VHbb STXS analysis with Run 2 in CMS

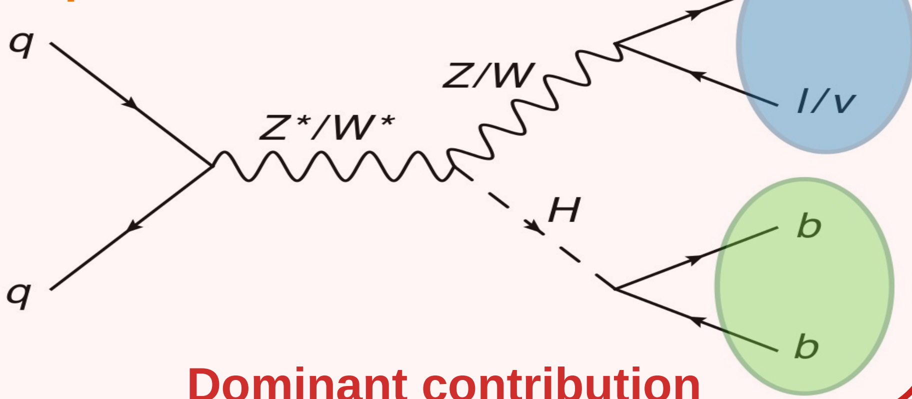
Krunal Gedia,
on behalf of the CMS Collaboration

ATLAS-CMS cross talk
22nd November 2022

- **Introduction (Signal, background, goal, samples)**
- Object reconstruction
- Event selection
- Fit model
- Cross check analysis results (VZbb & dijet mass analysis)
- VHbb unblinded results

Higgs produced with an associated vector boson

quark-induced ZH$b\bar{b}$ & WH$b\bar{b}$



Dominant contribution

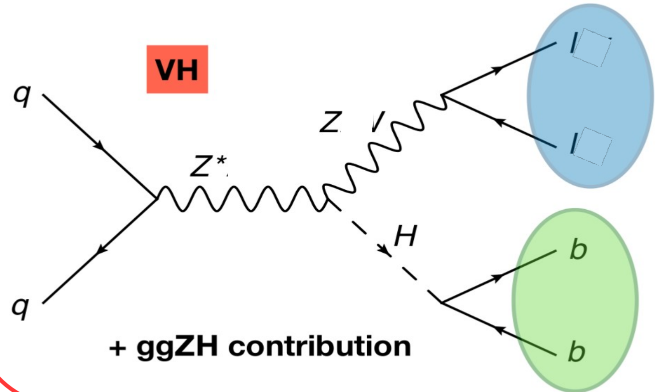
Why VHbb to study H → bb coupling ?

- boost of the V-boson → QCD/V+Jets background
- Leptonic V decay → Trigger
- Large MET → Trigger

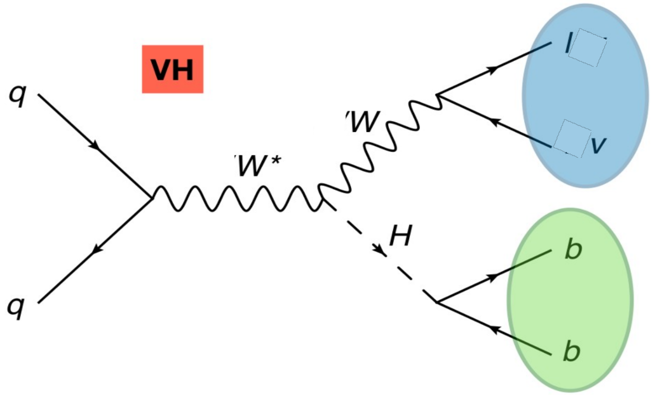
- VHbb → Dominant decay mode of Higgs decay to bottom quark. (60% BR fraction)

+ggZH contribution

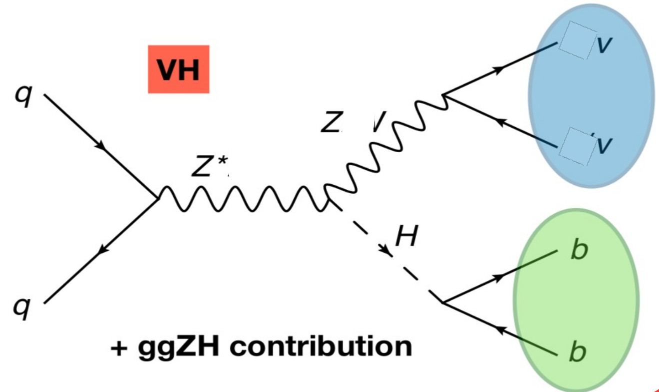
2-lepton channel



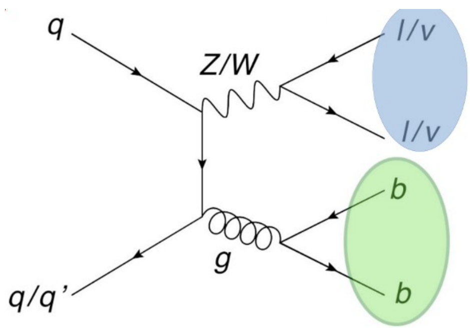
1-lepton channel



0-lepton channel

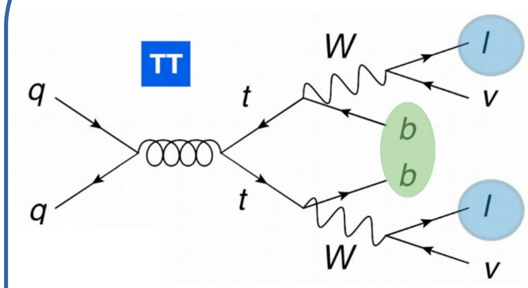


Major backgrounds



falling mass distribution without peak, little softer pT

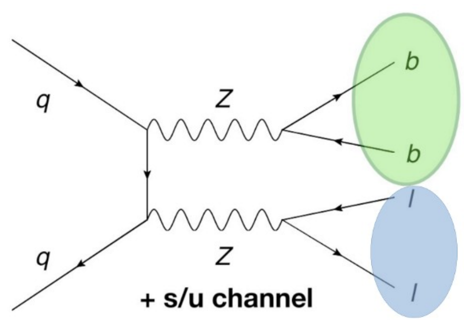
V+jets



invariant mass has flat distribution

If W decays hadronically, more additional jets

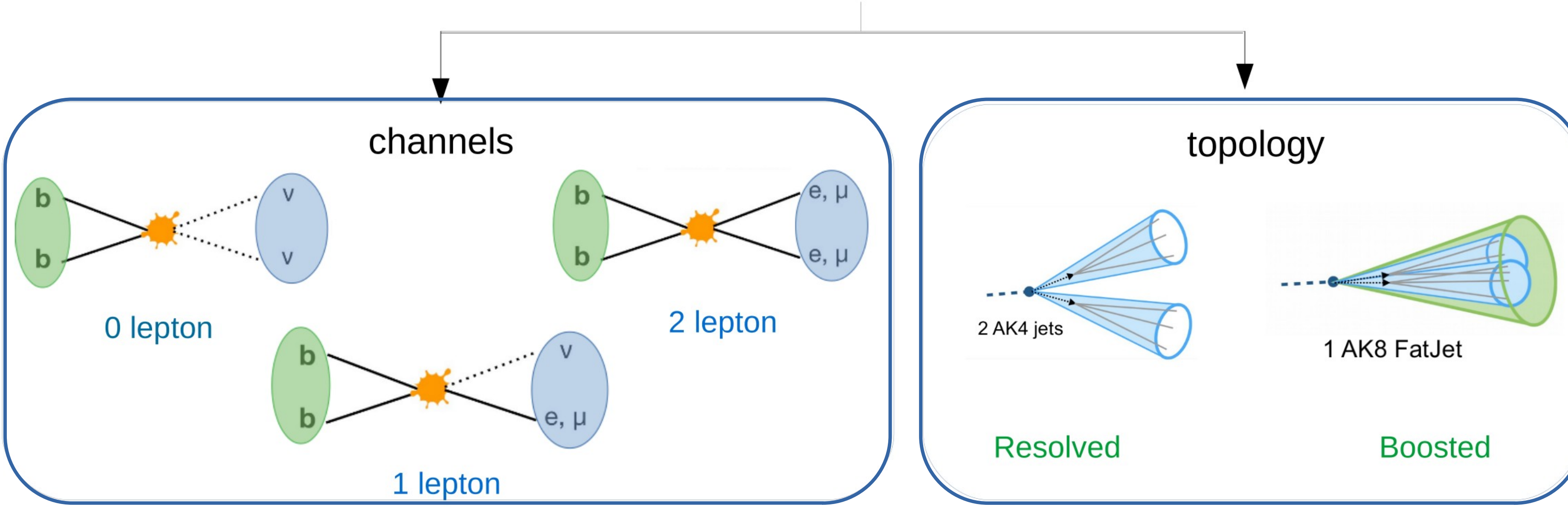
TT + ST

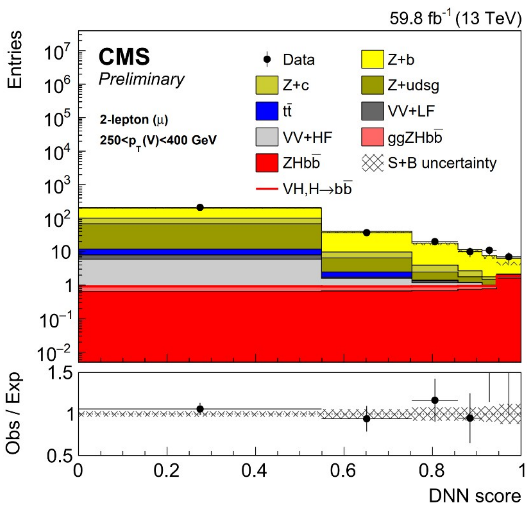


invariant mass has a peak at mV

Diboson

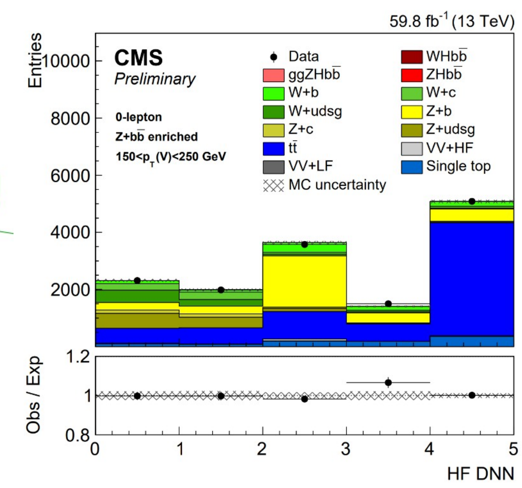
Analysis strategy





Extract signal strength using MVA in SR!

