CMS Draft Analysis Note

The content of this note is intended for CMS internal use and distribution only

2023/03/08
Archive Hash: 3f2c4e0-D
Archive Date: 2023/02/24

Search for Anomalous Higgs Boson Couplings in the Production of $W^{\pm}H$ via Vector Boson Scattering

Philip Chang[2], Leonardo Giannini[1], Jonathan Guiang[1], Avi Yagil[1], and Frank Würthwein[1]

[1] University of California, San Diego
[2] University of Florida

Abstract

This note describes the search for negative couplings between the Higgs boson with the W and Z bosons in the production of $W^{\pm}H$ via vector boson scattering. The target of the search is the rejection of the $\lambda_{WZ} = \frac{\kappa_W}{\kappa_Z} = -1$ hypothesis, where κ_W and κ_Z are the coupling modifiers as defined in the κ-framework between the the Higgs boson and the W and Z bosons respectively. The search is performed in the one lepton, two b quark, and two jet final state using a cut-based approach. This work is based on a data sample of proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ recorded by the CMS experiment during 2016, 2017, and 2018, corresponding to a total integrated luminosity of $138\,\text{fb}^{-1}$. The hypothesis of $\kappa_W = -1$, $\kappa_Z = +1$ is specifically excluded with a significance of 9.0σ. Moreover, 625 such κ_W, κ_Z hypotheses are analyzed, and the $\lambda_{WZ} < 0$ scenario is excluded.

Changelog

Version 1

- Incomplete first draft

Version 2

- Filled in missing sections
- Added abstract

Version 3

- Adressed internal comments and made changes to most sections
- First complete draft

Version 4

Section 1

- Changed sentence to make it more clear that this analysis targets the merged $H \rightarrow b\bar{b}$ final state

Section 4

- Added table with signal and background yields in the Preselection and signal region (Table 12)
- Added some missing MET filters
- Added a sentence to clarify that the Control Regions are only used to ensure that there is no significant mismodeling in the Monte Carlo and that they are not otherwise used in the analysis

Appendix C

- Added another systematic to the background extrapolation for the diboson composition

Version 5

General

- Added changelog
- Removed SM VBS WH sample from set of backgrounds

Section 1

- Replaced duplicate diagram in Fig. 1 with a previously missing diagram

Section 2

- Added description of 2D κ_W, κ_Z scan signal samples
- Added plot of all κ_W, κ_Z points scanned (Fig. 5)
- Added plot of all κ_W, κ_Z point cross sections (Fig. 6)
- Added validation plot of κ_W, κ_Z scan cross sections (Fig. 7)

Section 7

- Added description of 2D exclusion
- Added 2D exclusion result (Fig. 18)

Version 6

Section 2

- Added PYTHIA version and tune information
- Added description of MADGRAPH phase space cuts

Section 7

- Added table with estimated background yields for each year of UL NanoAOD (Table 17)

Version 7

General

- Every appendix is now referenced at least once in the text

Section 4

- Added some additional text to explain how the SR selections were chosen
- Changed caption for Table 12 for clarity

Section 5

- Added table with A, B, C, and D yields (Table 13)

Section 6, 7

- Decorrelated ParticleNet Xbb scale factors across years
- Added 1.27% uncertainty on the $H \to b\bar{b}$ BR
- Minor language fixes

Contents

1 Introduction

The Higgs boson was the last missing particle to complete the standard model (SM). After its discovery [1, 2], the LHC experiments have measured many of its properties and have found no significant deviations from SM predictions [3, 4].

One commonly used framework used to quantify the deviations from the SM is the so-called k-framework [5], which introduces coupling modifiers for each fermion and boson coupling. If we consider in particular the gauge bosons (W^+, W^-, and Z) and their coupling modifiers (κ_W and κ_Z respectively), which are object of study of this note, we see that their magnitudes have been constrained to $|\kappa_W| = 1.02^{+0.11}_{-0.10}$ and $|\kappa_Z| = 1.04^{+0.07}_{-0.07}$ by the CMS Collaboration, showing precise agreement with the Standard Model [3]. However, there is no constraint on the sign of κ_W or κ_Z, nor the relative sign between them.

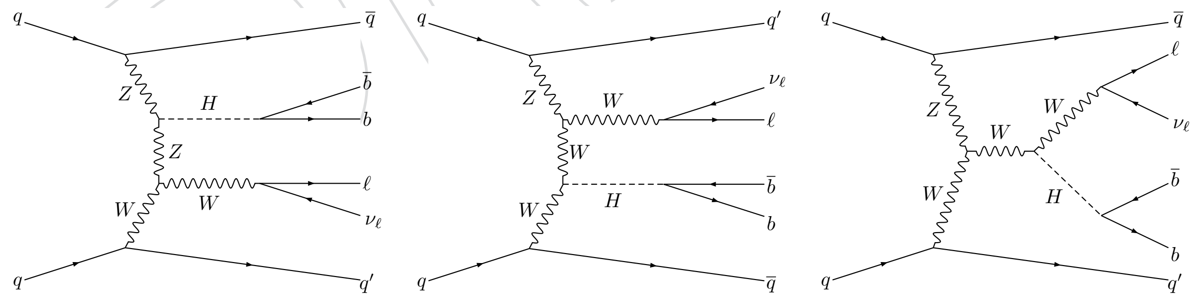
The relative sign between κ_W and κ_Z can be expressed as their ratio:

$$\lambda_{WZ} = \frac{\kappa_W}{\kappa_Z} = \pm 1 \tag{1}$$

These couplings have the same sign in the SM ($\lambda_{WZ} = 1$) in order to preserve the custodial symmetry. This property, however, has not yet been confirmed with experimental data, as the processes studied so far are quadratic in κ_W or κ_Z and therefore only sensitive to their magnitudes. A possible channel to directly probe the λ_{WZ} ratio at the LHC is the production of VH via vector-boson scattering (VBS) [6]. Such a channel is sensitive to the relative sign of κ_W and κ_Z since the since the cross section σ has an interference term that is linear in both κ_W and κ_Z:

$$\sigma \propto |\mathcal{M}|^2 = \kappa_W^2 |\mathcal{M}_W|^2 + \kappa_W \kappa_Z \mathcal{M}_{WZ}^2 + \kappa_Z^2 |\mathcal{M}_Z|^2 \tag{2}$$

Therefore, this channel provides the possibility to determine with certainty that λ_{WZ} is indeed positive, as the SM predicts.



(a) (b) (c)

Figure 1: Target Feynman diagrams of $pp \to W^{\pm}H + jj$ for this analysis.

A search for negative λ_{WZ} via the VBS production of VH with the CMS detector is documented in this note. In particular, we present a search for $pp \to W^{\pm}H + jj$ in the one lepton, two b quark (merged into a single, large-cone jet), and two jet final state (Fig. 1), where $\kappa_W = -1$. We do not consider ZH production here, since $Z \to \ell^+\ell^-$ has a less favorable branching ratio and $Z \to q\bar{q}$ would require an entirely different analysis. We also do not explicitly perform a search for $\kappa_Z = -1$, however the cross-section and resultant kinematics are identical (Fig. 2).

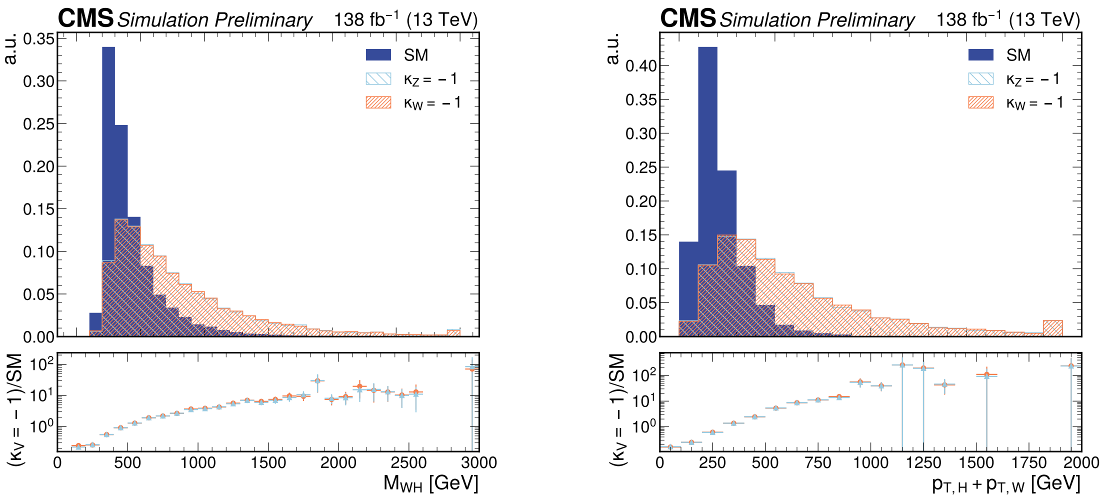


Figure 2: The invariant mass of the $W^{\pm}H$ system (left) and S_T (right) are plotted using the $pp \rightarrow W^{\pm}H + jj$ Monte Carlo simulation. The data plotted is taken directly from MADGRAPH5_aMC@NLO.

The $pp \rightarrow W^{\pm}H + jj$ signal process for this analysis is accessible at the LHC because setting $\kappa_W = \pm 1$ and $\kappa_Z = \mp 1$ catastrophically changes the interference between Feynman diagrams that contribute to the final state of interest. This firstly results in a 6-fold increase in the overall cross section versus the SM (Fig. 3). The cross section importantly changes most dramatically at high values of the combined invariant mass of the scattering bosons–e.g. $M_{W^{\pm}Z}$ in the case of VBS $W^{\pm}H$. This can be seen explicitly for $W^{+}W^{-} \rightarrow ZH$ in Fig. 1 in the paper by D. Stolarski and Y. Wu [6], or equivalently in Fig. 4a in this note, where the combined invariant mass of the outgoing W and H is plotted using Monte Carlo simulation at the generator level. Importantly, this implies that the decay products of W and H receive a significant boost, providing a powerful discriminating feature in addition to the already distinct VBS signature of $pp \rightarrow W^{\pm}H + jj$. See Section 4.3 for more details.

This note is structured as follows. The CMS proton-proton collision data and Monte Carlo (MC) simulation samples used in this analysis are listed in Section 2. Section 3 details how, using this data, leptons and jets are defined and more precisely selected for this analysis. The definition of the kinematic "signal region" (SR) for this analysis is documented in Section 4, and Section 5 illustrates how background contamination in this region is estimated. The systematic uncertainties on the expected signal and background yields in the signal region are cataloged in Section 6. Finally, Section 7 presents the yields in the signal region along with the corresponding statistical interpretation, followed by a complete summary in Section 8.

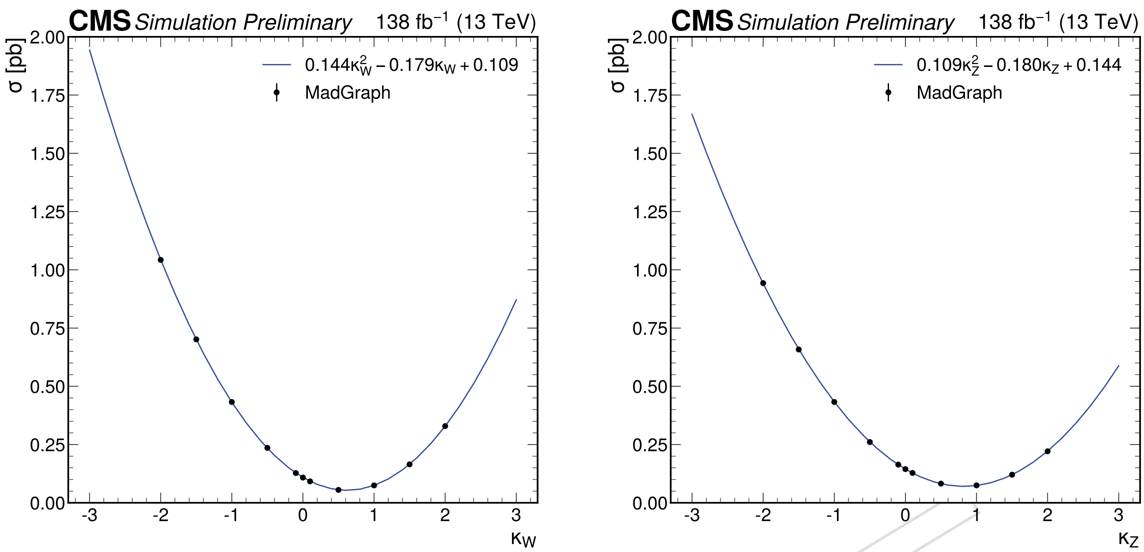


Figure 3: The cross section is plotted as a function of enhancements to κ_W (keeping $\kappa_Z = +1$) on the left and to κ_Z (keeping $\kappa_W = +1$) on the right. The black points are taken from MADGRAPH5_aMC@NLO, and the blue curve is the best fit of a 2nd order polynomial to those points. The errors are also plotted, but are smaller than the width of the black points. Importantly, the cross section for $\kappa_W = -1, \kappa_Z = +1$ and $\kappa_W = +1, \kappa_Z = -1$ are exactly the same.

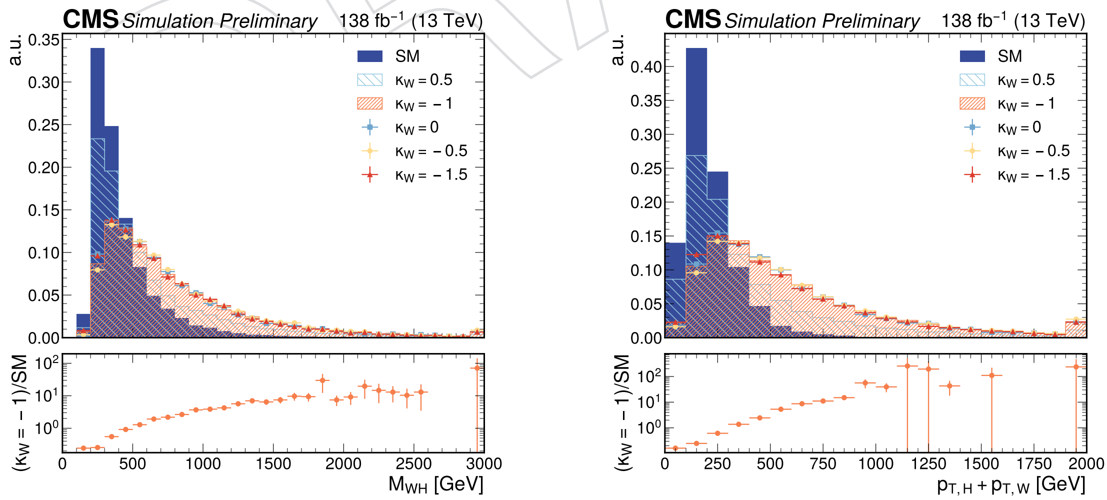


Figure 4: The invariant mass of the $W^{\pm}H$ system (left) and S_T (right) are plotted using the $pp \rightarrow W^{\pm}H + jj$ Monte Carlo simulation. The data plotted is taken directly from MADGRAPH5_aMC@NLO.

2 Samples

2.1 Collision data

This analysis uses an amount of CMS proton-proton collision data corresponding to an integrated luminosity of $138\,\mathrm{fb}^{-1}$. In particular, the SingleMuon and SingleElectron datasets are analyzed using the NanoAOD data tier (version 9) under the "Ultra Legacy" (UL) reconstruction campaign. The "golden" JSON files used to certify the events used in this analysis are tabulated in Table 1, and the triggers used to filter this data are listed in Table 11.

Year	Golden JSON file
2016 (pre-VFP)	Cert_271036-284044_13TeV_Legacy2016_Collisions16_JSON.txt
2016 (post-VFP)	Cert_271036-284044_13TeV_Legacy2016_Collisions16_JSON.txt
2017	Cert_294927-306462_13TeV_UL2017_Collisions17_GoldenJSON.txt
2018	Cert_314472-325175_13TeV_Legacy2018_Collisions18_JSON.txt

Table 1: Golden JSON files used in to certify the proton-proton collision events used in this analysis.

2.2 Monte Carlo simulation

2.2.1 Signal

The MC simulation of $\mathrm{pp} \to \mathrm{W}^{\pm}\mathrm{H} + \mathrm{jj}$ was generated at leading order (LO) using the MADGRAPH5_aMC@NLO generator [7] using a modified version of the Standard Model MADGRAPH model (see Appendix A), where $\kappa_{\mathrm{W}} = -1$ and $\kappa_{\mathrm{Z}} = +1$ in order to test $\lambda_{\mathrm{WZ}} = -1$. Importantly, the W and H bosons are decayed inclusively, and parton showers are handled by PYTHIA (version 8) with the CP5 tune[8]. Standard MADGRAPH phase space cuts are also applied, but two cuts are specifically tightened for generating this signal process. In particular, the pseudorapidity of any jet is required to be less than 6.5 and the invariant mass of any two jets is required to be larger than 100 GeV. Additional settings are taken from centrally generated CMS MC samples at LO. The signal simulation was then processed through the same workflow as other UL MC samples. This analysis has been specifically optimized for this point ($\kappa_{\mathrm{W}} = -1, \kappa_{\mathrm{Z}} = +1$).

An additional set of signal samples were generated using the reweighting feature of MADGRAPH5_aMC@NLO such that the values of κ_{W} and κ_{Z} could be varied in a two-dimensional scan. Because the reweighting is done around a central point, two signal samples were generated: one reweighted around $\kappa_{\mathrm{W}} = -1$, $\kappa_{\mathrm{Z}} = +1$ such that $\lambda_{\mathrm{WZ}} \leq 0$ and another reweighted around $\kappa_{\mathrm{W}} = \kappa_{\mathrm{Z}} = +1$ such that $\lambda_{\mathrm{WZ}} > 0$. These samples are generated using the same version and settings of PYTHIA as the $\kappa_{\mathrm{W}} = -1$, $\kappa_{\mathrm{Z}} = +1$ sample. Moreover, the points scanned are plotted in Fig. 5, and the reweighting has been validated against samples generated for individual κ_{W}, κ_{Z} points (Fig. 7).

2.2.2 Background

The MC simulation samples of background processes are listed in Tables 2, 3, 4, and 5 corresponding to the 2016 pre-VFP, 2016 post-VFP, 2017, and 2018 detector conditions respectively. The selected MC samples are intended to represent the major backgrounds for this analysis, and the completeness of this set is validated against data in sidebands of relevant phase space–see Section 4.5. All MC samples also use the NanoAOD data tier under the UL reconstruction campaign. Notably, the samples containing the production of a leptonically decaying W via vector

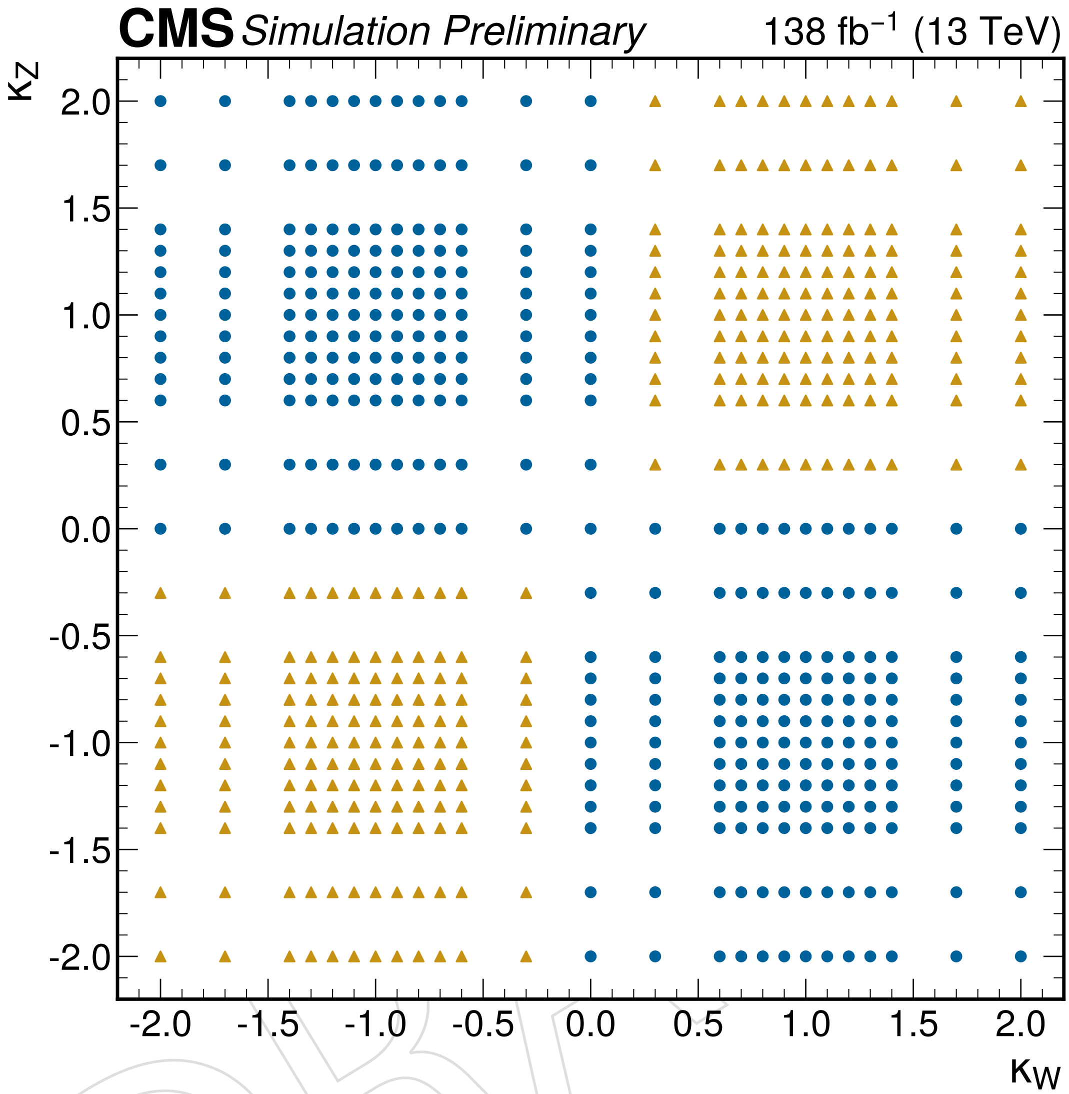


Figure 5: Scatter plot of the κ_W, κ_Z reweighting points that were scanned for this analysis, where the blue points were reweighted from $\kappa_W = -1$, $\kappa_Z = +1$ and the gold points were reweighted from $\kappa_W = \kappa_Z = +1$.

boson scattering were erroneously produced and had to be corrected manually (Appendix D). However, these samples are a very small background for this analysis.

