

CMS Draft Analysis Note

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

2020/07/16

Archive Hash: f509d3a-D

Archive Date: 2020/07/16

Search for Higgs boson decays into muons in proton-proton collisions at 13 TeV using Run2 data sets

Darin Acosta², Mohammad Alhusseini⁹, Paolo Azzurri⁷, Pierluigi Bortignon⁵, Dimitri Bourilkov², Andrew Brinkerhoff², Stephane Cooperstein⁴, Silvio Donato⁷, Irene Dutta¹¹, Raffaele Gerosa⁴, Leonardo Giannini⁸, Markus Klute¹, Dmitry Kondratyev¹⁰, Tobias Kramer³, Torben Lange³, Nan Lu¹¹, Giulio Mandorli⁸, Andrea Carlo Marini¹, Jane Nachtman⁹, Norbert Neumeister¹⁰, Yasar Onel⁹, Joosep Pata¹¹, Oliver Rieger³, Andrea Rizzi⁶, Vivek Sharma⁴, Peter Schleper³, Maria Spiropulu¹¹, Adish Vartak⁵, Jean-Roch Vlimant¹¹, Si Xie¹¹, and Xunwu Zuo²

¹ Massachusetts Institute of Technology, Cambridge, US

² University of Florida, Gainesville, US

³ University of Hamburg, Hamburg, Germany

⁴ University of California San Diego, San Diego, US

⁵ CERN, Geneva, CH

⁶ University of Pisa, Pisa, IT

⁷ INFN Sezione di Pisa, Pisa, IT

⁸ Scuola Normale Superiore di Pisa, Pisa, IT

⁹ University of Iowa, Iowa City, US

¹⁰ Purdue University, West Lafayette, US

¹¹ California Institute of Technology, Pasadena, US

Abstract

A measurement of the Higgs boson decay to a pair of muons is presented. This result combines searches in four exclusive categories targeting the production of the Higgs boson via gluon fusion, via vector boson fusion, in association with a weak vector boson, and in association with a pair of top quarks. The measurement is performed using $\sqrt{s} = 13$ TeV proton-proton (pp) collision data, corresponding to an integrated luminosity of 137 fb^{-1} , recorded by the CMS experiment at the CERN LHC. An excess of events is observed in data with a significance of 2.9 standard deviations, where the expectation for the standard model (SM) Higgs boson with $m_H = 125.38 \text{ GeV}$ is 2.5. The measured signal strength, relative to the SM expectation, is $1.18^{+0.41}_{-0.39}(\text{stat})^{+0.17}_{-0.15}(\text{sys})$. Assuming SM production cross sections for the various modes, the $H \rightarrow \mu\mu$ branching fraction is constrained at 95% CL to be within $0.7 \times 10^{-4} < \mathcal{B}(H \rightarrow \mu\mu) < 4.5 \times 10^{-4}$.

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

PDFAuthor: Hmm team
PDFTitle: Search for Higgs boson decays into muons in proton-proton collisions at 13 TeV using Run2 data sets
PDFSubject: CMS
PDFKeywords: CMS, physics, Higgs, muons, 13TeV

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

DRAFT

1 Contents

2	1	Introduction	5
3	2	Data sets and simulation	6
4	2.1	Simulation overview	6
5	2.2	Signal simulation	7
6	2.3	Background simulation	8
7	2.4	Triggers	10
8	2.5	Data sets	11
9	3	Event reconstruction	15
10	3.1	Primary vertex	15
11	3.2	Muons	15
12	3.3	Muon FSR recovery	15
13	3.4	Electrons	18
14	3.5	Jets	18
15	3.6	B-tagged jets	19
16	3.7	Missing transverse momentum	19
17	3.8	Prompt lepton identification	19
18	3.9	Muon and electron IDs used in the ttH channel	20
19	3.10	Muon and electron IDs used in the VH channels	21
20	3.11	Additional “soft” hadronic activity	22
21	4	Corrections to data and simulation	23
22	4.1	Pileup reweighting	23
23	4.2	B-tagging scale factors	24
24	4.3	L1 EGamma prefiring corrections	25
25	4.4	Muon efficiency and trigger scale factors	25
26	4.5	Correction to Z boson transverse momentum	26
27	4.6	Muon momentum scale and resolution	26
28	4.7	GeoFit correction to muon p_T	29
29	4.8	Event-by-event mass resolution	30
30	5	Analysis strategy	36
31	5.1	Main differences w.r.t. HIG-17-019	36
32	5.2	Signal model in ggH, VH, and ttH channels	37
33	5.3	Background model in ggH, VH, and ttH channels	38
34	5.3.1	Physics-inspired models	38
35	5.3.2	Agnostic models	39
36	5.4	Sources of systematic uncertainty	39
37	5.4.1	Experimental uncertainties	39
38	5.4.2	Theoretical uncertainties	41
39	6	Gluon-fusion channel	42
40	6.1	Multivariate discriminant	42
41	6.1.1	Input variables	43
42	6.1.2	BDT discriminant training	45
43	6.2	Event categorization	49

44	6.2.1	Category optimization	51
45	6.3	Signal extraction	55
46	6.3.1	Signal lineshape and yields	55
47	6.3.2	Uncertainties on the signal model	58
48	6.4	Background estimation	66
49	6.4.1	Uncertainty on the background model	68
50	6.4.2	The core-PDF method	71
51	6.5	Results	75
52	6.6	Effect of systematic uncertainties on the background model	78
53	7	VBF category based on MC template fit	80
54	7.1	Event selection: signal and control regions	80
55	7.2	Data vs MC comparisons	80
56	7.2.1	lepton and dilepton kinematics	82
57	7.2.2	jet and dijet kinematics	83
58	7.2.3	Other relevant variables	85
59	7.3	Gluon/Quark jet discrimination	85
60	7.4	Signal vs. background discrimination	88
61	7.4.1	Neural Network	88
62	7.5	Systematic uncertainties	91
63	7.5.1	Experimental uncertainties	91
64	7.5.2	Theory uncertainties	93
65	7.6	Results	95
66	7.6.1	Expected results	95
67	7.6.2	Observed results	96
68	7.6.3	Mass scan	96
69	8	ttH channel	103
70	8.1	Event pre-selection and categorization	103
71	8.2	Multivariate discriminants	105
72	8.2.1	Input observables	105
73	8.2.2	Training	108
74	8.2.3	BDT validation	110
75	8.2.4	BDT categorization	115
76	8.3	Signal extraction	116
77	8.3.1	Signal model	116
78	8.3.2	Uncertainties on the signal model	116
79	8.4	Background model	120
80	8.5	Results	121
81	9	VH Leptonic Categories	123
82	9.1	Event Selections	124
83	9.1.1	WH \rightarrow leptons Event Categories	124
84	9.1.2	ZH $\rightarrow \ell\ell + \mu\mu$ event selection	124
85	9.2	MVA discriminants in VH leptonic categories	125
86	9.2.1	BDT targeting WH $\rightarrow \ell + \mu\mu$ signal	125

87	9.2.2	BDT targeting ZH $\rightarrow \ell\ell + \mu\mu$ signal	126
88	9.2.3	Validation of WH BDT and ZH BDT	131
89	9.2.4	BDT categorization	135
90	9.3	VH signal models	135
91	9.3.1	VH signal systematic uncertainties	136
92	9.4	VH background models	138
93	9.4.1	VH Background systematic uncertainty	138
94	9.5	VH Results	139
95	10	Combined result	142
96	10.1	Combination procedure	142
97	10.2	Results for $m_H = 125.38$ GeV	144
98	10.3	Combination between Run1 and Run2 results	154
99	10.4	Measurement of Higgs boson coupling to muons	154
100	References	References	155
101	A	VBF “data-driven” analysis	157
102	A.1	Input variables for the VBF BDT	157
103	A.2	BDT discriminant trainings	158
104	A.3	Event categorization	163
105	A.3.1	Category optimization	165
106	A.4	Signal extraction	168
107	A.4.1	Signal lineshape and yields	169
108	A.5	Uncertainties on the signal model	170
109	A.5.1	Shape uncertainties	171
110	A.5.2	Rate uncertainties from theory	171
111	A.5.3	Rate uncertainties from experimental sources	173
112	A.6	Background estimation	177
113	A.7	Uncertainty on the background model	180
114	A.7.1	Toy generation for bias studies	180
115	A.7.2	Bias of the various background functions	180
116	A.8	Results	182
117	A.9	Effect of systematic uncertainties on the background model	186
118	B	Compatibility between VBF template and data-driven analyses	187
119	C	Bias studies for $m_H = 120$ and 130 GeV	190
120	D	GeoFit Corrections	194
121	E	LeptonMVA	204
122	F	Models for signal paramterization	212
123	G	Hadronic top quark reconstruction	215
124	H	Resolution categories	218
125	I	Studies of Drell-Yan background with FEWZ	220
126	I.1	FEWZ Calculations	220
127	I.2	Fit Functions	220
128	I.3	Results and Conclusions	221
129	I.3.1	NNLO Results	221

130	I.3.2 Conclusion	221
131	J Double Muon trigger for 2017	224
132	K HEM effect in 2018	225
133	L Boosted Decision Tree for VBF MC-templates	226

DRAFT

134 1 Introduction

135 Since the discovery of the 125 GeV Higgs boson in 2012 [1, 2] various measurements of its
 136 interactions with standard model (SM) particles have been performed. The interactions of the
 137 Higgs boson with the electroweak gauge bosons and charged fermions belonging to the third
 138 generation of SM particles have now been observed, and have been found to be consistent
 139 with SM predictions [3–11]. The direct interactions of the Higgs boson with first and second
 140 generation fermions, however, have yet to be established experimentally. In fact, since the
 141 coupling of the Higgs boson to fermions, as predicted by the SM, is proportional to the fermion
 142 mass, the $\mathcal{B}(H \rightarrow f\bar{f})$ is expected to be small for fermions belonging to the first and second
 143 generations. In particular, the $H \rightarrow \mu\mu$ branching ratio for the 125 GeV Higgs boson is predicted
 144 by the SM to be 2.18×10^{-4} [12].

145 The most recent search from the CMS Collaboration for $H \rightarrow \mu\mu$ decays was performed using
 146 a combination of proton-proton collision data collected at 7, 8 and 13 TeV, corresponding to in-
 147 tegrated luminosities of 5.0, 19.8 and 35.9 fb^{-1} , respectively [13]. The expected significance for
 148 this search was 1.0 standard deviation w.r.t. the background, whereas the expected upper limit
 149 on the production cross section times $\mathcal{B}(H \rightarrow \mu\mu)$ was 2.2 times the SM prediction. The ob-
 150 served significance and the observed upper limit were 0.9 standard deviations and 2.9 times the
 151 SM prediction, respectively. The measured signal strength is $\mu = 1.0 \pm 1.0 \text{ (stat)} \pm 0.1 \text{ (syst)}$.
 152 The most recent published search for $H \rightarrow \mu\mu$ decays from the ATLAS Collaboration has an-
 153 alyzed 13 TeV proton-proton collision data corresponding to a total integrated luminosity of
 154 80 fb^{-1} . This search sets an expected (observed) upper limit on the production cross section
 155 times $\mathcal{B}(H \rightarrow \mu\mu)$ of 2.0 (2.1) w.r.t. the SM prediction at 95% confidence level [14], and the
 156 corresponding measured signal strength is found to be $\mu = -0.1 \pm 1.5 \text{ (stat)}$. A preliminary
 157 analysis of the full Run 2 dataset has been also performed in Ref. [15], with an upper limit
 158 on the strength modifier of 1.3 (1.7). The corresponding best fit value for the signal strength
 159 parameters is $\mu = 0.5 \pm 0.7 \text{ (stat)}$.

160 In this analysis note, we describe the search for $H \rightarrow \mu\mu$ decays performed using the data set
 161 collected by the CMS detector during Run-2, at a center of mass energy of 13 TeV. This corre-
 162 sponds to a total integrated luminosity of 137.2 fb^{-1} . The analysis described here targets the
 163 Higgs boson production via gluon fusion (ggH), vector boson fusion (VBF), and in association
 164 with a vector boson (VH , $V = Z, W^\pm$) or with a top-antitop pair ($t\bar{t}H$). The final states of interest
 165 comprise a pair of two prompt, isolated and opposite charge muons produced in association
 166 with 0, 1 or 2 hadronic jets or in association with one or more additional leptons (μ, e). The
 167 analysis strategy consists in dividing the events into several categories in order to increase the
 168 signal to background ratio, as well as to target the production modes previously described. The
 169 signal is typically extract via a fit to the invariant mass distribution in each category, similarly
 170 to what done in Ref. [16]. Only in a specific section of the phase space, enriched in VBF signal
 171 events, the signal is extracted via a MC-based template fit to the output of an MVA discrimi-
 172 nant.

173 In the ggH and VBF enriched channels, the largest background in this analysis consists of Drell-
 174 Yan events in which an off-shell Z boson decays to a pair of muons. The vector boson fusion
 175 signal, although smaller in cross section compared to gluon fusion by an order of magnitude,
 176 provides a distinctive signature wherein the two muons from the Higgs boson decay are pro-
 177 duced in association with two jets that have a large pseudorapidity gap ($\Delta\eta_{jj}$), and form a large
 178 dijet invariant mass (m_{jj}). This feature helps to suppress the large Drell-Yan background, mak-
 179 ing the vector boson fusion channel a very sensitive probe of the $H \rightarrow \mu\mu$ signal. In contrast,
 180 in the VH and $t\bar{t}H$ channels the main backgrounds are represented by diboson (WZ, ZZ) and

181 $t\bar{t}$ -based ($t\bar{t}$, $t\bar{t}Z$) processes.

182 The description of the search for $H \rightarrow \mu\mu$ decays in the ggH channel, in which the total back-
 183 ground is extracted in a data-driven manner by fitting the dimuon mass ($m_{\mu\mu}$) spectrum with
 184 analytical functions, is taken from AN-19-090 [17]. Similarly, the description of the search for
 185 $H \rightarrow \mu\mu$ decays in the VBF final state via a MC-based template fit is derived from Refs. [18–20].
 186 Finally, an alternative analysis strategy for the VBF channel consists in extracting the total back-
 187 ground from data following a similar strategy to the one employed in the ggH channel. This
 188 alternative method yields to a 15% less performant result compared to the MC-based template
 189 fit. Results from this search are reported in Appendix A and are taken from AN-19-090 [17].

190 2 Data sets and simulation

191 Since this analysis targets a multitude of different final states, we take into account signal con-
 192 tributions from the main Higgs boson production modes at the LHC that pass our analysis
 193 selection requirements. The considered production modes are: gluon fusion (ggH), vector bo-
 194 son fusion (VBF), associated production with a W boson (WH), associated production with a Z
 195 boson (ZH), associated production with a top-quark pair (ttH), associated production with a b-
 196 quark pair (bbH), and associated production with single top (tHq and tHW). The WH and ZH
 197 processes are sometimes commonly referred to as VH modes. Table 2.1 provides the cross sec-
 198 tions for each of these processes for the 125 GeV SM Higgs boson, along with the respective the-
 199 oretical uncertainties, as recommended by the LHC Higgs Cross Section Working Group [12].
 200 The SM $H \rightarrow \mu\mu$ branching ratio for a 125 GeV Higgs boson is predicted to be 2.18×10^{-4} [12].
 201 The uncertainty on this branching ratio is estimated to be $\pm 1.23\%$.

Table 2.1: Higgs boson production cross sections for various modes at $\sqrt{s} = 13$ TeV.

Process	Cross section (pb)	Perturbative Order	+QCD scale unc. (%)	-QCD scale unc. (%)	+(PDF+ α_s) unc. (%)	-(PDF+ α_s) unc. (%)
ggH	48.58	N3LO(QCD) NLO (EWK)	+4.6	-6.7	+3.2	-3.2
VBF	3.782	NNLO (QCD) NLO (EWK)	+0.4	-0.3	+2.1	-2.1
WH	1.373	NNLO (QCD) NLO (EWK)	+0.5	-0.7	+1.9	-1.9
qq \rightarrow ZH	0.761	NNLO (QCD) NLO (EWK)	+0.5	-0.6	+1.9	-1.9
gg \rightarrow ZH	0.123	NLO (QCD)	+25.1	-18.9	+2.4	-2.4
ttH	0.507	NLO (QCD) NLO (EWK)	+5.8	-9.2	+3.6	-3.6
bbH	0.488	NNLO (QCD)	+20.2	-23.9	—	—
tHq	0.074	NLO (QCD)	+6.5	-14.9	+3.7	-3.7
tHW	0.015	NLO (QCD)	+4.9	-6.7	+6.3	-6.3

202 2.1 Simulation overview

203 Simulations are extremely useful and important to study the kinematic features of the signal
 204 and background processes, as well as to assess the impact of detector (mis)reconstruction on
 205 the physics objects used in the analysis. Simulated samples of signal and background events
 206 are used to train MVA discriminants that helps to isolate the Higgs boson signal from the SM
 207 backgrounds.

208 The processes considered in this analysis have been simulated using either the MADGRAPH5_aMC@NLO [21]
 209 or the POWHEG (v2) [22] generators. The matrix element level Monte Carlo (MC) events from

- 210 these generators are then interfaced with PYTHIA (v8.2 or greater) [23] in order to simulate the
 211 fragmentation and hadronization of partons in the initial and final states along with the under-
 212 lying event. This is done using the CUETP8M1 tune [24] for simulations corresponding to the
 213 2016 data taking era, and using the CP5 tune [25] for the 2017 and 2018 data taking periods.
 214 In the case of the processes simulated with the MADGRAPH5_aMC@NLO generator at leading
 215 order (next-to-leading order), jets from the matrix element calculations are matched to the par-
 216 ton shower produced by PYTHIA following the MLM (FxPx) prescription [26, 27]. The 2016
 217 (2017/18) era simulations use the NNPDF 3.0 (3.1) parton distribution functions [28, 29]. The
 218 interactions of all final state particles with the CMS detector are simulated using GEANT4 [30].

219 2.2 Signal simulation

220 The MC samples for the ggH production are simulated at next-to-leading order accuracy in
 221 QCD using both MADGRAPH5_aMC@NLO and POWHEG generators. Since the POWHEG simula-
 222 tion only contains events with positive weights, POWHEG samples have been used in the train-
 223 ing of the BDT multivariate discriminants, which cannot correctly handle negative weights.
 224 On the other hand, the MADGRAPH5_aMC@NLO samples are used in the final signal extrac-
 225 tion. Simulated events for VBF, VH, ttH, and bbH are produced with NLO precision in QCD
 226 theory via the POWHEG generator. Additional signal samples, obtained by varying the tune
 227 parameters for the underlying event simulation, are also produced and used to estimate the
 228 corresponding systematic uncertainty.

229 The Higgs boson signal produced through gluon-fusion is simulated by Monte Carlo gener-
 230 ators at NLO in QCD and then matched to the parton shower. The p_T of the Higgs boson is
 231 corrected to the prediction from the NNLOPS generator [31, 32], which is the highest order par-
 232 ton shower matched ggH simulation available that includes the finite top quark mass effects.
 233 The reweighting is performed in exclusive jet bins, $N_{\text{jets}} = 0, 1, 2, 3$ or more jets, where jets are
 234 defined at the generator level and required to have $p_T > 30 \text{ GeV}$. More details can be found in
 235 Section 4.1 of Ref. [6].

236 In the case of VBF processes, that have a distinct feature of two jets with a large $\Delta\eta_{jj}$ separation
 237 and no colour connection in the rapidity gap, it has been observed a significant difference in
 238 the parton showering done by PYTHIA and HERWIG [33] generators. The centrally produced
 239 VBF samples showered with PYTHIA have been used only for the training of the MVAs. In
 240 fact, as indicated by phenomenological studies shown in Ref. [34], the standard p_T -ordered
 241 shower implemented in PYTHIA, even when interfaced with a NLO generator like POWHEG,
 242 does not provide a good description of the additional hadronic activity in VBF events. In order
 243 to have a more correct description of the kinematics of VBF signal events when the final result
 244 is derived, VBF signal samples have been produced privately and showered by either PYTHIA
 245 in dipole shower mode¹ or HERWIG programs. The former configuration is used for the cen-
 246 tral prediction in each era, while the latter (events showered with HERWIG) is used to assign
 247 a conservative systematic uncertainty related to modelling of the parton shower. A detailed
 248 explanation of the VBF H signal modeling treatment including comparisons among many pos-
 249 sible parton shower simulation configurations is available in Appendix A of Ref. [18]. Parton
 250 showering with HERWIG is done using HERWIG++ (HERWIG7) with the UE-EE-5C tune [24]
 251 (CH3 tune) in the 2016 (2017 and 2018) era. For all the other signal production modes (ggH,
 252 VH, and ttH), only PYTHIA has been considered for the parton showering. The signal sim-
 253 ulated samples for 2016, 2017 and 2018 eras used in the analysis are listed in Tables 2.2, 2.3

¹By default PYTHIA produces parton shower following a p_T -ordered logic. However, it can also produce “dipole-shower” by setting the parameter “dipoleRecoil” to ON in the configuration.

254 and 2.4, respectively. In addition, simulated events have also been produced for Higgs boson
 255 masses of 120 and 130 GeV, allowing to interpolate signal acceptance and lineshape parameters
 256 over a 10 GeV mass range.

257 Simulated events for ggZH, tHq, and tHW processes have been produced privately because no
 258 official full-SIM samples were produced with exclusive decays of the Higgs boson into a pair
 259 of muons. The ggZ H process is produced at LO via POWHEG, while the remaining ones are
 260 also produced with LO accuracy with the MADGRAPH5_aMC@NLO generator.

Table 2.2: Signal simulated samples for the 2016 data taking era. The (*) stands for PUMoriond17_94X_mcRun2_asymptotic_v3.

Process	Sample	$\sigma \times BR$ (pb)
ggH	/GluGlu_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v2/MINIAODSIM	0.01057 pb
	/GluGluHToMuMu_M125_TuneCP5_PSweights_13TeV_amcatnloFXFX_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.01057 pb
	/GluGluHToMuMu_M125_TuneCP5up_PSweights_13TeV_amcatnloFXFX_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.01057 pb
	/GluGluHToMuMu_M125_TuneCP5down_PSweights_13TeV_amcatnloFXFX_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.01057 pb
	/GluGluHToMuMu_M120_TuneCP5_PSweights_13TeV_amcatnloFXFX_pythia8/*-v1/MINIAODSIM	0.012635 pb
	/GluGluHToMuMu_M130_TuneCP5_PSweights_13TeV_amcatnloFXFX_pythia8/*-v1/MINIAODSIM	0.008505 pb
VBF	/VBFHToMuMu_M-125_TuneCP5_PSweights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0008228 pb
	/VBFHToMuMu_M-125*-powheg-pythia_dipole/rgerosa-RunIISummer16MiniAODv3*/USER	0.0008228 pb
	/VBFHToMuMu_M-125*-powheg-herwigpp/rgerosa-RunIISummer16MiniAODv3*/USER	0.0008228 pb
	/VBFHToMuMu_M-125_TuneCP5up_PSweights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0008228 pb
	/VBFHToMuMu_M-125_TuneCP5down_PSweights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0008228 pb
	/VBFHToMuMu_M-120_TuneCP5_PSweights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000953 pb
	/VBFHToMuMu_M-130_TuneCP5_PSweights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000683 pb
WH	/WplusH_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v2/MINIAODSIM	0.0001858 pb
	/WplusH_HToMuMu_WToAll_M125_TuneCP5up_PSweights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0001858 pb
	/WplusH_HToMuMu_WToAll_M125_TuneCP5down_PSweights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0001858 pb
	/WplusH_HToMuMu_WToAll_M120_TuneCP5_PSweights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000232 pb
	/WplusH_HToMuMu_WToAll_M130_TuneCP5_PSweights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000139 pb
	/WminusH_HToMuMu_WToAll_M125_TuneCP5up_PSweights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v2/MINIAODSIM	0.0001164 pb
	/WminusH_HToMuMu_WToAll_M125_TuneCP5down_PSweights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0001164 pb
	/WminusH_HToMuMu_WToAll_M120_TuneCP5_PSweights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000148 pb
	/WminusH_HToMuMu_WToAll_M130_TuneCP5_PSweights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000088 pb
qqZH	/ZH_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v2/MINIAODSIM	0.0001923 pb
	/ZH_HToMuMu_ZToAll_M125_TuneCP5up_PSweights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0001923 pb
	/ZH_HToMuMu_ZToAll_M125_TuneCP5down_PSweights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0001923 pb
	/ZH_HToMuMu_ZToAll_M120_TuneCP5_PSweights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000241 pb
	/ZH_HToMuMu_ZToAll_M130_TuneCP5_PSweights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000148 pb
ggZH	/ggZH_HToMuMu_ZToLL_M125_13TeV_powheg_pythia8/rgerosa-RunIISummer16MiniAODv3*/USER	2.684e-06 pb
tH	/ttHToMuMu_M125_TuneCP5_PSweights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0001103 pb
	/ttHToMuMu_M125_TuneCP5up_PSweights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0001103 pb
	/ttHToMuMu_M125_TuneCP5down_PSweights_13TeV_powheg_pythia8/RunIISummer16MiniAODv3*-v1/MINIAODSIM	0.0001103 pb
	/ttHToMuMu_M120_TuneCP5_PSweights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000138 pb
	/ttHToMuMu_M130_TuneCP5_PSweights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000085 pb
bbH	/bbHToMuMu_M-125_4FS_yb2_TuneCP5_13TeV_amcatnlo_pythia8/RunIISummer16MiniAODv3*/MINIAODSIM	0.0001059 pb
tHq	/THQ_HToMuMu_TuneCP5_13TeV-madgraph_pythia/rgerosa-RunIISummer16MiniAODv3*/USER	1.612e-05 pb
tHW	/THW_HToMuMu_TuneCP5_13TeV-madgraph_pythia/rgerosa-RunIISummer16MiniAODv3*/USER	3.292e-06 pb

261 2.3 Background simulation

262 The largest contribution to the background in this search comes from Drell-Yan events. We
 263 are particularly interested in the off-shell production of the Z boson in the mass range of 110
 264 to 150 GeV. This is the typical range in which a fit is performed to the dimuon mass distri-
 265 bution in order to estimate the background from data and search for the Higgs boson signal
 266 peak. Therefore, a Drell-Yan MC sample has been generated via MADGRAPH5_aMC@NLO,
 267 with NLO precision in QCD and with up to 2 partons (jets) at the matrix element, applying a
 268 dimuon mass cut on LHE leptons in the range between 105 and 160 GeV. In addition, in order to
 269 gain statistics in a VBF-like phase space, a NLO QCD DY+2-jet sample has been produced via
 270 MADGRAPH5_aMC@NLO requiring two jets with $m_{jj} > 350$ GeV at the generator level. Dimuon
 271 events under the Z peak, however, are also of interest to study the kinematic properties of the
 272 Drell-Yan background, to compute certain data-based corrections, and, in general, to assess the
 273 reliability of the simulation. For such studies, a Drell-Yan MC sample with a dimuon mass cut
 274 of 50 GeV at the matrix element level, produced again using MADGRAPH5_aMC@NLO at NLO
 275 in QCD with up to 2 jets in the final state, is used.

276 When dimuon events with VBF-like jets are considered, the contribution from the electroweak

Table 2.3: Signal simulation samples for the 2017 data taking era. The (*) stands for PU2017_12Apr2018_94X_mc2017_realistic_v14.

Process	Sample	$\sigma \times \text{BR (pb)}$
ggH	/GluGluHToMuMu_M-125_TuneCP5_PSweights_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.01057 pb
	/GluGluHToMuMu_M125_TuneCP5_PSweights_13TeV_amcatnloFXFX_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.01057 pb
	/GluGluHToMuMu_M125_TuneCP5up_PSweights_13TeV_amcatnloFXFX_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.01057 pb
	/GluGluHToMuMu_M125_TuneCP5down_PSweights_13TeV_amcatnloFXFX_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.01057 pb
	/GluGluHToMuMu_M120_TuneCP5_PSweights_13TeV_amcatnloFXFX_pythia8/*-v1/MINIAODSIM	0.012653 pb
	/GluGluHToMuMu_M130_TuneCP5_PSweights_13TeV_amcatnloFXFX_pythia8/*-v1/MINIAODSIM	0.008505 pb
VBF	/VBFHToMuMu_M-125_TuneCP5_PSweights_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.0008228 pb
	/VBFHToMuMu_M-125*powheg-pythia_dipole/rgerosa-RunIIFall17MiniAODv2+MINIAODSIM*/USER	0.0008228 pb
	/VBFHToMuMu_M-125*powheg-herwig7.fixed/rgerosa-RunIIFall17MiniAODv2+MINIAODSIM*/USER	0.0008228 pb
	/VBFHToMuMu_M-125_TuneCP5up_PSweights_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.0008228 pb
	/VBFHToMuMu_M-125_TuneCP5down_PSweights_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.0008228 pb
	/VBFHToMuMu_M-120_TuneCP5_PSweights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000953 pb
WH	/WplusH_HToMuMu_WToAll_M125_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v3/MINIAODSIM	0.0001858 pb
	/WplusH_HToMuMu_WToAll_M125_TuneCP5up_PSweights_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.0001858 pb
	/WplusH_HToMuMu_WToAll_M125_TuneCP5down_PSweights_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.0001858 pb
	/WplusH_HToMuMu_WToAll_M120_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000232 pb
	/WplusH_HToMuMu_WToAll_M130_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000139 pb
	/WminusH_HToMuMu_WToAll_M125_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.0001164 pb
ZH	/WminusH_HToMuMu_WToAll_M125_TuneCP5up_PSweights_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.0001164 pb
	/WminusH_HToMuMu_WToAll_M125_TuneCP5down_PSweights_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.0001164 pb
	/WminusH_HToMuMu_WToAll_M120_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000148 pb
	/WminusH_HToMuMu_WToAll_M130_13TeV_powheg_pythia8/*-v1_1.MINIAODSIM	0.000088 pb
	/ZH_HToMuMu_ZToAll_M125_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.0001923 pb
	/ZH_HToMuMu_ZToAll_M125_TuneCP5up_PSweights_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.0001923 pb
ggZH	/ZH_HToMuMu_ZToAll_M125_TuneCP5down_PSweights_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.0001923 pb
	/ZH_HToMuMu_ZToAll_M120_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.000241 pb
	/ZH_HToMuMu_ZToAll_M130_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.000148 pb
	/ggZH_HToMuMu_ZToLL_M125_13TeV_powheg_pythia8/rgerosa*RunIIFall17MiniAODv2*/USER	2.684e-06 pb
	/ttH_HToMuMu_M125_TuneCP5_PSweights_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.0001103 pb
	/ttH_HToMuMu_M125_TuneCP5up_PSweights_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.0001103 pb
tHq	/ttH_HToMuMu_M125_TuneCP5down_PSweights_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.0001103 pb
	/ttH_HToMuMu_M120_TuneCP5_PSweights_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.000138 pb
	/ttH_HToMuMu_M130_TuneCP5_PSweights_13TeV_powheg_pythia8/RunIIFall17MiniAODv2*-v1/MINIAODSIM	0.000085 pb
	/bbH_bbToMuMu_M-125_4FS_yb2_TuneCP5_13TeV-amcatnlo-pythia8/RunIIFall17MiniAODv2*/USER	0.0001059 pb
	/THQ_HToMuMu_TuneCP5_13TeV-madgraph-pythia/rgerosa-RunIIFall17MiniAODv2*/USER	1.612e-05 pb
	/THW_HToMuMu_TuneCP5_13TeV-madgraph-pythia/rgerosa-RunIIFall17MiniAODv2*/USER	3.292e-06 pb

277 production of the Z boson becomes significant (Z_{jj} -EW). This process has been simulated at
278 LO using MADGRAPH5_aMC@NLO and showered via HERWIG ++ (2016 simulated samples)
279 or HERWIG7 (2017 and 2018 MC). The HERWIG parton shower is adopted instead of PYTHIA
280 because it is known to better model purely electroweak process without color connection as
281 reported in Ref. [35]. Two alternative set of samples for the Z_{jj} -EW production are available:
282 one in which the invariant mass of the two LHE muons is required to be larger than 50 GeV,
283 the other one in which $m_{\mu\mu}$ is between 105 and 160 GeV. The former sample is used when
284 data and simulation are compared under the Z-peak, while the latter is used for the signal
285 region. A significant bias was observed for the prediction from the original centrally produced
286 EW Z_{jj} MC samples because of a generator-level selection applied that removed events that
287 would have later passed the offline selections. This effect was studied in detail and a new set
288 of EW Z_{jj} samples was produced covering the full offline phase space, with the quark-level p_T
289 requirement loosened from 25 GeV to 1 GeV. A detailed description of these studies is available
290 in Appendix C of Ref. [18]. In order to estimate an uncertainty due to the parton shower model,
291 an alternative set of Z_{jj} -EW samples have been produced privately showering the same events
292 at ME with PYTHIA in dipole shower mode.

293 After Drell-Yan, one of the most significant background contribution comes from $t\bar{t}$ events in
294 which both the top quarks decay leptonically. The process has been generated at NLO using the
295 POWHEG generator. Furthermore, there are minor contributions from other top quark processes
296 such as single top (tW, t, and s-channel) production, ttZ, ttW, ttWW, tZq, and semi-leptonic $t\bar{t}$
297 decays. These contributions are also taken into account via dedicated NLO simulation pro-
298 duced with MADGRAPH5_aMC@NLO or POWHEG. Remaining background can be attributed to
299 the diboson processes (ZZ, WZ, WW) with some very small contributions from triboson pro-
300 duction (WWW, WWZ, WZZ, ZZZ) that have also been taken into account. The diboson pro-

Table 2.4: Signal simulation samples for the 2018 data taking era. The (*) stands for 102X_upgrade2018_realistic_v15.

Process	Sample	$\sigma \times \text{BR} (\text{pb})$
ggH	/GluGluHToMuMu_M125_TuneCP5_PSweights_13TeV_amcatnloFXFX_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.01057 pb
	/GluGluHToMuMu_M-125_TuneCP5_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.01057 pb
	/GluGluHToMuMu_M125_TuneCP5up_PSweights_13TeV_amcatnloFXFX_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.01057 pb
	/GluGluHToMuMu_M125_TuneCP5down_PSweights_13TeV_amcatnloFXFX_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.01057 pb
	/GluGluHToMuMu_M120_TuneCP5_PSweights_13TeV_amcatnloFXFX_pythia8/*-v1/MINIAODSIM	0.012653 pb
	/GluGluHToMuMu_M130_TuneCP5_PSweights_13TeV_amcatnloFXFX_pythia8/*-v1/MINIAODSIM	0.008505 pb
VBF	/VBFHToMuMu_M-125_TuneCP5_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v1/MINIAODSIM	0.0008228 pb
	/VBFHToMuMu_M-125*-powheg_pythia_dipole_fixed/rgerosa-RunIIAutumn18MiniAOD*-MINIAODSIM*/USER	0.0008228 pb
	/VBFHToMuMu_M-125*-powheg_herwig7/rgerosa-RunIIAutumn18MiniAOD*-MINIAODSIM*/USER	0.0008228 pb
	/VBFHToMuMu_M125_TuneCP5up_PSweights_13TeV_amcatnlo_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.0008228 pb
	/VBFHToMuMu_M125_TuneCP5down_PSweights_13TeV_amcatnlo_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.0008228 pb
	/VBFHToMuMu_M120_TuneCP5_PSweights_13TeV_amcatnlo_pythia8/RunIIAutumn18MiniAOD*-v1/MINIAODSIM	0.000953 pb
	/VBFHToMuMu_M130_TuneCP5_PSweights_13TeV_amcatnlo_pythia8/*-v1/MINIAODSIM	0.000683 pb
WH	/WplusH_HToMuMu_WToAll_M125_TuneCP5_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.0001858 pb
	/WplusH_HToMuMu_WToAll_M125_TuneCP5up_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.0001858 pb
	/WplusH_HToMuMu_WToAll_M125_TuneCP5down_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.0001858 pb
	/WplusH_HToMuMu_WToAll_M120_TuneCP5_PSweights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000232 pb
	/WplusH_HToMuMu_WToAll_M130_TuneCP5_PSweights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000139 pb
	/WminusH_HToMuMu_WToAll_M125_TuneCP5_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.0001164 pb
	/WminusH_HToMuMu_WToAll_M125_TuneCP5up_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.0001164 pb
	/WminusH_HToMuMu_WToAll_M125_TuneCP5down_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.0001164 pb
	/WminusH_HToMuMu_WToAll_M120_TuneCP5_PSweights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000148 pb
	/WminusH_HToMuMu_WToAll_M130_TuneCP5_PSweights_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.000088 pb
ZH	/ZH_HToMuMu_ZToAll_M125_TuneCP5_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.0001923 pb
	/ZH_HToMuMu_ZToAll_M125_TuneCP5up_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.0001923 pb
	/ZH_HToMuMu_ZToAll_M125_TuneCP5down_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.0001923 pb
	/ZH_HToMuMu_ZToAll_M120_TuneCP5_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.000241 pb
	/ZH_HToMuMu_ZToAll_M130_TuneCP5_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.000148 pb
ggZH	/ggZH_HToMuMu_ZToLL_M125_13TeV_powheg_pythia8/rgerosa-RunIIAutumn18MiniAOD*/USER	2.68e-06 pb
tH	/ttHToMuMu_M125_TuneCP5_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.0001103 pb
	/ttHToMuMu_M125_TuneCP5up_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.0001103 pb
	/ttHToMuMu_M125_TuneCP5down_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.0001103 pb
	/ttHToMuMu_M120_TuneCP5_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.000138 pb
	/ttHToMuMu_M130_TuneCP5_PSweights_13TeV_powheg_pythia8/RunIIAutumn18MiniAOD*-v2/MINIAODSIM	0.000085 pb
bbH	/bbHToMuMu_M-125_4FS_yb2_TuneCP5_13TeV_amcatnlo_pythia8/RunIIAutumn18MiniAOD*/MINIAODSIM	0.0001059 pb
THq	/THQ_HToMuMu_TuneCP5_13TeV-madgraph_pythia/rgerosa-RunIIAutumn18MiniAOD*/USER	1.612e-05 pb
THW	/THW_HToMuMu_TuneCP5_13TeV-madgraph_pythia/rgerosa-RunIIAutumn18MiniAOD*/USER	3.292e-06 pb

cesses have been simulated using either **MADGRAPH5_aMC@NLO** or **POWHEG** at NLO in QCD, whereas the triboson processes have been simulated using **MADGRAPH5_aMC@NLO** with NLO precision in QCD corrections.

The centrally produced background simulated samples for 2016, 2017 and 2018 eras are listed in Tables 2.5, 2.6 and 2.7, respectively. The cross sections used for normalizing the background expected yields are obtained from the best available theoretical predictions. In particular, the Drell-Yan cross section has been obtained from the **FEWZ** [36–38] generator at NNLO accuracy in QCD, and NLO accuracy in electroweak corrections. Similarly, the cross section for the $t\bar{t}$ has been computed at NNLO+NNLL accuracy in QCD [39]. The cross section for WW production is taken from Ref. [40] calculated at NNLO in QCD, while the NLO cross section for the WZ process, as obtained from the simulated MC samples, is corrected with NNLO/NLO k-factors provided in Ref. [41]. The fully differential cross section for the $q\bar{q} \rightarrow ZZ$ process has been computed at NNLO in QCD [42], and the NNLO/NLO K-factor as a function of m_{ZZ} has been applied to the simulated samples. This K-factor varies from 1.0 to 1.2 as a function of m_{ZZ} . Finally, the $gg \rightarrow ZZ$ cross section is corrected with NNLO/LO k-factors following the same strategy detailed in Ref. [43].

2.4 Triggers

Signal events in this analysis contain two prompt, isolated, and high p_T muons. When the Higgs boson is produced at rest in the transverse plane ($p_T^H \approx 0$), the two muons are emitted back-to-back with a transverse momentum of about $m_H/2 \approx 60 \text{ GeV}$. At the generator level, the p_T distribution for the leading muon produced in both ggH and VBF signal events turns on at about 40 GeV, peaking around 60 GeV as previously explained. Therefore, signal events

Table 2.5: Background simulation samples for the 2016 data taking era. The (*) stands for RunIIISummer16MiniAODv3-PUMoriond17.94X_mcRun2_asymptotic_v3.

Process	Sample	Cross section (pb)	Cross section Precision
Drell-Yan	/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/*_ext2-v1/MINIAODSIM	5954	NNLO (QCD) NLO (EWK)
	/DYJetsToLL_M-105To160_TuneCUETP8M1_PSweights_13TeV-amcatnloFXFX-pythia8/*-v1/MINIAODSIM	47.12	NNLO (QCD) NLO (EWK)
	/DYJetsToLL_M-105To160_VBFFilter_TuneCUETP8M1_PSweights_13TeV-amcatnloFXFX-pythia8/*/MINIAODSIM	1.774	NNLO (QCD) NLO (EWK)
Zjj-EW	/EWK_LLJJ_MLL_105-160_SM_5f_LO_TuneCH3_13TeV-madgraph-herwig7_corrected/*/MINIAODSIM	0.075	LO
	/EWK_LLJJ_MLL_105-160_TuneCP5_13TeV-madgraph-pythia.dipole/rgerosa-RunIIISummer16MiniAODv3*/USER	0.075	LO
	/EWK_LLJJ_MLL50_MJJ-120_TuneCH3_PSweights_13TeV-madgraph-herwig7_corrected/*/MINIAODSIM	0.983	LO
Top	/TTTo2LNu_TuneCUETP8M2_ttHreanche3_13TeV-powheg-pythia8/*-v2/MINIAODSIM	86.61	NNLO (QCD)
	/TTToSemilepton_TuneCUETP8M2_ttHreanche3_13TeV-powheg-pythia8/*-v2/MINIAODSIM	358.57	NNLO (QCD)
	/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/*_ext1-v1/MINIAODSIM	35.9	NNLO (QCD)
	/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/*_ext1-v1/MINIAODSIM	35.9	NNLO (QCD)
	/ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1/*-v1/MINIAODSIM	80.95	NLO (QCD)
	/ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1/*-v1/MINIAODSIM	136.02	NLO (QCD)
	/ST_s-channel_4f_leptonDecays_13TeV-amcatnlo_pythia8_TuneCUETP8M1/*-v1/MINIAODSIM	3.40	NLO (QCD)
	/tZq_ll_4f_13TeV-amcatnlo_pythia8/*_ext1-v1/MINIAODSIM	0.0758	NLO (QCD)
	/TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8/*_ext2-v1/MINIAODSIM	0.2001	NLO (QCD)
	/TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo_pythia8/*_ext2-v1/MINIAODSIM	0.2529	NLO (QCD)
	/TTTT_TuneCUETP8M2T4_13TeV-amcatnlo_pythia8/*-v1/MINIAODSIM	0.0091	NLO (QCD)
	/TTWW_TuneCUETP8M2T4_13TeV-amcatnlo_pythia8/*_ext1-v1/MINIAODSIM	0.00073	NLO (QCD)
	/TTWW_TuneCUETP8M2T4_13TeV-amcatnlo_pythia8/*_ext1-v1/MINIAODSIM	0.0070	NLO (QCD)
	/TTJJ_TuneCUETP8M2T4_13TeV-amcatnlo_pythia8/*_ext1-v1/MINIAODSIM	0.000039	NLO (QCD)
Diboson	/WWTo2L2Nu_13TeV-powheg/*-v2/MINIAODSIM	12.178	NNLO (QCD)
	/WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8/*_ext1-v1/MINIAODSIM	4.102	NLO (QCD)
	/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/*-v2/MINIAODSIM	5.606	NLO (QCD)
	/ZZTo2L2Nu_13TeV_powheg_pythia8_ext1/*-v2/MINIAODSIM	0.5644	NLO (QCD)
	/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/*-v3/MINIAODSIM	3.224	NLO (QCD)
	/ZZTo4L_13TeV_powheg_pythia8/*-v1/MINIAODSIM	1.256	NLO (QCD)
	/GluGluToContInToZZTo_2e2mu_2nu2tau_2e2tau_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.00318	LO
	/GluGluToContInToZZTo_4mu_4tau_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.00157	LO
	/GluGluToContInToZZTo_2mu2nu_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.00149	LO
Triboson	/WWW_4F_TuneCUETP8M1_13TeV-amcatnlo_pythia8/*-v2/MINIAODSIM	0.2086	NLO (QCD)
	/WWZ_TuneCUETP8M1_13TeV-amcatnlo_pythia8/*-v2/MINIAODSIM	0.1651	NLO (QCD)
	/WZZ_TuneCUETP8M1_13TeV-amcatnlo_pythia8/*-v2/MINIAODSIM	0.05565	NLO (QCD)
	/ZZZ_TuneCUETP8M1_13TeV-amcatnlo_pythia8/*-v2/MINIAODSIM	0.01398	NLO (QCD)

323 can be selected with high efficiency in data using online single muon triggers. The high level
 324 trigger paths for the single muon triggers used in this search are listed in Table 2.8. These
 325 triggers impose a loose isolation requirement on each muon candidate, and a p_T threshold of
 326 27 (24) GeV in 2017 (2016, 2018).

327 2.5 Data sets

328 The analysis is performed using dimuon events in which at least one of the two muons fires
 329 the single muon triggers listed in Section 2.4. For this reason, offline events are selected from
 330 the SingleMuon data set in all the three years – 2016, 2017 and 2018.

331 The 2016 data used in this analysis correspond to an integrated luminosity of 35.9 fb^{-1} . The
 332 data are re-analyzed with the MiniAODv3 and with the new analysis strategy, described in the
 333 following sections. Good run ranges and luminosity blocks are selected using the JSON file
 334 Cert_271036-284044_13TeV_ReReco_07Aug2017_Collisions16_JSON.txt.

335 The 2017 data used in this analysis correspond to an integrated luminosity of 41.5 fb^{-1} . Good
 336 run ranges and luminosity blocks are selected using the JSON file
 337 Cert_294927-306462_13TeV_EOY2017ReReco_Collisions17_JSON_v1.txt.

338 The 2018 data used in this analysis correspond to an integrated luminosity of 59.7 fb^{-1} . Good
 339 run ranges and luminosity blocks are selected using the JSON file
 340 Cert_314472-325175_13TeV_17SeptEarlyReReco2018ABC_PromptEraD_Collisions18_JSON.txt.

Table 2.6: Background simulation samples for the 2017 data taking era. The (*) stands for RunIIFall17MiniAODv2-PU2017_12Apr2018_94X_mc2017_realistic_v14.

Process	Sample	Cross section (pb)	Cross section Precision
Drell-Yan	/DYJetsToLL_M-50.TuneCP5_13TeV-amcatnloFXFX-pythia8/*v1/MINIAODSIM	6225.4	NNLO (QCD)
	/DYJetsToLL_M-105To160.TuneCP5_PSweights_13TeV-amcatnloFXFX-pythia8/*-v1/MINIAODSIM	47.12	NLO (EWK)
	/DYJetsToLL_M-105To160_VBFFilter_TuneCP5_PSweights_13TeV-amcatnloFXFX-pythia8/*/MINIAODSIM	2.043	NNLO (QCD)
Zjj-EW	/EWK_LJJJ_MLL_105-160_SM_5f_LO_TuneCH3_13TeV-madgraph-herwig7_corrected/*/MINIAODSIM	0.078	LO
	/EWK_LJJJ_MLL_105-160_TuneCP5_13TeV-madgraph-pythia_dipole_rgerosa_RunIIFall17MiniAODv2*/USER	0.078	LO
	/EWK_LJJJ_MLL50_MJJ-120_TuneCH3_PSweights_13TeV-madgraph-herwig7_corrected/*/MINIAODSIM	1.029	LO
Top	/TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8/*-v1/MINIAODSIM	86.61	NNLO (QCD)
	/TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8/*-v2/MINIAODSIM	358.57	NNLO (QCD)
	/ST_tW_top_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8/*-v1/MINIAODSIM	35.9	NNLO (QCD)
	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8/*-v2/MINIAODSIM	35.9	NNLO (QCD)
	/ST_t-channel_antitop_5f_TuneCP5_PSweights_13TeV-powheg-pythia8/*-v1/MINIAODSIM	80.95	NLO (QCD)
	/ST_t-channel_top_5f_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM	136.02	NLO (QCD)
	/ST_s-channel_antitop_leptonDecays_13TeV-PSweights_powheg-pythia/*-v1/MINIAODSIM	1.293	NLO (QCD)
	/ST_s-channel_top_leptonDecays_13TeV-PSweights_powheg-pythia/*-v1/MINIAODSIM	2.069	NLO (QCD)
	/tZq_ll_4f_ckm_NLO_TuneCP5_PSweights_13TeV-amcatnlo_pythia8/*-v1/MINIAODSIM	0.0758	NLO (QCD)
	/TTWJetsToLNu_TuneCP5_PSweights_13TeV-amcatnloFXFX-madspin-pythia8/*-v1/MINIAODSIM	0.2001	NLO (QCD)
Diboson	/TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo_pythia8/*/MINIAODSIM	0.2529	NLO (QCD)
	/TTT_TuneCP5_13TeV-amcatnlo_pythia8/*-v1/MINIAODSIM	0.0091	NLO (QCD)
	/TTTW_TuneCP5_13TeV-amcatnlo_pythia8/*-v1/MINIAODSIM	0.00073	NLO (QCD)
	/TTTJ_TuneCP5_13TeV-amcatnlo_pythia8/*-v1/MINIAODSIM	0.0070	NLO (QCD)
	/TTWW_TuneCP5_13TeV-amcatnlo_pythia8/*-v1/MINIAODSIM	0.000039	NLO (QCD)
	/WWTo2L2Nu_NNPDF31_TuneCP5_PSweights_13TeV-powheg-pythia8/*_ext1-v1/MINIAODSIM	12.178	NNLO (QCD)
	/WZTo3LNu_13TeV-powheg-pythia8/*-v2/MINIAODSIM	4.658	NLO (QCD)
Triboson	/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/*-v1/MINIAODSIM	6.331	NLO (QCD)
	/ZZTo2L2Nu_13TeV_powheg_pythia8/*-v1/MINIAODSIM	0.601	NLO (QCD)
	/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/*-v1/MINIAODSIM	3.688	NLO (QCD)
	/ZZTo4L_13TeV_powheg_pythia8/*-v1/MINIAODSIM	1.325	NLO (QCD)
	/GluGluToContinToZZTo_(2e2mu,2mu2tau,2e2tau)_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.00329	LO
	/GluGluToContinToZZTo_(4mu,4tau)_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.00140	LO
	/GluGluToContinToZZTo_(2mu2nu)_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.00177	LO
	/WWW_4F_TuneCP5_13TeV-amcatnlo_pythia8/*-v2/MINIAODSIM	0.2086	NLO (QCD)
	/WWZ_4F_TuneCP5_13TeV-amcatnlo_pythia8/*-v2/MINIAODSIM	0.1651	NLO (QCD)
	/WZZ_TuneCP5_13TeV-amcatnlo_pythia8/*-v1/MINIAODSIM	0.05565	NLO (QCD)
	/ZZZ_TuneCP5_13TeV-amcatnlo_pythia8/*-v2/MINIAODSIM	0.01398	NLO (QCD)

Table 2.7: Background simulation samples for the 2018 data taking era. The (*) stands for RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15.

Process	Sample	Cross section (pb)	Cross section Precision
Drell-Yan	/DYJetsToLL_M-50.TuneCP5.13TeV-amcatnloFXFX-pythia8/*v1/MINIAODSIM	6225.4	NNLO (QCD)
	/DYJetsToLL_M-105To160.TuneCP5_PSweights_13TeV-amcatnloFXFX-pythia8/*-v1/MINIAODSIM	47.12	NLO (EWK)
	/DYJetsToLL_M-105To160_VBFFilter_TuneCP5_PSweights_13TeV-amcatnloFXFX-pythia8/*/MINIAODSIM	2.029	NNLO (QCD)
Zjj-EW	/EWK_LJJJ_MLL_105-160_SM_5f_LO_TuneCH3_13TeV-madgraph-herwig7_corrected/*/MINIAODSIM	0.078	LO
	/EWK_LJJJ_MLL_105-160_TuneCP5_13TeV-madgraph-pythia.dipole/rgerosa-RunIIAutumn18MiniAOD*/USER	0.078	LO
	/EWK_LJJJ_MLL50_MJJ_120_TuneCH3_PSweights_13TeV-madgraph-herwig7_corrected/*/MINIAODSIM	1.029	LO
Top	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM	86.61	NNLO (QCD)
	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM	358.57	NNLO (QCD)
	/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/*_ext1-v1/MINIAODSIM	35.9	NNLO (QCD)
	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/*_ext1-v1/MINIAODSIM	35.9	NNLO (QCD)
	/ST_t-channel_antitop_5f_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM	80.95	NLO (QCD)
	/ST_t-channel_top_5f_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM	136.02	NLO (QCD)
	/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-madgraph-pythia8/*_ext1-v4/MINIAODSIM	3.40	NLO (QCD)
	/tZq_ll_4f_ckm_NLO_TuneCP5_13TeV-madgraph-pythia8/*_ext1-v2/MINIAODSIM	0.0758	NLO (QCD)
	/TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/*_ext1-v2/MINIAODSIM	0.2001	NLO (QCD)
	/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM	0.2529	NLO (QCD)
	/TTW_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM	0.0091	NLO (QCD)
	/TTWW_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM	0.00073	NLO (QCD)
	/TTJ_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM	0.00698	NLO (QCD)
		0.00039	NLO (QCD)
Diboson	/WWTo2L2Nu_NNPDF31_TuneCP5_13TeV-powheg-pythia8/*-v1/MINIAODSIM	12.178	NNLO (QCD)
	/WZTo3LNu_TuneCP5_13TeV-powheg-pythia8/*_ext1-v2/MINIAODSIM	4.658	NLO (QCD)
	/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/*-v1/MINIAODSIM	6.321	NLO (QCD)
	/ZZTo2L2Nu_TuneCP5_13TeV_powheg_pythia8/*_ext1-v2/MINIAODSIM	0.601	NLO (QCD)
	/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/*-v1/MINIAODSIM	3.696	NLO (QCD)
	/ZZTo4L_TuneCP5_13TeV_powheg_pythia8/*_ext1-v2/MINIAODSIM	1.325	NLO (QCD)
	/GluGluToContinToZZTo_(2e2mu,2mu2tau,2e2tau)_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.00329	LO
Triboson	/GluGluToContinToZZTo_(4mu,4tau)_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.001402	LO
	/GluGluToContinToZZTo_(2mu2nu)_13TeV_MCFM701_pythia8/*-v2/MINIAODSIM	0.001772	LO
	/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM	0.2086	NLO (QCD)
	/WWZ_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM	0.1651	NLO (QCD)
/WZZ_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM		0.05565	NLO (QCD)
	/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8/*_ext1-v2/MINIAODSIM	0.01398	NLO (QCD)

Table 2.8: Single muon trigger paths used for selecting events for further analysis in 2016, 2017 and 2018 eras.

Year	Data set path
2016	HLT_IsoMu24_v* OR HLT_IsoTkMu24_v*
2017	HLT_IsoMu27_v*
2018	HLT_IsoMu24_v*

Table 2.9: Single muon data sets used in this analysis for the 2016 era.

Dataset	Run Range	Int. Lum.
/SingleMuon/Run2016B-17Jul2018-v1/MINIAOD	272007—275376	5.8 fb ⁻¹
/SingleMuon/Run2016C-17Jul2018-v1/MINIAOD	275657—276283	2.6 fb ⁻¹
/SingleMuon/Run2016D-17Jul2018-v1/MINIAOD	276315—276811	4.2 fb ⁻¹
/SingleMuon/Run2016E-17Jul2018-v1/MINIAOD	276831—277420	4.0 fb ⁻¹
/SingleMuon/Run2016F-17Jul2018-v1/MINIAOD	277772—278808	3.1 fb ⁻¹
/SingleMuon/Run2016G-17Jul2018-v1/MINIAOD	278820—280385	7.5 fb ⁻¹
/SingleMuon/Run2016H-17Jul2018-v1/MINIAOD	280919—284044	8.6 fb ⁻¹

Table 2.10: Single muon data sets used in this analysis for the 2017 era.

Dataset	Run Range	Int. Lum.
/SingleMuon/Run2017B-31Mar2018-v1/MINIAOD	297047—299329	4.8 fb ⁻¹
/SingleMuon/Run2017C-31Mar2018-v1/MINIAOD	299368—302029	9.6 fb ⁻¹
/SingleMuon/Run2017D-31Mar2018-v1/MINIAOD	302031—302663	4.2 fb ⁻¹
/SingleMuon/Run2017E-31Mar2018-v1/MINIAOD	303824—304797	9.2 fb ⁻¹
/SingleMuon/Run2017F-31Mar2018-v1/MINIAOD	305040—306462	13.4 fb ⁻¹

Table 2.11: Single muon data sets used in this analysis for the 2018 era.

Dataset	Run Range	Int. Lum.
/SingleMuon/Run2018A-17Sep2018-v2/MINIAOD	315257—316995	14.00 fb ⁻¹
/SingleMuon/Run2018B-17Sep2018-v1/MINIAOD	317080—319310	7.10 fb ⁻¹
/SingleMuon/Run2018C-17Sep2018-v1/MINIAOD	319337—320655	6.94 fb ⁻¹
/SingleMuon/Run2018D-31Mar2018-v1/MINIAOD	320500—325175	31.93 fb ⁻¹

341 3 Event reconstruction

342 The analysis primarily relies on muons and jets in order to target the final state signature of
 343 $H \rightarrow \mu\mu$ decays, but it also makes use of other physics objects such as additional leptons
 344 (muons or electrons), hadronic jets, and b-tagged jets in order to tag the different Higgs boson
 345 production modes as well as to suppress certain backgrounds like $t\bar{t}$ and dibosons.

346 3.1 Primary vertex

347 Events are required to contain at least one well-identified primary vertex. The “good vertex”
 348 filter is applied selecting vertexes with the following properties:

- 349 • $|z| < 24$ mm
- 350 • $\rho < 2$ mm
- 351 • Number of degrees of freedom > 4

352 Among them, the one with the largest value of summed physics-object p_T^2 is taken to be the
 353 *primary interaction vertex*.

354 3.2 Muons

355 Events in this analysis are required to have at least two opposite charge muons with $p_T > 20$ GeV²,
 356 $|\eta| < 2.4$, and passing the medium identification (ID) criteria defined by the MUO POG [44].
 357 Each of the muons is required to pass the loose isolation requirement [45] in which the particle
 358 flow (PF) [46] based isolation sum of the p_T of charged and neutral PF hadrons, and PF pho-
 359 tons, lying in a $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ cone of 0.4 around the muon, is required to be less than
 360 25% of the p_T of the muon. The charged hadrons included in this sum are required to originate
 361 from the primary interaction vertex. If a PF photon in the isolation cone is associated with the
 362 final state radiation (FSR) of the muon (see Section 3.3), then it is not included in the isolation
 363 sum. Furthermore, the $\Delta\beta$ -correction is applied to the isolation sum which subtracts half of the
 364 p_T sum of charged hadron candidates, within the isolation cone, that belong to vertices other
 365 than the primary interaction one. This correction aims to deconvolve from the isolation calcu-
 366 lation the contamination due to neutral particles produced by interactions other than the hard
 367 scattering (primary interaction vertex). At this stage, the p_T of each muons passing the afore-
 368 mentioned identification strategy is corrected by the so-called *Rochester corrections*, described
 369 in more details in Section 4.6. Furthermore, at least one of the two muons is required to be
 370 geometrically matched to the level-3 muon candidate corresponding to a single muon trigger
 371 that was fired in the event. This muon is required to have $p_T > 29(26)$ GeV for 2017 (2016, 2018)
 372 eras to ensure that it lies on the single muon trigger plateau.

373 These thresholds are imposed to the p_T of identified muons before final-state-radiation (FSR)
 374 recovery, so that photons produced by final state radiation do not change (bias) the acceptance.

375 3.3 Muon FSR recovery

376 The sensitivity of this search depends critically on the resolution of the $H \rightarrow \mu\mu$ mass peak. In
 377 a small fraction of signal events, a muon in the final state may radiate a photon, thereby losing
 378 some of its momentum. This causes a slight degradation in the resolution of the signal peak.
 379 A procedure has been employed that attempts to recover this FSR by looking for PF photons
 380 in the vicinity of the muon. This strategy is based on the one developed in the context of the
 381 $H \rightarrow ZZ$ searches [47] with some modifications tuned specifically for the $H \rightarrow \mu\mu$ search, which

²This pre-selection is applied on the p_T defined at the “RECO” level i.e. as provided in *slimmedMuons* collection

are described in more details in the following slides [48]. The procedure can be summarized as follows:

- Consider all reconstructed muons with $p_T > 20 \text{ GeV}$ and $|\eta| < 2.4$ as candidates for FSR recovery.
- For a given muon, consider all PF photons with $0.0001 < \Delta R(\mu, \gamma) < 0.5$, $p_T > 2 \text{ GeV}$, $0.0 < |\eta| < 1.4442$ OR $1.566 < |\eta| < 2.5$ as possible FSR candidates.
- Ignore PF photons that are associated with a reconstructed electron, i.e. belong to the list of photons associated with the bremsstrahlung of the electron.
- Only consider a photon with PF isolation sum, without pileup correction and computed in a cone of $\Delta R < 0.3$, less than 180% of its p_T as an FSR candidate (I_γ / p_T^μ).
- Require $\Delta R(\mu, \gamma) / p_T^2 < 0.012$.
- In order to strongly suppress the contamination from $H \rightarrow Z\gamma \rightarrow \mu\mu\gamma$ decays, FSR photon candidates are required to have $p_T^\gamma / p_T^\mu < 0.4$. The residual contamination of $H \rightarrow Z\gamma \rightarrow \mu\mu\gamma$ events in the $H \rightarrow \mu\mu$ analysis is estimated to be around 0.1–0.2% depending on the era.
- In case of multiple FSR photon candidates, only the one with the smallest value of $\Delta R(\mu, \gamma) / p_T^2$ (p_T is that of the photon) is considered.

If an FSR photon is associated with a muon, its momentum is added to that of the original muon. The FSR-corrected momentum is promoted to be the new momentum of the muon. Furthermore, as already described in Section 3.2, this photon is ignored from the isolation sum of the muon when imposing the loose isolation requirement.

As described above, the FSR strategy needs to maximize the recovery of photons radiated from high p_T muons and, at the same time, minimize the contamination from Higgs boson decays to $Z\gamma$, where a prompt high p_T photon is produced in association with a Z boson decaying into a muon pair. Therefore, when FSR photons in SM Higgs decays to muons are used to reconstruct the Higgs boson mass, $H \rightarrow Z(\mu\mu)\gamma$ events are considered as a resonant background. In order to reduce this background component three kinematic observables, which offer a very high signal efficiency and strong rejection of $H \rightarrow Z(\mu\mu)\gamma$ events, are tested: p_T^γ / p_T^μ , p_T^γ , and $\min - \Delta R(\mu, \gamma)$. These studies are performed on simulated $ggH \rightarrow \mu\mu$ (signal) and $ggH \rightarrow Z(\mu\mu)\gamma$ (background) events. Reconstructed muons and photons are matched to the corresponding generator level particles to make sure that, for signal events, the kinematics of FSR photons radiated by a muon and, for background events, the kinematics of a three-body decay are correctly represented. The variable showing the best rejection power with a simple rectangular cut, expressed as $\epsilon_{\text{sig}} \times (1 - \epsilon_{\text{bkg}})$, is found to be p_T^γ / p_T^μ . Figure 3.1 shows the evolution of $\epsilon_{\text{sig}} \times (1 - \epsilon_{\text{bkg}})$ (black line) as a function of p_T^γ / p_T^μ . The maximum significance is obtained by requiring $p_T^\gamma / p_T^\mu < 0.4$.

In addition, the thresholds applied on the relative photon isolation (I_γ / p_T^μ) and on the ratio $\Delta R(\mu, \gamma) / p_T^2$ have been optimized in order to improve the analysis sensitivity. After an inclusive dimuon selection, the optimization is obtained by performing a rectangular 2D scan in the $I_\gamma / p_T^\mu - \Delta R(\mu, \gamma) / p_T^2$ plane. For each pair of cuts, $\sum_i^n S_i^2 / B_i$ is used as a figure of merit where S_i and B_i are the number of expected signal and background events in the i -th mass bin from 120 to 130 GeV with $n = 10$. Signal is represented by $ggH \rightarrow \mu\mu$ events, while background events are obtained by summing the expected contributions from $Z/\gamma^* + \text{jets}$, $t\bar{t}$, and Dibosons. Figure 3.2 shows the evolution of the signal significance (z-axis) in the $I_\gamma / p_T^\mu - \Delta R(\mu, \gamma) / p_T^2$ plane, along with optimal working points (rectangles): blue for 2016, black for 2017, and light

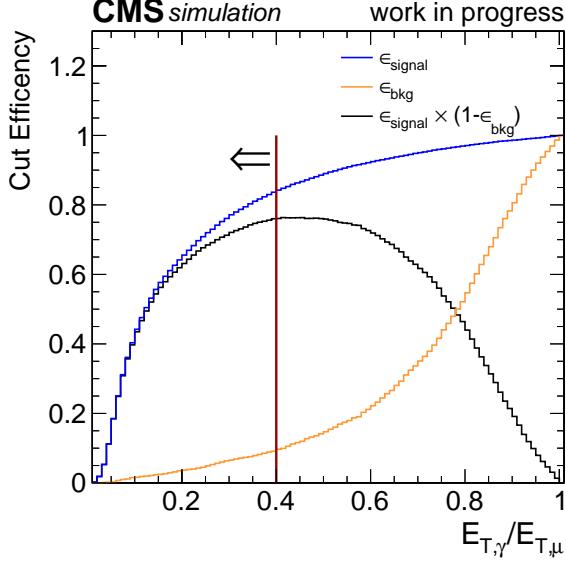


Figure 3.1: Signal ($\text{ggH} \rightarrow \mu\mu$, blue solid line) and background ($H \rightarrow Z(\mu\mu)\gamma$, orange solid line) efficiency trends as a function of p_T^γ/p_T^μ in simulated Higgs boson events in which an FSR photon candidate is found close to one of the two identified muons. The signal significance, defined as $\epsilon_{\text{sig}} \times (1 - \epsilon_{\text{bkg}})$, is reported in black. The signal significance is maximized by requiring $p_T^\gamma/p_T^\mu < 0.4$.

427 blue for 2018 era. The red rectangle, instead, indicates the best working point based on full
 428 Run2 data (sum of 2016, 2017 and 2018 predictions) and corresponds to the following selection:
 429 $I_\gamma/p_T^\mu < 1.8$ and $\Delta R(\mu, \gamma)/p_T^2 < 0.012$.

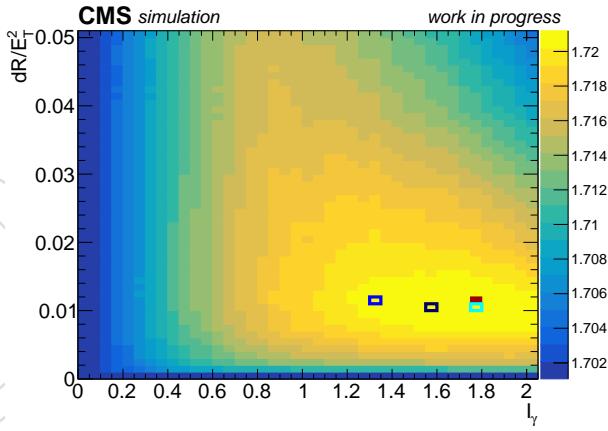


Figure 3.2: Signal significance, defined as S/\sqrt{B} , obtained by performing a rectangular 2D scan in the $I_\gamma/p_T^\mu - \Delta R(\mu, \gamma)/p_T^2$ plane on simulated events in which there is at least one FSR photon candidate. The maximum significance obtained on the basis of full Run2 data is obtained by requiring $I_\gamma/p_T^\mu < 1.8$ and $\Delta R(\mu, \gamma)/p_T^2 < 0.012$ and is indicated by the red solid rectangle. The blue boxes show the best cut values for the different data-taking periods 2016 (dark blue), 2017 (blue), and 2018 (light blue).

430 Finally, in order to check the effect of the FSR recovery strategy on the signal lineshape, sim-
 431 ulated $\text{ggH} \rightarrow \mu\mu$ events are divided into two groups based on the presence of a FSR photon
 432 at the generator level with $\Delta R(\mu, \gamma) < 0.3$. In Fig. 3.3 (left), the signal lineshape at the recov-
 433 ered level in simulated $\text{ggH} \rightarrow \mu\mu$ events, modelled with a double-sided Crystal-Ball function, is re-
 434 ported in red. Signal events without FSR (blue line) exhibits a symmetric peak around 125 GeV,
 435 while events with a FSR photon prior to recovery shows a broad $m_{\mu\mu}$ distribution (green line).
 436 Fig. 3.3 (right) shows that the FSR recovery procedure, in events with one photon candidate at

437 the generator level with $\Delta R(\mu, \gamma) < 0.3$, allows to obtain a signal lineshape similar to the one
 438 characterizing events without FSR.

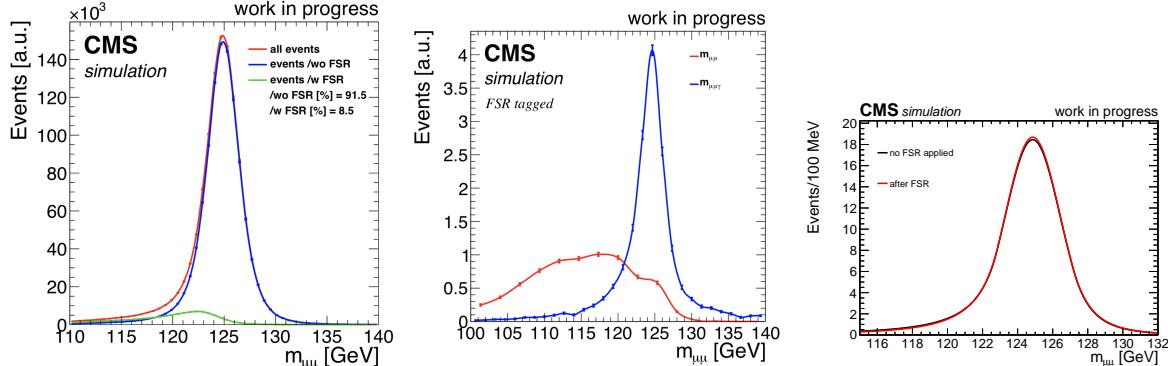


Figure 3.3: Impact of FSR photon recovery on signal events. Left: signal divided into events with at least one FSR photon radiation (green) and without (blue) at the generator level. Right: reconstructed dimuon mass of events with at least one FSR photon close to a muon before (red) and after (blue) including the FSR photon into the dimuon mass computation. Bottom: Impact of FSR photon recovery on ggH events after an inclusive dimuon selection. The recovery increases the expected yield by 2% and improves the resolution by 0.5%.

439 3.4 Electrons

440 Electrons used in the analysis are taken from the `slimmedElectrons` collection contained
 441 in `MiniAOD`, and are required to have $p_T > 20 \text{ GeV}$, $|\eta| < 2.5$. Furthermore, electrons are
 442 identified using the WP90 MVA based ID developed by the EGamma POG [49], which com-
 443 bines GSF-track, ECAL super-cluster, and isolation information into a single discriminator via
 444 a Boosted Decision Tree (BDT).

445 3.5 Jets

446 Jets used in the analysis are reconstructed by clustering the PF candidates in the event using
 447 the anti- k_t algorithm [50] with a distance parameters of 0.4. While clustering the jets, only the
 448 charged PF candidates that are associated with the primary interaction vertex are considered
 449 (charged hadron subtraction). Jet momentum is determined as the vector sum of all particle
 450 momenta inside the jet, and is found from simulation to vary, on average, between 5 and 10%
 451 of the true momentum over the whole p_T spectrum and detector acceptance. An offset cor-
 452 rection (L1FastJet correction) is applied to jet energies in data and MC to take into account
 453 the contribution from additional pp interactions within the same or adjacent bunch crossings
 454 (pileup) [51]. Additional energy corrections are derived from simulation and applied as a func-
 455 tion of jet η and p_T (L2, L3 corrections), again, both in data and MC. Finally, residual corrections
 456 are measured *in situ* from the energy balance in dijet, multijet, γ +jets, and leptonically decay-
 457 ing Z+jets events [52]. These are applied to jets in data in order to match the jet response with
 458 the simulation [53]. The levels of the applied corrections are summarized in Tab. 3.1. The jet-
 459 energy-correction tags used in the analysis are: `Summer16_07Aug2017_V11` for the 2016 era,
 460 `Fall117_17Nov2017_V32` for the 2017 era, and `Autumn18_V19` for 2018.

Data	L1FastJet, L2Relative, L3Absolute, L2L3Residual
Monte Carlo	L1FastJet, L2Relative, L3Absolute

Table 3.1: Levels of jet energy corrections applied in data and MC.

461 Offline, jets are required to have $p_T > 25 \text{ GeV}$ and $|\eta| < 4.7$. Furthermore, jets are required to
 462 pass the loose pileup jet ID and the loose (tight) jet ID requirements as recommended by the

463 JME POG for the 2016 (2017, 2018) era. These requirements help to suppress the contamination
 464 from pileup jets and misreconstructed jets from detector noise in the fiducial region beyond the
 465 coverage of the tracker, respectively. Jets are also cleaned w.r.t. the selected muon candidates
 466 by requiring a geometrical separation of $\Delta R(j, \mu) > 0.4$. In order to mitigate the impact of jet
 467 misreconstruction due to ECAL endcap noise in 2017, jets that lie in the region $2.6 < |\eta| <$
 468 3.0 are required to pass the tight pileup jet ID w.p. in both data and MC. Detailed studies
 469 explaining how this selection has been tuned in order to reject spurious jets in data and, at
 470 the same time, ensure the minimal loss in performance (signal acceptance) are reported in the
 471 following set of slides presented in one JME (JMAR) meeting [54].

472 3.6 B-tagged jets

473 Jets arising from the hadronization of bottom quarks are identified, within the tracker fiducial
 474 region ($|\eta| < 2.5$), with the DeepCSV b-tagging algorithm developed by the BTV POG. Both
 475 the loose and the medium w.p. are used in the analysis [55–57] according to the following
 476 logic: events which contain at least one (two) jets passing the medium (loose) b-tagging w.p.
 477 are classified into the ttH enriched category. Simulated events are corrected in order to take
 478 into account differences between data and MC measured in the b-tagging efficiency for both
 479 bottom quark and light flavour jets.

480 3.7 Missing transverse momentum

481 The p_T^{miss} is computed by taking the negative vector sum of the transverse momenta of all the
 482 reconstructed PF candidates in a given event. Corrections to the momenta of jets reconstructed
 483 in the event are propagated to the p_T^{miss} (Type-1 correction). The analysis also employs all the
 484 p_T^{miss} filters recommended by the JME POG [58] to remove events with spurious p_T^{miss} due to
 485 misreconstruction, detector noise and non-collision background.

486 3.8 Prompt lepton identification

487 The separation between leptons originating from the decay of a massive particle (prompt lep-
 488 tons) from those produced in the decays of charm or bottom quarks (non-prompt) becomes one
 489 of the key features in the multilepton channels (ttH, WH, and ZH channels) to reduce certain
 490 backgrounds from top or DY processes. Therefore, a multivariate lepton identification method
 491 referred to as *LeptonMVA* is introduced. Several analyses within CMS have used the Lepton-
 492 MVA approach so far. For the analysis presented in this note, the version developed in context
 493 of the tZq analysis is used [59]. A BDT was trained on MC samples with 2016 and 2017 detec-
 494 tor conditions, separately for electrons and muons. Three working points which correspond to
 495 different values of the identification efficiency and purity are provided. For each of them, the
 496 selection efficiencies in data and simulation were measured independently for 2016 and 2017.
 497 The version of the training from 2017 and the corresponding selection efficiencies are used also
 498 for 2018. The medium working point offers the best performance and is applied to both elec-
 499 trons and muons requiring the leptonMVA score to be larger than 0.4 for each lepton selected
 500 in the VH and ttH channels.

501 The presence of a jet close to the lepton is an important feature to separate prompt leptons
 502 from fake leptons. For the matching of jets to leptons a maximum distance between the jet
 503 and the lepton of < 0.4 and a lower threshold on the transverse jet momentum p_T of 15 GeV
 504 are required. The p_T of the jet is computed by applying the *lepton-aware* jet energy corrections
 505 (JEC). Here the JEC are applied to the difference of the un-calibrated jet and the lepton, and
 506 subsequently adding the lepton momentum to the calibrated jet again. In case no matching

507 jet is found, the jet related observables are set to a default value. A complete list of input
 508 observables to the BDT is:

- 509 • $I_{\text{rel},\text{charged}}^{\text{mini}}$: The isolation of the lepton with respect to charged particles.
- 510 • $I_{\text{rel},\text{neutral}}^{\text{mini}}$: The isolation of the lepton with respect to neutral particles.
- 511 • p_T^{ratio} : The ratio of the lepton p_T to the p_T of the nearest jet.
- 512 • $\text{deepCSV}^{\text{closest}}$: The discriminant value of the deepCSV algorithm of the nearest jet.
- 513 • $N_{\text{tracks}}^{\text{jet}}$: The number of tracks of charged particles within the jet.
- 514 • p_T^{rel} : The component of the lepton momentum in direction transverse to the jet.
- 515 • d_{xy} and d_z : The transverse and longitudinal impact parameters of the lepton track
 516 with respect to the PV.
- 517 • SIP_{3D} : The three dimensional impact parameter of the lepton track with respect to
 518 the PV, divided by its uncertainty, corresponding to its significance.
- 519 • MVA_{ID} : The response of the BDT trained by the EGamma POG to separate electrons
 520 from jets (electrons only).
- 521 • $\text{segmentCompatibility}$: The compatibility of track segments in the muon system
 522 with the pattern expected for a minimum ionizing particle (Muons only).

523 The corresponding control plots can be found in Appendix E. Further validation is presented
 524 in [59].

525 3.9 Muon and electron IDs used in the ttH channel

526 In contrast to the ggH, VBF, and VH channels, the ttH analysis requires also an adapted base-
 527 line lepton selection. The identification and pre-selection of muons and electrons³ is performed
 528 using a cut-based approach. These requirements have been optimized in the context of the tZq
 529 analysis and are found to be more efficient than the standard muon and electron POG IDs also
 530 in the context of the ttH, $H \rightarrow \mu\mu$ search.

531 Leptons in signal events are expected to be isolated. The isolation of a lepton I^{mini} is defined as
 532 the sum of scalar p_T values of charged particles, neutral hadrons, and photons, that are recon-
 533 structed within a narrow cone centred around the lepton direction. To increase the efficiency of
 534 leptons reconstructed in events with slightly boosted kinematics or topologies of high hadronic
 535 event activity, the cone size R decreases dynamically with the p_T of the lepton.

$$R = \begin{cases} 0.05 & \text{if } p_T > 200 \text{ GeV} \\ 10 \text{ GeV}/p_T & \text{if } 50 < p_T < 200 \text{ GeV} \\ 0.2 & \text{if } p_T < 50 \text{ GeV} \end{cases}$$

536 Residual PU contributions to the neutral component of the isolation sum are taken into account
 537 using effective area corrections

$$I^{\text{mini}} = \sum_{\text{charged}} p_T + \max\left(0, \sum_{\text{neutrals}} p_T - \rho A \left(\frac{R}{0.3}\right)^2\right)$$

538 where ρ is the energy density of neutral particles reconstructed within the geometric acceptance

³final state τ_{had} leptons are not considered

539 of the tracking detectors (`fixedGridRhoFastjetCentralNeutral`) and \mathcal{A} represents the
540 effective areas which are listed for muons in Tab. 3.2 and electrons in Tab. 3.3.

Table 3.2: Muon effective areas

Range	2016	2017/2018
$0.0 < \eta \leq 0.8$	0.0735	0.0566
$0.8 < \eta \leq 1.3$	0.0619	0.0562
$1.3 < \eta \leq 2.0$	0.0465	0.0363
$2.0 < \eta \leq 2.2$	0.0433	0.0119
$2.2 < \eta \leq 2.4$	0.0577	0.0064

Table 3.3: Electron effective areas

Range	2016	2017/2018
$0.0 < \eta \leq 1.0$	0.1703	0.1440
$1.0 < \eta \leq 1.479$	0.1715	0.1562
$1.479 < \eta \leq 2.0$	0.1213	0.1032
$2.0 < \eta \leq 2.2$	0.1230	0.0859
$2.2 < \eta \leq 2.3$	0.1635	0.1116
$2.3 < \eta \leq 2.4$	0.1937	0.1321

541 The criteria employed in the ttH analysis to select events are summarized in Table 3.4.

Table 3.4: Full overview of the electron and muon identification requirements. The cuts in parentheses are applied according to the $|\eta|$ of the electrons. The first value for $0 < |\eta| < 1.479$ and the second for $1.479 < |\eta| < 2.5$. The ‘standard’ muon ID used in the ggH and VBF channel is given for reference.

Muon requirements	Selection	ggH/VBF	ttH
p_T	$> 20\text{ GeV}$	✓	✓
$ \eta $	< 2.4	✓	✓
POG ID	Medium	✓	✓
I_{rel}	< 0.25	✓	-
$I_{\text{rel}}^{\text{mini}}$	< 0.4	-	✓
d_{xy}	< 0.05	-	✓
d_z	< 0.1	-	✓
SIP _{3D}	< 8	-	✓
deepCSV ^{closest}	$< 0.8958_{(2016)} / < 0.8001_{(2017,2018)}$	-	✓

Electron requirements	Selection	ttH
p_T	$> 20\text{ GeV}$	✓
$ \eta $	< 2.5	✓
Ecal Gap Veto	$1.4442 < \eta < 1.566$	✓
$I_{\text{rel}}^{\text{mini}}$	< 0.4	✓
d_{xy}	< 0.05	✓
d_z	< 0.1	✓
SIP _{3D}	< 8	✓
Number of Missing Hits	< 2	✓
$\sigma_{i\eta i\eta}$	$< (0.011, 0.030)$	✓
H/E	$< (0.10, 0.07)$	✓
$\Delta\eta_{in}$	$< (0.010, 0.008)$	✓
$\Delta\phi_{in}$	$< (0.04, 0.07)$	✓
$-0.05 < 1/E - 1/p$	$< (0.010, 0.005)$	✓
deepCSV ^{closest}	$< 0.8958_{(2016)} / < 0.8001_{(2017,2018)}$	✓

542 3.10 Muon and electron IDs used in the VH channels

543 In the WH and the ZH channels, the selection criteria for muons and electrons need to be
544 aligned with those in the ggH and the VBF channels to ensure the orthogonality on the event
545 selection. Thus in the VH channels, the PF loose isolation is used to select muons, and the MVA
546 WP90 ID is used to select electron. The leptonMVA is used in the VH channels to reduce the
547 non-prompt backgrounds. However, some selection cuts from the tZq analysis (see Table 3.4)

548 are not applied, for example the lepton I^{\min} and the ECAL super-cluster variables for electrons.
 549 These cuts serve the same purpose as the standard PF loose isolation for muons and the MVA
 550 WP90 ID for electrons, and they yield similar efficiency in the VH phasespace.

551 Details of the lepton selection criteria in the VH channels are listed in Table 3.5 and 3.6.

Table 3.5: Muon selection requirements in VH

Variable	ggH and VBF	WH 3 ℓ and ZH 4 ℓ
Muon type	Tracker or global	Tracker or global
p_T	$> 20 \text{ GeV}$	$> 20 \text{ GeV}$
$ \eta $	< 2.4	< 2.4
POG ID	Medium	Medium
PFIso0.4	< 0.25	< 0.25
$d_{xy}(\text{PV})$	-	$< 0.05 \text{ cm}$
$d_z(\text{PV})$	-	$< 0.10 \text{ cm}$
SIP	-	< 8.0
lepMVA	-	> 0.4

Table 3.6: Electron selection requirements in VH

Variable	ggH and VBF	WH3 ℓ and ZH4 ℓ
p_T	$> 20 \text{ GeV}$	$> 20 \text{ GeV}$
$ \eta $	< 2.5	< 2.5
POG MVA ID	wp90	wp90
Ecal gap veto	(1.444, 1.566)	(1.444, 1.566)
$d_{xy}(\text{PV})$	-	$< 0.05 \text{ cm}$
$d_z(\text{PV})$	-	$< 0.10 \text{ cm}$
SIP	-	< 8.0
Conversion Veto	-	✓
Number of Missing Hits	-	< 2
lepMVA	-	> 0.4

552 3.11 Additional “soft” hadronic activity

553 In VBF production, a suppression of jet activity is expected in the pseudorapidity region be-
 554 tween the two jets Ref. [35]. Since the amount of additional radiation is expected to be small
 555 (soft) pile-up contribution become very relevant in additional hadron energy. Therefore the
 556 contribution of pile-up is simply avoided by considering only use of charged tracks that clearly
 557 originate from the event main interaction point to monitor the additional radiation.

558 The additional soft activity is defined as follows. At first a collection of “Additional tracks” is
 559 built using reconstructed tracks that

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

566 After this track selection a collection of “soft track-jets” is build clustering the “Additional
 567 tracks” collection with the anti- k_T clustering algorithm [50] with distance parameter $R = 0.4$.

- 568 The use of track-jets represents a clean and commissioned method [60] to reconstruct the had-
 569 ronization of partons with very low energies, down to few GeV [61].
- 570 For the purpose of separating the signal from backgrounds, clustered soft track-jets are consid-
 571 ered, and eventually consider
- 572 • the scalar p_T sum of the soft track-jets with transverse momentum $p_T > 1$ GeV, H_T^{soft} ;
- 573 • the soft track-jet multiplicity N_1^{soft} with transverse momentum $p_T > 2$ GeV, N_2^{soft} ;
- 574 • the soft track-jet multiplicity N_5^{soft} with transverse momentum $p_T > 5$ GeV, N_5^{soft} ;
- 575 • the soft track-jet multiplicity N_{10}^{soft} with transverse momentum $p_T > 10$ GeV, N_{10}^{soft} ;

576 4 Corrections to data and simulation

577 In this analysis we apply corrections to several physical observables in both data and simula-
 578 tion. This is essential to correct for certain biases in detector response, mismodeling of physical
 579 observables in simulation, and to ensure that the simulation correctly models the key features
 580 of the signal and background processes. In most of cases, the background in this search is de-
 581 termined in an entirely data-driven manner. However, we do categorize events as a function of
 582 a multivariate discriminant that is trained on simulated signal and background events. In the
 583 VBF-template analysis, the background model fully relies on the prediction from simulation
 584 that can be adjusted by means of nuisance parameters which are estimated via a simultane-
 585 ous fit performed across the Higgs signal region and the Z+jets enriched control sample. We
 586 also rely entirely on simulation to predict all features of our expected signal. Therefore, it is
 587 desirable to ensure that our signal and background processes are faithfully modelled by the
 588 simulation as much as possible.

589 4.1 Pileup reweighting

590 The simulation is produced with a pileup profile that is approximately similar to that mea-
 591 sured in data. However, a reweighting is performed to ensure that the profile of the pileup
 592 interactions in simulation matches exactly with the estimated pileup distribution integrated
 593 over the data taking period. The LUM POG provides the luminosity information for each certi-
 594 fied lumi-section [62] which is then multiplied by the estimated minimum bias cross section of
 595 $69.2 \text{ mb} \pm 5\%$ to obtain the pileup profile in data. In the case of the simulation, the true num-
 596 ber of in-time pileup interactions are known, and these are used to obtain the pileup profile
 597 with which the simulated sample was produced. The ratio of the data and simulation profiles
 598 gives the weight to be applied to a simulated event as a function of the true number of in-time
 599 interactions. These pileup weights are derived separately for the 2016, 2017 and 2018 eras.

600 The number of reconstructed primary vertices in a collision event is directly correlated with
 601 the pileup, and therefore, the distribution of the number of reconstructed vertices is indicative
 602 of the amount of pileup in a certain data taking period. Figure 4.1 shows the distribution of
 603 the number of vertices in the signal fit region (SR) in data and simulation, shown before (top
 604 row) after (bottom row) pileup reweighting for the 2016, 2017 and 2018 eras. These are dimuon
 605 events passing the trigger and muon selection requirements in which the dimuon mass falls
 606 between 110–150 GeV. Pileup reweighting improves visibly the agreement only in the 2016 era.

607 However, the remaining disagreements between data and simulation observed in the distribu-
 608 tion of number of vertexes is expected to have a negligible impact in analysis. Therefore, no
 609 additional corrections, like re-weight the N_{vtx} in MC to match the one observed in data, are
 610 applied. In fact, all the experimental corrections for muons, electrons, jets, and b-jets have been

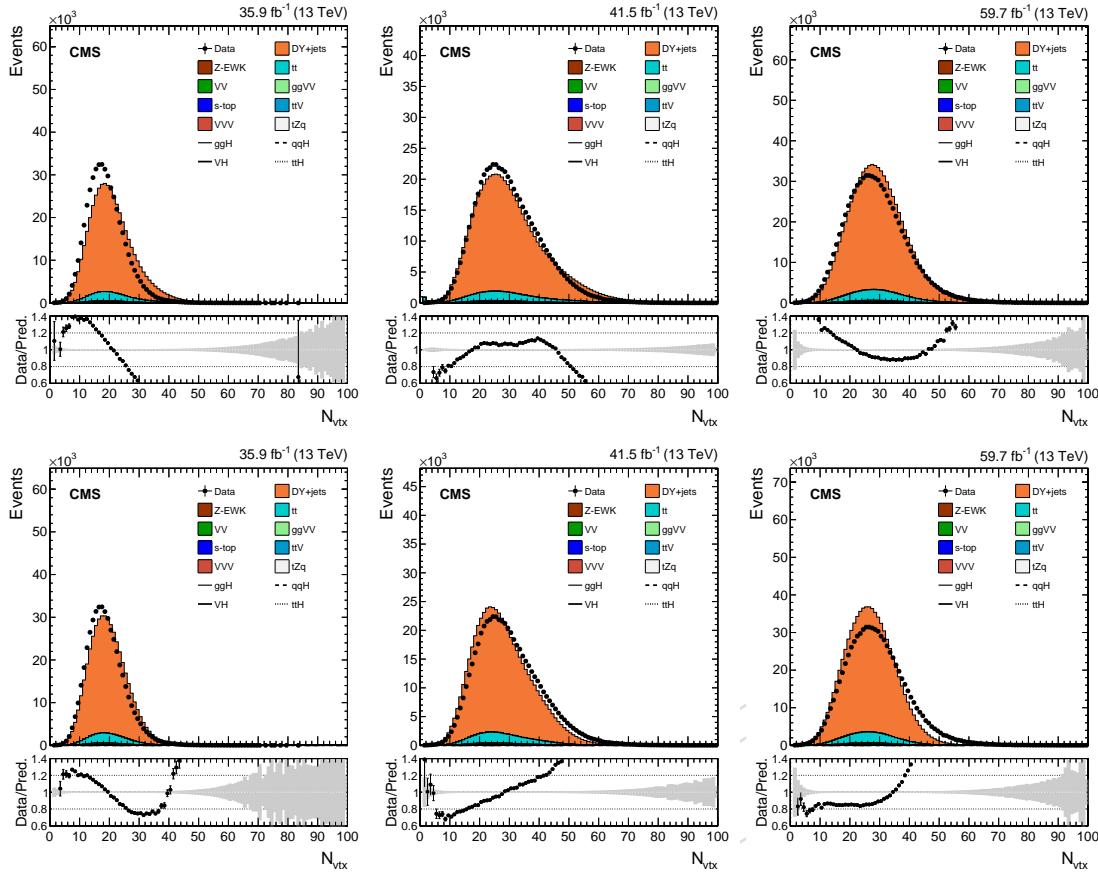


Figure 4.1: Distribution of the number of reconstructed primary vertexes in 2016 (left), 2017 (middle) and 2018 (right) eras reported for both data (black points) and simulation (solid histograms) before (first row) and after (second row) the pileup reweighting for events with $m_{\mu\mu}$ between 110 and 150 GeV. The pileup reweighting improves the agreement only in the 2016 era.

611 derived applying only the official pileup reweight, hence they should already take into account
 612 the effect of residual differences in pileup between data and simulation. Moreover, observables
 613 more prone to the pileup model like p_T^{miss} are not used for the event classification.

614 4.2 B-tagging scale factors

615 The probability distribution for a jet to come from the hadronisation of a bottom quark is de-
 616 rived in CMS using supervised machine learning trained on MC simulation. The b-tagging
 617 efficiency is calculated for data and MC in bins by jet p_T , η , in event samples enriched either in
 618 jets from b hadrons or in “light flavour” jets (light quarks and gluons). Different methods are
 619 provided by BTV POG to correct the difference in efficiency between MC and data, as described
 620 in Ref. [63], however a very good closure is observed between these alternative techniques.

621 We use the method that calibrates the b-tagging discriminant shape by applying event weights
 622 to simulated events, following the recipe described in Ref. [64]. Jets used for scale factor ap-
 623 plication are required to pass all selections listed in Section 3.5, and additionally to be within
 624 tracker acceptance region $|\eta| < 2.5$. Scale factors are extracted from the lookup tables provided
 625 by BTV POG [55–57]:

- 626 • 2016: DeepCSV_2016LegacySF_V1
- 627 • 2017: DeepCSV_94XSF_V5_B_F

628 • 2018: DeepCSV_102XSF_V1

629 4.3 L1 EGamma prefireing corrections

630 The ECAL endcap time alignment of the detector drifted during the data-taking, reaching a
 631 sever condition in 2017 RunF. The L1 trigger system of CMS has been designed to forbid trig-
 632 gering events in consecutive bunch crossings (trigger rules). The trigger primitives generated
 633 by the ECAL deposits and reconstructed as belonging to the previous bunch crossings are there-
 634 fore able to generate an inefficiency in the L1 trigger decision.

635 For what concern the $H \rightarrow \mu\mu$ search, the main source of prefireing are events with forward
 636 jets used to tag the VBF production. The ggH, ttH, and VH enriched categories see a small
 637 efficiency loss due to prefireing, while the VBF dedicated ones have inefficiencies of the order of
 638 2–5% in the 2016 era, and 4–10% in 2017 while no significant loss was observed in the 2018 era.
 639 Prefireing inefficiency is measured using a special set of triggered events called “un-prefirable”
 640 events. This inefficiency has been measured by the JetMet POG as a function of both photon
 641 and jet p_T in the forward η region of endcap, $2.5 < |\eta| < 3.0$. Figure 4.2 shows the central maps
 642 of prefireing probabilities ($\mathcal{P}_{\text{prefire}}(j)$) as function of the pseudo-rapidity and the electromagnetic
 643 component of the jet transverse momentum.

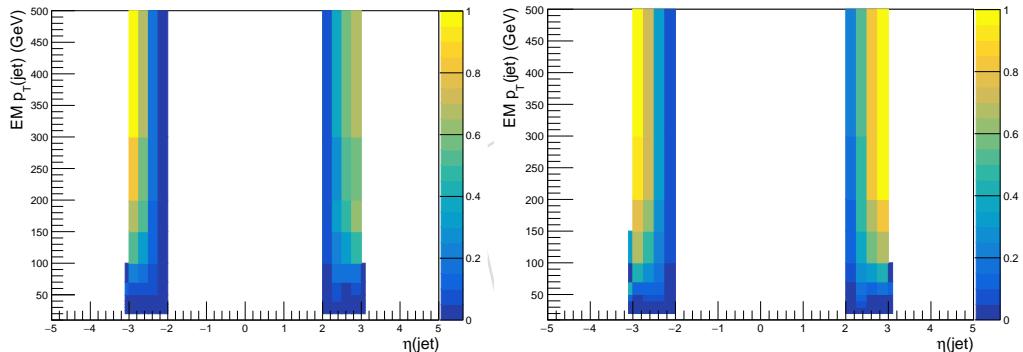


Figure 4.2: Centralized pre-firing probability provided per-jet as function of its electromagnetic trans-
 verse momentum and η for 2016 (left) and 2017 (right).

The event efficiency factor ($\epsilon_j = 1 - \mathcal{P}$) is applied to the MC simulation and it is accounted using all jets in the event, as:

$$\epsilon = \prod_{j \in \text{jets}} 1 - \mathcal{P}_{\text{prefire}}(j)$$

644 4.4 Muon efficiency and trigger scale factors

645 The MUO POG has measured the efficiencies, in both data and simulation, for the identification
 646 and isolation requirements used in this analysis for selecting muons. These efficiencies have
 647 been measured using a tag-and-probe technique [65] on Z+jets events in data and simulation,
 648 as a function of the muon p_T and η . These results are used to compute data-to-simulation scale
 649 factors that are then applied per muon, in simulated events, given its p_T and η . Therefore,
 650 the identification and isolation scale factors are applied as a logical AND (product) of all the
 651 selected muons.

652 The MUO POG also provides the efficiencies for a muon passing the baseline ID and isolation
 653 requirements to fire the single muon trigger. These are also computed as a function of the muon
 654 p_T and η . In a dimuon event, if the the per muon trigger efficiencies are ϵ_1 and ϵ_2 , then the per-
 655 event efficiency to fire the trigger is given by $\epsilon = 1 - (1 - \epsilon_1)(1 - \epsilon_2)$. These efficiencies can

be computed separately for data and for simulation, and their ratio gives the scale factor that is then applied as a weight to a simulated dimuon event.

4.5 Correction to Z boson transverse momentum

A reweighting of the dimuon p_T distribution of the Drell-Yan background simulation is performed. This reweighting is derived from data in the “Z control region” ($ZCR, 70 < m_{\mu\mu} < 110 \text{ GeV}$). All corrections described so far are applied to the simulated Drell-Yan events in the ZCR and their total yield is normalized to that observed in data. Figure 4.3 shows the dimuon p_T distribution in data and simulation for the 2016 (left), 2017 (middle) and 2018 (right) eras prior to any specific correction of the $Z/\gamma^* + \text{jets}$ p_T spectrum and prior to N_{jet} binning. The distributions in data and simulation have visible differences, particularly at low p_T . These can be attributed to missing resummation effects that are not modelled in simulation. A correction for this effect is derived as follows:

- In each era, events are divided into three groups according to the number of jets, $N_{\text{jets}} = 0, = 1, \geq 2$.
- The normalized data-to-simulation ratio in each jet bin is smoothed with a 12-th order polynomial in order to have a continuous correction as a function of the dimuon p_T up to 200 GeV.

Figure 4.4, 4.5, and 4.6 show the dimuon p_T distributions in the SR ($110 < m_{\mu\mu} < 150 \text{ GeV}$) for 2016, 2017, and 2018 eras, respectively, obtained after the p_T reweighting for events with $N_{\text{jets}} = 0$ (left), $N_{\text{jets}} = 1$ (middle), and $N_{\text{jets}} \geq 2$ (right). The re-weight allows to obtain a very good shape agreement between data and simulation in all eras and jet multiplicity.

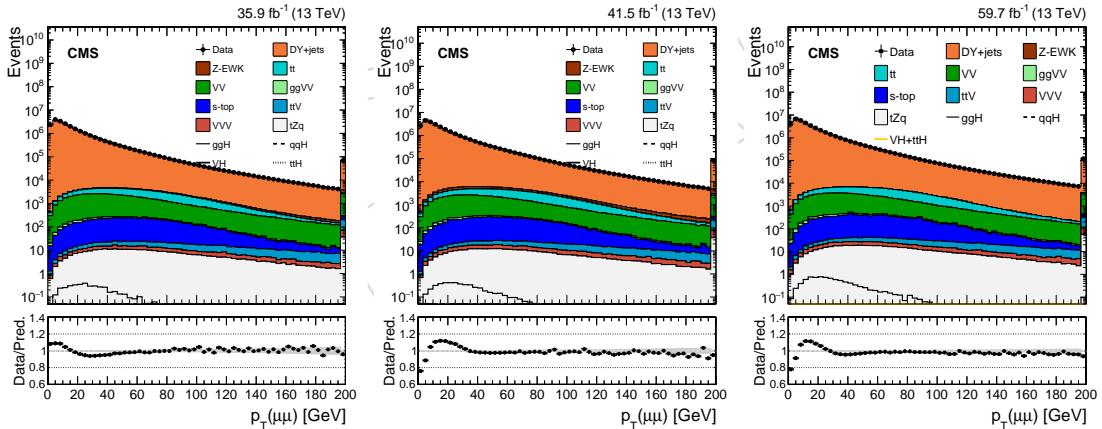


Figure 4.3: The dimuon p_T distribution in 2016 (left), 2017 (middle) and 2018 (right) eras reported for both data (black points) and simulation (solid histograms) for events with $m_{\mu\mu}$ between 70 and 110 GeV. A visible disagreement between data and MC is present for events with dimuon p_T smaller than 40 GeV. The different trend observed in the 2016 era compared to 2017 and 2018 is due to different parton shower and underlying event tunes used in the simulation as described in Section 2.1.

4.6 Muon momentum scale and resolution

There are differences in both the muon momentum scale and resolution between data and simulation. This causes a shift in the peak position of a dimuon resonance and a difference in its resolution between data and simulation. This is corrected by the *Rochester corrections*, which are applied to both data and simulation. The muon momentum scale is corrected in both data and simulation so that the Z boson peak matches the PDG value, while an additional smearing is applied to the muon momentum in the simulation so that the Z boson peak resolution in data

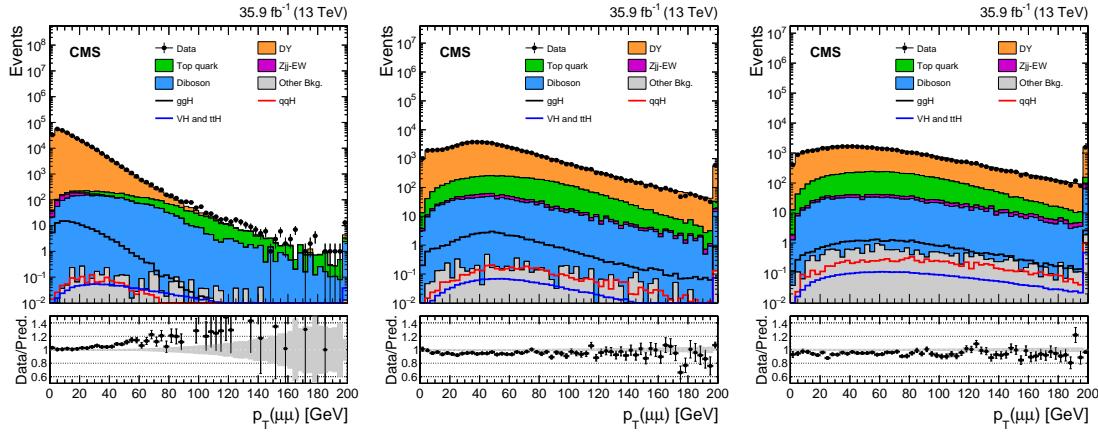


Figure 4.4: The dimuon p_T distribution in the 2016 era in events with $N_{\text{jets}} = 0$ (left), $N_{\text{jets}} = 1$ (middle), and $N_{\text{jets}} \geq 2$ reported for both data (black points) and simulation (solid histograms) for events in the signal fit region after the $Z/\gamma^* + \text{jets}$ p_T reweight. The agreement between data and simulation is significantly improved by the reweight defined in the ZCR.

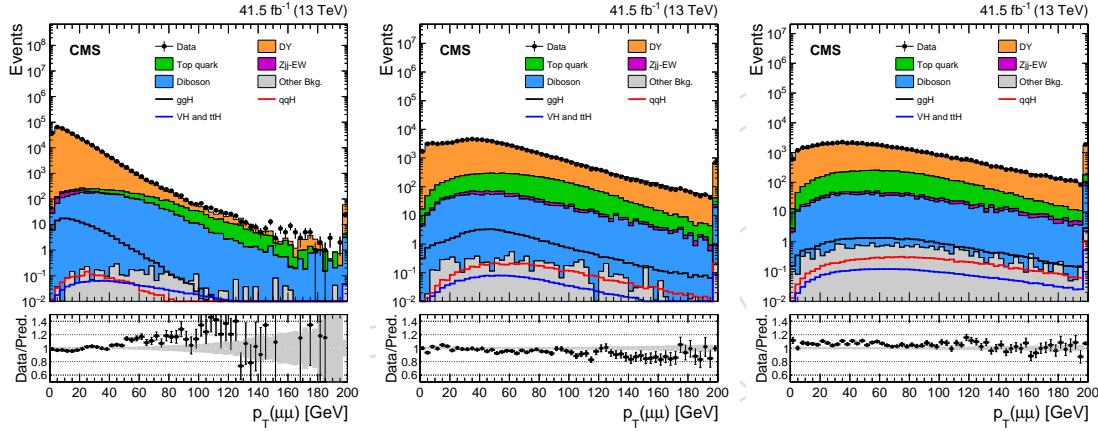


Figure 4.5: The dimuon p_T distribution in the 2017 era in events with $N_{\text{jets}} = 0$ (left), $N_{\text{jets}} = 1$ (middle), and $N_{\text{jets}} \geq 2$ reported for both data (black points) and simulation (solid histograms) for events in the signal fit region after the $Z/\gamma^* + \text{jets}$ p_T reweight. The agreement between data and simulation is significantly improved by the reweight defined in the ZCR.

and simulation agree. We use the Run2 v3 corrections as reported in Ref. [66]. Figure 4.7 shows the dimuon mass distributions for each era in the ZCR after applying the Rochester corrections.

In order to check in more details whether the *Rochester corrections* are effectively resolving the disagreements in scale and resolution between data and simulation, the Z boson peak in data and MC is fitted with analytical functions after applying the *Rochester corrections* in the mass range $75 < m_{\mu\mu} < 105$ GeV.

Dimuon events are categorized based on the η of each of the two muons. Three η categories are considered: barrel (B) with $|\eta| < 0.9$, overlap (O) with $0.9 < |\eta| < 1.8$, and endcap (E) $1.8 < |\eta| < 2.4$. This gives a total of nine possible η categories for a dimuon event. In each η -category, the dimuon mass distribution is fit with a functional form given by the convolution of a Breit-Wigner (Eq. 4.1) and a double-sided Crystal Ball (Eq. 4.2). The peak and width parameters of the Breit-Wigner function are fixed to the nominal mass and width of the Z boson, whereas the other parameters are allowed to float. Moreover, a background component is also included and is modelled with an exponential function. Figure 4.8 shows the fitted Z

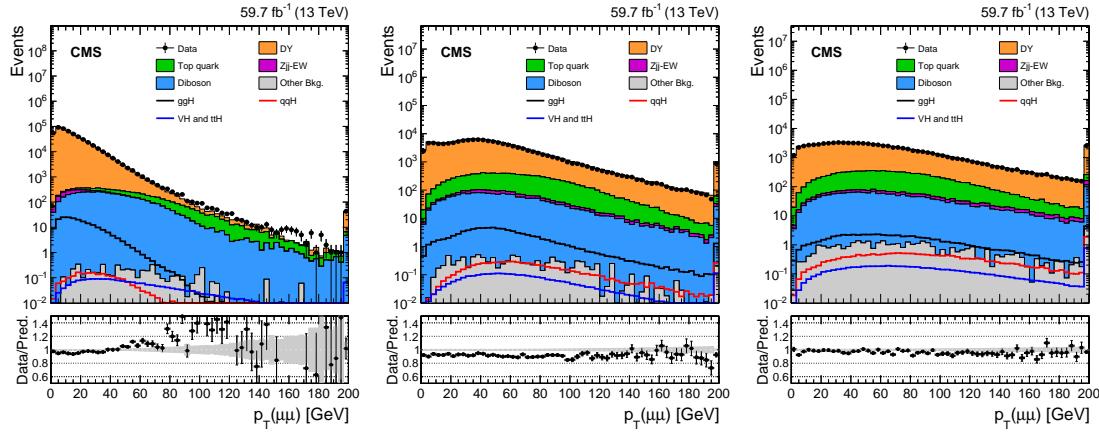


Figure 4.6: The dimuon p_T distribution in the 2018 era in events with $N_{\text{jets}} = 0$ (left), $N_{\text{jets}} = 1$ (middle), and $N_{\text{jets}} \geq 2$ reported for both data (black points) and simulation (solid histograms) for events in the signal fit region after the $Z/\gamma^*+\text{jets}$ p_T reweight. The agreement between data and simulation is significantly improved by the reweight defined in the ZCR.

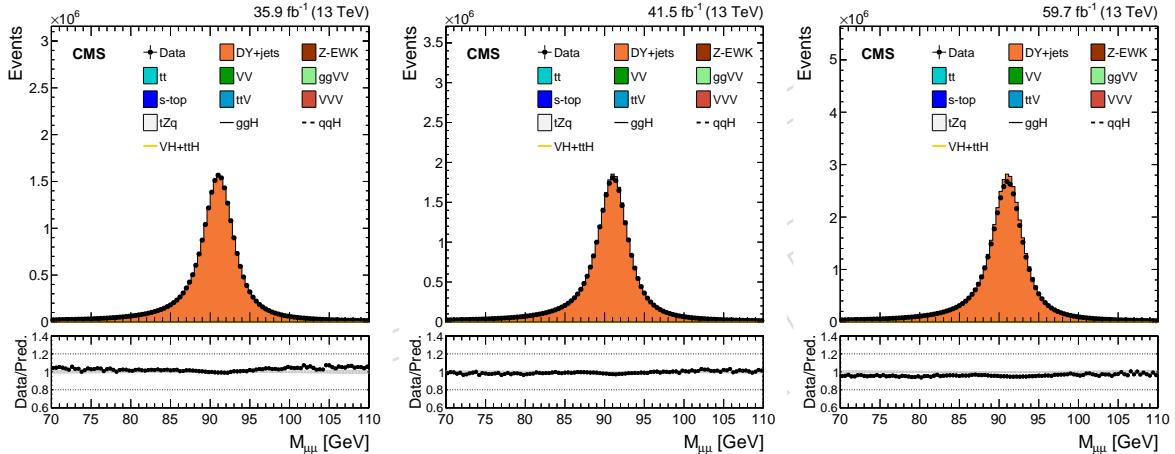


Figure 4.7: The dimuon mass distribution in 2016 (left), 2017 (middle) and 2018 (right) eras reported for both data (black points) and simulation (solid histograms) after the application of the Rochester corrections for events with $m_{\mu\mu}$ between 70 and 110 GeV. Besides a few percent shift in the overall normalization, the agreement in both peak position and resolution between data and simulation is good in all data taking periods.

698 peak values in data and MC for each of the nine η -categories for 2016 (left), 2017 (middle), and
 699 2018 (right) eras. Similarly, Fig. 4.9 reports the trends observed for the σ of the double-sided
 700 Crystal Ball, which is used as a proxy for the dimuon mass resolution. After the application of
 701 the *Rochester corrections*, the muon momentum scale, represented by the position of the Z boson
 702 peak, agrees within 0.1% between data and simulation. In contrast, the dimuon mass reso-
 703 lution in MC remains slightly optimistic, in particular for barrel muons in the 2016 era. Overall,
 704 the residual discrepancy between data and MC (not covered by the *Rochester energy smearing*)
 705 remains smaller than 5% across η -categories and eras. The maximum difference observed be-
 706 tween data and simulation is covered by the mass resolution uncertainty assigned in the final
 707 fit, as described later in Section 8.3.

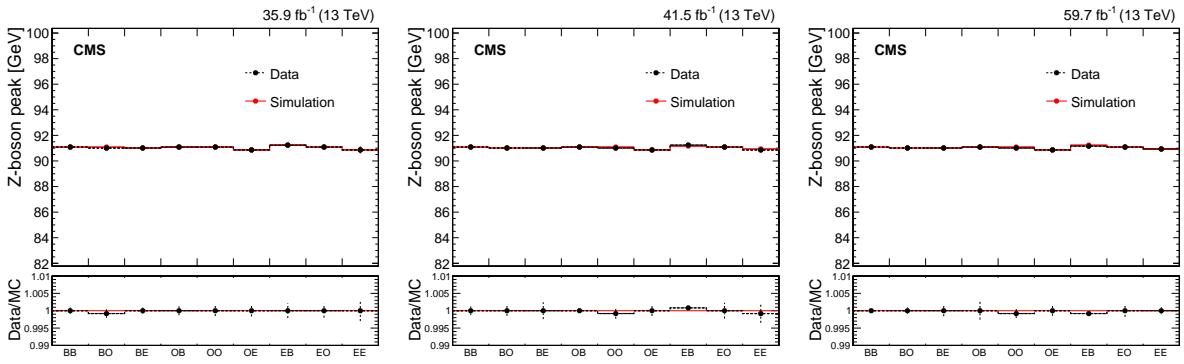


Figure 4.8: Z boson peak position extracted from an analytical fit to the dimuon mass distribution of events categorized in nine exclusive η -bins for 2016 (left), 2017 (middle), and 2018 (right) eras. Values are extracted independently for data and simulated events.

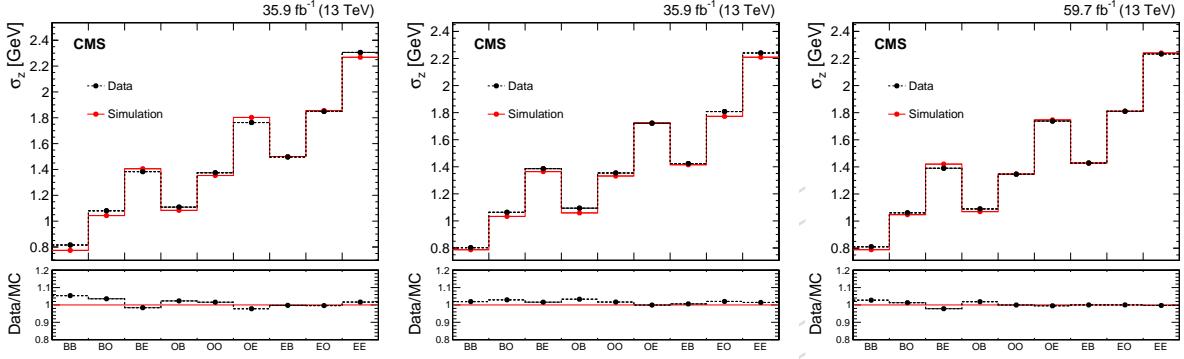


Figure 4.9: Z boson peak resolution extracted from an analytical fit to the dimuon mass distribution of events categorized in nine exclusive η -bins for 2016 (left), 2017 (middle), and 2018 (right) eras. Values are extracted independently for data and simulated events.

708 4.7 GeoFit correction to muon p_T

709 In CMS, the reconstructed particle tracks can be displaced from the primary pp interaction vertex. This displacement can be measured by transverse and longitudinal impact parameters
710 denoted by d_0 and z_0 respectively. These parameters are used in differentiating particle tracks
711 originating from the primary vertex (*prompt*) from the tracks originating from a secondary ver-
712 tex such as in b-quark decays. Muon selections that are used in this analysis guarantee that the
713 d_0 values of the muons are close to zero. However, it is possible to have small but non-zero
714 d_0 values. The effect of this residual d_0 is observable as a trend in reconstructed p_T values of
715 these muons, which in turn leads to a increase in reconstructed dimuon invariant mass res-
716 olution. This effect is corrected in both data and MC by *GeoFit Corrections* that are derived in
717 Z+jets MC samples, which uses muon d_0 values calculated by using the beam spot information
718 ($d_{0\text{BS}}$). Figure 4.10 shows the trend in reconstructed dimuon mass when plotted against $d_{0\text{BS},\mu^+}$
719 before and after the corrections. The corrections improve the dimuon mass resolution in both
720 data and MC without introducing an overall mass scale shift or any other bias. Figure 4.11
721 shows dimuon mass distributions for ggH signal MC before and after the *GeoFit Corrections*.
722 The details of *GeoFit Corrections* can be found in Section D.

724 The GeoFit corrections provide a simple method (first order approximation) to include the as-
725 sumption that muon tracks originated from the collision vertex. Including the beamspot con-
726 straint in the muon track fit provides, from first principles, a more precise correction. The latter
727 method is tested using a ggH 2018 sample. These corrections are derived by modifying the co-

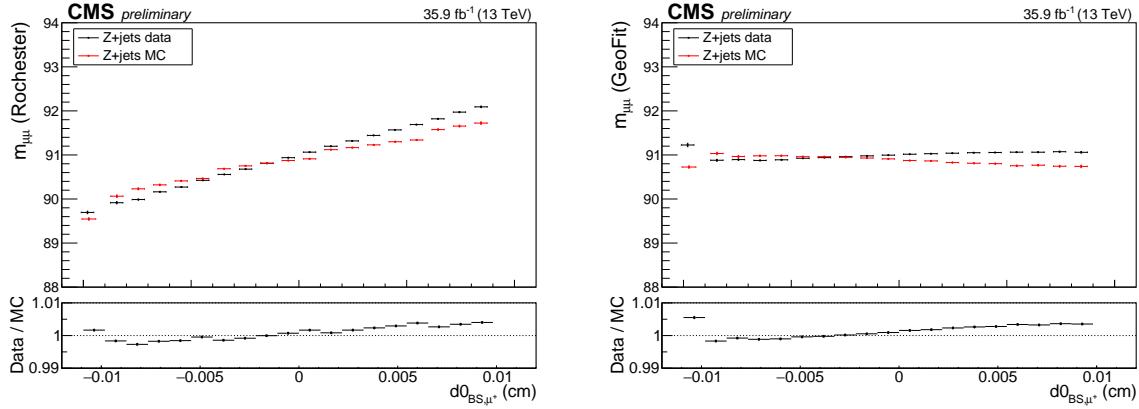


Figure 4.10: Dimuon mass peak mean value vs. $d0_{BS,\mu^+}$ before (left) and after (right) *GeoFit Corrections* showing in 2016 Z+jets data and MC.

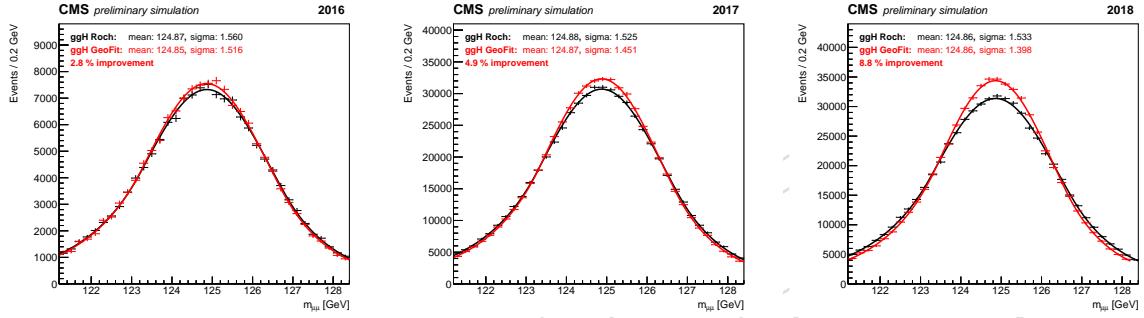


Figure 4.11: Dimuon mass peak around 125 GeV for ggH signal MC samples in 2016 (left), 2017 (center) and 2018 (right) fitted by using a double-sided crystal ball function.

728 variance matrix of the muons track adding the measured beamspot as an additional constrain.
 729 CMSW provides an interface for this operation via the `SingleTrackVertexConstraint`
 730 package. The results are shown in Fig. 4.12 and in Fig. 4.13. The corrections to the muon
 731 trasverse momentum are very similar to the ones provided by the GeoFit method. The invari-
 732 ant mass resolution improvement using this more sofisticated technique is equivalent to the
 733 one obtained using the GeoFit corrections. These results strongly suggest that the GeoFit cor-
 734 rections are a phenomenological equivalent way to the muon track refit with beamspot. Both
 735 techniques provide a gain in the invariant mass resolution by including the assumption that
 736 muon tracks generated at the collision vertex. GeoFit corrections come with a more analysis
 737 friendly interface, and are easier to share and implement by the various frameworks used in
 738 this analysis. For this reason they are the preferred in this analysis compared to the track refit
 739 with the beamspot. In the future (Run3), with some coordination among the analysis teams, it
 740 could be possible to include the muon refitting with beamspot in the various frameworks and
 741 use it instead of the GeoFit.

742 4.8 Event-by-event mass resolution

743 The sensitivity of the search depends on the resolution of the Higgs boson mass peak, which
 744 in turn depends on the muon momentum resolution. The muon p_T resolution varies primarily
 745 depending on the η of the muon. Muons that pass through the central barrel region of the
 746 detector ($|\eta| < 0.9$) have the best p_T resolution, whereas muons passing through the forward
 747 region ($|\eta| > 1.8$) tend to have the worst resolution. In order to optimize the sensitivity of the
 748 search it is important to separate out events with good dimuon mass resolution from those in

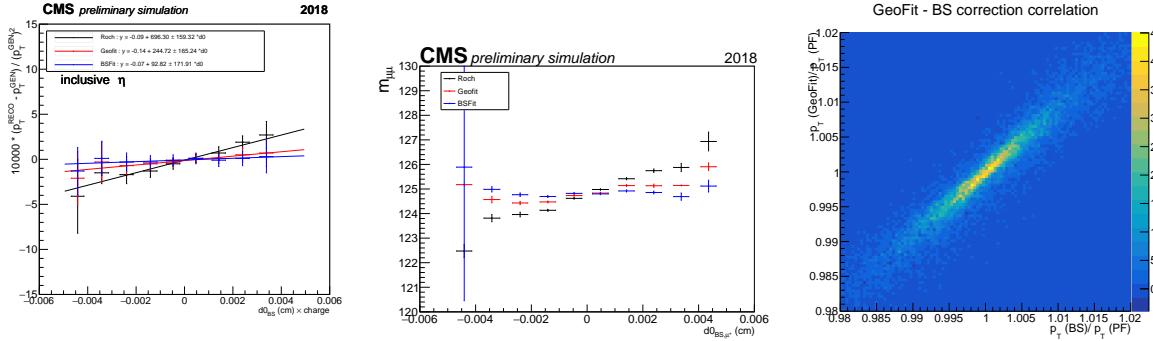


Figure 4.12: Muon p_T (left) and dimuon mass (center) profiled against the muon transverse impact parameters times the charge. Correlation between muon p_T correction from GeoFit and beamspot fit procedures (right). Data come from a ggH signal MC samples in 2018.

