

CMS Draft Analysis Note

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

2018/07/18

Head Id: 469071

Archive Id: 469071

Archive Date: 2018/07/18

Archive Tag: trunk

Observation of the Higgs boson produced in association with a W or Z boson and decaying to a bottom quark-anti-quark pair

Caterina Vernieri¹, Jacobo Konigsberg², Sean-Jiun Wang², Andrea Rizzi³, Leonardo Giannini³, Stephane Brunet Cooperstein⁴, James Olsen⁴, Christopher Palmer⁴, Pirmin Berger⁵, Christoph Grab⁵, Gael Ludovic Perrin⁵, Luca Perrozzi⁵, Rainer Mankel⁶, Heiner Tholen⁶, Adinda de Wit⁶, Xavier Coubez⁷, Luca Mastrolorenzo⁷, Andrzej Novak⁷, and Alexander Schmidt⁷

¹ Fermi National Accelerator Laboratory, Batavia, Illinois, USA

² University of Florida, Gainesville, Florida, USA

³ INFN, Scuola Normale and Università di Pisa, Italy

⁴ Princeton University, Princeton, New Jersey, USA

⁵ Institute for Particle Physics, ETH Zurich, Zurich, Switzerland

⁶ DESY, Hamburg, Germany

⁷ RWTH Aachen University, Aachen, Germany

Abstract

A search for the standard model Higgs boson decaying to bottom quarks and produced in association with either a W or Z boson decaying leptonically is presented. A data sample comprising up to 41.2/fb from 2017 data taking with $\sqrt{s} = 13$ TeV has been analyzed in final states including 0, 1, or 2 charged leptons and two identified b-jets. Five distinct trigger strategies and analysis regions target the channels $Z(\mu\mu)H$, $Z(ee)H$, $Z(\nu\nu)H$, $W(\mu\nu)H$, and $W(e\nu)H$. The note focuses mainly on the description and analysis of the latest data, collected in 2017, highlighting the differences with respect to HIG-16-044. The overall 2017 fitted Higgs signal strength is found to be $\mu = \sigma/\sigma_{SM} = 1.10^{+0.34}_{-0.33}$, while the observed (expected) local significance of the excess over background prediction is quantified as 3.5σ (3.2σ). Results are also combined with the ones obtained with Run1 data, yielding a fitted signal strength of $\mu = \sigma/\sigma_{SM} = 1.02^{+0.22}_{-0.22}$ with the observed (expected) local significance of the excess over background predictions quantified as 4.9σ (4.9σ).

This box is only visible in draft mode. Please make sure the values below make sense.

PDFAuthor: VHbb Team
PDFTitle: Search for the Standard Model Higgs Boson Produced in Association with W and Z and Decaying to Bottom Quarks with 2017 data
PDFSubject: CMS
PDFKeywords: CMS, physics, software, computing

Please also verify that the abstract does not use any user defined symbols

DRAFT

1 Contents

1	1	Introduction	4
2	1.1	Analysis Strategy	6
3	1.2	Diboson cross check and cross section measurement	6
4	2	Data and Simulated Samples	7
5	2.1	Data	7
6	2.2	Monte Carlo samples	7
7	3	Triggers	11
8	3.1	Dedicated trigger for $Z(\nu\nu)H$	14
9	4	Physics Objects	16
10	4.1	Primary vertex selection and pile-up treatment	16
11	4.2	Electrons	18
12	4.3	Muons	18
13	4.4	Lepton isolation	19
14	4.5	Jets	20
15	4.6	Identification of b-jets	20
16	4.7	Missing transverse energy	22
17	4.8	Additional “soft” hadronic activity	22
18	5	Vector Boson Reconstruction	24
19	5.1	Vector boson transverse momentum reweighting in V+jets events	24
20	5.2	EWK signal VH corrections	25
21	6	Higgs Boson Reconstruction	27
22	6.1	Regression	27
23	6.2	ISR/FSR radiation recovery	31
24	6.3	Kinematic fit in 2-lepton channel	33
25	7	Residual MC corrections	38
26	7.1	$M(jj)$ reweighting in LO V+Jets	38
27	7.2	W boson and $t\bar{t}$ transverse momentum reweighting	38
28	8	Event Selection	40
29	8.1	Signal and background characteristics	40
30	8.2	Discriminating variables	41
31	8.3	Signal region pre-selection	42
32	9	Binned Multivariate Shape Analysis	44
33	9.1	Multivariate discriminator: input variables	44
34	9.2	Multivariate discriminator training	45
35	9.3	BDT performance	50
36	9.4	DNN performance	52
37	10	Signal Efficiency	57
38	10.1	Lepton reconstruction and ID efficiencies	58
39	11	Background Control Regions	62
40	11.1	0-lepton channel control regions	63
41	11.2	1-lepton channel control regions	67
42	11.3	2-lepton channel control regions	71

44	11.4	DNN background classifier	78
45	11.5	Scale factor determination	80
46	12	Systematics	82
47	12.1	Jet p_T, η decorrelation in JEC, JER and B-tagging systematics	84
48	12.2	Implementation of PDF Uncertainty Systematics	84
49	13	Results	86
50	13.1	Signal and Control Regions fits	86
51	13.2	Signal extraction calculation	88
52	13.3	Results VH	88
53	14	Conclusions	95
54	A	BDT distributions for VH discrimination	102
55	B	Results of the VZ(bb) cross-check analysis	103
56	B.1	Signal and Control Regions fits	103
57	B.2	Results VZ	105

DRAFT

58 **Notes and objectives for AN v11 (July 18th)**

- 59 • Corrected cross sections for VV samples
60 • Recomputed expected and observed numbers for VH analysis, updated tables and
61 plots
62 • Disclaimer: the numbers are not final, they might be subject to minor
63 changes due to alternative correlation scheme or other fit adjustments de-
64 cided together with the Higgs combination group
65 • Implemented and unblinded VZ cross-check analysis
66 • To be done by approval time
67 • Massless DNN (i.e. mass plot)

68 **Notes and objectives after pre-approval (AN v10, July 15th)**

- 69 • Consistent use of DNN discriminator in all the channels
70 • Factorized JECs into 26 independent nuisances as suggested at pre-approval
71 • Decorrelated b-tag systematics in p_T, η as suggested at pre-approval
72 • Reshaped DNN to lower the yield of the left-most bin as suggested at pre-approval
73 • Used b-enriched MC to reduce statistical uncertainty and increase signal sensitivity
74 • Used 2-muon channel trigger efficiency SFs
75 • Assigned full $\Delta\eta_{jj}$ reweighting described in Sec. 7 as shape systematic uncertainty
76 • Consolidated fit and made results more solid:
77 • Pruned low stat templates variations
78 • Reduced number of bins in high statistics control regions, as suggested
79 by ARC
80 • Improved discriminating power in the Heavy Flavour CR for 0- and 1-
81 lepton with dedicated multi-background DNN (cfr Sec. 11.4)
82 • Solved issues observed on:
83 • Wrongly named MC sample (labeled as low mass, while being high mass)
84 • Incorrect stitching of b-enriched V+jets samples due to use of incompati-
85 ble cross sections
86 • Incorrect reweighting of $\Delta\eta_{jj}$ distribution applied
87 • Updated SF and numbers with un-blinded fit results
88 • To be done by approval time
89 • Diboson analysis (VZ) and massless DNN (i.e. mass plot)

1 Introduction

In the Standard Model (SM) [1] [2] [3] the Brout-Englert-Higgs mechanism [4] [5] [6] explains the electroweak symmetry breaking and allows electroweak gauge bosons to acquire mass. The mechanism predicts the existence of a Higgs scalar field, and its observation in 2012 with the LHC Run1 proton-proton collision data by both the CMS [7] and ATLAS [8] collaborations achieved one of the main goals of the LHC physics program.

The observed Higgs boson mass is now precisely determined to be $m_H = 125.26 \pm 0.21$ GeV [9], i.e. with a 0.2% precision level. On the other hand, the observed properties and couplings are consistent with those of a minimal SM Higgs boson but only with a precision at the level of 10% or worse [10]. In particular, while a SM Higgs boson with a mass of 125.1 GeV is predicted to have a dominant decay branching ratios to b-quarks (58.1% [11]), the LHC Run1 data did not yet clearly establish the coupling of the Higgs boson to b-quarks, nor in general to down-type quarks.

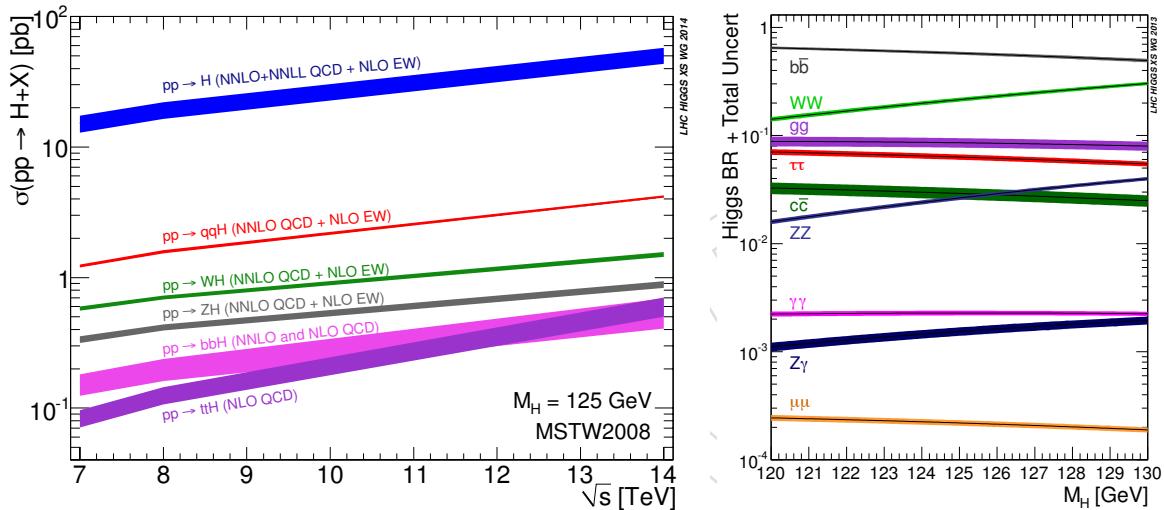


Figure 1: Minimal Standard Model Higgs production and decay at the LHC [12]. (left) Production cross-sections at $\sqrt{s} = 7$ –14 TeV, for $m_H = 125$ GeV. (right) Decay Branching Fractions for $m_H = 120$ –130 GeV.

Figure 1 shows the expected production cross sections as a function of the LHC proton-proton collision energy, and the expected decay modes Branching Fractions for minimal SM Higgs boson with $m_H \simeq 125$ GeV. The most abundant LHC channel for a SM Higgs boson after the gluon-fusion (GF) production is the Vector Boson Fusion (VBF) production followed by the vector boson associated (VH) production.

The mass at which the Higgs boson has been observed is particularly challenging at hadron colliders. Figure 1 shows the Higgs boson branching fractions to accessible final states near its measured mass. The dominant decay mode is $H \rightarrow b\bar{b}$, but in the dominant GF production mode the resulting irreducible background from QCD production of b quarks makes the inclusive search for the SM Higgs boson decaying to $b\bar{b}$ pairs much more difficult and not the most sensitive in proton collisions.

In the search for the decay of $H \rightarrow b\bar{b}$ as predicted by the SM at the LHC, the most sensitive channel is where the Higgs boson is produced in association with a W or Z boson, and recoiling with large momentum transverse to the beam line [13]. The presence of a boosted vector boson in the final state highly suppresses the QCD background, while also providing

118 an efficient trigger path when the W or Z decays to charged leptons. Requiring a large boost
119 provides several additional advantages: it further reduces the large backgrounds from W and
120 Z production in association with jets; it helps in reducing the large background from top-quark
121 production in the signal channels including neutrinos; it makes accessible the Z($\nu\nu$)H channel
122 via large missing transverse energy (E_T^{miss}); and it generally improves the mass resolution of
123 the reconstructed Higgs candidates.

124 The first search results from CMS for five channels ($Z(\ell\ell)H$, $Z(\nu\nu)H$, $W(\ell\nu)H$, with $\ell = e, \mu$) us-
125 ing 1.1 fb^{-1} of data and a cut analysis technique were presented in 2011 [14, 15]. Subsequently,
126 CMS published updated search results on the full 5 fb^{-1} 2011 dataset [16, 17] using similar
127 techniques achieving sensitivity of roughly four times the SM expectation at $M_H = 125 \text{ GeV}$.
128 These results were published in Ref. [18]. Further CMS updates were presented at the ICHEP
129 conference in summer 2012 [19, 20] based on the full 7 TeV data, and 5.1 fb^{-1} of 2012 8 TeV
130 data, and for the HCP 2012 conference [21, 22] based on the full 2011 7 TeV data and 12.1 fb^{-1}
131 of 2012 8 TeV data. Results comprising the full Run 1 data ($5 + 19 \text{ fb}^{-1}$ at 7 and 8 TeV) were
132 produced for the LHCP 2013 conference [23, 24], and published [25]. The Run 1 results yielded
133 an excess of events observed above the expected background, with a local significance of 2.1
134 standard deviations for a Higgs boson mass of 125 GeV, consistent with the expectation from
135 a SM Higgs boson production. The corresponding signal strength relative to the SM predicted
136 one was fitted to $\mu = 1.0 \pm 0.5$.

137 The VH, $H \rightarrow b\bar{b}$ Run 1 results have been corrected for the expected contribution of ZH pro-
138 ductions initiated by gluon fusion, yielding a lower $\mu = 0.89 \pm 0.43$ fitted signal strength, but
139 preserving the signal significance [26]. These results have been further combined with CMS
140 $H \rightarrow b\bar{b}$ searches in the VBF production mode [27], and in the top quark pair $t\bar{t}H$ associ-
141 ated production mode [28, 29]. For $m_H = 125 \text{ GeV}$ the final CMS Run1 combination yielded a
142 $H \rightarrow b\bar{b}$ signal strength $\mu = 1.03^{+0.44}_{-0.42}$ with a significance of 2.6 standard deviations.

143 In 2016, CMS collected $35.9/\text{fb}$ of pp data at 13 TeV. It was subsequently analyzed for the first
144 time and the $H \rightarrow b\bar{b}$ signal was extracted with a signal strength of $\mu = 1.2 \pm 0.4$ and an
145 observed (expected) significance of 3.3σ (2.8σ). This was combined with the Run 1 analysis
146 for an overall signal strength of $\mu = 1.06^{+0.31}_{-0.29}$ with observed (expected) significance of 3.8σ
147 (3.8σ) [30]. The ATLAS collaboration achieved the first evidence of the same process with a
148 similar dataset [31].

149 This analysis note describes the observation of the standard model Higgs boson with $H \rightarrow$
150 $b\bar{b}$ decay, produced in association with a W or Z boson decaying leptonically. Final states
151 including 0, 1, or 2 charged leptons and two identified b-jets are presented. They are divided
152 in five distinct analysis regions targeting the $Z(\mu\mu)H$, $Z(ee)H$, $Z(\nu\nu)H$, $W(\mu\nu)H$, and $W(e\nu)H$
153 channels.

154 The note focuses on the first analysis of the $41.3/\text{fb}$ of data collected in 2017 at $\sqrt{s} = 13 \text{ TeV}$.
155 Finally, a combination of the Run2+Run1 results is presented.

156 1.1 Analysis Strategy

157 The strategy follows closely the Run 1 and Run 2 analyses described in detail in
158 AN-12-181 [20], AN-12-349 [21], AN-13-069 [23], and AN-15-168 [32].

159 The event selection begins by identifying W and Z bosons decaying to charged leptons using
160 standard methods, and Z bosons decaying to neutrinos by requiring large E_T^{miss} . The $H \rightarrow b\bar{b}$
161 decay is reconstructed as two separate particle-flow jets clustered using the anti- k_T algorithm
162 with a size parameter of 0.4 (AK04). To identify candidates that are consistent with the b-jet
163 hypothesis, the best discriminating b-taggers among those commissioned is used for 2017 data,
164 i.e. the novel deep learning CSV (DeepCSV) tagger [33].

165 Backgrounds arise from production of W and Z bosons in association with one or more jets,
166 singly and pair-produced top quarks, and dibosons. Smaller residual QCD background is
167 present in the 0-lepton and 1-lepton channels. High-purity control regions for the V + udscg
168 and $t\bar{t}$ backgrounds are identified in data and used to estimate expected yields in the signal re-
169 gion. Samples orthogonal to the signal region in dijet mass and enhanced in V + $b\bar{b}$ production
170 are used to constrain the irreducible $b\bar{b}$ background from data. To increase the invariant mass
171 resolution, jets close-by the Higgs daughter candidates are identified as final state radiation
172 and added in the computation of the Higgs invariant mass (**).

173 In all channels a sufficient boost of the vector boson is required to overcome the majority of
174 QCD and low p_T irreducible background. In the channels with larger backgrounds, the V and
175 H candidates are required to have a larger boost transverse to the beam, i.e. 170 GeV in the
176 0-lepton and 150 GeV (**) in the 1-leptonIn the 2-lepton channel, where the background from
177 $t\bar{t}$ production is much less (and the effective signal cross section is also lower) lower boost
178 regions and a kinematic fit (**) to exploit the absence of event missing energy are also consid-
179 ered. Further background rejection is achieved by exploiting signal properties as the resonant
180 dijet mass, the back-to-back VH topology, and reduced additional hadronic activity. The most
181 discriminating variables are combined into a multivariate discriminator, and a Deep Neural
182 Network (DNN) is found to provide slightly better performance with respect to the previously
183 used BDT [34] (**).

184 1.2 Diboson cross check and cross section measurement

185 The WZ/ZZ (VZ) associated production channel with Z decaying to bottom quarks is a back-
186 ground for the VH search and partially reducible by making use of the $M(b\bar{b})$ distribution. This
187 well known SM process represents also a standard candle one can use to validate the analysis
188 strategy developed for the Higgs, since the value of the Z boson mass is close to the Higgs one
189 and the cross section is about 10 times larger.

190 The extraction of the Diboson signal has to be accounted as a validation of the Higgs analysis
191 strategy and the few differences from the analysis workflow defined for the analogue Higgs
192 measurement will be highlighted in the following sections of this note.

193 The diboson cross check measurement has been already used in previous analyses and led to
194 the observation of the VZ(bb) process at hadron colliders [30].

(**)^{**}New with respect to the previous Run2 analysis on 2016 data described in HIG-16-044 [30] and AN-15-168 [32].

2 Data and Simulated Samples

2.1 Data

The proton proton collision data with $\sqrt{s} = 13$ TeV and 25 ns minimum bunch spacing collected in 2017 have been used in this analysis, in the 31Mar2018 re-reco version with re-miniAODv2. For the $Z(\mu\mu)H$ channel we use the DoubleMuon dataset, for $Z(ee)H$ we use DoubleEG, for $W(\mu\nu)H$ we use SingleMuon, for $W(e\nu)H$ we use SingleElectron while for the $Z(\nu\nu)H$ channel we use the MET dataset.

Table 1 summarizes the datasets with the example of the SingleMuon datasets used in the current analysis and the approximate luminosity as reported from the golden JSON

`Cert_294927 – 306462_13TeV_EOY2017ReReco_Collisions17_JSON_v1.txt`

with `brilcalc` and the `normtag`.

Table 1: List of 2017 data samples used for the SingleMuon dataset. An uncertainty of 2.3% is assigned for the 2017 luminosity.

Dataset	$\int \mathcal{L} (\text{fb}^{-1})$
SingleMuon/Run2017B-31Mar2018-v1	~4.8
SingleMuon/Run2017C-31Mar2018-v1	~9.6
SingleMuon/Run2017D-31Mar2018-v1	~4.2
SingleMuon/Run2017E-31Mar2018-v1	~9.3
SingleMuon/Run2017F-31Mar2018-v1	~13.5
Total Lumi	41.3

2.2 Monte Carlo samples

Monte Carlo samples in CMSSW 94X are taken from the RunIIFall17 production with the `mc2017_realistic_v10` conditions and re-miniAODv2 to match 2017 data conditions. The list of samples with the number of events, cross sections and integrated luminosities is given below for Higgs boson signal (Table 3), di-boson production (Table 4), vector boson plus jets (Table 5), $t\bar{t}$ and single top with QCD multi-jet (Table 8).

Samples were produced using one or more of the following programs: PYTHIA 8 [35, 36], POWHEG [37], and MADGRAPH 5.aMC@NLO [38, 39] with MLM merging [40] or FxFx merging scheme [41]. Parton shower and hadronisation are performed with PYTHIA 8 [36] using the CP5 tune [42]. The NNPDF3.1 parton distribution functions (PDF) [43] at NNLO are used for all samples.

2.2.1 Signal simulation

Table 2 summarizes the cross sections and branching fractions assumed for each signal channel and with $M_H = 125$ GeV. The cross sections are computed at NNLO, as described in Ref. [11, 12, 44].

The list of signal Monte Carlo samples is given in Table 3, including $H \rightarrow b\bar{b}$ (HToBB) decays for Higgs bosons produced in vector boson associated processes initiated either by quark fusion or gluon fusion processes.

Table 2: Signal cross sections and branching ratios for $M_H = 125$ GeV at $\sqrt{s} = 13$ TeV [11].

process	$\sigma(\text{pb})$	QCDScale	PDF
ggH	44.14	+7.6-8.1%	$\pm 3.1\%$
VBF H	3.782	+0.4-0.3%	$\pm 2.1\%$
WH	1.373 (0.840+0.533)	+0.5-0.7%	$\pm 1.9\%$
ZH	0.8839	+3.8-3.1%	$\pm 1.6\%$
decay	BR	Uncertainty	
$H \rightarrow b\bar{b}$	58.24%	+0.72-0.74%	

Table 3: Signal Monte Carlo samples with $M_H = 125$ GeV

Sample	Generator	$m_H(\text{GeV}/c^2)$	$\sigma(\text{pb})$	events	$\int \mathcal{L}(\text{fb}^{-1})$
WplusH_HToBB_WToLNu_M125_13TeV_powheg_pythia8 POWHEG +PYTHIA 8	125	0.840 * 0.108535 * 0.5824	2 360 806	44 462.	
WminusH_HToBB_WToLNu_M125_13TeV_powheg_pythia8 POWHEG +PYTHIA 8	125	0.533 * 0.108535 * 0.5824	2 382 500	70 715.	
ZH_HToBB_ZToLL_M125_13TeV_powheg_pythia8 POWHEG +PYTHIA 8	125 (0.8839 - 0.1227) * 0.10974 * 0.5824	4 728 560	97 195.		
ZH_HToBB_ZToNuNu_M125_13TeV_powheg_pythia8/ POWHEG +PYTHIA 8	125 (0.8839 - 0.1227) * 0.20103 * 0.5824	2 973 469	33 364.		
ggZH_HToBB_ZToLL_M125_13TeV_powheg_pythia8 POWHEG +PYTHIA 8	125	0.1227 * 0.10974 * 0.5824	2 863 006	365 082.	
ggZH_HToBB_ZToNuNu_M125_13TeV_powheg_pythia8 POWHEG +PYTHIA 8	125	0.1227 * 0.20103 * 0.5824	2 976 978	207 228.	

2.2.2 Background simulations

- Tables 4–8 summarize the simulated background samples and their equivalent luminosities and cross sections where applicable. The production cross sections for W+jets and Z+jets are rescaled to next-to-next-to-leading-order (NNLO) cross sections calculated using the FEWZ 3.1 program [45–47]. The $t\bar{t}$ and single top quark samples are also rescaled to their cross sections based on NNLO calculations [48, 49].

Table 4: List of Monte Carlo diboson samples. The values are taken from inclusive measurements performed by CMS [50–52] and multiplied by the corresponding BR computed from PDG to obtain the exclusive values apart from the sample ZZTo2L2Q, taken from the NLO calculation and marked with the symbol (*).

Sample	Generator	$\sigma(\text{pb})$	events	$\int \mathcal{L}(\text{fb}^{-1})$
WW_TuneCP5_13TeV_pythia8 PYTHIA 8	115.3	7 547 722	63.6	
WZ_TuneCP5_13TeV_pythia8 PYTHIA 8	48.1	3 928 630	83.4	
ZZ_TuneCP5_13TeV_pythia8 PYTHIA 8	14.6	985 600	118.0	
WWTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8 MADGRAPH 5+PYTHIA 8	50.86	-	-	
WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8 MADGRAPH 5+PYTHIA 8	10.88	-	-	
ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8 MADGRAPH 5+PYTHIA 8	3.69(*)	-	-	

Table 5: List of Monte Carlo Z+jets ($Z \rightarrow ll$) leading order samples

Sample	Generator	σ (pb)	events	$\int \mathcal{L} (\text{fb}^{-1})$
DYJetsToLL_M-4to50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	204.0*1.23	8 796 005	43.3
DYJetsToLL_M-4to50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	54.4*1.23	1 816 239	34.6
DYJetsToLL_M-4to50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	5.70*1.23	-	-
DYJetsToLL_M-4to50_HT-600toInf_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	1.85*1.23	1 771 556	940.
DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	5343.0*1.23	-	-
DYJetsToLL_M-50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	161.1*1.23	10 031 487	55.3
DYJetsToLL_M-50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	48.66*1.23	9 917 255	201.
DYJetsToLL_M-50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	6.97*1.23	9 348 901	1339.
DYJetsToLL_M-50_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	1.743*1.23	8 003 554	4760
DYJetsToLL_M-50_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	0.805*1.23	3 065 315	3953
DYJetsToLL_M-50_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	0.193*1.23	-	-
DYJetsToLL_M-50_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	0.00347*1.23	388 323	88600
DYBJetsToLL_M-50_Zpt-100to200_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	4.042*1.23*1.085	4 080 000	1320
DYBJetsToLL_M-50_Zpt-200toInf_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	0.4286*1.23*1.085	2 110 000	6670
DYJetsToLL_BGenFilter_Zpt-100to200_M-50_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	3.384*1.23*1.15	-	-
DYJetsToLL_BGenFilter_Zpt-200toInf_M-50_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	0.5327*1.23*1.15	-	-

Table 6: List of Monte Carlo Z+jets ($Z \rightarrow \nu\nu$) leading order samples

Sample	Generator	σ (pb)	events	$\int \mathcal{L} (\text{fb}^{-1})$
ZJetsToNuNu_HT-100To200_13TeV-madgraph	MADGRAPH 5+PYTHIA8	304.2*1.23	16 588 827	48.1
ZJetsToNuNu_HT-200To400_13TeV-madgraph	MADGRAPH 5+PYTHIA8	91.92*1.23	21 722 423	413.
ZJetsToNuNu_HT-400To600_13TeV-madgraph	MADGRAPH 5+PYTHIA8	13.18*1.23	9 784 358	471.
ZJetsToNuNu_HT-600To800_13TeV-madgraph	MADGRAPH 5+PYTHIA8	3.258*1.23	5 585 160	6547.
ZJetsToNuNu_HT-800To1200_13TeV-madgraph	MADGRAPH 5+PYTHIA8	1.496*1.23	2 038 645	5172.
ZJetsToNuNu_HT-1200To2500_13TeV-madgraph	MADGRAPH 5+PYTHIA8	0.3419*1.23	363 874	3736.
ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph	MADGRAPH 5+PYTHIA8	0.005112*1.23	6734	2928.
ZBJetsToNuNu_M-50_Zpt-100to200_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	7.7*1.23*1.085	4 080 000	1320
ZBJetsToNuNu_M-50_Zpt-200toInf_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	0.8131*1.23*1.085	2 110 000	6670
ZJetsToNuNu_BGenFilter_Zpt-100to200_M-50_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	2.139*3*1.23*1.11	-	-
ZJetsToNuNu_BGenFilter_Zpt-200toInf_M-50_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA8	0.3287*3*1.23*1.11	-	-

Table 7: List of Monte Carlo W+jets leading order samples

Sample	Generator	σ (pb)	events	$\int \mathcal{L} (\text{fb}^{-1})$
WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	52940.0*1.21		
WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	1395.0 *1.21		
WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	407.9 *1.21		
WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	57.48*1.21		
WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	12.87*1.21		
WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	5.366*1.21		
WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	1.074*1.21		
WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	0.03216*1.21		
WBJetsToLNu_Wpt-100to200_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	7.35*1.21*1.5	3 979 072	662.7
WBJetsToLNu_Wpt-200toInf_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	1.1*1.21*1.5	2 892 981	3393.3
WJetsToLNu_EGenFilter_Wpt-100to200_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	26.6*1.21*1.12	6 690 000	256.3
WJetsToLNu_EGenFilter_Wpt-200toInf_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	3.9*1.21*1.12	11 650 000	3286.3

Table 8: $t\bar{t}$, single top and QCD Monte Carlo samples

Sample	Generator	σ (pb)	events	$\int \mathcal{L} (\text{fb}^{-1})$
TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8	POWHEG +PYTHIA 8	88.29	66979742	789.5
TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	POWHEG +PYTHIA 8	365.34	111325048	304.7
TTToHadronic_TuneCP5_PSweights_13TeV-powheg-pythia8	POWHEG +PYTHIA 8	377.96	130725364	346
ST_tW_top_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	POWHEG +PYTHIA 8	35.85	7660001	215.
ST_tW_antitop_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8	POWHEG +PYTHIA 8	35.85	7993682	224.5
ST_t-channel_top_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8	POWHEG +PYTHIA 8	136.02*0.325	5865875	133.
ST_t-channel_antitop_4f_inclusiveDecays_TuneCP5_13TeV-powhegV2-madspin-pythia8	POWHEG +PYTHIA 8	80.95*0.325	3939990	150.
ST_s-channel_4f_leptonDecays_TuneCP5_PSweights_13TeV-amcatnlo-pythia8	POWHEG +PYTHIA 8	10.32*0.325	9906720	???
QCD_HT100to200_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	27990000	82 095 800	0.003
QCD_HT200to300_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	1547000	18 784 379	0.011
QCD_HT300to500_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	322600	54 267 650	0.16
QCD_HT500to700_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	29980	19 542 847	0.66
QCD_HT700to1000_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	6334	45 100 675	6.60
QCD_HT1000to1500_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	1088	15 193 645	12.59
QCD_HT1500to2000_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	99.11	3 939 077	32.85
QCD_HT2000toInf_TuneCP5_13TeV-madgraphMLM-pythia8	MADGRAPH 5+PYTHIA 8	20.23	1 961 774	77.2

230 3 Triggers

231 A mix of different triggers is used to collect events consistent with the signal hypothesis in each
 232 channel. Table 9 summarizes the triggers used in this analysis for 2017 data.

Table 9: List of L1 and HLT triggers used for the 2017 data set, and the channels to which they apply.

Channel	L1 Seeds	HLT Paths
W($\mu\nu$)H	L1_SingleMu22	HLT_IsoMu27
Z($\mu\mu$)H	L1_DoubleMu_12_5	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8
W($e\nu$)H	L1_SingleEG38 OR L1_SingleIsoEG30 OR L1_SingleIsoEG28er2p1 OR L1_DoubleEG_25_12	HLT_Ele32_WPTight_Gsf_L1DoubleEG
Z(ee)H	L1_SingleEG30 OR L1_SingleIsoEG22er OR L1_SingleIsoEG24 OR L1_DoubleEG_15_10	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ
Z($\nu\bar{\nu}$)H	L1_ETM110 L1_ETMHF120 L1_ETMHF110_HTT60er	HLT_PFMET120_PFMHT120_IDTight OR HLT_PFMET120_PFMHT120_IDTight_PFHT60

233 The W($\mu\nu$)H and W($e\nu$)H channels utilize single lepton triggers. The Z($\mu\mu$)H and Z(ee)H
 234 channels are instead based on di-lepton triggers, which provide lower p_T cut on the higher
 235 leg with respect to the single lepton triggers. All the triggers are emulated in the MC, and all
 236 the MC events are required to satisfy the same trigger conditions used in data. Differences in
 237 trigger efficiencies between data and MC are taken into account by correction scale factors
 238 derived using the tag-and-probe method exploiting di-lepton events from Z bosons decays.
 239 Given the very tight tag lepton selection requirements, as well as the narrow di-lepton invariant
 240 mass cut around the Z peak, the probe lepton is very pure with minimal selection.

241 The trigger efficiencies are measured on top of the offline lepton identification and isolation
 242 selections. For di-lepton triggers, the scale factors for each leg are computed separately because
 243 of the different selection requirements. The efficiency correction scale factors are measured
 244 to be around 0.97 for single lepton triggers and for each leg of the di-lepton triggers. The
 245 correction scale factors and their uncertainties are shown in Figures 2 and 3 for the electron
 246 triggers as a function of the electron p_T and η .

247 A different selection on the dimuon invariant mass is used for run B and runs C, D, E, F:
 248 the HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8 HLT path is used for run B and the
 249 HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8 trigger path is used for runs C, D, E, F.
 250 Corrections are applied on the HLT path by evaluating the data and MC efficiencies separately
 251 for each muon leg, computed on the whole B, C, D, E, and F datasets. Additional corrections
 252 are evaluated separately on run B and runs C, D, E, F to take into account the cut on the dZ
 253 between the tracks and the invariant mass of the two muon candidates for run B and run C, D,
 254 E, F. They are studied separately for run B and C, D, E, F as the dimuon mass cut is different for
 255 those two datasets. The two sets of corrections are factorised: a first scale factor is computed on
 256 the HLT path without the additional dZ and mass cut on the whole run B, C, D, E, F. A second
 257 scale factor is then computed to account for the dZ and mass selection, requiring the numerator
 258 to pass the full

259 $\text{HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8}$

260 (respectively $\text{HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8}$) trigger path for run B (re-
 261 spectively run C, D, E, F), and the denominator to pass the trigger path without any dZ and

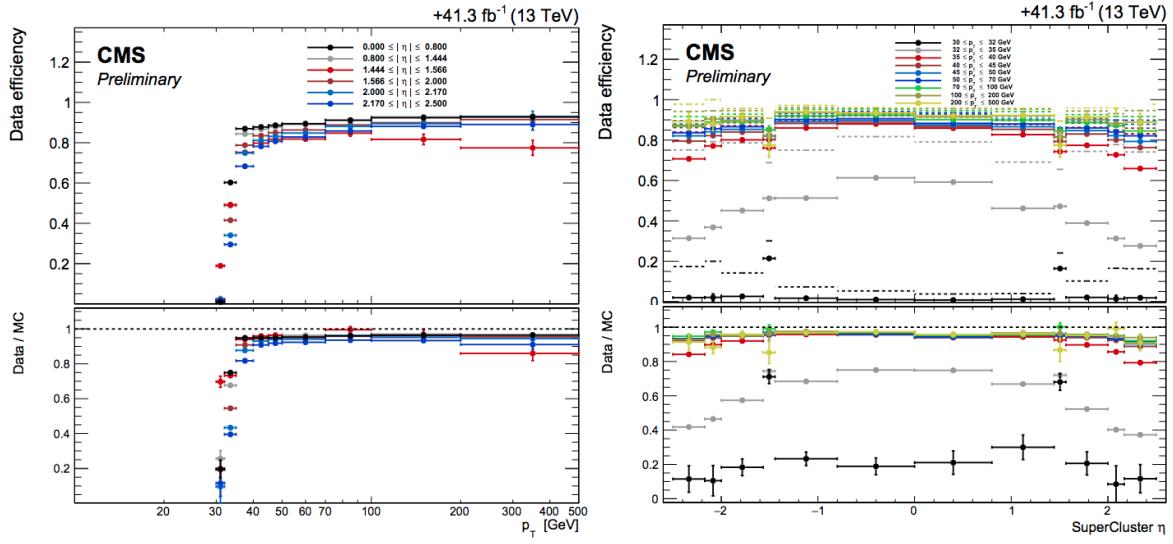


Figure 2: Distributions of HLT_Ele32_WPTight_Gsf efficiency as function of p_T and η for 2017 data and simulation. The efficiencies are measured after applying WP80 in the general purpose electron MVA IDs plus isolation selection. The turn-on can be seen as rising efficiency in p_T above 32 GeV.

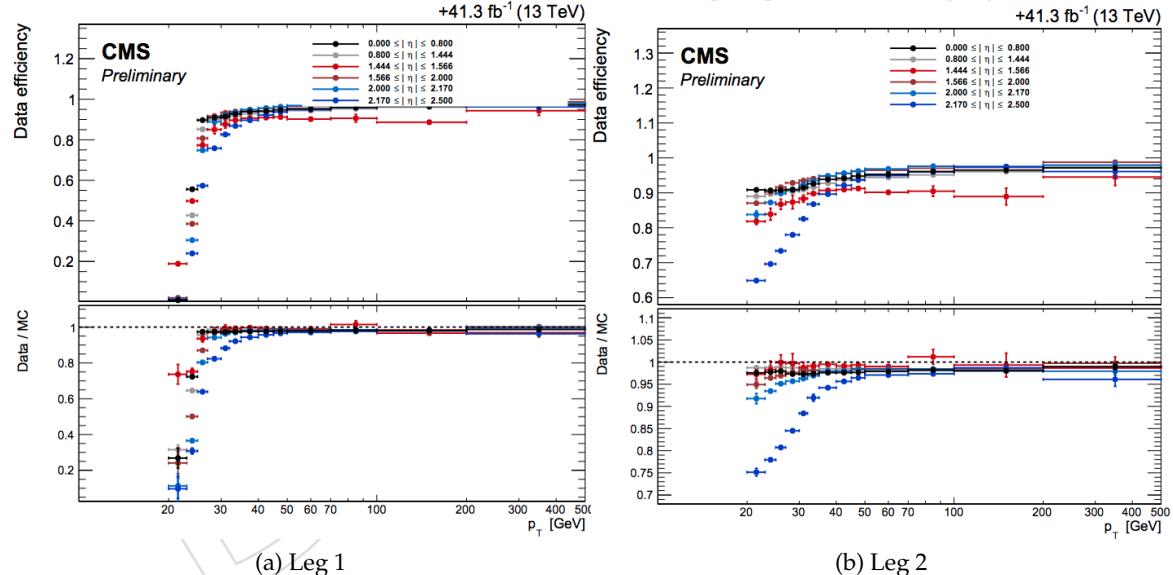


Figure 3: Distributions of HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ efficiency as function of p_T and η for 2017 data. The efficiencies are measured after applying WP90 in the general purpose electron MVA IDs plus isolation selection.

262 mass requirement. The corresponding scale factors are shown in Fig. 4.

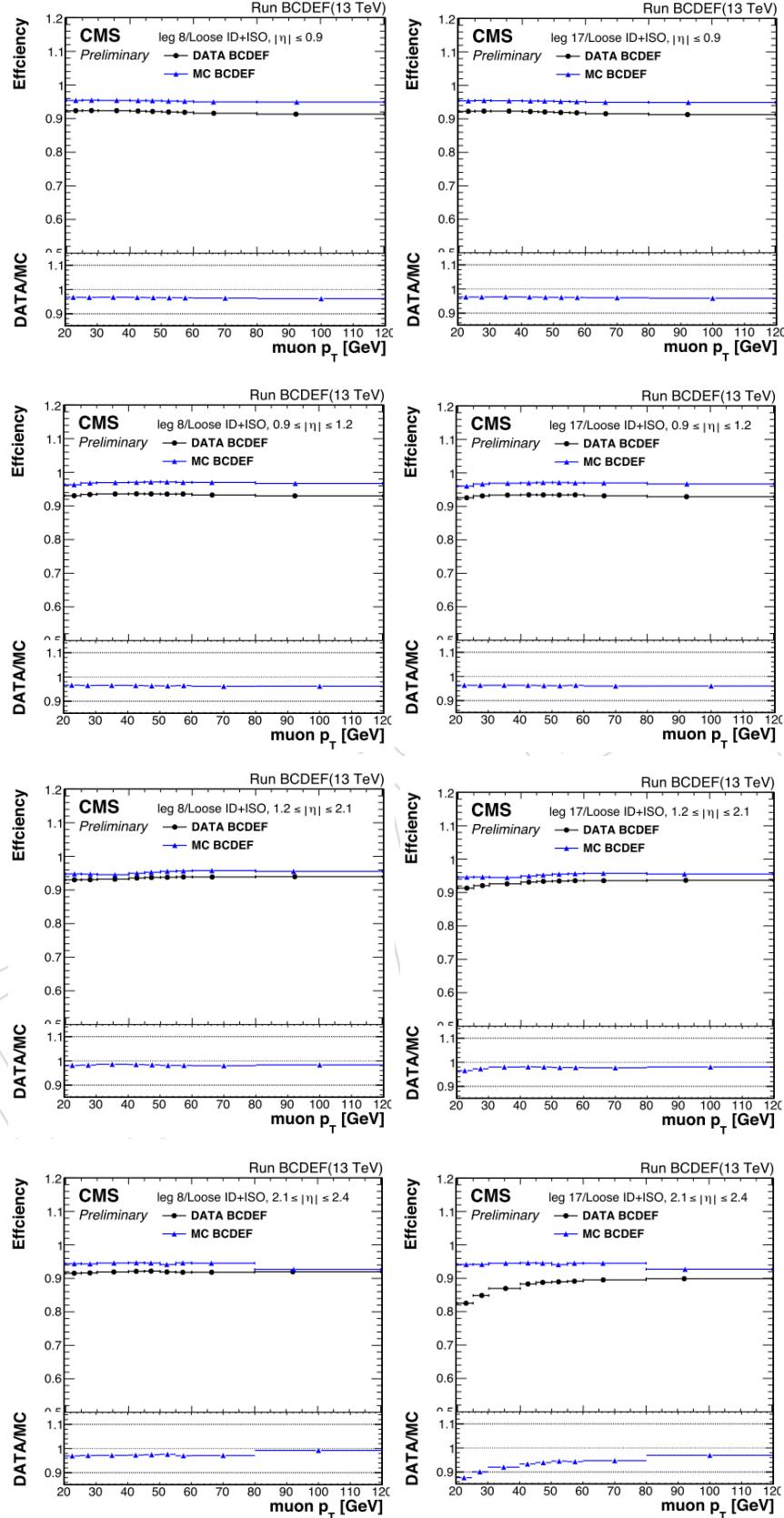


Figure 4: Distributions for each leg of the HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8_v* OR HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8_v* trigger as function of p_T and η for Run B, C, D, E, F. The left figure corresponds to the 8 GeV leg scale factor. The right figure correspond to the 17 GeV leg scale factor, requiring the second muon to pass the 8 GeV leg.

