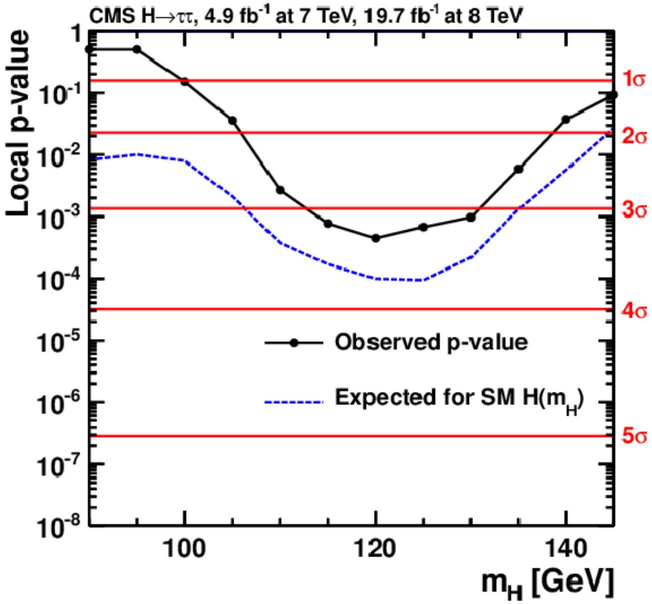
Decay Mode	BR in [%]
$\tau_{had}\tau_{had}$	42
$\tau_{\mu}\tau_{had}$	23
$\tau_{e}\tau_{had}$	23
$\tau_{e}\tau_{\mu}$	6
$\tau_{e}\tau_{e}$	3
$\boldsymbol{\tau_{\mu}\tau_{\mu}}$	**3**

- **Challenges**: Small Branching ratio and $Z \rightarrow \mu\mu$ background

Run I and Run II Overview

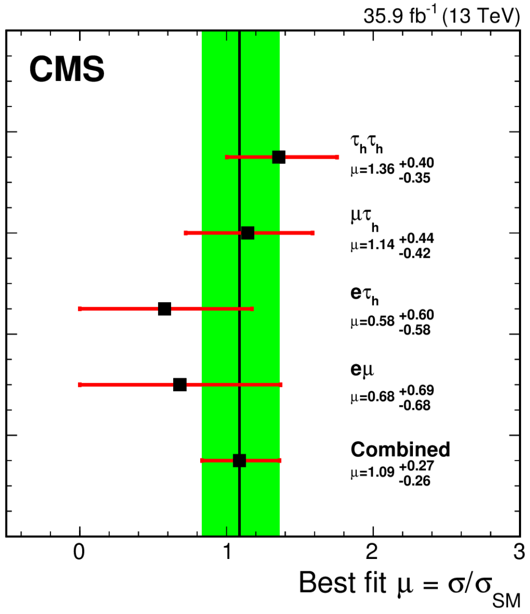
Run I (Results with all six decay modes)

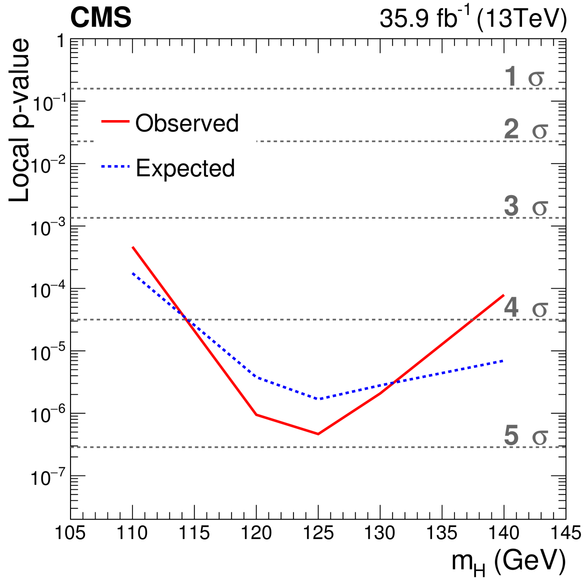




- CMS obtained evidence at 3.2σ with signal strength $\hat{\mu} = 0.78 \pm 0.27$ [JHEP 05 (2014) 104]
- CMS + ATLAS combined results shows evidence at 5.5σ [JHEP 08 (2016) 045]

Run II (Results with $\tau_h\tau_h$, $\tau_\mu\tau_h$, $\tau_e\tau_h$, $\tau_e\tau_\mu$)





- CMS observed evidence at 4.7σ, with signal strength relative to SM $\hat{\mu} = 1.09^{+0.27}_{-0.26}$ Submitted to Phys. Lett. B (CMS-HIG-16-043 ; CERN-EP-2017-181)

SM $\boldsymbol{H} \rightarrow \boldsymbol{\tau\tau} \rightarrow \boldsymbol{\mu\mu}$ Analysis with Run II 2016 Data

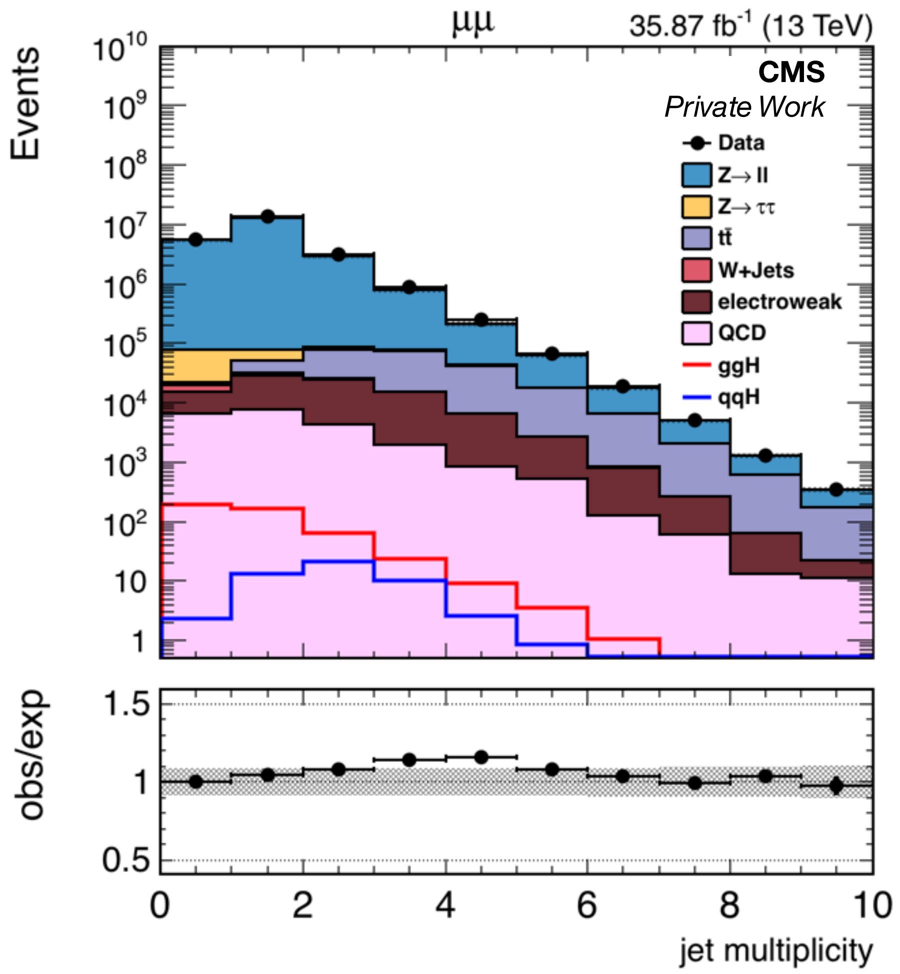
Event Selection

- Run II full 2016 dataset RunB-H
- $\sqrt{s} = 13$ TeV, $\int \mathcal{L}\, dt = 35.9$ fb^{-1}
- Bunch crossing 25ns
- Single Muon Data used with HLT single muon trigger *HLT_IsoMu24_V*
- $d_{xy} < 0.045$cm
- $d_Z < 0.2$cm
- $\triangle R(\mu^+,\mu^-) \geq 0.5$
- $\triangle\beta$ relIso < 0.15

Channel Specific Selection

- Two muons with opposite charges
- $p_T > 10$ GeV and $|\eta| < 2.4$
- Leading muon $p_T > 25$ GeV and match HLT muon objects with $\triangle R < 0.5$
- Must satisfy medium Particle Flow (PF) muon identification criteria for simulated events and for data from Run G-H
- Special 2016 (ICHEP) medium PF criteria for data from Run B-F
- Visible mass $m_{vis} > 20$ GeV

Event Categorization



Three event categories based on the jet multiplicity are considered for signal extraction:

- **0-jet** Targets the gluon-gluon fusion production (ggH) (0 jet multiplicity)
- **Boosted** Targeting ggH events with the Higgs recoiling against an additional jet and qqH events that fail the di-jet cut (jet multiplicity is 1 or ≥2 with di-jet mass m_{jj}<300 GeV)
- **VBF** Targets the vector boson fusion (qqH) via di-jet cut (jet multiplicity is exactly 2 with di-jet mass m_{jj}>300 GeV)

Data/MC Corrections and Event Weights

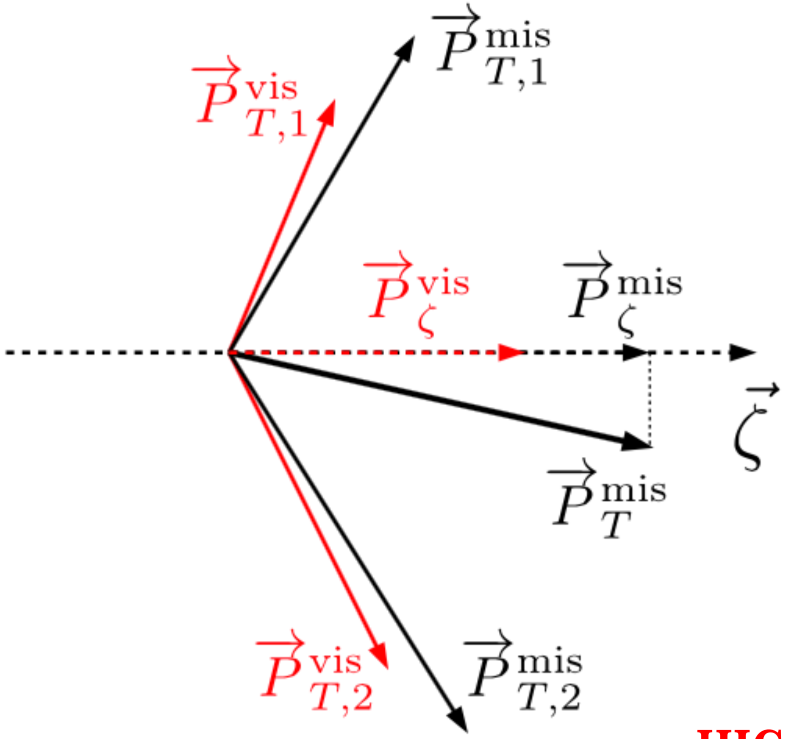
- **Generator event weights** applied to all MC samples on the event-by-event basis.
- **Pile-Up (PU)** reweighting applied to all MC samples (minimum bias xsec 69.2 mb).
- **Lepton ID/Iso scale factors** obtained using "Tag and Probe" method using $Z \rightarrow \mu\mu$ MC samples and applied to all simulated events.
- **Trigger Efficiency** applied to all MC events.
- **DY reweighting** evaluated separately for each jet category in bins of dimuon p_T, η, and mass and applied only to the DY events with mass > 50 GeV.
- **Recoil corrections** applied to DY, W+jets, and Higgs MC samples
- **Top p_T reweighting** applied to $t\bar{t}$ events.