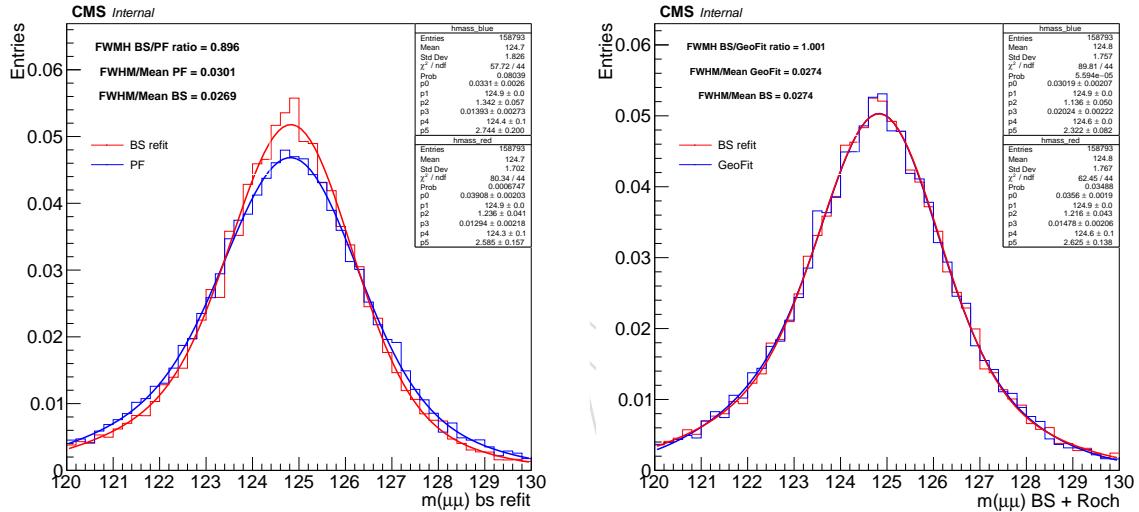


Figure 4.13: Dimuon mass peak around 125 GeV for ggH signal MC samples in 2018 fitted by using a triple gaussian function. Dimuon mass calculated with muons tracks refitted with beamspot constraint is compared with its particle flow (left) and GeoFit (right) estimate.

749 which the mass is measured less precisely.

750 In a given event, it is possible to estimate the uncertainty on the measured mass of the dimuon
751 system from the covariance matrix of the fitted track of the individual muons. Figure 4.14
752 shows the shape of the distribution of the relative dimuon mass uncertainty in ggH signal
753 events for different geometrical regions of the detector through which the muons traverse.
754 We can see from this figure that the mass resolution tends to get worse, on average, as the
755 maximum η of the muons increases. However, the mass resolution distribution for a given
756 geometrical configuration of the two muons has a certain spread, therefore it is more suitable
757 to make use of the per-event mass uncertainty information rather than categorizing events
758 based on the muon η , as was done in Ref. [13].

759 The Higgs boson mass peak is typically modelled via an analytical function that contains a
760 Gaussian core and some non-Gaussian tails. We assume that the σ of this Gaussian core de-
761 pends directly on the per-event mass resolution. The intrinsic width of the Higgs boson is
762 expected to be extremely small, and its contribution can be neglected. However, the mass
763 uncertainty obtained from the muon track fit covariance matrix is still an approximation and

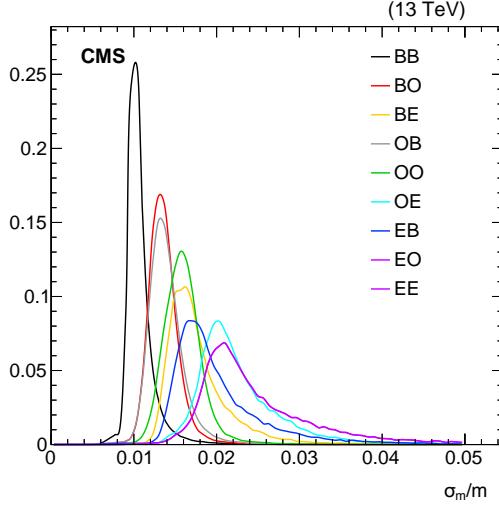


Figure 4.14: Distribution of the per-event dimuon mass resolution in ggH signal events for different combinations of the leading and subleading muon η : “B” corresponds to $|\eta| < 0.9$, “O” to $0.9 < |\eta| < 1.8$, and “E” to $|\eta| > 1.8$. Distributions are scaled in order to have unit area.

may have some bias w.r.t. the *actual* mass uncertainty. Therefore, we perform a calibration of the event-by-event mass uncertainty, separately in data and simulation, to ensure it correctly reflects the dimuon mass resolution in our events of interest.

This calibration is performed using events in the ZCR, where the dimuon mass requirement used here is a bit tighter i.e. $75 < m_{\mu\mu} < 105$ GeV. Furthermore, we categorize events based on the η of each of the two muons. Three η categories are considered: barrel (B) with $|\eta| < 0.9$, overlap (O) with $0.9 < |\eta| < 1.8$, and endcap (E) $1.8 < |\eta| < 2.4$. This gives a total of nine possible η categories for a dimuon event. Each η category is further subdivided into four p_T bins of the leading muon: [30, 45, 52, 62, 200] GeV.

For events with leading muon p_T smaller than 52 GeV, the dimuon mass distribution is characterized by asymmetric tails. Therefore, the mass distribution of events in these low p_T categories is fit using a functional form given by the convolution of a Breit-Wigner function (Equation 4.1) and the double-sided Crystal Ball (Equation 4.2). In contrast, for events with leading muon $p_T > 52$ GeV, the mass peak is more symmetric and therefore is modelled with a Voigtian function, which corresponds to the convolution of a Breit-Wigner and a Gaussian function.

$$\text{BW}(m_{\mu\mu}; m_Z, \Gamma_Z) = \frac{1}{(m_{\mu\mu} - m_Z)^2 + \Gamma_Z^2/4} \quad (4.1)$$

$$\text{DCB}(m_{\mu\mu}) = \begin{cases} e^{-(m_{\mu\mu}-s)^2/(2\sigma^2)} & -\alpha_L < (m_{\mu\mu} - s)/\sigma < \alpha_R \\ \left(\frac{n_L}{|\alpha_L|}\right)^{n_L} \times e^{-\alpha_L^2/2} \times \left(\frac{n_L}{|\alpha_L|} - |\alpha_L| - (m_{\mu\mu} - s)/\sigma\right)^{-n_L} & (m_{\mu\mu} - s)/\sigma \leq -\alpha_L \\ \left(\frac{n_R}{|\alpha_R|}\right)^{n_R} \times e^{-\alpha_R^2/2} \times \left(\frac{n_R}{|\alpha_R|} - |\alpha_R| - (m_{\mu\mu} - s)/\sigma\right)^{-n_R} & (m_{\mu\mu} - s)/\sigma \geq \alpha_R \end{cases} \quad (4.2)$$

The peak and width parameters of the Breit-Wigner function are fixed to the nominal mass and width of the Z boson, whereas the other parameters are allowed to float. Moreover, an additional component for the background is also included and is modelled with an exponential function in the two lowest p_T bins ($p_T < 52$ GeV), while a $\text{Erf} \times \text{Exp}$ is adopted in events with an “high” p_T leading muon ($p_T > 52$ GeV). The σ parameter of the Gaussian core can

be compared to the median value of the event-by-event mass uncertainty for all the events in that category. The ratio of the two defines the calibration for that specific (η , p_T) category. Calibration factors are extracted separately in data and simulation and for each eras (2016, 2017, and 2018). Figure 4.15 shows three examples of fits to the dimuon invariant mass distribution in data in different η categories for events with leading muon p_T between 52 and 62 GeV recorded in 2016. The estimated calibration factors in data and MC for each era and for each (η , p_T) category are listed in Table 4.1.

Overall, in both data and MC, the calibration factors in each (η , p_T) category are positive, in the range between 1.05 and 1.30, indicating that the per-event mass resolution estimated from the covariance matrix of the track fit tends to overestimate the mass resolution of the Z peak by about 5–30%. In particular, the dimuon events that require the largest correction are those in which at least one of the two muons belong to the barrel region and the leading muon has a p_T above 50 GeV. The calibration factors extracted from simulated events are a few % larger than those from data. These corrections are also not identical across eras, in fact, in both data and MC, they progressively decrease moving from the 2016 to the 2018 era.

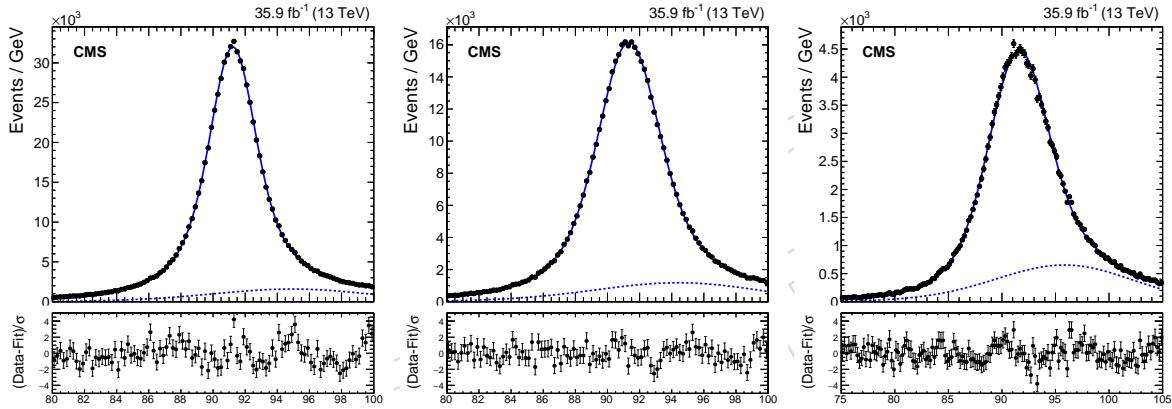


Figure 4.15: Example of fits to the the dimuon invariant mass distribution in data for events with leading muon p_T between 52 and 62 GeV recorded in 2016. From left to right, the distributions correspond to BB, OO, and EE categories.

In order to understand how well we are able to model the mass resolution, a closure test for the calibration procedure is performed. Using the same set of events in the ZCR, we divide the distribution of the calibrated per-event mass resolution is several bins. We then fit the dimuon events in each resolution bin with the same analytical model previously described. Figure 4.16 shows the comparison between the fitted value of the σ parameter in each resolution bin with the corresponding median value of the per-event mass uncertainty for both data (blue) and simulation (red). As can be seen, the calibrated per-event mass uncertainty and the measured resolution agree within 10% in both data and simulation.

Finally, Fig. 4.17 shows the distribution of the per-event mass resolution in data and simulation for events in the signal fit region before (top row) and after (bottom row) the application of calibration factors. It is clear that the calibration procedure allows to significantly improve the agreement between data and MC. The residual discrepancy observed for low resolution events ($\sigma_m/m > 0.035$) has a negligible impact on the analysis.

Table 4.1: Estimated calibration factors, as extracted from the fit to the dimuon mass distribution in Z+jets events previously described, in data and MC for each era and for each (η , p_T) bin.

Leading muon p_T -bin (GeV)	Muon $ \eta $ -bin	Calibration factor in Data			Calibration factor in MC		
		2016	2017	2018	2016	2017	2018
30–45	BB+OB+EB	1.24	1.19	1.16	1.25	1.16	1.17
30–45	BO+OO+EO	1.18	1.13	1.11	1.19	1.11	1.12
30–45	BE+OE+EE	1.15	1.11	1.07	1.17	1.07	1.09
45–52	BB	1.15	1.19	1.16	1.14	1.16	1.10
45–52	BO	1.16	1.13	1.12	1.17	1.12	1.13
45–52	BE	1.16	1.10	1.08	1.18	1.08	1.12
45–52	OB	1.14	1.12	1.08	1.16	1.08	1.10
45–52	OO	1.12	1.09	1.07	1.11	1.11	1.09
45–52	OE	1.07	1.07	1.06	1.10	1.09	1.09
45–52	EB	1.11	1.05	1.03	1.15	1.09	1.05
45–52	EO	1.13	1.03	1.05	1.16	1.10	1.04
45–52	EE	1.00	0.98	0.96	1.04	1.02	1.00
52–62	BB	1.25	1.21	1.19	1.26	1.24	1.21
52–62	BO	1.26	1.24	1.24	1.28	1.28	1.26
52–62	BE	1.19	1.14	1.15	1.22	1.16	1.16
52–62	OB	1.24	1.24	1.22	1.26	1.27	1.24
52–62	OO	1.17	1.16	1.15	1.19	1.18	1.17
52–62	OE	1.17	1.17	1.15	1.20	1.21	1.17
52–62	EB	1.19	1.18	1.13	1.23	1.23	1.18
52–62	EO	1.19	1.15	1.13	1.23	1.20	1.17
52–62	EE	1.11	1.10	1.08	1.15	1.14	1.12
62–200	BB	1.24	1.21	1.18	1.26	1.25	1.19
62–200	BO	1.27	1.24	1.22	1.29	1.28	1.23
62–200	BE	1.15	1.12	1.14	1.17	1.17	1.16
62–200	OB	1.23	1.22	1.21	1.26	1.27	1.23
62–200	OO	1.16	1.15	1.14	1.18	1.18	1.16
62–200	OE	1.18	1.15	1.15	1.21	1.18	1.18
62–200	EB	1.15	1.10	1.12	1.17	1.16	1.16
62–200	EO	1.16	1.11	1.10	1.20	1.18	1.13
62–200	EE	1.09	1.08	1.06	1.14	1.15	1.10

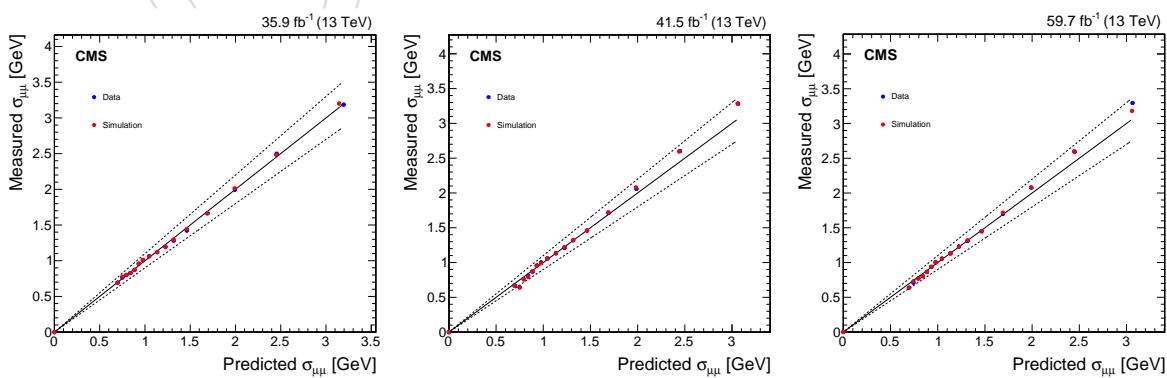


Figure 4.16: Results of the closure test for the per-event mass uncertainty calibration performed in the ZCR for 2016 (left), 2017 (middle), and 2018 (right) eras. After categorizing the events in bins of resolution, the x-axis reports the median calibrated mass resolution in each bin, while the y-axis shows the resolution of the Z boson peak as extracted from an analytical fit to the $m_{\mu\mu}$ distribution.

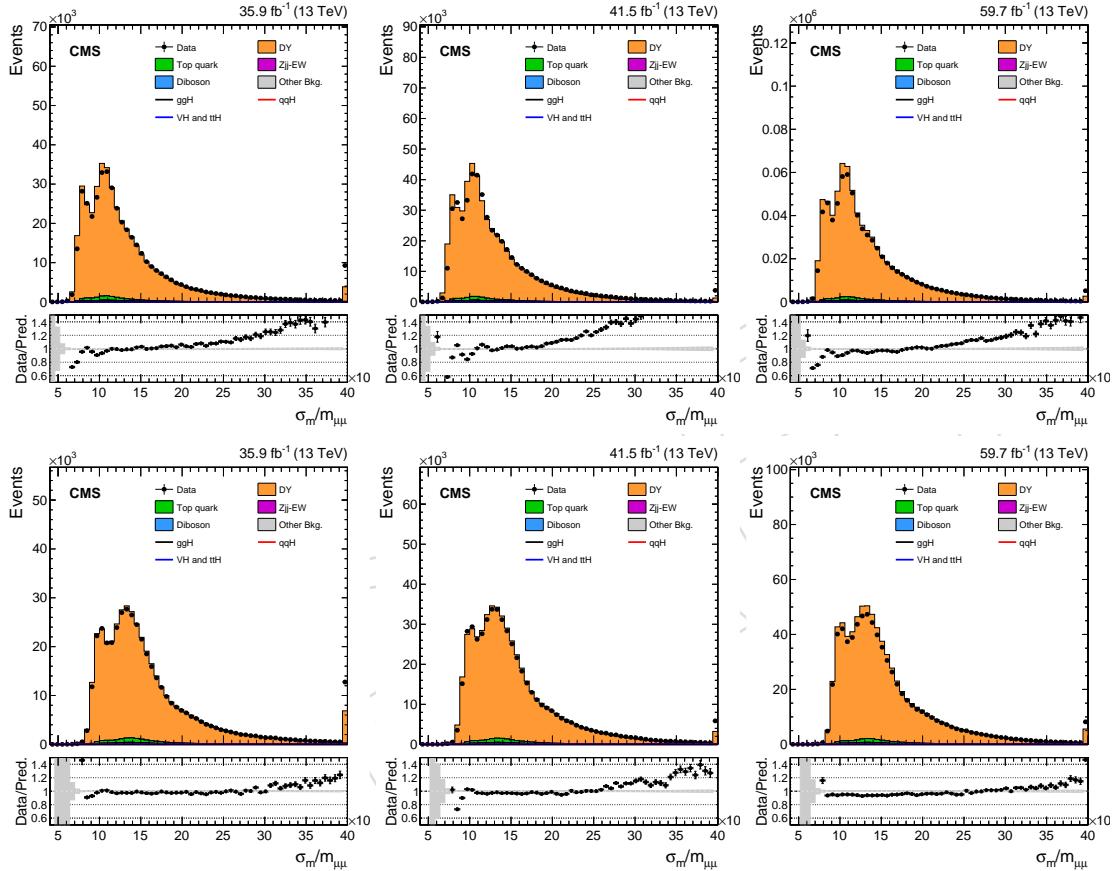


Figure 4.17: The per-event mass resolution distribution in 2016 (left), 2017 (middle) and 2018 (right) eras reported for both data (black points) and simulation (solid histograms) for events in the signal fit region before (top) and after (bottom) the calibration performed in the ZCR. The agreement between data and simulation is significantly improved by applying the calibration coefficients listed in Table 4.1.

812 5 Analysis strategy

813 We start by selecting events accepted by the single muon trigger in which there are exactly two
 814 opposite charged muons passing the requirements described in Section 3.2. These pre-selected
 815 events are further divided into exclusive categories which aim to separate the five different
 816 Higgs boson production modes: ttH, WH, ZH, VBF, and ggH. As a first step, events containing
 817 either one b-tagged jet according to the medium DeepCSV w.p. or two loosely identified b-
 818 tagged jets are classified into the ttH-tag. The ttH category is further divided into two sub-
 819 classes: events containing electrons or additional muons passing the conditions described in
 820 Sections 3.2 and 3.4 belong to the ttH leptonic category, while events without additional leptons
 821 are grouped in the so-called ttH hadronic category. Events without b-tagged jets but containing
 822 at least one additional and identified lepton, either a muon or an electron, are classified into
 823 the VH-tag. The VH category is further divided into two sub-classes: events containing only
 824 one additional electron or muon belong to the WH leptonic tag, while events containing two
 825 additional electrons or muons forms the ZH category. Moreover, events without additional b-
 826 tagged jets or identified leptons (electrons or muons), in which there are at least two hadronic
 827 jets with p_T larger than 25 GeV, leading jet p_T passing $p_T > 35$ GeV, $m_{jj} > 400$ GeV, and $\Delta\eta_{jj} >$
 828 2.5 are classified into the VBF-enriched category. Finally, the ggH category collects all the events
 829 belonging to the residual phase space (no additional b-tagged jets or identified leptons, no pairs
 830 of jets with $m_{jj} > 400$ GeV and $\Delta\eta_{jj} > 2.5$), made mostly by events with either no additional
 831 jets or with just one jet in the final state. Figure 5.1 gives a simple graphical sketch of the event
 832 categorization described above. In each exclusive category of events, three regions are defined
 833 based on the reconstructed dimuon invariant mass ($m_{\mu\mu}$) of the muon pair selected to be the
 834 one coming from $H \rightarrow \mu\mu$ decays. Events with $70 < m_{\mu\mu} < 110$ GeV belong to the *Z control*
 835 *region* as usually these events are largely dominated from on-shell decays of a Z-boson. Events
 836 with $110 < m_{\mu\mu} < 150$ GeV belong to the *signal fit region* as this is the mass range that is used
 837 in the final fit for the signal extraction in analysis where the total background is estimated,
 838 in a data-driven way, by fitting the $m_{\mu\mu}$ spectrum via analytical functions. Lastly events with
 839 $115 < m_{\mu\mu} < 135$ GeV belong to the *signal region*, as the $H \rightarrow \mu\mu$ signal resides here.

840 More details about the event selection and the analysis strategy in each of the aforementioned
 841 classes of events can be found in later Sections of the note.

842 5.1 Main differences w.r.t. HIG-17-019

843 Several changes and improvements in the $H \rightarrow \mu\mu$ analysis strategy have been introduced
 844 with respect to the earlier publication based on 2016 data described in Ref. [13]:

- 845 • **Event classification:** In HIG-17-019 [13], a single BDT was trained in order to pro-
 846 vide the best separation between the total expected signal (ggH, VBF, VH, and ttH)
 847 and the total expected background (DY, top quark, diboson, ttV, triboson, etc.). Fif-
 848 teen event categories were defined in order to maximize the analysis significance on
 849 the basis of η of the forward muon (used as proxy for $m_{\mu\mu}$ resolution) and BDT-score.
 850 This analysis (HIG-17-019) was only sensitive to ggH and VBF modes. In contrast,
 851 the analysis strategy presented in this note is based on four exclusive channels tar-
 852 geting the main Higgs boson production modes. In each channel, specific MVAs
 853 are trained to provide the best separation between the expected contributions from
 854 signal and background processes. In each channel, the choice of input variables has
 855 been optimized to achieve the expected significance to $H \rightarrow \mu\mu$ decays, without re-
 856 lying on observables known to be badly modelled by the simulation. Eventually, the
 857 signal is extracted from a simultaneous fit performed over twenty one independent

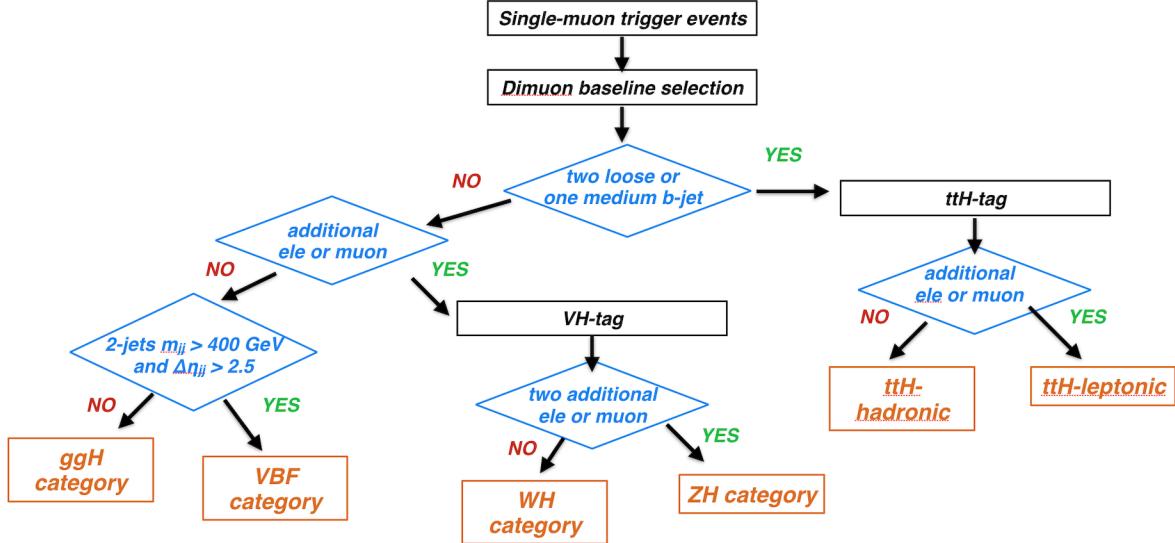


Figure 5.1: Graphical summary of the logical definition of each event category from the baseline dimuon selection. Exclusive channels are defined to target the following production modes: ttH (lepton or hadronic $t\bar{t}$ final states), leptonic WH and ZH, VBF, and ggH.

event categories: five in the ggH channel, three in VBF, six in WH, two in ZH, three in ttH hadronic, and two in ttH leptonic (detailed description will be found in the next Sections).

- **Per event mass uncertainties:** In HIG-17-019 [13], the MVA discriminator was blind to the information of the dimuon mass resolution. The η of the most forward muon in the event was used to gauge the $m_{\mu\mu}$ resolution. In contrast, the dimuon mass resolution is now embedded in the training. In all channels except for the VBF one, where the mass resolution is used as an input variable along with $m_{\mu\mu}$, signal events are weighted proportionally to the per-event mass uncertainty estimated from the covariance matrix of the muon tracks. This choice allows the MVA to learn how the mass resolution evolves with muon p_T and η .
- **Category boundaries:** In order to maximize the sensitivity of data-driven based searches (ggH, VH, and ttH), a new iterative method is used to define categories along the MVA score. This allows to reduce the number of event categories while improving the analysis sensitivity.
- **Background model in ggH:** The background is estimated directly from data (data-driven) by fitting the dimuon mass spectrum with analytical functions. A new strategy has been developed (core-PDF) which allows to keep the bias on the signal strength under control and, at the same time, correlate some of the background parameters across categories decreasing the number of freely floating parameters in the final fit.

The new analysis strategy proposed in this note allows for substantial improvements in the performance. In fact, both the expected significance and upper limit on $\mathcal{B}(H \rightarrow \mu\mu)$ decays for $m_H = 125$ GeV improved by about 35% compared to HIG-17-019.

5.2 Signal model in ggH, VH, and ttH channels

The ggH, VH, and ttH channels follow the same analysis strategy, in which the signal is extracted from a fit to the $m_{\mu\mu}$ distribution across several event categories. The expected $m_{\mu\mu}$

885 distribution in signal events, in each category and for each production mode, is modelled us-
 886 ing a double-sided Crystal Ball function (DCB) which has been already defined in Eq. 4.2. This
 887 functional form provides, at the same time, a robust description of the Higgs dimuon mass
 888 spectrum and a simple way to implement the shape uncertainties, since there is only one pa-
 889 rameter associated to the peak and one to the width of the lineshape. Alternative models can
 890 also be used to model the $m_{\mu\mu}$ spectrum in signal events, as described in Appendix F.

891 5.3 Background model in ggH, VH, and ttH channels

892 In the ggH channel, the expected background in each category is mainly composed of DY
 893 events, with smaller contaminations arising from $t\bar{t}$, single-top, and diboson processes. Pro-
 894 cesses involving dimuon decays of a Z boson represent more than 90% of the total expected
 895 background across the ggH phase space. Similarly, in the WH category the background is
 896 dominated by WZ and DY processes which also involve dimuon decays of on-shell Z bosons.
 897 Finally, in the ZH phase space, the background comes entirely from ZZ and ggZZ processes.
 898 The $m_{\mu\mu}$ distribution for the total background in these categories is expected to be smoothly
 899 falling, driven by the Breit-Wigner nature of the Z boson lineshape. The background is esti-
 900 mated by fitting the dimuon mass spectrum in the signal fit region ($110 < m_{\mu\mu} < 150$ GeV) via
 901 empirical functional forms. According to this strategy, the background contribution under the
 902 peak is therefore estimated from signal-free mass sidebands. We assume no prior knowledge
 903 of the parameters of the fit functions (in both shape and normalization), i.e. they are allowed
 904 to float freely in the fit to the data.

905 In contrast to the other categories, in the ttH categories the major sources of background are
 906 $t\bar{t}$, ttZ, and ttW (W) processes. Therefore, the expected $m_{\mu\mu}$ distribution of background events
 907 does not show a “simple” Breit-Wigner like behaviour and “agnostic functions” are used to
 908 estimate the background.

909 In order to model the background $m_{\mu\mu}$ spectrum in each category, two classes of analytical
 910 functional forms are considered, based on their ability to fit the mass spectrum with a χ^2/dof
 911 of about 1.

912 5.3.1 Physics-inspired models

Given that the majority of the background comes from dimuon decays of a Z boson, this class
 of functions is characterized by a Breit-Wigner core modified by either polynomial, exponential,
 or power-law terms. These functions have been validated by fitting the Drell-Yan mass
 spectrum predicted by the FEWZ [36–38] generator at NNLO accuracy in QCD, and NLO ac-
 curacy in electroweak corrections, within $110 < m_{\mu\mu} < 150$ GeV as described in Ref. [13]. The
 following parametrizations are considered:

$$\text{BWZ} \times \text{Bernstein}(m_{\mu\mu}) = \frac{\Gamma_Z \cdot e^{a \cdot m_{\mu\mu}}}{(m_{\mu\mu} - m_Z)^2 + (\Gamma_Z/2)^2} \times \text{Bern}_n(m_{\mu\mu}) \quad (5.1)$$

$$\text{BWZGamma}(m_{\mu\mu}) = f \cdot \frac{\Gamma_Z \cdot e^{a \cdot m_{\mu\mu}}}{(m_{\mu\mu} - m_Z)^2 + (\Gamma_Z/2)^2} + (1 - f) \cdot \frac{e^{a \cdot m_{\mu\mu}}}{m_{\mu\mu}^2} \quad (5.2)$$

$$\text{BWZRedux}(m_{\mu\mu}) = \frac{\Gamma_Z \cdot e^{a \cdot m_{\mu\mu} + b \cdot m_{\mu\mu}^2}}{(m_{\mu\mu} - m_Z)^c + (\Gamma_Z/2)^c} \quad (5.3)$$

$$\text{BWZ}(m_{\mu\mu}) = \frac{\Gamma_Z \cdot e^{a \cdot m_{\mu\mu}}}{(m_{\mu\mu} - m_Z)^2 + (\Gamma_Z/2)^2} \quad (5.4)$$

where $\text{Bern}_n(m_{\mu\mu})$ corresponds to a Bernstein polynomial of order n , whose coefficients are chosen such that the integral is normalized to unity. In addition, the prediction from FEWZ has been also considered as alternatives to the Breit-Wigner core. The FEWZ prediction in discrete bins of $m_{\mu\mu}$ is first interpolated by means of a spline in the range $110 < m_{\mu\mu} < 150$ GeV. Then, in order to account for the presence of non $Z/\gamma^*+\text{jets}$ backgrounds as well as for detector effects, the spline is multiplied by a Bernstein polynomial as follows:

$$\text{FEWZ} \times \text{Bernstein}(m_{\mu\mu}) = \text{FEWZ-spline}(m_{\mu\mu}) \times \text{Bern}_n(m_{\mu\mu}) \quad (5.5)$$

913 5.3.2 Agnostic models

This class of functions consists in a set of general parametrizations able, in principle, to describe any smoothly falling distribution by choosing the right number of degrees of freedom (*dof*). We consider three main families: polynomials in the Bernstein basis, series of power laws, and series of exponentials, which are defined as follows:

$$\text{Bernstein}(m_{\mu\mu}) = \sum_i^n a_i \cdot \binom{n}{i} m_{\mu\mu}^i (1 - m_{\mu\mu})^{n-i} \quad (5.6)$$

$$\text{S-exponential}(m_{\mu\mu}) = \sum_i^n a_i \cdot e^{b_i \cdot m_{\mu\mu}} \quad (5.7)$$

$$\text{S-power-law}(m_{\mu\mu}) = \sum_i^n a_i \cdot m_{\mu\mu}^{b_i} \quad (5.8)$$

$$\text{Exponential+constant}(m_{\mu\mu}) = f + (1 - f) \times e^{a \cdot m_{\mu\mu}} \quad (5.9)$$

$$\text{Power-law+constant}(m_{\mu\mu}) = f + (1 - f) \times m_{\mu\mu}^a \quad (5.10)$$

914 In addition, another non Breit-Wigner like function, able to fit well the background spectrum
915 in each category, is given by a Landau PDF modulated by polynomials, defined as follows:

$$\text{Landau} \times \text{Bernstein}(m_{\mu\mu}) = \text{Landau}(m_{\mu\mu}, m_Z, a) \times \text{Bern}_n(m_{\mu\mu}) \quad (5.11)$$

916 5.4 Sources of systematic uncertainty

917 In this Section, a summary of the main sources of systematic uncertainty affecting all channels
918 of the search presented in the next Sections is given. The goal is to describe them once, trying to
919 avoid useless repetitions, so that in the next Sections dedicated to each channel only their effect
920 will be reported. These uncertainties are divided into two classes depending if they originate
921 from theory or experimental sources.

922 5.4.1 Experimental uncertainties

- 923 • **Single-muon trigger efficiency:** to account for differences in the single-muon trigger
924 efficiency between data and simulated events, scale factors are provided by the
925 Muon POG and are used to correct the prediction from MC. The corresponding un-
926 certainty is obtained by varying these scale factors within their uncertainties.
- 927 • **Muon identification and isolation efficiency:** to account for differences in the ef-
928 ficiency between data and simulated events of the chosen muon identification and
929 isolation working points, scale factors are provided by the Muon POG and are used
930 to correct the prediction from MC. The corresponding uncertainty is obtained by
931 varying these scale factors within their uncertainties.

- 932 • **Muon energy scale and resolution:** uncertainties in the Rochester scale and smearing
933 corrections are propagated to the event selection in each channel. The corresponding variation in the expected rate for each process in each bin is used as an
934 estimate of the uncertainty.
935
- 936 • **Jet energy scale (JES):** the recommended jet energy corrections by JetMET POG are
937 applied in the analysis. For each era, the JetMET POG provided a list of uncertainty
938 sources [67], which are varied independently for all jets in simulated events. The
939 event selection in each channel is re-applied accounting for the shifted jet energies.
940 The corresponding variations in the expected rate for each process in each bin are
941 used as an estimate of the uncertainty.
- 942 • **Jet energy resolution (JER):** in the analysis presented in this note, jets are not smeared
943 by default with the JER scale factors provided by the JetMET POG. This choice is mo-
944 tivated by the fact that these smearing factors are measured in samples enriched in
945 high- p_T jets, while the phase space explored by the $H \rightarrow \mu\mu$ search is dominated by
946 jets with p_T in the range 25–60 GeV. This choice was made after a discussion involv-
947 ing Higgs conveners and JME experts. However, JER uncertainties are considered
948 in the analysis by taking the variation in the expected rate for each process in each
949 bin caused by the application of the JER smearing. Following the indications from
950 JetMET POG, JER uncertainties are divided in six exclusive categories. Central jets
951 are divided in two categories by η : the first is composed by jets with $|\eta| < 1.93$,
952 while the second is composed by jets with $1.93 < |\eta| < 2.5$. Forward jets are divided
953 in four categories by η and p_T : $p_T < 50$ GeV and $2.5 < |\eta| < 3.139$, $p_T > 50$ GeV and
954 $2.5 < |\eta| < 3.139$, $p_T < 50$ GeV and $|\eta| > 3.139$, finally $p_T > 50$ GeV and $|\eta| > 3.139$.
- 955 • **B-tagging efficiency:** to account for differences in the efficiency between data and
956 simulated events of the chosen b-tagging working points, scale factors are provided
957 by the BTV POG and are used to correct the prediction from MC [63]. The corre-
958 sponding uncertainty is obtained by varying these scale factors within their uncer-
959 tainties.
- 960 • **Pileup model:** an uncertainty in the number of pileup events is obtained by varying
961 the minimum bias cross section used in the pileup re-weighting applied to simulated
962 events. We follow the official recommendation for CMS analyses.
- 963 • **Prefiring correction:** the prefiring inefficiency is emulated in MC by weighting sim-
964 ulated events with precomputed prefiring probability maps, as a function of jet η
965 and p_T , provided by the JetMET POG [68]. The corresponding uncertainty is ob-
966 tained by varying these probabilities within their associated uncertainties, taken to
967 be as large as 20% of the prefiring correction.
- 968 • **Electron identification and isolation efficiency:** to account for differences in the ef-
969 ficiency between data and simulated events of the chosen electron identification and
970 isolation working points, scale factors are provided by the EGamma POG and are
971 used to correct the prediction from MC. The corresponding uncertainty is obtained
972 by varying these scale factors within their uncertainties. This uncertainty only af-
973 fects WH, ZH, and ttH-leptonic categories.
- 974 • **Electron energy scale:** uncertainties in the EGamma scale and smearing corrections
975 are propagated to the event selection in the WH, ZH, and ttH-leptonic channels. The
976 corresponding variation in the expected rate for each process in each bin is used as
977 an estimate of the uncertainty.
- 978 • **LeptonMVA efficiency:** to account for differences in the efficiency between data and

simulated events of the chosen leptonMVA working point, scale factors are taken from the TOP PAG and are used to correct the prediction from MC. The corresponding uncertainty is obtained by varying these scale factors within their uncertainties. This uncertainty only affects WH, ZH, and ttH-leptonic categories.

- **Luminosity:** this uncertainty affects directly the predicted signal and background yields from simulation. The recommended uncertainties from LUM POG for each year are used.
- **Quark-Gluon-Likelihood discriminator:** uncertainties on quark-gluon-likelihood discriminator are evaluated by applying the recipe provided by JetMET POG reported in this Twiki [69].

5.4.2 Theoretical uncertainties

- **Signal inclusive cross section:** uncertainties in the production cross section for ggH, VBF, VH, and ttH processes from QCD scale and PDF variations are taken from Ref. [12] and are listed in Tab. 2.1.
- **ggH STXS stage 1:** these are a set of uncertainties specific for the ggH process and are estimated following the recommendations provided by the LHC Higgs cross-section working group (WG1) [70]. This recipe provides a set of independent sources of uncertainty defined in STXS bins. Each source is modelled via a log-Normal nuisance parameter correlated across categories and eras. These sources account for variations in the estimate of the ggH acceptance in bins of Higgs boson p_T and N_{jets} .
- **VBF STXS stage 1:** acceptance uncertainties specifically designed for the VBF production mode are estimated following the recommendations provided by the LHC Higgs cross-section working (WG2) [71]. This recipe provides a set of independent sources of uncertainty defined in different STXS bins, which are modelled via log-Normal nuisance parameters correlated across categories and eras. They account for variation of the VBF signal acceptance as a function of Higgs boson p_T , N_{jets} , and m_{jj} .
- **QCD scale variations:** the background prediction in the VBF channel is taken from MC. Therefore, theoretical uncertainties arising from choice of renormalization and factorization scale values are evaluated by varying them by a factor 2 and 1/2 compared to their default values. The acceptance only effect from scale variations on VH and ttH events is also considered.
- **PDF variations:** the background prediction in the VBF channel is taken from MC. Therefore, the uncertainty due to the PDFs is evaluated by taking the RMS of the predictions from the replicas available for NNPDF3.0 in the 2016 samples, while for 2017/18 samples the sum-in-quadrature of the variations coming from the Hessian components of NNPDF3.1. The acceptance only effect from PDF variations on ggH, VBF, VH, and ttH events is also considered.
- **Parton shower acceptance:** the acceptance uncertainty from parton shower is computed only for signal processes and for the Zjj-EW sample (for the VBF signal and Zjj-EW processes, more info are given in Section 7). In the case of ggH, WH, WH, and ttH signal events, a number of PYTHIA tune parameters have been varied independently during the shower and their effect is stored as variation in the per-event weight assigned at the generator level. The sum-in-quadrature of such variations compared to the nominal prediction is used to assign a parton shower uncertainty.
- **Underlying event tune:** is estimated for signal processes only with alternative simulated samples generated by varying parameters of the underlying event tune.

1025 6 Gluon-fusion channel

1026 In this Section, we describe the search for $H \rightarrow \mu\mu$ decays performed using the data set col-
 1027 lected by the CMS detector during Run-2, at a centre of mass energy of 13 TeV. This corresponds
 1028 to a total integrated luminosity of 137.2 fb^{-1} . The analysis described here specifically targets
 1029 the ggH production mode that offer the largest Higgs boson production cross sections at the
 1030 LHC. The final states of interest comprise a pair of two prompt, isolated and opposite charge
 1031 muons produced in association with 0, 1, or 2 and more jets (with m_{jj} and $\Delta\eta_{jj}$). The largest
 1032 background in this channel consists of Drell-Yan events in which an off-shell Z boson decays
 1033 to a pair of muons. The next most significant background contributions come from $t\bar{t}$ events,
 1034 in which both the top quarks decay to muons, and from diboson production (WW, WZ, ZZ).
 1035 The gluon fusion signal is quite similar to the Drell-Yan background. The most important dis-
 1036 criminating feature, apart from the dimuon resonance peak, is the transverse momentum (p_T)
 1037 of the dimuon system, which has a harder spectrum in the case of the Higgs boson signal w.r.t.
 1038 the Drell-Yan background. The analysis takes advantage of all of these features, and searches
 1039 for a narrow dimuon resonance at the Higgs boson mass of 125 GeV on top of the background
 1040 continuum.

1041 In addition, a similar analysis strategy has been developed targeting the VBF phase space. The
 1042 VBF signal, although smaller in cross section compared to ggH by an order of magnitude, pro-
 1043 vides a distinctive signature wherein the two muons from the Higgs boson decay are produced
 1044 in association with two jets that have a large pseudorapidity gap ($\Delta\eta_{jj}$), and form a large dijet
 1045 invariant mass (m_{jj}). This feature helps to suppress the large Drell-Yan background, making
 1046 the VBF channel a very sensitive probe of the $H \rightarrow \mu\mu$ signal. This analysis is detailed in Ap-
 1047 pendix A and serves as an orthogonal cross check of the VBF template fit described in Section 7.
 1048 The advantage of such strategy consists in a fully data-driven background estimation from fits
 1049 to the $m_{\mu\mu}$ distribution in VBF enriched categories via analytical functions. On the other hand,
 1050 the constraint power of the observed data in the mass sidebands is weaker than the one coming
 1051 from MC. Therefore, the VBF “data-driven” analysis results to be about 15% less sensitive than
 1052 the template method.

1053 6.1 Multivariate discriminant

1054 The most important discriminating variable between signal and background processes in this
 1055 search is the dimuon invariant mass, which forms a sharp peak around 125 GeV for the signal
 1056 as opposed to the background in which it decreases monotonically within the same mass range.
 1057 This feature also allows for a completely data-driven estimation of the background through an
 1058 analytical fit to the dimuon mass spectrum. However, there are other kinematic variables that
 1059 can be used to enhance the signal-to-background separation. As has been previously men-
 1060 tioned, the VBF signal offers a distinctive signature involving two jets with a large η -gap and
 1061 large invariant mass. In the case of the gluon fusion signal, the p_T of the dimuon system rep-
 1062 resents an important discriminating observable since it shows a harder spectrum in the signal
 1063 compared to the Drell-Yan background.

1064 Therefore, in order to enhance the signal-to-background separation, we construct a multivariate
 1065 discriminant using boosted decision trees (BDT), which takes as input several variables that
 1066 capture the distinctive kinematic features of the signal and the background. This BDT is trained
 1067 on events belonging to the “residual” phase space not selected by ttH, leptonic VH, and VBF
 1068 channels described in Section 8, 9, and 7, respectively. This corresponds to events with no
 1069 additional leptons, no b-tagged jets according the criterion described in Section 5, and, in the
 1070 case of events with two or more jets, the m_{jj} of the two p_T -leading jets is required to be smaller

than 400 GeV or their $\Delta\eta_{jj}$ is required to be smaller than 2.5. 0 or 1 additional jets, and events with two or more jets in which the m_{jj} of the two leading jets is smaller than 400 GeV. The dominant source of signal is ggH, while the main background comes from Drell-Yan. This discriminator is referred to as the “ggH BDT” or “non-VBF BDT”. The input variables have been specifically chosen so that the BDT output remains highly uncorrelated with the dimuon mass. We then categorize events based on BDT score in order to maximize the search sensitivity. Since the BDT output is highly uncorrelated with the dimuon mass, this categorization does not sculpt the background to peak around the Higgs boson mass. This in turn allows us to fit the dimuon mass distribution in each of the categories without a significant bias. Finally, we perform a statistical combination of all the categories to extract a final estimate of the analysis performance.

6.1.1 Input variables

We start with a suite of input variables constructed entirely from the two muons. A dimuon system essentially has six degrees of freedom corresponding to the momentum 3-vectors of the two muons. The ϕ -symmetry of the detector makes one of the degrees of freedom redundant. Since we want to explicitly avoid the use of the dimuon mass in the BDT discriminant, that leaves us with four degrees of freedom i.e. four possible input variables. Two of these variables are chosen to be the $p_T^{\mu\mu}$ and $y_{\mu\mu}$, and they correspond to the p_T and rapidity of the dimuon system. The distributions of these variables are sensitive to the production mechanism in the case of the signal and the Drell-Yan background. The $p_T^{\mu\mu}$ variable in particular is found to be the most discriminating between ggH and Drell-Yan events. The remaining two variables correspond to the decay of the boson in the rest frame of the dimuon system. We specifically choose the Collins-Sopper rest frame [72] for this purpose, and take the azimuthal angle (ϕ_{CS}) and the cosine of the polar angle ($\cos\theta_{CS}$) to be the two variables. These decay angles carry some discrimination owing to the difference in the spin properties of the signal (scalar) and the Drell-Yan background (spin-1). In addition to the four aforementioned variables, we also include single muon information represented by the transverse momenta of the two muons relative the dimuon mass, $p_T^{\mu_1}/m_{\mu\mu}$ and $p_T^{\mu_2}/m_{\mu\mu}$, as well as the individual muon η , η^{μ_1} and η^{μ_2} . Figure 6.1 shows the shapes of all the muon related variables used as inputs to the BDT training for signal and background events, where the signal is given by the sum of the expected contributions from the different Higgs boson production mechanisms (ggH, VBF, VH, and ttH) while the background is the total expected background (DY, top, ttV, diboson, triboson, Z+jets EW, etc.).

In addition to the observables describing the dimuon system, we also include variables carrying information about jets in the event. In the case of events with only one jet in the final state with p_T larger than 25 GeV, the following variables are used to enhance the signal-to-background separation: jet p_T , jet η , the η and the ϕ gaps between the dimuon system and the jet ($\Delta\eta(\mu\mu, j)$, $\Delta\phi(\mu\mu, j)$). In events with two or more jets, other variables help to target the presence of VBF and V(qq)H signals, as well as to further exploit differences in the jet kinematics between ggH and Drell-Yan events. These are represented by the η -gap between the two leading jets in the event ($\Delta\eta_{jj}$), the dijet invariant mass (m_{jj}), their azimuthal angular separation ($\Delta\phi_{jj}$) which carries some spin-discrimination power [73], the Zeppenfeld variable [74] defined in Equation 6.1, where y_s correspond to the rapidity of the dimuon system, the leading and the subleading jets, the p_T of the subleading jet, and the minimum difference in η and ϕ between the dimuon system and the two jets ($\min-\Delta\eta(\mu\mu, j_1j_2)$, $\min-\Delta\phi(\mu\mu, j_1j_2)$). Lastly, the distribution of the number of jets of the event is also used as input. In events with $N_{jet} = 0$, jet variables are set to zero apart from $\Delta\phi_{jj}$ which is set to be -1. Similarly, in one jet events dijet observables

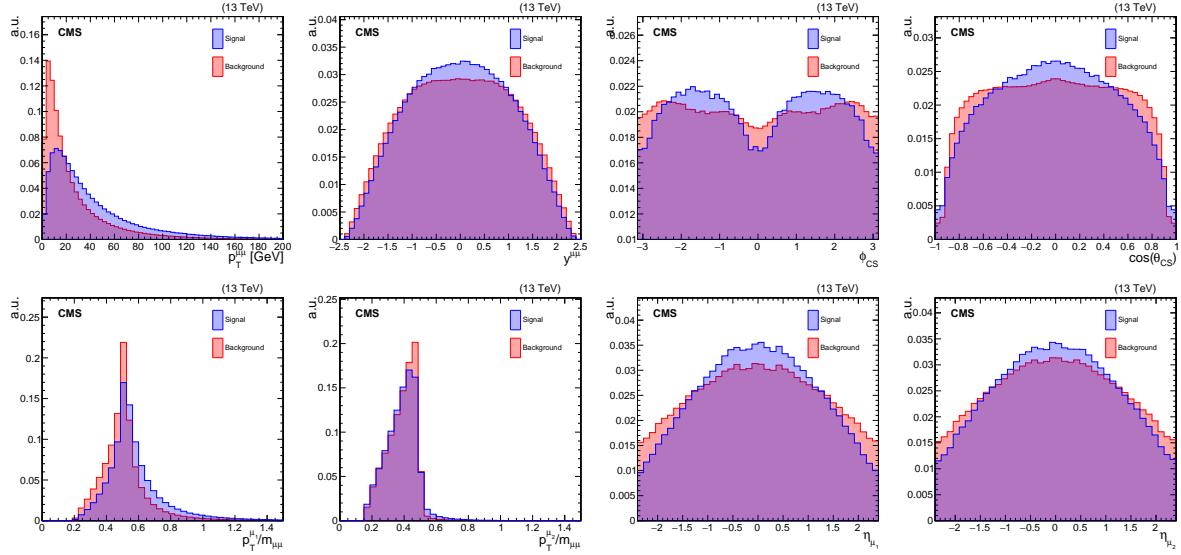


Figure 6.1: Distribution of the dimuon (top) and muon (bottom) variables used as input in the ggH BDT in signal (blue) and background (red) events. Distributions are scaled in order to have unit area. On top, from left to right: $p_T^{\mu\mu}$, $y_{\mu\mu}$, ϕ_{CS} , and $\cos \theta_{\text{CS}}$. Bottom, from left to right: $p_T^{\mu_1}/m_{\mu\mu}$, $p_T^{\mu_2}/m_{\mu\mu}$, η_{μ_1} , η_{μ_2} .

are defaulted to zero apart from $\Delta\phi_{jj}$ which is set again to be equal to -1.

$$z^*(H) = \frac{y_{\mu\mu} - (y_{j_1} + y_{j_2})/2}{|y_{j_1} - y_{j_2}|} \quad (6.1)$$

Fig 6.2 shows the shapes in signal and background events of jet related variables, as well as angular distances ratios between the dimuon system and the jets which are used as inputs to the ggH BDT trained on events with m_{jj} smaller than 400 GeV or $\Delta\eta_{jj}$ smaller than 2.5.

Figure 6.3 and 6.4 show the distributions of the ggH BDT inputs in data and simulation for events in the signal fit region after requiring that no events with two or more jets have m_{jj} larger than 400 GeV, $\Delta\eta_{jj}$ above 2.5 units, and leading jet $p_T > 35$ GeV. The observation and the expectation in the three eras (2016, 2017, and 2018) have been summed together. Overall, the agreement between data and MC for both muon and jet variables used in input is good. In the $N_{\text{jet}} \geq 1$ phase-space, an overall normalization discrepancy of about 5% is observed between data and MC as reported by the N_{jets} distribution. Since the background in this search is directly estimated from data, this small discrepancy in normalization between data and simulation does not have an impact in the analysis results. The largest differences between data and simulation are observed in the leading and subleading jet η distributions, when the jet falls either in the HF detector ($|\eta| > 3.0$) or is reconstructed within the most forward sectors of EE/HE ($2.7 < |\eta| < 3.0$). During Run2, this region registered a large increase in the noise vs time due to a strong radiation damage of both active materials (ECAL crystals) and light sensors. Moreover, discrepancies are also observed in the tails of m_{jj} and $\Delta\eta_{jj}$ distributions as reported in Fig. 6.4. However, these events are very peculiar and are affected by a large uncertainty from jet energy scale and resolution sources. These events characterized by high m_{jj} and $\Delta\eta_{jj}$ but they are not accepted by the VBF category because the leading jet p_T is smaller than 35 GeV. Comparisons are also reported independently for each era in the Appendix of Ref. [17], where a more detailed description of the observed trends in the main kinematic distributions is given.

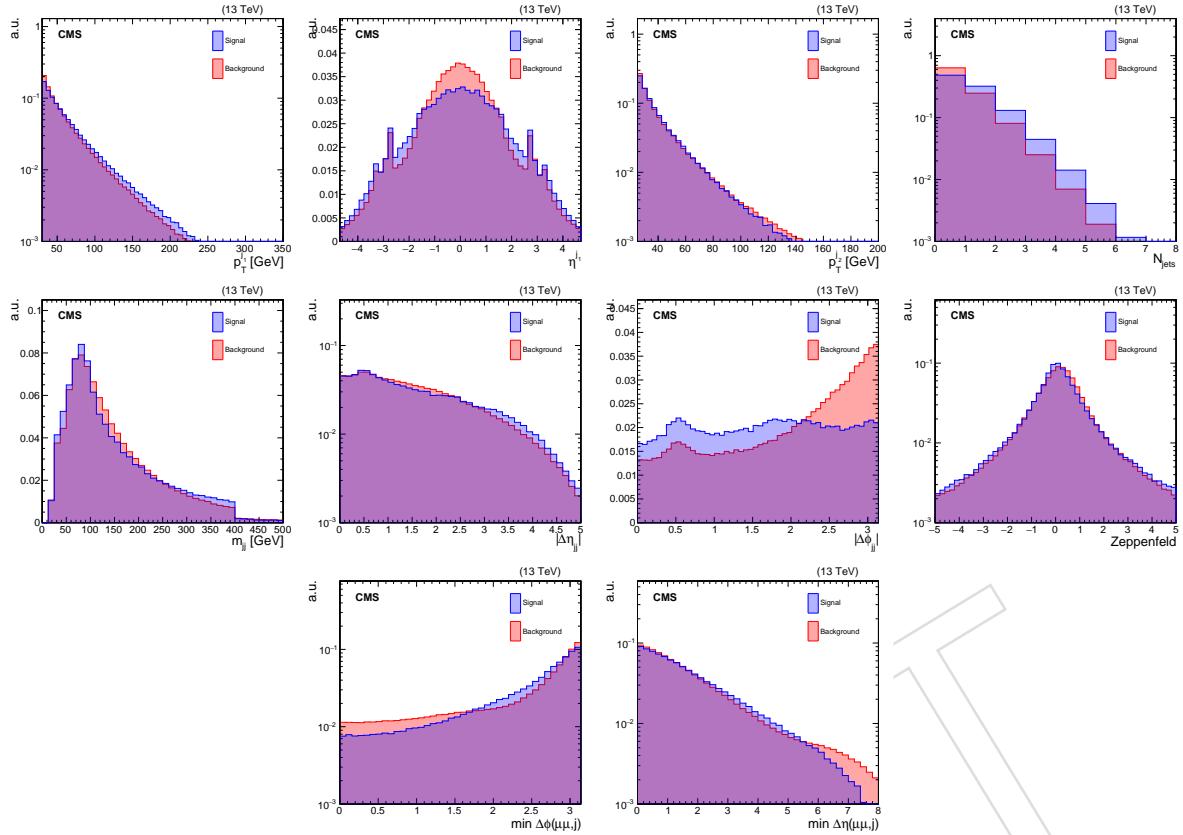


Figure 6.2: Distribution of the jet related variables used as input in the ggH BDT in signal (blue) and background (red) events. Distributions are scaled in order to have unit area. First row, from left to right: p_T^{j1} , η^{j1} , p_T^{j2} , and N_{jets} . Second row, from left to right: m_{jj} , $\Delta\eta_{jj}$, $\Delta\phi_{jj}$, and Zeppenfeld. Last row, from left to right: $\min|\Delta\phi(\mu\mu, js)|$, and $\min|\Delta\eta(\mu\mu, js)|$.

6.1.2 BDT discriminant training

The training is performed separately on the MC samples for the three eras: 2016, 2017, and 2018. This helps to capture the impact of the differences in reconstruction and detector performance across the three data taking periods. For example, the occupancy and properties of forward jets $|\eta| > 2.6$ is pretty different between 2016 and the other years. In addition, in the 2017 era, a tight requirement on the pileup jet ID is adopted for jets in the range $2.6 < |\eta| < 3.0$ in order to mitigate the noise from the ECAL detector. The BDT discriminants are trained using the ‘Gradient Boost’ method. The training is done using simulated signal and background events, using approximatively half of the available statistics corresponding to events with odd event number. Events with even event number are used instead for testing purposes. In the case of the signal, POWHEG samples (centrally produced) for ggH and VBF are used in the training because they don’t have negative weights. For background samples with negative weighted events (Drell-Yan for example), we observed similar performance on the testing sample following three alternative strategies: use both positive and negative weighted events in the training, discard events with a negative weight, and take the absolute value of the per-event weight⁴. In order to maximize the training statistics, the final BDTs have been trained by flipping the sign of negative weighted events. The training is performed targeting an Higgs boson signal with a

⁴In order not to bias the background composition, a correction factor is introduced to ensure that the same amount of background events are expected when either negative weighted events are discarded or they are considered with a positive weight.

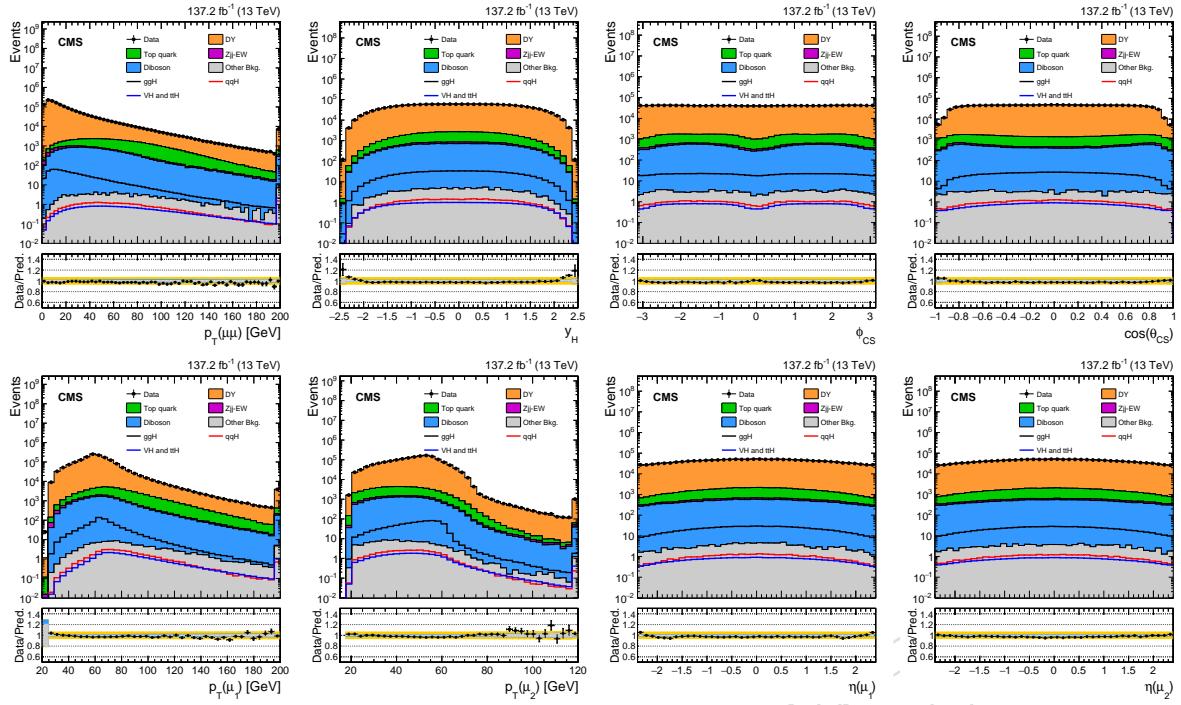


Figure 6.3: Distributions of muon variables used in input to the ggH BDT training for data (black points) and simulation (solid histograms) obtained from events in the signal fit region. First line, from left to right, $p_T^{\mu\mu}$, $y_{\mu\mu}$, ϕ_{CS} , and $\cos \theta_{\text{CS}}$. Second line, from left to right: $p_T^{\mu_1}$, $p_T^{\mu_2}$, η^{μ_1} , and η^{μ_2} . In the ratio panel, the grey band indicates the uncertainty coming from the limited statistics of the simulated samples, the azure one corresponds to the sum in quadrature between statistical and systematic uncertainties from experimental sources, while the orange band includes the theoretical uncertainties in the background prediction. The asymmetry in the experimental uncertainty band is due to the one-sided effect of jet energy resolution.

mass of 125 GeV. The settings used for the BDT training are as follows:

- Number of trees = 1000.
- Minimum node size = 3%.
- Shrinkage = 0.10.
- Bagged sample fraction = 0.5.
- Number of cuts = 30.
- Maximum depth = 4.
- Transformation for inputs: (I, N).
- Separation type: CrossEntropy.

As explained in Section 6.1, the input variables have been specifically chosen so that the BDT output remains highly uncorrelated with the dimuon mass. The training is performed in a tighter mass range compared to the signal fit region, i.e. $115 < m_{\mu\mu} < 135$ GeV, which corresponds to a ± 10 GeV window around the true Higgs boson mass. We perform several alternative trainings: independent BDTs in each jet bin ($N_{\text{jets}} = 0, 1, \geq 2$), binary or multi-classifier based on a deep neural network. However, we did not observe any significant gain from these alternative methods, therefore we have preferred to stick with the “original” strategy.

The analysis sensitivity depends on the dimuon mass resolution, which in turn depends strongly on the muon kinematics as described in Section 4.8, particularly the η of the muon. There-

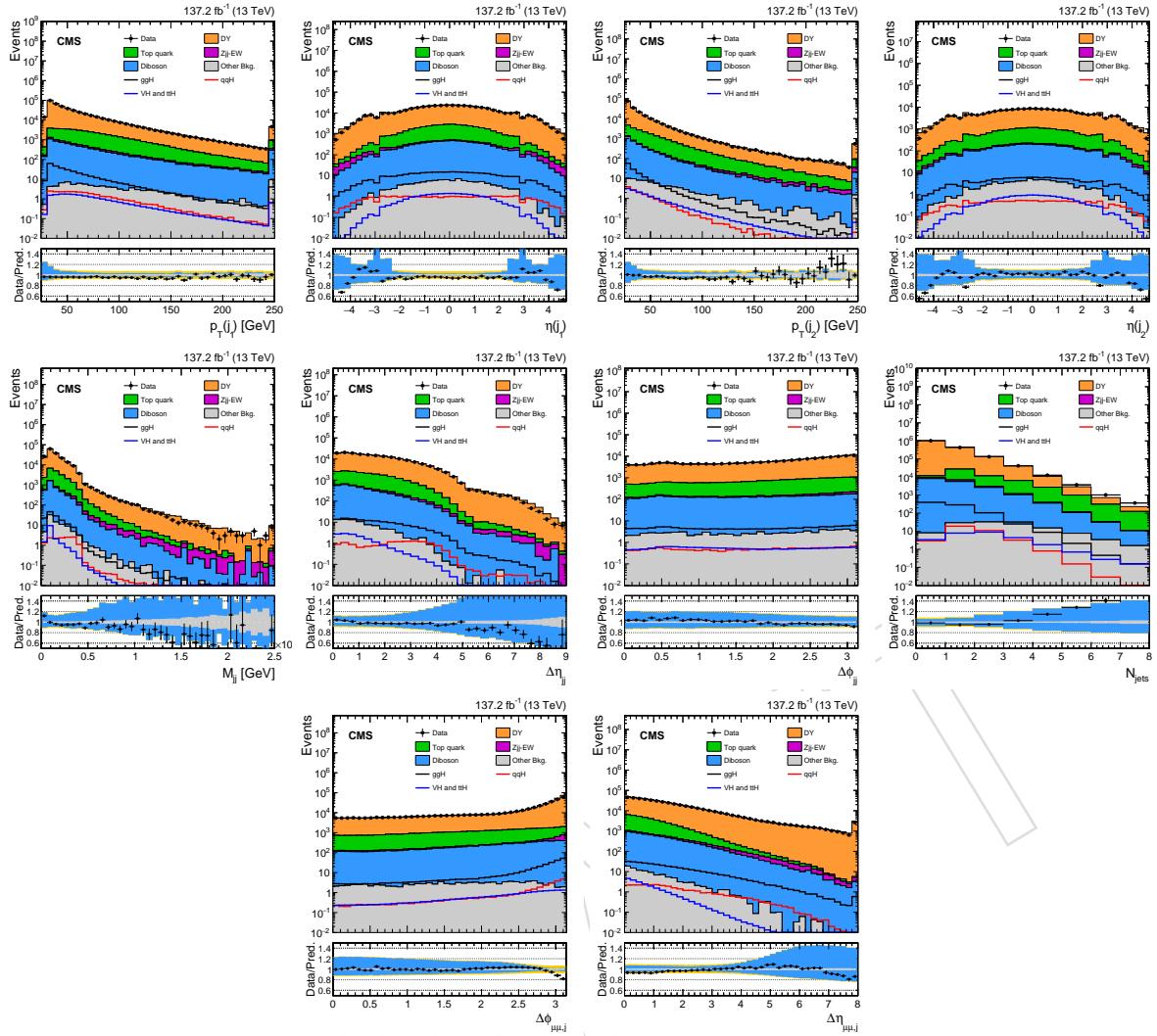


Figure 6.4: Distributions of jet variables used in input to ggH BDT training for data (black points) and simulation (solid histograms) obtained from events in the signal fit region. First row, from left to right: $p_T^{j_1}$, η^{j_1} , $p_T^{j_2}$, and η^{j_2} . Second row, from left to right: m_{jj} , $\Delta\eta_{jj}$, $\Delta\phi_{jj}$, and N_{jets} . Last row, from left to right: $\min - \Delta\phi(\mu\mu, js)$, $\min - \Delta\eta(\mu\mu, js)$. The description of the ratio panel is the same of Fig. 6.3.

fore, we make the BDT training aware of the dimuon mass resolution by weighting the signal events proportionally to $1/\sigma_{\mu\mu'}$, where $\sigma_{\mu\mu'}$ is the calibrated per-event dimuon mass resolution. It should be noted that the $\sigma_{\mu\mu'}$ distribution is not used as an input to the BDT, but is only used as a weighting factor in the training. The weight is not applied in the evaluation of the BDT score. The performance of the BDT discriminant does not change significantly if we include the per-event mass resolution as one of the inputs, probably because the BDT can infer it already from single muon variables ($p_T/m_{\mu\mu}$ and η). For this reason, we have chosen not to include the per-event mass resolution in the list of input variables. The adopted use of the per-event mass uncertainties is a significant departure from the approach adopted in the earlier iteration of the analysis [13]. In fact, in the previous iteration, events were categorized both on the basis of their BDT score and the maximum of the $|\eta|$ of the two muons. The approach adopted here of using per-event mass uncertainties as a weight for signal events during the BDT training reduces the categorization to a single variable i.e. the BDT score. In addition, this also carries more precisely the information about the muon momentum resolution and its correlation with the event kinematics (muon p_T and η). Consequently, we will end up using significantly fewer

1191 categories but still having better performance (25% improved expected significance) on 2016
 1192 data w.r.t. Ref. [13].

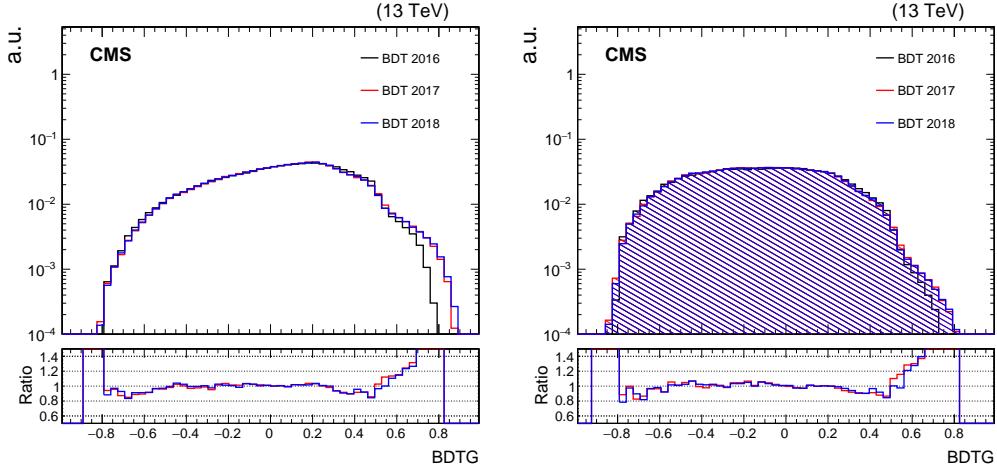


Figure 6.5: Distribution of the ggH BDT output in signal (left) and background (right) simulated events taken from the testing sample for three eras: 2016 (black), 2017 (red), and 2018 (blue). Distributions are scaled in order to have unit area. The red (blue) distributions in lower pad is defined as the ratio between the 2018 (2017) BDT shape and the 2016 one, which is used as a reference. The shapes of the BDT outputs are found to be quite similar across the three eras. Relevant differences are observed only in the high score region for both signal and background events.

1193 Figure 6.5 shows a comparison between the output of the BDT discriminant in the case of the
 1194 signal (left) and background (right) simulations for the three eras. The shapes of the ggH BDT
 1195 outputs are found to be in good agreement for both signal and background processes apart
 1196 from the high score region (BDT-score larger than 0.6), where the 2017 and 2018 BDTs show
 1197 longer tails in both signal and background events. Overall, the BDT discriminators related to
 1198 2017 and 2018 eras are quite similar in terms of shapes while, in contrast, the 2016 shapes are
 1199 more different particularly in the high score region. Since the properties of jets in the endcap
 1200 and forward regions of the detector changed significantly after the 2016 data-taking, it is ex-
 1201 pected to see differences in regions more sensitive to the forward jet kinematics between the
 1202 2016 BDTs and those from the other eras.

1203 Figure 6.6 (left) shows the "ROC" (Receiver Operating Characteristic) curves for the ggH BDT
 1204 discriminant. The performance of the ggH BDT discriminant is found to be very similar across
 1205 the three eras. Figure 6.6 (right) shows the dimuon mass resolution, expressed in terms of full
 1206 width half maximum (FWHM), as a function of the BDT score. The dimuon mass resolution
 1207 is extracted from a fit to the $m_{\mu\mu}$ distribution with a Double-sided Crystal Ball function. The
 1208 signal resolution improves as a function of the BDT score, thereby demonstrating that the BDT
 1209 is indeed aware of the importance of the dimuon mass resolution. In addition, the dimuon
 1210 mass resolution is found to improve progressively moving from 2016 to 2018 eras, and it is also
 1211 observed to be stable as a function of the BDT outputs. This gain in the muon p_T resolution
 1212 is due to a combination of sources: installation of an additional layer in the Pixel detector in
 1213 early 2017, improvements in muon and tracker detector alignments, as well as improvements
 1214 in beam-spot and muon tracker reconstruction (particularly for high p_T muons) allowing the
 1215 geo-fit corrections to be more effective in 2018 rather than in the 2016 data-taking period.

1216 In order to understand how the BDT scores are correlated with the input observables, Fig. 6.7
 1217 shows, for signal events, the evolution of the shapes of some of the input variables as a function
 1218 of the ggH BDT score. In events belonging to the ggH phase space, the high score region probes

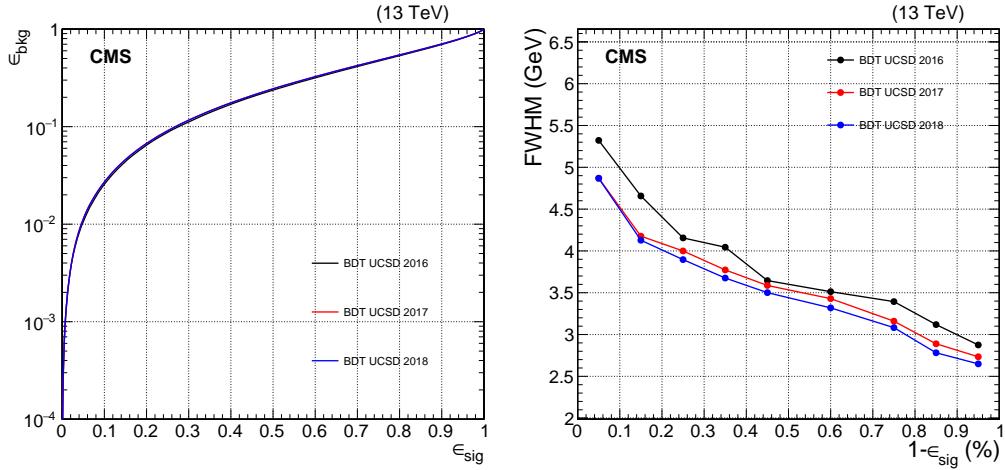


Figure 6.6: On the left, signal efficiency (ϵ_{sig}) versus background efficiency (ϵ_{bkg}) ROC curves obtained from the ggH BDT output scores for events in the testing sample. The performance of the BDT discriminant is found to be stable across the three eras. On the right, FWHM of the dimuon mass distribution in signal events computed as a function of the “inclusive” BDT score. The signal resolution improves similarly in the three eras as a function of the BDT score, proving that the BDT is able to assign high score to events with good mass resolution.

events with good dimuon mass resolution, as well as a large dimuon p_T . A good fraction of events at high are characterized by the presence of one jet or two jets in the final state with a large η -separation with respect to the dimuon system.

Moreover, it is also crucial to make sure that the BDT has not learnt the signal mass hypothesis used in the training, $m_H = 125 \text{ GeV}$, via hidden correlations between the invariant mass of the two muons and the other input variables. To do so, two independent cross checks are performed:

- Fig. 6.8 (left) shows, for background events, the evolution of the dimuon mass shape as a function of the ggH BDT score where no sculpting of the expected mass distribution is observed either at low or at high BDT score. This suggests that the BDT output is not sensitive to the dimuon mass distribution.
- Fig. 6.8 (right) shows, for ggH signal events, the evolution of the ggH BDT shape as a function of the signal mass hypothesis, i.e. for $m_H = 120, 125$, and 130 GeV . No differences as a function of the mass are observed for the BDT output.

Therefore, we can conclude that the ggH BDT is highly uncorrelated with the dimuon mass. Events can be further categorized based on the BDT scores in order to maximize the search sensitivity without sculpting or biasing the expected invariant mass distribution for signal and background events.

Finally, Fig. 6.9 shows a comparison between the ROC curves obtained in each era and in each training from training (black) and testing (red) events. The ROC curves are essentially identical between training and testing events for all eras, indicating that no over-training is observed.

6.2 Event categorization

At this stage, the ggH phase space collects all the remaining events which are not belonging to the VBF channel or to the other exclusive modes (VH-leptonic and ttH). Figure 6.10 shows the distributions of the ggH BDT score in data and simulation for events belonging to the signal fit region (left), $110 < m_{\mu\mu} < 150 \text{ GeV}$, and to the Z control region (ZCR, right), i.e. in events

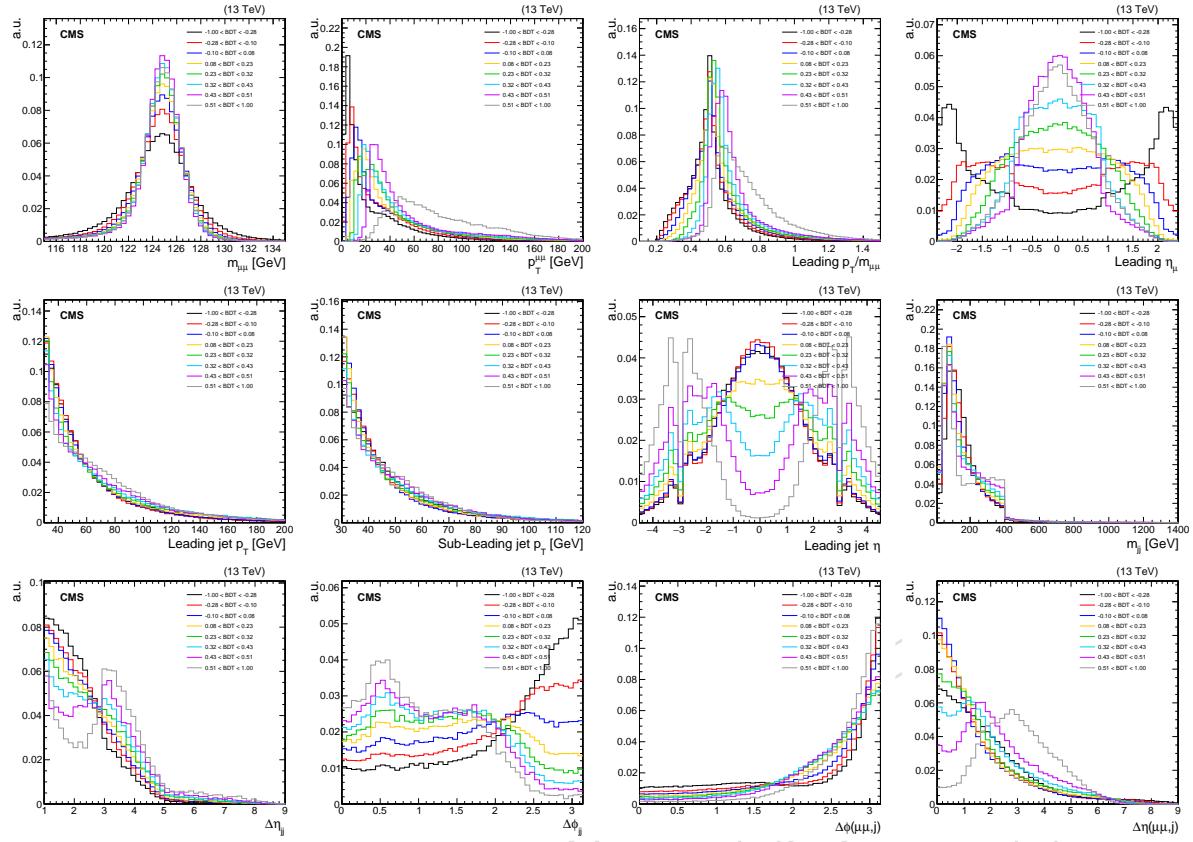


Figure 6.7: Evolution of the distributions of some of the key kinematic observables as a function of the ggH BDT output for signal events (ggH, VBF, VH, and ttH summed together). Distributions are scaled in order to have unit area. On top, from left to right: $m_{\mu\mu}$, $p_T^{\mu\mu}$, $p_T^{\mu_1}/m_{\mu\mu}$, and η_{μ_1} . Middle, from left to right: $p_T^{j_1}$, $p_T^{j_2}$, η^{j_1} , and m_{jj} . Bottom, from left to right: $\Delta\eta_{jj}$, $\Delta\phi_{jj}$, min- $\Delta\phi(\mu\mu, j)$, and min- $\Delta\eta(\mu\mu, j)$.

passing the analysis selections with $m_{\mu\mu}$ between 70 and 110 GeV. The ZCR is used in order to test more precisely how the simulation models the data in the high score region, in fact its statistical power is about one-hundred times larger than the fit region. The observation and the expectation in the three eras (2016, 2017, and 2018) have been summed together. In the ratio plot, the grey band corresponds to the uncertainty coming from the limited statistics of the simulated samples, while the azure one corresponds to the sum in quadrature between statistical and systematic uncertainties from experimental sources. The uncertainty due to jet energy resolution has been symmetrized in this plot. However, it will be implemented as a one-sided uncertainty in the datacards as explained later in this document. Finally, the orange band includes an estimate of the theoretical uncertainties, which is added in quadrature to the statistical and the experimental systematic uncertainties. Overall, a good agreement between data and MC is observed in both cases and most of the discrepancies are covered by systematic uncertainties assigned to the background prediction. An excess in data of about 10-15% resides in the high score region (BDT-score larger than 0.65) of the ggH BDT. The high score part of the ggH BDT integrates mainly 1-jet events in which the jet belongs to the endcap region, or two-jet events with low m_{jj} but either a small $\Delta\phi_{jj}$ or a large min- $\Delta\eta(\mu\mu, j)$. The simulation does not provide a perfect description of data in these regions of the phase space as reported in Fig. 6.4. Therefore, the origin of the observed trend is understood and mainly related to jet occupancy in the forward part of the endcaps.

Comparisons are also reported independently for each year in Fig. 6.11, where the first column

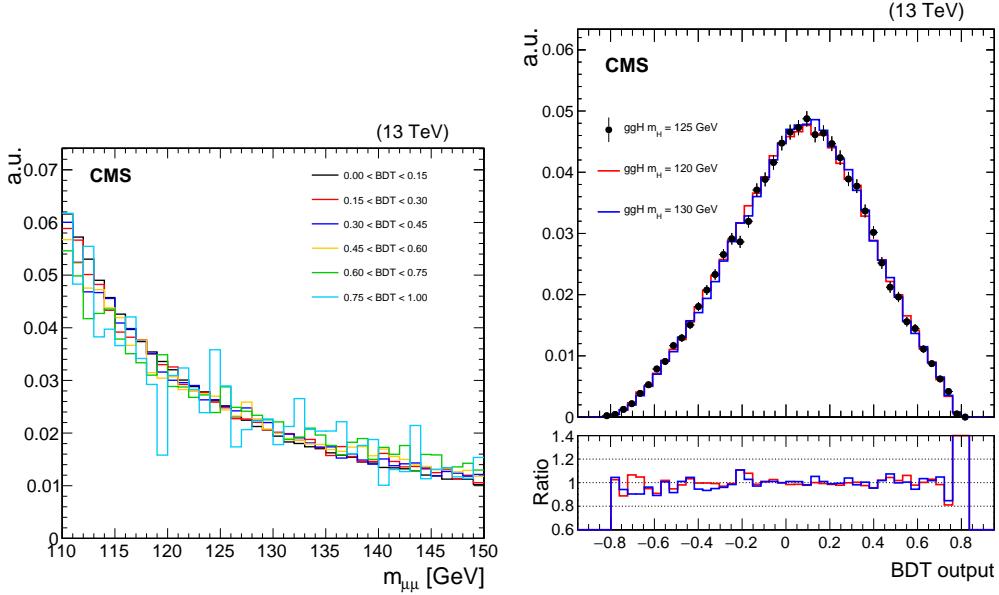


Figure 6.8: On the left, dimuon mass distribution in simulated background events computed in a set of orthogonal ranges defined along the ggH BDT score. On the right, distribution of the ggH BDT output in ggH simulated signal events for three different mass hypothesis: $m_H = 125$ GeV (black), $m_H = 120$ GeV (red), and $m_H = 130$ GeV (blue). Distributions are scaled in order to have unit area.

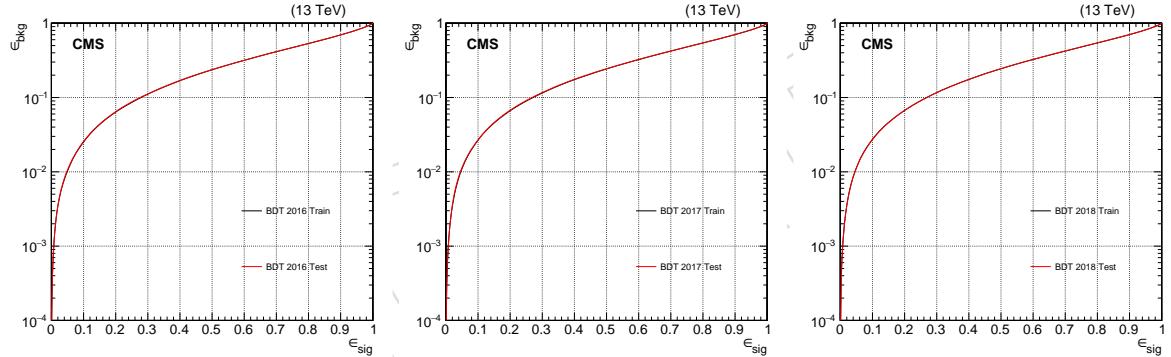


Figure 6.9: Signal efficiency (ϵ_{sig}) vs background efficiency (ϵ_{bkg}) ROC curves obtained in each eras from the ggH BDT output scores for events in the training (black) and testing (red) samples. The ROC curves shown in the first column refers to the 2016 era, while in the second and third ones the ROCs for 2017 and 2018 eras are reported. No differences in performance are observed between training and testing events.

1265 corresponds to the 2016, the second to the 2017, and the third to the 2018 era. The first row
 1266 reports the ggH BDT outputs in the signal fit region, while the second row refers to the ZCR.
 1267 A good agreement between data and MC is observed across all eras in both regions. Overall,
 1268 the trends observed in the ZCR are similar to those observed and already described for the
 1269 signal region. The excess in data observed in the high score region of the ggH BDT reported in
 1270 Fig. 6.10 originates from 2017 and 2018 eras, where the description of endcap jets from simula-
 1271 tion is known to be not as good as in 2016.

1272 6.2.1 Category optimization

1273 Based on the output of the trained BDT discriminant reported in Fig. 6.10, we can further cat-
 1274 egorize events belonging to the ggH phase space. In order to do that, we need to set up a

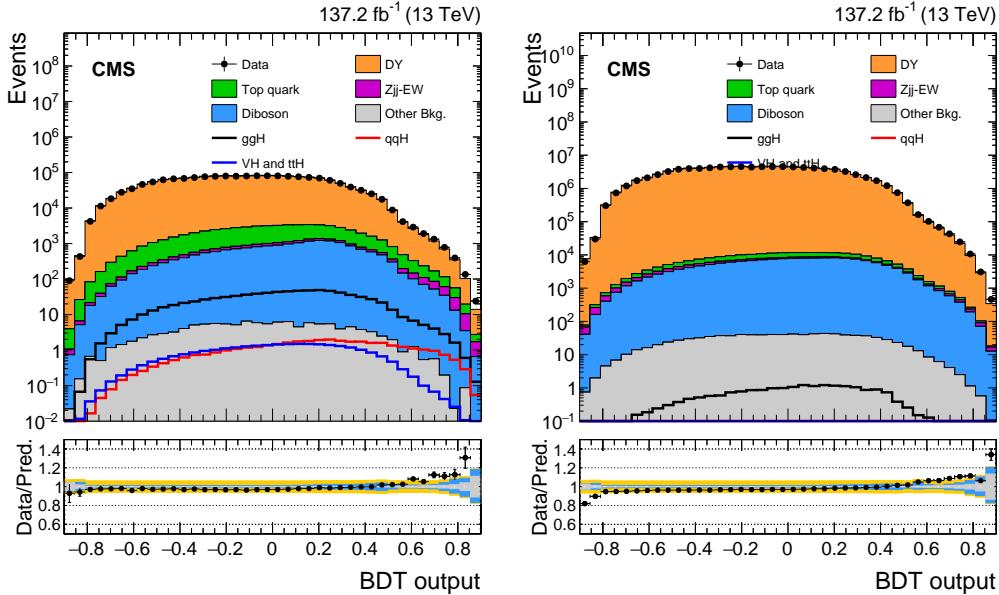


Figure 6.10: Distribution of ggH BDT scores reported for both data (black points) and simulation (solid histograms) for events in the signal fit region (left), where $m_{\mu\mu}$ is between 110 and 150 GeV, and in the ZCR (right), $70 < m_{\mu\mu} < 110$ GeV. The observation and the expectation in the three eras (2016, 2017, and 2018) have been summed together. In the ratio plot, the grey band indicates the uncertainty coming from the limited statistics of the simulated samples, the azure one corresponds to the sum in quadrature between statistical and systematic uncertainties from experimental sources, while the orange band includes the theoretical uncertainties in the background prediction.

procedure and choose a figure of merit to decide where to put a boundary along the BDT output. The designed procedure is iterative and the optimal boundaries are obtained by summing up the signal and background contributions expected from the three eras.

We start with a single category collecting all the events belonging to the ggH phase space, and we compute the expected significance for a Higgs boson with $m_H = 125$ GeV. This is obtained by modelling the $m_{\mu\mu}$ distribution of the total signal (all production modes are added together into a single distribution) with a double-sided Crystal Ball function. The mass range of the fit is chosen to be 110–150 GeV. In this optimization procedure, ggH events from the MADGRAPH5_aMC@NLO generator are considered because these simulated samples are the ones used for the final signal extraction. In order to avoid an over-optimization, MADGRAPH5_aMC@NLO events with odd event-number are considered in this optimization procedure, while even ones are used to build the signal model in each category for the final signal extraction. In contrast, VBF events are taken from the centrally produced POWHEG samples showered with PYTHIA and only even event-number are considered, which correspond to those belonging to the MVA testing sample. The total expected background yield is also obtained from simulation (background simulated sample have been listed in Section 2) and we model the shape of the dimuon mass distribution in background events via an analytical function called BWZRedux, defined in Eq. 5.3, where m_Z and Γ_Z are the nominal values for peak and width of the Z boson line-shape, respectively. From earlier iterations of this analysis, the BWZRedux function is known to fit the background very well across the full phase-space.

After this, we compare the significance obtained with a single category with the one obtained by dividing the events into two categories. Since the BDT shapes for signal and background events are slightly different across eras, as shown in Fig. 6.5, the BDT output in the ggH is scanned in signal quantiles, i.e. in steps of 1% in signal efficiency. The optimal position for the

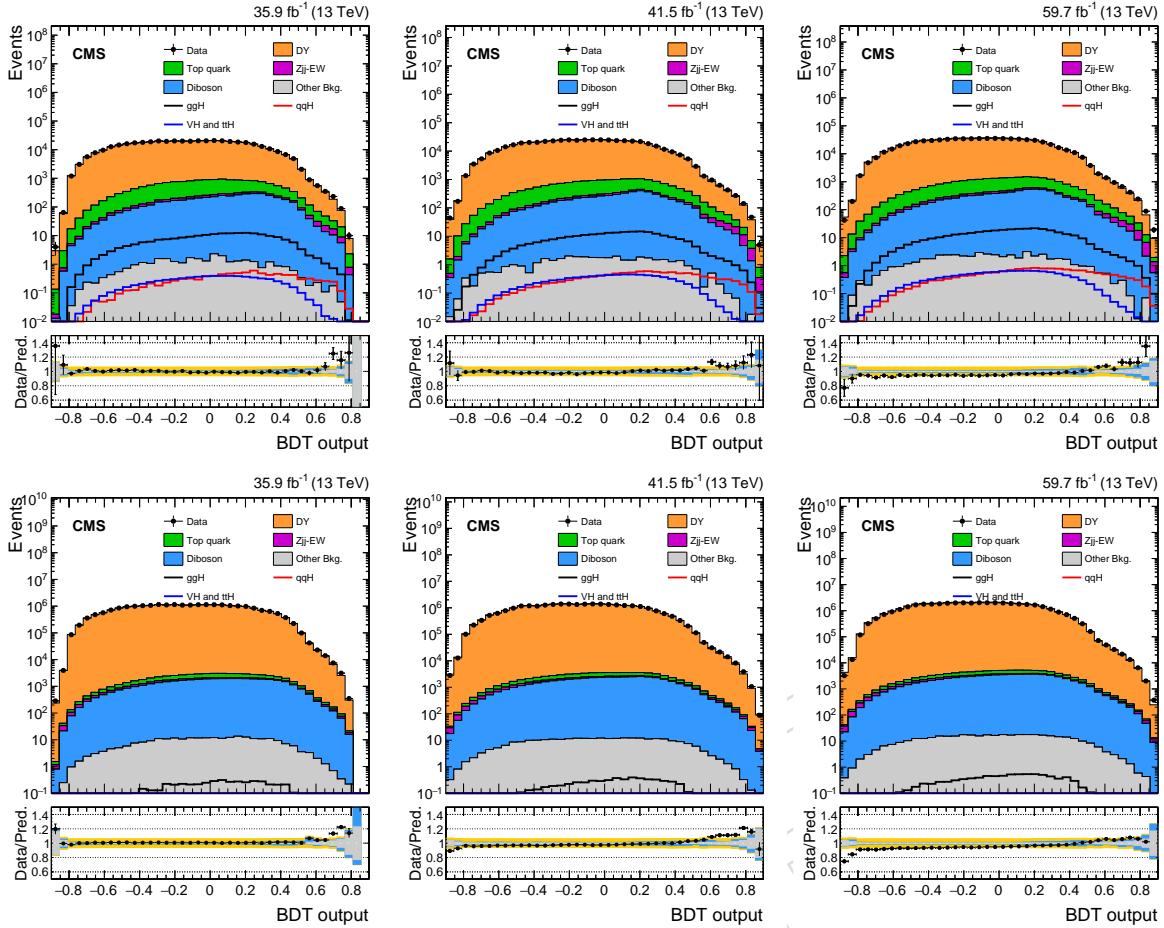


Figure 6.11: Distribution of the ggH BDT score in 2016 (left), 2017 (middle) and 2018 (right) eras reported for both data (black points) and simulation (solid histograms) for events with $m_{\mu\mu}$ between 110 and 150 GeV (first row) and for events in the ZCR (second row). In the ratio plot, the grey band indicates the uncertainty coming from the limited statistics of the simulated samples, the azure one corresponds to the sum in quadrature between statistical and systematic uncertainties from experimental sources, while the orange band includes the theoretical uncertainties in the background prediction.

category boundary corresponds to the one that gives the maximum significance obtained from a combined fit of the two categories, where the total signal and background contributions are modelled as described above. Once the optimal boundary has been found and two categories are defined, the iterative algorithm searches for a third optimal category, etc. This procedure stops when the n -th category does not provide an additional gain in significance larger than 1%.

By using this iterative strategy, five event categories are defined. The category boundaries defined in terms of signal efficiency quantiles and BDT ranges for the ggH BDT are reported in Table 6.1. However, since predictions from MADGRAPH5_aMC@NLO and POWHEG generators for the ggH process can be different by a few % depending on the phase space, we have repeated the same optimization procedure using POWHEG signal samples instead of MADGRAPH5_aMC@NLO ones. The same number of optimal categories are found, and the corresponding ranges in signal efficiency are in agreement within $\pm 2\%$, suggesting that the optimal categories listed in Table 6.1 are robust and universal, i.e. not specific to a particular MC generator.

Figure 6.12 shows the evolution of the combined expected significance from categories defined

Table 6.1: Optimized category boundaries defined along the output of the ggH BDT discriminant.

Name	Signal eff. range	BDT range 2016	BDT range 2017	BDT range 2018
ggH-cat0	[100, 70)%	[-1.00, -0.125)	[-1.00, -0.113)	[-1.00, -0.115)
ggH-cat1	[70, 35)%	[-0.125, 0.185)	[-0.113, 0.191)	[-0.115, 0.191)
ggH-cat2	[35, 20)%	[0.185, 0.309)	[0.191, 0.315)	[0.191, 0.312)
ggH-cat3	[20, 5)%	[0.309, 0.471)	[0.315, 0.498)	[0.312, 0.496)
ggH-cat4	$\leq 5\%$	[0.471, 1.00)	[0.498, 1.00)	[0.496, 1.00)

1315 along the ggH BDT output during the optimization procedure. The first plot on the left reports
 1316 the trend of the significance as a function of the first boundary, defined in terms of signal ef-
 1317 ficiency quantiles. As described before, expectations in signal and background from the three
 1318 eras are summed together. The significance is maximized by dividing the ggH phase space
 1319 into two categories in a signal efficiency range between. Moreover, the other graphs reported
 1320 in Fig. 6.12 show how the other category boundaries listed in Table 6.1 have been defined dur-
 1321 ing the iterative scan. The combined significance extracted during the optimization procedure
 1322 from the five ggH categories is of about 1.63σ .

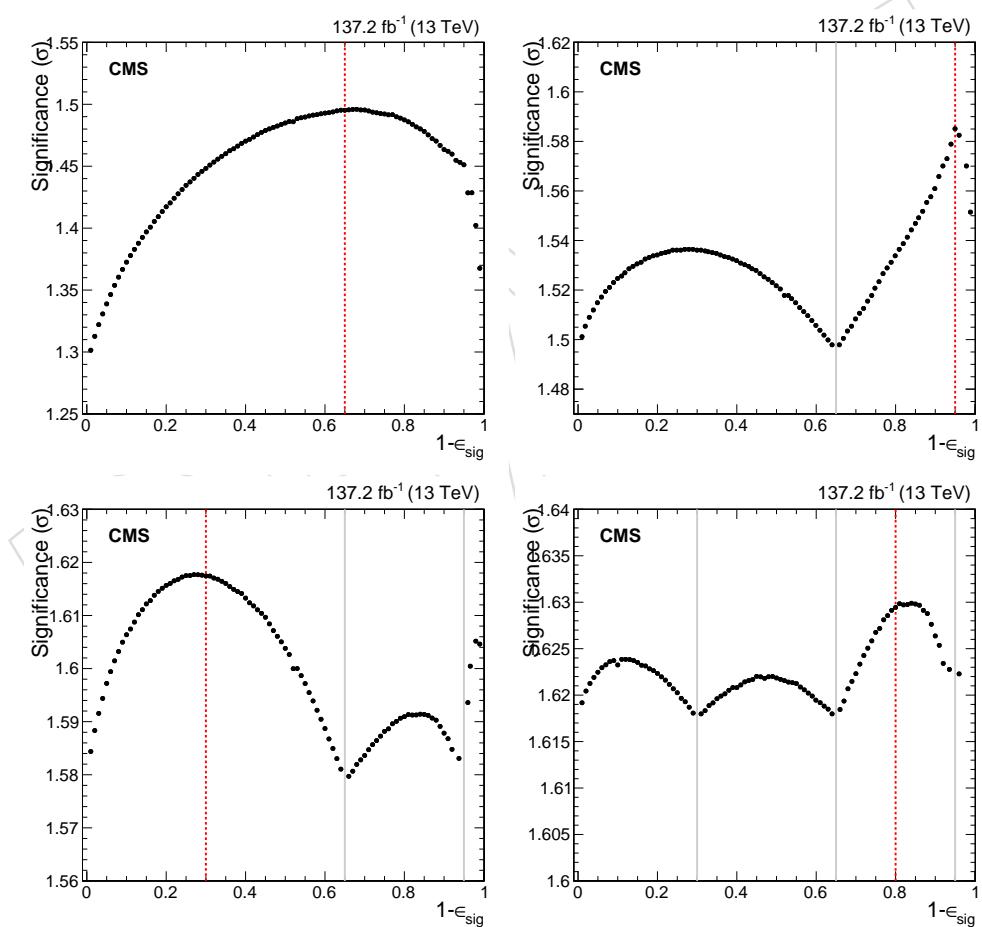


Figure 6.12: Evolution of the combined significance for a Higgs boson signal with $m_H = 125\text{ GeV}$ defined by scanning the ggH BDT output in signal efficiency quantiles. Top row, from left to right: improvements in the total expected significance by using two categories instead of one, and three categories instead of two. Second row, from left to right: improvements in the total expected significance with four categories instead of three, and five optimized categories instead of four. Dashed red lines indicate the category boundary giving optimal performance, while the solid grey ones indicate already defined boundaries from previous steps of the iterative procedure.

1323 Figure 6.13 shows a comparison between the outputs of the BDT discriminant for signal and
 1324 background simulated events. The black dashed vertical lines indicate, approximately, the cat-
 1325 egory boundaries optimized with the iterative procedure previously described. In addition,
 1326 inside the lower pad of each figure, the significance expressed as S/\sqrt{B} of each category, cal-
 1327 culated by integrating signal and background expected events inside the half width at half
 1328 maximum (HWHM, where HWHM=FWHM/2) of the expected signal peak, is also reported.

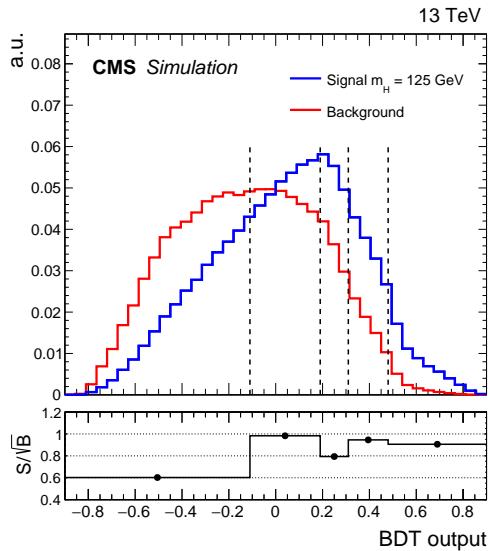


Figure 6.13: Distribution of the ggH BDT output in signal (blue) and background (red) simulated events. The black dashed vertical lines indicate, approximately, the category boundaries optimized with the iterative procedure previously described. In the lower pad, the significance expressed as S/\sqrt{B} of each category, calculated by integrating signal and background expected events inside the HWHM of the expected signal peak, is also reported.

1329 Finally, Fig. 6.14 show the distributions of the dimuon invariant mass in the five ggH categories
 1330 for data and simulated events in the range $110 < m_{\mu\mu} < 150 \text{ GeV}$. The agreement between
 1331 data and simulation in both shape and normalization is good across all categories apart from a
 1332 normalization offset of about 5% observed in the last category. However, since the background
 1333 in this search is directly extracted from data, as described later in Section 6.4, this discrepancy
 1334 in normalization does not have any effect on the analysis results.

1335 6.3 Signal extraction

1336 In this search the signal is extracted by fitting the dimuon invariant mass spectrum in data via
 1337 analytical functions able to model the expected signal and background processes. In the case
 1338 of the signal (ggH, VBF, VH, and ttH), these analytical functions contain several parameters
 1339 whose values are determined from simulation and are not allowed to float in the final fit, i.e.
 1340 they are treated as constants or constrained parameters when the fit to the data is performed.
 1341 In the case of the background, we assume no prior knowledge of the parameters of the fit
 1342 functions, therefore they are allowed to float freely in the signal extraction fit. The overall
 1343 normalizations of the signal and the backgrounds are also floated freely.

1344 6.3.1 Signal lineshape and yields

1345 The signal line shape is modelled using a double-sided Crystal Ball (DCB) function defined in
 1346 Equation 4.2. In each of the five categories listed in Table 6.1, we fit the signal contribution

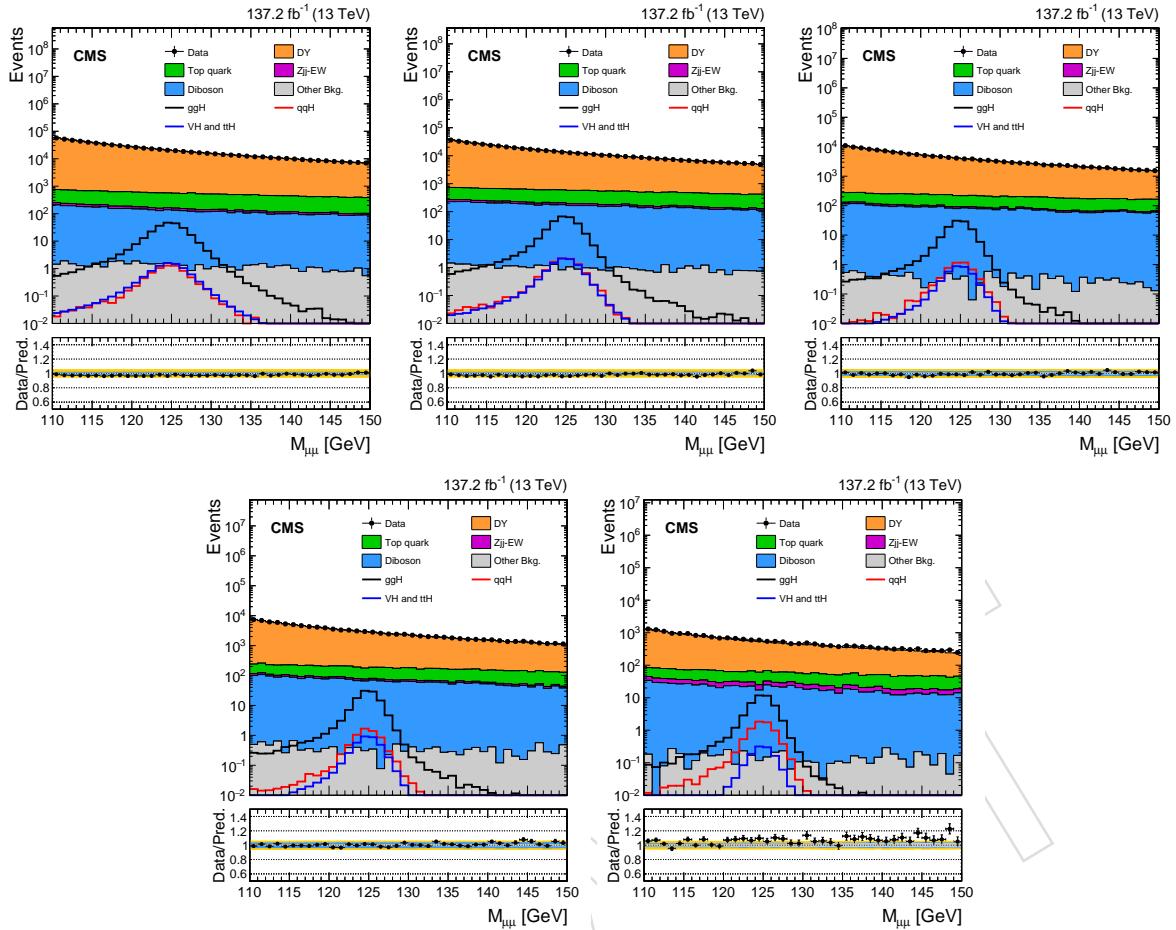


Figure 6.14: Distribution of the dimuon invariant mass in the ggH categories reported for both data (black points) and simulation (solid histograms) for events with $m_{\mu\mu}$ between 110 and 150 GeV. Top row: ggH-cat0 (left), ggH-cat1 (middle), and ggH-cat1 (right). Bottom row: ggH-cat3 (left), and ggH-cat4 (right). The expected signal distributions from the different Higgs boson production modes are also reported. In the ratio plot, the grey band indicates the uncertainty coming from the limited statistics of the simulated samples, the azure one corresponds to the sum in quadrature between statistical and systematic uncertainties from experimental sources, while the orange band includes the theoretical uncertainties in the background prediction.

from each the production modes (ggH, VBF, WH, ZH, and ttH) independently and we fix the corresponding shape parameters. Since POWHEG simulated events for ggH have been used in the ggH and VBF BDT trainings, both the lineshapes and the expected yields used in the ultimate signal extraction for this process are obtained from MADGRAPH5_aMC@NLO simulated samples. Predictions for the VBF production in the ggH BDT categories are taken instead from POWHEG interfaced with PYTHIA in dipole shower mode. Figure 6.15 shows the resulting signal shapes for the ggH process in ggH categories obtained by fitting the expected dimuon mass spectrum with a double-sided Crystal Ball, where the expectations in the three eras (2016, 2017, and 2018) have been summed together. Overall, the double-sided Crystal Ball provides a good description of both the bulk and the tails of the expected signal distributions in each category, $\chi^2/ndf \approx 1$, while one width parameter allows the shape uncertainties to be implemented in a straightforward way. A comparison of DCB with alternative models is presented in Appendix F.

The Higgs boson signal rates expected in 137.2 fb^{-1} of data are reported in Table 6.2 for each

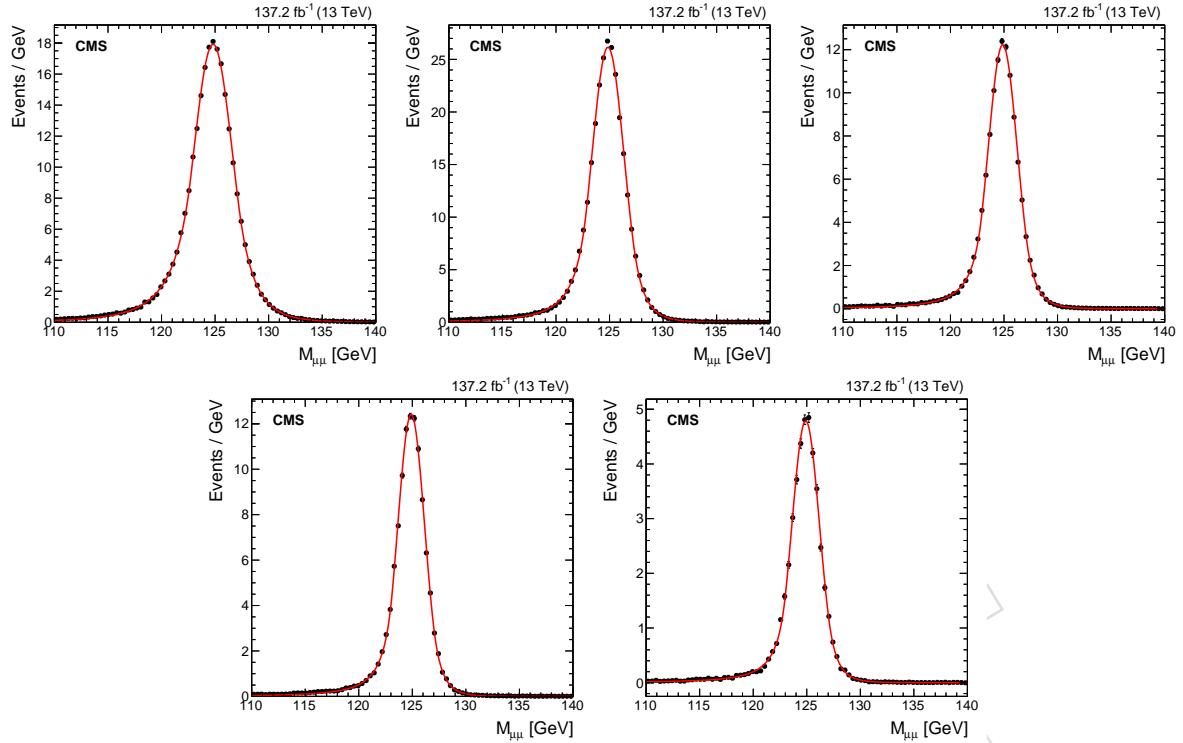


Figure 6.15: Distributions of the dimuon invariant mass for ggH signal events produced via the MADGRAPH5_aMC@NLO generator in the ggH BDT categories. Top row: cat0_inclusive (left), ggH-cat1 (middle), and ggH-cat2 (right). Bottom row: ggH-cat3 (left), and ggH-cat4 (right). The expectations in the three eras (2016, 2017, and 2018) have been summed together.

category defined along the ggH output, as well as for each Higgs boson production mode. The ggH signal represents the dominant contribution in all the categories.

Table 6.2: Higgs boson signal rates expected in 137.2 fb^{-1} of data in each category defined along the ggH and VBF BDT outputs for ggH, VBF, WH, ZH, and ttH production modes.

Category	ggH	VBF	WH	ZH	ttH	Total
ggH-cat0	254.52	7.81	5.48	3.42	0.27	271.53
ggH-cat1	295.21	10.84	5.83	3.67	0.19	315.74
ggH-cat2	124.04	5.35	2.19	1.50	0.07	133.14
ggH-cat3	116.49	7.03	2.07	1.60	0.07	127.26
ggH-cat4	45.42	7.70	0.60	0.61	0.02	54.36

The same procedure is also applied to alternative signal samples produced with a different Higgs boson mass hypothesis, namely $m_H = 120$ and 130 GeV . This allows to interpolate both the expected signal rates and the double-sided Crystal Ball shape parameters as a function of the Higgs boson mass within the mass range of $120\text{--}130 \text{ GeV}$. The interpolation vs m_H is performed via a smooth spline function. The interpolation is crucial in order to compute local p-values (upper limits) scanning the interesting mass range of a few GeV around the measured Higgs boson mass of 125.09 GeV . Figure 6.16 shows the interpolated signal shapes, normalized to the expected rate in 137.2 fb^{-1} of data, for the ggH process in the ggH categories. Similar trends are also observed for the other production modes i.e. VBF, WH, ZH, and ttH.

In addition, in order to better track the evolution of expected signal yield and shape as a function of m_H , the trends of the dimuon peak width (σ_{CB}) and of the acceptance times efficiency are reported for the ggH signal in the ggH categories in Fig. 6.17 and 6.18, respec-

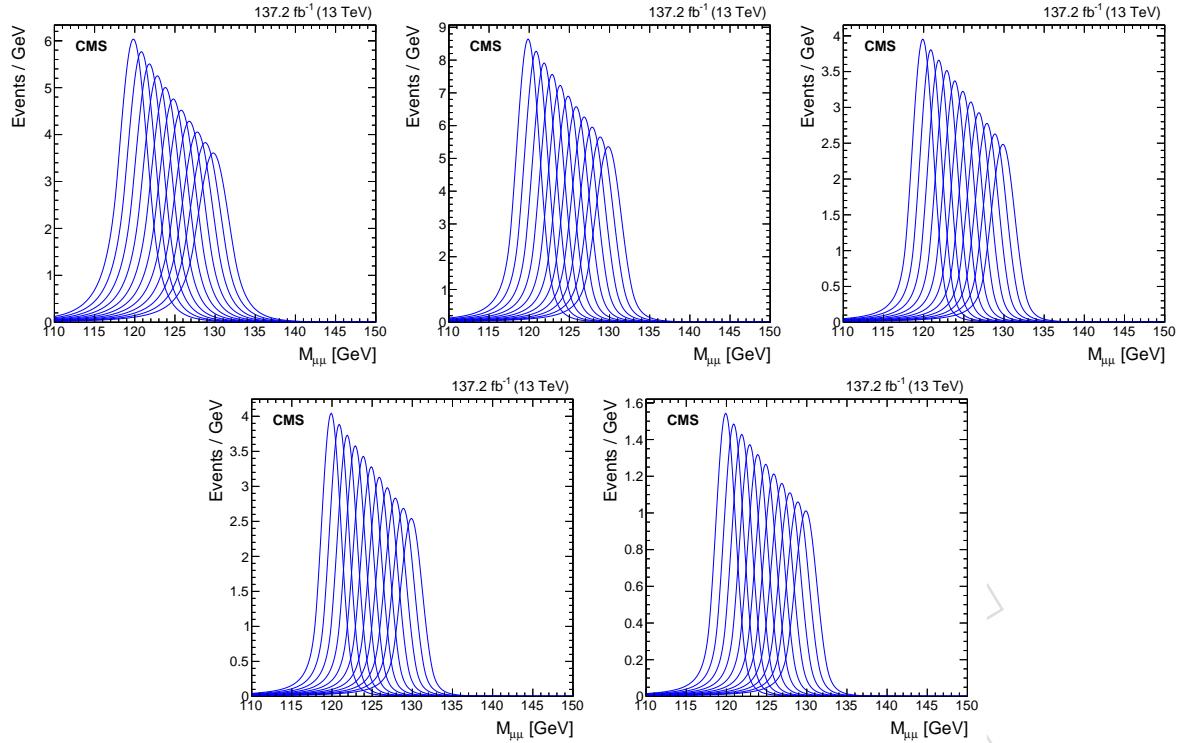


Figure 6.16: Distributions of the dimuon invariant mass for the ggH process in the ggH BDT categories as a function of m_H , resulting from an interpolation of the expected rates and shape parameters across the generated Higgs signal samples for $m_H = 120, 125$, and 130 GeV . Top row: ggH-cat0 (left), ggH-cat1 (middle), and ggH-cat2 (right). Bottom row: ggH-cat3 (left), and ggH-cat4 (right).

tively. As expected, the signal width increases linearly with the mass but the mass resolution, defined as σ_{CB}/m_H , remains constant. The product of acceptance and efficiency also slightly increases with the mass in each category. However, since $\sigma \times \mathcal{B}(H \rightarrow \mu\mu)$ decreases by about 30% moving from $m_H = 120$ to $m_H = 130 \text{ GeV}$, the total expected signal rate given by $\sigma \times \mathcal{B}(H \rightarrow \mu\mu) \times \text{acceptance} \times \text{efficiency}$ also decreases as already shown in Fig. 6.16.

6.3.2 Uncertainties on the signal model

Systematic uncertainties affecting of the expected signal prediction in each category are taken into account in the signal extraction and distinguished between *shape* and *rate* uncertainties. The *shape* uncertainties are those affecting the modelling of the signal peak, while the *rate* uncertainties are those affecting the predicted number of signal events in each category.

Shape uncertainties

For all Higgs boson production modes, the expected $m_{\mu\mu}$ signal shape is primarily affected by the uncertainties in the muon energy scale and resolution. The Rochester corrections, described in Section 4.6, are meant to correct for differences in both scale and resolution between data and simulated events. However, the uncertainties associated to the Rochester corrections, the uncertainties from the calibration of the per-event mass errors (closure test), as well as the assumption of a given analytical signal shape (double-sided Crystal Ball), have to be taken into account in the signal extraction.

The residual muon energy scale uncertainty has been estimated to be 0.1% of the muon momentum, and it affects the signal peak by changing its position w.r.t. the nominal expected value from simulation. This uncertainty is correlated across all signal production processes

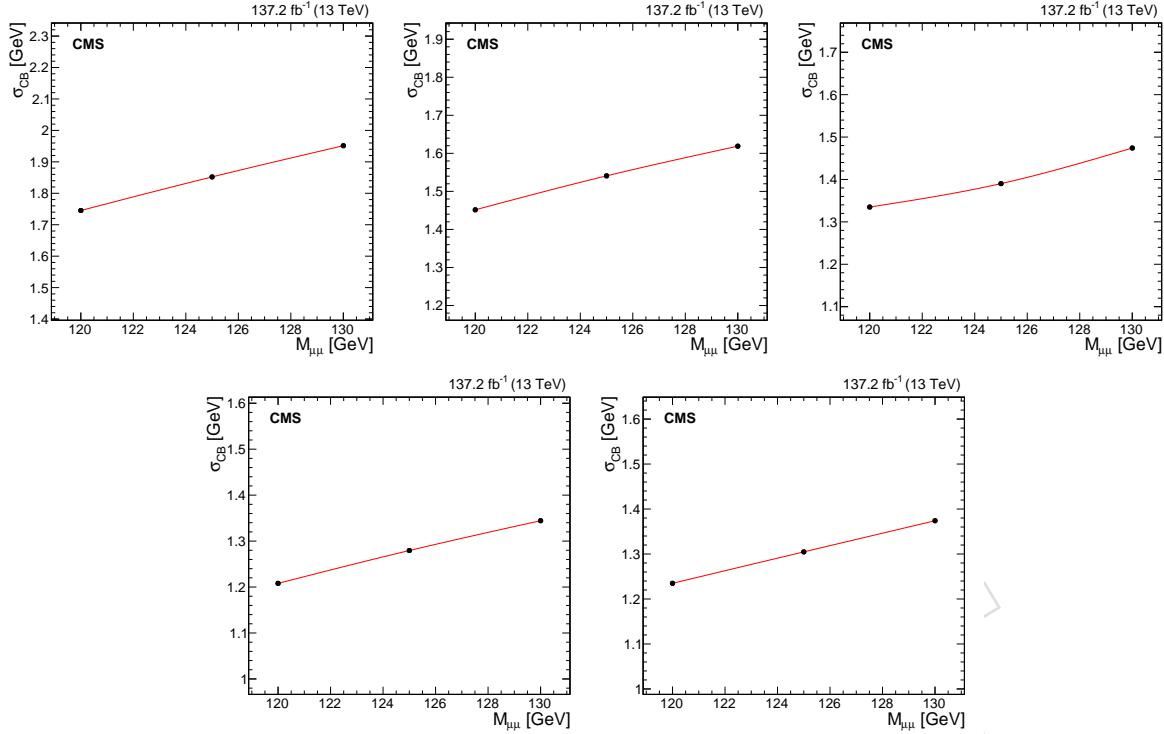


Figure 6.17: Evolution of the dimuon peak width (σ_{CB}) for the ggH process in the ggH BDT categories as a function of m_H , in the range $120 < m_{\mu\mu} < 130$ GeV. Top row: ggH-cat0 (left), ggH-cat1 (middle), and ggH-cat2 (right). Bottom row: ggH-cat3 (left), and ggH-cat4 (right).

