V + jets modelling in CMS

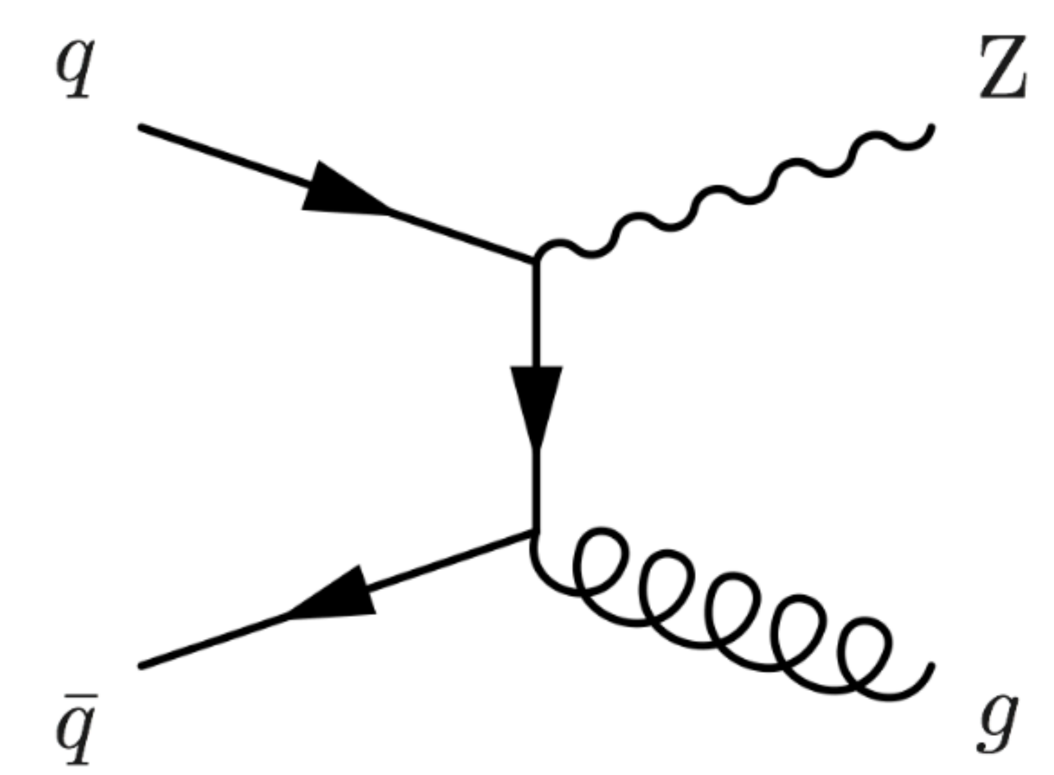
VH($c\bar{c}$) and VH($b\bar{b}$) analyses experience

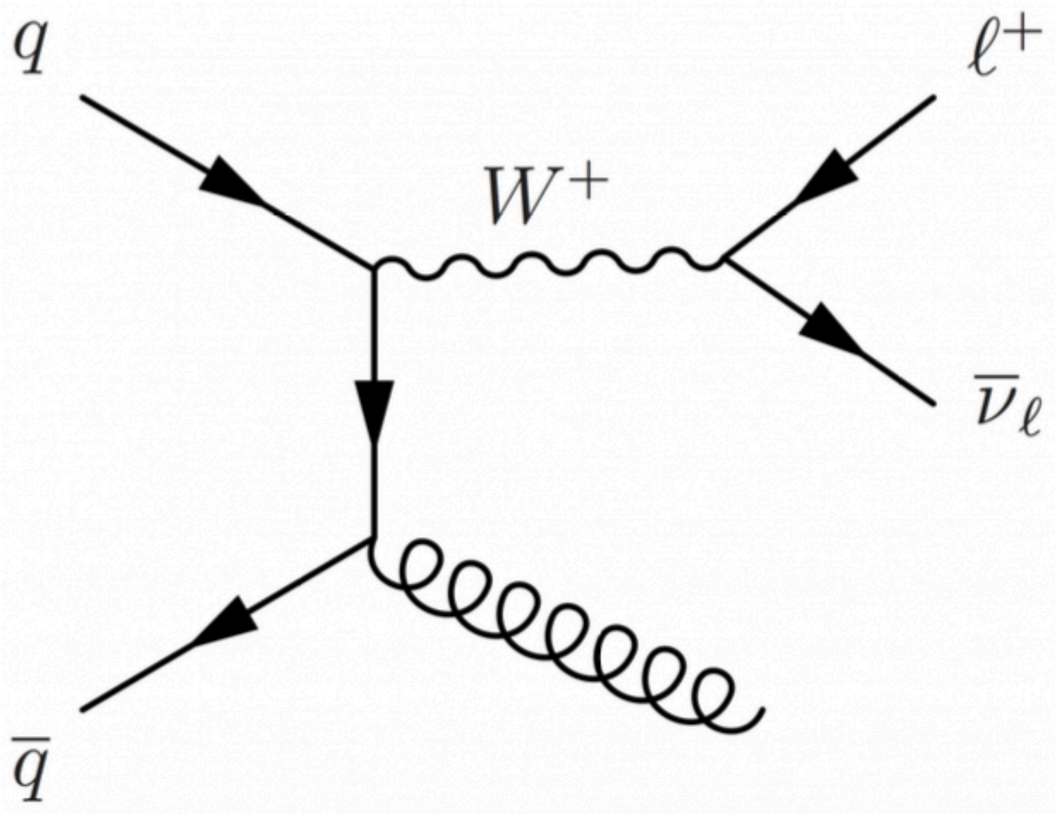
Aliya Nigamova (UHH), on behalf of the CMS collaboration

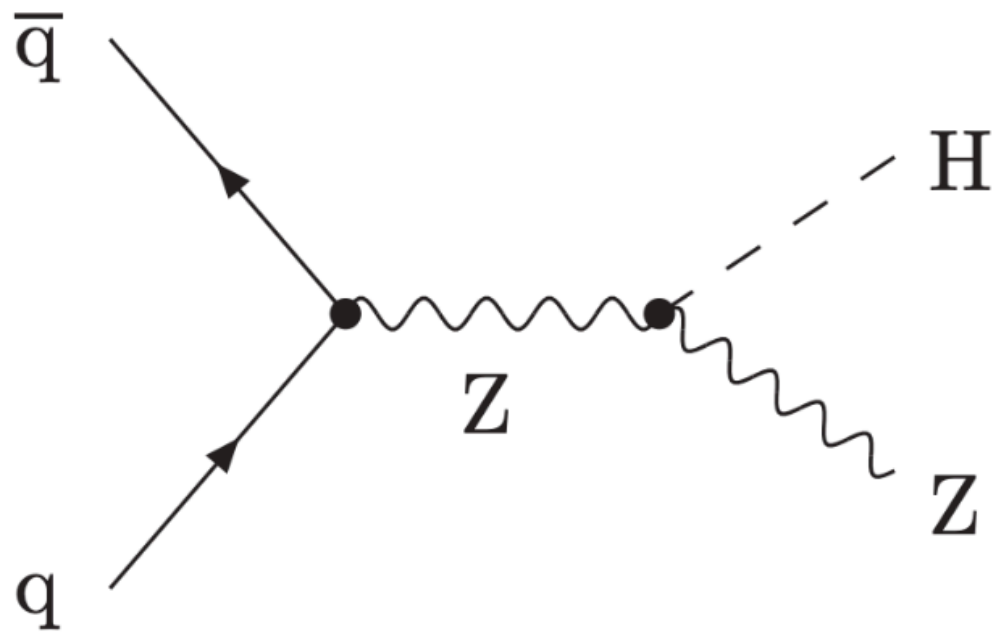
The 19th Workshop of the LHC Higgs Working Group | 28-30 November 2022