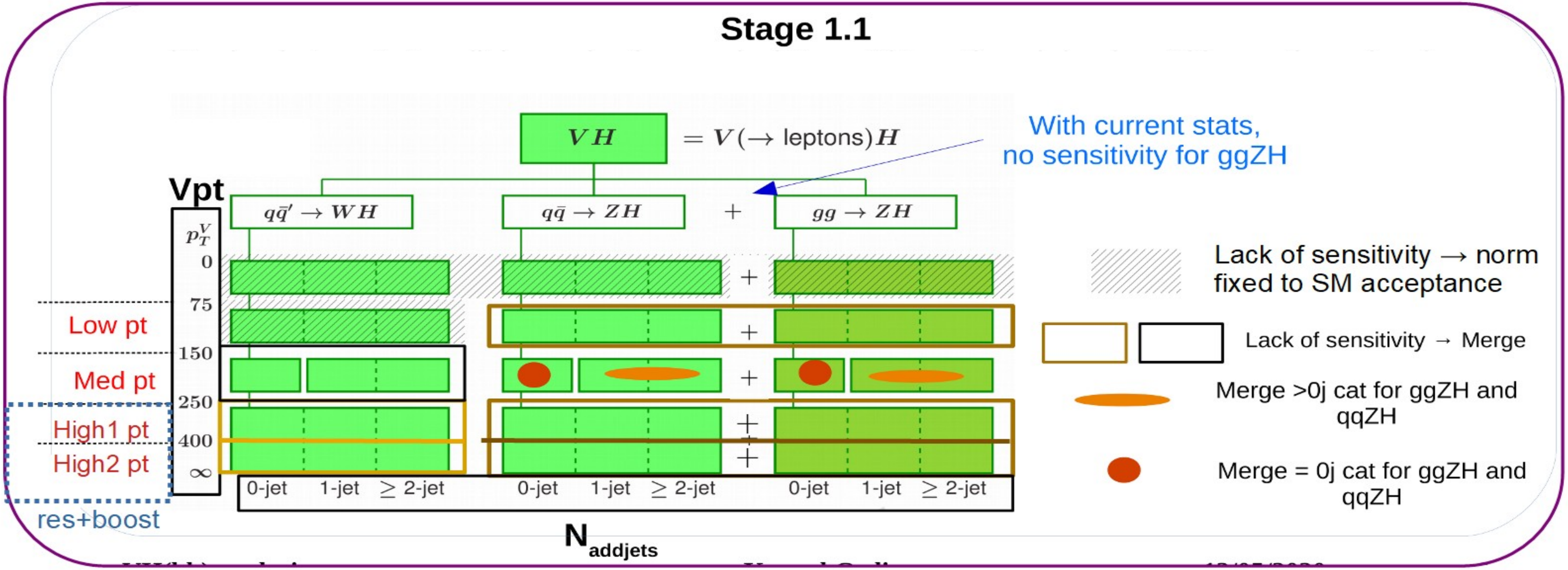
Background normalization from orthogonal CR!
TT, V+LF, V+HF CRs



Goal: STXS measurements with Run 2 data

Goal: STXS measurements

- Measure cross-section in mutually exclusive phase space.
- Reduction of theoretical uncertainties.
- Possibility to combine ATLAS and CMS measurements easily



STXS POI in fit model

					Topology
Process		qqZH+ggZH		qqWH	
p_T(V)	75-150	1			Resolved
	150-250	2	3	6	Resolved
	250-400	4		7	Resolved+Boosted
	> 400	5		8	Resolved+Boosted
nAdd(jets)		=0	>=1	>=0	

Signal MC samples

Sample	Generator	Order	Comments
qqZH, qqWH	POWHEG v2 + MiNLO procedure	NLO	XSec reweighted NLO EWK in $p_T(V)$
ggZH	POWHEG v2	LO	-

Background MC samples

Sample	Year	Generator	Order	Comments
Z+Jets W+Jets	2016	LO MadGraph With MLM matching	LO	LO reweighted to NLO in etabb + XSec reweighted to NNLO QCD + NLO EWK in $p_T(V)$
	2017, 2018	aMC@NLO with FXFX merging	NLO	XSec reweighted to NNLO QCD + NLO EWK in $p_T(V)$
ZZ, WW, WZ	2016, 2017, 2018	aMC@NLO with FXFX merging	NLO	XSec from CMS measurement
tt, Single top	2016, 2017, 2018	POWHEG	NLO	XSec reweighted to NNLO + NNLL

Types of V+jets samples and stitching

2016

- HT binned and b-enriched LO samples available.
- Stitching of samples to increase MC stats especially in b-enriched phase space.
- LO samples reweighted to NLO samples in deta(bb) and in bins of number of b-jets. Nuisances corresponding to reweighting procedure included in the fit.

2017 & 2018

Several sets of V+jets NLO samples available for the analysis depending on channel and year.

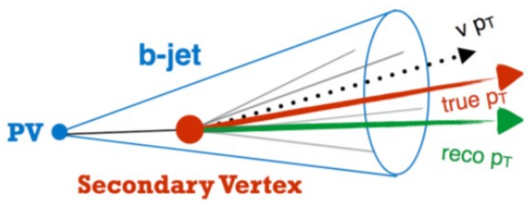
- Inclusive
- Split in jet multiplicity,
- Split in vector boson transverse momentum,
- Split in vector boson transverse momentum and jet multiplicity

Performed stitching of all available the samples to improve effective statistics for the analysis (details on stitching procedure in back-up)

- Introduction (Signal, background, goal, samples)
- **Object reconstruction**
- Event selection
- Fit model
- Cross check analysis results (VZbb & dijet mass analysis)
- VHbb unblinded results

Vector boson: Isolated leptons &/or MET

Electrons: MVA-based electron ID & Iso. cut
Muon: Cut-based ID and Isolation cut
MET: PF + CHS* MET



Higgs boson: Leading & sub-leading b-tagged jets

Jets:
- AK4/AK8 CHS* jets, JEC/JER & dedicated smearing applied
- AK4 jets DeepCSV classifier: External efficiency SF available for b,c and light jets.
- Mass-decorrelated AK8 jet DeepAK8 classifier: External efficiency SF available for bb signal output node (extrapolated from g→bb).

Improving reconstructed dijet mass resolutions
- DNN-based b-jet energy regression → dedicated smearing derived
- Kinematic fit (2 lepton channel)
- FSR recovery

*Charge Hadron Subtracted

Object reconstruction: Improving (di-)jet mass resolution

DNN-based b-jet energy regression

- Energy correction due to escaping neutrino from semi-leptonic decay, calibration mis-match.

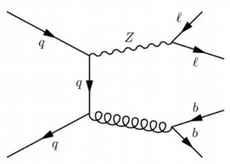
FSR recovery: (all channels)

- Add close jets within dR<0.8, p_T>30 to closest of the two selected b-jets

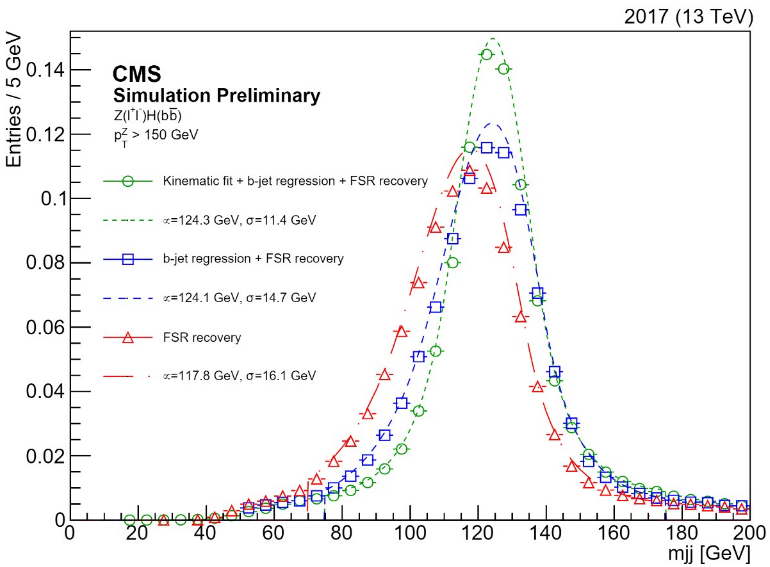
Kinematic fit: (only in two lepton channel)

- leptons have better momentum resolution than jets no intrinsic MET
- Fit leptons and jets with uncertainties under constraints m(ll) = m(Z) and pT(total) = 0.
- Thus, get constraints on individual jets resolution.

Dedicated smearing



- Good detector resolution of leptons allows us to use Z(ll)bb process to estimate jet scaling/smearing factors.
- Dedicated nuisances in fit for jet energy scale and smear factors.



bTag mismodelling

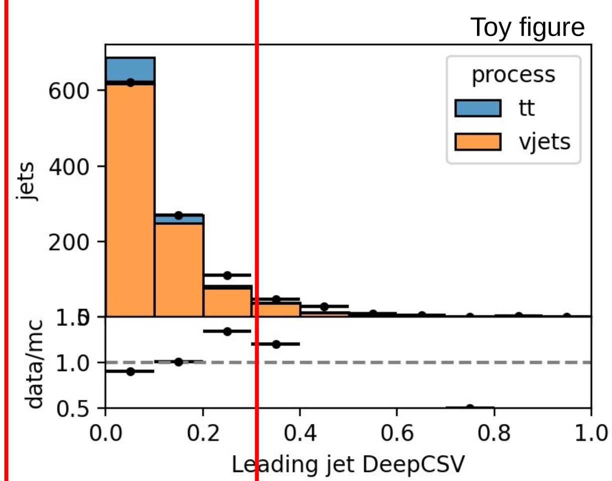
DeepCSV < loose WP

dRbb mismodelling

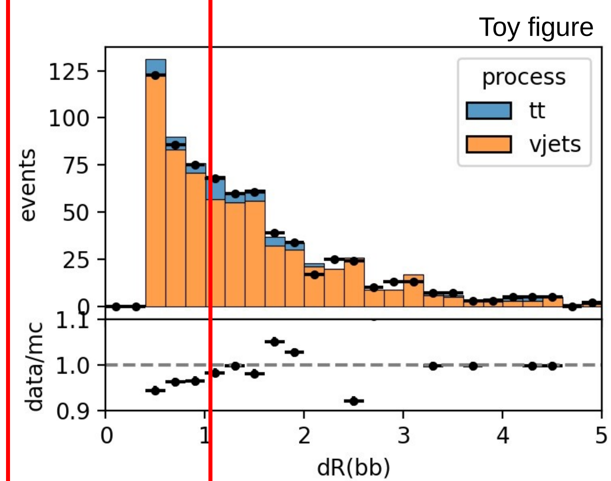
dRbb < 1

V+jets with deepCSV < loose WP
are mis-modelled even after btag SFs application

V+jets with dRbb < 1.0
are mis-modelled due to well-known MG generator feature at low dRbb



Loose WP ◁— B-jets classifier score corresponding to light-jets mis-tag rate of 10%.

