VBF_Hbb analysis with CMS(LHC) 2018 data of *pp*-collisions at 13 TeV

A. ALIKHANYAN
National Laboratory

Armen Tumasyan

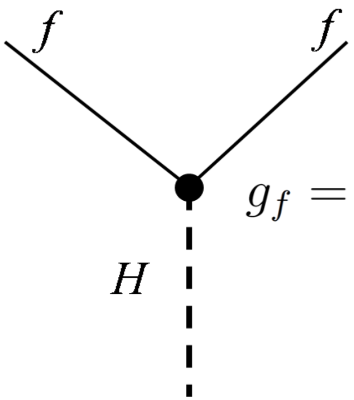
ANLS-EPD, 23.06.2021

Outline

- Introduction
- Analysis strategy and event selection
- Data/MC comparison
- Signal-background discrimination
- Signal and background models
- Systematic uncertainties
- Results and expected significance
- b-jet energy regression
- Validation with ″$Z \rightarrow$ bb +2 jets″ as standard candle
- Upcoming works

Introduction

Higgs-fermion vertex



$$g_f = \frac{g_2 m_f}{2 m_W}$$

$H \rightarrow b\bar{b}$ features:

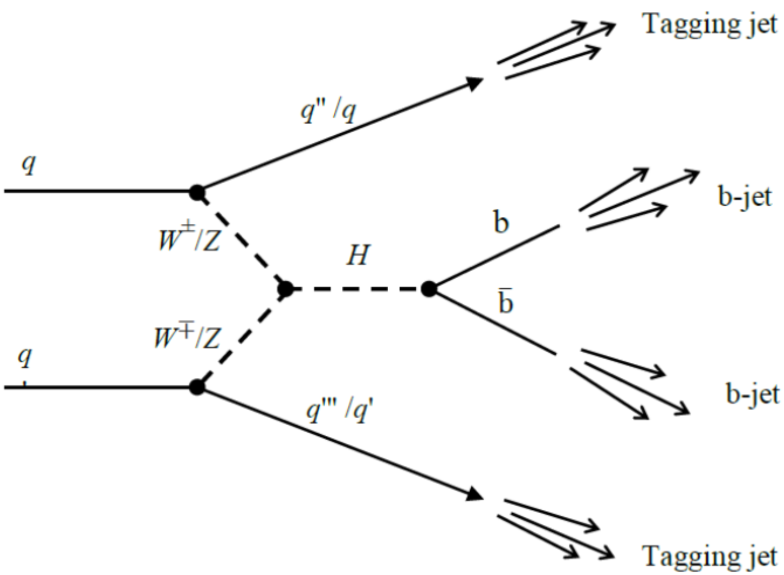
- ❑ branching ratio ~ 58 %
- ❑ direct access to *H-b* coupling
- ❑ difficulties with event reconstruction

One of the main challenges for LHC Run_2 (2015-18):

Observation of Higgs boson in 3rd generation fermions decay modes.

Features of VBF $H \rightarrow b\bar{b}$: ($\sigma \approx 2.2$ pb at 13 TeV)

- ❑ two *b*-quark jets from Higgs boson decay mainly in central region of CMS
- ❑ two light-quark jets from scattered quarks with large $\Delta\eta$ and inv. mass and forward-backward tend
- ❑ low additional hadron activity
- ❑ well suppressed background



H→bb results in ATLAS and CMS

- **H→bb observation** (**main contribution from VH_Hbb channel**)

ATLAS: μ = **1.02** ± 0.12(stat.) ± 0.14(syst.), **6,7 σ** (*with 2018 data*)

ATLAS Collaboration, Eur.Phys.J.C 81(2021)178

CMS: μ = **1.04** ± 0.20, **5,6 σ** (*without 2018 data*)

CMS Collaboration, Phys.Rev.Lett. 121(2018)121801

- **VBF H→bb status**

ATLAS: μ = **0.95** ± 0.31(stat.) (+ 0.20 – 0.17) (syst.), **2,7 σ** **(VBF inclusive)**

μ = **0.99** (+ 0.36 – 0.34), **3.0 σ** **(VBF combined)** (Run_2 ~ **126** fb^{-1})

(complementary measurement of VBF_Hbb in association with photon)

ATLAS Collaboration, arXiv:2011.08280v1 [hep-ex] 16 Nov 2020

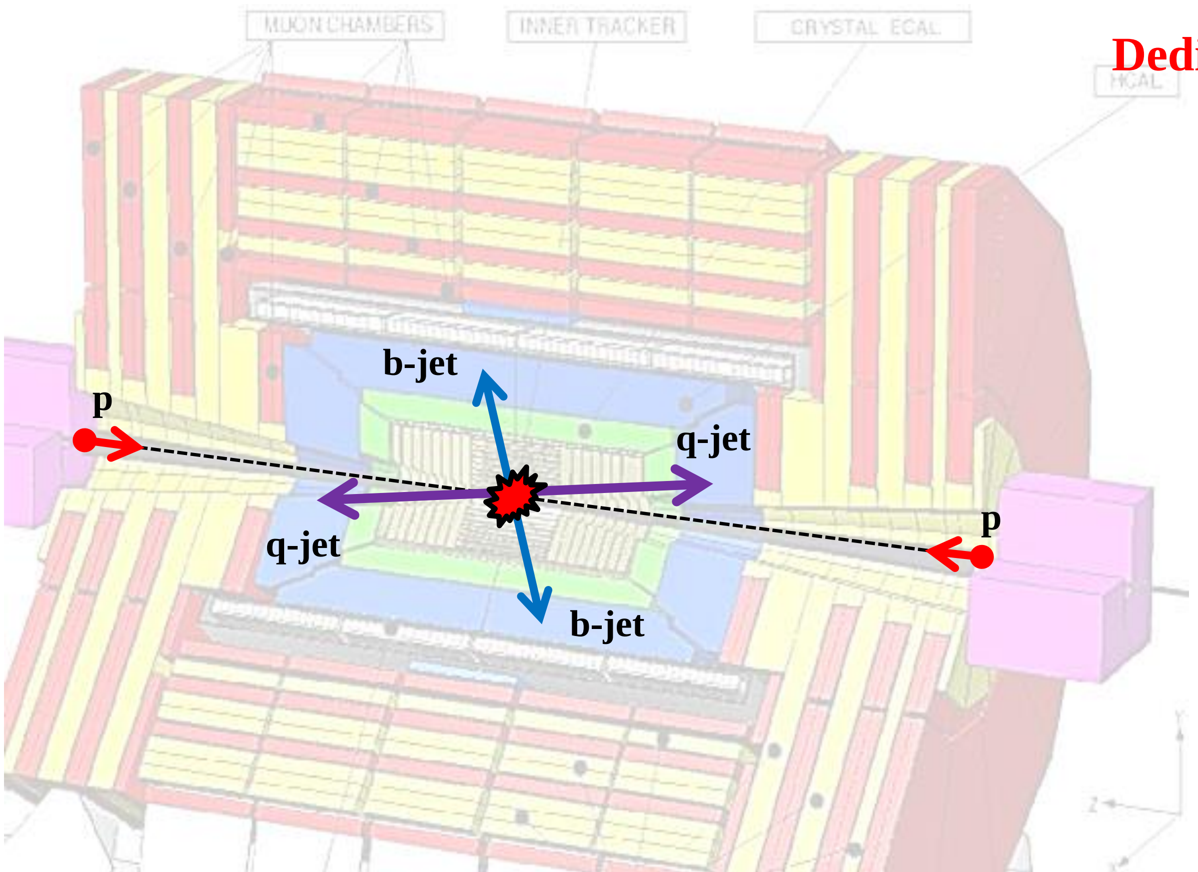
CMS : Run_2 analysis: **ongoing**

Statistics: *2016*: **35.9** fb^{-1} *2017*: **7.7** fb^{-1} *2018*: **59.7** fb^{-1}

Analysis strategy

Two main analysis categories based on two main features of VBF_Hbb process:

- ❑ **SingleB** relies on **tight** VBF topology and **soft** b-tagging
- ❑ **DoubleB** relies on **soft** VBF topology and **tight** b-tagging



Dedicated online triggers (L1 – HLT)

Online requirements:

		soft	tight	
❑ VBF topology:				
$\Delta\eta_{qq}$	>	2.4	3.5	
m_{qq}	>	240	460	(GeV)
❑ online b-tagging:				
$N_{b\text{-}jets}$	≥	1	2	

Event selection and interpretation

Online Triggers:
(*recorded integral luminosity in 2018 ~ 54.4 fb^{-1}*)

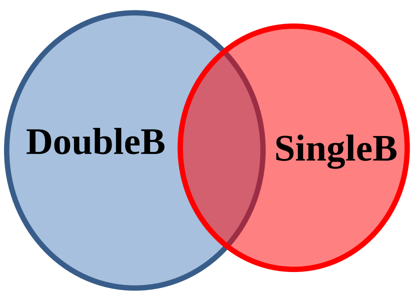
L1: L1_TripleJet_100_80_70_DoubleJet_80_70_er2p5 (SingleB & DoubleB)

HLT: HLT_QuadPFJet105_88_76_15_PFBTagDeepCSV_1p3_VBF2 (SingleB)

HLT_QuadPFJet105_88_76_15_DoublePFBTagDeepCSV_1p3_7p7_VBF (DoubleB)

Offline selections: (*follows to online triggers logic*)

- All considered jets are within $|\eta| < 4.7$, with $P_T > 30$ GeV and passing *loose* PileUp-condition
- 4 offline jets with $P_T > 120, 100, 85, 45$ GeV
- 2 most b-tagged jets among 4 (6) P_T-leading jets with $|\eta| < 2.4$ selected as b-jets (1^{st} and 2^{nd} b-jets)
- 1^{st} b-jet *tight*-tagged, 2^{nd} b-jet *medium*-tagged, $\Delta\varphi_{bb} < 1.6$ (2.1) in SingleB (DoubleB)
- 2 remaining jets among 4 P_T-leading jets selected as q-jets, $M_{qq} > 500$ (250) GeV, $\Delta\eta_{qq} > 3.8$ (2.5) in SingleB (DoubleB)
- Isolated lepton veto: **NO** e with $P_T > 7$ GeV or μ with $P_T > 5$ GeV



SingleB: signal selection efficiency ~ 0.5 %

DoubleB: signal selection efficiency ~ 0.6 %

Inclusive SingleB and exclusive DoubleB selections

General information

Reconstruction:

