Standard Model Higgs to Fermions

with Run-2 data in CMS experiment

Yiwen Wen
on behalf of CMS collaboration
DESY

Kobe, DIS2018, 2018.04.17

HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

DIS 2018

Overview

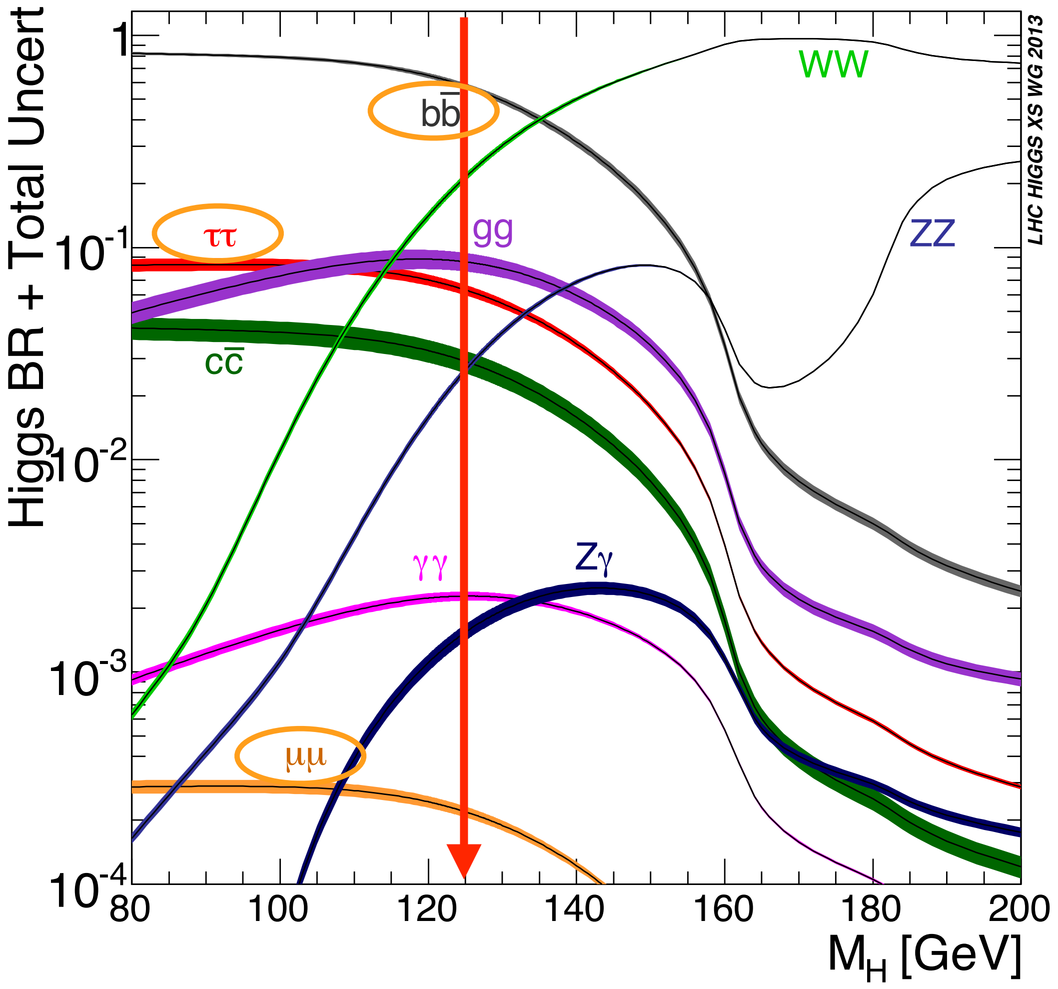
decay of the Higgs boson

in Run-I, a 125 GeV Higgs boson was discovered by CMS and ATLAS

- based on the bosonic decay channel: $\gamma\gamma$, ZZ and WW
- the properties measured so far are consistent with SM

SM also predicts fermonic decay of Higgs boson

- large branching ratios for bb and tautau
- analysis less sensitive due to overwhelming background
- observation of Higgs to fermions decays is crucial to test Yukawa coupling



contents

35.9/fb, full 2016 data

- **observation of H to tau tau (Phys. Lett. B 779 (2018) 283)**
- **evidence for VH, H to bb (Phys. Lett. B 780 (2018) 501)**
- **search for boosted H to bb (Phys. Rev. Lett. 120 (2018) 071802)**
- **search for H to μμ (CMS-PAS-HIG-17-019)**

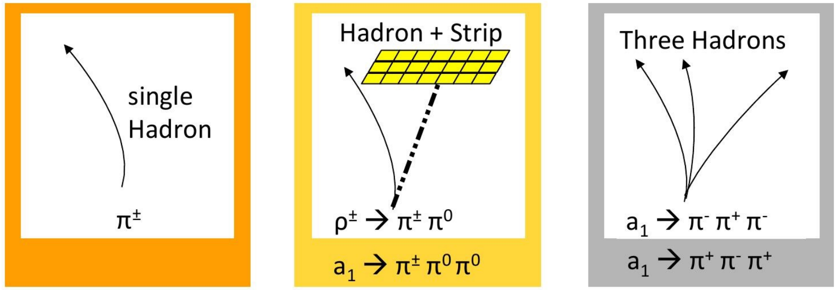
*see Aruna’s presentation for ttH, H to bb results

observation of H to tau tau

Phys. Lett. B 779 (2018) 283

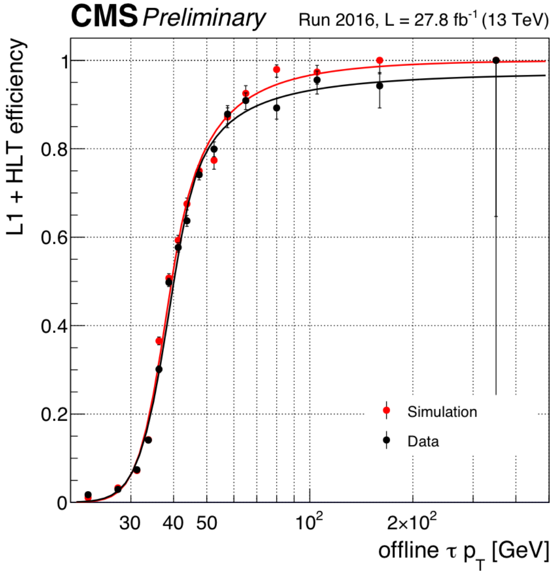
advantage to study Yukawa coupling: **higher event rate** than Higgs to mumu and **less backgrounds** than Higgs to bb

hadron plus strip algorithm to identify hadronic tau decay modes



exploit the new developments of **hadronic tau id and triggers** since Run-I:

- **dynamic strip reconstruction,** strip size adjusted dynamically as a function of the pT of e/γ
  - better acceptance at low pT, better jet rejection at high pT
- MVA-based **discriminator rejecting jet faking tau**
- **L1 trigger upgraded**:
  - increases the algorithm complexity and the readout granularity
  - improving hadronic tau-id with dynamic clustering technique at the hardware level
  - maintaining the **same pT threshold as Run-I** but **improve the turn-on and efficiency** despite of high pile-up

