http://cds.cern.ch/record/1668410
arXiv:1403.3047, submitted to EPJC

Measurement of **WZ and ZZ** production in pp collisions at 8 TeV in final states with **b-tagged jets** with the CMS experiment

Young Scientist Forum

Caterina Vernieri
on behalf of the CMS collaboration

Les Rencontres de Moriond on “EW Interactions and Unified Theories” - 15-22 March 2014

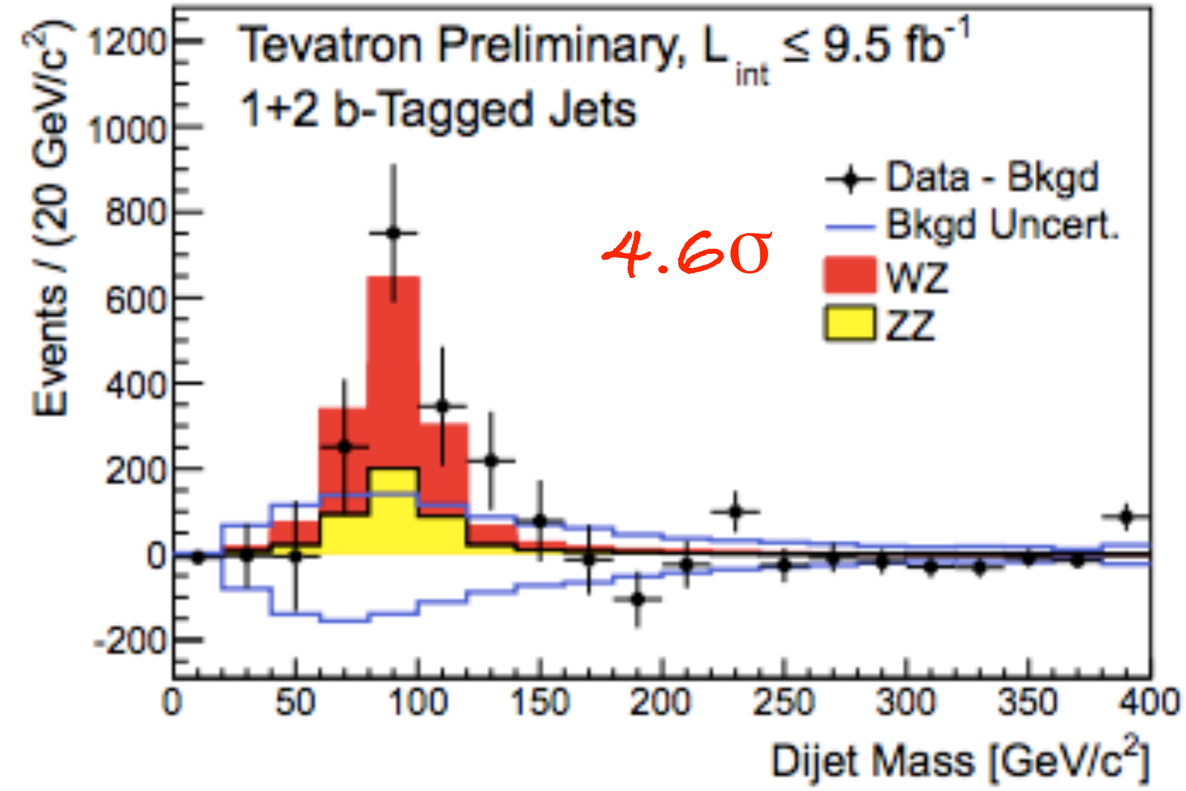
Istituto Nazionale di Fisica Nucleare

SCUOLA NORMALE SUPERIORE

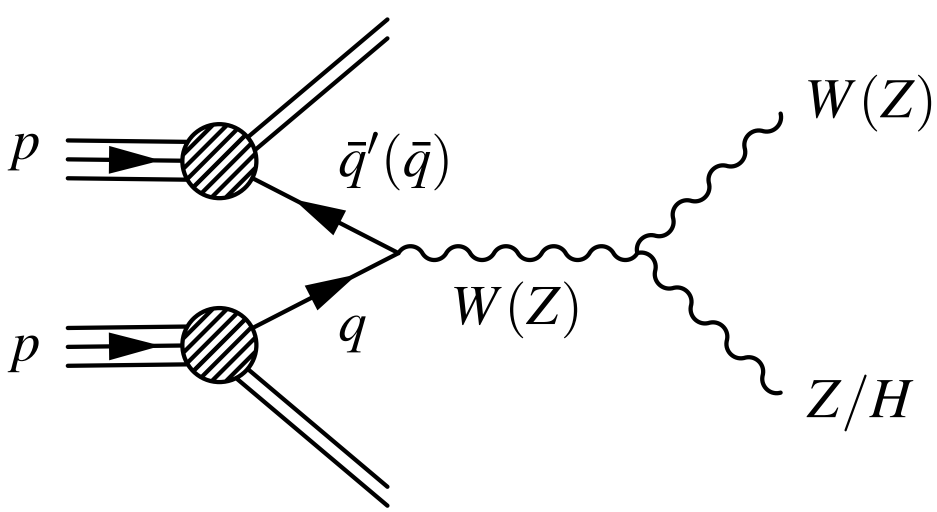
Diboson, a Standard Model Candle

Diboson production in the $b\bar{b}$ final state

- Evidence at Tevatron
- Purest bb resonance
- A standard candle to validate the Higgs signal in the boosted regime



