

A Collider Search for Dark Matter Produced in Association
with a Higgs Boson with the CMS Detector at the 13 TeV LHC

By

DUSTIN RAY BURNS

B.S. (Georgia Institute of Technology) 2011

M.S. (University of California, Davis) 2012

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Physics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Chair Michael Mulhearn

Robin Erbacher

Albert De Roeck

Committee in Charge

2017

Copyright © 2017 by
Dustin Ray Burns
All rights reserved.

CONTENTS

List of Figures	iv
List of Tables	viii
1 Event reconstruction and simulation	1
1.1 Particle reconstruction	1
1.2 Monte Carlo event simulation	4
2 Datasets and Monte Carlo Simulations	7
2.1 Data	7
2.1.1 Triggers and Datasets	7
2.1.2 Trigger Efficiency	9
2.2 Simulation	10
2.2.1 Signal Samples	10
2.2.2 Background Samples	11
3 Physics Objects	14
3.1 Electrons	14
3.1.1 Electron Reconstruction	14
3.1.2 Electron Identification	14
3.1.3 Electron Isolation	18
3.1.4 Electron Energy Calibrations	18
3.1.5 Electron Efficiency Measurements	19
3.2 Muons	21
3.2.1 Muon Reconstruction and Identification	21
3.2.2 Muon Isolation	22
3.2.3 Muon Efficiency Measurements	23
3.3 Photons for FSR recovery	26
3.4 Jets	27
3.5 MET	29

4 Event Selection	33
4.1 Trigger Selection	33
4.2 Vertex Selection	33
4.3 ZZ Candidate Selection	33
4.4 Choice of the best ZZ Candidate	35
5 Background Estimation	36
5.1 Irreducible Backgrounds	36
5.1.1 $q\bar{q} \rightarrow ZZ$ Modelling	36
5.1.2 $gg \rightarrow ZZ$ Modelling	37
5.2 Reducible Background	38
6 Signal region and blinding	45
6.0.1 Cut-and-count based signal region	45
6.0.2 MVA based signal region	46
6.1 Yields and distributions	47
7 Results	50
7.1 Statistical analysis	50
7.1.1 Systematic uncertainties	50
7.1.2 Limit setting	56
7.2 Results	59
7.2.1 Cut-and-count based	59
7.2.2 MVA based	60
8 Conclusions	64
Appendices	66
A Production Cross Sections for Benchmark Signal Models	67

LIST OF FIGURES

1.1	Event display of hadronic jet in the x-y plane, with solid arcs at the ECAL and HCAL surfaces. The locations of clusters are given by the solid dots.	3
1.2	Event display of hadronic jet in the $\eta - \phi$ plane for the ECAL (a) and HCAL (b). The locations of clusters are given by the solid dots.	4
1.3	MET reconstruction resolution.	4
2.1	Trigger efficiency measured in data using 4ℓ events collected by single lepton triggers for the $4e$ (top left), 4μ (top right), $2e2\mu$ (bottom left) and 4ℓ (bottom right) final states.	10
2.2	Number of pileup vertices before and after reweighted is applied.	13
3.1	Electron reconstruction efficiencies efficiency in data versus η and data/MC scale factors as provided by the EGM POG.	15
3.2	Performance comparison of the MVA trained for the 2015 analysis and the retraining for 2016 conditions. The respective working points are indicated by the markers.	16
3.3	BDT output for the training and testing sample for true and fake electrons in the high- p_T endcap training bins.	17
3.4	(a): electron energy scale measured in the $Z + \ell$ control region for EB and EE electrons. The results of the Crystall-ball fit are reported in the figure. (b): lepton energy scales per 500 pb luminosity block.	19
3.5	Electron selection efficiencies measured using the Tag-and-Probe technique described in the text, non-gap electrons (left) and gap electrons (right). .	20
3.6	Electron selection efficiencies measured using the Tag-and-Probe technique described in the text, non-gap electrons (top) and gap electrons (bottom)	20

3.7 Muon reconstruction and identification efficiency at low p_T , measured with the tag&probe method on J/Ψ events, as function of p_T in the barrel (left) and endcaps (center), and as function of η for $p_T > 7\text{GeV}$ (right). In the upper panel, the larger error bars include also the systematical uncertainties, while the smaller ones are purely statistical. In the lower panel showing the ratio of the two efficiencies, the black error bars are for the statistical uncertainty, the orange rectangles for the systematical uncertainty and the violet rectangles include both uncertainties.	24
3.8 Efficiency of the muon impact parameter requirements, measured with the tag&probe method on Z events, as function of p_T in the barrel (left) and endcaps (center), and as function of η for $p_T > 20\text{GeV}$ (right). In the upper panel, the larger error bars include also the systematical uncertainties, while the smaller ones are purely statistical. In the lower panel showing the ratio of the two efficiencies, the black error bars are for the statistical uncertainty, the orange rectangles for the systematical uncertainty and the violet rectangles include both uncertainties.	24
3.9 Efficiency of the muon isolation requirement, measured with the tag&probe method on Z events, as function of p_T in the barrel (left) and endcaps (right). In the upper panel, the larger error bars include also the systematical uncertainties, while the smaller ones are purely statistical. In the lower panel showing the ratio of the two efficiencies, the black error bars are for the statistical uncertainty, the orange rectangles for the systematical uncertainty and the violet rectangles include both uncertainties.	25
3.10 Tracking efficiency in data and simulation as a function of η for muon $p_T < 10\text{GeV}$ (left) and $p_T > 10\text{GeV}$ (right) with ReReco data.	26
3.11 Left: Overall data to simulation scale factors for muons, as function of p_T and η . Right: Uncertainties on data to simulation scale factors for muons, as function of p_T and η	26
3.12 Missing transverse energy after SM selection before removal of fake MET.	31

3.13 Maximum azimuthal angle difference between MET and jets.	32
3.14 Minimum azimuthal angle difference between MET and jets.	32
5.1 Left: NNLO/NLO QCD k -factor for the $q\bar{q} \rightarrow ZZ$ background as a function of $m(ZZ)$ for the 4ℓ and $2\ell 2\ell'$ final states. Right: NLO/NLO electroweak k -factor for the $q\bar{q} \rightarrow ZZ$ background as a function of $m(ZZ)$	37
5.2 $gg \rightarrow H \rightarrow 2\ell 2\ell'$ cross sections at NNLO, NLO and LO at each H boson pole mass using the SM H boson decay width (top) or at the fixed and small decay width of 4.07 MeV (top). The cross sections using the fixed value are used to obtain the K factor for both the signal and the continuum background contributions as a function of $m_{4\ell}$ (bottom).	38
5.3 Fake rates as a function of the probe p_T for electrons (a) and muons (b) which satisfy the loose selection criteria, measured in a $Z(\ell\ell) + \ell$ sample in the 13 TeV data. The barrel selection includes electrons (muons) up to $ \eta = 1.479$ (1.2).	40
5.4 Invariant mass distribution of the events selected in the 2P+2F control sample in the 13 TeV dataset, (top left) 4μ , (top right) $4e$, (bottom left) $2\mu 2e$ and (bottom right) $2e 2\mu$ channels.	41
5.5 Invariant mass distribution of the events selected in the 3P+1F control sample in the 13 TeV dataset, (top left) 4μ , (top right) $4e$, (bottom left) $2\mu 2e$ and (bottom right) $2e 2\mu$ channels.	42
6.1 PLACEHOLDER BDT response.	47
7.1 Difference between the $Z \rightarrow \ell\ell$ mass peak positions in data and simulation normalized by the nominal Z boson mass obtained as a function of the p_T and $ \eta $ of one of the leptons regardless of the second for muons (left) and electrons (right).	54

7.2	Different $m_{4\ell}$ distributions after propagating the biases in Fig. 7.1 to Higgs boson events. The change in the mean of the double crystal ball is used to determine the systematic uncertainty due to the lepton momentum scale. The middle plot shows the nominal distribution, while the left (right) plots show the down (up) systematic variations. The 4μ channel is shown in the top row and the $4e$ channel is shown in the bottom row.	55
7.3	PLACEHOLDER 1D μ limits for $m_{A0} = 300$ GeV for Zp2HDM.	59
7.4	PLACEHOLDER 1D cross section times BR limits for $m_{A0} = 300$ GeV for Zp2HDM.	60
7.5	PLACEHOLDER 1D μ limits for $m_\chi = 1$ GeV for ZpBaryonic.	61
7.6	PLACEHOLDER 1D cross section times BR limits for $m_\chi = 1$ GeV for ZpBaryonic.	61
7.7	PLACEHOLDER 1D μ limits for $m_{A0} = 300$ GeV for Zp2HDM.	62
7.8	PLACEHOLDER 1D cross section times BR limits for $m_{A0} = 300$ GeV for Zp2HDM.	62
7.9	PLACEHOLDER 1D μ limits for $m_\chi = 1$ GeV for ZpBaryonic.	63
7.10	PLACEHOLDER 1D cross section times BR limits for $m_\chi = 1$ GeV for ZpBaryonic.	63

LIST OF TABLES

2.1	Datasets used in the analysis.	8
2.2	Trigger paths used in 2016 collision data.	9
2.3	Trigger efficiencies measured using 4ℓ events.	11
2.4	Benchmark signal samples analyzed.	11
2.5	Higgs signal samples and cross sections.	12
2.6	Background Monte Carlo samples and cross sections.	12
3.1	Overview of input variables to the identification classifier. Variables not used in the run I MVA are marked with (\cdot)	17
3.2	Minimum BDT score required for passing the electron identification. . . .	17
3.3	The requirements for a muon to pass the Tracker High- p_T ID. Note that these are equivalent to the Muon POG High- p_T ID with the global track requirements removed.	22
5.1	Cross sections for $q\bar{q} \rightarrow ZZ$ production at 13 TeV	36
5.2	The contribution of reducible background processes in the signal region predicted from measurements in data using the Opposite-Sign Leptons method. The predictions correspond to 35.9 fb^{-1} of data at 13 TeV. . . .	44
6.1	PLACEHOLDER Event yields after SM selection, with $m_{4l} > 70 \text{ GeV}$. .	48
6.2	PLACEHOLDER Cut flow for $4e$ channel. Example signal is Z'2HDM with $m_{Z'} = 600 \text{ GeV}$	48
6.3	PLACEHOLDER Cut flow for 4μ channel. Example signal is Z'2HDM with $m_{Z'} = 600 \text{ GeV}$	49
6.4	PLACEHOLDER Cut flow for $2e2\mu$ channel. Example signal is Z'2HDM with $m_{Z'} = 600 \text{ GeV}$	49
7.3	Summary of the experimental systematic uncertainties in the $H \rightarrow 4\ell$ measurements.	56

7.4	Summary of the theory systematic uncertainties in the $H \rightarrow 4\ell$ measurements for the inclusive analysis	57
A.1	EFT production cross sections [pb]	68
A.2	Scalar production cross sections [pb] corresponding to mass points in Table ??	69
A.3	ZpBaryonic production cross sections [pb] corresponding to mass points in Table ??	70
A.4	ZpHS production cross sections [pb] corresponding to mass points in Table ??	71
A.5	Zp2HDM production cross sections [pb] corresponding to mass points in Table ??	72

Chapter 1

Event reconstruction and simulation

This chapter overviews the algorithms used to reconstruct the trajectories and identify (ID) types of particles produced in proton-proton collisions in CMS, collectively known as Particle Flow (PF), and how these collisions are simulated. Unless otherwise noted, the material in the first section comes from reference [?].

1.1 Particle reconstruction

The PF algorithms combine information from all of the CMS sub-detectors discussed in the previous chapter to reconstruct the particles produced in the collision event. Since many of the particles produced initially in the collision are unstable, decaying before they have time to interact with the sub-detectors, PF reconstructs the stable particles: electrons, muons, photons, and hadrons. The remaining physics objects of interest, jets, missing energy, taus, etc, can be determined from the information provided by the stable PF-IDed particles.

The different particles are reconstructed and IDed using information from individual sub-detectors, or combinations of sub-detectors. The direction and momentum of charged particles is measured by the tracker. Electrons are reconstructed using tracks and energy deposits in the ECAL. Muons are reconstructed from a combination of tracker and muon chamber data. Photons are reconstructed from energy deposits in the ECAL. Finally, charged and neutral hadrons are reconstructed from energy deposited primarily in the HCAL, with a contribution from energy deposits in the ECAL. The missing transverse

energy, an observable of particular importance to this analysis, used to identify DM that does not interact with the detector material, is the modulus of the sum of transverse momenta of all the PF reconstructed particles.

The basic pieces of information from the subdetectors used by PF are called elements, and consist of charged-particle tracks, muon tracks, and calorimeter clusters. The tracker provides charged-particle track elements. Since the tracker has the best momentum resolution of the subdetectors, it is of critical importance that the tracking efficiency be nearly 100%, with as low a fake rate as possible, to reduce an excess in reconstructed energy. These two goals are accomplished using an iterative algorithm: first, tracks are seeded using very tight criteria, yielding a low efficiency, but negligible fake rate, then track seed criteria are loosened and hits that clearly belong to a track are removed, resulting in increasing efficiency. The ECAL and HCAL subsystems (ECAL barrel, HCAL barrel, HCAL endcap, PS first layer, and PS second layer) provide cluster elements. The calorimeter clustering algorithm measures the energy and direction of neutral particles (e.g. photons, neutral hadrons), differentiates energy deposits from neutral and charged hadrons, reconstructs electrons, and contributes to the reconstruction of charged hadrons. The algorithm is summarized as follows: cluster seeds are identified as energy deposit peaks over a given energy, from which topological clusters are grown by appending adjacent cells, and last, topological clusters seed PF clusters. An example is shown in Figure 1.1 and Figure 1.2, where a simple jet is reconstructed into four clusters, shown as dots.

Once the PF elements are determined, they are linked together into blocks, which correspond to the signatures left in the sub-detectors of a single particle. Single particles typically leave one to three elements. The linking algorithm determines the quality of the link between all pairwise elements in an event, then forms blocks from the highest quality links, starting from the tracker, and proceeding outward through the calorimeters and muon chambers. Once the blocks are formed, PF associates a global event particle with each block. PF muons are formed from global muon candidates if its momentum is consistent across all track elements. PF electrons are IDed from electron candidates using tracker and ECAL cluster variables, accounting for the Bremsstrahlung photons produced

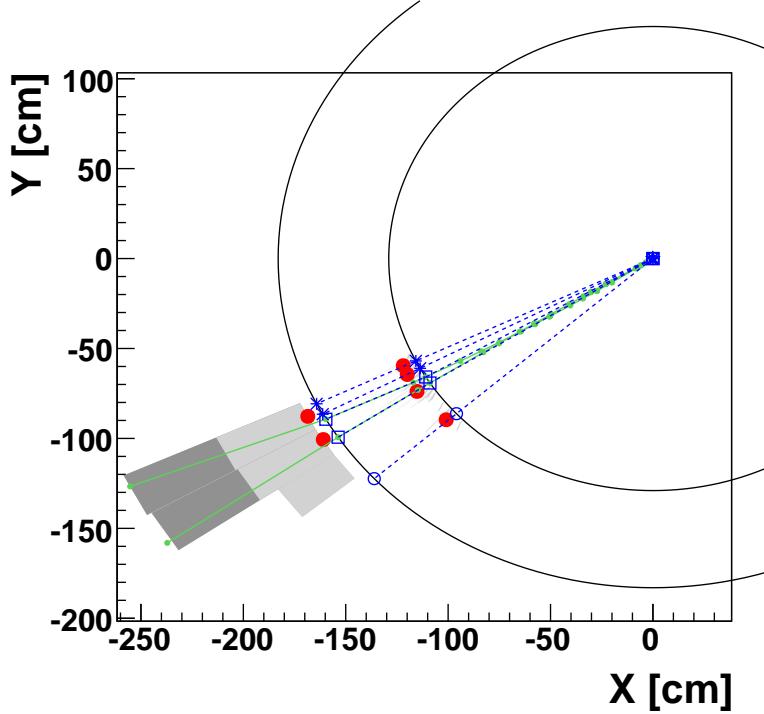


Figure 1.1: Event display of hadronic jet in the x-y plane, with solid arcs at the ECAL and HCAL surfaces. The locations of clusters are given by the solid dots.

when the electron passes through the tracker material. Once the elements associated to PF muons and PF electrons are removed, the remaining elements are analyzed to ID charged hadrons, photons, or neutral hadrons. PF charged hadrons are associated to remaining tracks if the linked clusters are consistent with the measured momenta. If the energy of the linked clusters is much larger than the track momentum, accounting for uncertainties, a PF photon or PF neutral hadron is formed. Any remaining clusters without linked tracks form PF photons or PF neutral hadrons.

As previously discussed, once the PF particles are identified, additional information about the event can be inferred. A quantity of particular importance to this analysis is the missing transverse energy (MET), defined above. The performance of the PF algorithms determination of the MET is shown in Figure 1.3 by the resolution of PF measured MET as a function of the true MET to be within $\pm 5\%$ above 20 GeV.

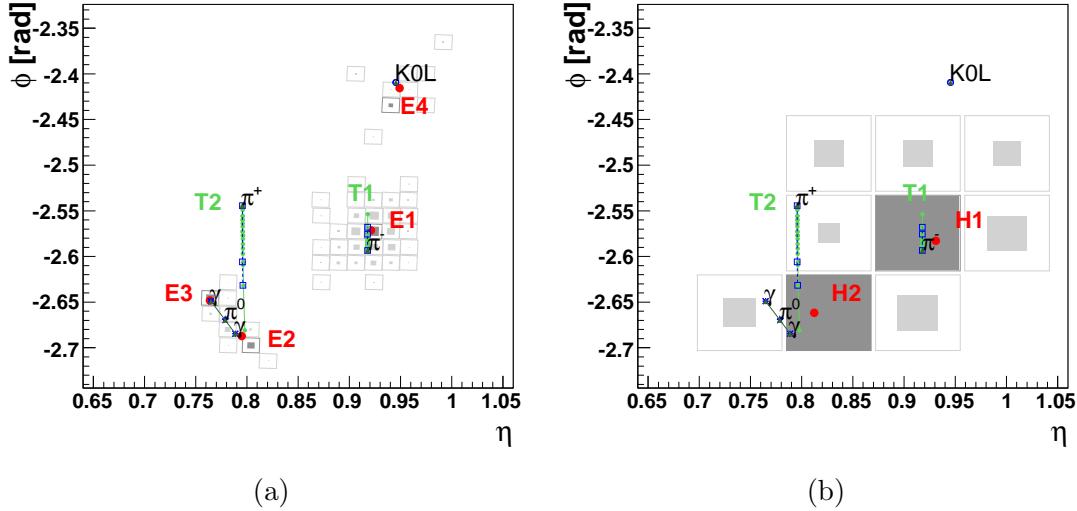


Figure 1.2: Event display of hadronic jet in the $\eta - \phi$ plane for the ECAL (a) and HCAL (b). The locations of clusters are given by the solid dots.

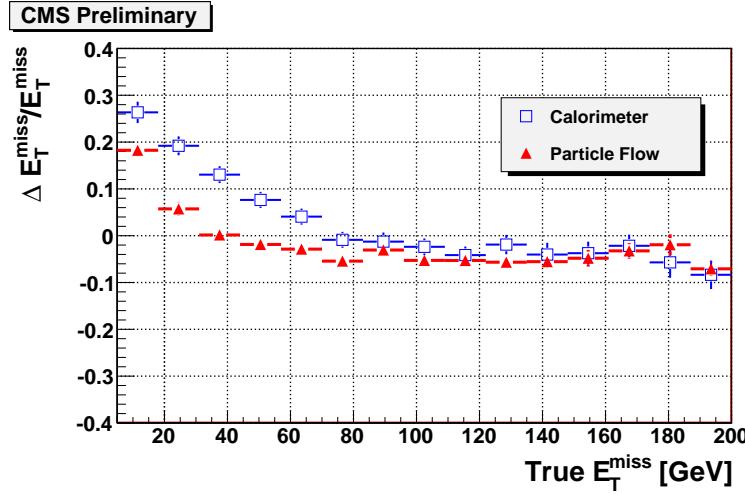


Figure 1.3: MET reconstruction resolution.

1.2 Monte Carlo event simulation

The simulation of proton collision events and their detection using Monte Carlo (MC) techniques is useful for several purposes. In addition to using the simulated events to test the detector hardware and software performance without collecting true data, simulations are used to build background models when searching the data for new physics processes. New physics signatures usually appear as excesses in data above a SM background. The background model consists of SM processes which produce the same or similar signature

as the new signal being searched for. These processes are modelled either using purely simulated events, or a combination of simulated events and data-driven techniques. In either case, it is often necessary to weight the background events by correction scale factors measured using data, which account for shortcomings of the simulations, such as the inability to perform perturbative QCD calculations for low momentum transfer processes. MC event generation can be factored into two parts: modeling the initial particles produced in a collision event and modeling how these initial particles evolve in and interact with the detector.

The first part of MC event generation is modeling the proton-proton collision and the initial particles produced at the primary vertex. Several software packages are used by CMS analysts to generate collision events and calculate the cross sections of the simulated processes, including PYTHIA [?], MADGRAPH [?], BlackHat and Sherpa [?], and POWHEG [?]. The packages have different implementations, but the underlying principles are the same. The momenta of the proton partons (quarks and gluons) that interact in the initial scatter are determined probabilistically by random sampling from the parton distribution functions (PDFs), which give the probability that a parton will carry a fraction x of the proton momentum. This is straightforward for processes with two incoming and one outgoing particle ($2 \rightarrow 1$) and two incoming and two outgoing particles ($2 \rightarrow 2$), in which the outcomes are weighted by their relative cross sections and determined probabilistically. For radiative processes such as ISR and FSR of a photon or gluon, generally $1 \rightarrow 2$ processes, higher order matrix elements must be calculated or approximated. Once the initial particles are determined, their fragmentation and decays are simulated in a process called hadronization, until the final stable particles are produced.