- ❑ Particle-Flow (PF) global reconstruction using information from all subsystems: e, μ, γ, charged and neutral hadrons, τ, missing-$E_{T,}$ etc.
- ❑ AntiKt4 jet algorithm with PF-reconstructed objects
- ❑ b-tagging (online/offline) with DeepCSV: DNN inputs: displaced secondary vertex, charged hadrons multiplicity, invariant mass, etc.   (total 66 features)

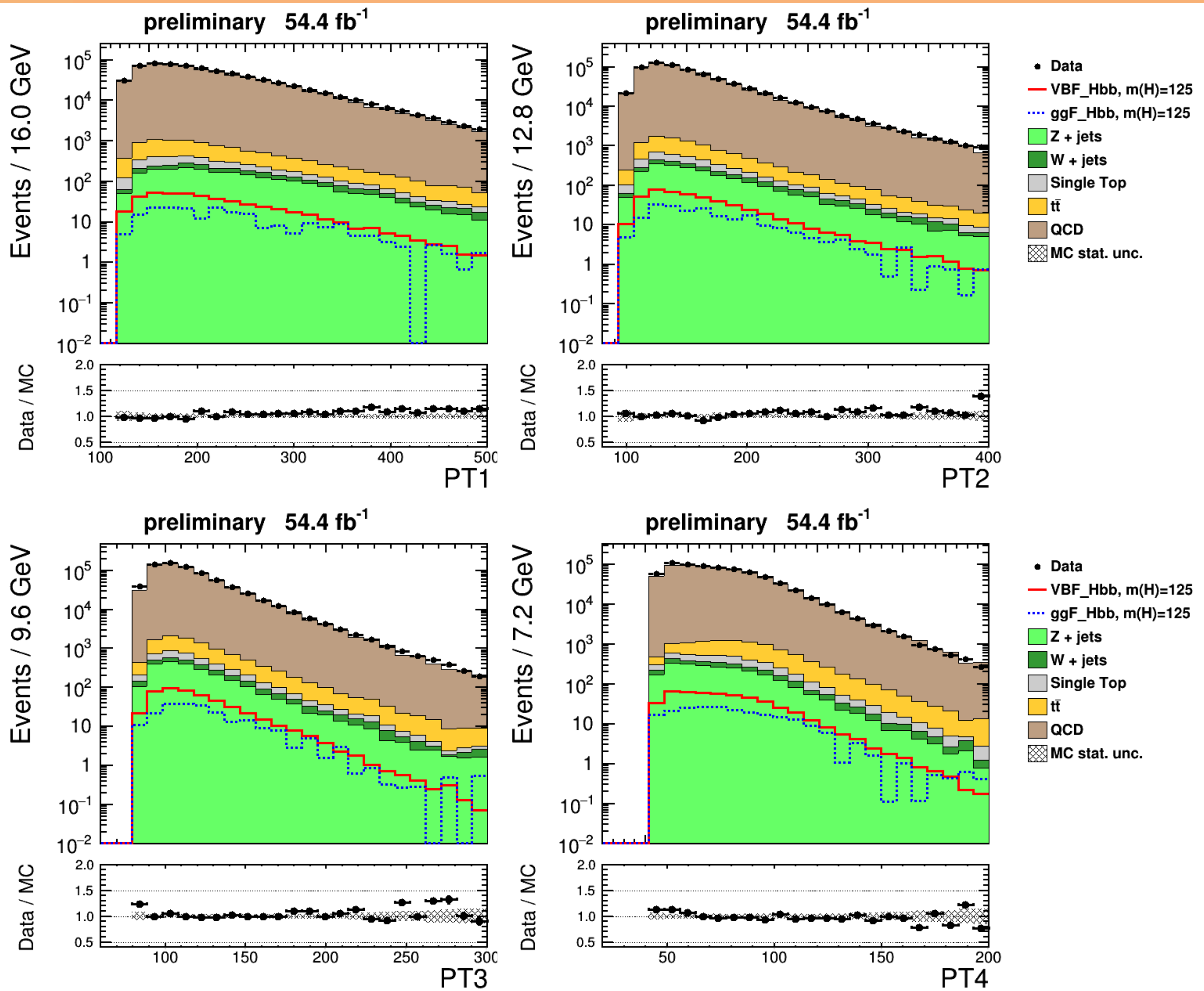
MC samples:

- **Signal**: VBF_Hbb, ggF_Hbb
- **Background**: QCD (>95%),   TTbar,   SingleTop,   Z+jets,   W+jets
- **Contribution of other signal or background processes are negligible**

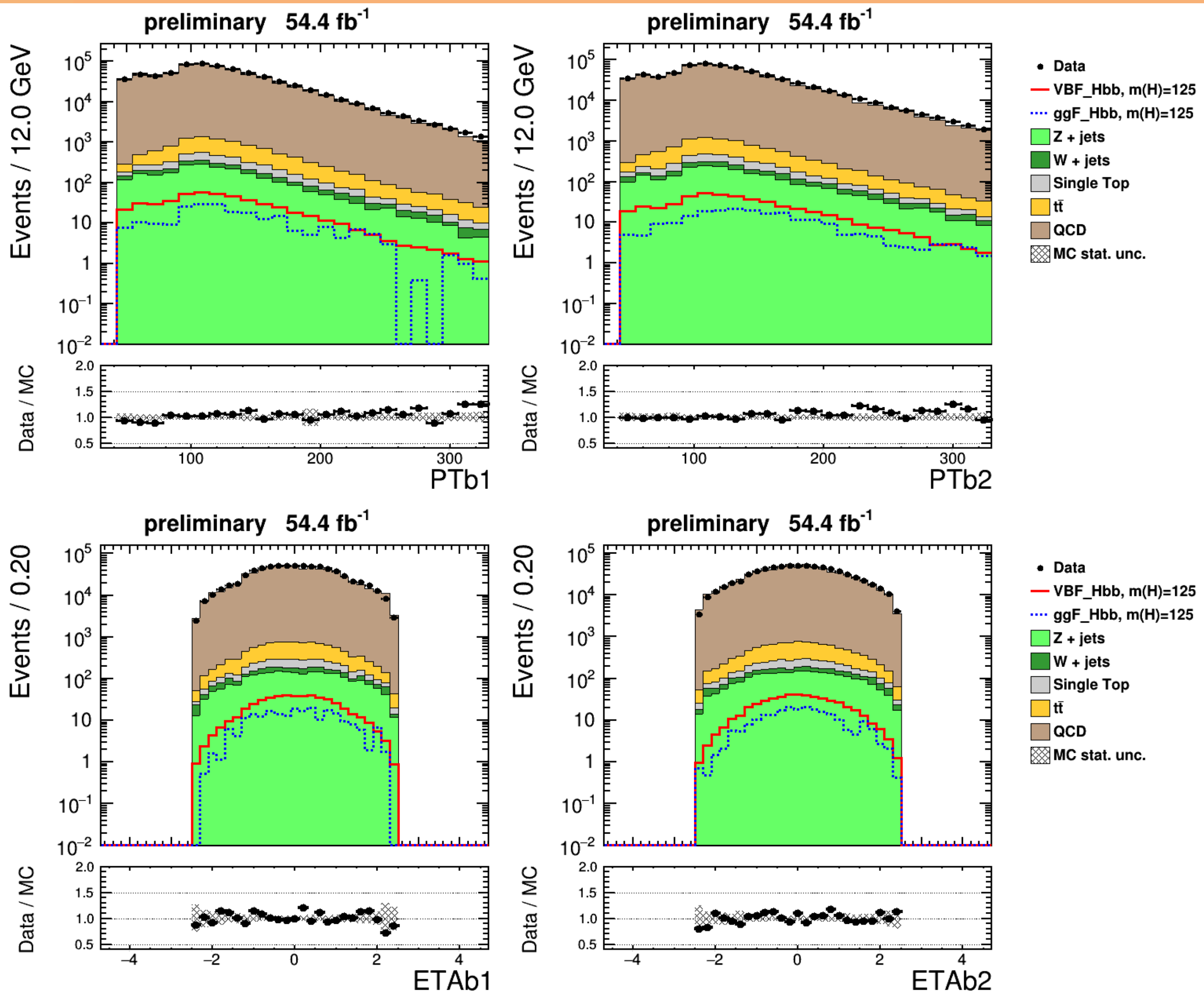
MC weights:

- Genweights and XSec normalization
- QCD normalization (k-factor) ~ **1.28** (**1.16**) for SingleB (DoubleB) selection
- PileUp reweighting, PileUpJetID SFs
- Trigger scale factors (SF) on P_T jets, online b-tagging SFs. online VBF-requirement SFs
- Offline b-tagging SFs