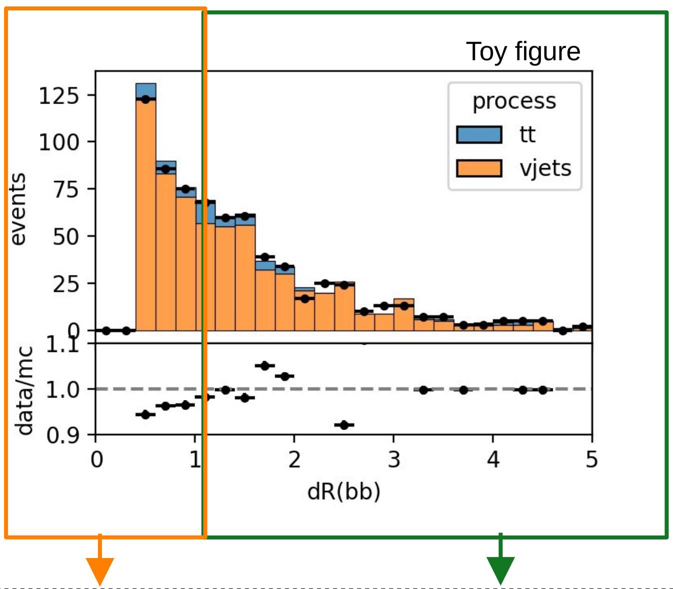


dRbb< 1

How are btag and dRbb corrections derived ?

- LF CR used for derivation of “btag” and “dRbb” mis-modelling corrections.
  - “btag” corrections are derived for LF CR with additional cut dRbb > 1.0 & extrapolated to V+jets in dRbb<1.0
  - “dRbb” corrections are derived for LF CR & extrapolated to V+jets in HF CR and SR.
- Assumption: “dRbb” corrections are independent of V+jets flavor. (as shown in the studies carried out by the VHcc analysis)
- Corrections are derived separately for each channel and Vpt bin.
- Pre-fit closure of this technique with data was found satisfactory.
- Systematic uncertainties associated to the dRbb reweighting are propagated in the fit model.

If LF CR has leading or subleading jet < loose WP

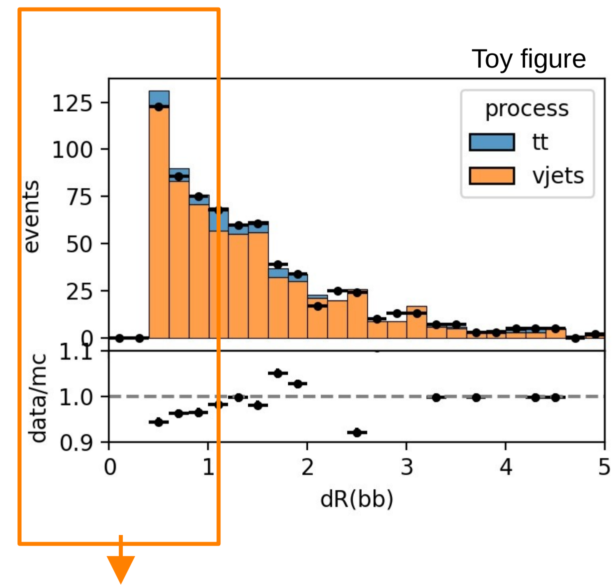


Step 1 bTag mismodelling   bTag mismodelling

+

Step 2 dRbb mismodelling

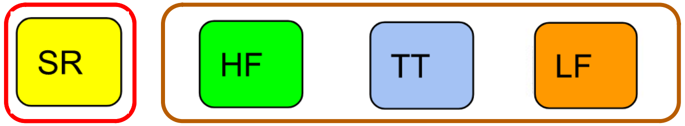
If LF CR has both leading and subleading jets > loose WP



dRbb mismodelling   Step 1

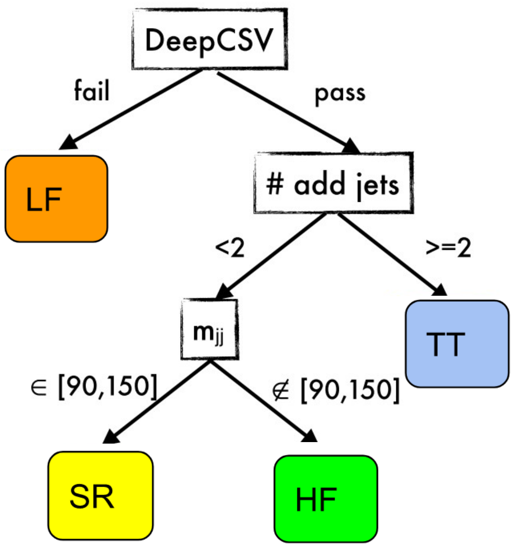
- Introduction (Signal, background, goal, samples)
- Object reconstruction
- **Event selection**
- Fit model
- Cross check analysis results (VZbb & dijet mass analysis)
- VHbb unblinded results

Event selection

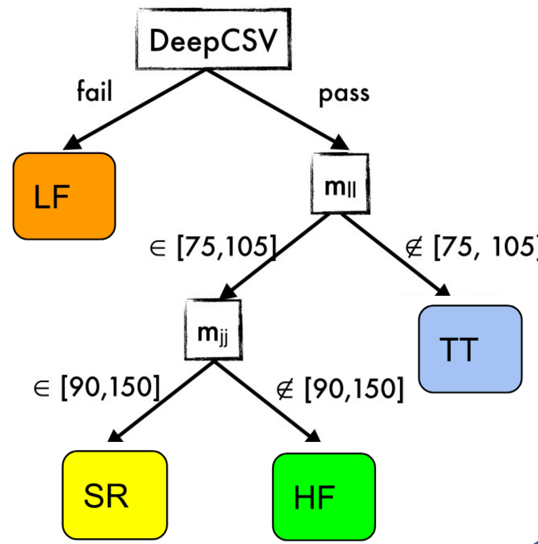


Resolved topology

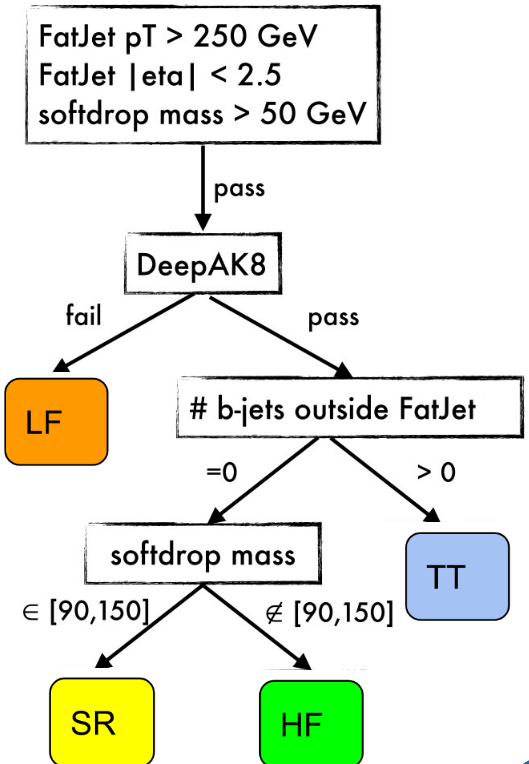
simplified strategy 0- & 1-lepton



simplified strategy 2-lepton



Boosted topology



Analysis regions for Resolved topology

	Channels					
$p_T(V)$	Zee	Zmm	Wen	Wmn	Znn	$p_T(V)$
75.0	SR, TT, HF, LF	SR, TT, HF, LF				75.0 low
150.0	0 Addj, >=1 Addj, TT, HF, LF	0 Addj, >=1 Addj, TT, HF, LF	SR, TT, HF, LF	SR, TT, HF, LF	0 Addj, >=1 Addj, TT, HF, LF	150.0 med
250.0						250.0 high1
400.0						400.0 high2

*for Znn channel, med bin $p_T(V)$ ranges from 170-250GeV due to trigger selection

Analysis regions for Boosted topology

	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

*for Znn channel, med bin ranges from 170-250GeV due to trigger selection

Overlap event strategy

Events passing both resolved and boosted cuts above $p_T(V) > 250$ GeV are assigned according to following priority order:

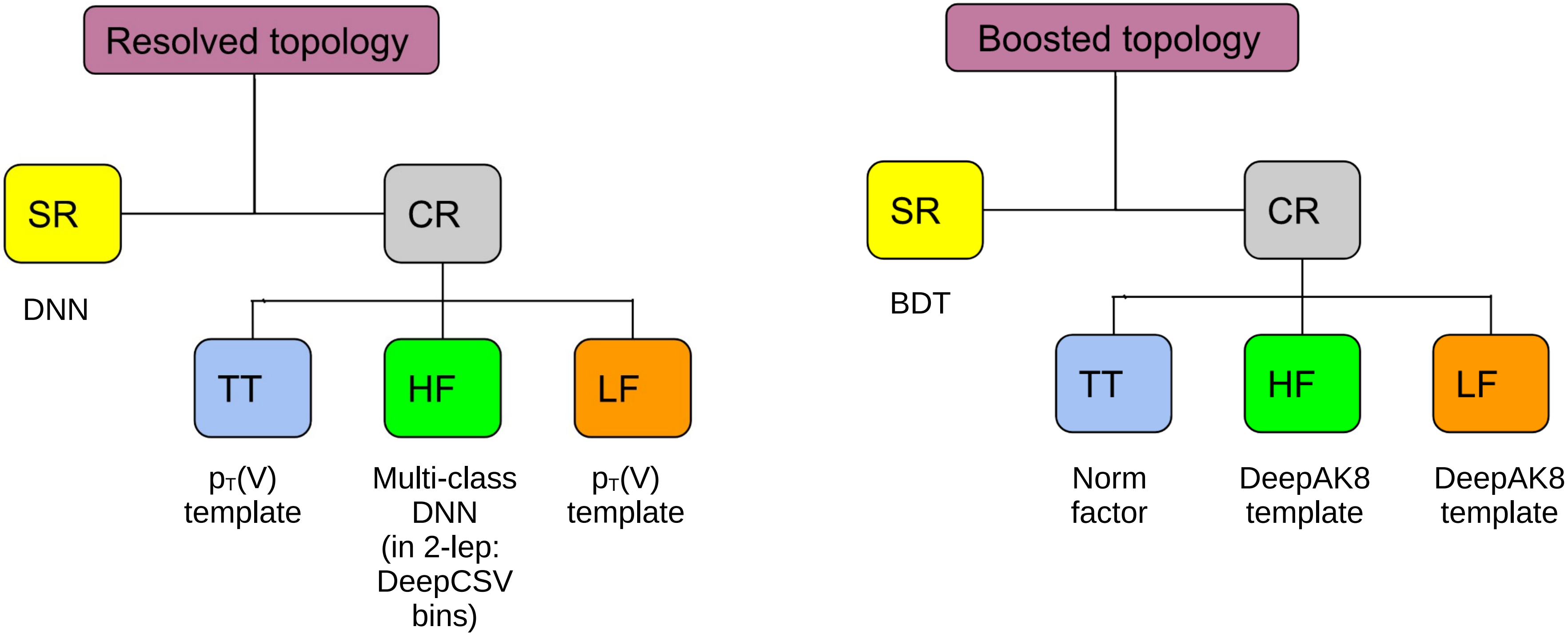
1. Resolved SR
2. Boosted SR
3. Resolved CR
4. Boosted CR.

