

A SEARCH FOR THE STANDARD MODEL HIGGS BOSON DECAYING INTO TWO MUONS
IN THE VECTOR BOSON ASSOCIATED PRODUCTION MODE AT THE CMS EXPERIMENT

By

XUNWU ZUO

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTORATE OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2021

© 2021 Xunwu Zuo

TABLE OF CONTENTS

	<u>page</u>
LIST OF TABLES.....	6
LIST OF FIGURES	10
LIST OF OBJECTS.....	11
CHAPTER	
1 INTRODUCTION	13
1.1 The standard model of particle physics.....	13
1.2 The Higgs coupling to fermions	13
1.3 Potential anomaly coupling of Higgs to muons.....	13
2 THE LHC AND CMS	14
2.1 The Large Hadron Collider	14
2.2 The Compact Muon Solenoid experiment	14
2.3 Trigger system in CMS	14
2.4 Level-1 Trigger	14
3 OVERVIEW OF THE SEARCH OF $H \rightarrow \mu\mu$ DECAY AT CMS.....	15
3.1 Data and simulation samples.....	19
3.1.1 The simulation of the signal processes.....	20
3.1.2 The simulation of the background processes.....	21
3.2 Exclusive analyses and their strategies	22
4 OBJECT RECONSTRUCTION AND IDENTIFICATION	27
4.1 CMS object reconstruction	27
4.2 Object selectoin in the H to muons analysis	27
5 MUON MOMENTUM CORRECTION AND CALIBRATION	28
5.1 Rochester correction.....	29
5.2 FSR recovery	30
5.3 GeoFit correction.....	32
5.3.1 Geometry of the track displacement	33
5.3.2 Development of GeoFit	37
5.3.3 Performance and validation.....	40
5.3.4 GeoFit vs track re-fit	41
5.4 Muon calibration results.....	42
6 SEARCH FOR $H^2\mu\mu$ TARGETING THE VH PRODUCTION MODE	52
6.1 Event selection	55
6.2 MVA discrimination	56
6.2.1 BDT targeting $WH \rightarrow \ell\nu + \mu\mu$ signal.....	58
6.2.2 BDT targeting $ZH \rightarrow \ell\ell + \mu\mu$ signal.....	60
6.2.3 Validation of the BDTs	65

6.3	Event Categorization	70
6.4	Signal and background modeling	71
6.4.1	Signal modeling	72
6.4.2	Background modeling	74
6.5	Systematic uncertainties.....	77
6.5.1	Signal shape uncertainties	78
6.5.2	Signal rate uncertainties	78
6.5.3	Background systematic bias.....	83
6.6	Results of the VH analysis	87
7	RESULTS OF THE H2MU SEARCH	89
7.1	projection of search at 14 TeV	89
	REFERENCES	89

LIST OF TABLES

<u>Tables</u>	<u>page</u>
3-1 Production modes of the Higgs boson in pp collisions at the LHC, their cross sections for $m_H = 125\text{GeV}$, and the corresponding expected number of events in the dataset for this analysis (the cross sections multiplying $\mathcal{B}(H \rightarrow \mu\mu)$ and the integrated luminosity (137fb^{-1}). The leptons (ℓ) in the table refer to electrons or muons.	19
3-2 Summary of the specification for the simulated Higgs signal samples.	21
3-3 Summary of the specification for the simulated background samples.	22
6-1 Selection criteria on muons and electrons in the VH analysis.	55
6-2 Event selections for the WH and ZH categories.	56
6-3 List of input variables used to train the signal-background separation BDT in the WH category. In this table, μ_H is the Higgs candidate, ℓ is the lepton from the W decay, μ_{OS} (μ_{SS}) refers to the muons in the Higgs candidate which OS (SS) to the lepton.	60
6-4 List of input variables used to train the signal-background separation BDT in the ZH category. In this table, μ_H is the Higgs candidate, and $\ell\ell_Z$ is the Z candidate.	63
6-5 Normalization uncertainties on the Higgs boson production cross sections for various modes at $\sqrt{s} = 13\text{TeV}$	79
6-6 Uncertainties on different signal components in the WH and ZH channels related to pileup re-weight and L1 prefiring re-weight. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected.	80
6-7 Uncertainties on different signal components in the WH and ZH channels related to the muon energy scale. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected.	80
6-8 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 modeling. The ID scale factor and Isolation scale factor are covered by the lepMVA scale factors. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected.	81
6-9 Uncertainties on different signal components in the WH and ZH channels related to B-jet vetoing. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected.	81

6-10Uncertainties on different signal components in the WH Cat0 related to jet energy calibration. 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. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected.....	82
6-11Summary of the expect and observed limits and significance in each individual sub-category and the combination.....	88

LIST OF FIGURES

<u>Figures</u>	<u>page</u>
3-1 A conceptual plot for the dimuon mass shapes for the signal and the background. The blue line shows the expected background shape, while the red line shows the expected signal shape on top of it.	16
3-2 Main production modes of the Higgs boson.	17
3-3 Examples of minor Higgs production modes....	18
3-4 Scheme of the procedure of assigning events to different categories. All events passing the common baseline selection are divided into four mutually exclusive categories: ggH, VBF, VH (WH and ZH), and t̄tH (leptonic and hadronic).	24
5-1 Taken from Ref. [43]. Performance of the <i>FSR recovery</i> in the simulated $H \rightarrow \mu\mu$ events. The $m_{\mu\mu}$ before and after the <i>FSR recovery</i> are shown for the events that contain at least one FSR photon (left), and for the inclusive signal events (right). .	31
5-2 Taken from Ref. [43]. Performance of the <i>FSR recovery</i> in the $Z \rightarrow \mu\mu$ events that contain FSR photons, in both data and simulation. A good agreement between simulation and data is observed, before or after the correction.	31
5-3 Scheme of the d0 definition in CMS. The (x_0, y_0) is the coordinate of a point near the vertex in the frame where the vertex is at $(0, 0)$, and the ϕ_0 is the azimuthal angle of the track at (x_0, y_0)	34
5-4 Scheme of the track geometry in the transverse plane. The blue lines show the geometry of the reconstructed track, compared to the black lines which are the geometry of the true track. The difference between the blue track and black track is exaggerated in this scheme. The blue track and the black track must intersect at two points. l is the distance between the two intersections, and x is the distance between the true vertex and the first intersection. s is the distance between the circular centers of the two tracks.	35
5-5 Plots credit to Efe Yigitbasi. Example plots showing the correlation between $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$ and $d0 \cdot \text{charge}$. The vertices used for the d0 calculation are the PV (left) and the BS (right). The PV plot shows a modulated dependence from expectation while the BS plot shows a linear shape as expected. Only barrel tracks from 2016 data are shown as examples. Plots of other $ \eta $ regions and other data-taking periods show similar behaviors.	38
5-6 Plots credit to Efe Yigitbasi. Plots for the $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$ vs d0 correlation in the 2016 DY simulation, and the linear fits to them. Muon tracks are divided into three different $ \eta $ regions: $ \eta < 0.9$ (left), $0.9 < \eta < 1.7$ (middle), and $1.7 < \eta $ (right).	39
5-7 Plots credit to Efe Yigitbasi. Plots for the $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$ vs d0 correlation in the 2017 DY simulation, and the linear fits to them. Muon tracks are divided into three different $ \eta $ regions: $ \eta < 0.9$ (left), $0.9 < \eta < 1.7$ (middle), and $1.7 < \eta $ (right).	39

5-8 Plots credit to Efe Yigitbasi. Plots for the $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$ vs d0 correlation in the 2018 DY simulation, and the linear fits to them. Muon tracks are divided into three different $ \eta $ regions: $ \eta < 0.9$ (left), $0.9 < \eta < 1.7$ (middle), and $1.7 < \eta $ (right).	39
5-9 Plots showing the p_T dependence on the d0 value with different stages of muon correction. The plots compare the $Z \rightarrow \mu\mu$ peak in data and simulation for three years (2016-2018) combined. All positively charged muons are put in the left plot and all negatively charged ones are put in the right plot. The p_T -d0 dependence is reversed for positive and negative muons.	40
5-10Plots showing the <i>GeoFit correction</i> improvement on the four main $H \rightarrow \mu\mu$ signal modes, ggH and VBF plotted on the left, and VH and $t\bar{t}H$ plotted on the right. The plots are made combining the expected signal in all three years of data-taking (2016-2018). The relative improvements on $m_{\mu\mu}$ resolution for ggH, VBF, VH, and $t\bar{t}H$ modes are, respectively, 6.1%, 7.8%, 8.0%, and 9.8%.	41
5-11Plots credit to Pierluigi Bortignon. Plots of the $m_{\mu\mu}$ shape of the 2018 ggH simulation sample, comparing different muon correction methods. The left plot shows the $m_{\mu\mu}$ distribution calculated with muon tracks re-fit with the additional BS constraint, compared with the particle flow shape (left plot). The <i>Rochester correction</i> is not applied in the left plot for both the red and the blue lines. The right plot shows the $m_{\mu\mu}$ distribution from the re-fit method, with the <i>Rochester correction</i> applied, compared with the shape from <i>GeoFit correction + Rochester correction</i> (right plot).	42
5-12Muon calibration plots vs $\eta(\mu^+)$, for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution.	44
5-13Muon calibration plots vs $\phi(\mu^+)$, for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution.	45
5-14Muon calibration plots vs $\phi(\mu^-)$, for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution.	46
5-15Muon calibration plots vs $p_T(\mu^+)$, for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution. The p_T binning sculpts the shape of the $m_{\mu\mu}$ peak, which leads to a jump at the $p_T = 45$ GeV in the plots.....	47
5-16Muon calibration plots vs $p_T(\mu\mu)$, for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution.	48

5-17 Muon calibration plots vs $\eta(\mu\mu)$, for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution.	49
5-18 Muon calibration plots vs $d0(\mu^+)$, for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution.	50
5-19 Muon calibration plots vs $d0(\mu^-)$, for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution.	51
6-1 Plots of the performance of the $WH \rightarrow 3\ell$. On the left, the BDT output score, with signal in blue and background in red. On the right, the receiver operating characteristic (ROC) curve, with training sample in red and testing sample 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. 6-5, the BDT does not sculpt the shape of $m_{\mu\mu}$	61
6-2 Input variables to the $WH \rightarrow 3\ell$ BDT, with signal in blue and background in red. ..	62
6-3 Plots of the performance of the $ZH \rightarrow 4\ell$ BDT. On the left, the BDT output score, 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.....	63
6-4 Input variables to the $ZH \rightarrow 4\ell$ BDT, with signal in blue and background in red.	64
6-5 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).	65
6-6 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).	66
6-7 The WH BDT output (left) and the ZH BDT output (right) in full Run 2 in the signal region $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$	66
6-8 Input variables to the WH BDT in full Run 2 in the signal region $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$	67
6-9 Input variables to the ZH BDT in full Run 2 in the signal region $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$	68
6-10 The input variable $M_T(\mu_{SS}, \text{MHT})$ to the WH BDT (left), and the BDT output for signal and background in different $M_T(\mu_{SS}, \text{MHT})$ bins (right). A mild disagreement is seen between the simulation and data in the low bins of $M_T(\mu_{SS}, \text{MHT})$, while the BDT is not sensitive to the $M_T(\mu_{SS}, \text{MHT})$ values in that region.	69

6-11 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.....	70
6-12 Scans for the BDT cut in the ZH channel. The BDT cut is chosen at -0.1.....	71
6-13 The signal modeling in the $WH \rightarrow \ell + \mu\mu$ 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).	73
6-14 The signal modelin in the $ZH \rightarrow \ell\ell + \mu\mu$ inclusive category. Considered signals modes are qqZH (left) and ggZH (right).	73
6-15 Schemes on how the bias affects the uncertainty on the signal strength measurement. The left plot is an illustration of the $[-\sigma, \sigma]$ coverage of the biased signal strength measurement. The middle plot shows this coverage becomes less as the bias gets larger. The right plot shows how the bias impacts the best fit signal strength uncertainty σ_{fit} . The gray dash lines in the right plot indicates the conventional acceptable range of the bias in this analysis.....	84
6-16 Bias in different BDT-based WH sub-categories. The sub-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.	86
6-17 Bias in different BDT-based ZH sub-categories. The sub-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.....	86
6-18 Plots taken from Ref. [21]. Post-fit $m_{\mu\mu}$ distribution of the WH sub-categories. The sub-categories are: Cat0 BDT [-1.0, -0.1] (left), Cat1 BDT [-0.1, 0.3] (middle), Cat2 BDT [0.3, 1.0] (right). The upper panel in the plots shows the distribution of observed data and the shape of the signal-plus-background fit. The lower panel in the plots shows the residual distribution after subtracting the background component in the fits. The green and yellow bands show the one and two standard deviation of the background component uncertainty.	87
6-19 Plots taken from Ref. [21]. Post-fit $m_{\mu\mu}$ distribution of the ZH sub-categories. The sub-categories are: Cat0 BDT [-1.0, -0.1] (left), Cat1 BDT [-0.1, 1.0] (right). The upper panel in the plots shows the distribution of observed data and the shape of the signal-plus-background fit. The lower panel in the plots shows the residual distribution after subtracting the background component in the fits. The green and yellow bands show the one and two standard deviation of the background component uncertainty.	88

LIST OF OBJECTS

Objects

page

CHAPTER 1 INTRODUCTION

1.1 The standard model of particle physics

1.2 The Higgs coupling to fermions

1.3 Potential anomaly coupling of Higgs to muons

CHAPTER 2
THE LHC AND CMS

2.1 The Large Hadron Collider

2.2 The Compact Muon Solenoid experiment

2.3 Trigger system in CMS

2.4 Level-1 Trigger

CHAPTER 3

OVERVIEW OF THE SEARCH OF $H \rightarrow \mu\mu$ DECAY AT CMS

In 2012, a new boson at 125 GeV was discovered by ATLAS and CMS at the LHC [3, 19, 20]. Various measurements have been performed to probe the properties of this boson ever since, and the boson was later acknowledged as the Higgs boson predicted by the SM. Up to now, the couplings between the Higgs boson and the electroweak (EW) gauge bosons have been observed to be consistent with the SM prediction, while the Yukawa couplings between the Higgs boson and the fermions have only been established for the third generation fermions. The first and second generation fermions are less massive than their third generation counterparts and thus weaker coupling to the Higgs boson, as the coupling strength is proportional to the mass of the fermion. This leads to significantly smaller branching fractions of the decay modes of the Higgs boson to the first or second generation fermions, and poses a substantial challenge to the searches for such decays. The $H \rightarrow \mu\mu$ decay in particular, has a branching ratio of $\mathcal{B}(H \rightarrow \mu\mu) = 2.18 \times 10^{-4}$, which corresponds to an expectation of about 1000 event instances in all the data collected by CMS from 2016 to 2018 (called the LHC Run 2 data-taking period). In contrast, these 1000 so-called signal events are engulfed by millions of events produced through other processes (background events) that mimic their experimental signature. The search for the $H \rightarrow \mu\mu$ decay, in a nutshell, is a struggle to make the signal events stand out from the vast backgrounds with statistical significance.

The search for the $H \rightarrow \mu\mu$ decay has been conducted using proton-proton (pp) collision data collected at center-of-mass energies of 7, 8, and 13 TeV by the CMS Collaboration [37, 45] and the ATLAS Collaboration [2, 1, 4]. The latest result [45] from CMS prior to this work reported an observed (expected in absence of $H \rightarrow \mu\mu$ decay) upper limit of 2.9 (2.2) times the SM prediction of the Higgs boson production and the $\mathcal{B}(H \rightarrow \mu\mu)$, at the 95% confidence level (CL).

Despite the challenging nature of the $H \rightarrow \mu\mu$ analysis, many efforts can be made to improve the separation of signal and background processes, allowing for a sizable refinement of the result. First of all, the Higgs boson has a narrow natural width, and the muons

are well identified objects in the CMS detector. This means that the two muons from the Higgs decay always compose an invariate mass near the nominal Higgs mass, 125 GeV. While for the background processes, the muons are created either by the decay of a Z boson, or by decays of multiple particles, for example a t and \bar{t} quark pair, each resulting in a single muon. The Z boson has a mean mass at 91.2 GeV and a natural width of 2.5 GeV, making a slowly falling tail in the mass spectrum around 125 GeV. For the muons that come from two different sources, their invariant mass follows a flat random distribution near the mass range of interest. Therefore, the signal can be strongly distinguished against the background in the dimuon invariant mass spectrum, as a sharp peak against a smooth falling shape. Figure 3-1 shows the conceptual shape of these mass spectra.

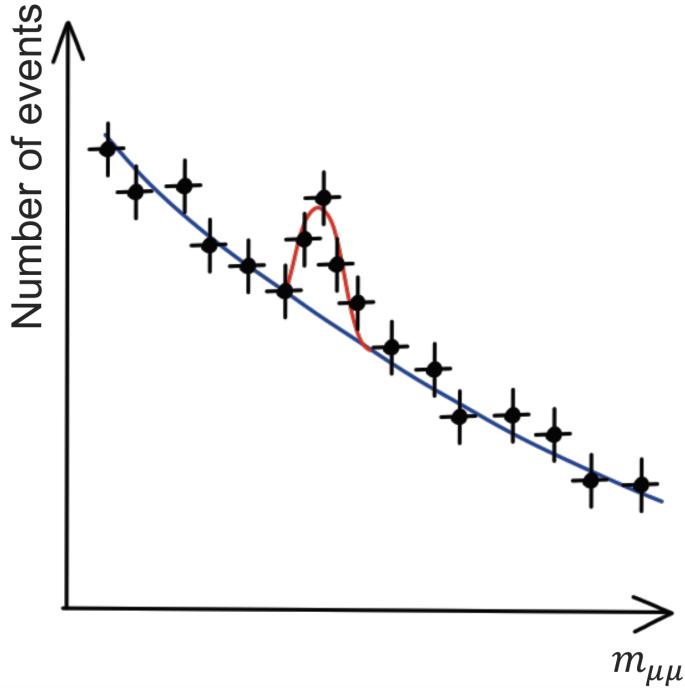


Figure 3-1. A conceptual plot for the dimuon mass shapes for the signal and the background. The blue line shows the expected background shape, while the red line shows the expected signal shape on top of it.

Furthermore, the Higgs boson is produced via several distinct production modes, each with some unique kinematic characteristics. By applying selection criteria targeting a certain signal production mode, it is possible to select a specific part of the kinematic phase

space that is enriched with that signal, and reject many background processes that do not share the same kinematic features. There are four main production modes considered in this analysis, ordered by their production cross sections: gluon fusion (ggH), vector boson fusion (VBF or qqH), associated production with a weak vector boson (VH), and associated production with a pair of top quarks ($t\bar{t}H$). The Feynman diagrams for these main production modes are shown in Figure 3-2. Some other minor production modes are also considered as signal contributions, including associated production with a pair of bottom quarks ($b\bar{b}H$), associated production with a Z boson through gluon fusion ($ggZH$), associated production with a top quark and a W boson (tHW), and associated production with a top quark and a light quark (tHq). The Feynman diagrams for these minor production modes are shown in Figure 3-3. Table 3-1 summarizes the cross sections for all these production modes, along with the expected number of events in the Run 2 dataset (137 fb^{-1}).

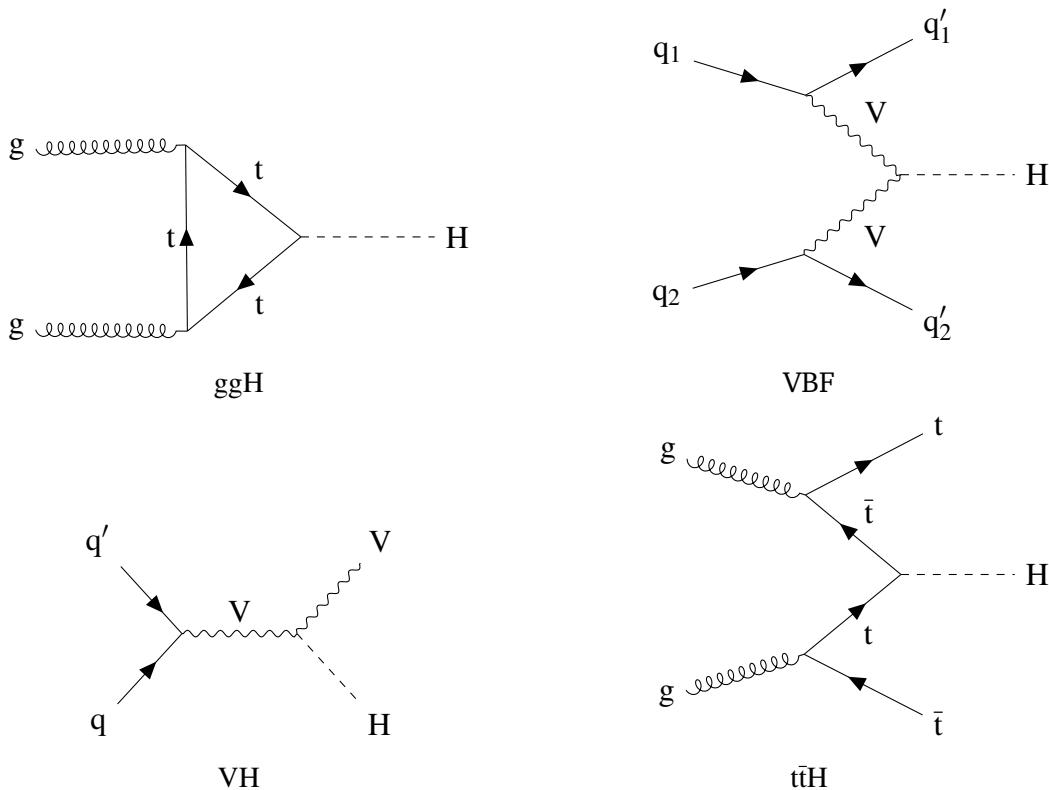


Figure 3-2. Main production modes of the Higgs boson.

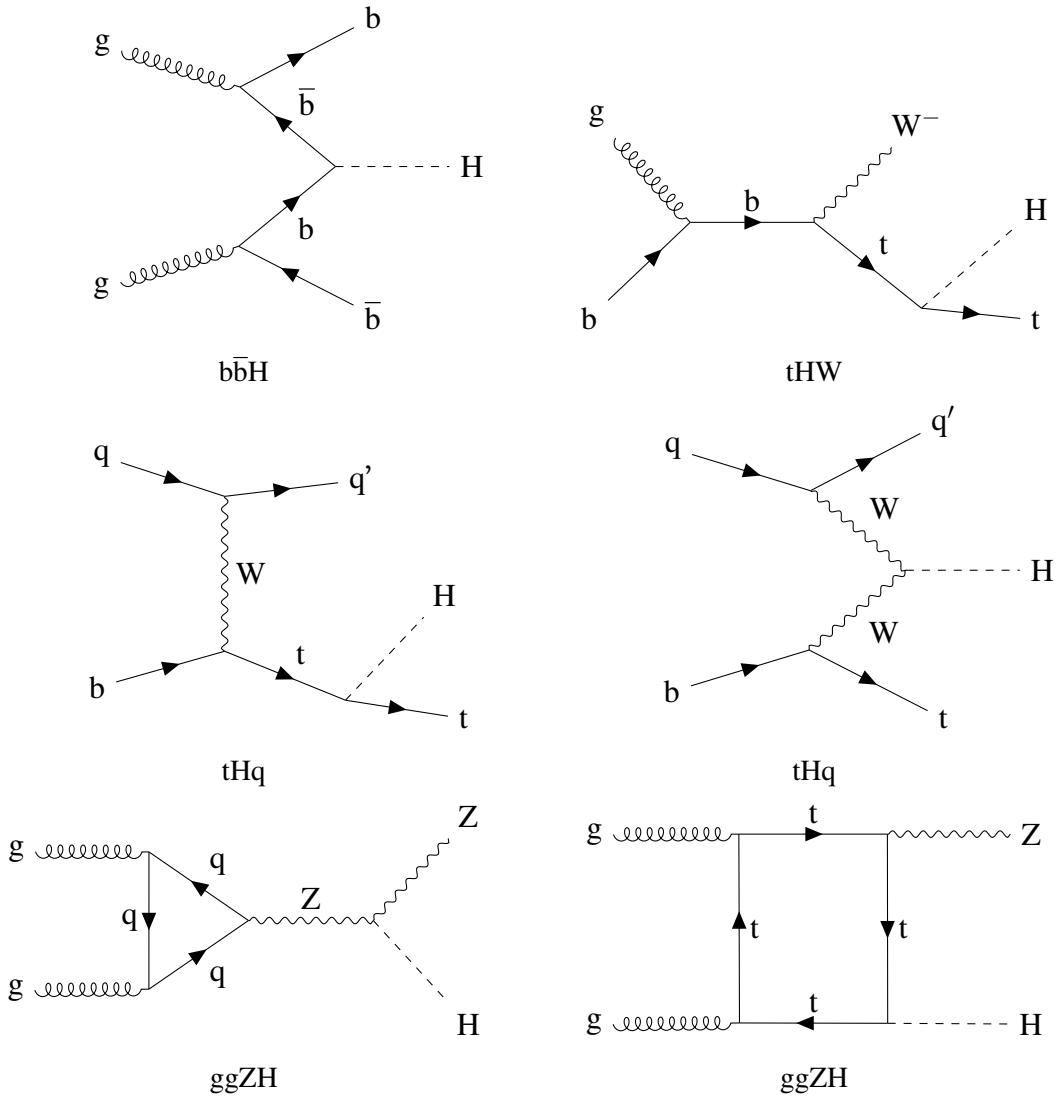


Figure 3-3. Examples of minor Higgs production modes.

Four event categories are defined in this analysis targeting each of the four main Higgs production modes, all of them requiring two oppositely charged (OS) muons that form a candidate for the Higgs boson decay. The additional requirements for each category are: the $t\bar{t}H$ category requires the presence of additional b -jets (from the decay of the top quarks) in the event, the VH category targets events with additional leptons (e or μ , from the decay of the vector boson), the VBF category selects events with two energetic and forward jets, while the ggH category does not specify additional objects, and collects all events that are not selected by the other three categories. No dedicated category is made

Table 3-1. Production modes of the Higgs boson in pp collisions at the LHC, their cross sections for $m_H = 125\text{GeV}$, and the corresponding expected number of events in the dataset for this analysis (the cross sections multiplying $\mathcal{B}(H \rightarrow \mu\mu)$ and the integrated luminosity (137fb^{-1}). The leptons (ℓ) in the table refer to electrons or muons.

signal mode	decay mode	Cross section (pb)	Expected number of events
ggH	inclusive	48.58	1450
VBF	inclusive	3.782	113
WH	inclusive	1.373	41.0
	$W \rightarrow \ell\nu$	0.293	8.75
qq \rightarrow ZH	inclusive	0.761	22.7
	$Z \rightarrow \ell\ell$	0.051	1.53
ggZH	inclusive	0.123	3.67
	$Z \rightarrow \ell\ell$	0.008	0.25
t \bar{t} H	inclusive	0.507	15.1
	$\geq 1 t \rightarrow \text{leptons}$	0.193	5.76
	Both $t \rightarrow \text{hadrons}$	0.230	6.86
sum of above	inclusive	55.13	1646
b \bar{b} H	inclusive	0.488	14.6
tHq	inclusive	0.074	2.21
tHW	inclusive	0.015	0.45
sum of all	inclusive	55.70	1663

for the minor signal modes, either because they have very similar features to one of the main modes, or because their cross sections are much smaller than the main modes and do not make a difference.