VZ,VH: same event topology

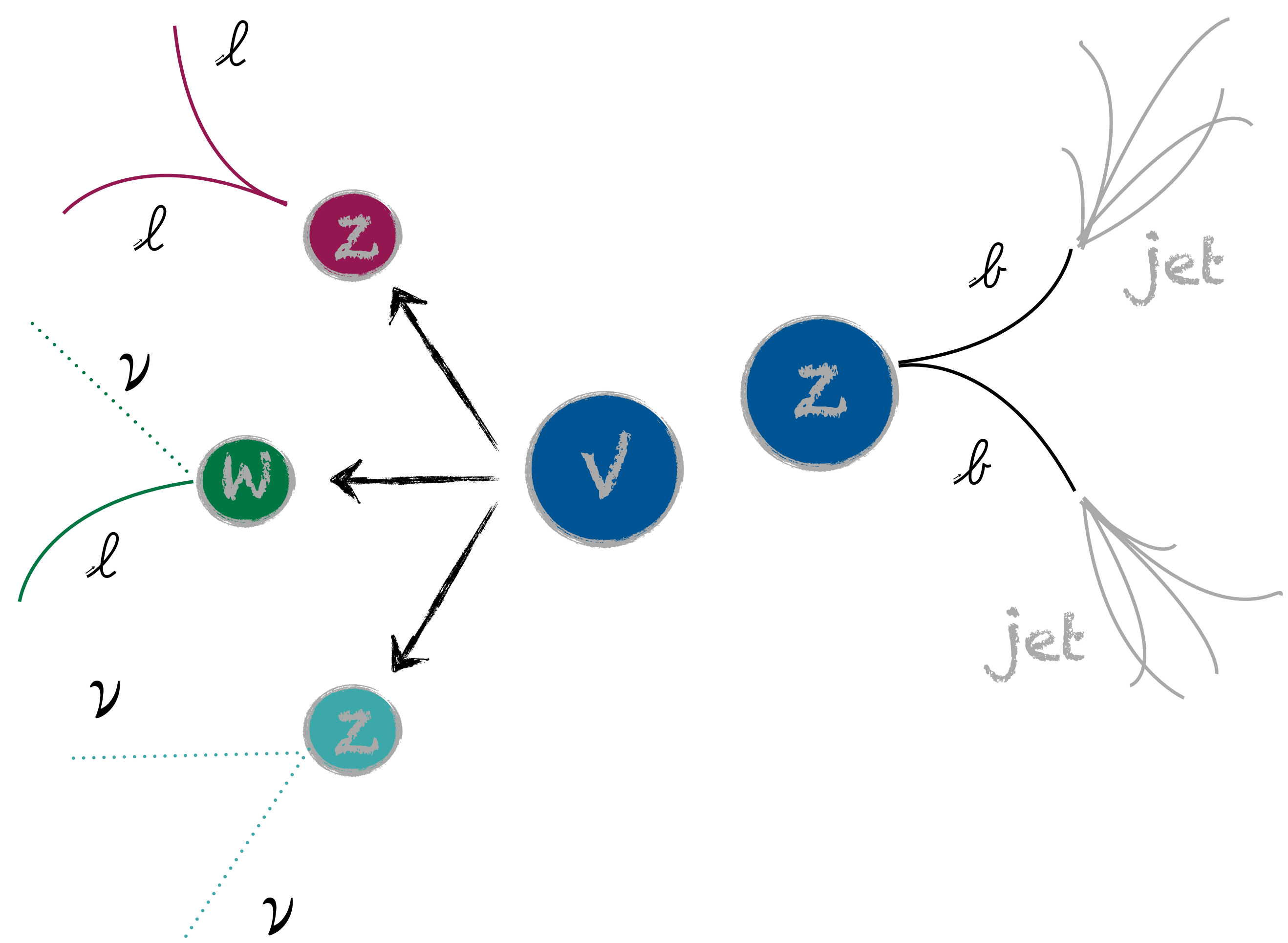


	$\sigma\cdot$BR at $\sqrt{s}$ =8 TeV [pb]		
bb	W($\ell\nu$)	Z($\ell\ell$)	Z($\nu\nu$)
Z	1.13	0.08	0.24
H	0.13	0.01	0.04

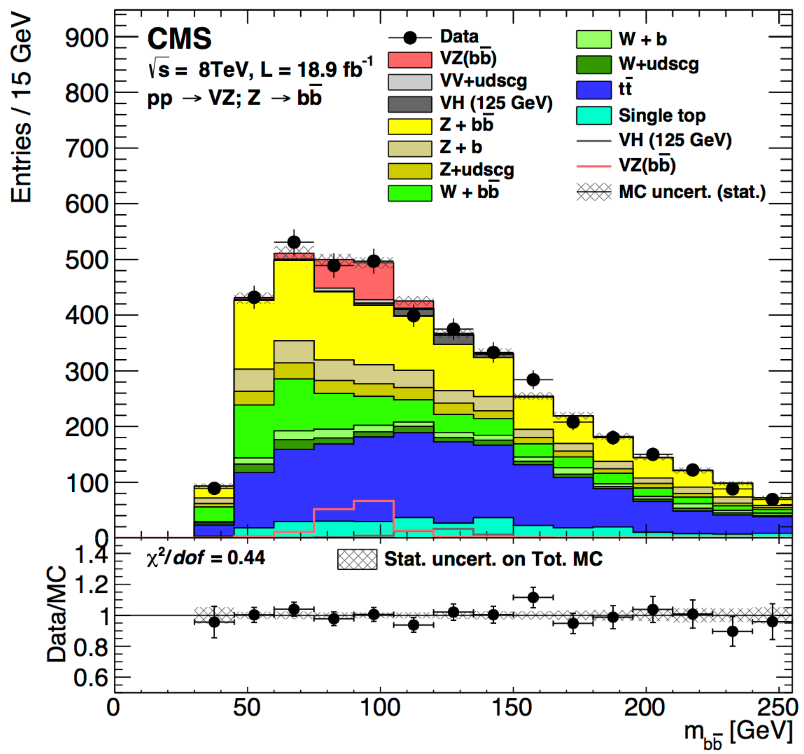
Moving from VH to VZ

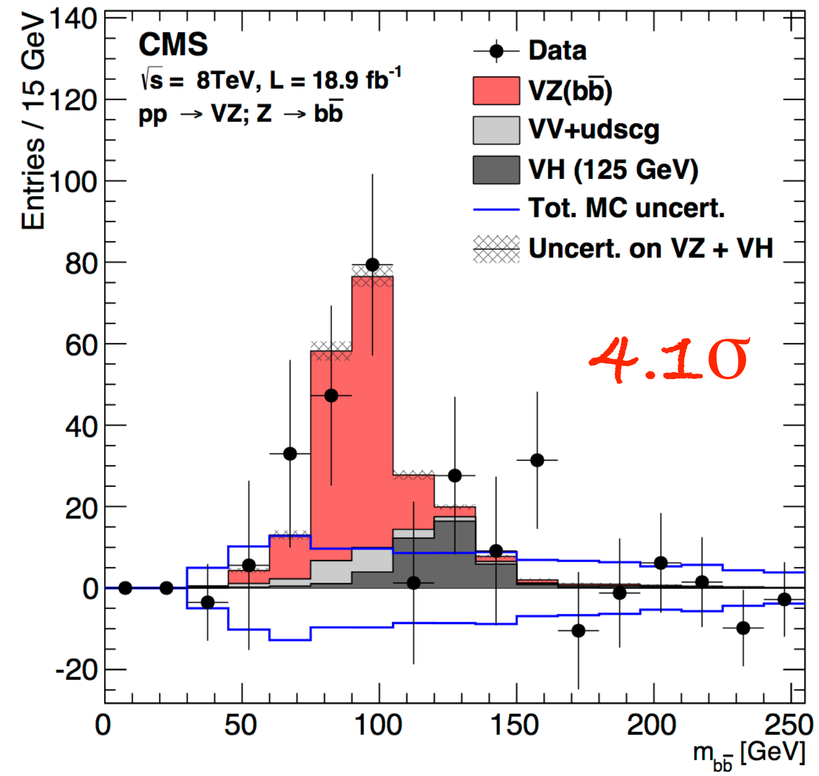
- H→ $b\bar{b}$ at LHC is searched in events where H is produced in association with a W or Z boson with high boost (~ 100 GeV)
  - events are triggered by the leptonic decay of the W/Z (e, μ, MET)
  - multi-jet QCD background is highly suppressed

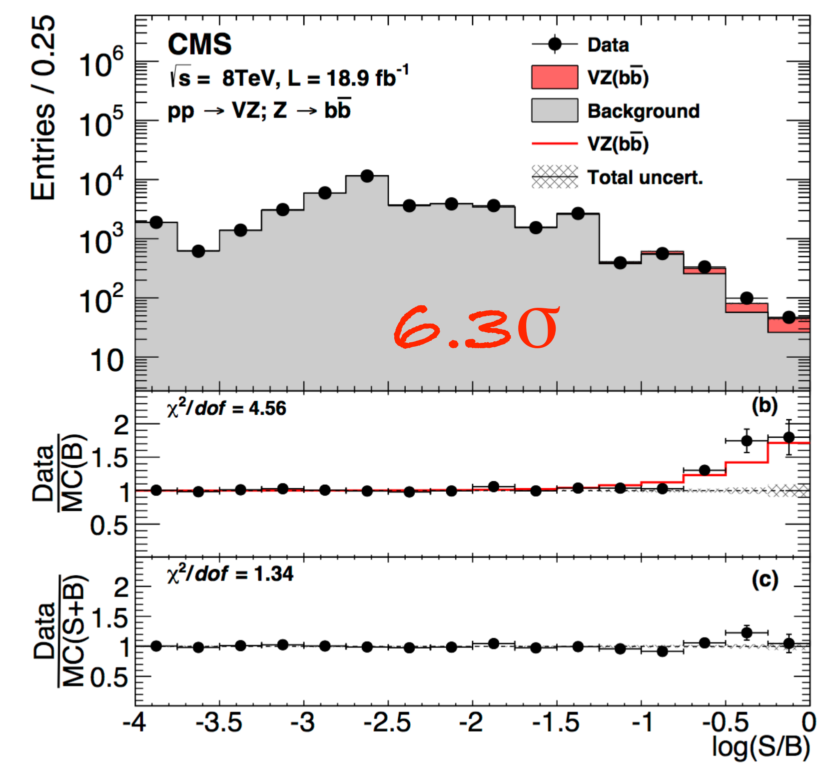
0-lepton (MET)	[100-130] [130-170] [>170]
1-lepton	[100-130] [130-180] [>180]
2-OSSF leptons	[>100]



