

Searches for invisibly decaying Higgs bosons produced through vector boson fusion at 13 TeV and cloud computing for high energy physics with the Compact Muon Solenoid experiment

Riccardo Di Maria

Imperial College London
Department of Physics

A dissertation submitted to Imperial College London
for the degree of Doctor of Philosophy

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Abstract

Searches for invisible decays of the Higgs boson, produced via vector boson fusion, are presented in this thesis. Data from proton-proton collisions at $\sqrt{s} = 13$ TeV, collected with the CMS detector during 2016 and corresponding to an integrated luminosity of 35.9 fb^{-1} , are used. The data are in agreement with the standard model prediction, and the observed (expected) upper limits on the invisibly decaying Higgs boson branching ratio $\mathcal{B}(\text{H} \rightarrow \text{inv.})$ are 0.58(0.30) and 0.33(0.25) at 95% CL for the ‘cut-based’ and ‘shape-based’ approaches, respectively, as published in Ref. [1]. Several combinations with other relevant analyses are performed using data from Run-1, Run-2 (2015), and 2016 to further improve the sensitivity to $\mathcal{B}(\text{H} \rightarrow \text{inv.})$, and the results are published in Refs. [1, 2]. The observed (expected) upper limit on $\mathcal{B}(\text{H} \rightarrow \text{inv.})$ is set at 0.19(0.15) at 95% CL from the combination performed with all datasets, and it is the most stringent limit to date. Interpretations of the result under non-SM production assumptions and Higgs-portal models are also provided. The observed (expected) 95% CL upper limit on $\frac{\sigma}{\sigma_{\text{SM}}} \times \mathcal{B}(\text{H} \rightarrow \text{inv.})$ varies in the range $[0.14, 0.24]([0.11, 0.19])$, assuming non-SM production cross-sections. The observed 90% CL upper limit of $\mathcal{B}(\text{H} \rightarrow \text{inv.}) < 0.16$ is translated into an upper limit on the spin-independent DM-nucleon elastic scattering cross-section. This limit is the most stringent constraint for $m_\chi < 18 \text{ GeV}$ or $< 7 \text{ GeV}$ assuming a fermion or a scalar DM candidate, respectively.

R&D computing activities within the Computing & Offline Project at CMS are presented in this thesis. The results from the pioneering implementation of the dynamic on demand analysis service (DODAS) of CMS in different cloud environments and the benchmarking of high energy physics use-cases are an important milestone in CMS, as published in Ref. [3]. The feasibility studies for the integration of production workflows are some examples to which the work in this thesis significantly contributes. Moreover, the exploitation of cloud resources from UK’s Amazon Web Services demonstrates the full exportability of DODAS, which thus provisions on-demand resources in the Grid without any constraint on their physical location.

Declaration

I, the author of this thesis, declare that this document is the result of my own work. The main body of the text describes the studies conducted and the results produced, which are supported by the referenced publications. Figures labelled ‘CMS’ are sourced from publications of the CMS Collaboration, including those from my own work. Figures, tables, and studies, which are not labelled or referenced have been internally approved within the CMS Collaboration as part of the internal review of the associated publications. Figures, tables, studies, and results taken from external sources are referenced appropriately throughout this document.

Riccardo Di Maria

Acknowledgements

To my parents and my brother, without whom I wouldn't be the man I am. You have always supported me, from swimming to physics, and encouraged me to stand up, and work hard. I will never be grateful enough.

To Chiara, my future wife. You put up with my quirks, and that is quite something. You have borne with and stood by me during these years of my PhD, and beyond. I can't wait for spending the rest of my life with you. When I am with you, there's no place I'd rather be.

To my supervisors, Gavin and Dave, who gave me this chance and showed me the way, in physics and in the real world.

To the ICL HEP group, especially Nick and Anne-Marie, who walked me through everything, being extremely patient.

To the Italian counterparts, Daniele B., Claudio G., and Daniele S., the 'computing' support everyone should have in life.

To my office mates, now friends, who were the only ones that could deeply understand what was in my head during these almost four years since they were living the same crazy experience that humans call a PhD.

To the Italian posse, summarised in one person, friend, Vito. Sometimes, silence speaks louder than words.

...a profound thank you.

Contents

List of figures	xiii
List of tables	xix
1 Introduction	1
2 High Energy Particle Physics at the LHC	3
2.1 The Large Hadron Collider	3
2.2 The Compact Muon Solenoid Experiment	7
2.3 The CMS Detector	9
2.3.1 The Tracker System	10
2.3.2 The Electromagnetic Calorimeter	13
2.3.3 The Hadron Calorimeter	14
2.3.4 The Magnet	15
2.3.5 The Muon System	16
2.3.6 The Trigger and Data Acquisition System	17
3 The Higgs Mechanism and Dark Matter	19
3.1 The Standard Model of Particle Physics	19
3.1.1 Elementary Particles in Nature	20
3.2 Gauge Theory, Symmetries, and the Goldstone Model	21
3.2.1 Spontaneous Symmetry Breaking (SSB)	23
3.3 Spontaneous Symmetry Breaking in Gauge Theories	24
3.4 Spontaneous Symmetry Breaking of the Electroweak Theory and the Higgs Mechanism	26
3.4.1 The Yukawa Mechanism	31
3.5 Dark Matter and Invisibly Decaying Higgs Bosons	31
3.5.1 Higgs to Invisible Decays as a Search for Dark Matter	33
3.6 Higgs Production and Decay Modes	34

4	Event Reconstruction and Analysis Objects	39
4.1	Particle-Flow	39
4.2	Physics Objects	40
4.2.1	Tracks	40
4.2.2	Primary Vertex	41
4.2.3	Electrons	41
4.2.4	Photons	43
4.2.5	Muons	44
4.2.6	Tau Leptons	45
4.2.7	Jets	45
4.2.8	B-tagged Jets	47
4.2.9	MET	47
5	Search for VBF Higgs Bosons Decaying to Invisible Final States	53
5.1	Samples	56
5.1.1	Datasets	56
5.1.2	Signal Samples	56
5.1.3	Background Samples	57
5.2	Trigger	58
5.2.1	MET Triggers	58
5.2.2	Electron Triggers	60
5.3	Rewighting Procedure	61
5.3.1	Trigger Efficiency and Pileup Weight	61
5.3.2	Lepton and Photon Efficiency and Reconstruction SF	61
5.3.3	B-tagging and Tau Vetos Scale Factors	62
5.3.4	NLO Corrections for V+jets QCD Production	62
5.3.5	NLO Corrections for V+jets EW Production	63
5.4	Event Selection	66
5.5	Control Regions	70
5.5.1	Z Control Regions	70
5.5.2	W Control Regions	74
5.5.3	The QCD-multijet Background	77
5.6	Systematic Uncertainties	79
5.6.1	Experimental Uncertainties on V+jets	80
5.6.2	Theory Uncertainties on V+jets	83
5.6.3	Systematic Uncertainties on Minor Backgrounds	84

5.6.4	Systematic Uncertainties on Higgs to Invisible Signals	85
5.7	Confidence Levels and Exclusion Limits	86
5.8	Signal Kinematics and Data-MC Optimisation	87
5.8.1	Optimisation Procedure	87
5.9	Results	92
5.9.1	The Cut-based Approach	92
5.9.2	The Shape-based Approach	96
5.9.3	Constraints on SM-like Higgs Bosons	98
6	Combinations of Higgs to Invisible Searches	99
6.1	Searches for Invisibly Decaying Higgs Bosons	99
6.2	Combination of Searches in Run-1 and Run-2 (2015)	100
6.2.1	The VBF Analysis	101
6.2.2	The $Z \rightarrow \ell^+ \ell^-$ Analysis	102
6.2.3	The $V \rightarrow jj$ and Mono-jet Analyses	104
6.2.4	Results	105
6.3	Combination of Searches in Run-1 and Run-2 (2015+2016)	108
6.3.1	Non-SM Production	109
6.3.2	DM Interpretations	109
7	Cloud Computing for High Energy Physics	113
7.1	Grid Computing at the LHC	113
7.1.1	Services and Elements of the Grid	114
7.1.2	The Tiered Structure of the WLCG	115
7.2	Computing in CMS	116
7.2.1	The CMS Workload Management System	116
7.2.2	The CMS Data Management System	117
7.2.3	Cloud Computing and Opportunistic Resources	118
7.3	Dynamic Resource Provisioning and Dynamic Clusters	120
7.3.1	Dynamic On-Demand Analysis Service	120
7.4	DODAS at Imperial College London, UK	123
7.4.1	Underlying Infrastructure and Implementation	123
7.4.2	CMS Workflows	127
7.4.3	Hammer Cloud Workflow	128
7.4.4	Results	128
7.5	Amazon Web Services for DODAS	131
7.5.1	Implementation	131

7.5.2 Results	132
8 Conclusions and Outlook	135
8.1 Physics Analyses	135
8.2 Cloud Computing	137
A Z Control Regions	139
B W Control Regions	143
Bibliography	147
Acronyms	159

List of figures

2.1	The LHC's injection chain with its multiple smaller pre-accelerators [7].	4
2.2	The measured luminosity delivered by the LHC to CMS during stable beams in Run-1 and Run-2 for pp collisions [15].	6
2.3	The LHC baseline plan from 2010 to 2023, including a forecast of the integrated and peak luminosities [16].	7
2.4	The cross-sections expressed in nb for several processes in pp (or p $\bar{p}$) collisions as a function of the centre-of-mass energy $\sqrt{s}$ in TeV [17].	8
2.5	A schematic longitudinal view of the five wheels in the barrel region, of the three discs in the end-caps, and of the sub-detectors in the very-forward region of the CMS experiment [20].	10
2.6	An illustrative transverse section of the CMS experiment showing the path of different particles [21].	11
2.7	A schematic view in the r - z plane of the CMS tracker where each detector module is pictured as a line-element [23].	12
2.8	The CMS Level-1 trigger for LHC Run-2.	18
3.1	The illustration of the potential $V(\phi)$ for $\mu^2 < 0$. SSB of $U(1)$ symmetry occurs when a particular minimum is chosen [48].	25
3.2	The rotational velocity of the galaxy NGC 6503 as a function of the distance from the galactic centre. The disc and gas contributions necessitate a dark matter halo to match the observed data [49].	32
3.3	(left): Contours of the Bullet Cluster mass distribution from gravitational lensing (optical images from the Magellan telescope). (right): Data from Chandra X-ray that traces hot plasma in a galaxy, plotted over the same contours [52].	33
3.4	The Feynman diagrams of the three production modes targeted in the search for invisibly decaying Higgs bosons: (a) ggH (with ISR), (b) VBF, and (c) VH [2].	34
3.5	SM Higgs boson production cross-sections at $\sqrt{s} = 13$ TeV [53].	35

3.6	The measurements on the Higgs boson couplings with SM particles and the estimated $\mathcal{B}(\text{H} \rightarrow \text{inv.})$ (B_{BSM}) from visible channels [57].	38
4.1	The distribution of MET for data and simulated backgrounds in the $Z \rightarrow \mu^+ \mu^-$ (left) and $Z \rightarrow e^+ e^-$ (right) CRs.	48
4.2	The distribution of MET for data and simulated backgrounds in the $W \rightarrow \mu^\pm \nu$ (left) and $W \rightarrow e^\pm \nu$ (right) CRs.	48
4.3	The distribution of $E_{\text{T,no-lep}}^{\text{miss}}$ for data and simulated backgrounds in the $Z \rightarrow \mu^+ \mu^-$ (left) and $Z \rightarrow e^+ e^-$ (right) CRs.	50
4.4	The distribution of $E_{\text{T,no-lep}}^{\text{miss}}$ for data and simulated backgrounds in the $W \rightarrow \mu^\pm \nu$ (left) and $W \rightarrow e^\pm \nu$ (right) CRs.	50
4.5	The difference between the calorimetric and PF MET relative to the recoil for data and simulated backgrounds passing a very loose VBF selection in the $W \rightarrow e^\pm \nu$ CR.	51
5.1	Schematic overview of Chapter 5 and the approach used to present the VBF Higgs to invisible analysis.	55
5.2	The comparison between $p_{\text{T}}^{\text{miss}} + H_{\text{T}}^{\text{miss}}$ (a) and VBF (b) trigger turn-on curves in simulated VBF Higgs to invisible events.	59
5.3	The trigger efficiencies of single-muon events selected from the SINGLE-MUON-triggered dataset for different jet η categories as a function of m_{jj}	60
5.4	The higher-order corrections on the Z (left) and W (right) p_{T} distributions after an inclusive selection.	63
5.5	The comparison of higher-order QCD corrections on the Z+jets (top left), $Z(\ell^+ \ell^-)$ +jets (top right), and W+jets (bottom) processes as a function of boson p_{T} . The black distribution refers to inclusive NLO k-factors, while corrections obtained after applying the VBF cut-and-count (loose) selection are shown by the red (blue) histogram.	64
5.6	The higher-order NLO-QCD corrections on the Z+jets-EW (top) and W+jets-EW (bottom) samples as a function of m_{jj} and p_{T}^{V} . The vertical axis on the right of each plot indicates the magnitude of the NLO-QCD k-factors applied.	65
5.7	The distributions of $\min \Delta \phi(E_{\text{T}}^{\text{miss}}, j)$ (left), obtained by removing its corresponding cut, and $E_{\text{T,no-}\mu}^{\text{miss}}$ (right) for data and simulated signal and backgrounds in the signal region.	67

5.8	The distributions of m_{jj} (left) and $\Delta\eta_{jj}$ (right) for data and simulated signal and backgrounds in the signal region.	68
5.9	The distributions of p_T (top) and η (bottom) for the leading (left) and sub-leading (right) jets for data and simulated signal and backgrounds in the signal region.	69
5.10	The distributions of p_T (top) and η (central) for the leading (left) and sub-leading (right) jets, and invariant mass (bottom left) and p_T (bottom right) for the muon pair for data and simulated signal and backgrounds in the $Z \rightarrow \mu^+\mu^-$ CR.	72
5.11	The distributions of p_T (top) and η (central) for the leading (left) and sub-leading (right) jets, and invariant mass (bottom left) and p_T (bottom right) for the electron pair for data and simulated signal and backgrounds in the $Z \rightarrow e^+e^-$ CR.	73
5.12	The distributions of p_T (top) and η (central) for the leading (left) and sub-leading (right) jets, transverse mass of the muon+MET system (bottom left), and $\min\Delta\phi(E_T^{\text{miss}}, j)$ (bottom right) for data and simulated signal and backgrounds in the $W \rightarrow \mu^\pm\nu$ CR.	75
5.13	The distributions of p_T (top) and η (central) for the leading (left) and sub-leading (right) jets, transverse mass of the electron+MET system (bottom left), and $\min\Delta\phi(E_T^{\text{miss}}, j)$ (bottom right) for data and simulated signal and backgrounds in the $W \rightarrow e^\pm\nu$ CR.	76
5.14	The distribution of m_{jj} for QCD-multijet simulated events in Region-A (left) and Region-B (right). The points show the yields in each bin from simulation, while the red lines shows the fitted exponential functions. . .	78
5.15	The expected yield of QCD-multijet above a given m_{jj} threshold in Region-A (left) and Region-B (centre). The blue points show the prediction from simulation, whereas the black lines refer to the prediction using the fitted exponential functions. The ratio $R(m_{jj})$ as a function of the threshold (right). The uncertainties represent the statistical uncertainty from the fit to simulation.	78
5.16	The distribution of p_T^V for Z (top) and W (bottom) simulated samples in the SR and CRs.	81
5.17	The distribution of number of vertices for Z (top) and W (bottom) simulated samples in the SR and CRs.	82

5.18	The Z/W ratio obtained from simulation as a function of p_T^V for several phase space selections. The grey band represents the total systematic uncertainty from theory.	83
5.19	The comparison of various observables between signal events, produced through VBF (black) and ggH (red) modes and decaying invisibly, and Z+jets (blue) and W+jets (green) backgrounds.	88
5.20	The expected upper limits on $\mathcal{B}(H \rightarrow \text{inv.})$ as a function of the optimisation variables. The fit includes the ggH channel in the signal prediction and all systematic uncertainties (top), no systematic uncertainties (central), and no systematic uncertainties and only the signal region (bottom).	90
5.21	The distribution of $\Delta\phi_{jj}$ for data and simulated signal and backgrounds in the signal region.	91
5.22	The impacts of the nuisance parameters from a fit to data in the SR and CRs.	94
5.23	The observed distribution of m_{jj} in the SR compared to the yields of various SM background processes resulting from the CR-only fit (left) and b-only fit (right), following the shape-based approach. The distribution of a signal for an invisibly decaying Higgs boson is overlaid for both VBF and ggH production modes, assuming $m_H = 125 \text{ GeV}$ and $\mathcal{B}(H \rightarrow \text{inv.}) = 100\%$ [1].	97
5.24	The observed (solid black) and expected (dashed black) 95% CL upper limits on $\frac{\sigma}{\sigma_{\text{SM}}} \times \mathcal{B}(H \rightarrow \text{inv.})$ as a function of the SM-like Higgs boson mass m_H for the cut-based (left) and shaped-based (right) approaches. The inner (green) and outer (yellow) band indicate the regions containing 68% and 95%, respectively, of the distribution of limits expected under the background-only hypothesis [1].	98
6.1	The observed (solid black) and expected (dashed black) upper limits at 95% CL on $\frac{\sigma}{\sigma_{\text{SM}}} \times \mathcal{B}(H \rightarrow \text{inv.})$ for individual combinations of categories, i.e. qqH, VH, and ggH tagged, and for the full combination, assuming $m_H = 125 \text{ GeV}$. The inner (green) and outer (yellow) band indicate the regions containing 68% and 95%, respectively, of the distribution of limits expected under the background-only hypothesis [2].	106

6.2	The observed (solid black) and expected (dashed black) profile likelihood ratios as a function of $\mathcal{B}(\text{H} \rightarrow \text{inv.})$, assuming SM production cross-sections of a Higgs boson with $m_{\text{H}} = 125 \text{ GeV}$. The results are reported for partial combinations of 7+8 TeV and 13 TeV searches and for the full combination [2].	107
6.3	The observed 95% CL upper limits on $\frac{\sigma}{\sigma_{\text{SM}}} \times \mathcal{B}(\text{H} \rightarrow \text{inv.})$ as a function of κ_{F} and κ_{V} , assuming $m_{\text{H}} = 125.09 \text{ GeV}$. The best estimate from LHC and its 68% and 95% CL regions for the κ -model from Ref. [57] are reported. The SM production rates are obtained for $\kappa_{\text{F}} = \kappa_{\text{V}} = 1$ [1].	110
6.4	The 90% CL upper limit on the spin-independent DM-nucleon elastic scattering cross-section as a function of m_{χ} , for both fermion (dashed red) and scalar (solid orange) DM candidates in Higgs-portal models. The limits from the XENON1T [89], PandaX-II [90], LUX [91], CDEX-10 [92], CDMSlite [93], and CRESST-II [94] direct-detection experiments are also reported [1].	111
7.1	A schematic view of the tiered structure of the WLCG within the CMS computing model [107].	116
7.2	A schematic view of the DODAS architecture, where the IaaS resource (orange), PaaS core service (blue), and user domain (white) layers are shown [3].	121
7.3	The jobs running at the T3_UK_Opportunistic site between May 1st and 4th, 2019.	130
7.4	The jobs completed at the T3_UK_Opportunistic site between March 1st and May 4th, 2019.	133
7.5	The average CPU efficiency at the T3_UK_Opportunistic site between April 20th and May 10th, 2019.	133
A.1	The distribution of m_{jj} for data and simulated signal and backgrounds in the $Z \rightarrow \mu^+ \mu^-$ (left) and $Z \rightarrow e^+ e^-$ (right) CRs.	140
A.2	The distributions of p_{T} (top) and η (bottom) for the leading (left) and sub-leading (right) muons for data and simulated signal and backgrounds in the $Z \rightarrow \mu^+ \mu^-$ CR.	141
A.3	The distributions of p_{T} (top) and η (bottom) for the leading (left) and sub-leading (right) electrons for data and simulated signal and backgrounds in the $Z \rightarrow e^+ e^-$ CR.	142

B.1	The distribution of m_{jj} for data and simulated signal and backgrounds in the $W \rightarrow \mu^\pm \nu$ (left) and $W \rightarrow e^\pm \nu$ (right) CRs.	144
B.2	The distributions of p_T (left) and η (right) for the leading muon for data and simulated signal and backgrounds in the $W \rightarrow \mu^\pm \nu$ CR.	145
B.3	The distributions of p_T (left) and η (right) for the leading electron for data and simulated signal and backgrounds in the $W \rightarrow e^\pm \nu$ CR.	145

List of tables

3.1	The elementary fermions observed in nature categorised into their three generations. Each of these particles has an antiparticle characterised by an opposite charge but identical mass. The magnitude of the electron charge is used as a unit for the electric charge of all particles [40].	21
3.2	The elementary bosons observed in nature, mediators of the related force. The magnitude of the electron charge is used as a unit for the electric charge of all particles [40].	21
4.1	The veto and tight electron identification criteria.	43
4.2	The loose photon identification criteria.	44
5.1	The summary of the selection criteria applied in the analysis for both shape-based and cut-based approaches.	66
5.2	The experimental uncertainties affecting transfer factors used in the analysis to estimate V+jets backgrounds in the SR. In the case of lepton vetoes, uncertainties are applied to the appropriate process in each region.	81
5.3	The systematic uncertainties on minor backgrounds populating the SR and CRs. In addition, all simulated samples include statistical uncertainties not shown here.	84
5.4	The production cross-sections and relative uncertainties for the SM Higgs boson.	85
5.5	The expected and observed event yields in the SR and CRs of the cut-and-count analysis for various SM processes. The background yields and the corresponding uncertainties are obtained from: the MC prediction (pre-fit), a combined fit to data in all CRs, but excluding data in the SR (CR-only fit), a combined fit to data in the SR and CRs assuming the absence of a signal (b-only fit), and a combined fit to data in the SR and CRs allowing for the presence of a signal for an invisibly decaying Higgs boson (s+b fit).	95

6.1	The expected composition of production modes of a 125 GeV Higgs boson for each analysis included in the combination, assuming SM production cross-sections [2].	100
6.2	The expected and observed event yields in the SR and CRs of the 2015 13 TeV analysis for various SM processes. The predicted backgrounds from a CR-only fit are shown with statistical and systematic uncertainties. The SM VBF and ggH production rates are assumed for an invisibly decaying Higgs boson with a mass $m_H = 125$ GeV and $\mathcal{B}(H \rightarrow \text{inv.}) = 100\%$ [2]. . .	103
6.3	The impact of the systematic and statistical uncertainties, and the total uncertainty on the fitted value of $\mathcal{B}(H \rightarrow \text{inv.})$ in the 13 TeV analyses [2].	107
6.4	The expected composition of production modes of a 125 GeV Higgs boson, and observed and expected upper limits on $\mathcal{B}(H \rightarrow \text{inv.})$ for each analysis performed in 2016 and included in the combination, assuming SM production cross-sections [1].	108

*“Your future hasn’t been written yet. No one’s has.
Your future is whatever you make it.
So make it a good one.”*
— Dr. Emmett Brown

Chapter 1

Introduction

Understanding the nature of dark matter is considered by many as the next puzzle to solve within particle physics at colliders after the discovery of the Higgs boson. The energy content of the Universe is 95% in the dark sector (dark energy and dark matter), whereas the visible component contributes only 5%. The Large Hadron Collider (LHC) can contribute to the search for dark matter in various ways. Dark matter can manifest itself as missing energy in the final state, arising from either the decay of the Higgs boson or of a new particle as postulated in several beyond the standard model theories. The direct production of dark matter at the LHC may be realised in Higgs-portal models, where the Higgs boson acts as a mediator between the standard model (SM) and the dark matter particles. In this context, the final state appears as ‘invisible’ for the Compact Muon Solenoid (CMS) detector, and the presence of visible objects, such as jets, is necessary to tag a dark matter event. The decay of the Higgs boson to dark matter or non-SM particles increases the branching fraction $\mathcal{B}(\text{H} \rightarrow \text{inv.})$, which is about 10^{-3} in the SM through the $\text{H} \rightarrow \text{ZZ} \rightarrow 4\nu$ process. Therefore, searching for invisible Higgs boson decays is a powerful method to probe for dark matter at the LHC.

Chapter 2 describes the LHC at CERN and the CMS experiment, with which the data used in this thesis were collected. The detector and its sub-systems are described, with the focus on detailing their characteristics and usage with respect to the search for invisible Higgs boson decays.

Chapter 3 describes the fundamental constituents of matter and their interactions. Gauge theories and symmetries are introduced, and the generalisation to the electroweak theory and the Higgs mechanism is discussed. This Chapter also motivates the search for an invisibly decaying Higgs boson, and the characteristics of the production and decay modes are highlighted.

Chapter 4 presents the event reconstruction and objects used in the Higgs to invisible analysis. The ‘particle-flow’ algorithm used to combine the sub-detector information into individual particles is also discussed, with the latter constituting the input for an analysis.

The main analysis conducted by the author is detailed in Chapter 5: the search for invisibly decaying Higgs bosons produced in the vector boson fusion (VBF) mode. Proton-proton collision data of the LHC Run-2, collected by CMS during 2016, are used to carry out the analysis. The focus of the Chapter is on one of the two methods used in the analysis: the cut-based, or cut-and-count, approach. The author led all aspects of the cut-and-count approach, and also significantly contributed to the shape-based method by performing synchronisation studies and background modelling. The data-driven method used for the background prediction and the reweighting procedure performed to match simulation to data are discussed in detail. A simultaneous fit across the signal and control regions constrains the simulation to data for the extraction of the final result. This Chapter concludes with results from both methods and the limits set on $\mathcal{B}(\text{H} \rightarrow \text{inv.})$, published in Ref. [1].

Combining analyses improves the sensitivity to invisible Higgs boson decays, and various combinations are presented in Chapter 6. The combination of searches in Run-1 and in the first part of Run-2 (2015) was conducted by the author and is published in Ref. [2]. This is further combined with searches performed with the Run-2 (2016) dataset, among which the VBF Higgs to invisible analysis described in Chapter 5 is the most sensitive. The limits are also re-interpreted under non-SM production assumptions and Higgs-portal models, and the results from this combination are published in Ref. [1].

Being able to optimally handle such a vast quantity of data, as collected by CMS, has become an essential pillar for physics analyses at the LHC. The computing activities carried out at CERN are thus vital to perform these dark matter searches. Chapter 7 presents the R&D activities performed by the author within the Computing & Offline Project at CMS, published in Ref. [3]. The dynamic on-demand analysis service (DODAS) in CMS provides opportunistic resources, whose integration with user and production workflows is discussed. A pioneering implementation of the service in different cloud environments and the benchmarking of high energy physics use-cases are presented, including the exploitation of opportunistic resources from UK’s Amazon Web Services.

Finally, Chapter 8 summarises the results obtained by the author in Chapters 5 to 7. In addition, an outlook is given with respect to potential future objectives and improvements.

Chapter 2

High Energy Particle Physics at the LHC

This Chapter describes the Large Hadron Collider at CERN and the Compact Muon Solenoid experiment. Section 2.1 gives an overview of the collider and the chain of accelerators used to reach a centre-of-mass energy of $\sqrt{s} = 13$ TeV. Section 2.2 focuses on the general organisation of the experiment and its scope, and Section 2.3 describes the detector and its sub-systems.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [4] is located at CERN (European Organisation for Nuclear Research), at about 100 m beneath the France-Switzerland border near Geneva. The LHC is the largest superconducting proton-proton (pp) and heavy ions (PbPb) collider, and it is installed in the 26.7 km tunnel that previously hosted the Large Electron-Positron Collider (LEP) [5]. The work presented in this thesis uses pp collision data from the 2011, 2012, 2015, and 2016 data-taking periods. The LHC has been designed and built to explore new energy frontiers, to test the standard model of particle physics (SM), and to look for new physics beyond it.

The mechanism to produce, inject, and accelerate protons up to 6.5 TeV proceeds in several steps, as shown in Figure 2.1. Hydrogen atoms are ionised to produce protons, that are accelerated in two steps by the linear accelerator (LINAC2) up to 50 MeV, and by the proton synchrotron booster (PSB or BOOSTER) up to 1.4 GeV. The protons enter the proton synchrotron (PS) which pushes them to 25 GeV, and subsequently

the super proton synchrotron (SPS), where their energy reaches 450 GeV. Finally, two proton beams are injected in opposite directions into the two separate LHC beam pipes, circulating clockwise (Beam-1) and counterclockwise (Beam-2), reaching a centre-of-mass energy of 13 TeV (14 TeV design) [6].

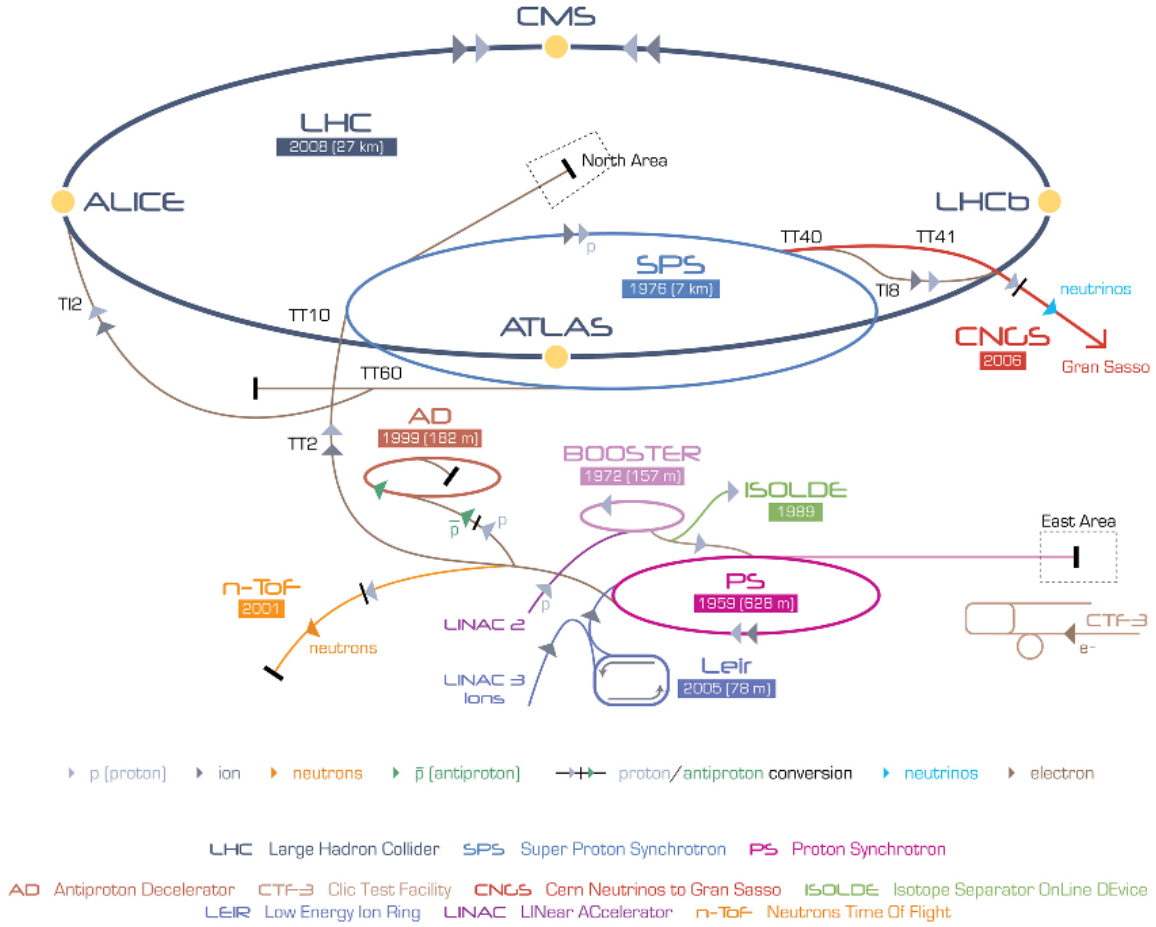


Figure 2.1: The LHC's injection chain with its multiple smaller pre-accelerators [7].

The two adjacent parallel beam pipes are separated by 194 mm, and host about 2556 bunches of protons per ring. Each bunch contains 1.15×10^{11} protons in a beam size of 2.5 μm . The LHC vacuum system is necessary to prevent particles losing energy in the acceleration process due to impacts with air molecules. The two beams interact at four different points at which the main LHC experiments are located: A Large Ion Collider Experiment (ALICE), studying the quark-gluon plasma state of matter through p-Pb or PbPb collisions [8]; A Toroidal LHC ApparatuS (ATLAS), a general-purpose experiment [9]; Compact Muon Solenoid (CMS), a general-purpose experiment [10]; and LHC-beauty (LHCb), the b-quark physics devoted experiment [11].

The superconducting magnets of the LHC are able to produce a strong magnetic field B needed to maintain the two beams in a circular trajectory, obeying the equation:

$$B \text{ [T]} = \frac{p \text{ [TeV]}}{0.3 r \text{ [km]}} , \quad (2.1)$$

where p is the proton momentum and r is the LHC radius, $r \simeq 4.2$ km. The magnets are twin bore coil dipoles hosting the two pipelines, and a temperature of 1.9 K is maintained by a cryogenic system using over 96 tonnes of liquid helium. The LHC ring is equipped with 1232 magnet coils made of copper-clad niobium-titanium cables, each measuring 14.3 m, and 392 quadrupole magnets used to focus the beams approaching the detectors.

The event rate R is:

$$R = \mathcal{L} \cdot \sigma , \quad (2.2)$$

where σ is the production cross-section related to the particular physics process of study, and $\mathcal{L}$ is the instantaneous luminosity expressed by:

$$\mathcal{L} = f \frac{n_1 n_2}{4\pi \sigma_x \sigma_y} , \quad (2.3)$$

where n_X is the number of particles contained in Bunch-X, f is the frequency of the bunch crossing, and σ_x and σ_y are the transverse dimensions of the beam. The design luminosity is $\mathcal{L} \simeq 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ with a bunch crossing interval of 25 ns.

The LHC is interested in studying rare processes, that have very low production cross-sections. Thus, it is necessary to use datasets with very high integrated luminosities. The integrated luminosity is defined as the instantaneous luminosity integrated over time, and it is expressed in fb^{-1} in the following, where $1 \text{ fb}^{-1} = 10^{39} \text{ cm}^{-2}$.

The SM was the focus of study for many years for the LEP collider at CERN and the Tevatron [12] at Fermilab. Remarkable insights into the SM have been provided by both colliders. However, a number of unanswered questions have been left as a legacy to the LHC. The main concern was the missing experimental evidence of a mechanism responsible for the masses of elementary particles. In 2012, the ATLAS and CMS experiments discovered a particle whose properties are consistent with the Higgs boson [13, 14]. This particle results from the Higgs mechanism which provides an explanation of how elementary particles acquire mass. Characterising the Higgs boson elucidates the nature of the electroweak (EW) symmetry breaking. The Higgs boson

thus represents a unique tool to look for new physics, and is discussed in further detail in Chapter 3. Furthermore, the LHC has other unanswered questions to consider, including large uncertainties at high energy scales in the SM, the observed matter-antimatter imbalance in the Universe, and the absence of any particle physics explanation for dark matter (DM).

Figure 2.2 shows the luminosity delivered by the LHC and recorded by CMS from the start of Run-1 (2010) to the end of Run-2 (2018). It is cumulative over all years for pp collisions during stable beams. The LHC delivered an integrated luminosity of 40.99 fb^{-1} in 2016. CMS recorded 37.80 fb^{-1} of which 35.90 fb^{-1} are used for physics analysis since the detector was fully operational at the time of data-taking.

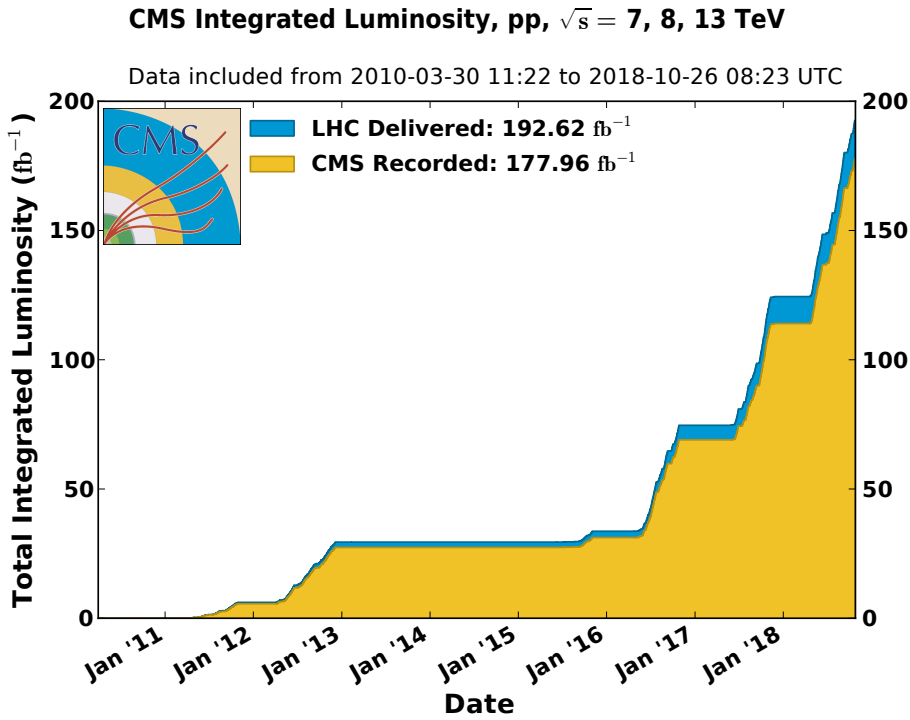


Figure 2.2: The measured luminosity delivered by the LHC to CMS during stable beams in Run-1 and Run-2 for pp collisions [15].

Figure 2.3 details the LHC schedule until 2023, and the expected peak and integrated luminosities. The LHC operates in cycles of 3-year data-taking, followed by a long shutdown (LS) period. This is used to maintain and upgrade the collider. The LSs are also exploited by the experiments to improve and substitute several parts of their sub-detectors in order to cope with the increasing LHC performance. In fact, the hard radiation from the LHC causes detector ageing, jeopardising the precision of the physics measurements.

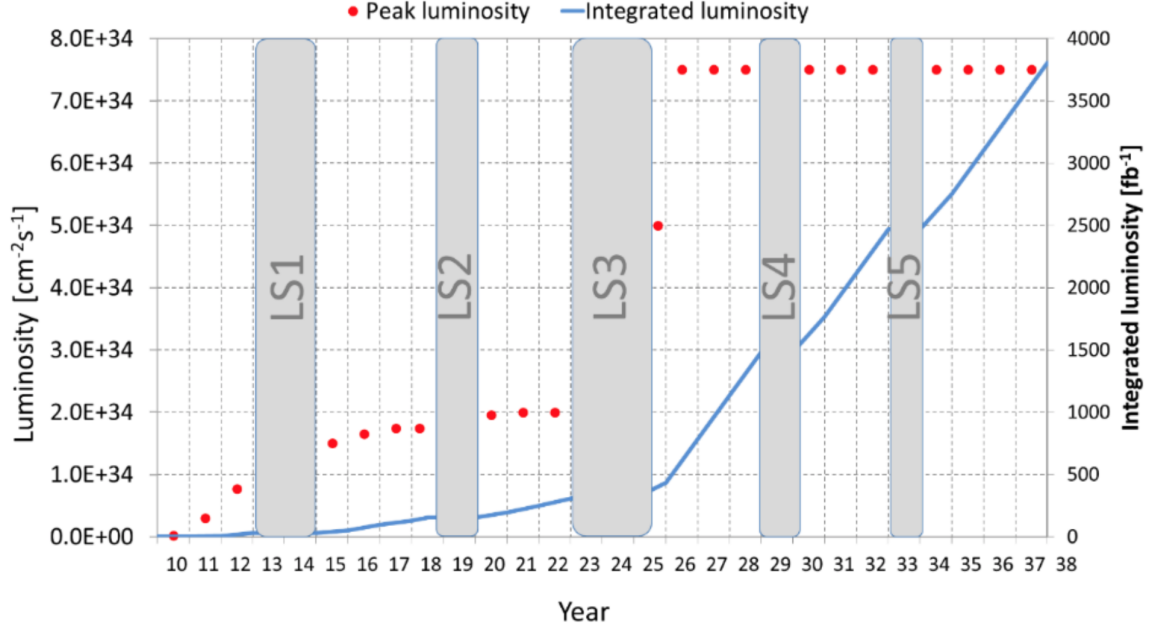


Figure 2.3: The LHC baseline plan from 2010 to 2023, including a forecast of the integrated and peak luminosities [16].