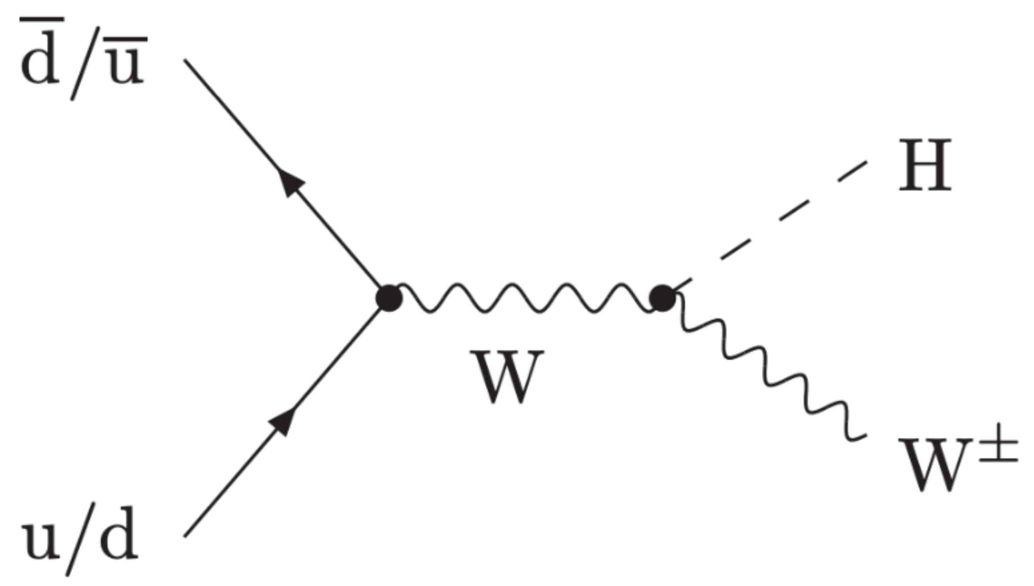
Introduction

- V+jets production is very abundant at LHC
- Similar topology to VH($H \rightarrow q\bar{q}$) processes: irreducible background
- Leading irreducible background for VH($c\bar{c}$) and VH($b\bar{b}$)
- Accurate prediction and normalisation estimation is crucial
- MVA methods for the jets tagging and signal-background separation
- Both VH($c\bar{c}$) and VH($b\bar{b}$) use the process-enriched control regions (CR) to constrain the backgrounds

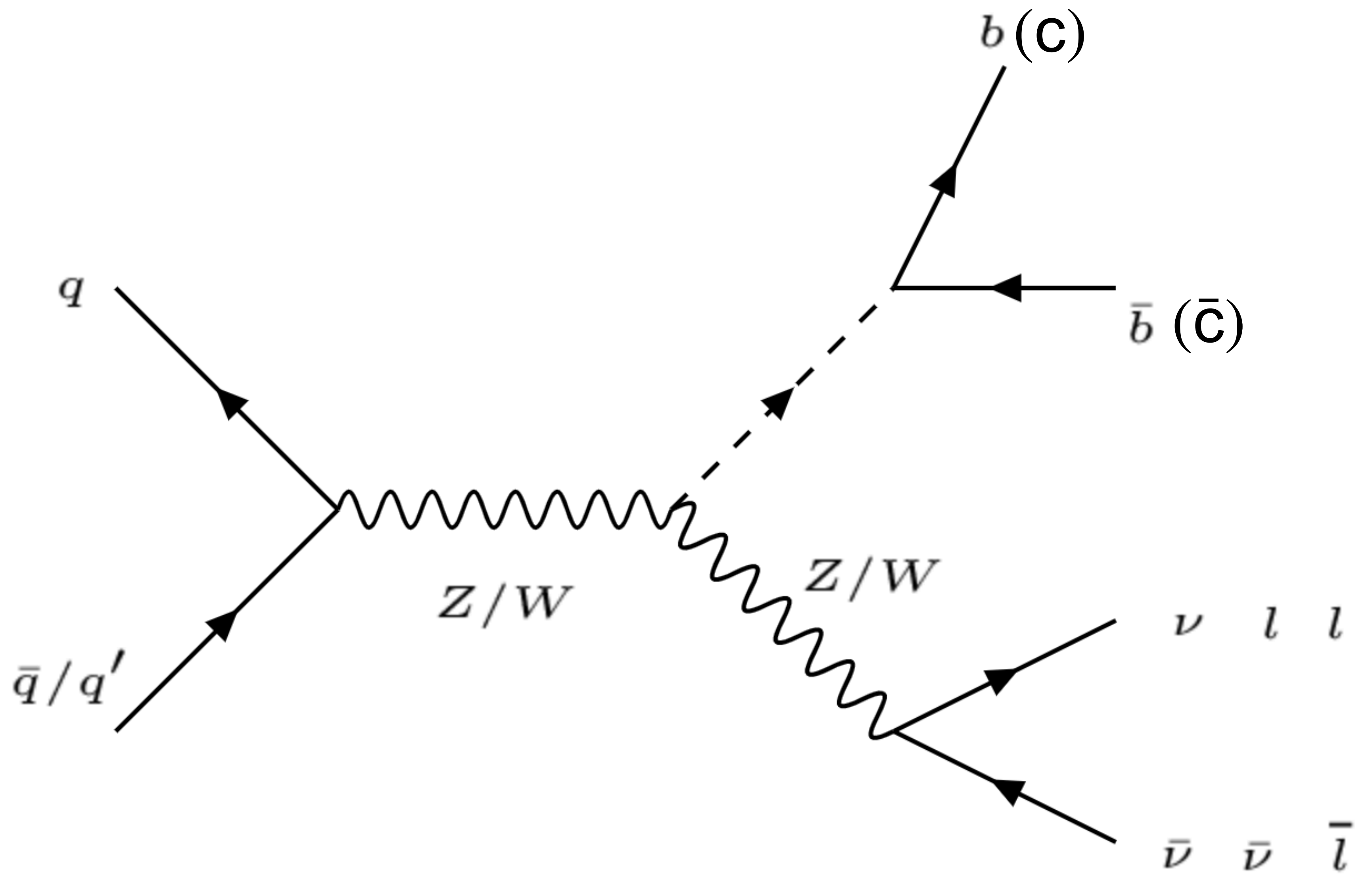








VH strategy



3 channels (V-boson decay mode): 0-lepton (Z $\rightarrow \nu\nu$), 1-lepton (W $\rightarrow$ e$\nu/\mu\nu$) and 2 lepton (Z $\rightarrow$ ee$/\mu\mu$)

VH($b\bar{b}$)

