H→b\bar{b} at CMS

Caterina Vernieri
on behalf of the CMS collaboration
H → b¯b at LHC
$H \rightarrow b\bar{b}$ at LHC

- Unique final state to measure the **coupling with down-type quark**
H → b¯b at LHC

• Unique final state to measure the coupling with down-type quark
• large branching fraction ~58%, dominates total width
H → b̄b at LHC

- Unique final state to measure the coupling with down-type quark
- Large branching fraction ~58%, dominates total width
$H \rightarrow b\bar{b}$ at LHC

- Unique final state to measure the **coupling with down-type quark**
- Large branching fraction $\sim 58\%$, dominates total width
- Overwhelming background from QCD production of $b$ quarks
H → b̄b at LHC

- Unique final state to measure the **coupling with down-type quark**
- Large branching fraction ~58%, dominates total width
- Overwhelming background from QCD production of b quarks
  - $10^7$ bigger, GF cannot be used
H → b̄b at LHC

- Unique final state to measure the coupling with down-type quark
- Large branching fraction ~58%, dominates total width
- Overwhelming background from QCD production of b quarks
  - 10^7 bigger, GF cannot be used
- Signal topology of the production mechanism is exploited
H → b\bar{b} at LHC

- Unique final state to measure the **coupling with down-type quark**
- Large branching fraction ~58%, dominates total width
- Overwhelming background from QCD production of b quarks
  - 10^7 bigger, GF cannot be used
- Signal topology of the production mechanism is exploited
  - VH associated production, V decaying leptonically
H → b\bar{b} at LHC

- Unique final state to measure the **coupling with down-type quark**
- Large branching fraction ~58%, dominates total width
- Overwhelming background from QCD production of b quarks
  - $10^7$ bigger, GF cannot be used
- Signal topology of the production mechanism is exploited
  - **VH** associated production, V decaying leptonically
  - **VBF** mechanism, a very peculiar topology but no leptons
H → b\bar{b} at LHC

- Unique final state to measure the coupling with down-type quark
- Large branching fraction ~58%, dominates total width
- Overwhelming background from QCD production of b quarks
  - 10^7 bigger, GF cannot be used
- Signal topology of the production mechanism is exploited
  - VH associated production, V decaying leptonically
  - VBF mechanism, a very peculiar topology but no leptons
  - ttH complementary to the VH channel
    - dominant backgrounds is t\bar{t} + jets instead of V + jets
H → b¯b at LHC

- Unique final state to measure the coupling with down-type quark
- Large branching fraction ~58%, dominates total width
- Overwhelming background from QCD production of b quarks
  - 10^7 bigger, GF cannot be used
- Signal topology of the production mechanism is exploited
  - VH associated production, V decaying leptonically
  - VBF mechanism, a very peculiar topology but no leptons
  - ttH complementary to the VH channel
    - Dominant backgrounds is t¯t + jets instead of V + jets

---

GF \(87\%\)
VH \(5\%\)
VBF \(7\%\)

See Bianchini's talk ttH \(0.6\%\)
Why so challenging?

Comparison with the discovery channel

**H(bb) searches need:**

- exploit all possible information from the event to improve S/B
- improve m(bb) resolution

<table>
<thead>
<tr>
<th></th>
<th>$H \rightarrow 4 \ell$</th>
<th>$H \rightarrow b\bar{b}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BR</td>
<td>0.013%</td>
<td>58%</td>
</tr>
<tr>
<td>mass resolution</td>
<td>1%</td>
<td>10%</td>
</tr>
<tr>
<td>signal efficiency</td>
<td>30%</td>
<td>1.3%</td>
</tr>
<tr>
<td>S/B</td>
<td>2</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Events / 3 GeV
Quick look at the backgrounds

VH example

**signal**

**irreducible backgrounds**

0-lepton (MET)
1-lepton [e,μ,τ]
2-OSSF leptons [ee,μμ]
Quick look at the backgrounds

VH example

**signal**

**irreducible backgrounds**

and diboson, of course

0-lepton (MET)
1-lepton \([e, \mu, \tau]\)
2-OSSF leptons \([ee, \mu\mu]\)
Quick look at the backgrounds