263 **3.1 Dedicated trigger for Z($\nu\nu$)H**

264 The main trigger of the Z($\nu\nu$)H channel is HLT_PFMET120_PFMHT120_IDTight which is seeded
265 at L1 by an OR of L1_ETM triggers with thresholds spanning from 100 to 120 GeV depending
266 on the instantaneous luminosity of the LHC. The trigger cuts on both PF MET and PF MHT in
267 order to reduce the rate and maintain these thresholds.

268 An additional trigger, HLT_PFMET120_PFMHT120_IDTight_PFHT60, is used to complement
269 the main trigger which was inactive for a small part of run F. The online PF MHT is evaluated
270 using PF corrected jets with $p_T > 20 \text{ GeV}$, $|\eta| < 5.2$, and passing the following ID requirements:

- 271 • Neutral Hadronic Fraction < 0.9 ;
- 272 • Neutral Electromagnetic Fraction < 0.99
- 273 • Number of Constituents > 1 ;
- 274 • Charged Hadronic Fraction > 0 ;
- 275 • Charged multiplicity > 0 .
- 276 • Charged Electromagnetic Fraction < 0.99 ;

277 The last three cuts are only applied to jets within the tracker acceptance ($|\eta| < 2.4$).

278 The HLT_PFMET120_PFMHT120_IDTight trigger cuts both on MET and MHT at the same
279 threshold. For this reason, the trigger efficiency has been studied as a function of the of-
280 fline $\min(\text{MET}, \text{MHT})$. The additional trigger HLT_PFMET120_PFMHT120_IDTight_PFHT60
281 contains an additional cut on PF HT at 60 GeV. This has been verified not to have a signifi-
282 cant impact on the measured efficiency due to the required presence of two jets. Therefore the
283 efficiency of the logical OR of the two triggers is measured.

284 The overall trigger efficiency has been measured using the data collected by the single-electron
285 triggers and requiring the presence of two jets in the tracker acceptance in the event. In order
286 to avoid bias from the L1 MET (calo-MET), the lepton is required not to be back to back with
287 the reconstructed MET.

288 Figure 5 shows the trigger efficiency for this trigger as a function of the offline $\min(\text{MET}, \text{MHT})$
289 distribution obtained in data in the single-electron dataset (left) and in a simulated W-boson+3-
290 jet sample (right). The points are fitted with the convolution of a crystal ball function and a step
291 function, which has five fit parameters.

292 An efficiency correction is derived as the ratio of the fitted functions for data and simulation.
293 Figure 6 shows this efficiency correction as a function of $\min(\text{MET}, \text{MHT})$. The uncertainty in
294 the efficiency correction measurement is determined by performing eigen-vector decomposi-
295 tions of the covariance matrices of the fitted functions for data and simulation. This allows us
296 to extract five independent uncertainties for each fitted function, and therefore determine inde-
297 pendent uncertainties in the efficiency corrections. Figure 6 (left) shows the nominal efficiency
298 correction, as well as the variation of this correction under the leading four uncertainties asso-
299 ciated with the fit of the efficiency curve in simulation. Figure 6 (right) also shows the nominal
300 efficiency correction in addition to the variation of this correction under the leading four un-
301 certainties associated with the fit of the efficiency curve in data. The uncertainties that have a
302 non-negligible effect on the shape of the final discriminant or the process normalisations are
303 taken into account as systematic uncertainties.

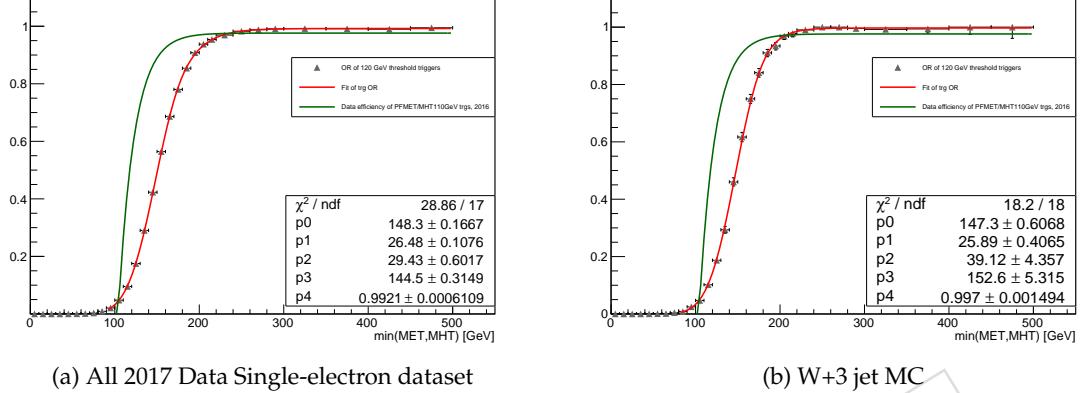


Figure 5: Distributions of trigger efficiency as function of $\min(\text{MET}, \text{MHT})$ for the data in the single-electron for 2017 full dataset is shown on the left. The same measurement is shown in MC on the right.

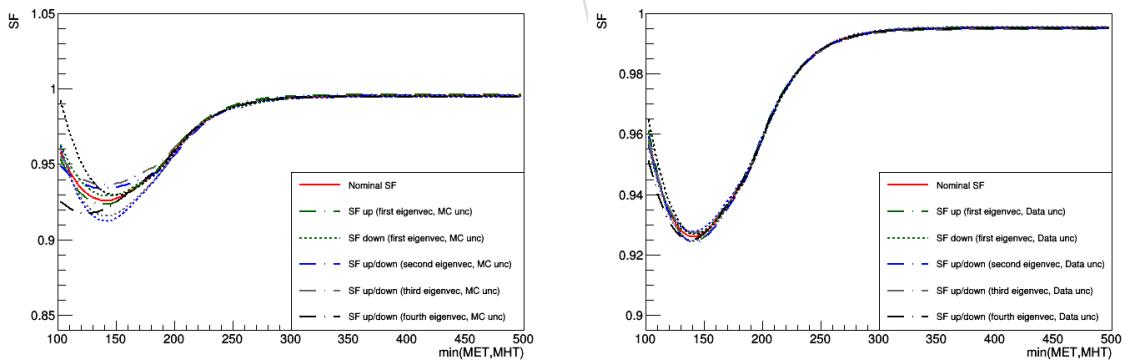


Figure 6: The efficiency correction applied to simulation is shown as a function of the $\min(\text{MET}, \text{MHT})$. In addition to the nominal efficiency correction, the variation in the correction due to uncertainties in the function fitted to the efficiency in MC (left) and in data (right) are shown. The efficiency correction is very close to 1, except in the turn-on where the difference in efficiency between data and simulation is up to 8%.

304 4 Physics Objects

305 This analysis takes as input the miniAODv2 data format of all the Data and Monte Carlo sam-
 306 ples listed in Section 2. In a first step, the events are processes applying the Jet Energy Correc-
 307 tions and the Jet Energy Regression described in this section, and storing the output in the ultra
 308 slim data tier called NanoAOD. In a second step, called post-processor, a python framework
 309 is used to select standard physics objects and recipes approved and provided by the relevant
 310 POGs. The post-processor also applies a first loose selection and classification of the events,
 311 producing plain ROOT trees for the final part of the analysis.

312 Sections 4.1–4.7 describe the reconstruction, identification, and selection of electrons, muons,
 313 jets, b-jets, and missing transverse momentum. Since pile-up affects all of these physics objects,
 314 we begin with a description of the primary vertex selection and the methods applied to mitigate
 315 the effects of pile-up.

316 4.1 Primary vertex selection and pile-up treatment

317 Primary vertices positions where proton-proton collisions occurred are reconstructed using
 318 tracks clustered with the Deterministic Annealing algorithm [53]. We use the miniAOD
 319 “offlineSlimmedPrimaryVertices” collection. Reconstructed primary vertices are re-
 320 quired to be valid, to have a z position within 24 cm of the nominal detector center, a radial
 321 position within 2 cm of the beamspot axis, and a vertex fit exceeding four degrees of freedom.

322 During the run-1, the signal vertex was chosen as the primary vertex with the largest $\sum_{\text{tracks}} p_T^2$.
 323 This algorithm has been improved during the Long Shutdown 1. The idea is to choose the
 324 vertex depending on the $\sum p_T^2$ of the elementary particles produced in the collision (quarks,
 325 gluons, charged leptons, photons, and neutrinos). To do that, the $\sum p_T^2$ is evaluated for each
 326 primary vertex using track-jets, track-MET, and charged leptons, instead of tracks. Moreover,
 327 displaced tracks originating from a B hadron decay are associated to the proper vertex exploit-
 328 ing the respective jet direction. This improved algorithm increases the probability to choose the
 329 correct vertex as signal vertex especially in $Z(\nu\nu)H$ channel, where only two b -jets are present
 330 in the final state.

331 Given the collisions’ instantaneous luminosity the data sample contains a significant number
 332 of additional interactions per bunch crossing—an effect known as pile-up (PU). The number of
 333 reconstructed primary vertices is related to the number of PU interactions in each triggered
 334 event, with a primary vertex reconstruction efficiency that is roughly around 70%. The pile-up
 335 distribution in collected data decreases as a function of time during each LHC fill and varies
 336 along the year because of different LHC collisions settings.

337 Over the course of the LHC operation in 2017, in-time PU as well as out-of-time pile-up in-
 338 creased. Effects from out-of-time pile-up are particularly more relevant after the LHC switched
 339 to bunch filling schemes with 25 ns spacing (as opposed to 50 ns minimum spacing employed
 340 throughout Run1 collisions). Near the end of 2017 data-taking LHC introduced luminosity
 341 “leveling”. The beams were separated and gradually brought together such that average PU
 342 was essentially constant (near 60) until the beams were once again head-on. Afterwards the
 343 average PU gradually declined over the course of the rest of the fill. Figure 7 shows the distri-
 344 bution of the average jet energy density ρ (highly correlated with PU) in the $t\bar{t}$ control region
 345 for $W(\mu\nu)H$ (see Sec. 11).

346 The presence of PU interactions affects the reconstruction of jets in general, in particular the
 347 jet momentum resolution and the reconstruction of the Higgs mass, and the use of vetoes on
 348 additional jet activity. It also affects the MET reconstruction, lepton isolation and b-tagging.

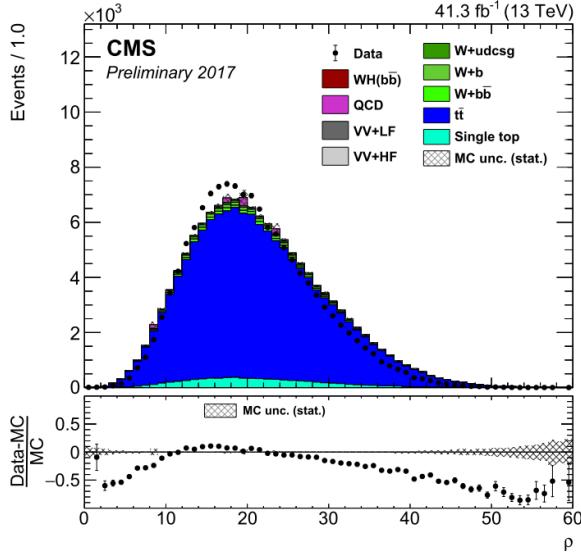


Figure 7: Distribution of the average jet energy density, ρ , which highly correlated with PU. Data are compared to simulation in the $t\bar{t}$ control region for $W(\mu\nu)H$. See Sec. 11 for the definition of this control region.

349 There are two distinct approaches to address all these effects (apart from MET):

- 350 • **PFnoPU:** also known as Charged Hadron Subtraction (CHS), PFPU is an algorithm
351 embedded in the PF jet processing chain that attempts to filter all charged hadrons
352 that do not appear to originate from the primary interaction. This approach is very
353 effective but only works in the pseudo-rapidity region covered by the Tracker. Al-
354 gorithms for tagging b jets are not impacted, since they apply their own track pre-
355 filtering that is also designed to be PU-resistant.
- 356 • **Fastjet:** is an external software package from which CMSSW takes virtually all its
357 jet reconstruction services [54]. In particular it provides the means to calculate the
358 momentum density per unit area, ρ , due to PU for each event, which can be used to
359 subtract the contamination of jets and lepton isolation cones based on their respec-
360 tive areas. These methods are therefore referred to as "Fastjet Subtraction."

361 Ideally, charged hadrons from PU interactions are filtered from the event first before the appli-
362 cation of Fastjet. In this analysis, both the PFnoPU and Fastjet Subtraction methods are applied
363 consistently in the reconstruction and identification of jets, and in the calculation of lepton
364 isolation.

365 The Run II Fall17 Monte Carlo samples, reconstructed with CMSSW_9_4 were generated with
366 the so-called

2017_25ns_WinterMC_PUScenarioV1_Poisson00TPU

367 PU-distribution. That PU distribution was a wide-coverage PU scenario to account for a range
368 of potential effective Minimum Bias cross sections. Changes the generator were not understood
369 at the time and so a range of options were considered equal. In order to make MC match
370 data we must reweight the simulated events by the number of generated PU. The standard
371 reweighting technique [55] is used in this analysis. We use a central value of 69.2 mb for the
372 total inelastic pp cross section, and we estimate the systematic uncertainties on the weights by
373 changing the central value to 66.02 and 72.38 mb, respectively.

374 4.2 Electrons

375 Electrons are reconstructed with the Gaussian Sum Filter algorithm (GSF Electrons) [56]. They
 376 are preselected by requiring $p_T > 7\text{ GeV}$, $|\eta| < 2.4$, $d_{xy} < 0.05\text{ cm}$, $d_z < 0.2\text{ cm}$ (where both
 377 distances are taken with respect to the primary vertex), and a very loose relative isolation cut
 378 of 0.4, where the ρ -subtracted PF isolation in a cone of radius 0.3 is used (Sec. 4.4).

379 A tighter identification is then applied using a multivariate approach recommended by the
 380 electron-gamma (EGM) POG as documented here:

381 <https://twiki.cern.ch/twiki/bin/viewauth/CMS/MultivariateElectronIdentificationRun2>.

382 A general purpose multivariate discriminator is trained for electrons that pass a set of cuts
 383 meant to include all electrons desired for standard analyses. A set of offline cuts on ECAL-
 384 based electron quantities is applied on top of the multivariate discriminator to reproduce the
 385 conditions of the training sample:

```
386 pt>15 & (
387   (abs(superCluster().eta)<1.4442 & full5x5_sigmaIetaIeta<0.012 &
388   hcalOverEcal<0.09 &
389   (ecalPFClusterIso/pt)<0.4 & (hcalPFClusterIso/pt)<0.25 &
390   (dr03TkSumPt/pt)<0.18 & abs(deltaEtaSuperClusterTrackAtVtx)<0.0095 &
391   abs(deltaPhiSuperClusterTrackAtVtx)<0.065) ||
392   (abs(superCluster().eta)>1.5660 & full5x5_sigmaIetaIeta<0.033 &
393   hcalOverEcal<0.09 &
394   (ecalPFClusterIso/pt)<0.45 & (hcalPFClusterIso/pt)<0.28 &
395   (dr03TkSumPt/pt)<0.18)
396 ).
```

397 Two cuts on the MVA ID discriminator [57] are applied defining two different working points
 398 based on the expected selection efficiency of either 90% (loose, WP90) or 80% (tight, WP80).

399 The loose WP90 working point is used for global event classification (based on vector boson
 400 type and decay), in the counting of additional leptons for the veto requirement, and in the event
 401 selection of the Z(ee)H channel.

402 The tighter WP80 working point is used in the W($e\nu$)H channel to suppress the fake back-
 403 ground in that final state. The p_T threshold in the W($e\nu$)H channel is 30 GeV. For the Z(ee)H
 404 channel thresholds are looser at 25 GeV and 17 GeV for the two electrons.

405 For 2017 analysis we use mvaEleID-Fall17-GeneralPurpose-V1-wpXX working points
 406 with channel specific isolation cuts (0.06 for W($e\nu$)H). Working points and isolation cuts for
 407 2-lepton channels are generally looser because requiring two leptons virtually eliminates all
 408 QCD background. Whereas in the 1-lepton channels tighter cuts are required to remove QCD
 409 background.

410 4.3 Muons

411 Muons are reconstructed from combined tracker and muon-chamber information (global muons) [58,
 412 59] They are preselected by requiring the loose muon POG ID (see below), $p_T > 5\text{ GeV}$, $|\eta| <$
 413 2.4 , $d_{xy} < 0.5\text{ cm}$, $d_z < 1.0\text{ cm}$ (where both distances are taken with respect to the primary ver-
 414 tex), and a very loose relative isolation cut of 0.4, where the $\Delta\beta$ -subtracted PF isolation in a cone
 415 of radius 0.4 is used (Sec. 4.4). They are further required to pass standard criteria suggested by
 416 the Muon POG. Two WP are used, a loose and tight:

- 417 • Loose muon:
- 418 • Particle-Flow Muon:
 419 `isPfMuon()`
- 420 • is Global or Tracker Muon:
 421 `isGlobalMuon() || isTrackerMuon()`
- 422 • Tight muon:
- 423 • the candidate is reconstructed as a Global Muon:
 424 `isGlobalMuon()`
- 425 • Particle-Flow Muon:
 426 `isPfMuon()`
- 427 • $\chi^2/ndof$ of the global-muon track fit:
 428 `globalTrack() -> normalizedChi2() < 10.`
- 429 • at least one muon-chamber hit included in the global-muon track fit:
 430 `globalTrack() -> hitPattern().numberOfValidMuonHits() > 0`
- 431 • muon segments in at least two muon stations; this implies that the muon
 432 is also an arbitrated tracker muon:
 433 `numberOfMatchedStations() > 1`
- 434 • tracker track transverse impact parameter w.r.t. the primary vertex:
 435 `fabs(muonBestTrack() -> dxy(vertex->position())) < 0.2`
- 436 • longitudinal distance of the tracker track wrt. the primary vertex:
 437 `fabs(muonBestTrack() -> dz(vertex->position())) < 0.5`
- 438 • number of pixel hits:
 439 `innerTrack() -> hitPattern().numberOfValidPixelHits() > 0`
- 440 • cut on number of tracker layers with hits:
 441 `innerTrack() -> hitPattern().trackerLayersWithMeasurement() > 5`

443 The p_T threshold in the $W(\mu\nu)H$ channel is 25 GeV. For the $Z(\mu\mu)H$ channel thresholds are
 444 looser at 25 GeV and 15 GeV for the two muons.

445 4.4 Lepton isolation

446 Lepton isolation is the main handle for selecting prompt muons and electrons produced in
 447 the electroweak decay of massive particles such as Z or W bosons and for rejecting the large
 448 number of leptons produced in jets through the decay of heavy-flavour hadrons or the decay in
 449 flight of charged pions and kaons. The isolation is quantified by estimating the total p_T of the
 450 particles emitted around the direction of the lepton. The particle-flow based isolation relative
 451 to the lepton p_T is defined as [9, 60]

$$I_{PF} \equiv \frac{1}{p_T^\ell} \left(\sum p_T^{\text{charged}} + \max \left[0, \sum p_T^{\text{neutral}} + \sum p_T^\gamma - p_T^{\text{PU}}(\ell) \right] \right) \quad (1)$$

452 where the sums run over the charged hadrons (p_T^{charged}), photons (p_T^γ), and neutral hadrons
 453 (p_T^{neutral}) with a distance ΔR to the lepton smaller than 0.3 (0.4) in the (η, φ) plane around the
 454 electron (muon) momentum.

455 The $\sum p_T^{\text{charged}}$ is the scalar sum of the transverse momenta of charged hadrons originating
 456 from the chosen primary vertex of the event. The $\sum p_T^{\text{neutral}}$ and $\sum p_T^\gamma$ are the scalar sums of the
 457 transverse momenta for neutral hadrons and photons, respectively. Since the isolation variable

is particularly sensitive to energy deposits from pileup interactions, a $p_T^{\text{PU}}(\ell)$ contribution is subtracted, using two different techniques. For muons, we define $p_T^{\text{PU}}(\mu) \equiv 0.5 \times \sum_i p_T^{\text{PU},i}$, where i runs over the momenta of the charged hadron PF candidates not originating from the primary vertex, and the factor of 0.5 corrects for the different fraction of charged and neutral particles in the cone. For electrons, the FASTJET technique [54] is used, in which $p_T^{\text{PU}}(e) \equiv \rho \times A_{\text{eff}}$, where the effective area A_{eff} is the geometric area of the isolation cone scaled by a factor that accounts for the residual dependence of the average pileup deposition on the η of the electron, and ρ is the median of the p_T density distribution of neutral particles within the area of any jet in the event.

Both muon and electron (WH) channels are required to have a relative isolation smaller than 0.06. They are the same by chance. In both cases the cuts were tightened to remove excess of data in poorly isolated distribution tails. Although the cuts are tight there is no loss in expected sensitivity and the data/mc agreement is very good in the bulk of the isolation distribution. In the $Z(\mu\mu)\text{H}$ ($Z(\text{ee})\text{H}$) channel, the threshold is relaxed to 0.12 (0.15).

4.5 Jets

Jets are reconstructed from particle-flow candidates [61, 62], using the anti- k_T clustering algorithm with distance parameter $R = 0.4$ [63, 64]. Reconstructed jets require a small additional energy correction, mostly due to thresholds on reconstructed tracks and clusters in the PF algorithm and various reconstruction inefficiencies [65]. We use the recommended JET/MET corrections described in the twiki

https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC#Jet_Energy_Corrections_in_Run2.

In particular, for the 2017 datasets we use the corrections

Fall17_17Nov2017BCDEF_V6_DATA Fall17_17Nov2017_V6_MC

for data and MC respectively. Loose jet identification criteria are also applied to reject misreconstructed jets resulting from detector noise, as well as jets heavily contaminated with pileup energy (clustering of energy deposits not associated with a parton from the primary pp interaction) [66]. Jets that overlap geometrically ($\Delta R < 0.4$) with preselected electrons or muons are discarded. Jets calibrated using the official JME POG prescriptions are considered for the analysis if $|\eta| < 2.5$. In the WH and $Z(\nu\nu)\text{H}$ channels a minimum threshold of $p_T > 25\text{ GeV}$ is used, while a looser selection ($p_T > 20\text{ GeV}$) is applied in the cleaner $Z(\ell\ell)\text{H}$ channels with lower dijet boost.

After applying jet energy scale corrections and a b-jet regression, jet energies must be smeared in order for simulation to agree with the resolution of equivalent jets in data, as described in Sec. 6.1.

4.6 Identification of b-jets

The identification of jets that originate from the hadronization of b quarks for 2017 data is performed with the output of a deep neural network (DNN) called Deep Combined Secondary Vertex (DeepCSV) [33]. This DNN has several probability outputs for different jets resulting from quarks of various flavors. In this note DeepCSV will generally refer to the summed probability that a jet originated from a b-quark or from 2 b-quarks:

```
bDiscriminator('pfDeepCSVJetTags:prob') + bDiscriminator('pfDeepCSVJetTags:probbb')
```

The DeepCSV algorithm provides a continuous discriminator output combining in an optimal way the information about track impact parameters and identified secondary vertices within

499 jets, even when full vertex information is not available, and information of any soft lepton
500 present in the jet.

501 The DeepCSV output that can be used to select optimal working points with respect to the VH
502 analyses, in addition to the standard Loose/Medium/Tight working points defined by the BTV
503 POG and are listed in Table 10. Independent optimizations of the selection criteria in all five
504 channels arrive at roughly the same optimal selection for the jet in the Higgs decay; that is, the
505 leading b-jet tends to pass selection tighter than the tight WPs. For the second jet, the optimal
506 selection typically falls between the loose and medium working points.

Table 10: Definition of b-tagging working points (WP). The WPs are selected such that a specific target for the mis-tag is achieved. That efficiency target is listed in the table per WP.

WP Name	Mistag efficiency	CSVv2	CMVA	DeepCSV
Loose (2017)	10%	0.5803	n/a	0.1522
Medium(2017)	1%	0.8838	n/a	0.4941
Tight (2017)	0.1%	0.9693	n/a	0.8001
Loose (2016)	10%	0.5426	-0.5884	0.2219
Medium(2016)	1%	0.8484	0.4432	0.6324
Tight (2016)	0.1%	0.9535	0.9432	0.8958

507 Although the b-tagging discriminator used for the 2016 analysis was the CMVAv2 in HIG-16-
508 044 [30], the DeepCSV algorithm has been back-ported and its score is present in the 2016 ntuples,
509 playing an important role in the training of the final signal vs background discriminator
510 described in Sec. 9.

511 4.6.1 DATA/MC reshaping of the b-jet discriminator

512 The calibration of the DeepCSV discriminator is determined using the recommended tag-and-
513 probe method as documented in Ref. [67]. This method, originally developed in the context
514 of ttH(bb) and applied to the previous version of the VH(bb) analysis [32], attempts at cor-
515 recting the distribution of the DeepCSV discriminator for simulated jets as to match the dis-
516 tribution observed in data control regions, extending the usual integrated SF approach. These
517 control regions are preselected by the requiring at least two opposite-sign leptons plus at least
518 two jets. Two exclusive set of selections based on the dilepton mass and the DeepCSV dis-
519 criminator or a “tag” jet are further imposed to enrich the control regions in Z+jets or t̄t, re-
520 spectively. The binned DeepCSV distribution of the “probe” jet is then compared to the one
521 expected from simulation. An iterative procedure is then initiated to scale every bin content
522 in the simulation simultaneously for light and heavy flavour jets (`hadronFlavour()=0` and
523 `hadronFlavour()=5`, respectively). This procedure is carried out for various p_T and $|\eta|$ bins.
524 The ratio between the re-scaled distribution in the simulated sample and the original one is
525 used as a jet-by-jet weight, w_j , defined as

$$w_j(\text{DeepCSV}_j; p_{T,j}, |\eta_j|, \text{flavour}_j). \quad (2)$$

526 By construction, the weights average to one when sampling DeepCSV using the distribution
527 predicted by the simulation: $N^{-1} \sum_{i=1}^N w_j \rightarrow 1$, for $N \gg 1$ jets. In an event with N_{jet} selected
528 jets whose DeepCSV discriminator is used in the analysis, an event weight w is defined starting
529 from the jet weights as:

$$w = \prod_{j=1}^{N_{\text{jet}}} w_j(\text{DeepCSV}_j; p_{T,j}, |\eta_j|, \text{flavour}_j). \quad (3)$$

530 **4.7 Missing transverse energy**

531 The use of missing transverse energy is central to the analyses presented in this note. It is
532 critical in the reconstruction of the $W \rightarrow \ell\nu$ and $Z \rightarrow \nu\bar{\nu}$ decays, and is used in the $Z(\mu\mu)H$
533 channel as a control sample to evaluate the comparison of data and simulation in the $Z(\nu\nu)H$
534 channel. It is also used in the $Z(\ell\ell)H$ to increase the purity of the $t\bar{t}$ control sample. For the
535 offline analysis, missing transverse energy is computed from the list of particle-flow objects
536 with the method described in Ref. [68]. The vector \vec{E}_T^{miss} is calculated as the negative of the
537 vectorial sum of transverse momenta of all particle-flow objects identified in the event, and
538 the magnitude of this vector is referred to as “pfMET”. The estimation of the \vec{E}_T^{miss} vector in
539 simulated events is improved by correcting it for the difference between raw (i.e. uncorrected)
540 and calibrated jets (Sec. 4.5), including the scale and resolution corrections, with $p_T >$
541 15 GeV , $|\eta| < 4.7$, and passing a set of filters meant to remove electron and muon candidates as
542 explained in

<https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETRun2Corrections>.

543 In addition, we apply a set of recommended filters to remove known issues of instrumental
544 noise and problematic events as explained in

<https://twiki.cern.ch/twiki/bin/view/CMS/MissingETOptionalFiltersRun2>.

545 The pfMET value divided by the scalar sum of E_T of all particle-flow objects is referred to as the
546 “pfMET significance.”. For the $Z(\nu\nu)H$ analysis, pfMET is required to be greater than 150 GeV .
547 In $Z(\nu\nu)H$ analysis, pfMHT is used too. It is defined as: $\text{pfMHT} = \sum_i \vec{p}_T(\text{jet})$, considering jets
548 with $p_T > 30$ and $|\eta| < 2.4$.

549 **4.8 Additional “soft” hadronic activity**

550 In the signal VH events, not much additional hadronic activity is expected after excluding the
551 V and H decay products. Since the amount of additional radiation is expected to be soft, and to
552 avoid the contributions from pileup interactions, only the charged tracks that clearly originate
553 from the event main interaction point (see Sec. 4.1) are used to monitor this event property.

554 The additional soft activity is defined as follows. First, a collection of “Additional tracks” is
555 built using reconstructed tracks that:

- 556 • have a *high purity* quality flag,
- 557 • have $p_T > 300 \text{ MeV}/c$,
- 558 • are not associated to the vector decay leptons, nor to the selected two b-jets in the
559 event (through the PF candidates components track references),
- 560 • make minimum $|d_z(\text{PV})|$ when associated to the main primary vertex (PV) in the
561 event,
- 562 • satisfy $|d_z(\text{PV})| < 2 \text{ mm}$ with respect to the main PV in the event.

563 In addition to removing the tracks in the jets, also tracks in the region between the two b-jets
564 are removed by defining an ellipse in the (η, ϕ) plane around the two b-jets with axes $(a; b) =$
565 $(\Delta R(\text{bb}) + 1; 1)$, and excluding all tracks pointing within the ellipse from the additional tracks
566 collection.

567 After this track selection, a collection of “soft track-jets” is build clustering the “Additional
568 tracks” collection with the anti- k_T clustering algorithm [64] with distance parameter $R = 0.4$.

569 The use of track-jets represents a clean and commissioned method [69] to reconstruct the had-
570 ronization of partons with very low energies, down to few GeV [70].

571 For the purpose of separating the signal from backgrounds, we make use of clustered soft
572 TrackJets, and eventually consider

- 573 • the scalar $p_{T\text{sum}}$ of the soft TrackJets with transverse momentum $p_T > 1$ GeV, H_T^{soft} ;
574 • the soft TrackJet multiplicity N^{soft} with transverse momentum $p_T > 2$ GeV, N_2^{soft} ;
575 • the soft TrackJet multiplicity N^{soft} with transverse momentum $p_T > 5$ GeV, N_5^{soft} ;
576 • the soft TrackJet multiplicity N^{soft} with transverse momentum $p_T > 10$ GeV, N_{10}^{soft} ;

577 The soft hadronic activity is used as discriminating variable in the regression for all the chan-
578 nels, as discussed in Section 9.

DRAFT

5 Vector Boson Reconstruction

- Reconstruction of W and Z bosons begins with the identification and selection of charged leptons and pfMET described in the previous section. Given the unique signature of a boosted vector boson recoiling from two jets, the dominant background is from real W and Z decays. Therefore, a minimal selection is sufficient to identify highly pure samples of V+jets events.
- Candidate $Z \rightarrow \ell\ell$ decays are reconstructed by combining isolated electrons (`mvaSpring16GP_WP90` and `pfRelIso03_all < 0.15`) and muons (`pfRelIso04_all < 0.25`) with $p_T > 20$ GeV, and requiring the dilepton invariant mass to satisfy $75 < M_{\ell\ell} < 105$ GeV. This analysis is performed in two bins of $p_T(V)$. The "low $p_T(V)$ " category where $50 < p_T(Z) < 150$ GeV and the "high $p_T(V)$ " category where $p_T(Z) > 150$ GeV.
- Candidate $W \rightarrow \ell\nu$ decays are identified primarily by the topology of a single isolated electron (`mvaSpring16GP_WP80` and `pfRelIso03_all < 0.12`) or muon (`pfRelIso04_all < 0.15`) with $p_T > 25$ GeV. The transverse momentum $p_T(W)$ and mass M_T of the W candidate are computed as:

$$p_T(W) = \sqrt{(p_{\text{fMET}}^x + p_x^\ell)^2 + (p_{\text{fMET}}^y + p_y^\ell)^2}, \quad \text{and} \quad (4)$$

$$M_T = \sqrt{(p_{\text{fMET}} + p_T^\ell)^2 - p_T(W)^2}. \quad (5)$$

It is observed that in the boosted regime, where the QCD background is much reduced, simply requiring $p_T(W) > \sim 100$ GeV is sufficient to select a relatively clean sample of real W decays. This analysis is performed in one category with $p_T(W) > 150$ GeV (**).

For inclusive W production, the distribution of M_T reflects the characteristic Jacobian peak and is very effective at separating signal from the large background of generic QCD production at small values of the transverse mass. In contrast, for the high boost used in this analysis, the neutrino begins to overlap with the lepton in azimuth, creating a broad flat region in M_T between 0–50 GeV that reduces the effectiveness of this variable in rejecting QCD background. This feature can be appreciated in Fig. 8, where the distribution below the peak is flat. Therefore, no selection is applied on M_T in the reconstruction of W candidates in the signal region. However, M_T remains effective at reducing QCD background in the low-boost W + bb control region (see Sec. 11), and in generally cleaning up the background in all of the control regions for the electron mode.

Candidate $Z \rightarrow \nu\bar{\nu}$ decays are reconstructed simply requiring $\text{pfMET} > 150$ GeV. The transverse momentum of the Z candidate is defined as $p_T(Z) = \min(\text{pfMET}, \text{pfMHT})$. The $Z(\nu\nu)H$ analysis is performed in one $p_T(Z)$ category with $p_T(Z) > 170$ GeV.

5.1 Vector boson transverse momentum reweighting in V+jets events

In an inclusive selection enriched in vector boson production a shape difference in the vector boson transverse momentum of the simulated LO samples with respect to data has been observed, both in W and Z events. The observed data has a softer spectrum than the data and a negative correction with increasing $p_T(V)$ is necessary to correct for the effect.

This negative correction is expected to stem from higher order QCD and electroweak corrections to the vector boson production. The former correction is derived for the different H_T bins while the latter is derived as a function of boson $p_T(V)$ at generator level. The correction factors are multiplicative factors, more details could be found in Ref. [71].

(**) New with respect to the previous Run2 analysis on 2016 data described in HIG-16-044 [30] and AN-15-168 [32].

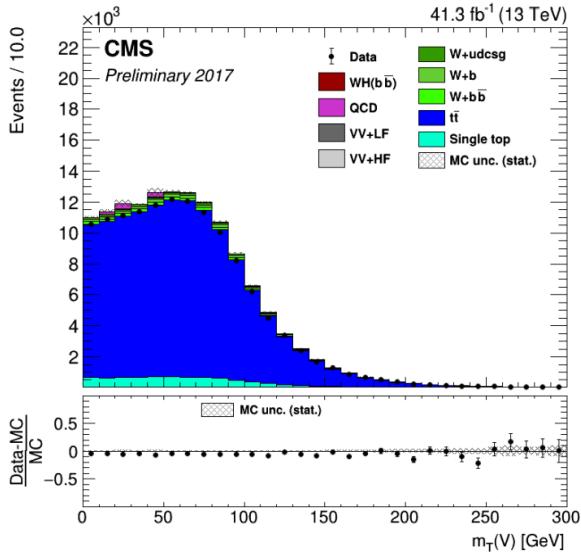


Figure 8: M_T distribution in the $t\bar{t}$ electron control region for the 1-lepton (see Sec. 11). It can be observed that there is broad flat region in M_T between 0–50 GeV.

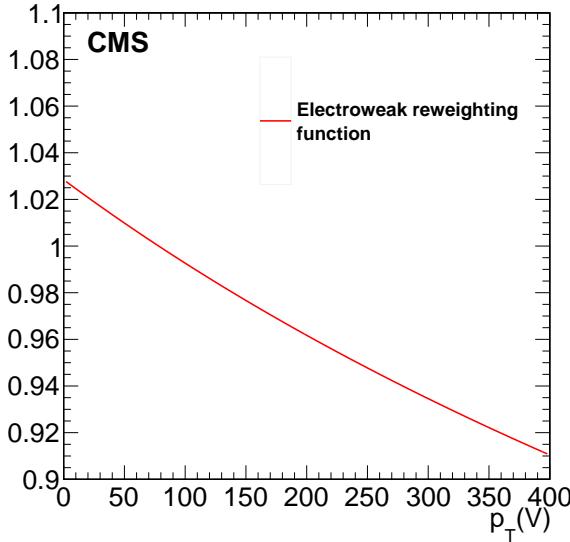


Figure 9: Electroweak correction as a function of boson $p_T(V)$ for the V+jets samples.

618 The function used for the electroweak correction is shown in Fig. 9.

619 5.2 EWK signal VH corrections

620 The signal Monte Carlo sample for $qqVH$ is produced with POWHEG+MiNLO and then rescaled
621 to NNLO QCD. The total cross section σ^{VH} is given by [11]:

$$\sigma^{WH} = \sigma_{NNLOQCD}^{WH,DY}(1 + \delta_{EW}) + \sigma_{t\text{-loop}} + \sigma_\gamma, \quad (6)$$

$$\sigma^{ZH} = \sigma_{NNLOQCD}^{ZH,DY}(1 + \delta_{EW}) + \sigma_{t\text{-loop}} + \sigma_\gamma + \sigma^{ggZH} \quad (7)$$

622 Since up to NLO, the electroweak corrections factorize, they can be applied differentially (in
623 $p_T(V)$) as multiplicative weight $(1 + \delta_{EW})$ to the signal. Figure 10 shows the shape of these
624 corrections for two of the channels.

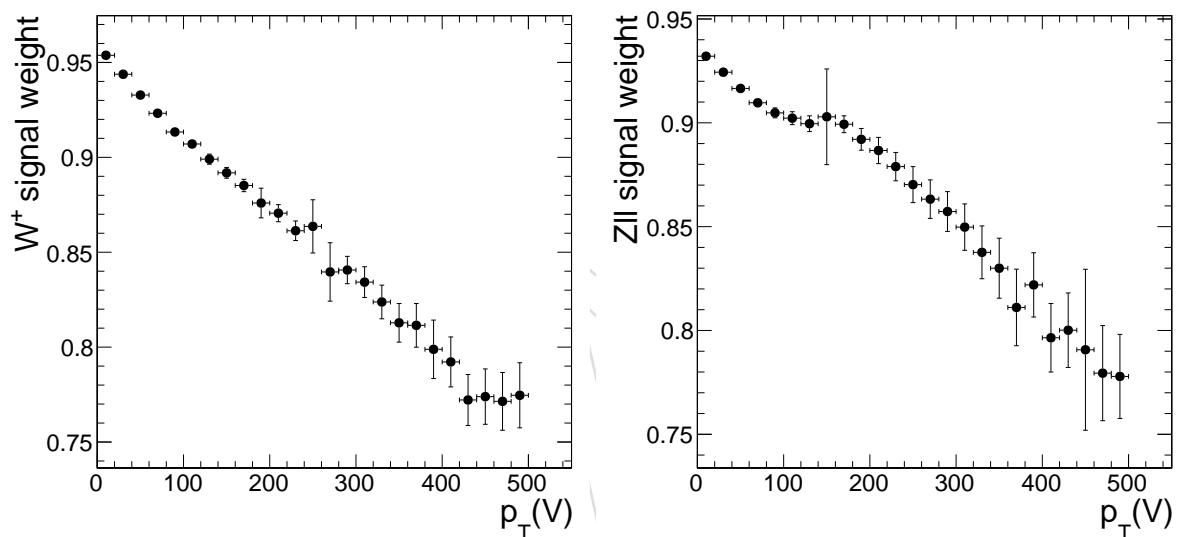


Figure 10: Multiplicative weights to apply the differential NLO electroweak signal correction for $W^+ H$ (left) and $Z H$ (right) processes.

6 Higgs Boson Reconstruction

The reconstruction of the $H \rightarrow b\bar{b}$ decay is a fundamental element of this analysis. In order to maximise the signal sensitivity it is critical to achieve the best possible dijet mass resolution for the Higgs candidate, while optimizing the correct selection of the two signal jets, rejecting wrong combinations.

Generally speaking, requiring a boost for the dijet system (or equivalently on the vector boson), or a high threshold separately on the p_T of each jet, improves the dijet mass resolution. However, selecting $H \rightarrow bb$ candidates by identifying the dijet combination with the largest associated b-tagging discriminate values ensures higher efficiency. This feature has been studied extensively in previous versions of the analysis (cfr Sec. 6 of AN-15-168 [32]). The signal efficiency of this selection is around 20% for events with $p_T(V) \sim 0$ and increases up to 80% for $p_T(V) > 250$ GeV.

The final Higgs candidates are the result of this selection with additional improvements on the invariant mass resolution due to a novel b-jet energy regression procedure for the three channels (**), a kinematic fit to exploit the absence of true missing transverse energy of the event in the 2-lepton channel (**), and a procedure to recover final state radiation jets (**).

6.1 Regression

Regression techniques to improve significantly the resolution of the dijet invariant mass have been developed at the Tevatron [72] and further optimized at LHC [30]. The regression correction is computed for individual jets, attempting to accurately estimate the true b-jet energy, thus improving the resolution and reducing the scale bias on a per-jet basis. The resulting improvement in the dijet mass resolution is approximately 15–20% depending on the channel. All searched channels use a common set of input variables for the b-jet energy regression as explained in the following.

Compared to the previous version of the VH analysis [30], the regression has been completely re-designed making use of a Deep Neural Network architecture and providing an estimate of the final b-jet energy resolution.

6.1.1 Configuration and training

A neural network with 6 hidden layers is used to perform the regression. The target in the training is the ratio between the generator level p_T of jets including neutrinos (i.e., the p_T of the “genJet” associated with the reco jet) and the reconstructed jet p_T after JEC corrections. The regression is trained on $t\bar{t}$ events in order to avoid biases towards the signal properties. Only b-jets identified by matching at parton level with $p_T > 20$ GeV, $|\eta| < 2.4$ are used. Additionally, the p_T distribution is mildly reweighted, thus reducing the bias towards the natural p_T spectrum for b-jets coming from top decays and stabilizing the performance for high p_T jets. Further details on the training, and full validation of the tool are contained in AN-2018/092 [73].

