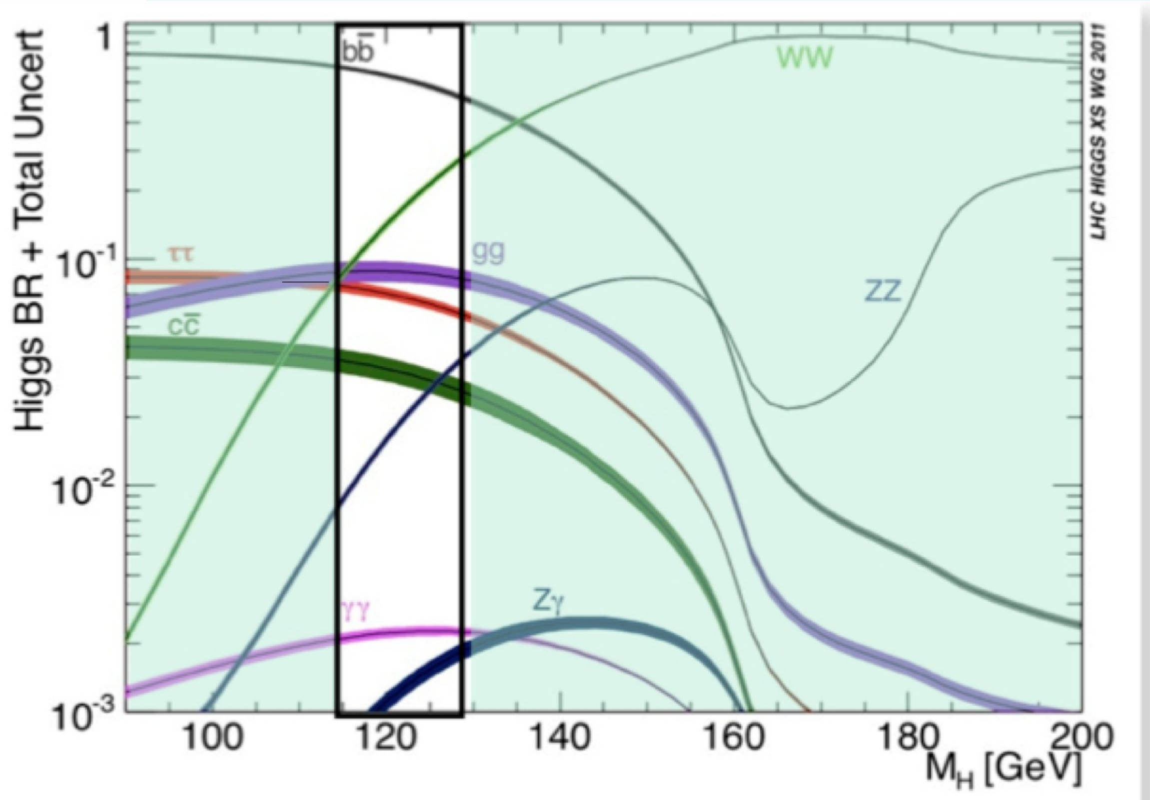
Search for the SM Higgs Boson decaying to $b\bar{b}$ at CMS

David Lopes Pegna (Princeton University-LPC FNAL)
On behalf of the CMS Collaboration

LHC Physics Workshop, Chicago
14 November 2012

If it is the SM Higgs....



@125 GeV:

BR($H \rightarrow bb$) ~ 58%

BR($H \rightarrow WW$) ~ 22%

BR($H \rightarrow \tau\tau$) ~ 6%

BR($H \rightarrow ZZ^*$) ~ 3%

BR(H-> $\gamma\gamma$) ~0.22%

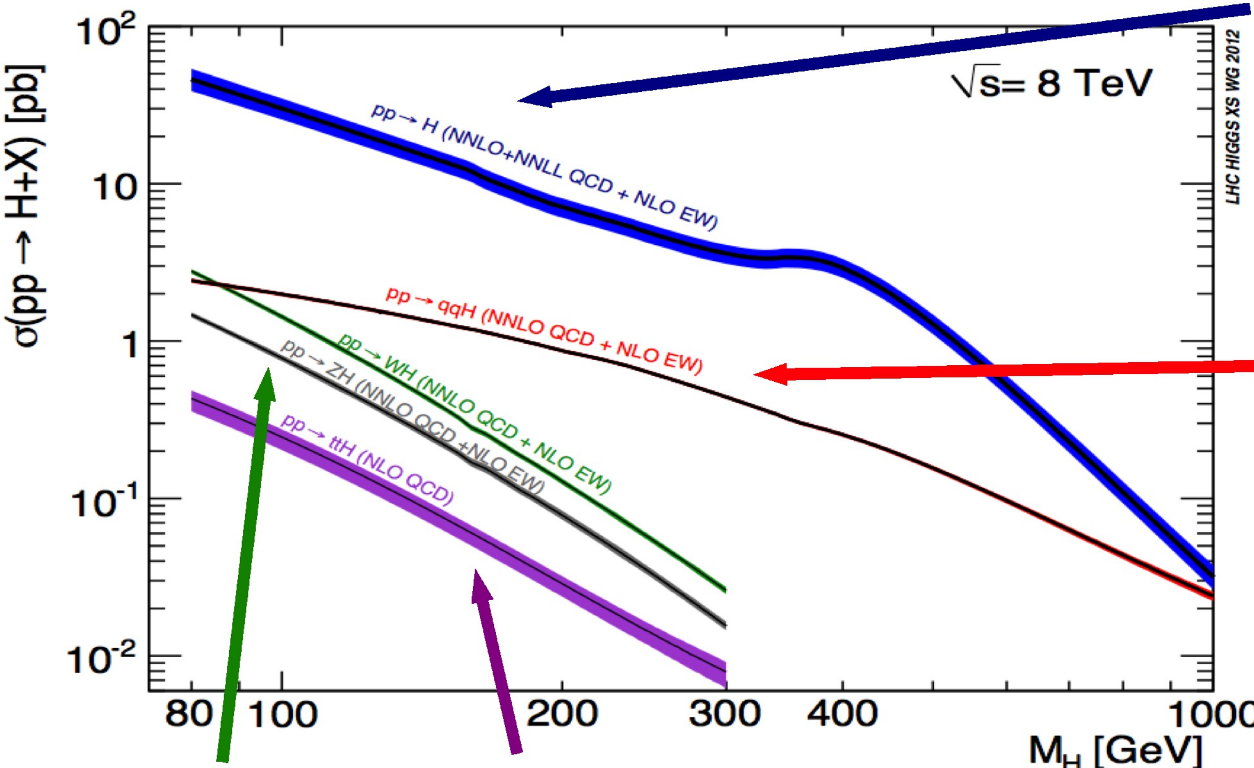
Our goal now is to confirm or exclude it's the Standard Model Higgs

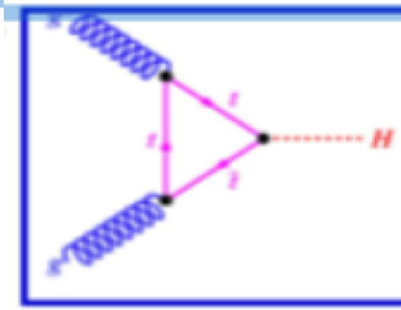
→ need complementary information from as many channels as possible

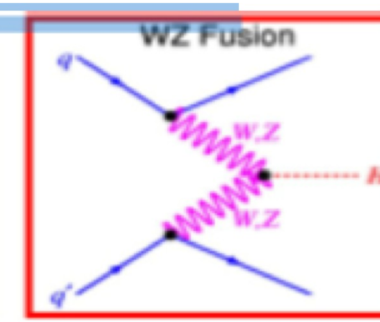
→ $H \rightarrow b\bar{b}$ largest Branching Ratio by far below 130 GeV

→ BR($H \rightarrow gg$) + BR($H \rightarrow c\bar{c}$) ~13%, w/o $H \rightarrow b\bar{b}$, ¾ of the width would be invisible!

Production Modes







gg fusion

Vector Boson Fusion

H → bb

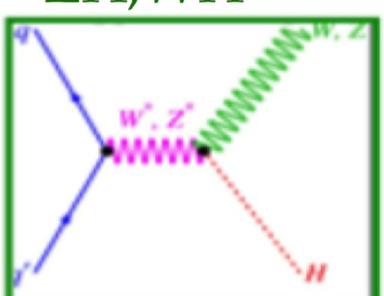
Overwhelming QCD Background makes gg → H → $b\bar{b}$ impossible

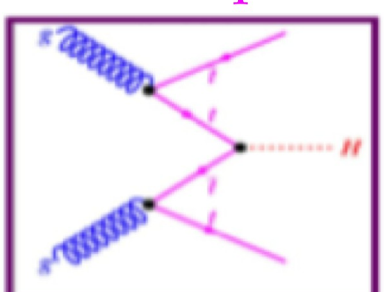
Look at associated Production

ttH: even less rate, but Accessible with large ∫Ldt

Higgs-strahlung: ZH,WH

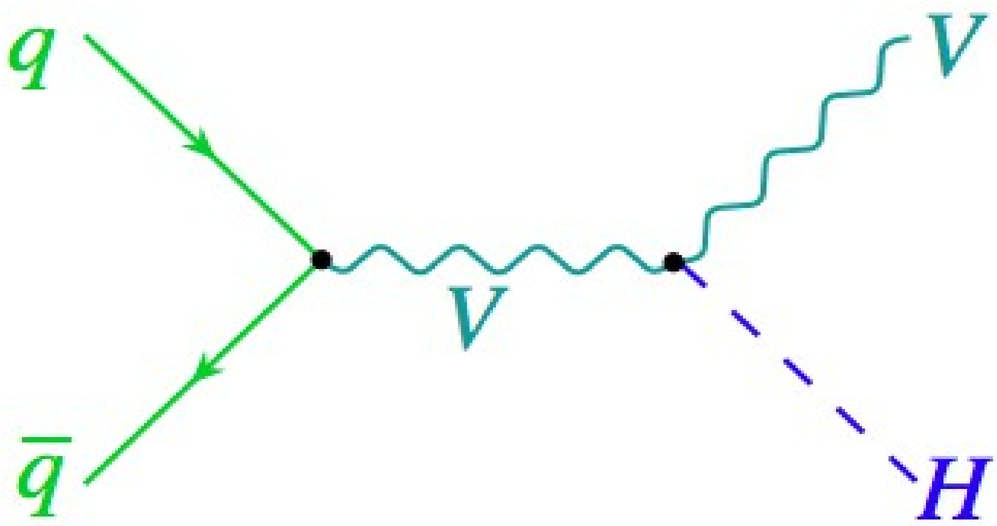
Associated production With ttbar pair





Updated for HCP (HIG-12--044) on 12.1 fb^{-1} @ $\sqrt{s}$= 8 TeV

http://cdsweb.cern.ch/record/1493618?ln=en

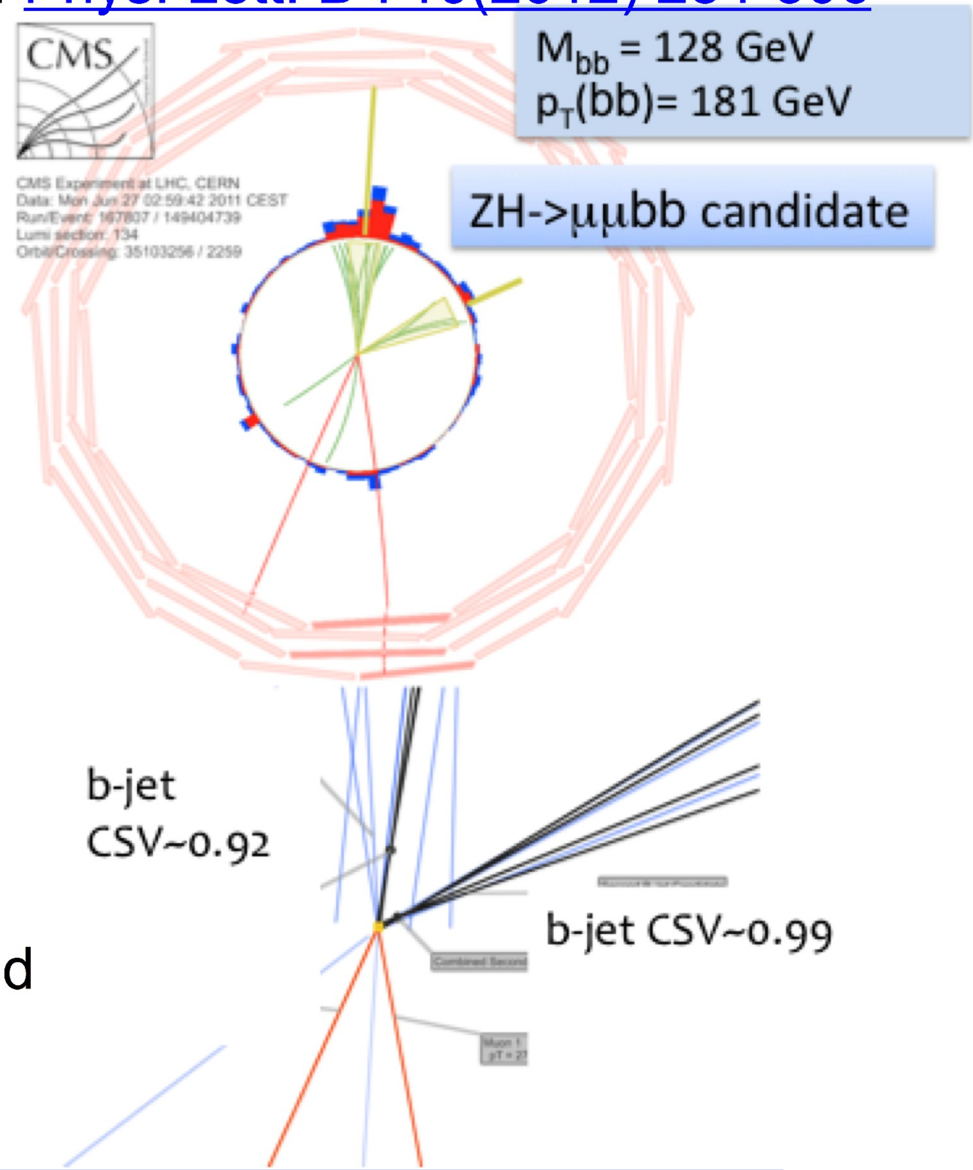


$VH \rightarrow Vb\bar{b}$
$V \rightarrow ll, l\nu, \nu\nu$

Most sensitive channel with b in final states
Intriguing excess in the Tevatron VH → $b\bar{b}$ analysis

VH Analysis in a nutshell

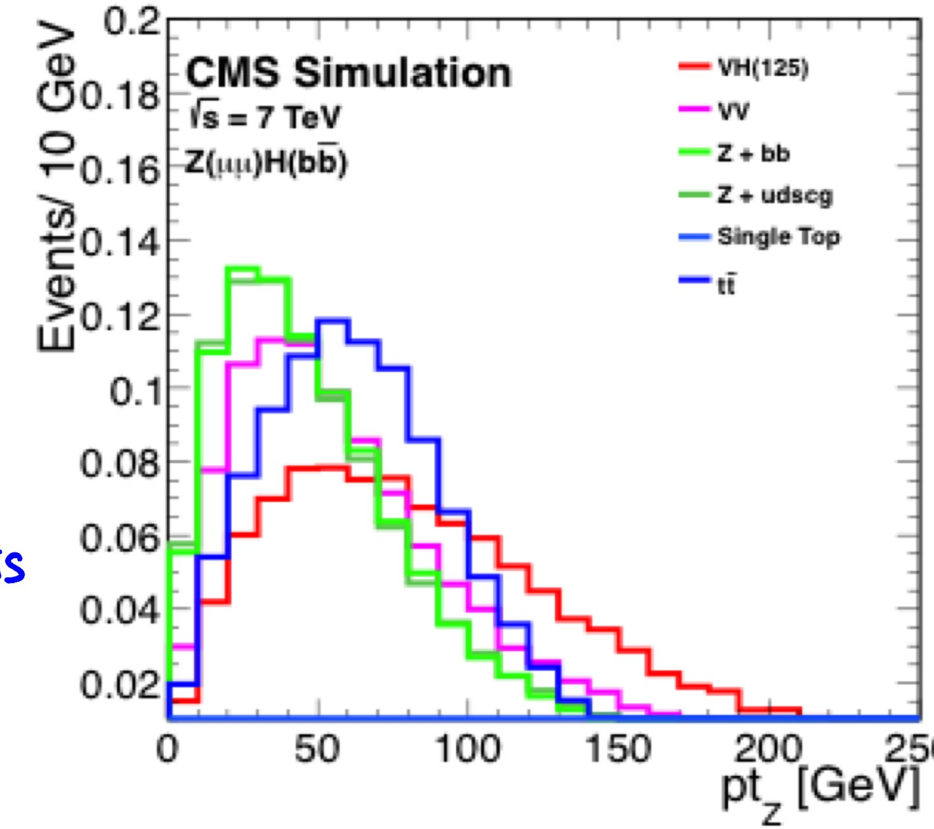
- First CMS Vhbb analysis on 7 TeV data: Phys. Lett. B 710(2012) 284-306
- 5 modes under study:
  Z(ll)H, W(lν)H, Z(νν)H, l= e,μ
- Boosted analysis (better S/B):
  → Require high momentum vector
  boson and 2-b tagged jet,
  back-to -back
- Use Data control regions to constrain
  most important backgrounds
  (V+jet, Light or Heavy, ttbar)
- Boosted Decision Tree algorithm (BDT)
  to discriminate signal versus background



Event Categories

- Boost topology requirement is the name of the game
  → original proposal by Butterworth et al. in 2008 in the context of substructure analysis

- Split events in two categories based on $p_T(V)$
  New since 2011 Analysis
  → increase acceptance in lower boost region, backgrounds still manageable
  → Lower threshold possible in Z(ll)H due to additional ttbar suppression



New since ICHEP 2012

- Further increase acceptance with a loose b-tag category (one tight, one loose tag) In WH and Z(νν)H

Channel	Medium Boost	High Boost	loose b-tag
$Z(\ell\ell)H$	$50 < p_T(Z) < 100$	$p_T(Z) > 100$	No
$W(\ell\nu)H$	$120 < p_T(W) < 170$	$p_T(W) > 170$	Yes
$Z(\nu\nu)H$	$130 < p_T(Z) < 170$	$p_T(Z) < 170$	Yes