0-lepton (MET)
1-lepton [e, μ, τ]
2-OSSF leptons [ee, μμ]

irreducible backgrounds

and diboson, of course
b-tagging
Combined Secondary Vertex

- The CSV through multivariate technique combines

  **Track information**
  - Impact parameter significance of the most energetic tracks

  **Vertex information**
  - if available or pseudo vertex from displaced tracks

<table>
<thead>
<tr>
<th></th>
<th>b</th>
<th>c</th>
<th>light [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loose</td>
<td>85</td>
<td>32</td>
<td>10</td>
</tr>
<tr>
<td>Medium</td>
<td>70</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>Tight</td>
<td>50</td>
<td>6</td>
<td>0.1</td>
</tr>
</tbody>
</table>
after b-tagging, mostly irreducible background is left

$m(b\bar{b})$ raw dijet resolution is ~ 10%

improvable to achieve a better signal discrimination
b-jet energy MVA regression

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

Multidimensional calibration targeting the jet \( p_T \) at generator level

✓ Basic kinematic and jet structure
✓ Secondary Vertex and soft lepton information
✓ MET related (as kinematic constraint)

✓ Final resolution improves by 15-25%
✓ The sensitivity increases by 10-20%
✓ The VZ/VH separation power improves

ICHEP 2-9 July 2014 Valencia                  Caterina Vernieri (Pisa)
Validation on Data, Diboson

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

Purest $b\bar{b}$ resonance

A standard candle to validate the Higgs search

<table>
<thead>
<tr>
<th>$\sigma \cdot$BR at $\sqrt{s} = 8$ TeV [pb]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$bb$</td>
</tr>
<tr>
<td>$Z$</td>
</tr>
<tr>
<td>$H$</td>
</tr>
</tbody>
</table>

VZ, VH: same event topology

6.3σ, first observation of the VZ($b\bar{b}$) at an hadron collider

$\sigma (pp \rightarrow WZ) = 4.8 \pm 1.4$ (stat.) $\pm 1.1$ (syst.) pb
$\sigma (pp \rightarrow ZZ) = 0.90 \pm 0.23$ (stat.) $\pm 0.16$ (syst.) pb
VH, Key elements

- V boson is required to have high boost (~100 GeV) - categorization in $p_T(V)$ bins
- multi-jet QCD background is highly suppressed
- $m(b\bar{b})$ invariant resolution improved in this phase space
- Extract normalization for the dominant and irreducible backgrounds from the data V+0b/1b/2b and top pair production
- Use of a multivariate discriminant, BDT
  14 BDTs - shape analyses (for each lepton mode and boost category)
VH(b¯b) reported an excess of $2.1\,\sigma$ in agreement with SM H expectation at 125 GeV

<table>
<thead>
<tr>
<th></th>
<th>$\mu = \sigma/\sigma_{\text{SM}}$</th>
<th>exp. sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CMS</strong></td>
<td>1.0 ± 0.5</td>
<td>2.1 $\sigma$</td>
</tr>
<tr>
<td><strong>CDF</strong></td>
<td>2.5 ± 1.0</td>
<td>1.3 $\sigma$</td>
</tr>
<tr>
<td><strong>D0</strong></td>
<td>1.2 ± 1.1</td>
<td>1.5 $\sigma$</td>
</tr>
<tr>
<td><strong>D0+CDF</strong></td>
<td>1.95 ± 0.75</td>
<td>1.9 $\sigma$</td>
</tr>
<tr>
<td><strong>ATLAS</strong></td>
<td>0.2 ± 0.9</td>
<td>1.6 $\sigma$</td>
</tr>
</tbody>
</table>

**Coupling results**

MVA analysis

S/B sorted

Entries / 0.25

$\chi^2$/dof = 0.98

$\chi^2$/dof = 0.84

ICHEP 2-9 July 2014 Valencia

Caterina Vernieri (Pisa)
VH, ZH new theory cross section available

- ZH production at LHC is mostly $qqZH$ (~95%)
- **NNLO** QCD corrections to $qqZH$ are included in the VH result just presented