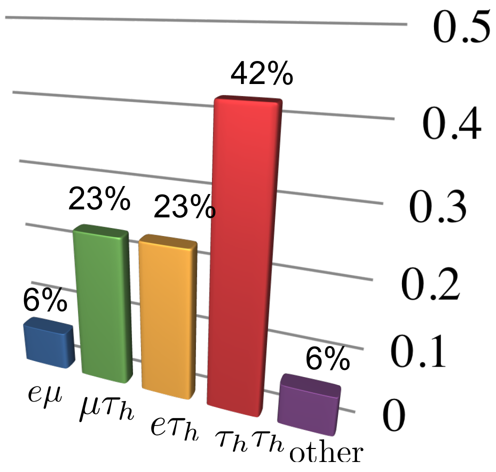


observation of H to tau tau

Phys. Lett. B 779 (2018) 283

analysis strategies

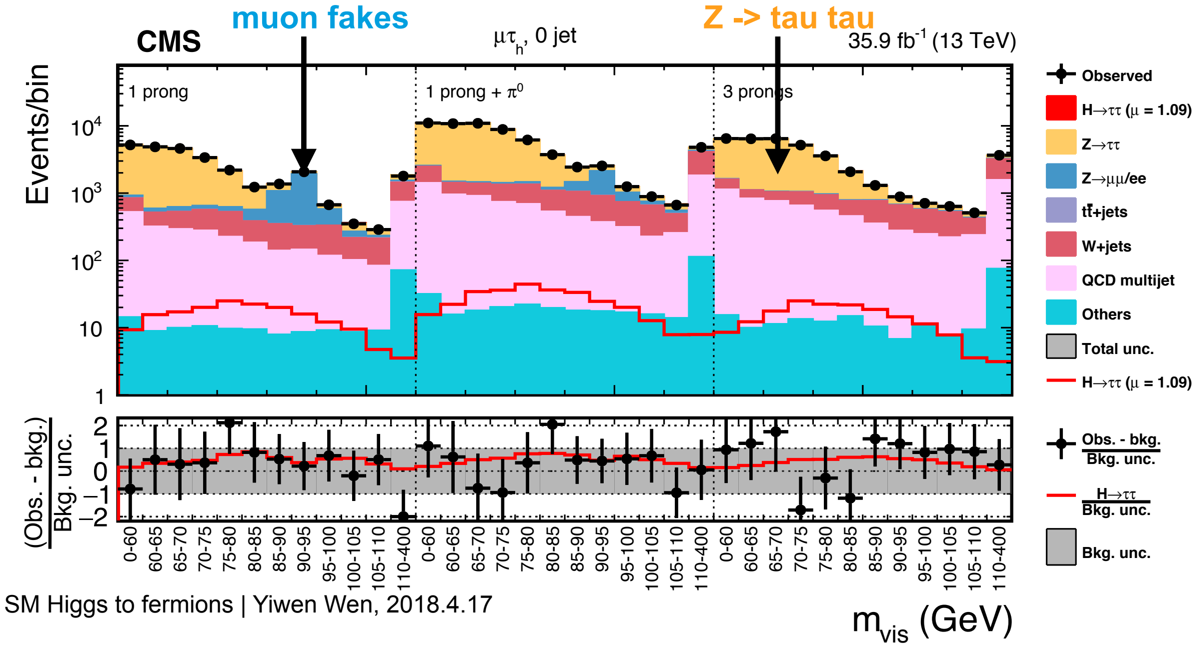
cover **four channels of di-tau decay**

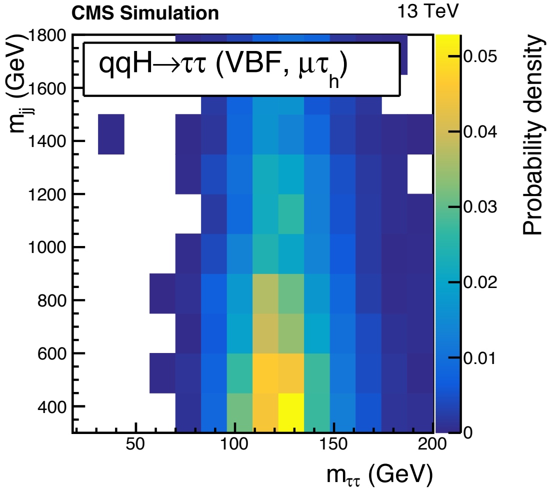


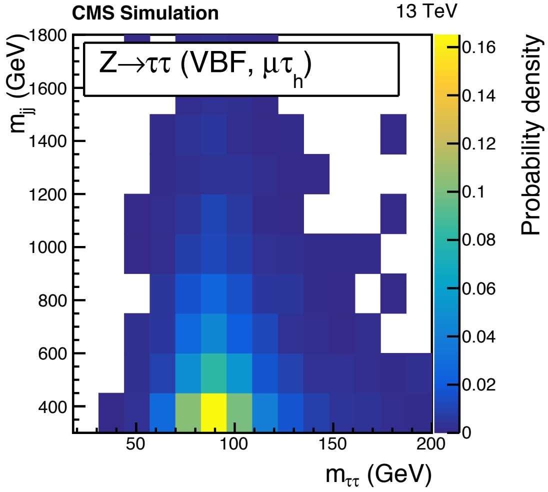
3 event categories focused on two most sensitive production mode **ggH** and **VBF**:

- **0-jet**: targeting ggH and allowing for systematics to be constrained in other categories
- **VBF**: targeting VBF production by requiring mjj cut
- **boosted**: targeting ggH events with a Higgs recoiling against a jet