B-jet energy Regression

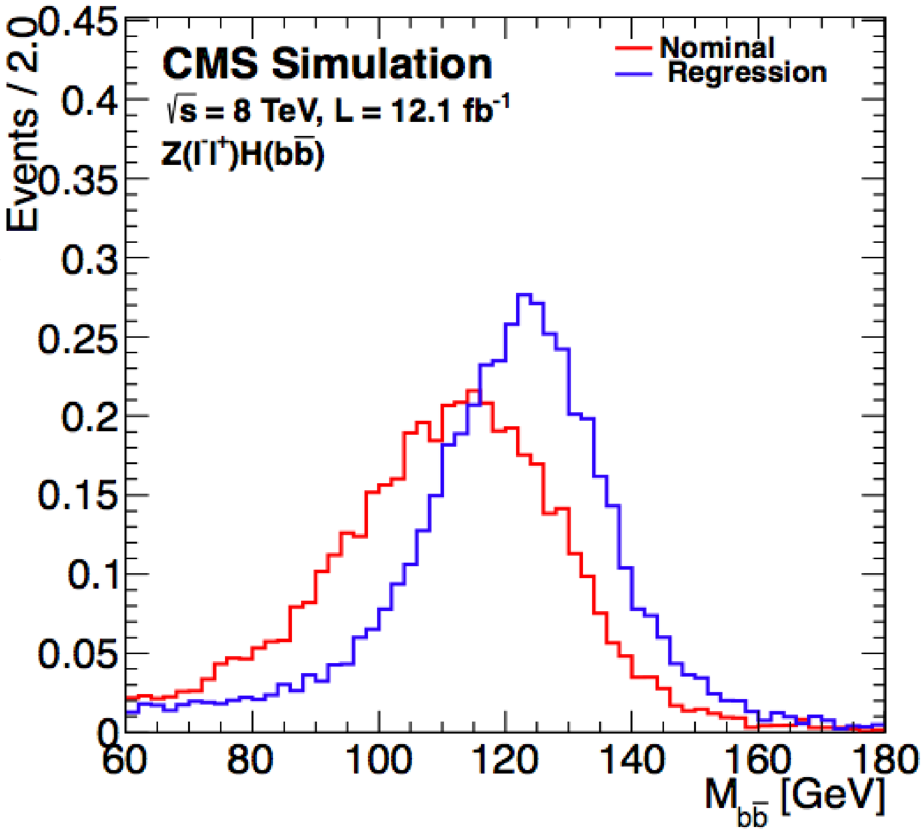
- Implementation based on NN method developed at CDF for b-jet energy corrections: http://arxiv.org/pdf/1107.3026.pdf

New since 2011 Analysis

- Multivariate Regression (BDT) trained on VH signal events using several (b)-jet variables

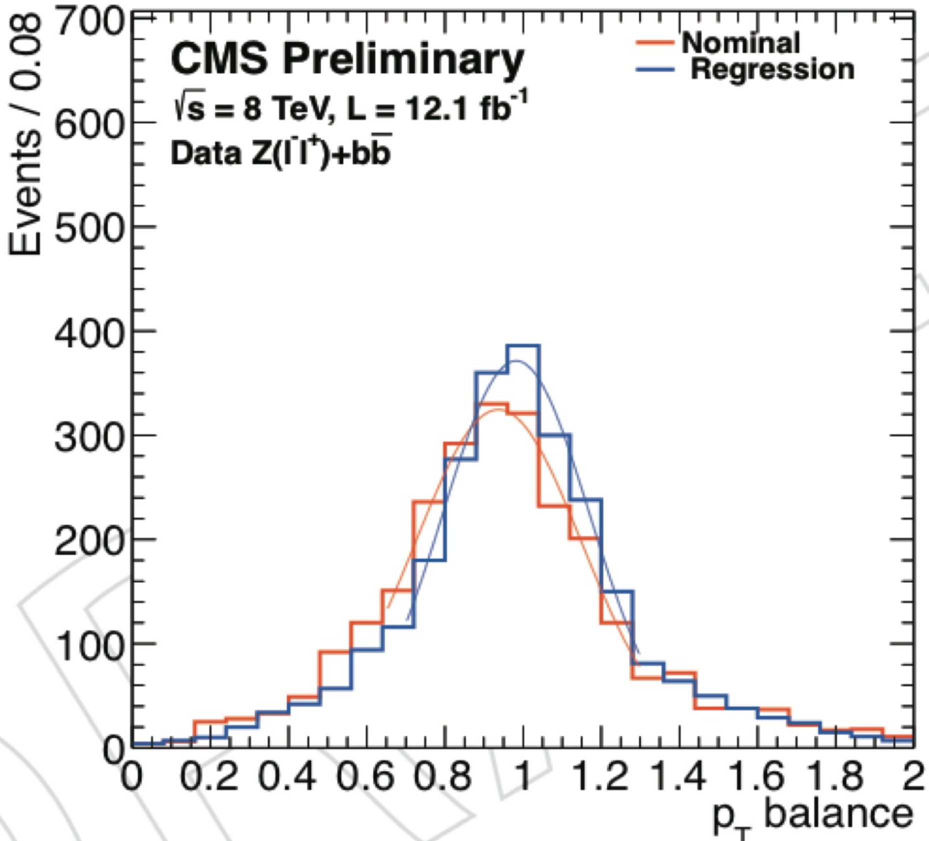
- Regression mainly corrects b-semileptonic decays, lepton-tag variables included in training   New since ICHEP 2012

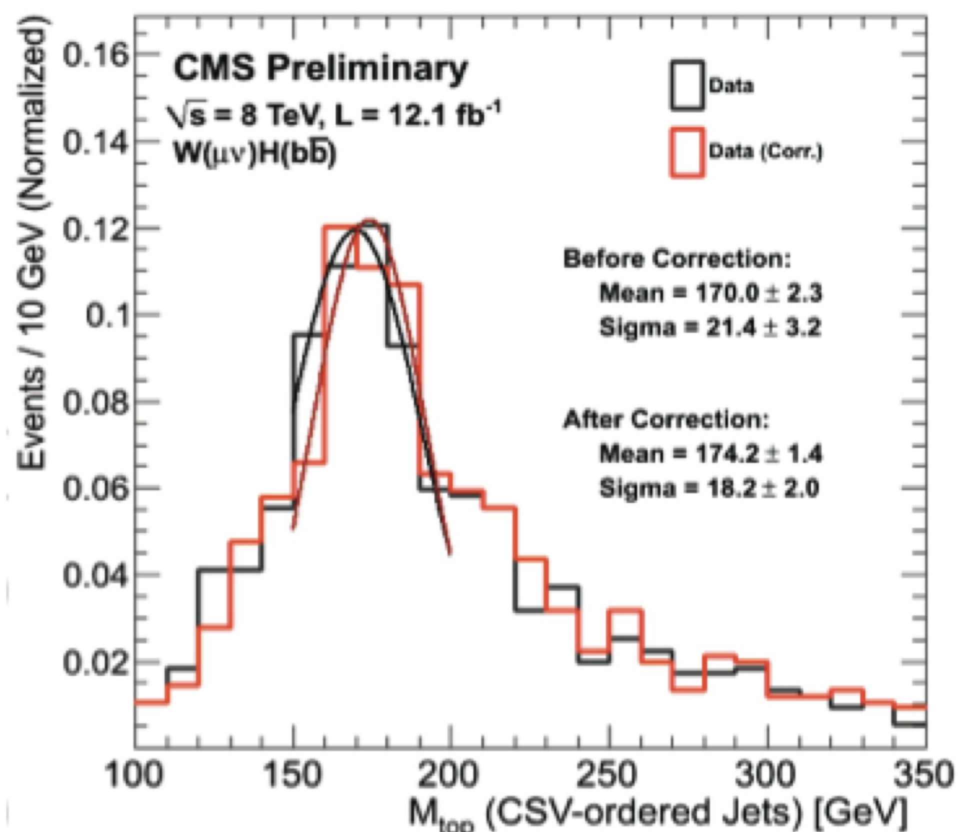
- Improvements in resolution of the order of 20% for Z(ll)H, 15% for W(lν)H and Z(νν)



Regression Validation

- Extensively validated on simulation and Data Control Regions
  → check of data/MC agreement of variables input to the regression in all control regions
  → p_T balance in Z(ll)+$b\bar{b}$
  → full reconstruction of top mass in ttbar and Single Top samples

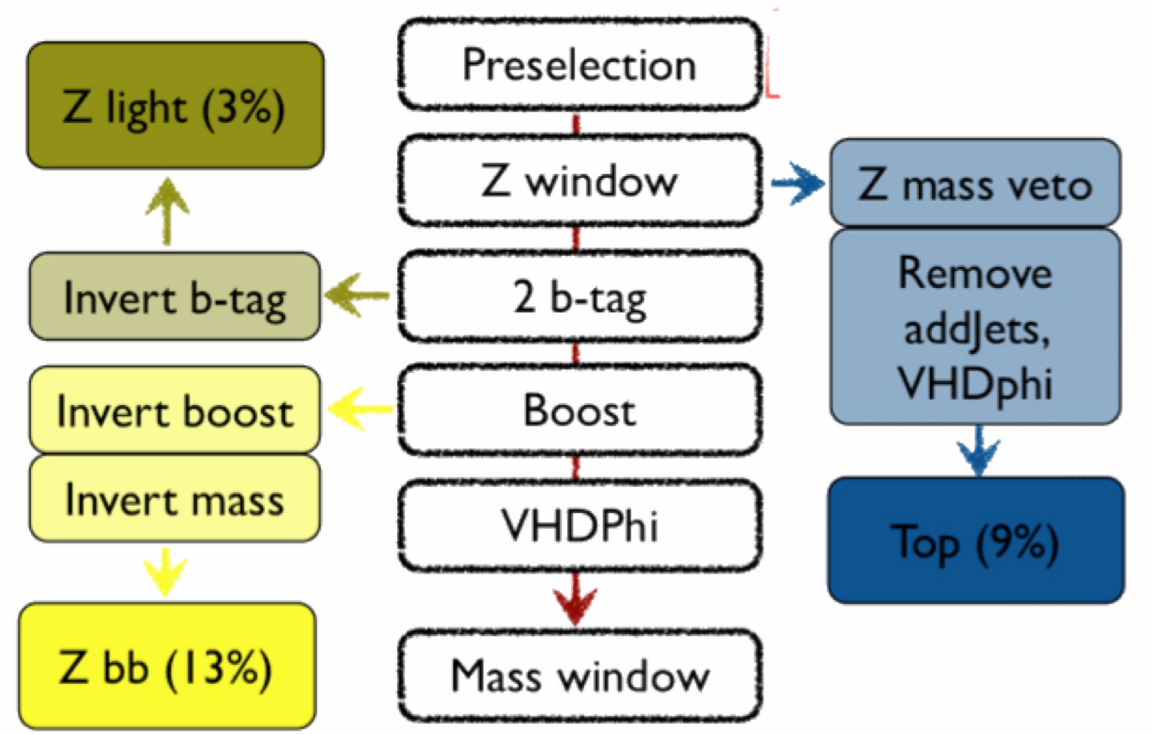




Background Control Regions

- Define several CRs enriched in different background components
- Kinematic selection as close as possible to the one for the Signal Region (SR)
- Scale Factors (SF) for V+light jets, ttbar and V+heavy jets determined simultaneously in each mode from simultaneous binned Maximum Likelihood fit

Example: Zee control region definition



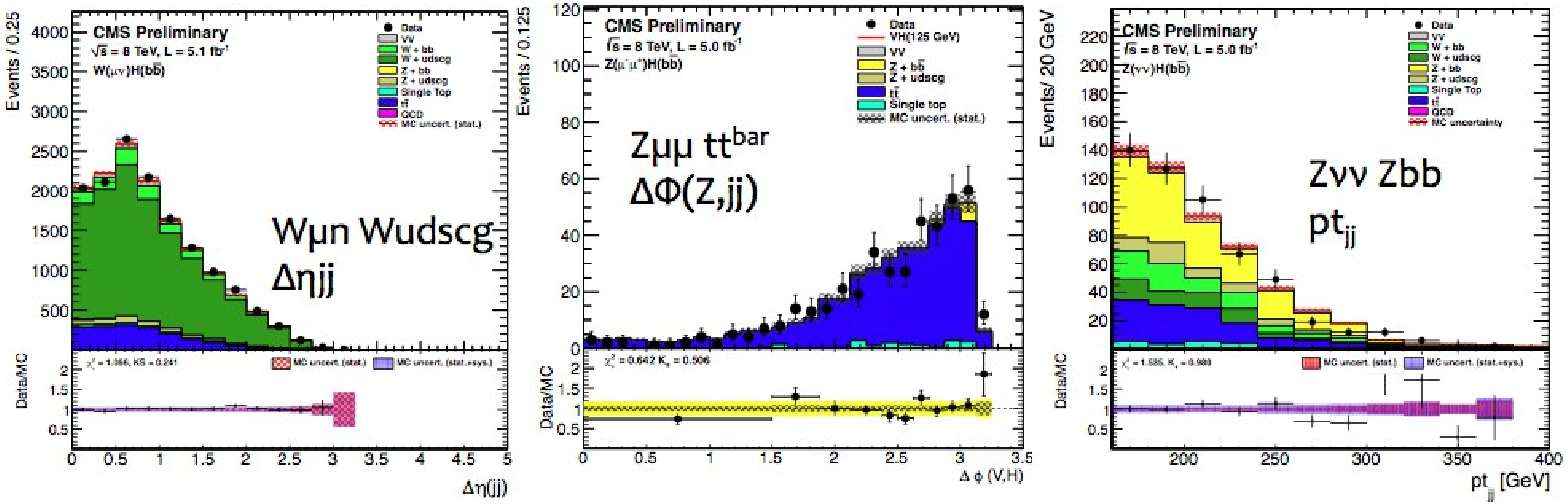
New since 2011 Analysis

- Renormalize background estimates in Signal region based on Scale Factors: **B(SR) = SF(CR) * B_{MC}(SR)**

Background Control Regions

- Example of data/MC agreement in the Control Regions for variables used in the analysis
  → Many more in backup

- Calibrate most important backgrounds, test analysis robustness



BDT: Event Selection

- Preselection cuts on:
  → boost topology
  → b-tag enriched

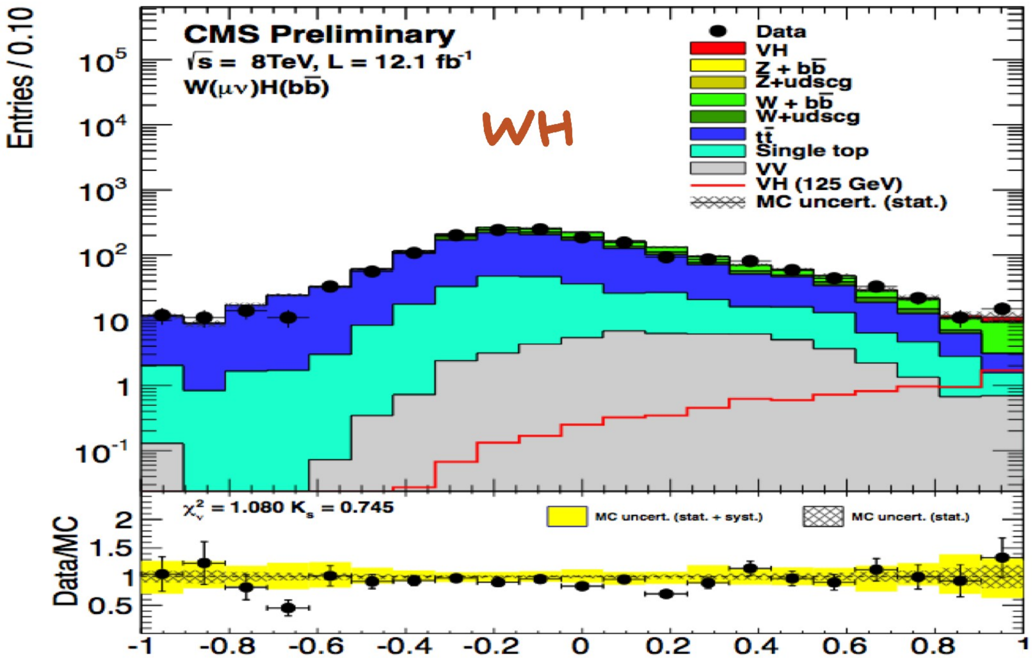
Variable	W(ℓν)H	Z(ℓℓ)H	Z(νν)H
$m_{\ell\ell}$	–	[75 – 105]	–
$p_T(j_1)$	> 30	> 20	> 60
$p_T(j_2)$	> 30	> 20	> 30
$p_T(jj)$	> 120	–	> 130
$M(jj)$	< 250	[80 – 150] (–)	< 250
$p_T(V)$	[120 – 170] (> 170)	[50 – 100] (> 100)	–
CSV_{max}	> 0.40	> 0.50 (> 0.244)	> 0.679
CSV_{min}	> 0.40	> 0.244	> 0.244
CSV_{min}^{loose}	– (< 0.40)	–	– (< 0.244)
N_{al}	= 0	–	= 0
E_T^{miss}	> 45 (elec)	–	[130 – 170] (> 170)
$\Delta\phi$(pfMET, J)	–	–	> 0.5
$\Delta\phi$(pfMET, trkMET)	–	–	< 0.5
$\Delta\phi$(V, H)	–	–	> 2.0

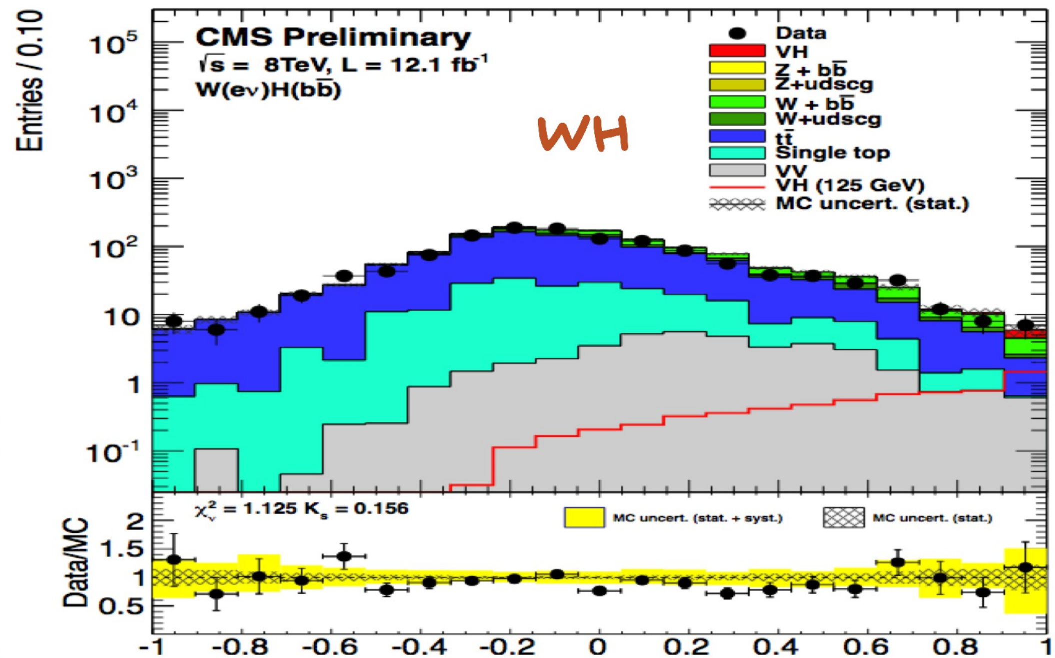
- Set of variables in the BDT largely overlapping with 2011 analysis