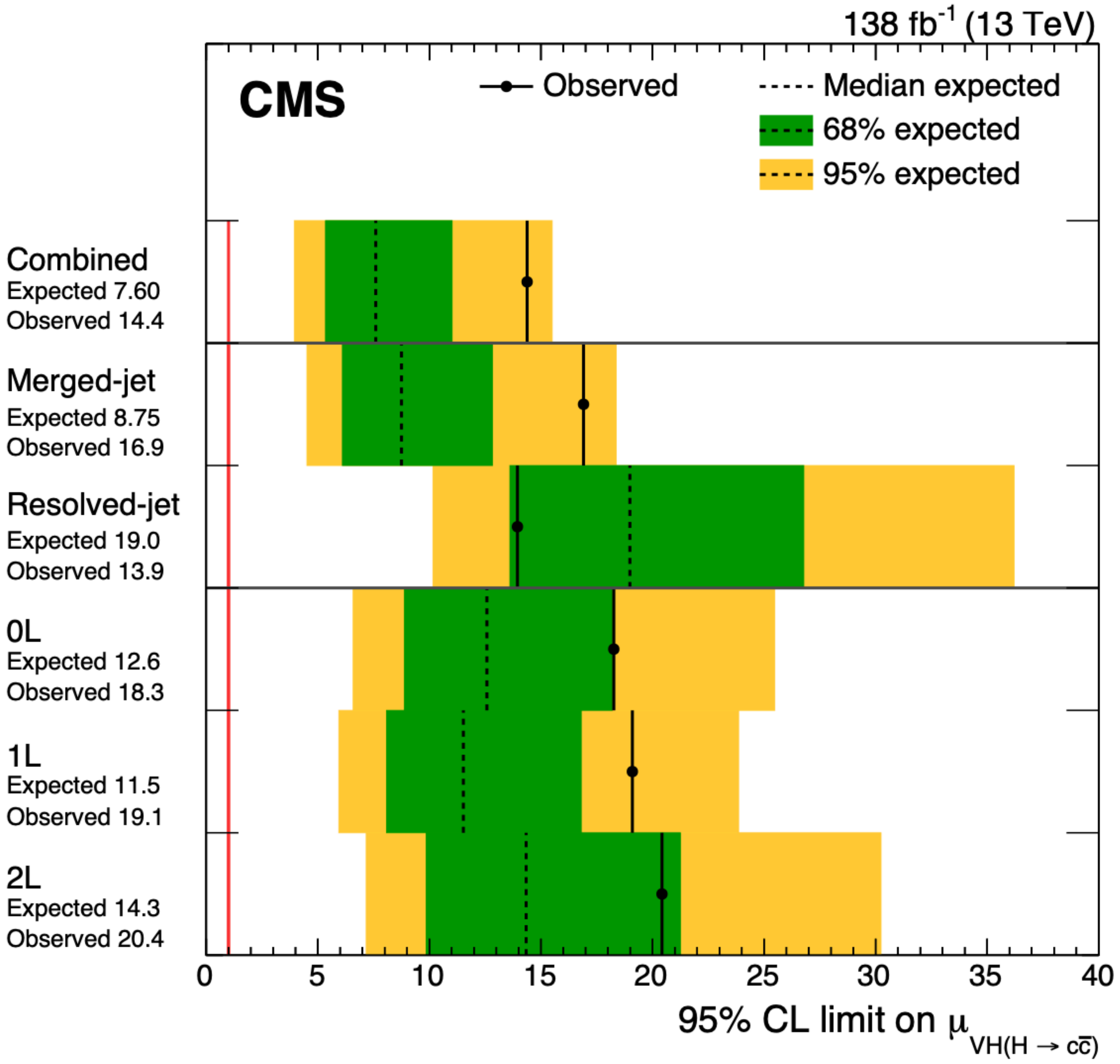
- Combine 2 Higgs decay topologies: resolved (DeepAK4) and boosted (DeepAK8)
- Simultaneous fit of boosted and resolved regions + overlap events are included with resolved priority
- Signal is extracted using DNN output (BDT in boosted analysis)
- Measuring cross-sections in STXS stage 1.2
    - SR are defined to match the STXS regions

VH($c\bar{c}$)

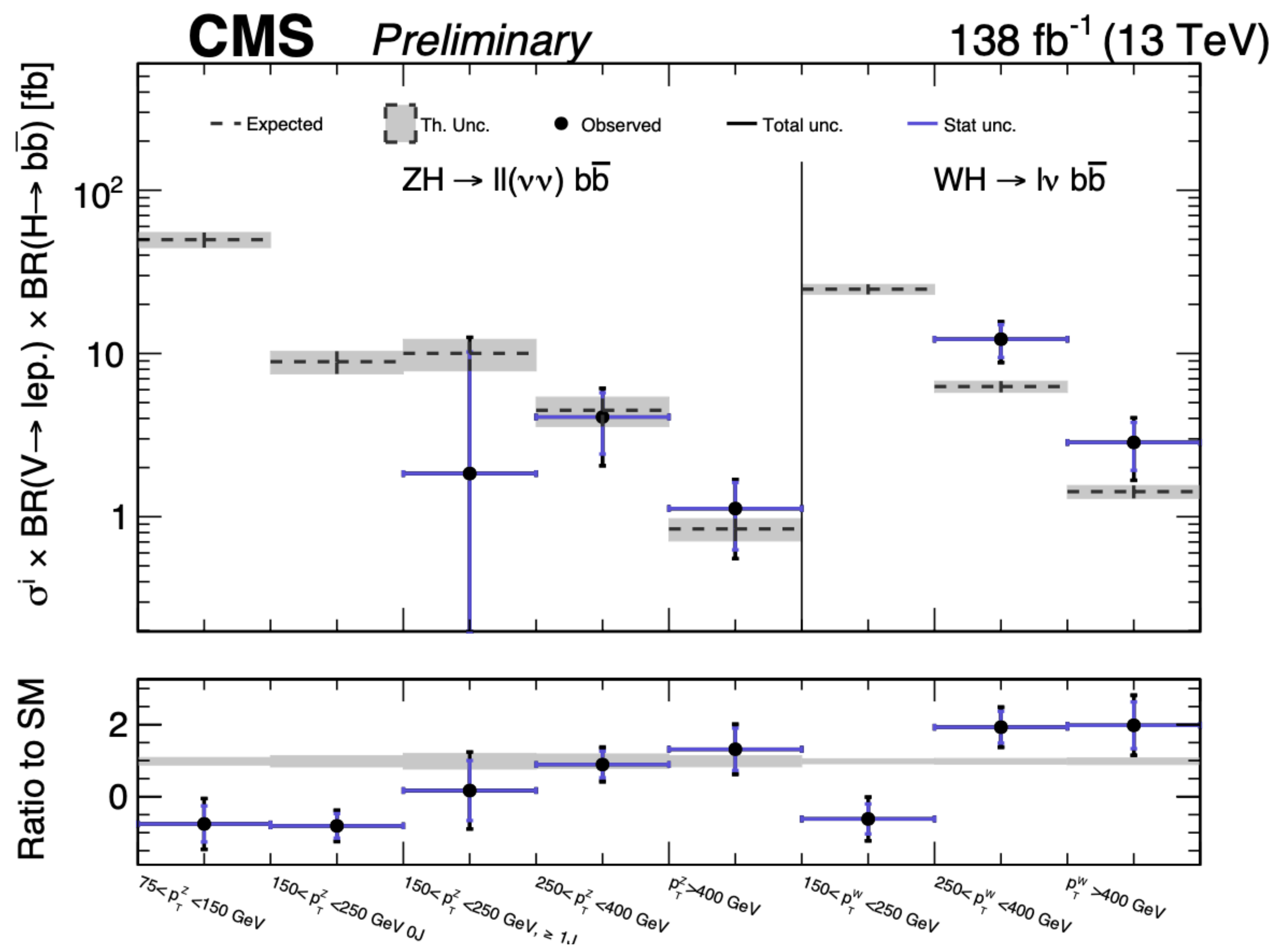
- Separate strategy for boosted and resolved analyses (orthogonality ensured with p_T^V cut)
- DeepJet in resolved and ParticleNet in boosted for c-tagging
- 2 lepton channel: split low and high p_T^V regions
- Fit BDT in resolved; m(jj) in boosted (BDT used for categorisation);