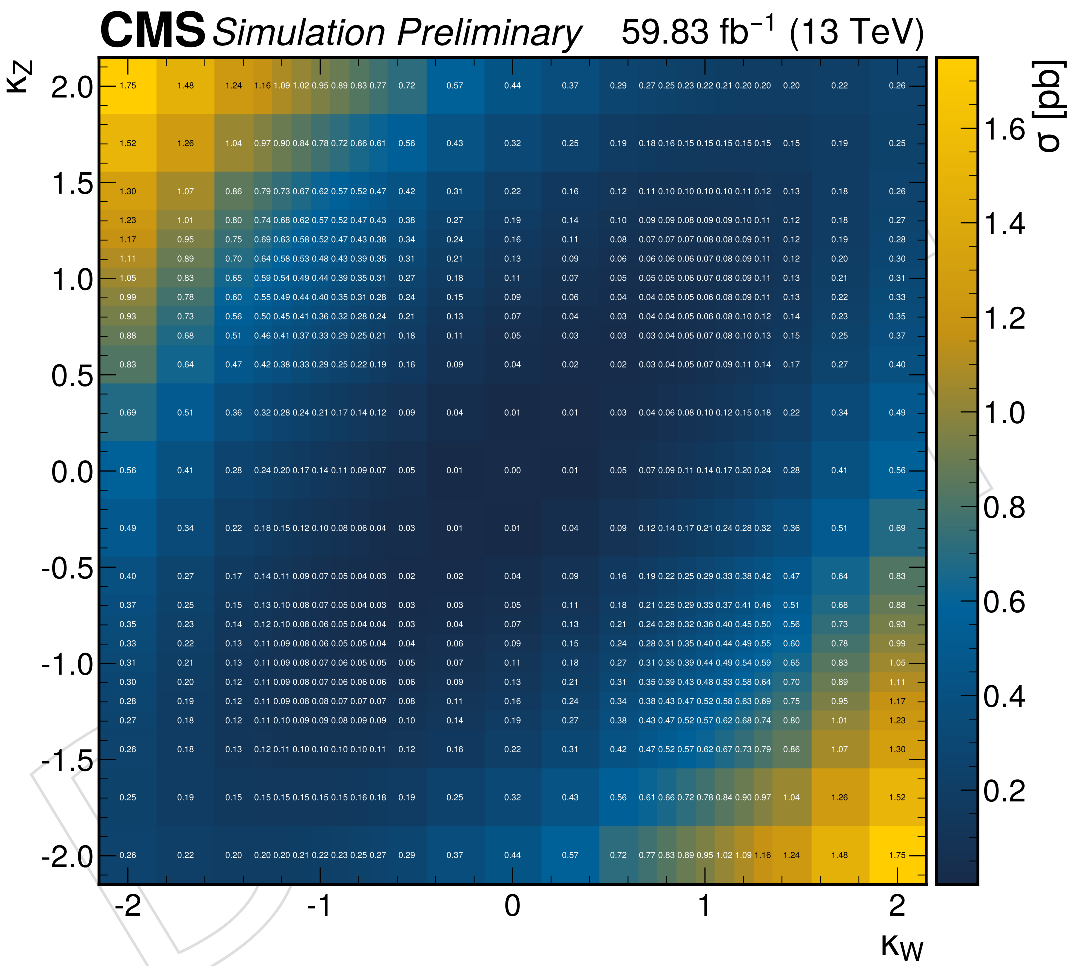


Figure 6: Two-dimensional histogram with bins centered at each κ_W, κ_Z point scanned for this analysis where the corresponding reweighted cross section is plotted on the z-axis.

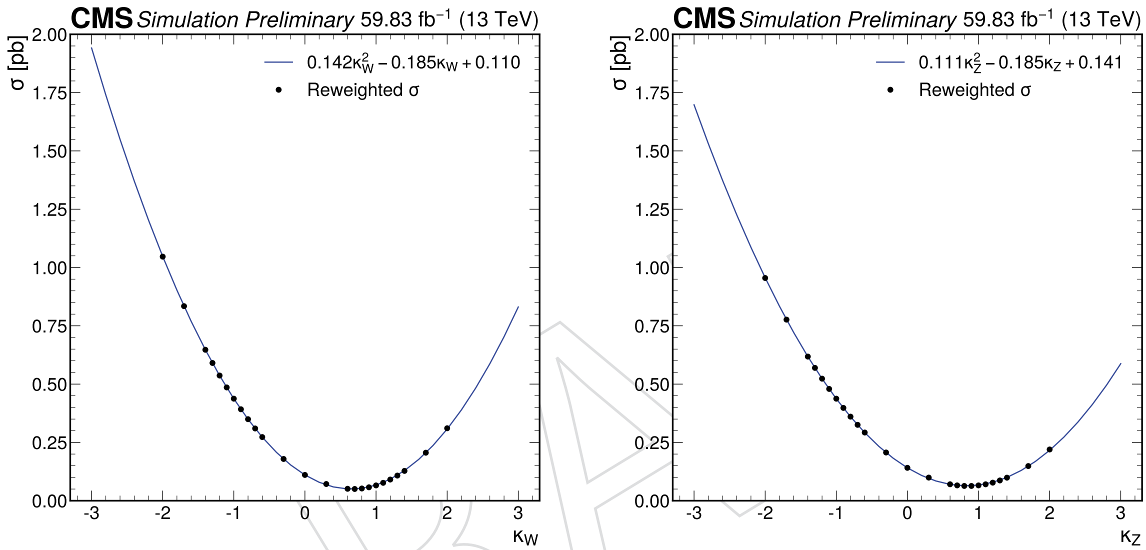


Figure 7: The cross section is plotted as a function of enhancements to κ_W (keeping $\kappa_Z = +1$) on the left and to κ_Z (keeping $\kappa_W = +1$) on the right. The black points are obtained by reweighting the central κ_W, κ_Z point by the MADGRAPH5_aMC@NLO reweighting, and the blue curve is the best fit of a 2nd order polynomial to those points. Based on the best fit, it is clear that the cross sections are nearly identical to those obtained by generating separate signal samples at a fixed κ_W, κ_Z point (Fig. 3).

Process	Sample Name	σ [pb]
SingleTop	/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8[0]	136.02
	/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8[0]	19.559
	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8[0]	19.559
	/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8[0]	80.95
TTbar1L	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8[0]	365.34
TTbar2L	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8[0]	88.29
TTX	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8[0]	377.96
	/TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8[1]	0.2043
	/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8[0]	0.2529
	/ttHTobb_M125_TuneCP5_13TeV-powheg-pythia8[1]	0.1279
	/ttHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8[1]	0.215
	/TTWZ_TuneCP5_13TeV-madgraph-pythia8[0]	0.003884
	/TTWW_TuneCP5_13TeV-madgraph-pythia8[0]	0.0115
	/TTbb_4f_TTTo2L2Nu_TuneCP5-Powheg-Openloops-Pythia8[0]	0.04
	/TTbb_4f_TTToSemiLeptonic_TuneCP5-Powheg-Openloops-Pythia8[0]	0.62
WJets	/WJetsToLNu_HT-70To100_TuneCP5_13TeV-madgraphMLM-pythia8[0]	1310.78
	/WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8[0]	1325.9
	/WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8[0]	348.5703
	/WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8[0]	47.308275
	/WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8[0]	11.358487
	/WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8[0]	5.2086934
	/WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8[0]	1.1880809
	/WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8[1]	0.024098031
Bosons	/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8[0]	20657.0
	/DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8[0]	6198.0
	/EWKWplus2Jets_WToQQ_dipoleRecoilOn_TuneCP5_13TeV-madgraph-pythia8[0]	10.67
	/EWKWminus2Jets_WToQQ_dipoleRecoilOn_TuneCP5_13TeV-madgraph-pythia8[0]	10.67
	/EWKZ2Jets_ZToLL_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8[0]	6.22
	/EWKZ2Jets_ZToNuNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8[1]	10.72
	/EWKZ2Jets_ZToQQ_dipoleRecoilOn_TuneCP5_13TeV-madgraph-pythia8[0]	10.67
	/WWTo1L1Nu2Q_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8[0]	51.65
	/WWTo2L2Nu_TuneCP5_13TeV-powheg-pythia8[0]	12.178
	/WZTo1L1Nu2Q_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8[0]	49.997
	/WZTo1L3Nu_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8[0]	3.054024
	/WZTo2Q2L_mllmin4p0_TuneCP5_13TeV-amcatnloFXFX-pythia8[1]	5.6
	/WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8[0]	4.42965
	/ZZTo2Q2L_mllmin4p0_TuneCP5_13TeV-amcatnloFXFX-pythia8[0]	3.28
	/ZZTo2L2Nu_TuneCP5_13TeV_powheg_pythia8[0]	0.564
	/ZZTo4L_M-1toInf_TuneCP5_13TeV_powheg_pythia8[0]	1.256
	/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8[2]	0.2086
	/WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8[2]	0.1651
	/WZZ_TuneCP5_13TeV-amcatnlo-pythia8[2]	0.05565
	/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8[2]	0.01398
	/WWJJToLNuLNu_EWK_noTop_TuneCP5_13TeV-madgraph-pythia8[0]	0.284
	/WZJJ_EWK_InclusivePolarization_TuneCP5_13TeV_madgraph-madspin-pythia8[0]	0.01701
	/ZZJJTo4L_TuneCP5_13TeV-madgraph-pythia8[1]	0.00884
EWKWLep	/EWKWPlus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8[0]	39.33
	/EWKWMinus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8[0]	32.26
VH	/VHToNonbb_M125_TuneCP5_13TeV-amcatnloFXFX_madspin_pythia8[1]	2.207
	/WminusH_HToBB_WToLNu_M-125_TuneCP5_13TeV-powheg-pythia8[0]	0.04901236122
	/WplusH_HToBB_WToLNu_M-125_TuneCP5_13TeV-powheg-pythia8[0]	0.08487599411
	/ZH_HToBB_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8[0]	0.02627486511
	/ggZH_HToBB_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8[0]	0.002461395396

[0] /RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v1/NANOAODSIM
[1] /RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11-v2/NANOAODSIM
[2] /RunIISummer20UL16NanoAODAPVv9-106X_mcRun2_asymptotic_preVFP_v11_ext1-v1/NANOAODSIM

Table 2: Background MC samples used in this analysis, corresponding to 2016preVFP detector conditions (UL), with their respective cross sections in picobarns. The cross sections for the H_{T}-binned W+jets samples are scaled by a "stitching" factor such that the samples together fill a continuous H_{T} distribution.

Process	Sample Name	σ [pb]
SingleTop	/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8[0]	80.95
	/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8[0]	136.02
	/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8[1]	19.559
	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8[1]	19.559
TTbar1L	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8[0]	365.34
TTbar2L	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8[0]	88.29
TTX	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8[0]	377.96
	/TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8[0]	0.2043
	/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8[0]	0.2529
	/ttHTobb_M125_TuneCP5_13TeV-powheg-pythia8[1]	0.1279
	/ttHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8[1]	0.215
	/TTWZ_TuneCP5_13TeV-madgraph-pythia8[0]	0.003884
	/TTWW_TuneCP5_13TeV-madgraph-pythia8[0]	0.0115
	/TTbb_4f_TTTo2L2Nu_TuneCP5-Powheg-Openloops-Pythia8[0]	0.04
	/TTbb_4f_TTToSemiLeptonic_TuneCP5-Powheg-Openloops-Pythia8[0]	0.62
WJets	/WJetsToLNu_HT-70To100_TuneCP5_13TeV-madgraphMLM-pythia8[0]	1283.91
	/WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8[0]	1303.06
	/WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8[0]	341.046
	/WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8[0]	45.4362
	/WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8[0]	11.0051
	/WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8[0]	4.94177
	/WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8[0]	1.15544
	/WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8[1]	0.0216234
Bosons	/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8[0]	20657.0
	/DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8[0]	6198.0
	/EWKWplus2Jets_WToQQ_dipoleRecoilOn_TuneCP5_13TeV-madgraph-pythia8[0]	10.67
	/EWKWminus2Jets_WToQQ_dipoleRecoilOn_TuneCP5_13TeV-madgraph-pythia8[0]	10.67
	/EWKZ2Jets_ZToLL_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8[0]	6.22
	/EWKZ2Jets_ZToNuNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8[1]	10.72
	/EWKZ2Jets_ZToQQ_dipoleRecoilOn_TuneCP5_13TeV-madgraph-pythia8[0]	10.67
	/WWTo1L1Nu2Q_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8[0]	51.65
	/WWTo2L2Nu_TuneCP5_13TeV-powheg-pythia8[0]	12.178
	/WZTo1L1Nu2Q_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8[0]	49.997
	/WZTo1L3Nu_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8[0]	3.054024
	/WZTo2Q2L_mllmin4p0_TuneCP5_13TeV-amcatnloFXFX-pythia8[1]	5.6
	/WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8[0]	4.42965
	/ZZTo2Q2L_mllmin4p0_TuneCP5_13TeV-amcatnloFXFX-pythia8[0]	3.28
	/ZZTo2L2Nu_TuneCP5_13TeV_powheg_pythia8[0]	0.564
	/ZZTo4L_M-1toInf_TuneCP5_13TeV_powheg_pythia8[0]	1.256
	/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8[2]	0.2086
	/WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8[2]	0.1651
	/WZZ_TuneCP5_13TeV-amcatnlo-pythia8[2]	0.05565
	/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8[2]	0.01398
	/WWJJToLNuLNu_EWK_noTop_TuneCP5_13TeV-madgraph-pythia8[0]	0.284
	/WZJJ_EWK_InclusivePolarization_TuneCP5_13TeV_madgraph-madspin-pythia8[0]	0.01701
	/ZZJJTo4L_TuneCP5_13TeV-madgraph-pythia8[1]	0.00884
EWKWLep	/EWKWPlus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8[0]	39.33
	/EWKWMinus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8[0]	32.26
VH	/VHToNonbb_M125_TuneCP5_13TeV-amcatnloFXFX_madspin_pythia8[1]	2.207
	/WminusH_HToBB_WToLNu_M-125_TuneCP5_13TeV-powheg-pythia8[0]	0.04901236122
	/WplusH_HToBB_WToLNu_M-125_TuneCP5_13TeV-powheg-pythia8[0]	0.08487599411
	/ZH_HToBB_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8[0]	0.02627486511
	/ggZH_HToBB_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8[0]	0.002461395396

[0] /RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v1/NANOAODSIM
[1] /RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17-v2/NANOAODSIM
[2] /RunIISummer20UL16NanoAODv9-106X_mcRun2_asymptotic_v17_ext1-v1/NANOAODSIM

Table 3: Background MC samples used in this analysis, corresponding to 2016postVFP detector conditions (UL), with their respective cross sections in picobarns. The cross sections for the H_T-binned W+jets samples are scaled by a "stitching" factor such that the samples together fill a continuous H_T distribution.

Process	Sample Name	σ [pb]
SingleTop	/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8[0]	136.02
	/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8[0]	80.95
	/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8[1]	19.559
	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8[1]	19.559
TTbar1L	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8[0]	365.34
TTbar2L	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8[0]	88.29
TTX	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8[0]	377.96
	/TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8[0]	0.2043
	/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8[0]	0.2529
	/ttHTobb_M125_TuneCP5_13TeV-powheg-pythia8[1]	0.1279
	/ttHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8[1]	0.215
	/TTWZ_TuneCP5_13TeV-madgraph-pythia8[0]	0.003884
	/TTWW_TuneCP5_13TeV-madgraph-pythia8[0]	0.0115
	/TTbb_4f_TTTo2L2Nu_TuneCP5-Powheg-Openloops-Pythia8[0]	0.04
	/TTbb_4f_TTToSemiLeptonic_TuneCP5-Powheg-Openloops-Pythia8[0]	0.62
WJets	/WJetsToLNu_HT-70To100_TuneCP5_13TeV-madgraphMLM-pythia8[0]	1319.76
	/WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8[0]	1334.74
	/WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8[0]	350.435
	/WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8[0]	46.5726
	/WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8[0]	11.1485
	/WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8[2]	5.02246
	/WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8[0]	1.183
	/WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8[1]	0.0258083
Bosons	/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8[0]	20657.0
	/DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8[0]	6198.0
	/EWKWplus2Jets_WToQQ_dipoleRecoilOn_TuneCP5_13TeV-madgraph-pythia8[0]	10.67
	/EWKWminus2Jets_WToQQ_dipoleRecoilOn_TuneCP5_13TeV-madgraph-pythia8[0]	10.67
	/EWKZ2Jets_ZToLL_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8[1]	6.22
	/EWKZ2Jets_ZToNuNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8[1]	10.72
	/EWKZ2Jets_ZToQQ_dipoleRecoilOn_TuneCP5_13TeV-madgraph-pythia8[0]	10.67
	/WWTo1L1Nu2Q_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8[0]	51.65
	/WWTo2L2Nu_TuneCP5_13TeV-powheg-pythia8[1]	12.178
	/WZTo1L1Nu2Q_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8[0]	49.997
	/WZTo1L3Nu_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8[0]	3.054024
	/WZTo2Q2L_mllmin4p0_TuneCP5_13TeV-amcatnloFXFX-pythia8[1]	5.6
	/WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8[1]	4.42965
	/ZZTo2Q2L_mllmin4p0_TuneCP5_13TeV-amcatnloFXFX-pythia8[0]	3.28
	/ZZTo2L2Nu_TuneCP5_13TeV_powheg_pythia8[0]	0.564
	/ZZTo4L_M-1toInf_TuneCP5_13TeV_powheg_pythia8[0]	1.256
	/WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8[0]	0.1651
	/WZZ_TuneCP5_13TeV-amcatnlo-pythia8[3]	0.05565
	/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8[3]	0.01398
	/WWJJToLNuLNu_EWK_noTop_TuneCP5_13TeV-madgraph-pythia8[0]	0.284
	/WZJJ_EWK_InclusivePolarization_TuneCP5_13TeV_madgraph-madspin-pythia8[0]	0.01701
	/ZZJJTo4L_TuneCP5_13TeV-madgraph-pythia8[1]	0.00884
	/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8[3]	0.2086
EWKWLep	/EWKWPlus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8[1]	39.33
	/EWKWMinus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8[1]	32.26
VH	/VHToNonbb_M125_TuneCP5_13TeV-amcatnloFXFX_madspin_pythia8[1]	2.207
	/WminusH_HToBB_WToLNu_M-125_TuneCP5_13TeV-powheg-pythia8[0]	0.04901236122
	/WplusH_HToBB_WToLNu_M-125_TuneCP5_13TeV-powheg-pythia8[0]	0.08487599411
	/ZH_HToBB_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8[0]	0.02627486511
	/ggZH_HToBB_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8[0]	0.002461395396

[0] /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v1/NANOAODSIM
[1] /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v2/NANOAODSIM
[2] /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9-v3/NANOAODSIM
[3] /RunIISummer20UL17NanoAODv9-106X_mc2017_realistic_v9_ext1-v2/NANOAODSIM

Table 4: Background MC samples used in this analysis, corresponding to 2017 detector conditions (UL), with their respective cross sections in picobarns. The cross sections for the H_T-binned W+jets samples are scaled by a "stitching" factor such that the samples together fill a continuous H_T distribution.