Analysis Strategy, as for the Higgs search







Key points:

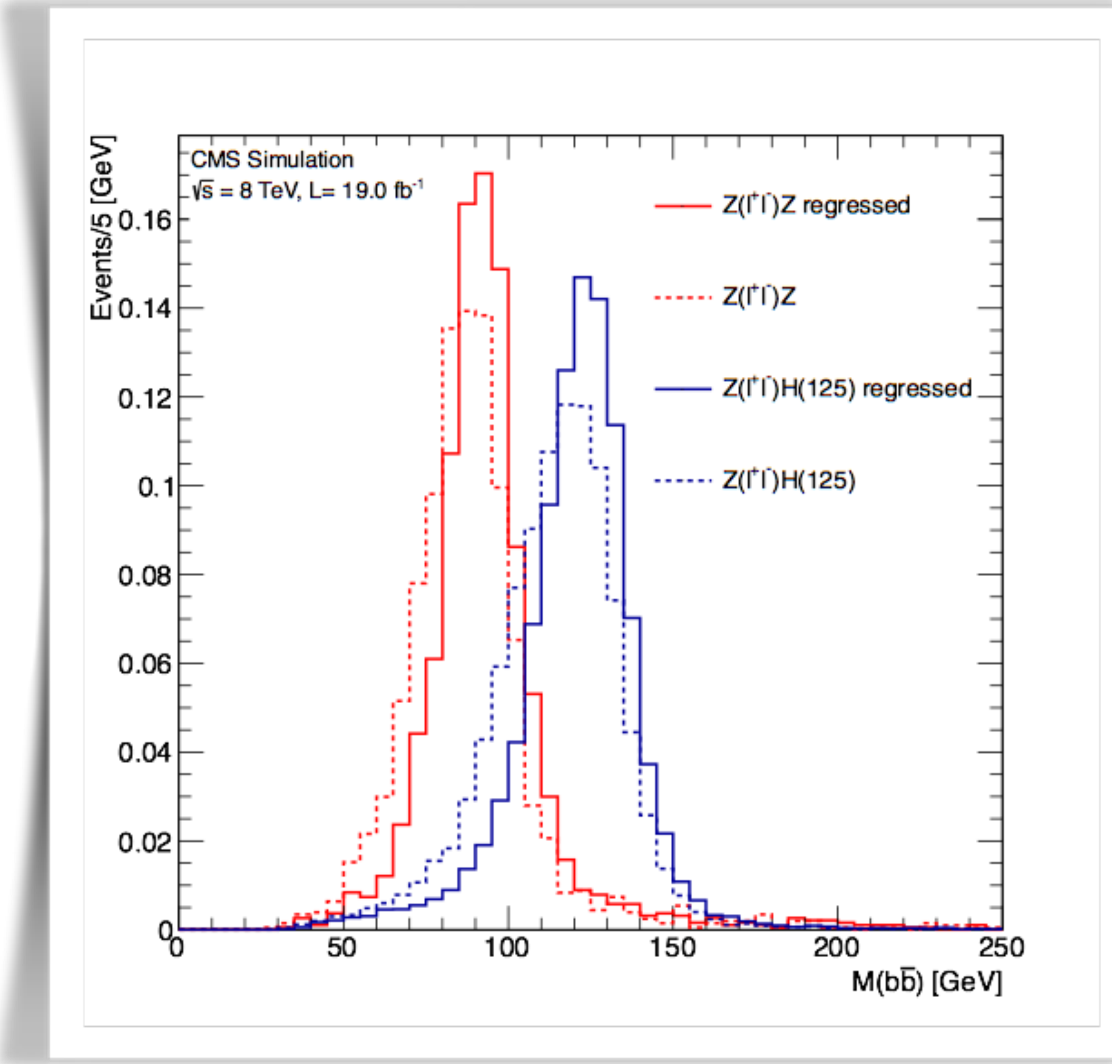
1. Extract normalization for the dominant backgrounds from the data V+0b/1b/2b and top pair production
2. A multivariate analysis, BDT
3. b-jet energy specific corrections (**regression**)

Strategy, b-jet energy MVA regression

B.R. 35% $b \rightarrow l + \nu + X$

Multidimensional calibration targeting the gen jet p_T

- ✓ Basic kinematic and jet properties
- ✓ b-tag and soft lepton information
- ✓ MET related



- Final resolution is ~15%
- The sensitivity increases by **10-20**%

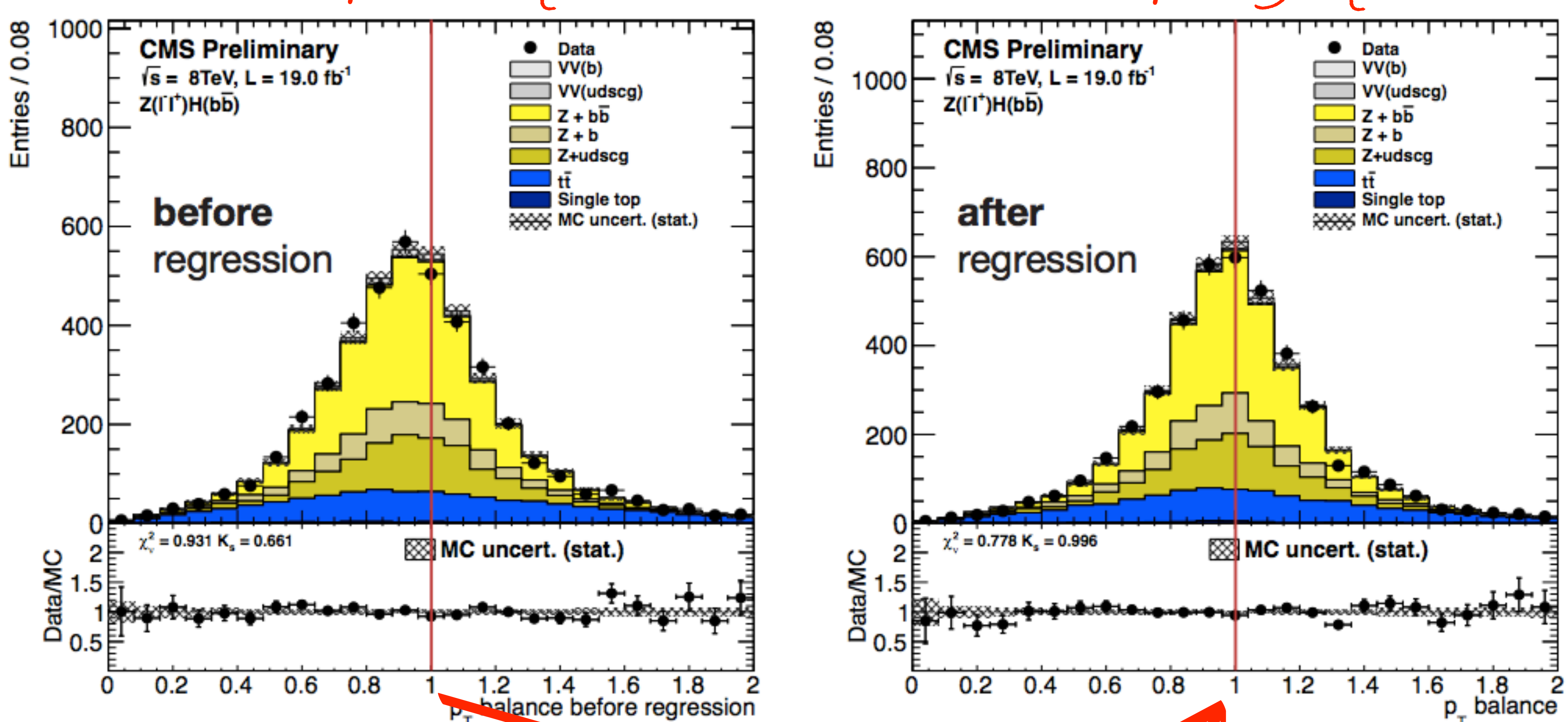
b-jet energy regression, Validation on Data