Recent Run 2 results

Leading limits for VH(c$\bar{c}$) process



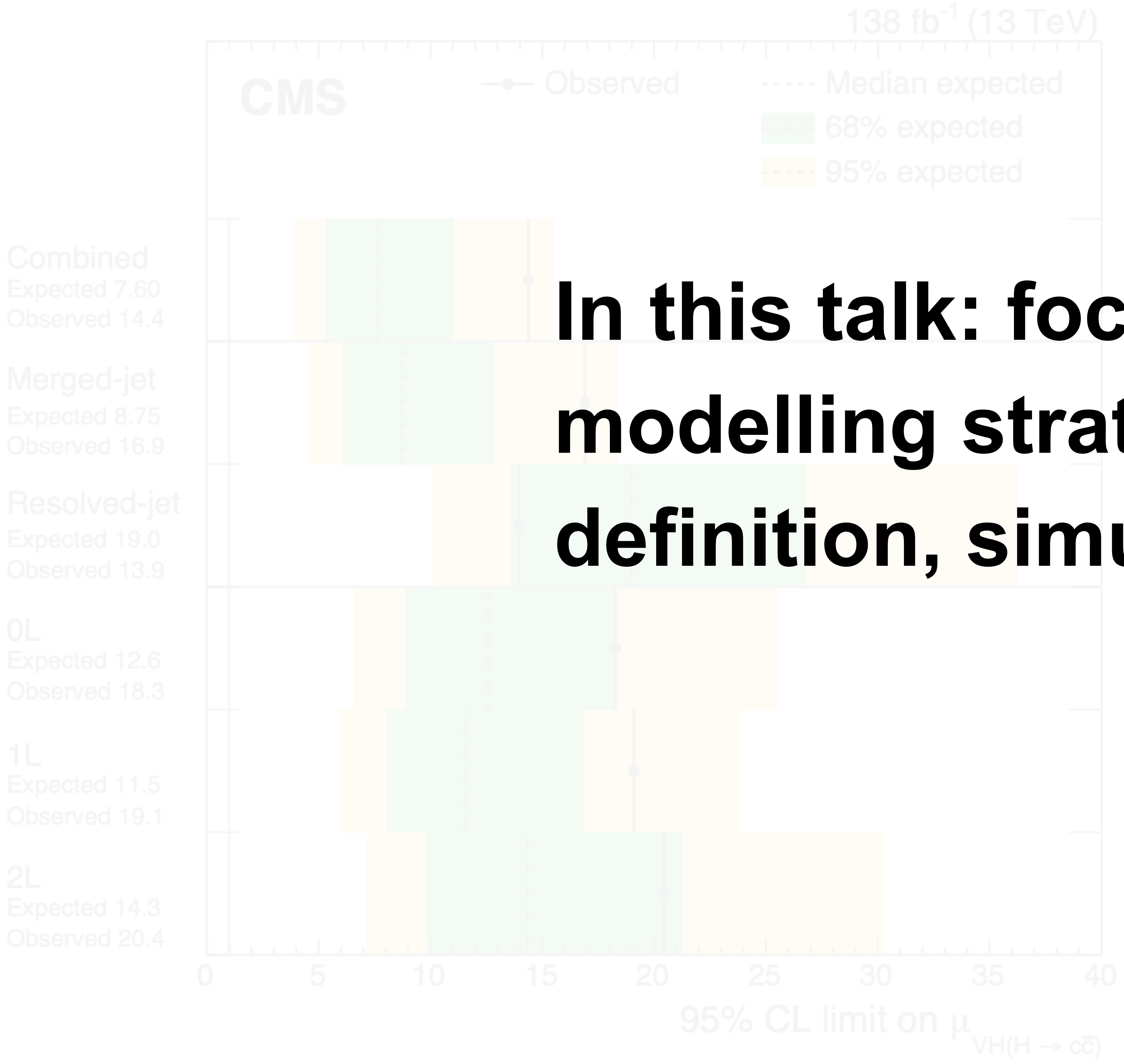
STXS measurement for VH(b$\bar{b}$)



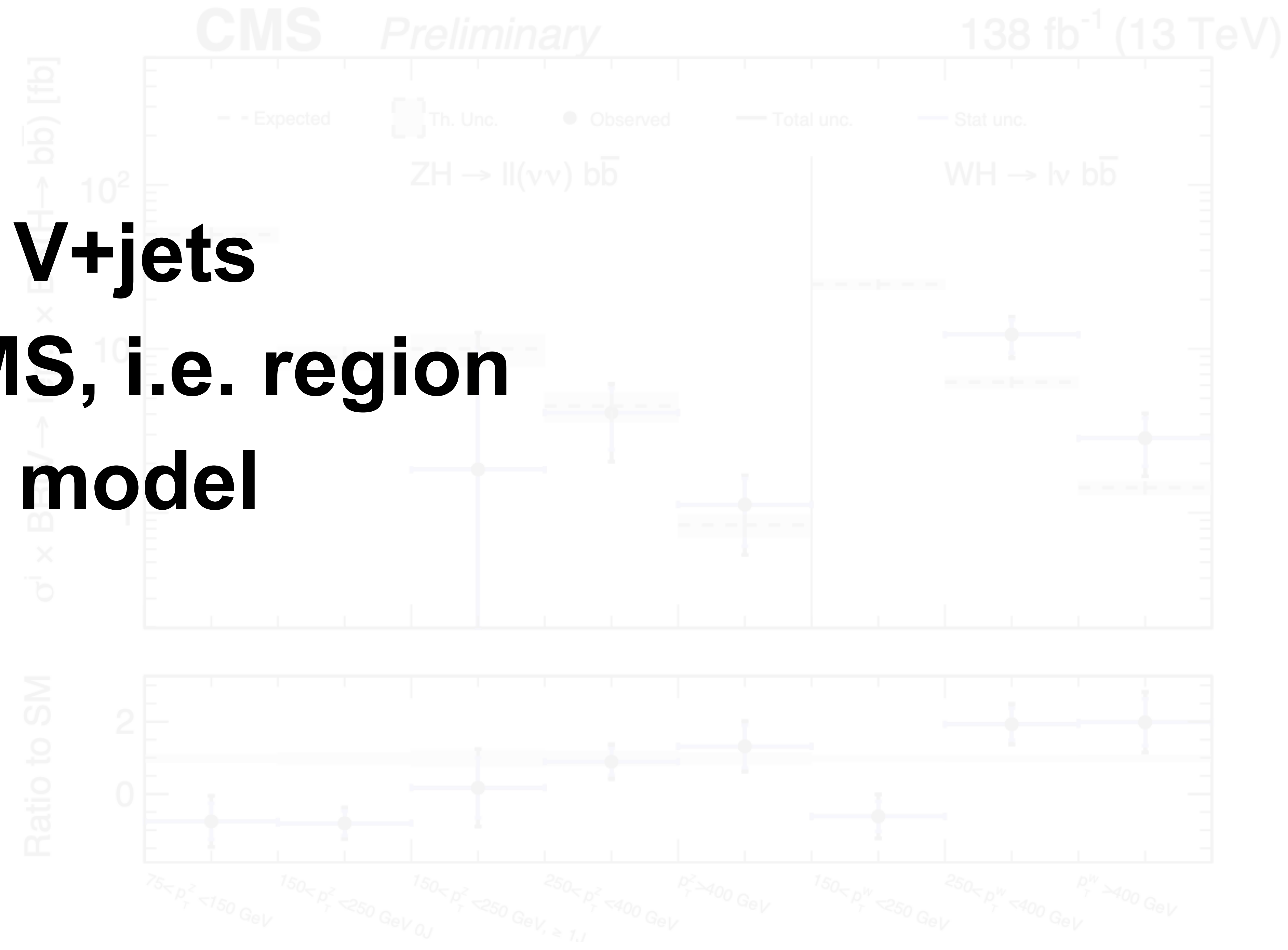
Initiated exchange between ATLAS and CMS to compare the results and strategies with the focus on background modelling: [VH(cc) meeting], [VH(bb) meeting]

Recent Run 2 results

Leading limits for VH($c\bar{c}$)



STXS measurement for VH($b\bar{b}$)



In this talk: focus on the V+jets modelling strategy in CMS, i.e. region definition, simulation, fit model

Initiated exchange between ATLAS and CMS to compare the results and strategies with the focus on background modelling: [VH(cc) meeting], [VH(bb) meeting]

Systematics

- Major contributions to the systematic uncertainties in both VH($c\bar{c}$) and VH($b\bar{b}$) analyses
  - Limited MC statistics, the uncertainties from scale variations, V+jets modelling

VH($c\bar{c}$)

Uncertainty source	$\Delta\mu / (\Delta\mu)_{tot}$
Statistical	85%
Background normalizations	37%
Experimental	48%
Sizes of the simulated samples	37%
c jet identification efficiencies	23%
Jet energy scale and resolution	15%
Simulation modeling	11%
Integrated luminosity	6%
Lepton identification efficiencies	4%
Theory	22%
Backgrounds	17%
Signal	15%

VH($b\bar{b}$)

	$\Delta\ \mu$
Background (theory)	+0.067 −0.064
Signal (theory)	+0.082 −0.060
MC stats.	+0.092 −0.093
Sim. modelling	+0.070 −0.066
b tagging	+0.059 −0.041
Jet energy resolution	+0.045 −0.057
Luminosity	+0.041 −0.034
Jet energy scale	+0.029 −0.036
LeptonID	+0.016 −0.002
Trigger(MET)	+0.001 −0.001

Simulation

VH(c$\bar{c}$) and 2017-2018 VH(b$\bar{b}$) analyses:

- NLO Madgraph aMC@NLO prediction with the inclusive reweighting to NNLO accurate in QCD
- Samples are split in p_T^V and n-jets
  - Stitched in case of phase space overlap to maximise the statistics
- $\Delta R(j_1 j_2)$ reweighting to account for residual disagreement (NLO feature)

