Search for the Standard Model Higgs boson decaying into a bottom quark pair with the CMS experiment

Arnaud Pin[1]
On behalf of the CMS Collaboration

1 : for the center for Cosmology, Particle Physics and Phenomenology,
Université Catholique de Louvain, Belgium

arnaud.pin@cern.ch

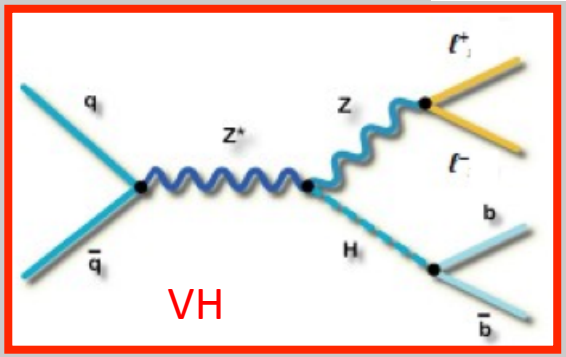
5th May – Phenomenology Symposium 2014 – University of Pittsburgh, USA

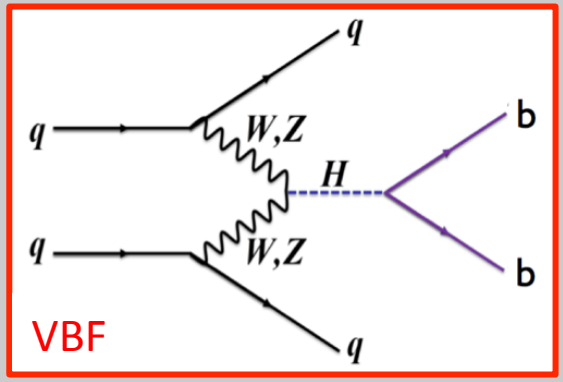
Searches and Measurement of the SM Higgs boson:

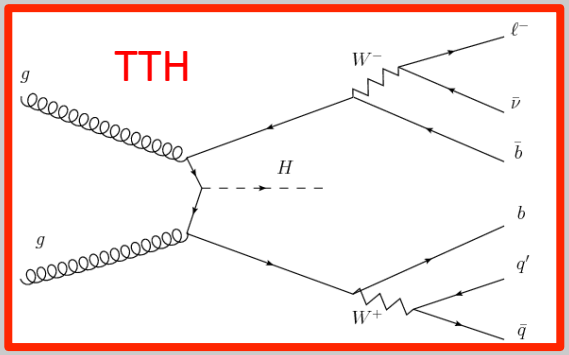
- Search for H→bb: test for fermionic final state.
- Large QCD background → Inclusive H(bb) production too challenging

3 Search Analysis:

- Need to search for H→bb in associated production:
  - Less background.
  - Lower Cross-section.
  - Associated production with:
    - Vector boson: **VH(bb)** [PhysRevD.89.012003]
    - Quarks: **VBF H to bb** [CMS-PAS-HIG-13-011]
    - Top quarks: **ttH H to bb** [CMS-PAS-HIG-13-019]







Analysis strategy:

1) Event Selection and Categorization

- Select boosted events on P_t V.
- Vector boson decaying leptonically:
  - **6 channels**: 2x **W(lν)**, 2x **Z(ll)**, **Z(νν)**, **W(τν)**

- Categorization depends on **P_t V**:

Pt V [GeV]	W(lν)	W(τν)	Z(ll)	Z(νν)
low	100 – 130	> 120	50 – 100	100 – 130
intermediate	130 – 180			130 – 170
high	> 180		> 100	> 170

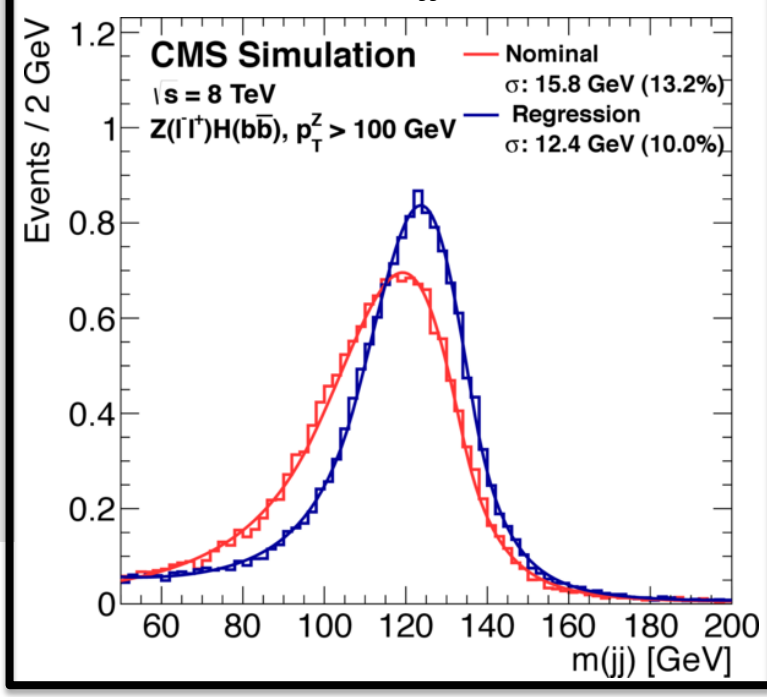
- Large azimuthal opening angle, $\Delta\varphi$(V,H)
- Total of 14 Categories (e/μ).

2) Multivariate analysis: BDT

Improved mass resolution:

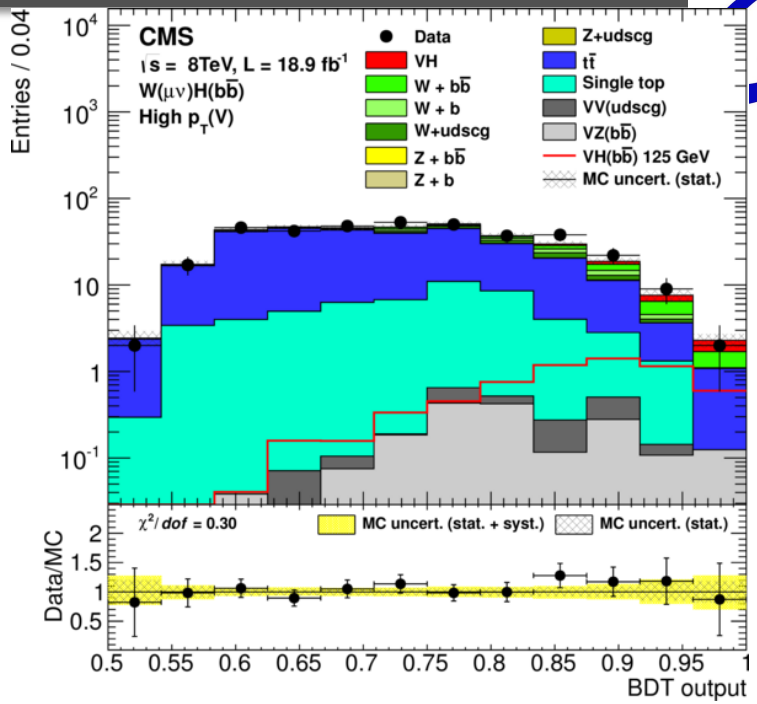
- using multivariate (BDT) jet-energy regression.
- Trained with ZH signal M.C.
- validated with Z+jets data.