- Dijet balance in Z(ll)+bb data.

$$p_{T_{balance}} = \frac{p_T(jj)}{p_T(\ell\ell)}$$

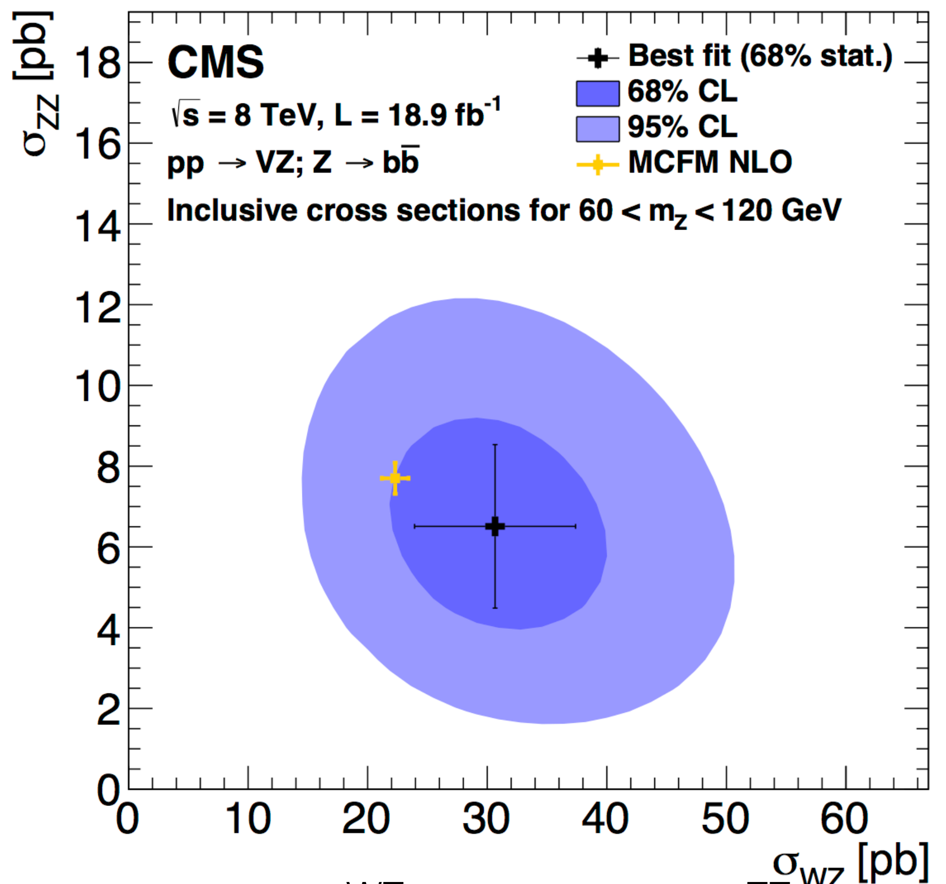
- An improved resolution and scale is observed
- Data/MC agreement improves after regression

$\sigma/\mu = 24.3\%$ $\sigma/\mu = 19.5\%$

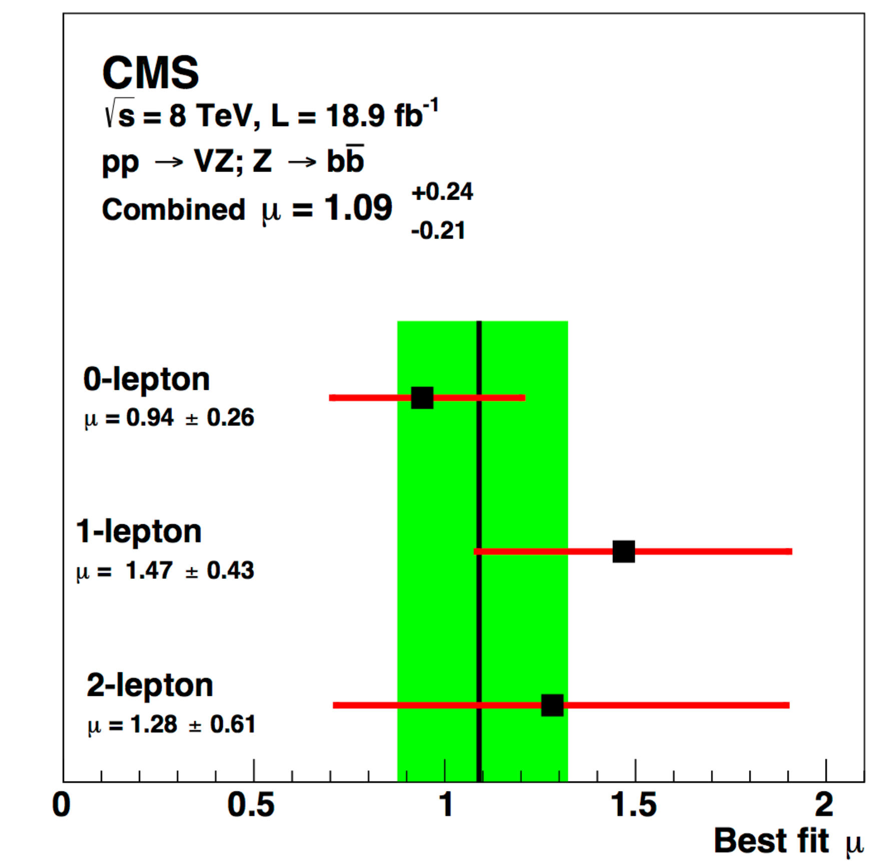


25% of improvement is found

arXiv:1403.3047, Results



Best fit: μ^{WZ}=1.37 and μ^{ZZ} =0.85



μ with respect to LO MC rescaled to NLO

First 5σ observation of the VZ(bb) at an hadron collider

Cross-sections found are in agreement with NLO MCFM prediction

σ (pp → WZ) = 4.8 ± 1.4 (stat.) ± 1.1(syst.) pb
σ (pp → ZZ) = 0.90 ± 0.23(stat.) ± 0.16(syst.) pb

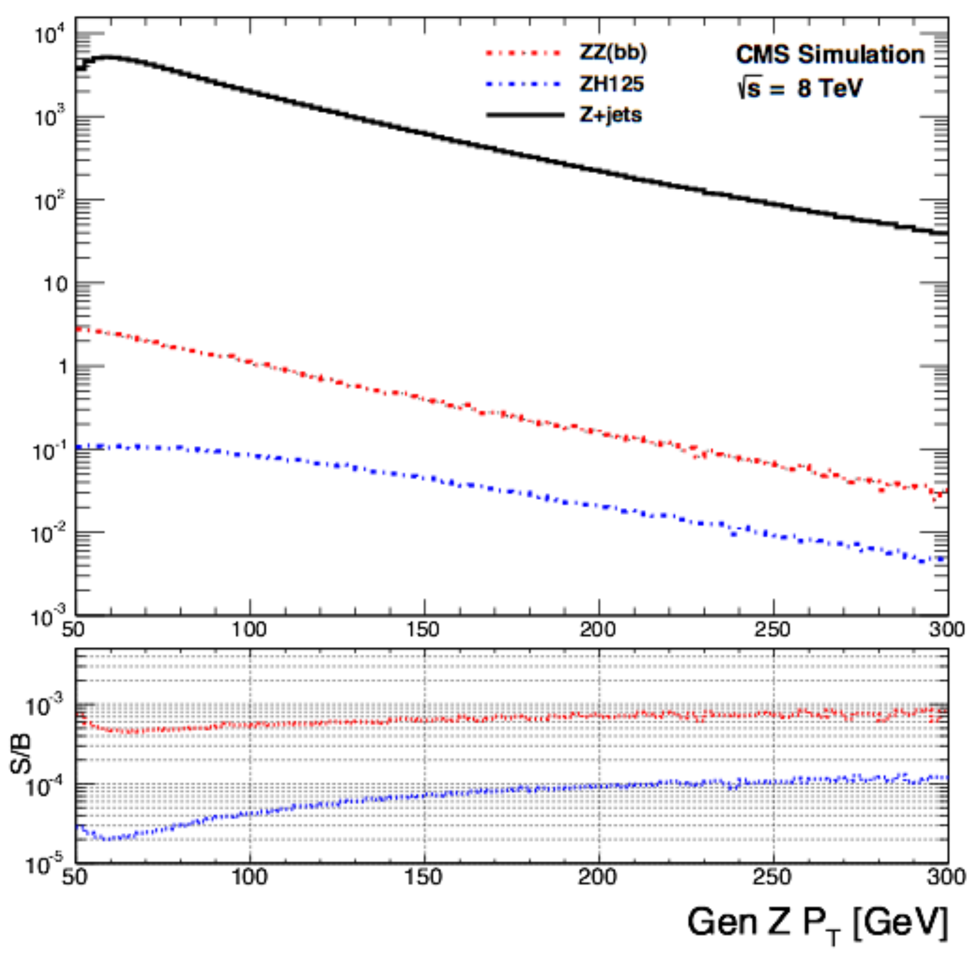
Validation of the analysis strategy for the VH(bb) search