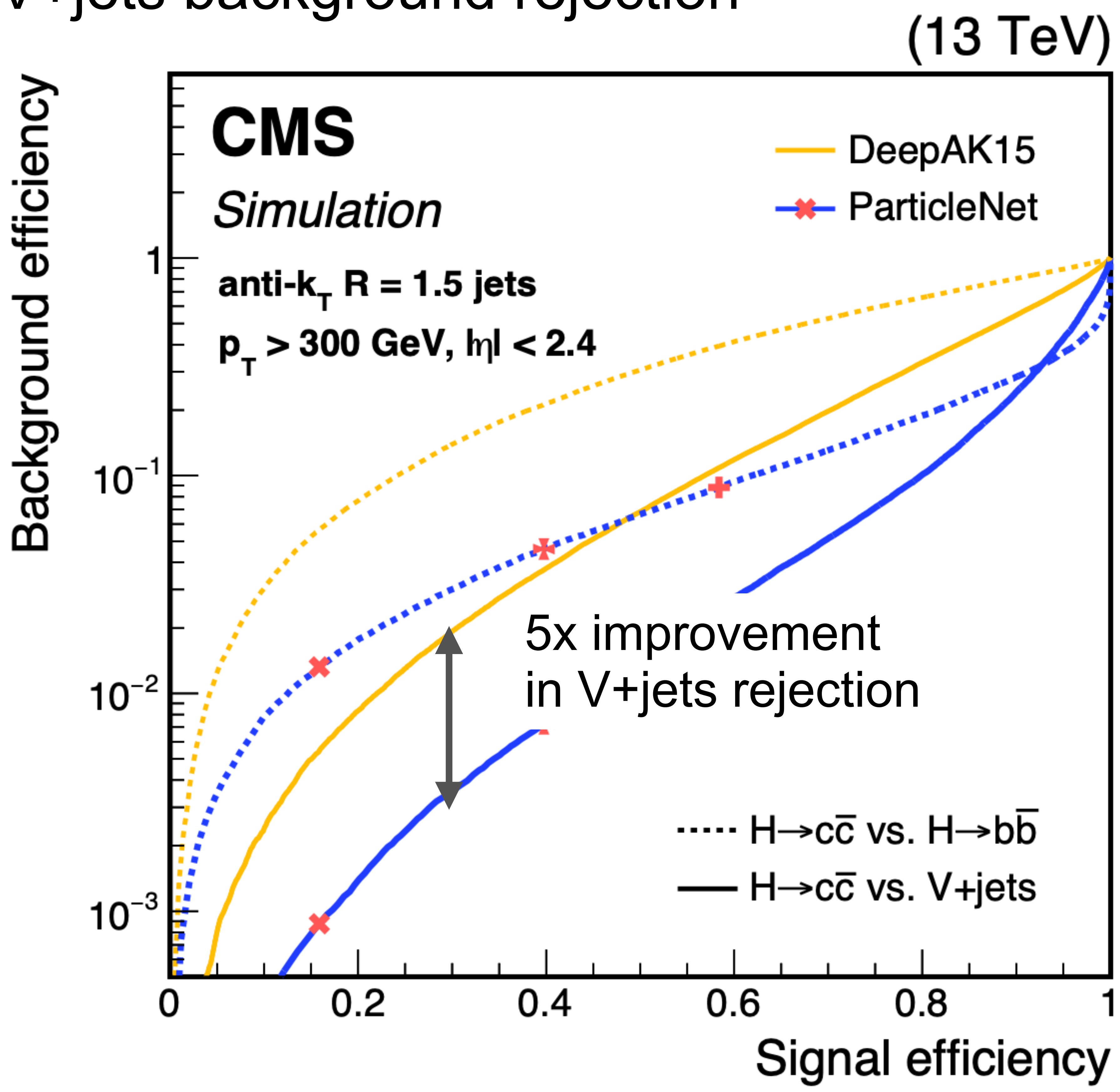
VH(b$\bar{b}$) 2016 analysis

- LO Madgraph aMC@NLO prediction with the inclusive reweighting to NNLO accurate in QCD
- HT-binned + b-enriched enriched binned in p_T^V
  - Stitched in case of phase space overlap to maximise the statistics
- $\Delta\eta(j_1 j_2)$ to reweighting to NLO accuracy

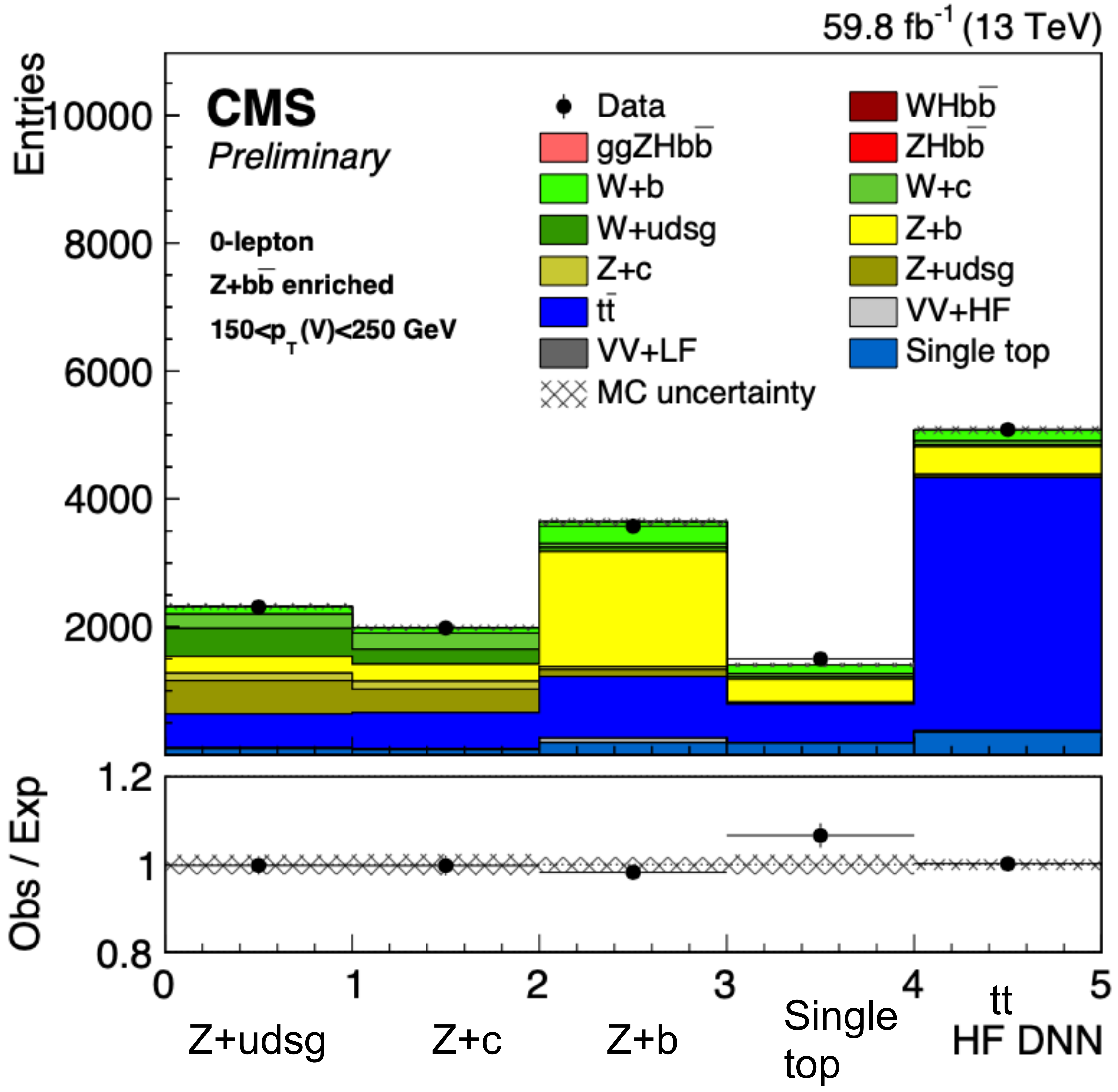
Additional systematic uncertainties assigned

Improving background rejection and modelling with NN

VH($c\bar{c}$): Major improvement from ParticleNet in in V+jets background rejection



