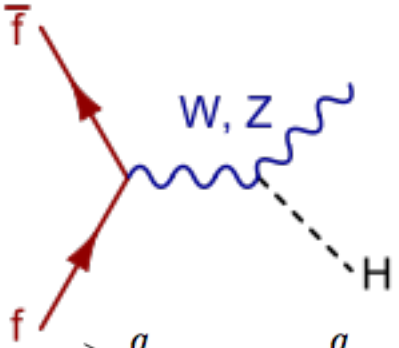
Searches for the Higgs Boson Decaying to Bottom Quarks in the CMS Detector

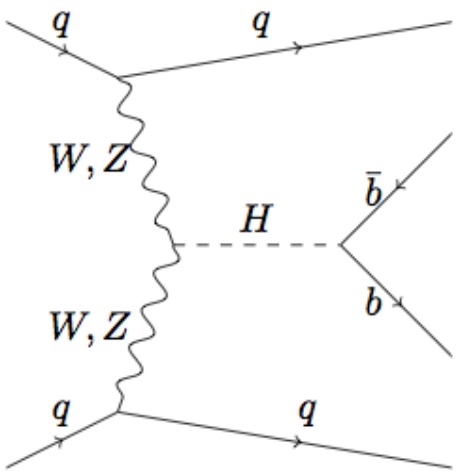
Seth Zenz
Princeton University
On Behalf of the CMS Collaboration

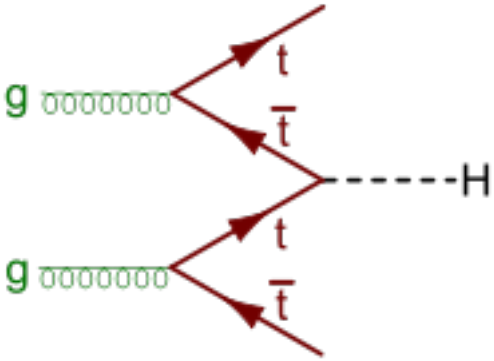
Higgs Hunting 2013
Orsay, France
26 July 2013

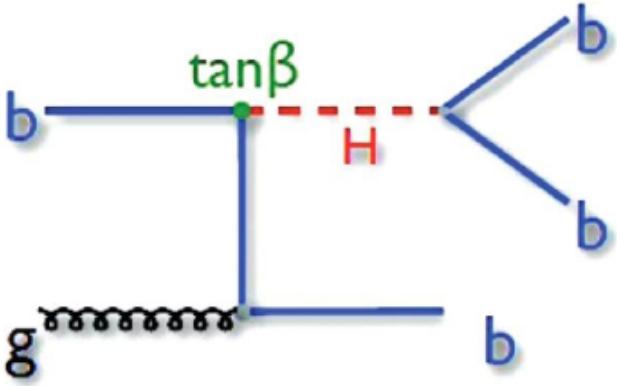
Overview

- H → bb in the Standard Model
  - VH → Vbb (5+19 fb^{-1})
  - VBF H → bb (19 fb^{-1})
  - ttH → ttbb (5 + New 19 fb^{-1})
- … And Beyond
  - bH+X, H → bb (2.7 - 4.8 fb^{-1})

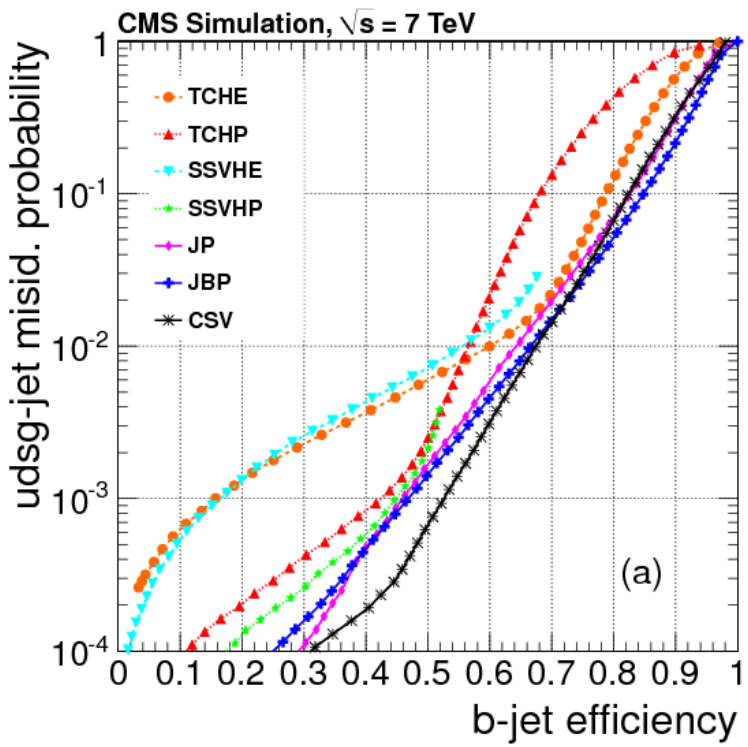


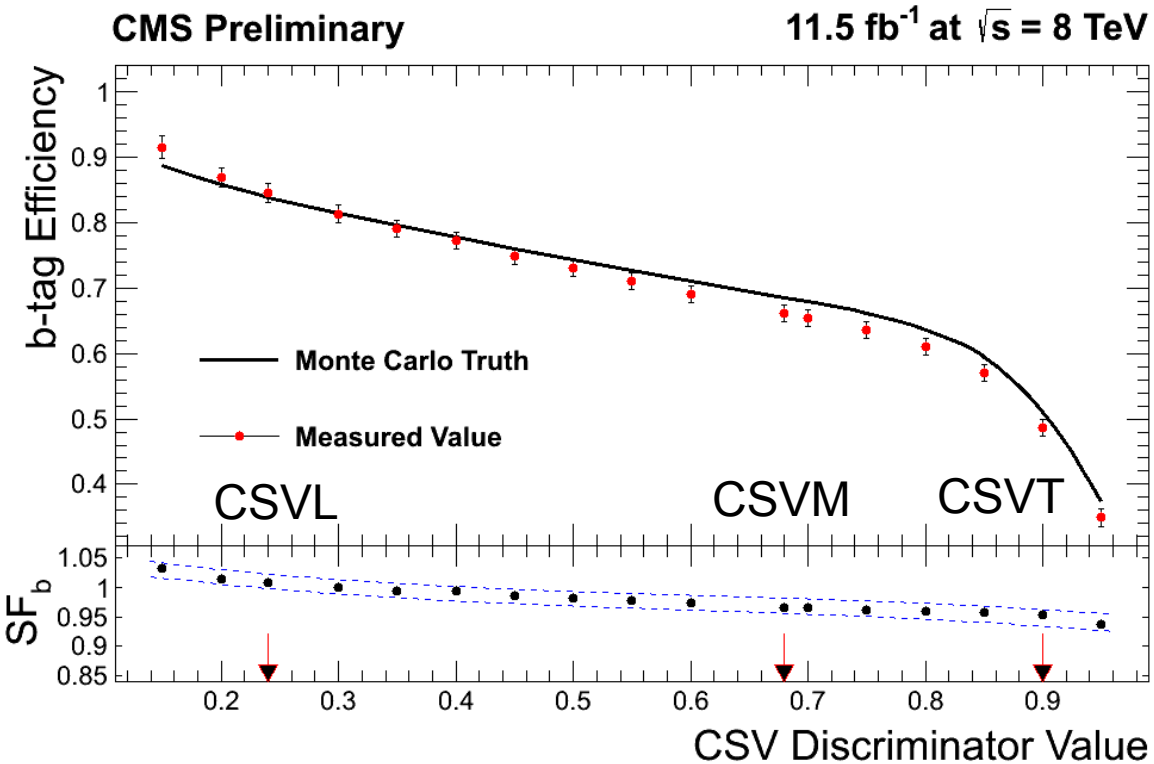






B-Tagging in CMS

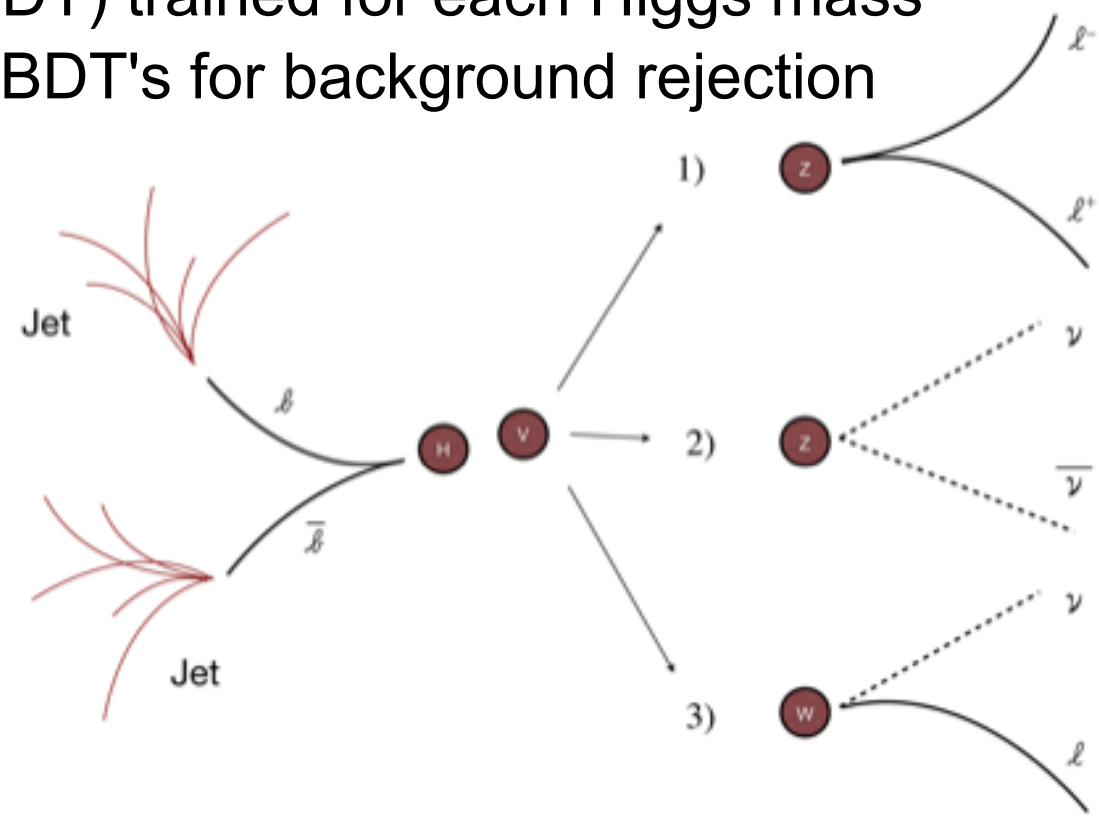




- Jets: particle flow inputs to anti-kT algorithm, R = 0.5
- B-tagging: Combined Secondary Vertex (CSV) discriminant
  - Explicitly reconstructed secondary vertex information
  - “Pseudovertex”: tracks with high impact parameter significance
  - Individual track IP's, jet and track kinematics

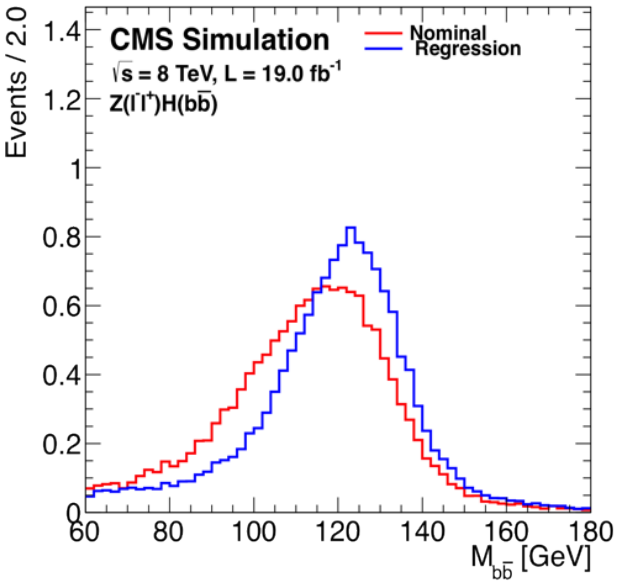
Analysis Strategy: VH → Vbb

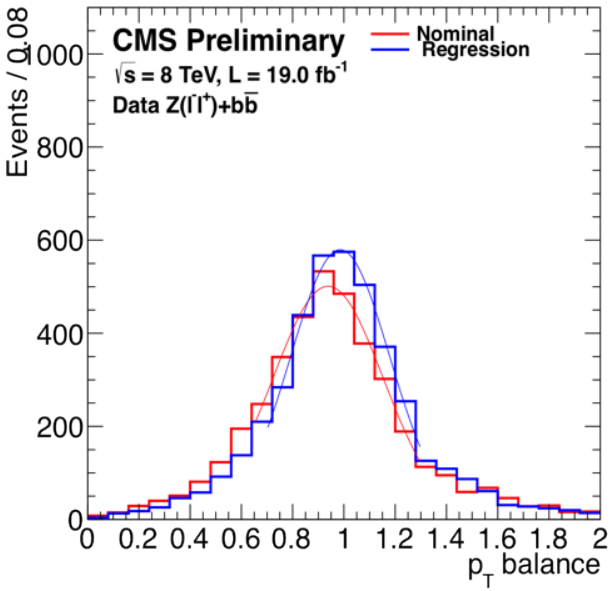
- Categorize by vector boson (V) decay: Z(νν), Z(ℓℓ), W(ℓν), W(τν)
- Require 2 b-tagged jets, leptons & MET as appropriate
- Bin in pT(V): signal-to-background increases at high pT
- Multivariate b-jet regression to improve bb mass resolution
- Boosted Decision Tree (BDT) trained for each Higgs mass
  - Additional specialized BDT's for background rejection
- Likelihood fit to BDT distributions
- Backgrounds: tt, V+bb, V+b, V+udscg, QCD, single top, VZ
- Dominant systematics: b-tagging, MC statistics, data-driven background normalization