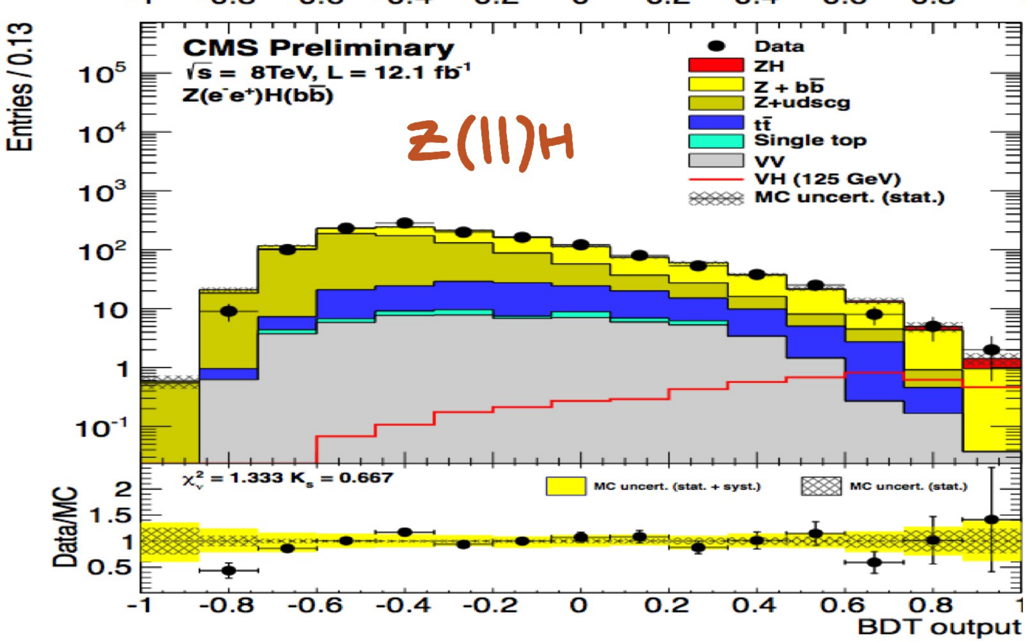
Variable
p_{Tj}: transverse momentum of each Higgs daughter
$m(jj)$: dijet invariant mass
$p_T(jj)$: dijet transverse momentum
$p_T(V)$: vector boson transverse momentum (or E_T^{miss})
CSV_{max}: value of CSV for the Higgs daughter with largest CSV value
CSV_{min}: value of CSV for the Higgs daughter 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 daughters
$\Delta R(jj)$: distance in η–ϕ between Higgs daughters
N_{aj}: number of additional jets
$\Delta\phi(E_T^{miss}, jet)$: azimuthal angle between E_T^{miss} and the closest jet (only for Z(νν)H)
$\Delta\theta_{pull}$: color pull angle

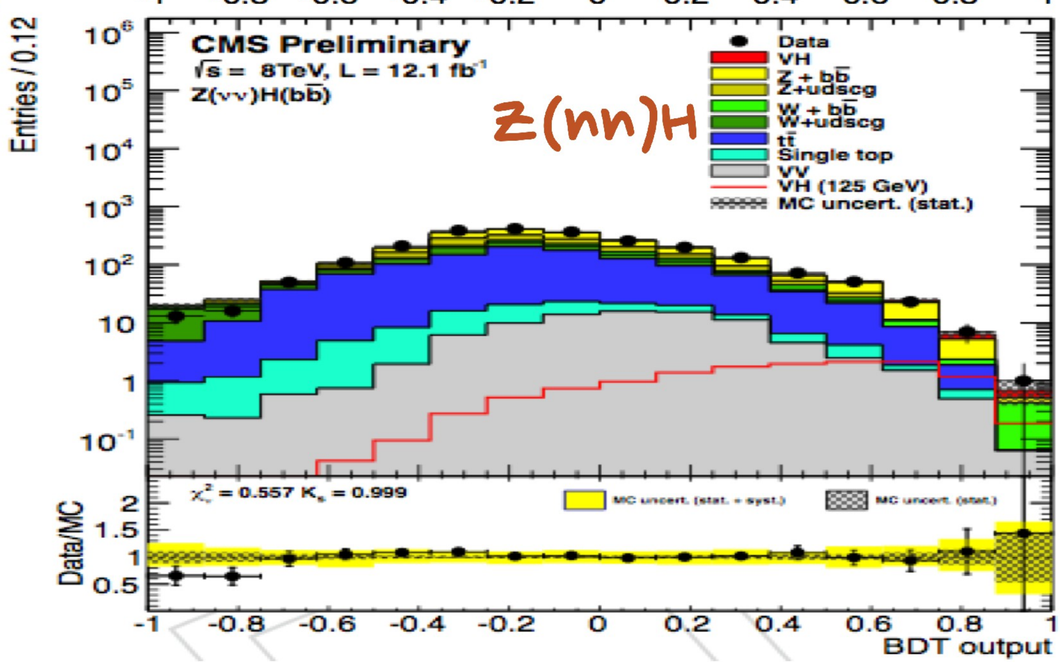
- Limit extraction based on shape analysis on BDT output:
  About 20% improvement in expected limit w.r.t. 2011 Cut and count in Signal enriched region

BDT Analysis (8 TeV)







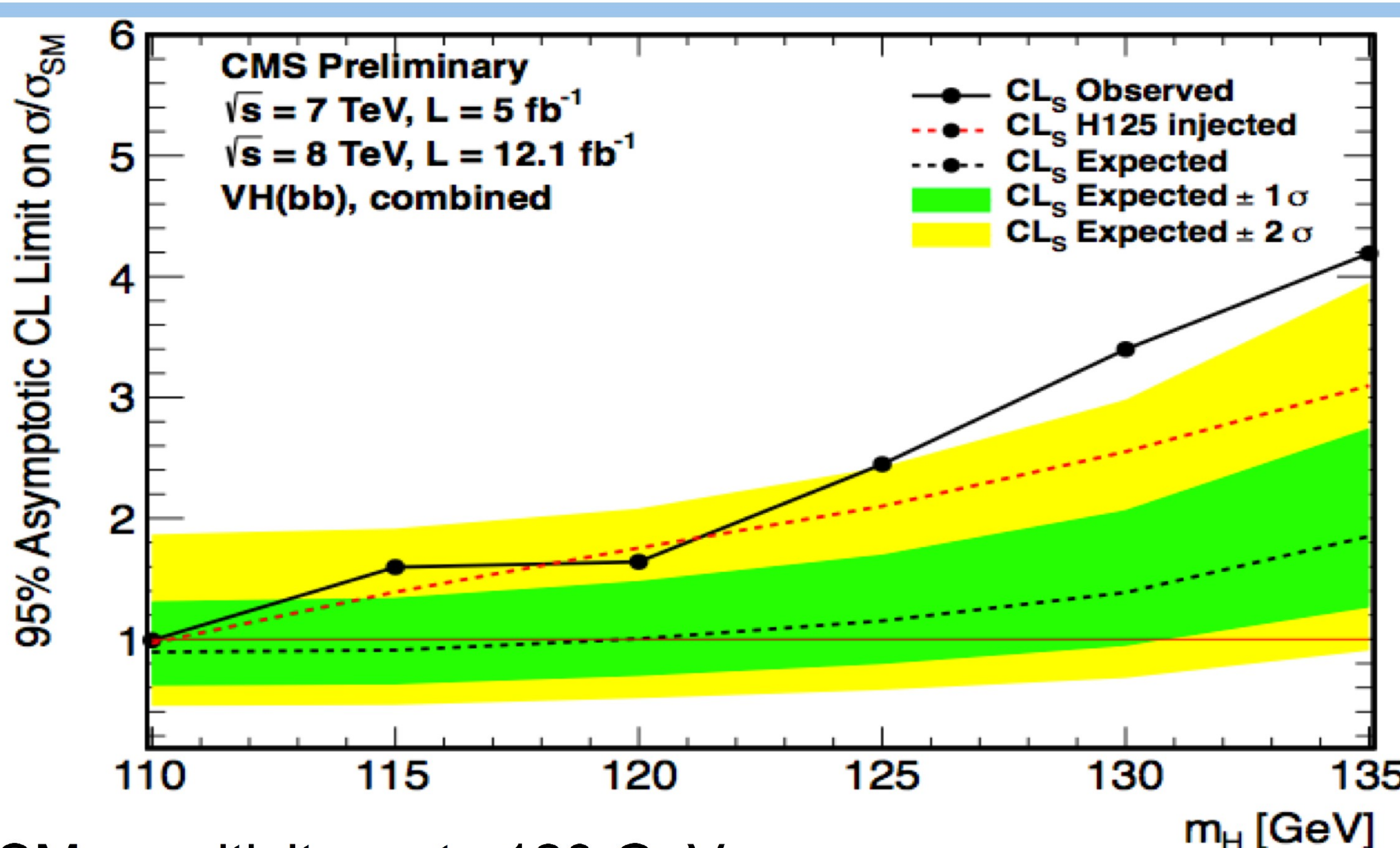


Systematic Uncertainties

Source	Range
Luminosity	2.2-4.4%
Lepton efficiency and trigger (per lepton)	3%
$Z(\nu\nu)H$ triggers	3%
Jet energy scale	2–3%
Jet energy resolution	3–6%
Missing transverse energy	3%
b-tagging	3–15%
Signal cross section (scale and PDF)	4%
Signal cross section (p_T boost, EWK/QCD)	5–10% / 10%
Signal Monte Carlo statistics	1-5%
Backgrounds (data estimate)	$\approx$ 10%
Single-top (simulation estimate)	15–30%
Dibosons (simulation estimate)	30%

Dominant uncertainties: b-tagging, background modeling, signal cross-section

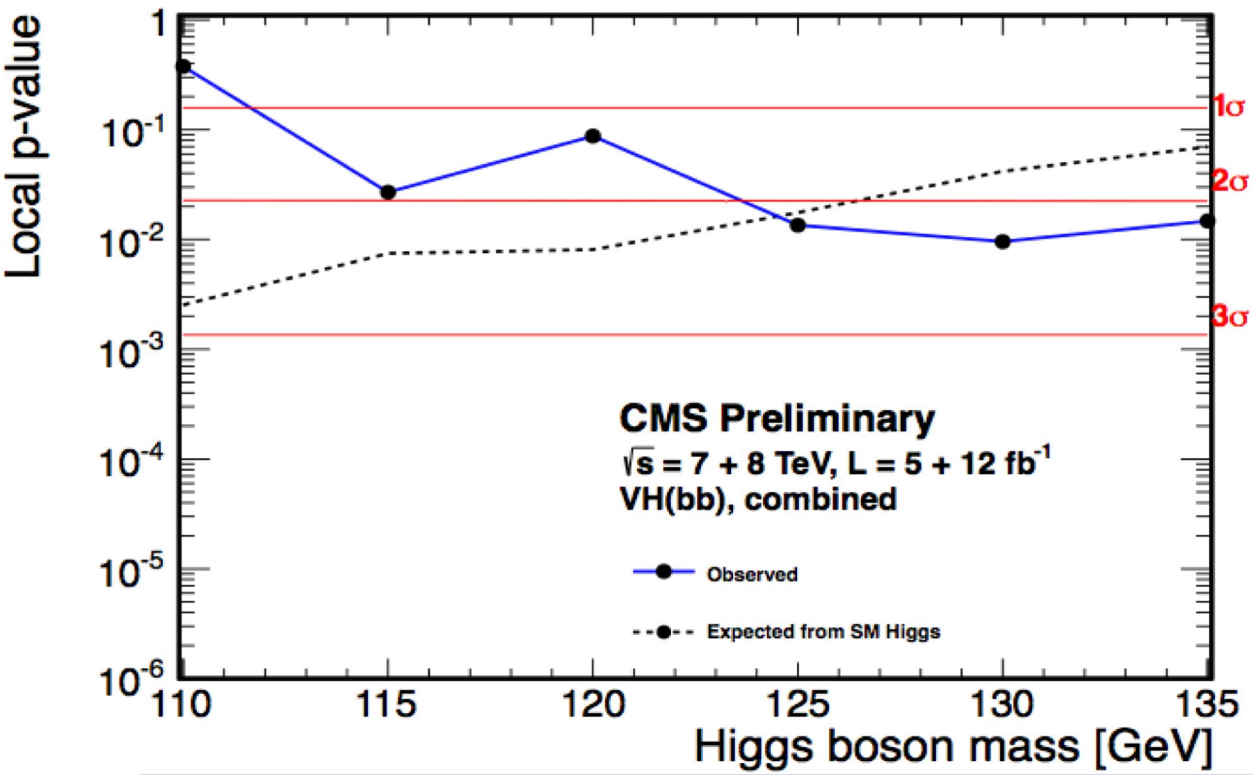
Results: SM Exclusion Limits

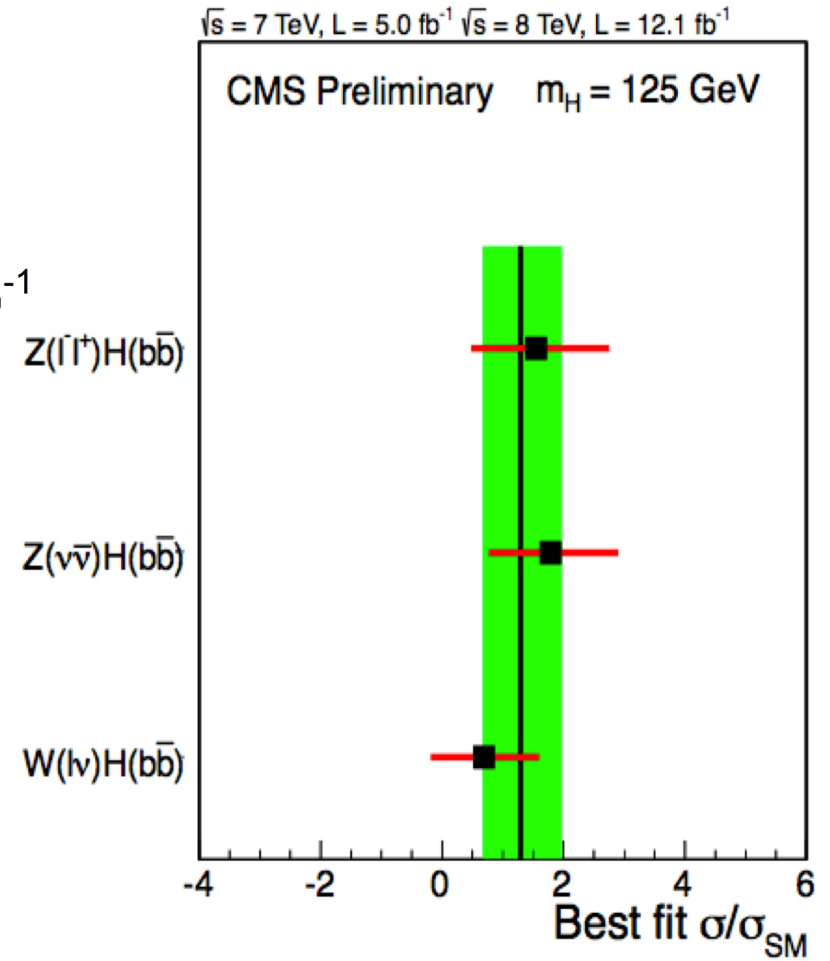


- SM sensitivity up to 120 GeV
- Expected sensitivity around 1.2 x σ_{SM} for m_H=125 GeV
- Observe broad excess across the full mass range considered
  → compatible with SM Higgs signal injection
  → Expect 1.15 x σ_{SM} at 125 GeV, observe 2.45 x σ_{SM}

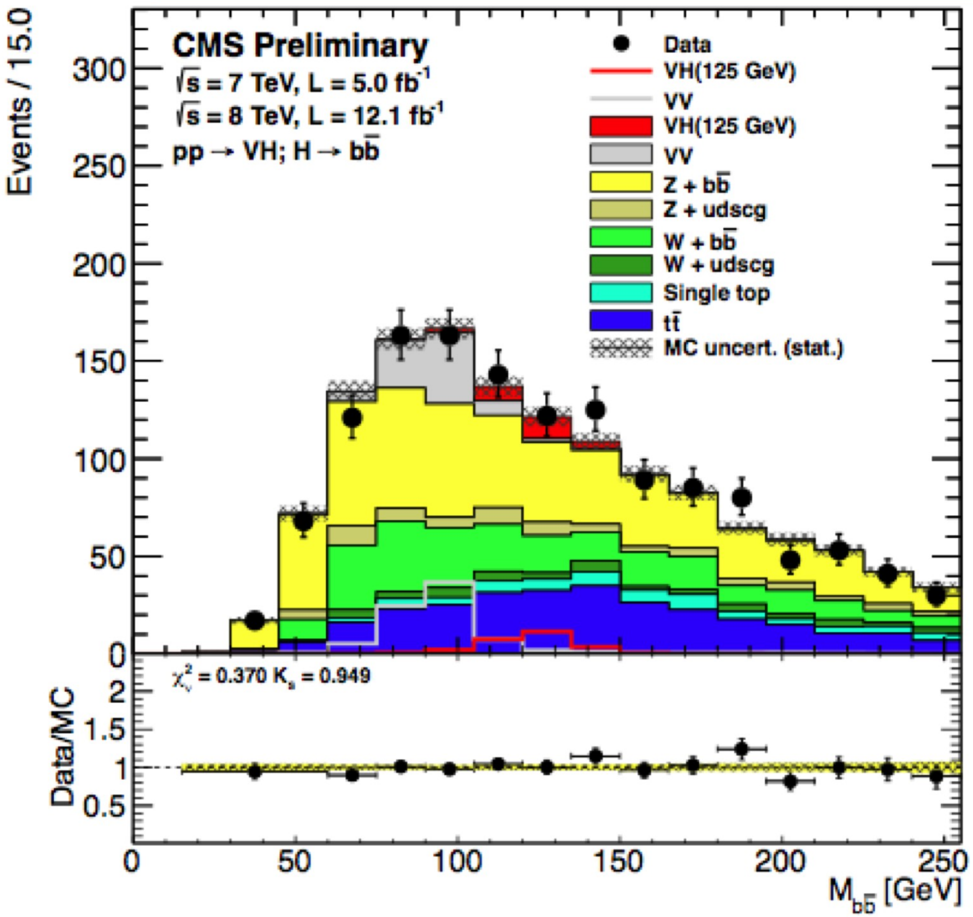
Signal Strength and p-values

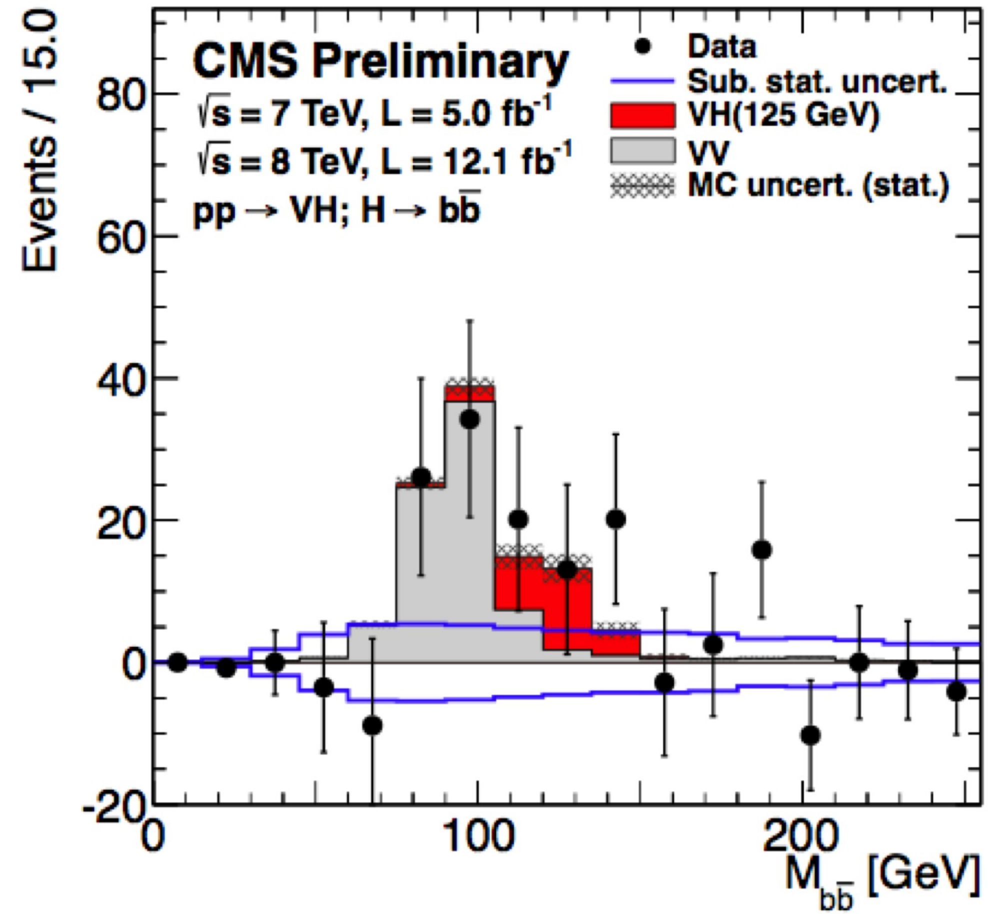
- **Observed Excess 2.2σ at 125 GeV!**
- Consistent signal strength for different channels, combined fit gives $\mu=1.3^{+0.7}_{-0.6}$
  → Expect O(30%) uncertainty on μ with 30 fb^{-1}





