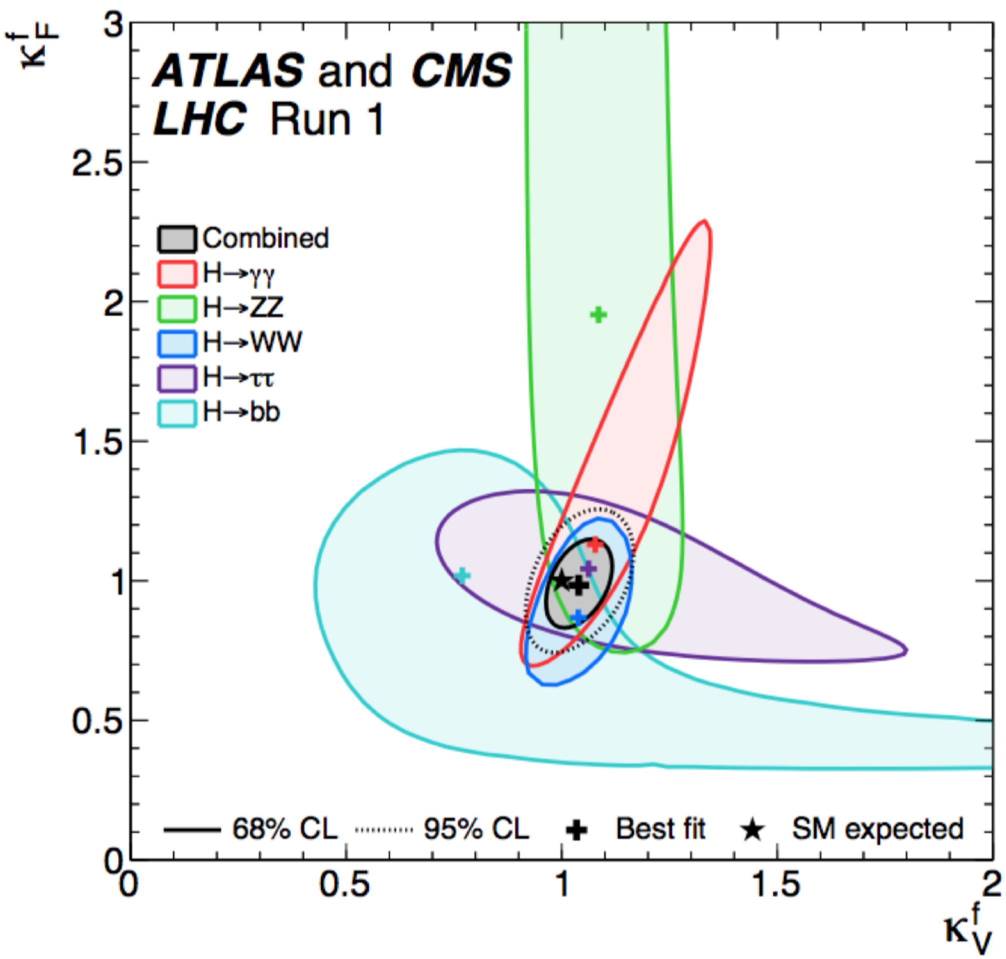
Standard Model Higgs Boson Decaying to Fermions at CMS

SUSY17 – TIFR, Mumbai, India
11 December 2017

Tyler Ruggles
on behalf of the CMS Collaboration

The Standard Model Higgs Boson

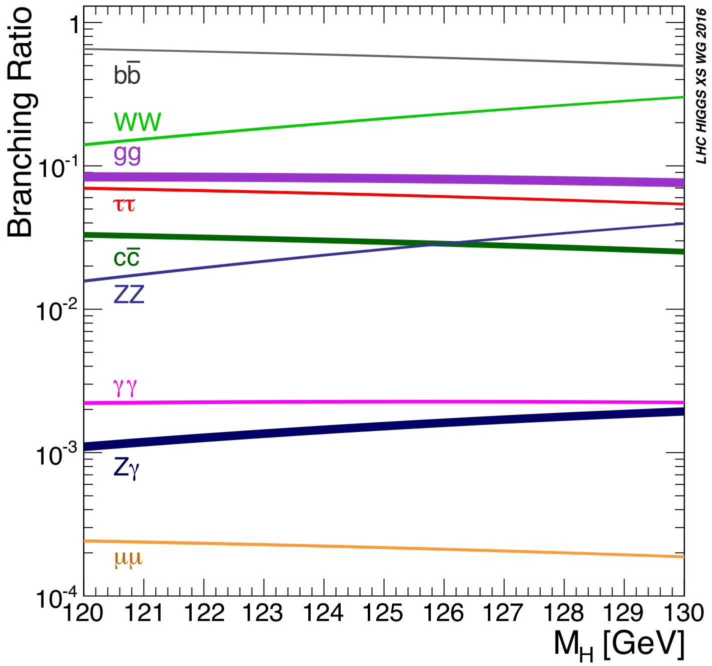
- Discovered with very high significance in multiple decay channels by both CMS and ATLAS
  - Highest significances in H→ZZ/WW/γγ final states accounting for **< 25% of Higgs decay branching ratio**

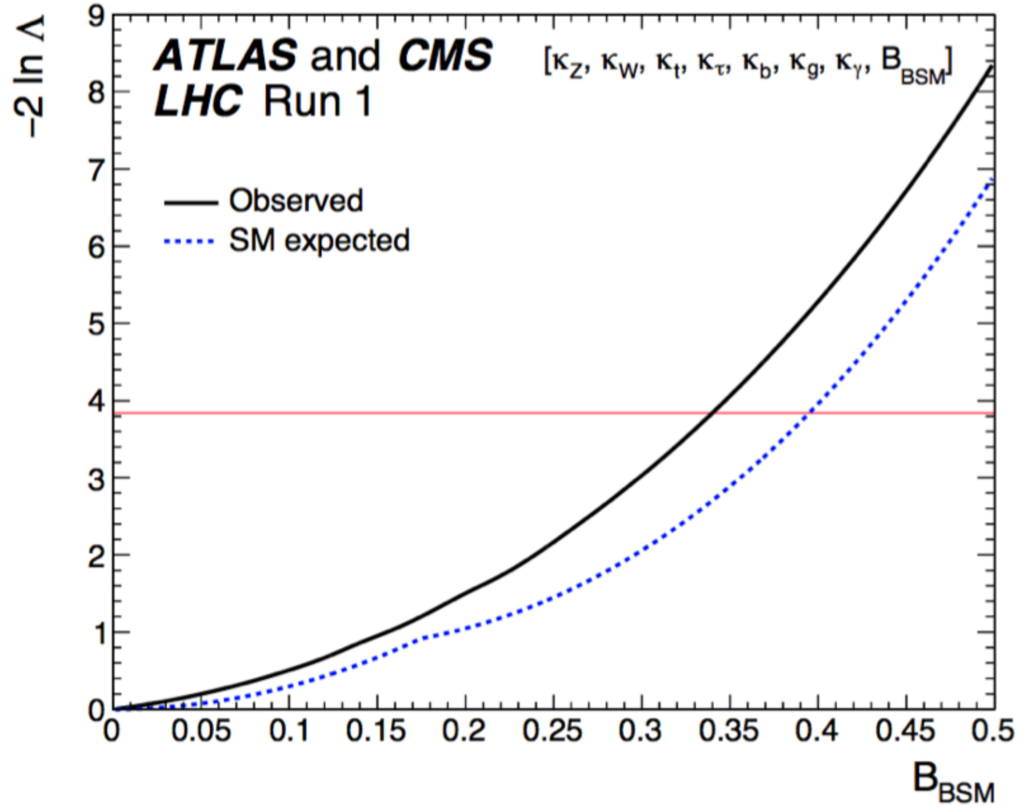


arXiv:1606.02266: https://arxiv.org/abs/1606.02266

The Standard Model Higgs Boson

- The SM-Higgs decaying to fermion analyses, H→bb/ττ, face unique analysis challenges despite composing **65% of the Higgs decay BR**
  - Provide a unique window to test the **Yukawa coupling to fermions**
  - H(bb) – unique final state to measure coupling to down type quarks
  - In the spirit of a SUSY conference, the more constraint we have on all Higgs decay processes, the less room there is for BSM physics from H(125)





arXiv:1606.02266: https://arxiv.org/abs/1606.02266

VH(bb)

- VHbb: arXiv:1709.07497 – submitted to PLB
- CMS-HIG-16-044
- Experimental challenges
  - Pileup, Jet Energy, and MET
  - B-jet tagging
  - Trigger thresholds

VH(bb)

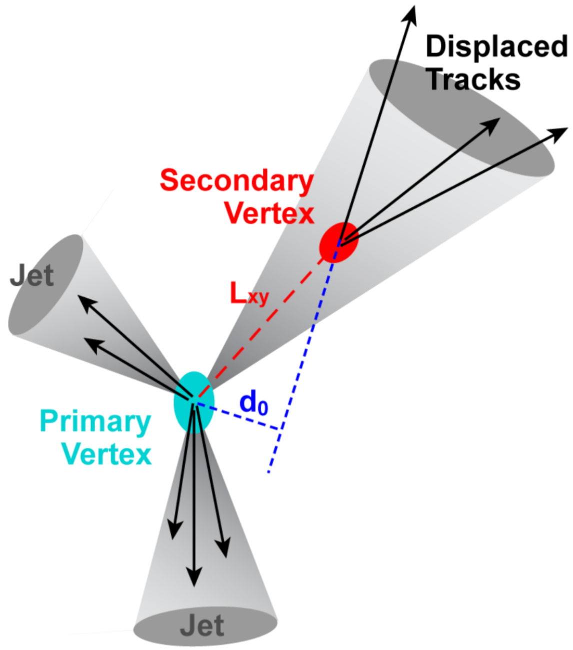
Fundamentals:

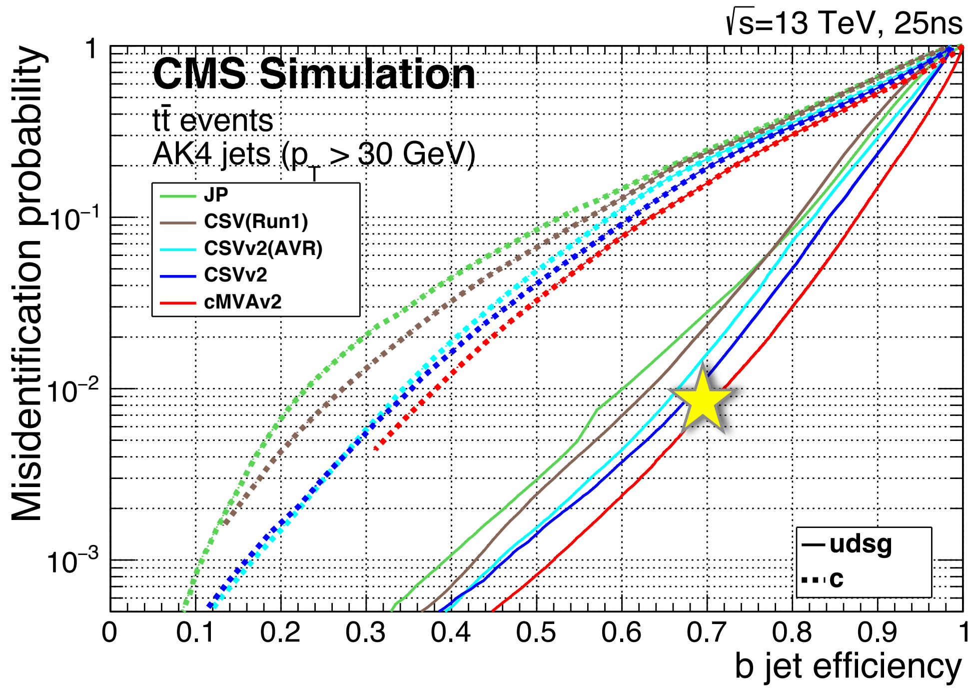
- Multiple final states targeting events with 2 b-tagged jets
  - Z(νν)H, W(e,μ)H, Z(ee,μμ)H
- Higgs candidate defined as jet pair with highest b-tagged
- Categorize based on pT(V)
  - 1-lepton: pT(V)>100 GeV
  - 2-lepton, 2 categories: 50 GeV < pT(V) < 150 GeV, pT(V) > 150 GeV
- Event BDT including di-jet mass of b-jets, b-tag values, kinematic variables
- Signal and Control Regions defined with multiple additional cuts including:
  - Number of jets, Number of b-tagged jets, m_{jj}, Z mass window
- Simultaneous final fit of signal region and many control regions

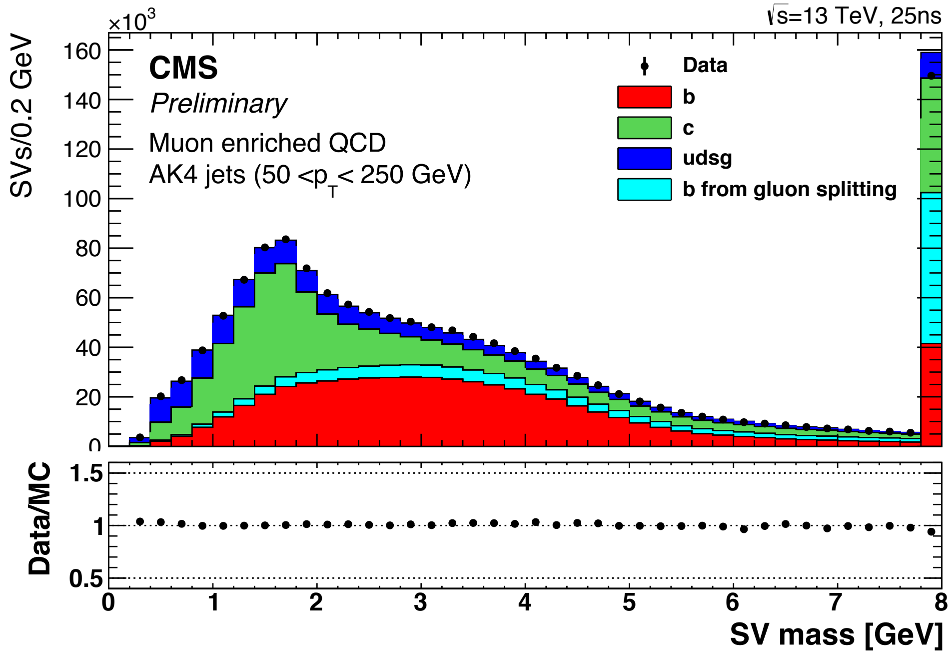
Category	Online Trig. Threshold (GeV)	Offline Trig. Threshold (GeV)
0-lepton	MET & HT Missing > 110	MET > 170
1-lepton	pT(μ) > 24 pT(e) > 27	pT(μ) > 25 pT(e) > 30
2-lepton	2 muon (17,8) 2 electron (23,12)	2 muon (20,20) 2 electron (25,20)

b-tagging

- Best separation of udscg jets vs. b-jets seen with the cMVAv2 algorithm
  - cMVAv2 combines 6 difference b-jet discriminants into BDT
- Includes secondary vertex (if any), tracks, impact parameters, soft leptons and more
- **~70% efficiency for <1% misidentification rate**



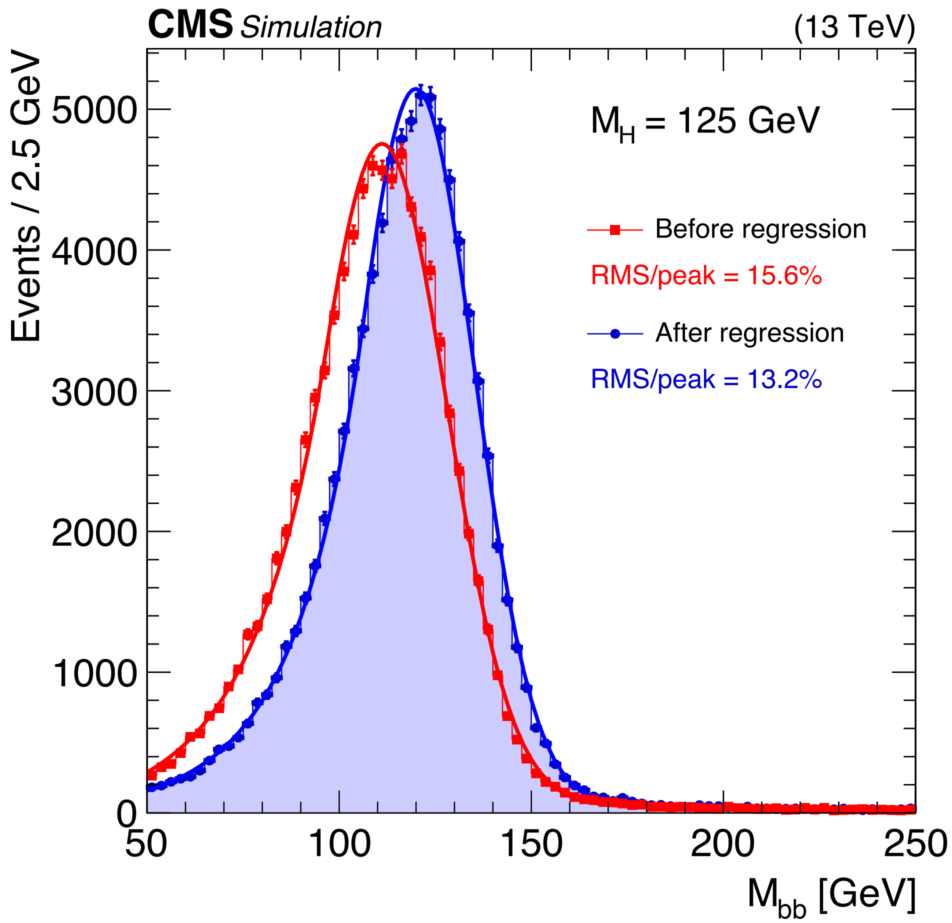




CMS-PAS-BTV-15-001

b-jet Energy Regression

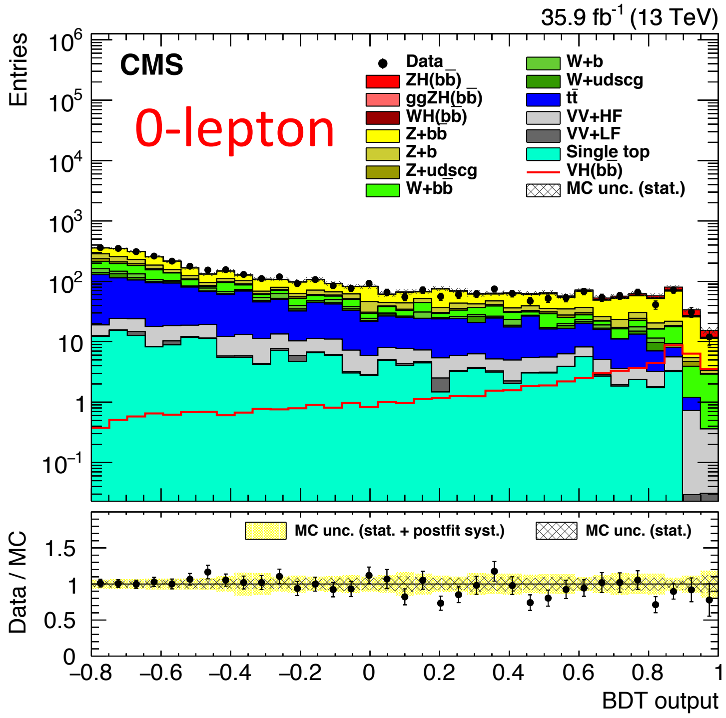
- Improve b-jet energy resolution with BDT regression
- Trained with gen-matched b-jets in ttbar MC
- Input kinematics: pT(b), energy, leading charged tracks, soft lepton pT, secondary vertex info
- **15% improvement in resolution (wrt pre-regression resolution) & correction of peak position!**
- **Right,** m_{jj} distributions for MC Z(ℓℓ)H(bb) events (*m*H= 125 GeV), before (red) and after (blue) the energy correction from the regression procedure is applied