Figure 2.4 provides cross-sections for several processes as a function of $\sqrt{s}$, including that referring to the production of a Higgs boson through the vector boson fusion (VBF) mode. This channel, detailed in Section 3.6, is the main process used in the search for a Higgs boson decaying invisibly, with a cross-section $\sigma_{\text{VBF}} \simeq 1.5 \times 10^3 \text{ fb}$. Considering an integrated luminosity of 35.9 fb^{-1} , about 5.4×10^4 Higgs bosons were produced through the VBF mode in the 2016 run.

2.2 The Compact Muon Solenoid Experiment

The Compact Muon Solenoid (CMS) is one of the two general-purpose experiments at the LHC. The CMS Collaboration is formed by over 4000 people, including particle physicists, computer scientists, technicians, engineers, and students representing around 200 scientific institutes and universities from more than 40 countries.

The exploration and investigation of a wide range of physics at the TeV scale is the main goal of the CMS experiment. At this scale, it is possible to study the EW symmetry breaking and the Higgs mechanism, along with the properties of the Higgs boson, and to precisely measure SM physics phenomena. The experiment also allows a better understanding of quantum chromodynamics (QCD) at extremely high temperature,

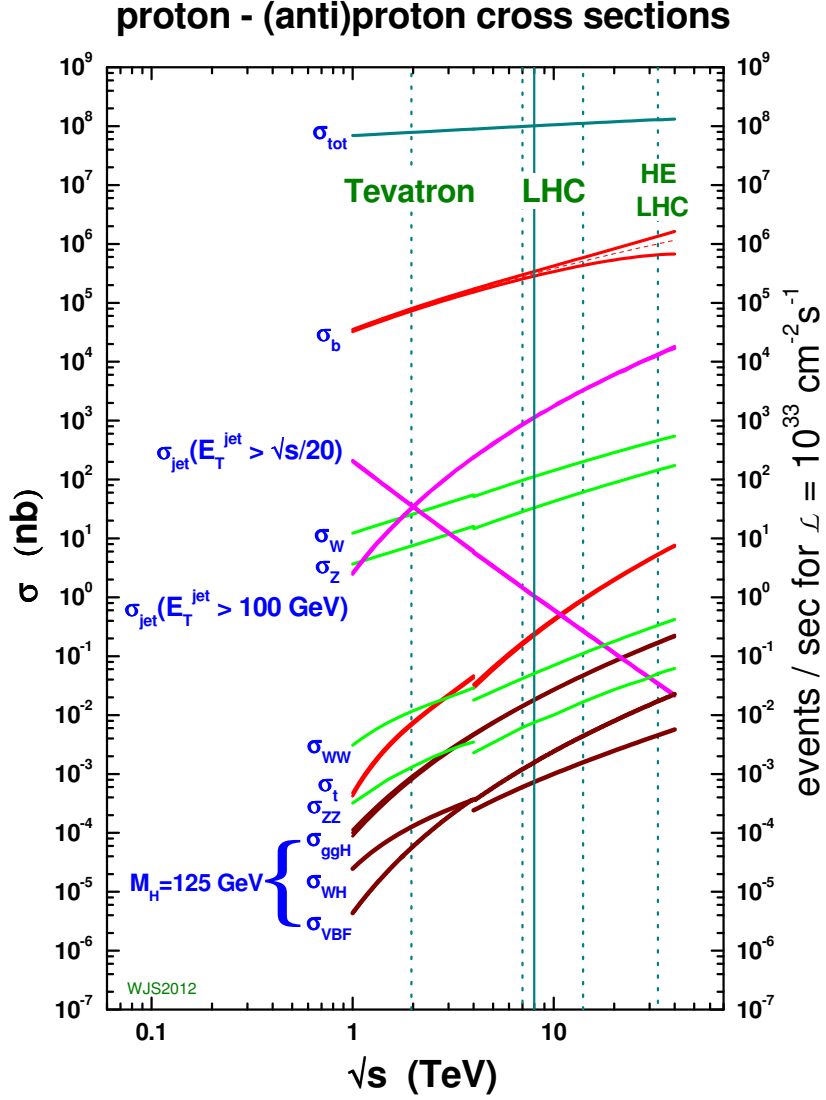


Figure 2.4: The cross-sections expressed in nb for several processes in pp (or $p\bar{p}$) collisions as a function of the centre-of-mass energy $\sqrt{s}$ in TeV [17].

density, or energy. Moreover, searches for beyond the standard model (BSM) physics are carried out, including possible Z'/W' new heavy gauge bosons, DM candidates, or particles theorised by supersymmetry (SUSY).

2.3 The CMS Detector

The aforementioned goals are met by the modular design of the CMS experiment. The CMS detector [18,19] is located in an underground cavern (Point 5), near Cessy in France, and is 14.6 m across, 21.6 m long, and weighs approximately 14 000 tonnes.

A right-handed cartesian reference frame with its origin at the CMS interaction point is used and defined as follows: the x -axis is horizontal and points towards the centre of the LHC ring; the y -axis is vertical and points upwards; and the z -axis is tangential to the beam line. As a result, the x - y plane is orthogonal to the beam pipe whereas the z -axis defines the longitudinal direction. These are used together with cylindrical coordinates since the CMS detector has a cylindrical structure and symmetry. The additional set of coordinates uses $r = \sqrt{x^2 + y^2}$ as the distance from the interaction point in the transverse plane. The azimuthal angle ϕ is measured from the x -axis in the transverse plane and the polar angle θ is measured from the z -axis in the longitudinal (y - z) plane. Physics quantities projected in the transverse plane r - ϕ are frequently used in the following, e.g. the transverse momentum (p_T) and the transverse energy (E_T) of a particle. The p_T is also used to sort objects, referring to leading and sub-leading as the highest and second highest p_T objects in an event, respectively. Another important quantity is the missing transverse momentum vector ($\vec{p}_T^{\text{miss}}$) defined as the negative vector sum of the momentum of all particles in an event projected in the transverse plane. The magnitude of $\vec{p}_T^{\text{miss}}$ is denoted as p_T^{miss} , or, as in this thesis, as the missing transverse energy (MET) E_T^{miss} . Finally, the pseudo-rapidity is defined as:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) . \quad (2.4)$$

The CMS detector faces numerous challenges due to the LHC operational regime. An example is the radiation that causes aging and damage of the sensitive materials of the detector. Moreover, the expected interaction rate is $R = \mathcal{L} \cdot \sigma_{\text{pp}} \simeq 10^9 \text{ Hz}$, considering $\mathcal{L} \simeq 10^{34} \text{ cm}^{-2}\text{s}^{-1}$ and expected pp cross-section $\sigma_{\text{pp}} \simeq 100 \text{ mb}$ at 13 TeV. This high instantaneous luminosity together with large total cross-section leads to a pileup effect, that is the overlapping of many events in the same time interval of data acquisition. These effects are a direct consequence of a bunch crossing interval of 25 ns, or a bunch crossing rate of 40 MHz. The sub-detectors and systems of CMS have been designed and developed to cope with these challenges.

The structure of the detector is characterised by a near-hermetic design and by three main sections: barrel, end-cap, and very-forward region. The first section is the central region and is composed of five ‘wheels’ coaxial to the beam, as shown by the schematic view in Figure 2.5. The barrel is closed at both ends by two structures making up the second section. These end-caps are structured as three discs on each side. The third section consists of sub-detectors close to the beam axis.

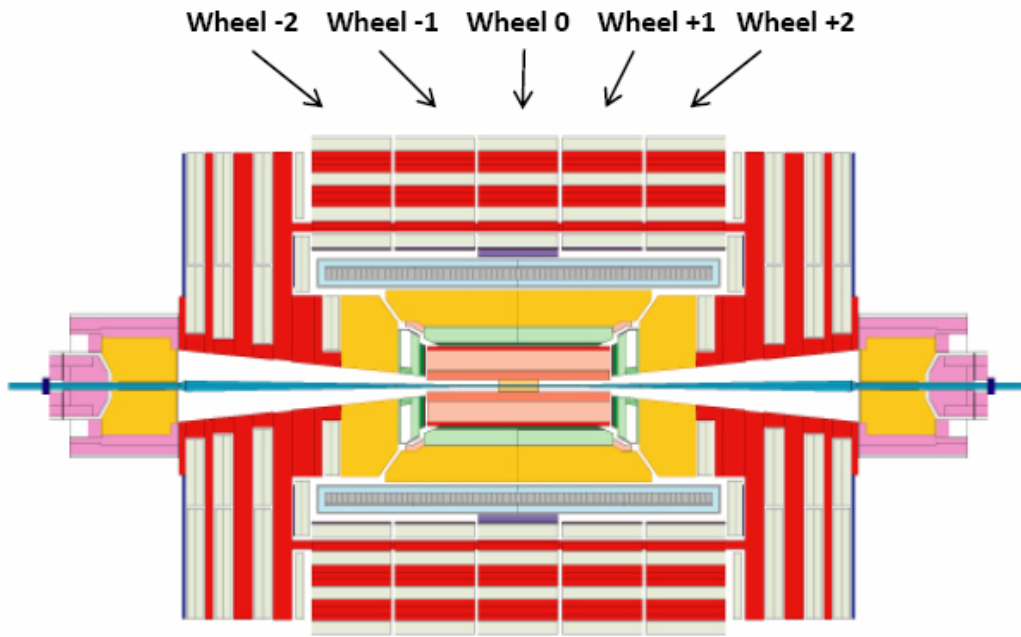


Figure 2.5: A schematic longitudinal view of the five wheels in the barrel region, of the three discs in the end-caps, and of the sub-detectors in the very-forward region of the CMS experiment [20].

The sub-detectors and systems are sensitive to different particles as shown in Figure 2.6. A more detailed picture of the sub-detectors and systems is provided in Sections 2.3.1–6. A more exhaustive description of the CMS detector is provided in Ref. [10].

2.3.1 The Tracker System

The tracking system of CMS [22] is designed to reconstruct the tracks of charged particles and primary and secondary vertices of interaction, and is located around the interaction point. It is composed of high-granularity silicon-pixel detectors in the inner region

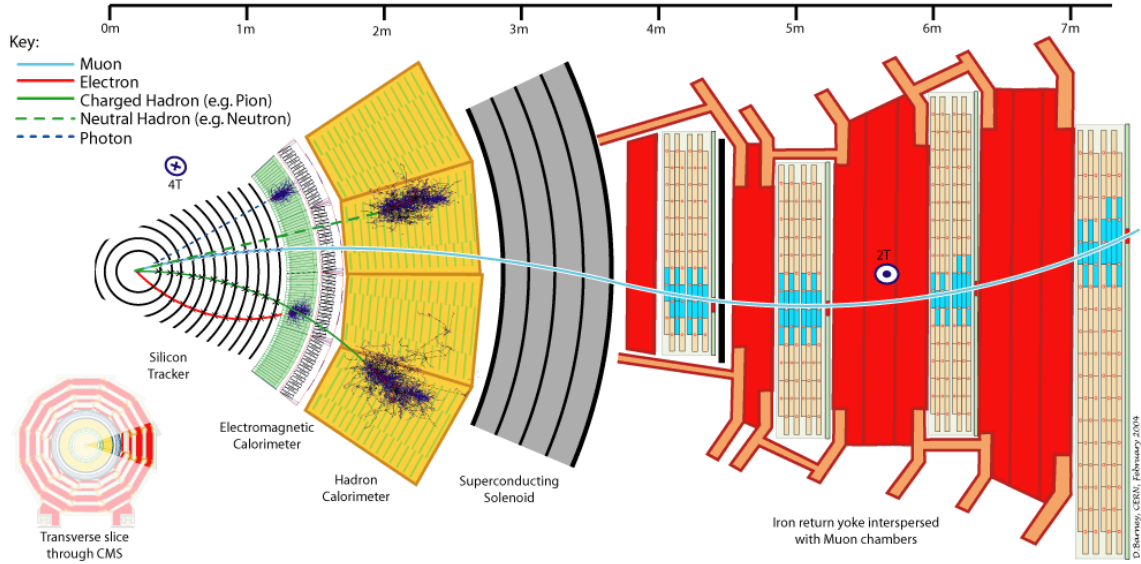


Figure 2.6: An illustrative transverse section of the CMS experiment showing the path of different particles [21].

and of silicon micro-strip detectors in the outer region. They cover a region with a pseudo-rapidity $|\eta| < 2.5$ and with a radius $r < 1.2$ m. The tracker is 5.8 m in length and 2.5 m in diameter and has a total surface of about 200 m^2 . The thickness of the sensors, that can be expressed in radiation lengths X_0 , varies as a function of $|\eta|$: e.g. $0.35 X_0$ at $|\eta| < 0.5$, $1.8 X_0$ in the transition region between barrel and end-cap ($|\eta| \simeq 1.48$), and $1.1 X_0$ at $|\eta| \simeq 2.5$.

This system and the strong magnetic field provided by the CMS magnet (Section 2.3.4) are essential to measure the momentum of the particles in an event. The design p_T resolution of this sub-detector is 1-2% at 100 GeV. The charged particles create electron-hole pairs while passing through the silicon. An electric field is thus applied to separate these pairs causing a current pulse.

The high-granularity of the tracker system leads to a low occupancy per component of the system. This sub-detector is also characterised by a fast response and a large redundancy. There are 66×10^6 pixel cells covering the region close to the beam pipe,

each of $100\text{ }\mu\text{m} \times 150\text{ }\mu\text{m}$ size. They are clustered in about 1400 modules and have a total surface area of about 1 m^2 . There are three layers, each 53 cm long, in the barrel region at radii of 4.4 cm, 7.3 cm, and 10.2 cm, respectively. Both end-cap regions host two discs at $|z| = 34.5\text{ cm}$ and $|z| = 46.5\text{ cm}$ consisting of 24 blades in a turbine-like shape. The spatial resolution is about $10\text{ }\mu\text{m}$ in the transverse $r\text{-}\phi$ plane and $20 - 40\text{ }\mu\text{m}$ along the z -axis.

The outer region hosting the silicon micro-strip detectors is further characterised by an inner and an outer section. Figure 2.7 shows the layout of the CMS tracker in the $r\text{-}z$ plane, highlighting each module detailed in the following. The first section is composed of four layers in the barrel (Tracker Inner Barrel or TIB) with a radius between 20 cm and 55 cm, and three discs in each end-cap (Tracker Inner Disc or TID), both at $|z| < 118\text{ cm}$. The outer section has six layers in the barrel (Tracker Outer Barrel or TOB) with a radius between 55 cm and 120 cm at $|z| < 118\text{ cm}$, and nine discs in each end-cap (Tracker End-Cap or TEC) covering the region $124\text{ cm} < |z| < 282\text{ cm}$. The strips are $80 - 180\text{ }\mu\text{m}$ wide and $10 - 20\text{ cm}$ long, increasing in size as a function of the radius r since the flux of particles decreases. The spatial resolution of these detectors is between $13\text{ }\mu\text{m}$ and $47\text{ }\mu\text{m}$ in the transverse $r\text{-}\phi$ plane and about $500\text{ }\mu\text{m}$ along the z -axis [23]. The $r\text{-}\phi$ resolution is important to precisely measure the p_{T} of a particle since this is the plane in which the track is bent by the magnetic field.

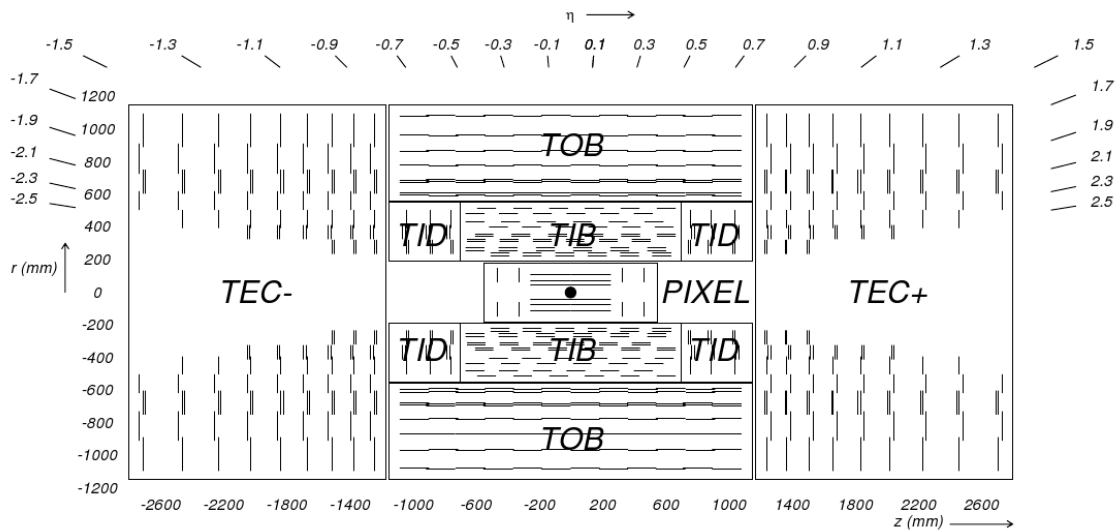


Figure 2.7: A schematic view in the $r\text{-}z$ plane of the CMS tracker where each detector module is pictured as a line-element [23].

2.3.2 The Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) of CMS [24] is designed to identify photons and electrons by measuring the energy deposits of these particles. This sub-detector is homogeneous and hermetic, and is 7.8 m long and 3.5 m wide. It is divided in two regions, namely the barrel and the end-cap. There are more than 75000 lead tungstate (PbWO_4) scintillating crystals designed to detect an electromagnetic (EM) shower produced through pair production and Bremsstrahlung processes from incident photons (γ) and electrons (e^-). Scintillation light is emitted due to the excitation of the crystals, with the number of photons being proportional to the amount of energy deposited in the ECAL. The light is collected by photo-detectors and is used to identify and reconstruct these neutral and charged particles along with the information provided by the tracker system (Section 2.3.1).

PbWO_4 was chosen as the scintillator because of its suitable characteristics, namely high density ($\rho = 8.28 \text{ g/cm}^3$), short radiation length ($X_0 = 0.89 \text{ cm}$), small Molière radius¹ ($R_M = 21.2 \frac{X_0}{\epsilon_c} \text{ MeV} = 2.2 \text{ cm}$, where $\epsilon_c = 8.58 \text{ MeV}$), and a short scintillation time ($\tau = 25 \text{ ns}$). Silicon avalanche photodiodes in the barrel and vacuum phototriodes in the end-caps collect almost 80% of the light from the EM shower. Since the LHC bunch crossing interval is 25 ns, particles of an event are thus properly associated with the correct bunch crossing.

The barrel region of the ECAL covers the pseudo-rapidity range $|\eta| < 1.479$. A thin walled alveolar-shaped structure contains about 61200 crystals at $r = 1.29 \text{ m}$. Each crystal has a front surface of $22 \text{ mm} \times 22 \text{ mm}$, a length of $25.8 X_0 = 230 \text{ mm}$, and a truncated-pyramid geometry. They are mounted with a 3° rotation with respect to the interaction vertex to avoid neutral particle trajectories being aligned with the gaps between crystals. The end-cap region covers the pseudo-rapidity range $1.479 < |\eta| < 3$, where there are more than 7000 crystals each with a front surface of $28.6 \text{ mm} \times 28.6 \text{ mm}$ and a length of $24.7 X_0 = 220 \text{ mm}$, clustered in super-crystals. Pre-shower detectors precede the end-cap ECAL in the range $1.653 < |\eta| < 2.6$ and are able to discriminate between single photons and pairs produced in neutral pion (π^0) decays. They are sampling calorimeters with a total thickness of 20 cm consisting of two layers: lead radiators to initiate the shower, and silicon-strip sensors to measure the transverse profile of the shower and the deposited energy.

¹The Molière radius measures the transverse dimension of a fully contained EM shower. It is a characteristic constant of the material used to contain the shower. An average of 90% and 95% of the energy deposition of a shower is contained in a cylinder of one and two R_M , respectively.

The energy resolution of the ECAL can be parametrised using a stochastic (S), a noise (N), and a constant (C) term, as follows:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2, \quad (2.5)$$

where E is the energy of an incident particle. The fluctuations in the lateral containment of the shower and in the number of photo-electrons produced are considered in the term S . The pileup, electronic and digital noise are accounted for by the term N . The energy leakage from the calorimeter, the non-uniformity of light collection along the crystals, and the calibration of the calorimeter enter in the term C . Test beams with electrons at $20 < p_T < 250$ GeV and without an applied magnetic field were used to calibrate those terms resulting in $S = 2.8\%$, $N = 12\%$, and $C = 0.3\%$ when the energy is expressed in GeV [25].

2.3.3 The Hadron Calorimeter

The hadron calorimeter (HCAL) of CMS [26] is designed to identify hadrons, or strongly interacting particles, and measure their energy deposits. Clustering these deposits allows for the reconstruction of the final state hadronisation of quarks, known as jets. The HCAL is vital to perform this measurement in particular for neutral hadrons since they are ‘invisible’ for the tracker (Section 2.3.1). The near-hermetic construction of the detector offers coverage in the pseudo-rapidity range $|\eta| < 5$.

The HCAL is divided in three regions: barrel, end-cap, and forward. The first covers the pseudo-rapidity range $|\eta| < 1.26$ whereas the second covers $|\eta| < 3$. Both these regions host brass-scintillator sampling calorimeters, hadronic barrel calorimeters (HB) and hadronic end-cap calorimeters (HE), respectively, positioned between the ECAL (Section 2.3.2) and the magnet (Section 2.3.4) from a radius $r = 1.77$ m to $r = 2.95$ m. The brass material has a short nuclear interaction length ($\lambda_0 = 16.42$ cm). This is required to obtain a shower of small dimension. The absorber layers are responsible for initiating the showers that create light signals in the plastic scintillators. Hybrid photodiodes using wavelength-shifting fibres are coupled to these calorimeters to detect these pulses. In addition, the hadronic outer barrel calorimeters (HO) are situated outside the magnet coil. They are used as absorbers to achieve full containment of the shower, improving the energy resolution of the HB by extending the coverage.

The forward region covers the $3 < |\eta| < 5$ range at $|z| = 11.2$ m around the beam pipe. Radiation-hard materials are used to build the hadronic forward calorimeters (HF) in this region due to the proximity to the beam line. The absorber material consists of plates of steel that initiate the shower. The active material uses quartz fibres to produce Cherenkov radiation when a particle above the Cherenkov threshold passes through. This signal is then collected by photomultiplier tubes.

The depth of the HCAL changes as a function of η , being, for example, $5.25 \lambda_0$, $9.1 \lambda_0$, and $10.5 \lambda_0$ at $\eta \simeq 0$, 1.3, and 5, respectively. Test beams of pions, protons, muons, and electrons were used to measure the energy resolution of the HCAL (Equation (2.6)). The setup included an ECAL module and its combination with the HCAL is parametrised as:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + C^2, \quad (2.6)$$

where S and C are about 85% and 7.5%, and 200% and 9% for the barrel and forward regions, respectively [27], and E is the energy of the particle. The energy resolution in the end-cap is comparable with that in the barrel region.

2.3.4 The Magnet

The magnet of CMS [28] is made of a superconducting solenoidal coil and is characterised by a strong axial magnetic field of $B = 3.8$ T. This significant bending power is used to identify and measure the p_T of charged particles in both the tracker (Section 2.3.1) and the iron yoke.

The structure has a weight of about 220 tonnes, a diameter of 6 m, and a length of 12.5 m, and is isolated from the outside through a vacuum cylinder. The magnet consists of niobium-titanium cables wrapped with copper, at a temperature of 4 K to ensure superconductivity. The architecture is completed by a 14 m long external yoke that is responsible for the return flux of the magnetic field. This yoke weighs about 10 000 tonnes and consists of five layers in the barrel region and three discs for each end-cap. It provides near-full coverage by absorbing the particles that pass through the HCAL (Section 2.3.3) with the exception of muons (μ) and neutrinos (ν).

2.3.5 The Muon System

The muon system of CMS [29] is the outer part of the CMS detector. It is primarily considered as a tracking system since muons are not frequently stopped and they do not leave high energy deposits due to their high mass. The overall structure consists of two regions: the barrel and the end-cap. The system covers the $|\eta| < 2.4$ region and is mounted in the return-yoke region of the magnet (Section 2.3.4). It is used, along with the magnetic bending of the return flux ($B = 1.8$ T), to measure the p_T of a muon, and it has a spatial resolution of $250\text{ }\mu\text{m}$ in the transverse r - ϕ plane and $500\text{ }\mu\text{m}$ along the z -axis. Tracker information is used and combined with muon system measurements to improve the momentum resolution. The muon identification, measurement, and reconstruction are described in further details in Section 4.2.5.

This sub-detector entirely consists of gaseous detectors: drift tubes (DTs) in the barrel covering $|\eta| < 1.2$, and cathode strip chambers (CSCs) in the end-cap regions that cover the range $0.9 < |\eta| < 2.4$. A charged particle creates ionisation electrons as it passes through the gas, and these produce electric signals when reaching the anode of the detector. Additional detectors, the resistive plate chambers (RPCs), ensure redundancy by matching with CSCs and DTs hits. They cover the pseudo-rapidity region $|\eta| < 1.6$. They also provide additional bunch crossing identification and trigger information.

The muon barrel hosts four stations of DTs and RPCs, divided into five wheels along the z -axis. The stations have eight layers of DTs each to measure the position of the muon in the transverse plane r - ϕ . There are four additional layers used to measure the z coordinate in the first three stations. The DTs contain an Ar(85%) - CO₂(15%) mixture, that is reliable and with good quenching properties, and they have a size of $4.2\text{ cm} \times 1.3\text{ cm}$ each and are 2.4 m long. The RPCs are made of four bakelite planes forming two gaps. These are filled with a C₂H₂F₄(95.2%) - iC₄H₁₀(4.5%) - SF₆(0.3%) mixture providing higher efficiency for low electric fields with respect to a single-gap chamber. The RPCs are characterised by a fast response time of 3 ns ensuring a time resolution suitable for triggering and for the assignment of muons to a bunch crossing.

The muon end-cap consists of four discs orthogonal to the beam axis where CSCs and RPCs are mounted. The RPCs have the same features as those mounted in the barrel region. The CSCs are generic multi-wire proportional chambers with the cathode plane divided in strips orthogonal to the anode wire. The radiation hardness and fine segmentation of the CSCs enable them to cope with the high particle rate that characterises the end-cap region. The structure is designed to measure the position of

the muon in a two-dimensional plane. These chambers consist of trapezoidal panels mounted on four discs in each end-cap, interspersed with iron return-yoke plates, and use an Ar(40%) - CO₂(50%) - CF₄(10%) mixture.

A p_T resolution of about 10% is reached for muons in the barrel region with $p_T < 200$ GeV using only the muon system information. Depending on η , the resolution ranges between 15% and 40% for 1 TeV muons. However, using the tracker information, the p_T resolution improves to 5% and 10% in the barrel and end-caps, respectively.

2.3.6 The Trigger and Data Acquisition System

The trigger and data-acquisition system (TriDAS) of CMS [30,31] is designed to collect and analyse the data from the detector every bunch crossing interval of 25 ns. During this time interval, the system selects events of potential interest for further physics analysis. In the LHC operational regime, there are about 10^9 interactions per second including pileup events, and about 10^8 channels for the read-out of the event. The size of each event is about 1 MB, which would result in about 40 TB to store per second. This is not feasible, nor is it possible to read-out the detector at this rate. Therefore, the trigger is a fundamental system for the selection of which events to store.

The overall structure consists of a multi-level trigger system: the Level-1 (L1) trigger and the High Level Trigger (HLT). The first is hardware-based with a $4\mu\text{s}$ latency and reduces the event rate from 40 MHz to 100 kHz. The L1 uses information from several sub-detectors to perform a first particle identification, as shown in Figure 2.8 [32].

The calorimeter trigger focuses on electrons, photons, jets, and MET objects, whereas the muon trigger uses muon information from the muon system. These sub-detectors are made of several sub-structures to increase the system performance. The global trigger makes use of calorimeter and muon trigger information along with algorithms designed to provide a final decision on an event. The system is designed such that the data are temporarily stored in pipeline memories for the $4\mu\text{s}$ following the collision.

The HLT is responsible for software event building, event selection, and event reconstruction reducing the event rate from 100 kHz to about 1 kHz. It relies on data filtered by the L1 and it assigns events to specific datasets using related signatures. The result of the selection process, or stream, contains the detector, L1, and HLT information to be used for the forthcoming reconstruction step. This step exploits the Worldwide LHC Computing Grid (WLCG) [33] that is detailed in Chapter 7.

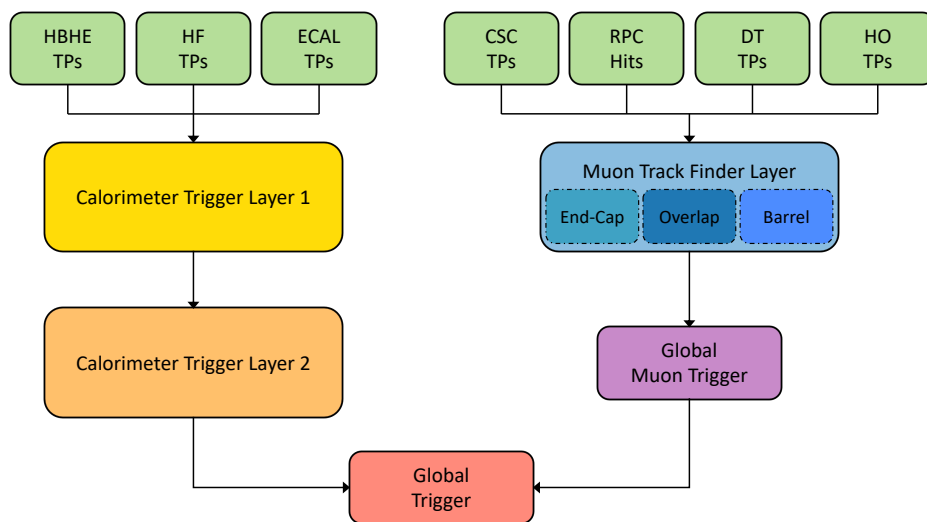


Figure 2.8: The CMS Level-1 trigger for LHC Run-2.

Chapter 3

The Higgs Mechanism and Dark Matter

This Chapter describes the theory that motivates the search for invisible decays of the Higgs boson, which is detailed in Chapters 5 and 6. Section 3.1 gives an overview of the standard model of particle physics and of elementary particles. Section 3.2 focuses on gauge theories and symmetries, introducing the concept of spontaneous symmetry breaking. Section 3.3 generalises the Goldstone model to the Abelian Higgs model, whereas Section 3.4 describes the electroweak theory and the Higgs mechanism, providing an extension to the Yukawa mechanism. Section 3.5 outlines the motivations behind invisibly decaying Higgs boson searches, exploring physics beyond the standard model and the connection with dark matter. Finally, Section 3.6 highlights the characteristics of the Higgs boson production and decay modes.

3.1 The Standard Model of Particle Physics

The standard model of particle physics (SM) describes elementary particles and their interactions through three of the four fundamental forces of nature: the electromagnetic, weak, and strong forces [34–37]. The SM is based on the concept of symmetries, that could be respected or broken by the theory, and is highly predictive. The SM has been extensively tested experimentally, e.g. by CMS, seeking additional elementary particles and constraining free parameters of the model. However, the SM faces challenges to explain some phenomena observed in nature, including the existence of DM.

The SM is a quantum field theory (QFT) characterised by gauge invariances, and consequently by symmetries and fields, as detailed in Section 3.2. These symmetries play a key role since a continuously differentiable symmetry of the Lagrangian describing the theory leads to a conservation law. As an example, the invariance of a system under translations in space, time transformations, and rotations in space correspond to the conservation of linear momentum, mass-energy, and angular momentum, respectively. This one-to-one correspondence between symmetries and conservation laws is manifest in Noether’s theorem [38]. The necessity of additional fields to preserve the invariance of the system characterises specific symmetries [39]. This is treated in further detail in Section 3.2.

QFT describes fields which give rise to the elementary particles in nature: the particles correspond to quantised excitations of these fields. The addition of a new field thus requires a corresponding particle. The features of the particles observed in nature and their representation in the SM are detailed in the following.

3.1.1 Elementary Particles in Nature

There are two types of elementary particles in nature as categorised by their spin: fermions have half-integer spin, and bosons have integer spin. The fermions include leptons, which are subject to the electromagnetic (with the exclusion of neutrinos) and weak forces, and quarks, which additionally interact via the strong force. Quarks exist in doublets, i.e. up and a down types, whereas leptons are categorised as charged or neutral. These particles are arranged in three generations, each including one particle per type. Table 3.1 summarises this structure and gives the mass and electric charge for each particle.

The bosons include the vector mediators of the three fundamental forces described by the SM: the photon (γ), gluon (g), $W^\pm$, and Z^0 bosons. Table 3.2 summarises these particles and gives the mass and electric charge for each boson.

The SM requires the Higgs mechanism [41–46], detailed in Section 3.4, to explain the masses of these particles. From this mechanism, a scalar boson naturally arises: the Higgs boson. In the context of this thesis, it can be exploited to search for DM, or generic BSM physics, as discussed in Section 3.5.1. A picture of gauge theories is necessary to fully understand these particles and their interactions, and how they fit in the SM.

Table 3.1: The elementary fermions observed in nature categorised into their three generations. Each of these particles has an antiparticle characterised by an opposite charge but identical mass. The magnitude of the electron charge is used as a unit for the electric charge of all particles [40].

Generation	Quarks			Leptons		
	Particle	Mass	Electric Charge	Particle	Mass	Electric Charge
I	u	2.2 MeV	$+\frac{2}{3}$	e^-	0.511 MeV	-1
	d	4.7 MeV	$-\frac{1}{3}$	ν_e	~ 0	0
II	c	1.275 GeV	$+\frac{2}{3}$	μ^-	105.658 MeV	-1
	s	95 MeV	$-\frac{1}{3}$	ν_μ	~ 0	0
III	t	173.0 GeV	$+\frac{2}{3}$	τ^-	1776.8 MeV	-1
	b	4.18 GeV	$-\frac{1}{3}$	ν_τ	~ 0	0

Table 3.2: The elementary bosons observed in nature, mediators of the related force. The magnitude of the electron charge is used as a unit for the electric charge of all particles [40].

Force	Particle	Mass	Electric Charge
Electromagnetic	γ	0	0
Weak	$W^\pm$	80.385 GeV	± 1
	Z^0	91.188 GeV	0
Strong	g	0	0

3.2 Gauge Theory, Symmetries, and the Goldstone Model

A symmetry of a physical system can be defined as a physical or mathematical feature of the system which is preserved after a generic transformation. Symmetries are classified as local if they have different symmetry transformations at different points in space-time, or as global if they hold at all points in space-time. The former play a key role in physics as they represent the basis for gauge theories. Symmetry principles or, more precisely, local gauge symmetries govern particle interactions.

From the Lagrangian $\mathcal{L} = \mathcal{L}(q_i, \dot{q}_i)$ of a physical system depending only on generalised coordinates q_i and their first derivatives $\dot{q}_i$, the equations of motion can be obtained by

substituting it into the Euler-Lagrange's equations (Equation (3.1)) as follows:

$$\frac{\partial \mathcal{L}}{\partial q_i} = \frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i} \right) . \quad (3.1)$$

The unitary Abelian $U(1)$ group consists of all phase transformations $U(\alpha) = e^{i\alpha}$ where the parameter α can assume any continuous real value. Considering a generic phase transformation of a field $\psi(x) \rightarrow \psi'(x) = e^{i\alpha}\psi(x)$, then the Dirac Lagrangian (Equation (3.2)) remains invariant under it [47]. The Dirac Lagrangian for a fermion ψ with mass m is defined as:

$$\mathcal{L} = \bar{\psi}(x)(i\gamma^\mu \partial_\mu - m)\psi(x) , \quad (3.2)$$

where γ_μ are the Dirac matrices:

$$\gamma^0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} , \quad \gamma^i = \begin{pmatrix} 0 & \sigma^i \\ -\sigma^i & 0 \end{pmatrix} \quad (3.3)$$

that use the Pauli matrices σ^i , with index $i = 1, 2, 3$ for both γ^i and σ^i :

$$\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \quad \sigma^2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} , \quad \sigma^3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} . \quad (3.4)$$

The exact symmetries are characterised by an invariance of the Lagrangian density ($\delta\mathcal{L} = 0$) and of the unique physical vacuum state under the exact symmetry transformation. Moreover, the $U(1)$ invariance of Equation (3.2) implies the conservation of a current density, $\partial_\mu J^\mu = 0$, where J^μ corresponds to the EM charge current density.

The existence of a symmetry leads to the impossibility of measuring a specific quantity, that is the phase α for $U(1)$. Thus, it is possible to arbitrarily choose and fix this quantity, that will have the same value for all points in space-time, leading to a global gauge invariance. The term ‘gauge’ refers to a degree of freedom within a theory whose external effects are not observable. The transformations between possible gauges of this degree of freedom do not change any observable physical property.

A local phase transformation of a field $\psi(x)$ is now considered:

$$\psi(x) \rightarrow \psi'(x) = e^{i\alpha(x)}\psi(x) , \quad (3.5)$$

where $\alpha(x)$ is a function of the space-time coordinates. The Dirac Lagrangian (Equation (3.2)) is not invariant under Equation (3.5), since it produces an additional term: $-(\partial_\mu \alpha(x))\bar{\psi}(x)\gamma^\mu\psi(x)$. A gauge-covariant derivative is introduced to restore the invariance of Equation (3.2):

$$\mathcal{D}_\mu = \partial_\mu + iqA_\mu , \quad (3.6)$$

where q is the charge of the particle in natural units described by the field $\psi(x)$. The gauge field $A_\mu(x)$ transforms according to the local phase transformation:

$$A_\mu(x) \rightarrow A'_\mu(x) = A_\mu(x) - \frac{1}{q}\partial_\mu\alpha(x) , \quad (3.7)$$

producing an interaction term with the fermion field in Equation (3.2): $-q(\bar{\psi}(x)\gamma^\mu\psi(x))A_\mu$. The invariance of the Dirac Lagrangian is thus restored under the local gauge transformation (Equation (3.5)), leading to a conserved current and consequently to the conservation of the charge. As a result, quantum electrodynamics (QED) is invariant under a gauge transformation of the $U(1)$ group.

Following a similar treatment to the above, other symmetry groups, such as $SU(2)$ for weak interactions and $SU(3)$ for strong interactions, can be considered. However, not all symmetries are exact, and approximate symmetries can lead to the invariance of the Lagrangian but not of the physical vacuum [47]. For this specific case, there is spontaneous symmetry breaking (SSB) of the Lagrangian, detailed in the following.

3.2.1 Spontaneous Symmetry Breaking (SSB)

A Lagrangian for two scalar fields ϕ_1 and ϕ_2 :

$$\mathcal{L} = \frac{1}{2}[(\partial_\mu\phi_1)(\partial^\mu\phi_1) + (\partial_\mu\phi_2)(\partial^\mu\phi_2)] - V(\phi_1^2 + \phi_2^2) , \quad (3.8)$$

where V is the effective potential, is considered to introduce the mechanism of SSB. This Lagrangian is invariant under $SO(2)$ rotations. Considering the effective potential $V(\phi^2) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}|\lambda|(\phi^2)^2$, where $\phi^2 = \phi_1^2 + \phi_2^2$ and μ is the reduced mass, there are two different cases depending on the value of μ^2 . For $\mu^2 > 0$, the potential has a unique minimum corresponding to the vacuum state $\langle\phi\rangle_0^{\text{T}1} = (0 \ 0)$. For $\mu^2 < 0$, the absolute minimum of the potential is $\langle|\phi|^2\rangle_0 = -\frac{\mu^2}{|\lambda|} = v^2$, where λ is the coupling and v the

¹T refers to the transpose of a matrix.

vacuum state. This minimum is a continuum of distinct vacuum states, degenerate in energy, due to the $SO(2)$ symmetry of the effective potential.

SSB occurs through the choice of one of these states, for example $\langle\phi\rangle_0^T = (v \ 0)$. A new field ϕ' can be defined as:

$$\phi' = \phi - \langle\phi\rangle_0 = \begin{pmatrix} \eta \\ \zeta \end{pmatrix} \quad (3.9)$$

and the Lagrangian for small oscillations is thus:

$$\mathcal{L}_{s.o.} = \frac{1}{2}[(\partial_\mu\eta)(\partial^\mu\eta) + 2\mu^2\eta^2] + \frac{1}{2}[(\partial_\mu\zeta)(\partial^\mu\zeta)] + \dots \quad (3.10)$$

A particle η is associated with radial oscillations, and acquires a mass $m^2 = -2\mu^2 > 0$ because of the restoring force of the potential against radial oscillations. A massless particle ζ results from the $SO(2)$ invariance of the Lagrangian for two scalar fields ϕ_1 and ϕ_2 . This is known as the ‘Goldstone phenomenon’ [47]: it involves the splitting of the spectrum and the appearance of a massless particle. There is no consumption of potential energy in the movement along the circular bottom of the potential, so that the energy of this particle is pure kinetic energy.