![Diagram showing $qqZH@NLO$ and $qqZH@NNLO$]

- $ggZH$ calculations were not ready and not included
  - spectrum available - it peaks at $p_T(H) \sim 150$ GeV
  - $\sigma$ corrections up to 30% at the highest $p_T$ category
  - **Back of the envelope**: folding in the corrected ZH $p_T$ spectrum, combining with WH as is, overall VH theory prediction scales up by 10%

**Estimated effect:**
- roughly 10% decrease (increase) in $\mu$ (sensitivity)
Search for the standard model Higgs boson produced in vector boson fusion, and decaying to bottom quarks

**CMS-HIG-13-011**

The CMS Collaboration

---

**VBF**

- 4 relatively hard jets
- jet flavor tagging
  - 2 central **b-jets**
  - 2 light quark jets (quark/gluon id)
- large $m_{qq}$ and $\Delta \eta_{qq}$
- suppressed color flow between the $b\bar{b}$ and VBF jets

---

**Key points:**

1. A topological trigger on signal main properties
2. Neural Network - categorization in S/B bins
3. Fit the $m(b\bar{b})$ spectrum in each bin

work in progress

final results soon

---

3.6 x SM @ 125
Conclusion & Perspectives

CMS searches for $H(bb\bar{b})$ consistent with the SM prediction of a Yukawa structure

Evidence for $H$ decaying to fermions

$H \rightarrow \tau\tau$ and $VH(bb\bar{b})$, 3.8 $\sigma$ (exp. 4.4 $\sigma$)

see Steggemann’s talk

An exciting program is expected to start at the LHC
Conclusion & Perspectives

CMS searches for $H(b\bar{b})$ consistent with the SM prediction of a Yukawa structure

Evidence for $H$ decaying to fermions
$H \rightarrow \tau\tau$ and $VH(b\bar{b})$, $3.8 \sigma$ (exp. $4.4 \sigma$)
see Steggemann’s talk

An exciting program is expected to start at the LHC
$H \rightarrow b\bar{b}$ evidence
Conclusion & Perspectives

CMS searches for $H(b\bar{b})$ consistent with the SM prediction of a Yukawa structure

Evidence for $H$ decaying to fermions

$H \rightarrow \tau\tau$ and $VH(b\bar{b})$, $3.8\,\sigma$ (exp. $4.4\,\sigma$)

see Stegemann’s talk

An exciting program is expected to start at the LHC

$H \rightarrow b\bar{b}$ evidence

Precision Measurement
Additional Material
b-jet energy regression, Validation on Data

Dijet balance in Z(ℓℓ)+ b̅b data

Data/MC agreement improves after regression

An improved resolution and scale is observed

25% of improvement is found
Impact of the b-jet energy regression

The resulting improvement in the Higgs mass resolution is 20\% (Zll), 15\% (Z\nu\nu and Wl\nu), 15\% (VBF)
(dependency on the event topology)
**VZ, Strategy**

**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**)
Systematic Uncertainties

- Shape systematic
  - btag, JER, JES, trigger, generator modeling
- logNormal systematic
  - SF, signal cross section
- The systematics effect is in total of ~25% on the signal strength