B-Jet Regression

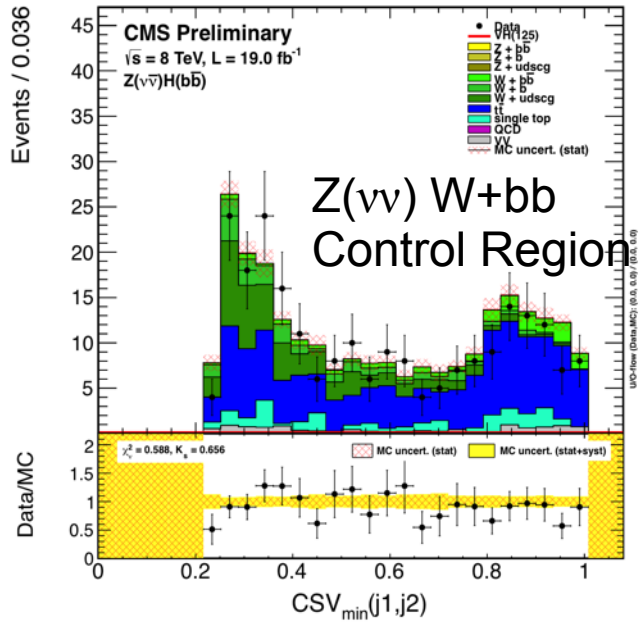
- Additional correction of b jet to parton energy, using BDT, to improve mass resolution
- Trained on simulated H → bb events
- Variables used:
  - Jet kinematics
  - Charged/neutral fraction, lead track pT, etc.
  - Vertex properties
  - Soft lepton properties
  - MET – Z(ll) only
- Improves Higgs mass resolution by ~15%
  → analysis sensitivity improved by 10-15%

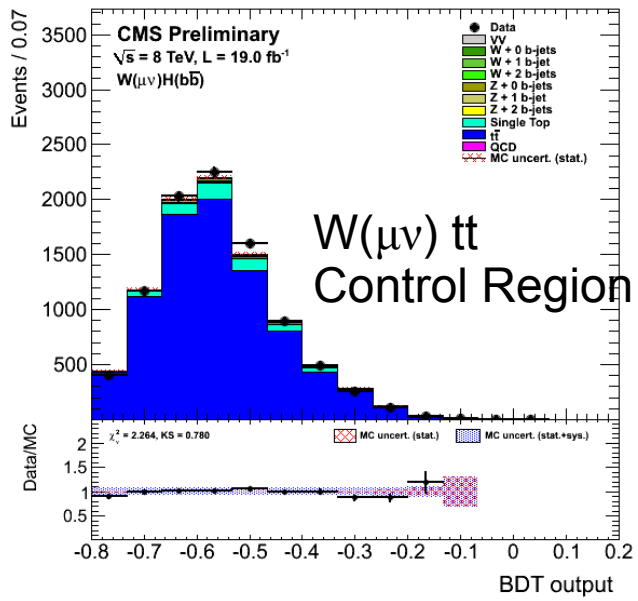


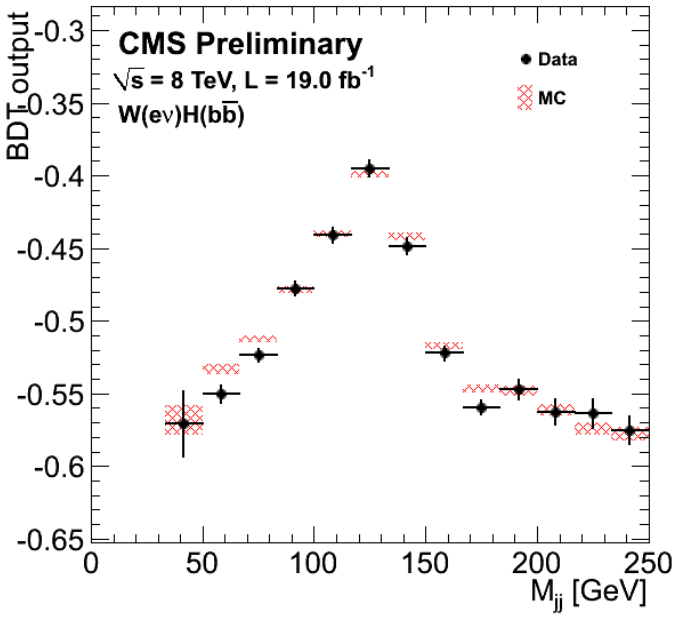


Boosted Decision Tree

- Trained for each channel and each Higgs mass
- Inputs
  - Higgs candidate jet pT and CSV
  - Dijet properties: ***m(jj)***, pT(jj), dEta(jj), dR(jj), color pull angle
  - Additional jets: Naj, max CSV, closest dR to Higgs candidate
  - dPhi(V,H), dPhi(E_T^{miss},j)
  - Angular variables of VH system