With such selections, naturally, each of the selected phase spaces presents a distinct event topology and contains a distinct background composition. Therefore, the analysis is performed independently in each of the event categories, following different optimized strategies. The detailed description of the analysis strategies in each event category is given in Section 3.2.

3.1 Data and simulation samples

This analysis uses the proton-proton collision data collected by the CMS detector during Run 2, which corresponds to a total integrated luminosity of 137.2fb^{-1} .

The triggers used in this analysis are the single muon triggers, which impose some loose isolation requirements and a p_T threshold on the HLT muon candidates. The p_T

cut is 27 (24) GeV for data collected in 2017 (2016, 2018). For the muons selected in this analysis, as explained in Chapter 4, the efficiencies of these triggers are above 95%, making the selection efficiency for the events with two muons close to 100%.

Simulated events from the Monte Carlo (MC) event generators for the signal and the dominant background processes are used to optimize the analysis strategy and assess the systematics uncertainties. The generated events are processed through a detailed simulation of the CMS detector based on GEANT4 [5] and are reconstructed with the same algorithms that are used for data. All MC samples except the EW $Z + jj$ samples use PYTHIA 8.2 [49] to model the parton showering (PS), hadronization, and the underlying event (UE), while the EW $Z + jj$ samples use HERWIG++ and HERWIG7 [13] for the same purpose. The effect of pileup interactions is modelled by overlaying simulated inelastic pp collisions on the hard-scattering event.

3.1.1 The simulation of the signal processes

The ggH signal process is simulated at next-to-leading order (NLO) accuracy in perturbative QCD, using both the MADGRAPH5_amc@NLO v2.4.2 [7] and POWHEG v2.0 [41, 29, 6, 10] MC event generators. The p_T distribution of the Higgs boson in ggH process is then reweighted to match the POWHEG NNLOPS prediction [33, 34]. The VBF, WH, qqZH, and $t\bar{t}H$ processes are simulated with POWHEG v2.0 [42, 40, 35] at NLO precision in QCD. The $b\bar{b}H$ process is simulated at NLO precision in QCD with POWHEG. The tHq, and tHW processes are generated at leading order (LO) with the MADGRAPH5_amc@NLO generator. The ggZH process is simulated at LO with the POWHEG generator. Simulated signal events are generated, for each production mode, at m_H values of 120, 125, 130 GeV. A table summarizing the simulation for signals is shown in Table 3-2.

Expected signal yields are normalized to the production cross sections and $\mathcal{B}(H \rightarrow \mu\mu)$ values taken from the recommendations of LHC Yellow Report [38]. The ggH production cross section is computed at next-to-next-to-NLO (N3LO) precision in QCD, and at NLO in EW theory [9]. The cross section of Higgs boson production in the VBF [16] and

Table 3-2. Summary of the specification for the simulated Higgs signal samples.

Sample	Generator (Perturbative order)	Parton Shower	Cross section	Additional corrections
ggH	MADGRAPH5_amc@NLO (NLO QCD)	PYTHIA	N3LO QCD, NLO EW	$p_T(H)$ from NNLOPS
VBF	POWHEG (NLO QCD)	PYTHIA dipole shower	NNLO QCD, NLO EW	-
qq → VH	POWHEG (NLO QCD)	PYTHIA	NNLO QCD, NLO EW	-
ggZH	POWHEG (LO)	PYTHIA	NNLO QCD, NLO EW	-
tH	POWHEG (NLO QCD)	PYTHIA	NLO QCD, NLO EW	-
bH	POWHEG (NLO QCD)	PYTHIA	NLO QCD	-
tHq	MADGRAPH5_amc@NLO (LO)	PYTHIA	NLO QCD	-
tHW	MADGRAPH5_amc@NLO (LO)	PYTHIA	NLO QCD	-

$\text{qq} \rightarrow \text{VH}$ [15] modes is calculated at next-to-NLO (NNLO) in QCD, including NLO EW corrections, while the $t\bar{t}H$ cross section is computed at NLO in QCD and EW theory [23, 28]. The $b\bar{b}H$, tHq , and tHW cross sections are computed at NLO in QCD without including higher-order EW corrections [38, 25, 24]. The $H \rightarrow \mu\mu$ partial width is computed with HDECAY [26, 50] at NLO in QCD and EW theory.

3.1.2 The simulation of the background processes

The background is modeled considering various SM processes, summarized in Table 3-3. The main background in the ggH and VBF categories is the DY process, which is simulated at NLO in QCD using the MADGRAPH5_amc@NLO generator. The corresponding cross section is calculated with FEWZ v3.1b2 [39] at NNLO in QCD and NLO accuracy in EW theory. The EW production of a Z boson in association with two jets ($Z + jj$) is an important background in the VBF category. This process is simulated at LO using the MADGRAPH5_amc@NLO v2.6.5 generator. The WZ, $q\bar{q} \rightarrow ZZ$, and WW processes, which constitute the main backgrounds in the VH category, are simulated at NLO in QCD using either the POWHEG or MADGRAPH5_amc@NLO generators. Their production cross sections are corrected with the NNLO/NLO K factors taken from Refs. [32], [31], and [30]. The gluon-initiated loop-induced ZZ process (ggZZ) is simulated with the MCFM v7.0 generator [17] at LO and the corresponding production cross section is corrected to match higher-order QCD predictions, following the strategy detailed in Ref. [46]. Minor contributions from triboson processes (WWW, WWZ, WZZ, and ZZZ) are also taken into account and are simulated at NLO in QCD using the MADGRAPH5_amc@NLO generator.

The main backgrounds in the $t\bar{t}H$ category involve the production of top quarks. The $t\bar{t}$ background is simulated with NLO precision in QCD using the `POWHEG` generator, and its cross section is obtained from the `TOP++ v2.0` [22] prediction that includes NNLO corrections in QCD and resummation of next-to-next-to-leading logarithmic (NNLL) soft gluon terms. The single top quark processes are simulated at NLO in QCD via either `POWHEG` or `MADGRAPH5_amc@NLO` and their cross sections are computed, at the same order of precision, using `HATHOR` [36]. Finally, contributions from the $t\bar{t}Z$, $t\bar{t}W$, $t\bar{t}WW$, $t\bar{t}t\bar{t}$, and tZq processes are also considered and are simulated using the `MADGRAPH5_amc@NLO` generator at NLO precision in QCD. For the simulated samples corresponding to the 2016 (2017–2018) data-taking periods, the NNPDF v3.0 (v3.1) NLO (NNLO) parton distribution functions (PDFs) are used [11, 12]. For processes simulated at NLO (LO) in QCD with the `MADGRAPH5_amc@NLO` generator, events from the matrix element (ME) characterized by different parton multiplicities are merged via the FxFx (MLM) prescription [8, 27].

Table 3-3. Summary of the specification for the simulated background samples.

Sample	Generator (Perturbative order)	Parton Shower	Cross section	Additional corrections
Drell-Yan	<code>MADGRAPH5_amc@NLO</code> (NLO QCD)	PYTHIA	NNLO QCD, NLO EW	-
Zjj-EW	<code>MADGRAPH5_amc@NLO</code> (LO)	HERWIG++/HERWIG7	LO	-
$t\bar{t}$	<code>POWHEG</code> (NLO QCD)	PYTHIA	NNLO QCD	-
Single top quark	<code>POWHEG/MADGRAPH5_amc@NLO</code> (NLO QCD)	PYTHIA	NLO QCD	-
Diboson (VV)	<code>POWHEG/MADGRAPH5_amc@NLO</code> (NLO QCD)	PYTHIA	NLO QCD	NNLO/NLO K factors
ggZZ	<code>MCFM</code> (LO)	PYTHIA	LO	NNLO/LO K factors
$t\bar{t}V$, $t\bar{t}VV$	<code>MADGRAPH5_amc@NLO</code> (NLO QCD)	PYTHIA	NLO QCD	-
Triboson (VVV)	<code>MADGRAPH5_amc@NLO</code> (LO)	PYTHIA	NLO QCD	-

3.2 Exclusive analyses and their strategies

In order to maximally harness the kinematic features in the different production modes of the Higgs boson, the analysis is conducted in four independent event categories: the ggH, VBF, VH and $t\bar{t}H$ categories. The workflow to divide events into different categories is shown in Figure 3-4. As a common prerequisite in this analysis, all events should contain two opposite-charged (or opposite-sign, OS) muons that make the candidate for the Higgs boson decay. Then, as a first step, events containing b-tagged jets (either one medium tag or two loose tag of the DeepCSV [44] working points) are classified into the $t\bar{t}H$ category. The $t\bar{t}H$ category is further divided into the $t\bar{t}H$ leptonic category and the

$t\bar{t}H$ hadronic category depending on whether the event contains electrons or additional muons (leptonic category), or whether it contains at least three jets (hadronic category). Some events containing b-tagged jets may not pass the secondary selections for either the leptonic or hadronic $t\bar{t}H$ categories. They are most likely background events, and are therefore discarded. The events without b-tagged jets, fall into the VH category if they contain additional leptons (electrons or muons). Inside the VH category, events are further tagged as WH events if there is one and only one extra lepton in the event, or tagged as ZH events if there are two same-flavor opposite-sign (SFOS) extra leptons. A small collection of events may pass the primary VH selection but fail the secondary WH or ZH selections, for example events that contain one extra electron and one extra muon. These events are most likely not from the signal processes, and are discarded. For the events with neither b-tagged jets nor additional leptons, if there are at least two energetic jets, composing a jet pair with $m_{jj} > 400\text{GeV}$ and $\Delta\eta_{jj} > 2.5$, the events are tagged as the VBF events. Finally, the ggH category collects all events that are not assigned to other categories. Most events in the ggH category are profiled to have either no additional objects or just one additional jet other than the Higgs candidate muons. The detailed definition of the different objects used in this categorization is given in Chapter 4.

The analysis is performed independently in each event category. Since these categories have distinct profiles in the expected signal yield, the signal purity, and the background composition, the optimal approaches to perform the analyses are also different. As a result, two different strategies to perform the analysis are considered:

- **Data-driven parametric fit to the $m_{\mu\mu}$ spectrum:** As is done in the previously published analyses on the the data collected prior to 2017 [37, 45], a multivariate analysis (MVA) method is used to profile the separation between the signal and the background processes. The MVA can be either cut-based as in the Run 1 analysis [37], or machine learning (ML) based as in the analysis on the 2016 data [45]. The MVA considers the kinematic information that is uncorrelated with the $m_{\mu\mu}$, and is used to divide the events into several regions with different signal-to-background-ratios (S/B), called the MVA-categories. In each MVA-category, the signal strength is evaluated from fits to the $m_{\mu\mu}$ spectrum in data, in what is called the *signal fit region*, for example $110 < m_{\mu\mu} < 150\text{GeV}$. Both the signal and the background are

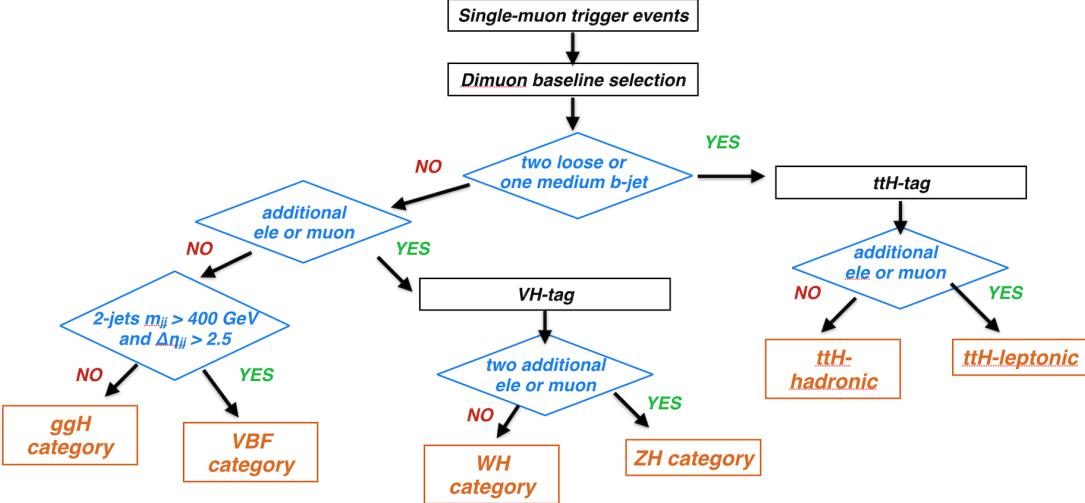


Figure 3-4. Scheme of the procedure of assigning events to different categories. All events passing the common baseline selection are divided into four mutually exclusive categories: ggH, VBF, VH (WH and ZH), and ttH (leptonic and hadronic).

modeled by parametric functions that are carefully studied to provide a truthful description of the distributions of physics processes. The total yield of the background is unconstrained in the fit and is determined entirely by the data. The effects of the systematic uncertainties from various sources on either the signal yield or the signal shape are assessed and propagated to the fit result. The systematic uncertainties do not affect the background estimation since it is based on data rather than predictions from simulations.

- **MC-based template fit to the Neural Network discriminator:** This approach is also based on an MVA, for which a ML algorithm, Deep Neural Network (DNN), is taken. The DNN takes all the kinematic variables *including* $m_{\mu\mu}$, and profiles the discrimination between the signal and the background. Without making further categories, the binned template of the DNN output in the whole phase space is used for the signal strength evaluation. Since the fit is applied to the DNN output rather than the $m_{\mu\mu}$ distribution, the *signal fit region* is further divided into two parts: the *signal region*, $115 < m_{\mu\mu} < 135$ GeV, and the *sideband region*, $110 < m_{\mu\mu} < 115$ GeV or $135 < m_{\mu\mu} < 150$ GeV. The data are fit simultaneously in both regions using the DNN templates of the signal and background simulation. The systematic uncertainties affect both the signal and the background prediction, and are employed as variations in either the yield or the shape of the templates. The background yield is estimated from simulation and is allowed to vary within its uncertainty in the fit, in the same manner as the other systematic uncertainties. The signal strength is extracted from the fit in the *signal region*. The *sideband region* does not contain any signal contribution, but is nonetheless used in the fit, to enhance the constraint on the background estimation.

These two strategies should give comparable results in the ideal case, where there are abundant statistics in both data and simulation, and where the data are well described by simulation. However these conditions are usually not met in real analyses, and one strategy may become preferable over the other. The pp collision is a very noisy environment, making it difficult to achieve an accurate modeling of many kinematic aspects, for example the pile-up events, the parton shower, and the production of leptons through bottom or charm quarks (nonprompt leptons). The modeling of these features usually involves extensive work in the validation of simulated samples, and is generally associated with large systematic uncertainties. In the scenarios where the MC does not model the data very well, or where the uncertainties from MC modeling are not much smaller than the statistical uncertainty in the data, it is more advantageous to follow the data-driven approach. On the other hand, if a phase space lacks enough statistics in data but can be well described by the MC, it is more beneficial to perform a MC-based analysis there.

The ggH category contains the majority of the events in the $H \rightarrow \mu\mu$ analysis, and has very low S/B . In all its MVA-based sub-categories, there are abundant data that the statistical uncertainty of data is smaller than the systematic uncertainties of the background prediction from simulation. Therefore the data-driven strategy is taken in the analysis in the ggH category. The VBF category is featured with a good amount of events, although much less than the ggH category, and a good S/B . This makes it possible to enhance the sensitivity of the analysis by picking very high S/B regions with the help of MVA discriminators. Naturally, the number of events in the high S/B regions is very low. Therefore the MC-based strategy is used in the VBF category. The VH and the $t\bar{t}H$ categories both have very few events, but high S/B , which seems like a good playground for the MC-based approach. However, one of the main backgrounds in the VH and $t\bar{t}H$ categories involves extra lepton(s) from nonprompt sources, and lacks accurate MC modeling. Moreover, the expected signal yields in these categories are low. It is impractical to make DNN templates with many bins, otherwise only a small fraction of one data event is expected in each bin.

Given the dataset used in this analysis, the data-driven method is the preferred choice in both the VH and the $t\bar{t}H$ categories. Overall, the ggH , VH, and $t\bar{t}H$ categories follow the data-driven strategy, while the VBF category takes the MC-based approach.

More details of the analysis strategy can be found in the paper describing this analysis [21], recently submitted for publication by CMS. The following chapters will cover the object definition and the muon corrections that are common to the whole analysis, then the detailed steps of the analysis in the VH category, and finally the results of both the VH category and of the combination of all four categories.

CHAPTER 4
OBJECT RECONSTRUCTION AND IDENTIFICATION

4.1 CMS object reconstruction

4.2 Object selectoin in the H to muons analysis

CHAPTER 5

MUON MOMENTUM CORRECTION AND CALIBRATION

This analysis aims to find a sharp signal peak on top of a smooth background in the $m_{\mu\mu}$ distribution. It is of crucial importance to correct the mismeasurement in muon momentum scale and improve the momentum resolution. The sharper muon peak, the better sensitivity of this analysis. It is also crucial to reduce the differences in the muon momentum scale and resolution between data and simulation, so that there is no significant bias in the modeling of the signal.

Three sets of corrections are applied in this analysis: the *Rochester correction* [14], the recovery of the final state radiation (FSR) photons, and the *GeoFit correction*. The *Rochester correction* is a centrally provided correction (by CMS) which corrects the biases in the muon momentum resulted from the mismodeling of detector alignment and magnetic field. A brief description of the *Rochester correction* is given in Section 5.1, while the technical details can be found in the Ref. [14]. The *FSR recovery* is a common practice in many CMS analyses which corrects the muon energy loss via FSR radiation. The recovery scheme in this analysis is optimized specifically for the $H \rightarrow \mu\mu$ decay, which is described in Section 5.2 and in more details in Ref. [43]. The *GeoFit correction* is developed by the author in the context of the $H \rightarrow \mu\mu$ analysis and approved by the CMS collaboration. It uses information of muon vertexing to correct the biases in muon momentum of the reconstructed muon tracks. The development of the *GeoFit correction* is described in details in Section 5.3. The effects these three corrections focus on are orthogonal, and are even mutually exclusive between the *FSR recovery* and *GeoFit correction*. In practice, the *Rochester correction* is applied to all muons, then each muon is surveyed for FSR photons. If a FSR photon is found associated to the muon, the *FSR recovery* is applied, if not, the *GeoFit correction* is applied.

The *Rochester correction*, *FSR recovery*, and *GeoFit correction* are applied to both data and simulation. The performance of the muon corrections is examined with the study on the $Z \rightarrow \mu\mu$ peak, which is listed in details in Section 5.4. These corrections fix all the known biases in muon measurement, and ensure a per-mille-level agreement between

data and simulation.

5.1 Rochester correction

In reality, the CMS detector can have various imperfections, such as the misalignment of the detector components, and the uncertainties in the magnetic field. Sometimes these imperfections are not correctly emulated in reconstruction software, and as a result, the reconstructed muons can be inaccurate. These measurement biases are reflected as the dependences of the muon momentum on its η , ϕ coordinates, and its charge. These dependences, in turn, smear the inclusive muon resolution and lead to sub-optimal physics results.

On the other hand, in the simulation of CMS events, none of the imperfections is assumed, which leads to slightly different detector responses from those in data, and eventually mismodelings (usually over-optimistic modelings) in muon measurement. The muon correction also needs to be applied to the simulation, in a manner not exactly the same as that to data. This also means that the reconstructed muons in simulation cannot be used as the reference for the correction. Instead, the correction is derived from the generated muon information, smeared with some functional forms to match the experimental resolution. This provides a set of reference muons that is free from any bias in reconstruction.

The well-understood $Z \rightarrow \mu\mu$ events are used to develop the *Rochester correction*. The idea of the correction is briefly summarized as follows:

- For data, reconstructed simulation (reco-sim), and the reference simulation (ref-sim), muons are divided into different η and ϕ bins, separately for μ^+ and μ^- . In each bin, the $1/p_T$ distributions of data and reco-sim are corrected so that the mean value of the distribution becomes the same as that in the ref-sim.
- The $1/p_T$ in reco-sim is usually narrower than that in data. A smearing is applied to the reco-sim $1/p_T$ distribution so that it matches the resolution in data.
- After the steps above, the $m_{\mu\mu}$ in each bin may still be off from the expected distribution by some small amounts. The ratio between this offset and the nominal Z mass is applied to the muon p_T as a correction factor iteratively, until the offset is minimized.

The *Rochester correction* removes the $m_{\mu\mu}$ dependences on muon η , ϕ , and charge, as well as the $m_{\mu\mu}$ resolution differences between data and the simulation. Details of the performance of the *Rochester correction* can be found in Section 5.4.

5.2 FSR recovery

In CMS, muons produced in pp collisions may radiate photons and lose energy, which is referred to as the final state radiation (FSR). The radiation may carry substantial energy and lead to an underestimation of the muon momentum. In an analysis that relies on the dimuon mass $m_{\mu\mu}$, the FSR may lead to two effects that degrade the sensitivity: a loss of event acceptance, and a smearing of the $m_{\mu\mu}$ resolution. To mitigate these effects, some of the FSR photons can be identified and added back to the muon energy, which is called the *FSR recovery*.

The selection for the FSR photons is modified on top of the strategy developed in the CMS H → ZZ analyses [46, 47]. The selection criteria is summarized as follows:

- Photons with transverse energy $E_T^\gamma > 2$ GeV and $|\eta| < 1.4$, $1.6 < |\eta| < 2.4$ are considered as FSR candidates.
- The photon is required to be within the cone of $\Delta R < 0.5$ around its closest muon which satisfies $p_T > 20$ GeV and $|\eta| < 2.4$.
- The photon is not identified as a bremsstrahlung photon associated with a reconstructed electron.
- The PF isolation of the photon in a cone of $\Delta R < 0.3$ should be less than 1.8, i.e. $\sum_i p_T^i(\Delta R(\gamma, i) < 0.3)/p_T(\gamma) < 1.8$, where i iterates the PF objects around the photon other than the candidate muon.
- The separation between the photon and the muon satisfies $\Delta R(\mu, \gamma)/p_T^2(\gamma) < 0.012$.
- In order to suppress energetic photons from the $H \rightarrow Z\gamma \rightarrow \mu\mu\gamma$ process, the p_T ratio between the photon and the muon is required to be less than 0.4, i.e. $p_T(\gamma)/p_T(\mu) < 0.4$.
- If multiple FSR photons are associated to a same muon, only the photon with the smallest $\Delta R(\mu, \gamma)/p_T^2(\gamma)$ is taken.

With this set of selection, about 3% of signal events are tagged with FSR photons. The momentum of the FSR photons are added to the muon momentum, while the photons

themselves are removed from the calculation of the muon isolation. The *FSR recovery* significantly improves the $m_{\mu\mu}$ reconstruction in the FSR tagged events, as shown in the left plot of Figure 5-1. And overall, as shown in the right plot of Figure 5-1, the effect on the inclusive signal is a 3% improvement on the $m_{\mu\mu}$ resolution, and a 1.7% increase in the total signal yield. The performance of *FSR recovery* is also validated with the $Z \rightarrow \mu\mu$ events, shown in Figure 5-2, where a good agreement is kept between simulation and data with or without the *FSR recovery*. The *FSR recovery* is expected to perform the same way on data as on simulation, and no bias is introduced by the application of the *FSR recovery*.

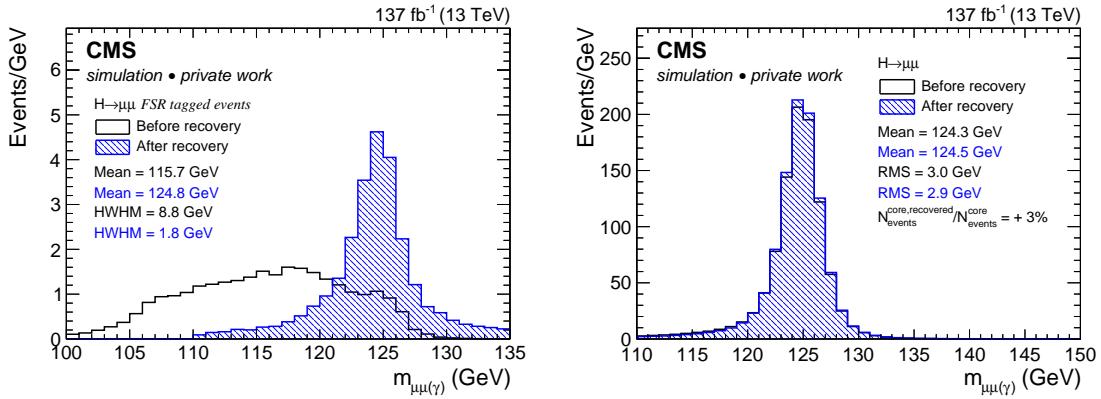


Figure 5-1. Taken from Ref. [43]. Performance of the *FSR recovery* in the simulated $H \rightarrow \mu\mu$ events. The $m_{\mu\mu}$ before and after the *FSR recovery* are shown for the events that contain at least one FSR photon (left), and for the inclusive signal events (right).

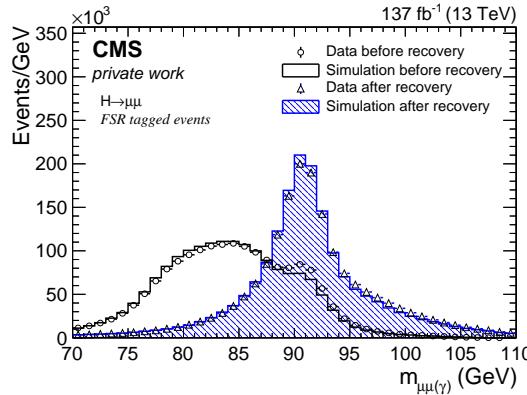


Figure 5-2. Taken from Ref. [43]. Performance of the *FSR recovery* in the $Z \rightarrow \mu\mu$ events that contain FSR photons, in both data and simulation. A good agreement between simulation and data is observed, before or after the correction.