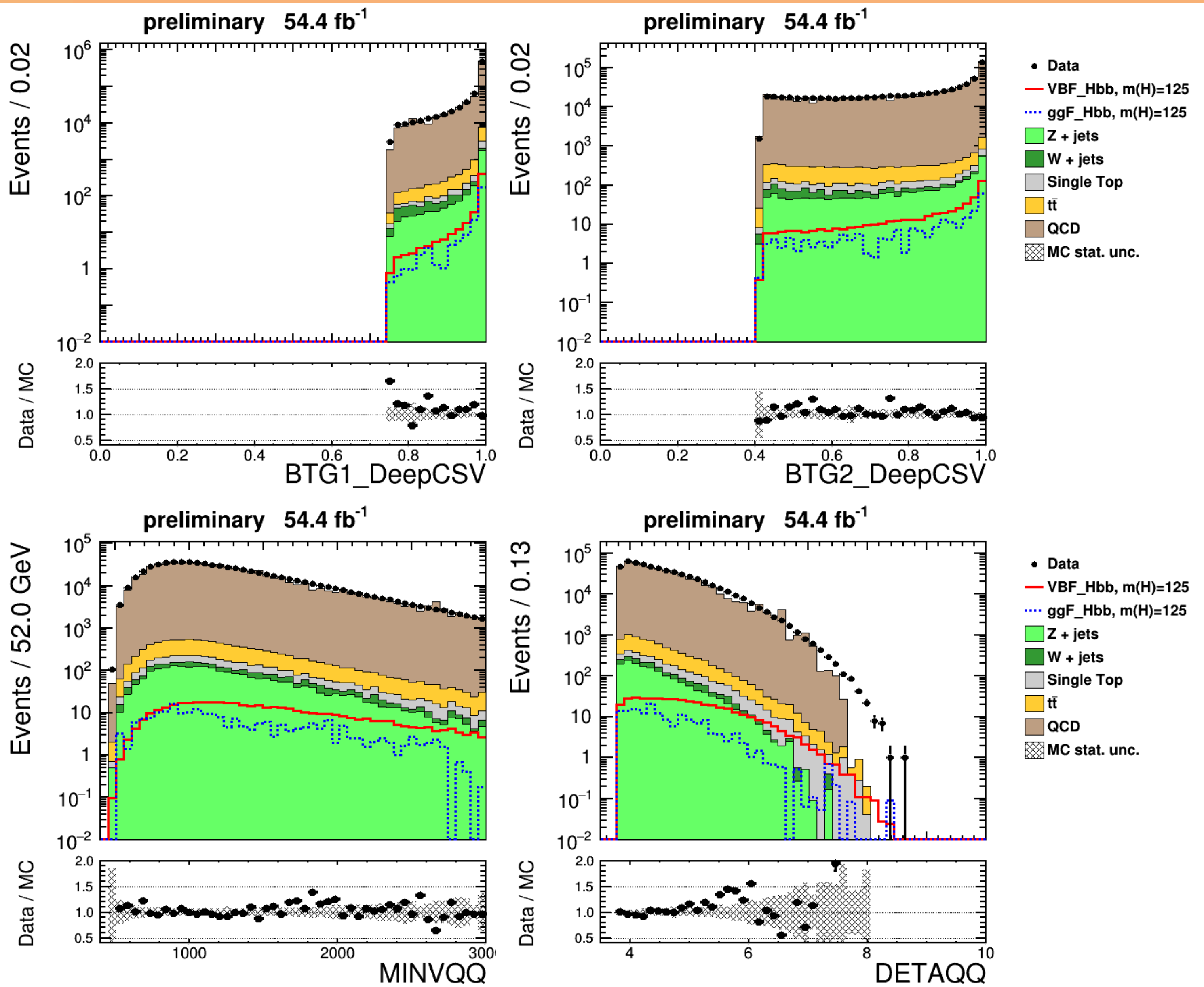
Data vs MC (SingleB)



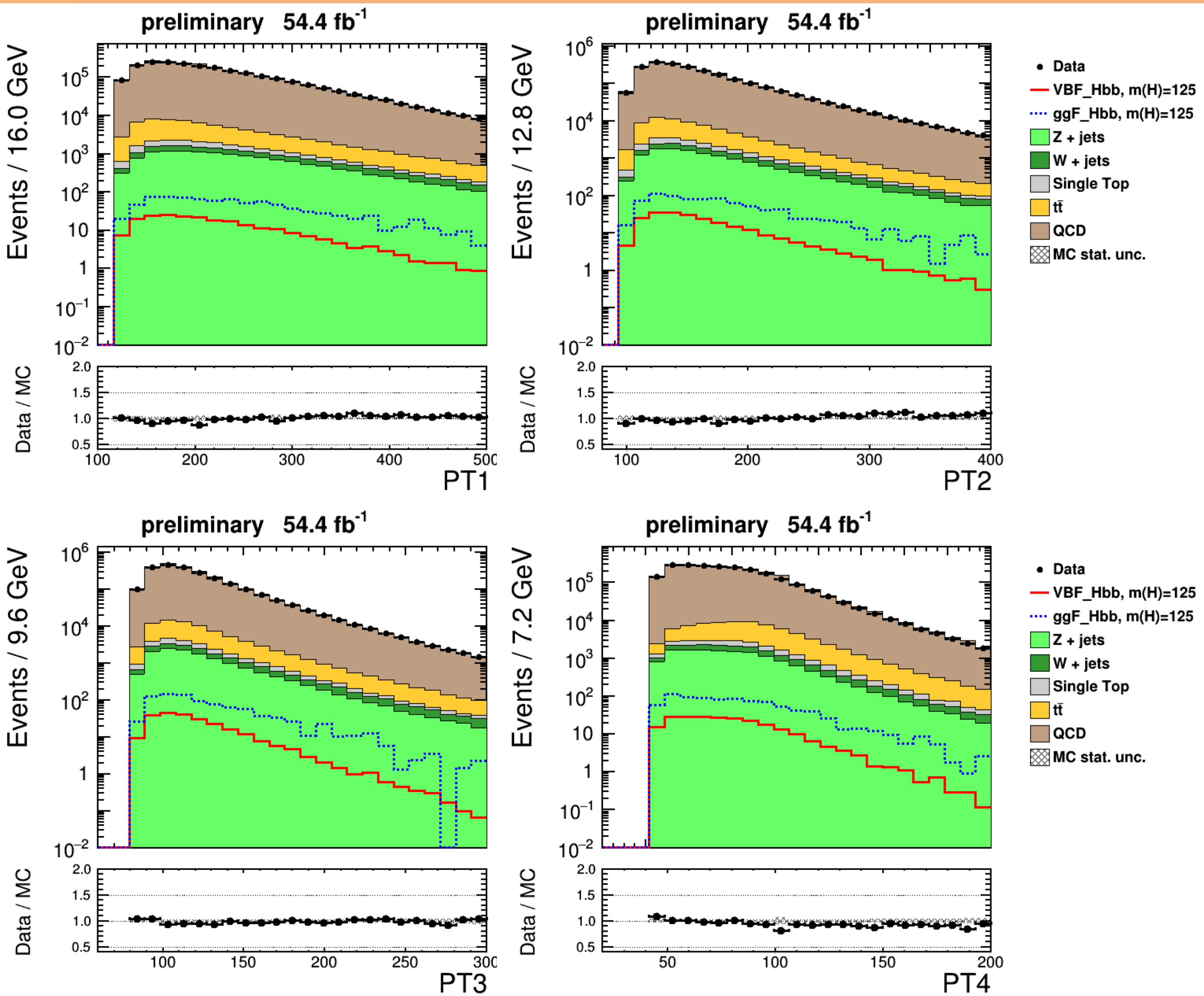
Data vs MC (SingleB)



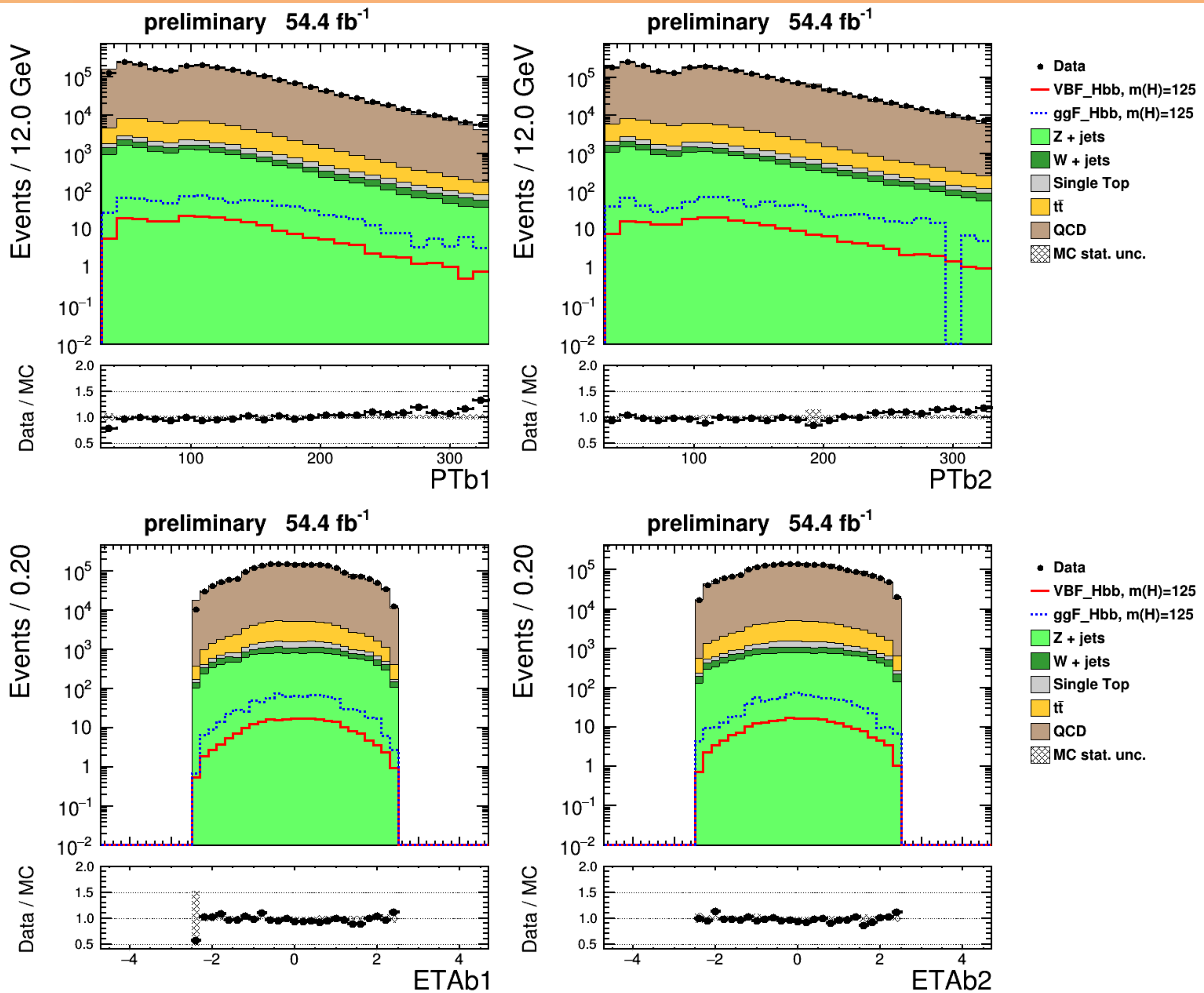
Data vs MC (SingleB)