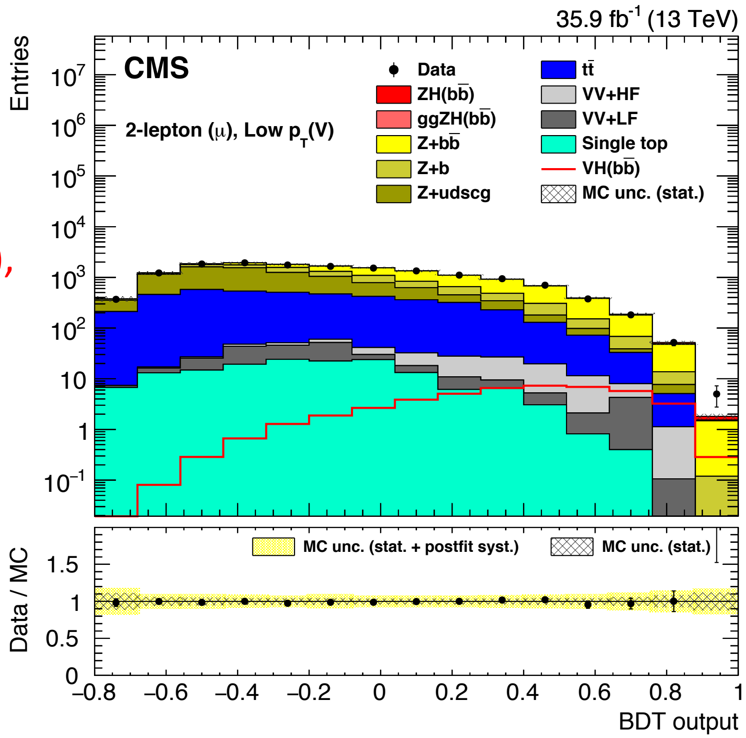
CMS-HIG-16-044

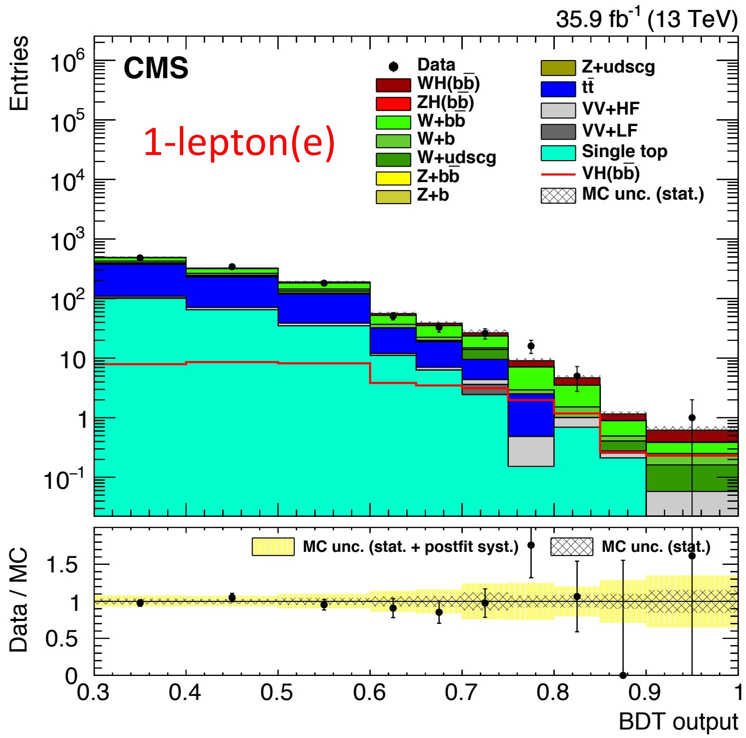
Backgrounds

- Dominant:
  - W/Z+Jets for 0/2-lepton
  - ttbar & single-top for 1-lepton
  - di-Bosons
- pT(V)-based corrections applied to W +Jets and ttbar
  - Derived such that pT(W) matches data in all control regions simultaneously



2-lepton(μ),
Low pT(V)



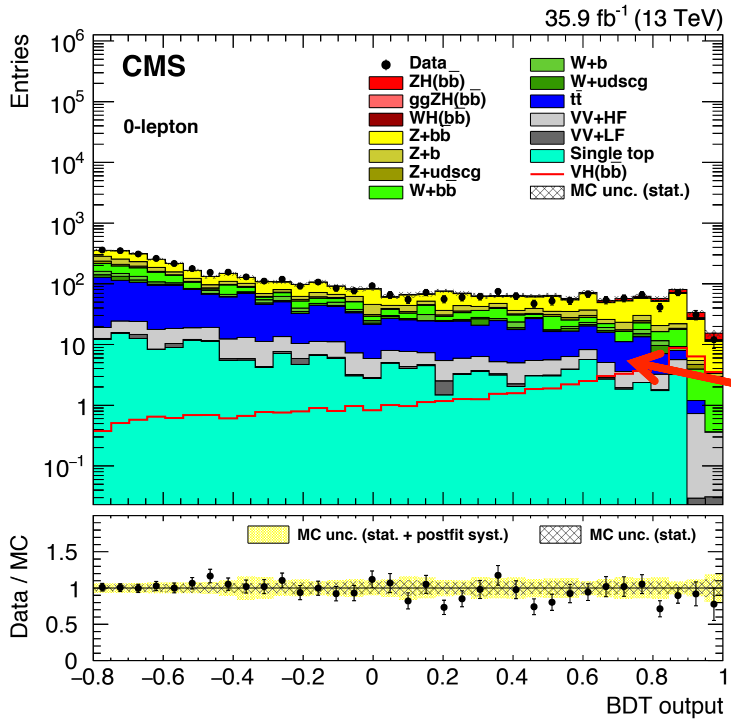


CMS-HIG-16-044

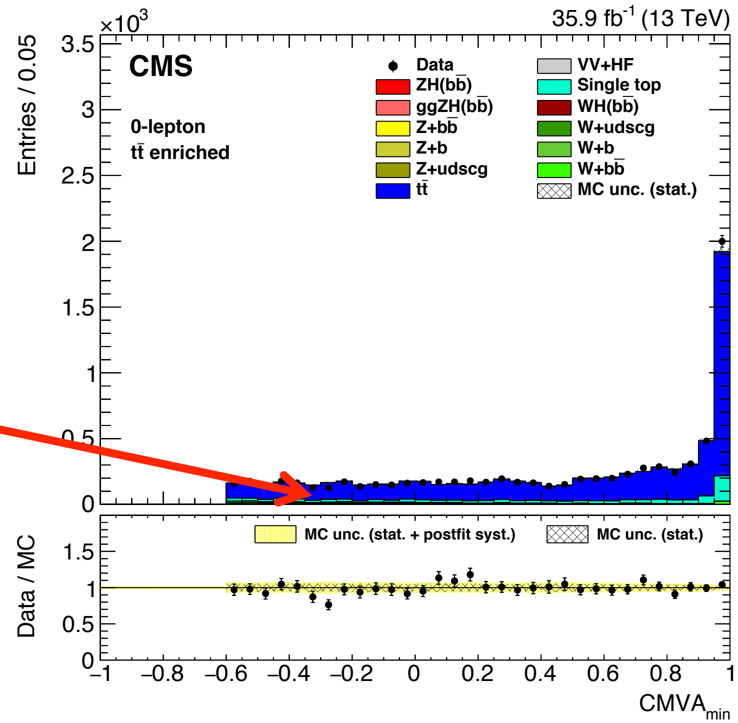
Control Regions – ttbar

- **Control regions** (CR) are are used to help **scale the yields of key background process** in the final fit
- One signal region (SR) cut in the 0-lepton category requires nJets <= 1
- Inverting this criteria provides a very pure ttbar CR
- Link normalization between SR and CR ttbar in final fit

0-lepton SR



0-lepton ttbar CR

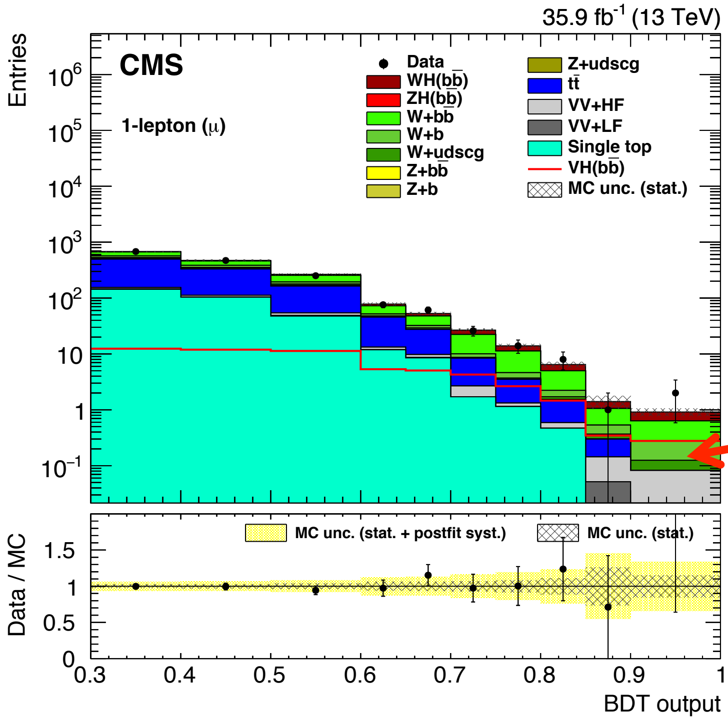


CMS-HIG-16-044

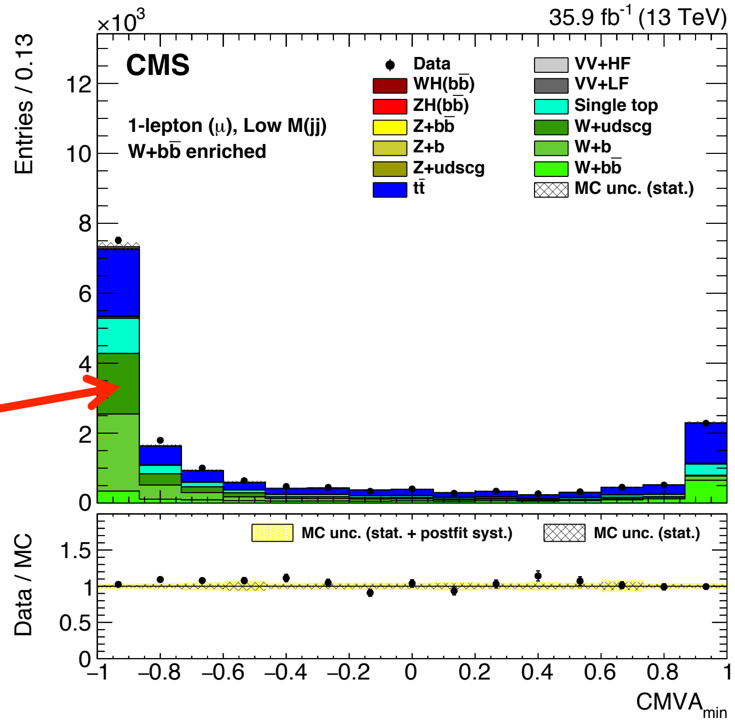
Control Regions – W+b(b)

- **Control regions** (CR) are are used to help **scale the yields of key background process** in the final fit
- In the 1-lepton(μ) category a W+b(b) CR can be defined by inverting the m_{jj} window SR cut
- Link normalization between SR and CR W+b(b) processes in final fit

1-lepton(μ) SR



1-lepton(μ) W+b(b) CR

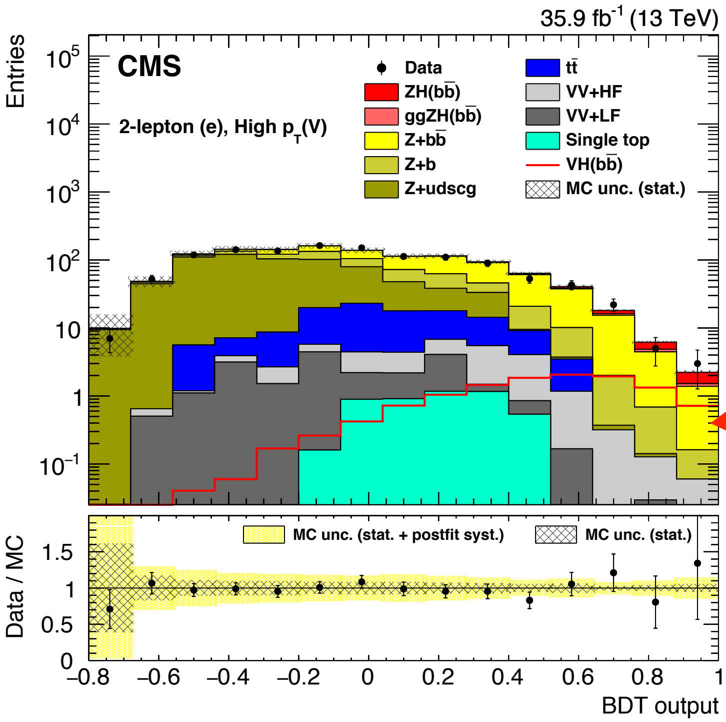


CMS-HIG-16-044

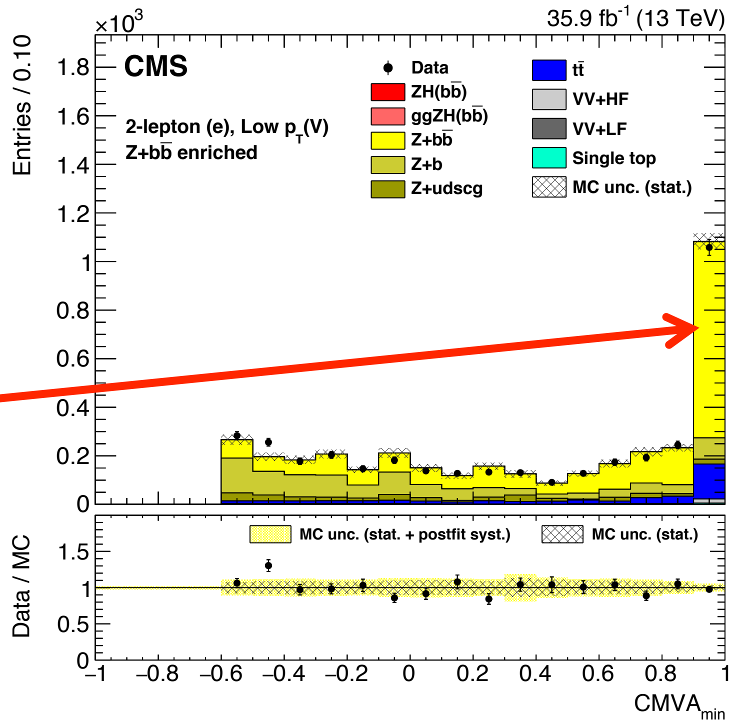
Control Regions – Z+b(b)

- **Control regions** (CR) are are used to help **scale the yields of key background process** in the final fit
- In the 2-lepton(e) category a Z+b(b) CR can be defined by inverting the m_{jj} window SR cut
- Link normalization between SR and CR Z+b(b) processes in final fit