→ **Impact on M_{jj} spectrum:**



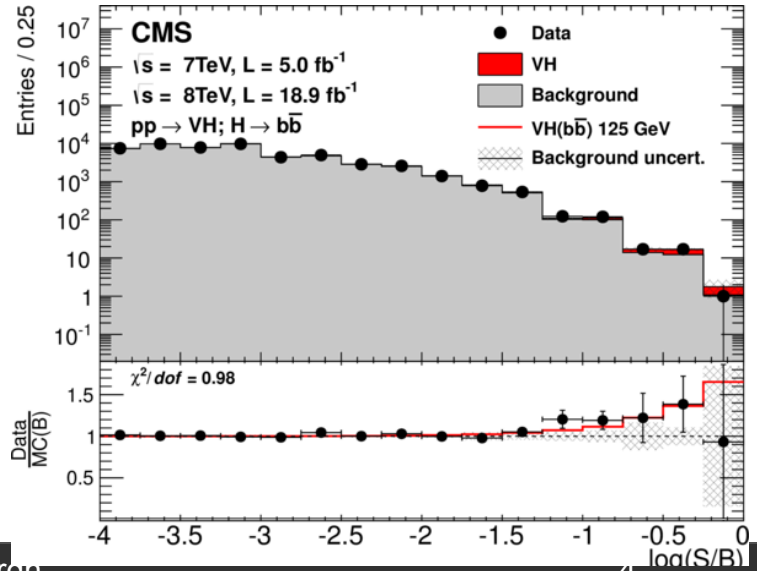
BDT approach

- Cascade BDT.
- Samples divided in **four** subsets.
- Each enriched in: tt, V+jets, dibosons, and VH
- **14 BDT distributions are considered**
  - each P_T V categories
  - Electron/muon for Z(ll) & W(lν)



Background:

- Control regions identified in data.
- Used to validate the simulation modeling of the distributions used as input to the BDT discriminants
- To obtain scale factors used to adjust the simulation event yield estimates for the most important background processes



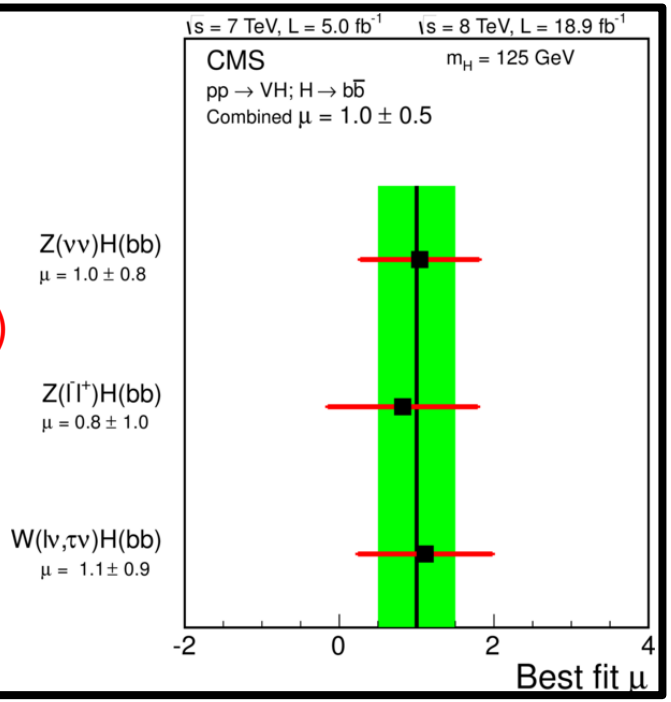
Z/W + H(bb): **Results**

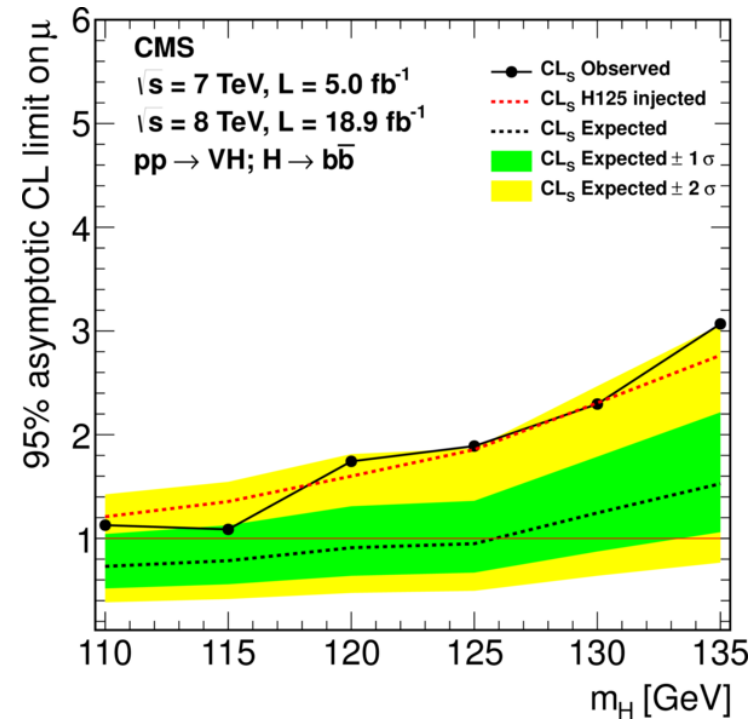
$m_H = 125$ GeV	σ/σ_{SM} (95% CL) median expected	σ/σ_{SM} (95% CL) observed	Significance expected	Significance observed
$W(\ell\nu,\tau\nu)H$	1.6	2.3	1.3	1.4
$Z(\ell\ell)H$	1.9	2.8	1.1	0.8
$Z(\nu\nu)H$	1.6	2.6	1.3	1.3
All channels	0.95	1.89	2.1	2.1

- **Excess** of events is **observed** above the expected background.
- Observation **agrees** with Standard Model expectation for Higgs boson @ 125 GeV.