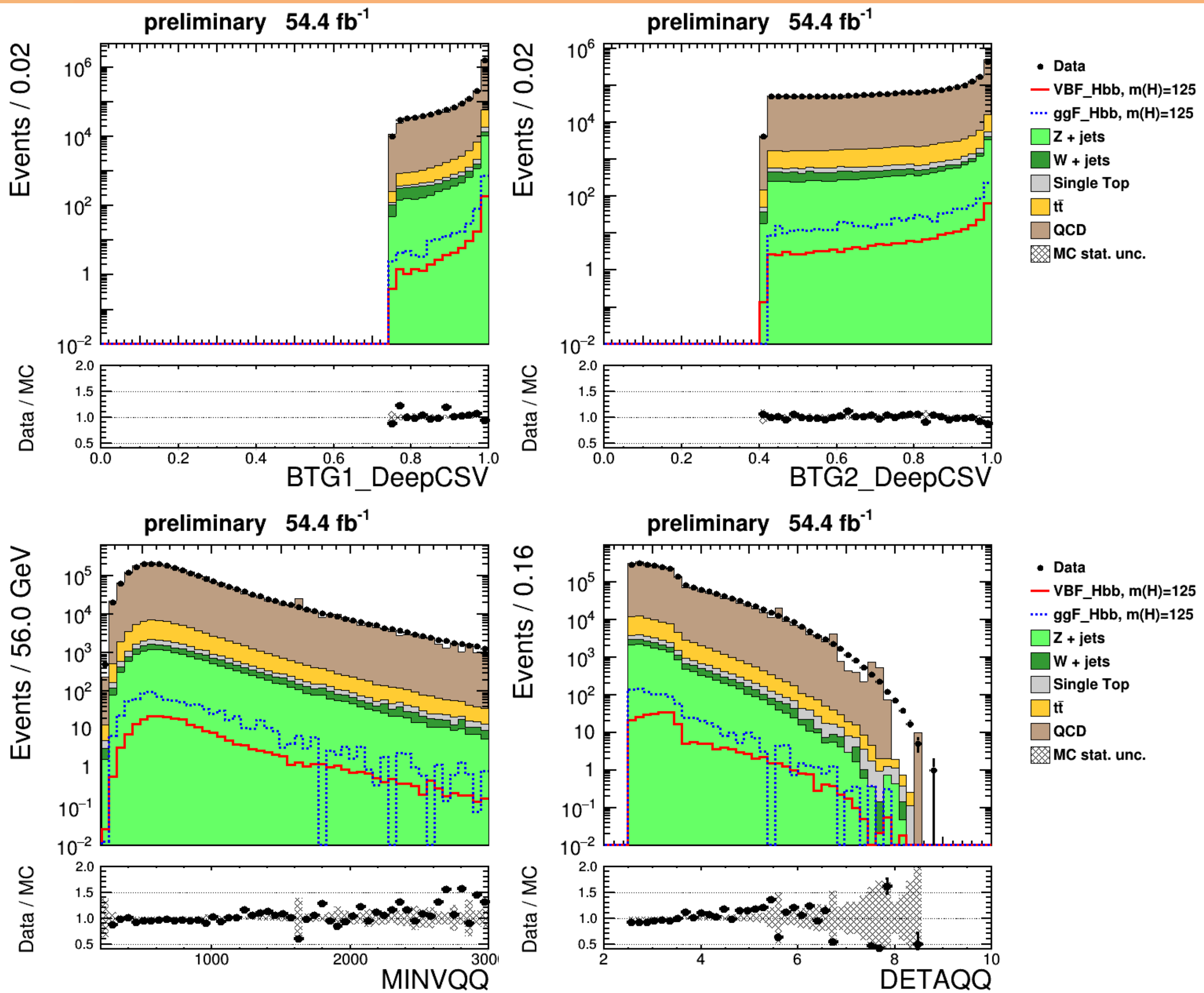
Data vs MC (DoubleB)



Data vs MC (DoubleB)



Data vs MC (DoubleB)



Signal-background discrimination

Even after all selections there is extremely large ratio of background to signal (10^3 - 10^4).

ML discrimination of signal vs background is the best way of weak signal extraction.

MVA Boosted decision trees method was used.

Signal: **VBF_Hbb (MC)**

Background: **5 % of 2018 data**

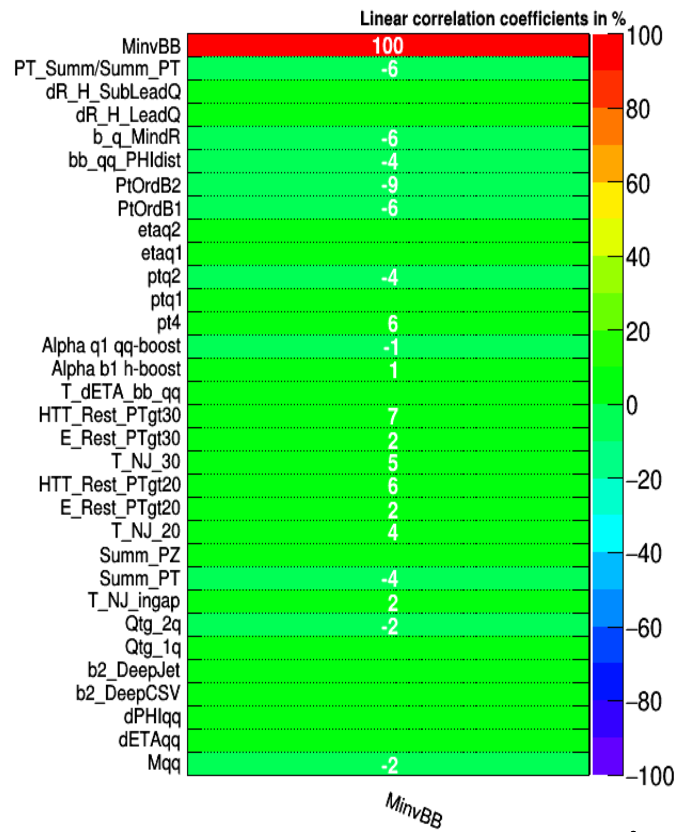
Input parameters:

- Kinematics of q-jets
- Correlation of q-jets
- Origin of q-jets
- q-jets and bb-system correlation
- Origin of b-jets
- Additional hadron activity
- Process momentum fractions
- bb-system (scalar vs vector)
- qq-system (nan vs vector)

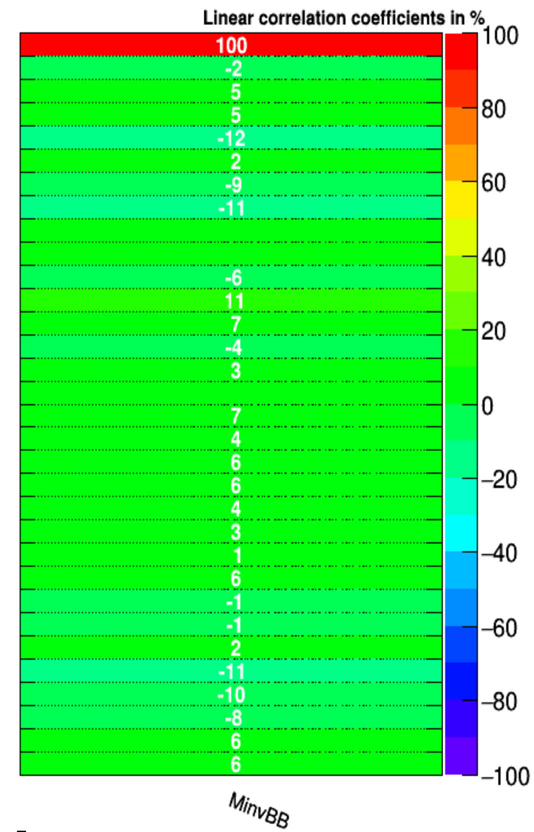
Requirement to input parameters:

- No Correlation with m_{bb}

Correlation Matrix (signal)

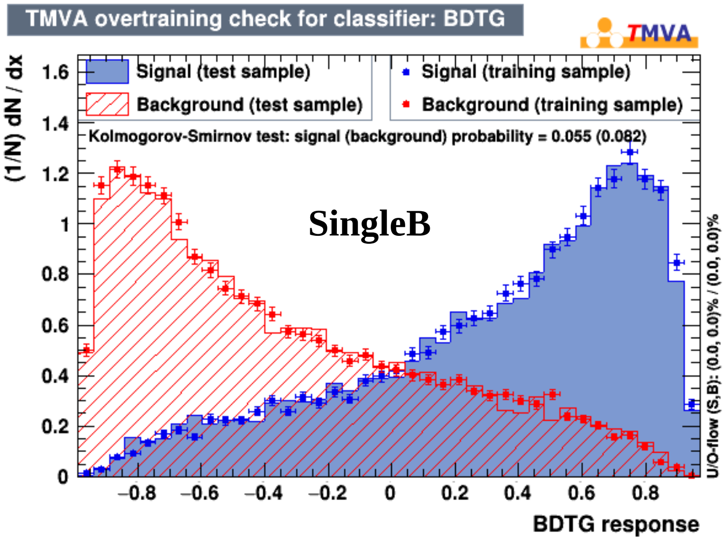


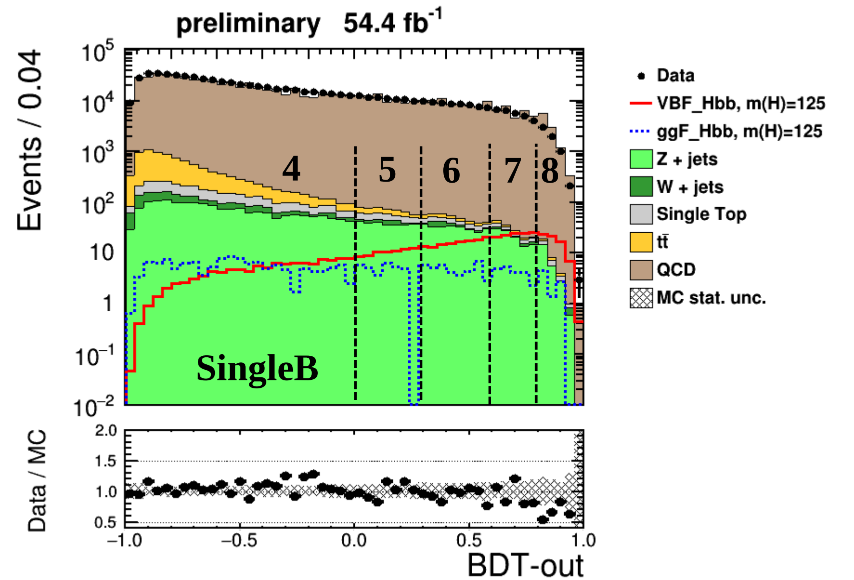
Matrix (background)