thanks!

http://cds.cern.ch/record/1668410
arXiv:1403.3047, submitted to EPJC

Motivations, Boosted Regime

- The VH search requires a large boost (~100 GeV) for bb pair for a further reduction of the Z/W+jets and tt backgrounds
  - Different boost categories for each channel are exploited separately
- Also an improvement in the resolution of the reconstructed Z/H(bb) candidate is achieved.



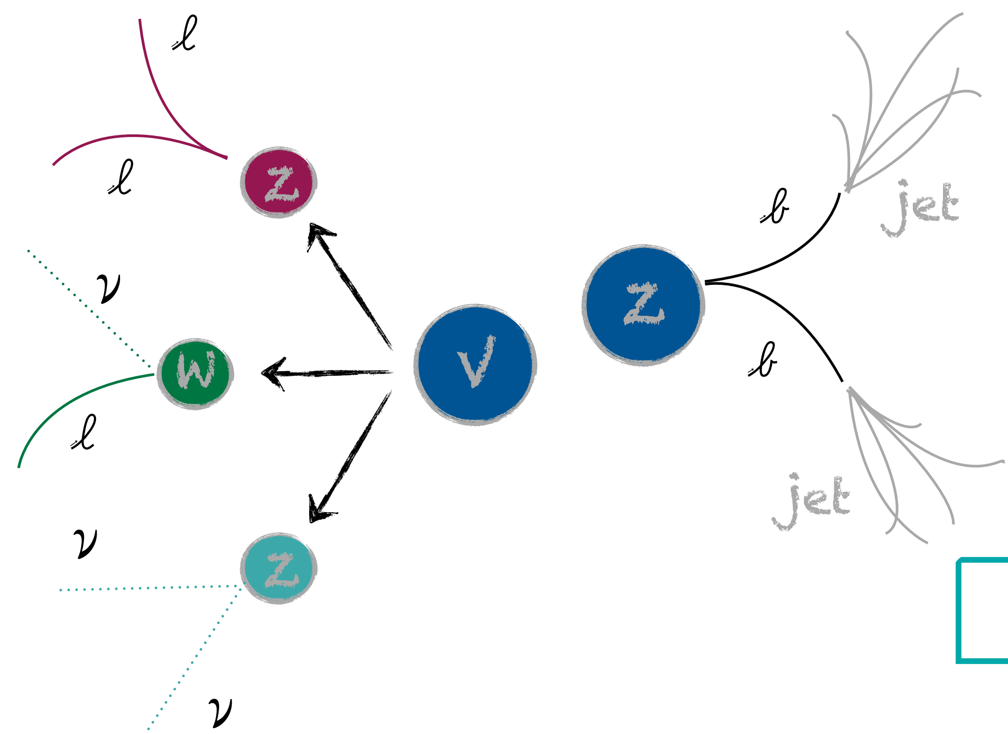
- Given the $p_T(V)$-dependency of the cross section, for VH the boosted regime improves S/B
- S/B for VZ has almost a flat dependency from $p_T(V)$
  - The trigger paths developed for VH search do not allow for $p_T(V)<100$ GeV for W(lv) and Z(vv)
  - VZ cross section measurement is then performed in the boosted regime as well
  - The phase space defined by $p_T(V)>100$ GeV is chosen in order to have the same acceptance across the different modes

Signal Topology

- Inclusive muon and electron triggers for the charged leptons mode
- Dedicated strategy for Z(νν):
  - Exploiting triggers with b-tag on-line requirements, allows to accept events with MET down to ~100 GeV

two isolated e or μ
M(ee) or M(μμ)~ M(Z)

one isolated e or μ
and MET



- (b)jets 20-30 GeV AK5 PFJets,
- The highest di-jet p_T combination in the event is selected.
- The CSV tagger is used to select the b-jets

no isolated lepton and large MET

Discriminating signal using:

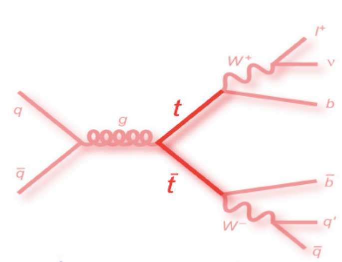
- M(bb) resolution
- Vector p_T (boost)
- **back-to-back topology** and **minimal additional jet** activity

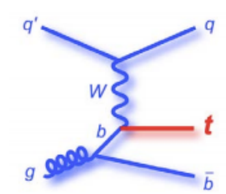
Reducible background

- W/Z+udscg (b-tagging)
- top (jet activity, Z mass window)
- single top (jet activity)

Resonant untagged

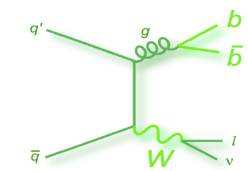
- WW, W/ZZ(jj), i.e. light

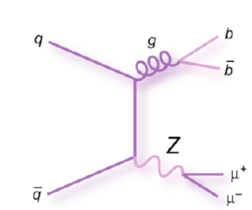




Irreducible backgrounds

- V+bb at high p_T
- VH(bb)





m(bb), VH vs. VBF

Variables can be grouped in 5:

- JET KINEMATIC
- ENERGY FRACTION
- MET RELATED
- JET STRUCTURE
- CSV-RELATED

Table 11: Input Variables to Regression

VBF	VH
	raw p_T
p_T	p_T
	E_T
	M_T
	JECUnc
η	
charged-hadron energy fraction	
neutral-hadron energy fraction	
photon energy fraction	
muon energy fraction	
electron energy fraction	CeFr
secondary vertex p_T	vtxPt
secondary vertex dL	vtx3dL
error on jet secondary vertex dL	vtx3deL
	vtxMass
CSV	
	SoftLeptPtRel
	SoftLeptPt
	SoftLeptdR
$d\phi$(MET, p_T)	$d\phi$(MET, p_T)
MET	MET
$p_{T,D}$	ptLeadTrk
	Ntot
ρ	

Background Estimate

The contributing backgrounds are :

- W/Z+jets splitted in **V+bb** and **V+udscg**
- tt pair production (**tt**)
- Single top, WW and VH
- QCD multijet

} Data-driven normalization to signal region (W/Z+jets, tt)

} MC (Single top, WW and VH)

} Negligible (QCD multijet)

- Control regions (CR) for the main backgrounds: **V+bb**, **tt**, **V+udscg** - are identified in data and used to adjust Monte Carlo estimates.
- A set of simultaneous fits is performed to the CR separately in each channel to obtain consistent data/MC **scale factors**.
  - Also a different fit among the different $p_T(V)$ categories except Z(ll) channel
- Based on CMS&Atlas studies events are split into 0/1/2b content at generator level.

Scale Factors

- Data/MC SF for each CR in each decay mode.
- Electron and muons samples are fit simultaneously to determine average SF.