two dimensional distributions for signal extraction



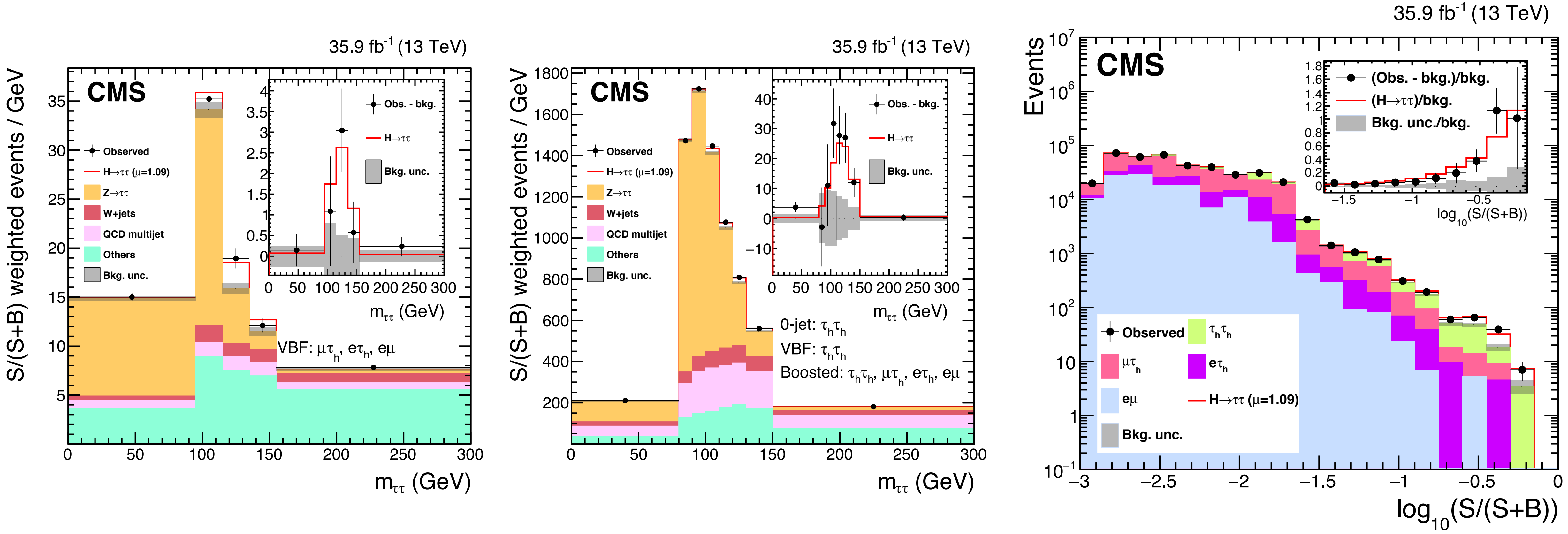




observation of H to tau tau

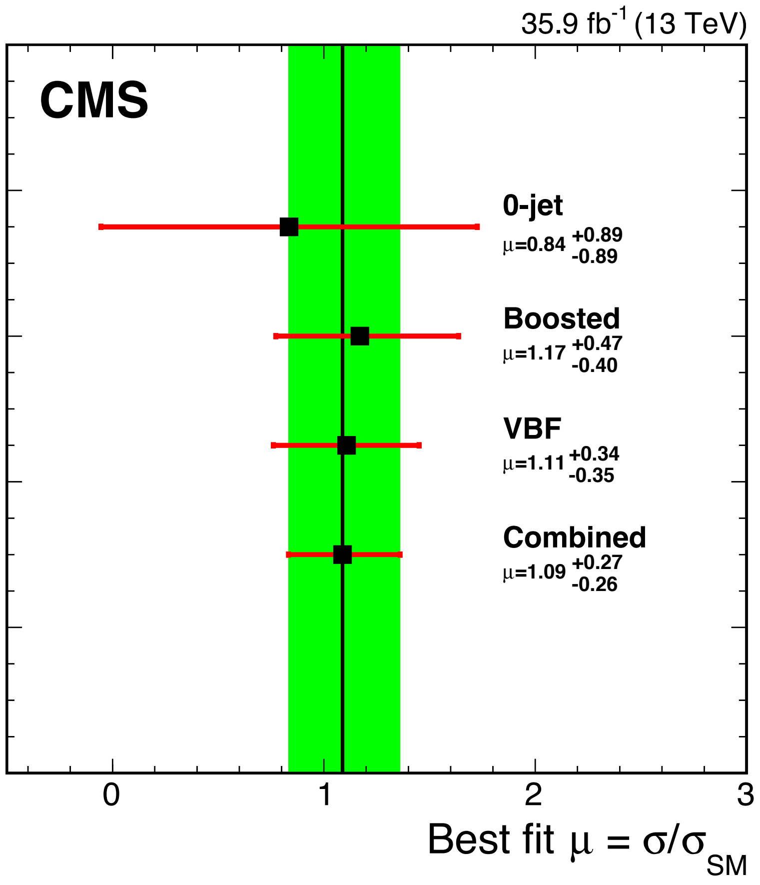
Phys. Lett. B 779 (2018) 283

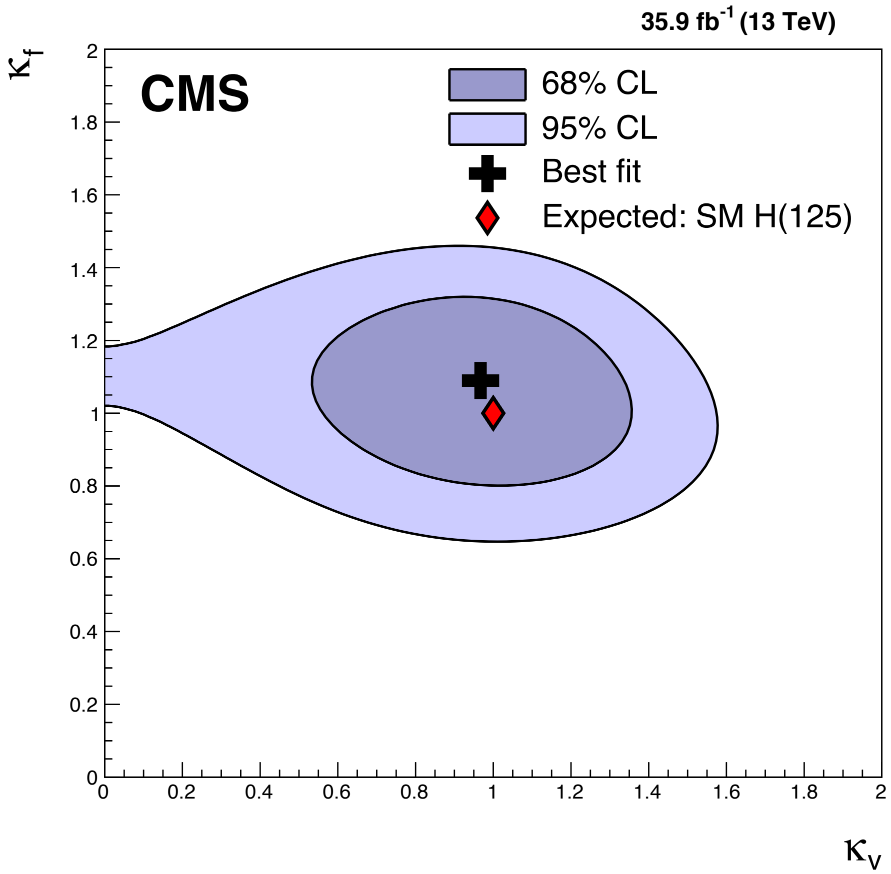
visualizations of signals



observation of H to tau tau

Phys. Lett. B 779 (2018) 283



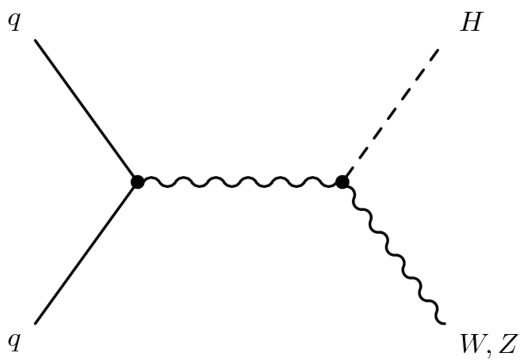


$$\mu = 1.09^{+0.15}_{-0.15}(stat)^{+0.16}_{-0.15}(syst)^{+0.10}_{-0.08}(theo)^{+0.13}_{-0.12}(bbb)$$

- 4.9(4.7)σ observe (expected) significance
- 5.9σ observed significance when combining Run-I and Run-II (7+8+13 TeV)

evidence for VH, H to bb

Phys. Lett. B 780 (2018) 501



analysis strategies

- highly **suppress QCD BG** because of requiring the presence of of a vector boson
- also providing an **efficient trigger** path when it leptonically decays
- **Higgs candidate** is required to have pT>100 GeV
    1. reduces large backgrounds from W+jets, DYJets and top
    2. makes accessible the Z(vv)H channel via large missing transverse energy
    3. improves mass resolution of the Higgs candidate

Z(ll)H(bb), 2 leptons

- **cleanest channel** due to the requirement of two leptons to tag the event
- further divided into low and high pT(V) region
- lower statistics due to relatively low Z→ll branching fraction

W(lv)H(bb), 1 lepton

- requires one lepton in the final state as well as MET
- large background contributions from top and W+jets

Z(vv)H(bb), 0 lepton

- triggers include b-tagging and require large MET
- QCD is negligible

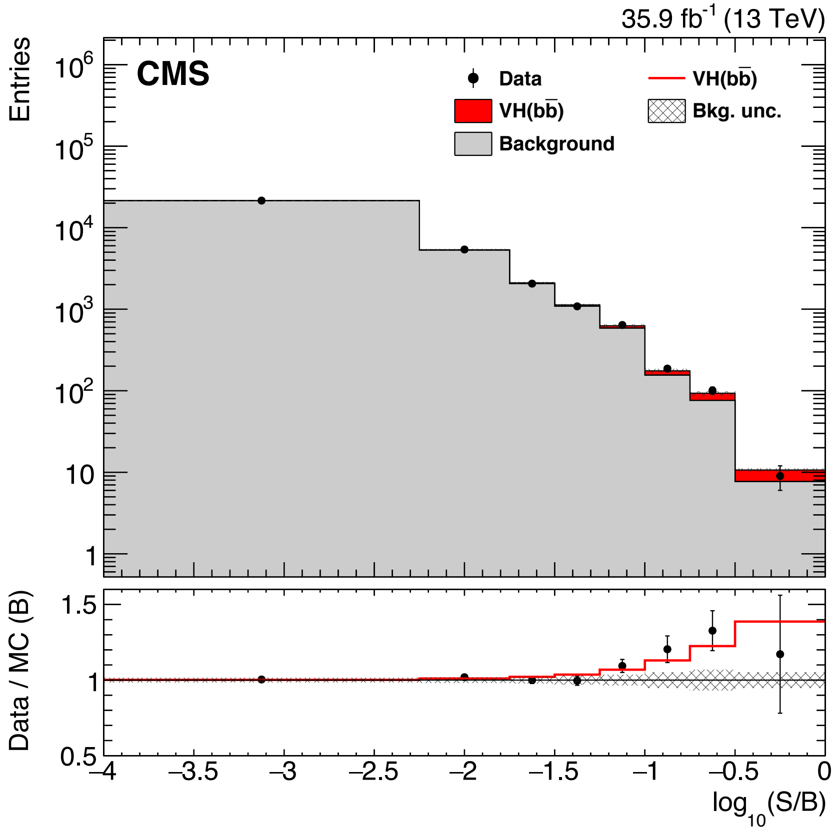
evidence for VH, H to bb

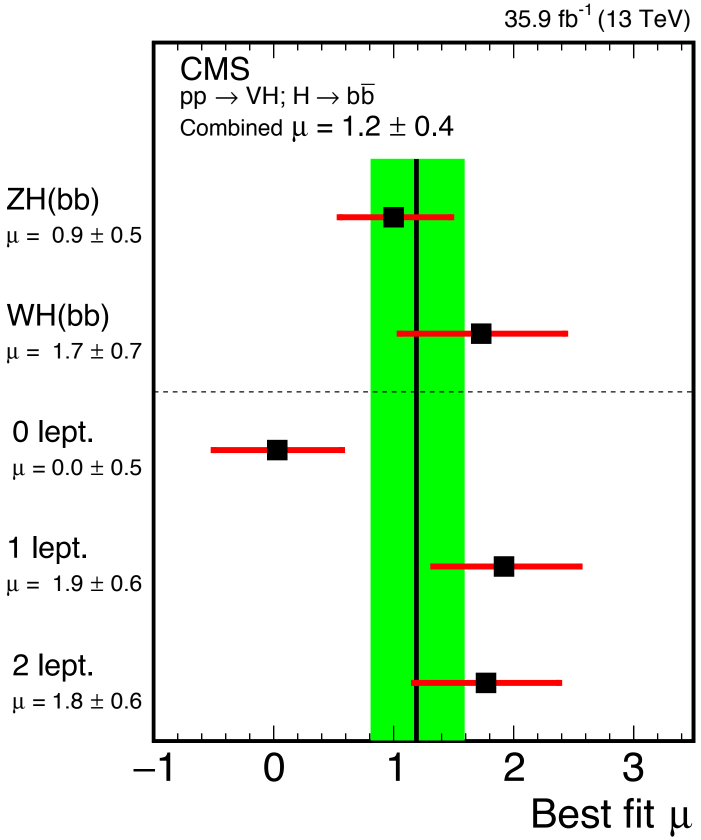
Phys. Lett. B 780 (2018) 501

- using **BDT output to discriminate** signals and background
- control regions helped to scale the yield of BG are defined with multiple additional cuts including:
  - number of jets, number of b-tagged jets, mjj, Z mass window
- maximum likelihood fit is performed for all categories simultaneously to extract the signal