The training used in the regression is agnostic about the data and Monte Carlo jet energy resolution scale factors, an estimate of which is provided a posteriori.

The complete set of variables used in the regression training are:

- Jet kinematics:
 - Jet raw p_T – transverse momentum of the jet without corrections;

(**) New with respect to the previous Run2 analysis on 2016 data described in HIG-16-044 [30] and AN-15-168 [32].

- 666 • η – pseudo-rapidity of the jet;
- 667 • Jet mass;
- 668 • M_T – transverse mass of the jet without corrections;
- 669 • Leading track p_T – transverse momentum of the leading track in the jet;
- 670 • Jet momentum dispersion (p_{tD}) – $\sqrt{\sum_i(p_{T,i}^2) / \sum_i(p_{T,i})}$;
- 671 • Jet vertex information:
 - 672 • vtx3dL – 3-d flight length of the jet secondary vertex;
 - 673 • vtx3deL – error on the 3-d flight length of the jet secondary vertex;
 - 674 • vtxMass – mass of the jet secondary vertex;
 - 675 • vtxPt – transverse momentum of the jet secondary vertex;
 - 676 • vtxNtrk – number of tracks associated with the jet secondary vertex;
- 677 • Jet constituent properties:
 - 678 • neEmEf – energy fraction of the neutral constituents detected in the ECAL;
 - 679 • neHEF – energy fraction of the neutral constituents detected in the HCAL;
 - 680 • chEmEf – energy fraction of the charged constituents detected in the ECAL;
 - 681 • chHEF – energy fraction of the charged constituents detected in the HCAL;
 - 682 • Jet_leptonDeltaR – distance in $\eta - \phi$ space of soft lepton candidate (if present) with respect to the jet axis;
 - 683 • $\text{Jet_leptonPtRelInv}$ – magnitude of jet (raw) momentum transverse to lepton axis;
 - 684 • Jet_leptonPtRel – magnitude of lepton momentum transverse to jet axis;
- 685 • Further Jet constituent properties: ← NEW!
 - 686 • Lepton flavor (Electron, Muon, no lepton);
 - 687 • Number of jet daughters with $p_T > 0.3$ GeV;
 - 688 • Energy fractions in rings of increasing radius about the jet axis ([0,0.05], [0.05,0.1], [0.1,0.2], [0.2,0.3], [0.3,0.4]) coming from:
 - 689 • Electrons and photons;
 - 690 • Muons;
 - 691 • Other charged particles;
 - 692 • Other neutral particles;
 - 693 • ρ – average jet energy density (highly correlated with PU).

697 All variables are consistently used across all the VH channels.

698 6.1.2 Validation

699 Full details of the validation are presented in AN-2018/092 [73].

700 The results are evaluated in Z+Jets control regions where the responses of b-jets and light jets
 701 are evaluated separately. Fig. 11 shows fits to the b-jets response for data and MC in the $p_{T,j2} <$
 702 15 GeV case, pre-regression (a) and post-regression (b).

703 The most discriminating variables across all modes are kinematic, and this is due to the fact
 704 that most of the power of the regression is coming from the neutrinos present in semileptonic
 705 B decays, which appear as a mis-measurement of the b-jet energy. This effect is illustrated
 706 in Fig. 12, which shows the dijet mass in $Z(\ell\ell)H$ signal events before and after regression
 707 for the case where the b-jet contains a lepton, and when it does not. This comparison shows

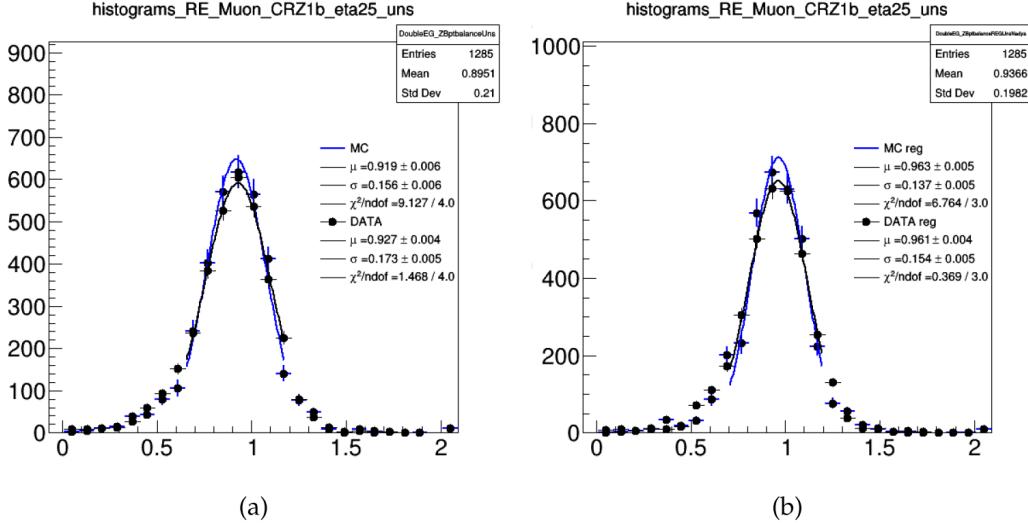


Figure 11: Fits to the b-jets response for data (black) and MC (blue) in the $p_{T,j2} < 15$ GeV case, pre regression (a) and post-regression (b). See AN-2018/092 [73] for details.

708 essentially that all of the improvement comes from the subset of events where the b-jet contains
 709 an identified lepton from semileptonic B decay. Therefore soft lepton variables are included in
 710 the regression across all modes.

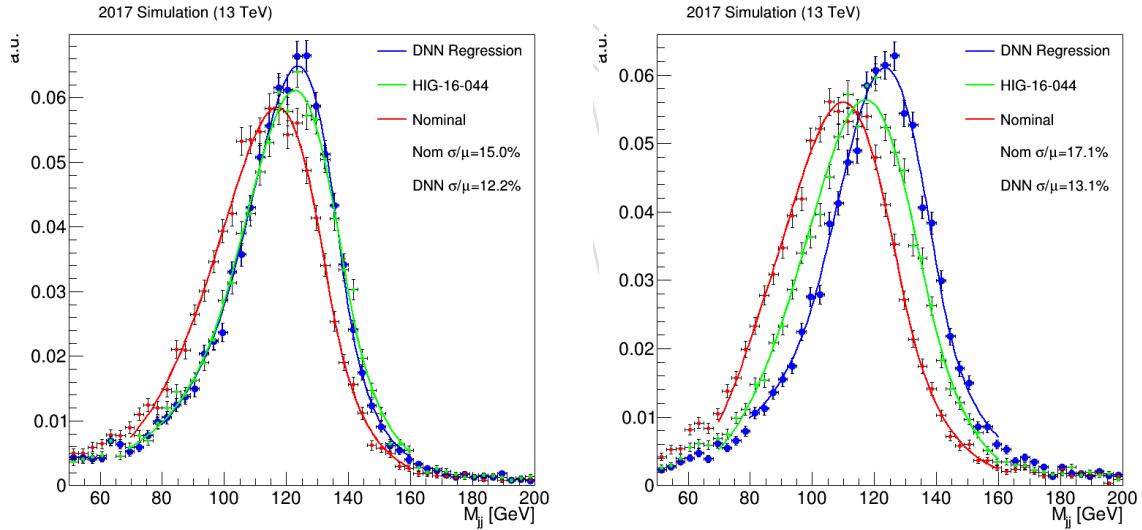


Figure 12: Comparison of the reconstructed dijet invariant mass for Higgs candidates in $Z(\ell\ell)H$ signal events before and after the DNN regression, compared to the regression used in the previous version of the analysis (HIG-16-044 [30]). Separate plots are shown for the case where the b-jet does not (left) or does (right) contain a lepton from semileptonic B decay. The same regression is used across all the analysis channels.

711 As further validation, we check the effect of the regression on the top mass in events selected to
 712 enhance single top production (details of the top reconstruction procedure in the WH channel
 713 are described in Sec. 9.1.1). Figure 13 shows the reconstructed top mass in this sample, com-
 714 paring data and simulation before and after regression, and the individual histograms for the
 715 single top and $t\bar{t}$ events overlaying the distributions before/after the regression is applied. A

716 clear improvement in the resolution is observed in the single top and $t\bar{t}$ events both in data and
 717 simulation.

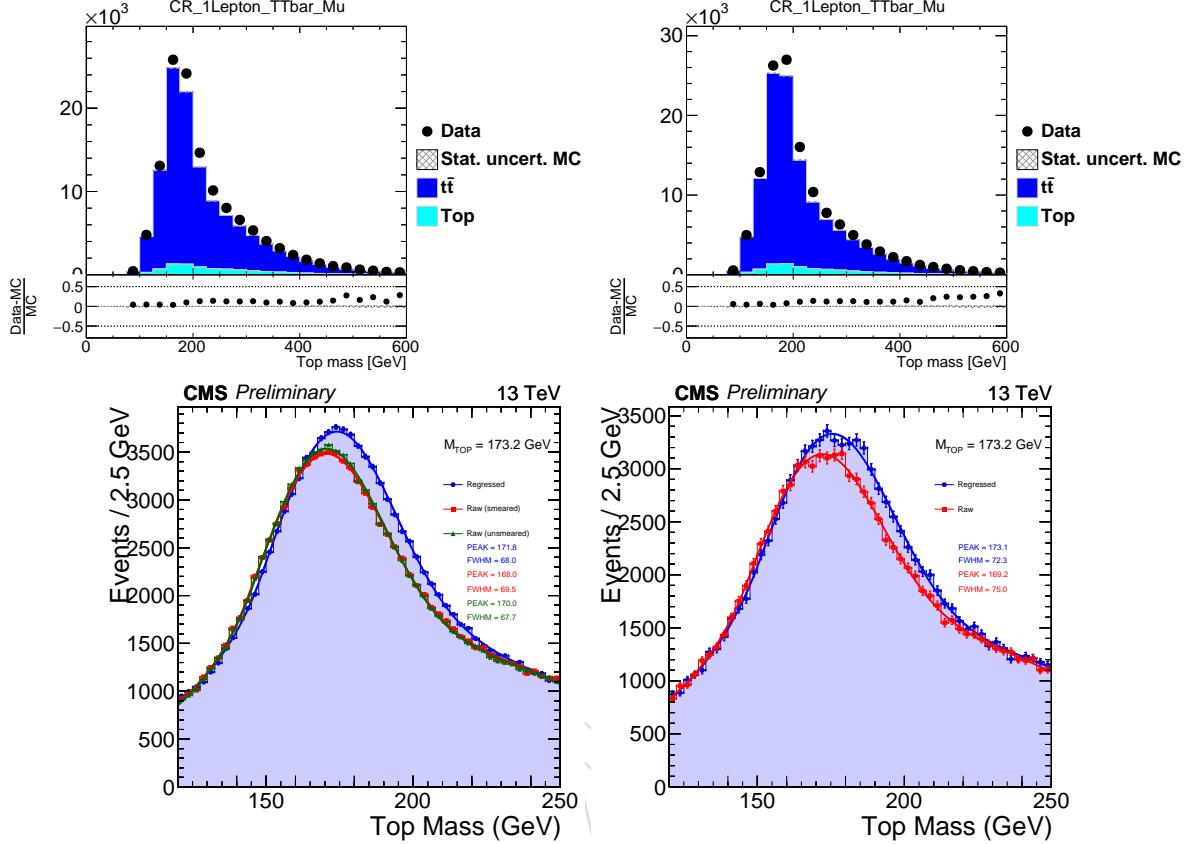


Figure 13: Distributions of the reconstructed top mass in a sample of events selected to enhance top, anti-top production. The top two figures show the data/MC agreement before (left) and after (right) the regression is applied. The bottom figures show a comparison of the top mass before and after the regression in simulation (left, where smeared and unsmeared refer to the additional smearing to the jet energy resolution to match the to data) and data (right).

718 6.2 ISR/FSR radiation recovery

719 Partons in the proton can emit initial state radiation (ISR), generally soft/collinear, before to
 720 scatter. Colored final state particles can also be emitted as soft/collinear final state radiation
 721 (FSR). These emissions can be identified together as additional jets in the event, other than the
 722 two Higgs candidate b-jets. In order to get the best estimation of the Higgs boson mass, the 4-
 723 vectors of the Higgs candidates are corrected by adding the 4-vector of additional jets selected
 724 among those within $\Delta R < 0.8$ of either Higgs candidate b-jet and passing the $p_T > 20$ GeV and
 725 $|\eta| < 3.0$ selection cuts.

726 Recovering the FSR jets improves the mass resolution of signal events by 2%, without sculpting
 727 the background shape. Validation plots are shown in Fig. 14–15 for the three channels.

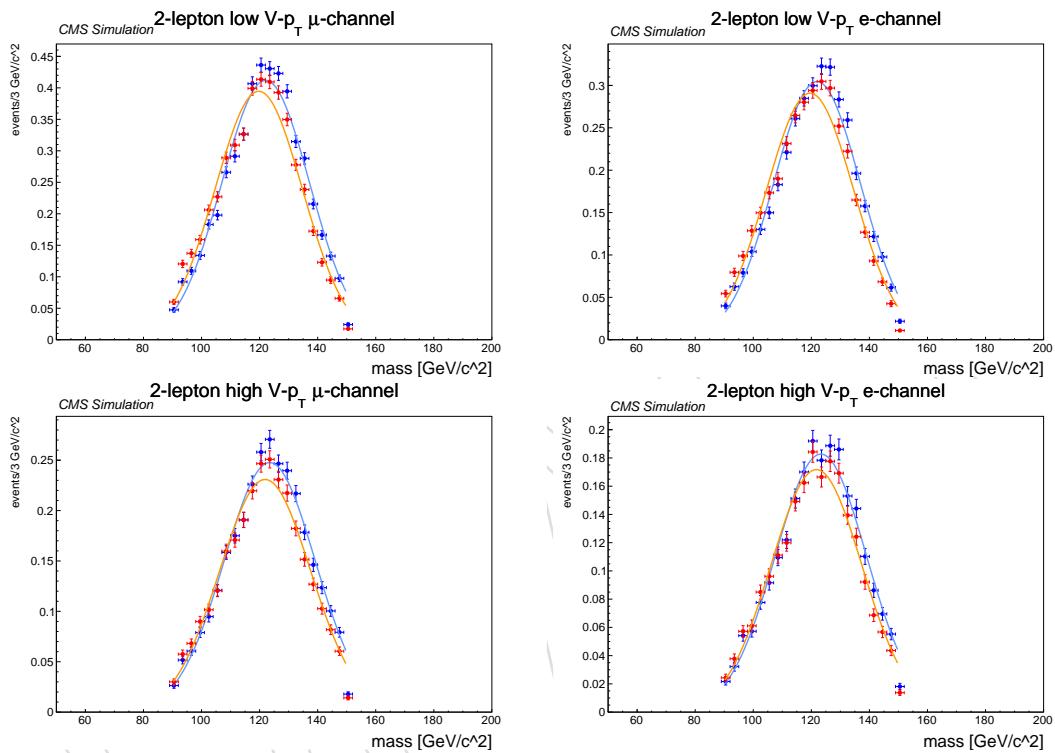


Figure 14: Reconstructed Higgs invariant mass distribution before (red) and after (blue) the FSR jet recovering with gaussian fit overlaid for the 2-lepton category in the low (top row) and high (bottom row) $p_T(V)$ categories, for the $Z \rightarrow \mu\mu$ (left) and $Z \rightarrow ee$ (right) channels.

728 In Table 11, a table is shown summarizing the mean value and the sigma from a gaussian fit
 729 to each category. The relative improvements have been calculated both in mass resolution and
 730 event yield due to signal region (cfr Sec. 8.3) kinematic cuts.

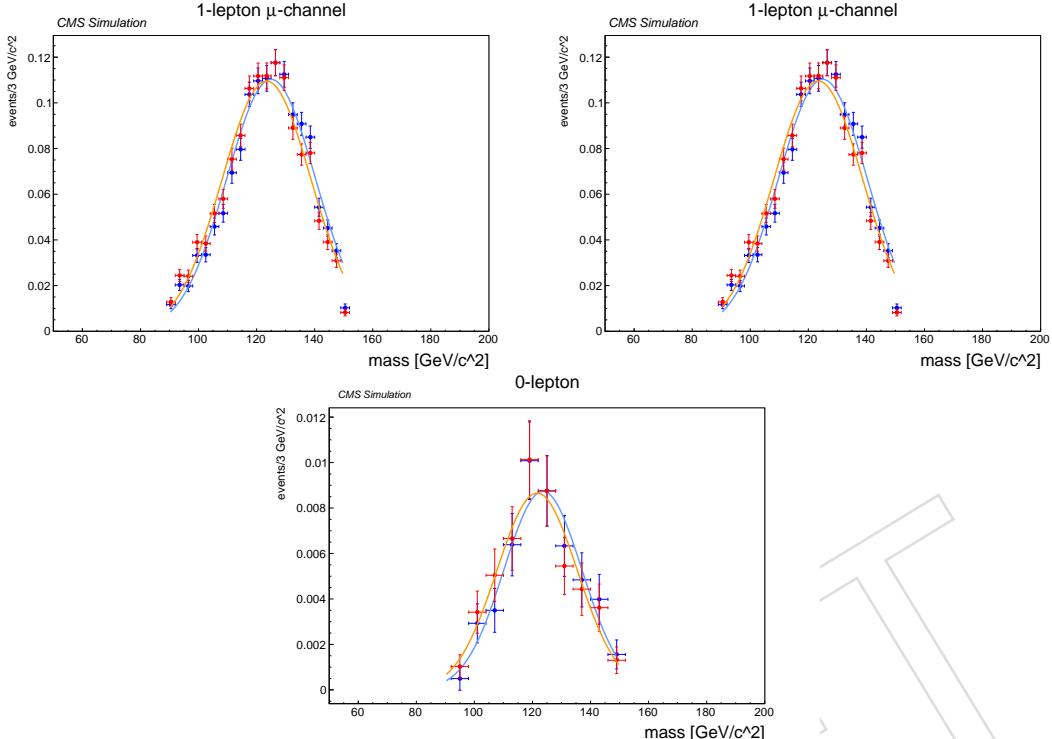


Figure 15: Reconstructed Higgs invariant mass distribution before (red) and after (blue) the FSR jet recovering with gaussian fit overlaid for the 1-lepton category (top row) in the $W \rightarrow \mu\nu$ (left) and $W \rightarrow e\nu$ (right) categories, and in the 0-lepton category (bottom row).

Table 11: Table summarizing the relative improvement brought by the FSR recovery in the Higgs candidate mass distribution. For each category, the number of events and the mean and σ resulting from a gaussian fit are shown, together with the relative improvement in the yield (imp. evt) and resolution (imp. rso).

category	Before ISR/FSR recovery			After ISR/FSR recovery			imp. evt [%]	imp. rso [%]
	N evt	mean	sigma	N evt	mean	sigma		
2-lep high Zmm	2.87	121.91	15.71	3.06	123.49	15.69	6.5	1.4
2-lep high Zee	2.14	121.75	15.65	2.27	123.29	15.70	6.0	1.0
2-lep low Zmm	4.79	119.71	15.05	4.98	122.09	15.07	4.0	1.8
2-lep low Zee	3.53	119.57	15.05	3.66	121.94	14.94	3.7	2.6
1-lep Wmnu	0.81	124.18	15.20	0.80	125.63	14.89	-1.0	3.1
1-lep Wenu	1.34	123.43	15.33	1.34	124.99	15.29	-0.3	1.6
0-lep	0.05	121.60	13.87	0.05	123.60	13.43	-1.9	4.8

731 6.3 Kinematic fit in 2-lepton channel

732 It has been shown (cfr [74] and references therein) that the resolution of the measured objects in
 733 the final state of proton collisions can be improved by forcing well-defined kinematic hypotheses
 734 through an event-by-event least square fitting technique. The resulting chi-square of the fit
 735 can be interpreted as the probability of the proposed kinematic hypotheses to be true for the
 736 observed event.

737 The vector sum of the transverse momenta of all particles events from pp collisions at LHC
 738 should be null. In events where high resolution final state particles (like charged leptons) are
 739 present and no undetected particles (like neutrinos) are produced, this kinematic constraint
 740 can be used to improve the energy estimate of other objects otherwise reconstructed with poor
 741 resolution (like jets). In the 2-lepton channel this opportunity occurs, as only 2 leptons, 2 b-jets
 742 and possibly jets from ISR and/or FSR are present.

743 The kinematic fit procedure consists in constraining the dilepton system to the Z boson mass
 744 and, subsequently, constraining the dilepton-dijet system in the transverse plane. Finally a fit is
 745 performed to the lepton and jet transverse momentum allowing their values to fluctuate within
 746 their uncertainties.

747 The fit is implemented in the PhysicsTools/KinFitter package and can be found in CMSSW_10_2_0_pre3.
 748 Particle candidates are instantiated as “FitParticles”, where the candidate 4-momentum and a
 749 covariance matrix in p_T , η and ϕ is handed to the fitting procedure. Kinematic relations be-
 750 tween the candidates are passed as “FitConstraint” instances. The kinematic fitting procedure
 751 is configured as follows:

752 FitParticles

753 **Two Higgs candidate jets:** b-jets from the Higgs candidates are used, after the jet energy re-
 754 gression and FSR recovery are applied. The b-jets resolutions in p_T (as a function of η and
 755 ϕ), η and ϕ (both as a function of p_T) are taken from AN-2015-108.

756 **Two lepton candidates:** electrons and muons are used, after corrections to the energy scale
 757 and resolution as prescribed by the respective POGs. The prescriptions also include an
 758 estimate of the per-lepton momentum uncertainty, which is used for the p_T variance. It
 759 is assumed that the largest contribution to the lepton 4-vector covariance matrix in the
 760 kinematic fit comes from its momentum, so the angular variances are set to a negligible
 761 value (10^{-4}).

762 **One recoil vector:** jets identified as ISR are used. They are required not to be Higgs-daughter
 763 candidates or FSR jets, to have $p_T > 20 \text{ GeV}$, and to pass loose pileup-, jet-ID-, and lepton
 764 filters. The complete detector coverage up to $\eta = 5$ is allowed for ISR jets. The recoil
 765 vector is computed as the vectorial sum over all ISR jets. The covariance in $p_{x,y}$ of the
 766 recoil vector is derived using the recommended POG resolutions. A value of $(8 \text{ GeV})^2$ in
 767 $p_{x,y}$, as determined in a $Z(l\bar{l})H(bb)$ signal sample, is added to the recoil covariance matrix.

768 FitConstraints

769 **Z boson mass:** in the signal process, the pair of high-momentum leptons originates from the
 770 decay of a Z boson. Therefore, the invariant mass of the 4-vector sum of the two lepton
 771 candidates is constrained to 91 GeV. A Gaussian uncertainty of 5 GeV is found to result
 772 in a realistic post-fit width of the Z boson candidate (**).

(**)⁷⁷³In the future, the Z mass constraint may be implemented using a Breit-Wigner prior pdf instead of Gaussian.

773 **Transverse momentum of the $llbbr$ system:** as the signal process does not exhibit MET in the
 774 hard process, the vector sum of all fitted particles (i.e. the $llbbr$ system) is constraint to
 775 zero in the transverse plane ($p_T(llbbr) = 0$).

776 In the following, the events are selected according to the signal region definition of the 2-lepton
 777 channel (cfr Sec. 8.3). Furthermore, both Higgs b-jets candidates are required to be matched to
 778 their generator equivalents within $\Delta R < 0.2$. If not specified otherwise, events in the high
 779 $p_T(V)$ region are shown.

780 The figure of merit used hereafter to measure the improvements is the pull of the candidate
 781 p_T , defined as $(p_{T,reco} - p_{T,gen}) / (\sigma_{p_{T,reco}})$, where $p_{T,reco}$ ($p_{T,gen}$) is the reconstructed (generated)
 782 energy and $\sigma_{p_{T,reco}}$ is the reconstructed energy resolution. The pull distributions for electrons
 783 and muons are shown in Fig. 16, while it is shown in Fig. 17 for the Higgs b-jet candidates with
 784 higher b-tag discriminator value (equivalent to the one obtained for the lower b-tag discrimi-
 785 nator ones).

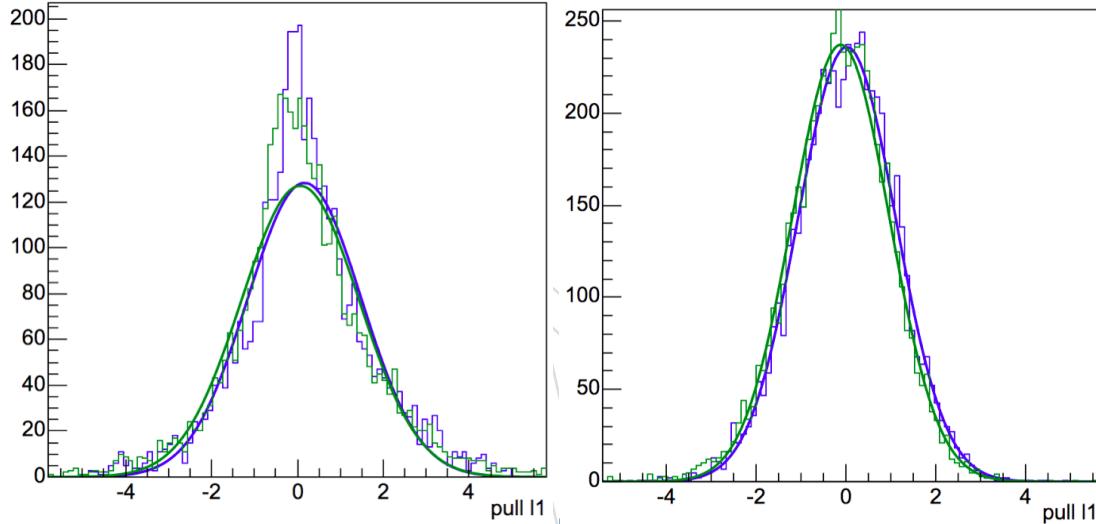


Figure 16: The pull distribution defined in the text is shown for electrons (left) and muons (right), before (blue) and after (green) the kinematic fit. Please note that the denominator for the calculation of the post-fit pull is the pre-fit uncertainty.

786 The Higgs invariant mass $M(jj)$ distribution is shown in Fig. 18 before and after the kinematic
 787 fit, for different number of ISR jets, while Table 12 details the resolution before and after the fit
 788 in bins of $p_T(V)$ and number of ISR jets. It is possible to observe that the resolution improve-
 789 ment is inversely proportional to the number of ISR jets, but sizeable even for more than 1 ISR
 790 jets.

Table 12: Higgs invariant mass resolution before and after the kinematic fit in bins of $p_T(V)$ and number of ISR jets

$p_T(V)$	#ISR Jets	σ_{reg}	σ_{fit}	Improvement (%)
> 150	0	14.3	9.2	36
> 150	1	15.3	11.5	25
> 150	> 1	14.7	12.8	13

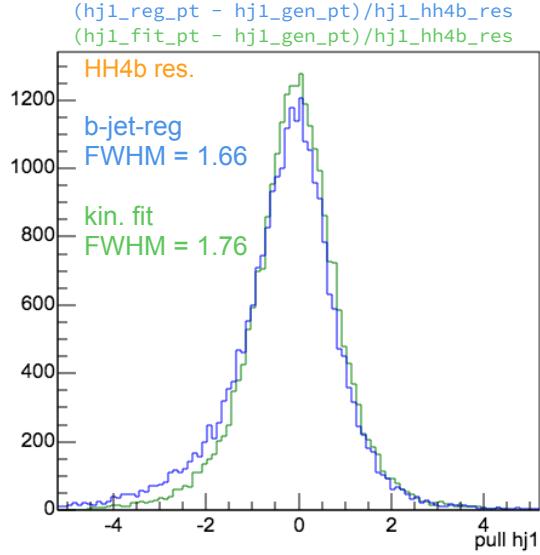


Figure 17: The pull distribution described in the text is shown for the Higgs candidate jets, before (blue) and after (green) the kinematic fit. Please note that the denominator for the calculation of the post-fit pull is the pre-fit uncertainty.

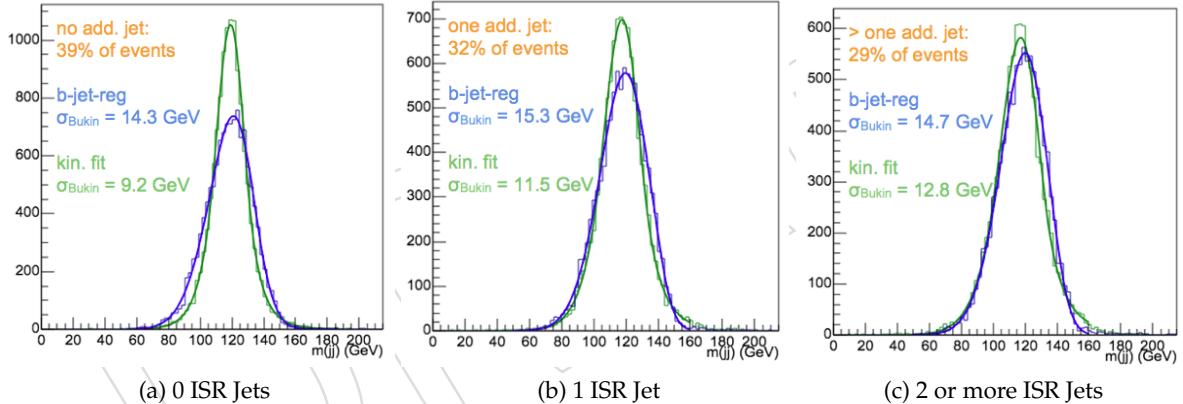


Figure 18: $m(jj)$ distributions before (with regression, blue curve) and after (green curve) the kinematic fit. The percentage of events shown in top is derived with respect to the number of events where both reconstructed Higgs jets are matched to their generator counterparts.

791 The fit is also performed in control regions (cfr Sec. 11.3) to check its impact on the background
 792 mass distributions and to verify that the jet resolution improves as expected in data. Finally,
 793 Figs. 19–20 show the p_T -balance of the Z and H boson candidates, i.e. the ratio $p_{Tjj}/p_T(V)$, for
 794 the high- and low- $p_T(V)$ regions in 2017 data. More distributions including the di-jet invariant
 795 mass $M(jj)$ are shown in Sec. 11.3 after the kinematic fit. All the distributions are well modelled
 796 and no mass-sculpting of the background shapes is observed.

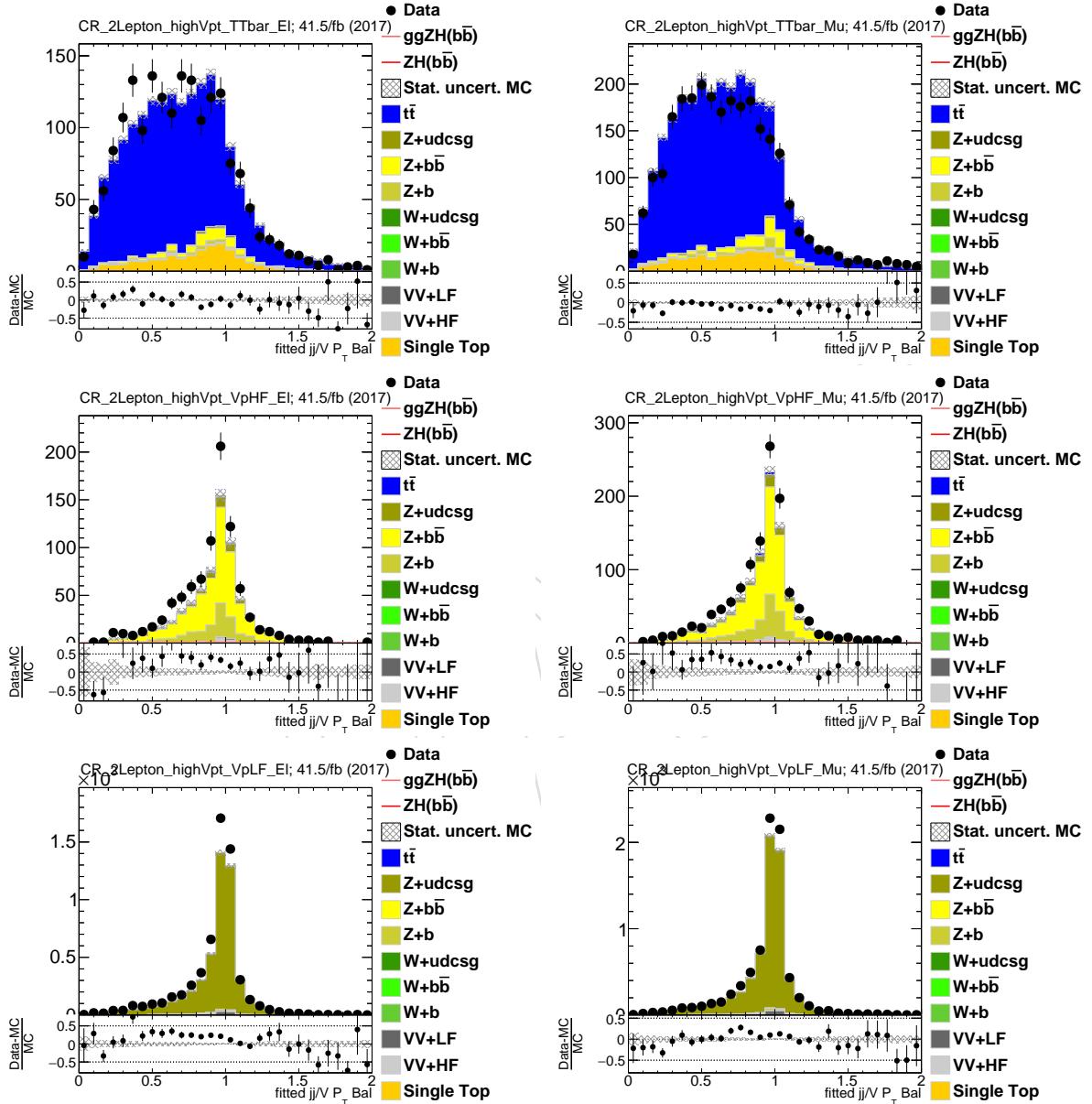


Figure 19: Distributions of the H-Z p_T balance after the kinematic fit in the high- $p_T(V)$ region of the 2-lepton channel in 2017 data, for the electron (left) and muon (right) channels in the $t\bar{t}$ (top), $Z + \text{heavy}$ (middle) and $Z + \text{light}$ (bottom) flavour control regions.

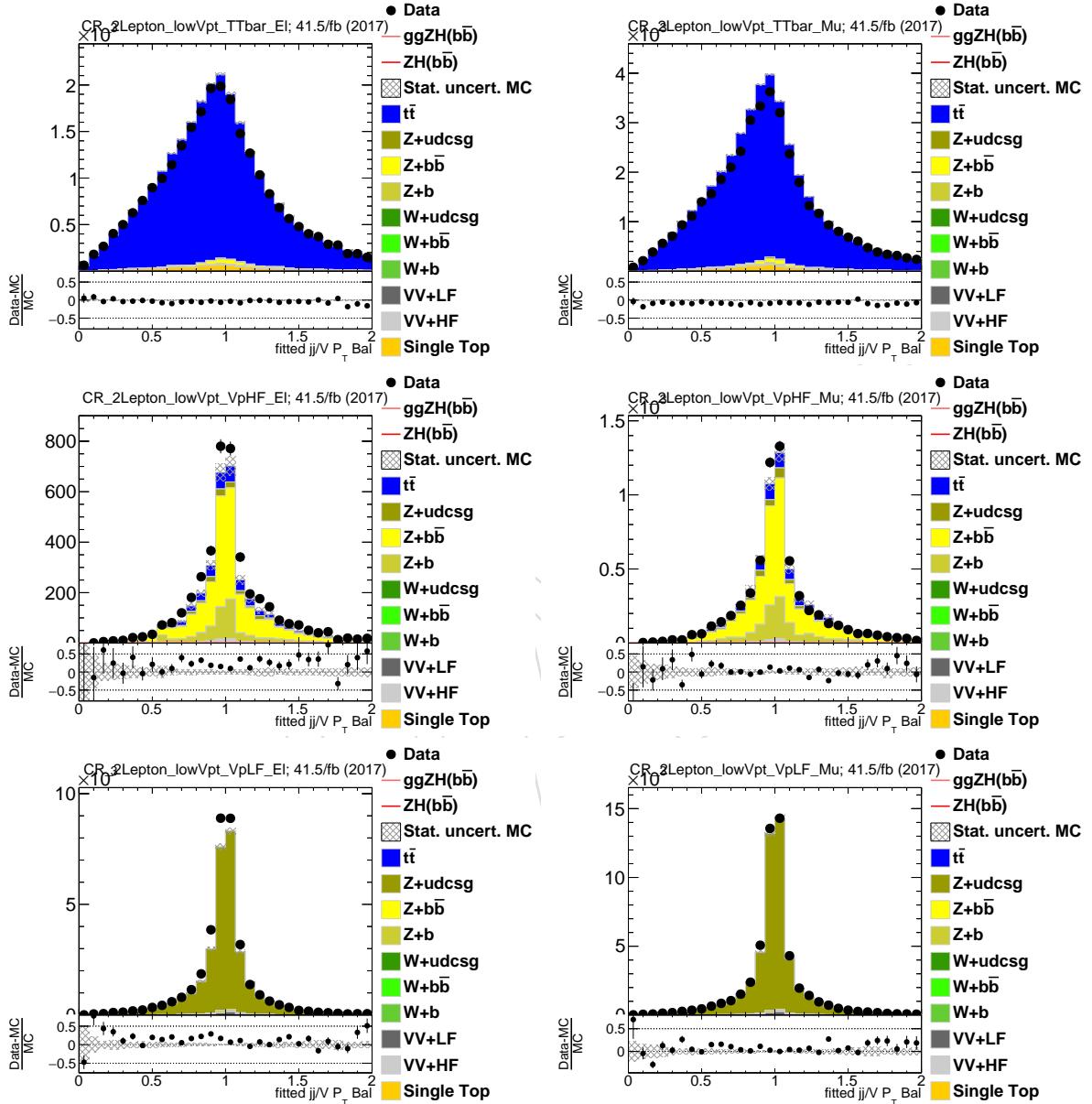


Figure 20: Distributions of the H-Z p_T balance after the kinematic fit in the low- $p_T(V)$ region of the 2-lepton channel in 2017 data, for the electron (left) and muon (right) channels in the $t\bar{t}$ (top), Z+heavy (middle) and Z+light (bottom) flavour control regions.

7 Residual MC corrections

Similarly to what has been done in the previous version of the analysis and documented in AN-15-168 [32], residual corrections are applied to the MC.

In particular, although there is good agreement in the $M(jj)$ and $p_T(V)$ distributions between data and NLO V+jets MC without any additional corrections needed, the statistical power of the available V+jets NLO samples is significantly less than the available LO V+jets MC. This difference in statistical power translates to a $> 10\%$ reduction in expected sensitivity when using the NLO V+jets samples as opposed to the LO V+jets samples. Therefore, we decided to use the LO V+jets sample in the analysis.

7.1 $M(jj)$ reweighting in LO V+Jets

We observe that the NLO DY+jets MC describes fairly well the Data, in particular the invariant mass of the di-jet system, while there is a discrepancy when using the LO sample. Reconstruction level comparisons of NLO and LO $\Delta\eta$ are used to derive event re-weighting. The NLO/LO ratio calculated in the inclusive phase space is shown in Fig. 21 for the $Z + 0b$, $Z + 1b$, $Z + 2b$ cases. The ratio is unchanged if generator level quantities are used and if the phase space is varied, even largely. The same scale factor is used for W+jets. After the reweighting is applied, the invariant mass distributions show better agreement in the LO MC. Other distributions are unaffected, apart from a slight improvement in the jet p_T distribution. The full reweighting is assigned as systematic uncertainty.

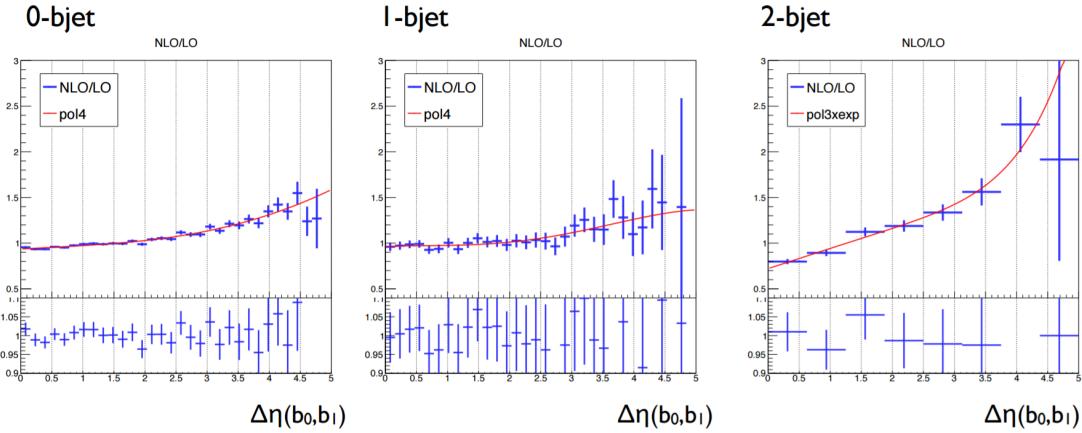


Figure 21: Ratio of the NLO/LO reconstruction level $\Delta\eta$ for DY+Jets MC for the $Z + 0b$ (left), $Z + 1b$ (center), $Z + 2b$ (right) cases.

7.2 W boson and $t\bar{t}$ transverse momentum reweighting

In the $W(\ell\nu)H$ control regions a downward slope in the data/MC ratio is observed for the reconstructed $p_T(V)$ after the full set of corrections and scale factors have been applied. Independent linear re-weighting functions are derived to correct this slope for $t\bar{t}$, $W + udscg$, and the combination of $W + bb$ and single top via a simultaneous fit of the reconstructed $p_T(V)$ in the $W(\ell\nu)H$ control regions to data. The input PDF for the fit in each control region is a sum of the MC prediction for each process corrected by a linear function of the reconstructed $p_T(V)$ with a slope that is allowed to float in the fit. The relative composition of the fitted processes in each control region is fixed. Table 13 lists the fitted slopes for each process with the uncertainties from RooFit.