3.3 Spontaneous Symmetry Breaking in Gauge Theories

A local gauge-invariant extension can be applied to the Goldstone model detailed in Section 3.2, obtaining the Abelian Higgs model. It is a $U(1)$ -invariant theory describing the electrodynamics of a charged scalar in the absence of SSB. The Lagrangian is:

$$\mathcal{L} = |\mathcal{D}^\mu\phi|^2 - \mu^2|\phi|^2 - |\lambda|(\phi^*\phi)^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (3.11)$$

where $\phi = \frac{\phi_1 + i\phi_2}{\sqrt{2}}$ is a complex scalar field, $\mathcal{D}^\mu$ is the gauge-invariant derivative, and $F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu$ is the kinetic term related to A_μ . It is invariant under both $U(1)$ rotations and local gauge transformations. For $\mu^2 > 0$, the potential has a unique minimum corresponding to $\phi = 0$ and the Lagrangian remains symmetric. The outcome consists of a single massless photon represented by the A_μ field, and of two scalar particles, ϕ and ϕ^* , both with mass μ . For $\mu^2 = -|\mu^2| < 0$, the symmetry is spontaneously broken.

The potential $V(\phi) = \mu^2|\phi|^2 + |\lambda|(\phi^*\phi)^2$ has the form of the so-called ‘Mexican Hat’, as shown in Figure 3.1, and this case is detailed in the following.

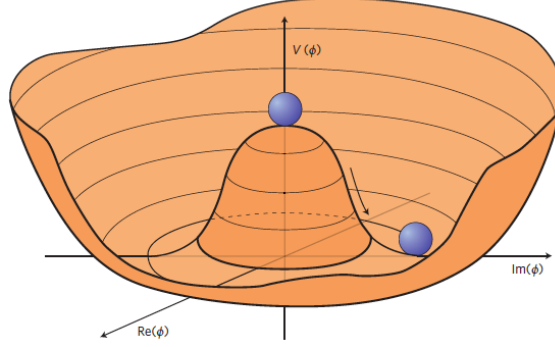


Figure 3.1: The illustration of the potential $V(\phi)$ for $\mu^2 < 0$. SSB of $U(1)$ symmetry occurs when a particular minimum is chosen [48].

The absolute minimum of the potential corresponds to a continuum of distinct vacuum states degenerate in energy $\langle |\phi|^2 \rangle_0 = -\frac{\mu^2}{2|\lambda|} = \frac{v^2}{2}$, where $\langle \phi \rangle_0 = \frac{v}{\sqrt{2}}$, with $v > 0$ a real number, is chosen as the vacuum state. Considering the translation $\phi' = \phi - \langle \phi \rangle_0$, where the field is parametrised as:

$$\phi = \frac{e^{i\frac{\zeta}{v}}(v + \eta)}{\sqrt{2}} \approx \frac{v + \eta + i\zeta}{\sqrt{2}}, \quad (3.12)$$

the Lagrangian for small oscillations (modulo additional constants) obtained is:

$$\mathcal{L}_{s.o.} = \frac{1}{2}[(\partial_\mu \eta)(\partial^\mu \eta) + 2\mu^2 \eta^2] + \frac{1}{2}[(\partial_\mu \zeta)(\partial^\mu \zeta)] - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + qvA_\mu(\partial^\mu \zeta) + \frac{q^2 v^2}{2}A_\mu A^\mu. \quad (3.13)$$

The η field is related to radial oscillations and has a mass $m^2 = -2\mu^2 > 0$. The A_μ field acquires mass, as shown by the term $qvA_\mu(\partial^\mu \zeta)$. The terms involving A_μ and ζ can be rewritten as:

$$\frac{q^2 v^2}{2} \left(A_\mu + \frac{1}{qv} \partial_\mu \zeta \right) \left(A^\mu + \frac{1}{qv} \partial^\mu \zeta \right), \quad (3.14)$$

and the gauge transformation:

$$A_\mu \rightarrow A'_\mu = A_\mu + \frac{1}{qv} \partial_\mu \zeta \quad (3.15)$$

can be considered. This corresponds to a phase rotation $\phi \rightarrow \phi' = e^{-i\frac{\zeta(x)}{v}}\phi(x) = \frac{v+\eta}{\sqrt{2}}$. In this gauge, Equation (3.13) can be written as:

$$\mathcal{L}_{s.o.} = \frac{1}{2}[(\partial_\mu \eta)(\partial^\mu \eta) + 2\mu^2 \eta^2] - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \frac{q^2 v^2}{2}A'_\mu A'^\mu + \dots \quad (3.16)$$

The outcome consists of: the η field with a mass $m^2 = -2\mu^2 > 0$; the A'_μ massive vector field with a mass $m = qv$; the absence of the ζ field. The latter disappears due to the choice of gauge, becoming the longitudinal component of A'_μ .

Before SSB, the theory has four degrees of freedom: two scalars ϕ and ϕ^* , and two helicity states of the massless gauge field A_μ . These degrees of freedom are conserved after SSB: one scalar particle η , and three helicity states of the massive gauge field A'_μ . This is called the Higgs mechanism [41–46]. The Lagrangian thus describes two interacting massive particles, a vector gauge boson, and a massive scalar, known as the Higgs boson.

The SSB of a non-Abelian gauge theory, such as $SU(2)$ with a scalar field $\phi^T = (\phi_1 \ \phi_2 \ \phi_3)$, follows the same principle as the Abelian Higgs model. It is therefore possible to build up spontaneously broken gauge theories in which the interactions are not mediated by massless vector bosons of the unbroken theory, but by massive vector bosons.

3.4 Spontaneous Symmetry Breaking of the Electroweak Theory and the Higgs Mechanism

The procedure highlighted in Section 3.3 can be generalised to describe electromagnetic and weak interactions in a unique model, namely $SU(2)_L \times U(1)_Y$. $SU(2)_L$ couples to left-handed fermions, whereas $U(1)_Y$ to weak hypercharge $Y = 2(Q - T_3)$, where Q is the electric charge and T_3 is the third component of the weak isospin.

The Dirac Lagrangian for a free fermion (Equation (3.2)) is invariant under $SU(N)$ rotations, $\psi \rightarrow \psi' = U\psi$, where U is a unitary matrix. Local gauge invariance is achieved by replacing the derivative ∂_μ with the covariant derivative of the $SU(2)_L \times U(1)_Y$ group:

$$\mathcal{D}_\mu = \partial_\mu + i\frac{g'}{2}B_\mu \mathbb{1} + i\frac{g}{2}W_\mu^a \sigma_a, \quad (3.17)$$

where σ_a are the non-commutative Pauli matrices, and B_μ is the massless mediator of the field of the $U(1)$ group (coupling constant g'), transforming as $B_\mu \rightarrow B'_\mu = B_\mu - \frac{1}{e}\partial_\mu\alpha(x)$. The W_μ^a are the three massless quanta of $SU(2)$ (coupling constant g), transforming as $W_\mu^a \rightarrow W'^a_\mu = W_\mu^a - \frac{1}{g}\partial_\mu\varepsilon^a(x) - \varepsilon_{abc}\varepsilon^b(x)W_\mu^c$, where ε^a are infinitesimal arbitrary parameters, and ε_{abc} is the total antisymmetric tensor. The Dirac Lagrangian thus transforms as:

$$\mathcal{L}'(\psi') = i\bar{\psi}U^\dagger\gamma^\mu U\mathcal{D}_\mu\psi - m\bar{\psi}U^\dagger U\psi = i\bar{\psi}\gamma^\mu\mathcal{D}_\mu\psi - m\bar{\psi}\psi = \mathcal{L}(\psi) \quad (3.18)$$

and it describes the interaction between matter particles through the exchange of gauge bosons related to the gauge fields. An additional term, referred to as the Yang-Mills Lagrangian density, accounts for the propagation of gauge fields:

$$\mathcal{L}_{YM} = -\frac{1}{4}(W^a)_{\mu\nu}(W^a)^{\mu\nu} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (3.19)$$

where $W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu - g\varepsilon_{abc}W^b_\mu W^c_\nu$ and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. The Lagrangian density does not contain any quadratic terms for the gauge fields. Therefore, the gauge bosons associated with these gauge fields are massless.

The real fields γ , Z^0 , and $W^\pm$ emerge from these massless fields through the SSB. The Higgs mechanism allows the mediators of the weak interaction to become massive and the mediator of the electromagnetic interaction to remain massless. The scalar Higgs boson and the corresponding Higgs field necessarily emerge from this mechanism. The scalar Higgs field is located at each point in space-time so that bosons and fermions become massive through their interaction with this field. The Higgs Lagrangian $\mathcal{L}_H$ is gauge invariant and corresponds to a self-interacting scalar isodoublet φ :

$$\mathcal{L}_H = \mathcal{L}_D - \mathcal{L}_V + \mathcal{L}_{YM} = (\mathcal{D}_\mu\varphi)^\dagger(\mathcal{D}^\mu\varphi) - V(\varphi^\dagger\varphi) - \frac{1}{4}(W^a)_{\mu\nu}(W^a)^{\mu\nu} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (3.20)$$

where $\mathcal{L}_D$ is the Dirac Lagrangian, $\mathcal{L}_V$ refers to the Higgs potential, and $\mathcal{L}_{YM}$ is the Yang-Mills Lagrangian density term. The Higgs potential can be written as:

$$V(\varphi^\dagger\varphi) = \mu^2\varphi^\dagger\varphi + \lambda(\varphi^\dagger\varphi)^2, \quad (3.21)$$

where μ^2 and λ are generic constants. For $\mu^2 > 0$, the potential has a parabolic shape. For $\mu^2 < 0$, the potential has the form of the ‘Mexican Hat’, and this case is detailed in the following.

The potential does not have a unique minimum, and the ground state $\varphi = 0$ corresponds to a local maximum of the potential. The system is invariant under global rotations but it is no longer invariant under local transformations. Considering φ as a complex doublet with a given hypercharge $Y_W = 1$, as:

$$\varphi = \begin{pmatrix} \varphi^a = \frac{1}{\sqrt{2}}(\varphi_1 + \varphi_2) \\ \varphi^b = \frac{1}{\sqrt{2}}(\varphi_3 + \varphi_4) \end{pmatrix}, \quad (3.22)$$

a generic choice of $\varphi_1 = \varphi_2 = \varphi_4 = 0$ and $\varphi_3 = v$ is made. This leads to $\varphi^T = \frac{1}{\sqrt{2}}(0 \ v)$, where $v = \sqrt{\frac{-\mu^2}{\lambda}}$ is the vacuum expectation value of the Higgs field. The $SU(2)$ symmetry is spontaneously broken as a particular minimum has been chosen. The φ field can be expanded about the vacuum state:

$$\varphi = e^{i\xi(x) \cdot \sigma} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}, \quad (3.23)$$

where $\xi(x)$ are real fields related to excitations along the minimum of the potential. A rotation:

$$\varphi' = e^{-i\xi(x) \cdot \sigma} \varphi(x) = \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (3.24)$$

is able to remove $\xi(x)$. The real field $h(x)$ does not vanish after the gauge transformation, and it is interpreted as the Higgs boson, a real particle.

Considering the expression of the scalar field φ' , taking into account only the second order terms of the fields B_μ , W_μ^a and h , and overlooking higher order terms, the Lagrangians are:

$$\begin{aligned} \mathcal{L}_D = (\mathcal{D}_\mu \varphi')^\dagger (\mathcal{D}^\mu \varphi') &= \frac{1}{2}(\partial_\mu h)(\partial^\mu h) + \frac{1}{2} \left(\frac{g^2 v^2}{4} \right) [(W^1)_\mu (W^1)^\mu + (W^2)_\mu (W^2)^\mu] \\ &+ \frac{1}{8} v^2 [g(W^3)_\mu - g' B_\mu] [g(W^3)^\mu - g' B^\mu], \end{aligned} \quad (3.25)$$

$$\mathcal{L}_V = V(\varphi'^\dagger \varphi') = \frac{1}{2}(-2\mu^2)h^2 + \text{const.}, \quad (3.26)$$

$$\mathcal{L}_{YM} = -\frac{1}{4}(W^a)_{\mu\nu}(W^a)^{\mu\nu} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} , \quad (3.27)$$

where $W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a - g\varepsilon_{abc}W_\mu^b W_\nu^c = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + O(W^2)$. The fields B_μ and W_μ^3 arise in mixed products in $\mathcal{L}_D$, so that the related bosons cannot have a physical mass. Orthogonal combinations of them can be defined as:

$$\mathcal{Z}_\mu = \cos\theta_W W_\mu^3 - \sin\theta_W B_\mu , \quad (3.28)$$

$$\mathcal{A}_\mu = \sin\theta_W W_\mu^3 + \cos\theta_W B_\mu , \quad (3.29)$$

where θ_W is the Weinberg angle, and $\tan\theta_W = \frac{g'}{g}$. The mixing angle θ_W is chosen in order to make the mixed products of $\mathcal{Z}_\mu$ and $\mathcal{A}_\mu$ disappear.

The Higgs Lagrangian is thus:

$$\begin{aligned} \mathcal{L}_H = & \frac{1}{2}(\partial_\mu h)(\partial^\mu h) - \frac{1}{2}(-2\mu^2)h^2 \\ & - \frac{1}{4}(W^1)_{\mu\nu}(W^1)^{\mu\nu} + \frac{1}{2}\left(\frac{g^2 v^2}{4}\right)(W^1)_\mu(W^1)^\mu \\ & - \frac{1}{4}(W^2)_{\mu\nu}(W^2)^{\mu\nu} + \frac{1}{2}\left(\frac{g^2 v^2}{4}\right)(W^2)_\mu(W^2)^\mu \\ & - \frac{1}{4}Z_{\mu\nu}Z^{\mu\nu} + \frac{1}{2}\left(\frac{g^2 v^2}{4\cos^2\theta_W}\right)\mathcal{Z}_\mu\mathcal{Z}^\mu \\ & - \frac{1}{4}A_{\mu\nu}A^{\mu\nu} \\ & + \mathcal{L}_{mixed} . \end{aligned} \quad (3.30)$$

$\mathcal{L}_{mixed}$ contains mixed terms, more explicitly:

$$\begin{aligned} \mathcal{L}_{mixed} = & \frac{g^2 v}{4} h (W^1)_\mu(W^1)^\mu + \frac{g^2 v}{4} h (W^2)_\mu(W^2)^\mu + \frac{g^2 v}{4\cos^2\theta_W} h \mathcal{Z}_\mu\mathcal{Z}^\mu \\ = & \frac{g^2 v}{4} h (W^+)_\mu(W^+)^\mu + \frac{g^2 v}{4} h (W^-)_\mu(W^-)^\mu + \frac{g^2 v}{4\cos^2\theta_W} h \mathcal{Z}_\mu\mathcal{Z}^\mu . \end{aligned} \quad (3.31)$$

The mass term for the gauge field $\mathcal{A}_\mu$ is not contained in $\mathcal{L}_H$ since the photon is a massless boson. Therefore, the Higgs boson does not directly couple with $\mathcal{A}_\mu$, whose kinetic term $A_{\mu\nu}A^{\mu\nu}$ in Equation (3.30) corresponds to the propagation of the photon.

The mass terms in the Higgs Lagrangian are as follows: $m_W^2 = \frac{g^2 v^2}{4}$, $m_Z^2 = \frac{(g^2 + g'^2)v^2}{4} = \frac{g^2 v^2}{4 \cos^2 \theta_W} = \frac{m_W^2}{\cos^2 \theta_W}$, $m_\gamma^2 = 0$, and $m_H = \sqrt{-2\mu^2} = \sqrt{2\lambda} v$. Since λ is an external parameter, the mass of the Higgs boson is not predicted by the theory.

The complex fields $W_\mu^\pm$ are defined as:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) , \quad (3.32)$$

deriving from the last term of the gauge invariant derivative

$$W_\mu^a \sigma_a = \begin{pmatrix} W_\mu^3 & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & -W_\mu^3 \end{pmatrix} = \begin{pmatrix} W_\mu^3 & \sqrt{2} W_\mu^+ \\ \sqrt{2} W_\mu^- & -W_\mu^3 \end{pmatrix} . \quad (3.33)$$

Considering $\mathcal{L}_{mixed}$ with higher order terms, it is found that the Higgs coupling to the massive gauge bosons $W^\pm$ and Z^0 is proportional to their mass squared: $g_{WWH} = \frac{g^2 v}{2} = \frac{2m_W^2}{v}$ and $g_{ZZH} = \frac{g^2 v}{4 \cos^2 \theta_W} = \frac{m_Z^2}{v}$. From these mass terms, the relation between the ratios of the masses and the couplings of the gauge bosons can be determined as follows:

$$\frac{m_W^2}{m_Z^2} = \frac{g^2}{g^2 + g'^2} = \cos^2 \theta_W , \quad (3.34)$$

$$1 - \frac{m_W^2}{m_Z^2} = \frac{g'^2}{g^2 + g'^2} = \sin^2 \theta_W . \quad (3.35)$$

The VBF production mode of the Higgs boson is characterised by the coupling of the vector bosons, either Z or W , to the Higgs boson, as detailed in Section 3.6. The cross-section of the VBF process must be precisely predicted to interpret the results of the VBF Higgs to invisible analysis in terms of an upper limit on the branching ratio of the Higgs boson decaying to invisible particles, as further discussed in Chapter 5. Therefore, knowledge of the relation between the masses of the vector bosons and Higgs boson couplings is vital in the context of this thesis. This relation is also useful to find the numerical value of θ_W , through different experiments, such as precision measurements of the Z^0 boson or the ratio between the masses of the $W^\pm$ and Z^0 bosons (Table 3.2).

The ATLAS and CMS Collaborations discovered a scalar particle with a mass of approximately 125 GeV in 2012 [13, 14]. This particle is consistent with the SM Higgs boson prediction, within the experimental uncertainty.

3.4.1 The Yukawa Mechanism

The Higgs field is also responsible for the mass of the fermions through an extension of the Higgs mechanism to the Yukawa interaction. In fact, when the Higgs field acquires a non-zero vacuum expectation value, SSB of the chiral symmetry occurs. In this scenario, the Lagrangian contains an interaction term:

$$\mathcal{L}_{\text{Yukawa}} = -\lambda_f \bar{\psi} H \psi , \quad (3.36)$$

where λ_f is the coupling between the Higgs field and the fermion, ψ is the Dirac field, and H is the Higgs field. The mass of a fermion is $\frac{\lambda_f v}{\sqrt{2}}$, proportional to its Yukawa coupling, leading the top quark to have the strongest couplings.

However, the ‘Yukawa mechanism’ is less predictive than the Higgs mechanism of the electroweak interaction. For the latter, the parameters of the theory have a clear theoretical interpretation, whereas the parameters of the Yukawa interaction are introduced ad-hoc in the SM. In fact, they are determined from the masses of the fermions observed by experiments.

The introduction of a possible fermion DM candidate within BSM theories that couples directly to the Higgs boson would require an interaction term similar to $\mathcal{L}_{\text{Yukawa}}$, as presented in Section 6.3.2 and contextualised in the following Sections.

3.5 Dark Matter and Invisibly Decaying Higgs Bosons

Cosmological observations suggest that the Universe consists of visible matter only at 5%. The energy content of the Universe is 95% in the form of dark energy (about 70%) and dark matter (about 25%) [49].

The SM is a highly successful theory that can explain many phenomena in nature. However, it is not able to predict the non-zero masses of neutrinos, necessary to explain neutrino oscillation [50], and violation of the charge-parity symmetry, explaining the asymmetry of matter and anti-matter observed in the Universe [51]. Another important weakness of the SM is the lack of a DM candidate.

Numerous studies have shown evidence for DM, two of which are presented as representative examples. The study of the rotational velocities of galaxies as a function of the distance from the galactic centre led to an evidence for DM. Figure 3.2 shows

the rotational velocity of the galaxy NGC 6503 as a function of the distance from the galactic centre. The prediction for these velocities contradicts the observation without the addition of non-luminous matter. The disc and gas contribution necessitates a dark matter halo to match the observed data.

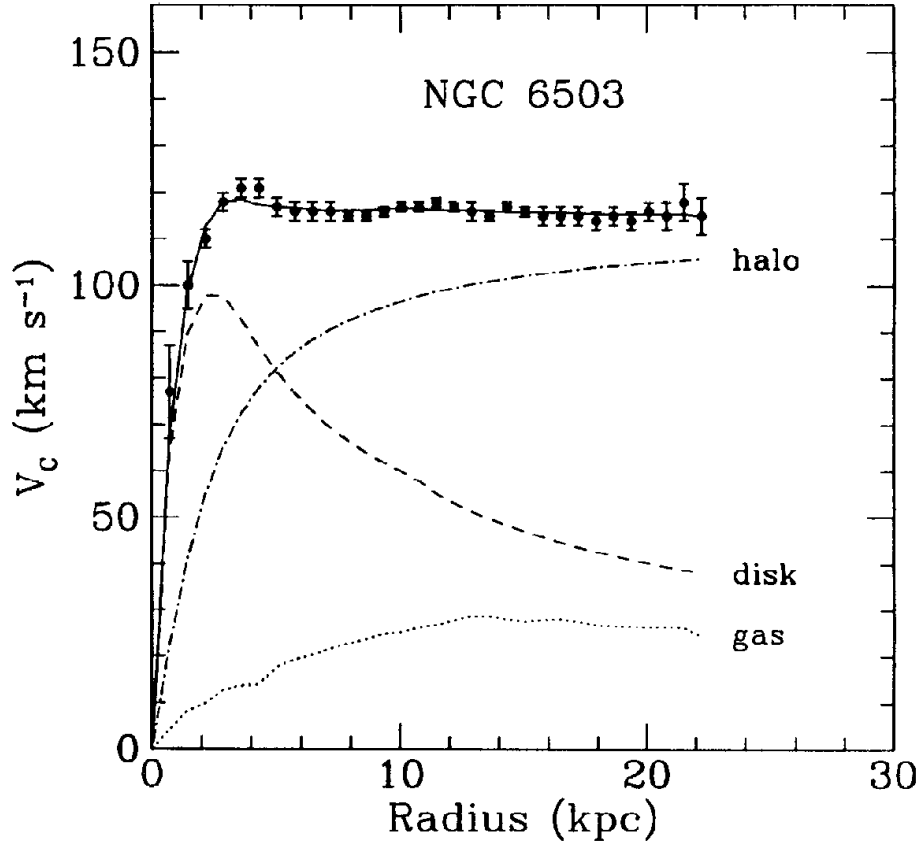


Figure 3.2: The rotational velocity of the galaxy NGC 6503 as a function of the distance from the galactic centre. The disc and gas contributions necessitate a dark matter halo to match the observed data [49].

The X-ray images of the Bullet Cluster (Figure 3.3), showing two colliding galaxies, provide further support for the existence of DM. The green contours indicating the distribution of gravitational mass are displaced with respect to the hot intergalactic plasma, indicated by the colour-scale. This observation suggests that most of the mass in the Bullet Cluster is less affected by the collision and is not visible, thus DM.

In the Early Universe, collisions between particles were frequent due to the high density and temperature. Photons could not travel freely in this plasma, and the ordinary matter was tightly coupled to them, leading to an ‘opaque’ Universe. The radiation pressure caused by photons opposed the creation of any concentration of matter under the effect

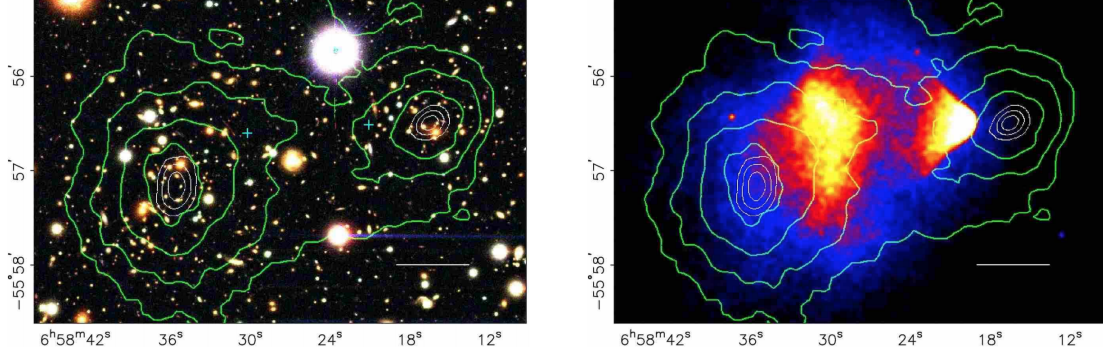


Figure 3.3: (left): Contours of the Bullet Cluster mass distribution from gravitational lensing (optical images from the Magellan telescope). (right): Data from Chandra X-ray that traces hot plasma in a galaxy, plotted over the same contours [52].

of gravity. Thus, fluctuations in the distribution of ordinary matter were prevented to grow denser as long as the baryonic matter was coupled to the photons. The temperature anisotropies observed in the cosmic microwave background are too small to explain the large scale structure that is observed today. A new massive component, that is non-baryonic and whose density fluctuations could have started growing earlier than ordinary matter, needs to be taken into account to explain how the initial seeds gave rise to this large scale structure. The existence of DM particles adds an extra density amplitude not visible in the cosmic microwave background since DM does not couple directly to photons. Therefore, the theorisation of DM as a particle is justified by the fluctuations in its distribution having grown denser even before the decoupling of the cosmic microwave radiation.

3.5.1 Higgs to Invisible Decays as a Search for Dark Matter

The SM predicts the Higgs boson to decay to four neutrinos with a branching ratio (BR) of approximately $\mathcal{B}(H \rightarrow \text{inv.}) = 0.12\%$ [53]. This final state, produced through an intermediate Z boson pair, is ‘invisible’ at collider experiments, such as CMS. This experimental issue is overcome by exploiting the additional final state objects produced, along with the Higgs boson, as detailed in Section 3.6.

A large $\mathcal{B}(H \rightarrow \text{inv.})$ would be strong evidence for new physics, specifically indicating coupling to BSM particles. One or more of these particles could be a candidate for DM. By considering the conservation of momentum, the momentum in the plane transverse to the LHC beams is zero before the interaction. After, the momentum of the visible system

is non-zero due to the invisible particles, and the magnitude of the momentum in the transverse plane is denoted as MET. Therefore, this MET is used as the observable in such searches, which are referred to as ‘direct searches’, and the Higgs boson is required to recoil against a visible system that is necessary to tag these events.

Indirect searches measure the total Higgs boson decay width and the sum of all visible decay widths. Any difference in the measurements represents an indication of an invisibly decaying Higgs boson. Understanding Higgs boson production and decay mechanisms is thus vital for both direct and indirect searches.

3.6 Higgs Production and Decay Modes

The LHC collisions produce many different final states, including the particles in Tables 3.1 and 3.2. This thesis focuses on those produced with and from a Higgs boson. Thus, the production and decay mechanisms of the Higgs boson are described in the following.

The dominant production modes are: gluon fusion (ggH), vector boson fusion (VBF), vector boson associated production (VH), and top quark associated production (ttH). Figure 3.4 shows the Feynman diagrams for the ggH, VBF, and VH channels, used in this thesis to search for invisibly decaying Higgs bosons. The cross-sections for each production mode as a function of the mass of the Higgs boson are shown in Figure 3.5 and discussed in the following.

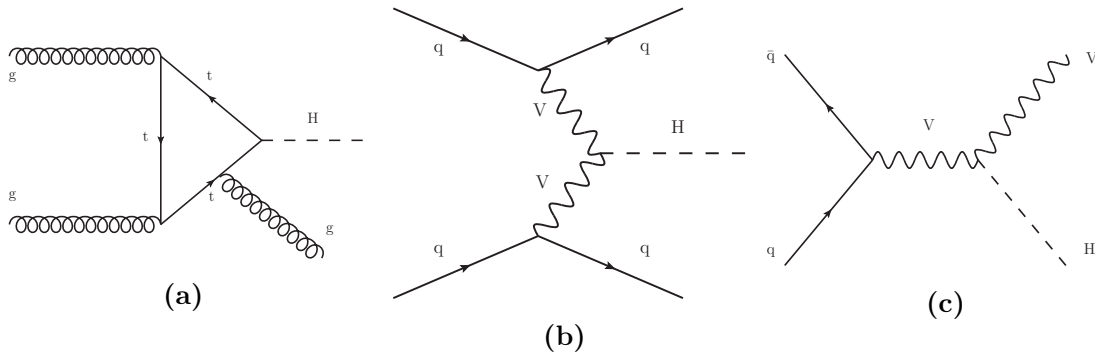


Figure 3.4: The Feynman diagrams of the three production modes targeted in the search for invisibly decaying Higgs bosons: (a) ggH (with ISR), (b) VBF, and (c) VH [2].

Although the ggH mode has the highest production rate, the initial state radiation (ISR) required to form a visible recoil system in invisibly decaying Higgs boson searches significantly decreases the rate to about 28% of the total inclusive ggH cross-section [54].

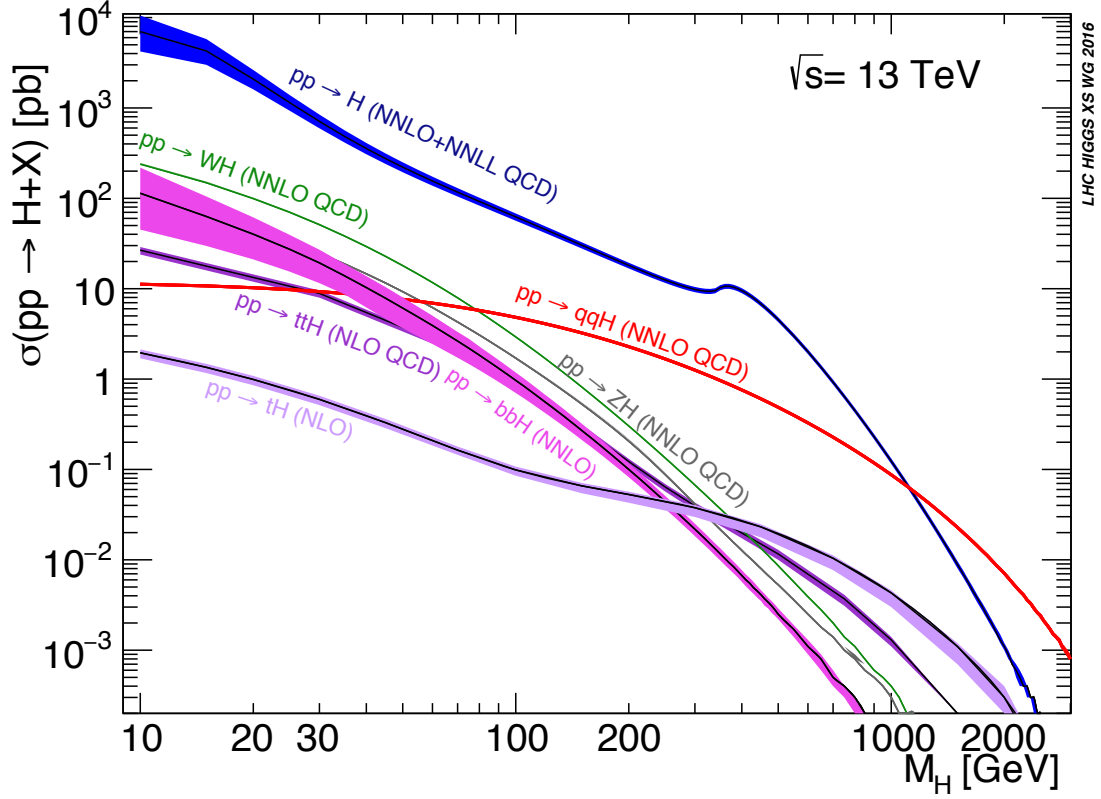


Figure 3.5: SM Higgs boson production cross-sections at $\sqrt{s} = 13$ TeV [53].

The ISR is QCD radiation from the initial state of a ggH process that results in gluons and quarks in the final state. Experimentally, these particles are observed as collimated jets of hadrons due to the colour confinement phenomenon [55]. These jets are then used to tag the event of interest, requiring a careful signal-background discrimination since they are similar to other QCD background processes.

The VBF channel has the second highest production cross-section. It is characterised by two quarks radiating vector bosons that produce a Higgs boson after fusing. The two quarks hadronising in jets in the final state represent the visible system necessary for Higgs to invisible searches. There is no connection in terms of the strong force (also referred to as ‘colour connection’) between the two quarks. This results in the jets having a distinctive topology, being well separated in η , with low hadronic activity between them. For these reasons, the VBF mode is the most sensitive channel for Higgs to invisible searches, and it is the main production mode used in this thesis.

The VH channel features a vector boson (V) produced in association with the Higgs boson. It can decay either leptonically, characterised by a clean final state where particle

identification is easy to perform, or hadronically, producing QCD final states similar to ggH . However, the lower VH cross-section penalises this channel when performing DM searches.

The $t\bar{t}H$ production mode is characterised by two top quarks produced with the Higgs boson. The hadronised jets in the final state are used to tag these events when a Higgs boson decays to invisible particles. The backgrounds are higher compared to the previous channels, and the cross-section is the lowest among the four. Thus, performing Higgs to invisible searches is extremely challenging without a massive amount of data.

Measuring the difference between the total Higgs boson decay width and the sum of all visible decay widths, as stated in Section 3.5.1, is one method to place limits on the coupling of the Higgs boson to invisible particles. The coupling of a particle to the Higgs boson is greater for particles with larger mass, leading to heavier particles having larger BRs. However, the current precision for a measurement of the total Higgs boson decay width is not sufficient to reach the SM prediction of a few MeV. An assumption on the Higgs boson width is thus necessary to constrain $\mathcal{B}(H \rightarrow \text{inv.})$ using visible widths.

The Higgs boson couplings can be parametrised in terms of coupling modifiers (κ) to account for BSM physics in the interpretation of the LHC data. The factorisation of the production and decay of the Higgs boson results as follows:

$$\sigma_i \cdot \text{BR}^f = \frac{\sigma_i(\vec{\kappa}) \cdot \Gamma^f(\vec{\kappa})}{\Gamma_H}, \quad (3.37)$$

where σ_i is the production cross-section of the Higgs boson ($i \rightarrow H$), and BR^f is the branching fraction for Higgs boson decay to the final state f ($H \rightarrow f$). The partial width for $H \rightarrow f$ is Γ^f , and Γ_H is the total width of the Higgs boson. The set of coupling modifiers $\vec{\kappa}$ accounts for possible BSM couplings of the Higgs boson, where a coupling modifier κ_i is defined as:

$$\kappa_i^2 = \frac{\sigma_i}{\sigma_{i,\text{SM}}} \quad \text{and} \quad \kappa_i^2 = \frac{\Gamma^i}{\Gamma_{\text{SM}}^i} \quad (3.38)$$

for a given production process and decay mode, respectively, with $\kappa_i = 1$ referring to the SM couplings. This κ -framework is used in Section 6.3 to interpret results for non-SM production of an invisibly decaying Higgs boson. A full treatment of the κ -framework can be found in Refs. [56, 57].

Figure 3.6 shows the agreement between the observed BRs of several final states and their SM prediction for two different parameterisations. The first (green) does not allow for any additional BSM contribution to the width of the Higgs boson, i.e. $B_{\text{BSM}} = 0$. The second (black) assumes $|\kappa_V| \leq 1$, where V refers to the Z or W boson, and $B_{\text{BSM}} \geq 1$. The results with their uncertainties are reported using data collected by both the ATLAS and CMS experiments. The hatched area indicates that κ_t is constrained to be positive without loss of generality. From these measurements, an upper limit on $\mathcal{B}(H \rightarrow \text{inv.})$ is set at 0.34 at 95% confidence level (CL) [57]. This allows for additional decay modes, so far unconstrained, to be present, where the decay of the Higgs boson to invisible particles is a possible component in BSM interpretations.

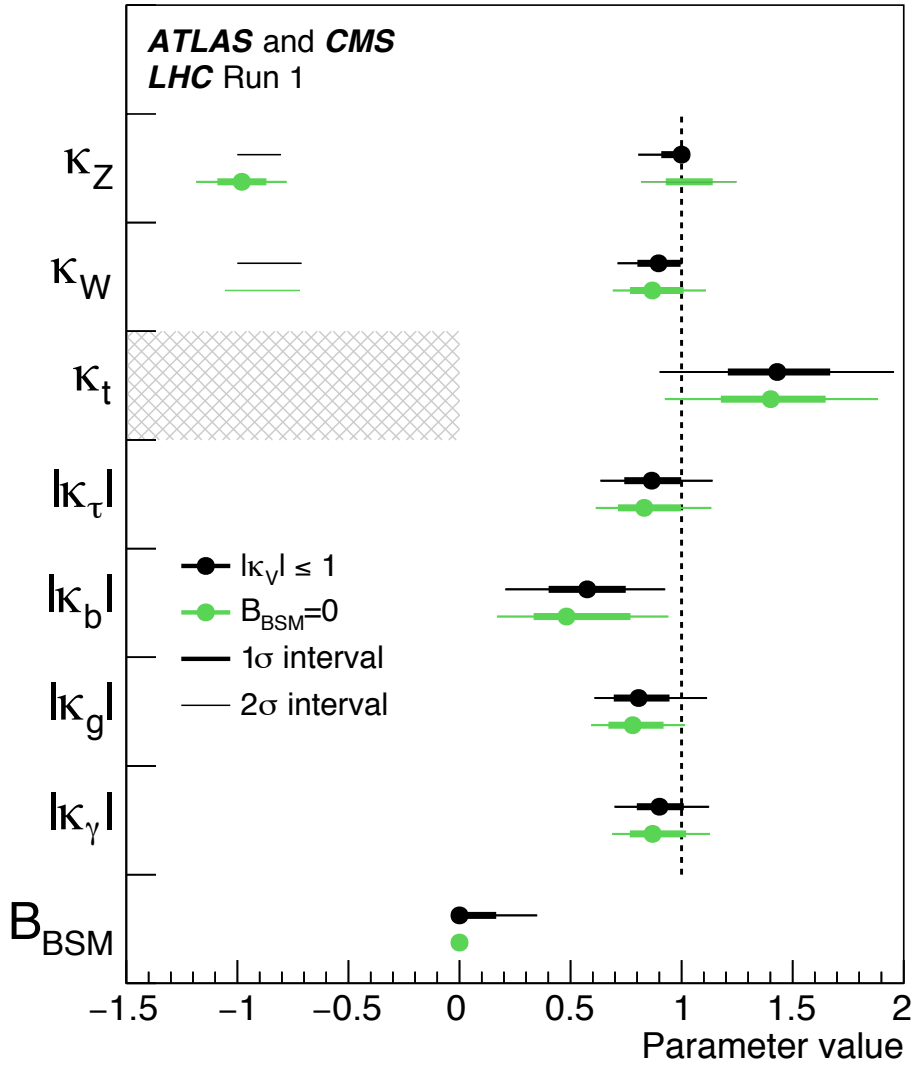


Figure 3.6: The measurements on the Higgs boson couplings with SM particles and the estimated $\mathcal{B}(\text{H} \rightarrow \text{inv.})$ (B_{BSM}) from visible channels [57].

Chapter 4

Event Reconstruction and Analysis Objects

This Chapter gives an overview of the event reconstruction for the objects required in searches for invisible decays of the Higgs boson. Section 4.1 presents the algorithm used to combine the CMS sub-detector information into individual particles. Section 4.2 details the objects used in the analysis and their reconstruction using the CMS detector. Each reconstructed object is required to pass a very loose selection to considerably reduce the number of fake objects, from mis-identification, entering the VBF Higgs to invisible analysis presented in Chapter 5.

4.1 Particle-Flow

The CMS event reconstruction combines different sub-detector information into individual particles using a specific algorithm: particle-flow (PF) [58]. The particles are classified as charged hadrons, neutral hadrons, photons, muons, or electrons, and they are referred to as PF candidates. They are used, for example, to calculate E_T^{miss} or lepton isolation, and to perform jet reconstruction. The PF approach is also used to make better estimates of various particle properties across the CMS sub-systems. As an example, the energy measurement of an HCAL object is improved by measuring the momentum in the tracker and the deposits in the ECAL.

The first step performed by the PF algorithm uses tracks, detailed in Section 4.2.1, and clusters in the calorimeter system. These clusters are reconstructed separately in the ECAL and HCAL from the energy deposits in their cells. The local maximum energy of a

cell must be greater than twice the expected noise, which is 80 MeV in the ECAL barrel, 300 MeV in the ECAL end-caps, and 800 MeV in the HCAL. Adjacent cells reaching these energies are clustered together around the cell with the highest energy, which is referred to as a ‘seed’.

A charged particle is identified using a cluster, in either the ECAL or HCAL, and a compatible track trajectory in the tracker system. Once the identification of charged objects is complete, the remaining calorimetric clusters are associated with neutral particles. The neutral candidates are matched with clusters in either the ECAL or HCAL depending on the type of particle, and there must be no compatible tracks in the tracker.