2-lepton(e) SR

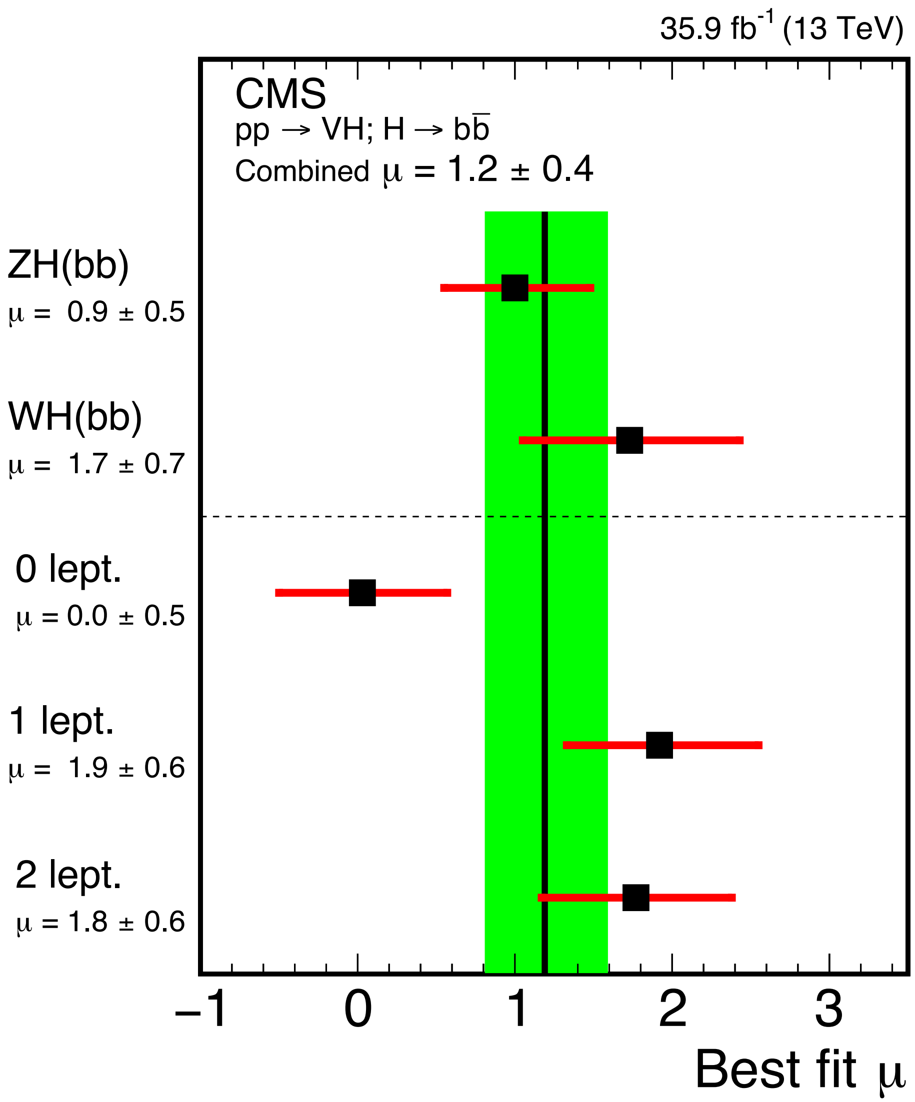


2-lepton(e) Z+b(b) CR



CMS-HIG-16-044

VH(bb) Results



- 3.3σ obs. Run-II only
- **3.8σ obs. evidence for H(bb), Run-I + Run-II**

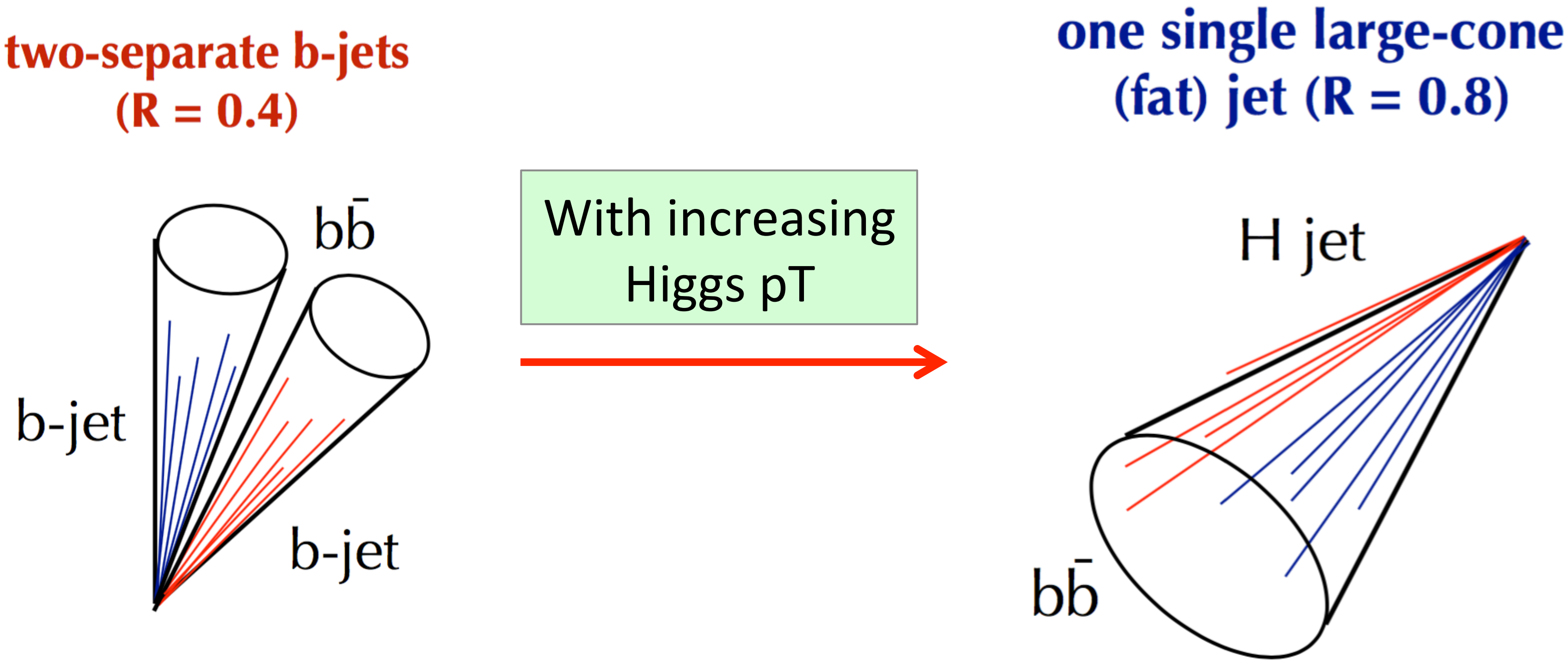
$$\mu_{CMS}^{run2} = 1.19^{+0.21}_{-0.20}(stat)^{+0.34}_{-0.32}(syst)$$

Boosted H(bb)

- arXiv:1709.05543 – submitted to PRL
- CMS-PAS-HIG-17-010
- First inclusive search for H(bb) where the Higgs boson is produced with high-pT

Boosted H(bb): Fundamentals

- Search for a heavily boosted gluon fusion produced Higgs(bb) with decay products contained in **a single wide, Anti-k_T "fat jet"**
- Trigger on total transverse energy or jet pT
  - Triggers 100% efficient for signal above 450 GeV jet pT and |eta|<2.5
- Veto event on presence of electrons, muon, and taus
- Veto on large missing ET (> 140 GeV)
- Categorize in 6 jet pT bins from 450 GeV – 1 TeV



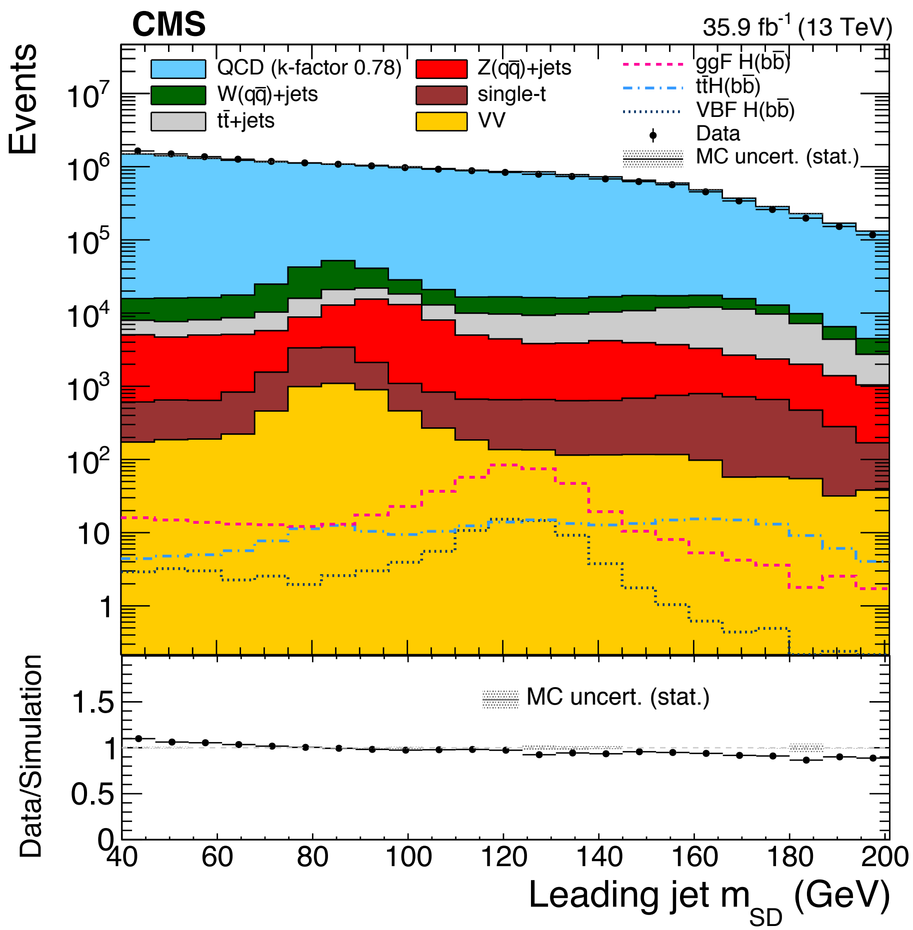
Background Composition

Mass distribution for the leading signal jet, after the online selection and the jet pT> 450 GeV requirement.

QCD: Fully derived from data, more detail following

W+Jets: LO MC based estimation, corrected with pT dependent NLO QCD k-factor & NLO Electroweak k-factor

VV: MC-based



ttbar: MC-based, normalized from dedicated control region

Z+Jets: LO MC based estimation, corrected with pT dependent NLO QCD k-factor & NLO Electroweak k-factor

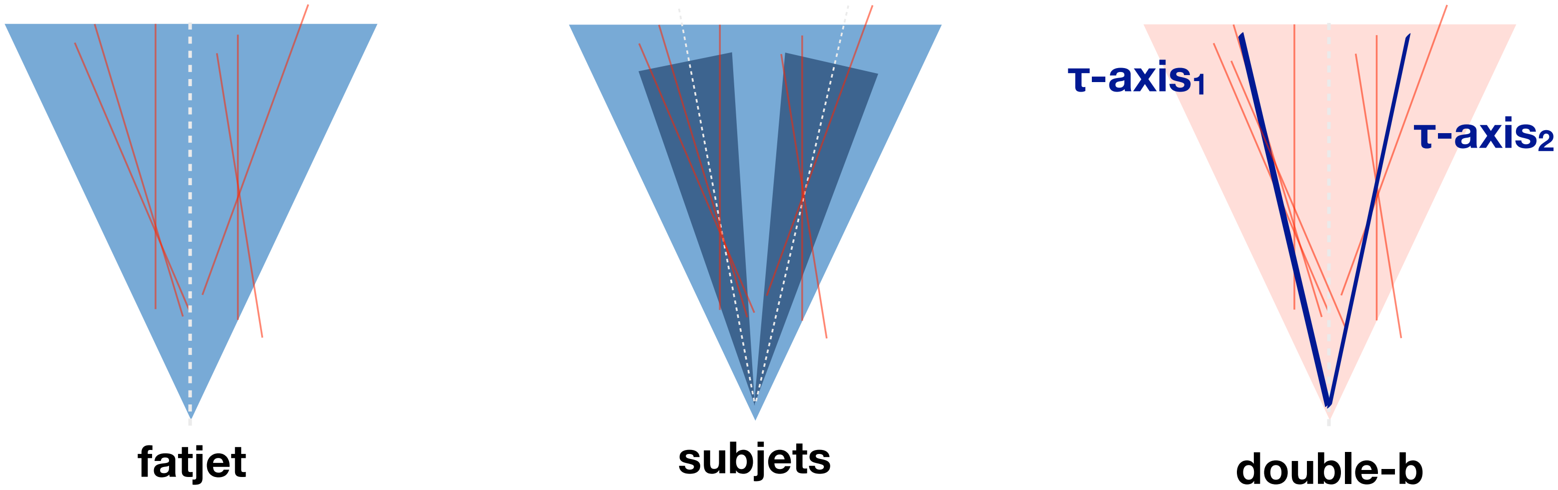
Single-top: MC-based

ggH Signal: NLO Powheg, normalized to N^{3}LO cross section + pT dependent higher order corrections incorporating finite top mass loop (1.27±0.38 correction)

Jet Requirements

Begin with a single "Fat Jet"

Jet required to have **2 prong substructure**: Energy correlation functions → N^1_2 provides additional splitting between 2-prong signal jets and QCD jets



Double-b tagger: IDs two B hadron decays from b and bbar within the same fat jet

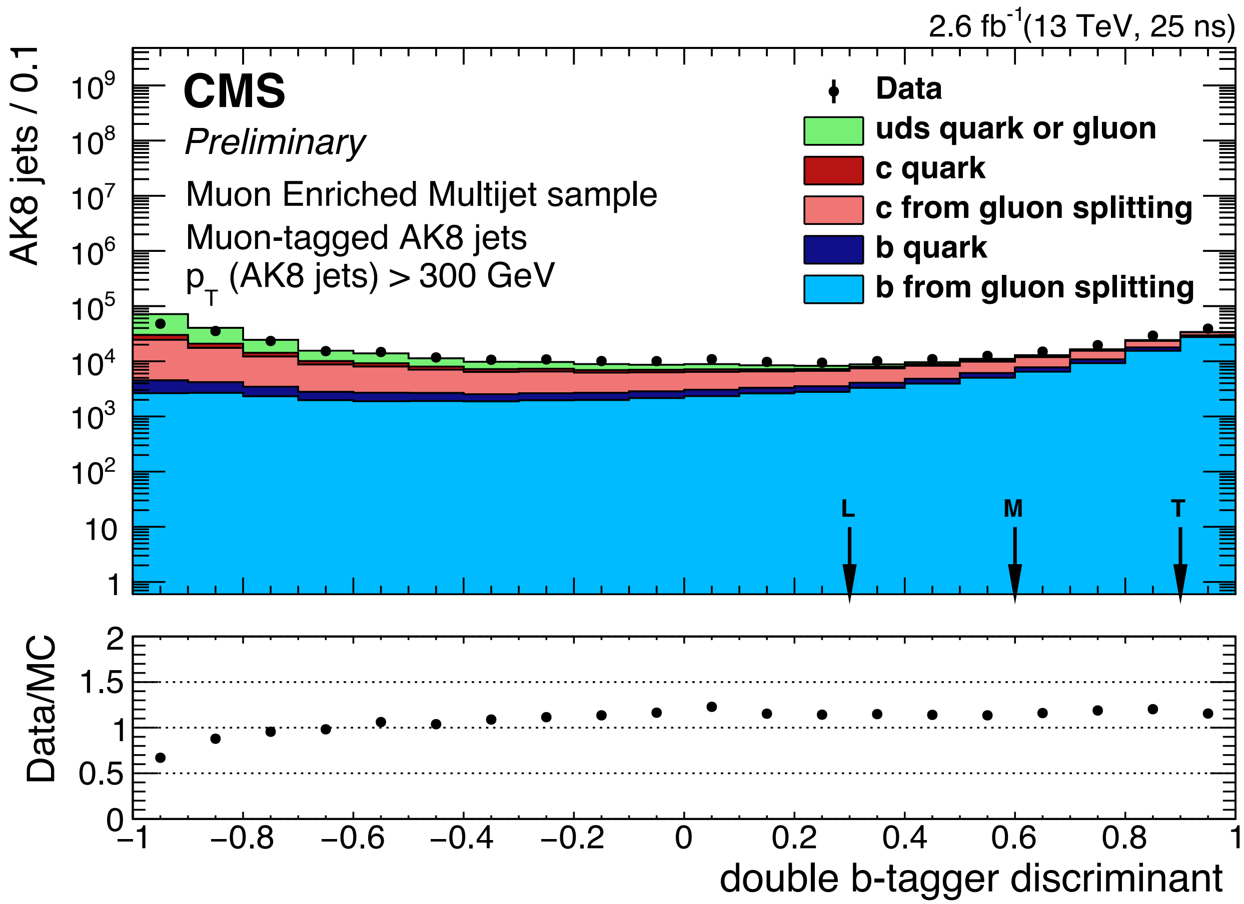
Jet Cleaning grooms away soft and wide angle jet constituents dominant in QCD jets → "Soft-Drop" Mass

*T-axes are Nsubjettiness axis

CMS-PAS-BTV-16-002

Double-b Tagger MVA

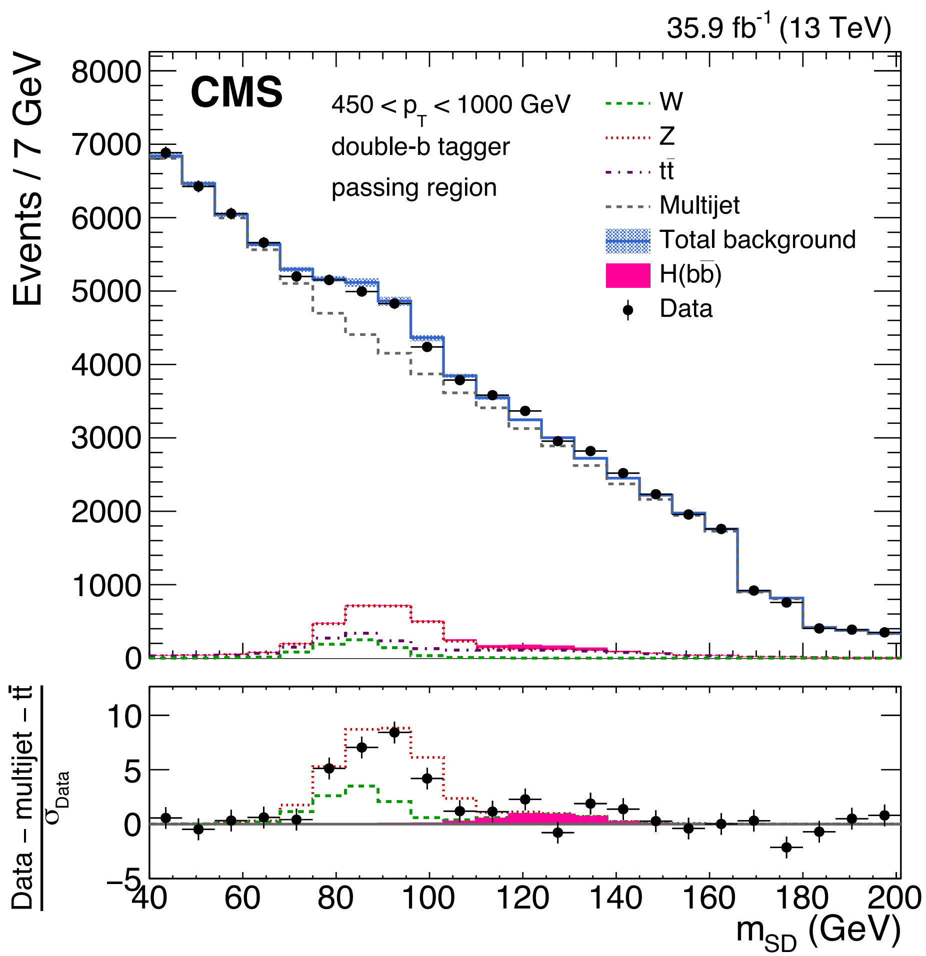
- Tracking and vertex information; 27 observables in total
- A double-b tagger working point is chosen giving **33% efficiency** for H(bb) and a **1% misidentification** rate for QCD jets
- Aimed to be mass and pT independent



CMS-PAS-BTV-16-002

Signal Region Composition

- Signal region, composed of events passing double-b tagger selection, dominated by QCD
- Resonant backgrounds from W/Z+jets
- Z(bb) provides handle to constrain nuisances for H(bb) signal & validate methodology