Table 13: Linear correction factors obtained from a simultaneous fit to the $p_T(V)$ distribution in data in the $W(\ell\nu)H$ control regions.

Process	$t\bar{t}$	$W + udscg$	$W + bb + \text{single top}$
Fitted Slope (/ GeV)	0.00061 ± 0.00008	0.00064 ± 0.00004	0.0016 ± 0.0001
Norm preserving constant	1.103	1.115	1.337

826 The TOP group has observed similar MC mis-modelling of the reconstructed $p_T(V)$ distribu-
 827 tion for $t\bar{t}$ events simulated with Powheg in 2016 conditions (**), and the centrally provided
 828 correction perfectly matches the result obtained with our fit procedure. No central recipe is
 829 provided yet for 2017 simulation.

830 In the $Z(\nu\nu)H$ and $Z(\ell\ell)H$ channels the TOP re-weighting is applied to $t\bar{t}$ simulation. It has
 831 been verified that the result of simultaneous fit is not sensitive to changes in the definition of
 832 the fitted $W+HF$ control region such as loosening the additional jet multiplicity requirement or
 833 adjusting the $M(jj)$ selection.

834 The systematic uncertainties on the $p_T(V)$ corrections are taken from the uncertainties on the
 835 fitted slopes given by RooFit. This corresponds to a 13% uncertainty on the fitted slope for
 836 $t\bar{t}$ and an 6% uncertainty for both $W + udscg$ and $W+HF + \text{single top}$. It has been checked
 837 that this uncertainty band covers the residual data/MC differences in the $p_T(V)$ distribution
 838 after applying the corrections. The overall effect of this systematic on the $W(\ell\nu)H$ analysis
 839 sensitivity is less than 3%.

(**)cfr <https://twiki.cern.ch/twiki/bin/view/CMS/TopPtReweighting>

8 Event Selection

The building blocks of the analysis, described in the following section, are discriminating variables whose distributions differ between signal and background events, allowing for an optimal separation of the two.

In order to maximize the signal purity, a dedicated pre-selection is applied to each channel, primarily based on the characteristics of the vector bosons decays as well as limitation from the triggers.

8.1 Signal and background characteristics

Signal events are characterized by the presence of a vector boson recoiling against two b-jets with an expected invariant mass in the range $100 < M_H < 150$ GeV. Based on current experimental data and theory input, a few preliminary considerations can be made. The vector boson and dijet objects are expected to be central and back-to-back in the transverse plane, so that the azimuthal opening angle between them is peaked at π . The dijet p_T spectrum is harder than the background one, and the transverse momentum of each b-jet peaks at roughly $M_H/2$. Additional jet activity in the event is low, and isolated leptons not arising from the decay of a W or Z boson are expected to be negligible.

The dominant backgrounds arise from three general sources. Each of them can be mitigated exploiting distinct features observed when compared to signal events:

- **V+jets:** production of W and Z bosons in association with one or more jets. This background has large cross section and looks very much like signal topologically, but has a generally softer p_T spectrum, a sharply falling dijet mass distribution, and decay characteristics (effective spin and color radiation) that, in principle, differ significantly from the signal. The contributions from udscg jets dominates is much reduced after the application of b-tagging on both Higgs daughters. In the highest purity phase space of the analysis contributions from V + bb is nearly indistinguishable with signal.
- **Top quarks:** production of t̄t pairs, as well as single top quarks in the tW, t-channel, and s-channel processes, represent a particularly challenging background at the Run2 13 TeV energy, where the production cross section is 3.4 times larger than at Run1 7-8 TeV, whereas Higgs boson production only increases a factor of 2. These backgrounds include one or two real W decays, at least two b-jets (for t̄t), with intrinsic mass and momentum scales close to the Higgs mass scale. The primary handles to reduce the t̄t background are topological: the additional jet multiplicity extends well beyond two (typically from the hadronic decay of the second W), the azimuthal opening angle between the vector boson and dijet, which is more broadly distributed than in signal events, and finally (in the W(ℓν)H(bb̄) channels) the reconstructed top mass can be used to eliminate events with mass similar to the top. Single top events are more difficult to reject relative to signal, but the cross section is such that it typically represents only 10-20% of the total background in WH and even less in the other channels.
- **Dibosons (WW, WZ, ZZ):** production of vector-boson pairs is another important background, since they can produce a resonant dijet system having a mass within a few standard deviations of the signal. The dominant contribution arises when one boson decays leptonically (including $Z \rightarrow \nu\bar{\nu}$), and the second boson decays to jets. The ZZ and WZ modes can lead directly to a V + bb combination with two

real b-jets coming from the hadronic $Z \rightarrow b\bar{b}$ decay. This background is virtually indistinguishable from signal events, apart from the peak position of the dijet mass. Good mass resolution is the key handle to separate signal from this background.

- **Multijet QCD:** In the Run 2 analysis, we reported also a larger multijet QCD background compared to Run 1, especially for the $Z(\nu\nu)H$ channel. In multijet events, MET is mainly produced by the mis-measurement of the energy of a jet (fake MET) or by the emission of a high- p_T neutrino originating from a hadron decay. The fraction of events with high- p_T neutrinos is predicted with large uncertainty since it depends on production rates of heavy hadrons and branching ratios that are known with limited precision.

8.2 Discriminating variables

To better separate the signal events from background, the discrimination power of the following variables has been studied:

- $M(jj)$: dijet invariant mass; it peaks at M_H for VH(bb) signal and M_Z for VZ(bb) diboson events, falls sharply for V+jets, and peaks broadly over the region 100–160 GeV for $t\bar{t}$ events.
- $p_T(jj)$: transverse momentum of the Higgs candidate.
- $p_{T,j}$: transverse momentum of the Higgs candidate daughters.
- $p_T(V)$: vector boson transverse momentum, as defined in Sec. 5, highly correlated with $p_T(jj)$ for signal and most backgrounds.
- **b-tagging shape:** continuous output of the b-tagging discriminant (DeepCSV, cfr Sec. 4.6), optimized separately for the jet with the higher value ($btag_{max}$), and the one with the lower value ($btag_{min}$).
- M_t : the top mass reconstructed in events with a leptonic decaying W and one of the b-jets.
- $\Delta\phi(V,H)$: azimuthal opening angle between the momenta of the vector boson and the Higgs candidate.
- $\Delta\eta(jj)$: distance in pseudo-rapidity between the two Higgs candidate b-jets.
- $\Delta R(jj)$: distance in $\eta-\phi$ space between the two Higgs candidate b-jets.
- N_{aj} : number of additional jets in the event apart from the Higgs candidate ones. Only central jets with $|\eta| < 2.5$ are considered, and their p_T threshold and multiplicity are optimized separately for each channel. In practice, the optimal threshold is found to be $p_T > 20$ GeV in all channels where a jet veto is applied.
- N_{al} : number of additional isolated leptons (as defined in Sec. 4) apart from those associated with the W or Z decay. Only leptons satisfying $p_T > 20$ GeV and $|\eta| < 2.5$ are considered in the count.
- **pfMET:** E_T^{miss} calculated with particle-flow objects (as defined in Sec. 4).
- $\Delta\phi(\text{pfMET}, J)$: azimuthal opening angle between the pfMET vector direction and the transverse momentum of the closest central jet in azimuth. Only jets satisfying $p_T > 30$ GeV and $|\eta| < 2.5$ are considered. This variable helps in reducing residual QCD background in the $Z(\nu\nu)H$ channel, where the source of the missing transverse energy is typically from fluctuations in the measured energy of a single jet.
- $\Delta\phi(\text{pfMET}, \text{lept.})$: azimuthal opening angle between the pfMET vector direction and the leading lepton direction. This variable helps in discriminating against di-

929 leptonic $t\bar{t}$ decays in the $Z(\nu\nu)H$ channel.

- 930 • $btag_{max,aj}$: maximum b-tagging of the additional jets in the event. This variable
931 helps in reducing the $t\bar{t}$ background in the $Z(\nu\nu)H$ and $W(\ell\nu)H$ channels.
- 932 • $\text{min}\Delta R(H, aj)$: minimum distance between an additional jet and the Higgs candi-
933 date. This variable helps in reducing the $t\bar{t}$ background in the $Z(\nu\nu)H$ and $W(\ell\nu)H$
934 channels.
- 935 • **soft – activity N_5^{soft}** : number of additional soft track-jets with $p_T > 5 \text{ GeV}$ as defined
936 in Sec. 4.8.
- 937 • **Angular variables**: Various kinematic and angular variables: HV system mass, an-
938 gle $Z-Z^*$, angle $Z-l$, angle H -jet (only for $Z(\ell\ell)H$). These variables help to discrimi-
939 nate against ZZ and $Z + b\bar{b}$ backgrounds.

940 8.3 Signal region pre-selection

941 For each channel we define a loose pre-selection region to reduce the background contami-
942 nation and perform the final steps of the analysis, including the training of the multivariate
943 discriminators and the statistical analysis leading to the final result. In particular, the 0-lepton
944 uses a cut denoted as “Anti-QCD” described in Sec. 8.3.1. Following re-optimization of the
945 analysis strategy, a $p_T(V) > 150 \text{ GeV}$ cut (instead of $p_T(V) > 100 \text{ GeV}$ used in HIG-16-044) is
946 used in the 1-lepton channel to suppress further the multijet contamination in the 2017 analysis
947 without any significant loss of sensitivity.

948 These regions also contain cuts aimed at excluding part of the phase space where the triggers
949 are inefficient or difficult to model in MC.

950 Table 14 lists the signal pre-selection cuts for each channel, for 2017 analysis.

Table 14: Signal region pre-selection cuts for each channel. The values listed for kinematical variables are in units of GeV.

Variable	$Z(\nu\nu)H$	$W(\ell\nu)H$	$Z(\ell\ell)H$
$p_T(V)$	> 170	> 150 (**)	$[50 - 150], > 150$
$m_{\ell\ell}$	–	–	$[75 - 105]$
p_T^ℓ	–	$(> 25, > 30)$	> 20
$p_T(j_1)$	> 60	> 25	> 20
$p_T(j_2)$	> 35	> 25	> 20
$p_T(jj)$	> 120	> 100	–
$M(jj)$	$[60 - 160]$	$[90 - 150]$	$[90 - 150]$
$btag_{max}$	$>\text{Tight}$	$>\text{Tight}$	$>\text{Loose}$
$btag_{min}$	$>\text{Loose}$	$>\text{Loose}$	$>\text{Loose}$
N_{aj}	–	< 2	–
N_{al}	$= 0$	$= 0$	–
E_T^{miss}	> 170	–	–
Anti-QCD	Yes	–	–
$\Delta\phi(V, H)(\text{rad})$	> 2.0	> 2.5	> 2.5
$\Delta\phi(\text{pfMET, trkMET})(\text{rad})$	< 0.5	–	–
$\Delta\phi(\text{pfMET, lep})(\text{rad})$	–	< 2.0	–
Tightened Lepton Iso.	–	$(0.06, 0.06)$	–

(**) New with respect to the previous Run2 analysis on 2016 data described in HIG-16-044 [30] and AN-15-168 [32].

951 **8.3.1 Anti-QCD cut**

952 In order to reject the multijet QCD background in the 0-lepton channel, a cut of $\Delta\phi(\text{jet}, \text{MET}) >$
953 0.5 is applied to all jets with $p_T > 30 \text{ GeV}$.

954 Tested in the region $500 \text{ GeV} < \text{genHT} < 700 \text{ GeV}$ the cut rejects $\sim 93\%$ of multijet QCD events,
955 with a signal efficiency of $\sim 96\%$. This cut is also used to increase the purity of the $Z + \text{light-jets}$
956 and $Z + b\text{-jets}$ in the respective control regions defined in Sec. 11.1.

DRAFT

9 Binned Multivariate Shape Analysis

The ultimate goal of the analysis presented in this note is to perform a fit to the shape of a multivariate discriminator able to distinguish the VH signal from the SM background.

In the previous version of the analysis, like the one described in HIG-16-044 [30] and the corresponding AN-15-168 [32], the discriminator was based on a Boost Decision Tree (BDT) trained and optimized for each channel on a selected set of variables. In order to increase further the final significance, a re-optimization of the the multivariate strategy has been performed including all the analysis improvements in the input variables (cfr Sec. 6), adding an alternative approach based on the Deep Neural Network (DNN) architecture in addition to the BDT, and determining a more accurate choice of the output discriminator binning.

9.1 Multivariate discriminator: input variables

Compared to HIG-16-044 [30] an updated set of input variables has been used in each category (0-, 1- and 2-lepton) for the 2017 analysis, and is reported in Table 15.

The Higgs candidate jet variables are considered after regression and FSR recovery for all the analysis categories, and after kinematic fit for the 2-lepton category. The additional jets are defined as those not identified as $H \rightarrow b\bar{b}$ candidates and satisfying the following criteria: $p_T > 30 \text{ GeV}$ and $|\eta| < 2.4$ for the 0-lepton and 2-lepton channels, and $p_T > 25 \text{ GeV}$ and $|\eta| < 2.9$ for the 1-lepton channel. The reconstructed top quark mass is described in Sec. 9.1.1.

Table 15: List of input variables used in the training of the multivariate discriminators, and their use in the different lepton category.

Variable	Description	0-lepton	1-lepton	2-lepton
$M(jj)$	dijet invariant mass	✓	✓	✓
$p_T(jj)$	dijet transverse momentum	✓	✓	✓
$p_T(j_1), p_T(j_2)$	transverse momentum of each jet	✓		✓
$\Delta R(jj)$	distance in η - ϕ between jets			✓
$\Delta\eta(jj)$	difference in η between jets		✓	✓
$\Delta\phi(jj)$	azimuthal angle between jets	✓		
$p_T(V)$	vector boson transverse momentum		✓	✓
$\Delta\phi(V, H)$	azimuthal angle between vector boson and dijet directions	✓	✓	✓
$p_T(jj)/p_T(V)$	p_T ratio between dijet and vector boson			✓
M_Z	reconstructed Z boson mass			✓
$btag_{max}$	value of the b-tagging discriminant (DeepCSV) for the jet with highest score	✓		✓
$btag_{min}$	value of the b-tagging discriminant (DeepCSV) for the jet with second highest score	✓	✓	✓
$btag_{add}$	value of b-tagging discriminant for the additional jet with highest value	✓		
E_T^{miss}	missing transverse momentum	✓	✓	✓
$\Delta\phi(E_T^{\text{miss}}, j)$	azimuthal angle between E_T^{miss} and closest jet with $p_T > 30 \text{ GeV}$	✓		
$\Delta\phi(E_T^{\text{miss}}, \ell)$	azimuthal angle between E_T^{miss} and lepton		✓	
m_T	mass of lepton $\vec{p}_T + E_T^{\text{miss}}$		✓	
M_t	reconstructed top quark mass		✓	
N_{aj}	number of additional jets		✓	✓
$p_T(\text{add})$	transverse momentum of leading additional jet	✓		
SA5	number of soft-track jets with $p_T > 5 \text{ GeV}$	✓	✓	✓

9.1.1 Reconstructed top mass in events with a lepton and MET

In the 1-lepton category, $W(\ell\nu)H(b\bar{b})$ events are characterized by the presence of a well isolated lepton, MET and two b-jets. Unfortunately, the same signature arises in semi-leptonic $t\bar{t}$ events, making the background from this process very prominent.

979 Several variables were analyzed to help discriminating against $t\bar{t}$, and the reconstructed top
 980 mass was found to be the most powerful. Since the charged lepton and the neutrino directly
 981 come from the decay of the W , using the constraint of the W mass, the lepton p_T , and the
 982 known p_T of the neutrino (assumed to be equal to MET) is possible to solve for the unknown
 983 longitudinal neutrino momentum with:

$$M_W^2 = (E_\nu + E_\ell)^2 - (\vec{p}_\nu + \vec{p}_\ell)^2 \quad (8)$$

984 There are always two solutions to this equation. When both are real, the solution with the
 985 smaller longitudinal neutrino momentum is selected. When the solutions are imaginary, the
 986 real part is taken as the longitudinal neutrino momentum.

987 Once done, the energy-momentum 4-vectors of the neutrino, the lepton and the closest b-jet are
 988 added together and the resulting 4-vector is assumed to be the top quark, from which the mass
 989 is computed.

990 9.2 Multivariate discriminator training

991 A separate multivariate training for each lepton category has been carried out on the variables
 992 listed in Table 15 using two BDT boosting structures (Gradient, BDTG, and Adaptive, BDTA),
 993 and the DNN.

994 The training background sample is obtained from the weighted sum of all the background
 995 sample listed in Sec. 2 (namely W+Jets, Drell-Yan, di-boson, QCD, $t\bar{t}$ and single-Top). For the
 996 1-lepton and 2-lepton categories, the training is performed inclusively on electron and muon
 997 events.

998 To obtain statistically independent samples for training and the testing the multivariate algo-
 999 rithms, the signal and background samples are splitted according to their event numbers: even
 1000 for training and odd for testing.

1001 Given the relatively low MC statistics available for 2017, it has been decided to train the mul-
 1002 tivariate discriminators on the 2016 dataset for the 2017 analysis. To ensure full consistency, in
 1003 the training the Higgs-candidate b-jets ordering was done according to the DeepCSV score.

1004 9.2.1 Two-lepton category

1005 The distributions of the input variables for signal and background used to train the multivariate
 1006 discriminator in the 2-lepton channels are reported in Fig. 22 for the high $p_T(V)$ region and in
 1007 Fig. 23 for the low $p_T(V)$ region.

1008 Given the fully reconstructed Z boson, this category is the most kinematically constrained and
 1009 can rely on the improved jet-related variables coming from the kinematic fit.

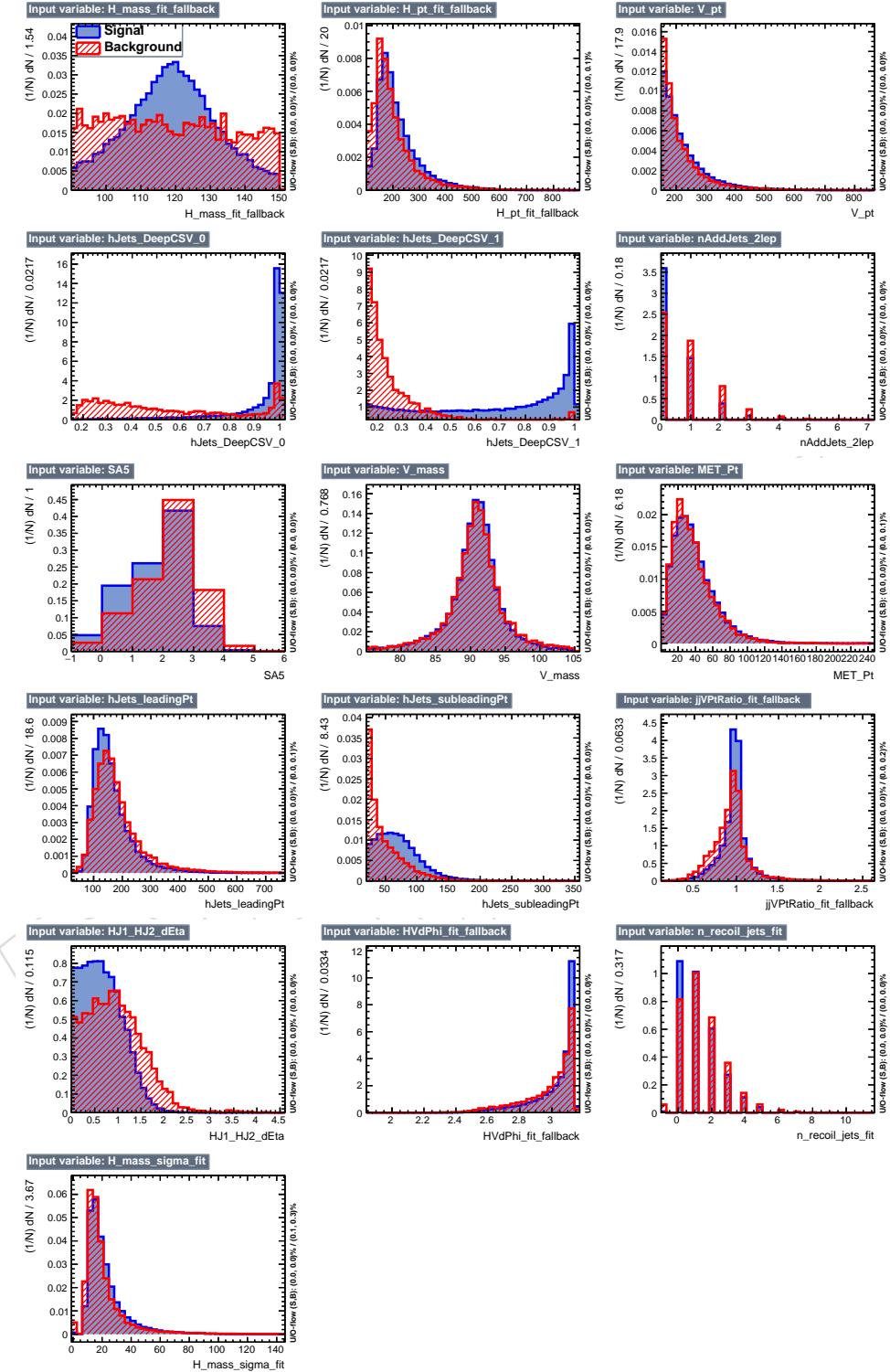
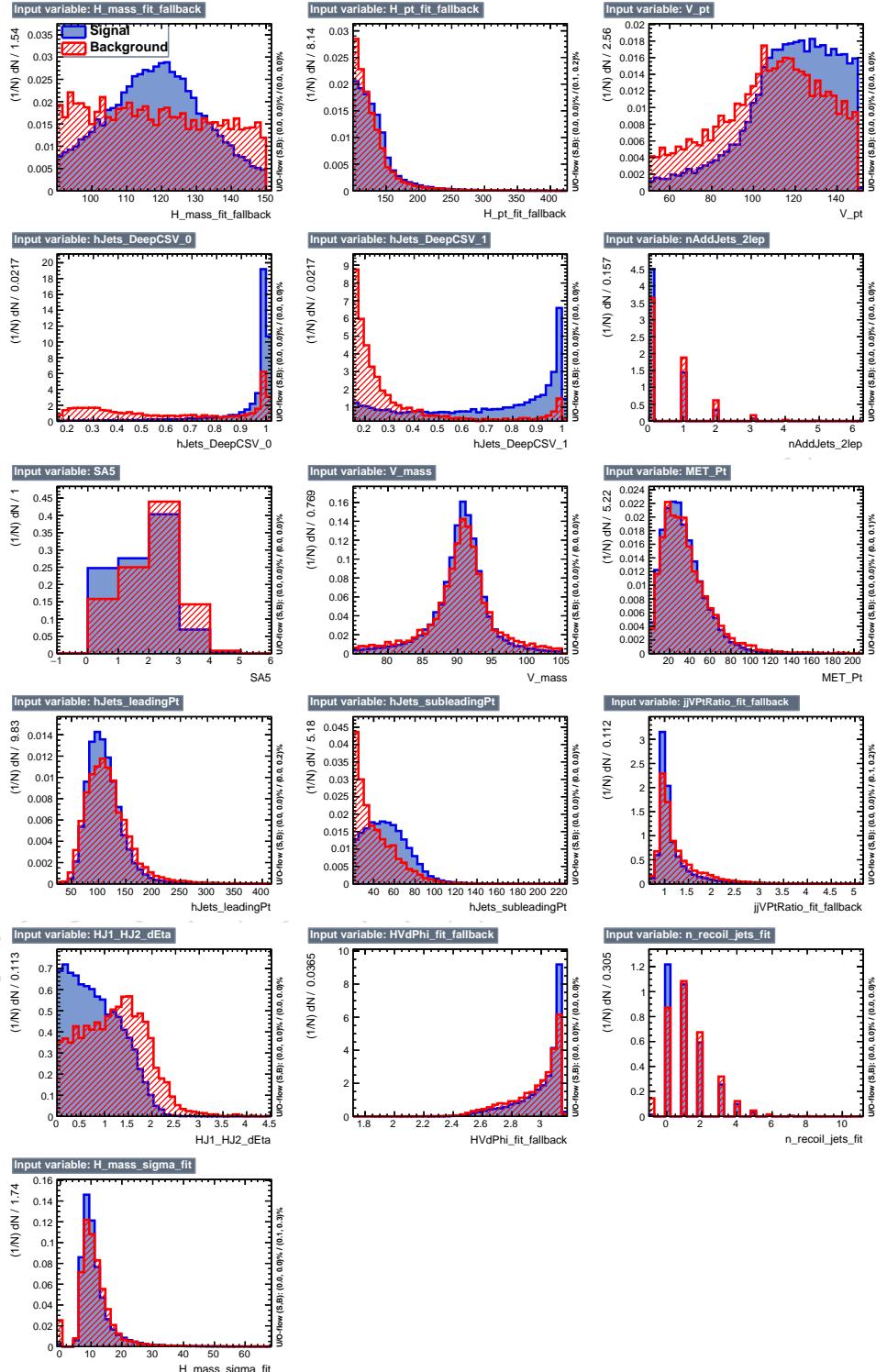


Figure 22: Set of input variables for the two-lepton high $p_T(V)$ BDT training

Figure 23: Set of input variables for the two-lepton low $p_T(V)$ BDT training

1010 **9.2.2 One-lepton category**

1011 The distributions of the input variables used to train the multivariate discriminator in the 1-
 1012 lepton category are shown for signal and background in Fig. 24.

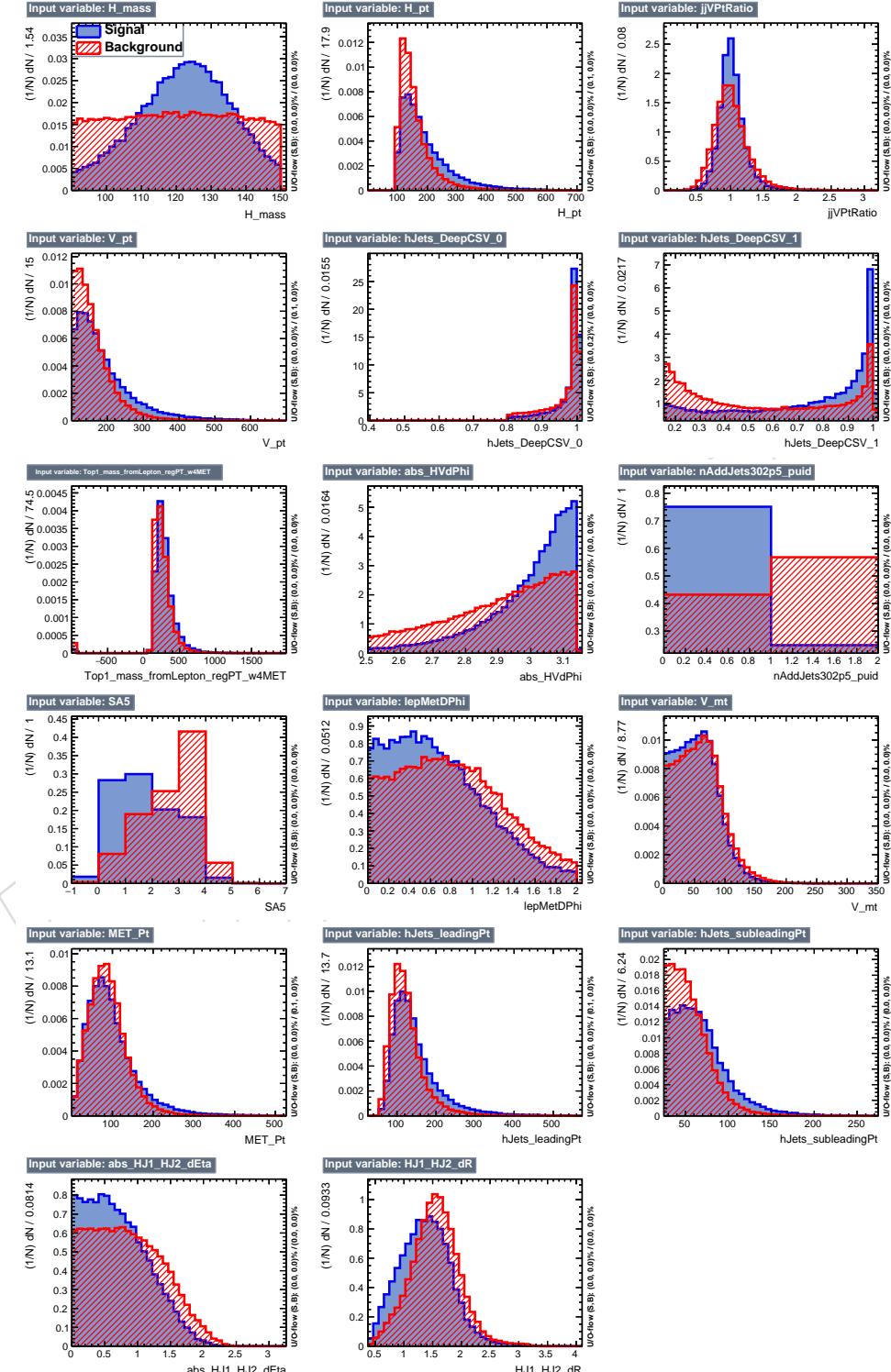


Figure 24: Set of input variables for the one-lepton BDT training

9.2.3 Zero-lepton category

The distributions of the input variables used to train the multivariate discriminator in the 0-lepton category are shown for signal and background in Fig. 25.

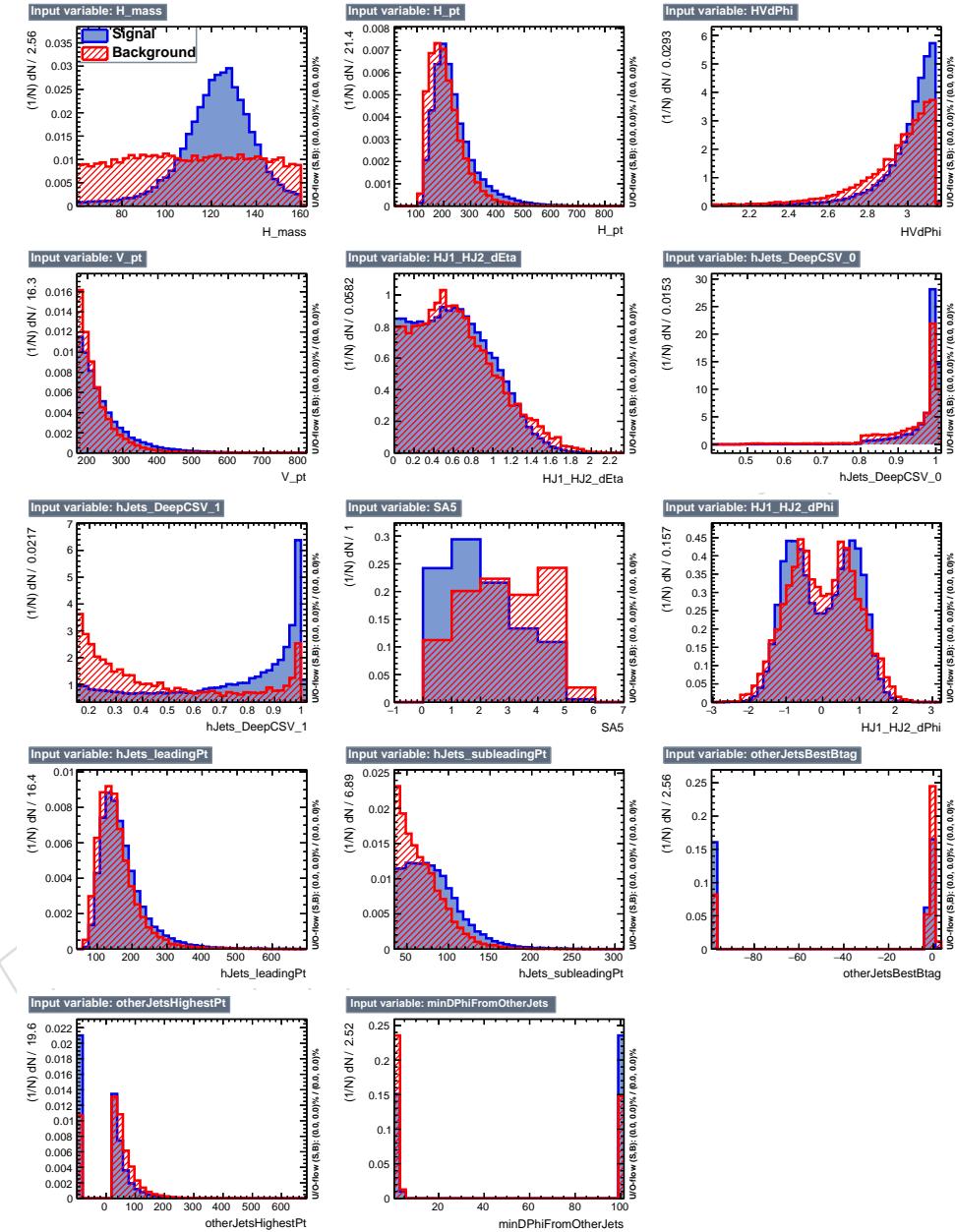


Figure 25: Set of input variables for the zero-lepton BDT training

9.3 BDT performance

The ranking of the input variables obtained in the BDT training is shown in Table 16 for 2017.

Table 16: Variable ranking for the BDT classifiers trained for the 2017 analysis (top variable is best ranked and has highest importance in the classifier).

Rank	$W(\ell\nu)H$	$Z(\nu\nu)H$	$Z(\ell\ell)H$ low p_T	$Z(\ell\ell)H$ high p_T
1	$\Delta\eta(jj)$	$M(jj)$	DeepCSV _{max}	DeepCSV _{max}
2	$M(jj)$	DeepCSV _{min}	$M(jj)$	$M(jj)$
3	DeepCSV _{min}	$\Delta\varphi(jj)$	DeepCSV _{min}	DeepCSV _{min}
4	M_t	$\Delta\eta(jj)$	M_Z	$\Delta\eta(jj)$
5	SA5	DeepCSV _{max}	$\Delta\eta(jj)$	M_Z
6	E_T^{miss}	SA5	$p_T(Z)$	E_T^{miss}
7	$ \Delta\phi(V, H) $	Sub-leading jet p_T	E_T^{miss}	Sub-leading jet p_T
8	$\Delta\varphi(\text{pfMET, lept.})$	$\Delta\phi(V, H)$	Sub-leading jet p_T	Leading jet p_T
9	p_T balance	Max $p_T(\text{add.jet})$	$\Delta\phi(V, H)$	N_{aj}
10	$\Delta R(jj)$	$p_T(jj)$	Leading jet p_T	$p_T(jj)$
11	DeepCSV _{max}	$p_T(Z)$	SA5	SA5
12	$p_T(Z)$	Leading jet p_T	$N_{\text{recoil jets}}$	$\Delta\phi(V, H)$
13	Leading jet p_T	Max DeepCSV _{add}	N_{aj}	$p_T(Z)$
14	Sub-leading jet p_T	Min $\Delta\phi_{\text{add}}$	$p_T(jj)$	$\sigma(H_{\text{mass}})$
15	$p_T(jj)$	—	p_T balance	p_T balance
16	N_{aj}	—	$\sigma(H_{\text{mass}})$	$N_{\text{recoil jets}}$
17	$m_T(W)$	—	—	—

The Gradient Boost (BDTG) algorithm has been found to outperform the Adaptive Boost (BDTA) and is used to extract the result. The corresponding BDT distributions obtained from the signal and background events in the training and testing sets are shown in Appendix A.

1021 **9.3.1 Performance improvement due with FSR recovery and kinematic fit**

1022 In order to estimate the performance improvement on the BDT ROC curve due to the FSR-jets
 1023 recovery and the kinematic fit, a training has been performed in the high $p_T(V)$ region of the
 1024 2-lepton category after removing these two features. The two trainings have been compared,
 1025 as shown in Fig. 26. It is possible to observe a substantial improvement in signal efficiency for
 1026 any given background rejection working point when the FSR recovery and kinematic fit are
 1027 included.

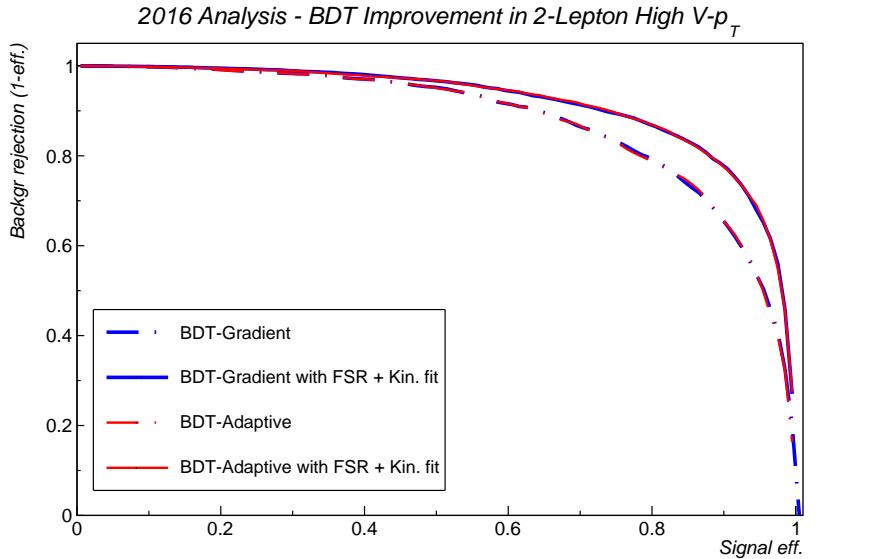


Figure 26: Comparison of the ROC curves obtained for BDT training in the two-lepton high $p_T(V)$ with and without the FSR-jets recovery and the kinematic fit. It is possible to see a substantial gain in signal efficiency for any given background rejection working point.

1028 **9.4 DNN performance**

1029 Multivariate techniques are used to further enhance the signal from background separation in
 1030 the analysis by making full use of correlations between discriminating variables in signal and
 1031 background events. The Deep Learning Neural Network approach (DNN) implemented in the
 1032 TensorFlow/Keras frameworks is used.

1033 Separate DNNs have been trained in each channel, after applying the b-jet energy regression
 1034 described in Sec. 6 and the FSR recovery in all the channels, and the Kinematic fit in the 2-
 1035 lepton channel. Electron modes are trained together with muon modes in the $Z(\ell\ell)H$ case but
 1036 separately in the $W(\ell\nu)H$ because of different performance. Table 15 lists the discriminat-
 1037 ing variables used in the training of the DNN analyzers. Only events passing the relatively
 1038 loose pre-selection, listed in Table 14, are used for training. All MC events have "event num-
 1039 bers". The even events are used for training the DNNs, while the odds events are saved for the
 1040 evaluation of the DNN performance.

1041 **9.4.1 Description of the DNN architecture and hyper-parameter optimization**

1042 The typical number of events used for the training is 200k, evenly divided by signal and back-
 1043 ground and corresponding, for the 2-lepton case, roughly to 40 signal and 3000 background
 1044 events when normalized to the luminosity collected in 2017. The implementation of the net-
 1045 work has been tested both in Keras and pure TensorFlow to customize the loss function. GPUs
 1046 are used for training (Tesla P100 at CSCS T2, GTX 970) requiring 5-10 minutes for each training
 1047 with large time gain with respect to CPUs (time: mobile i5 / GTX970 \approx 25 – 30).

1048 The chosen architecture consists of 5 hidden layers (2 to 5 were tested). The number of nodes
 for each layer is 32 (8 to 512 tested). A sketch of the architecture is shown in Fig. 27. At

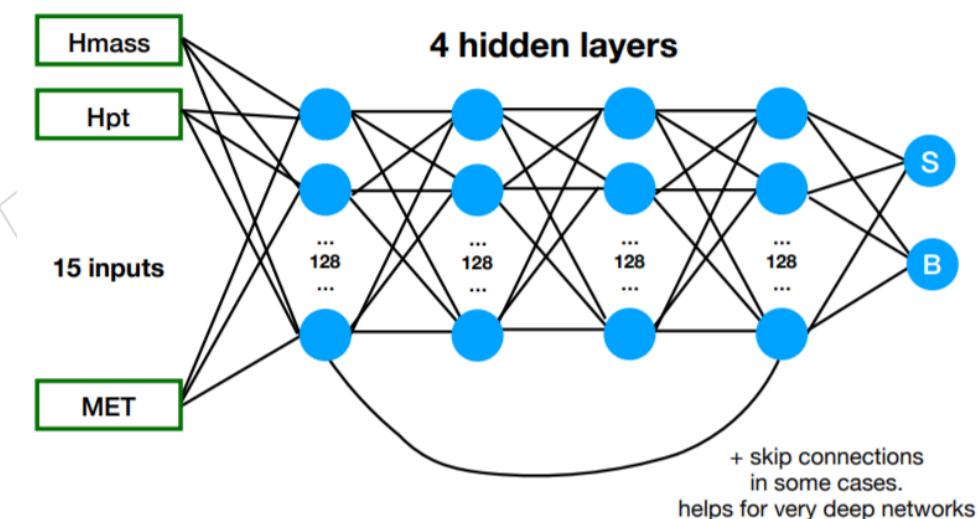


Figure 27: Sketch of the DNN architecture (left)

1049 each layer, dropout (i.e. during learning phase randomly set weights of node to 0) is used to
 1050 ensure regularization and make the network more robust. The leaky ReLU is used as activation
 1051 function (no big differences observed compared to others). Two loss functions were compared.
 1052 The first is named cross-entropy (standard for DNN) defined as $-(y \log(p) + (1 - y) \log(1 - p))$
 1053 which is equivalent to $-(\log -likelihood)$ for the data y under model p assuming a binomial
 1054 distribution, assuring fast convergence but not ideal for signal extraction in case of low Signal
 1055

1056 to background ratio ($S \ll B$), therefore requiring accurate rebinning.
 1057 The second loss is the counting experiment median significance (a.k.a Asimov-like or “Bin-
 1058 aware” loss function), defined as $\text{med}[Z_0|1] = \sqrt{q_{0,A}} = \sqrt{2((s+b)\ln(1+s/b) - s)}$ and using
 1059 fixed binning at training time. To get a fully differentiable loss function, the bins have been
 1060 approximated by smooth kernel functions. For both choices, the loss function is minimized
 1061 with the *Adam* algorithm.