4.2 Physics Objects

This thesis uses several physics objects in the search for an invisibly decaying Higgs boson to identify signal-like events, suppress backgrounds, and define control regions (CRs) in data for the background estimation. The recommendations of the CMS physics object groups (POGs) are followed, with a few exceptions justified by analysis related constraints. These physics objects and the selection requirements imposed on them are described in the following.

4.2.1 Tracks

Tracks are vital for the reconstruction of most objects described in the following sections. They are reconstructed in the CMS tracker using a combinatorial track finder (CTF) algorithm, described in further detail in Ref. [23]. Hits in the pixel tracker generate seeds used to initiate the CTF algorithm. In the case of seeds with only two hits, the initial momentum related to the track is constrained using the nominal crossing point of the beams. This procedure is iterated for each layer of the tracker, adding the most compatible hit to the track, and performing a fit to the selected hits before moving to the next layer. Tracks with the lowest number of hits are discarded. Furthermore, the algorithm checks for tracks sharing the same hits, performing a best-fit on each of them. This process is iterated, and hits associated with a fully-reconstructed track related to previous iterations are not considered to avoid a non-physical overlap. The CTF algorithm efficiency is measured in data, using tracks from $Z \rightarrow \mu^+ \mu^-$ events, to be above 99% for muons with $1 < p_T < 100$ GeV.

A different algorithm is used to re-fit tracks once the previous stage is complete, proceeding from the inner to the outer layers [23]. The aim is to reduce the bias from the seed of the track and from the beam crossing point constraint. The parameters of the best-fit track are then re-estimated by a third algorithm that iterates from the outer to the inner layers of the sub-detector to ensure redundancy in the procedure. Quality criteria are applied on the final reconstructed objects to reject fake tracks.

4.2.2 Primary Vertex

The large probability of multiple pp interactions per bunch crossing at the LHC luminosity leads to the necessity of identifying the primary vertex (PV) of an interaction as the vertex most likely to contain the hard, or highest energy, interaction. The algorithm implementing the PV reconstruction consists of three steps: selection of tracks, clustering of tracks into vertices, and fitting of vertices positions. The first step selects tracks in the primary interaction region which have a large transverse impact parameter, defined as the transverse distance of closest approach to the PV or to the beam line. The second step uses prototype vertices and an adaptive vertex filter to determine the position [59]. The position of the vertex is then fitted and weights are assigned to each track according to the probability that it belongs to that vertex. This process is iterated for each prototype vertex. Further details can be found in Ref. [23].

The PV is defined as the vertex with the highest sum of the transverse momenta of the constituent tracks. Events with at least one jet ($p_T > 20$ GeV) are used to measure the reconstruction performance, resulting in an efficiency greater than 99% to reconstruct at least one PV with at least three tracks [23]. The VBF Higgs to invisible analysis, described in Chapter 5, requires a maximum displacement of the PV of 24 cm in the z -direction and 2 cm in the x - y plane from the centre of the detector.

4.2.3 Electrons

An electron candidate is reconstructed by using tracking and calorimetric information. However, an electron interacting with a magnetic field loses energy in the form of Bremsstrahlung photons. It is estimated that about 35% of electrons lose more than 70% of their initial energy through this effect [60]. The electron and the Bremsstrahlung photons are spread in the ϕ direction because of the strong magnetic field of CMS, and both create EM showers in the ECAL. Therefore, the reconstruction of electrons uses superclusters

combining EM deposits from the showering electrons and the Bremsstrahlung photons. The barrel and end-cap regions use different algorithms for superclusters due to the difference in their geometry [61]. A hybrid clustering algorithm is implemented for the barrel, adding arrays of five crystals around the seed within a distance of 17 crystals from it in an η - ϕ grid configuration. The end-cap region implements the so-called multi- 5×5 algorithm [61], using the identified seed as the centre for a 5×5 square of crystals defining the cluster. Deposits in the pre-shower detector are also used and matched to the clusters of a supercluster under certain selections. The centre of the supercluster is chosen as the average position, weighted in energy, of all the clusters considered that satisfy several requirements. The position and energy of the supercluster are used together with the information from the innermost layer of the tracker to extrapolate the track of an electron. The supercluster algorithms are described in further detail in Ref. [61].

The electrons are required to be isolated from other activity in the detector. Considering a cone of $\Delta R < 0.3$ ($\Delta R^2 = \Delta\phi^2 + \Delta\eta^2$) around the electron direction, the PF isolation (ISO) is defined as the p_T sum of all PF candidates whose trajectories are within the cone. This is corrected for pileup by subtracting the event-by-event energy density (ρ) times an effective area.

Several identification (ID) criteria are necessary to reject fake electrons, having the same signature as real electrons but produced from other particles. The VBF Higgs to invisible analysis uses two categorisations for electrons, both selected for $|\eta| < 2.5$: ‘veto’ and ‘tight’. The first categorisation uses loose ID criteria to veto events containing electrons. The tight categorisation has more stringent criteria, and is used to select events with electrons in the final state. These electrons are required to be separated by $\Delta R > 0.3$ from any veto lepton. Electrons with $p_T > 50$ GeV and $|\eta| < 1.479$ are reconstructed with efficiencies of 93% and 85% for the veto and tight selections, respectively [61]. This thesis uses electrons with $p_T > 40$ GeV. During the first part of 2016 data-taking, the tracking efficiency and performance were strongly affected by dead-time in the silicon-strip sensors due to a saturation effect of the pre-amplifiers. This problem lowered the reconstruction efficiency for both electrons and muons, as well as for the b-tagging efficiency. For this reason, an overall 1% systematic uncertainty on the electron reconstruction scale factors is assigned in the analysis.

Table 4.1 summarises the main selection used by the VBF Higgs to invisible analysis for both veto and tight electron ID. The quantities $\Delta\phi_{in-seed}$ and $\Delta\eta_{in-seed}$ are the distances between the position of the supercluster and the electron trajectory extrapolated to the ECAL in the ϕ and η directions, respectively. The η width of a cluster, weighted in

energy, is given by $\sigma_{i\eta i\eta}$, whereas $\frac{H}{E}$ is the ratio between the energy deposited in the HCAL and the ECAL within the region of the seed cluster. The distance $|d_{xy}(\text{vtx})|$ is the transverse impact parameter and $|d_z(\text{vtx})|$ is the longitudinal impact parameter with respect to the primary vertex, both expressed in cm.

Table 4.1: The veto and tight electron identification criteria.

Variable	Veto Electrons		Tight Electrons	
	Barrel	End-Caps	Barrel	End-Caps
Full 5x5 $\sigma_{i\eta i\eta}$	< 0.0115	< 0.0370	< 0.00998	< 0.0292
$ \Delta\eta_{in-seed} $	< 0.00749	< 0.0089	< 0.00308	< 0.00605
$ \Delta\phi_{in-seed} $	< 0.228	< 0.213	< 0.0816	< 0.0394
H/E	< 0.356	< 0.211	< 0.0414	< 0.0641
ISO (ρ -corrected)	< 0.175	< 0.159	< 0.0588	< 0.0571
$ 1/E - 1/p $	< 0.299	< 0.150	< 0.0129	< 0.0129
$ d_{xy}(\text{vtx}) $	< 0.050	< 0.100	< 0.050	< 0.100
$ d_z(\text{vtx}) $	< 0.100	< 0.200	< 0.100	< 0.200
Expected Inner Missing Hits	≤ 2	≤ 3	≤ 2	≤ 1

Events containing well-reconstructed electrons that pass the veto selection are rejected to suppress EW backgrounds, in the signal region (SR). Electrons are also used to estimate these backgrounds by defining CRs, namely single-electron and double-electron, for each key process: $W \rightarrow e^\pm \nu$ and $Z \rightarrow e^+ e^-$. This data-driven approach uses the tight selection, and it is further detailed in Section 5.5.

4.2.4 Photons

A photon veto is applied to suppress backgrounds such as $Z(\nu\bar{\nu}) + \gamma + \text{jet}$ or $W(\ell^\pm \nu) + \gamma + \text{jet}$. Events with one or more photons with $p_T > 15 \text{ GeV}$ and $|\eta| < 2.5$ that pass loose ID criteria (Table 4.2) are rejected. The isolation value is computed in a ΔR cone of 0.3 and is corrected for pileup using the ρ -correction as in the case of electrons. The criteria shown in Table 4.2, including the dependence on the p_T of the photon, are provided by the POGs for all CMS analyses.

Table 4.2: The loose photon identification criteria.

Variable	Loose Photons	
	Barrel	End-Caps
Full 5x5 $\sigma_{i\eta i\eta}$	< 0.0103	< 0.0301
H/E	< 0.0597	< 0.0481
Charged Hadron ISO	< 1.295	< 1.011
Neutral Hadron ISO	$< 10.92 + 0.0148 \times p_T + 1.7 \times 10^{-5} \times p_T^2$	$< 5.931 + 0.0163 \times p_T + 1.4 \times 10^{-5} \times p_T^2$
Photon ISO	$< 3.630 + 0.0047 \times p_T$	$< 6.641 + 0.0034 \times p_T$

4.2.5 Muons

A muon candidate has high mass and is not subject to the strong interaction. As a result, energy deposits from muons in the calorimeters are low, and they pass through the muon system and leave the detector. For this reason, a muon can be reconstructed by matching compatible tracks in the tracker and in the muon system. This approach, referred to as global muon reconstruction, also helps to discriminate between muons and hadronic activity. The algorithm behind the reconstruction runs on each track in the muon system to look for compatible tracks in the tracker. A track fit is then performed using all the hits of both sub-detectors and accounting for possible energy loss as the muon leaves the detector [62].

In this thesis, muons are required to be within $|\eta| < 2.4$ and isolated, considering a cone of $\Delta R < 0.4$. A muon passing the global reconstruction and satisfying the ISO criteria is referred to as a veto muon. These muons are used to veto events containing muons, suppressing EW backgrounds.

The study of events containing muons in the final state is performed using tight muons. The identification of these muons requires hits in at least five inner tracker layers, one of which must be a pixel layer. Moreover, at least one muon chamber hit must be included in the global track fit and muon segments in at least two muon stations are required in this selection. The transverse impact parameter is $|d_{xy}(\text{vtx})| < 0.2 \text{ cm}$ and the longitudinal is $|d_z(\text{vtx})| < 0.5 \text{ cm}$ with respect to the primary vertex. Tight muons are used to estimate EW backgrounds by defining two CRs: single-muon for $W \rightarrow \mu^\pm \nu$ and double-muon for $Z \rightarrow \mu^+ \mu^-$ processes. This thesis uses muon candidates with $p_T > 20 \text{ GeV}$, which are reconstructed with an efficiency greater than 99% [62]. An overall 1% systematic uncertainty on the muon reconstruction scale factors is assigned in the analysis, as in the case of electrons detailed in Section 4.2.3.

4.2.6 Tau Leptons

A tau candidate can be categorised as either a leptonically or hadronically decaying tau lepton. The first category comprises about 35% of tau leptons, with a final state characterised by electrons, muons, and neutrinos [40]. The charged leptons are reconstructed as prompt leptons contributing to the electron and muon candidates as described in Section 4.2.3 and Section 4.2.5, respectively. Therefore, a specific reconstruction is not necessary for leptonically decaying tau leptons. The neutrinos are not reconstructed and are included in the MET calculation.

The remaining 65% of the tau leptons (τ_h) decay to hadrons and tau neutrinos. A reconstruction for this category is necessary, and the hadron plus strips (HPS) algorithm [63] is used for this purpose. Possible final states include charged hadrons, and neutral pions decaying into photon pairs [40]. The HPS features the reconstruction of both charged hadrons and photons, and it uses a PF jet seed looking for the most energetic EM PF candidate in the jet. The algorithm creates a strip with the related four-momentum, and a 0.05×0.2 η - ϕ window at the strip centre is used to select other high-energy EM candidates to be added to the strip using the same logic. The procedure is iterated, updating the four-momentum until there are no compatible candidates present in the window. A τ_h candidate is then defined as a compatible combination of strips and charged hadrons consistent with tau lepton decay final states.

As in the case of muons and electrons, events containing identified tau leptons are rejected to suppress EW backgrounds. The tau leptons are required to have $p_T > 18$ GeV and $|\eta| < 2.3$. A data-to-Monte Carlo (MC) efficiency correction of 0.99 ± 0.05 is measured on $Z \rightarrow \tau^+ \tau^-$ events, and used to estimate the tau veto uncertainty. Considering a cone of $\Delta R < 0.5$ around the tau lepton direction, the tau ISO is defined as the p_T sum of all photon and hadron PF candidates from the PV whose trajectories are contained within the cone. Fake tau leptons are rejected using information from other sub-detectors, e.g. tracks and hits in the muon system, leading to a fake rate of about 2% in the barrel and 3% in the end-caps [63].

4.2.7 Jets

A specific signature of two quarks in the final state characterises an invisibly decaying Higgs boson produced through the VBF mode. These two quarks hadronise in highly collimated jets. The anti- k_T clustering algorithm [64] is used to combine particles into

jets. These particles are the product of a particle shower from hadronisation. The algorithm is able to distinguish sub-processes as soft gluon radiation or gluon production during the hadronisation phase. Two metrics are used: the distance between particles (d_{ij}) and the distance from a particle to a nominal beam line particle (d_{iB}). In particular, they are used to: (i) calculate d_{ij} of all possible particle pairs in an event including the nominal beam line; (ii.a) combine particles i and j into a single particle for a minimum d_{ij} , i.e. if $d_{ij} = \min(d_{ij})$, and return to stage (i); (ii.b) in case d_{iB} is minimum, consider i as a jet candidate, remove it from the list of particles, and return to stage (i); (iii) continue until all jet candidates are found. The metrics are defined as:

$$d_{iB} = \frac{1}{p_{T,i}^2}, \quad d_{ij} = \min\left(\frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2}\right) \frac{\Delta R_{ij}^2}{R^2}, \quad (4.1)$$

where R is a parameter related to the jet cone radius and ΔR_{ij} is the η - ϕ distance between particles. In this context, $R = 0.4$ and the resulting clustered jets are referred to as AK4 jets.

The anti- k_T algorithm uses PF candidates as input, producing PF jets as output. A charged hadron subtraction is performed during the jet clustering, using only charged PF candidates that are associated with the chosen PV. Jets are corrected for the contribution of neutral particles from pileup interactions by using an event-by-event energy density estimation. Further energy calibrations are applied to jets in data to match their response with the simulation. A similar procedure is used on MC events after the generator step and before starting the detector simulation stage.

The reconstructed jets are required to have $p_T > 30 \text{ GeV}$, and to be fully contained within the detector, i.e. $|\eta| < 4.7$. A set of loose ID criteria is also used to reject mis-reconstructed jets, detector noise, or beam background. At least two PF candidates are required to be contained in a jet, with a contribution from both the ECAL and HCAL. If information from the tracker is available, at least one charged object must contribute to the energy of the jet, but this must be less than 99% of the jet energy in the case of electrons. Both the neutral hadron and photon contributions to the energy of the jet are required to be less than 99% of the energy of the jet. Finally, if the leading jet is within the tracker acceptance, i.e. $|\eta| < 2.4$, the energy fraction attributed to charged and neutral hadrons is required to be greater than 10% and less than 80%, respectively. An efficiency above 99% is estimated for jets from gluons or quarks passing these criteria. In addition, jets with $\Delta R < 0.4$ from any veto lepton are rejected to reduce possible fake jets from mis-reconstructed leptons.

4.2.8 B-tagged Jets

The jets arising from a b-quark, also referred to as b-tagged jets or b-jets, are tagged and identified from the collection of AK4 jets by requiring that the value of the combined secondary vertex version-2 (CSVv2) tagger is greater than 0.85 [65, 66]. The jets are required to have $p_T > 20$ GeV and to be inside the tracker coverage ($|\eta| < 2.4$) during identification.

Events containing b-jets are vetoed, reducing the top background by a factor of 3 inside the SR. This also has a minor impact on the VBF Higgs to invisible signal rate, which decreases by roughly 6%. Finally, efficiency scale factors are used on simulated events to improve the data-to-MC agreement.

4.2.9 MET

A particle passing through the CMS detector and not interacting cannot be reconstructed, and thus leaves a transverse momentum imbalance as a signature. A particularly large amount of MET is interpreted as evidence for weakly-interacting particles, that are either neutrinos or BSM candidates, such as DM.

The MET, defined in Section 2.3, is vital in searches for invisible decays of the Higgs boson, and is measured accurately as a result of the CMS detector's high hermeticity. A set of filters is applied to avoid spurious MET. They account for the particle halo surrounding the proton beams, noise from ECAL or HCAL, mis-reconstructed backgrounds, and corrections from pileup contributions.

For processes such as $Z \rightarrow \mu^+ \mu^-$ or $Z \rightarrow e^+ e^-$, MET should not be present in the final state since all decay products are visible. However, these events feature MET as shown by Figure 4.1. The main reason for this is the relatively poor resolution of the p_T measurements of the jet candidates. Corrections to the momenta of PF jets reconstructed in the event are further propagated to the MET to minimise this effect. Figure 4.2 shows a similar effect for the $W \rightarrow \mu^\pm \nu$ and $W \rightarrow e^\pm \nu$ processes. The distribution of MET has a different shape in comparison to that in the double-lepton CRs since real MET from neutrinos is expected to be present in the final state of the single-lepton processes. The last bin in Figure 4.1 (Figure 4.2) includes events passing the requirement $\text{MET} > 160(500)$ GeV, and is referred to as the ‘overflow bin’ of the distribution.

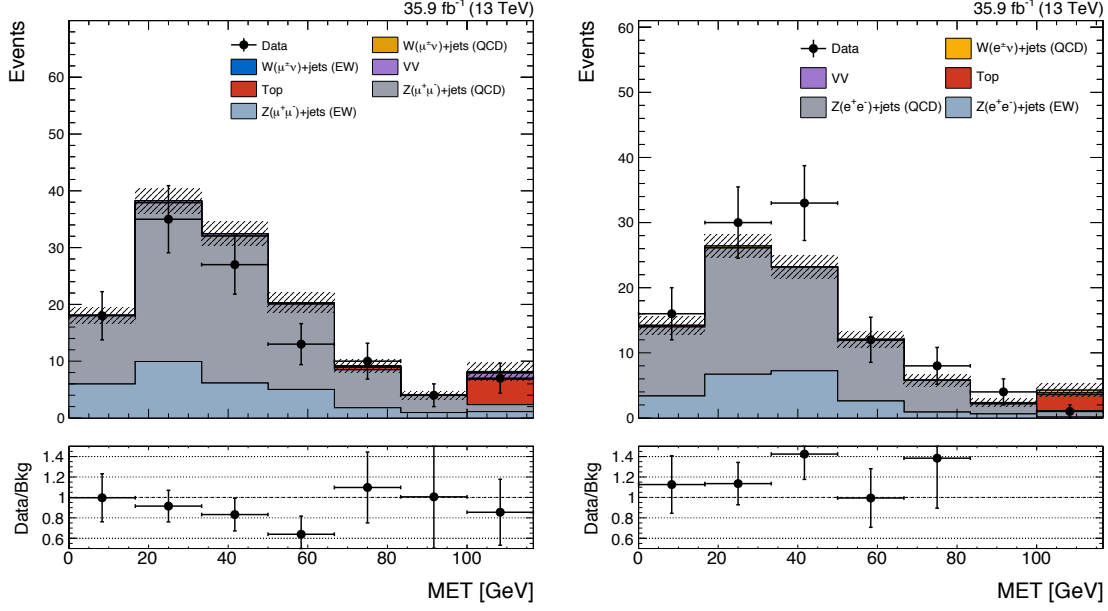


Figure 4.1: The distribution of MET for data and simulated backgrounds in the $Z \rightarrow \mu^+\mu^-$ (left) and $Z \rightarrow e^+e^-$ (right) CRs.

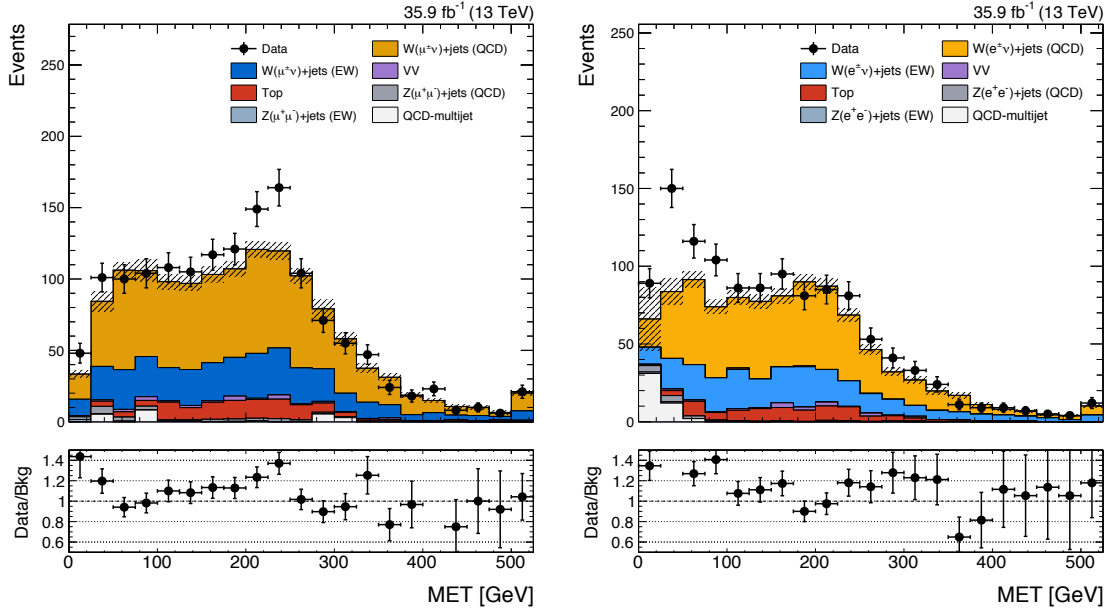


Figure 4.2: The distribution of MET for data and simulated backgrounds in the $W \rightarrow \mu^\pm\nu$ (left) and $W \rightarrow e^\pm\nu$ (right) CRs.

Events containing leptons from W or Z boson decays are used to estimate the major backgrounds for invisibly decaying Higgs boson searches, as mentioned in Sections 4.2.3 and 4.2.5. These CRs, characterised by the dominant production of W or Z bosons, are constructed by selecting either a single lepton or a pair of same flavour, opposite sign leptons. These (tight) leptons are removed during the calculation of MET, and the variable $E_{T,\text{no-lep}}^{\text{miss}}$, also referred to as hadronic recoil U , is used for the background estimation. The hadronic recoil against the Z and W bosons is calculated as:

$$\vec{U} = \vec{p}_T^{\text{miss}} + \vec{p}_T^{\ell^+ \ell^- / \ell^\pm}, \quad (4.2)$$

where the last term is double-lepton ($\ell^+ \ell^-$) or single-lepton ($\ell^\pm$) depending on the CR. In the case of $Z \rightarrow \nu \bar{\nu}$ decays, $\vec{U} = \vec{p}_T^{\text{miss}}$, and the recoil in the CRs can be used to mimic the MET in $Z \rightarrow \nu \bar{\nu}$ events that fulfil the signal selection.

Figure 4.3 shows the distribution of $E_{T,\text{no-lep}}^{\text{miss}}$ for the $Z \rightarrow \mu^+ \mu^-$ (left) and $Z \rightarrow e^+ e^-$ (right) processes. These CRs suffer from a low number of events passing the final selection detailed in Section 5.5.1. For this reason, the distribution of $E_{T,\text{no-lep}}^{\text{miss}}$ in single-lepton CRs is jointly used as proxy for the MET in the SR. Figure 4.4 shows the distribution of $E_{T,\text{no-lep}}^{\text{miss}}$ for the $W \rightarrow \mu^\pm \nu$ (left) and $W \rightarrow e^\pm \nu$ (right) processes. The $W \rightarrow \mu^\pm \nu$ CR shows improved agreement between data and simulated events in comparison with the $W \rightarrow e^\pm \nu$ CR, particularly in the low- $E_{T,\text{no-lep}}^{\text{miss}}$ region. The statistical uncertainty on the data in the high- $E_{T,\text{no-lep}}^{\text{miss}}$ region is comparatively high due to the low number of events, but nonetheless is consistent with simulation.

To minimise the effect of fake PF MET, a requirement on $|E_{T,\text{calorimetric}}^{\text{miss}} - E_{T,\text{PF}}^{\text{miss}}|/U < 0.5$ is added. This is the difference between the calorimetric MET and PF MET relative to the recoil, and it typically has a large value for events in which the PF MET is mis-measured or due to fake tracks. Figure 4.5 shows this distribution in the single-electron CR after a very loose VBF selection, i.e. requiring only the two highest- p_T jets to be in opposite hemispheres of the detector. The backgrounds are estimated using MC samples, and a mis-modelling of the QCD-multijet production causes a poor agreement between data and simulation in the low- $|E_{T,\text{calorimetric}}^{\text{miss}} - E_{T,\text{PF}}^{\text{miss}}|/U$ region. This specific effect is detailed in Section 5.5.3.

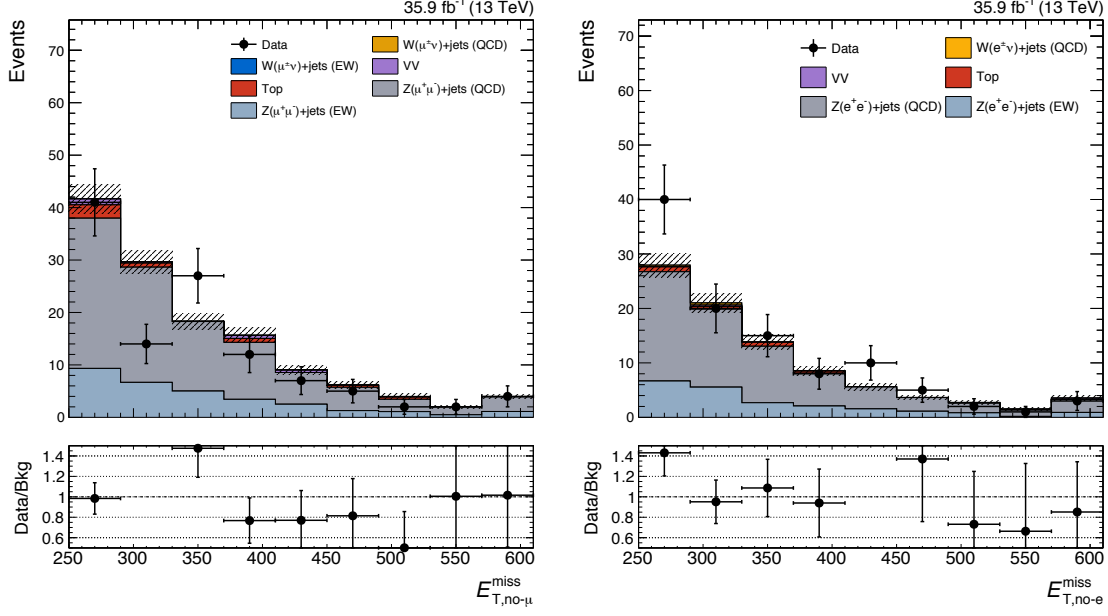


Figure 4.3: The distribution of $E_{T,no-lep}^{miss}$ for data and simulated backgrounds in the $Z \rightarrow \mu^+\mu^-$ (left) and $Z \rightarrow e^+e^-$ (right) CRs.

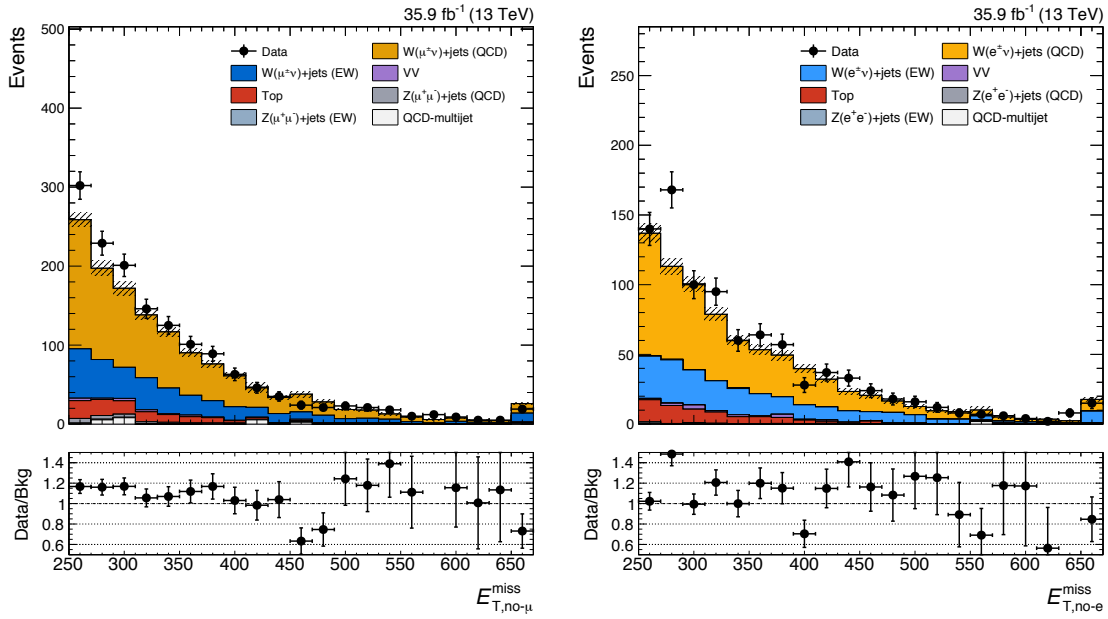


Figure 4.4: The distribution of $E_{T,no-lep}^{miss}$ for data and simulated backgrounds in the $W \rightarrow \mu^\pm\nu$ (left) and $W \rightarrow e^\pm\nu$ (right) CRs.

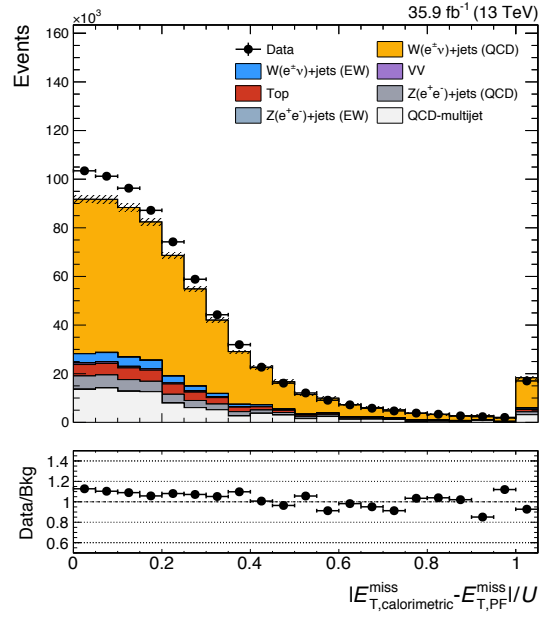


Figure 4.5: The difference between the calorimetric and PF MET relative to the recoil for data and simulated backgrounds passing a very loose VBF selection in the $W \rightarrow e^\pm \nu$ CR.

Chapter 5

Search for VBF Higgs Bosons Decaying to Invisible Final States in 2016 Run-2 Data

A search for invisibly decaying Higgs bosons produced in the VBF mode, using the 2016 data of the LHC Run-2, is presented in this Chapter. The VBF Higgs to invisible analysis is the main work conducted by the author, and is published in Ref. [1].

The experimental signature of an invisibly decaying Higgs boson consists of a significant amount of E_T^{miss} , and two VBF jets with large invariant mass and large pseudo-rapidity separation. The analysis presented in this thesis features two methods: a cut-based (cut-and-count) and a shape-based approach. The author led all aspects of the cut-based analysis, which features an easy re-interpretation of the results using different phenomenological models that predict the same signature in the final state. The simplicity and flexibility of this approach were fundamental to design and develop the second method. The shape-based approach is used to improve the sensitivity to an invisibly decaying Higgs boson by fully exploiting the kinematics of the VBF topology. The author was also responsible for several aspects of this analysis.

Section 5.1 reports the data and simulated samples, including signals and backgrounds, used in the analysis. Section 5.2 details the triggers used to select the data, whereas Section 5.3 highlights the procedure performed to match simulated events to data. Sections 5.4 and 5.5 describe the final selection of events implemented to extract the signal, and the control regions used to estimate the major backgrounds, respectively. Section 5.6 discusses the systematic uncertainties included in the analysis. An overview of the statistical method used to place upper limits on $\mathcal{B}(\text{H} \rightarrow \text{inv.})$ is given in Section 5.7.

Section 5.8 details the kinematics of the signal for an invisibly decaying Higgs boson, and the optimisation procedure used to obtain the final selection in Section 5.4. These kinematics also drive the specific choice of the cut-based selection. Finally, Section 5.9 shows the results and limits set on $\mathcal{B}(\text{H} \rightarrow \text{inv.})$.

Figure 5.1 gives a schematic overview of the analysis procedure and composition of this Chapter. Data are selected by different triggers and categorised in four datasets. These serve several purposes: to calculate the trigger efficiency, estimate the QCD background, and provide data for the SR and CRs. Simulated samples are reweighted to improve the agreement with data, and systematic uncertainties are included to account for limitations in the reweighting procedure. The signal samples, background samples, and data are grouped into the SR and CRs, and a simultaneous fit is preformed across these regions to optimise the analysis and produce the final results.

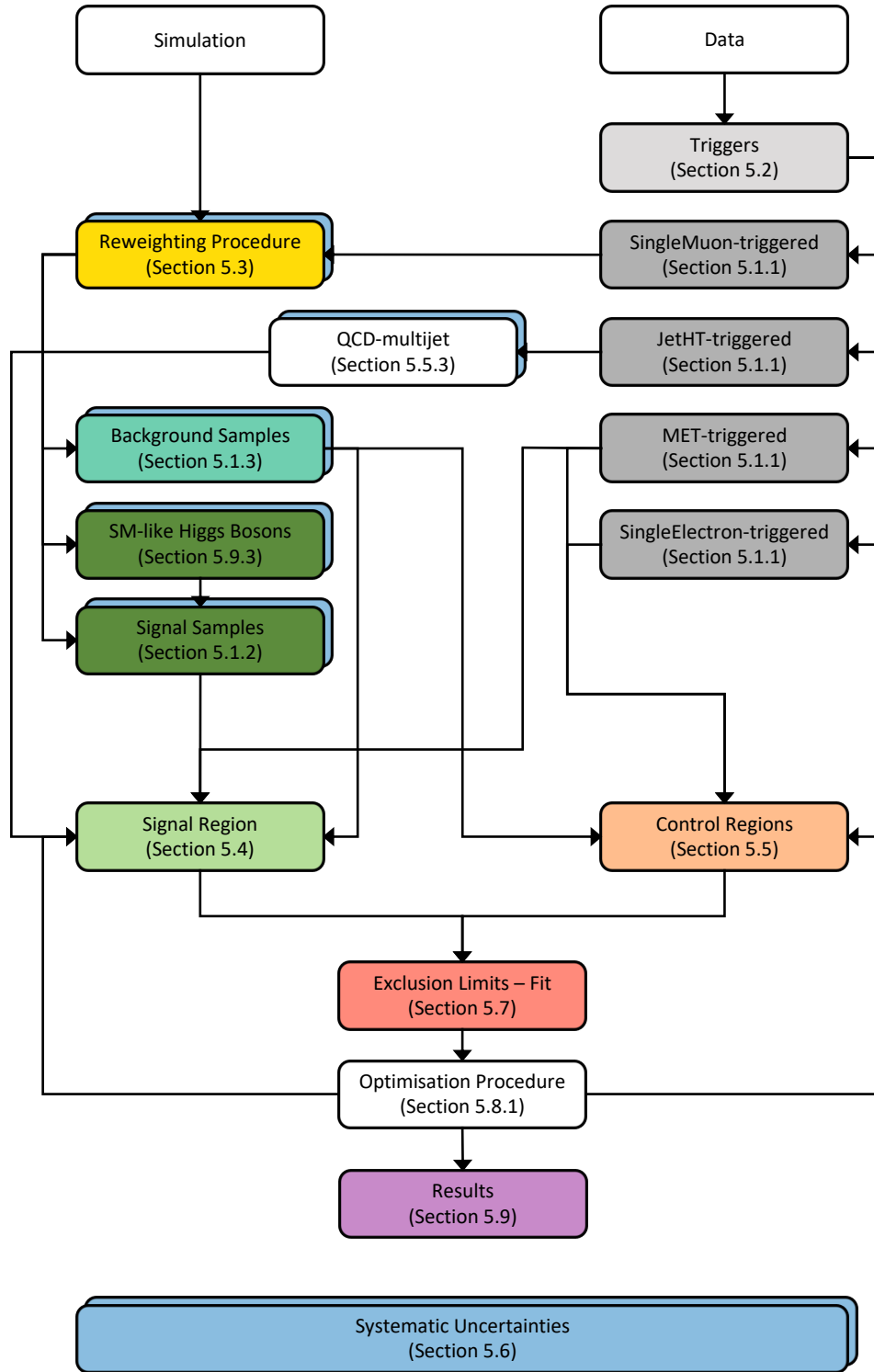


Figure 5.1: Schematic overview of Chapter 5 and the approach used to present the VBF Higgs to invisible analysis.

5.1 Samples

The analysis presented in this Chapter exploits different datasets and several simulated samples to perform the search for invisibly decaying Higgs bosons. These samples are discussed in the following Sections, providing details on the choices made and on their usage throughout the analysis.

5.1.1 Datasets

The events collected by CMS are stored in different datasets corresponding to the trigger paths used to select tailored process-enriched data samples. The VBF Higgs to invisible analysis uses datasets corresponding to an integrated luminosity of 35.9 fb^{-1} recorded in 2016. The MET-triggered dataset is used to select events in data that pass the signal event selection, and to define two CRs in data used to estimate the key backgrounds $W \rightarrow \mu^\pm \nu$ and $Z \rightarrow \mu^+ \mu^-$. The SINGLEELECTRON-triggered dataset is used to estimate backgrounds from $W \rightarrow e^\pm \nu$ and $Z \rightarrow e^+ e^-$ processes by defining two additional CRs, and to perform measurements on physics objects used in the analysis in order to obtain, for example, trigger efficiencies. A full study and estimation of the trigger efficiency for the muon CRs and for the SR is performed using the SINGLEMUON-triggered dataset. This is possible due to the fact that muons are not considered while using the MET-triggered path, as detailed in Section 4.2.9. Finally, the JETHHT-triggered dataset is used to estimate the QCD background, implementing a data-driven method.

5.1.2 Signal Samples

The simulated signal events, produced through the VBF and ggH modes and decaying invisibly, are generated using POWHEG [67–71]. The inclusive cross-sections are taken from the values provided by Ref. [53]: the VBF cross-section is calculated at next-to-leading order (NLO) EW and at NNLO+NNLL QCD, whereas the ggH cross-section is calculated at NLO EW and at N3LO QCD. This analysis uses 10^5 simulated VBF Higgs to invisible events with $\sigma = 3.782 \text{ pb}$ (equivalent luminosity of 26.44 fb^{-1}), and 2×10^5 simulated ggH Higgs to invisible events with $\sigma = 48.58 \text{ pb}$ (equivalent luminosity of 4.17 fb^{-1}).

5.1.3 Background Samples

There are several SM background processes for the VBF Higgs to invisible analysis. These mimic the experimental signature consisting of two VBF jets with large pseudo-rapidity separation and large invariant mass, and a significant amount of E_T^{miss} coming from the invisible Higgs boson decay. Simulated events for these backgrounds are detailed in the following.

Z($\nu\bar{\nu}$)+jets-QCD: This represents the main irreducible background and consists of a Z boson and two or more jets arising from QCD vertices, i.e. $O(\alpha_{\text{EW}}^4 \alpha_{\text{QCD}}^2)$ at leading order (LO).

Z($\nu\bar{\nu}$)+jets-EW: This is another irreducible background and consists of a Z boson and two or more jets coming from purely EW vertices, i.e. $O(\alpha_{\text{EW}}^6)$ at LO.

W($\ell^\pm\nu$)+jets-QCD and EW: These processes constitute the second largest background and consist of a W boson and two or more jets coming from either QCD or EW vertices. The contribution of this background arises when the leptons are outside of the detector acceptance, and can be reduced by applying a veto on events containing leptons.

Z($\ell^+\ell^-$)+jets-QCD and EW: These processes mimic signal-like events when the leptons from the Z boson decay are not reconstructed. The Z(e^+e^-)+jets and Z($\mu^+\mu^-$)+jets events are used as control samples for estimating the Z($\nu\bar{\nu}$)+jets background.

Top: The top-quark decays (both $t\bar{t}$ and single-top) constitute a background since the W boson produced in these processes can further decay leptonically, producing genuine E_T^{miss} in the event.

Dibosons: The decays of a diboson pair (VV), i.e. WW, WZ, or ZZ, also constitute a background. Typically, one boson decays leptonically ($W \rightarrow \ell^\pm\nu$ or $Z \rightarrow \nu\bar{\nu}$) whereas the second boson decays hadronically. This process produces jets and E_T^{miss} in the final state.