		Resolved		
		SR	CR	
Boosted	SR	R	B	B
	CR	R	R	B
		R	R	

~15% gain compared to having just resolved events with $p_T(V)<250$GeV and boosted thereafter

- Introduction (Signal, background, goal, samples)
- Object reconstruction
- Event selection
- **Fit model**
- Cross check analysis results (VZbb & dijet mass analysis)
- VHbb unblinded results

What do we fit?

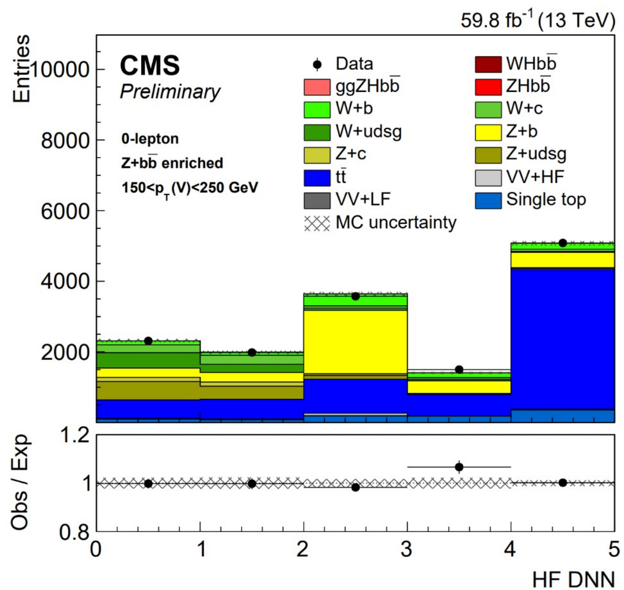


Multivariate variables in fit

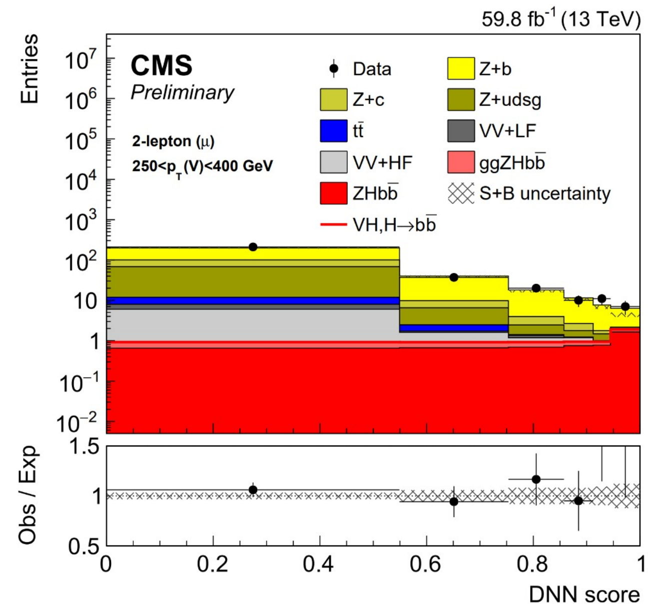
DNN/ HF DNN :

- Channel dependent 15-27 high-level input features whose data/MC is verified in CR.
- Separate DNN for low (75< $p_T V$< 150) and med+high (250< $p_T V$) STXS bins.

HF DNN



S/B DNN



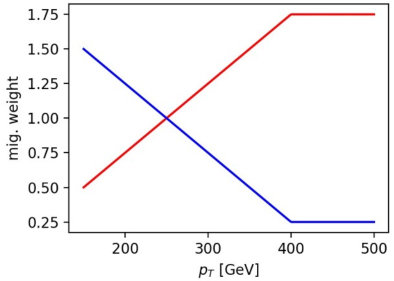
BDT:

- Uses FatJet kinematic variables + deepAK8.
- Overlap events have resolved features as well.

MVA binning: 1 signal event/bin as it improves the Asimov sensitivity

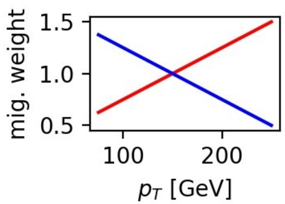
Free-floating SF scheme

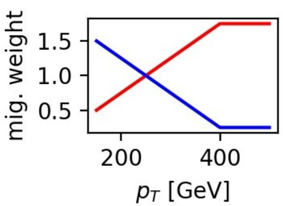
- Since MVA output are highly dependent on $p_T(V)$, SF per STXS $p_T(V)$ bin could result in ~accurate yields for the bin however, process yield as a function of $p_T(V)$ is not continuous.
- To ensure continuous postfit $p_T(V)$ distribution, we use 1 SF with category migration nuisances at bin boundary.
- Implementing linear category migration uncertainties help in getting better post-fit data/MC closure in the $p_T(V)$.
- The category migration uncertainties have been given large prior uncertainties.
- Uncorrelated SF and category migration among the years.
- SF are split in electron and muon channel for 2-lepton and 1-lepton channel in the fit to have additional degrees of freedom to improve the overall goodness of fit.
  - We do not observe any inconsistency in the postfit results for electron and muon SF.
- CR->SR extrapolations:
  - No explicit extrapolation factor from CR to SR.
  - Postfit background yields/SF consistent in CR-only and CR+SR fit.
  - Also, as definition of SR is relatively loose, the background constraints are also partially exploited in SR.



SF_{mig_250} = 1-0.005*(min(400.0,max(Vp$_T$,150.0))-250.0)

Process SF for TT



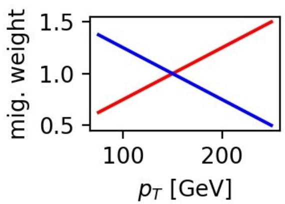


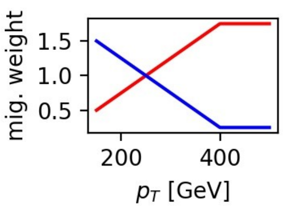
	Channels					
$p_T(V)$	Zee	Zmm	Wen	Wmn	Znn	$p_T(V)$
75.0						75.0 low
150.0						150.0 med
250.0						250.0 high1
400.0						400.0 high2

1 free-floating
- TT SF in TT, V+HF, V+LF CR & SR of all the channels

Linear $p_T(V)$ shape migration uncertainty

Process SF for Z+udsg OR Z+c OR Z + b/bb process





	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

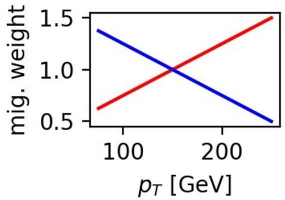
Linear p_T(V) shape migration uncertainty

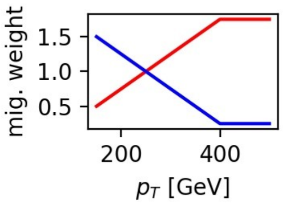
1 free-floating
- Z+udsg SF in TT, V+HF, V+LF CR & SR of Zee, Zmm & Znn channel
- Z+c SF in TT, V+HF, V+LF CR & SR of Zee, Zmm & Znn channel
- Z+b/bb SF in TT, V+HF, V+LF CR & SR of Zee, Zmm & Znn channel

V+b and V+bb SF are merged since sub-leading b-jet is usually soft and its inclusion depends on the acceptance of kinematic cuts.

Process SF for W+udsg OR W+c OR W+b/bb

	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





Linear $p_T(V)$ shape migration uncertainty

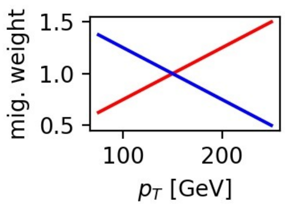
1 free-floating
- W+udsg SF in TT, V+HF, V+LF CR & SR of Wen & Wmn channel
- W+c SF in TT, V+HF, V+LF CR & SR of Wen & Wmn channel
- W+b/bb SF in TT, V+HF, V+LF CR & SR of Wen & Wmn channel

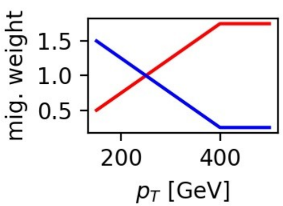
V+b and V+bb SF are merged since sub-leading b-jet is usually soft and its inclusion depends on the acceptance of kinematic cuts.

In-situ SF for AK8 TT FatJet

These SF account for efficiency measurements of DeepAK8 since external SF for tt FatJet are unavailable

	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



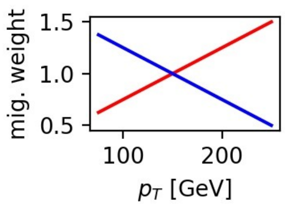


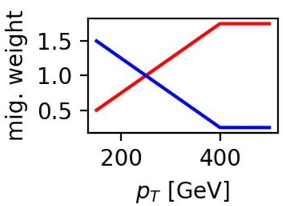
1 correlated free-floating TT AK8 SF in Boosted TT, V+HF CR & SR of
- All the channels

Linear $p_T(V)$ shape migration uncertainty

In-situ SF for AK8 V+udsg FatJet

These SF account for efficiency measurements of DeepAK8 since external SF for V+udsg FatJet are unavailable





	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

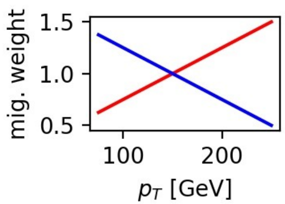
1 correlated free-floating V+udsg AK8 SF in Boosted V+LF CR & SR of
- Zee and Wen channel
- Zmm and Wmn channel
- Znn channel.