<table>
<thead>
<tr>
<th>Source</th>
<th>Type</th>
<th>Yield uncertainty (%)</th>
<th>Individual contribution to μ uncertainty (%)</th>
<th>Effect of removal on μ uncertainty (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luminosity</td>
<td>norm.</td>
<td>4.4</td>
<td>12.3</td>
<td>4.4</td>
</tr>
<tr>
<td>Lepton efficiency and trigger</td>
<td>norm.</td>
<td>3</td>
<td>&lt; 2</td>
<td>&lt; 0.1</td>
</tr>
<tr>
<td>Z → νν triggers</td>
<td>shape</td>
<td>3</td>
<td>&lt; 2</td>
<td>&lt; 0.1</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>shape</td>
<td>2–3</td>
<td>8.6</td>
<td>2.0</td>
</tr>
<tr>
<td>Jet energy resolution</td>
<td>shape</td>
<td>3–6</td>
<td>6.4</td>
<td>1.1</td>
</tr>
<tr>
<td>Missing transverse energy</td>
<td>shape</td>
<td>3</td>
<td>2.3</td>
<td>0.2</td>
</tr>
<tr>
<td>b-tagging</td>
<td>shape</td>
<td>3–15</td>
<td>8.7</td>
<td>2.0</td>
</tr>
<tr>
<td>Signal cross section (scale and PDF)</td>
<td>norm.</td>
<td>5</td>
<td>10.7</td>
<td>3.9</td>
</tr>
<tr>
<td>Monte Carlo statistics</td>
<td>shape</td>
<td>1–5</td>
<td>5.0</td>
<td>0.7</td>
</tr>
<tr>
<td>Backgrounds (data estimate)</td>
<td>norm.</td>
<td>10</td>
<td>10.6</td>
<td>3.6</td>
</tr>
<tr>
<td>Single-top and Higgs (MC estimate)</td>
<td>norm.</td>
<td>15</td>
<td>4.6</td>
<td>0.6</td>
</tr>
<tr>
<td>MC modeling (V+jets and t$t$)</td>
<td>shape</td>
<td>10</td>
<td>2.7</td>
<td>1.1</td>
</tr>
</tbody>
</table>
VZ, Results

Best fit: $\mu_{WZ} = 1.37$ and $\mu_{ZZ} = 0.85$

$\sigma (pp \rightarrow ZZ) = 6.5 \pm 1.7$ (stat.) $\pm 1.0$ (syst.) $\pm 0.9$ (th.) $\pm 0.2$ (lum.) pb
$\sigma (pp \rightarrow WZ) = 30.7 \pm 9.3$ (stat.) $\pm 7.1$ (syst.) $\pm 4.1$ (th.) $\pm 1.0$ (lum.) pb

$\mu$ with respect to LO MC rescaled to NLO
**Signal Topology**

<table>
<thead>
<tr>
<th>Phys obj</th>
<th>( Z \rightarrow ll )</th>
<th>( W \rightarrow l\nu )</th>
<th>( Z \rightarrow vv )</th>
<th>( Z \rightarrow ll )</th>
<th>( W \rightarrow l\nu )</th>
<th>( Z \rightarrow vv )</th>
</tr>
</thead>
<tbody>
<tr>
<td>PF muon</td>
<td>20,</td>
<td>( \eta</td>
<td>&lt;2.4 )</td>
<td>20,</td>
<td>( \eta</td>
<td>&lt;2.4 )</td>
</tr>
<tr>
<td>MVA electron</td>
<td>20,</td>
<td>( \eta</td>
<td>&lt;2.4 )</td>
<td>30,</td>
<td>( \eta</td>
<td>&lt;2.5 )</td>
</tr>
<tr>
<td>HPS tau</td>
<td>-</td>
<td>40,</td>
<td>( \eta</td>
<td>&lt;2.1 )</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>AK5 PF jets</td>
<td>20,</td>
<td>( \eta</td>
<td>&lt;2.4 )</td>
<td>30,</td>
<td>( \eta</td>
<td>&lt;2.4 )</td>
</tr>
<tr>
<td>Type I MET</td>
<td>-</td>
<td>45</td>
<td>100</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

- (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**

**Two isolated 20 GeV e or \( \mu \) with**
\( 75<M(ee,\mu\mu)<105 \)

**One isolated 30 GeV lepton and additional missing transverse energy (MET)**

**No isolated lepton and large MET**
**VH, Background Estimate**

The contributing backgrounds are:

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

Data-driven normalization to signal region

Control regions (CR) for the main backgrounds: **V+bb**, **t̅t̅**, **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.
- Electron and muons samples are fit simultaneously to determine average SF.
**VH, Scale Factors**