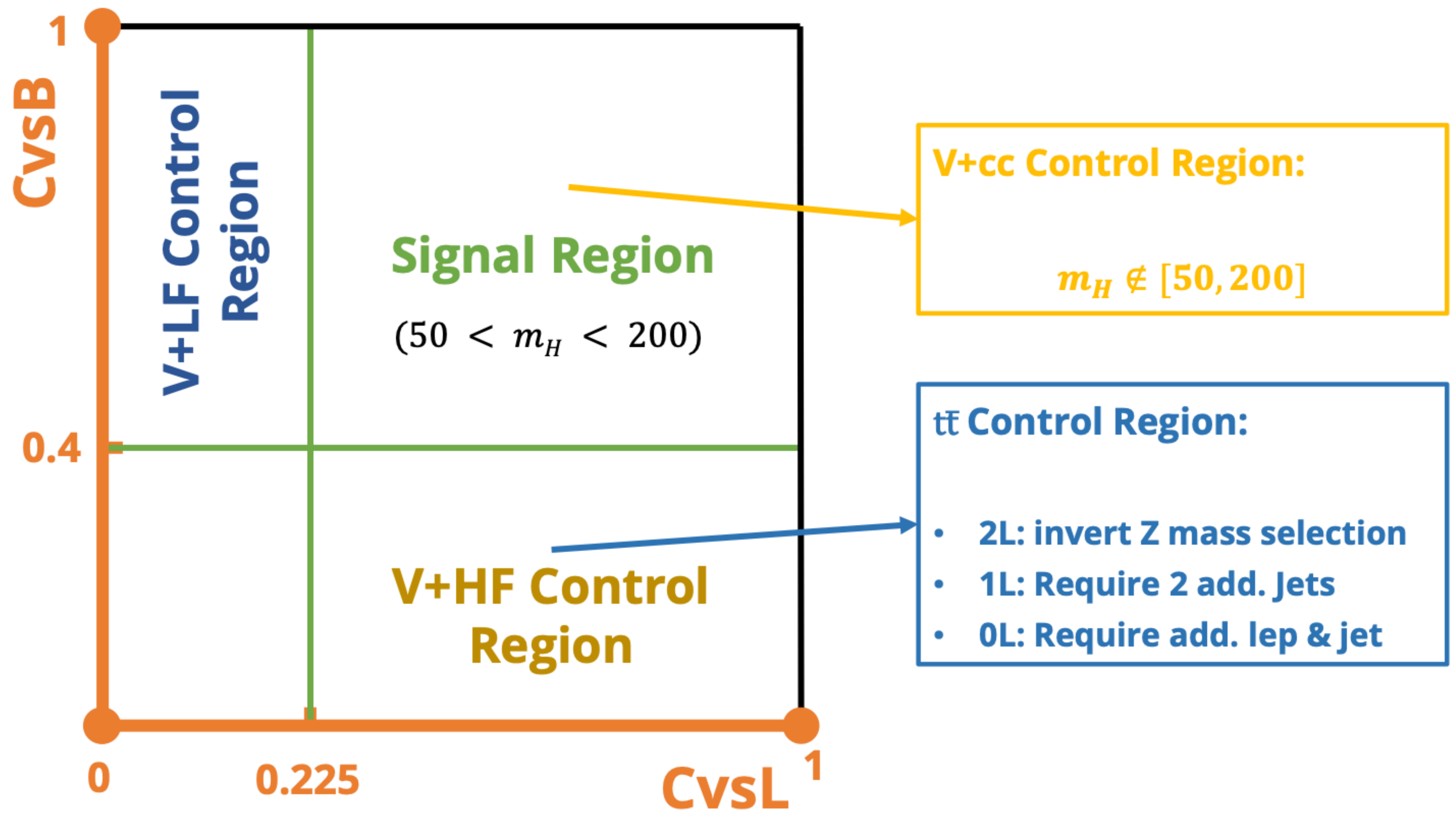
VH($b\bar{b}$): In V+HF control region the multi-class DNN was trained to improve the V+jets separation from the other major backgrounds