- the combined best-fit signal strength is

  μ = 1.19 ± 0.21(stat.) ± 0.33(syst.)
- 3.3(2.8)σ observed (expected) significance
- **3.8σ observed significance** when combining Run-I and Run-II (8+13 TeV)



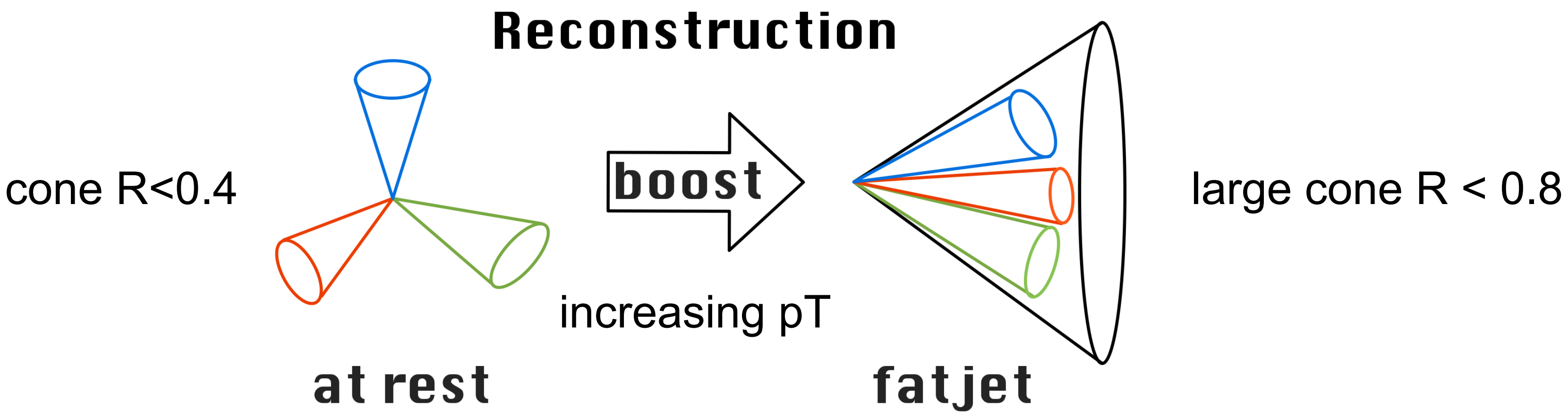


search for boosted H to bb

Phys. Rev. Lett. 120 (2018) 071802

analysis strategies

- inclusive production of H->bb usually considered inaccessible due to **the overwhelming QCD** background
- introduced a new idea **“jet substructure”** to perform the first inclusive search
- sensitivity gained in a high pT(H) region for pt>450 GeV
- selected H->bb candidate recoiling against ISR jets and veto electron/muon/taus/MET

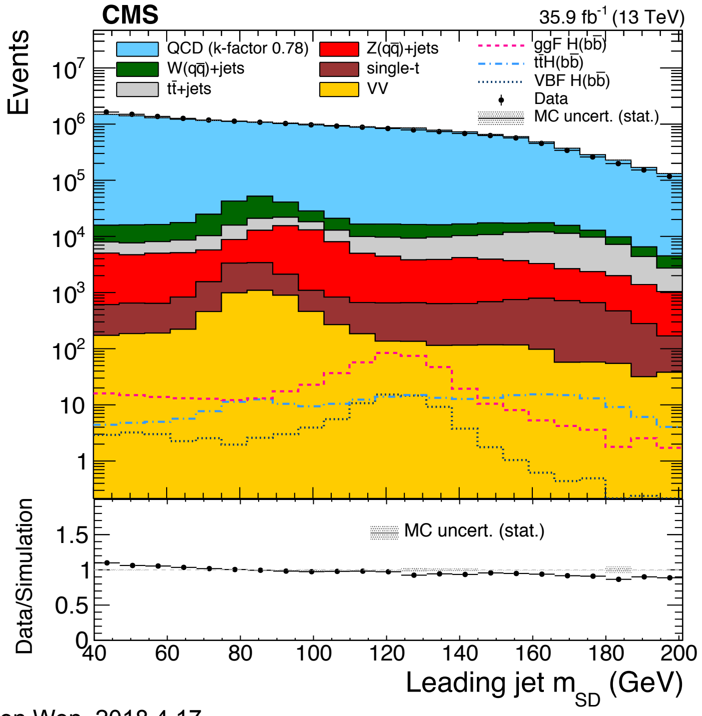


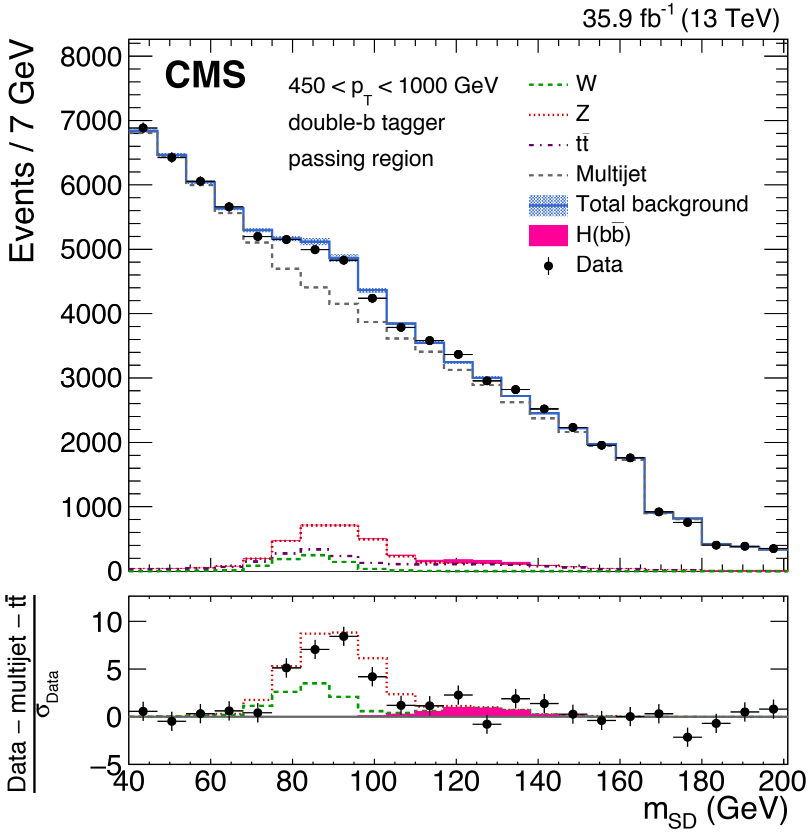
search for boosted H to bb

Phys. Rev. Lett. 120 (2018) 071802

selecting H to bb candidate

- exploit the jet-substructure techniques:
  - find 2-prong substructure in a fat jet
  - using **double-b MVA tagger** (33% efficiency for 1% fake rates for QCD) to identify 2 b-jets
  - **jet grooming**, to remove the soft and wide angle jet
- "soft-drop mass" of Higgs candidate as discriminant