Boosted Decision Trees (BDT)

- BDTs are trained and evaluated independently for each jet category using Higgs signal samples ggH (0-jets and boosted category), qqH (VBF category) and $Z \rightarrow \mu\mu$ and $Z \rightarrow \tau\tau$ background samples.
- The variables used in the BDT training are:
  - $\boldsymbol{\eta}_{\mathbf{dimuon}}$ The pseudo-rapidity of the dimuon system.
  - $\mathbf{p_T(2\mu)/\ [p_T(\mu^+)+p_T(\mu^-)]}$ The ratio of the transverse momentum of the dimuon system to the scalar sum of the transverse momenta of the positive and the negative muon.
  - $\boldsymbol{E_T^{miss}}$ The missing transverse energy of the system.
  - $\boldsymbol{\Delta\Phi(\mu^+, E_T^{miss})}$ The azimuthal angle between the positive muon transverse momentum and the missing transverse momentum.
  - $\mathbf{cos(\boldsymbol{\theta^*})}$ The polar angle of the μ^+ in the rest frame of the dimuon system

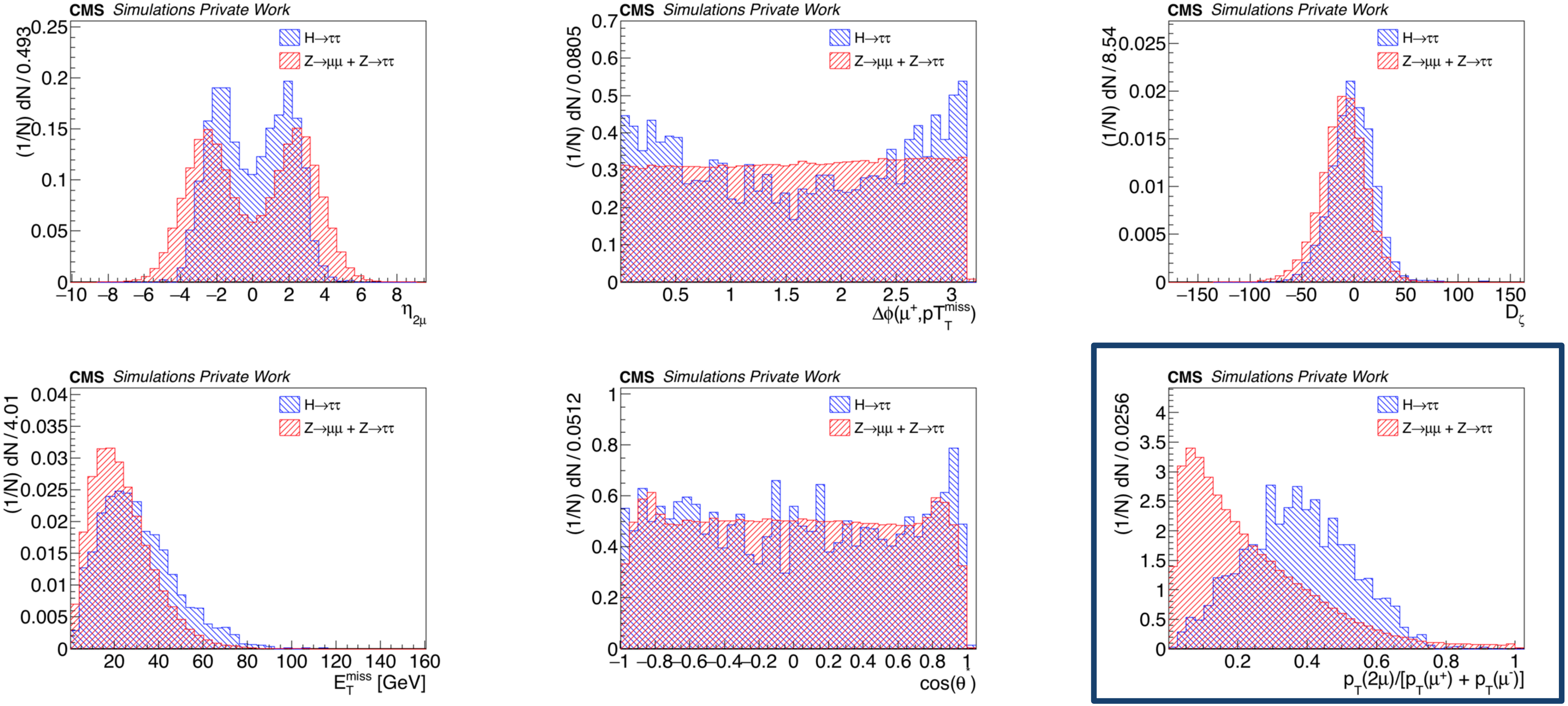
BDT continued ..

- **D_ζ** is defined as the difference of the projection of the visible transverse momentum of the τ decay products plus missing transverse momentum and visible transverse momentum of the τ decay products on the ζ axis which is the center line between visible momenta.
- **p_T^{tot}** (only for boosted category) Total transverse momentum of the system.
- **m_{jj}** (only for VBF category) Di-jet mass

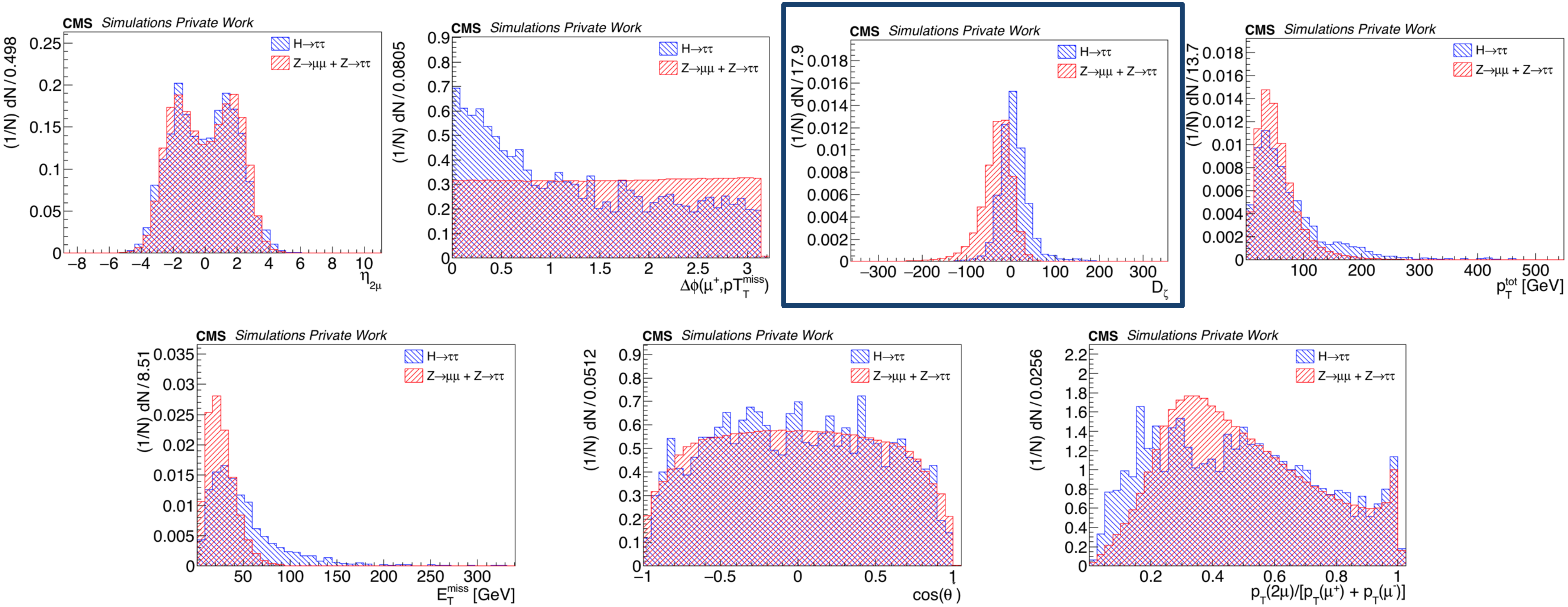


HIG-16-043

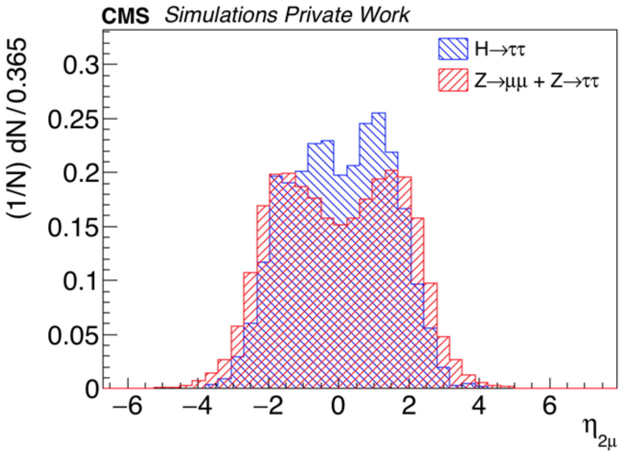
BDT Input Variables for 0-jet Category

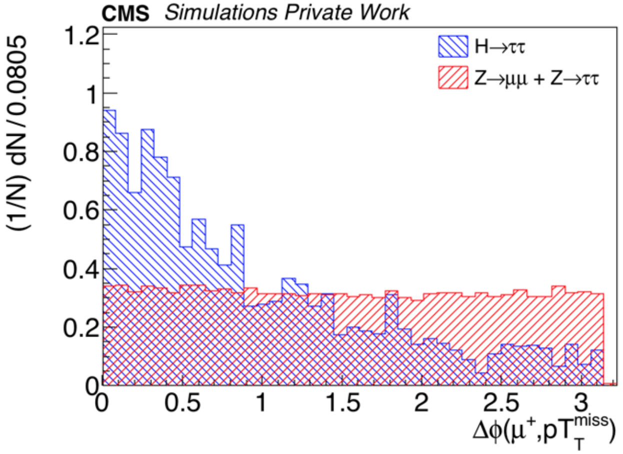


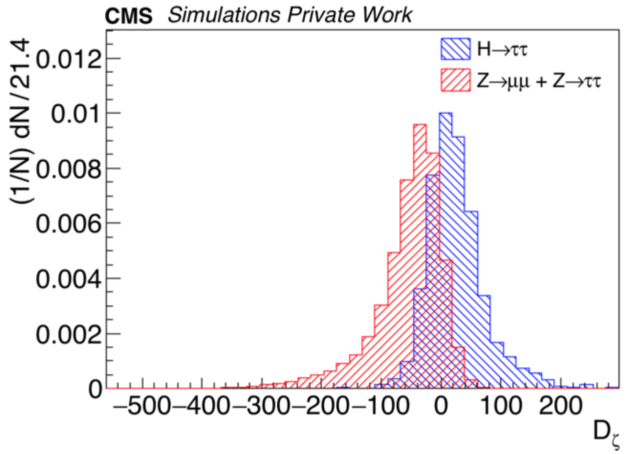
BDT Input Variables for Boosted Category