Process	Sample Name	σ [pb]
SingleTop	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8[0]	19.559
	/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8[2]	136.02
	/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8[2]	80.95
	/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8[0]	19.559
TTbar1L	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8[2]	365.34
TTbar2L	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8[2]	88.29
TTX	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8[2]	377.96
	/TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8[2]	0.2043
	/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8[2]	0.2529
	/ttHTobb_M125_TuneCP5_13TeV-powheg-pythia8[0]	0.1279
	/ttHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8[0]	0.215
	/TTWZ_TuneCP5_13TeV-madgraph-pythia8[2]	0.003884
	/TTWW_TuneCP5_13TeV-madgraph-pythia8[2]	0.0115
	/TTbb_4f_TTTo2L2Nu_TuneCP5-Powheg-Openloops-Pythia8[2]	0.04
	/TTbb_4f_TTToSemiLeptonic_TuneCP5-Powheg-Openloops-Pythia8[2]	0.62
WJets	/WJetsToLNu_HT-70To100_TuneCP5_13TeV-madgraphMLM-pythia8[2]	1321.16
	/WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8[2]	1335.7
	/WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8[2]	351.689
	/WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8[2]	47.1663
	/WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8[2]	11.4196
	/WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8[2]	5.12389
	/WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8[2]	1.18295
	/WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8[0]	0.0255202
Bosons	/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8[1]	0.2086
	/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8[2]	20657.0
	/DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8[2]	6198.0
	/EWKWplus2Jets_WToQQ_dipoleRecoilOn_TuneCP5_13TeV-madgraph-pythia8[2]	10.67
	/EWKWminus2Jets_WToQQ_dipoleRecoilOn_TuneCP5_13TeV-madgraph-pythia8[2]	10.67
	/EWKZ2Jets_ZToLL_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8[0]	6.22
	/EWKZ2Jets_ZToNuNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8[0]	10.72
	/EWKZ2Jets_ZToQQ_dipoleRecoilOn_TuneCP5_13TeV-madgraph-pythia8[2]	10.67
	/WWTo1L1Nu2Q_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8[2]	51.65
	/WWTo2L2Nu_TuneCP5_13TeV-powheg-pythia8[0]	12.178
	/WZTo1L1Nu2Q_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8[2]	49.997
	/WZTo1L3Nu_4f_TuneCP5_13TeV-amcatnloFXFX-pythia8[2]	3.054024
	/WZTo2Q2L_mllmin4p0_TuneCP5_13TeV-amcatnloFXFX-pythia8[2]	5.6
	/WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8[0]	4.42965
	/ZZTo2Q2L_mllmin4p0_TuneCP5_13TeV-amcatnloFXFX-pythia8[2]	3.28
	/ZZTo2L2Nu_TuneCP5_13TeV_powheg_pythia8[2]	0.564
	/ZZTo4L_M-1toInf_TuneCP5_13TeV_powheg_pythia8[2]	1.256
	/WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8[1]	0.1651
	/WZZ_TuneCP5_13TeV-amcatnlo-pythia8[1]	0.05565
	/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8[1]	0.01398
	/WWJJToLNuLNu_EWK_noTop_TuneCP5_13TeV-madgraph-pythia8[2]	0.284
	/WZJJ_EWK_InclusivePolarization_TuneCP5_13TeV_madgraph-madspin-pythia8[2]	0.01701
	/ZZJJTo4L_TuneCP5_13TeV-madgraph-pythia8[0]	0.00884
EWKWLep	/EWKWPlus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8[0]	39.33
	/EWKWMinus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8[0]	32.26
VH	/VHToNonbb_M125_TuneCP5_13TeV-amcatnloFXFX_madspin_pythia8[0]	2.207
	/WminusH_HToBB_WToLNu_M-125_TuneCP5_13TeV-powheg-pythia8[2]	0.04901236122
	/WplusH_HToBB_WToLNu_M-125_TuneCP5_13TeV-powheg-pythia8[2]	0.08487599411
	/ZH_HToBB_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8[2]	0.02627486511
	/ggZH_HToBB_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8[2]	0.002461395396

[0] /RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v2/NANOAODSIM
[1] /RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1_ext1-v2/NANOAODSIM
[2] /RunIISummer20UL18NanoAODv9-106X_upgrade2018_realistic_v16_L1v1-v1/NANOAODSIM

Table 5: Background MC samples used in this analysis, corresponding to 2018 detector conditions (UL), with their respective cross sections in picobarns. The cross sections for the H_{T}-binned W+jets samples are scaled by a "stitching" factor such that the samples together fill a continuous H_{T} distribution.

3 Objects

3.1 Leptons

The lepton identification criteria for this analysis are shared with a complimentary effort targeting the same final state, but with both jets from the Higgs boson resolved as AK4 jets–rather than as a single AK8 “fat” jet. These criteria are primarily based on the central cut-based criteria developed by the E/Gamma and Muon Physics Object Groups (POGs), but some additional selections are made as detailed in Tables 6 and 7 for electrons and muons respectively. Leptons are furthermore considered as coming from the W boson in the $\mathrm{pp} \to \mathrm{W}^{\pm}\mathrm{H} + \mathrm{jj}$ final state if they pass the Tight lepton ID and have p_{T} greater than 40 GeV. Finally, tau leptons are not considered in this analysis for simplicity.

Electrons			
Observable	Condition	Veto	Tight
p_T	–	> 10 GeV	> 35 GeV
E/Gamma POG cut-based ID	–	Veto	Medium
$\lvert\eta\rvert$	–	–	< 2.5
$\lvert dz\rvert$	$\lvert\eta\rvert \geq 1.429$	–	< 0.2
$\lvert dz\rvert$	$\lvert\eta\rvert < 1.429$	–	< 0.1
$\lvert dxy\rvert$	$\lvert\eta\rvert \geq 1.429$	–	< 0.1
$\lvert dxy\rvert$	$\lvert\eta\rvert < 1.429$	–	< 0.05

Table 6: Veto and tight selection criteria for identifying reconstructed electrons. An emdash (–) indicates selection criteria that are not applied.

Muons		
Observable	Veto	Tight
p_T	> 10 GeV	> 26 GeV
Muon POG cut-based ID	Tight	Tight
Particle Flow relative isolation	0.4	0.15
$\lvert\eta\rvert$	–	< 2.4

Table 7: Veto and tight selection criteria for identifying reconstructed muons. An emdash (–) indicates selection criteria that are not applied.

3.2 Jets and missing E_{T}

AK4 jets and AK8 “fat” jets reconstructed from Particle Flow (PF) candidates using anti-$\mathrm{k_T}$ algorithm with a cone of $\Delta R < 0.4$ and $\Delta R < 0.8$ respectively are considered in this analysis. In particular, a single AK8 fat jet is selected as the $\mathrm{H} \to \mathrm{b\overline{b}}$ candidate, since the Higgs boson is expected to be significantly boosted when $\lambda_{\mathrm{WZ}} = -1$, and two AK4 jets are selected as the VBS quark candidates. The jet energy corrections (JECs) applied to AK4 and AK8 jets as well as the prescriptions used for AK4 jet energy resolution smearing–such that the resolution in MC better matches what is measured in data–are tabulated in Table 8. Importantly, jets that overlap with any lepton passing the Veto lepton ID are not considered. This overlap is computed for AK4 jets in the NanoAOD processing, where two objects are considered as overlapping if they share a packed PF candidate. For AK8 jets, a lepton is considered to be overlapping if $\Delta R(\ell, \mathrm{jet}) < 0.8$. The type-1 PF-MET is used in the analysis, and jet energy corrections (JECs) are propagated to $E_{\mathrm{T}}^{\mathrm{miss}}$. Next, those AK4 and AK8 jets that remain are further processed using the selections detailed below.

Year	JEC era	JER era
2016 (pre-VFP)	Summer19UL16APV_V7_MC Summer19UL16APV_RunBCDEF_V7_DATA	Summer20UL16APV_JRV3_MC
2016 (post-VFP)	Summer19UL16_V7_MC Summer19UL16_RunBCDEFGH_Combined_V7_DATA	Summer20UL16_JRV3_MC
2017	Summer19UL17_V5_MC Summer19UL17_Run[BCDEF]_V5_DATA	Summer19UL17_JRV2_MC
2018	Summer19UL18_V5_MC Summer19UL18_Run[ABCD]_V5_DATA	Summer19UL18_JRV2_MC

Table 8: Jet energy corrections (JECs) and applied to AK4 and AK8 jets for data and MC are sorted by "era," corresponding to a year of NanoAOD UL processing. The equivalent era for jet energy resolution corrections are also tabulated. Brackets indicate that there are separate eras for each character in the brackets.

3.2.1 AK8 jets

AK8 fat jets used in this analysis must firstly have $p_{\mathrm{T}} > 300\,\mathrm{GeV}$ and be within the tracker acceptance ($|\eta| < 2.5$). In addition, they must have $M > 50\,\mathrm{GeV}$ and $M_{\mathrm{SD}} > 40\,\mathrm{GeV}$, where M is the invariant mass of the fat jet and M_{SD} is the soft drop mass corrected with PUPPI.

3.2.2 AK8 jet originating from b quarks

Since the only AK8 fat jet of interest to this analysis originates from $\mathrm{H} \to \mathrm{b\overline{b}}$, just one AK8 candidate is ultimately selected. A graph neural network (GNN) referred to as "ParticleNet" [9] is used to tag fat jets as having come from a b quark jet and anti-b quark jet merged into a single fat jet. Specifically, ParticleNet yields an "Xbb" discriminant for each fat jet, and the fat jet with the highest Xbb score is selected as the $\mathrm{H} \to \mathrm{b\overline{b}}$ candidate. In NanoAOD, this discriminant is stored as follows:

$$\frac{\text{FatJet\_particleNetMD\_Xbb}}{\text{FatJet\_particleNetMD\_Xbb} + \text{FatJet\_particleNetMD\_QCD}} \tag{3}$$

3.2.3 AK4 jets

First, only only jets with $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ are considered. Then, any jet that overlaps with the $\mathrm{H} \to \mathrm{b\overline{b}}$ AK8 fat jet candidate are ignored, where an AK4 and AK8 jet are considered to be overlapping if $\Delta R(\text{jet}, \text{fat jet}) < 0.8$. The remaining jets are further cleaned using the Tight Lep-Veto PF jet ID detailed in Tables 9 and 10 following the JetMET (JME) POG recommendation.

3.2.4 AK4 jets originating from VBS quarks

The event must have at least two AK4 jets that pass the aforementioned selections and also have $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\Delta\eta_{\mathrm{jj}}| < 4.7$ to be considered in this analysis. If there are exactly two such jets, they are taken as the VBS quark candidates. However, if there are more than two, then the following prescription is followed. First, the jets are split by whether they are located in the positive or negative η hemisphere, then sorted by the magnitude of the three-momenta. If all of the jets are in one η hemisphere, then the leading and trailing jets in that hemisphere are taken as the VBS candidates. Otherwise, the leading jet from each η hemisphere is selected.

3.2.5 AK4 jets originating from b quarks

The signal process for this analysis has no b quarks resolved as AK4 jets in the final state, except in the subleading case where one of the VBS quarks is a b quark, as the $\mathrm{H} \to \mathrm{b\overline{b}}$ candidate is

2016APV/2016 UL (106X) AK4 Jets				
Variable	$\|\eta\| \leq 2.4$	$2.4 < \|\eta\| \leq 2.7$	$2.7 < \|\eta\| \leq 3.0$	$3.0 < \|\eta\| \leq 5.0$
Neutral Hadron Fraction	< 0.90	< 0.90	< 0.90	> 0.2
Neutral EM Fraction	< 0.90	< 0.99	> 0 and < 0.99	< 0.9
Number of Constituents	> 1	–	–	–
Muon Fraction	$< 0.80^{\dagger}$	–	–	–
Charged Hadron Fraction	> 0	–	–	–
Charged Multiplicity	> 0	–	–	–
Charged EM Fraction	$< 0.80^{\dagger}$	–	–	–
Number of Neutral Particles	–	–	> 1	> 10

† For analyses that veto jets based on lepton overlap, referred to as "LepVeto" by the JetMET POG.

Table 9: Tight JetMET POG PF jet ID criteria for AK4 jets in 2016 post-VFP and pre-VFP NanoAOD UL samples. An emdash (–) indicates that no selection is applied.

2017/2018 UL (106X) AK4 Jets				
Variable	$\|\eta\| \leq 2.6$	$2.6 < \|\eta\| \leq 2.7$	$2.7 < \|\eta\| \leq 3.0$	$3.0 < \|\eta\| \leq 5.0$
Neutral Hadron Fraction	< 0.90	< 0.90	–	> 0.2
Neutral EM Fraction	< 0.90	< 0.99	> 0.01 and < 0.99	< 0.9
Number of Constituents	> 1	–	–	–
Muon Fraction	$< 0.80^{\dagger}$	$< 0.80^{\dagger}$	–	–
Charged Hadron Fraction	> 0	–	–	–
Charged Multiplicity	> 0	> 0	–	–
Charged EM Fraction	$< 0.80^{\dagger}$	$< 0.80^{\dagger}$	–	–
Number of Neutral Particles	–	–	> 1	> 10

† For analyses that veto jets based on lepton overlap, referred to as "LepVeto" by the JetMET POG.

Table 10: Tight JetMET POG PF jet ID criteria for AK4 jets in 2017 and 2018 NanoAOD UL samples. An emdash (–) indicates that no selection is applied.

expected to be resolved as an AK8 fat jet. However, the main backgrounds for this analysis do have at least one b quark in the final state. Thus, any event with a "b-tagged" jet can be vetoed. In particular, a deep neural network referred to as "DeepJet" is used to tag jets as having come from a b quark, using the Medium working point. Only jets within the tracker acceptance ($|\eta| < 2.4$) are considered for this tagging.

4 Selections

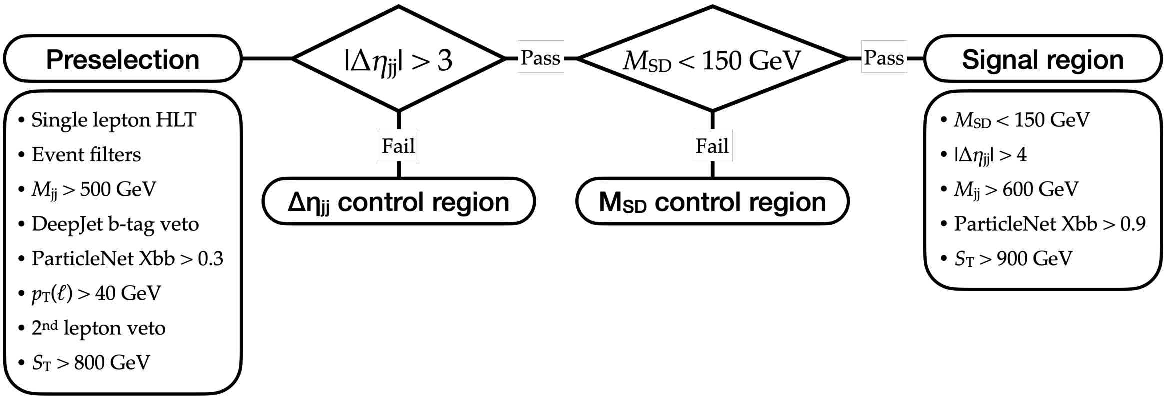


Figure 8: Summary of selections applied in this analysis.

4.1 Triggers and data quality cuts

To start, the single lepton High Level Triggers (HLTs) tabulated in Table 11 are applied. Notably, since there is ultimately a selection of $p_T > 40\,\text{GeV}$ applied to the leptons in the event, this analysis should sit on the plateau of the trigger efficiency curve. For collision data, events from the SingleMuon dataset are required to pass the single muon triggers, while events from the SingleElectron dataset and required to pass the single electron triggers and fail the single muon triggers. Since the trigger information is available in MC as well as data, the simulated events are required to pass the same triggers as the actual collision data. In particular, the single electron and single muon triggers are both applied, and the event is rejected only if both sets of triggers fail. A centrally derived scale factor is applied in MC such that the efficiency of the triggers in MC matches that measured in data. Finally, the following event filters recommended by the JME are applied in data in order to remove detector noise and unphysical events [10]:

- Flag_goodVertices
- Flag_HBHENoiseFilter
- Flag_HBHENoiseIsoFilter
- Flag_EcalDeadCellTriggerPrimitiveFilter
- Flag_BadPFMuonFilter
- Flag_BadPFMuonDzFilter
- Flag_hfNoisyHitsFilter
- Flag_eeBadScFilter
- Flag_ecalBadCalibFilter (2016 only)
- Flag_globalSuperTightHalo2016Filter

The last filter is applied to specifically remove beam-halo events. These filters are also applied to MC events, but have no effect.

4.2 Preselection

In addition to the triggers and JME-recommended event filters, a set of basic selections referred to as the "Preselection" are applied. These cuts are common to the signal regions and control regions defined later in this text. That is, the Preselection is applied to all events used in this analysis. In general, it selects events with one lepton, two AK4 jets, and one AK8 jet passing

Year	Single lepton HLT path
2016	HLT_IsoMu24 HLT_IsoTkMu24 HLT_Ele27_WPTight_Gsf
2017	HLT_IsoMu27 HLT_Ele32_WPTight_Gsf_L1DoubleEG
2018	HLT_IsoMu24 HLT_Ele32_WPTight_Gsf

Table 11: Triggers used to filter events for this analysis.

the object selections defined in the previous section. It also applies more stringent selections on each object in order to narrow in on a signal-like phase space.

First, a loose selection is made on the combined VBS jet invariant mass: $M_{\mathrm{jj}} > 500\,\mathrm{GeV}$. Next, the ParticleNet Xbb score of the $\mathrm{H} \rightarrow \mathrm{b\bar{b}}$ fat jet candidate is required to be greater than 0.3. The event is also required to have no AK4 jets passing the Medium DeepJet working point, as mentioned in Section 3.2.5. The event must furthermore have one and only one lepton with $p_{\mathrm{T}} > 40\,\mathrm{GeV}$ that passes the Tight lepton ID. If there are any additional leptons that pass the Veto lepton ID, the event is vetoed–these leptons are not required pass the p_{T} threshold. Finally, the event must have $S_{\mathrm{T}} > 800\,\mathrm{GeV}$ (where S_{T} is defined in Eq. 5).

4.3 Signal signature

4.3.1 VBS signature

The VBS signature of $\mathrm{pp} \rightarrow \mathrm{W^{\pm}H} + \mathrm{jj}$ provides a distinct kinematic signature, namely two nearly back-to-back jets–i.e. a large absolute difference in pseudorapidity $|\Delta\eta_{\mathrm{jj}}|$–with a large combined invariant mass M_{jj}. In particular, the background processes fall off exponentially in M_{jj} whereas the signal process is more flatly distributed. Combined with the fact that the signal has a distinctly larger average value of $|\Delta\eta_{\mathrm{jj}}|$ than background, these VBS characteristics form a strong handle for distinguishing signal from background. These variables are plotted after reconstruction, and after applying the Preselection, in Fig. 9.

4.3.2 $\mathrm{W^{\pm}H}$ signature

Because they are a proxy for the hard scattering energy of the event, the following variables provide access to the enhancement of M_{WH} from $\lambda_{\mathrm{WZ}} = -1$:

$$L_{\mathrm{T}} = p_{\mathrm{T}}(\ell) + E_{\mathrm{T}}^{\mathrm{miss}} \quad (4)$$

$$S_{\mathrm{T}} = L_{\mathrm{T}} + p_{\mathrm{T}}(\mathrm{H} \rightarrow \mathrm{b\bar{b}}\ \text{fat jet}) \quad (5)$$

where L_{T} is the scalar sum of "leptonic" transverse energy, and S_{T} is the scalar sum of the transverse energy of the $\mathrm{W^{\pm}H}$ decay products. As mentioned in Section 1, the cross section for signal events with large M_{WH} is significantly increased when $\lambda_{\mathrm{WZ}} = -1$, which is seen in the distribution of L_{T} and S_{T}. In particular, the background falls off exponentially in these variables, whereas there are a significant number of signal events in the high L_{T} and S_{T} tails. These variables are plotted after reconstruction in Fig. 10, where it is clear that they provide powerful signal-versus-background discrimination, even after already applying the Preselection detailed in Section 4.2 below.

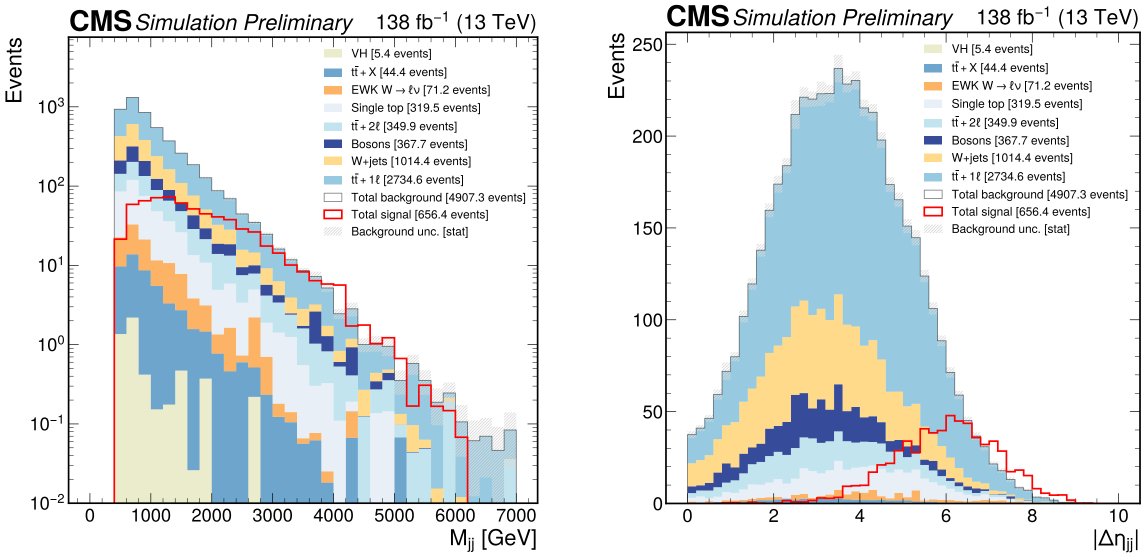


Figure 9: The VBS jet combined invariant mass (left) and absolute difference in pseudorapidity (right) are plotted after applying the Preselection.

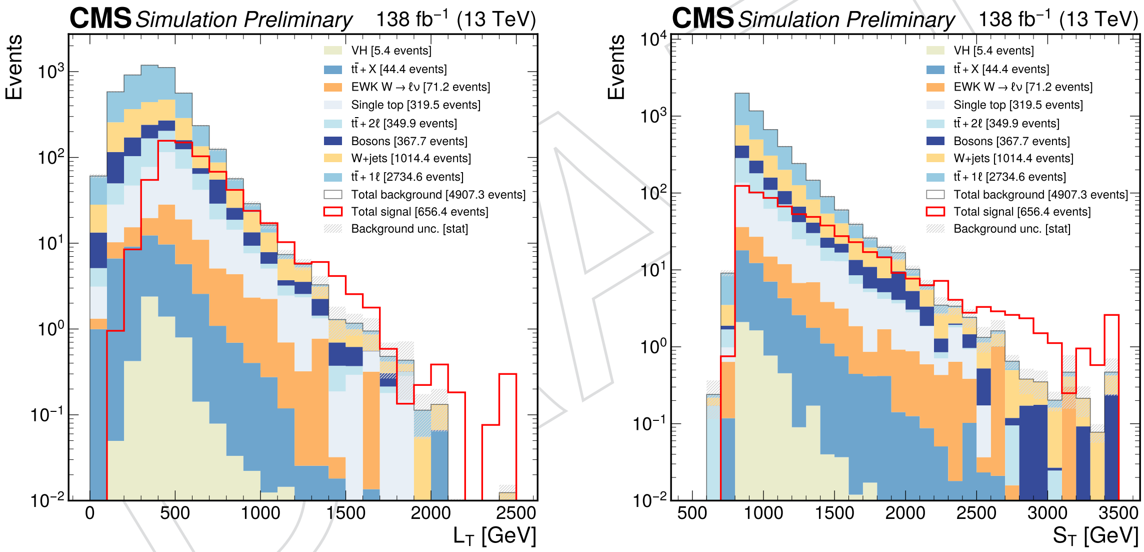


Figure 10: The variables L_{T} (left) and S_{T} (right) are plotted after applying the Preselection.