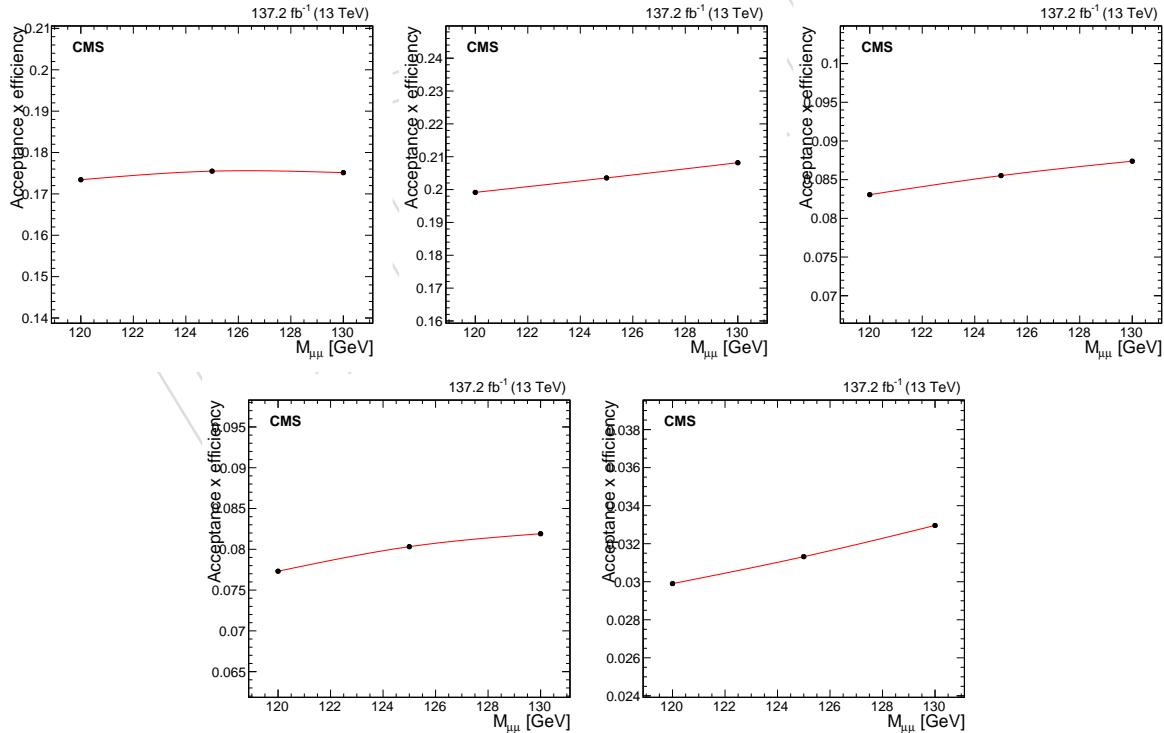


Figure 6.18: Evolution of the acceptance times efficiency for the ggH process in the ggH BDT categories as a function of m_H , in the range $120 < m_{\mu\mu} < 130$ GeV. Top row: ggH-cat0 (left), ggH-cat1 (middle), and ggH-cat2 (right). Bottom row: ggH-cat3 (left), and ggH-cat4 (right).

and eras, but decorrelated across categories. It is modelled via Gaussian constrained nuisance parameters. On the other hand, the uncertainty in the dimuon mass resolution has been estimated, conservatively, to be as large as 10% in each category. It is modelled via a Gaussian constrained nuisance parameter, correlated across processes and eras, but uncorrelated across categories whose effect is to shrink or enlarge the width of the signal peak (σ_{CB}).

Rate uncertainties from theory

Several sources from theory introduce an uncertainty in the predicted signal rate in each event category and for each production mode. Some of them affect the overall signal prediction, i.e. they affect either the production cross section or the value of $\mathcal{B}(H \rightarrow \mu\mu)$, and are referred to as *normalization* uncertainties. In contrast, others only affect the prediction of the acceptance in single categories, causing migrations of events across ggH categories, as well as migrations across ggH, VBF, and ttH channels. These are referred to as *acceptance* uncertainties.

As described in Section 2, the central values of the Higgs boson production cross section and $\mathcal{B}(H \rightarrow \mu\mu)$ are taken from Ref. [12]. The associated uncertainties, listed in Table 2.1, are modelled via constrained log-Normal nuisance parameters and they are fully correlated across categories and eras. The uncertainty on $\mathcal{B}(H \rightarrow \mu\mu)$ is also fully correlated across Higgs production modes. In addition, the uncertainty in the total cross section from PDF+ α_s can be correlated only between VBF and VH modes because both processes are quark-initiated.

The uncertainties affecting the ggH process from the sources defined by the LHC cross-section WG1, already introduced in Section 5.4, are estimated directly via the simulated sample used in the analysis. They are found to be in good agreement between MADGRAPH5_aMC@NLO and POWHEG generators. In addition, the uncertainty in the ggH acceptance due to the limited knowledge of the parton densities (PDF) has been also investigated. Table 6.3 summarizes the aforementioned per-category uncertainties for the ggH production mode, as well as how they are correlated across categories: positive sign means fully correlated, negative one means fully anti-correlated. The total uncertainty in the acceptance from these sources is of about 6–8% in the ggH categories. As expected, specific p_T^H and VBF phase-space uncertainties play either a limited or no role in the ggH BDT categories. Uncertainties related to p_T^H are also strongly anti-correlated in each category. Finally, uncertainties related to jet migrations (Mig-01, and Mig-12) are slightly anti-correlated between low-score ggH BDT categories (dominated by 0-jet events) and high-score ones (dominated by 1 and 2-jet events).

Table 6.3: Acceptance uncertainties for ggH signal events in ggH BDT categories from WG1 uncertainties and PDF. Uncertainties smaller than 0.1% are not reported.

Category	ggH-Mu	ggH-Res	Mig-01	Mig-12	VBF-2j	VBF-3j	Pt60	Pt120	Mtop	PDF
ggH-cat0	0.954/1.046	0.980/1.020	1.004/0.996	—	—	—	1.008/0.992	0.998/1.002	0.999/1.001	1.003/0.997
ggH-cat1	0.953/1.047	0.978/1.022	0.998/1.002	1.004/0.996	0.999/1.001	—	—	1.001/0.999	0.999/1.001	1.002/0.998
ggH-cat2	0.953/1.047	0.979/1.021	1.001/0.999	1.001/0.999	0.999/1.001	—	0.995/1.005	1.005/0.995	0.998/1.002	1.002/0.998
ggH-cat3	0.952/1.048	0.975/1.025	0.995/1.005	0.999/1.001	0.999/1.001	—	0.991/1.009	1.008/0.992	0.997/1.003	1.002/0.998
ggH-cat4	0.945/1.055	0.957/1.043	0.955/1.045	1.004/0.996	0.997/1.003	—	0.971/1.029	1.017/0.983	0.996/1.004	0.997/1.003

Furthermore, acceptance uncertainties designed for the VBF process and obtained following the LHC Higgs cross-section WG2 recommendations, as described in Section 5.4, are estimated directly via the simulated sample used in the analysis, and are found to be in good agreement between MADGRAPH5_aMC@NLO and POWHEG generators. In addition, the uncertainty in the VBF signal acceptance due to the limited knowledge of the PDFs has been also investigated. Table 6.4 summarizes the aforementioned per-category uncertainties for the VBF production mode, as well as how they are correlated across categories. The total uncertainty in the acceptance from these sources is of about 0.8–1.2% in the ggH categories. As expected from the coarse binning adopted in the STXS prescription, only the low- p_T^H uncertainty (Pt25) plays a role. For

what concerns m_{jj} uncertainties, those related to low m_{jj} bins yield to a 0.5–1.5% effect while $M_{jj}-700$, $M_{jj}-1000$, and $M_{jj}-1500$ can be neglected since the ggH categories are only sensitive to low m_{jj} VBF signal events. The effect of $M_{jj}-60$, $M_{jj}-120$, and $M_{jj}-350$ is correlated across ggH categories. Finally, the Jet-01 nuisance is observed to be, as expected, anti-correlated between low-score and high-score categories.

Table 6.4: Acceptance uncertainties for VBF signal events in ggH BDT categories from WG2 uncertainties and PDF. Uncertainties smaller than 0.1% are not reported.

Category	Pt25	Pt200	$M_{jj}-60$	$M_{jj}-120$	$M_{jj}-350$	$M_{jj}-700$	$M_{jj}-1000$	$M_{jj}-1500$	Jet-01	PDF
ggH-cat0	1.008/0.992	—	—	1.006/0.994	1.005/0.995	—	—	—	1.001/0.998	1.002/0.996
ggH-cat1	1.005/0.995	—	1.002/0.998	1.005/0.995	1.003/0.997	—	—	—	1.004/0.996	1.003/0.997
ggH-cat2	1.004/0.996	—	1.004/0.996	1.005/0.995	1.003/0.997	—	—	—	1.004/0.996	1.004/0.996
ggH-cat3	1.004/0.996	—	1.005/0.995	1.006/0.994	1.002/0.998	—	—	—	1.004/0.996	1.002/0.998
ggH-cat4	1.003/0.997	—	1.006/0.994	1.005/0.995	1.001/0.999	1.001/0.999	—	—	1.005/0.995	1.001/0.999

Uncertainties related to the parton shower model are derived by varying the per-event weights stored internally in each signal sample. Variations are those related to the CP5 tune [25] of PYTHIA (v8.2 or greater). Table 6.5 reports the parton shower related uncertainties in each category defined along the ggH BDT for the different Higgs boson production modes. Uncertainties are fully correlated across processes and eras. The parton shower uncertainty is small, at the level of 0.5–1.5%, for the VBF production, while for ggH ranges from 1.5–5.5% depending on the ggH category. The parton shower effect is anti-correlated between low-score and high-score ggH categories, which means that the associated nuisance parameter primarily induces migration of events across N_{jet} bins.

Table 6.5: Acceptance uncertainties for ggH, VBF, WH, ZH, and ttH signal events in ggH BDT categories from variations of the parton shower tune parameters. Uncertainties smaller than 0.1% are not reported.

Category	ggH	VBF	WH	ZH	ttH
ggH-cat0	1.044/0.975	0.992/1.003	0.934/1.018	0.931/1.022	0.979/1.016
ggH-cat1	1.005/0.994	0.996/1.002	0.955/1.011	0.956/1.010	0.979/1.025
ggH-cat2	0.974/1.015	0.998/1.001	0.971/1.010	0.930/1.012	0.995/1.001
ggH-cat3	0.947/1.032	1.001/0.997	0.957/1.012	1.006/0.998	0.978/1.015
ggH-cat4	0.879/1.054	1.010/0.987	0.967/1.014	1.009/0.998	1.006/0.975

Finally, the uncertainty related to variations of the underlying event tune is estimated by using the alternative signal samples listed in Section 2.2. Values of this uncertainty in each category and for each production mode are reported in Table 6.6, and are fully correlated across processes and eras. In order to minimize the impact of the limited size of MC statistics available for the signal samples generated with the Up/Down variations of the underlying event tune, the observed variations in the signal yield have been symmetrized. For both ggH and VBF processes, the estimated uncertainty ranges between $\pm 3\%$ across ggH categories.

Table 6.6: Acceptance uncertainties for ggH, VBF, WH, ZH, and ttH signal events in ggH BDT categories from variations of the underlying event tune parameters. Uncertainties smaller than 0.1% are not reported.

Category	ggH	VBF	WH	ZH	ttH
ggH-cat0	0.998/1.002	0.996/1.004	0.997/1.003	0.999/1.001	0.992/1.008
ggH-cat1	1.005/0.995	1.008/0.992	0.998/1.002	1.001/0.999	1.006/0.994
ggH-cat2	0.998/1.002	0.990/1.010	0.997/1.003	1.003/0.997	1.021/0.979
ggH-cat3	1.005/0.995	1.015/0.985	1.002/0.998	1.004/0.994	0.992/1.008
ggH-cat4	1.002/0.998	0.990/1.010	0.994/1.004	1.005/0.995	0.982/1.018

1457 Rate uncertainties from experimental sources

1458 Several experimental sources of uncertainty affect the expected signal rate in each event cate-
 1459 gory and for each production mode. A first set is represented by muon-related sources. Un-
 1460 certainties in the knowledge of muon identification and isolation efficiencies, as well those
 1461 related to the single muon trigger efficiency, are propagated in the analysis assuming full cor-
 1462 relation between the two muons in the event. Their effects are modelled via two independent
 1463 log-Normal nuisance parameters correlated across processes, categories, and eras. Table 6.7
 1464 reports their values described for each production mode as well as for each BDT category. The
 1465 uncertainty in the muon identification and isolation (single muon trigger) scale factors trans-
 1466 lates into an overall $\pm 1\%$ ($\pm 0.2\%$) uncertainty in the signal in each category. Moreover, the
 1467 effect of the uncertainty in the knowledge of the muon p_T coming from the Rochester correc-
 1468 tions is also taken into account and modelled via a log-Normal nuisance parameter correlated
 1469 across processes, categories, and eras. Table 6.8 reports the estimate for this uncertainty for
 1470 each production mode as well as for each BDT category. The muon p_T uncertainty induces a
 1471 few per-mile effect mostly anti-correlated as a function of the dimuon p_T spectrum, i.e. anti-
 1472 correlated moving from low to high score categories of the ggH BDT.

Table 6.7: Uncertainties for ggH, VBF, WH, ZH, and ttH processes in ggH BDT categories obtained from the uncertainties related to muon identification and isolation scale factors, and single muon trigger scale factors. Uncertainties smaller than 0.1% are not reported.

Category	Muon ID and Iso uncertainty					Muon trigger uncertainty				
	ggH	VBF	WH	ZH	ttH	ggH	VBF	WH	ZH	ttH
ggH-cat0	0.994/1.006	0.993/1.008	0.992/1.008	0.992/1.008	0.992/1.008	0.998/1.002	0.998/1.002	0.998/1.002	0.998/1.002	0.998/1.002
ggH-cat1	0.993/1.007	0.992/1.008	0.992/1.008	0.992/1.008	0.991/1.009	0.998/1.002	0.998/1.002	0.998/1.002	0.998/1.002	0.998/1.002
ggH-cat2	0.993/1.007	0.992/1.008	0.991/1.009	0.992/1.008	0.990/1.010	0.999/1.001	0.999/1.001	0.998/1.002	0.998/1.002	0.998/1.002
ggH-cat3	0.992/1.008	0.992/1.008	0.991/1.009	0.991/1.009	0.990/1.011	0.999/1.001	0.999/1.001	0.998/1.002	0.998/1.002	0.998/1.002
ggH-cat4	0.991/1.009	0.991/1.009	0.990/1.010	0.990/1.010	0.990/1.010	0.999/1.001	0.998/1.002	0.999/1.001	0.999/1.001	0.998/1.002

Table 6.8: Uncertainties for ggH, VBF, WH, ZH, and ttH processes in ggH BDT categories obtained from the uncertainties related to muon energy scale. Uncertainties smaller than 0.1% are not reported.

Category	Muon scale uncertainty				
	ggH	VBF	WH	ZH	ttH
ggH-cat0	1.002/0.998	1.001/0.998	1.001/0.999	1.001/0.999	1.003/0.998
ggH-cat1	1.001/0.999	1.001/0.999	1.001/0.999	1.001/0.999	1.002/0.999
ggH-cat2	0.999/1.002	0.998/1.002	0.999/1.001	0.998/1.002	0.999/1.001
ggH-cat3	0.998/1.002	0.998/1.002	0.998/1.002	0.999/1.001	0.999/1.001
ggH-cat4	0.998/1.003	0.999/1.001	0.998/1.003	0.998/1.002	0.999/1.003

1473 The dominant experimental uncertainties are represented by the limited knowledge in the en-
 1474 ergy scale and resolution of the hadronic jets in the event, especially for those belonging to the
 1475 forward region of the detector ($|\eta| > 2.5$). Each source of uncertainty is varied independently
 1476 up and down in order to change the p_T of all the jets found in the event. The analysis selec-
 1477 tions are re-applied after each variation, and the ggH BDT is re-evaluated to account for the
 1478 corresponding changes in the jet kinematics. Variations in the expected signal rate in each cat-
 1479 egory are modelled via Log-Normal nuisance parameters, whose effects are correlated across
 1480 categories, signal production modes, and eras (depending on the uncertainty source). Simi-
 1481 larly, the uncertainty in the jet energy resolution is used to smear the p_T of the jets in simulated
 1482 events in the six exclusive categories listed in Section 5.4. The corresponding variations in the
 1483 expected signal rate are modelled via independent log-Normal nuisance parameters, whose
 1484 effect is correlated across categories, signal processes, but not correlated across eras. The un-
 1485 certainties related to jet-energy-resolution are found to be one-sided because we don't apply
 1486 the resolution scale-factors (smearing) to correct for the central value, but energy smearing is
 1487 applied only to estimate an associated uncertainty (conservative).

1488 Therefore, following the recipe from JetMET POG, twenty-one nuisance parameters are con-
1489 sidered in the final fit which are either correlated or uncorrelated across eras. Table 6.9 and
1490 Table 6.10 list for each ggH BDT category the estimated values for each jet-energy-scale and
1491 jet-energy-resolution uncertainty for the ggH and VBF processes, respectively. Uncertainties in
1492 the jet energy scale cause significant migration of events between the ggH and the VBF cate-
1493 gories, therefore the effect of these nuisances is found to be largely anti-correlated across N_{jet}
1494 bins. The sources showing the largest impact on the predicted signal yields are those correlated
1495 across eras and not specific to a particular detector region: flavorQCD, relativeBal, relativeSam-
1496 ple, and absolute. According to Table 6.9, the total uncertainty due to jet-energy-scale affecting
1497 the ggH process ranges between 0.2–2.5% as a function of the ggH BDT-score. In contrast, the
1498 total uncertainty on the predict VBF signal yields due to jet-energy-scale is found to be larger
1499 than ggH in the ggH categories, ranging from 1.5 to 7%. On the other hand, the uncertainty
1500 due to jet energy resolution has a slightly smaller impact on the predicted signal yields. In the
1501 ggH categories, the total jet-energy-resolution uncertainty affecting the ggH (VBF) prediction
1502 ranges between 0.1–1% (0.1–2.5%) as a function of the BDT score.

1503 The jet energy scale and resolution uncertainties affecting the other production modes (WH,
1504 ZH, and ttH) are not reported in this note because they have a small impact on the final result.
1505 However, they are included in the “datacards” used in the final fit for the signal extraction and,
1506 overall, their size is similar to what has been measured for the ggH process and reported in
1507 Table 6.9.

1508 Other sources of uncertainty are represented by the uncertainty in the ECAL pre-firing correc-
1509 tion introduced to emulate an inefficiency in the L1 trigger efficiency affecting 2016 and 2017
1510 eras, and the uncertainty in the minimum bias cross section which affects the pileup re-weight
1511 procedure described in Section 4.1. The pre-firing uncertainty has a small impact ($\leq 0.3\%$) on
1512 the predicted signal rate in the ggH categories. The pileup uncertainty instead is found to be
1513 quite flat as a function of the ggH, ranging between 0.1% and 0.5%. Table 6.11 summarizes the
1514 estimated value of pre-firing and pileup uncertainties for each production mode and for each
1515 category. Finally, the uncertainty on the luminosity measurement is of about 2.5% in each era,
1516 and is modelled via a log-Normal nuisance parameter whose effect is correlated across BDT
1517 categories and processes.

Table 6.9: Uncertainties for the ggH process in both the ggH BDT categories obtained by propagating the effect of each source of uncertainty related to jet energy scale and jet energy resolution to the expected signal yields. Uncertainties smaller than 0.1% are not reported.

Source of Uncertainty	ggH Categories				
	cat0	cat1	cat2	cat3	cat4
Jet energy scale					
flavorQCD	0.994/1.006	—	1.009/0.991	1.012/0.989	1.013/0.988
relativeBal	—	0.999/1.001	0.998/1.003	0.996/1.004	0.997/1.002
relativeSample	—	0.999/1.001	0.995/1.005	0.995/1.005	0.997/1.002
absolute	0.998/1.002	1.001/0.999	1.008/0.992	1.008/0.992	1.003/0.996
BBEC1	0.998/1.002	0.999/1.000	1.002/0.998	1.004/0.996	1.005/0.996
EC2	—	—	1.001/0.999	1.001/0.999	—
HF	1.001/0.998	1.001/0.999	1.002/0.998	—	0.996/1.004
relativeSample_2016	—	—	1.001/0.998	1.002/0.998	1.001/0.999
absolute_2016	—	—	1.001/0.999	1.001/0.999	—
BBEC1_2016	—	—	—	1.001/0.999	1.001/0.999
EC2_2016	—	—	—	1.001/0.999	—
HF_2016	—	—	—	—	0.998/1.001
relativeSample_2017	—	1.001/0.999	1.002/0.999	1.001/0.999	1.001/0.999
absolute_2017	—	—	1.001/0.999	1.001/0.999	1.001/0.999
BBEC1_2017	—	—	—	—	1.001/0.999
EC2_2017	—	—	1.001/0.999	1.001/0.999	1.001/0.999
HF_2017	—	—	—	—	—
relativeSample_2018	0.999/1.001	1.001/0.999	1.004/0.995	1.006/0.994	1.005/0.994
absolute_2018	0.999/1.001	—	1.002/0.997	1.002/0.998	1.004/0.997
BBEC1_2018	—	—	—	—	1.001/0.999
EC2_2018	—	—	—	—	1.001/0.999
HF_2018	—	—	—	—	—
Jet energy resolution	cat0	cat1	cat2	cat3	cat4
jerBarrel	—/1.005	—	—/0.996	—/0.994	—/0.994
jerEC1	—/1.001	—	—/0.999	—/0.997	—/0.997
jerEC2low	—	—/0.999	—/0.997	—/0.997	—/0.995
jerEC2high	—	—	—/0.999	—	—/0.999
jerHFlow	—/0.998	—/0.999	—	—	—/1.004
jerHFhigh	—	—/0.999	—	—	—

Table 6.10: Uncertainties for the VBF process in both the ggH BDT categories obtained by propagating the effect of each source of uncertainty related to jet energy scale and jet energy resolution to the expected signal yields. Uncertainties smaller than 0.1% are not reported.

Source of Uncertainty	ggH Categories				
	cat0	cat1	cat2	cat3	cat4
Jet energy scale					
flavorQCD	0.997/1.001	1.009/0.996	1.013/0.980	1.014/0.978	1.026/0.976
relativeBal	0.996/1.004	0.993/1.014	0.982/1.008	0.984/1.006	0.982/1.022
relativeSample	0.991/1.008	0.993/1.017	0.977/1.011	0.982/1.016	0.976/1.026
absolute	1.001/0.997	1.015/0.990	1.019/0.977	1.016/0.974	1.029/0.972
BBEC1	0.998/1.001	1.002/0.999	0.999/0.997	1.004/0.996	1.004/0.995
EC2	—	1.002/0.999	1.002/0.995	1.001/0.998	1.004/0.997
HF	1.004/0.996	1.006/0.994	1.010/0.990	1.006/0.990	1.009/0.990
relativeSample_2016	1.003/1.000	1.007/0.998	1.000/0.992	1.002/0.995	1.011/0.992
absolute_2016	—	1.002/1.000	1.000/0.994	—	1.003/0.999
BBEC1_2016	—	1.001/0.999	—	1.000/0.997	—
EC2_2016	—	1.002/0.999	0.999/0.997	—	1.002/0.999
HF_2016	—	—	1.001/0.998	0.999/1.000	—
relativeSample_2017	1.002/0.997	1.005/0.997	1.005/0.997	1.006/0.995	1.009/0.993
absolute_2017	—	1.003/0.999	1.002/0.996	1.002/0.996	1.005/0.997
BBEC1_2017	—	—	—	—	1.001/0.999
EC2_2017	—	—	1.002/0.997	1.002/0.998	1.003/0.998
HF_2017	—	—	—	—	—
relativeSample_2018	1.004/0.996	1.008/0.993	1.015/0.986	1.013/0.986	1.020/0.978
absolute_2018	0.999/1.001	1.003/0.999	1.003/0.997	1.004/0.995	1.005/0.997
BBEC1_2018	—	—	1.001/0.998	1.001/0.998	1.002/0.999
EC2_2018	1.001/0.999	—	1.001/0.998	1.003/0.998	1.003/0.996
HF_2018	—	—	1.002/0.999	1.001/0.998	1.001/0.998
Jet energy resolution	cat0	cat1	cat2	cat3	cat4
jerBarrel	—/1.005	—	—/0.996	—/0.994	—/0.994
jerEC1	—/1.001	—	—/0.999	—/0.997	—/0.997
jerEC2low	—	—/0.999	—/0.997	—/0.997	—/0.995
jerEC2high	—	—	—/0.999	—	—/0.999
jerHFlow	—/0.998	—/0.999	—	—	—/1.004
jerHFhigh	—	—/0.999	—	—	—

Table 6.11: Uncertainties for ggH, VBF, WH, ZH, and ttH processes in ggH BDT categories obtained from the uncertainties related to pre-firing correction and pileup re-weight. Uncertainties smaller than 0.1% are not reported.

Category	Pileup re-weight					Pre-firing				
	ggH	VBF	WH	ZH	ttH	ggH	VBF	WH	ZH	ttH
ggH-cat0	1.001/0.998	1.001/0.999	1.002/0.998	1.002/0.997	1.004/0.996	1.002/0.998	1.004/0.996	1.003/0.997	1.004/0.996	1.005/0.995
ggH-cat1	1.001/0.999	1.001/0.998	1.002/0.997	1.003/0.997	1.005/0.995	1.001/0.999	1.002/0.998	1.001/0.999	1.001/0.999	1.004/0.996
ggH-cat2	1.001/0.999	1.001/0.999	1.002/0.998	1.002/0.998	1.006/0.994	1.001/0.999	1.002/0.998	1.001/0.999	1.001/0.999	1.003/0.997
ggH-cat3	1.001/0.999	1.001/0.999	1.002/0.998	1.002/0.998	1.005/0.995	1.001/0.999	1.003/0.997	1.001/0.999	1.001/0.999	1.003/0.997
ggH-cat4	1.002/0.997	1.002/0.997	1.001/0.999	1.001/0.998	1.006/0.994	1.002/0.998	1.005/0.995	1.002/0.998	1.001/0.999	1.005/0.995

6.4 Background estimation

Given the low signal-to-background ratio expected in most of the event categories defined in Section 6.2, the key aspect of this search is represented by a robust estimation of the $m_{\mu\mu}$ shape of the total background. The expected background in each category is mainly composed of $Z/\gamma^* + \text{jets}$ events, with smaller contaminations arising from $t\bar{t}$, single-top, and diboson processes. Processes involving dimuon decays of a Z boson represent more than 90% of the total expected background across the ggH phase space.

Table 6.12: Background composition in (%) expected in the ggH BDT categories for events in the signal fit region, $110 < m_{\mu\mu} < 150 \text{ GeV}$. Event yields are normalized to the measured luminosity of 137.2 fb^{-1} .

Category	$Z/\gamma^* + \text{jets}$	$Z + \text{jets EW}$	$t\bar{t}$	Diboson	Other (s-top, VVV, etc)
ggH-cat0	97.5%	—	1.5%	0.6%	0.4%
ggH-cat1	96.0%	0.1%	2.1%	1.2%	0.6%
ggH-cat2	95.0%	0.2%	2.3%	1.9%	0.6%
ggH-cat3	93.7%	0.3%	2.8%	2.2%	1.0%
ggH-cat4	88.8%	1.3%	4.6%	3.6%	1.7%

As described in Section 2.1, simulated samples for the main background processes have been generated with NLO QCD precision but they cannot be directly used to model the background for several reasons:

- Simulated samples are missing higher order corrections in both QCD (NNLO) and EW (NLO) theory which are known to have a sizable impact on the predicted dimuon mass spectrum.
- Theory uncertainties due to resummation effects, QCD scale variations, and PDF are typically large enough to be over-constrained by data. They primarily provide an acceptance uncertainty in each category but they cannot be sufficiently trusted to provide a shape uncertainty for the background $m_{\mu\mu}$ spectrum.

The $m_{\mu\mu}$ distribution for the total background in each category is expected to be smoothly falling, driven by the Breit-Wigner nature of the Z boson lineshape. In this context, we choose to estimate the background via a data driven approach, smoothing the dimuon mass spectrum in the signal fit region ($110 < m_{\mu\mu} < 150 \text{ GeV}$) via empirical functional forms. According to this strategy, the background contribution under the peak is therefore estimated from signal-free mass sidebands. We assume no prior knowledge of the parameters of the fit functions (in both shape and normalization), i.e. they are allowed to float freely in the fit to the data. Therefore, the uncertainty in the estimate of the background under the peak is mainly affected by the statistical power of the mass sidebands, as well as on the assumption of the background functional form. The functions used have been already defined in Section 5.4, they can be either Breit-Wigner-like parametrizations (physics inspired) or agnostic ones.

The number of degrees of freedom for the polynomial terms used to modulate some of the physics-inspired functions, as well as the order of agnostic models like Bernstein polynomials, series of power-laws and exponentials, are selected by means of a likelihood ratio test as recommended by the CMS statistics committee [75]. This test works as follows: assuming that, within a given family of functions, n parameters are enough to describe the $m_{\mu\mu}$ spectrum in a chosen category, we first fit the $m_{\mu\mu}$ distribution in data with n and $n+1$ degrees of freedom. Then, we compute $2 \cdot (\text{Log}\mathcal{L}_{n+1} - \text{Log}\mathcal{L}_n)$ and, under the assumption that this quantity is distributed according to the χ^2 distribution with one degree of freedom, the corresponding p-value is used to accept or reject the null hypothesis. If the p-value is smaller than 5%, the null hypothesis is rejected and we apply the same test to the function with $n+1$ parameters. In contrast, if the

1556 p-value is larger than 5%, a good compatibility is observed between the two models, therefore
 1557 the function with n degrees of freedom is selected.

1558 In the cases of $BWZ \times Bernstein(m_{\mu\mu})$, $FEWZ \times Bernstein(m_{\mu\mu})$, and $Landau \times Bernstein(m_{\mu\mu})$,
 1559 this test indicates that, depending on the category, either one or two or three polynomial
 1560 degrees of freedom are enough to have an accurate description of the $m_{\mu\mu}$ spectrum. On the other
 1561 hand, for what concerns sum of power-laws and exponentials, a series of order two ($n = 2$)
 1562 is selected via the likelihood ratio test. Finally, when Bernstein polynomials are used to fit the
 1563 $m_{\mu\mu}$ spectrum in each category, the likelihood ratio test suggests that from five to six degrees
 1564 of freedom are necessary in order to have a good description of the observed spectra. Given
 1565 that all the other families require about three shape degrees of freedom per category, plain
 1566 Bernstein polynomials appear as an unnatural set of functions for modelling the total expected
 1567 background and therefore they are not considered in the studies presented in later Sections.

1568 Table 6.13 reports for each category defined along ggH BDT output and for each family of func-
 1569 tions the chosen number of shape degrees of freedom along with the χ^2/dof obtained from a
 1570 background-only fit to the data. In each category and for each function the associated χ^2 is
 1571 computed with the binning reported in Fig. 6.19. In each category, the χ^2/dof is about one for
 1572 the chosen functions, indicating that they can well describe the dimuon mass spectrum in the
 1573 range $110 < m_{\mu\mu} < 150$ GeV. Moreover, Fig. 6.19 shows the results of b-only fits performed
 1574 independently in the ggH categories with the functions listed in Table 6.13. The lower pad
 1575 in each plot reports the ratio between the background prediction obtained with the aforemen-
 1576 tioned functions and the one coming from a fit via the BWZRedux.

Table 6.13: Number of shape degrees of freedom selected, in each category and for each family of functions previously described, by means of a likelihood ratio test. The χ^2/dof obtained for each function from a background-only fit to the data in the range $110 < m_{\mu\mu} < 150$ GeV and computed following the recommendations of Ref. [76] is also reported.

Category	$BWZ \times Bernstein$	$BWZGamma$	$BWZRedux$	$FEWZ \times Bernstein$	$Landau \times Bernstein$	S-exponential	S-power-law
ggH-cat0	3-dof, $\chi^2 = 1.07$	2-dof, $\chi^2 = 1.05$	3-dof, $\chi^2 = 1.09$	2-dof, $\chi^2 = 1.07$	3-dof, $\chi^2 = 1.07$	3-dof, $\chi^2 = 1.12$	3-dof, $\chi^2 = 1.06$
ggH-cat1	3-dof, $\chi^2 = 1.10$	2-dof, $\chi^2 = 1.05$	3-dof, $\chi^2 = 1.13$	2-dof, $\chi^2 = 1.09$	3-dof, $\chi^2 = 1.10$	3-dof, $\chi^2 = 1.09$	3-dof, $\chi^2 = 1.06$
ggH-cat2	3-dof, $\chi^2 = 1.18$	2-dof, $\chi^2 = 1.16$	3-dof, $\chi^2 = 1.19$	2-dof, $\chi^2 = 1.16$	3-dof, $\chi^2 = 1.19$	3-dof, $\chi^2 = 1.16$	3-dof, $\chi^2 = 1.16$
ggH-cat3	3-dof, $\chi^2 = 0.84$	2-dof, $\chi^2 = 0.84$	3-dof, $\chi^2 = 0.84$	2-dof, $\chi^2 = 0.84$	3-dof, $\chi^2 = 0.84$	3-dof, $\chi^2 = 0.84$	3-dof, $\chi^2 = 0.84$
ggH-cat4	3-dof, $\chi^2 = 0.95$	2-dof, $\chi^2 = 0.95$	3-dof, $\chi^2 = 0.95$	2-dof, $\chi^2 = 0.95$	3-dof, $\chi^2 = 0.97$	3-dof, $\chi^2 = 0.96$	3-dof, $\chi^2 = 0.94$

1577 In each category and for each function selected by the F-test and reported in Table 6.13, it is
 1578 important to check the behaviour of the post-fit shape uncertainty on the total background as
 1579 a function of $m_{\mu\mu}$, as obtained from a background-only fit to the data. These comparisons are
 1580 reported in the Appendix of Ref. [17]. As expected, the relative size of the post-fit shape uncer-
 1581 tainties is proportional to the amount of data events in the category as well as to the number
 1582 degrees of freedom of the chosen analytical function. Breit-Wigner and power-law based func-
 1583 tion not modulated by polynomials show a “pole” in the mass region between 120–130 GeV,
 1584 therefore they are expected to lead to a better performance in terms of significance. On the
 1585 other hand, functions like $FEWZ \times Bernstein$, $BWZ \times Bernstein$, and sum of exponentials are
 1586 showing a shape uncertainty less dependent on $m_{\mu\mu}$ depending on the number of polynomial
 1587 degrees of freedom. However, it is important to remind that background shape uncertainties
 1588 have a “important” effect only in categories with high data statistics, i.e. those for which the
 1589 total background normalization can be precisely constrained by the $m_{\mu\mu}$ sidebands.

1590 Finally, Table 6.14 reports the expected significances in each category for a SM Higgs boson
 1591 with $m_H = 125$ GeV, extracted from a s+b fit to Asimov toys in which the background pre-
 1592 dictions are taken from b-only fits to data with the functions reported in Table 6.13. Similarly,
 1593 Table 6.15 reports the expected significances in each category extracted from s+b fit to Asimov
 1594 toys in which the background predictions are taken from simulated events. The expected sig-

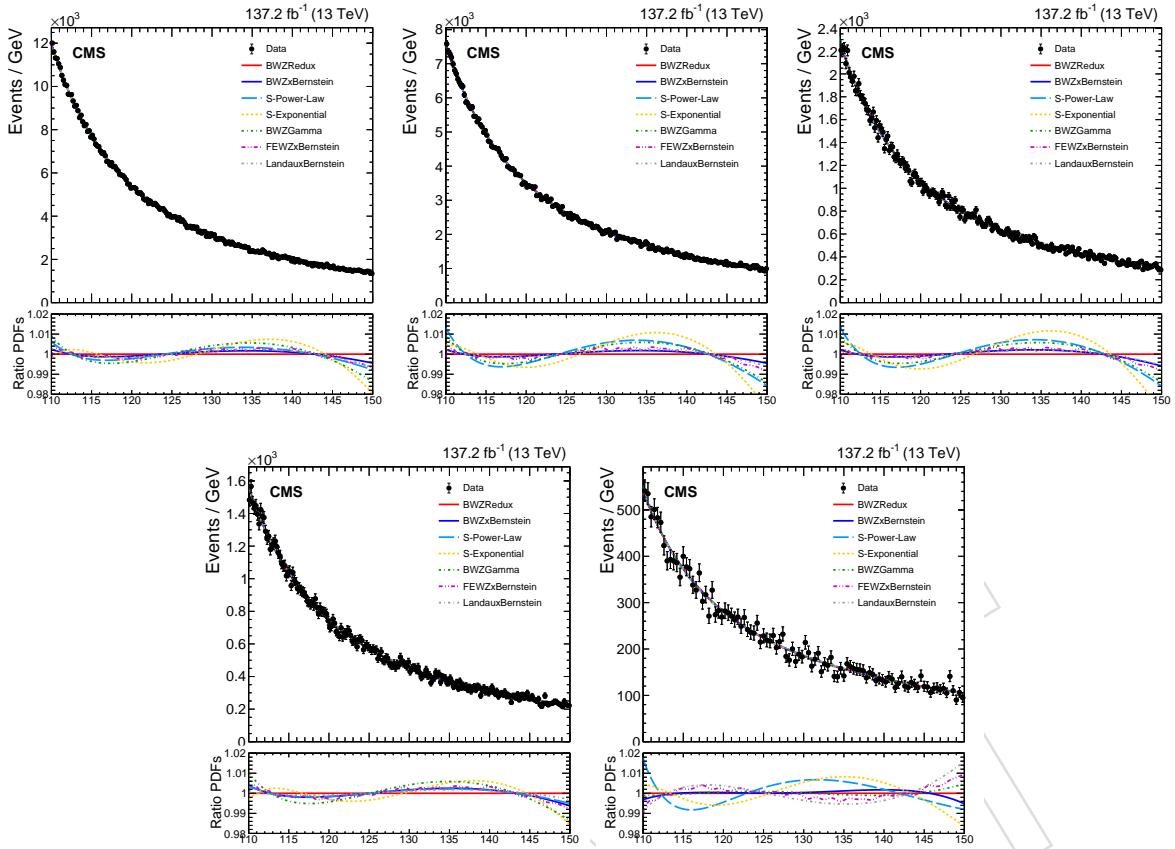


Figure 6.19: Dimuon invariant mass distributions in data for events belonging to the ggH BDT categories. The solid lines correspond to alternative background predictions obtained by fitting the mass spectra, under the background-only hypothesis, with the functional forms listed in Table 6.13. Top row, from left to right: ggH-cat0, ggH-cat1, and ggH-cat2. Bottom row, from left to right: ggH-cat3 and ggH-cat4. The lower pad in each plot reports the ratio between the background prediction obtained with each function and the one coming from a fit via the BWZRedux.

significance obtained from simulated events is used as figure of merit when choosing of the best performing functional form in each category. The prediction from simulation is found to be in good agreement with data as shown in Fig. 6.14. Across the ggH categories, BWZRedux, S-exponential, and S-power-law are showing very similar performances, compatible within $\pm 1\%$. The BWZGamma improves the performance of the BWZRedux only by few per-cent in some categories. The FEWZ \times Bernstein instead shows a consistent improvement of about 1–4% across all categories. However, in all categories the background parametrization which provides the best expected significance is represented by the BWZ \times Bernstein, whose performance improves the one obtained with the BWZRedux function by about 1–5%. These results, along with the outcome of the bias studies presented in Section 6.4.1, will be used to choose the background fit function in each category. In fact, in each category, the function providing the best expected significance without inducing a significant bias in the estimate of the signal strength is the one chosen for the final fit.

6.4.1 Uncertainty on the background model

As presented in Table 6.13, the functional forms used to model the dimuon mass distribution in background events belong to two main classes: physics-inspired models and agnostic ones. Within these families, the choice of particular analytical function to model the background in

Table 6.14: A posteriori expected significance in each category and for each family of functions obtained for a SM Higgs boson with $m_H = 125$ GeV. Systematic uncertainties on the signal prediction have not be added into the datacards used to extract these significances, however their impact is expected to be quite small at the level of few %.

Category	BWZ × Bernstein	BWZGamma	BWZRedux	FEWZ × Bernstein	Landau × Bernstein	S-exponential	S-power-law
ggH-cat0	$\sigma = 0.48$	$\sigma = 0.47$	$\sigma = 0.45$	$\sigma = 0.49$	$\sigma = 0.47$	$\sigma = 0.47$	$\sigma = 0.48$
ggH-cat1	$\sigma = 0.88$	$\sigma = 0.85$	$\sigma = 0.83$	$\sigma = 0.87$	$\sigma = 0.85$	$\sigma = 0.84$	$\sigma = 0.86$
ggH-cat2	$\sigma = 0.73$	$\sigma = 0.71$	$\sigma = 0.70$	$\sigma = 0.72$	$\sigma = 0.71$	$\sigma = 0.70$	$\sigma = 0.71$
ggH-cat3	$\sigma = 0.85$	$\sigma = 0.85$	$\sigma = 0.85$	$\sigma = 0.86$	$\sigma = 0.85$	$\sigma = 0.84$	$\sigma = 0.85$
ggH-cat4	$\sigma = 0.84$	$\sigma = 0.81$	$\sigma = 0.81$	$\sigma = 0.82$	$\sigma = 0.81$	$\sigma = 0.80$	$\sigma = 0.80$
Combined ggH	$\sigma = 1.71$	$\sigma = 1.68$	$\sigma = 1.66$	$\sigma = 1.71$	$\sigma = 1.69$	$\sigma = 1.67$	$\sigma = 1.68$

Table 6.15: A posteriori expected significance in each category and for each family of functions obtained for a SM Higgs boson with $m_H = 125$ GeV. Systematic uncertainties on the signal prediction have not be added into the datacards used to extract these significances, however their impact is expected to be quite small at the level of few %.

Category	BWZ × Bernstein	BWZGamma	BWZRedux	FEWZ × Bernstein	Landau × Bernstein	S-Exponential	S-PowerLaw
ggH-cat0	$\sigma = 0.46$	$\sigma = 0.46$	$\sigma = 0.44$	$\sigma = 0.48$	$\sigma = 0.46$	$\sigma = 0.47$	$\sigma = 0.47$
ggH-cat1	$\sigma = 0.83$	$\sigma = 0.83$	$\sigma = 0.82$	$\sigma = 0.85$	$\sigma = 0.84$	$\sigma = 0.83$	$\sigma = 0.84$
ggH-cat2	$\sigma = 0.70$	$\sigma = 0.70$	$\sigma = 0.69$	$\sigma = 0.71$	$\sigma = 0.70$	$\sigma = 0.70$	$\sigma = 0.70$
ggH-cat3	$\sigma = 0.85$	$\sigma = 0.84$	$\sigma = 0.84$	$\sigma = 0.86$	$\sigma = 0.85$	$\sigma = 0.84$	$\sigma = 0.85$
ggH-cat4	$\sigma = 0.86$	$\sigma = 0.83$	$\sigma = 0.83$	$\sigma = 0.83$	$\sigma = 0.83$	$\sigma = 0.82$	$\sigma = 0.83$
Combined ggH	$\sigma = 1.69$	$\sigma = 1.68$	$\sigma = 1.66$	$\sigma = 1.70$	$\sigma = 1.68$	$\sigma = 1.67$	$\sigma = 1.68$

a given category may potentially inject a bias in the estimated signal strength, when a signal plus background fit is performed. Therefore, as done in earlier iteration of this analysis [13] as well as in other searches following a similar approach for the signal extraction [77], a detailed set of bias studies are performed.

Toy generation for bias studies

The first step in performing bias studies for the signal extraction consists in generating toy experiments for each category for each functional form selected as a possible *truth* function that describes the $m_{\mu\mu}$ spectrum in background events. Starting from the background shapes reported in Fig. 6.19, which have been obtained from a b-only fit to the data, and the expected signal shapes shown in Fig. 6.16, 2000 toys are generated for each background functional form in each category. The number of toys is chosen in order to have an uncertainty of about 2% on the estimated bias. Toys are generated by injecting an amount of signal event proportional to the expected rate in each category for $m_H = 125$ GeV. The same amount of toy data have been generated also for other Higgs boson masses of 120 and 130 GeV. Since we are only interested in estimating the bias on the signal strength induced by the assumption of a specific background shape, systematic uncertainties affecting the signal prediction, described in Section 6.3.2 and 6.3.2 are not considered in the bias studies.

Bias of the various background functions

The dimuon mass distributions obtained from the toy-experiments previously described are fitted with a signal plus background model, where the background function is one of those reported in Table 6.13. In each fit, the bias on the signal strength is defined as follows:

$$\text{Bias}(\mu_{fit}, \hat{\mu}_{inj}, \sigma_\mu) = \frac{\mu_{fit} - \hat{\mu}_{inj}}{\sigma_\mu} \quad (6.2)$$

where μ_{fit} is the value of the signal strength extracted from the fit, $\hat{\mu}_{inj}$ is the signal strength injected is in each toy ($\hat{\mu}_{inj} = 1$)⁵, and σ_μ is its post-fit uncertainty estimated as the crossing at

⁵In each toy, the amount of signal events injected correspond to a Poissonian “smearing” around a mean value

— $\Delta\text{Log}\mathcal{L} = 1$ of the likelihood scan as a function of μ . In absence of systematic effects, the bias is expected to be Gaussian distributed, centred at zero, and with a width of one. The mean value of the bias distribution is used to quote the bias associated to a given fit function. Good functional forms, which can be used in the final signal extraction, are those characterized by a small bias on the signal strength in each category ($< 20\%$ as defined in Equation 6.2). The threshold of 20% is chosen because such bias guarantees that the systematic uncertainty on the quoted upper limit or on the expected significance is smaller than 2%.

Figure 6.20 shows the measured bias on the signal strength for $m_H = 125\text{ GeV}$ in ggH categories, obtained by fitting the toys generated via several truth function with BWZRedux, BWZGamma, FEWZ \times Bernstein, S-exponential, S-power-law, and Landau \times Bernstein. The number of degrees of freedom considered for each functional form and in each category are listed in Table 6.13. A large tension is observed between the series of exponentials and all the other truth models. BWZRedux, BWZGamma, BWZ \times Bernstein, and FEWZ \times Bernstein exhibit a bias in single categories within 20% against non exponential truth models. The bias w.r.t. the sum of exponentials is larger than 50% in ggH-cat0, ggH-cat1, and ggH-cat2, while it is larger than 20% in ggH-cat3 and ggH-cat4.

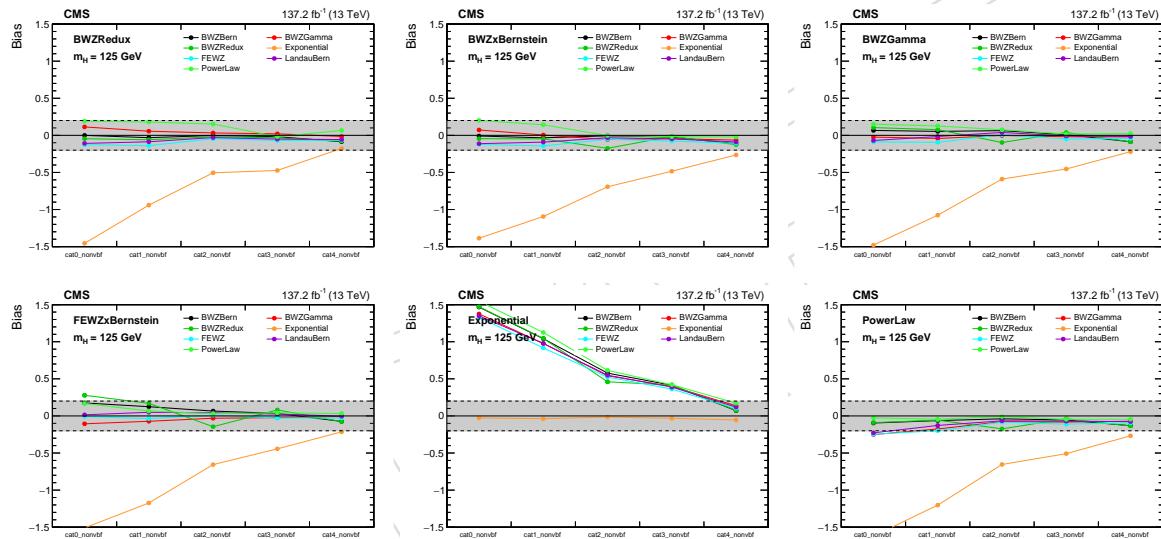


Figure 6.20: Mean value of the bias distribution in ggH BDT categories obtained by fitting toys, generated according to different background truth models and injecting a nominal signal at $m_H = 125\text{ GeV}$, with the following functional forms: BWZRedux (top left), BWZxBernstein (top middle), BWZGamma (top right), FEWZ \times Bernstein (bottom right), S-exponential (bottom middle), and S-power-law (bottom right).

An additional concern is the potential bias on the signal strength extracted from the simultaneous fit performed across multiple categories (*correlated bias*). In principle, since each category represents an independent entity from the background estimate point of view, we would need to generate toys according to all possible combinations of truth background functions. However, having nine event categories and at least six alternative truth models makes this test computationally not feasible. Therefore, we consider only the cases in which the type of function used to generate and fit the toys is the same across all categories. For example, when the BWZRedux is chosen as the truth model, pseudo-experiments are thrown using this background shape in all categories. On the other hand, when the BWZRedux is tested as fit function, pseudo-experiments generated according to every possible truth function are fitted with

given by the expected signal rate for $\mu = 1$

1658 a signal-plus-background model in which the BWZRedux parametrization is used to describe
 1659 the background in all categories. Figure 6.21 shows the estimated bias on the signal strength
 1660 for $m_H = 125\text{ GeV}$ obtained when the ggH categories are simultaneously fitted using either a
 1661 BWZRedux (left), a BWZGamma (middle), or a FEWZ \times Bernstein (right) function to model
 1662 the background. When ggH categories are fitted, a large bias is observed against the series of
 1663 exponentials. This was already expected from the measurements reported in Fig. 6.20, whereas
 1664 the bias w.r.t all the other truth models can be reasonably controlled within 20%.

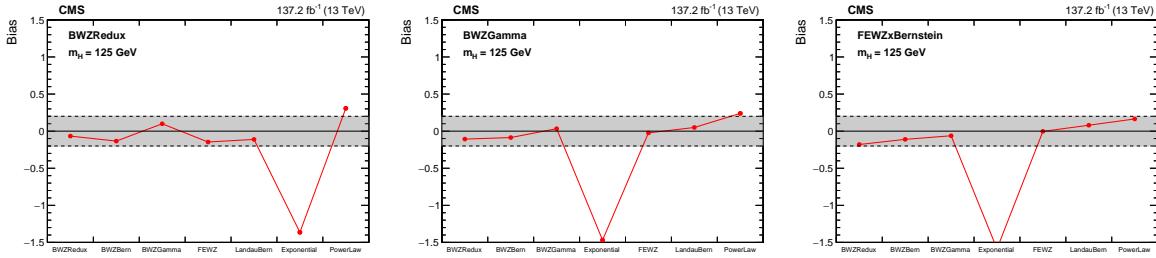


Figure 6.21: Mean value of the bias distribution on the signal strength obtained by fitting toys, generated across all the ggH categories according to alternative background truth models and injecting a nominal signal at $m_H = 125\text{ GeV}$, with the following functional forms: BWZRedux (left), BWZGamma (middle), and FEWZ \times Bernstein (right).

1665 In conclusion, in ggH categories the bias on the signal strength cannot be controlled within an
 1666 acceptable level by using single functions to fit for the background in each category. Therefore,
 1667 if possible, an alternative bias-free strategy to estimate the total background is required that
 1668 can control better the bias against the series of exponentials.

1669 6.4.2 The core-PDF method

1670 The simplest way to estimate the background via analytical fits to the observed $m_{\mu\mu}$ distri-
 1671 butions relies on using independent functions for each category. However, as described in
 1672 Section 6.4.1, this approach does not allow to obtain a robust control of the bias on the sig-
 1673 nal strength in the ggH categories. Therefore, we have developed an alternative strategy to
 1674 estimate the background in these categories that tries to maximize the expected sensitivity
 1675 while introducing a negligible bias on the measured signal yield. This method is referred to
 1676 as *core-PDF*.

1677 The main motivation for this method originates from the fact that the background composition
 1678 across categories is quite stable, i.e. the background purity of dimuon decays of Z bosons
 1679 (DY+jets or Z-EW) is fairly constant along the BDT outputs. Therefore, it is natural to expect
 1680 that background shape parameters can be correlated across BDT categories. On the other hand,
 1681 this correlation can only be partial because, as shown in Fig. 6.7, going from low to high score
 1682 along the BDT outputs one moves from events with small p_T of the dimuon system to more
 1683 boosted signatures, and this has a direct effect on the $m_{\mu\mu}$ shape of background events. In fact,
 1684 the $m_{\mu\mu}$ distribution of Z/ γ^* +jets events becomes less steep for events with larger $p_T^{\mu\mu}$. The
 1685 *core-PDF* method tries to resolve this difference in the following way:

- 1686 • The background in the ggH categories is extracted using a model that contains a
 1687 common function, called *core-function*, whose parameters are estimated from a si-
 1688 multaneous fit to the data across all the categories. Therefore, these parameters are
 1689 fully correlated across categories.
- 1690 • In order to give to the background function additional flexibility to better fit the data
 1691 in each category, the *core-function* is modulated (multiplied) in each category by a

1692 polynomial term called “shape modifier function (SMF)”. We start with a first order
 1693 polynomial but the ultimate order of the SMFs is selected by performing bias studies
 1694 for the signal strength.

1695 Figure 6.22 shows a comparison between the background shapes in ggH BDT categories as
 1696 obtained from a background-only fit to the data via a BWZRedux function. This confirms
 1697 that the $m_{\mu\mu}$ distribution in background events becomes progressively “softer” moving from
 1698 low to high score, and that this effect is mainly correlated with an increase in the dimuon p_T .
 1699 Moreover, Fig. 6.23 shows proxies for the SMFs in data for the ggH categories. These SMFs are
 1700 defined as the bin-by-bin ratio between the $m_{\mu\mu}$ spectrum in the chosen category and inclusive
 1701 obtained from by summing together all the events belonging to the ggH phase space. Results
 1702 from a fit to the SMF using either a third or a second order polynomial are also reported.

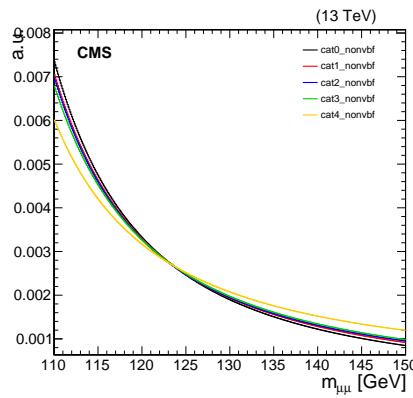


Figure 6.22: Comparison between the $m_{\mu\mu}$ shapes obtained from a background-only fit to the data using a BWZRedux function in ggH categories. In both cases, the $m_{\mu\mu}$ distribution becomes softer moving from low (cat0) to high (cat3 or cat4) ggH BDT score.

1703 In order to measure the bias of the *core-PDF* model, which always performs a fit across mul-
 1704 tiple categories, one needs to test both the robustness of the method against alternative truth
 1705 functions as well as the fact that polynomial SMFs are enough to model differences in the
 1706 background shape across categories. Pseudo-experiments are generated following the same ap-
 1707 proach described in Section 6.4.1 in the context of testing the bias for the combined fit (*correlated bias*).
 1708 Toys are generated independently in each category according to a chosen truth function. When
 1709 we want to measure the bias in a single category, the following procedure is adopted:

- 1710 • In the toy generation, the signal is only injected in that specific category while, in the
 1711 other ones, background-only $m_{\mu\mu}$ distributions are produced.
- 1712 • Toys are then fitted with a signal-plus-background model which uses a *core-PDF*
 1713 to describe the background and the presence of the signal is only allowed in the
 1714 category where the signal injection was performed. In this way, the parameters of
 1715 the *core-function* are constrained by all the data included in the fit, while the signal
 1716 strength is extracted only by looking at the category of interest.

1717 On the other hand, when the bias on the overall signal strength is tested, the injection is per-
 1718 formed in every category and therefore the presence of the signal in the fit is allowed every-
 1719 where. In order to control the bias in the ggH categories within an acceptable threshold of 20%,
 1720 the following setup is used:

- 1721 • As described in Fig. 6.23, in both ggH-cat0 and ggH-cat1, SMFs are modelled with a
 1722 third-order polynomial while in the other three categories second-order polynomials
 1723 are considered.

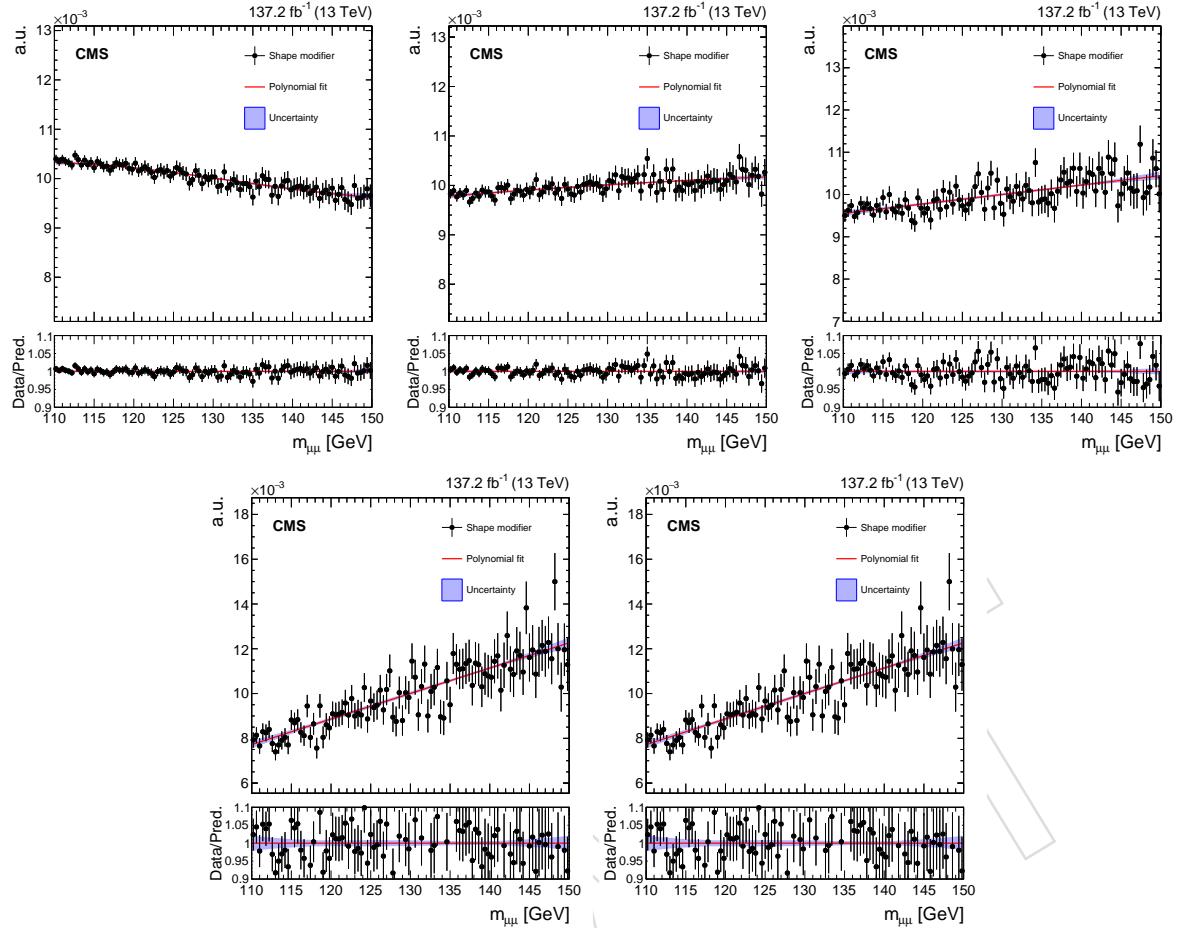


Figure 6.23: Shape modifiers in data used in the *core-PDF* for events belonging to the ggH categories. These shape modifiers are defined as the bin-by-bin ratio between the $m_{\mu\mu}$ spectrum in a given category and the inclusive one, obtained by merging the ggH categories into a single one. Top row: ggH-cat0 (left), ggH-cat1 (middle), and ggH-cat2 (right). Bottom row: ggH-cat3 (left), and ggH-cat4 (right). The solid red lines indicate the results of a fit to the shape modifiers via third-order (ggH-cat0 and ggH-cat1) or second-order polynomials, while the blue area corresponds to the 68% uncertainty band from the fit parameters.

- With the goal of substantially reducing the bias between Breit-Wigner like functions and sum of exponentials, an envelope of three PDFs with three degrees of freedom is used as *core-function*: BWZRedux, FEWZ \times Bernstein, and S-exponential. The discrete profiling method is described in more details in Ref. [78]. This method was designed in the context of the $H \rightarrow \gamma\gamma$ search in order to estimate the systematic uncertainty associated to the choice of a particular analytical function used to fit for the background. The method treats the choice of the background function as a discrete nuisance parameter in the likelihood fit to the data. This allows to dynamically select the best core-function that can be used to model the background at each point of the likelihood scan as a function of the signal strength. Given that the three functions considered in the envelope have the same number of freely floating parameters (3-dof), no penalization term as a function of the degrees of freedom is used.

Figure 6.24 (left) shows the measured bias on the signal strength for $m_H = 125$ GeV in each of the ggH categories when the background is modelled using the *core-PDF* method. The bias is observed to be smaller than 20% in each category and for each truth function used to generate

1739 the toys. In addition, Fig. 6.24 (right) reports the bias measured in toys where the signal has
 1740 been injected in all the ggH categories. Also in this case, the bias measured as a function of
 1741 different truth functions can be controlled within 20%.

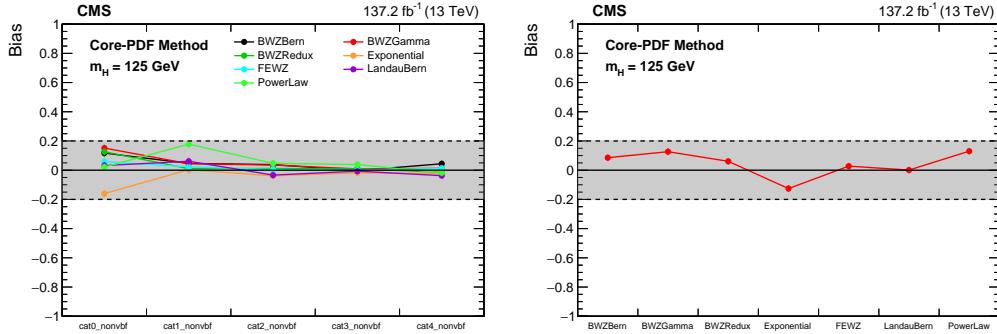


Figure 6.24: Mean bias on the signal strength obtained from the *core-PDF* method used to model the background in ggH categories. On the left, the measured bias in each of the ggH categories when a nominal signal at $m_H = 125\text{ GeV}$ is injected. On the right, the bias measured when the signal is injected in all the ggH categories simultaneously.

1742 Given that a discrete likelihood profile is used to define the *core-PDF* component, the coverage
 1743 of this method is checked on the same toys used for the bias studies. The coverage within
 1744 the 68% CL ($\approx 1\sigma$) interval is defined as the ratio between the observed number of toys for
 1745 which $-2 \times \Delta \text{Log} \mathcal{L}(\hat{\mu}_{inj}) < 1$ and the total number of toys generated, where a signal with
 1746 $\hat{\mu}_{inj} = 1$ and $m_{\{H\}} = 125\text{ GeV}$ is injected. Similarly, the coverage for 16% CL ($\approx 0.2\sigma$), 38% CL
 1747 ($\approx 0.5\sigma$), 86% CL ($\approx 1.5\sigma$), and 95% CL ($\approx 2\sigma$) intervals have also been tested. Figure 6.25
 1748 shows that the method designed for the background estimation has good coverage properties
 1749 for the tested intervals. The *core-PDF* fit shows, as expected from a discrete likelihood profile,
 1750 a small over-coverage (by few %) in categories and against some truth models.

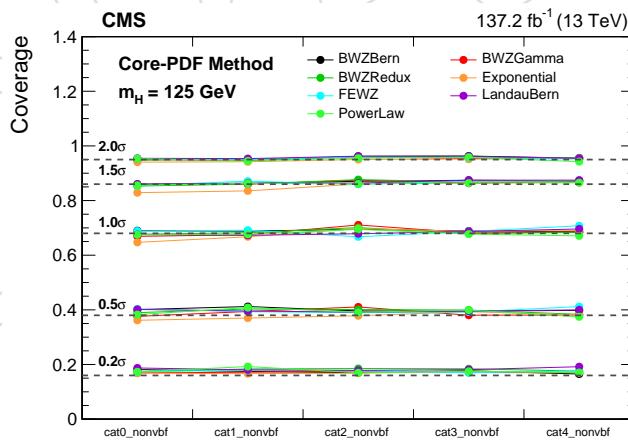


Figure 6.25: Coverage observed with the *core-PDF* method for 16%, 38%, 68%, 86%, and 95% CL
 intervals in toys where a signal at $m_{\{H\}} = 125\text{ GeV}$ is injected with $\hat{\mu}_{inj} = 1$.

1751 Finally, Appendix C shows results from bias studies performed in ggH categories for different
 1752 mass points of $m_H = 120\text{ GeV}$ and $m_H = 130\text{ GeV}$, which correspond to the smallest and largest
 1753 Higgs boson mass hypothesis tested by this search, respectively. Overall, similar results and
 1754 conclusions to those presented for $m_H = 125\text{ GeV}$ are found.

6.5 Results

In order to extract the final results, binned maximum-likelihood fits are performed to the observed $m_{\mu\mu}$ distributions of each category, in the range $110 < m_{\mu\mu} < 150$ GeV, where the expected signal from the different production modes is scaled by common strength modifier (μ). Fits are performed fixing the expected Higgs boson mass to 125 GeV. Systematic uncertainties from theoretical and experimental sources, as described in Section 8.3, are included in the likelihood via Log-Normal nuisance parameters. Binned fits are used to speedup the computation, and the chosen bin size of 50 MeV is sufficiently small compared to the mass resolution that no information is lost. The data and the result from a signal-plus-background model performed simultaneously across categories are shown in Fig. 6.26. The one (green) and two (yellow) standard deviation bands shown for the background component of the fit include the uncertainty in the fitted parameters. The background in the ggH categories is modelled via the *core-PDF* method detailed in Section 6.4.2.

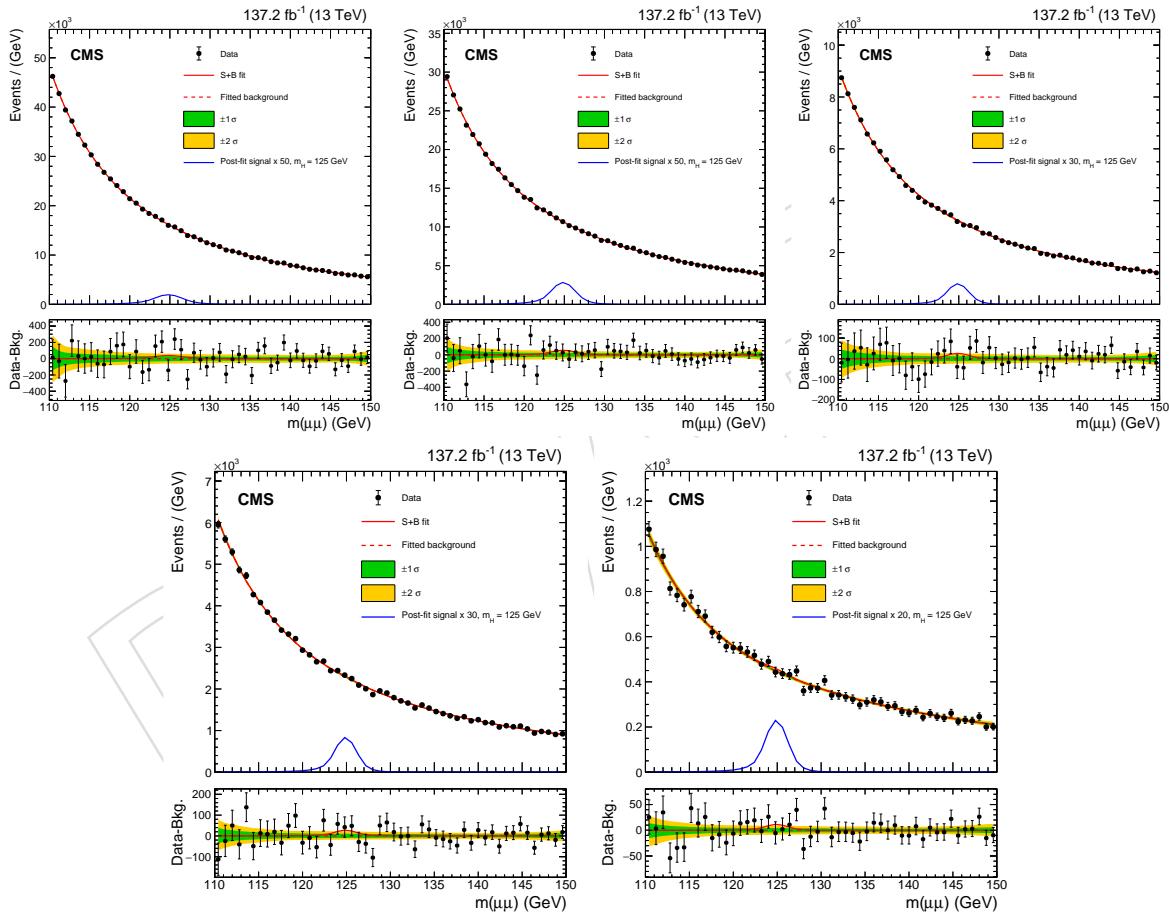


Figure 6.26: Comparison between observed data and the total background extracted from a signal-plus-background fit performed across all event categories. Top row: ggH-cat0 (left), ggH-cat1 (middle), and ggH-cat2 (right). Bottom row: ggH-cat3 (left), and ggH-cat4 (right). The one (green) and two (yellow) standard deviation bands include the uncertainties in the background component of the fit. The lower panel in each plot shows the residuals after the background subtraction.

Likelihood scans of the signal strength modifier are performed, in which the other parameters of the signal and background models are allowed to vary. Systematic uncertainties are included in the form of nuisance parameters. Figure 6.27 (left) shows the observed likelihood scan obtained by a combined fit performed across ggH categories. In contrast, Fig. 6.27 (right) reports the expected likelihood scan obtained by fitting an Asimov dataset in which nuisance

parameters are estimated from a background-only fit to the data and a signal is also injected, with a rate corresponding to the one predicted by the SM. The crossing point of this likelihood scan at $\mu = 0$ gives the expected significance of the signal-plus-background hypothesis w.r.t. the background-only one. In the Appendix of Ref. [17], the same expected likelihood scans, performed either in presence or in absence of a SM Higgs boson signal in the toy, are presented for each of the five event categories defined along the ggH BDT output.

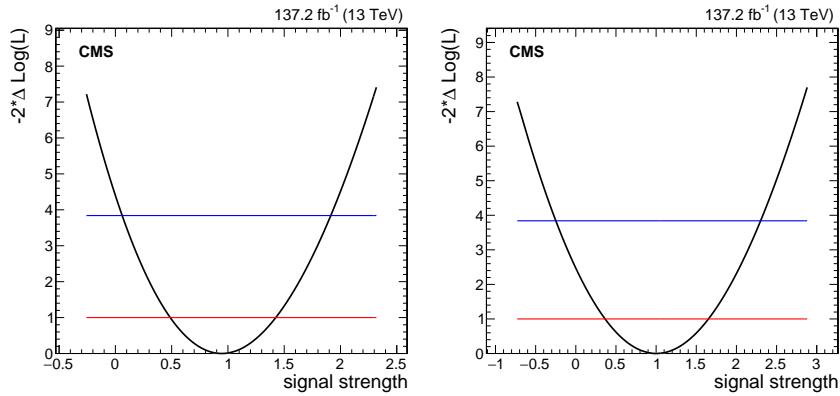


Figure 6.27: Observed and expected profile likelihood ratios as a function of the signal strength modifier (μ) for the combined signal-plus-background fit performed across ggH event categories. On the left, likelihood scan obtained from a signal-plus-background fit to the observed data. On the right, the likelihood profile vs μ is obtained by fitting an Asimov dataset in which the expected Higgs boson signal, assuming SM production cross sections, has been injected in the toy and nuisances parameters are fixed to the values extracted from a background-only fit to the data.

Figure 6.28 shows the results of a goodness-of-fit test for the combined fit performed across ggH categories, in which the saturated likelihood ratio is used as test-statistics. The expected distribution for the test-statistics is obtained by performing signal-plus-background fits, with floating μ , on toys thrown from the best-fit estimate of the nuisance parameters (a-posteriori toys). The observed p-value, i.e. probability of having a goodness-of-fit value larger than the observed one, is 0.21. This indicates a good compatibility between the observed data and the signal-plus-background model.

Table 6.16 lists both the expected (observed) upper limits on the signal strength modifier, computed at the 95% CL with the Asymptotic CL_s method [79–81], and the significance of the incompatibility between the expected signal-plus-background hypothesis and the background-only hypothesis for each BDT category, as well as for their combination. The *core-PDF* allows to control the bias on the signal strength within 20% in each category at the price of loosing about 6% sensitivity in terms of expected significance and upper limit. The expected (observed) upper limit in the absence of a SM decay in this channel is $\mu < 1.27(1.88) \times \text{SM}$, which corresponds to an expected (observed) significance of the Higgs boson decaying into two muons of 1.57 (1.23) standard deviations. Moreover, in Table 28 in the Appendix of Ref. [17] expected asymptotic upper limits on the signal strength are reported independently for each era and in each category. They have been computed with simplified datacards neglecting systematic uncertainties in the signal prediction and modelling the background via a BWZRedux function.

The systematic uncertainties affecting the signal prediction in each category have a small impact on the ultimate analysis sensitivity, causing a loss of about 3% in both the expected significance and expected upper limit. To assess the importance of each source of systematic uncertainty (constrained nuisance parameter) on the signal strength, extracted from the combined across ggH BDT categories, a signal-plus-background fit is performed to estimate the best fit

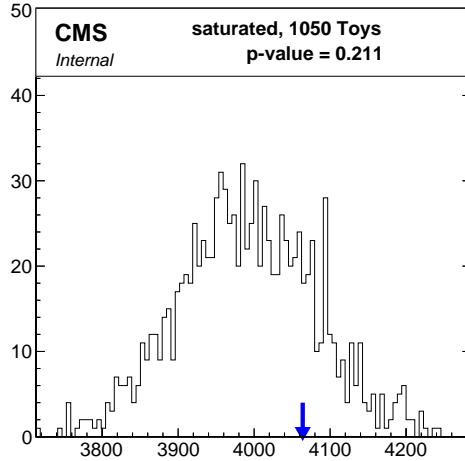


Figure 6.28: Results from a goodness-of-fit test for the combined fit performed across ggH categories. The observed value for the test-statistics is indicated by the blue arrow, while the histogram (solid black) indicates the expected distribution obtained in signal-plus-background toys. The observed p-value, i.e. probability of having a goodness-of-fit value larger than the observed one, is 0.21.

Table 6.16: Summary of the expected and observed upper limits on μ , computed at the 95% CL for $m_H = 125\text{ GeV}$, as well as the expected and observed signal significance for each BDT category and for their combination.

Category	Expected UL	Observed UL	Expected Signif.	Observed Signif.
ggH-cat0	< 4.641	5.750	0.440σ	0.707
ggH-cat1	< 2.477	2.340	0.806σ	0.000
ggH-cat2	< 2.773	3.383	0.724σ	0.532
ggH-cat3	< 2.305	2.996	0.877σ	0.787
ggH-cat4	< 2.477	3.345	0.829σ	0.827
Combination ggH	< 1.266	1.885	1.571σ	1.227

1803 value of each nuisance parameter, along with their related post-fit uncertainty. Data are then re-
 1804 fitted shifting independently each nuisance parameter by $\pm 1\sigma$ (post-fit) with respect to the best
 1805 fit value. The impact of each source of uncertainty on the signal strength is defined as the differ-
 1806 ence between the values of μ obtained after moving the nuisance by $\pm 1\sigma$ and the best fit result.
 1807 Figure 6.29 (left) shows the result of the impact study in the context of the combined fit on
 1808 the observed data, where nuisances are ranked according to their impact on the measured sig-
 1809 nificant strength. None of these nuisance parameters are significantly pulled or constrained by the
 1810 signal-plus-background fit. The most important systematics are represented by the uncertainty
 1811 in the dimuon mass resolution (up to 10% uncertainty correlated across categories as described
 1812 in Section 6.3.2), the uncertainties in the ggH inclusive cross section, the uncertainty related to
 1813 the parton shower parameters, and those affecting the ggH acceptance from the STXS stage 1.1
 1814 prescription. Finally, the uncertainties in measurement of the integrated luminosity and in the
 1815 prediction of $\mathcal{B}(H \rightarrow \mu\mu)$ are also playing a relevant role. On the other hand, Fig. 6.29 (right)
 1816 reports the result of the impact study in the combined fit to the observed data when also freely
 1817 floating background parameters (normalization and shape) are considered. The parameters
 1818 with the largest impact on the performance of the final fit are the background normalization in
 1819 each category, as well as some of the freely floating shape parameters of the shape modifiers.

1820 The 95% CL upper limits on the Higgs boson production cross section times $\mathcal{B}(H \rightarrow \mu\mu)$ are

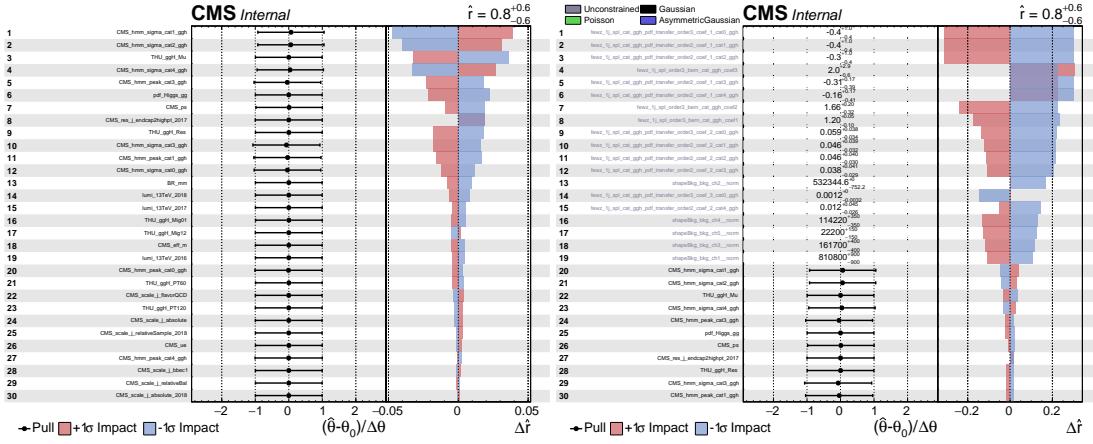


Figure 6.29: Result of the impact study on μ for the combined fit performed across ggH categories on the observed data.

computed in the range $120 < m_H < 130$ GeV using an asymptotic approximation [81] of the CL_s method [80], and are shown in Figure 6.30 (left). The observed and expected upper limits for a 125.09 GeV $H \rightarrow \mu\mu$ signal, assuming SM production rates, are found to be 1.88 and 1.27 times the SM expectation, respectively. Figure 6.30 (right) shows the observed and expected local p-values computed in the mass range $120 < m_H < 130$ GeV for a 125.09 GeV Higgs boson signal. The expected significance is found to be 1.57 standard deviations for $m_H = 125$ GeV, while the observed significance in data is 1.23 standard deviations.

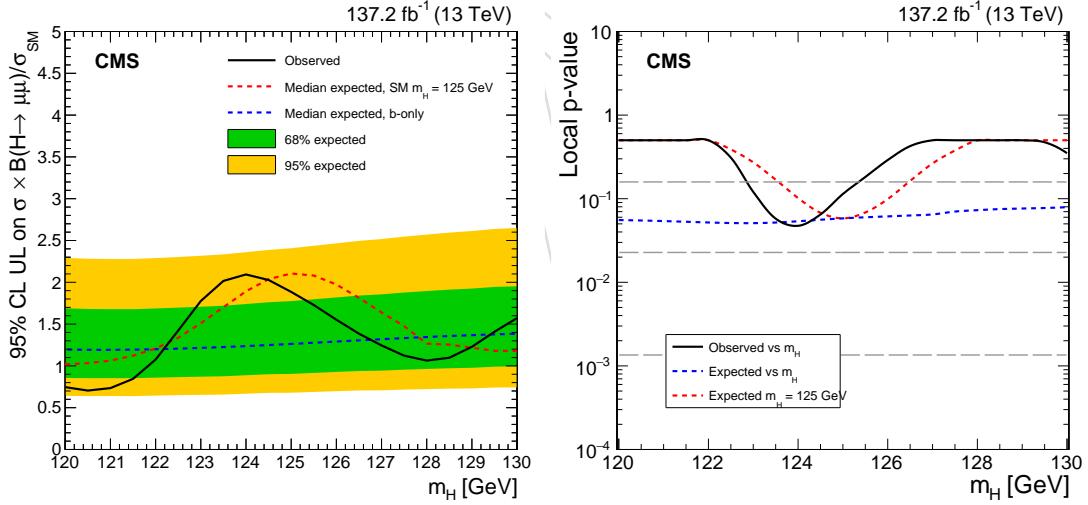


Figure 6.30: On the left, the observed 95% CL upper limit on the on the Higgs boson production cross section times $B(H \rightarrow \mu\mu)$, relative to the SM expectation, in the region around the Higgs boson mass along with the expected exclusion limit obtained in the background-only hypothesis (dashed blue line) and in the signal-plus-background one (dashed red line) for the SM Higgs boson with $m_H = 125.09$ GeV. On the right, the observed and expected local p-values as a function of m_H in the range $120 < m_H < 130$ GeV.

6.6 Effect of systematic uncertainties on the background model

As presented in Section 6.4.2, no significant bias (< 20%) on the signal strength from the chosen parametric background model in each ggH category is observed, therefore no extra nuisance parameters are introduced in the final fit to account the residual per-category bias. However, in order to verify that this missing systematic uncertainty has a small impact (< 2%) on the

overall significance, spurious signals are injected in each ggH category proportionally to the maximum observed bias. Table 6.17 reports, for each category, the chosen function used in the signal extraction, the observed post-fit uncertainty on the signal strength ($\Delta\mu$), and the maximum bias w.r.t. the chosen functional form.

Table 6.17: In each ggH category, the chosen function used in the signal extraction, the observed post-fit uncertainty on the signal strength ($\Delta\mu$), and the maximum bias for the selected function form as measured in the studies presented in Section 6.4.2.

Category	Functional form	Post-fit $\Delta\mu$	Maximum bias (%)
ggH-cat0	Core-PDF	2.09	16.1%
ggH-cat1	Core-PDF	1.18	17.4%
ggH-cat2	Core-PDF	1.47	4.7%
ggH-cat3	Core-PDF	1.18	3.8%
ggH-cat4	Core-PDF	1.25	3.7%

The spurious signal considered in each ggH category to account for the measured bias on the signal strength is added as a new background process and its $m_{\mu\mu}$ shape is chosen to be identical to the one expected for the ggH process. The total pre-fit normalization for this spurious process is set to be zero in each category. The normalization is allowed to vary in the fit to the data within a Gaussian prior with width (ΔB) defined, for each ggH category, as:

$$\Delta B = \hat{b} \times \Delta\mu \times N_{\text{sig}} \quad (6.3)$$

where \hat{b} is the measured bias on μ , $\Delta\mu$ is the post-fit uncertainty on the signal strength, and N_{sig} is the expected number of signal events in the chosen category. Table 6.18 reports the difference in the expected significance for Higgs boson decay to muon pairs with $m_H = 125$ GeV between the “baseline” result already presented in Table 6.16 and the one obtained considering the uncertainty in the background model (implemented via spurious signals). The difference in the combined expected significance is found to be 0.4% and therefore neglected in the final result.

Table 6.18: Difference in the expected significance to $H \rightarrow \mu\mu$ decays for $m_H = 125$ GeV in each category between the “baseline” result, which ignores systematic uncertainties in the background model, and the one obtained adding spurious signals with rate proportional to the measured bias on μ from the studies presented in Section 6.4.2.

Category	Signif. w/o bias	Signif. with bias	Difference (%)
ggH-cat0	0.440σ	0.435σ	1.15%
ggH-cat1	0.806σ	0.794σ	1.51%
ggH-cat2	0.724σ	0.723σ	0.14%
ggH-cat3	0.877σ	0.876σ	0.11%
ggH-cat4	0.829σ	0.828σ	0.13%
Combination ggH	1.571σ	1.565σ	0.38%

1849 7 VBF category based on MC template fit

1850 As described in the analysis strategy (Sec. 5) and event selections (Sec. 3), the events without
 1851 additional b-tagged jets nor additional leptons (i.e. not included in the ttH or VH categories),
 1852 but with two additional jets with p_T larger than 35 and 25 GeV, respectively for the leading and
 1853 subleading jet, with invariant mass $m_{jj} > 400$ GeV, and pseudorapidity separation $\Delta\eta_{jj} > 2.5$
 1854 are classified into the VBF category.

1855 Given the low statistics of events falling in this category and the good precision of Monte Carlo
 1856 simulations in modelling the relevant backgrounds, a signal extraction method fully based on
 1857 Monte Carlo is exploited. This part of the analysis follows the strategy employed by CMS
 1858 to study and observe the Zjj-EW (VBF Z) production in dilepton plus dijet final states with
 1859 7 TeV [82] and 8 TeV [83] data, and then with better precision with 13 TeV [35] data. Similar
 1860 measurements have been performed with single-lepton plus two jets to establish also the
 1861 existence of the Wjj-EW (VBF W) production with 8 TeV [84] and 13 TeV [85] data.

1862 7.1 Event selection: signal and control regions

1863 A Signal region and two control regions (Z region and Sideband) are fixed by $m_{\mu\mu}$ cuts. In
 1864 the Signal region muons are required to have an invariant mass within 10 GeV of the Higgs
 1865 mass: $|m_{\mu\mu} - 125| < 10$ GeV. The SideBand contains events with $110 < m_{\mu\mu} < 115$ GeV or
 1866 $135 < m_{\mu\mu} < 150$ GeV. Events with dimuon invariant mass within 15 GeV of the Z mass
 1867 $|m_{\mu\mu} - 91| < 15$ GeV belong to Z region.

1868 A pre-selection common to all three regions is defined requiring the events to pass the following
 1869 cuts:

- 1870 • Two muons as described in Sec.3.2
- 1871 • At most one b-tagged loose jet and no b-tagged-medium jets
- 1872 • Leading jet $p_T > 35$ GeV and sub-leading jet $p_T > 25$ GeV
- 1873 • Two p_T -leading jets $m_{jj} > 400$ GeV $|\Delta\eta_{jj}| > 2.5$

1874 Then the three orthogonal regions are defined as follows:

- 1875 • *SignalRegion*: $115 < m_{\mu\mu} < 135$ GeV
- 1876 • *SideBand*: $110 < m_{\mu\mu} < 115$ GeV or $135 < m_{\mu\mu} < 150$ GeV
- 1877 • *ZRegion*: $76 < m_{\mu\mu} < 106$ GeV

1878 Figure 7.1 shows the m_{jj} and $\Delta\eta_{jj}$ distributions for the VBF H, ggH, and Z+jets (EW + QCD)
 1879 processes in a preselected inclusive region before any VBF selections. The purple vertical lines
 1880 indicate the selections applied to enrich in VBF H, as described in this section. After applying
 1881 the VBF selections, the remaining signal is about half VBF H and half ggH. Note that the pre-
 1882 selected events failing the VBF selections are considered in the data-driven inclusive analysis
 1883 (Sec. 6).

1884 Figure 7.2 shows the dimuon mass distribution for each year of the selected muons in the VBF-
 1885 enriched preselection region before splitting regions by dimuon mass.

1886 Tables 7.1, 7.2 and 7.3 show event yields after the selection for the three years.

1887 7.2 Data vs MC comparisons

1888 Distributions of main kinematic properties of pre-selected events are shown in the following
 1889 figures. The DY simulation model is provided by the AMC@NLO samples (ME with dileptons

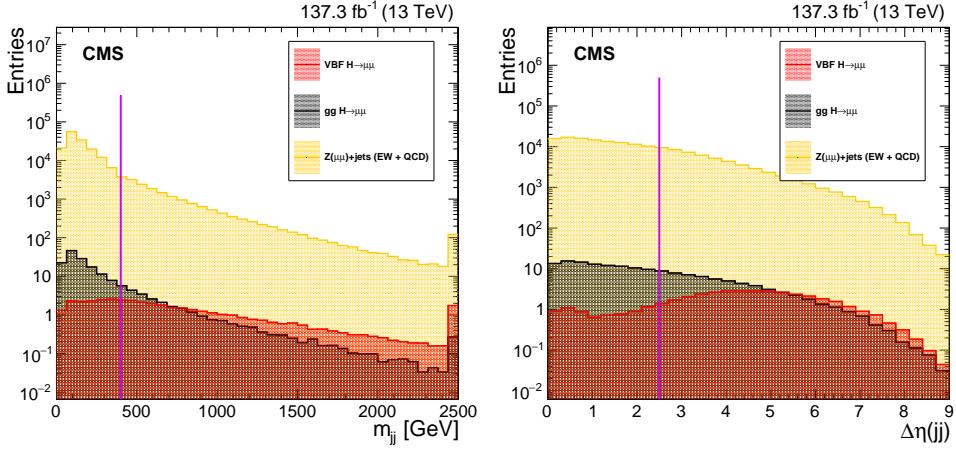


Figure 7.1: Dijet invariant mass (top) and pseudorapidity separation (bottom) in an inclusive pre-selected region before applying VBF selections. The distribution for the VBF H (red), ggH (black), and Z+jets (EW + QCD, yellow) processes are shown weighted to the expected number of events. The purple line indicates the selection applied to define the VBF-enriched region described in this section.

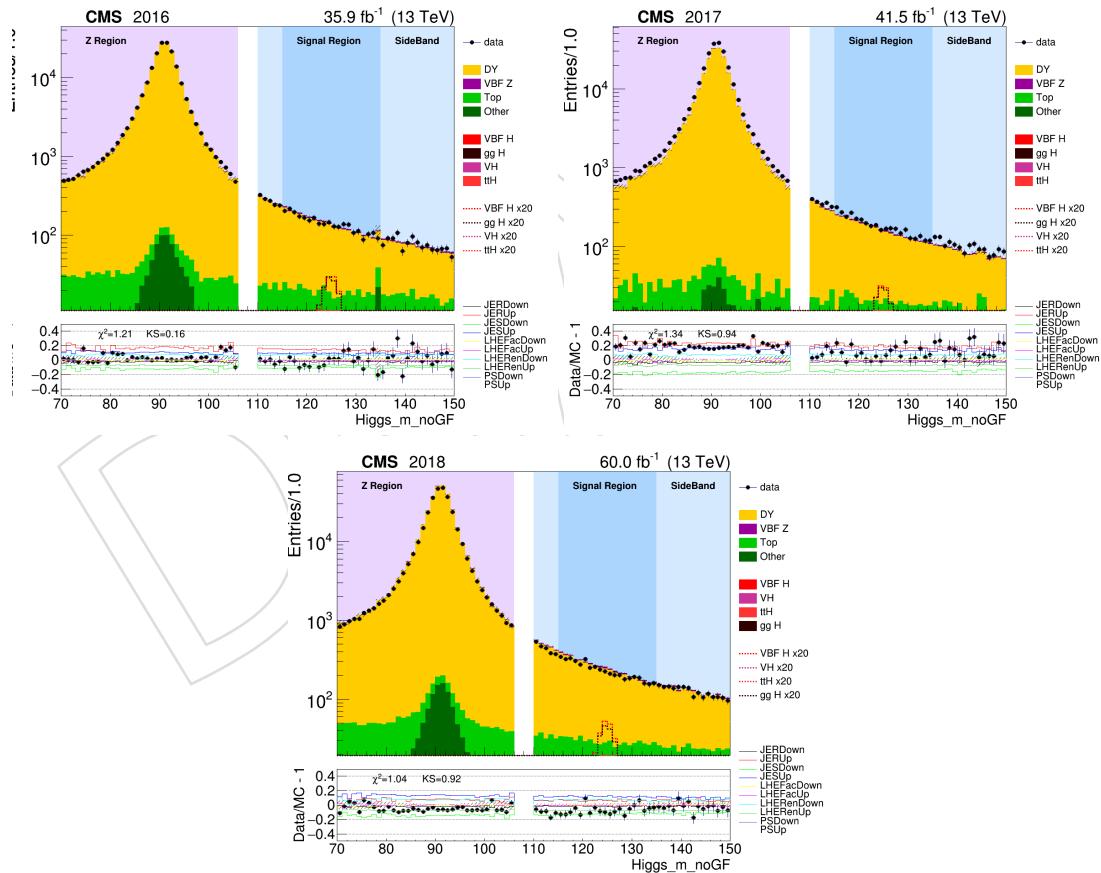


Figure 7.2: Dimuon invariant mass distributions after the event selection for 2016 (top), 2017 (center) and 2018 (bottom). Events in different regions are shown in the same plot for every year.

plus up to two final state partons, and NLO QCD corrections). All plots indicate a simple signal versus background separation power d is defined as

Table 7.1: Event yields in Data and for different MC sources in Signal region for 2016, 2017 and 2018.

Sample	Yields 2016	Yields 2017	Yields 2018
Data	2360	3133	3895
qqH	7.1	7.5	11.7
ggH	6.8	6.8	10.1
DY	1950	2720	3270
VBF Z	75	79	125
Top	275	288	410
Total MC	2300	3090	3800

Table 7.2: Event yields in Data and for different MC sources in Sideband for 2016, 2017 and 2018.

Sample	Yields 2016	Yields 2017	Yields 2018
Data	2051	2708	3310
DY	1690	2360	2780
VBF Z	66	68	107
Top	254	250	366
Total MC	2010	2680	3250

Table 7.3: Event yields in Data and for different MC sources in Z region for 2016, 2017 and 2018.

Sample	Yields 2016	Yields 2017	Yields 2018
Data	157276	219427	270443
DY + jets	155284	182616	280894
VBF Z	4094	4049	6625
Top	547	543	819
Other	275	290	448

$$d = \frac{1}{2} \int |s(x) - b(x)|$$

where $s(x)$ and $b(x)$ are respectively the normalised signal and background distributions (with unit integral). A $d = 0$ value indicates perfectly overlapping signal and background shapes, while a $d = 1$ value indicates perfectly separated signal and background shapes.

All plots have a lower panel showing the corresponding data/MC ratio with uncertainty bands corresponding to (i) data and MC statistical fluctuations (hatched), (ii) jet energy scale (JES) uncertainties on MC, (iii) QCD scale uncertainties, and (iv) PS uncertainties, as described in more detail in Section 7.5.

7.2.1 lepton and dilepton kinematics

Figure 7.3 show the p_T distribution of the selected muons after the event selection in the Signal Region for all the years.

Figure 7.4 show the p_T distribution of the dimuon system for selected events in the Signal region for all the years.

The signal events consist in a narrow peak in the invariant mass spectrum and invariant mass information will be used to extract the signal. Therefore resolution of $m_{\mu\mu}$ has been used to better exploit the invariant mass information. Dimuon Mass resolution is computable by:

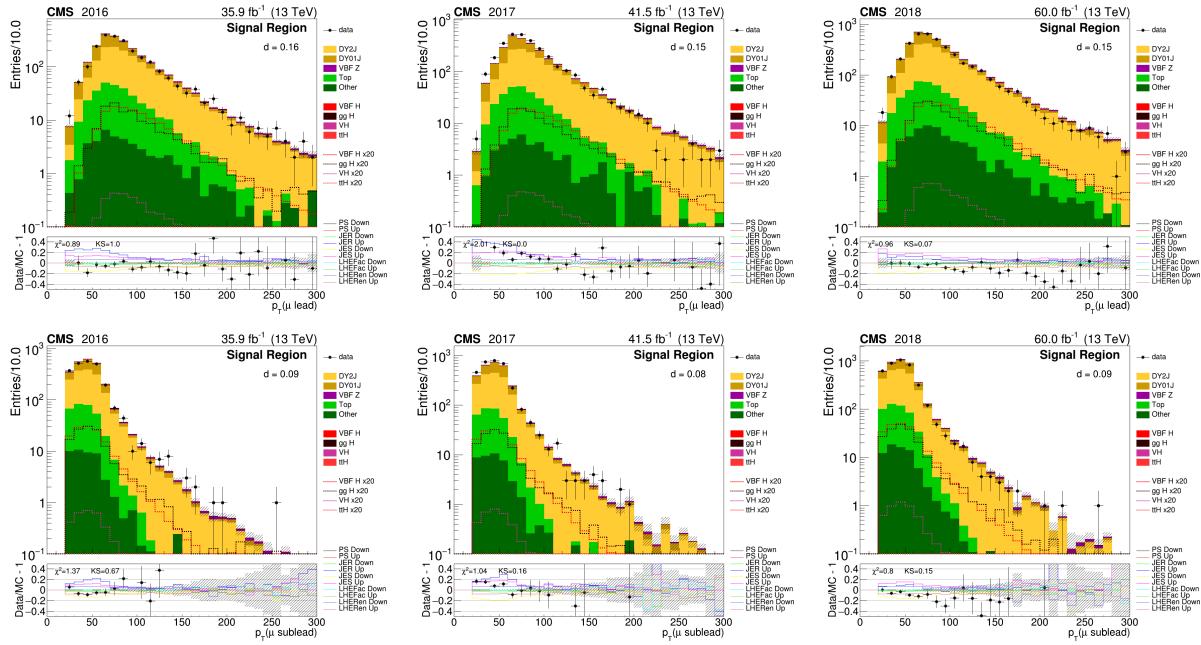


Figure 7.3: Leading (top) and subleading muon (bottom) transverse momentum distributions after the event selection in Signal Region for 2016 (left), 2017 (center), and 2018 (right).

$$\Delta m_{\mu\mu}^{rel} \equiv \frac{\Delta m_{\mu\mu}}{m_{\mu\mu}} = \frac{1}{2} \sqrt{\left(\left(\frac{\Delta p_T(\mu_1)}{p_T(\mu_1)} \right)^2 + \left(\frac{\Delta p_T(\mu_2)}{p_T(\mu_2)} \right)^2 \right)}$$

where $p_T(\mu_1)$ and $p_T(\mu_2)$ are the transverse momenta and $\Delta p_T(\mu_1)$ and $\Delta p_T(\mu_2)$ are the transverse momenta uncertainty of the p_T -leading and subleading muon. In order to estimate $\Delta m_{\mu\mu}^{rel}$ simulated events are used to parametrize the average $\langle \Delta p_T(\mu) \rangle$ as a function of $p_T(\mu)$ and $\eta(\mu)$. The parametrization has been used to compute:

$$\Delta m_{\mu\mu}^{rel} = \frac{1}{2} \sqrt{\left(\left(\frac{\langle \Delta p_T(\mu_1) \rangle(p_T(\mu_1), \eta(\mu_1))}{p_T(\mu_1)} \right)^2 + \left(\frac{\langle \Delta p_T(\mu_2) \rangle(p_T(\mu_2), \eta(\mu_2))}{p_T(\mu_2)} \right)^2 \right)}$$

Figure 7.5 shows $\Delta m_{\mu\mu}^{rel}$ distribution computed as described.

7.2.2 jet and dijet kinematics

Kinematic properties of di-jets in selected events are shown in the following figures.

Figures 7.6, 7.7 shows the p_T distribution of the p_T -leading and subleading jets after the event selection in the Signal region for all the years. Figures 7.7 shows the pseudorapidity distributions of the p_T -subleading and subleading jets after the event selection in the Signal region for all the years.

Figure 7.8 shows the dijet invariant mass distributions after the selection in the three regions for all the years.

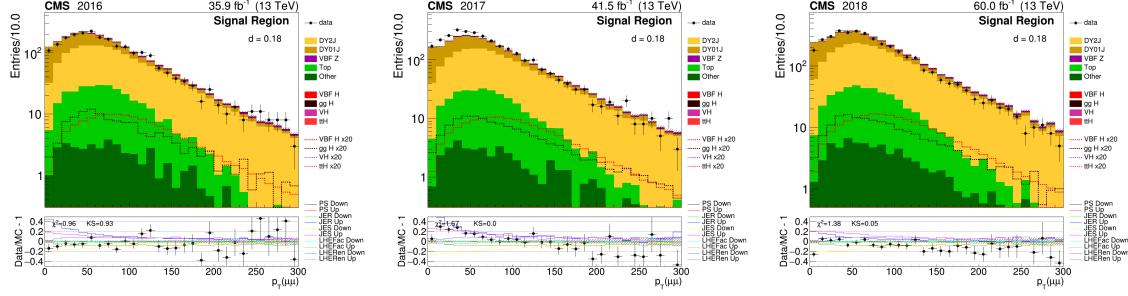


Figure 7.4: Transverse momentum distribution of the dimuon system after the event selection in in Signal Region for 2016 (left), 2017 (center) and 2018 (right).

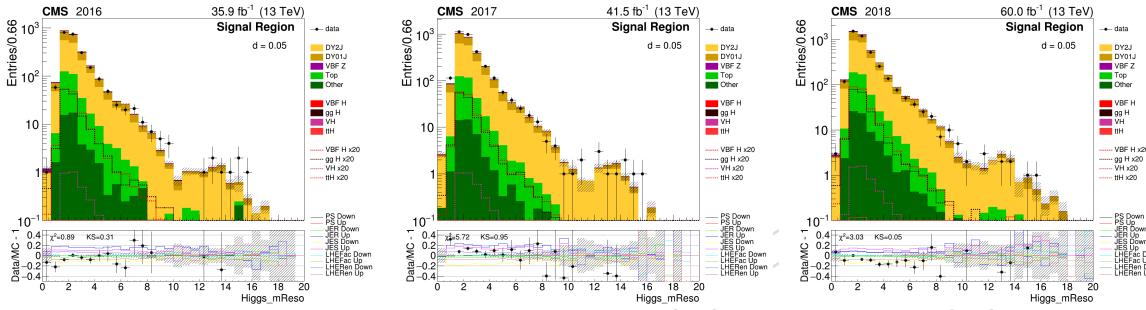


Figure 7.5: Dimuon mass uncertainty $\Delta m_{\mu\mu}$ after the event selection in in Signal Region for 2016 (left), 2017 (center) and 2018 (right).

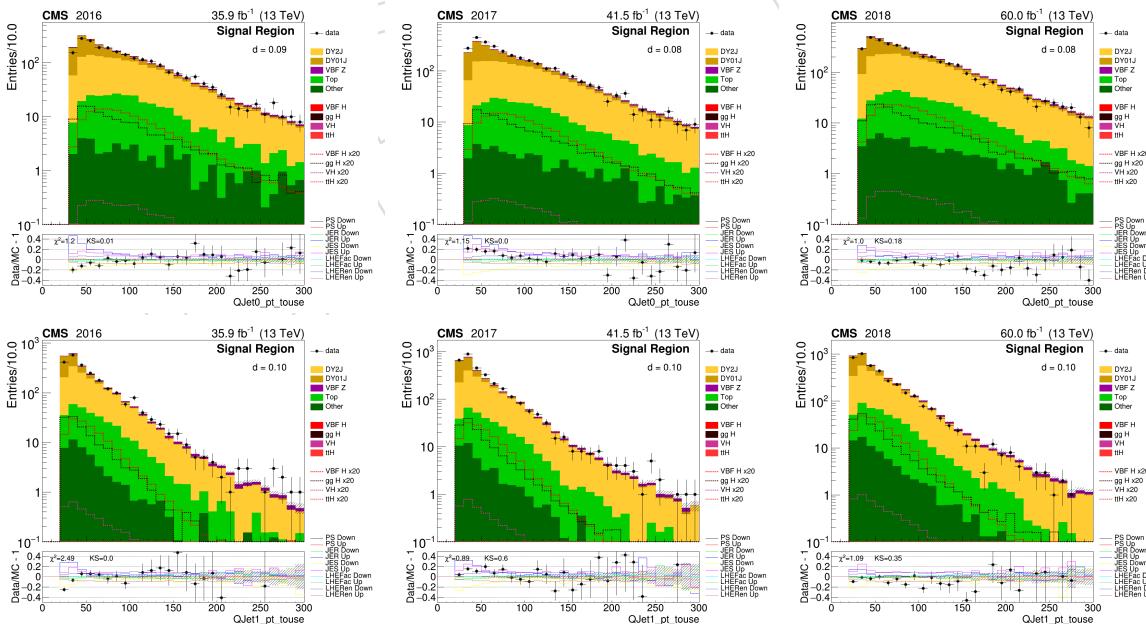


Figure 7.6: Leading and subleading jets transverse momentum distributions after the event selection in Signal Region for 2016 (left), 2017 (center) and 2018 (right).

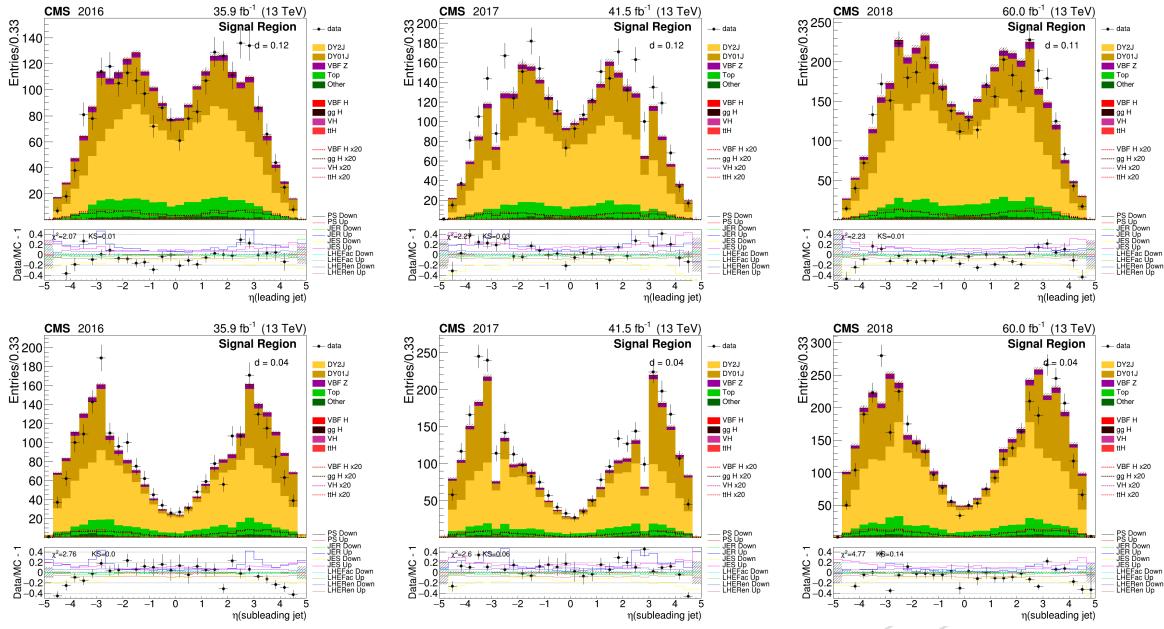


Figure 7.7: Leading and subleading jets pseudorapidity distribution after the event selection in Signal Region for 2016 (left), 2017 (center) and 2018 (right).

1920 7.2.3 Other relevant variables

1921 The “Zeppenfeld” variables $y^*(H)$ and $z^*(H)$ are defined by

$$y^*(H) = y_H - \frac{y_1 + y_2}{2} \quad z^*(H) = \frac{y^*(H)}{|y_1 - y_2|}$$

1922 where y_1 and y_2 are the rapidity of the two selected jets and y_H is the one of the dimuon system.
 1923 The variables $z^*(H)$ results in $|z^*(H)| < 0.5$ for events where the H is located in the rapidity
 1924 opening of the two jets.

1925 Figure 7.9 shows $z^*(H)$ after the event selection in Signal region for all the years.

The $\mu\mu jj$ transverse momentum balance $R(p_T)$ is defined by

$$R(p_T) = \frac{|\vec{p}_T(jj) + \vec{p}_T(\mu\mu)|}{|\vec{p}_T(j_1)| + |\vec{p}_T(j_2)| + |\vec{p}_T(\mu\mu)|}$$

1926 Figure 7.10 shows $R(p_T)$ after the event selection in Signal region for all the years.

1927 7.3 Gluon/Quark jet discrimination

1928 In the mixed electroweak+QCD production of DY+jets events, that is the main background of
 1929 the analysis, about one half of the jets produced in association with the Z/γ boson originate
 1930 from gluons. Is the case for the VBF final state, only jets initiated by final states with quarks are
 1931 possible. Differences in the showering and fragmentation of gluons and quarks have been well
 1932 established in data from LEP [86], these differences in the internal jet composition and structure
 1933 are exploited in this analysis to obtain an enhanced separation of the signal events.

1934 A comprehensive study of quark/gluon jet separation in CMS and validation of the method

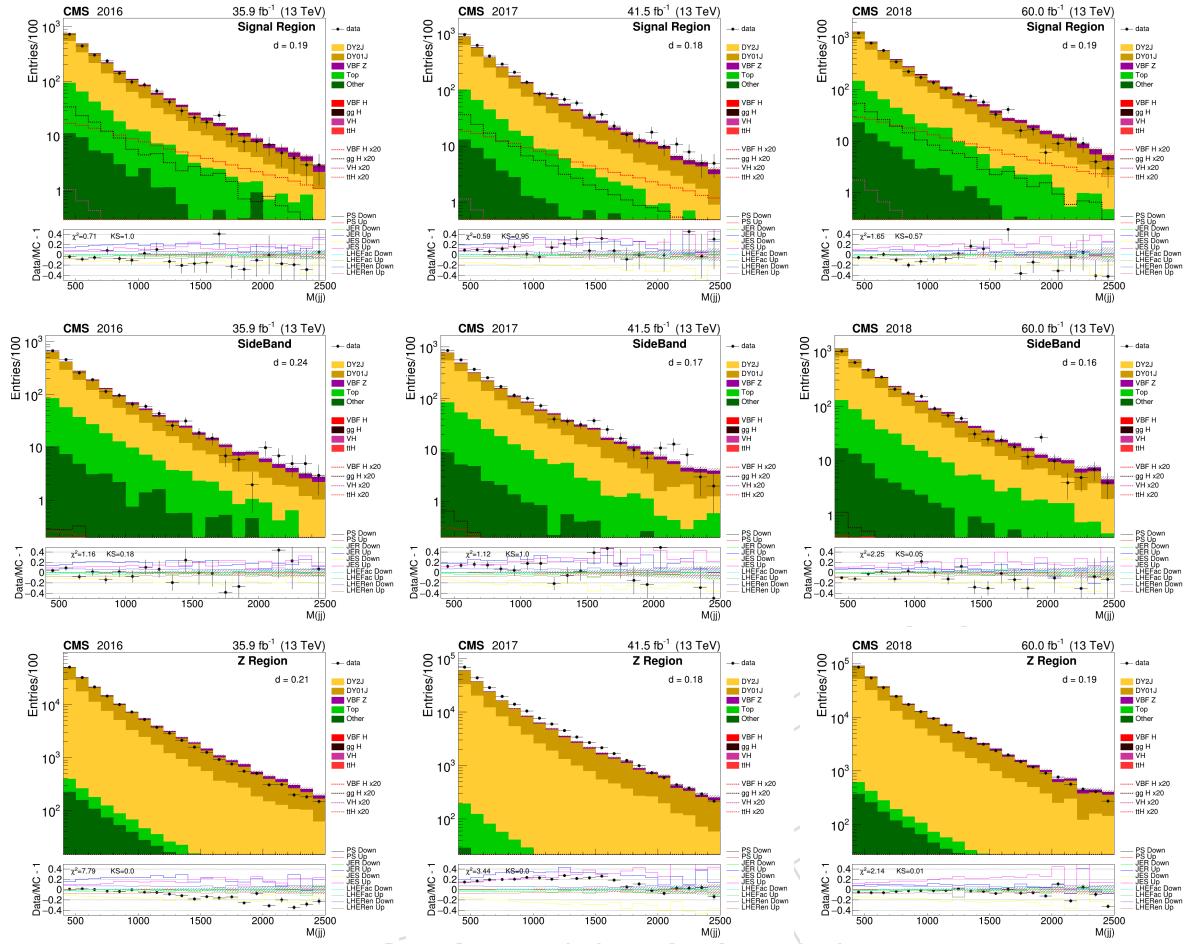


Figure 7.8: Dijet invariant mass distributions after the event selection in the three regions for all the years. From top to bottom: Signal region, Sideband, Z region. From left to right: 2016, 2017 and 2018 events.

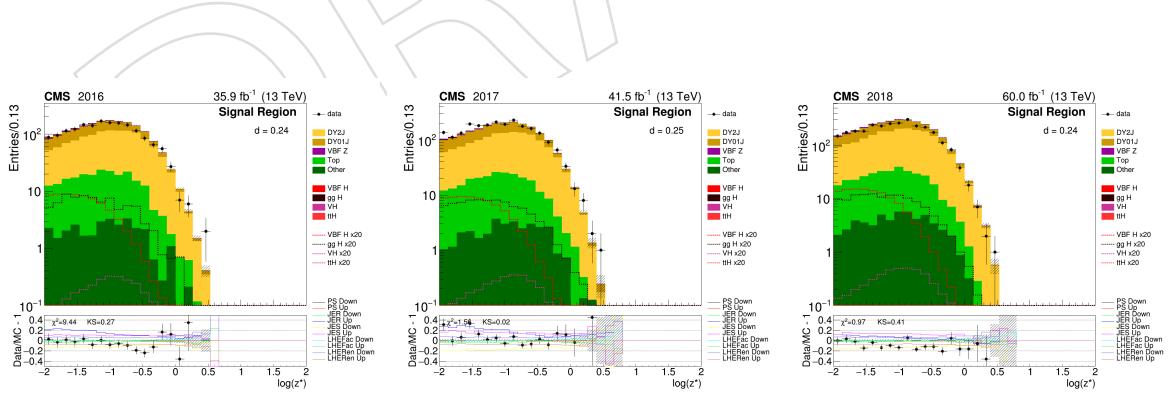


Figure 7.9: Distributions of the “Zeppenfeld” variable $z^*(H)$ after the event selection in Signal Region for 2016 (left), 2017 (center), and 2018 (right).

with independent dijet, Z+jet and γ +jet samples is available in Ref. [87] and published in Ref. [88].

The distributions of the Gluon/Quark discriminator (QGL) input variables of the generic JME POG tool [88] are

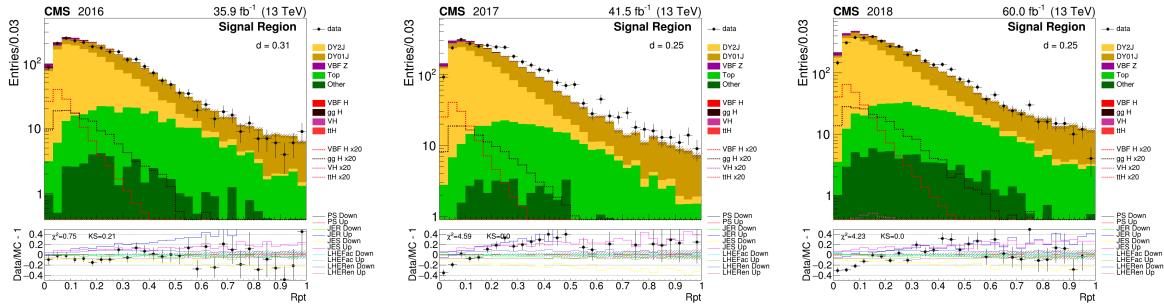


Figure 7.10: Distributions of the transverse momentum balance $R(p_T)$ after the event selection in Signal Region for 2016 (left), 2017 (center) and 2018 (right).

1939 1. the jet constituents minor RMS in the $\eta\phi$ plane,

1940 2. the jet particle multiplicity,

1941 3. the jet internal p_T distribution $p_T D$.

1942 All jet-property variables are based on the PF constituents of the jet, and the RMS variable is
1943 calculated weighting each jet constituent with its squared transverse momentum ($w_i = p_{Ti}^2$).

1944 The Gluon/Quark discriminator output values are shown in Figure 7.11.

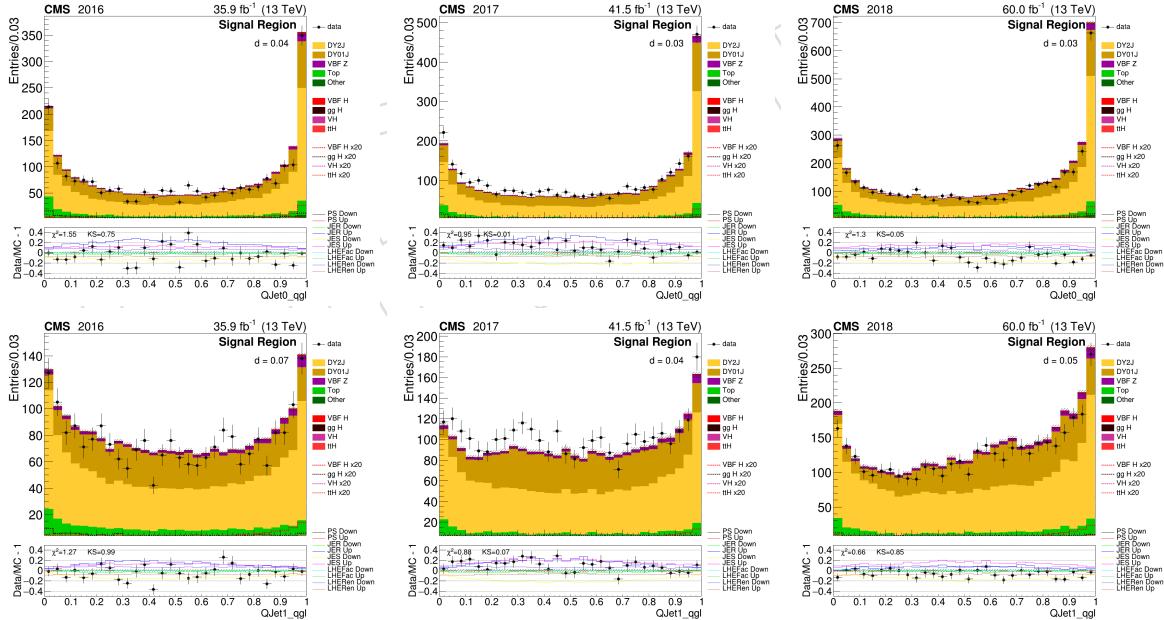


Figure 7.11: Gluon/Quark discriminator output distributions for p_T -leading and subleading jets after the event selection in the Signal Region for 2016 (left), 2017 (center) and 2018 (right).

1945 Moreover data-driven scale factors for the QGL output have been derived with balanced di-jet
1946 and Z+jet events in 13 TeV collision data and fitted with functional forms to be applied
1947 separately for quark and gluon jets in MC events [89]. The method is been also validated
1948 with different parton shower simulations [90, 91]. As recommended by the QGL authors, all

1949 nominal MC samples include the data-driven QGL corrections and the full difference with the
1950 uncorrected events is used to define the QGL systematic uncertainty band.

1951 7.4 Signal vs. background discrimination

1952 Many of the variables shown in the previous Sections 7.2 and 7.3 have been tried to built multi-
1953 variate discriminators (MVA) to optimally separate signal and background events, and mea-
1954 sure the signal presence from the MVA output distribution. In this analysis two multi-variate
1955 discriminators have been developed: a Boosted Decision Tree (BDT) and a Deep Feed-Forward
1956 Neural Network (DNN). The training of the BDT is performed with the TMVA package [92].
1957 The training of the DNN is performed with the KERAS package [93] and Tensorflow backend
1958 [94].

1959 The performances of the trained DNN are found to be slightly better than those obtained with
1960 the BDT discriminators. The DNN is therefore chosen for the analysis and described in the
1961 following, while details and results obtained with the BDT are described in Appendix L.

1962 7.4.1 Neural Network

1963 A Deep Neural Network (DNN) has been optimized in parallel with the BDT described in Ap-
1964 pendix L. In general DNNs are powerful for processing lower level inputs. This is not the case
1965 in the VBF $H \rightarrow \mu\mu$ analysis, as the multivariate analysis is performed at the end of the analysis
1966 chain, however using a DNN allowed to seek extra improvement by adding variables to the
1967 ones found by the N+1 tests. Moreover several training setups were tested in order to leverage
1968 on the flexibility of the DNNs compared to the BDTs.

1969

1970 The training is performed using simulated samples for the three years all mixed together. This
1971 allows to exploit the higher statistics of the simulation, compared to dedicated trainings for
1972 each year of the data taking. The variable "year" is added to the training and it serves as a flag
1973 for the training, so that possible discrepancies due to the simulation of different data taking
1974 conditions can be taken into account.

1975 The simulated samples used in the training are:

- 1976 • the background samples:
 - 1977 • DYJetsToLL_M-105To160*_13TeV-madgraphMLM-pythia8
 - 1978 • DYJetsToLL_M-105To160_VBFFilter*_13TeV-madgraphMLM-pythia8
 - 1979 • EWK_LLJJ_MLL_105-160*_13TeV-madgraph-herwig*
 - 1980 • TT_*-13TeV-powheg-pythia8 (inclusive W decays)
- 1981 • the signal sample:
 - 1982 • VBF_HToMuMu_M125*_13TeV_powheg_pythia8

1983 The signal and the VBF Z (EWK_LLJJ_MLL_105-160*_13TeV-madgraph-herwig*) samples
1984 are used both in the training and in the final evaluation of the DNN. In order not lose half of the
1985 statistical power of the sample in the final evaluation, a 4-fold procedure is used at training time.
1986 Half of the training samples are used for the training. A quarter is used for the validation,
1987 i.e. to choose the best performing model. The model is evaluated at test time, i.e. for the
1988 final evaluation and extraction of the significance, on the remaining quarter of the events. This
1989 procedure is performed for all the samples used in the training, including the ones which are
1990 not used later. This allows to partially regularize the final result, as the fluctuations of single
1991 subsets of events are reduced by a factor 2. The four subsamples are defined using the event

number modulus remainder, the same procedure is used for samples not used in training (e.g. data and other MC samples) in order to decide which fold to apply in the final evaluation.