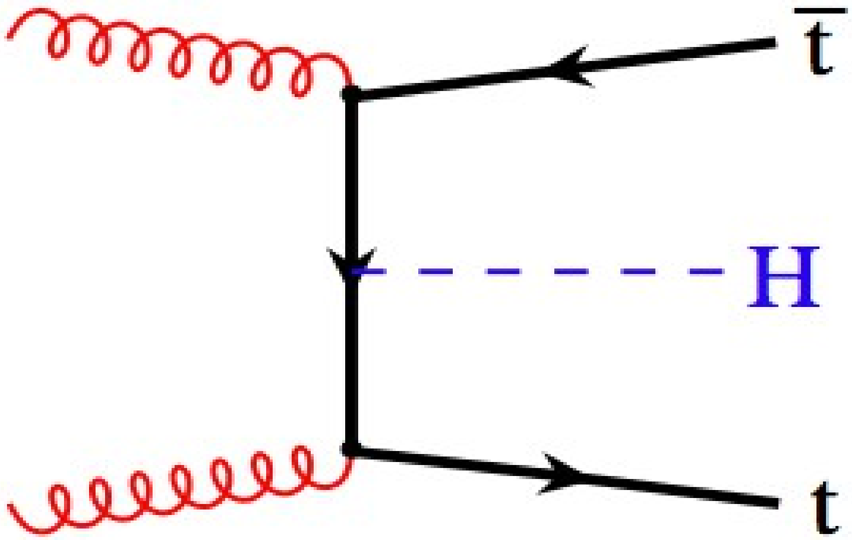
7 + 8 TeV di-jet mass distributions





- Tighter cuts, stronger background rejection, used as cross-check
- Show combination of 5 channels, overall nice Data/MC agreement
  → Small excess also showing-up here
- Good description of di-boson peak from W(Z)Z(bb)!

5.0 fb^{-1} @ $\sqrt{s}$= 7 TeV
HIG-12-025



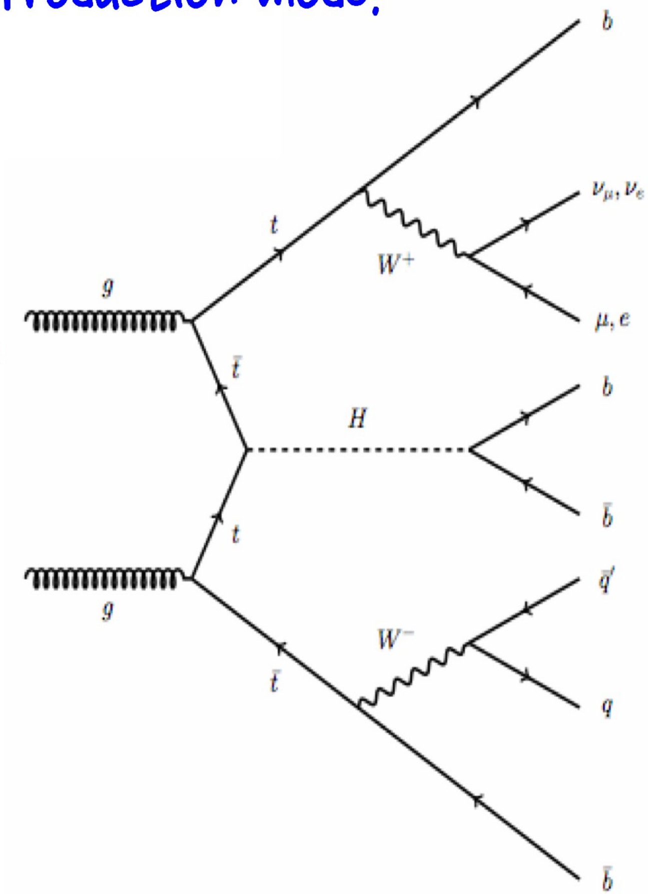
$gg \rightarrow t\bar{t}H$
$H \rightarrow b\bar{b}$

Presented for the first time at ICHEP
Test new production mechanism

ttH Analysis Overview

- Additional information in overall Higgs search
- Study lepton+jet (LJ) or di-lepton (DIL) top decays
- Major background from ttbar (+jet) events
- Split events by top decay and by number of jets and b-tags
- ANN to separate ttbar and ttbarH
  → Use simultaneous fit of ANN shape in each jet/tag category for search
  → Very different S/B, categories with low sensitivity help constraining B

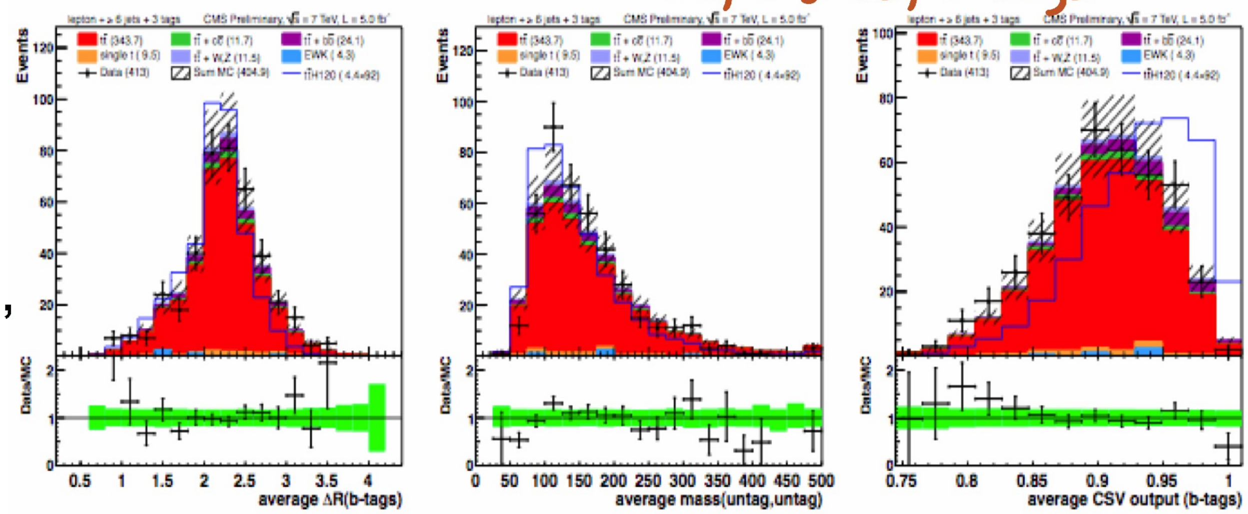
New Analysis,
First LHC study of this
Production mode!



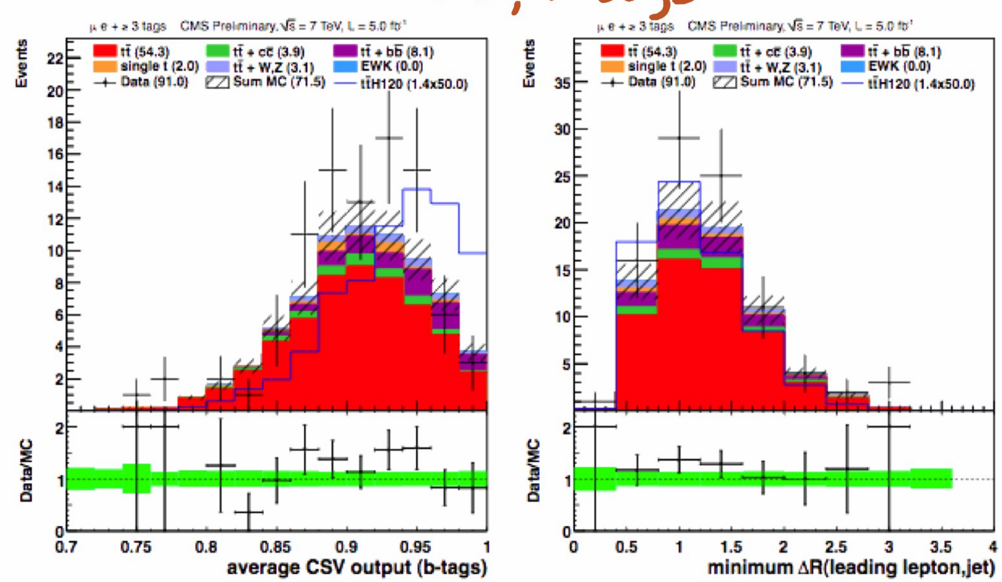
ANN Analysis Validation

- Build ANN discriminant for each (LJ or DIL) category
- Most relevant variables: b-tag, kinematic and angular correlation (e.g. min ΔR between all pairs of b-tagged jets)
  → Check data/MC agreement
- Irreducible background from tt+bb events studied with dedicated control region
  Built from ad-hoc ANN