The second part of MC event generation is simulating the detector response to the stable particles produced in the first step, including their interaction with the detector material itself, both active elements and structural material. The primary software package used in this step by CMS is GEANT [?], in which a complete digital representation of the CMS detector is built. GEANT simulates the passage of each each stable particle, step-by-step, outward through the detector, probabilistically determining the interaction

that occurs at each step depending on the particle's energy, material it is in, and the EM field present. Since the detector is not perfectly efficient, both because the acceptance is less than one and the reconstruction efficiency of the individual detector elements is suboptimal, calibration values must be measured at CMS and fed back in to the simulations, so that the performance of the detector can be accurately simulated. Once the final response of the detector is simulated, the resulting MC may be weighted by scale factors measured using real data, in order to correct for mismodelling of the detector.

Chapter 2

Datasets and Monte Carlo Simulations

2.1 Data

2.1.1 Triggers and Datasets

This analysis uses a data sample recorded by the CMS experiment during 2016, corresponding to 35.9 fb^{-1} of data. The datasets are listed in Table 2.1, along with the integrated luminosity. The analysis relies on five different primary datasets (PDs), *DoubleEG*, *DoubleMuon*, *MuEG*, *SingleElectron*, and *SingleMuon*, each of which combines a certain collection of HLT paths. To avoid duplicate events from different primary datasets, events are taken:

- from DoubleEG if they pass the diEle or triEle triggers,
- from DoubleMuon if they pass the diMuon or triMuon triggers and fail the diEle and triEle triggers,
- from MuEG if they pass the MuEle or MuDiEle or DiMuEle triggers and fail the diEle, triEle, diMuon and triMuon triggers,
- from SingleElectron if they pass the singleElectron trigger and fail all the above triggers.
- from SingleMuon if they pass the singleMuon trigger and fail all the above triggers.

The HLT paths used for 2016 collision data are listed in Table 2.2, together with their L1 seed, prescale value and the associated primary dataset.

Run-range	Dataset	Integrated luminosity
273150-275376	/DoubleMuon/Run2016B-23Sep2016-v3/AOD /DoubleEG/Run2016B-23Sep2016-v3/AOD /MuonEG/Run2016B-23Sep2016-v3/AOD /SingleElectron/Run2016B-23Sep2016-v3/AOD /SingleMuon/Run2016B-23Sep2016-v3/AOD	5.892 fb ⁻¹
275656-276283	/DoubleMuon/Run2016C-23Sep2016-v1/AOD /DoubleEG/Run2016C-23Sep2016-v1/AOD /MuonEG/Run2016C-23Sep2016-v1/AOD /SingleElectron/Run2016C-23Sep2016-v1/AOD /SingleMuon/Run2016C-23Sep2016-v1/AOD	2.646 fb ⁻¹
276315-276811	/DoubleMuon/Run2016D-23Sep2016-v1/AOD /DoubleEG/Run2016D-23Sep2016-v1/AOD /MuonEG/Run2016D-23Sep2016-v1/AOD /SingleElectron/Run2016D-23Sep2016-v1/AOD /SingleMuon/Run2016D-23Sep2016-v1/AOD	4.353 fb ⁻¹
276831-277420	/DoubleMuon/Run2016E-23Sep2016-v1/AOD /DoubleEG/Run2016E-23Sep2016-v1/AOD /MuonEG/Run2016E-23Sep2016-v1/AOD /SingleElectron/Run2016E-23Sep2016-v1/AOD /SingleMuon/Run2016E-23Sep2016-v1/AOD	4.117 fb ⁻¹
277932-278808	/DoubleMuon/Run2016F-23Sep2016-v1/AOD /DoubleEG/Run2016F-23Sep2016-v1/AOD /MuonEG/Run2016F-23Sep2016-v1/AOD /SingleElectron/Run2016F-23Sep2016-v1/AOD /SingleMuon/Run2016F-23Sep2016-v1/AOD	3.186 fb ⁻¹
278820-280385	/DoubleMuon/Run2016G-23Sep2016-v1/AOD /DoubleEG/Run2016G-23Sep2016-v1/AOD /MuonEG/Run2016G-23Sep2016-v1/AOD /SingleElectron/Run2016G-23Sep2016-v1/AOD /SingleMuon/Run2016G-23Sep2016-v1/AOD	7.721 fb ⁻¹
281207-284068	/DoubleMuon/Run2016H-PromptReco-v1/AOD /DoubleEG/Run2016H-PromptReco-v1/AOD /MuonEG/Run2016H-PromptReco-v1/AOD /SingleElectron/Run2016H-PromptReco-v1/AOD /SingleMuon/Run2016H-PromptReco-v1/AOD /DoubleMuon/Run2016H-PromptReco-v2/AOD /DoubleEG/Run2016H-PromptReco-v2/AOD /MuonEG/Run2016H-PromptReco-v2/AOD /SingleElectron/Run2016H-PromptReco-v2/AOD /SingleMuon/Run2016H-PromptReco-v2/AOD /DoubleMuon/Run2016H-PromptReco-v3/AOD /DoubleEG/Run2016H-PromptReco-v3/AOD /MuonEG/Run2016H-PromptReco-v3/AOD /SingleElectron/Run2016H-PromptReco-v3/AOD /SingleMuon/Run2016H-PromptReco-v3/AOD	8.857 fb ⁻¹

Table 2.1: Datasets used in the analysis.

HLT path	L1 seed	prescale	primary dataset
HLT_Ele17_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	L1_DoubleEG_15_10	1	DoubleEG
HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	L1_DoubleEG_22_10	1	DoubleEG
HLT_DoubleEle33_CaloIdL_GsfTrkIdVL	(Multiple)	1	DoubleEG
HLT_Ele16_Ele12_Ele8_CaloIdL_TrackIdL	L1_TripleEG_14_10_8	1	DoubleEG
HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL	L1_DoubleMu_11_4	1	DoubleMuon
HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL	L1_DoubleMu_11_4	1	DoubleMuon
HLT_TripleMu_12_10_5	L1_TripleMu_5_5_3	1	DoubleMuon
HLT_Mu8_TrkIsoVVL_Ele17_CaloIdL_TrackIdL_IsoVL	L1_Mu5_EG15	1	MuonEG
HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL	L1_Mu5_EG20	1	MuonEG
HLT_Mu17_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	L1_Mu12_EG10	1	MuonEG
HLT_Mu23_TrkIsoVVL_Ele12_CaloIdL_TrackIdL_IsoVL	L1_Mu20_EG10	1	MuonEG
HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL	L1_SingleMu*	1	MuonEG
HLT_Mu8_DiEle12_CaloIdL_TrackIdL	L1_Mu6_DoubleEG10	1	MuonEG
HLT_DiMu9_Ele9_CaloIdL_TrackIdL	L1_DoubleMu7_EG7	1	MuonEG
HLT_Ele25_eta2p1_WPTight	L1_SingleEG*	1	SingleElectron
HLT_Ele27_WPTight	L1_SingleEG*	1	SingleElectron
HLT_Ele27_eta2p1_WPLoose_Gsf	L1_SingleEG*	1	SingleElectron
HLT_IsoMu20 OR HLT_IsoTkMu20	L1_SingleMu*	1	SingleMuon
HLT_IsoMu22 OR HLT_IsoTkMu22	L1_SingleMu*	1	SingleMuon

Table 2.2: Trigger paths used in 2016 collision data.

2.1.2 Trigger Efficiency

The efficiency in data of the combination of triggers used in the analysis with respect to the offline reconstruction and selection is measured by considering 4ℓ events triggered by single lepton triggers. One of the four reconstructed leptons (the “tag”) is geometrically matched to a trigger object passing the final filter of one of the single muon or single electron triggers. The other three leptons are used as “probes”. In each 4ℓ event there are up to 4 possible tag-probe combinations, and all possible combinations are counted in the denominator of the efficiency. For each of the three probe leptons all matching trigger filter objects are collected. Then the matched trigger filter objects of the three probe leptons are combined in attempt to reconstruct any of the triggers used in the analysis. If any of the analysis triggers can be formed using the probe leptons, the set of probes is also counted in the numerator of the efficiency.

This method does not have a perfect closure in MC events due to the fact that the presence of a fourth lepton increases the trigger efficiency, and this effect is not accounted for. Also, in the $2e2\mu$ final state, the three probe leptons cannot be combined to form all possible triggers which can collect events with two electrons and two muons (e.g. if the tag lepton is an electron, the three remaining leptons cannot pass a double electron trigger). Therefore the method is also exercised on MC and the difference between data

and MC is used to determine the reliability of the simulation. The efficiency plotted as a function of the minimum p_T of the three probe leptons in data and MC using this method can be seen in Fig.2.1. The MC efficiency describes well the data within the statistical uncertainties.

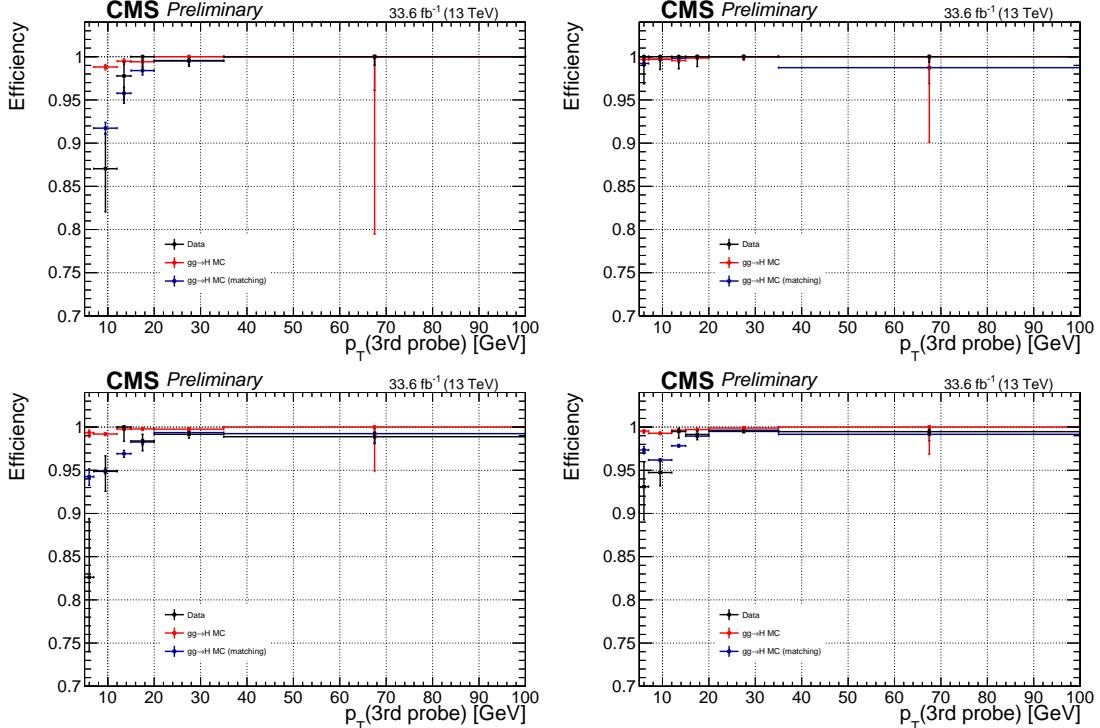


Figure 2.1: Trigger efficiency measured in data using 4ℓ events collected by single lepton triggers for the $4e$ (top left), 4μ (top right), $2e2\mu$ (bottom left) and 4ℓ (bottom right) final states.

A summary of the trigger efficiencies in MC truth, and in MC and data using the tag and probe method are summarized in table 2.3. The trigger efficiency in simulation is found to be $> 99\%$ in each final state.

2.2 Simulation

2.2.1 Signal Samples

The signal samples used are centrally produced for the benchmarks defined in the previous section and are summarized in Table 2.4.

Final State	$gg \rightarrow H$ MC	$gg \rightarrow H$ MC (matching)	Data (matching)
4e	$0.991^{+.002}_{-.002}$	$0.948^{+.004}_{-.004}$	$0.982^{+.005}_{-.007}$
4 μ	$0.997^{+.001}_{-.001}$	$0.997^{+.001}_{-.001}$	$1.000^{+.000}_{-.001}$
2e2 μ	$0.995^{+.001}_{-.001}$	$0.964^{+.002}_{-.002}$	$0.983^{+.003}_{-.004}$

Table 2.3: Trigger efficiencies measured using 4ℓ events.

Dataset	Parameters
/ZprimeToA0hToA0chichihZZTo4l_2HDM_MZp-*_MA0-300_13TeV-madgraph-pythia8/[1]	$m_{A^0} = 300$ GeV
/ZprimeToA0hToA0chichihZZTo4l_2HDM_MZp-*_MA0-*_13TeV-madgraph/[2]	$m_{A^0} \neq 300$ GeV
/MonoHZZ4l_ZpBaryonic_MZp-*_MChi-*_13TeV-madgraph/[3]	
[1] RunIISpring16DR80-premix_withHLT_80X_mcRun2_asymptotic_v14-v1/AODSIM	
[2] RunIISpring16reHLT80-PUSpring16RAWAODSIM_reHLT_80X_mcRun2_asymptotic_v14-v1/AODSIM	
[3] RunIISpring16DR80-premix_withHLT_80X_mcRun2_asymptotic_v14-v1/AODSIM	

Table 2.4: Benchmark signal samples analyzed.

2.2.2 Background Samples

Descriptions of the SM Higgs boson production are obtained using the POWHEG V2 [?, ?, ?] generator for the five main production modes: gluon fusion ($gg \rightarrow H$) including quark mass effects [?], vector boson fusion (VBF) [?], and associated production (WH, ZH and $t\bar{t}H$ [?]). In the case of WH and ZH the MiNLO HVJ extension of POWHEG is used [?]. The description of the decay of the Higgs boson to four leptons is obtained using the JHUGEN generator [?]. In the case of WH, ZH and $t\bar{t}H$, the Higgs boson is allowed to decay to $H \rightarrow ZZ \rightarrow 2\ell 2X$ such that 4-lepton events where two leptons originate from the decay of associated Z , W bosons or top quarks are also taken into account in the simulation. Showering of parton-level events is done using PYTHIA8.209, and in all cases matching is performed by allowing QCD emissions at all energies in the shower and vetoing them afterwards according to the POWHEG internal scale. All samples are generated with the NNPDF 3.0 NLO parton distribution functions (PDFs) [?]. The list of Higgs signal samples and their cross sections are shown in Table 2.5.

Production of ZZ via quark-antiquark annihilation is generated at next-to-leading order (NLO) using POWHEG V2 [?] and PYTHIA8, with the same settings as for the Higgs signal. As this simulation covers a large range of ZZ invariant masses, dynamical QCD

Process	Dataset Name	$\sigma \times BR(\times \epsilon_{\text{filter}})$
$gg \rightarrow H \rightarrow ZZ \rightarrow 4\ell$	/GluGluHToZZTo4L_M125_13TeV_powheg2_JHUGenV6_pythia8	12.18 fb
$qq \rightarrow Hqq \rightarrow ZZqq \rightarrow 4\ell qq$	/VBF_HToZZTo4L_M125_13TeV_powheg2_JHUGenV6_pythia8	1.044 fb
$q\bar{q} \rightarrow W^+H \rightarrow W^+ZZ \rightarrow 4\ell + X$	/WplusH_HToZZTo4L_M125_13TeV_powheg2-minlo-HWJ_JHUGenV6_pythia8	0.232 fb
$q\bar{q} \rightarrow W^-H \rightarrow W^-ZZ \rightarrow 4\ell + X$	/WminusH_HToZZTo4L_M125_13TeV_powheg2-minlo-HWJ_JHUGenV6_pythia8	0.147 fb
$q\bar{q} \rightarrow ZH \rightarrow ZZZ \rightarrow 4\ell + X$	/ZH_HToZZ_4LFilter_M125_13TeV_powheg2-minlo-HZJ_JHUGenV6_pythia8	0.668 fb
$gg \rightarrow ttH \rightarrow ttZZ \rightarrow 4\ell + X$	/ttH_HToZZ_4LFilter_M125_13TeV_powheg_JHUGen_pythia8	0.393 fb

Table 2.5: Higgs signal samples and cross sections.

Process	Dataset Name	$\sigma \cdot BR$
$qq \rightarrow ZZ \rightarrow 4\ell$	/ZZTo4L_13TeV_powheg_pythia8	1.256pb
$qq \rightarrow ZZ \rightarrow 4\ell$	/ZZTo4L_13TeV-amcatnloFXFX-pythia8	1.212pb
$gg \rightarrow ZZ \rightarrow 4e$	/GluGluToContinToZZTo4e_13TeV_MCFM701	0.00159pb
$gg \rightarrow ZZ \rightarrow 4\mu$	/GluGluToContinToZZTo4mu_13TeV_MCFM701	0.00159pb
$gg \rightarrow ZZ \rightarrow 4\tau$	/GluGluToContinToZZTo4tau_13TeV_MCFM701	0.00159pb
$gg \rightarrow ZZ \rightarrow 2e2\mu$	/GluGluToContinToZZTo2e2mu_13TeV_MCFM701	0.00319pb
$gg \rightarrow ZZ \rightarrow 2e2\tau$	/GluGluToContinToZZTo2e2tau_13TeV_MCFM701	0.00319pb
$gg \rightarrow ZZ \rightarrow 2\mu2\tau$	/GluGluToContinToZZTo2mu2tau_13TeV_MCFM701	0.00319pb
$Z \rightarrow \ell\ell + \text{jets}$	/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	6104pb
$Z \rightarrow \ell\ell + \text{jets}$	/DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	18610pb
$WZ \rightarrow 3\ell\nu$	/WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8	4.430pb
$t\bar{t}$	/TTJets_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	815.96pb
$t\bar{t} \rightarrow 2\ell 2\nu 2b$	/TTTo2L2Nu_13TeV-powheg	87.31pb

Table 2.6: Background Monte Carlo samples and cross sections.

factorization and renormalization scales have been chosen, equal to m_{ZZ} .

The $gg \rightarrow ZZ$ process is simulated at leading order (LO) with MCFM [2, 4]. In order to match the $gg \rightarrow H \rightarrow ZZ$ transverse momentum spectra predicted by POWHEG at NLO, the showering for MCFM samples is performed with different PYTHIA8 settings, allowing only emissions up to the parton-level scale (“wimpy” shower).

Although not directly used to model data observations, additional MC samples of WZ , Drell-Yan+jets, $t\bar{t}$, and tribosons are generated using MADGRAPH5_AMCATNLO [?] either inclusively or merging several jet multiplicities, as detailed in the table. Table 2.6 summarizes the MC simulation datasets used for this analysis.

2.2.2.1 Pileup Reweighting

The MC samples are reweighted to match the pileup distribution measured in 2016 data. Scale factors are measured and applied to each event weight before histograms are filled and yields are calculated, based on the number of pileup vertices present in the event. The mean number of pileup vertices for data measured in 2016 is about 20. Figure 2.2 shows the distributions of the numbers of pileup vertices for data and MC before and after the events are reweighted.

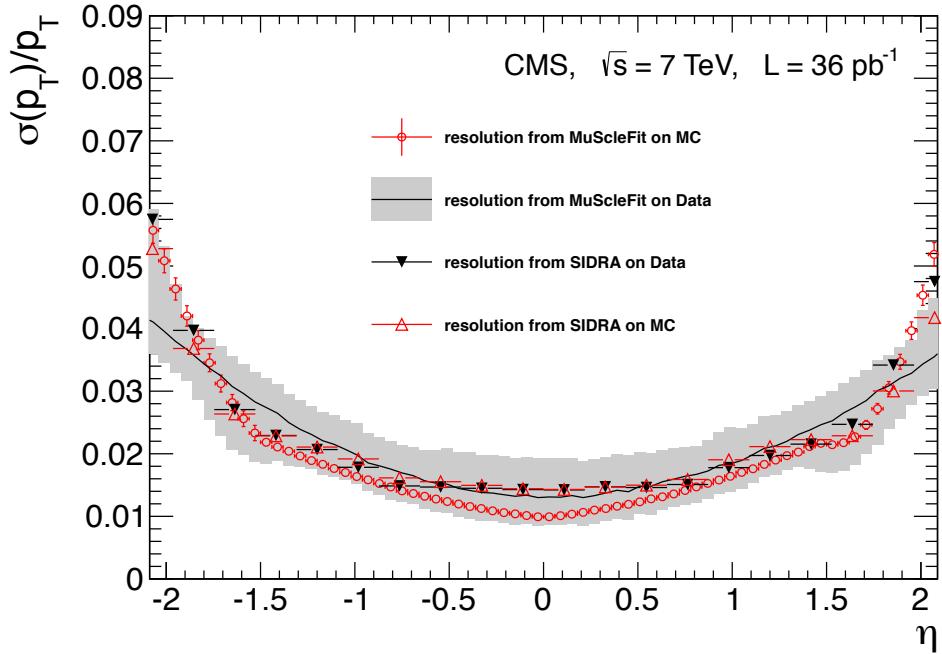


Figure 2.2: Number of pileup vertices before and after reweighted is applied.

Chapter 3

Physics Objects

3.1 Electrons

3.1.1 Electron Reconstruction

Electron candidates are preselected using loose cuts on track-cluster matching observables, so as to preserve the highest possible efficiency while rejecting part of the QCD background. To be considered for the analysis, electrons are required to have a transverse momentum $p_T^e > 7$ GeV, a reconstructed $|\eta^e| < 2.5$, and to satisfy a loose primary vertex constraint defined as $d_{xy} < 0.5$ and $d_z < 1$. Such electrons are called **loose electrons**.