A schematic representation of the k-fold procedure with 4 folds is shown if figure 7.12.

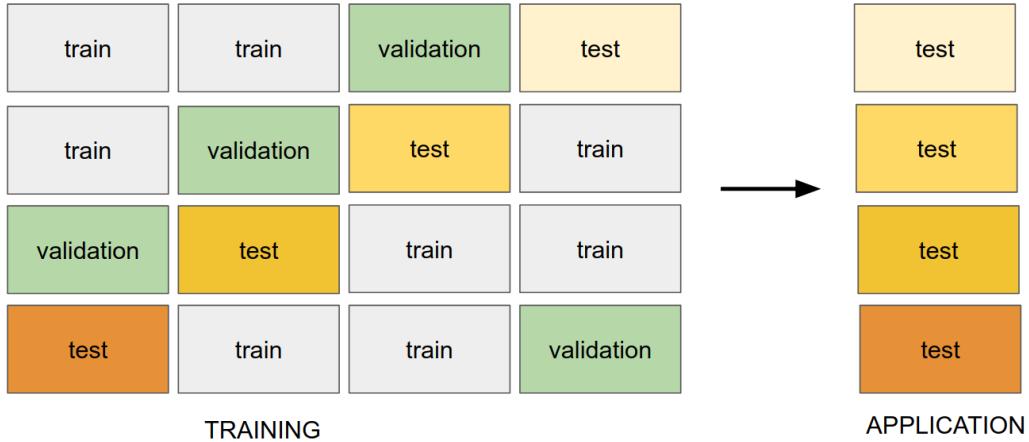


Figure 7.12: Scheme of the 4-fold training, validation and evaluation procedure

1994

1995 The events are weighted at training time by applying the event weights used in the anal-
1996 ysis. The weights are then divided by the average weight of each class (signal and back-
1997 ground), in order to obtain weights of order 1 and the relative weights are consistent in the
1998 background class. The top events (`TT_*`_13TeV-powheg-pythia8) are down-weighted, as
1999 they have larger weights compared to the other background simulations, and they are a rel-
2000 atively easy background for the final discriminator. Class weights can be applied depending on
2001 the training setup. Before being fed to the DNN all the variables are standardized. The sample
2002 mean is subtracted from each value and the result is divided by the sample standard deviation:
2003 as a result, the new distributions have mean 0 and standard deviation 1.

2004 The training is performed in multiple steps and with a suitable architecture.

2005 Four networks are first optimized independently with different inputs and for different tasks.
2006 The outputs of the last hidden layer nodes, are then merged and combined to solve the ac-
2007 tual classification problem. The final stage of the training consists in fine-tuning the model by
2008 unfreezing some of the upstream layers and training them together with the down stream ones.

2009 The loss function used at each step is the binary cross-entropy or "log-loss", which results in
2010 the maximum likelihood estimator for binary classification problems.

2011 Each networks is made of 3 to 4 hidden layers with a few tenths of nodes in each hidden
2012 layer and a pyramidal architecture. A 20% dropout is used after each hidden layer in order to
2013 regularize the model. The learning rate is also gradually decreased at training time based on
2014 the validation loss, thus preventing over-training.

2015 A schematic representation of the DNN architecture is shown in figure 7.13. The grey block
2016 indicate the DNNs optimized for single tasks, with their output in blue. The last hidden layer
2017 outputs for the 4 networks are merged in as single vector and used as input for a combination,
2018 whose output is shown in red. The preliminary tasks of the training are aimed at optimizing
2019 single backgrounds rejection and exploiting the event topology independently of the mass. The
2020 four preliminary steps are:

- 2021 • (1) signal -vs- VBF Z
 2022 • (2) signal -vs- DY
 2023 • (3) mass independent signal -vs- background
 2024 • (4) mass + mass resolution (a quick pre-training is performed just as input to the
 2025 combination)

2026 The combination step (5.a) uses all the information with coming from the networks (1,2,3),
 2027 which are frozen. The network (4) weights are left unfrozen at this stage. A fine tuning is then
 2028 performed (5.b), where the weights of the network (3) are also unfrozen. All the stages use a
 2029 mini-batch size of 1024 events, while in the final step a few epochs have the same mini-batch
 2030 size, and the very final epochs have a 10240 events mini-batch size.

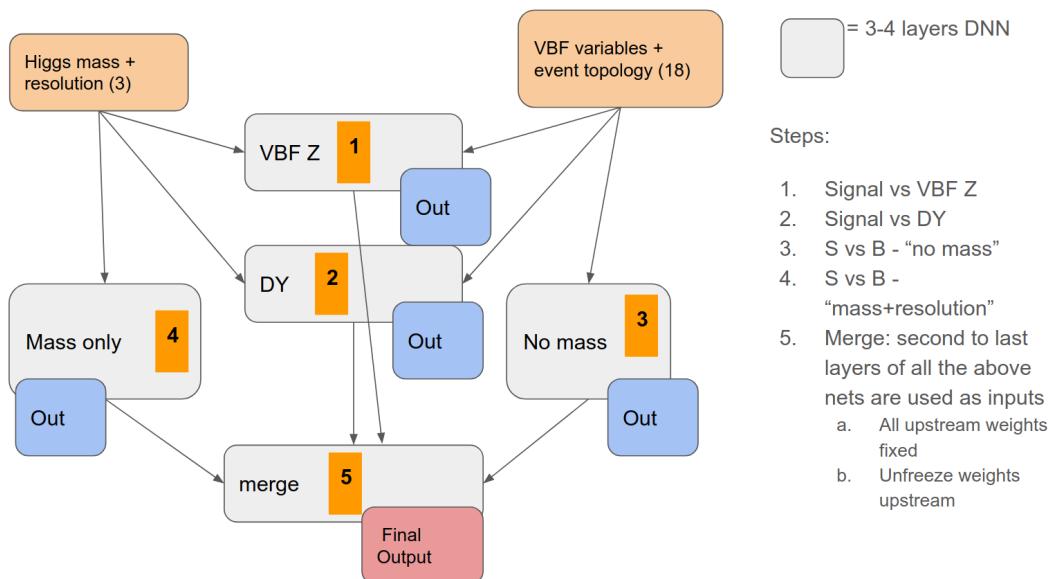


Figure 7.13: Schematic representation of the DNN architecture: the training procedure consists in optimizing for single tasks, combining the outputs and fine-tuning the model by unfreezing upstream weights with appropriate learning rate.

2031 The variable used in the current setup of the DNN are:

- 2032 • $m(\mu\mu)$, $\Delta m(\mu\mu)_{rel}$, $\Delta m(\mu\mu)$ - the dimuon mass and the relative and absolute mass
 2033 resolutions
 2034 • $m(jj)$, $\log m(jj)$ - the dijet mass and its logarithm
 2035 • $R(p_T)$
 2036 • Z^*
 2037 • $\Delta\eta(jj)$ - the pseudorapidity difference between the 2 selected jets
 2038 • N_5^{soft} - # soft jet with $p_T > 5 \text{ GeV}$
 2039 • $\min_j \Delta\eta(\mu\mu, j)$ - the minimum pseudorapidity difference between a jet and the dimuon
 2040 system
 2041 • $p_T(\mu\mu)$, $\log p_T(\mu\mu)$, $\eta(\mu\mu)$ - dimuon 4-vector components
 2042 • $p_T(j_1)$, $p_T(j_2)$, $\eta(j_1)$, $\eta(j_2)$, $\phi(j_1)$, $\phi(j_2)$ - jets' 4-vectors components
 2043 • $qgl(j_1)$, $qgl(j_2)$ - the quark-gluon likelihood discriminators for the selected jets.

2044 The first three variables are the ones used by the network (4) as shown in figure 7.13, while the

2045 remaining 18 together with the year are used in the stage (3). All the variables are employed
 2046 when training against single backgrounds (1,2), and in the for the output classifier (5).

2047 The best training is chosen using the estimated significance in simulation. The significance is
 2048 computed both for the training and validation fold, and the minimum significance is used as
 2049 to pick the best model. The significance is evaluated using the "Asimov" significance in bins
 2050 containing 0.5 expected signal events.

2051 The signal and background DNN output distributions are shown in figure 7.14

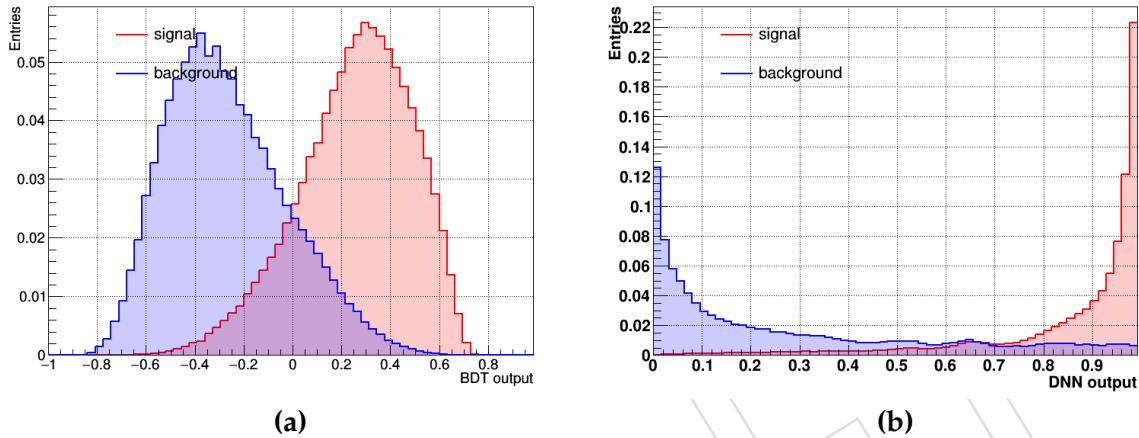


Figure 7.14: Plot of the signal and background normalized distributions for the BDT and the DNN. The simulated samples of 2016, 2017 and 2018 are used all together.

2052 Several training setups were tested: the results are shown in figure 7.15. The DNN is also
 2053 compared to the BDT trained using the same samples.

2054 The data and Monte Carlo distributions of the DNN output values are shown in Figure 7.16 for
 2055 all the region for all the years; Drell-Yan is computed using NLO. DNN output in Figure 7.16
 2056 is computed with a fixed value of $m(\mu\mu)=125$ GeV for control region events.

2057 7.5 Systematic uncertainties

2058 The systematic uncertainties are divided into experimental and theoretical, they can affect both
 2059 the overall signal acceptance and the shape of the DNN output. Experimental uncertainties
 2060 take into account differences between a measurements and the true values of observables. The-
 2061 thetical uncertainties are related to mismodelling in the MC sample generation. In the follow-
 2062 ing Sections, details on the various systematic uncertainties are given. A generic introduction to
 2063 various sources of systematic uncertainties considered in this analysis was given in Section 5.4.

2064 7.5.1 Experimental uncertainties

2065 This section describes experimental uncertainties considered in this analysis.

- 2066 • **Luminosity (norm. only):** the luminosity affects directly the predicted signal and
 2067 background yields. We use the currently recommended recipe given by the Lumi-
 2068 POG [95].
- 2069 • **Single muon trigger and muon selection efficiencies (norm. and shape):** uncer-
 2070 tainties on the data driven corrections for the trigger, and offline muon selections

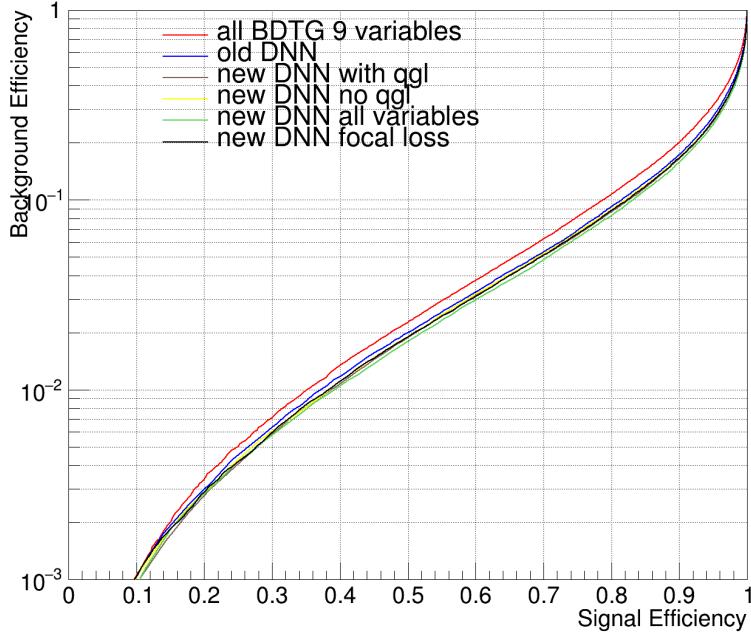


Figure 7.15: Roc curves comparing the signal versus background efficiency for the BDT, the DNN trained with 2018 samples only and several versions of the DNN trained on the 2016+17+18 simulation.

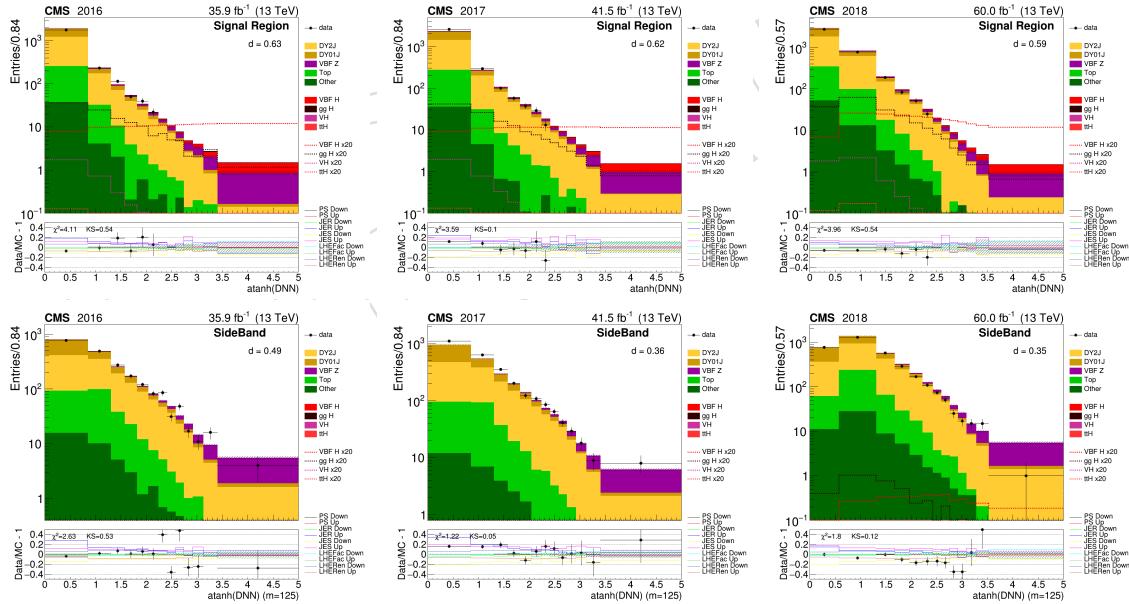


Figure 7.16: DNN output distribution after the event selection in Signal (top) and Sideband (bottom) regions for 2016 (left), 2017 (center) and 2018 (right). DNN output distributions are computed with $m(\mu\mu)=125$ GeV in the Sideband region.

are taken from the measurements performed by the Muon POG and amount to a total of 2-3%. Their effect is correlated across DNN bins, analysis regions, and eras.

- **Muon scale and resolution (norm. and shape):** the muon momentum scale and resolution correction uncertainties, as computed by the Rochester method, are used. Their effect is correlated across DNN bins, analysis regions, and eras.
- **Pileup uncertainty (shape):** an uncertainty on the number of primary vertices is derived by varying the minimum bias cross section used in the pileup re-weighting applied to MC samples. The corresponding effect is correlated across DNN bins, analysis regions, and eras.
- **Jet energy scale (norm. and shape):** obtained by varying the transverse momentum of each jet by the uncertainties provided by the JetMet POG as described in Section 5.4.1. This results in a set of uncertainties (nuisance parameters) affecting both signal and background acceptance as well as the shape of the DNN output. The corresponding nuisance parameters are correlated across regions and eras (when recommended by JetMet), and induce significant correlations between DNN bins.
- **Jet energy resolution (norm. and shape):** obtained by smearing the transverse momentum of each jet by the smearing factors provided by the JetMet group, following the recipe already described in Section 5.4.1. This results in a set of uncertainties (nuisance parameters) affecting both signal and background acceptance as well as the shape of the DNN output. The corresponding nuisance parameters are correlated across DNN bins and regions but uncorrelated across eras.
- **Quark-gluon likelihood (shape):** uncertainties on QGL discriminator are evaluated following the recommended recipe from the JetMet POG. The corresponding effect is correlated across DNN bins, analysis regions, and eras.
- **Prefiring (norm. and shape):** the uncertainty due to the pre-firing corrections affect only 2016 and 2017 eras and is estimated following the official prescription from the JetMet group. This uncertainty varies from 0.3–1.5% (0.7–2%) as a function of the DNN score in the 2016 (2017) era, and its effect is correlated across DNN bins and regions.
- **Statistics of simulation:** for all signal and background processes, the per-bin statistical uncertainty arising from the limited size of the simulated samples is taken into account and affects mostly the high score region of the discriminators. This statistical uncertainty is modelled using the autoMCStat algorithm implemented in the Higgs combine tool.
- **Drell-Yan contribution from pileup and noise:** a significant fraction (about 30–40%) of the DY background populating the low score DNN bins comprises events in which either the leading or the subleading jet fall in the forward region of the detector ($|\eta| > 3.0$) and are not matched with a jet at the generator level. These jets originate either from soft emission produced by the parton shower or from pileup interactions, and are promoted above the p_T thresholds used in the analysis by the detector response. The remainder of the DY events contain two jets matched to generator-level jets primarily arising from the quarks at matrix element level. The normalizations of these two distinct components of the DY background are allowed to float freely in the signal extraction, constrained only by low score events of the VBF-SR and VBF-SB. A detailed description of the DY composition and studies of the robustness of the fit against bias arising from a potential large miscalibration of the forward jet resolution is available in Appendix B of Ref. [18].

2118 7.5.2 Theory uncertainties

- 2119 • **VBF STXS stage 1 (norm. and shape):** these uncertainties are evaluated following
2120 the LHCHXSWG recommendations as described in Section 5.4.2.
- 2121 • **ggH STXS stage 1 (norm. and shape):** these uncertainties are evaluated following
2122 the LHCHXSWG recommendations as described in Section 5.4.2.
- 2123 • **Renormalization and factorization scale (norm. and shape):** uncertainties from
2124 the choice of renormalization and factorization scales are evaluated, for background
2125 processes, as described in Section 5.4.2.
- 2126 • **Parton distribution functions (norm. and shape):** the uncertainty due to the parton
2127 distribution functions is evaluated by applying the recipe described in Section 5.4.2.
2128 For each process and PDF replica, the expected yield in each DNN bin is computed.
2129 The ratios between the replicas and the nominal histogram are used to compute the
2130 corresponding uncertainty. The envelope of these ratios is fitted with two different
2131 functions: $y = mx + b$ and $y = ax^2 + b$. In order to build a variation of the nominal
2132 histogram, the bin contents of the nominal distributions are multiplied by either b' ,
2133 $m'x$ or $a'x^2$, where x is the value of the bin center, m' , a' , b' are the RMS of the para-
2134 meters m , a , b as given by the previous fits. The normalization effects are removed
2135 for the two non constant variations. Down variations are obtained flipping the sign
2136 of m' , a' , and b' . The corresponding nuisance parameters are named $PDFX0,1,2$,
2137 respectively. Their effects are correlated across DNN bins, regions, and eras but un-
2138 correlated between processes. In other words, the PDF uncertainties ($PDFX0,1,2$)
2139 are defined to be process dependent.
- 2140 • **Parton shower acceptance uncertainty (norm. and shape) for VBF-H:** PS uncer-
2141 tainty is computed for signal processes and for the VBF-Z samples only. For the
2142 signal samples order of 5-10% acceptance and/or shape differences are observed in
2143 the predictions obtained with different combinations of ME generator (AMC@NLO
2144 and POWHEG) and parton shower (different versions and configurations of PYTHIA
2145 and HERWIG), as already described in Section 2.2. We decided to use as central pre-
2146 diction the one obtained with POWHEG +PYTHIA with dipole recoil and to take the
2147 full difference with an alternative setup, POWHEG +HERWIG 7, as uncertainty. A
2148 detailed description is available in Appendix A of Ref. [18]. The uncertainty is com-
2149 puted for 2017 and 2018 datasets and the ratios obtained are also applied to 2016
2150 for which the HERWIG 7 sample is not available. The uncertainty is symmetrized by
2151 inverting the ratio in order to have a two sided nuisance.
- 2152 • **Parton shower acceptance uncertainty (norm. and shape) for VBF-Z:** for the VBF-
2153 Z sample a specific recipe to assess the acceptance uncertainty from parton-shower
2154 has been designed. Two alternative samples showered with different PS programs
2155 (PYTHIA dipole shower and HERWIG angular-ordered shower) are used to estimate
2156 the PS uncertainty. From SMP studies [35, 85], the PYTHIA shower in global recoil
2157 mode (standard p_T -ordered shower) is known to mismodel the additional hadronic
2158 activity in VBF-Z and VBF-W events, while predictions from HERWIG are in better
2159 agreement with the observed data. In contrast, events showered with PYTHIA in
2160 dipole mode are expected to show a good agreement with data at a level comparable
2161 with HERWIG based predictions. However, a comparison between data and PYTHIA
2162 dipole shower for VBF-Z events has not been performed so far. Therefore, HERWIG
2163 is used to derive the central prediction for the VBF-Z process and the full difference
2164 between the predictions from the two PS programs is considered as an uncertainty,
2165 which varies as a function of the DNN score from 2 to 8%.

- **Additional uncertainty on DY model:** it has been observed that some DNN input variables have slightly different distribution, at generator level, for the 2016 DY sample compared to the 2017/18 DY samples. A much better closure for the reconstruction level variable is achieved in 2016 (see e.g. Figure 7.10). For this reason an additional shape uncertainty for the DY component has been taken from the ratio of the 2016 vs 2017/18 predictions.

7.6 Results

Results for signal yields are expressed in terms of the ratio of the measured signal cross-section over the Monte Carlo expected cross-section μ_S .

The signal is extracted by performing a binned maximum likelihood fit of the data distributions using the Monte Carlo signal and background histograms shapes. The fit is simultaneously performed in the SignalRegion and SideBand as defined in section 7.1. Systematic uncertainties are considered as fully correlated between the two regions. In each region the DNN output is fit. In the case of the SideBand region the DNN is evaluated by replacing the input variable corresponding to the dilepton mass with a fixed value of 125 GeV.

The distributions have been re-binned attempting to minimize the total number of bins in order to avoid spurious constraints in the nuisance parameters. The binning has been defined to have a constant number of signal events per bin. As no improvement was observed with finer binning, a choice was made to have 0.6 signal events per bin.

The data is split in the three different years of data taking (2016,17,18) and fit independently. A combined fit is then performed to the global μ_S keeping theoretical systematic uncertainties as correlated and decorrelating the experimental ones as suggested by the various physics object groups. The fits have been performed using a RooStats-based statistics combine tool developed by the Higgs PAG [96]. Limits on the signal strength are computed with the Asymptotic CLs method [81]. Systematic uncertainties sources described in Sec. 7.5 are taken into account with additional nuisance parameters that modify the likelihood function in a frequentist manner [97].

7.6.1 Expected results

This section summarizes the expected sensitivity to the SM signal with $m_H = 125$ GeV. In order to have a more accurate prediction, the nuisance parameter central values are adjusted from initial fits to the control regions.

The expected significance for each of the fits (the separated three years and the combined one) are shown in table 7.4.

Year	Expected significance			
	Dipole recoil sample		central production sample	
	Prefit	Postfit	Prefit	Postfit
2016	1.04	1.04	0.95	0.94
2017	0.89	0.88	0.81	0.81
2018	1.17	1.21	1.09	1.12
Combined	1.78	1.80	1.63	1.66

Table 7.4: Expected significance using POWHEG +PYTHIA with and without dipole recoil in PYTHIA showering.

2199 The expected impact of the various sources of systematic uncertainties is shown in Figures 7.17,
 2200 together with fitted value of each nuisance in the control regions only fit. The leading theoretical
 2201 uncertainty are those originating from parton shower modelling of the signal and the EWK
 2202 Z_{jj} production, while the most relevant experimental effect is due to the large uncertainty in the
 2203 prediction of the DY with pileup/noise jets. The sensitivity of the analysis to the latter effect is
 2204 not surprising as the VBF signature is characterized by a considerable amount of events with
 2205 soft jets (subleading jet distribution peaking at about 30 GeV) in the forward region where no
 2206 tracker coverage is available. This phase space is particularly difficult from an experimental
 2207 point of view and a large uncertainty has been used to predict the DY background in that re-
 2208 gion. Overall the uncertainty is dominated by statistical effects and the impact of no nuisance
 2209 is highly reduced by strong constraints.

2210 7.6.2 Observed results

2211 The data and simulation for the fitted observables are shown in Figures 7.18, 7.19, and 7.20,
 2212 both for the pre-fit expectation and the post-fit prediction. The fit is performed simultaneously
 2213 across regions (VBF-SB, VBF-SR) and data-taking periods. In the top panel, the blue histogram
 2214 stacked on top of the post-fit backgrounds indicates the total signal ($ggH + VBF$) extracted from
 2215 the fit. In contrast, the non-stacked solid black and red distributions indicate the pre-fit (SM)
 2216 expectations for the VBF and ggH signals, respectively. The observed data are in agreement
 2217 with the signal-plus-background prediction in all regions within uncertainties.

2218 Figure 7.21 shows the best-fit nuisance parameter values and their impacts on the measured
 2219 signal strength for the unblinded fit to the combined Run-2 data. No significant pulls or con-
 2220 straints are observed. The nuisance parameters with leadings impacts are very similar to the
 2221 blinded control region-only fit. Note that the expected impacts shown in the previous sec-
 2222 tion follow an older nuisance parameter naming convention from before the overall $H \rightarrow \mu\mu$
 2223 combination. Also note that some impacts are expected to be highly asymmetric because of
 2224 their asymmetric definitions, particularly for the parameters related to the jet energy resolu-
 2225 tion. Figure 7.22 shows the expected distribution of the goodness-of-fit test statistic for fits
 2226 to 1050 toys thrown with nuisance parameter values obtained from a background-only fit to
 2227 the data. The value of the test statistic for the observed data is indicated by the blue arrow.
 2228 The observed data are consistent with the expectation from the toys, with an observed p-value
 2229 of 0.40, demonstrating consistency between the predicted background and the observed data
 2230 while remaining agnostic to the potential observed signal.

2231 The observed significances and best-fit signal strengths from the unblinded fit to the VBF Signal
 2232 and Sideband regions are summarized in Table 7.5. An excess is observed over the background-
 2233 only expectation in the 2016 and 2018 data, while a small deficit is observed in 2017 data. The
 2234 measured signal rate for the combined Run-2 fit is consistent with the expectation for the SM
 2235 Higgs boson with $m_H = 125$ GeV, with a signal strength of $\mu = 0.95 \pm 0.60$. The corresponding
 2236 significance of this excess over the background-only expectation is 1.72 standard deviations.

Year	Observed significance	Best-fit signal strength
2016	1.27	$\mu = 1.39^{+1.31}_{-1.11}$
2017	0	$\mu = -0.30^{+0.94}_{-0.79}$
2018	1.89	$\mu = 1.58^{+1.03}_{-0.88}$
Combined	1.72	$\mu = 0.95^{+0.64}_{-0.56}$

Table 7.5: Observed significances and best-fit signal strengths for the VBF-only fit to data.

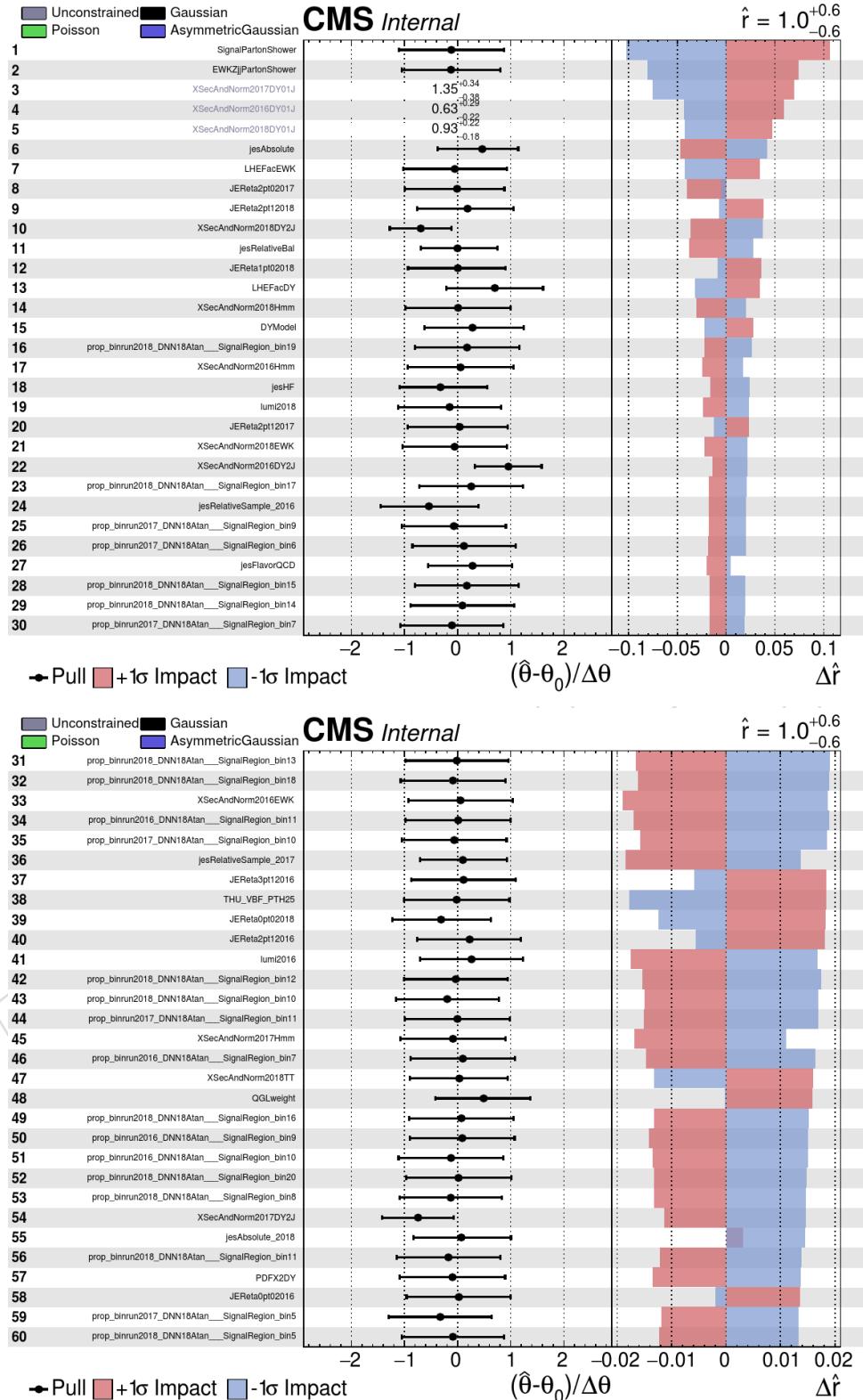


Figure 7.17: Nuisance parameters fitted values (in the control region only fit) and their impact on the μ_S measurement

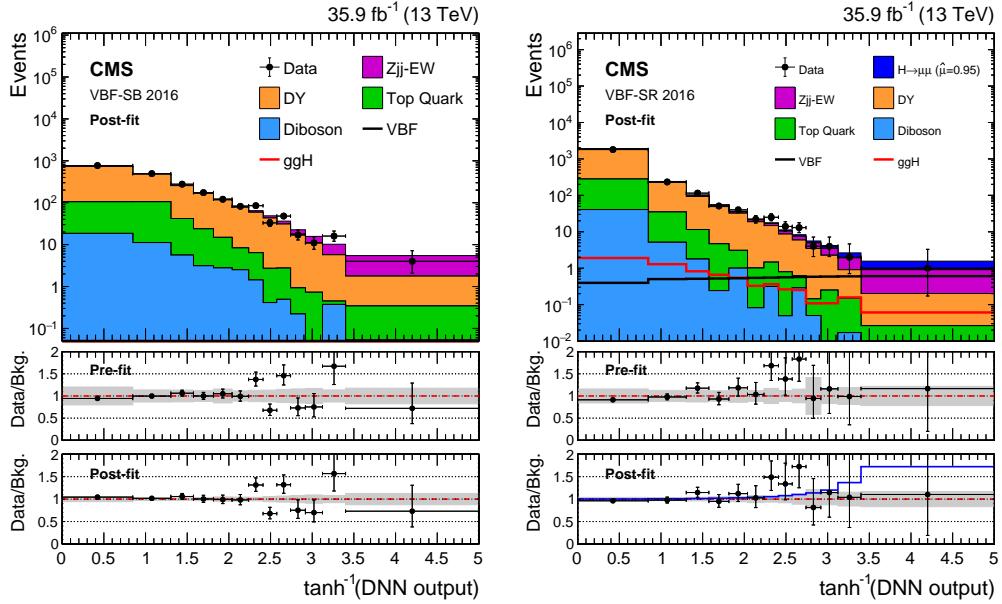


Figure 7.18: The observed DNN output distribution in the VBF-SB (left) and VBF-SR (right) regions compared to the post-fit background estimate for various SM processes. The predicted backgrounds are obtained from a combined signal-plus-background fit performed across analysis regions and eras. Distributions reported are related to the 2016 data-taking period. In the second panel, the ratio between data and the pre-fit background prediction is shown. The gray band indicates the total pre-fit uncertainty obtained from the systematic sources previously described. The third panel reports the ratio between data and the post-fit background prediction from the signal-plus-background fit. The gray band indicates the total background uncertainty after performing the fit, while the blue histogram refers to the total signal extracted from the fit.

2237 7.6.3 Mass scan

2238 Given that most of the DNN input variables have very little correlation with the dilepton mass,
 2239 the DNN can be used without retraining to probe different values of the Higgs mass. This
 2240 is achieved, in the range [120,130] GeV, by biasing the dilepton mass, both in data and in the
 2241 MC samples, with a known value ΔM so that the DNN become sensitive to a Higgs mass of
 2242 $m' = 125\text{GeV} - \Delta M$. A scan of ΔM in steps of 0.5 GeV has been performed.

2243 The expected p-value, in presence of a 125 GeV Higgs signal, as a function of the probed
 2244 mass is shown in Figure 7.23

2245 The Likelihood model depends on the Higgs mass in a discontinuous way since it may change
 2246 abruptly when an event moves across a bin edge in the DNN scores. In order to understand
 2247 how statistical fluctuations would show up in the p-value in the mass scan, the procedure of
 2248 the mass scan is tested with a few fullsim MC toys. These toys are generated by obtaining
 2249 the expected numbers of events for each physics process and in both the signal and sideband
 2250 regions by tossing a Poisson distribution using the expected yields, then randomly sampling
 2251 the numbers of events obtained in the previous step from the corresponding MC samples, and
 2252 finally putting all those MC events together as the MC toy. This procedure is tested using ten
 2253 toys with the 125 GeV Higgs boson as the signal and six examples are shown in Figure 7.24. This
 2254 study shows the level of statistical fluctuations that may present in data and also shows that
 2255 the mass scan cannot resolve well Higgs boson mass around 123-127 GeV, which is expected
 2256 from the FWHM of the dimuon mass resolution of about 3.9 GeV and the limited number of
 2257 signal events in the DNN bins with the largest signal-to-background ratio.

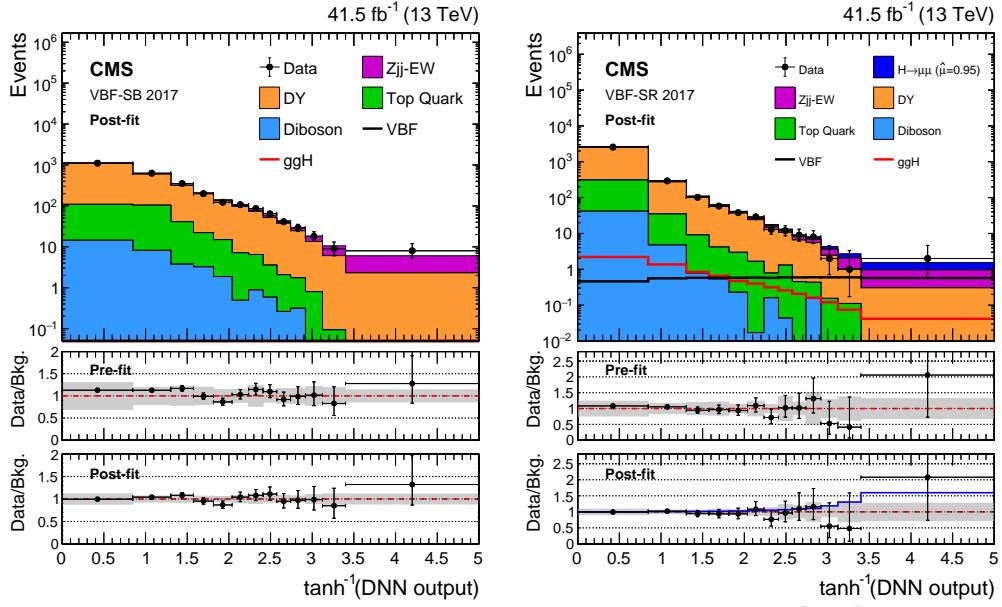


Figure 7.19: The observed DNN output distribution in the VBF-SB (left) and VBF-SR (right) regions compared to the post-fit background estimate for various SM processes. The predicted backgrounds are obtained from a combined signal-plus-background fit performed across analysis regions and eras. Distributions reported are related to the 2017 data-taking period. The description of ratio panels is the same as in Fig. 7.18

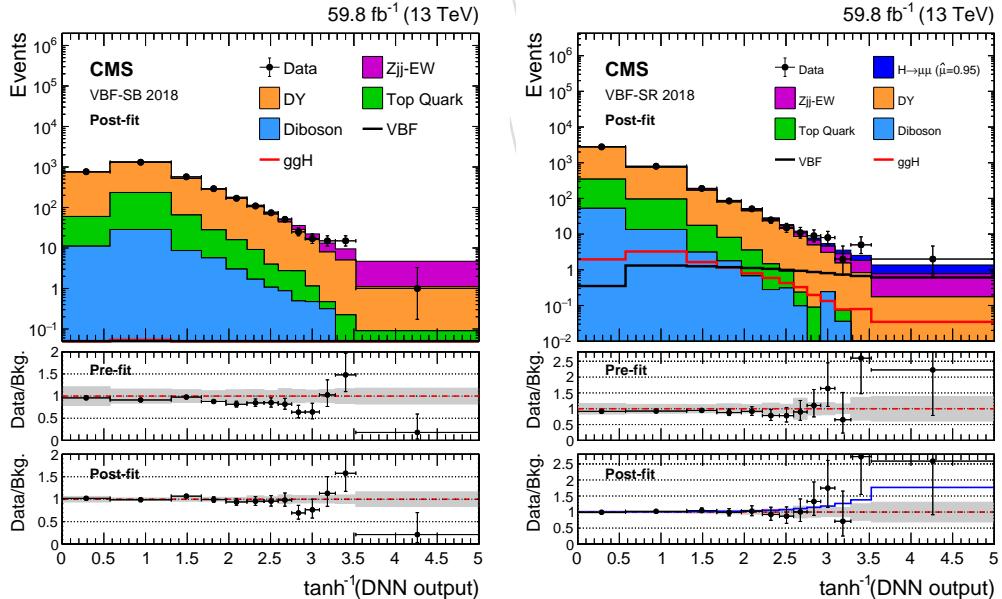


Figure 7.20: The observed DNN output distribution in the VBF-SB (left) and VBF-SR (right) regions compared to the post-fit background estimate for various SM processes. The predicted backgrounds are obtained from a combined signal-plus-background fit performed across analysis regions and eras. Distributions reported are related to the 2018 data-taking period. The description of ratio panels is the same as in Fig. 7.18.

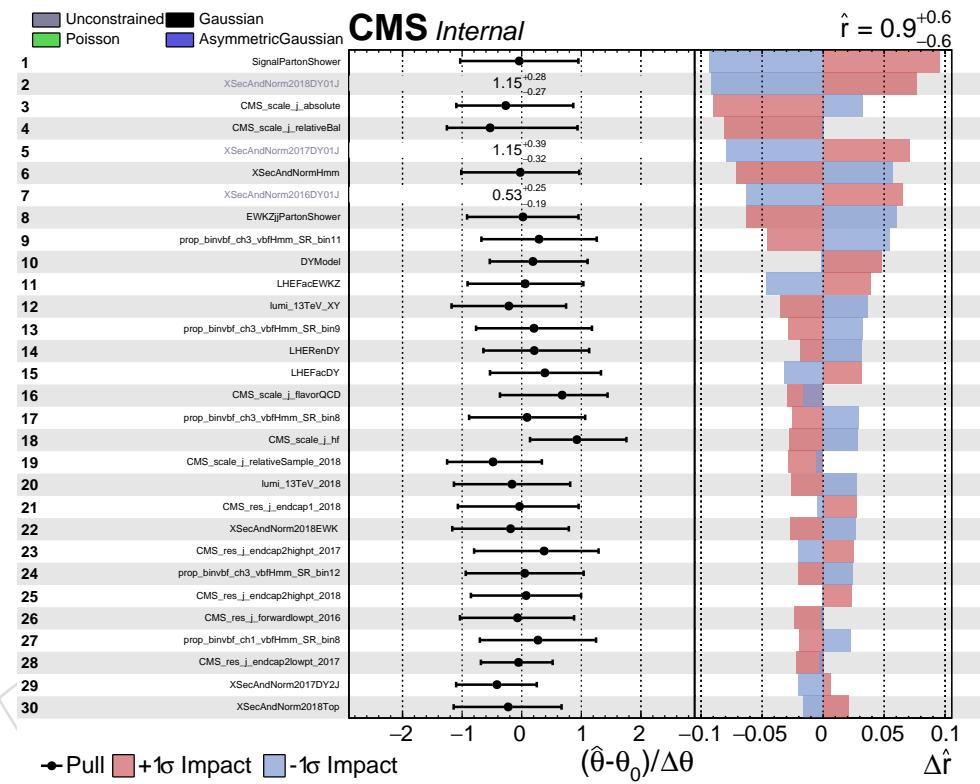


Figure 7.21: Nuisance parameter fitted values and their impact on the μ_S measurement for the unblinded fit to the combined Run-2 data.

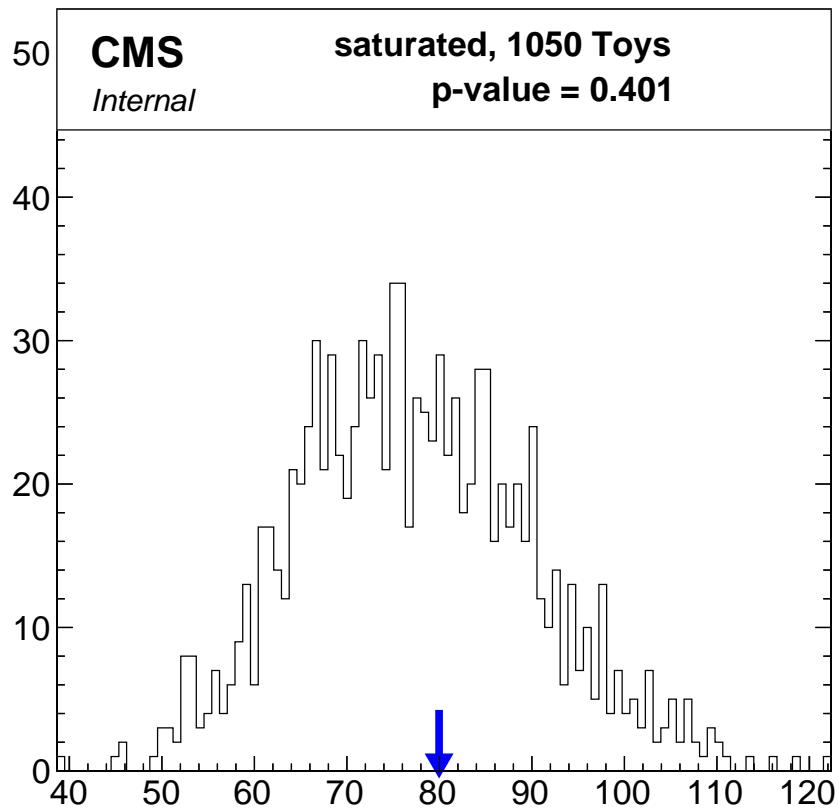


Figure 7.22: Distribution of the goodness-of-fit test statistic for 1050 toys thrown from the background-only fit result. The test statistic for the observed data, indicated by the blue arrow, is consistent with the expectation from the toys, with a p-value of 0.40.

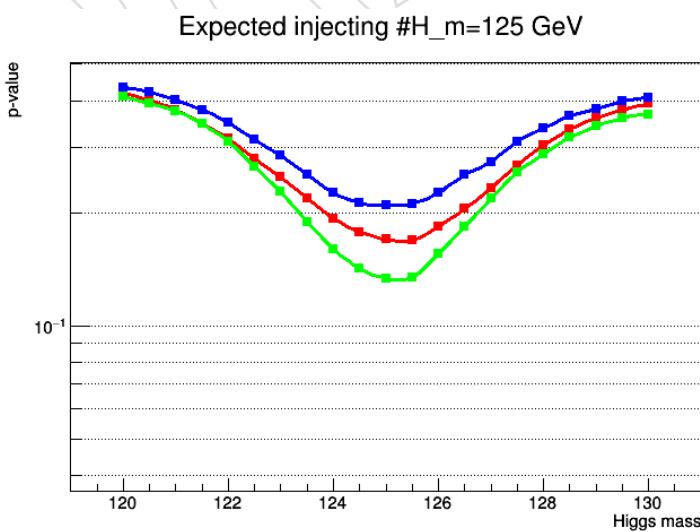


Figure 7.23: p-value in the mass scan injecting a SM signal at $m = 125$ GeV for the VBF analysis (different colors correspond to different years: 2016 in red, 2017 in blue, 2018 in green).

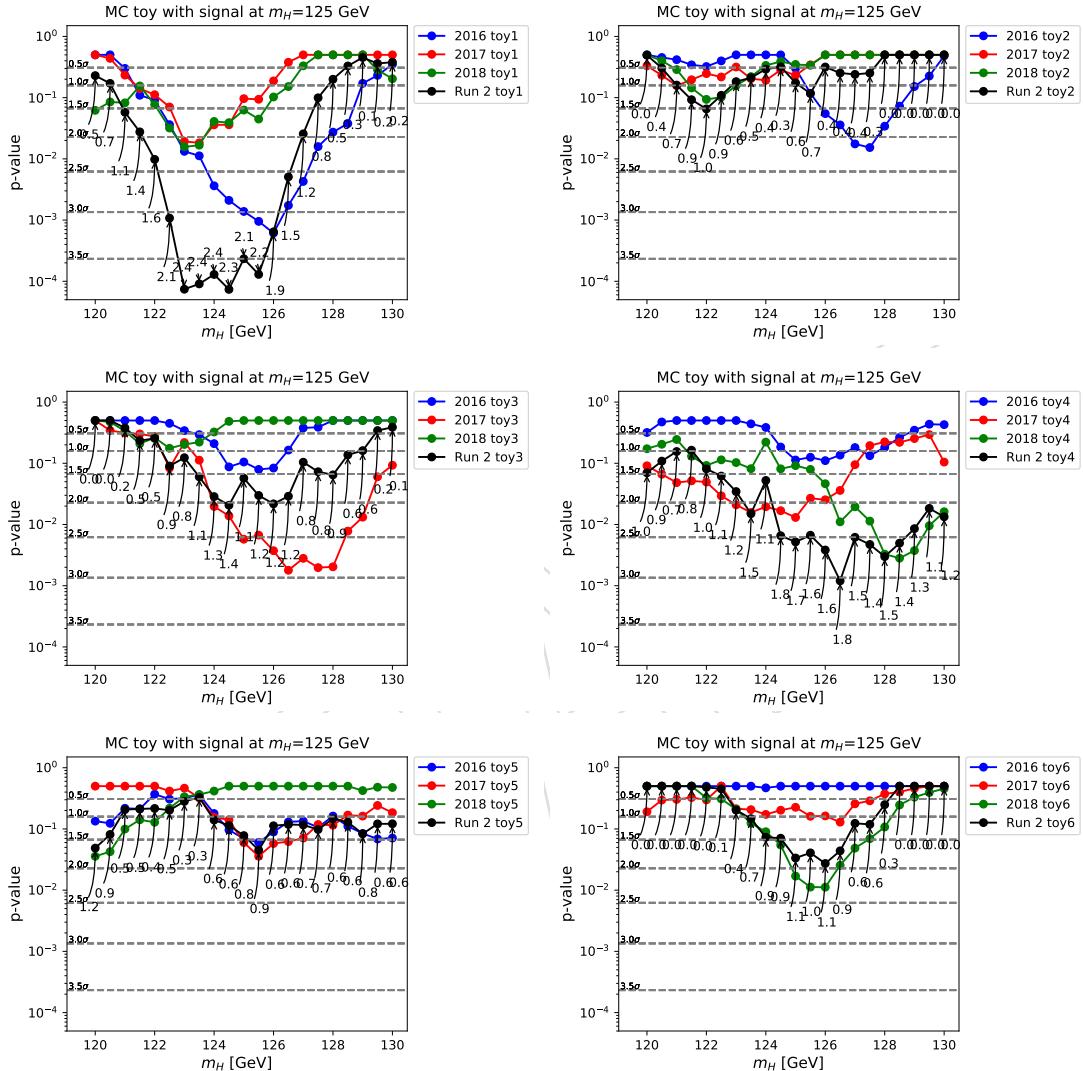


Figure 7.24: p-value in the mass scan injecting fullsim MC toy by including a SM Higgs boson signal at $m = 125 \text{ GeV}$ for the VBF analysis for 2016, 2017, 2018 and the combined Run 2. The best fit signal strength μ values for the toy under different Higgs mass hypotheses are written next to the points showing the p-values.

2258 8 ttH channel

2259 The production of the Higgs boson in association with a pair of top quarks has the smallest
 2260 cross section among the considered production modes. The main challenge is to separate signal
 2261 from reducible background events which are about three orders of magnitude larger than the
 2262 signal. However, distinct kinematic properties of ttH events help to reduce the SM background
 2263 to a level comparable to the expected signal yield. In the following, the pre-selection and cate-
 2264 gorization of events, the optimization using multivariate discriminants based on boosted deci-
 2265 sion trees (BDTs) as well as the final signal extraction are described. Fig. 8.1 shows the Feynman
 2266 diagram of the ttH process. The particles in the signal final state comprise a pair of oppositely
 2267 charged muons, as well as a pair of top quarks. Top quarks decay dominantly into a b quark
 2268 and W boson ($t \rightarrow bW$). The W boson then decays further via a leptonic decay ($W \rightarrow l\nu$) or an
 2269 hadronic decay ($W \rightarrow q\bar{q}'$). In each of these cases, the b quark hadronizes and is detected as a
 2270 jet. For this reason, ttH events are tagged by the presence of b-jets. This requirement reduces
 2271 background events and ensures mutual exclusivity between ttH events and all other channels
 2272 considered in this search.

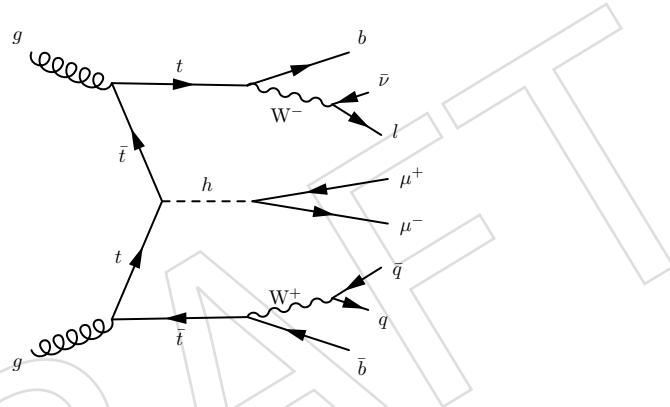


Figure 8.1: Leading order Feynman diagram of the ttH process. The signal process is shown where a Higgs boson decays to a pair of muons and the top quarks decay either semi-leptonically ($t \rightarrow bl\nu$) or hadronically ($t \rightarrow b\bar{q}\bar{q}'$).

2273 8.1 Event pre-selection and categorization

2274 Single muon triggers are used to select events, as described in Section 2.4. The trigger muon
 2275 is required to pass tight ID and tight isolation criteria defined by MUO POG. Events with at
 2276 least one medium or two loose b-tagged jets as well as a pair of oppositely charged muons
 2277 passing the ttH lepton identification requirements are selected. The dominant background at
 2278 this point are dileptonic $t\bar{t}$ events. The ratio of signal and background events can be improved
 2279 when events are categorized according to the decay products of the top quark. An overview of
 2280 the selection requirements and category definition for ttH events is given in Table 8.1.

2281 Events with exactly two muons are assigned to the hadronic category, aiming at events where
 2282 both top quarks decay hadronically and at least one of them can be reconstructed. Events with
 2283 three to four leptons are grouped into the leptonic category selecting events where at least one
 2284 top quark decays leptonically. The leptonMVA helps to suppress non-prompt or fake lepton
 2285 background. Charge requirements ensure lepton pairs to have opposite charge. An important
 2286 feature of the pre-selection is the dimuon pair finding. In case of more than two muons, at
 2287 least one opposite charged dimuon pair is required to have an invariant mass between 110
 2288 - 150 GeV. If two dimuon pairs fall into the 110 - 150 GeV mass window, the pair with the
 2289 higher dimuon transverse momentum is chosen as the Higgs candidate. Backgrounds from

2290 quarkonium decays and combinatorial background are reduced by a veto on events where the
 2291 invariant mass of any oppositely charged same flavour lepton pair is close to the mass of the Z
 2292 boson ($m_Z \pm 10$ GeV) or below a threshold of 12 GeV.

Figure 8.2: Dimuon mass of the ttH leptonic and ttH hadronic category after pre-selection with full Run2 data.

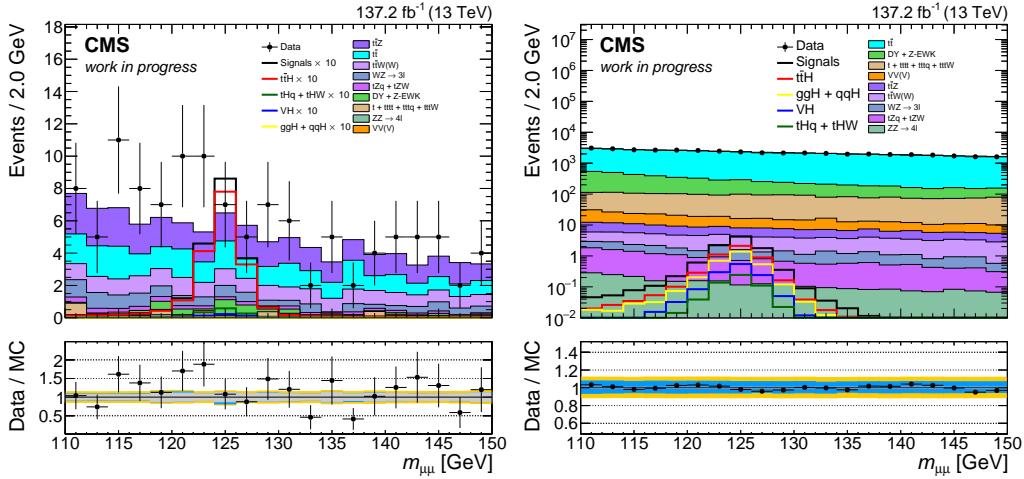


Table 8.1: Summary of the event selection criteria for the ttH hadronic and leptonic categories.

Event Selection	ttH hadronic	ttH leptonic	
		semileptonic	dileptonic
B-tag Requirement	$N_{\text{medium}} > 0$	$N_{\text{loose}} > 1$	
Number of Leptons	2	3	4
Charge Requirements	$\sum q_i = 0$	$\sum q_i = \pm 1$	$\sum q_i = 0$
Jet Multiplicity	≥ 3	≥ 2	
Leading Jet p_T	≥ 50 GeV	≥ 35 GeV	
Jet-triplet Convergence	$100 \text{ GeV} < m_{jjj} < 300 \text{ GeV}$	-	
Z Mass Veto	-	$ m_{\ell\ell} - m_Z < 10 \text{ GeV}$	
Low Mass Resonance Veto	-	$m_{\ell\ell} < 12 \text{ GeV}$	

Table 8.2: Expected and observed event yields in the sideband and signal region for 137.2 fb^{-1} of data in the ttH leptonic and ttH hadronic category.

Process	ttH leptonic	ttH hadronic
$t\bar{t}Z$	32.52	74.85
$t\bar{t}$	26.23	38875.10
$t\bar{t}W(W)$	18.00	53.87
$WZ \rightarrow 3l$	11.01	14.71
Top (t + ttq + ttW)	3.20	1467.78
tZq/tZW	3.81	13.50
DY	3.47	3670.42
$ZZ \rightarrow 4\ell$	1.83	2.55
VV(V)	0.61	175.58
Total background	100.68	44348.40
observation	118	44377

8.2 Multivariate discriminants

After the event pre-selection and categorization the leading backgrounds in the ttH hadronic category are $t\bar{t}$ and DY+Jets events. In the ttH leptonic category $t\bar{t}$, ttZ, ttW(W) and WZ are the most important backgrounds. To improve the separation between signal and background events, multivariate discriminants based on BDTs are used. The strongest discrimination is provided by the dimuon mass. For this reason, the dimuon mass is used for the final signal extraction, and all input observables are chosen to be highly uncorrelated with the dimuon mass.

8.2.1 Input observables

The observables sensitive to the kinematics of ttH events can be divided into three groups. Firstly, observables describing the dimuon system, secondly, observables containing information on the hadronic activity in the event, and thirdly, observables which aim to reconstruct the specific kinematics of the top quark decay products. The corresponding distributions are presented in Fig. 8.3, 8.4 and 8.5, respectively.

The Higgs boson candidate is described by the dimuon transverse momentum $p_T^{\mu\mu}$, the rapidity of the dimuon system $y_{\mu\mu}$, the azimuthal angle ϕ_{CS} and the cosine of the polar angle $\cos\theta_{CS}$ in the Collins-Soper rest frame. Additionally, the η of the two muons from the Higgs boson candidate and their p_T relative to $m_{\mu\mu}$ are used. To take the per-event Higgs candidate mass resolution into account, signal events are weighted according to the dimuon mass resolution as described in 4.8.

Further H_T and H_T^{miss} , which are the vectorial(scalar) transverse momentum sum of all selected leptons, and all selected jets with $|\eta| < 2.5$ in the event, the missing transverse momentum E_T^{miss} , the number of jets N_{jets} and the single jet p_T and η of the leading jets are included. The projection of the missing transverse momentum onto the bisector of the dimuon system is considered as well, which is referred to the $\Delta\zeta$ observable.

Additionally to the aforementioned inputs, observables which in particular describe the kinematics of the top quark decay products are considered. In ttH leptonic events these are represented by the $\Delta\phi$ between the reconstructed direction of the Higgs boson and the additional lepton, the invariant mass of the jet with the highest b-tag score and the additional lepton $m_{\ell b}$ and the transverse mass between the additional lepton and the missing transverse momentum $m_T^{\ell, \text{MET}}$. To distinguish the $\mu\mu\mu$ from the $\mu\mu e$ final state also the number of electrons is an input to the BDT. In ttH hadronic events the hadronic decay of a top quark can potentially be reconstructed from the combination of three jets. An algorithm dedicated to this reconstruction has been developed. The resolved hadronic top tagger RHTT is described in appendix G. For each combination of a jet-triplet the compatibility with the hypothesis of a hadronic top quark decay is tested. From the best top quark candidate, which is defined by the jet-triplet with the highest RHTT score, the following observables are constructed: The transverse top quark momentum of the top quark, the p_T balance between the top quark and the Higgs candidate. Finally, also the RHTT BDT score of the best top candidate is included.

The full list of observables is reported in Table 8.3.

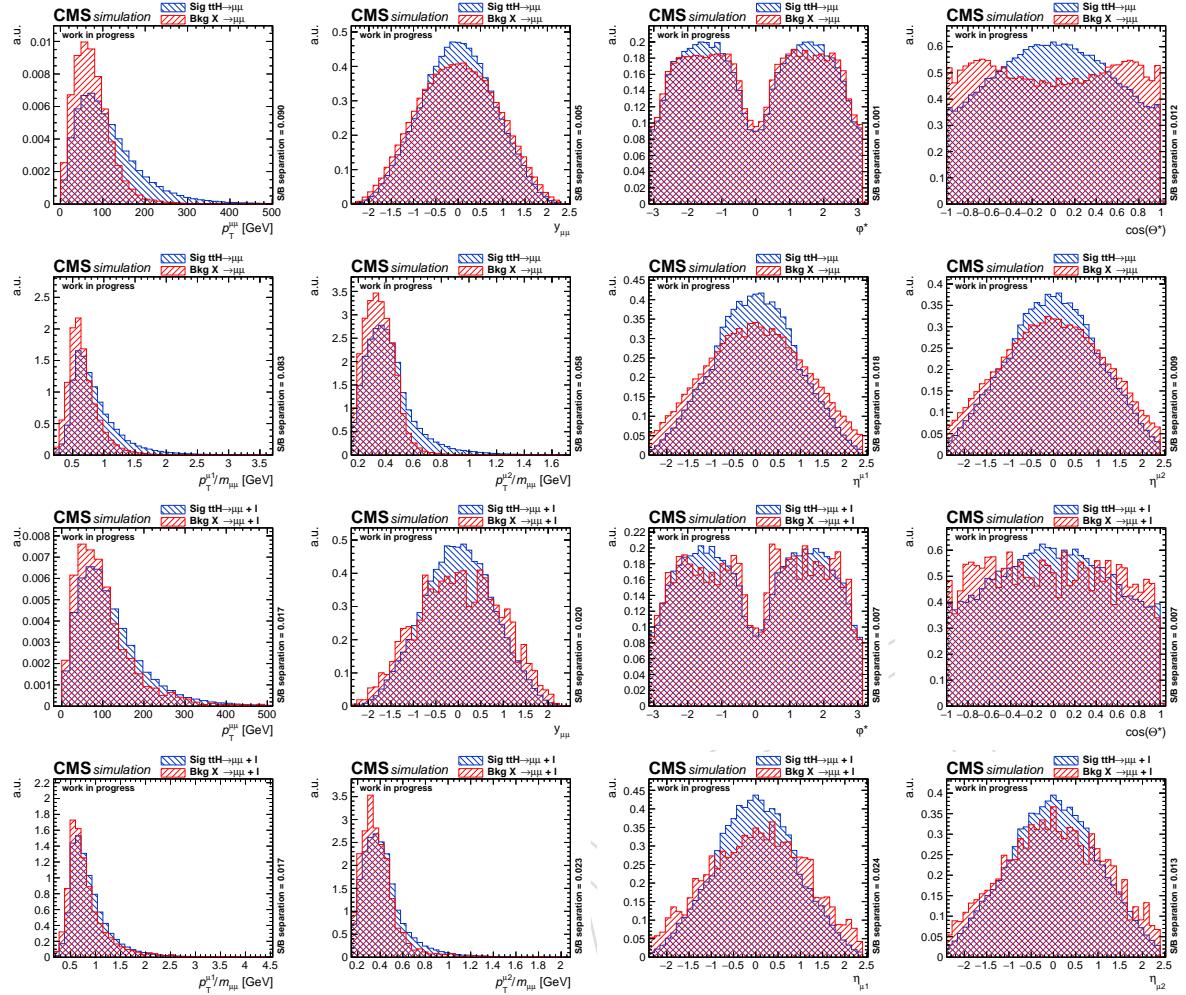


Figure 8.3: Observables describing the kinematics of the dimuon system are shown for signal (blue) and background (red) events for events in the ttH hadronic (two top rows) and ttH leptonic (two bottom rows) categories.

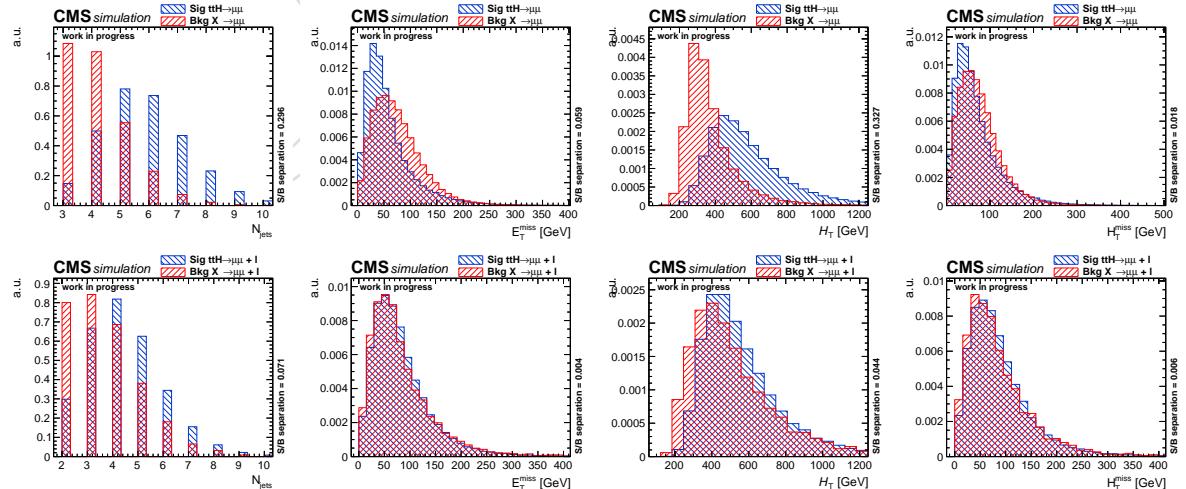


Figure 8.4: Observables describing the hadronic activity in the event for the ttH hadronic (top row) and ttH leptonic (bottom row) categories.

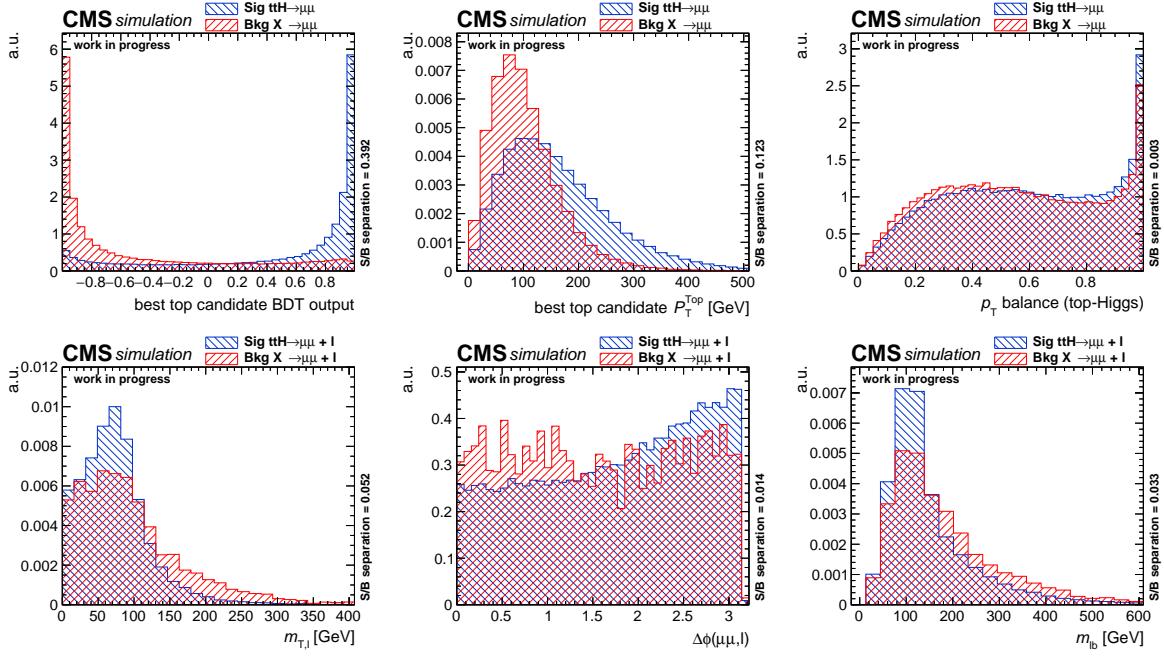


Figure 8.5: Observables targeting the distinct kinematic properties of the top quark decay products for the ttH hadronic (top row) and ttH leptonic (bottom row) categories.

Table 8.3: Input observables used in BDT training. A set of common observables is used in both the ttH hadronic and the ttH leptonic BDT. Three observables aiming at the kinematics of the leptonic decay of the top quark, and four observables characterizing the kinematics of the hadronic top quark decay are added, respectively. This is indicated by the (✓).

Observables	Formula	ttH leptonic	ttH hadronic
Transverse momentum of Higgs boson candidate	$p_T^{\mu\mu}$	✓	✓
Rapidity of Higgs boson candidate	$y_{\mu\mu}$	✓	✓
Azimuthal angle in CS-frame	$\Delta\Phi_{CS}$	✓	✓
Polar angle in CS-frame	$\cos\theta_{CS}$	✓	✓
Single muon transverse momentum	$p_T^\mu/m_{\mu\mu}$	✓	✓
Single muon pseudorapidity	η_μ	✓	✓
Missing transverse momentum	E_T^{miss}	✓	✓
Momentum sum (vectorial)	H_T^{miss}	✓	✓
Momentum sum (scalar)	H_T	✓	✓
Number of jets	N_{jets}	✓	✓
E_T^{miss} projection onto dimuon bisector	$\Delta\zeta$	✓	✓
$\Delta\Phi$ Higgs candidate and add. leading lepton	$\Delta\Phi(\mu\mu, \ell)$	✓	-
Mass of b-tagged jet and add. leading lepton	$m_{\ell b}$	✓	-
Transverse mass of additional lepton and E_T^{miss}	$m_T^{\ell, \text{MET}}$	✓	-
Max. RHTT output	max. RHTT	-	✓
Transverse momentum of top candidate	p_T^{top}	-	✓
Balance of Higgs and Top candidate momentum	$p_T^{H,\text{top}}$	-	✓
Single jet transverse momentum of leading jets	p_T^{jet}	✓	✓
Single jet η of leading jets	η^{jet}	✓	✓

2333 **8.2.2 Training**

2334 To ensure a stable training, and to increase the robustness of the BDT, the training is performed
 2335 on a combination of MC samples of all eras as the available MC statistics of the considered
 2336 background samples become small in the selected phase space. As explained in Section 6.1.2,
 2337 the training is performed on a tight mass window $115 \text{ GeV} < m_{\mu\mu} < 135 \text{ GeV}$. Since kinematics
 2338 of ttH events are very different compared to the other Higgs production modes, only ttH with
 2339 $H \rightarrow \mu\mu$ at $m_H = 125 \text{ GeV}$ is considered in the training. For the background all contributions
 2340 are taken into account. The signal sample is divided into events with odd and even event num-
 2341 bers, so that only events with even event numbers are used for training. Due to much smaller
 2342 MC statistics in the leptonic category compared to the hadronic category, the pre-selection re-
 2343 quirements are relaxed by removing the Z mass veto and broadening the mass window to
 2344 $110 \text{ GeV} < m_{\mu\mu} < 140 \text{ GeV}$. The BDT settings used for the ttH leptonic and ttH hadronic cate-
 2345 gory are reported in Table 8.4. Figures 8.6,8.7 show the performance of the training via a ROC
 2346 curve. Both, the training in the hadronic and leptonic show strong separation power against
 2347 backgrounds. A Kolmogorov Smirnov test is performed to prevent over-training. The BDT
 2348 distributions between test and the trainings samples agree well within uncertainties, from this
 2349 it can be concluded that the BDTs are not over-trained. Figures 8.8,8.9 show that the training is
 2350 unbiased with respect to the dimuon mass for signal events with $m_H = 120, 125, 130 \text{ GeV}$ and
 2351 for background events in five quantiles of the BDT output distribution.

Table 8.4: BDT settings of the ttH leptonic and ttH hadronic BDTs.

Category	Trees	Shrinkage	Bagged fraction	Cuts	Depth	Separation
ttH leptonic	200	0.1	0.5	40	3	CrossEntropy
ttH hadronic	600	0.1	0.5	100	3	CrossEntropy

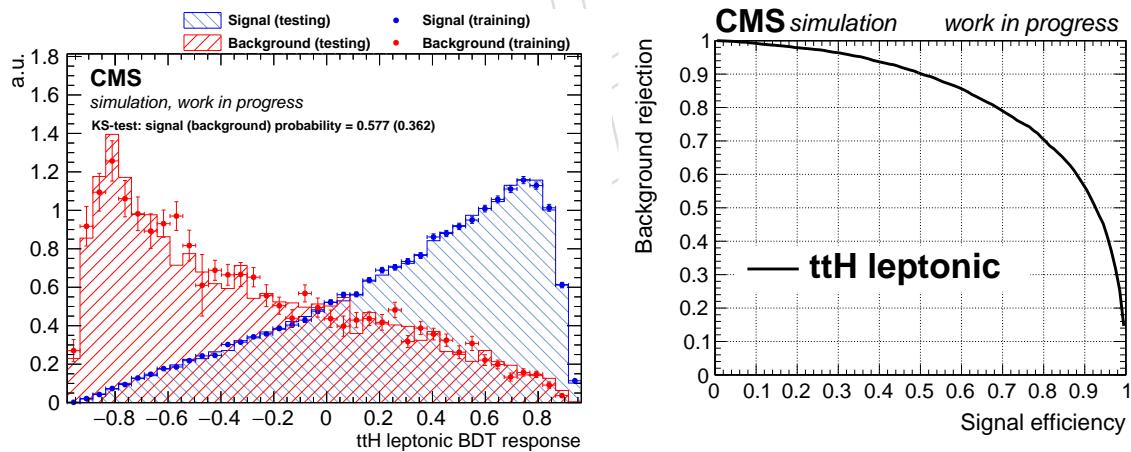


Figure 8.6: Performance and over-training plots for the ttH leptonic category

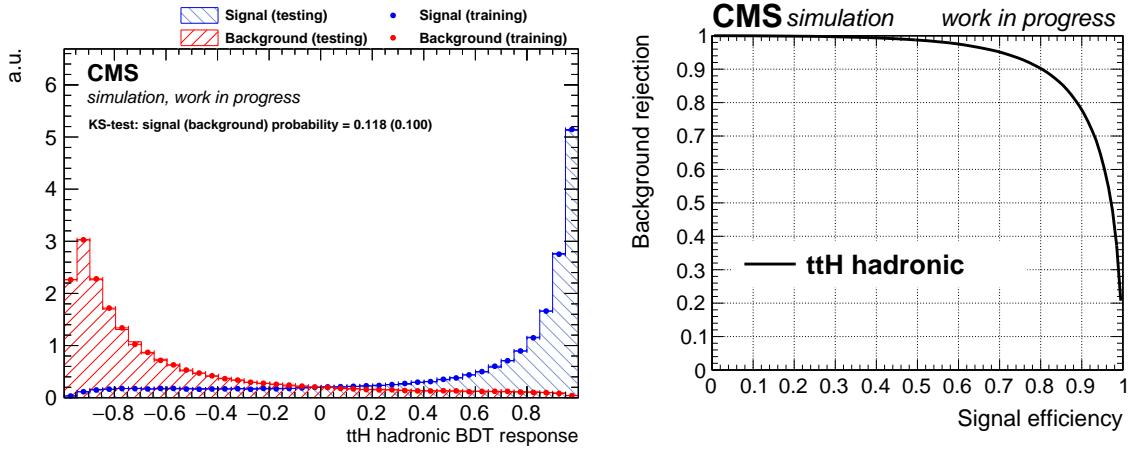


Figure 8.7: Performance and over-training plots for the ttH hadronic category

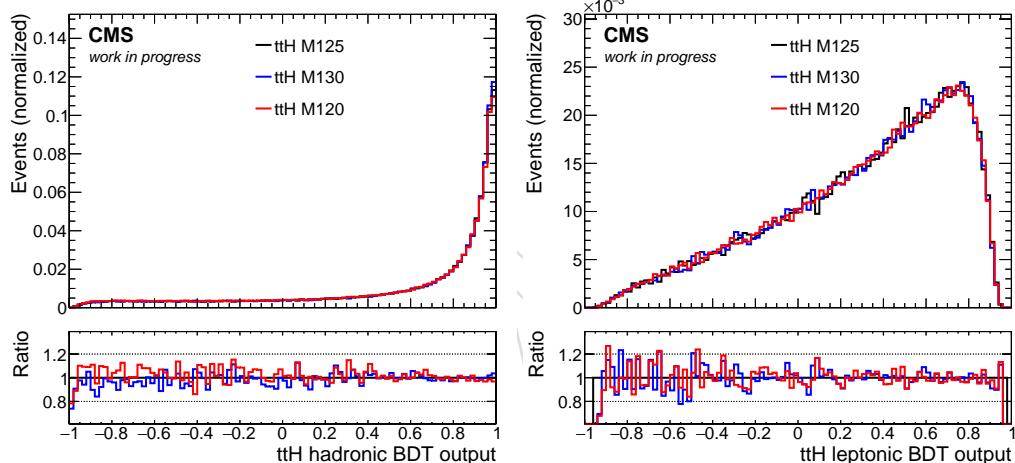


Figure 8.8: BDT output for three signal mass points of the ttH signal sample. The distributions are normalized to have unit area.

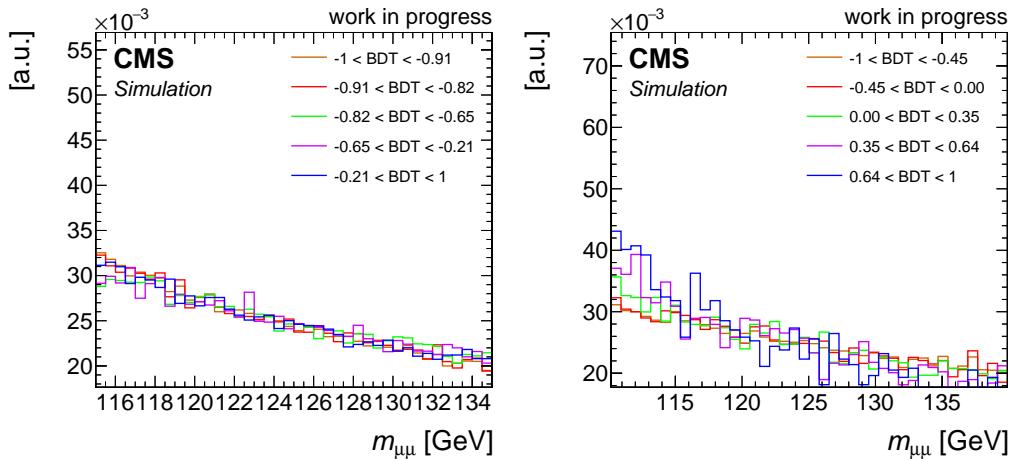


Figure 8.9: Distribution of the dimuon mass for five quantiles of the BDT output. No mass sculpting visible.

2352 8.2.3 BDT validation

2353 The validation of the BDT is performed on events with a dimuon mass of $110 \text{ GeV} < m_{\mu\mu} <$
 2354 150 GeV . The final BDT discriminates for the ttH hadronic and leptonic categories are shown
 2355 in Figure 8.10. The BDT output of the ttH hadronic category shows a small trend in the data-
 2356 MC agreement. The discrepancy increases towards the high BDT score, as events with high jet
 2357 multiplicities are preferred and events with three or more jets are strongly underestimated in
 2358 MC prediction. Most of the MC backgrounds rely on jets from PS, the modelling of these jets
 2359 is expected to be less accurate compared to jets from ME. For this reason, the signal modelling
 2360 is not corrected and the background is estimated in a fully data-driven way via a sideband fit
 2361 to the dimuon mass distribution. Figures 8.11, 8.12, and 8.13 present a set of input observables
 2362 used in the ttH hadronic BDT. The BDT output of the ttH leptonic category shows a good
 2363 agreement of the shape, although the normalization is underestimated by 10%(18%) in the
 2364 SB(SB+SR). Figure 8.14, 8.15, and 8.16 show a set of corresponding input observables of the ttH
 2365 leptonic BDT.

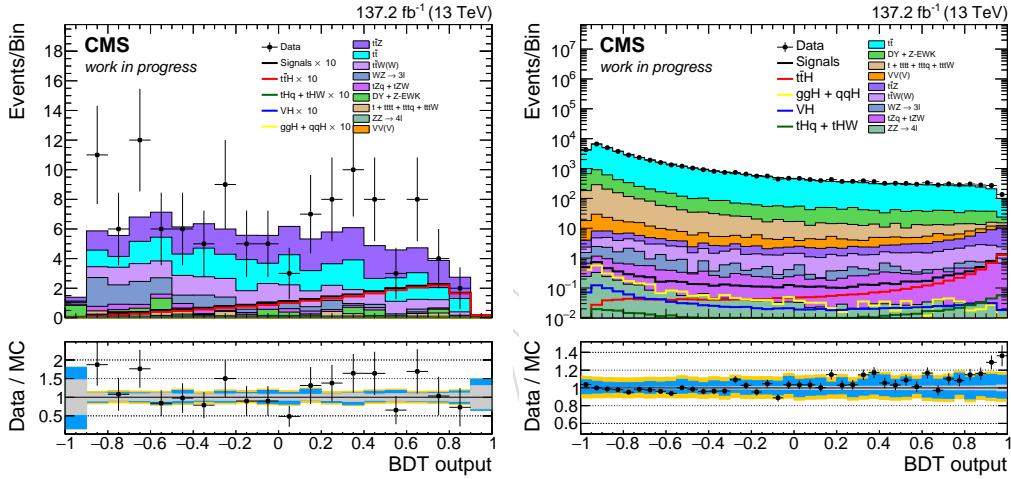


Figure 8.10: Final discriminants of the ttH leptonic and hadronic categories with full Run2 dataset for events in the dimuon sideband. The stat + syst uncertainties are indicated by the blue filled area in the ratio panel, the gray area is the stat-only component of the uncertainty.

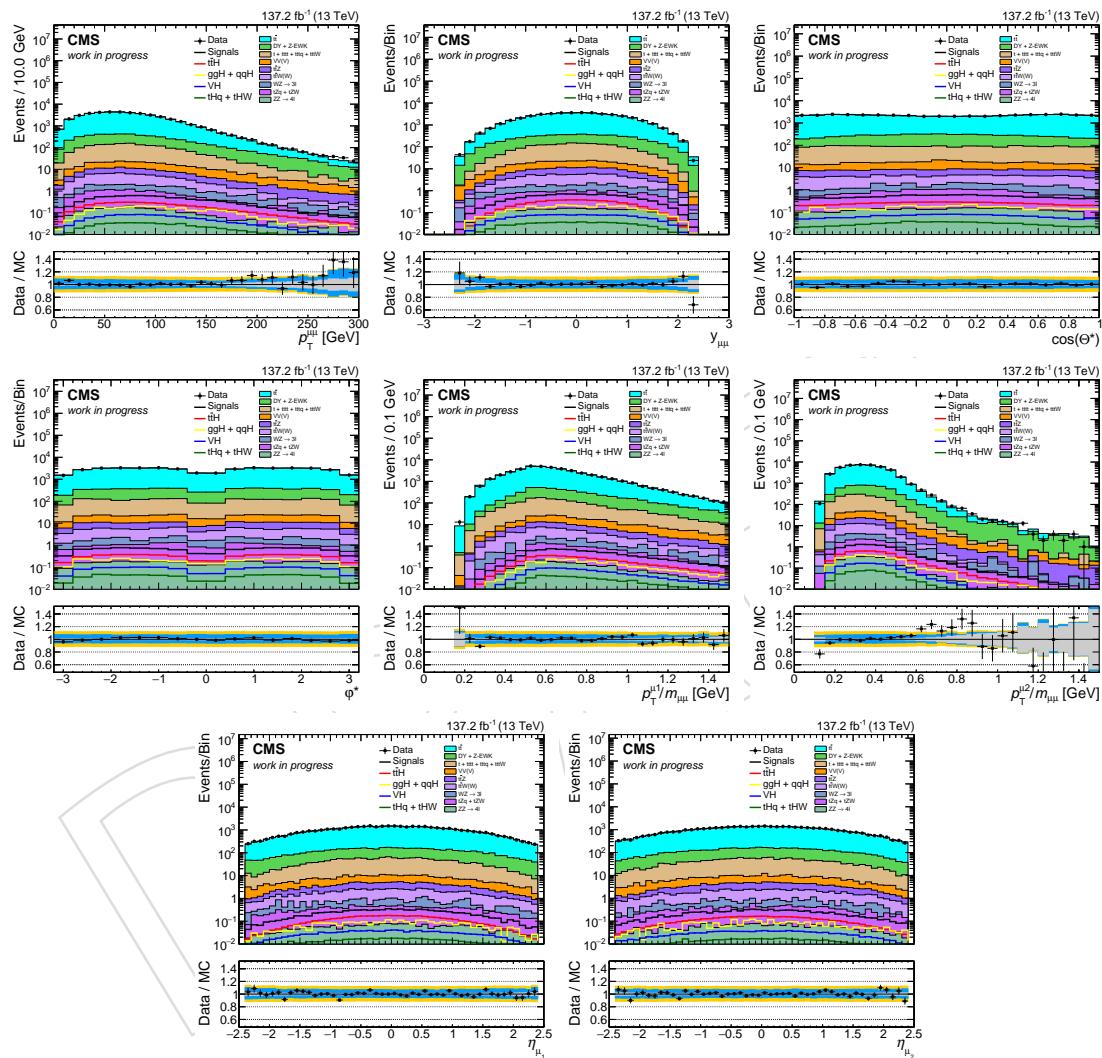


Figure 8.11: Set of input observables related to the muon kinematics of the ttH hadronic BDT with the full Run2 dataset.

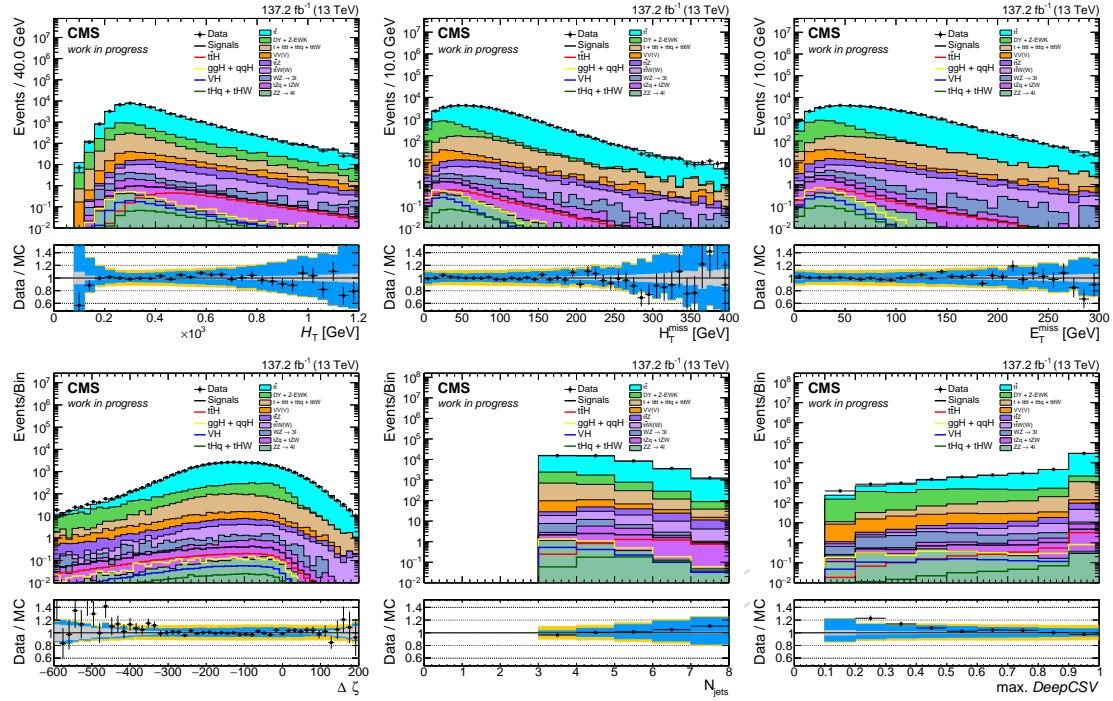


Figure 8.12: Set of input observables related to the hadronic activity of the $t\bar{t}H$ hadronic BDT with the full Run2 dataset.

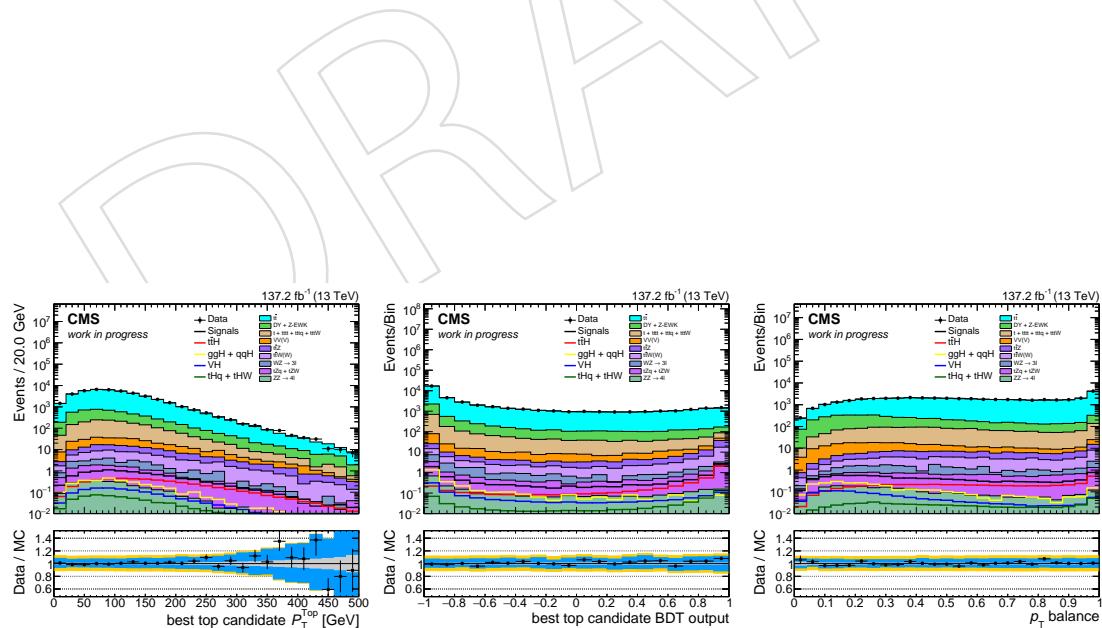


Figure 8.13: Set of input observables related to the kinematics of the reconstructed top quark candidate with the full Run2 dataset.

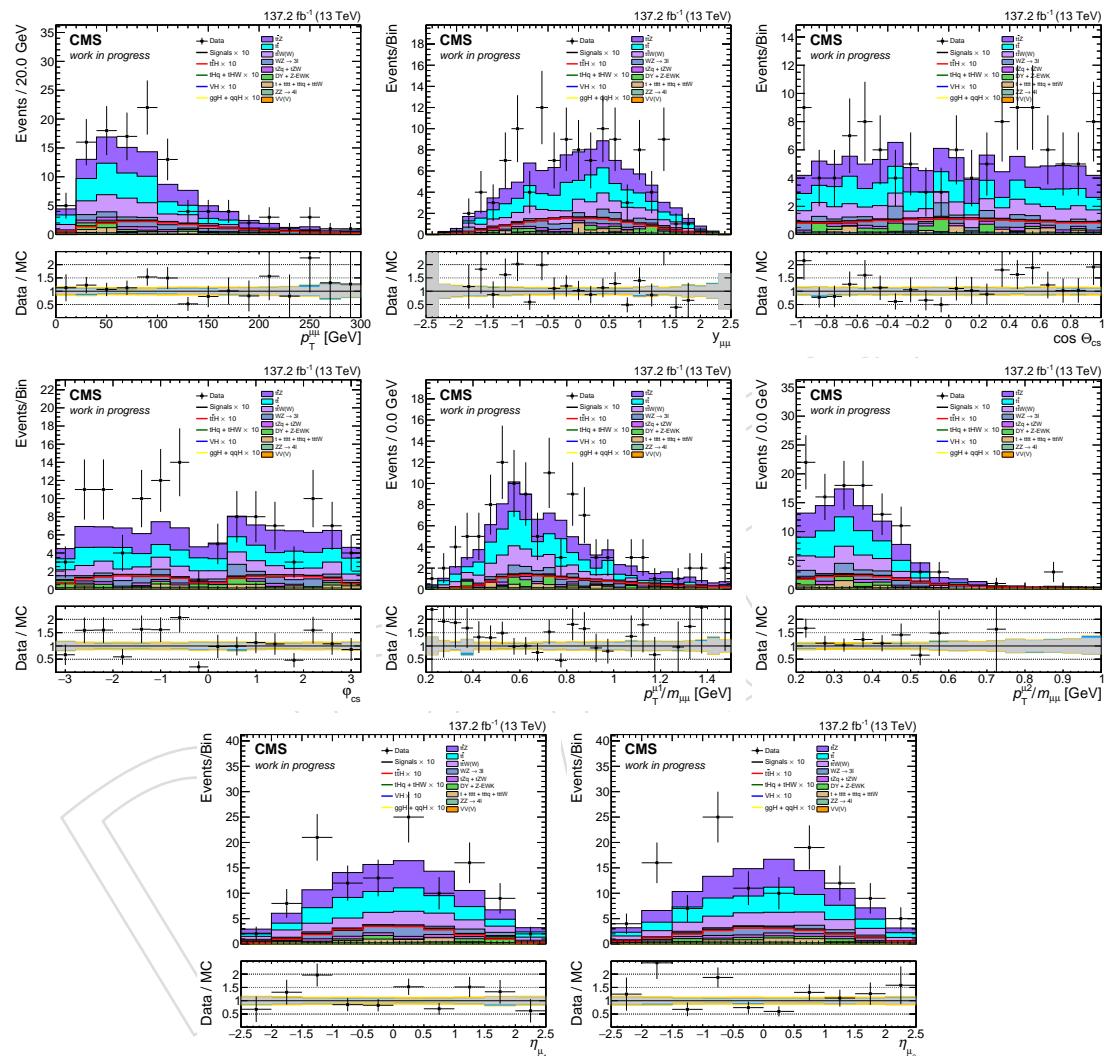


Figure 8.14: Set of input observables related to the muon kinematics of the ttH leptonic BDT with the full Run2 dataset.

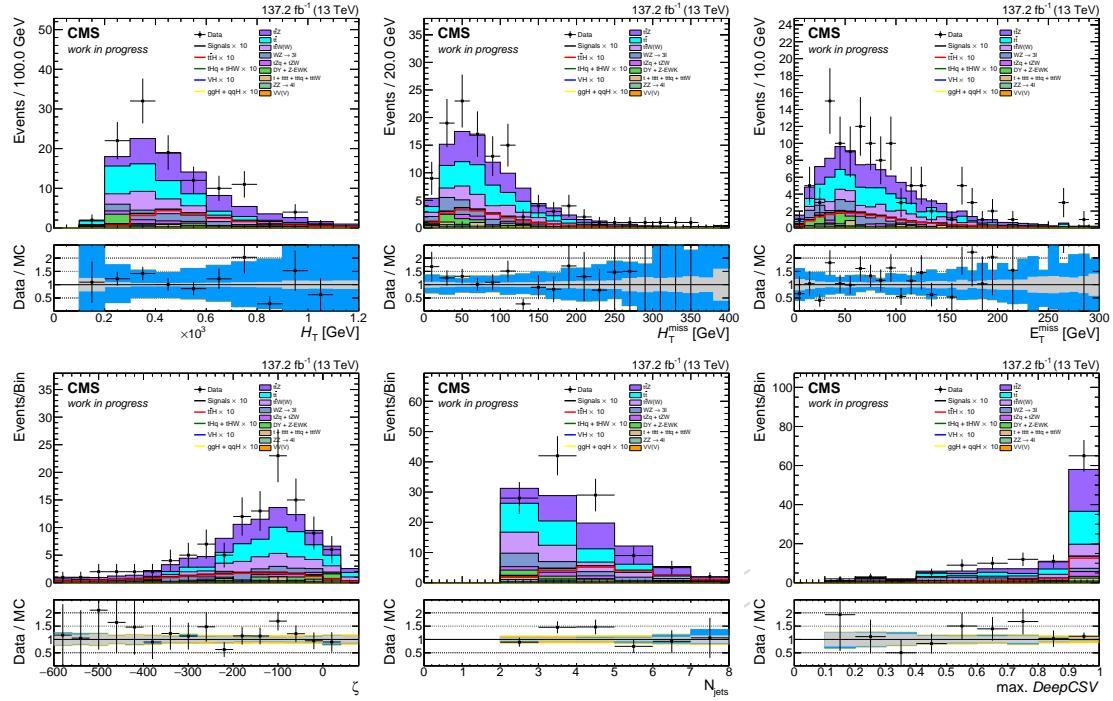


Figure 8.15: Set of input observables related to the leptonic activity of the $t\bar{t}H$ leptonic BDT with the full Run2 dataset.

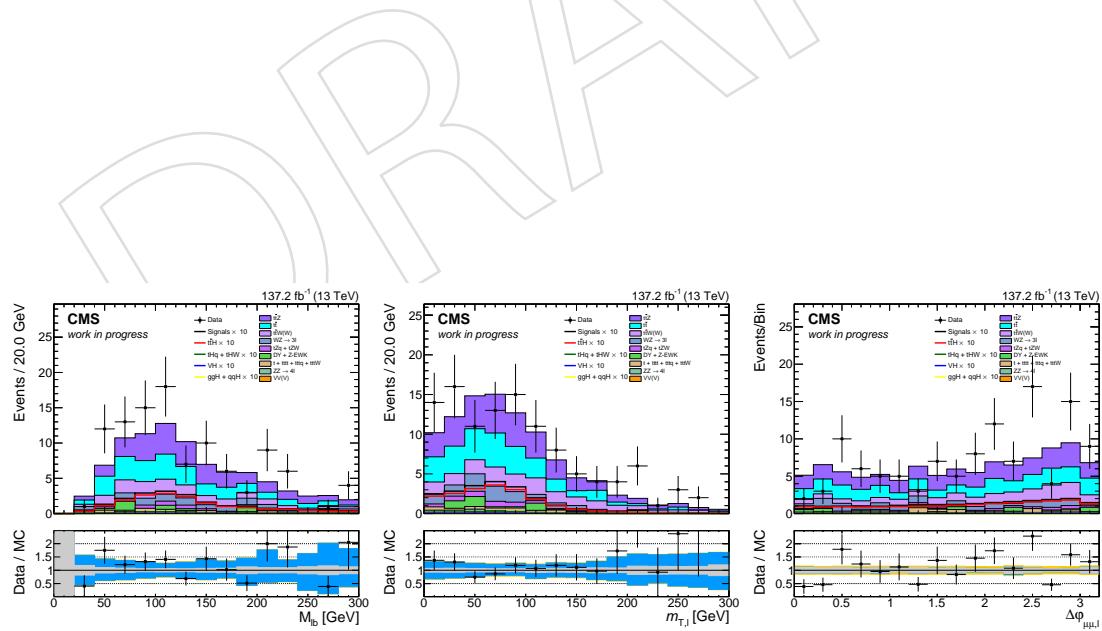


Figure 8.16: Set of input observables related to the kinematics of the top quark decay products with the full Run2 dataset.

2366 **8.2.4 BDT categorization**

2367 The obtained response of the BDTs are now used to classify events according to their BDT
 2368 score. The optimization is based on the full limit setting strategy without the consideration
 2369 of systematic uncertainties. Fig 8.17 shows the significance of the categories when splitting
 2370 into several subcategories. For this purpose a simple exponential function is used to model the
 2371 background. Since in the ttH leptonic category the reducible background is already strongly
 2372 reduced by the LeptonMVA, the gain from splitting into BDT subcategories is 15%, while in
 2373 the ttH hadronic category the dominant backgrounds are still reducible dileptonic $t\bar{t}$ events,
 2374 and the impact of a BDT splitting is much stronger with a relative improvement of 50%. The
 2375 iterative splitting algorithm results in two ttH leptonic subcategories, and three ttH hadronic
 2376 categories. The boundaries are summarized in Tab. 8.5.

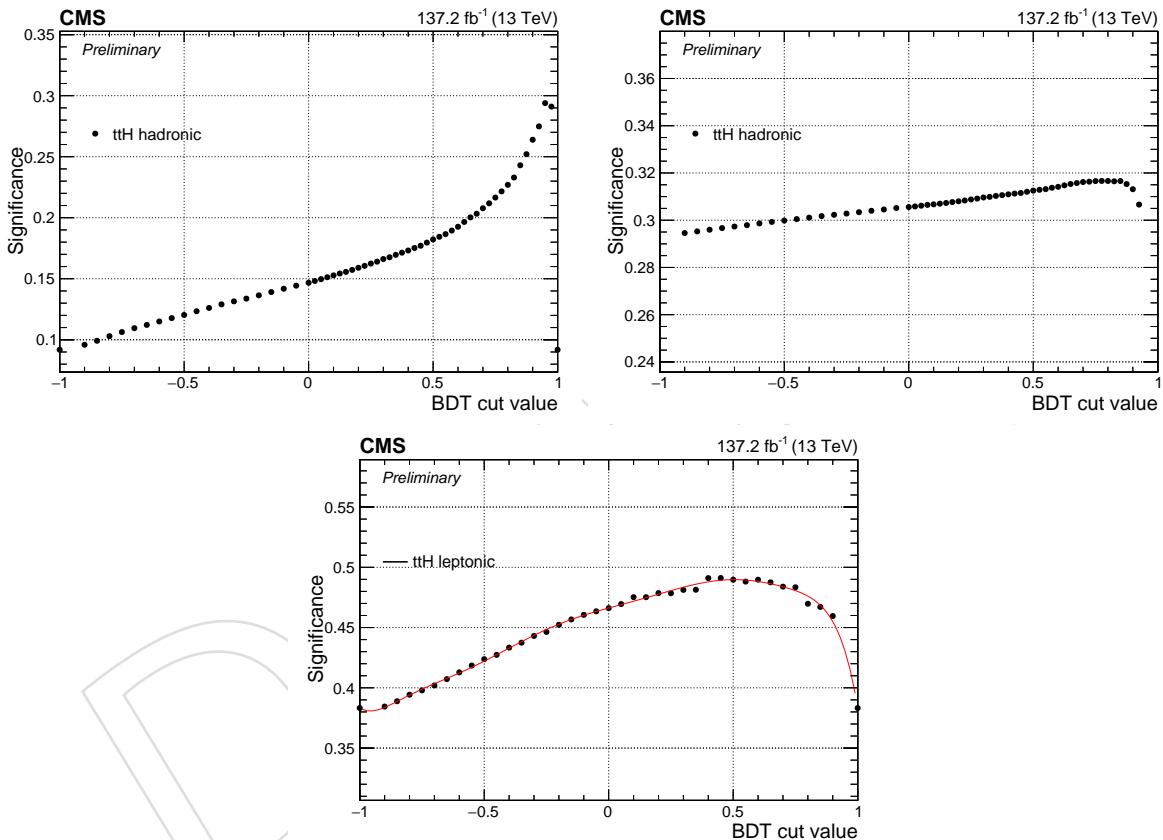


Figure 8.17: The significance after BDT splitting into subcategories. All three optimization steps for the ttH hadronic (two top plots) and ttH leptonic (bottom) categories are shown.

Table 8.5: Optimized category boundaries for the ttH categories

Name	Discriminant Range
ttH leptonic cat1	[0.4, 1.0]
ttH leptonic cat0	[-1.0, 0.4]
ttH hadronic cat2	[0.95, 1.0]
ttH hadronic cat1	[0.8, 0.95]
ttH hadronic cat0	[0.8, -1.0]

2377 8.3 Signal extraction

2378 As described in Sec. 8.3 the signal is extracted by fitting the dimuon invariant mass spectrum
 2379 in data. This is done via analytical functions able to model the expected signal and background
 2380 processes.

2381 8.3.1 Signal model

2382 The signal line shape is modelled using a double-sided Crystal Ball (DCB) function defined in
 2383 Equation 4.2. In each of the ttH categories the signal contribution from each production modes
 2384 is fitted independently and the corresponding shape parameters are fixed. The lineshapes used
 2385 in the final fit for the ttH signal are reported in Fig. 8.18 and 8.19 for the three hadronic and the
 2386 two leptonic categories, respectively.

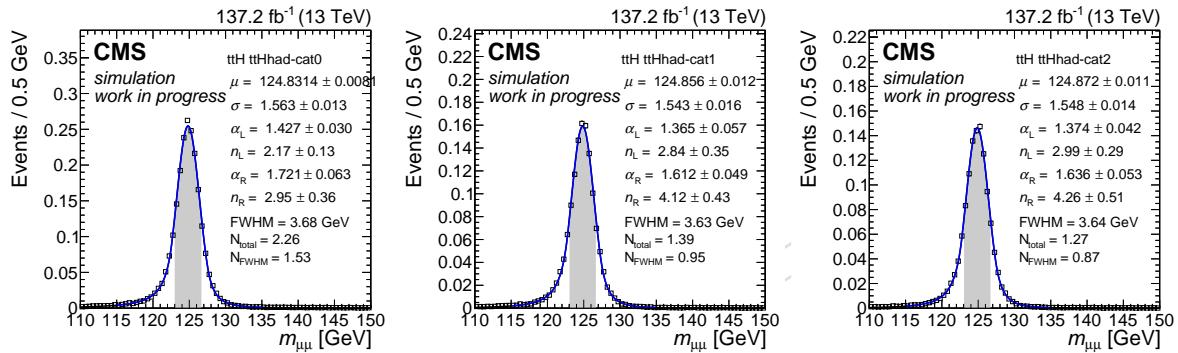


Figure 8.18: Distributions of the dimuon invariant mass for ttH signal events in the ttH hadronic BDT categories. From left to right, ttH hadronic cat0 (left), ttH hadronic cat1 (middle), and ttH hadronic cat2 (right). The expectations in the three eras (2016, 2017, and 2018) have been summed together.

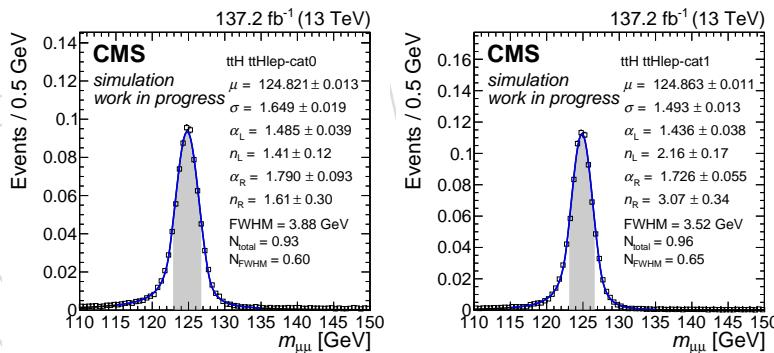


Figure 8.19: Distributions of the dimuon invariant mass for ttH signal events in the ttH leptonic BDT categories: ttH leptonic cat0 (left), and ttH leptonic cat1 (right). The expectations in the three eras (2016, 2017, and 2018) have been summed together.

2387 8.3.2 Uncertainties on the signal model

2388 The *shape* uncertainties affecting the modelling of the signal peak are considered as described in
 2389 Sec. 6.3.2. The *rate* uncertainties take changes of the signal acceptance into account as described
 2390 in Sec. 6.3.2 and Sec. 6.3.2. In the following, the corresponding theoretical rate uncertainties of
 2391 the ttH categories are summarized in Tab. 8.6, 8.7, and the rate uncertainties from experimental
 2392 sources are given in Tab 8.8, 8.9, 8.10, 8.11, and 8.12. The uncertainties are given for the most
 2393 important signal processes in terms of their sensitivity.

Table 8.6: Acceptance uncertainties for ttH, WH, ZH, VBF and ggH signal events in ttH BDT categories from variations of the parton shower tune parameters. Uncertainties smaller than 0.1% are not reported.

Category	ISR					FSR				
	WH	ZH	ttH	tHq	tHW	WH	ZH	ttH	tHq	tHW
ttH lep-cat1	0.959/1.054	0.971/1.039	0.998/1.003	0.998/1.003	0.998/1.003	0.924/1.172	1.059/0.939	1.003/0.998	1.003/0.998	1.003/0.998
ttH lep-cat0	0.988/1.014	0.996/1.005	1.002/0.997	1.002/0.997	1.002/0.997	0.967/1.038	0.998/1.012	1.000/0.999	1.000/0.999	1.000/0.999
ttH had-cat2	0.961/1.055	0.968/1.041	0.999/1.001	0.999/1.001	0.999/1.001	0.952/1.057	1.018/0.982	1.015/0.975	1.015/0.975	1.015/0.975
ttH had-cat1	0.977/1.032	0.971/1.039	1.000/1.000	1.000/1.000	1.000/1.000	0.990/1.019	1.020/0.952	1.002/0.993	1.002/0.993	1.002/0.993
ttH had-cat0	0.993/1.010	0.995/1.007	1.001/0.999	1.001/0.999	1.001/0.999	0.996/1.003	1.007/0.986	0.998/1.003	0.998/1.003	0.998/1.003

Table 8.7: Uncertainties for the ttH process in both the ttH BDT categories obtained by propagating the effect of variations in the renormalization and factorization scale, as well as the PDF distributions to the expected signal yields. Uncertainties smaller than 0.1% are not reported.

Source of Uncertainty	ttH Categories				
	ttH lep-cat1	ttH lep-cat0	ttH had-cat2	ttH had-cat1	ttH had-cat0
Scale QCD	0.998/1.006	0.999/1.004	0.994/1.011	1.000/1.001	0.995/1.005
PDF	1.000/1.000	1.000/1.000	1.000/1.000	1.000/1.000	1.000/1.000

Table 8.8: Acceptance uncertainties for WH, ZH, ttH, tHq and tHW signal events in ttH BDT categories related to muon identification and isolation scale factors. Uncertainties smaller than 0.1% are not reported.

Category	Muon ID					Muon Iso				
	WH	ZH	ttH	tHq	tHW	WH	ZH	ttH	tHq	tHW
ttH lep-cat1	0.995/1.005	0.995/1.005	0.995/1.005	0.995/1.005	0.995/1.005	0.999/1.001	0.999/1.001	0.998/1.002	0.998/1.002	0.999/1.001
ttH lep-cat0	0.994/1.006	0.995/1.005	0.994/1.006	0.994/1.006	0.994/1.006	0.998/1.002	0.998/1.002	0.998/1.002	0.998/1.002	0.998/1.002
ttH had-cat2	0.996/1.004	0.996/1.004	0.996/1.004	0.996/1.004	0.996/1.004	0.999/1.001	0.999/1.001	0.999/1.001	0.999/1.001	0.999/1.001
ttH had-cat1	0.996/1.004	0.996/1.004	0.996/1.004	0.996/1.004	0.996/1.004	0.999/1.001	0.999/1.001	0.999/1.001	0.999/1.001	0.999/1.001
ttH had-cat0	0.995/1.005	0.995/1.005	0.995/1.005	0.995/1.005	0.996/1.004	0.999/1.001	0.998/1.002	0.998/1.002	0.998/1.002	0.999/1.001

Table 8.9: Acceptance uncertainties for WH, ZH, ttH, tHq and tHW signal events in ttH BDT categories related to LeptonMVA scale factors for muons and electrons. Uncertainties smaller than 0.1% are not reported.

Category	Muon LepMVA					Electron LepMVA				
	ttH	WH	ZH	VBF	ggH	ttH	WH	ZH	VBF	ggH
ttH lep-cat1	0.997/1.003	0.999/1.001	0.997/1.003	0.998/1.002	0.997/1.003	0.993/1.007	0.990/1.010	0.993/1.007	0.992/1.008	0.994/1.006
ttH lep-cat0	0.997/1.003	0.997/1.003	0.995/1.005	0.996/1.004	0.996/1.004	0.997/1.003	0.992/1.008	0.998/1.002	0.997/1.003	0.997/1.003
ttH had-cat2	0.997/1.003	0.997/1.003	0.997/1.003	0.996/1.004	0.997/1.003	—	—	—	—	—
ttH had-cat1	0.998/1.002	0.997/1.003	0.997/1.003	0.998/1.002	0.997/1.003	—	—	—	—	—
ttH had-cat0	0.998/1.002	0.998/1.002	0.997/1.003	0.998/1.002	0.997/1.003	—	—	—	—	—

Table 8.10: Uncertainties for WH, ZH, ttH, tHq and tHW processes in ttH BDT categories obtained from the uncertainties related to muon energy scale. Uncertainties smaller than 0.1% are not reported.

Category	Muon scale uncertainty				
	WH	ZH	ttH	tHq	tHW
ttH lep-cat1	0.998/1.000	1.000/1.000	1.000/1.000	1.000/1.001	0.997/1.000
ttH lep-cat0	0.999/1.000	1.000/1.002	1.000/1.000	1.000/1.000	1.000/0.997
ttH had-cat2	1.000/1.001	0.997/1.000	0.999/1.000	0.999/1.000	0.999/1.001
ttH had-cat1	1.000/1.000	0.999/1.001	1.000/1.000	1.000/1.002	0.999/1.000
ttH had-cat0	1.000/1.000	1.000/1.000	1.000/1.000	1.000/1.000	1.000/1.000

Table 8.11: Uncertainties for the ttH process in both the ttH BDT categories obtained by propagating the effect of each source of uncertainty related to jet energy scale and jet energy resolution to the expected signal yields. Uncertainties smaller than 0.1% are not reported.

Source of Uncertainty	ttH Categories				
	ttH lep-cat1	ttH lep-cat0	ttH had-cat2	ttH had-cat1	ttH had-cat0
JES/JER					
flavorQCD	0.989/1.011	1.004/0.996	0.978/1.025	0.989/1.008	1.010/0.991
relativeBal	0.995/1.006	1.001/0.999	0.991/1.010	0.998/1.003	1.001/0.997
absolute	0.990/1.011	1.004/0.996	0.982/1.019	0.993/1.008	1.006/0.993
BBEC1	0.997/1.004	1.001/0.999	0.993/1.008	0.997/1.004	1.002/0.997
EC2	0.999/1.001	1.000/1.000	0.999/1.001	0.999/1.001	1.000/1.000
HF	0.999/1.002	1.001/0.998	0.999/1.001	1.002/1.000	1.000/1.000
relativeSample_2016	0.992/1.005	1.005/0.999	0.987/1.011	0.996/1.006	1.004/0.996
absolute_2016	0.998/1.001	1.000/1.001	0.996/1.003	0.998/1.003	1.003/0.998
BBEC1_2016	0.999/1.001	1.000/1.000	0.998/1.003	1.000/1.001	1.000/0.999
EC2_2016	1.000/1.001	1.000/1.000	1.000/1.000	1.001/1.000	0.999/1.000
HF_2016	0.999/1.000	1.001/1.000	1.000/1.001	1.001/0.999	0.999/1.000
jerBarrel_2016	0.996	1.005	0.980	1.000	1.013
jerEC1_2016	1.001	0.999	1.000	1.001	0.999
jerEC2low_2016	1.005	0.996	1.003	0.999	1.000
jerEC2high_2016	0.999	1.001	0.999	1.000	1.001
jerHFlow_2016	1.001	0.999	1.002	0.998	1.000
jerHFhigh_2016	1.000	1.000	1.000	1.000	1.000
relativeSample_2017	0.996/1.006	1.002/0.997	0.995/1.001	0.998/1.006	1.002/0.998
absolute_2017	0.997/1.004	1.001/0.998	0.993/1.002	1.000/1.007	1.002/0.997
BBEC1_2017	0.999/1.001	1.000/1.000	0.998/1.001	1.000/1.003	1.000/0.999
EC2_2017	0.997/1.002	1.002/0.999	0.998/1.001	0.999/1.002	1.001/0.999
HF_2017	1.000/1.001	1.000/1.000	0.999/1.000	1.001/1.000	1.000/1.000
jerBarrel_2017	0.999	1.002	0.980	1.004	1.013
jerEC1_2017	1.002	0.999	0.999	1.000	1.002
jerEC2low_2017	1.002	1.000	0.997	0.998	1.004
jerEC2high_2017	0.999	1.001	0.996	0.999	1.002
jerHFlow_2017	1.002	0.998	1.000	1.004	0.999
jerHFhigh_2017	1.000	1.000	1.000	1.000	1.000
relativeSample_2018	0.987/1.014	1.003/0.998	0.975/1.025	0.990/1.006	1.007/0.994
absolute_2018	0.996/1.005	1.000/1.000	0.990/1.012	0.994/1.004	1.005/0.995
BBEC1_2018	1.000/1.001	0.999/1.000	0.999/1.003	0.999/1.000	1.000/0.999
EC2_2018	1.000/1.001	1.000/1.000	1.000/1.001	0.999/1.000	1.000/1.000
HF_2018	1.000/1.001	1.000/0.999	0.999/1.001	1.001/1.000	1.000/1.000
jerBarrel_2018	0.999	1.002	0.982	1.005	1.010
jerEC1_2018	0.999	1.001	1.001	1.000	1.001
jerEC2low_2018	1.003	0.997	1.003	0.999	1.000
jerEC2high_2018	1.000	1.000	0.999	0.998	1.001
jerHFlow_2018	1.001	0.999	1.001	1.000	0.999
jerHFhigh_2018	1.000	1.000	1.000	1.000	1.000

Table 8.12: Uncertainties for ttH, WH, ZH, VBF and ggH processes in ttH BDT categories obtained from the uncertainties related to pre-firing correction and pileup re-weight. Uncertainties smaller than 0.1% are not reported.

Uncertainty	Category	WH	ZH	ttH	tHq	tHW
pileup 2016	ttH lep-cat1	0.978/1.022	0.973/1.022	1.006/0.994	1.011/0.990	1.006/0.993
	ttH lep-cat0	0.991/1.010	0.982/1.018	1.002/0.998	1.002/0.997	1.007/0.993
	ttH had-cat2	0.995/1.003	0.971/1.031	1.004/0.997	0.939/1.051	1.005/0.995
	ttH had-cat1	0.979/1.020	1.000/1.000	1.002/0.999	1.011/0.991	0.996/1.003
	ttH had-cat0	0.983/1.018	0.994/1.006	0.997/1.003	1.001/0.999	0.993/1.007
pileup 2017	ttH lep-cat1	1.013/0.987	0.987/1.013	1.004/0.995	1.002/0.993	1.015/0.987
	ttH lep-cat0	0.980/1.018	0.992/1.004	1.001/0.998	1.006/0.993	1.004/0.996
	ttH had-cat2	0.990/1.018	0.981/1.024	1.000/0.999	1.000/1.002	1.000/0.999
	ttH had-cat1	0.977/1.031	0.982/1.003	1.000/1.000	1.004/0.997	0.997/1.004
	ttH had-cat0	0.980/1.012	0.992/1.011	0.997/1.004	1.001/0.999	0.996/1.005
pileup 2018	ttH lep-cat1	0.984/1.020	1.025/0.977	1.005/0.995	1.000/1.000	1.001/0.998
	ttH lep-cat0	0.968/1.028	0.962/1.044	1.002/0.997	1.004/0.996	1.007/0.993
	ttH had-cat2	0.997/1.006	0.981/1.029	0.999/1.002	1.003/0.997	0.997/1.002
	ttH had-cat1	0.974/1.028	0.979/1.026	0.996/1.004	0.999/1.002	0.998/1.002
	ttH had-cat0	0.988/1.005	0.985/1.016	0.995/1.005	0.997/1.003	0.993/1.007
prefire 2016	ttH lep-cat1	1.004/0.996	1.004/0.996	1.004/0.996	1.007/0.993	1.003/0.997
	ttH lep-cat0	1.005/0.995	1.005/0.995	1.004/0.996	1.007/0.993	1.004/0.996
	ttH had-cat2	1.003/0.997	1.004/0.996	1.004/0.996	1.006/0.994	1.003/0.997
	ttH had-cat1	1.004/0.996	1.004/0.996	1.004/0.996	1.008/0.992	1.004/0.996
	ttH had-cat0	1.004/0.996	1.003/0.997	1.005/0.995	1.007/0.993	1.004/0.996
prefire 2017	ttH lep-cat1	1.006/0.994	1.016/0.984	1.005/0.994	1.012/0.988	1.005/0.995
	ttH lep-cat0	1.006/0.994	1.004/0.995	1.007/0.993	1.011/0.989	1.007/0.993
	ttH had-cat2	1.004/0.996	1.006/0.994	1.006/0.994	1.012/0.988	1.006/0.994
	ttH had-cat1	1.007/0.992	1.006/0.994	1.007/0.993	1.013/0.987	1.007/0.993
	ttH had-cat0	1.005/0.994	1.005/0.995	1.007/0.992	1.012/0.988	1.006/0.994

2394 **8.4 Background model**

2395 The background is modelled via agnostic functions with one to three degrees of freedom, these
 2396 functions have been introduced in Sec. 5.3, and are presented for all five ttH categories in
 2397 Figure 8.20 and 8.21.