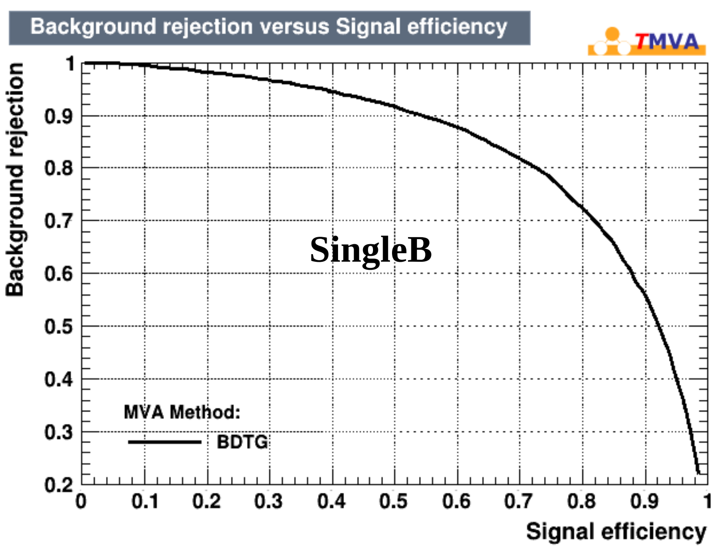


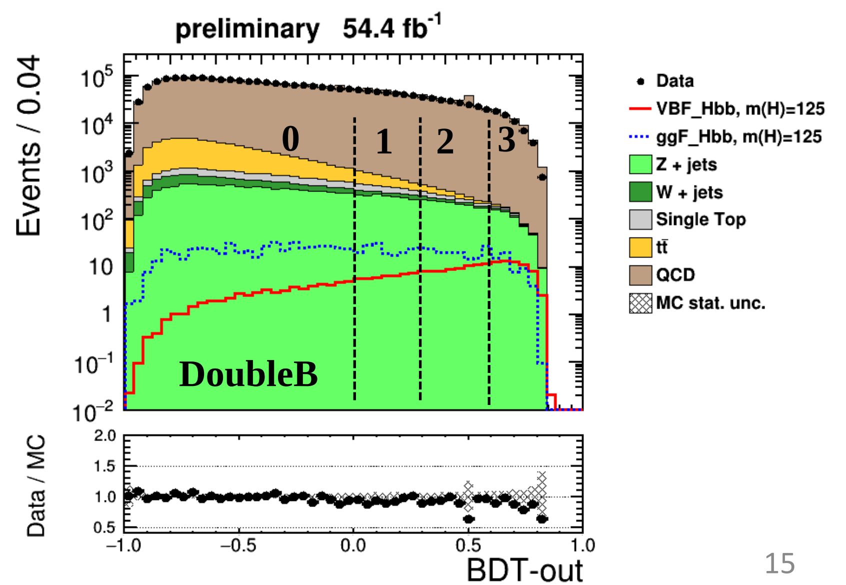
SingleB

Signal-background discrimination





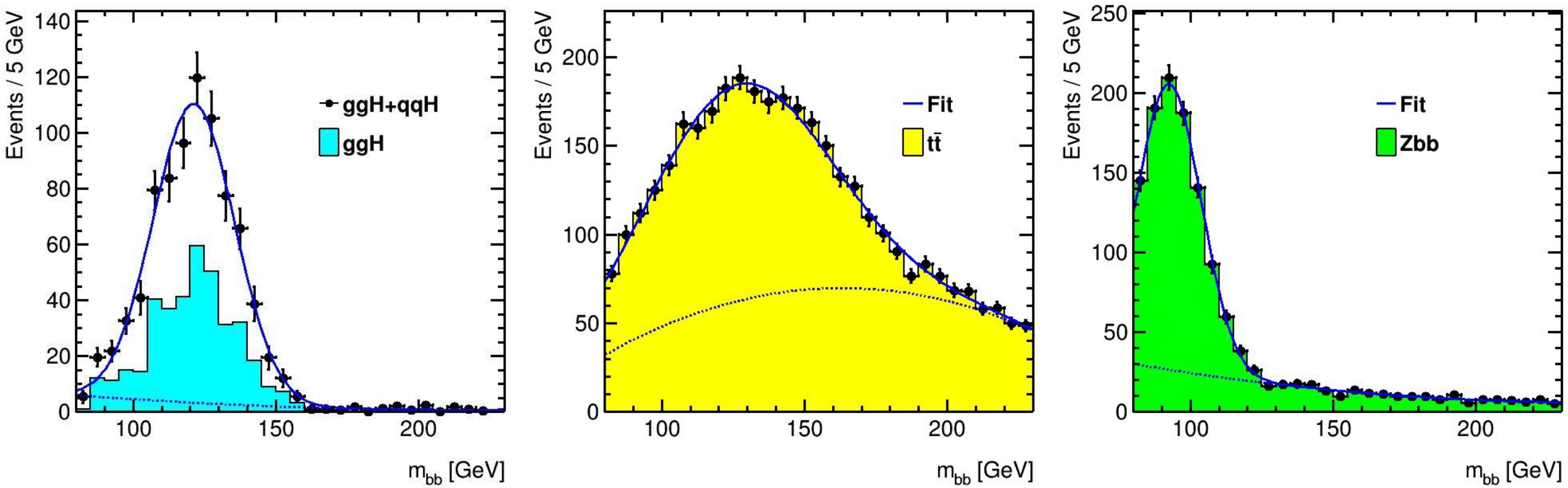




Signal and peaking backgrounds models

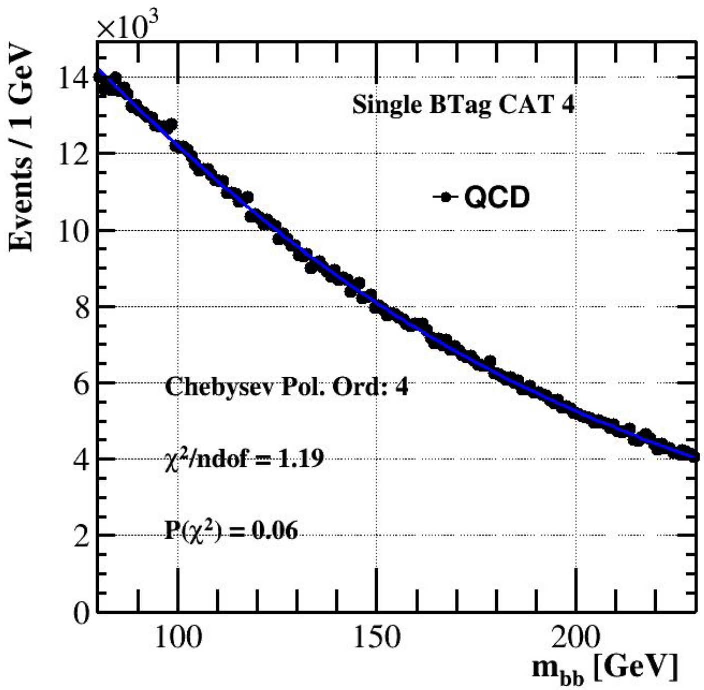
- Contribution of signal (VBF/ggF H→ bb) and peaking background processes (Z → bb, TTbar) are estimated from MC simulations.
- Shape of m_{bb} distribution modelled (fitted) by superposition of **Crystal Ball** function and **Bersntein polynomial** of 2^{nd} order.
- Fits are performed in **$80 < m_{bb} < 230$ GeV** region
- No significant dependence of pdf on BDT score: use same pdf for each category with different normalization value

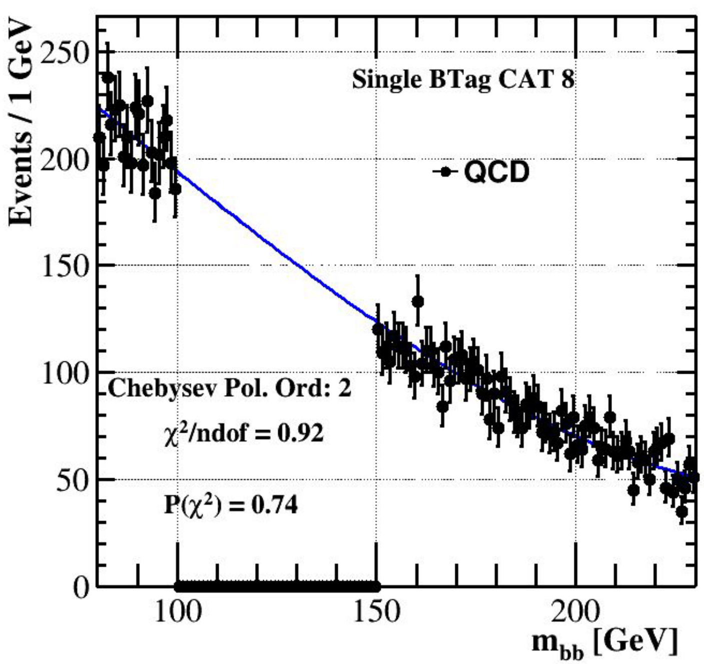
Combined by all categories



QCD model

- QCD modelling done directly from data separately in each category.
- Contribution from non-QCD processes (W+Jets, Z+Jets, TTbar, Single-Top, VBF H, ggF H) is estimated from simulation and subtracted from data.
- The m_{bb} spectra are fitted with Chebyshev polynomials, order selected by assessing χ^2/ ndof.





- Small (~ 5%) bias has been found.

Systematic uncertainties

Systematic uncertainties mainly affect the analysis in 2 ways

- **Shape uncertainties:** effects on m_{bb} distribution
- **Normalization uncertainties:** effects on cross sections, event yields

Systematic uncertainties

Machine unc.
- Luminosity.