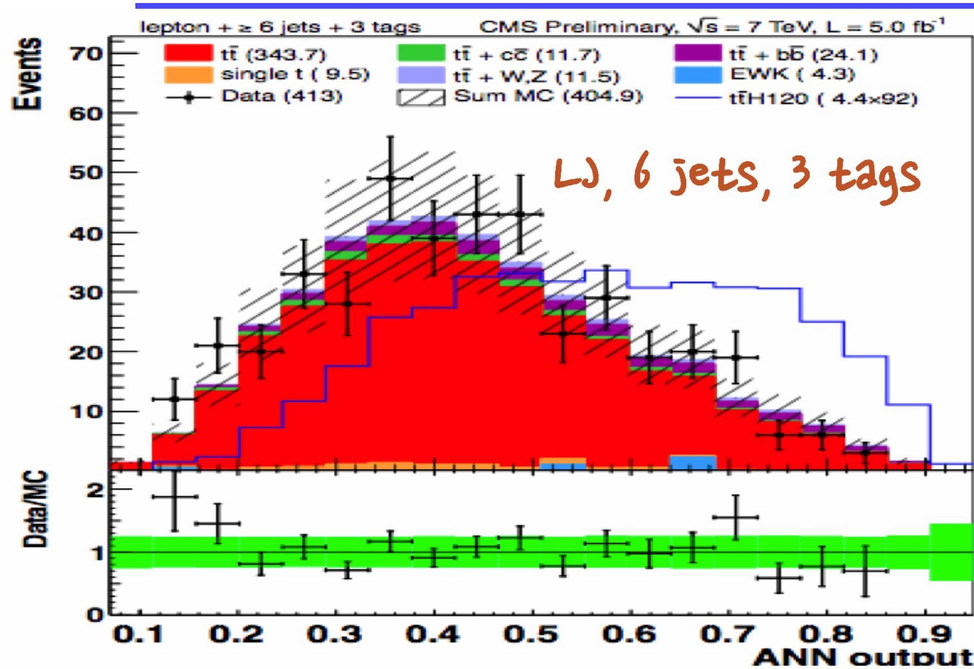
LJ, 6 jets, 3 tags

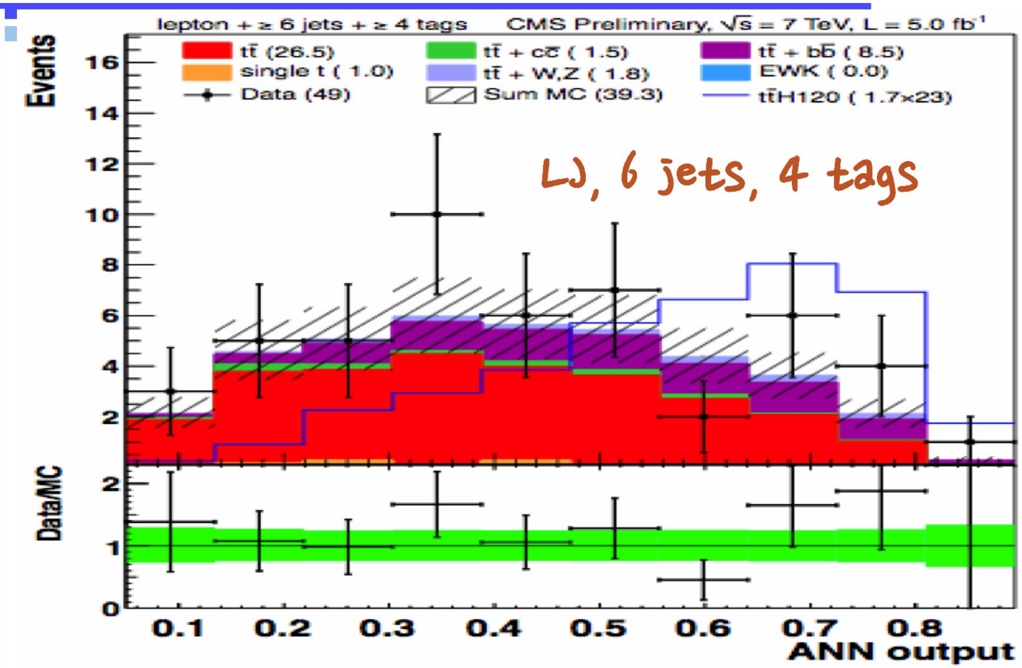


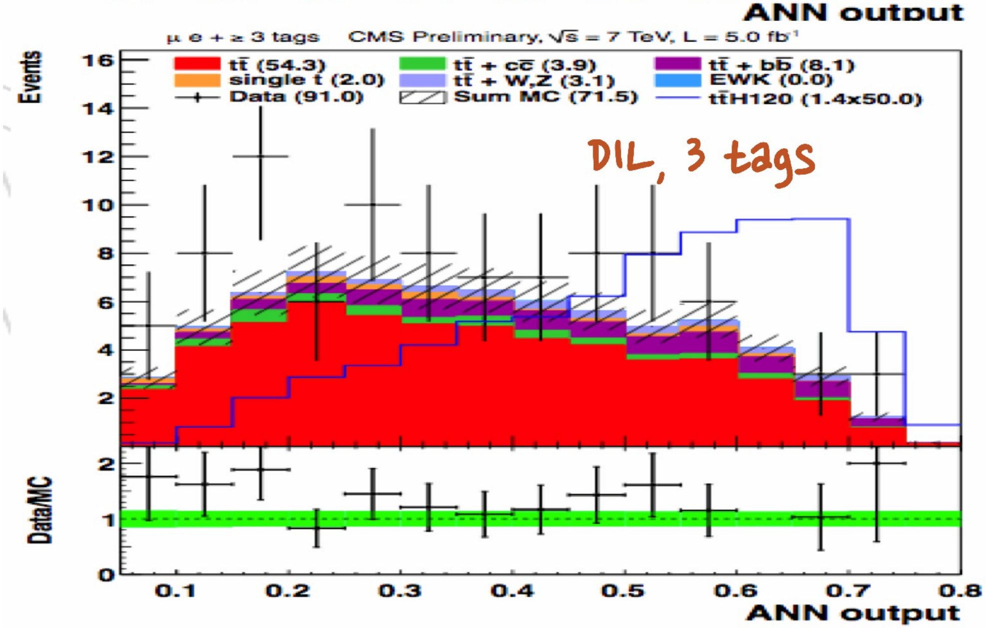
DIL, 3 tags



ANN Output Distributions



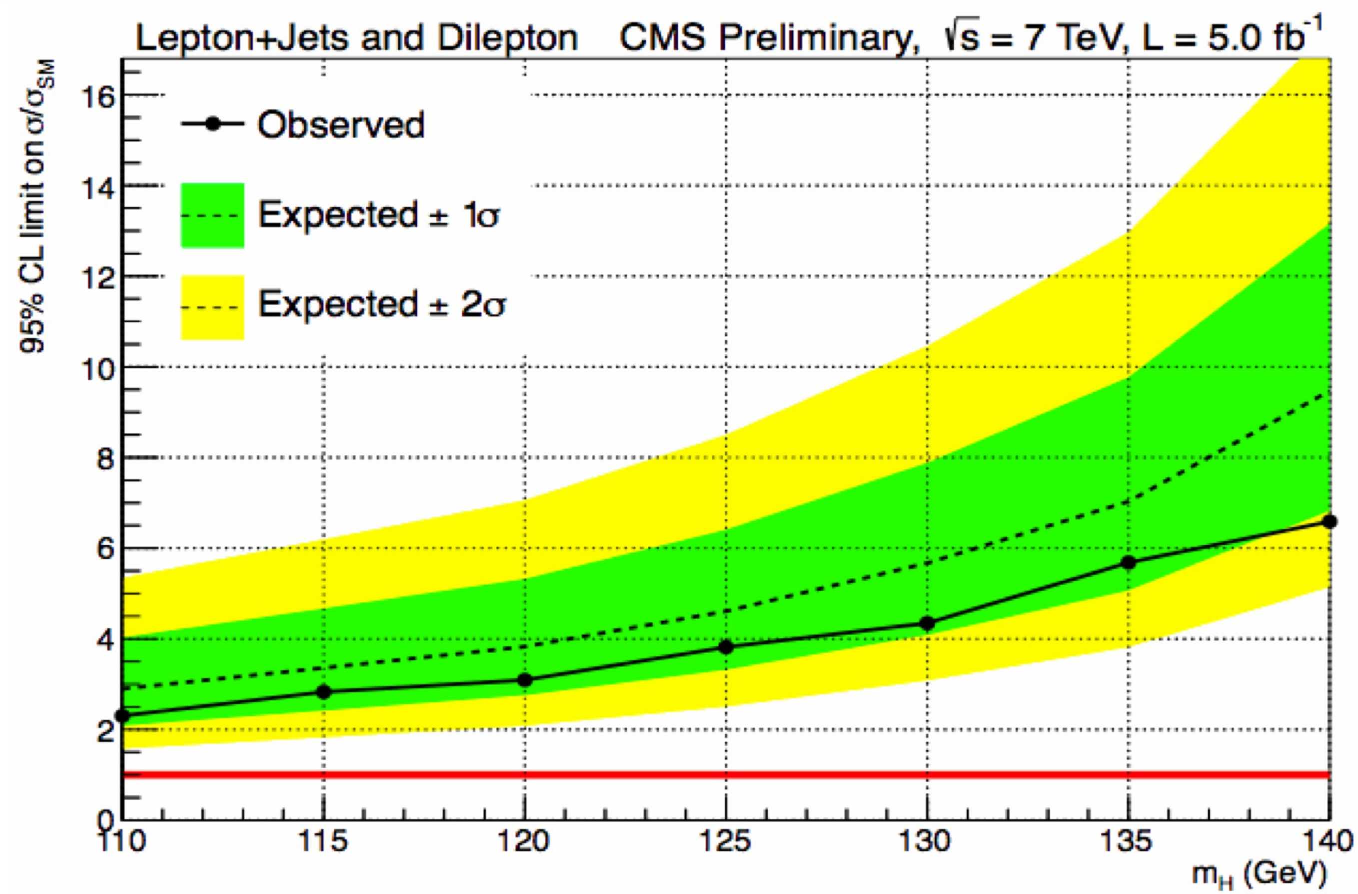




S/B strongly dependent on # tags
DIL: 2-3 tag categories
LJ: 2-4 tags, 4-6 jets

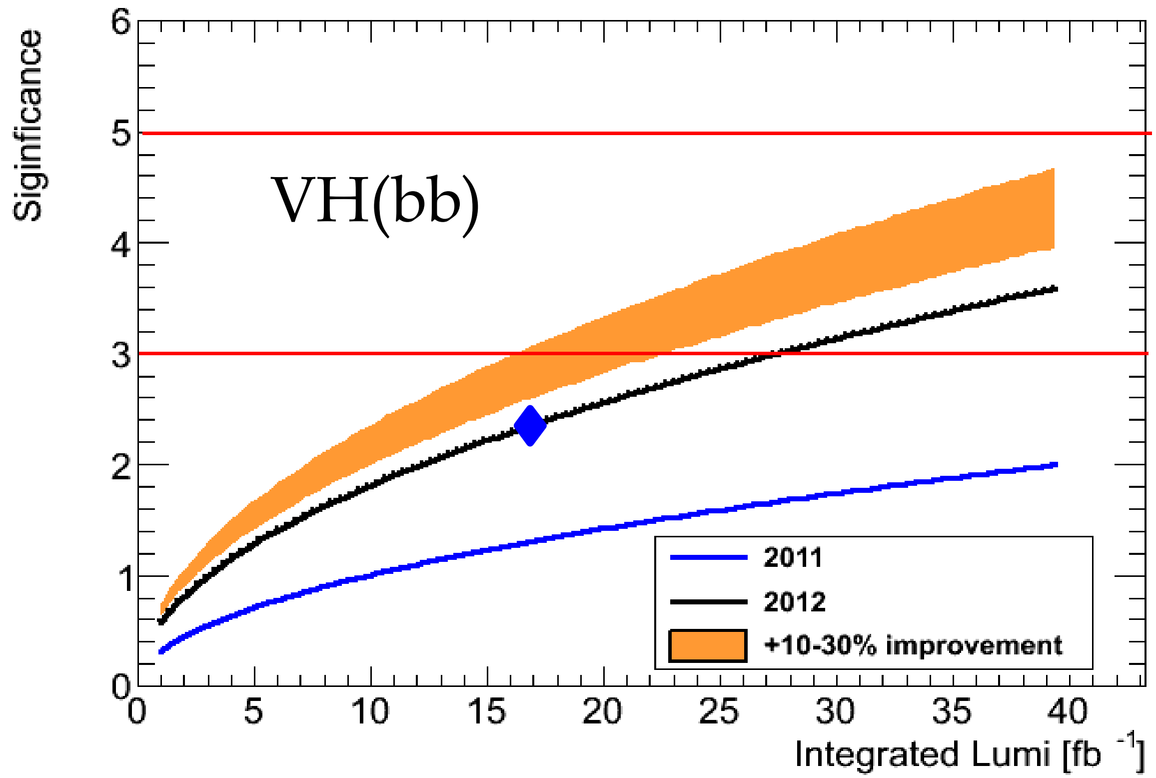
Signal expectation rescaled to
Σ (background)

Results: SM Exclusion Limits



- Sensitivity dominated by lepton+jet mode, 5-10% improvement from di-lepton mode
- Dominant uncertainties: b-tag, JES in LJ, factorization scale in DIL
- No excess seen, expect 4.6 x σ_{SM} at 125 GeV, observe 3.8 x σ_{SM}

Prospects



- VH analysis entering in the (re)discovery realm
- Sensitivity improves better than $\sqrt{L}$, but does not scale up with $\sqrt{s}$!
- More data available by the end of the year, new ideas being explored (e.g. substructure, but many more in the bucket)
- Also large room for improvement in ttH

Conclusions

- The characterization of the new boson discovered at m=125 GeV at the LHC is top the priorities of the CMS and ATLAS physics programs

- Outstanding performances of the LHC will allow to shed some light on the nature of this new particle by the end of the year

- Presented most recent results on search for SM $H \rightarrow b\bar{b}$ in CMS
  $\rightarrow$ Test coupling to fermions
  $\rightarrow H \rightarrow b\bar{b}$ largest BR for m_H=125 GeV

- 2σ excess in $H \rightarrow b\bar{b}$, under-fluctuation in ttH (7 TeV data), update coming soon

Backup Slides

Data Samples and Triggers

- Analysis presented here based on full 2011 data sample (5 $fb^{-1,}$ VH+ttH) and 2012 Data collected until June TS (5 $fb^{-1,}$, VH)

2011

Mode	Lepton Trigger	Cross-Trigger (Jet, MET)
$W(\mu\nu)H$	(Isolated) muon, 17-40 GeV	-
$Z(\mu\mu)H$	(Isolated) muon, 17-40 GeV	-
$W(e\nu)H$	Isolated electron, ID cuts, 17-32 GeV	2 jets (25-30 GeV) + MHT (15-25 GeV)
$Z(ee)H$	Di-electron, 17-8 GeV	-
$Z(\nu\bar{\nu})H$	-	MET (80-100 GeV) + 2 jets (20 GeV) OR MHT (150 GeV)
$t\bar{t}H$	Isolated muon, 24 GeV	-
$t\bar{t}H$	Isolated electron, ID cuts, 25 GeV	3 jets (30 GeV)
$t\bar{t}H$	two leptons (electron and/or muon), 17-8 GeV	-

2012

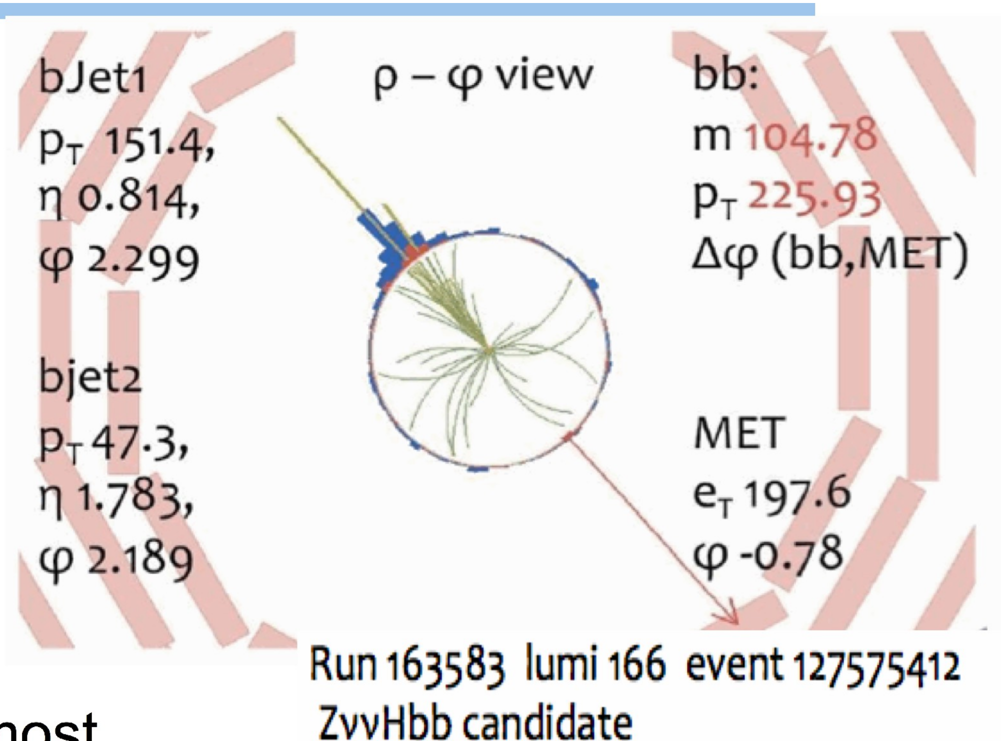
Mode	Lepton Trigger	Cross-Trigger (Jet, MET)
$W(\mu\nu)H$	(Isolated) muon, 24-40 GeV	-
$Z(\mu\mu)H$	(Isolated) muon, 24-40 GeV	-
$W(e\nu)H$	Isolated electron, ID cuts, 27 GeV	-
$Z(ee)H$	Di-electron, 17-8 GeV	-
$Z(\nu\bar{\nu})H$	-	MET (80 GeV) + 2 jets (25-60 GeV), $\Delta\phi$ cuts OR MHT (150 GeV)

- Lepton efficiencies determined directly on data using Z events
  InVH, trigger Efficiencies well above 90% w.r.t. offline cuts (Boost)

VH Analysis in a nutshell

- 5 modes under study:
  Z(ll)H, W(lν)H, Z($\nu\nu$)H, l= e,μ

- Boosted analysis:
  → Require high momentum vector
  boson and 2-b tagged jets,
  back-to -back
  → Better signal to background ratio
  → Two p_T(V) bins



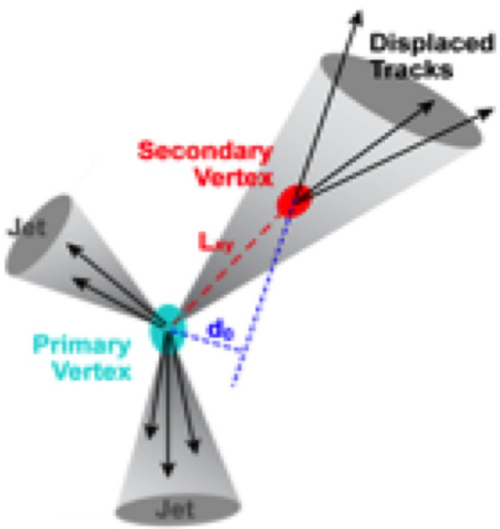
- Use Data control regions to constrain most
  Important backgrounds (V+jet, Light or Heavy, ttbar)

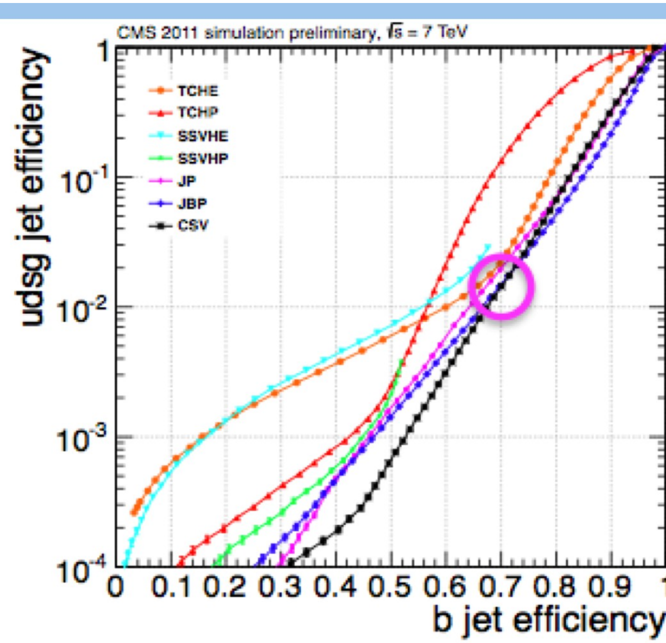
- b-jet energy regression
  → Mass resolution improvement

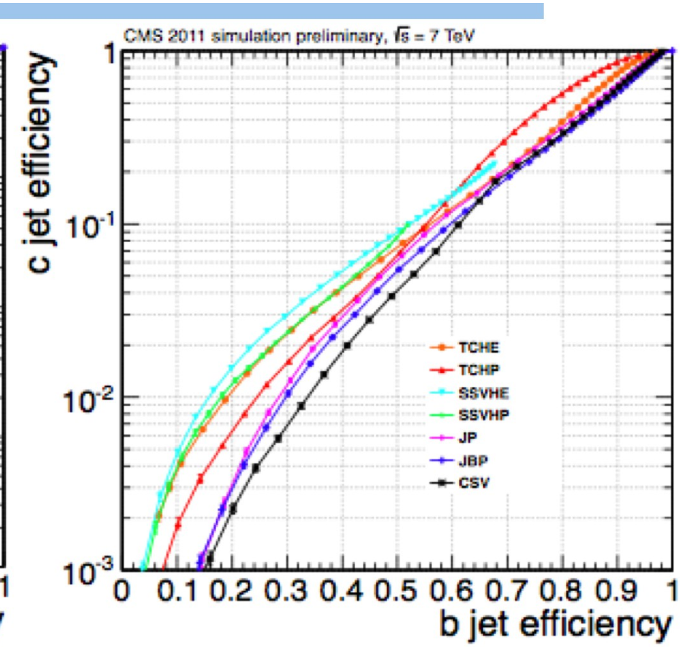
Channel	Medium boost	High boost
ZllH	50<Zpt<100	Zpt>100
WlnH	120< Wpt<170	Wpt>170
ZnnH	120<Zpt<160	Zpt>160

- Boosted Decision Tree algorithm (BDT)
  to discriminate signal versus background

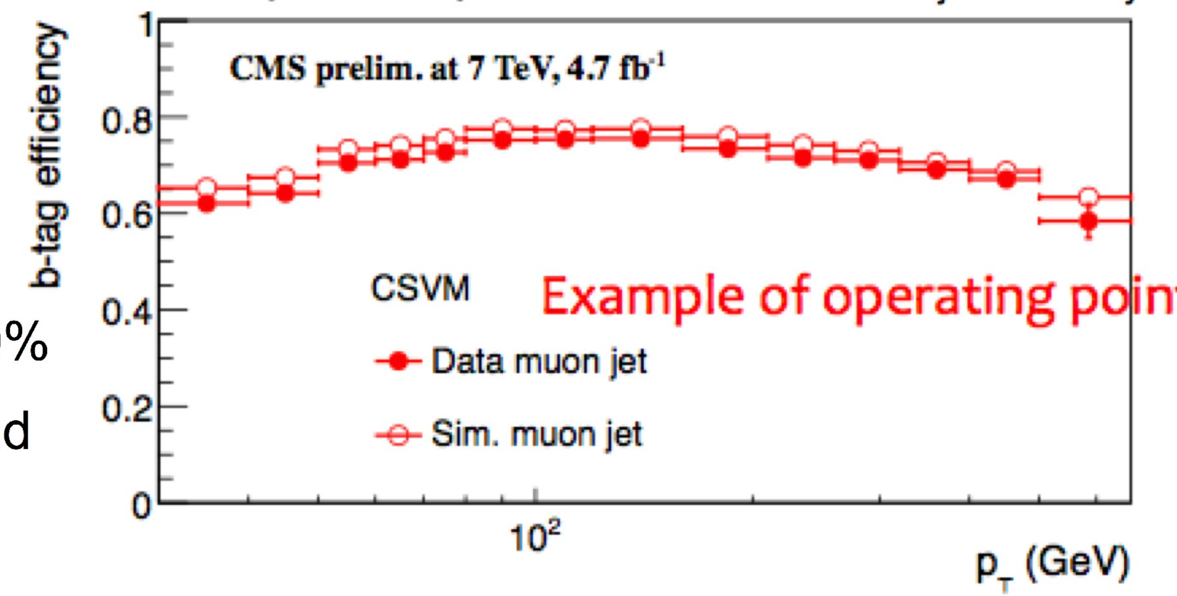
b-tagging at CMS



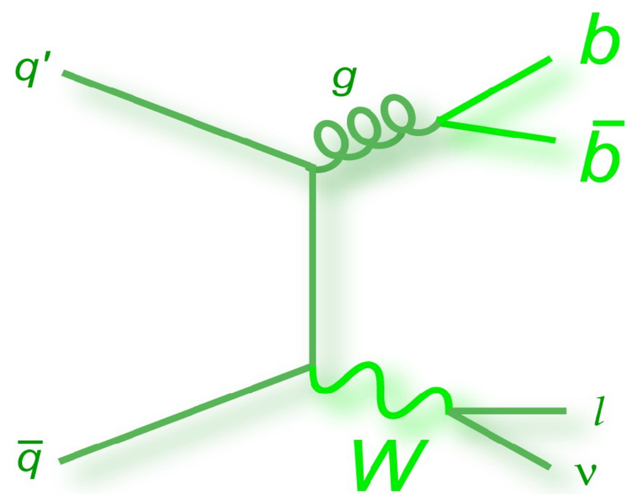


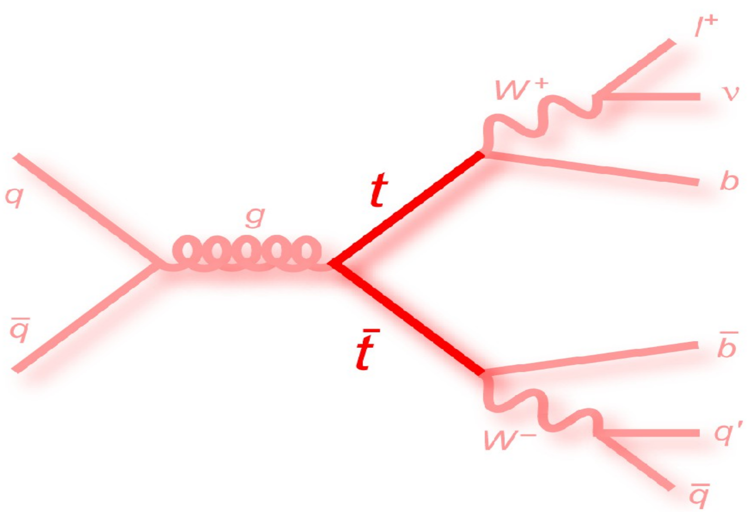


- CSV: Likelihood tagger using SV (if any), track IP etc.
- Eff ~70% for udsg ~2%, c-jet eff~20%
- Eff and fake rate from data: ttbar and μ+jet events