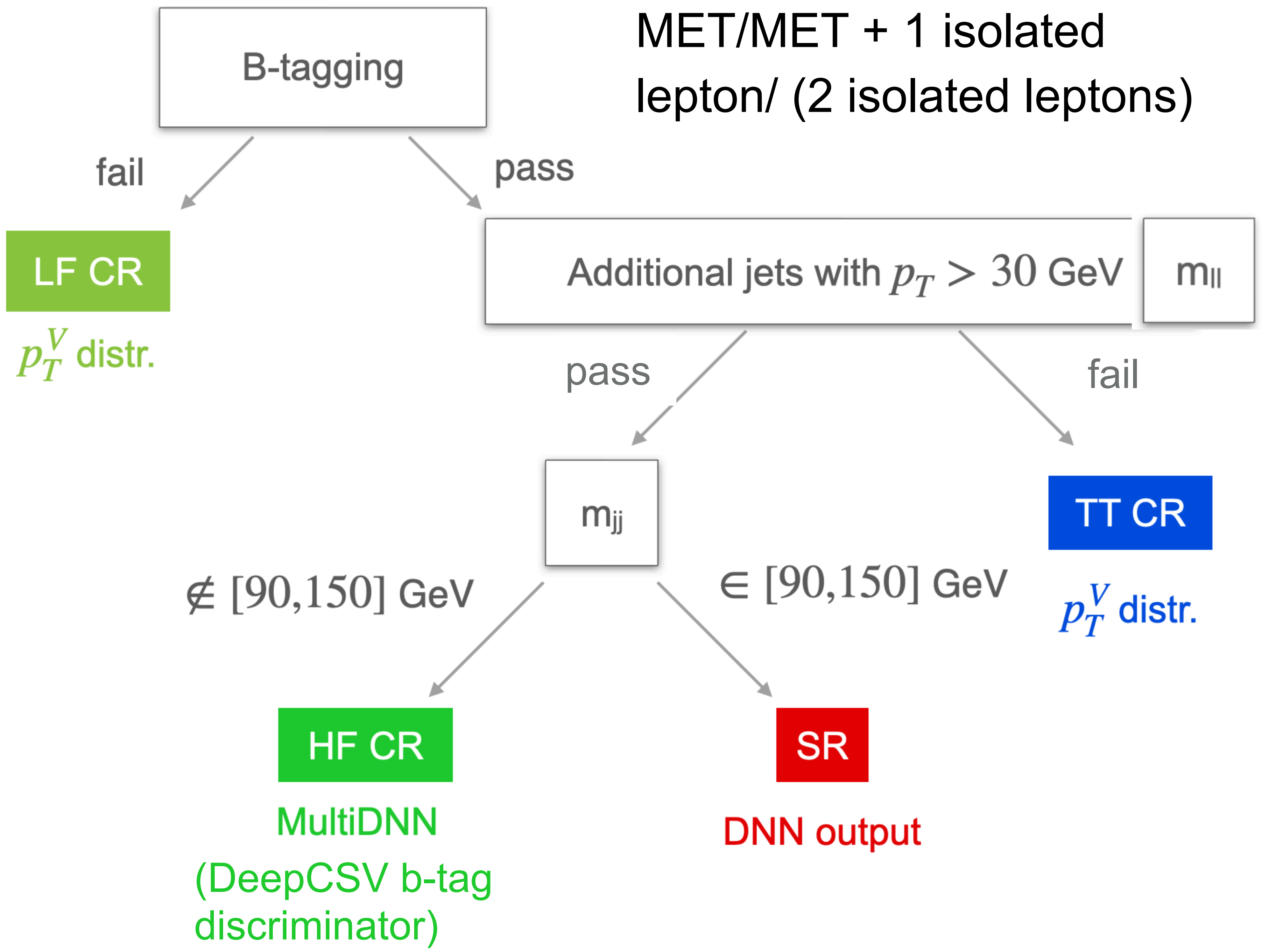
Regions definition

Resolved

VH($c\bar{c}$) regions

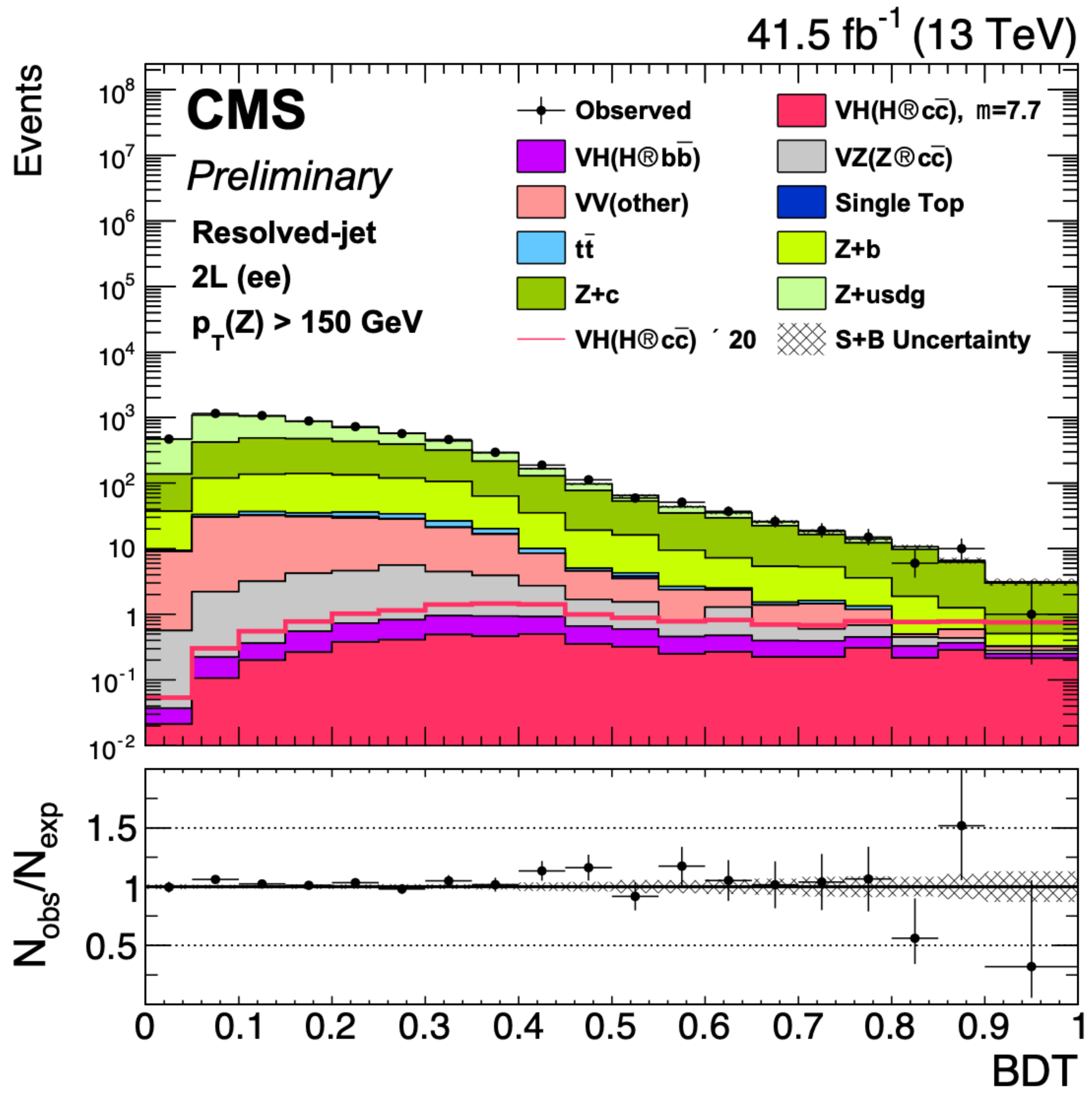


VH($b\bar{b}$) regions

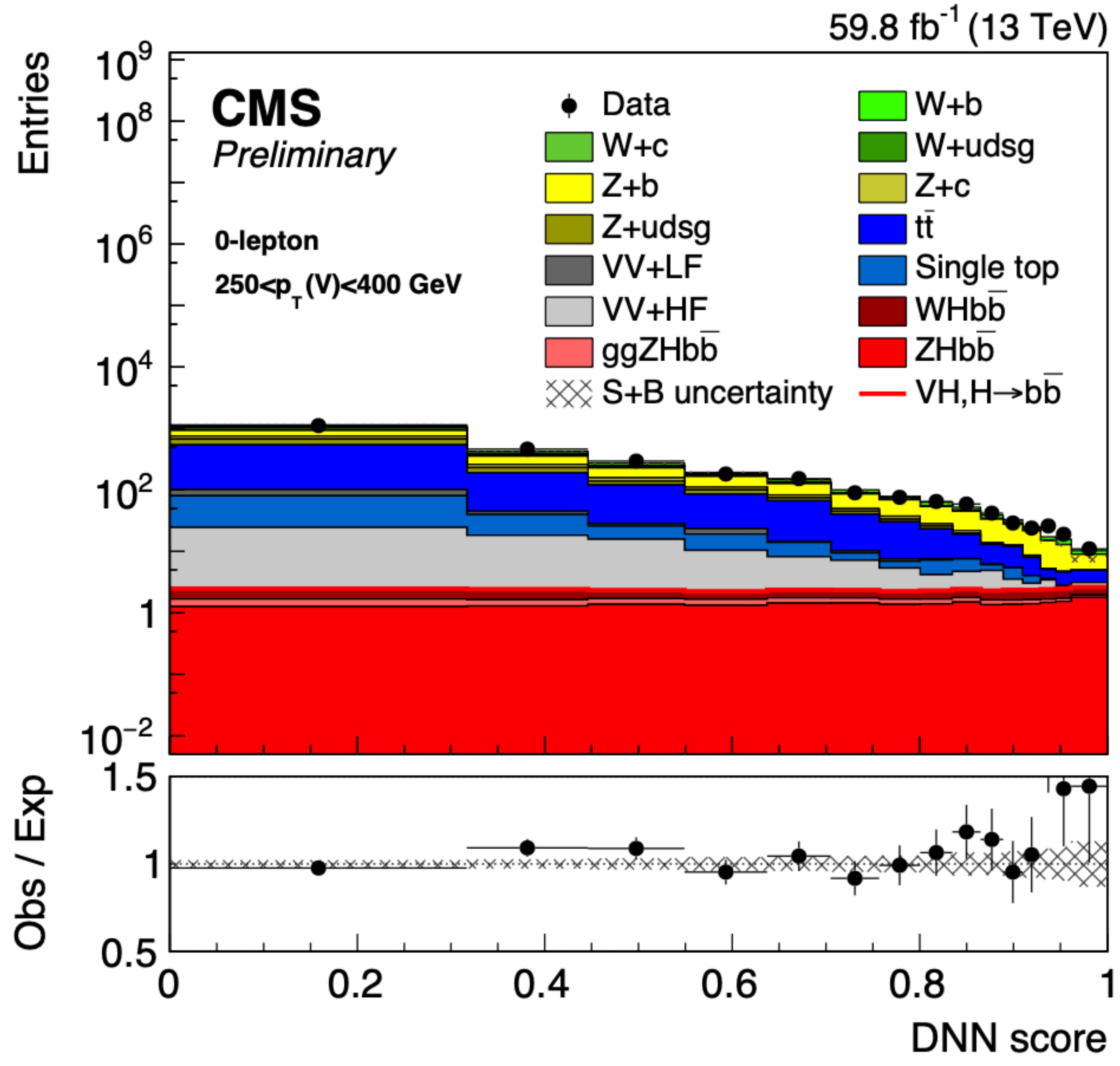


Signal region distributions

VH(cc̄), 2 lepton resolved region



VH(bb̄) regions, 0 lepton resolved region



Fit model VH($c\bar{c}$)

Resolved

V+jets process is subcategorised:
- V+b - at least one b-hadron (HF CR)
- V+c - at least one c-hadron, no b-hadrons (cc CR)
- V+udsg - no b- or c-hadrons (LF CR)

Channel		SR	LF CR	HF CR	cc CR	$t\bar{t}$ CR
0L		✓	✓	✓	✓	✓
1L	e	✓	✓		✓	✓
	μ	✓	✓		✓	✓
2LL	ee	✓	✓	✓	✓	✓
	μμ	✓	✓	✓	✓	✓
2LH	ee	✓	✓	✓	✓	✓
	μμ	✓	✓	✓	✓	✓