QCD: QCD events do not have large genuine E_T^{miss} but have a large production cross-section. Thus, events in which the energy of a high- p_T jet is mis-measured serve as a background to the search for invisibly decaying Higgs bosons.

The generation of certain simulated samples is performed as a function of H_T . This observable is defined as the scalar sum of the transverse momenta of all jets in an event.

MC samples are produced at LO in QCD using the MADGRAPH generator [72, 73] in several bins of H_T for the $Z(\nu\bar{\nu})$ +jets-QCD, $W(\ell^\pm\nu)$ +jets-QCD, $Z(\ell^+\ell^-)$ +jets-QCD, and QCD processes. Inclusive simulated events are instead used for EW processes. POWHEG is used to generate MC samples at NLO for the Top and WW processes, whereas PYTHIA8 [74] at LO is used for the WZ and ZZ MC samples. All MC events are passed through a GEANT4 based simulation [75] of the CMS detector to reproduce the interactions between final state particles and the different sub-detectors. Finally, reconstruction algorithms used for data are also applied to the simulated events.

5.2 Trigger

The data used for the VBF Higgs to invisible analysis are collected using several trigger paths. These triggers are designed to select events with large $p_T^{\text{miss}1}$ and large H_T^{miss} (missing transverse hadronic momentum). These HLT trigger variables rely on the PF algorithm, and identified muons are removed from the event before p_T^{miss} and H_T^{miss} are calculated. As a consequence, the same triggers select $Z \rightarrow \mu^+\mu^-$ and $W \rightarrow \mu^\pm\nu$ events used as control samples for the $Z(\nu\bar{\nu})$ +jets and $W(\ell^\pm\nu)$ +jets backgrounds, respectively. The $Z \rightarrow e^+e^-$ and $W \rightarrow e^\pm\nu$ events used to define the double-electron and single-electron CRs, respectively, are selected using SINGLEELECTRON-triggered dataset.

5.2.1 MET Triggers

The $p_T^{\text{miss}} + H_T^{\text{miss}}$ based signal triggers are designed to maximise the signal acceptance, given the rate and timing restrictions of the HLT. The p_T^{miss} and H_T^{miss} quantities are computed using PF candidates, and filters are applied to minimise the average time taken for the path. Additional ID criteria are imposed on jets that are used in the calculation of the H_T^{miss} variable to minimise anomalous MET affecting the rate and the timing of the trigger. All PF jets with $p_T > 20$ GeV are required to have a neutral hadron energy fraction below 90%.

The VBF Higgs to invisible analyses performed with data collected before 2016 are also used in this thesis when combined with other production channels of the Higgs boson among different Run-eras, as presented in Chapter 6. These analyses use dedicated VBF

¹The use of the variable p_T^{miss} in Section 5.2 is made to avoid confusion with the variable E_T^{miss} used during the analysis.

triggers with a selection based on VBF-jet p_T , di-jet invariant mass (m_{jj}), and di-jet pseudo-rapidity difference ($\Delta\eta_{jj}$). However, $p_T^{\text{miss}} + H_T^{\text{miss}}$ based triggers give higher signal efficiency and sharper efficiency turn-on curves compared to these dedicated triggers for VBF Higgs to invisible signal events, leading to the decision to not use VBF triggers in the 2016 dataset.

Figure 5.2 shows efficiencies for the VBF triggers and for $p_T^{\text{miss}} + H_T^{\text{miss}}$ based triggers for a VBF-like selection on simulated VBF Higgs to invisible events. In the efficiency calculations, the denominator is defined to be all events passing a relatively tight VBF selection, similar to the 2015 analysis, whereas the numerator is events passing the selection and the related trigger requirements. The $p_T^{\text{miss}} + H_T^{\text{miss}}$ based triggers give sharper turn-on curves and earlier plateaus, and allow one to relax the sub-leading jet p_T selection with a consequent gain in signal acceptance.

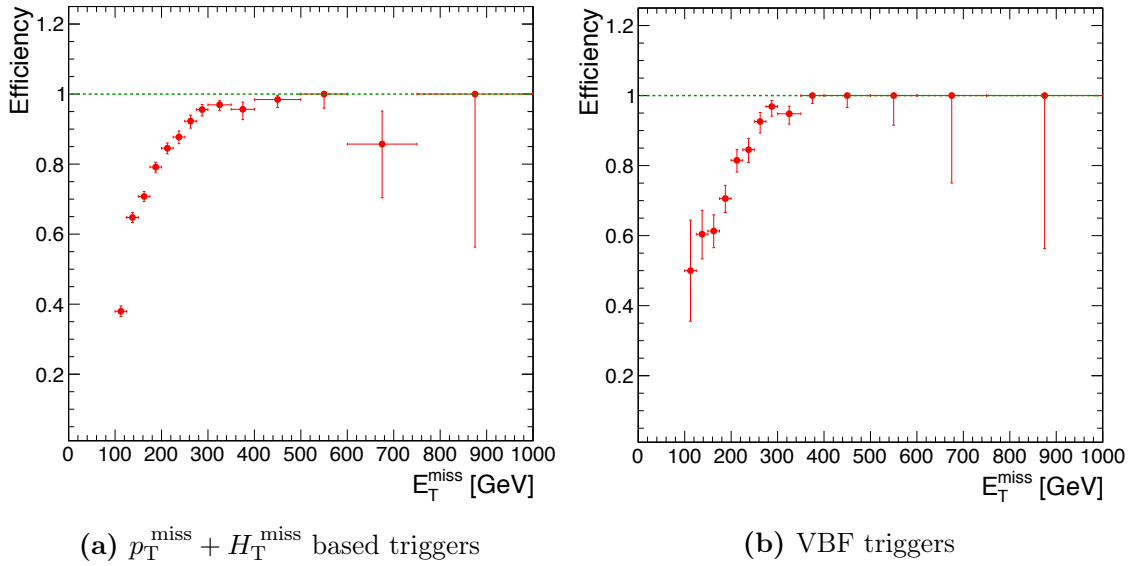


Figure 5.2: The comparison between $p_T^{\text{miss}} + H_T^{\text{miss}}$ (a) and VBF (b) trigger turn-on curves in simulated VBF Higgs to invisible events.

The performance of the $p_T^{\text{miss}} + H_T^{\text{miss}}$ triggers is measured using both double-muon and single-muon events selected with the respective CR requirements from the SINGLEMUON-triggered dataset. These signal triggers are seeded at L1 using a logical OR of available E_T^{miss} -based seeds. Energy deposits in the forward calorimeter, i.e. $|\eta| > 3$, are not included in the computation of the p_T^{miss} . This causes an inefficiency in the tails of the m_{jj} spectrum that is studied categorising events with the leading and sub-leading jets in: barrel-barrel, barrel-forward (or forward-barrel), and forward-forward. Figure 5.3 shows

efficiencies measured in single-muon events in each of these categories. The efficiency is around 100% in the barrel-barrel region, and more than 80% in the barrel-forward region for events with $m_{jj} > 300$ GeV. The L1 seed requirements lead to an efficiency less than 50% in the forward-forward category for events with $m_{jj} > 2000$ GeV. For this reason, events in this category are explicitly removed by imposing a cut in the analysis selection.

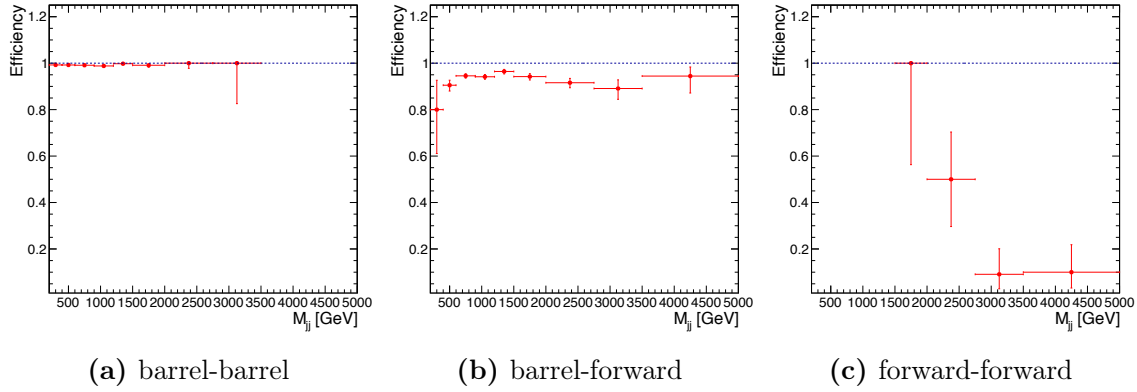


Figure 5.3: The trigger efficiencies of single-muon events selected from the SINGLEMUON-triggered dataset for different jet η categories as a function of m_{jj} .

A measurement of efficiency as a function of H_T^{miss} , computed using jets with $|\eta| < 3.0$, is performed to account for any dependency of the efficiency on the kinematics of the event. The turn-on curves for the $p_T^{\text{miss}} + H_T^{\text{miss}}$ triggers are measured in double-muon and single-muon events selected from the SINGLEMUON-triggered dataset for different jet η categories, with an overall agreement within 1.5 – 2.5%. This is assigned as a systematic uncertainty to all processes in the signal region. For both sets of triggers, the measured efficiencies are applied as an event weight in simulation.

5.2.2 Electron Triggers

The single and double-electron CRs use a logical OR of two electron triggers: the first (ELE27) has tight ID and ISO requirements whereas the second (ELE105) has just a loose ID and no ISO requirement. The ELE27 trigger suffers from some inefficiency for $Z \rightarrow e^+e^-$ events in which the Z boson is boosted. In this configuration, the two electrons can spoil each other's ISO at the trigger level. The ELE105 trigger recovers part of this inefficiency.

The efficiency of the electron triggers is measured in data. For high- p_T electrons, i.e. $p_T > 100$ GeV, the efficiency is around 100% regardless the value of η . For low- p_T

electrons, the efficiency is measured using a tag-and-probe technique. Events are selected requiring one tag electron which fires the single-electron triggers and passes the tight ID requirements. Then the second probe electron is also required to pass the tight selection, and the invariant mass of the double-electron pair is required to be consistent with the Z boson mass. The fraction of events in which the probe electron fires the trigger gives the trigger efficiency, and is measured in bins of p_T and η of the electron. The efficiency is around 20% for electrons with $20 < p_T < 30$ GeV and $|\eta| > 2.1$, whereas it is more than 50% in the remaining region. The measured efficiencies are applied as an event weight in simulation in bins of p_T and η of the electron.

5.3 Reweighting Procedure

Several corrections are applied to the simulated events to improve the data-to-MC agreement. These corrections and the resulting weights are described in the following. Limitations in this procedure are accounted for including systematic uncertainties, as presented in Section 5.6, discussing also the nature and scale of the related discrepancies.

5.3.1 Trigger Efficiency and Pileup Weight

Simulated events are corrected to account for the L1 and HLT efficiency, measured directly in data as detailed in Section 5.2. They are also weighted to account for pileup effects using the recommended value of the total pp cross-section of $69.2 \text{ mb} \pm 4.6\%$ [76].

5.3.2 Lepton and Photon Efficiency and Reconstruction SF

Data-to-MC scale factors (SFs) are derived from efficiencies measured for the electrons, muons, and photons passing ID and ISO criteria in both data and simulation, and then applied to simulated events in the four CRs.

For veto electrons, the weight is calculated as $\frac{1 - \text{SF} \times \varepsilon_{\text{MC}}}{1 - \varepsilon_{\text{MC}}}$, where ε_{MC} is the MC efficiency. For the veto muons, the p_T - η binned MC efficiency is roughly 1, and large effects due to statistical fluctuations are observed when no reconstructed muons are required. Thus, the chosen method is to not apply the veto but to reweight events using the formula $(1 - \text{SF})^n$, where n is the number of veto muons, whose SF is p_T and η dependent. This method is also used for the tau and b-jet vetoes.

5.3.3 B-tagging and Tau Vetos Scale Factors

Events with b-tag jets or tau leptons are not vetoed but reweighted using the method detailed in Section 5.3.2. For tau leptons the SF is uniform at 0.99 ± 0.05 , whereas for b-jets the weight is dependent on the p_T and η of the jet and applied to simulated events as such.

5.3.4 NLO Corrections for V+jets QCD Production

The Z+jets-QCD and W+jets-QCD processes play an extremely important role in the analysis, being the key backgrounds and control samples. The MC samples for these processes are produced at LO in independent H_T bins, and their prediction is significantly improved when NLO corrections are applied. Event-by-event weights (k-factors) are applied to the LO MC samples such that the generator-level distributions of the p_T of the Z or W boson, referred to as p_T^V , match the corresponding NLO prediction. In addition, the LO samples are further reweighted to incorporate the reduction in the cross-section at high p_T^V due to higher-order EW effects. Calculations of Z+jets and W+jets differential cross-sections presented in Refs. [77,78] are used to achieve this NLO-QCD + NLO-EW correction.

Figure 5.4 shows the comparisons of the p_T^V distributions of the base (LO) MC, NLO-QCD, and NLO QCD + EW for an inclusive selection, referred to as mono-jet, i.e. after requiring a high- p_T central jet not overlapping with generator-level leptons and neutrinos.

Other observables, such as leading and sub-leading jet p_T , m_{jj} , and $\Delta\phi_{jj}$, show differences between the LO and NLO V+jets-QCD samples. As a consequence, a comparison between the NLO-QCD and LO p_T^V spectra is carried out in a VBF-like phase space. In particular, the final higher-order corrections are derived after applying the following selection on the generator-level objects:

- leading (sub-leading) jet with $p_T > 80(40)$ GeV and $|\eta| < 4.7$;
- $\eta_{j_1} \cdot \eta_{j_2} < 0$ and at least one jet with $|\eta| < 3$;
- $\min\Delta\phi(j, p_T^V) > 0.5$, that considers the first four jets with $p_T > 30$ GeV and not overlapping with leptons within a cone of $\Delta R < 0.4$;
- $\Delta\phi_{jj} < 1.5$, $m_{jj} > 1300$ GeV, and $\Delta\eta_{jj} > 4.0$.

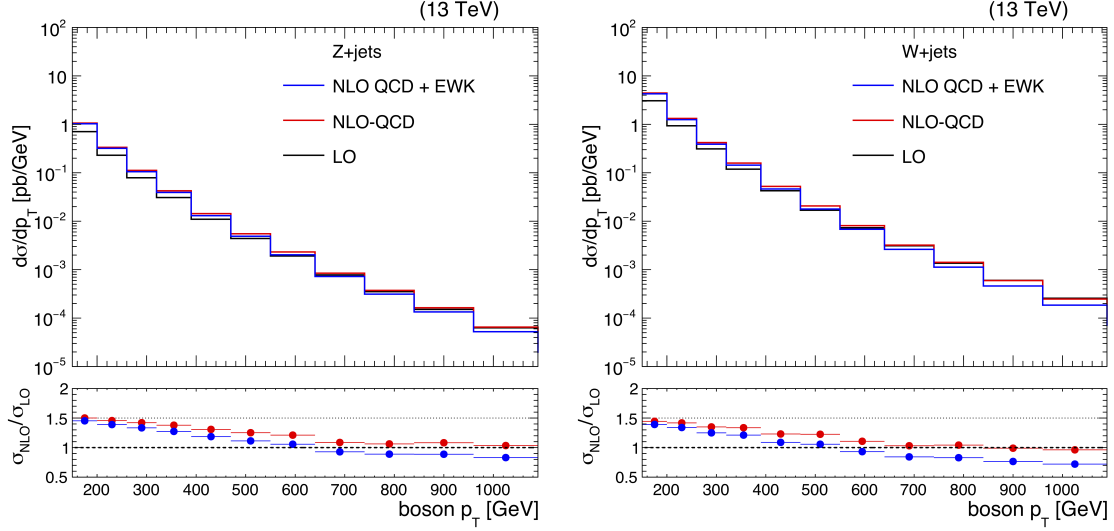


Figure 5.4: The higher-order corrections on the Z (left) and W (right) p_T distributions after an inclusive selection.

A generator-level jet refers to the cluster of all objects coming from the interaction vertex that lie within the jet cone ($\Delta R < 0.4$) using the anti- k_T algorithm. Electrons, photons, and muons are removed from the jet collection after the clustering for the VBF Higgs to invisible analysis. In addition to the previous selection, the leptons are required to have $|\eta| < 2.4$ and the leading lepton to have $p_T > 20$ GeV when computing k-factors for W+jets and $Z(\ell^+\ell^-)$ +jets.

Figure 5.5 shows the evolution of the NLO-QCD k-factors for Z+jets (top left), $Z(\ell^+\ell^-)$ +jets (top right), and W+jets (bottom) as a function of boson p_T , obtained requiring only one central jet in the event (black), applying a loose $\Delta\eta_{jj} > 1$ requirement (blue), and applying the full set of selections (red). Overall, as also indicated by Figure 5.4, the NLO-QCD corrections are larger in the low- p_T^V region. In addition, they are smaller for events with large m_{jj} for both Z+jets and W+jets production modes. This study also demonstrates that a tight $\Delta\eta_{jj}$ requirement has no impact on the k-factors since the differences are compatible within the statistical uncertainty. Instead, requirements on jet p_T , $\Delta\phi_{jj}$, and m_{jj} have a significant impact.

5.3.5 NLO Corrections for V+jets EW Production

The EW production of V+jets starts to dominate at large values of m_{jj} and $\Delta\eta_{jj}$, leading to significant NLO-QCD corrections. For this reason, Z+jets-EW and W+jets-EW

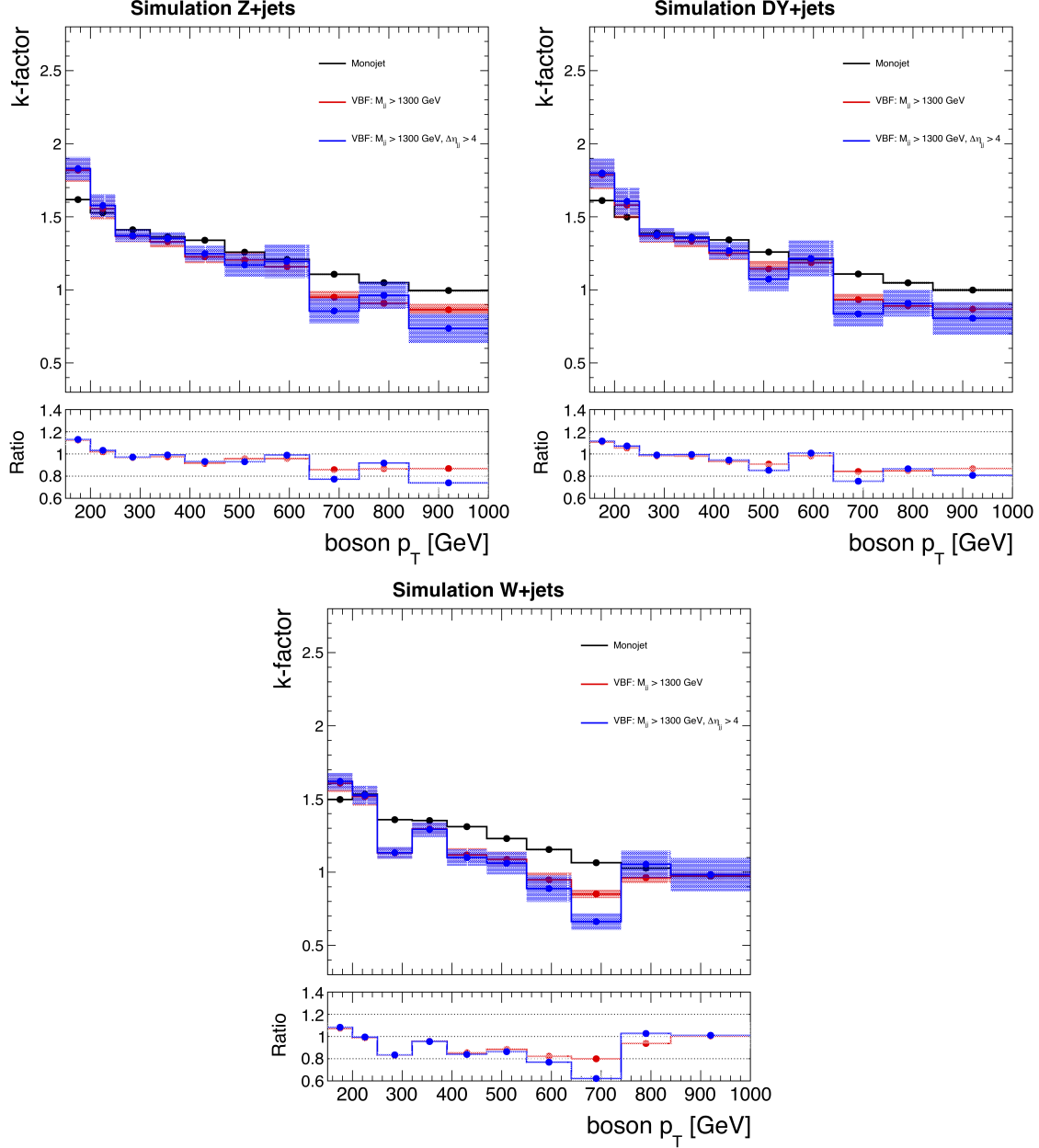


Figure 5.5: The comparison of higher-order QCD corrections on the Z+jets (top left), $Z(\ell^+\ell^-)$ +jets (top right), and W+jets (bottom) processes as a function of boson p_T . The black distribution refers to inclusive NLO k-factors, while corrections obtained after applying the VBF cut-and-count (loose) selection are shown by the red (blue) histogram.

samples are reweighted using additional k-factors derived in VBFNLO [79]. These weights are calculated at parton-level by requiring: two leading partons with $p_T > 80$ and 40 GeV, opposite signs in η , $\Delta\eta > 1.5$, $\Delta\phi < 1.5$, and an invariant mass larger than 250 GeV.

Figure 5.6 shows the NLO-QCD k-factors applied to the EW production of W+jets and Z+jets as a function of m_{jj} and p_T^V .

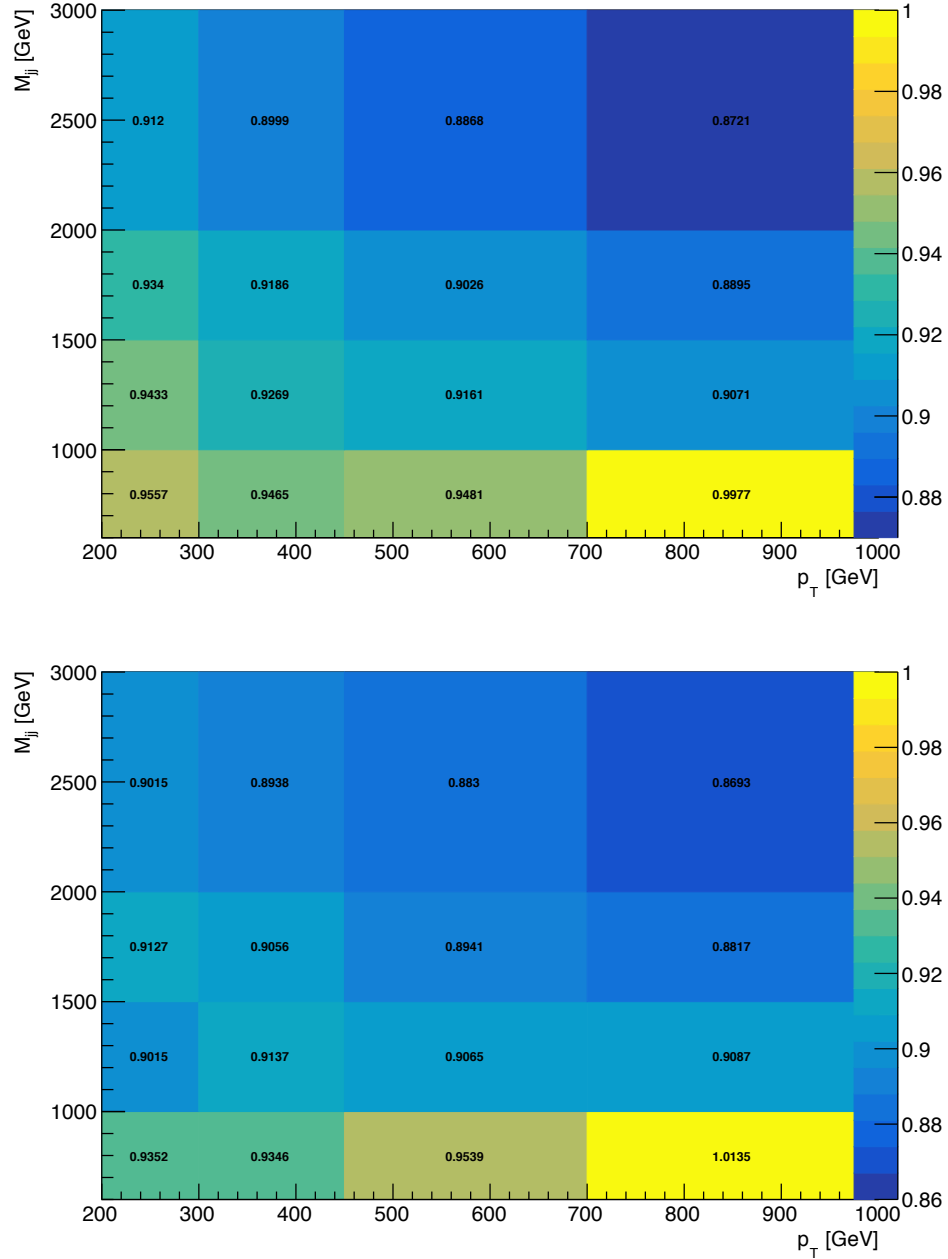


Figure 5.6: The higher-order NLO-QCD corrections on the Z+jets-EW (top) and W+jets-EW (bottom) samples as a function of m_{jj} and p_T^V . The vertical axis on the right of each plot indicates the magnitude of the NLO-QCD k-factors applied.

5.4 Event Selection

As mentioned, the VBF Higgs to invisible analysis described here features two approaches: a cut-based (cut-and-count) and a shape-based. Both these are optimised in terms of the expected 95% CL upper limit on the branching ratio of the SM 125 GeV Higgs boson decaying invisibly, $\mathcal{B}(H \rightarrow \text{inv.})$. More details on the fit procedure and optimisation are given in Section 5.7 and Section 5.8, respectively. An exhaustive presentation of the cut-based approach is provided in the following, whereas only final results and limits for the shape-based method are reported and used in the combination with other production channels, further detailed in Chapter 6.

Table 5.1: The summary of the selection criteria applied in the analysis for both shape-based and cut-based approaches.

Observable	Shape-based	Cut-based	Target Background
Leading, Sub-leading Jet	$p_T^{j_1, j_2} > 80, 40 \text{ GeV}, \eta_{j_1, j_2} < 4.7$		All
$\eta_{j_1} \cdot \eta_{j_2}$	< 0		$Z(\nu\bar{\nu}) + \text{jets}, W(\ell^\pm\nu) + \text{jets}$
MET	$E_{T, \text{no-}\mu}^{\text{miss}} > 250 \text{ GeV}$		QCD-multijet, $t\bar{t}$, $\gamma + \text{jets}$, $V + \text{jets}$
$ E_{T, \text{calorimetric}}^{\text{miss}} - E_{T, \text{PF}}^{\text{miss}} /U$	< 0.5		QCD-multijet
$\min \Delta\phi(E_T^{\text{miss}}, j)$	$> 0.5 \text{ rad}$		QCD-multijet, $\gamma + \text{jets}$
$ \Delta\phi_{jj} $	$< 1.5 \text{ rad}$		$Z(\nu\bar{\nu}) + \text{jets}, W(\ell^\pm\nu) + \text{jets}$
Electron	$N_e = 0$ with $p_T > 10 \text{ GeV}, \eta < 2.5$		$W(\ell^\pm\nu) + \text{jets}$
Muon	$N_\mu = 0$ with $p_T > 10 \text{ GeV}, \eta < 2.4$		$W(\ell^\pm\nu) + \text{jets}$
τ_h Candidate	$N_{\tau_h} = 0$ with $p_T > 18 \text{ GeV}, \eta < 2.3$		$W(\ell^\pm\nu) + \text{jets}$
Photon	$N_\gamma = 0$ with $p_T > 15 \text{ GeV}, \eta < 2.5$		$\gamma + \text{jets}, V\gamma$
b-tagged Jet	$N_{\text{jet}} = 0$ with $p_T > 20 \text{ GeV}, \text{CSV}2 > 0.848$		$t\bar{t}$, single-top
$ \Delta\eta_{jj} $	> 1.0	> 4.0	$Z(\nu\bar{\nu}) + \text{jets}, W(\ell^\pm\nu) + \text{jets}$
m_{jj}	$> 200 \text{ GeV}$	$> 1300 \text{ GeV}$	$Z(\nu\bar{\nu}) + \text{jets}, W(\ell^\pm\nu) + \text{jets}$

The SR selection criteria are driven by several considerations, detailed in the following, and are summarised in Table 5.1 for both approaches. The chosen triggers drive the MET criteria, $E_T^{\text{miss}} > 250 \text{ GeV}$, to be above the 90% efficiency point in Figure 5.2a. The signal final state must not contain leptons, $N_\ell = 0$, where ℓ is an electron, muon, or hadronic tau candidate. As a consequence, $E_{T, \text{no-lep}}^{\text{miss}} = E_T^{\text{miss}} = U$. It is important to use $E_{T, \text{no-lep}}^{\text{miss}}$ for the background estimation method, such that the kinematics of the recoil are then as similar as possible in the SR and CRs. A complete treatment of the four CRs used in the analysis is provided in Section 5.5. Figure 5.7 (right) shows the distribution of $E_{T, \text{no-}\mu}^{\text{miss}}$ for both data and simulated signal and backgrounds after the cut-based selection in the SR. The backgrounds are estimated using MC samples and do

not accurately model the data. The data-driven approach used to estimate the V+jets production aims to improve this data-to-MC agreement, as detailed in Section 5.5.

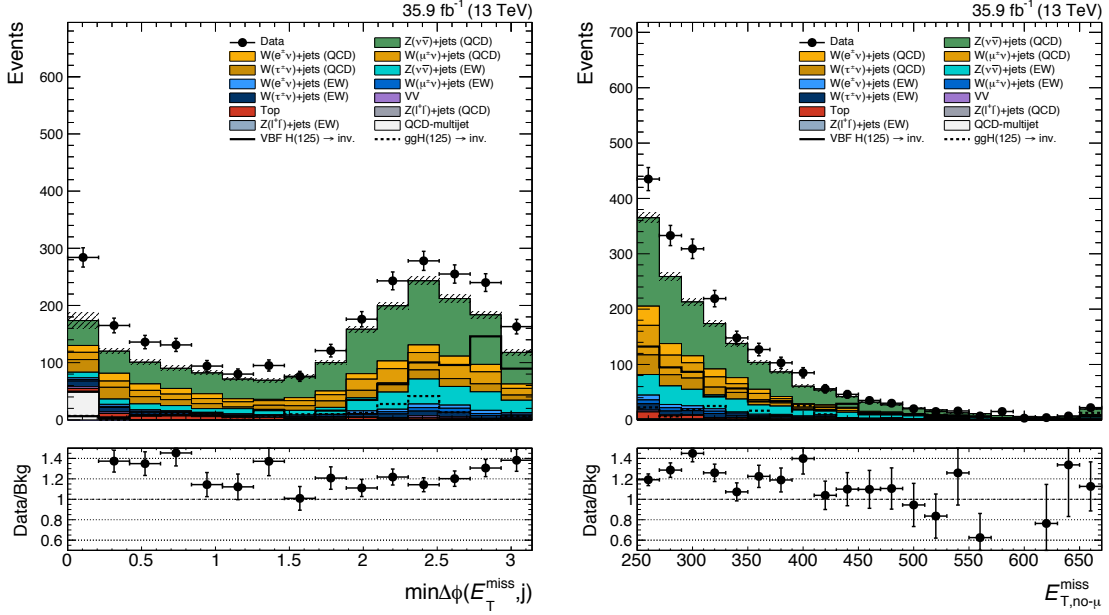


Figure 5.7: The distributions of $\min\Delta\phi(E_T^{\text{miss}}, j)$ (left), obtained by removing its corresponding cut, and $E_{T,\text{no-}\mu}^{\text{miss}}$ (right) for data and simulated signal and backgrounds in the signal region.

A VBF topology of the jets is expected from the signal, i.e. two leading jets in opposite hemispheres of the detector with a large pseudo-rapidity separation and a large invariant mass. The two highest- p_T jets in an event define the VBF-tag pair, requiring the leading (sub-leading) jet to have $p_T > 80(40)$ GeV. An azimuthal separation between them of $\Delta\phi_{jj} < 1.5$, chosen using the optimisation procedure, further reduces the V+jets background. Photons and b-jets must also not be present in the final state, hence the requirements $N_\gamma = 0$ and $N_{\text{jet}} = 0$, respectively. The signal kinematics presented in Section 5.8 further justify the chosen selection of Table 5.1.

In addition to the variables defining the VBF + MET topology, selection criteria are applied to reject backgrounds where the MET arises from jet mis-measurement, i.e. fake MET. The QCD background, QCD-multijet events with fake MET, is reduced by requiring the minimum separation in ϕ between the MET direction and jets with $p_T > 30$ GeV to be $\min\Delta\phi(E_T^{\text{miss}}, j) > 0.5$. This quantity peaks at zero for jets arising from mis-measurement, as shown by the region $\min\Delta\phi(E_T^{\text{miss}}, j) < 0.5$ in Figure 5.7 (left). This distribution is obtained by removing the related cut in the selection criteria, referred to as an ‘N-1’ selection, for the cut-based approach. The prediction of the QCD

background using MC samples suffers from mis-modelling that causes the poor agreement between data and simulation in the low- $\min\Delta\phi(E_T^{\text{miss}}, j)$ region. For this reason, the $\Delta\phi$ extrapolation method detailed in Section 5.5.3 is used to estimate the QCD-multijet background.

Finally, the signal region selection is optimised by calculating the expected limit on $\mathcal{B}(H \rightarrow \text{inv.})$ for a wide range of m_{jj} , $\Delta\phi_{jj}$, and $\Delta\eta_{jj}$ thresholds, as detailed in Section 5.8. This optimisation study requires the introduction of CRs, systematic uncertainties, and the statistical procedure, outlined in the following Sections. Figure 5.8 shows the distributions of m_{jj} and $\Delta\eta_{jj}$ after the optimisation but using MC samples to predict the backgrounds. This causes a poor agreement between data and simulation in the regions $m_{jj} < 1500$ GeV and $4.5 < \Delta\eta_{jj} < 5.5$. The objective of the CRs is to improve this agreement by using a data-driven method to estimate the V+jets background.

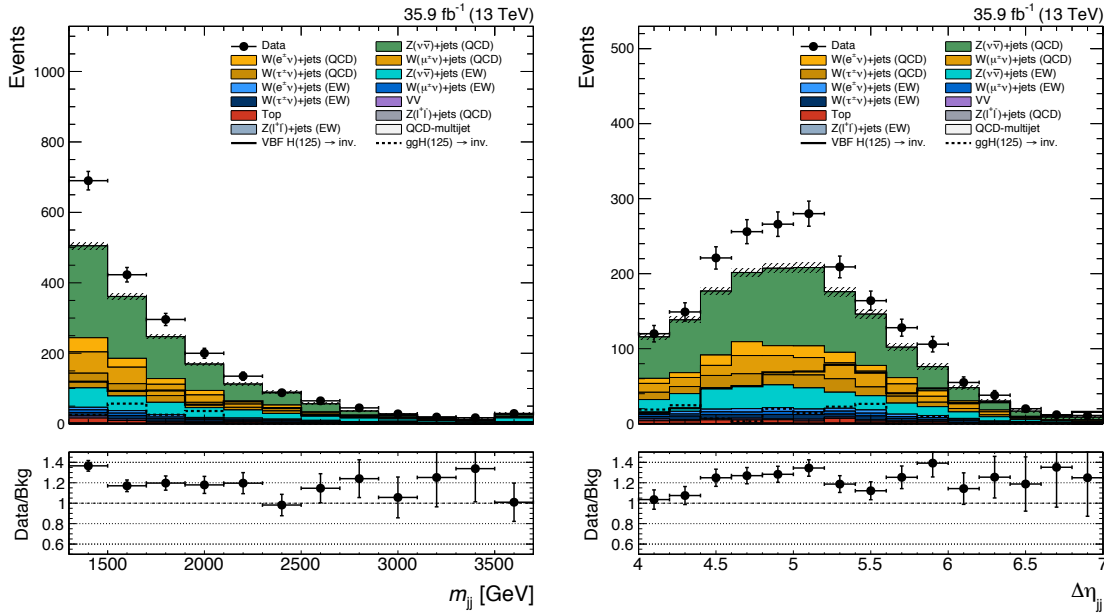


Figure 5.8: The distributions of m_{jj} (left) and $\Delta\eta_{jj}$ (right) for data and simulated signal and backgrounds in the signal region.

Figure 5.9 shows the distributions of p_T and η for the leading (j_1) and sub-leading (j_2) jets for both data and simulated signal and backgrounds after the cut-based selection in the SR. The agreement between data and simulation varies within the range of p_T and η for both jets. In fact, MC samples suffer from mis-modelling that causes a disagreement up to 40%. A data-driven method is used to improve the background estimation, in particular the contribution from V+jets production, as detailed in Section 5.5.

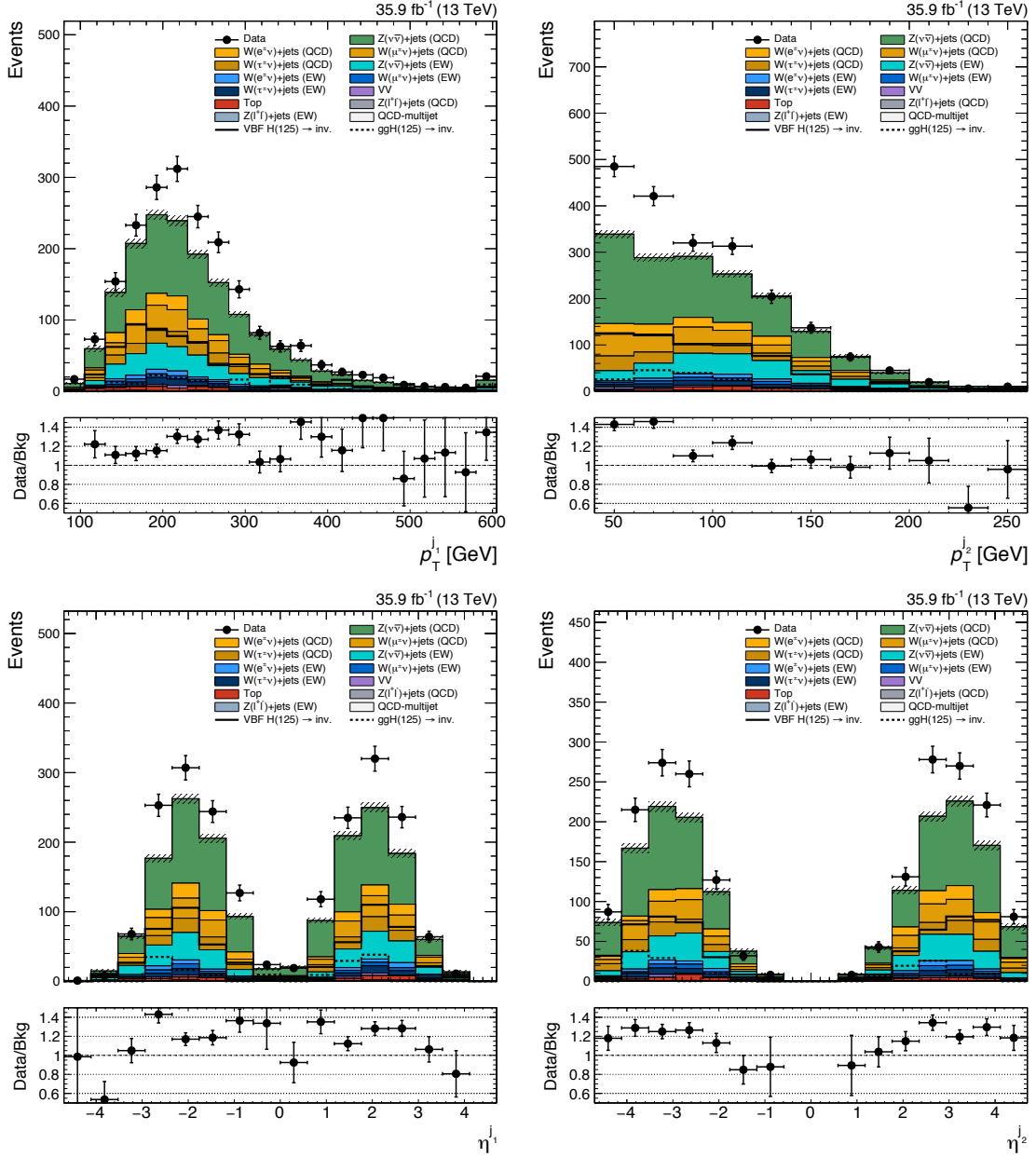


Figure 5.9: The distributions of p_T (top) and η (bottom) for the leading (left) and sub-leading (right) jets for data and simulated signal and backgrounds in the signal region.