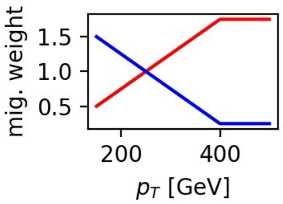
Linear $p_T(V)$ shape migration uncertainty

In-situ SF for AK8 V+b/bb FatJet

These SF account for efficiency measurements of DeepAK8 since external SF for V+b/bb FatJet are unavailable

	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





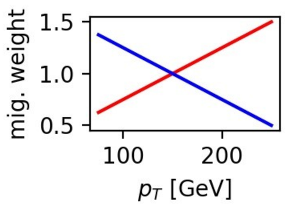
Linear $p_T(V)$ shape migration uncertainty

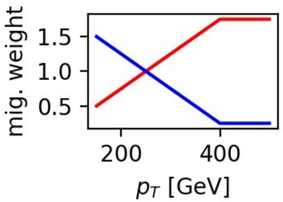
1 correlated free-floating V+b/bb AK8 SF in Boosted V+HF CR & SR of
- Zee and Wen channel
- Zmm and Wmn channel
- Znn channel.

V+b and V+bb SF are merged since sub-leading b-jet is soft and its inclusion depends on the acceptance of kinematic cuts.

In-situ SF for AK8 V+c FatJet

These SF account for efficiency measurements of DeepAK8 since external SF for V+c FatJet are unavailable





	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

1 correlated free-floating C AK8 SF in Boosted V+LF CR of

- All the channels.

1 correlated free-floating C AK8 SF in Boosted V+HF CR & SR of

- All the channels.

Linear $p_T(V)$ shape migration uncertainty

VHbb signal STXS uncertainty

Migration uncertainty

- Calculated for maximum split scheme.
- Each bin boundary acts as a source of uncertainty.
- 6 sources/NP: Each $p_T(V)$ bin boundary : Δ_X, X= 75, 150, 250, 400 GeV
  Each njet bin boundary : Δ_X, X= 1, 2
- Δ_X evaluation: Max. absolute deviation of 7-point scale of μ_R and μ_F variations
  $[\mu_R/\mu_{Rnom}, \mu_F/\mu_{Fnom}]$ = (0.5,0.5), (0.5,1), (1, 0.5), (1, 1), (1, 2), (2, 1), (2, 2)
- Δ_X distribution among STXS bins: 'short range scheme-2'
  - Anti-correlated variation of Δ_X for $p_T(V)$ > X and $-\Delta_X$ for $p_T(V)$ < X
  - Impact of $-\Delta_X$ only on the first bin just below X.

$p_T(V)$ bin (GeV)	Δ_{75}	Δ_{150}	Δ_{250}	Δ_{400}
[0, 75[	$-\Delta_{75}/\sigma_{[0,75[}$	0	0	0
[75, 150[	$\Delta_{75}/\sigma_{[75,\infty[}$	$-\Delta_{150}/\sigma_{[75,150[}$	0	0
[150, 250[	$\Delta_{75}/\sigma_{[75,\infty[}$	$\Delta_{150}/\sigma_{[150,\infty[}$	$-\Delta_{250}/\sigma_{[150,250[}$	0
[250, 400[	$\Delta_{75}/\sigma_{[75,\infty[}$	$\Delta_{150}/\sigma_{[150,\infty[}$	$\Delta_{250}/\sigma_{[250,\infty[}$	$-\Delta_{400}/\sigma_{[250,400[}$
[400, ∞[	$\Delta_{75}/\sigma_{[75,\infty[}$	$\Delta_{150}/\sigma_{[150,\infty[}$	$\Delta_{250}/\sigma_{[250,\infty[}$	$\Delta_{400}/\sigma_{[400,\infty[}$

For each $p_T(V)$ bin

n_{jets} bin	Δ_1	Δ_2
0 jets	$-\Delta_1/\sigma_{n_{jets}=0}$	0
1 jet	$\Delta_1/\sigma_{n_{jets}\geq 1}$	$-\Delta_2/\sigma_{n_{jets}=1}$
$\geq$ 2 jets	$\Delta_1/\sigma_{n_{jets}\geq 1}$	$\Delta_2/\sigma_{n_{jets}\geq 2}$

Acceptance uncertainty

- Scale variations in a bin keeping cross section of the bin same.
- If the bins are merged, we estimate residual uncertainty from STXS migration uncertainty.
- If the bin is not merged we take the simultaneous muR, muF variation as an estimate.
- In both cases, only shape effects are considered in each STXS bin,

Other nuisances in the fit model

Systematic	Type	Comments
Diboson cross section uncertainty	Log normal	15%
Single top cross section uncertainty	Log normal	15%
LHE PDF	Log normal	Uncorrelated across proc.
Renormalisation scale	Shape	Uncorrelated across proc.
Factorisation scale	Shape	Uncorrelated across proc.
DeepCSV SF – related nuisances (BTV-POG)	Shape	Applied to AK4 jets and decorrelated across jet pT and eta.
DeepAK8 SF – related nuisances (BTV-POG)	Shape	Applied to boosted Hbb process and decorrelated in bins of FatJet pT.
Jet Energy Calibration	Shape	Applied to AK4 jets
Jet Energy Scale (VHbb specific derivation)	Shape	Applied to AK4 jets
Jet Energy Smear (VHbb specific derivation)	Shape	Applied to AK4 jets
JMS (Jet mass scale)	Shape	Applied to AK8 jets
JMR (Jet mass resolution)	Shape	Applied to AK8 jets

Other nuisances in the fit model

Systematic	Type	Comments
dR(bb)	Shape	Uncorrelated across V+jets process (only in 2017/2018).
deta(bb)	Shape	Uncorrelated across V+jets process (only in 2016).
PU	Shape	Correlated across proc.
Lumi (Lumi POG)	Log normal	Uncorrelated & partial-correlated b/w years recommended scheme
VHbb theoretical migration uncertainties	Log normal	Uncorrelated across Hbb signals.
VHbb theoretical acceptance uncertainties (scale variations)	Shape	Uncorrelated across Hbb signals.
Lepton efficiency (ID, ISO, trigger)	Log normal	Uncorrelated across lepton flavour and channel,
MET trigger efficiency	Log normal	For Znn lepton channel
PDF and alpha_s uncertainty	Log normal	1.6% for ZH and 1.9% for WH
Hbb branching ratio (LHC HXSWG)	Log normal	5%

Fit Results: SFs close to unity (TT, V+jets) with no major pulls on nuisances.

Impact of different nuissances on inclusive signal strength

	$\Delta\ \mu$
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)	$+0.001\ -0.001$

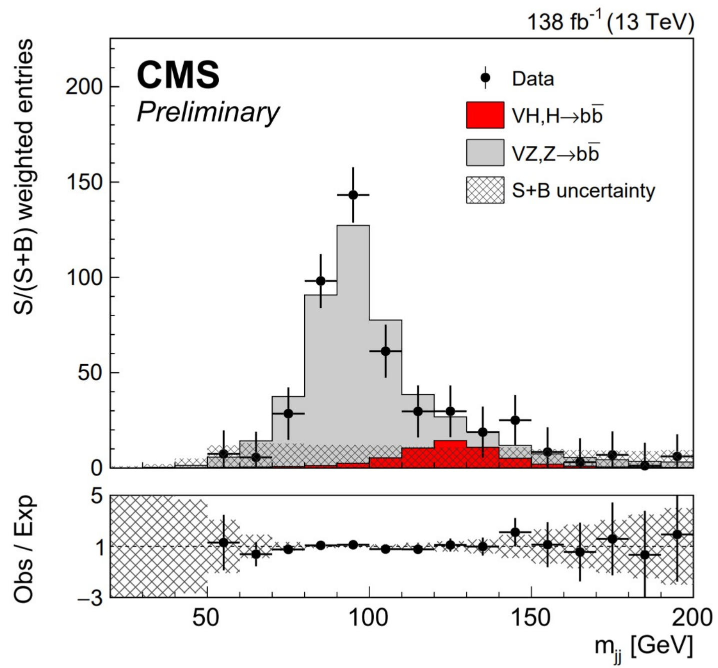
MC stats.: Mainly from V+jets sample size

Sim. modelling:
deta(bb) uncertainties
dR(bb) uncertainties
$p_T(V)$ category migration

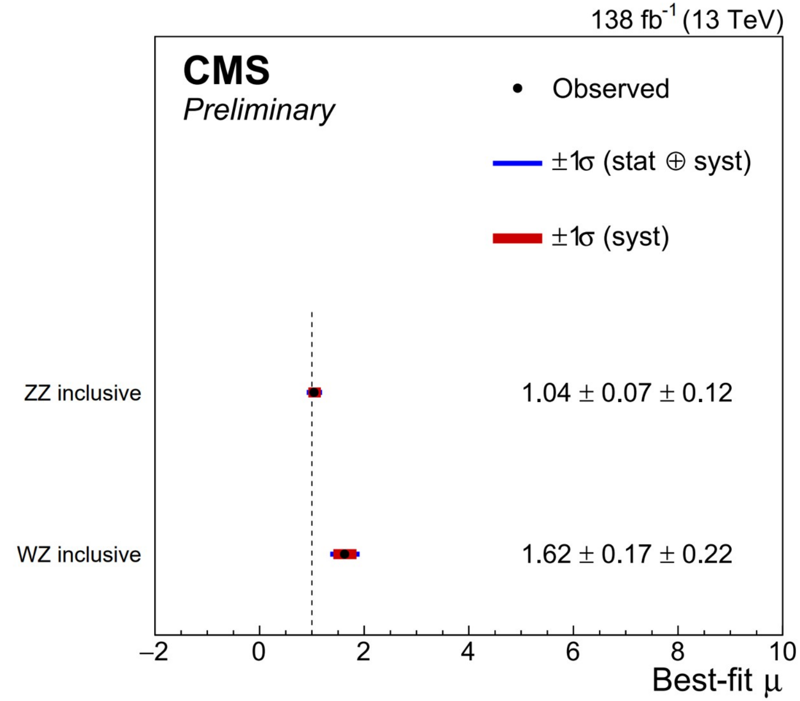
- Introduction (Signal, background, goal, samples)
- Object reconstruction
- Event selection
- Fit model
- **Cross check analysis results (VZbb & dijet mass analysis)**
- VHbb unblinded results