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

<table>
<thead>
<tr>
<th>Process</th>
<th>$W(\ell\nu)$</th>
<th>$Z(\ell\ell)$</th>
<th>$Z(\nu\nu)$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Low</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W0b</td>
<td>$1.03 \pm 0.01 \pm 0.05$</td>
<td>$0.83 \pm 0.02 \pm 0.04$</td>
<td></td>
</tr>
<tr>
<td>W1b</td>
<td>$2.22 \pm 0.25 \pm 0.20$</td>
<td>$2.30 \pm 0.21 \pm 0.11$</td>
<td></td>
</tr>
<tr>
<td>W2b</td>
<td>$1.58 \pm 0.26 \pm 0.24$</td>
<td>$0.85 \pm 0.24 \pm 0.14$</td>
<td></td>
</tr>
<tr>
<td>Z0b</td>
<td>$1.11 \pm 0.04 \pm 0.06$</td>
<td>$1.24 \pm 0.03 \pm 0.09$</td>
<td></td>
</tr>
<tr>
<td>Z1b</td>
<td>$1.59 \pm 0.07 \pm 0.08$</td>
<td>$2.06 \pm 0.06 \pm 0.09$</td>
<td></td>
</tr>
<tr>
<td>Z2b</td>
<td>$0.98 \pm 0.10 \pm 0.08$</td>
<td>$1.25 \pm 0.05 \pm 0.11$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$1.03 \pm 0.01 \pm 0.04$</td>
<td>$1.10 \pm 0.05 \pm 0.06$</td>
<td>$1.01 \pm 0.02 \pm 0.04$</td>
</tr>
<tr>
<td><strong>Intermediate</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W0b</td>
<td>$1.02 \pm 0.01 \pm 0.07$</td>
<td>$0.93 \pm 0.02 \pm 0.04$</td>
<td></td>
</tr>
<tr>
<td>W1b</td>
<td>$2.90 \pm 0.26 \pm 0.20$</td>
<td>$2.08 \pm 0.20 \pm 0.12$</td>
<td></td>
</tr>
<tr>
<td>W2b</td>
<td>$1.30 \pm 0.23 \pm 0.14$</td>
<td>$0.75 \pm 0.26 \pm 0.11$</td>
<td></td>
</tr>
<tr>
<td>Z0b</td>
<td>$1.19 \pm 0.03 \pm 0.07$</td>
<td>$1.11 \pm 0.06 \pm 0.12$</td>
<td></td>
</tr>
<tr>
<td>Z1b</td>
<td>$2.30 \pm 0.07 \pm 0.08$</td>
<td>$1.11 \pm 0.06 \pm 0.12$</td>
<td></td>
</tr>
<tr>
<td>Z2b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$1.02 \pm 0.01 \pm 0.15$</td>
<td>$0.99 \pm 0.02 \pm 0.03$</td>
<td></td>
</tr>
<tr>
<td><strong>High</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W0b</td>
<td>$1.04 \pm 0.01 \pm 0.07$</td>
<td>$0.93 \pm 0.02 \pm 0.03$</td>
<td></td>
</tr>
<tr>
<td>W1b</td>
<td>$2.46 \pm 0.33 \pm 0.22$</td>
<td>$2.12 \pm 0.22 \pm 0.10$</td>
<td></td>
</tr>
<tr>
<td>W2b</td>
<td>$0.77 \pm 0.25 \pm 0.08$</td>
<td>$0.71 \pm 0.25 \pm 0.15$</td>
<td></td>
</tr>
<tr>
<td>Z0b</td>
<td>$1.11 \pm 0.04 \pm 0.06$</td>
<td>$1.17 \pm 0.02 \pm 0.08$</td>
<td></td>
</tr>
<tr>
<td>Z1b</td>
<td>$1.59 \pm 0.07 \pm 0.08$</td>
<td>$2.13 \pm 0.05 \pm 0.07$</td>
<td></td>
</tr>
<tr>
<td>Z2b</td>
<td>$0.98 \pm 0.10 \pm 0.08$</td>
<td>$1.12 \pm 0.04 \pm 0.10$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$1.00 \pm 0.01 \pm 0.11$</td>
<td>$1.10 \pm 0.05 \pm 0.06$</td>
<td>$0.99 \pm 0.02 \pm 0.03$</td>
</tr>
</tbody>
</table>
VH, 0/1/2 b splitting