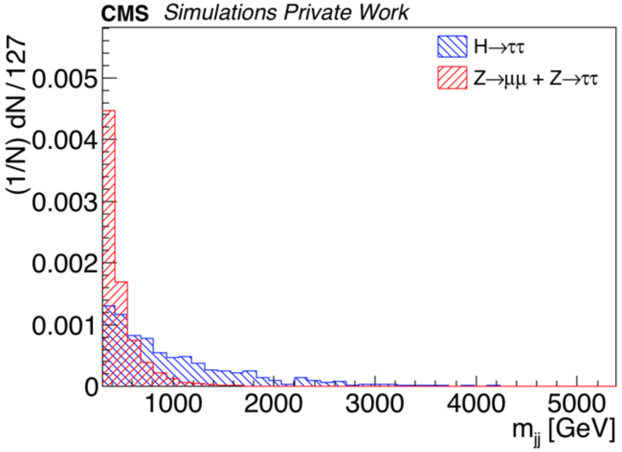


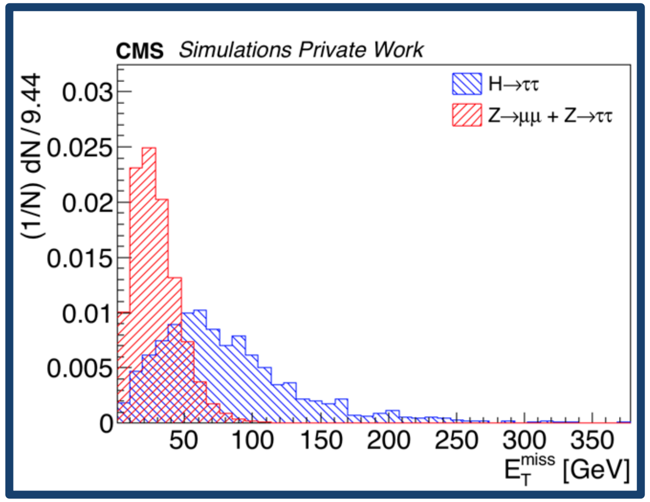
BDT Input Variables for VBF Category

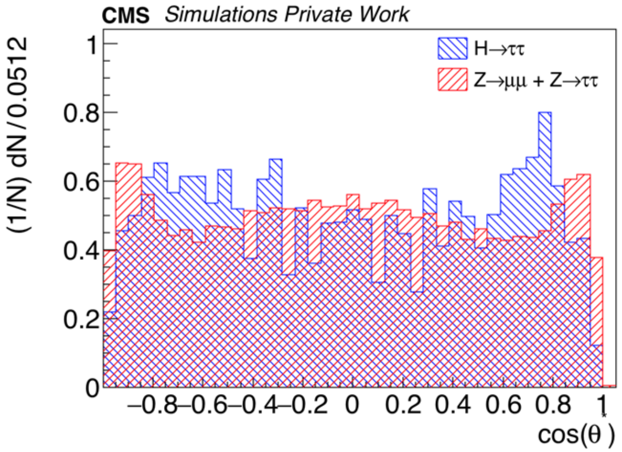


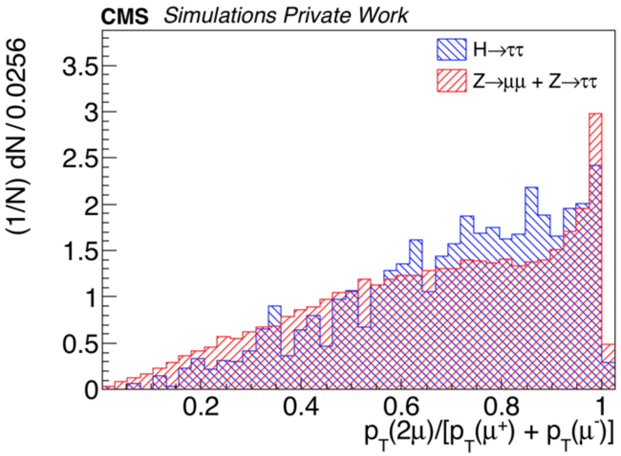






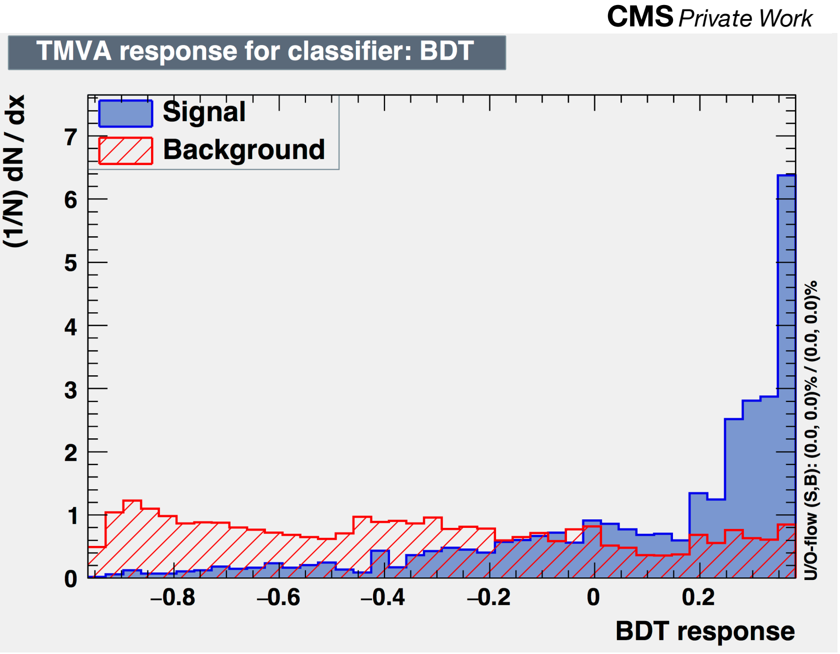




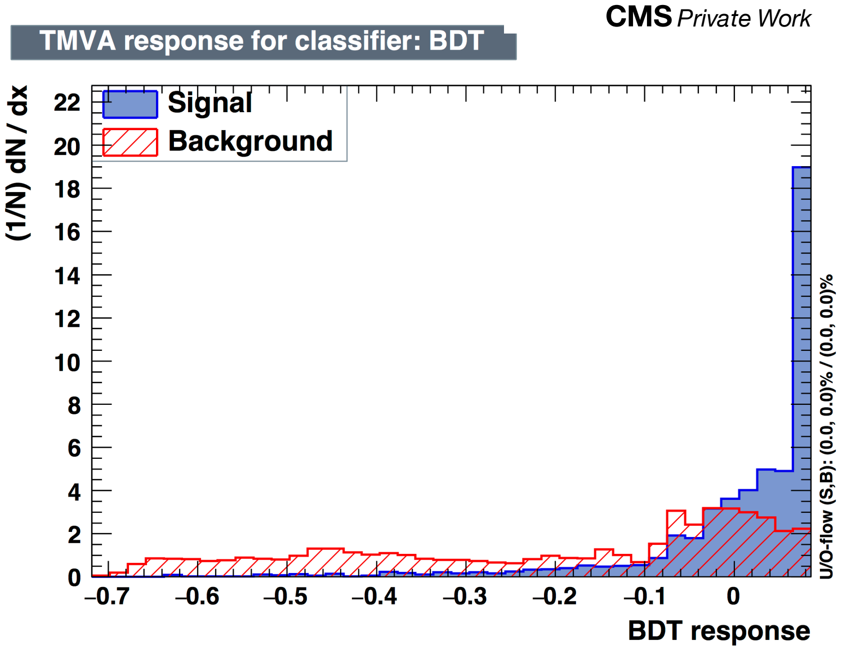


BDT Response for Each Category (TMVA)

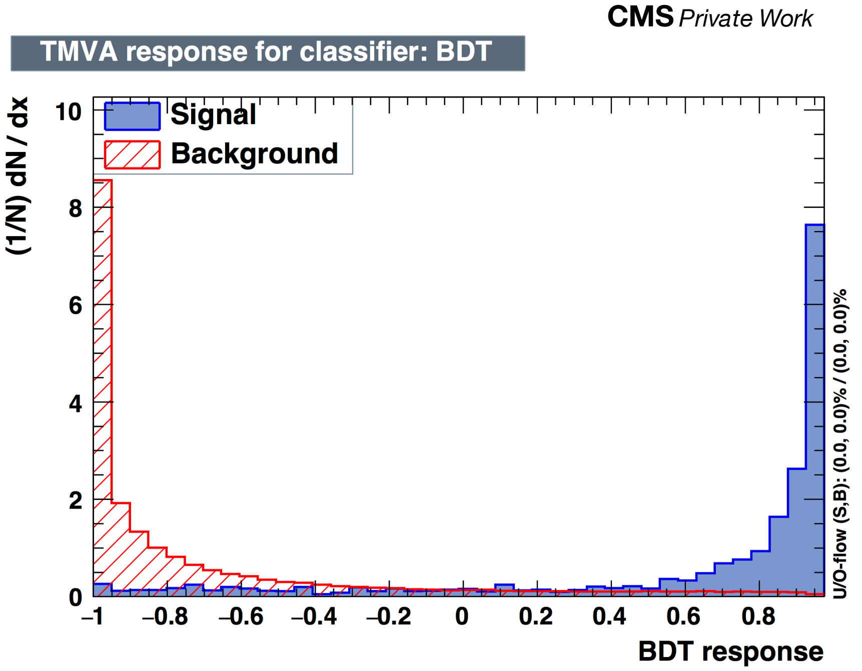
0-Jet



boosted



VBF

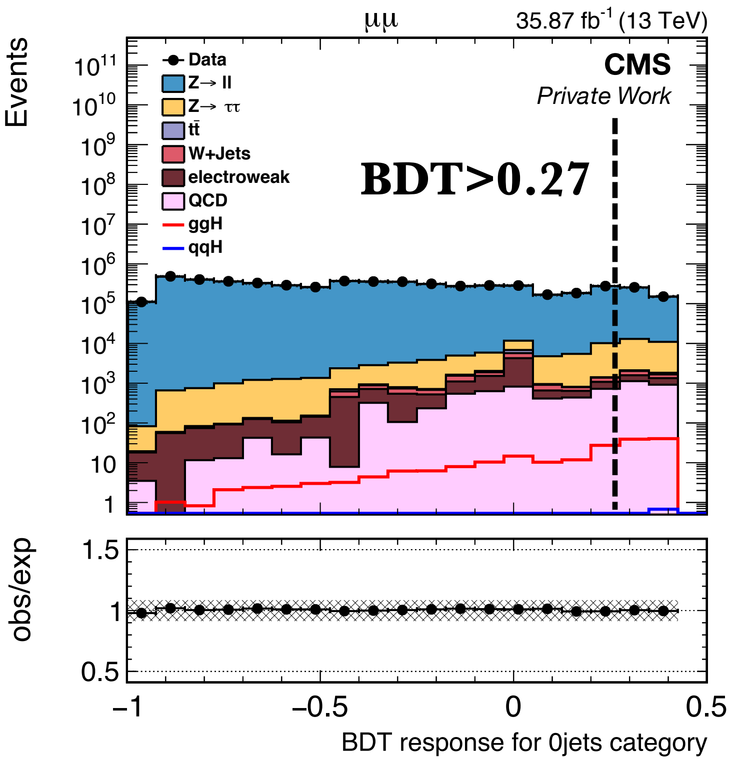


Background Estimation

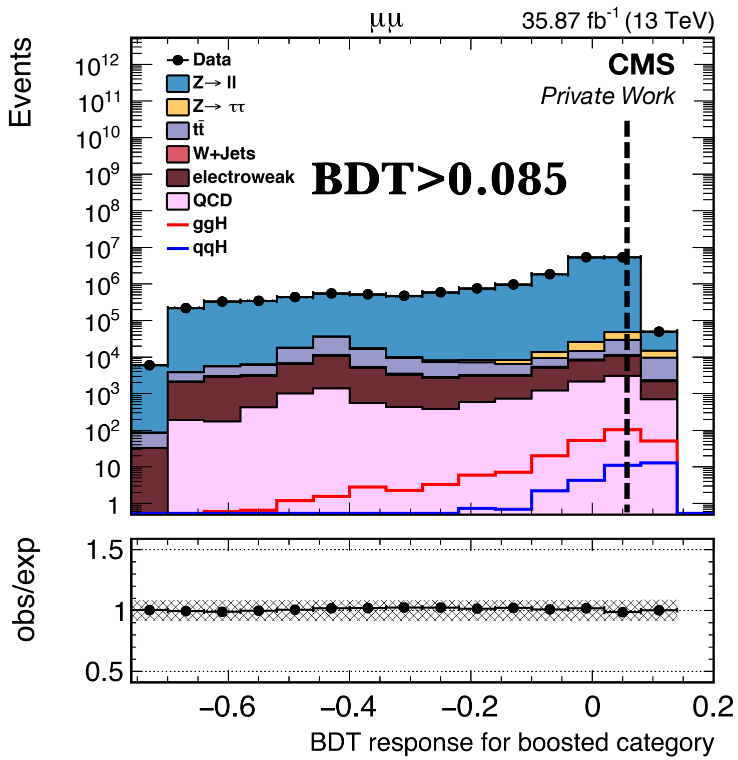
- The dominating background for this channel is **DY** and this background is estimated using the simulated MC events after applying all required corrections.
- The **QCD multijet** background is estimated from the same-sign (SS) data events, and then normalized by applying the extrapolation factor derived from the ratio of the opposite-sign events to same-sign events .
- **Diboson, Single top, and W+jets** together are referred to as the **"Electroweak"** background. It is estimated directly from MC after applying all the necessary corrections and reweighting.
- $t\bar{t}$ background is estimated from the MC events after applying the top p_T reweighting
- Events yield for respective backgrounds are derived from the prefit plots.

BDT Response for Each Category

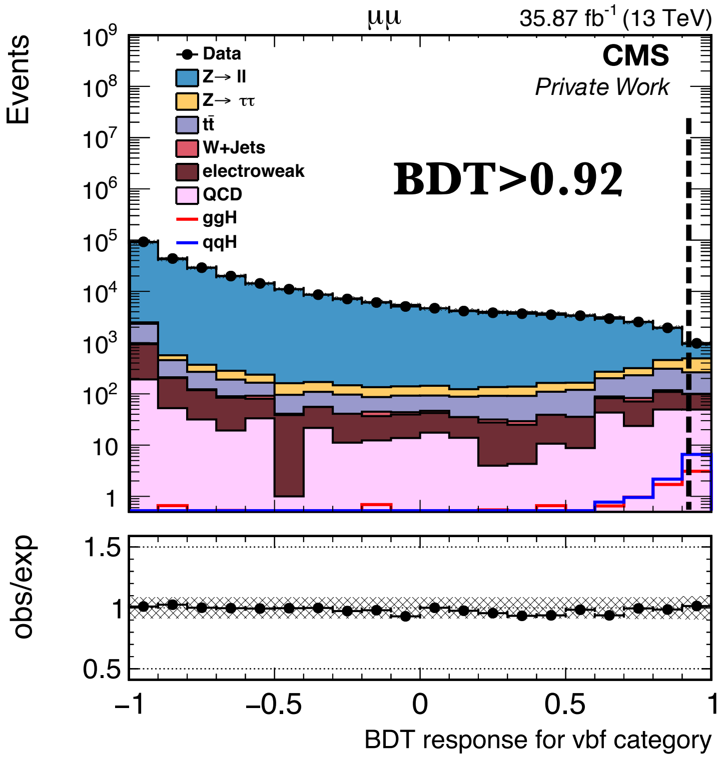
0-Jet



boosted

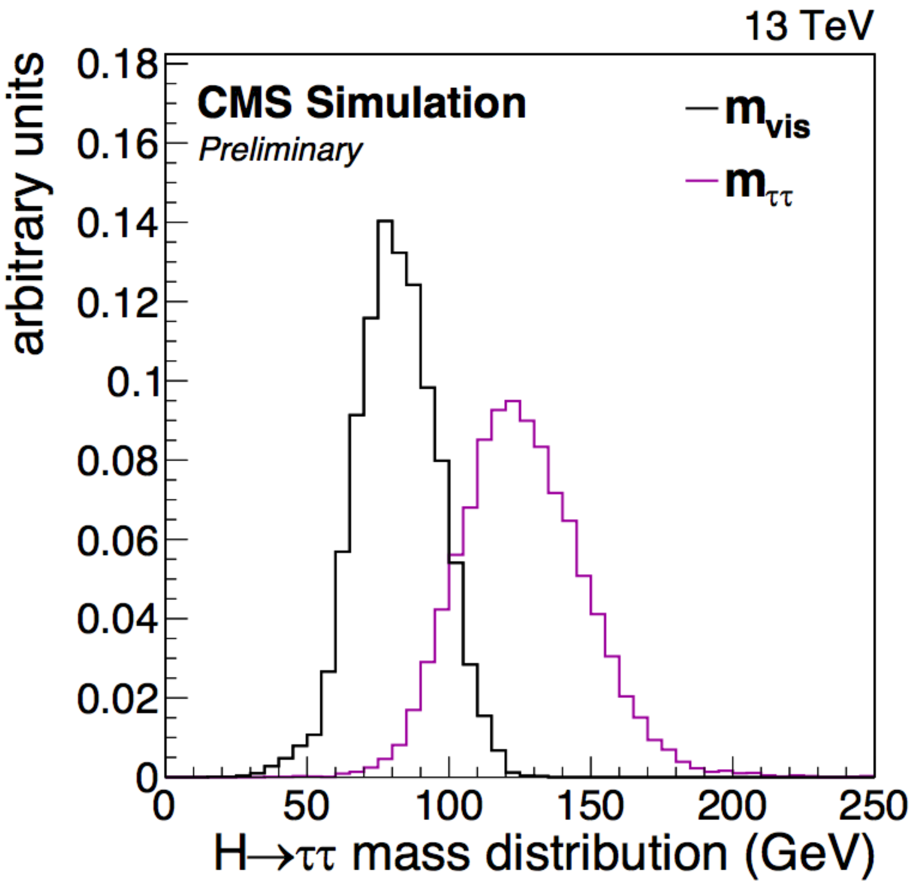


VBF



- BDT cuts are optimized by maximizing the sensitivity for signal and minimizing the upper limit on $\mu = \sigma/\sigma_{SM}$
- For SVFit calculation a loose cut on the BDT is applied