Local Significance:

- **Significance:**
  2.1σ obs. (2.1σ exp.)
- **Best fit:**
  μ=1.0±0.5



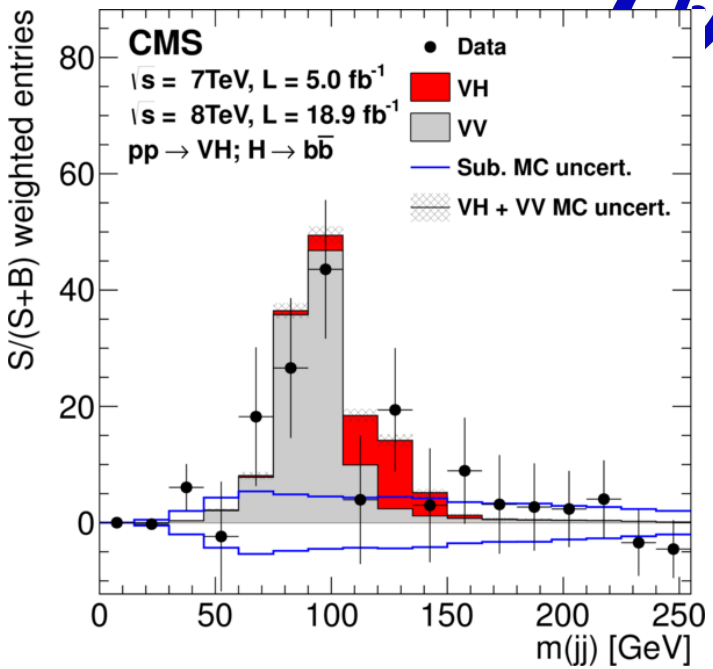


Z/W + H(bb): Cross Check

M_{bb} search:

- more restrictive selection optimized for sensitivity in this single variable

- Higgs boson of mass 125 GeV.
- 95% CL upper limits on σ/σ_{SM}
  Expected: 1.4 x SM
  Observed: 2.0 x SM

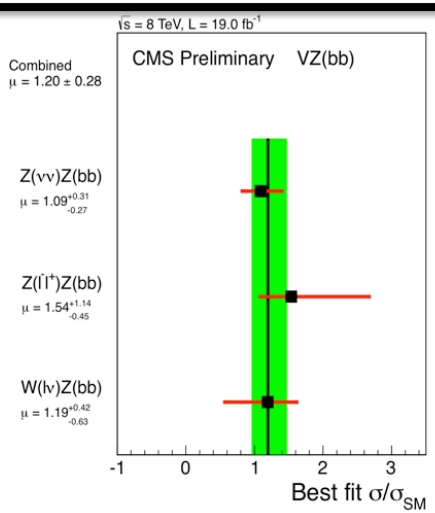


→ **Single variable gives consistent results**

Dibosons (VZ, Z→(bb)) analysis: (arXiv:1403.304)

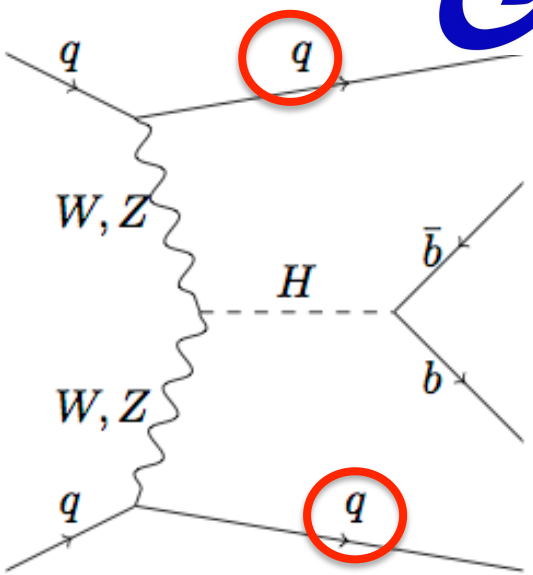
- **validation** of the multivariate approach
- BDT discriminants also trained to **find diboson** signals (ZZ and WZ, with Z → bb)

- In agreement with expectation.
- Check and measure of the VV cross section

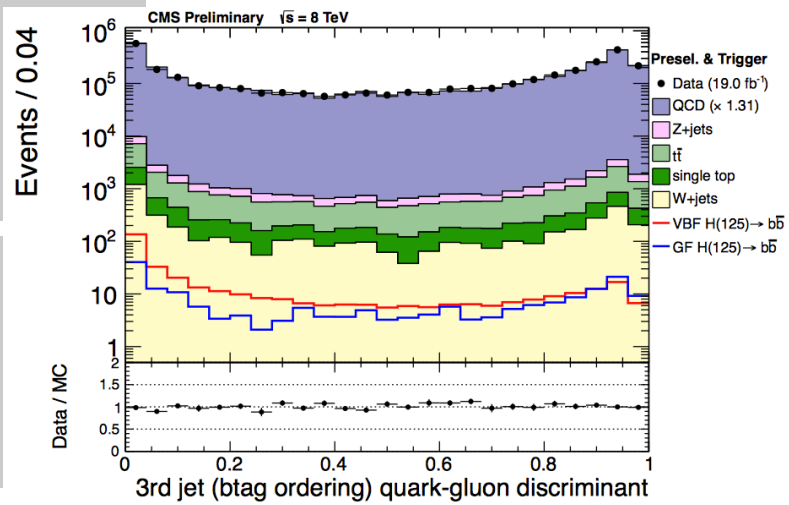


VBF H(bb)

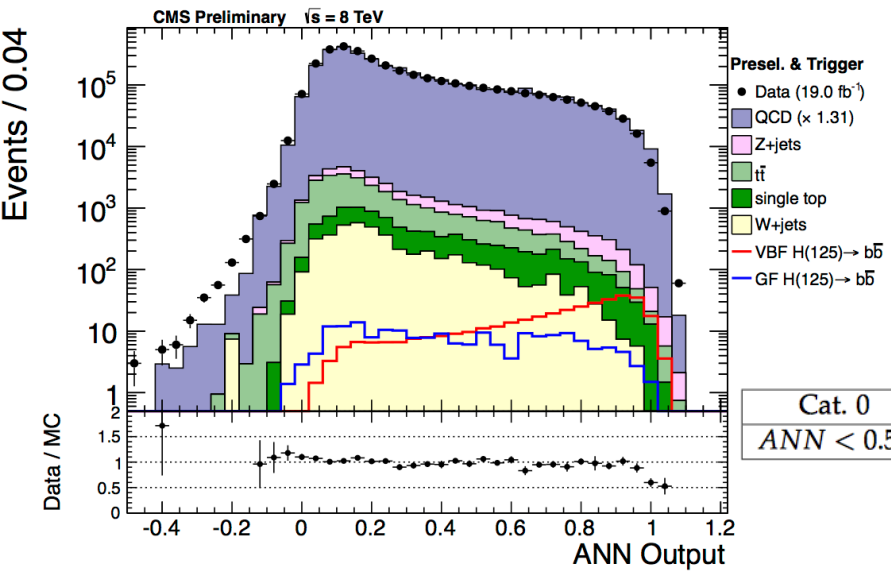
Fully hadronic final state:



- **VBF selection**
  - 4 jets
  - Dedicated trigger:
    - CMS L1: 3 jets.
    - Only one of the two leading jets allowed to be forward
  - HLT: four jets (one b-tagged), one pair with large m_{jj} $\Delta\eta_{jj}$
- Use of Quark/gluon-jet tagger
  - Distinguish quarks/gluon jet up to $|\eta|<4.7$
  - Reconstruct VBF jet pair "qq"
    1. Pair of least b-tagged jets
    2. Largest pseudorapidity separation.
- **Discriminate** VBF against QCD
  - ANN: MVA considering distributions and correlation of various discriminating variables.