In addition, the softdrop mass of the $\mathrm{H} \to \mathrm{b\bar{b}}$ candidate fat jet shows a distinct peak around the Higgs mass and is highly concentrated in the high ParticleNet Xbb score (Fig. 11).

4.4 Signal region

The signal region for this analysis is defined on top of the Preselection with similar, but tighter selections. First, the S_{T} threshold is increased to $S_{\mathrm{T}} > 900\,\mathrm{GeV}$. Then, the selections on the VBS jet variables are tightened to $M_{\mathrm{jj}} > 600\,\mathrm{GeV}$ and $|\Delta\eta_{\mathrm{jj}}| > 4$. Finally, the selections on the $\mathrm{H} \to \mathrm{b\bar{b}}$ fat jet candidate are made much more strict, where the ParticleNet Xbb score is required to be greater than 0.9 and M_{SD} is required to be less than 150 GeV.

The background in this region is estimated using a data-driven technique, as described in Section 5. Moreover, the selections described above were specifically designed such that this could

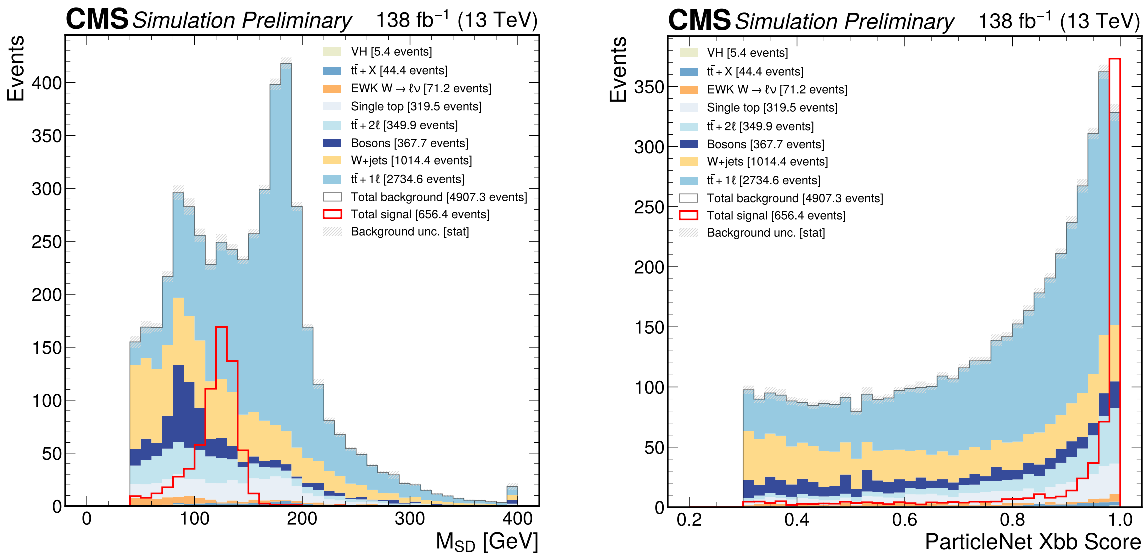


Figure 11: The $H \to b\bar{b}$ fat jet candidate soft drop mass (left) and ParticleNet Xbb score (right) are plotted after applying the Preselection.

be done at all. Looser cuts are preferred in particular, as it was found that the "arms" of the background extrapolation become correlated in a more restricted phase space. Therefore, the signal region selections are not optimized for maximal purity, though such a region can be formed (see Appendix E).

The Monte Carlo yields of both signal and background after applying the Preselection and signal region selections are each tabulated in Table 12.

Cut	VH	EWK W	Bosons	W+jets	Single top	$t\bar{t} + X$	$t\bar{t} + 1\ell$	$t\bar{t} + 2\ell$
Preselection	5.37	71.20	367.67	1014.45	319.55	44.45	2734.64	349.92
SR	0.85	2.96	5.98	21.21	13.05	0.95	49.90	21.43

Cut	Total bkg.	VBS $W^{\pm}H$ ($\lambda_{WZ} = -1$)
Preselection	4907.25	656.39
SR	116.33	397.44

Table 12: The the background yields separated by sample (top) as well as the total background and signal yields (bottom) are tabulated after applying the Preselection followed by the signal region (SR) selections. All yields are taken from Monte Carlo and weighted by cross section to the total integrated luminosity of Run 2 (138 fb^{-1}).

4.5 Control regions

Two regions orthogonal to the signal region are defined as "control" regions. Both are signal depleted, so data and Monte Carlo can be compared in order to test the validity of the simulation in otherwise relevant phase space. In particular, these regions are used only to ensure that there is no significant mismodeling or missing background Monte Carlo. The first is a control region in $\Delta\eta_{jj}$ that consists of the Preselection and $|\Delta\eta_{jj}| \leq 3$. The second is a control region in M_{SD} that consists of the Preselection, $|\Delta\eta_{jj}| > 3$, and $M_{SD} \geq 150$ GeV. See Fig. 8 for a pictorial representation of these regions. In Figures 12 and 13, it can be seen that there is satisfactory agreement in both regions between data and MC for variables relevant to this analysis. Additional control region plots can be found in Appendix B.1.

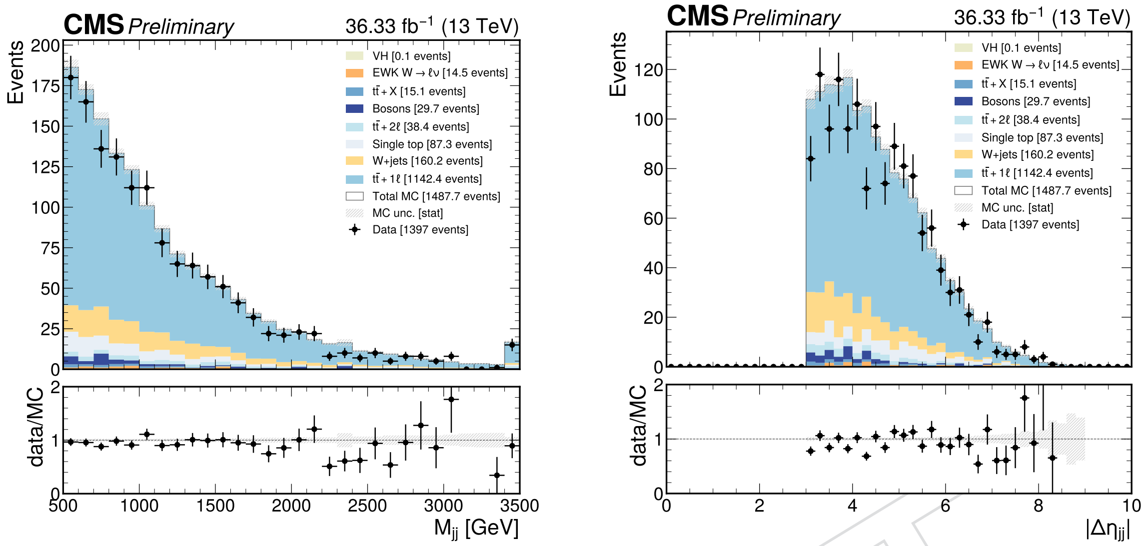


Figure 12: The VBS jet combined invariant mass (left) and absolute difference in pseudorapidity (right) are plotted in the M_{SD} control region.

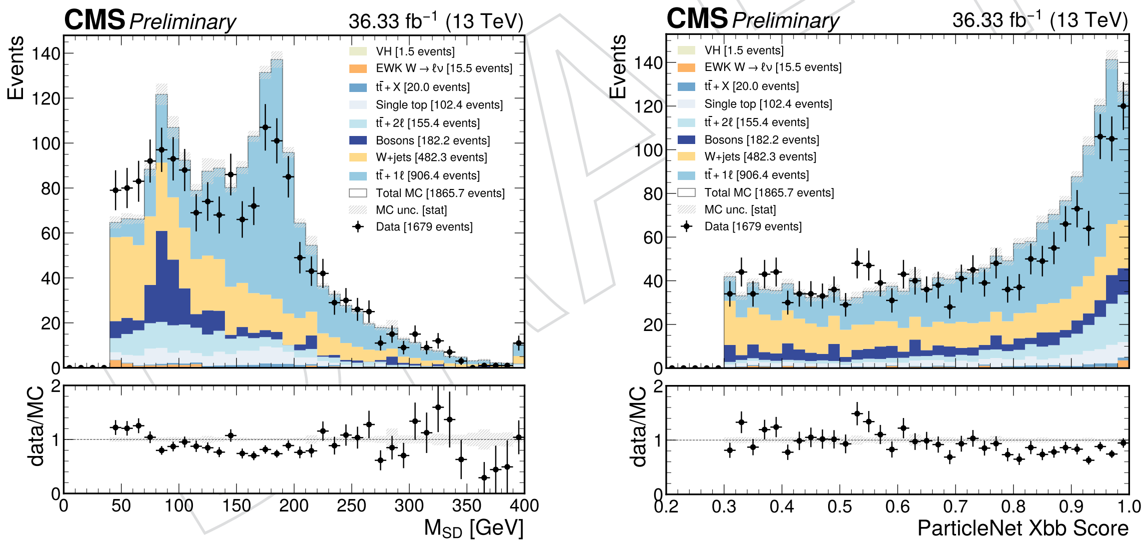


Figure 13: The $H \to b\bar{b}$ fat jet candidate soft drop mass (left) and ParticleNet Xbb score (right) are plotted in the $\Delta\eta_{\text{jj}}$ control region.

5 Background estimation

The background in the signal region is estimated using the "ABCD" method, where regions A, B, C, and D are illustrated in Fig. 14. First, let the background yield in regions A, B, C, and D in Monte Carlo be defined as A_{MC}, B_{MC}, C_{MC}, and D_{MC}. Likewise, let the the same yields in data be defined as A_{data}, B_{data}, C_{data}, and D_{data}. Under these definitions, the estimated background yield in Region D, which will be referred to as D_{data}^{pred}, can be computed with data as follows:

$$D_{data}^{pred} = C_{data} \times \frac{A_{data}}{B_{data}} \tag{6}$$

where the same can be done in MC, yielding D_{MC}^{pred}. First, it can be seen in Fig. 15 that data and MC agree reasonably well in regions A, B, and C which have already been unblinded since they have nearly zero contribution from signal by design. More plots comparing data and MC can be found in Appendix B.2. In addition, it has been verified that the "transfer factor" that scales the actual yield in Region C to the estimated yield in region D is consistent within statistical uncertainty across data and MC:

$$\frac{A_{MC}}{B_{MC}} = 0.71 \pm 3.1\% \qquad \frac{A_{data}}{B_{data}} = 0.71 \pm 11.0\%$$

The closure of the ABCD method described here is tested by comparing D_{MC}^{pred} to D_{MC}. This checks how closely the estimation in Monte Carlo predicts the actual yield in Monte Carlo.

$$D_{MC}^{pred} = C_{MC} \times \frac{A_{MC}}{B_{MC}} = 129.4 \qquad D_{MC} = 116$$

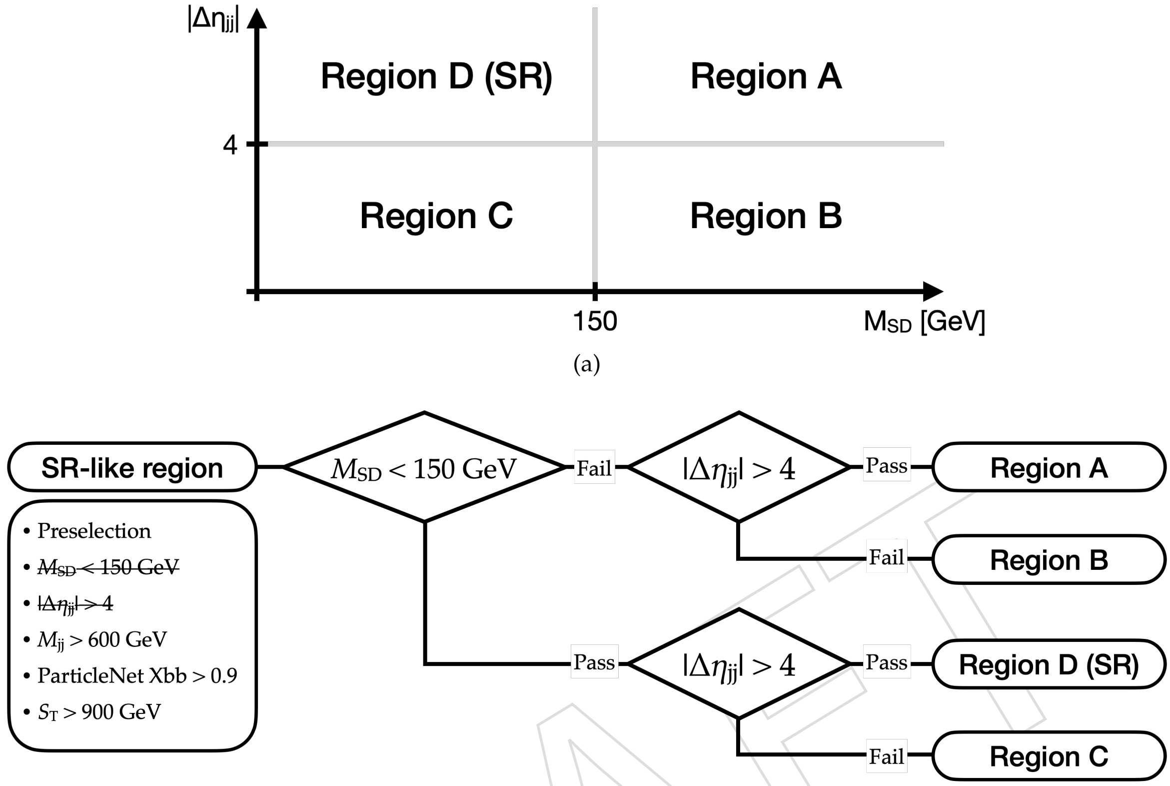
It is clear that the ABCD method for this analysis systematically over-predicts the background yield in Region D. The difference between the predicted and actual yield in MC is therefore taken as a systematic on this method. An additional 6.0% systematic is added to account for uncertainty in the W+jets background composition (see Appendix C). Thus, the systematic and statistical uncertainties ϵ_{syst} and ϵ_{stat} are

$$\epsilon_{syst} = \left| 1 - \frac{D_{MC}^{pred}}{D_{MC}} \right| \oplus 6.0\% \approx 12.7\%$$
$$\epsilon_{stat} = \frac{\sqrt{A_{data}}}{A_{data}} \oplus \frac{\sqrt{B_{data}}}{B_{data}} \oplus \frac{\sqrt{C_{data}}}{C_{data}} \approx 13.4\%$$

and the final estimated background yield in the signal region is therefore given by

$$D_{data}^{pred} = 120 \pm 16.1 \pm 15.3$$

Finally, the yields in Regions A, B, C, and D are tabulated in Table 13, where it can be seen that there is negligable signal pollution in Regions A, B, and C. In particular, while there are some signal events in Regions A and C, they are much smaller than the signal yield in Region D, the signal region.



(b)

Figure 14: The ABCD configuration for estimation of background in the signal region is shown graphically (a) and as a flowchart (b). The set of selections applied to all regions is referred to as the "SR-like" region in (b).

Region	Total bkg. (MC)	Total sig.	Total data
A	172.8 ± 3.2	12.2 ± 1.5	142 ± 11.9
B	241.7 ± 5.8	0.9 ± 0.4	201 ± 14.2
C	181.0 ± 4.4	16.7 ± 1.8	170 ± 13.0
D	116.3 ± 3.8	397.4 ± 8.7	–

Table 13: The signal and background MC yields are tabulated in Regions A, B, C, and D alongside the data yields in Regions A, B, and C.

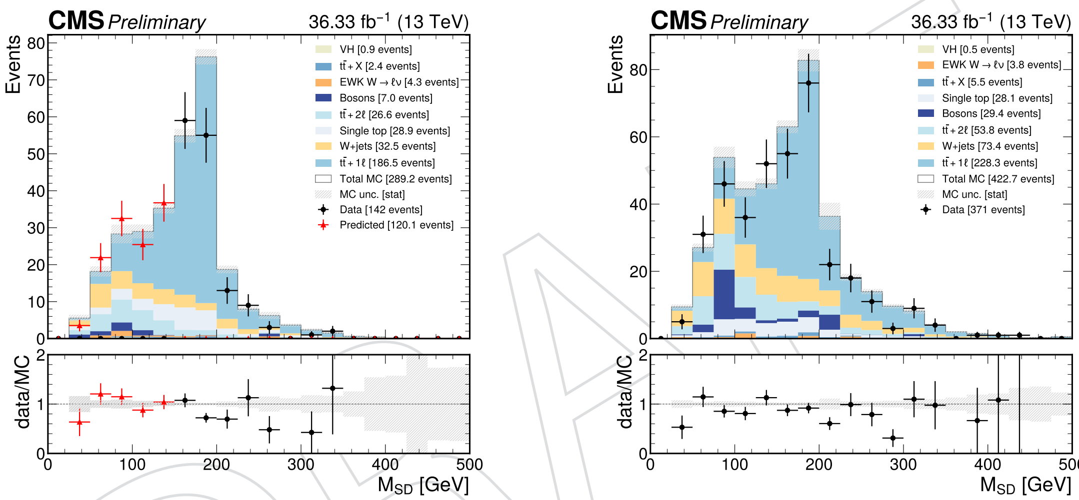


Figure 15: The softdrop mass of the H $\rightarrow b\bar{b}$ fat jet candidate is plotted in Regions A and D (left), which share the selection $|\Delta\eta_{jj}| > 0.4$, and Regions B and C (right), which share the selection $|\Delta\eta_{jj}| \leq 0.4$, showing closure of the ABCD method.

6 Systematic uncertainties

The systematics on the background yield in the signal region are already derived in Section 5. Since the estimation strategy is data-driven, the systematics on the Monte Carlo, which are more numerous, only need to be evaluated for the signal yield. Most sources of systematic uncertainty are derived by varying the renormalization and factorization scales (μ_R and μ_F), PDF, and various experimental corrections by one standard deviation and taking the maximal difference in yield as the error. In particular, the corrections and their uncertainties are typically derived centrally in order to augment the efficiency of a specific selection in MC to match that measured in data. In general, these corrections are applied as an event weight ω, such that the weighted contribution W of each raw Monte Carlo event is given by the product of the event weights for that same event. The yield in a given signal region y containing N raw Monte Carlo events is therefore given by

$$y = \sum_{i=1}^{N} W_i \tag{7}$$

Then, the yield y_{var} is computed after applying a systematic variation (up or down) of each source of systematic uncertainty independently:

$$y_{var} = \sum_{i=1}^{N} W_i \times \frac{\omega_{var}}{\omega} \tag{8}$$

Finally, the maximum of the percent differences δ_{up} or δ_{down} are taken as the systematic uncertainty for that source, where

$$\delta_{var} = \left| 1 - \frac{y_{var}}{y} \right| \tag{9}$$

Although most of the systematic uncertainties in this analysis are centrally computed, and thus their values in Table 14 are derived as shown above, not all of them are. Those systematics that are privately computed or are otherwise derived following a different prescription from that described already are discussed in more detail below.

6.1 PDF systematic uncertainty

The signal Monte Carlo simulation was generated using a Hessian Parton Distribution Function (PDF). After generation, 100 eigenvectors of a covariance matrix are stored as event weights ω_i^{PDF}, where $i = 1, 2, ..., 100$ [11]. These variations are derived such that the systematic uncertainty is derived as follows. First, in a signal region with N events, the ratio R_i of the sum of the i^{th} PDF variation and the sum of the MADGRAPH generator weights ω^{gen} is computed:

$$R_i = \frac{\sum_{j=1}^{N} \omega_{i,j}^{PDF}}{\sum_{j=1}^{N} \omega_j^{gen}} \tag{10}$$

Next, the yield y_i^{var} for the i^{th} PDF variation is given by

$$y_i^{var} = \sum_{j=1}^{N} W_j \times \frac{\omega_i^{PDF}}{R_i} \tag{11}$$

This decouples the systematic on the PDF from the systematic on the cross section, which would otherwise be double counted. Finally, the systematic uncertainty on the PDF is given by

$$\delta_{PDF} = \left[\sum_{i=1}^{100} \left| 1 - \frac{y_i^{var}}{y} \right|^2 \right]^{\frac{1}{2}} \tag{12}$$

Type	Systematic	Value
Background	ABCD syst. unc.	12.7%
	Data stat. unc.	13.4%
Signal	PDF variations	2.2%
	μ_F scale	17.7%
	Parton shower ISR weights	0.2%
	Parton shower FSR weights	1.5%
	Pileup reweighting	0.1%
	Pileup jet ID	1.0%
	L1 pre-fire corrections	1.0%
	HLT scale factors	0.8%
	Simulation stat. unc.	2.2%
	Lepton scale factors	0.03% - 1.6%
	ParticleNet Xbb scale factors	1.0% - 1.9%
	DeepJet b-tagging scale factors	0.3%
	MET unc.	0.3%
	Jet energy scale	6.4%
	Jet energy resolution	0.6%
	Luminosity	1.6%
	$\mathrm{H} \rightarrow \mathrm{b\overline{b}}$ BR	1.27%

Table 14: The size of the systematic uncertainties on the signal and estimated background yield in the signal region are tabulated. The systematics tabulated here represent the full set of nuisance parameters used in the fitting procedure that produced the final result.