Backgrounds





Reducible backgrounds

QCD (strongly suppressed by iso and boost)

V+udscg,V+$b\bar{b}$ @ low p_T and mass

W(lν)W(jj)

ttbar and single top (Wb)

Irreducible backgrounds

V+$b\bar{b}$ @ high p_T and mass

ZZ($b\bar{b}$), W(lν)Z($b\bar{b}$)

Important discriminating variables

Mass resolution (separation of VH from VV)

b-tagging (suppression of V+light)

Back-to-back topology

Additional jet activity

Physics Objects (2011)

- Particle Flow based Analysis
  PileUp removal using PFNoPU
  PV selected as the one with highest activity

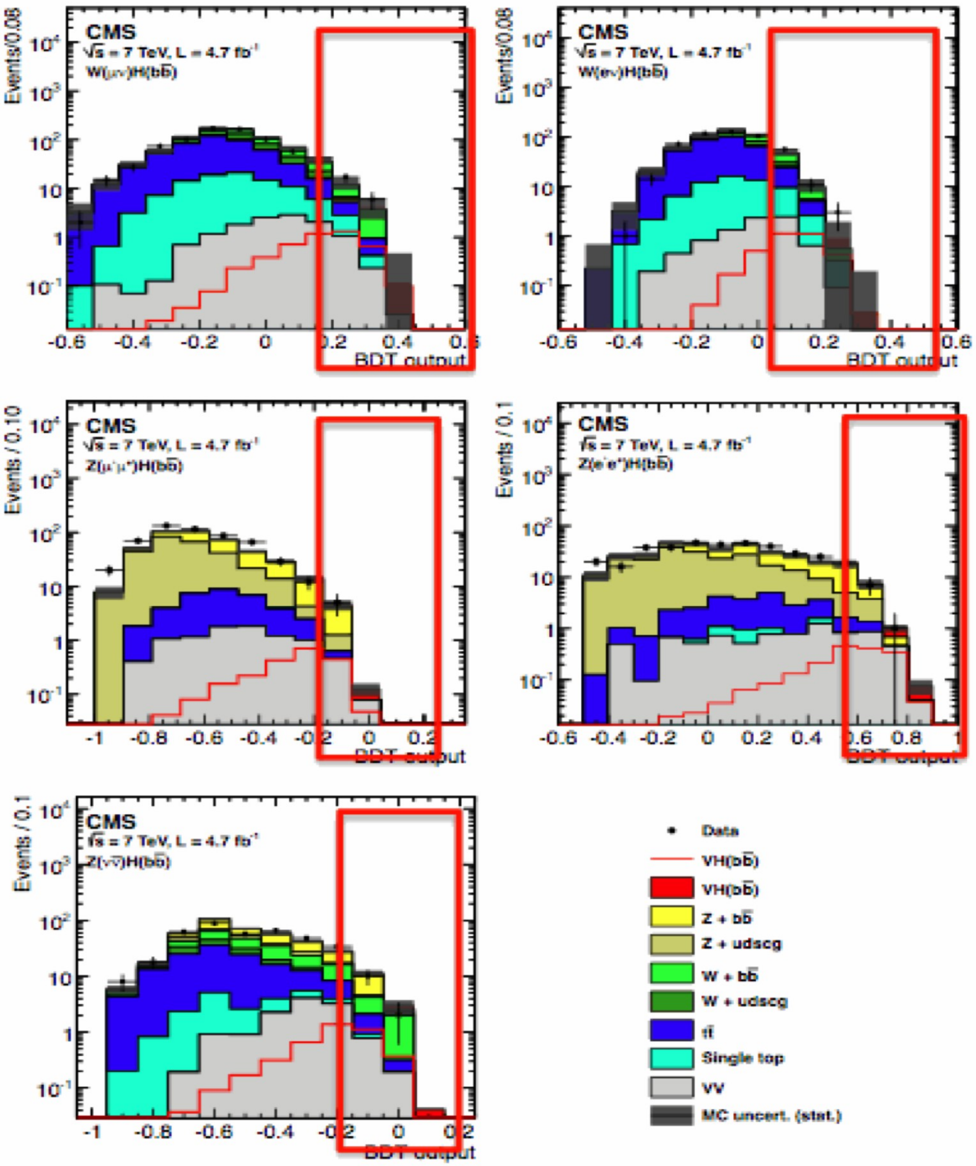
	$Z \to \ell\ell$	$W \to \ell\nu$	$Z \to \nu\nu$	$Z \to \ell\ell$	$W \to \ell\nu$	$Z \to \nu\nu$
Physics Object	p_T (GeV)			ID,Iso		
PF Muon	20, $\lvert\eta\rvert$ <2.4	20, $\lvert\eta\rvert$ <2.4	-	VBTF, PFiso < 0.15		-
PF Electron	20, $\lvert\eta\rvert$ < 2.5, NoGap	30, $\lvert\eta\rvert$ < 2.5, NoGap	-	WP95	WP80	-
AK5 PF Jets	20, $\lvert\eta\rvert$ <2.4	30, $\lvert\eta\rvert$ <2.4	80/30, $\lvert\eta\rvert$ <2.4	Loose		Tight
PFMET	-	35 ($W \to e\nu$)	160	-	-	-
$p_T(V, H)$	100	150-165	160	-	-	-

- MC re-weighted to match PU distribution on data
- Z(ℓℓ): 75 < m(ℓℓ) <105 GeV,
- Z(νν): PFMET cut and lepton veto
- W(ℓν): Combine PFMET and lepton
  No additional leptons

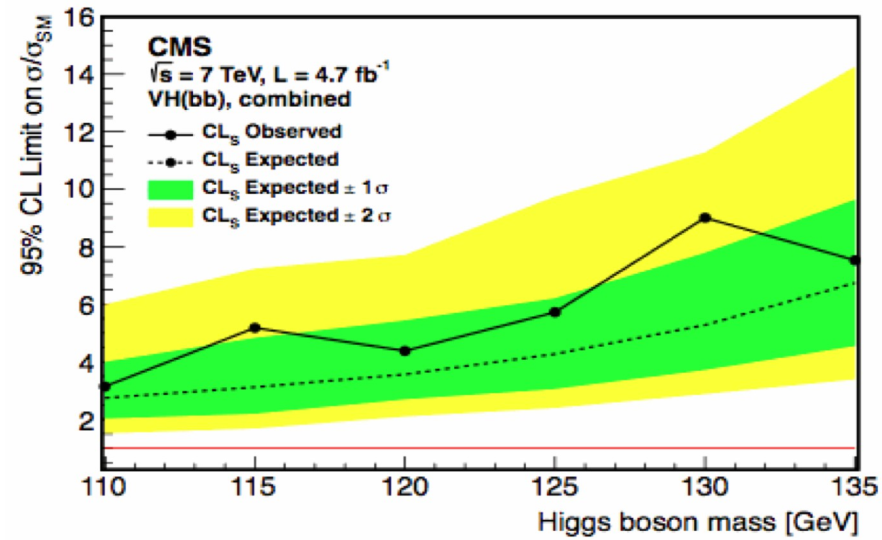
Muon selection:

- Global and Tracker;
- χ^2/ndof < 10 for the global muon fit;
- Tracks associated to muons must satisfy:
  - at least one pixel hit
  - at least ten total hits (strip + pixel)
  - at least one valid hit in the muon chambers
  - at least two muon stations
  - impact parameter in the transverse plane $d_{xy} < 2\,\text{mm}$

VHbb 2011 Results



m_H (GeV)	110	115	120	125	130	135
BDT Exp.	2.7	3.1	3.6	4.3	5.3	6.7
BDT Obs.	3.1	5.2	4.4	5.7	9.0	7.5
$m(jj)$ Exp.	3.0	3.2	4.4	4.7	6.4	7.7
$m(jj)$ Obs.	3.4	5.6	6.7	6.3	10.5	8.9



Final yield estimate based on Cut and Count on the BDT discriminant

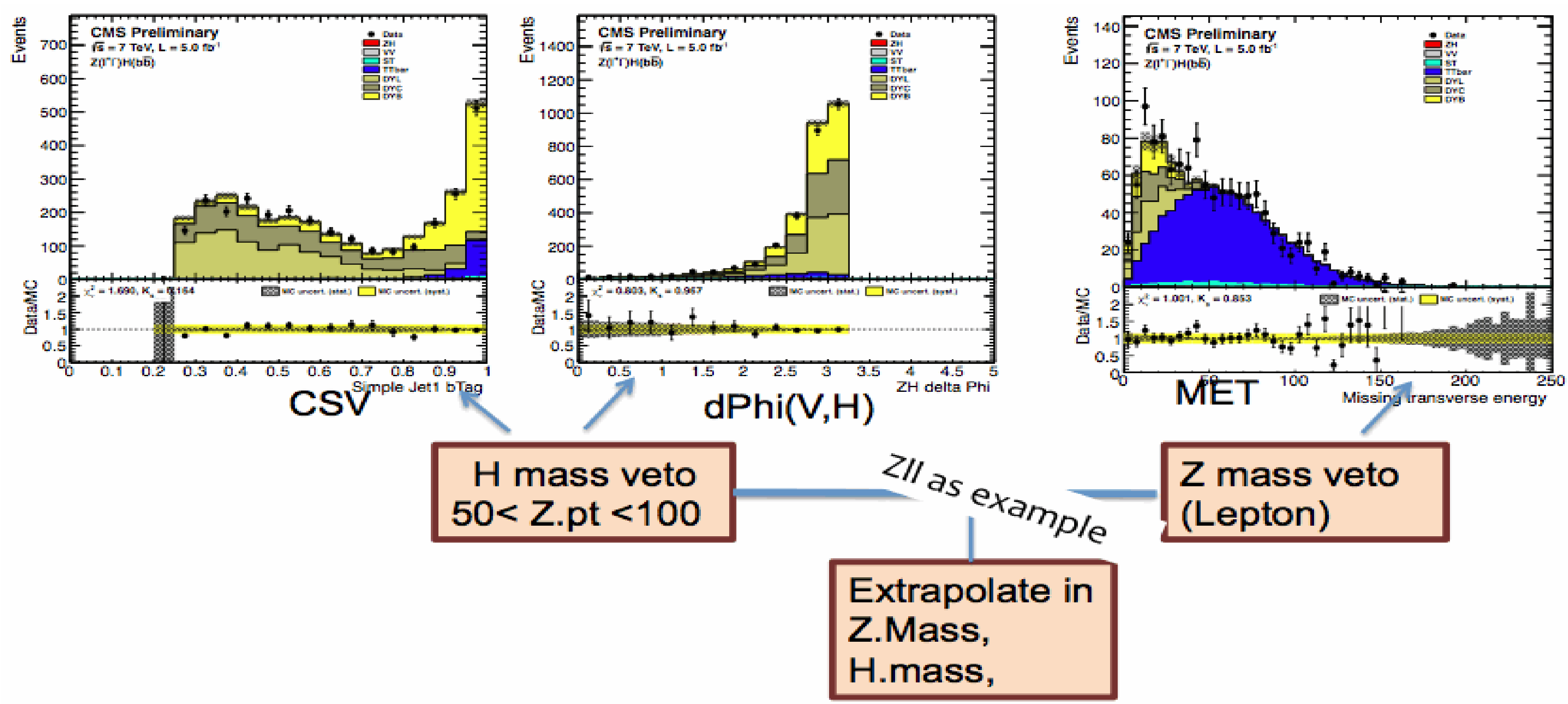
Simple Cut and Count analysis on di-jet invariant Mass (MJJ) as a cross-check

PLB 710(2012)284-306

2011 Improvements

Category	2011	ICHEP 2012	Sensitivity Gain
Background Treatment	Event Count in Control Regions	Fit shapes in Control Regions	
Higgs Reconstruction	AK5PF di-jet with standard corrections	Regression	10-20%
Boost	Single bin, high boost analysis	Two bins (add medium boost)	10%
BDT && MJJ	Cut and Count	Shape Analysis	20%

Control Region Shape Fit



- Scale Factors for V+light(heavy) and ttbar background re-weighting extracted from simultaneous binned Maximum Likelihood fit in 3 control regions
- Control regions defined as kinematically close to Signal Region, still independent

Background Scale Factors

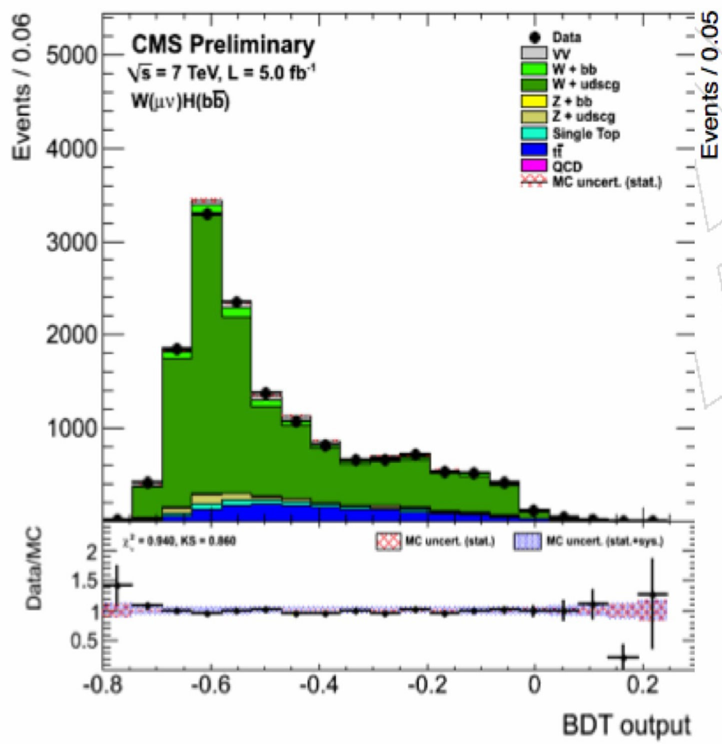
- Scale factors for background re-weighting largely consistent between 7 and 8 TeV analysis

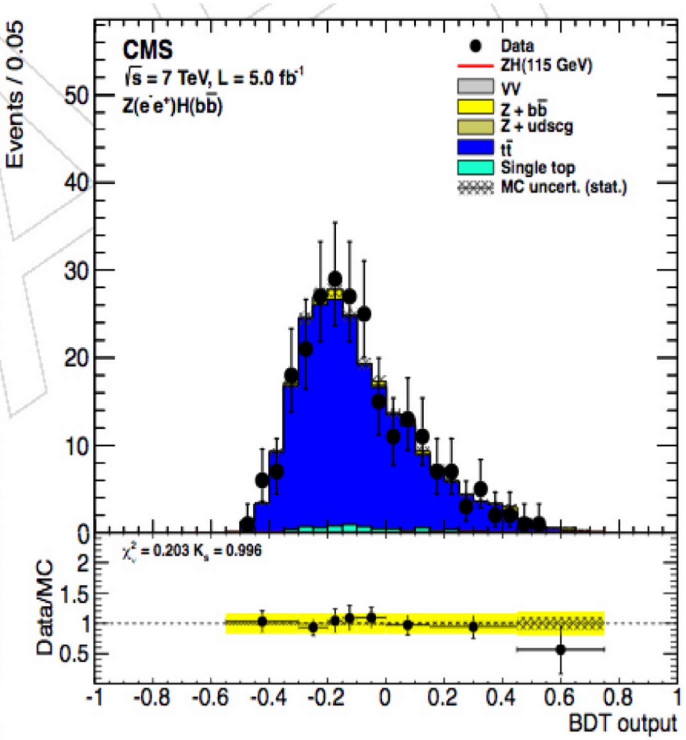
Process	W($\ell\nu$)H	W($\ell\nu$)H	Z($\ell\ell$)H	Z($\ell\ell$)H	Z($\nu\nu$)H	Z($\nu\nu$)H
Low p_T	7 TeV	8 TeV	7 TeV	8 TeV	7 TeV	8 TeV
W + udscg	$0.88 \pm 0.01 \pm 0.03$	$1.01 \pm 0.02 \pm 0.01$	–	–	$0.89 \pm 0.01 \pm 0.03$	$0.96 \pm 0.06 \pm 0.03$
$Wb\bar{b}$	$1.91 \pm 0.14 \pm 0.31$	$2.07 \pm 0.15 \pm 0.10$	–	–	$1.36 \pm 0.10 \pm 0.15$	$1.30 \pm 0.17 \pm 0.10$
Z + udscg	–	–	$1.11 \pm 0.03 \pm 0.11$	$1.10 \pm 0.02 \pm 0.06$	$0.87 \pm 0.01 \pm 0.03$	$1.15 \pm 0.07 \pm 0.03$
$Zb\bar{b}$	–	–	$0.98 \pm 0.05 \pm 0.12$	$1.08 \pm 0.04 \pm 0.08$	$0.96 \pm 0.02 \pm 0.03$	$1.12 \pm 0.10 \pm 0.04$
$t\bar{t}$	$0.93 \pm 0.02 \pm 0.05$	$1.07 \pm 0.01 \pm 0.01$	$1.03 \pm 0.04 \pm 0.11$	$1.01 \pm 0.02 \pm 0.06$	$0.97 \pm 0.02 \pm 0.04$	$1.05 \pm 0.07 \pm 0.03$
High p_T	7 TeV	8 TeV	7 TeV	8 TeV	7 TeV	8 TeV
W + udscg	$0.79 \pm 0.01 \pm 0.02$	$0.94 \pm 0.02 \pm 0.01$	–	–	$0.78 \pm 0.02 \pm 0.03$	$0.95 \pm 0.05 \pm 0.02$
$Wb\bar{b}$	$1.49 \pm 0.14 \pm 0.19$	$1.72 \pm 0.16 \pm 0.08$	–	–	$1.48 \pm 0.15 \pm 0.20$	$1.27 \pm 0.18 \pm 0.10$
Z + udscg	–	–	$1.11 \pm 0.03 \pm 0.11$	$1.10 \pm 0.02 \pm 0.06$	$0.97 \pm 0.02 \pm 0.04$	$1.04 \pm 0.07 \pm 0.02$
$Zb\bar{b}$	–	–	$0.98 \pm 0.05 \pm 0.12$	$1.08 \pm 0.04 \pm 0.08$	$1.08 \pm 0.09 \pm 0.06$	$1.15 \pm 0.10 \pm 0.04$
$t\bar{t}$	$0.84 \pm 0.02 \pm 0.03$	$0.99 \pm 0.01 \pm 0.01$	$1.03 \pm 0.04 \pm 0.11$	$1.01 \pm 0.02 \pm 0.06$	$0.97 \pm 0.02 \pm 0.04$	$1.03 \pm 0.07 \pm 0.03$