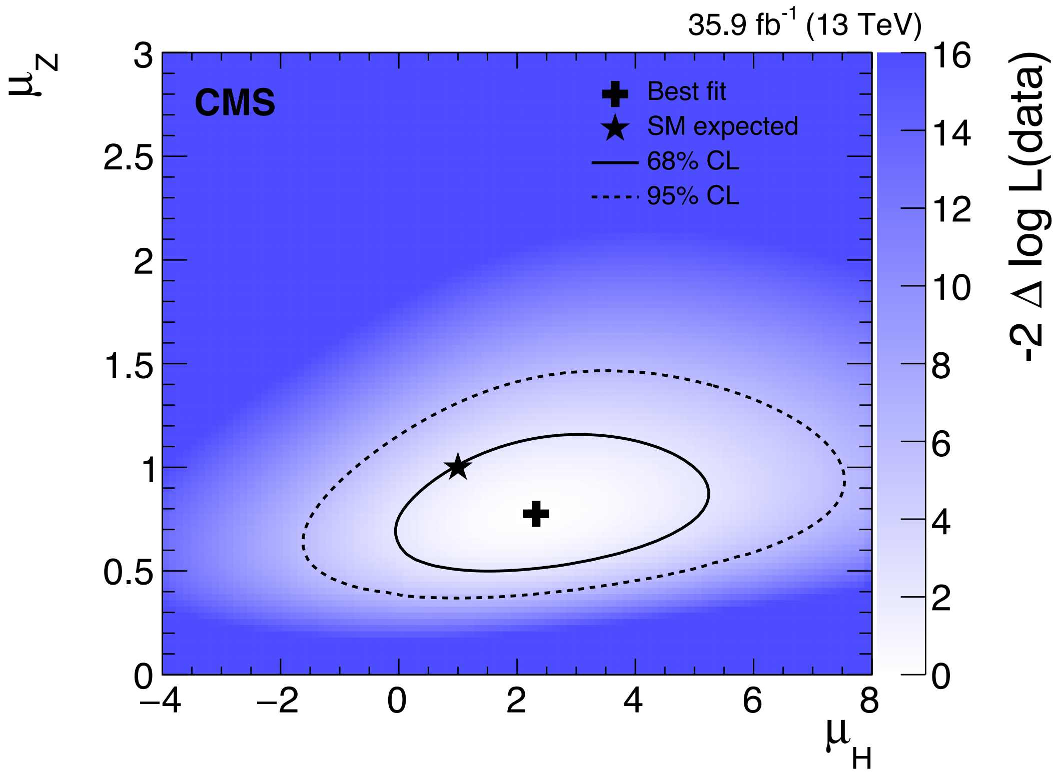
search for boosted H to bb

Phys. Rev. Lett. 120 (2018) 071802

results

- **first observation of Z->bb, 5.8σ !**
- **1.5σ significance for H(bb)**
- agrees with SM prediction

	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 UL signal strength	< 3.3	< 4.1	—
Observed UL signal strength	< 5.8	< 7.2	—
Expected significance	0.7σ	0.5σ	5.8σ
Observed significance	1.5σ	1.6σ	5.1σ



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

search for H to μμ

CMS-PAS-HIG-17-019

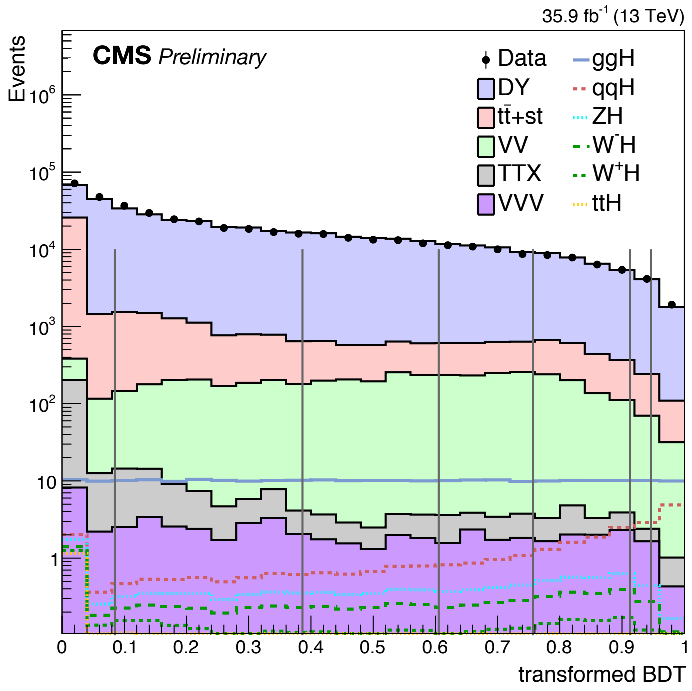
motivation

- Higgs boson decay to pair of muons extends the study of its **couplings to 2nd generation leptons**
- benefit from good muon resolution

analysis strategies

- events from MC are used to train **by BDT** and its output will be used further to optimize event categorization
- variables:
  - dimuon variables to distinguish between ggH signal and DY background
  - jet variables to identify VBF signal
  - b-tagged jets — most likely from top production

Index	BDT quantile	Max. muon $\|\eta\|$	ggH [%]	VBF [%]	WH [%]	ZH [%]	ttH [%]	Signal	Bkg./GeV @125GeV	FWHM [GeV]	Bkg. functional fit form	$S/\sqrt{B}$ @ FWHM
0	0 – 8%	$\|\eta\| < 2.4$	4.9	1.3	3.3	6.3	31.9	21.2	3150.5	4.2	mBW $\cdot B_{deg4}$	0.12
1	8 – 39%	$1.9 < \|\eta\| < 2.4$	5.6	1.7	3.9	3.5	1.3	22.3	1327.5	7.3	mBW $\cdot B_{deg4}$	0.16
2	8 – 39%	$0.9 < \|\eta\| < 1.9$	10.3	2.8	6.5	6.4	5.2	41.1	2222.2	4.1	mBW $\cdot B_{deg4}$	0.29
3	8 – 39%	$\|\eta\| < 0.9$	3.2	0.8	1.9	2.1	3.5	12.7	775.9	2.9	mBW $\cdot B_{deg4}$	0.17
4	39 – 61%	$1.9 < \|\eta\| < 2.4$	2.9	1.7	2.7	2.7	0.3	11.8	435.0	7.0	mBW $\cdot B_{deg4}$	0.14
5	39 – 61%	$0.9 < \|\eta\| < 1.9$	7.2	3.3	6.1	5.2	1.3	29.2	955.9	4.1	mBW $\cdot B_{deg4}$	0.31
6	39 – 61%	$\|\eta\| < 0.9$	3.6	1.1	2.6	2.2	0.9	14.5	479.3	2.8	mBW $\cdot B_{deg4}$	0.26
7	61 – 76%	$1.9 < \|\eta\| < 2.4$	1.2	1.5	1.8	1.7	0.2	5.2	146.6	7.6	mBW $\cdot B_{deg4}$	0.11
8	61 – 76%	$0.9 < \|\eta\| < 1.9$	4.8	3.6	4.5	4.4	0.7	20.3	514.3	4.2	mBW $\cdot B_{deg4}$	0.29
9	61 – 76%	$\|\eta\| < 0.9$	3.2	1.6	2.3	2.1	0.6	13.1	319.7	3.0	mBW	0.28
10	76 – 91%	$1.9 < \|\eta\| < 2.4$	1.2	3.1	2.2	2.1	0.2	5.8	102.4	7.2	Sum Exp(n=2)	0.14
11	76 – 91%	$0.9 < \|\eta\| < 1.9$	4.4	8.7	6.2	6.0	1.1	20.3	363.3	4.2	mBW	0.34
12	76 – 91%	$\|\eta\| < 0.9$	3.1	4.0	3.8	3.6	0.9	13.7	230.0	3.2	mBW $\cdot B_{deg4}$	0.34
13	91 – 95%	$\|\eta\| < 2.4$	1.7	6.4	2.5	2.6	0.5	8.6	95.5	4.0	mBW	0.28
14	95 – 100%	$\|\eta\| < 2.4$	2.0	19.4	1.5	1.4	0.7	13.7	82.4	4.2	mBW	0.47
overall			59.1	61.1	51.8	52.3	49.2	253.3	12961.5	3.9		