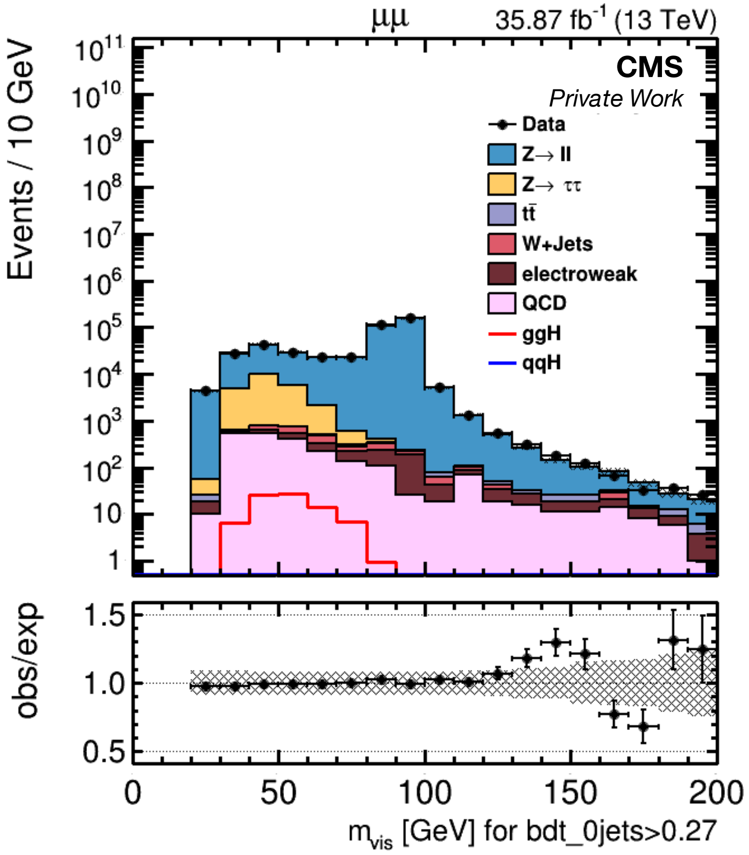
Signal Extraction Method



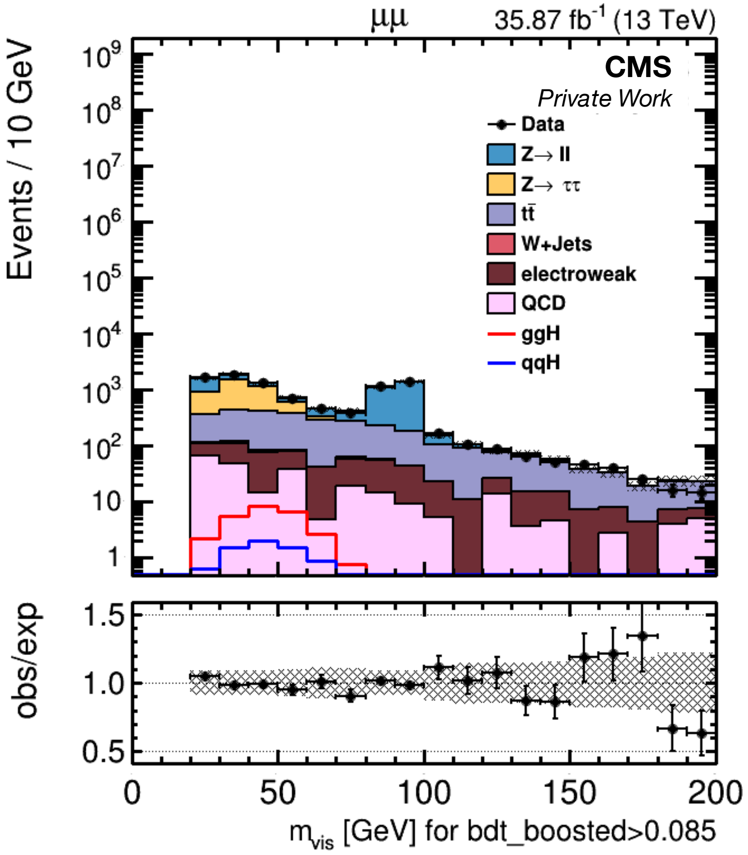
- Di-tau mass ($m_{\tau\tau}$) is reconstructed using **S**econdary **V**ertex **Fit** (SVFit)
  - It is a likelihood-based estimation of the parent boson mass
  - Inputs used for the calculations are: MET, MET uncertainties and four-vectors of the muon candidate
- Signal is extracted using 2-dimensional distribution in the plane of reconstructed SVFit mass $m_{\tau\tau}$ and visible mass $m_{\mu\mu}$

Mass Distributions – Visible mass ($m_{\mu\mu}$)

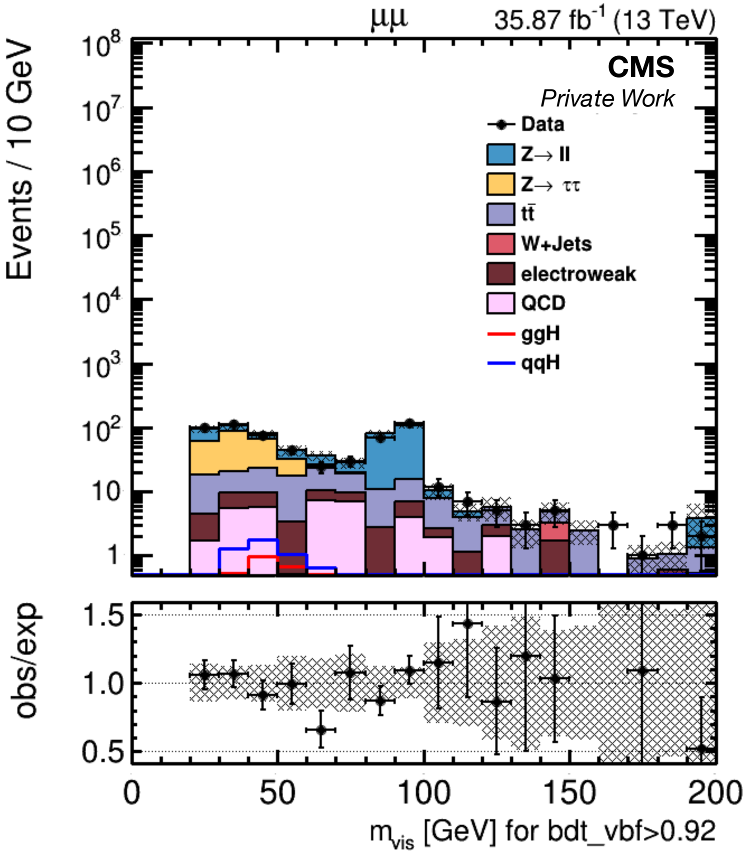
0-Jet



boosted

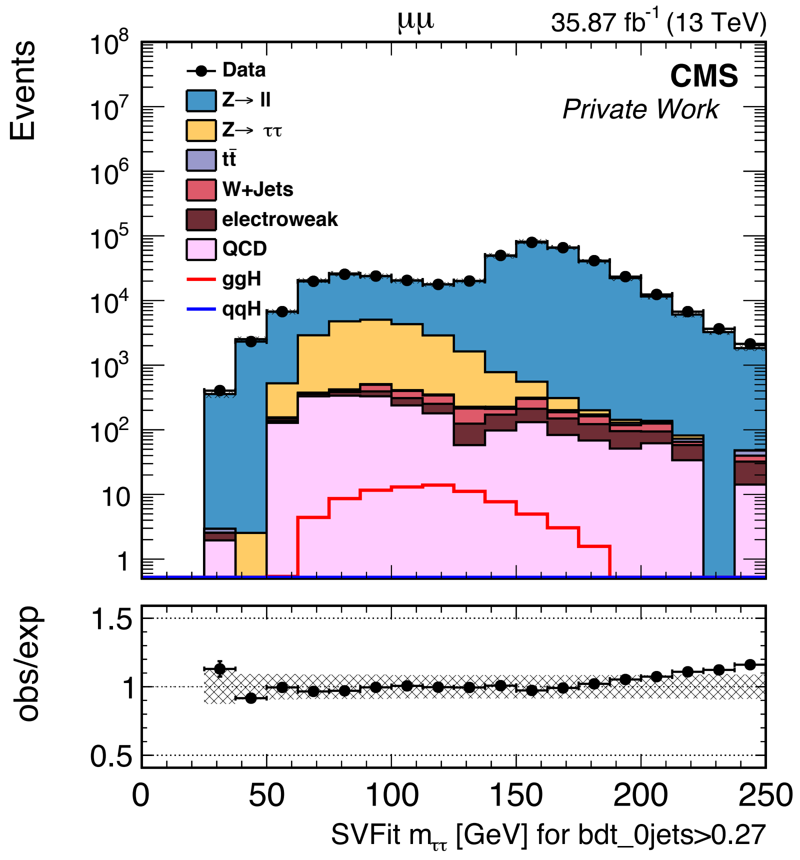


VBF

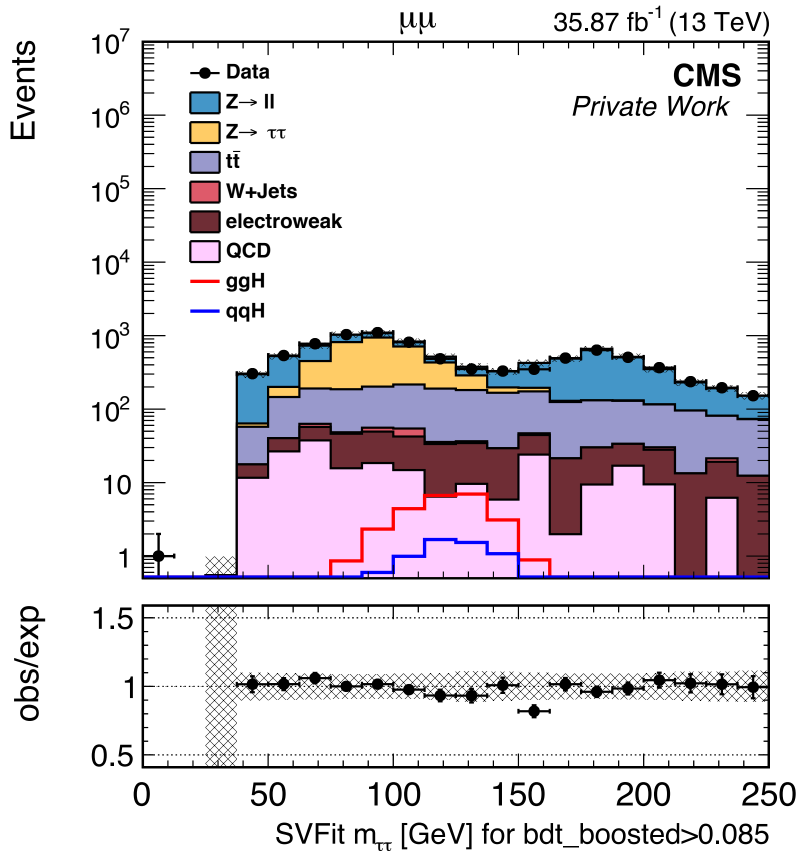


Mass Distributions - SVFit mass ($m_{\tau\tau}$)