5.3 GeoFit correction

In CMS, charged tracks are reconstructed from the hits in the tracker system. The tracks may originate from primary vertices (PV), which are vertices of the pp interactions, or secondary vertices, which are the decay vertices of particles with a non-infinitesimal lifetime. The track reconstruction does not assume any vertex information since the vertices are reconstructed as the intersections of groups of tracks. This practice is essential for the studies of b quarks, τ leptons, and other long-lived particles that produce a vertex displaced from the colliding region in the transverse plane at the scale of a millemeter or above. However, for the tracks from prompt interactions, such as the decay of a Z boson, W boson, or H boson, called the prompt tracks, even if they are expected not to have a visible displacement from the PV, a non-zero displacement may still appear in reconstruction, due to the uncertainties in the track fit. In other words, if a track is known to be prompt, the posterior information on the colliding position can be used to improve the quality of the track fit, and in turn the measurement on the track momentum.

The study shown in this section reports the finding that this false displacement in prompt tracks has a strong geometrical correlation with the mis-measurement of the track momentum. A simple analytic function can be derived from the geometry and can be verified by fitting the displacement vs the p_T bias in the simulated samples. This analytic form is applied as a correction to p_T and is therefore named the *GeoFit correction*. Section 5.3.1 explains the geometry of the track displacement and the correlation between different variables. Section 5.3.2 describes the studies on simulated samples in order to find the best fit parameters in that correlation. The *GeoFit correction* is developed using only muon tracks in the context of the $H \rightarrow \mu\mu$ search. It removes the dependence of $m_{\mu\mu}$ on track displacement which leads to an improvement on the $m_{\mu\mu}$ resolution of the combined signal ranging from 3% to 10%, depending on the data-taking period. Details of the *GeoFit correction* performance, along with validation studies are shown in Section 5.3.3. In addition, an alternative way to correct this p_T bias is to redo the track fit including the

colliding vertex as an additional hit in the track, which should achieve a more fundamental correction at the cost of more computational resources. A preliminary set of study comparing the *GeoFit* correction with the track re-fit shows the two correction methods give almost equivalent results, detailed in Section 5.3.4.

5.3.1 Geometry of the track displacement

The displacement of a track from a vertex is usually measured as the impact parameters, d_{xy} and d_z , which are the signed distance between the vertex and its point of closest approach (PCA) on the track, in the transverse and longitudinal directions. In CMS, because most studies only care about the transverse impact parameter, the PCA is defined as the point on the 2D-projection of the track in the transverse plane that is the closest to the vertex. (It is not necessarily the PCA in the 3D-space. The d_z is calculated at the 3D-point corresponding to the 2D-PCA, rather than the 3D-PCA.) As the d_z is not used in our studies, the term "impace parameter", if not otherwise stated, refers specifically to the transverse impact parameter d_{xy} , also denoted as d_0 . The definition of d_0 can be expressed as

$$d_0 = -x_0 \cdot \sin(\phi_0) + y_0 \cdot \cos(\phi_0) \quad (5-1)$$

where (x_0, y_0) is the coordinate of a point near the vertex in the frame in which the vertex is at $(0, 0)$, and ϕ_0 is the azimuthal angle of the track at (x_0, y_0) . A scheme for this definition is shown in Figure 5-3.

In the track geometry, illustrated in Figure 5-4, the reconstructed track is very close to, but slightly deviated from, the true track, which leads to a small d_0 between the reco track and the true vertex, as well as a small distance between the circular centers of the reco track and the true track. The circular centers of the two tracks are labeled as O and O' for the true track and the reco track, and s is the distance between O and O' . The radii of the two tracks are r and r' , with $\Delta r = r' - r$. The two tracks must intersect at two points, labeled as point M and N , with the distance between M and N denoted as l . β is half of the central angle spanned by the chord l in the true track, while α is the angle $\angle O'MO$. The

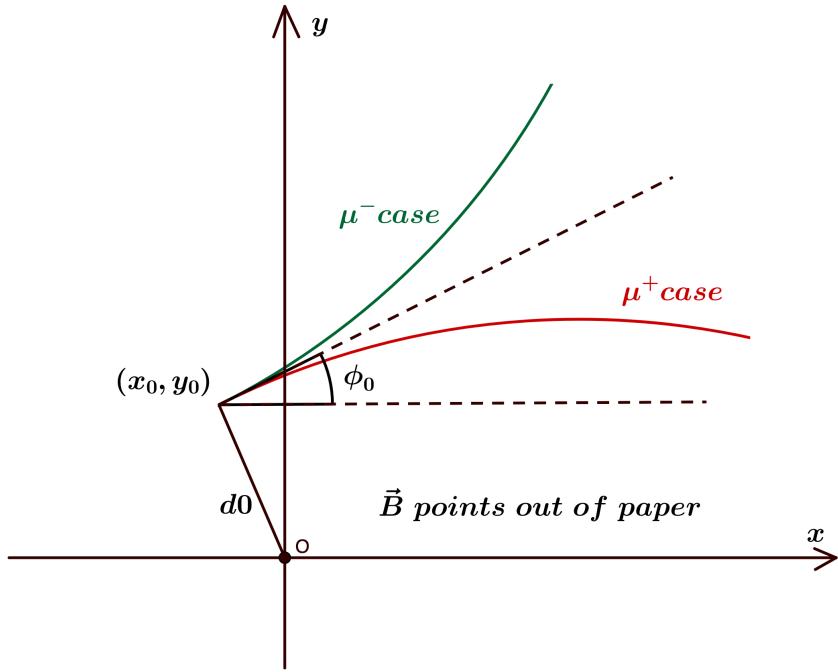


Figure 5-3. Scheme of the d_0 definition in CMS. The (x_0, y_0) is the coordinate of a point near the vertex in the frame where the vertex is at $(0, 0)$, and the ϕ_0 is the azimuthal angle of the track at (x_0, y_0) .

true vertex is denoted as V , with d_0 as the impact parameter of the reco track to it, while the PCA on the reco track is denoted as P . The distance between M and V is marked as x . As this scheme represents typical muon tracks in CMS, the radii of the tracks under study are at the scale of several tens of meters, and the Δr is expected to be much smaller than r . Points M and N are expected to be around the coverage of the CMS tracker system, which is about a meter. Therefore x and l are expected to be much smaller than r as well. Finally, the d_0 scale of the tracks under study is about ten microns, which is much smaller than x , l , and r .

In this setup, a few geometric relationships can be found between different variables, listed as follows: Since $x \ll r$, arc \widehat{VM} and \widehat{PM} can be viewed as line segments which are respectively perpendicular to OM and $O'M$. Therefore in triangle $\triangle VMP$,

$$d_0 = x \cdot \sin\alpha \quad (5-2)$$

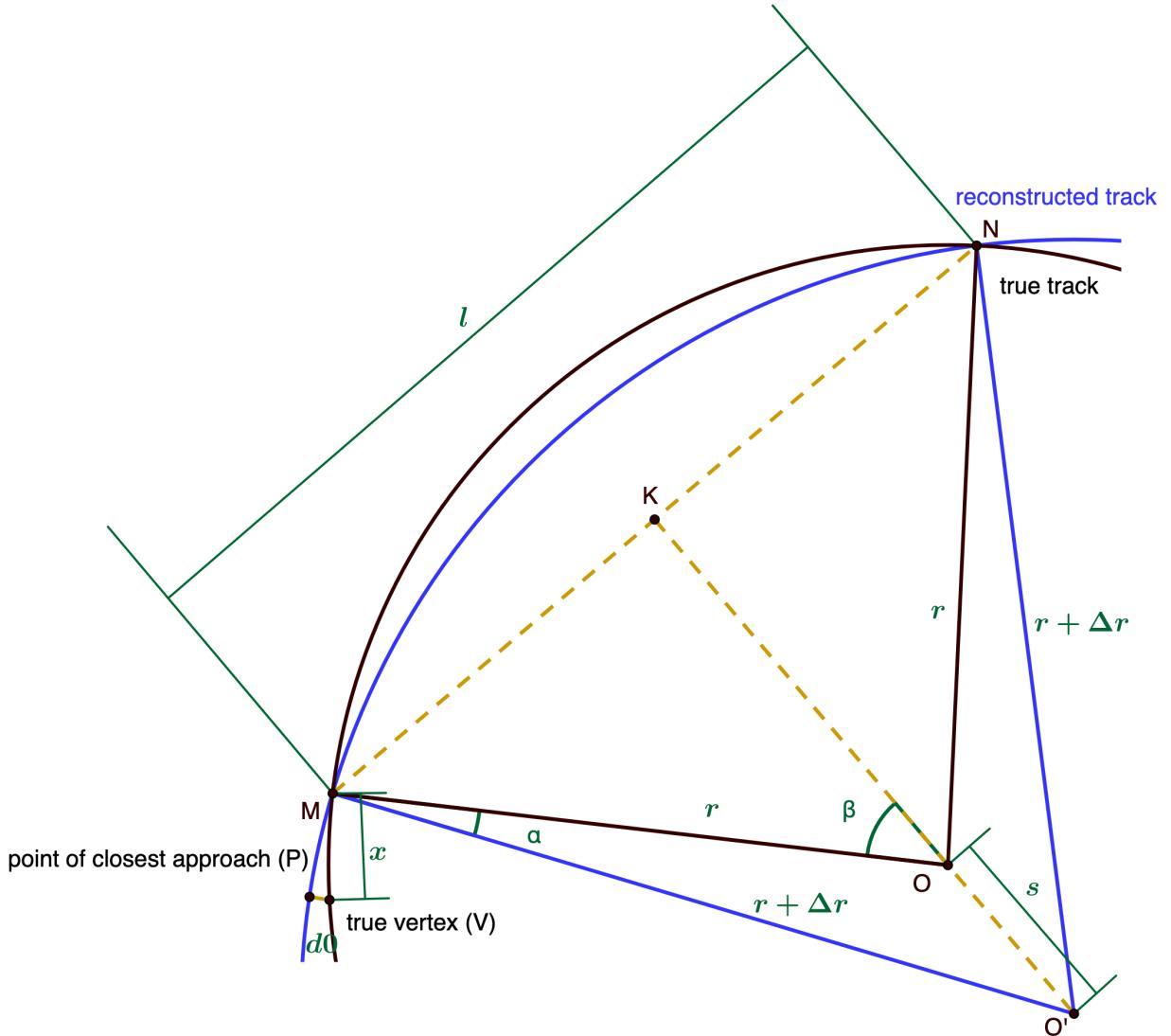


Figure 5-4. Scheme of the track geometry in the transverse plane. The blue lines show the geometry of the reconstructed track, compared to the black lines which are the geometry of the true track. The difference between the blue track and black track is exaggerated in this scheme. The blue track and the black track must intersect at two points. l is the distance between the two intersections, and x is the distance between the true vertex and the first intersection. s is the distance between the circular centers of the two tracks.

In triangle $\triangle O'MO$, the sine law gives

$$\frac{s}{\sin \alpha} = \frac{r + \Delta r}{\sin \beta} \quad (5-3)$$

And in triangle $\triangle O'MK$,

$$\sin\beta = \frac{l/2}{r} \quad (5-4)$$

Then, using the Pythagorean theorem in both triangle $\triangle O'MK$ and triangle $\triangle OMK$, there is

$$s = \sqrt{(r + \Delta r)^2 - (l/2)^2} - \sqrt{r^2 - (l/2)^2} \quad (5-5)$$

Combining Equation 5-2 to 5-5 and assuming $r \gg l$, one can get

$$d0 = \frac{xl}{2} \cdot \frac{\Delta r}{r^2} \quad (5-6)$$

Note that in CMS, under the 3.8T magnetic field, tracks follow

$$p_T \text{ (in GeV)} = 1.14 \cdot r \text{ (in meter)} \quad (5-7)$$

We reach

$$d0 \propto \frac{\Delta p_T}{p_T^2} \quad (5-8)$$

Now a quantitative relationship is extracted between $d0$ and p_T , but with one caveat: the variables x and l in the scheme above may vary track by track, and are impossible to measure in real data, meaning that the coefficient in the proportionality is not a constant for different tracks, and Equation 5-8 can be smeared. Therefore, to validate this proportionality, studies are performed on simulated samples comparing the reconstructed p_T and the generated p_T of muon tracks. Plots of $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$ vs $d0$ are made, to see whether the proportionality can be observed after the smearing, which is the topic of Section 5.3.2.

Another remark needs to be made, that in Figure 5-4 the p_T mismeasurement is related to the relative position of the true vertex to the reconstructed track. To be more specific, if the true vertex is inside of the reco track, the p_T is overestimated, while if the true vertex is outside of the reco track, the p_T is underestimated. However, in the CMS definition of $d0$ shown in Figure 5-3, the sign of $d0$ corresponds to an opposite

relative position between the vertex and the track for the positively charged muons and the negatively charged muons. A positive $d0$ value means the true vertex is inside of the reco track if the muon is positive, but outside of the reco track if the muon is negative. Therefore in CMS convention the $d0-p_T$ correlation is expected to be reversed for different muon charges, and in Section 5.3.2 studies are always performed evaluating $d0 \cdot \text{charge}$ rather than just $d0$.

5.3.2 Development of GeoFit

The $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$ vs $d0$ plots are made with the following steps: The values p_T^{reco} , p_T^{gen} , and $d0$ are extracted for each track in simulated samples. The distribution of $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$ is made for tracks in different $d0 \cdot \text{charge}$ bins. The maximum position and the corresponding full-width-half-maximum (FWHM) is found for each fine-binned $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$ distribution and set as the value and the uncertainty of one data point in the plots in Figure 5-5. The plots are then fit with analytic functions which are considered as the experimental realization of Equation 5-8.

In CMS, the colliding position can be measured as two different physics objects, the primary vertex (PV) and the beam spot (BS). A primary vertex is a 3D-point which is compatible with several tracks in the same event. Usually several PVs are reconstructed per event, each of them considered as the position of a pp collision instance. The beam spot, on the other hand, is actually a 3D-region in which most of the pp collisions happen, and is reconstructed with all tracks in all the events in many consecutive luminosity sections. The beam spot is usually around 10-20 μm wide in the x, y directions, and about 7-9 cm long in the z direction.

Both the PV and the BS are reasonable representation of the colliding position and are useful in different cases. The PV has a good z coordinate precision and can be used to judge if two tracks originate from the same interaction. But the position of the PV can be biased by the few energetic tracks associated to it, as the tracks are weighted by their p_T^2 in the reconstruction of the PV. The BS is wide-spread in z direction, but is less

affected by individual tracks in its x, y coordinates. To compare the two types of vertices, the d0 is measured regarding each of them, and examples of the $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$ vs d0 dependences are shown in Figure 5-5. The dependence in the PV plot is not linear as the reconstructed PV is pulled towards the energetic muon tracks, while the dependence in the BS plot follows a linear trend as predicted in Equation 5-1. Therefore, the BS is considered as the position of the true vertex in the rest of the study.

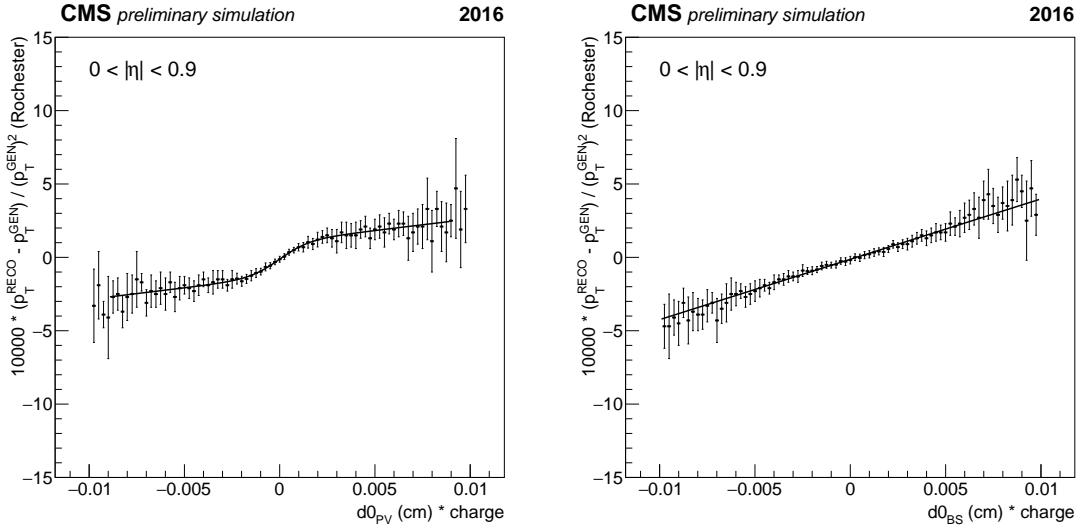


Figure 5-5. Plots credit to Efe Yigitbasi. Example plots showing the correlation between $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$ and $d0 \cdot \text{charge}$. The vertices used for the d0 calculation are the PV (left) and the BS (right). The PV plot shows a modulated dependence from expectation while the BS plot shows a linear shape as expected. Only barrel tracks from 2016 data are shown as examples. Plots of other $|\eta|$ regions and other data-taking periods show similar behaviors.

The $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$ vs d0 correlation is found to be different in different $|\eta|$ regions and data-taking periods: different $|\eta|$ regions are covered by different detector components, and there have been upgrades on the detector and the reconstruction algorithm between different data-taking periods. Overall, the $(p_T^{reco} - p_T^{gen})/(p_T^{gen})^2$ vs d0 correlation is evaluated by three years (2016, 2017, 2018) and three $|\eta|$ regions (barrel, overlap, endcap), shown in Figure 5-6 for 2016, 5-7 for 2017, and 5-8 for 2018. Each of the plot is fit with a linear function, whose best fit parameters are also shown in the plot.

These fit results are applied as the analytic correction to muon p_T , based on the d0,

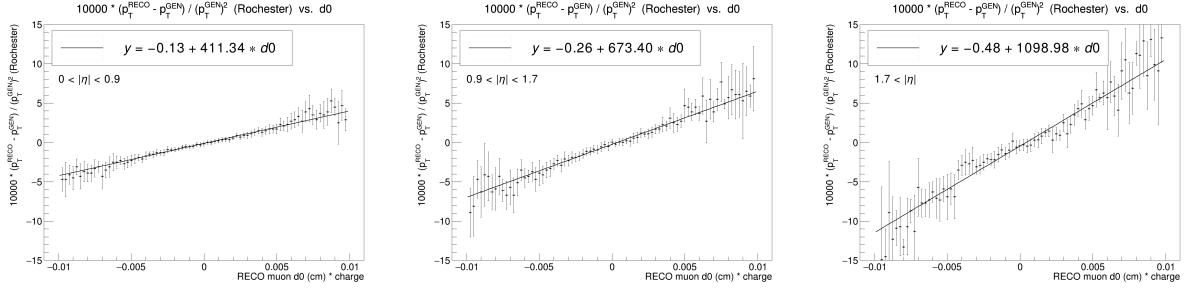


Figure 5-6. Plots credit to Efe Yigitbasi. Plots for the $(p_T^{reco} - p_T^{gen}) / (p_T^{gen})^2$ vs d0 correlation in the 2016 DY simulation, and the linear fits to them. Muon tracks are divided into three different $|\eta|$ regions: $|\eta| < 0.9$ (left), $0.9 < |\eta| < 1.7$ (middle), and $1.7 < |\eta|$ (right).

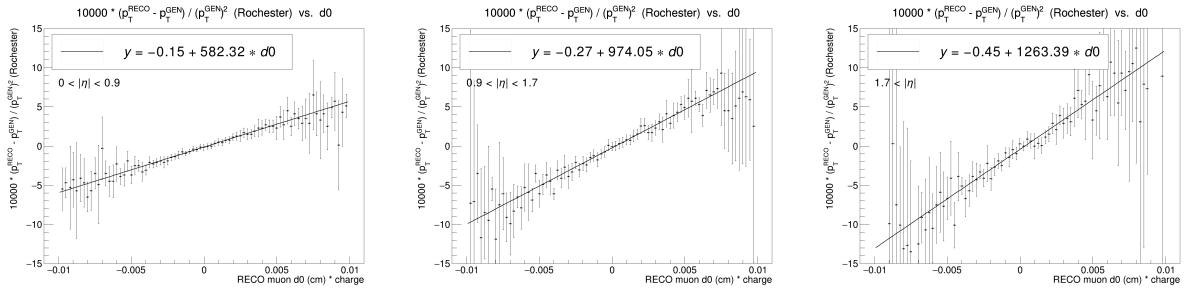


Figure 5-7. Plots credit to Efe Yigitbasi. Plots for the $(p_T^{reco} - p_T^{gen}) / (p_T^{gen})^2$ vs d0 correlation in the 2017 DY simulation, and the linear fits to them. Muon tracks are divided into three different $|\eta|$ regions: $|\eta| < 0.9$ (left), $0.9 < |\eta| < 1.7$ (middle), and $1.7 < |\eta|$ (right).

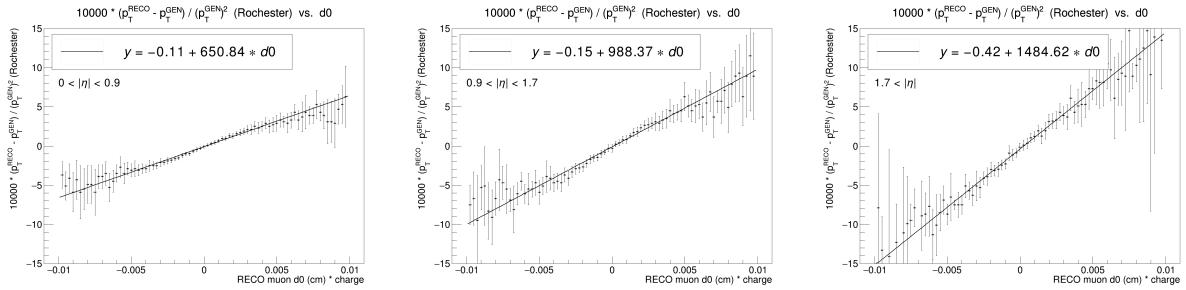


Figure 5-8. Plots credit to Efe Yigitbasi. Plots for the $(p_T^{reco} - p_T^{gen}) / (p_T^{gen})^2$ vs d0 correlation in the 2018 DY simulation, and the linear fits to them. Muon tracks are divided into three different $|\eta|$ regions: $|\eta| < 0.9$ (left), $0.9 < |\eta| < 1.7$ (middle), and $1.7 < |\eta|$ (right).

p_T , $|\eta|$, and charge of the muon. The correction is applied to all muons in data and simulation in all categories in the $H \rightarrow \mu\mu$ analysis, unless the muon is tagged for FSR recovery. The performance of this correction is detailed in Section 5.3.3.

5.3.3 Performance and validation

The *GeoFit correction* removes the p_T dependence on $d0$, the overall effect of which on the $Z \rightarrow \mu\mu$ peak is illustrated in Figure 5-9. A clear trend in the $m_{\mu\mu}$ is seen regarding to $d0$ before the *GeoFit correction*, while no significant dependence remains after the correction. As a side remark, the $m_{\mu\mu}$ mismeasurement in Figure 5-9 can be as large as 1.5 GeV for extreme $d0$ values, but in data and simulation the distribution of the muon $d0$ is roughly a Gaussian shape with a standard deviation around 15 μm . So most of the events are near the center of the plots, and the size of the correction is not as exaggerated as the values at the tails.

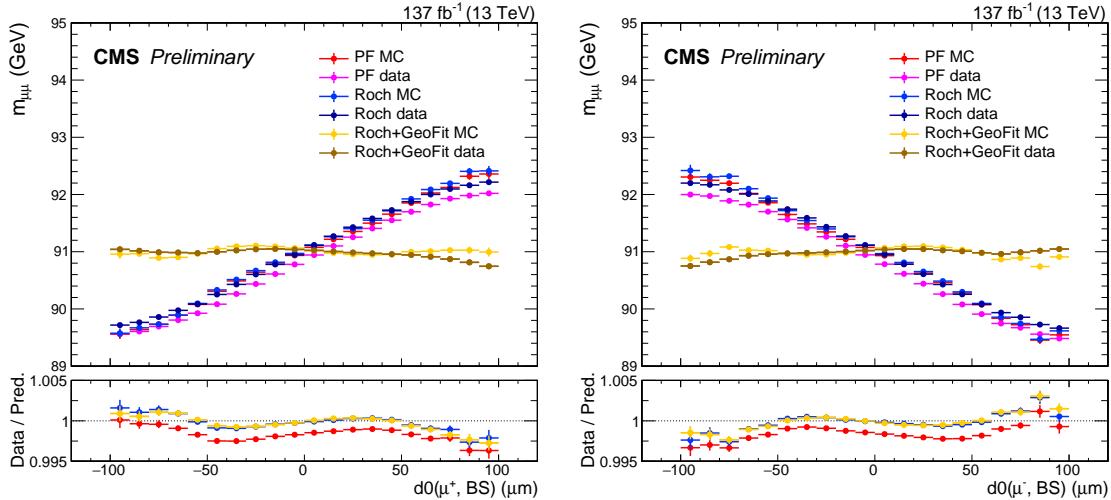


Figure 5-9. Plots showing the p_T dependence on the $d0$ value with different stages of muon correction. The plots compare the $Z \rightarrow \mu\mu$ peak in data and simulation for three years (2016-2018) combined. All positively charged muons are put in the left plot and all negatively charged ones are put in the right plot. The p_T - $d0$ dependence is reversed for positive and negative muons.

Overall, the removal of the p_T - $d0$ dependence leads to an improvement on the inclusive $m_{\mu\mu}$ resolution. This improvement is different for different processes depending on their kinematic profiles in p_T and $|\eta|$. Figure 5-10 shows the improvement on $m_{\mu\mu}$ resolution in the four main expected signal modes, ggH, VBF, VH, and $t\bar{t}H$. The relative improvements on $m_{\mu\mu}$ resolution for ggH, VBF, VH, and $t\bar{t}H$ modes are, respectively, 6.1%,

7.8%, 8.0%, and 9.8%. This improvement on signal resolution translates into about 5% improvement on the significance of the inclusive $H \rightarrow \mu\mu$ analysis.