Process	W($\ell\nu$)	Z($\ell\ell$)	Z($\nu\nu$)
Low			
W0b	$1.03 \pm 0.01 \pm 0.05$	–	$0.83 \pm 0.02 \pm 0.04$
W1b	$2.22 \pm 0.25 \pm 0.20$	–	$2.30 \pm 0.21 \pm 0.11$
W2b	$1.58 \pm 0.26 \pm 0.24$	–	$0.85 \pm 0.24 \pm 0.14$
Z0b	–	$1.11 \pm 0.04 \pm 0.06$	$1.24 \pm 0.03 \pm 0.09$
Z1b	–	$1.59 \pm 0.07 \pm 0.08$	$2.06 \pm 0.06 \pm 0.09$
Z2b	–	$0.98 \pm 0.10 \pm 0.08$	$1.25 \pm 0.05 \pm 0.11$
	$1.03 \pm 0.01 \pm 0.04$	$1.10 \pm 0.05 \pm 0.06$	$1.01 \pm 0.02 \pm 0.04$
Intermediate			
W0b	$1.02 \pm 0.01 \pm 0.07$	–	$0.93 \pm 0.02 \pm 0.04$
W1b	$2.90 \pm 0.26 \pm 0.20$	–	$2.08 \pm 0.20 \pm 0.12$
W2b	$1.30 \pm 0.23 \pm 0.14$	–	$0.75 \pm 0.26 \pm 0.11$
Z0b	–	–	$1.19 \pm 0.03 \pm 0.07$
Z1b	–	–	$2.30 \pm 0.07 \pm 0.08$
Z2b	–	–	$1.11 \pm 0.06 \pm 0.12$
	$1.02 \pm 0.01 \pm 0.15$	–	$0.99 \pm 0.02 \pm 0.03$
High			
W0b	$1.04 \pm 0.01 \pm 0.07$	–	$0.93 \pm 0.02 \pm 0.03$
W1b	$2.46 \pm 0.33 \pm 0.22$	–	$2.12 \pm 0.22 \pm 0.10$
W2b	$0.77 \pm 0.25 \pm 0.08$	–	$0.71 \pm 0.25 \pm 0.15$
Z0b	–	$1.11 \pm 0.04 \pm 0.06$	$1.17 \pm 0.02 \pm 0.08$
Z1b	–	$1.59 \pm 0.07 \pm 0.08$	$2.13 \pm 0.05 \pm 0.07$
Z2b	–	$0.98 \pm 0.10 \pm 0.08$	$1.12 \pm 0.04 \pm 0.10$
	$1.00 \pm 0.01 \pm 0.11$	$1.10 \pm 0.05 \pm 0.06$	$0.99 \pm 0.02 \pm 0.03$

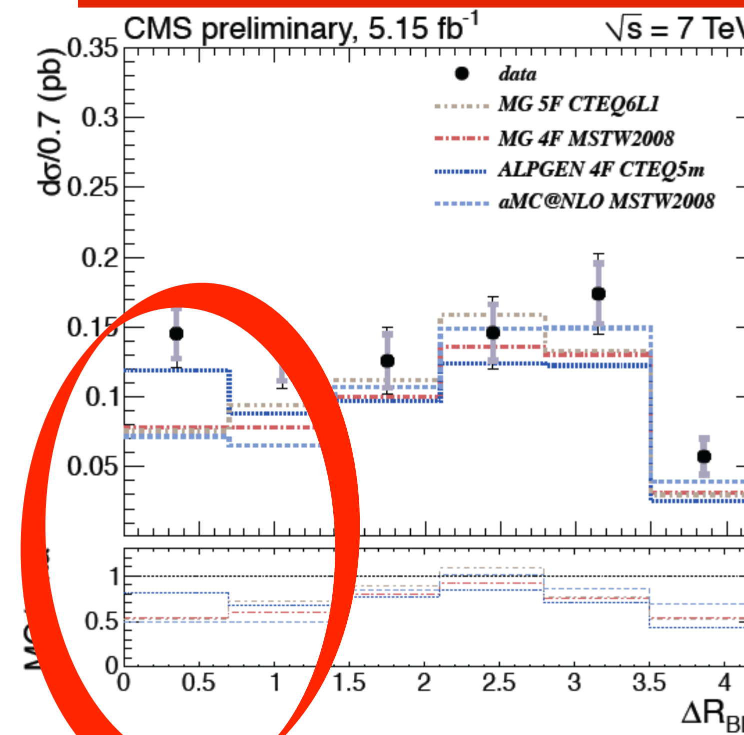
All SFs are in good agreement across the different modes

- The major part is close to 1
- V+1b is typically ~2, but:

- Not dominant background
- consistent with other CMS/Atlas studies

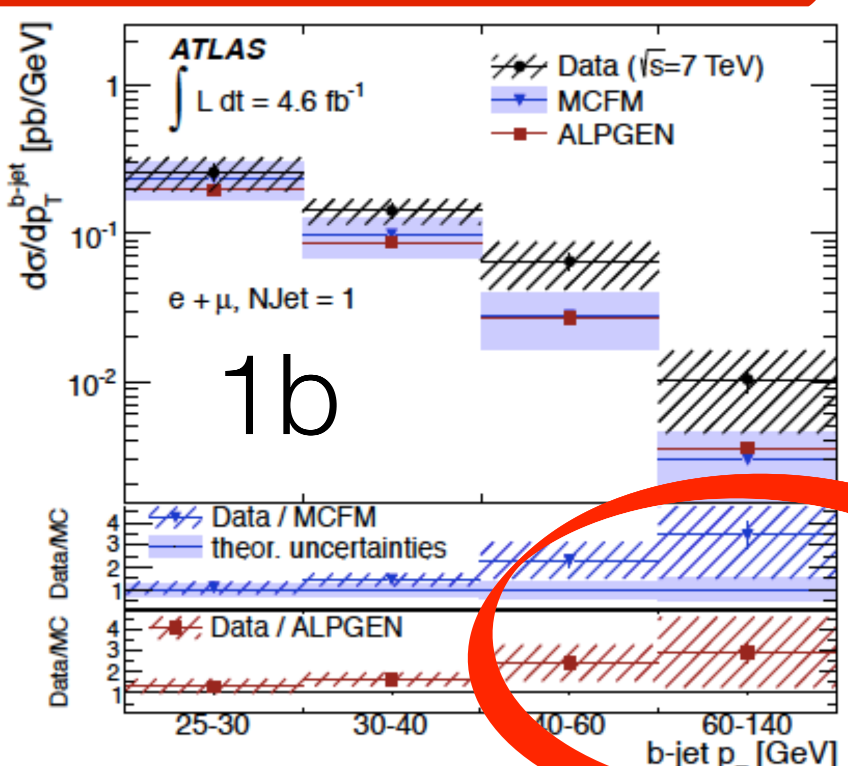
0/1/2 b splitting

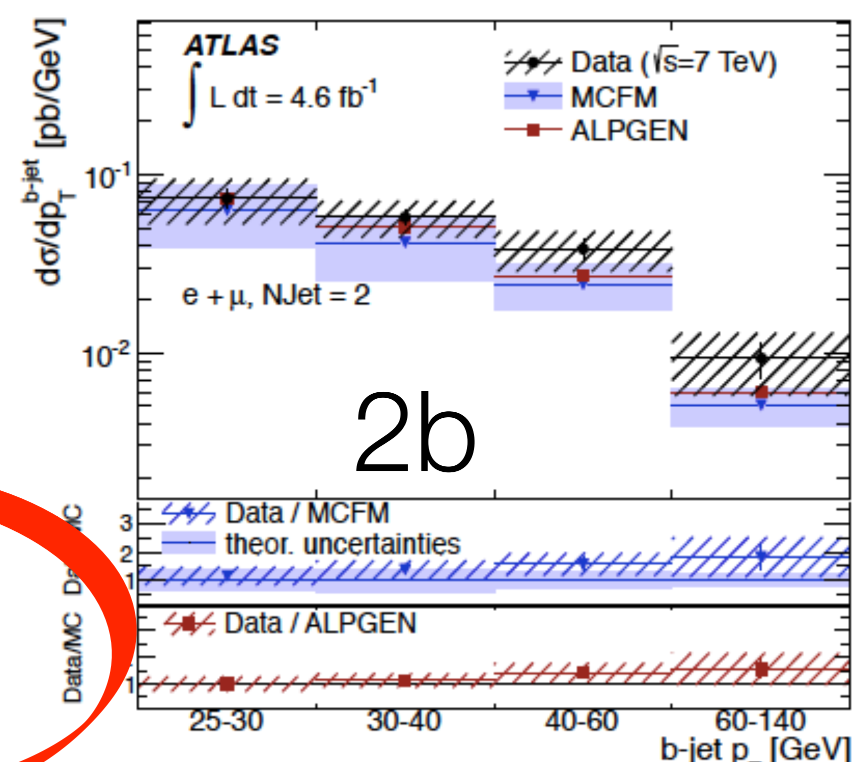
CMS, BB angular correlations, AN-11-179



Indication of gluon splitting, i.e. two b's end up in the same jet.
- SF(1b) ~ 2 is then motivated