5.5 Control Regions

There are several processes with an experimental signature similar to an invisibly decaying Higgs boson produced through the VBF mode, as detailed in Section 5.1.2. The dominant background is V+jets production, with a cross-section around three orders of magnitude higher than the Higgs boson production assuming $\mathcal{B}(H \rightarrow \text{inv.}) = 100\%$. Thus, any uncertainty on V+jets has a huge impact on the analysis. The MET arises from processes such as $Z \rightarrow \nu\bar{\nu}$, or $W \rightarrow \ell^\pm \nu$ where the lepton is outside of the fiducial reconstruction area of the detector. For this reason, a data-driven approach is used and CRs are exploited to estimate the V+jets background. The CRs contain well identified leptons and are used to check data-to-MC agreement, as detailed in Sections 5.5.1 and 5.5.2. They are also used in a simultaneous fit with the SR to constrain the MC to data for the extraction of the final result.

The QCD production of multiple jets (QCD-multijet) has a large cross-section, leading to a significant number of VBF-like jets. The MET from real invisible particles is low except from heavy-flavour decays, but the mis-measurement of these jets can cause a significant fake MET. Additional minor backgrounds are the production of two vector bosons and top-quark decays, both characterised by low cross-sections. The b-jet veto decreases further the contributions from these processes. Simulated events are weighted to match observation by accounting for pileup and trigger efficiency, and corrections are applied for lepton ID efficiency, as detailed in Section 5.3.

5.5.1 Z Control Regions

The background from $Z(\nu\bar{\nu})$ +jets processes is irreducible, but can be controlled by looking at the leptonic decays of the Z boson into e^+e^- or $\mu^+\mu^-$. The kinematics of the events are expected to be very similar, except for biases introduced by the requirements on the leptons. The $Z \rightarrow \mu^+\mu^-$ and $Z \rightarrow e^+e^-$ CRs are defined by removing the lepton veto from the SR selection and instead requiring:

$Z \rightarrow \mu^+\mu^-$ CR: a pair of muons with $p_T > 20, 10 \text{ GeV}$, $|\eta| < 2.4$, $60 < m_{\mu\mu} < 120 \text{ GeV}$
i.e. exactly two loose muons with at least one tight, and no other loose muons;

$Z \rightarrow e^+e^-$ CR: a pair of electrons with $p_T > 40, 10 \text{ GeV}$, $|\eta| < 2.5$, $60 < m_{ee} < 120 \text{ GeV}$
i.e. exactly two veto electrons with at least one tight, and no other veto electrons.

A requirement on the invariant mass of the two leptons is applied to remove the contribution from $\gamma^* \rightarrow \ell^+ \ell^-$ events.

The number of $Z \rightarrow \nu\bar{\nu}$ events in the SR, $N_{\text{expected}}^{\text{SR}}$, is estimated by using the formula:

$$N_{\text{expected}}^{\text{SR}} = \frac{\sigma(Z \rightarrow \nu\bar{\nu})}{\sigma(Z \rightarrow \ell^+ \ell^-)} \cdot \frac{\epsilon^{\text{SR}}}{\epsilon^{\text{CR}}} \cdot (N_{\text{data}}^{\text{CR}} - N_{\text{background}}^{\text{CR}}), \quad (5.1)$$

where σ is the cross-section of the related process, ϵ^{SR} is the efficiency for $Z \rightarrow \nu\bar{\nu}$ events to pass the SR selection, and ϵ^{CR} is the efficiency for $Z \rightarrow \ell^+ \ell^-$ events to pass the related CR selection. The number of observed events in the CR is given by $N_{\text{data}}^{\text{CR}}$, whereas $N_{\text{background}}^{\text{CR}}$ is the number of simulated MC events from other minor backgrounds in the CR. Both QCD and EW processes can produce Z bosons and have different cross-sections. The efficiencies ϵ^{SR} and ϵ^{CR} are cross-section weighted averages of the efficiencies for both QCD and EW productions in the SR and in the related CR, respectively, calculated as:

$$\epsilon^{\text{SR}} = \frac{\sigma(Z \rightarrow \nu\bar{\nu}, \text{EW}) \cdot \frac{N_{\text{MC}}^{\text{SR}}(\text{EW})}{N_{\text{gen}}(\text{M}_Z, \text{EW})} + \sigma(Z \rightarrow \nu\bar{\nu}, \text{QCD}) \cdot \frac{N_{\text{MC}}^{\text{SR}}(\text{QCD})}{N_{\text{gen}}(\text{M}_Z, \text{QCD})}}{\sigma(Z \rightarrow \nu\bar{\nu}, \text{EW}) + \sigma(Z \rightarrow \nu\bar{\nu}, \text{QCD})}, \quad (5.2)$$

$$\epsilon^{\text{CR}} = \frac{\sigma(Z \rightarrow \ell^+ \ell^-, \text{EW}) \cdot \frac{N_{\text{MC}}^{\text{CR}}(\text{EW})}{N_{\text{gen}}(\text{EW})} + \sigma(Z \rightarrow \ell^+ \ell^-, \text{QCD}) \cdot \frac{N_{\text{MC}}^{\text{CR}}(\text{QCD})}{N_{\text{gen}}(\text{QCD})}}{\sigma(Z \rightarrow \ell^+ \ell^-, \text{EW}) + \sigma(Z \rightarrow \ell^+ \ell^-, \text{QCD})},$$

where EW and QCD refer to the type of production of the Z boson, $N_{\text{MC}}^{\text{SR}}$ and $N_{\text{MC}}^{\text{CR}}$ are the numbers of MC events predicted to be in the SR and CR, respectively, and N_{gen} is the number of events of the MC Z+jets process at generator-level.

Figure 5.10 (Figure 5.11) shows the distributions of p_{T} and η for the leading and sub-leading jets, and invariant mass and p_{T} for the muon (electron) pair for both data and simulated signal and backgrounds in the $Z \rightarrow \mu^+ \mu^-$ ($Z \rightarrow e^+ e^-$) CR. The data reasonably agree with simulation given the high statistical uncertainty from the limited amount of data in these regions. The kinematics of the jets are vital to show that a VBF double-jet pair is tagged, whereas the distributions of the invariant mass and p_{T} for the muon (electron) pair ensure that the leptons are produced from Z boson decays. The distributions of other kinematic variables in the double-lepton CRs are shown in Appendix A. The distributions of p_{T} and η for the leading and sub-leading leptons provide topological information on the Z+jets processes. The distribution of m_{jj} is used during the fit in the shape-based approach, and shows a reasonable agreement between data and MC.

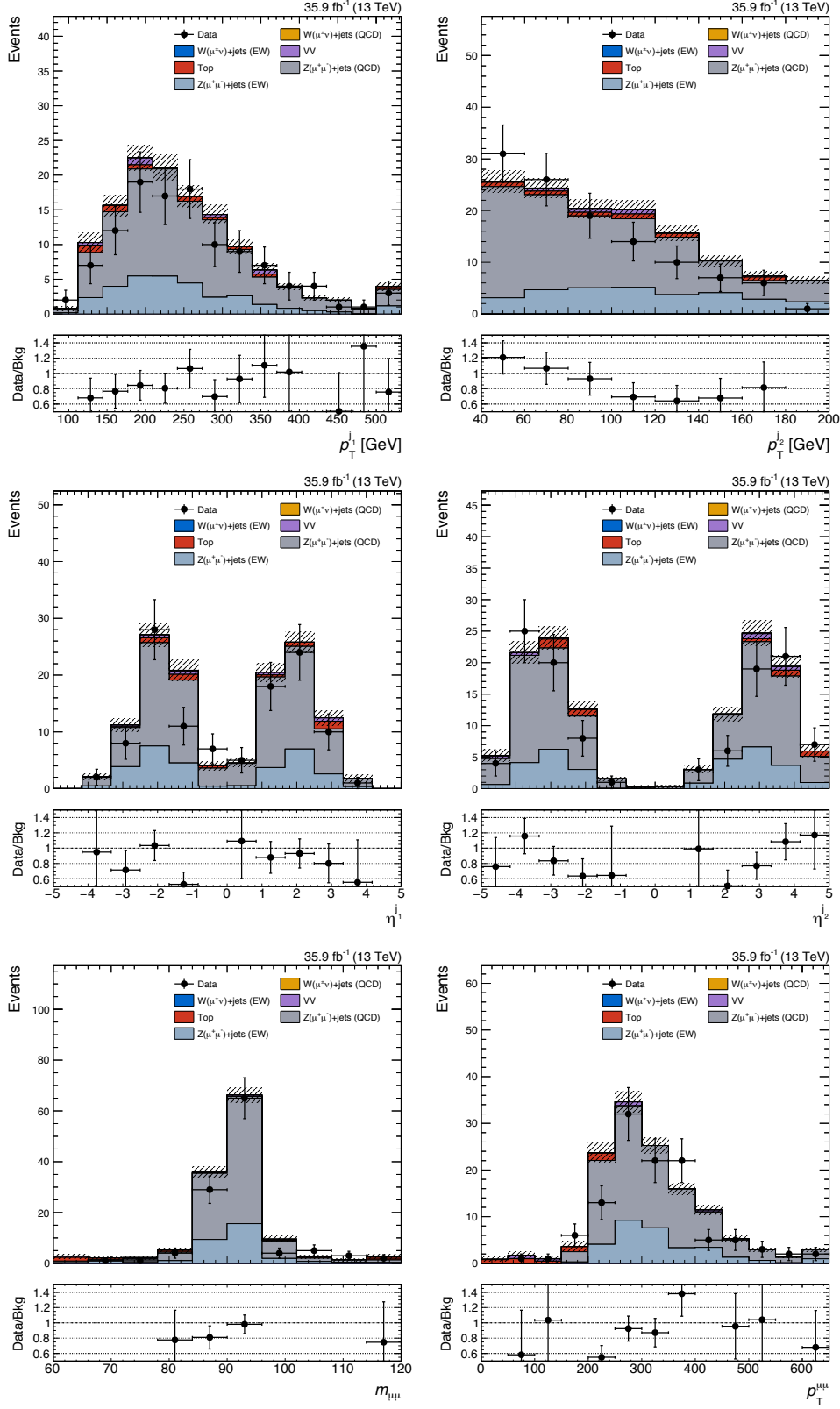


Figure 5.10: The distributions of p_T (top) and η (central) for the leading (left) and sub-leading (right) jets, and invariant mass (bottom left) and p_T (bottom right) for the muon pair for data and simulated signal and backgrounds in the $Z \rightarrow \mu^+ \mu^-$ CR.

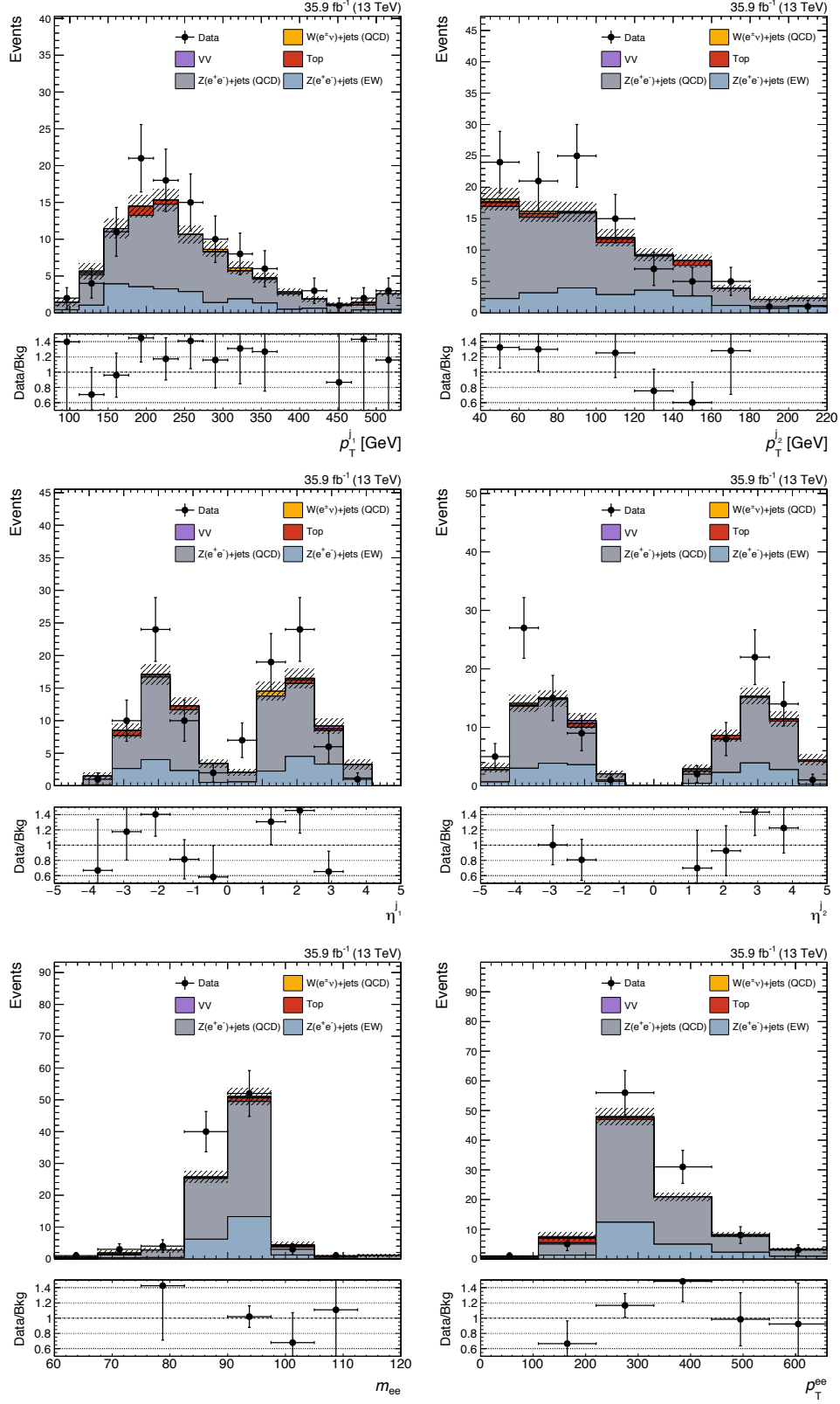


Figure 5.11: The distributions of p_T (top) and η (central) for the leading (left) and sub-leading (right) jets, and invariant mass (bottom left) and p_T (bottom right) for the electron pair for data and simulated signal and backgrounds in the $Z \rightarrow e^+e^-$ CR.

5.5.2 W Control Regions

The $W \rightarrow \ell^\pm \nu$ backgrounds, where the W boson decays to a lepton, i.e. $e^\pm$ or $\mu^\pm$, and a neutrino of the same flavour, are estimated using single-lepton events. The kinematics of these events can be different between the SR and CR. The lepton from the W boson is required to be within the detector acceptance and to have a large p_T in the CR, whereas it is outside of the η acceptance or has low p_T in the SR. This difference is accounted for in the transfer factors from the CRs to the SR, and systematic uncertainties on the ratios between the SR and CRs are included, as documented in Section 5.6.

The $W \rightarrow \mu^\pm \nu$ and $W \rightarrow e^\pm \nu$ CRs are defined by removing the lepton veto from the SR selection and instead requiring:

$W \rightarrow \mu^\pm \nu$ CR: a single muon with $p_T > 20$ GeV and $|\eta| < 2.4$
i.e. exactly one tight muon and no other veto muons;

$W \rightarrow e^\pm \nu$ CR: a single electron with $p_T > 40$ GeV and $|\eta| < 2.5$
i.e. exactly one tight electron and no other veto electrons.

The transverse mass (m_T) of the lepton+MET system is defined as:

$$m_T = \sqrt{2p_T^{\text{miss}} p_T^\ell (1 - \cos \Delta\phi)} , \quad (5.3)$$

where $\Delta\phi$ is the angle between p_T^{miss} and lepton p_T in the transverse plane. To reduce the contamination from QCD-multijet events in these regions, a selection on $m_T < 160$ GeV is required. In addition, a $\text{MET} > 60$ GeV cut is included in the $W \rightarrow e^\pm \nu$ CR to further reduce the QCD-multijet contribution as shown in Figure 4.2 (right). The number of $W \rightarrow \ell^\pm \nu$ events in the SR is estimated by using the procedure detailed in Section 5.5.1.

Figure 5.12 (Figure 5.13) shows the distributions of p_T and η for the leading and sub-leading jets, transverse mass of the muon(electron)+MET system, and $\min\Delta\phi(E_T^{\text{miss}}, j)$ for both data and simulated signal and backgrounds in the $W \rightarrow \mu^\pm \nu$ ($W \rightarrow e^\pm \nu$) CR. The data are overall consistent with simulation. The kinematics of the jets show that a VBF double-jet pair is tagged, whereas the distribution of $\min\Delta\phi(E_T^{\text{miss}}, j)$ ensures that $E_{T,\text{no-lep}}^{\text{miss}}$ recoils against the jets. The variable m_T replaces the invariant mass observable since the neutrino produced with the lepton is not detected. The distributions of other kinematic variables in the single-lepton CRs are shown in Appendix B. The agreement between the data and simulation is more evident in comparison with the double-lepton CRs since the single-lepton CRs have a higher statistical power.

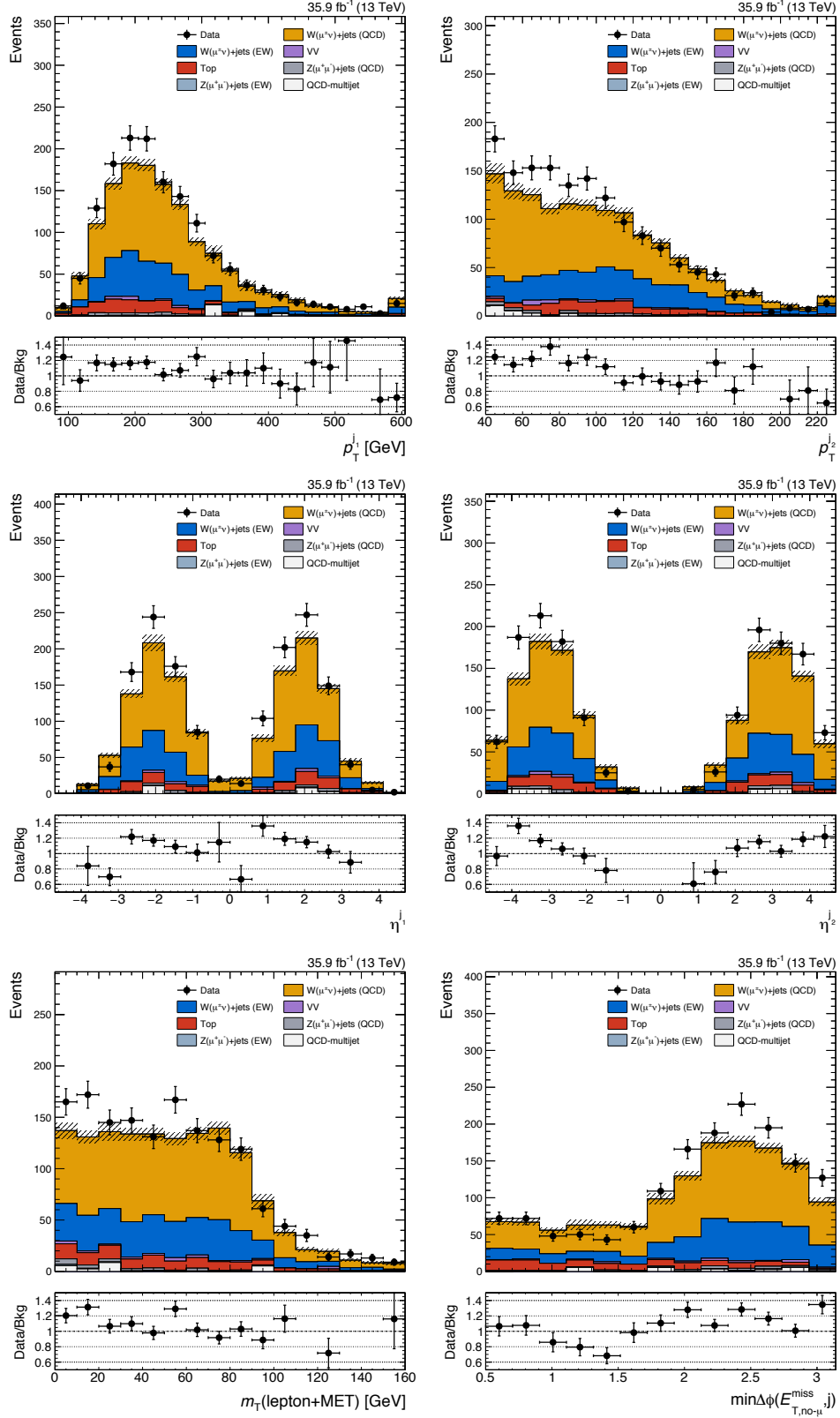


Figure 5.12: The distributions of p_T (top) and η (central) for the leading (left) and sub-leading (right) jets, transverse mass of the muon+MET system (bottom left), and $\min \Delta\phi(E_T^{\text{miss}}, j)$ (bottom right) for data and simulated signal and backgrounds in the $W \rightarrow \mu^\pm \nu$ CR.

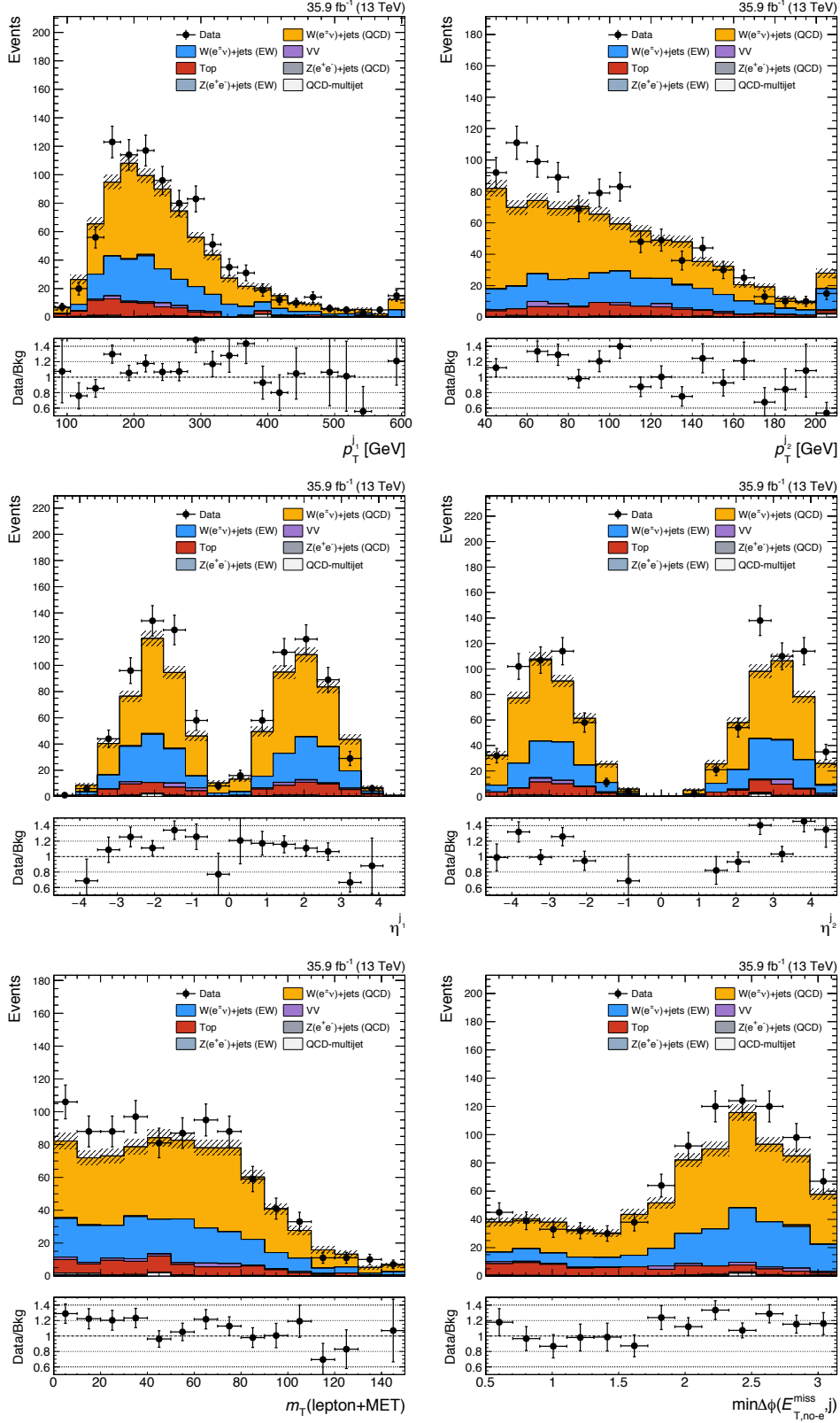


Figure 5.13: The distributions of p_T (top) and η (central) for the leading (left) and sub-leading (right) jets, transverse mass of the electron+MET system (bottom left), and $\min \Delta\phi(E_T^{\text{miss}}, j)$ (bottom right) for data and simulated signal and backgrounds in the $W \rightarrow e^+ \nu$ CR.

5.5.3 The QCD-multijet Background

QCD-multijet production typically features events that are well-balanced in the transverse plane, i.e. $\Delta\phi$ between E_T^{miss} and the jets tend to assume values towards π . This balance can be broken by mis-measurement of the energy of the jets, un-instrumented or non-functional detector regions, or neutrinos from semi-leptonic decays of heavy-flavour mesons. Although such effects are accounted for during the calibration, the very high QCD production cross-section possibly leads to a significant QCD background contribution in $E_T^{\text{miss}} + \text{jets}$ searches.

Control regions in data are defined to predict and extrapolate the QCD contribution into the SR. In fact, QCD MC samples suffer from a limited accuracy in simulating these processes, as shown in Figure 5.7 (left). A mis-measurement of the jet energy leads to higher MET and brings $\vec{p}_T^{\text{miss}}$ closer to the $\vec{p}_T$ of the mis-measured jet. The $\min\Delta\phi(E_T^{\text{miss}}, j)$ distribution in QCD-multijet events depends on the MET in the event, and this dependence translates into a fall-off of the ratio r as a function of MET, where r is defined as:

$$r = \frac{\min\Delta\phi(E_T^{\text{miss}}, j) > 0.5}{\min\Delta\phi(E_T^{\text{miss}}, j) < 0.5} \quad (5.4)$$

that is used to predict the contribution of QCD in the high $\Delta\phi(\vec{p}_T^{\text{miss}}, \vec{p}_T^{\text{jet}})$ region from events with low $\Delta\phi(\vec{p}_T^{\text{miss}}, \vec{p}_T^{\text{jet}})$.

The $\Delta\phi$ extrapolation method and the closure test used to estimate the QCD background contribution in the SR are performed by defining the following regions after the full VBF selection:

Region-A: $m_{jj} > 300 \text{ GeV}$ and $\Delta\eta_{jj} > 1$;

Region-B: $m_{jj} > 300 \text{ GeV}$, $\Delta\eta_{jj} > 1$, and $\Delta\phi(\vec{p}_T^{\text{miss}}, \vec{p}_T^{\text{jet}}) < 0.5$;

Region-C: $m_{jj} > 300 \text{ GeV}$, $\Delta\eta_{jj} > 1$, $\Delta\phi(\vec{p}_T^{\text{miss}}, \vec{p}_T^{\text{jet}}) < 0.5$, and $\Delta\eta_{jj} > 4$.

The selection $\Delta\phi(\vec{p}_T^{\text{miss}}, \vec{p}_T^{\text{jet}}) < 0.5$ is required to enhance the QCD contribution, whereas the m_{jj} and $\Delta\eta_{jj}$ criteria are loosened with respect to the VBF selection to ensure a higher number of events is used during the procedure. An exponential function is fit to the m_{jj} spectrum of each of the QCD MC samples in Region-A and Region-B. Figure 5.14 shows the fitted functions, for which there is a sufficient agreement given the large uncertainties of the MC samples.

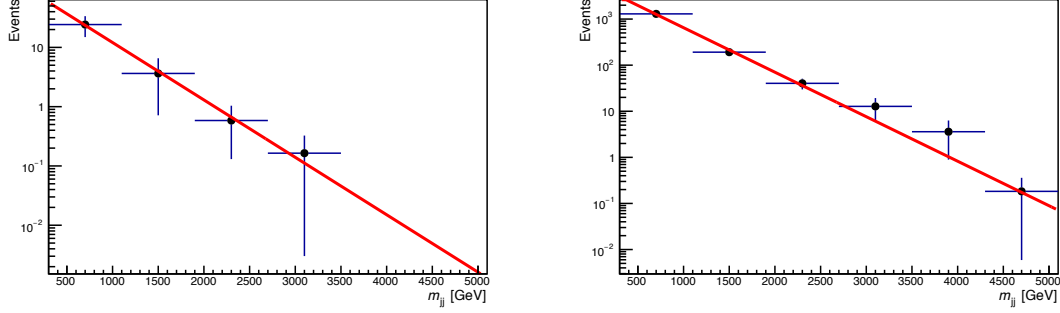


Figure 5.14: The distribution of m_{jj} for QCD-multijet simulated events in Region-A (left) and Region-B (right). The points show the yields in each bin from simulation, while the red lines shows the fitted exponential functions.

The rate of QCD events in each region above a given m_{jj} threshold is estimated by integrating the function above that threshold. Figure 5.15 (left and centre) shows the expected QCD yield extrapolated from the fit functions.

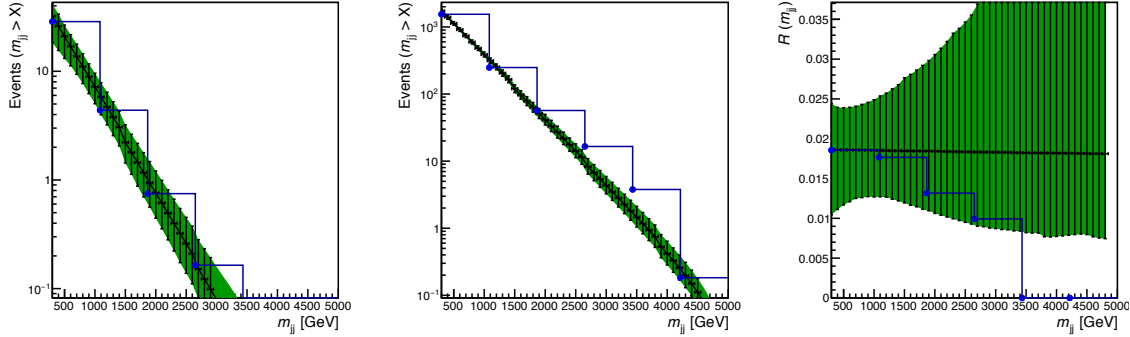


Figure 5.15: The expected yield of QCD-multijet above a given m_{jj} threshold in Region-A (left) and Region-B (centre). The blue points show the prediction from simulation, whereas the black lines refer to the prediction using the fitted exponential functions. The ratio $R(m_{jj})$ as a function of the threshold (right). The uncertainties represent the statistical uncertainty from the fit to simulation.

The ratio $R(m_{jj})$ is defined as the ratio of the extrapolated QCD-multijet yields:

$$R(X) = \frac{\int_X^{+\infty} f_A(m_{jj})}{\int_X^{+\infty} f_B(m_{jj})}, \quad (5.5)$$

where X is the threshold on m_{jj} , and f_A and f_B are the two exponential fits in Region-A and Region-B, respectively. Figure 5.15 (right) shows the value of R as a function of the m_{jj} cut, that for the VBF selection is 1300 GeV as presented in Section 5.8.

The resulting value of R is $0.019^{+0.008}_{-0.006}$, where the uncertainty is calculated using pseudo-data, or toys. For each toy, the parameters of the exponential fits are randomised according to the covariance matrices from the initial fits, and R is recalculated using the new values of the parameters. The total uncertainty on R is then calculated using the root mean square of the results from the multiple toys. Further details on the statistical procedure can be found in Section 5.7.

Region-C is used to estimate the contribution of the QCD-multijet background in the SR by counting events in data passing the $m_{jj} > 1300$ GeV requirement (N_{obs}). The expected contribution from V+jets, $t\bar{t}$, and VV backgrounds ($N_{\text{background}}$) is then subtracted, and the result is scaled by the factor R . The values of $N_{\text{obs}} = 533$ and $N_{\text{bkg}} = 338.9$ yield a predicted QCD-multijet rate of 3.3 ± 1.4 where the uncertainty is taken as 42% from the fit. Since this estimate assumes that the value of R is independent of the choice of $\Delta\eta_{jj}$, a further uncertainty of 50% is included in the QCD-multijet estimate. However, this additional uncertainty does not have a significant impact on the analysis since the QCD background is very small, compared to the other backgrounds.

5.6 Systematic Uncertainties

Systematic uncertainties are included in the analysis to account for limitations in the reweighting procedure, such as the efficiency of reconstruction for leptons. All sources are expected to be similar in the SR and CRs by construction of these regions, and thus to cancel out on the ratios between the SR and CRs. The systematic uncertainties presented in the following Sections account for possible remaining effects, and they are measured using data or estimated from simulation as part of this analysis, or provided by the CMS POGs.

Experimental and theoretical uncertainties on the V+jets processes impact significantly the analysis through the ratios of their cross-sections. Thus, it is important to estimate and reduce them to optimise the results. Furthermore, systematic uncertainties on minor backgrounds and Higgs to invisible signals are included in the analysis and described in Sections 5.6.3 and 5.6.4.

5.6.1 Experimental Uncertainties on V+jets

The kinematic distributions of several variables used in the analysis are compared to understand whether experimental uncertainties cancel between SR and CRs, as presented in the following. Table 5.2 summarises the experimental uncertainties included in the transfer factors when yields are translated from a CR into the SR.

The lepton kinematic selections in the CRs have a direct impact on the boson p_T and jet part, leading to possible remaining effects for sources of uncertainty related to jets. The systematic uncertainties accounting for the jet energy resolution (JER) do not affect the transfer factors since the hadronic content of the event is almost identical between the SR and CRs. Figure 5.16 shows the distribution of p_T for the vector boson for different processes, confirming that jet energy scale (JES) systematic uncertainties do not always cancel perfectly between SR and CRs. For the Z CRs, a large overlap in the phase-space analysed is present when comparing $Z \rightarrow \nu\bar{\nu}$ and $Z \rightarrow \ell^+\ell^-$ processes. This is caused by the Z boson decaying into high- p_T products in the central region, where the lepton reconstruction is efficient. Thus, JES uncertainties cancel within the statistical precision on the ratio of Z+jets in the SR to Z+jets in CRs, and on the ratio of Z+jets to W+jets in the SR. Therefore, the remaining effect is considered negligible and no systematic uncertainty is included in the analysis. For the W CRs, the phase-spaces analysed are mutually exclusive. The CRs feature a high- p_T central lepton, whereas there are low- p_T or forward leptons in the SR. Furthermore, electrons in the forward region are reconstructed as jets and affected by the selection differently than muons. A 2% systematic uncertainty is included on the ratios between the SR and W CRs to account for these effects.

Systematic uncertainties from 0.5% to 5% account for efficiencies of reconstruction, ID, ISO, and veto for leptons. They are provided by the CMS POGs and included on the ratios between the SR and CRs. The lepton veto is the only experimental uncertainty affecting the Z/W ratio in the SR.

A 1(2)% uncertainty is included in the analysis to account for the electron (MET) trigger efficiency. This is measured using data and is further detailed in Section 5.2.

Table 5.2: The experimental uncertainties affecting transfer factors used in the analysis to estimate V+jets backgrounds in the SR. In the case of lepton vetoes, uncertainties are applied to the appropriate process in each region.

Source of Uncertainty	Process	Uncertainty
Pileup	All	0.1 – 2%
Electron Trigger	$W_{\text{SR}}/W_{e\nu}, Z_{\nu\nu}/Z_{ee}$	1%
MET Trigger	$W_{\text{SR}}/W_{\text{CR}}, Z_{\nu\nu}/Z_{\text{CR}}, Z/W$	2%
Muon Reconstruction Efficiency	$W_{\text{SR}}/W_{\mu\nu}, Z_{\nu\nu}/Z_{\mu\mu}$	1% (per lepton)
Muon ID Efficiency	$W_{\text{SR}}/W_{\mu\nu}, Z_{\nu\nu}/Z_{\mu\mu}$	1% (per lepton)
Muon ISO Efficiency	$W_{\text{SR}}/W_{\mu\nu}, Z_{\nu\nu}/Z_{\mu\mu}$	0.5% (per lepton)
Electron Reconstruction Efficiency	$W_{\text{SR}}/W_{e\nu}, Z_{\nu\nu}/Z_{ee}$	1% (per lepton)
Electron ID/ISO Efficiency	$W_{\text{SR}}/W_{e\nu}, Z_{\nu\nu}/Z_{ee}$	1.5% (per lepton)
Lepton Veto	W_{SR}	3.1% (QCD), 5.0% (EW)
JES	$W_{\text{SR}}/W_{\text{CR}}$	2%

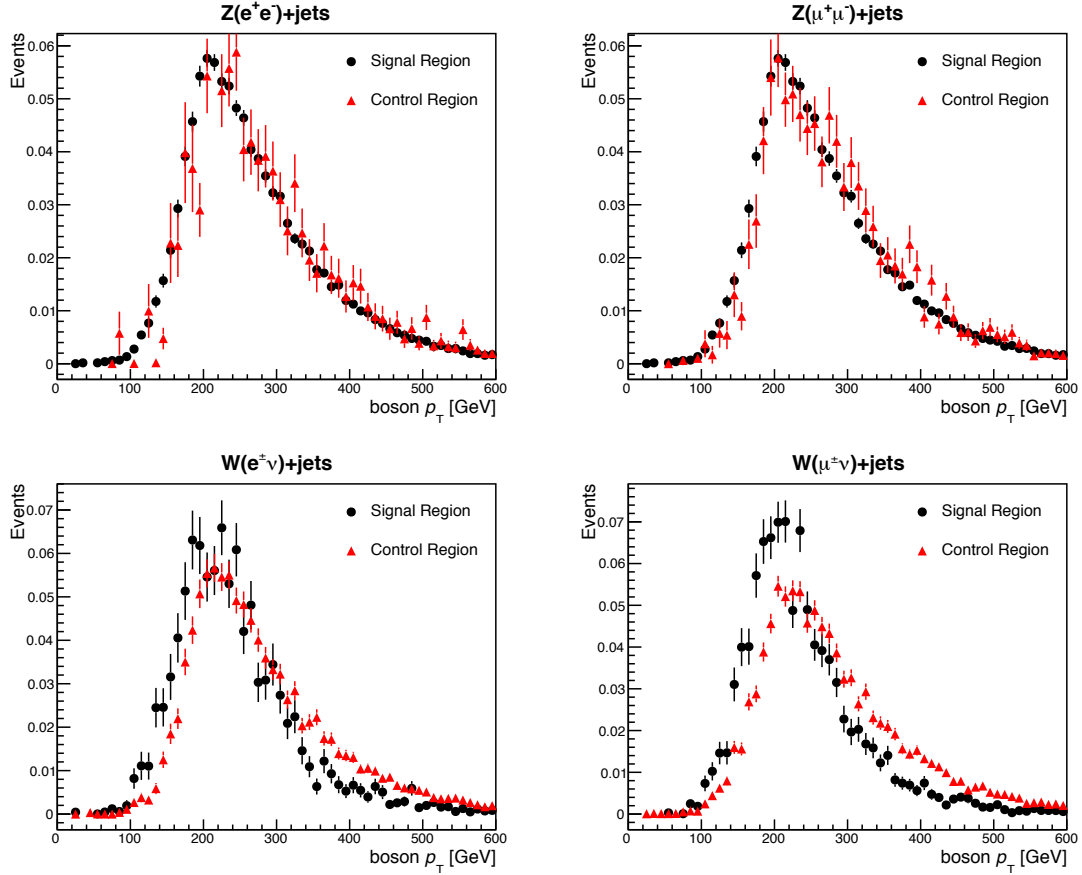


Figure 5.16: The distribution of p_T^V for Z (top) and W (bottom) simulated samples in the SR and CRs.