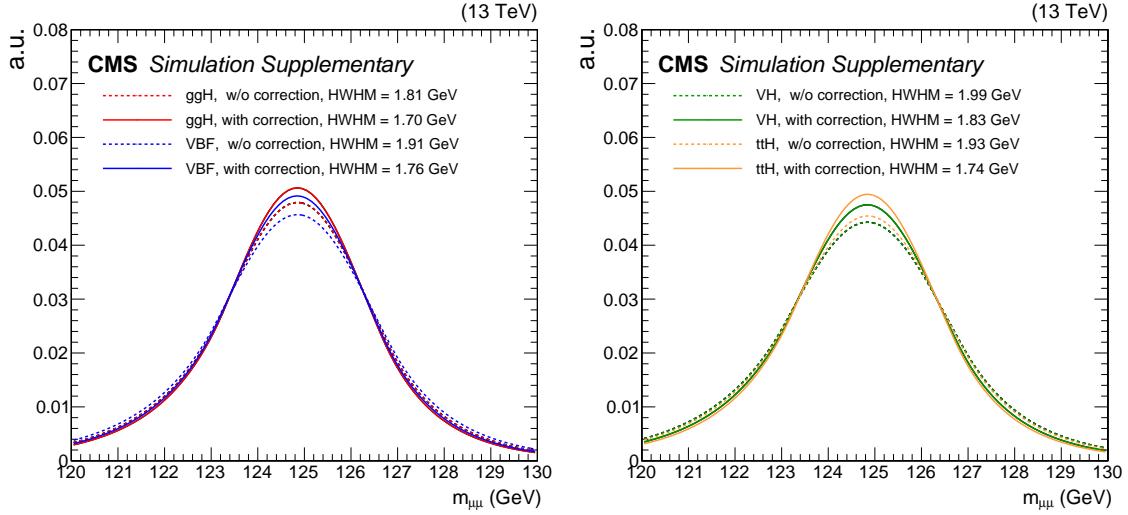


Figure 5-10. Plots showing the *GeoFit correction* improvement on the four main $H \rightarrow \mu\mu$ signal modes, ggH and VBF plotted on the left, and VH and $t\bar{t}H$ plotted on the right. The plots are made combining the expected signal in all three years of data-taking (2016-2018). The relative improvements on $m_{\mu\mu}$ resolution for ggH, VBF, VH, and $t\bar{t}H$ modes are, respectively, 6.1%, 7.8%, 8.0%, and 9.8%.

5.3.4 GeoFit vs track re-fit

The *GeoFit correction* provides a simple method to correct the p_T dependence on d0 based on high level physics variables. Since the origin of this p_T dependence is well-understood, it is also possible to derive a more fundamental correction by re-fitting each muon track including the BS position as an additional constraint to the track. This method requires lower-level information of muon reconstruction and is computationally more expensive, but is in principle more precise. To compare the performance of the *GeoFit correction* and the re-fitting method, a preliminary study is made on the 2018 ggH signal simulation. The $m_{\mu\mu}$ shape of the inclusive signal is plotted applying the track re-fit method vs applying the *GeoFit correction*, shown in Figure 5-11. This comparison shows that the $m_{\mu\mu}$ shapes from the two methods are almost equivalent. The *GeoFit correction*, although an approximation method, captures most of the effect and provides about the same improve-

ment in $m_{\mu\mu}$ resolution as the re-fitting method. The *GeoFit correction* is therefore chosen in the $H \rightarrow \mu\mu$ analysis to speed up the workflow. In the meantime, the possibility of CMS centrally providing the re-fit results for more general use cases is under discussion, which may become an official option in Run 3 (2022-2024) of CMS data-taking.

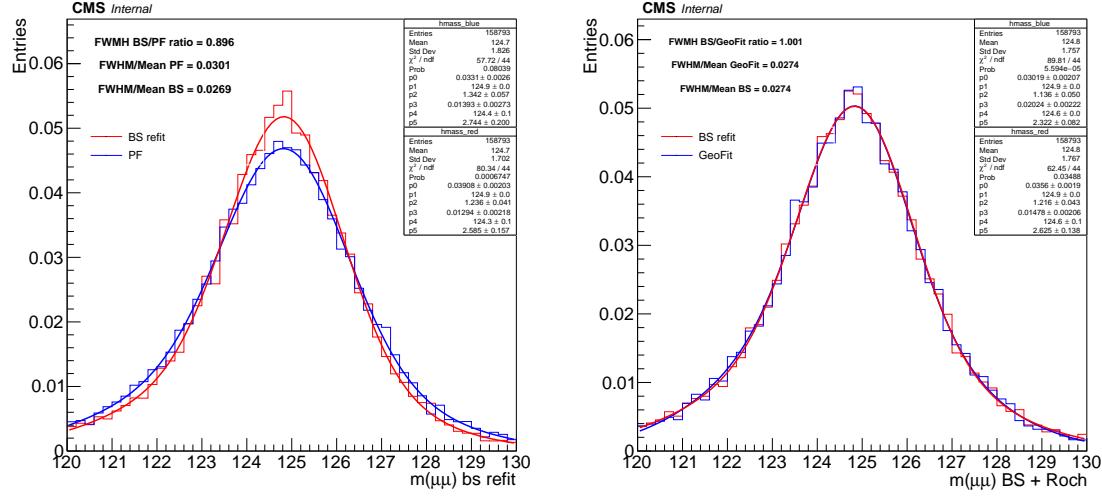


Figure 5-11. Plots credit to Pierluigi Bortignon. Plots of the $m_{\mu\mu}$ shape of the 2018 ggH simulation sample, comparing different muon correction methods. The left plot shows the $m_{\mu\mu}$ distribution calculated with muon tracks re-fit with the additional BS constraint, compared with the particle flow shape (left plot). The *Rochester correction* is not applied in the left plot for both the red and the blue lines. The right plot shows the $m_{\mu\mu}$ distribution from the re-fit method, with the *Rochester correction* applied, compared with the shape from *GeoFit correction* + *Rochester correction* (right plot).

5.4 Muon calibration results

The $Z \rightarrow \mu\mu$ is a well-understood process with a mass scale not far from the Higgs boson and with a much larger number of events in CMS. It is therefore used as a candle to monitor the performance of the *Rochester correction* and the *GeoFit correction*, and validate that these corrections do not introduce new biases. In this study, the distribution of the $m_{\mu\mu}$ is plotted in different bins of some dimuon kinematic variables. The $m_{\mu\mu}$ distributions are fit with a Voigtian + Exponential function, the Voigtian part being a convolution of a Breit-Wigner function and a Gaussian function. The parameters mean mass from the Breit-Wigner part and standard deviation from the Gaussian part are taken as the mean

value and the experimental resolution of the $m_{\mu\mu}$ distribution, and are plotted against the dimuon kinematic variable of interest to check for potential trends.

The calibration plots are made by year as the corrections are provided by year. Events containing *FSR recovery* are removed from this study as it is a separate effect. Different variables are tested in the Figures listed: Figure 5-12 for the η of the positive muon, Figure 5-13 for the ϕ of the positive muon, Figure 5-14 for the ϕ of the negative muon, Figure 5-15 for the p_T of the positive muon, Figure 5-16 for the p_T of the dimuon system, Figure 5-17 for the η of the dimuon system, Figure 5-18 for the d_0 of the positive muon, and Figure 5-19 for the d_0 of the negative muon.

From these plots, it can be concluded that all the known biases in muon p_T are removed and no new bias has been introduced. The *Rochester correction* and *GeoFit correction* correct orthogonal effects, and do not interfere the performance of each other. After the corrections, a per-mille level agreement is achieved between data and simulation in the $m_{\mu\mu}$ value, while the agreement in $m_{\mu\mu}$ resolution is about a few percent.

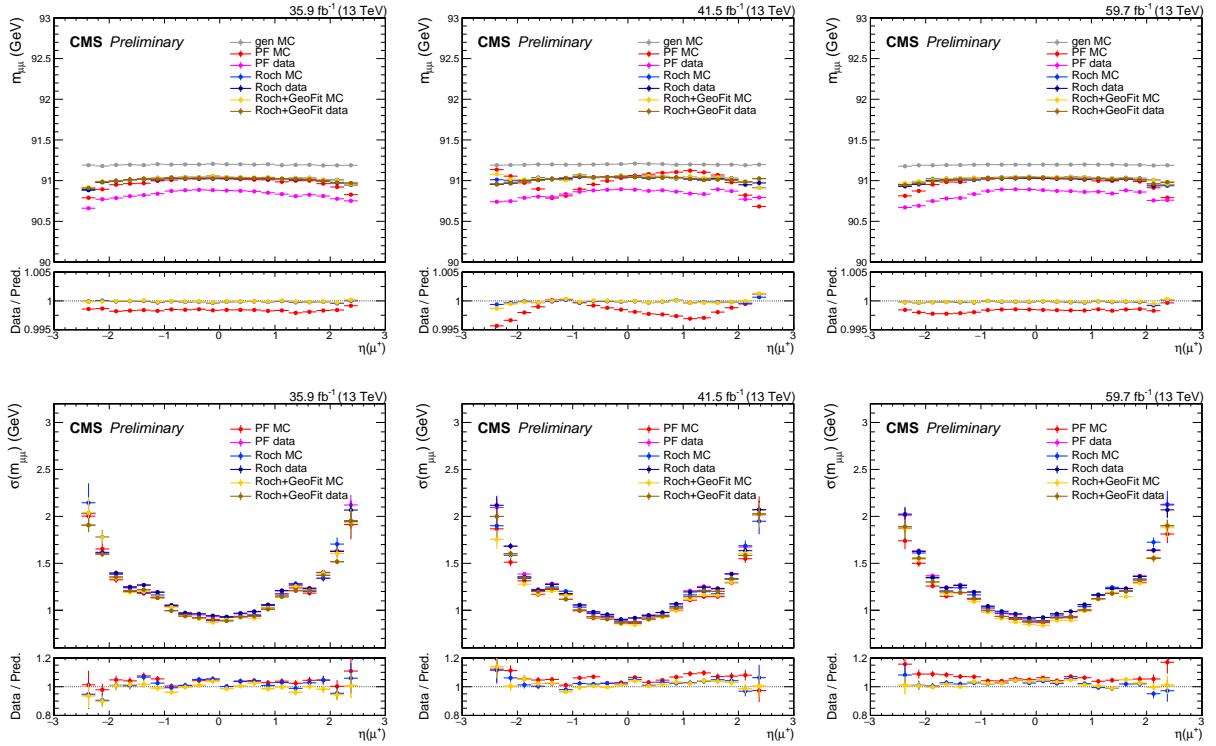


Figure 5-12. Muon calibration plots vs $\eta(\mu^+)$, for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution.

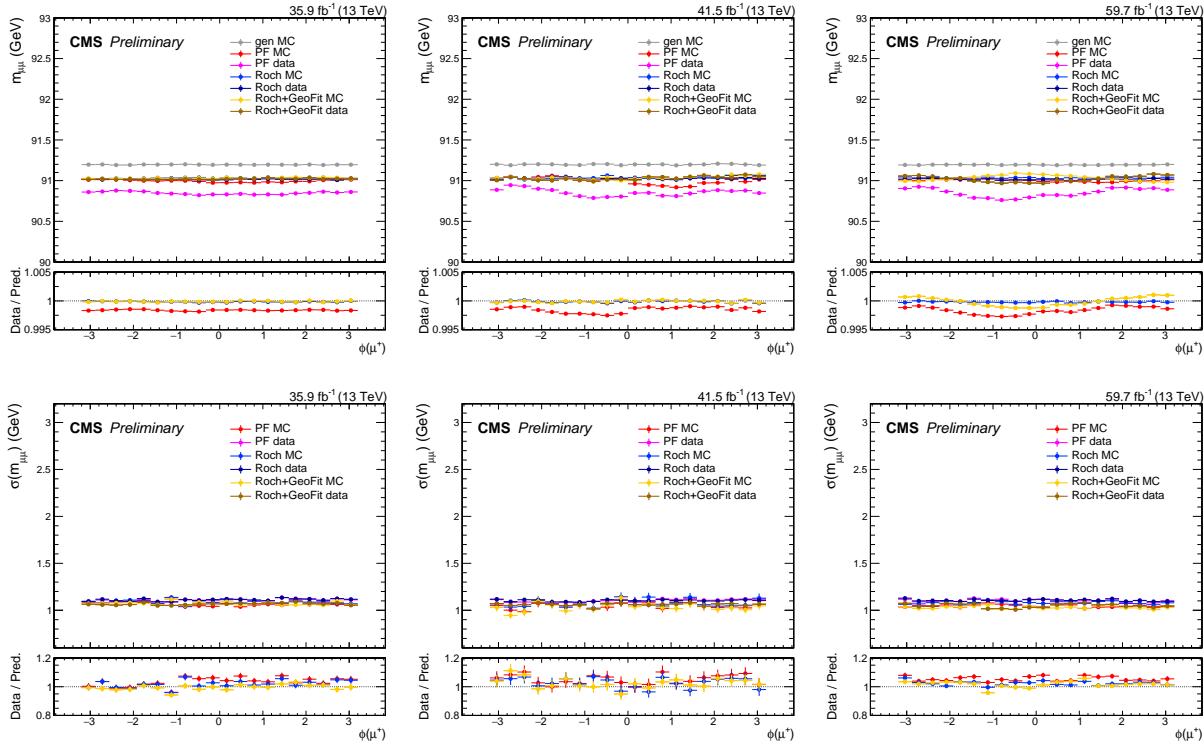


Figure 5-13. Muon calibration plots vs $\phi(\mu^+)$, for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution.

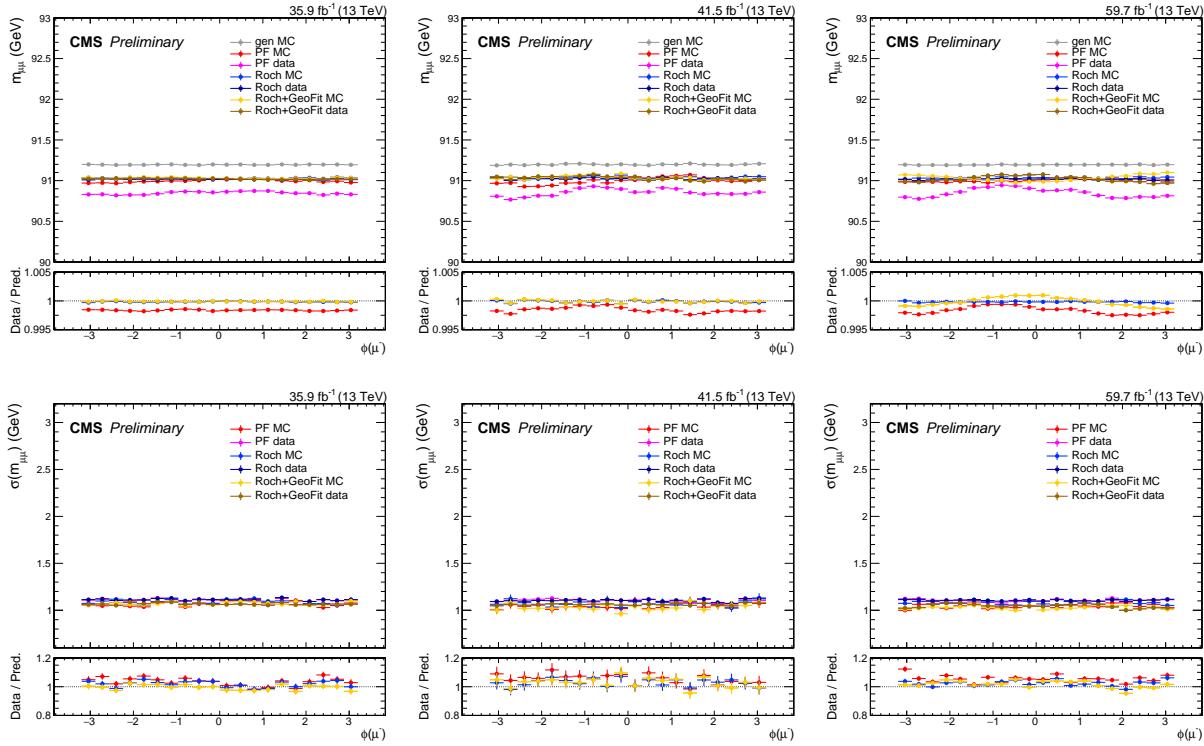


Figure 5-14. Muon calibration plots vs $\phi(\mu^-)$, for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution.

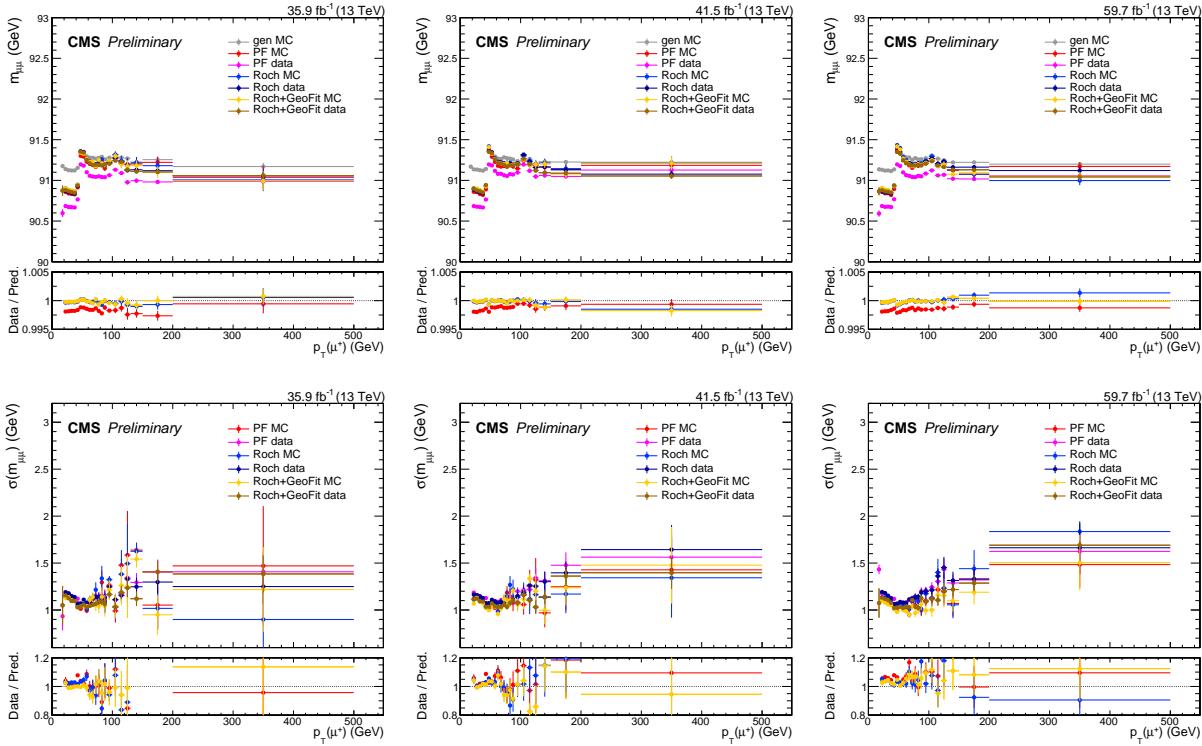


Figure 5-15. Muon calibration plots vs $p_T(\mu^+)$, for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution. The p_T binning sculpts the shape of the $m_{\mu\mu}$ peak, which leads to a jump at the $p_T = 45$ GeV in the plots.

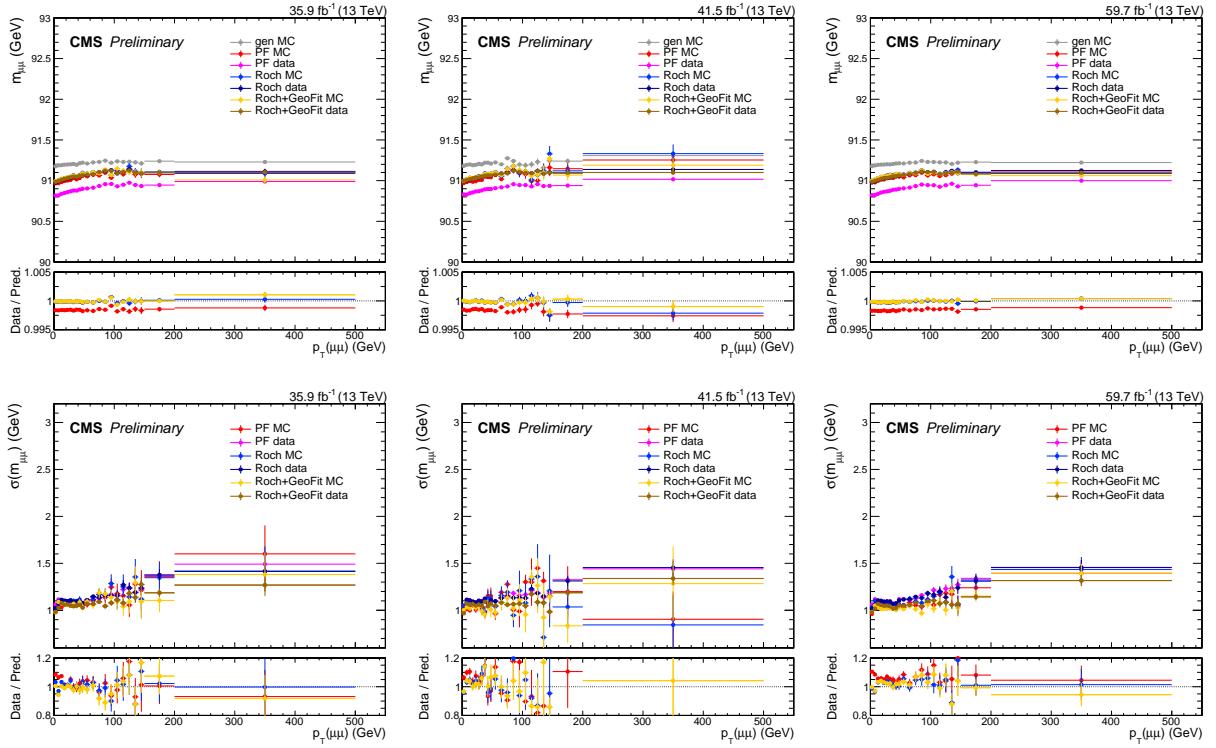


Figure 5-16. Muon calibration plots vs $p_T(\mu\mu)$, for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution.

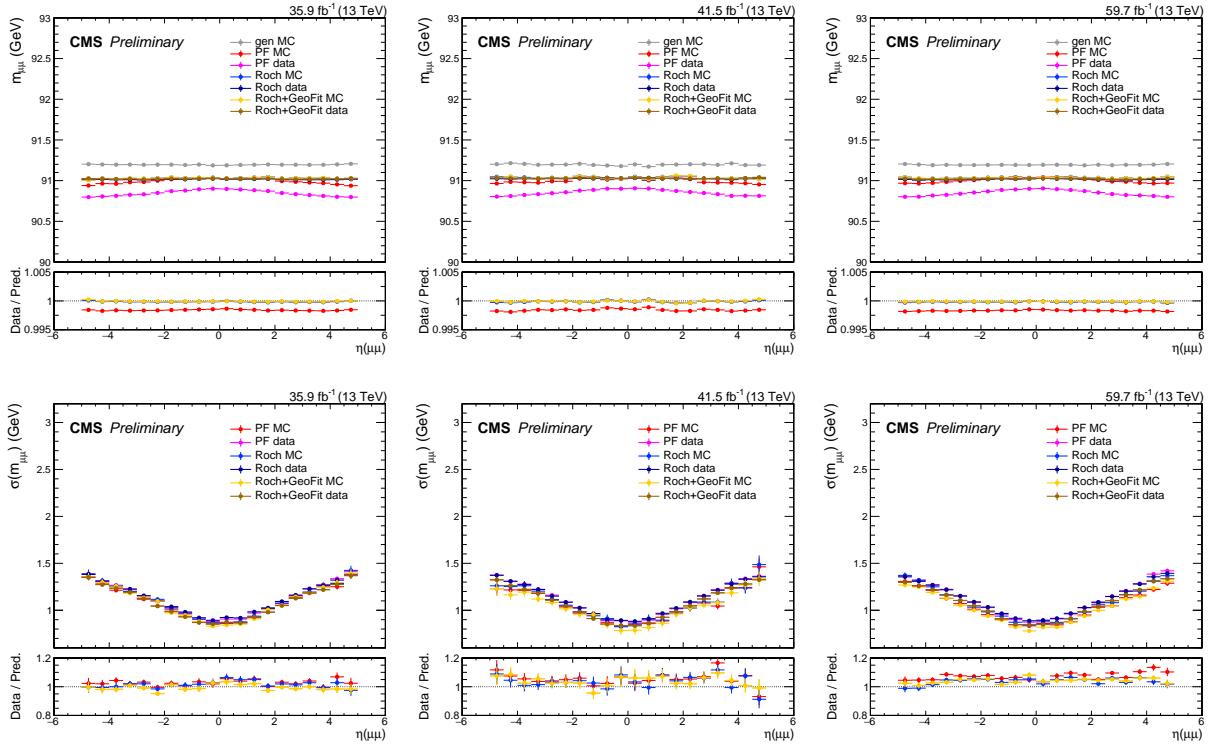


Figure 5-17. Muon calibration plots vs $\eta(\mu\mu)$, for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution.

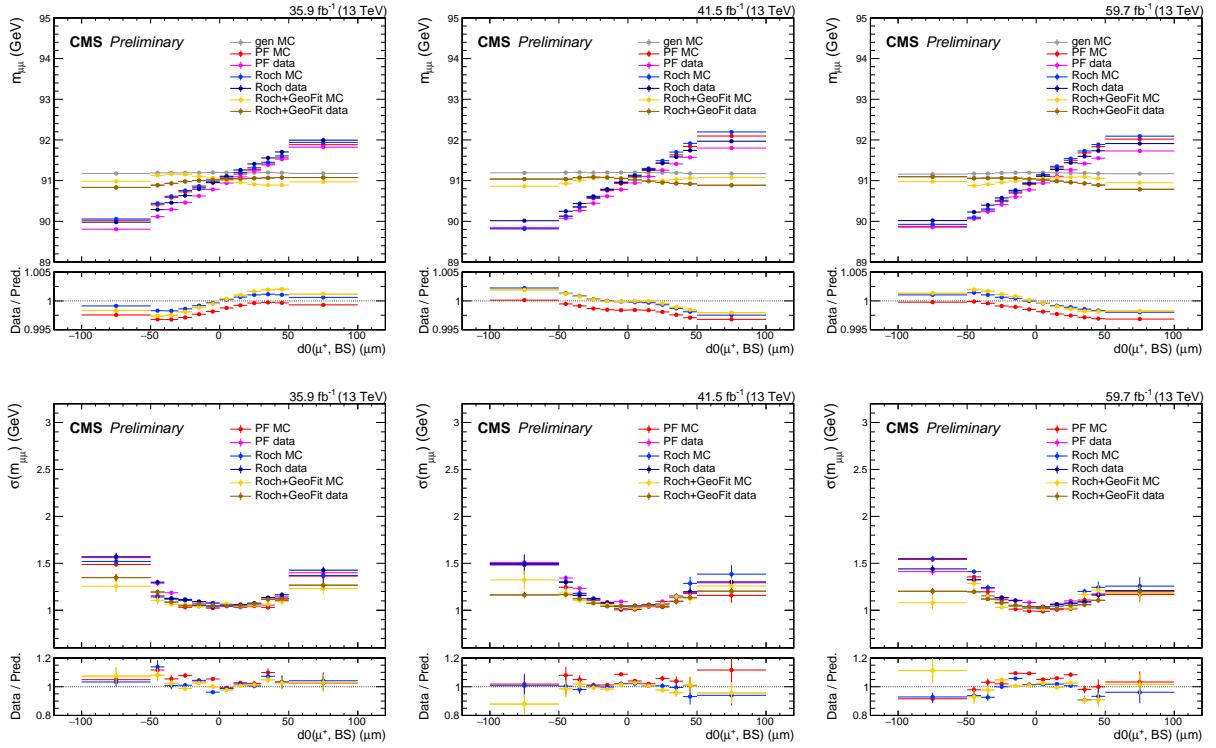


Figure 5-18. Muon calibration plots vs $d0(\mu^+)$, for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution.