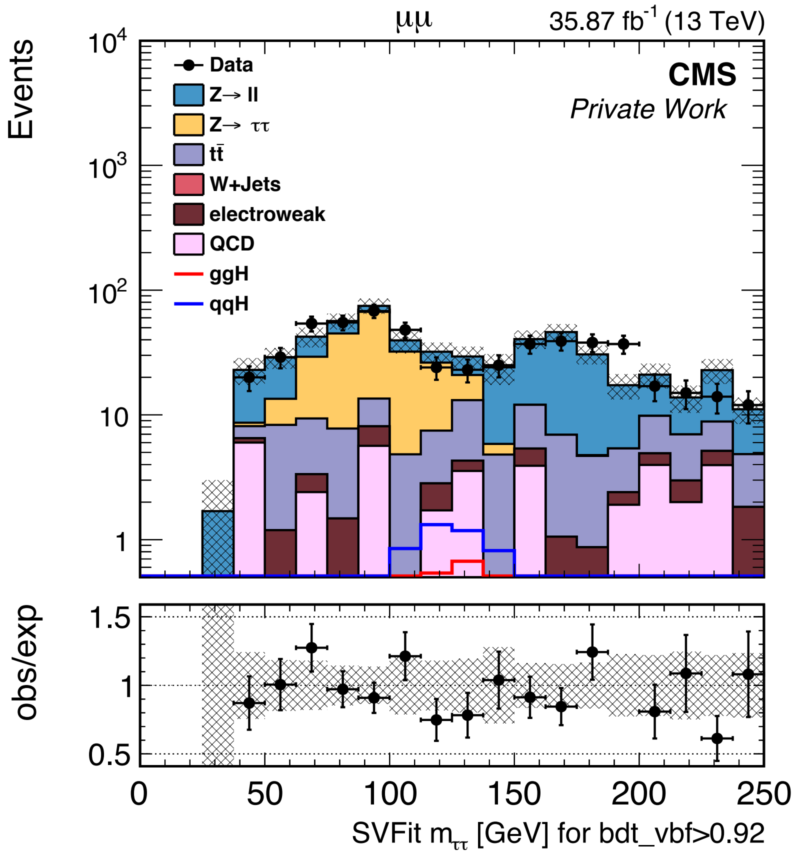
0-Jet



boosted

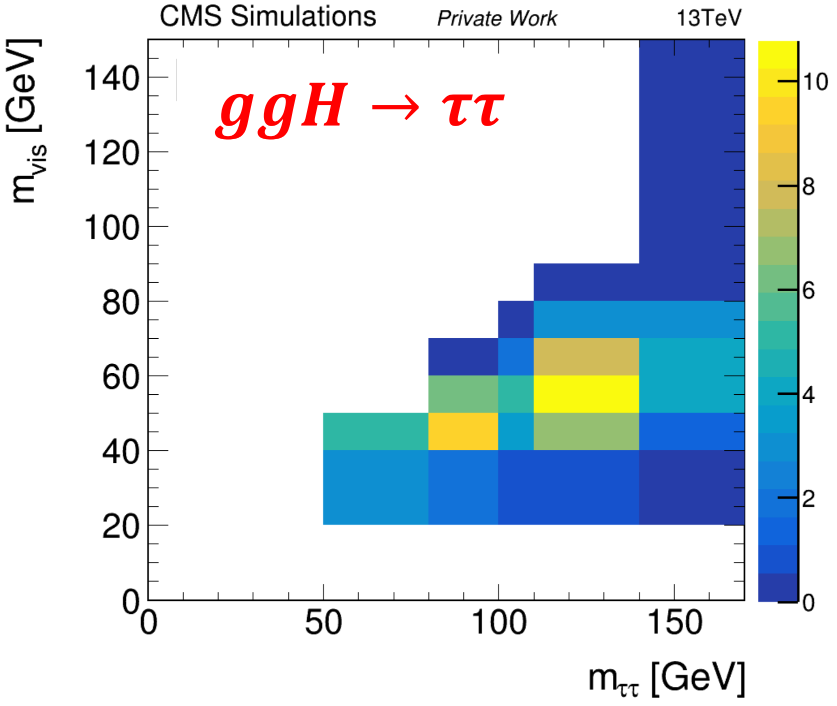


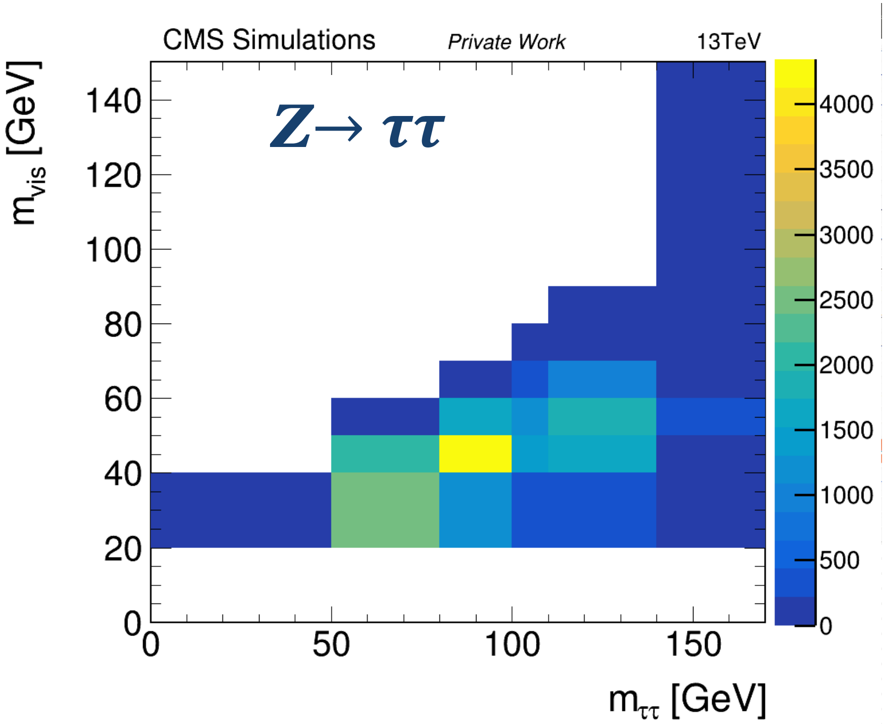
VBF

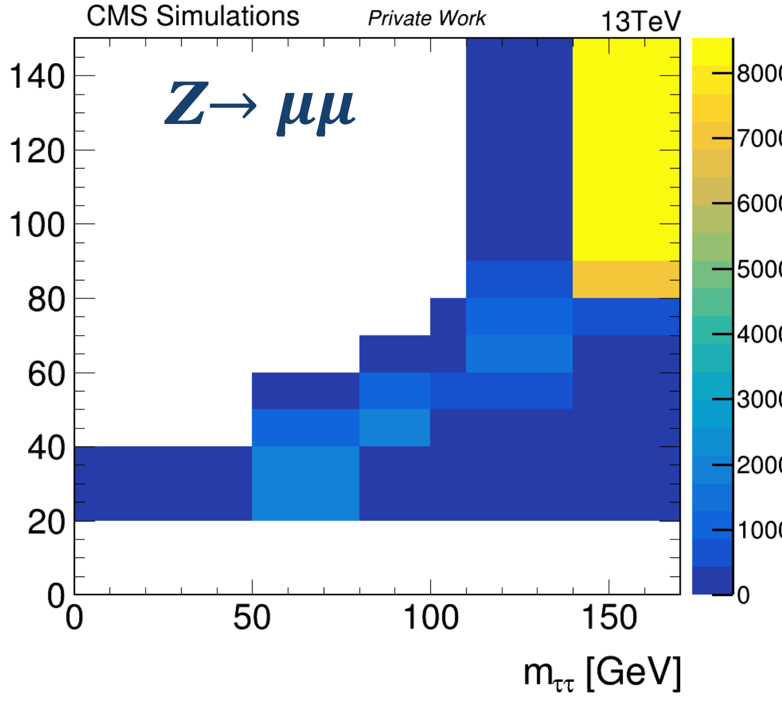


2-D Mass distribution (0-jet)

After applying all selections

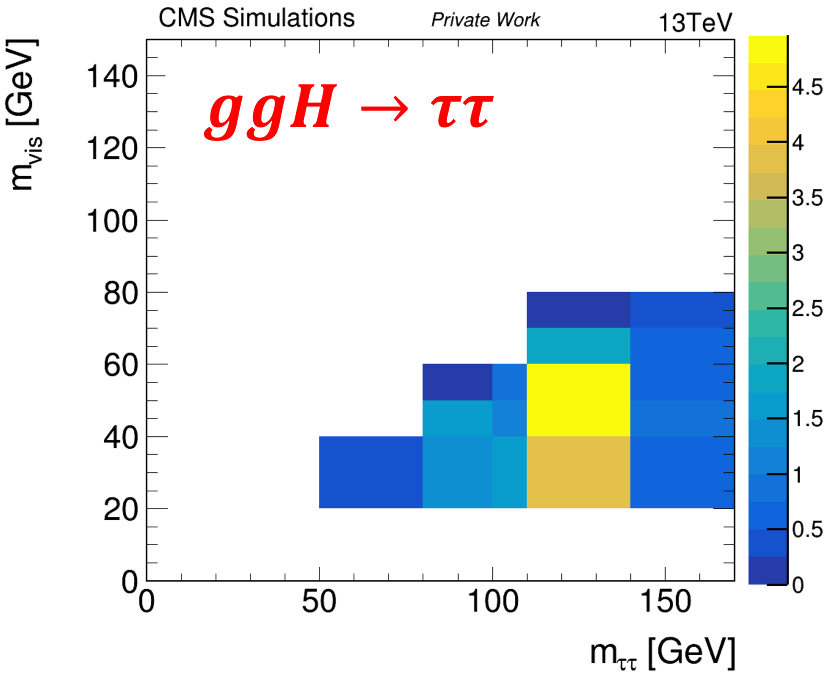


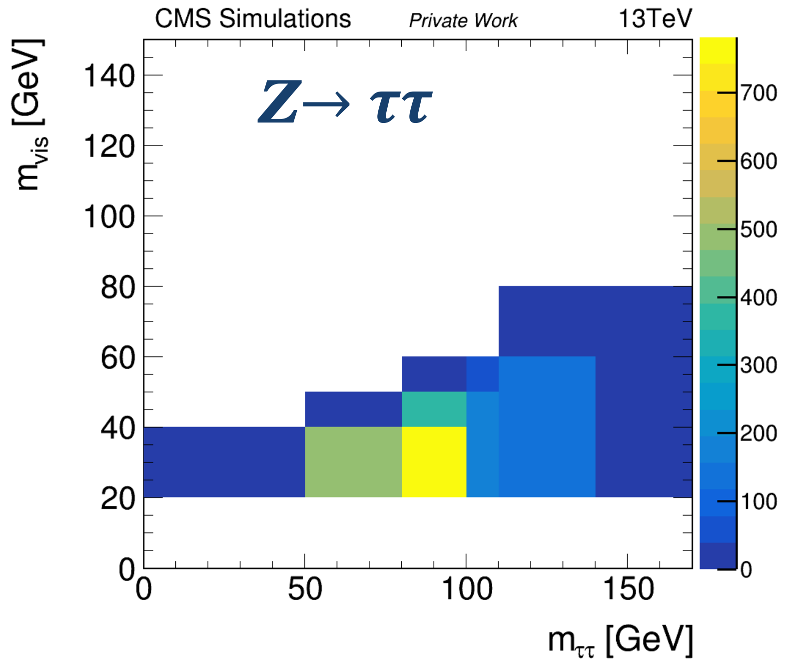


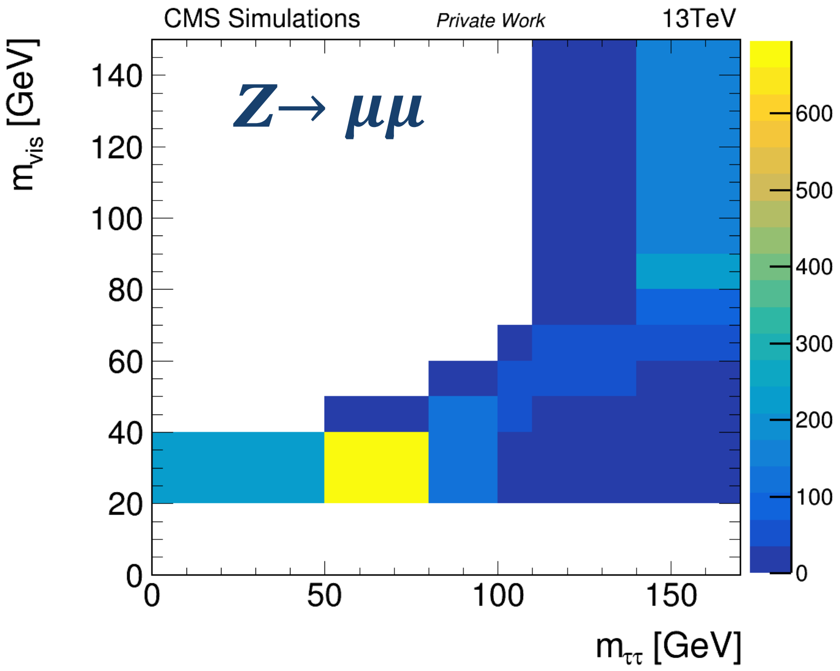


2-D Mass distribution (boosted)