6.2 ParticleNet Xbb scale factor uncertainty

Corrections for the ParticleNet Xbb discriminator distribution shape are computed for signal using a central tool originally developed for the $\mathrm{H} \rightarrow \mathrm{c\overline{c}}$ analysis [12, 13]. This tool utilizes a Boosted Decision Tree (BDT), referred to as "sfBDT," that isolates the phase space populated by $\mathrm{H} \rightarrow \mathrm{b\overline{b}}$ jets in signal, then selects $\mathrm{g} \rightarrow \mathrm{b\overline{b}}$ jets in that phase space from Monte Carlo simulation of QCD multijet events. The sfBDT is therefore trained to select suitable $\mathrm{g} \rightarrow \mathrm{b\overline{b}}$ jets to serve as proxies for the $\mathrm{H} \rightarrow \mathrm{b\overline{b}}$ jets in signal, in particular by vetoing jets with a high gluon contamination rate. Moreover, the sfBDT receives as input the variables involving the basic kinematics of the subjets and secondary vertices associated with the jet. Thus trained, the sfBDT can also be used to select the same kind of proxy jets from data.

The signal-like jets selected by the sfBDT can then be used to measure the efficiency of the ParticleNet Xbb discriminant in QCD Monte Carlo and data. A "pass" and "fail" region is defined by jets passing or failing the ParticleNet Xbb tagging threshold–corresponding to the requirement of Xbb > 0.9 in the signal region for this analysis. In each region, a fit of the mass of the secondary vertex with the maximum impact parameter d_{xy} significance is performed in order to distinguish the contribution of b-type ($\mathrm{g} \rightarrow \mathrm{b\overline{b}}$), c-type, and light-type jets. Three scale factors, one for each type of jet, are allowed to float in this fit, where each scale factor is defined as follows:

$$SF_i = \frac{\epsilon_{\mathrm{data},i}}{\epsilon_{\mathrm{MC},i}} \tag{13}$$

where i stands for the jet category and ϵ is the efficiency of the ParticleNet tagger. This measurement is done in bins of p_{T} for a more robust correction, and it is repeated for each year of UL NanoAOD individually. The post-fit distributions of the secondary vertex mass variable for 2018 are shown in Fig. 16. Additional plots can be found in Appendix B.3

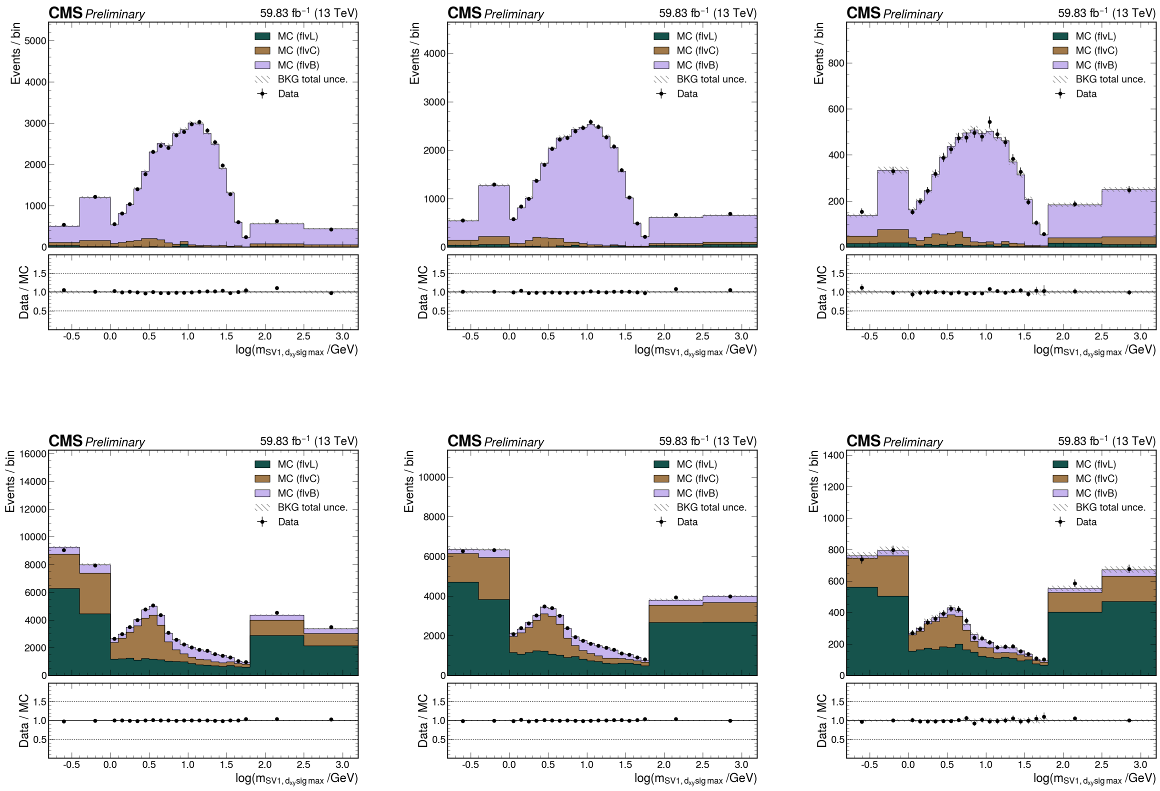


Figure 16: The post-fit distribution of the natural logarithm of the secondary vertex mass $m_{\mathrm{SV}_1,d_{xy}\,\mathrm{sig\,max}}$ is plotted in the "pass" (top) and "fail" (bottom) region for 2018 in the $[250, 500)$ GeV (left), $[500, 700)$ GeV (center), and $[700, \infty)$ GeV (right) p_{T} bins. The subscript "SV_1, d_{xy} sig max" stands for the secondary vertex with the maximum impact parameter d_{xy} significance, and the natural logarithm is taken to account for the variable's long tail due to limited resolution.

The scale factors are finally collected (Table 15) and applied to signal as an event weight to correct the efficiency of ParticleNet in Monte Carlo to match that in data, and the systematic uncertainty is computed using the upwards and downwards 1σ fluctuations of the event weights.

Year	p_{T} range in GeV:		
	[250, 500)	[500, 700)	[700, ∞)
2018	$0.990^{+0.027}_{-0.031}$	$1.040^{+0.038}_{-0.034}$	$1.069^{+0.056}_{-0.038}$
2017	$1.020^{+0.027}_{-0.025}$	$1.049^{+0.041}_{-0.031}$	$1.030^{+0.041}_{-0.030}$
2016 (post-VFP)	$1.028^{+0.046}_{-0.047}$	$1.090^{+0.104}_{-0.098}$	$1.045^{+0.102}_{-0.087}$
2016 (pre-VFP)	$1.038^{+0.116}_{-0.115}$	$1.084^{+0.137}_{-0.132}$	$1.027^{+0.145}_{-0.142}$

Table 15: Summary of the ParticleNet Xbb-tagging scale factors for the Xbb $>$ 0.9 working point.

7 Results

The background yield in the signal region estimated from data and the signal yield predicted by Monte Carlo simulation are tabulated in Table 16. Using these yields, and the systematics tabulated in Table 14, we perform a maximum-likelihood fit using the Combine statistical tool maintained by the Higgs Physics Analysis Group (PAG). The tool is, in particular, run with the following parameters:

```
combine -M MultiDimFit -d vbswh.root -m 125 -t -1 \
    --expectSignal=0 \
    --setParameters r_VBSWH_mKW=0 \
    --setParameterRanges r_VBSWH_mKW=0.0,2.0 \
    --saveNLL \
    --algo grid \
    --points 101 \
    --rMin 0 \
    --rMax 5 \
    --alignEdges 1
```

where `vbswh.root` is the datacard (Fig. 19) translated into a Combine workspace ROOT file. The fit is performed using an "observed" yield that is artificially set to be equal to the predicted background yield. Under this background-only hypothesis, we expect to exclude $\kappa_W = -1, \kappa_Z = +1$ at 9.0σ (Fig. 17).

A two-dimensional exclusion is also performed, where each κ_W, κ_Z point is processed by Combine using same the parameters as the single-point result above, such that each point is treated as a distinct signal model. The exclusions for a signal strength of 1 are thus derived and plotted on the z-axis of a two-dimensional (κ_W, κ_Z) histogram, which is then interpolated such that smooth exclusion contours can be obtained (Fig. 18). Additional details can be found in Appendix F.

Type	Yield	±	stat.	±	syst.
Signal	397	±	8.7	±	77.9
Background	120	±	16.1	±	15.3

Table 16: The background yield estimated from data and signal yield predicted by Monte Carlo simulation in the signal region are tabulated with their associated statistical and systematic uncertainties. The systematic uncertainty for signal quoted here is the sum of all of the independent systematics (percent errors) listed in Table 14 in quadrature multiplied by the total yield.

Type	Yield	±	stat.	±	syst.
Bkg. 2016 (pre-VFP)	16	±	6.6	±	7.5
Bkg. 2016 (post-VFP)	6	±	2.4	±	1.4
Bkg. 2017	36	±	8.5	±	2.3
Bkg. 2018	69	±	13.9	±	5.6

Table 17: The background yield for each year of UL NanoAOD estimated from data in the signal region is tabulated with its associated statistical and systematic uncertainties. The sum of the background yields tabulated here does not match the total yield in Table 16, because they are each individually estimated using only data from each data taking period.

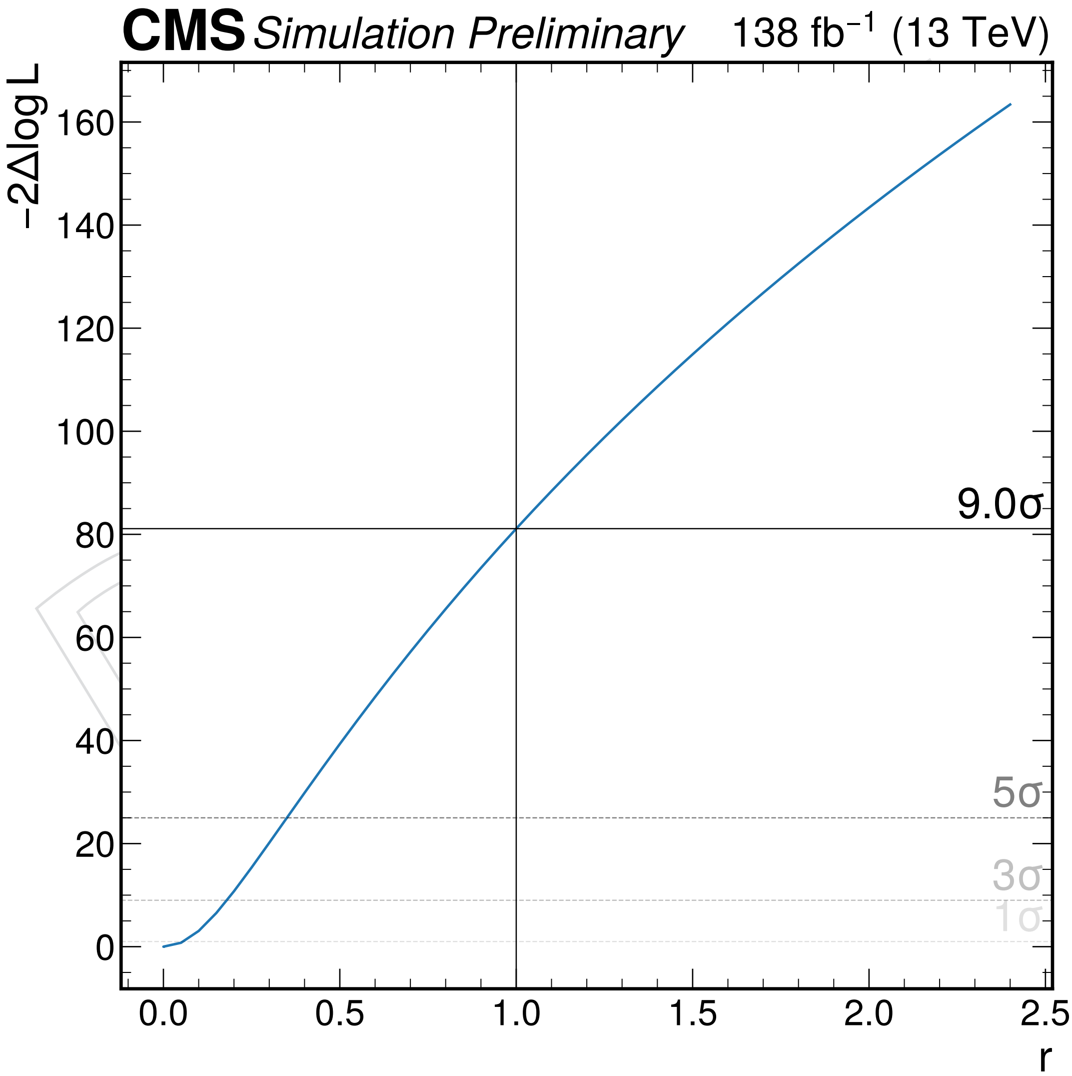


Figure 17: The maximum-likelihood fit for the background-only hypothesis, where the observed yield is artifically set to be equal to the predicted background yield, is plotted for the signal region. This shows a strong exclusion of $r = 1$ with a significance of 9.0σ.

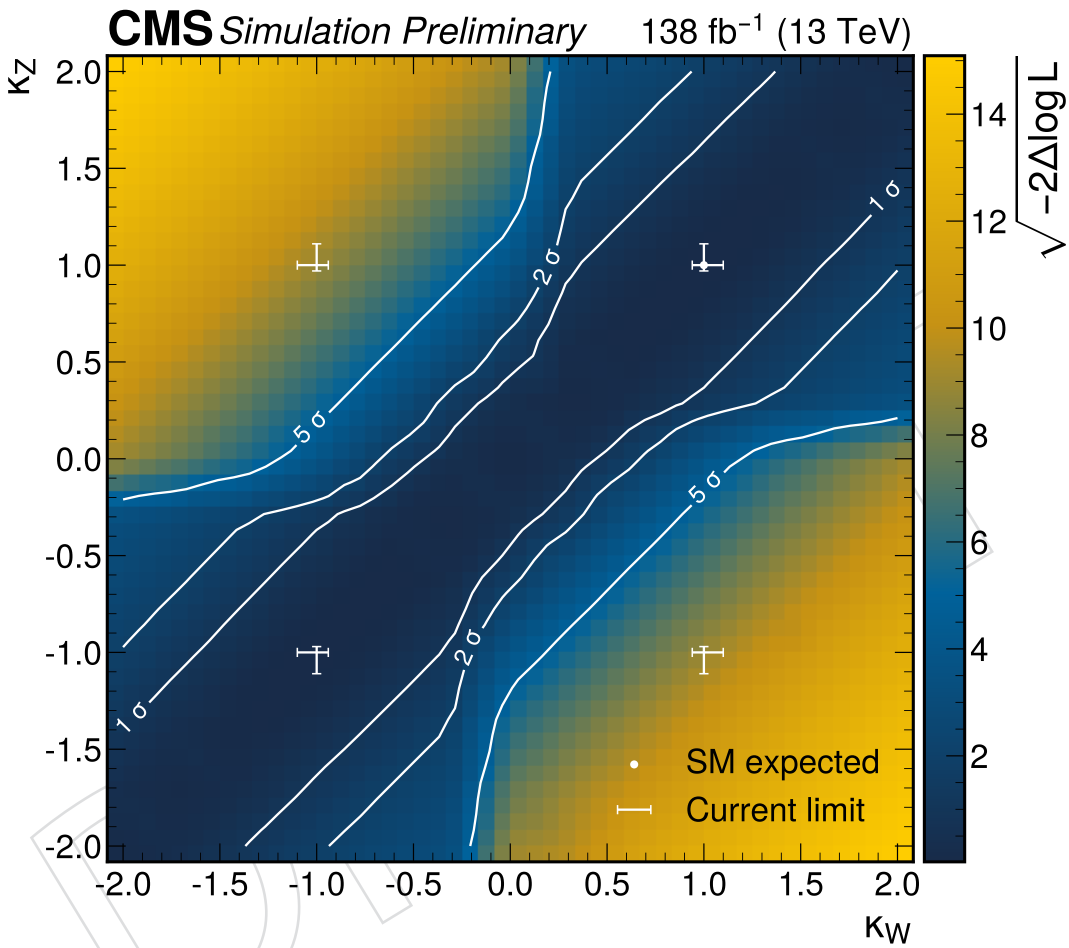


Figure 18: The interpolated exclusion of κ_W, κ_Z values with $\sigma = 1, 2, 5$ boundaries plotted as white contours. In addition, the current best limits ($|\kappa_W| = 1.02 \pm 0.08$, $|\kappa_Z| = 1.04 \pm 0.07$) are plotted as capped error bars. As was done for the single-point exclusion, the plot is made for the background-only hypothesis, where the observed yield is artificially set to be equal to the predicted background yield. Together with the exclusion contours, this plot clearly shows $\lambda_{WZ} < 0$ is excluded.

```
imax 1 number of channels
jmax 1 number of backgrounds
kmax 24 number of nuisance parameters
--------------------------------------------------
bin                                    bin1
observation                             120
--------------------------------------------------
bin                                    bin1         bin1
process                           VBSWH_mkW     TotalBkg
process                                   0            1
rate                                 397.44       120.10
--------------------------------------------------
abcd_syst               lnN               -       1.1270
abcd_stat               lnN               -       1.1340
pdf_vars                lnN          1.0215            -
muF_scale               lnN          1.1771            -
isr_weights             lnN          1.0019            -
fsr_weights             lnN          1.0153            -
pu_rwgt                 lnN          1.0012            -
puid_sf                 lnN          1.0100            -
L1_prefire              lnN          1.0097            -
hlt_sfs                 lnN          1.0079            -
mc_stat                 lnN          1.0218            -
lep_id                  lnN          1.0156            -
elec_reco               lnN          1.0031            -
muon_iso                lnN          1.0003            -
xbb_sfs_2016preVFP      lnN          1.0177            -
xbb_sfs_2016postVFP     lnN          1.0102            -
xbb_sfs_2017            lnN          1.0104            -
xbb_sfs_2018            lnN          1.0185            -
btag_sfs                lnN          1.0030            -
met_unc                 lnN          1.0029            -
jes                     lnN          1.0642            -
jer                     lnN          1.0059            -
lumi                    lnN          1.0160            -
hbb_br                  lnN          1.0127            -
```

Figure 19: Combine datacard used to produce the final result under the null hypothesis, where the observed count is artificially set to be equal to the predicted background yield.

8 Summary

An analysis has been performed as described in this note, searching for anomalous values of the Higgs boson couplings to W and Z bosons in the production of $W^{\pm}H$ via vector boson scattering. The work is based on a data sample of proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ recorded by the CMS experiment during 2016, 2017, and 2018, corresponding to a total integrated luminosity of $138\,\text{fb}^{-1}$. In particular, the search was done in the one lepton, two b quark, two jet final state where we ultimately exclude $\kappa_W = -1$ at 9.0σ. A two-dimensional exclusion is also presented, showing a wide exclusion of opposite-sign κ_W, κ_Z values. This, in conjunction with previous work, provides strong evidence that $\lambda_{WZ} \neq -1$.

References

[1] CMS Collaboration, "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC", *Phys. Lett. B* **716** (2012) 30–61, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.

[2] ATLAS Collaboration, "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett. B* **716** (2012) 1–29, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.

[3] CMS Collaboration, "A portrait of the Higgs boson by the CMS experiment ten years after the discovery", *Nature* **607** (Jul, 2022) 60–68, doi:10.1038/s41586-022-04892-x.

[4] ATLAS Collaboration, "A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery", *Nature* **607** (Jul, 2022) 52–59, doi:10.1038/s41586-022-04893-w.

[5] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties", doi:10.5170/CERN-2013-004, arXiv:1307.1347.

[6] D. Stolarski and Y. Wu, "Tree-level interference in vector boson fusion production of Vh", *Phys. Rev. D* **102** (Aug, 2020) 033006, doi:10.1103/PhysRevD.102.033006.

[7] J. Alwall et al., "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations", *Journal of High Energy Physics* **2014** (Jul, 2014) 79, doi:10.1007/JHEP07(2014)079.

[8] CMS Collaboration, "Extraction and validation of a new set of cms pythia8 tunes from underlying-event measurements", *The European Physical Journal C* **80** (Jan, 2020) 4, doi:10.1140/epjc/s10052-019-7499-4.

[9] H. Qu and L. Gouskos, "Jet tagging via particle clouds", *Phys. Rev. D* **101** (Mar, 2020) 056019, doi:10.1103/PhysRevD.101.056019.

[10] CMS Collaboration, "Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector", *Journal of Instrumentation* **14** (jul, 2019) P07004, doi:10.1088/1748-0221/14/07/P07004.

[11] J. Butterworth et al., "PDF4LHC recommendations for LHC Run II", *Journal of Physics G: Nuclear and Particle Physics* **43** (jan, 2016) 023001, doi:10.1088/0954-3899/43/2/023001.

[12] P. Bortignon et al., "Calibration of the ParticleNet boosted jet flavour tagger in full Run 2 data", CMS Analysis Note 2021/005, 2021.