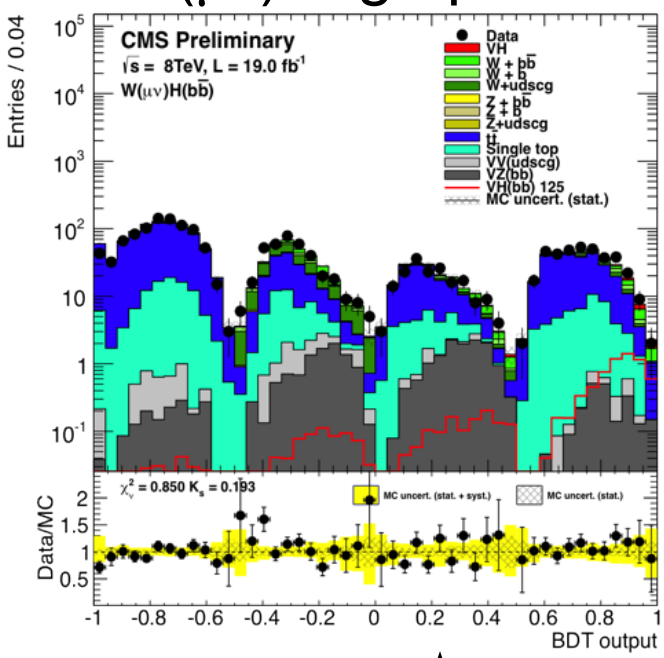




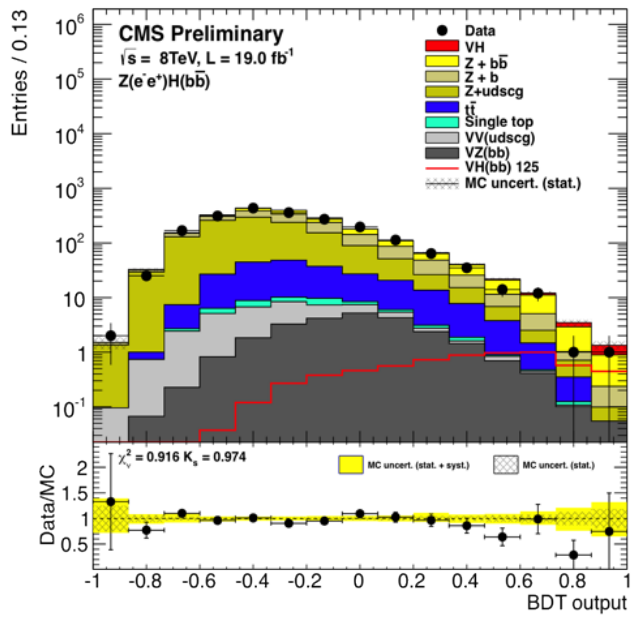


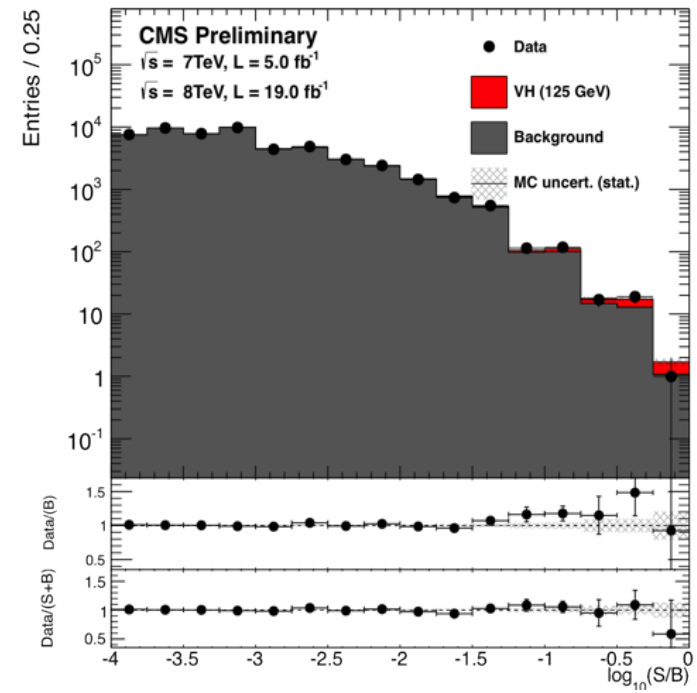
Signal Extraction

W(μν) High pT



Z(ee) High pT





"Fails" anti-tt MVA

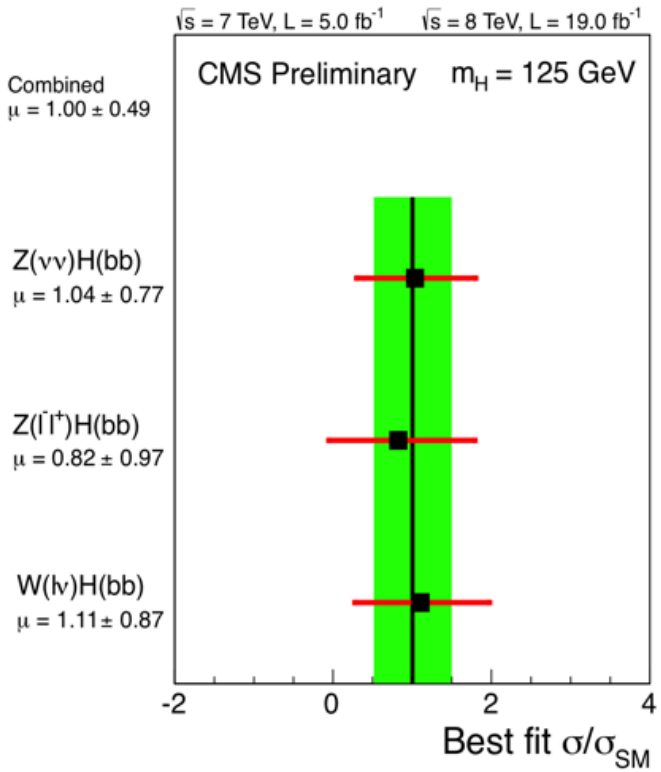
"Fails" anti-V+jets MVA

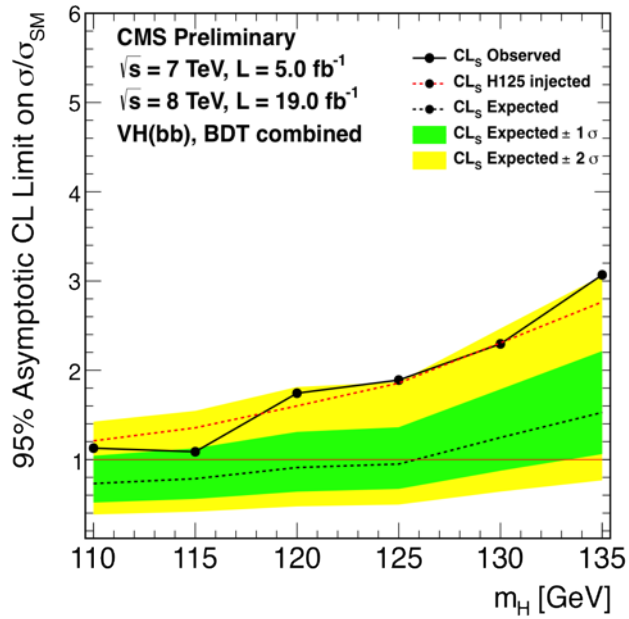
"Fails" Anti-VV MVA

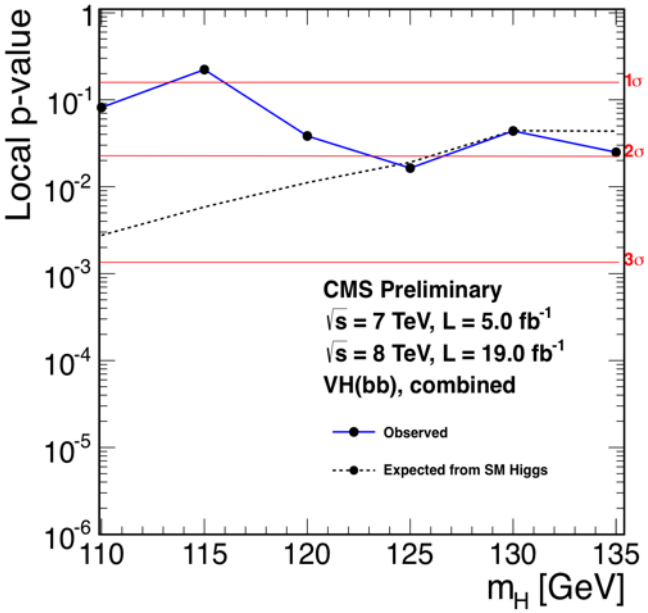
"Passes" all Background-Specific MVA's

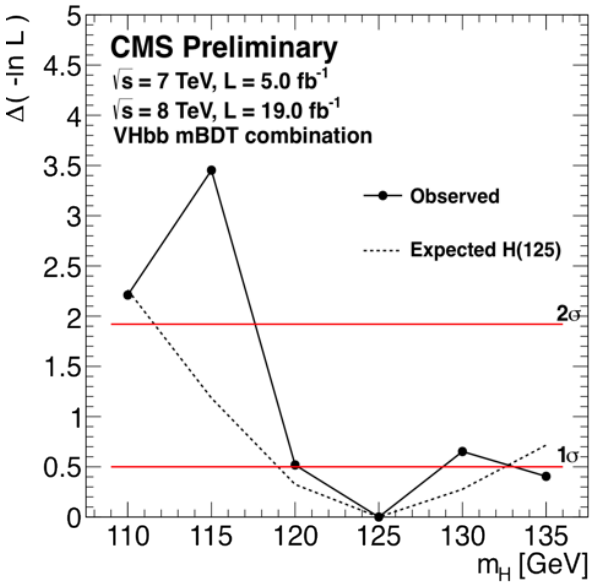
- Simultaneous fit in all 14 signal regions
  - Z(νν): 3, Z(ℓℓ): 4, W(ℓν): 6, W(τν): 1
- Visualize fit by summing bins by log(S/B)

VH → Vbb Results





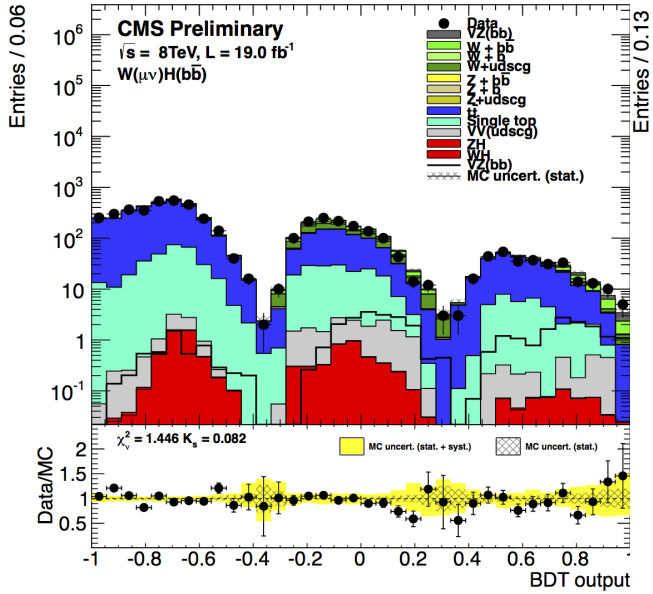


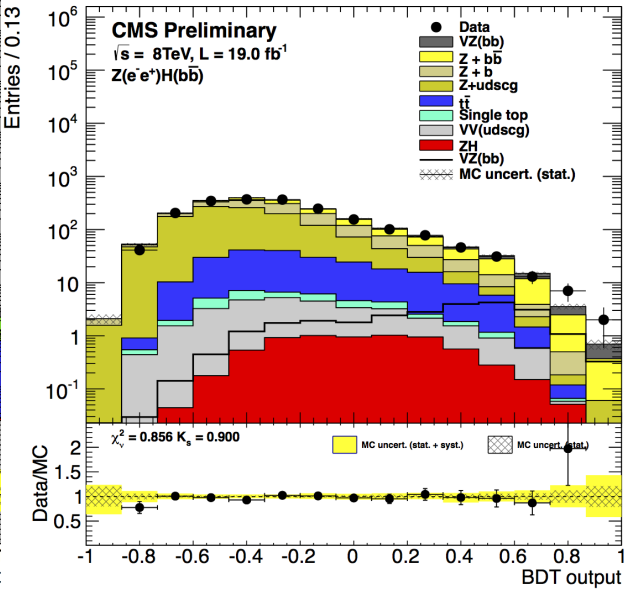


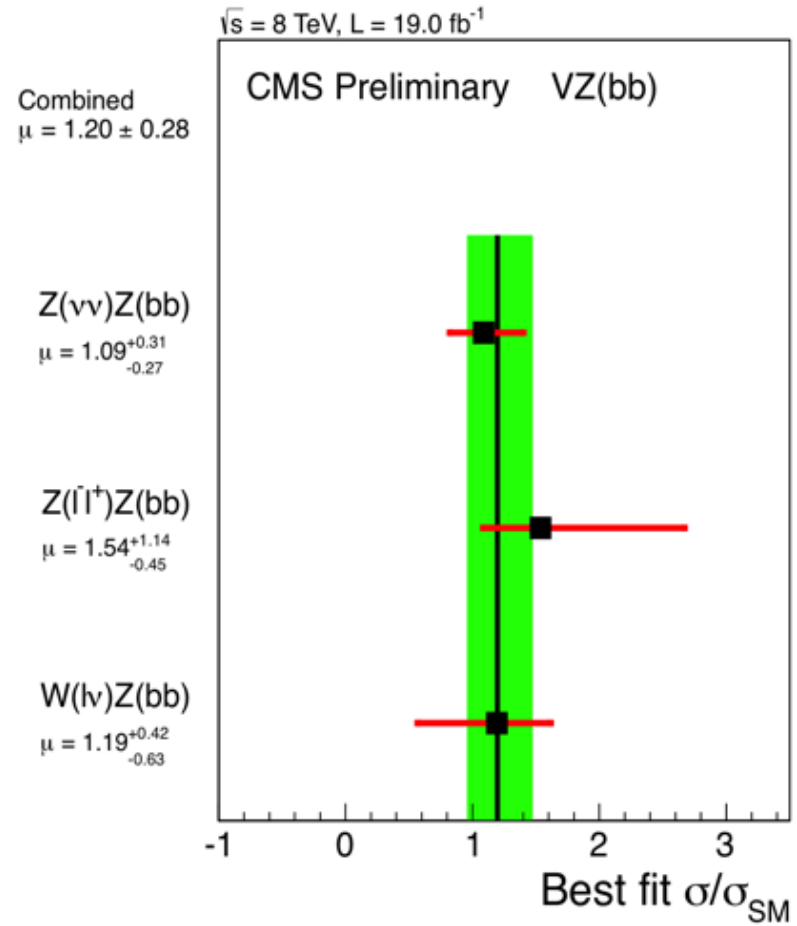
Results of BDT likelihood fit for 125 GeV Higgs:

- ***$\mu = \sigma/\sigma_{SM} = 1.00 \pm 0.49$***
- Observed (expected) 95% CL limit: 0.95 (1.89)
- Observed (expected) significance: ***2.1σ***
- In combination with H → ττ: ***3.4σ***

Validation with VZ → Vbb



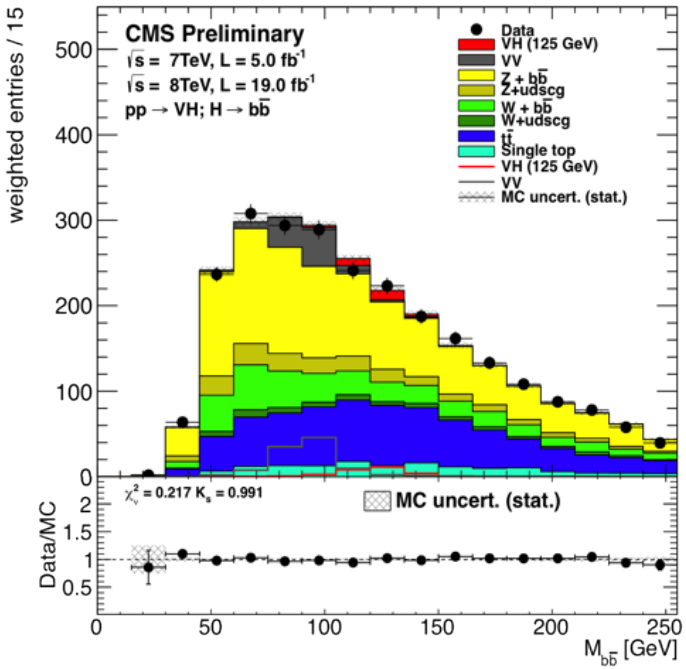


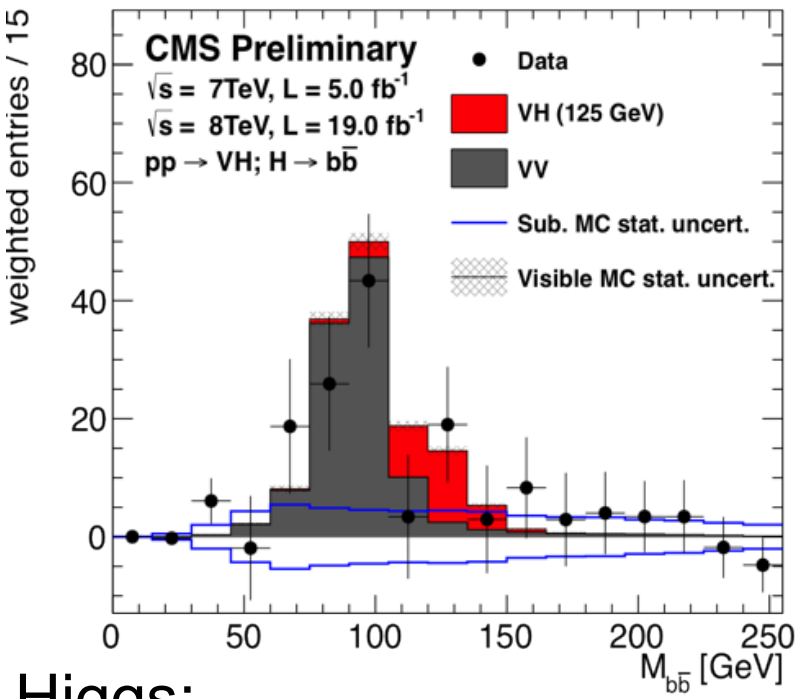


- Expected: 6.3σ
- Observed: 7.5σ
- Best fit μ: 1.20 ± 0.28
- Method gives correct SM result!

VH → Vbb Mass Analysis

- Tighten selection of BDT input variables
- To visualize: weight each category by S/(S+B)

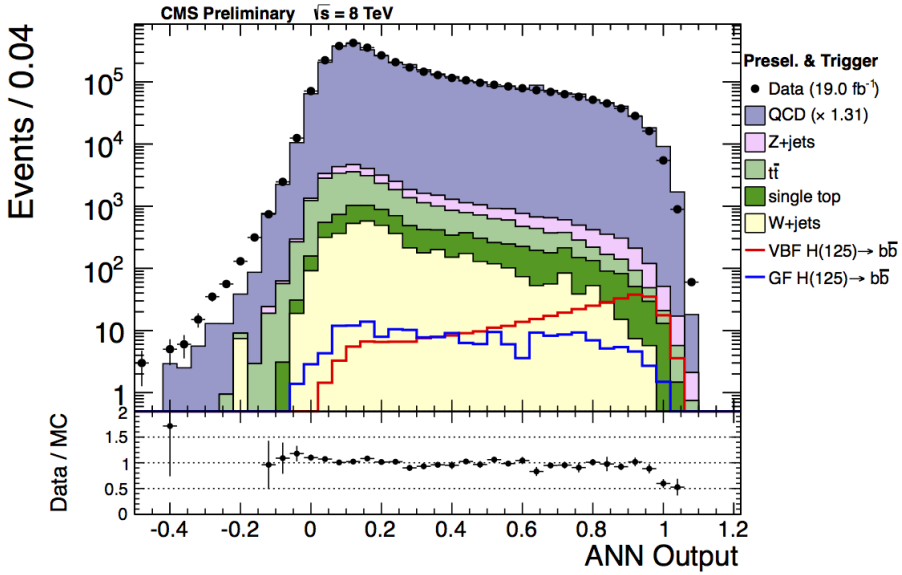


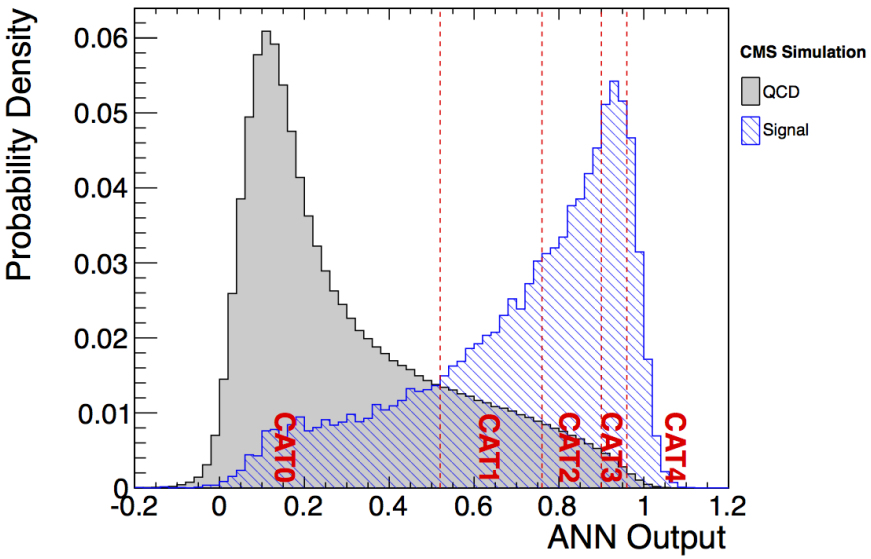


- Results of fit in M_{bb} for 125 GeV Higgs:
  - $\mu = 0.76^{+0.68}_{-0.66}$
  - Observed (expected) 95% CL Limit: 2.0 (1.4)
  - Similar to latest ATLAS results! (ATLAS-CONF-2013-079)