Experimental unc.
- JES
- JER
- Trigger SF
- B-Tag
- Pileup (PU) Jet ID
- QGL re-weighting
- PU re-weighting

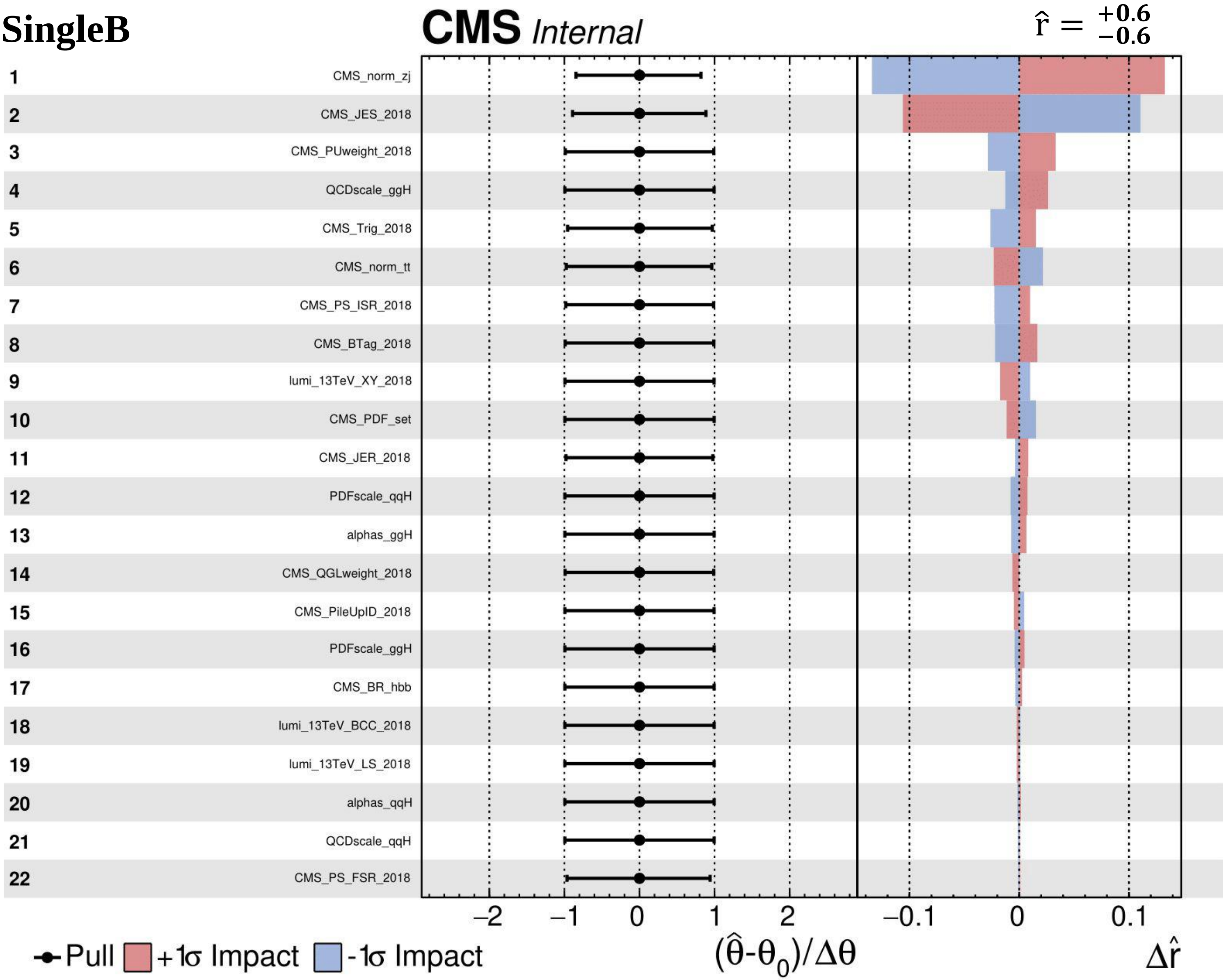
unc. in acceptance (Generator specific)
- Parton shower
  ISR
  FSR
- PDF set

Theory unc. in xsec
- PDF in qqH (VBF)
- QCD_scale in qqH
- α_S value in qqH
- PDF in ggH
- QCD scale in ggH
- α_S value in ggH
- BR ($H \rightarrow bb$)

- Additional uncertainty of 30% added for cross section of tt and Z+jets
- **Listed uncertainties do not affect modelling of QCD background (from data)**

Impacts of nuisance parameters on the signal strength

DoubleB + SingleB



Expected results and Likelihood scan

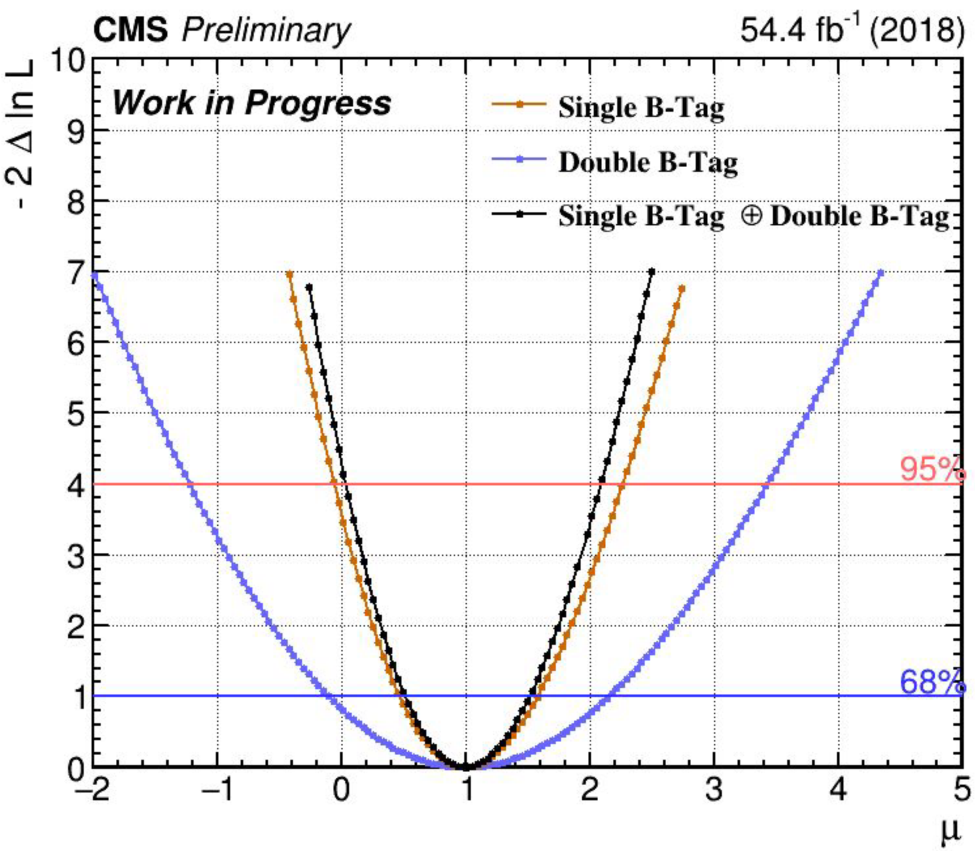
Expected significance

SingleB	
Stat only	Stat + Syst
1.67	**1.63**

DoubleB	
Stat only	Stat + Syst
0.63	**0.58**

Combined	
Stat only	Stat + Syst
1.78	**1.72**

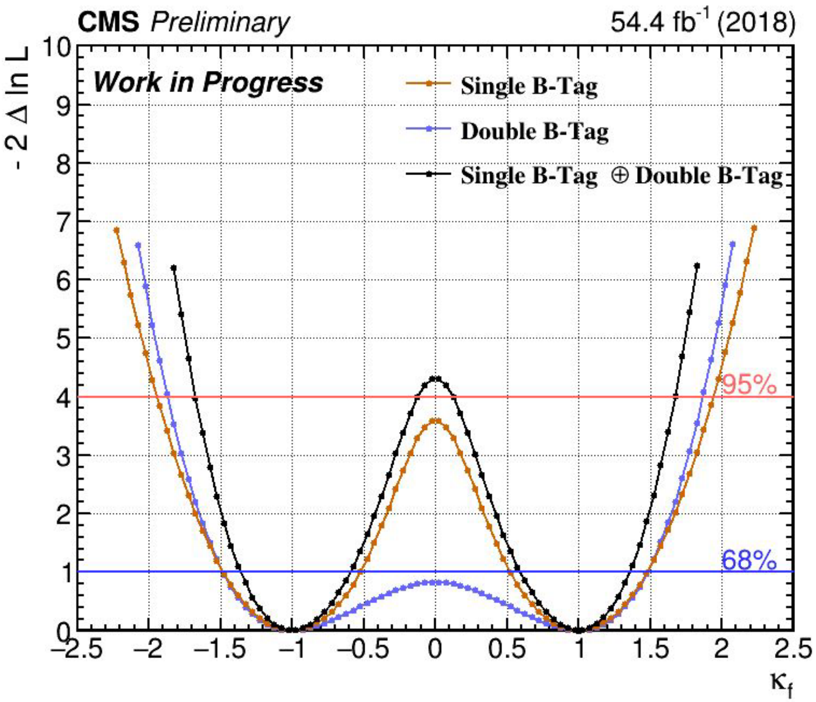
1D Likelihood scan of the signal strength (μ)