After applying all selections

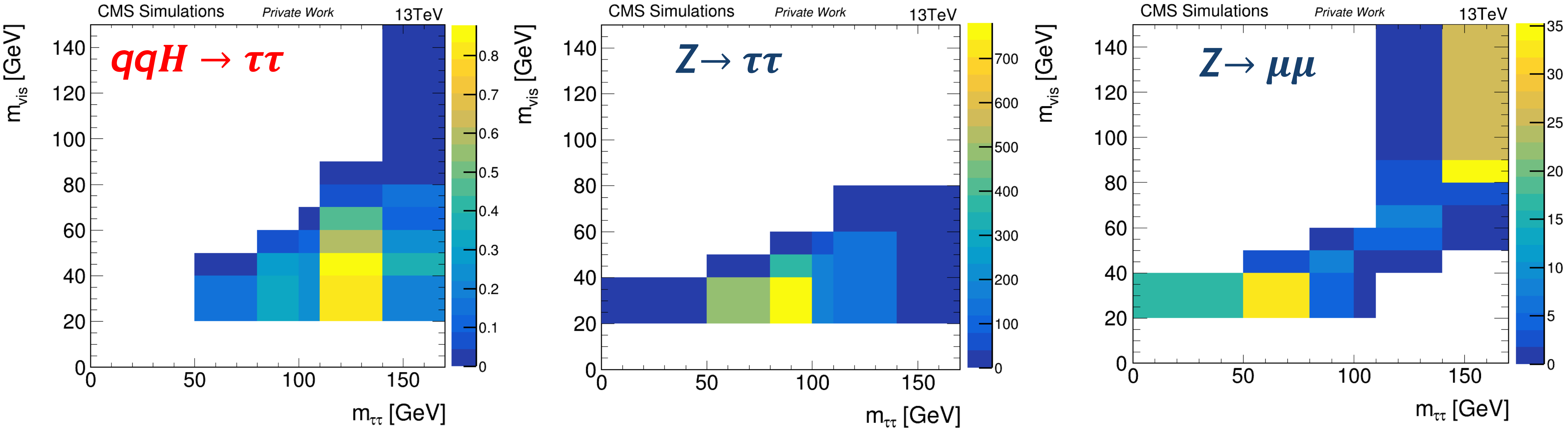






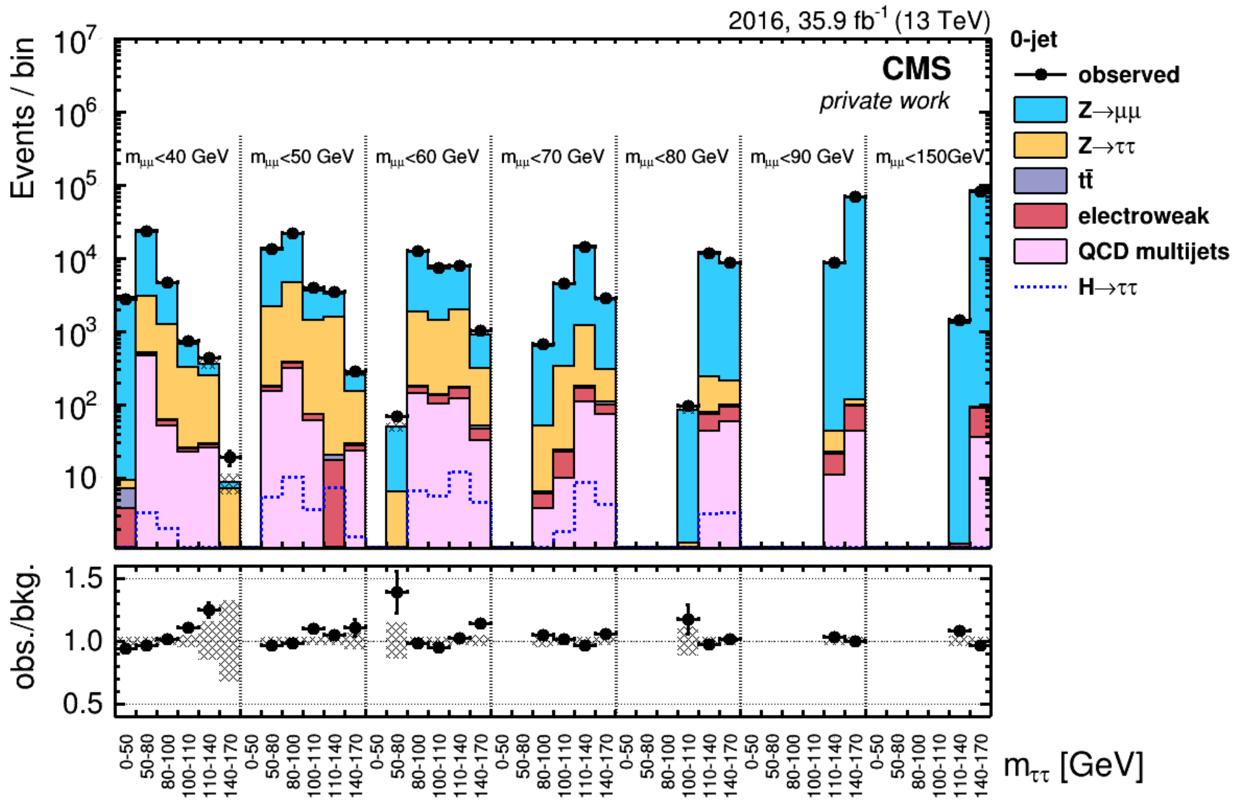
2-D Mass distribution (VBF)

After applying all selections

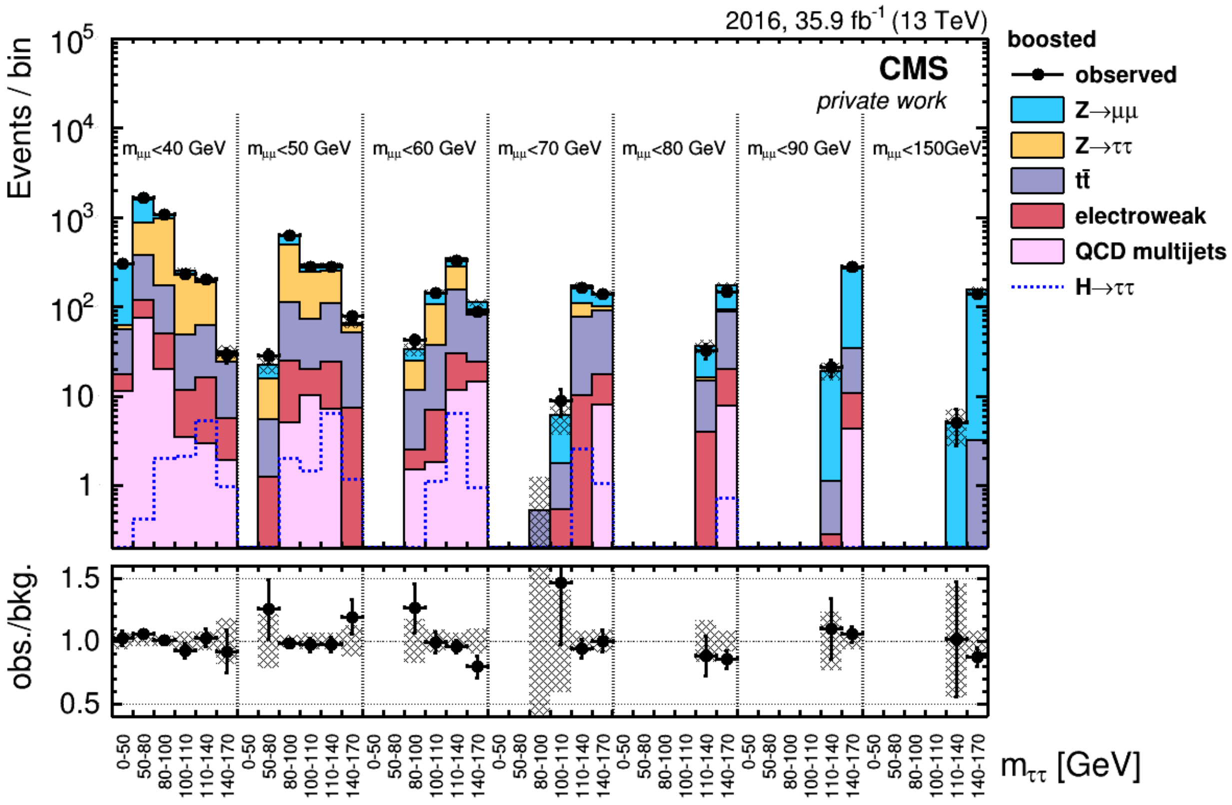


Prefit Plots

0-Jet

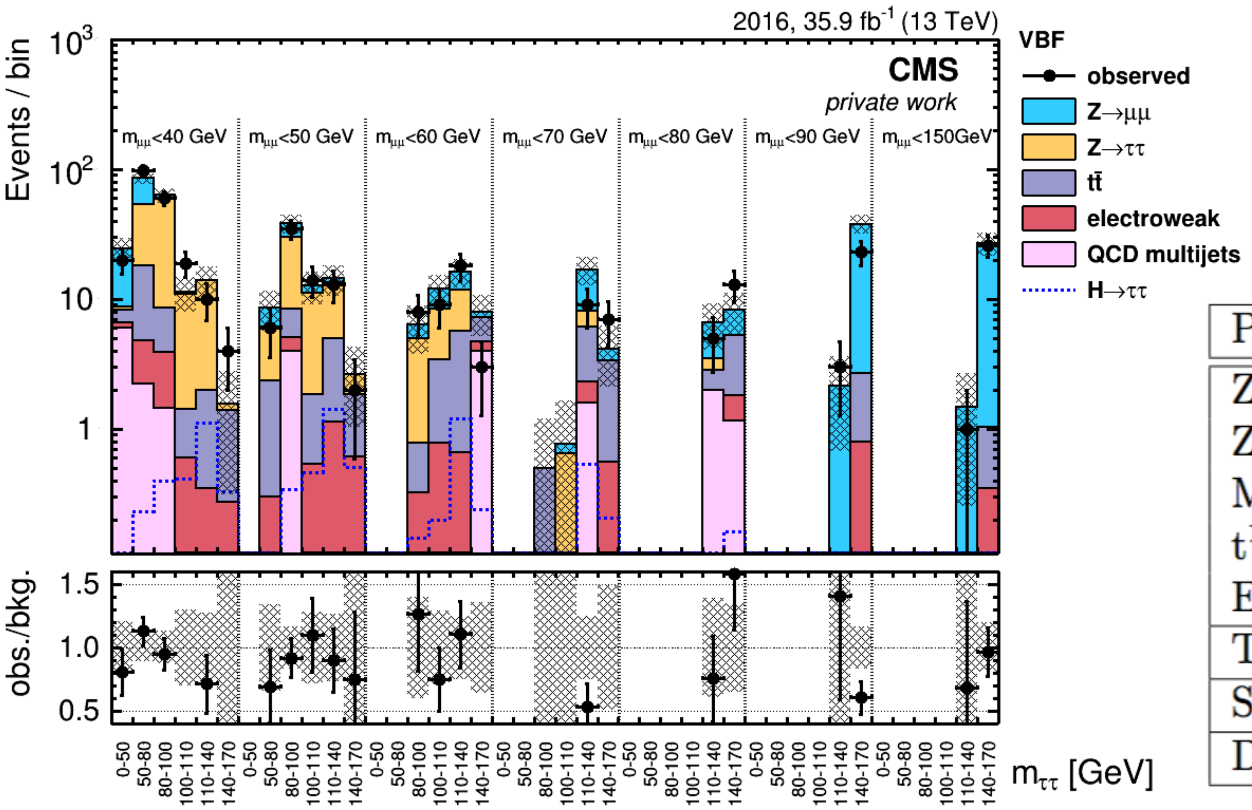


boosted



Prefit Plots

VBF



- The expected number of background events in this channel in the data corresponding to the integrated luminosity 35.9 fb^{-1}

Process	0-jet	boosted	VBF
$Z \rightarrow \tau\tau$	$20\,502 \pm 156$	$2\,585 \pm 37$	171 ± 12
$Z \rightarrow \mu\mu$	$288\,856 \pm 928$	$2\,027 \pm 37$	155 ± 11
Multijet	$1\,900 \pm 150$	182 ± 36	22 ± 10
$t\bar{t}$	111 ± 7	$1\,266 \pm 23$	60 ± 5
Electroweak	560 ± 7	240 ± 7	16 ± 2
Total Expected Background	$31\,2495 \pm 561$	$6\,336 \pm 80$	425 ± 21
Signal $H \rightarrow \tau\tau$	86 ± 3	28 ± 2	6 ± 0
Data	308 013	6 342	405