Constrained using 3 CR with the floating rate parameters:

<table>
<tr><th>Channel</th><th>Z+LF</th><th>Z+HF</th><th>Z+cc</th><th>W+LF</th><th>W+HF</th><th>W+cc</th><th>$t\bar{t}$</th></tr>
<tr><td>0L</td><td>✓</td><td>✓</td><td>✓</td><td rowspan="2">✓</td><td rowspan="2" colspan="2">✓</td><td>✓</td></tr>
<tr><td>1L</td><td></td><td></td><td></td><td>✓</td></tr>
<tr><td>2LL</td><td>✓</td><td>✓</td><td>✓</td><td></td><td></td><td></td><td>✓</td></tr>
<tr><td>2LH</td><td>✓</td><td>✓</td><td>✓</td><td></td><td></td><td></td><td>✓</td></tr>
</table>

[Details]

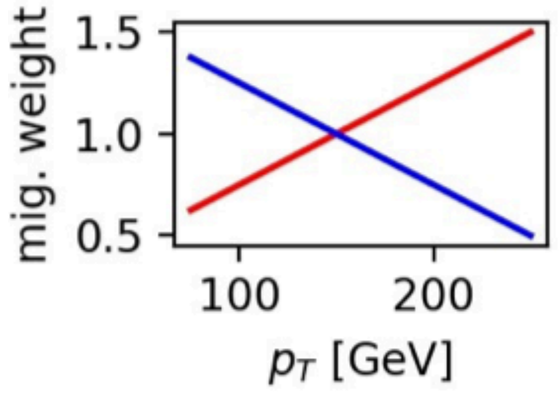
Fit model VH($b\bar{b}$)

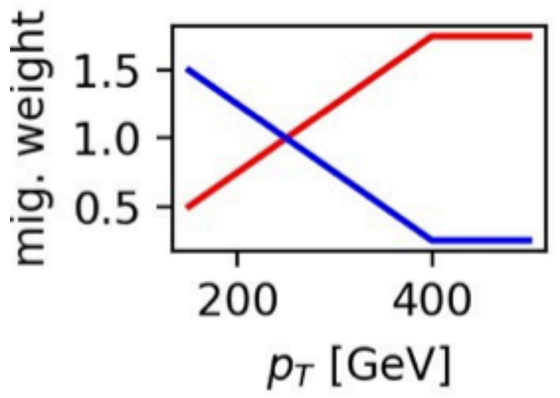
V+jets process is subcategorised:
- V+b - at least one b-hadron
- V+c - at least one c-hadron, no b-hadrons
- V+udsg - no b- or c-hadrons

STXS measurement (binned in p_T^V and n-jets)

➡Control the background in p_T^V bins

- Inclusive in p_T^V free-floating RP defined for each channel
- p_T^V migration shape uncertainty:





- Large priors, constrained in the fit
- Ensures the smoothness of post-fit p_T^V distribution

Z+b/c/udsg

	Channels					
pT(V)	Zee	Zmm	Wen	Wmn	Znn	pT(V)
75.0						75.0 low
150.0						150.0 med
250.0						250.0 high1
400.0						400.0 high2

[Details]

Resolved

Fit model VH($b\bar{b}$)

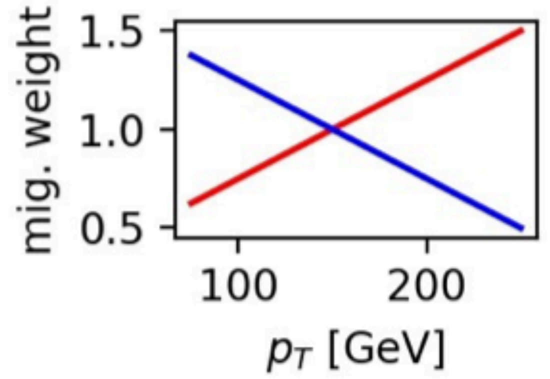
V+jets process is subcategorised:

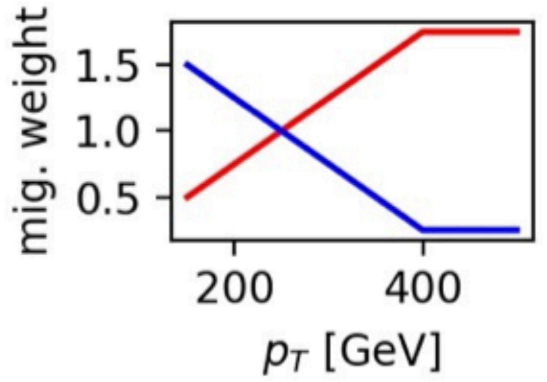
V+b - at least one b-hadron
V+c - at least one c-hadron, no b-hadrons
V+udsg - no b- or c-hadrons

STXS measurement (binned in p_T^V and n-jets)

➡Control the background in p_T^V bins

- Inclusive in p_T^V free-floating rate parameters defined for each channel
- p_T^V migration shape uncertainty:





- Large priors, constrained in the fit
- Ensures the smoothness of post-fit p_T^V distribution

W+b/c/udsg

	Channels					
p_T(V)	Zee	Zmm	Wen	Wmn	Znn	p_T(V)
75.0						75.0 low
150.0						150.0 med
250.0						250.0 high1
400.0						400.0 high2

[Details]

Resolved

Summary

- Summarised the V+jets background modelling strategy in VH($c\bar{c}$) and VH($b\bar{b}$) analyses in CMS
- NLO prediction is used whenever possible
- Control region method with free floating rate parameters extract the normalisation
- Relying on the flexible fit model to improve data-MC agreement
  - Majority of the extracted rate parameters are close to unity
- Corrections applied ($\Delta R(j_1 j_2)$ for NLO and $\Delta\eta(j_1 j_2)$ for LO) to improve data-MC agreement, with additional systematic uncertainties
- Exchange with ATLAS has been initiated, very useful and constructive discussion in the VH WG1 subgroup

Backup

Boosted VHcc

Merged

All regions

Channel		SR		V+Jets CR	$t\bar{t}$ CR
0L		✓		✓	✓
1L	e	Low BDT ✓	High BDT ✓	✓	✓
	μ	Low BDT ✓	High BDT ✓	✓	✓
2L	ee	✓		✓	
	μμ	✓		✓	

Each further split by purity: LP, MP, HP

Normalization factors

Channel	W+j	Z+j	$t\bar{t}$ (top)	$t\bar{t}$ (W)	$t\bar{t}$ (other) nJet<2	$t\bar{t}$ (other) nJet≥2
0L	✓	✓	✓	✓	✓	✓
1L		↕				
2L		✓		VW		

- $t\bar{t}$ further split by which component lies inside the AK15 jet
- Same RP used for SingleTop (and VW in 2L)
- Each RP further split by purity: LP, MP, HP

Boosted VHbb

0-lepton				
Variable	SR	Z + b jets	Z + light jets	$t\bar{t}$
DeepAK8 (bbVsLight)	> 0.8	> 0.8	< 0.8	> 0.8
$M(jj)$	$\in[90,150]$	$\notin[90,150]$	> 50	> 50
N_{al}	$= 0$	$= 0$	$= 0$	> 0
N_{aj}	$= 0$	$= 0$	$= 0$	> 1

1-lepton				
Variable	SR	W + b jets	W + light jets	$t\bar{t}$
DeepAK8 (bbVsLight)	> 0.8	> 0.8	< 0.8	> 0.8
$M(jj)$	$\in[90,150]$	$\notin[90,150]$	> 50	> 50
N_{al}	$= 0$	$= 0$	$= 0$	> 0
N_{aj}	$= 0$	$= 0$	$= 0$	> 1

2-lepton				
Variable	SR	Z + b jets	Z + light jets	$t\bar{t}$
DeepAK8 (bbVsLight)	> 0.8	> 0.8	< 0.8	> 0.8
$M(jj)$	$\in[90,150]$	$\notin[90,150]$	> 50	> 50
$M(V)$	$\in[75,105]$	$\in[75,105]$	$\in[75,105]$	$\notin[90,150]$