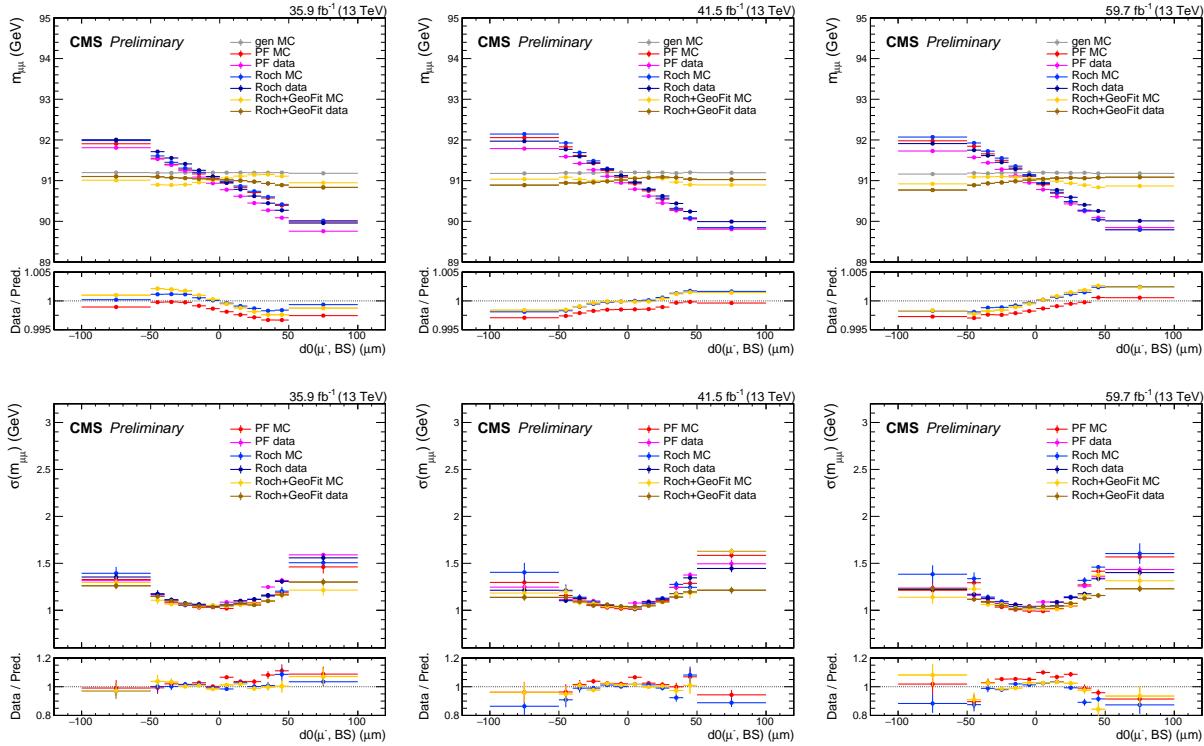


Figure 5-19. Muon calibration plots vs $d0(\mu^-)$, for 2016 (left column), 2017 (middle column) and 2018 (right column). The top row shows the mean value of the Voigtian fit to the $m_{\mu\mu}$ distribution, while the bottom row shows its experimental resolution.

CHAPTER 6

SEARCH FOR H2MU TARGETING THE VH PRODUCTION MODE

As described in Section 3.2, the analysis in the VH category targets the VH production modes of the Higgs boson, and is established as two independent parts, the WH category and the ZH category. This chapter provides a full description of the procedures and the results in these two categories. In this chapter the VH category will sometimes be referred to as the VH categories, when we focus more on the individual specifications of the WH and ZH categories, rather than their common characteristics.

The VH analyses focus on the leptonic (e or μ) decay modes of the V boson (W or Z), which leads to distinct final states involving extra well-reconstructed charged lepton(s) in addition to the two muons from the Higgs decay. By requiring these additional leptons, the main backgrounds in the generic $\mu\mu$ phase-space, the DY and the $t\bar{t}$ processes, are greatly suppressed, leading to a high S/B in the VH categories, and ensuring a good expected significance with only a handful of signal events. The other decay modes of the V boson are disregarded for different reasons:

The V bosons decay to tau leptons ($W \rightarrow \tau + \nu_\tau$, or $Z \rightarrow \tau\bar{\tau}$) at the same rate as they decay to electrons or muons. However, in CMS, the reconstruction of τ is not as efficient as that of electrons or muons: the reconstruction of the hadronic decays of τ suffers from a sizable fake rate, while the leptonic decays of τ would just appear as an electron or a muon of much lower p_T . Tagging the hadronic decays of taus from the W or Z decay would lead to a much lower S/B than the e and μ tags in the current VH analyses, bringing negligible contribution to the overall sensitivity. As for the leptonic decays of τ , they are not explicitly vetoed. If any VH signal events containing such decays pass the selection, they are considered as part of the signal contribution. Although, this contribution is small, as the e or μ from the τ decay would most likely fail the p_T or the vertex proximity cuts.

The hadronic decays of the W or the Z boson lead to two jets in the event, making an invariant mass near the mean mass of the boson. Although amounting to a larger branching ratio than the leptonic decay modes, these hadronic decays turned out not as helpful in enriching the VH events. A selection based on the dijet invariant mass could

pick most of the VH events, but in the meantime collect a much larger amount of ggH+jets and VBF events, as well as an immense background of DY+jets and $t\bar{t}$ processes. The S/B from this selection is not very high, as this phase-space is dominated by ggH+jets vs DY+jets events. Up to the time of this report, no kinematic handle is found to be particularly effective in enriching the hadronic VH signal. Therefore, the hadronic VH events are considered as minor signal contributions in the ggH category, and no dedicated VH hadronic tag is deployed.

The Z boson can decay to a pair of neutrinos at a branching ratio of 20%, which leave no electronic signal in the CMS detector and appear as a missing transverse momentum (E_T^{miss}) in the event. This E_T^{miss} equals to the p_T of the Z boson and can provide discrimination against some backgrounds. However, due to the complex hadronic activity in pp collisions and the large uncertainty in the hadronic calorimetry in CMS, almost all events are reconstructed with some nonzero E_T^{miss} , even for events without any real E_T^{miss} , like the DY events. As a result, the purity of the $ZH \rightarrow vv + \mu\mu$ signal can only be enhanced if a very tight cut on the E_T^{miss} is applied. The signal efficiency for this cut is, unfortunately, low, because only a small fraction of the Z bosons in ZH events have high p_T . Furthermore, after this cut, some large irreducible backgrounds still remain, like the $t\bar{t}$ and diboson processes. Overall, a tag for the $ZH \rightarrow vv + \mu\mu$ would be much less sensitive than the existing $ZH \rightarrow \ell\ell + \mu\mu$ tag, and is not deployed. The $ZH \rightarrow vv + \mu\mu$ signal events are considered as minor signal contributions in the ggH category.

After requiring additional lepton(s) in the event, the main background in the resulting VH phase-spaces becomes the WZ and ZZ processes, for the WH and ZH categories respectively. For both WZ and ZZ backgrounds, if more than two muons are present in the event, there are different possibilities of the association between the muons and their parent particles. In fact, the wrong pairing of muons, called the combinatorial background, yields the majority of the WZ and ZZ background. For example, in an on-shell WZ event, the muon from the W decay can be falsely paired with the oppositely signed (OS) muon

from the Z decay, making an invariant mass near the Higgs mass value. A set of cut-based event selection is optimized to reduced the combinatorial background, further improving the S/B . More details are given in Section 6.1.

The minor backgrounds may include the triboson processes, the DY process accompanied with additional nonprompt leptons, or the t quark associated processes, for example the $t\bar{t}$, tW , and $t\bar{t}V$ processes, where the b quarks from the top decays either fall out of the acceptance of the detector or fails the b-tagging. All these minor backgrounds have different kinematic profiles from the signal and can be reduced to different extents. Boosted Decision Tree (BDT) discriminators are trained in both the WH and the ZH categories to account for the differences between the signal and the inclusive background as much as possible.

With the BDT discriminators, events are further divided in to several sub-categories with different S/B . In each sub-category, the $m_{\mu\mu}$ spectrum is plotted, and analytic functions are used to model the shapes of signal and background. The strength of the $H \rightarrow \mu\mu$ signal is evaluated by fitting the signal and the background functions to the $m_{\mu\mu}$ spectrum of data. Combining the results in all WH and ZH sub-categories, an observed (expected in absence of $H \rightarrow \mu\mu$ decay) upper limit of 10.8 (5.13) times the SM prediction is set at the 95% confidence level (CL) on the product of the Higgs boson production cross section and $\mathcal{B}(H \rightarrow \mu\mu)$. The corresponding signal strength, relative to the SM prediction, is $\mu = 5.48^{+3.10}_{-2.83}$. An excess of signal events is observed (expected with the SM prediction) with a significance of 2.02 (0.42) standard deviations.

The following sections of this chapter are organized as follows: Section 6.1 describes the event selection. Section 6.2 discusses the details of the training of the BDTs, and the determination of sub-categories based on them. Section 6.4 shows the performances of different analytic functions on the signal and background modeling. Section 6.5 lists different sources of systematic uncertainties considered in this analysis. And Section 6.6 gives the statistical interpretation of the results.

6.1 Event selection

The VH analysis takes physics objects reconstructed by the particle-flow (PF) algorithm [48]. Selections on the objects are described in details in Chapter 4. The selection criteria for electrons and muons, which are the most important for this analysis, are also summarized in Tabel 6-1.

Table 6-1. Selection criteria on muons and electrons in the VH analysis.

Variable	Muon	Electron
p_T	$> 20\text{GeV}$	$> 20\text{GeV}$
$ \eta $	< 2.4	< 2.5
ID and Iso	Medium ID + Loose Iso	MVA wp90
ECal gap veto	-	(1.444, 1.566)
$d_{xy}(\text{PV})$	$< 0.05 \text{ cm}$	$< 0.05 \text{ cm}$
$d_z(\text{PV})$	$< 0.10 \text{ cm}$	$< 0.10 \text{ cm}$
SIP	< 8.0	< 8.0
Conversion Veto	-	✓
Number of Missing Hits	-	< 2
lepMVA	> 0.4	> 0.4

The event selection targets the leptonic decays of the W or the Z bosons in the VH signals. The selection steps are deviced to suppress different background processes and optimize the S/B in the WH and ZH categories respectively. The event selection in the $\text{WH} \rightarrow \ell\nu + \mu\mu$ category is described as follows:

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

The event selection in the $ZH \rightarrow \ell\ell + \mu\mu$ category is described as follows:

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

These selection criteria are also summarized in Table 6-2.

Table 6-2. Event selections for the WH and ZH categories.

Criterion	WH category	ZH category
Muon trig match	✓	✓
b-jets veto, 0 medium and < 2 loose	✓	✓
$\mu^+\mu^-$ pair with $110 < m_{\mu\mu} < 150\text{ GeV}$	✓	✓
Additional lepton(s)	1	1 SFOS pair
Low-mass resonance veto $m_{\ell\ell} > 12\text{ GeV}$	✓	✓
Number of $ m_{\mu\mu} - m_Z < 10\text{ GeV}$ or $ m_{ee} - m_Z < 20\text{ GeV}$	= 0	= 1
Choice of muon combination	Highest $p_T(\mu\mu)$ as $\mu\mu_H$	Smallest $ m_{\mu\mu} - m_Z $ as $\mu\mu_Z$

efficiency of sig combo and bkg combo, plots before and after.

6.2 MVA discrimination

After the event selection of the WH or the ZH categories, the remaining background processes resemble the kinematic signatures of the signals and cannot be decisively reduced by simple selection cuts. To further suppress the backgrounds and enhance the S/B , a BDT is trained in each category, making use of the many lesser discriminating variables.

Different variables can be effective in separating different background processes. For example, the muon pair from t quark associated processes usually have more p_T than those from the WH signal process, while the E_T^{miss} in the DY process is likely to be smaller than that in the WH events. The main background, WZ or ZZ in WH and ZH categories respectively, in which the Higgs candidate $\mu\mu$ pair comes from an off-shell Z boson, is a more complicated story as it looks almost kinematically identical to the signal. The key in discriminating them lies in the spin difference between the Z and the H bosons, which is measured as a difference in the helicity angle θ^* between the decay products of the Z (H) boson.

The helicity angle is defined in a decay system in the frame in which the parent particle is at rest, as the angle between the direction of the decay and the boost direction of the parent particle. The distribution of this angle is determined by the spins of the parent particle and the decay products. In the case of $W \rightarrow WZ$ ($Z \rightarrow ZZ$) vs $W \rightarrow WH$ ($Z \rightarrow ZH$), the helicity angle between the W (Z_1) and the Z_2 follows the distribution $1 + \cos^2 \theta^*$ ^a [illustrative figure](#), while the the helicity angle between the W (Z) and the H follows a flat distribution. Similarly, for $Z \rightarrow \mu\mu$ vs $H \rightarrow \mu\mu$, the muons from a Z decay tend to align with the polarization of the Z boson, while the muons from a H decay do not have a preference in direction.

All these helicity angles, θ_{WH}^* , θ_{ZH}^* , and $\theta_{\mu\mu}^*$, can in principle provide significant discrimination between the signal and the background. However, in practice, θ_{WH}^* cannot be reconstructed due to the lack of information of the neutrino from the W decay, and the distribution of $\theta_{\mu\mu}^*$ is severely sculpted by the acceptance of the CMS detector, rendering a similar shape for the $Z \rightarrow \mu\mu$ and $H \rightarrow \mu\mu$ processes. On the other hand, the spin information can be partially captured in other variables, for example the helicity angle between the W-lepton and the $\mu\mu_H$ system, and some other angular correlations like $\Delta\phi$, $\Delta\eta$. All these variables are tested as inputs to the BDT for the performance. Some of them turned out insignificant and are later trimmed off from the input collection. To make sure the

BDTs do not sculpt the $m_{\mu\mu}$ shape, which will be used for the signal extraction, variables which are strongly correlated with $m_{\mu\mu}$, for example the p_T of the muons, are not used in the BDTs.

6.2.1 BDT targeting $\text{WH} \rightarrow \ell\nu + \mu\mu$ signal

The WH BDT takes variables of three different kinds, the kinematic variables of leptons which are uncorrelated with $m_{\mu\mu}$, the angular correlations between different leptons as discussed above, and the variables reflecting the missing energy in the event. In total, there are 16 input variables to the BDT, which are listed in Table 6-3. The variables related to the missing energy includes the missing energy itself, the transverse mass (M_T) between a lepton and the missing energy, and the angular separation between a lepton and the missing energy. Two types of missing energy are tested, E_T^{miss} , which is the inverse of the sum of the transverse energy of all PF candidates, and H_T^{miss} , which only considers well-defined jets, photons, and leptons in the similar calculation. The E_T^{miss} -related and H_T^{miss} -related variables are expected to play interchangeable roles in the BDT. The H_T^{miss} -related variables turned out to be slightly better performing and are kept in the final BDT, while the E_T^{miss} -related variables are trimmed.

Another important feature in the WH category is that about 40% of the WZ background are from the wrong combination of muons in 3μ events. The duplication of some variables calculated with the alternative combination of the muons in each event may also help with the discrimination. For example, apart from the transvers mass between the nominal W-lepton and the the H_T^{miss} , another transverse mass is also considered, using the H_T^{miss} and the same-sign muon from the nominal Higgs candidate, which turns out to be effective.

The sensitivity of this analysis depends largely on the resolution of the $m_{\mu\mu}$ peak, which is determined by muon momentum resolution, which in turn primarily depends on η of the muons as the detector condition differs in different η regions in CMS. It is important to divide events with different resolution into different categories, which leads

to an enhancement to the overall S/B . In the previous $H \rightarrow \mu\mu$ analysis [45, 18], the categorization is achieved by dividing events based on both the BDT output and the η value of the muons. While in this work, the resolution information is incorporated into the BDT, not as an input variable, but by weighting the signal events by $1/\sigma(m_{\mu\mu})$, which is the per-event experimental dimuon mass resolution, calculated from the p_T uncertainty of the muon tracks. In this way, the BDT output encapsulates both the kinematic information and the resolution information in its output, allowing for a categorization based on a single variable, achieving a better significance with fewer categories. The resolution is not used as an direct input to the BDT because its distribution is not very different between signal and background. The weights are only applied in the training on signal events, and not applied in the evaluation of the BDT score.

The BDT is trained with a collection of simulated samples from all eras in Run 2. The training is performed in the mass window of $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$. To make sure the BDT is not sensitive to the $m_{\mu\mu}$ value, signal samples with different Higgs mass assumptions, $m_H = 120, 125, 130 \text{ GeV}$, are all used as signals in the training. Signal events are only used if the candidate $\mu\mu$ pair truly originates from the Higgs decay, so that the BDT only picks the true kinematic signatures of the signal. No parent matching is required for backgrounds. To benefit from the maximal statistics in simulation while keeping sensitive to all kinematic features, events with $e + \mu\mu$ and $\mu + \mu\mu$ are used together in the training, but can be distinguished by the "number of electrons" as one of the input variables to the BDT. To increase the statistics of the non-prompt backgrounds in the simulated events, the lep-MVA cut is loosened from 0.4 to -0.4 for the training collection, and the non-prompt yields are scaled by a factor of 0.5 to account for the increased non-prompt lepton efficiency. For both the signal and background samples, half of the events is used for the training, while the other half is used for the testing.

The BDT output and its Receiver Operating Characteristic (ROC) curve are shown in Figure 6-1, in which the BDT performs the same on training and testing samples, indicating

no over-training. Distributions of the BDT input variables are shown in Figures 6-2.

Table 6-3. 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 is the lepton from the W decay, μ_{OS} (μ_{SS}) refers to the muons in the Higgs candidate which OS (SS) to the lepton.

Variable	Description
$p_T(\mu\mu_H)$	p_T of the Higgs candidate
$ \eta(\mu_1) $	η of the leading muon in the Higgs candidate
$ \eta(\mu_2) $	η of the trailing muon in the Higgs candidate
$\Delta R(\mu_{SS}, \mu_{OS})$	ΔR between the two muons in the Higgs candidate
$p_T(\ell)$	p_T of the extra lepton in the event
Number of electrons	Number of electrons in the event
$\Delta R(\ell, \mu\mu_H)$	ΔR between the extra lepton and the Higgs candidate
$\Delta\eta(\ell, \mu\mu_H)$	$\Delta\eta$ between the extra lepton and the Higgs candidate
$\Delta\eta(\ell, \mu_{SS})$	$\Delta\eta$ between the extra lepton and the SS muon
$\cos\theta^*(\ell, \mu_{SS})$	$\cos\theta^*$ between the extra lepton and the SS muon
$\Delta R(\ell, \mu_{OS})$	ΔR between the extra lepton and the OS muon
$\Delta\eta(\ell, \mu_{OS})$	$\Delta\eta$ between the extra lepton and the OS muon
$\cos\theta^*(\ell, \mu_{OS})$	$\cos\theta^*$ between the extra lepton and the OS muon
$M_T(\mu_{SS}, MHT)$	transverse mass of the H_T^{miss} and the SS muon
$M_T(\ell, MHT)$	transverse mass of the H_T^{miss} and the extra lepton
$ \Delta\phi(\ell, MHT) $	$ \Delta\phi $ between the H_T^{miss} and the extra lepton

6.2.2 BDT targeting $ZH \rightarrow \ell\ell + \mu\mu$ signal

After the event selection of the ZH category, the background is almost purely composed of $ZZ \rightarrow 4\ell$ and $ggZZ \rightarrow 4\ell$ processes. Other backgrounds, prompt or non-prompt, have negligible contribution in this channel. Both ZZ and ggZZ processes have the identical final states as the ZH signal. Apart from the dimuon mass, which is used in the last stage for signal extraction, the most distinct discrimination between the signal and the background lies in the helicity angles, between the leptons from the H (Z) decay, and between the H (Z_1) and the Z₂ bosons.

The input variables to the BDT are listed in Table 6-4 and shown in Figure 6-4, among which, $\cos\theta^*(\mu\mu_H, \ell\ell_Z)$, the helicity angle between the Higgs candidate and the Z candidate, is one of the most discriminating. In the ZZ background process, a propagator Z boson (Z_0) decays to two Z bosons (Z_1 and Z_2), which in turn decays to lepton pairs. Since

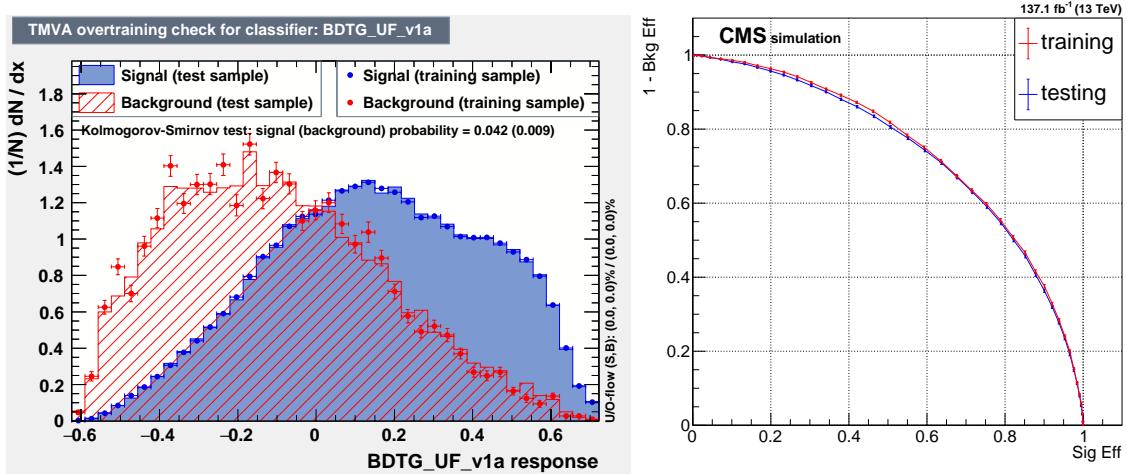


Figure 6-1. Plots of the performance of the $\text{WH} \rightarrow 3\ell$. On the left, the BDT output score, with signal in blue and background in red. On the right, the receiver operating characteristic (ROC) curve, with training sample in red and testing sample in blue. A slight over-training is observed in the region of low signal efficiency, due to the fluctuation in background. As will be shown in Fig. 6-5, the BDT does not sculpt the shape of $m_{\mu\mu}$.

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 Z_0 . Whereas in the ZH events, since the Higgs bosons are spin-0 particles, there is no preferred direction for the $Z_0 \rightarrow Z_1 H$ decay. **double check the distribution**. 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, where in the Z decay the muons prefer to align with the momentum of their parent and in the H decay they follow a flat distribution. 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 signal and 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 6.2.1, the training is performed with simulated 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, so only qqZH samples are used as signal events. Signal

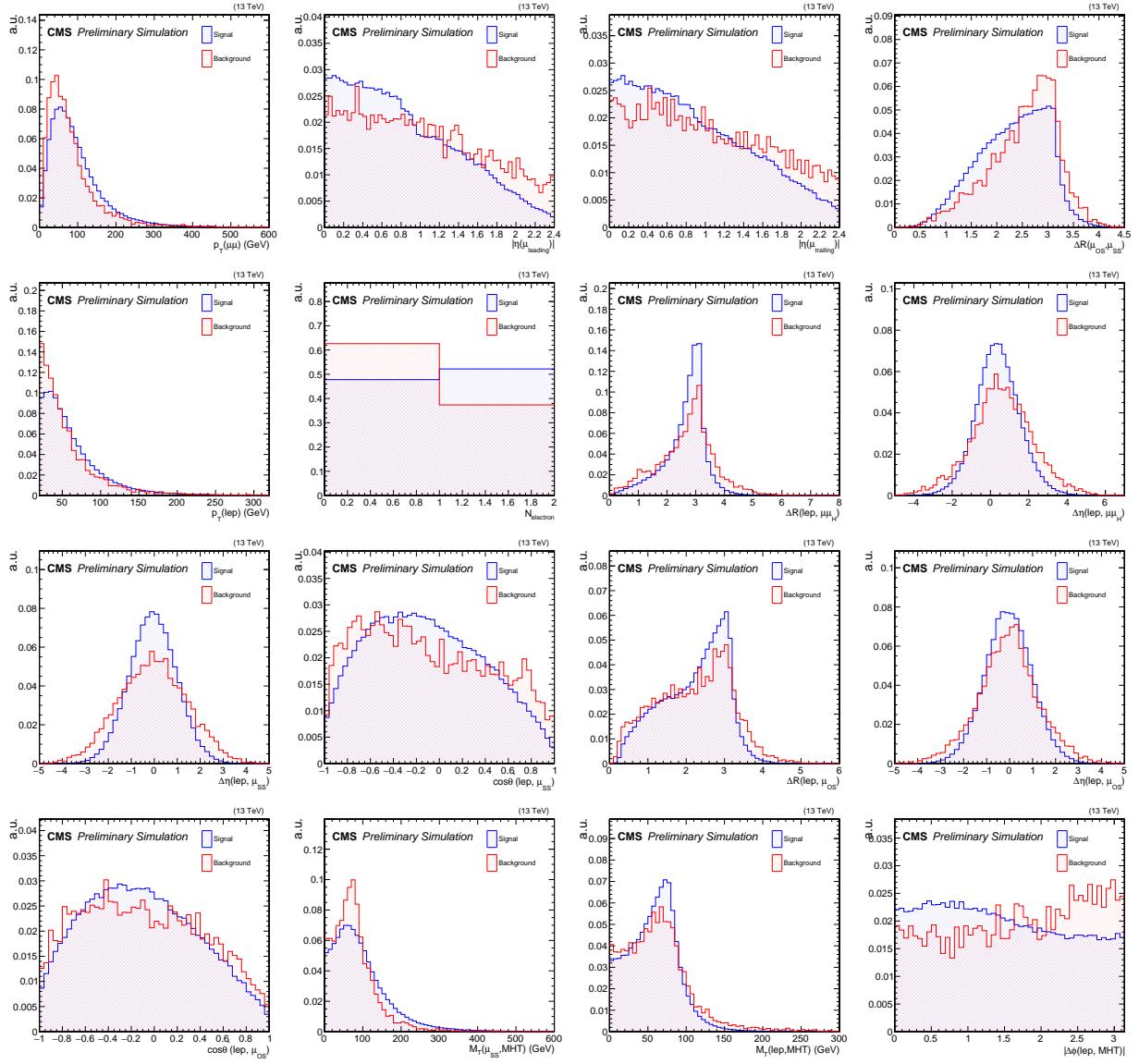


Figure 6-2. Input variables to the $\text{WH} \rightarrow 3\ell$ BDT, with signal in blue and background in red.

samples with different Higgs mass assumptions, $m_H = 120, 125, 130$ 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 flavor" as one of the input variables. To increase the statistics of training events, the lepMVA cut is loosened from 0.4 to -0.4. Even so, there is no non-prompt background component passing the loosened selection.