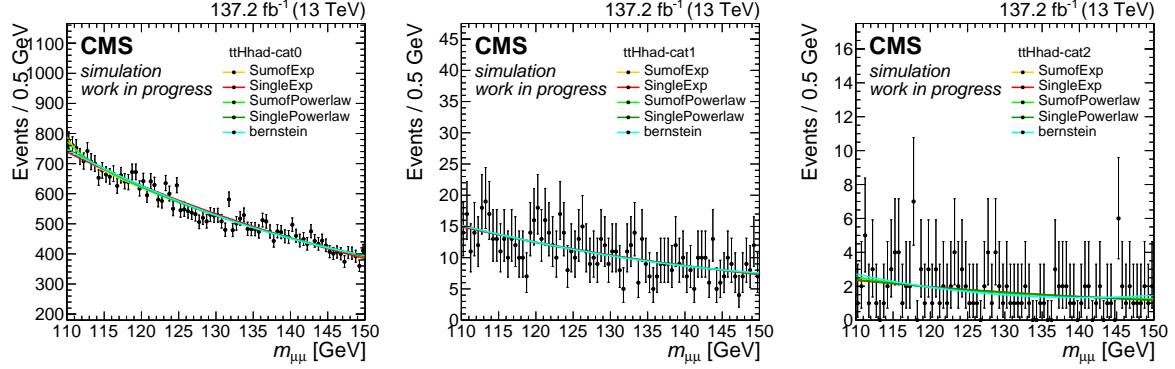


Figure 8.20: Distributions of the dimuon invariant mass for unblinded data events in the ttH hadronic BDT categories. From left to right, ttH hadronic cat0 (left), ttH hadronic cat1 (middle), and ttH hadronic cat2 (right). The expectations in the three eras (2016, 2017, and 2018) have been summed together.

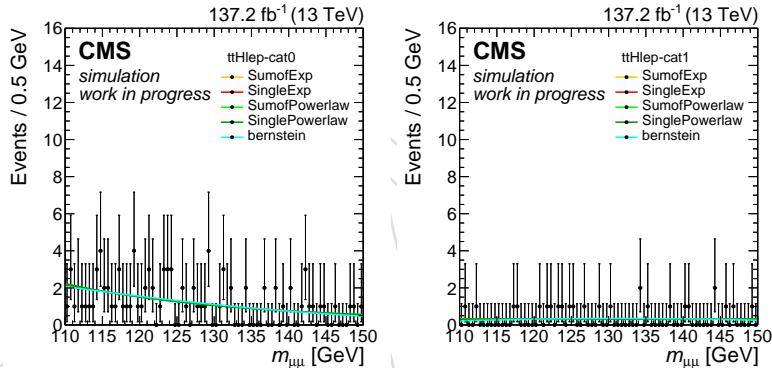


Figure 8.21: Distributions of the dimuon invariant mass for unblinded data events in the ttH leptonic BDT categories: ttH leptonic cat0 (left), and ttH leptonic cat1 (right). The expectations in the three eras (2016, 2017, and 2018) have been summed together.

2398 Since the underlying model of the background spectrum is not known, the choice of the back-
 2399 ground function may inject a bias in the extracted signal strength. This bias effects the result
 2400 of the signal+background fit and is examined in the following. The so-called bias study is per-
 2401 formed using toy experiments in each category considering each function as a possible truth
 2402 function. The background shape is extracted from a fit to data, while the signal shape as well as
 2403 the expected signal event yield is obtained from MC simulation. From this a signal plus back-
 2404 ground model is build and used to generate 2000 toy experiments, which are fitted by each
 2405 function. The bias is then defined as

$$\text{Bias}(\mu_{\text{fit}}, \mu_{\text{inj}}, \sigma_{\mu}) = \frac{\mu_{\text{fit}} - \mu_{\text{inj}}}{\sigma_{\mu}}$$

2406 where μ_{fit} is the signal strength extracted from the fit, μ_{inj} is the injected signal strength in each
 2407 toy and σ_{μ} is the uncertainty on the signal strength. A bias of less than 20% corresponds to
 2408 an impact on the final limit of around 2%, and can be neglected. The results of these studies
 2409 can be seen for each of the defined sub-categories in Fig. 8.22 for ttH hadronic and Fig. 8.23

for ttH leptonic. From these studies we choose the background model to be a one dimensional exponential function in the ttH leptonic categories and a sum of two exponential functions in the ttH hadronic cat2, while in the the ttH hadronic cat1/cat0 we use a berstein polynomial of order 2. These functions show stable behavoir and are robust against bias as they fulfill the constrain to stay below 20% in all cases.

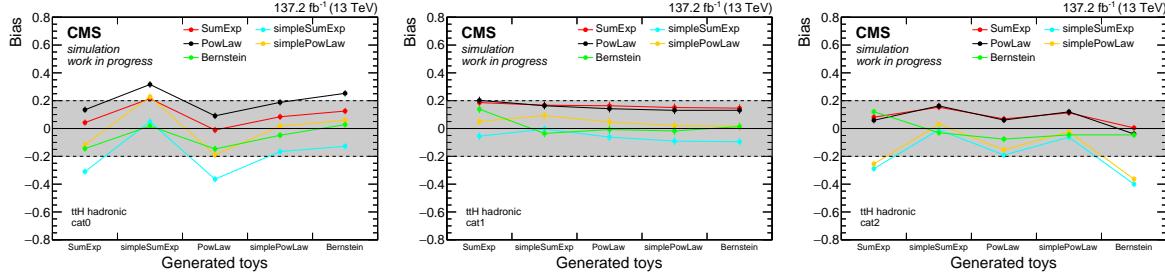


Figure 8.22: Bias studies for all sub-categories in the ttH hadronic channel.

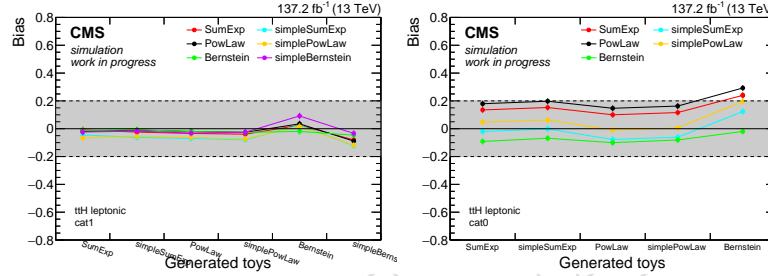


Figure 8.23: Bias studies for all sub-categories in the ttH leptonic channel.

8.5 Results

The limits for the ttH channels are obtained using a parametric background fit in the dimuon mass. The background is modelled by a set of agnostic background functions, which shows stable behaviour and is unbiased against other test functions, as described in Sec. 8.4. The signal is described by a double Crystal Ball distribution. The corresponding expected and observed limits as wells as the expected and observed significances for each ttH category are reported in Tab. 8.13. The corresponding impacts are shown in Fig. 8.24 and a godness-of-fit test is provided by Fig. 8.25.

Table 8.13: Summary of the expected and observed upper limits on μ , computed at the 95% CL for $m_H = 125$ GeV, as well as the expected and observed signal significance for each BDT category and for their combination.

Category	Expected UL	Observed UL	Expected Signif.	Observed Signif.
ttH-lep-cat1	< 6.11	< 10.62	0.42σ	1.49σ
ttH-lep-cat0	< 10.69	< 13.91	0.21σ	0.61σ
ttH-had-cat2	< 9.53	< 11.08	0.22σ	0.33σ
ttH-had-cat1	< 19.00	< 23.53	0.11σ	0.57σ
ttH-had-cat0	< 31.88	< 23.84	0.06σ	0.00σ
Combination ttH	< 4.14	< 7.04	0.53σ	1.55σ

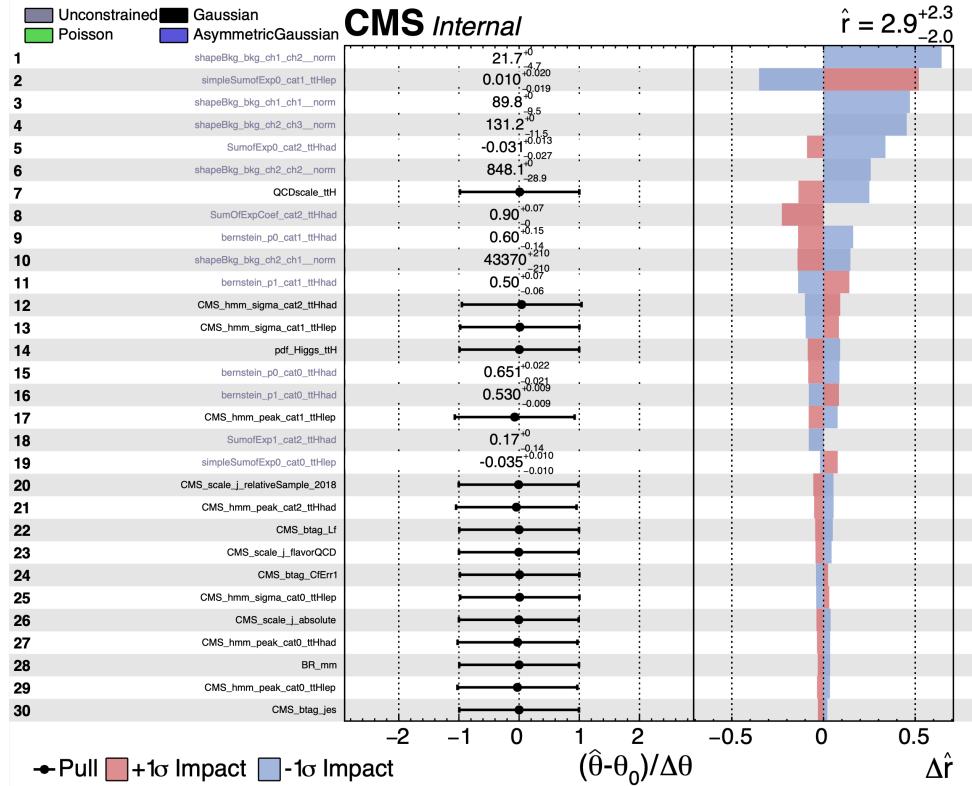


Figure 8.24: Impact study on μ for the combined fit performed in the ttH channel.

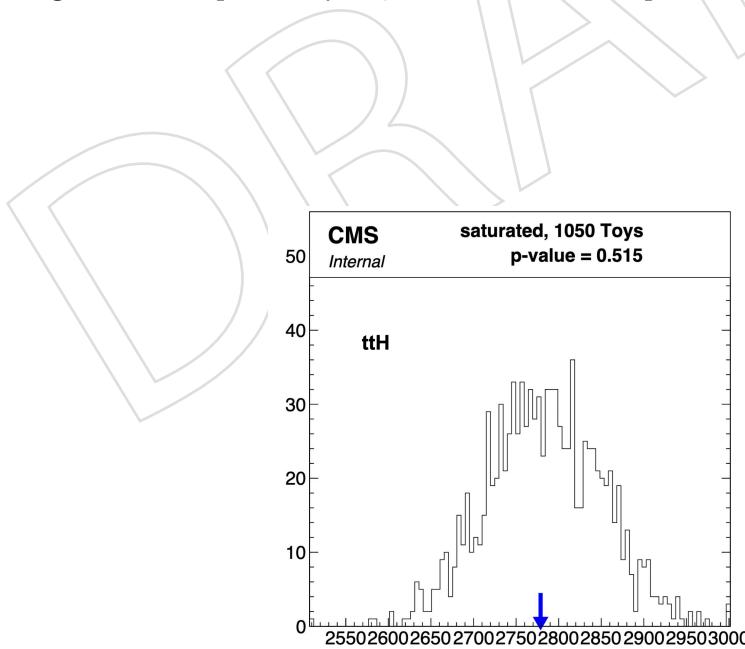


Figure 8.25: Goodness-of-fit tests for the ttH channel. The observed value for the test-statistics is indicated by the blue arrow. The histograms indicate the expected distribution obtained from signal-plus-background toys.

2423 9 VH Leptonic Categories

2424 In the 13 TeV pp collisions in CMS, the vast majority of Higgs bosons are produced through the
 2425 gluon fusion (ggH) process, with smaller contributions from Vector Boson Fusion (VBF) and
 2426 the associated production of a Higgs boson with a W or Z boson (WH and ZH) or a top quark
 2427 pair ($t\bar{t}H$).

2428 In previous searches for $H \rightarrow \mu^+\mu^-$ by CMS and ATLAS, the dominant background is the
 2429 Drell-Yan-produced $\mu\mu$ process. While the WH and ZH production cross sections are 40 –
 2430 90 times smaller than ggH , the Drell-Yan background is greatly suppressed if we only select
 2431 events with at least one extra reconstructed charged lepton (e or μ) in the event, in addition to
 2432 the two muons from the Higgs decay. In CMS, τ leptons almost always decay before exiting
 2433 the inner tracker, and thus cannot be completely reconstructed. This leaves much smaller irre-
 2434ducible backgrounds, the largest of which are WZ and ZZ, with the high-mass $\mu\mu$ pair (close
 2435 to 125 GeV) coming from an off-shell Z or from two separate W or Z boson decays.

2436 The hadronic decays of the W and the Z bosons lead to two jets in the event. These two jets form
 2437 an invariant mass around the on-shell mass of the W or Z boson, but the event is otherwise not
 2438 easily distinguishable from the ggH +jets or the VBF signal events. Thus no exclusive analysis
 2439 is conducted targeting the hadronic decays of the W or the Z boson in the WH or the ZH signal
 2440 modes. Those events are included in the ggH or the VBF analyses.

2441 In the ZH signal events, an invisible decay of a boosted Z boson would also have distinct
 2442 kinematic characteristics, with a real missing transverse momentum. However, most Z bosons
 2443 have low p_T , making it hard to distinguish a true $Z \rightarrow \nu\nu$ decay from the instrumental E_T^{miss} ,
 2444 which comes from the large uncertainty in hadronic calorimetry in CMS. Although it is possible
 2445 to enrich the purity of the $ZH \rightarrow \nu\nu + \mu\mu$ events by posing a tight cut on the E_T^{miss} , it results
 2446 in a very low yield of the ZH signal events (even smaller than the expected signal in the ZH
 2447 $\rightarrow ll + \mu\mu$ analysis). Furthermore, the $ZH \rightarrow \nu\nu + \mu\mu$ phasespace suffers from the irreducible
 2448 backgrounds like $t\bar{t}$ and diboson processes, whereas in the $ZH \rightarrow ll + \mu\mu$ analysis the only
 2449 irreducible background are the ZZ and ggZZ processes. A dedicated analysis targeting the ZH
 2450 $\rightarrow \nu\nu + \mu\mu$ signal would have a much weaker sensitivity than the $ZH \rightarrow ll + \mu\mu$ analysis, and
 2451 is not performed in this paper.

2452 The analysis in the VH channel focuses on the leptonic decays of the vector bosons. The basic
 2453 analysis strategy is to identify events consistent with WH or ZH production where the Higgs
 2454 boson decays to muons, and at least one extra charged lepton ℓ (where $\ell = \mu$ or e) comes from a
 2455 W or Z boson decay. Events are selected such that WH candidates must have one extra charged
 2456 lepton, while the ZH candidates must have a same-flavour opposite-sign (SFOS) dilepton pair
 2457 with an invariant mass close to the Z boson mass. No b-jets are expected from these events,
 2458 thus an additional veto of b-tagged jets is applied to ensure the exclusivity between the VH
 2459 channel and the $t\bar{t}H$ channel.

2460 Single Muon triggers are used for the VH analysis as described in Section 2.4. Object definitions
 2461 are described in Section 3. In particular, the specific selections on muons and electrons in the
 2462 VH channel are described in Section 3.10.

2463 With the aforementioned selections, events are further sub-divided by multivariate discrimina-
 2464 tors designed to separate Higgs signal events from the background processes using variables
 2465 uncorrelated with Higgs candidate mass $m_{\mu\mu}$. Finally, the dimuon mass spectrum is fit be-
 2466 tween 110 and 150 GeV in each sub-category, using analytic functions to model the signal and
 2467 background shapes.

2468 9.1 Event Selections

2469 9.1.1 WH → leptons Event Categories

2470 The WH category contains one additional selected charged lepton (e or μ) and missing trans-
 2471 verse momentum (E_T^{miss}) from the W boson decay. The dominant backgrounds to the WH sig-
 2472 nal are WZ and Z+jets events, where WZ produces 3 prompt leptons, and Z+jets produces two
 2473 muons and one non-prompt lepton. In background events, either the Z boson is off-shell with
 2474 a high invariant mass, or (in the 3μ case) one of the muons included in the Higgs candidate $\mu\mu$
 2475 pair is not from the Z decay.

2476 The following selections are used to define the $\text{WH} \rightarrow \ell + \mu\mu$ category:

- 2477 • At least one muon must have $p_T > 26 \text{ GeV}/29 \text{ GeV}/26 \text{ GeV}$ for year 2016/2017/2018
 2478 respectively, which is matched to a single-muon trigger object
- 2479 • All SFOS lepton pairs must have an invariant mass $> 12 \text{ GeV}$
- 2480 • The charge of the three leptons must add up to ± 1
- 2481 • At least one $\mu^+\mu^-$ pair must have an invariant mass between 110 and 150 GeV
- 2482 • If two $\mu^+\mu^-$ pairs fall in the 110 - 150 GeV mass window, the pair with the higher p_T
 2483 is chosen as the Higgs candidate (denoted $\mu\mu_H$)
- 2484 • The event must contain exactly 0 medium b-tagged jets and less than 2 loose b-
 2485 tagged jets
- 2486 • In 3μ events, the non-Higgs-candidate $\mu^+\mu^-$ pair ($\mu\mu_{\text{OS}}$) must not have an invariant
 2487 mass between 81 and 101 GeV, to suppress WZ and Z+jets backgrounds

2488 9.1.2 ZH → $\ell\ell + \mu\mu$ event selection

2489 The category contains two additional SFOS leptons (e^+e^- or $\mu^+\mu^-$) with an invariant mass
 2490 close to the Z boson mass of 91.2 GeV. The dominant background is ZZ production, where one
 2491 of the Z bosons is off-shell with a high invariant mass, or (in the 4μ case) two muons from
 2492 separate Z decays are chosen as the Higgs candidate $\mu\mu$ pair.

2493 The following selection cuts are used to define the $\text{ZH} \rightarrow \ell\ell + \mu\mu$ category:

- 2494 • At least one muon must have $p_T > 26 \text{ GeV}/29 \text{ GeV}/26 \text{ GeV}$ for year 2016/2017/2018
 2495 respectively, and the event must contain an unprescaled single-muon trigger object
- 2496 • The charge of the four leptons must add up to 0.
- 2497 • All SFOS lepton pairs must have an invariant mass $> 12 \text{ GeV}$.
- 2498 • In $\mu\mu ee$ events, the e^+e^- pair must have invariant mass between 70 and 110 GeV,
 2499 and the $\mu^+\mu^-$ must have invariant mass between 110 and 150 GeV.
- 2500 • In 4μ events, if it is possible to form two distinct $\mu^+\mu^-$ pairs each with a mass be-
 2501 tween 81 and 101 GeV, the event is discarded.
- 2502 • In 4μ events, one muon pair must have mass between 110 and 150 GeV, and the
 2503 other muon pair must have mass between 81 and 101 GeV.
- 2504 • In 4μ events, if both combinations have a muon pair in the Z-mass window and a
 2505 muon pair in the signal-mass window, the combination in which the mass of the Z
 2506 candidate is closer to 91 GeV is chosen.
- 2507 • The event must contain exactly 0 medium b-tagged jets and less than 2 loose b-
 2508 tagged jets.

2509 Figure 9.1 shows the dimuon mass distribution in the WH and ZH channels after the event
 2510 selections. The data and MC agree within their uncertainties. NNLO k-factors are applied to
 WZ, ZZ and ggZZ samples as described in Section 2.

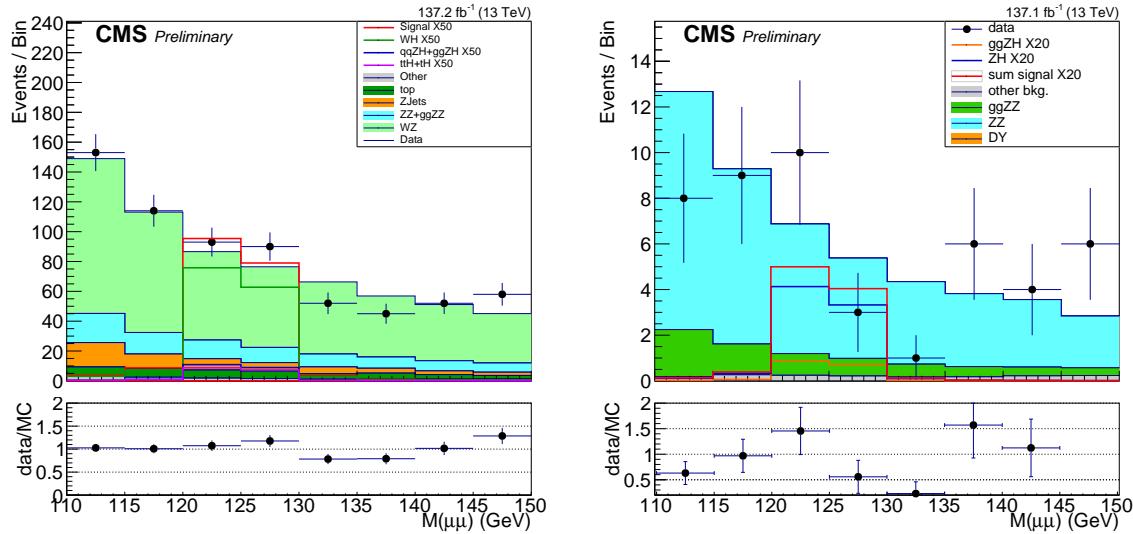


Figure 9.1: Dimuon mass distribution of the WH and the ZH channels after event selection with full Run2 data.

2511

2512 9.2 MVA discriminants in VH leptonic categories

2513 9.2.1 BDT targeting $\text{WH} \rightarrow \ell + \mu\mu$ signal

2514 After the baseline WH category selection, the dominant background is the WZ process, fol-
 2515 lowed by contributions from ZZ and Z+jets events. In WZ events where the Higgs candidate
 2516 $\mu\mu$ pair comes from an off-shell Z boson, the kinematics look almost identical to the WH signal.
 2517 However, some angular correlations between the leptons (including $\Delta\phi$, $\Delta\eta$, and the helicity
 2518 angle θ^*) have been found to provide some discriminating power. The definition of the helicity
 2519 angle is given in the Section 9.2.2. Furthermore, in 3 μ events, a Higgs candidate pair can also
 2520 be formed using one muon from the W decay, and one from the Z decay. Although, most of the
 2521 events are vetoed by removing events with $81 < M(\mu\mu_{OS}) < 101$ GeV, almost half of the WZ
 2522 background in 3 μ events still come from the "wrong combination" WZ decays. This is also true
 2523 for the ZZ background, where one of the four leptons in the event falls out of the acceptance,
 2524 resulting in a final state very much like the WZ events.

2525 A BDT with several input variables is trained to further separate the WH signal from different
 2526 sources of background. The BDT is trained with a combination of MC samples from all eras
 2527 in Run2. The training is performed in the mass window of $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$. Signal
 2528 MC samples with different Higgs mass assumptions $m_H = 120, 125, 130 \text{ GeV}$ are all used as
 2529 signals in the training. Signal events are only used if the candidate $\mu\mu$ pair truly originates from
 2530 the Higgs decay. Signal events are weighted by $1/\sigma[M(\mu\mu_H)]$, as explained in Section 6.1.2 in
 2531 detail. Events with $e + \mu\mu$ and $\mu + \mu\mu$ are used together in the training, but can be distinguished
 2532 by using the "number of electrons" as one of the input variables to the BDT. To increase the
 2533 statistics of non-prompt backgrounds in MC events for training, the lepMVA cut is loosened
 2534 from 0.4 to -0.4, and the non-prompt yields are scaled by a factor of 0.5 to account for the
 2535 increased non-prompt lepton efficiency. With both signal and background samples, half of the
 2536 events is used for training, while the other half is used for testing.

2537 The current BDT includes 16 input variables, which are listed in Table 9.1 and shown in Figures 9.5. Many of them are angular variables between different leptons or between leptons
 2538 and MHT, as each of them provides some discrimination against the prompt WZ and ZZ back-
 2539 ground. The transverse mass of MHT and leptons are included to ensure the discrimination
 2540 against the Z+jets background. The impacts of both MET and MHT related input variables
 2541 on signal and background separation were tested, and MHT related variables were shown to
 2542 perform slightly better. Therefore, in order to keep the number of input variables to the BDT
 2543 reasonable, MET related input variables are not included.
 2544

2545 The BDT output and the ROC curve are shown in Figure 9.2.

2546 The hyperparameters for the WH BDT are:

- 2547 • number of trees: 500,
- 2548 • number of cuts: 20,
- 2549 • maximum depth: 4,
- 2550 • minimum node size: 3%,
- 2551 • Shrinkage = 0.1
- 2552 • Bagged sample fraction = 0.5

Table 9.1: List of input variables used to train the signal-background separation BDT in the WH category. In this table, $\mu\mu_H$ is the Higgs candidate, ℓ_W is the lepton from the W decay, muOS (muSS) refers to the muons in the Higgs candidate which OS (SS) to the lepton.

Variable	TMVA name	Description
$p_T(\mu\mu_H)$	H_pair_pt	p_T of the Higgs candidate
$ \eta(\mu_1) $	muH1_eta_abs	η of the leading muon in the Higgs candidate
$ \eta(\mu_2) $	muH2_eta_abs	η of the trailing muon in the Higgs candidate
$\Delta R(\mu_{SS}, \mu_{OS})$	muSS_muOS_dR	ΔR between the two muons in the Higgs candidate
$p_T(\ell_W)$	lep_pt	p_T of the extra lepton in the event
Number of electrons	Number of electrons	Number of electrons in the event
$\Delta R(\ell_W, \mu\mu_H)$	lep_H_pair_dR	ΔR between the extra lepton and the Higgs candidate
$\Delta\eta(\ell_W, \mu\mu_H)$	lep_H_pair_dEta	$\Delta\eta$ between the extra lepton and the Higgs candidate
$\Delta\eta(\ell_W, \mu_{SS})$	lep_muSS_dEta	$\Delta\eta$ between the extra lepton and the SS muon
$\cos\theta^*(\ell_W, \mu_{SS})$	lep_muSS_cosThStar	$\cos\theta^*$ between the extra lepton and the SS muon
$\Delta R(\ell_W, \mu_{OS})$	lep_muOS_dR	ΔR between the extra lepton and the OS muon
$\Delta\eta(\ell_W, \mu_{OS})$	lep_muOS_dEta	$\Delta\eta$ between the extra lepton and the OS muon
$\cos\theta^*(\ell_W, \mu_{OS})$	lep_muOS_cosThStar	$\cos\theta^*$ between the extra lepton and the OS muon
$M_T(\mu_{SS}, \text{MHT})$	muSS_MHT_MT	transverse mass of the MHT and the SS muon
$M_T(\ell_W, \text{MHT})$	lep_MHT_MT	transverse mass of the MHT and the extra lepton
$ \Delta\phi(\ell_W, \text{MHT}) $	lep_MHT_dPhi_abs	$ \Delta\phi $ between the MHT and the extra lepton

2553 9.2.2 BDT targeting $ZH \rightarrow \ell\ell + \mu\mu$ signal

2554 After the baseline selection and category selection, the dominant backgrounds in the ZH chan-
 2555 nel are the ZZ to 4 leptons and ggZZ to 4 leptons processes. Other backgrounds, prompt or
 2556 non-prompt, have negligible contribution in this channel. Both the ZZ and ggZZ processes
 2557 have the identical final states as the ZH signal. Apart from the dimuon mass, which is used
 2558 in the last stage for signal extraction, the kinematic discrimination between the signal and the
 2559 background processes lies in the kinematic angles between the leptons in the event.

2560 The input variables to the BDT are listed in Table 9.2 and shown in Figure 9.4. In particular,

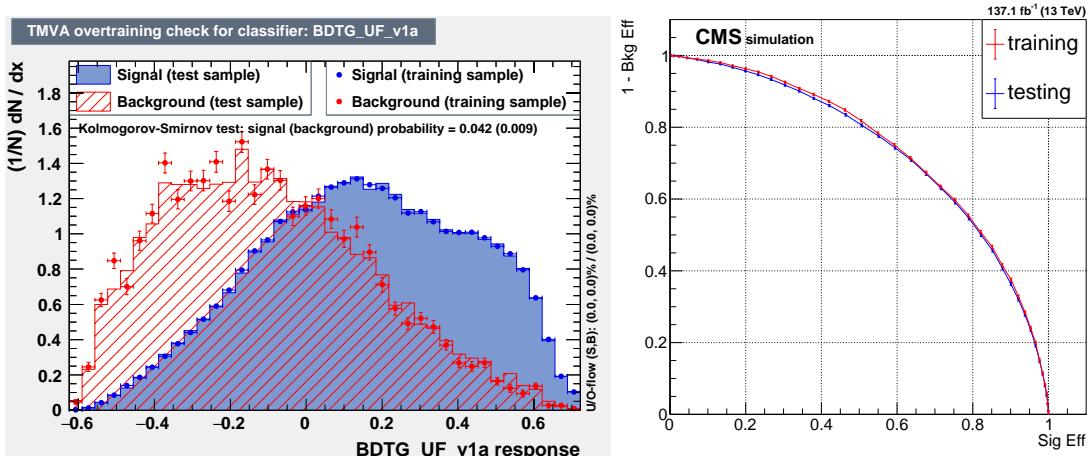


Figure 9.2: On the left, the BDT output score from the $W H \rightarrow 3\ell$ BDT, with signal in blue and background in red. On the right, the receiver operating characteristic (ROC) curve, with training in red and testing in blue. A slight over-training is observed in the region of low signal efficiency, due to the fluctuation in background. As will be shown in Fig. 9.6, the BDT does not sculpt the shape of $m_{\mu\mu}$.

one of the strongest discriminating variables is $\cos\theta^*(\mu\mu_H, \ell\ell_Z)$, the helicity angle between the Higgs candidate and the Z candidate in the event. The helicity angle θ^* is defined between two particles in the frame in which their parent particle is at rest. In that frame, the helicity angle is the angle between the direction of the decay and the direction of the boost of the parent. In the ZZ background process, a propagator Z boson (Z_0) decays to two Z bosons (Z_1 and Z_2), each of which decays to a lepton pair. Since Z bosons are spin-1 particles, in the $Z_0 \rightarrow Z_1 Z_2$ process, the direction of the decay is more likely to align with the direction of the momentum of the Z_0 . Whereas in the ZH events, since Higgs bosons are spin-0 particles, there is no preferred direction for the $Z_0 \rightarrow Z_1 H$ decay. A similar kinematic discrimination is also present in the helicity angle $\cos\theta^*(\mu_1, \mu_2)$, between the $Z \rightarrow \mu\mu$ decay and the $H \rightarrow \mu\mu$ decay. However, in this analysis, due to the acceptance of the CMS detector, the distribution of $\cos\theta^*(\mu_1, \mu_2)$ is sculpted and turns out not very different between the signal and the background. This variable is included in the initial training and later discarded during the variable trimming process.

Similar to the WH BDT training, as described in Section 9.2.1, the training is performed with MC samples from all eras in Run2. The training is performed in the mass window of $110\text{ GeV} < m_{\mu\mu} < 150\text{ GeV}$. This training was performed prior to the production of ggZH signal samples, thus only qqZH MC is used as signal events. Signal MC samples with different Higgs mass assumptions $m_H = 120, 125, 130\text{ GeV}$ are used. Signal events are only used if the candidate $\mu\mu$ pair truly originates from the Higgs decay. Signal events are weighted by $1/\sigma[M(\mu\mu_H)]$. Events with $ee + \mu\mu$ and $\mu\mu + \mu\mu$ are used together in the training, but can be distinguished with the "lepton flavour" as one of the input variables. To increase the statistics of MC events for training, the lepMVA cut is loosened from 0.4 to -0.4, but no rescaling is necessary, since there is no non-prompt background component.

The BDT output and the ROC curve are shown in Figure 9.3.

The hyperparameters for the ZH BDT are:

- number of trees: 400,
- number of cuts: 10,
- maximum depth: 3,

- 2589 • minimum node size: 3%,
 2590 • Shrinkage = 0.1
 2591 • Bagged sample fraction = 0.5

Table 9.2: List of input variables used to train the signal-background separation BDT in the ZH category.

Variable	TMVA name	Description
$p_T(\mu\mu_H)$	dimu_pt	p_T of the Higgs candidate
$ \eta(\mu\mu_H) $	dimu_abs_eta	$ \eta $ of the Higgs candidate
$ \Delta\phi(\mu\mu_H) $	dimuon_abs_dPhi	$ \Delta\phi $ between the muons in the Higgs candidate
$M(\ell\ell_Z)$	dilep_mass	invariant mass of the Z candidate
$p_T(\ell\ell_Z)$	dilep_pt	p_T of the Z candidate
$ \eta(\ell\ell_Z) $	dilep_abs_eta	$ \eta $ of the Z candidate
$\Delta R(\ell\ell_Z)$	dilep_dR	ΔR between the leptons in the Z candidate
lepton flavor	lep_ID	flavor of the Z candidate lepton pair
$\cos\theta^*(\mu\mu_H, \ell\ell_Z)$	cts_dipair_H	cosine helicity angle between the Higgs and the Z candidates
$\Delta\eta(\mu\mu_H, \ell\ell_Z)$	dipair_dEta_H	$\Delta\eta$ between the Higgs and the Z candidates

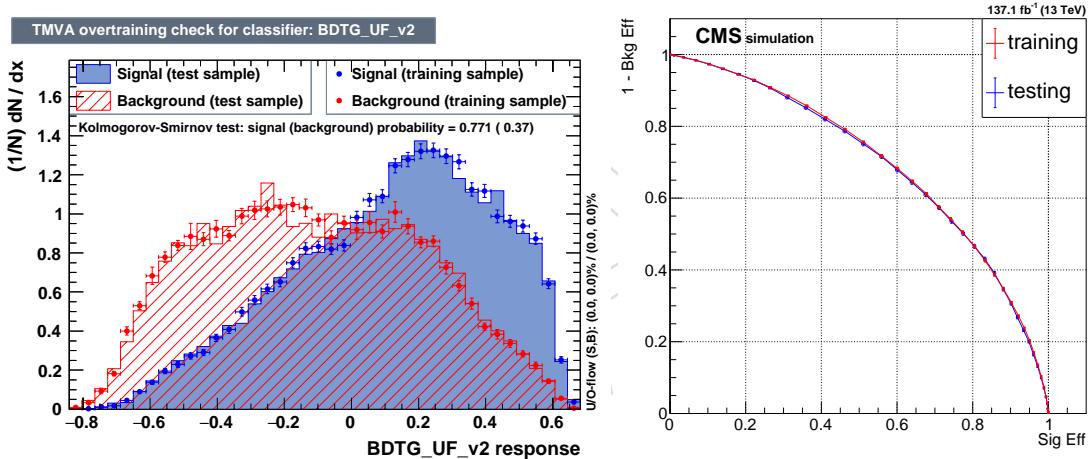


Figure 9.3: On the left, the BDT output score, with signal in blue and background in red. On the right, the receiver operating characteristic (ROC) curve showing the $ZH \rightarrow 4\ell$ BDT performance, with training in red and testing in blue.

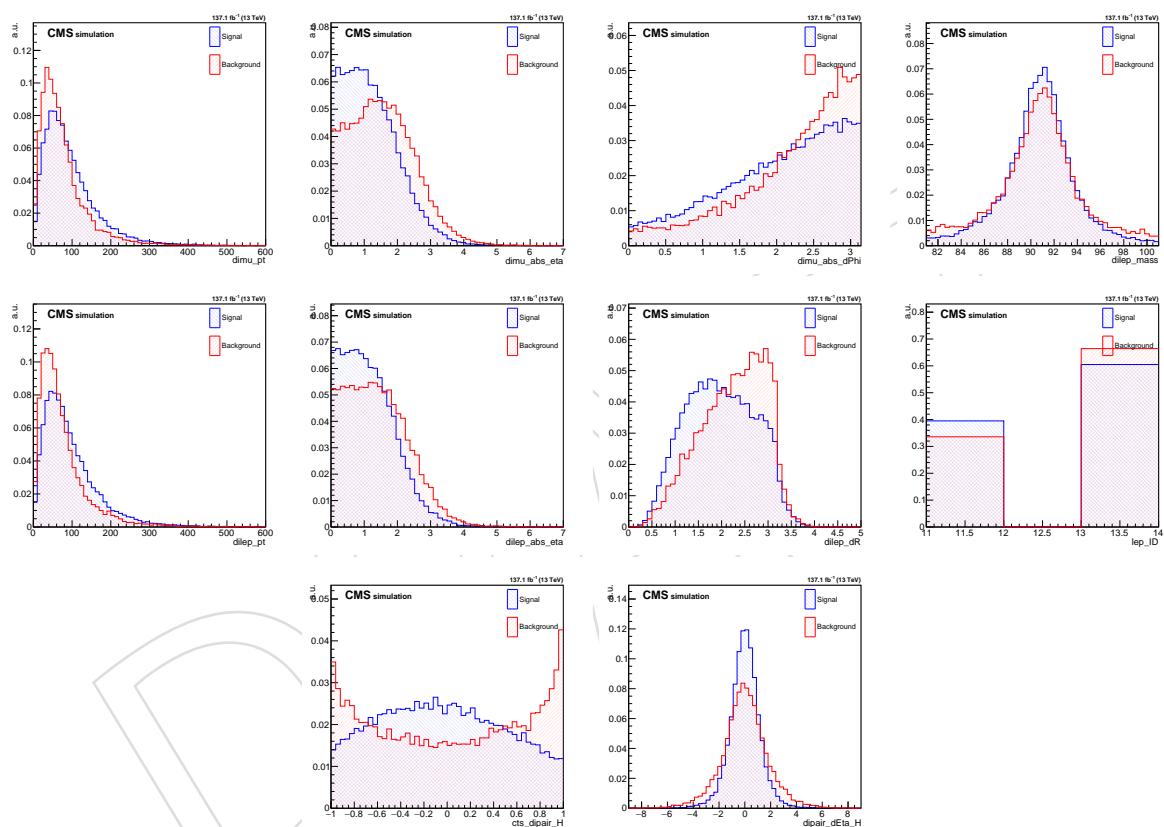


Figure 9.4: Input variables to the $ZH \rightarrow 4\ell$ BDT, with signal in blue and background in red.

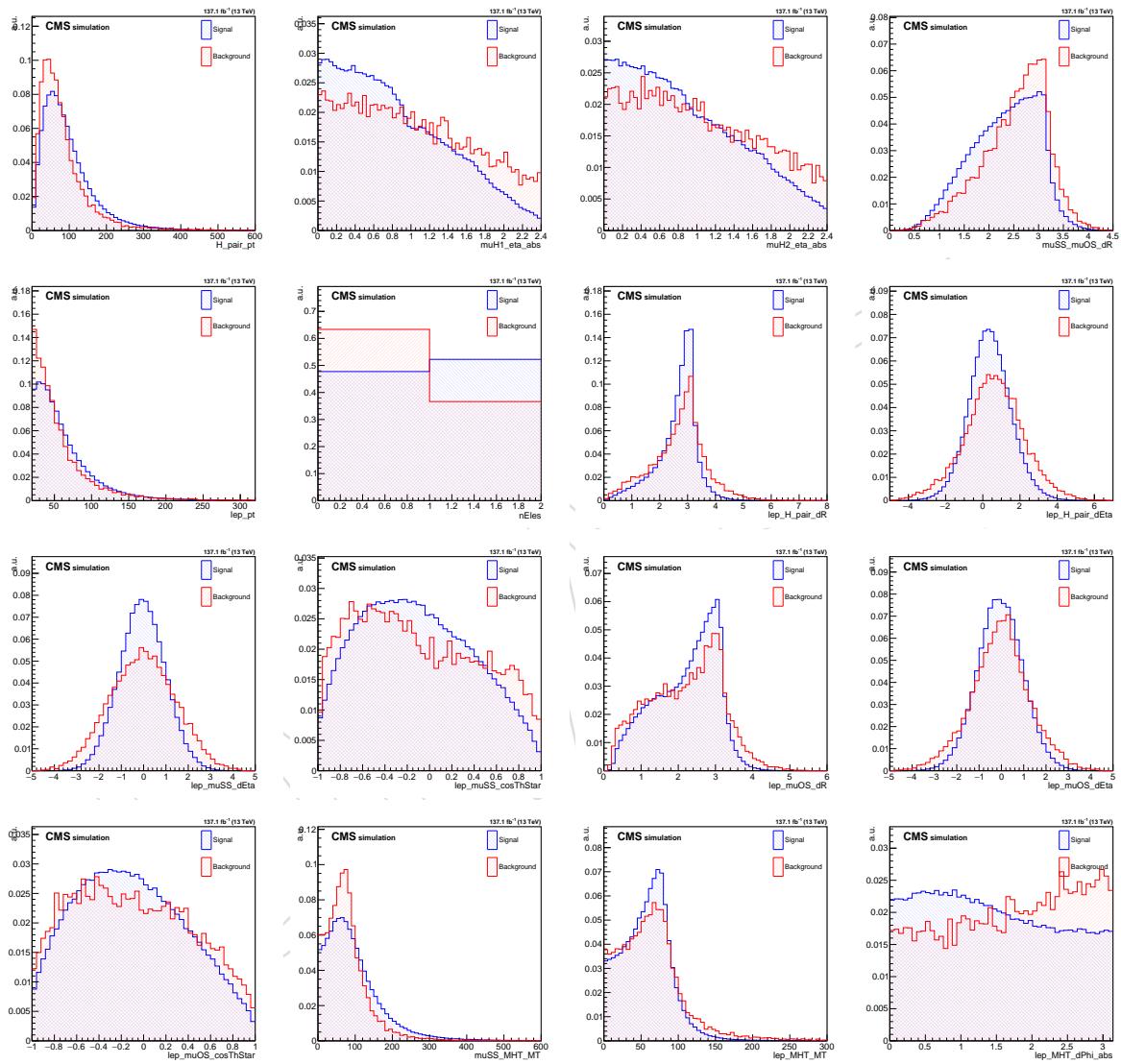


Figure 9.5: Input variables to the $WH \rightarrow 3\ell$ BDT, with signal in blue and background in red.

2592 **9.2.3 Validation of WH BDT and ZH BDT**

2593 As described in Section 5, the BDT discriminants are used to divide events into different cat-
 2594 egories, in which a data-driven shape analysis on the dimuon mass is performed. For this
 2595 purpose, it is important that the BDT output is not correlated to the dimuon mass. Also, the
 2596 BDT discriminants need to perform the same way on data as they do on MC.

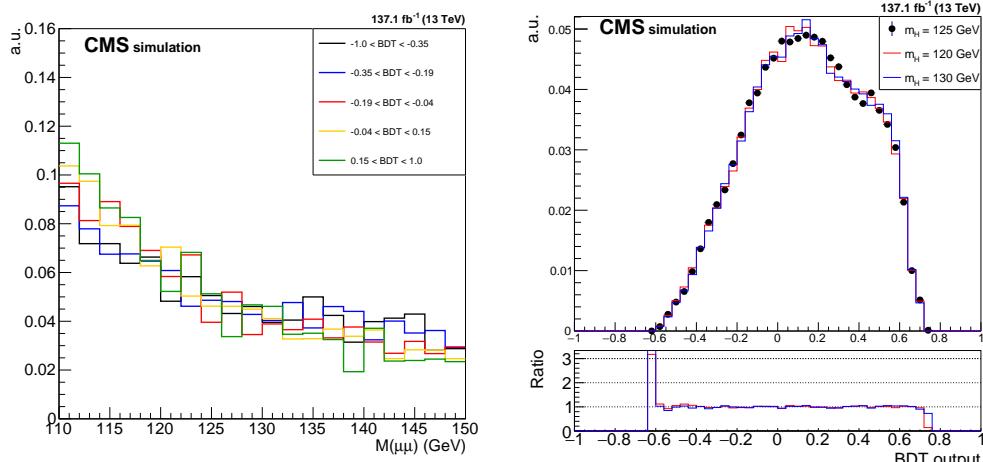


Figure 9.6: For the WH BDT, the distribution of the dimuon mass shape in the background for five different BDT quantile (left), and the distribution of the BDT output for three different signal mass assumptions (right).

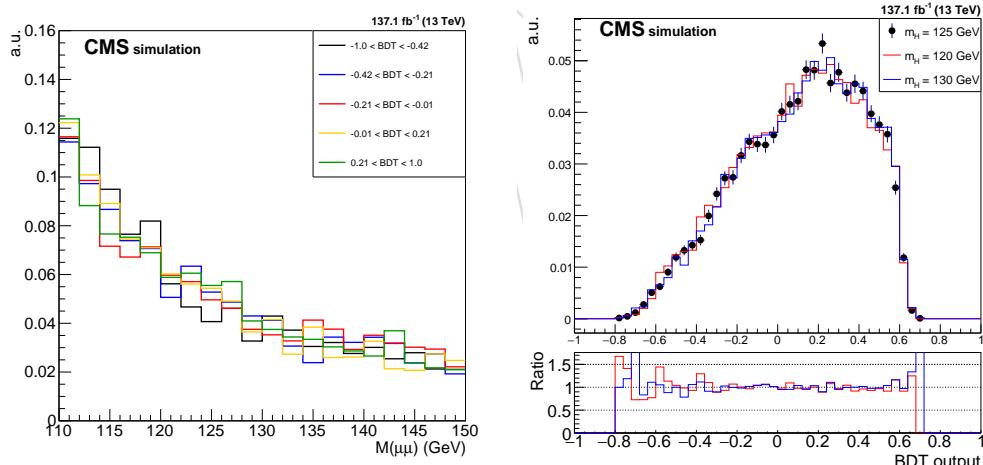


Figure 9.7: For the ZH BDT, the distribution of the dimuon mass shape in the background for five different BDT quantile (left), and the distribution of the BDT output for three different signal mass assumptions (right).

2597 Figure 9.6 and 9.7 show that the WH BDT and ZH BDT discriminants are not correlated with
 2598 the dimuon mass. Figure 9.8 shows the BDT output of both the WH BDT and the ZH BDT. The
 2599 BDT outputs agree between data and MC within uncertainties. Figure 9.9 and 9.10 show the
 2600 input variables of the two BDTs. The BDT input variables agree between data and MC within
 2601 uncertainties.

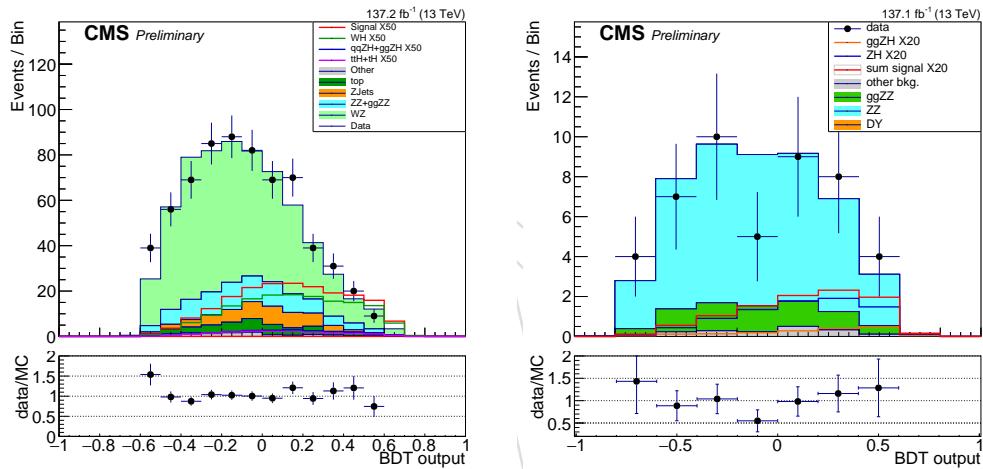


Figure 9.8: The WH BDT output (left) and the ZH BDT output (right) in full Run 2 in the signal region $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$, data not blinded.

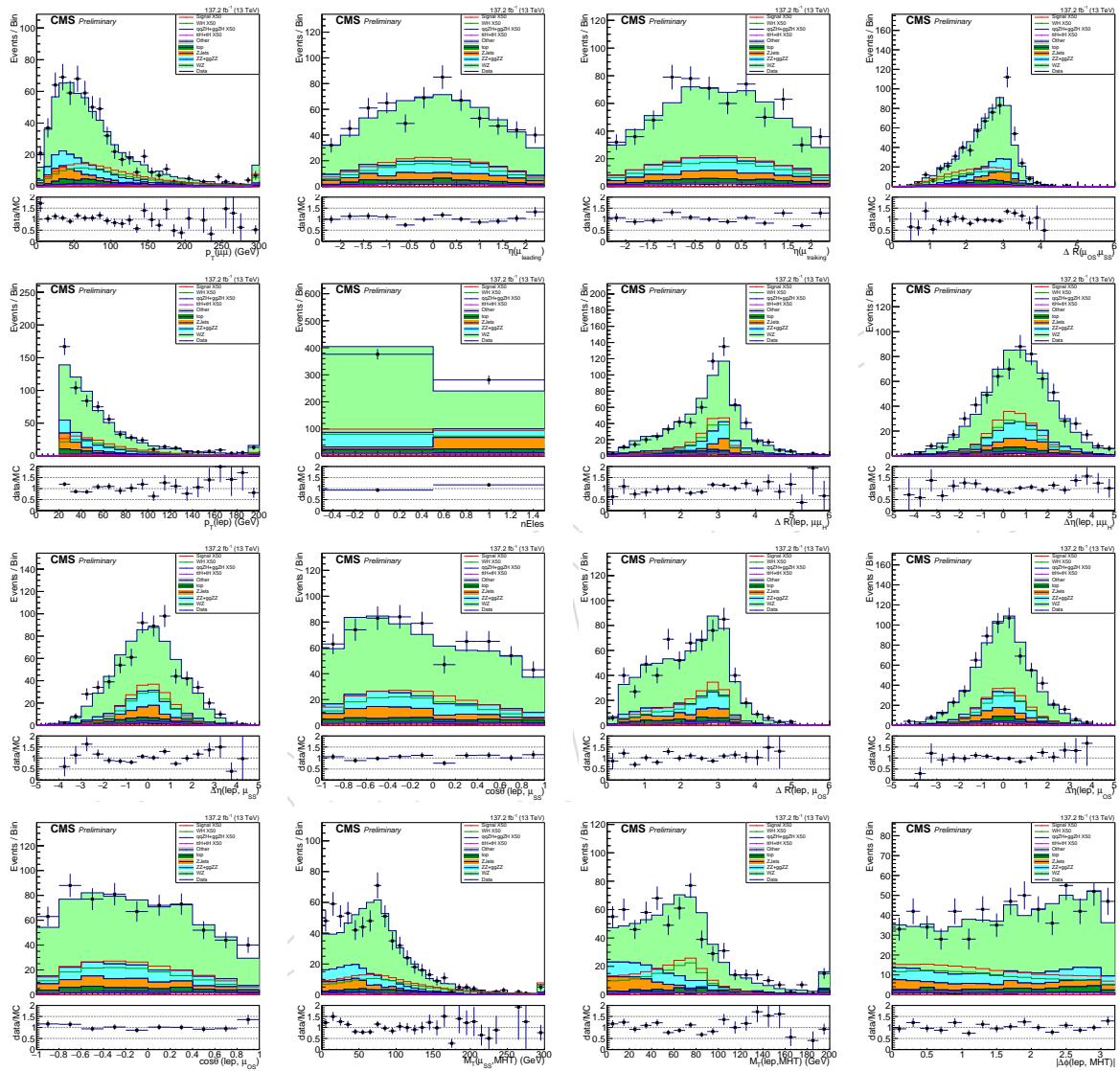


Figure 9.9: Input variables to the WH BDT in full Run 2 in the signal region 110 GeV < m_{μμ} < 150 GeV, data not blinded.

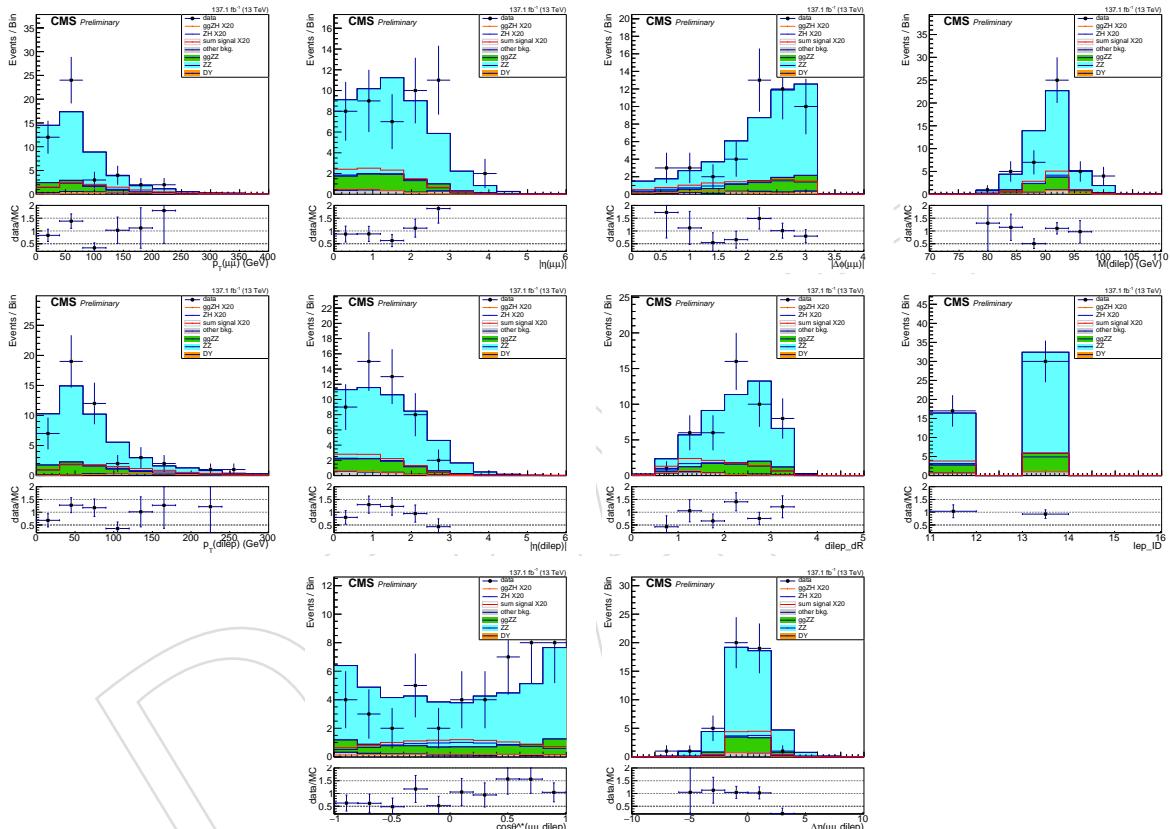


Figure 9.10: Input variables to the ZH BDT in full Run 2 in the signal region $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$, data not blinded.

9.2.4 BDT categorization

Based on the output of the BDT discriminants, both the WH and the ZH phase-space can be further divided into several categories with different S/B ratios, to optimize the overall sensitivity of the VH analysis. To achieve that, an iterative procedure is taken. In each iteration, a cut is scanned at a step of 0.01 of the BDT value and the sum of significance of the corresponding categories is calculated as the figure of merit. The figure of merit is chosen as the S/\sqrt{B} in each category summed in quadrature, where S and B represent the expected signal and background yields within the FWHM of the signal peak in that category. In addition, to ensure that there are enough events in each category to perform a shape analysis, all categories have to meet a minimal total event yield requirement during the BDT scanning process.

Figure 9.11 shows the iterations performed on the WH BDT. The minimum number of events in each category is set to be 30. With this procedure, two BDT boundaries are set, dividing the WH phase-space into 3 categories, BDT [-1.0, -0.1], BDT [-0.1, 0.3] and BDT [0.3, 1.0].

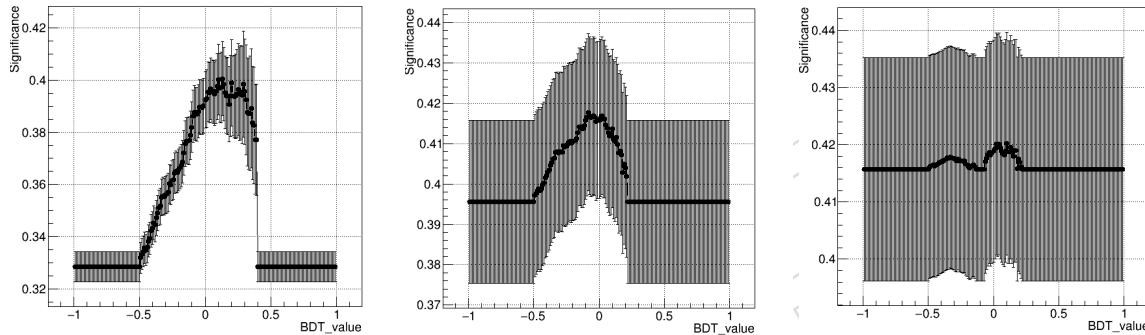


Figure 9.11: Scans for the first (left), second (middle) and a potential third (right) BDT cut in the WH channel. The first BDT cut is chosen at 0.3. The second BDT cut is chosen at -0.1. A third BDT cut is not necessary.

Similarly, Figure 9.12 shows the scan performed on the ZH BDT. The minimum number of events in each category is set to be 16. With this procedure, one BDT boundary is set, dividing the ZH phase-space into 2 categories, BDT [-1.0, -0.1], and BDT [-0.1, 1.0].

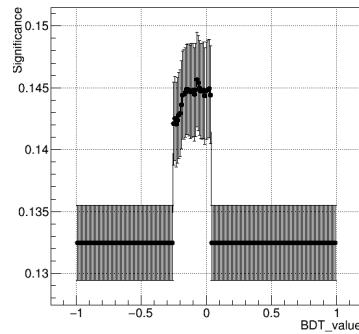


Figure 9.12: Scans for the BDT cut in the ZH channel. The BDT cut is chosen at -0.1

9.3 VH signal models

In all sub-categories, signals are modeled independently by different production modes. In particular, qqZH and ggZH signals are modeled separately in the VH channels. Each of the components is modelled with a Double-sided Crystal Ball function (DCB), as described in

2622 Equation 4.2. In all DCB functions, the parameters n_L and n_R are fixed to 2.0, since they only
 2623 affect the shape in tails and can take values in a large range without changing the quality of the
 2624 fit by much. Other parameters are freely floating.

2625 Examples of signal modeling are shown in Figures 9.13 and 9.14. Please note that the plots
 2626 shown are the signals in the inclusive WH and ZH channels. The actual models used in each
 2627 category are slightly different. ggH, VBF and bbH contributions to the WH channel are negligi-
 2628 ble and are not considered. Similarly, in the ZH channel only qqZH and ggZH are considered
 2629 since all other contributions are negligible.

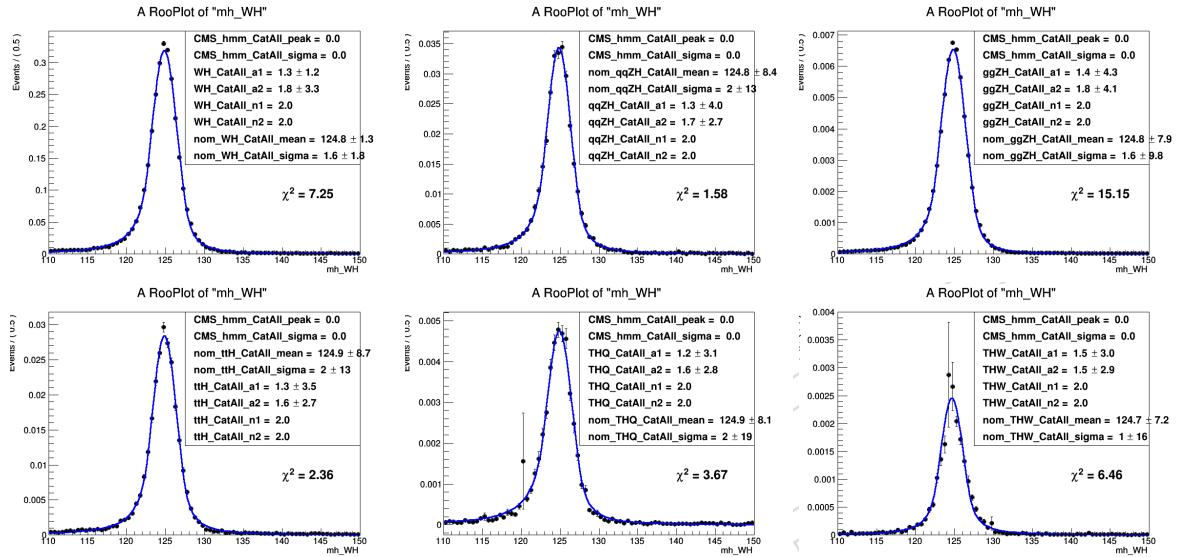


Figure 9.13: The signal modeling in the $\text{WH} \rightarrow \ell + \mu\mu$ inclusive category. Considered signal modes are WH (top left), qqZH (top middle), ggZH (top right), ttH (bottom left), THQ (bottom middle), and THW (bottom right).

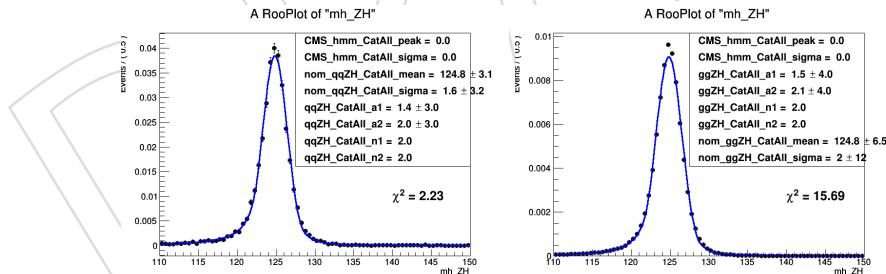


Figure 9.14: The signal modeling in the $\text{ZH} \rightarrow \ell\ell + \mu\mu$ inclusive category. Considered signals modes are qqZH (left) and ggZH (right).

2630 9.3.1 VH signal systematic uncertainties

2631 Several sources of signal systematic uncertainties are considered. The *shape* uncertainties af-
 2632 fecting the modelling of the signal peak are considered as described in Sec. 6.3.2. The *rate* un-
 2633 certainties are implemented as uncertainties on the signal yield. The luminosity uncertainties
 2634 for each year are used as recommended from LUM POG. The impacts from theoretical un-
 2635 certainties in the WH and ZH channels are shown in Table 2.1. The impacts from experimental
 2636 uncertainties in the WH and ZH channel are shown in Table 9.3 to Table 9.7.

Table 9.3: Uncertainties on different signal components in the WH and ZH channels related to pileup re-weight and L1 prefire re-weight. Uncertainties smaller than 0.1% are neglected.

Uncertainty	Category	WH	qqZH	ggZH	ttH	THQ	THW
pileup 2016	WH cat0	1.008/0.993	1.008/0.993	1.006/0.995	1.009/0.993	1.010/0.991	1.004/0.996
	WH cat1	1.008/0.993	1.006/0.995	1.006/0.995	1.006/0.995	1.010/0.991	1.007/0.994
	WH cat2	1.006/0.995	1.004/0.997	1.006/0.996	1.007/0.994	1.002/0.997	1.005/0.996
	ZH cat0	-	1.007/0.993	1.008/0.993	-	-	-
	ZH cat1	-	1.008/0.993	1.007/0.994	-	-	-
pileup 2017	WH cat0	1.006/0.995	1.002/0.998	1.003/0.997	1.003/0.998	1.003/0.995	1.006/0.993
	WH cat1	1.004/0.996	1.004/0.996	1.003/0.997	1.003/0.997	1.004/0.995	1.003/0.996
	WH cat2	1.005/0.995	1.005/0.997	1.003/0.997	1.006/0.994	1.005/0.995	1.004/0.998
	ZH cat0	-	1.004/0.996	1.004/0.995	-	-	-
	ZH cat1	-	1.003/0.996	1.004/0.996	-	-	-
pileup 2018	WH cat0	1.006/0.994	1.004/0.996	1.005/0.995	1.006/0.994	1.005/0.995	1.005/0.995
	WH cat1	1.005/0.995	1.003/0.997	1.004/0.996	1.005/0.995	1.004/0.996	1.008/0.992
	WH cat2	1.004/0.996	1.002/0.997	1.004/0.996	1.004/0.997	1.005/0.995	1.008/0.992
	ZH cat0	-	1.006/0.994	1.005/0.995	-	-	-
	ZH cat1	-	1.006/0.994	1.005/0.995	-	-	-
prefire 2016	WH cat0	1.001/0.999	1.001/0.999	1.002/0.998	1.002/0.998	1.002/0.998	1.002/0.998
	WH cat1	1.001/0.999	1.001/0.999	1.002/0.998	1.002/0.998	1.002/0.998	1.001/0.999
	WH cat2	-	1.001/0.999	1.001/0.999	1.001/0.999	1.002/0.998	1.001/0.999
	ZH cat0	-	1.001/0.999	1.001/0.999	-	-	-
	ZH cat1	-	1.001/0.999	1.001/0.999	-	-	-
prefire 2017	WH cat0	1.002/0.998	1.003/0.997	1.003/0.997	1.004/0.996	1.003/0.997	1.002/0.997
	WH cat1	1.001/0.999	1.003/0.997	1.003/0.997	1.003/0.997	1.003/0.996	1.002/0.998
	WH cat2	1.001/0.999	1.002/0.998	1.002/0.998	1.002/0.998	1.002/0.998	1.001/0.999
	ZH cat0	-	1.002/0.998	1.003/0.997	-	-	-
	ZH cat1	-	1.001/0.999	1.002/0.998	-	-	-

Table 9.4: Uncertainties on different signal components in the WH and ZH channels related to the muon energy scale. Uncertainties smaller than 0.1% are neglected.

Category	WH	qqZH	ggZH	ttH	THQ	THW
WH cat0	-	1.000/1.001	-	0.999/1.003	1.000/0.999	1.000/1.003
WH cat1	-	0.999/1.000	1.001/0.999	1.001/1.000	0.999/1.002	0.999/0.998
WH cat2	1.000/1.001	0.998/1.000	0.998/1.001	0.998/1.000	1.000/1.004	1.001/1.000
ZH cat0	-	1.000/1.001	1.000/1.001	-	-	-
ZH cat1	-	0.999/1.000	-	-	-	-

Table 9.5: Uncertainties on different signal components in the WH and ZH channels related to lepMVA scale factor. The lepMVA scale factor is the only scale factor applied to correct for the lepton efficiency modelling. The ID scale factor and Isolation scale factor are covered by the lepMVA scale factors. Uncertainties smaller than 0.1% are neglected.

Uncertainty	Category	WH	qqZH	ggZH	ttH	THQ	THW
muon SF	WH cat0	0.982/1.018	0.983/1.018	0.977/1.023	0.977/1.023	0.979/1.021	0.975/1.025
	WH cat1	0.983/1.018	0.983/1.017	0.976/1.024	0.980/1.020	0.981/1.020	0.977/1.023
	WH cat2	0.977/1.023	0.981/1.020	0.975/1.025	0.975/1.025	0.977/1.024	0.972/1.029
	ZH cat0	-	0.981/1.019	0.975/1.026	-	-	-
	ZH cat1	-	0.975/1.026	0.967/1.034	-	-	-
electron SF	WH cat0	0.997/1.003	0.996/1.004	0.996/1.004	0.997/1.003	0.996/1.004	0.998/1.002
	WH cat1	0.995/1.005	0.994/1.006	0.995/1.005	0.995/1.005	0.994/1.006	0.995/1.005
	WH cat2	0.995/1.005	0.994/1.006	0.994/1.006	0.995/1.005	0.994/1.006	0.995/1.005
	ZH cat0	-	0.994/1.006	0.994/1.006	-	-	-
	ZH cat1	-	0.992/1.008	0.994/1.006	-	-	-

Table 9.6: Uncertainties on different signal components in the WH and ZH channels related to B-jet vetoing. Uncertainties smaller than 0.1% are neglected.

Category	WH	qqZH	ggZH	ttH	THQ	THW
WH cat0	1.001/0.999	1.001/0.999	0.992/1.008	1.055/0.947	0.991/1.009	0.986/1.014
WH cat1	1.001/0.999	1.001/0.999	0.991/1.009	1.057/0.945	0.992/1.008	0.987/1.013
WH cat2	1.001/0.999	1.000/0.999	0.990/1.010	1.053/0.949	0.991/1.009	0.986/1.014
ZH cat0	-	1.001/0.999	0.994/1.006	-	-	-
ZH cat1	-	1.002/0.998	0.995/1.005	-	-	-

Table 9.7: Uncertainties on different signal components in the WH Cat0 related to jet energy scale. JEC uncertainties are in general small for the main signals in the WH and ZH channels. WH Cat0 is shown as an example. Numbers in other categories are similar. Uncertainties smaller than 0.1% are neglected.

Uncertainty	WH	qqZH	ggZH	ttH	THQ	THW
flavorQCD	1.001/1.000	0.996/1.000	1.001/1.000	1.006/0.991	1.002/0.995	1.015/0.992
relativeBal	1.001/1.000	0.998/1.000	1.001/1.000	1.007/0.995	1.000/0.998	1.007/0.995
absolute	-	1.000/1.001	-	-	1.000/1.001	1.001/1.004
BBEC1	1.001/1.000	0.996/1.000	1.001/0.999	1.015/0.989	1.002/0.993	1.012/0.993
EC2	-	-	-	0.999/1.000	0.999/0.997	1.002/1.001
HF	0.999/1.001	0.998/1.000	1.001/1.001	1.005/0.994	1.000/0.995	1.005/1.003
relativeSample_2016	-	0.999/1.000	-	1.001/0.997	-	1.001/0.997
absolute_2016	-	1.000/0.999	-	1.001/0.999	1.000/1.001	-
BBEC1_2016	-	1.000/0.999	-	1.000/0.999	1.000/1.001	1.001/1.000
EC2_2016	-	-	-	-	1.000/1.001	1.001/1.000
HF_2016	-	-	-	-	-	1.001/1.000
relativeSample_2017	-	0.999/1.002	1.001/1.000	1.001/0.998	0.998/1.001	1.003/1.003
absolute_2017	-	0.999/1.002	-	1.002/0.999	-	1.003/1.002
BBEC1_2017	-	0.999/1.001	-	1.002/1.000	0.999/1.001	1.002/1.000
EC2_2017	-	1.001/1.000	-	-	1.001/1.003	1.001/1.003
HF_2017	-	-	-	-	0.999/0.998	-
relativeSample_2018	0.999/1.001	1.001/1.002	1.001/1.000	1.004/0.995	1.001/0.996	1.007/0.993
absolute_2018	1.000/1.001	0.998/1.000	-	1.000/0.999	1.002/0.996	1.003/0.994
BBEC1_2018	1.000/1.001	0.998/1.000	-	1.000/0.999	-	1.001/0.999
EC2_2018	1.000/1.001	0.999/1.001	-	1.000/1.001	1.000/1.001	1.000/1.001
HF_2018	-	1.000/0.999	-	1.000/1.001	1.000/0.999	-

9.4 VH background models

The backgrounds in the VH categories are modelled by fitting analytic functions to the background shape. As described in Section 5.3, both Breit-Wigner-like shapes and agnostic function sets are studied in the VH categories. Specifically, as all VH categories have low event yield, some order-1 agnostic functions are not always stable in the fitting. As a result, a BWZ times Bernstein function (2 DOF) is used instead of the plain Bernstein polynomials, and a single power law (or exponential) plus a constant function (2 DOF) is used along with the plain single power law (or exponential) function (1 DOF). This ensures the stability of all background functions as well as their flexibility.

9.4.1 VH Background systematic uncertainty

The VH analysis is performed in a data-driven fashion, so that the uncertainty in the background estimation comes from the statistical uncertainty in data and the prior assumption of the function form of the background. Each of the background function candidates implies some prior assumption of the true background shape, which could be a source of bias. To evaluate such biases, studies are conducted between different groups of background functions. In each

category, 3000 toys are generated for each background function candidate as the "truth model" and with injected signal at either 0 or 1 SM signal strength. Different functions are taken as the "fit models" and fitted to the toys in the process of signal extraction. The extracted signal strength is compared with the injected signal strength as an evaluation of the bias between the "fit model" and the "truth model". The bias is defined as

$$\text{Bias}(\mu_{\text{fit}}, \mu_{\text{inj}}, \sigma_{\mu}) = \frac{\mu_{\text{fit}} - \mu_{\text{inj}}}{\sigma_{\mu}}$$

where μ_{fit} is the signal strength extracted from the fit, μ_{inj} is the injected signal strength in each toy and σ_{μ} is the uncertainty on the signal strength. A bias of 20% corresponds to an impact on the final limit of about 2% and can be neglected. The results of the bias studies for the WH categories are shown in Figure 9.15. The results of the bias studies for the ZH categories are shown in Figure 9.16.

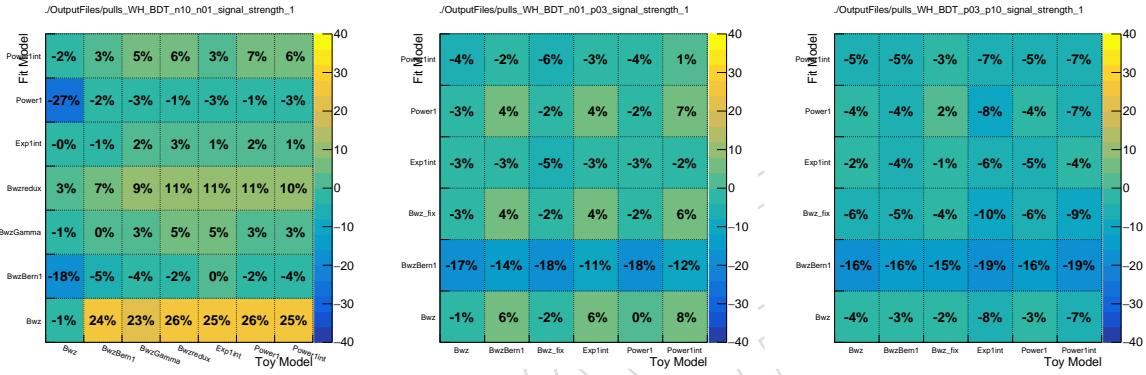


Figure 9.15: Bias in different BDT-based WH categories. The categories are: Cat0 BDT [-1.0, -0.1] (left), Cat1 BDT [-0.1, 0.3] (middle), Cat2 BDT [0.3, 1.0] (right). In the tables, the Power stands for a single "Power" function and the "PowerInt" stands for a single power function plus a constant.

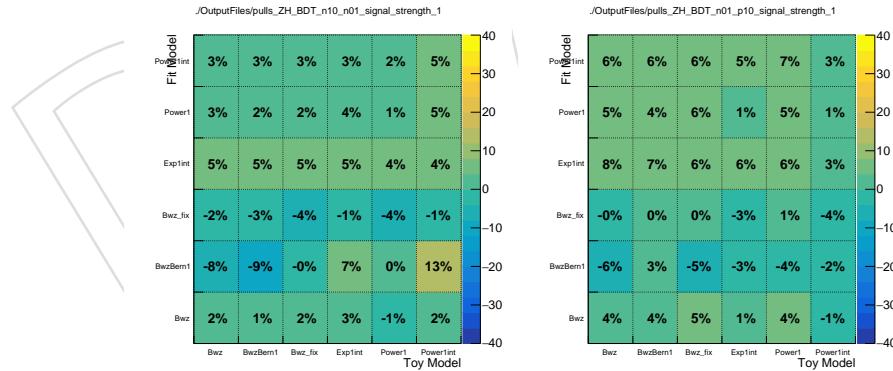


Figure 9.16: Bias in different BDT-based ZH sub-categories. The categories are: Cat0 BDT [-1.0, -0.1] (left), Cat1 BDT [-0.1, 1.0] (right). In the tables, the "Power" stands for a single power function and the "PowerInt" stands for a single power function plus a constant.

Based on these results, in the WH [-1.0, -0.1] category, the BWZGamma function (2 DOF) is chosen as the background model, and in all other categories, the BWZ (1 DOF) is chosen to be the background model. The fits on the blinded data are shown in Figure 9.17 and 9.18.

9.5 VH Results

Figure 9.19 shows the goodness-of-fit test for the combined fit performed in the WH and the ZH channels. Figure 9.20 shows the impact study in the WH and the ZH channels. The full

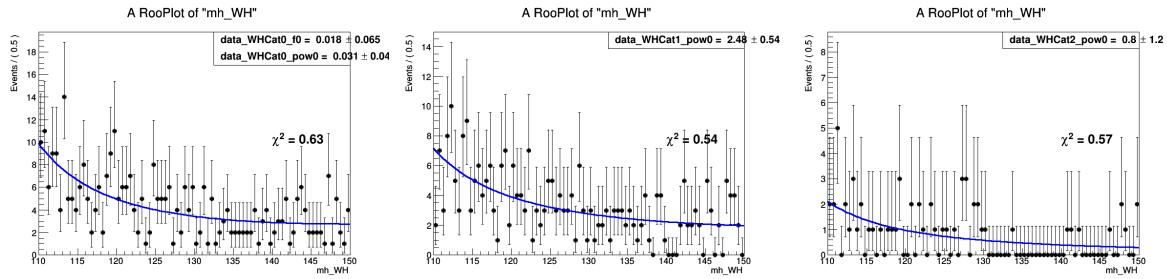


Figure 9.17: Background fits on the blinded data in different BDT-based WH categories. The categories are: Cat0 BDT [-1.0, -0.1] (left), Cat1 BDT [-0.1, 0.3] (middle), Cat2 BDT [0.3, 1.0] (right).

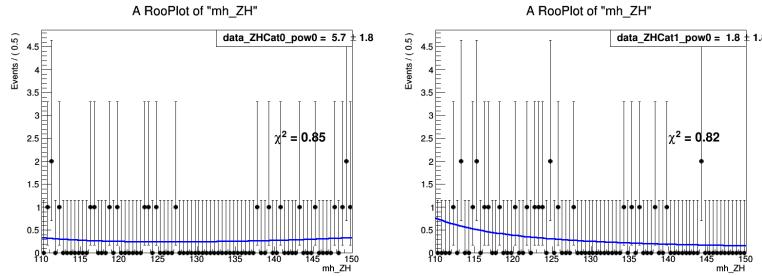


Figure 9.18: Background fits on the blinded data in different BDT-based ZH categories. The categories are: Cat0 BDT [-1.0, -0.1] (left) and Cat1 BDT [-0.1, 1.0] (right).

Run 2 expected and observed limits and significance for the WH and ZH categories are shown in Table 9.8.

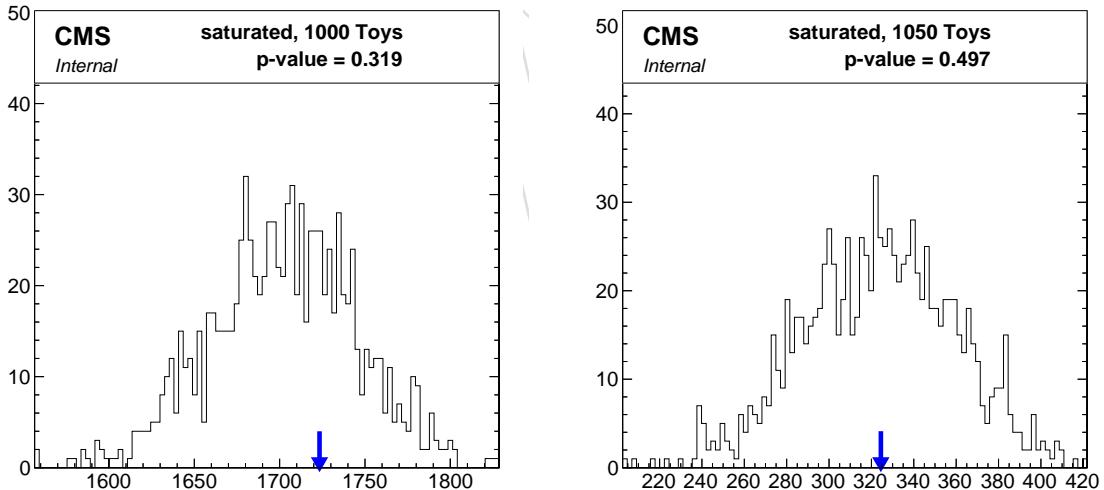


Figure 9.19: The goodness-of-fit tests for the WH channel (left) and the ZH channel (right). The observed value for the test-statistics is indicated by the blue arrows, while the histograms (solid black) indicate the expected distribution obtained in signal-plus-background toys. The observed p-value, i.t. probability of having a goodness-pf-fit value larger than the observed one is 0.319 for WH and 0.497 for ZH.

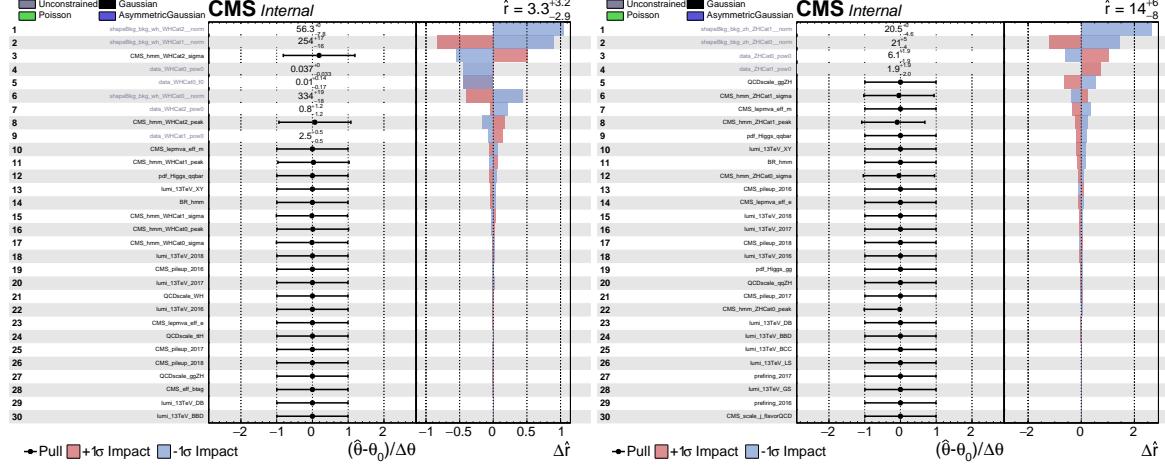


Figure 9.20: The impact study on μ for the combined fit performed in the WH channel (left) and the ZH channel (right).

Table 9.8: Summary of the expect and observed limits and significance in each individual category and the combination.

Category	Expected UL	Observed UL	Expected Signif.	Observed Signif.
WH $\rightarrow 3\ell$ Cat0	27	37.4	0.08	0.89
WH $\rightarrow 3\ell$ Cat1	9.8	12.8	0.22	0.79
WH $\rightarrow 3\ell$ Cat2	7.3	10.4	0.33	0.64
ZH $\rightarrow 4\ell$ Cat0	58	68.2	0.05	0.65
ZH $\rightarrow 4\ell$ Cat1	19	34.0	0.14	1.81
WH combine	5.5	8.9	0.40	1.15
ZH combine	17	31.7	0.15	1.92
WH and ZH combine	5.1	10.2	0.43	1.86

2670 10 Combined result

2671 The binned likelihood function is constructed using data and the MC/analytical predictions for
 2672 signal and background processes. The systematic uncertainties are incorporated into the like-
 2673 lihood function as nuisance parameters ($\vec{\theta}$), with their corresponding probability density func-
 2674 tion. The analytical functional forms contains free parameters that are constrained by solely
 2675 the data during the fit. In each channel, constrained rate uncertainties are model with a log-
 2676 normal distribution. Variation on the MC shapes within $\pm 1\sigma$ are obtained using a fifth order
 2677 polynomial interpolation in each observable bin such that the likelihood function is double-
 2678 differentiable everywhere. On the other hand, extrapolation outside the $\pm 1\sigma$ interval are ob-
 2679 tained with a linear function.

2680 The confidence intervals of the parameters of interest (POIs), $\vec{\alpha}$, are estimated using a profile
 2681 likelihood ratio test statistic $q(\vec{\alpha})$ [81],

$$2682 q(\vec{\alpha}) = -2 \ln \left(\frac{L(\vec{\alpha}, \hat{\vec{\theta}}_{\vec{\alpha}})}{L(\hat{\vec{\alpha}}, \hat{\vec{\theta}})} \right). \quad (10.1)$$

2683 The quantities $\hat{\vec{\alpha}}$ and $\hat{\vec{\theta}}$ denote the unconditional maximum likelihood estimates of the param-
 2684 eter values, while $\hat{\vec{\theta}}_{\vec{\alpha}}$ denotes the conditional maximum likelihood estimate for fixed values
 2685 of the parameters of interest $\vec{\alpha}$. The 1σ , 95%, and 2σ confidence level (CL) intervals for one-
 2686 dimensional measurements of each POI are determined as the interval for which $q(\vec{\alpha}) < 1$,
 2687 $q(\vec{\alpha}) < 3.84$, and $q(\vec{\alpha}) < 4$, respectively, unless otherwise stated.

2688 In order to determine a possible excess in the data, a local p-value (p) is derived from the test
 2689 statistics evaluated at the background only hypothesis (q_0) using the asymptotic properties of
 2690 the likelihood functions (Wilk's theorem, $q_0 \sim \chi^2$). The p-value is converted into a significance,
 2691 Z , defined such that a Gaussian distributed variable found Z standard deviations above its
 2692 mean has an upper-tail probability equal to p :

$$Z = \Phi^{-1}(1 - p), \quad (10.2)$$

2693 where Φ^{-1} is the inverse of the cumulative distribution of a standard Gaussian.

Upper limits on the POIs are derived using the asymptotic approximation [81] and the CLs
 criterion [79, 80], using the test statistics q_μ modified for upper limits:

$$2694 q_\mu = \begin{cases} q & \mu > \hat{\mu} \\ 0 & \mu \leq \hat{\mu} \end{cases} \quad (10.3)$$

2695 The analytical model are constructed as function of the mass of the Higgs boson (m_H). The mass
 2696 is fixed during the fit to the tested value in order to obtain a local significance. The template
 2697 categories use the mass as input to the multivariate discriminators, assuming a 125 GeV Higgs
 2698 boson.

2697 10.1 Combination procedure

2698 Theoretical uncertainties are considered correlated across channel and data-taking periods apart
 2699 from some of those affecting background processes in the VBF-channel. The uncertainty on the
 2700 branching fraction of $H \rightarrow \mu\mu$ is taken from Ref. [12]. The uncertainty on the modelling of

2701 the ggH signal are taken from the “ad-interim” recommendation scheme for inclusive mea-
 2702 surements [70]. A total of nine nuisance parameters are used in this scheme (THU_gg\PH:
 2703 Mu, Res, Mig01, Mig12, VBF2j, VBF3j, PT60, PT120, qmtop) suitable for inclusive mea-
 2704 surements and accounting for migration between the different bins in term of STXS phase
 2705 space [12]. For channels that do not implement this scheme (VBF, ttH, WH, and ZH), the
 2706 corresponding QCD scale uncertainty on the production cross section is correlated with the
 2707 THU_gg\PH_Mu parameter, since correspond to this degree of freedom. A similar scheme has
 2708 been deployed by the LHC HXS WG2 for the VBF processes [71]. A total of ten nuisance pa-
 2709 rameters are added to model the VBF theoretical uncertainties (THU_VBF: PTH200, Mjj60,
 2710 Mjj120, Mjj350, Mjj700, Mjj1000, Mjj1500, PTH25, JET01, TOT). For channels that do
 2711 not implement this scheme (ttH, WH, and ZH) the QCD scale uncertainty on the VBF Higgs
 2712 signal is correlated with the TOT parameter. For WH, ZH, and ttH the QCD scale uncertainties
 2713 reported in Ref. [12] on the production cross section are correlated among channels. PDF un-
 2714 certainties on the Higgs boson production are considered correlated among the channels and
 2715 deployed as four separate nuisances (pdf_Higgs_gg, pdf_Higgs_qqbar, pdf_Higgs_qg,
 2716 pdf_Higgs_tth) accordingly to the main parton in the corresponding production mode.
 2717 Uncertainties arising from the parton shower and underlying event modelling are correlated
 2718 among the ggH, VH, and ttH channels. In contrast, the parton-shower uncertainty affecting
 2719 the VBF-H and VBF-Z predictions in the VBF channel are kept uncorrelated (see description in
 2720 Section 7.5).

2721 The channel impacted most by experimental uncertainties is the VBF one, where some con-
 2722 straints on experimental uncertainties are expected. Uncertainties affecting the integrated lumi-
 2723 nosity measurement are considered correlated among the channels with an year-by-year corre-
 2724 lation scheme recommended by the Lumi-POG [95]. The pileup uncertainties are derived using
 2725 a $\pm 5\%$ variation on minimum-bias cross section in the pileup reweighting procedure [62]. The
 2726 uncertainty is considered correlated among channels, except for VBF, but uncorrelated across
 2727 years. In fact, since the VBF channel is more sensitive to pileup through both signal and back-
 2728 ground processes, the corresponding uncertainties are not correlated with those affecting other
 2729 channels (ggH, WH, ZH, and ttH). The jet-energy-scale (JES) and the jet-energy-resolution
 2730 (JER) uncertainties are implemented through the splitting recommended by the POG based
 2731 on the JES derivation methods (12 nuisances for the JES, some of them correlated across data-
 2732 taking periods) and on the p_T - η -year (18 nuisances for the JER), respectively [53]. Jet-energy-
 2733 scale and resolution uncertainties are considered correlated across all channels. Two channels
 2734 make use of the quark-gluon discriminators (ttH, VBF). The corresponding uncertainty is cor-
 2735 related between them and is also further divided into three independent nuisance parameters
 2736 depending on the parton shower model used to simulate MC events: PYTHIA global-recoil,
 2737 PYTHIA dipole-shower, and HERWIG (angular ordered). The uncertainty on the b-jet identifi-
 2738 cation impacts mostly the ttH and VBF channels, and just minimally WH, ZH, and ggH. A
 2739 single nuisance is used in WH, ZH, and ggH channels, while in ttH and VBF the DeepCSV
 2740 reweighting uncertainties (8 nuisance parameters) are considered, correlated across year and
 2741 processes.

2742 The muon trigger efficiency and identification uncertainties are considered correlated among
 2743 the channels and data-taking periods. The uncertainty on the muon energy scale and reso-
 2744 lution has a dual effect. Firstly, it changes the expected signal and background rates. This
 2745 effect is correlated among the channels except for VBF that further decorrelate it among the
 2746 years. Secondly, it changes the position and resolution of the dimuon-invariant-mass peak in
 2747 ggH, WH, ZH, and ttH categories. The corresponding shape uncertainties are considered un-
 2748 correlated across channels and event categories. The WH, ZH, and ttH channels further use

the lepton-MVA to enhance the purity of the additionally required muons. The corresponding uncertainty is correlated among these channels. Eventually, in 2016 and 2017 data-taking periods, an additional inefficiency was found due to a timing mis-calibration of the ECAL detector, known as pre-firing. Trigger pre-firing occurs when real energy deposits are wrongly assigned to the previous bunch crossing, effectively vetoing the trigger firing on the event. This inefficiency is accounted for in the analysis by applying a set of dedicated corrections provided by the JME-POG. The corresponding uncertainties are correlated across channels and processes but uncorrelated across years.

10.2 Results for $m_H = 125.38 \text{ GeV}$

While the results described in the previous Sections have been derived assuming the Higgs boson mass to be $m_H = 125 \text{ GeV}$, the final results described in the paper/pas are produced for $m_H = 125.38 \text{ GeV}$ that corresponds to the best-fit mass measured by CMS combining Run1 and Run2 (2016) data [98]. A summary of the signal composition, dimuon mass resolution, and purity expressed with the $S/(S + B)$ rate in the event categories included in the analysis is provided in Fig. 10.1. The same information are also reported in Tables 10.1, 10.2, and 10.3 for the ggH, VH, and ttH channels, respectively. In addition, Table 10.4 reports for each bin or group of bins of the DNN output in the VBF-SR the expected number of signal events, the signal composition, the observed number of events in data, the estimated background yields with the associated post-fit uncertainty, $S/(S + B)$ and the S/\sqrt{B} ratios summing the post-fit estimates from each of the three data-taking periods.

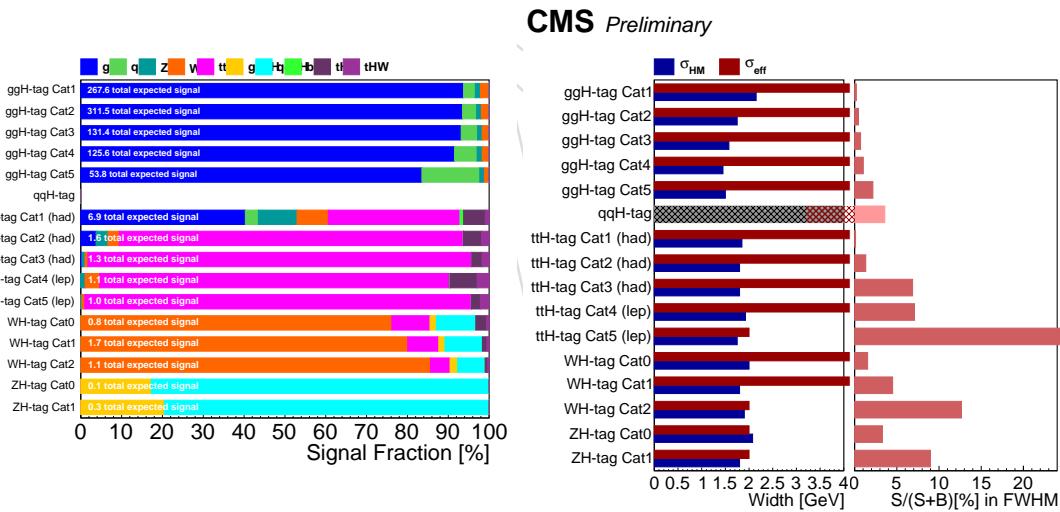


Figure 10.1: On the left, summary of the expected signal composition for $m_H = 125.38 \text{ GeV}$ in each category included in the combined fit. On the right, the dimuon mass resolution (expressed as full-width-half maximum or minimum interval that contains 68% of signal events) and the $S/(S + B)$, computed within the FWHM of the signal peak, of each event category.

A simultaneous fit is performed across all channels and event categories, with a single overall signal strength modifier (μ) with a flat prior. The signal strength modifier is defined as the ratio between the observed Higgs boson rate in the $H \rightarrow \mu\mu$ decay channel and the SM expectation, $\mu = (\sigma\mathcal{B}(H \rightarrow \mu\mu))_{\text{obs}} / (\sigma\mathcal{B}(H \rightarrow \mu\mu))_{\text{SM}}$. The relative contributions from the different Higgs boson production modes are fixed to the SM prediction within uncertainties. A total of 21 event categories included in the combined fit.

Figure 10.2 shows the observed and the predicted distributions of the DNN discriminant in the VBF-SB after applying the VBF event selection. The background prediction is obtained from

Table 10.1: The product of acceptance and selection efficiency for the different signal production processes, the total expected number of signal events with $m_H = 125.38$ GeV, the HWHM of the signal peak, the estimated number of background events and the observation in data within \pm HWHM, and the $S/(S+B)$ and the S/\sqrt{B} ratios within \pm HWHM, for each of the optimized ggH event categories.

Category	Sig.	ggH (%)	VBF (%)	VH + t̄H (%)	HWHM (GeV)	Bkg. in HWHM	$S/(S+B)$ (%) in HWHM	S/\sqrt{B} in HWHM	Data in HWHM
ggH-cat1	267.6	93.7	2.9	3.4	2.12	86362	0.20	0.60	86632
ggH-cat2	311.5	93.5	3.4	3.1	1.75	46349	0.46	0.98	46393
ggH-cat3	131.4	93.2	4.0	2.8	1.60	12656	0.70	0.80	12738
ggH-cat4	125.6	91.5	5.5	3.0	1.47	8260	1.03	0.96	8377
ggH-cat5	53.8	83.5	14.3	2.2	1.50	1678	2.16	0.91	1711

Table 10.2: The product of acceptance and selection efficiency for the different signal production processes, the total expected number of signal events with $m_H = 125.38$ GeV, the HWHM of the signal peak, the estimated number of background events and the observed number of events within \pm HWHM, and the $S/(S+B)$ and the S/\sqrt{B} ratios computed within the HWHM of the signal peak for each of the optimized event categories defined along the WH and ZH BDT outputs.

Category	Sig.	WH (%)	qqZH (%)	ggZH (%)	t̄H + tH (%)	HWHM (GeV)	Bkg. in HWHM	$S/(S+B)$ (%) in HWHM	S/\sqrt{B} in HWHM	Data in HWHM
WH-cat1	0.82	76.2	9.6	1.6	12.6	2.00	32.0	1.54	0.09	34
WH-cat2	1.72	80.1	9.1	1.5	9.3	1.80	23.1	4.50	0.23	27
WH-cat3	1.14	85.7	6.7	1.8	4.8	1.90	5.48	12.6	0.35	4
ZH-cat1	0.11	—	82.8	17.2	—	2.07	2.05	3.29	0.05	4
ZH-cat2	0.31	—	79.6	20.4	—	1.80	2.19	8.98	0.14	4

Table 10.3: The product of acceptance and selection efficiency for the different signal production processes, the total expected number of signal events with $m_H = 125.38$ GeV, the HWHM of the signal peak, the estimated number of background events and the observed number of events within \pm HWHM, and the $S/(S+B)$ and S/\sqrt{B} ratios computed within the HWHM of the signal peak, for each of the optimized event categories defined along the t̄H hadronic and leptonic BDT outputs.

Category	Sig.	t̄H (%)	ggH (%)	VH (%)	tH (%)	VBF+bb̄H (%)	HWHM (GeV)	Bkg. in HWHM	$S/(S+B)$ (%) in HWHM	S/\sqrt{B} in HWHM	Data in HWHM
t̄Hhad-cat1	6.87	32.3	40.3	17.2	6.2	4.0	1.85	4298	1.07	0.07	4251
t̄Hhad-cat2	1.62	84.3	3.8	5.6	6.2	—	1.81	82.0	1.32	0.12	89
t̄Hhad-cat3	1.33	94.0	0.3	1.3	4.2	0.2	1.80	12.3	6.87	0.26	12
t̄Hlep-cat1	1.06	85.8	—	4.7	9.5	—	1.92	9.00	7.09	0.22	13
t̄Hlep-cat2	0.99	94.7	—	1.0	4.3	—	1.75	2.08	24.5	0.47	4

Table 10.4: Event yields in each bin or in group of bins defined along the DNN output in the VBF-SR for various processes. The background yields and the corresponding uncertainties are obtained after performing a combined signal-plus-background fit across analysis regions and data-taking periods. The observed event yields and the expected signal contribution at $m_H = 125.38$ GeV, produced via VBF and ggH modes and assuming SM cross sections and $\mathcal{B}(H \rightarrow \mu\mu)$, are also reported.

DNN bin	Signal	VBF (%)	ggH (%)	Bkg. $\pm \Delta B$	$S/(S+B)$ (%)	S/\sqrt{B}	Data
1–3	19.8	30	70	8893 ± 66	0.22	0.21	8810
4–6	11.8	57	43	395 ± 8	2.90	0.59	392
7–9	8.55	73	27	103 ± 4	7.66	0.84	120
10	2.34	85	15	15.2 ± 1.4	13.2	0.60	18
11	2.18	88	12	9.1 ± 1.2	19.2	0.72	10
12	2.13	87	13	5.8 ± 1.1	26.7	0.88	6
13	1.90	94	6	2.6 ± 0.9	41.8	1.18	7

²⁷⁷⁷ a simultaneous signal-plus-background fit performed across the VBF-SR and VBF-SB regions

as well as data-taking periods. Similarly, Fig. 10.3 shows the distributions of the transformed DNN discriminator in the VBF-SR, obtained after performing the same signal-plus-background fit. The expected distributions for the Higgs boson signal produced via ggH and VBF production, assuming SM production cross sections and branching fraction to a pair of muons, are overlaid. Figure 10.5, Fig. 10.6, and Fig. 10.7 shows the $m_{\mu\mu}$ distributions in each of the ggH, VH, and ttH categories, respectively, along with the result from the combined signal-plus-background fit. Figure 10.4 shows the observed and predicted DNN score distributions in the VBF-SB (left) and VBF-SR (right) regions for the combination of 2016, 2017, and 2018 data. The lower panel shows the ratio between the data and the post-fit background prediction, with the best-fit signal contribution indicated by the blue line in the VBF-SR.

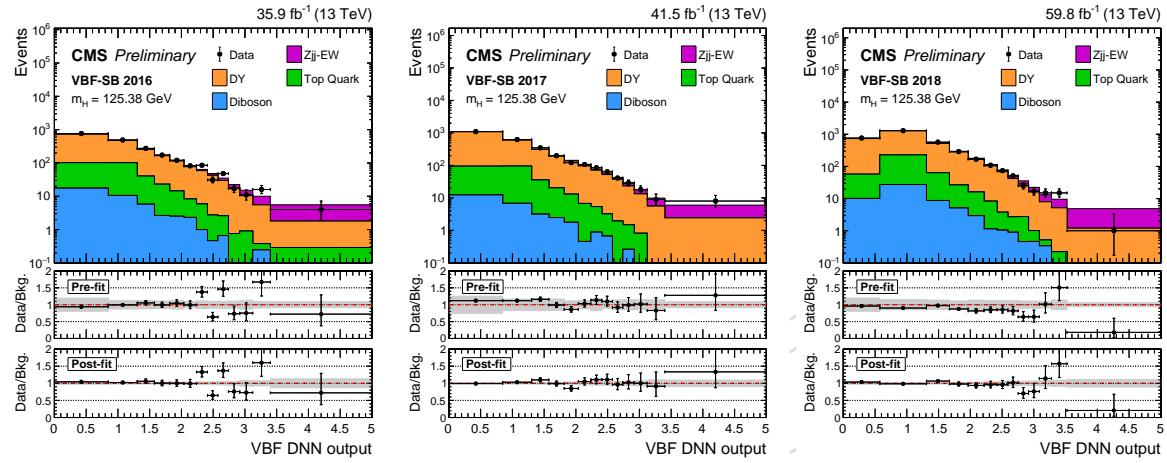


Figure 10.2: The observed DNN output distribution for data collected in 2016 (left), 2017 (middle), and 2018 (right) in the VBF-SB region compared to the post-fit background estimate from SM processes. The predicted backgrounds are obtained from a signal-plus-background fit performed across analysis regions and years. In the second panel, the ratio between data and the pre-fit background prediction is shown. The gray band indicates the total pre-fit uncertainty obtained from the systematic sources previously described. The third panel shows the ratio between data and the post-fit background prediction from the signal-plus-background fit. The gray band indicates the total background uncertainty after performing the fit.

An unbiased mass distribution representative of the fit result in the VBF category is obtained by weighting both simulated and data events from the VBF-SR and VBF-SB regions by the per-event $S/(S+B)$ ratio, computed as a function of the mass-decorrelated DNN output, defined in Section 6, for events within $m_{\mu\mu} = 125.38 \text{ GeV} \pm \text{HWHM}$. The best-fit estimates for the nuisance parameters and signal strength are propagated to the $m_{\mu\mu}$ distribution. Figure 10.8 (left) shows the observed and predicted weighted $m_{\mu\mu}$ distributions for events in the VBF-SB and VBF-SR regions combining 2016, 2017, and 2018 data. The lower panel shows the residuals between the data and the post-fit background prediction, along with the post-fit uncertainty obtained from the b-only fit. The best-fit signal contribution with $m_H = 125.38 \text{ GeV}$ is indicated by the blue line. The signal and background distributions are then interpolated with a spline function in order to obtain a continuous spectrum that can be summed with the analytical fit results in the ggH, WH, ZH, and $t\bar{t}H$ categories. Figure 10.8 (right) shows the $m_{\mu\mu}$ distribution for the weighted combination of all event categories. The ggH, VH, and $t\bar{t}H$ categories are weighted proportionally to the corresponding $S/(S+B)$ ratio, where S and B are the number of expected signal and background events with mass within $\pm \text{HWHM}$ of the expected signal peak with $m_H = 125.38 \text{ GeV}$. The upper panel is dominated by the ggH categories with many data events but relatively small $S/(S+B)$. The lower panel shows the residuals after background subtraction, with the best-fit SM signal contribution with $m_H = 125.38 \text{ GeV}$ indicated

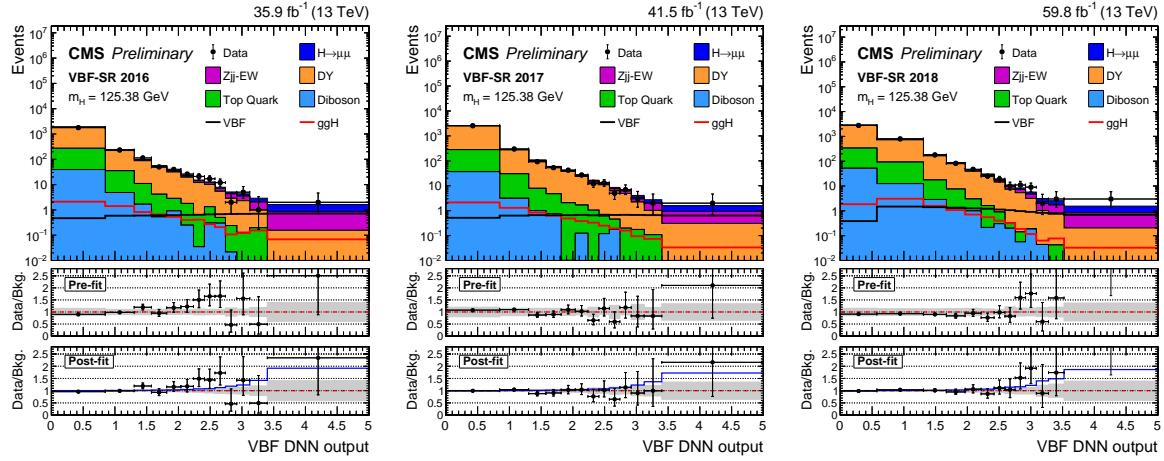


Figure 10.3: The observed DNN output distribution in the VBF-SR region compared to the post-fit background estimate for the contributing SM processes. The expected distributions for the Higgs boson signal produced via ggH and VBF modes with $m_H = 125.38$ GeV are overlaid, assuming SM production cross sections and $\mathcal{B}(H \rightarrow \mu\mu)$. The predicted backgrounds are obtained from a signal-plus-background fit performed across analysis regions and years. The description of the three panels is the same as in Fig. 10.2. The blue histogram (first panel) and solid line (third panel) indicate the total signal extracted from the fit with $m_H = 125.38$ GeV.

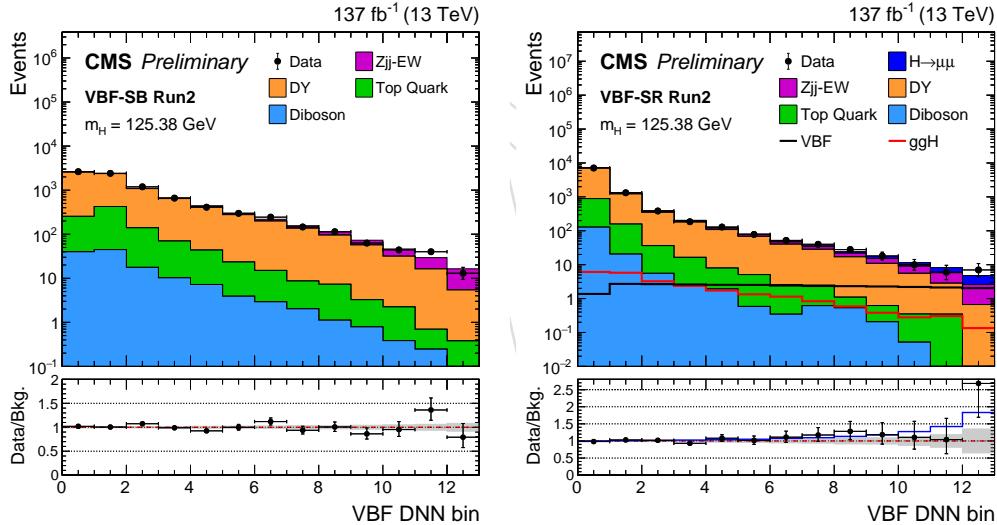


Figure 10.4: The observed DNN output distribution in the VBF-SB (left) and VBF-SR (right) regions for the combination of 2016, 2017, and 2018 data, compared to the post-fit prediction from SM processes. The lower panel shows the ratio between data and the post-fit background prediction from the signal-plus-background fit. The best-fit $H \rightarrow \mu\mu$ signal contribution is indicated by the blue line, and the gray band indicates the total background uncertainty.

2806 by the red line. An excess of events over the background-only expectation is observed near
 2807 $m_{\mu\mu} = 125$ GeV.

2808 The observed (expected for $\mu = 1$) significance at $m_H = 125.38$ GeV of the incompatibility
 2809 with the background-only hypothesis is 2.9 (2.5) standard deviations (s.d.). The 95% CL upper
 2810 limit (UL) on the signal strength, computed with the asymptotic CL_s criterion [79, 81, 99], is
 2811 also derived from the combined fit performed across all channels and event categories. The
 2812 observed (expected for $\mu = 0$) upper limit on μ at 95% CL for $m_H = 125$ GeV is 1.9 (0.8).
 2813 Table 10.5 reports the observed and expected 95% CL upper limits on μ and the significances

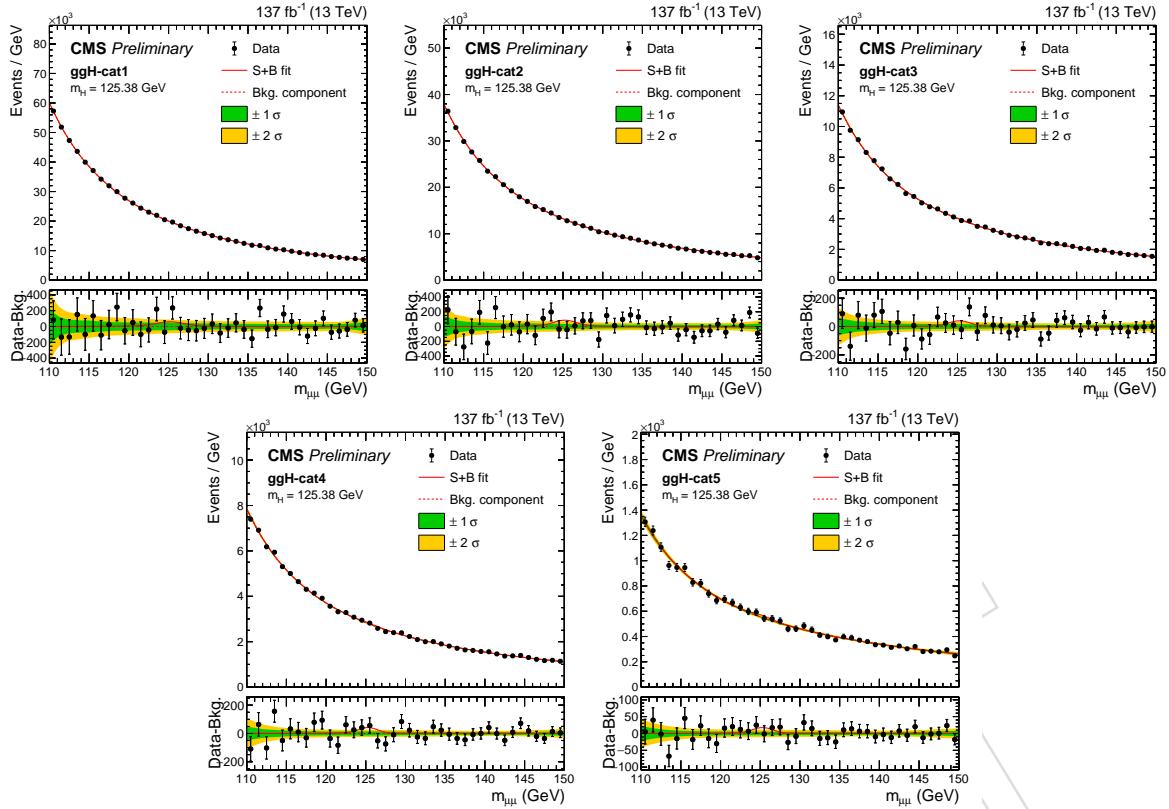


Figure 10.5: Comparison between the observed data and the total background extracted from a signal-plus-background fit performed across the ggH categories. First row, from left to right: ggH-cat1, ggH-cat2, and ggH-cat3. Second row, from left to right: ggH-cat4 and ggH-cat5. The one (green) and two (yellow) standard deviation bands include the uncertainties in the background component of the fit. The lower panel shows the residuals after background subtraction and the red line indicates the signal with $m_H = 125.38 \text{ GeV}$ extracted from the fit.

over the background-only expectation at $m_H = 125 \text{ GeV}$ in each channel as well as for the combined fit. The same quantities are also reported in Fig. 10.9: upper-limits on the left and significances on the right.

Table 10.5: Observed and expected (in absence of $H \rightarrow \mu\mu$ decays) 95% CL upper limits on μ for $m_H = 125.38 \text{ GeV}$ and the corresponding significances for the incompatibility with the background-only hypothesis, for each production category as well as for the combined fit.

Production category	Observed (expected)	Signif.	Observed (expected) UL on μ
VBF	2.35 (1.77)		2.53 (1.22)
ggH	0.99 (1.56)		1.77 (1.28)
t̄tH	1.20 (0.54)		6.48 (4.20)
VH	2.02 (0.43)		10.8 (5.13)
Combination	2.92 (2.46)		1.92 (0.82)

Profile likelihood ratios as a function of μ , for both the combined fit and each individual search channel, are reported in Fig. 10.10 (left) for $m_H = 125.38 \text{ GeV}$. Results from the combined fit are shown for both data and an Asimov pseudodata set [81] in which a signal with strength $\mu = 1$ is injected, while the background and constrained nuisance parameters are fixed to the maximum-likelihood estimate obtained from a background-only fit to the data. The best-fit signal strength for the Higgs boson with mass of 125.38 GeV , and the corresponding 68% CL interval, is $\hat{\mu} = 1.18^{+0.41}_{-0.39} (\text{stat})^{+0.17}_{-0.15} (\text{syst})$. The statistical component of the post-fit uncertainty

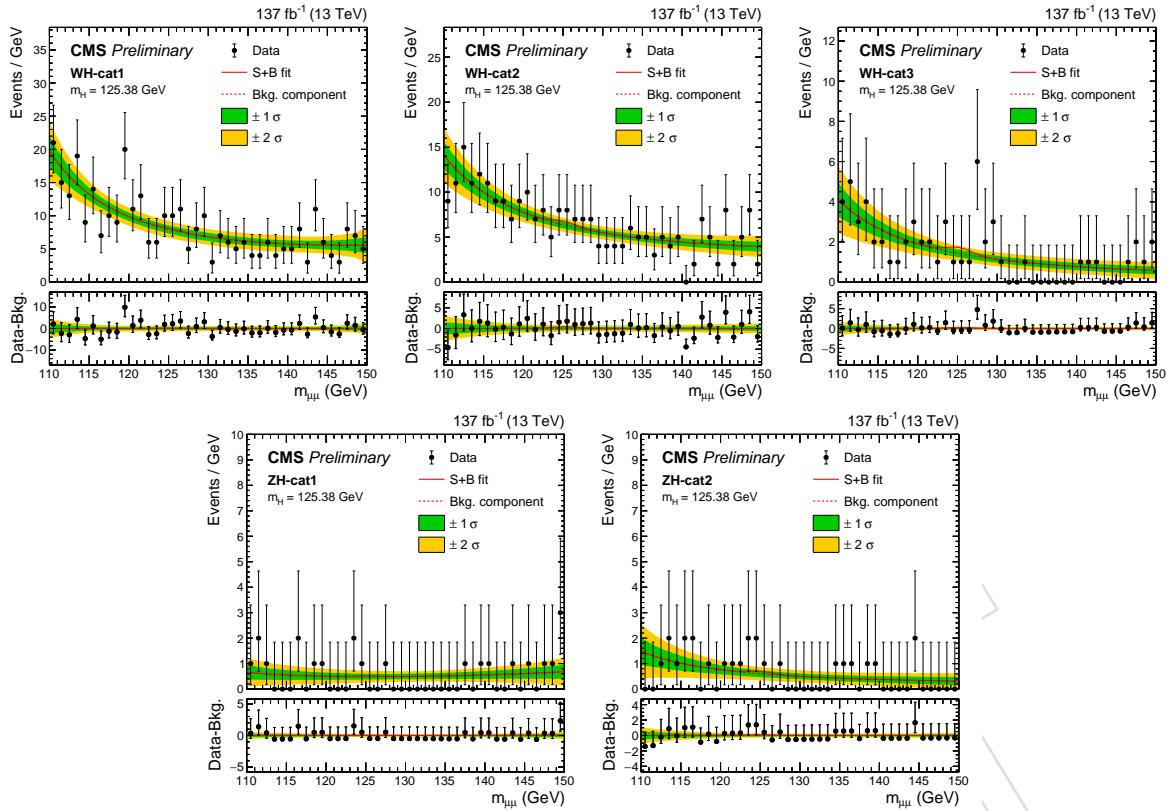


Figure 10.6: Comparison between the observed data and the total background extracted from a signal-plus-background fit performed across WH (first row) and ZH (second row) event categories. First row, from left to right: WH-cat1, WH-cat2, and WH-cat3. Second row, from left to right: ZH-cat1 and ZH-cat2. The one (green) and two (yellow) standard deviation bands include the uncertainties in the background component of the fit. The lower panel shows the residuals after the background subtraction, where the red line indicates the signal with $m_H = 125.38$ GeV extracted from the fit.

is separated by performing a likelihood scan as a function of μ in which systematic uncertainties are removed. The systematic uncertainty component is then taken as the difference in quadrature between the total and the statistical uncertainties. A comparison between the statistical-only and the full observed likelihood scans as a function of μ is given in Fig. 10.10 (right). The systematic uncertainties affecting the signal (in each category) and the background (in the VBF channel) prediction have a limited impact on the ultimate analysis sensitivity. In fact, the statistical uncertainty is about 2.5 times larger than the systematic one.

Figure 10.11 (left) shows the impact of both constrained (systematic uncertainties) and unconstrained (freely floating background parameters) parameters on the fitted signal strength. Impacts are derived following the same procedure detailed in Section 6.5. None of the constrained nuisance parameters are significantly pulled or constrained by the signal-plus-background fit. The parameters with highest impact on the analysis performance are the freely floating shape and rate parameters affecting the background prediction in ggH categories. They are followed by parton shower uncertainties affecting VBF-H and VBF-Z predictions in the VBF channel, DY rate uncertainties in VBF categories, some of the sources affecting the knowledge of the jet energy scale, and the theoretical uncertainty on the inclusive ggH production cross section. Moreover, Fig. 10.11 shows the results of a goodness-of-fit test for the combined fit in which the saturated likelihood ratio is used as test-statistics. The expected distribution for the test-statistics is obtained by performing signal-plus-background fits, with floating μ , on toys thrown from the best-fit estimate of the nuisance parameters a-posteriori toys). The observed

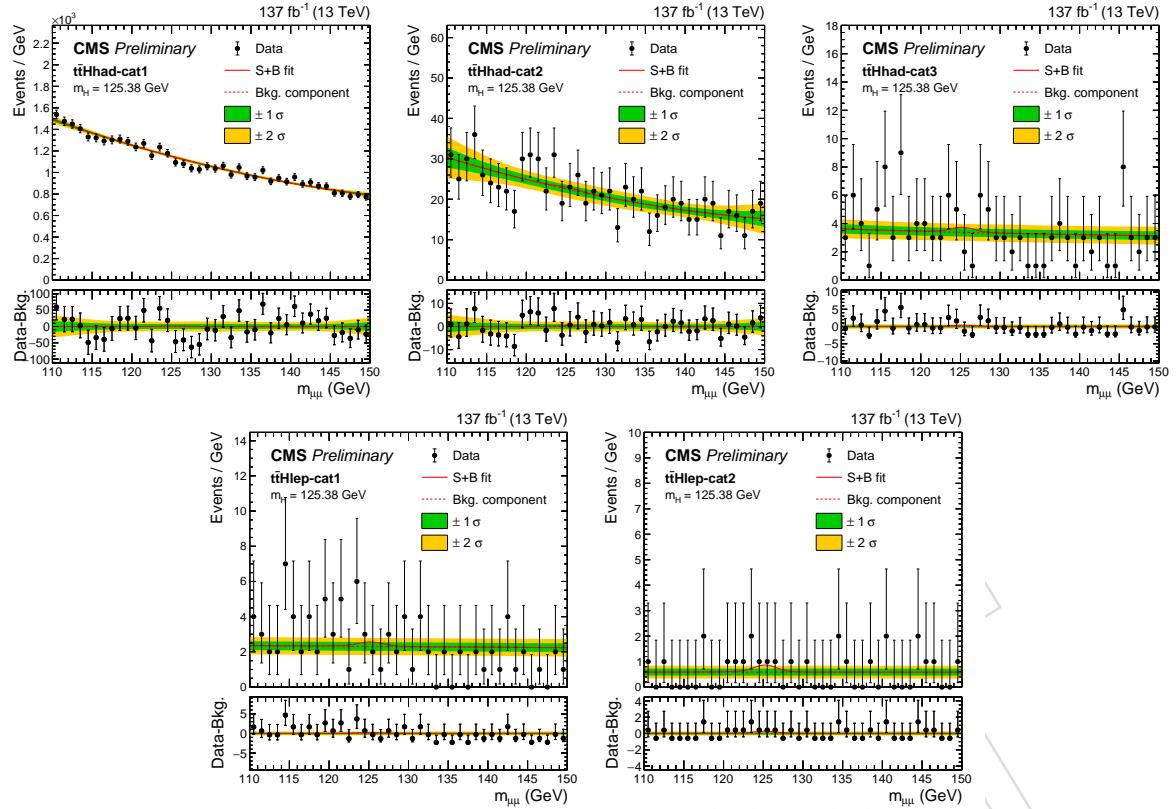


Figure 10.7: Comparison between the observed data and the total background extracted from a signal-plus-background fit performed across $t\bar{t}H$ hadronic (first row) and leptonic (second row) event categories. First row, from left to right: $t\bar{t}H$ had-cat1, $t\bar{t}H$ had-cat2, and $t\bar{t}H$ had-cat3. Second row, from left to right: $t\bar{t}H$ lep-cat1 and $t\bar{t}H$ lep-cat2. The one (green) and two (yellow) standard deviation bands include the uncertainties in the background component of the fit. The lower panel shows the residuals after the background subtraction, where the red line indicates the signal with $m_H = 125.38$ GeV extracted from the fit.