Considering the effect of pileup, a weight is applied event-by-event to ensure the distribution of the number of vertices in MC reproduces the data measurement for the full data-taking period. The uncertainty on this weight is calculated by varying the minimum bias cross-section by $\pm 4.6\%$ [76] and recalculating the data distribution. The pileup also affects the lepton ISO since the probability to have activity around the lepton increases as a function of pileup. As a result, the efficiency to select a lepton passing ID and ISO criteria in the CRs decreases for high pileup, whereas the number of events passing the veto requirement in the SR increases. Figure 5.17 shows the average numbers of vertices in the SR and CRs. The SR selects events with a slightly higher number of vertices compared to the CRs. This effect is accounted for by uncertainties from 0.1% to 2% depending on both the process and region under consideration.

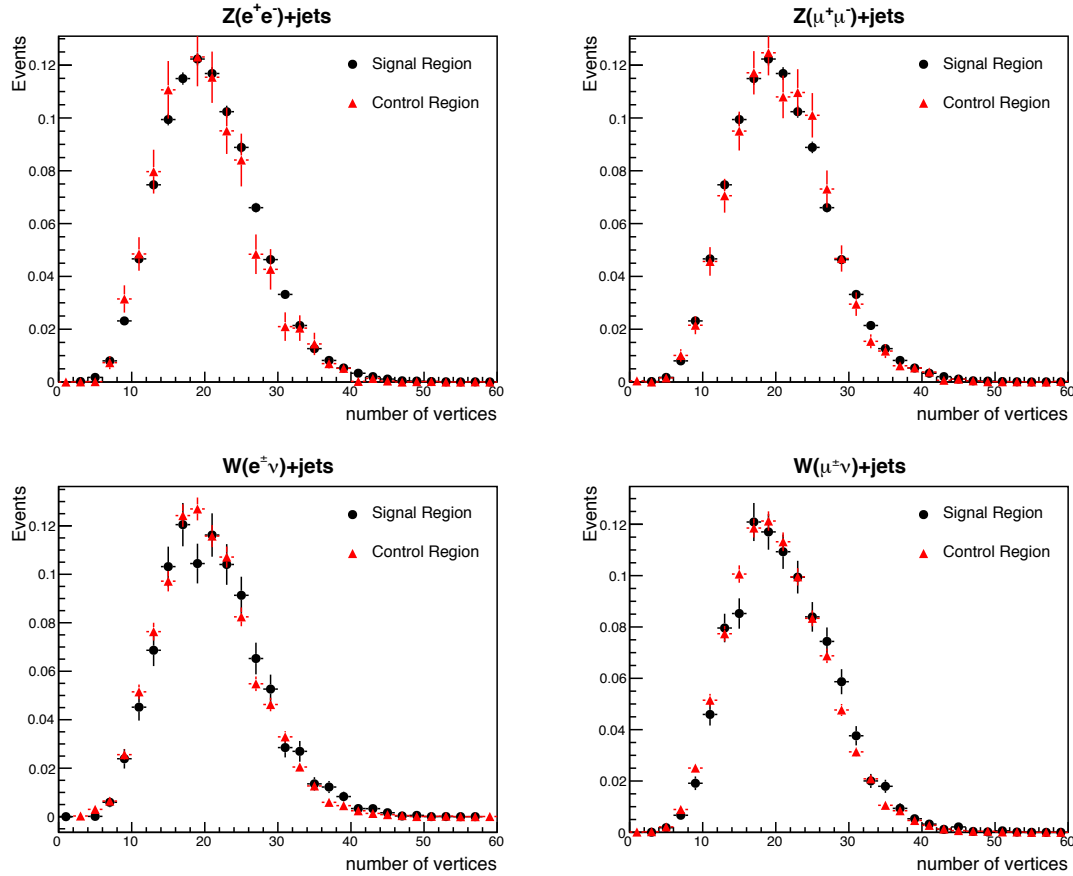


Figure 5.17: The distribution of number of vertices for Z (top) and W (bottom) simulated samples in the SR and CRs.

5.6.2 Theory Uncertainties on V+jets

The LO and NLO predictions for the W+jets and Z+jets processes are compared after applying the full VBF kinematic selection, using generator-level quantities. This allows the derivation of a theoretical systematic uncertainty on the $\frac{Z(\nu\bar{\nu})+\text{jets}}{W(\ell^\pm\nu)+\text{jets}}$ ratio at $\approx 12.5\%$. The systematic uncertainty covers contributions to the p_T^V distribution from the choice of the parton distribution functions (PDFs), and QCD-EW higher-order corrections to the ratio which are not included in the calculation.

A cross-check on the total systematic uncertainty assigned to the QCD Z/W ratio is performed by studying the variations of the ratio while tightening VBF-like selections. Results are shown in Figure 5.18 as a function of p_T^V . The grey band represents the total uncertainty assigned to the ratio and covers the change in the Z/W ratio from a simple two-jet selection up to the tight VBF criteria defined in Section 5.3.4.

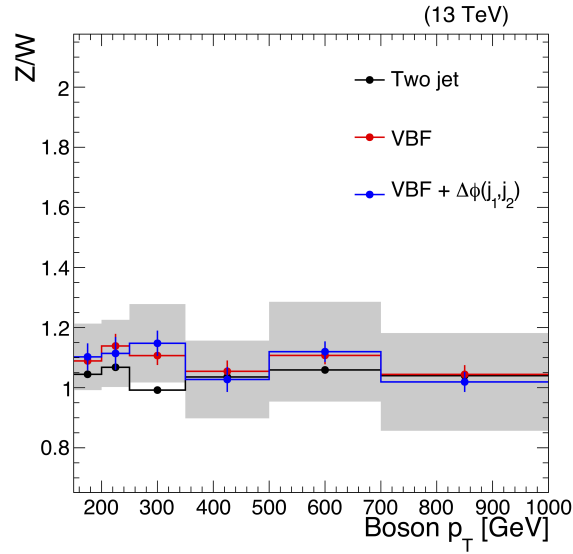


Figure 5.18: The Z/W ratio obtained from simulation as a function of p_T^V for several phase space selections. The grey band represents the total systematic uncertainty from theory.

These uncertainties are derived exclusively for QCD V+jets production. However, they are also applied as a systematic uncertainty to the EW V+jets processes and, specifically, to the EW Z/W ratio. This should be considered as a conservative approach since higher-order corrections for purely EW processes tend to be smaller than those for QCD effects. The higher-order uncertainties on the EW processes are treated as uncorrelated with respect to the QCD V+jets uncertainties.

Finally, the Z/W ratio is not sensitive to NLO corrections detailed in Sections 5.3.4 and 5.3.5, since the k-factors for the W+jets and Z+jets processes are similar across the kinematic ranges, such as for m_{jj} and p_T^V .

5.6.3 Systematic Uncertainties on Minor Backgrounds

The contributions from minor backgrounds, such as single-top, $t\bar{t}$, and VV, are predicted using MC samples, whereas the QCD-multijet background is estimated from data, as detailed in Section 5.5.3. The systematic uncertainties on these backgrounds are summarised in Table 5.3 per source of uncertainty, and are detailed in the following.

Table 5.3: The systematic uncertainties on minor backgrounds populating the SR and CRs. In addition, all simulated samples include statistical uncertainties not shown here.

Source of Uncertainty	Process	Uncertainty
Luminosity	All except for QCD-multijet background	2.5%
Electron Trigger	MC backgrounds in $W \rightarrow e^\pm \nu$ and $Z \rightarrow e^+ e^-$ CRs	$\approx 1\%$
MET Trigger	MC backgrounds in SR, $W \rightarrow \mu^\pm \nu$ and $Z \rightarrow \mu^+ \mu^-$ CRs	$\approx 2\%$
JES and JER	MC backgrounds and signal in SR and CRs	$\approx 5\%$
Muon Reconstruction Efficiency	MC backgrounds in $W \rightarrow \mu^\pm \nu$ and $Z \rightarrow \mu^+ \mu^-$ CRs	$\approx 1\%$
Muon ID/ISO Efficiency	MC backgrounds in $W \rightarrow \mu^\pm \nu$ and $Z \rightarrow \mu^+ \mu^-$ CRs	$\approx 2\%$
Electron Reconstruction Efficiency	MC backgrounds in $W \rightarrow e^\pm \nu$ and $Z \rightarrow e^+ e^-$ CRs	$\approx 1\%$
Electron ID/ISO Efficiency	MC backgrounds in $W \rightarrow e^\pm \nu$ and $Z \rightarrow e^+ e^-$ CRs	$\approx 2\%$
Lepton Veto	MC backgrounds in SR like $Z \rightarrow \ell^+ \ell^-$, $t\bar{t}$, VV	$\approx 3\%$
b-tagged Jet Veto	Top in SR and all CRs	$\approx 3\%$
	All remaining in SR and all CRs	$\approx 1 - 2\%$
Top p_T Reweighting	Top	10%
Top Normalisation	Top	10%
VV Normalisation	VV	15%
$Z(\ell^+ \ell^-)$ +jets Normalisation	$Z(\ell^+ \ell^-)$ +jets (SR)	20%
QCD Statistics	QCD (SR)	42%
QCD Closure	QCD (SR)	50%

A 2.5% luminosity uncertainty is included on the single-top, $t\bar{t}$, and VV backgrounds following the recommendations of the CMS POGs. The QCD-multijet contribution is not affected by this uncertainty, due to the data-driven approach used. The uncertainties accounting for this estimation method, i.e. 42% and 50%, are discussed in Section 5.5.3.

The POGs recommendations were also followed for uncertainties accounting for p_T -reweighting and normalisation of the Top contribution, normalisation of the VV and $Z(\ell^+\ell^-)$ +jets backgrounds, and efficiency of veto for b-jets.

JES and JER uncertainties were estimated from simulation using the procedure detailed in Section 5.6.1. An uncertainty of about 5% is included since accounts for both effects on all minor backgrounds.

Finally, Section 5.6.1 provided details of the systematic uncertainties accounting for lepton reconstruction, ID, ISO, and veto efficiencies, and for the electron and MET trigger efficiencies.

5.6.4 Systematic Uncertainties on Higgs to Invisible Signals

The expected signal composition for the VBF Higgs to invisible analysis is around 20(48)% ggH and 80(52)% VBF for the cut-based (shape-based) approach, as further detailed in Section 5.9. The contribution from associated production channels, i.e. VH and ttH, is expected to be negligible. Therefore, theoretical uncertainties on Higgs boson production are included for both ggH and VBF modes and summarised in Table 5.4.

Table 5.4: The production cross-sections and relative uncertainties for the SM Higgs boson.

Production Mode	Cross-Section	QCD-Scale [%]	PDF + α_{QCD}	PDF Acceptance	QCD-Scale Acceptance	ggH 2-jet
ggH	48.6 pb	$^{+4.6}_{-6.7}$	$\pm 3.2\%$	1%	-	45%
VBF	3.78 pb	$^{+0.4}_{-0.3}$	$\pm 2.1\%$	1%	2%	-

The QCD-scale uncertainty is estimated from factorisation and renormalisation scale variations for both production modes to account for higher-order terms in the cross-section calculation. The PDF + α_{QCD} uncertainty covers contributions to the signals from the choice of the PDFs and coupling constants. These systematic uncertainties are provided by Ref. [53]. Uncertainties accounting for the change in acceptance as a result of these variations with respect to the phase space of the VBF Higgs to invisible analysis are also included, and are described in detail in Ref. [1]. Likewise, the ggH 2-jet uncertainty accounting for limitations of the theoretical prediction for ggH in a VBF-like phase space with two jets is detailed in Ref. [1].

5.7 Confidence Levels and Exclusion Limits

A procedure based on the CL_S prescription [80] is used to set upper limits on the branching ratio of the Higgs boson decay to invisible particles. The signal hypothesis (H_1) assumes $\mathcal{B}(H \rightarrow \text{inv.}) = 100\%$, whereas $\mathcal{B}(H \rightarrow \text{inv.})$ is equal to 0 for the background-only hypothesis (H_0). A likelihood function $\mathcal{L}$ is defined for an observation given its expectation under a certain hypothesis for different categories i , i.e. the SR and CRs:

$$\mathcal{L} = \prod_i P(n_i | \nu_i(\mathcal{B}(H \rightarrow \text{inv.}), \boldsymbol{\theta})) \cdot \prod_k^N e^{-(\theta_k - \bar{\theta}_k)^2}. \quad (5.6)$$

The Poisson term, $P(n_i | \nu_i(\mathcal{B}(H \rightarrow \text{inv.}), \boldsymbol{\theta})) = \nu_i^{n_i} e^{-\nu_i} / n_i!$, gives the probability to observe n_i events in the i -th category given ν_i expected events assuming a set of parameters: $\mathcal{B}(H \rightarrow \text{inv.})$ and $\boldsymbol{\theta} = \{\theta_k\}_{k=1}^N$. The uncertainties on ν_i are accounted for by the nuisance parameters θ_k . The Gaussian term, $e^{-(\theta_k - \bar{\theta}_k)^2}$, provides the constraints on the values allowed for θ_k , with $\bar{\theta}_k$ the best estimate of θ_k . All systematic uncertainties are modelled as log-normal rate modifiers, i.e. for a generic process Q, $Q(\boldsymbol{\theta}) = n_Q^{\text{MC}} \cdot \prod_k (1 + \kappa_k)^{\theta_k}$, where $\kappa = \frac{\delta n_Q^{\text{MC}}}{n_Q^{\text{MC}}}$ and n_Q^{MC} indicates the number of simulated events from the process Q.

The VBF Higgs to invisible analysis considers the yields in the SR and CRs simultaneously. The likelihood function is constructed as a product of Poisson probabilities, one for each region. The scale factor f_V , that multiplies the V+jets yield, and the $\mathcal{B}(H \rightarrow \text{inv.})$ are free parameters in the fit. The likelihood for the complete model is given by:

$$\nu = \mathcal{B}(H \rightarrow \text{inv.}) (S_{\text{ggH}}(\boldsymbol{\theta}) + S_{\text{VBF}}(\boldsymbol{\theta})) + \sum_j V_j^{\text{SR}}(\boldsymbol{\theta}) f_V + \sum_j B_j^{\text{SR}} \quad (5.7)$$

for the SR, where S_{ggH} and S_{VBF} refer to the ggH and VBF signals, respectively, and

$$\nu_i = \sum_j V_j^{\text{CR}}(\boldsymbol{\theta}) f_V + \sum_j B_j^{\text{CR}} \quad (5.8)$$

for the CRs $i = Z \rightarrow e^+e^-$, $Z \rightarrow \mu^+\mu^-$, $W \rightarrow e^\pm\nu$, and $W \rightarrow \mu^\pm\nu$. The quantities V_j and B_j indicate the V+jets processes and all other minor backgrounds, respectively, in the region under consideration.

A profile likelihood ratio $q_{\mathcal{B}(\text{H} \rightarrow \text{inv.})}$ is used as the test statistic, and is defined as:

$$q_{\mathcal{B}(\text{H} \rightarrow \text{inv.})} = -2 \ln \frac{\mathcal{L}(\text{observed} | \mathcal{B}(\text{H} \rightarrow \text{inv.}), \hat{\theta}_{\mathcal{B}(\text{H} \rightarrow \text{inv.})})}{\mathcal{L}(\text{observed} | \hat{\mathcal{B}}(\text{H} \rightarrow \text{inv.}), \hat{\theta}_{\hat{\mathcal{B}}(\text{H} \rightarrow \text{inv.})})}, \quad (5.9)$$

where $\hat{\theta}$ and $\hat{\mathcal{B}}(\text{H} \rightarrow \text{inv.})$ are the values of θ and $\mathcal{B}(\text{H} \rightarrow \text{inv.})$ for which the likelihood function is maximised, whereas $\hat{\theta}_{\mathcal{B}(\text{H} \rightarrow \text{inv.})}$ is the value of the nuisance parameter that maximises the likelihood for a given $\mathcal{B}(\text{H} \rightarrow \text{inv.})$.

The CL_S criterion is used to set limits on $\mathcal{B}(\text{H} \rightarrow \text{inv.})$, and is defined as:

$$\text{CL}_\text{S} = \frac{P\left(q_{\mathcal{B}(\text{H} \rightarrow \text{inv.})} \geq q_{\mathcal{B}(\text{H} \rightarrow \text{inv.})}^{\text{observed}} | H_1\right)}{P\left(q_{\mathcal{B}(\text{H} \rightarrow \text{inv.})} \geq q_{\mathcal{B}(\text{H} \rightarrow \text{inv.})}^{\text{observed}} | H_0\right)}, \quad (5.10)$$

where the probability P is calculated using the asymptotic approximations for the distributions of $q_{\mathcal{B}(\text{H} \rightarrow \text{inv.})}$ under the hypotheses H_1 and H_0 [81]. The values of $\mathcal{B}(\text{H} \rightarrow \text{inv.})$ for which $\text{CL}_\text{S} \leq 0.05$ are all excluded at 95% confidence level (CL) or greater. The upper limit is the smallest value among these values of $\mathcal{B}(\text{H} \rightarrow \text{inv.})$.

5.8 Signal Kinematics and Data-MC Optimisation

Various observables are compared between signal events, produced through the VBF and ggH modes and decaying invisibly, and backgrounds before obtaining the selection presented in Table 5.1. This shape comparison is performed by requiring only two jets in the opposite hemispheres with $p_\text{T} > 80$ and 40 GeV.

Figure 5.19 shows this study for signal-like events and for Z+jets and W+jets backgrounds. The distribution of η for the leading and sub-leading jets shows that a large fraction of the VBF signal is found in the forward region of the detector. Furthermore, the variables with most discriminating power are found to be m_{jj} , $\Delta\phi_{\text{jj}}$, and $\Delta\eta_{\text{jj}}$, chosen as reference observables for the optimisation of the analysis.

5.8.1 Optimisation Procedure

An optimisation study is performed to determine the cuts on the m_{jj} , $\Delta\phi_{\text{jj}}$, and $\Delta\eta_{\text{jj}}$ variables. The expected sensitivity of the analysis is computed for various selections

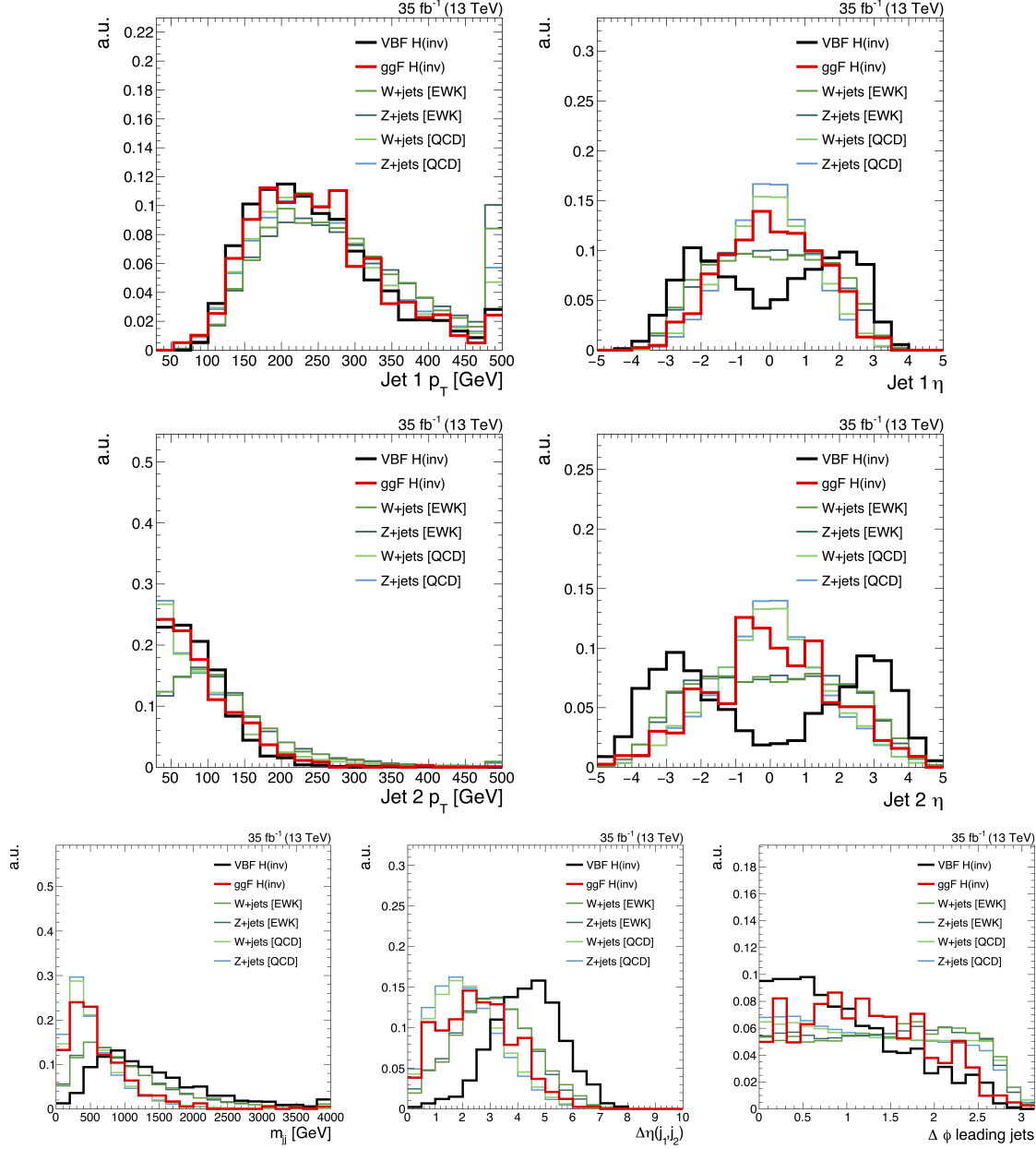


Figure 5.19: The comparison of various observables between signal events, produced through VBF (black) and ggH (red) modes and decaying invisibly, and Z+jets (blue) and W+jets (green) backgrounds.

and two scenarios. The VBF production mode is used at first to optimise the sensitivity specifically to the VBF production. Then, both VBF and ggH modes are used to include the improvement in sensitivity from the ggH contribution, which is around 20(48)% for the cut-based (shape-based) approach, as discussed in Section 5.9.

The procedure features three steps: (i) including all systematic uncertainties; (ii) removing all systematic uncertainties but using the lepton CRs to estimate the V+jets background; (iii) removing all systematic uncertainties, considering only the SR, and estimating the V+jets contribution from simulation. The first step uses all information necessary to obtain the final results presented in Section 5.9, whereas the second step quantifies the effect from the systematic uncertainties discussed in Section 5.6. Finally, the third step includes only the SR and uses simulation to estimate the V+jets contribution, assuming perfect modelling of these processes. An Asimov dataset derived from simulation is used to perform the fit in the SR and CRs and set an upper limit on $\mathcal{B}(\text{H} \rightarrow \text{inv.})$. The maximum-likelihood estimates of all parameters from a fit to the Asimov dataset are equal to their true values [81].

Figure 5.20 shows the expected upper limit on $\mathcal{B}(\text{H} \rightarrow \text{inv.})$ for a SM Higgs boson as a function of different cuts on the optimisation variables for each of the three steps of the procedure. Table 5.1 reports the chosen values from the optimisation procedure, where the loose cuts are selected to increase the MC statistics in each region, i.e. $m_{\text{jj}} > 1300 \text{ GeV}$, $\Delta\phi_{\text{jj}} < 1.5$, and $\Delta\eta_{\text{jj}} > 4.0$.

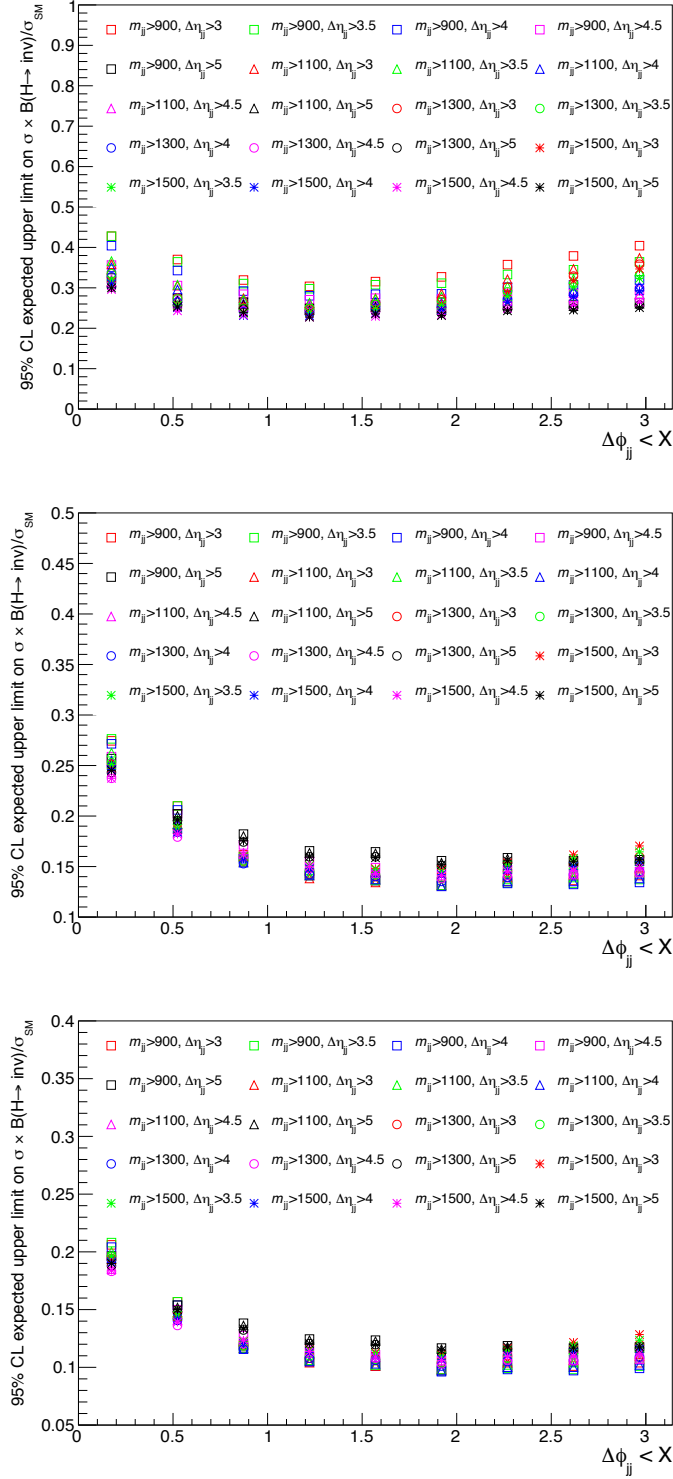


Figure 5.20: The expected upper limits on $\mathcal{B}(H \rightarrow \text{inv.})$ as a function of the optimisation variables. The fit includes the ggH channel in the signal prediction and all systematic uncertainties (top), no systematic uncertainties (central), and no systematic uncertainties and only the signal region (bottom).

Figure 5.21 shows the distribution of $\Delta\phi_{jj}$ for both data and simulated signal and backgrounds in the SR after the optimisation procedure. The backgrounds are estimated using MC samples, leading to an overall disagreement between data and simulation of about 20%. The distributions of m_{jj} and $\Delta\eta_{jj}$ are shown in Figure 5.8.

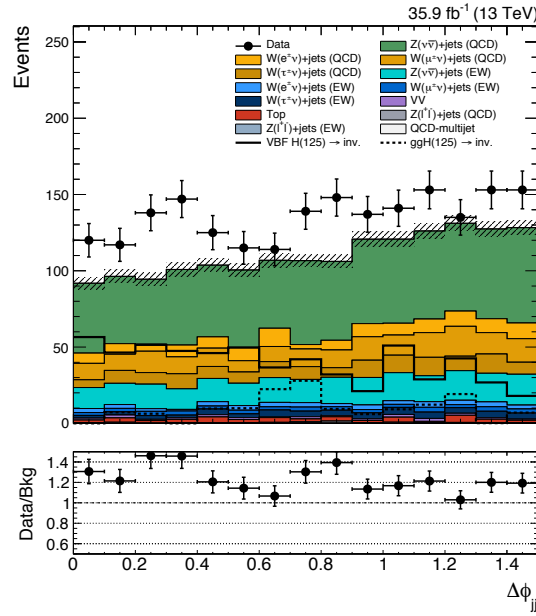


Figure 5.21: The distribution of $\Delta\phi_{jj}$ for data and simulated signal and backgrounds in the signal region.

5.9 Results

The agreement between data and MC in the SR is shown in Figures 5.7 to 5.9 and 5.21, using MC samples to predict the backgrounds, for the cut-based approach, and in Figure 5.23 for the shape-based method. Table 5.5 summarises the expected and observed event yields in the SR and CRs for various fit procedures using the cut-based approach, as discussed in Section 5.9.1.

The results obtained from both of the cut-based and shape-based analyses are interpreted in terms of an upper limit on the Higgs boson production cross-section times $\mathcal{B}(H \rightarrow \text{inv.})$, relative to the predicted cross-section (σ_{SM}) assuming SM production.

An interpretation of the search for invisibly decaying Higgs bosons is also presented for two distinctive BSM models. The first model allows for the presence of SM-like Higgs bosons with masses in the range $110 < m_H < 1000 \text{ GeV}$. The second sets upper limits on the spin-independent DM-nucleon cross-sections in a Higgs-portal DM interpretation [82], and is detailed in Chapter 6, as it uses the combination of Higgs to invisible searches.

5.9.1 The Cut-based Approach

The cut-based analysis ‘cuts and counts’ events in the SR and CRs using the selections detailed in Section 5.4 and Section 5.5, respectively. A simultaneous maximum-likelihood fit to data is performed across all four CRs, namely $Z \rightarrow e^+e^-$, $Z \rightarrow \mu^+\mu^-$, $W \rightarrow e^\pm\nu$, and $W \rightarrow \mu^\pm\nu$, to estimate the contribution from the $Z(\nu\bar{\nu})$ +jets and $W(\ell^\pm\nu)$ +jets backgrounds. As a result, a scale factor of $f_V = 1.18 \pm 0.06$ is measured for the V+jets processes. The expected and observed event yields in the SR and CRs are summarised in Table 5.5 and detailed in the following. Statistical and systematic uncertainties are reported for all the fits with the exclusion of the pre-fit.

The predicted backgrounds from MC are referred to as ‘pre-fit’ expected yields, which are used to show the agreement between data and simulation in the SR and CRs in all distributions presented in this thesis, such as Figures 5.9 and 5.10. The total background estimated in the double-muon CR is around 15% higher than data. The SR and the other three CRs are characterised by a background prediction between 10% and 23% lower than the number of observed events in each region.

The event yields in the double-lepton CRs are considerably smaller than the $Z(\nu\bar{\nu})$ +jets contribution in the SR. In fact, the $Z \rightarrow \ell^+\ell^-$ BR is approximately six times smaller

than the $Z \rightarrow \nu\bar{\nu}$ BR, leading to a limited statistical power to constrain the $Z(\nu\bar{\nu})$ +jets background. For this reason, the single-lepton CRs, that feature higher statistics, are jointly used to estimate the $Z(\nu\bar{\nu})$ +jets background in the SR. The post-fit, or ‘CR-only fit’, shows the estimate of the V+jets background obtained by fitting the data across all the CRs but excluding the SR. The agreement between data and simulation is significantly improved in the single-lepton and double-electron CRs, whereas it is slightly reduced in the double-muon CR since the high statistical power of the single-lepton CRs drives the fit. The total background in the SR is around 14% lower than data.

The V+jets background estimate obtained by a simultaneous fit across the SR and CRs, under the background-only hypothesis, is referred to as the ‘b-only fit’. The fit includes data from the SR and improves the agreement between data and simulation from 14% to 2% by scaling up the total background in this region. The simultaneous fit across all regions significantly improves the agreement in the electron CRs, whereas the muon CRs are penalised.

Finally, the ‘s+b fit’ provides event yields allowing for the presence of an invisibly decaying Higgs boson signal, whose normalisation is a free parameter in the fit as discussed in Section 5.7. The expected signal composition of the production modes is around 20% ggH and 80% VBF, as mentioned in Section 5.6.4. The background predictions in all regions are equal to the results of the CR-only fit, and the s+b fit adjusts the signal to cover the 14% difference between data and simulation shown by the CR-only fit in the SR.

An excess of events in the SR is observed corresponding to a significance of 2.5 standard deviations compared to the background prediction. This is mostly due to events in the low- m_{jj} region which are not compatible with a signal for an invisibly decaying Higgs boson produced through the VBF mode, as shown in Figure 5.19 (bottom left). The observed (expected) upper limit is measured to be $\mathcal{B}(H \rightarrow \text{inv.}) < 0.58(0.30)$ at 95% CL. The regions containing 68% and 95% of the distribution of limits expected under the background-only hypothesis are $[0.22, 0.43]$ and $[0.17, 0.58]$, respectively.

Figure 5.22 shows the impacts of the nuisance parameters after a fit to data in the SR and CRs, and the post-fit nuisance parameter constraints. The impact is defined as the difference in the best-fit value of $\mathcal{B}(H \rightarrow \text{inv.})$, denoted as $\hat{r}$, when fixing each nuisance parameter to its nominal $+1\sigma_{\text{up}}$ and nominal $-1\sigma_{\text{down}}$ values. The best-fit value of $\mathcal{B}(H \rightarrow \text{inv.})$ is measured to be $\hat{r} = 0.31^{+0.13}_{-0.12}$.

The first and ninth nuisance parameters in the list refer to the theoretical systematic uncertainty on the QCD and EW parts, respectively, of the $\frac{Z(\nu\bar{\nu})+\text{jets}}{W(e^\pm\nu)+\text{jets}}$ ratio, as detailed in Section 5.6.2. The pre-fit value of both parameters are $\theta_k = 0$, where θ_k is defined in Section 5.7, whereas the post-fit best estimate values are approximately -0.7 for QCD and -0.2 for EW. The QCD part is also slightly constrained by the fit, as shown by the decrease in size of its error bars, i.e. from ± 1 to ± 0.7 . The impact of the nuisance parameters on $\hat{r}$ is $\pm 5\%$ and $\pm 1\%$ for the QCD and EW parts, respectively, when their nominal values are shifted by $\mp 1\sigma_{\text{up}}^{\text{down}}$. In fact, both parameters directly affect the background in the SR, leading $\hat{r}$ to decrease (increase) when the background is scaled up (down).

A similar argument is valid for systematic uncertainties accounting for the MET trigger efficiency (2nd), QCD scale variation in a ggH 2-jet phase space (3rd), and finite ggH simulated events (10th). In contrast, the nuisance parameters affecting backgrounds in the CR have an inverted impact on $\hat{r}$. For example, these uncertainties account for ID (4th, 5th) and tracking (6th, 8th) efficiencies for muons, and the electron reconstruction efficiency (7th).

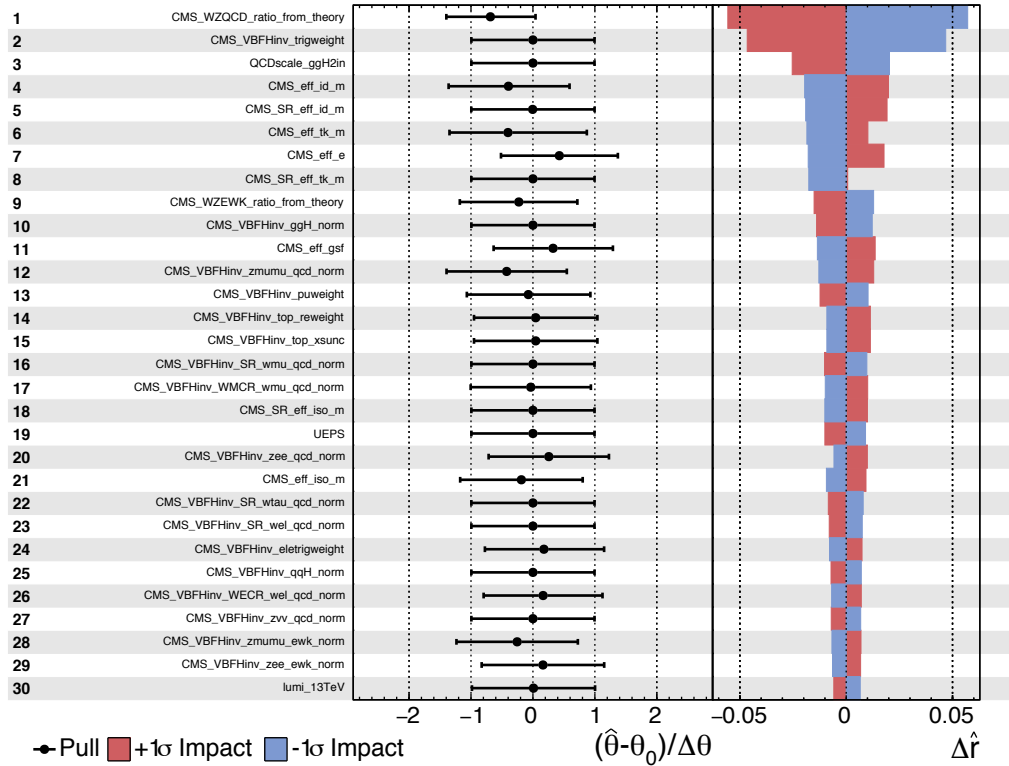


Figure 5.22: The impacts of the nuisance parameters from a fit to data in the SR and CRs.

Table 5.5: The expected and observed event yields in the SR and CRs of the cut-and-count analysis for various SM processes. The background yields and the corresponding uncertainties are obtained from: the MC prediction (pre-fit), a combined fit to data in all CRs, but excluding data in the SR (CR-only fit), a combined fit to data in the SR and CRs assuming the absence of a signal (b-only fit), and a combined fit to data in the SR and CRs allowing for the presence of a signal for an invisibly decaying Higgs boson (s+b fit).