The BDT output and the ROC curve are shown in Figure 6-3, in which the BDT performs the same on training and testing samples, indicating no over-training. Distributions of the BDT input variables are shown in Figure 6-4.

Table 6-4. List of input variables used to train the signal-background separation BDT in the ZH category. In this table, $\mu\mu_H$ is the Higgs candidate, and $\ell\ell_Z$ is the Z candidate.

Variable	Description
$p_T(\mu\mu_H)$	p_T of the Higgs candidate
$ \eta(\mu\mu_H) $	$ \eta $ of the Higgs candidate
$ \Delta\phi(\mu\mu_H) $	$ \Delta\phi $ between the muons in the Higgs candidate
$M(\ell\ell_Z)$	invariant mass of the Z candidate
$p_T(\ell\ell_Z)$	p_T of the Z candidate
$ \eta(\ell\ell_Z) $	$ \eta $ of the Z candidate
$\Delta R(\ell\ell_Z)$	ΔR between the leptons in the Z candidate
lepton flavor	flavor of the Z candidate lepton pair
$\cos\theta^*(\mu\mu_H, \ell\ell_Z)$	cosine helicity angle between the Higgs and the Z candidates
$\Delta\eta(\mu\mu_H, \ell\ell_Z)$	$\Delta\eta$ between the Higgs and the Z candidates

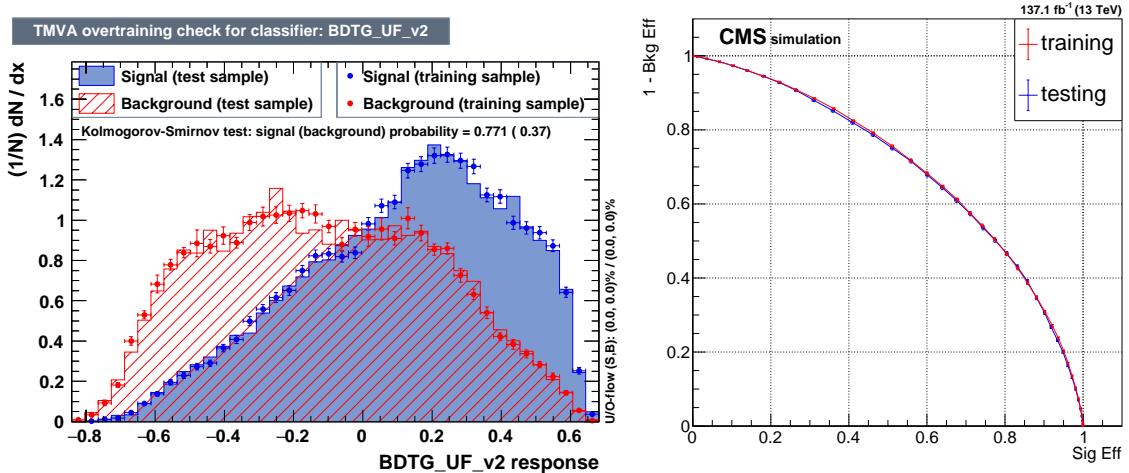


Figure 6-3. Plots of the performance of the $ZH \rightarrow 4\ell$ BDT. On the left, the BDT output score, with signal in blue and background in red. On the right, the receiver operating characteristic (ROC) curve, with training in red and testing in blue.

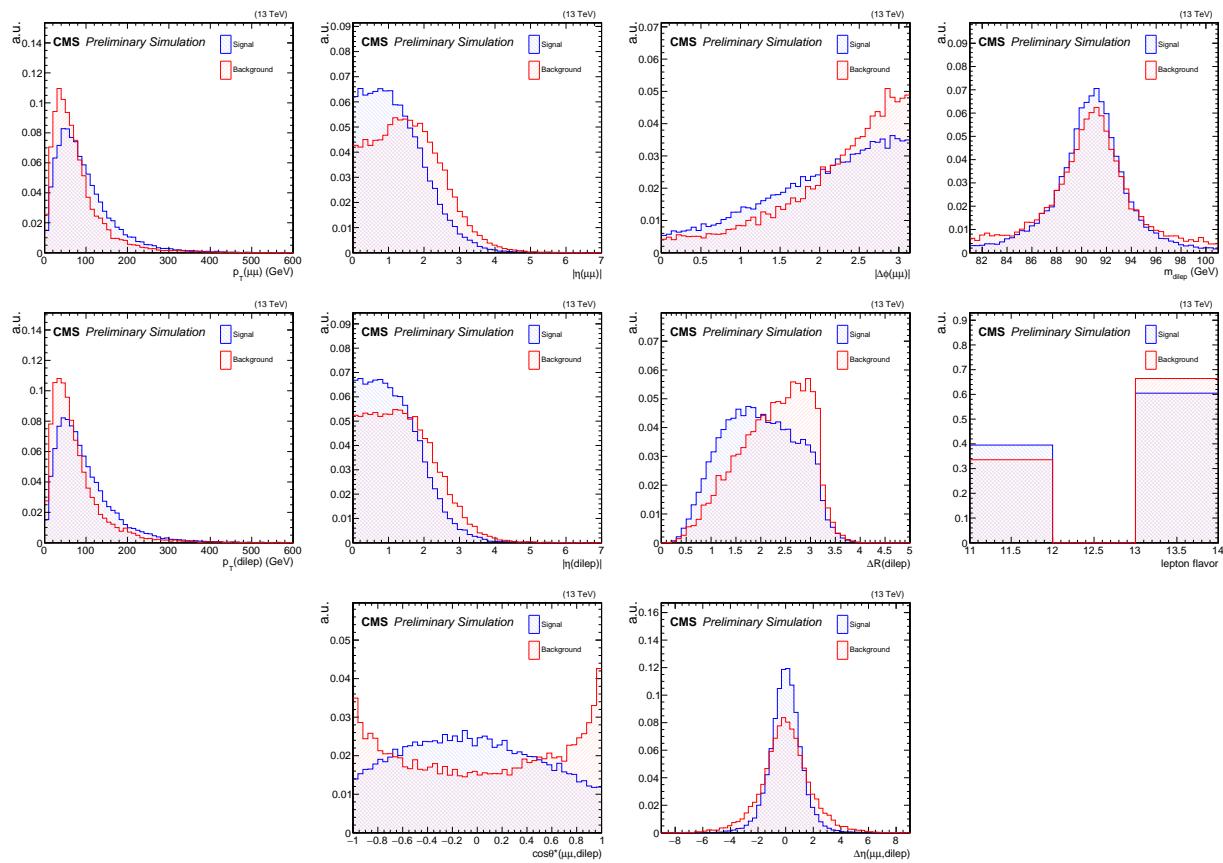


Figure 6-4. Input variables to the $ZH \rightarrow 4\ell$ BDT, with signal in blue and background in red.

6.2.3 Validation of the BDTs

As discussed in 3.2, the strategy of this analysis is to divide events into sub-categories with different S/B , and consequently maximize the overall sensitivity. The signal extraction is performed by fitting the $m_{\mu\mu}$ spectrum, therefore it is crucial that any selection cut applied to the BDT score should not sculpt the $m_{\mu\mu}$ shape. Two checks are performed for this purpose:

- The $m_{\mu\mu}$ shape of the background is compared between events in different BDT quantiles, shown as the left plots of Figures 6-5 and 6-6.
- The BDT output is compared between several signal samples with different m_H assumptions, shown as the right plots of Figures 6-5 and 6-6.

From these plots, no sign of correlation between BDT and $m_{\mu\mu}$ is seen. Therefore the WH and ZH BDTs can be used for categorization without introducing sculpting of the $m_{\mu\mu}$ shape.

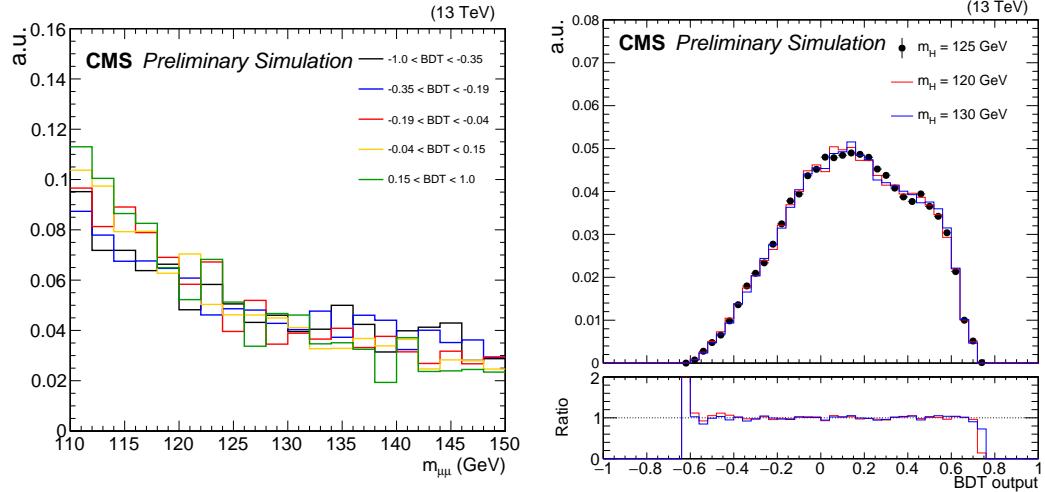


Figure 6-5. For the WH BDT, the distribution of the dimuon mass shape in the background for five different BDT quantile (left), and the distribution of the BDT output for three different signal mass assumptions (right).

Furthermore, it is also important to make sure the BDT would perform the same way on data as they do on the simulation. In order to do this, the inputs and output of the BDTs are plotted comparing between data and the simulation. Figure 6-7 shows the

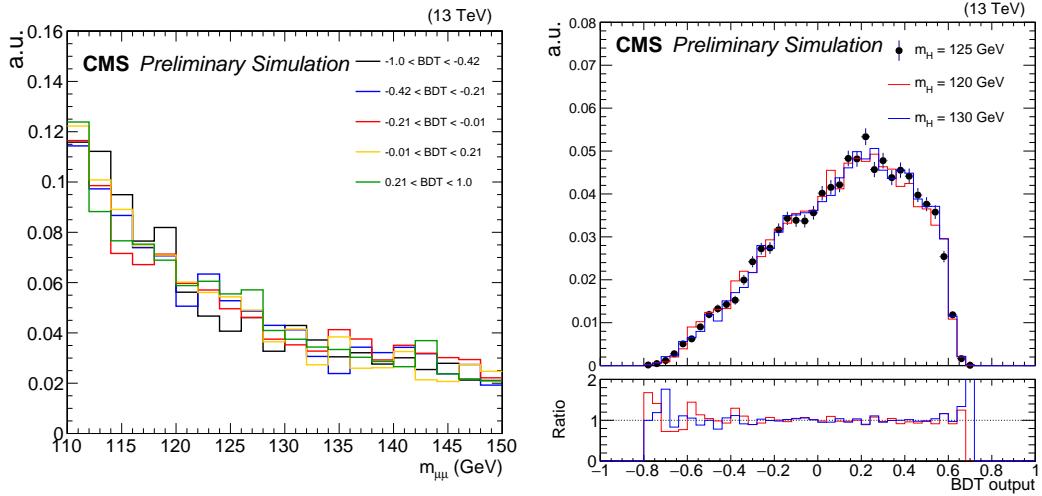


Figure 6-6. For the ZH BDT, the distribution of the dimuon mass shape in the background for five different BDT quantile (left), and the distribution of the BDT output for three different signal mass assumptions (right).

output of the WH BDT and the ZH BDT, and Figure 6-8 and 6-9 show the input variables to the WH and ZH BDTs respectively. Overall, data and simulation agree with each other within the uncertainties for the BDT outputs and most of the inputs. Some fluctuations are seen in data, especially in the ZH category, as the total number of events is small. These fluctuations are expected within the statistical uncertainty and do not indicate any systematic disagreement.

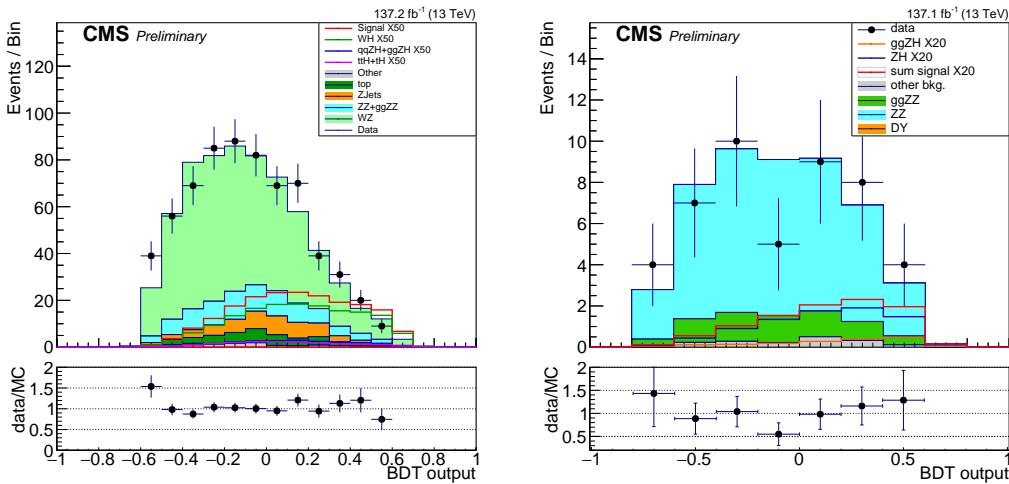


Figure 6-7. The WH BDT output (left) and the ZH BDT output (right) in full Run 2 in the signal region $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$.

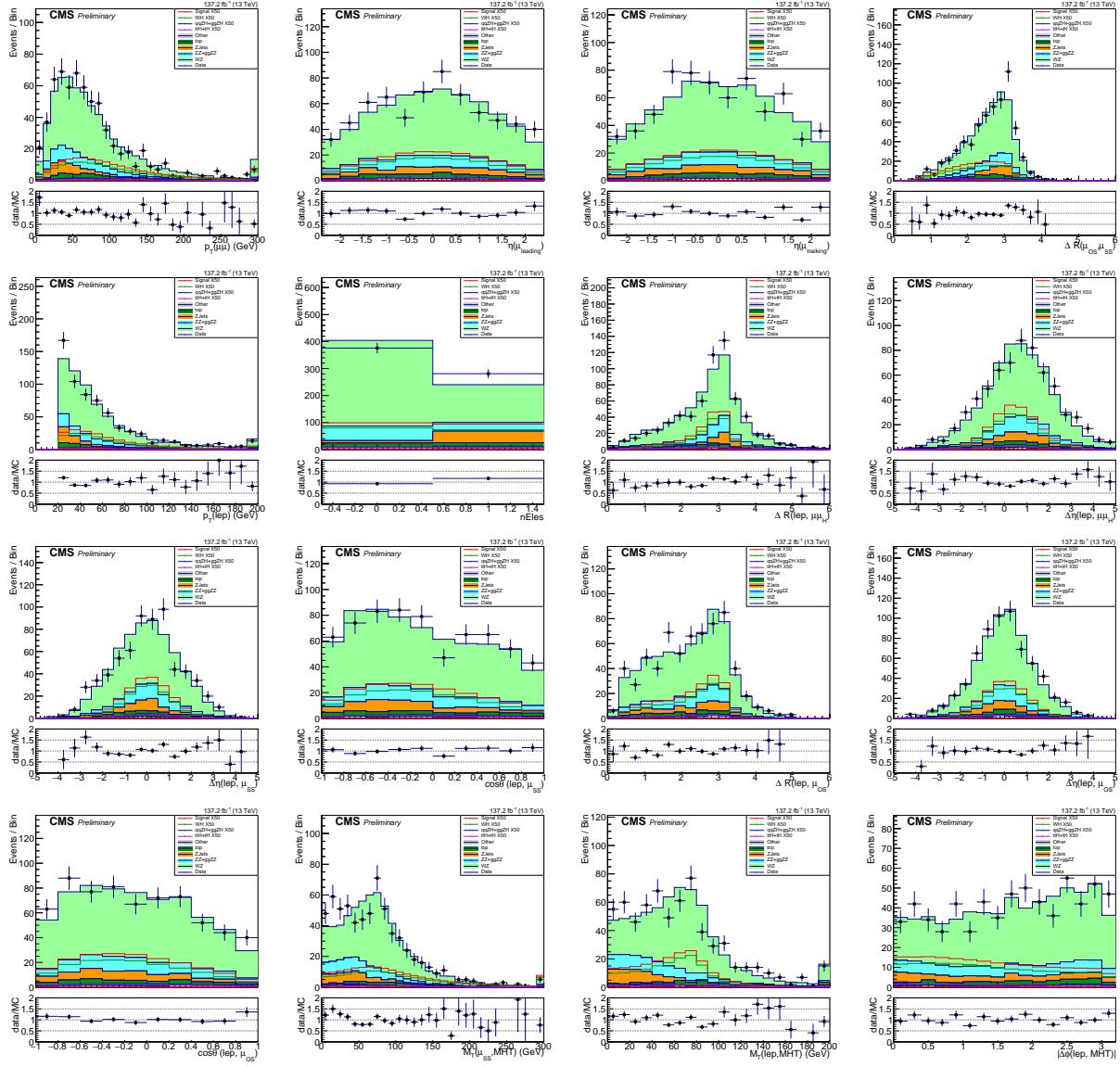


Figure 6-8. Input variables to the WH BDT in full Run 2 in the signal region $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$.

An arguable disagreement is seen in the leftmost region in one of the inputs to the WH BDT, $M_T(\mu_{SS}, \text{MHT})$, shown as the second plot in the bottom row of Figure 6-8, also put separately in Figure 6-10. To understand if this disagreement would translate into a mismodeling of the WH BDT, the BDT output is plotted for both signal and background in different $M_T(\mu_{SS}, \text{MHT})$ bins, shown as the right plot in Figure 6-10. In this plot, for both signal and background, the BDT profile is almost the same for events with $M_T(\mu_{SS}, \text{MHT})$

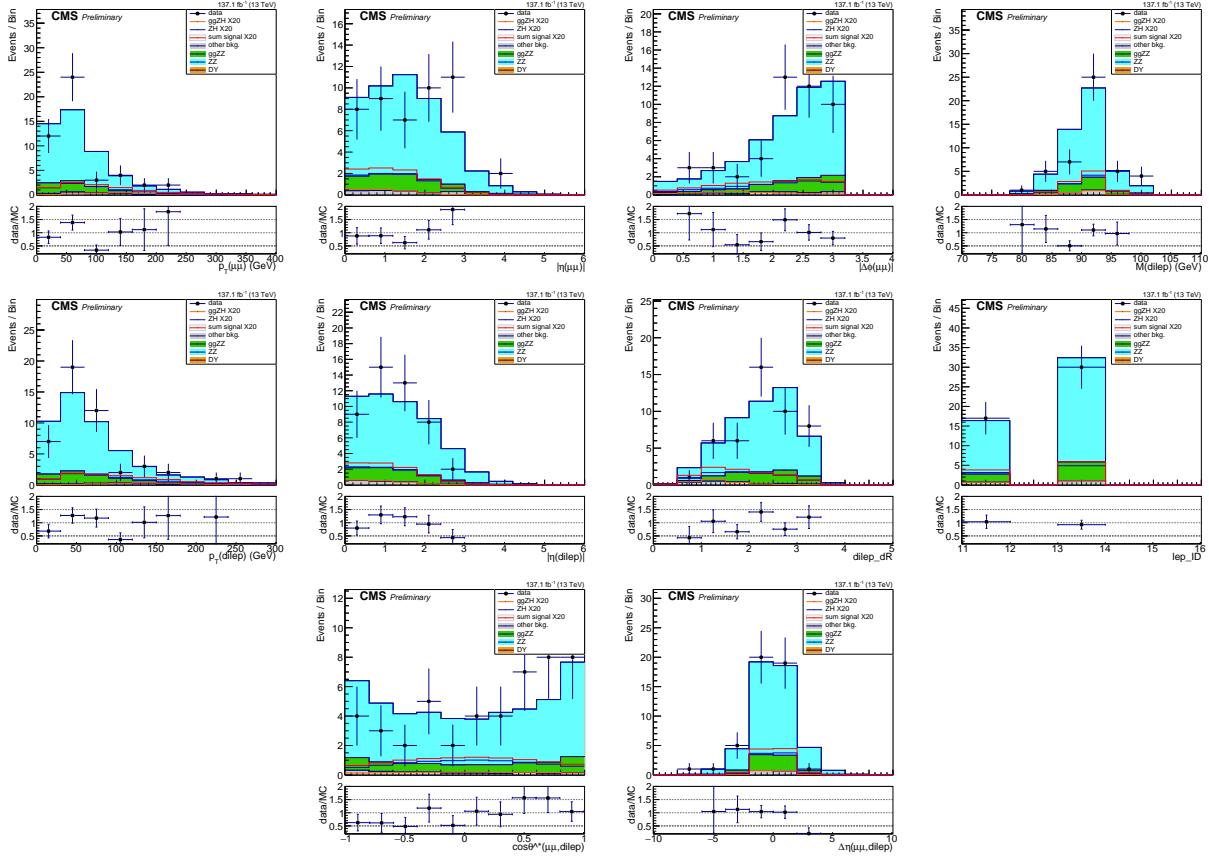


Figure 6-9. Input variables to the ZH BDT in full Run 2 in the signal region $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$.

$< 40 \text{ GeV}$ and events with $40 < M_T(\mu_{SS}, \text{MHT}) < 80 \text{ GeV}$, while it is different between events with $M_T(\mu_{SS}, \text{MHT}) < 80 \text{ GeV}$ from events with $M_T(\mu_{SS}, \text{MHT}) > 80 \text{ GeV}$. The BDT is sensitive to whether the $M_T(\mu_{SS}, \text{MHT})$ is greater or smaller than 80 GeV, but does not further distinguish events if the $M_T(\mu_{SS}, \text{MHT})$ is less than 80 GeV. Once the bins below 80 GeV are merged in the left plot of Figure 6-10, there is no significant disagreement, therefore it should not cause any mismodeling of the BDT.

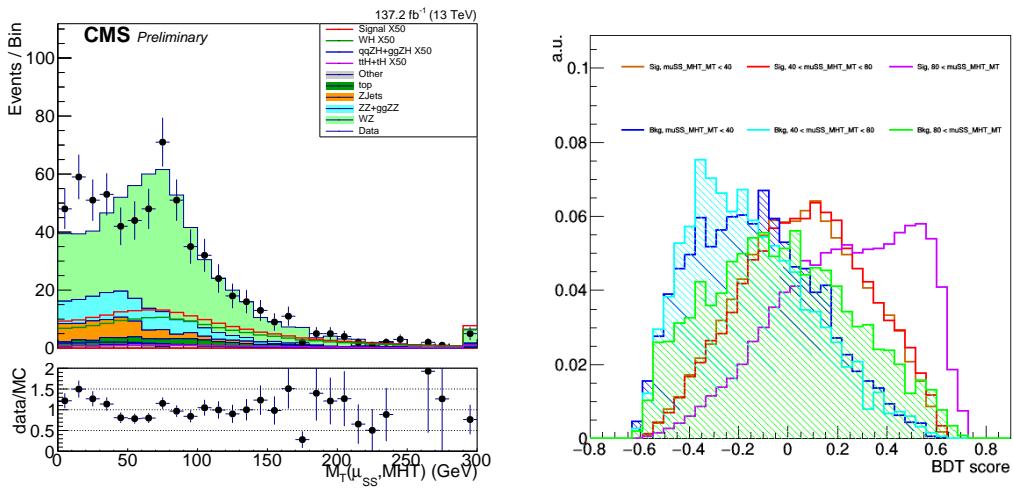


Figure 6-10. The input variable $M_T(\mu_{SS}, \text{MHT})$ to the WH BDT (left), and the BDT output for signal and background in different $M_T(\mu_{SS}, \text{MHT})$ bins (right). A mild disagreement is seen between the simulation and data in the low bins of $M_T(\mu_{SS}, \text{MHT})$, while the BDT is not sensitive to the $M_T(\mu_{SS}, \text{MHT})$ values in that region.

6.3 Event Categorization

To optimize the overall sensitivity of the VH analyses, the WH and the ZH phase-spaces are divided into several sub-categories with different S/B ratios, based on the BDT discriminants described in Section 6.2.

To achieve the maximal sensitivity with a reasonable number of sub-categories, 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 the significance of the resulting sub-categories is calculated as the figure of merit. The figure of merit is defined as the S/\sqrt{B} in each sub-category summed in quadrature, where S and B represent the expected signal and background yields within the FWHM of the signal peak in each sub-category. In addition, to ensure that there are enough events in each sub-category to perform a shape analysis, all sub-categories have to meet a minimal total event yield requirement during the BDT scanning process.

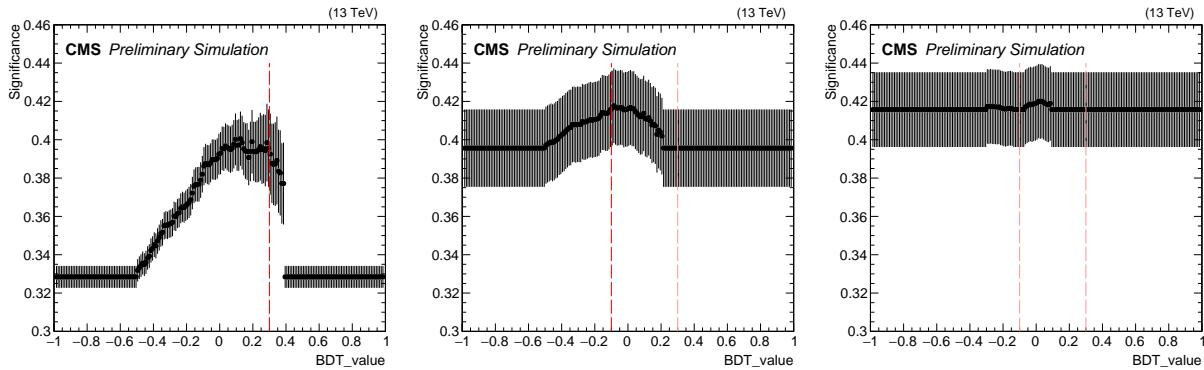


Figure 6-11. Scans for the first (left), second (middle) and a potential third (right) BDT cut in the WH channel. The first BDT cut is chosen at 0.3. The second BDT cut is chosen at -0.1. A third BDT cut is not necessary.