1062 To get a preliminary estimate of the performance of each DNN training, and compare them, a
 1063 simple figure of merit can be used: $S/\sqrt{S+B}$, dividing the discriminator output in 15 evenly
 1064 spaced bins. The difference in performance between the training and the test set provides
 1065 a metric to compare DNN performance variance and stability. To further refine the hyper-
 1066 parameters, checks are performed choosing different dropout factors. It has been observed
 1067 that using drop out favors a reduction of the training/testing difference therefore providing
 1068 more robust results.

1069 Another test has been carried to check the performance as a function of the learning rate, find-
 1070 ing no striking dependence or improvement with respect to the default value (0.001).

1071 A thorough comparison of the performance of the two loss functions requires an optimization
 1072 of the hyper-parameters in the two cases, as well as a careful choice of the binning. Several
 1073 studies show that the two loss function ensure similar performance. For simplicity, the cross-
 1074 entropy loss is used to quote the results.

1075 9.4.2 Training and Validation

1076 A comparison of the training and testing performance in the 0-lepton channel as a function of
 the training epoch is reported in Fig. 28.

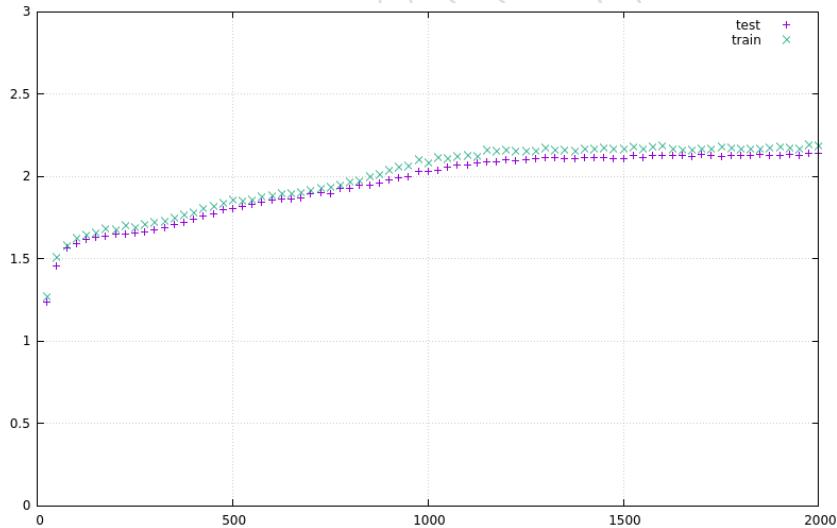


Figure 28: A comparison of the training and testing performance in the 0-lepton channel as a function of the training epoch.

1077

1078 The signal and background distributions obtained for the various channels with dedicated
 1079 trainings are shown in Fig. 29. The ROC curves of the trainings are instead shown in Fig. 30.

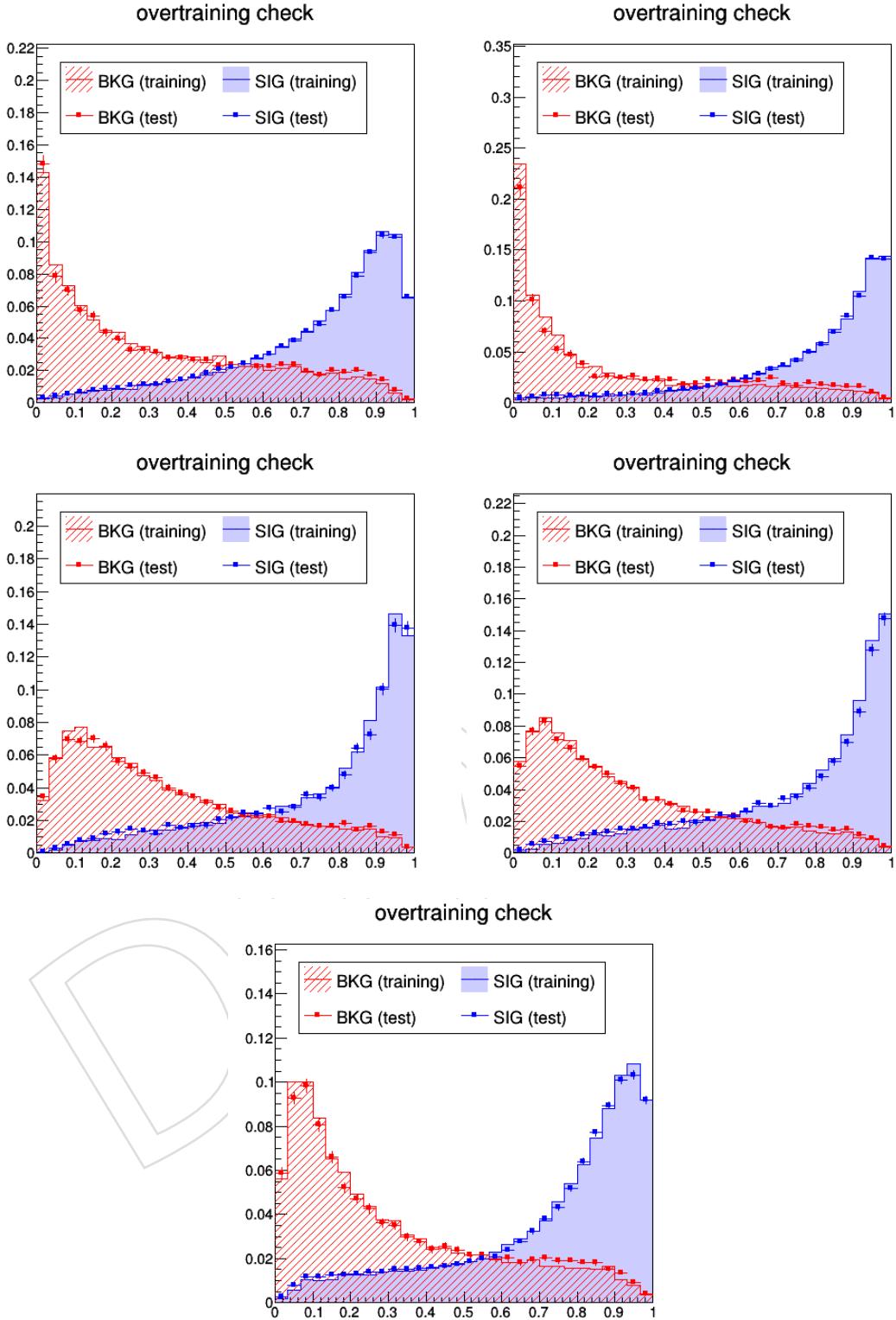


Figure 29: The plots above show the DNN output for signal and background simulation. Top row: 2-lepton low- $p_T(V)$ (left) and low- $p_T(V)$ (right). Middle row: 1-lepton electrons (left) and muons (right). Bottom row: 0-lepton.

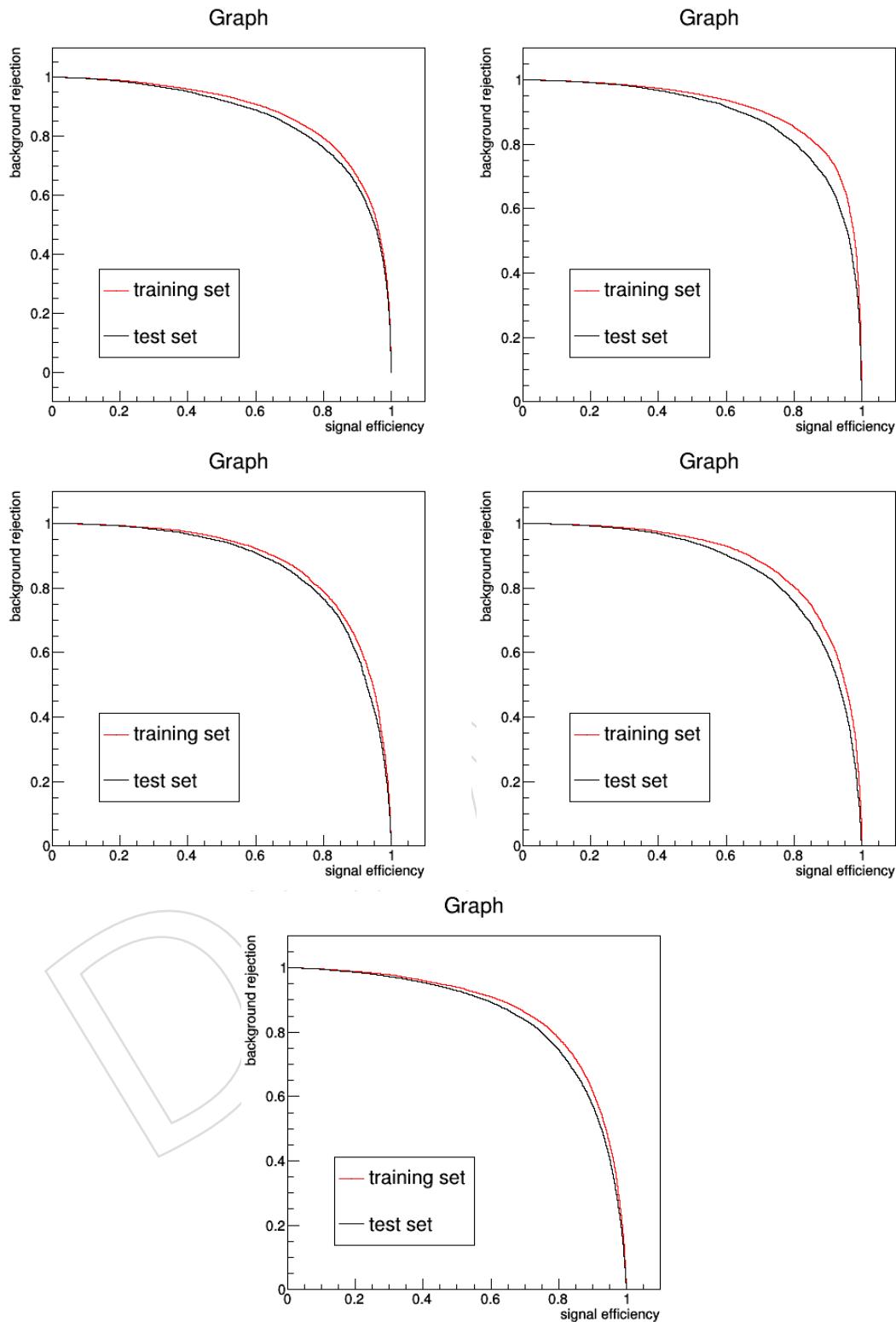


Figure 30: The plots above show the ROC curves for the DNN trainings. Top row: 2-lepton low- $p_T(V)$ (left) and low- $p_T(V)$ (right). Middle row: 1-lepton electrons (left) and muons (right). Bottom row: 0-lepton.

1080 **9.4.3 DNN reshaping**

1081 Using 15 equidistant bins between 0 and 1 for the DNN score, most of the background ends up
 1082 in the first and most of the signal in the last bin, a rebinning is applied to distribute signal and
 1083 background over multiple bins with different S/B. The bins for the shape of the DNN score
 1084 are equivalent to equidistant bins, with the score transformed with the function $\frac{\sqrt{x} + x^{12}}{2}$, shown
 1085 in figure 31.

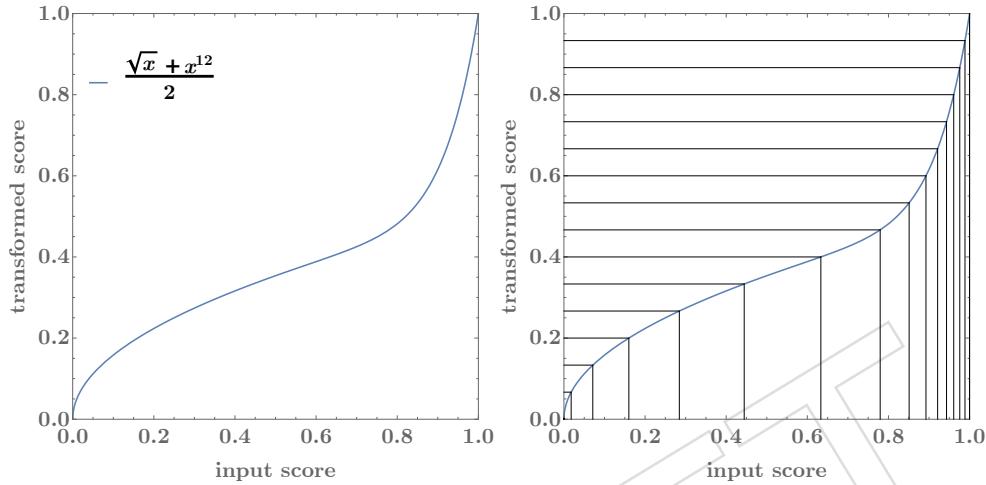


Figure 31: Derivation of binning for DNN. Left plot shows the transformation function and the right plot shows the obtained binning.

1086 10 Signal Efficiency

1087 The relative signal and background efficiency after each cut defining the signal regions de-
 1088 scribed in Sec. 8.3 are shown in Fig. 32 for the 1-lepton and 2-lepton categories.

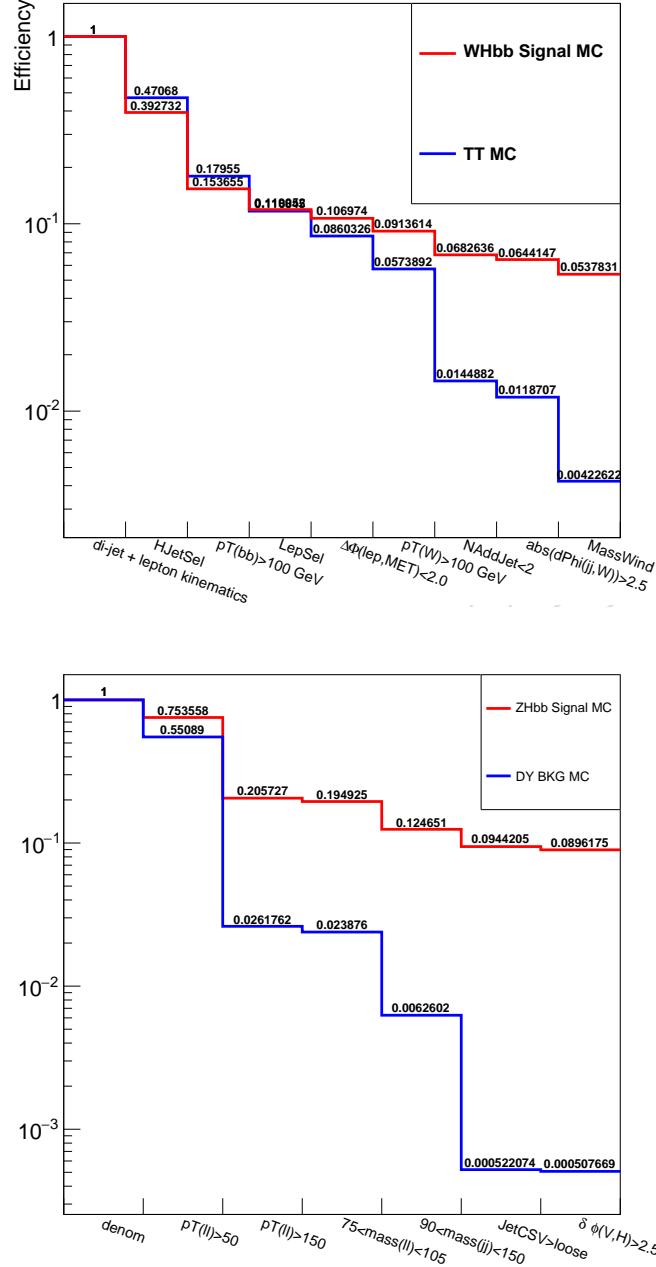


Figure 32: Relative efficiency after each cut defining the signal regions, for signal and the dominant background. Top: relative efficiency for $W(\ell\nu)H$ and $t\bar{t}$ in the 1-lepton category. The plot has been obtained requiring at least two jets and one lepton with $p_T > 25$ GeV. Bottom: relative efficiency for $Z(\ell\ell)H$ and $Z+jets$ in the high $p_T(V)$ region of the 2-lepton category. The plot has been obtained requiring at least two jets and one lepton with $p_T > 20$ GeV

1089 10.1 Lepton reconstruction and ID efficiencies

1090 Muon reconstruction efficiency data/MC scale factors in 2017 data are computed in two steps.
1091 First, the efficiency for any global muon to be reconstructed as a loose (tight) muon is found,
1092 followed by the efficiency for a loose (tight) muon to pass the isolation cuts. The efficiency for
1093 the entire 2017 dataset is measured in a single analysis. The results for the loose identification
1094 and the loose isolation on top of a loose identified muon are shown in Fig. 33. The results for the
1095 tight identification and the tight isolation on top of a loose identified muon are shown in Fig. 34
1096 for Run BCDEF. For both working points, the scale factors in the ratio plot are not strongly
1097 dependent on p_T and appear only as a function of η , although these factors are parameterized
1098 in both variables when applied in the analysis.

1099 Electron reconstruction efficiency data/MC scale factors in 2017 data are computed together for
1100 the MVA-based identification cut and for the isolation cut. Two working points are considered
1101 WP80 and WP90. The data efficiency corrections for identification and isolation requirement
1102 for these working points appear in Figure 35.



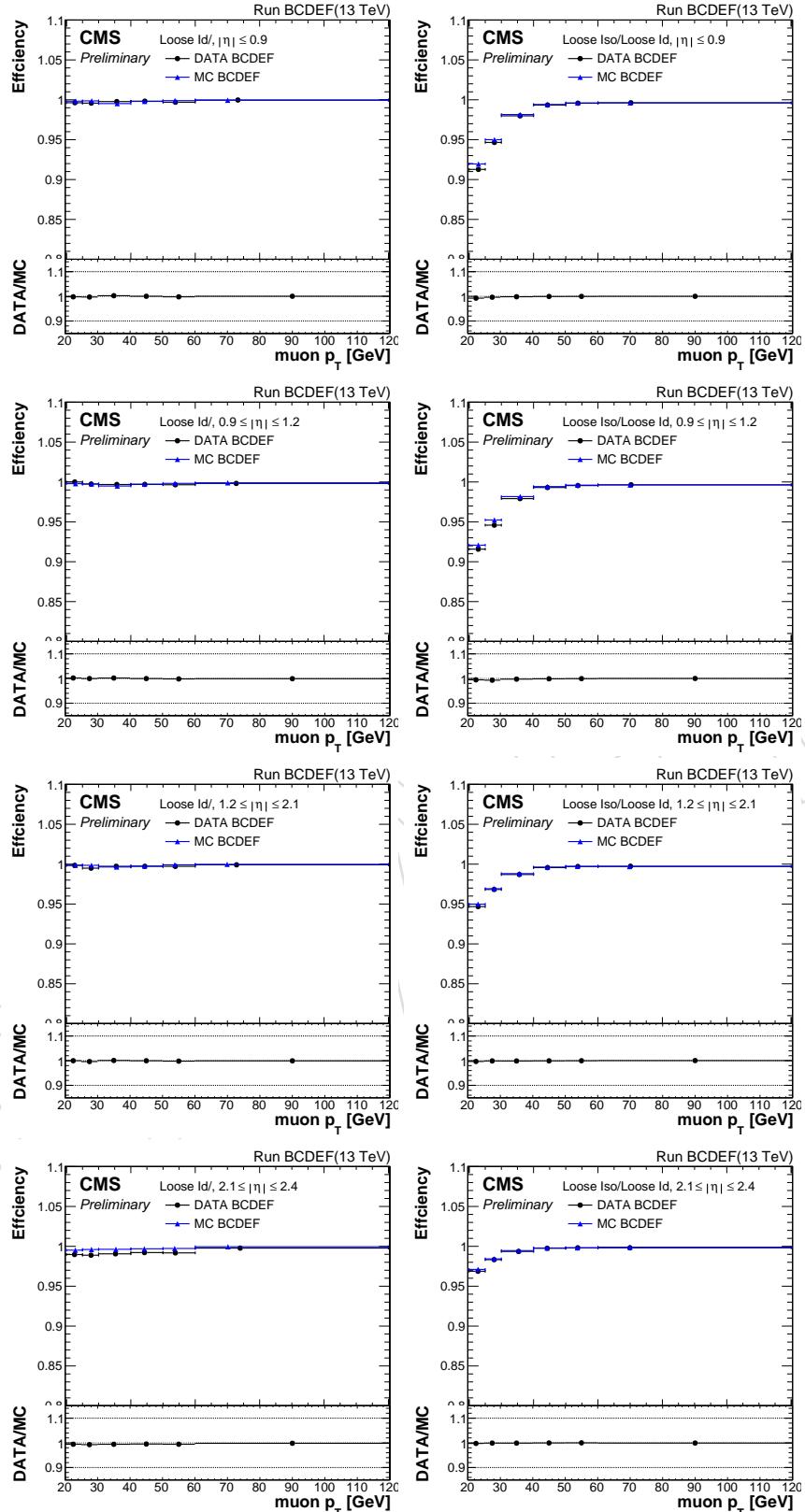


Figure 33: 2017 data and MC efficiency for a muon to pass the loose ID (left) and for it to pass the loose ISO after loose ID (right), distributed in the muon p_T . The efficiency have been computed in 4 $|\eta|$ bins. From top to bottom: $|\eta| < 0.9$, $0.9 < |\eta| < 1.2$, $1.2 < |\eta| < 2.1$, $2.1 < |\eta| < 2.4$

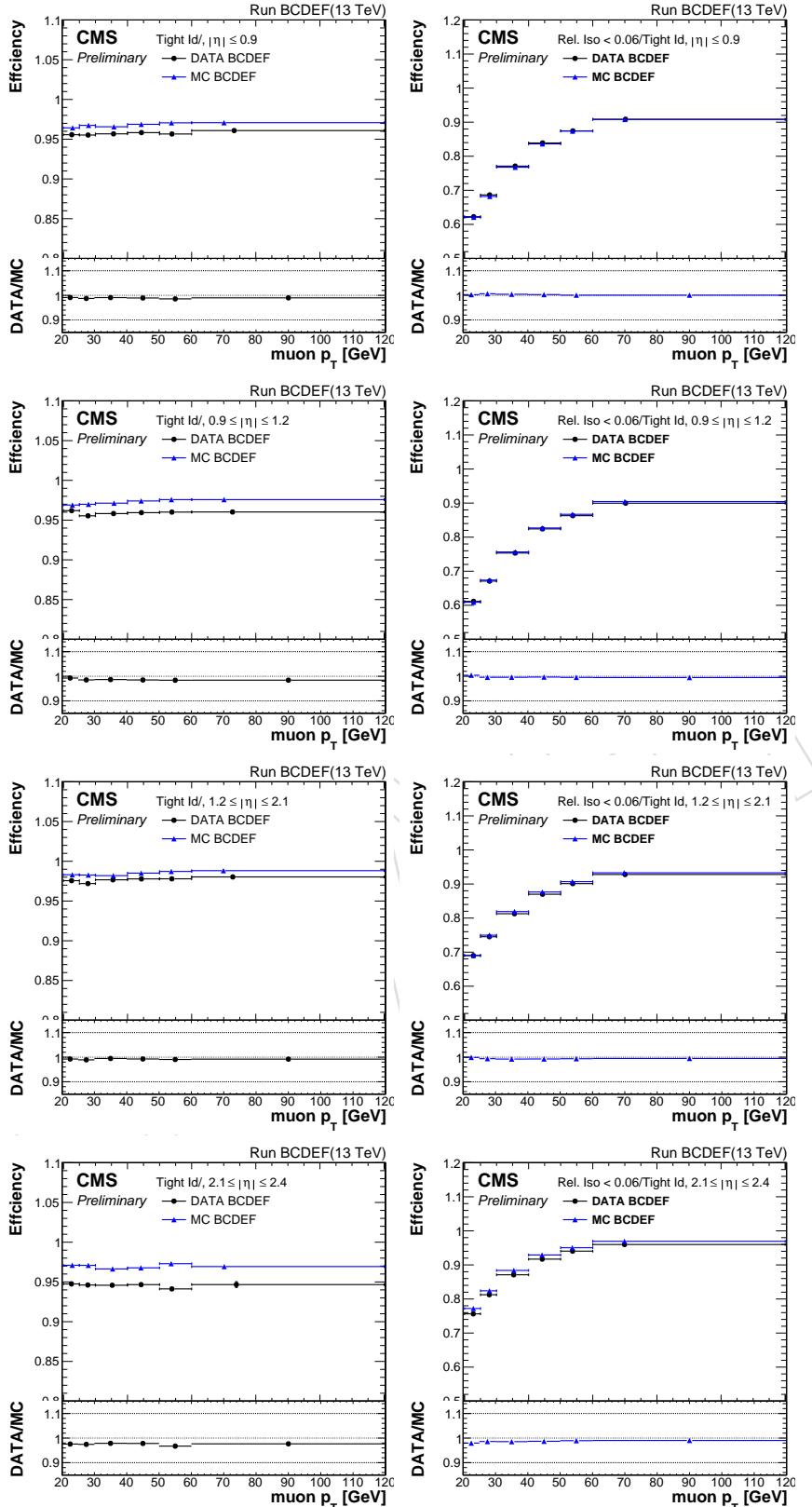


Figure 34: 2017 data and MC efficiency for a muon to pass the tight ID (left), and for it to pass tight ISO after tight ID (right), distributed in the muon p_T . The efficiency have been computed in 4 $|\eta|$ bins. From top to bottom: $|\eta| < 0.9$, $0.9 < |\eta| < 1.2$, $1.2 < |\eta| < 2.1$, $2.1 < |\eta| < 2.4$

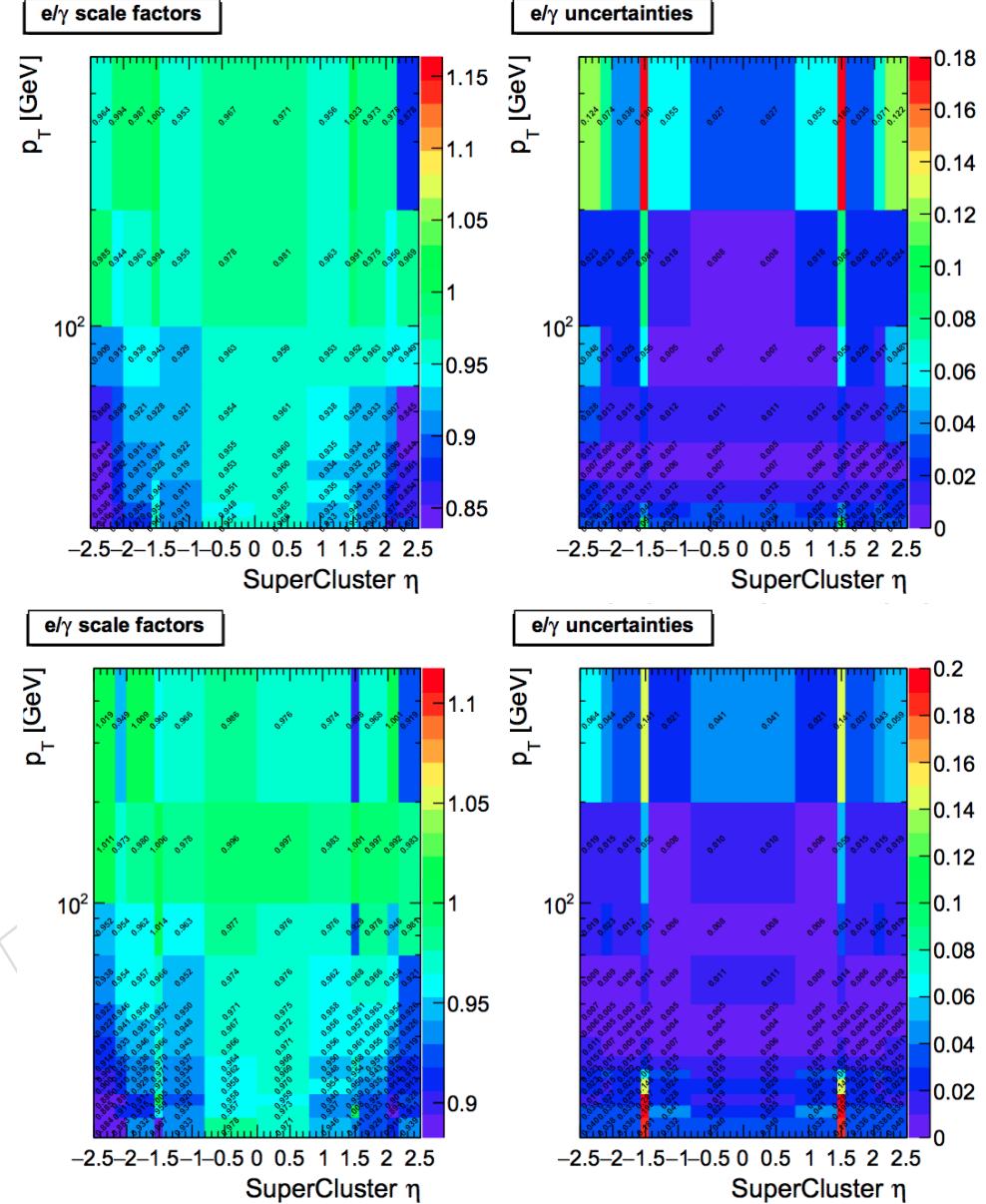


Figure 35: Efficiency corrections in 2017 data for identification MVA cut for the electron selection for the WP80 (above) and WP90 (below) working points and isolation cuts computed in bins of electron η and p_T . The uncertainties per correction is shown in the left. The uncertainties are generally under 2%.

1103 11 Background Control Regions

1104 Appropriate control regions (CR) are identified in data and used to adjust Monte Carlo nor-
 1105 malization predictions for several of the most important background processes, including pro-
 1106 duction of W and Z bosons in association with jets, and $t\bar{t}$.

1107 All the control regions are expressly designed to be mutually exclusive with each other and
 1108 with respect to the signal regions (SR), at same time allowing for minimal extrapolation uncer-
 1109 tainties. A cartoon to illustrate the basic properties of the CR design for the three channels is
 1110 shown in Fig. 36.

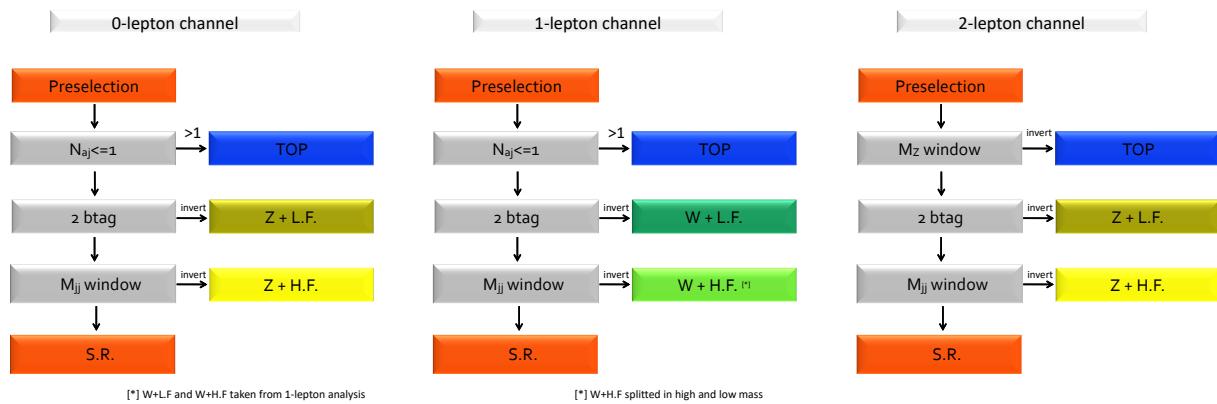


Figure 36: Cartoon to illustrate the basic properties of the CR design for the three channels. The W+jets CR for the 0-lepton channel are not shown as they are not used in the fits, as described in Sec. 11.5.

1111 This section contains the control region plots for the 2017 analysis.

11.1 0-lepton channel control regions

This section describes the control regions for $Z + \text{udscg}$, $Z + b\bar{b}$, and $t\bar{t}$ production as reconstructed in the 0-lepton channel, characterized by the $E_T^{\text{miss}} > 170 \text{ GeV}$ cut.

The cuts used to define the control regions are reported in Tab. 17. For comparison, the signal region selection is reported in Table 14.

- The $Z + b$ -jets control region is the most similar to the signal region, only requiring inverted m_{jj} selection. The anti-QCD cut is used to increase the $Z + b$ -jets purity.
- The $Z + \text{light-jets}$ control region is defined inverting the b -tagging cut and removing the m_{jj} cut. The remaining cuts are identical to the $Z + b$ -jets control region.
- The $t\bar{t}$ control region is defined by requiring at least two additional jets (besides the two b -jets) with $p_T > 30 \text{ GeV}$, at least a medium b -tagged jet, and at least one isolated lepton.

The corresponding plots are shown in Fig. 37–39 for 2017 data.

Table 17: Definition of control regions for $Z(\nu\nu)\text{H}$. The values listed for kinematical variables are in units of GeV.

Variable	$t\bar{t}$	$Z + \text{udscg}$	$Z + b\bar{b}$
V Decay Category	$W(\ell\nu)$	$Z(\nu\nu)$	$Z(\nu\nu)$
$p_T(j_1)$	> 60	> 60	> 60
$p_T(j_2)$	> 35	> 35	> 35
$p_T(jj)$	> 120	> 120	> 120
E_T^{miss}	> 170	> 170	> 170
$\Delta\phi(V, H)$	> 2	> 2	> 2
N_{al}	≥ 1	$= 0$	$= 0$
N_{aj}	≥ 2	≤ 1	≤ 1
$M(jj)$	—	—	$\notin [60 - 160]$
$btag_{max}$	$>\text{Medium}$	$<\text{Medium}$	$>\text{Tight}$
$btag_{min}$	$>\text{Loose}$	Loose	$>\text{Loose}$
$\Delta\varphi(j, E_T^{\text{miss}})$	—	> 0.5	> 0.5
$\Delta\varphi(\text{tkMET}, E_T^{\text{miss}})$	—	< 0.5	< 0.5
$\min \Delta\varphi(j, E_T^{\text{miss}})$	$< \pi/2$	—	—

1125 **11.1.1 2017**

1126 $t\bar{t}$ -enriched control region in the 0-lepton channel for the 2017 analysis is shown in Fig. 37.

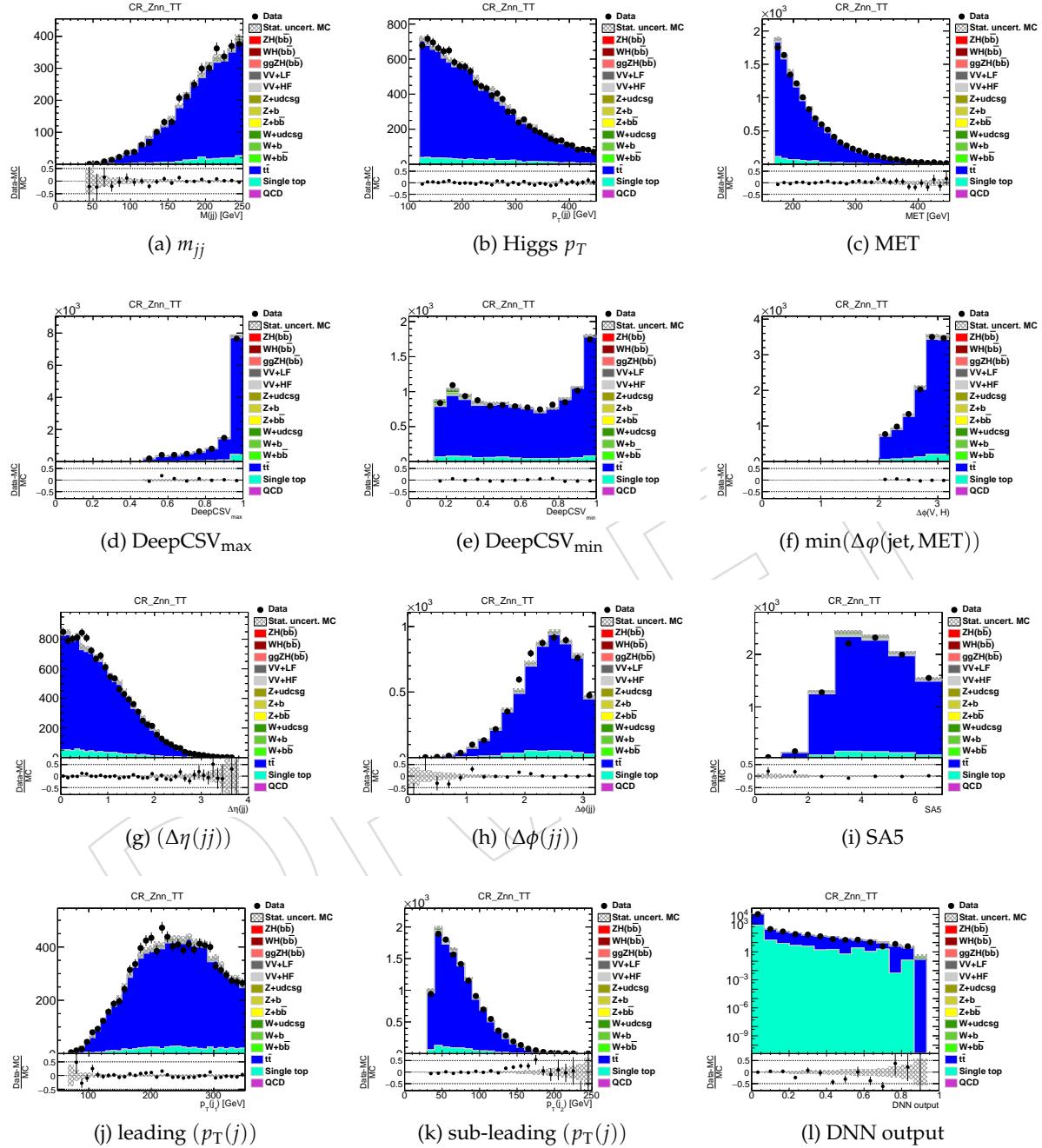


Figure 37: Distributions of some variables for the $t\bar{t}$ -enriched control region in the 0-lepton channel for the 2017 analysis. The plots use the scale factors reported in Sect. 11.5.

1127 Z+light-jets control region in the 0-lepton channel for the 2017 analysis is shown in Fig. 38.

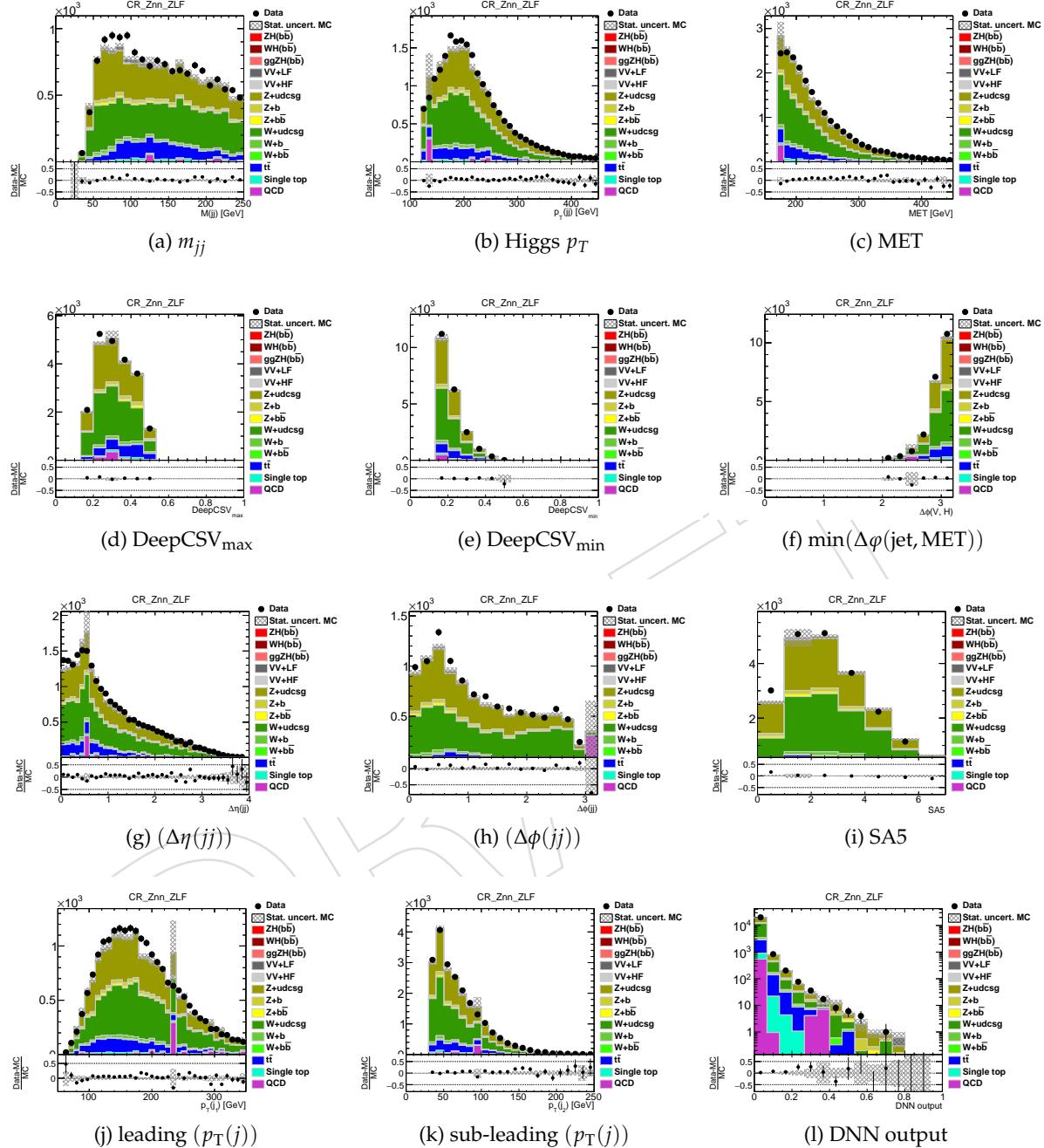


Figure 38: Distributions of some variables for the Z+light-jets control region in the 0-lepton channel for the 2017 analysis. The plots use the scale factors reported in Sect. 11.5.