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
Good agreement in several control regions for all modes after applying SFs.
**VH, Event Selections**

- **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(\ell\ell) optimized on \( p_T(V) \), selecting different regions and then wrt the cut on \( \Delta R(bb\bar{b}) \)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>( W(\ell\nu)Z )</th>
<th>( Z(\ell\ell)Z )</th>
<th>( Z(\nu\nu)Z )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_\ell )</td>
<td>( &gt; 30, &gt; 30 )</td>
<td>( &gt; 20, &gt; 20 )</td>
<td>( &lt; 75 &lt; m_\ell &lt; 105 )</td>
</tr>
<tr>
<td>( p_T(j_1), p_T(j_2) )</td>
<td>( &gt; 100 )</td>
<td>( &gt; 110 )</td>
<td>( &gt; 60 )</td>
</tr>
<tr>
<td>( p_T(\ell) )</td>
<td>( &gt; 20 )</td>
<td>( &gt; 150 )</td>
<td>( &gt; 170 )</td>
</tr>
<tr>
<td>( p_T(V) )</td>
<td>( &gt; 100 )</td>
<td>( &gt; 150 )</td>
<td>( &gt; 180 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>( W(\ell\nu)Z )</th>
<th>( Z(\ell\ell)Z )</th>
<th>( Z(\nu\nu)Z )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( m_\ell )</td>
<td>( &gt; 30, &gt; 30 )</td>
<td>( &gt; 20, &gt; 20 )</td>
<td>( &gt; 60 )</td>
</tr>
<tr>
<td>( p_T(j_1), p_T(j_2) )</td>
<td>( &gt; 100 )</td>
<td>( &gt; 110 )</td>
<td>( &gt; 60 )</td>
</tr>
<tr>
<td>( p_T(\ell) )</td>
<td>( &gt; 20 )</td>
<td>( &gt; 150 )</td>
<td>( &gt; 170 )</td>
</tr>
<tr>
<td>( p_T(V) )</td>
<td>( &gt; 100 )</td>
<td>( &gt; 150 )</td>
<td>( &gt; 180 )</td>
</tr>
</tbody>
</table>

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.**
VH, 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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_T(j)$</td>
<td>transverse momentum of each $Z(b\bar{b})$ daughter</td>
</tr>
<tr>
<td>$M(jj)$</td>
<td>dijet invariant mass</td>
</tr>
<tr>
<td>$p_T(jj)$</td>
<td>dijet transverse momentum</td>
</tr>
<tr>
<td>$p_T(V)$</td>
<td>vector boson transverse momentum (or $E_T^{\text{miss}}$)</td>
</tr>
<tr>
<td>CSV$_{\text{max}}$</td>
<td>value of CSV for the $Z(b\bar{b})$ daughter with largest CSV value</td>
</tr>
<tr>
<td>CSV$_{\text{min}}$</td>
<td>value of CSV for the $Z(b\bar{b})$ daughter with second largest CSV value</td>
</tr>
<tr>
<td>$\Delta \phi(V,H)$</td>
<td>azimuthal angle between V and dijet</td>
</tr>
<tr>
<td>$\Delta \eta(jj)$</td>
<td>difference in $\eta$ between $Z(b\bar{b})$ daughters</td>
</tr>
<tr>
<td>$\Delta R(jj)$</td>
<td>distance in $\eta-\phi$ between $Z(b\bar{b})$ daughters</td>
</tr>
<tr>
<td>$N_{\text{adj}}$</td>
<td>number of additional jets</td>
</tr>
<tr>
<td>$\Delta \theta_{\text{pull}}$</td>
<td>color pull angle [2]</td>
</tr>
<tr>
<td>$\Delta \phi(E_T^{\text{miss}},\text{jet})$</td>
<td>azimuthal angle between $E_T^{\text{miss}}$ and the closest jet (only for $Z(\nu\nu)$)</td>
</tr>
<tr>
<td>maxCSV$_{\text{adj}}$</td>
<td>maximum CSV of the additional jets in an event (only for $Z(\nu\nu)$ and $W(\ell\nu)$)</td>
</tr>
<tr>
<td>min$\Delta R(H,a_{\text{adj}})$</td>
<td>minimum distance between an additional jet and the $Z(b\bar{b})$ candidate (only for $Z(\nu\nu)$ and $W(\ell\nu)$)</td>
</tr>
<tr>
<td>Angular variables</td>
<td>VZ system mass, Angle $Z-Z^*$, Angle $Z-1$, Angle $Z$-jet (only for $Z(\ell\ell)$)</td>
</tr>
</tbody>
</table>
VH, Systematic Uncertainties