[13] CMS Collaboration, "Calibration of the mass-decorrelated ParticleNet tagger for boosted $b\bar{b}$ and $c\bar{c}$ jets using LHC Run 2 data", CMS Detector Performance Note 2022/005, 2022.

A Signal MADGRAPH model

The MADGRAPH model representing $pp \to W^{\pm}H + jj$ was created by making the following modification to the Standard Model:

```
GC_72 = Coupling(name = 'GC_72',
                 value = '-((ee**2*complex(0,1)*vev)/(2.*sw**2))',
                 order = {'QED':1})
```

where `GC_72` represents κ_W. In particular, a minus sign was placed in front of the value of the `value` key word argument. This sets the value of κ_W to be the opposite of κ_Z, thus enforcing $\lambda_{WZ} = -1$. A MC sample using this model was generated as follows:

```
import model sm_mkw

define w = w+ w-

generate p p > w h j j QCD=0
```

where the model `sm_mkw` has the aforementioned modification applied. Finally, PYTHIA was used to handle parton showers with the same settings used to generated other CMS MC samples at LO.

419 B Additional plots

420 B.1 Control regions

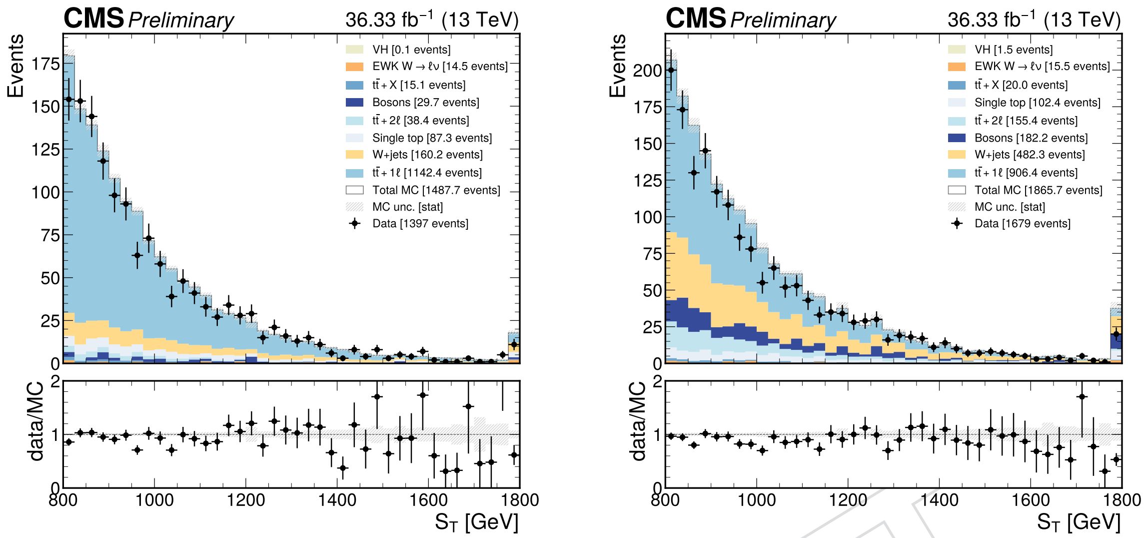


Figure 20: S_T is plotted in the M_{SD} (left) and $\Delta\eta_{jj}$ (right) control regions.

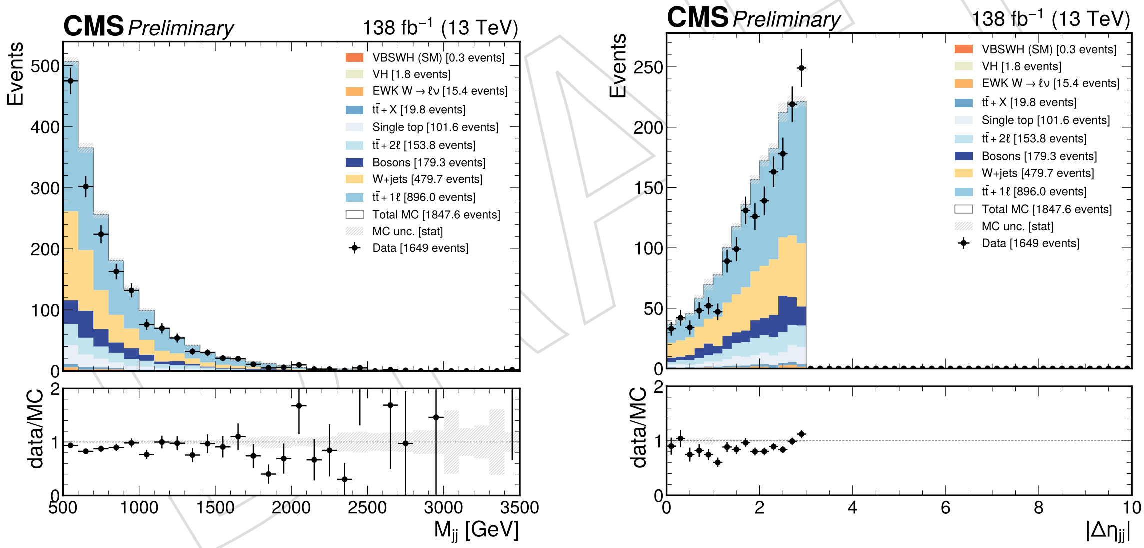


Figure 21: The VBS jet combined invariant mass (left) and absolute difference in pseudorapidity (right) are plotted in the $\Delta\eta_{jj}$ control region.

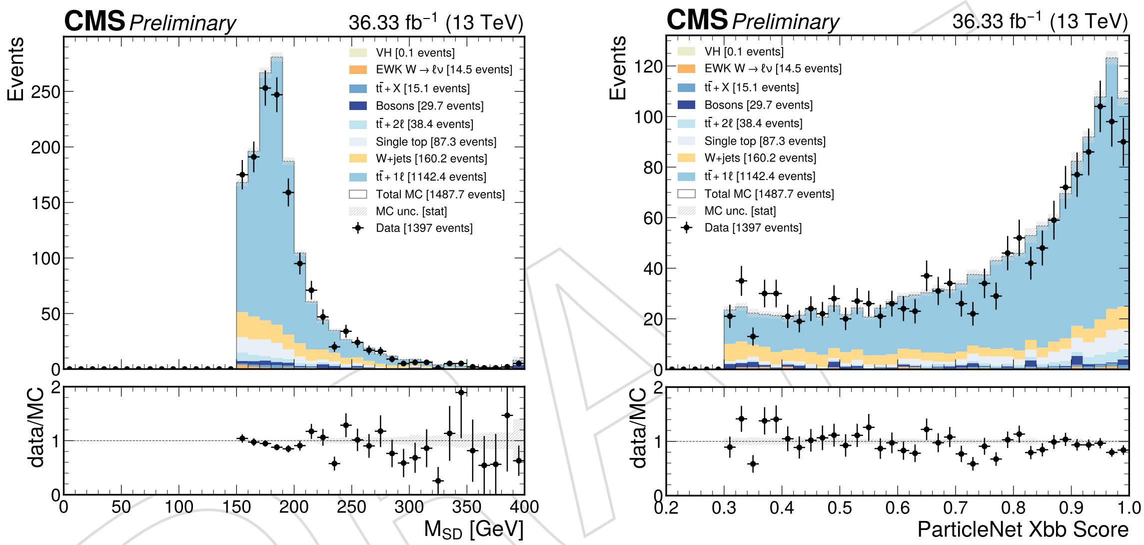


Figure 22: The H $\rightarrow$ b$\overline{\text{b}}$ fat jet candidate soft drop mass (left) and ParticleNet Xbb score (right) are plotted in the M_{SD} control region.

B.2 Background estimation

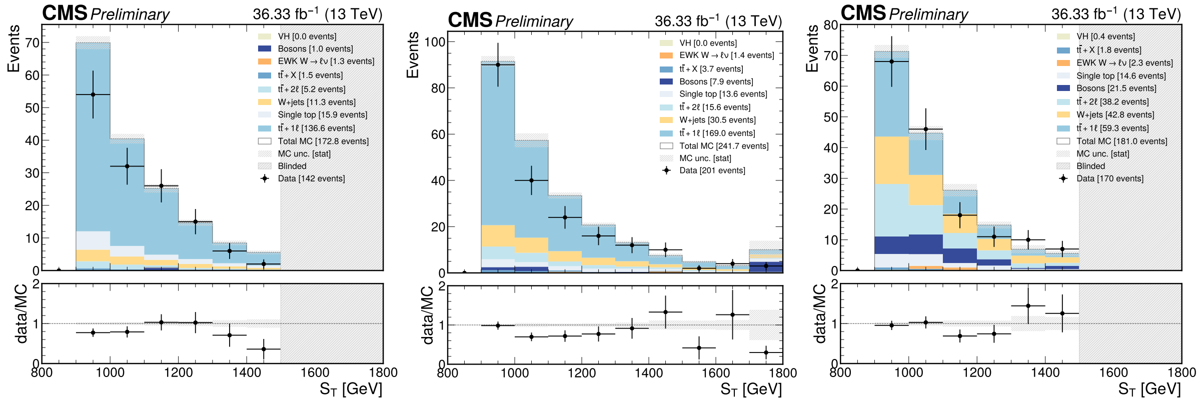


Figure 23: The variable S_T is plotted in regions A (left), B (center), and C (right) in both data and Monte Carlo, showing fair agreement in each region. The signal contamination is negligible in the total yield, but concentrated at high S_T, where there is little background. As such, the high S_T tail in Regions A and C are blinded. The same tail in Region B has little signal, so it has been unblinded.

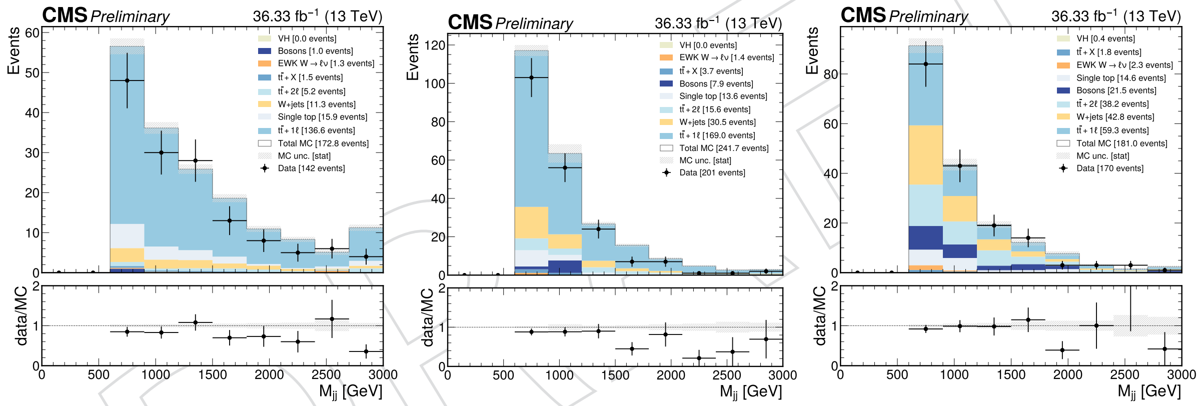


Figure 24: The VBS jet combined invariant mass is plotted in Regions A (left), B (center), and C (right) in both data and Monte Carlo, showing fair agreement in each region.

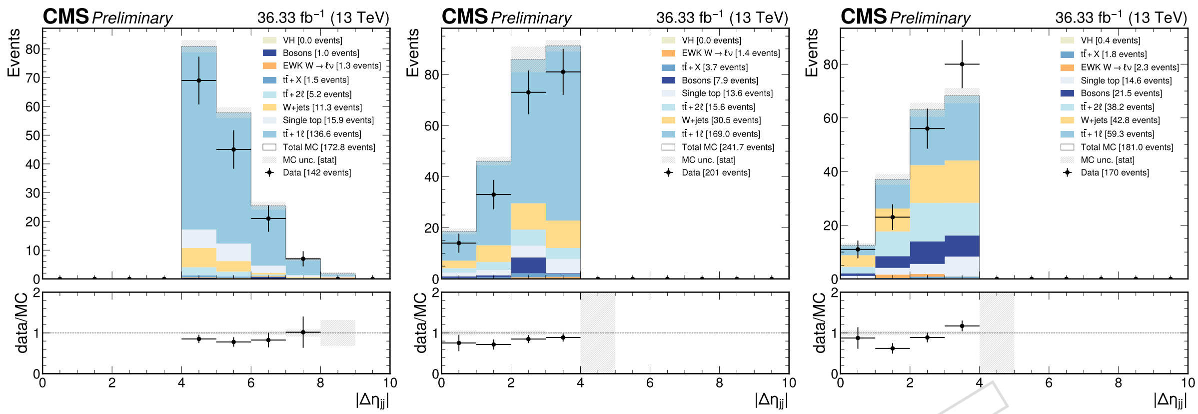


Figure 25: The VBS jet absolute difference in pseudorapidity is plotted in Regions A (left), B (center), and C (right) in both data and Monte Carlo, showing fair agreement in each region.

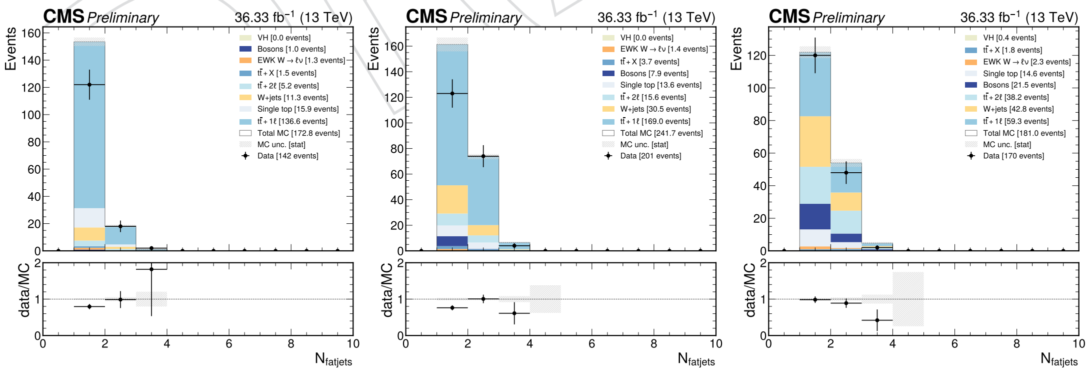


Figure 26: The number of AK8 fat jets passing the selections detailed in Section 3.2.1 is plotted in Regions A (left), B (center), and C (right) in both data and Monte Carlo, showing fair agreement in each region.

B.3 ParticleNet Xbb scale factor uncertainty

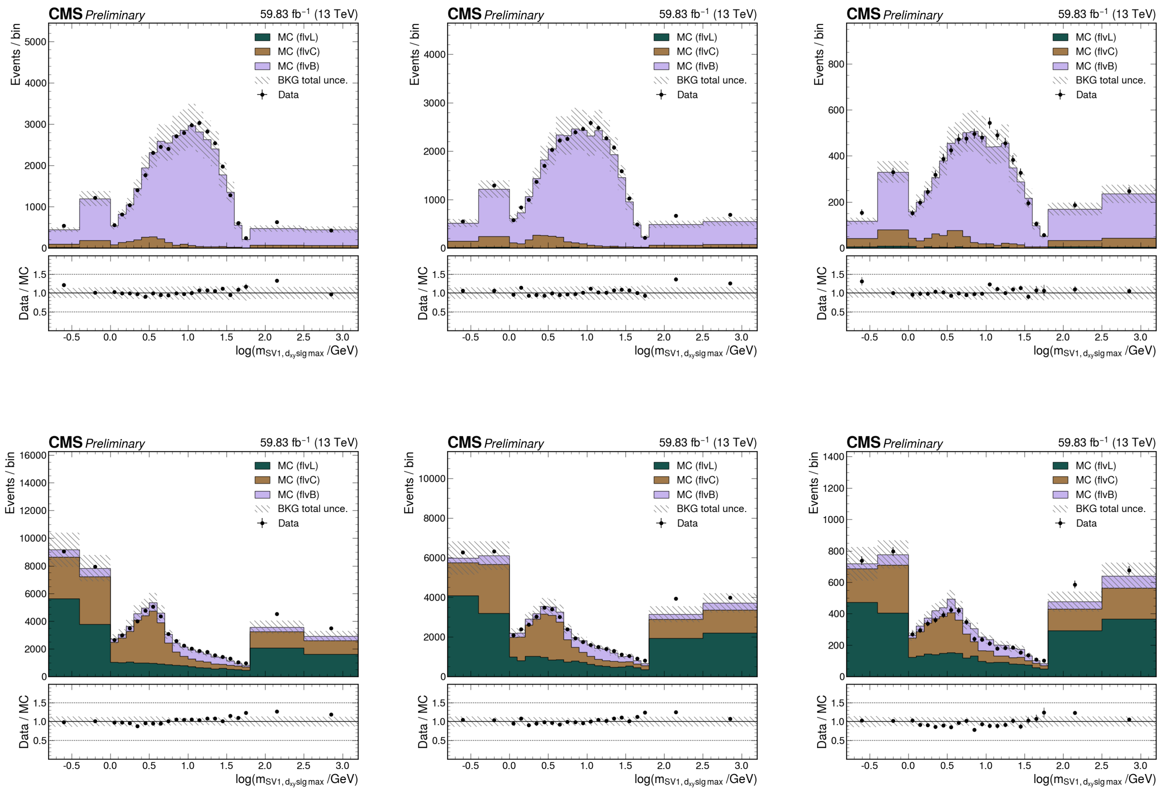


Figure 27: The pre-fit distribution of the natural logarithm of the secondary vertex mass $m_{\mathrm{SV}_1,d_{xy}\,\mathrm{sig\,max}}$ is plotted in the “pass” (top) and “fail” (bottom) region for 2018 in the $[250, 500)$ GeV (left), $[500, 700)$ GeV (center), and $[700, \infty)$ GeV (right) p_{T} bins. The subscript “SV_1, d_{xy} sig max” stands for the secondary vertex with the maximum impact parameter d_{xy} significance, and the natural logarithm is taken to account for the variable’s long tail due to limited resolution.

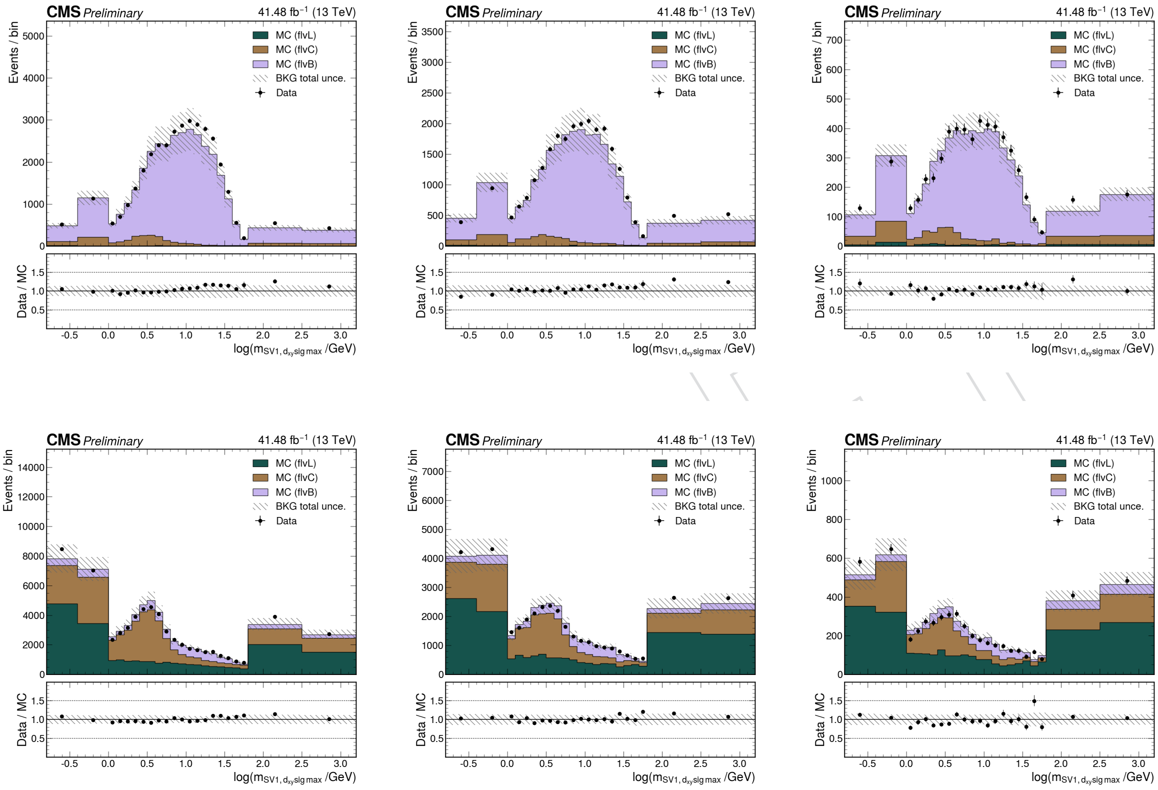


Figure 28: The pre-fit distribution of the natural logarithm of the secondary vertex mass $m_{\mathrm{SV}_1, d_{xy}\,\mathrm{sig\,max}}$ is plotted in the "pass" (top) and "fail" (bottom) region for 2017 in the $[250, 500)$ GeV (left), $[500, 700)$ GeV (center), and $[700, \infty)$ GeV (right) p_{T} bins. The subscript "SV_1, d_{xy} sig max" stands for the secondary vertex with the maximum impact parameter d_{xy} significance, and the natural logarithm is taken to account for the variable's long tail due to limited resolution.