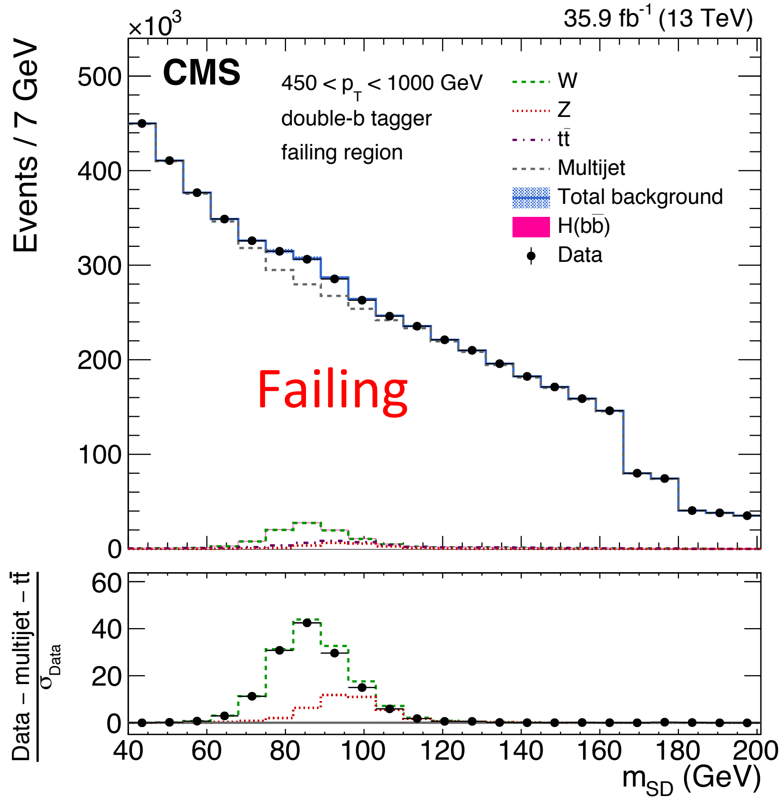
Background Composition	Percent of Bkg.
QCD	90%
ttbar+jets	3%
W/Z+jets	5%
Single-top, VV	<1%

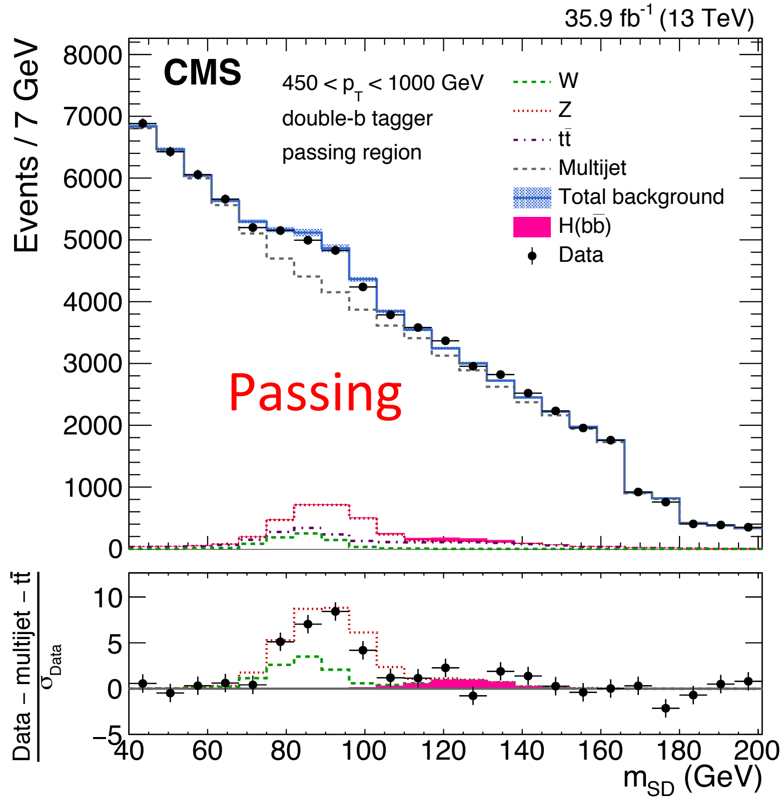
Higgs signal in SR is dominated by gluon fusion

Higgs Composition	Percentage of Signal
ggH	87%
VBF	6%
VH	4%
ttH	2%

Passing and Failing Double-b Tag

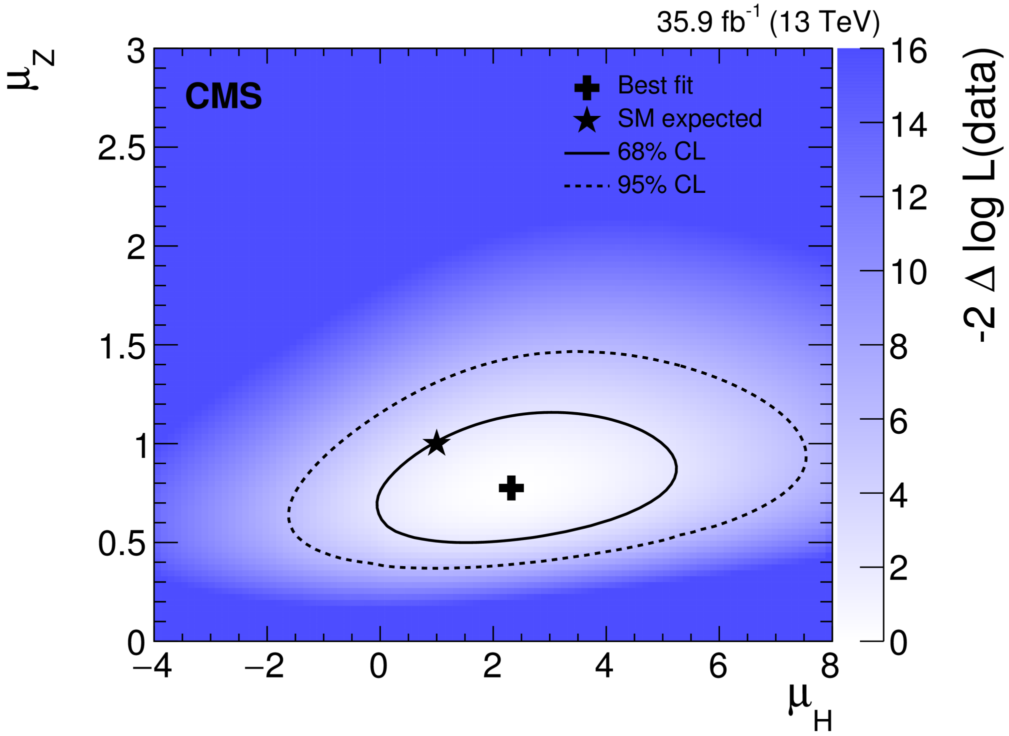
- QCD estimated from data in this failing, **inverting b tag region**
  - Ratio passing region / failing region approx. constant as function of jet pT
  - Corrections applied as function of ρ and jet pT
  - $\rho = \log(m^2_{SD} / pT^2)$, describes correlation between b tagging discriminator, jet mass, and jet pT





Ratio of data to its statistical uncertainty, after subtracting the non-resonant backgrounds (QCD+ttbar)

Boosted H(bb) Results



Likelihood scan of Z Boson signal strength and Higgs Boson signal strength

- First **observation of Z(bb)** in single jet topology (5.1σ observed, 5.8σ expected)
- 1.5σ observed significance of SM-Higgs
- Both Z and Higgs signal strengths agree with SM
- Analysis opens new door for H(bb) physics

	H	H no p_T corr.	Z
Observed signal strength	$2.3^{+1.8}_{-1.6}$	$3.2^{+2.2}_{-2.0}$	$0.78^{+0.23}_{-0.19}$
Expected significance	0.7σ	0.5σ	5.8σ
Observed significance	1.5σ	1.6σ	5.1σ

$H(\tau\tau)$

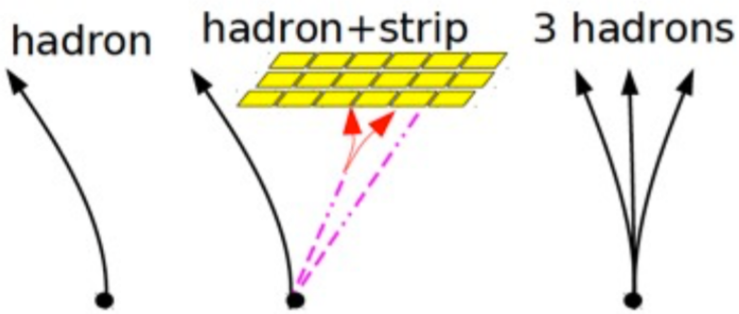
- Htt: arXiv:1708.00373 – submitted to PLB
- CMS-HIG-16-043
- Analysis Overview:
  - Four different decay channels depending on tau decays:
    - **$e\tau_h$, $\mu\tau_h$, $\tau_h\tau_h$, $e\mu$**
  - Three different signal region categories:
    - 0-Jet, VBF, Boosted
- Main changes from Run-I analysis:
  - Tau ID moved from cut-based to MVA-based
  - Use unrolled 2D distributions for signal extraction
  - Simultaneous fit of signal regions and controls regions

Tau Decays and Reconstruction

- Tau leptons decay either leptonically (35%) or hadronically (65%)
- Hadronic decays are reconstructed from charged and neutral jet constituents in CMS's Particle Flow algorithm
  - 1-prong, 1-prong + π^0, 3-prong
- Hadrons Plus Strips algorithm:
  - Can ID each τ_h decay mode
  - Exploits intermediate resonances $\rho(770)$ and $a_1(1260)$

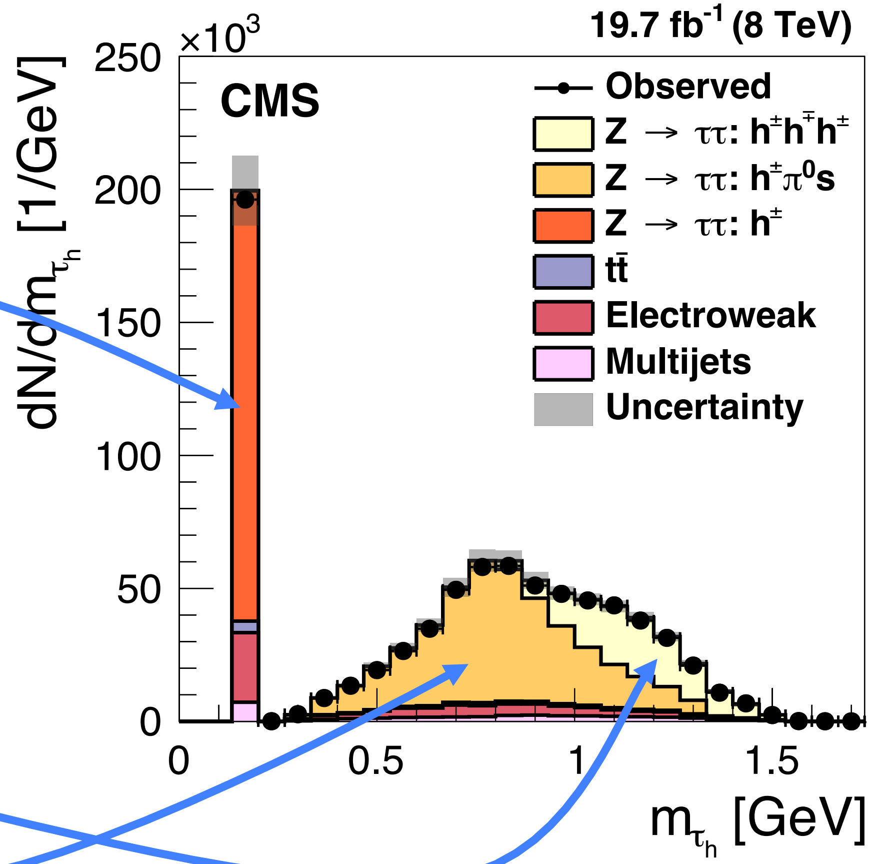
Decay mode	Meson resonance	$\mathcal{B}$ [%]
$\tau^- \to e^- \overline{\nu}_e \nu_\tau$		17.8
$\tau^- \to \mu^- \overline{\nu}_\mu \nu_\tau$		17.4
$\tau^- \to h^- \nu_\tau$		11.5
$\tau^- \to h^- \pi^0 \nu_\tau$	$\rho(770)$	26.0
$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	9.5
$\tau^- \to h^- h^+ h^- \nu_\tau$	$a_1(1260)$	9.8
$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$		4.8
Other modes with hadrons		3.2
All modes containing hadrons		64.8

1-prong



1-prong + π^0s

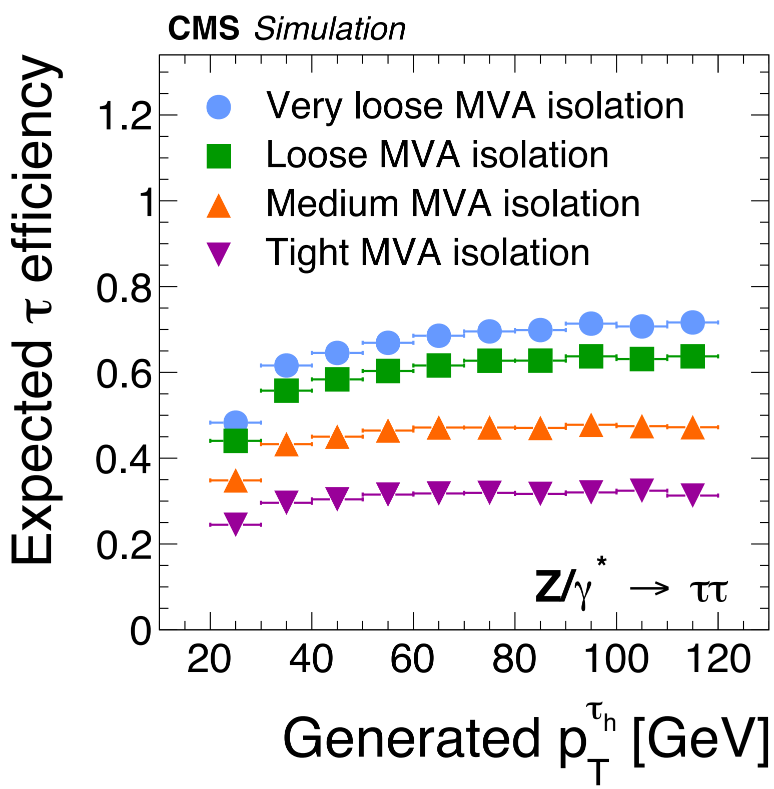
3-prong

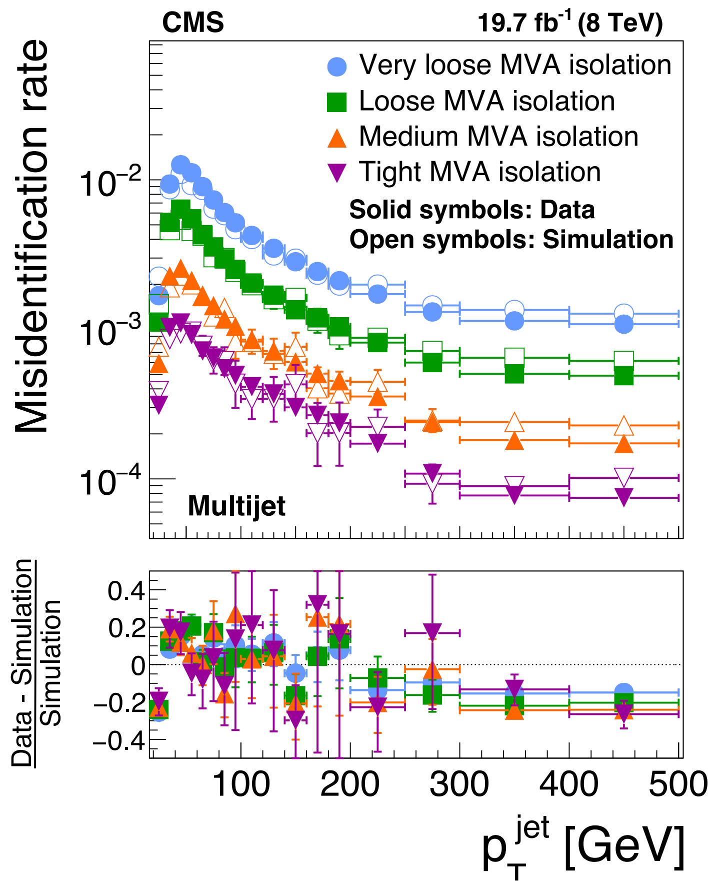


CMS-TAU-14-001

Tau Identification

- We need best tau identification with strong jet rejection
- H(ττ) uses an MVA-based ID that has large performance gains compared to Run 1 H(ττ) cut-based ID
  - Inputs: Isolation (pT sums), τ lifetime, distribution of signal and isolation candidates within the jet in ΔR, Δη, Δφ, and e/γ multiplicies