Cross check analysis (mjj and VZbb)

μ = 0.34 +/- 0.34



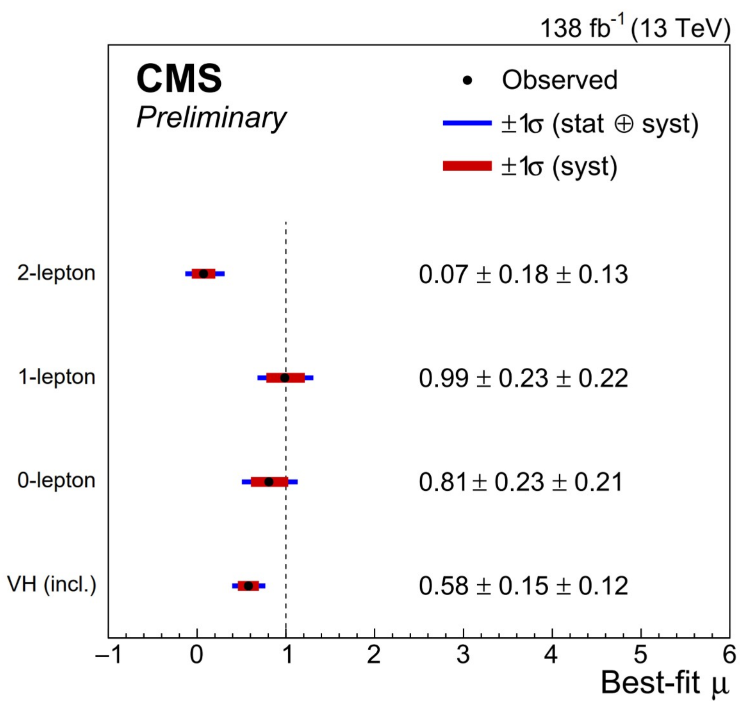
μ = 1.16 +/- 0.13

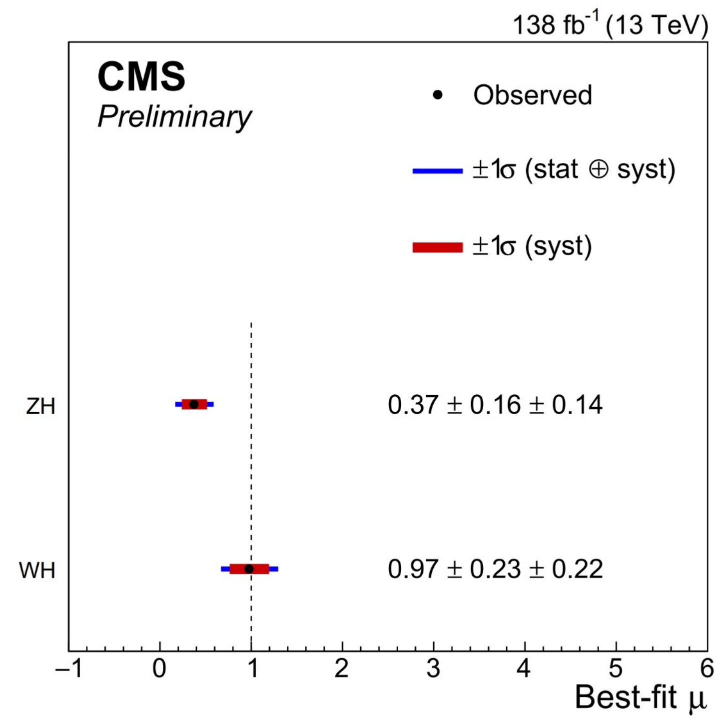


- Fit mjj distribution in 4 different bins of DNN (same DNN but with mass-correlated inputs fixed) score for SR.
- Same CR used in the fit.
- Combine SR post-fit mjj distribution of all channels by weighting events with S/(S+B).
- Sensitivity little lower than for fit with DNN score.

- VZ(bb) as signal instead of VH(bb).
- Same final state, similar kinematics but different dijet invariant mass.

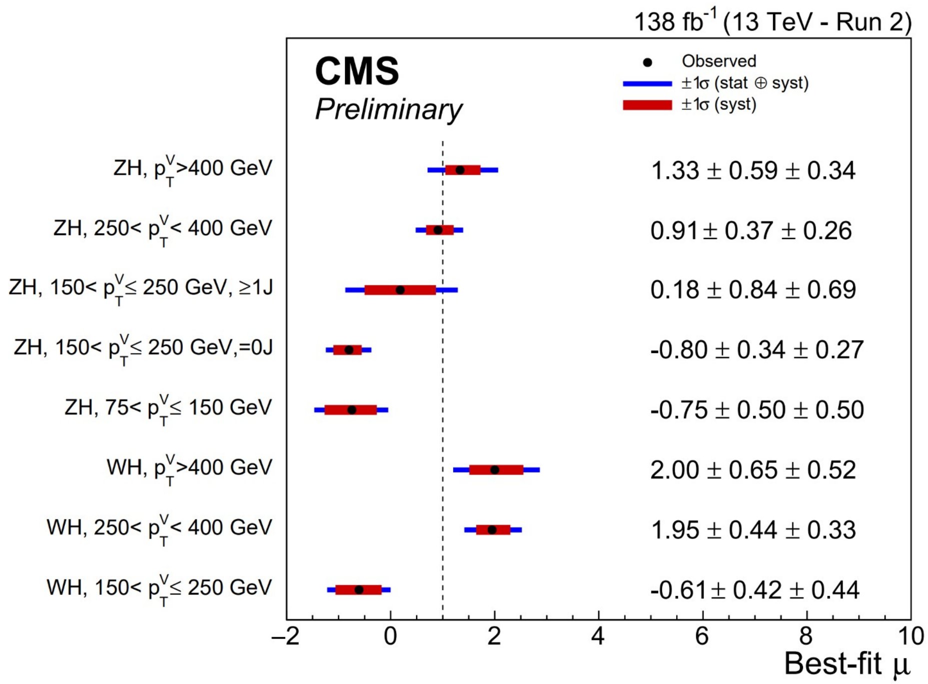
- Introduction (Signal, background, goal, samples)
- Object reconstruction
- Event selection
- Fit model
- Cross check analysis results (VZbb & dijet mass analysis)
- **VHbb unblinded results**

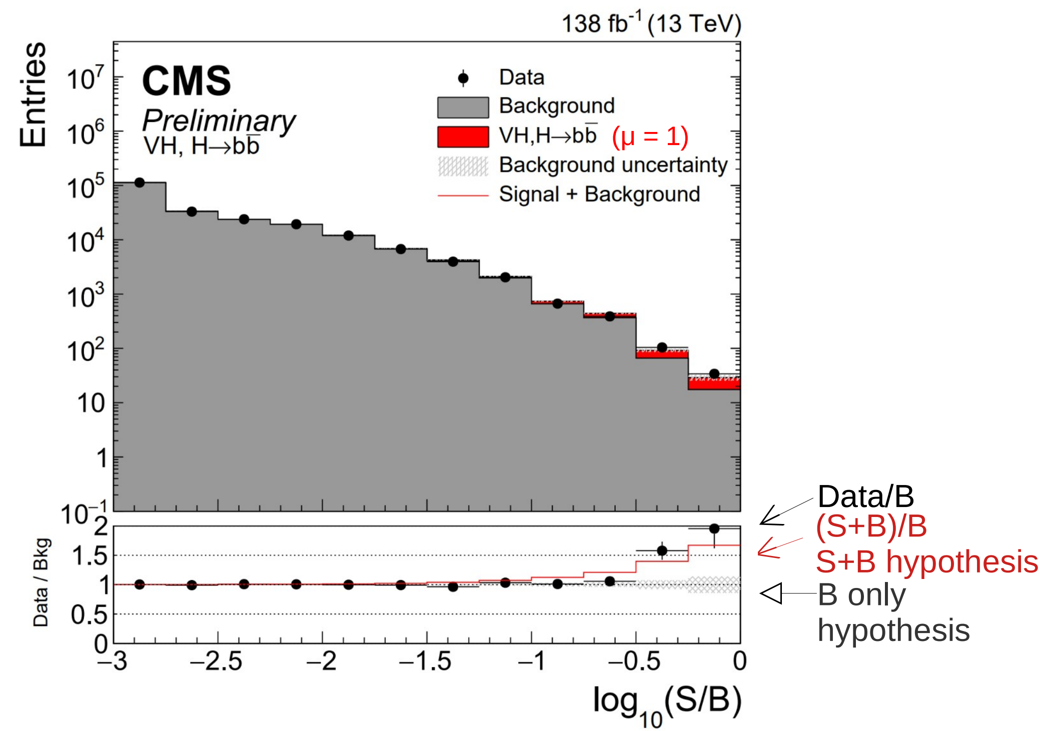




- The compatibility p-value between the Z(ll)H/Z(νν)H production modes against the inclusive ZH signal extraction is 20%.
- The local observed (expected) significance of the measured ZH and WH signals, over the background-only expectation, is found to be 3.3 standard deviations (5.2 standard deviations)
- Jacknife procedure results in 2 standard deviations between the results shown here and the previously published results, when taking only the 2017 data set into account. (see back-up)

VHbb Run 2 results





Summary

- CMS VHbb STXS analysis using full Run 2 data discussed.
- Focus on modeling of MC events and fit model.
- Analysis strategy with analysis regions and fit input variables discussed.
- Usage of SF with category migration and other nuisances in fit model discussed.

Back-Up

Types of V+jets NLO samples and stitching

Several sets of V+jets NLO samples available for the analysis depending on channel and year.