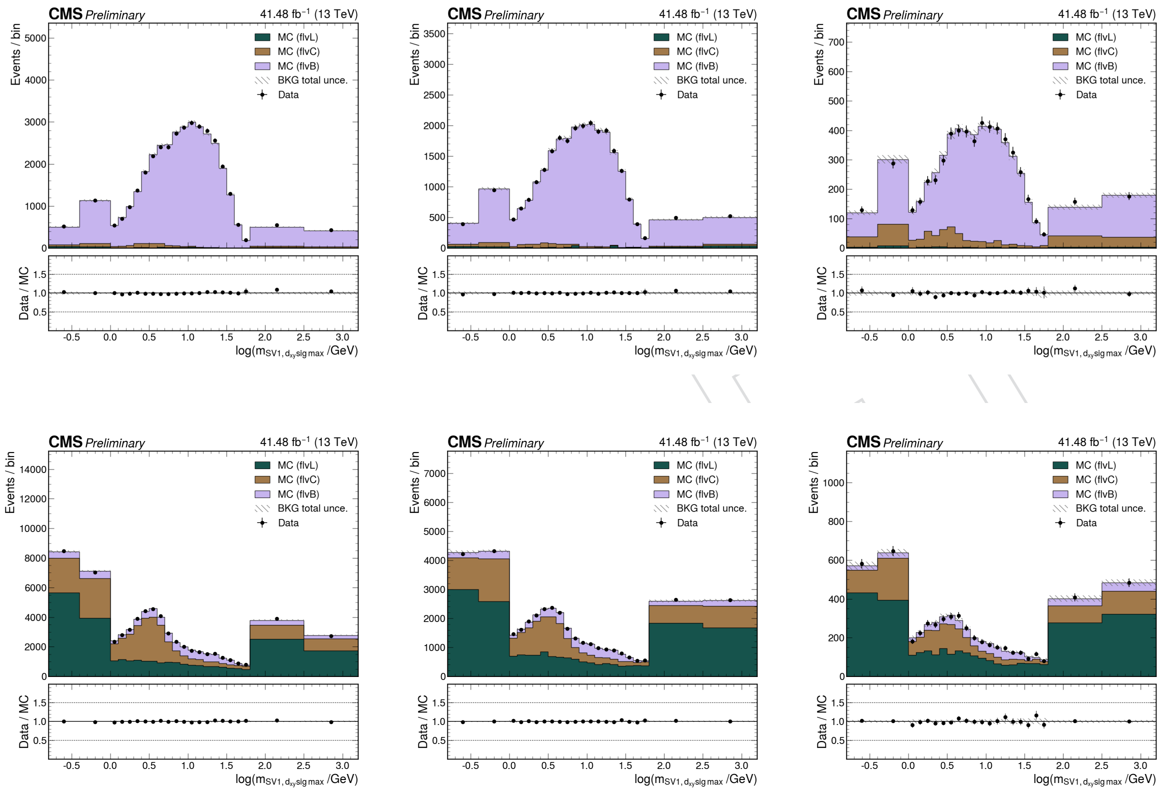


Figure 29: The post-fit distribution of the natural logarithm of the secondary vertex mass $m_{\mathrm{SV}_1,d_{xy}\,\mathrm{sig\,max}}$ is plotted in the “pass” (top) and “fail” (bottom) region for 2017 in the $[250, 500)$ GeV (left), $[500, 700)$ GeV (center), and $[700, \infty)$ GeV (right) p_{T} bins. The subscript “SV_1, d_{xy} sig max” stands for the secondary vertex with the maximum impact parameter d_{xy} significance, and the natural logarithm is taken to account for the variable’s long tail due to limited resolution.

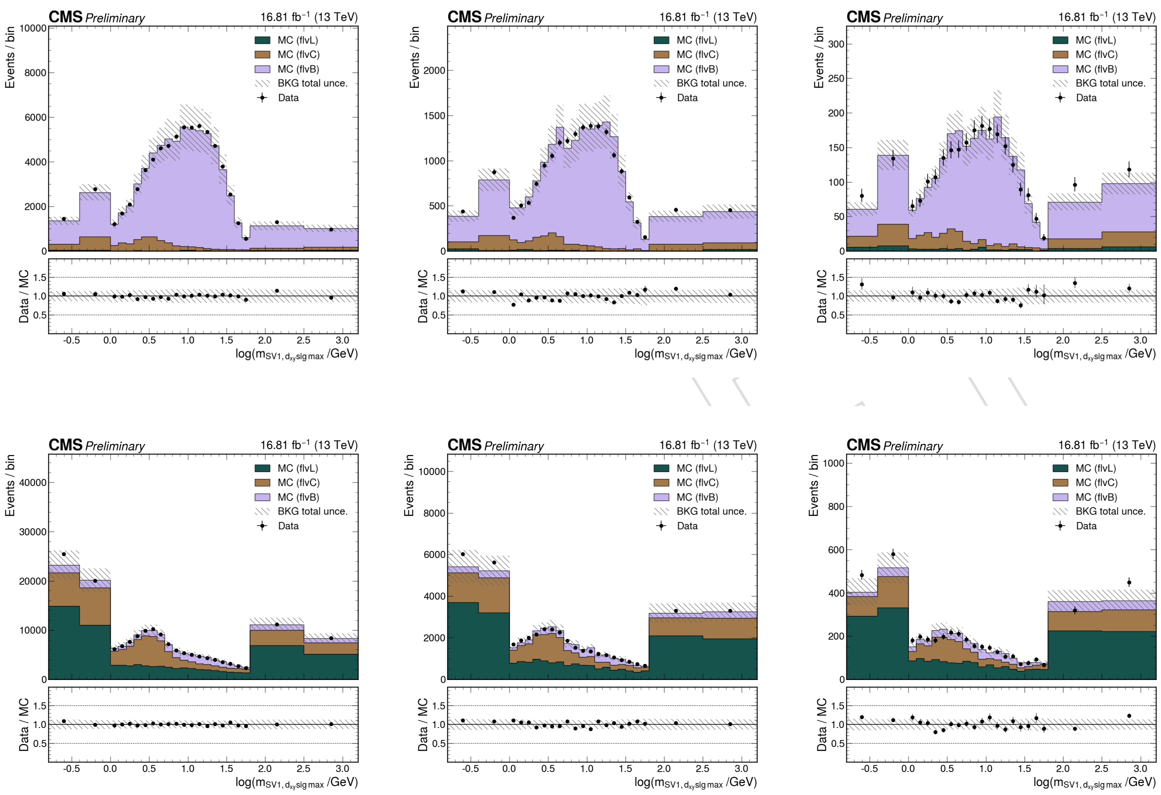


Figure 30: The pre-fit distribution of the natural logarithm of the secondary vertex mass $m_{\mathrm{SV}_1,d_{xy}\,\mathrm{sig\,max}}$ is plotted in the "pass" (top) and "fail" (bottom) region for 2016 (post-VFP) in the $[250, 500)$ GeV (left), $[500, 700)$ GeV (center), and $[700, \infty)$ GeV (right) p_{T} bins. The subscript "SV_1, d_{xy} sig max" stands for the secondary vertex with the maximum impact parameter d_{xy} significance, and the natural logarithm is taken to account for the variable's long tail due to limited resolution.

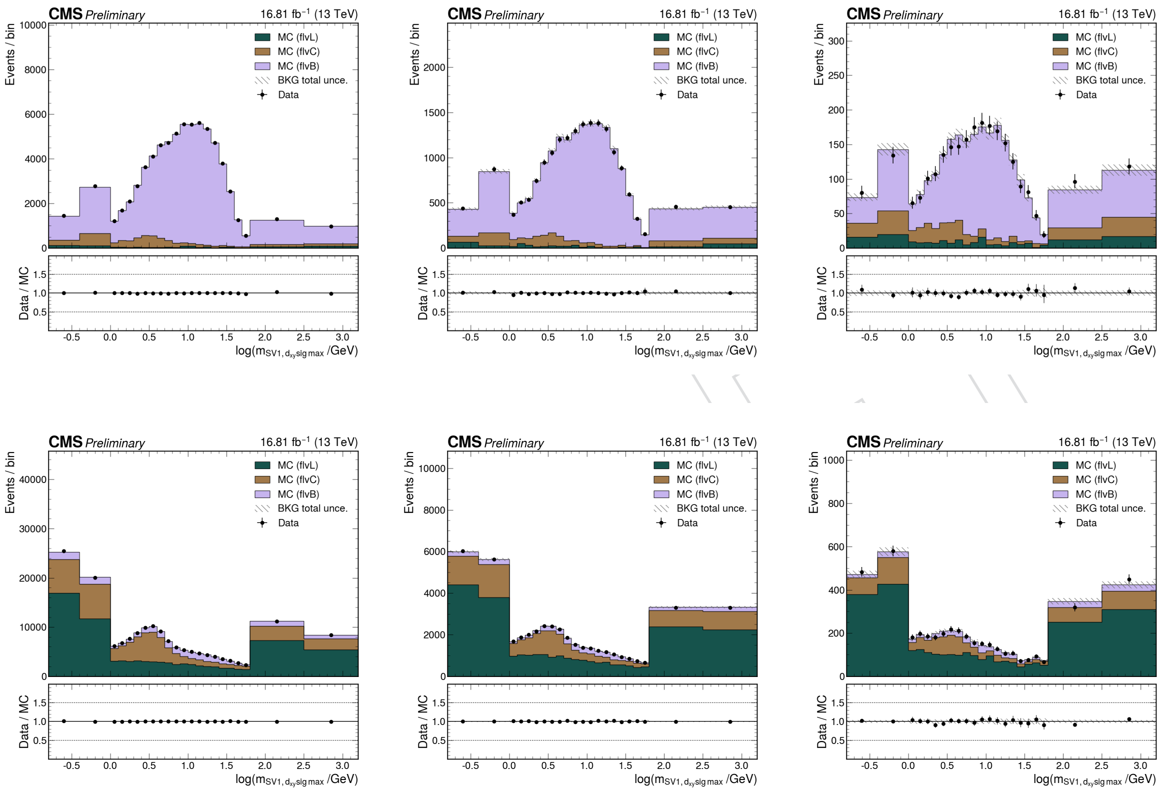


Figure 31: The post-fit distribution of the natural logarithm of the secondary vertex mass $m_{\mathrm{SV}_1, d_{xy}\,\mathrm{sig\,max}}$ is plotted in the "pass" (top) and "fail" (bottom) region for 2016 (post-VFP) in the $[250, 500)$ GeV (left), $[500, 700)$ GeV (center), and $[700, \infty)$ GeV (right) p_{T} bins. The subscript "SV_1, d_{xy} sig max" stands for the secondary vertex with the maximum impact parameter d_{xy} significance, and the natural logarithm is taken to account for the variable's long tail due to limited resolution.

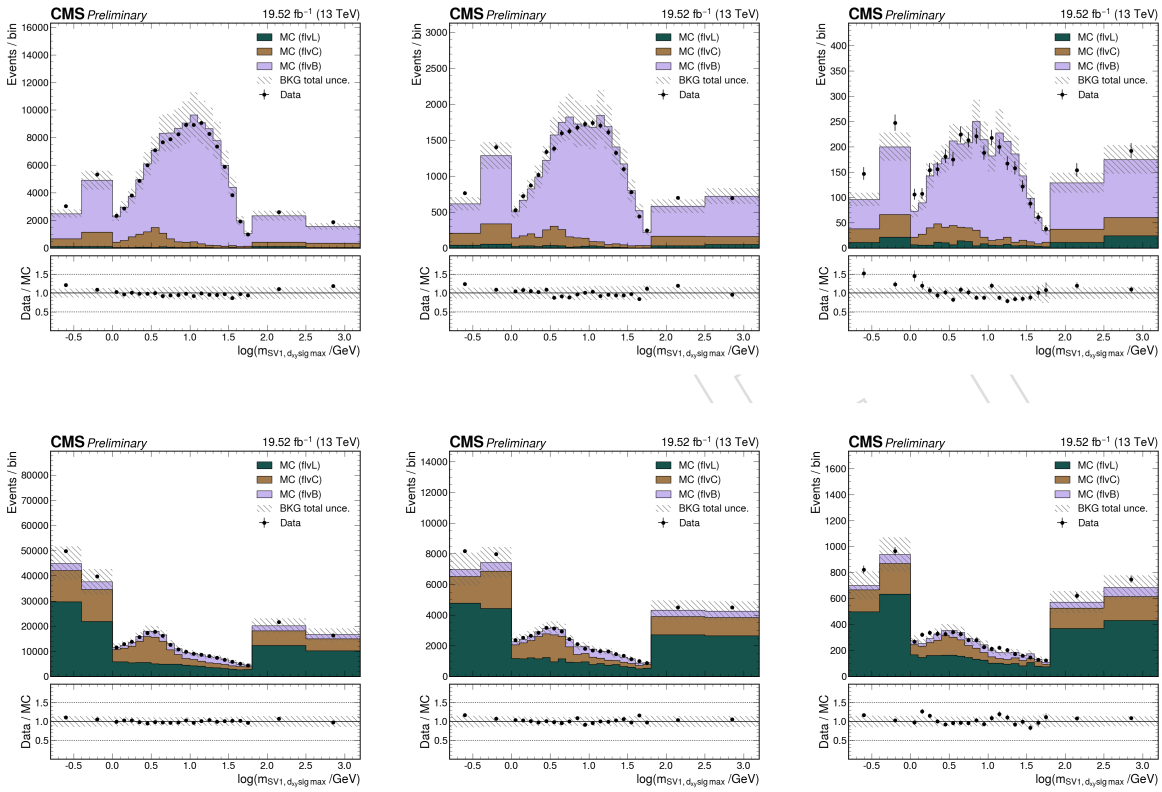


Figure 32: The pre-fit distribution of the natural logarithm of the secondary vertex mass $m_{\mathrm{SV}_1, d_{xy}\,\mathrm{sig\,max}}$ is plotted in the “pass” (top) and “fail” (bottom) region for 2016 (pre-VFP) in the $[250, 500)$ GeV (left), $[500, 700)$ GeV (center), and $[700, \infty)$ GeV (right) p_{T} bins. The subscript “SV_1, d_{xy} sig max” stands for the secondary vertex with the maximum impact parameter d_{xy} significance, and the natural logarithm is taken to account for the variable’s long tail due to limited resolution.

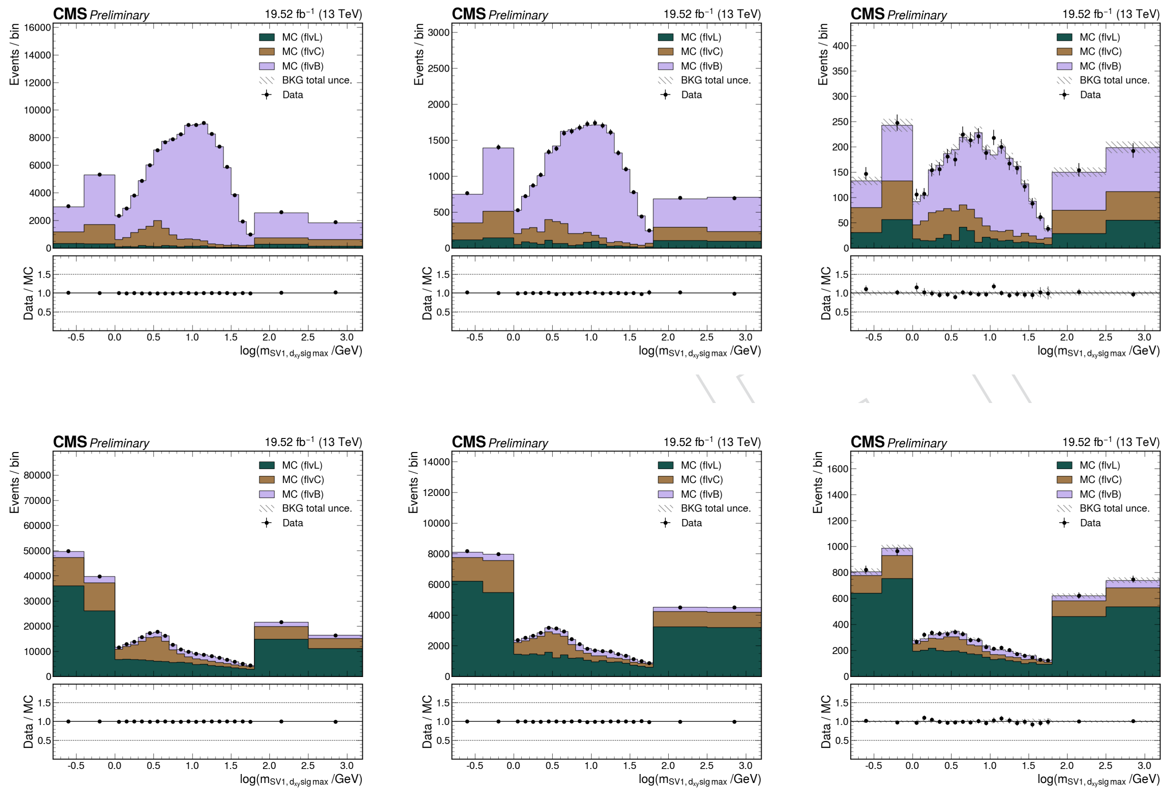


Figure 33: The post-fit distribution of the natural logarithm of the secondary vertex mass $m_{\mathrm{SV}_1,d_{xy}\,\mathrm{sig\,max}}$ is plotted in the "pass" (top) and "fail" (bottom) region for 2016 (pre-VFP) in the $[250, 500)$ GeV (left), $[500, 700)$ GeV (center), and $[700, \infty)$ GeV (right) p_{T} bins. The subscript "SV_1, d_{xy} sig max" stands for the secondary vertex with the maximum impact parameter d_{xy} significance, and the natural logarithm is taken to account for the variable's long tail due to limited resolution.

C ABCD subleading background composition uncertainty

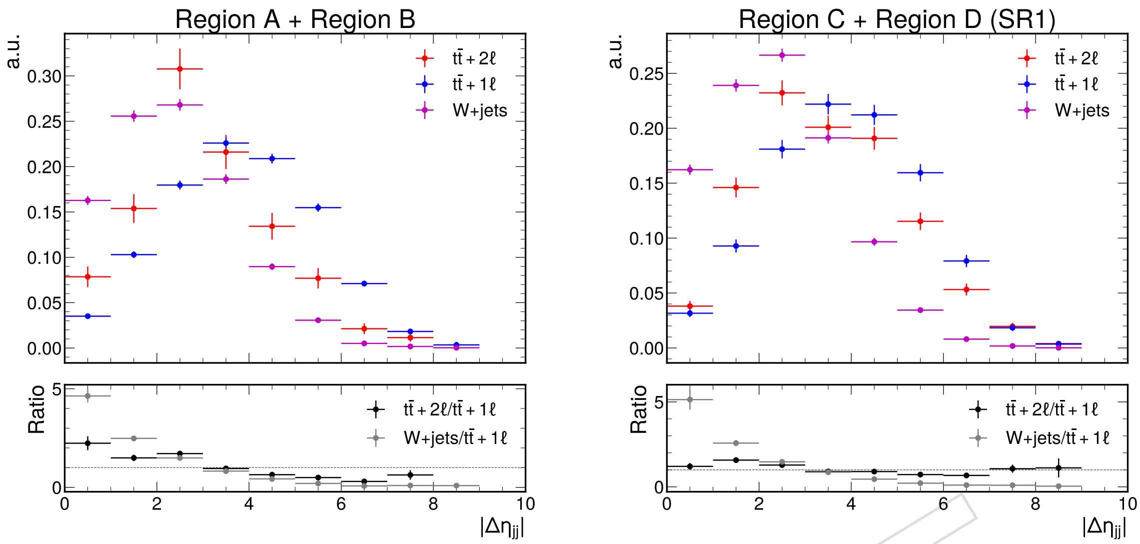


Figure 34: The VBS jet absolute difference in pseudorapidity is plotted in Regions A and B (left) versus Regions C and D (right) in Monte Carlo. It is clear that the subdominant W+Jets background is distributed in $\Delta\eta_{\text{jj}}$ in a significantly different way from the dominant TTbar backgrounds, which are consistent with each other.

Based on Fig. 34, it is clear that the W+Jets background is distributed in $\Delta\eta_{\text{jj}}$ in a significantly different way from the dominant TTbar backgrounds. To account for this, the W+Jets background yield is first varied up and down by a factor of 2 in Regions A and B, which in turn affects the expected value of the transfer factor used to perform the estimation of the background yield in the signal region:

$$\frac{A_{MC}}{B_{MC}} = 0.71 \pm 3.1\% \qquad \frac{A^{up}_{MC}}{B^{up}_{MC}} = 0.68 \pm 2.9\% \qquad \frac{A^{down}_{MC}}{B^{down}_{MC}} = 0.74 \pm 3.2\%$$

The maximal the percent difference between the up and down variations versus the nominal transfer factor in Monte Carlo is taken as a systematic uncertainty, amounting to about 5.4%.

Upon close inspection of the plots in Appendix 5, it is clear that the Bosons background (mostly diboson processes) is also distributed differently between the regions. As such, it is appropriate to derive a systematic uncertainty on the Bosons background composition as was done for W+jets above:

$$\frac{A_{MC}}{B_{MC}} = 0.71 \pm 3.1\% \qquad \frac{A^{up}_{MC}}{B^{up}_{MC}} = 0.7 \pm 4.5\% \qquad \frac{A^{down}_{MC}}{B^{down}_{MC}} = 0.72 \pm 2.6\%$$

Again, the maximal the percent difference between the up and down variations versus the nominal transfer factor in Monte Carlo is taken as a systematic uncertainty, amounting to about 2.6%.