Measurement Strategy:



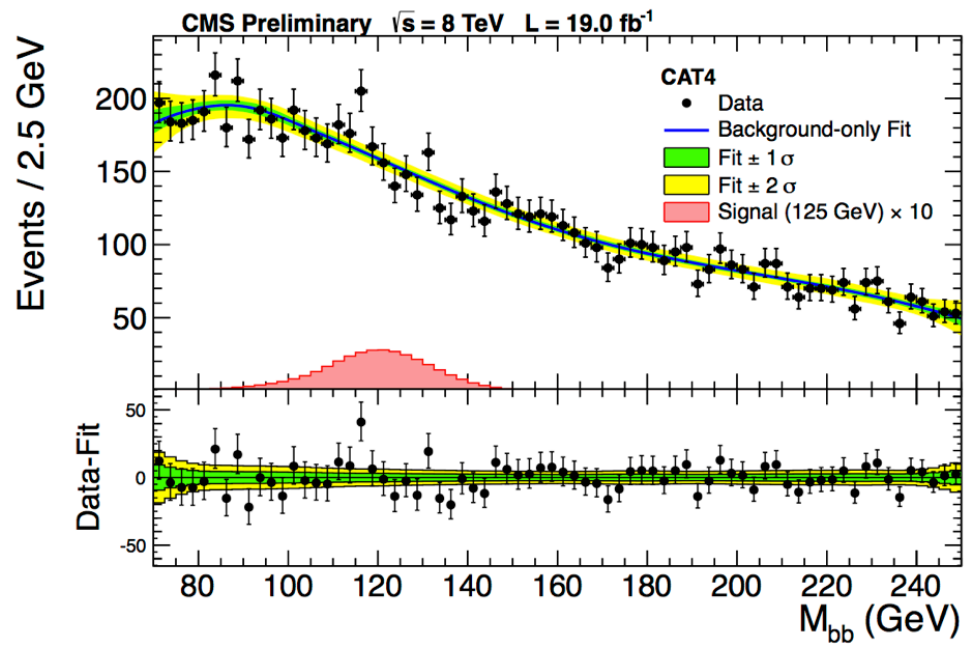
Categorization according to neural net output:

- No kinematic information of b-tagged jets
- **Minimal correlation** with M(bb)

Cat. 0	Cat. 1	Cat. 2	Cat. 3	Cat. 4
$ANN < 0.52$	$0.52 \leq ANN < 0.76$	$0.76 \leq ANN < 0.90$	$0.90 \leq ANN < 0.96$	$ANN \geq 0.96$

→ **Use M(bb) to extract signal:**

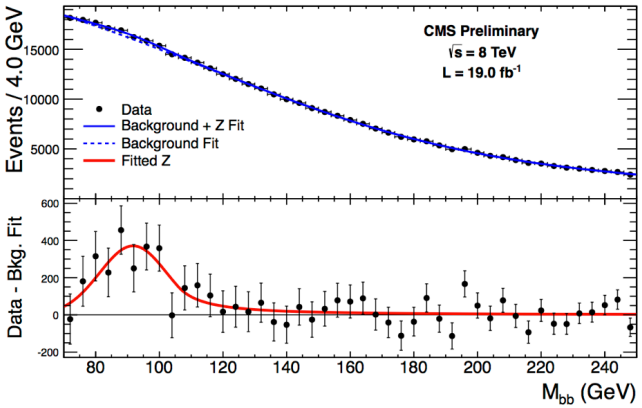
- In the 4 categories
- B-jet energy regression to improve resolution. (same as VH analysis)



VBF H(bb): Cross-check & **Results**

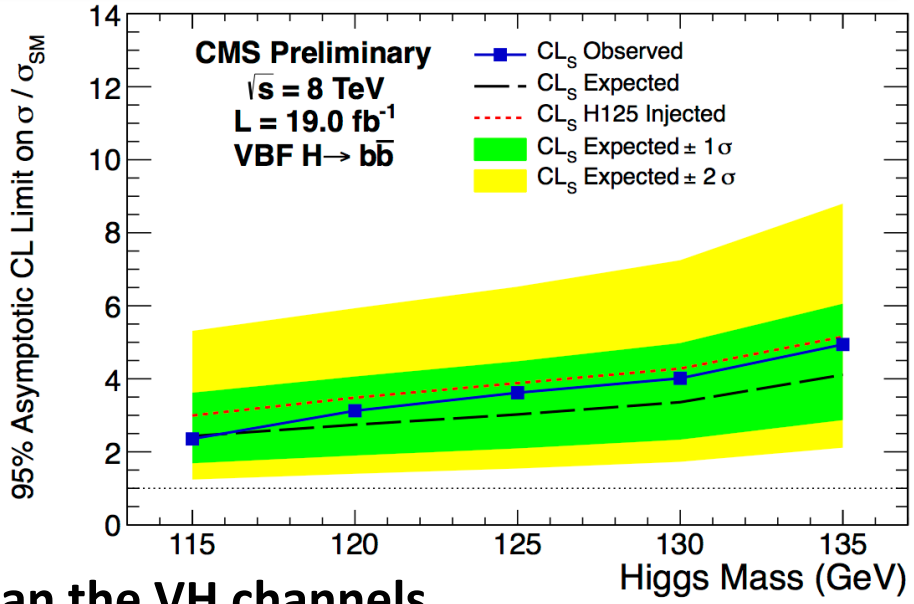
Cross-check : Search for **Z→bb** peak in the m(**bb**) spectrum.