The early runs in the 2016 data-taking exhibit an tracking inefficiency originating from a reduced hit reconstruction efficiency in the strip detector (“HIP” effect). The resulting data-MC discrepancy is corrected using scale factors as is done for the electron selection with data efficiencies measured using the same tag-and-probe technique outlined later (see Section 3.1.5). These studies are carried out by the EGM POG and the results are summarized here.

The electron reconstruction scale factors are shown Fig. 3.1 as a one-dimensional function of the super cluster η only, as it was shown that the p_T dependence of the scale factor is negligible. More details on electron reconstruction can be found in Ref. [?].

3.1.2 Electron Identification

Reconstructed electrons are identified by means of a Gradient Boosted Decision Tree (GBDT) multivariate classifier algorithm, which exploits observables from the electro-

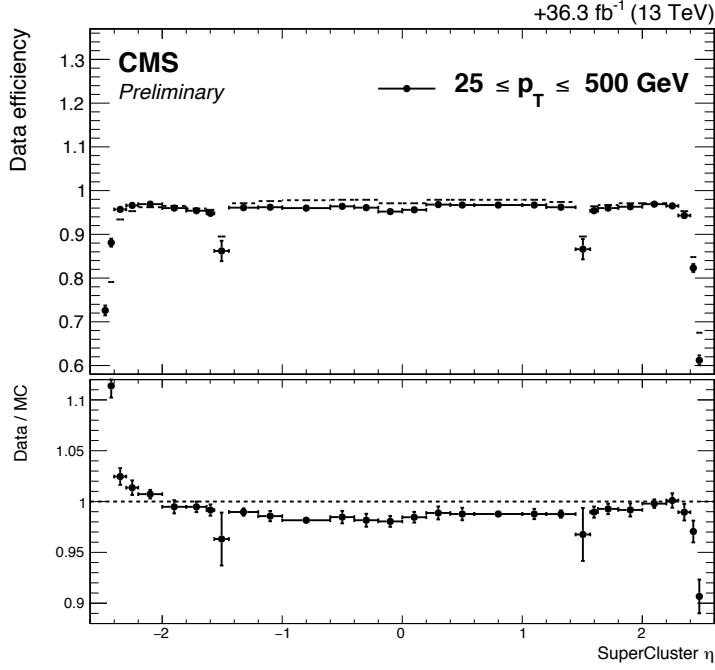


Figure 3.1: Electron reconstruction efficiencies efficiency in data versus η and data/MC scale factors as provided by the EGM POG.

magnetic cluster, the matching between the cluster and the electron track as well as observables based exclusively on tracking measurements. The BDT has been retrained using CMSSW_8_0_X samples. The classifier is trained on Drell-Yan plus jets MC sample for both signal and background:

The impact of the retraining of the ID for the 2016 conditions is illustrated in the ROC curves shown in Fig. 3.2. Several studies to improve the performance of the MVA for the harsher 2016 running conditions were performed. One study considered a new splitting of the BDT training bins, where electrons falling into the gap regions of the ECAL, e.g. the EB-EE transition region, were trained separately from the non-gap electrons. No improvement for either population is observed, indicating that the current setup is already able to properly take the significantly differing input distributions in those regions into account. Also studied where additional variables including more cluster-shape observables. None of these variables helped to improve the performance in the relevant $> 95\%$ signal efficiency regime, though up to 20% improved background rejection was seen for 80%

working points. Finally, the hyper-parameters of the MVA were systematically scanned for their optimal values. The resulting configuration was found to improve the overall performance only marginally by < 10%, however, introducing a significant overtraining effect. Due to the small gains and large overtraining, it was decided to not modify the hyper-parameters beyond the interface changes coming from changing to the latest 4.2.0 version of the TMVA package.

Figure 3.3 shows the output of the BDT on the training and testing samples for true and fake electrons for the high- p_T training bin in the endcap. The good agreement between the training and testing distributions is similar across the 6 training bins and indicates that the classifier has not been overtrained.

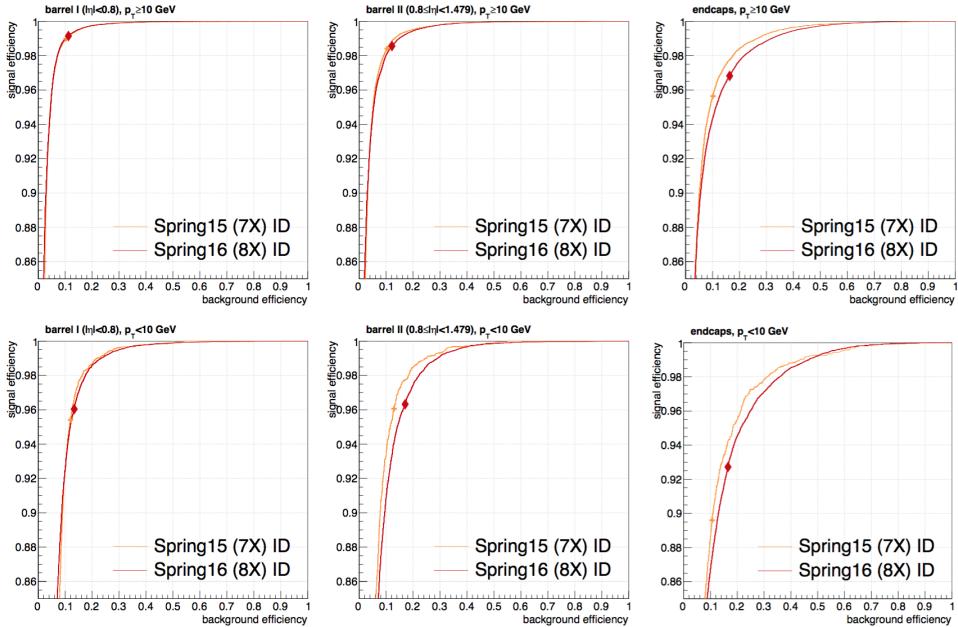


Figure 3.2: Performance comparison of the MVA trained for the 2015 analysis and the retraining for 2016 conditions. The respective working points are indicated by the markers.

Table 3.1 summarizes the full list of observables used as input to the classifier and table 3.2 lists the cut values applied to the BDT score for the chosen working point. For the analysis, we define **tight electrons** as the loose electrons that pass this MVA identification working point.

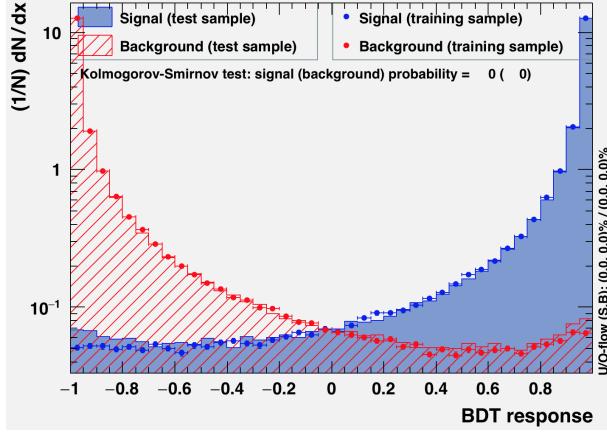


Figure 3.3: BDT output for the training and testing sample for true and fake electrons in the high- p_T endcap training bins.

observable type	observable name
cluster shape	RMS of the energy-crystal number spectrum along η and φ ; $\sigma_{i\eta i\eta}$, $\sigma_{i\varphi i\varphi}$
	super cluster width along η and ϕ
	'ratio of the hadronic energy behind the electron supercluster to the supercluster energy, H/E
	circularity $(E_{5 \times 5} - E_{5 \times 1})/E_{5 \times 5}$
track-cluster matching	sum of the seed and adjacent crystal over the super cluster energy R_9
	for endcap traing bins: energy fraction in pre-shower E_{PS}/E_{raw}
tracking	energy-momentum agreement E_{tot}/p_{in} , E_{ele}/p_{out} , $1/E_{tot} - 1/p_{in}$
	position matching $\Delta\eta_{in}$, $\Delta\varphi_{in}$, $\Delta\eta_{seed}$
	fractional momentum loss $f_{brem} = 1 - p_{out}/p_{in}$
	number of hits of the KF and GSF track N_{KF} , N_{GSF} (·)
	reduced χ^2 of the KF and GSF track χ^2_{KF} , χ^2_{GSF}
	number of expected but missing inner hits (·)
	probability transform of conversion vertex fit χ^2 (·)

Table 3.1: Overview of input variables to the identification classifier. Variables not used in the run I MVA are marked with (·).

minimum BDT score	$ \eta < 0.8$	$0.8 < \eta < 1.479$	$ \eta > 1.479$
$5 < p_T < 10$ GeV	-0.211	-0.396	-0.215
$p_T > 10$ GeV	-0.870	-0.838	-0.763

Table 3.2: Minimum BDT score required for passing the electron identification.

3.1.3 Electron Isolation

The relative isolation for electrons is defined as:

$$\text{RelPFiso} = \left(\sum_{\text{charged}} p_T + \sum_{\text{neutral}}^{\text{corr}} p_T \right) / p_T^{\text{lepton}}. \quad (3.1)$$

where the corrected neutral component of isolation is computed using the formula :

$$\sum_{\text{neutral}}^{\text{corr}} p_T = \max\left(\sum_{\text{neutral}}^{\text{uncorr}} p_T - \rho \times A_{\text{eff}}, 0\text{GeV}\right). \quad (3.2)$$

and the mean pile-up contribution to the isolation cone is obtained as :

$$PU = \rho \times A_{\text{eff}} \quad (3.3)$$

where ρ is the mean energy density in the event and the effective area A_{eff} is defined as the ratio between the slope of the average isolation and that of ρ as a function of the number of vertices.

The electron isolation working point was optimized in Ref. [?] and the electron isolation working was chosen to be $\text{RelPFiso}(\Delta R = 0.3) < 0.35$.

3.1.4 Electron Energy Calibrations

Electrons in data are corrected for features in ECAL energy scale in bins of p_T and $|\eta|$. Corrections are calculated on a $Z \rightarrow ee$ sample to align the dielectron mass spectrum in the data to that in the MC, and to minimize its width.

The $Z \rightarrow ee$ mass resolution in Monte Carlo is made to match data by applying a pseudorandom Gaussian smearing to electron energies, with Gaussian parameters varying in bins of p_T and $|\eta|$. This has the effect of convoluting the electron energy spectrum with a Gaussian.

The electron energy scale is measured in data by fitting a Crystall-ball function to the di-electron mass spectrum around the Z peak in the $Z + \ell$ control region. The energy scale for the full 2016 dataset is shown in Fig. 3.4(a) and agrees with the MC with 100 MeV. The stability of the energy scale across different run periods is shown in Fig. 3.4(b), where the data is binned into approximately 500 pb luminosity blocks.

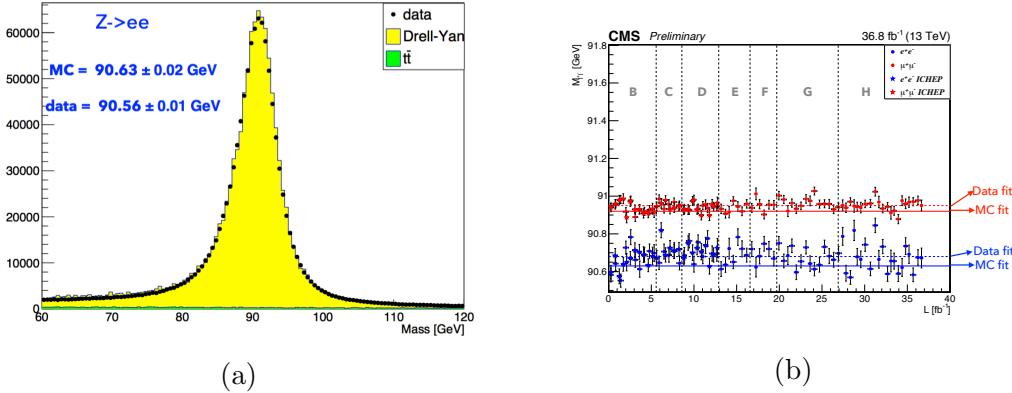


Figure 3.4: (a): electron energy scale measured in the $Z + \ell$ control region for EB and EE electrons. The results of the Crystall-ball fit are reported in the figure. (b): lepton energy scales per 500 pb luminosity block.

3.1.5 Electron Efficiency Measurements

The Tag-and-Probe study was performed on the single electron primary datasets listed in table 2.1 using the same golden JSON of 36.8 fb^{-1} as for the main analysis. More details on the Tag-and-Probe method can be found in Ref. [?].

Tag electrons need to satisfy the following quality requirements:

- trigger matched to HLT_Ele27_eta2p1_WPTight_Gsf_v*
- $p_T > 30 \text{ GeV}$, super cluster (SC) $\eta < 2.1$ but on in EB-EE gap ($1.4442 < |\eta| < 1.566$)
- tight working point of the Spring16 cut-based electron ID

Probe electrons only need to be reconstructed as GsfElectron. The FSR recovery algorithm used in the main analysis is used consistently throughout the efficiency measurement: the isolations are calculated without any FSR photons matched to electrons and the probe electron p_{Tas} as well as the di-electron invariant mass include the FSR photons, if any.

The nominal MC efficiencies are evaluated from the LO MadGraph Drell-Yan sample, while the NLO systematics use the 0,1 jet MadGraph_AMCatNLO sample listed in Table 2.6.

In contrast to previous efficiency measurements, a template fit is used here. The m_{ee} signal shape of the passing and failing probes is taken from MC and convoluted with a

Gaussian. The data is then fitted with the convoluted MC template and a CMSShape (an Error-function with a one-sided exponential tail). This change follows from the usage of the new T&P tool developed by the EGM POG.

The electron selection efficiency is measured as a function of the probe electron p_T and its SC η , and separately for electrons falling in the ECAL gaps. Figure 3.5 shows the p_T turn-on curves measured in data, and the final 2D scale factor is shown in Fig. 3.6 together with the systematic uncertainties. These scale factors are very similar to the ICHEP figures, but more homogenous across η and p_T because of the higher statistics and the usage of more stable fitting routines in the new T&P tool.

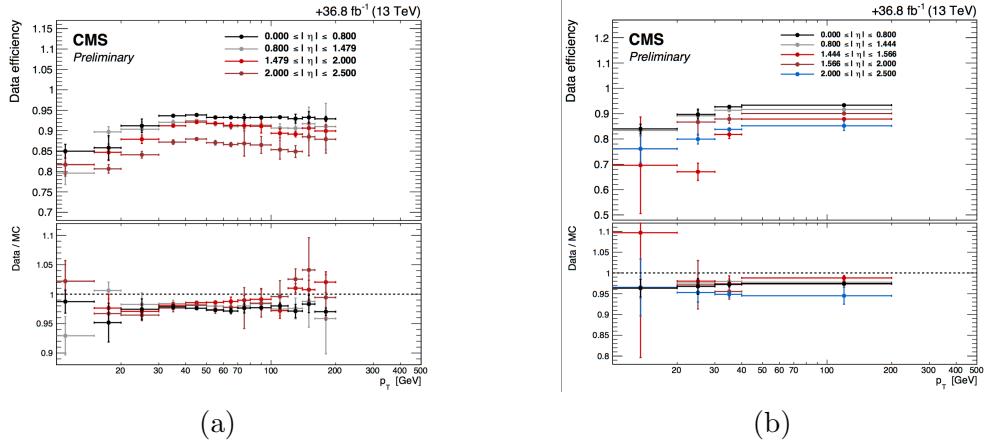


Figure 3.5: Electron selection efficiencies measured using the Tag-and-Probe technique described in the text, non-gap electrons (left) and gap electrons (right).

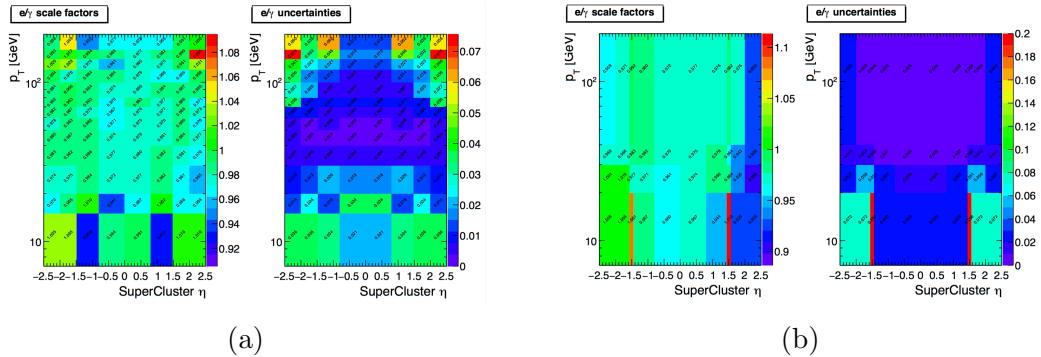


Figure 3.6: Electron selection efficiencies measured using the Tag-and-Probe technique described in the text, non-gap electrons (top) and gap electrons (bottom)

The EGM recommendations on the evaluation of Tag-and-Probe uncertainties for ef-

ficiency measurements are followed. Specifically,

- Variation of the signal shape from a MC shape to an analytic shape (Crystal Ball) fitted to the MC
- Variation of the background shape from a CMS-shape to a simple exponential in fits to data
- Variation of the tag selection: tag $p_T > 35$ GeV and passes MVA-based 8X ID
- Using an NLO MC sample for the signal templates

The total uncertainty for the measurement of the scale factors is the quadratic sum of the statistical uncertainties returned from the fit and the aforementioned systematic uncertainties.

3.2 Muons

3.2.1 Muon Reconstruction and Identification

More details on muon reconstruction can be found in Ref. [?]. We define **loose muons** as the muons that satisfy $p_T > 5$, $|\eta| < 2.4$, $d_{xy} < 0.5$, $d_z < 1$, where d_{xy} and d_z are defined w.r.t. the PV and using the ‘muonBestTrack’. Muons have to be reconstructed by either the Global Muon or Tracker Muon algorithm. Standalone Muon tracks that are only reconstructed in the muon system are rejected. Muons with `muonBestTrackType==2` (standalone) are discarded even if they are marked as global or tracker muons.

Loose muons with p_T below 200GeV are considered **tight muons** if they also pass the PF muon ID (note that the naming convention used for these IDs differs from the muon POG naming scheme, in which the “tight ID” used here is called the “loose ID”). Loose muons with p_T above 200GeV are considered tight muons if they pass the PF ID or the Tracker High- p_T ID, the definition of which is shown in Table 3.3. This relaxed definition is used to increase signal efficiency for the high-mass search. When a very heavy resonance decays to two Z bosons, both bosons will be very boosted. In the lab frame, the leptons coming from the decay of a highly boosted Z will be nearly collinear, and the PF ID loses

efficiency for muons separated by approximately $\Delta R < 0.4$, which roughly corresponds to muons originating from Z bosons with $p_T > 500\text{GeV}$.

Plain-text description	Technical description
Muon station matching	Muon is matched to segments in at least two muon stations NB: this implies the muon is an arbitrated tracker muon.
Good p_T measurement	$\frac{p_T}{\sigma_{p_T}} < 0.3$
Vertex compatibility ($x - y$)	$d_{xy} < 2 \text{ mm}$
Vertex compatibility (z)	$d_z < 5 \text{ mm}$
Pixel hits	At least one pixel hit
Tracker hits	Hits in at least six tracker layers

Table 3.3: The requirements for a muon to pass the Tracker High- p_T ID. Note that these are equivalent to the Muon POG High- p_T ID with the global track requirements removed.

An additional “ghost-cleaning” step is performed to deal with situations when a single muon can be incorrectly reconstructed as two or more muons:

- Tracker Muons that are not Global Muons are required to be arbitrated.
- If two muons are sharing 50% or more of their segments then the muon with lower quality is removed.

3.2.2 Muon Isolation

Particle-Flow based isolation, described for electrons in section 3.1.3, is also used for the muons. The only difference with electrons is the way the pileup contribution is subtracted: for the muons, $\Delta\beta$ correction is applied, whereby $\Delta\beta = \frac{1}{2} \sum_{\text{PU}}^{\text{charged had.}} p_T$ gives an estimate of the energy deposit of neutral particles (hadrons and photons) from pile-up vertices. The relative isolation for muons is then defined as:

$$\text{RelPFiso} = \frac{\sum^{\text{charged had.}} p_T + \max(\sum^{\text{neutral had.}} E_T + \sum^{\text{photon}} E_T - \Delta\beta, 0)}{p_T^{\text{lepton}}} \quad (3.4)$$

The isolation working point for muons was optimized in Ref. [?] and the working point was chosen to be the same as electrons, namely $\text{RelPFiso}(\Delta R = 0.3) < 0.35$.

3.2.3 Muon Efficiency Measurements

Muon efficiencies are measured with the Tag and Probe (T&P) method performed on $Z \rightarrow \mu\mu$ and $J/\Psi \rightarrow \mu\mu$ events in bins of p_T and η . More details on the methodology can be found in Ref. [?]. The Z sample is used to measure the muon reconstruction and identification efficiency at high p_T , and the efficiency of the isolation and impact parameter requirements at all p_T . The J/Ψ sample is used to measure the reconstruction efficiency at low p_T , as it benefits from a better purity in that kinematic regime. In this case, events are collected using `HLT_Mu7p5_Track2_Jpsi_v*` when probing the reconstruction and identification efficiency in the muon system, and using the `HLT_Mu7p5_L2Mu2_Jpsi_v*` when probing the tracking efficiency.