- Inclusive (used as target for LOtoNLO $\Delta\eta(bb)$ reweighting),
- Split in jet multiplicity,
- Split in vector boson momentum,
- Split in vector boson momentum and jet multiplicity

Performed stitching of all available samples to improve effective statistics for the analysis

Stitching procedure (for reference):

Example: region A (largest possible subset common between all the samples available in that region)

sample X: n unweighted events
sample Y: m unweighted events
sample Z: g unweighted events

stitching weight of a sample in region A:
sample X: n/(n+m+g)
sample Y: m/(n+m+g)
sample Z: g/(n+m+g)

Event selection

Variable	SR	Z + b-jets	Z + light-jets	$t\bar{t}$
Common selection:				
min(MET,MHT)	> 100	-//-	-//-	-//-
E_T^{miss}	> 170	-//-	-//-	-//-
p_T^{j1}	> 60	-//-	-//-	-//-
p_T^{j2}	> 35	-//-	-//-	-//-
$p_T(jj)$	> 120	-//-	-//-	-//-
$\Delta\phi(Z,H)$	> 2.0	-//-	-//-	-//-
$M(jj)$	$> 50, < 500$	-//-	-//-	-//-
N_{al}	< 1	-//-	-//-	-//-
Njets close to MET	0	-//-	-//-	-//-
Different between SR and CRs:				
N_{aj}	≤ 1	≤ 1	≤ 1	≥ 2
$M(jj)$	[90-150]	$\notin$ [90 – 150]	-	-
$btag_{max}$	> medium	> medium	< medium	> medium
$btag_{min}$	> loose	> loose	> loose	> loose
$\Delta\phi$(pfMET,trkMET)	< 0.5	< 0.5	< 0.5	-
min $\Delta\phi$(pfMET,J)	-	-	-	$< \pi/2$

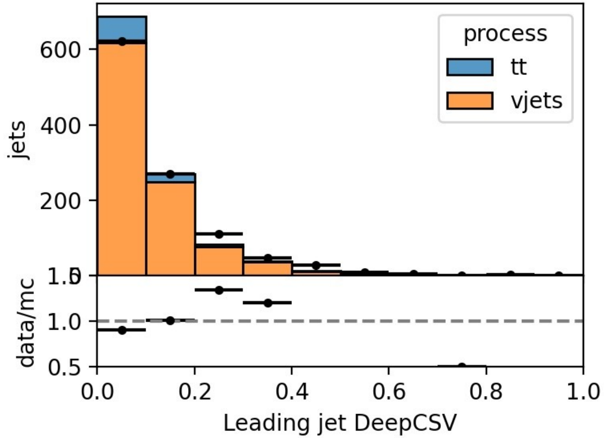
Table 16: 0-lepton channel selection for control and signal regions

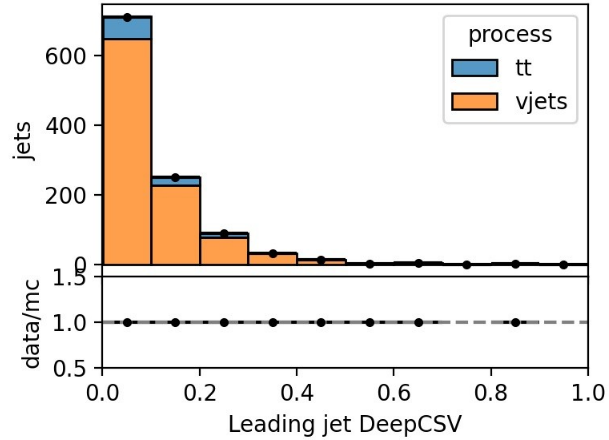
Variable	SR	W + b-jets	W + light-jets	$t\bar{t}$
Common selection:				
$p_T(jj)$	> 100	-//-	-//-	-//-
$p_T(V)$	> 150	-//-	-//-	-//-
N_{lep}	< 1	-//-	-//-	-//-
p_T^{j1}	> 25	-//-	-//-	-//-
p_T^{j2}	> 25	-//-	-//-	-//-
$\Delta\phi$(lep, pfMET)	< 2	-//-	-//-	-//-
Difference between SR and CRs:				
$btag_{max}$	> *medium*	>medium	[loose-medium]	>tight
$btag_{min}$	> *loose*	-	-	-
$M(jj)$	[90,150]	[150,250] and <90	<250	< 250
N_{aj}	< 2	< 2	-	>1
σ(MET)	-	> 2	> 2	-
$\Delta\phi(H,V)$	< 2.5	-	-	-

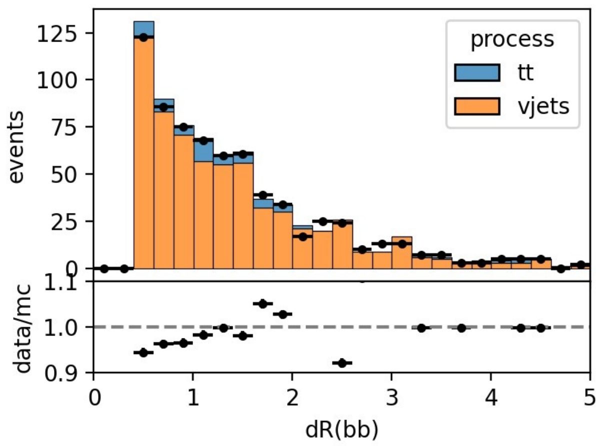
Table 17: Definition of the SR and CR for the 1-lepton channel resolved selection.

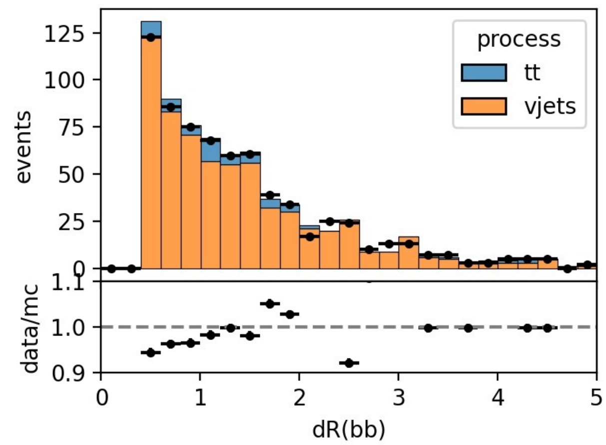
Variable	SR	Z + b-jets	Z + light-jets	$t\bar{t}$
Common selection:				
p_T^{j1}	> 20	-//-	-//-	-//-
p_T^{j2}	> 20	-//-	-//-	-//-
$p_T(V)$	> 75	-//-	-//-	-//-
$M(jj)$	> 50	-//-	-//-	-//-
$btag_{max}$	>medium	>medium	<loose	>tight
$btag_{min}$	>loose	>loose	<loose	>loose
$M(V)$	[75,105]	[85,97]	[75,105]	[10,75] and <120
$M(jj)$	[90,150]	$\notin$[90,150]	[90,150]	-
p_T(MET)	-	<60	-	-
$\Delta\phi(H,V)$	-	> 2.5	> 2.5	-

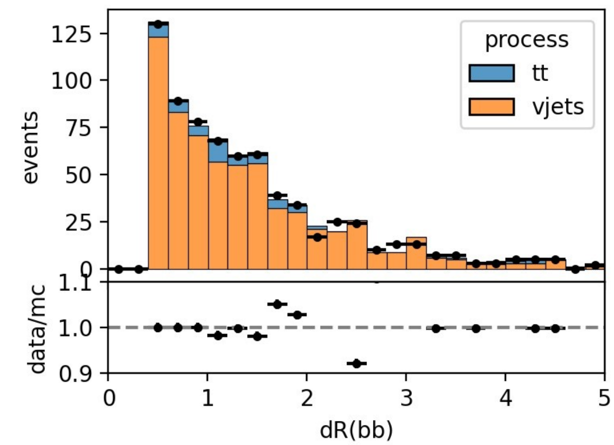
Table 18: Definition of the SR and CR for the 2-lepton channel resolved selection.


process
tt
vjets
jets
data/mc
Leading jet DeepCSV


process
tt
vjets
jets
data/mc
Leading jet DeepCSV


process
tt
vjets
events
data/mc
dR(bb)


process
tt
vjets
events
data/mc
dR(bb)

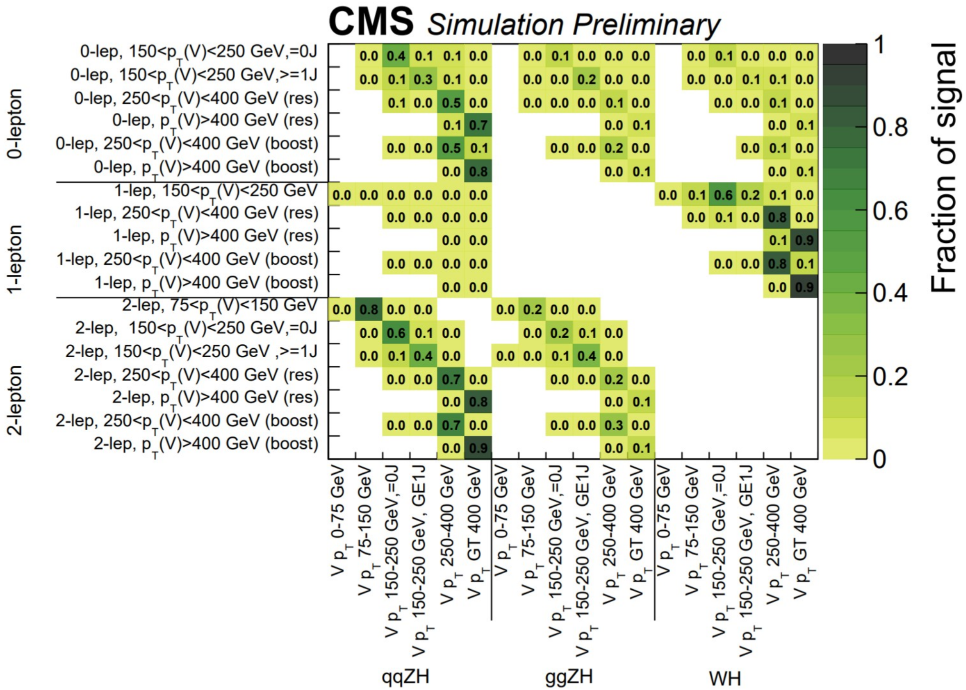

process
tt
vjets
events
data/mc
dR(bb)