2844 p-value, i.e. probability of having a goodness-of-fit value larger than the observed one, is 0.22.
 2845 This indicates a good compatibility between the observed data and the signal-plus-background
 2846 model. Figure 10.12 (left) reports the outcome of the channel compatibility test which consists
 2847 in a χ^2 -like probability extracted from the best-fit values obtained in each category involved
 2848 in the final fit. The test is performed as follows: a signal strength modifier is defined in each
 2849 event category except for the VBF-SB regions, then a global fit is performed in order to estimate
 2850 the signal strengths per-category and the corresponding post-fit uncertainties (68% CL), finally
 2851 a χ^2 -like test-statistics is computed from the measured strengths. The reduced χ^2 is of about
 2852 0.6 indicating a good compatibility between the observations from signal-plus-background fits
 2853 performed in single channels. Figure 10.12 (right) reports a summary of the best-fit values for
 2854 the signal strength and the corresponding 68% CL intervals obtained from the profile likelihood
 2855 scan in each channel.

2856 An additional likelihood scan is performed in which the four main Higgs boson production
 2857 mechanisms are associated to either fermion (ggH and $t\bar{t}H$) or vector boson (VBF and VH)
 2858 couplings. Two signal strength modifiers, denoted as $\mu_{ggH,t\bar{t}H}$ and $\mu_{VBF,VH}$, are varied in-
 2859 dependently as unconstrained parameters in the fit. Figure 10.13 (left) shows the 1σ and 2σ
 2860 contours, computed as variations around the likelihood maximum for $m_H = 125.38$ GeV, for
 2861 the signal strength modifiers $\mu_{ggH,t\bar{t}H}$ and $\mu_{VBF,VH}$. The best-fit values for these parameters
 2862 are $\hat{\mu}_{ggH,t\bar{t}H} = 0.66^{+0.67}_{-0.66}$ and $\hat{\mu}_{VBF,VH} = 1.80^{+0.89}_{-0.77}$. In addition, Fig.10.13 (right) shows a two-

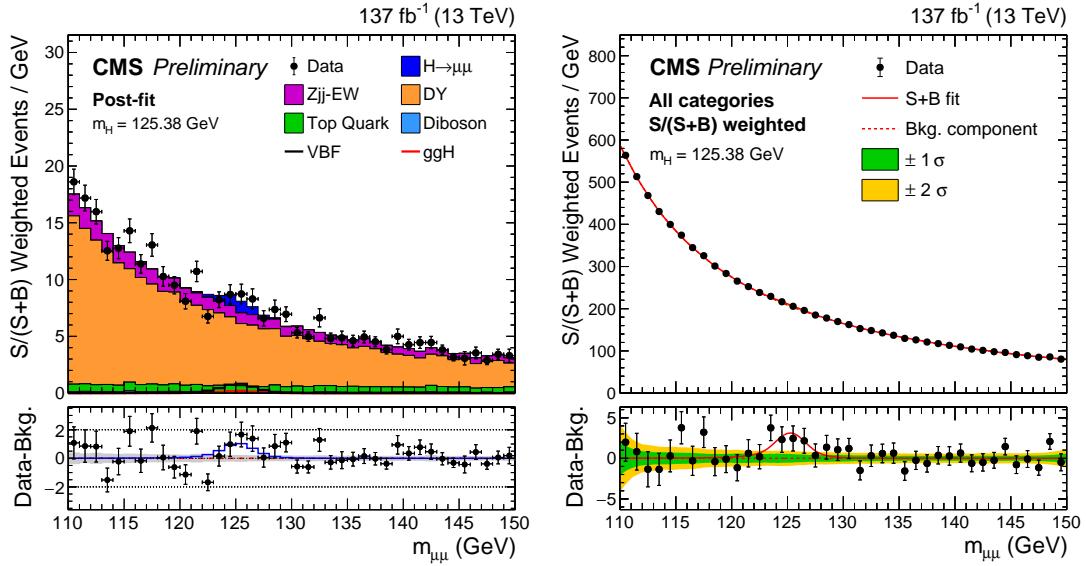


Figure 10.8: Left: the $m_{\mu\mu}$ distribution for the weighted combination of VBF-SB and VBF-SR events. Each event is weighted proportionally to the $S/(S + B)$ calculated as a function of the mass-decorrelated DNN output. The lower panel shows the residuals after subtracting the background prediction from the signal-plus-background fit. The best-fit $H \rightarrow \mu\mu$ signal contribution is indicated by the blue line, and the gray band indicates the total background uncertainty from the b-only fit. Right: the $m_{\mu\mu}$ distribution for the weighted combination of all event categories. The upper panel is dominated by the ggH categories with many data events but relatively small $S/(S + B)$. The lower panel shows the residuals after background subtraction, with the best-fit SM $H \rightarrow \mu\mu$ signal contribution with $m_H = 125.38$ GeV indicated by the red line.

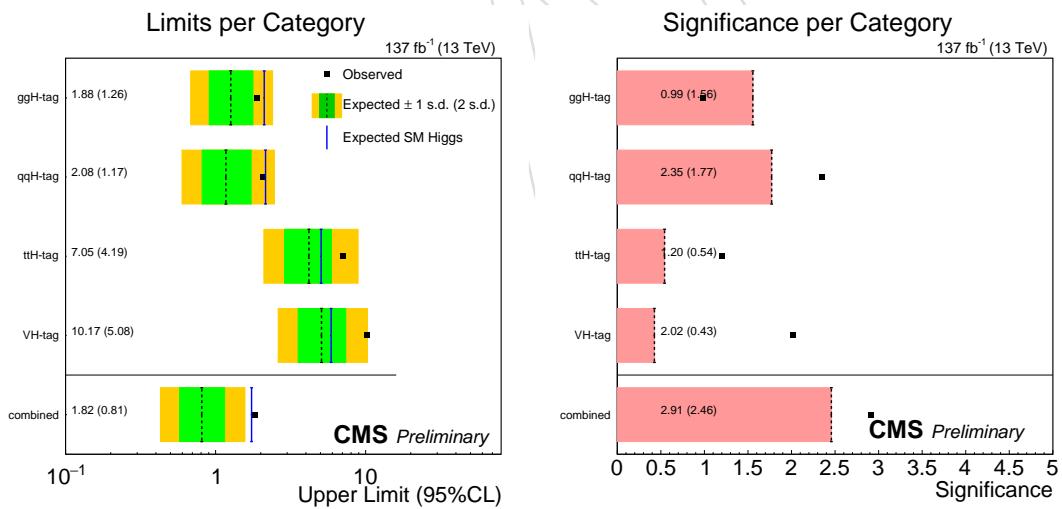


Figure 10.9: On the left, 95% CL asymptotic limits per tag and combined for $m_H = 125.38$ GeV. On the right, expected and observed significances for $m_H = 125.38$ GeV.

dimensional likelihood ratio scan as a function of the coupling strength modifiers κ_F and κ_V for $m_H = 125.38$ GeV [100]. The best-fit estimates for the coupling modifiers are $\hat{\kappa}_F = 0.97^{+0.35}_{-0.37}$ and $\hat{\kappa}_V = 1.74^{+\inf}_{-0.77}$.

The local p-value quantifies the probability for the background to produce a fluctuation larger than the apparent signal observed in the search region. Figure 10.14 (left) shows the observed local p-value for the combined fit and for each individual production category as a function

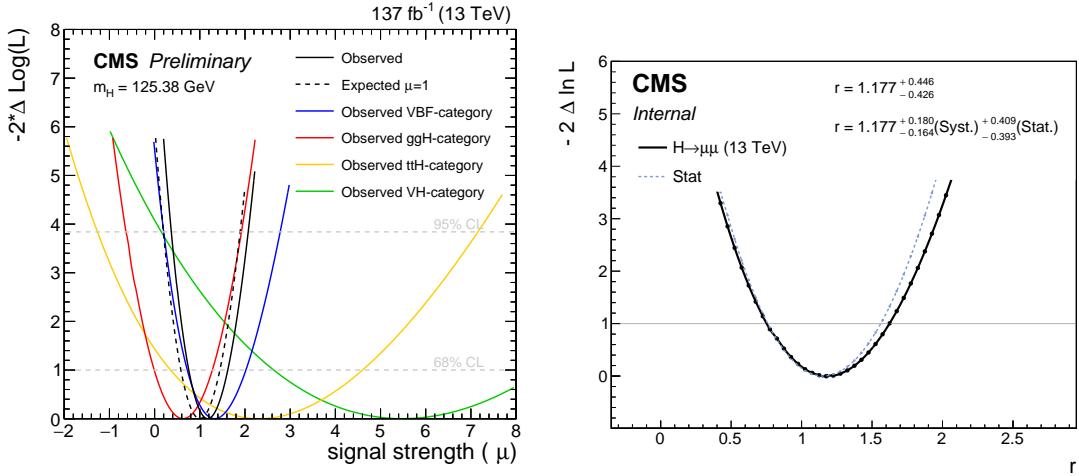


Figure 10.10: On the left, profile likelihood ratios as a function of μ for $m_H = 125.38 \text{ GeV}$, where the solid curves represent the observation in data and the dashed line represents the expected result from the combined signal-plus-background fit. The observed likelihood scans are reported for the full combination (black), as well as for the individual ggH, VBF, VH, and ttH channels. On the right, full likelihood scan as a function of μ for $m_H = 125.38 \text{ GeV}$ (black) and statistical-only one (dashed blue) used to separate the statistical and the systematic components of the post-fit uncertainty on μ .

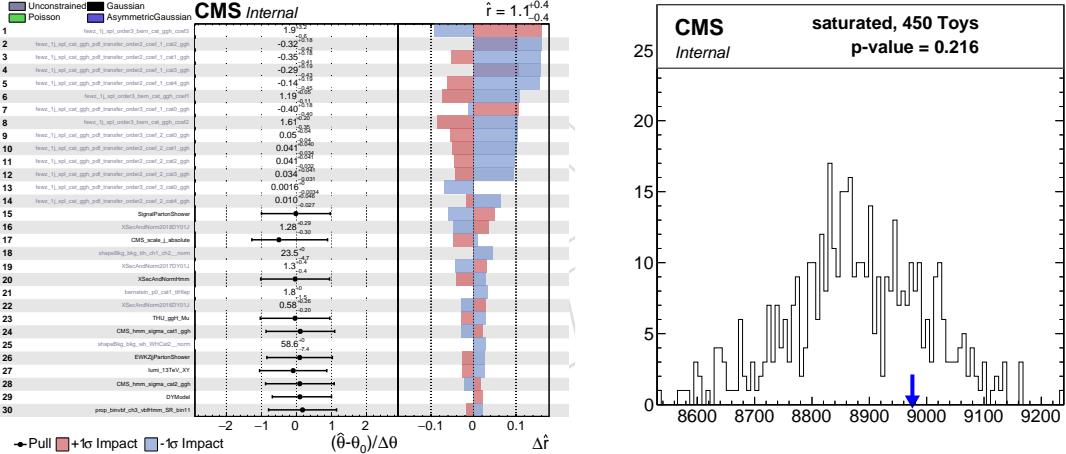


Figure 10.11: On the left, result of the impact study on μ for the combined fit performed across all event categories to the observed data. On the left, results from a goodness-of-fit test for the combined fit. The observed value for the test-statistics is indicated by the blue arrow, while the histogram (solid black) indicates the expected distribution obtained in signal-plus-background toys. The observed p-value, i.e. probability of having a goodness-of-fit value larger than the observed one, is 0.22.

of m_H in a 5 GeV window around the expected Higgs boson mass. Figure 10.14 (right) shows the expected p-values computed for the combined fit and for each production category computed on an Asimov data set [81] generated from the background expectation obtained from the signal-plus-background fit injecting a signal at $m_H = 125.38 \text{ GeV}$. The solid markers indicates the mass points for which expected and observed p-values are computed. In the ggH, VH, and ttH category, in order to evaluate p-values for masses different than 125 GeV, signal models are derived with alternative $H \rightarrow \mu\mu$ signal samples generated with m_H fixed to 120 and 130 GeV. Signal shape parameters and the expected rate for each production mode in each event category are then interpolated within $120 < m_H < 130 \text{ GeV}$, providing a signal model for any mass value in the $m_H = 125 \pm 5 \text{ GeV}$ range. A different strategy is employed

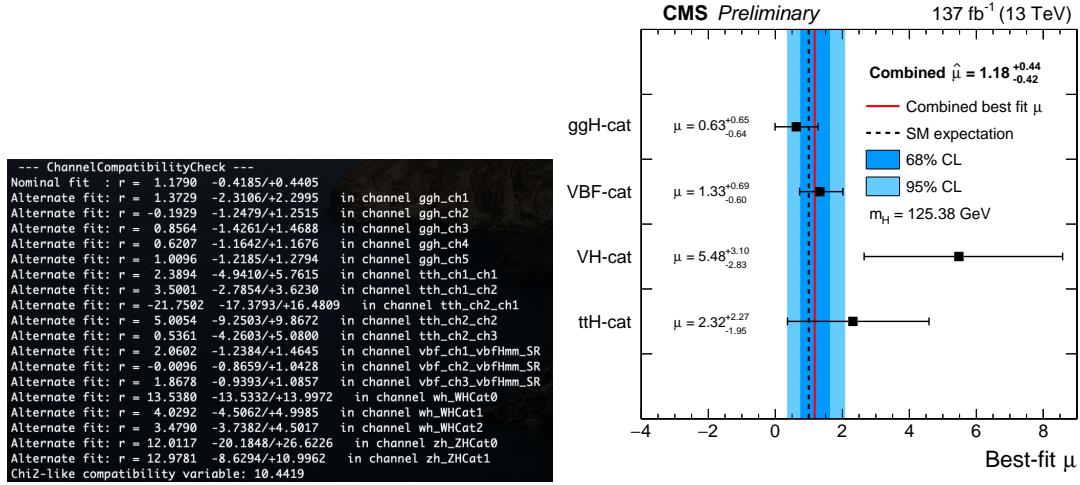


Figure 10.12: On the left, outcome from the channel compatibility test performed on top of the combined fit by defining independent signal strengths in each channel except for VBF-SB regions. On the right, signal strength modifiers measured for $m_H = 125$ GeV in each channel (black points) are compared to the result of the combined fit (solid red line) and the SM expectation (dashed gray line).

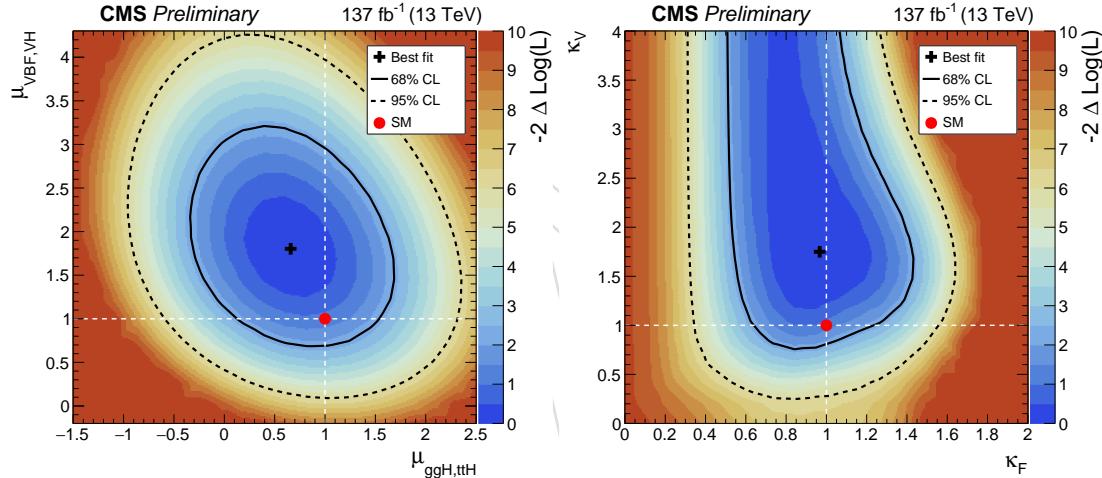


Figure 10.13: On the left, scan of the profiled likelihood ratio as a function of $\mu_{ggH,ttH}$ and $\mu_{VBF,VH}$ with the corresponding 1σ and 2σ uncertainty contours. The black cross indicates the best-fit values ($\hat{\mu}_{ggH,ttH}, \hat{\mu}_{VBF,VH}$) = (0.66, 1.80) while the red circle represents the SM expectation. On the right, scan of the profiled likelihood ratio as a function of the coupling strength modifiers κ_f and κ_V along with their 1σ and 2σ uncertainty contours. The black cross indicates the best-fit values ($\hat{\kappa}_f, \hat{\kappa}_V$) = (0.97, 1.74) while the red circle represents the SM expectation.

in the VBF category since $m_{\mu\mu}$ is a DNN input variable. As described in Section 7, the DNN output can be decorrelated from the dimuon mass information by fixing the value of $m_{\mu\mu}$ to 125 GeV. Therefore, a potential signal with mass m' different from 125 GeV can be extracted by fitting the data with an alternative set of signal and background templates, obtained by shifting the mass value used as input to the DNN evaluation by $\Delta m = 125 \text{ GeV} - m'$ and adjusting the expected signal yields by the corresponding differences in the production cross section, decay rate, and acceptance. This procedure is also applied to the data, yielding for each tested mass hypothesis a different observed DNN distribution to fit. Throughout the explored mass range, $120 < m_H < 130$ GeV, the VBF category has the highest expected sensitivity to $H \rightarrow \mu\mu$ decays, followed by the ggH, ttH, and VH categories, respectively. The observed (expected for $\mu = 1$) significance at $m_H = 125.38$ GeV of the incompatibility with the background-only hypothesis

is 2.9 (2.5) σ . Fluctuations in the observed p-value of the VBF category and for the combined fit are due to the nature of the signal extraction fit used in the VBF analysis. When evaluating the DNN for each tested mass point, event migrations in data between neighbouring bins in the high score DNN region can produce discrete variations in the observed significance. In contrast, the smooth behaviour of the observed p-value in the ggH, VH, and ttH categories derives from the stronger correlation between mass points caused by the limited $m_{\mu\mu}$ resolution. Table 10.6 reports the evolution of the observed p-value and best-fit signal strength as a function of the Higgs mass hypothesis as obtained from the combined fit performed across all event categories.

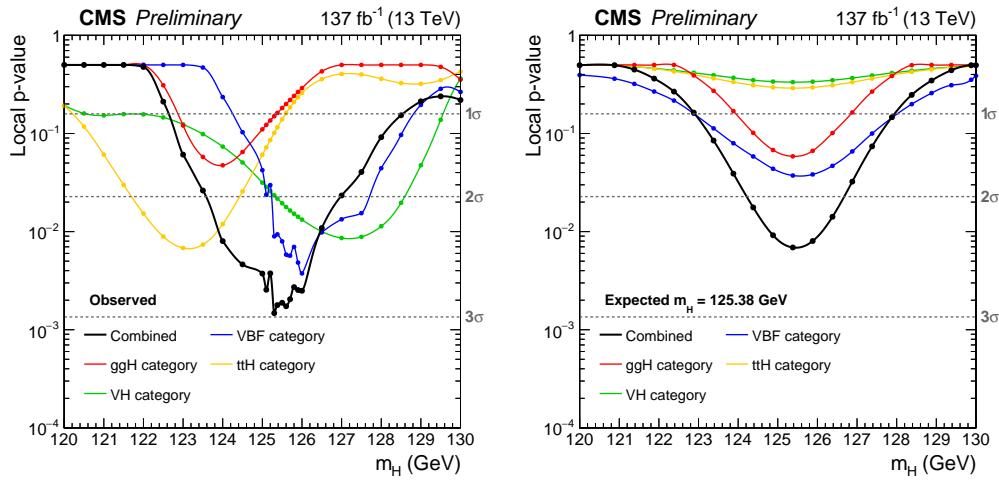


Figure 10.14: Left: observed local p-values as a function of m_H , extracted from the combined fit as well as from each individual production category, are shown. Right: the expected p-values are calculated using the background expectation obtained from the signal-plus-background fit and injecting a signal with $m_H = 125.38$ GeV and $\mu = 1$.

10.3 Combination between Run1 and Run2 results

The combination with the results obtained in Run I (7 and 8 TeV) [101] is performed following the strategy already defined in Ref. [102]. The yields in each category are scaled to match the improvements in the theoretical predictions from the YR3 to the YR4 [12]. The theoretical uncertainties are conservatively unchanged and taken as fully correlated between Run2 and Run2. The experimental uncertainties are conservatively taken as uncorrelated. The observed (expected) combined significance for $m_H = 125.38$ GeV is 2.94 (2.49) standard deviations, hence the combination with Run1 results improves the 13 TeV measurement by about 1%.

10.4 Measurement of Higgs boson coupling to muons

The results presented in this note are the most precise measurements in the $H \rightarrow \mu\mu$ decay channel reported to date, and can be used to improve constraints on the coupling between the Higgs boson and fermions of the second generation. The signal strength measured in the $H \rightarrow \mu\mu$ analysis cannot be translated directly into a measurement of the Higgs boson coupling to muons because it is also sensitive to the interactions between the Higgs boson and several SM particles involved in the considered production processes, primarily the top quark and vector boson couplings. These Higgs boson couplings to other particles are constrained by combining the result of this analysis with those presented in Ref. [103], based on pp collision data recorded by the CMS experiment at $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of up to 137 fb^{-1} . Under the assumption that there are no BSM particles contributing

Table 10.6: Observed p-value and best-fit signal strength obtained from the combined fit as a function of m_H in the range $123 < m_H < 127$ GeV. In this range, Higgs boson mass hypothesis are tested with a step size of 0.5 GeV. Values reported for $m_H = 125.09$ GeV and $m_H = 125.38$ GeV are obtained through a smooth spline interpolation.

m_H	Observed p-value	Observed signif.	Best-fit $\mu \pm \Delta\mu$
123.0	0.0609028	1.547	$0.633^{+0.430}_{-0.409}$
123.5	0.0262234	1.939	$0.814^{+0.444}_{-0.429}$
124.0	0.0080025	2.409	$1.007^{+0.450}_{-0.427}$
124.5	0.0046340	2.602	$1.072^{+0.449}_{-0.427}$
125.0	0.0037353	2.675	$1.101^{+0.445}_{-0.430}$
125.1	0.0025561	2.800	$1.140^{+0.450}_{-0.419}$
125.2	0.0037554	2.673	$1.087^{+0.443}_{-0.416}$
125.3	0.0014760	2.973	$1.195^{+0.442}_{-0.418}$
125.38	0.0017833	2.915	$1.177^{+0.440}_{-0.418}$
125.4	0.0017828	2.914	$1.176^{+0.443}_{-0.417}$
125.5	0.0018830	2.897	$1.159^{+0.436}_{-0.429}$
125.6	0.0017351	2.923	$1.171^{+0.437}_{-0.419}$
125.7	0.0020479	2.871	$1.163^{+0.443}_{-0.428}$
125.8	0.0027200	2.780	$1.113^{+0.438}_{-0.415}$
125.9	0.0025399	2.802	$1.135^{+0.439}_{-0.423}$
126.0	0.0024945	2.808	$1.130^{+0.441}_{-0.419}$
126.5	0.0108817	2.294	$0.941^{+0.439}_{-0.420}$
127.0	0.0233476	1.989	$0.811^{+0.437}_{-0.415}$

to the Higgs boson total width, Higgs boson production and decay rates in each category are expressed in terms of coupling modifiers within the so-called κ -framework. Six free coupling parameters are introduced in the likelihood (κ_W , κ_Z , κ_t , κ_τ , κ_b , and κ_μ) and are extracted from a simultaneous fit across all categories. In the combined fit, the event categories of the $H \rightarrow \mu\mu$ analysis described in this document supersede those considered in Ref. [103]. Figure 10.15 (left) shows the observed profile likelihood ratio as a function of κ_μ for $m_H = 125.38$ GeV. The best-fit value for κ_μ , as well as those for the other couplings, are compatible with the SM prediction. The corresponding 95% CL interval for the κ_μ parameter is $0.62 < \kappa_\mu < 1.52$. An additional fit is performed using a phenomenological parametrization relating the masses of the fermions and vector bosons to the corresponding coupling modifiers using two parameters, denoted M and ε [104]. The fermion and the weak boson coupling modifiers are expressed as $\kappa_F = v m_F^\varepsilon / M^{1+\varepsilon}$ and $\kappa_V = v m_V^{2\varepsilon} / M^{1+2\varepsilon}$, where v is the SM Higgs boson vacuum expectation value of 246.22 GeV. The SM expectation, corresponding to coupling modifiers equal to unity, is obtained when $M = v$ and $\varepsilon = 0$. The lepton, vector boson, and quark masses are taken from Ref. [105]. Figure 10.15 (right) shows the best-fit estimates for the six coupling modifiers as a function of particle mass as well as the the result of the (M, ε) fit. The result is consistent with the SM prediction within 1σ over the entire mass spectrum.

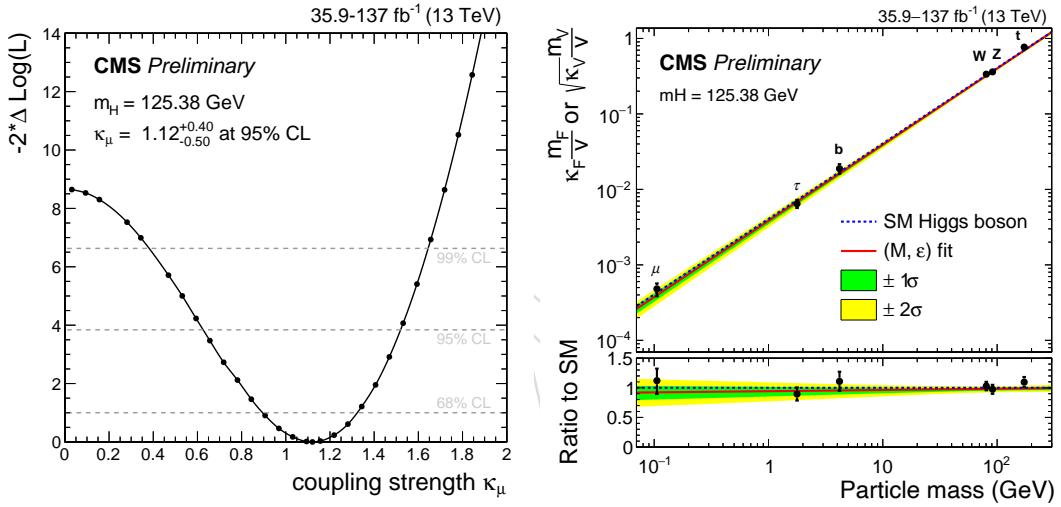


Figure 10.15: Left: observed profile likelihood ratio as a function of κ_μ for $m_H = 125.38 \text{ GeV}$, obtained from a combined fit with Ref. [6] in the κ -framework model. The best-fit value for κ_μ is 1.12 and the corresponding observed 95% CL interval is $0.62 < \kappa_\mu < 1.52$. Right: the best-fit estimates for the coupling modifiers extracted from the resolved κ -framework model overlayed to the result of the phenomenological (M, ε) fit.

2935 A VBF “data-driven” analysis

2936 The most important discriminating variable between signal and background processes in the
 2937 $H \rightarrow \mu\mu$ search is the dimuon invariant mass, which forms a sharp peak around 125 GeV for
 2938 the signal as opposed to the background in which it decreases monotonically within the same
 2939 mass range. This feature also allows for a completely data-driven estimation of the background
 2940 through an analytical fit to the dimuon mass spectrum. As previously mentioned, the VBF sig-
 2941 nals offers a distinctive signature involving two jets with a large η -gap and large m_{jj} . Therefore,
 2942 in order to enhance the signal-to-background separation, we construct a multivariate discrim-
 2943 inant using boosted decision trees, which take as input several variables that capture the dis-
 2944 tinctive kinematic features of the signal and the background. This BDT is trained exclusively
 2945 on events containing 2 or more jets with m_{jj} larger than 400 GeV (VBF BDT). In terms of
 2946 signal, this category contains a significant contribution from VBF events (about 55% of the total
 2947 expected signal). In the case of the background, these events are still largely dominated by the
 2948 Drell-Yan process. However, the contribution from $t\bar{t}$ and electroweak $Z+jets$ production starts
 2949 to become significant, particularly in events with a VBF-like topology. The input variables have
 2950 been specifically chosen so that the BDT output remains highly uncorrelated with the dimuon
 2951 mass. We then categorize events based on BDT score in order to maximize the search sensi-
 2952 tivity. Since the BDT output is highly uncorrelated with the dimuon mass, this categorization
 2953 does not sculpt the background to peak around the Higgs boson mass. This in turn allows us to
 2954 fit the dimuon mass distribution in each of the categories without a significant bias. Finally, we
 2955 perform a statistical combination of all the categories to extract a final estimate of the analysis
 2956 performance.

2957 A.1 Input variables for the VBF BDT

2958 The BDT discriminant used for events with 2 or more jets with m_{jj} larger than 400 GeV takes
 2959 as input the same variables used in the ggH MVA described in Section 6.1.1. However, ad-
 2960 ditional observables are included to further increase the separation between VBF and ggH
 2961 signals. These are represented by the p_T -balance ratio (R_{p_T}), the p_T -centrality, the number of
 2962 soft-jets with p_T larger than 2 GeV (N_{soft}), and the related H_T (H_T^{soft}). The R_{p_T} and p_T -centrality
 2963 are defined in Equation A.1. A soft-jet, instead, is defined as an anti- k_t jet, with a distance pa-
 2964 rameter R of 0.4 and p_T larger than 2 GeV, obtained by clustering only charged PF candidates
 2965 assigned from the primary vertex. The two identified muons and the charged PF candidates
 2966 associated to the two leading jets in the event are not included in the soft-jet definition.

$$R_{p_T} = \frac{|\vec{p}_T^{\mu_1} + \vec{p}_T^{\mu_2} + \vec{p}_T^{j_1} + \vec{p}_T^{j_2}|}{p_T^{\mu_1} + p_T^{\mu_2} + p_T^{j_1} + p_T^{j_2}}, \quad p_T - \text{centrality} = \frac{p_T^{\mu\mu} - |\vec{p}_T^{j_1} + \vec{p}_T^{j_2}|/2}{|\vec{p}_T^{j_1} - \vec{p}_T^{j_2}|} \quad (\text{A.1})$$

2967 Figure A.1 shows the shapes in signal and background events of dijet and soft-jet related vari-
 2968 ables, as well as angular distances and p_T ratios between the dimuon system and jets which
 2969 are used as inputs to the VBF BDT trained on events with m_{jj} larger than 400 GeV. Both the
 2970 shapes and the separation power of muon observables are similar to what was shown for the
 2971 ggH BDT in Fig. 6.1.

2972 Figure A.2 and A.3 show the distributions of the VBF BDT inputs in data and simulation for
 2973 events in the signal fit region after requiring two or more jets in the event with m_{jj} larger than
 2974 400 GeV, $\Delta\eta_{jj} > 2.5$, and leading jet p_T greater than 35 GeV. The observation and the expectation
 2975 in the three eras (2016, 2017, and 2018) have been summed together. Overall, the agreement be-
 2976 tween data and MC for muon and jet variables is good, but there are some visible discrepancies

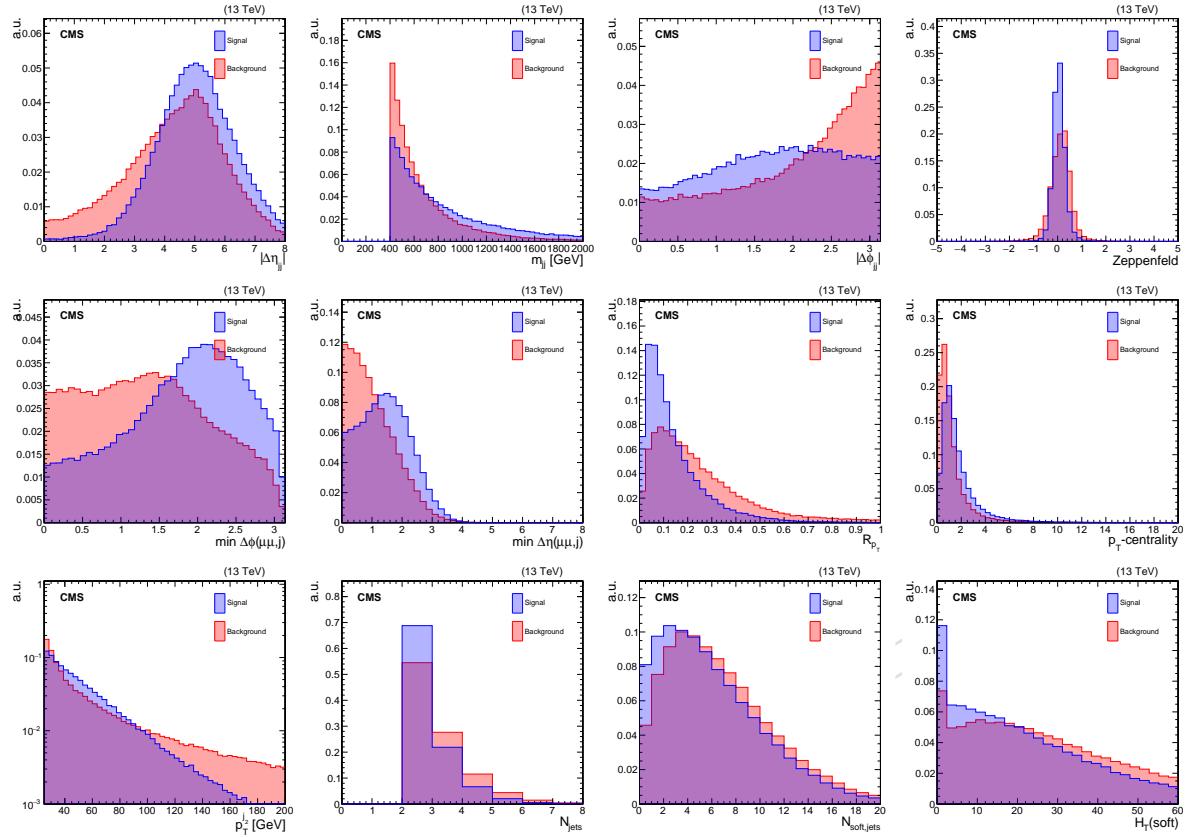


Figure A.1: Distribution of the jet related variables used as inputs in the VBF BDT trained on events with m_{jj} larger than 400 GeV for signal (blue) and background (red) processes. Distributions are scaled in order to have unit area. From left to right on the top row: $\Delta\eta_{jj}$, m_{jj} , $\Delta\phi_{jj}$, and Zeppenfeld variable. From left to right on the middle row: $\min\Delta\phi(\mu\mu, j_1j_2)$, $\min\Delta\eta(\mu\mu, j_1j_2)$, R_{p_T} , and p_T – centrality. From left to right on the bottom row: second jet p_T , number of jets, number of soft-jets with p_T larger than 2 GeV, and lastly their H_T .

in the leading and subleading jet η distributions, in the $\Delta\phi_{jj}$, in the N_{jets} and soft-jet variables. Comparisons are also reported independently for each era in the Appendix of Ref [17], where a more detailed description of the observed trends in the main kinematic distributions is reported.

A.2 BDT discriminant trainings

The training is performed separately on the MC samples for the three eras: 2016, 2017, and 2018. This helps to capture the impact of the differences in reconstruction and detector performance across the three data taking periods. For example, the occupancy and properties of forward jets $|\eta| > 2.6$ is pretty different between 2016 and the other years. In addition, in the 2017 era, a tight requirement on the pileup jet ID is adopted for jets in the range $2.6 < |\eta| < 3.0$ in order to mitigate the noise from the ECAL detector. The BDT discriminant is trained using the ‘Gradient Boost’ method. The training is done using simulated signal and background events, using approximatively half of the available statistics corresponding to events with odd event number. Events with even event number are used instead for testing purposes. In the case of the signal, POWHEG samples (centrally produced) for ggH and VBF are used in the training because they don’t have negative weights. For background samples with negative weighted events (Drell-Yan for example), we observed similar performance on the testing sample following three alternative strategies: use both positive and negative weighted events in the training,

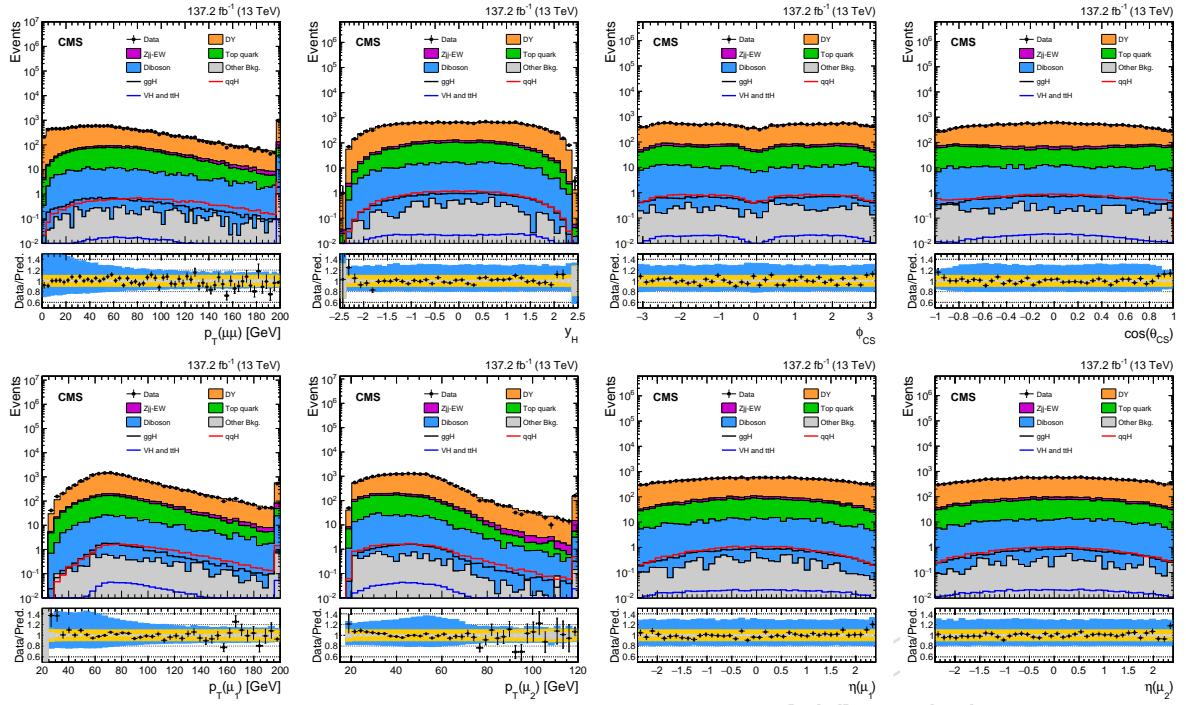


Figure A.2: Distributions for some of the variables used in input to the VBF BDT training for data (black points) and simulation (solid histograms) obtained from events in the signal fit region with $N_{\text{jets}} \geq 2$, $m_{jj} > 400 \text{ GeV}$ and $\Delta\eta_{jj} > 2.5$ and leading jet p_T greater than 35 GeV . First line, from left to right, $p_T^{\mu\mu}$, $y_{\mu\mu}$, ϕ_{CS} , and $\cos\theta_{\text{CS}}$. Second line, from left to right: $p_T^{\mu_1}$, $p_T^{\mu_2}$, η^{μ_1} , and η^{μ_2} . In the ratio panel, the grey band indicates the uncertainty coming from the limited statistics of the simulated samples, the orange one corresponds to the sum in quadrature between statistical and systematic uncertainties from theoretical sources, while the azure band includes the theoretical uncertainties in the background prediction. The asymmetry in the experimental uncertainty band is due to the one-sided effect of jet energy resolution.

2995 discard events with a negative weight, and take the absolute value of the per-event weight⁶. In
 2996 order to maximize the training statistics, the final BDTs have been trained by flipping the sign
 2997 of negative weighted events. The training is performed targeting an Higgs boson signal with a
 2998 mass of 125 GeV . The settings used for the BDT training are as follows:

- 2999 • Number of trees = 1000.
- 3000 • Minimum node size = 3%.
- 3001 • Shrinkage = 0.10.
- 3002 • Bagged sample fraction = 0.5.
- 3003 • Number of cuts = 30.
- 3004 • Maximum depth = 4.
- 3005 • Transformation for inputs: (I, N).
- 3006 • Separation type: CrossEntropy.

3007 As explained in Section 6.1, the input variables have been specifically chosen so that the BDT
 3008 output remains highly uncorrelated with the dimuon mass. The training is performed in a
 3009 tighter mass range compared to the signal fit region, i.e. $115 < m_{\mu\mu} < 135 \text{ GeV}$, which corre-

⁶In order not to bias the background composition, a correction factor is introduced to ensure that the same amount of background events are expected when negative weighted events are either discarded or they are considered with a positive weight.

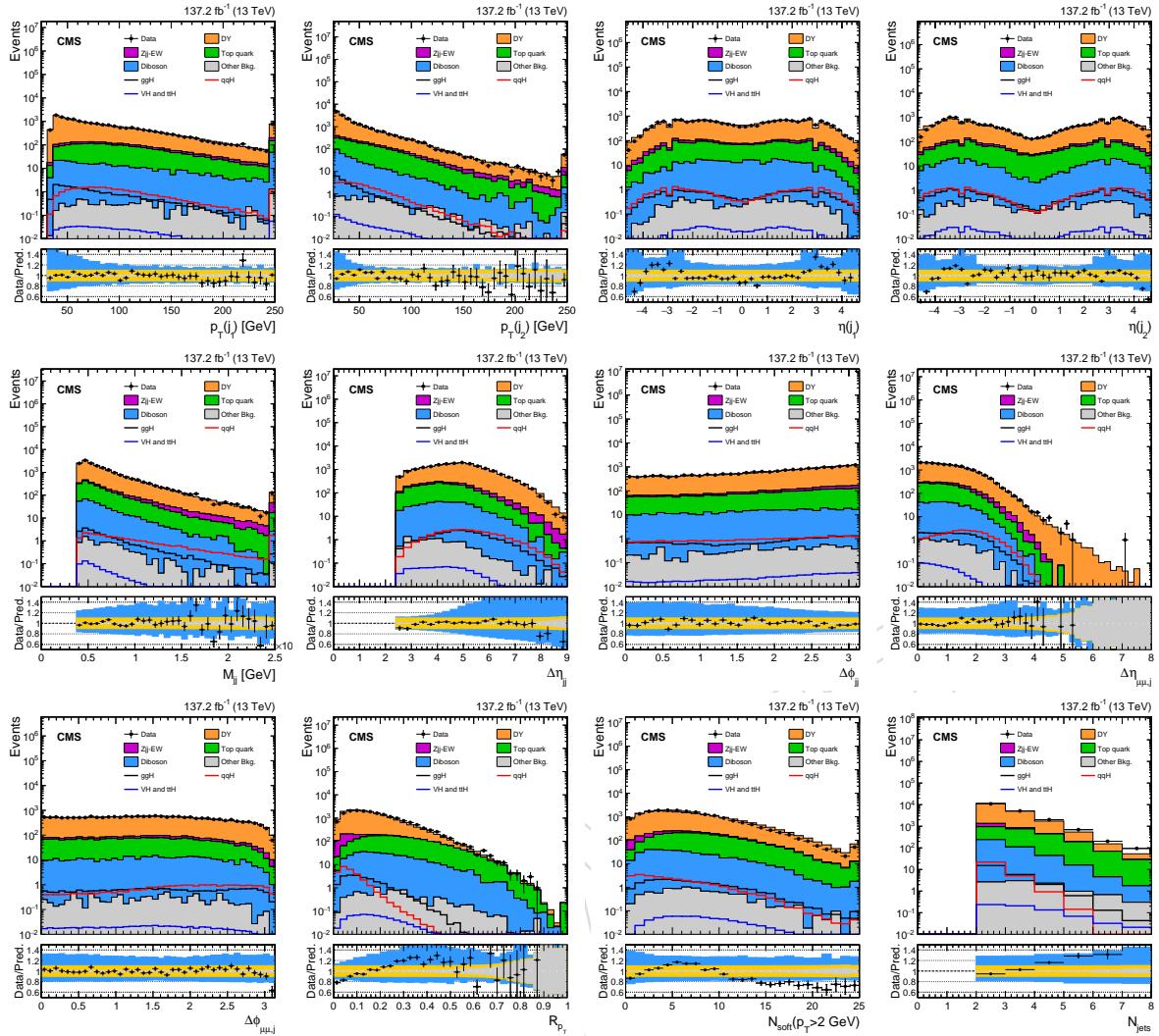


Figure A.3: Distributions for some of the variables used in input to the 2-jet VBF BDT training for data (black points) and simulation (solid histograms) obtained from events in the signal fit region with $N_{\text{jets}} \geq 2$, $m_{jj} > 400 \text{ GeV}$, $\Delta\eta_{jj} > 2.5$ and leading jet p_T larger than 35 GeV . First line, from left to right, $p_T^{j_1}$, $p_T^{j_2}$, η^{j_1} , and η^{j_2} . Second line, from left to right: m_{jj} , $\Delta\eta_{jj}$, $\Delta\phi_{jj}$, min- $\Delta\eta(\mu\mu, j_1 j_2)$. Third line, from left to right, min- $\Delta\phi(\mu\mu, j_1 j_2)$, R_{p_T} , N_{soft} , and N_{jets} . The description of the ratio panel is the same of Fig. A.2.

sponds to a $\pm 10 \text{ GeV}$ window around the true Higgs boson mass.

The analysis sensitivity depends on the dimuon mass resolution, which in turn depends strongly on the muon kinematics as described in Section 4.8, particularly the η of the muon. Therefore, we make the BDT training aware of the dimuon mass resolution by weighting the signal events proportionally to $1/\sigma_{\mu\mu}$, where $\sigma_{\mu\mu}$ is the calibrated per-event dimuon mass resolution. It should be noted that the $\sigma_{\mu\mu}$ distribution is not used as an input to the BDT, but is only used as a weighting factor in the training. The weight is not applied in the evaluation of the BDT score. The performance of the BDT discriminant does not change significantly if we include the per-event mass resolution as one of the inputs. This is to be expected because the shape of the $\sigma_{\mu\mu}$ distribution is very similar between signal and background events. Therefore, the adopted use of the per-event mass uncertainties is a significant departure from the approach adopted in the earlier iteration of the analysis [13]. In fact, in the previous iteration, events were categorized both on the basis of their BDT score and the maximum of the $|\eta|$ of the two

3023 muons. The approach adopted here of using per-event mass uncertainties as a weight for sig-
 3024 nals events during the BDT training reduces the categorization to a single variable i.e. the BDT
 3025 score. In addition, this also carries more precisely the information about the muon momen-
 3026 tum resolution and its correlation with the event kinematics (muon p_T and η). Consequently,
 3027 we will end up using significantly fewer categories but still having better performance (25%
 3028 improved expected significance) on 2016 data w.r.t. Ref. [13].

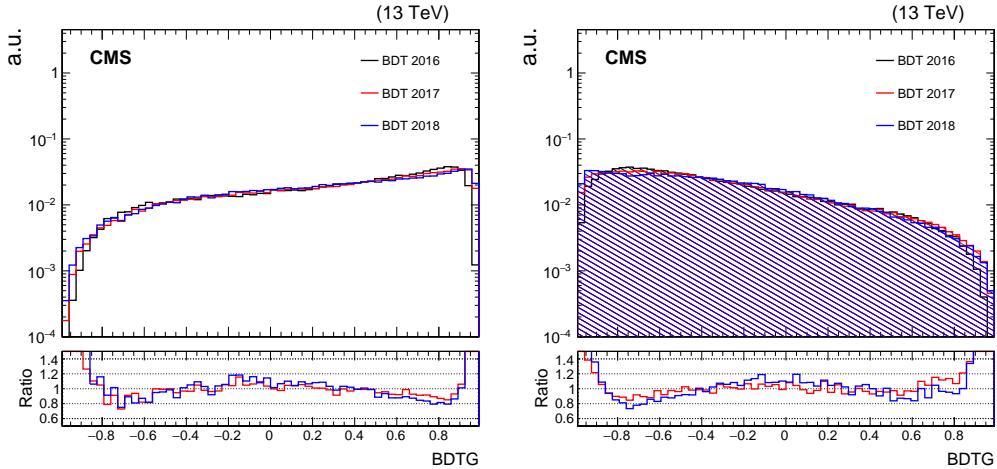


Figure A.4: Distribution of the VBF BDT outputs in signal (left) and background (right) simulated events taken from the testing sample for three eras: 2016 (black), 2017 (red), and 2018 (blue). Distributions are scaled in order to have unit area. The red (blue) distribution in lower pad is defined as the ratio between the 2018 (2017) BDT shape and the 2016 one, which is used as reference. The shapes of the BDT outputs are found to be quite similar across the three eras. The largest differences are observed in the high score region.

3029 Figure A.4 shows a comparison between the output of the two BDT discriminants (left: ggH
 3030 BDT, right: VBF BDT) in the case of the signal (top) and background (bottom) simulations for
 3031 the three eras. The shapes of the ggH BDT outputs are found to be in good agreement for both
 3032 signal and background processes apart from the high score region (BDT-score larger than 0.6),
 3033 where the 2017 and 2018 BDTs show longer tails in both signal and background events. Similar
 3034 differences in the high score region are also observed in the VBF BDTs (BDT-score greater than
 3035 0.85). Overall, the BDT discriminators related to 2017 and 2018 eras are quite similar in terms
 3036 of shapes in both ggH and VBF trainings. In contrast, as already mentioned before, the 2016
 3037 shapes are more different particularly in the high score region. Since the properties of jets in the
 3038 endcap and forward regions of the detector changed significantly after the 2016 data-taking, it
 3039 is expected to see differences in regions more sensitive to the forward jet kinematics between
 3040 the 2016 BDTs and those from the other eras.

3041 Figure A.5 (left) shows the "ROC" (Receiver Operating Characteristic) curves for the VBF BDT
 3042 discriminant. The performance of 2016 training outperforms both the 2017 and the 2018 ones.
 3043 Figure A.5 (right) shows the dimuon mass resolution, expressed in terms of full width half max-
 3044 imum (FWHM), as a function of the BDT score. The dimuon mass resolution is extracted from
 3045 a fit to the $m_{\mu\mu}$ distribution with a Double-sided Crystal Ball function. The signal resolution
 3046 improves as a function of the BDT score, thereby demonstrating that the BDT is indeed aware
 3047 of the importance of the dimuon mass resolution. This improvement in the resolution is more
 3048 visible along the ggH BDT (Fig. 6.6) output than the VBF one. In fact, in the VBF phase space,
 3049 there is a strong interplay between the per-event mass resolution and other kinematic variables
 3050 used in input to the BDT (VBF sensitive variables), therefore the resolution more or less flattens
 3051 out at high scores because a better separation between signal and background can be achieved

3052 by looking at kinematic properties rather than the mass resolution. In addition, the dimuon
 3053 mass resolution is found to improve progressively moving from 2016 to 2018 eras, and it is also
 3054 observed to be stable as a function of the BDT outputs. This gain in the muon p_T resolution
 3055 is due to a combination of sources: installation of an additional layer in the Pixel detector in
 3056 early 2017, improvements in muon and tracker detector alignments, as well as improvements
 3057 in beam-spot and muon tracker reconstruction (particularly for high p_T muons) allowing the
 3058 geo-fit corrections to be more effective in 2018 rather than in the 2016 data-taking period.

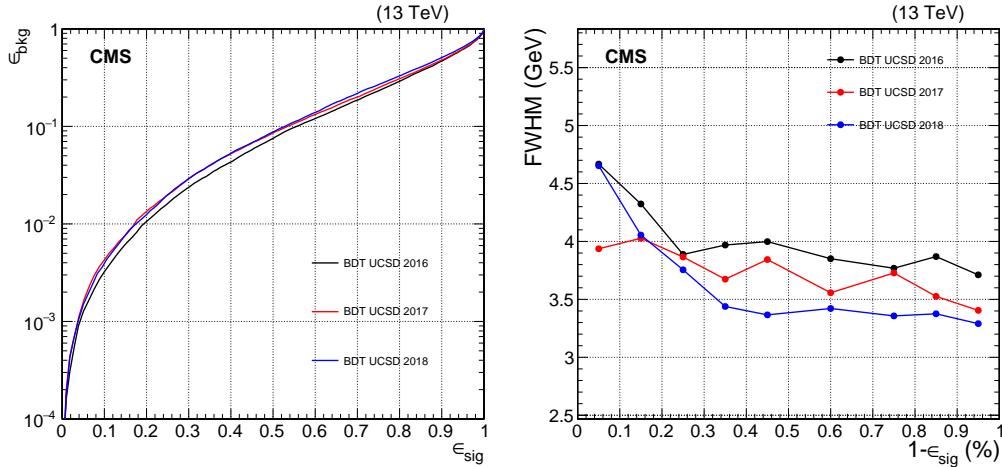


Figure A.5: On the left, signal efficiency (ϵ_{sig}) vs background efficiency (ϵ_{bkg}) ROC curves obtained from the VBF BDT output scores for events in the testing sample. The performance of the BDT discriminant in the 2016 training outperforms the MVAs obtained from the other years. On the right, the FWHM of the dimuon mass distribution in signal events computed as a function of the VBF BDT score. The signal resolution improves similarly in the three eras as a function of the BDT score, proving that the BDT is able to assign high score to events with good mass resolution.

3059 In order to understand how the BDT score is correlated with the input observables, Fig. A.6
 3060 shows, for signal events, the evolution of the shapes of some of the input variables as a function
 3061 of the VBF BDT score. The high score region of the VBF BDT, which looks at events with at least
 3062 two jets with m_{jj} larger than 400 GeV, probes events with a large dimuon p_T , large m_{jj} , large m_{jj} ,
 3063 and a small R_{p_T} .

3064 Moreover, it is also crucial to make sure that both the these BDTs haven't learnt the signal mass
 3065 hypothesis used in the training, $m_H = 125$ GeV, via hidden correlations between the invariant
 3066 mass of the two muons and the other input variables. To do so, two independent cross checks
 3067 are performed:

- Fig. A.7 (left) shows, for background events, the evolution of the dimuon mass shape as a function of the VBF BDT scores where no sculpting of the expected mass distribution is observed either at low or at high BDT score. This suggests that the BDT output is not sensitive to the dimuon mass distribution.
- Fig. A.7 (right) shows, for VBF signal events, the evolution of the VBF BDT shapes as a function of the signal mass hypothesis, i.e. for $m_H = 120, 125$, and 130 GeV. No differences as a function of the mass are observed for the two BDT outputs.

3075 Therefore, we can conclude that the VBF BDT is highly uncorrelated with the dimuon mass.
 3076 Events can be further categorized based on the BDT scores in order to maximize the search
 3077 sensitivity without sculpting or biasing the expected invariant mass distribution for signal and
 3078 background events.

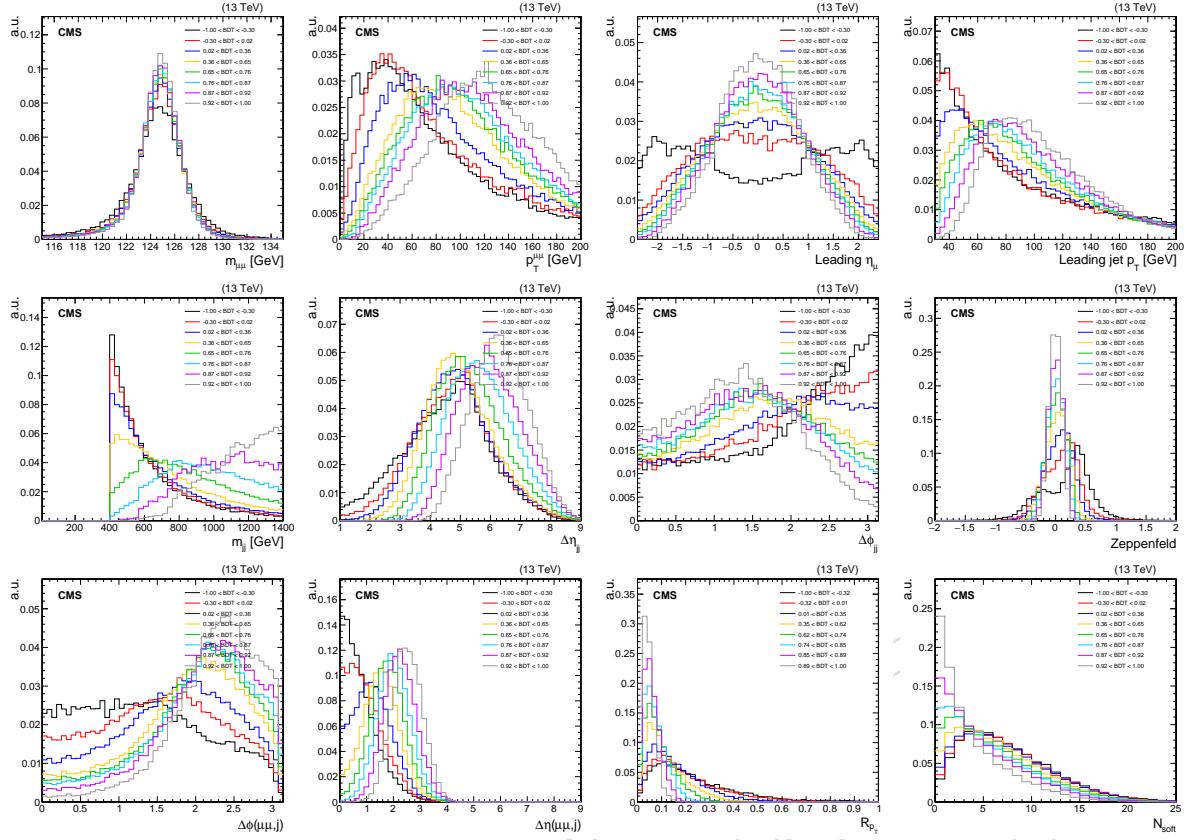


Figure A.6: Evolution of the distributions of some of the key kinematic observables as a function of the VBF BDT output for signal events (ggH, VBF, VH, and ttH summed together). Distributions are scaled in order to have unit area. On top, from left to right: $m_{\mu\mu}$, $p_T^{\mu\mu}$, η_{μ_1} , and $p_T^{j_1}$. Middle, from left to right: m_{jj} , $\Delta\eta_{jj}$, $\Delta\phi_{jj}$, and Zeppenfeld. Bottom, from left to right: min- $\Delta\phi(\mu\mu, j)$, min- $\Delta\eta(\mu\mu, j)$, R_{p_T} , and N_{soft} .

Finally, Fig. A.8 shows a comparison between the ROC curves obtained in each era and in each training from training (black) and testing (red) events. A small over-training is observed in 2017 and 2018 eras because the available Drell-Yan MC statistics is about 20% smaller than in 2016.

A.3 Event categorization

At this stage, we have pre-selected events into two categories. VBF events are those with at least two jets in the final state with $m_{jj} > 400 \text{ GeV}$, $\Delta\eta_{jj} > 2.5$, $p_T^{j_1} > 35 \text{ GeV}$, and $p_T^{j_2} > 25 \text{ GeV}$. In contrast, the “inclusive” phase space collects all the remaining events which are not belonging to the VBF channel or to the other exclusive modes (VH-leptonic and ttH). Figure A.9 (left) shows the distributions of the VBF BDT score in data and simulation for events belonging to the signal fit region, $110 < m_{\mu\mu} < 150 \text{ GeV}$. In contrast, Fig. A.9 (right) reports a comparison between data and MC in the ZCR, i.e. in events passing the analysis selections with $m_{\mu\mu}$ between 70 and 110 GeV. The ZCR is used in order to test more precisely how the simulation models the data in the high score region, in fact its statistical power is about one-hundred times larger than the fit region. The observation and the expectation in the three eras (2016, 2017, and 2018) have been summed together. The MVA-score the output score has been analytically transformed in order to better highlight events belonging to the high score region, where the VBF signal purity

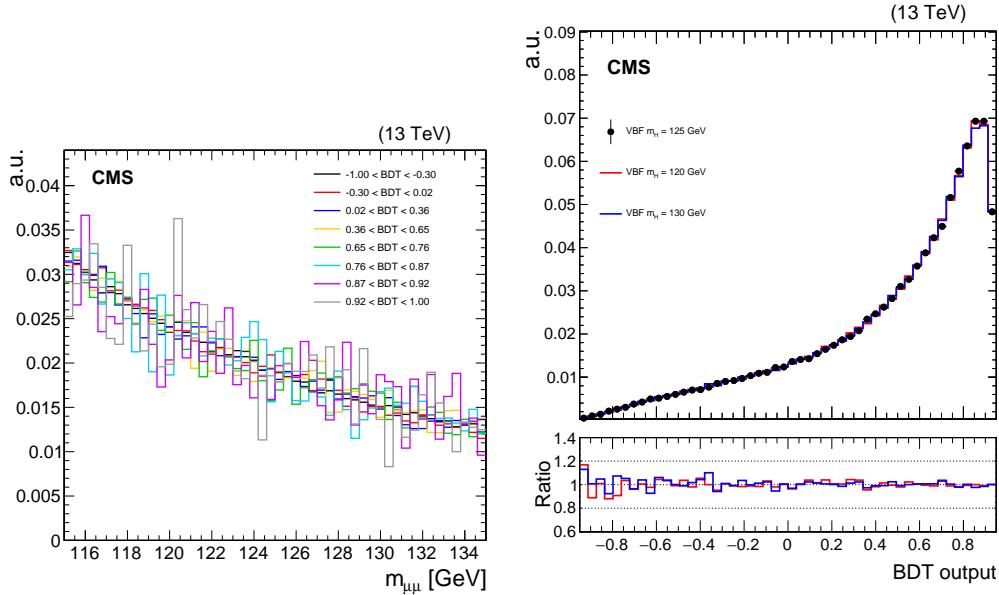


Figure A.7: On the left, dimuon mass distribution in simulated background events computed in a set of orthogonal ranges defined along VBF BDT score. On the right, distribution of the VBF BDT outputs in VBF simulated signal events for three different mass hypothesis: $m_H = 125$ GeV (black), $m_H = 120$ GeV (red), and $m_H = 130$ GeV (blue). Distributions are scaled in order to have unit area.

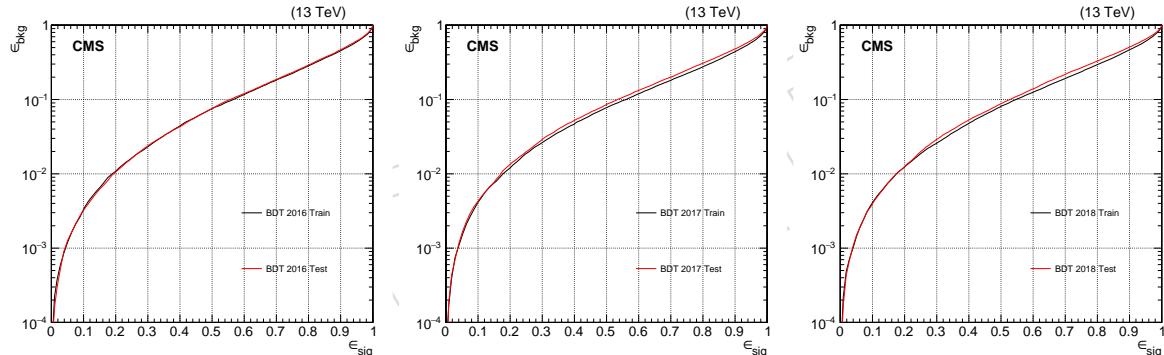


Figure A.8: Signal efficiency (ϵ_{sig}) vs background efficiency (ϵ_{bkg}) ROC curves obtained in each eras from the VBF BDT output scores for events in the training (black) and testing (red) samples. The ROC curves shown in the first column refers to the 2016 era, while in the second and third ones the ROCs for 2017 and 2018 eras are reported. A small difference between training and testing events is observed for the BDTs trained on 2017 and 2018 MC, while no over-fitting is observed in the 2016 era. Therefore, the trained BDTs are at most slightly affected by over-training.

increases. The transformed BDT output is defined as follows:

$$\text{transformed-BDT} = \text{BDT}' = \tanh^{-1} \left(\frac{\text{BDT-score} + 1}{2} \right)$$

3084 In the ratio plot, the grey band corresponds to the uncertainty coming from the limited statistics
 3085 of the simulated samples, while the orange one corresponds to the sum in quadrature be-
 3086 tween statistical and systematic uncertainties from theoretical sources. Finally, the azure band
 3087 includes an estimate of the experimental uncertainties, which is added in quadrature to the
 3088 statistical and the theoretical systematic ones. The uncertainty due to jet energy resolution is
 3089 one-sided as described in more details in Section A.5.3. Overall, a good agreement between

3090 data and MC is observed in both cases and most of the discrepancies are covered by systematic
 3091 uncertainties assigned to the background prediction.

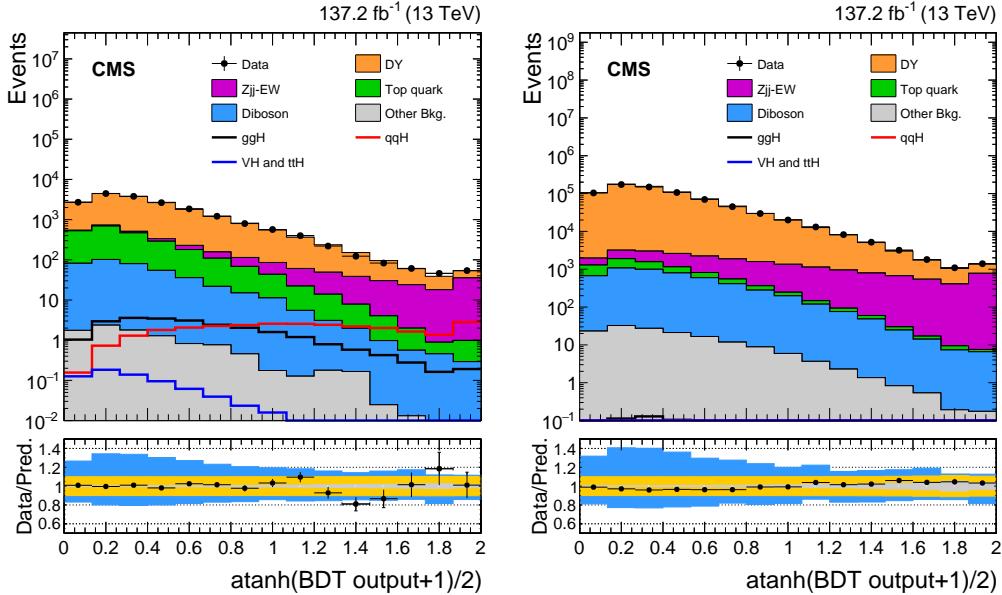


Figure A.9: Distribution of VBF BDT scores reported for both data (black points) and simulation (solid histograms) for events with $m_{\mu\mu}$ between 110 and 150 GeV (left), as well as for events in the ZCR (right) where $70 < m_{\mu\mu} < 110$ GeV. The observation and the expectation in the three eras (2016, 2017, and 2018) have been summed together. In the ratio plot, the grey band indicates the uncertainty coming from the limited statistics of the simulated samples, the orange one corresponds to the sum in quadrature between statistical and systematic uncertainties from theoretical sources, while the azure band includes the experimental uncertainties in the background prediction.

3092 Comparisons are also reported independently for each year in Fig. A.10, where the first column
 3093 corresponds to the 2016, the second to the 2017, and the third to the 2018 era. A good agreement
 3094 between data and MC is observed across all eras. A 10% offset in the overall normalization is
 3095 observed in the 2017 era which is well covered by both the theoretical uncertainties in the
 3096 background prediction and the uncertainties attributed to the knowledge of the jet energy scale
 3097 and resolution. In the 2016 era, an excess of data is observed in the high score region of both
 3098 the signal and the Z-peak control sample.

3099 A.3.1 Category optimization

3100 Based on the output of the trained BDT discriminant reported in Fig. A.9, we can further cat-
 3101 egorize events belonging to the VBF phase space. In order to do that, we have followed a
 3102 strategy identical to the one adopted to optimize ggH categories described in Section 6.2.1.

3103 We start with a single category collecting all the events belonging to the VBF phase space,
 3104 and we compute the expected significance for a Higgs boson with $m_H = 125$ GeV. This is ob-
 3105 tained by modelling the $m_{\mu\mu}$ distribution of the total signal (all production modes are added
 3106 together into a single distribution) with a double-sided Crystal Ball function. The mass range
 3107 of the fit is chosen to be 110–150 GeV. In this optimization procedure, ggH events from the
 3108 MADGRAPH5_aMC@NLO generator are considered because these simulated samples are the
 3109 ones used for the final signal extraction. In contrast, VBF events are taken from the centrally
 3110 produced POWHEG samples showered with PYTHIA and only even event-number are consid-
 3111 ered, which correspond to those belonging to the MVA testing sample. The total expected
 3112 background yield is also obtained from simulation (background simulated samples listed in

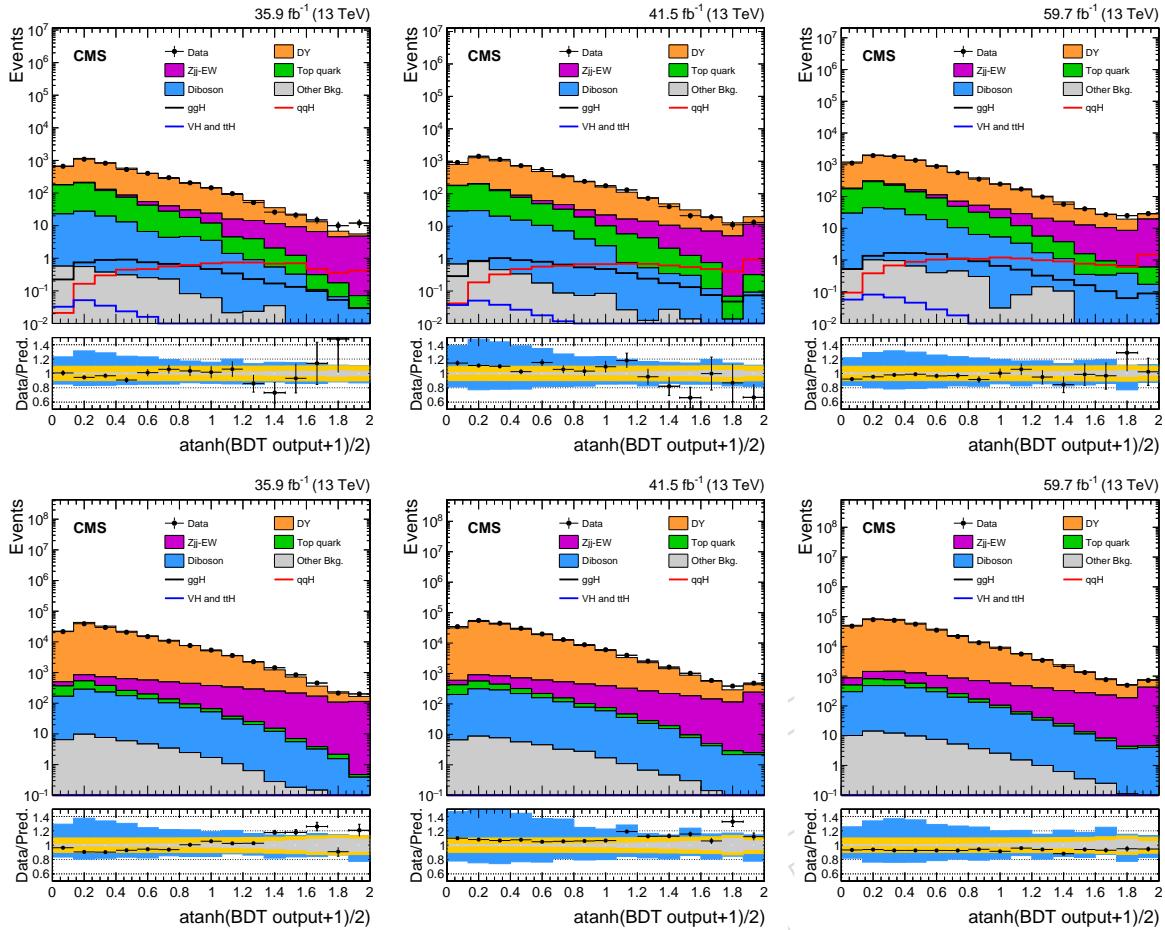


Figure A.10: Distribution of the VBF BDT scores in 2016 (left), 2017 (middle) and 2018 (right) eras reported for both data (black points) and simulation (solid histograms) for events with $m_{\mu\mu}$ between 110 and 150 GeV (first row), as well as for events in the ZCR (second row) where $m_{\mu\mu} > 110$ GeV. In the ratio plot, the grey band indicates the uncertainty coming from the limited statistics of the simulated samples, the orange one corresponds to the sum in quadrature between statistical and systematic uncertainties from theoretical sources, while the azure band includes the experimental uncertainties in the background prediction.

3113 Section 2) and we model the shape of the dimuon mass distribution in background events via
 3114 an analytical function called BWZGamma (two shapes degrees of freedom), defined in Equa-
 3115 tion 5.2.

3116 After this, we compare the significance obtained with a single category with the one obtained
 3117 by dividing the events into two categories. Since the BDT shapes for signal and background
 3118 events are different across eras, as shown in Fig. A.4, the BDT outputs in the VBF phase space
 3119 are scanned in signal quantiles, i.e. in steps of 1% in signal efficiency. The optimal position for
 3120 the category boundary corresponds to the one that gives the maximum significance obtained
 3121 from a combined fit of the two categories, where the total signal and background contribu-
 3122 tions are modelled as described above. Once the optimal boundary has been found and two
 3123 categories are defined, the iterative algorithm searches for a third optimal category, etc. This
 3124 procedure stops when the n -th category does not provide an additional gain in significance
 3125 larger than 1%.

3126 By using this iterative strategy, four categories are defined along the VBF BDT score. The cat-
 3127 egory boundaries defined in terms of signal efficiency quantiles and BDT ranges are reported

in Table A.1. However, since predictions from MADGRAPH5_aMC@NLO and POWHEG generators for both the ggH process can be different by a few % depending on the phase space, we have repeated the same optimization procedure using POWHEG signal samples instead of MADGRAPH5_aMC@NLO ones. The same number of optimal categories is found, and the corresponding ranges in signal efficiency are in agreement within $\pm 2\%$, suggesting that the optimal categories listed in Table A.1 are robust and universal, i.e. not specific to a particular MC generator.

Table A.1: Optimized category boundaries defined along the output of the transformed VBF BDT discriminant.

Name	Signal eff. range	BDT range 2016	BDT range 2017	BDT range 2018
VBF-cat0	[100, 45)%	[0.000, 0.948)	[0.000, 0.932)	[0.000, 0.866)
VBF-cat1	[45, 22)%	[0.948, 1.347)	[0.932, 1.372)	[0.866, 1.310)
VBF-cat2	[22, 10)%	[1.347, 1.610)	[1.372, 1.728)	[1.310, 1.701)
VBF-cat3	$\leq 10\%$	≥ 1.610	≥ 1.728	≥ 1.701

Figure A.11 shows the evolution of the combined expected significance from categories defined along the VBF BDT output during the optimization procedure. The first plot on the left reports the trend of the significance as a function of the first boundary, defined in terms of signal efficiency quantiles. As described before, expectations in signal and background from the three eras are summed together. The significance is maximized by dividing the VBF phase space into two categories in a signal efficiency range between. Moreover, the other graphs reported in Fig. A.11 show how the other category boundaries listed in Table A.1 have been defined during the iterative scan. The combined significance extracted during the optimization procedure from the four BDT categories is of about 1.45σ .

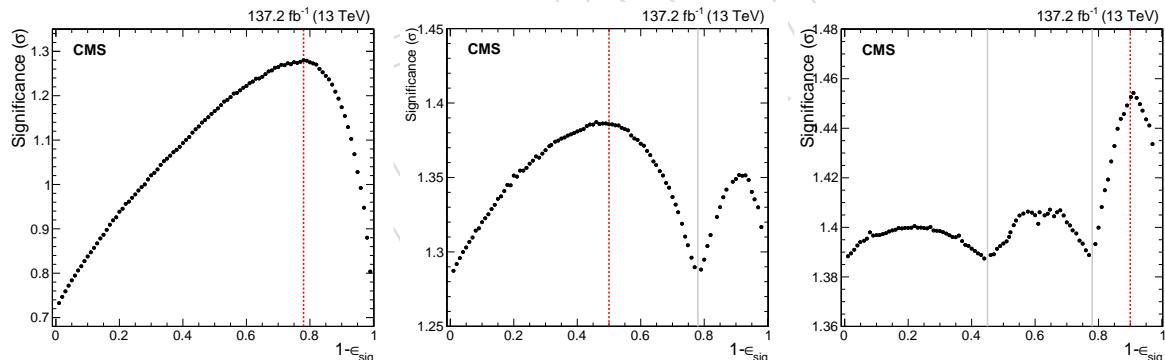


Figure A.11: Evolution of the combined significance for a Higgs boson signal with $m_H = 125\text{ GeV}$ defined by scanning the transformed VBF BDT output in signal efficiency quantiles. From left to right: improvements in the total expected significance by using two categories instead of one, three categories instead of two, and four categories instead of three. Dashed red lines indicate the category boundary giving optimal performance, while the solid gray ones indicate already defined boundaries from previous steps of the iterative procedure.

Figure A.12 shows a comparison between the outputs of the VBF BDT discriminant for signal and background simulated events. The black dashed vertical lines indicate, approximately, the category boundaries optimized with the iterative procedure previously described. In addition, inside the lower pad of each figure, the significance expressed as S/\sqrt{B} of each category, calculated by integrating signal and background expected events inside the half width at half maximum (HWHM, $\text{HWHM}=\text{FWHM}/2$) of the expected signal peak, is also reported.

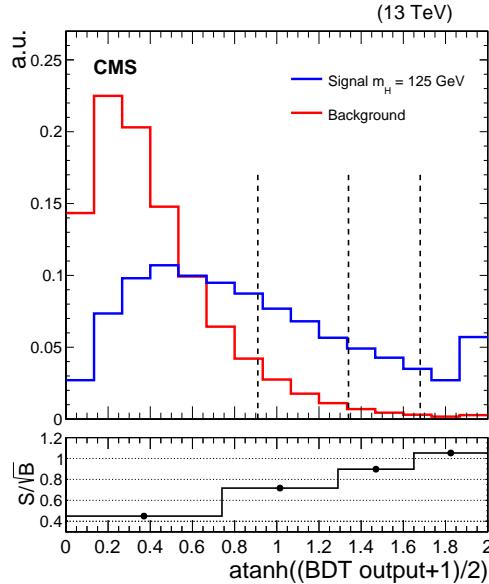


Figure A.12: Distribution of the transformed VBF BDT output in signal (blue) and background (red) simulated events. The black dashed vertical lines indicate, approximately, the category boundaries optimized with the iterative procedure previously described. In the lower pad, the significance expressed as S/\sqrt{B} of each category, calculated by integrating signal and background expected events inside the HWHM of the expected signal peak, is also reported.

Finally, Fig. A.13 shows the distributions of the dimuon invariant mass in the four VBF categories for data and simulated events in the range $110 < m_{\mu\mu} < 150$ GeV. The agreement between data and simulation in both shape and normalization, apart from statistical fluctuations in high score region of the VBF BDT, is good across all categories. However, since the background in this search is directly extracted from data, as described later in Section 6.4, this discrepancy in normalization does not have any effect on the analysis results. For the time being, the signal peak region remains blinded ($120 < m_{\mu\mu} < 130$ GeV).

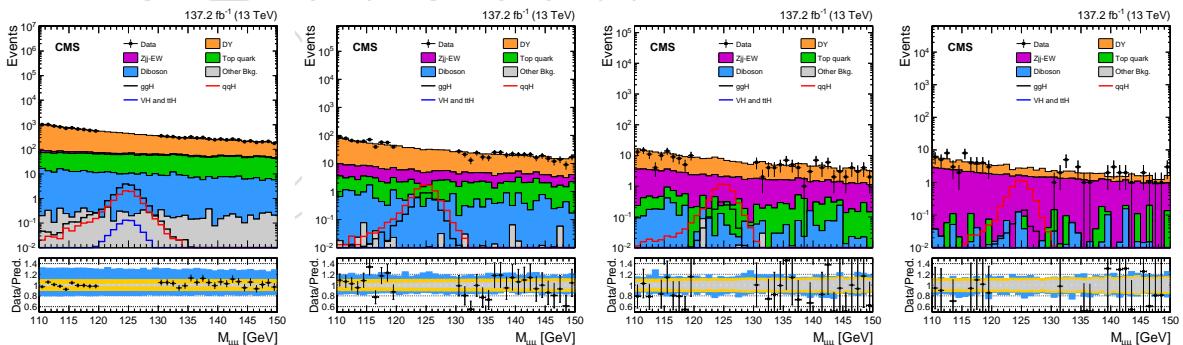


Figure A.13: Distribution of the dimuon invariant mass in the VBF categories reported for both data (black points) and simulation (solid histograms) for events with $m_{\mu\mu}$ between 110 and 150 GeV. From left to right: VBF-cat0, VBF-cat1, VBF-cat2, and VBF-cat3. The expected signal distributions from the different Higgs boson production modes are also reported.

A.4 Signal extraction

In this search the signal is extracted by fitting the dimuon invariant mass spectrum in data via analytical functions able to model the expected signal and background processes. In the case of the signal (ggH, VBF, VH, and ttH), these analytical functions contain several parameters

whose values are determined from simulation and are not allowed to float in the final fit, i.e. they are treated as constants or constrained parameters when the fit to the data is performed. In the case of the background, we assume no prior knowledge of the parameters of the fit functions, therefore they are allowed to float freely in the signal extraction fit. The overall normalizations of the signal and the backgrounds are also floated freely.

A.4.1 Signal lineshape and yields

The signal line shape is modelled using a double-sided Crystal Ball function defined in Equation 4.2. In each of the four categories listed in Table A.1, we fit the signal contribution from each the production modes (ggH, VBF, WH, ZH, and ttH) independently and we fix the corresponding shape parameters. Since POWHEG simulated events for ggH have been used in the ggH and VBF BDT trainings, both the lineshapes and the expected yields used in the ultimate signal extraction for this process are obtained from MADGRAPH5_aMC@NLO simulated samples. Predictions for the VBF production in the VBF BDT categories are taken instead from POWHEG interfaced with PYTHIA in dipole shower mode. Figure A.14 shows the resulting signal shapes for the VBF process in VBF categories obtained by fitting the expected dimuon mass spectrum with a double-sided Crystal Ball, where the expectations in the three eras (2016, 2017, and 2018) have been summed together. Overall, the double-sided Crystal Ball provides a good description of both the bulk and the tails of the expected signal distributions in each category, $\chi^2/ndf \approx 1$.

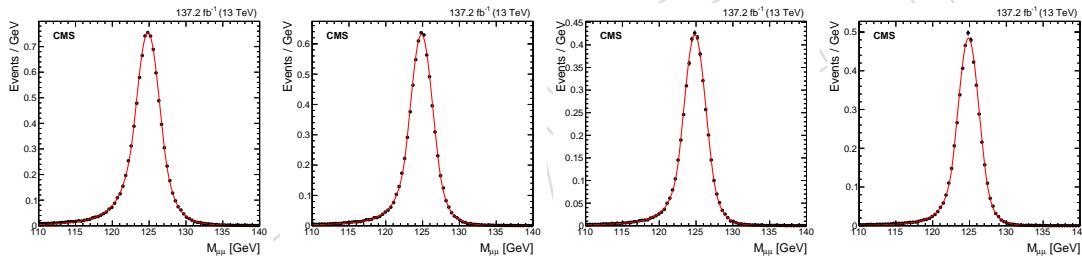


Figure A.14: Distributions of the dimuon invariant mass for VBF signal events produced via the POWHEG generator (showered with PYTHIA in dipole mode) in the VBF BDT categories. From left to right: VBF-cat0, VBF-cat1, VBF-cat2, and VBF-cat3, where the expectations in the three eras (2016, 2017, and 2018) have been summed together.

The Higgs boson signal rates expected in 137.2 fb^{-1} of data are reported in Table A.2 for each category defined along the VBF BDT output, as well as for each Higgs boson production mode. The VBF signal overcomes the ggH one in the last three categories defined in the VBF phase space.

Table A.2: Higgs boson signal rates expected in 137.2 fb^{-1} of data in each category defined along VBF BDT output for ggH, VBF, WH, ZH, and ttH production modes.

Category	ggH	VBF	WH	ZH	ttH	Total
VBF-cat0	18.38	9.74	0.39	0.23	0.04	28.79
VBF-cat1	3.99	7.38	0.02	0.01	< 0.01	11.42
VBF-cat2	1.16	4.75	< 0.01	< 0.01	< 0.01	5.94
VBF-cat3	0.48	5.17	< 0.01	< 0.01	< 0.01	5.67

The same procedure is also applied to alternative signal samples produced with a different Higgs boson mass hypothesis, namely $m_H = 120$ and 130 GeV . This allows to interpolate both the expected signal rates and the double-sided Crystal Ball shape parameters as a function of the Higgs boson mass within the mass range of $120\text{--}130 \text{ GeV}$. The interpolation vs m_H is

3188 performed via a smooth spline function. The interpolation is crucial in order to compute local
 3189 p-values (upper limits) scanning the interesting mass range of a few GeV around the measured
 3190 Higgs boson mass of 125.09 GeV. Figure A.15 shows the interpolated signal shapes, normalized
 3191 to the expected rate in 137.2 fb^{-1} of data, for the VBF process in the VBF categories.

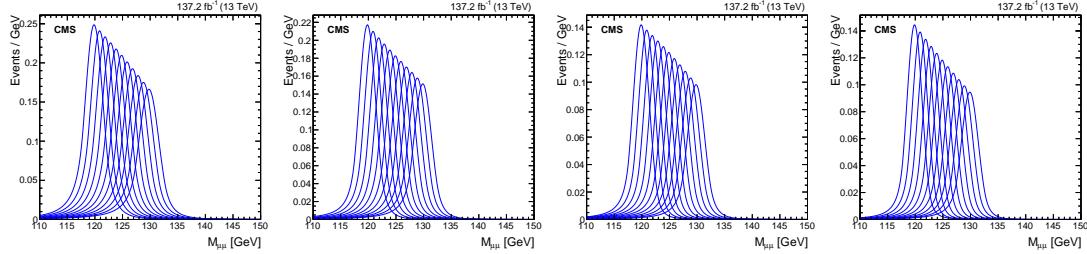


Figure A.15: Distributions of the dimuon invariant mass for the VBF process in the VBF BDT categories as a function of m_H , resulting from an interpolation of the expected rates and shape parameters across the generated Higgs signal samples for $m_H = 120, 125$, and 130 GeV . From left to right: VBF-cat0, VBF-cat1, VBF-cat2, and VBF-cat3.

3192 In addition, in order to better track the evolution of expected signal yield and shape as a
 3193 function of m_H , the trends of the dimuon peak width (σ_{CB}) and of the acceptance times ef-
 3194 ficiency are reported for the VBF signal in the VBF categories in Fig. A.16 and A.17, respec-
 3195 tively. As expected, the signal width increases linearly with the mass but the mass resolu-
 3196 tion, defined as σ_{CB}/m_H , remains constant. The product of acceptance and efficiency also
 3197 slightly increases with the mass in each category. However, since $\sigma \times \mathcal{B}(H \rightarrow \mu\mu)$ decreases
 3198 by about 30% moving from $m_H = 120$ to $m_H = 130 \text{ GeV}$, the total expected signal rate given by
 3199 $\sigma \times \mathcal{B}(H \rightarrow \mu\mu) \times \text{acceptance} \times \text{efficiency}$ also decreases as already shown in Fig. A.15.

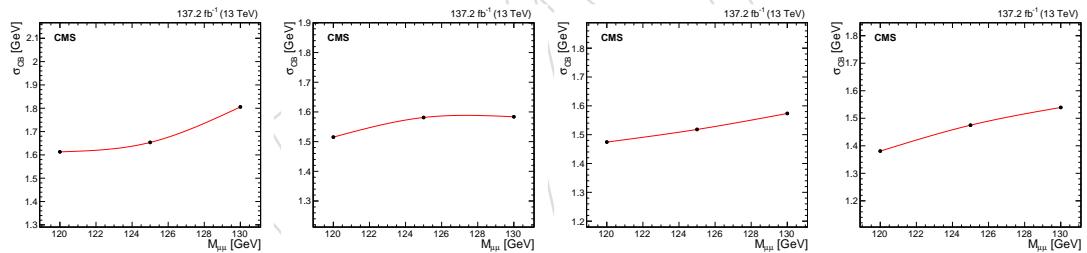


Figure A.16: Evolution of the dimuon peak width (σ_{CB}) for the VBF process in the VBF BDT categories as a function of m_H , in the range $120 < m_{\mu\mu} < 130 \text{ GeV}$. From left to right: VBF-cat0, VBF-cat1, VBF-cat2, and VBF-cat3.

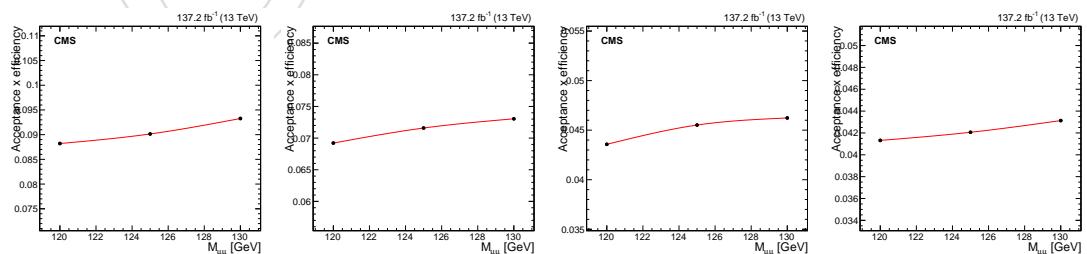


Figure A.17: Evolution of the acceptance times efficiency for the VBF process in the VBF BDT categories as a function of m_H , in the range $120 < m_{\mu\mu} < 130 \text{ GeV}$. From left to right: VBF-cat0, VBF-cat1, VBF-cat2, and VBF-cat3.

3200 **A.5 Uncertainties on the signal model**

3201 Systematic uncertainties affecting of the expected signal prediction in each category are taken
 3202 into account in the signal extraction and distinguished between *shape* and *rate* uncertainties.

3203 The *shape* uncertainties are those affecting the modelling of the signal peak, while the *rate*
 3204 uncertainties are those affecting the predicted number of signal events in each category.

3205 A.5.1 Shape uncertainties

3206 For all Higgs boson production modes, the expected $m_{\mu\mu}$ signal shape is primarily affected by
 3207 the uncertainties in the muon energy scale and resolution. The Rochester corrections, described
 3208 in Section 4.6, are meant to correct for differences in both scale and resolution between data
 3209 and simulated events. However, the uncertainties associated to the Rochester corrections, the
 3210 uncertainties from the calibration of the per-event mass errors (closure test), as well as the
 3211 assumption of a given analytical signal shape (double-sided Crystal Ball), have to be taken into
 3212 account in the signal extraction.

3213 The residual muon energy scale uncertainty has been estimated to be 0.1% of the muon mo-
 3214 mentum, and it affects the signal peak by changing its position w.r.t. the nominal expected
 3215 value from simulation. This uncertainty is correlated across all signal production processes
 3216 and eras, but decorrelated across categories. It is modelled via Gaussian constrained nuisance
 3217 parameters. On the other hand, the uncertainty in the dimuon mass resolution has been es-
 3218 timated, conservatively, to be as large as 10% in each category. It is modelled via a Gaussian
 3219 constrained nuisance parameter, correlated across processes and eras, but uncorrelated across
 3220 categories whose effect is to shrink or enlarge the width of the signal peak (σ_{CB}).

3221 A.5.2 Rate uncertainties from theory

3222 Several sources from theory introduce an uncertainty in the predicted signal rate in each event
 3223 category and for each production mode. Some of them affect the overall signal prediction, i.e.
 3224 they affect either the production cross section or the value of $\mathcal{B}(H \rightarrow \mu\mu)$, and are referred to
 3225 as *normalization* uncertainties. In contrast, others only affect the prediction of the acceptance
 3226 in single categories, causing migrations of events across ggH and VBF BDT categories. These
 3227 are referred to as *acceptance* uncertainties.

3228 As described in Section 2, the central values of the Higgs boson production cross section and
 3229 $\mathcal{B}(H \rightarrow \mu\mu)$ are taken from Ref. [12]. The associated uncertainties, listed in Table 2.1, are mod-
 3230 elled via constrained log-Normal nuisance parameters and they are fully correlated across cat-
 3231 egories and eras. The uncertainty on $\mathcal{B}(H \rightarrow \mu\mu)$ is also fully correlated across Higgs produc-
 3232 tion modes. In addition, the uncertainty in the total cross section from PDF+ α_s can be correlated
 3233 only between VBF and VH modes because both processes are quark-initiated.

3234 The uncertainties affecting the ggH process from the sources defined by the LHC cross-section
 3235 WG1, already introduced in Section 5.4, are estimated directly via the simulated sample used
 3236 in the analysis. They are found to be in good agreement between MADGRAPH5_aMC@NLO
 3237 and POWHEG generators. In addition, the uncertainty in the ggH acceptance due to the lim-
 3238 ited knowledge of the parton densities (PDF) has been also investigated using variations of
 3239 the per-event weight stored inside in the simulated samples. Table A.3 summarizes the afore-
 3240 mentioned per-category uncertainties for the ggH production mode, as well as how they are
 3241 correlated across categories. The total uncertainty in the acceptance from these sources is of
 3242 about 12–15% in the VBF categories. As expected, specific p_T^H and VBF phase-space uncertain-
 3243 ties represent the major sources of uncertainty in the VBF BDT categories, where on the VBF-3j
 3244 nuisance is strongly anti-correlated because the high-score VBF region is dominated by 2-jet
 3245 events. Finally, uncertainties related to jet migrations (Mig-01, and Mig-12) show a correlated
 3246 effect in the acceptance of the VBF categories.

3247 Furthermore, acceptance uncertainties designed for the VBF process and obtained following

Table A.3: Acceptance uncertainties for ggH signal events in VBF BDT categories from WG1 uncertainties and PDF. Uncertainties smaller than 0.1% are not reported.

Category	ggH-Mu	ggH-Res	Mig-01	Mig-12	VBF-2j	VBF-3j	Pt60	Pt120	Mtop	PDF
VBF-cat0	0.963/1.037	0.968/1.032	0.967/1.033	0.991/1.009	0.919/1.081	0.999/1.001	0.996/1.004	1.019/0.981	0.990/1.010	0.993/1.007
VBF-cat1	0.969/1.031	0.972/1.028	0.968/1.032	0.997/1.003	0.901/1.099	1.047/0.953	0.974/1.026	1.033/0.967	0.989/1.011	0.990/1.010
VBF-cat2	0.975/1.025	0.978/1.022	0.974/1.026	0.999/1.001	0.880/1.120	1.108/0.892	0.969/1.031	1.039/0.961	0.990/1.010	0.983/1.021
VBF-cat3	0.977/1.023	0.979/1.021	0.977/1.023	0.993/1.007	0.872/1.128	1.133/0.867	0.959/1.041	1.055/0.945	0.986/1.014	0.976/1.043

the LHC Higgs cross-section WG2 recommendations, as described in Section 5.4, are estimated directly via the simulated sample used in the analysis, and are found to be in good agreement between MADGRAPH5_aMC@NLO and POWHEG generators. In addition, the uncertainty in the VBF signal acceptance due to the limited knowledge of the parton densities (PDF) has been also investigated using variations of the per-event weight stored inside in the simulated samples. Table A.4 summarizes the aforementioned per-category uncertainties for the VBF production mode, as well as how they are correlated across categories. The total uncertainty in the acceptance from these sources is of about 2.0–2.5% in the VBF categories. As expected from the coarse binning adopted in the STXS prescription, only the low- p_T^H uncertainty (Pt25) plays a role. For what concern m_{jj} uncertainties, those related to low m_{jj} bins are driving the overall acceptance uncertainty in the VBF categories, while M_{jj} -1000 and M_{jj} -1500 have no effect since the analysis is not very sensitive yet to the TeV tail of the m_{jj} distribution. The effect of M_{jj} -60, M_{jj} -120, and M_{jj} -350 is correlated across VBF categories. In fact, given the offline selection of $m_{jj} > 400\text{ GeV}$, this nuisance is mainly causing event migrations across the m_{jj} boundary. Similar behaviour is observed for Jet-01 which accounts for migrations across jet bins.

Table A.4: Acceptance uncertainties for VBF signal events in VBF BDT categories from WG2 uncertainties and PDF. Uncertainties smaller than 0.1% are not reported.

Category	Pt25	Pt200	M_{jj} -60	M_{jj} -120	M_{jj} -350	M_{jj} -700	M_{jj} -1000	M_{jj} -1500	Jet-01	PDF
VBF-cat0	0.984/1.016	—	1.011/0.989	1.023/0.977	—	1.002/0.998	—	—	0.995/1.005	1.002/0.997
VBF-cat1	0.991/1.009	—	1.013/0.987	1.018/0.982	0.998/1.002	1.003/0.997	—	—	0.994/1.006	1.003/0.996
VBF-cat2	1.003/0.997	—	1.015/0.985	1.007/0.993	0.997/1.003	1.004/0.996	—	—	0.994/1.007	1.001/0.999
VBF-cat3	1.012/0.988	—	1.019/0.981	0.999/1.001	0.996/1.004	1.005/0.995	—	—	0.992/1.008	0.998/1.002

The acceptance uncertainties affecting the expected signal yields per-category for the remaining production processes (WH, ZH, and ttH), coming from QCD-scale and PDF variations, have been also estimated. Table A.5 summarizes the aforementioned per-category uncertainties for the VBF, VH, and ttH production modes. In the final signal extraction, QCD-scale and PDF acceptance uncertainties are modelled via log-Normal nuisance parameters, and their effect is correlated across categories and eras. For the ggH process, PDF uncertainties are mostly anti-correlated between ggH and VBF categories, while for the other modes the effect appears to be correlated.

Table A.5: Acceptance uncertainties for VBF, WH, ZH, and ttH signal events in VBF BDT categories from QCD-scale and PDF variations. Uncertainties smaller than 0.1% are not reported. Uncertainties for WH, ZH, and ttH processes in the last two VBF categories are not estimated because the corresponding expected signal are smaller than 0.01 as reported in Table A.2.

Category	QCD-scale uncertainty			PDF uncertainty		
	WH	ZH	ttH	WH	ZH	ttH
VBF-cat0	1.035/0.990	1.032/0.985	0.991/1.006	1.003/0.997	1.003/0.997	0.999/1.001
VBF-cat1	1.058/0.995	1.049/0.992	0.986/1.007	1.002/0.997	1.002/0.997	0.999/1.001
VBF-cat2	—	—	—	—	—	—
VBF-cat3	—	—	—	—	—	—

Uncertainties related to the parton shower model are derived by varying the per-event weights stored internally in each signal sample. Variations are those related to the CP5 tune [25] of

3273 PYTHIA (v8.2 or greater). Table A.6 reports the parton shower related uncertainties in each
 3274 category defined along the VBF BDT for the different Higgs boson production modes. Uncer-
 3275 tainties are fully correlated across processes and eras. The parton shower uncertainty is small,
 3276 at the level of 2–3%, for the VBF production, while for ggH it ranges between 5–10% moving
 3277 from high to low score. The parton shower effect is anti-correlated between low-score ggH cat-
 3278 egories and VBF ones, which means that the associated nuisance parameter primarily induces
 3279 migration of events across N_{jet} bins.

Table A.6: Acceptance uncertainties for ggH, VBF, WH, ZH, and ttH signal events in VBF BDT cate-
 gories from variations of the parton shower tune parameters. Uncertainties smaller than 0.1% are not
 reported. Uncertainties for WH, ZH, and ttH processes in the last two VBF categories are not estimated
 because the corresponding expected signal are smaller than 0.01 as reported in Table A.2.

Category	ggH	VBF	WH	ZH	ttH
VBF-cat0	0.850/1.105	0.939/1.032	0.855/1.065	0.883/1.068	0.868/1.015
VBF-cat1	0.815/1.095	0.985/1.009	0.872/1.091	0.857/1.047	0.881/1.046
VBF-cat2	0.895/1.054	1.010/0.986	—	—	—
VBF-cat3	0.910/1.045	1.058/0.950	—	—	—

3280 Finally, the uncertainty related to variations of the underlying event tune is estimated by using
 3281 the alternative signal samples listed in Section 2.2. Values of this uncertainty in each category
 3282 and for each production mode are reported in Table A.7, and are fully correlated across pro-
 3283 cesses and eras. In order to minimize the impact of the limited size of MC statistics available
 3284 for the signal samples generated with the Up/Down variations of the underlying event tune,
 3285 the observed variations in the signal yield have been symmetrized. For both ggH and VBF
 3286 processes, the estimated uncertainty ranges between $\pm 3\%$ across VBF BDT categories.

Table A.7: Acceptance uncertainties for ggH, VBF, WH, ZH, and ttH signal events in VBF BDT cate-
 gories from variations of the underlying event tune parameters. Uncertainties smaller than 0.1% are not
 reported. Uncertainties for WH, ZH, and ttH processes in the last two VBF categories are not estimated
 because the corresponding expected signal are smaller than 0.01 as reported in Table A.2.

Category	ggH	VBF	WH	ZH	ttH
VBF-cat0	0.997/1.003	1.010/0.990	0.978/1.022	0.991/1.009	0.991/1.009
VBF-cat1	0.975/1.025	0.994/1.006	1.036/0.964	1.038/0.962	0.930/1.070
VBF-cat2	0.972/1.028	0.996/1.004	—	—	—
VBF-cat3	1.017/0.983	1.021/0.979	—	—	—

3287 A.5.3 Rate uncertainties from experimental sources

3288 Several experimental sources of uncertainty affect the expected signal rate in each event cate-
 3289 gory and for each production mode. A first set is represented by muon-related sources. Un-
 3290 certainties in the knowledge of muon identification and isolation efficiencies, as well those
 3291 related to the single muon trigger efficiency, are propagated in the analysis assuming full cor-
 3292 relation between the two muons in the event. Their effects are modelled via two independent
 3293 log-Normal nuisance parameters correlated across processes, categories, and eras. Table A.8
 3294 reports their values described for each production mode as well as for each BDT category. The
 3295 uncertainty in the muon identification and isolation (single muon trigger) scale factors trans-
 3296 lates into an overall $\pm 1\%$ ($\pm 0.2\%$) uncertainty in the signal in each category. Moreover, the
 3297 effect of the uncertainty in the knowledge of the muon p_T coming from the Rochester correc-
 3298 tions is also taken into account and modelled via a log-Normal nuisance parameter correlated
 3299 across processes, categories, and eras. Table A.9 reports the estimate for this uncertainty for
 3300 each production mode as well as for each BDT category. The muon p_T uncertainty induces a

3301 few per-mile effect mostly anti-correlated as a function of the dimuon p_T spectrum, i.e. anti-
 3302 correlated moving from low to high score categories of both ggH and VBF BDTs.

Table A.8: Uncertainties for ggH, VBF, WH, ZH, and ttH processes in VBF BDT categories obtained from the uncertainties related to muon identification and isolation scale factors, and single muon trigger scale factors. Uncertainties smaller than 0.1% are not reported. Uncertainties for WH, ZH, and ttH processes in the last two VBF categories are not estimated because the corresponding expected signal are smaller than 0.01 as reported in Table A.2.

Category	Muon ID and Iso uncertainty					Muon trigger uncertainty				
	ggH	VBF	WH	ZH	ttH	ggH	VBF	WH	ZH	ttH
VBF-cat0	0.992/1.008	0.992/1.008	0.992/1.008	0.992/1.008	0.990/1.010	0.998/1.002	0.998/1.002	0.998/1.002	0.998/1.002	0.998/1.002
VBF-cat1	0.991/1.009	0.991/1.009	0.990/1.010	0.990/1.010	0.990/1.010	0.999/1.001	0.998/1.002	0.998/1.002	0.998/1.002	0.998/1.002
VBF-cat2	0.990/1.010	0.990/1.010	—	—	—	0.998/1.002	0.998/1.002	—	—	—
VBF-cat3	0.990/1.010	0.991/1.009	—	—	—	0.998/1.002	0.998/1.002	—	—	—

Table A.9: Uncertainties for ggH, VBF, WH, ZH, and ttH processes in VBF BDT categories obtained from the uncertainties related to muon energy scale. Uncertainties smaller than 0.1% are not reported. Uncertainties for WH, ZH, and ttH processes in the last two VBF categories are not estimated because the corresponding expected signal are smaller than 0.01 as reported in Table A.2.

Category	Muon scale uncertainty				
	ggH	VBF	WH	ZH	ttH
VBF-cat0	1.001/0.999	1.002/0.999	1.001/0.999	1.001/0.999	1.002/0.999
VBF-cat1	1.001/0.998	1.001/0.999	0.997/1.003	0.998/1.002	0.996/1.002
VBF-cat2	0.995/1.003	0.998/1.002	—	—	—
VBF-cat3	0.998/1.008	0.998/1.003	—	—	—

3303 The dominant experimental uncertainties are represented by the limited knowledge in the en-
 3304 ergy scale and resolution of the hadronic jets in the event, especially for those belonging to the
 3305 forward region of the detector ($|\eta| > 2.5$). Each source of uncertainty is varied independently
 3306 up and down in order to change the p_T of all the jets found in the event. The analysis selec-
 3307 tions are re-applied after each variation, and the ggH BDT is re-evaluated to account for the
 3308 corresponding changes in the jet kinematics. Variations in the expected signal rate in each cat-
 3309 egory are modelled via Log-Normal nuisance parameters, whose effects are correlated across
 3310 categories, signal production modes, and eras (depending on the uncertainty source). Simi-
 3311 larly, the uncertainty in the jet energy resolution is used to smear the p_T of the jets in simulated
 3312 events in the six exclusive categories listed in Section 5.4. The corresponding variations in the
 3313 expected signal rate are modelled via independent log-Normal nuisance parameters, whose
 3314 effect is correlated across categories, signal processes, but not correlated across eras. The un-
 3315 certainties related to jet-energy-resolution are found to be one-sided because we don't apply
 3316 the resolution scale-factors (smearing) to correct for the central value, but energy smearing is
 3317 applied only to estimate an associated uncertainty (conservative).

3318 Therefore, following the recipe from JetMET POG, twenty-one nuisance parameters are con-
 3319 sidered in the final fit which are either correlated or uncorrelated across eras. Table A.10 and
 3320 Table A.11 list for each BDT category the estimated values for each jet-energy-scale and jet-
 3321 energy-resolution uncertainty belonging to the reduced set of sources for the ggH and VBF pro-
 3322 cesses, respectively. Uncertainties in the jet energy scale cause significant migration of events
 3323 between the ggH and the VBF categories, therefore the effect of these nuisances is found to
 3324 be largely anti-correlated across N_{jet} bins. The sources showing the largest impact on the pre-
 3325 dicted signal yields are those correlated across eras and not specific to a particular detector
 3326 region: flavorQCD, relativeBal, relativeSample, and absolute. According to Table A.10, the to-
 3327 tal uncertainty due to jet-energy-scale affecting the ggH process ranges between 10–15% in the
 3328 VBF categories. In contrast, the total uncertainty on the predict VBF signal yields due to jet-
 3329 energy-scale is found to be between 3–5%. On the other hand, the uncertainty due to jet energy

resolution has a slightly smaller impact on the predicted signal yields. In the VBF categories, the total jet-energy-resolution uncertainty affecting the ggH (VBF) prediction ranges between 8–5% (4–6%) as a function of the BDT score.

The jet energy scale and resolution uncertainties affecting the other production modes (WH, ZH, and ttH) are not reported in this note because they have a small impact on the final result. However, they are included in the “datacards” used in the final fit for the signal extraction and, overall, their size is similar to what has been measured for the ggH process and reported in Table A.10.

Table A.10: Uncertainties for the ggH process in the VBF BDT categories obtained by propagating the effect of each source of uncertainty related to jet energy scale and jet energy resolution to the expected signal yields. Uncertainties smaller than 0.1% are not reported.

Source of Uncertainty	VBF Categories			
	cat0	cat1	cat2	cat3
Jet energy scale				
flavorQCD	0.957/1.044	0.974/1.035	0.929/1.023	0.992/1.042
relativeBal	1.047/0.957	1.027/0.969	1.047/0.951	1.042/0.951
relativeSample	1.062/0.940	1.055/0.964	1.053/0.940	1.066/0.973
absolute	0.934/1.074	0.948/1.051	0.935/1.032	0.947/1.071
BBEC1	0.997/1.006	0.993/1.002	0.993/1.001	0.998/1.006
EC2	0.994/1.007	0.993/1.005	1.001/0.995	0.978/1.002
HF	0.970/1.032	0.983/1.026	0.952/1.009	0.993/1.045
relativeSample_2016	0.982/1.015	0.996/1.010	0.983/1.009	0.998/1.016
absolute_2016	0.993/1.003	0.999/1.002	—	0.998/1.021
BBEC1_2016	0.998/1.001	0.996/1.001	0.994/1.001	0.999/1.005
EC2_2016	0.995/1.005	0.998/1.001	0.999/1.003	0.990/1.004
HF_2016	0.997/1.001	—	1.001/0.997	0.995/1.013
relativeSample_2017	0.981/1.020	0.988/1.014	0.968/1.013	0.995/1.022
absolute_2017	0.991/1.010	0.993/1.003	0.978/1.008	0.999/1.014
BBEC1_2017	—	0.997/1.001	0.996/1.001	0.999/1.001
EC2_2017	0.994/1.006	0.995/1.003	0.997/1.009	0.991/1.002
HF_2017	0.997/1.003	0.996/1.001	0.998/1.001	0.999/1.006
relativeSample_2018	0.956/1.049	0.968/1.042	0.934/1.028	0.983/1.050
absolute_2018	0.990/1.011	0.989/1.007	0.981/1.020	1.011/0.983
BBEC1_2018	0.996/1.005	0.999/1.003	0.988/1.002	1.003/0.993
EC2_2018	0.991/1.008	0.998/1.004	0.992/1.001	0.999/1.008
HF_2018	0.996/1.005	0.994/1.003	0.987/1.013	1.008/0.997
Jet energy resolution	cat0	cat1	cat2	cat3
jerBarrel	—/1.007	—/0.998	—/0.986	—/1.012
jerEC1	—/1.010	—/1.006	—/0.986	—/1.013
jerEC2low	—/1.068	—/1.038	—/1.009	—/1.007
jerEC2high	—/1.002	—/1.002	—/0.983	—/0.977
jerHFlow	—/1.026	—/1.025	—/1.014	—/0.998
jerHFhigh	—/1.001	—/0.998	—/0.997	—/0.996

Other sources of uncertainty are represented by the uncertainty in the ECAL pre-firing correc-

Table A.11: Uncertainties for the VBF process in the VBF BDT categories obtained by propagating the effect of each source of uncertainty related to jet energy scale and jet energy resolution to the expected signal yields. Uncertainties smaller than 0.1% are not reported.

Source of Uncertainty	VBF Categories			
	cat0	cat1	cat2	cat3
Jet energy scale				
flavorQCD	0.985/1.023	0.986/1.008	0.984/1.012	0.982/1.011
relativeBal	1.023/0.983	1.001/0.995	1.014/0.982	1.017/0.985
relativeSample	1.022/0.984	1.021/0.984	1.006/0.975	1.022/0.972
absolute	0.978/1.030	0.984/1.013	0.981/1.017	0.983/1.019
BBEC1	0.999/1.005	0.998/1.001	0.997/1.001	0.999/1.002
EC2	0.997/1.004	0.997/1.001	—	0.999/1.002
HF	0.990/1.015	0.994/1.004	0.987/1.010	0.996/1.007
relativeSample_2016	0.990/1.009	0.999/1.005	0.990/1.001	0.999/1.004
absolute_2016	0.999/1.003	—	0.999/1.004	0.998/1.021
BBEC1_2016	—	—	0.998/1.002	—
EC2_2016	1.000/1.003	0.997/0.999	0.998/0.998	0.999/1.003
HF_2016	—	—	0.999/1.003	1.001/0.999
relativeSample_2017	0.993/1.009	0.997/1.003	0.991/1.005	0.990/1.007
absolute_2017	0.997/1.005	0.998/1.000	0.994/1.005	0.997/1.002
BBEC1_2017	0.999/1.001	—	—	—
EC2_2017	0.998/1.002	0.997/1.001	0.999/1.001	0.998/1.002
HF_2017	0.997/1.003	0.996/1.001	0.998/1.001	—
relativeSample_2018	0.985/1.020	0.991/1.012	0.987/1.010	0.982/1.011
absolute_2018	0.997/1.005	0.996/1.002	0.998/1.004	0.998/1.001
BBEC1_2018	0.998/1.002	—	0.997/1.002	—
EC2_2018	0.997/1.004	0.998/1.001	0.999/1.002	0.997/1.001
HF_2018	0.999/1.002	0.999/1.001	0.999/1.003	0.999/1.002
Jet energy resolution	cat0	cat1	cat2	cat3
jerBarrel	—/1.007	—/0.998	—/0.986	—/1.012
jerEC1	—/1.010	—/1.006	—/0.986	—/1.013
jerEC2low	—/1.068	—/1.038	—/1.009	—/1.007
jerEC2high	—/1.002	—/1.002	—/0.983	—/0.977
jerHFlow	—/1.026	—/1.025	—/1.014	—/0.998
jerHFhigh	—/1.001	—/0.998	—/0.997	—/0.996

tion introduced to emulate an inefficiency in the L1 trigger efficiency affecting 2016 and 2017 eras, and the uncertainty in the minimum bias cross section which affects the pileup re-weight procedure described in Section 4.1. The impact of the pre-firing uncertainty increases as a function of the VBF BDT output from 0.4% to 0.8%. In fact, in the high score region of the VBF BDT a good fraction of signal events have one energetic jet in the endcap region. The pileup uncertainty instead is found to be quite flat as a function of VBF categories, ranging between 0.1% and 0.5%. Table A.12 summarizes the estimated value of pre-firing and pileup uncertainties for each production mode and for each category. Finally, the uncertainty on the luminosity measurement is of about 2.5% in each era, and is modelled via a log-Normal nuisance parameter

3348 whose effect is correlated across BDT categories and processes.

Table A.12: Uncertainties for ggH, VBF, WH, ZH, and ttH processes in the VBF BDT categories obtained from the uncertainties related to pre-firing correction and pileup re-weight. Uncertainties smaller than 0.1% are not reported. Uncertainties for WH, ZH, and ttH processes in the last two VBF categories are not estimated because the corresponding expected signal are smaller than 0.01 as reported in Table A.2.

Category	Pileup re-weight					Pre-firing				
	ggH	VBF	WH	ZH	ttH	ggH	VBF	WH	ZH	ttH
VBF-cat0	0.997/1.003	0.999/1.001	0.993/1.007	0.994/1.006	1.003/0.995	1.004/0.996	1.006/0.994	1.006/0.994	1.006/0.994	1.011/0.989
VBF-cat1	0.996/1.004	1.003/0.997	0.991/1.009	0.996/1.005	1.006/0.993	1.004/0.996	1.005/0.995	1.006/0.994	1.005/0.995	1.012/0.987
VBF-cat2	0.990/1.010	1.003/0.997	—	—	—	1.006/0.993	1.007/0.993	—	—	—
VBF-cat3	0.998/1.004	1.003/0.996	—	—	—	1.006/0.994	1.008/0.992	—	—	—

3349 A.6 Background estimation

3350 Given the low signal-to-background ratio expected in most of the event categories defined in
 3351 Section A.3, the key aspect of this search is represented by a robust estimation of the $m_{\mu\mu}$ shape
 3352 of the total background. The expected background in each category is mainly composed of
 3353 Z/γ^* +jets events, with smaller contaminations arising from $t\bar{t}$, single-top, and diboson pro-
 3354 cesses. In the high score region of the VBF BDT, the contribution from the electroweak Z pro-
 3355 duction becomes also relevant. Table A.13 reports the expected background composition from
 3356 simulation in each category. Processes involving dimuon decays of a Z boson represent more
 3357 than 95% of the total expected background across the VBF phase space.

Table A.13: Background composition in (%) expected in the VBF BDT categories for events in the signal fit region, $110 < m_{\mu\mu} < 150$ GeV. Event yields are normalized to the measured luminosity of 137.2 fb^{-1} .

Category	Z/γ^* +jets	Z +jets EW	$t\bar{t}$	Diboson	Other (s-top, VVV, etc)
VBF-cat0	84.9%	1.6%	10.0%	2.2%	1.2%
VBF-cat1	83.2%	9.6%	4.8%	1.7%	0.6%
VBF-cat2	70.8%	24.6%	2.9%	1.2%	0.4%
VBF-cat3	43.6%	54.3%	1.2%	0.1%	—

3358 As described in Section 2.1, simulated samples for the main background processes have been
 3359 generated with NLO QCD precision but they cannot be directly used to model the background
 3360 for several reasons:

- 3361 • Simulated samples are missing higher order corrections in both QCD (NNLO) and
 3362 EW (NLO) theory which are known to have a sizable impact on the predicted dimuon
 3363 mass spectrum.
- 3364 • Theory uncertainties due to resummation effects, QCD scale variations, and PDF
 3365 are typically large enough to be over-constrained by data. They primarily provide
 3366 an acceptance uncertainty in each category but they cannot be sufficiently trusted to
 3367 provide a shape uncertainty for the background $m_{\mu\mu}$ spectrum.

3368 The $m_{\mu\mu}$ distribution for the total background in each category is expected to be smoothly
 3369 falling, driven by the Breit-Wigner nature of the Z boson lineshape. In this context, we choose
 3370 to estimate the background via a data driven approach, smoothing the dimuon mass spectrum
 3371 in the signal fit region ($110 < m_{\mu\mu} < 150$ GeV) via empirical functional forms. According to
 3372 this strategy, the background contribution under the peak is therefore estimated from signal-
 3373 free mass sidebands. We assume no prior knowledge of the parameters of the fit functions
 3374 (in both shape and normalization), i.e. they are allowed to float freely in the fit to the data.
 3375 Therefore, the uncertainty in the estimate of the background under the peak is mainly affected
 3376 by the statistical power of the mass sidebands, as well as on the assumption of the background

functional form. The functions used have been already defined in Section 5.4, they can be either Breit-Wigner-like parametrizations (physics inspired) or agnostic ones.

The number of degrees of freedom for the polynomial terms used to modulate some of the physics-inspired functions, as well as the order of agnostic models like Bernstein polynomials, series of power-laws and exponentials, are selected by means of a likelihood ratio test as recommended by the CMS statistics committee [75]. This test works as follows: assuming that, within a given family of functions, n parameters are enough to describe the $m_{\mu\mu}$ spectrum in a chosen category, we first fit the $m_{\mu\mu}$ distribution in data with n and $n + 1$ degrees of freedom. Then, we compute $2 \cdot (\text{Log}\mathcal{L}_{n+1} - \text{Log}\mathcal{L}_n)$ and, under the assumption that this quantity is distributed according to the χ^2 distribution with one degree of freedom, the corresponding p-value is used to accept or reject the null hypothesis. If the p-value is smaller than 5%, the null hypothesis is rejected and we apply the same test to the function with $n + 1$ parameters. In contrast, if the p-value is larger than 5%, a good compatibility is observed between the two models, therefore the function with n degrees of freedom is selected.