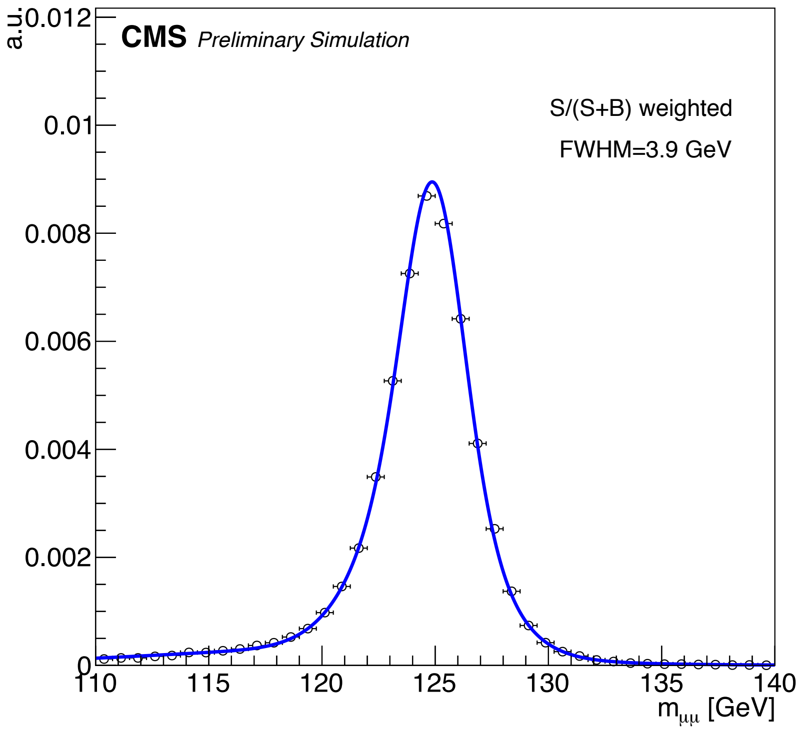


search for H to μμ

CMS-PAS-HIG-17-019

signal modeling:

- a sum of up to three Gaussian functions
- good description of the distributions



background modeling:

- functions chosen separately for each category
- choice: based on minimizing the possible bias in the fitted signal yield

$$\text{Bernsteins } (B_{deg\ n}): B(x) = \sum_{i=0}^{n} \alpha_i \left[\binom{n}{i} x^i (1-x)^{n-i} \right]$$

$$\text{Sum of exponentials (Sum Exp)}: B(x) = \sum_{i=1}^{n} \beta_i e^{\alpha_i x}$$

$$\text{Breit–Wigner}: B(x) = \frac{e^{ax}\sigma_z}{(x-\mu_z)^2 + (\frac{\sigma_z}{2})^2}$$

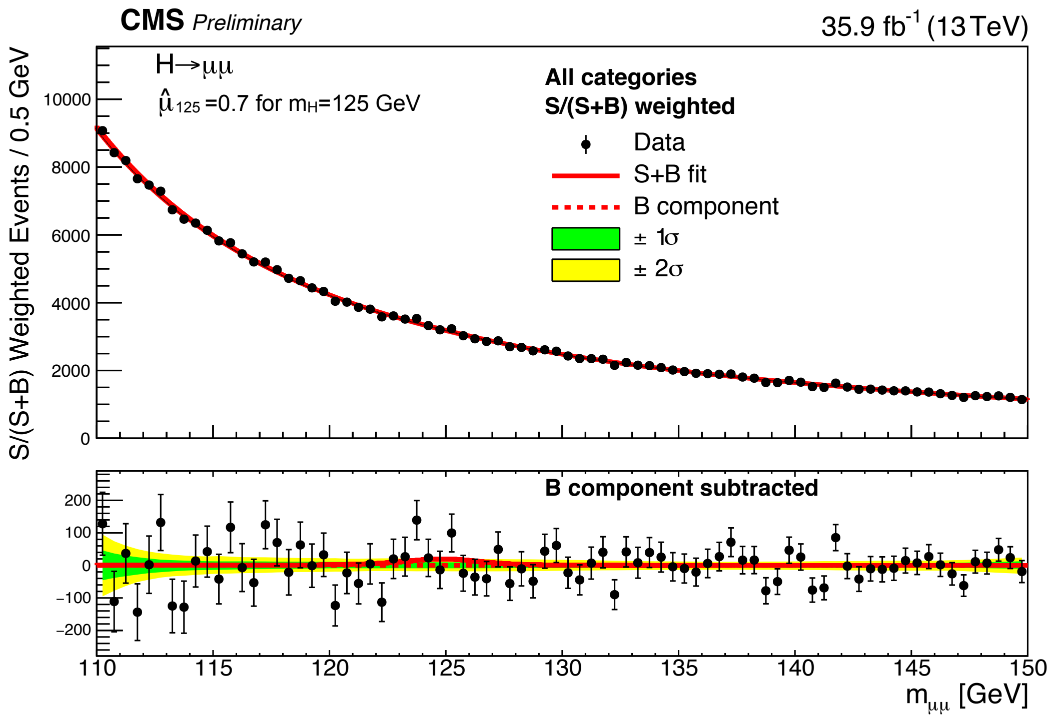
$$\text{Modified Breit–Wigner (mBW)}: B(x) = \frac{e^{a_2 x + a_3 x^2}}{(x-\mu_z)^{a_1} + (\frac{\sigma_z}{2})^{a_1}}$$

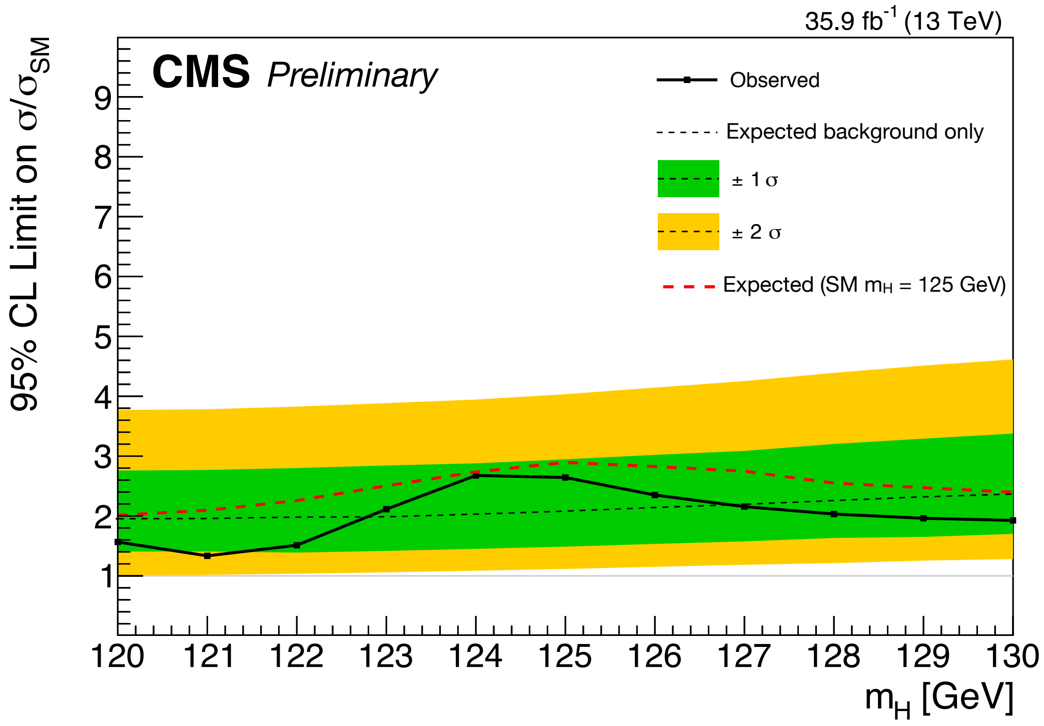
search for H to μμ

CMS-PAS-HIG-17-019

results

- **no significant excess**
- run-II: upper limit: 2.68 (2.08) times SM prediction, signal strength: μ = 0.7 ± 1.0
- run-I + run-II: upper limit: 2.68 (1.89) times SM prediction, signal strength: μ = 0.9 ± 1.0 and **significance: 0.9σ**





Summary

SM Higgs to fermions

using Run-II 13 TeV data collected by CMS, analysis enthusiasts have been pushing Higgs decaying to fermions searches and measurements into **new territory**

- **first observation of Higgs to fermion decay** by a single experiment: **Higgs to tau tau**
- **first analysis searching inclusively for Higgs to bb**
- VH, H to bb showed us **evidence for Higgs to bb** decay
- updated search results of Higgs to mumu

more exciting results will be coming

- with 2017 data collected and the on-going 2018 data-taking, will reach ~100/fb in total for run-II
- new results will pop-up anytime in this link:

  http://cms-results.web.cern.ch/cms-results/public-results/publications/HIG/index.html

back up

search for VBF, H-> bb

CMS-PAS-HIG-16-003

definitions of categories

	SingleB	DoubleB
Trigger	one b-tagged jet	two b-tagged jets
jets p_T	$p_T^{1,2,3,4} > 92, 76, 64, 30$ GeV	
jets $\lvert\eta\rvert$	<4.7	
b tag	no cut	two jets with CSV>0.5
$\Delta\phi_{bb}$	<1.6 radians	<2.4 radians
VBF topology	$m_{qq} > 460$ GeV $\lvert\Delta\eta_{qq}\rvert > 4.1$	$m_{qq} > 200$ GeV $\lvert\Delta\eta_{qq}\rvert > 1.2$
Veto	None	Events that belong to SingleB