- Same event pre-selection.
- Same fitting procedure.
- Result in agreement with the expectations:
  - observed (expected) significance: 8.0 (6.8)



Results:

- **Limit at 125 GeV:**
  - Observed: **3.6** x SM
  - Expected **3.0** x SM
- **Best fit**: μ = **0.7 ± 1.4**

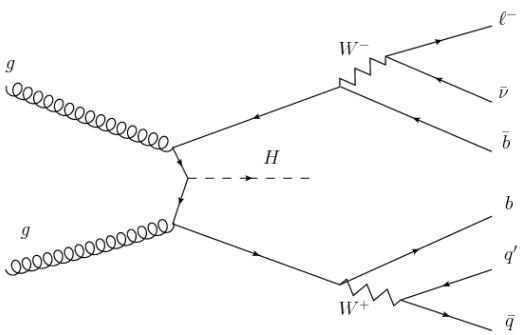


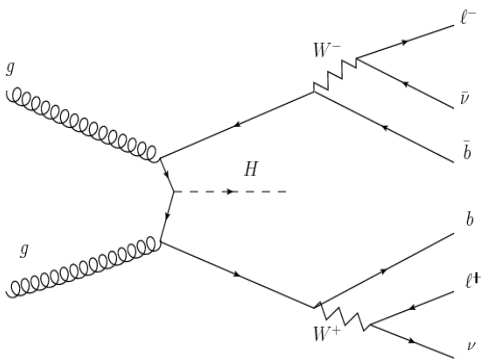
Less sensitivity than the VH channels

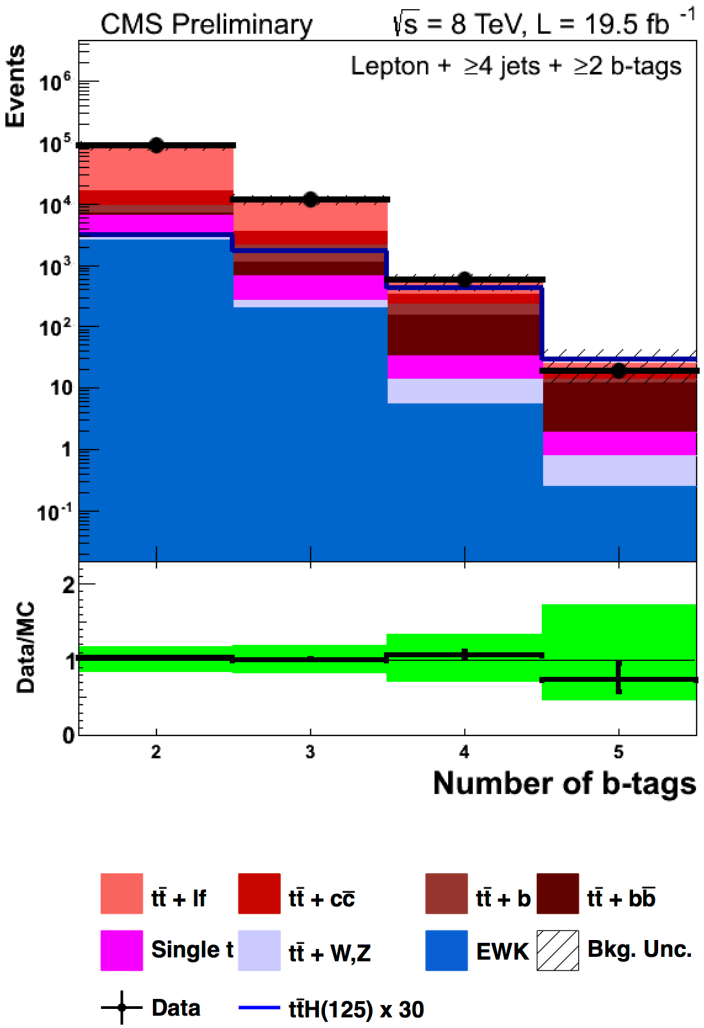
Analysis strategy:

$t\bar{t}$ system considered decay channels: 1-2 leptons

- **Categorization** depends on
  - number of jets
  - number of b-tagged jets
- Semi-leptonic: **7bins** (4 jets-3b's to 6 jets-2b's)
- Dileptonic: **3 bins** (3jets-2b's - ≥4 jets-2b's - ≥3b's)





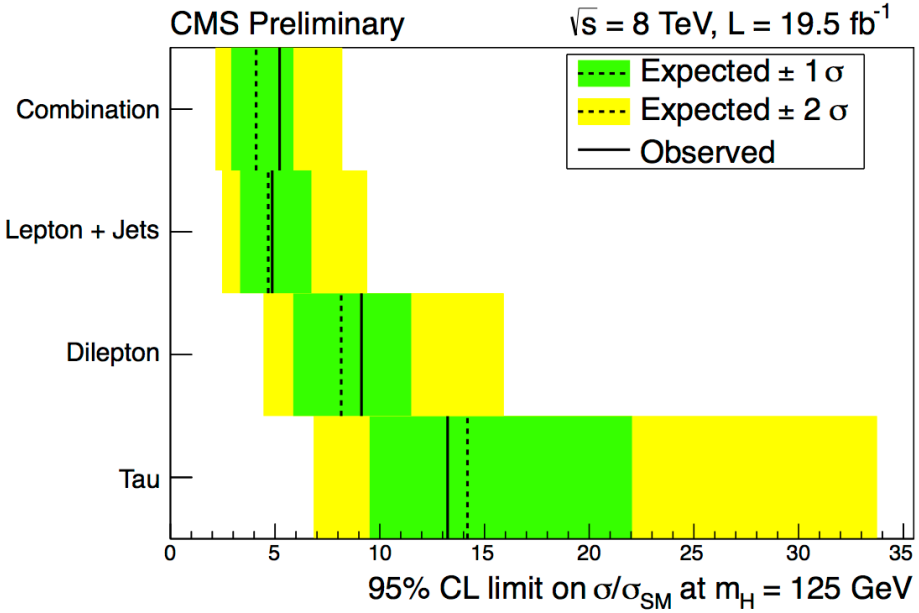


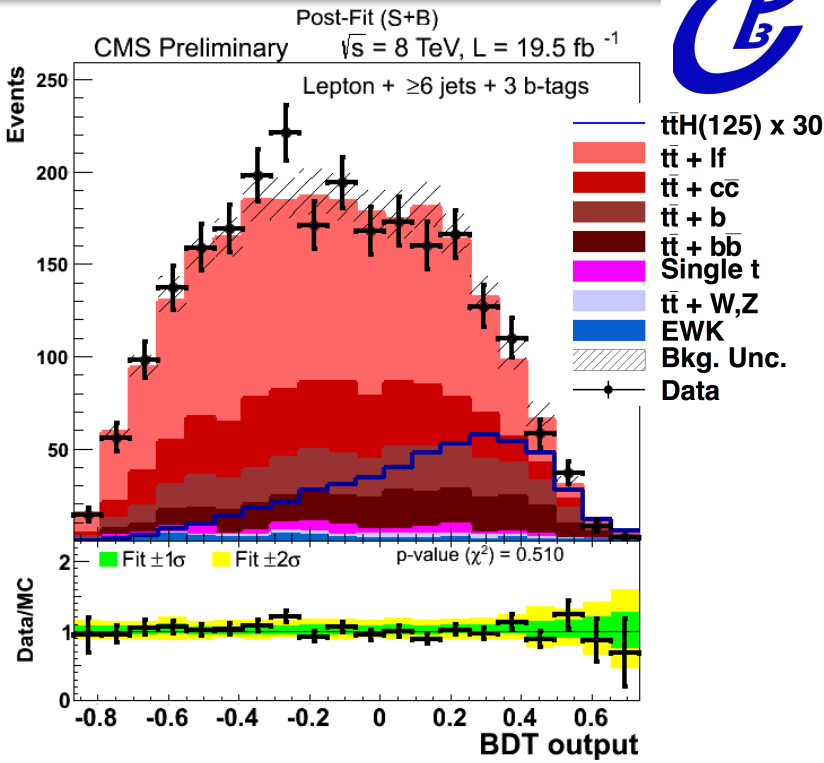
- Use of Boosted Decision Tree:
  - Trained separetely in each categories.
  - Object kinematic, event shape, CSV b-tag.
  - Up to 10 variables used (depends on cat.).