ATLAS, W+bjets xsec measurement arXiv:1302.2929v1-13-069





Mjj Selection

Same selection criteria as applied in the VH analysis for each channel, optimized wrt purity:

Variable	W(ℓν)Z	Z(ℓℓ)Z	Z(νν)Z
$m_{\ell\ell}$	–	–	$75 < m_{\ell\ell} < 105$
$p_T(j_1)$, $p_T(j_2)$	$> 30, > 30$	$> 20, > 20$	$> 60 > 30$
p_T(jj)	> 100	–	> 110
$p_T(\ell)$	> 30	> 20	–
p_T(V)	$[100-130]$ (μ) $[100-150]$ (e) $[130-180]$ (μ) > 150 (e) > 180 (μ)	– $[100-150]$ > 150	$[100-130]$ $[130-170]$ > 170
CSV(j_1), CSV(j_2)	CSVT, > 0.5	CSVM, > 0.5	CSVT, > 0.5
$\Delta\phi$(V, H)	> 2.95	–	> 2.95
ΔR(jj)	–	$-(-, < 1.6)$	–
$N_{\rm aj}$	$= 0$	–	$= 0$
$N_{\rm al}$	$= 0$	–	$= 0$
$E_T^{\rm miss}$	> 45	$< 60.$	–
$\Delta\phi(E_T^{\rm miss}$, jet)	–	–	$> 0.7\,(> 0.7, > 0.5)$
$\Delta\phi(E_T^{\rm miss}$, trkMET)	–	–	< 0.5
$\Delta\phi(E_T^{\rm miss}$, lep)	$< \pi/2$	–	–

- Subset of events used in the BDT analysis
  - Tighter selection in b-tagging and other additional selection
- Different binning in boson-p_T
  - For Z(ll) optimized on p_T (V), selecting different regions and then wrt the cut on dR(bb)

BDT Selection

Final selection criteria optimized for each channel for the Higgs search in order to maximize signal efficiency

- Z/W selection plus loose b-tag requirements than for Mjj selection.

Variable	$W(\ell\nu)Z$	$Z(\ell\ell)Z$	$Z(\nu\nu)Z$
$m_{\ell\ell}$	–	[75 – 105]	–
$p_T(j_1)$, $p_T(j_2)$	> 30, > 30	> 20, > 20	> 60, > 30
$p_T(\mathrm{jj})$	> 100	–	> 100
$M(\mathrm{jj})$	< 250	[40 – 250]	< 250
$p_T(\ell)$	> 30	> 20	–
$p_T(\mathrm{V})$	[100 – 130] (μ) [100 – 150] (e) [130 – 180] (μ) > 150 (e) > 180 (μ)	– – > 100	[100 – 130] [130 – 170] > 170
$\mathrm{CSV_{max}}$, $\mathrm{CSV_{min}}$	> 0.40, > 0.40	> 0.50, > 0.24	> 0.67, > 0.24
N_{aj}	–	–	< 2 (–,–)
N_{al}	$= 0$	–	$= 0$
$\mathrm{E_T^{miss}}$	> 80 (τ)	–	–
$\Delta\phi(\mathrm{V,H})$	–	–	> 2.0
$\Delta\phi(\mathrm{E_T^{miss}}, \mathrm{jet})$	–	–	> 0.7 (> 0.7, > 0.5)
$\Delta\phi(\mathrm{E_T^{miss}}, \mathrm{trkMET})$	–	–	< 0.5
pfMET significance	–	–	> 3 (–,–)
$\Delta\phi(\mathrm{E_T^{miss}}, \mathrm{lep})$	$< \pi/2$	–	–
Tightened Lepton Iso.	< 0.075 (–, –)	–	–

Strategy, BDT Training

- Separate BDTs trained in each channel and for each boost category
- **Inputs** used allow to exploit the complete kinematic information of the event.
- All variables are monitored in the CR, as well the outputs distribution
- VZ(jj) and VH are treated as backgrounds in the training.

Variable
p_T(j): transverse momentum of each Z($b\bar{b}$) 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 Z($b\bar{b}$) daughter with largest CSV value
CSV_{min}: value of CSV for the Z($b\bar{b}$) daughter with second largest CSV value
$\Delta\phi$(V, H): azimuthal angle between V and dijet
$\Delta\eta$(jj): difference in η between Z($b\bar{b}$) daughters
ΔR(jj): distance in η–ϕ between Z($b\bar{b}$) daughters
N_{aj}: number of additional jets
$\Delta\theta_{pull}$: color pull angle [2]
$\Delta\phi$(E_T^{miss}, jet): azimuthal angle between $E_T^{miss}$and the closest jet (only for Z($\nu\nu$))
$maxCSV_{aj}$: maximum CSV of the additional jets in an event (only for Z($\nu\nu$)$andW$($\ell\nu$))
minΔR(H, aj): mimimum distance between an additional jet and the Z($b\bar{b}$) candidate (only for Z($\nu\nu$) and W($\ell\nu$))
Angular variables: VZ system mass, Angle Z-Z*, Angle Z-l, Angle Z-jet (only for Z($\ell\ell$))

Strategy, multi BDT

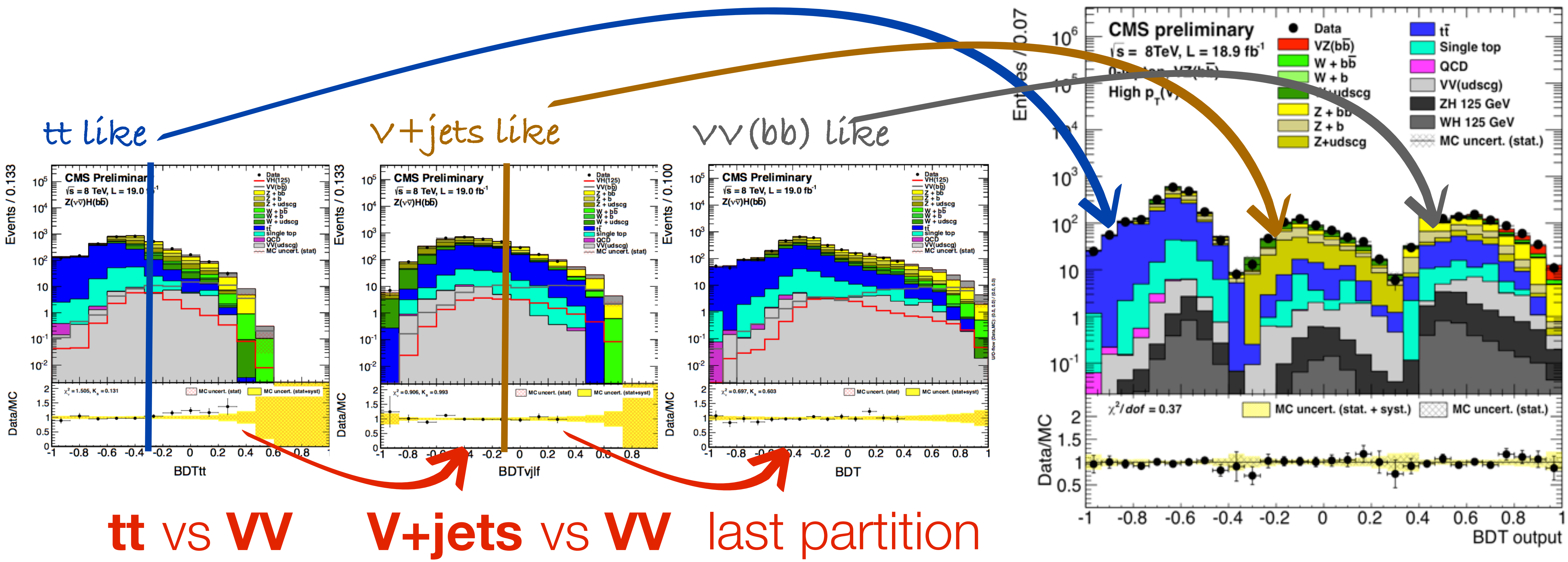
- A background specific BDT training in the Zνν and W(eν,μν) channels is performed to achieve a better discrimination of the signal.
- A series of BDTs is trained targeting 2 different backgrounds: tt and V+jets.
- The final discriminant is made of 3 regions:

  tt- enriched

  V+jets-enriched

  Signal-like region

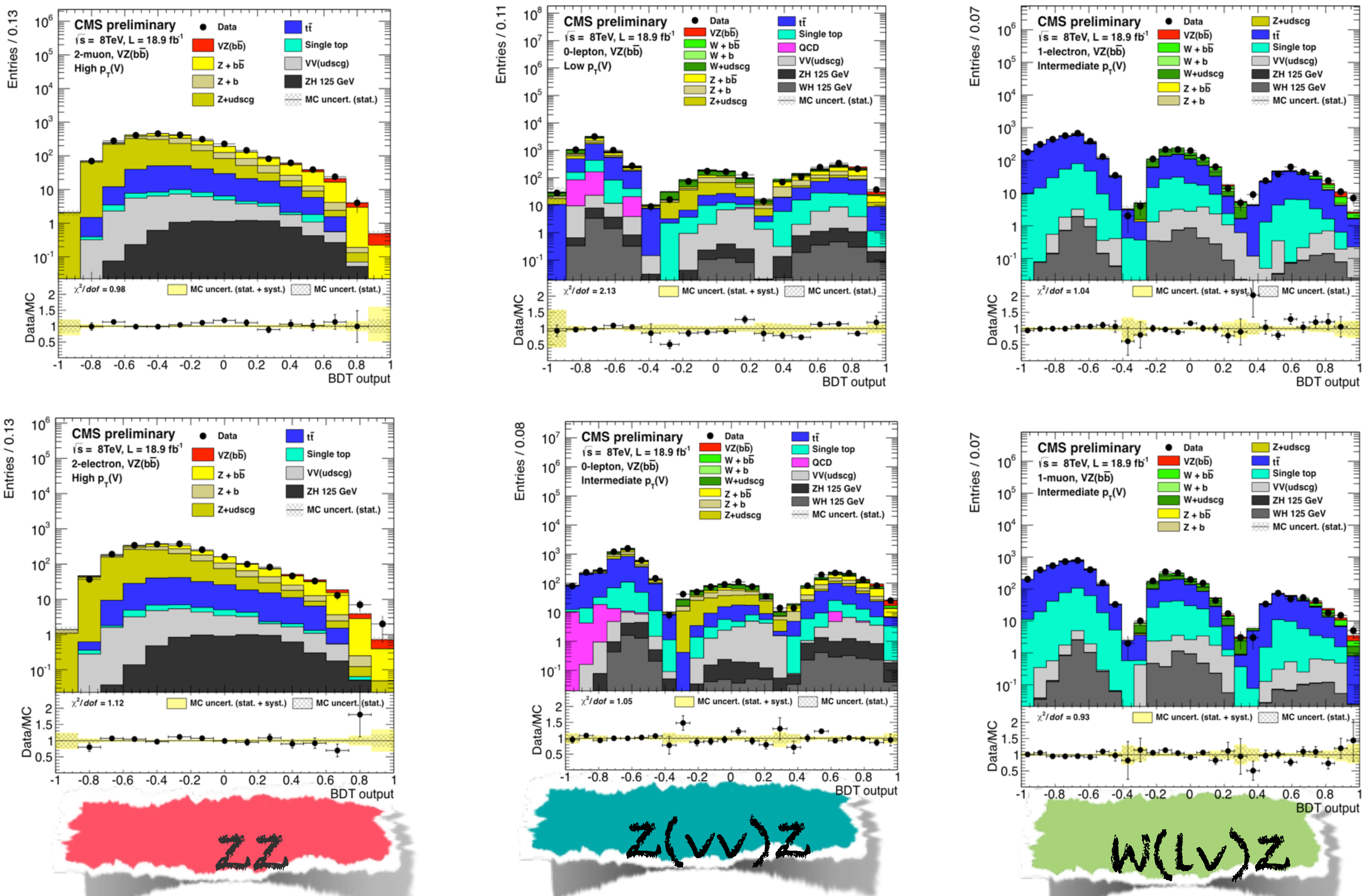
[*5-10% gain in exp. sensitivity wrt the single BDT approach*]



Strategy, BDT

Separate BDTs trained in each channel and for each boost category, separately for e and μ:

- BDT: 2 Z(ll)
- multi-BDT: 6 W(lν), 3 Z(νν)



Systematics

- Shape systematic
  - btag, JER, JES, trigger, generator modeling, bin-by-bin stats
- logNormal systematic
  - SF, signal cross section

		Individual contributions to uncertainty	
Source of uncertainty	Type	μ_{WZ} (%)	μ_{ZZ} (%)
Luminosity	norm	3.3	3.2
Lepton efficiency and trigger	norm	1.9	0.6
0-lepton triggers	dist	—	1.6
Jet energy scale	dist	7.2	6.4
Jet energy resolution	dist	6.1	5.9
E_T^{miss}	dist	3.3	1.8
b tagging	dist	7.7	5.7
VZ cross section (theory)	norm	13.4	13.4
Monte Carlo statistics	dist	5.5	3.6
Backgrounds (from data)	norm	12.5	11.5
Single-top and VH (from simulation)	norm	1.9	—
MC modeling of V+jets and $t\bar{t}$	dist	4.7	4.8

Diboson Physics at 8 TeV

CMS

$\sigma (pp \rightarrow ZZ \rightarrow 2l2l') = 8.4 \pm 1.0(\text{stat.}) \pm 0.7(\text{syst.}) \pm 0.4(\text{lumi.})$ pb — SMP-12-024

$\sigma (pp \rightarrow WZ \rightarrow 2ll'\nu) = 24.61 \pm 0.76(\text{stat.}) \pm 1.13(\text{syst.}) \pm 1.08(\text{lumi.})$ pb — SMP-12-006

$\sigma (pp \rightarrow ZZ(bb)) = 6.5 \pm 1.7(\text{stat.}) \pm 1.0(\text{syst.}) \pm 0.9(\text{th.}) \pm 0.2(\text{lum.})$ pb
$\sigma (pp \rightarrow WZ(bb)) = 30.7 \pm 9.3(\text{stat.}) \pm 7.1(\text{syst.}) \pm 4.1(\text{th.}) \pm 1.0(\text{lum.})$ pb — SMP-13-011

ATLAS

$\sigma (pp \rightarrow ZZ \rightarrow 2l2l') = 7.1 \pm 0.5(\text{stat.}) \pm 0.3(\text{syst.}) \pm 0.2(\text{lumi.})$ pb — CONF-13-020

$\sigma (pp \rightarrow WZ \rightarrow 2ll'\nu) = 20.3 \pm 0.8(\text{stat.}) \pm 1.2(\text{syst.}) \pm 0.7(\text{lumi.})$ pb — CONF-13-021