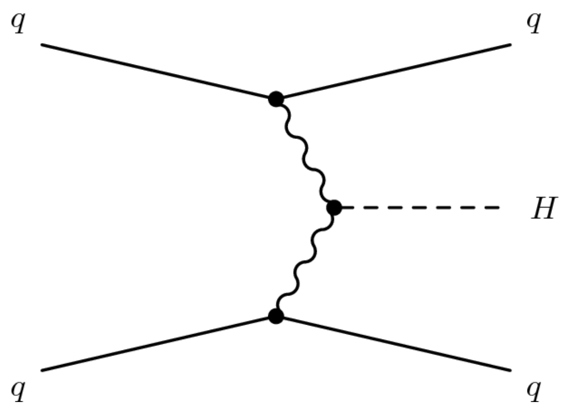
BDT boundary values	SingleB				DoubleB		
	Cat. 1	Cat. 2	Cat. 3	Cat. 4	Cat. 5	Cat. 6	Cat. 7
	0.28 – 0.72	0.72 – 0.87	0.87 – 0.93	0.93 – 1.0	0.36 – 0.76	0.76 – 0.89	0.89 – 1.0
Data	25298	5834	1281	302	69963	9831	1462
Z +jets	49± 4	12.5± 2.0	4.1± 1.1	1.7± 0.7	448± 11	50± 4	8.4± 1.7
W +jets	25.8± 3.5	1.6± 0.9	0.1± 0.1	<0.1	74± 6	4.6± 1.3	0.9± 0.6
$t\bar{t}$	53± 1	5.1± 0.2	0.7± 0.1	0.2± 0.04	534± 2	22.6± 0.4	1.1± 0.1
Single t	52± 1	9.7± 0.5	1.8± 0.2	0.4± 0.1	221± 3	23.2± 0.8	1.8± 0.2
VBF $m_H(125)$	19.5± 0.2	13.7± 0.1	7.2± 0.1	4.2± 0.1	21.7± 0.2	10.5± 0.1	3.8± 0.1
GF $m_H(125)$	5.5± 0.2	1.8± 0.1	0.6± 0.07	0.2± 0.04	18.7± 0.4	3.1± 0.1	0.6± 0.07

search for VBF, H-> bb

CMS-PAS-HIG-16-003

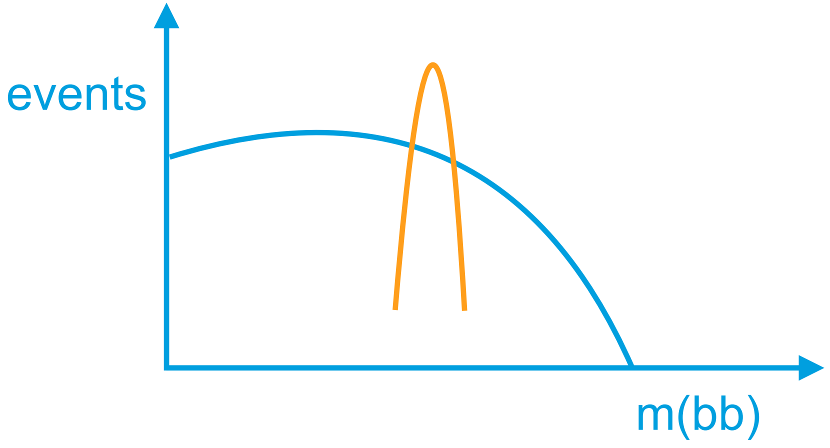
VBF H-> bb signature

- 2 central b-jets
- 2 light q-jets with large $\Delta\eta$ and m(jj)
- suppress color-flow between VBF jets



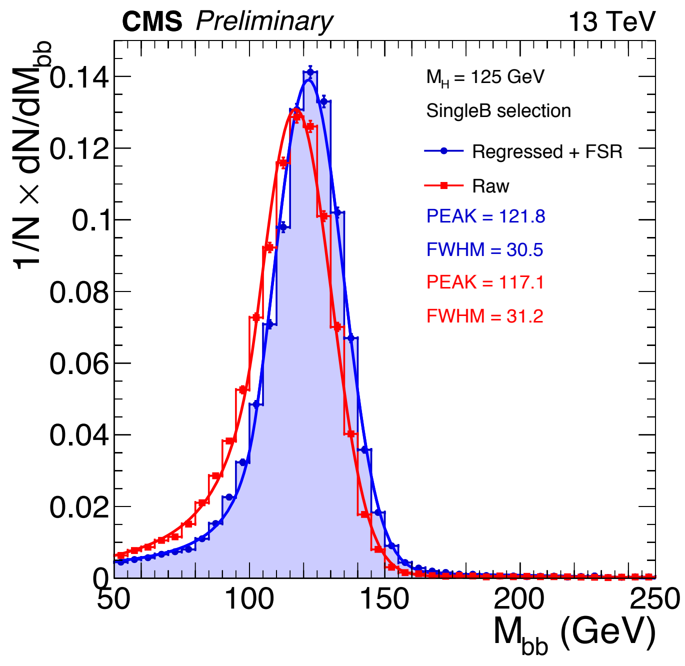
Analysis strategies

- topological triggers
- **use BDT** to exploit the difference between signals and QCD. BDT is orthogonal to b-jet kinematics
- perform fits of m(bb) spectra in different MVA categories
- search for a m(bb) bump on a smoothly falling background
- analysis is split into two parts from complementary trigger strategies: **SingleB** and **DoubleB**



search for VBF, H-> bb

CMS-PAS-HIG-16-003

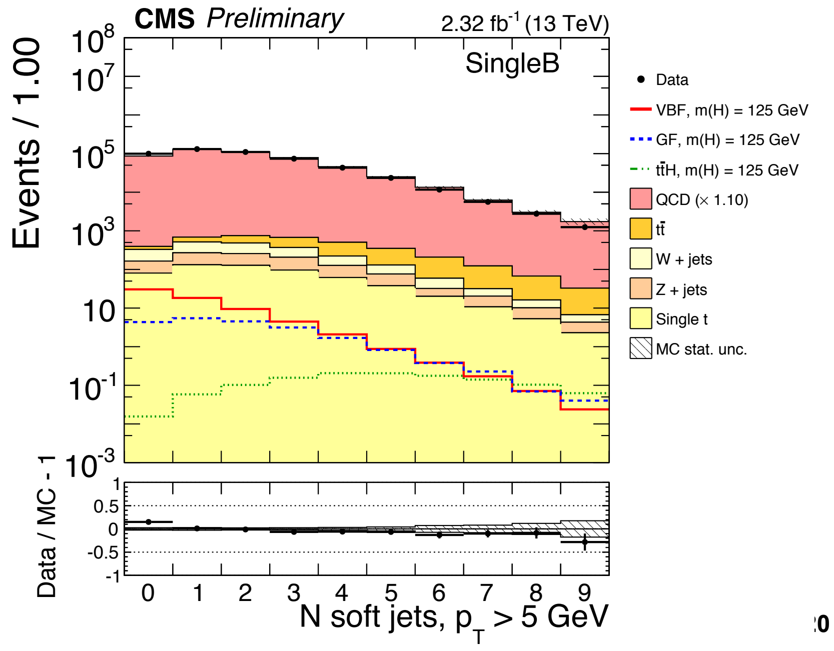


calibration of b-jet pT

- regression MVA technique is used
- trained using tt events, validated with Z+jets
- 7% improved m(bb) resolution