1128 Z+b-jets control region in the 0-lepton channel for the 2017 analysis is shown in Fig. 39.

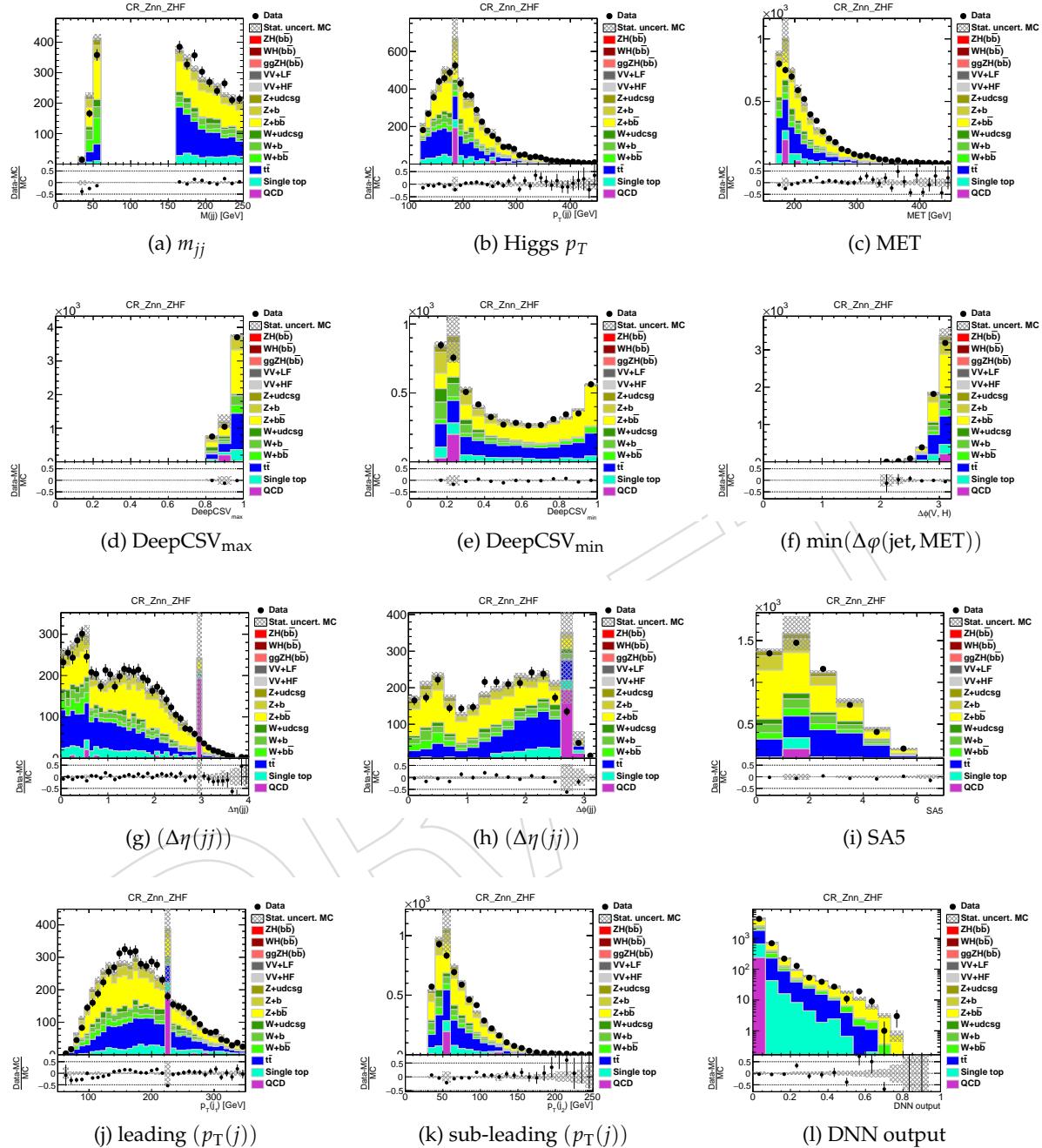


Figure 39: Distributions of some variables for the Z+b-jets control region in the 0-lepton channel for the 2017 analysis. The plots use the scale factors reported in Sect. 11.5.

11.2 1-lepton channel control regions

This section describes the control regions for $W + \text{udscg}$, $W + b\bar{b}$, and $t\bar{t}$ production as reconstructed in the 1-lepton channel. The selection criteria, common for the electrons and muons, are summarized in Table 18. For comparison, the signal region selection is reported in Table 14.

- The $W + \text{udscg}$ control region is defined by inverting the b-tagging requirement of the signal region, to enhance the light flavor jets contribution.
- The $t\bar{t}$ control region is defined by requiring one tight b-tag and increasing the requirement on number of additional jets to > 1 (besides the two b-jets).
- The $W + b\bar{b}$ control region is the most similar to the signal region. We apply the same criteria on the jet with highest b-tagging score as for $t\bar{t}$, but require that there are no additional jets in the event (besides the two b-jets). In addition, we apply a dijet invariant mass veto to remove overlap with the signal region, removing the $W(\ell\nu)H$ contamination.

The corresponding plots are shown in Fig. 40–42 for 2017 data.

Following re-optimization of the analysis strategy, a $p_T(V) > 150 \text{ GeV}$ cut is used consistently with the signal region definition (cfr Sec. 8.3), instead of the $p_T(V) > 100 \text{ GeV}$ used in HIG-16-044.

Table 18: Definition of control regions for the 0-lepton channel, common for the electrons and muons. N_{al} is the number of additional isolated leptons in the event. METsig is the significance of the E_T^{miss} . The values listed for kinematical variables are in units of GeV.

Variable	$W + \text{udscg}$	$t\bar{t}$	$W + b\bar{b}$
$p_T(j_1)$	> 25	> 25	> 25
$p_T(j_2)$	> 25	> 25	> 25
$p_T(jj)$	> 100	> 100	> 100
$p_T(V)$	$> 150^{(**)}$	$> 150^{(**)}$	$> 150^{(**)}$
$btag_{\text{max}}$	< Medium	> Tight	> Tight
N_{aj}	–	> 1	= 0
N_{al}	= 0	= 0	= 0
METsig	> 2.0	–	> 2.0
$\Delta\phi(\text{pfMET}, \text{lep})$	< 2	< 2	< 2
$M(jj)$	< 250	< 250	$< 250, \text{veto } [90 - 150]$

(**) New with respect to the previous Run2 analysis on 2016 data described in HIG-16-044 [30] and AN-15-168 [32].

11.2.1 2017

1146 $t\bar{t}$ -enriched muon control region in the 1-lepton channel for the 2017 analysis is shown in Fig. 40.

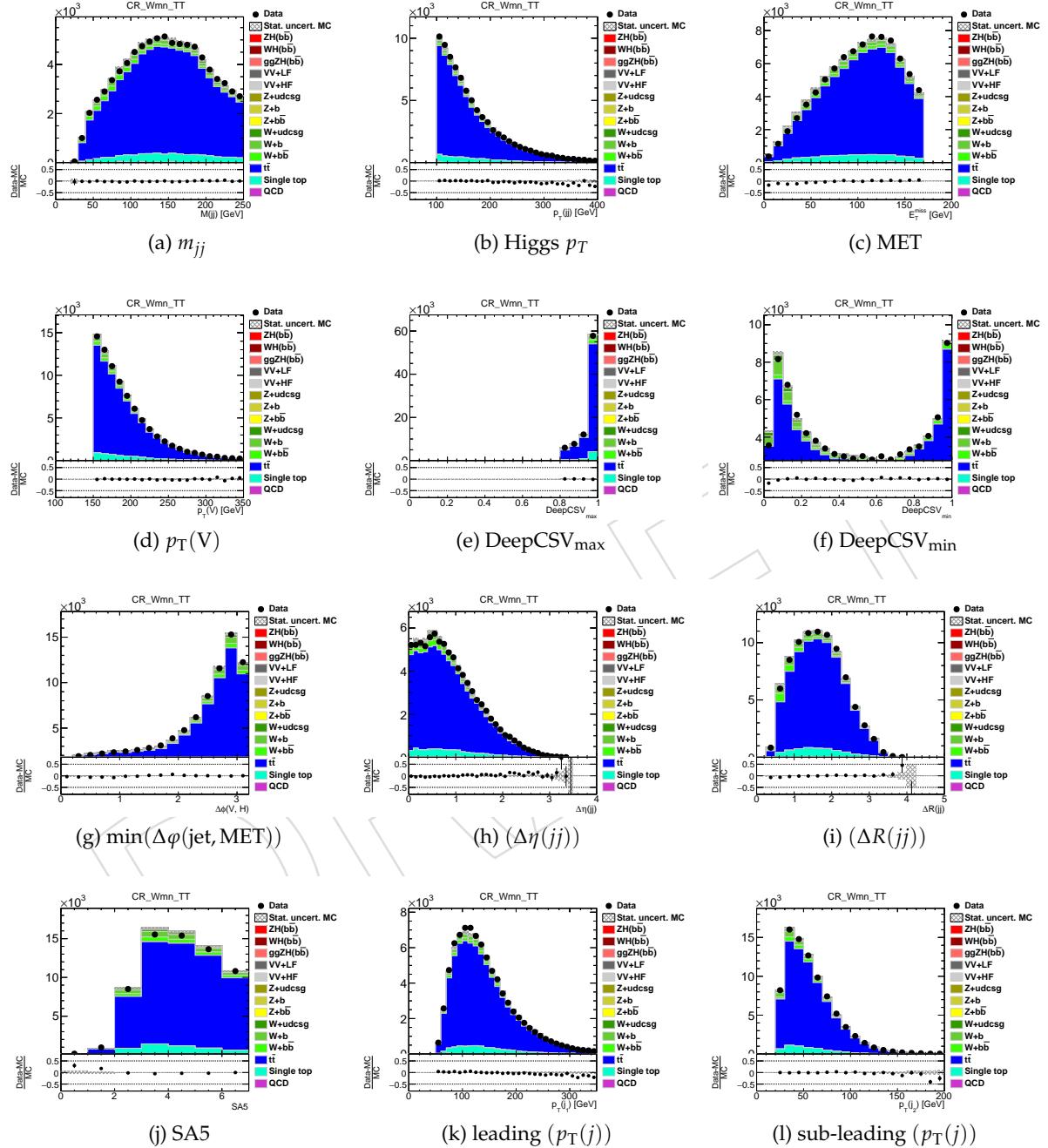


Figure 40: Distributions of some variables for the $t\bar{t}$ -enriched muon control region in the 1-lepton channel for the 2017 analysis. The plots use the scale factors reported in Sect. 11.5.

1148 W($\mu\nu$) + udscg-enriched electron control region in the 1-lepton channel for the 2017 analysis is
 1149 shown in Fig. 41.

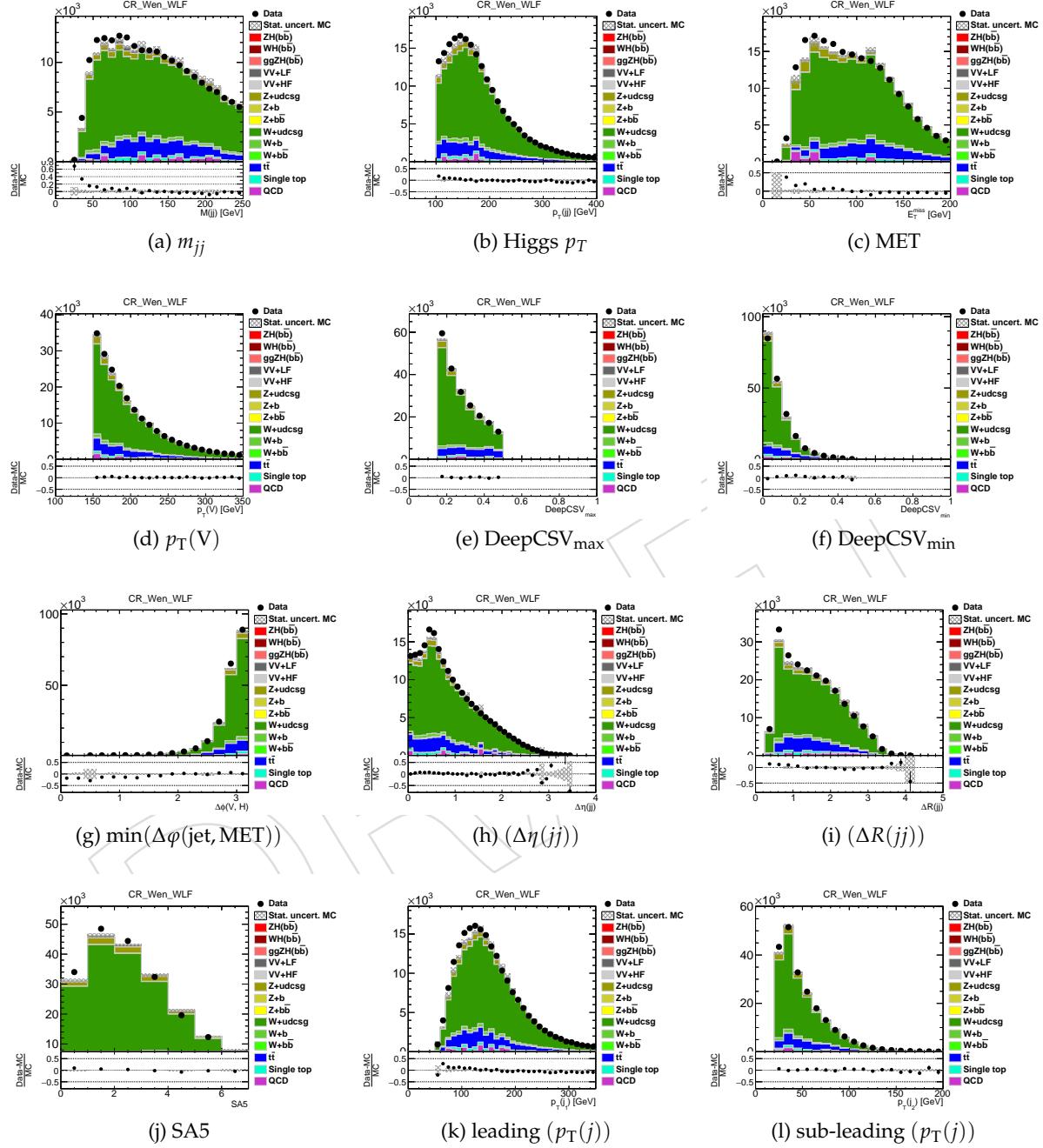


Figure 41: Distributions of some variables for the W(e ν) + udscg electron control region in the 1-lepton channel for the 2017 analysis. The plots use the scale factors reported in Sect. 11.5.

1150 W + $\bar{b}b$ -enriched muon control region in the 1-lepton channel for the 2017 analysis is shown in
 1151 Fig. 42.

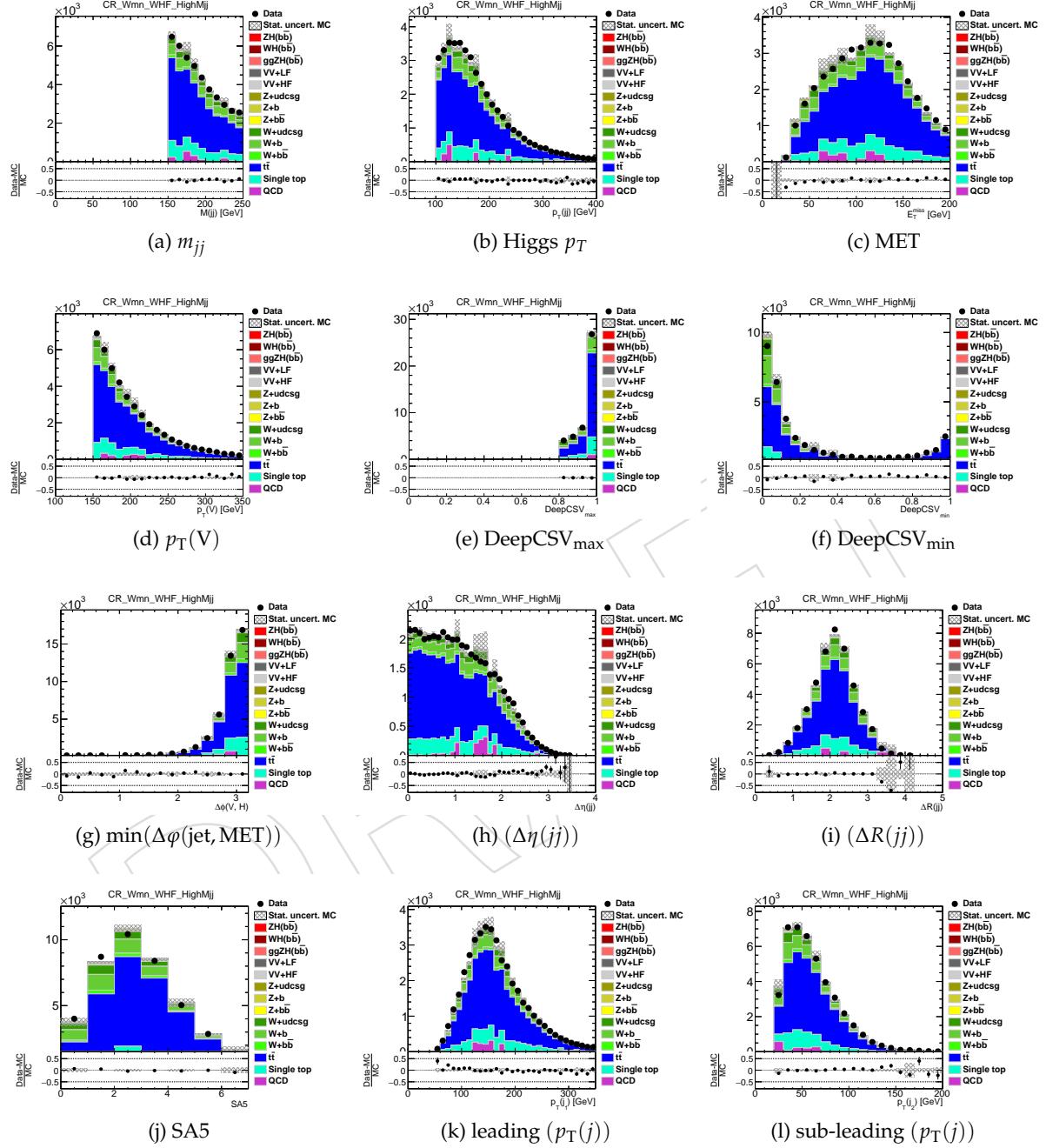


Figure 42: Distributions of some variables for the $W + \bar{b}b$ muon control region in the 1-lepton channel for the 2017 analysis. The plots use the scale factors reported in Sect. 11.5.

1152 11.3 2-lepton channel control regions

1153 This section describes the control regions for $Z + \text{udscg}$, $t\bar{t}$, and $Z + b\bar{b}$ production as recon-
 1154 structed in the 2-lepton channel. The selection criteria, common for the electrons and muons,
 1155 are summarized in Table 19. For comparison, the signal region selection is reported in Table 14.

- 1156 • The $Z + \text{udscg}$ control region is defined by inverting the b-tagging requirement of
 1157 the signal region, to enhance the light flavor jets contribution.
- 1158 • The $t\bar{t}$ control region is defined by inverting the dilepton invariant mass cut.
- 1159 • The $Z + b\bar{b}$ control region is the most similar to the signal region. We apply an
 1160 inverted m_{jj} selection to remove any overlap.

1161 The corresponding plots are shown in Fig. 43–48 for 2017 data.

Table 19: Definition of control regions for the 2-lepton channel. The values listed for kinematical variables are in units of GeV.

Variable	$t\bar{t}$	$Z + \text{udscg}$	$Z + b\bar{b}$
$p_T(j_1)$	> 20	> 20	> 20
$p_T(j_2)$	> 20	> 20	> 20
$p_T(V)$	$[50, 150], > 150$	$[50, 150], > 150$	$[50, 150], > 150$
$btag_{max}$	$>\text{Tight}$	$<\text{Loose}$	$>\text{Tight}$
$btag_{min}$	$>\text{Loose}$	$<\text{Loose}$	$>\text{Loose}$
E_T^{miss}	–	–	< 60
$\Delta\phi(V, H)$	–	> 2.5	> 2.5
$m_{\ell\ell}$	$\notin [0, 10], \notin [75, 120]$	$[75, 105]$	$[85, 97]$
$M(jj)$	–	$[90, 150]$	$\notin [90, 150]$

11.3.1 2017

1162 $t\bar{t}$ -enriched muon control region in the low $p_T(V)$ bin for 2-lepton channel for the 2017 analysis
 1163 is shown in Fig. 43.

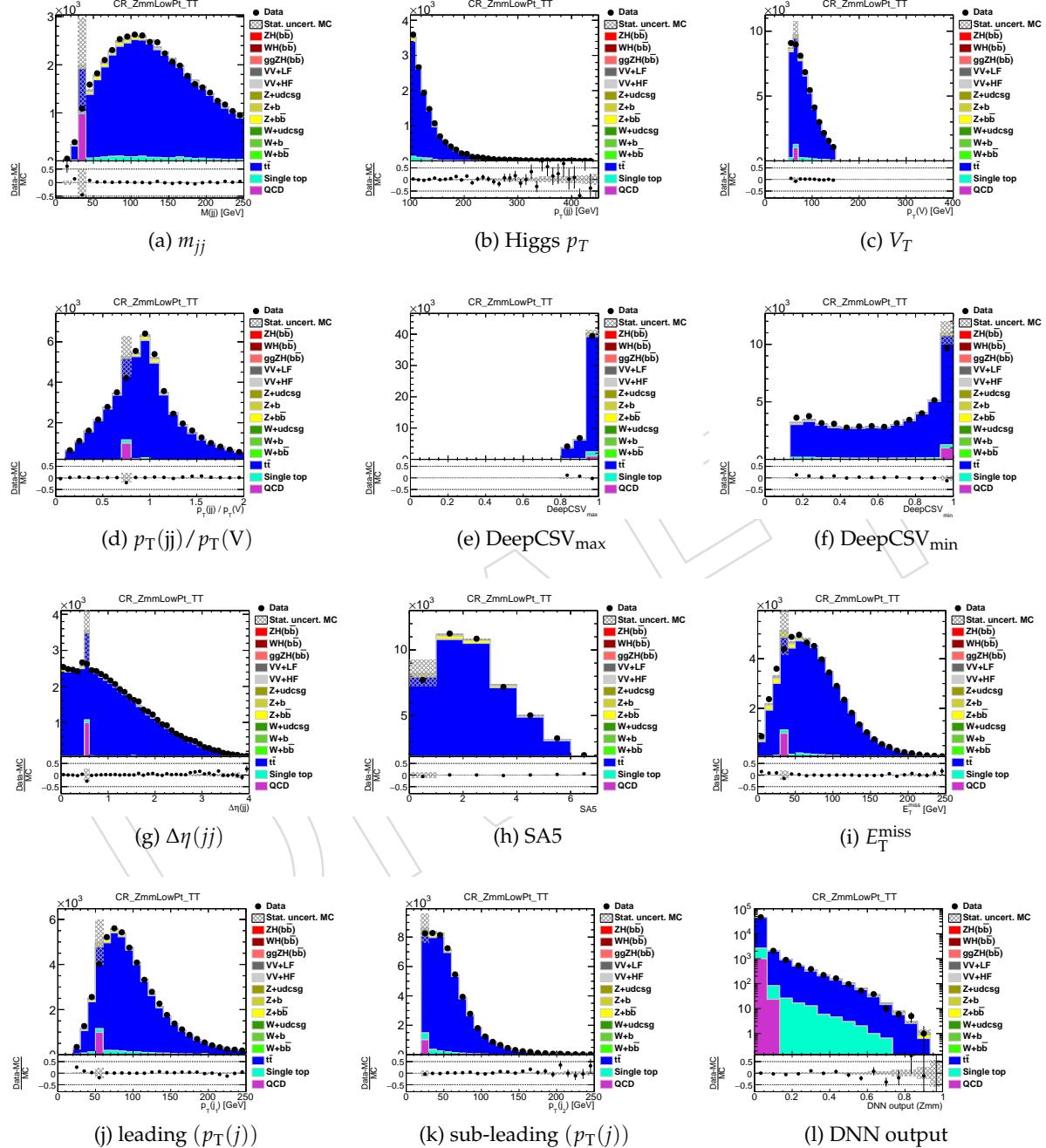


Figure 43: Distributions of variables in data and simulated samples in the $t\bar{t}$ muon control region for 2-lepton channel for the 2017 analysis in the low $p_T(V)$ bin. The plots use the scale factors reported in Sect. 11.5.

1165 $t\bar{t}$ -enriched muon control region in the high $p_T(V)$ bin for 2-lepton channel for the 2017 analysis
 1166 is shown in Fig. 44.

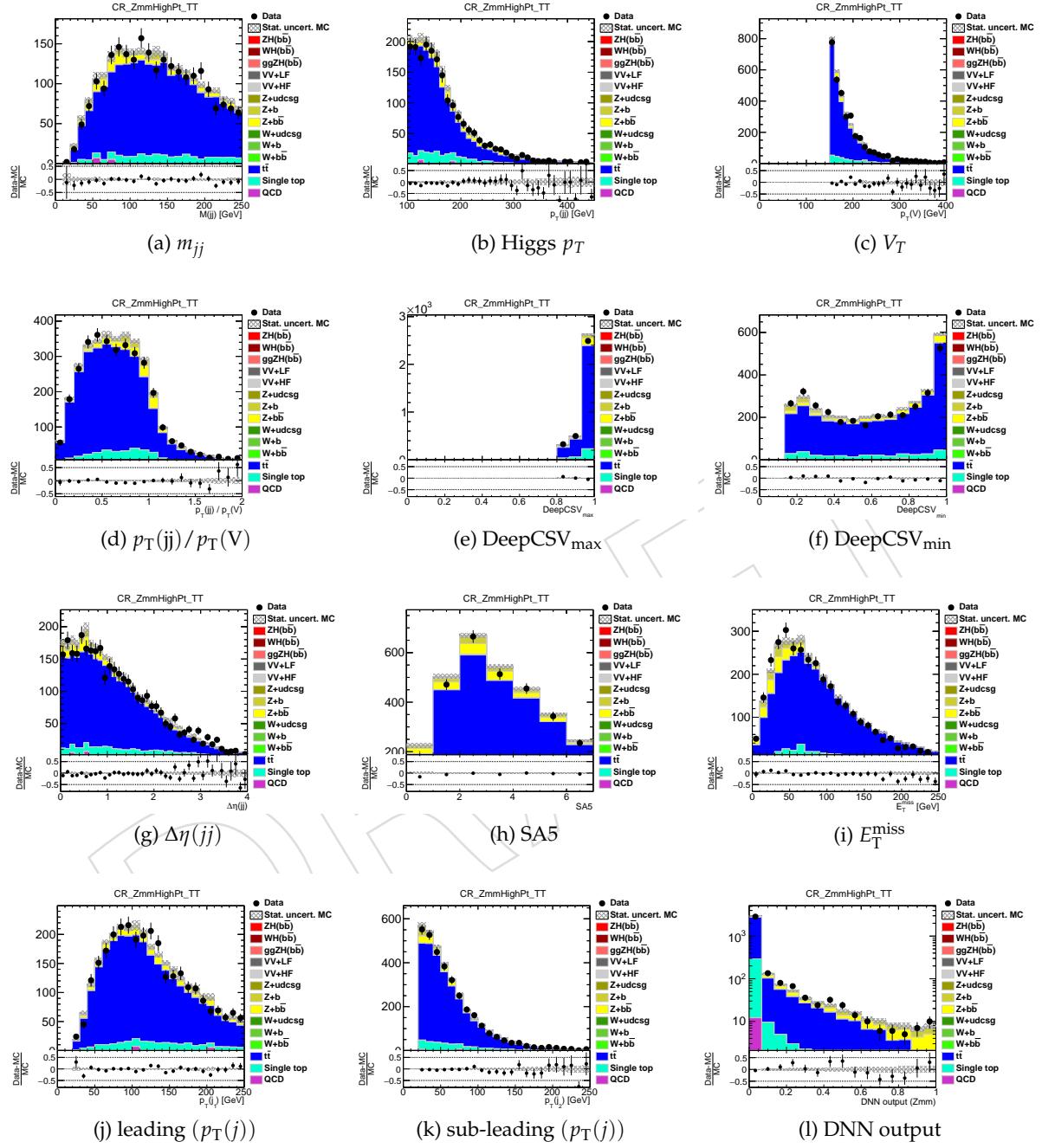


Figure 44: Distributions of variables in data and simulated samples in the $t\bar{t}$ muon control region for 2-lepton channel for the 2017 analysis in the high $p_T(V)$ bin. The plots use the scale factors reported in Sect. 11.5.

1167 Z($\mu\mu$) + udscg-enriched electron control region in the low $p_T(V)$ bin for 2-lepton channel for
 1168 the 2017 analysis is shown in Fig. 45.

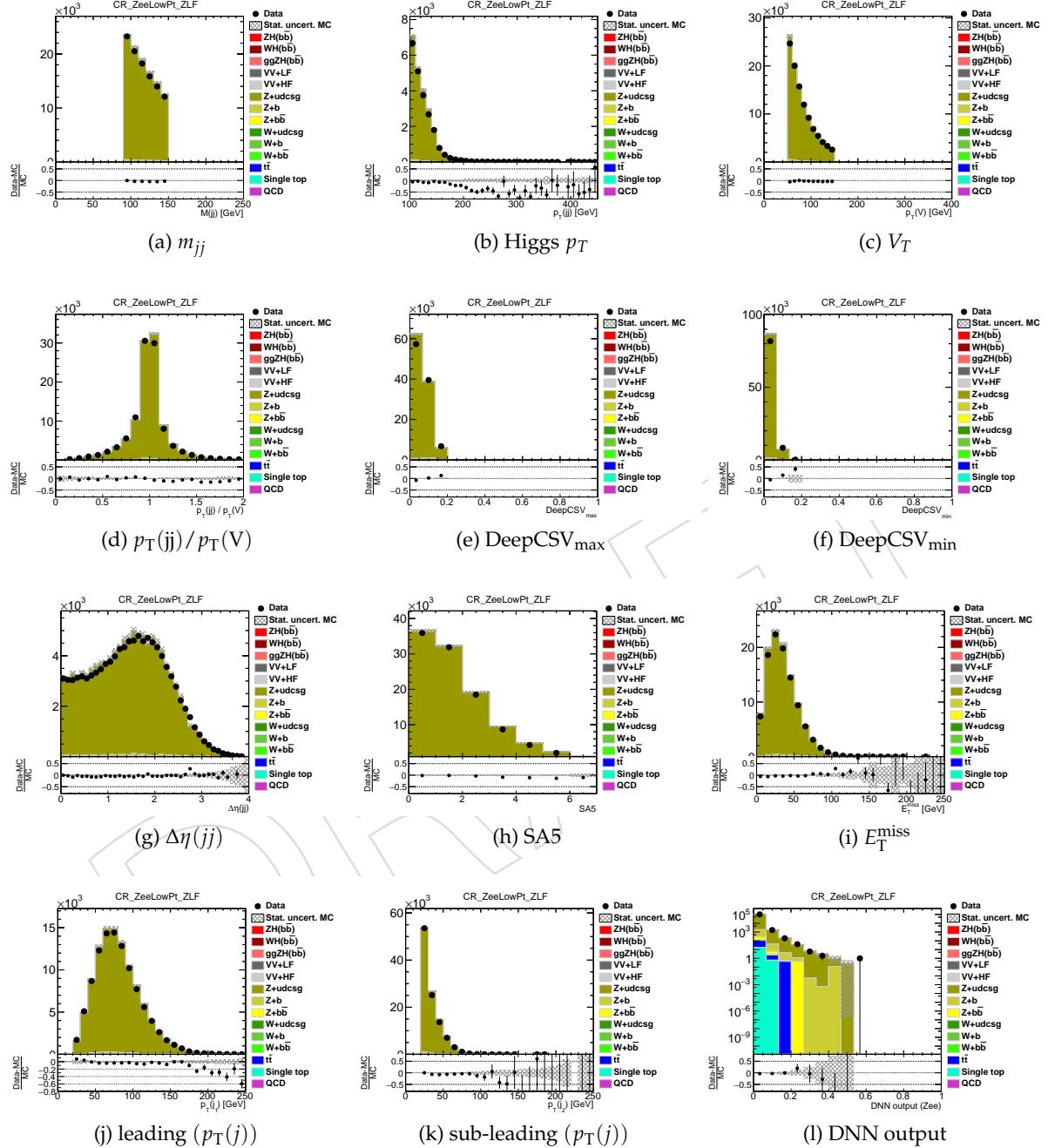


Figure 45: Distributions of variables in data and simulated samples in the Z($\mu\mu$) + udscg electron control region for 2-lepton channel for the 2017 analysis in the low $p_T(V)$ bin. The plots use the scale factors reported in Sect. 11.5.

1169 Z($\mu\mu$) + udscg-enriched electron control region in the high $p_T(V)$ bin for 2-lepton channel for
 1170 the 2017 analysis is shown in Fig. 46.

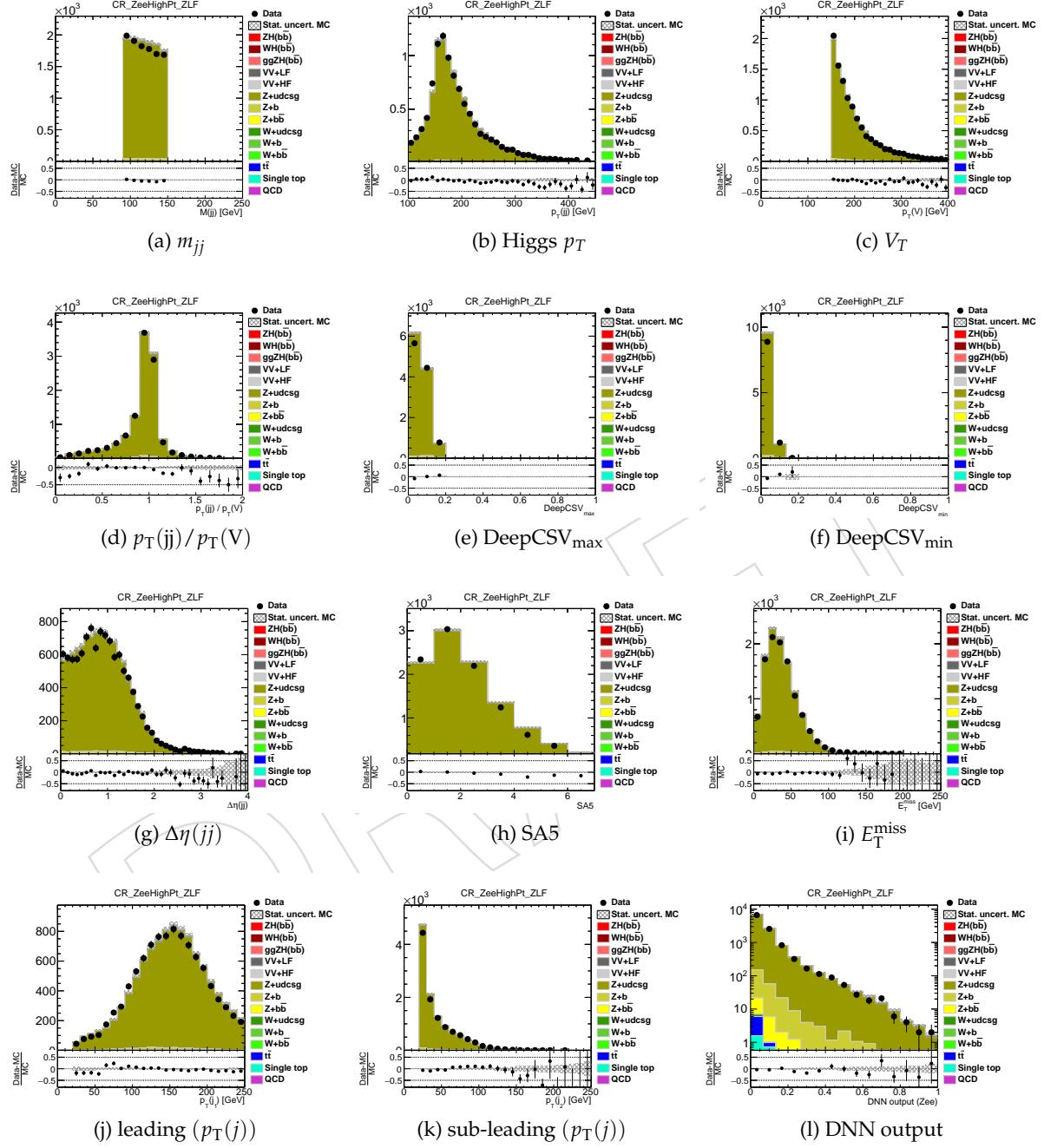


Figure 46: Distributions of variables in data and simulated samples in the Z($\mu\mu$) + udscg electron control region for 2-lepton channel for the 2017 analysis in the high $p_T(V)$ bin. The plots use the scale factors reported in Sect. 11.5.

1171 $Z(\mu\mu) + b\bar{b}$ -enriched muon control region in the low $p_T(V)$ bin for 2-lepton channel for the 2017
 1172 analysis is shown in Fig. 47.

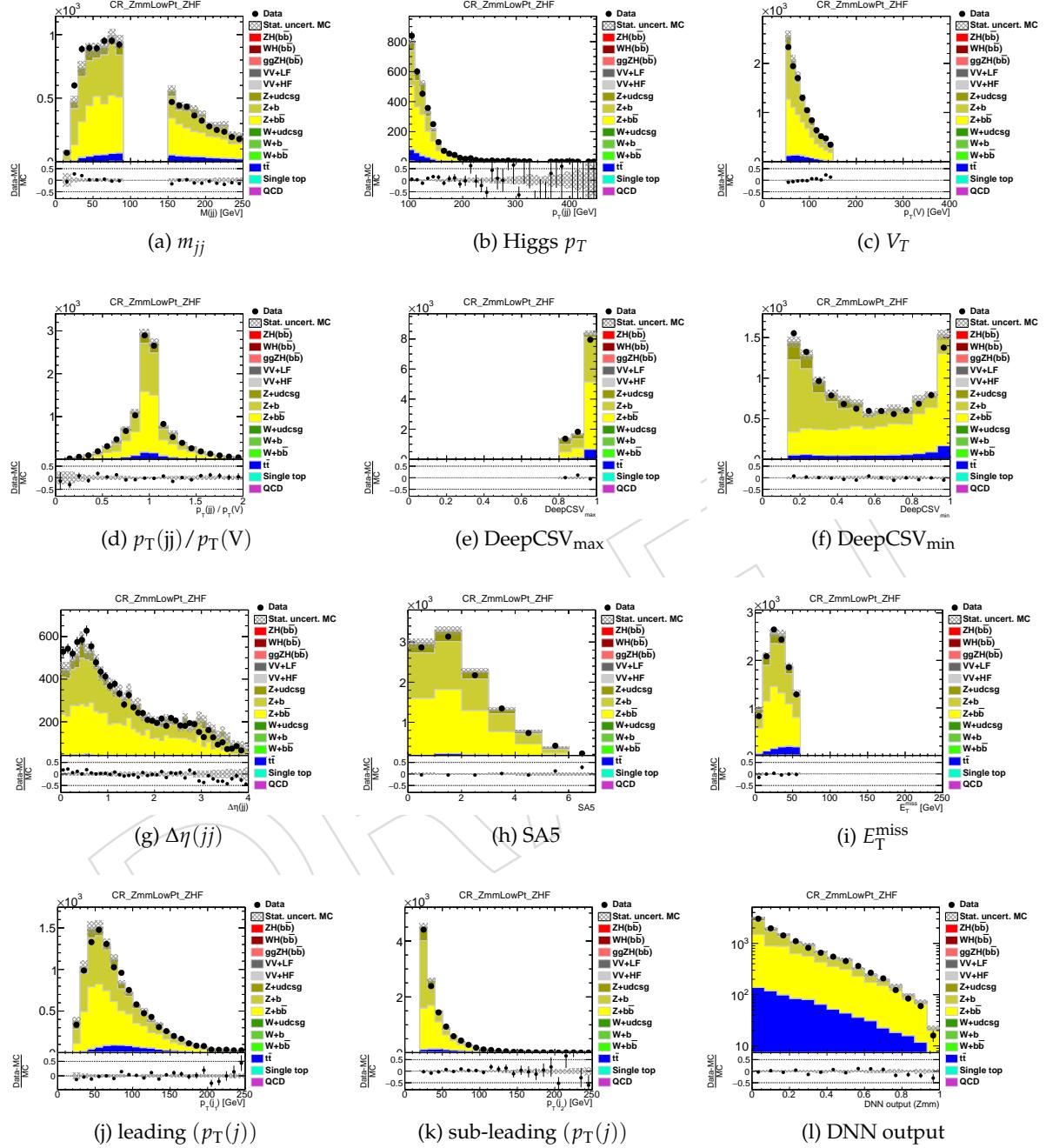


Figure 47: Distributions of variables in data and simulated samples in the $Z(\mu\mu) + b\bar{b}$ muon control region for 2-lepton channel for the 2017 analysis in the low $p_T(V)$ bin. The plots use the scale factors reported in Sect. 11.5.