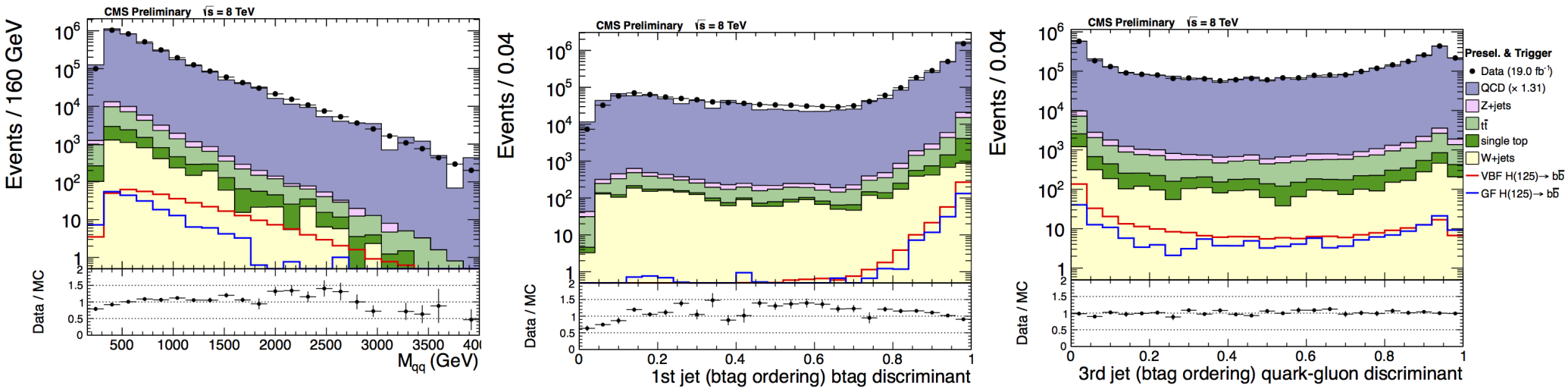
Analysis Strategy: VBF H → bb

- Jet pT > (85,70,60,40) GeV
- Quark jets from VBF required to have $\Delta\eta_{qq} > 2.5$, $m_{qq} > 300$ GeV
- Higgs candidate jets: $d\varphi_{bb} < 2$
- Construct Artificial Neural Network (ANN) to separate ***QCD-dominated*** background from VBF qqH → qqbb events
- Separate into categories and fit for background in each
  - Polynomial for QCD, templates for SM backgrounds



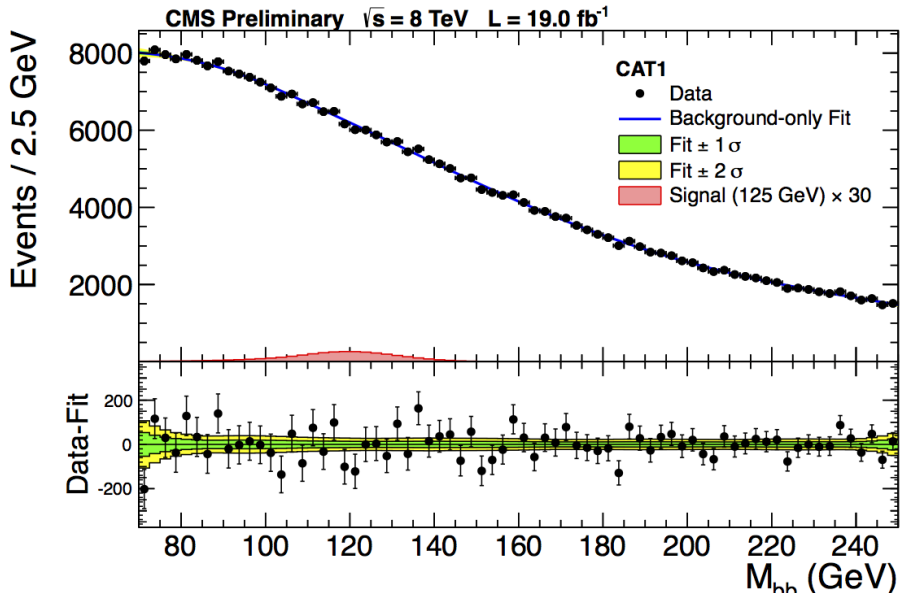


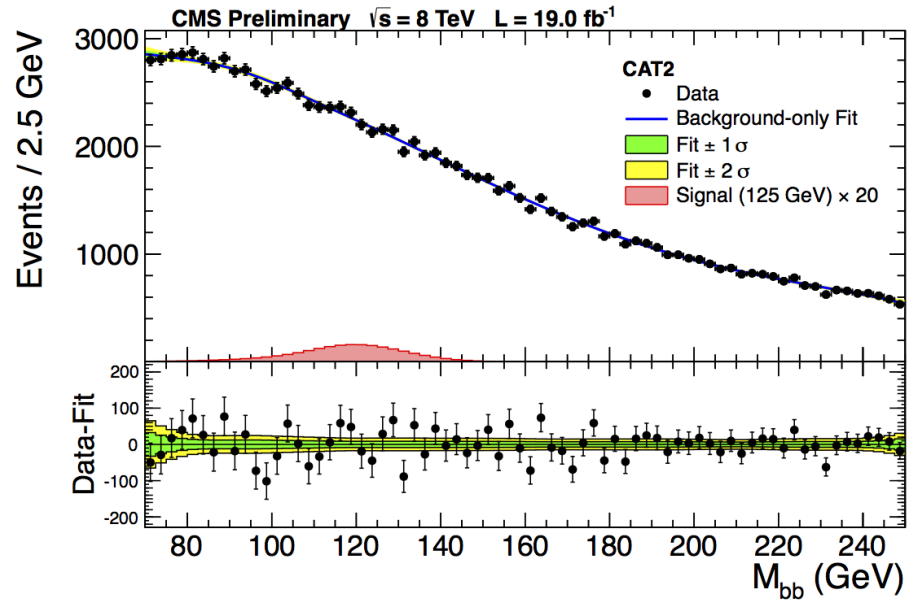
ANN Training and Inputs

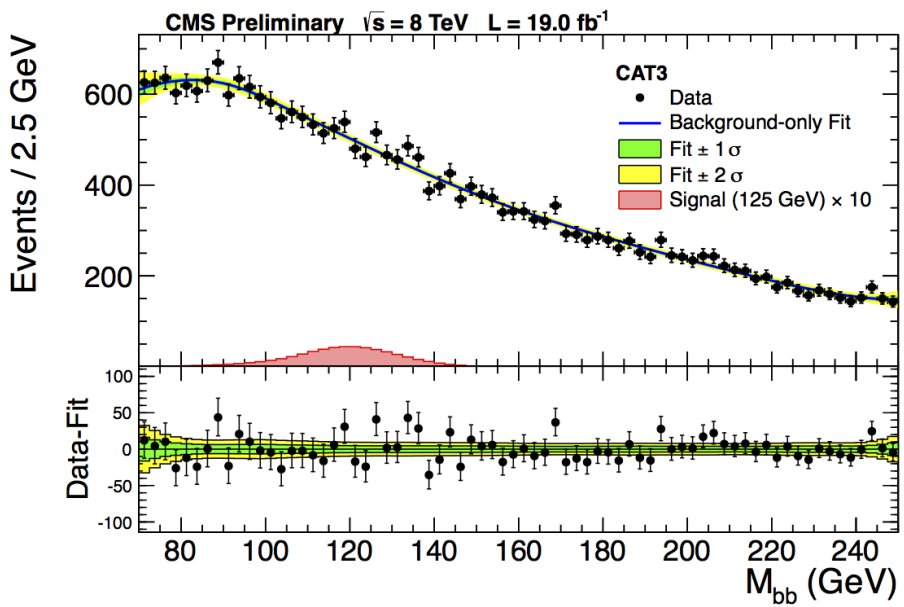


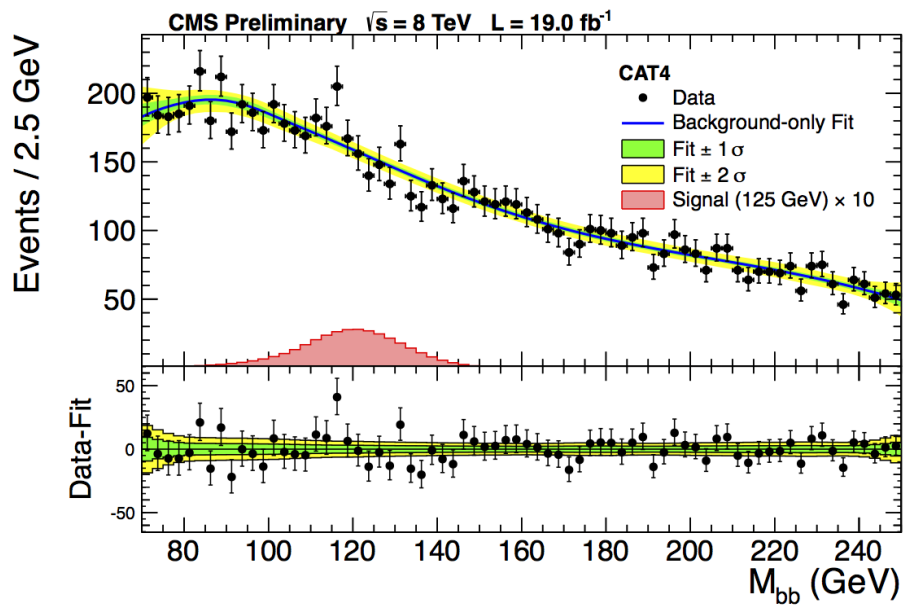
- Input variables – no strong correlations with m_{bb}
  - $\Delta\eta_{qq}$, $\delta\Delta\eta_{qq}$: separation b/w qq jets, difference from b-jets
  - Mass and pseudorapidity of qq pair system, η of 3rd jet
  - CSVs of b-tagged jets
  - Quark/gluon discriminant of q jets (constituent kinematics)
  - H_T – scalar sum of soft track jets in the event
  - Kinematics of bb system w.r.t. qq system (H/g separation)

Signal Extraction





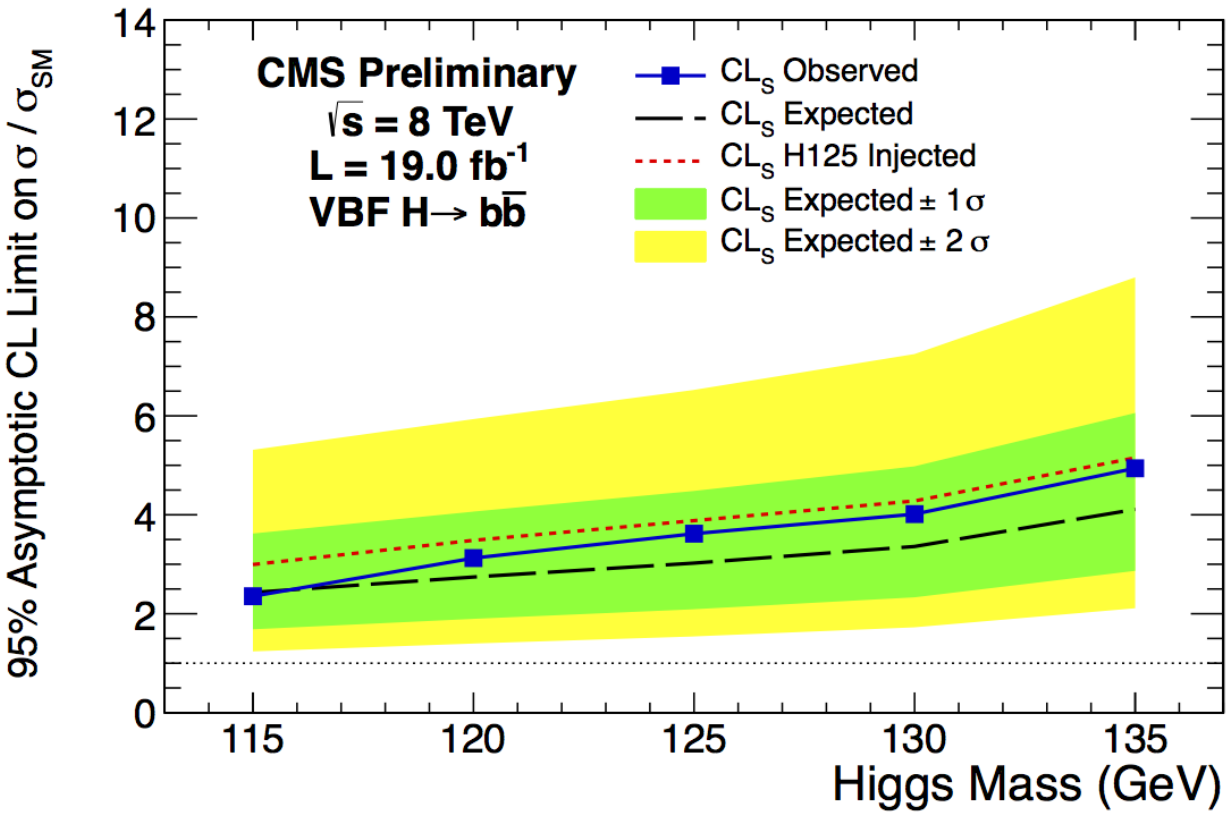




VBF Results

Results of fit for 125 GeV Higgs:

- $\mu = \sigma/\sigma_{SM} = 0.7 \pm 1.4$
- Observed (expected) 95% CL limit: 3.6 (3.0)
- Observed (expected) significance: 0.5σ (0.7σ)



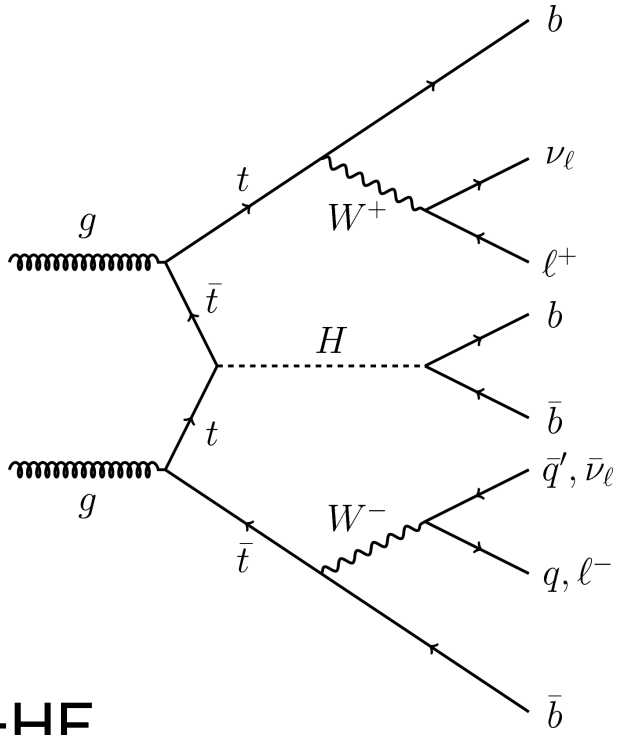
- Uncertainty dominated by background statistics
- All systematic uncertainties combined worsen limit by ~15%

Analysis Strategy: ttH → ttbb

- Lepton+Jets (lvlvbb), Dilepton (lvqqbb)
  - Subcategories: number of jets and b-tags (CSVM, p_T > 30 GeV)

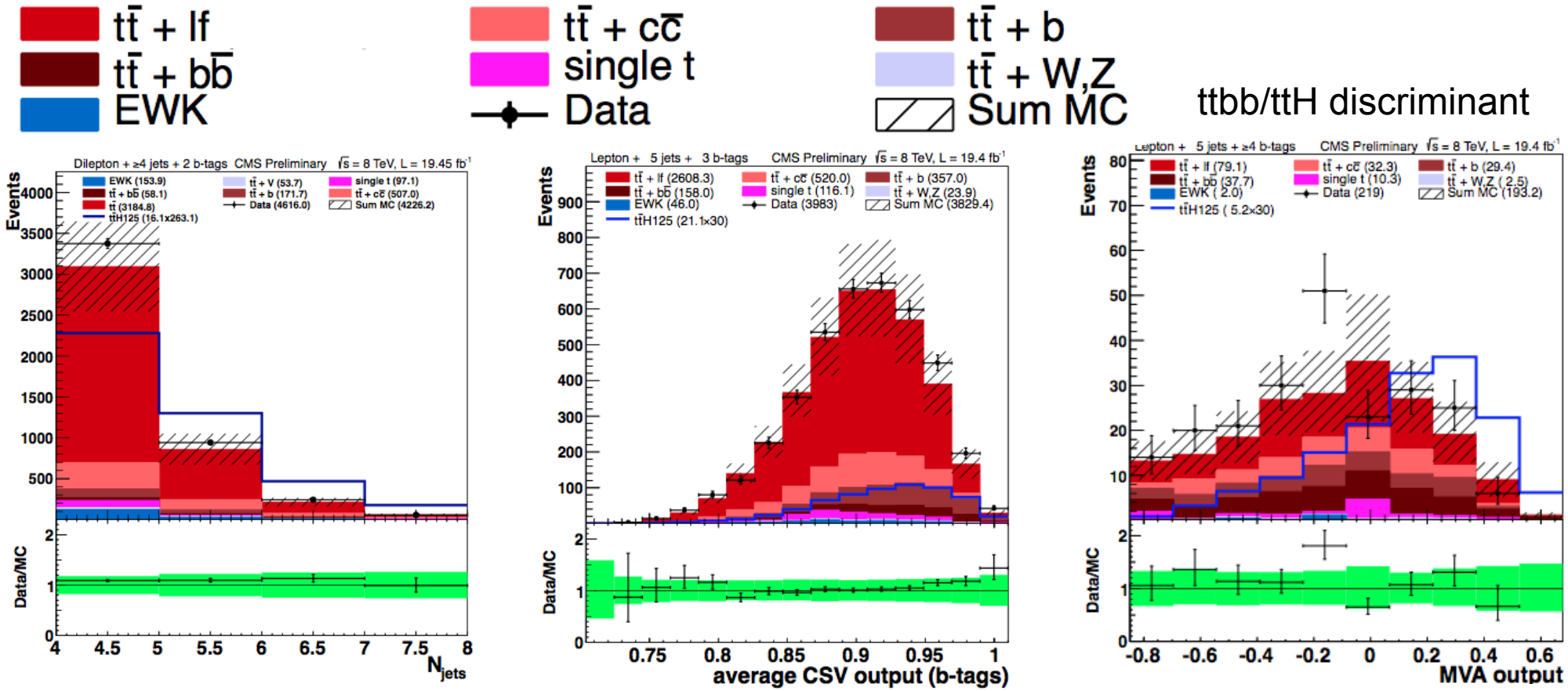
	Lepton+Jets							Dilepton	
Jets	≥6	4	5	≥6	4	5	≥6	2	≥3
Tags	2	3	3	3	4	≥4	≥4	2	≥3

- BDT tuned in each subcategory
  - Inputs: kinematics of single or combined objects, “best Higgs mass”, event shape, CSV discriminants of jets
- Global fit of BDT distributions
  - Includes ttH → ttττ for 8 TeV
- Systematic uncertainties
  - Experimental: b-tag efficiency and fake rate, Jet Energy Scale, ...
  - Theoretical: Q^2 on tt+jets, extra 50% on tt+HF, ...

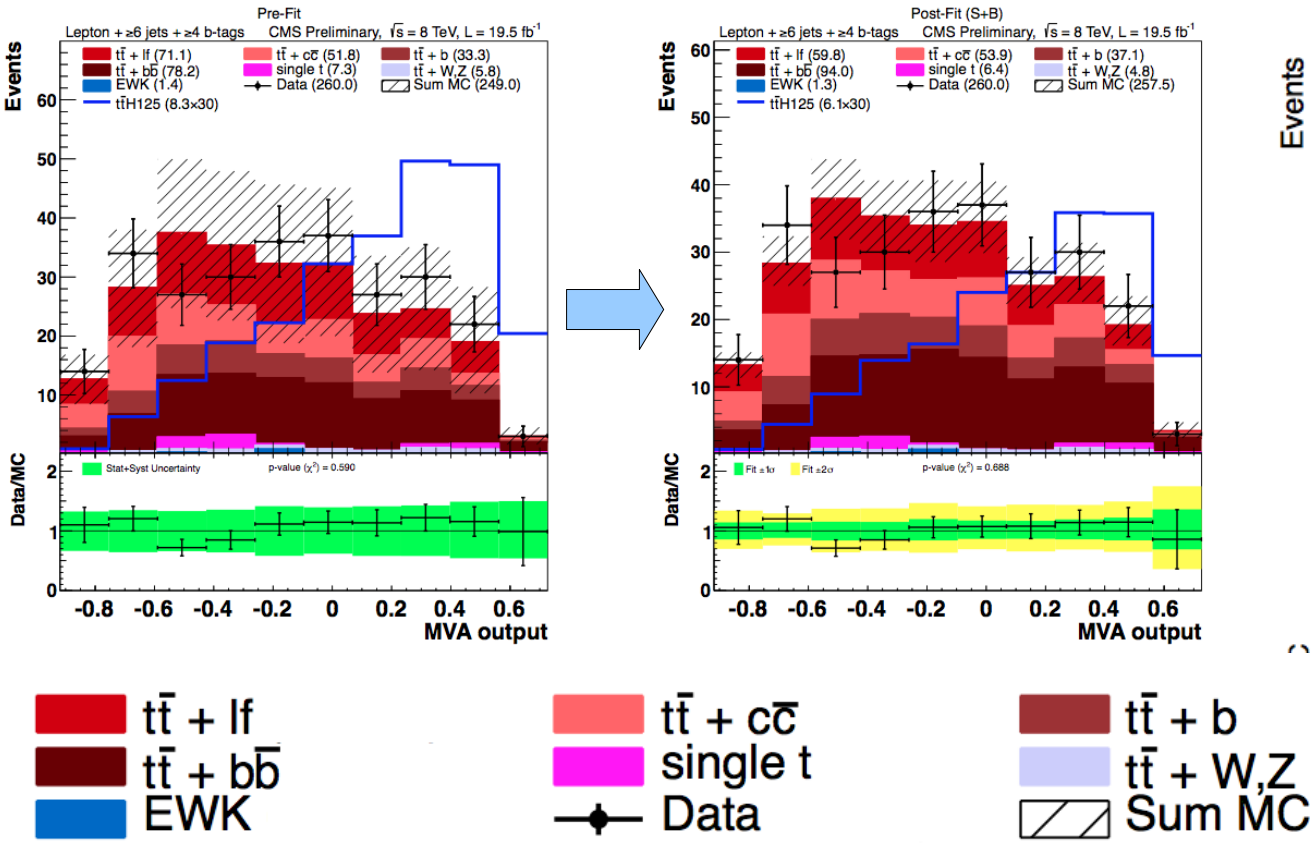


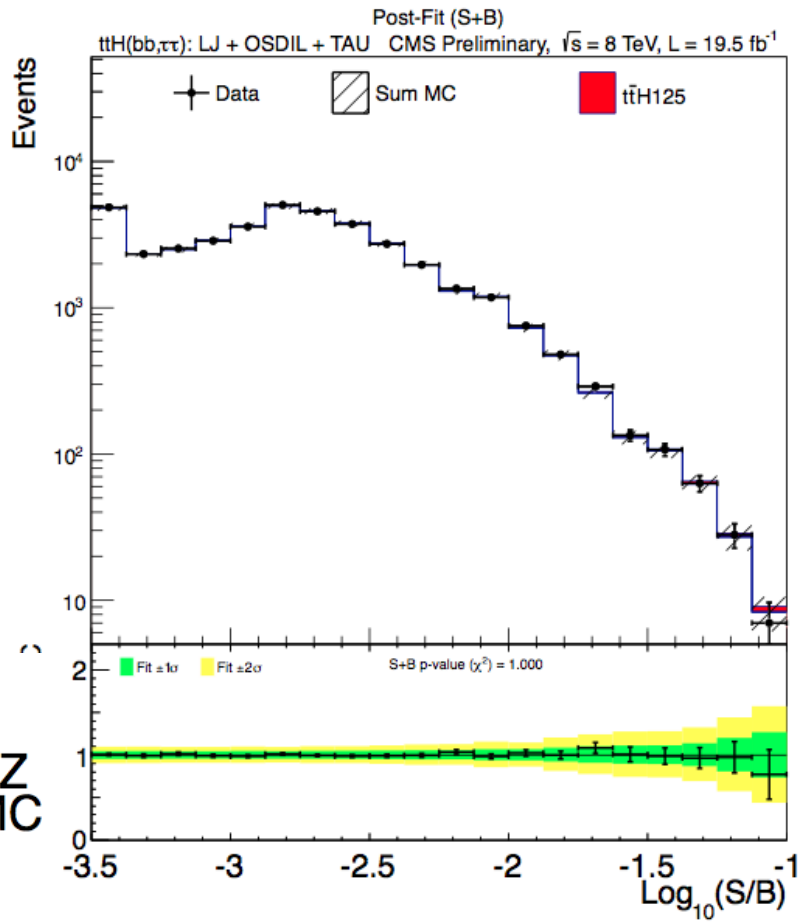
BDT Inputs and Training

- Training and variable selection separate for each channel
  - Up to 10 variables with highest 1D signal/background separation
- Specialized BDT trained against ttbb used as input for event-wide discriminant in most sensitive categories: 13% overall improvement