Reconstruction and identification Results for the muon reconstruction and identification efficiency for $p_T > 20\text{GeV}$ have been derived by the Muon POG. The probe in this measurement are tracks reconstructed in the inner tracker, and the passing probes are those that are also reconstructed as a global or tracker muon and passing the Muon POG Loose muon identification. Results for low p_T muons were derived using J/Ψ events, with the same definitions of probe and passing probes. The systematic uncertainties are estimated by varying the analytical signal and background shape models used to fit the dimuon invariant mass. Details on the procedure can be found in Ref. [?]. The efficiency and scale factors used for low p_T muons are the ones derived using single muon prompt-reco dataset.

The efficiency in data and simulation is shown in Fig. 3.7.

Impact parameter requirements The measurement is performed using Z events. Events are selected with `HLT_IsoMu20_v*` or `HLT_IsoMu22_v*` triggers. For this measurement, the probe is a muon passing the POG Loose identification criteria, and it is considered a passing probe if satisfies the SIP3D, dxy , dz cuts of this analysis. The results are shown in Fig. 3.8.

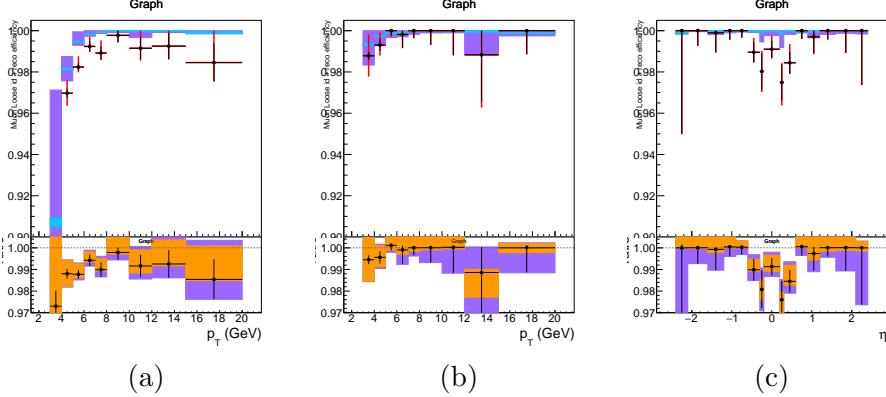


Figure 3.7: Muon reconstruction and identification efficiency at low p_T , measured with the tag&probe method on J/Ψ events, as function of p_T in the barrel (left) and endcaps (center), and as function of η for $p_T > 7\text{GeV}$ (right). In the upper panel, the larger error bars include also the systematical uncertainties, while the smaller ones are purely statistical. In the lower panel showing the ratio of the two efficiencies, the black error bars are for the statistical uncertainty, the orange rectangles for the systematical uncertainty and the violet rectangles include both uncertainties.

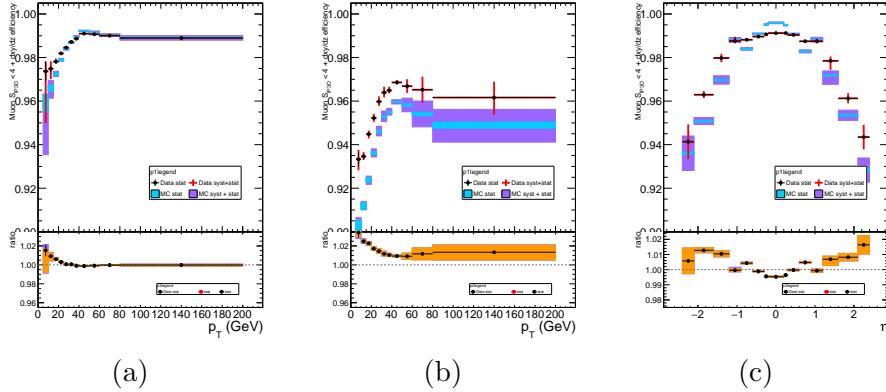


Figure 3.8: Efficiency of the muon impact parameter requirements, measured with the tag&probe method on Z events, as function of p_T in the barrel (left) and endcaps (center), and as function of η for $p_T > 20\text{GeV}$ (right). In the upper panel, the larger error bars include also the systematical uncertainties, while the smaller ones are purely statistical. In the lower panel showing the ratio of the two efficiencies, the black error bars are for the statistical uncertainty, the orange rectangles for the systematical uncertainty and the violet rectangles include both uncertainties.

Isolation requirements The isolation efficiency is measured using events from the Z decay for any p_T , selected with either of `HLT_IsoMu20_v*` or `HLT_IsoMu22_v*` triggers. The isolation of the muons are calculated after recovery of the FSR photons and subtracting their contribution to the isolation cone of the muons. More detailed description

of the method can be found in Ref. [?].

The results are shown in Fig. 3.9.

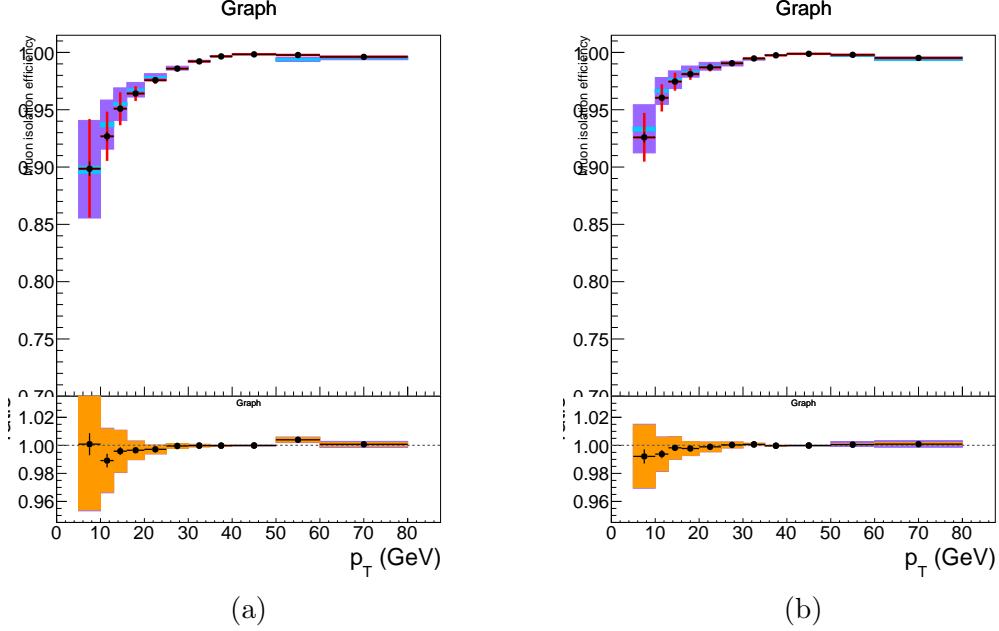


Figure 3.9: Efficiency of the muon isolation requirement, measured with the tag&probe method on Z events, as function of p_T in the barrel (left) and endcaps (right). In the upper panel, the larger error bars include also the systematical uncertainties, while the smaller ones are purely statistical. In the lower panel showing the ratio of the two efficiencies, the black error bars are for the statistical uncertainty, the orange rectangles for the systematical uncertainty and the violet rectangles include both uncertainties.

Tracking The efficiency to reconstruct a muon track in the inner detector is measured using as probes tracks reconstructed in the muon system alone. The method for measuring the tracking efficiency is the same as in Ref. [?], and the results on 2016 data are briefly discussed here. The efficiency and data to mc scale factors are measured from Z events as a function of η for $p_T > 10\text{GeV}$ and $p_T < 10\text{GeV}$. The values of data to mc scale factors used are from the ReReco version of the full dataset collected in 2016.

The tracking efficiency in data and simulation as a function of η is shown in Fig. 3.10.

Overall results The product of all the data to simulation scale factors for muon tracking, reconstruction, identification, impact parameter and isolation requirements is shown in Fig. 3.11.

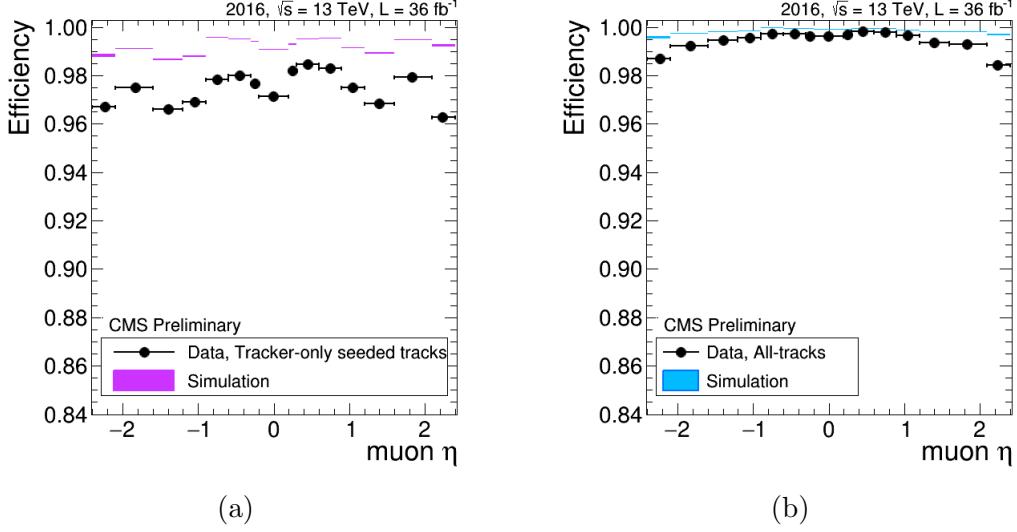


Figure 3.10: Tracking efficiency in data and simulation as a function of η for muon $p_T < 10 \text{ GeV}$ (left) and $p_T > 10 \text{ GeV}$ (right) with ReReco data.

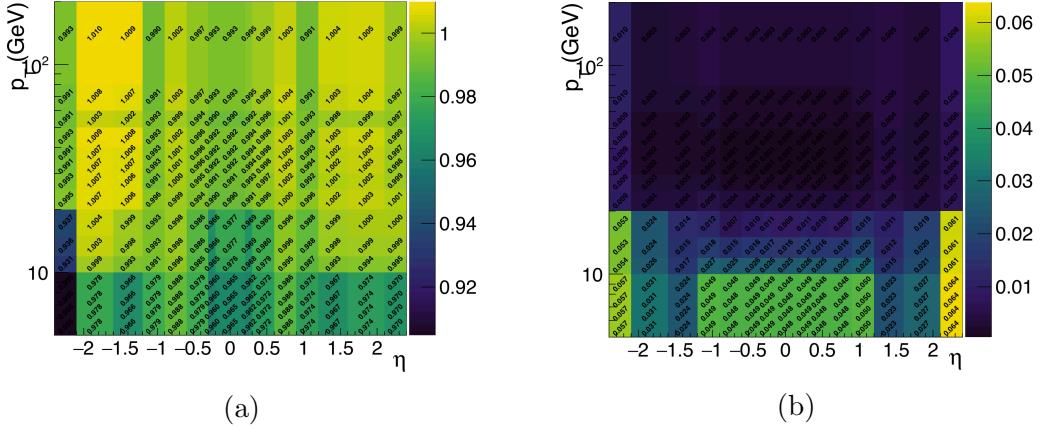


Figure 3.11: Left: Overall data to simulation scale factors for muons, as function of p_T and η . Right: Uncertainties on data to simulation scale factors for muons, as function of p_T and η .

3.3 Photons for FSR recovery

The FSR recovery algorithm was considerably simplified with respect to what was done in Run I, while maintaining similar performance. The selection of FSR photons is now only done per-lepton and no longer depends on any Z mass criterion, thus much simplifying the subsequent ZZ candidate building and selection. As regards the association of photons with leptons, the rectangular cuts on $\Delta R(\gamma, l)$ and $E_{T,\gamma}$ have been replaced by a cut on $\Delta R(\gamma, l)/E_{T,\gamma}^2$.

Starting from the collection of 'PF photons' provided by the particle-flow algorithm, the selection of photons and their association to a lepton proceeds as follows:

1. The preselection of PF photons is done by requiring $p_{T,\gamma} > 2$ GeV, $|\eta^\gamma| < 2.4$, and a relative Particle-flow isolation smaller than 1.8. The latter variable is computed using a cone of radius $R = 0.3$, a threshold of 0.2 GeV on charged hadrons with a veto cone of 0.0001, and 0.5 GeV on neutral hadrons and photons with a veto cone of 0.01, also including the contribution from pileup vertices (with the same radius and threshold as per charged isolation) .
2. Supercluster veto: we remove all PF photons that match with any electron passing both the loose ID and SIP cuts. The matching is performed by directly associating the two PF candidates.
3. Photons are associated to the closest lepton in the event among all those pass both the loose ID and SIP cuts.
4. We discard photons that fail the cuts $\Delta R(\gamma, l)/E_{T,\gamma}^2 < 0.012$, and $\Delta R(\gamma, l) < 0.5$.
5. If more than one photon is associated to the same lepton, the lowest- $\Delta R(\gamma, l)/E_{T,\gamma}^2$ is selected.
6. For each FSR photon that was selected, we exclude that photon from the isolation sum of all the leptons in the event that pass both the loose ID and SIP cuts. This concerns the photons that are in the isolation cone and outside the isolation veto of said leptons ($\Delta R < 0.4$ AND $\Delta R > 0.01$ for muons and $\Delta R < 0.4$ AND $(\eta^{\text{SC}} < 1.479 \text{ OR } \Delta R > 0.08)$ for electrons).

More details on the optimization of the FSR photon selection can be found in Ref. [?, ?].

3.4 Jets

Vector Boson Fusion (VBF) and other production mechanisms of Higgs Boson normally differ as regards the jet kinematics. In this analysis, jets are thus used for the event

categorization, which will be introduced in Section ??.

3.4.0.1 Jet Identification

Jets are reconstructed by using the anti- k_T clustering algorithm out of particle flow candidates, with a distance parameter $R = 0.4$, after rejecting the charged hadrons that are associated to a pileup primary vertex.

To reduce instrumental background, the loose working point jet ID suggested by the JetMET Physics Object Group is applied. In this analysis, the jets are required to be within $|\eta| < 4.7$ area and have a transverse momentum above 30 GeV. In addition, the jets are cleaned from any of the tight leptons (passing the SIP and isolation cut computed after FSR correction) and FSR photons by a separation criterion: $\Delta R(\text{jet}, \text{lepton}/\text{photon}) > 0.4$.

3.4.0.2 Jet Energy Corrections

The calorimeter response to particles is not linear and it is not straightforward to translate the measured jet energy to the true particle or parton energy, therefore we need Jet Energy Corrections. In this analysis, standard jet energy corrections are applied to the reconstructed jets, which consist of L1 Pileup, L2 Relative Jet Correction, L3 Absolute Jet Correction for both Monte Carlo samples and data, and also residual calibration for data.

3.4.0.3 B-tagging

For categorization purpose, we need to distinguish whether a jet is b-jet or not. The *Combined Secondary Vertex* algorithm is used as our b-tagging algorithm. It combines information about impact parameter significance, the secondary vertex and jet kinematics. The variables are combined using a likelihood ratio technique to compute the b-tag discriminator. In this analysis, a jet is considered to be b-tagged if it passes the *CSVv2M* working point, i.e. if its `pfCombinedInclusiveSecondaryVertexV2BJetTags` discriminator is greater than 0.8484 [?].

Data to simulation scale factors for b-tagging efficiency are provided for this working point for the full dataset as a function of jet p_T , η and flavour. They are applied to simulated jets by downgrading (upgrading) the b-tagging status of a fraction of the b-

tagged (untagged) jets that have a scale factor smaller (larger) than one.

3.5 MET

The missing transverse energy, E_T^{MISS} or MET, of an event is defined as the magnitude of the imbalance of momentum in the plane transverse to the beam line. Since momentum is conserved in this plane, any imbalance in momentum is attributed to particles escaping the detector without interacting with the detector material, such as neutrinos or hypothetical dark matter candidates. Raw MET or particle flow MET (PFMET) is defined as the magnitude of the negative vectorial sum of the transverse momentum of all reconstructed particle flow candidates, or

$$\vec{E}_T^{\text{MISS}} = - \sum_{i \in \text{all}} \vec{p}_{T,i} \quad (3.5)$$

The vector quantity that is the negative sum of reconstructed particle momenta is sometimes called the missing transverse momentum, although this term is used interchangeably with its magnitude, the MET.

An alternative definition of the MET, called the type-I corrected MET, takes into account the jet energy corrections (JEC), correcting for mismeasurement of MET due to detector inefficiencies and non linear responses in the calorimeters. The type-I corrected MET definition is given in Equation 3.6. Systematic uncertainties related to modelling real MET are obtained by varying the JEC and jet energy resolution (JER) and measuring the propagation of these variations to the MET uncertainty. These measurements are described in greater detail in the next section.

$$\vec{E}_T^{\text{MISS Type-I}} = - \sum_{\text{jet}} \vec{p}_{T,\text{jet}}^{\text{JEC}} - \sum_{i \in \text{uncl.}} \vec{p}_{T,i} \quad (3.6)$$

where the total contribution has been split into contributions from jets (first term) and contributions from unclustered objects (second term). The transverse momenta of jets in the first term is then replaced with the JEC transverse momenta.

3.5.0.4 MET filters

Due to detector and instrumental noise, several filters are applied to veto noisy events [?]:

- HBHENoiseFilter
- HBHENoiseIsoFilter
- EcalDeadCellTriggerPrimitiveFilter
- goodVertices
- eeBadScFilter
- globalTightHalo2016Filter
- BadPFMuonFilter
- BadChargedCandidateFilter

The first two filters remove noisy events from the HCAL, where the HBHE scintillator produce anomalous signals with pulse shapes and pixel multiplicities discrepant from those from a clean signal. The EcalDeadCellTriggerPrimitiveFilter removes events with down ECAL data links, comparing the sum of energy deposited in each cell of a supercluster to the trigger primitive saturation energy. goodVertices removes events with noisy vertex reconstruction from pileup effects. The eeBadScFilter removes events with noisy ECAL end cap super clusters. globalTightHalo2016Filter removes events with enhanced MET from beam-halo particles which are in time with the beam. The last two filters remove events with mis-reconstructed muon and charged hadron particle flow candidates.

3.5.0.5 Fake MET modeling

Figure 3.12 shows a discrepancy between data and MC in the high- E_T^{MISS} tail. These events typically contain a high p_T object either back-to-back or collinear with the E_T^{MISS} , pointing to artificially high E_T^{MISS} from mismeasurement of the object. These fake events are identified and removed by studying distributions of the transverse angular difference between the MET and various objects in the event [?].

Several variables are studied in order to understand how to remove these fake events from data:

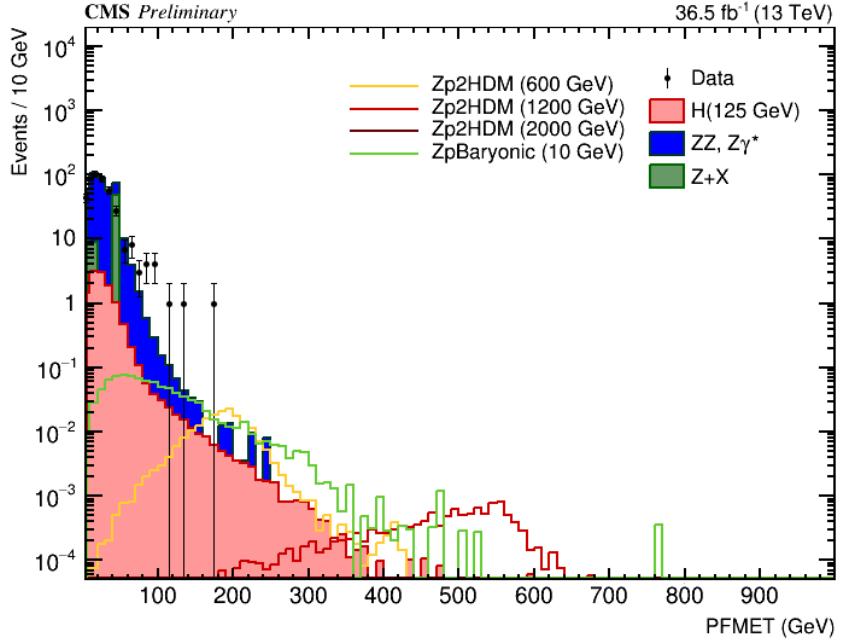


Figure 3.12: Missing transverse energy after SM selection before removal of fake MET.

- $\max|\Delta\phi(\text{jet}, E_T^{\text{MISS}})|$ with the maximum taken over selected jets, Figure 3.13
- $\min|\Delta\phi(\text{jet}, E_T^{\text{MISS}})|$ with the minimum taken over selected jets, Figure 3.14

The max variables are to check for the occurrence of objects with mismeasured energy back-to-back with the MET, the min variable is to check for jets with mismeasured energy collinear with the MET.

For jets with a high transverse momentum, greater than 50 GeV, it is required that $\max|\Delta\phi(\text{jet}, E_T^{\text{MISS}})| < 2.7$ and $\min|\Delta\phi(\text{jet}, E_T^{\text{MISS}})| > 0.5$ to exclude events with large MET from mismeasurement of jet energies. These cuts are based on the Run 1 SM analysis selection, chosen to balance the small loss in signal efficiency with the increased systematic uncertainty from mismodelling of the MET in background MC simulations (described in Section 7.1.1.2).