BDT and Results

Limits:

on σ/σ_{SM} for SM Higgs of m_H =125GeV at 95%Cl extracted from the BDT output





Dataset: 5.0 fb^{-1} (@ 7 TeV – 2011)
19.5 fb^{-1} (@ 8 TeV – 2012)

Channel	Exp (σ/σ_{SM})	Obs (σ/σ_{SM})
LJ	4.7	4.9
DIL	8.2	9.1
TAU	14.2	13.2
7 TeV LJ + DIL	6.7	5.9
$\gamma\gamma$	5.4	5.5
COMB	2.7	3.4

H→ bb Summary

- CMS Collaboration performed searches for a Standard Model Higgs boson decaying in a bottom quark pair.

- Three production modes investigated:
  - Associate production with vector boson (WH / ZH).
    → **Most sensitive analysis**
    [PhysRevD.89.012003] (5.0 + 19.5 fb^{-1})
  - Higgs production by vector boson fusion.
    [CMS-PAS-HIG-13-011] (19.5 fb^{-1})
  - Associate production with top quark pair.
    [CMS-PAS-HIG-13-019] (5.0 + 19.5 fb^{-1}) and [10.1007/JHEP05(2013)145] (5.0 + 5.1 fb^{-1})

CMS: H→bb in VH with significance of 2.1σ (2.1σ exp.)

H→bb is doable at the LHC and
we look forward towards measuring this decay unequivocally

Thank you

Next round with LHC run II …

Backup

Event Selection for VH analysis

Variable	$W(\ell\nu)H$	$W(\tau\nu)H$	$Z(\ell\ell)H$	$Z(\nu\nu)H$
$p_T(V)$	[100–130] [130–180] [>180]	[>120]	[50–100] [>100]	[100–130] [130–170] [>170]
$m_{\ell\ell}$	–	–	[75–105]	–
$p_T(j_1)$	>30	>30	>20	>60
$p_T(j_2)$	>30	>30	>20	>30
$p_T(jj)$	>100	>120	–	[>100] [>130] [>130]
$m(jj)$	<250	<250	[40–250] [< 250]	<250
E_T^{miss}	>45	>80	–	[100–130] [130–170] [> 170]
$p_T(\tau)$	–	>40	–	–
$p_T(track)$	–	>20	–	–
CSV_{max}	>0.40	>0.40	[>0.50] [>0.244]	>0.679
CSV_{min}	>0.40	>0.40	>0.244	>0.244
N_{aj}	–	–	–	[< 2] [–] [–]
$N_{a\ell}$	=0	=0	–	=0
$\Delta\phi(V,H)$	–	–	–	>2.0
$\Delta\phi(E_T^{miss}, jet)$	–	–	–	[>0.7] [>0.7] [>0.5]
$\Delta\phi(E_T^{miss}, E_T^{miss}(tracks))$	–	–	–	<0.5
E_T^{miss} significance	–	–	–	[>3] [–] [–]
$\Delta\phi(E_T^{miss}, \ell)$	$< \pi/2$	–	–	–

VH analysis: BDT variables

Variable
$p_T(j_1), p_T(j_2)$: transverse momentum of each Higgs boson daughter
$m(jj)$: dijet invariant mass
$p_T(jj)$: dijet transverse momentum
$p_T(V)$: vector boson transverse momentum (or E_T^{miss})
N_{aj}: number of additional jets (see caption)
CSV_{max}: value of CSV for the Higgs boson daughter with largest CSV value
CSV_{min}: value of CSV for the Higgs boson daughter with second largest CSV value
$\Delta\phi(V, H)$: azimuthal angle between V (or E_T^{miss}) and dijet
$\lvert\Delta\eta(jj)\rvert$: difference in η between Higgs boson daughters
$\Delta R(jj)$: distance in η–ϕ between Higgs boson daughters
$\Delta\theta_{pull}$: color pull angle [43]
$\Delta\phi(E_T^{miss}, jet)$: azimuthal angle between E_T^{miss} and the closest jet (only for $Z(\nu\nu)H$)
$maxCSV_{aj}$: maximum CSV of the additional jets in an event (only for $Z(\nu\nu)H$ and $W(\ell\nu)H$)
$minΔR(H, aj)$: minimum distance between an additional jet and the Higgs boson candidate (only for $Z(\nu\nu)H$ and $W(\ell\nu)H$)
Invariant mass of the VH system (only for $Z(\ell\ell)H$)
Cosine of the angle between the direction of the V boson in the rest frame of the VH system and the direction of the VH system in the laboratory frame (only for $Z(\ell\ell)H$)
Cosine of the angle between the direction of one of the leptons in the rest frame of the Z boson and the direction of the Z boson in the laboratory frame (only for $Z(\ell\ell)H$)
Cosine of the angle between the direction of one of the jets in the rest frame of the reconstructed Higgs boson and the direction of the reconstructed Higgs boson in the laboratory frame (only for $Z(\ell\ell)H$)

Z/W + H(bb): Sources of systematic uncertainties

For the BDT analysis:

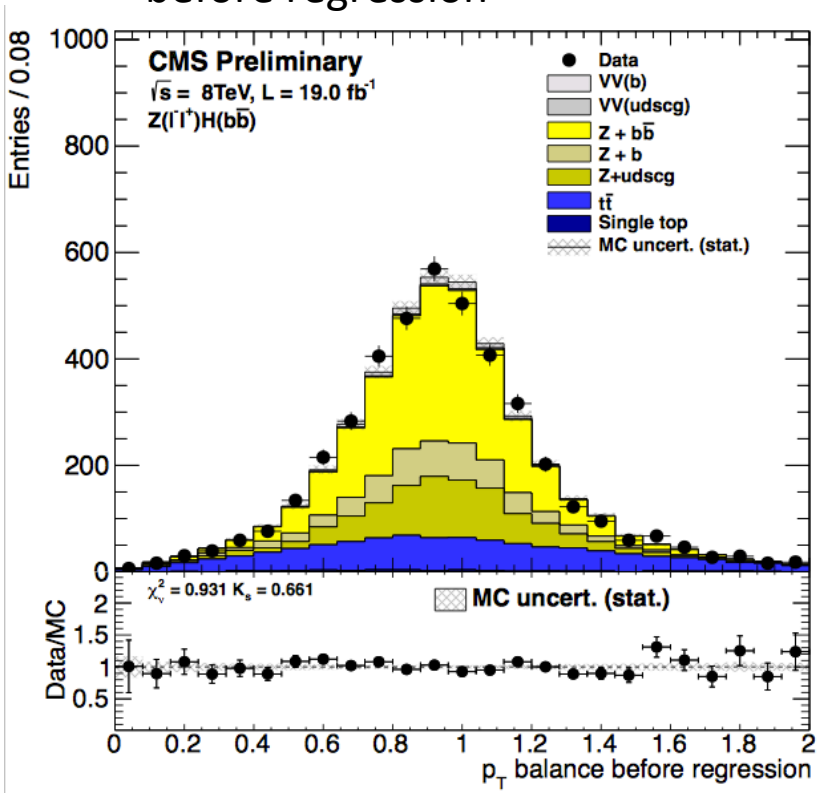
Source	Type	Event yield uncertainty range (%)	Individual contribution to μ uncertainty (%)	Effect of removal on μ uncertainty (%)
Luminosity	norm.	2.2–2.6	<2	<0.1
Lepton efficiency and trigger (per lepton)	norm.	3	<2	<0.1
$Z(\nu\nu)H$ triggers	shape	3	<2	<0.1
Jet energy scale	shape	2–3	5.0	0.5
Jet energy 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)	norm.	4	3.9	0.3
Signal cross section (p_T boost, EW/QCD)	norm.	2/5	3.9	0.3
Monte Carlo statistics	shape	1–5	13.3	3.6
Backgrounds (data estimate)	norm.	10	15.9	5.2
Single-top-quark (simulation estimate)	norm.	15	5.0	0.5
Dibosons (simulation estimate)	norm.	15	5.0	0.5
MC modeling (V+jets and $t\bar{t}$)	shape	10	7.4	1.1

- Dominant systematic uncertainties:
  - **b-tagging.**
  - **MC statistic.**
  - **Background** measured in the control regions.

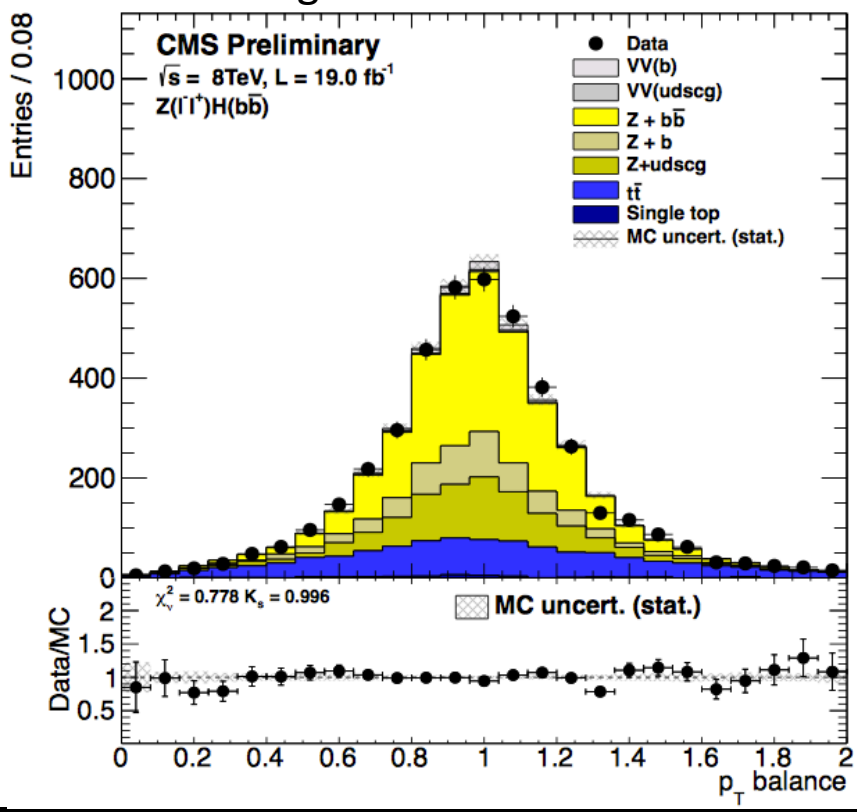
Z/W + H(bb): Regression data validation

Distribution of the ratio between the pT(jj) and the pT of the dilepton system on data versus MC.

before regression



After regression



VBF systematics

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	$\pm 50\%$

VBF Event yields

Expected number of events in the *mbb* interval [70,250GeV], in each ANN output category.

Sample/ANN range	< 0.52	$0.52 - 0.76$	$0.76 - 0.90$	$0.90 - 0.96$	> 0.96
QCD	1.9e+6	3.2e+5	1.1e+5	2.7e+4	8.7e+3
$Z + jets$	5531	1222	531	124	54
$t\bar{t}$	12730	1032	190	33	15
t	1839	383	128	25	10
$\bar{t}$	895	226	73	15	7
$W + jets$	2033	226	50	4	<1
VBF $M_H(125)$	66	79	84	49	33
GF $M_H(125)$	94	37	18	6	2

TTH

tth BDT variables Per categories.

Lepton+jets

4 jets, 3 b-tags	4 jets, 4 b-tags
jet 1 p_T jet 2 p_T jet 3 p_T jet 4 p_T M3 $\sum p_T$(jets,lepton,MET) HT lowest CSV (tags) MHT MET	jet 1 p_T jet 2 p_T jet 4 p_T HT $\sum p_T$(jets,lepton,MET) M3 ave CSV (tags) second-highest CSV (tags) third-highest CSV (tags) lowest CSV (tags)
5 jets, 3 b-tags	**5 jets, $\geq$ 4 b-tags**
jet 1 p_T jet 2 p_T jet 3 p_T jet 4 p_T $\sum p_T$(jets,lepton,MET) (Σ jet p_T)/(Σ jet E) HT ave CSV (tags) third-highest CSV (tags) fourth-highest CSV (jets)	max $\Delta\eta$ (tag, ave jet η) $\sum p_T$(jets,lepton,MET) (Σ jet p_T)/(Σ jet E) ave ΔR(tag,tag) ave CSV (tags) dev from ave CSV (tags) second-highest CSV (tags) third-highest CSV (tags) lowest CSV (tags) ttbb/ttH BDT