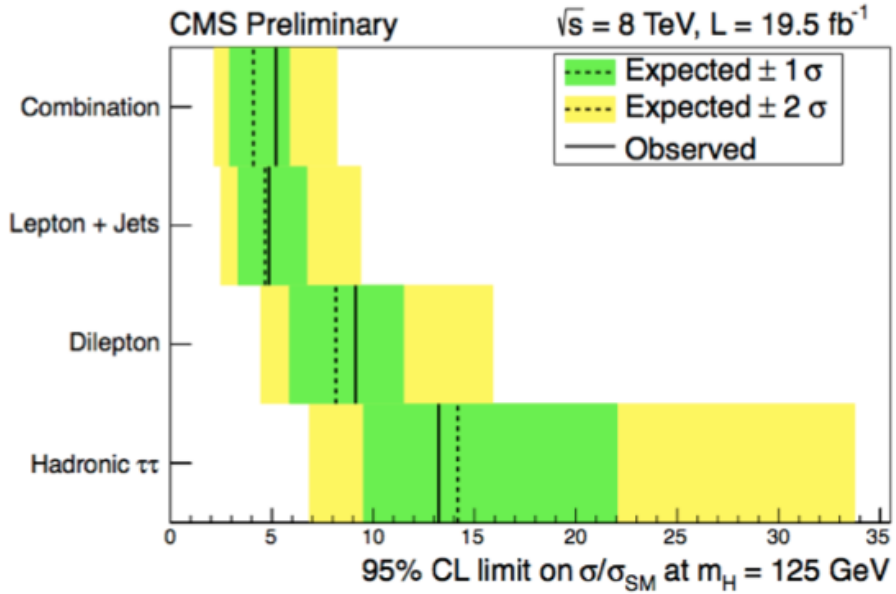
Signal Extraction

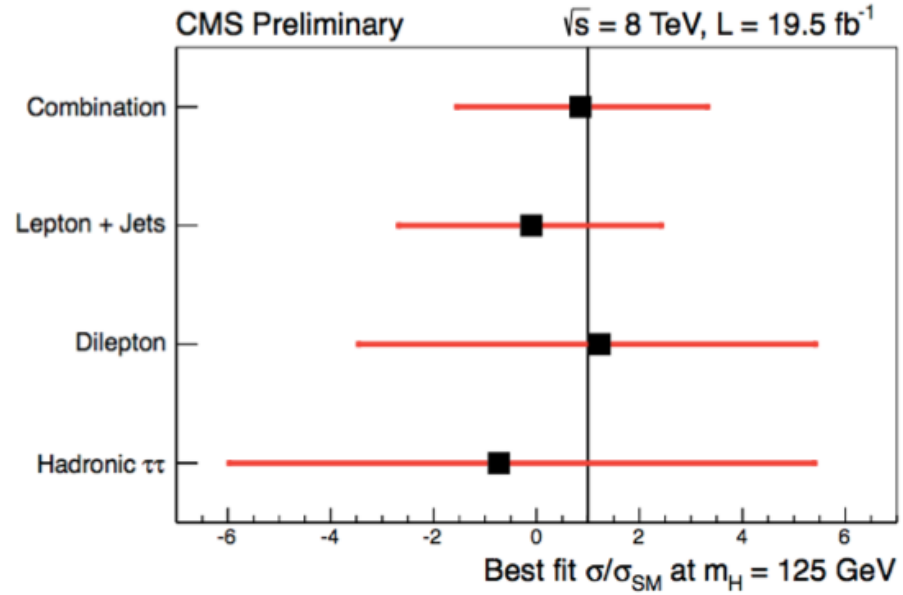




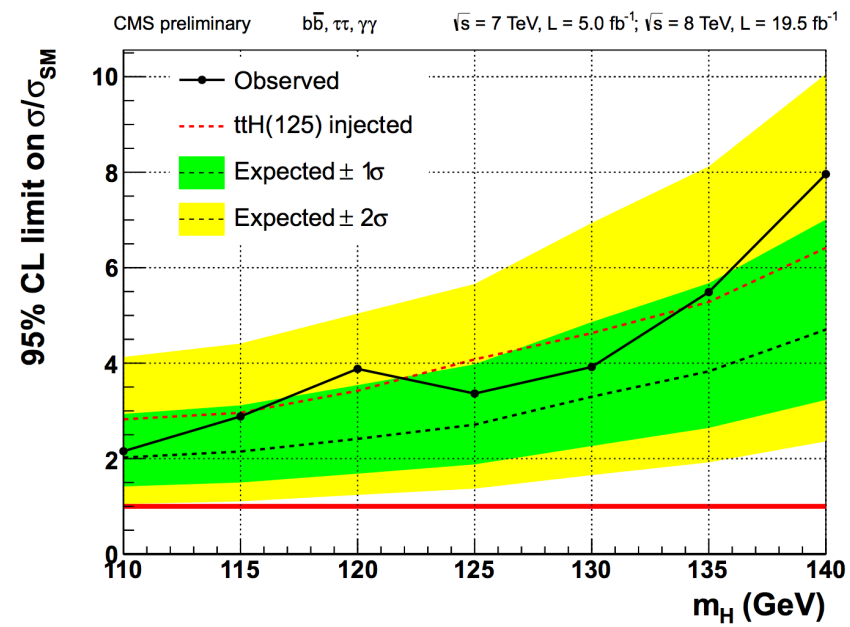
- Simultaneous fit in all categories, including tau
- Visualize fit by summing bins by log(S/B)

ttH → ttbb Results





- Observed (expected) 95% CL limit for 8 TeV at 125 GeV:
  - Combined [ττ,bb]: 5.2 (4.1)
  - Lepton+Jets: 4.9 (4.7)
  - Dilepton: 9.1 (8.2)
- Combined 7+8 TeV ttH(bb) + 8 TeV ttH(ττ,γγ): 3.3 (3.9)

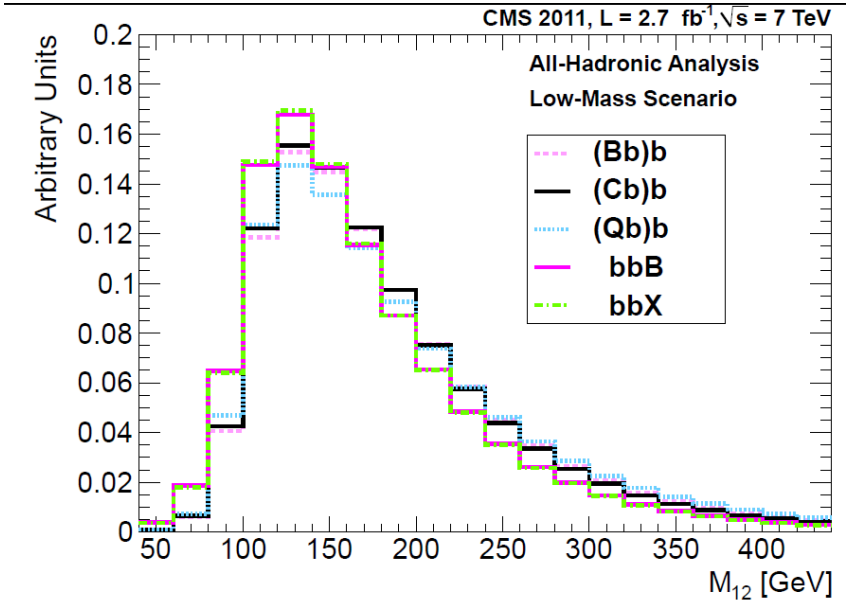


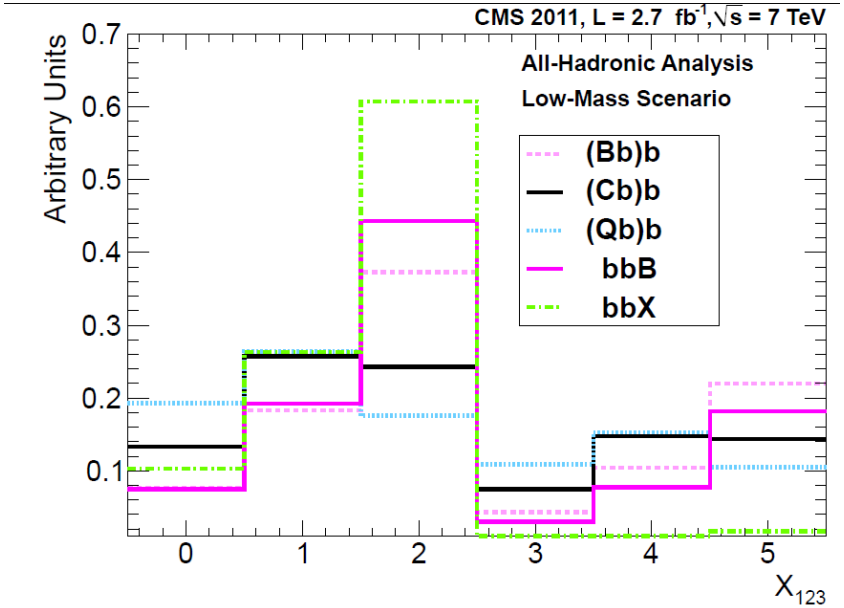
Analysis Strategy: bH → bbb

- All-Hadronic analysis
  - 3 reconstructed jets with |eta| < 2.2
  - 2 online b-tags required for trigger
  - Low-mass: pT(j1,j2,j3) > (46,38,20) GeV – 2.7 fb^{-1}
  - Medium-mass: pT(j1,j2,j3) > (60,53,20) GeV – 4.0 fb^{-1}
- Semi-leptonic analysis
  - Require muon (pT > 15 GeV) from semi-leptonic b decay
  - Reduce trigger rate, allow jet pT(j1,j2,j3) < (30,30,20)
- Derive QCD background in 3 b-tag signal sample using events with 1-2 b-tags
- Dominant systematics: trigger and b-tagging efficiency, jet energy scale, factorization/renormalization QCD scale

All-Hadronic Background

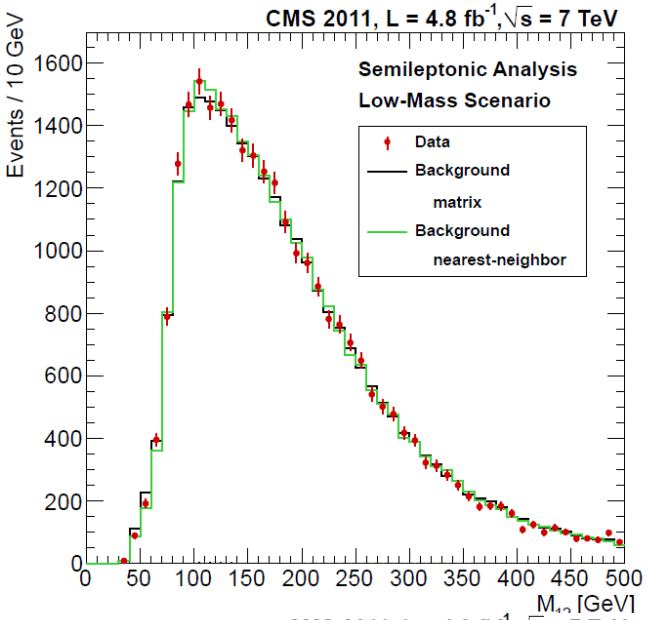
- Construct 2 variables for fits:
  - Dijet mass of 2 leading jets
  - Event Btag variable X_{123}
    - Defined based on masses of reconstructed b-vertices
    - Sufficient statistics in all bins
- In 2 b-tag sample, define templates using assumed flavor of untagged jet
  - B-tagging efficiency from MC
    - Function of (p_T,η)
    - Data/MC scale factors
  - Additional simulation-derived corrections for non-bb contribution (3-4%), online btagging