Figure 6-11 shows the iterations performed on the WH BDT. The minimum number of events in each category is set to be 30. In the first scan, the overall significance maximizes around $0 \sim 0.3$, and the cut is chosen at 0.3 so that there are enough events on its left side for a second cut. In the second scan, the overall significance maximizes around $-0.1 \sim 0.05$, and the position of the second cut can be any value in this range. To help decide the second cut, several third scans are performed under different assumptions of the second

cut, all showing negligible changes of the overall significance (similar to the right plot in Figure 6-11). Therefore, there is no need for a third cut, and the choice for the second cut can be somewhat arbitrary. The second cut is decided at -0.1 so that there are a good number of events in the middle sub-category to ensure a stable shape analysis. As a result, two BDT boundaries are set, dividing the WH phase-space into 3 sub-categories, BDT within [-1.0, -0.1], BDT within [-0.1, 0.3] and BDT within [0.3, 1.0].

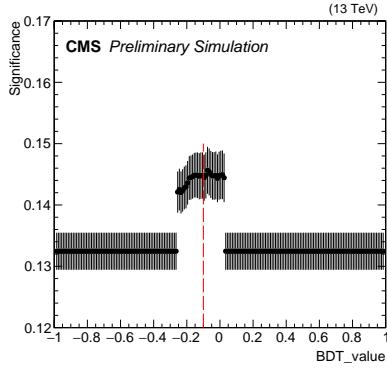


Figure 6-12. Scans for the BDT cut in the ZH channel. The BDT cut is chosen at -0.1.

Similarly, Figure 6-12 shows the scan performed on the ZH BDT. The minimum number of events in each category is set to be 16, as the total number of events in the ZH category is less than 50. In the BDT scan, the overall significance maximizes around $-0.15 \sim 0.05$. The BDT cut is chosen at -0.1, dividing the ZH events into two roughly equal halves. A second cut is not needed as the number of events is not enough for a further division. The resulting 2 ZH sub-categories are, BDT within [-1.0, -0.1], and BDT within [-0.1, 1.0].

6.4 Signal and background modeling

The extraction of signal is performed by fitting analytic functions to the $m_{\mu\mu}$ spectrum in each sub-category. Different functions are used to model the expected signal and background shapes: a sharp signal peak near 125 GeV, and a smooth falling background shape in $110 \text{ GeV} < m_{\mu\mu} < 150 \text{ GeV}$. Functions are first tested on the simulated samples to make sure they perform well in describing the shapes. The parameters of the signal function are constrained, with systematic uncertainties described in Section 6.5.1, to the best fit values

to simulation as the expectation of the SM signal. The parameters of the background function are allowed to float freely, so that no prior assumption on the background is imposed, and the background prediction relies completely on data. The final evaluation of the signal strength is achieved by fitting the signal + background functions to data, where the normalizations of both the signal function and the background function are allowed to float freely. The normalization of the signal, in particular, is called the signal strength modifier and represents the signal strength relative to the SM prediction.

Function modeling of signal and background are described in Sections [6.4.1](#) and [6.4.2](#) respectively.

6.4.1 Signal modeling

In all sub-categories, signals are modeled independently by different production modes, with the contributions from three years (2016, 2017, 2018) summed together. In particular, qqZH and ggZH signals are modeled separately as there are no other signal component in the ZH category. Each of the components is modeled with a Double-sided Crystal Ball function (DCB), as described in Equation [6-1](#). In all DCB functions, the parameters n_L and n_R are fixed to 2.0, since they only affect the shape in tails and can take values in a large range without changing the quality of the fit by much. Other parameters are allowed to float freely.

$$\text{DCB}(m_{\mu\mu}) = \begin{cases} e^{-(m_{\mu\mu}-s)^2/(2\sigma^2)} & -\alpha_L < (m_{\mu\mu}-s)/\sigma < \alpha_R \\ (\frac{n_L}{|\alpha_L|})^{n_L} \times e^{-\alpha_L^2/2} \times (\frac{n_L}{|\alpha_L|} - |\alpha_L| - (m_{\mu\mu}-s)/\sigma)^{-n_L} & (m_{\mu\mu}-s)/\sigma \leq -\alpha_L \\ (\frac{n_R}{|\alpha_R|})^{n_R} \times e^{-\alpha_R^2/2} \times (\frac{n_R}{|\alpha_R|} - |\alpha_R| + (m_{\mu\mu}-s)/\sigma)^{-n_R} & (m_{\mu\mu}-s)/\sigma \geq \alpha_R \end{cases} \quad (6-1)$$

Examples of signal modeling are shown in Figures [6-13](#) and [6-14](#). Please note that the plots shown are the signals in the inclusive WH and ZH categories. The actual models used in each sub-category are slightly different. ggH, VBF and b̄bH have negligible contributions to the WH category and are not considered. Similarly, in the ZH category only

qqZH and ggZH are considered since all other contributions are negligible.

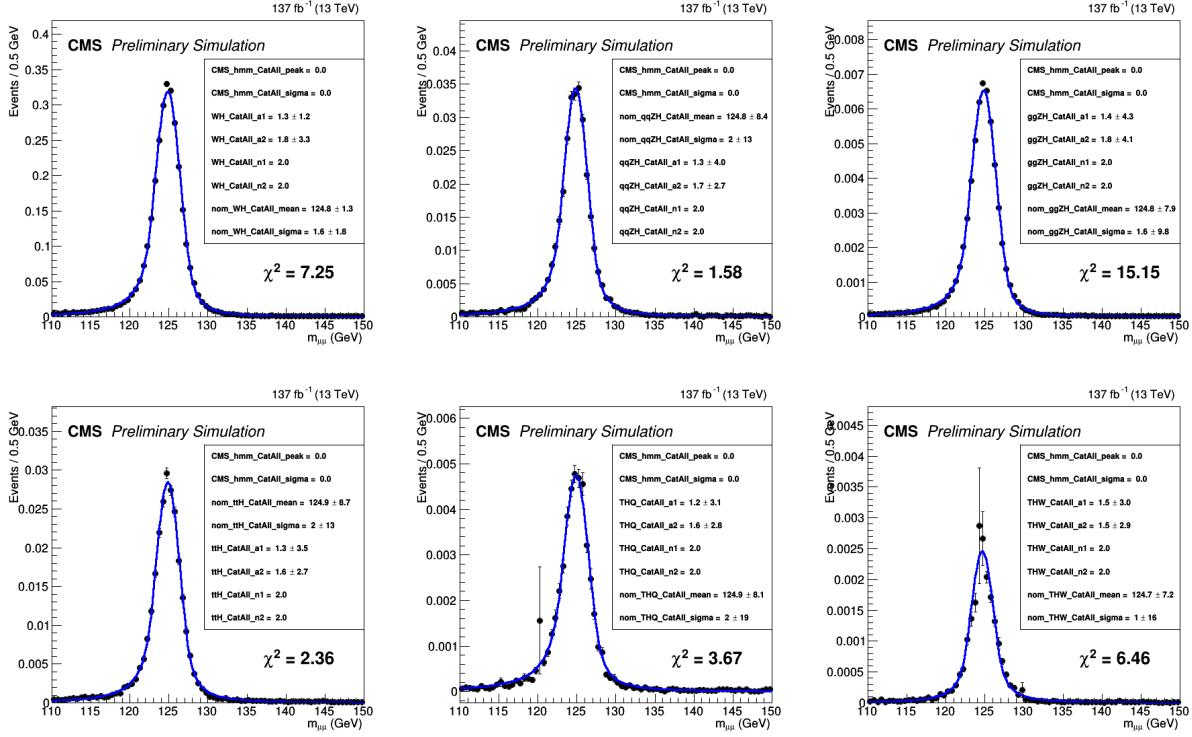


Figure 6-13. The signal modeling in the $WH \rightarrow \ell + \mu\mu$ inclusive category. Considered signal modes are WH (top left), qqZH (top middle), ggZH (top right), ttH (bottom left), THQ (bottom middle), and THW (bottom right).

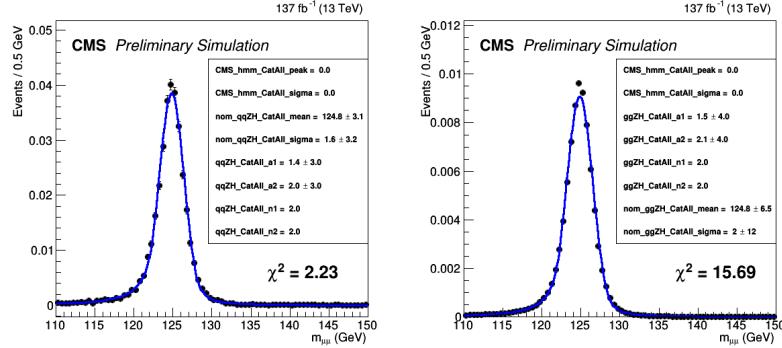


Figure 6-14. The signal modelin in the $ZH \rightarrow ll + \mu\mu$ inclusive category. Considered signals modes are qqZH (left) and ggZH (right).

6.4.2 Background modeling

As discussed in Section 6.1, the main background in WH (ZH) category is the WZ (ZZ) process, a fraction of which consists wrong pairing of the muons. The $m_{\mu\mu}$ spectrum of the correctly paired WZ (ZZ) events follows a Breit-Wigner tail of the Z boson, while the spectrum of the wrongly paired WZ (ZZ) events is rather flat. Overall the background shapes in WH and ZH categories are smoothly falling, and Breit-Wigner like to some extent. Therefore, a group of different functional forms are considered as candidates for the background modeling. Some of them are physics-inspired, meaning that they are modified from the form of the Breit-Wigner function, the others are agnostic, which take the form of some general functional bases. The physics-inspired function candidates include the *BWZ* function (Equation 6-2), which is a Breit-Wigner core times an exponential term , the *BWZRedux* (Equation 6-3), which is a Breit-Wigner core times a exponential term with more degrees of freedom (DOFs), the *BWZGamma* (Equation 6-4), which is a linear combination of the *BWZ* function and an exponential function, and the *BWZ × Bernstein* (Equation 6-5), which is the *BWZ* function times a Bernstein polynomial.

$$\text{BWZ}(m_{\mu\mu}) = \frac{\Gamma_Z \cdot e^{a \cdot m_{\mu\mu}}}{(m_{\mu\mu} - m_Z)^2 + (\Gamma_Z/2)^2} \quad (6-2)$$

$$\text{BWZRedux}(m_{\mu\mu}) = \frac{\Gamma_Z \cdot e^{a \cdot m_{\mu\mu} + b \cdot m_{\mu\mu}^2}}{(m_{\mu\mu} - m_Z)^c + (\Gamma_Z/2)^c} \quad (6-3)$$

$$\text{BWZGamma}(m_{\mu\mu}) = f \cdot \frac{\Gamma_Z \cdot e^{a \cdot m_{\mu\mu}}}{(m_{\mu\mu} - m_Z)^2 + (\Gamma_Z/2)^2} + (1-f) \cdot \frac{e^{a \cdot m_{\mu\mu}}}{m_{\mu\mu}^2} \quad (6-4)$$

$$\text{BWZ} \times \text{Bernstein}(m_{\mu\mu}) = \frac{\Gamma_Z \cdot e^{a \cdot m_{\mu\mu}}}{(m_{\mu\mu} - m_Z)^2 + (\Gamma_Z/2)^2} \times \text{Bern}_n(m_{\mu\mu}) \quad (6-5)$$

The agnostic function candidates include the Bernstein polynomials (Equation 6-6), a series of exponential functions (Equation 6-7), and a series of power functions (Equation 6-8). In the actual fits, given the low statistics in the VH sub-categories, the sum

of exponential or power functions are usually reduced to a single exponential or power function plus a constant (Equation 6-9 and 6-10).

$$\text{Bernstein}(m_{\mu\mu}) = \sum_i^n a_i \cdot \binom{n}{i} m_{\mu\mu}^i (1 - m_{\mu\mu})^{n-i} \quad (6-6)$$

$$\text{S-exponential}(m_{\mu\mu}) = \sum_i^n a_i \cdot e^{b_i \cdot m_{\mu\mu}} \quad (6-7)$$

$$\text{S-power-law}(m_{\mu\mu}) = \sum_i^n a_i \cdot m_{\mu\mu}^{b_i} \quad (6-8)$$

$$\text{Exponential+constant}(m_{\mu\mu}) = f + (1 - f) \times e^{a \cdot m_{\mu\mu}} \quad (6-9)$$

$$\text{Power-law+constant}(m_{\mu\mu}) = f + (1 - f) \times m_{\mu\mu}^a \quad (6-10)$$

In each sub-category, the function candidates are fit to the $m_{\mu\mu}$ shape in the range of $110 < m_{\mu\mu} < 150$ GeV, with events blinded in the signal region $120 < m_{\mu\mu} < 130$ GeV, so the functions are not aware of the existence of the signal. Because the limited number of events in VH sub-categories, the distribution of data is subject to large fluctuations. The $m_{\mu\mu}$ shape of data reflects both the underlying physics shape, as well as the specific features from the fluctuation of this particular dataset. It is important to make sure the modeling of background does not over-fit these specific features. On the other hand, as shown in Section 6.1 and 6.2, the simulated samples are known to provide a good modeling of data, and the $m_{\mu\mu}$ shape from the simulation can be assumed to be a good representation of the true physics shape. Therefore, the simulation can be used to study the performance of the background function candidates to learn how they would model generic expected physics shapes, and the data is treated as a particular realization of these physics distributions. The fit to simulation takes the $m_{\mu\mu}$ shape of the simulation, but assumes the statistical error in each bin as the Poisson error of the expected number of events in that bin rather than

the number of simulated events in the sample. If a function candidate provides a good fit to the simulation, it is then tested on the real dataset, to make sure the fit does not break down because of the fluctuation. If the function gives consistently good fit performances on the simulation and data, it is considered as a good candidate. It is worth a remark that the fit to data does not assume any parameter information from the fit to simulation, so the simulation is only used to study the performance of the background functions, but not used to constrain the specific shapes.

All the functional forms listed above can be used with different DOFs. For the physics-inspired funtions, the m_Z and Γ_Z can either be fixed at the nomial value for the Z boson or allowed to float freely, while for the agnostic functions, the order of the serie can be adjusted. To find the right DOFs, each functional form is tested with different setups, and the optimal DOFs is determined following the idea of the likelihood ratio test. A standard likelihood ratio test compares the likelihood ratio between the fits with n and $n+1$ DOFs, usually calculated as $2(\text{Log}\mathcal{L}_{n+1} - \text{Log}\mathcal{L}_n)$, where the $\text{Log}\mathcal{L}_n$ is the likelihood of the fit with n DOFs. This quantity should follow the χ^2_1 distribution, the chi-square distribution with one degree of freedom, whose p-value is then used to decide whether adding one more DOF in the fit leads to a significantly better fit quality.

In the practice of background fitting in the VH sub-categories, which all have low expected number of events, it turns out in most cases two DOFs are enough, one for overall normalization and one for shape variation. In some sub-categories with very low statistics, even functions without any shape DOF give good performances, namely, a function with all its shape parameters fixed at the best fit values to the simulation can be a good fit to data. These fixed shapes are included as some candidates along side with their freely floating versions. On the other hand, when the shape parameters are allowed to float, agnostic functions with low DOFs do not always fit well as they lack enough flexibility. For example, an order-1 Bernstein polynomial (2 shape DOFs) is just a straight line and is obviously not the true $m_{\mu\mu}$ shape, and the fits with a single exponential or power function are not stable

and sometimes do not converge. To mitigate these behaviors, a BWZ \times order-1 Bernstein (2 shape DOF in total), in which the BWZ part is fixed with the nominal Z boson shape, is used instead of the plain Bernstein, and a single exponential (power) function plus a free constant (2 shape DOF in total) is used instead of the plain exponential (power) function. Overall, the good function candidates include fixed forms (1 normalization DOF + 0 shape DOF) and floating forms (1 normalization DOF + 1 ~2 shape DOFs).

The final choices of the background function in each sub-category is decided based on the bias it may have against other possibilities, described in detail in Section 6.5.3. If several functions pass the bias requirement, the function with the fewest DOF is chosen, as it leads to the highest significance in statistical analysis.

6.5 Systematic uncertainties

A crucial task in the statistical analysis is to evaluate all the systematic uncertainties that affect the signal and background estimation. In this analysis, both the signal and background are described by analytic functions, and the statistical analysis is performed based on the fits of them. All sources of systematic uncertainties are therefore translated into the variations of the parameters of signal and background functions.

Several sources of signal systematic uncertainties are considered, divided into two types, the *shape* and the *rate* uncertainties. The *shape* uncertainties, described in Section 6.5.1, account for the factors affecting the expected shape of the signal peak, while the *rate* uncertainties, described in Section 6.5.2, are those affecting the expected signal yield.

A different approach is taken to evaluate the systematic uncertainty in background. The background estimation always takes the best fit to data and does not rely on simulation, therefore none of the theoretical or experimental uncertainties considered for the signal needs to be considered for the background. However, by fitting the background shape with an analytical function, a potential bias could be introduced between the chosen background model and the underlying real distribution. A bias between the background

estimation and the true background appearing at the position of the signal is essentially a spurious signal. This bias has to be small so that it does not impact the validity of the signal strength evaluation. The study to evaluate this potential bias is described in details in Section 6.5.3.

6.5.1 Signal shape uncertainties

For all Higgs boson production modes, the expected $m_{\mu\mu}$ signal shape is primarily affected by the uncertainties in muon energy scale and resolution, in other words the mean and sigma values in the DCB fits of the signal peak. As described in Chapter 5, the *Rochester correction* is implemented to correct for differences in both scale and resolution between data and simulated events, while the *FSR recovery* and *GeoFit correction* are not expected to introduce new differences between data and simulation. In the meantime, Section 5.4 shows that the simulation of the DY peak agrees with the data upto a per-mille level in scale and a percent level in resolution. The shape uncertainties can be estimated accordingly.

The muon energy scale shape uncertainty is estimated to be 0.1% of the mean value of the $m_{\mu\mu}$ peak, and the muon energy resolution uncertainty is conservatively estimated to be 10% of the resolution of the $m_{\mu\mu}$ peak. The effect of the scale uncertainty is an overall shift of the $m_{\mu\mu}$ peak to higher or lower mass value, while the effect of the resolution uncertainty is a stretching or squeezing of the width of the $m_{\mu\mu}$ peak. Both uncertainties are modeled as a Gaussian constrained nuisance parameter which is correlated across different production modes but uncorrelated between different sub-categories.

6.5.2 Signal rate uncertainties

The rate uncertainties are the ones that affect the signal yield in each sub-category, and may come from various sources. Some of them affect the overall prediction of the signal and act as a factor on the overall normalization of the signal. The normalization uncertainties include the theoretical uncertainties on the cross sections of signal productions, and theoretical uncertainties on the $\mathcal{B}(H \rightarrow \mu\mu)$, as well as the uncertainties on

the CMS luminosity measurement. Other uncertainties tweak the event kinematics and affect the acceptance of signals in each sub-category. The acceptance uncertainties include the uncertainties on all the event weights in the simulation, the uncertainties from all the efficiency scale factors applied in the analysis, and the uncertainties from all the physics object calibrations and corrections.

The impacts from theoretical uncertainties are shown in Table 6-5. The uncertainty on the $\mathcal{B}(H \rightarrow \mu\mu)$ is $\pm 1.23\%$, independent from the production modes. The luminosity uncertainty for each year is set following the official recommendation of CMS, which is 2.5%, 2.3% and 2.5% for 2016, 2017 and 2018 respectively. Since the signals are modeled summing all years, the luminosity uncertainty in each year reflected in the overall signal yield is 0.7%, 0.7% and 1.1%, for 2016, 2017, and 2018.

Table 6-5. Normalization uncertainties on the Higgs boson production cross sections for various modes at $\sqrt{s} = 13\text{TeV}$.

Process	Perturbative Order	+QCD scale unc. (%)	-QCD scale unc. (%)	+ (PDF + α_s) unc. (%)	- (PDF + α_s) unc. (%)
WH	NNLO (QCD)	+0.5	-0.7	+1.9	-1.9
	NLO (EWK)				
qqZH	NNLO (QCD)	+0.5	-0.6	+1.9	-1.9
	NLO (EWK)				
ggZH	NLO (QCD)	+25.1	-18.9	+2.4	-2.4
$t\bar{t}H$	NLO (QCD)	+5.8	-9.2	+3.6	-3.6
	NLO (EWK)				
tHq	NLO (QCD)	+6.5	-14.9	+3.7	-3.7
tHW	NLO (QCD)	+4.9	-6.7	+6.3	-6.3

The impacts from the pileup re-weight and ECAL L1 trigger prefiring re-weight are shown in Table 6-6. The acceptance impacts from the muon energy scale corrections are shown in Table 6-7. The impacts from the muon and electron ID scale factors (the lepMVA scale factor) are shown in Table 6-8. The impacts from the B-jet ID scale factors (for the B-jet vetoing) are shown in Table 6-9. The impacts from the jet energy calibrations are shown in Table 6-10.

Table 6-6. Uncertainties on different signal components in the WH and ZH channels related to pileup re-weight and L1 prefiring re-weight. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected.

Uncertainty	Category	WH	qqZH	ggZH	ttH	THQ	THW
pileup 2016 (%)	WH cat0	+0.8/-0.7	+0.8/-0.7	+0.6/-0.5	+0.9/-0.7	+1.0/-0.9	+0.4/-0.4
	WH cat1	+0.8/-0.7	+0.6/-0.5	+0.6/-0.5	+0.6/-0.5	+1.0/-0.9	+0.7/-0.6
	WH cat2	+0.6/-0.5	+0.4/-0.3	+0.6/-0.4	+0.7/-0.6	+0.2/-0.3	+0.5/-0.4
	ZH cat0	-	+0.7/-0.7	+0.8/-0.7	-	-	-
	ZH cat1	-	+0.8/-0.7	+0.7/-0.6	-	-	-
pileup 2017 (%)	WH cat0	+0.6/-0.5	+0.2/-0.2	+0.3/-0.3	+0.3/-0.2	+0.3/-0.5	+0.6/-0.7
	WH cat1	+0.4/-0.4	+0.4/-0.4	+0.3/-0.3	+0.3/-0.3	+0.4/-0.5	+0.3/-0.4
	WH cat2	+0.5/-0.5	+0.5/-0.3	+0.3/-0.3	+0.6/-0.6	+0.5/-0.5	+0.4/-0.2
	ZH cat0	-	+0.4/-0.4	+0.4/-0.5	-	-	-
	ZH cat1	-	+0.3/-0.4	+0.4/-0.4	-	-	-
pileup 2018 (%)	WH cat0	+0.6/-0.6	+0.4/-0.4	+0.5/-0.5	+0.6/-0.6	+0.5/-0.5	+0.5/-0.5
	WH cat1	+0.5/-0.5	+0.3/-0.3	+0.4/-0.4	+0.5/-0.5	+0.4/-0.4	+0.8/-0.8
	WH cat2	+0.4/-0.4	+0.2/-0.3	+0.4/-0.4	+0.4/-0.3	+0.5/-0.5	+0.8/-0.8
	ZH cat0	-	+0.6/-0.6	+0.5/-0.5	-	-	-
	ZH cat1	-	+0.6/-0.6	+0.5/-0.5	-	-	-
prefire 2016 (%)	WH cat0	+0.1/-0.1	+0.1/-0.1	+0.2/-0.2	+0.2/-0.2	+0.2/-0.2	+0.2/-0.2
	WH cat1	+0.1/-0.1	+0.1/-0.1	+0.2/-0.2	+0.2/-0.2	+0.2/-0.2	+0.1/-0.1
	WH cat2	-	+0.1/-0.1	+0.1/-0.1	+0.1/-0.1	+0.2/-0.2	+0.1/-0.1
	ZH cat0	-	+0.1/-0.1	+0.1/-0.1	-	-	-
	ZH cat1	-	+0.1/-0.1	+0.1/-0.1	-	-	-
prefire 2017 (%)	WH cat0	+0.2/-0.2	+0.3/-0.3	+0.3/-0.3	+0.4/-0.4	+0.3/-0.3	+0.2/-0.3
	WH cat1	+0.1/-0.1	+0.3/-0.3	+0.3/-0.3	+0.3/-0.3	+0.3/-0.4	+0.2/-0.2
	WH cat2	+0.1/-0.1	+0.2/-0.2	+0.2/-0.2	+0.2/-0.2	+0.2/-0.2	+0.1/-0.1
	ZH cat0	-	+0.2/-0.2	+0.3/-0.3	-	-	-
	ZH cat1	-	+0.1/-0.1	+0.2/-0.2	-	-	-

Table 6-7. Uncertainties on different signal components in the WH and ZH channels related to the muon energy scale. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected.

Uncertainty (%)	WH	qqZH	ggZH	ttH	THQ	THW
WH cat0	-	+0.0/+0.1	-	-0.1/+0.3	+0.0/-0.1	+0.0/+0.3
WH cat1	-	-0.1/-0.0	+0.1/-0.1	+0.1/-0.0	-0.1/+0.2	-0.1/-0.2
WH cat2	+0.0/+0.1	-0.2/-0.0	-0.2/+0.1	-0.2/-0.0	+0.0/+0.4	+0.1/-0.0
ZH cat0	-	+0.0/+0.1	+0.0/+0.1	-	-	-
ZH cat1	-	-0.1/-0	-	-	-	-

Table 6-8. 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 modeling. The ID scale factor and Isolation scale factor are covered by the lepMVA scale factors. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected.

Uncertainty	Category	WH	qqZH	ggZH	ttH	THQ	THW
muon SF (%)	WH cat0	-1.8/+1.8	-1.7/+1.8	-2.3/+2.3	-2.3/+2.3	-2.1/+2.1	-2.5/+2.5
	WH cat1	-1.7/+1.8	-1.7/+1.7	-2.4/+2.4	-2.0/+2.0	-1.9/+2.0	-2.3/+2.3
	WH cat2	-2.3/+2.3	-1.9/+2.0	-2.5/+2.5	-2.5/+2.5	-2.3/+2.4	-2.8/+2.9
	ZH cat0	-	-1.9/+1.9	-2.5/+2.6	-	-	-
	ZH cat1	-	-2.5/+2.6	-3.3/+3.4	-	-	-
electron SF (%)	WH cat0	-0.3/+0.3	-0.4/+0.4	-0.4/+0.4	-0.3/+0.3	-0.4/+0.4	-0.2/+0.2
	WH cat1	-0.5/+0.5	-0.6/+0.6	-0.5/+0.5	-0.5/+0.5	-0.6/+0.6	-0.5/+0.5
	WH cat2	-0.5/+0.5	-0.6/+0.6	-0.6/+0.6	-0.5/+0.5	-0.6/+0.6	-0.5/+0.5
	ZH cat0	-	-0.6/+0.6	-0.6/+0.6	-	-	-
	ZH cat1	-	-0.8/+0.8	-0.6/+0.6	-	-	-

Table 6-9. Uncertainties on different signal components in the WH and ZH channels related to B-jet vetoing. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected.