Uncertainty Model

Normalization Uncertainties

- Luminosity 2.6%
- Identification/Isolation Efficiencies 2% for each muon
- Background (Systematic uncertainties on the sideband)
  - QCD multijets 20%
  - $t\bar{t}$ 7%
  - Electroweak 15%

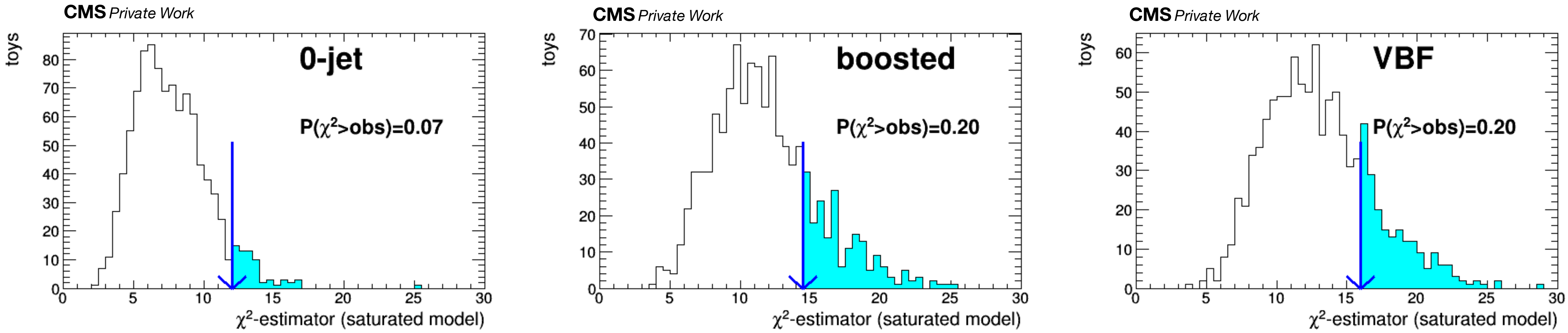
Shape Uncertainties

- Muon momentum scale 1% (conserved number)
- Top p_T reweighting
- DY reweighting
- b-tag efficiency
- Jet Energy Scale (JES)
- Missing Transverse Energy response

These uncertainties are used as nuisance parameters in likelihood function. The expected and observed limits are extracted by maximizing the likelihood function for each bin of the unrolled mass.

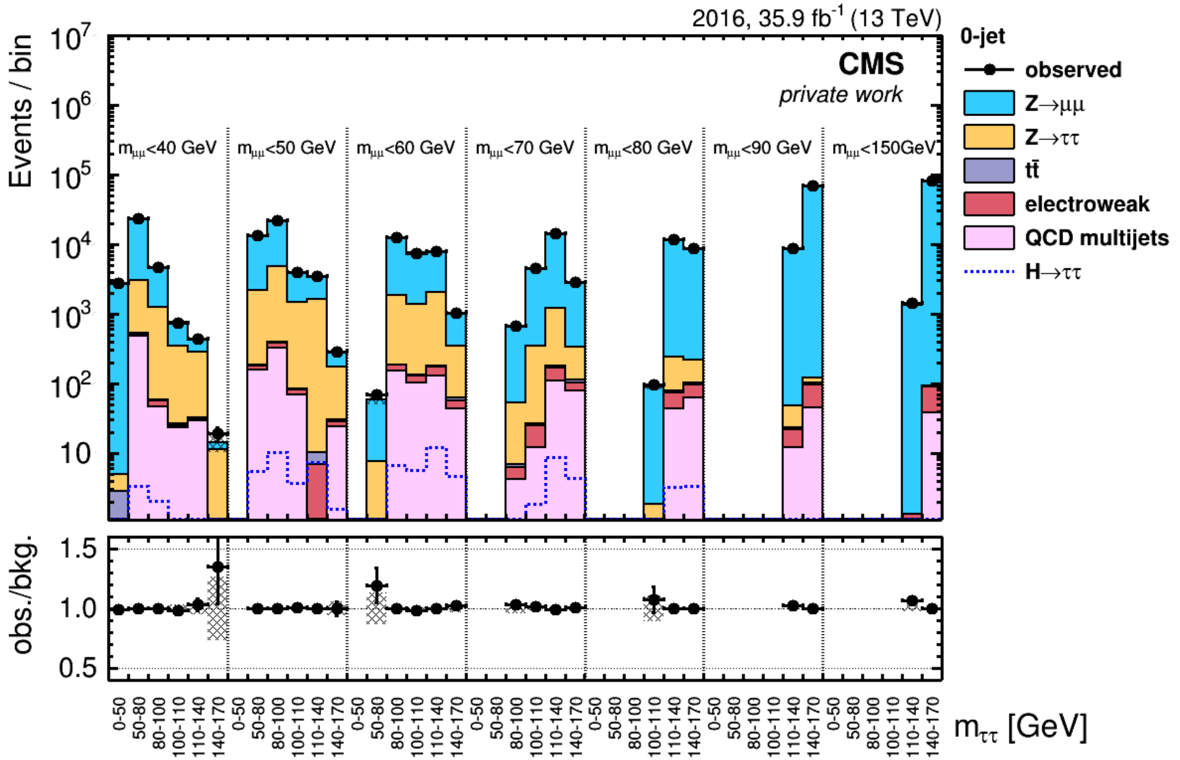
Goodness of Fit

Using 1000 toy runs

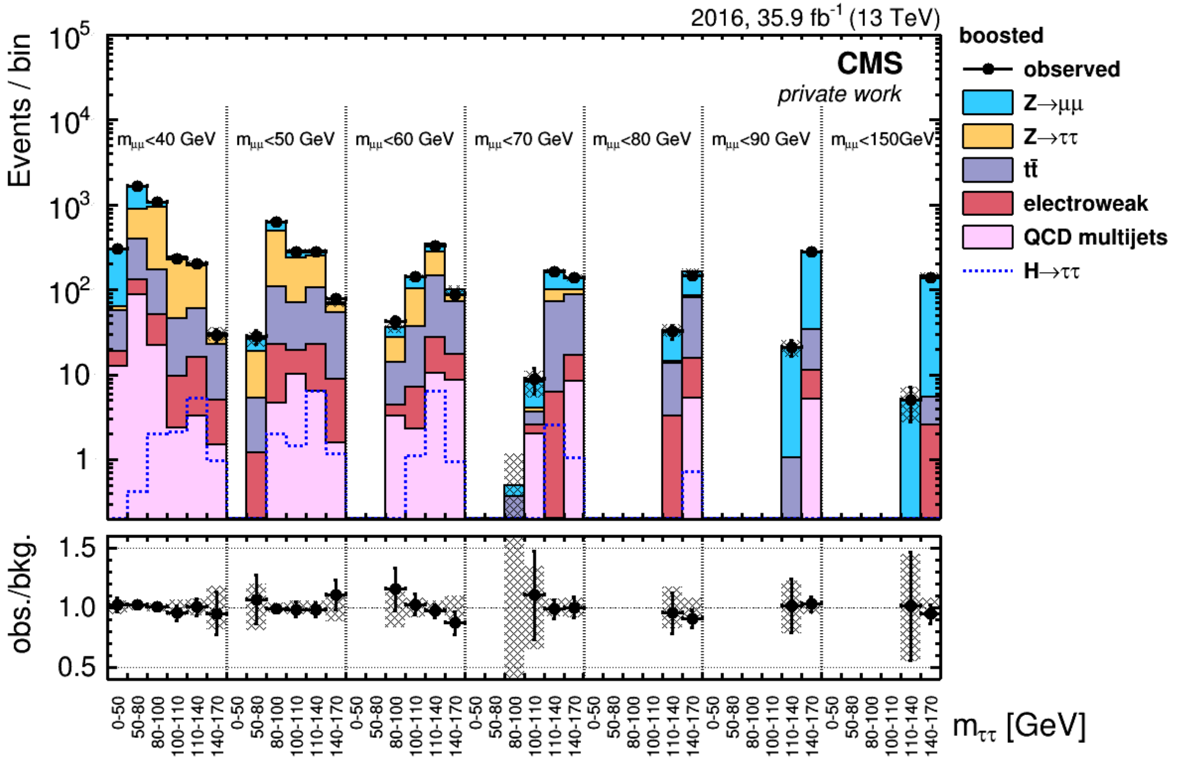


Postfit Plots

0-Jet

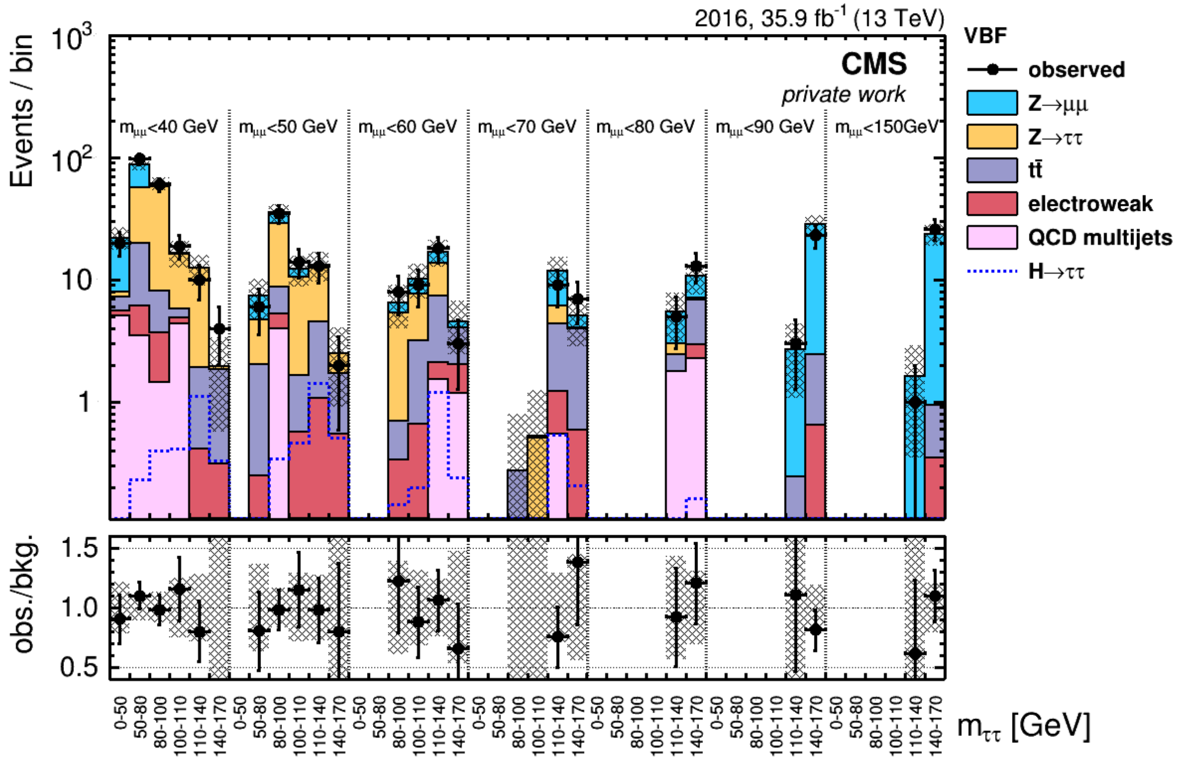


boosted



Postfit Plots

VBF



- The event yield in the $H \rightarrow \tau\tau$ signal and respective background events in this channel in the data corresponding to the integrated luminosity 35.9 fb^{-1}

Process	0-jet	boosted	VBF
$Z \rightarrow \tau\tau$	20 851 ± 144	2 585 ± 51	167 ± 13
$Z \rightarrow \mu\mu$	283 967 ± 533	2 044 ± 45	134 ± 12
Multijet	1 973 ± 45	194 ± 14	26 ± 5
$t\bar{t}$	110 ± 11	1 248 ± 35	59 ± 8
Electroweak	550 ± 23	229 ± 15	16 ± 4
Total Background	308 032 ± 555	6332 ± 80	402 ± 20
Signal $H \rightarrow \tau\tau$	86 ± 3	28 ± 2	6 ± 0
Data	308 013	6 342	405