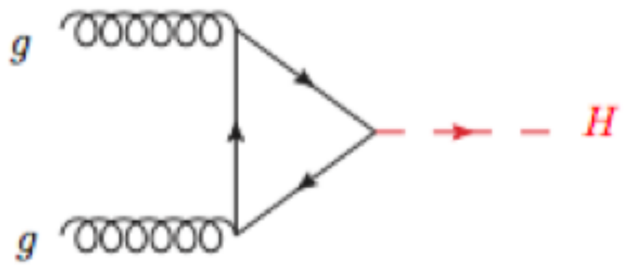


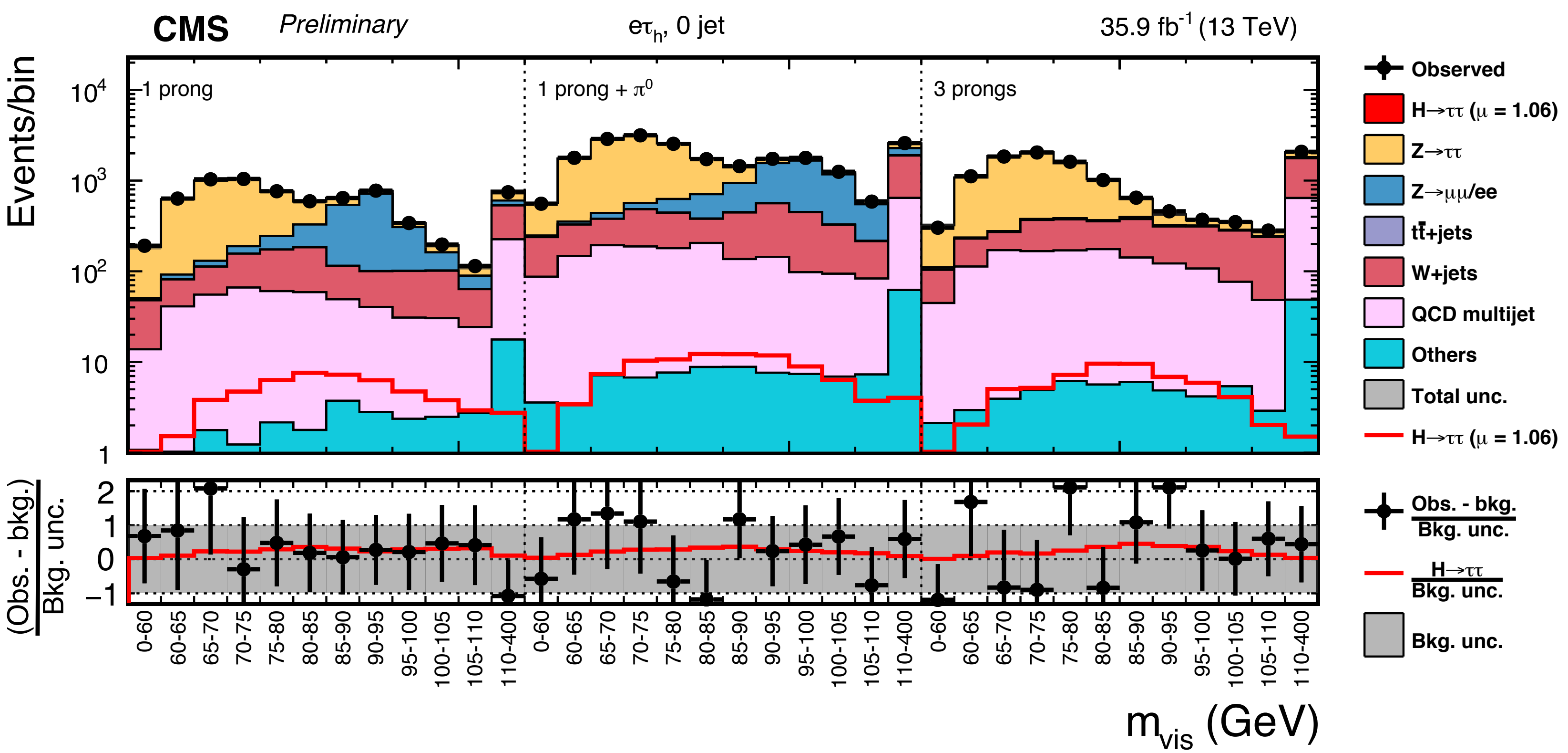


CMS-TAU-14-001

Signal Extraction Categories

	0-jet	VBF	Boosted
		Selection	
$\tau_h\tau_h$	No jet	$\geq$2 jets, $p_T^{\tau\tau} > 100\,\mathrm{GeV}$, $\Delta\eta_{jj} > 2.5$	Others
$\mu\tau_h$	No jet	$\geq$2 jets, $m_{jj} > 300\,\mathrm{GeV}$, $p_T^{\tau\tau} > 50\,\mathrm{GeV}$, $p_T^{\tau_h} > 40\,\mathrm{GeV}$	Others
$e\tau_h$	No jet	$\geq$2 jets, $m_{jj} > 300\,\mathrm{GeV}$, $p_T^{\tau\tau} > 50\,\mathrm{GeV}$	Others
$e\mu$	No jet	2 jets, $m_{jj} > 300\,\mathrm{GeV}$	Others
		Observables	
$\tau_h\tau_h$	$m_{\tau\tau}$	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$
$\mu\tau_h$	τ_h decay mode, m_{vis}	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$
$e\tau_h$	τ_h decay mode, m_{vis}	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$
$e\mu$	p_T^{μ}, m_{vis}	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$





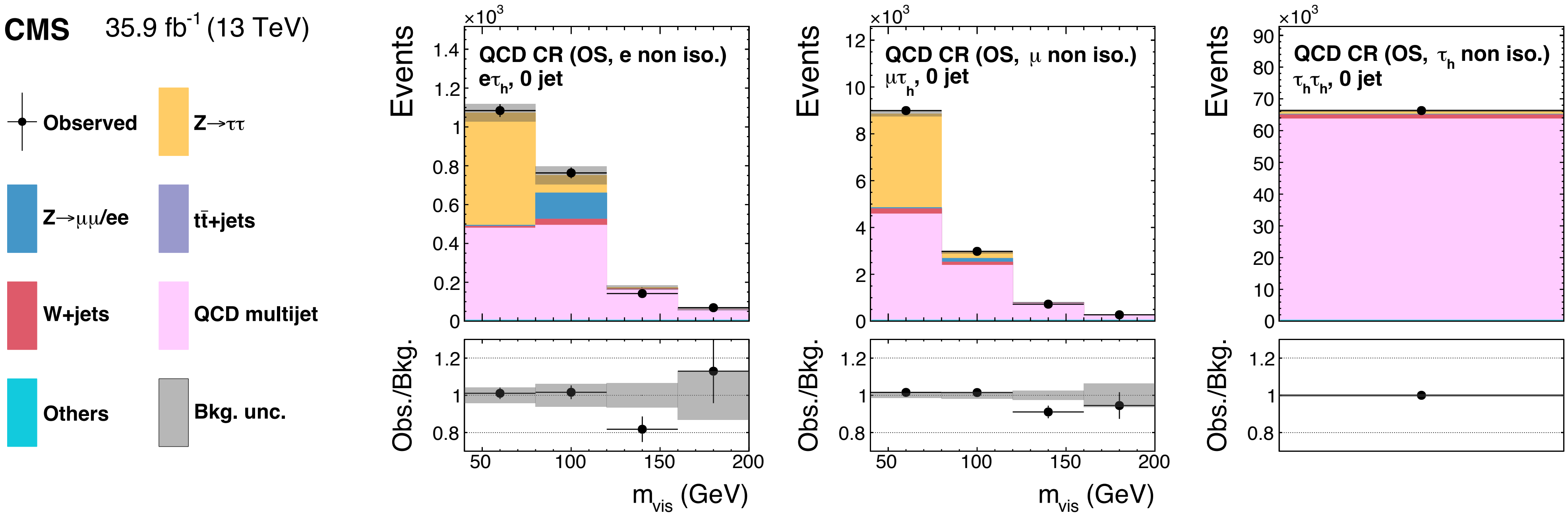
Main Backgrounds & Control Regions

$Z \rightarrow ee/\mu\mu/\tau\tau$:

- LO MC sample
- Corrections p_T^Z, $mass_Z$, & m_{jj} from $Z \rightarrow \mu\mu$ CR

QCD multijet:

- Fully data driven
- $e\tau_h/\mu\tau_h$ – taken from region with same-charge taus
- $\tau_h\tau_h$ – taken from region with opposite charge, anti-isolated taus
- CR included in simultaneous fit



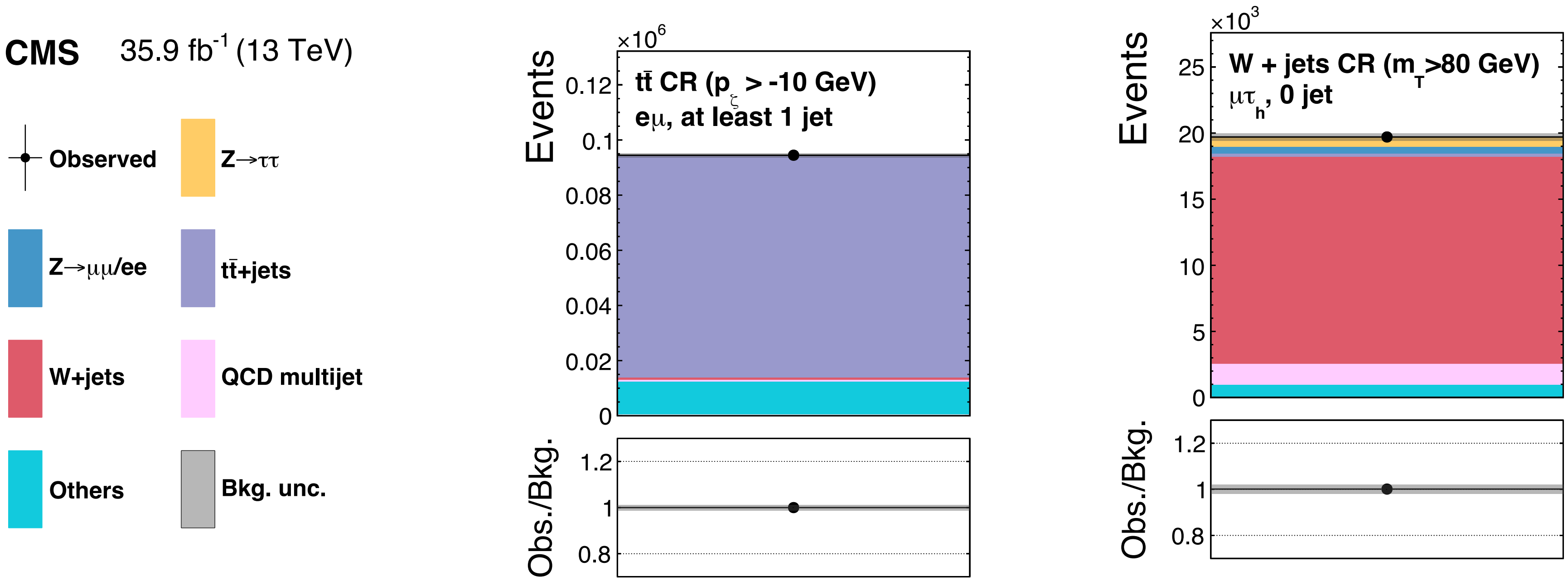
Main Backgrounds

ttbar:

- MC sample
- Single CR in eμ ttbar enriched CR

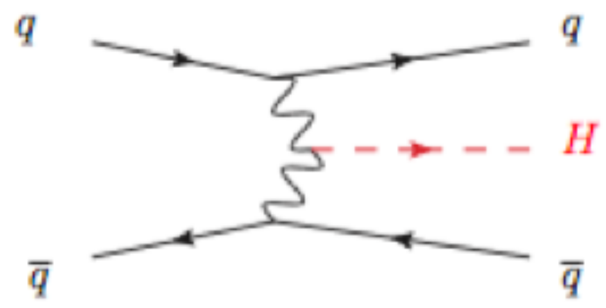
W+Jets:

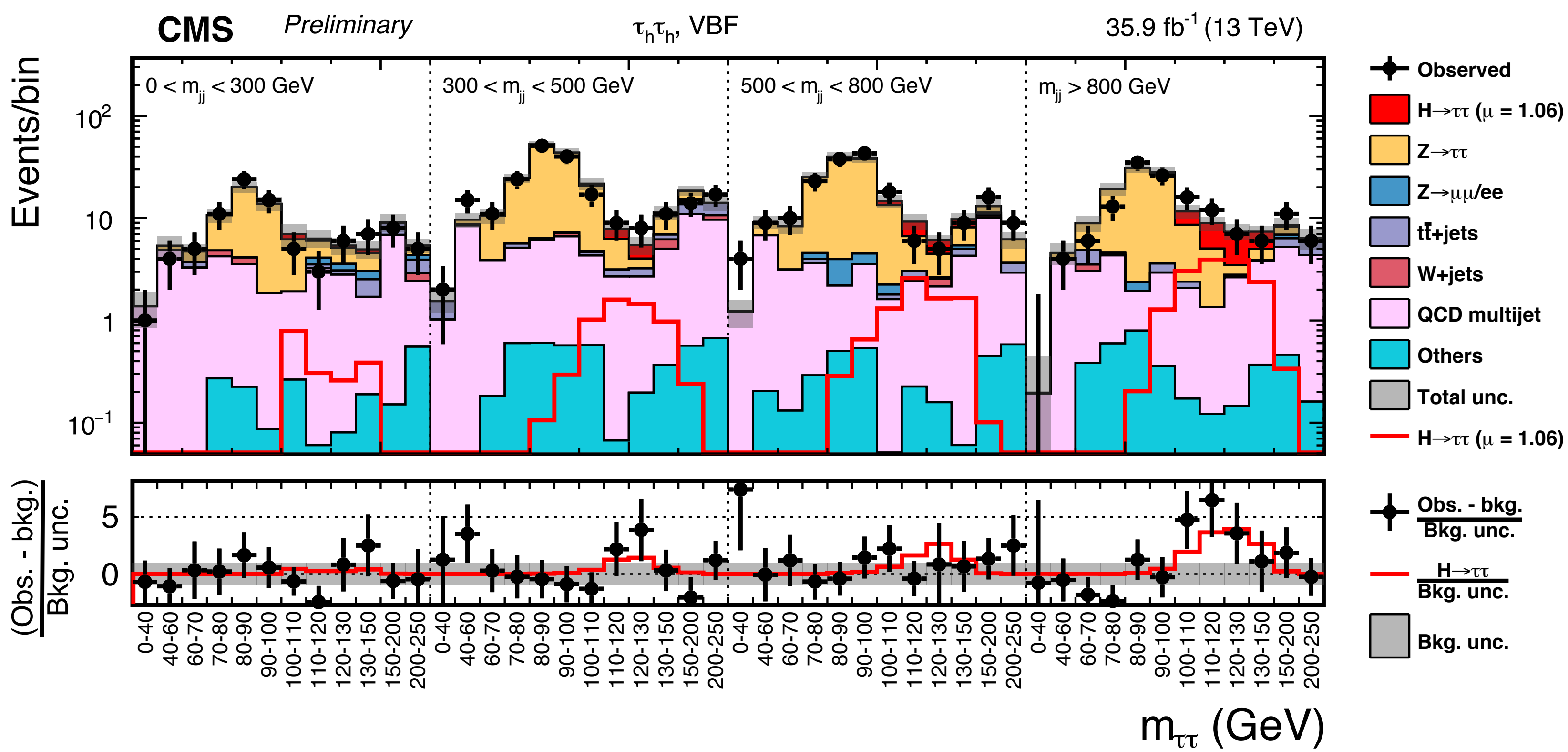
- Shape from LO MC
- $e\tau_h/\mu\tau_h$ – Normalized with data in high m_T CR [mT(lep,MET)>80 GeV]
- $\tau_h\tau_h$/eμ – from MC with corrections



Signal Extraction Categories

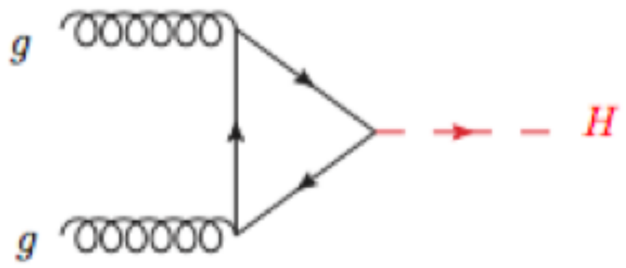
	0-jet	VBF	Boosted
		Selection	
$\tau_h\tau_h$	No jet	$\geq$2 jets, $p_T^{\tau\tau} > 100\,\text{GeV}$, $\Delta\eta_{jj} > 2.5$	Others
$\mu\tau_h$	No jet	$\geq$2 jets, $m_{jj} > 300\,\text{GeV}$, $p_T^{\tau\tau} > 50\,\text{GeV}$, $p_T^{\tau_h} > 40\,\text{GeV}$	Others
$e\tau_h$	No jet	$\geq$2 jets, $m_{jj} > 300\,\text{GeV}$, $p_T^{\tau\tau} > 50\,\text{GeV}$	Others
$e\mu$	No jet	2 jets, $m_{jj} > 300\,\text{GeV}$	Others
		Observables	
$\tau_h\tau_h$	$m_{\tau\tau}$	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$
$\mu\tau_h$	τ_h decay mode, m_{vis}	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$
$e\tau_h$	τ_h decay mode, m_{vis}	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$
$e\mu$	p_T^{μ}, m_{vis}	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$