¹¹⁷³ $Z(\mu\mu) + b\bar{b}$ -enriched muon control region in the high $p_T(V)$ bin for 2-lepton channel for the
¹¹⁷⁴ 2017 analysis is shown in Fig. 48.

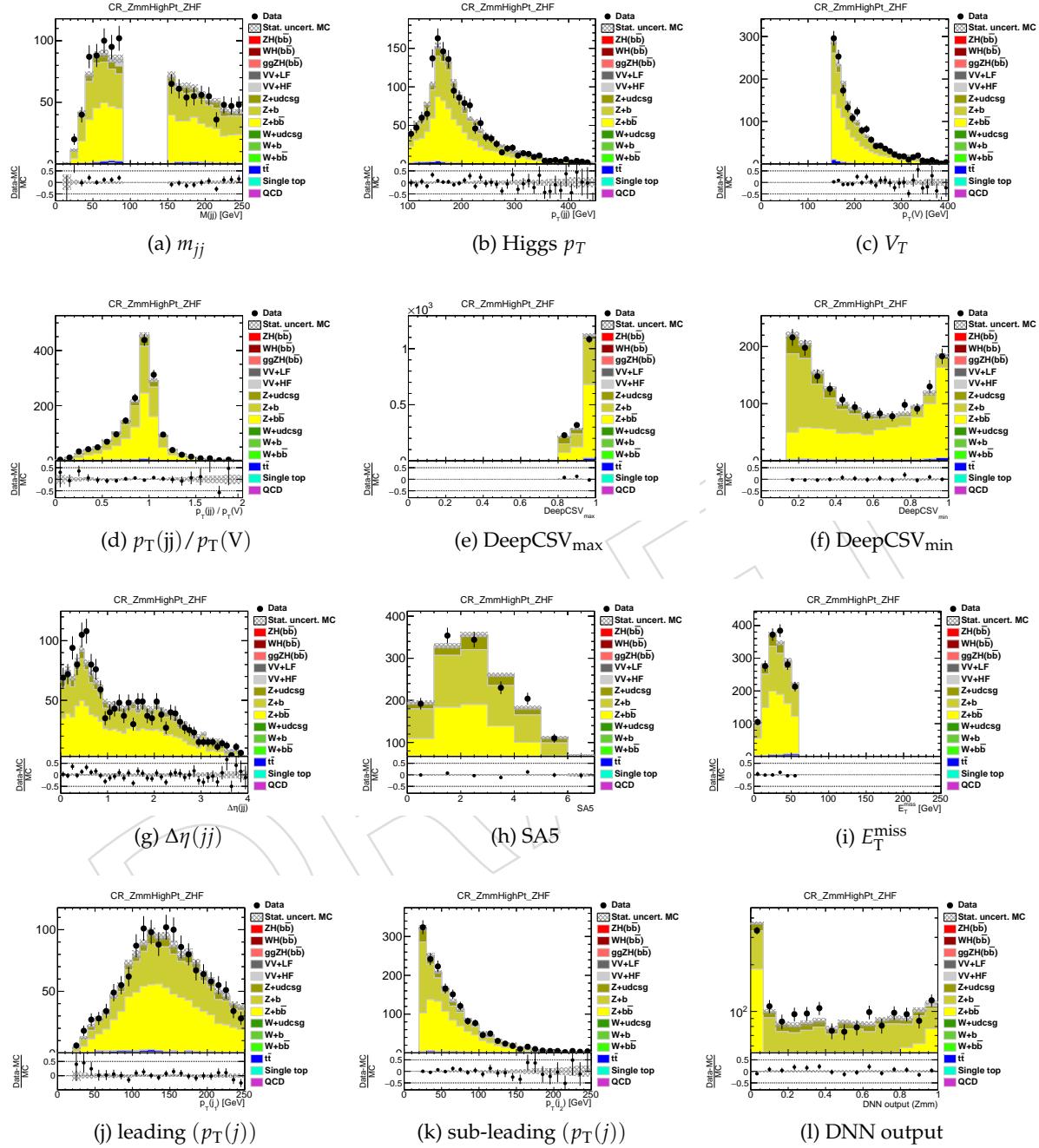


Figure 48: Distributions of variables in data and simulated samples in the $Z(\mu\mu) + b\bar{b}$ muon control region for 2-lepton channel for the 2017 analysis in the high $p_T(V)$ bin. The plots use the scale factors reported in Sect. 11.5.

11.4 DNN background classifier

1176 A multi-output DNN is trained in the heavy flavor control regions for the 0 and 1-lepton channel,
 1177 using the same input variables as for the signal background discrimination, given in table
 1178 15. This background classifier returns the 5-vector p of the probabilities of an event belonging
 1179 to the categories: V+BB, V+B, TT, ST, V+light (which is mostly V+light, but including also the
 1180 little contributions from Diboson and QCD.)

category	V+BB	TT	V+light	V+B	ST
bin	0	1	2	3	4

For use in the fit, shapes are filled with the argmax of the probabilities, resulting in 5 bins. (The convention for the bin numbers is given in table 11.4). For validation only, the individual bins are further split into regions of different purity with the formula

$$x = \text{argmax}(p) + (1 - 2(1 - \max(p))^4) \quad (9)$$

1182 and shown with in total 50 bins in Fig. 49.

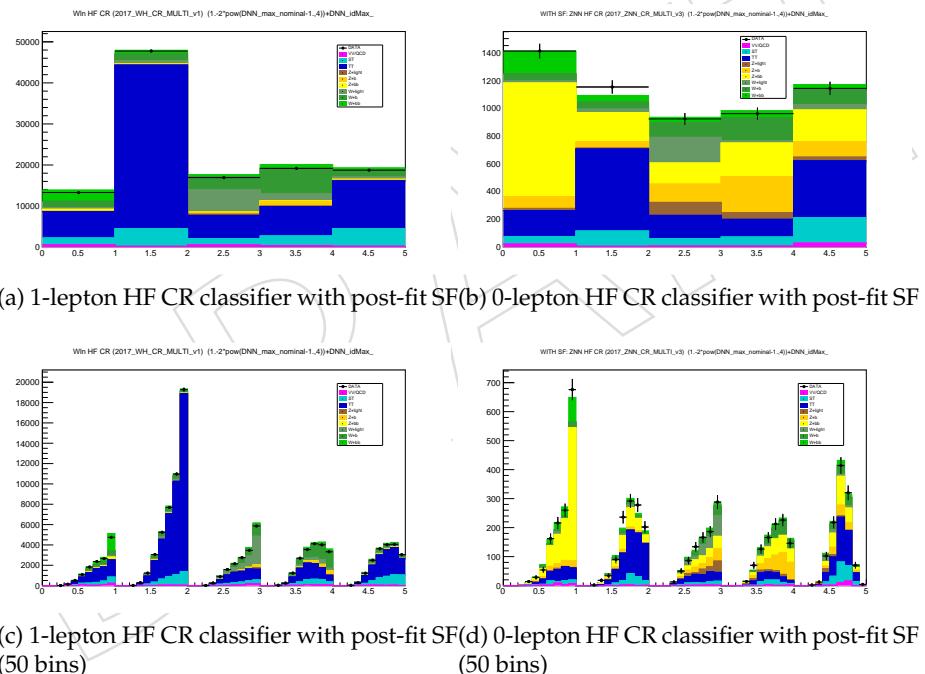


Figure 49: Output of multi-class background classifier in 1-lepton (left) and 0-lepton (right) channel for the 2017 analysis. The plots use the scale factors from the fit described in Sec. 11.5. First row shows the binning used in the fit, bottom rows shows a finer binning for validation only.

1183 The output of the multi-class background classifier is shown in Fig. 50. The blue curves show
 1184 the probabilities for the samples of the corresponding class whereas the red curves show the
 1185 probabilities for all others.

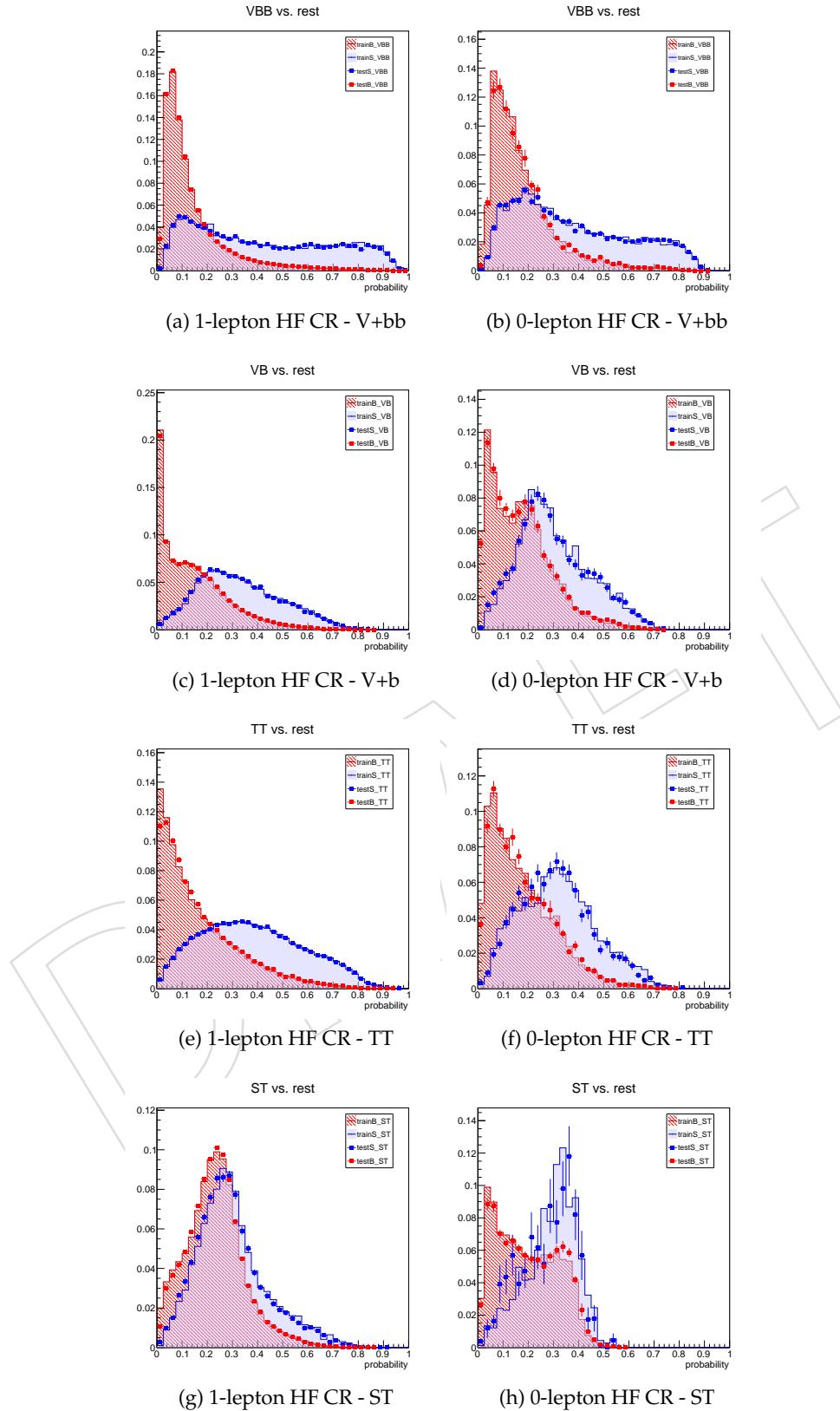


Figure 50: Output of multi-class background classifier in 1-lepton (left) and 0-lepton (right) channels for the 2017 analysis. The blue curves show the probabilities for the samples of the corresponding class whereas the red curves show the probabilities for all others. Test and training samples are drawn superimposed.

1186 11.5 Scale factor determination

1187 The normalizations of the $t\bar{t}$, $W+jets$, and $Z+jets$ backgrounds are fitted as free parameters in the
 1188 signal extraction (cfr. Sec. 13.1), for which all the control and signal regions are used (CR+SR
 1189 fit). The ratios of the fitted normalizations to those predicted by MC are usually referred to as
 1190 Scale Factors (SF).

1191 Preliminary SF are needed also before the signal extraction, to train the multivariate discriminant
 1192 using the proper background yields and to test the MC modelling of its input variables.
 1193 For this purpose, a fit is performed using all the control regions simultaneously (CR-only fit).

1194 In all cases, the fits to data are performed through binned maximum likelihood (cfr. Sec. 13.1)
 1195 using templates derived from simulation. All the systematic uncertainties described in Sec. 12
 1196 are included, to take into account potential shape differences between data and MC.

1197 In the spirit of capturing the possible MC prediction mis-modelling in the most accurate way
 1198 for each process, we use different SF for the Vector bosons produced with a single b ($Z + b$,
 1199 $W + b$) or a $b\bar{b}$ pair ($Z + b\bar{b}$, $W + b\bar{b}$), given that they come from different sets of Feynman
 1200 diagrams. Furthermore, the SF do not account only for cross section discrepancies, but also
 1201 potential residual differences in the physics object selection. Therefore, different scale factors
 1202 are used for the same processes in different channels, with the exception of the $W+jets$ ones
 1203 in the 0-lepton channel which are in common (i.e. correlated in the fits) with the 1-lepton
 1204 channel. Finally, in the 1- and 2-lepton lepton channels the scale factors are correlated between
 1205 muons and electrons, given that the lepton efficiencies are taken into account as systematic
 1206 uncertainties.

1207 In previous versions of this analysis, the most discriminating variable to disentangle the vari-
 1208 ous backgrounds had been identified as the $btag_{min}$ distribution, which was used to fit the CR.
 1209 In the current version of the analysis, a different approach is used, aimed at further stabilizing
 1210 the fit and maximizing the analysis sensitivity:

- 1211 • **V+light flavor and TTbar CR:** Fit to DeepCSV distribution reduced to 1 bin (i.e.
 1212 yield) for each channel

1213 - Fit to DeepCSV distribution reduced to 1 bin (i.e. yield) for each channel

- 1214 • **V+heavy flavor:**

1215 - Fit to DeepCSV distribution reduced to 2 bins for 2-lepton channel

1216 - Fit to background multi-classifier DNN described in Sec. 11.4 reduced to 5 bins
 1217 (1 bin per category) for 0- and 1-lepton channels

1218 Table 20 summarizes the fit results for the 2017 data for the SR+CRs fits.

Table 20: Data/MC scale factors for the 2017 analysis in the 0-, 1- and 2-lepton channels from SR+CRs fit. The errors include both statistical and systematic uncertainties. Compatible fitted values are obtained from the CR-only fit.

Process	$Z(\nu\nu)H$	$W(\ell\nu)H$	$Z(\ell\ell)H$ low- p_T	$Z(\ell\ell)H$ high- p_T
$W + udscg$	1.05 ± 0.08	1.05 ± 0.08	–	–
$W + b$	2.13 ± 0.15	2.13 ± 0.15	–	–
$W + b\bar{b}$	1.90 ± 0.19	1.90 ± 0.19	–	–
$Z + udscg$	0.95 ± 0.09	–	0.88 ± 0.06	0.81 ± 0.05
$Z + b$	0.98 ± 0.17	–	0.93 ± 0.12	1.16 ± 0.10
$Z + b\bar{b}$	1.19 ± 0.10	–	0.82 ± 0.06	0.88 ± 0.08
$t\bar{t}$	1.00 ± 0.08	0.94 ± 0.08	0.89 ± 0.07	0.91 ± 0.08

DRAFT

12 Systematics

1220 The primary physics results described in this note are:

- 1221 • the signal strength and uncertainty of an excess compatible with the signature of
1222 a standard model Higgs boson produced in association with a vector boson and
1223 decaying to a $b\bar{b}$ pair
- 1224 • the significance of a signal-like excess.

1225 Both are affected by systematic uncertainties on the expected signal and background yields
1226 and shapes, which need to be properly taken into account. Several sources have been con-
1227 sidered, which are propagated consistently to the DNN output in the signal regions and the
1228 $btag_{min}$ discriminator in the control regions. The systematic uncertainties can affect either the
1229 normalization of a process, or its shape, or both:

- 1230 • **Luminosity**: an uncertainty of 2.3% is assessed for 2017 luminosity [55].
- 1231 • **Lepton Efficiency**: muon and electron trigger, reconstruction, and identification ef-
1232 ficiencies are determined in data using the standard tag-and-probe technique with
1233 Z bosons. The corresponding scale factors are applied to the Monte Carlo samples.
1234 The systematic uncertainty is evaluated from the statistical uncertainties in the bin-
1235 by-bin efficiencies and efficiency differences using alternative samples and selection.
- 1236 • **Unclustered MET**: we follow the suggested procedure from the JetMET POG and
1237 find a 3% systematic uncertainty on the calibration of unclustered MET (ie, missing
1238 energy associated with particles not clustered into jets). No shape uncertainties are
1239 considered.
- 1240 • **MET+jets Trigger**: the parameters describing the trigger efficiency curve have been
1241 varied within their statistical uncertainties. An uncertainty of 1% is estimated, and
1242 no shape uncertainties are considered.
- 1243 • **Jet Energy Scale**: the energy scale for each jet is varied up and down within one
1244 standard deviation, individually for each source of uncertainty assessed by the JET-
1245 MET group. A decorrelation scheme in p_T and η is also implemented, as described
1246 in Section 12.1.
- 1247 • **Jet Energy Resolution**: we smear the energy resolution for each b-jet after energy
1248 regression using the JetMET approved prescription (10% uncertainty). For the shape
1249 uncertainty, the DNN is recomputed after the smearing and the modified output is
1250 used to define the shape variation.
- 1251 • **B-jet Tagging**: official b-tagging scale factors are applied consistently to jets in signal
1252 and background events. An average systematic uncertainty of 6% per b-jet, 12%
1253 per c jet, and 15% per fake tag (light quarks and gluons) are used to account for
1254 the normalization uncertainty. For the shape, we vary the b-tagging discriminator
1255 output, described in Sec. 4.6.1, by the uncertainties centrally provided by the BTV
1256 POG. The “up” and “down” variations are then used as alternative inputs to the
1257 DNN, and the resulting DNN outputs are used as the shape variations in the fit. A
1258 decorrelation scheme in p_T and η is also implemented, as described in Section 12.1.
- 1259 • **Cross section**: the total signal cross section has been calculated to next-to-next-to-
1260 leading order accuracy, and the total uncertainty is 4% [75], including the effect of
1261 scale and PDF variations.
- 1262 • **$H \rightarrow b\bar{b}$ Branching ratio**: an uncertainty of 0.5% is used [11], for 2017 luminosity.
- 1263 • **Theoretical p_T Spectrum**: this analysis is performed in the boosted regime, and

thus, potential differences in the p_T spectrum of the V and H between data and Monte Carlo generators could introduce systematic effects in the signal acceptance and efficiency estimates. Two calculations are become available that estimate the NLO electroweak [76–78] and NNLO QCD [79] corrections to VH production in the boosted regime. Both the EWK and NNLO QCD corrections have been applied to the signal MC samples. The estimated effect from NNLO electroweak corrections are 2% for ZH and 2% for WH. For the remaining QCD correction an uncertainty of 5% for both ZH and WH is estimated.

- $\Delta\phi(V,H)$: systematic uncertainty on the jet angular resolution is assumed to be negligible, and this is confirmed by the good agreement observed in the control regions.
- N_{al} : the efficiency of the lepton veto is found to be 100% in the simulation, and no additional uncertainty is assigned.
- $\Delta\eta$ reweighting: The entire $\Delta\eta$ reweighting in LO V+Jets, applied to account for an $M(jj)$ residual data/MC discrepancy as documented in Sec. 7, is assumed as shape systematic uncertainty.
- W boson and $t\bar{t}$ transverse momentum reweighting: The systematic uncertainties on the $p_T(V)$ corrections are taken from the uncertainties on the fitted slopes, corresponding to 13% for $t\bar{t}$ and 6% for both W + udscg and W+HF + single top, as documented in Sec. 7.
- **Background Estimate:** a mix of data-driven methods, simulation, and theory uncertainties contribute to the total uncertainty on the background estimates. Correlated (luminosity, b-tagging, JEC/JER, and TnP efficiencies) and uncorrelated uncertainties (statistical, control region, and cross section) are combined separately. An uncertainty of 15% is assumed for single top and dibosons (approximately the uncertainty on the measured cross section). The other backgrounds are taken directly from data, with the associated uncertainties from the control regions.
- **Monte Carlo Statistics:** the finite size of the signal and background MC samples is included in the normalization uncertainties. In addition, the shape of the DNN is allowed to vary within the bin-by-bin statistical uncertainties from the MC samples (in a coherent way), while also constraining the total integral within its uncertainty.
- **PDF uncertainties:** the imperfect knowledge of the proton quark content is encoded in a set of NNPDF MC replicas. For each process, the RMS of all the variations is checked in each bin of the DNN distribution and the largest variation is used as normalization nuisance in the datacards. For more details refer to section 12.2.
- **pQCD scale variations:** The perturbative QCD renormalization (μ_R) and factorization (μ_F) scale variations of 1/2 and 2 times the nominal values are considered separately for each process and taken as uncorrelated sources of systematic uncertainties (shape + normalization).

As noted above, for particular sources of uncertainty, DNN shapes are taken into account. The “up” and “down” variations of DNN inputs are used, and the resulting alternative DNN outputs are used as the shape variations in the fit.

In other words, these uncertainties define the 1σ envelope within which the DNN shape is allowed to vary. This variation is performed in the following way: given a nominal shape $n(x)$ and two variations for “up” and “down”, $p(x)$ and $m(x)$, the code introduces a nuisance parameter “ t ”, and the final function $f(x)$ used to describe the variable is defined as:

$$t > 0 : f(x;t) = n(x) + (p(x) - n(x)) * a(t) \quad t < 0 : f(x;t) = n(x) + (m(x) - n(x)) * a(t) \quad (10)$$

1309 Here, $a(t)$ is quadratic for $|t| < 1$; linear for $|t| \geq 1$; is everywhere continuous; and has
 1310 continuous first and second derivatives.

1311 12.1 Jet p_T , η decorrelation in JEC, JER and B-tagging systematics

1312 In the Run1 version of this analysis, the nuisance parameters associated to jet energy scale
 1313 and resolution, as well as those associated to b-tagging, were prone to over-constraining in the
 1314 signal extraction fit.

1315 After careful studies, this was attributed to the presence of a single nuisance parameter for all
 1316 jets, regardless of p_T and η , although the size of the nuisance varies as a function of p_T and η .
 1317 The net effect was to artificially constraining the pull of these nuisances in the most signal-like
 1318 regions from the high statistics background regions, where typically jets have lower p_T than the
 1319 most sensitive phase space of the signal region. In other words, the observed over-constraining
 1320 behaviour is symptomatic of a wrong assumption on the full correlation of the systematics
 1321 over the entire phase space, which instead should be partitioned in an uncorrelated way, or the
 1322 sources of uncertainties should be decomposed.

1323 Starting from HIG-16-044, and still pursued in the current version of the analysis, the jet en-
 1324 ergy scale uncertainties were decomposed into 27 independent sources as recommended by
 1325 the JET/MET POG and described in

1326 <https://twiki.cern.ch/twiki/bin/viewauth/CMS/JECUncertaintySources>

1327 A single nuisance parameter is used for jet energy resolution, and no significant constraining
 1328 is observed.

1329 The systematic uncertainties for the b-tagging, instead, were de-correlated in five p_T and three
 1330 η bins, matching the binning originally chosen by the BTV POG to compute them.

1331 12.2 Implementation of PDF Uncertainty Systematics

1332 The imperfect knowledge of the proton quark content is encoded in a set of 101 NNPDF MC
 1333 replicas. For each of these replicas, the DNN score in the signal region is re-calculated sepa-
 1334 rately for each simulation sample. The RMS and mean over the replicas is evaluated for each
 1335 DNN score bin. Table 21 shows the resulting effective log-normal uncertainty value for each
 1336 sample.

Table 21: The effective log-normal yield uncertainty per channel due to the imperfect knowl-
 edge of the proton quark content. These values are derived by taking the maximum value
 of the RMS / mean of the 101 NNPDF MC replicas over all DNN score bins for each sample
 separately.

Sample	TT	W+bb	W+b	W+udscg	WH	ZH
Effective Log-normal Yield Unc. (%)	0.5	2	3	5	1	1
Sample	VV+HF	VV+udscg	Z+bb	Z+b	Z+udcsg	
Effective Log-normal Yield Unc. (%)	2	3	2	3	5	

1337 This method approximates that the effect of each individual replica is a flat yield scaling. This
 1338 approximation was also made in the Run-1 VH(bb) analysis. However, it is now apparent that
 1339 for some of the MC replicas there is a clear shape dependence of the variation with respect
 1340 to the nominal DNN score. The log-normal implementation of this systematic is therefore not
 1341 entirely correct, and needs to be revised. The recommended prescription is to consider an inde-
 1342 pendent shape nuisance for each of the Hessian eigenvariations. Unfortunately, in the currently

1343 available production of ntuples only the replicas are available. An overly conservative imple-
1344 mentation of each of the 101 replicas as independent shape nuisances resulted in approximately
1345 a five percent reduction in the analysis expected sensitivity. Therefore it is not expected that
1346 the final implementation of this systematic will have a large effect on the analysis sensitivity.

DRAFT

13 Results

The signal extraction is performed with a simultaneous fit of the background enriched control regions (CRs), described in Sec. 11, and the signal region (SR), described in Sec. 8.3. In the CR the fitted distribution is the output of the b-tagging discriminator, described in Sec. 4.6, while in the signal region we use the output of the multivariate discriminator described in Sec. 9. The 2-lepton channel is split into two $p_T(V)$ categories (low- and high- $p_T(V)$), while the 0- and 1-lepton channels are inclusive in $p_T(V)$.

In the 2017 analysis we use the multivariate discriminator from the DNN output for the three channels.

This section summarizes the results obtained from the fits in 2017 analysis for the VH process, and their combination with both the 2016 analysis included in HIG-16-044 and Run1.

The results obtained for the 2017 VZ cross-check analysis are contained in Appendix B.

13.1 Signal and Control Regions fits

The unblinded distributions resulting from the combined fit of all channels are shown in Figure 52 for the multivariate discriminator output in the signal regions, and in Figure 51 for the multi-background DNN distributions in the control regions. All scale factors from Sec. 11.5 and post-fit nuisances values have been applied. In particular, the signal cross section is rescaled by the observed combined signal strength.

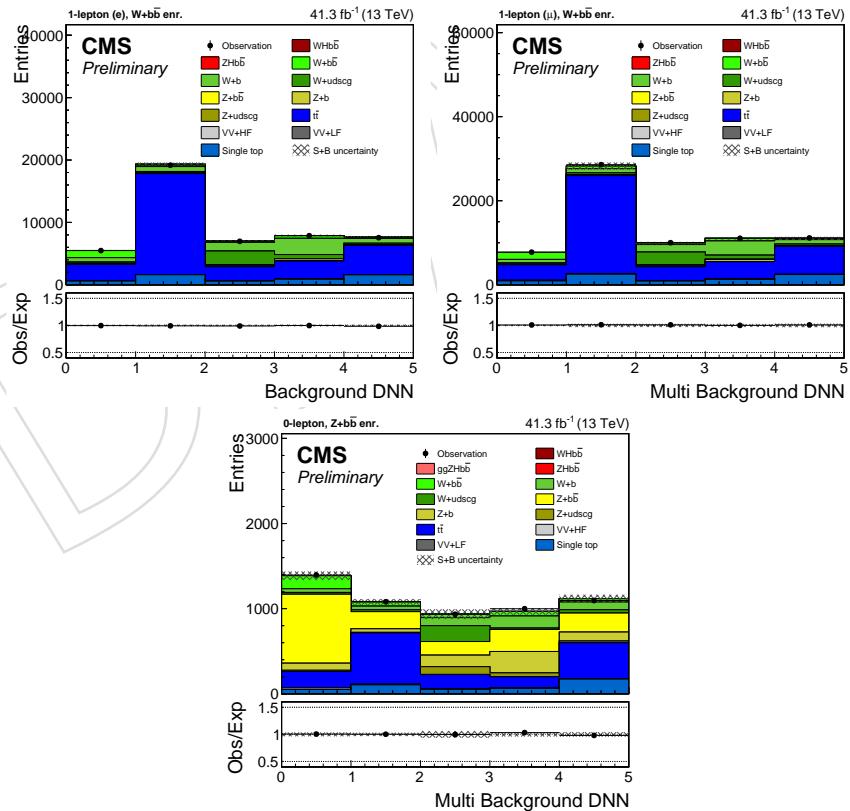


Figure 51: Post-fit distributions of the Multi-background DNN fit variable for 2017 analysis in the 1-lepton channel (top row) for muon (left) and electron (right) control regions, and for the 0-lepton channel (bottom row).

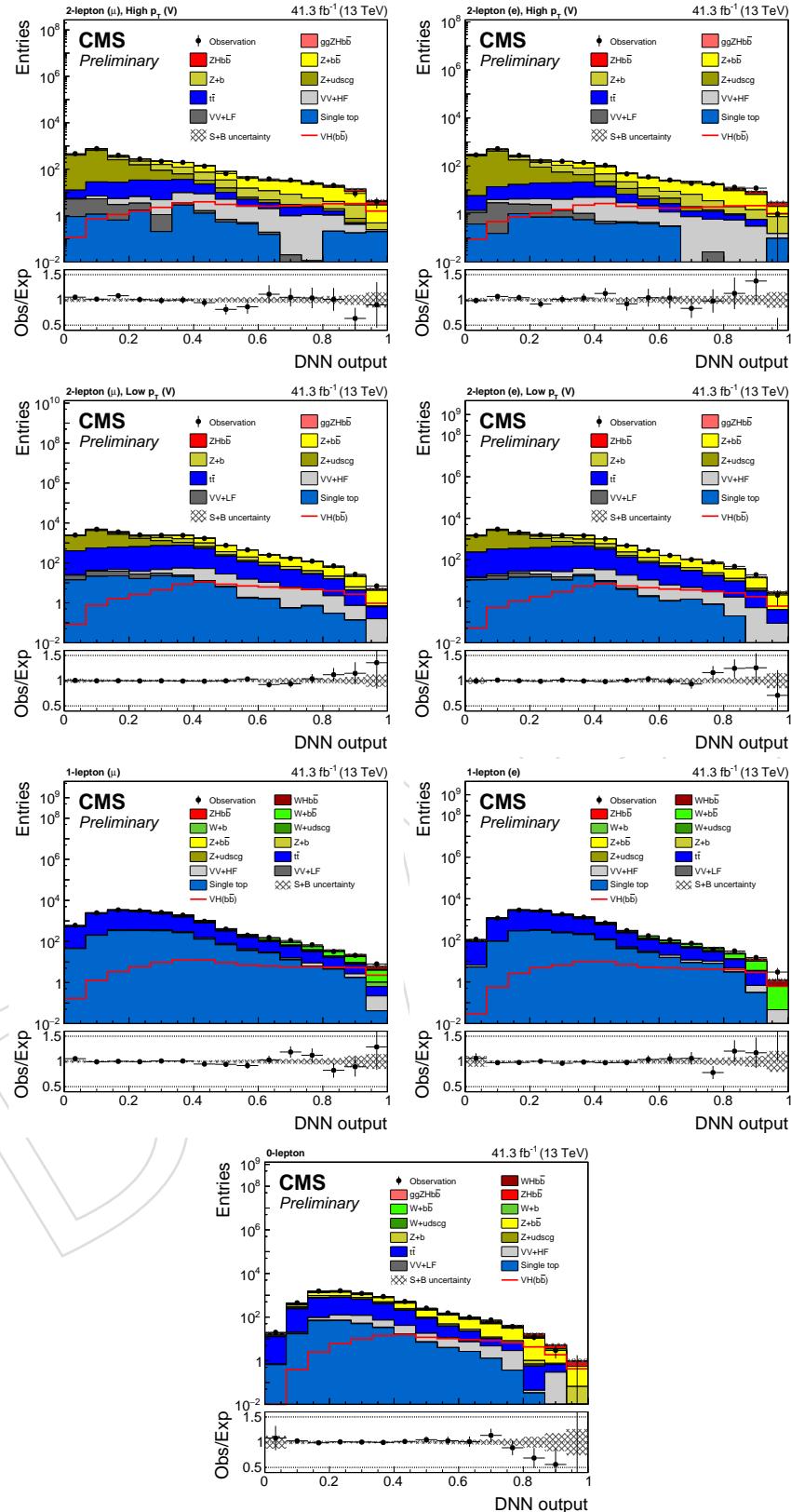


Figure 52: Post-fit distributions (VH signal is post-fit) of multivariate discriminator output channels for 2017 analysis, after all signal region pre-selection criteria have been applied. First row: 2-lepton muon (left) and electron (right) channel for high $p_T(V)$ region, in the second row the low $p_T(V)$ is shown. Third row: 1-lepton muon (left) and electron (right) channel. Fourth row: 0-lepton channel. 0-lepton (bottom)

1365 13.2 Signal extraction calculation

1366 The maximum likelihood estimates or best-fit-values of μ and θ are denoted $\hat{\mu}$ and $\hat{\theta}$, while $\hat{\theta}_{\mu}$
1367 denotes the conditional maximum likelihood estimate of all nuisance parameters with μ fixed.
1368 In this analysis the range of μ is restricted to the physically meaningful regime, *i.e.* it is not
1369 allowed to be negative.

1370 Following the fully frequentist methodology, Monte Carlo pseudo-experiments that include
1371 pseudo-data and $\hat{\theta}$ values of the nuisance parameters are generated to construct the pdfs.

1372 The results are obtained from the standard shape analysis implementation within RooStats,
1373 where the DNN shapes are allowed to vary within the envelope defined by the nuisance vari-
1374 ations.

1375 13.3 Results VH

1376 The exact command used to get the signal strength and significance is:

1377 Significance: `combineTool.py -M Significance --cminDefaultMinimizerStrategy`
1378 `1 --cminPreFit=1 --significance -d ws.root`

1379 The impact of the most significant nuisances in the signal extraction fit is shown in Fig. 60.

1380 A comparison of the pre- and post-fit nuisances pulls and constraints is available in Fig. 54 and
1381 Fig. 55. As it can be seen, very few nuisances move by an appreciable amount or are slightly
1382 over-constrained, as highlighted by those in red color.

DRAFT

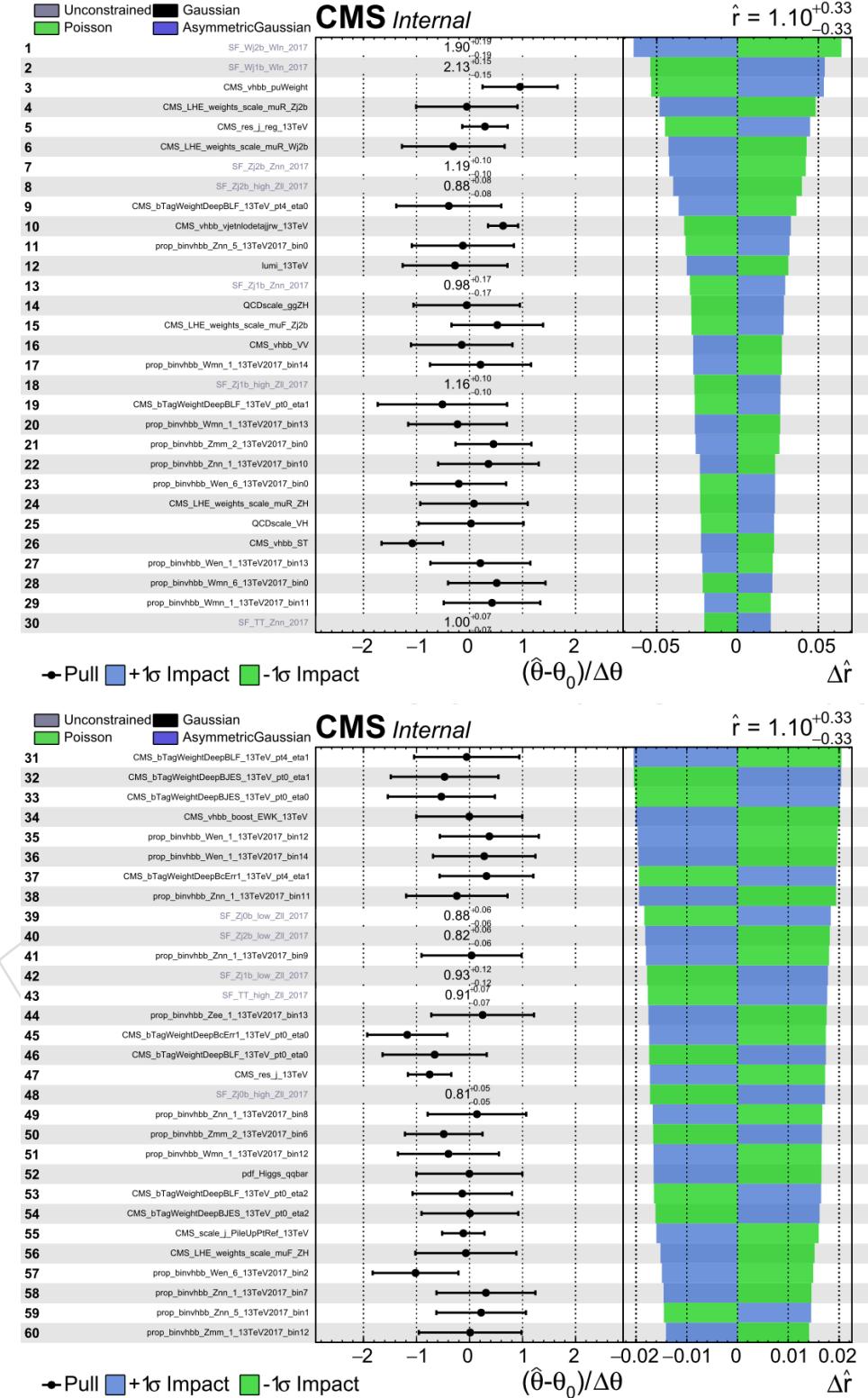


Figure 53: Impact plot of the most significant nuisances in the signal extraction fit.

Comparison of nuisances

Figure 54: Impact plot of the most significant nuisances in the signal extraction fit. As it can be seen, very few nuisances move by an appreciable amount or are slightly over-constrained, as highlighted by those in red color.

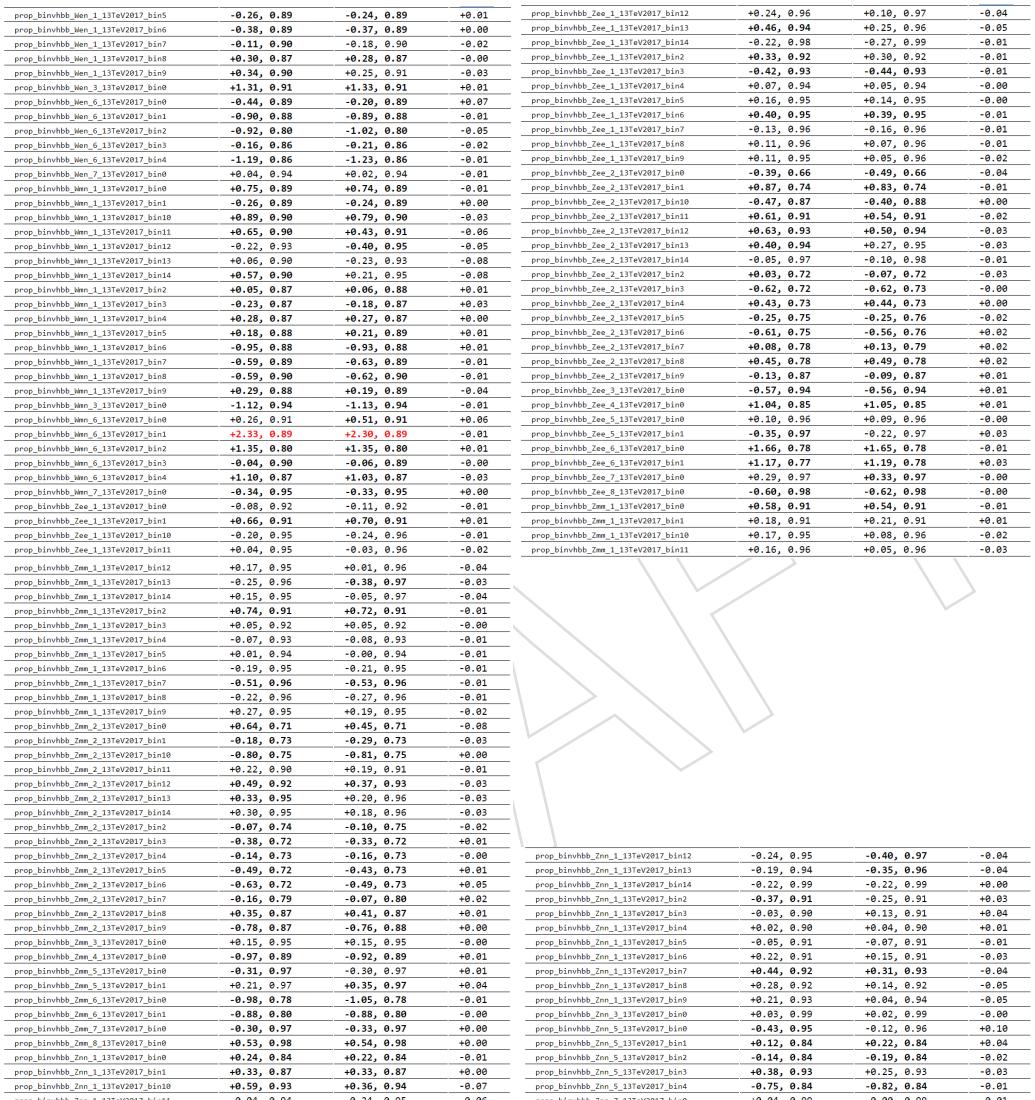


Figure 55: Impact plot of the most significant nuisances in the signal extraction fit. As it can be seen, very few nuisances move by an appreciable amount or are slightly over-constrained, as highlighted by those in red color.

¹³⁸³ The observed (expected post-fit) local significance of the excess over background prediction is
¹³⁸⁴ quantified as 3.5σ (3.2σ).