In the cases of $\text{BWZ} \times \text{Bernstein}(m_{\mu\mu})$, $\text{FEWZ} \times \text{Bernstein}(m_{\mu\mu})$, and $\text{Landau} \times \text{Bernstein}(m_{\mu\mu})$, this test indicates that, depending on the category, either one or two polynomial degrees of freedom are enough to have an accurate description of the $m_{\mu\mu}$ spectrum. On the other hand, for what concerns sum of power-laws and exponentials, a series of order two ($n = 2$) is selected via the likelihood ratio test. Finally, when Bernstein polynomials are used to fit the $m_{\mu\mu}$ spectrum in each category, the likelihood ratio test suggests up to four degrees of freedom are necessary in order to have a good description of the observed spectra. Given that all the other families require about three shape degrees of freedom per category, plain Bernstein polynomials appear as an unnatural set of functions for modelling the total expected background, and therefore they are not considered in the studies presented in later Sections.

Table A.14 reports for each category defined along the VBF BDT output and for each family of functions the chosen number of shape degrees of freedom along with the χ^2/dof obtained from a background-only fit to the data. In each category and for each function the associated χ^2 is computed with the binning reported in Fig. A.18. Moreover, in order to correctly account for the low statistics regime of the high score VBF categories, the Baker-Cousins χ^2 definition is adopted [76]. In each category, the χ^2/dof is about one for the chosen functions, indicating that they can well describe the dimuon mass spectrum in the range $110 < m_{\mu\mu} < 150 \text{ GeV}$. Moreover, Fig. A.18 show the results of b-only fits performed independently in the VBF categories with the functions listed in Table A.14. The lower pad in each plot reports the ratio between the background prediction obtained with the aforementioned functions and the one coming from a fit via the BWZRedux.

Table A.14: Number of shape degrees of freedom selected, in each category and for each family of functions previously described, by means of a likelihood ratio test. The χ^2/dof obtained for each function from a background-only fit to the data in the range $110 < m_{\mu\mu} < 150 \text{ GeV}$ and computed following the recommendations of Ref. [76] is also reported.

Category	$\text{BWZ} \times \text{Bernstein}$	BWZGamma	BWZRedux	$\text{FEWZ} \times \text{Bernstein}$	$\text{Landau} \times \text{Bernstein}$	S-exponential	S-power-law
VBF-cat0	3-dof, $\hat{\chi}^2 = 1.11$	2-dof, $\hat{\chi}^2 = 1.10$	3-dof, $\hat{\chi}^2 = 1.11$	2-dof, $\hat{\chi}^2 = 1.11$	3-dof, $\hat{\chi}^2 = 1.13$	3-dof, $\hat{\chi}^2 = 1.10$	3-dof, $\hat{\chi}^2 = 1.11$
VBF-cat1	3-dof, $\hat{\chi}^2 = 0.85$	2-dof, $\hat{\chi}^2 = 0.84$	3-dof, $\hat{\chi}^2 = 0.85$	2-dof, $\hat{\chi}^2 = 0.84$	3-dof, $\hat{\chi}^2 = 0.87$	3-dof, $\hat{\chi}^2 = 0.85$	3-dof, $\hat{\chi}^2 = 0.85$
VBF-cat2	3-dof, $\hat{\chi}^2 = 1.02$	2-dof, $\hat{\chi}^2 = 1.01$	3-dof, $\hat{\chi}^2 = 1.02$	2-dof, $\hat{\chi}^2 = 1.01$	3-dof, $\hat{\chi}^2 = 1.04$	3-dof, $\hat{\chi}^2 = 1.02$	3-dof, $\hat{\chi}^2 = 1.02$
VBF-cat3	2-dof, $\hat{\chi}^2 = 1.03$	2-dof, $\hat{\chi}^2 = 1.02$	3-dof, $\hat{\chi}^2 = 1.03$	2-dof, $\hat{\chi}^2 = 1.02$	2-dof, $\hat{\chi}^2 = 1.03$	3-dof, $\hat{\chi}^2 = 1.03$	3-dof, $\hat{\chi}^2 = 1.03$

In each category and for each function selected by the F-test and reported in Table A.14, it is important to check the behaviour of the post-fit shape uncertainty on the total background as a function of $m_{\mu\mu}$, as obtained from a background-only fit to the data. These comparisons are reported in Appendix of Ref. [17]. As expected, the relative size of the post-fit shape uncertainties is proportional to the amount of data events in the category as well as to the number

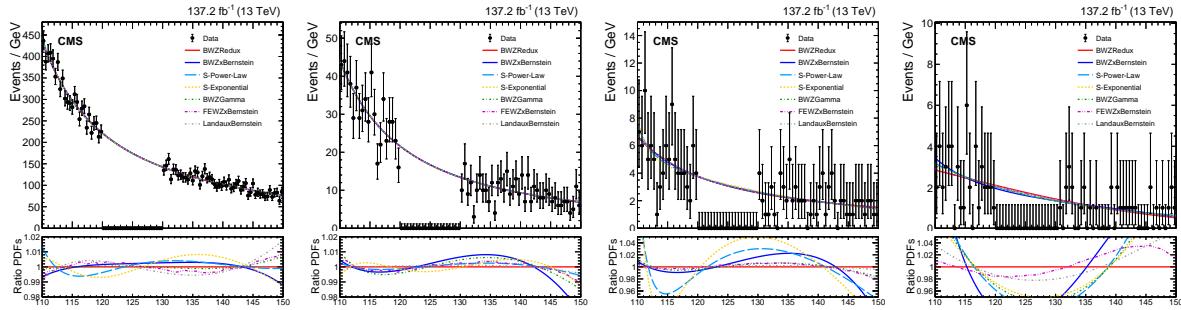


Figure A.18: Dimuon invariant mass distributions in data for events belonging to the VBF BDT categories. The solid lines correspond to alternative background predictions obtained by fitting the mass spectra, under the background-only hypothesis, with the functional forms listed in Table A.14. From left to right: VBF-cat0, VBF-cat1, VBF-cat2, and VBF-cat3. The lower pad in each plot reports the ratio between the background prediction obtained with each function and the one coming from a fit via the BWZRedux.

3417 degrees of freedom of the chosen analytical function. Breit-Wigner and power-law based function
 3418 not modulated by polynomials show a “pole” in the mass region between 120–130 GeV,
 3419 therefore they are expected to lead to a better performance in terms of significance. On the
 3420 other hand, functions like FEWZ \times Bernstein, BWZ \times Bernstein, and sum of exponentials are
 3421 showing a shape uncertainty less dependent on $m_{\mu\mu}$ depending on the number of polynomial
 3422 degrees of freedom. However, it is important to remind that background shape uncertainties
 3423 have a “important” effect only in categories with high data statistics, i.e. those for which the
 3424 total background normalization can be precisely constrained by the $m_{\mu\mu}$ sidebands. In contrast,
 3425 the performance of the high score VBF categories (VBF-cat1, VBF-cat2, and VBF-cat3) is mostly
 3426 limited by the large uncertainty on the total background prediction, rather than to the choice
 3427 of the background functional shape and the related shape uncertainty.

3428 Finally, Table A.15 reports the expected significances in each category for a SM Higgs boson
 3429 with $m_H = 125$ GeV, extracted from a s+b fit to Asimov toys in which the background predictions
 3430 are taken from b-only fits to data with the functions reported in Table A.14. Similarly,
 3431 Table A.16 reports the expected significances in each category extracted from s+b fit to Asimov
 3432 toys in which the background predictions are taken from simulated events. The expected
 3433 significance obtained from simulated events is used as figure of merit when choosing of the
 3434 best performing functional form in each category. The prediction from simulation is found to
 3435 be in good agreement with data as shown in Fig. A.13. This reflects in very similar expected
 3436 significances obtained in each category and for each functional form. Across both VBF
 3437 categories, BWZRedux, and S-exponential are showing performances compatible within $\pm 3\%$. The
 3438 BWZGamma improves the performance of the BWZRedux by about 10% in VBF-cat3. The
 3439 FEWZ \times Bernstein instead shows the same performance of BWZRedux across the VBF
 3440 categories. The background parametrizations which provide the best expected significance are
 3441 represented by the BWZ \times Bernstein and the S-PowerLaw, whose performance improve the
 3442 one obtained with the BWZRedux function by about 1–8% depending on the categories. These
 3443 results, along with the outcome of the bias studies presented in Section A.7.2, will be used to
 3444 choose the background fit function in each category. In fact, in each category, the function pro-
 3445 viding the best expected significance without inducing a significant bias in the estimate of the
 3446 signal strength is the one chosen for the final fit.

Table A.15: A posteriori expected significance in each category and for each family of functions obtained for a SM Higgs boson with $m_H = 125\text{ GeV}$. Systematic uncertainties on the signal prediction have not be added into the datacards used to extract these significances, however their impact is expected to be quite small at the level of few %.

Category	BWZ × Bernstein	BWZGamma	BWZRedux	FEWZ × Bernstein	Landau × Bernstein	S-exponential	S-power-law
VBF-cat0	$\sigma = 0.41$	$\sigma = 0.39$	$\sigma = 0.39$	$\sigma = 0.39$	$\sigma = 0.39$	$\sigma = 0.38$	$\sigma = 0.38$
VBF-cat1	$\sigma = 0.62$	$\sigma = 0.60$	$\sigma = 0.60$	$\sigma = 0.61$	$\sigma = 0.60$	$\sigma = 0.60$	$\sigma = 0.60$
VBF-cat2	$\sigma = 0.74$	$\sigma = 0.75$	$\sigma = 0.75$	$\sigma = 0.75$	$\sigma = 0.75$	$\sigma = 0.72$	$\sigma = 0.78$
VBF-cat3	$\sigma = 1.08$	$\sigma = 1.11$	$\sigma = 0.98$	$\sigma = 1.00$	$\sigma = 1.00$	$\sigma = 1.11$	$\sigma = 1.12$
Combined VBF	$\sigma = 1.51$	$\sigma = 1.52$	$\sigma = 1.43$	$\sigma = 1.44$	$\sigma = 1.45$	$\sigma = 1.50$	$\sigma = 1.51$

Table A.16: A posteriori expected significance in each category and for each family of functions obtained for a SM Higgs boson with $m_H = 125\text{ GeV}$. Systematic uncertainties on the signal prediction have not be added into the datacards used to extract these significances, however their impact is expected to be quite small at the level of few %.

Category	BWZ × Bernstein	BWZGamma	BWZRedux	FEWZ × Bernstein	Landau × Bernstein	S-Exponential	S-PowerLaw
VBF-cat0	$\sigma = 0.37$	$\sigma = 0.39$	$\sigma = 0.39$	$\sigma = 0.39$	$\sigma = 0.38$	$\sigma = 0.38$	$\sigma = 0.39$
VBF-cat1	$\sigma = 0.63$	$\sigma = 0.61$	$\sigma = 0.61$	$\sigma = 0.61$	$\sigma = 0.60$	$\sigma = 0.60$	$\sigma = 0.61$
VBF-cat2	$\sigma = 0.72$	$\sigma = 0.70$	$\sigma = 0.70$	$\sigma = 0.70$	$\sigma = 0.70$	$\sigma = 0.69$	$\sigma = 0.70$
VBF-cat3	$\sigma = 1.21$	$\sigma = 1.22$	$\sigma = 1.08$	$\sigma = 1.09$	$\sigma = 1.07$	$\sigma = 1.15$	$\sigma = 1.18$
Combined VBF	$\sigma = 1.51$	$\sigma = 1.54$	$\sigma = 1.47$	$\sigma = 1.48$	$\sigma = 1.46$	$\sigma = 1.49$	$\sigma = 1.50$

3447 A.7 Uncertainty on the background model

3448 As presented in Table A.14, the functional forms used to model the dimuon mass distribution
 3449 in background events belong to two main classes: physics-inspired models and agnostic ones.
 3450 Within these families, the choice of particular analytical function to model the background in
 3451 a given category may potentially inject a bias in the estimated signal strength, when a signal
 3452 plus background fit is performed. Therefore, as done in earlier iteration of this analysis [13] as
 3453 well as in other searches following a similar approach for the signal extraction [77], a detailed
 3454 set of *bias studies* are performed.

3455 A.7.1 Toy generation for bias studies

3456 The first step in performing bias studies for the signal extraction consists in generating toy ex-
 3457 periments for each category for each functional form selected as a possible *truth* function that
 3458 describes the $m_{\mu\mu}$ spectrum in background events. Starting from the background shapes re-
 3459 ported in Fig. A.18, which have been obtained from a b-only fit to the data, and the expected
 3460 signal shapes shown in Fig. A.15, 2000 toys are generated for each background functional form
 3461 in each category. The number of toys is chosen in order to have an uncertainty of about 2%
 3462 on the estimated bias. Toys are generated by injecting an amount of signal event proportional
 3463 to the expected rate in each category for $m_H = 125\text{ GeV}$. The same amount of toy data have
 3464 been generated also for other Higgs boson masses of 120 and 130 GeV. Since we are only in-
 3465 terested in estimating the bias on the signal strength induced by the assumption of a specific
 3466 background shape, systematic uncertainties affecting the signal prediction, described in Sec-
 3467 tion A.5.2 and A.5.3 are not considered in the bias studies.

3468 A.7.2 Bias of the various background functions

The dimuon mass distributions obtained from the toy-experiments previously described are fitted with a signal plus background model, where the background function is one of those reported in Table A.14. In each fit, the bias on the signal strength is defined as follows:

$$\text{Bias}(\mu_{fit}, \hat{\mu}_{inj}, \sigma_\mu) = \frac{\mu_{fit} - \hat{\mu}_{inj}}{\sigma_\mu} \quad (\text{A.2})$$

where μ_{fit} is the value of the signal strength extracted from the fit, $\hat{\mu}_{inj}$ is the signal strength injected in each toy ($\hat{\mu}_{inj} = 1$)⁷, and σ_μ is its post-fit uncertainty estimated as the crossing at $-2\Delta\text{Log}\mathcal{L} = 1$ of the likelihood scan as a function of μ . In absence of systematic effects, the bias is expected to be Gaussian distributed, centred at zero, and with a width of one. The mean value of the bias distribution is used to quote the bias associated to a given fit function. Good functional forms, which can be used in the final signal extraction, are those characterized by a small bias on the signal strength in each category (< 20% as defined in Equation A.2). The threshold of 20% is chosen because such bias guarantees that the systematic uncertainty on the quoted upper limit or on the expected significance is smaller than 2%.

Figure A.19 shows the measured bias on the signal strength for $m_H = 125$ GeV in the VBF categories, obtained by fitting the toys generated via several truth function with BWZRedux, BWZGamma, FEWZ \times Bernstein, S-exponential, S-power-law, and Landau \times Bernstein. The number of degrees of freedom considered for each functional form and in each category are listed in Table A.14. In contrast with the ggH categories, no large tension is observed between sum-of-exponentials and the other functions thanks to a better signal-to-background ratio, a smaller statistical constraint from events in the mass sidebands, and a larger contamination from combinatorial backgrounds that can be usually well described by a series of exponentials.

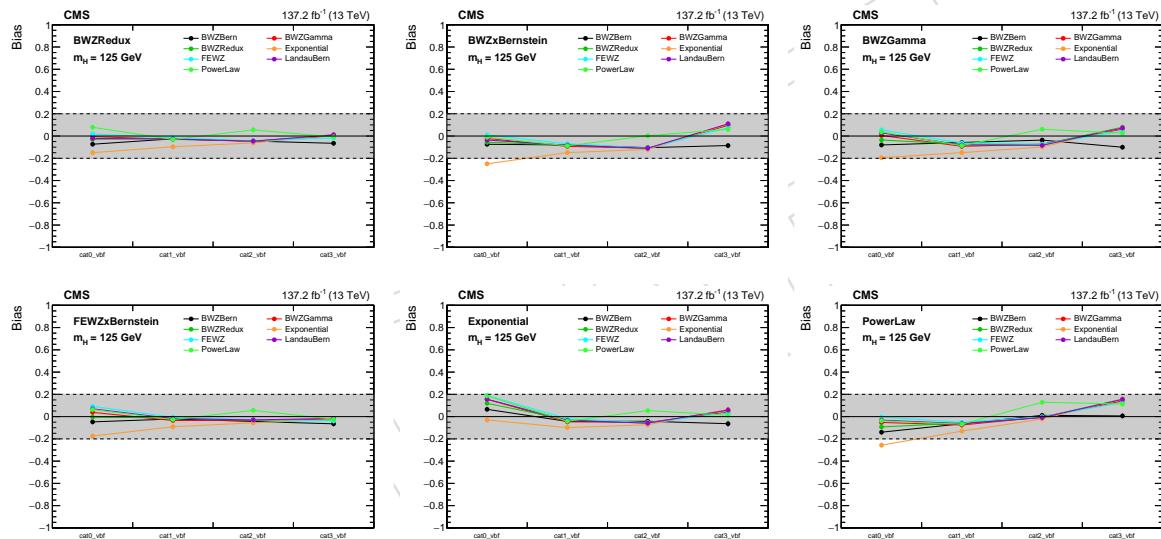


Figure A.19: Mean value of the bias distribution in VBF BDT categories obtained by fitting toys, generated according to different truth background models and injecting a nominal signal at $m_H = 125$ GeV, with the following functional forms: BWZRedux (top left), BWZxBernstein (top middle), BWZGamma (top right), FEWZ \times Bernstein (bottom right), S-exponential (bottom middle), and S-power-law (bottom left).

An additional concern is the potential bias on the signal strength extracted from the simultaneous fit performed across multiple categories (*correlated bias*). In principle, since each category represents an independent entity from the background estimate point of view, we would need to generate toys according to all possible combinations of truth background functions. However, having nine event categories and at least six alternative truth models makes this test computationally not feasible. Therefore, we consider only the cases in which the type of function used to generate and fit the toys is the same across all categories. For example, when the BWZRedux is chosen as the truth model, pseudo-experiments are thrown using this

⁷In each toy, the amount of signal events injected correspond to a Poissonian “smearing” around a mean value given by the expected signal rate for $\mu = 1$

background shape in all categories. On the other hand, when the BWZRedux is tested as fit function, pseudo-experiments generated according to every possible truth function are fitted with a signal-plus-background model in which the BWZRedux parametrization is used to describe the background in all categories. Figure A.20 shows the estimated bias on the signal strength for $m_H = 125$ GeV obtained when the VBF categories are simultaneously fitted using either a BWZRedux (left), a BWZGamma (middle), or a FEWZ \times Bernstein (right) function to model the background. When all VBF categories are simultaneously fitted no significant bias is observed for any of the truth model.

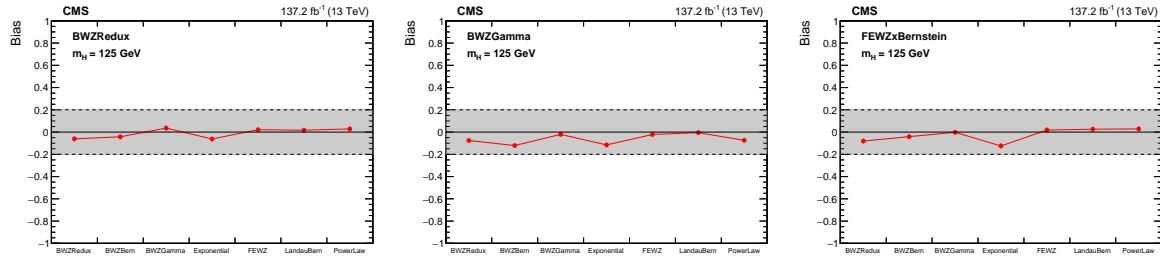


Figure A.20: Mean value of the bias distribution on the signal strength obtained by fitting toys, generated across all the VBF categories according to alternative background truth models and injecting a nominal signal at $m_H = 125$ GeV, with the following functional forms: BWZRedux (left), BWZGamma (middle), and FEWZ \times Bernstein (right).

In conclusion, the bias in VBF categories can be well controlled by using different alternative fit functions. Based on the performance listed in Table 6.14, as well as the observed bias reported in Fig. A.20 and Fig. A.20, the following functions are chosen in each VBF category to estimate the total background: BWZRedux in VBF-cat0 and VBF-cat1, while BWZGamma in VBF-cat2 and VBF-cat3.

Finally, Fig. A.21 shows the bias measured in toys where the signal has been injected in both ggH and VBF categories. Toys are fitted with a signal-plus-background model in which the presence of the signal is allowed in each categories, instead the background is estimated as follows:

- The *core-PDF* previously described is used to model the background in the ggH categories.
- In the VBF categories, the background is modelled via single and independent functions: BWZRedux in VBF-cat0 and VBF-cat1, while BWZGamma is used in VBF-cat2 and VBF-cat3.

The bias measured as a function of different truth functions can be controlled within 20%.

Finally, Appendix C shows results from bias studies performed in VBF categories for different mass points of $m_H = 120$ GeV and $m_H = 130$ GeV, which correspond to the smallest and largest Higgs boson mass hypothesis tested by this search, respectively. Overall, similar results and conclusions to those presented for $m_H = 125$ GeV are found.

A.8 Results

In order to extract the final results, binned maximum-likelihood fits are performed to the observed $m_{\mu\mu}$ distributions of each category, in the range $110 < m_{\mu\mu} < 150$ GeV, where the expected signal from the different production modes is scaled by common strength modifier (μ). Fits are performed fixing the expected Higgs boson mass to 125 GeV. Systematic uncertainties from theoretical and experimental sources, as described in Section A.4, are included in the like-

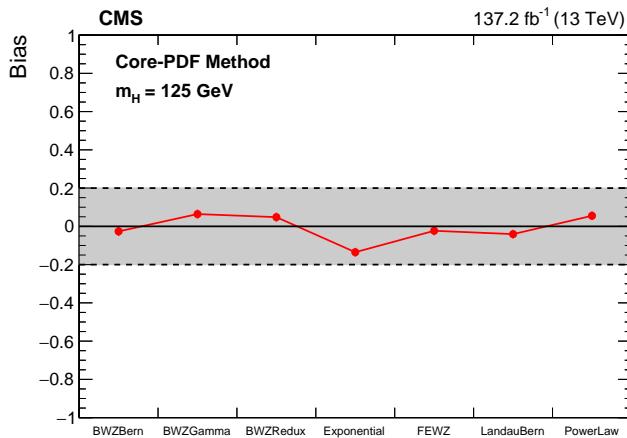


Figure A.21: Mean bias on the signal strength obtained from the *core-PDF* method used to model the background in all categories in toys where the signal has been injected in all categories simultaneously.

lihood via Log-Normal nuisance parameters. Binned fits are used to speed-up the computation, and the chosen bin size of 50 MeV is sufficiently small compared to the mass resolution that no information is lost. The data and the result from a signal-plus-background model performed simultaneously across categories are shown in Fig. A.22 for the VBF categories, respectively. The one (green) and two (yellow) standard deviation bands shown for the background component of the fit include the uncertainty in the fitted parameters. The background in the VBF categories the functions listed in Section A.7.2 are adopted. In addition, Fig. A.23 shows the background component and the signal-plus-background fit to the data combining ggH and VBF categories, weighting them by the expected ratio between signal and signal-plus-background expectations in each category.

Likelihood scans of the signal strength modifier for the combined fit across ggH and VBF categories are performed, in which the other parameters of the signal and background models are allowed to vary. Systematic uncertainties are included in the form of nuisance parameters. Figure A.24 (left) shows the expected likelihood scan obtained by fitting an Asimov dataset, defined by fixing the nuisances parameters to their maximum likelihood estimate obtained from a fit to the data in which $\mu = 0$ is assumed. The likelihood scan crossing at $2 \cdot \Delta \text{Log} \mathcal{L} = 3.84$ corresponds to the expected asymptotic upper limit. In contrast, Fig. A.24 (right) reports the expected likelihood scan obtained by fitting an alternative Asimov dataset in which nuisance parameters are still estimated from a background-only fit to the data but a signal is also injected, with a rate corresponding to the one predicted by the SM. The crossing point of this likelihood scan at $\mu = 0$ gives the expected significance of the signal-plus-background hypothesis w.r.t. the background-only one. In the Appendix of Ref. [17], the same expected likelihood scans, performed either in presence or in absence of a SM Higgs boson signal in the toy, are presented for each of the nine event categories defined along the ggH and VBF BDT outputs.

Table A.17 lists both the expected upper limits on the signal strength modifier, computed at the 95% CL with the Asymptotic CL_s method [79–81], and the significance of the incompatibility between the expected signal-plus-background hypothesis and the background-only hypothesis for each VBF BDT category, as well as for their combination. The *core-PDF* allows to control the bias on the signal strength within 20% in each category at the price of loosing about 4% sensitivity in terms of expected significance and upper limit. The expected upper limit in the absence of a SM decay in this channel is $\mu < 0.93 \times \text{SM}$, which corresponds to an expected significance of the Higgs boson decaying into two muons of 2.18 standard deviations. More-

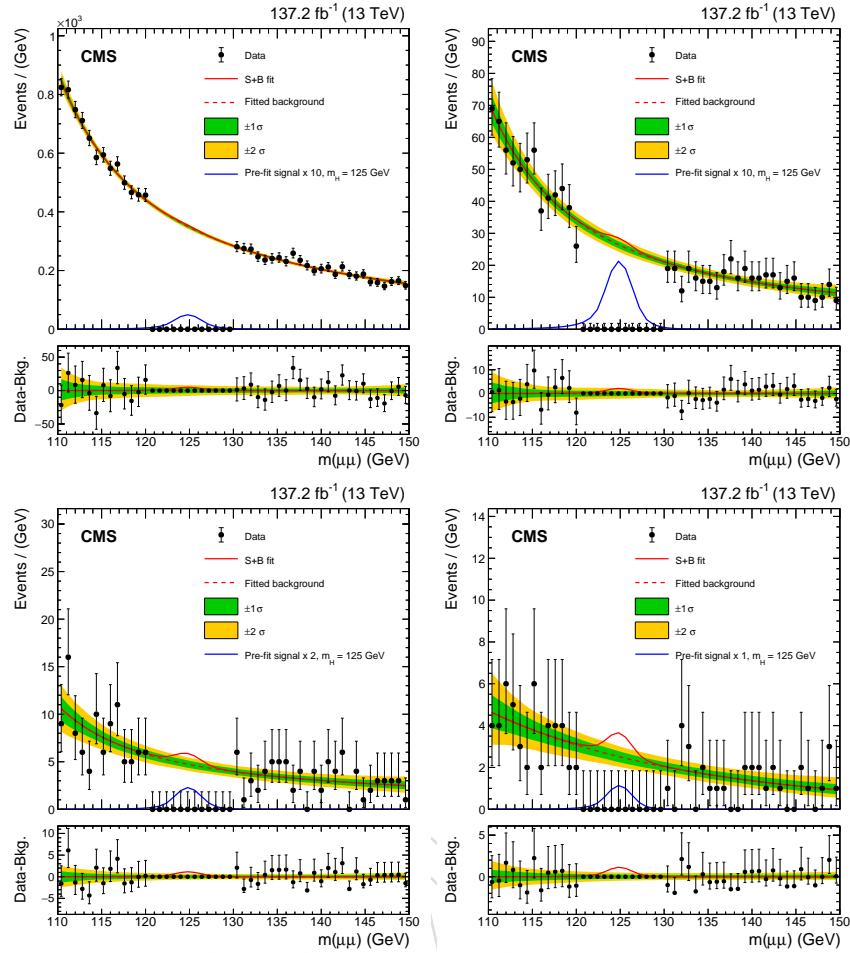


Figure A.22: Comparison between observed data and the total background extracted from a signal-plus-background fit performed across all event categories. Top row: VBF-cat0 (left), and VBF-cat1 (right). Bottom row: VBF-cat2 (left), and VBF-cat3 (right). The one (green) and two (yellow) standard deviation bands include the uncertainties in the background component of the fit. The lower panel in each plot shows the residuals after the background subtraction.

over, in the Appendix of Ref. [17] expected asymptotic upper limits on the signal strength are reported independently for each era and in each category. They have been computed with simplified datacards neglecting systematic uncertainties in the signal prediction and modelling the background via a BWZRedux or a BWZGamma function depending on the analysis categories.

Table A.17: Summary of the expected and observed upper limits on μ , computed at the 95% CL for $m_H = 125\text{ GeV}$, as well as the expected and observed signal significance for each BDT category and for their combination.

Category	Expected UL	Observed UL	Expected Signif.	Observed Signif.
VBF-cat0	< 5.359	—	0.380σ	—
VBF-cat1	< 3.422	—	0.606σ	—
VBF-cat2	< 2.789	—	0.739σ	—
VBF-cat3	< 2.195	—	1.120σ	—
Combination VBF	< 1.418	—	1.519σ	—
Combination ggH and VBF	< 0.926	—	2.183σ	—

The 95% CL upper limits on the Higgs boson production cross section times $\mathcal{B}(H \rightarrow \mu\mu)$ are computed in the range $120 < m_H < 130\text{ GeV}$ using an asymptotic approximation [81] of the

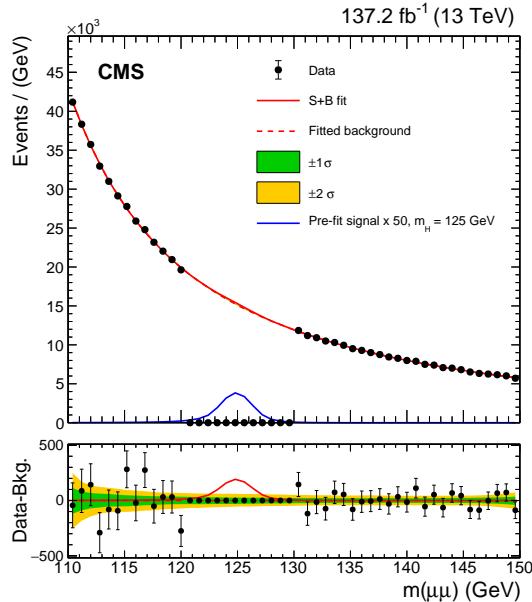


Figure A.23: Comparison between observed data and the total background from a signal-plus-background fit performed across all event categories. Events are weighted according to the expected ratio of signal to signal-plus-background in the category to which they belong. The solid blue line represents the expectation from the SM Higgs boson multiplied by a factor of 10. The one (green) and two (yellow) standard deviation bands include the uncertainties in the background component of the fit. The lower panel in each plot shows the residuals after the background subtraction.

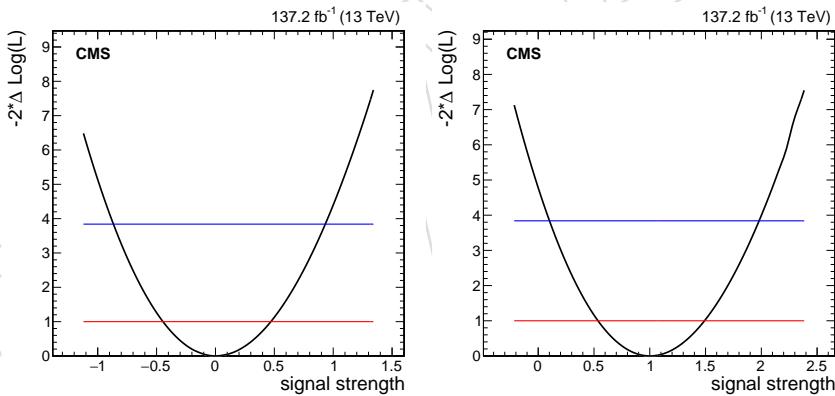


Figure A.24: Expected profile likelihood ratios as a function of the signal strength modifier (μ) for the combined signal-plus-background fit performed across all event categories. On the left, likelihood scan obtained from an Asimov dataset defined by fixing nuisance parameters to the values extracted from a background-only fit to the data. On the right, the likelihood profile vs μ is obtained by fitting an Asimov dataset in which the expected Higgs boson signal, assuming SM production cross sections, has been injected in the toy.

3565 CL_s method [80], and are shown in Figure A.25 (left). The observed and expected upper limits
 3566 for a 125.09 GeV $H \rightarrow \mu\mu$ signal, assuming SM production rates, are found to be X.XX and 0.93
 3567 times the SM expectation, respectively. Figure A.25 (right) shows the observed and expected
 3568 local p-values computed in the mass range $120 < m_H < 130$ GeV for a 125.09 GeV Higgs boson
 3569 signal. The expected significance is found to be 2.18 standard deviations, while the observed
 3570 significance in data is X.XX standard deviations.

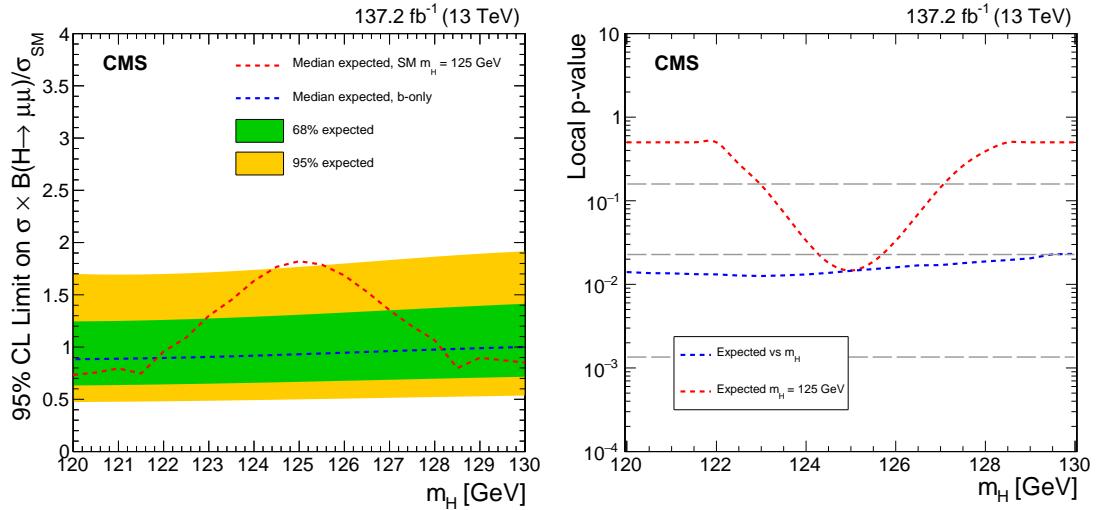


Figure A.25: On the left, the observed 95% CL upper limit on the on the Higgs boson production cross section times $\mathcal{B}(H \rightarrow \mu\mu)$, relative to the SM expectation, in the region around the Higgs boson mass along with the expected exclusion limit obtained in the background-only hypothesis (dashed blue line) and in the signal-plus-background one (dashed red line) for the SM Higgs boson with $m_H = 125.09$ GeV. On the right, the observed and expected local p-values as a function of m_H in the range $120 < m_H < 130$ GeV.

3571 A.9 Effect of systematic uncertainties on the background model

3572 As presented in Section A.7.2, no significant bias (< 20%) on the signal strength from the chosen
 3573 parametric background model in each VBF category is observed, therefore no extra nuisance
 3574 parameters are introduced in the final fit to account the residual per-category bias. However, in
 3575 order to verify that this missing systematic uncertainty has a small impact (< 2%) on the overall
 3576 significance, spurious signals are injected in each category proportionally to the maximum
 3577 observed bias. Table A.18 reports, for each VBF category, the chosen function used in the signal
 3578 extraction, the observed post-fit uncertainty on the signal strength ($\Delta\mu$), and the maximum bias
 3579 w.r.t. the chosen functional form.

Table A.18: In each VBF category, the chosen function used in the signal extraction, the observed post-fit uncertainty on the signal strength ($\Delta\mu$), and the maximum bias for the selected function form as measured in the studies presented in Section A.7.2.

Category	Functional form	Post-fit $\Delta\mu$	Maximum bias (%)
VBF-cat0	BWZRedux	2.56	15%
VBF-cat1	BWZRedux	1.71	9.7%
VBF-cat2	BWZGamma	1.32	9.9%
VBF-cat3	BWZGamma	1.24	7.9%

The spurious signal considered in each category to account for the measured bias on the signal strength is added as a new background process and its $m_{\mu\mu}$ shape is chosen to be identical to the one expected for the VBF process. The total pre-fit normalization for this spurious process is set to be zero in each category. The normalization is allowed to vary in the fit to the data within a Gaussian prior with width (ΔB) defined, for each VBF category, as:

$$\Delta B = \hat{b} \times \Delta\mu \times N_{\text{sig}} \quad (\text{A.3})$$

3580 where \hat{b} is the measured bias on μ , $\Delta\mu$ is the post-fit uncertainty on the signal strength, and N_{sig}
 3581 is the expected number of signal events in the chosen category. Table A.19 reports the difference

3582 in the expected significance for Higgs boson decay to muon pairs with $m_H = 125\text{ GeV}$ between
 3583 the “baseline” result already presented in Table A.17 and the one obtained considering the
 3584 uncertainty in the background model (implemented via spurious signals). The difference in
 3585 the combined expected significance is found to be 0.6% and therefore neglected in the final
 3586 result.

Table A.19: Difference in the expected significance to $H \rightarrow \mu\mu$ decays for $m_H = 125\text{ GeV}$ in each category between the “baseline” result, which ignores systematic uncertainties in the background model, and the one obtained adding spurious signals with rate proportional to the measured bias on μ from the studies presented in Section A.7.2 and 6.4.2.

Category	Signif. w/o bias	Signif. with bias	Difference (%)
VBF-cat0	0.380σ	0.378σ	0.53%
VBF-cat1	0.606σ	0.604σ	0.33%
VBF-cat2	0.739σ	0.736σ	0.41%
VBF-cat3	1.120σ	1.110σ	0.09%
Combination VBF	1.519σ	1.509σ	0.66%
Combination ggH and VBF	2.183σ	2.170σ	0.59%

3587 B Compatibility between VBF template and data-driven analyses

3588 Two independent strategies have been developed to extract the VBF $H \rightarrow \mu\mu$ signal. Both
 3589 analyses consider dimuon events with $110 < m_{\mu\mu} < 150\text{ GeV}$, no additional leptons, and two
 3590 or more jets in the final state, not b-tagged, with leading jet p_T greater than 30 GeV , second
 3591 jet p_T larger than 25 GeV , $m_{jj} > 400\text{ GeV}$, and $|\Delta\eta_{jj}| > 2.5$. The most performing strategy,
 3592 described in Section 7, relies on a template fit to the output of a DNN distribution designed
 3593 to separate at best signal from background events. The alternative analysis, instead, extracts
 3594 the signal by searching for a narrow peak over a smooth background along $m_{\mu\mu}$ distributions
 3595 defined by categorizing events according to the output of a BDT discriminant as detailed in
 3596 Section A. Apart from the differences in the signal extraction strategy, different types of MVA
 3597 are employed in the these two searches (DNN including $m_{\mu\mu}$ in one case, massless BDT in
 3598 the other), as well as different MC samples for the DY background are used in their training
 3599 (LO in the DNN, NLO for in the massless BDT) and slightly different input variables. Besides
 3600 these differences, results from two analyses are expected to be correlated in the sense that, in
 3601 presence of a significant signal contribution, they should be able to detect it giving comparable
 3602 results. However, the current expected sensitivity to SM $H \rightarrow \mu\mu$ decays of the VBF channel
 3603 is smaller than 2σ , with a post-fit uncertainty on the signal strength mainly limited by the
 3604 statistics of the current explored data set. Therefore, in order to quantify the level of correlation
 3605 and compatibility between the expected results from the two strategies, the following test based
 3606 on “full-SIM” toys has been performed:

- 3607 • The DNN template fit is performed simultaneously across the three data-taking pe-
 3608 riods (2016, 2017, and 2018) and the two mass regions (VBF-SR and VBF-SB). This
 3609 corresponds to six independent DNN distributions. The predicted signal and back-
 3610 ground yields in each bin of these DNN output distributions are used as a starting
 3611 point for the toy generation. The VBF and ggH signal rates are taken from the SM
 3612 expectation, while the predicted rate for each background process in each bin is ob-
 3613 tained from a “CR-only” fit performed simultaneously across data-taking periods
 3614 excluding data in the VBF-SR.
- 3615 • For each DNN bin and each process, a random Poisson number (N) is generated
 3616 with mean value given by the expected rate previously described.

- For each DNN bin and each process, N simulated “full-SIM” events corresponding to the same process and the same DNN bin are randomly selected.
- The selected events for signal and background processes are used to fill six DNN templates (one for each mass region and data-taking period) and four $m_{\mu\mu}$ histograms according to categories defined along the massless BDT output described in Section A.3.
- 1500 signal-plus-background toys are generated.

The DNN output and $m_{\mu\mu}$ histograms obtained with the procedure described above are then considered as pseudodata. These pseudodata are fitted either with the DNN templates and corresponding uncertainties described in Section 7.5 and 10 or with the analytical functional forms reported in Section A.3. The corresponding signal strengths extracted from each toy can be directly compared in order to assess the level of correlation and compatibility between the two searches. Figure B.1 (left) shows the distribution of fitted signal strength with the DNN template fit, while Fig. B.1 (right) shows the same distribution as obtained from the data-driven $m_{\mu\mu}$ analysis. In both cases, the signal strength obtained from the toys follows a Gaussian distribution with σ very close to the expected post-fit uncertainty obtained from fitting an Asimov toy with $\mu = 1$. The mean value is slightly smaller than one in both cases: $\bar{\mu} = 0.98$ for the DNN template and $\bar{\mu} = 0.93$ for the data-driven analysis. The small bias observed in the estimated signal strength originates from two approximations implicitly assumed in the toy generation:

- Pseudodata don’t account for systematic uncertainties in the signal and background predictions, while they are included in the final fit.
- The kinematics of signal and background events considered in the toys is not exactly the same of the one used to create the datacards for the final fit. In fact, simulated events for some processes (ggH, DY, WZ, single top) have both positive and negative weighted events. In the toy generation, MC events are randomly picked up without explicitly conserving the expected fraction between positive and negative weighted events.

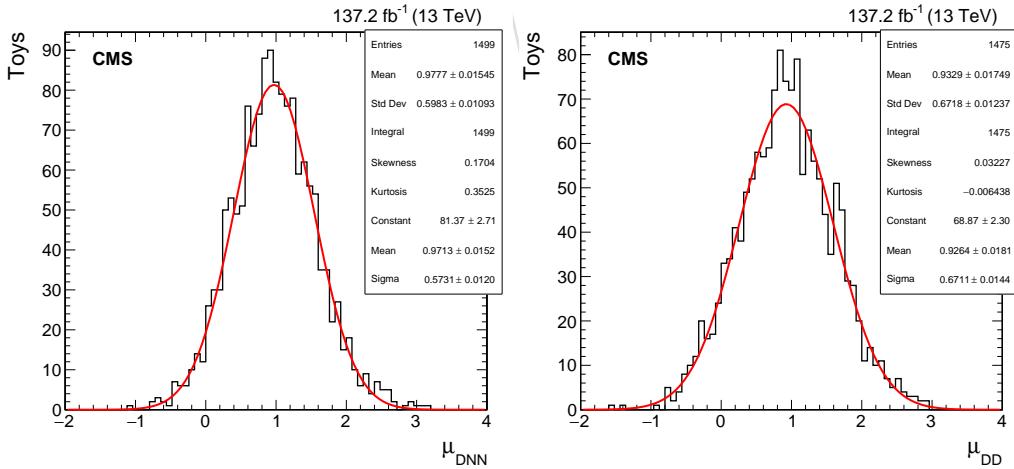


Figure B.1: Distributions of the signal strength extracted by fitting 1500 pseudodata sets generated using full-SIM MC events for signal and background processes for the DNN template analysis (left) and the data-driven one (right). Distributions are fitted with a Gaussian function (red lines).

Figure B.2 (left) shows the difference, on a toy-by-toy basis, between the signal strengths extracted from the data-driven and the DNN template fits. Apart from a small shift in the mean value which can be neglected and is related to the reasons described above, the distribution is observed to be Gaussian-like with a width (σ) of about 0.45. In addition, Fig. B.2 (right) reports

3648 a scatter-plot of the signal strengths extracted from the data-driven (x-axis) and DNN template
 3649 (y-axis) fits. Results from the two searches appear to be quite correlated, with a spread smaller
 3650 than the post-fit uncertainty. In order to better quantify the level of correlation, Fig. B.3 shows
 3651 the pull distribution defined as the difference, computed on a toy-by-toy basis, of the two sig-
 3652 nals strengths divide by the mean post-fit uncertainty, $(\sigma_{DD} + \sigma_{DNN})/2$. The width of the pull
 3653 distribution is 0.64, i.e. the observed spread between the two results is about 60% of the post-fit
 3654 uncertainty. In other words, the post-fit uncertainty on the signal strength has a coverage of
 3655 about 85%. This implies that, with a probability of 85%, the observed signal strengths obtained
 3656 from the DNN template and the data-driven analyses are expected to be compatible within the
 3657 associated post-fit uncertainty.

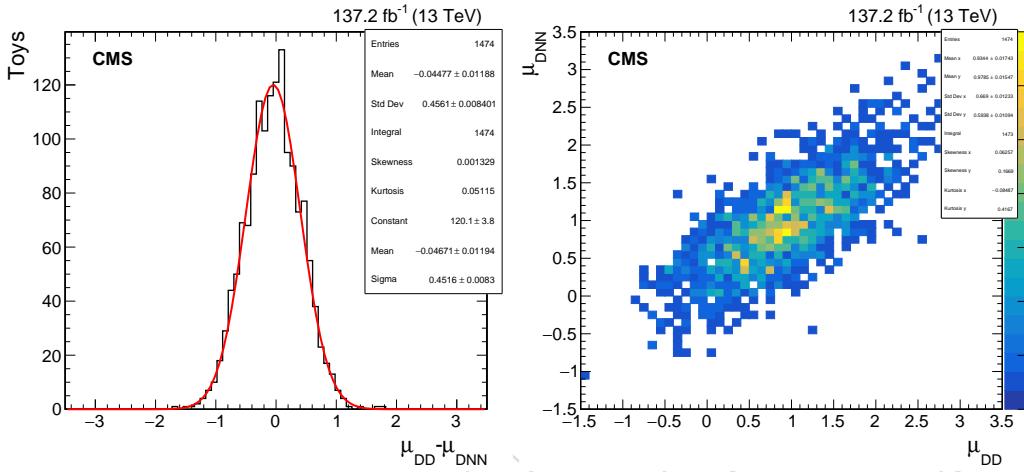


Figure B.2: On the left, distribution of the difference between the signal strengths extracted from the data-driven and the DNN template analyses by fitting 1500 pseudodata sets generated with the procedure described above. On the right, the correlation between the two fitted signal strengths is reported in the form of a 2D histogram.

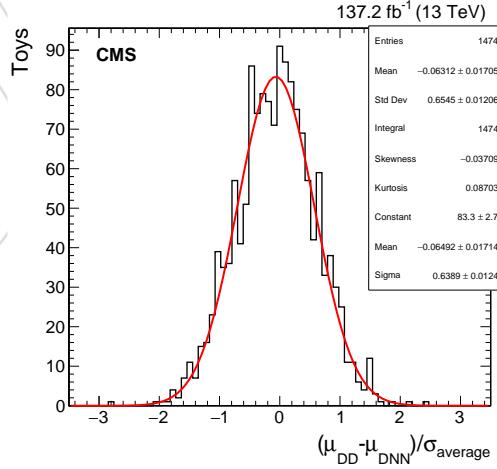


Figure B.3: Pull distribution defined as $(\mu_{DD} - \mu_{DNN})/\sigma$, where $\sigma = (\sigma_{DD} + \sigma_{DNN})/2$, obtained by fitting 1500 pseudodata sets generated by using full-SIM simulated events for signal and background processes. The distribution is fitted with a Gaussian function (red line) with a width of 0.64 ± 0.01 .

3658 A similar study has been performed in order to assess the level of correlation between the DNN
 3659 template fit described in Section 7 and a similar analysis documented in Ref. [18], in which the
 3660 signal is extracted from DNN discriminator trained by the UCSD group. Results from this
 3661 study are reported in Appendix D of Ref. [18], showing that the variance in the measured

3662 signal rate from the two DNN fits is about 60% of the post-fit uncertainty, which is similar to
 3663 the level of compatibility presented before between the DNN fit and the data-driven analysis.

3664 C Bias studies for $m_H = 120$ and 130 GeV

3665 Figure C.1 (left) shows the measured bias on the signal strength for $m_H = 120$ GeV in each of
 3666 the ggH categories when the background is modelled using the *core-PDF* method. The bias is
 3667 observed to be smaller than 20% in each category and for each truth function used to generate
 3668 the toys. In addition, Fig. C.1 (right) reports the bias measured in toys where the signal has been
 3669 injected in all the ggH categories. Also in this case, the bias measured as a function of different
 3670 truth functions can be controlled within 20%. Bias studies have been performed using the same
 3671 strategy/method detailed in Section 6.4.2, where the only difference is the tested Higgs boson
 3672 mass hypothesis. Even though the bias is still within the upper threshold of 20%, the maximum
 3673 bias in each category is larger for $m_H = 120$ GeV than $m_H = 125$ GeV. This is expected because
 3674 of two reasons:

- 3675 • At $m_H = 120$ GeV the resonance is not placed anymore in the centred of the fit range
 3676 (110–150 GeV).
- 3677 • The background shape around 120 GeV is steeper and the corresponding background
 3678 prediction from the tested functional forms appear to have a larger spread in that re-
 3679 gion.

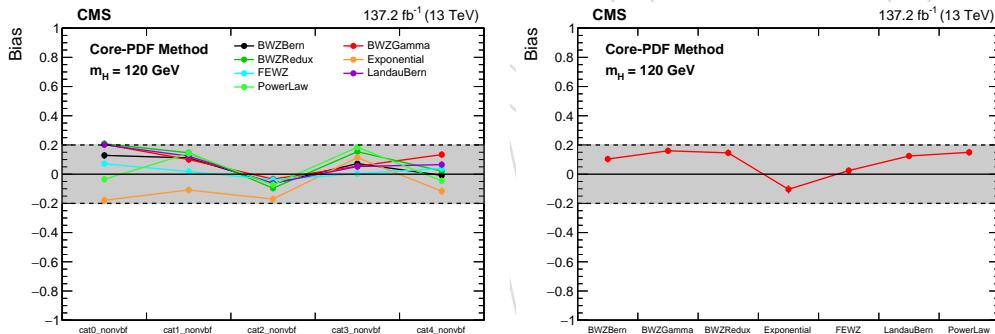


Figure C.1: Mean bias on the signal strength obtained from the *core-PDF* method used to model the background in ggH categories. On the left, the measured bias in each of the ggH categories when a nominal signal at $m_H = 120$ GeV is injected. On the right, the bias measured when the signal is injected in all the ggH categories simultaneously.

3680 The coverage within the 68% CL ($\approx 1\sigma$) interval is defined as the ratio between the observed
 3681 number of toys for which $-2 \times \Delta \text{Log} \mathcal{L}(\hat{\mu}_{inj}) < 1$ and the total number of toys generated,
 3682 where a signal with $\hat{\mu}_{inj} = 1$ and $m_H = 120$ GeV is injected. Similarly, the coverage for 16%
 3683 CL ($\approx 0.2\sigma$), 38% CL ($\approx 0.5\sigma$), 86% CL ($\approx 1.5\sigma$), and 95% CL ($\approx 2\sigma$) intervals have also been
 3684 tested. Figure C.2 shows that the method designed for the background estimation has good
 3685 coverage properties for the tested intervals. The *core-PDF* fit shows, as expected from a dis-
 3686 crete likelihood profile, a small over-coverage (by few %) in categories and against some truth
 3687 models.

3688 Figure C.3 shows the measured bias on the signal strength for $m_H = 120$ GeV in the VBF
 3689 categories obtained by fitting the toys generated via several truth function with BWZRedux,
 3690 BWZGamma, and FEWZ \times Bernstein. The number of degrees of freedom considered for each
 3691 functional form and in each category are listed in Table A.14, optimized for $m_H = 125$ GeV.
 3692 The BWZRedux, BWZGamma, and FEWZ \times Bernstein exhibit a bias within 20% against all

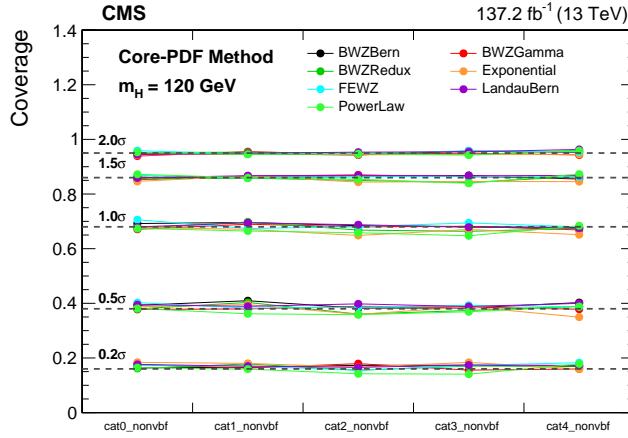


Figure C.2: Coverage observed with the *core-PDF* method for 16%, 38%, 68%, 86%, and 95% CL intervals in toys where a signal at $m_{\{H} = 120 \text{ GeV}$ is injected with $\hat{\mu}_{inj} = 1$.

possible truth models in VBF cat-1, VBF-cat2, and VBF-cat3. Differently from what was already observed for $m_H = 125 \text{ GeV}$ in Section 6.4.1, the maximum bias measured in VBF cat-0 (against sum of exponential) is around 30% for BWZRedux, 35% for BWZGamma, and 25% for FEWZ \times Bernstein.

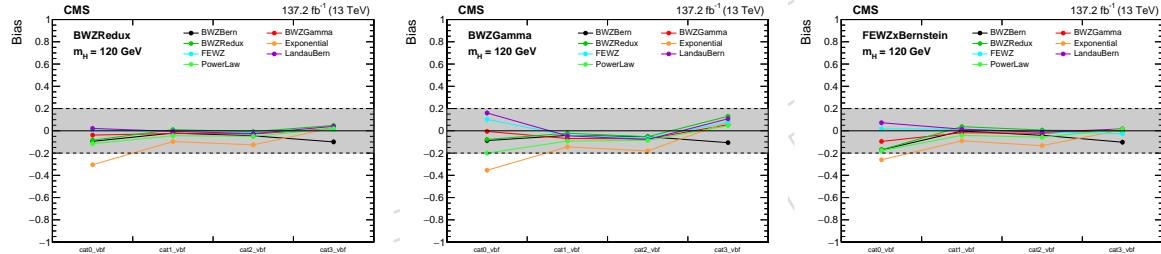


Figure C.3: Mean value of the bias distribution in VBF BDT categories obtained by fitting toys, generated according to different truth background models and injecting a nominal signal at $m_H = 120 \text{ GeV}$, with the following functional forms: BWZRedux (left), BWZGamma (middle), and FEWZ \times Bernstein (right).

Figure C.4 shows the estimated bias on the signal strength for $m_H = 120 \text{ GeV}$ obtained when the VBF categories are simultaneously fitted using either a BWZRedux (left), a BWZGamma (middle), or a FEWZ \times Bernstein (right) function to model the background. When all VBF categories are simultaneously fitted no significant bias is observed for any of the truth model despite the tension observed in VBF cat-0 as reported in the studies presented in Fig. C.3.

Figure C.5 (left) shows the measured bias on the signal strength for $m_H = 130 \text{ GeV}$ in each of the ggH categories when the background is modelled using the *core-PDF* method. The bias is observed to be smaller than 20% in each category and for each truth function used to generate the toys. In addition, Fig. C.5 (right) reports the bias measured in toys where the signal has been injected in all the ggH categories. Also in this case, the bias measured as a function of different truth functions can be controlled within 20%. Bias studies have been performed using the same strategy/method detailed in Section 6.4.2, where the only difference is the tested Higgs boson mass hypothesis.

The coverage within the 68% CL ($\approx 1\sigma$) interval is defined as the ratio between the observed number of toys for which $-2 \times \Delta \log \mathcal{L}(\hat{\mu}_{inj}) < 1$ and the total number of toys generated, where a signal with $\hat{\mu}_{inj} = 1$ and $m_{\{H} = 130 \text{ GeV}$ is injected. Similarly, the coverage for 16%

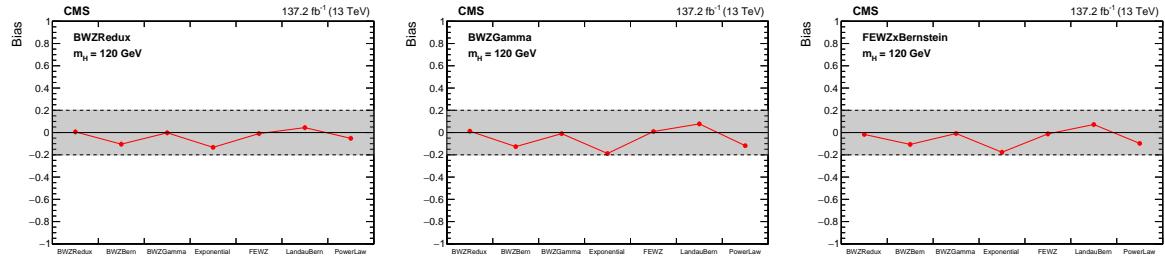


Figure C.4: Mean value of the bias distribution on the signal strength obtained by fitting toys, generated across all the VBF categories according to alternative background truth models and injecting a nominal signal at $m_H = 120$ GeV, with the following functional forms: BWZRedux (left), BWZGamma (middle), and FEWZ \times Bernstein (right).

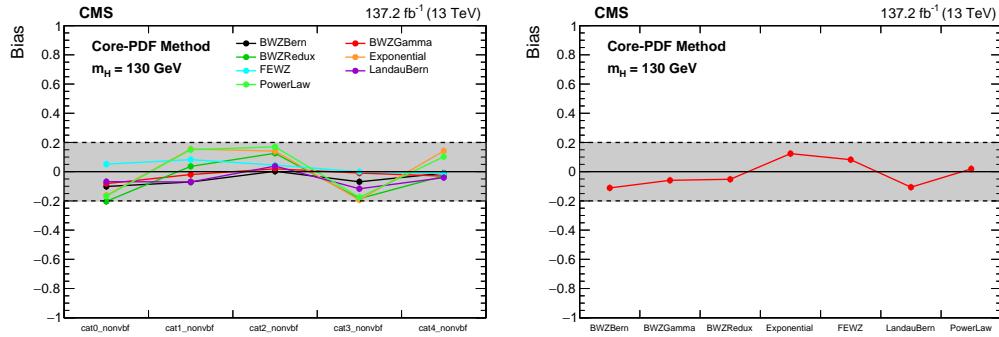


Figure C.5: Mean bias on the signal strength obtained from the *core-PDF* method used to model the background in ggH categories. On the left, the measured bias in each of the ggH categories when a nominal signal at $m_H = 130$ GeV is injected. On the right, the bias measured when the signal is injected in all the ggH categories simultaneously.

CL ($\approx 0.2\sigma$), 38% CL ($\approx 0.5\sigma$), 86% CL ($\approx 1.5\sigma$), and 95% CL ($\approx 2\sigma$) intervals have also been tested. Figure C.6 shows that the method designed for the background estimation has good coverage properties for the tested intervals. The *core-PDF* fit shows, as expected from a discrete likelihood profile, a small over-coverage (by few %) in categories and against some truth models.

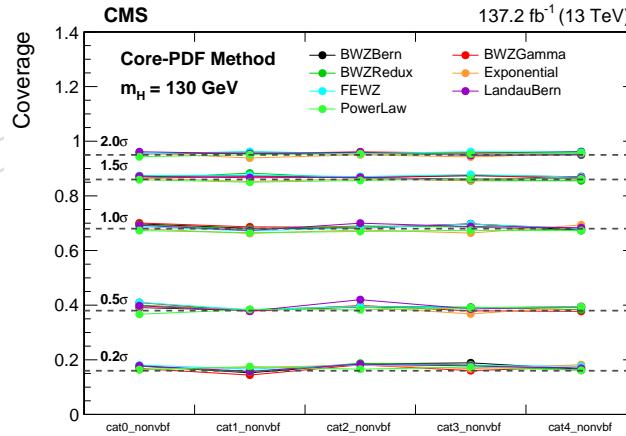


Figure C.6: Coverage observed with the *core-PDF* method for 16%, 38%, 68%, 86%, and 95% CL intervals in toys where a signal at $m_H = 130$ GeV is injected with $\mu_{inj} = 1$.

Figure C.7 shows the measured bias on the signal strength for $m_H = 130$ GeV in the VBF

3719 categories obtained by fitting the toys generated via several truth function with BWZRedux,
 3720 BWZGamma, FEWZ \times Bernstein. The number of degrees of freedom considered for each func-
 3721 tional form and in each category are listed in Table A.14, optimized for $m_H = 125\text{ GeV}$. The
 3722 BWZRedux, BWZGamma, and FEWZ \times Bernstein exhibit a bias in single categories within
 3723 20% against all possible truth models, as already observed for $m_H = 125\text{ GeV}$ in Section 6.4.1.

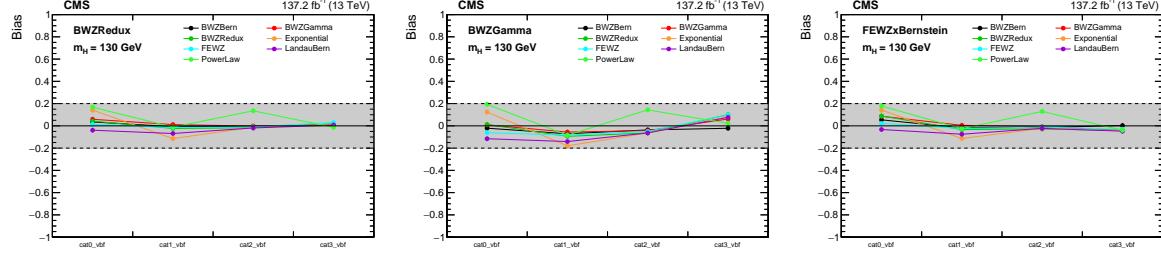


Figure C.7: Mean value of the bias distribution in VBF BDT categories obtained by fitting toys, generated according to different truth background models and injecting a nominal signal at $m_H = 130\text{ GeV}$, with the following functional forms: BWZRedux (left), BWZGamma (middle), and FEWZ \times Bernstein (right).

3724 Figure C.8 shows the estimated bias on the signal strength for $m_H = 130\text{ GeV}$ obtained when
 3725 the VBF categories are simultaneously fitted using either a BWZRedux (left), a BWZGamma
 3726 (middle), or a FEWZ \times Bernstein (right) function to model the background. As expected from
 3727 the studies presented in Fig. C.7, when all VBF categories are simultaneously fitted no signifi-
 3728 cant bias is observed for any of the truth model.

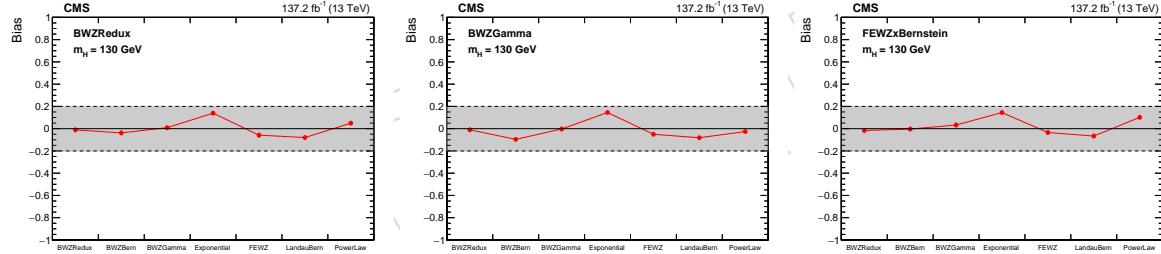


Figure C.8: Mean value of the bias distribution on the signal strength obtained by fitting toys, generated across all the VBF categories according to alternative background truth models and injecting a nominal signal at $m_H = 130\text{ GeV}$, with the following functional forms: BWZRedux (left), BWZGamma (middle), and FEWZ \times Bernstein (right).

3729 D GeoFit Corrections

3730 In CMS, muon p_T values are primarily computed using the measured radius of curvature R of
 3731 the reconstructed muon track in the inner tracker. This reconstruction has inherent uncertainties,
 3732 which affect both the track trajectory and p_T value. Prompt muons originate directly from
 3733 the collision vertex, so the measured point of closest approach between the muon track and
 3734 the collision vertex in the $x - y$ plane (the "d0" value) should be exactly 0 for muons from W,
 3735 Z, and Higgs boson decays. If the muon track's radius is mismeasured, however, the d0 value
 3736 will be non-0, measured "inside" the true collision vertex if the radius (and thus p_T) is under-
 3737 estimated, and "outside" if the radius and p_T are overestimated. Thus the measured muon d0
 3738 value in Z and Higgs boson decays should be correlated with the size and direction of the p_T
 3739 mismeasurement. Fig. D.1 (left) is derived from a Z+jets MC sample, and shows the pull in
 3740 fitted dimuon mass peak values versus μ^+ and μ^- d0. There is a clear trend in the mass peak
 3741 values, increasing along the $d0_{\mu^+} - d0_{\mu^-}$ direction. The right-hand plot shows that this trend
 3742 is nearly identical in data and MC.

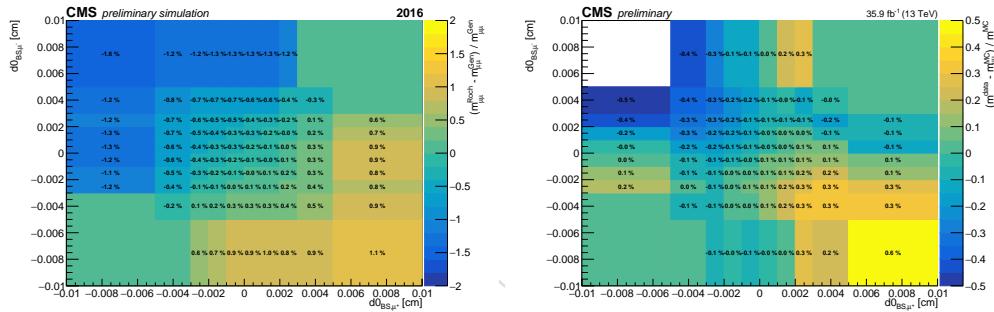


Figure D.1: 2D maps showing pulls on dimuon mass around Z boson mass peak, comparing reconstructed *Rochester Corrected* (Roch) and generator level masses (Gen) on the left, and the difference between data and MC mass peaks on the right, binned in $d0_{\mu^+}$ vs. $d0_{\mu^-}$ for Z+jets MC and data samples from 2016.

3743 A detailed geometric inspection shows that prompt muon track d0 values are related to the
 3744 mismeasurement in radius of curvature (ΔR) as follows:

$$|d0| \sim \frac{|\Delta R|}{R^2} \quad (\text{D.1})$$

3745 Since in homogeneous magnetic fields $p_T \sim R$, we can write the above equation as:

$$|d0| \sim \frac{|\Delta p_T|}{p_T^2} \quad (\text{D.2})$$

3746 Therefore the mismeasurement in p_T of the muons due to this effect can be corrected by finding
 3747 the constant of proportionality in Eq. D.2. The correction derived using this procedure is named
 3748 *GeoFit Corrections* due to its geometric motivation.

3749 The constant of proportionality of *GeoFit Corrections* may depend on factors such as the muon η ,
 3750 since the constant of proportionality between d0 and p_T depends on the magnetic field. Since
 3751 muon charge affects the direction the muon bends in magnetic field we expect this effect to
 3752 have reversed sign for μ^- compared to μ^+ . In this study the value of d0 is always multiplied by
 3753 muon charge to take this effect into account. The effects of other variables on *GeoFit Corrections*
 3754 are harder to guess, and in order to understand these effects many dependencies are studied.

3755 First, the definition of d_0 that is used in these corrections needs to be determined. In general,
 3756 d_0 values for muons are defined as the distance of closest approach of the muon track to the
 3757 reconstructed primary vertex ($d_{0_{PV}}$). It is calculated as $d_0 = -x_0 \sin(\phi_0) + y_0 \cos(\phi_0)$ where
 3758 (x_0, y_0) are the coordinates of the point on the track closest to the PV, and ϕ_0 is the azimuthal
 3759 angle of the track at that point. However, since the reconstructed PV is determined by the
 3760 tracks of physics objects in an event, in events with few high p_T tracks the PV position tends
 3761 to be closer to these tracks. This results in calculated $d_{0_{PV}}$ values to be smaller than the true
 3762 values. Another useful d_0 parameter can be calculated by using the *beam spot* (BS) instead of PV,
 3763 and is denoted as $d_{0_{BS}}$. Since the BS is calculated over 23–1400s of collisions [106], it has a more
 3764 averaged out position which is less biased towards high p_T tracks in a single event. Figure D.2
 3765 shows the distribution of $d_{0_{PV}}$ and $d_{0_{BS}}$ values of the leading muon for data and MC samples.
 3766 As can be seen from these figures, $d_{0_{PV}}$ distributions have a sharper peak between $\pm 30\mu\text{m}$
 3767 which is stacked on top of a wider peak, in line with our reasoning.

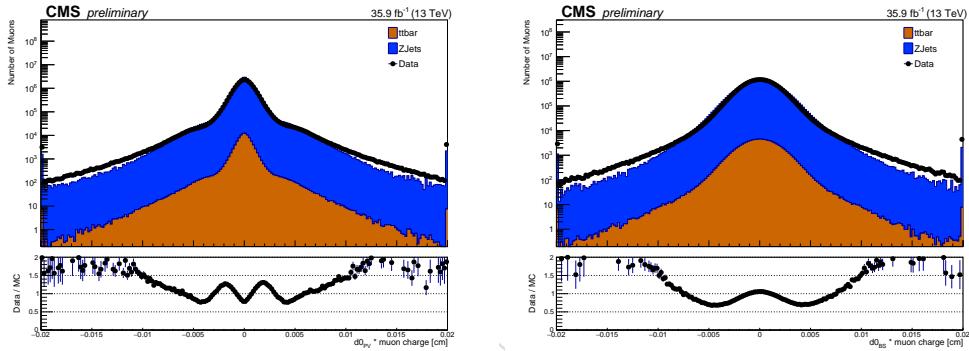


Figure D.2: Plots showing comparisons of $d_{0_{PV}}$ (left) and $d_{0_{BS}}$ (right) of leading muon between data and MC. $d_{0_{PV}}$ plot shows the narrow peak around zero, while $d_{0_{BS}}$ distribution does not have a similar feature.

3768 In order to inspect the effect of using $d_{0_{PV}}$ or $d_{0_{BS}}$ further, Figure D.3 contains 2D plots derived
 3769 from Z+jets MC samples that display the number of muons in bins of $(p_T^{Roch} - p_T^{Gen})/p_T^2$ and
 3770 $d_{0_{PV}}$ or $(p_T^{Roch} - p_T^{Gen})/p_T^2$ and $d_{0_{BS}}$. From these plots we can see that the linear relationship
 3771 between $(p_T^{Roch} - p_T^{Gen})/p_T^2$ and $d_{0_{BS}}$ is more pronounced compared to $d_{0_{PV}}$.

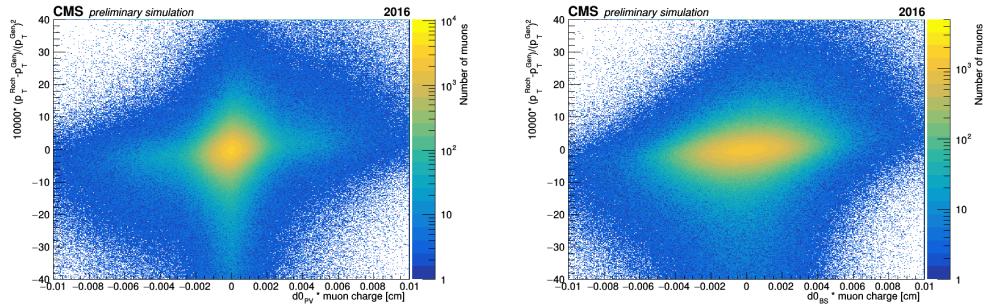


Figure D.3: 2D maps showing number of muons in bins of $(p_T^{Roch} - p_T^{Gen})/p_T^2$ and $d_{0_{PV}}$ (left) and $(p_T^{Roch} - p_T^{Gen})/p_T^2$ and $d_{0_{BS}}$ (right). The linear relationship is more pronounced in the right plot, whereas the left plot have a more concentrated and circular distribution near the center.

3772 In order to extract the constant of proportionality, we take slices of $d_{0_{PV}}$ or $d_{0_{BS}}$ and find the
 3773 peak and width of Gaussian-like $(p_T^{Roch} - p_T^{Gen})/p_T^2$ distributions in these slices. Figure D.4
 3774 shows data points which represents the positions of the peaks for each of the d_0 slices with
 3775 error bars representing the width of these distributions in each slice. As can be seen from these

plots, the distributions with $d0_{BS}$ show a clear linear dependence between $(p_T^{Roch} - p_T^{Gen}) / p_T^2$ and $d0_{BS}$ whereas the distributions with $d0_{PV}$ show a non-linear dependence. Therefore $d0_{BS}$ is a better handle to calculate the effect of non-zero muon d0 on muon p_T , and is chosen to be used throughout these corrections.

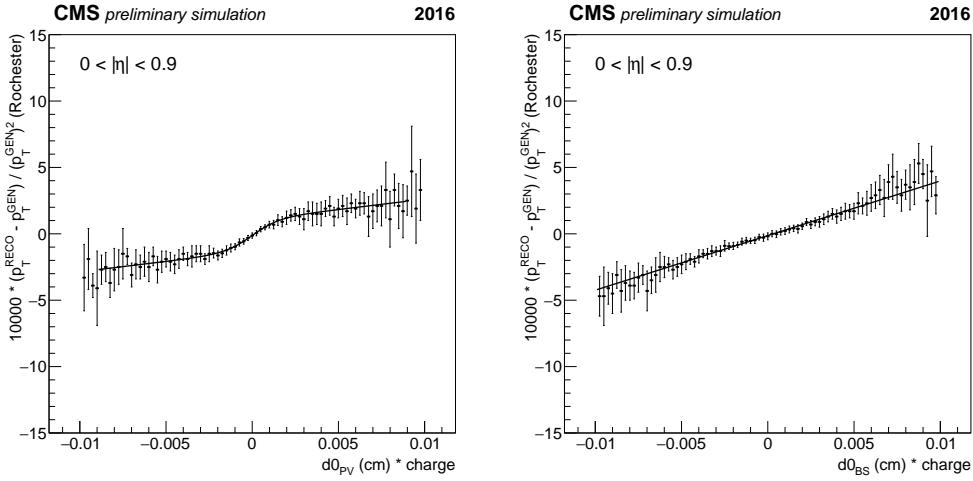


Figure D.4: Plots showing the dependence of $(p_T^{Roch} - p_T^{Gen}) / p_T^2$ on $d0_{PV}$ (left) and $d0_{BS}$ (right). The dependence is clearly non-linear for $d0_{PV}$ while being linear for $d0_{BS}$.

After deciding which d0 definition to use, it is still necessary to understand if these corrections are correlated with other event parameters. Figure D.5 shows the dependence of the dimuon mass peak shift vs. muon η . This effect is more pronounced in higher $|\eta|$ regions compared to lower $|\eta|$ regions and is consistent between data and MC. Figure D.5 suggests that when deriving the constant of proportionality for *GeoFit Corrections* we need to consider three separate η regions: $|\eta| < 0.9$, $0.9 < |\eta| < 1.7$, and $|\eta| > 1.7$.

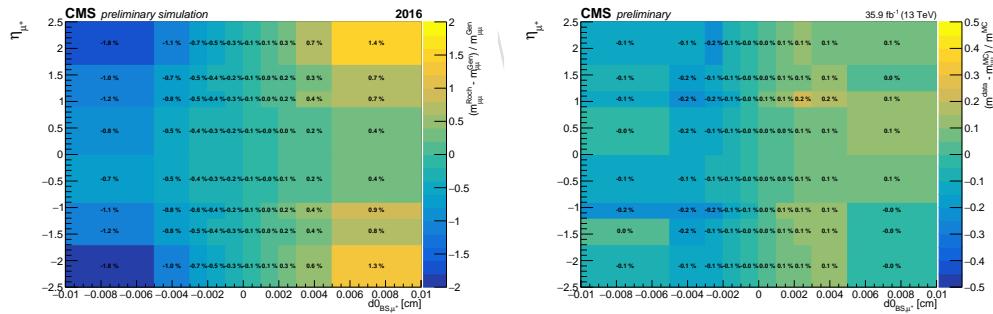


Figure D.5: The dependence of p_T bias on $d0$ and muon η (left), and the difference between data and MC (right). The $d0$ dependence becomes more pronounced as $|\eta|$ increases.

Figure D.6 shows the dependence of the dimuon mass peak shift vs muon ϕ . The $d0$ effect on p_T seems to be independent of muon ϕ and therefore we do not need to consider different regions of ϕ .

The remaining studies are done by using $(p_T^{Roch} - p_T^{Gen}) / p_T^2$ vs. $d0_{BS}$ distributions in the three $|\eta|$ regions defined above.

Figures D.7, D.8 and D.9 show the dependence of *GeoFit Corrections* on pileup in 2016 – 2018. The constant of proportionality derived from events with low number of vertices in the event is consistent with the constant of proportionality derived from events with high number of

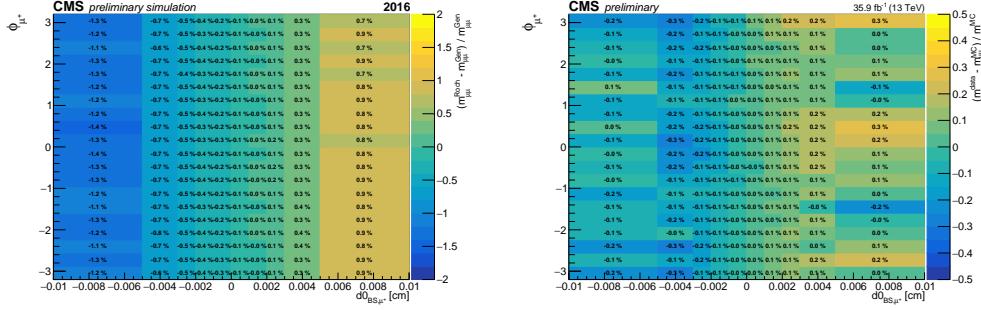


Figure D.6: Plots showing the dependence of d_0 effect on muon ϕ . There is no clear dependence on ϕ .

3794 vertices. Therefore it is understood that *GeoFit Corrections* do not depend on the amount of
 3795 pileup in the event.

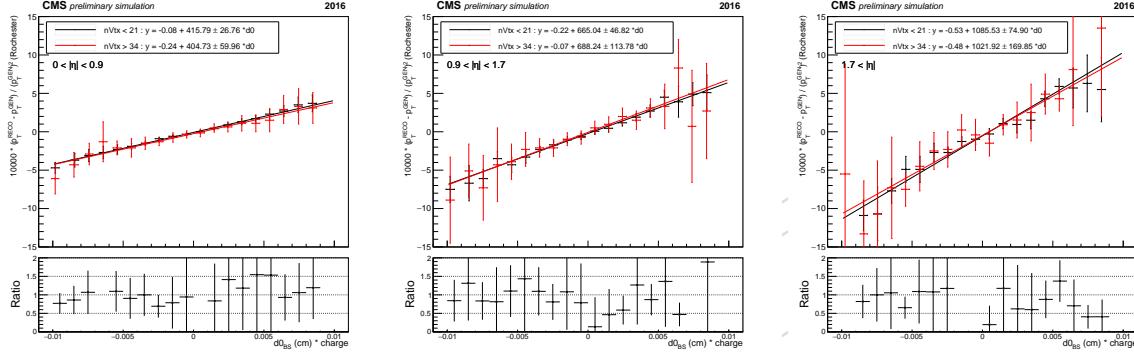


Figure D.7: Plots comparing $(p_T^{Roch} - p_T^{Gen}) / p_T^2$ vs $d_{0\text{BS}}$ distributions for events with less than 21 and more than 34 vertices for each $|\eta|$ region, derived from Z+jets MC in 2016. The ratio plot show the ratio between the black data points ($nVtx < 21$) and the red data points ($nVtx > 34$).

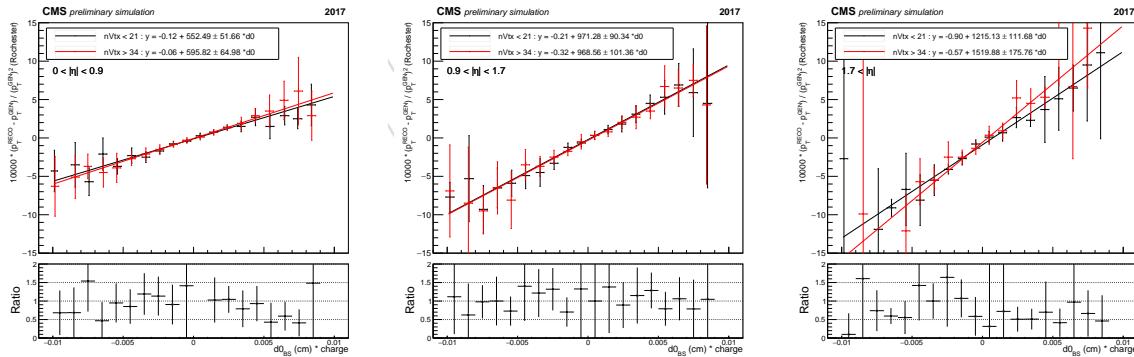


Figure D.8: Plots comparing $(p_T^{Roch} - p_T^{Gen}) / p_T^2$ vs $d_{0\text{BS}}$ distributions for events with number of vertices less than 21 and number of vertices more than 34 for each $|\eta|$ region derived from Z+jets MC in 2017. The ratio plot show the ratio between the black data points ($nVtx < 21$) and the red data points ($nVtx > 34$).

3796 Figures D.10, D.11, and D.12 show the dependence of *GeoFit Corrections* on muon charge in
 3797 2016 – 2018. The constants of proportionality derived by using only μ^- or μ^+ are consistent
 3798 with each other. Therefore *GeoFit Corrections* do not depend on the muon charge.

3799 Figures D.13, D.14, and D.15 show the dependence of *GeoFit Corrections* on number of tracks in
 3800 the vertex in 2016 – 2018. This is achieved by considering ttH MC samples instead of Z+jets MC
 3801 samples due to ttH samples having more high p_T prompt objects in the event. The constants of

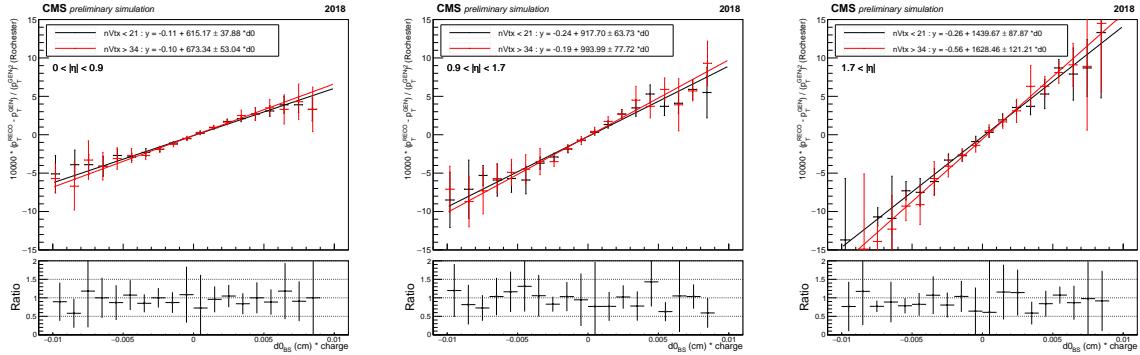


Figure D.9: Plots comparing $(p_T^{Roch} - p_T^{Gen}) / p_T^2$ vs $d0_{BS}$ distributions for events with number of vertices less than 21 and number of vertices more than 34 for each $|\eta|$ region derived from Z+jets MC in 2018. The ratio plot show the ratio between the black data points ($nVtx < 21$) and the red data points ($nVtx > 34$).

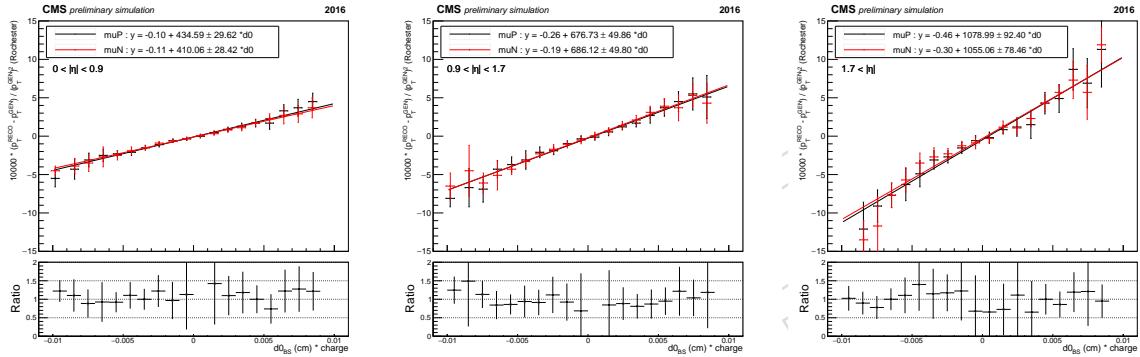


Figure D.10: Plots comparing $(p_T^{Roch} - p_T^{Gen}) / p_T^2$ vs $d0_{BS}$ distributions for μ^+ and μ^- for each $|\eta|$ region derived from Z+jets MC in 2016. The ratio plot show the ratio between the black data points (μ^+) and the red data points (μ^-).

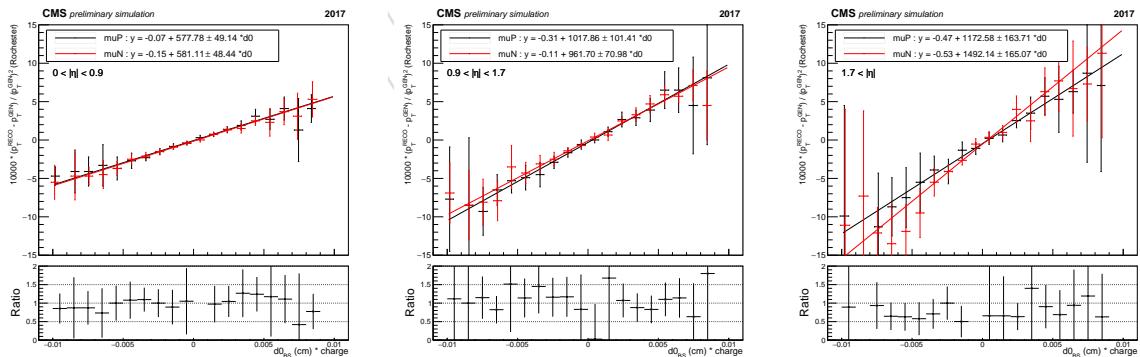


Figure D.11: Plots comparing $(p_T^{Roch} - p_T^{Gen}) / p_T^2$ vs $d0_{BS}$ distributions for μ^+ and μ^- for each $|\eta|$ region derived from Z+jets MC in 2017. The ratio plot show the ratio between the black data points (μ^+) and the red data points (μ^-).

proportionality derived from each sample is consistent with each other, which shows that the *GeoFit Corrections* do not depend strongly on the number of tracks in the vertex.

Finally, Figures D.16 and D.17 show comparison of p_T distributions between data and MC for leading and subleading muons, before and after *GeoFit Corrections*. We can see that *GeoFit Corrections* do not change the agreement between p_T distributions in data and MC.

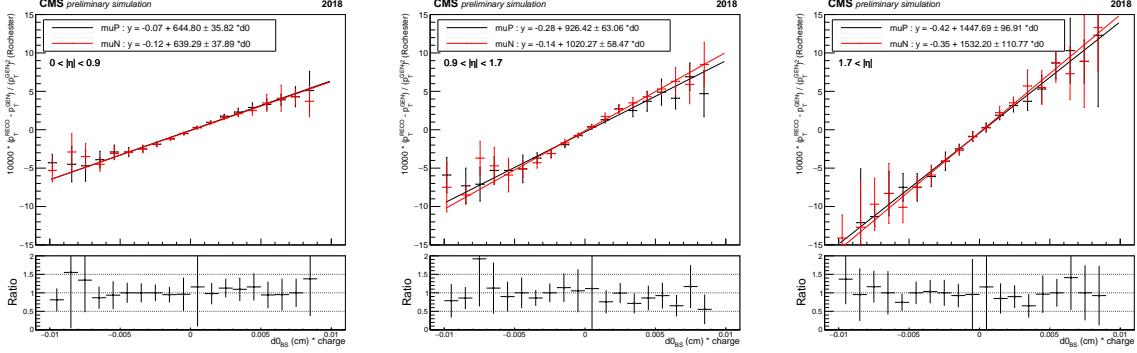


Figure D.12: Plots comparing $(p_T^{Roch} - p_T^{Gen}) / p_T^2$ vs $d0_{BS}$ distributions for μ^+ and μ^- for each $|\eta|$ region derived from Z+jets MC in 2018. The ratio plot show the ratio between the black data points (μ^+) and the red data points (μ^-).

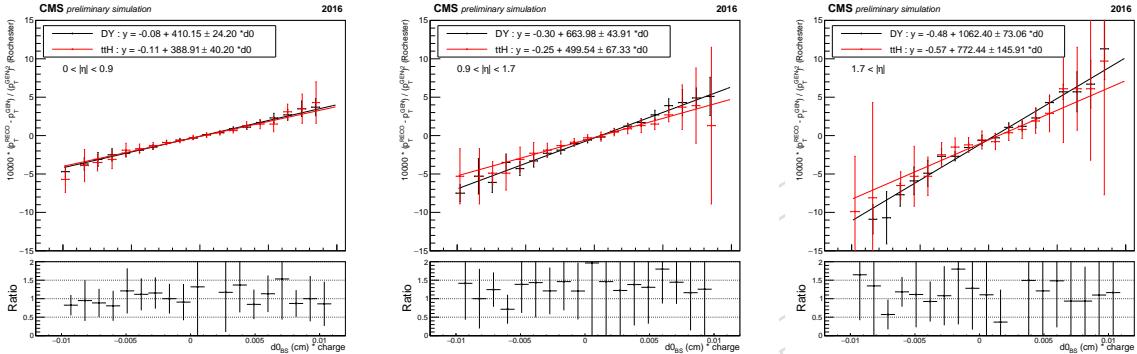


Figure D.13: Plots comparing $(p_T^{Roch} - p_T^{Gen}) / p_T^2$ vs $d0_{BS}$ distributions in each $|\eta|$ region derived from Z+jets MC and ttH MC in 2016. The ratio plot show the ratio between the black data points (Z+jets MC) and the red data points (ttH MC).

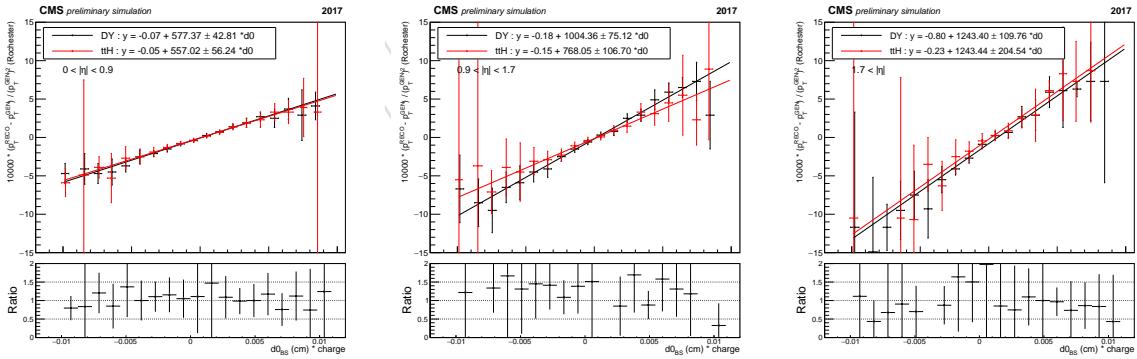


Figure D.14: Plots comparing $(p_T^{Roch} - p_T^{Gen}) / p_T^2$ vs $d0_{BS}$ distributions in each $|\eta|$ region derived from Z+jets MC and ttH MC in 2017. The ratio plot show the ratio between the black data points (Z+jets MC) and the red data points (ttH MC).

After checking all the dependencies we can determine that the *GeoFit Corrections* only depend strongly on muon η and the era of the MC samples. The latter dependency can be explained by different detector conditions and reconstructed algorithms used in different eras. Due to improvements in inner tracker and BS calculation algorithms we expect the performance of *GeoFit Corrections* to improve from 2016 to 2018.

Figures D.18, D.19, D.20, D.21, D.22, D.23, and D.24 show the comparison of dimuon mass

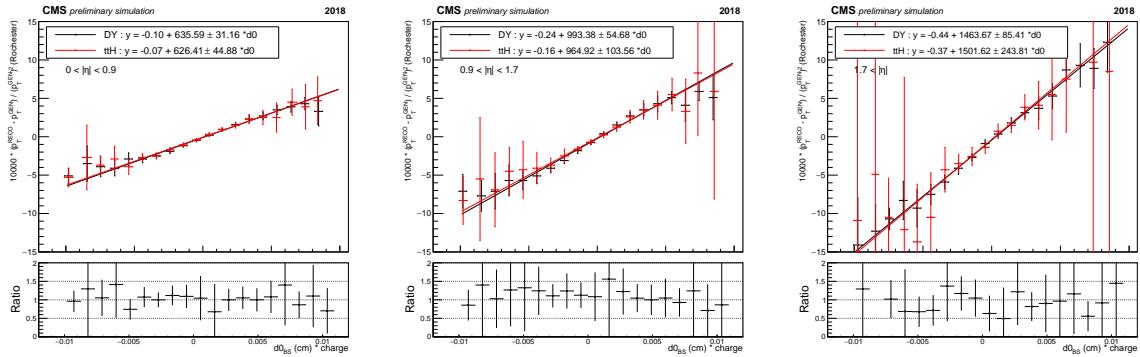


Figure D.15: Plots comparing $(p_T^{\text{Roch}} - p_T^{\text{Gen}}) / p_T^2$ vs $d0_{\text{BS}}$ distributions in each $|\eta|$ region derived from Z+jets MC and ttH MC in 2018. The ratio plot show the ratio between the black data points (Z+jets MC) and the red data points (ttH MC).

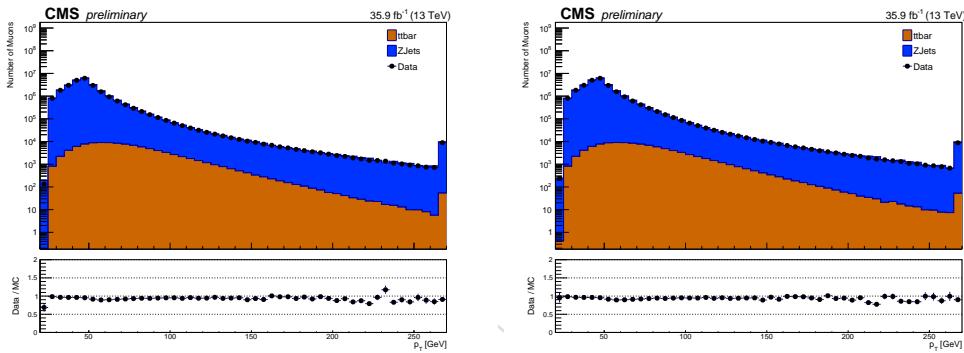


Figure D.16: Plots showing the comparison of leading muon p_T between data and MC before (left) and after (right) *GeoFit Corrections*.

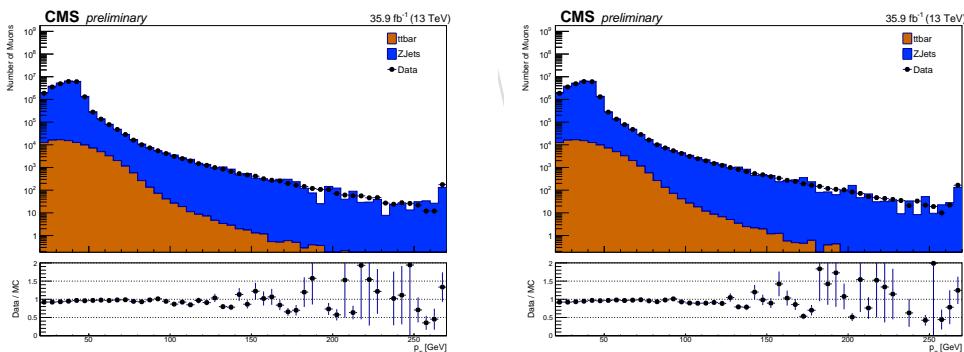


Figure D.17: Plots showing the comparison of subleading muon p_T between data and MC before (left) and after (right) *GeoFit Corrections*.

peaks for Z+jets, ggH, VBF, VH and ttH MC and Z+jets data before and after the *GeoFit Corrections*. The resulting improvements for MC samples are compatible with our expectations with the performance improving from 2016 to 2018. The improvements in Z+jets data are also consistent with Z+jets MC with data performing slightly better overall.

One interesting aspect of *GeoFit Corrections* is that the performance of these corrections improve with increasing average p_T of muons. This is consistent with the geometric nature of these corrections since this effect scales with p_T^2 . This results in a better performance for *GeoFit Corrections* in ggH, VBF, VH, and ttH signal samples compared to Z+jets samples.

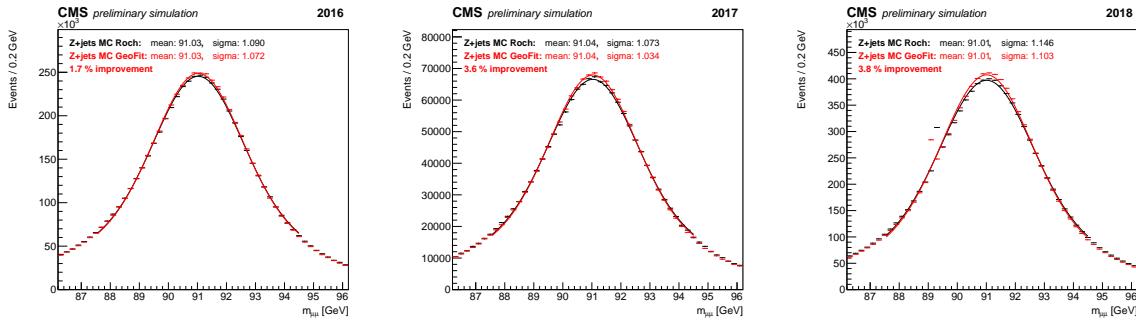


Figure D.18: Dimuon mass peak around 91 GeV for Z+jets MC samples in 2016 (left), 2017 (center) and 2018 (right) fitted by using a Voigtian+exponential function.

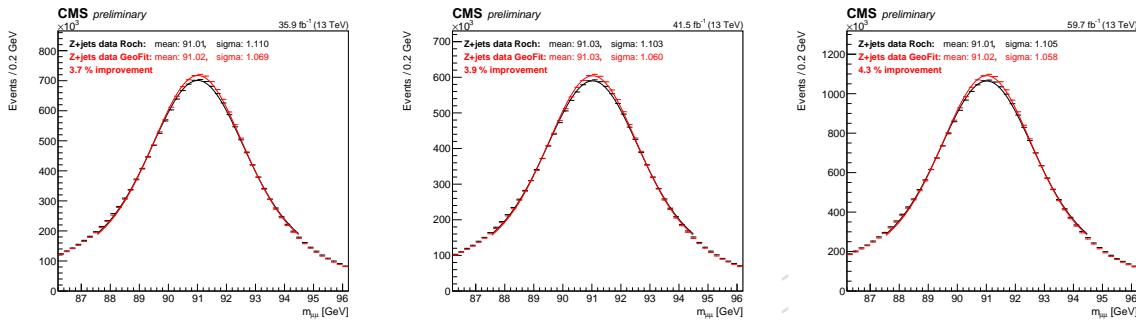


Figure D.19: Dimuon mass peak around 91 GeV for Z+jets data samples in 2016 (left), 2017 (center) and 2018 (right) fitted by using a Voigtian+exponential function.

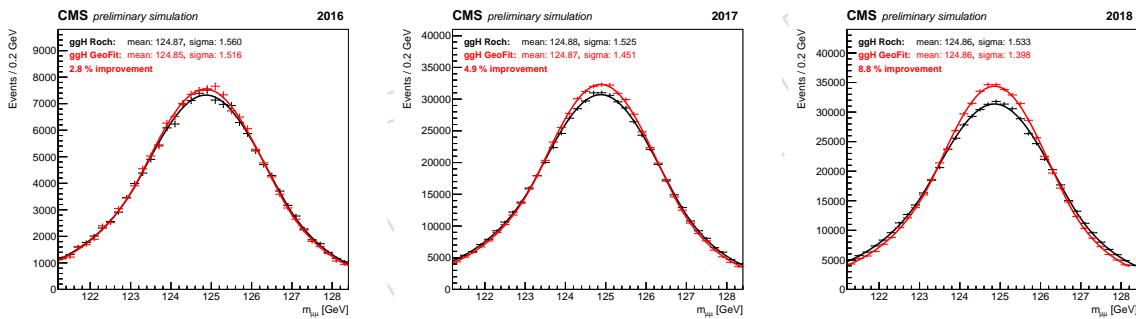


Figure D.20: Dimuon mass peak around 125 GeV for ggH signal MC samples in 2016 (left), 2017 (center) and 2018 (right) fitted by using a double-sided crystal ball function.

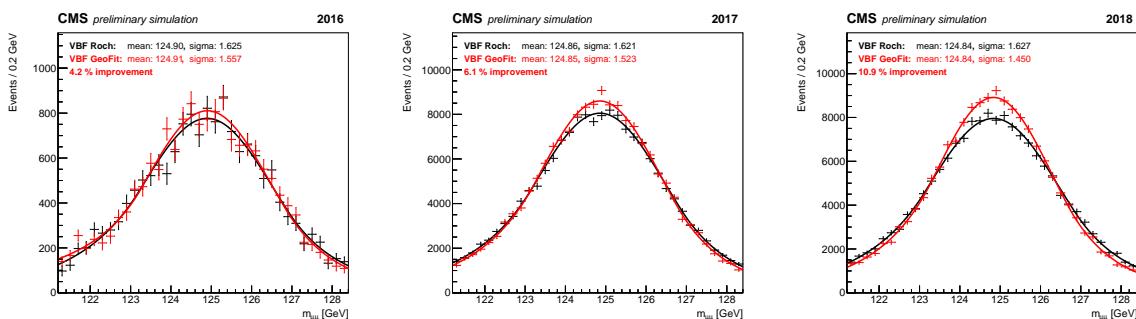


Figure D.21: Dimuon mass peak around 125 GeV for VBF signal MC samples in 2016 (left), 2017 (center) and 2018 (right) fitted by using a double-sided crystal ball function.

3821 The fitting uncertainties for the *GeoFit Correction* constants of proportionality are on the order of
 3822 10%. Since these corrections cannot (by construction) change the mass scale, the only impact on

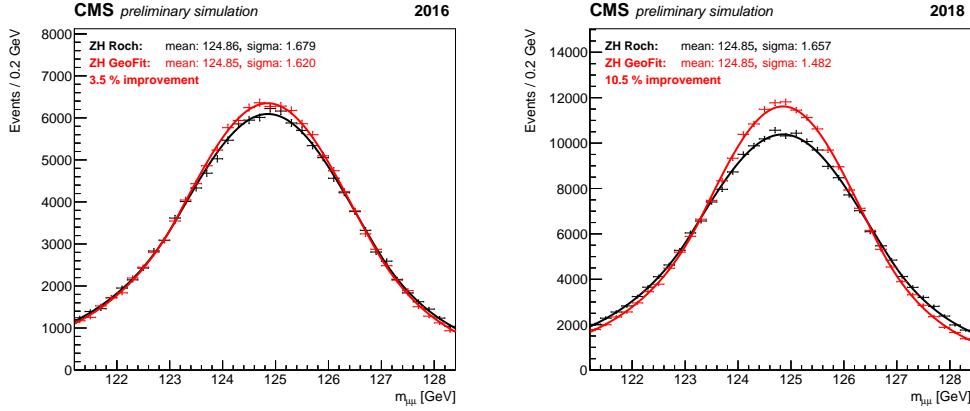


Figure D.22: Dimuon mass peak around 125 GeV for ZH signal MC samples in 2016 (left) and 2018 (right) fitted by using a double-sided crystal ball function.

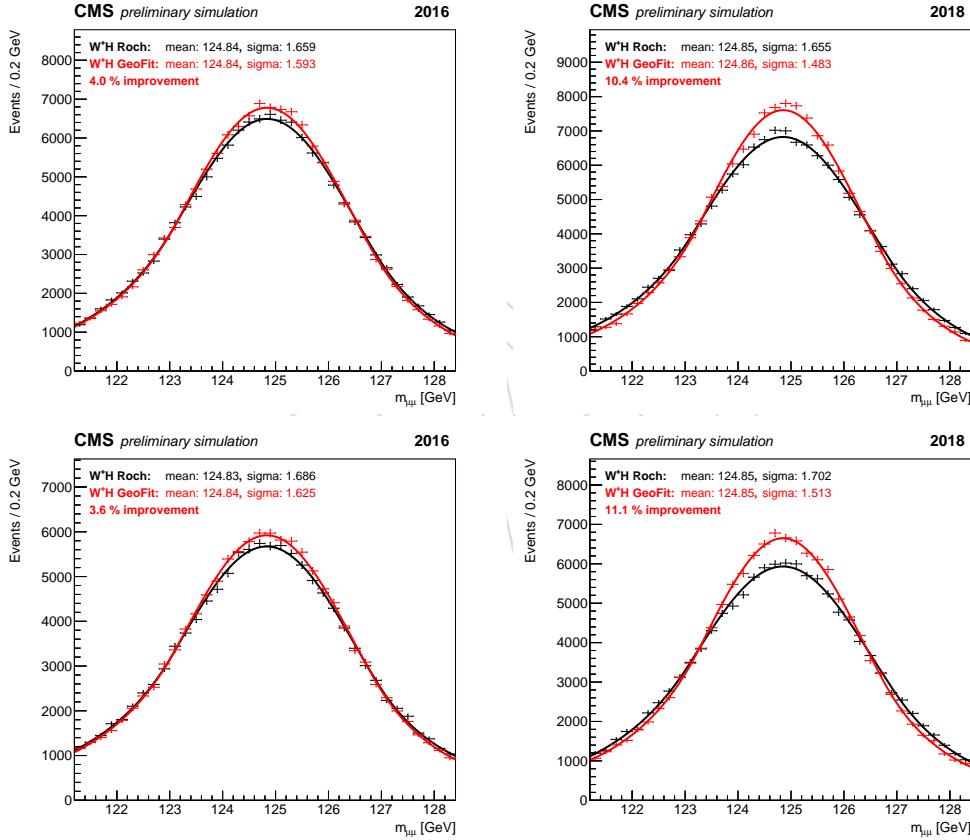


Figure D.23: Dimuon mass peak around 125 GeV for VH signal MC samples for WH negative in 2016 (top left), WH negative in 2018 (top right), WH positive in 2016 (bottom left), WH positive in 2018 (bottom right) fitted by using a double-sided crystal ball function.

the Higgs to dimuon measurement uncertainty is in the agreement between data and MC mass resolution. As shown in Fig. D.18 and D.19, the corrections actually bring the measured Z boson width in data and MC closer together. Thus the 10% uncertainty on measured Higgs mass resolution derived from Z-peak fits (Section 4.6) is sufficient to cover the total mass resolution uncertainty after *GeoFit Corrections*.

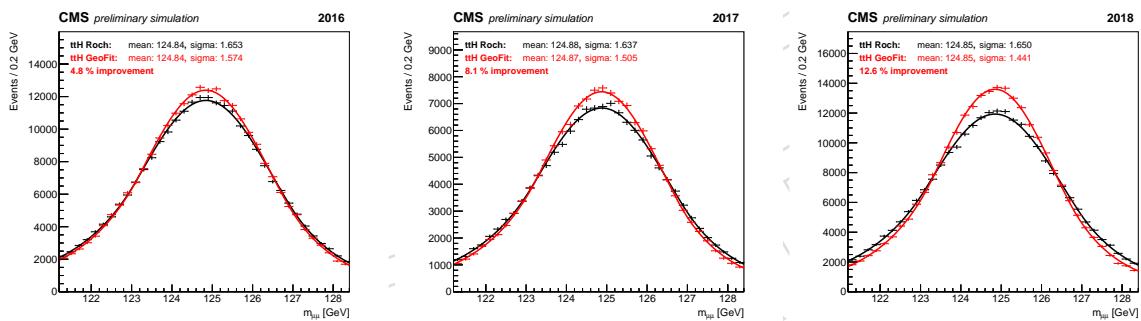


Figure D.24: Dimuon mass peak around 125 GeV for ttH signal MC samples in 2016 (left), 2017 (center) and 2018 (right) fitted by using a double-sided crystal ball function.

3828 E LeptonMVA

3829 Figures E.1-E.15 show the input variables for muons across the three years before any cut on
 3830 the *LeptonMVA* output, figures E.17-E.30 the ones for electrons. The resulting outputs can be
 3831 seen in figure E.16 (muons) and figure E.31 (electrons).