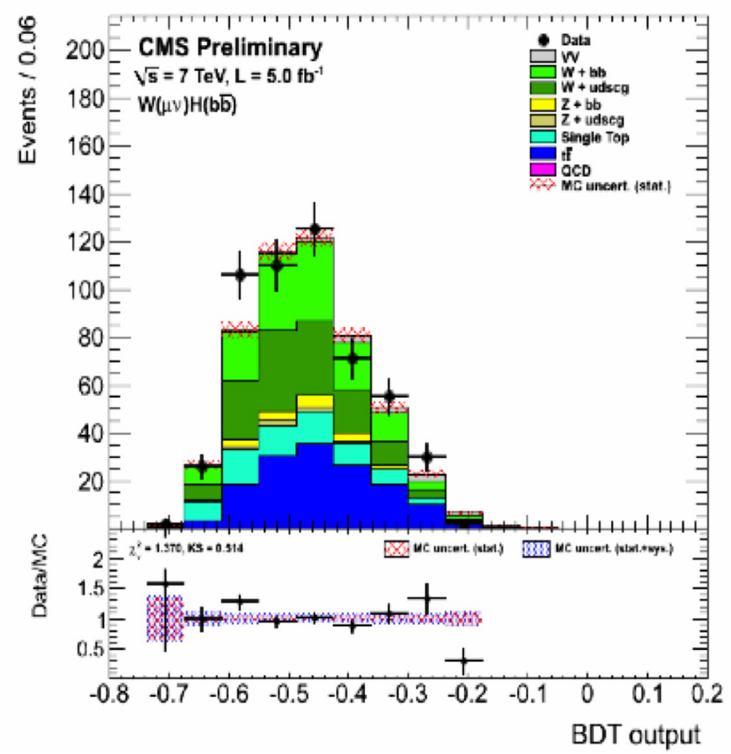
- Uncertainties include: MC statistics, detector effect (jet resolution and scale, b-tag efficiency and mis-id) and estimated by repeating the fit with template variations

BDT Test In Control Regions

W(μν)H — **Z(ee)H** — **W(μν)H**







W+light — **ttbar** — **W+heavy**

Excellent agreement of BDT output in different kinematic regions and background composition proves BDT robustness

B-jet energy Regression

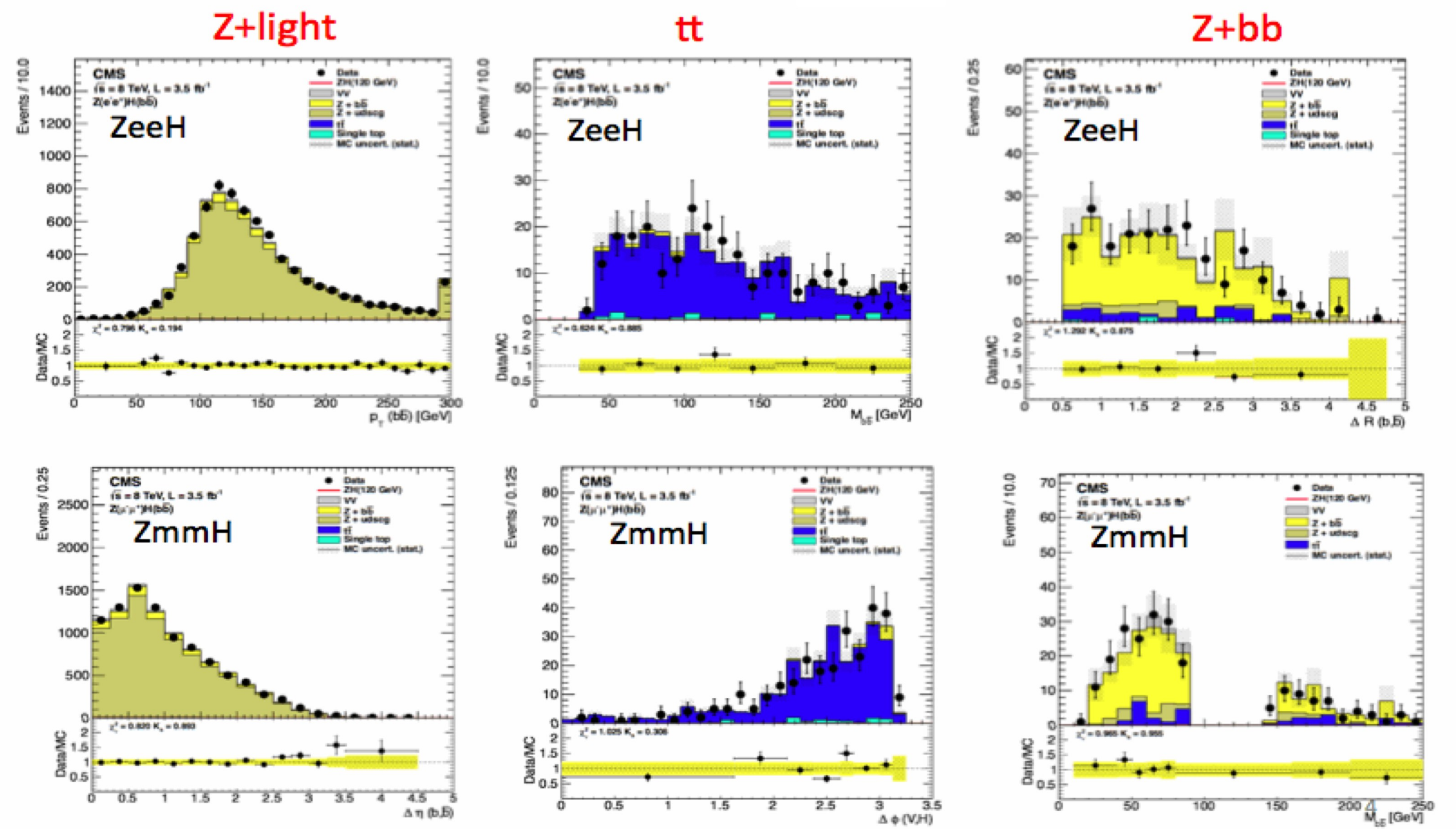
New since 2011 Analysis

- Implementation based on NN method developed at CDF for b-jet energy corrections: http://arxiv.org/pdf/1107.3026.pdf
- Multivariate Regression (BDT) trained on VH signal events using several (b)-jet variables

Variable category	Variable
Jet kinematics	p_T, raw p_T, E_T, m_T
Jet-related properties	p_T(lead track), cef, N(constituents), JEC uncertainty, ρ25
Secondary vertex	p_T, mass, L_{3D}, ΔL_{3D}
Soft lepton	p_T, $p_T$rel, ΔR(jet,lep)
ZllH specific	MET, ΔΦ(jet.MET)

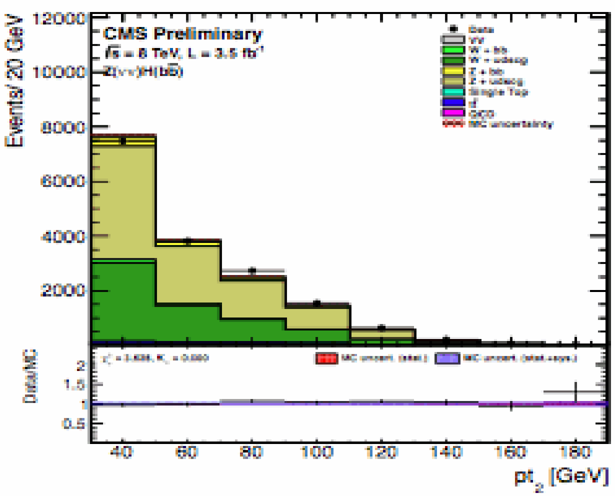
→ Training at all mass points simultaneously to avoid mass bias

Control Regions Data/MC

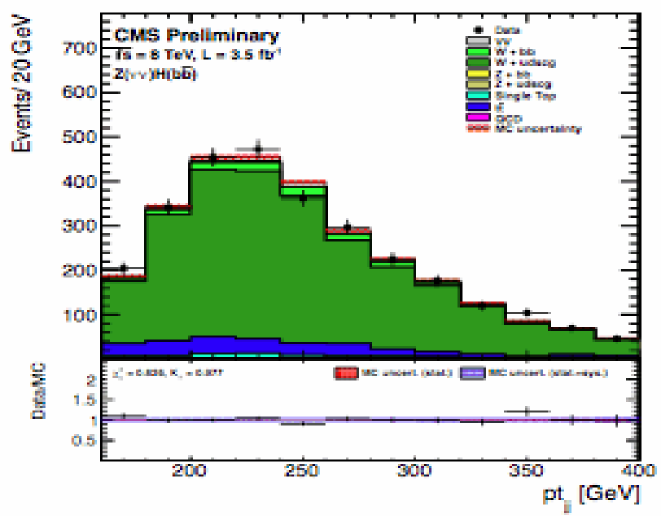


Control Regions Data/MC

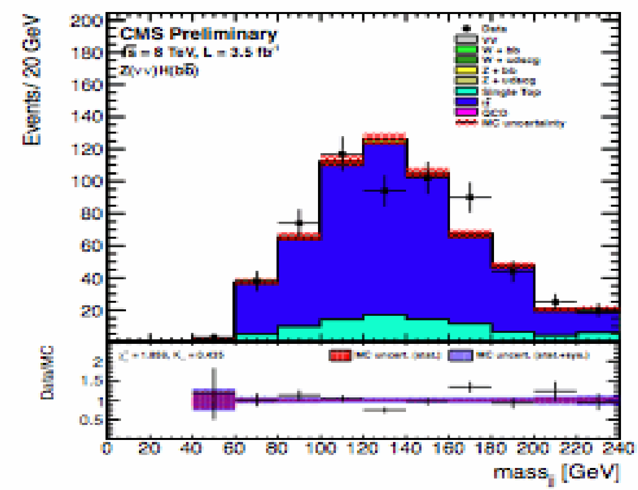
Z+light



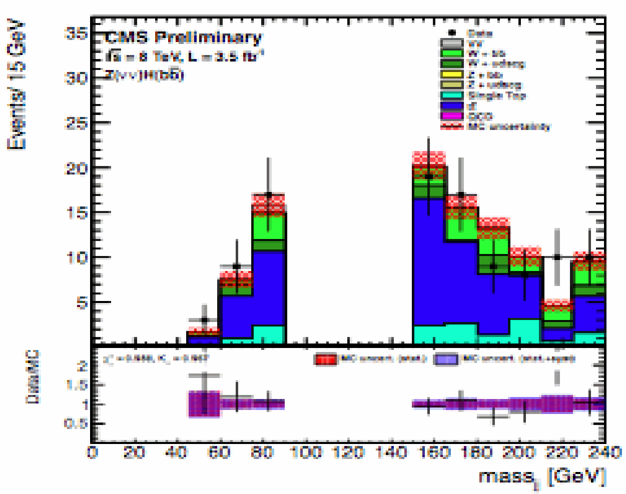
W+light

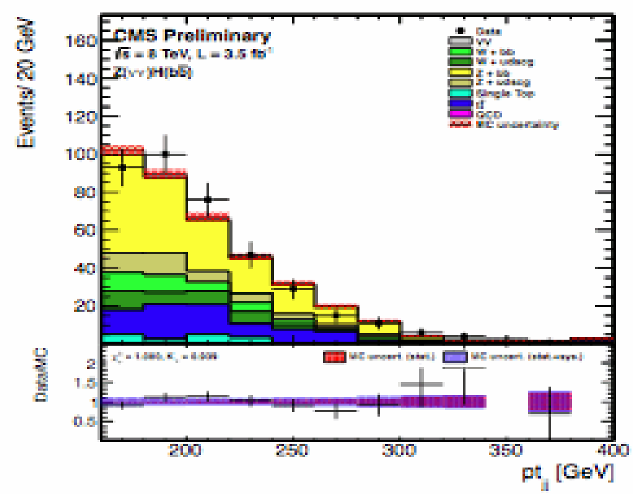


tt



W+bb

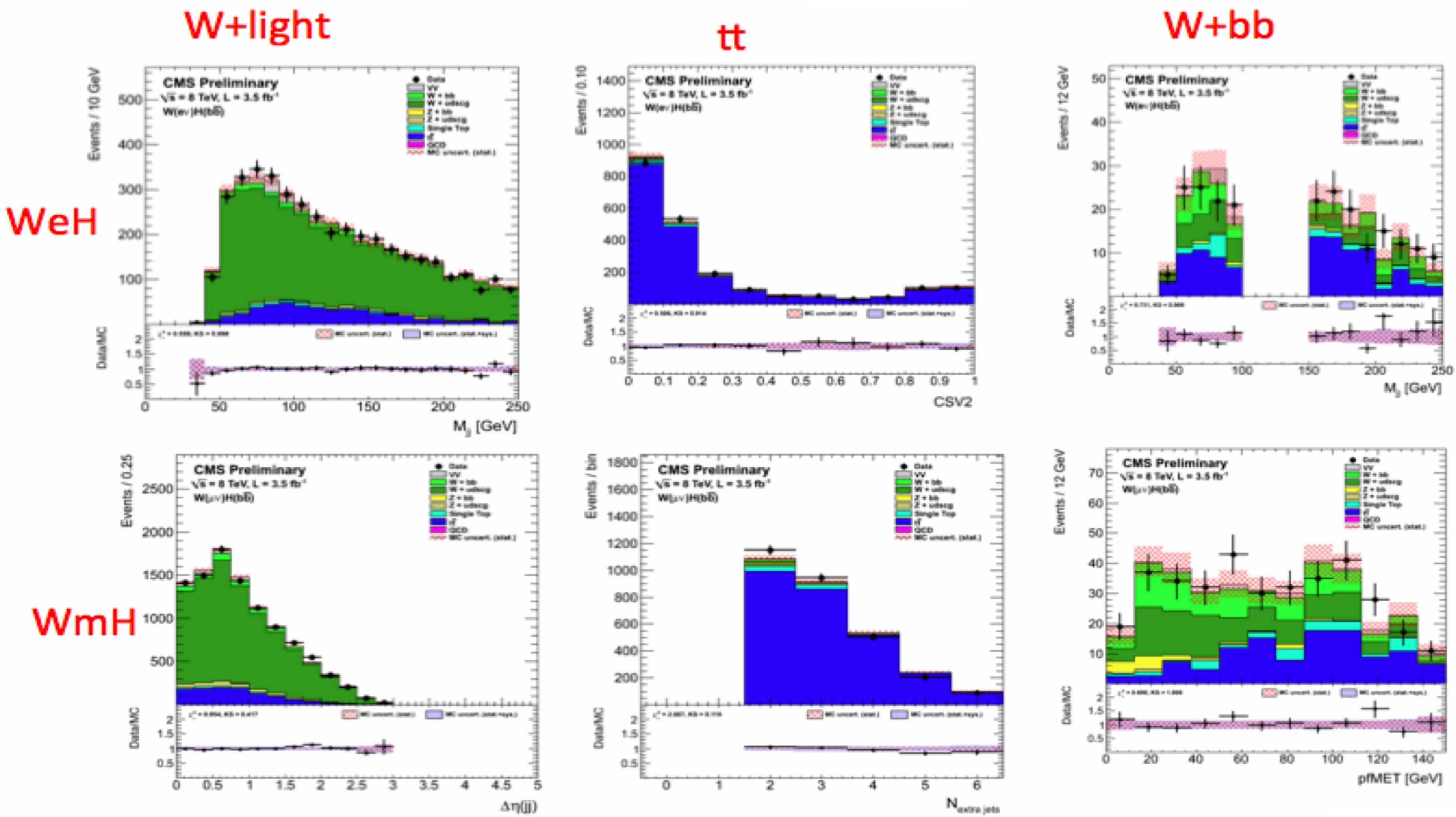




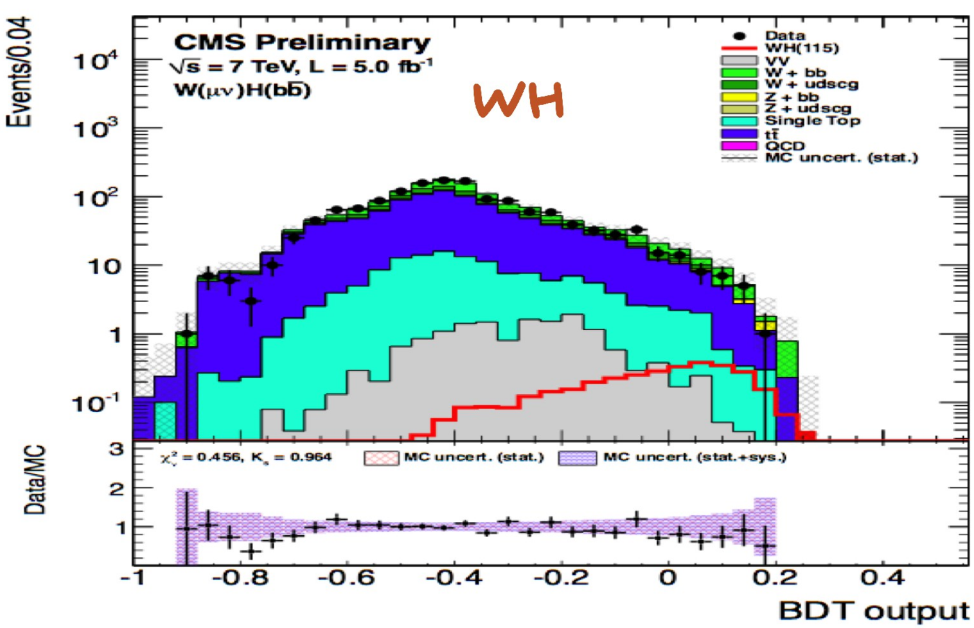
Z+bb

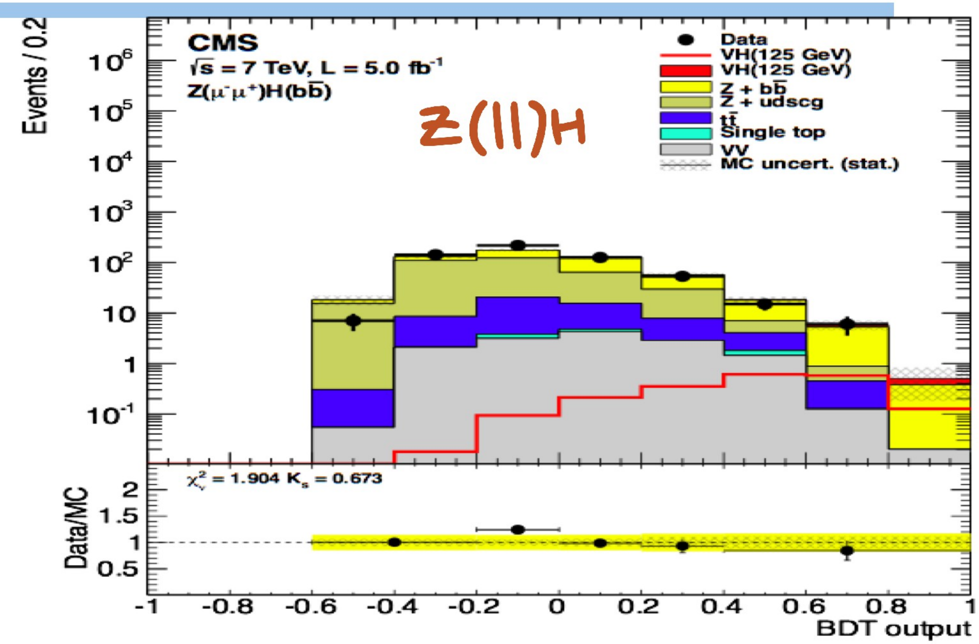
Good agreement across the board

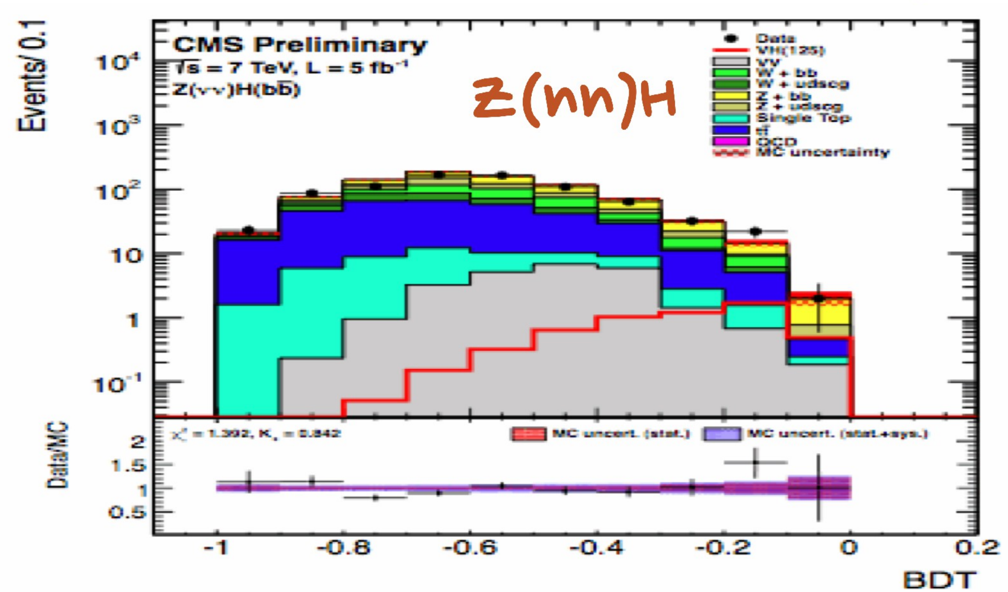
Control Regions Data/MC



BDT Analysis (7 TeV)