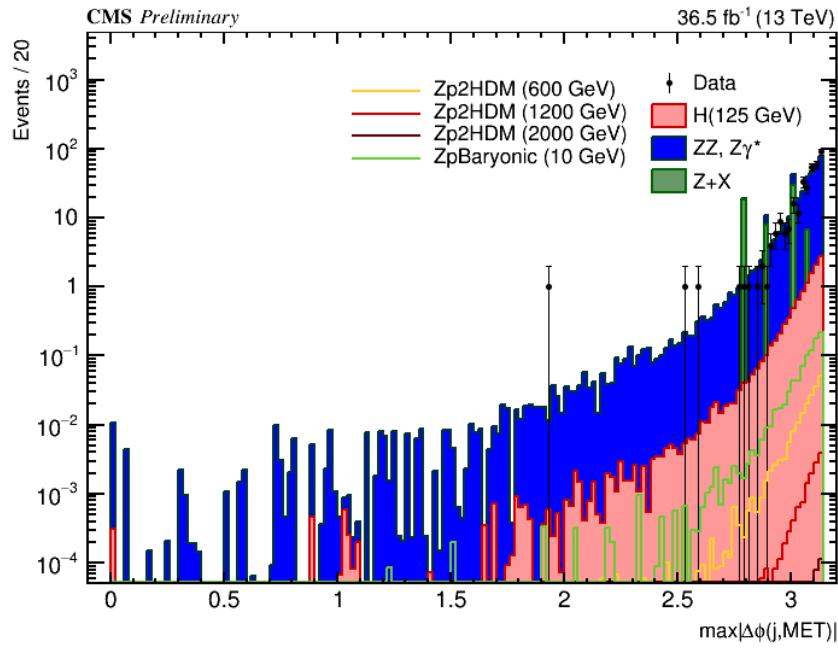


Figure 3.13: Maximum azimuthal angle difference between MET and jets.

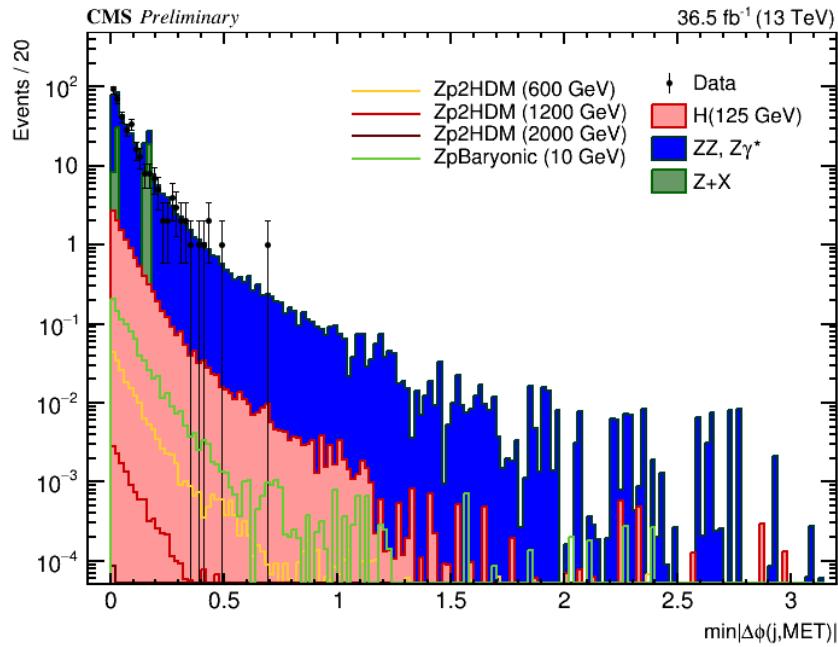


Figure 3.14: Minimum azimuthal angle difference between MET and jets.

Chapter 4

Event Selection

4.1 Trigger Selection

The events are required to have fired the High-Level Trigger paths described in section 2.1.1. Unlike in the Run I analysis, the trigger requirement does not depend on the selected final state: it is always the OR of all 10 HLT paths. The reason is in Run II we will be targeting associated production modes that can come with additional leptons, thus improving trigger efficiency further.

4.2 Vertex Selection

The events are required to have at least one good primary vertex (PV) fulfilling the following criteria: high number of degree of freedom ($N_{PV} > 4$), collisions restricted along the z -axis ($z_{PV} < 24$ cm) and small radius of the PV ($r_{PV} < 2$ cm).

4.3 ZZ Candidate Selection

The four-lepton candidates are built from what we call **selected leptons**, which are the tight leptons (defined in sections 3.1.2 and 3.2.1) that pass the $SIP_{3D} < 4$ vertex constraint and the isolation cuts (defined in sections 3.1.3 and 3.2.2), where FSR photons are subtracted as described in Section 3.3. A lepton cross cleaning is applied by discarding electrons which are within $\Delta R < 0.05$ of selected muons.

The construction and selection of four-lepton candidates proceeds according to the following sequence:

1. **Z candidates** are defined as pairs of selected leptons of opposite charge and matching flavour (e^+e^- , $\mu^+\mu^-$) that satisfy $12 < m_{\ell\ell(\gamma)} < 120$ GeV, where the Z candidate mass includes the selected FSR photons if any.
2. **ZZ candidates** are defined as pairs of non-overlapping Z candidates. The Z candidate with reconstructed mass $m_{\ell\ell}$ closest to the nominal Z boson mass is denoted as Z_1 , and the second one is denoted as Z_2 . ZZ candidates are required to satisfy the following list of requirements:
 - **Ghost removal** : $\Delta R(\eta, \phi) > 0.02$ between each of the four leptons.
 - **lepton p_T** : Two of the four selected leptons should pass $p_{T,i} > 20$ GeV and $p_{T,j} > 10$ GeV.
 - **QCD suppression**: all four opposite-sign pairs that can be built with the four leptons (regardless of lepton flavor) must satisfy $m_{\ell\ell} > 4$ GeV. Here, selected FSR photons are not used in computing $m_{\ell\ell}$, since a QCD-induced low mass dilepton (eg. J/Ψ) may have photons nearby (e.g. from π_0).
 - **Z_1 mass**: $m_{Z_1} > 40$ GeV
 - **'smart cut'**: defining Z_a and Z_b as the mass-sorted alternative pairing Z candidates (Z_a being the one closest to the nominal Z boson mass), require $\text{NOT}(|m_{Z_a} - m_Z| < |m_{Z_1} - m_Z| \text{ AND } m_{Z_b} < 12)$. Selected FSR photons are included in m_Z 's computations. This cut discards 4μ and $4e$ candidates where the alternative pairing looks like an on-shell Z + low-mass $\ell^+\ell^-$. (NB. In Run I, such a situation was avoided by choosing the best ZZ candidate before applying kinematic cuts to it, most precisely before the $m_{Z_2} > 12$ GeV cut. The present smart cut allows to choose the best ZZ candidate after all kinematic cuts.)
 - **four-lepton invariant mass**: $m_{4\ell} > 70$ GeV
3. Events containing at least one selected ZZ candidate form the **SM signal region**.

4.4 Choice of the best ZZ Candidate

Unlike in the Run I analysis, the best ZZ candidate is now chosen after all kinematic cuts, a change that allows to test other selection strategies for this candidate choice. This is especially relevant for events with more than four selected leptons, having in mind the search for the Higgs boson production modes where associated particles can decay to leptons, such as VH and ttH.

For the current analysis, we adopt a different approach compared to Run 1: if more than one ZZ candidate survives the above selection, we choose the one with the highest value of $\mathcal{D}_{\text{bkg}}^{\text{kin}}$ (defined in Section ??, except if two candidates are composed of them four leptons in which case the candidate with Z_1 closest in mass to nominal Z boson mass is chosen.

Chapter 5

Background Estimation

5.1 Irreducible Backgrounds

5.1.1 $q\bar{q} \rightarrow ZZ$ Modelling

The $q\bar{q} \rightarrow ZZ$ background is generated at NLO, while the fully differential cross section has been computed at NNLO [1], but are not yet available in a partonic level event generator. Therefore NNLO/NLO k -factors for the $q\bar{q} \rightarrow ZZ$ background process are applied to the POWHEG sample. The inclusive cross sections obtained using the same PDF and renormalization and factorization scales as the POWHEG sample at LO, NLO, and NNLO are shown in Table 5.1. The NNLO/NLO k -factors are applied in the analysis differentially as a function of $m(ZZ)$.

Additional NLO electroweak corrections which depend on the initial state quark flavor and kinematics are also applied to the $q\bar{q} \rightarrow ZZ$ background process in the region $m(ZZ) > 2m(Z)$ where the corrections have been computed. The differential QCD and electroweak k -factors can be seen in Figure 5.1.

QCD Order	$\sigma_{2\ell 2\ell'}(\text{fb})$	$\sigma_{4\ell}(\text{fb})$
LO	$218.5^{+16\%}_{-15\%}$	$98.4^{+13\%}_{-13\%}$
NLO	$290.7^{+5\%}_{-8\%}$	$129.5^{+4\%}_{-6\%}$
NNLO	$324.0^{+2\%}_{-3\%}$	$141.2^{+2\%}_{-2\%}$

Table 5.1: Cross sections for $q\bar{q} \rightarrow ZZ$ production at 13 TeV

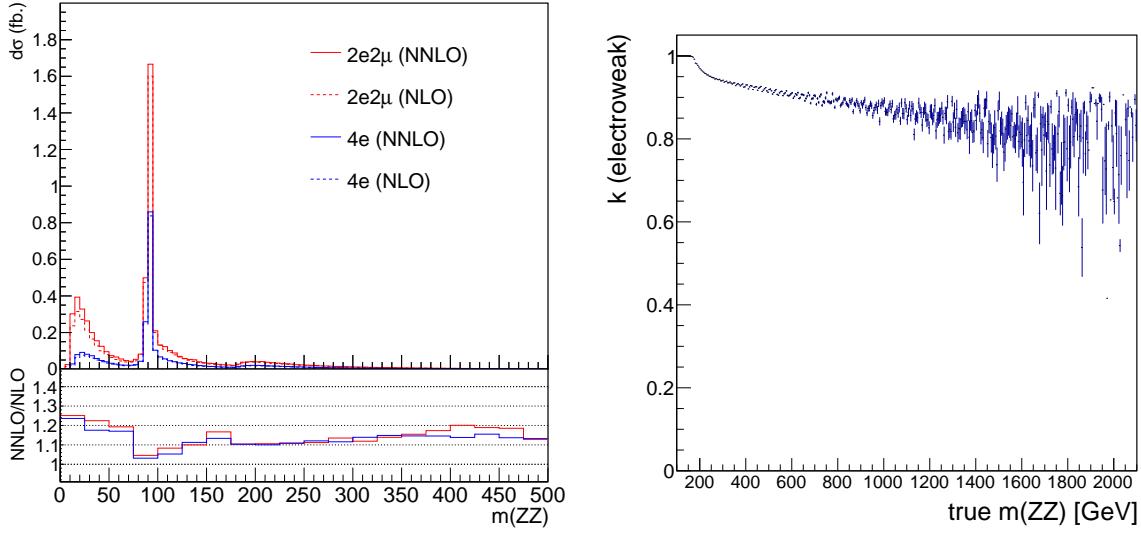


Figure 5.1: Left: NNLO/NLO QCD k -factor for the $q\bar{q} \rightarrow ZZ$ background as a function of $m(ZZ)$ for the 4ℓ and $2\ell 2\ell'$ final states. Right: NLO/NLO electroweak k -factor for the $q\bar{q} \rightarrow ZZ$ background as a function of $m(ZZ)$.

5.1.2 $gg \rightarrow ZZ$ Modelling

Event simulation for the $gg \rightarrow ZZ$ background is done at LO with the generator MCFM 7.0 [2, 3, 4]. Although no exact calculation exists beyond the LO for the $gg \rightarrow ZZ$ background, it has been recently shown [5] that the soft collinear approximation is able to describe the background cross section and the interference term at NNLO. Further calculations also show that the K factors are very similar at NLO for the signal and background [6] and at NNLO for the signal and interference terms [7]. Therefore, the same K factor is used for the signal and background [8]. The NNLO K factor for the signal is obtained as a function of $m_{4\ell}$ using the HNNLO v2 Monte Carlo program [9, 10, 11] by calculating the NNLO and LO $gg \rightarrow H \rightarrow 2\ell 2\ell'$ cross sections at the small H boson decay width of 4.07 MeV and taking their ratios. The NNLO as well as the NLO K factors and the cross sections from which they are derived are illustrated in Fig. 5.2, along with the NNLO, NLO and LO cross sections at the SM H boson decay width [12].

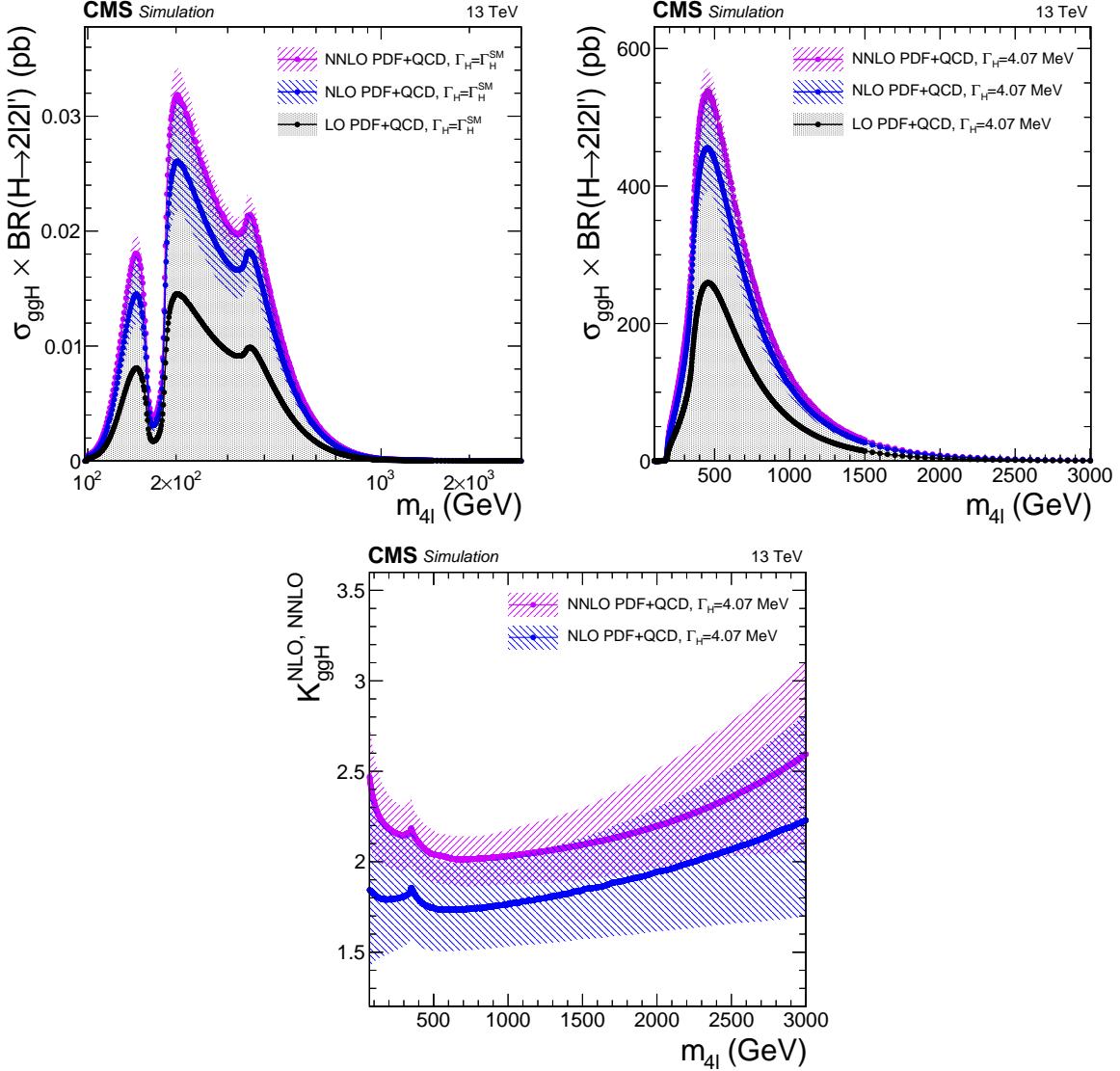


Figure 5.2: $gg \rightarrow H \rightarrow 2\ell 2\ell'$ cross sections at NNLO, NLO and LO at each H boson pole mass using the SM H boson decay width (top) or at the fixed and small decay width of 4.07 MeV (top). The cross sections using the fixed value are used to obtain the K factor for both the signal and the continuum background contributions as a function of $m_{4\ell}$ (bottom).

5.2 Reducible Background

The reducible background for the $H \rightarrow ZZ \rightarrow 4\ell$ analysis, hereafter called $Z + X$, originates from processes which contain one or more non-prompt leptons in the four-lepton final state. The main sources of non-prompt leptons are non-isolated electrons and muons coming from decays of heavy-flavour mesons, mis-reconstructed jets (usually originating

from light-flavour quarks) and electrons from γ conversions. In this discussion, we will consider a “fake lepton” any jet mis-reconstructed as a lepton and any lepton originating from a heavy meson decay. Similarly, any electron originating from a photon conversion will be considered a “fake electron”.

In the $H \rightarrow ZZ \rightarrow 4\ell$ analysis, the rate of these background processes is estimated by measuring the f_e and f_μ probabilities for fake electrons and fake muons which do pass the **loose** selection criteria (defined in Section 3.1.1 and 3.2.1) to also pass the final selection criteria (defined in Section 4.3). These probabilities, hereafter referred to as fake ratios or fake rates, are applied in dedicated control samples in order to extract the expected background yield in the signal region.

In the following section, two independent methods are presented to measure both the yields and shapes of the reducible background. The final result combines the outcome of the two approaches.

Fake Rate Determination (OS Method) In order to measure the lepton fake ratios f_e , f_μ , we select samples of $Z(\ell\ell) + e$ and $Z(\ell\ell) + \mu$ events that are expected to be completely dominated by final states which include a Z boson and a fake lepton. These events are required to have two same flavour, opposite charge leptons with $p_T > 20/10$ GeV passing the tight selection criteria, thus forming the Z candidate. In addition, there is exactly one lepton passing the loose selection criteria as defined above. This lepton is used as the probe lepton for the fake ratio measurement. The invariant mass of this lepton and the opposite sign lepton from the reconstructed Z candidate should satisfy $m_{2l} > 4$ GeV.

The fake ratios are evaluated using the tight requirement $|M_{inv}(\ell_1, \ell_2) - M_Z| < 7$ GeV, to reduce the contribution from photon (asymmetric) conversions populating low masses. The fake ratios are measured in bins of the transverse momentum of the loose lepton and barrel and endcap region.

The electron and muon fake rates are measured within $|M_{inv}(\ell_1, \ell_2) - M_Z| < 7$ GeV and $E_T^{\text{miss}} < 25$ GeV are shown in Figure 5.3.

Fake Rate Application (OS Method) Two control samples are obtained as subsets of four lepton events which pass the first step of the selection (*First Z* step, see section 4.3),

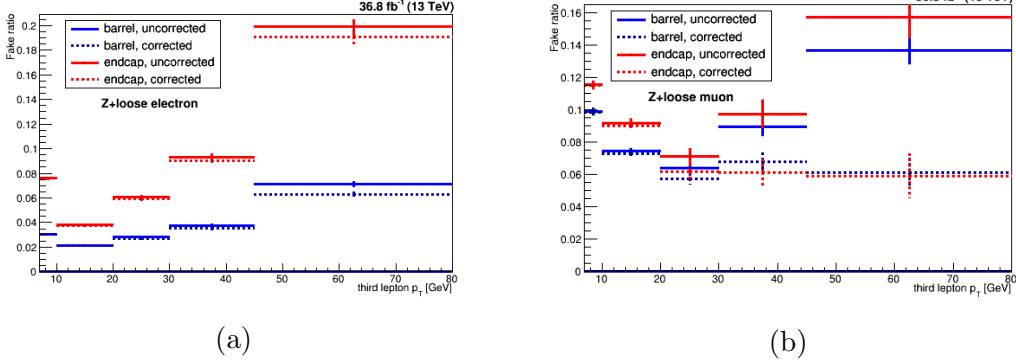


Figure 5.3: Fake rates as a function of the probe p_T for electrons (a) and muons (b) which satisfy the loose selection criteria, measured in a $Z(\ell\ell) + \ell$ sample in the 13 TeV data. The barrel selection includes electrons (muons) up to $|\eta| = 1.479$ (1.2).

requiring an additional pair of loose leptons of same flavour and opposite charge, that pass the SIP_{3D} cut. The events must satisfy all kinematic cuts applied for the *Higgs phase space* selection (see 4.3). The first control sample is obtained by requiring that the two loose leptons which do not make the Z_1 candidate do not pass the final identification and isolation criteria. The other two leptons pass the final selection criteria by definition of the Z_1 . This sample is denoted as “2 Prompt + 2 Fail” ($2P+2F$) sample. It is expected to be populated with events that intrinsically have only two prompt leptons (mostly DY , with a small fraction of $t\bar{t}$ and $Z\gamma$ events). The second control sample is obtained by requiring one of the four leptons not to pass the final identification and isolation criteria. The other three leptons should pass the final selection criteria. This control sample is denoted as “3 Prompt + 1 Fail” ($3P+1F$) sample. It is expected to be populated with the type of events that populate the $2P+2F$ region, albeit with different relative proportions, as well as with WZ events that intrinsically have three prompt leptons.

The control samples obtained in this way, orthogonal by construction to the signal region, are enriched with fake leptons and are used to estimate the reducible background in the signal region.

The invariant mass distribution of events selected in the $2P+2F$ control sample is shown in Fig. 5.4 for the 13 TeV dataset.