Uncertainty (%)	WH	qqZH	ggZH	ttH	THQ	THW
WH cat0	+0.1/-0.1	+0.1/-0.1	-0.8/+0.8	+5.5/-5.3	-0.9/+0.9	-1.4/+1.4
WH cat1	+0.1/-0.1	+0.1/-0.1	-0.9/+0.9	+5.7/-5.5	-0.8/+0.8	-1.3/+1.3
WH cat2	+0.1/-0.1	+0.0/-0.1	-1.0/+1.0	+5.3/-5.1	-0.9/+0.9	-1.4/+1.4
ZH cat0	-	+0.1/-0.1	-0.6/+0.6	-	-	-
ZH cat1	-	+0.2/-0.2	-0.5/+0.5	-	-	-

Table 6-10. Uncertainties on different signal components in the WH Cat0 related to jet energy calibration. 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. Numbers before/after are the affects from shifting the uncertainty source up/down. Uncertainties smaller than 0.1% are neglected.

Uncertainty (%)	WH	qqZH	ggZH	ttH	THQ	THW
flavorQCD	+0.1/-0.0	-0.4/-0.0	+0.1/-0.0	+0.6/-0.9	+0.2/-0.5	+1.5/-0.8
relativeBal	+0.1/-0.0	-0.2/-0.0	+0.1/-0.0	+0.7/-0.5	+0.0/-0.2	+0.7/-0.5
absolute	-	+0.0/+0.1	-	-	+0.0/+0.1	+0.1/+0.4
BBEC1	+0.1/-0.0	-0.4/-0.0	+0.1/-0.1	+1.5/-1.1	+0.2/-0.7	+1.2/-0.7
EC2	-	-	-	-0.1/-0.0	-0.1/-0.3	+0.2/+0.1
HF	-0.1/+0.1	-0.2/-0.0	+0.1/+0.1	+0.5/-0.6	+0.0/-0.5	+0.5/+0.3
relativeSample_2016	-	-0.1/-0.0	-	+0.1/-0.3	-	+0.1/-0.3
absolute_2016	-	+0.0/-0.1	-	+0.1/-0.1	+0.0/+0.1	-
BBEC1_2016	-	+0.0/-0.1	-	+0.0/-0.1	+0.0/+0.1	+0.1/-0.0
EC2_2016	-	-	-	-	+0.0/+0.1	+0.1/-0.0
HF_2016	-	-	-	-	-	+0.1/-0.0
relativeSample_2017	-	-0.1/+0.2	+0.1/-0.0	+0.1/-0.2	-0.2/+0.1	+0.3/+0.3
absolute_2017	-	-0.1/+0.2	-	+0.2/-0.1	-	+0.3/+0.2
BBEC1_2017	-	-0.1/+0.1	-	+0.2/+0.0	-0.1/+0.1	+0.2/+0.0
EC2_2017	-	+0.1/-0.0	-	-	+0.1/+0.3	+0.1/+0.3
HF_2017	-	-	-	-	-0.1/-0.2	-
relativeSample_2018	-0.1/+0.1	+0.1/+0.2	+0.1/-0.0	+0.4/-0.5	+0.1/-0.4	+0.7/-0.7
absolute_2018	+0.0/+0.1	-0.2/-0.0	-	+0.0/-0.1	+0.2/-0.4	+0.3/-0.6
BBEC1_2018	+0.0/+0.1	-0.2/-0.0	-	+0.0/-0.1	-	+0.1/-0.1
EC2_2018	+0.0/+0.1	-0.1/+0.1	-	+0.0/+0.1	+0.0/+0.1	+0.0/+0.1
HF_2018	-	+0.0/-0.1	-	+0.0/+0.1	+0.0/-0.1	-

6.5.3 Background systematic bias

As described in Section 6.4.2, the background modeling follows a data-driven approach, and is not affected by any systematic uncertainty in the simulation. Instead of evaluating the impacts of uncertainties as done for the signal modeling, the main task for background is to make sure it is robust against spurious signals.

A spurious signal is produced by the bias between the analytic background function and the true background shape at the position of the expected signal. For an analysis with finite statistics, there is a statistical uncertainty on the signal strength, σ_{stat} , resulted from the statistical fluctuation of background events. The best fit signal strength μ_{fit} under the null hypothesis, which is the hypothesis without the existence of a true signal, should follow $\mathcal{N}(0, \sigma_{stat})$, the normal distribution with a mean of 0 and a standard deviation of σ_{stat} . The 1σ or 2σ range of this distribution gives the 68.3% or 95.4% confidence intervals for the exclusion of this null hypothesis. If there is a systematic spurious signal $\hat{\mu}_{SS}$, namely a bias, the probability distribution for the μ_{fit} becomes $\mathcal{N}(\hat{\mu}_{SS}, \sigma_{stat})$, and the coverage of the $1\sigma_{stat}$ range becomes equation 6-11, which is not 68.3%.

$$\int_{-\sigma_{stat}}^{\sigma_{stat}} \mathcal{N}(\hat{\mu}_{SS}, \sigma_{stat}) = \frac{1}{2} [erf(\frac{\sigma_{stat} + \hat{\mu}_{SS}}{\sqrt{2} \sigma_{stat}}) + erf(\frac{\sigma_{stat} - \hat{\mu}_{SS}}{\sqrt{2} \sigma_{stat}})] \quad (6-11)$$

The left plot of Figure 6-15 illustrates this difference in coverage. The middle plot of Figure 6-15 shows how the coverage changes as the bias gets larger. As a result, to achieve the 68.3% confidence level, the signal uncertainty from the fit, σ_{fit} , needs to satisfy equation 6-12 and becomes larger than the σ_{stat} .

$$\frac{1}{2} [erf(\frac{\sigma_{fit} + \hat{\mu}_{SS}}{\sqrt{2} \sigma_{stat}}) + erf(\frac{\sigma_{fit} - \hat{\mu}_{SS}}{\sqrt{2} \sigma_{stat}})] = 68.3\% \quad (6-12)$$

In this way, the bias in the background modeling, even if it is not strictly an uncertainty, adds to the overall uncertainty of the signal strength measurement. The relationship between σ_{fit} and $\hat{\mu}_{SS}$ is shown in the right plot of Figure 6-15. As a convention in the $H \rightarrow \mu\mu$ analysis, biases below 20% of the statistical uncertainty of the signal are considered ac-

ceptable, which corresponds to less than 2% inflation of the signal uncertainty.

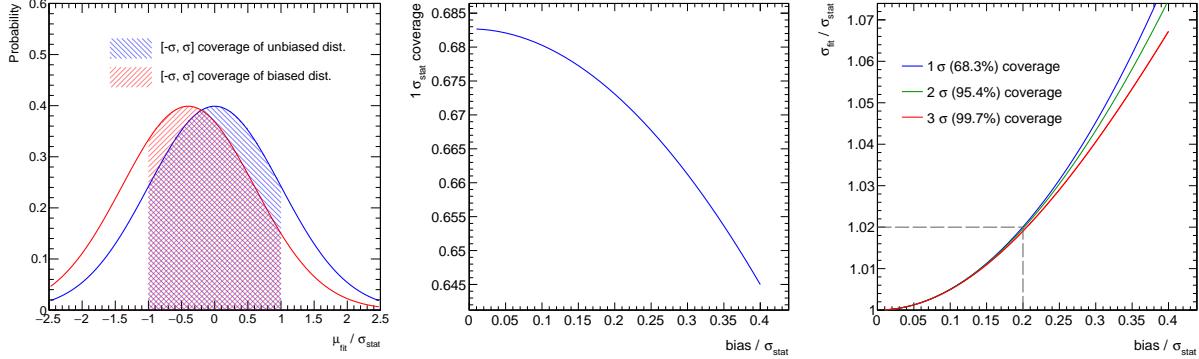


Figure 6-15. Schemes on how the bias affects the uncertainty on the signal strength measurement. The left plot is an illustration of the $[-\sigma, \sigma]$ coverage of the biased signal strength measurement. The middle plot shows this coverage becomes less as the bias gets larger. The right plot shows how the bias impacts the best fit signal strength uncertainty σ_{fit} . The gray dash lines in the right plot indicates the conventional acceptable range of the bias in this analysis.

In this analysis, the bias is evaluated between different function candidates via groups of toy studies. The true background shape is of course unknown, but is believed to be covered by the flexibility of the collective set of functional forms. Procedures for the bias evaluation are as follows:

1. Toy generation

- One function candidate $f(m_{\mu\mu})$ is fit to the background shape of the simulation to find the best fit parameters.
- The best fit shape of $f(m_{\mu\mu})$ is used as the Probability Density Function (PDF) to generate toy datasets. In each toy, the number of events in each bin is taken sampling the Poisson distribution of the expected number of events given by $f(m_{\mu\mu})$.
- For each selected function $f(m_{\mu\mu})$, 3000 toys are generated.

2. Signal injection

- For each background toy, an artificial signal is also generated following the Poisson distribution of a given signal strength $\hat{\mu}_{inj}$. In this set of study, two sets of artificial signal strength are tested, which are zero or the expected SM signal strength.

- The artificial signal toys are added to the background toys, which completes the signal + background toys.

3. Signal extraction

- For each signal + background toy of function $f(m_{\mu\mu})$, the shape analysis is performed using another function $g(m_{\mu\mu})$. In these toy analyses, systematic uncertainties on the signal modeling are not included, as they are unrelated to the bias estimation.
- From these fits, the best fit signal strength μ_{fit} and its standard deviation σ_{fit} are extracted.

4. Bias evaluation

- The spurious signal between function $f(m_{\mu\mu})$ and $g(m_{\mu\mu})$ in each toy is defined as:

$$\mu_{ss}(f,g) = \frac{\mu_{fit} - \hat{\mu}_{inj}}{\sigma_{fit}} \quad (6-13)$$

- The distribution (of 3000 toys) of this spurious signal is fit with a gaussian function. As stated above, this spurious signal should follow the Gaussian distribution $\mathcal{N}(\hat{b}, 1)$, where $\hat{b} = \hat{\mu}_{ss}/\sigma_{stat}$. The mean value from the Gaussian fit is the bias between function $f(m_{\mu\mu})$ and $g(m_{\mu\mu})$.

Such bias is evaluated in each sub-category between each combination of $f(m_{\mu\mu})$ and $g(m_{\mu\mu})$, and the results are summarized in Figure 6-16 and 6-17. Most of the functions have good bias response against other functions can be used for the analysis. In the final analysis, the BWZGamma function (2 shape DOF) is chosen as the background model in the WH [-1.0, -0.1] category, and the BWZ function (1 shape DOF) is chosen in all other categories.

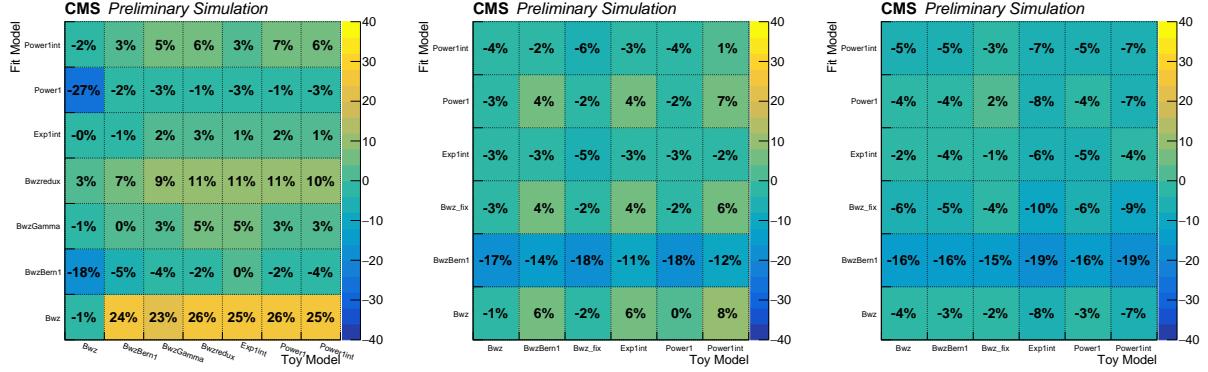


Figure 6-16. Bias in different BDT-based WH sub-categories. The sub-categories are: Cat0 BDT [-1.0, -0.1] (left), Cat1 BDT [-0.1, 0.3] (middle), Cat2 BDT [0.3, 1.0] (right). In the tables, the Power stands for a single "Power" function and the "PowerInt" stands for a single power function plus a constant.

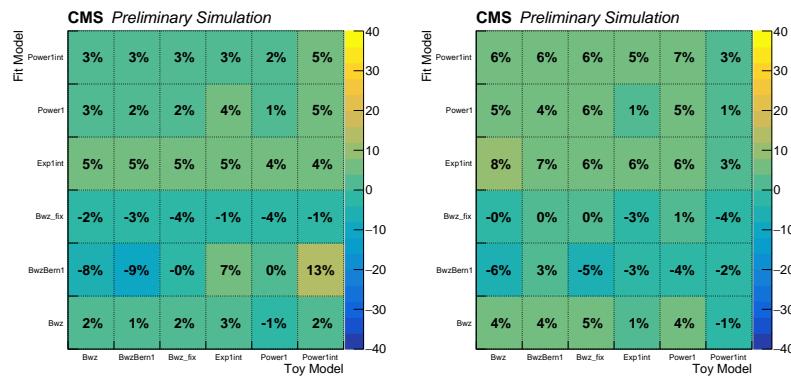


Figure 6-17. Bias in different BDT-based ZH sub-categories. The sub-categories are: Cat0 BDT [-1.0, -0.1] (left), Cat1 BDT [-0.1, 1.0] (right). In the tables, the "Power" stands for a single power function and the "PowerInt" stands for a single power function plus a constant.

6.6 Results of the VH analysis

The final results are extracted by performing a binned maximum-likelihood fit in each VH sub-category. The fit is performed on the observed $m_{\mu\mu}$ distribution in the range of $110 < m_{\mu\mu} < 150$ GeV. All the different signal modes in different sub-categories share a common signal strength modifier μ . Figure 6-18 and 6-19 show the post-fit results of the signal-plus-background fits in the WH and ZH sub-categories.

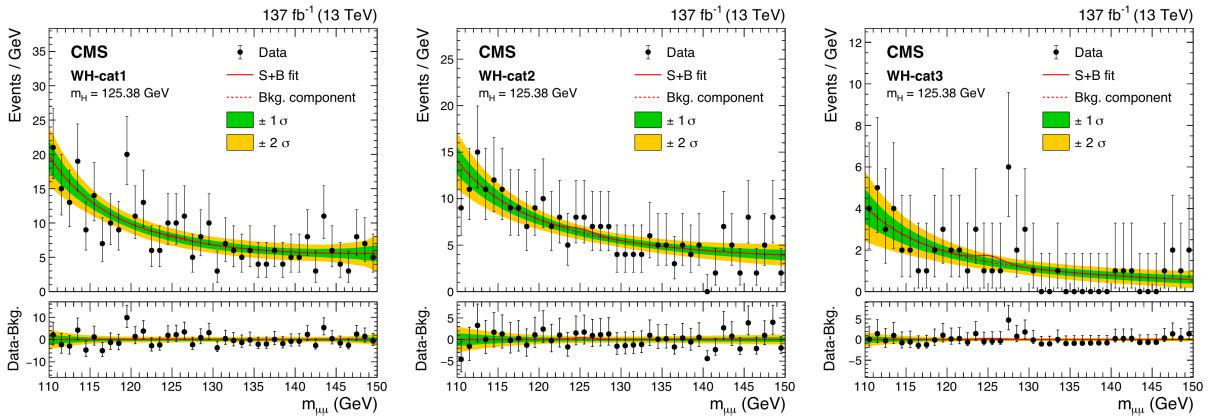


Figure 6-18. Plots taken from Ref. [21]. Post-fit $m_{\mu\mu}$ distribution of the WH sub-categories. The sub-categories are: Cat0 BDT [-1.0, -0.1] (left), Cat1 BDT [-0.1, 0.3] (middle), Cat2 BDT [0.3, 1.0] (right). The upper panel in the plots shows the distribution of observed data and the shape of the signal-plus-background fit. The lower panel in the plots shows the residual distribution after subtracting the background component in the fits. The green and yellow bands show the one and two standard deviation of the background component uncertainty.

As described in Section 3, the VH analysis is combined with other categories (ggH, VBF, and $t\bar{t}H$) to make the inclusive $H \rightarrow \mu\mu$ analysis. More studies in the statistical analysis are covered in more details in Chapter 7. The expected and observed limits and significance for the VH sub-categories are summarized in Table 6-11.

put likelihood scan, goodness of fit, and brazilian plots here or in the next chapter?

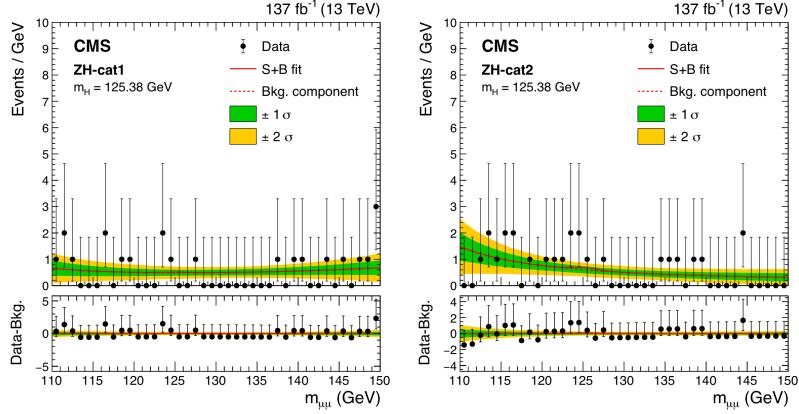


Figure 6-19. Plots taken from Ref. [21]. Post-fit $m_{\mu\mu}$ distribution of the ZH sub-categories. The sub-categories are: Cat0 BDT [-1.0, -0.1] (left), Cat1 BDT [-0.1, 1.0] (right). The upper panel in the plots shows the distribution of observed data and the shape of the signal-plus-background fit. The lower panel in the plots shows the residual distribution after subtracting the background component in the fits. The green and yellow bands show the one and two standard deviation of the background component uncertainty.

Table 6-11. Summary of the expect and observed limits and significance in each individual sub-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

CHAPTER 7 RESULTS OF THE H2MU SEARCH

modify from AN

7.1 projection of search at 14 TeV

REFERENCES

- [1] M. Aaboud et al., *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), 051802.
- [2] G. Aad et al., *Search for the standard model higgs boson decay to + with the atlas detector*, Physics Letters B **738** (2014), 68 – 86.
- [3] Georges Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012), 1.
- [4] _____, *A search for the dimuon decay of the standard model higgs boson with the atlas detector*, Submitted to *Phys. Lett. B*, 2020.
- [5] S. Agostinelli et al., *Geant4a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003), no. 3, 250 – 303.
- [6] Simone Alioli, Paolo Nason, Carlo Oleari, and Emanuele Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010), 043.
- [7] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *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.
- [8] Johan Alwall et al., *Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions*, Eur. Phys. J. C **53** (2008), 473.
- [9] Charalampos Anastasiou, Claude Duhr, Falko Dulat, Elisabetta Furlan, Thomas Gehrmann, Franz Herzog, Achilleas Lazopoulos, and Bernhard Mistlberger, *High precision determination of the gluon fusion Higgs boson cross-section at the LHC*, JHEP **05** (2016), 058.
- [10] E. Bagnaschi, G. Degrassi, P. Slavich, and A. Vicini, *Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM*, JHEP **02** (2012), 088.
- [11] Richard D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015), 040.
- [12] _____, *Parton distributions from high-precision collider data*, Eur. Phys. J. C **77** (2017), 663.
- [13] Johannes Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016), 196.
- [14] A. Bodek, A. van Dyne, J. Y. Han, W. Sakumoto, and A. Strelnikov, *Extracting muon momentum scale corrections for hadron collider experiments*, Eur. Phys. J. C **72** (2012), 2194.
- [15] Oliver Brein, Abdelhak Djouadi, and Robert Harlander, *NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders*, Phys. Lett. B **579** (2004), 149.

- [16] Matteo Cacciari, Frédéric A. Dreyer, Alexander Karlberg, Gavin P. Salam, and Giulia Zanderighi, *Fully differential vector-boson-fusion Higgs production at next-to-next-to-leading order*, Phys. Rev. Lett. **115** (2015), 082002, [Erratum: DOI10.1103/PhysRevLett.120.139901].
- [17] John M. Campbell, R. Keith Ellis, and Ciaran Williams, *Vector boson pair production at the LHC*, JHEP **07** (2011), 018.
- [18] Andrew M. Carnes, *A search for the standard model higgs boson decaying to two muons at the cms experiment*, Ph.D. thesis, University of Florida, 2018.
- [19] Serguei Chatrchyan et al., *Observation of a new boson at a mass of 125GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012), 30.
- [20] _____, *Observation of a new boson with mass near 125GeV in pp collisions at $\sqrt{s} = 7$ and 8TeV*, JHEP **06** (2013), 081.
- [21] CMS Collaboration, *Evidence for higgs boson decay to a pair of muons*, 2020.
- [22] Michal Czakon and Alexander Mitov, *Top++: A program for the calculation of the top-pair cross-section at hadron colliders*, Comput. Phys. Commun. **185** (2014), 2930.
- [23] S. Dawson, C. Jackson, L. H. Orr, L. Reina, and D. Wackerlo, *Associated Higgs production with top quarks at the large hadron collider: NLO QCD corrections*, Phys. Rev. D **68** (2003), 034022.
- [24] Federico Demartin, Benedikt Maier, Fabio Maltoni, Kentarou Mawatari, and Marco Zaro, *tWH associated production at the LHC*, Eur. Phys. J. C **77** (2017), 34.
- [25] Federico Demartin, Fabio Maltoni, Kentarou Mawatari, and Marco Zaro, *Higgs production in association with a single top quark at the LHC*, Eur. Phys. J. C **75** (2015), 267.
- [26] A. Djouadi, J. Kalinowski, and M. Spira, *HDECAY: A program for Higgs boson decays in the standard model and its supersymmetric extension*, Comput. Phys. Commun. **108** (1998), 56.
- [27] Rikkert Frederix and Stefano Frixione, *Merging meets matching in MC@NLO*, JHEP **12** (2012), 061.
- [28] S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro, *Weak corrections to Higgs hadroproduction in association with a top-quark pair*, JHEP **09** (2014), 065.
- [29] Stefano Frixione, Paolo Nason, and Carlo Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, Journal of High Energy Physics **2007** (2007), no. 11, 070–070.
- [30] T. Gehrmann, M. Grazzini, S. Kallweit, P. Maierhöfer, A. von Manteuffel, S. Pozzorini, D. Rathlev, and L. Tancredi, *W^+W^- production at hadron colliders in next-to-next-to-leading order QCD*, Phys. Rev. Lett. **113** (2014), 212001.

- [31] Massimiliano Grazzini, Stefan Kallweit, and Dirk Rathlev, *ZZ production at the LHC: fiducial cross sections and distributions in NNLO QCD*, Phys. Lett. B **750** (2015), 407.
- [32] Massimiliano Grazzini, Stefan Kallweit, Dirk Rathlev, and Marius Wiesemann, *$W^\pm Z$ production at the LHC: fiducial cross sections and distributions in NNLO QCD*, JHEP **05** (2017), 139.
- [33] Keith Hamilton, Paolo Nason, Emanuele Re, and Giulia Zanderighi, *NNLOPS simulation of Higgs boson production*, JHEP **10** (2013), 222.
- [34] Keith Hamilton, Paolo Nason, and Giulia Zanderighi, *Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator*, JHEP **05** (2015), 140.
- [35] Heribertus B. Hartanto, Barbara Jager, Laura Reina, and Doreen Wackerlo, *Higgs boson production in association with top quarks in the POWHEG BOX*, Phys. Rev. D **91** (2015), 094003.
- [36] P. Kant, O. M. Kind, T. Kintscher, T. Lohse, T. Martini, S. Mölbitz, P. Rieck, and P. Uwer, *HATHOR for single top-quark production: Updated predictions and uncertainty estimates for single top-quark production in hadronic collisions*, Comput. Phys. Commun. **191** (2015), 74.
- [37] V. Khachatryan et al., *Search for a standard model-like higgs boson in the + and e+e decay channels at the lhc*, Physics Letters B **744** (2015), 184 – 207.
- [38] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, 10 2016.
- [39] Ye Li and Frank Petriello, *Combining QCD and electroweak corrections to dilepton production in FEWZ*, Phys. Rev. D **86** (2012), 094034.
- [40] Gionata Luisoni, Paolo Nason, Carlo Oleari, and Francesco Tramontano, *$HW^\pm/HZ + 0$ and 1 jet at NLO with the POWHEG BOX interfaced to gosam and their merging within MiNLO*, JHEP **10** (2013), 083.
- [41] Paolo Nason, *A new method for combining NLO QCD with shower monte carlo algorithms*, Journal of High Energy Physics **2004** (2004), no. 11, 040–040.
- [42] Paolo Nason and Carlo Oleari, *NLO Higgs boson production via vector-boson-fusion matched with shower in POWHEG*, JHEP **02** (2010), 037.
- [43] Oliver Rieger, *Tbd*, Ph.D. thesis, University of Hamburg, 2020.
- [44] A. M. Sirunyan et al., *Identification of heavy-flavour jets with the CMS detector in pp collisions at 13TeV*, JINST **13** (2018), P05011.
- [45] ———, *Search for the higgs boson decaying to two muons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, Phys. Rev. Lett. **122** (2019), 021801.

- [46] Albert M Sirunyan et al., *Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13\text{TeV}$* , JHEP **11** (2017), 047.
- [47] _____, *Search for a new scalar resonance decaying to a pair of Z bosons in proton-proton collisions at $\sqrt{s} = 13\text{ TeV}$* , JHEP **06** (2018), 127, [Erratum: JHEP03,128(2019)].
- [48] A.M. Sirunyan et al., *Particle-flow reconstruction and global event description with the CMS detector*, Journal of Instrumentation **12** (2017), no. 10, P10003–P10003.
- [49] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands, *An introduction to pythia 8.2*, Computer Physics Communications **191** (2015), 159 – 177.
- [50] Michael Spira, *QCD effects in Higgs physics*, Fortsch. Phys. **46** (1998), 203.