event categorization based on BDT

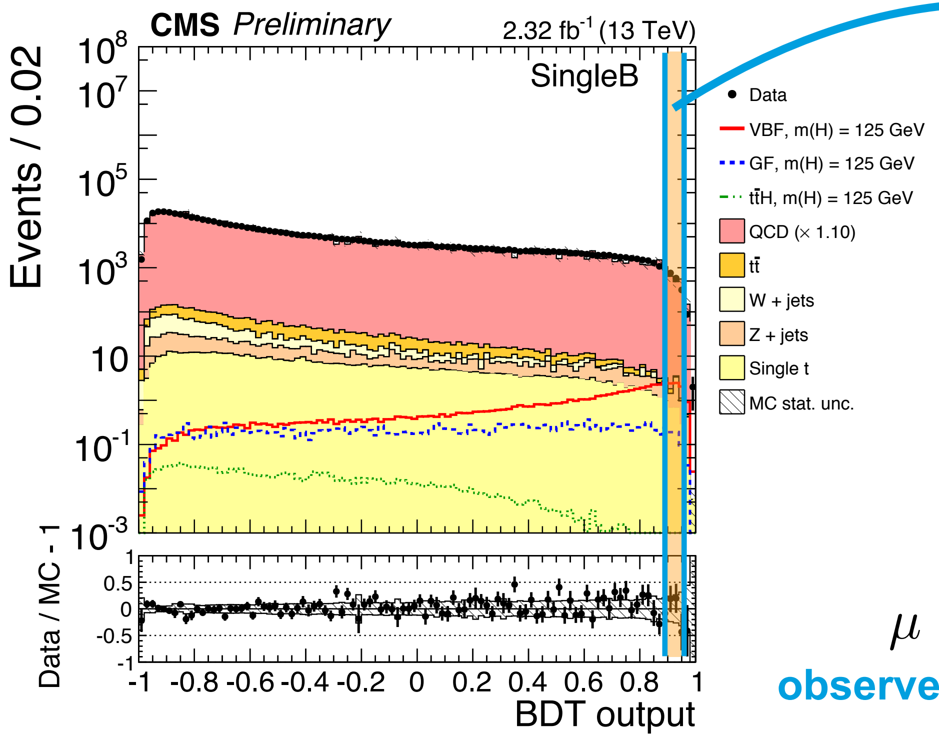
- discriminating variables:
  - soft track-jet multiplicity
  - q/g discrimination: minor RMS of jet constituents in the η-φ plane
  - VBF di-jet signature
  - b-tag
- separate trainings for **SingleB** and **DoubleB**

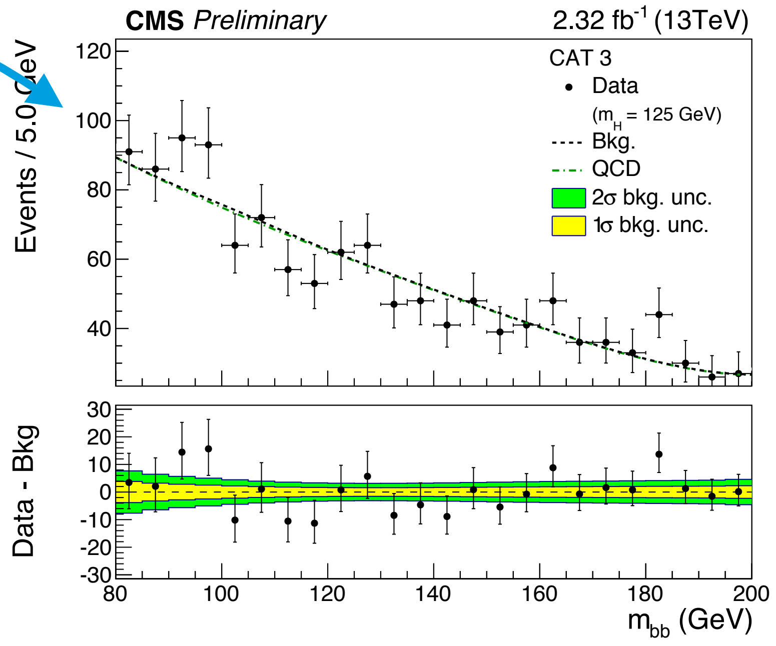


search for VBF, H-> bb

CMS-PAS-HIG-16-003

- events are divided into 7 categories based in BDT output to maximize the signal sensitivity
- QCD modeling taken from data in signal-free category and then transferred to the signal categories
- top and Z+jets are modeled from MC
- simultaneous fit in 7 signal categories





$$\mu = -3.72^{+2.39}_{-2.51}$$

observed(expected) upper limit @95% CL:

3.0 (5.0) times SM prediction

search for VBF, H-> bb

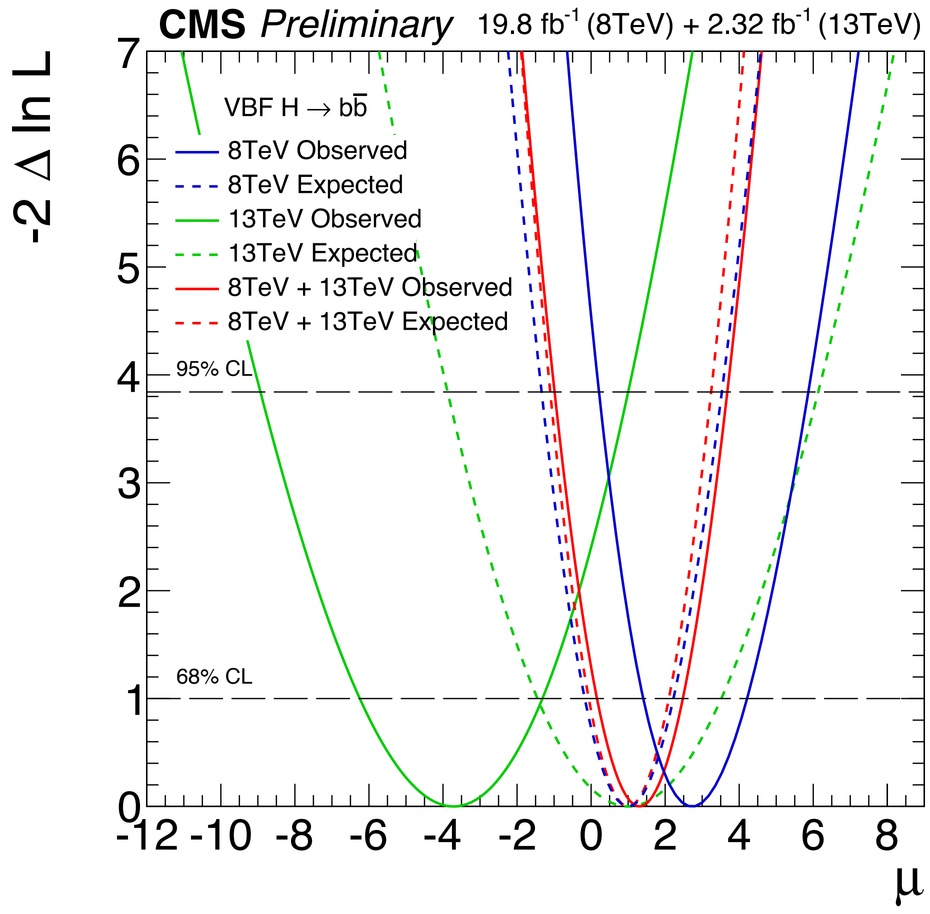
CMS-PAS-HIG-16-003

combination of Run-1 results

The combination of Run 1 and Run 2: **observed (expected) upper limit of 3.4 (2.3) times the SM prediction**, and a signal strength of

$$\mu = 1.3^{+1.2}_{-1.1}$$

with significance 1.2 standard deviation



search for VBF, H-> bb

CMS-PAS-HIG-16-003

systematics

Background uncertainties		
QCD shape parameters	determined by the fit	
QCD bkg. normalization	determined by the fit	
Top quark bkg. normalization	30%	
Z/W+jets bkg. normalization	30%	
Uncertainties affecting the signal	**VBF signal**	**GF signal**
JES (signal shape)	2%	
JER (signal shape)	2%	
Integrated luminosity	2.7%	
Branching fraction ($H \rightarrow b\bar{b}$)	1.3%	
JES (acceptance)	1–4%	2–11%
JER (acceptance)	1–2%	1–3%
b-jet tagging	3–9%	2–10%
Trigger	8–15%	6–11%
Theory uncertainties	**VBF signal**	**GF signal**
UE & PS	2–7%	10–45%
Scale variation (global)	0.4%	8%
Scale variation (categories)	1%	15%
PDF (global)	2%	3%
PDF (categories)	1–2%	1–2%

observation of H-> tau tau

Phys. Lett. B 779 (2018) 283

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%	—

evidence for VH, H-> bb

Phys. Lett. B 780 (2018) 501

systematics

Source	Type	Individual contribution to the μ uncertainty (%)	Effect of removal to the μ uncertainty (%)
Scale factors ($t\bar{t}$, V+jets)	norm.	9.4	3.5
Size of simulated samples	shape	8.1	3.1
Simulated samples' modeling	shape	4.1	2.9
b tagging efficiency	shape	7.9	1.8
Jet energy scale	shape	4.2	1.8
Signal cross sections	norm.	5.3	1.1
Cross section uncertainties (single-top, VV)	norm.	4.7	1.1
Jet energy resolution	shape	5.6	0.9
b tagging mistag rate	shape	4.6	0.9
Integrated luminosity	norm.	2.2	0.9
Unclustered energy	shape	1.3	0.2
Lepton efficiency and trigger	norm.	1.9	0.1

search for Boosted H-> bb

Phys. Rev. Lett. 120 (2018) 071802

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%	—

search for H-> μμ

Phys. Rev. Lett. 120 (2018) 071802