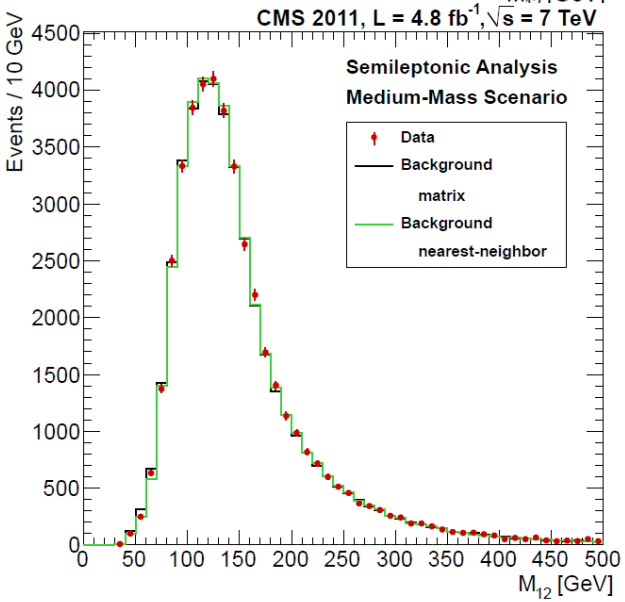




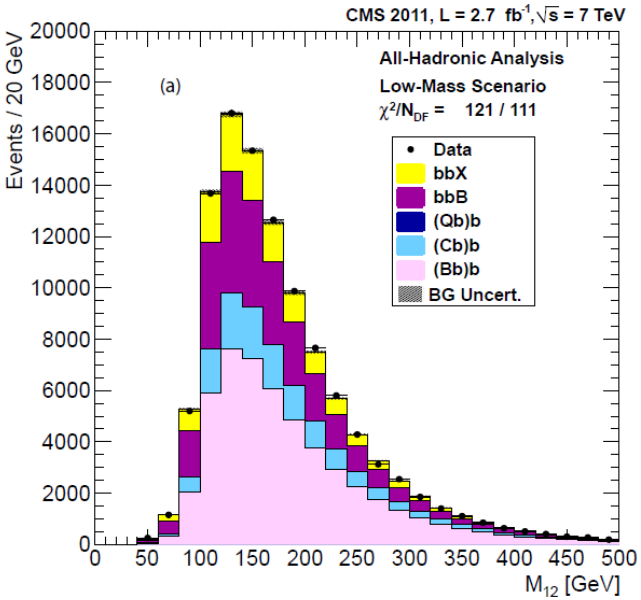
Semi-Leptonic Background

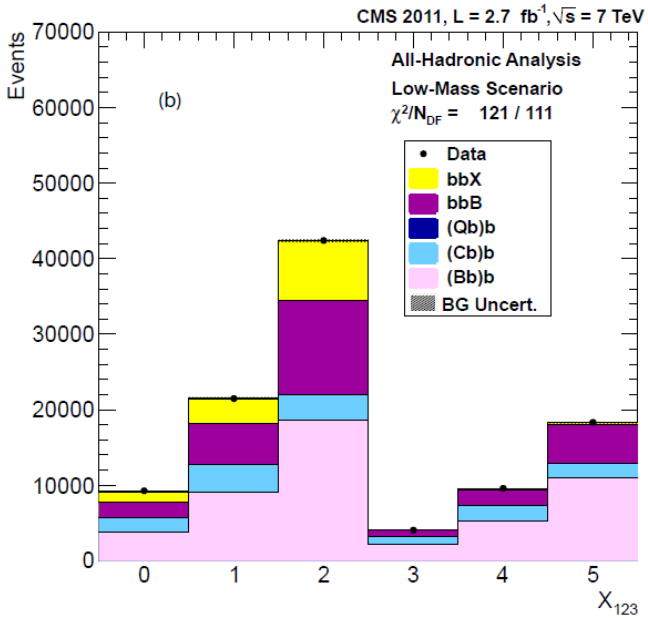
- Templates derived using two methods using disjoint samples: bbj, bjj
- Likelihood ratio to differentiate background-like from signal-like events
  - Jet pT, separation of lead jets in (η,φ), and separation from 3rd jet
- Matrix Method: template from bbj events; b-tagging probabilities for 3rd jet
- Nearest Neighbor method: use kinematically-similar background events in control region to predict bjj → bbb probability
- Good agreement between methods!

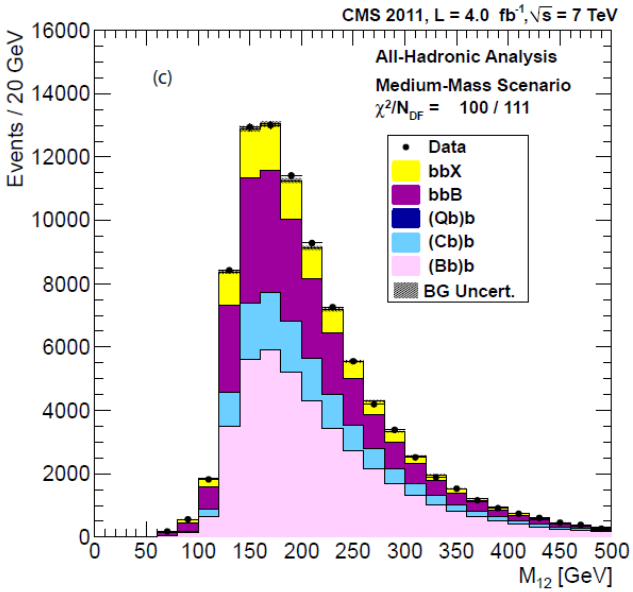


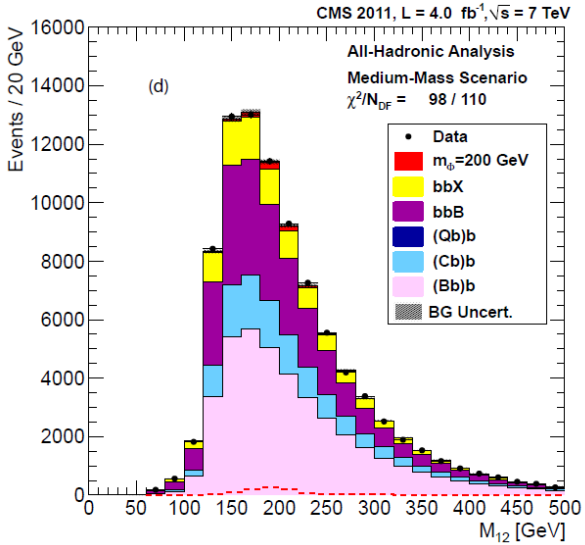


All-Hadronic Signal Extraction



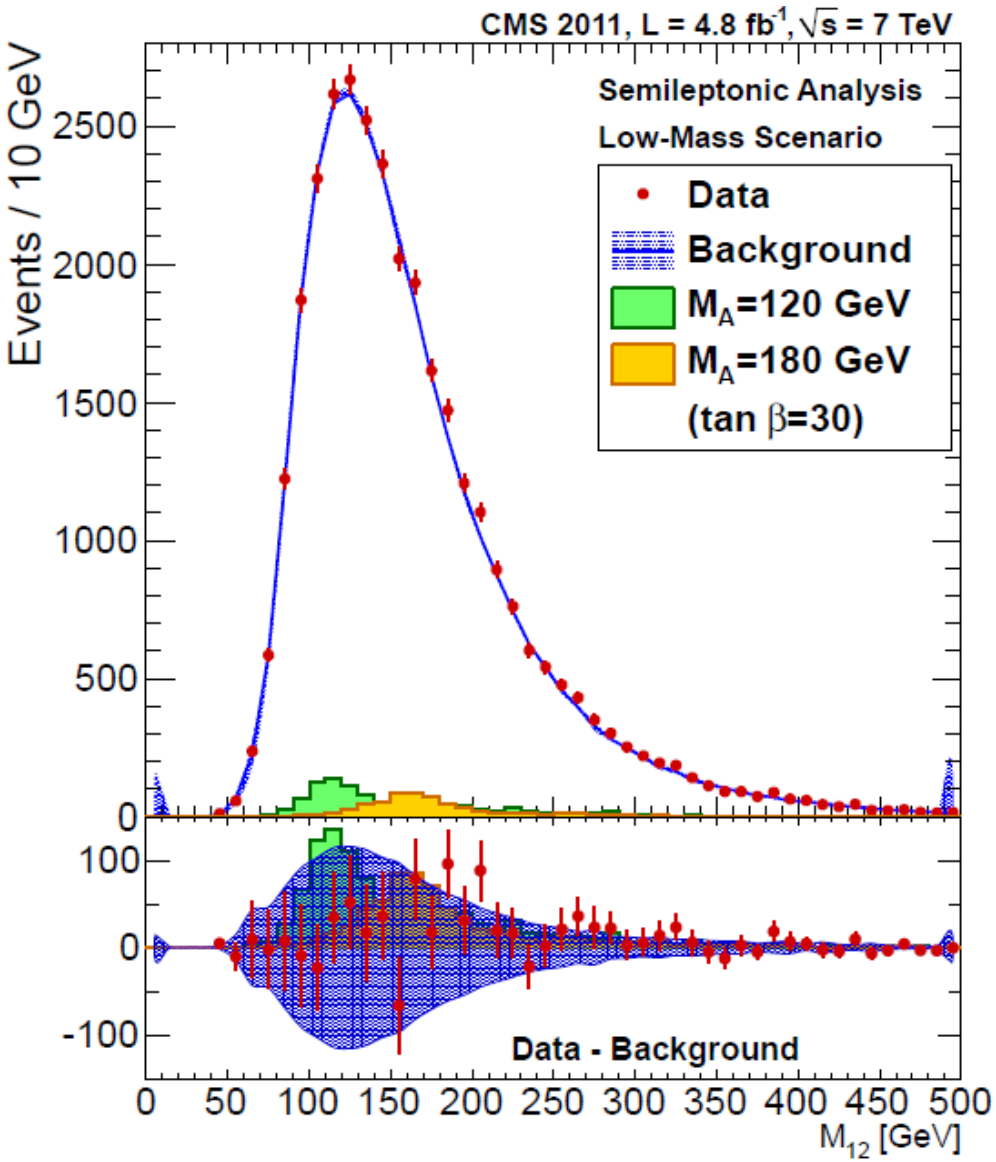


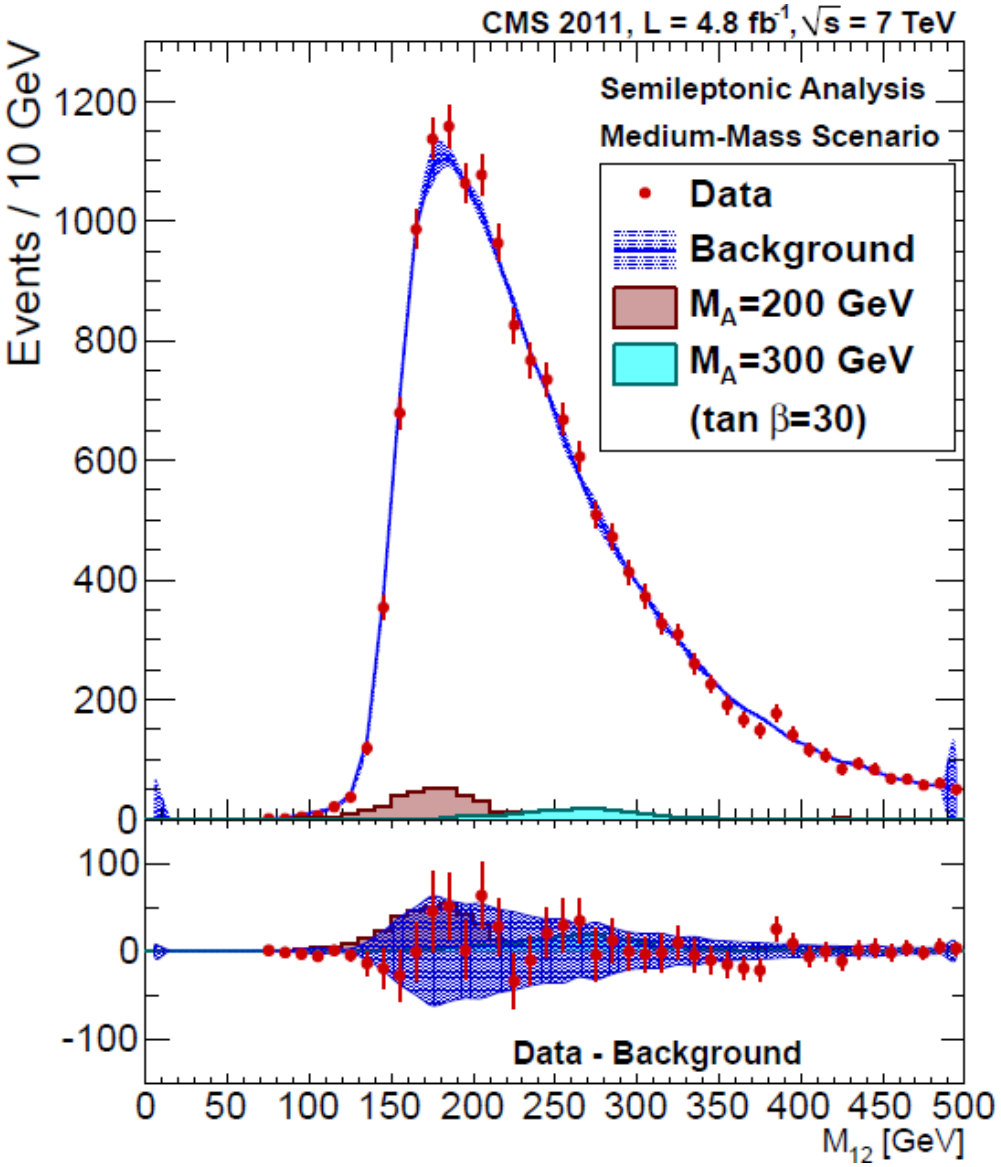




- Low-mass fit in M_{bb} and X_{123}
- High-mass fit in Mbb
  - No signal - left
  - With signal - right