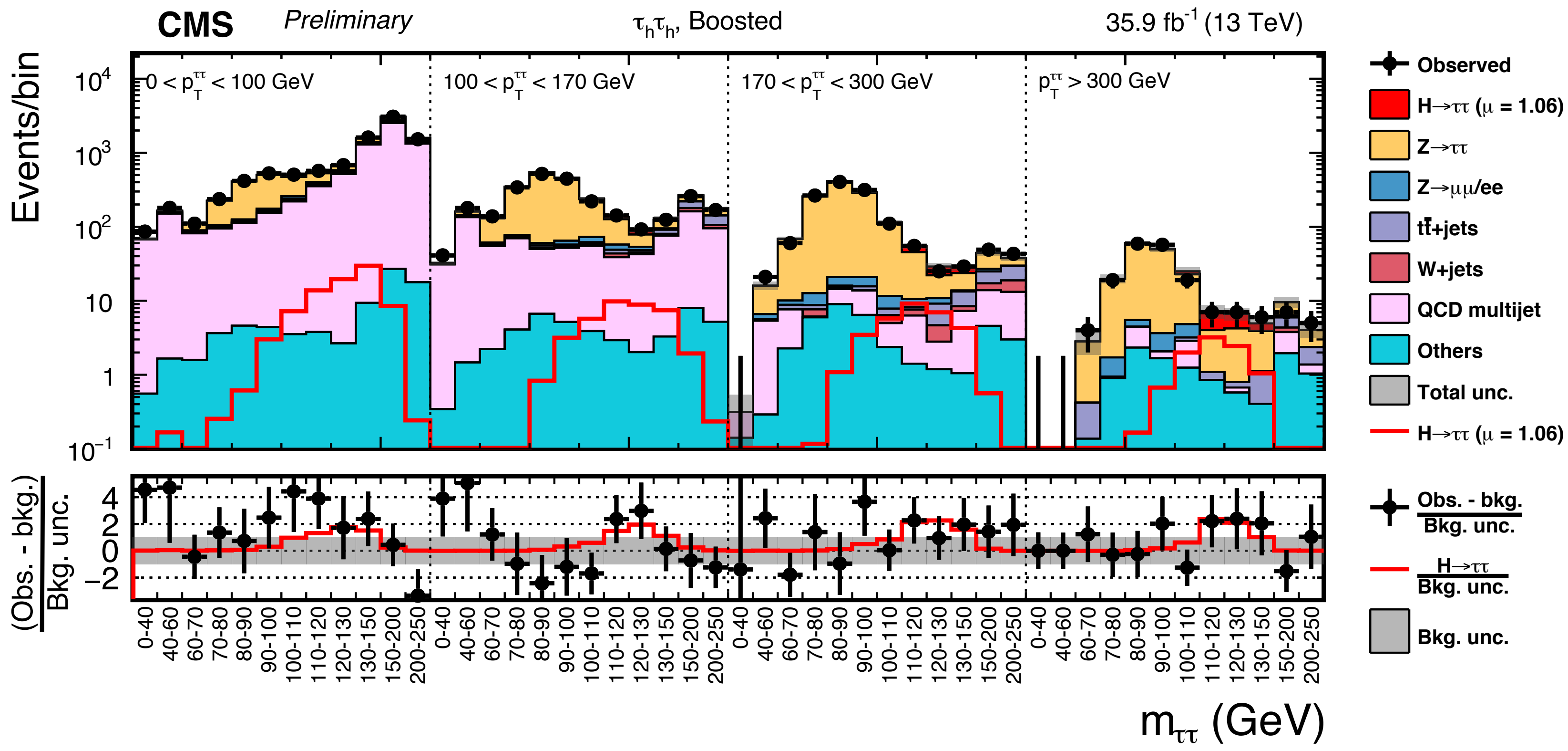




Signal Extraction Categories

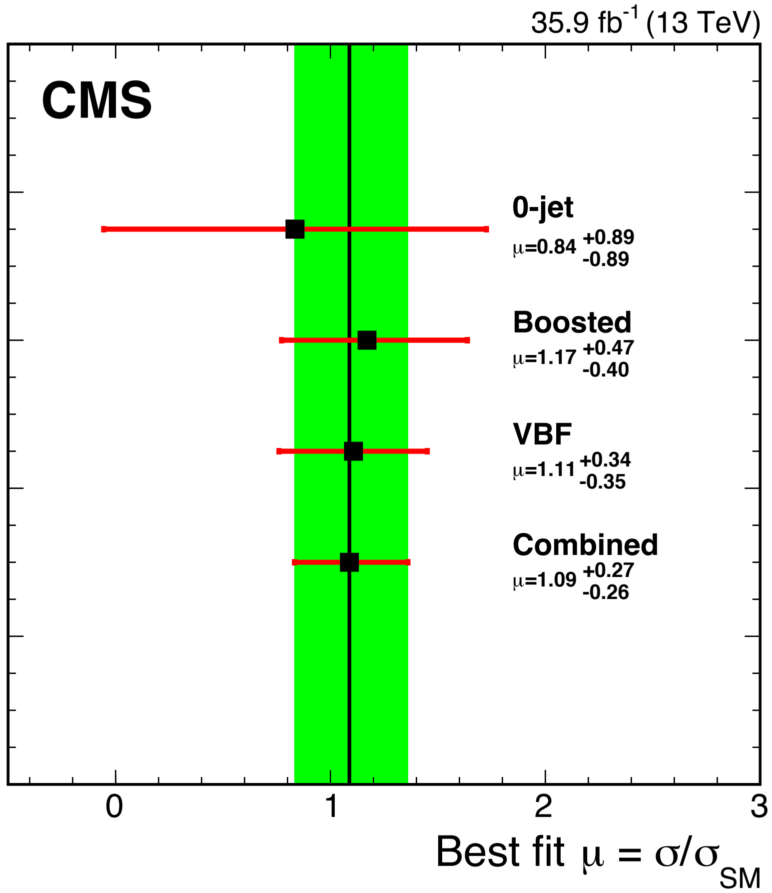
	0-jet	VBF	Boosted
	Selection		
$\tau_h\tau_h$	No jet	$\geq$2 jets, $p_T^{\tau\tau} > 100\,\text{GeV}$, $\Delta\eta_{jj} > 2.5$	Others
$\mu\tau_h$	No jet	$\geq$2 jets, $m_{jj} > 300\,\text{GeV}$, $p_T^{\tau\tau} > 50\,\text{GeV}$, $p_T^{\tau_h} > 40\,\text{GeV}$	Others
$e\tau_h$	No jet	$\geq$2 jets, $m_{jj} > 300\,\text{GeV}$, $p_T^{\tau\tau} > 50\,\text{GeV}$	Others
$e\mu$	No jet	2 jets, $m_{jj} > 300\,\text{GeV}$	Others
	Observables		
$\tau_h\tau_h$	$m_{\tau\tau}$	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$
$\mu\tau_h$	τ_h decay mode, m_{vis}	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$
$e\tau_h$	τ_h decay mode, m_{vis}	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$
$e\mu$	p_T^{μ}, m_{vis}	m_{jj}, $m_{\tau\tau}$	$p_T^{\tau\tau}$, $m_{\tau\tau}$

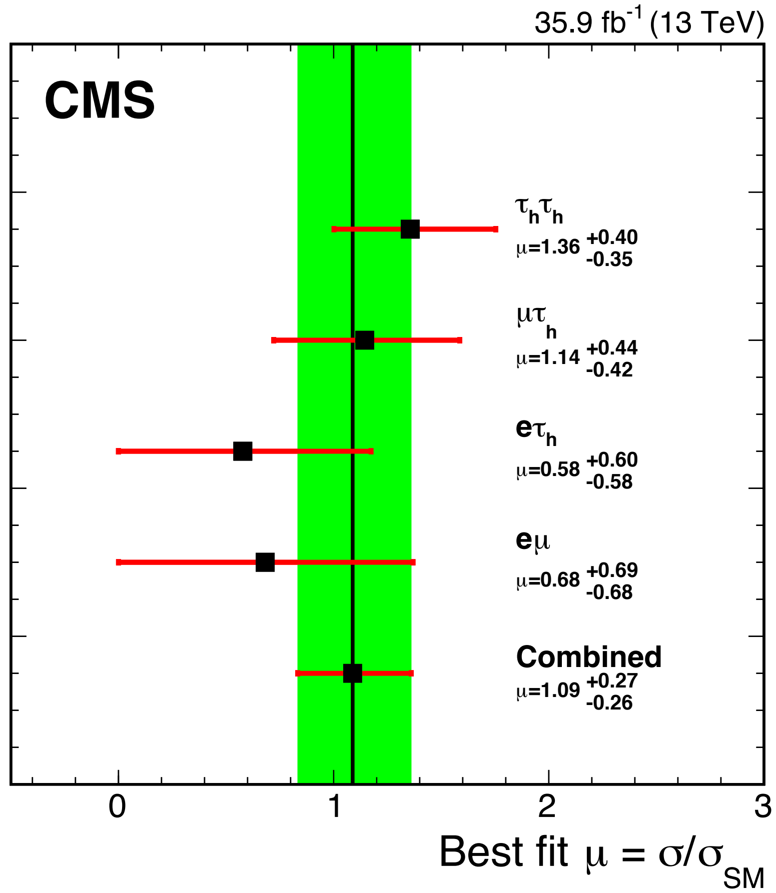




Results

- Best fit signal strengths all consistent with Standard Model
- Significance:
  - Run-II Obs. **4.9σ**
  - Run-II Expt. 4.7σ
  - CMS Run-I + Run-II Obs. **5.9σ**





Conclusion

- Using 2016, 13 TeV data gathered by CMS, analysts have been pushing Higgs decaying to fermion searches and measurements into new territory
- H(bb) is the first analysis searching inclusively for H(bb) and is the first observation of Z(bb) in the single jet topology (**5.1σ**)
- ZH(bb) analysis showed us evidence for ZH(bb) decays at CMS (**3.3σ**)
- H(ττ) analysis has provided CMS with the first single-experiment observation of H(ττ) decays when combined with Run-I (**5.9σ**)

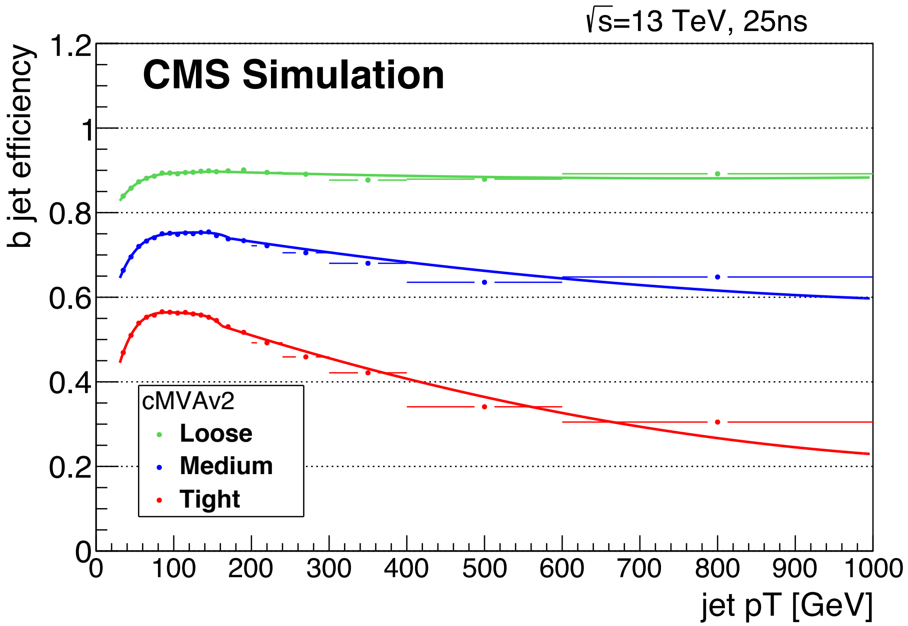
- With 2017 data taking complete we have more than doubled our 13 TeV data set

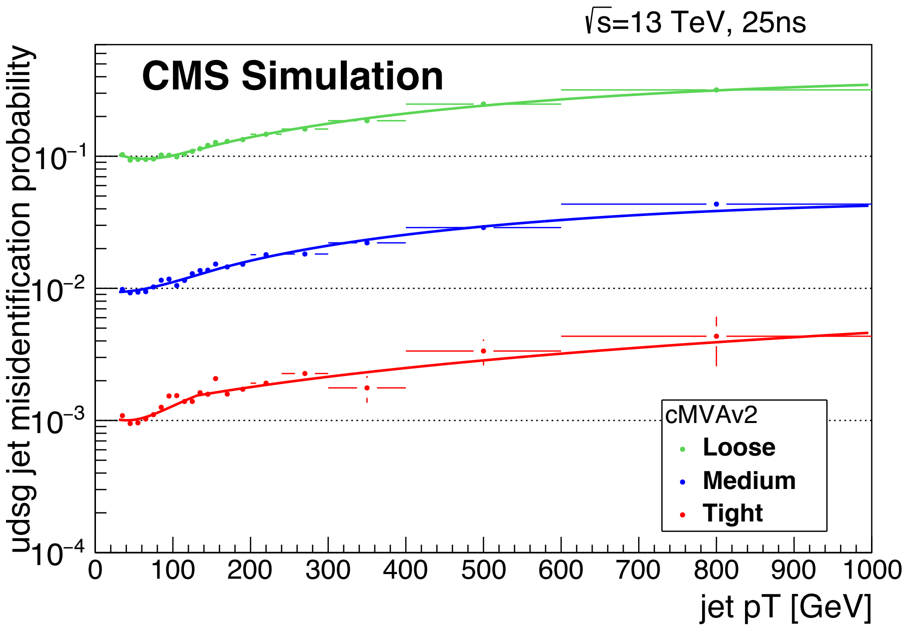
- Many thanks to the CMS collaboration and all the hard work from everyone

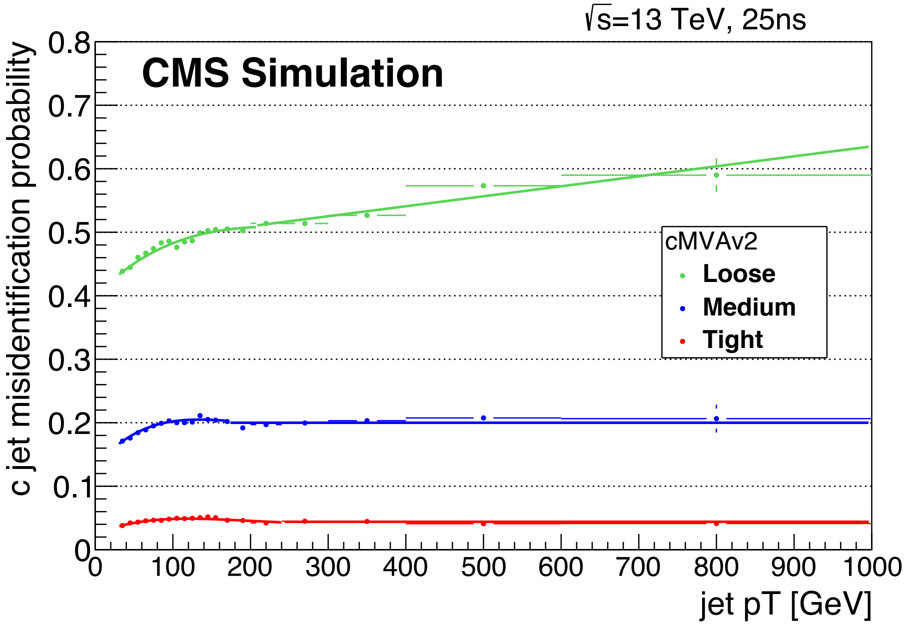
- More exciting results coming soon!

Backup

VH(bb) b-tagging Efficiencies and MisID







CMS-PAS-BTV-15-001

VH(bb) BDT Training

- Variables used in the training of the event BDT discriminant for the different channels. Jets are counted as additional jets to those selected to reconstruct the H(bb) decay if they satisfy the following: pT> 30 GeV and |η|< 2.4 for the 0- and 2-lepton channels, and pT> 25 GeV and |η|< 2.9 for the 1-lepton channel.

Variable	Description	Channels
$M(\mathrm{jj})$	dijet invariant mass	All
$p_{\mathrm{T}}(\mathrm{jj})$	dijet transverse momentum	All
$p_{\mathrm{T}}(\mathrm{j}_1)$, $p_{\mathrm{T}}(\mathrm{j}_2)$	transverse momentum of each jet	0- and 2-lepton
$\Delta R(\mathrm{jj})$	distance in η–ϕ between jets	2-lepton
$\Delta\eta(\mathrm{jj})$	difference in η between jets	0- and 2-lepton
$\Delta\phi(\mathrm{jj})$	azimuthal angle between jets	0-lepton
$p_{\mathrm{T}}(\mathrm{V})$	vector boson transverse momentum	All
$\Delta\phi(\mathrm{V},\mathrm{jj})$	azimuthal angle between vector boson and dijet directions	All
$p_{\mathrm{T}}(\mathrm{jj})/p_{\mathrm{T}}(\mathrm{V})$	p_{T} ratio between dijet and vector boson	2-lepton
$M(\ell\ell)$	reconstructed Z boson mass	2-lepton
$\mathrm{CMVA}_{\mathrm{max}}$	value of CMVA discriminant for the jet with highest CMVA value	0- and 2-lepton
$\mathrm{CMVA}_{\mathrm{min}}$	value of CMVA discriminant for the jet with second highest CMVA value	All
$\mathrm{CMVA}_{\mathrm{add}}$	value of CMVA for the additional jet with highest CMVA value	0-lepton
$p_{\mathrm{T}}^{\mathrm{miss}}$	missing transverse momentum	1- and 2-lepton
$\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}},\mathrm{j})$	azimuthal angle between $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and closest jet ($p_{\mathrm{T}} > 30\,\mathrm{GeV}$)	0-lepton
$\Delta\phi(\vec{p}_{\mathrm{T}}^{\mathrm{miss}},\ell)$	azimuthal angle between $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ and lepton	1-lepton
m_{T}	mass of lepton $\vec{p}_{\mathrm{T}} + \vec{p}_{\mathrm{T}}^{\mathrm{miss}}$	1-lepton
m_{top}	reconstructed top quark mass	1-lepton
N_{aj}	number of additional jets	1- and 2-lepton
$p_{\mathrm{T}}(\mathrm{add})$	transverse momentum of leading additional jet	0-lepton
SA5	number of soft-track jets with $p_{\mathrm{T}} > 5\,\mathrm{GeV}$	All

VH(bb) Control Region Scale Factors

- Data/MC scale factors for each of the main background processes in each channel, as obtained from the combined signal-extraction fit to control and signal region distributions. Electron and muon samples in the 1- and 2-lepton channels are fit simultaneously to determine average scale factors. The same scale factors for W+jets processes are used for the 0- and 1-lepton channels.

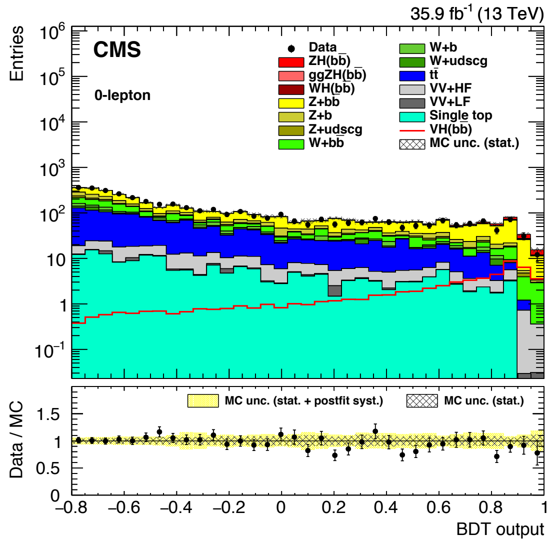
Process	0-lepton	1-lepton	2-lepton low-$p_T(V)$	2-lepton high-$p_T(V)$
W0b	1.14 ± 0.07	1.14 ± 0.07	—	—
W1b	1.66 ± 0.12	1.66 ± 0.12	—	—
W2b	1.49 ± 0.12	1.49 ± 0.12	—	—
Z0b	1.03 ± 0.07	—	1.01 ± 0.06	1.02 ± 0.06
Z1b	1.28 ± 0.17	—	0.98 ± 0.06	1.02 ± 0.11
Z2b	1.61 ± 0.10	—	1.09 ± 0.07	1.28 ± 0.09
$t\bar{t}$	0.78 ± 0.05	0.91 ± 0.03	1.00 ± 0.03	1.04 ± 0.05