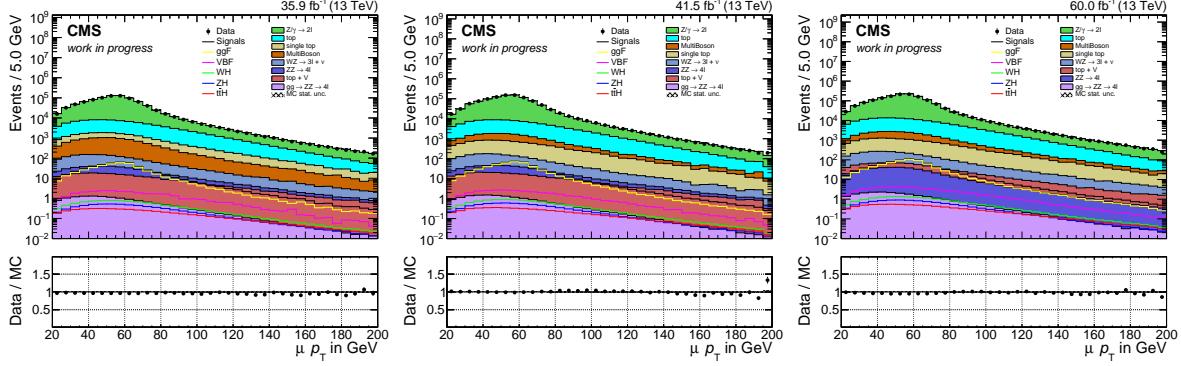


Figure E.1: Muon p_T for 2016 (left), 2017 (center), 2018 (right)

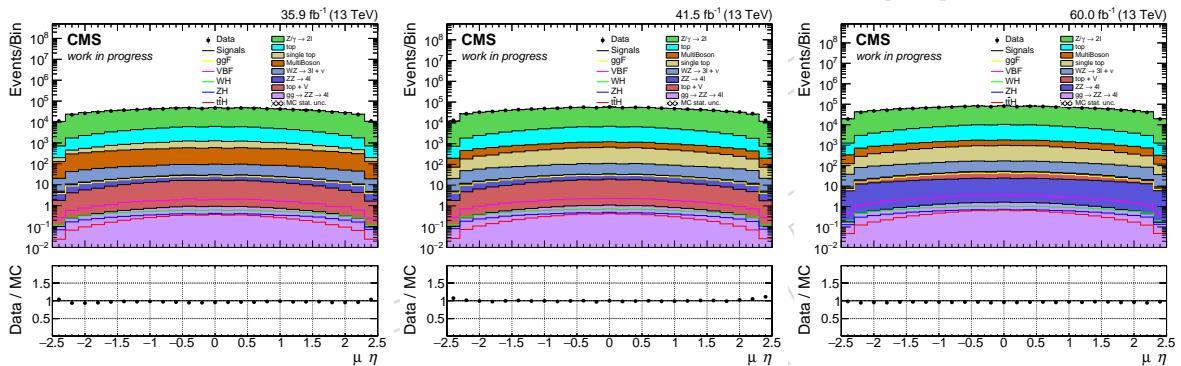


Figure E.2: Muon η for 2016 (left), 2017 (center), 2018 (right)

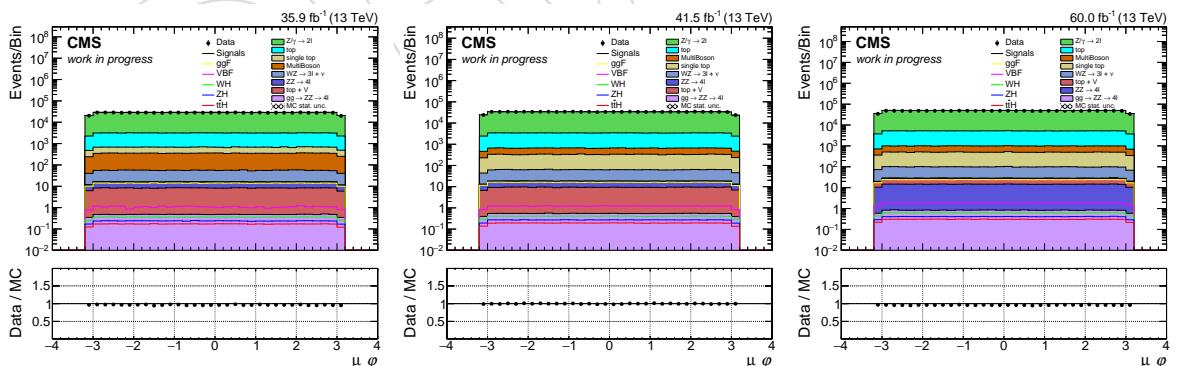


Figure E.3: Muon ϕ for 2016 (left), 2017 (center), 2018 (right)

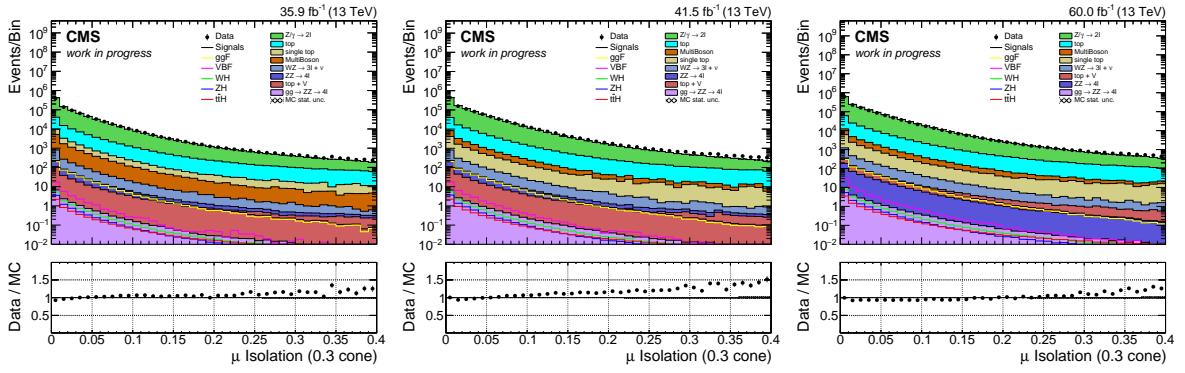


Figure E.4: Muon Isolation03 for 2016 (left), 2017 (center), 2018 (right)

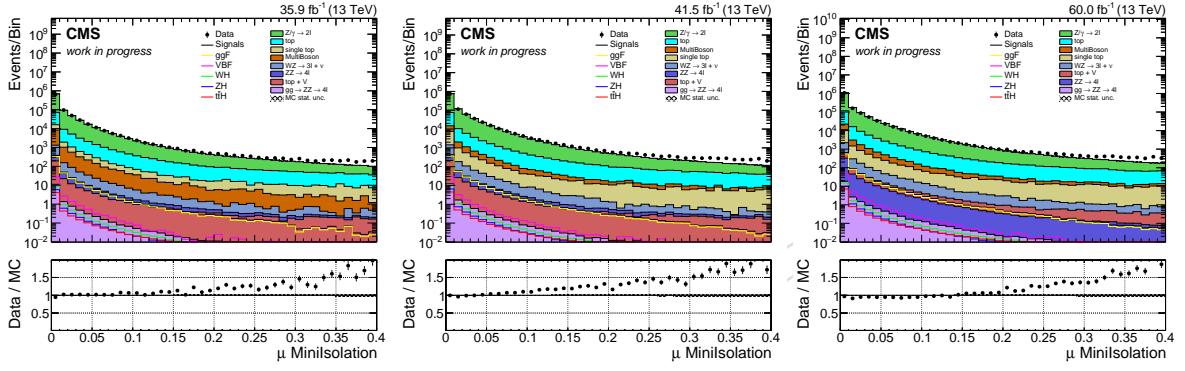


Figure E.5: Muon MiniIso for 2016 (left), 2017 (center), 2018 (right)

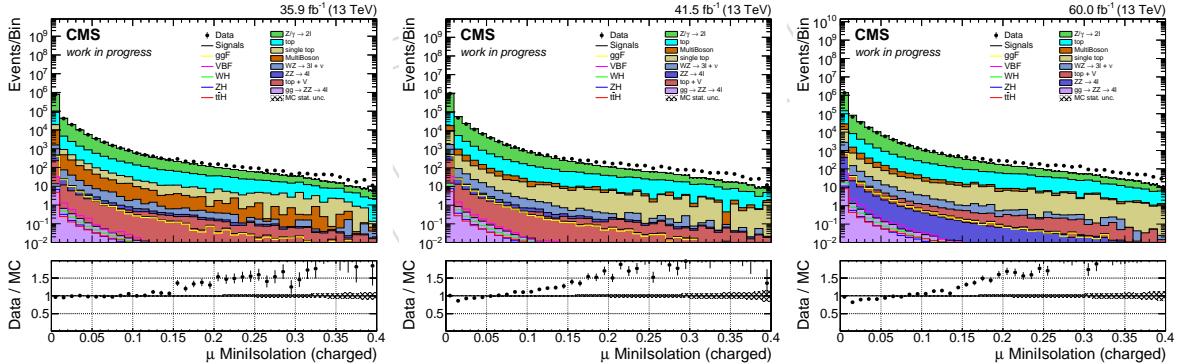


Figure E.6: Muon MiniIso (charged) for 2016 (left), 2017 (center), 2018 (right)

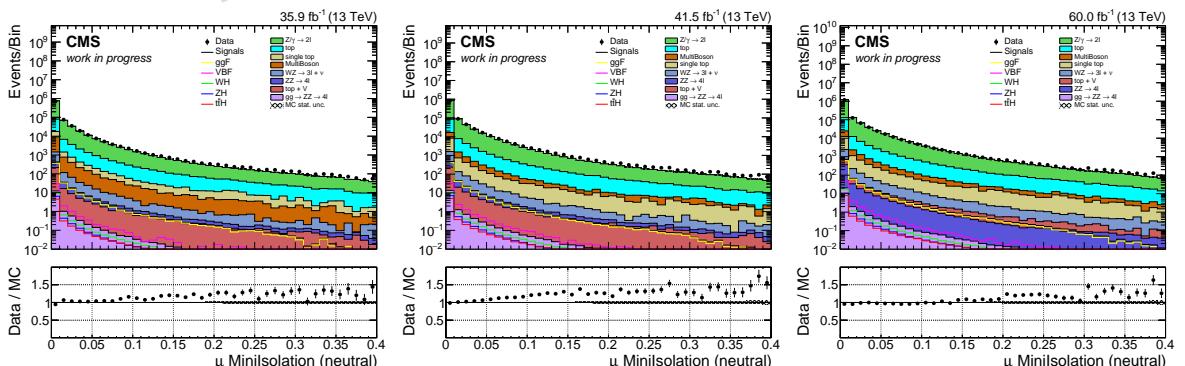


Figure E.7: Muon MiniIso (neutral) for 2016 (left), 2017 (center), 2018 (right)

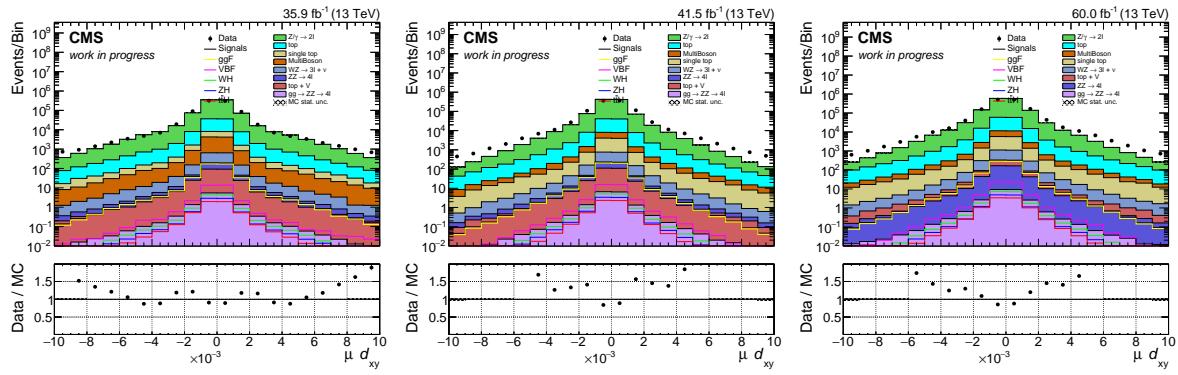
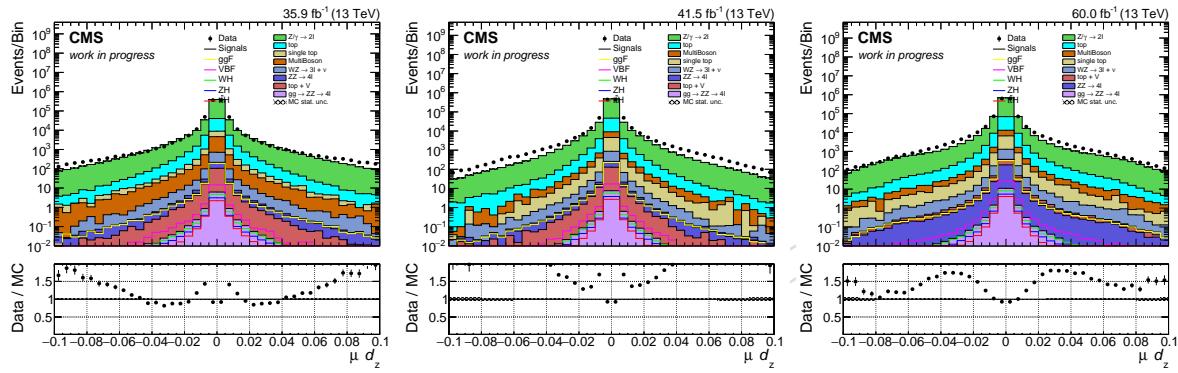
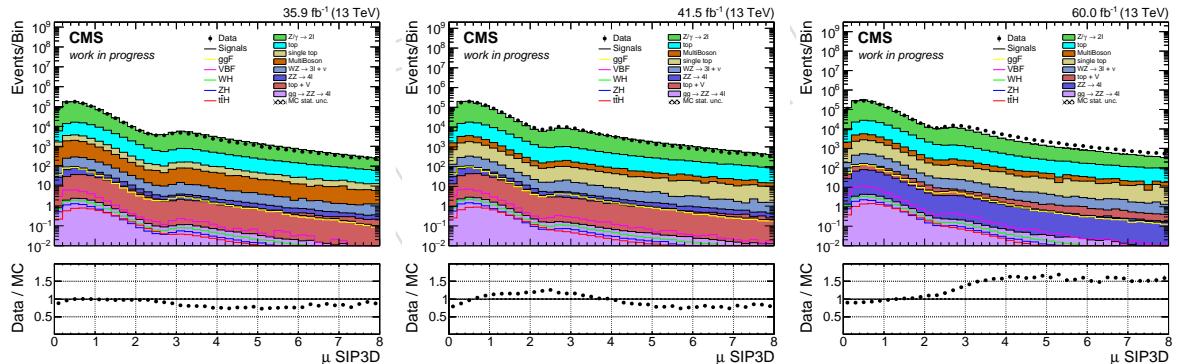
Figure E.8: Muon d_{xy} for 2016 (left), 2017 (center), 2018 (right)Figure E.9: Muon d_z for 2016 (left), 2017 (center), 2018 (right)

Figure E.10: Muon SIP3D for 2016 (left), 2017 (center), 2018 (right)

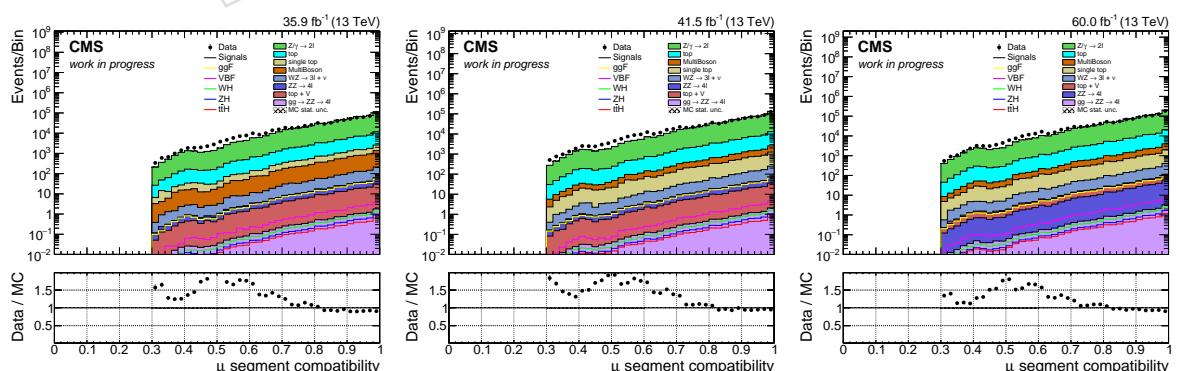


Figure E.11: Muon segment compatibility for 2016 (left), 2017 (center), 2018 (right)

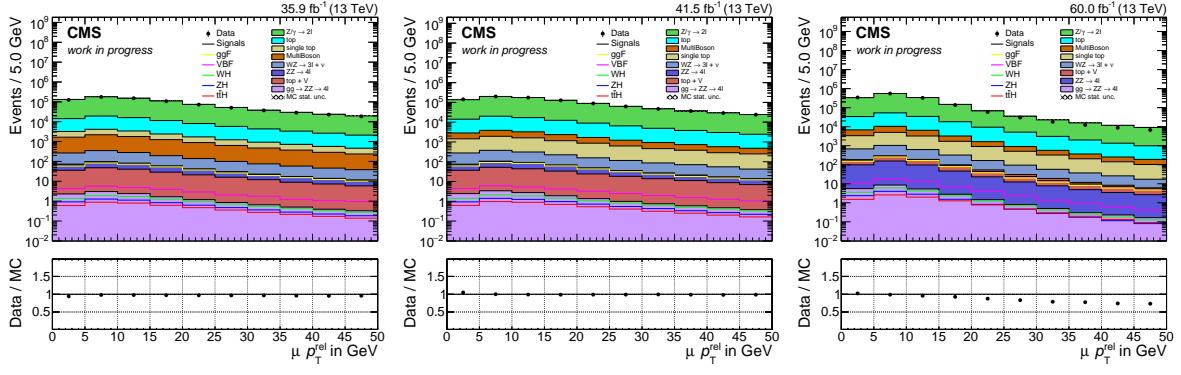


Figure E.12: Muon p_T^{rel} for 2016 (left), 2017 (center), 2018 (right)

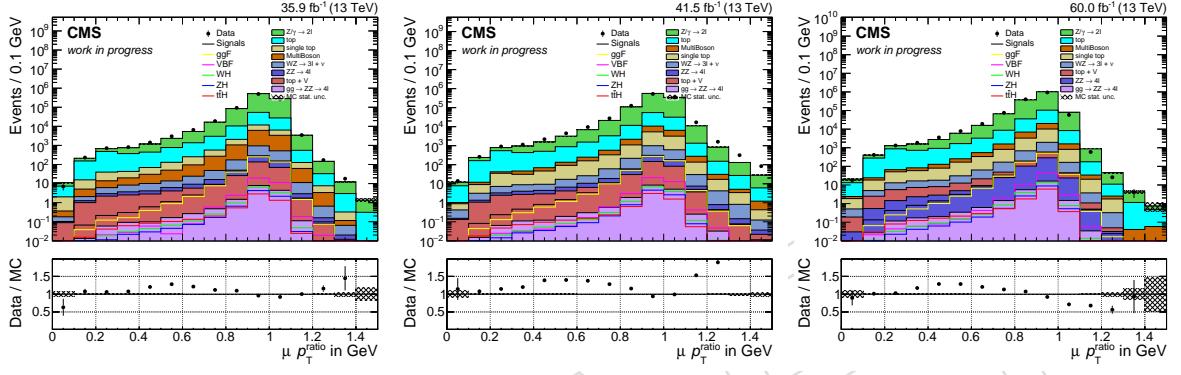


Figure E.13: Muon p_T^{ratio} for 2016 (left), 2017 (center), 2018 (right)

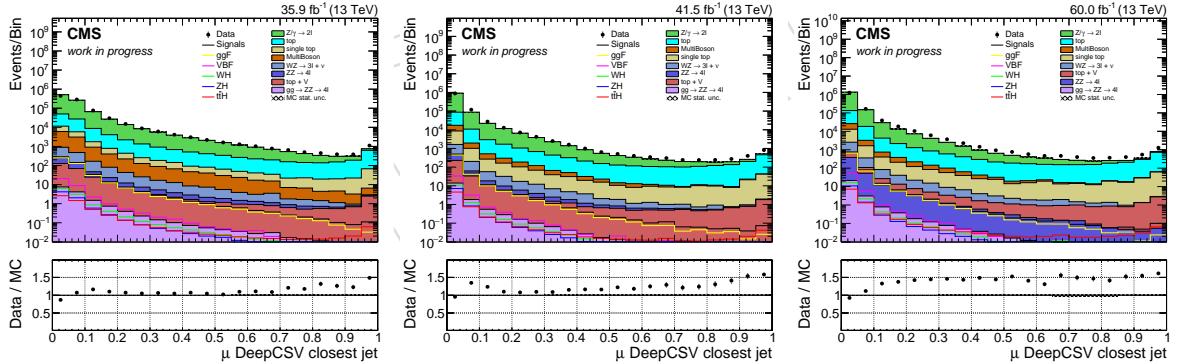


Figure E.14: DeepCSV value of the closest jet to a muon for 2016 (left), 2017 (center), 2018 (right)

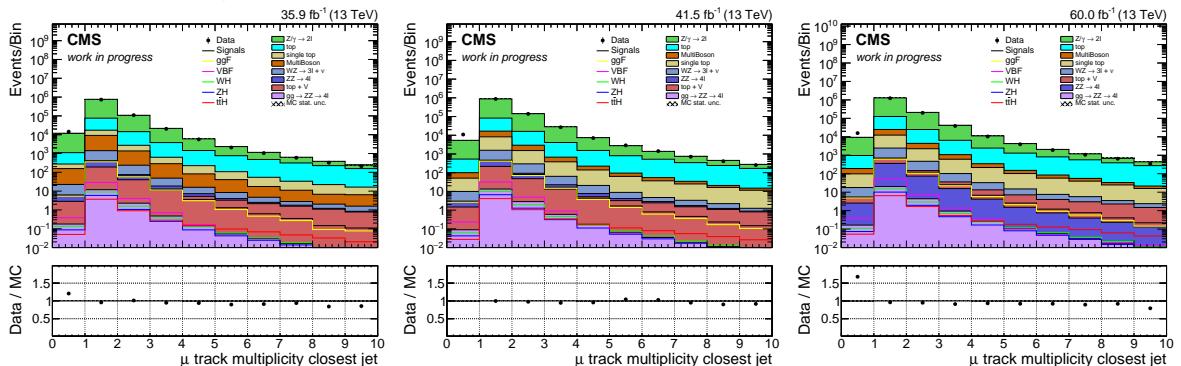


Figure E.15: Track multiplicity of the closest jet to a muon for 2016 (left), 2017 (center), 2018 (right)

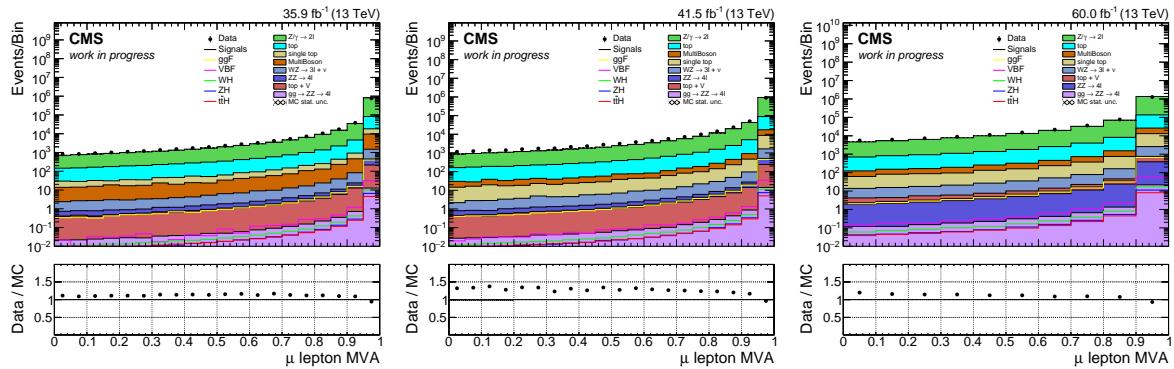


Figure E.16: Muon LeptonMVA output for 2016 (left), 2017 (center), 2018 (right)

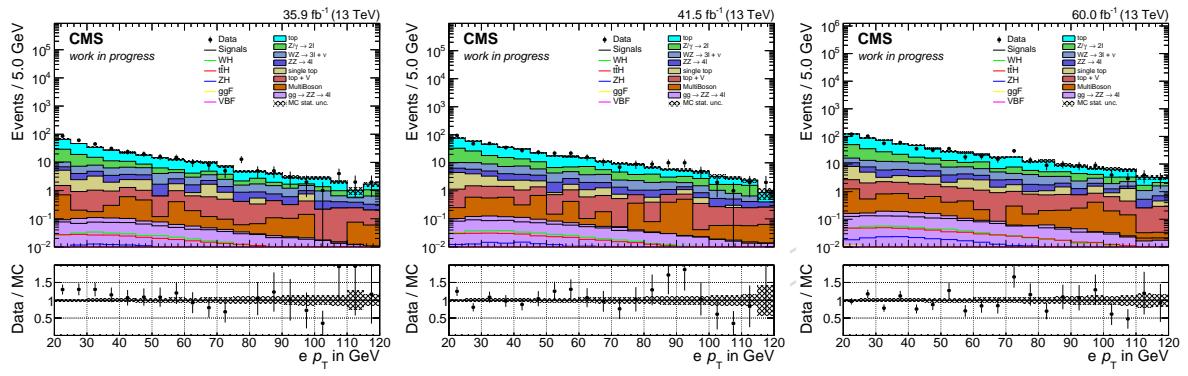


Figure E.17: Electron p_T for 2016 (left), 2017 (center), 2018 (right)

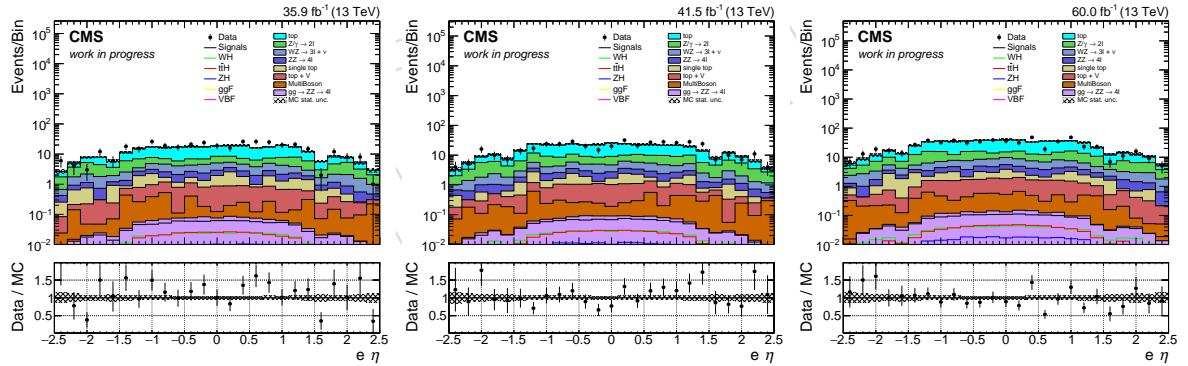


Figure E.18: Electron η for 2016 (left), 2017 (center), 2018 (right)

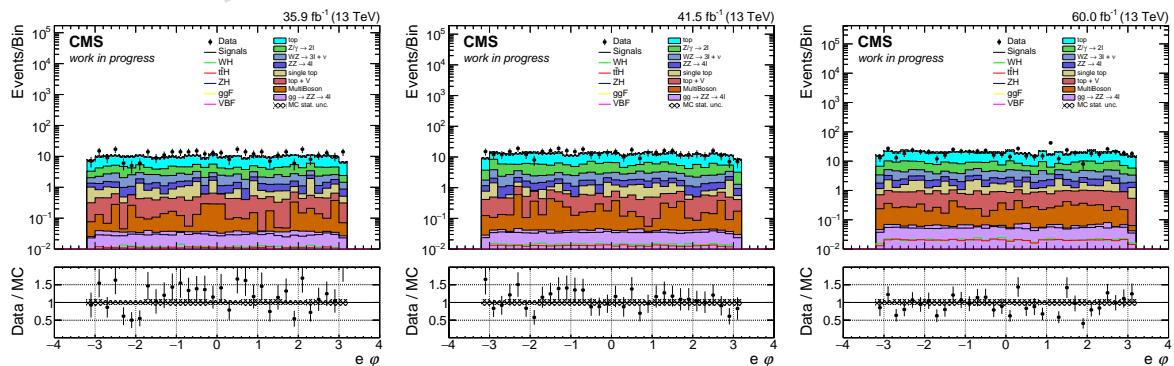


Figure E.19: Electron ϕ for 2016 (left), 2017 (center), 2018 (right)

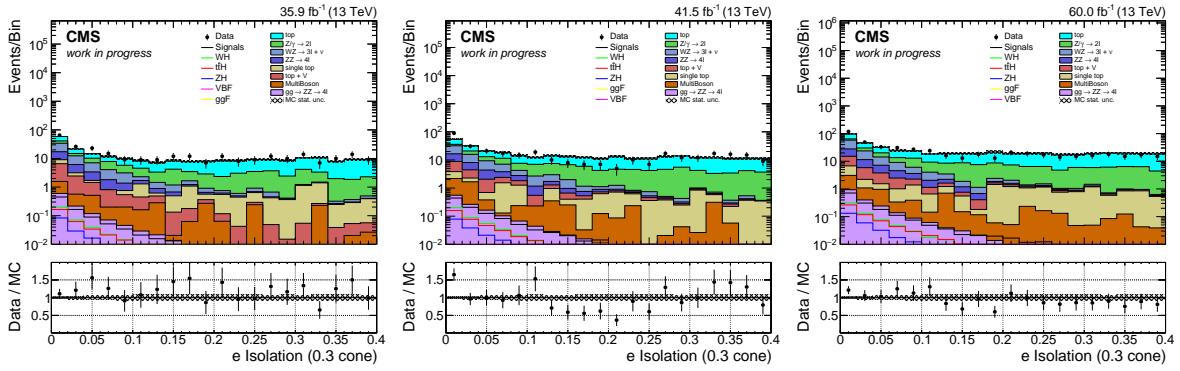


Figure E.20: Electron Isolation03 for 2016 (left), 2017 (center), 2018 (right)

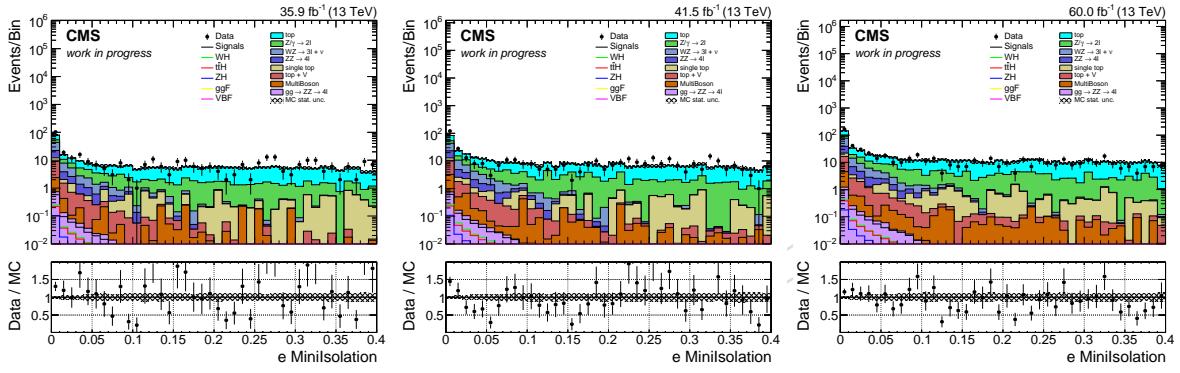


Figure E.21: Electron MiniIso for 2016 (left), 2017 (center), 2018 (right)

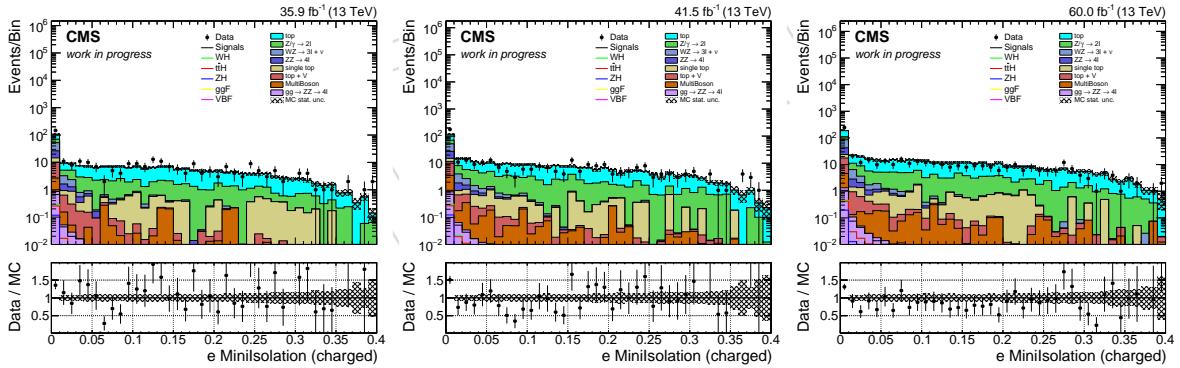


Figure E.22: Electron MiniIso (charged) for 2016 (left), 2017 (center), 2018 (right)

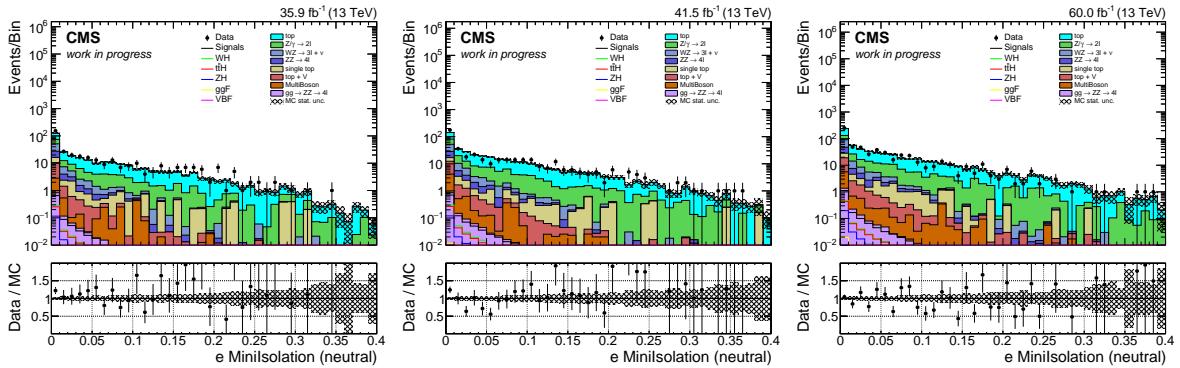


Figure E.23: Electron MiniIso (neutral) for 2016 (left), 2017 (center), 2018 (right)

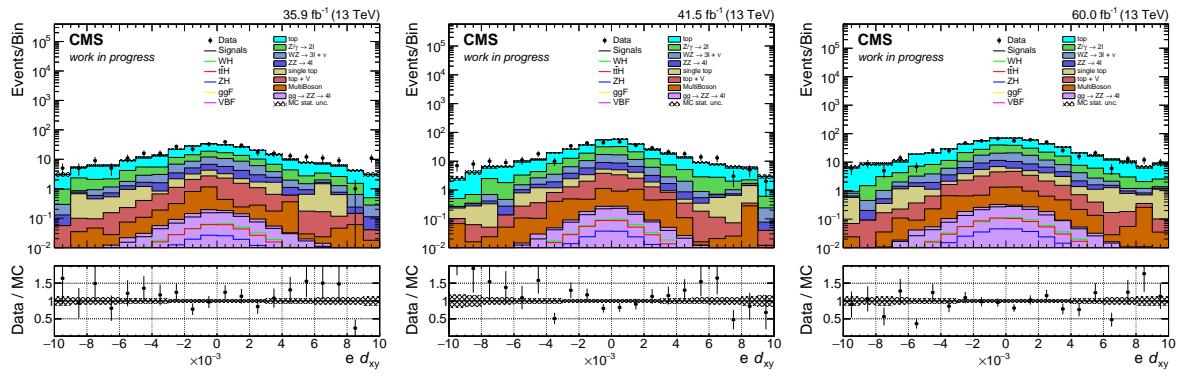
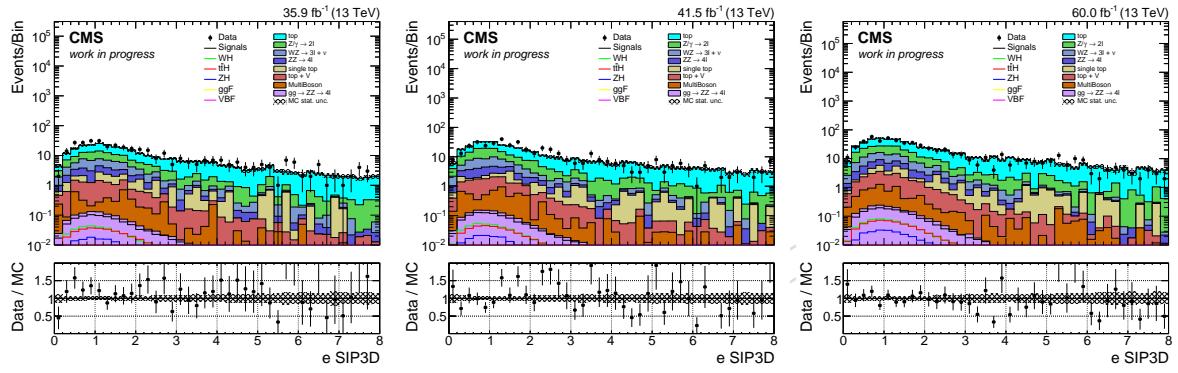
Figure E.24: Electron d_{xy} for 2016 (left), 2017 (center), 2018 (right)

Figure E.25: Electron SIP3D for 2016 (left), 2017 (center), 2018 (right)

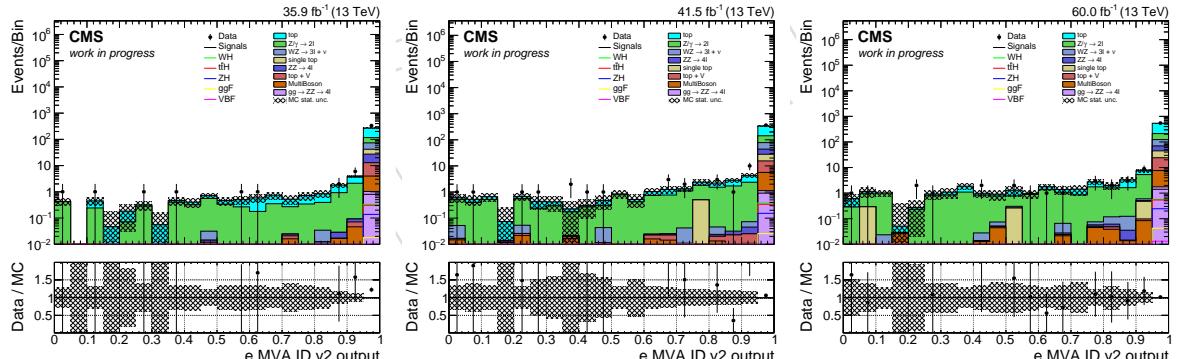
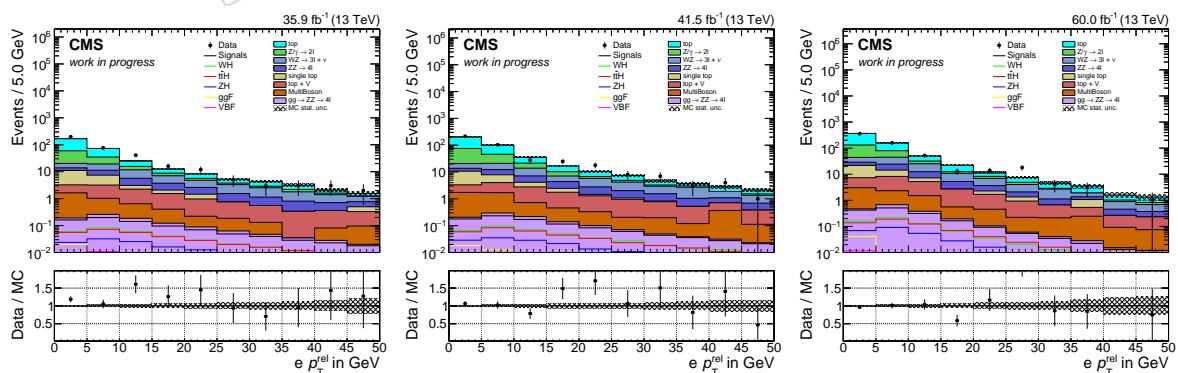


Figure E.26: Electron MVA Id (V2) value for 2016 (left), 2017 (center), 2018 (right)

Figure E.27: Electron p_T^{rel} for 2016 (left), 2017 (center), 2018 (right)

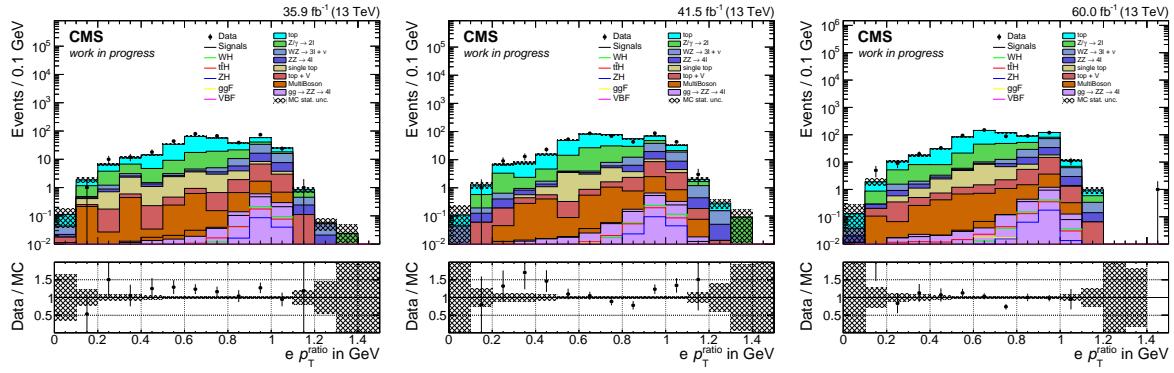


Figure E.28: Electron p_T^{ratio} for 2016 (left), 2017 (center), 2018 (right)

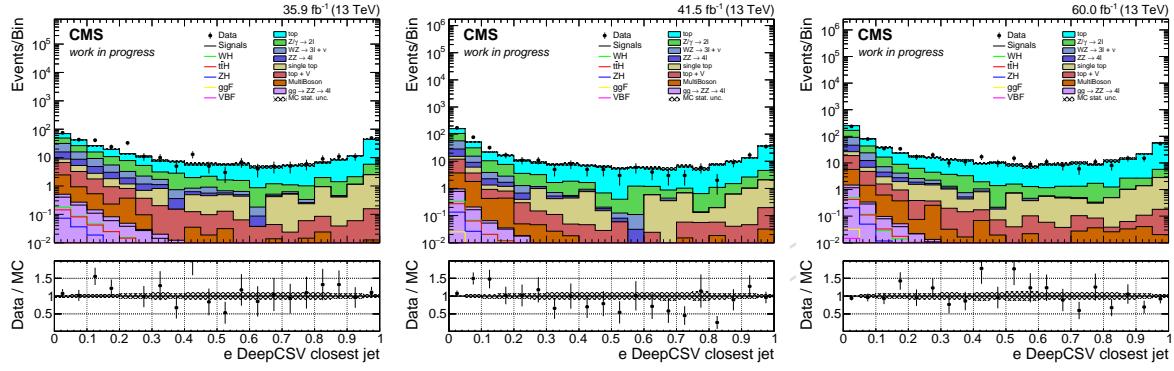


Figure E.29: DeepCSV value of the closest jet to an electron for 2016 (left), 2017 (center), 2018 (right)

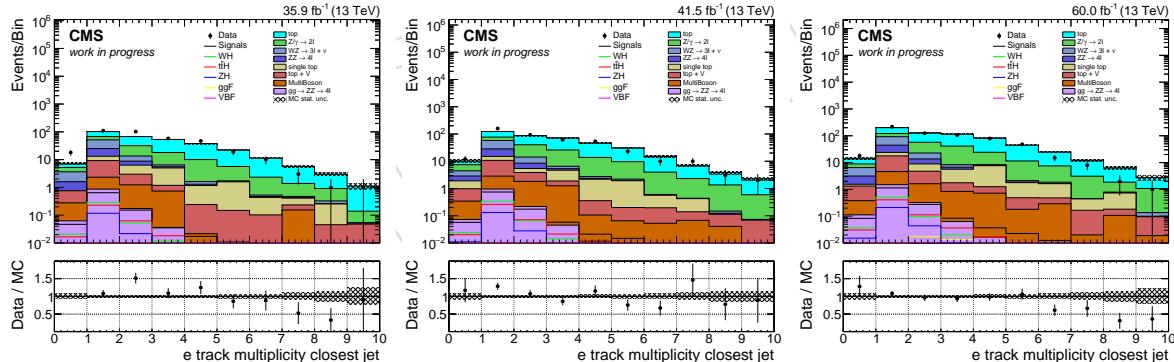


Figure E.30: Track multiplicity of the closest jet to an electron for 2016 (left), 2017 (center), 2018 (right)

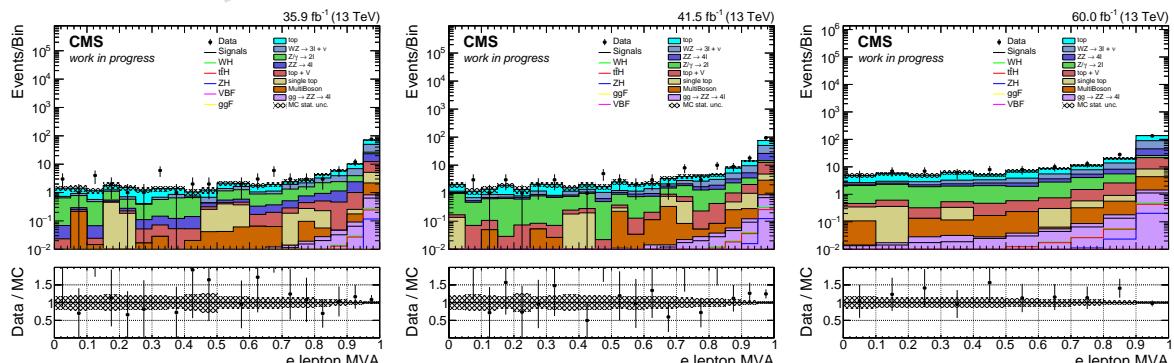


Figure E.31: Electron LeptonMVA output for 2016 (left), 2017 (center), 2018 (right)

3832 F Models for signal paramterization

3833 The data-driven VBF/ggH category (Section 6.2), VH category (Section 9) and ttH category
 3834 all use parametrized signal lineshapes, which are defined by fitting the dimuon
 3835 mass spectrum of signal MC, using identical selections to those applied to data. The particular
 3836 fit function has been chosen based on the study of multiple models, which is summarized in
 3837 this appendix.

3838 The shape of the signal is a narrow peak at 125 GeV in the dimuon mass spectrum. The width
 3839 of the Higgs boson in the standard model is ≈ 4 MeV, but in the detector the signal appears
 3840 wider due to:

- 3841 • **detector resolution**, that results into Gaussian-like widening of the peak by 1 –
 3842 4 GeV, depending on the pseudorapidity of the muons
- 3843 • **final state radiation (FSR)**, due to which approximately 8.5% of signal events mi-
 3844 grate to the tail in the lower-mass part of the spectrum.

3845 In order to take into account the detector resolution and the FSR effects, a set of parametric
 3846 models has been considered. Each of the considered models has a single or a double Gaussian
 3847 as a core, to model the detector resolution, and a broad Gaussian or a power-law function to
 3848 model the FSR tail. The considered options are listed below and shown in Figure F.1.

- 3849 1. **A sum of two Gaussians.** In this case one narrow Gaussian models the detector reso-
 3850 lution, and another broader Gaussian models the FSR tail. The model doesn't provide a
 3851 good fit of the signal spectrum, which is expressed by a high value of $\chi^2/\text{d.o.f}$ of the fit.
- 3852 2. **A sum of three Gaussians.** Here the core of the distribution is modelled by two Gaus-
 3853 sians, and the FSR tail is again modelled with a broad Gaussian. This model provides a
 3854 good fit quality and has been used in the past to model the signal shape in $H \rightarrow \mu\mu$ search
 3855 by CMS [13]. However, this model uses two width parameters to describe the width of
 3856 the peak, which is inconvenient for resolution studies as well as for implementing the
 3857 shape uncertainties.
- 3858 3. **Crystal Ball (CB)** [107]. This is a piecewise function, defined as a Gaussian with a power-
 3859 law tail at the lower-mass end. The model provides a good fit of data in the peak as well
 3860 as in the FSR tail, but fails to correctly model the higher-mass tail of the distribution.

$$\text{CB}(x; \mu, \sigma, \alpha, n) = \begin{cases} e^{-\frac{(x-\mu)^2}{2\sigma^2}} & \frac{x-\mu}{\sigma} > \alpha \\ \left(\frac{n}{|\alpha|}\right)^n \cdot e^{-\frac{\alpha^2}{2}} \cdot \left(\frac{n}{|\alpha|} - |\alpha| - \frac{x-\mu}{\sigma}\right)^{-n} & \frac{x-\mu}{\sigma} \leq -\alpha \end{cases}$$

- 3861 4. **Double-sided Crystal Ball (DCB).** This model is a piecewise function similar to Crystal
 3862 Ball, but here both of the tails are modelled with power-law functions. The model pro-
 3863 vides a good fit of the peak as well as both tails, and only has one width parameter (width
 3864 of the core Gaussian), which makes it a convenient choice for mass resolution studies. The
 3865 scale and resolution uncertainties can be implemented in a natural way, as uncertainties
 3866 on the peak position and the width of the core Gaussian.

$$\text{DCB}(x; \mu, \sigma, \alpha_{L/R}, n_{L/R}) = \begin{cases} e^{-\frac{(x-\mu)^2}{2\sigma^2}} & -\alpha_L < \frac{x-\mu}{\sigma} < \alpha_R \\ \left(\frac{n_L}{|\alpha_L|}\right)^{n_L} \cdot e^{-\frac{\alpha_L^2}{2}} \cdot \left(\frac{n_L}{|\alpha_L|} - |\alpha_L| - \frac{x-\mu}{\sigma}\right)^{-n_L} & \frac{x-\mu}{\sigma} \leq -\alpha_L \\ \left(\frac{n_R}{|\alpha_R|}\right)^{n_R} \cdot e^{-\frac{\alpha_R^2}{2}} \cdot \left(\frac{n_R}{|\alpha_R|} - |\alpha_R| - \frac{x-\mu}{\sigma}\right)^{-n_R} & \frac{x-\mu}{\sigma} \geq \alpha_R \end{cases}$$

- 3867 5. **A sum of Crystal Ball and a Gaussian.** This model has been previously used in $H \rightarrow \mu\mu$
 3868 search by ATLAS [14]. It provides a good description of the peak and both tails, but uses
 3869 two Gaussians to describe the core.
- 3870 6. **A sum of Double-sided Crystal Ball and a Gaussian.** This model provides the best fit
 3871 quality among the considered options, but also has the highest number of parameters.
 3872 The core is again described by two Gaussians.

3873 Basing on the features of the considered functions, Double-sided Crystal Ball has been chosen
 3874 for signal parametrization, due to its good fit quality, single width parameter in the model, and
 3875 straightforward implementation of shape uncertainties. The effect of switching from a sum of
 3876 three Gaussians to DCB was evaluated by fitting the same signal dataset with both functions
 3877 and calculating expected significance in both cases. The difference in expected significance
 3878 didn't exceed 0.5%.

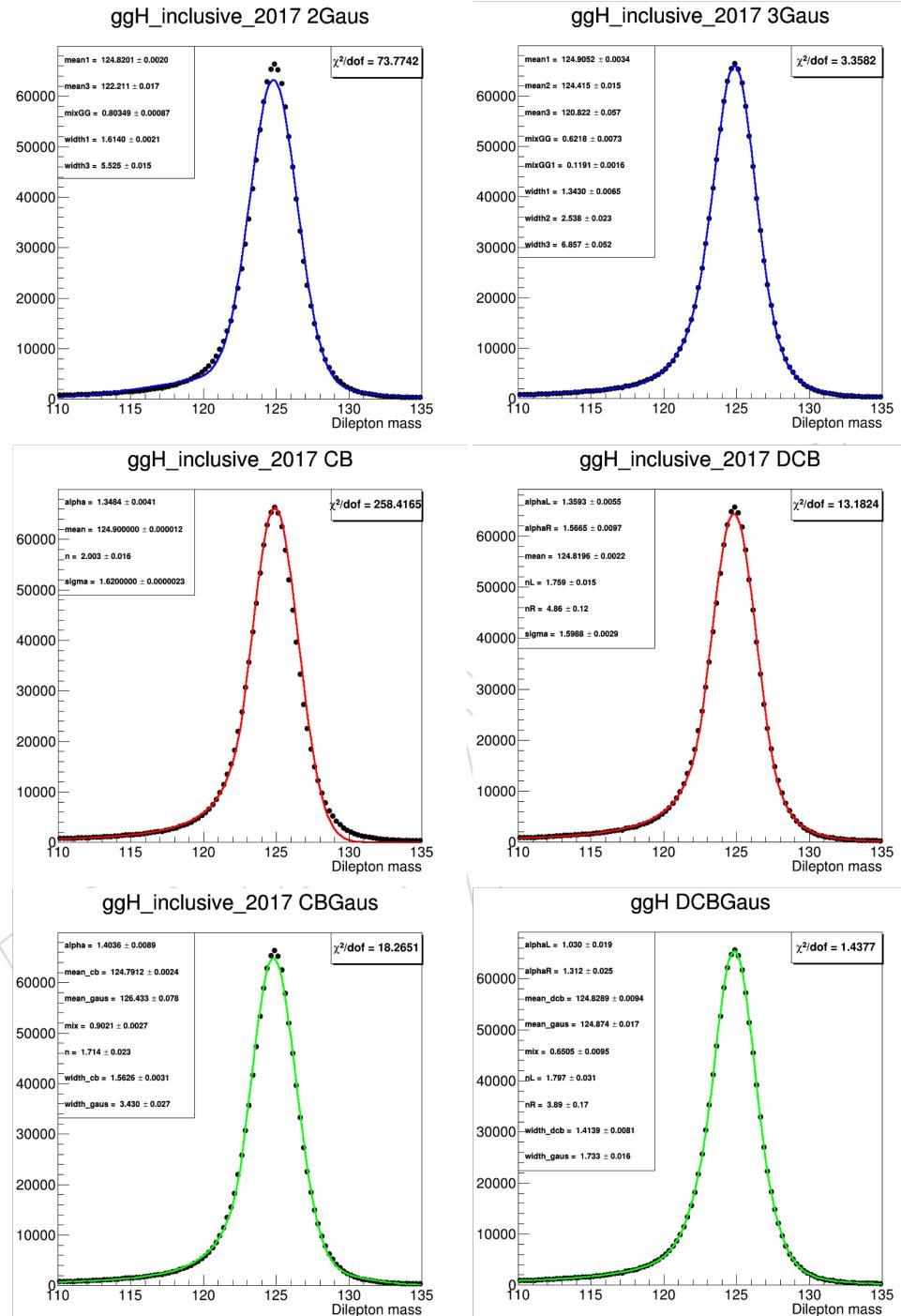


Figure F.1: Parametric fits of ggH MC. From left to right: **top row:** a sum of two Gaussians and a sum of three Gaussians; **middle row:** Crystal Ball (CB) and Double-sided Crystal Ball (DCB); **bottom row:** a sum of CB and a Gaussian and a sum of DCB and a Gaussian.

3879 G Hadronic top quark reconstruction

3880 The resolved hadronic top quark tagger (RHTT) was developed to reconstruct hadronically
 3881 decaying top quarks from a combination of three jets, referred to as 'jet-triplet' where one jet
 3882 originates from the b quark and the remaining two jets were produced in the W boson decay.
 3883 Each of these jets can be reconstructed as a resolved AK4 jet. This targets especially low to
 3884 moderately boosted hadronically decaying top quarks. The jet-triplets can be characterized by
 3885 different kinematic properties. This is used to separate jets matched to quarks emerging from
 3886 a top quark decay (signal) from arbitrary jet permutations (background). The RHTT algorithm
 3887 is based on a combination of a constrained kinematic fit [108] and a particle-level BDT. The
 3888 purpose of the kinematic fit is to measure the kinematic compatibility of a jet-triplet with the
 3889 hypothesis of a hadronic top quark decay, and the BDT enables the exploitation of further
 3890 intrinsic jet properties like e.g. the b-tag probability.

During the kinematic fit each jet four-momentum vector is varied within its uncertainties, where a general 1-dim χ^2 term has the following form

$$\chi^2 = \frac{y_{fit} - y_{meas}}{\sigma_y^{meas}}.$$

The minimization of the thereby introduced least square function is performed considering two Gaussian mass constraints

$$b + j1 + j2 = \text{top} \quad \text{with} \quad m_{top} = 173.0 \text{ GeV}, \sigma_{top} = 1.41 \text{ GeV}$$

and

$$j1 + j2 = W \quad \text{with} \quad m_W = 80.4 \text{ GeV}, \sigma_W = 2.085 \text{ GeV}.$$

3891 For validation of the performance of the kinematic fit the pull distributions of E_t , η and ϕ for
 3892 each jet can be seen in Fig. G.1, G.2, G.3. The pull distributions show a reasonable behaviour of
 3893 the kinematic fit.

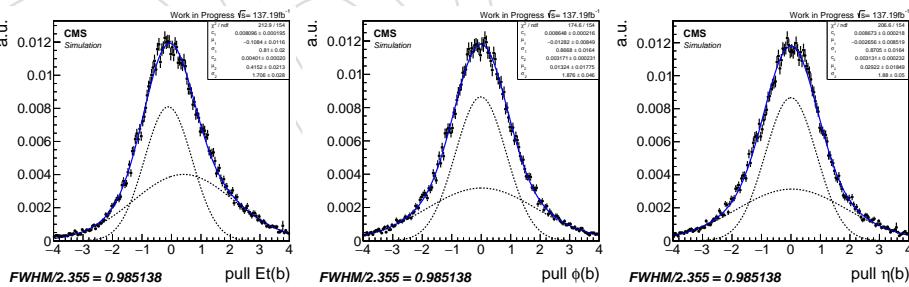


Figure G.1: Pull distributions for each jet of the jet triplet of the transverse energy E_t , ϕ and η . Here only generator matched top-quark jets are shown.

3894 The BDT is trained using a set of signal samples which contain actual hadronic top quark de-
 3895 cays (ttH) and background samples (DY+Jets, TTTTo2L2Nu) where only random jet combi-
 3896 nations occur. To separate signal from background the χ^2 normalized to the number of degrees of
 3897 freedom is used together with the modified four-momentum vectors of the jets after the kine-
 3898 matic fit is performed. The modified transverse momentum of the jets as well as their angular
 3899 relations provide information on the kinematic compatibility of the jet-triplet with the hypoth-
 3900 esis of a hadronically decaying top quark. Furthermore, the initially reconstructed values of the
 3901 top quark mass m_{top} as well as the W Boson mass m_W are included. Finally, the deepCSV and

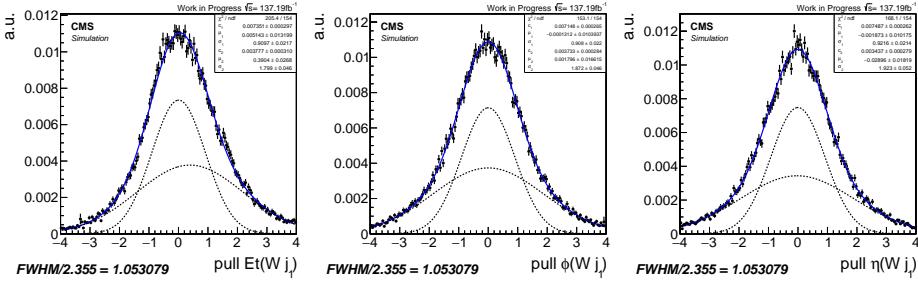


Figure G.2: Pull distributions for each jet of the jet triplet of the transverse energy E_t , ϕ and η . Here only generator matched top-quark jets are shown.

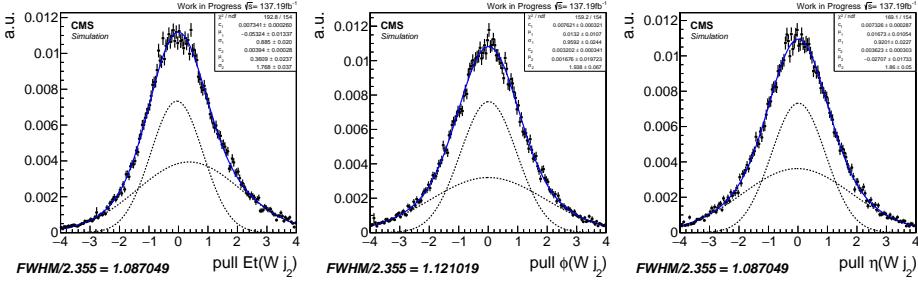


Figure G.3: Pull distributions for each jet of the jet triplet of the transverse energy E_t , ϕ and η . Here only generator matched top-quark jets are shown.

QuarkGluon likelihood are used to improve the correct matching of jet permutations, so that the jet from the b quark is considered correctly in the kinematic equations. The corresponding distributions of the input variables can be found in Fig. G.4, G.5, G.6, G.7, G.8, and G.9.

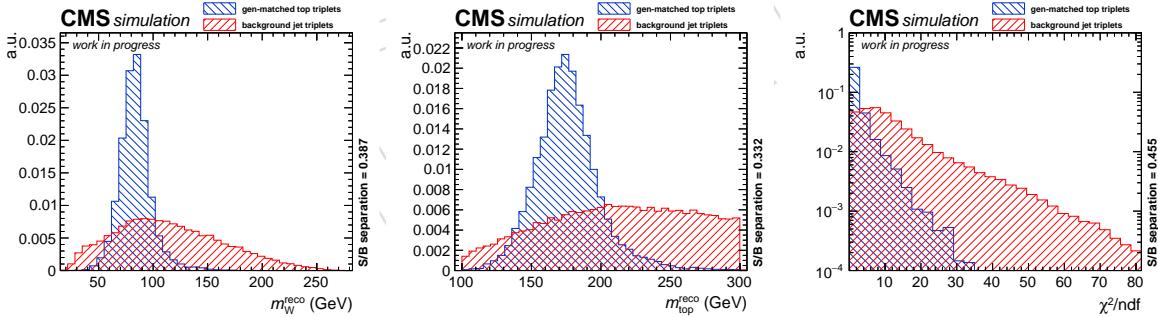


Figure G.4: Normalized input variables to the RHTT.

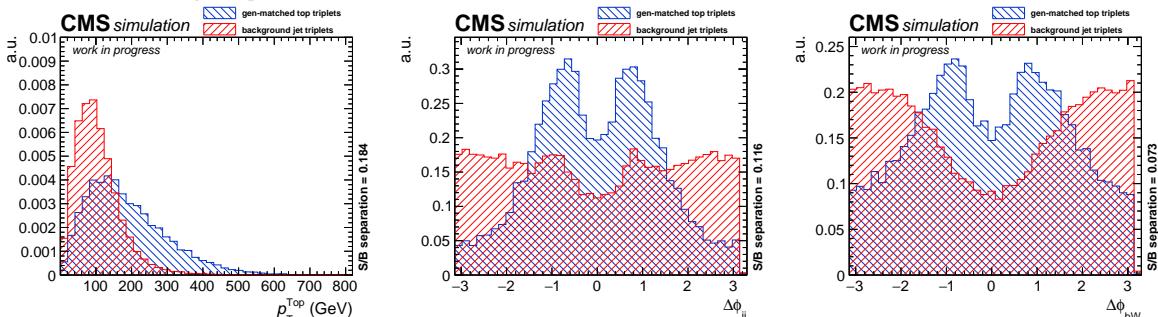


Figure G.5: Normalized input variables to the RHTT.

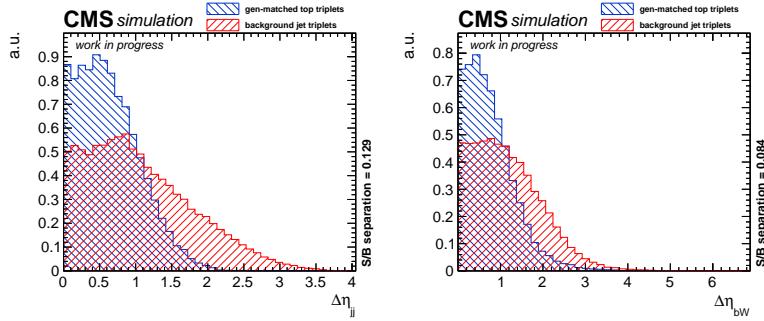


Figure G.6: Normalized input variables to the RHTT.

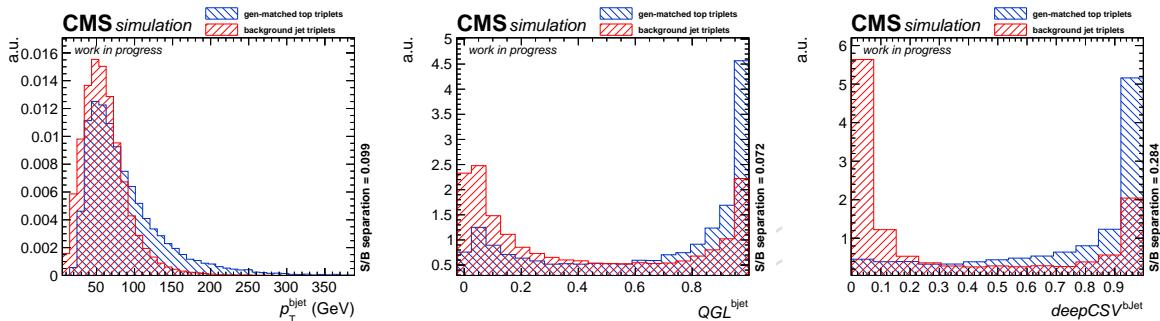


Figure G.7: Normalized input variables to the RHTT.

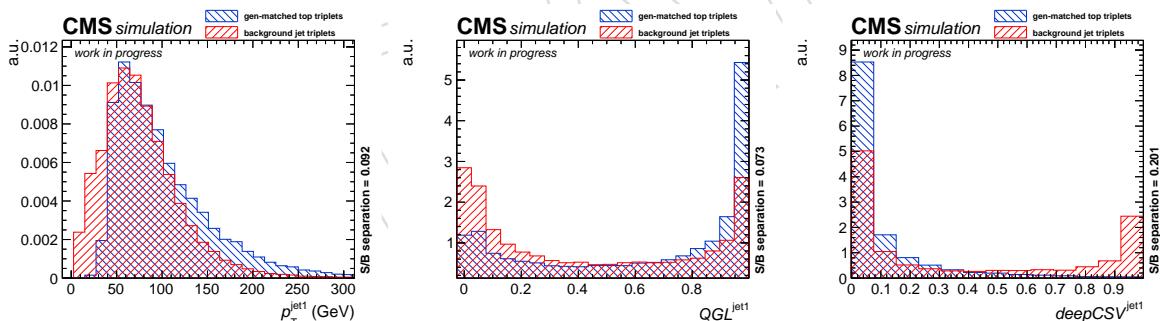


Figure G.8: Normalized input variables to the RHTT.

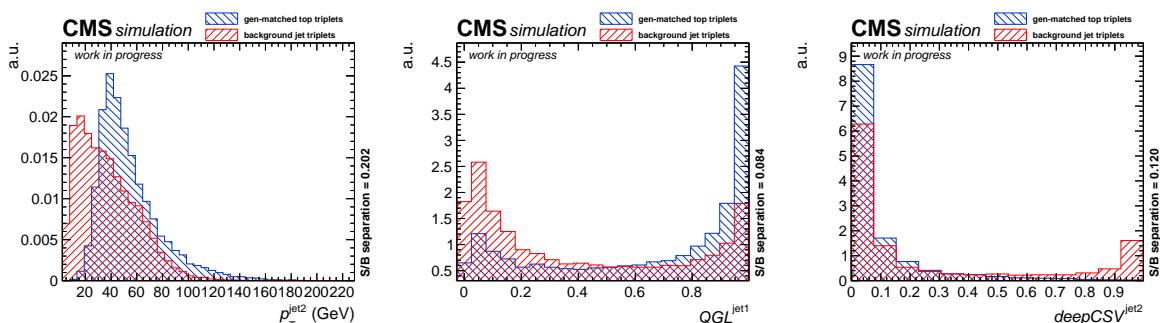


Figure G.9: Normalized input variables to the RHTT.

3905 H Resolution categories

3906 A different approach to the weighting of the BDT with the mass resolution is to further categorize
 3907 the data into the regions sensitive to the resolution. This strategy and the optimization of
 3908 these categories for the full Run 2 dataset is explained in this section.

3909 A conservative approach, where the resolution information is taken into account by splitting
 3910 the data into pseudorapidity categories, was considered as an alternative to the approach that
 3911 uses event-by-event mass resolution, as described in Section 4.8.

3912 The largest absolute value of pseudorapidity of the two muons was considered as a splitting
 3913 variable for categorization, because the dominating effect on the dimuon mass resolution
 3914 comes from the muon with the worse momentum resolution. This is illustrated in Figure H.1,
 3915 where the width parameter of signal fit with double-sided Crystal-Ball is plotted as a function
 3916 of the largest absolute value of pseudorapidity of the two muons.

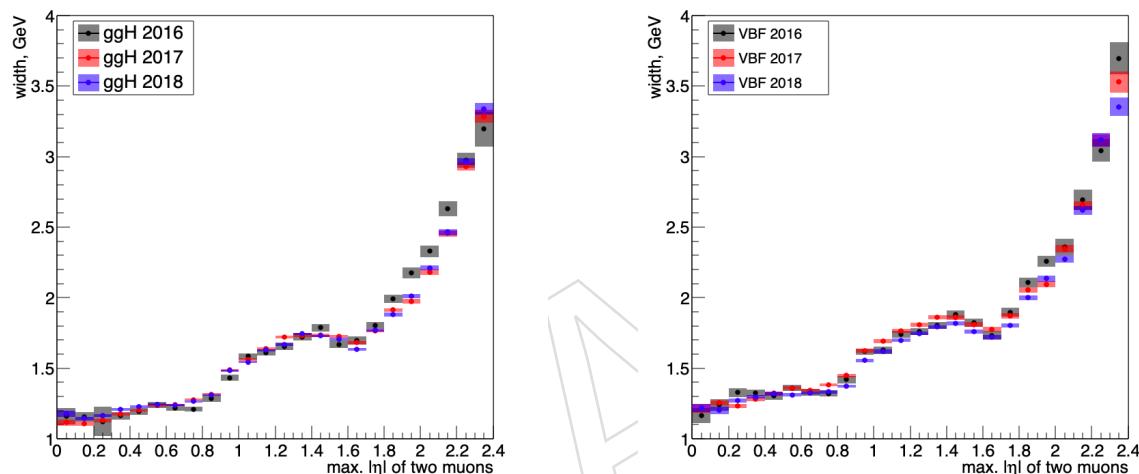


Figure H.1: Left: dimuon mass resolution of ggH events; right: dimuon mass resolution of VBF events.

3917 The number of the pseudorapidity categories was limited to three, and the optimization procedure
 3918 was performed in order to optimize the category boundaries.

3919 All possible ways to set two cuts between $|\eta| = 0$ and $|\eta| = 2.4$, where the cut values are limited
 3920 to the numbers with one decimal place have been considered (253 options in total). The figure
 3921 of merit was the expected significance evaluated using Asimov dataset, that was generated
 3922 using the parametric signal and background shapes. The signal shapes were obtained by fitting
 3923 the ggH MC mass spectrum with double-sided Crystal Ball, and the background shapes were
 3924 derived by fitting the sidebands of the mass spectrum of 2016 data. The signal MC was scaled
 3925 to 35.9fb^{-1} in all cases for the comparison purposes.

3926 Figure H.2 shows the two-dimensional plots of the expected significance as a function of the
 3927 cut values for 2016, 2017, and 2018 MC. Although the shape of the plots is slightly different
 3928 for different years, it has been decided to select the same cuts for all data-taking periods. The
 3929 option with the cuts placed at $|\eta| = 0.9$ and $|\eta| = 1.9$ (marked by the red squares on the plots)
 3930 gives in all cases the significance that is different from the maximum value by no more than
 0.1%.

3932 The strategy of using resolution categories didn't show any improvement in significance as
 3933 compared to the strategy with event-by-event mass resolution, while the total number of cate-

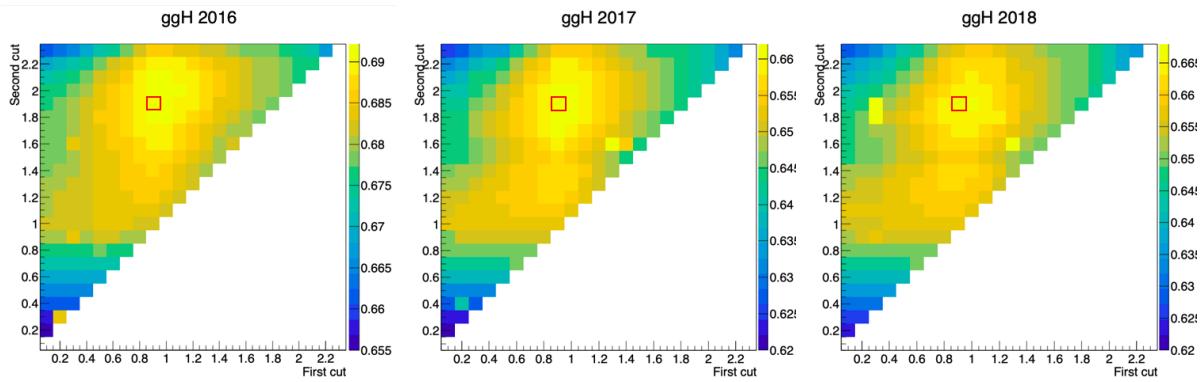


Figure H.2: Expected signal significance as a function of the pseudorapidity cuts for 2016, 2017, 2018 MC. The lower and the higher rapidity cuts correspond to the horizontal and the vertical axes, respectively. The red squares mark the option with the cuts at $|\eta| = 0.9$ and $|\eta| = 1.9$.

3934 gories was lower in the latter case. Therefore, it has been decided to take the mass resolution
 3935 information into account by using event-by-event resolution.

DRAFT

3936 I Studies of Drell-Yan background with FEWZ

3937 The Drell-Yan process is a major background contributor in the search for the Higgs boson de-
 3938 caying to dimuons. Thus, it is important to have a function that describes this process well. In
 3939 order to find such a function, the program FEWZ [36] was used to create next-to-leading order
 3940 (NLO) and next-to-next-to-leading order (NNLO) predictions for the dimuon mass spectrum
 3941 of the Drell-Yan process and then various functions were fitted to these histograms.

3942 Parton distribution functions (PDF) from the LHAPDF library [109–112] were used by FEWZ
 3943 to calculate the muon pair production in 13 TeV proton-proton collisions at the LHC: here
 3944 the LUXqed_plus_PDF4LHC15_nnlo_100 [113] PDF. It includes the complete QED component
 3945 (coherent and non-coherent) and enables the calculation of the photon-induced effects [37] in
 3946 addition to the Drell-Yan process (for a detailed discussion of all the effects see e.g. [38]). The
 3947 LUXqed is a NNLO PDF, we have used it both for NNLO and NLO calculations.

3948 I.1 FEWZ Calculations

3949 FEWZ was used to calculate the NLO and NNLO dimuon mass histograms for two different
 3950 mass windows: 110 to 160 GeV with 50 bins of one GeV, and 105 to 160 GeV with 55 bins. These
 3951 histograms were calculated at NNLO for the inclusive category. The definition is in Table I.1.

Table I.1: FEWZ category definitions for NLO and NNLO calculations.

	Inclusive
Higher- p_T muon	> 20 GeV
Lower- p_T muon	> 10 GeV
Muon η	$ \eta < 2.4$
Number of jets	≤ 2
Jet p_T	NA

3952 The evolution of the calculations was as follows:

- 3953 • Starting point: 1 histogram 110–160 GeV, 50x1 GeV bins
- 3954 • Intermediate: 3 histograms 105–125, 120–140, 140–160 GeV, 55x1 GeV bins, overlap
 3955 120–125 GeV
- 3956 • Final step: 6 histograms 105–125, 113–133, 120–140, 130–150, 140–160, 150–170 GeV,
 3957 55x1 GeV bins, overlaps 113–120, 120–130, 130–140, 140–150 and 150–160 GeV, al-
 3958 ways using the higher statistics (left) half of the steeply falling DY spectra

3959 I.2 Fit Functions

3960 The FEWZ calculated dimuon mass histograms were fitted with the different functions, where
 3961 p_i are the parameters and m is the mass. The Z boson parameters are fixed to $m_Z = 91.1876$ GeV
 3962 and $\Gamma_Z = 2.4952$ GeV. Results are shown for the modified relativistic Breit-Wigner form (MRBW)
 3963 multiplied by a two-parameter falling exponential function.

3964 Eq. I.1: A two parameter exponential distribution times a modified relativistic Breit-Wigner
 3965 distribution with the power p1 as a free parameter (MRBW)

$$3966 p0 * m^2 \frac{\exp(p2 \frac{m}{100} + p3 (\frac{m}{100})^2)}{(m^2 - m_Z^2)^{p1} + (\frac{\Gamma_Z \cdot m^2}{m_Z})^{p1}} \quad (I.1)$$

3966 **I.3 Results and Conclusions**

3967 The MRBW function detailed in the previous section is fitted to the FEWZ calculated dimuon
 3968 mass histograms using a least-square fit from the MINUIT package implemented in ROOT. The
 3969 MRBW fit does a good job for the 110–160 or 105–160 GeV cases with only 4 free parameters.

3970 **I.3.1 NNLO Results**

3971 The results of the calculations are tabulated in Table I.2.

3972 The MRBW fit results for Eq. I.1 are shown in Figures I.1 to I.3. The average of the positive and
 3973 negative PDF errors is used as a symmetric error for each mass bin. The MRBW fit function
 3974 describes well the dimuon mass region between 105 and 160 GeV. We can conclude that this fit
 3975 with 4 free parameters provides a good description of the NNLO Drell-Yan Spectra. The higher
 3976 χ^2 of the NNLO fits using just one histogram is due to the limits of the histogram method in
 3977 FEWZ, used to produce the 110–160 or 105–160 GeV differential distributions with one, very
 3978 time consuming, calculation for each case. This limits the precision of the one GeV bins and
 3979 induces “jitter” above the estimate of the statistical precision provided by FEWZ. By using 3
 3980 and then 6 calculations with 3, or 6 histograms respectively, we are able to substantially reduce
 3981 the “jitter”, increasing the smoothness of the differential distributions and improving the χ^2
 3982 of the fits. In summary, this “divide and conquer” strategy successfully solves the challenge
 3983 of producing high precision theory predictions for differential mass spectra with FEWZ in the
 3984 very computer intensive NNLO case.

3985 **I.3.2 Conclusion**

3986 The FEWZ calculations provide an up-to-date theory description of the Drell-Yan dimuon spec-
 3987 tra for different categories in the CMS acceptance for the Higgs search. They can be fitted with
 3988 physics motivated functions with only three or four free parameters. These functions can be
 3989 compared to the data or Monte-Carlo predictions, with only the normalization left as a free
 3990 parameter. Or, if needed, they can be adjusted to take experimental effects like resolution and
 3991 efficiency into account.

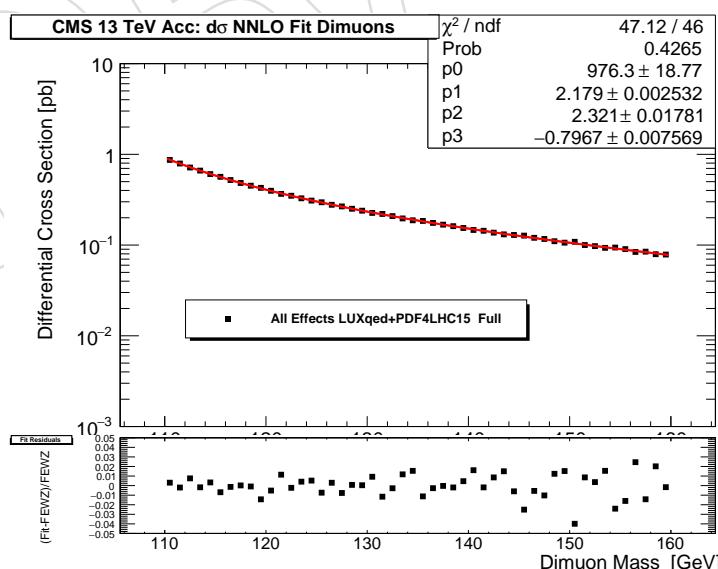


Figure I.1: The MRBW fit of Eq. I.1 to the differential NNLO cross section vs. dimuon mass between 110–160 GeV using one histogram.

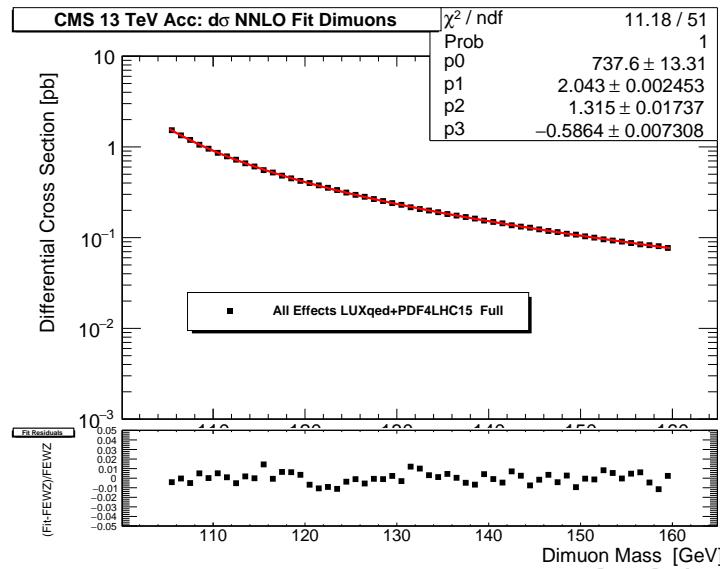


Figure I.2: The MRBW fit of Eq. I.1 to the differential NNLO cross section vs. dimuon mass between 105–160 GeV using three histograms.

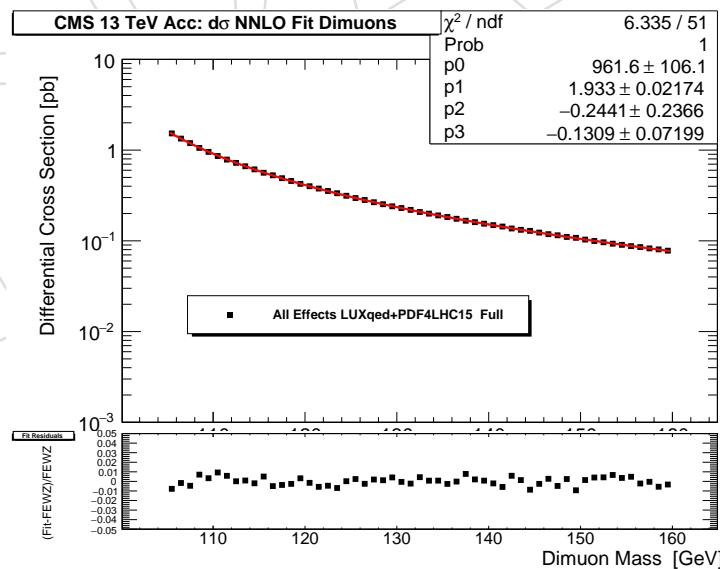


Figure I.3: The MRBW fit of Eq. I.1 to the differential NNLO cross section vs. dimuon mass between 105–160 GeV using six histograms.

Table I.2: Differential FEWZ spectrum at NNLO from six histograms. Cross sections in [pb].

Mass (GeV)	xsec (pb)	numerical error (pb)	+PDF error (pb)	-PDF error (pb)
105.50	1.52799	0.00315722	0.0216026	0.021008
106.50	1.33971	0.00307773	0.018819	0.0183896
107.50	1.1947	0.00269509	0.0166639	0.0162937
108.50	1.05731	0.0027113	0.0146619	0.0143865
109.50	0.956099	0.00251965	0.0131709	0.0129274
110.50	0.860989	0.00243462	0.011787	0.0116159
111.50	0.786549	0.0022344	0.0107113	0.0105746
112.50	0.72338	0.00215155	0.00976792	0.00969583
113.50	0.663691	0.00144907	0.00890965	0.0088634
114.50	0.613444	0.00136115	0.00818182	0.00816593
115.50	0.563279	0.00135635	0.00747735	0.00749429
116.50	0.527754	0.00115827	0.00695078	0.00696417
117.50	0.490348	0.00119192	0.00641746	0.00647491
118.50	0.456846	0.0011548	0.00594638	0.00602024
119.50	0.424666	0.00115472	0.00550524	0.00558815
120.50	0.39985	0.000927391	0.00515315	0.00525411
121.50	0.377076	0.000811078	0.0048303	0.00493814
122.50	0.354401	0.000817696	0.0045192	0.00463694
123.50	0.334895	0.000779854	0.00424225	0.00436977
124.50	0.313994	0.000796971	0.00396141	0.00409909
125.50	0.296318	0.000772704	0.00372586	0.00386698
126.50	0.282099	0.000726179	0.00352624	0.0036749
127.50	0.266413	0.000724196	0.00332048	0.00347042
128.50	0.25326	0.0006778	0.0031365	0.00329526
129.50	0.240175	0.000678717	0.00296817	0.00312408
130.50	0.22978	0.000531152	0.00282374	0.00298436
131.50	0.219459	0.00049732	0.00268947	0.0028458
132.50	0.208047	0.000494815	0.00254075	0.00269913
133.50	0.199488	0.000482045	0.00242671	0.00258573
134.50	0.190763	0.000459688	0.00231435	0.00246931
135.50	0.183222	0.000433943	0.00221037	0.00236529
136.50	0.175104	0.000431184	0.00210607	0.00226208
137.50	0.166567	0.000461638	0.00200093	0.00215562
138.50	0.160753	0.000405287	0.00192351	0.0020797
139.50	0.154576	0.000408747	0.00184387	0.00199822
140.50	0.148986	0.000330027	0.00177006	0.00192164
141.50	0.143821	0.000300118	0.00170486	0.00185269
142.50	0.1368	0.000319616	0.00162132	0.00176741
143.50	0.13233	0.000287473	0.0015611	0.00170176
144.50	0.128804	0.000292927	0.00151279	0.00165943
145.50	0.123426	0.000288925	0.00144767	0.00159186
146.50	0.118441	0.000290218	0.00138638	0.00152903
147.50	0.115181	0.000264258	0.00134667	0.00148101
148.50	0.11043	0.000276816	0.00128909	0.00141875
149.50	0.107968	0.00025658	0.00125878	0.00138563
150.50	0.103254	0.00022423	0.00120178	0.00132616
151.50	0.0995988	0.000220796	0.00115601	0.0012821
152.50	0.0963518	0.000225732	0.00111543	0.00124038
153.50	0.0930374	0.000216627	0.00107616	0.00119795
154.50	0.090376	0.000207501	0.00104454	0.00116202
155.50	0.0874481	0.000216818	0.0010082	0.00112557
156.50	0.0853507	0.000212262	0.000982915	0.0010952
157.50	0.0826025	0.000201768	0.000949715	0.00106242
158.50	0.0805298	0.000202462	0.000924765	0.00103331
159.50	0.0779493	0.000196159	0.00089621	0.0010005

3992 J Double Muon trigger for 2017

3993 The effect on the total acceptance using the Double Muon trigger in the 2017 has been quan-
 3994 tified. The changes in the total signal acceptance is provided in Fig. J.1. Events are merged
 3995 between the two primary datasets accordingly to the trigger that fired the event. The standard
 3996 SingleMuon selection is preferred. In the DoubleMuon primary dataset, if the singleMuon trig-
 3997 ger paths failed the double muons `HLT_DoubleIsoMu20_eta2p1_v` if they are in $|\eta| < 2.1$.
 3998 No changes in the p_T threshold is applied and the 26/20 threshold are kept in this case. No SF
 3999 are applied to the dimuon path, yet.

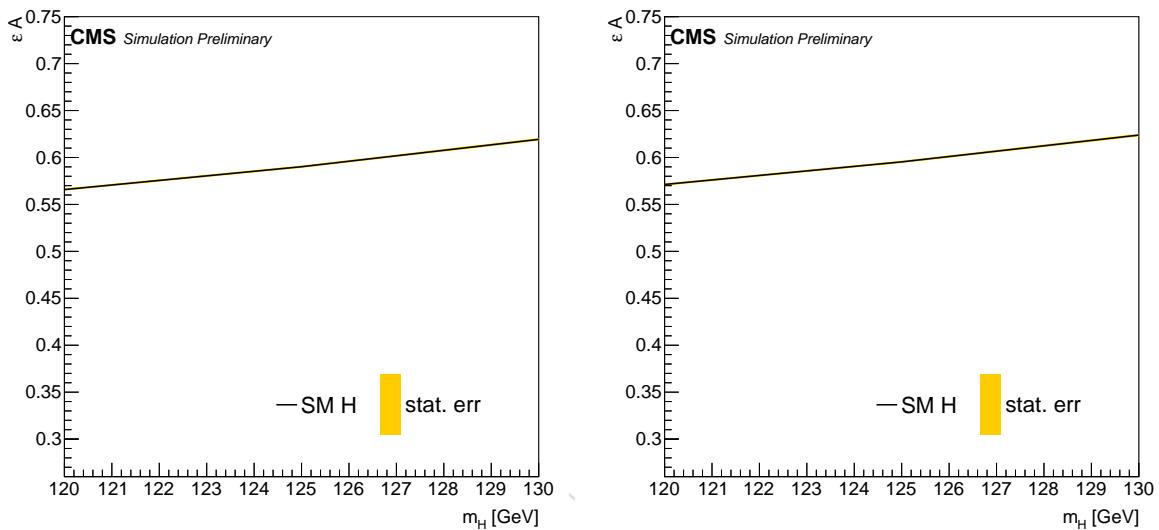


Figure J.1: Summary of the different categories yields and composition for the 2017 (left), 2017 with double Muon trigger (right) data-taking period.

4000 The increasing in signal events is quantified to be $\lesssim 2\%$, stable with respect to the classification.

4001 K HEM effect in 2018

4002 During the 2018 data-taking a sector of the hadronic calorimeter in the forward region was
 4003 turned off. This reflects in mis-calibration of jets and missing transverse energy. To assess the
 4004 magnitude of this effect, in MC, we scale down the jet energy by

- 4005 • 20 % for jets with $-1.57 < \phi < -0.87$ and $-2.5 < \eta < -1.3$
- 4006 • 35 % for jets with $-1.57 < \phi < -0.87$ and $-3.0 < \eta < -2.5$

4007 for all jets with $p_T > 15\text{ GeV}$ passing the tight ID (in order to reject muons/electrons), and
 4008 propagate this to the MET as well.

4009 We recompute the BDT for 01 jets and 2jets and assess the difference with respect to the nominal
 4010 procedure.

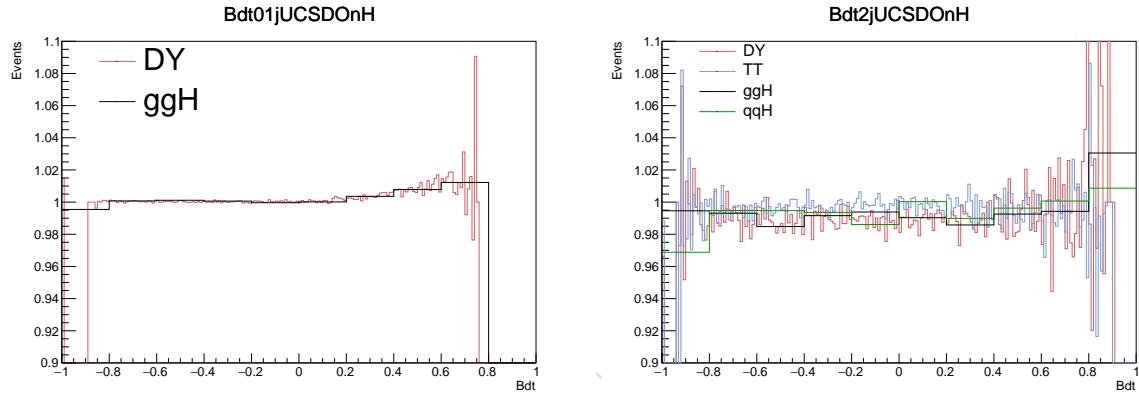


Figure K.1: Ratio as function of the BDT distribution (left, 01 jet, right 2jet) with and without the HEM corrections described above. A maximum of a 2% variation is observed in the high BDT score part, compatible with the different signal and background processes.

4011 A negligible effect is observed in the analysis, with changes in yields $\pm 1\text{--}2\%$ using the full cor-
 4012 rection. The 2018 categories will included a systematic uncertainty ...

4013 L Boosted Decision Tree for VBF MC-templates

4014 A large list of possible input variable has been used to perform N+1 tests. The optimal variable
 4015 choice has been performed in two steps: one N+1 test using 2016 MC samples and one N+1
 4016 test using 2016, 2017 and 2018 MC samples. Table L.1 and L.1 show all MC samples used in
 4017 both the steps.

Table L.1: Samples used for N+1 test in the first step of the variable optimization.

process	samples
DY	DYJetsToLL_M-105To160_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8 (2016)
	DYJetsToLL_M-105To160_VBFFilter_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (2016)
VBF Z	EWK_LLJJ_MLL-50_MJJ-120_13TeV-madgraph-pythia8 (2016)
VBF H	VBF_HToMuMu_M125_13TeV_powheg_pythia8 (2016)

Table L.2: Samples used for N+1 test in the second step of the variable optimization.

process	samples
DY	DYJetsToLL_M-105To160_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8 (2016, 2017, 2018)
	DYJetsToLL_M-105To160_VBFFilter_TuneCUETP8M1_13TeV-madgraphMLM-pythia8 (2016)
	DYJetsToLL_M-105To160_VBFFilter_TuneCP5_13TeV-madgraphMLM-pythia8 (2017)
	DYJetsToLL_M-105To160_VBFFilter_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8 (2018)
VBF Z	EWK_LLJJ_MLL_105-160_SM_5f_LO_TuneEEC5_13TeV-madgraph-herwigpp (2016)
	EWK_LLJJ_MLL_105-160_SM_5f_LO_TuneCH3_13TeV-madgraph-herwig7 (2017, 2018)
VBF H	VBF_HToMuMu_M125_13TeV_powheg_pythia8 (2016)
	VBFHToMuMu_M-125_TuneCP5_PSweights_13TeV_powheg_pythia8 (2017, 2018)

4018 Training has been performed using event of the Signal Region described in section 7.1: 120K
 4019 events have been selected from backgrounds and 130K events from signal in the first step while
 4020 630K events have been selected from backgrounds and 330K events from signal in the first step

4021 The N+1 tests start from an initial set of 2 variables ($m(\mu\mu)$ and $m(jj)$). At each N+1 step ad
 4022 additional variable among a large number of possibilities, is added and the performances of all
 4023 possible N+1 variables sets are evaluated with new independent BDT trainings. The variable
 4024 that provides the greatest separation power increase in terms of expected statistical significance
 4025 is then added to the list of chosen variables and the next N+1 step is performed. The addition of
 4026 discriminating variables is stopped when none of the remaining variables provide a significant
 4027 increase of separation. The first N+1 test ended up with a set of 7 variables that optimize the
 4028 first described BDT. The second N+1 test started from those 7 variables and returned a set of 9
 4029 variables optimizing the new BDT.

4030 The BDT has been trained with adaptive boosting algorithm. The number of Trees and the
 4031 maximum cell tree depth have been tuned to achieve a model that does not overfit the training
 4032 sample and has the highest possible performance. The final chosen options are

4033 Options chosen for the first step are

```
4034 BookMethod( TMVA::Types::kBDT, "BDTG", "!H:!V: NTrees=100:  

4035 MinNodeSize=5%: BoostType=AdaBoost: Shrinkage=.10: UseBaggedBoost:
```

```

4036 BaggedSampleFraction=0.6: nCuts=20: MaxDepth=2:
4037 NegWeightTreatment=IgnoreNegWeightsInTraining")
4038 Options chosen for the second step are
4039 factory->BookMethod( TMVA::Types::kBDT, "BDTG", "!H:!V:NTrees=300:
4040 fMinNodeSize=0.05:BoostType=AdaBoost:Shrinkage=0.1:UseBaggedBoost:
4041 fVarTransform=Decorrelate:NegWeightTreatment=IgnoreNegWeightsInTraining:
4042 fBaggedSampleFraction=0.6:MaxDepth=3:nCuts=20"
4043

```

The BDT output is a real number in range[-1,1], it is generally close to 1 for signal and close to -1 for background. The metric used to evaluate the BDT performance in N+1 tests is the expected sensitivity obtained by the BDT output distribution. In order to improve the sensitivity for the extraction of the signal component the transformation $BDT' = \tanh^{-1}((BDT + 1)/2)$ is applied to BDT output. This allows the purest signal region of the BDT output to be better sampled.

Sensitivity is computed by

$$\text{Total sensitivity} = \sqrt{\sum_i \frac{S_i^2}{B_i}} \quad (\text{L.1})$$

4044 where S_i and B_i are the expected events for signal and background in i^{th} bin.

4045 The contents B_i in the last bins are also strongly dependent from the training and from the binning. In order to settle this problem a minimum number of events is imposed all the bins. On 4046 one hand, this number must be big enough to make the significance relative error limited, on 4047 the other hand it cannot be too high because it will reduce its contribution to the total significance. For these reasons, in order to remove this dependence, every bin is required to satisfy 4048 the following condition:

- 4051 • each bin has to contain at least 0.3 signal events.
- 4052 • relative error of each bin event prediction has to be less than 30%

4053 It means that bin widths are not fixed but they change in order to respect the above conditions. 4054 Therefore, bin limits are computed for all the other bins with the abovementioned criteria. Such 4055 procedure yields a total significance that is almost independent from the binning.

4056 In order to reduce uncertainty in the sensitivity calculation, due again to the low MC statistics, 4057 signal and background MC samples are split randomly into two parts 20 times: 20 training 4058 samples and 20 test samples. Each N+1 test is repeated 20 times with different training samples 4059 and test samples. Each set of N+1 variables has 20 training and 20 sensitivities. The average 4060 sensitivity is computed and used for comparison in N+1 test.

4061 Sensitivities shown in figure L.1 are the averages of sensitivities in 20 different test.

4062 At the end of the N+1 tests, 9 variables have been selected to be the most discriminants and 4063 then used in the BDT:

- 4064 1. $m(\mu\mu)$
- 4065 2. $m(jj)$
- 4066 3. $R(p_T)$
- 4067 4. $\Delta\eta(j,j)$

- 4068 5. z^*
 4069 6. N^{soft}
 4070 7. $\min(|\eta(j_1, \mu\mu)|, |\eta(j_2, \mu\mu)|)$
 4071 8. $\Delta\phi(j, j)$
 4072 9. $p_T(j_2)$

4073 where the the variables from 3 to 7 have been chosen during the first step while have been
 4074 chosen in the second step. Figure L.1 show the second N+1 test.

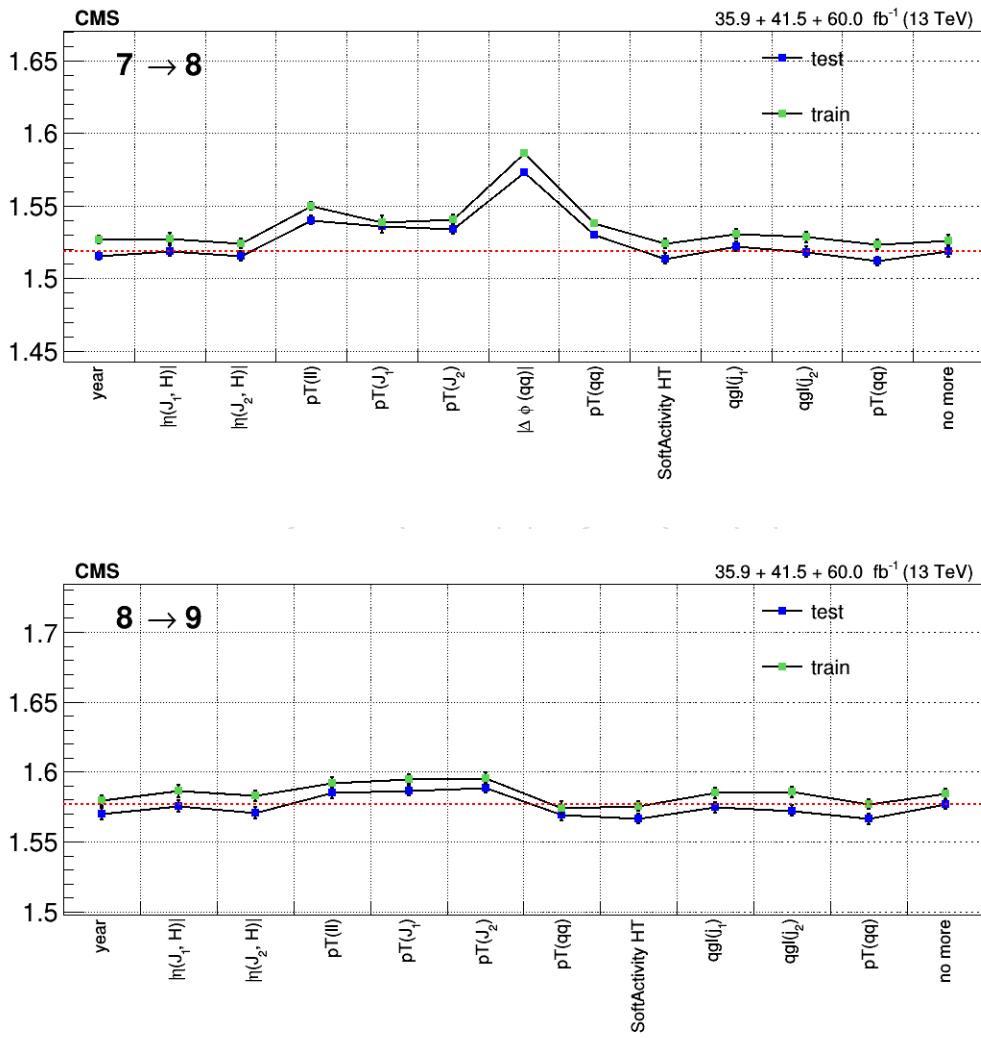


Figure L.1: Averages significances computed by (L.1) after first (up) and second (bottom) iteration of the second N+1 test. Each training is performed by adding one variable every time. Last points correspond to the training with 7 variables (up) and 3 variables (bottom).

4075 The data and Monte Carlo distributions of the BDT output values are shown in Figure L.2 for
 4076 all the region for all the years. Since $m(\mu\mu)$ is one of the main input variable, the BDT output
 4077 distribution peak at zero in control region and it has no event with high score. Therefore BDT

4078 output in Figure L.2 is computed with a fixed value of $m(\mu\mu)=125$ GeV for control region
 4079 events. Drell-Yan distribution in Figure L.2 is computed using NLO samples.

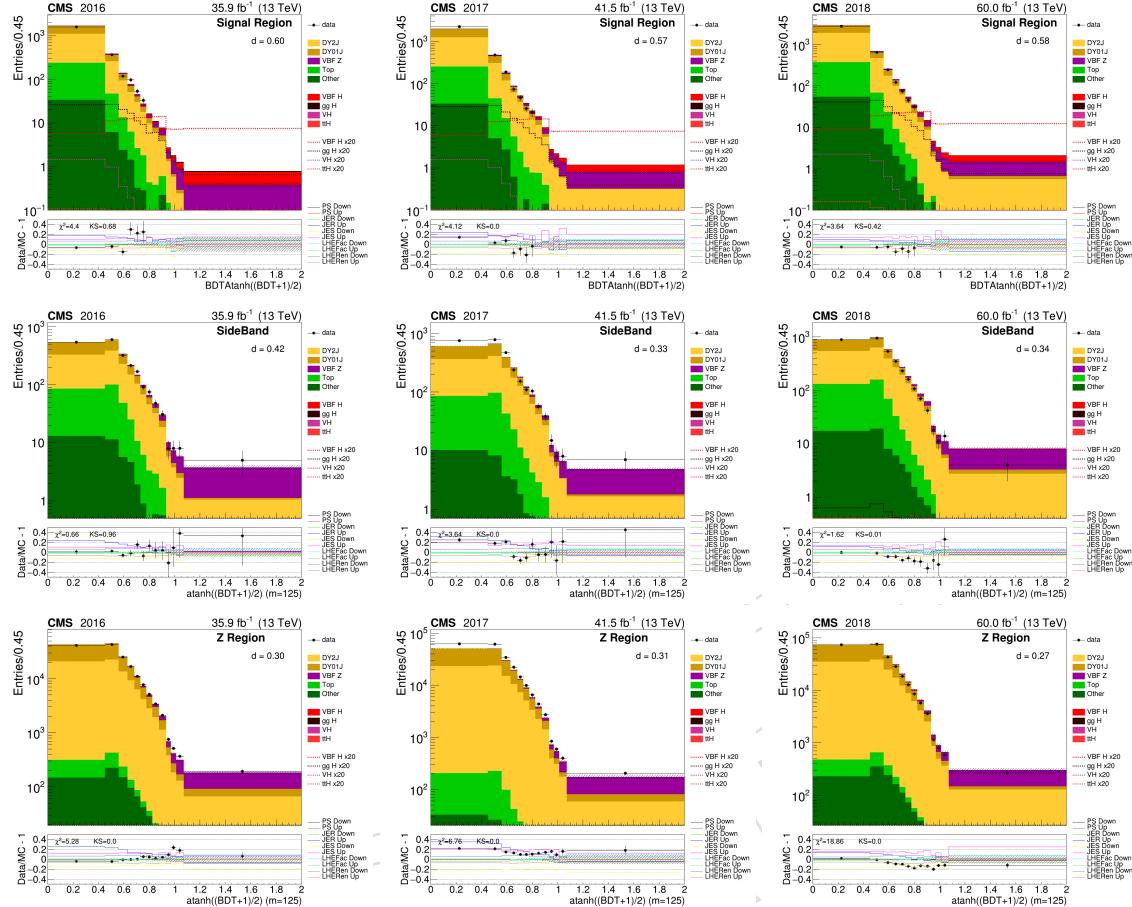


Figure L.2: BDT output distribution after the event selection in Signal Region for 2016 (left), 2017 (center) and 2018 (right). BDT output distributions are computed with $m(\mu\mu)=125$ GeV in control regions.

4080 **References**

- 4081 [1] ATLAS Collaboration, "Observation of a new particle in the search for the Standard
4082 Model Higgs boson with the ATLAS detector at the LHC", *Phys. Lett.* **B716** (2012) 1–29,
4083 doi:[10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020), arXiv:[1207.7214](https://arxiv.org/abs/1207.7214).
- 4084 [2] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS
4085 experiment at the LHC", *Phys. Lett.* **B716** (2012) 30–61,
4086 doi:[10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021), arXiv:[1207.7235](https://arxiv.org/abs/1207.7235).
- 4087 [3] CMS Collaboration, "Observation of the Higgs boson decay to a pair of τ leptons with
4088 the CMS detector", *Phys. Lett.* **B779** (2018) 283–316,
4089 doi:[10.1016/j.physletb.2018.02.004](https://doi.org/10.1016/j.physletb.2018.02.004), arXiv:[1708.00373](https://arxiv.org/abs/1708.00373).
- 4090 [4] CMS Collaboration, "Observation of $t\bar{t}H$ production", *Phys. Rev. Lett.* **120** (2018), no. 23,
4091 231801, doi:[10.1103/PhysRevLett.120.231801](https://doi.org/10.1103/PhysRevLett.120.231801), arXiv:[1804.02610](https://arxiv.org/abs/1804.02610).
- 4092 [5] CMS Collaboration, "Observation of Higgs boson decay to bottom quarks", *Phys. Rev.*
4093 *Lett.* **121** (2018), no. 12, 121801, doi:[10.1103/PhysRevLett.121.121801](https://doi.org/10.1103/PhysRevLett.121.121801),
4094 arXiv:[1808.08242](https://arxiv.org/abs/1808.08242).
- 4095 [6] CMS Collaboration, "Combined measurements of Higgs boson couplings in
4096 proton-proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J.* **C79** (2019), no. 5, 421,
4097 doi:[10.1140/epjc/s10052-019-6909-y](https://doi.org/10.1140/epjc/s10052-019-6909-y), arXiv:[1809.10733](https://arxiv.org/abs/1809.10733).
- 4098 [7] ATLAS Collaboration, "Cross-section measurements of the Higgs boson decaying into a
4099 pair of τ -leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector",
4100 *Phys. Rev.* **D99** (2019) 072001, doi:[10.1103/PhysRevD.99.072001](https://doi.org/10.1103/PhysRevD.99.072001),
4101 arXiv:[1811.08856](https://arxiv.org/abs/1811.08856).
- 4102 [8] ATLAS Collaboration, "Observation of Higgs boson production in association with a
4103 top quark pair at the LHC with the ATLAS detector", *Phys. Lett.* **B784** (2018) 173–191,
4104 doi:[10.1016/j.physletb.2018.07.035](https://doi.org/10.1016/j.physletb.2018.07.035), arXiv:[1806.00425](https://arxiv.org/abs/1806.00425).
- 4105 [9] ATLAS Collaboration, "Observation of $H \rightarrow b\bar{b}$ decays and VH production with the
4106 ATLAS detector", *Phys. Lett.* **B786** (2018) 59–86,
4107 doi:[10.1016/j.physletb.2018.09.013](https://doi.org/10.1016/j.physletb.2018.09.013), arXiv:[1808.08238](https://arxiv.org/abs/1808.08238).
- 4108 [10] ATLAS Collaboration, "Measurement of the Higgs boson mass in the $H \rightarrow ZZ^* \rightarrow 4\ell$
4109 and $H \rightarrow \gamma\gamma$ channels with $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector", *Phys.*
4110 *Lett.* **B784** (2018) 345–366, doi:[10.1016/j.physletb.2018.07.050](https://doi.org/10.1016/j.physletb.2018.07.050),
4111 arXiv:[1806.00242](https://arxiv.org/abs/1806.00242).
- 4112 [11] ATLAS Collaboration, "Measurements of gluon-gluon fusion and vector-boson fusion
4113 Higgs boson production cross-sections in the $H \rightarrow WW^* \rightarrow e\nu\mu\nu$ decay channel in pp
4114 collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector", *Phys. Lett.* **B789** (2019) 508–529,
4115 doi:[10.1016/j.physletb.2018.11.064](https://doi.org/10.1016/j.physletb.2018.11.064), arXiv:[1808.09054](https://arxiv.org/abs/1808.09054).
- 4116 [12] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs
4117 Cross Sections: 4. Deciphering the Nature of the Higgs Sector",
4118 doi:[10.23731/CYRM-2017-002](https://doi.org/10.23731/CYRM-2017-002), arXiv:[1610.07922](https://arxiv.org/abs/1610.07922).
- 4119 [13] CMS Collaboration, "Search for the Higgs boson decaying to two muons in
4120 proton-proton collisions at $\sqrt{s} = 13$ TeV", *Phys. Rev. Lett.* **122** (2019), no. 2, 021801,
4121 doi:[10.1103/PhysRevLett.122.021801](https://doi.org/10.1103/PhysRevLett.122.021801), arXiv:[1807.06325](https://arxiv.org/abs/1807.06325).

- [4122] [14] ATLAS Collaboration, “Search for the dimuon decay of the Higgs boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”, *Phys. Rev. Lett.* **119** (2017), no. 5, 051802, doi:10.1103/PhysRevLett.119.051802, arXiv:1705.04582.
- [4125] [15] ATLAS Collaboration Collaboration, “A search for the dimuon decay of the Standard Model Higgs boson in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector”, Technical Report ATLAS-CONF-2019-028, CERN, Geneva, Jul, 2019.
- [4128] [16] CMS Collaboration, “Observation of the Diphoton Decay of the Higgs Boson and Measurement of Its Properties”, *Eur. Phys. J.* **C74** (2014), no. 10, 3076, doi:10.1140/epjc/s10052-014-3076-z, arXiv:1407.0558.
- [4131] [17] CMS Collaboration Collaboration, “Search for Higgs boson decays into muons in proton-proton collisions at 13 TeV using full Run2 data”, Technical Report AN-19-090, Geneva, 2019.
- [4134] [18] CMS Collaboration Collaboration, “Template-based search for Higgs boson decay to muons targeting the VBF production mode with full Run-2 data”, Technical Report AN-19-205, Geneva, 2019.
- [4137] [19] CMS Collaboration Collaboration, “Search for the standard model Higgs boson produced in vector boson fusion and decaying to muons using the $\sqrt{s} = 13$ TeV data sample”, Technical Report AN-17-234, Geneva, 2019.
- [4140] [20] CMS Collaboration Collaboration, “Search for the Higgs boson decaying to two muons in pp collisions at 13 TeV using the full Run 2 data”, Technical Report AN-19-185, Geneva, 2019.
- [4143] [21] 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, arXiv:1405.0301.
- [4146] [22] P. Nason, “A New method for combining NLO QCD with shower Monte Carlo algorithms”, *JHEP* **11** (2004) 040, doi:10.1088/1126-6708/2004/11/040, arXiv:hep-ph/0409146.
- [4149] [23] 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, arXiv:0710.3820.
- [4152] [24] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J.* **C76** (2016), no. 3, 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.
- [4155] [25] CMS Collaboration, “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements”, arXiv:1903.12179.
- [4157] [26] J. Alwall et al., “Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions”, *Eur. Phys. J.* **C53** (2008) 473–500, doi:10.1140/epjc/s10052-007-0490-5, arXiv:0706.2569.
- [4160] [27] R. Frederix and S. Frixione, “Merging meets matching in MC@NLO”, *JHEP* **12** (2012) 061, doi:10.1007/JHEP12(2012)061, arXiv:1209.6215.

- [28] 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).
- [29] NNPDF Collaboration, "Parton distributions from high-precision collider data", *Eur. Phys. J.* **C77** (2017), no. 10, 663, doi:[10.1140/epjc/s10052-017-5199-5](https://doi.org/10.1140/epjc/s10052-017-5199-5), arXiv:[1706.00428](https://arxiv.org/abs/1706.00428).
- [30] S. Agostinelli et al., "Geant4 : a simulation toolkit", *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **506** (2003), no. 3, 250 – 303, doi:[https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [31] K. Hamilton, P. Nason, E. Re, and G. Zanderighi, "NNLOPS simulation of Higgs boson production", *JHEP* **10** (2013) 222, doi:[10.1007/JHEP10\(2013\)222](https://doi.org/10.1007/JHEP10(2013)222), arXiv:[1309.0017](https://arxiv.org/abs/1309.0017).
- [32] K. Hamilton, P. Nason, and G. Zanderighi, "Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator", *JHEP* **05** (2015) 140, doi:[10.1007/JHEP05\(2015\)140](https://doi.org/10.1007/JHEP05(2015)140), arXiv:[1501.04637](https://arxiv.org/abs/1501.04637).
- [33] J. Bellm et al., "Herwig 7.0/Herwig++ 3.0 release note", *Eur. Phys. J.* **C76** (2016), no. 4, 196, doi:[10.1140/epjc/s10052-016-4018-8](https://doi.org/10.1140/epjc/s10052-016-4018-8), arXiv:[1512.01178](https://arxiv.org/abs/1512.01178).
- [34] B. Jäger et al., "Parton-shower effects in Higgs production via Vector-Boson Fusion", arXiv:[2003.12435](https://arxiv.org/abs/2003.12435).
- [35] CMS Collaboration, "Electroweak production of two jets in association with a Z boson in proton–proton collisions at $\sqrt{s} = 13$ TeV", *Eur. Phys. J.* **C78** (2018), no. 7, 589, doi:[10.1140/epjc/s10052-018-6049-9](https://doi.org/10.1140/epjc/s10052-018-6049-9), arXiv:[1712.09814](https://arxiv.org/abs/1712.09814).
- [36] Y. Li and F. Petriello, "Combining QCD and electroweak corrections to dilepton production in FEWZ", *Phys. Rev.* **D86** (2012) 094034, doi:[10.1103/PhysRevD.86.094034](https://doi.org/10.1103/PhysRevD.86.094034), arXiv:[1208.5967](https://arxiv.org/abs/1208.5967).
- [37] D. Bourilkov, "Photon-induced Background for Dilepton Searches and Measurements in pp Collisions at 13 TeV", arXiv:[1606.00523](https://arxiv.org/abs/1606.00523).
- [38] D. Bourilkov, "Exploring the LHC Landscape with Dileptons", arXiv:[1609.08994](https://arxiv.org/abs/1609.08994).
- [39] 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](https://doi.org/10.1103/PhysRevLett.110.252004), arXiv:[1303.6254](https://arxiv.org/abs/1303.6254).
- [40] T. Gehrmann et al., " W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD", *Phys. Rev. Lett.* **113** (2014), no. 21, 212001, doi:[10.1103/PhysRevLett.113.212001](https://doi.org/10.1103/PhysRevLett.113.212001), arXiv:[1408.5243](https://arxiv.org/abs/1408.5243).
- [41] M. Grazzini, S. Kallweit, D. Rathlev, and M. Wiesemann, " $W^\pm Z$ production at the LHC: fiducial cross sections and distributions in NNLO QCD", *JHEP* **05** (2017) 139, doi:[10.1007/JHEP05\(2017\)139](https://doi.org/10.1007/JHEP05(2017)139), arXiv:[1703.09065](https://arxiv.org/abs/1703.09065).
- [42] M. Grazzini, S. Kallweit, and D. Rathlev, "ZZ production at the LHC: fiducial cross sections and distributions in NNLO QCD", *Phys. Lett. B* **750** (2015) 407–410, doi:[10.1016/j.physletb.2015.09.055](https://doi.org/10.1016/j.physletb.2015.09.055), arXiv:[1507.06257](https://arxiv.org/abs/1507.06257).

- [4202] [43] CMS Collaboration, “Measurements of properties of the Higgs boson decaying into the
4203 four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **11** (2017) 047,
4204 doi:10.1007/JHEP11(2017)047, arXiv:1706.09936.
- [4205] [44] CMS Muon POG, “Baseline muon selections for Run-II – Medium ID”.
4206 https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideMuonIdRun2#Medium_Muon.
- [4208] [45] CMS Muon POG, “Baseline muon selections for Run-II – Particle flow isolation”.
4209 https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideMuonIdRun2#Particle_Flow_isolation.
- [4211] [46] CMS Collaboration, “Particle-flow reconstruction and global event description with the
4212 CMS detector”, *JINST* **12** (2017), no. 10, P10003,
4213 doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.
- [4214] [47] CMS Collaboration, “Search for a new scalar resonance decaying to a pair of Z bosons in
4215 proton-proton collisions at $\sqrt{s} = 13$ TeV”, *JHEP* **06** (2018) 127,
4216 doi:10.1007/JHEP06(2018)127, 10.1007/JHEP03(2019)128,
4217 arXiv:1804.01939. [Erratum: JHEP03,128(2019)].
- [4218] [48] CMS Higgs to muons W.G., “FSR tune in order to minimize $H \rightarrow Z\gamma$ contamination”.
4219 [https://indico.cern.ch/event/867116/contributions/3654090/
4220 subcontributions/294675/attachments/1957523/3252192/h2mu_4221 workingmeetin05122019.pdf](https://indico.cern.ch/event/867116/contributions/3654090/subcontributions/294675/attachments/1957523/3252192/h2mu_workingmeetin05122019.pdf).
- [4222] [49] CMS EGM POG, “Multivariate Electron Identification for Run2”.
4223 <https://twiki.cern.ch/twiki/bin/view/CMS/MultivariateElectronIdentificationRun2>.
- [4225] [50] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm”, *JHEP* **04**
4226 (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- [4227] [51] M. Cacciari and G. P. Salam, “Pileup subtraction using jet areas”, *Phys. Lett. B* **659**
4228 (2008) 119, doi:10.1016/j.physletb.2007.09.077, arXiv:0707.1378.
- [4229] [52] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp
4230 collisions at 8 TeV”, *JINST* **12** (2017) P02014,
4231 doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.
- [4232] [53] CMS JetMET POG, “Recommended Jet Energy Corrections and Uncertainties For Data
4233 and MC”. <https://twiki.cern.ch/twiki/bin/view/CMS/JECDatamc>.
- [4234] [54] Raffaele Gerosa, “PUjetID in VBF $H \rightarrow \mu\mu$ analysis”.
4235 [https://indico.cern.ch/event/860457/contributions/3623773/
4236 attachments/1939388/3215059/Raffaele_05_11_2019.pdf](https://indico.cern.ch/event/860457/contributions/3623773/attachments/1939388/3215059/Raffaele_05_11_2019.pdf).
- [4237] [55] CMS BTV POG, “Usage of b/c Tag Objects for 13 TeV Data in 2016 data and
4238 simulation”. <https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation2016Legacy>.
- [4240] [56] CMS BTV POG, “Heavy flavour tagging for 13 TeV data in 2017 data and simulation”.
4241 <https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation94X>.

- 4243 [57] CMS BTV POG, "Heavy flavour tagging for 13 TeV data in 2018 data and simulation".
4244 https:
4245 //twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation102X.
- 4246 [58] CMS JetMET POG, "Recommended MET filters". https:
4247 //twiki.cern.ch/twiki/bin/view/CMS/MissingETOptionalFiltersRun2.
- 4248 [59] CMS Collaboration Collaboration, "tZq in the trilepton final state (2016 2017)",
4249 Technical Report AN-18-100, Geneva, 2018.
- 4250 [60] CMS Collaboration, "Commissioning of trackjets in pp collisions at $\sqrt{s}=7$ TeV", CMS
4251 Physics Analysis Summary CMS-PAS-JME-10-006, CERN, 2010.
- 4252 [61] CMS Collaboration, "Performance of jet reconstruction with charged tracks only", CMS
4253 Physics Analysis Summary CMS-PAS-JME-08-001, CERN, 2009.
- 4254 [62] CMS Lumi POG, "Utilities for Accessing Pileup Information for Data".
4255 https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData.
- 4256 [63] CMS BTV POG, "Methods to apply b-tagging efficiency scale factors".
4257 https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods.
- 4258 [64] CMS BTV POG, "B-tagging discriminant shape calibration using event weights with a
4259 tag-and-probe method".
4260 https://twiki.cern.ch/twiki/bin/view/CMS/BTagShapeCalibration.
- 4261 [65] CMS Muon POG, "https://twiki.cern.ch/twiki/bin/view/CMS/MuonPOG".
4262 https://twiki.cern.ch/twiki/bin/view/CMS/MuonPOG.
- 4263 [66] CMS Muon POG, "Rochester Correction".
4264 https://twiki.cern.ch/twiki/bin/viewauth/CMS/RochcorMuon.
- 4265 [67] CMS JetMET POG, "Reduced set of jet energy uncertainties for Run2 searches".
4266 https://twiki.cern.ch/twiki/bin/viewauth/CMS/
4267 JECUncertaintySources#Run_2_reduced_set_of_uncertainty.
- 4268 [68] CMS JetMET POG, "L1 ECAL prefiring corrections". https://twiki.cern.ch/
4269 twiki/bin/viewauth/CMS/L1ECALPrefiringWeightRecipe.
- 4270 [69] CMS JetMET POG, "Quark-Gluon Likelihood uncertainty prescription".
4271 https://twiki.cern.ch/twiki/bin/viewauth/CMS/
4272 QuarkGluonLikelihood#Systematics.
- 4273 [70] LHC Higgs cross section working group I, "Uncertainties for ggH process in different
4274 kinematic regions".
4275 https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG_GF_RUN2.
- 4276 [71] LHC Higgs cross section working group II, "Uncertainties for VBF-H process in
4277 different kinematic regions". https:
4278 //gitlab.cern.ch/LHCHIGGSXS/LHCHXSWG2/STXS/VBF-Uncertainties.
- 4279 [72] J. C. Collins and D. E. Soper, "Angular distribution of dileptons in high-energy hadron
4280 collisions", *Phys. Rev. D* **16** (Oct, 1977) 2219–2225,
4281 doi:10.1103/PhysRevD.16.2219.

- [73] J. R. Andersen, K. Arnold, and D. Zeppenfeld, "Azimuthal angle correlations for higgs boson plus multi-jet events", *Journal of High Energy Physics* **2010** (Jun, 2010) 91, doi:10.1007/JHEP06(2010)091.
- [74] D. Rainwater, R. Szalapski, and D. Zeppenfeld, "Probing color-singlet exchange in $z + 2$ -jet events at the cern lhc", *Phys. Rev. D* **54** (Dec, 1996) 6680–6689, doi:10.1103/PhysRevD.54.6680.
- [75] C. S. Committee, "Report on parametric density estimator".
<https://twiki.cern.ch/twiki/bin/view/CMS/DensityEstimation>.
- [76] S. Baker and R. D. Cousins, "Clarification of the Use of Chi Square and Likelihood Functions in Fits to Histograms", *Nucl. Instrum. Meth.* **221** (1984) 437–442, doi:10.1016/0167-5087(84)90016-4.
- [77] CMS Collaboration, "Measurements of Higgs boson properties in the diphoton decay channel in proton-proton collisions at $\sqrt{s} = 13$ TeV", *JHEP* **11** (2018) 185, doi:10.1007/JHEP11(2018)185, arXiv:1804.02716.
- [78] P. D. Dauncey, M. Kenzie, N. Wardle, and G. J. Davies, "Handling uncertainties in background shapes", *JINST* **10** (2015), no. 04, P04015, doi:10.1088/1748-0221/10/04/P04015, arXiv:1408.6865.
- [79] T. Junk, "Confidence level computation for combining searches with small statistics", *Nucl. Instrum. Meth.* **A434** (1999) 435–443, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006.
- [80] A. L. Read, "Presentation of search results: the CL_s technique", *Journal of Physics G: Nuclear and Particle Physics* **28** (sep, 2002) 2693–2704, doi:10.1088/0954-3899/28/10/313.
- [81] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, "Asymptotic formulae for likelihood-based tests of new physics", *Eur. Phys. J.* **C71** (2011) 1554, doi:10.1140/epjc/s10052-011-1554-0, 10.1140/epjc/s10052-013-2501-z, arXiv:1007.1727. [Erratum: Eur. Phys. J.C73,2501(2013)].
- [82] CMS Collaboration, "Measurement of the Hadronic Activity in Events with a Z and Two Jets and Extraction of the Cross Section for the Electroweak Production of a Z with Two Jets in pp Collisions at $\sqrt{s} = 7$ TeV", *JHEP* **10** (2013) 062, doi:10.1007/JHEP10(2013)062, arXiv:1305.7389.
- [83] CMS Collaboration, "Measurement of electroweak production of two jets in association with a Z boson in proton-proton collisions at $\sqrt{s} = 8$ TeV", *Eur. Phys. J.* **C75** (2015), no. 2, 66, doi:10.1140/epjc/s10052-014-3232-5, arXiv:1410.3153.
- [84] CMS Collaboration, "Measurement of electroweak production of a W boson and two forward jets in proton-proton collisions at $\sqrt{s} = 8$ TeV", *JHEP* **11** (2016) 147, doi:10.1007/JHEP11(2016)147, arXiv:1607.06975.
- [85] CMS Collaboration, "Measurement of electroweak production of a W boson in association with two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV", arXiv:1903.04040.
- [86] ALEPH Collaboration, "Studies of quantum chromodynamics with the ALEPH detector", *Phys.Rept.* **294** (1998) 1–165, doi:10.1016/S0370-1573(97)00045-8.

- [87] P. Azzurri et al., "New gluon/quark jet discriminator", CMS Analysis Note 2012/281, CERN, 2012.
- [88] CMS Collaboration, "Performance of quark/gluon discrimination using pp collision data at $\sqrt{s}=8$ TeV", CMS Physics Analysis Summary CMS-PAS-JME-13-002, CERN, 2013.
- [89] CMS JME POG, "Quark-gluon likelihood at 13 TeV". <https://twiki.cern.ch/twiki/bin/viewauth/CMS/QuarkGluonLikelihood>.
- [90] G. Rauco et al., "First look at quark-gluon jet discrimination with 13 tev data", CMS Analysis Note 2016/234, CERN, 2016.
- [91] CMS Collaboration, "Jet algorithms performance in 13 tev data", CMS Physics Analysis Summary CMS-PAS-JME-16-003, CERN, 2017.
- [92] 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.
- [93] F. Chollet et al., "Keras". <https://keras.io>, 2015.
- [94] M. Abadi et al., "TensorFlow: Large-scale machine learning on heterogeneous systems", 2015. Software available from tensorflow.org. <http://tensorflow.org/>.
- [95] L. POG, "Year-to-year combinations and correlations". <https://twiki.cern.ch/twiki/bin/view/CMS/TWikiLUM#LumiComb>.
- [96] CMS Collaboration, "Roostats-based statistic tool for higgs pag", cms twiki page.
- [97] The ATLAS and CMS Collaborations, "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV",
- [98] CMS Collaboration, "A measurement of the Higgs boson mass in the diphoton decay channel", *Phys. Lett. B* **805** (2020) 135425, doi:10.1016/j.physletb.2020.135425, arXiv:2002.06398.
- [99] A. L. Read, "Presentation of search results: the CL_s technique", *J. Phys. G* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313.
- [100] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties", doi:10.5170/CERN-2013-004, arXiv:1307.1347.
- [101] CMS Collaboration, "Search for a standard model-like Higgs boson in the $\mu^+\mu^-$ and e^+e^- decay channels at the LHC", *Phys. Lett. B* **744** (2015) 184–207, doi:10.1016/j.physletb.2015.03.048, arXiv:1410.6679.
- [102] CMS Collaboration, "Measurement of differential fiducial cross sections for Higgs boson production in the diphoton decay channel in pp collisions at $\sqrt{s} = 13$ TeV",
- [103] CMS Collaboration, "Combined Higgs boson production and decay measurements with up to 137 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV", Technical Report CMS-PAS-HIG-19-005, CERN, Geneva, 2020.

- 4363 [104] J. Ellis and T. You, "Updated Global Analysis of Higgs Couplings", *JHEP* **06** (2013) 103,
4364 doi:10.1007/JHEP06(2013)103, arXiv:1303.3879.
- 4365 [105] Particle Data Group Collaboration, "Review of Particle Physics", *Phys. Rev. D* **98** (2018)
4366 030001, doi:10.1103/PhysRevD.98.030001.
- 4367 [106] CMS Collaboration, "Description and performance of track and primary-vertex
4368 reconstruction with the CMS tracker", *JINST* **9** (2014), no. 10, P10009–P10009,
4369 doi:10.1088/1748-0221/9/10/p10009, arXiv:1405.6569.
- 4370 [107] M. Oreglia, "A Study of the Reactions $\psi' \rightarrow \gamma\gamma\psi$ ". PhD thesis, SLAC, 1980.
- 4371 [108] J. D'Hondt et al., "Fitting of Event Topologies with External Kinematic Constraints in
4372 CMS", Technical Report CMS-NOTE-2006-023, CERN, Geneva, Jan, 2006.
- 4373 [109] M. R. Whalley, D. Bourilkov, and R. C. Group, "The Les Houches accord PDFs
4374 (LHAPDF) and LHAGLUE", in *HERA and the LHC: A Workshop on the implications of
4375 HERA for LHC physics. Proceedings, Part B*, pp. 575–581. 2005.
4376 arXiv:hep-ph/0508110.
- 4377 [110] D. Bourilkov, R. C. Group, and M. R. Whalley, "LHAPDF: PDF use from the Tevatron to
4378 the LHC", in *TeV4LHC Workshop - 4th meeting Batavia, Illinois, October 20-22, 2005*. 2006.
4379 arXiv:hep-ph/0605240.
- 4380 [111] D. Bourilkov, "Study of parton density function uncertainties with LHAPDF and
4381 PYTHIA at LHC", in *LHC / LC Study Group Meeting Geneva, Switzerland, May 9, 2003*.
4382 2003. arXiv:hep-ph/0305126.
- 4383 [112] A. Buckley et al., "LHAPDF6: parton density access in the LHC precision era", *Eur.
4384 Phys. J.* **C75** (2015) 132, doi:10.1140/epjc/s10052-015-3318-8,
4385 arXiv:1412.7420.
- 4386 [113] A. Manohar, P. Nason, G. P. Salam, and G. Zanderighi, "How bright is the proton? A
4387 precise determination of the photon parton distribution function", *Phys. Rev. Lett.* **117**
4388 (2016), no. 24, 242002, doi:10.1103/PhysRevLett.117.242002,
4389 arXiv:1607.04266.