The expected number of reducible background events in the $3P+1F$ region, N_{3P1F}^{bkg} , can be computed from the number of events observed in the $2P+2F$ control region, N_{2P2F} , by

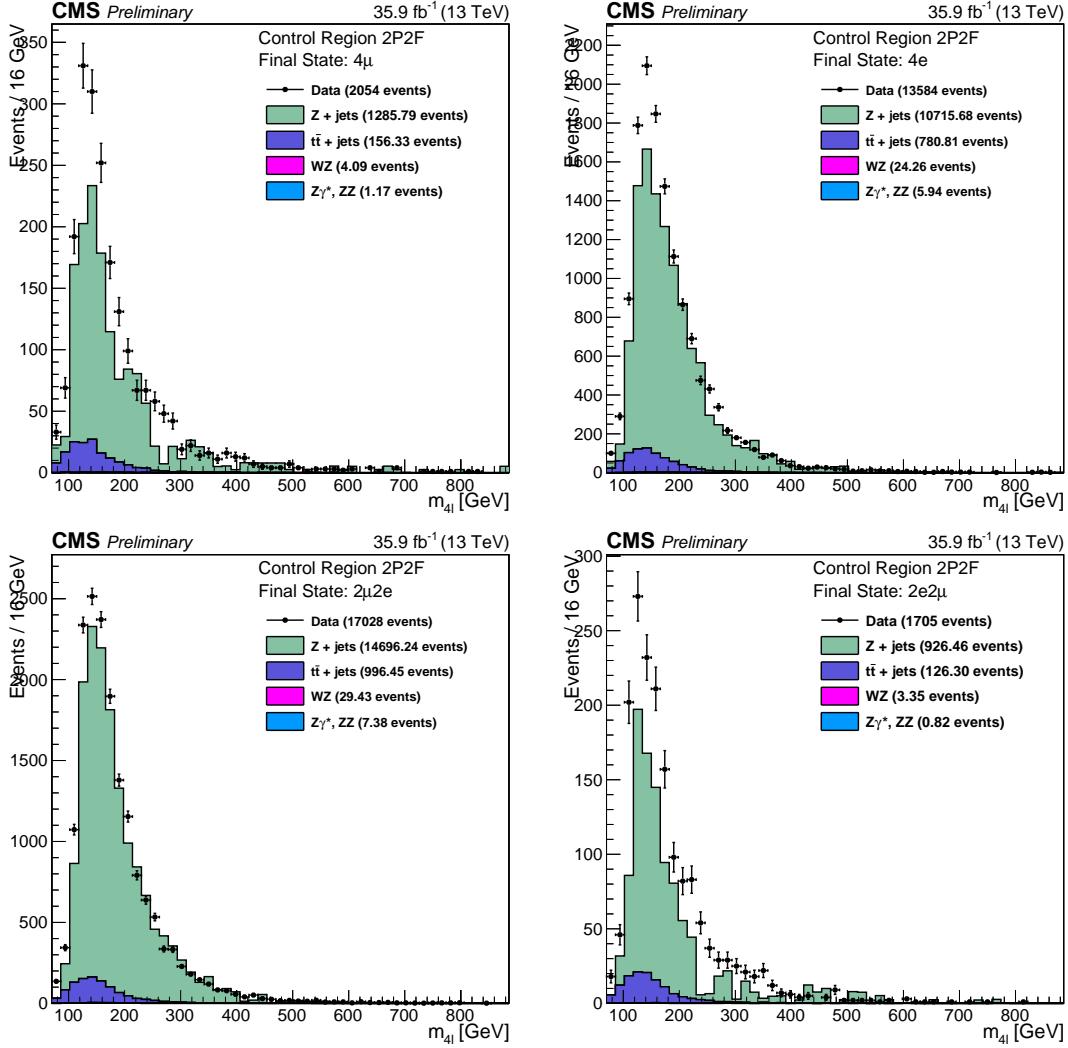


Figure 5.4: Invariant mass distribution of the events selected in the 2P+2F control sample in the 13 TeV dataset, (top left) 4μ , (top right) $4e$, (bottom left) $2\mu 2e$ and (bottom right) $2e 2\mu$ channels.

weighting each event in the region with the factor $(\frac{f_i}{1-f_i} + \frac{f_j}{1-f_j})$, where f_i and f_j correspond to the fake ratios of the two loose leptons:

$$N_{3P1F}^{\text{bkg}} = \sum \left(\frac{f_i}{1-f_i} + \frac{f_j}{1-f_j} \right) N_{2P2F} \quad (5.1)$$

Figure 5.5 shows the invariant mass distributions of the events selected in the 3P+1F control sample, together with the expected reducible background estimated from Eq. 5.1, stacked on the distribution of WZ and of irreducible background ($ZZ, Z\gamma^* \rightarrow 4\ell$) taken

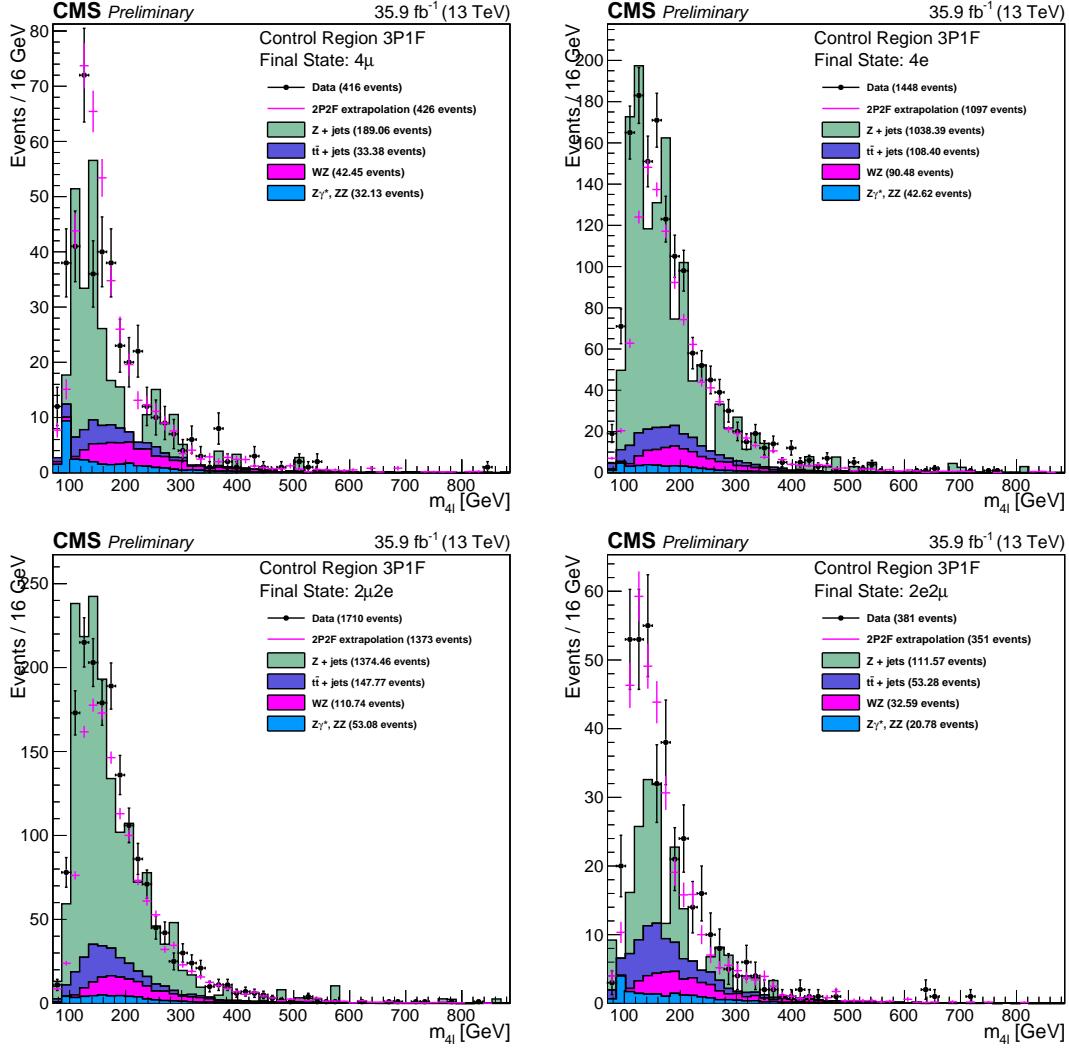


Figure 5.5: Invariant mass distribution of the events selected in the 3P+1F control sample in the 13 TeV dataset, (top left) 4μ , (top right) $4e$, (bottom left) $2\mu 2e$ and (bottom right) $2e 2\mu$ channels.

from the simulation.

Would the fake rates be measured in a sample that has exactly the same background composition as the 2P+2F sample, the difference between the observed number of events in the 3P+1F sample and the expected background predicted from the 2P+2F sample would solely amount to the (small) WZ and $Z\gamma_{conv}$ contribution. Large differences arise because the fake rates used in Eq. 5.1 do not properly account for the background composition of the 2P+2F control sample.

In particular, the difference seen in Fig. 5.5 between the observed 3P+1F distribution and the expectation from 2P+2F, in the channels with loose electrons ($4e$ and $2\mu 2e$), and concentrated at low masses, is due to photon conversions. This is confirmed explicitly by the simulation. The difference between the 3P+1F observation and the prediction from 2P+2F to recover the missing contribution from conversions - and more generally, in principle, to “correct” for the fact that the fake rates do not properly account for the background composition of the 2P+2F sample. More precisely, the expected reducible background in the signal region is given by the sum of two terms :

- a “2P2F component”, obtained from the number of events observed in the 2P+2F control region, N_{2P2F} , by weighting each event in that region with the factor $\frac{f_i}{1-f_i} \frac{f_j}{1-f_j}$, where f_i and f_j correspond to the fake ratios of the two loose leptons.
- a “3P1F component”, obtained from the difference between the number of observed events in the 3P+1F control region, N_{3P1F} , and the expected contribution from the 2P+2F region and ZZ processes in the signal region, $N_{3P1F}^{ZZ} + N_{3P1F}^{bkg}$. The N_{3P1F}^{bkg} is given by equation 5.1 and N_{3P1F}^{ZZ} is the contribution from ZZ which is taken from the simulation. The difference $N_{3P1F} - N_{3P1F}^{bkg} - N_{3P1F}^{ZZ}$, which may be negative, is obtained for each (p_T, η) bin of the “F” lepton, and is weighted by $\frac{f_i}{1-f_i}$, where f_i denotes the fake rate of this lepton. This “3P1F component” accounts for the contribution of reducible background processes with only one fake lepton (like WZ events), and for the contribution of other processes (e.g. photon conversions) that are not properly estimated by the 2P2F component, because of the fake rates used.

Therefore, the full expression for the prediction can be symbolically written as:

$$N_{SR}^{bkg} = \sum_i \frac{f_i}{(1-f_i)} (N_{3P1F} - N_{3P1F}^{bkg} - N_{3P1F}^{ZZ}) + \sum_j \frac{f_i}{(1-f_i)} \frac{f_j}{(1-f_j)} N_{2P2F} \quad (5.2)$$

Previous equation is equivalent to the following:

$$N_{SR}^{bkg} = \left(1 - \frac{N_{3P1F}^{ZZ}}{N_{3P1F}}\right) \sum_j^{N_{3P1F}} \frac{f_a^j}{1-f_a^j} - \sum_i^{N_{2P2F}} \frac{f_3^i}{1-f_3^i} \frac{f_4^i}{1-f_4^i} \quad (5.3)$$

For channels where the Z_2 candidate is made from two electrons, the contribution of the 3P1F component is positive, and amounts to typically 30% of the total predicted background.

For channels with loose muons (4μ and $2e2\mu$), the 3P+1F sample is rather well described by the prediction from 2P+2F, as seen in Fig. 5.5b, and the 3P1F component is mainly driven by statistical fluctuations in the 3P+1F sample, which are larger than the expectation from WZ production.

Table 5.2 shows the expected number of events in the signal regions from the reducible background processes for the 13 TeV.

baseline	4e	4μ	$2e2\mu$	$2\mu2e$
13 TeV	22.19	32.81	22.48	41.72

Table 5.2: The contribution of reducible background processes in the signal region predicted from measurements in data using the Opposite-Sign Leptons method. The predictions correspond to 35.9 fb^{-1} of data at 13 TeV.

Chapter 6

Signal region and blinding

Unlike the 2015 version of this analysis, which used a two-step signal region (SR) and optimization process, the current analysis uses one step, where a near-optimal SR is defined using all of the desired variables. This simplification allows for the use of the same SR for all signal models, including the same variable validation and background model. The tradeoff is a small loss in sensitivity for models with tighter optimal cuts, but the benefit of simplicity outweighs the cost of this loss. In principle, each benchmark signal point can be optimized individually to gain a few percent in sensitivity, but this is left to a future study.

Two strategies are employed to define the signal region: cut-and-count based and multivariate analysis (MVA) based. Several key discriminating variables are studied as inputs to both SR definitions: the missing transverse momentum, E_T^{MISS} , the four-lepton invariant mass, m_{4l} , the transverse mass of the Higgs and MET, $m_T(4l + E_T^{\text{MISS}}) \equiv m_T$, the difference in the transverse angle ϕ between the Higgs and MET, $|\Delta\phi(4l, E_T^{\text{MISS}})| \equiv |\Delta\phi|$, as well as the lepton and jet multiplicities.

6.0.1 Cut-and-count based signal region

The cut-and-count strategy is the simpler choice, and is used as a baseline to measure the performance of the MVA. First, the following selection is applied to isolate events with a Higgs from events with additional prompt particles (e.g. VBF Higgs production):

- Tight lepton multiplicity = 4

- b jet multiplicity ≤ 1
- VBF jet multiplicity ≤ 1

Next, the event selection is optimized by scanning over a range of cuts for the remaining variables and selecting the set of cuts that maximizes the sensitivity, measured directly by the cross section upper limit. The two variables with the greatest discriminating power between signal and background are m_{4l} and E_T^{MISS} , so we maximize the sensitivity over these two variables. The best sensitivity occurs where the upper limit is minimum, around:

- $|m_{4l} - m_H| < 10 \text{ GeV}$
- $E_T^{\text{MISS}} > 60 \text{ GeV}$

These cuts define the SR, which is applied to all of the signals, losing less than 10% sensitivity for the models with a tighter optimal cut on E_T^{MISS} . Since the signal used to define the SR has the softest E_T^{MISS} spectrum of all the signals, this SR corresponds to the most modest or most loose choice, meaning no signal will be cut on too hard, while most of the background is still removed.

In addition to blinding observations of data distributions in the SR until the selection is frozen, it should be noted that there is no additional blinding on the MET distribution above a certain threshold, as was the case previously in similar searches. This is due primarily to the need to understand events that contribute large amounts of fake MET. This is covered thoroughly, including the procedure for removing these events from data, in the next section. In order to validate the modelling in these SRs, they are split into control regions (CR) based on the MET being above or below 60 GeV, referred to as the high and low MET regions.

6.0.2 MVA based signal region

This search channel has the advantage of having backgrounds that are easily reduced by applying cuts on the discriminating variables. It was observed in the study of the 2015 data and MC samples that applying additional cuts did not improve the sensitivity

since the background levels were already sufficiently low. Applying cuts on additional variables reduced the signal efficiency, which in turn reduced the sensitivity. This is also observed with 2016 MC samples, where applying the $|\Delta\phi(l_{lll}, E_T^{\text{MISS}})|$ does not improve the sensitivity. These observations motivate the use of MVA techniques, which can take all of the desired variables as input, but do not reduce the signal efficiency. Although it is simpler to cut on these discriminating variables as described above, there is potential for significant improvement in the sensitivity with an MVA approach.

The SR event selection is optimized for the MVA based case by training a boosted decision tree (BDT) with the ROOT TMVA package with the input variables: m_{4l} and E_T^{MISS} . Including additional input features does not significantly improve the performance of the MVA. The training is done over the weighted set of backgrounds and an admixture of signal models to reduce bias toward a single signal model. The BDT response is shown in Figure 6.1, with the signal peaked toward 1 and the background peaked toward -1.

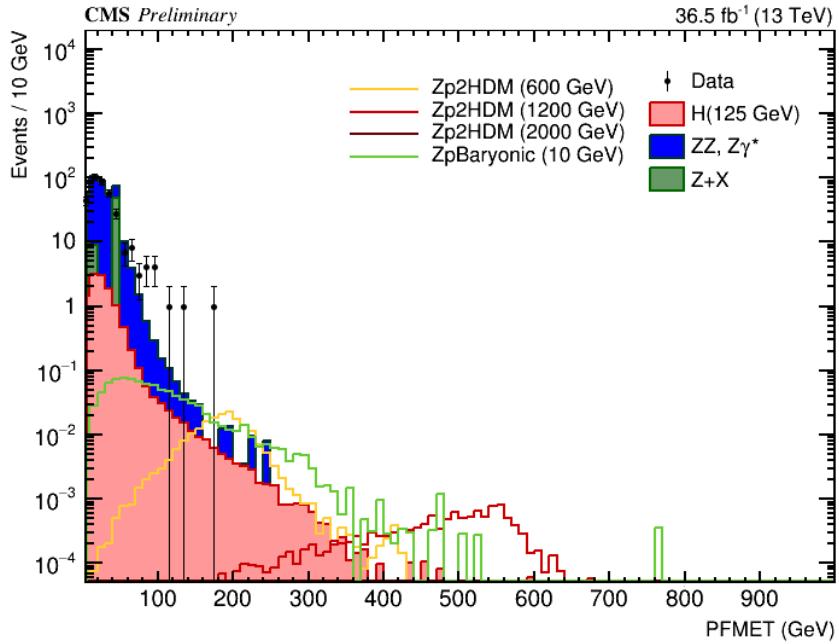


Figure 6.1: PLACEHOLDER BDT response.

6.1 Yields and distributions

Channel	4e	4 μ	2e2 μ	4l
$q\bar{q} \rightarrow ZZ$	0.029	328	1.58e+03	1.91e+03
$gg \rightarrow ZZ$	24.7	33.7	123	182
$Z + X$	14.5	20.9	31	66.5
ZH	0.672	0.905	1.51	3.09
Other Higgs	7.24	11	31.6	49.8
Total background	47.2	395	1.77e+03	2.21e+03
Signal ($m_\chi = 600$ GeV)	0.143	0.203	0.349	0.695
Total expected	47.3	395	1.77e+03	2.21e+03
Observed	103	177	0	280

Table 6.1: PLACEHOLDER Event yields after SM selection, with $m_{4l} > 70$ GeV.

Sample	$q\bar{q} \rightarrow ZZ$	$gg \rightarrow ZZ$	$Z + X$	ZH	Other H	Total	Signal	Observed
Initial	1.19e+04	934	3.17e+08	17.9	387	3.17e+08	3.62	4.16e+07
HLT	1.19e+04	934	3.17e+08	17.9	387	3.17e+08	3.62	4.16e+07
Z_1 lepton cuts	10.3	130	3.08e+03	2	44.4	3.26e+03	0.429	2.3e+04
m_{Z_1}	7.19	110	1.87e+03	1.77	42	2.03e+03	0.404	1.61e+04
At least one Z_2	0.0327	53.6	14.5	0.672	14.3	83.1	0.143	103
m_{Z_2}	0.0327	53.6	14.5	0.672	14.3	83.1	0.143	103
$m_{ll} > 4$ for OS-SF	0.029	24.7	14.5	0.672	7.24	47.2	0.143	103
$m_{lll} > 70$ GeV	0.029	24.7	14.5	0.672	7.24	47.2	0.143	103

Table 6.2: PLACEHOLDER Cut flow for 4e channel. Example signal is Z'2HDM with $m_{Z'} = 600$ GeV.

Sample	$q\bar{q} \rightarrow ZZ$	$gg \rightarrow ZZ$	$Z + X$	ZH	Other H	Total	Signal	Observed
Initial	1.19e+04	878	3.17e+08	17.9	387	3.17e+08	3.66	4.39e+07
HLT	1.19e+04	878	3.17e+08	17.9	387	3.17e+08	3.66	4.39e+07
Z_1 lepton cuts	2.03e+03	144	5.66e+04	2.37	57.5	5.89e+04	0.517	3.75e+04
m_{Z_1}	1.67e+03	128	4.31e+04	2.16	54.4	4.5e+04	0.486	1.74e+04
At least one Z_2	399	76.4	23	0.971	23.8	523	0.217	178
m_{Z_2}	399	76.4	23	0.971	23.8	523	0.217	178
$m_{ll} > 4$ for OS-SF	329	33.7	20.9	0.905	11	395	0.203	178
$m_{lll} > 70$ GeV	328	33.7	20.9	0.905	11	395	0.203	177

Table 6.3: PLACEHOLDER Cut flow for 4μ channel. Example signal is Z'2HDM with $m_{Z'} = 600$ GeV.

Sample	$q\bar{q} \rightarrow ZZ$	$gg \rightarrow ZZ$	$Z + X$	ZH	Other H	Total	Signal	Observed
Initial	3.7e+04	934	3.28e+06	17.9	387	3.32e+06	3.62	0
HLT	3.7e+04	934	3.28e+06	17.9	387	3.32e+06	3.62	0
Z_1 lepton cuts	8.73e+03	422	4.71e+03	5.83	136	1.4e+04	1.33	0
m_{Z_1}	6.96e+03	369	2.98e+03	5.3	127	1.04e+04	1.23	0
At least one Z_2	1.58e+03	123	31	1.51	31.6	1.77e+03	0.349	0
m_{Z_2}	1.58e+03	123	31	1.51	31.6	1.77e+03	0.349	0
$m_{ll} > 4$ for OS-SF	1.58e+03	123	31	1.51	31.6	1.77e+03	0.349	0
$m_{lll} > 70$ GeV	1.58e+03	123	31	1.51	31.6	1.77e+03	0.349	0

Table 6.4: PLACEHOLDER Cut flow for $2e2\mu$ channel. Example signal is Z'2HDM with $m_{Z'} = 600$ GeV.