$\geq$ 6 jets, 2 tags	$\geq$ 6 jets, 3 tags	$\geq$ 6 jets, $\geq$ 4 tags
$\sum p_T$(jets,lepton,MET) HT mass(lepton,closest tag) max $\Delta\eta$ (jet, ave jet η) min ΔR(lepton,jet) H_2 sphericity (Σ jet p_T)/(Σ jet E) third-highest CSV (jets) fourth-highest CSV (jets)	H_0 sphericity (Σ jet p_T)/(Σ jet E) max $\Delta\eta$ (jet, ave jet η) $\sum p_T$(jets,lepton,MET) ave CSV (tags) second-highest CSV (tags) third-highest CSV (tags) fourth-highest CSV (jets) ttbb/ttH BDT	(Σ jet p_T)/(Σ jet E) ave ΔR(tag,tag) product($\Delta\eta$(leptonic top, bb), $\Delta\eta$(hadronic top, bb)) closest tag mass max $\Delta\eta$ (tag, ave tag η) ave CSV (tags) third-highest CSV (tags) fourth-highest CSV (tags) best Higgs boson mass ttbb/ttH BDT

Event variables used in additional ttbb/ttH BDTs

5 jets, $\geq$ 4 tags	$\geq$ 6 jets, 3 tags	$\geq$ 6 jets, $\geq$ 4 tags
ave ΔR(tag,tag) max $\Delta\eta$ (tag, ave tag η) (Σ jet p_T)/(Σ jet E) tagged dijet mass closest to 125 H_1 H_3 $\sum p_T$(jets,lepton,MET) fourth-highest CSV (tags) aplanarity MET	tagged dijet mass closest to 125 (Σ jet p_T)/(Σ jet E) $\sqrt{\Delta\eta(t^{lep}, bb) \times \Delta\eta(t^{had}, bb)}$ H_1 H_3 M3 max $\Delta\eta$ (tag, ave tag η) max $\Delta\eta$ (tag, ave jet η) max $\Delta\eta$ (jet, ave jet η) abs $\Delta\eta$ (hadronic top, bb) abs $\Delta\eta$ (leptonic top, bb) sphericity aplanarity min ΔR(tag,tag) jet 3 p_T	H_3 ave ΔR(tag,tag) closest tagged dijet mass sphericity max $\Delta\eta$ (tag, ave jet η) max $\Delta\eta$ (tag, ave tag η) mass(lepton,jet,MET) (Σ jet p_T)/(Σ jet E) abs $\Delta\eta$ (leptonic top, bb) abs $\Delta\eta$ (hadronic top, bb) $\sqrt{\Delta\eta(t^{lep}, bb) \times \Delta\eta(t^{had}, bb)}$ ave CSV (tags) best ΔR(b,b) best Higgs boson mass median inv. mass (tag pairs)

tth BDT variables
Per categories.

Dilepton

3 jets, 2 b-tags	$\geq$4 jets, 2 b-tags	$\geq$3 jets, $\geq$3 b-tags
ave CSV (tags)	ave CSV (non-tags)	ave CSV (tags)
min ΔR(jet,jet)	min ΔR(jet,jet)	min ΔR(jet,jet)
$\sum p_{\mathrm{T}}$(jets,leptons)	$\sum p_{\mathrm{T}}$(jets,leptons)	$\sum p_{\mathrm{T}}$(jets,leptons)
ave CSV (non-tags)	Number of jets	Number of HiggsLike dijet 15
	HiggsLike dijet mass	HiggsLike dijet mass
	HiggsLike dijet mass2	HiggsLike dijet mass2

ttH expected Event Yields

L+ jets

	≥6 jets 2 b-tags	4 jets 3 b-tags	5 jets 3 b-tags	≥6 jets 3 b-tags	4 jets 4 b-tags	5 jets ≥4 b-tags	≥6 jets ≥4 b-tags
$t\bar{t}H(125)$	33.4 ± 8.1	14.0 ± 3.0	21.1 ± 4.5	23.1 ± 5.5	1.8 ± 0.5	5.2 ± 1.4	8.3 ± 2.3
$t\bar{t}$+lf	7650 ± 2000	4710 ± 820	2610 ± 530	1260 ± 340	74 ± 30	79 ± 34	71 ± 36
$t\bar{t}$+b	530 ± 300	350 ± 190	360 ± 200	280 ± 160	21 ± 12	29 ± 17	33 ± 20
$t\bar{t} + b\bar{b}$	220 ± 120	99 ± 52	158 ± 85	200 ± 110	13.1 ± 7.3	38 ± 21	78 ± 47
$t\bar{t} + c\bar{c}$	1710 ± 1110	440 ± 230	520 ± 290	470 ± 280	19 ± 11	32 ± 18	52 ± 31
$t\bar{t}V$	99 ± 27	16.2 ± 3.8	23.9 ± 5.7	28.8 ± 7.4	1.1 ± 0.4	2.5 ± 0.7	5.8 ± 1.8
Single t	264 ± 54	235 ± 41	116 ± 22	55 ± 14	3.4 ± 1.6	10.3 ± 5.3	7.3 ± 3.1
V+jets	160 ± 110	122 ± 95	44 ± 38	29 ± 27	2.1 ± 2.4	1.9 ± 1.7	1.2 ± 1.3
Diboson	5.9 ± 1.6	6.3 ± 1.4	2.4 ± 0.7	1.0 ± 0.4	0.3 ± 0.2	0.1 ± 0.1	0.2 ± 0.1
Total bkg	10630 ± 2790	5970 ± 1060	3830 ± 790	2310 ± 620	133 ± 44	193 ± 62	249 ± 90
Data	10724	5667	3983	2426	122	219	260

Dilepton

	3 jets + 2 b-tags	≥4 jets + 2 b-tags	≥3 b-tags
$t\bar{t}H(125)$	7.7 ± 1.4	16.1 ± 3.1	11.2 ± 2.5
$t\bar{t}$+lf	7460 ± 1060	3190 ± 680	289 ± 83
$t\bar{t}$+b	189 ± 97	172 ± 93	149 ± 82
$t\bar{t} + b\bar{b}$	38 ± 20	58 ± 31	80 ± 44
$t\bar{t} + c\bar{c}$	480 ± 260	510 ± 300	147 ± 79
$t\bar{t}V$	30.2 ± 6.3	54 ± 12	11.9 ± 2.9
Single t	229 ± 35	97 ± 16	17.3 ± 5.1
V+jets	350 ± 130	151 ± 66	40 ± 23
Diboson	10.4 ± 1.7	3.1 ± 0.6	0.7 ± 0.4
Total bkg	8770 ± 1250	4230 ± 850	740 ± 190
Data	9060	4616	774