Fitted Signal Strength and Limits

Signal Strength relative to SM

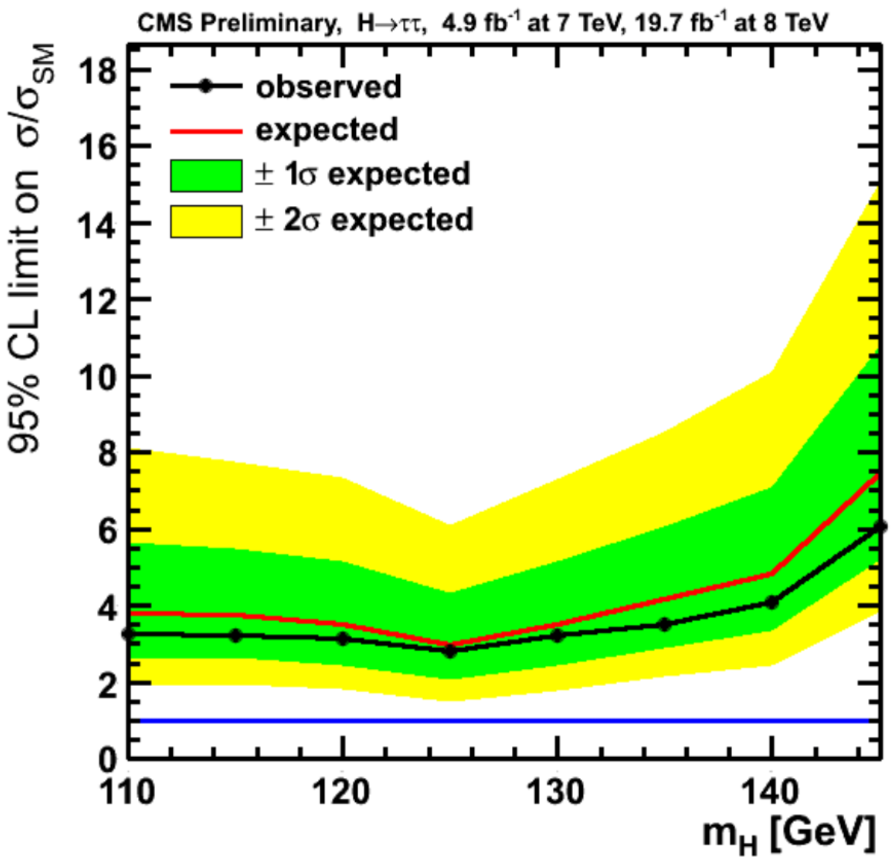
Category	Best Fit
0-jet	12.1±13.0
boosted	-2.3±2.8
VBF	0.4±1.9
Combination	-1.0±1.7

Upper 95% CL limits on the signal strength in mumu channel

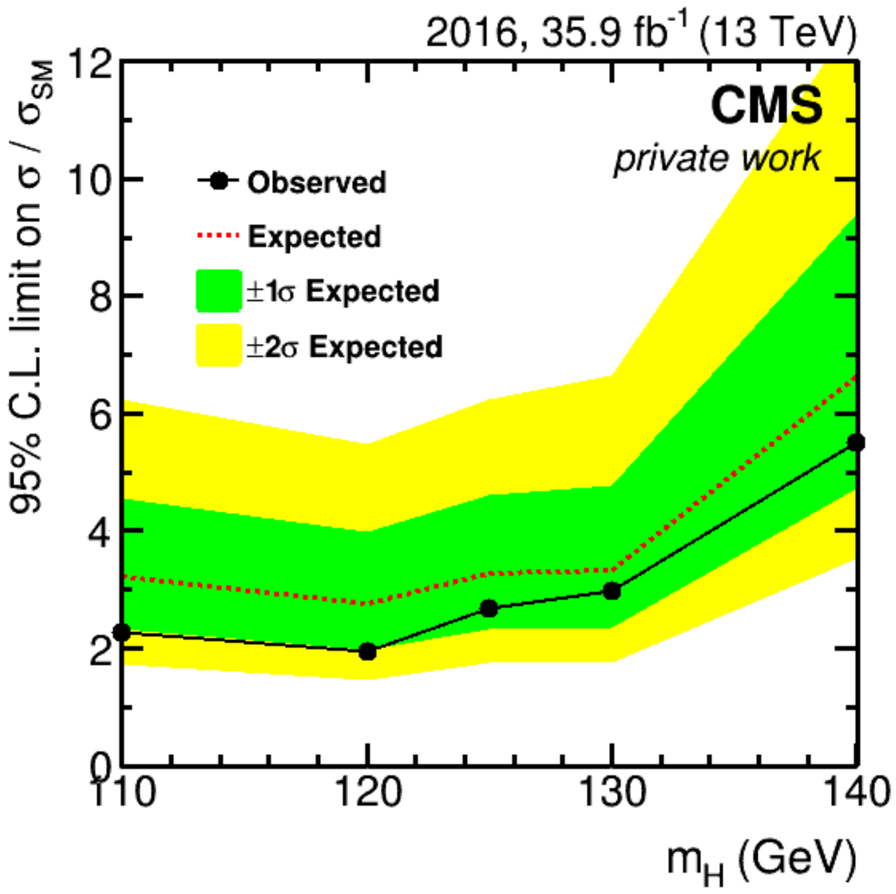
Category	-2σ	-1σ	exp.	$+1\sigma$	$+2\sigma$	**obs.**
0-jet	12.7	17.0	23.8	33.6	45.8	**34.9**
boosted	2.8	3.8	5.3	7.4	10.0	**4.0**
VBF	2.1	2.8	4.0	5.7	8.0	**4.2**
Combination	1.7	2.3	3.2	4.6	6.2	**2.7**

Limit Plots

Run I ($\tau_\mu \tau_\mu$ Channel)



Run II ($\tau_\mu \tau_\mu$ Channel)



Summary

- Successfully implemented Single BDT method to suppress the DY dominating background in each event category.
- The signal strength in $H \rightarrow \tau\tau \rightarrow \mu\mu$ channel relative to SM is

$$\boldsymbol{\hat{\mu} = \sigma/\sigma_{SM} = -1.00 \pm 1.7} \text{ (Run II)}$$

- The expected and observed 95% CL limits on signal strength in the combination category are

3.2 (Exp.) and 2.7(Obs.)

Thanks!!!!

Acknowledgement

My sincere thanks to:

Dr. Marcus Hohlmann (Thesis and research advisor at Florida Tech)

Dr. Alexei Raspereza (co-research advisor at DESY)

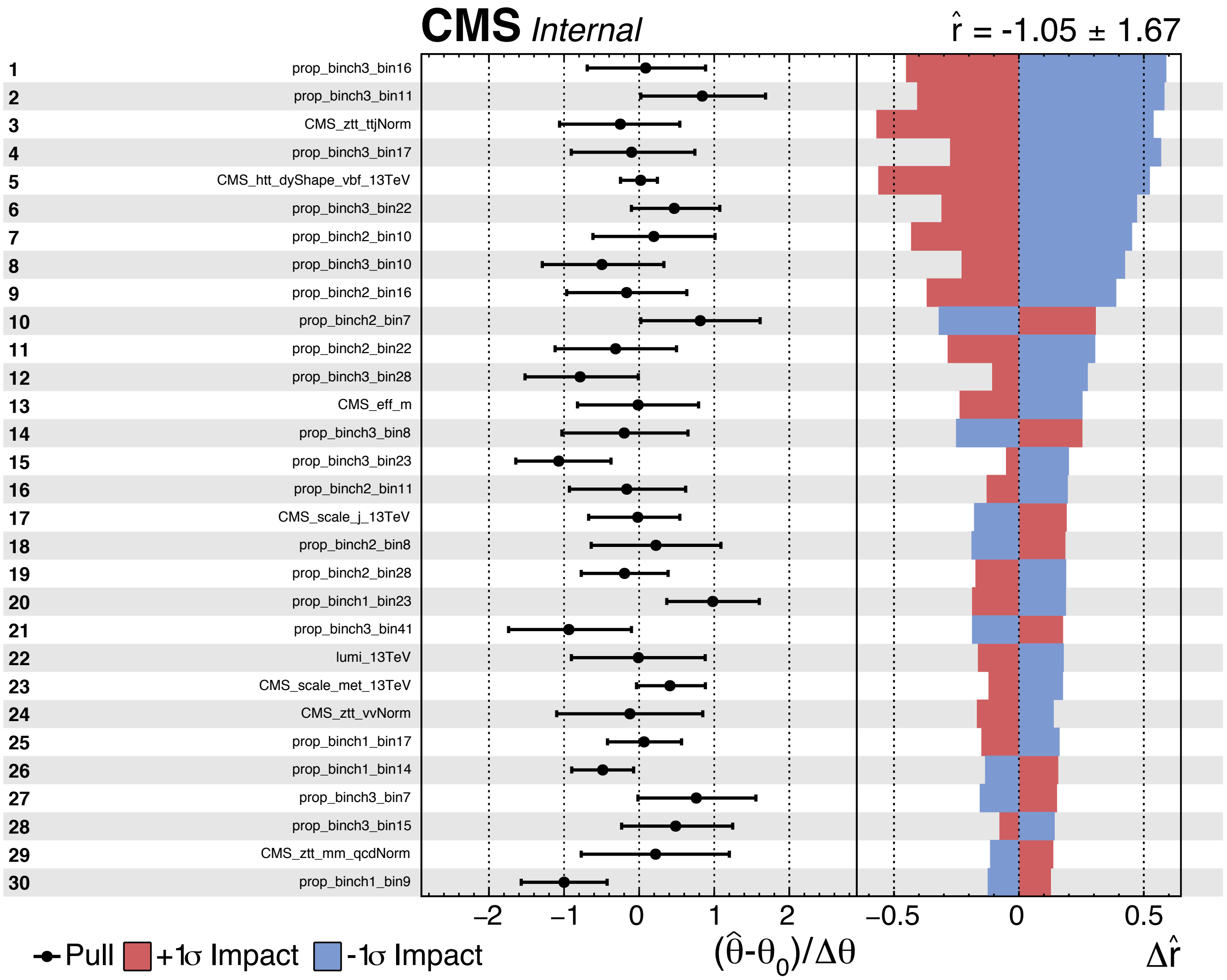
Dr. Elisabetta Gallo (co-research advisor at DESY)

DESY HTT group

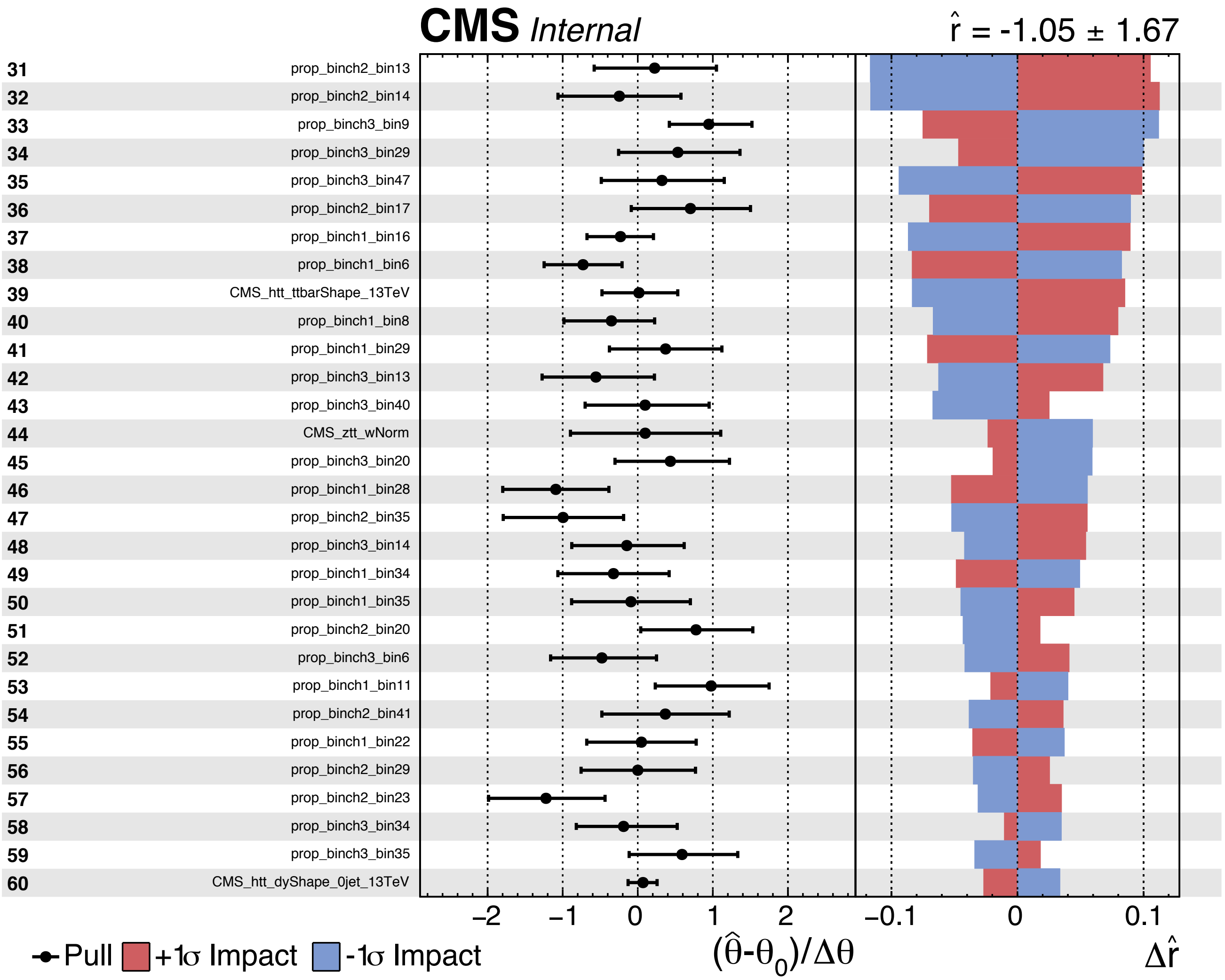
CMS HTT group

Backup Slides

Pulls and Impact (combine) Page1



Pulls and Impact (combine) Page 2



Pulls and Impact (combine) Page 3