Chapter 7

Results

7.1 Statistical analysis

7.1.1 Systematic uncertainties

This section covers the systematic, including statistical, errors associated with the analysis. The general strategy is to duplicate the systematics applied in the SM H to 4l search, including errors on the background estimation from data, and add the errors associated with MET modelling, guided by the mono-H to $\gamma\gamma$ strategy.

7.1.1.1 Uncertainties on Reducible Background estimation

Sources of systematic uncertainties for the methods presented above potentially arises from the different composition of background processes (DY , $t\bar{t}$, WZ , $Z\gamma$) in the region where we measure and where we apply the fake ratios. OS method corrects for the resulting bias via the “3P1F component” of its prediction. SS method corrects explicitly the electron fake rates for the correct fraction of photon conversions, but no attempt is made to correct the muon fake rates. The closure tests presented here are used to assess a possible residual bias in the methods.

Statistics in 4l Control Sample The limited size of the samples in the control regions where we measure and where we apply the fake ratios is the source of the statistical uncertainties of the method. The dominating statistical uncertainty is driven by the number of events in the control region and is typically in the range of 1-10%.

Sensitivity of Fake Ratios to Background Composition Compositions of reducible background processes (DY , $t\bar{t}$, WZ , $Z\gamma^{(*)}$) in the region where we measure and where we apply the fake ratio method are typically not the same. This is the main source of the systematic uncertainty of the fake ratio method.

This uncertainty can be estimated by measuring the fake ratios for individual background processes in the $Z + 1L$ region in simulation. The weighted average of these individual fake ratios is the fake ratio that we measure in this sample (in simulation). The exact composition of the background processes in the 2P+2F region where we plan to apply the fake ratios can be determined from simulation, and one can reweigh the individual fake ratios according to the 2P+2F composition. The difference between the reweighed fake ratio and the average one can be used as a measure of the uncertainty on the measurement of the fake ratios. The fake ratios for individual processes, the average fake ratio and the reweighed fake ratio determined by simulation are shown in table below. The effect of this systematic uncertainty is propagated to the final estimates, and it amounts to about 32% for $4e$, 33% for $2e2\mu$ and 35% for 4μ final state.

Shape Uncertainty In order to estimate the uncertainty on the m_{4l} shape we have looked at the differences between the shapes of predicted background distributions for all three channels, and between both of the two methods. The envelope of differences between these shapes of distributions is used as an estimate of the shape uncertainty. The uncertainty is estimated to be roughly in the range of 5% - 15%. Since the difference of the shapes slowly varies with m_{4l} , it is taken as a constant versus m_{4l} and is absorbed in the much larger uncertainty on the predicted yield of backgrounds.

7.1.1.2 MET systematics

There are two types of systematic uncertainties related to the modeling of MET, those from the measurement of real MET, as from the signal samples or backgrounds with neutrinos, and those from fake MET, due to the mismeasurement of jets and other objects. The fake MET systematics apply to the Higgs signals with no associated W production and the non-resonant backgrounds.

The uncertainties from the modeling of real met are measured by varying several

corrections used to calculate MET, propogating these variations to the efficiency of MC samples to pass the MET selection [?]. The list of corrections used in this calculation are given below. Each correction is varied up and down by one standard deviation of the input distribution, with the systematic uncertainty taken as the maximim difference in efficiencies accross all correction variations.

- Jet energy
- Jet resolution
- Muon energy
- Electron energy
- Tau energy
- Photon energy
- Unclustered jet energy

The efficiencies for the V+H and a benchmark signal sample to pass the MET selection of $E_T^{\text{MISS}} > 60 \text{ GeV}$ after each correacion is varied up and down are given in Table ?? and Table ??, respectively.

Correction	Efficiency Up	Efficiency Down
Original PFMET		
Jet energy		
Jet resolution		
Muon energy		
Electron energy		
Tau energy		
Photon energy		
Unclustered jet energy		

Correction	Efficiency Up	Efficiency Down
Original PFMET		
Jet energy		
Jet resolution		
Muon energy		
Electron energy		
Tau energy		
Photon energy		
Unclustered jet energy		

The second systematic uncertainty results from the modeling of fake MET, primarily due to the mismeasurement of jets (see Section 5.2). This systematic is measured in the sideband CR as the percent difference between the efficiency for the data and total background sample to pass the MET selection. The efficiency for data is XX, MC is XX, giving a difference of around 40% which is taken as the systematic.

7.1.1.3 Additional systematics

The main experimental uncertainties which affect both signal and background are the uncertainty on the integrated luminosity (2.6%) and the uncertainty on the lepton identification and reconstruction efficiency (ranging from 2.5–9% on the overall event yield for the 4μ and $4e$ channels, respectively). Experimental uncertainties for the reducible background estimation, described in Section 5.1.1, vary between 36% (4μ) and 43% ($4e$). The uncertainty on the lepton energy scale is determined by considering the $Z \rightarrow \ell\ell$ mass distributions in data and simulation. Events are separated into categories based on the p_T and η of one of the two leptons, determined randomly, and integrating over the other. The dilepton mass distributions are then fit to a Breit-Wigner parameterization convolved with a double-sided Crystal Ball function. The offset in the measured peak position with respect to the nominal Z boson mass in data and simulation are extracted, and the results are shown in Fig. 7.1. The relative difference between data and simulation is propagated

to the reconstructed four-lepton mass from simulated Higgs boson events. The results of the propagation can be seen in Fig. 7.2. In the case of electrons, since the same data set is used to derive and validate the momentum scale corrections, the size of the corrections are taken into account for the final value of the uncertainty. The uncertainty is determined to be 0.04% (0.3%) for the 4μ ($4e$) channels, respectively. The uncertainty on the 4ℓ mass resolution coming from the uncertainty on the per-lepton energy resolution is 20%, as described in Section ??.

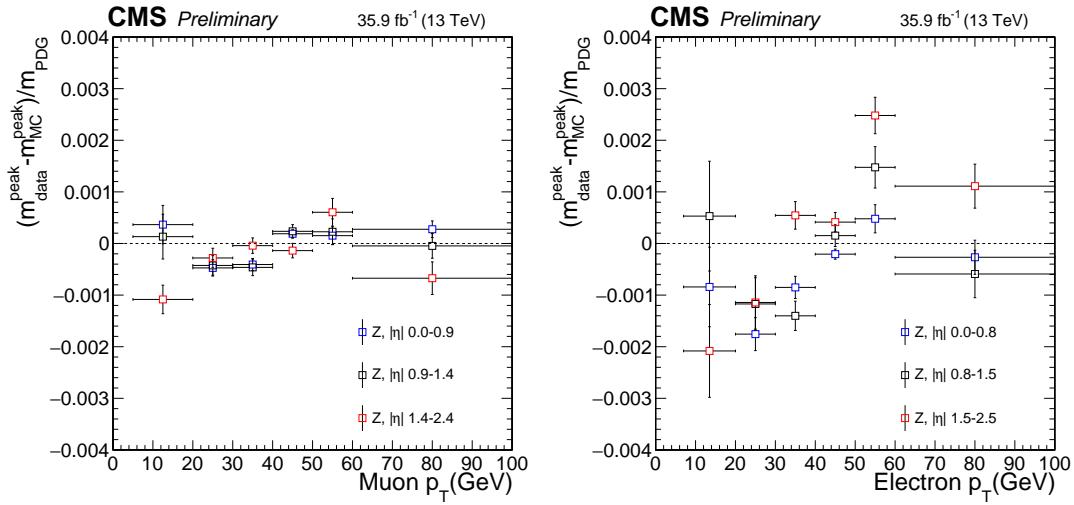


Figure 7.1: Difference between the $Z \rightarrow \ell\ell$ mass peak positions in data and simulation normalized by the nominal Z boson mass obtained as a function of the p_T and $|\eta|$ of one of the leptons regardless of the second for muons (left) and electrons (right).

Theoretical uncertainties which affect both the background signal and background estimation include uncertainties from the renormalization and factorization scale and choice of PDF set. The uncertainty from the renormalization and factorization scale is determined by varying these scales between 0.5 and 2 times their nominal value while keeping their ratio between 0.5 and 2. The uncertainty from the PDF set is determined by taking the root mean square of the variation when using different replicas of the default NNPDF set. An additional uncertainty of the 10% on the K factor used for the $gg \rightarrow ZZ$ prediction is applied as described in Section 5.1. A systematic uncertainty of 2% on the branching ratio of $H \rightarrow ZZ \rightarrow 4\ell$ only affects the signal yield. In the case of event categorization, experimental and theoretical uncertainties which account

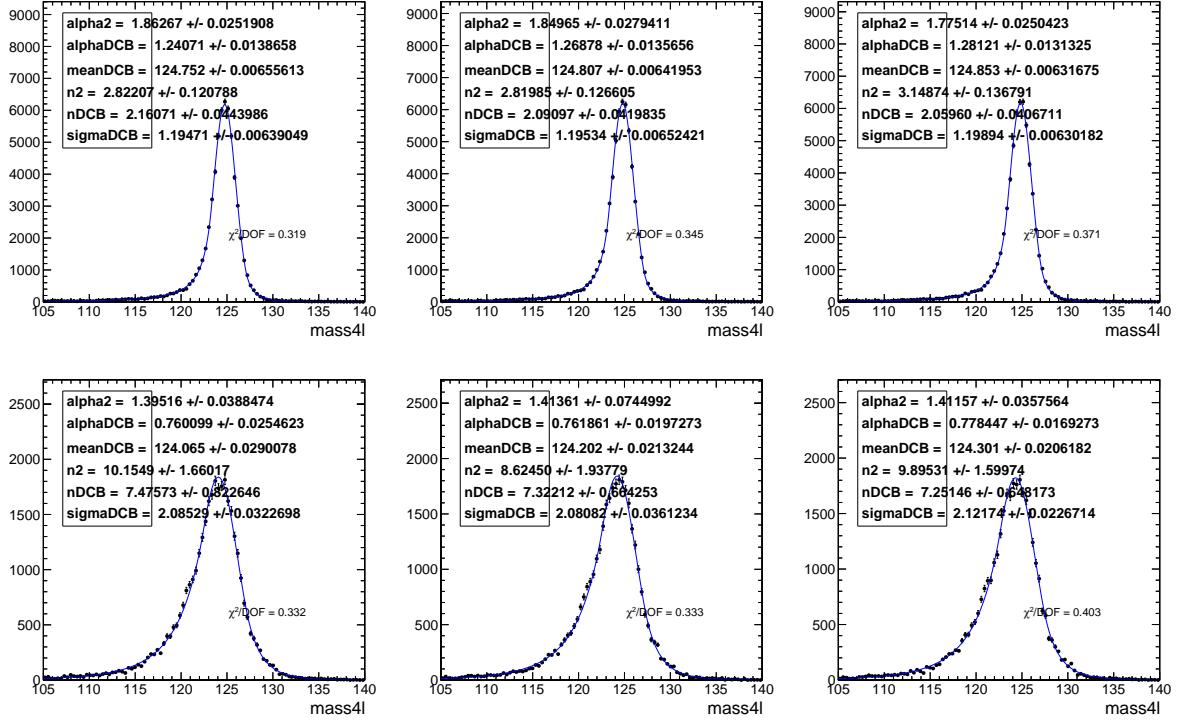


Figure 7.2: Different m_{4l} distributions after propagating the biases in Fig. 7.1 to Higgs boson events. The change in the mean of the double crystal ball is used to determine the systematic uncertainty due to the lepton momentum scale. The middle plot shows the nominal distribution, while the left (right) plots show the down (up) systematic variations. The 4μ channel is shown in the top row and the $4e$ channel is shown in the bottom row.

for possible migration of signal and background events between categories are included. The main sources of uncertainty on the event categorization include the QCD scale, PDF set, and the modeling of hadronization and the underlying event. These uncertainties amount to between 4–20% for the signal and 3–20% for the background depending on the category. The lower range corresponds to the VBF and VH processes and the upper range corresponds to the $gg \rightarrow H$ process yield in the VBF-2jet-tagged category. Additional uncertainties come from the imprecise knowledge of the jet energy scale (from 2% for the $gg \rightarrow H$ yield in the untagged category to 15% for $gg \rightarrow H$ yield in the VBF-2jet-tagged category) and b-tagging efficiency and mistag rate (up to 6% in the -tagged category).

Table 7.3: Summary of the experimental systematic uncertainties in the $H \rightarrow 4\ell$ measurements.

Summary of relative systematic uncertainties	
Common experimental uncertainties	
Luminosity	2.6 %
Lepton identification/reconstruction efficiencies	2.5 – 9 %
Background related uncertainties	
Reducible background ($Z+X$)	36 – 43 %
Signal related uncertainties	
Lepton energy scale	0.04 – 0.3 %
Lepton energy resolution	20 %

7.1.2 Limit setting

The primary tool used to interpret the analysis described above in the context of the signal models is the HiggsAnalysis-CombinedLimit package [?], a collection of RooStats-based software [?] used within the Higgs PAG [?], hereafter called the combine tool, or simply combine. Specifically, limits (one-sided Bayesian credible intervals) are set on the expected and observed signal production cross section times branching ratio BR of H to four leptons for the various benchmarks using the asymptotic CL_S method [?], an approach to calculating a profile likelihood ratio using an approximation of the LHC test-statistic distributions. The upper limit on the cross section gives the maximum number of events that can be attributed to the signal process, consistant with the data that is observed. Combine finds the upper limit as the numerical solution to Equation 7.1, which sets the integral of the posterior probability $p(\sigma|D)d\sigma$ equal to the desired confidence level for the measurement, typicall 95%. The posterior probability gives the degree of belief that σ lies in the interval $[\sigma, \sigma + d\sigma]$, and is formed by inverting a multi-Poisson model, $p(D|\sigma, \boldsymbol{\theta})$ using Bayes' theorem (Equation 7.2), after numerically marginalizing priors describing

Table 7.4: Summary of the theory systematic uncertainties in the $H \rightarrow 4\ell$ measurements for the inclusive analysis

Summary of inclusive theory uncertainties	
QCD scale (gg)	$\pm 3.9 \%$
PDF set (gg)	$\pm 3.2 \%$
Bkg K factor (gg)	$\pm 10 \%$
QCD scale (VBF)	$+0.4/-0.3 \%$
PDF set (VBF)	$\pm 2.1 \%$
QCD scale (WH)	$+0.5/-0.7 \%$
PDF set (WH)	$\pm 1.9 \%$
QCD scale (ZH)	$+3.8/-3.1 \%$
PDF set (ZH)	$\pm 1.6 \%$
QCD scale ($t\bar{t}H$)	$+5.8/-9.2 \%$
PDF set ($t\bar{t}H$)	$\pm 3.6 \%$
$BR(H \rightarrow ZZ \rightarrow 4\ell)$	2 %
QCD scale ($q\bar{q} \rightarrow ZZ$)	$+3.2/-4.2 \%$
PDF set ($q\bar{q} \rightarrow ZZ$)	$+3.1/-3.4 \%$
Electroweak corrections ($q\bar{q} \rightarrow ZZ$)	$\pm 0.1 \%$

the uncertainties $\pi(\boldsymbol{\theta})$ Equation 7.3.

$$\int_0^{\sigma_{\text{up}}} p(\sigma|D) d\sigma = 1 - \alpha \quad (7.1)$$

$$p(\sigma|D) = p(D|\sigma)\pi(\sigma)/p(D) \quad (7.2)$$

$$p(D|\sigma) = \int p(D|\sigma, \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \quad (7.3)$$

Combine takes as input specially formatted text files called data cards, which contain the signal, background, and data yields, along with associated systematic uncertainties, and point to the location of the file containing the shape distributions for each sample. When the shape distributions are included, combine computes the limits across all bins, then combines the results. So, a non-shape based limit is equivalent to a shape-based limit with one bin. This can be used as a cross check that the shape-based limit is behaving correctly, and to analyze the improvement of using the shape-based approach over non-shape based. The event yields and systematic uncertainties are discussed later in this section.

Combine returns the 95% confidence level (CL) expected and observed limits on the signal strength parameter, μ , as well as one and two sigma deviations of the expected limit, where μ is the signal scale factor in the i^{th} bin in the mean count, $n_i = \mu * s_i + b_i$, where s_i and b_i are the signal and background yields, respectively.

Limits are set for the three decay channels, $4e$, 4μ , and $2e2\mu$ individually, then combined for the $4l$ limit. The signal strength parameter gives the ratio of the 95% CL expected signal yield to the theoretical yield, or more generally,

$$\mu[\sigma_{\text{norm}} * BR] = \frac{\sigma_{95\% \text{CL}} * BR}{\sigma_{\text{norm}} * BR}, \quad (7.4)$$

where σ_{norm} is the cross section used to normalize the signal yields given as input to combine, typically the theoretical production cross section, but also set to 1 pb for comparison to other H decay channels. This formula is used to calculate the 95% CL cross section times branching ratio, which is the desired output variable. This limit can then be compared to the theoretical cross section times BR. The points where the upper limit is lower than the theory value are interpreted as exclusions.

7.2 Results

After the event selection is optimized and all of the uncertainties are accounted for, limits are set on the signal strength parameter, μ , and presented in two ways. For each model, the limits are shown for μ , to be easily compared with the limits obtained by the other H decay channels, and as cross section limit versus mass plots. These results are found for both the cut-and-count and MVA based strategies. Until the selection is frozen and the data unblinded, only expected limits will be plotted in the SR.

7.2.1 Cut-and-count based

7.2.1.1 Zp2HDM

The 1D slice of mass points fixing $m_{A0} = 300$ GeV is selected for the limits plots since this region has the highest cross section and the largest branching fraction to DM production. The 1D limits are presented for μ in Figure 7.7 and $\sigma_{95\%CL} * BR$ in Figure 7.8, obtained using the optimized selection given in Section 6.0.1.

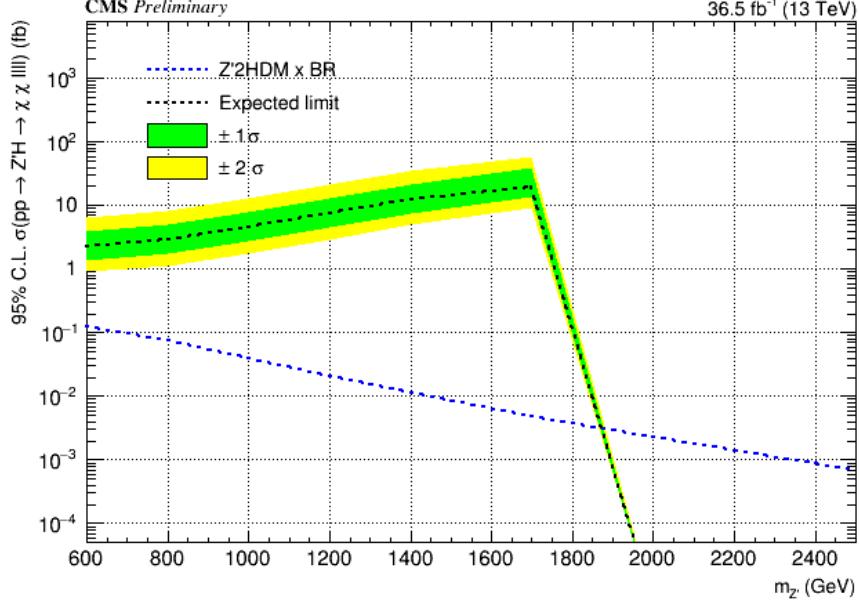


Figure 7.3: PLACEHOLDER 1D μ limits for $m_{A0} = 300$ GeV for Zp2HDM.

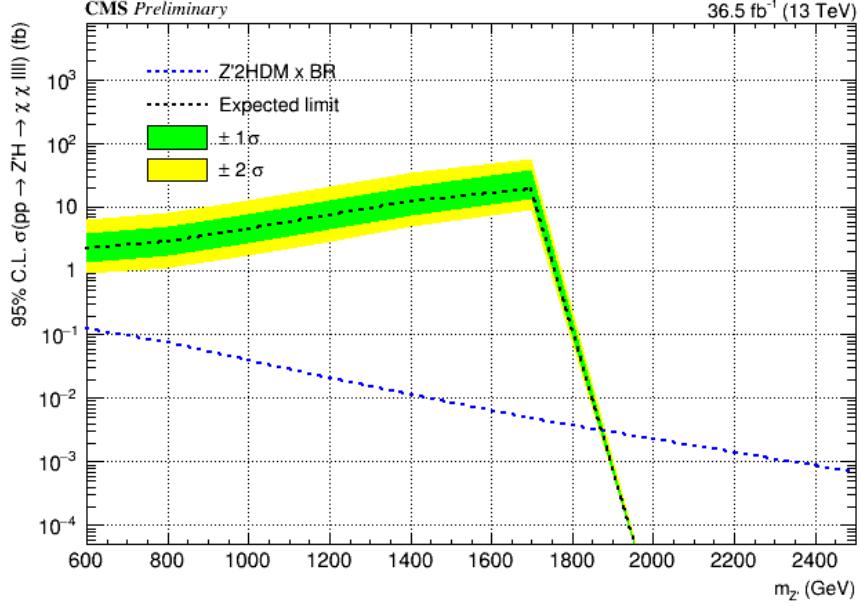


Figure 7.4: PLACEHOLDER 1D cross section times BR limits for $m_{A0} = 300$ GeV for Zp2HDM.