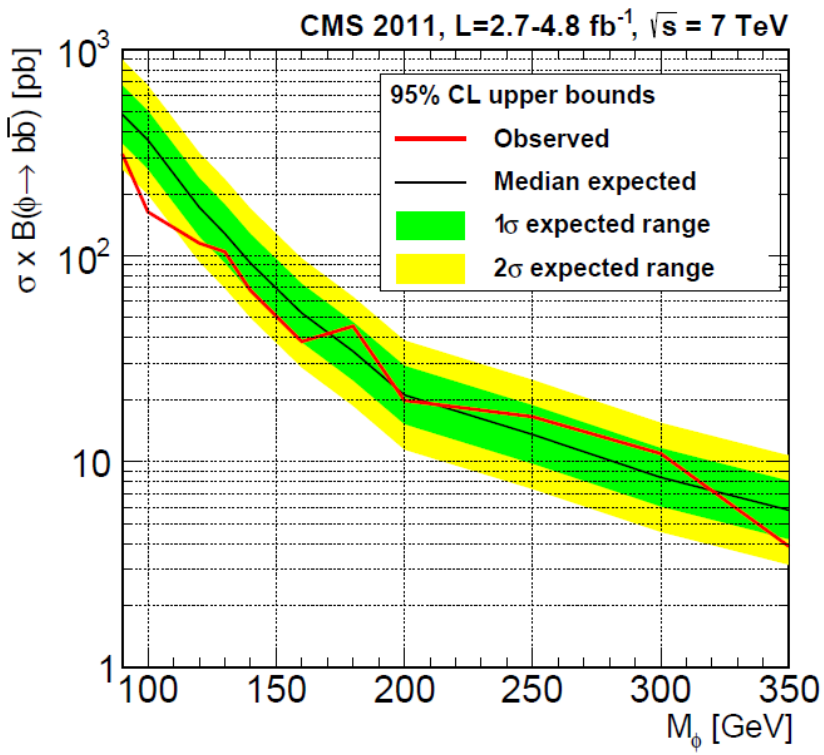
Semi-Leptonic Signal Extraction

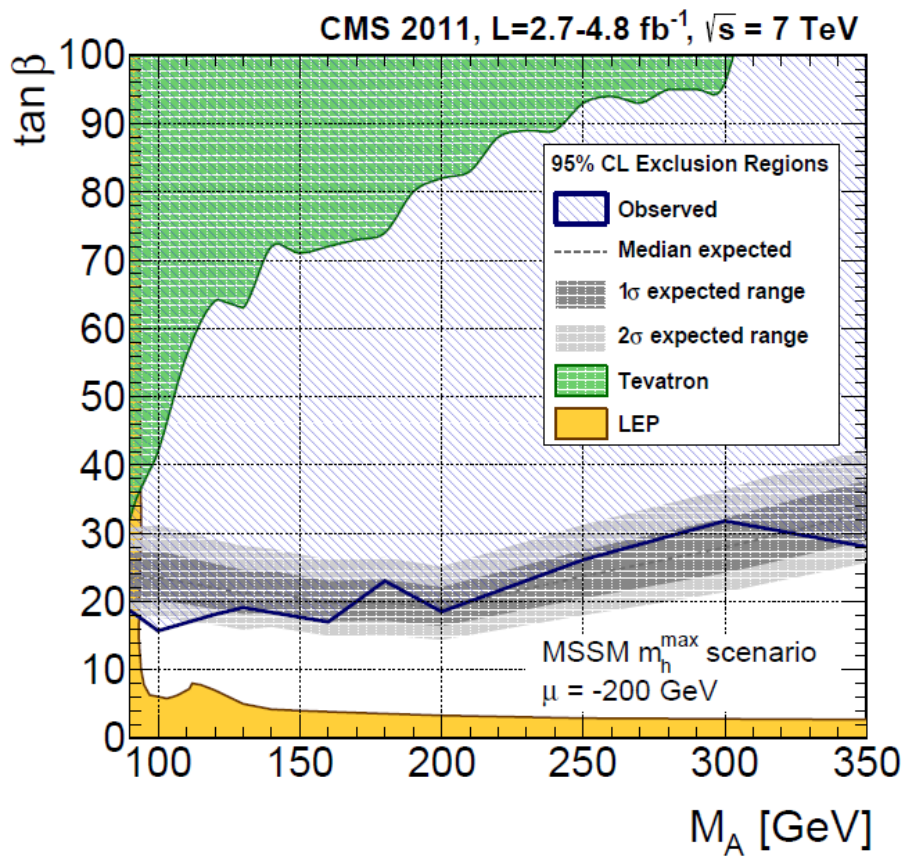




bH → bbb Results

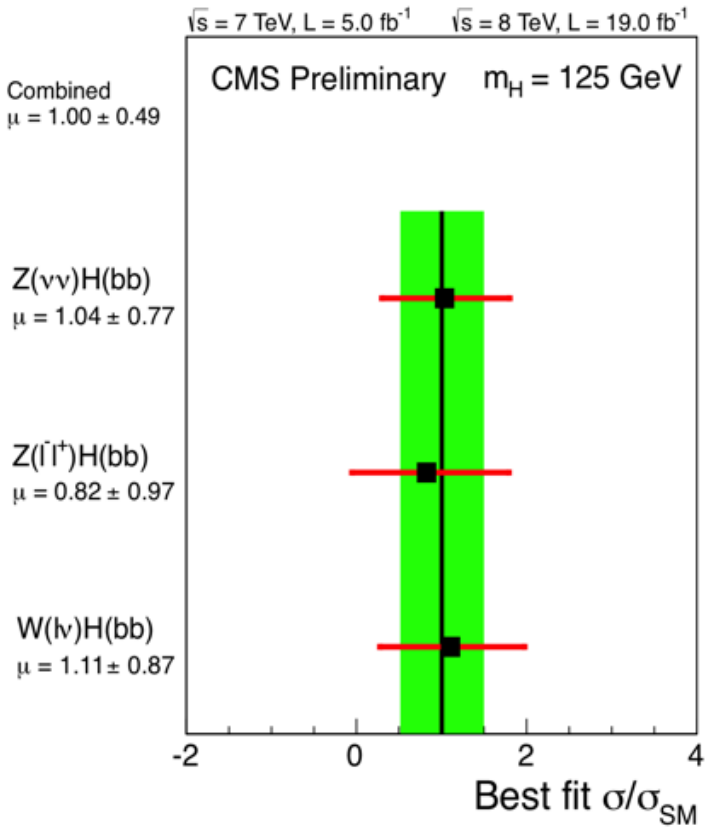
- Limits placed on cross section for bH+X cross section times H → bb branching ratio
- Interpreted as limits on parameters in the MSSM m_h^{max} benchmark scenario

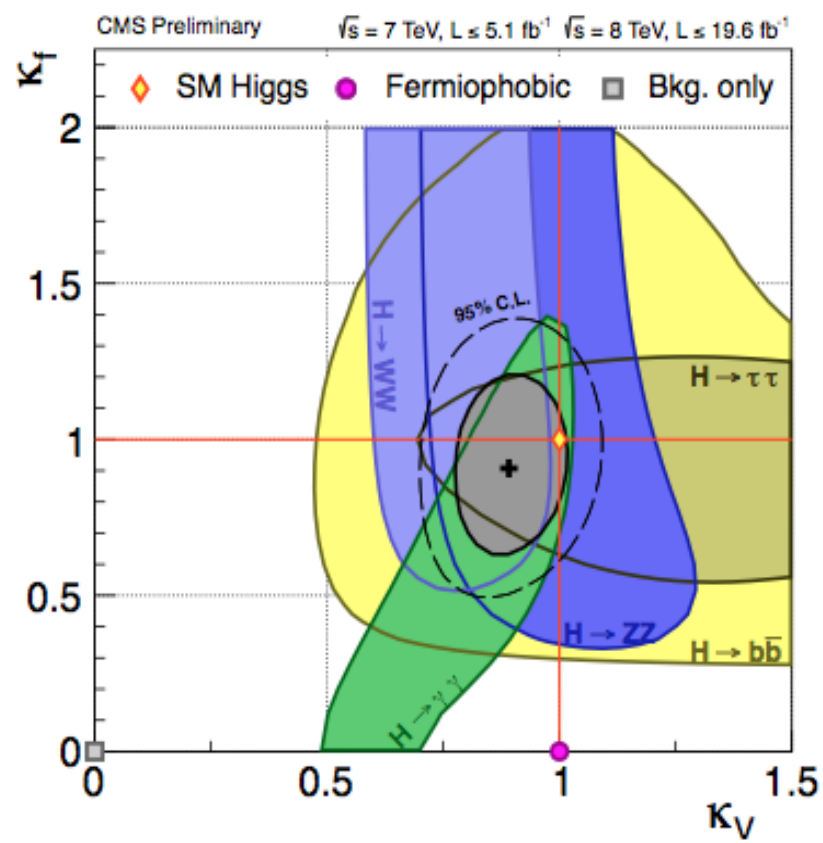




Conclusions

- Broad, well-developed CMS H → bb Program
  - All possible SM production mechanisms exploited
  - Multivariate techniques maximize kinematic information
- Everything consistent with Standard Model so far!





Extras

VH Systematic Uncertainties

Source	Type	Yield uncertainty (%) range	Contribution to uncertainty (%)	Removal effect on total uncertainty (%)
Luminosity	normalization	2.2-4.4	< 2	< 0.1
Lepton efficiency and trigger (per lepton)	normalization	3	< 2	< 0.1
$Z(\nu\nu)H$ triggers	shape	3	< 2	< 0.1
Jet energy scale	shape	2–3	5.0	0.5
Jet energy resolution	shape	3–6	5.9	0.7
Missing transverse energy	shape	3	3.2	0.2
b-tagging	shape	3–15	10.2	2.1
Signal cross section (scale and PDF)	normalization	4	3.9	0.3
Signal cross section (p_T boost, EWK/QCD)	normalization	2/5	3.9	0.3
Signal Monte Carlo statistics	shape	1–5	13.3	3.6
Backgrounds (data estimate)	normlization	10	15.9	5.2
Single-top (simulation estimate)	normalization	15	5.0	0.5
Dibosons (simulation estimate)	normalization	15	5.0	0.5
MC modeling (V+jets and $t\bar{t}$)	shape	10	7.4	1.1

VBF Systematic Uncertainties

Source	Uncertainty
Background fit	depending on the statistics of each category
Z+jets cross section	$\pm 20\%$
top cross section	$\pm 20\%$
Signal and Z peak position (JES)	$\pm 1.5\%$
Signal and Z resolution	$\pm 10\%$
Luminosity	$\pm 4.4\%$
Trigger efficiency	$\pm 5 - 8\%$
Signal acceptance due to JES	$\pm 10\%$
Signal acceptance due to JER	$\pm 2\%$
VBF cross section	$\pm 3\%$
VBF Monte Carlo acceptance	$\pm 10\%$
PDF	$\pm 5\%$
VBF ANN shape due to b-tag	$\pm 2\%$
VBF ANN shape due to quark-gluon discriminator	$\pm 2\%$
VBF ANN shape due to UE modeling	$-8 - +2\%$
GF cross section	$\pm 15\%$
GF Monte Carlo acceptance	$\pm 50\%$
GF ANN shape	$\pm 50\%$

ttH Systematic Uncertainties

Source	Shape	Remarks
Luminosity	No	Signal and all backgrounds
Lepton ID/Trig	No	Signal and all backgrounds
Pileup	No	Signal and all backgrounds
Top p_T reweighting	Yes	Only $t\bar{t}$ background
Jet Energy Resolution	No	Signal and all backgrounds
Jet Energy Scale	Yes	Signal and all backgrounds
b-Tag hf fraction	Yes	Signal and all backgrounds
b-Tag hf stats (linear)	Yes	Signal and all backgrounds
b-Tag hf stats (quadratic)	Yes	Signal and all backgrounds
b-Tag lf fraction	Yes	Signal and all backgrounds
b-Tag lf stats (linear)	Yes	Signal and all backgrounds
b-Tag lf stats (quadratic)	Yes	Signal and all backgrounds
b-Tag Charm (linear)	Yes	Signal and all backgrounds
b-Tag Charm (quadratic)	Yes	Signal and all backgrounds
QCD Scale ($t\bar{t}H$)	No	Scale uncertainty for NLO $t\bar{t}H$ prediction
QCD Scale ($t\bar{t}$)	No	Scale uncertainty for NLO $t\bar{t}$ and single top predictions
QCD Scale (V)	No	Scale uncertainty for NNLO W and Z prediction
QCD Scale (VV)	No	Scale uncertainty for NLO diboson prediction
pdf (gg)	No	Pdf uncertainty for gg initiated processes ($t\bar{t}$, $t\bar{t}Z$, $t\bar{t}H$)
pdf ($q\bar{q}$)	No	Pdf uncertainty for $q\bar{q}$ initiated processes ($t\bar{t}W$, W, Z).
pdf (qg)	No	Pdf uncertainty for qg initiated processes (single top)
Madgraph Q^2 Scale ($t\bar{t}$+0p,1p,2p)	Yes	Madgraph Q^2 scale uncertainty for $t\bar{t}$+jets split by parton number. There is one nuisance parameter per parton multiplicity and they are uncorrelated.
Madgraph Q^2 Scale ($t\bar{t}$+b/$b\bar{b}$/$c\bar{c}$)	Yes	Madgraph Q^2 scale uncertainty for $t\bar{t}$+b/$b\bar{b}$/$c\bar{c}$.
Madgraph Q^2 Scale (V)	No	Varies by jet bin.
Extra $t\bar{t}$+hf rate uncertainty	No	A 50% uncertainty on the rate of $t\bar{t}$+b, $t\bar{t}+b\bar{b}$, $t\bar{t}+c\bar{c}$.

bH Systematic Uncertainties

Source	all-hadronic	semi-leptonic	type
Trigger efficiency	10%	$3-5\%$	rate
Online b-tagging efficiency	32%	-	rate
b-tagging efficiency	$10-13\%^{\dagger}$	12%	shape/rate
b-tagging efficiency dependence on topology	6%	-	rate
Jet Energy Scale	$1.4-6.8\%$	$^{+2.5}_{-3.1}\%$	shape/rate
Jet Energy Resolution	$0.6-1.3\%$	1.9%	shape/rate
Muon momentum scale	-	0.2%	rate
Muon momentum resolution	-	0.6%	rate
Signal Monte Carlo statistics	$1.1-2.6\%$		rate
Integrated luminosity	2.2%		rate
PDF and α_s uncertainties*	$3-6\%$	$^{+(2.5-4.7)}_{-(2.7-4.4)}\%$	rate
Renormalisation and fragmentation QCD scale*	$6-28\%$		rate
Underlying event and parton showering*	4%		rate