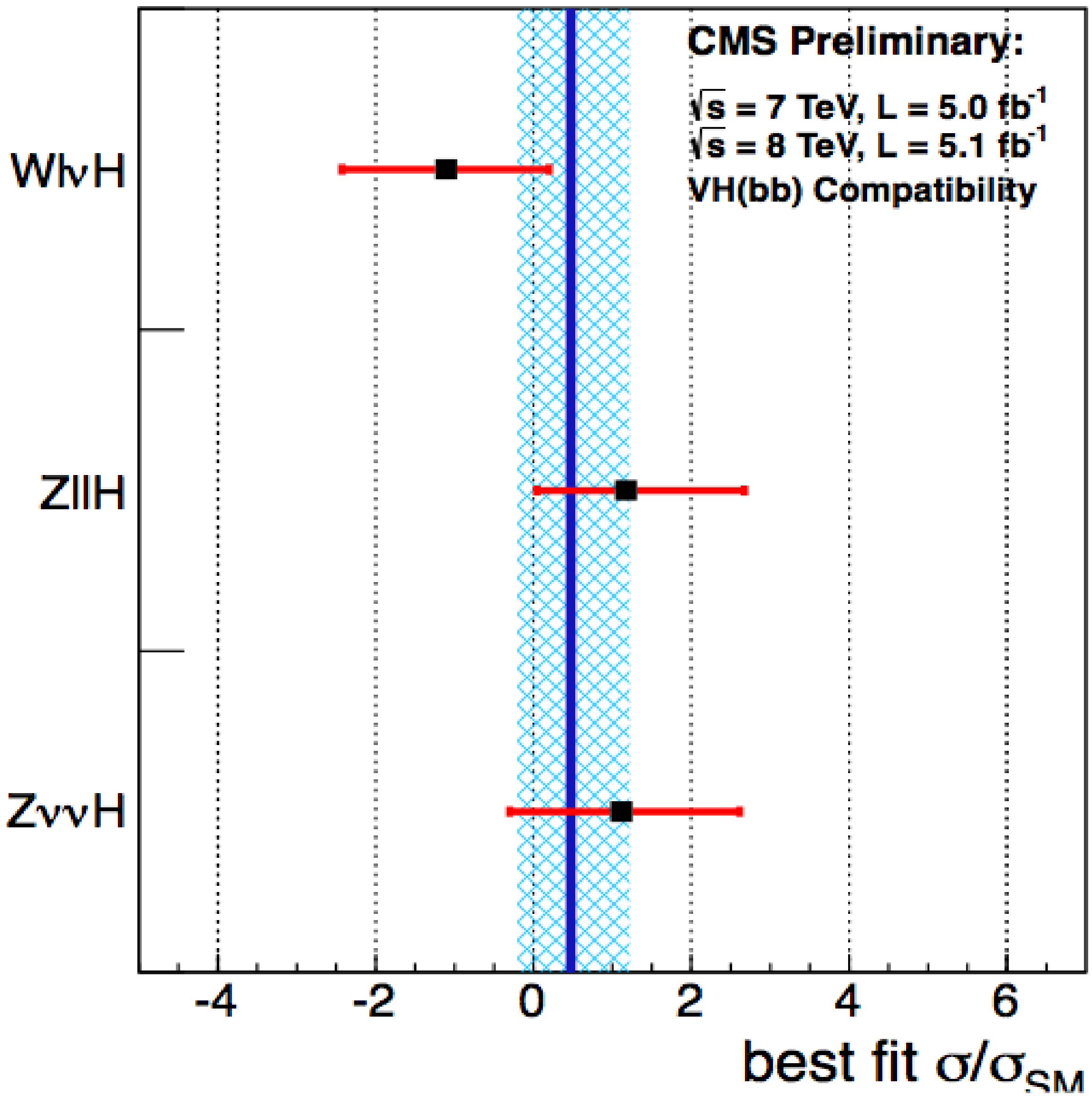


Analyses at 7 and 8 TeV carried on separately, final results from combination of:

5 (channels) x 2 (p_T bins) x 2 (7+8 TeV)

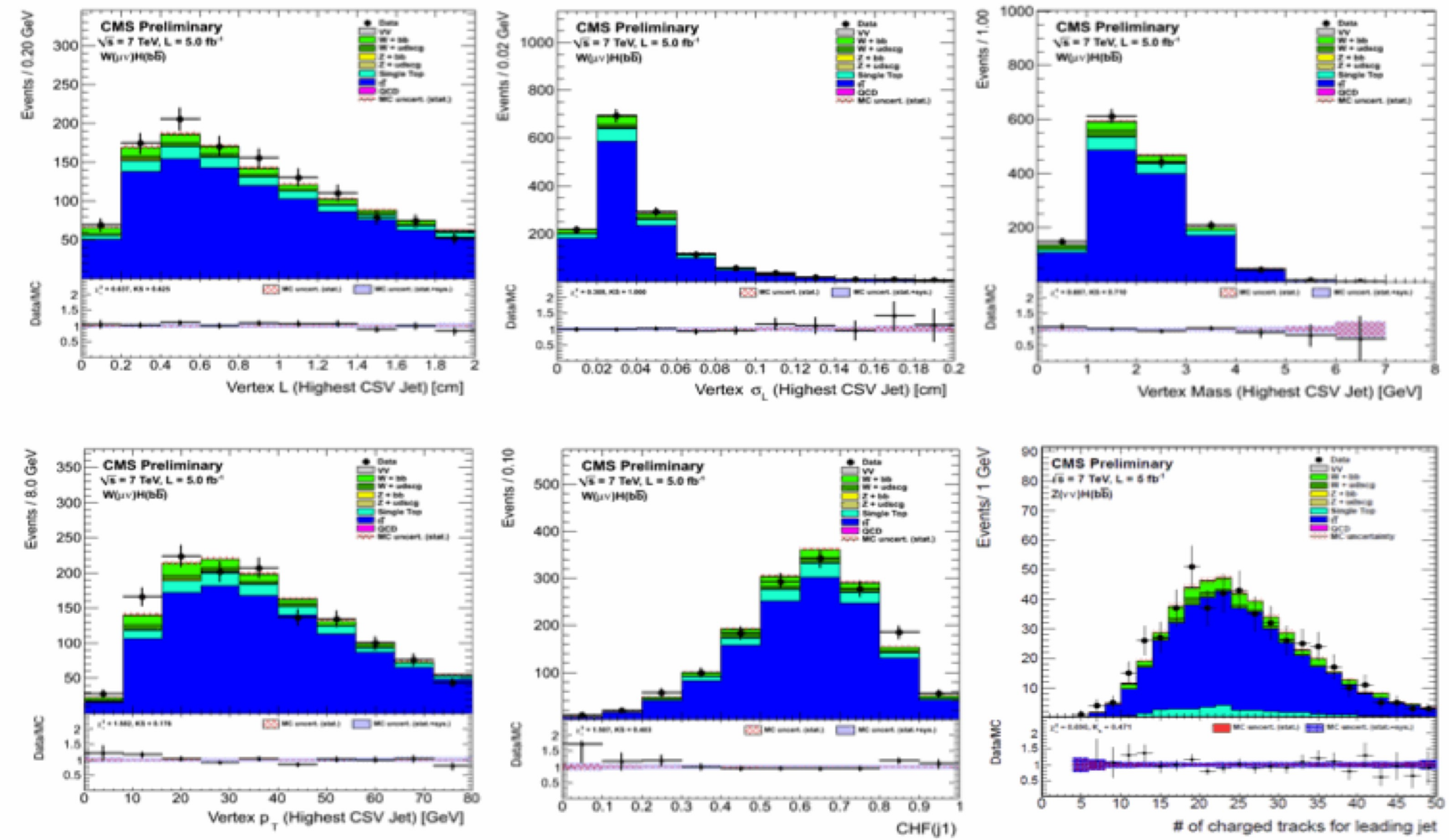
=20 BDT discriminant fits at each m_H (110-135 GeV)

σ/σ_SM compatibility

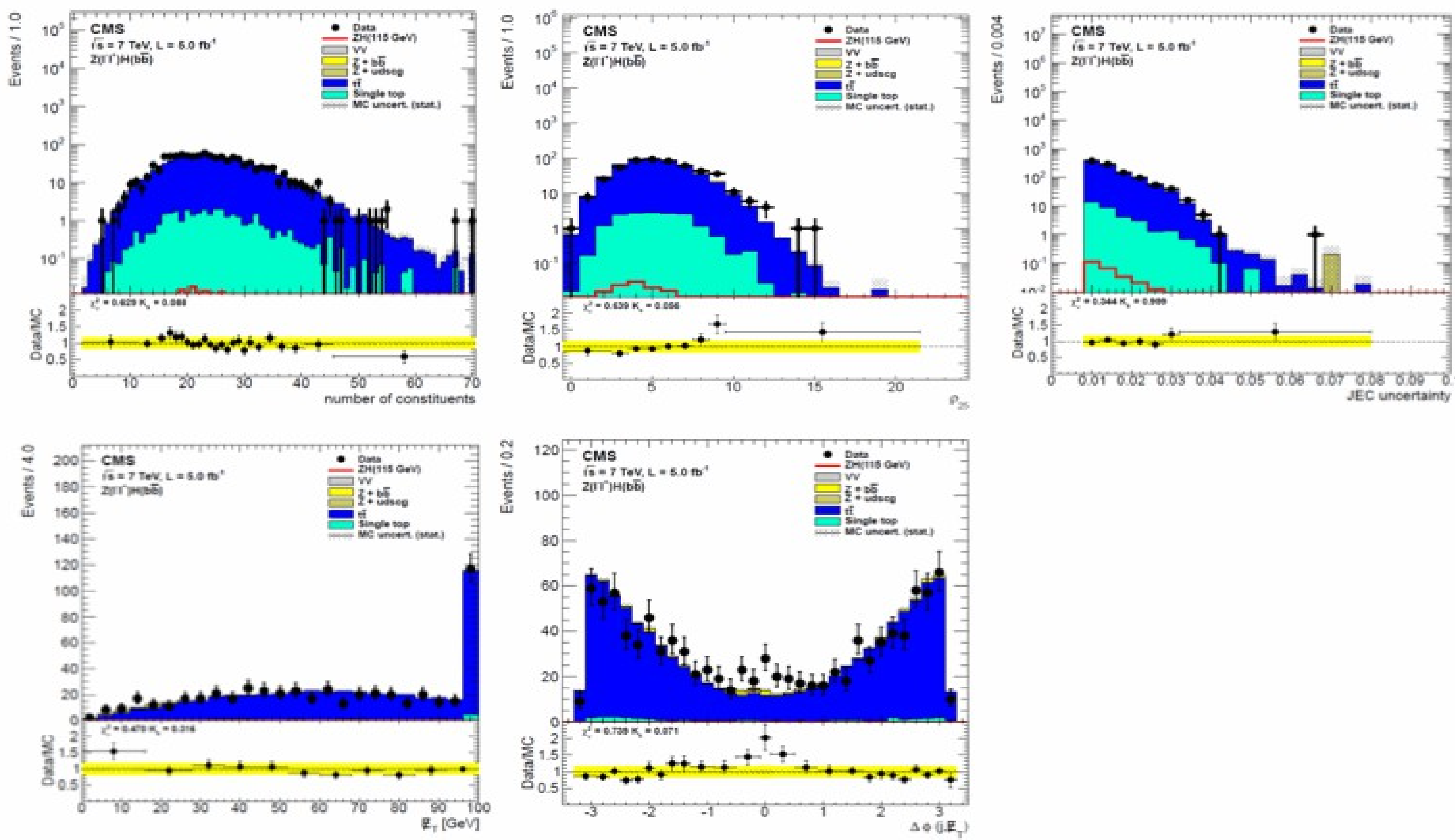


7 TeV
+ 8 TeV

Regression Input Variables



Regression Input Variables



Systematic Uncertainties

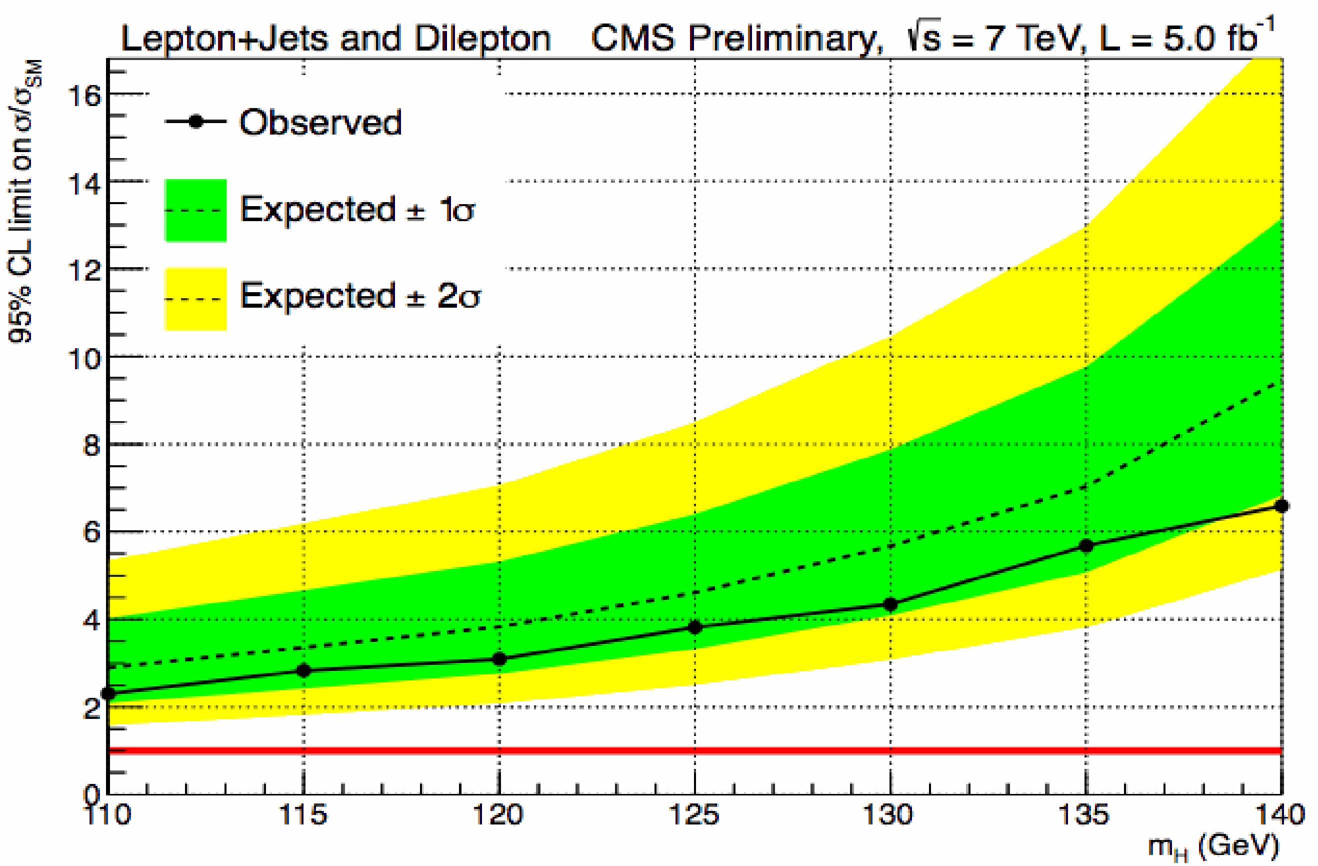
- Signal:
  Higgs cross-section: use NNLO from LHC WG, currently estimate 4% error (PDF+alphas, scale)

  p_T spectrum: recent theoretical calculations address our boosted regime: 5(10)% for Z(W)H due to electroweak corrections (http://arxiv.org/abs/0710.4749) and 10% from QCD (NNLO vs NLO, http://www.arxiv.org/abs/1107.1164)

- Background:
  Data-driven:
  Uncertainty on the SF determination →
  1) Statistical uncertainty
  2) systematic on CR definition
  From CR: V+jets (light: 2%, heavy: 12%), ttbar (3%)

  MC based: VV (30%), single top (30%)

ttH Exclusion Limits



Mass	Exp.	Obs.
110	2.90	2.30
115	3.36	2.83
120	3.83	3.09
125	4.61	3.82
130	5.67	4.35
135	7.03	5.68
140	9.47	6.59

- Sensitivity dominated by lepton+jet mode, 5-10% improvement from di-lepton mode
- Dominant uncertainties: b-tag, JES in LJ, factorization scale in DIL
- No excess seen, expect 4.6 x σ_{SM} at 125 GeV, observe 3.8 x σ_{SM}