D Erroneous EWK samples

While constructing this analysis, it became clear that an abnormally large yield of EWKWLep samples survived past the Preselection. In particular, introducing those samples ruined the otherwise satisfactory closure of the ABCD background estimation method–compare Figures 15 and 35. After applying only the Preselection and $|\Delta\eta_{jj}| > 3$, which contains the relevant EWKWLep events, it is clear that the EWKWLep samples are abnormal. First, the vast majority of EWKWLep events in this region had one or both incoming b quarks (Fig. 36a). Moreover, most EWKWLep events that are VBS W events at the generator level in this region have just one outgoing quark matched to a VBS quark, where the other outgoing quark is matched to the $H \to b\bar{b}$ fat jet candidate and that outgoing quark is predominantly a b quark. Therefore, it became apparent that the kinematics of these b initiated VBS events were somehow simulated incorrectly.

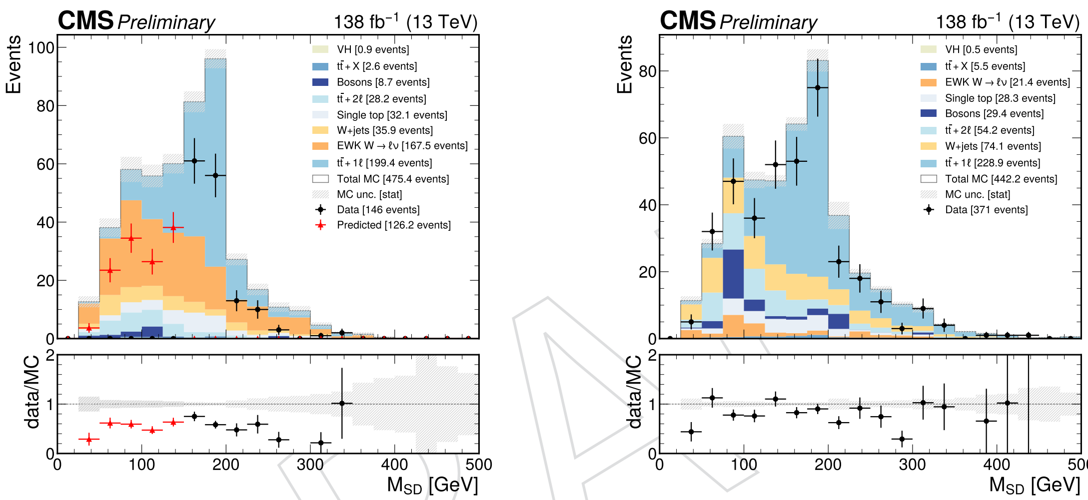


Figure 35: The softdrop mass of the $H \to b\bar{b}$ fat jet candidate is plotted in Regions A and D (left), which share the selection $|\Delta\eta_{jj}| > 0.4$, and Regions B and C (right), which share the selection $|\Delta\eta_{jj}| \leq 0.4$, showing poor closure of the ABCD method when compared to Fig. 15. An erroneously large yield of the EWK W leptonic sample is evident in Regions A and D in particular.

It was finally determined that there was an issue at the MadGraph level. In the process card, the following line was originally used to generate the EWKWLep samples:

```
pp > l vl j j / t t~ h QCD = 0
```

where the '/' excludes diagrams containing t, $\bar{t}$, and H diagrams from generation, and ignores any interference from them. However, the process should have instead been generated using

```
pp > l vl j j $ t t~ h QCD = 0
```

where the '$' excludes diagrams containing t, $\bar{t}$, and H diagrams from generation, but includes any interference from them. That is, the interference from diagrams containing t and $\bar{t}$ is significant. This can be seen in Fig. 37a, where it is clear that the p_T of outgoing b quarks is significantly boosted when using the incorrect MadGraph generation line. Put simply, the

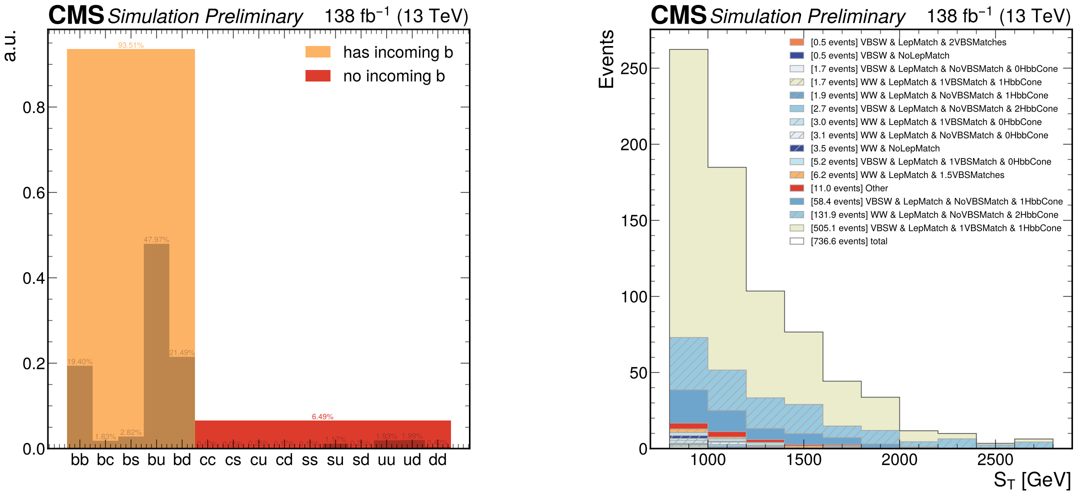


Figure 36: The incoming quark flavors (left) and event type (right) are plotted for only the EWKWLep samples after applying the Preselection and $|\Delta\eta_{jj}| > 3$. The event types are as follows. First, an event is classified as being a VBS W or WW event, as the EWKWLep samples contain both. Next, the generator-level lepton is matched using $\Delta R(\ell_{gen}, \ell_{reco}) < 0.4$ to the reconstructed lepton, and the event is classified as having or not having a matched lepton. Then, the generator-level quarks are matched to the reconstructed VBS quarks using $\Delta R(q_{gen}, q_{reco}) < 0.4$, and the event is classified according to the number of matches. Finally, the number of generator-level quarks in the $\mathrm{H} \to \mathrm{b\bar{b}}$ fat jet $\Delta R < 0.8$ cone are counted, and the event is classified accordingly.

EWKWLep samples are BSM samples where the top quark does not exist, so they inhabit a familiarly boosted phase space. As a temporary fix, a scale factor was derived by taking the ratio of the histograms of the outgoing b quark p_{T} for the correctly (numerator) and incorrectly (denominator) generated samples. This scale factor, binned in p_{T} of the outgoing b quark, is sufficient to completely remove the issue. In addition, the EWKWLep samples have no restriction on the minimum dijet mass, so diboson events are also included in the sample. Since this analysis uses a dedicated diboson sample already, these events are removed in order to avoid double counting. At the time of writing, the EWKWLep samples have been decommissioned and are being centrally generated with the correct MADGRAPH line.

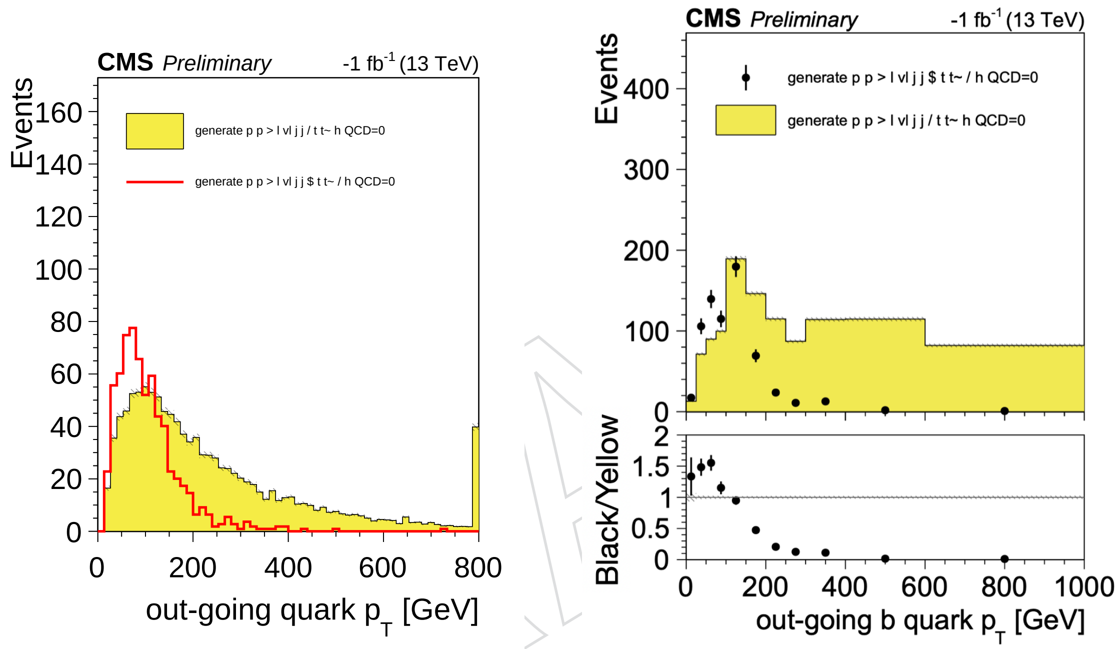


Figure 37: The p_T of outgoing b quarks are plotted for the correctly versus incorrectly generated samples. The plots are the same, except the left plot is binned evenly while the right plot is binned such that ratio of correctly to incorrectly generated samples can be taken as a scale factor.

E High-purity signal region

A high-purity signal region can be constructed by tightening the S_T cut in the signal region used in this analysis to $S_T > 1500\,\text{GeV}$. Doing an ABCD background estimation as was done in Section 5 is not possible, however, since Regions A and C are too signal polluted to be unblinded. Instead, Region B can be used to validate data-MC agreement, and the extrapolation factor that scales the actual yield in Region C to the estimated yield in Region D can be taken from MC, provided data and MC agree reasonably well in Region B–this is indeed the case. Then, the signal yield would again be taken from Monte Carlo. A table of systematics can then be produced as was done in Section 6.

Type	Systematic	Value
Background	Estimation syst. unc.	35.7%
	Data stat. unc.	13.4%
Signal	PDF variations	2.2%
	μ_F scale	21.1%
	Parton shower ISR weights	0.3%
	Parton shower FSR weights	0.8%
	Pileup reweighting	0.5%
	Pileup jet ID	1.0%
	L1 pre-fire corrections	1.0%
	HLT scale factors	0.8%
	Simulation stat. unc.	4.2%
	Lepton scale factors	0.03% - 1.5%
	ParticleNet Xbb scale factors	1.2% - 2.4%
	DeepJet b-tagging scale factors	0.3%
	MET unc.	0.2%
	Jet energy scale	8.0%
	Jet energy resolution	0.5%
	Luminosity	1.6%
	$\text{H} \to \text{b}\bar{\text{b}}$ BR	1.27%

Table 18: The size of the systematic uncertainties on the signal and estimated background yield in the high-purity signal region are tabulated. The systematics tabulated here represent the full set of nuisance parameters that could be used in a fitting procedure.

From just MC, it is clear that this region is very pure in signal, with 111 signal events expected versus 6 predicted background events. Such a region was devised in case the signal region used in the analysis was at all ambiguous, and it could be worth exploring in its own right. It was not considered for the final result of this analysis since it is not orthogonal to the main signal region, and the main signal region is anyway sufficient to exclude $\lambda_{WZ} = -1$.

Type	Yield	±	stat.	±	syst.
Signal	106	±	4.5	±	25.0
Background	5	±	0.7	±	1.9

Table 19: The background yield estimated from data and signal yield predicted by Monte Carlo simulation in the high-purity signal region are tabulated with their associated statistical and systematic uncertainties.

F Two-dimensional exclusion

A two-dimensional exclusion (i.e. an exclusion of λ_{WZ} values) is obtained by repeating the process used to obtain a single-point exclusion of $\kappa_W = -1, \kappa_Z = +1$ for each κ_W, κ_Z reweighting generated for this analysis. This effectively treats each point as a separate signal model. The exclusion of a signal strength of 1 for each point is then plotted on the z-axis of a two-dimensional (κ_W, κ_Z) histogram (Fig. 38). However, Combine did not scan enough signal strength values for a small number of points, so the exclusion was instead inferred by fitting the tail of the log-likelihood plot (Fig. 39, 40). In order to obtain the final result, the two-dimensional histogram of actual and inferred exclusions is interpolated such that a smooth exclusion contour can be drawn. Moreover, certain reweightings failed to correctly modify the signal acceptance, leading to discontinuities in the otherwise smoothly-varying histogram of $r = 1$ exclusions. These discontinuities, however, do not affect any of the exclusion contours and are anyway smoothed out by the interpolation.

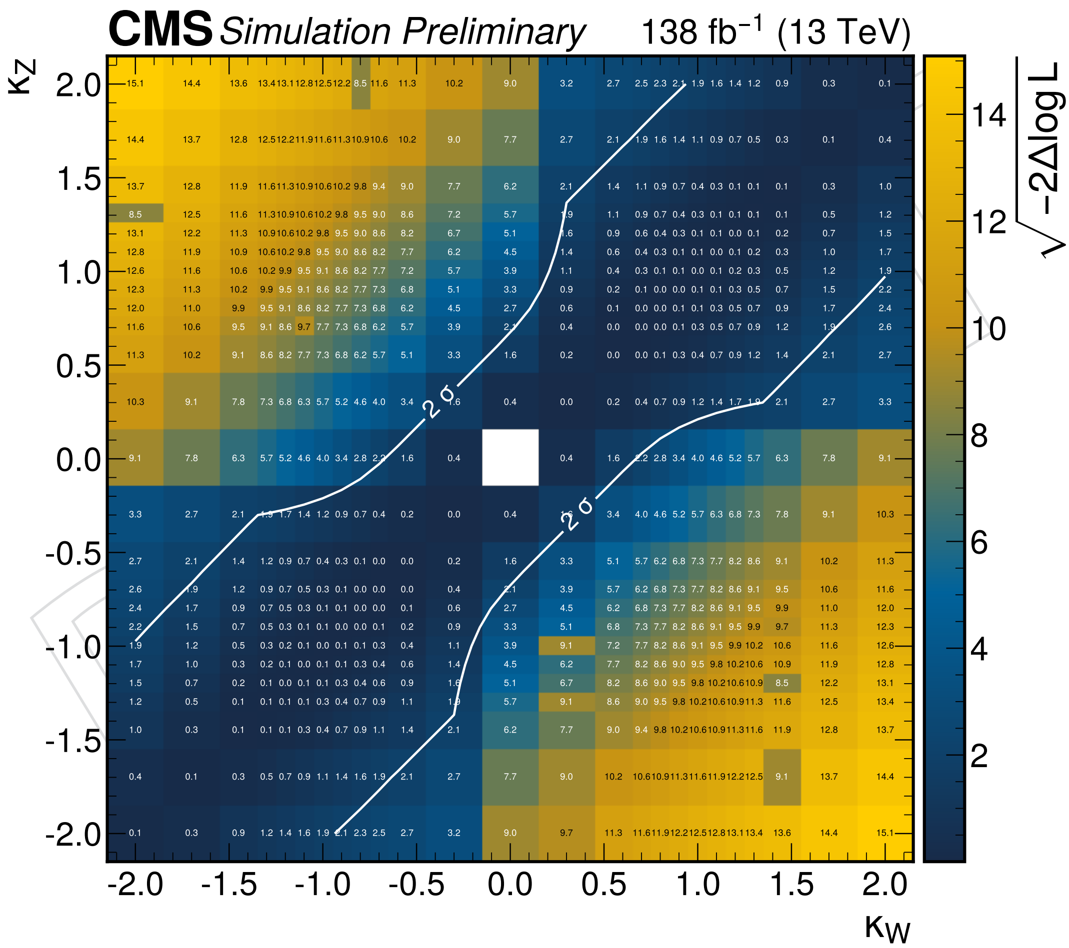


Figure 38: A two-dimensional histogram is plotted where the z-axis corresponds to the exclusion $\sigma = \sqrt{-2\Delta \log L}$ of a signal strength of 1 computed by Combine. The bins are centered on the κ_W, κ_Z plots scanned for this analysis.

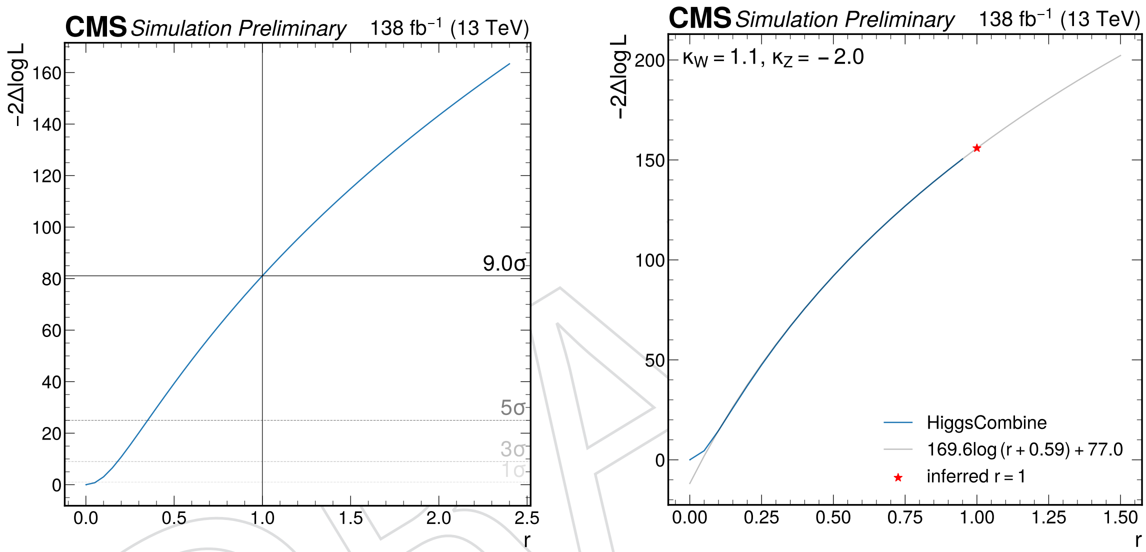


Figure 39: The single-point exclusion of $\kappa_W = -1$, $\kappa_Z = +1$ is plotted on the left as an example of an exclusion that can be taken directly from Combine. The exclusion of $\kappa_W = 1.1$, $\kappa_Z = -2.0$ is plotted on the right as an example of where the $r = 1$ exclusion must be inferred from the exclusions computed by Combine.

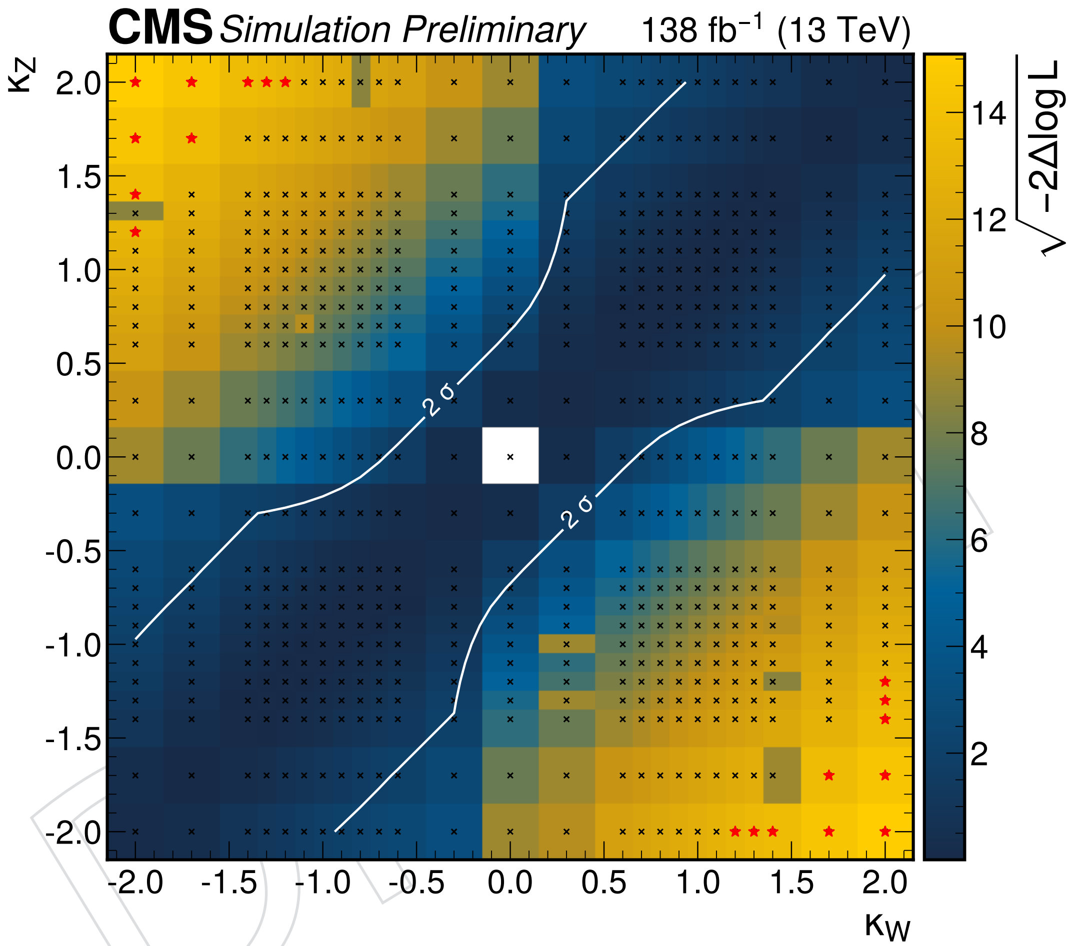


Figure 40: A two-dimensional histogram is plotted where the z-axis corresponds to the exclusion $\sigma = \sqrt{-2\Delta \log L}$ of a signal strength of 1 computed by Combine. The bins are centered on the κ_W, κ_Z plots scanned for this analysis. Red stars are plotted for κ_W, κ_Z points where the exclusion was inferred from the exclusions computed by Combine, while black x's are plotted for κ_W, κ_Z points are taken directly from Combine.