- Shape systematic
  - btag, JER, JES, trigger, generator modeling
- logNormal systematic
  - SF, signal cross section
- The systematics effect is in total of ~15% on the signal strength

<table>
<thead>
<tr>
<th>Source</th>
<th>Type</th>
<th>Event yield uncertainty range (%)</th>
<th>Individual contribution to $\mu$ uncertainty (%)</th>
<th>Effect of removal on $\mu$ uncertainty (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luminosity</td>
<td>norm.</td>
<td>2.2–2.6</td>
<td>&lt;2</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>Lepton efficiency and trigger (per lepton)</td>
<td>norm.</td>
<td>3</td>
<td>&lt;2</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>Z($\nu \bar{\nu}$)H triggers</td>
<td>shape</td>
<td>3</td>
<td>&lt;2</td>
<td>&lt;0.1</td>
</tr>
<tr>
<td>Jet energy scale</td>
<td>shape</td>
<td>2–3</td>
<td>5.0</td>
<td>0.5</td>
</tr>
<tr>
<td>Jet energy resolution</td>
<td>shape</td>
<td>3–6</td>
<td>5.9</td>
<td>0.7</td>
</tr>
<tr>
<td>Missing transverse energy</td>
<td>shape</td>
<td>3</td>
<td>3.2</td>
<td>0.2</td>
</tr>
<tr>
<td>b-tagging</td>
<td>shape</td>
<td>3–15</td>
<td>10.2</td>
<td>2.1</td>
</tr>
<tr>
<td>Signal cross section (scale and PDF)</td>
<td>norm.</td>
<td>4</td>
<td>3.9</td>
<td>0.3</td>
</tr>
<tr>
<td>Signal cross section ($p_T$ boost, EW/QCD)</td>
<td>norm.</td>
<td>2/5</td>
<td>3.9</td>
<td>0.3</td>
</tr>
<tr>
<td>Monte Carlo statistics</td>
<td>shape</td>
<td>1–5</td>
<td>13.3</td>
<td>3.6</td>
</tr>
<tr>
<td>Backgrounds (data estimate)</td>
<td>norm.</td>
<td>10</td>
<td>15.9</td>
<td>5.2</td>
</tr>
<tr>
<td>Single-top-quark (simulation estimate)</td>
<td>norm.</td>
<td>15</td>
<td>5.0</td>
<td>0.5</td>
</tr>
<tr>
<td>Dibosons (simulation estimate)</td>
<td>norm.</td>
<td>15</td>
<td>5.0</td>
<td>0.5</td>
</tr>
<tr>
<td>MC modeling (V+jets and tt)</td>
<td>shape</td>
<td>10</td>
<td>7.4</td>
<td>1.1</td>
</tr>
</tbody>
</table>
The most contributing backgrounds are:

- Z+jets
- top
- QCD

**MC normalization**
shape model: crystal ball

**data driven normalization**
shape model: Bernstein

The fit procedure has been validated using the Z(b\bar{b}) signal

H(b\bar{b}) signal, separately for each category


- Analytical matrix element method for S/B separation
  - integrate over unreconstructed or poorly measured particles
  - ttH vs. ttbb
- Event categories through MEM classification
  - select 4 jets most likely to come from b quark hadronization
  - 4 event categories
- The Ps/B is then computed as the ratio of ttH/ttb̅b̅ probability density
- Combined fit to Ps/B discriminant
  \( \mu = 0.7 \pm 1.4 \)
- 30% improvement on previous ttH(bb̅) CMS analysis
  - less sensitive to tt+HF modeling