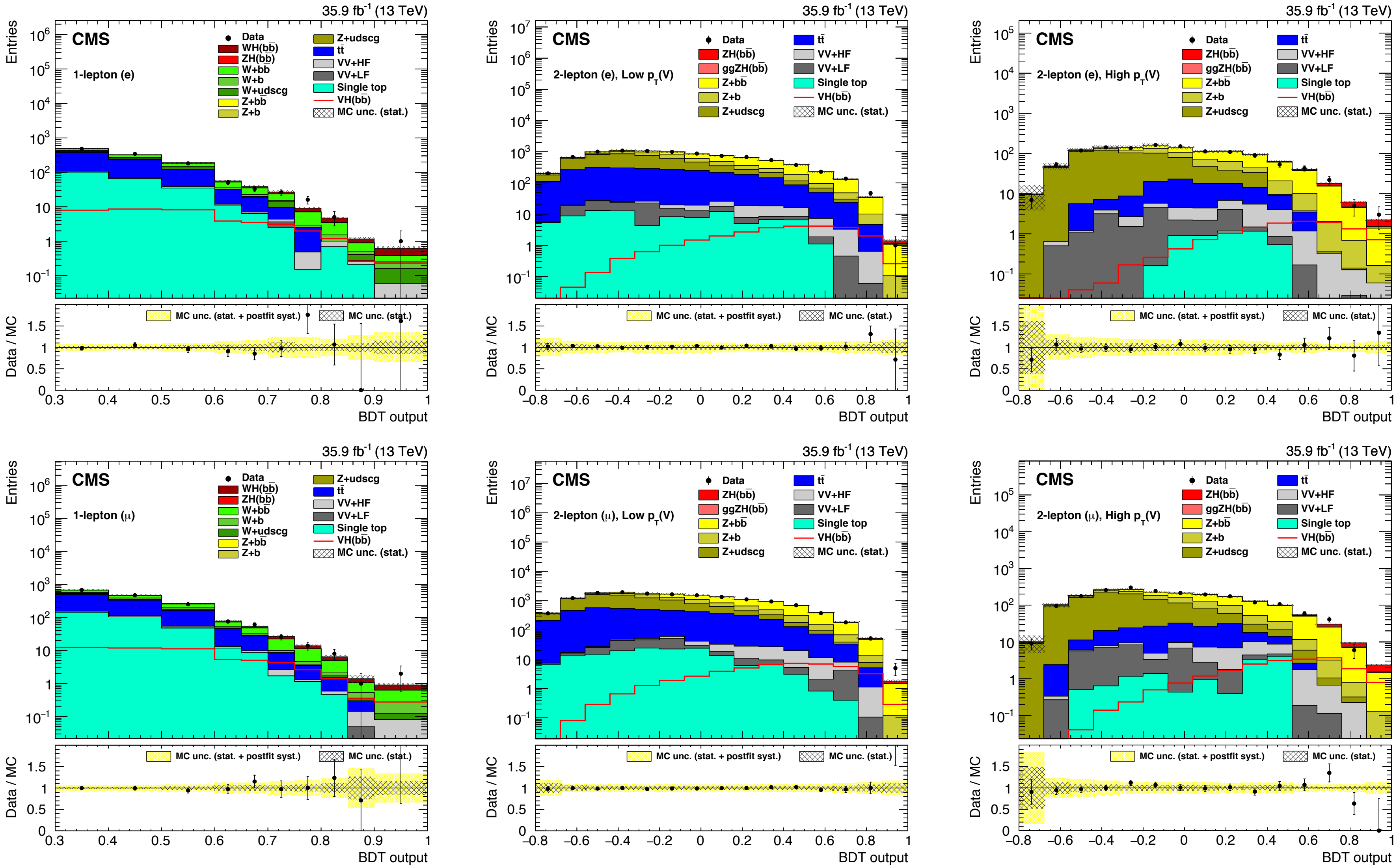
VH(bb) Signal Region Criteria

- Selection criteria that define the signal region

Variable	0-lepton	1-lepton	2-lepton
$p_{T}(V)$	>170	>100	$[50,150], >150$
$M(\ell\ell)$	—	—	$[75,105]$
p_{T}^{ℓ}	—	$(>25, >30)$	>20
$p_{T}(j_1)$	>60	>25	>20
$p_{T}(j_2)$	>35	>25	>20
$p_{T}(jj)$	>120	>100	—
$M(jj)$	$[60,160]$	$[90,150]$	$[90,150]$
$\Delta\phi(V,jj)$	>2.0	>2.5	>2.5
$CMVA_{max}$	$>CMVA_{T}$	$>CMVA_{T}$	$>CMVA_{L}$
$CMVA_{min}$	$>CMVA_{L}$	$>CMVA_{L}$	$>CMVA_{L}$
N_{aj}	<2	<2	—
$N_{a\ell}$	$=0$	$=0$	—
p_{T}^{miss}	>170	—	—
$\Delta\phi(\vec{p}_{T}^{miss}, j)$	>0.5	—	—
$\Delta\phi(\vec{p}_{T}^{miss}, \vec{p}_{T}^{miss}(trk))$	<0.5	—	—
$\Delta\phi(\vec{p}_{T}^{miss}, \ell)$	—	<2.0	—
Lepton isolation	—	<0.06	$(<0.25, <0.15)$
Event BDT	>-0.8	>0.3	>-0.8

VH(bb) All 7 Signal Regions





VH(bb) Systematics

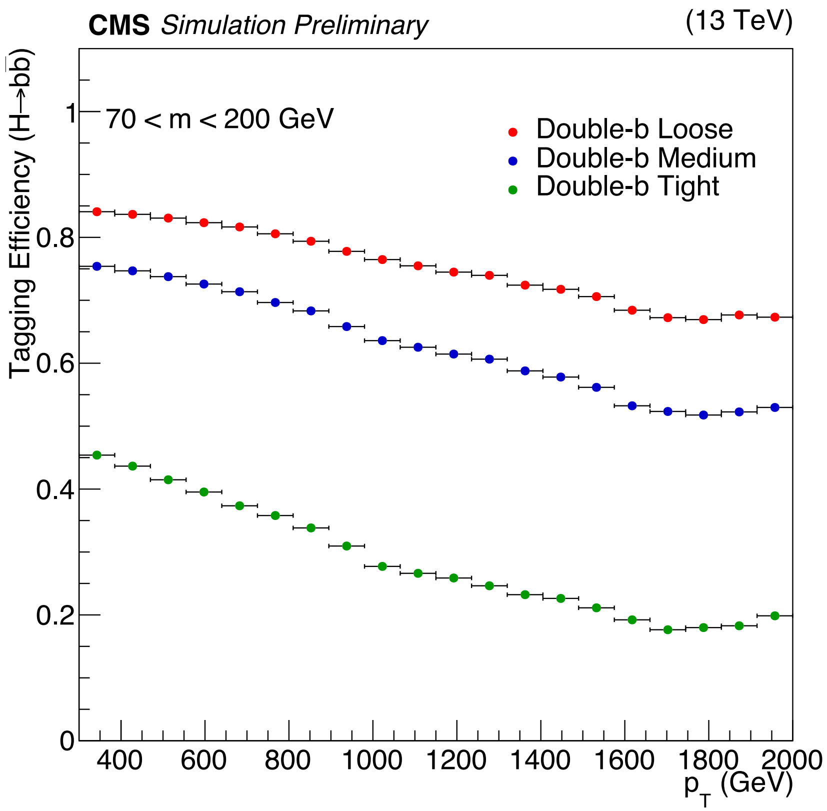
- Effect of each source of systematic uncertainty in the expected signal strength μ. The third column shows the uncertainty in μ from each source when only that particular source is considered.

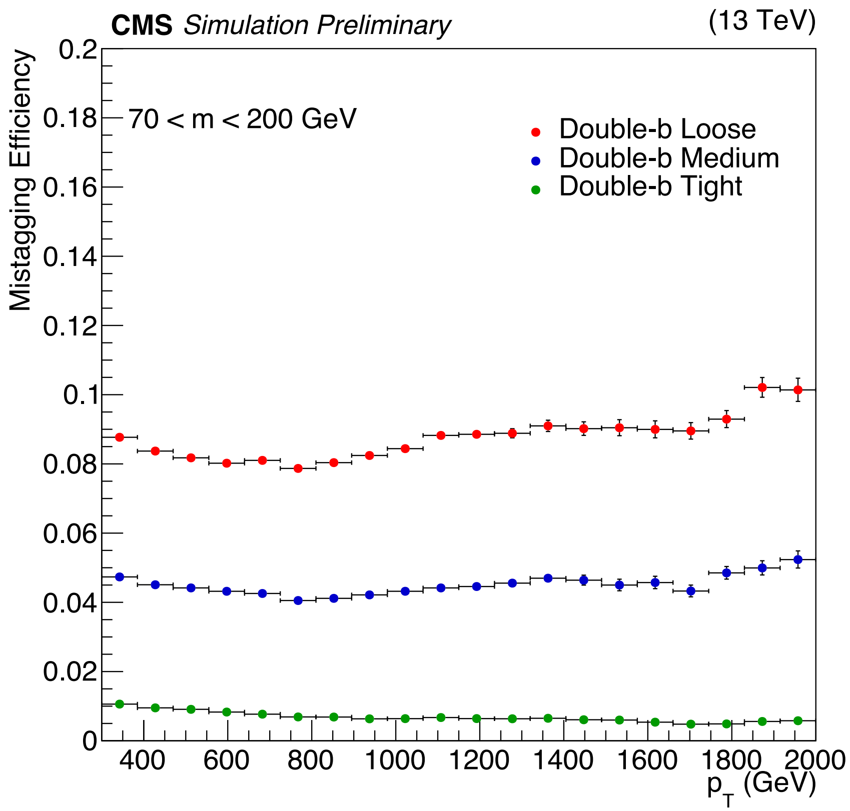
Experimental Uncertainties

Theory Uncertainties

Source	Type	Individual contribution to the μ uncertainty (%)
Scale factors ($t\bar{t}$, V+jets)	norm.	9.4
Size of simulated samples	shape	8.1
Simulated samples' modeling	shape	4.1
b tagging efficiency	shape	7.9
Jet energy scale	shape	4.2
Signal cross sections	norm.	5.3
Cross section uncertainties (single-top, VV)	norm.	4.7
Jet energy resolution	shape	5.6
b tagging mistag rate	shape	4.6
Integrated luminosity	norm.	2.2
Unclustered energy	shape	1.3
Lepton efficiency and trigger	norm.	1.9

Boosted H(bb), Double-b Tagging Eff and Mistag

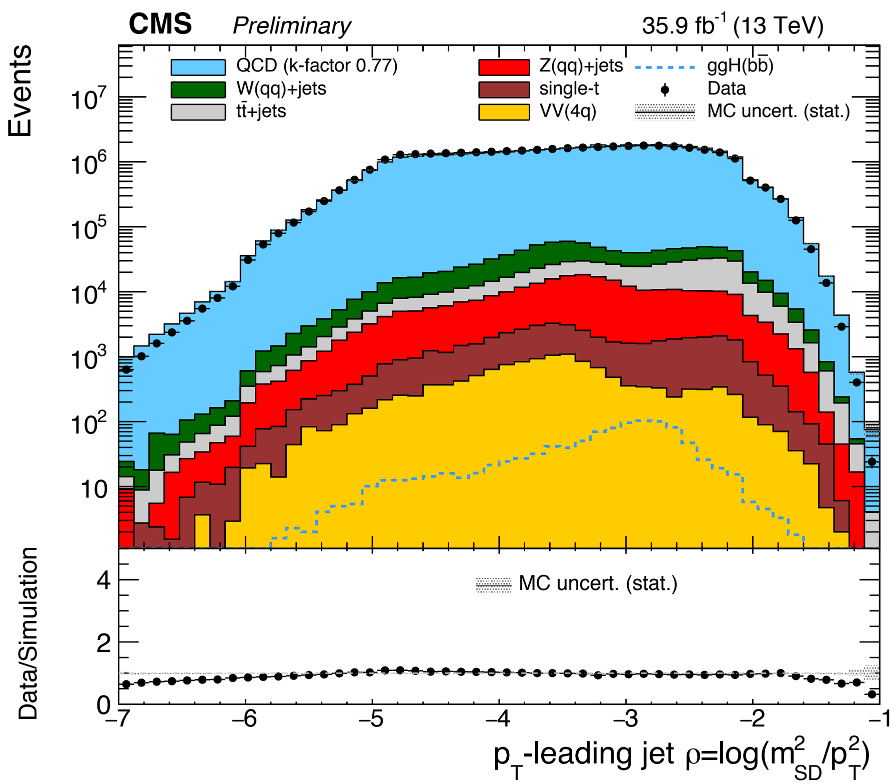


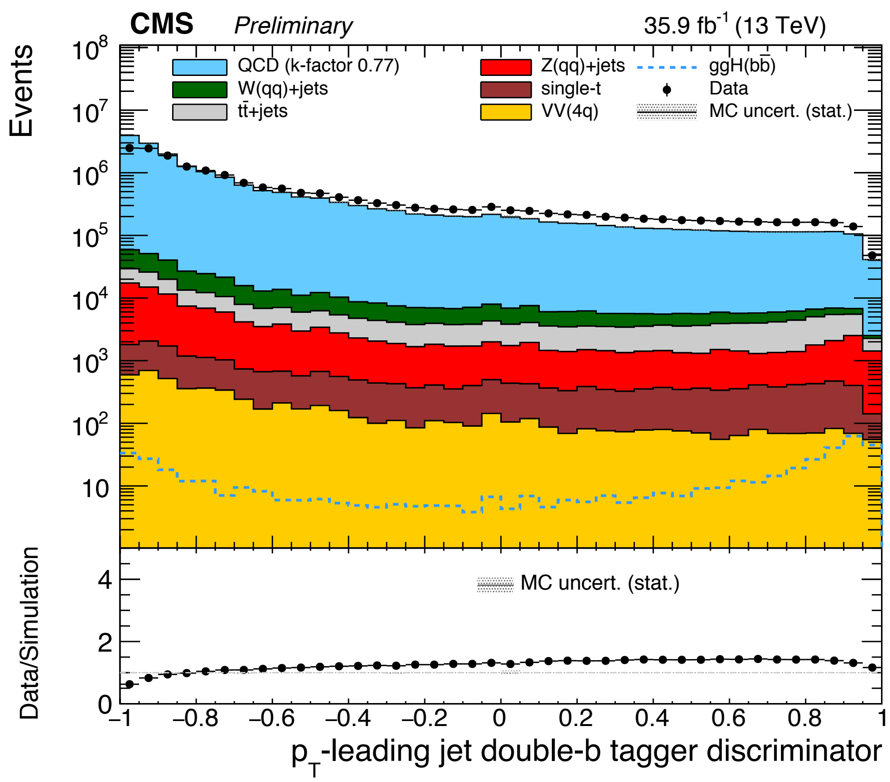


CMS-PAS-BTV-16-002

Boosted H(bb), Additional Distributions

- **Left:** QCD Mass Scale Variable
- **Right:** double-b tagger discriminator





CMS-PAS-HIG-17-010

Boosted H(bb), Full Systematics

Table 1: Systematic uncertainties and their relative size.

Systematic uncertainty source	Type (shape or normalization)	Relative size (or description)
QCD transfer factor	both	profile $a_{k\ell}$ and QCD normalization
Luminosity	normalization	2.5%
V-tag ($N_2^{1,DDT}$) efficiency	normalization	4.3%
Muon veto efficiency	normalization	0.5%
Electron veto efficiency	normalization	0.5%
Trigger efficiency	normalization	4%
Muon ID efficiency	shape	up to 0.2%
Muon isolation efficiency	shape	up to 0.1%
Muon trigger efficiency	shape	up to 8%
$t\bar{t}$ normalization SF	normalization	from 1μ CR: 8%
$t\bar{t}$ double-b mis-tag SF	normalization	from 1μ CR: 15%
W/Z NLO QCD corrections	normalization	10%
W/Z NLO EWK corrections	normalization	15% – 35%
W/Z NLO EWK ratio decorrelation	normalization	5% – 15%
double-b tagging efficiency	normalization	4%
Jet energy scale	normalization	up to 10%
Jet energy resolution	normalization	up to 15%
Jet mass scale	shape	shift m_{SD} peak by $\pm 0.4\%$
Jet mass resolution	shape	smear m_{SD} distribution by $\pm 9\%$
Jet mass scale p_T	normalization	0.4%/100 GeV (p_T)
Monte Carlo statistics	normalization	-
H p_T correction (gluon fusion)	both	30%

CMS-PAS-HIG-17-010

Boosted H(bb) Uncertainties

- The dominant uncertainties for the Higgs signal and W/Z are shown
- The Higgs pT-based corrections applied to correct for finite top mass contributions play a large role in signal uncertainties

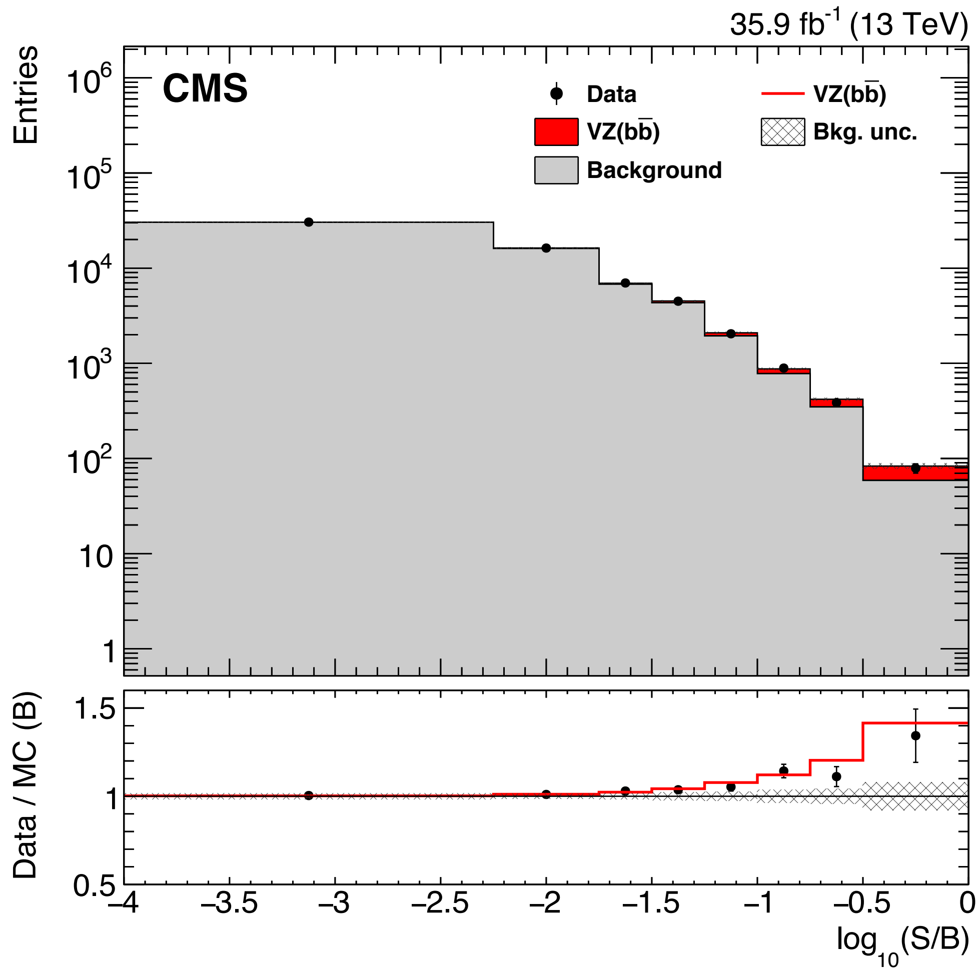
Systematic source	W/Z	H
Integrated luminosity	2.5%	2.5%
Trigger efficiency	4%	4%
Pileup	<1%	<1%
$N_2^{1,\mathrm{DDT}}$ selection efficiency	4.3%	4.3%
Double-b tag	4% (Z)	4%
Jet energy scale / resolution	10/15%	10/15%
Jet mass scale (p_T)	0.4%/100 GeV (p_T)	0.4%/100 GeV (p_T)
Simulation sample size	2–25%	4–20% (ggF)
H p_T correction	—	30% (ggF)
NLO QCD corrections	10%	—
NLO EW corrections	15–35%	—
NLO EW W/Z decorrelation	5-15%	—

Shape-based

Shape-based & norm

VH(bb) VZ(bb) Cross Check

Combination of all channels in the VZ search, with Z(bb) into a single event BDT distribution. The bottom inset shows the ratio of the data to the predicted background, with a red line overlaying the expected SM contribution from VZ with Z(bb).