7.2.1.2 ZpBaryonic

The 1D slice of mass points fixing $m_\chi = 1$ GeV is selected for the limits plots since this region has the highest cross section. The 1D limits are presented for μ in Figure 7.9 and $\sigma_{95\%CL} * BR$ in Figure 7.10, obtained using the optimized selection given in Section 6.0.1.

7.2.2 MVA based

7.2.2.1 Zp2HDM

The 1D limits are presented for μ in Figure 7.7 and $\sigma_{95\%CL} * BR$ in Figure 7.8, obtained using the optimized selection given in Section 6.0.2.

7.2.2.2 ZpBaryonic

The 1D limits are presented for μ in Figure 7.9 and $\sigma_{95\%CL} * BR$ in Figure 7.10, obtained using the optimized selection given in Section 6.0.2.

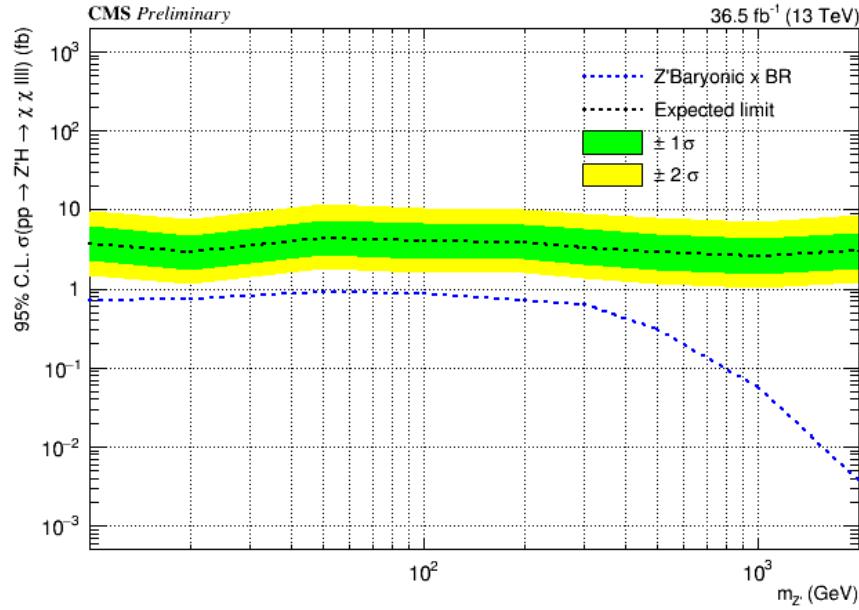


Figure 7.5: PLACEHOLDER 1D μ limits for $m_\chi = 1$ GeV for ZpBaryonic.

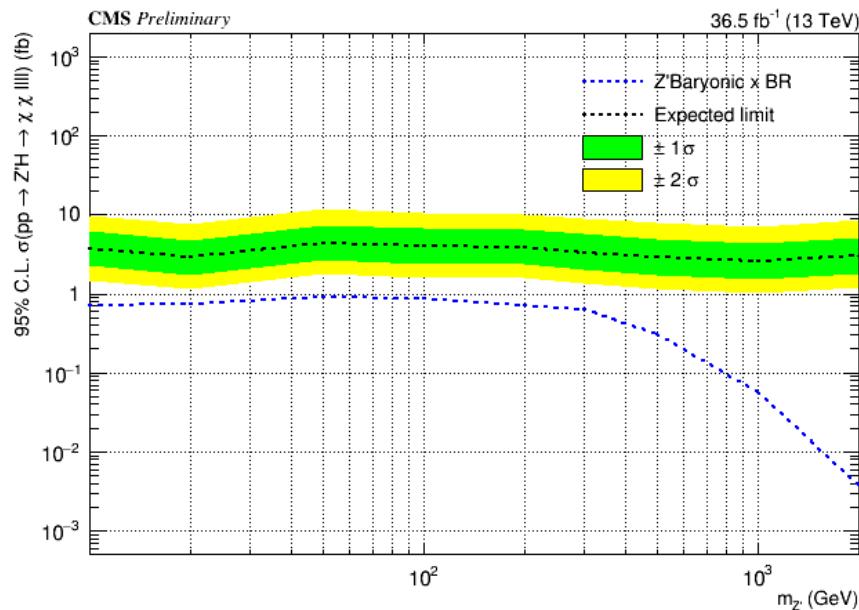
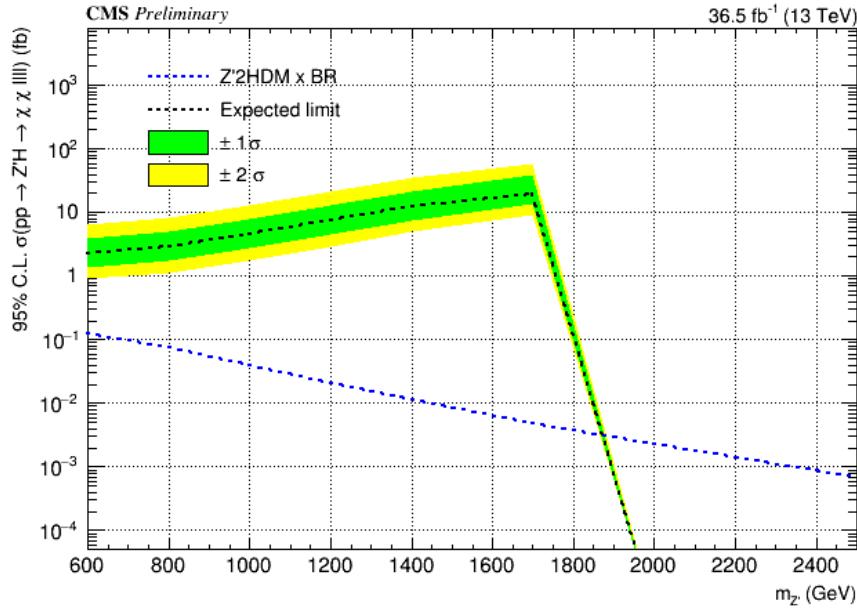
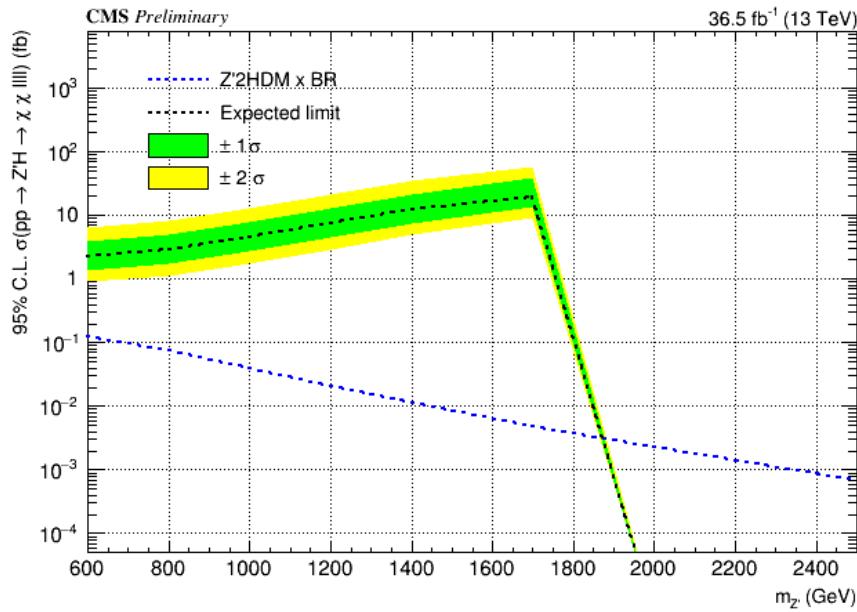


Figure 7.6: PLACEHOLDER 1D cross section times BR limits for $m_\chi = 1$ GeV for ZpBaryonic.

Figure 7.7: PLACEHOLDER 1D μ limits for $m_{A0} = 300$ GeV for Zp2HDM.Figure 7.8: PLACEHOLDER 1D cross section times BR limits for $m_{A0} = 300$ GeV for Zp2HDM.

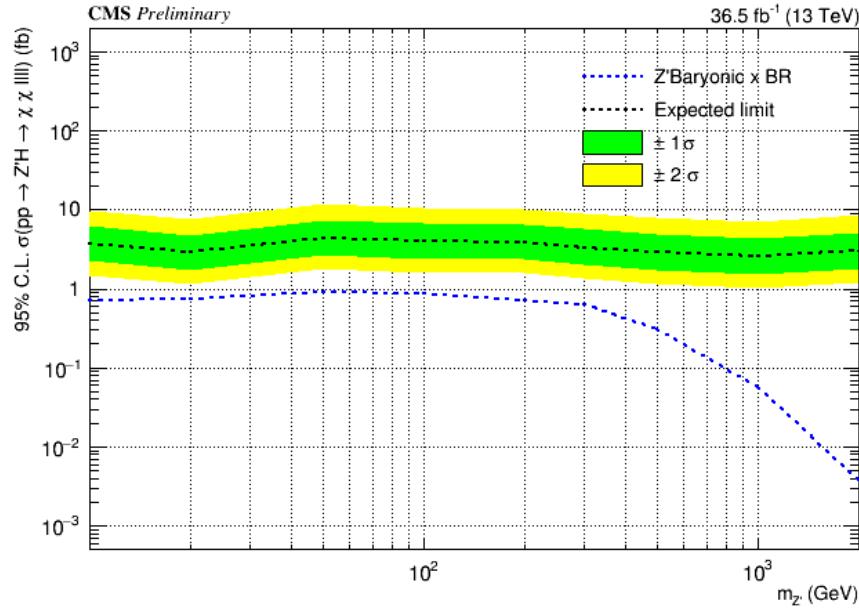


Figure 7.9: PLACEHOLDER 1D μ limits for $m_\chi = 1$ GeV for ZpBaryonic.

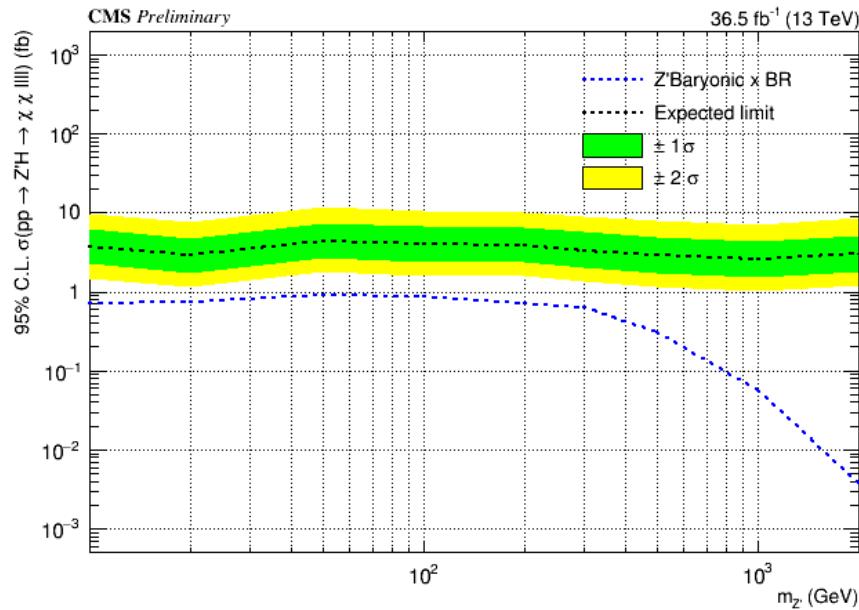


Figure 7.10: PLACEHOLDER 1D cross section times BR limits for $m_\chi = 1$ GeV for ZpBaryonic.

Chapter 8

Conclusions

The discovery of the Higgs boson was the highlight of the physics results of Run I at the LHC. In addition to measuring the properties and couplings of the Higgs, efforts were launched to use the Higgs as a probe for new physics, including to search for dark matter. In parallel with these studies, progress was made by complementary, non-collider searches. Although no direct detection has been confirmed, large areas of parameter space have been excluded. Collider searches, including the search for the mono-Higgs signature, have much better sensitivity to low-mass DM than direct searches, as well as the ability to study higher order couplings that are not accessible to direct searches. The main results of this thesis are the rediscovery of the Higgs boson with Run II data and the cross section limits on two DM models, contributing to the world-leading low-mass limits and mediator mass exclusions for these models.

The impact of this study is very high, being among the first ever searches for the mono-Higgs signature. Several key contributions were made to the SM Higgs search in the four lepton final state, including a cross-check exercise to validate the event selection with other groups, and the definition of a new event category for Higgs candidate events with large missing transverse momentum. The key ingredient that this study added to the SM search and collaboration documentation is the study of E_T^{MISS} . In particular, whether the background and statistical modeling techniques remain valid in the high E_T^{MISS} regime.

There are numerous extensions of this analysis that would have a large impact on the field. The limits found with the four-lepton final state can be combined with the results

from the other Higgs decay channel analyses to obtain even stronger results. This work is underway. There are additional models that predict a mono-Higgs signature that can be studied with the datasets currently available. The most commonly studied models are those that predict very hard E_T^{MISS} spectra, which gives an advantage to other Higgs decay channels. Many of the unstudied models with softer spectra would favor the four lepton channel. Finally, a more detailed study can be done to understand the relationship of the limits found here with those set by other search strategies in order to guide the design of future analyses and experiments.

There is strong motivation for a coupling between the dark sector and ordinary matter. Studies such as this one shed light on the nature of this interaction. With technology and analysis techniques constantly advancing, it is only a matter of time before DM is observed, expanding our understanding of the mass-energy content of the universe beyond the mere 5% that we currently know.

Appendices

Appendix A

Production Cross Sections for Benchmark Signal Models

m_χ [GeV]	1	10	50	65	100	200	400	800	1000	1300
EFT_HHxx_scalar										
EFT_HHxx_combined	0.10071E+01	0.99793E+00	0.60671E+00	0.48291E-04	0.22725E-05	0.11059E-06	0.36569E-08	0.40762E-10	0.64956E-11	0.51740E-12
EFT_HHxg5x	0.15731E+01	0.15194E+01	0.34134E+00	0.41039E-04	0.10581E-04	0.16553E-05	0.14628E-06	0.40608E-08	0.85950E-09	0.96480E-10
EFT_xdxHDHc	0.15735E+03	0.15394E+03	0.94804E+02	0.12990E-01	0.23075E-02	0.41820E-03	0.45743E-04	0.16734E-05	0.39327E-06	0.49769E-07
$\Lambda = 100$ GeV	0.29530E+00	0.29067E+00	0.10540E+00	0.89849E-01	0.64959E-01	0.30639E-01	0.88644E-02	0.97986E-03	0.33847E-03	0.68674E-04
$\Lambda = 1000$ GeV	0.16306E-04	0.15508E-04	0.12088E-05	0.88288E-06	0.53312E-06	0.18046E-06	0.34918E-07	0.27514E-08	0.90662E-09	0.19313E-09
EFT_xgxFHDH	0.57027E+00	0.57001E+00	0.56025E+00	0.55337E+00	0.53270E+00	0.45792E+00	0.29777E+00	0.10288E+00	0.57444E-01	0.23260E-01

Table A.1: EFT production cross sections [pb]

m_χ [GeV]	m_S [GeV]					
1	0.21915E+01	0.20798E+01	0.19192E+01	0.18118E+01	0.16735E+01	0.52244E+01
10	0.17416E+01	0.17420E+01	0.18581E+01	0.17510E+01	0.16735E+01	0.41877E+01
50	0.39053E+00	0.38877E+00	0.38409E+00	0.37097E+00	0.12861E+01	0.28732E+01
150	0.24136E-05			0.38372E-05	0.21922E-04	0.42337E-03
500	0.34099E-08					0.49399E-08
1000	0.17012E-10					0.55260E-10

Table A.2: Scalar production cross sections [pb] corresponding to mass points in Table ??

m_χ [GeV]	$m_{Z'}$ [GeV]					
1	0.26615E+01	0.27802E+01	0.33248E+01	0.32341E+01	0.26566E+01	0.23191E+01
10	0.21182E-01	0.74027E-01	0.32732E+01	0.32250E+01	0.10842E+01	0.18700E+00
50	0.36342E-03	0.12726E-01	0.31337E+00	0.21226E+01	0.20120E+01	0.11728E-01
150	0.55972E-05			0.56526E-02	0.18000E+00	0.17340E-07
500	0.80295E-08				0.36591E-04	0.16918E-07
1000	0.49387E-10				0.98079E-06	0.13179E-07
					0.57596E-03	0.80146E-08

Table A.3: ZpBaryonic production cross sections [pb] corresponding to mass points in Table ??

m_χ [GeV]	$m_{Z'}$ [GeV]					
1	0.61935E-02	0.63192E-02	0.82991E-02	0.11942E-01	0.19171E-01	0.21560E-01
10	0.58781E-02	0.58938E-02	0.82944E-02	0.11937E-01		0.16010E-01
50	0.10294E-03		0.77820E-04	0.15258E-04	0.12066E-01	0.12105E-01
150	0.28382E-06				0.83917E-05	0.72889E-03
500	0.34689E-09					0.43355E-06
1000	0.20703E-11					0.44782E-07

Table A.4: ZpHS production cross sections [pb] corresponding to mass points in Table ??

m_{A^0} [GeV]	$m_{Z'}$ [GeV]							
300	42.386	45.097	35.444	26.07	18.942	11.778	7.4456	3.6446
400	5.8513	14.847	14.534	11.792	9.029	5.851	3.7819	1.8758
500		5.9605	8.4961	7.9575	6.5515	4.5063	3.0028	1.5235
600		1.5853	4.6972	5.4808	4.9946	3.7044	2.5694	1.3447
700			2.1092	3.4848	3.6766	3.0253	2.2023	1.1984
800				0.65378	1.9638	2.5511	2.4077	1.8689
								1.0692

Table A.5: Zp2HDM production cross sections [pb] corresponding to mass points in Table ??

REFERENCES

- [1] Massimiliano Grazzini, Stefan Kallweit, and Dirk Rathlev. ZZ production at the LHC: Fiducial cross sections and distributions in NNLO QCD. *Phys. Lett. B*, 750:407 – 410, 2015. URL: <http://www.sciencedirect.com/science/article/pii/S0370269315007303>, arXiv:1507.06257, doi:<http://dx.doi.org/10.1016/j.physletb.2015.09.055>.
- [2] John M. Campbell and R. K. Ellis. MCFM for the Tevatron and the LHC. *Nucl. Phys. Proc. Suppl.*, 205:10, 2010. arXiv:1007.3492, doi:[10.1016/j.nuclphysbps.2010.08.011](https://doi.org/10.1016/j.nuclphysbps.2010.08.011).
- [3] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Vector boson pair production at the LHC. *JHEP*, 07:018, 2011. arXiv:1105.0020, doi:[10.1007/JHEP07\(2011\)018](https://doi.org/10.1007/JHEP07(2011)018).
- [4] John M. Campbell, R. Keith Ellis, and Ciaran Williams. Bounding the Higgs width at the LHC using full analytic results for $gg \rightarrow e^-e^+\mu^-\mu^+$. *JHEP*, 04:060, 2014. arXiv:1311.3589, doi:[10.1007/JHEP04\(2014\)060](https://doi.org/10.1007/JHEP04(2014)060).
- [5] Marco Bonvini, Fabrizio Caola, Stefano Forte, Kirill Melnikov, and Giovanni Ridolfi. Signal-background interference effects in $gg \rightarrow H \rightarrow WW$ beyond leading order. *Phys. Rev. D*, 88:034032, 2013. arXiv:1304.3053, doi:[10.1103/PhysRevD.88.034032](https://doi.org/10.1103/PhysRevD.88.034032).
- [6] Kirill Melnikov and Matthew Dowling. Production of two Z-bosons in gluon fusion in the heavy top quark approximation. *Phys. Lett. B*, 744:43, 2015. arXiv:1503.01274, doi:[10.1016/j.physletb.2015.03.030](https://doi.org/10.1016/j.physletb.2015.03.030).
- [7] Chong Sheng Li, Hai Tao Li, Ding Yu Shao, and Jian Wang. Soft gluon resummation in the signal-background interference process of $gg(\rightarrow h^*) \rightarrow ZZ$. 2015. arXiv:1504.02388.
- [8] Giampiero Passarino. Higgs CAT. *Eur. Phys. J. C*, 74:2866, 2014. arXiv:1312.2397, doi:[10.1140/epjc/s10052-014-2866-7](https://doi.org/10.1140/epjc/s10052-014-2866-7).
- [9] Stefano Catani and Massimiliano Grazzini. An NNLO subtraction formalism in hadron collisions and its application to Higgs boson production at the LHC. *Phys. Rev. Lett.*, 98:222002, 2007. arXiv:hep-ph/0703012, doi:[10.1103/PhysRevLett.98.222002](https://doi.org/10.1103/PhysRevLett.98.222002).
- [10] Massimiliano Grazzini. NNLO predictions for the Higgs boson signal in the $H \rightarrow WW \rightarrow l\nu l\nu$ and $H \rightarrow ZZ \rightarrow 4l$ decay channels. *JHEP*, 02:043, 2008. arXiv:0801.3232, doi:[10.1088/1126-6708/2008/02/043](https://doi.org/10.1088/1126-6708/2008/02/043).
- [11] Massimiliano Grazzini and Hayk Sargsyan. Heavy-quark mass effects in Higgs boson production at the LHC. *JHEP*, 09:129, 2013. arXiv:1306.4581, doi:[10.1007/JHEP09\(2013\)129](https://doi.org/10.1007/JHEP09(2013)129).

- [12] LHC Higgs Cross Section Working Group. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. CERN Report CERN-2013-004, 2013. [arXiv:1307.1347](https://arxiv.org/abs/1307.1347), doi: 10.5170/CERN-2013-004.