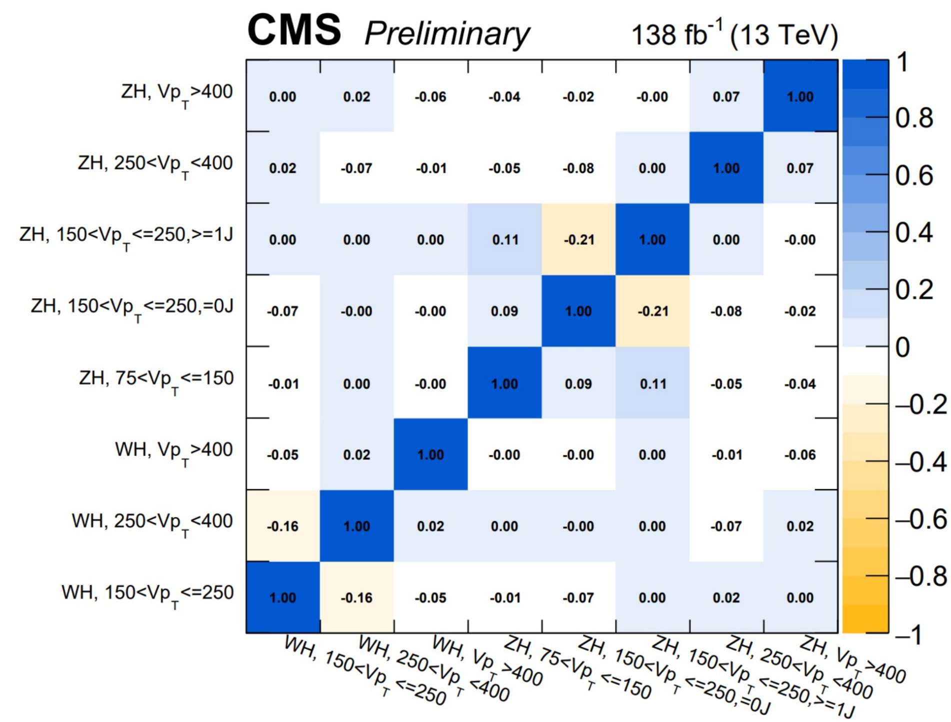
Input features of DNN

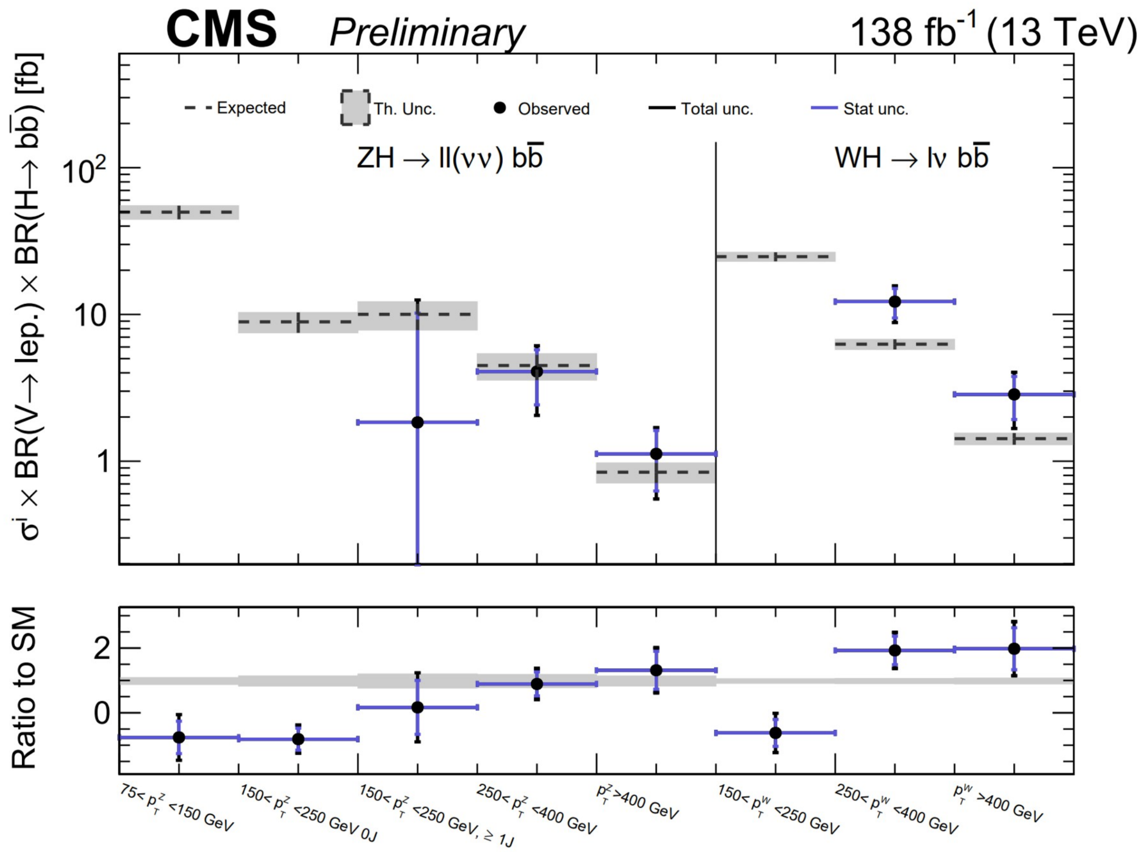
Variable	Description	0-lepton	1-lepton	2-lepton
$M(\mathrm{jj})$	Dijet invariant mass	✓	✓	✓
$p_{\mathrm{T}}(\mathrm{jj})$	Dijet transverse momentum	✓	✓	✓
$\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$	Missing transverse momentum	✓	✓	✓
$\mathrm{M_t(V)}$	Transverse mass of the vector boson		✓	
$p_{\mathrm{T}}(\mathrm{V})$	Transverse momentum of the vector boson		✓	✓
$p_{\mathrm{T}}(\mathrm{jj}) / p_{\mathrm{T}}(\mathrm{V})$	Ratio of transverse momenta of the vector boson and Higgs boson		✓	✓
$\Delta\phi(\mathrm{V,H})$	Azimuthal angle between the vector boson and the dijet directions	✓	✓	✓
$\mathrm{btag_{max}}$	b tagging score of leading jet	✓	✓	✓
$\mathrm{btag_{min}}$	b tagging score of subleading jet	✓	✓	✓
$\Delta\eta(\mathrm{jj})$	Pseudorapidity difference between leading and sub-leading jet	✓	✓	✓
$\Delta\phi(\mathrm{jj})$	Azimuthal angle between leading and sub-leading jet	✓	✓	
$p_T^{\mathrm{max}}(\mathrm{j_1,j_2})$	Maximum transverse momentum of jet between leading and subleading jet	✓	✓	
SA5	Number of soft-track jets with momentum greater than 5 GeV	✓	✓	✓
$\mathrm{N_{aj}}$	Number of additional jets	✓	✓	
$\mathrm{btag_{max}}$ (add)	Maximum btagging discriminant score among additional jets	✓		
$p_T^{\mathrm{max}}(\mathrm{add})$	Maximum transverse momentum among additional jets	✓		
$\Delta\phi(\mathrm{jet}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$	Azimuthal angle between additional jet and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$	✓		
$\Delta\phi(\mathrm{lep}, \vec{p}_{\mathrm{T}}^{\mathrm{miss}})$	Azimuthal angle between lepton and $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$		✓	
M_{t}	Reconstructed top quark mass		✓	
$p_{\mathrm{T}}(\mathrm{j_1})$	Transverse momentum of leading jet			✓
$p_{\mathrm{T}}(\mathrm{j_2})$	Transverse momentum of sub-leading jet			✓
$M(\mathrm{V})$	Reconstructed vector boson mass			✓
$\Delta R(\mathrm{V,H})$	Angular separation between the vector boson and Higgs boson			✓
$\Delta R(\mathrm{V,H})$ (kin)	Angular separation between the vector boson (reconstructed after kinematic fit) and Higgs boson			✓
$\sigma(\mathrm{M(jj)})$	Resolution of dijet invariant mass			✓
N_{rec}	Number of recoil jets			✓

BDT input features:

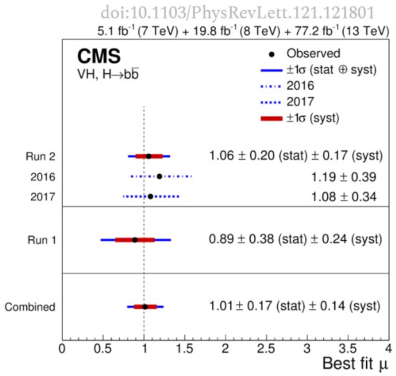
- FatJet softdrop mass
- FatJet DeepAK8 bbVsLight WP
- FatJet p_T
- V boson properties: mass(V), transverse mass (V)
- Resolved features for overlap events (except DeepCSV)


CMS Simulation Preliminary
Fraction of signal
0-lepton
1-lepton
2-lepton
0-lep, 150<p_T(V)<250 GeV,=0J
0-lep, 150<p_T(V)<250 GeV,>=1J
0-lep, 250<p_T(V)<400 GeV (res)
0-lep, p_T(V)>400 GeV (res)
0-lep, 250<p_T(V)<400 GeV (boost)
0-lep, p_T(V)>400 GeV (boost)
1-lep, 150<p_T(V)<250 GeV
1-lep, 250<p_T(V)<400 GeV (res)
1-lep, p_T(V)>400 GeV (res)
1-lep, 250<p_T(V)<400 GeV (boost)
1-lep, p_T(V)>400 GeV (boost)
2-lep, 75<p_T(V)<150 GeV
2-lep, 150<p_T(V)<250 GeV,=0J
2-lep, 150<p_T(V)<250 GeV ,>=1J
2-lep, 250<p_T(V)<400 GeV (res)
2-lep, p_T(V)>400 GeV (res)
2-lep, 250<p_T(V)<400 GeV (boost)
2-lep, p_T(V)>400 GeV (boost)
V p_T 0-75 GeV
V p_T 75-150 GeV
V p_T 150-250 GeV,=0J
V p_T 150-250 GeV, GE1J
V p_T 250-400 GeV
V p_T GT 400 GeV
qqZH
ggZH
WH


CMS Preliminary
138 fb^-1 (13 TeV)
ZH, Vp_T>400
ZH, 250<Vp_T<400
ZH, 150<Vp_T<=250,>=1J
ZH, 150<Vp_T<=250,=0J
ZH, 75<Vp_T<=150
WH, Vp_T>400
WH, 250<Vp_T<400
WH, 150<Vp_T<=250


CMS Preliminary
138 fb-1 (13 TeV)
Expected
Th. Unc.
Observed
Total unc.
Stat unc.
ZH → ll(νν) bb̄
WH → lν bb̄
σi × BR(V→ lep.) × BR(H→ bb̄) [fb]
Ratio to SM
75< pTZ <150 GeV
150< pTZ <250 GeV 0J
150< pTZ <250 GeV, ≥ 1J
250< pTZ <400 GeV
pTZ >400 GeV
150< pTW <250 GeV
250< pTW <400 GeV
pTW >400 GeV

VH->bb Comparisons



Jackknife Comparison

- Estimate correlation and agreement between correlated measurements
  - Remove a portion of data events, re-run both analyses, repeat. Extract compatibility from distribution of signal strengths
- **Previous vs current 2017** inclusive measurements
  - ~**50% correlation** between the 2017 measurements
  - ~**2σ compatibility**

Slide from Nick Haubrich, Higgs 2022