$$\mu = 1.0^{+0.61}_{-0.58} \quad \text{at } 68\% \text{ CL}$$

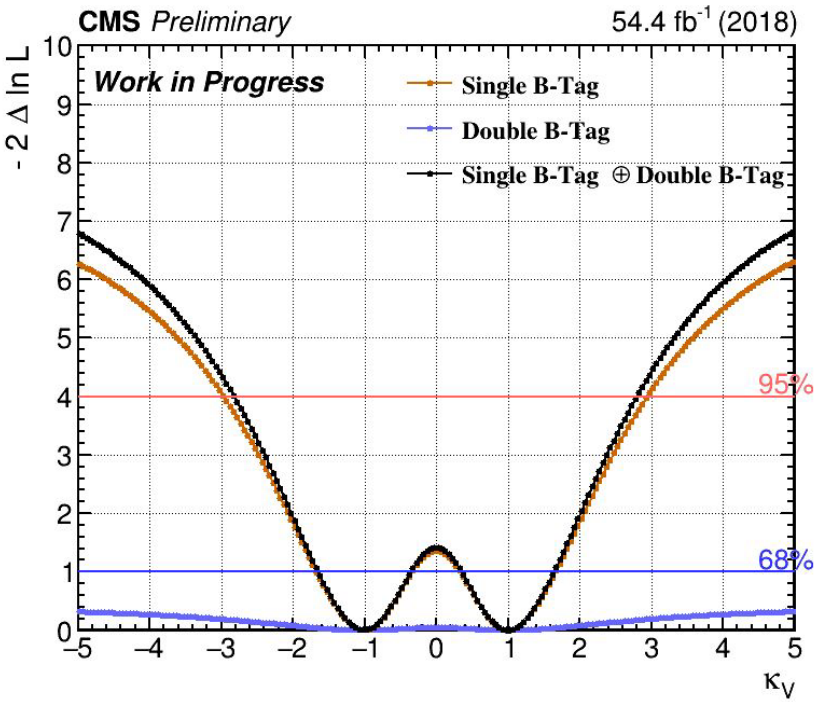
Couplings Likelihood scan

Hbb coupling modifier κ_f



$\kappa_f = 1.0^{+0.36}_{-0.48}$ at 68% CL

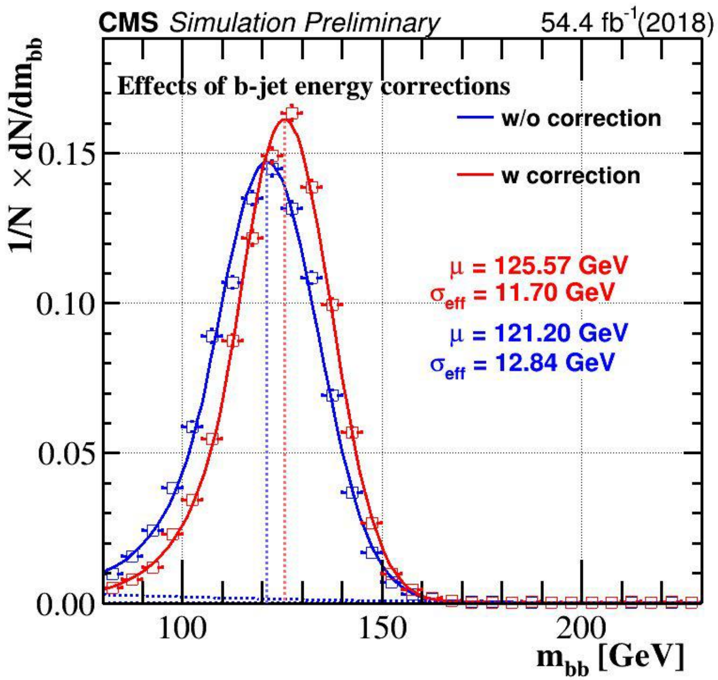
HVV coupling modifier κ_V



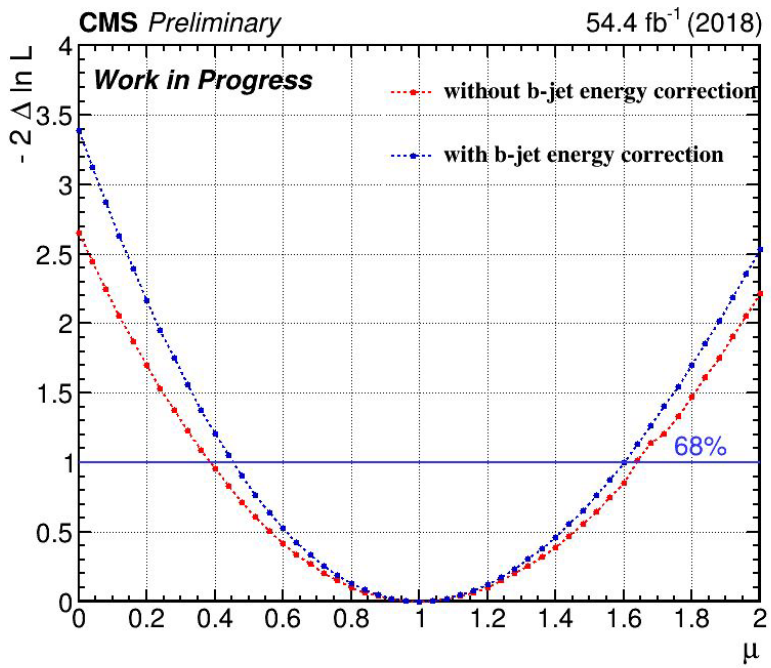
$\kappa_v = 1.0^{+0.65}_{-0.65}$ at 68% CL

b-jet energy regression

Additional correction on b-jet energy taking into account energy miscount due to neutrino in lepton decay modes of B-hadrons. DNN regression.



SingleB

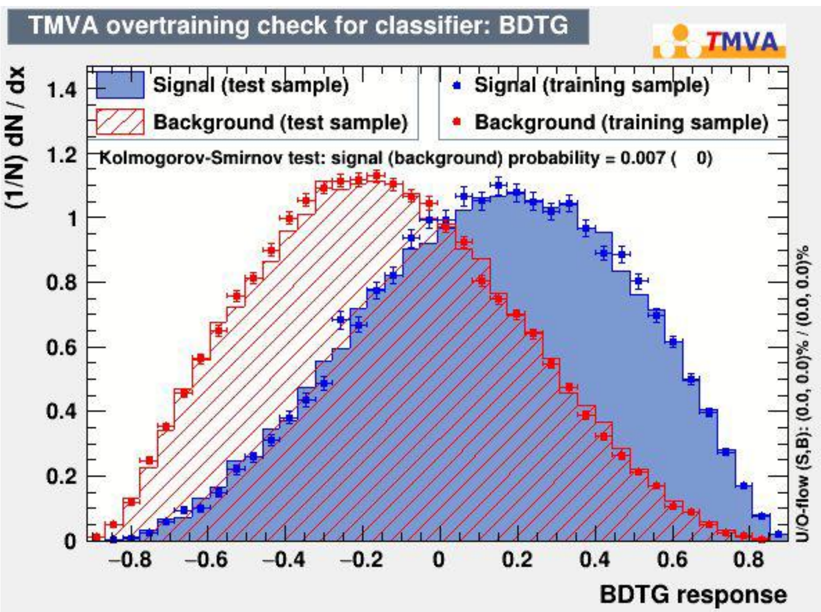


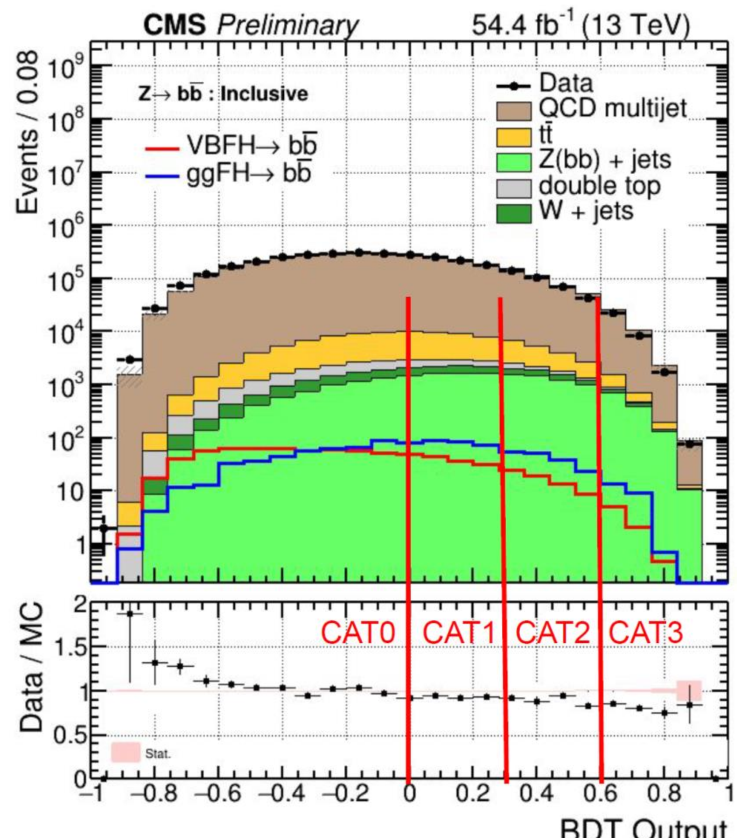
- **Peak of m_{bb} shifted closer to 125 GeV**
- **Relative resolution improved by 12%**

Result vs b-regression	Signal strength	Expected significance
Without	$1.0^{+0.64}_{-0.61}$	**1.63**
With	$1.0^{+0.60}_{-0.55}$	**1.84**

Measurement of Z→bb as standard candle

- Same events selection as inclusive VBF H→bb
- Same BDT inputs for signal vs background discrimination as in VBF H→bb
- Combined fit of the m_{bb} spectra in all event categories as in VBF H→bb
- Same approach of signal and background modelling as in VBF H→bb
- m_{bb} distribution (for signal and peaking backgrounds) fitting performed by combination of Crystal Ball (CB) function and Bernstein polynomial of 2nd order in **$60 < m_{bb} < 210$ GeV**

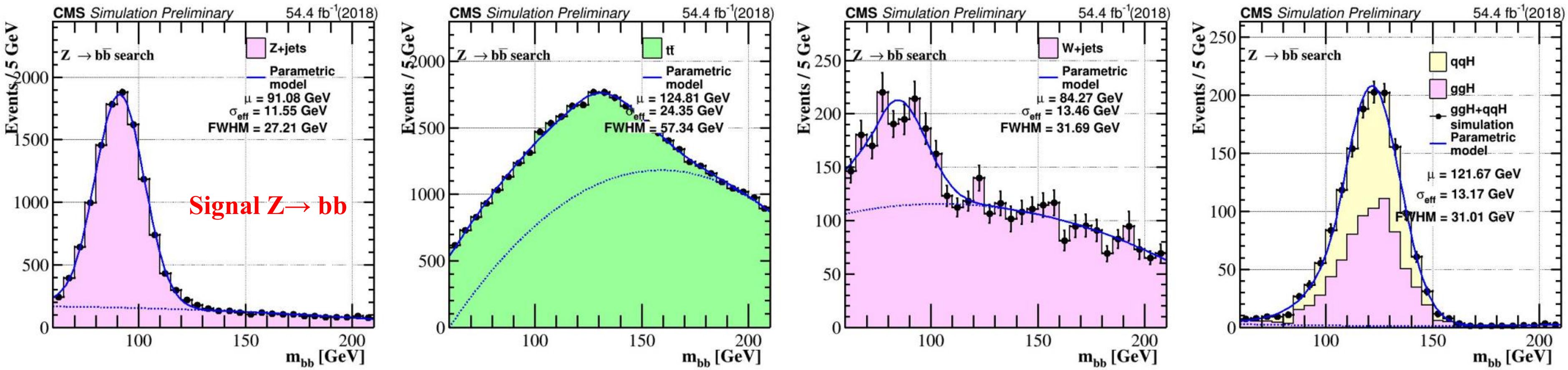




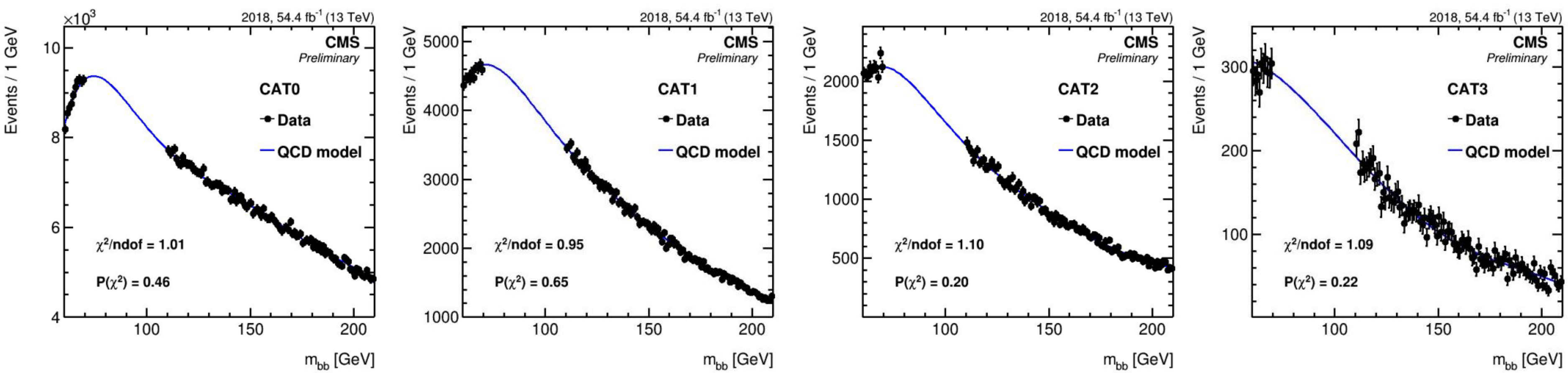
- **Signal: MC Z→bb**
- **Background: 5% of 2018 data**