Validation results for VZ production with Z(bb). Expected and observed significances, and the observed signal strengths. Significance values are given in numbers of standard deviations.

Channels	Significance expected	Significance observed	Signal strength observed
0-lepton	3.1	2.0	0.57 ± 0.32
1-lepton	2.6	3.7	1.67 ± 0.47
2-lepton	3.2	4.5	1.33 ± 0.34
Combined	4.9	5.0	1.02 ± 0.22

VH(bb) Results Cont.

- The total numbers of events in each channel, for the 20% most sensitive region of the event BDT output distribution is shown for all background processes, for the SM Higgs boson VH signal, and for data. The yields from simulated samples are computed with adjustments to the shapes and normalizations of the BDT distributions given by the signal extraction fit.

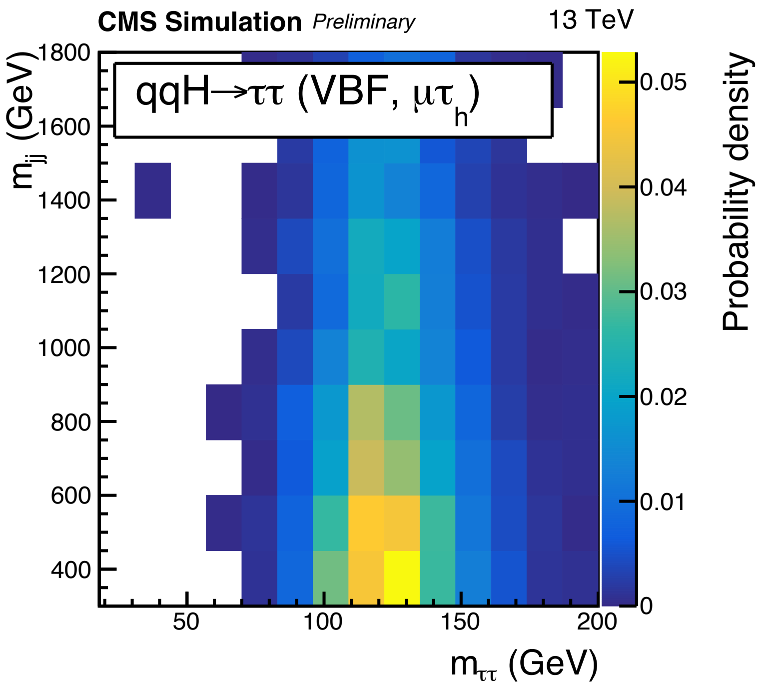
Process	0-lepton	1-lepton	2-lepton low-$p_T(V)$	2-lepton high-$p_T(V)$
Vbb	216.8	102.5	617.5	113.9
Vb	31.8	20.0	141.1	17.2
V+udscg	10.2	9.8	58.4	4.1
$t\bar{t}$	34.7	98.0	157.7	3.2
Single top quark	11.8	44.6	2.3	0.0
VV(udscg)	0.5	1.5	6.6	0.5
VZ(bb)	9.9	6.9	22.9	3.8
Total background	315.7	283.3	1006.5	142.7
VH	38.3	33.5	33.7	22.1
Data	334	320	1030	179
S/B	0.12	0.12	0.033	0.15

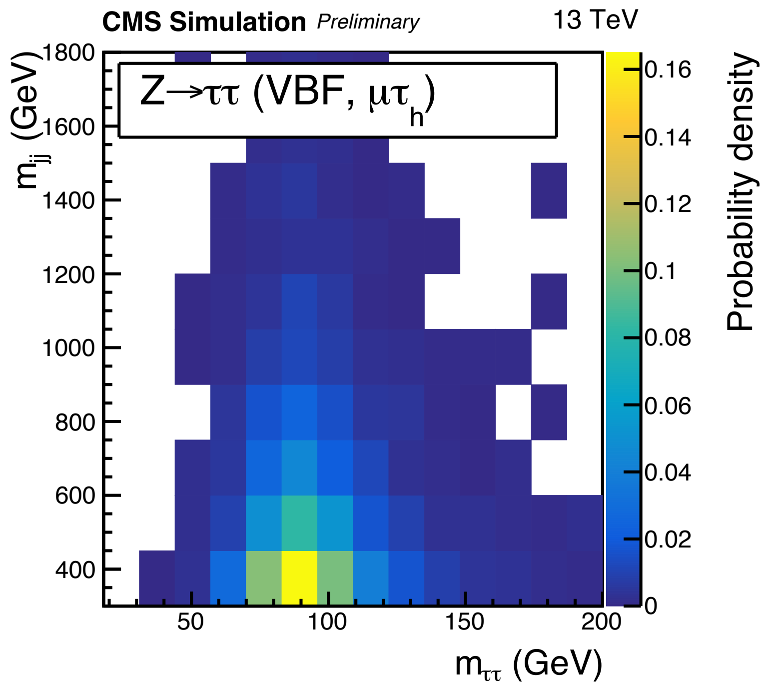
- The expected and observed significances and the observed signal strengths for VH production with H(bb) for Run1 data, Run2 (2016) data, and for the combination of the two. Significance values are given in numbers of standard deviations.

Data used	Significance expected	Significance observed	Signal strength observed
Run 1	2.5	2.1	$0.89^{+0.44}_{-0.42}$
Run 2	2.8	3.3	$1.19^{+0.40}_{-0.38}$
Combined	3.8	3.8	$1.06^{+0.31}_{-0.29}$

Separation of Signal and Background

- Additional separation of signal and background can be gained by specifically targeting different higgs production kinematics for VBF & gluon fusion
- In the VBF targeted category, we consider 2D distributions of di-tau mass ($m_{\tau\tau}$) and m_{jj}

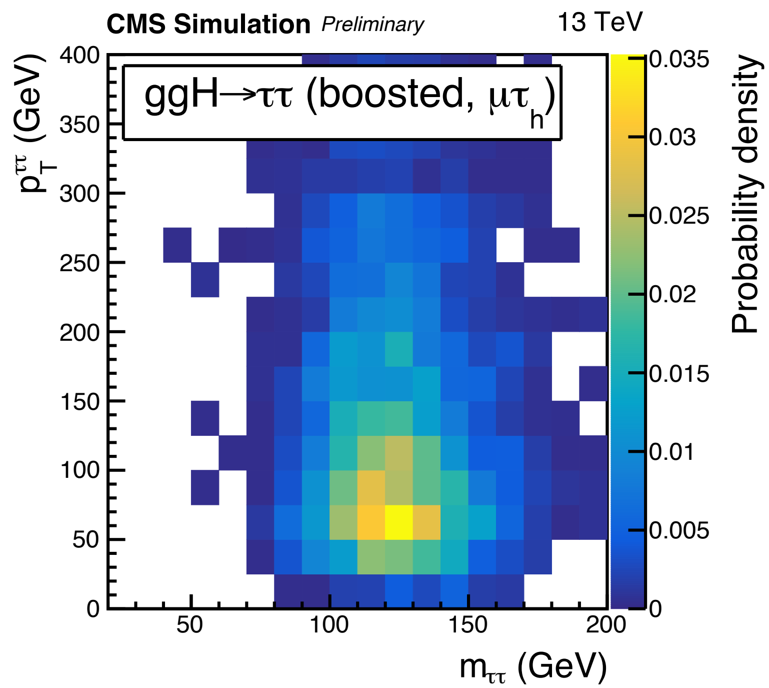


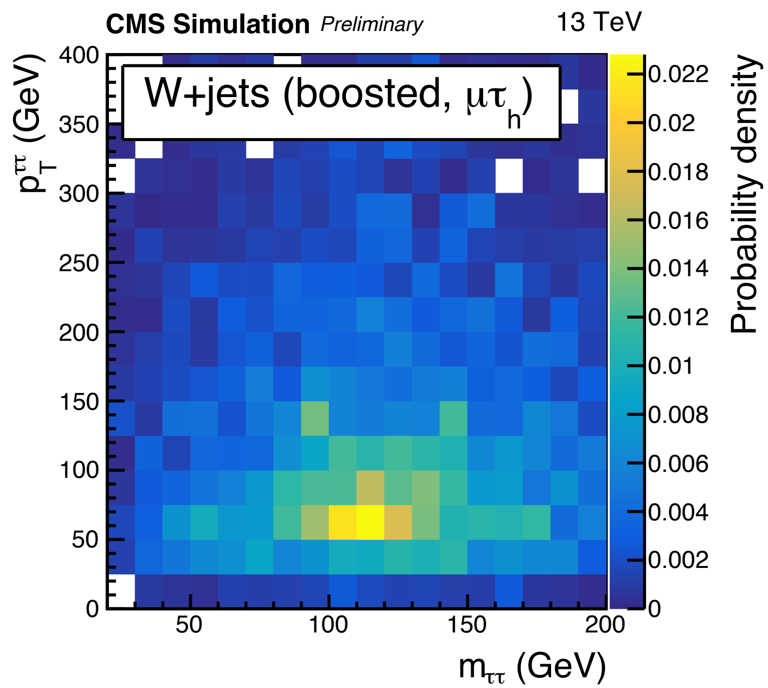


CMS-HIG-16-043

Separation of Signal and Background

- Additional separation of signal and background can be gained by specifically targeting different higgs production kinematics for VBF & gluon fusion
- In the Boosted category, we consider 2D distributions of di-tau mass ($m_{\tau\tau}$) and Higgs pT ($p_T^{\tau\tau}$)

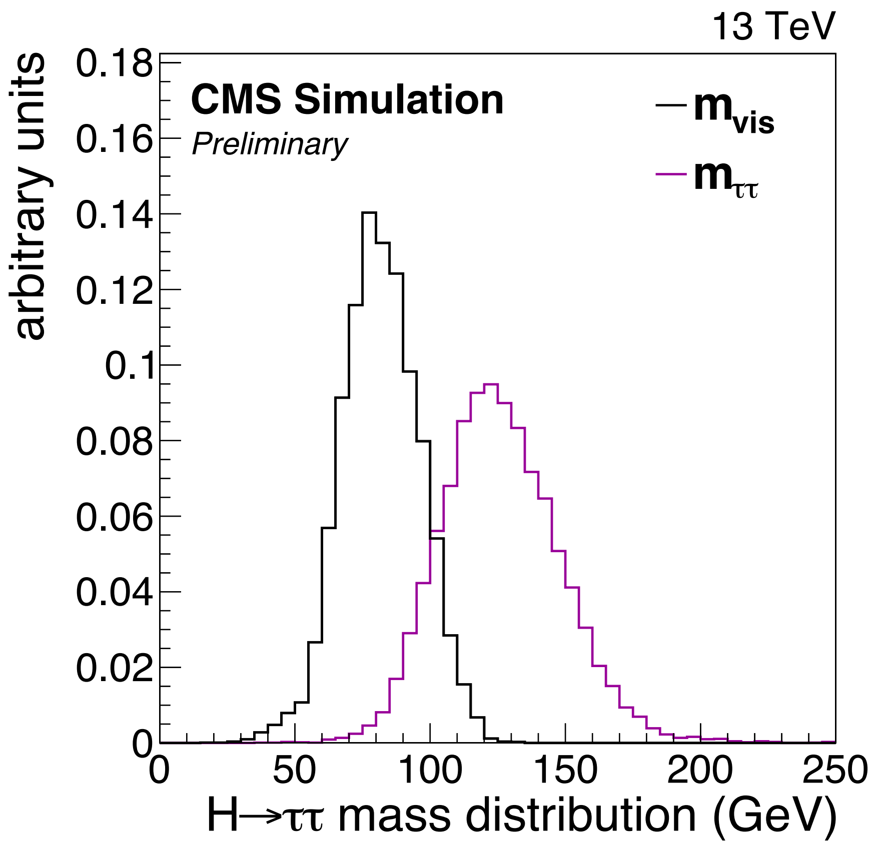




CMS-HIG-16-043

H(ττ) di-Tau Mass Reconstruction

- Missing neutrino energy complicates di-tau mass reconstruction
- Likelihood-based estimations of parent boson mass and kinematics is provided by SVFit algorithm improving signal separation
- Includes: tau lepton 4-vectors, MET, MET uncertainty



CMS-HIG-16-043

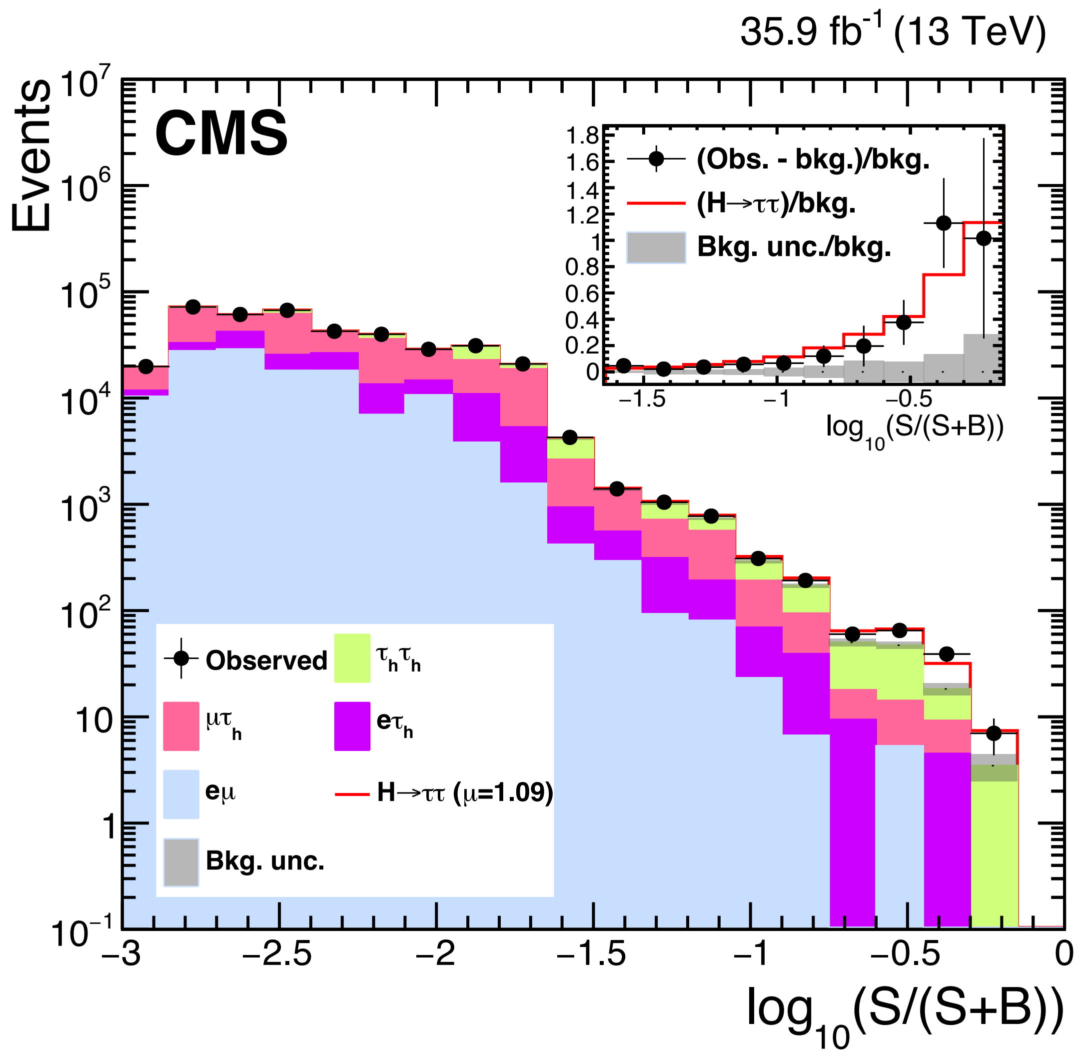
H(ττ) Systematics

Source of uncertainty	Prefit	Postfit (%)
τ_h energy scale	1.2% in energy scale	0.2–0.3
e energy scale	1–2.5% in energy scale	0.2–0.5
e misidentified as τ_h energy scale	3% in energy scale	0.6–0.8
μ misidentified as τ_h energy scale	1.5% in energy scale	0.3–1.0
Jet energy scale	Dependent upon p_T and η	—
$\vec{p}_T^{miss}$ energy scale	Dependent upon p_T and η	—
τ_h ID & isolation	5% per τ_h	3.5
τ_h trigger	5% per τ_h	3
τ_h reconstruction per decay mode	3% migration between decay modes	2
e ID & isolation & trigger	2%	—
μ ID & isolation & trigger	2%	—
e misidentified as τ_h rate	12%	5
μ misidentified as τ_h rate	25%	3–8
Jet misidentified as τ_h rate	20% per 100 GeV τ_h p_T	15
$Z \to \tau\tau/\ell\ell$ estimation	Normalization: 7–15%	3–15
	Uncertainty in $m_{\ell\ell/\tau\tau}$, $p_T(\ell\ell/\tau\tau)$, and m_{jj} corrections	—
W + jets estimation	Normalization (eμ, $\tau_h\tau_h$): 4–20%	—
	Unc. from CR (eτ_h, $\mu\tau_h$): $\simeq$5–15	—
	Extrap. from high-m_T CR (eτ_h, $\mu\tau_h$): 5–10%	—
QCD multijet estimation	Normalization (eμ): 10–20%	5–20%
	Unc. from CR (eτ_h, $\tau_h\tau_h$, $\mu\tau_h$): $\simeq$5–15%	—
	Extrap. from anti-iso. CR (eτ_h, $\mu\tau_h$): 20%	7–10
	Extrap. from anti-iso. CR ($\tau_h\tau_h$): 3–15%	3–10
Diboson normalization	5%	—
Single top quark normalization	5%	—
$t\bar{t}$ estimation	Normalization from CR: $\simeq$5%	—
	Uncertainty on top quark p_T reweighting	—
Integrated luminosity	2.5%	—
b-tagged jet rejection (eμ)	3.5–5.0%	—
Limited number of events	Statistical uncertainty in individual bins	—
Signal theoretical uncertainty	Up to 20%	—

CMS-HIG-16-043

H(ττ) Significance Distribution

- Observed signal strength compatible with SM in all categories (**left**)
- Merging all signal regions into the same plot and ordering the post-fit bins by log(S/B), a clear excess is seen in the most significant bins (**right**)



H(ττ) Couplings