¹³⁸⁵ When the 2017 analysis is combined with the 2016 analysis published in HIG-16-044 [30], and
¹³⁸⁶ the legacy Run1 analysis, the total significance becomes 4.9σ (4.9σ).

¹³⁸⁷ The results are summarized in Table 22.

Table 22: Expected significance for the VH production cross section times $H \rightarrow b\bar{b}$ branching ratio, for the combination and sub-channels separately.

Channel	0-lepton	1-lepton	2-lepton	combo
2016 analysis (HIG-16-044)				
Expected syst, no SFs (pre-fit)	1.90	1.55	1.92	3.07
Expected post-fit CR-only syst, CR-only fitted SFs	1.37	1.55	1.73	2.79
Expected post-fit SR+CR syst, SR+CR fitted SFs	1.50	1.50	1.80	2.80
Observed post-fit SR+CR syst, SR+CR fitted SFs	0.00	3.20	3.10	3.30
2017 analysis (HIG-18-016)				
Expected syst, no SFs (pre-fit)	1.94	2.08	1.81	3.38
Expected post-fit CR-only syst, CR-only fitted SFs	1.79	1.78	1.86	3.16
Expected post-fit SR+CR syst, SR+CR fitted SFs	1.85	1.79	1.93	3.22
Observed post-fit SR+CR syst, SR+CR fitted SFs	0.89	1.98	2.42	3.54
2016 (HIG-16-044) + 2017 analysis (HIG-18-016) (*)				
Expected syst, no SFs (pre-fit)	—	—	—	4.65
Expected post-fit CR-only syst, CR-only fitted SFs	—	—	—	4.25
Expected post-fit SR+CR syst, SR+CR fitted SFs	—	—	—	4.26
Observed post-fit SR+CR syst, SR+CR fitted SFs	—	—	—	4.44
Run1+2016+2017 analysis (HIG-18-016) (*)				
Expected syst, no SFs (pre-fit)	—	—	—	5.22
Expected post-fit CR-only syst, CR-only fitted SFs	—	—	—	4.86
Expected post-fit SR+CR syst, SR+CR fitted SFs	—	—	—	4.95
Observed post-fit SR+CR syst, SR+CR fitted SFs	—	—	—	4.92

¹³⁸⁸ The corresponding fitted signal strength is found to be $\mu = \sigma/\sigma_{SM} = 1.10^{+0.34}_{-0.33}$, as reported in
¹³⁸⁹ Table 23.

Table 23: Signal strength for the VH production cross section times $H \rightarrow b\bar{b}$ branching ratio, for the combination and sub-channels separately.

Channel	0-lepton	1-lepton	2-lepton	combo
2017 analysis (HIG-18-016) individual fits				
Observed post-fit SR+CR syst, SR+CR fitted SFs	$0.50^{+0.59}_{-0.50}$	$1.02^{+0.57}_{-0.53}$	$1.32^{+0.60}_{-0.56}$	—
2017 analysis (HIG-18-016) combined fit				
Observed post-fit SR+CR syst, SR+CR fitted SFs	0.7 ± 0.7	1.3 ± 0.6	1.2 ± 0.6	$1.10^{+0.34}_{-0.33}$
2016 (HIG-16-044) +2017 analysis (HIG-18-016) (*) combined fit				
Observed post-fit SR+CR syst, SR+CR fitted SFs	—	—	—	$1.05^{+0.25}_{-0.25}$
Run1+2016+2017 analysis (HIG-18-016) (*) combined fit				
Observed post-fit SR+CR syst, SR+CR fitted SFs	—	—	—	$1.02^{+0.22}_{-0.22}$

¹³⁹⁰ The 2017 channel compatibility and signal strength is shown in Fig. 56. At 125 GeV the signal
¹³⁹¹ strength for all modes is compatible with a p-value of 96%.

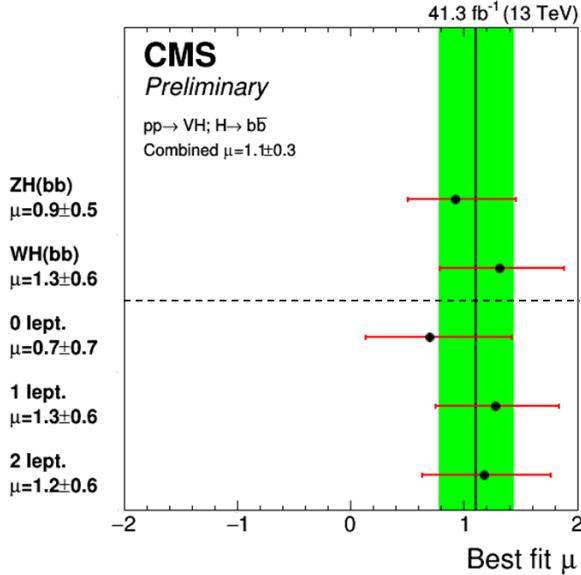


Figure 56: The most likely signal strength and uncertainty fitted in the three channels for 2017 analysis, together with the strength for WH and ZH.

¹³⁹² Figure 57 combines the DNN outputs of all channels for the 2017 analysis, where the events are
¹³⁹³ gathered in bins of similar expected signal-to-background ratio, as given by the value of the
¹³⁹⁴ output of their corresponding DNN discriminant.

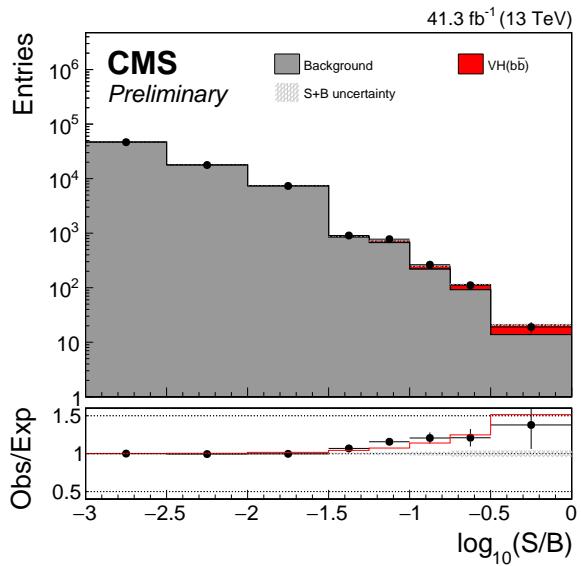


Figure 57: Combination of all channels into a single distribution. Events are sorted in bins of similar expected signal-to-background ratio, as given by the value of the output of their corresponding BDT discriminant (trained with a Higgs boson mass hypothesis of 125 GeV). The bottom inset show the ratio of the data to the predicted sum of background and SM Higgs boson signal with a mass of 125 GeV. All MC templates are post-fit except for the signal template, which is pre-fit.

1395 14 Conclusions

1396 This note reports the search for the standard model Higgs production in association with vector
1397 bosons and decaying into $b\bar{b}$ pairs. A data sample of approximately 41.2 fb^{-1} corresponding to
1398 the 2017 running period has been analyzed in five modes ($Z(\mu\mu)H$, $Z(ee)H$, $Z(\nu\nu)H$, $W(\mu\nu)H$,
1399 $W(e\nu)H$) and a signal is extracted. The fitted signal strength at 125GeV is $\mu = \sigma/\sigma_{\text{SM}} =$
1400 $1.10^{+0.34}_{-0.33}$, with an observed (expected) significance of 3.5 (3.2) standard deviations.

DRAFT

1401 References

- 1402 [1] A. Salam and J. C. Ward, "On a gauge theory of elementary interactions", *Nuovo Cim.* **19**
1403 (1961) 165–170, doi:10.1007/BF02812723.
- 1404 [2] S. Glashow, "Partial symmetries of weak interactions", *Nucl.Phys.* **22** (1961) 579–588,
1405 doi:10.1016/0029-5582(61)90469-2.
- 1406 [3] S. Weinberg, "A model of leptons", *Phys.Rev.Lett.* **19** (1967) 1264–1266,
1407 doi:10.1103/PhysRevLett.19.1264.
- 1408 [4] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons", *Phys.*
1409 *Rev. Lett.* **13** (1964) 321–323, doi:10.1103/PhysRevLett.13.321.
- 1410 [5] P. W. Higgs, "Broken symmetries and the masses of gauge bosons", *Phys. Rev. Lett.* **13**
1411 (1964) 508–509, doi:10.1103/PhysRevLett.13.508.
- 1412 [6] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, "Global conservation laws and massless
1413 particleless", *Phys. Rev. Lett.* **13** (1964) 585–587, doi:10.1103/PhysRevLett.13.585.
- 1414 [7] CMS Collaboration, "Observation of a new boson at a mass of 125 gev with the cms
1415 experiment at the lhc", *Phys.Lett.* **B716** (2012) 30–61,
1416 doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235.
- 1417 [8] ATLAS Collaboration, "Observation of a new particle in the search for the standard
1418 model higgs boson with the atlas detector at the lhc", *Phys.Lett.* **B716** (2012) 1–29,
1419 doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214.
- 1420 [9] CMS Collaboration, "Measurements of properties of the Higgs boson decaying into the
1421 four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV", *JHEP* **11** (2017) 047,
1422 doi:10.1007/JHEP11(2017)047, arXiv:1706.09936.
- 1423 [10] CMS Collaboration, "Combined measurements of the Higgs boson's couplings at
1424 $\sqrt{s} = 13$ TeV", Technical Report CMS-PAS-HIG-17-031, CERN, Geneva, 2018.
- 1425 [11] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs
1426 Cross Sections: 4. Deciphering the Nature of the Higgs Sector", arXiv:1610.07922.
- 1427 [12] LHC Higgs Cross Section Working Group et al., "Handbook of lhc higgs cross sections: 3.
1428 higgs properties", CERN-2013-004 (CERN, Geneva, 2013) arXiv:1307.1347.
- 1429 [13] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, "Jet substructure as a new
1430 Higgs search channel at the LHC", *Phys. Rev. Lett.* **100** (2008) 242001,
1431 doi:10.1103/PhysRevLett.100.242001, arXiv:0802.2470.
- 1432 [14] CMS Collaboration, "Search for the standard model Higgs boson decaying to bottom
1433 quarks and produced in association with a W or a Z boson", *CMS Physics Analysis*
1434 *Summary CMS-PAS-HIG-11-012* (2011).
- 1435 [15] VHbb Team, "Search for the standard model higgs boson produced in association with a
1436 w or z and decaying to bottom quarks", *CMS Note* **2011/240** (2011).
- 1437 [16] CMS Collaboration, "Search for the standard model Higgs boson decaying to bottom
1438 quarks", *CMS Physics Analysis Summary CMS-PAS-HIG-11-031* (2011).

- [1439] [17] VHbb Team, "Search for the standard model higgs boson produced in association with a
[1440] w or z and decaying to bottom quarks", *CMS Note* **2011/430** (2011).
- [1441] [18] CMS Collaboration, "Search for the standard model Higgs boson decaying to bottom
[1442] quarks in pp collisions at $\sqrt{s} = 7$ TeV", *Phys. Lett.* **B710** (2012) 284–306,
[1443] doi:10.1016/j.physletb.2012.02.085, arXiv:1202.4195.
- [1444] [19] CMS Collaboration, "Search for the standard model Higgs boson produced in association
[1445] with W or Z bosons, and decaying to bottom quarks (ICHEP 2012)", *CMS Physics
[1446] Analysis Summary CMS-PAS-HIG-12-019* (2012).
- [1447] [20] Hbb subgroup, "Search for the standard model higgs boson decaying to bottom quarks
[1448] and produced in association with a w or a z boson", *CMS Physics Analysis Note
[1449] AN-12-181* (2011).
- [1450] [21] Hbb subgroup, "Search for the standard model higgs boson decaying to bottom quarks
[1451] and produced in association with a w or a z boson (hcp)", *CMS Physics Analysis Note
[1452] AN-12-349* (2012).
- [1453] [22] CMS Collaboration, "Search for the standard model Higgs boson produced in association
[1454] with W or Z bosons, and decaying to bottom quarks", *CMS Physics Analysis Summary
[1455] CMS-PAS-HIG-12-044* (2012).
- [1456] [23] Hbb subgroup, "Search for the standard model higgs boson decaying to bottom quarks
[1457] and produced in association with a w or a z boson (lhcp 2013)", *CMS Physics Analysis
[1458] Note AN-13-069* (2013).
- [1459] [24] CMS Collaboration, "Search for the standard model Higgs boson produced in association
[1460] with W or Z bosons, and decaying to bottom quarks (LHCP 2013)", *CMS Physics Analysis
[1461] Summary CMS-PAS-HIG-13-012* (2013).
- [1462] [25] CMS Collaboration, "Search for the standard model Higgs boson produced in association
[1463] with a W or a Z boson and decaying to bottom quarks", *Phys. Rev.* **D89** (2014), no. 1,
[1464] 012003, doi:10.1103/PhysRevD.89.012003, arXiv:1310.3687.
- [1465] [26] CMS Collaboration, "Precise determination of the mass of the Higgs boson and tests of
[1466] compatibility of its couplings with the standard model predictions using proton
[1467] collisions at 7 and 8 TeV", *Eur. Phys. J.* **C75** (2015), no. 5, 212,
[1468] doi:10.1140/epjc/s10052-015-3351-7, arXiv:1412.8662.
- [1469] [27] CMS Collaboration, "Search for the standard model Higgs boson produced through
[1470] vector boson fusion and decaying to $b\bar{b}$ ", *Phys. Rev.* **D92** (2015), no. 3, 032008,
[1471] doi:10.1103/PhysRevD.92.032008, arXiv:1506.01010.
- [1472] [28] CMS Collaboration, "Search for the standard model Higgs boson produced in association
[1473] with a top-quark pair in pp collisions at the LHC", *JHEP* **05** (2013) 145,
[1474] doi:10.1007/JHEP05(2013)145, arXiv:1303.0763.
- [1475] [29] CMS Collaboration, "Search for the associated production of the Higgs boson with a
[1476] top-quark pair", *JHEP* **09** (2014) 087,
[1477] doi:10.1007/JHEP09(2014)087, 10.1007/JHEP10(2014)106,
[1478] arXiv:1408.1682. [Erratum: JHEP10,106(2014)].

- [30] CMS Collaboration, "Evidence for the higgs boson decay to a bottom quark,antiquark pair", *Physics Letters B* **780** (2018) 501 – 532,
doi:<https://doi.org/10.1016/j.physletb.2018.02.050>.
- [31] ATLAS Collaboration, "Evidence for the $H \rightarrow b\bar{b}$ decay with the ATLAS detector", *Journal of High Energy Physics* **12** (2017) 24, doi:[10.1007/JHEP12\(2017\)024](https://doi.org/10.1007/JHEP12(2017)024), arXiv:[1708.03299](https://arxiv.org/abs/1708.03299).
- [32] Hbb subgroup, "Search for the standard model higgs boson produced in association with w and z and decaying to bottom quarks", *CMS Physics Analysis Note AN-15-168* (2017).
- [33] CMS Collaboration, "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", *JINST* **13** (2018), no. 05, P05011, doi:[10.1088/1748-0221/13/05/P05011](https://doi.org/10.1088/1748-0221/13/05/P05011), arXiv:[1712.07158](https://arxiv.org/abs/1712.07158).
- [34] H. Voss, A. Höcker, J. Stelzer, and F. Tegenfeldt, "TMVA, the toolkit for multivariate data analysis with ROOT", in *XIth International Workshop on Advanced Computing and Analysis Techniques in Physics Research (ACAT)*, p. 40. 2007. arXiv:[physics/0703039](https://arxiv.org/abs/physics/0703039).
- [35] T. Sjöstrand, S. Mrenna, and P. Skands, "Pythia 6.4 physics and manual", *JHEP* **05** (2006) 026, doi:[10.1088/1126-6708/2006/05/026](https://doi.org/10.1088/1126-6708/2006/05/026), arXiv:[hep-ph/0603175](https://arxiv.org/abs/hep-ph/0603175).
- [36] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1", *Comput. Phys. Commun.* **178** (2008) 852–867, doi:[10.1016/j.cpc.2008.01.036](https://doi.org/10.1016/j.cpc.2008.01.036), arXiv:[0710.3820](https://arxiv.org/abs/0710.3820).
- [37] P. Nason, "A new method for combining nlo qcd with shower monte carlo algorithms", *JHEP* **0411** (2004) 040, doi:[10.1088/1126-6708/2004/11/040](https://doi.org/10.1088/1126-6708/2004/11/040), arXiv:[hep-ph/0409146](https://arxiv.org/abs/hep-ph/0409146).
- [38] J. Alwall et al., "Madgraph 5 : Going beyond", *JHEP* **1106** (2011) 128, doi:[10.1007/JHEP06\(2011\)128](https://doi.org/10.1007/JHEP06(2011)128), arXiv:[1106.0522](https://arxiv.org/abs/1106.0522).
- [39] 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", *JHEP* **07** (2014) 079, doi:[10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079), arXiv:[1405.0301](https://arxiv.org/abs/1405.0301).
- [40] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, "Matching matrix elements and shower evolution for top-quark production in hadronic collisions", *JHEP* **01** (2007) 013, doi:[10.1088/1126-6708/2007/01/013](https://doi.org/10.1088/1126-6708/2007/01/013), arXiv:[hep-ph/0611129](https://arxiv.org/abs/hep-ph/0611129).
- [41] R. Frederix and S. Frixione, "Merging meets matching in MC@NLO", *JHEP* **12** (2012) 061, doi:[10.1007/JHEP12\(2012\)061](https://doi.org/10.1007/JHEP12(2012)061), arXiv:[1209.6215](https://arxiv.org/abs/1209.6215).
- [42] CMS Collaboration, "CMS event generator tuning with 13 TeV LHC data", Technical Report CMS-PAS-GEN-17-001, CERN, Geneva, 2018.
- [43] NNPDF Collaboration, "Parton distributions for the LHC Run II", *JHEP* **04** (2015) 040, doi:[10.1007/JHEP04\(2015\)040](https://doi.org/10.1007/JHEP04(2015)040), arXiv:[1410.8849](https://arxiv.org/abs/1410.8849).
- [44] LHC Higgs Cross Section Working Group et al., "Handbook of lhc higgs cross sections: 1. inclusive observables", CERN-2011-002 (CERN, Geneva, 2011) arXiv:[1101.0593](https://arxiv.org/abs/1101.0593).
- [45] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order", *Comput. Phys. Commun.* **182** (2011) 2388, doi:[10.1016/j.cpc.2011.06.008](https://doi.org/10.1016/j.cpc.2011.06.008), arXiv:[1011.3540](https://arxiv.org/abs/1011.3540).

- [46] Y. Li and F. Petriello, “Combining QCD and electroweak corrections to dilepton production in the framework of the FEWZ simulation code”, *Phys. Rev. D* **86** (2012) 094034, doi:10.1103/PhysRevD.86.094034, arXiv:1208.5967.
- [47] R. Gavin, Y. Li, F. Petriello, and S. Quackenbush, “W physics at the LHC with FEWZ 2.1”, *Comput. Phys. Commun.* **184** (2013) 208, doi:10.1016/j.cpc.2012.09.005, arXiv:1201.5896.
- [48] M. Czakon, P. Fiedler, and A. Mitov, “Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\alpha_S^4)$ ”, *Phys. Rev. Lett.* **110** (2013) 252004, doi:10.1103/PhysRevLett.110.252004, arXiv:1303.6254.
- [49] N. Kidonakis, “Differential and total cross sections for top pair and single top production”, in *Proceedings, 20th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2012): Bonn, Germany, March 26-30, 2012*, pp. 831–834. 2012. arXiv:1205.3453. [,831(2012)]. doi:10.3204/DESY-PROC-2012-02/251.
- [50] CMS Collaboration, “Measurements of the $pp \rightarrow WZ$ inclusive and differential production cross section and constraints on charged anomalous triple gauge couplings at $\sqrt{s} = 13$ TeV.”, Technical Report CMS-PAS-SMP-18-002, CERN, Geneva, 2018.
- [51] CMS Collaboration, “Measurement of the WW cross section pp collisions at $\text{sqrt}(s)=13$ TeV”, Technical Report CMS-PAS-SMP-16-006, CERN, Geneva, 2016.
- [52] CMS Collaboration, “Measurement of the ZZ production cross section and $Z \rightarrow \ell^+\ell^-\ell'^+\ell'^-$ branching fraction in pp collisions at $\sqrt{s}=13$ TeV”, *Phys. Lett. B* **763** (2016) 280–303, doi:10.1016/j.physletb.2017.09.030, 10.1016/j.physletb.2016.10.054, arXiv:1607.08834. [Erratum: *Phys. Lett.B*772,884(2017)].
- [53] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014), no. 10, P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.
- [54] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet User Manual”, *Eur. Phys. J. C* **72** (2012) 1896, doi:10.1140/epjc/s10052-012-1896-2, arXiv:1111.6097.
- [55] CMS Collaboration, “CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV”, Technical Report CMS-PAS-LUM-17-004, CERN, 2018.
- [56] CMS Collaboration, “Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015), no. 06, P06005, doi:10.1088/1748-0221/10/06/P06005, arXiv:1502.02701.
- [57] CMS Electron/gamma Physics Object Group, “recipes for Run 2 ID criteria”. <https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2>.
- [58] CMS Collaboration, “Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV”.
- [59] CMS Collaboration, “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV”, *JINST* **7** (2012) P10002, doi:10.1088/1748-0221/7/10/P10002, arXiv:1206.4071.

- 1560 [60] CMS Collaboration, “Particle-flow reconstruction and global event description with the
 1561 CMS detector”, *JINST* **12** (2017), no. 10, P10003,
 1562 doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.
- 1563 [61] CMS Collaboration, “Particle-flow event reconstruction in CMS and performance for jets,
 1564 taus, and E_T^{miss} ”, CMS Physics Analysis Summary CMS-PAS-PFT-09-001, CERN, 2009.
- 1565 [62] CMS Collaboration, “Commissioning of the particle-flow event reconstruction with the
 1566 first lhc collisions recorded in the cms detector”, CMS Physics Analysis Summary
 1567 CMS-PAS-PFT-10-001, CERN, 2010.
- 1568 [63] M. Cacciari and G. P. Salam, “Dispelling the N^3 myth for the k_t jet-finder”, *Phys. Lett. B*
 1569 **641** (2006) 57, doi:10.1016/j.physletb.2006.08.037, arXiv:hep-ph/0512210.
- 1570 [64] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm”, *JHEP* **04**
 1571 (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- 1572 [65] CMS Collaboration, “Determination of Jet Energy Calibration and Transverse
 1573 Momentum Resolution in CMS”, *JINST* **6** (2011) P11002,
 1574 doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.
- 1575 [66] CMS Collaboration, “Pileup jet identification”, CMS Physics Analysis Summary
 1576 CMS-PAS-JME-13-005, CERN, 2013.
- 1577 [67] CMS BTV POG, “Event reweighting using scale factors calculated with a tag and probe
 1578 method”.
<https://twiki.cern.ch/twiki/bin/view/CMS/BTagShapeCalibration>.
- 1580 [68] CMS Collaboration, “Performance of the CMS missing transverse momentum
 1581 reconstruction in pp data at $\sqrt{s} = 8$ TeV”, *JINST* **10** (2015), no. 02, P02006,
 1582 doi:10.1088/1748-0221/10/02/P02006, arXiv:1411.0511.
- 1583 [69] CMS Collaboration, “Commissioning of trackjets in pp collisions at $\sqrt{s}=7$ TeV”, CMS
 1584 Physics Analysis Summary CMS-PAS-JME-10-006, CERN, 2010.
- 1585 [70] CMS Collaboration, “Performance of jet reconstruction with charged tracks only”, CMS
 1586 Physics Analysis Summary CMS-PAS-JME-08-001, CERN, 2009.
- 1587 [71] S. Kallweit et al., “NLO QCD+EW predictions for V + jets including off-shell
 1588 vector-boson decays and multijet merging”, *JHEP* **04** (2016) 021,
 1589 doi:10.1007/JHEP04(2016)021, arXiv:1511.08692.
- 1590 [72] CDF, D0 Collaboration, “Improved b -jet Energy Correction for $H \rightarrow b\bar{b}$ Searches at CDF”,
 1591 arXiv:1107.3026.
- 1592 [73] N. Chernavskaya, et al, “Deep neural network based simultaneous b-jet energy
 1593 correction and resolution estimator”, *CMS Note* **2018/092** (2018).
- 1594 [74] H. Ritter, E. v. Toerne, and N. Wermes, “Improvement of the Higgs mass resolution in the
 1595 $ZH \rightarrow l^+l^-bb$ channel using a Kinematic Fit with constraints”, 2014. Presented on Dec
 1596 2014. <https://cds.cern.ch/record/2318401>.
- 1597 [75] LHC Higgs Cross Section Working Group et al., “Handbook of lhc higgs cross sections:
 1598 Inclusive observables”, arXiv:1101.0593.

- 1599 [76] M. Ciccolini, A. Denner, and S. Dittmaier, “Strong and electroweak corrections to the
1600 production of Higgs+2jets via weak interactions at the LHC”, *Phys. Rev. Lett.* **99** (2007)
1601 161803, doi:10.1103/PhysRevLett.99.161803, arXiv:0707.0381.
- 1602 [77] M. Ciccolini, A. Denner, and S. Dittmaier, “Electroweak and QCD corrections to Higgs
1603 production via vector-boson fusion at the LHC”, *Phys. Rev.* **D77** (2008) 013002,
1604 doi:10.1103/PhysRevD.77.013002, arXiv:0710.4749.
- 1605 [78] Stefan Dittmaier, talk at EPS 2011.
- 1606 [79] G. Ferrera, M. Grazzini, and F. Tramontano, “Associated WH production at hadron
1607 colliders: a fully exclusive QCD calculation at NNLO”, arXiv:1107.1164. * Temporary
1608 entry *.

DRAFT

1609 A BDT distributions for VH discrimination

1610 The The Gradient Boost (BDTG) BDT distributions obtained from the signal and background
 1611 events in the training and testing sets are shown in Fig. 58 for 2017 analysis, for the 0-, 1-
 1612 and 2-lepton categories. Equivalent plots are obtained for the 2016 analysis. No particular
 1613 overtraining is observed.

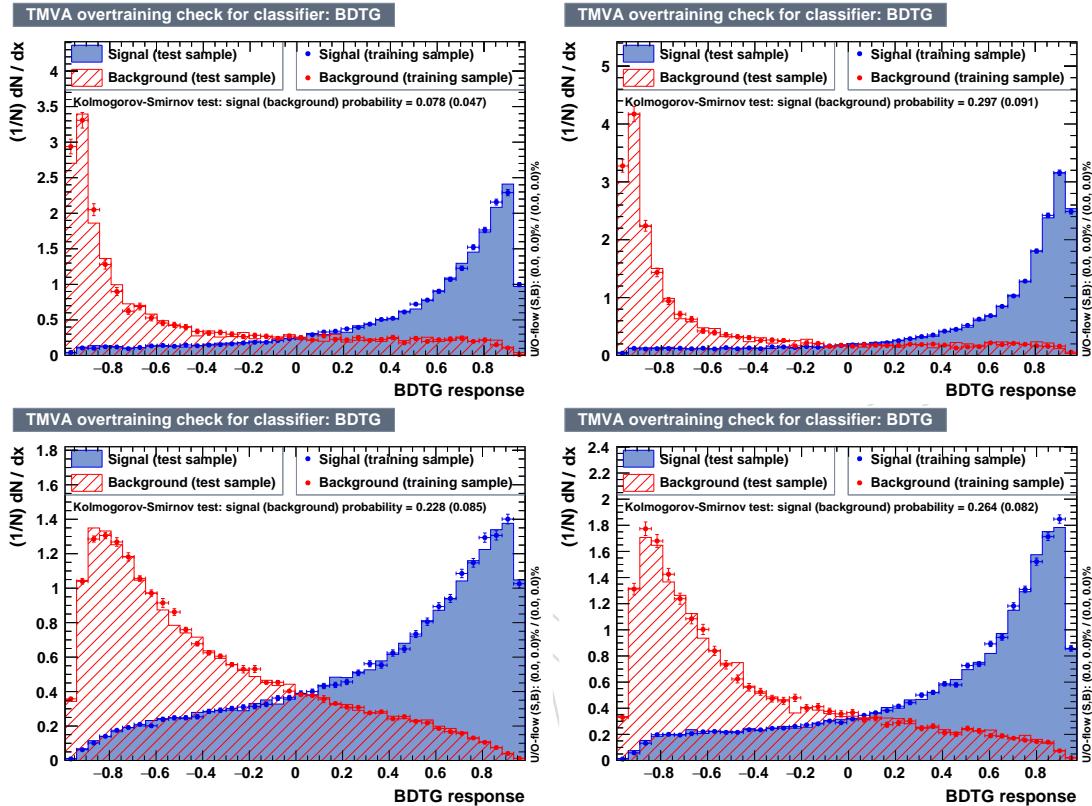


Figure 58: BDT output distributions for signal and background in the training and testing sets for 2017 analysis. In the top row, plots for the 2-lepton category in the low (left) and high (right) $p_T(V)$ regions. In the bottom row, plots for the 1-lepton (left) and 0-lepton (right) categories.

1614 B Results of the VZ(bb) cross-check analysis

1615 The VZ(bb) cross-check analysis is performed using all the objects, corrections and analysis
 1616 techniques used for VH, in order to validate it. The CR and SR definitions are very close to
 1617 those listed in Tables 17, 18, 19 (CR), and 14 (SR). The only difference lies in the SR invariant
 1618 mass boundaries for the 1- and 2-lepton channels, lowered to 60 GeV (instead of 90 GeV), and
 1619 corresponding change to CR to keep phase space exclusive. This change is not needed for the 0-
 1620 lepton channel which has already the boundary at 60 GeV for VH analysis. Finally, in the final
 1621 DNN discriminator described in Sec. 9.4 the role of VZ and ZH has been swapped, targeting
 1622 VZ as signal and including VH as background.

1623 B.1 Signal and Control Regions fits

1624 The unblinded distributions resulting from the combined fit of all channels are shown in Fig-
 1625 ure 59 for the multivariate discriminator output in the signal regions.

Table 24: Data/MC scale factors for the 2017 analysis in the 0-, 1- and 2-lepton channels from VZ SR+CRs fit. The errors include both statistical and systematic uncertainties. Compatible values are obtained from the VH fit.

Process	Z($\nu\nu$)H	W($\ell\nu$)H	Z($\ell\ell$)H low- p_T	Z($\ell\ell$)H high- p_T
W + udscg	1.04 ± 0.01	1.04 ± 0.01	–	–
W + b	2.02 ± 0.09	2.02 ± 0.09	–	–
W + $b\bar{b}$	2.02 ± 0.13	2.02 ± 0.13	–	–
Z + udscg	0.86 ± 0.05	–	0.88 ± 0.01	0.80 ± 0.01
Z + b	1.07 ± 0.14	–	0.89 ± 0.05	1.13 ± 0.08
Z + $b\bar{b}$	1.20 ± 0.07	–	0.84 ± 0.03	0.95 ± 0.05
t \bar{t}	0.97 ± 0.02	0.93 ± 0.01	0.88 ± 0.01	0.90 ± 0.02

1626 All scale factors from Table 24 and post-fit nuisances values have been applied. In particular,
 1627 the signal cross section is rescaled by the observed combined signal strength.

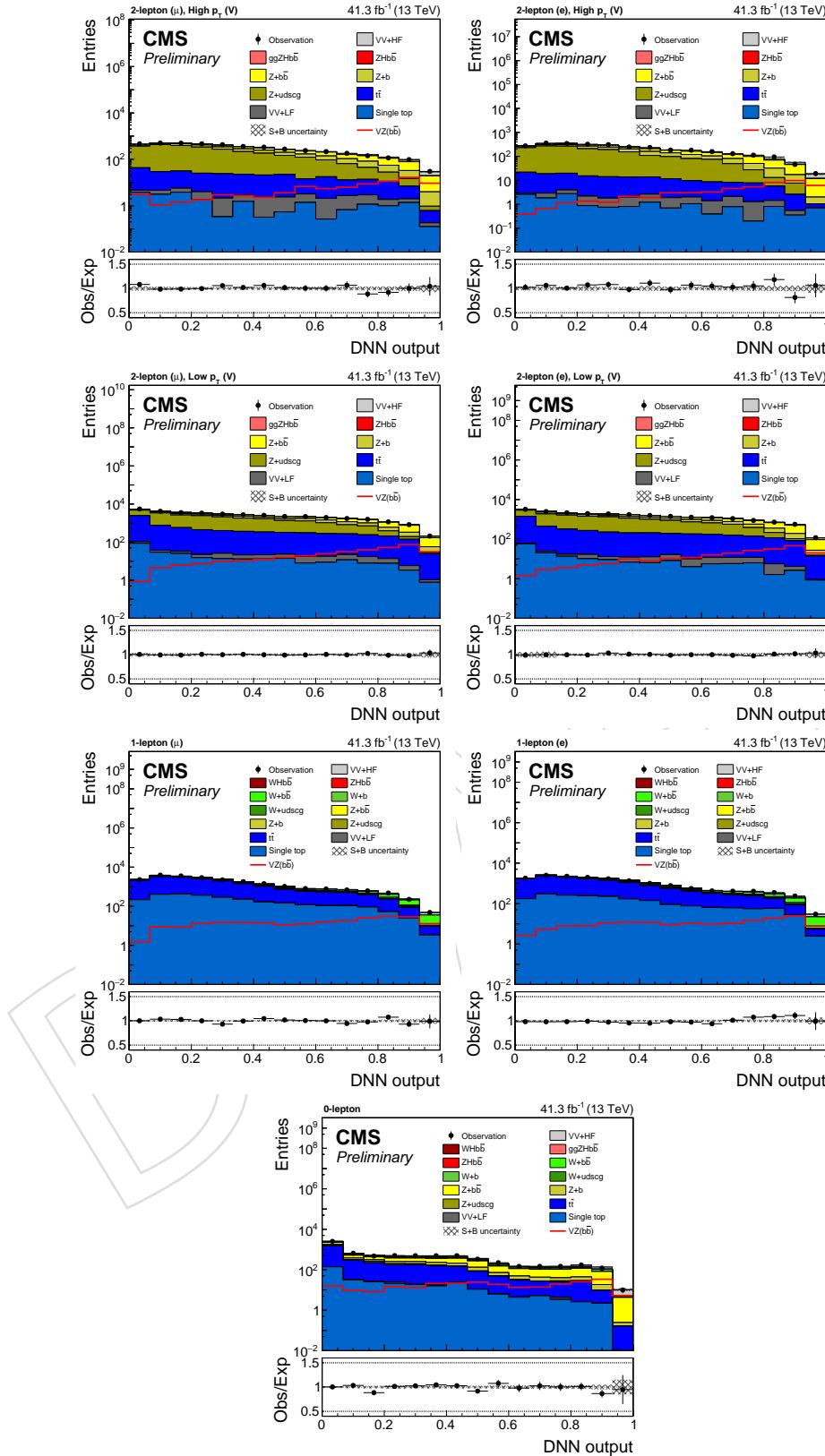


Figure 59: Post-fit distributions (VZ signal is post-fit) of multivariate discriminator output channels for 2017 analysis, after all signal region pre-selection criteria have been applied. First row: 2-lepton muon (left) and electron (right) channel for high $p_T(V)$ region, in the second row the low $p_T(V)$ is shown. Third row: 1-lepton muon (left) and electron (right) channel. Fourth row: 0-lepton channel. 0-lepton (bottom)

1628 **B.2 Results VZ**

1629 The impact of the most significant nuisances in the signal extraction fit is shown in Fig. 60.

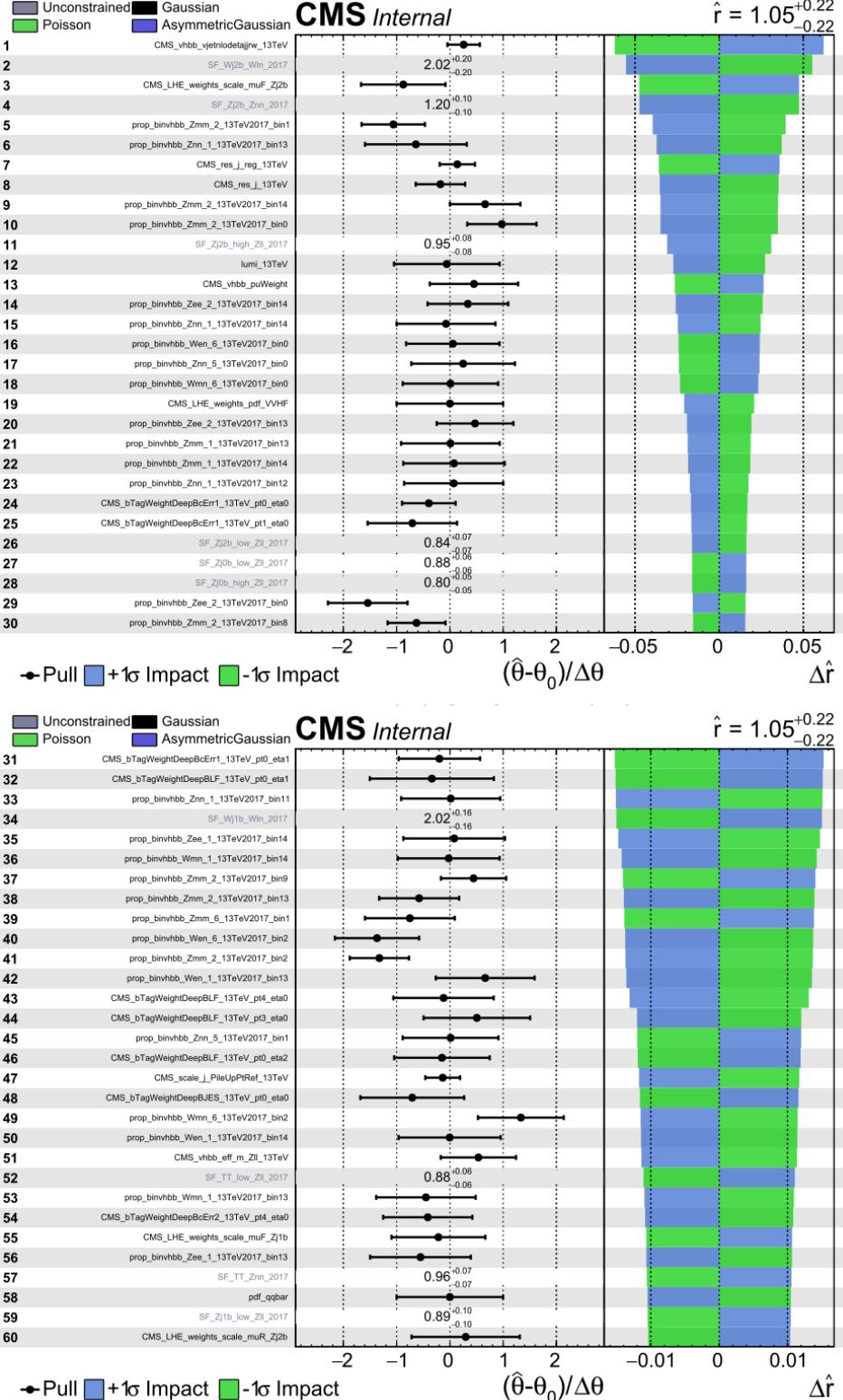


Figure 60: Impact plot of the most significant nuisances in the VZ signal extraction fit.

1630 The corresponding fitted signal strength is found to be $\mu = \sigma/\sigma_{\text{SM}} = 1.05^{+0.22}_{-0.21}$. The observed
 1631 (expected post-fit) local significance of the excess over background prediction is quantified as
 1632 5.2 σ (5.0 σ). The 2017 channel compatibility and signal strength is shown in Fig. 61. At 125 GeV
 1633 the signal strength for all modes is compatible with a p-value of 64%.

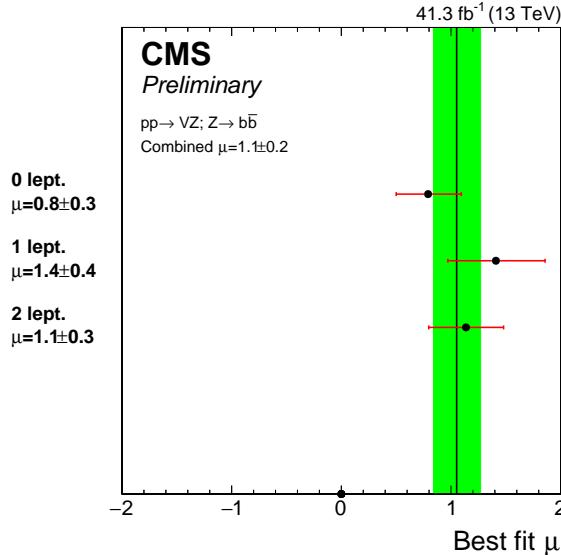


Figure 61: The most likely signal strength and uncertainty fitted in the three channels for 2017 analysis, together with the strength for WH and ZH.

1634 The results are summarized in Table 25, 26.

Table 25: Expected significance for the VH production cross section times $H \rightarrow b\bar{b}$ branching ratio, for the combination and sub-channels separately.

Channel	0-lepton	1-lepton	2-lepton	combo
2017 analysis (HIG-18-016)				
Expected post-fit SR+CR syst, SR+CR fitted SFs	3.46	2.14	3.02	4.99
Observed post-fit SR+CR syst, SR+CR fitted SFs	2.70	2.14	3.44	5.19

Table 26: Signal strength for the VH production cross section times $H \rightarrow b\bar{b}$ branching ratio, for the combination and sub-channels separately.

Channel	0-lepton	1-lepton	2-lepton	combo
2017 analysis (HIG-18-016) individual fits				
Observed post-fit SR+CR syst, SR+CR fitted SFs	$0.79^{+0.32}_{-0.30}$	$1.03^{+0.48}_{-0.47}$	$1.15^{+0.34}_{-0.33}$	–
2017 analysis (HIG-18-016) combined fit				
Observed post-fit SR+CR syst, SR+CR fitted SFs	0.8 ± 0.3	1.4 ± 0.4	1.1 ± 0.3	$1.05^{+0.22}_{-0.21}$