Measurement of Z→bb as standard candle

Signal and peaking background models

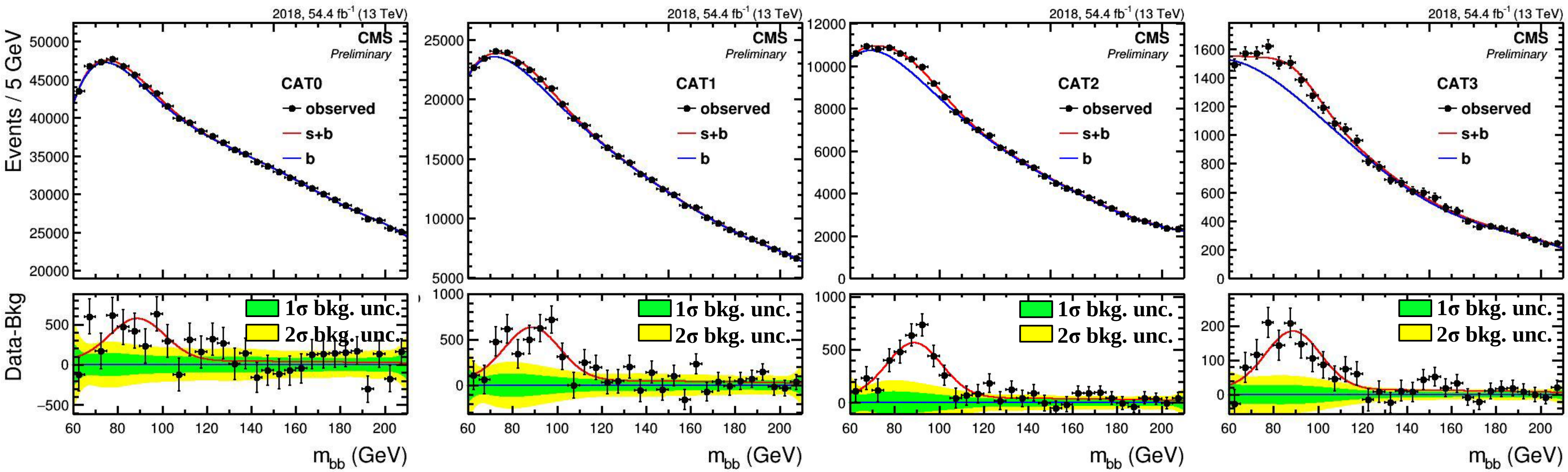


QCD model

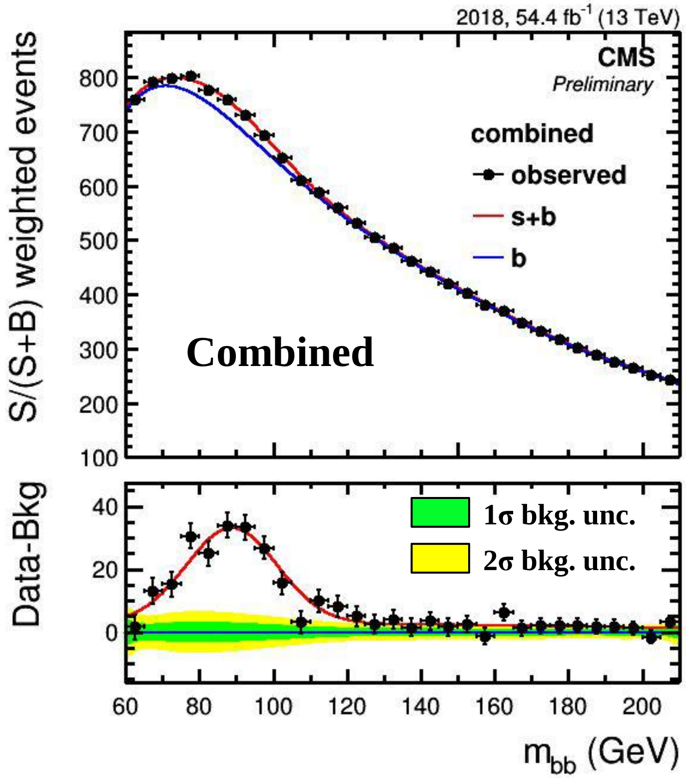


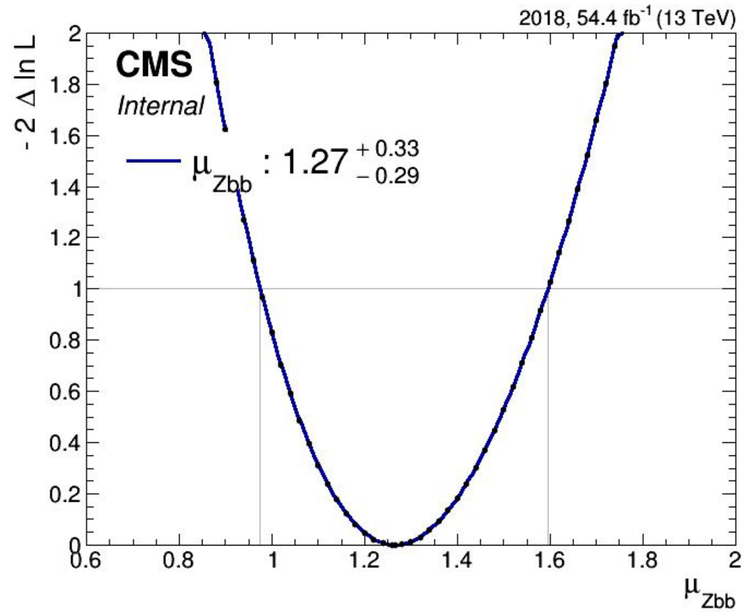
Measurement of Z→bb as standard candle

Results of fit in each BDT-category.

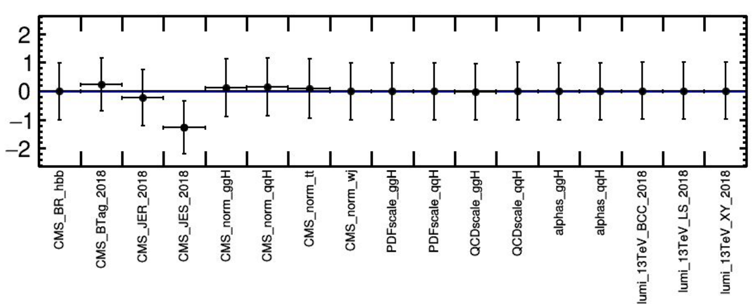


Measurement of Z→bb as standard candle





postfit nuisance parameters (systematics)



Z→bb signal significance in std. dev.

Expected	**4.77**
Observed	**4.99**

ongoing / upcoming

- b-jet energy regression in DoubleB
- Analysis Note: CMS AN-2021/045
- Recalculation with UL-campaign
- Preapproval
- Unblind analysis
- Approval
- Analysis with 2016 data and combining
- Publication

Thank you

backup

Bias study

SingleB

- ❖ Signal contribution is very small compared to background.
- ❖ Bias in background modelling may substantially affect extracted signal strength
- ❖ Bias test is important to test the convergence the background modelling

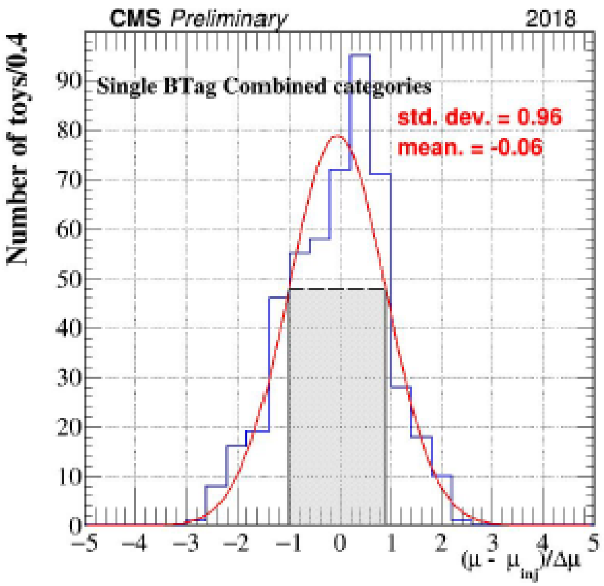
General approach:

→ Generate toy (~ 500) using nominal QCD background modelling with a definite injected signal strength (μ_{inj})

→ Fit with alternative background functions and determine:
(i) Fitted signal strength (μ_{fit})
(ii) Fit uncertainty ($\sigma_{up/down}$)

→ Finally bias is calculated as

$$B = \frac{\mu_{inj} - \mu_{fit}}{0.5.(\sigma_{fit}^{up} + \sigma_{fit}^{down})}$$



→ Bias has been measured for combination and each individual categories

→ Negligible bias has been found
→ 5-10% depending on category

Alternative model:
CAT-4 : Chebyshev pol. of 5th order
CAT-5 - 8 : Chebyshev pol. of 3rd order