Region	Process	Pre-fit (MC)	CR-only fit	b-only fit	s+b fit
Double-electron	Top	2.7	2.9 ± 1.0	2.3 ± 0.9	2.9 ± 1.2
	VV	0.6	0.7 ± 0.4	0.6 ± 0.4	0.7 ± 0.5
	$W(e^\pm\nu)+\text{jets}$ (QCD)	0.8	0.9 ± 0.6	0.9 ± 0.6	0.9 ± 0.6
	$Z(e^+e^-)+\text{jets}$ (EW)	21.5	24.1 ± 3.2	25.1 ± 3.3	24.1 ± 3.6
	$Z(e^+e^-)+\text{jets}$ (QCD)	62.6	66.5 ± 6.0	73.6 ± 5.5	66.5 ± 6.4
	Total Background	88.3	95.0 ± 6.3	102.6 ± 5.7	95.0 ± 6.0
	Observed Events		104	104	104
Single-electron	QCD-multijet	2.0	2.0 ± 2.0	2.0 ± 2.0	2.0 ± 2.1
	Top	71.8	74.2 ± 13.6	64.6 ± 12.4	74.2 ± 15.5
	VV	14.0	14.4 ± 3.7	13.2 ± 3.3	14.4 ± 3.8
	$W(e^\pm\nu)+\text{jets}$ (EW)	223.6	253.1 ± 11.0	259.3 ± 10.0	253.1 ± 9.5
	$W(e^\pm\nu)+\text{jets}$ (QCD)	477.9	541.5 ± 20.4	554.5 ± 19.1	541.5 ± 17.4
	$Z(e^+e^-)+\text{jets}$ (EW)	2.1	2.4 ± 0.2	2.4 ± 0.2	2.4 ± 0.2
	$Z(e^+e^-)+\text{jets}$ (QCD)	4.6	5.2 ± 0.3	5.4 ± 0.2	5.2 ± 0.2
	Total Background	798.5	895.6 ± 23.7	904.3 ± 23.5	895.6 ± 23.5
	Observed Events		902	902	902
Double-muon	Top	5.0	4.8 ± 1.4	3.9 ± 1.3	4.8 ± 1.5
	VV	2.4	2.3 ± 1.1	2.2 ± 0.9	2.3 ± 1.1
	$W(\mu^\pm\nu)+\text{jets}$ (EW)	0.1	0.1 ± 0.1	0.1 ± 0.1	0.1 ± 0.1
	$W(\mu^\pm\nu)+\text{jets}$ (QCD)	0.2	0.2 ± 0.2	0.2 ± 0.3	0.2 ± 0.2
	$Z(\mu^+\mu^-)+\text{jets}$ (EW)	30.9	32.5 ± 4.1	34.0 ± 4.2	32.5 ± 4.5
	$Z(\mu^+\mu^-)+\text{jets}$ (QCD)	91.9	91.5 ± 7.6	101.5 ± 6.8	91.5 ± 7.9
	Total Background	130.4	131.5 ± 7.8	142.0 ± 6.9	131.5 ± 7.4
	Observed Events		114	114	114
Single-muon	QCD-multijet	22.3	23.0 ± 13.1	21.6 ± 9.8	23.0 ± 16.9
	Top	110.4	112.1 ± 22.4	97.0 ± 20.7	112.1 ± 21.8
	VV	21.3	21.3 ± 4.4	19.9 ± 4.2	21.3 ± 4.5
	$W(\mu^\pm\nu)+\text{jets}$ (EW)	365.0	405.6 ± 14.8	415.9 ± 13.1	405.6 ± 14.0
	$W(\mu^\pm\nu)+\text{jets}$ (QCD)	816.2	906.9 ± 30.1	929.3 ± 27.1	906.9 ± 29.1
	$Z(\mu^+\mu^-)+\text{jets}$ (EW)	5.2	5.7 ± 0.3	5.9 ± 0.2	5.7 ± 0.2
	$Z(\mu^+\mu^-)+\text{jets}$ (QCD)	24.4	27.1 ± 1.2	28.2 ± 1.1	27.1 ± 1.1
	Total Background	1364.7	1501.6 ± 34.3	1517.8 ± 33.4	1501.6 ± 33.7
	Observed Events		1504	1504	1504
Signal Region	QCD-multijet	3.3	3.3 ± 2.3	3.4 ± 2.4	3.3 ± 2.3
	Top	36.8	37.8 ± 8.8	34.0 ± 9.0	37.8 ± 8.8
	VV	18.0	18.5 ± 6.2	16.6 ± 5.4	18.5 ± 5.7
	$W(e^\pm\nu)+\text{jets}$ (EW)	39.5	44.2 ± 3.8	47.3 ± 3.8	44.2 ± 3.5
	$W(e^\pm\nu)+\text{jets}$ (QCD)	117.6	131.4 ± 9.7	141.2 ± 9.1	131.4 ± 9.2
	$W(\mu^\pm\nu)+\text{jets}$ (EW)	35.8	40.0 ± 7.1	47.3 ± 7.7	40.0 ± 7.0
	$W(\mu^\pm\nu)+\text{jets}$ (QCD)	192.7	215.2 ± 20.5	241.6 ± 19.8	215.2 ± 20.0
	$W(\tau^\pm\nu)+\text{jets}$ (EW)	50.9	56.9 ± 4.3	60.9 ± 4.2	56.9 ± 4.3
	$W(\tau^\pm\nu)+\text{jets}$ (QCD)	136.3	152.2 ± 11.1	163.8 ± 10.1	152.2 ± 10.4
	$Z(\ell^+\ell^-)+\text{jets}$ (QCD)	7.5	8.4 ± 1.7	9.4 ± 1.8	8.4 ± 1.8
	$Z(\nu\bar{\nu})+\text{jets}$ (EW)	247.4	268.6 ± 32.2	294.3 ± 33.6	268.6 ± 34.9
	$Z(\nu\bar{\nu})+\text{jets}$ (QCD)	779.1	801.9 ± 72.0	933.6 ± 53.6	801.9 ± 70.5
	Total Background	1665.5	1778.8 ± 95.8	1993.9 ± 41.5	1778.8 ± 67.8
	Observed Events			2035	2035
	ggH (100% $\mathcal{B}(H \rightarrow \text{inv.})$, $m_H = 125.09 \text{ GeV}$)	146.8			50.5 ± 31.9
	VBF (100% $\mathcal{B}(H \rightarrow \text{inv.})$, $m_H = 125.09 \text{ GeV}$)	596.7			205.7 ± 82.3

5.9.2 The Shape-based Approach

The shape-based approach fully exploits the VBF kinematics to improve the sensitivity to an invisibly decaying Higgs boson. Table 5.1 shows that the m_{jj} and $\Delta\eta_{jj}$ selections are lowered with respect to the cut-based criteria, i.e. from 1300 GeV to 200 GeV and from 4.0 to 1.0, respectively. This allows one to incorporate the information from the shape of the m_{jj} distribution into the final fit procedure. The choice of this specific variable was determined by an optimisation study as reported in Ref. [1]. The signal and background shapes are fitted to the m_{jj} distribution observed in data, and used to extract the observed signal. This would be expected to manifest itself as an excess of events over the total background in the high- m_{jj} region, as shown in Figure 5.19 (bottom left), similarly to the cut-based approach.

Figure 5.23 shows the observed and expected distributions of m_{jj} in the SR after applying the event selection in Table 5.1. On the left, the background is estimated by fitting the data in all CRs, referred to as the CR-only fit. A combined fit to the data in the SR and CRs, assuming the absence of any signal (b-only fit), is shown on the right. The expected signal distributions for an invisibly decaying SM Higgs boson produced via ggH and VBF modes and with $\mathcal{B}(H \rightarrow \text{inv.}) = 100\%$ are overlaid. The upper ratio panel shows ratios of the data and pre-fit (post-fit) background prediction in red (blue), using the grey band to indicate the total uncertainty after performing the fit. This provides information on a possible disagreement between the data and simulation, and the improvement from the post-fit, by showing their ratios in various regions of the distribution. The lowest ratio panel shows the comparison between the data and post-fit background estimate relative to the uncertainty on the post-fit. This serves as a scale to measure the relative tension between the data and simulation after performing the post-fit.

The comparison between data and estimated backgrounds in the SR shows a 4 – 10% excess localised in the bulk of the m_{jj} distribution. As outlined in Section 5.9.1, this excess is inconsistent with a possible VBF signal since the discrepancy would be expected to increase as a function of m_{jj} . The data are thus compatible with the SM prediction, and the observed upper limit is $\mathcal{B}(H \rightarrow \text{inv.}) < 0.33$ at 95% CL. The expected upper limit is $\mathcal{B}(H \rightarrow \text{inv.}) < 0.25$ at 95% CL, and the regions containing 68% and 95% of the distribution of limits are $[0.18, 0.35]$ and $[0.14, 0.47]$, respectively.

The shape-based approach shows an improved expected upper limit on $\mathcal{B}(H \rightarrow \text{inv.})$ with respect to the cut-based approach since information from the shape of the m_{jj} distribution is used in the fit procedure. The loosened m_{jj} and $\Delta\eta_{jj}$ criteria allow a greater number

of events to pass the selection, and increase the contribution from the ggH production mode from 20% to 48%, leading to an improvement in the sensitivity of the analysis. These motivations also partially drive the argument on the observed upper limit on $\mathcal{B}(H \rightarrow \text{inv.})$, that decreases from 0.58 to 0.33. In addition, the statistical power of the distribution in the region $0.2 < m_{jj} < 1 \text{ TeV}$ mitigates the limited agreement between the data and simulation in the region $1 < m_{jj} < 1.5 \text{ TeV}$, consequently decreasing the significance of the excess of events detailed in Section 5.9.1.

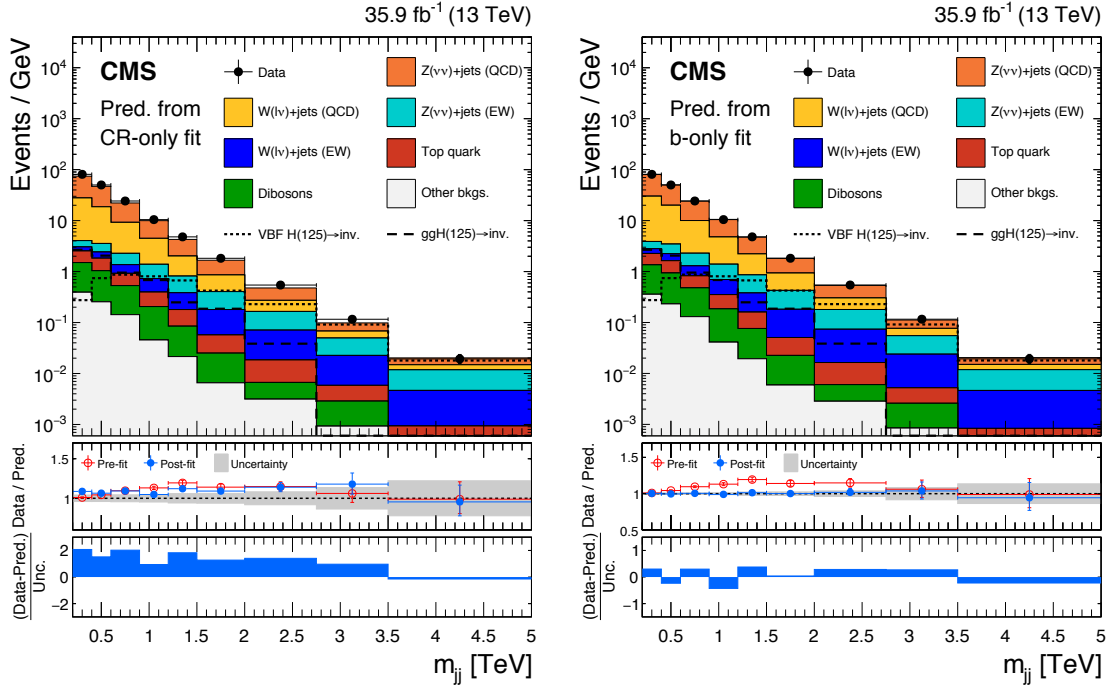


Figure 5.23: The observed distribution of m_{jj} in the SR compared to the yields of various SM background processes resulting from the CR-only fit (left) and b-only fit (right), following the shape-based approach. The distribution of a signal for an invisibly decaying Higgs boson is overlaid for both VBF and ggH production modes, assuming $m_H = 125 \text{ GeV}$ and $\mathcal{B}(H \rightarrow \text{inv.}) = 100\%$ [1].

5.9.3 Constraints on SM-like Higgs Bosons

A BSM scenario in which an additional SM-like Higgs boson (that does not mix with the SM 125 GeV Higgs boson) is also allowed to decay to invisible final states [56] is presented in this Section. The model includes VBF and ggH mechanisms in the production of the Higgs boson.

The results presented in Sections 5.9.1 and 5.9.2 are interpreted in the context of this model. Figure 5.24 shows 95% CL upper limits on $\frac{\sigma}{\sigma_{\text{SM}}} \times \mathcal{B}(\text{H} \rightarrow \text{inv.})$ as a function of the mass hypothesis of the SM-like Higgs boson (m_{H}) for both approaches.

Figure 5.24 (left) shows a constant significance of 2.5 standard deviations for the excess of events observed in the SR compared to the background prediction as a function of m_{H} . The production cross-section σ of the Higgs boson, assuming SM-like coupling, decreases as a function of the m_{H} hypothesis, as shown in Figure 3.5. As a consequence, the amount of expected signal also decreases as a function of m_{H} , and this is accounted for by an increase in the fitted $\mathcal{B}(\text{H} \rightarrow \text{inv.})$. A similar argument is valid for Figure 5.24 (right). However, the shape-based approach includes information from the distribution of m_{jj} . The shape of this observable is a free parameter in the fit, and affects the results since it can change as a function of m_{H} .

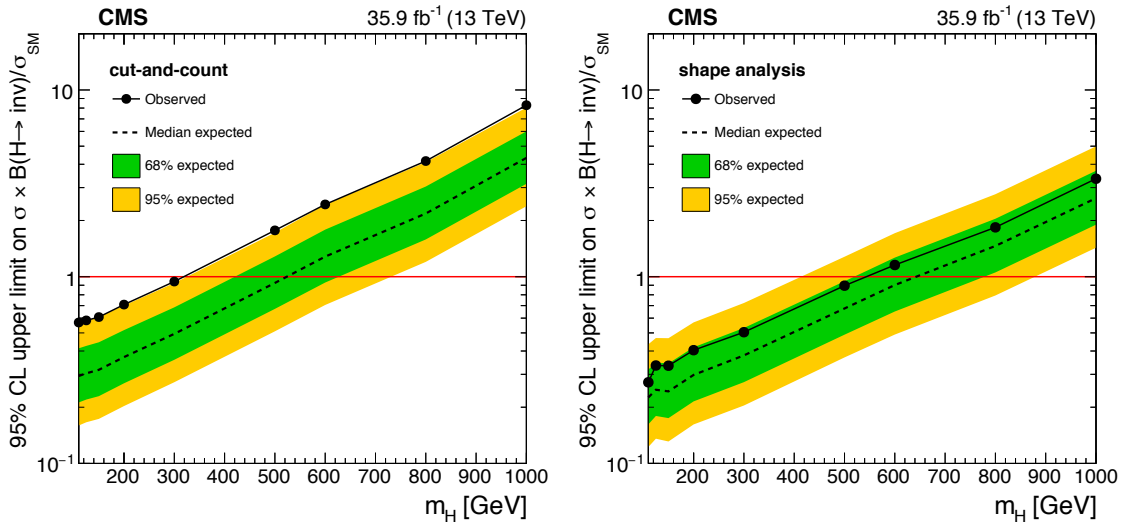


Figure 5.24: The observed (solid black) and expected (dashed black) 95% CL upper limits on $\frac{\sigma}{\sigma_{\text{SM}}} \times \mathcal{B}(\text{H} \rightarrow \text{inv.})$ as a function of the SM-like Higgs boson mass m_{H} for the cut-based (left) and shaped-based (right) approaches. The inner (green) and outer (yellow) band indicate the regions containing 68% and 95%, respectively, of the distribution of limits expected under the background-only hypothesis [1].

Chapter 6

Combinations of Higgs to Invisible Searches

This Chapter presents combinations of searches for invisibly decaying Higgs bosons. Section 6.1 gives an overview of the datasets and analyses used. Section 6.2 details the combination of searches in Run-1 and in the first part of Run-2 (2015), which was conducted by the author and is published in Ref. [2]. Section 6.3 focuses on the combination of searches using the 2016 dataset, to which the author contributed significantly. The results from this combination are further combined with those from Section 6.2 and published in Ref. [1], and interpretations under non-SM production assumptions and DM-Higgs portal models are provided. These interpretations are also provided for the combination in Ref. [2], but they are not presented in this thesis.

6.1 Searches for Invisibly Decaying Higgs Bosons

Direct searches for invisibly decaying Higgs bosons increase the sensitivity to the invisible width of the Higgs boson. Such measurements are complementary to the constraints provided by indirect searches, which use the visible decay channels of the Higgs boson. For this reason, a combination of several searches targeting specific invisible Higgs boson decay signatures is carried out by looking at final states with large MET recoiling against a distinctive visible system. The data collected by CMS in 2011 and 2012 at $\sqrt{s} = 7$ and 8 TeV, respectively, i.e. Run-1, and in 2015 and 2016 at $\sqrt{s} = 13$ TeV, i.e. first part of Run-2, are used in the combinations presented in this Chapter. The data collected during these data-taking periods correspond to 4.9, 19.7, 2.3, and 35.9 fb⁻¹ of pp collisions, respectively.

The combination targets the production of a Higgs boson in the VBF, VH, and ggH modes assuming SM production cross-sections. The Feynman diagrams for these channels are shown in Figure 3.4. A detailed description of the VBF Higgs to invisible topology is provided in Chapter 5. The VH searches target both ZH and WH productions where the vector bosons decay to light-flavour jets ($V \rightarrow jj$). The $Z \rightarrow e^+e^-$, $Z \rightarrow \mu^+\mu^-$, and $Z \rightarrow b\bar{b}$ decays are also included in these analyses, together with the contribution from $gg \rightarrow ZH$ processes. The 7 and 8 TeV $Z \rightarrow \ell^+\ell^-$ searches and the 8 TeV $Z(b\bar{b})H$ analysis are described in Ref. [83]. The ggH production mode features a Higgs boson and a high- p_T jet in the final state, and is referred to as the mono-jet final state. The combinations for these searches are detailed in the following Sections.

6.2 Combination of Searches in Run-1 and Run-2 (2015)

An invisibly decaying Higgs boson has a characteristic signature regardless of the production topology. The MET in the final state recoils against a visible system, whose properties are exploited to reduce the expected contribution from SM backgrounds. This visible system consists of either jets or leptons, and mutually exclusive categories are designed to target events from a specific production mode. Table 6.1 summarises the expected signal composition in each analysis included in the combination.

Table 6.1: The expected composition of production modes of a 125 GeV Higgs boson for each analysis included in the combination, assuming SM production cross-sections [2].

Analysis	Final State	Integrated $\mathcal{L}$ (fb^{-1})			Expected Signal Composition (%)	
		7 TeV	8 TeV	13 TeV	7 or 8 TeV	13 TeV
qqH-tagged	VBF-jets	-	19.2	2.3	7.8 ggH, 92.2 VBF	9.1 ggH, 90.9 VBF
VH-tagged	$Z \rightarrow \ell^+\ell^-$	4.9	19.7	2.3	100 ZH	
	$Z \rightarrow b\bar{b}$	-	18.9	-	100 ZH	
	$V \rightarrow jj$	-	19.7	2.3	25.1 ggH, 5.1 VBF, 23.0 ZH, 46.8 WH	38.7 ggH, 7.1 VBF, 21.3 ZH, 32.9 WH
ggH-tagged	Mono-jet	-	19.7	2.3	70.4 ggH, 20.4 VBF, 3.5 ZH, 5.7 WH	69.3 ggH, 21.9 VBF, 4.2 ZH, 4.6 WH

The composition of production modes in the 13 TeV 2015 VBF Higgs to invisible analysis differs to the results presented in Section 5.9.1. In fact, the selection detailed in Section 6.2.1 shows significant differences in the sub-leading jet p_T and $\min\Delta\phi(E_T^{\text{miss}}, j)$ criteria compared to the selection in Table 5.1, reducing the contribution from the ggH production mode. The limited discriminatory power of the jet ID used to categorise events in the $V \rightarrow jj$ and mono-jet analyses leads to a mixture of different production modes in the expected signal composition. The 7 or 8, and 13 TeV data-taking periods are expected to have similar signal compositions. However, a less efficient lepton veto requirement in the 13 TeV $V \rightarrow jj$ analysis causes a higher ZH to WH ratio with respect to the 8 TeV search.

An optimisation is carried out for each analysis with respect to the specific conditions and integrated luminosity of the data-taking period. This leads to differences in the kinematic requirements across the 7 or 8, and 13 TeV datasets that are discussed in the following.

6.2.1 The VBF Analysis

The topology of an invisibly decaying Higgs boson produced in the VBF mode is discussed in detail in Chapter 5. The selection criteria used for the analyses in this combination differ in comparison to the 2016 analysis (Table 5.1). The leading and sub-leading jets are required to have $p_T > 50(80)$ and $45(70)$ GeV, respectively, for the 8(13) TeV analysis. They are required to be separated by $\Delta\eta_{jj} > 3.6$ and have an invariant mass of $m_{jj} > 1200(1100)$ GeV at 8(13) TeV. Events are further required to have $\text{MET} > 90$ and 200 GeV for the 8 and 13 TeV analysis, respectively. A $\min\Delta\phi(E_T^{\text{miss}}, j) > 2.3$ requirement is included to reduce the QCD-multijet background. Furthermore, both these analyses use dedicated VBF triggers, pre-selecting on jet p_T , m_{jj} , and $\Delta\eta_{jj}$ [83].

The $Z(\nu\bar{\nu})+\text{jets}$ and $W(\ell^\pm\nu)+\text{jets}$ backgrounds are estimated using four CRs: single-muon, single-electron, single-tau-lepton, and double-muon. These regions are defined by removing the lepton veto from the SR selection and requiring exactly one isolated lepton. In the single-tau-lepton CR, a selection of $1 < \min\Delta\phi(E_T^{\text{miss}}, j) < 2.3$ is applied to ensure orthogonality with the SR since the latter does not explicitly veto on tau leptons. A theoretical systematic uncertainty of 30% on the $W(\ell^\pm\nu)+\text{jets}$ to $Z(\nu\bar{\nu})+\text{jets}$ ratio is included in the analysis, and is estimated by comparing the predictions at LO and NLO accuracy.

The QCD-multijet background is estimated using an additional CR with a requirement on $\min\Delta\phi(E_T^{\text{miss}}, j) < 0.5$ for the 13 TeV analysis, whereas the 8 TeV analysis uses a different approach not discussed in this thesis and further detailed in Ref. [83]. A 80(100)% systematic uncertainty is included for biases in the extrapolation to the signal region at 8(13) TeV. The VV, $t\bar{t}$, and single-top backgrounds are estimated directly from simulation.

Several experimental systematic uncertainties are included during the background prediction to account for JES and JER (8%), pileup description, and lepton reconstruction efficiencies. A 20% uncertainty is included on the prediction of the $W \rightarrow \tau^\pm \nu$ contribution due to a loose selection on the single-tau-lepton CR, and 7(10)% and 10(20)% uncertainties are included on the production cross-sections of the VV and top-quark backgrounds, respectively, for the 8(13) TeV analysis.

A maximum-likelihood fit is performed simultaneously across the SR and CRs to estimate the background contribution. The inputs to the fit are the expected MC background yields and observed event counts in the CRs. A SF for both of the W +jets and Z +jets processes, and a SF for QCD-multijet yields in the CRs, are allowed to vary as free parameters in the fit. Therefore, the fit constrains these contributions directly from the data.

Table 6.2 summarises the expected and observed yields in the SR and CRs of the 2015 13 TeV analysis. The background yields and the corresponding statistical and systematic uncertainties are obtained from a combined fit to data in all CRs, excluding data in the SR (CR-only fit).

6.2.2 The $Z \rightarrow \ell^+ \ell^-$ Analysis

Although the cross-section of the $Z(\ell^+ \ell^-)H$ production mode is smaller than that of the VBF production mode, the channel features a clean final state with a smaller contribution from SM backgrounds. This search targets events with an opposite-charge, same-flavour lepton pair, i.e. e^+e^- or $\mu^+\mu^-$, consistent with a leptonic decaying Z boson. Large MET is produced in association with the pair of leptons, which is used as the visible system for the Higgs to invisible analysis. Details of the analysis are reported in Ref. [2].

The events are categorised into 0-jet and 1-jet, with jets satisfying $p_T > 30 \text{ GeV}$ and $|\eta| < 4.7$. A 2% uncertainty on the signal process, anti-correlated between 0-jet and

Table 6.2: The expected and observed event yields in the SR and CRs of the 2015 13 TeV analysis for various SM processes. The predicted backgrounds from a CR-only fit are shown with statistical and systematic uncertainties. The SM VBF and ggH production rates are assumed for an invisibly decaying Higgs boson with a mass $m_H = 125$ GeV and $\mathcal{B}(H \rightarrow \text{inv.}) = 100\%$ [2].

Signal Region			Control Regions				
Process			Single-electron	Single-muon	Single-tau-lepton	Double-muon	QCD
$Z(\mu^+\mu^-)+\text{jets}$	QCD	-	-	-	-	4.2 ± 1.1	-
	EW	-	-	-	-	2.0 ± 0.7	-
$Z(\nu\bar{\nu})+\text{jets}$	QCD	47 ± 12	-	-	-	-	-
	EW	21 ± 7	-	-	-	-	-
$W(\mu^\pm\nu)+\text{jets}$	QCD	13 ± 2	-	53 ± 5	0.4 ± 0.2	-	45 ± 5
	EW	4.3 ± 0.8	-	27 ± 3	-	-	6.0 ± 0.9
$W(e^\pm\nu)+\text{jets}$	QCD	9.3 ± 1.5	17 ± 3	-	0.2 ± 2.2	-	39 ± 4
	EW	5.4 ± 1.1	7.8 ± 1.3	-	0.2 ± 0.1	-	6.1 ± 1.0
$W(\tau^\pm\nu)+\text{jets}$	QCD	13 ± 2	0.06 ± 0.06	-	12 ± 2	-	74 ± 9
	EW	5.5 ± 1.2	-	-	5.1 ± 1.2	-	24 ± 3
Top		2.3 ± 0.4	1.5 ± 0.3	6.8 ± 0.9	7.1 ± 1.0	0.22 ± 0.06	82 ± 11
QCD-multijet		3 ± 23	-	5 ± 3	0.4 ± 0.3	-	1200 ± 170
VV		0.7 ± 0.3	0.4 ± 0.4	0.8 ± 0.4	-	0.02 ± 0.02	1.8 ± 0.7
Total Background		125 ± 28	27 ± 3	91 ± 8	25 ± 4	6.4 ± 1.4	1500 ± 170
Observed Events		126	29	89	24	7	1461
Signal	VBF	53.6 ± 4.9					
$m_H = 125$ GeV	ggH	5.4 ± 3.6					

1-jet, is included to account for the event migration between these jet categories. The selected leptons are required to have $p_T > 20$ GeV with an invariant mass $76 < m_{\ell\ell} < 106$ GeV. The transverse mass of the lepton pair+MET system, m_T , is required to be greater than 200 GeV. The events are also required to have $E_T^{\text{miss}} > 120(100)$ GeV and $\Delta\phi(\ell^+\ell^-, E_T^{\text{miss}}) > 2.7(2.8)$ for the 7 and 8(13) TeV searches.

The VV background, i.e. $ZZ \rightarrow \ell^+\ell^-\nu\bar{\nu}$ and $ZW \rightarrow \ell^+\ell^-\ell^\pm\nu$, is suppressed by rejecting additional electrons or muons with $p_T > 10$ GeV. A veto on tau leptons with $p_T > 20$ GeV is also used to suppress the ZW contribution. A γ +jets CR is used to determine the Z+jets contribution, and a systematic uncertainty of 100% is included throughout to account for the limited data in the high- p_T region. A selection on $\min\Delta\phi(E_T^{\text{miss}}, j) > 0.5$ is applied in the 1-jet category for the 13 TeV dataset to further reduce this contribution.

The $W \rightarrow t\bar{t}$, tW , and WW backgrounds are estimated through opposite-charge, different-flavour lepton CRs in the 13 TeV analysis, including a 70% uncertainty for the extrapolation process. The 8 TeV analysis performs a similar method using sideband regions around the mass peak of the Z boson. A 10 – 15% systematic uncertainty is included to

cover the differences between the results of the two methods. Uncertainties accounting for JES and JER, MET energy scale and resolution, lepton efficiencies, and momentum scale are around 2% each, and statistical uncertainties are included for simulated samples.

6.2.3 The $V \rightarrow jj$ and Mono-jet Analyses

The mono-jet and $V \rightarrow jj$ searches select final states with central jets and MET to target the ggH and VH production modes, respectively. The search strategies for these channels are similar since both target events with large MET recoiling against jets from either gluon radiation (ggH mode) or a hadronically decaying vector boson (VH mode). The jet topology is used to classify events in two exclusive event categories. Full details of these analyses are given in Ref. [2].

A hadronically decaying vector boson with high p_T produces final state particles reconstructed as a single large-radius jet. The $V \rightarrow jj$ channel includes events required to have jet $p_T > 200(250)$ GeV and $|\eta| < 2(2.4)$, $\min\Delta\phi(E_T^{\text{miss}}, j) > 2(0.5)$ for the 8(13) TeV analysis, and $E_T^{\text{miss}} > 250$ GeV. An event failing this selection is included in the mono-jet category, which has a looser selection in jet $p_T > 150(100)$ GeV and $|\eta| < 2(2.5)$ for the 8(13) TeV analysis, and $E_T^{\text{miss}} > 200$ GeV. The overlap with the VBF search is avoided by removing events that pass the VBF selection.

These analyses suffer from large backgrounds, i.e. $Z(\nu\bar{\nu})+\text{jets}$ and $W(\ell^\pm\nu)+\text{jets}$, which are estimated using CRs in data consisting of double-muon, single-muon, and $\gamma+\text{jets}$ events. Additional double-electron and single-electron CRs are used in the 13 TeV analyses. The events in the CRs satisfy the SR selection criteria with the exception of lepton and photon vetoes, and are separated into $V \rightarrow jj$ and mono-jet categories. The remaining top-quark and VV backgrounds are estimated from simulation. The QCD-multijet background contribution is estimated using simulation in the 8 TeV dataset, and using a dedicated CR in the 13 TeV analyses.

Theoretical uncertainties, whose values are up to 20% for large MET, cover contributions to the W/Z and γ/Z differential cross-section ratios from the choice of the PDFs [84] for higher-order EW corrections. Systematic uncertainties are included to account for JES (2 – 5%), and efficiencies in the selection of leptons, τ_h , and photons (up to 3%). A 13% uncertainty on the yields in the $V \rightarrow jj$ and mono-jet categories, anti-correlated between the two, is included to account for uncertainties in the efficiency of the $V \rightarrow jj$ tagging.

Additional systematic uncertainties are included in the background cross-sections and to account for data-to-MC discrepancies.

A simultaneous maximum-likelihood fit across all regions is performed to constrain the MC to data for the extraction of the final result. The analysis features a shape-based approach, using information from the shape of the E_T^{miss} distribution in the fit procedure.

6.2.4 Results

The results of these searches show no significant deviations from the SM expectations [2], and are interpreted in terms of upper limits on $\mathcal{B}(\text{H} \rightarrow \text{inv.})$. The calculation of these limits assumes the various Higgs boson production cross-sections are as given in the SM, as illustrated in Figure 3.5 for $m_{\text{H}} = 125 \text{ GeV}$.

The analyses have several nuisance parameters in common, whose correlations are studied across the different searches and Run-eras. These correlations are accounted for in the statistical procedure used to extract the final upper limit on $\mathcal{B}(\text{H} \rightarrow \text{inv.})$. The uncertainties are correlated across all tagged analyses for a given dataset for sources such as the VV cross-sections, lepton efficiencies, momentum scales, and integrated luminosity. The inclusive signal cross-sections uncertainties are also correlated across the different data-taking eras. The JES and JER uncertainties are uncorrelated between the VBF channel and the $\text{V} \rightarrow \text{jj}$ and mono-jet searches due to the distinct kinematics of the jets. A different technique is used to estimate the JES and JER uncertainties for the b-tagged jets in the $\text{Z} \rightarrow \text{b}\bar{\text{b}}$ search, and they are thus not correlated with the other searches [85]. The uncertainties in the PDFs and in the inclusive production cross-sections of the VBF, ggH, and VH modes are correlated across the data-taking eras. The full correlation study on all the remaining uncertainties is reported in Ref. [2].

The expected and observed upper limits on $\frac{\sigma}{\sigma_{\text{SM}}} \times \mathcal{B}(\text{H} \rightarrow \text{inv.})$ at 95% CL are shown in Figure 6.1. The limits are obtained for individual combinations of the qqH, VH, and ggH tagged categories, as defined in Table 6.1, and for the full combination, assuming $m_{\text{H}} = 125 \text{ GeV}$ and SM production rates. The observed (expected) upper limits for each category at 95% CL are as follows: 0.44(0.31) qqH-tagged, 0.39(0.50) VH-tagged, and 0.51(0.56) ggH-tagged. The VBF production mode is the most sensitive channel for Higgs to invisible searches, as described in Section 3.6. The regions containing 68% and 95% of the distribution of limits expected under the background-only hypothesis are $[0.23, 0.42]$ and $[0.17, 0.55]$, respectively. The result from this search is thus within the

‘95% region’, whose significance is mitigated during the full combination by the inverted observations from the other categories. In fact, the VH-tagged combined analyses show a limit lower than the expected upper limit, but within the ‘68% region’: $[0.36, 0.71]$. The mono-jet analysis measures an upper limit slightly below the expectation under the background-only hypothesis, whose ‘68% region’ is $[0.39, 0.84]$. The observed (expected) upper limit from the full combination is $\mathcal{B}(H \rightarrow \text{inv.}) < 0.24(0.23)$ at 95% CL.

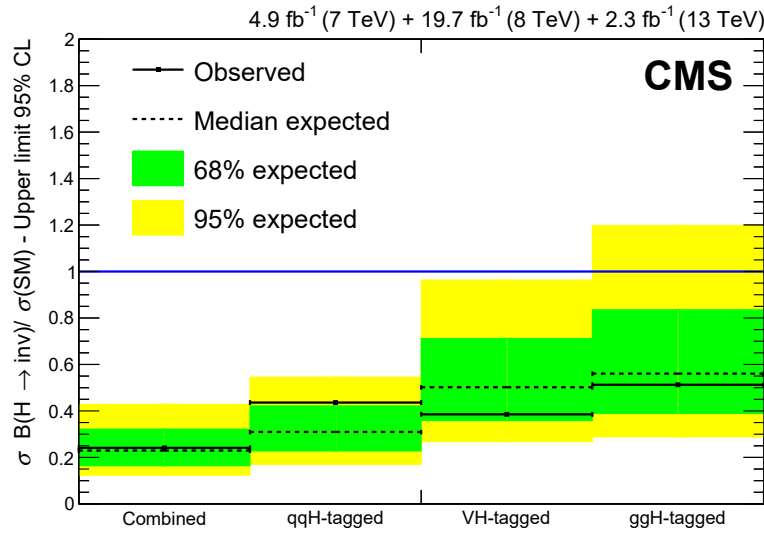


Figure 6.1: The observed (solid black) and expected (dashed black) upper limits at 95% CL on $\frac{\sigma}{\sigma_{\text{SM}}} \times \mathcal{B}(H \rightarrow \text{inv.})$ for individual combinations of categories, i.e. qqH, VH, and ggH tagged, and for the full combination, assuming $m_H = 125$ GeV. The inner (green) and outer (yellow) band indicate the regions containing 68% and 95%, respectively, of the distribution of limits expected under the background-only hypothesis [2].

Figure 6.2 shows the profile likelihood ratios as a function of $\mathcal{B}(H \rightarrow \text{inv.})$ for partial combinations of the Run-1 (7 + 8 TeV) and Run-2 2015 (13 TeV) searches, and for the full combination. The results are reported for the data, and for an Asimov dataset assuming no invisible decays of the Higgs boson. The sensitivity of the full combination is driven by the Run-1 searches, as shown by the steepness of the 7 + 8 TeV curve (solid red). The results at 13 TeV improve this sensitivity, and mitigate the observation from Run-1, whose best-fit value for $\mathcal{B}(H \rightarrow \text{inv.})$ is measured to be around 0.1.

Table 6.3 summarises the impact of the systematic and statistical uncertainties, and the total uncertainty on the fitted value of $\mathcal{B}(H \rightarrow \text{inv.})$ in the 13 TeV analyses of Table 6.1. The impact is calculated by varying each nuisance parameter independently within one standard deviation of its maximum-likelihood estimate, as presented in Section 5.9.1.

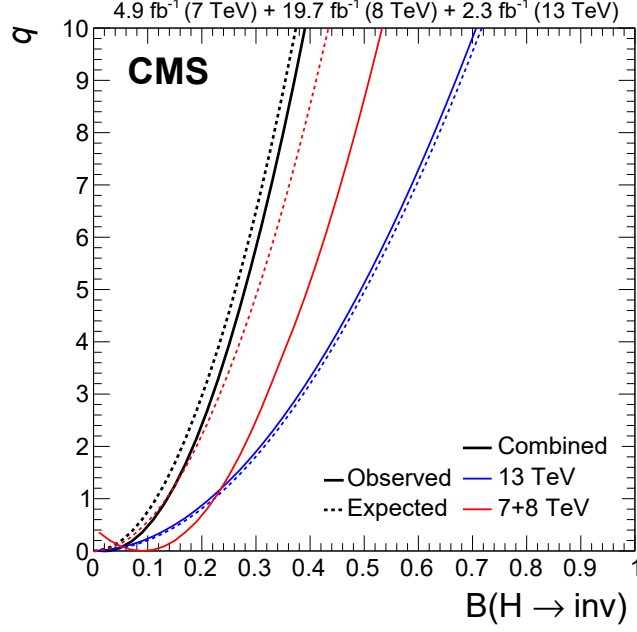


Figure 6.2: The observed (solid black) and expected (dashed black) profile likelihood ratios as a function of $\mathcal{B}(\text{H} \rightarrow \text{inv.})$, assuming SM production cross-sections of a Higgs boson with $m_{\text{H}} = 125 \text{ GeV}$. The results are reported for partial combinations of 7 + 8 TeV and 13 TeV searches and for the full combination [2].

The total statistical uncertainty is obtained by fixing all nuisance parameters to their maximum-likelihood estimates, whereas the total systematic uncertainty is calculated by subtracting in quadrature the statistical uncertainty from the total uncertainty. The statistical uncertainties refer to effects correlated with luminosity, and govern the sensitivity of the VBF and $\text{Z} \rightarrow \ell^+ \ell^-$ analyses. The $\text{V} \rightarrow \text{jj}$ and mono-jet analyses are dominated by systematic uncertainties on the $\text{Z}(\nu\bar{\nu}) + \text{jets}$ and $\text{W}(\ell^\pm \nu) + \text{jets}$ backgrounds.

Table 6.3: The impact of the systematic and statistical uncertainties, and the total uncertainty on the fitted value of $\mathcal{B}(\text{H} \rightarrow \text{inv.})$ in the 13 TeV analyses [2].

Final State	Impact of		
	Total Systematic	Total Statistical	Total Uncertainty
VBF	$+15\%$ -19%	$+28\%$ -27%	$+32\%$ -33%
$\text{Z} \rightarrow \ell^+ \ell^-$	$+27\%$ -23%	$+56\%$ -50%	$+62\%$ -55%
$\text{V} \rightarrow \text{jj}$	$+55\%$ -51%	$+50\%$ -46%	$+74\%$ -69%
Mono-jet	$+57\%$ -50%	$+25\%$ -22%	$+62\%$ -55%

6.3 Combination of Searches in Run-1 and Run-2 (2015+2016)

A combination of searches for invisibly decaying Higgs bosons, whose details are given in Ref. [1], was performed using only the 2016 dataset. Table 6.4 summarises the expected signal composition in each analysis, together with their individual upper limits on $\mathcal{B}(\text{H} \rightarrow \text{inv.})$. The VBF-tag search refers to the analysis presented in Chapter 5 using the shape-based approach. The $V \rightarrow \text{jj}$ and mono-jet analyses, and the $Z \rightarrow \ell^+ \ell^-$ analysis are not discussed in this thesis and further details can be found in Ref. [86] and Ref. [87], respectively. However, they can be considered as a significant improvement over the related 13 TeV analyses presented in Sections 6.2.2 and 6.2.3. The events of the $V \rightarrow \text{jj}$ and mono-jet analyses passing the VBF selection are removed to avoid double counting, and represent approximately 6(15)% of the total background for a 250(1000) GeV MET. As a consequence, a 5% loss in sensitivity compared to the results of Ref. [86] is measured.

Table 6.4: The expected composition of production modes of a 125 GeV Higgs boson, and observed and expected upper limits on $\mathcal{B}(\text{H} \rightarrow \text{inv.})$ for each analysis performed in 2016 and included in the combination, assuming SM production cross-sections [1].

Analysis	Final State	Expected Signal Composition (%)	Limit	
			Observed	Expected
qqH-tagged	VBF-jets	48 ggH, 52 VBF	0.33	0.25
VH-tagged	$Z \rightarrow \ell^+ \ell^-$	79 ZH, 21 ZH	0.40	0.42
	$V \rightarrow \text{jj}$	39 ggH, 6 VBF, 33 WH, 22 ZH	0.50	0.48
ggH-tagged	Mono-jet	80 ggH, 12 VBF, 5 WH, 3 ZH	0.66	0.59

A study on the correlation of systematic uncertainties between the searches was carried out, as motivated in Section 6.2.4, whose outcome is summarised in the following. The uncertainties accounting for veto efficiency of b-jet and τ_h candidates, integrated luminosity, momentum scale variations, and lepton and photon identification and reconstruction efficiencies are correlated among all searches. The uncertainties in the inclusive signal production cross-sections and in the cross-sections of VV and Top backgrounds are also correlated across the tagged categories. In contrast, systematic uncertainties related to jets are not correlated between the VBF search and the other categories since they have different jet kinematics.

The searches show no significant deviations from the SM expectations. The expected and observed 95% CL upper limits on $\frac{\sigma}{\sigma_{\text{SM}}} \times \mathcal{B}(\text{H} \rightarrow \text{inv.})$ are 0.20 and 0.26, respectively, assuming $m_{\text{H}} = 125.09 \text{ GeV}$ and SM production rates.

The results from the combination of these searches are further combined with the searches presented in Section 6.2, resulting in a grand combination detailed in Ref. [1]. An observed (expected) 95% CL upper limit is set at $\mathcal{B}(\text{H} \rightarrow \text{inv.}) < 0.19(0.15)$, which is the most sensitive limit to date on $\mathcal{B}(\text{H} \rightarrow \text{inv.})$ from either direct or indirect searches. The results from this combination are also used for interpretations under non-SM production assumptions and DM-Higgs portal models, as detailed in the following Sections.

6.3.1 Non-SM Production

The SM production rate assumptions govern the sensitivity of the searches considered in the combination. The VBF, VH, and ggH production cross-sections are parametrised in terms of κ_{F} and κ_{V} within the κ -framework, as presented in Section 3.6. These coupling strength modifiers scale the Higgs boson coupling to fermions and vector bosons, respectively, where $\kappa_{\text{F}} = \kappa_{\text{V}} = 1$ defines SM production rates [56].

Figure 6.4 shows the observed upper limit on $\frac{\sigma}{\sigma_{\text{SM}}} \times \mathcal{B}(\text{H} \rightarrow \text{inv.})$ at 95% CL calculated as a function of κ_{F} and κ_{V} . The best estimate from LHC and its 68% and 95% CL regions for the κ -model are also shown [57]. The observed (expected) 95% CL upper limit on $\mathcal{B}(\text{H} \rightarrow \text{inv.})$ varies in a range $[0.14, 0.24]([0.11, 0.19])$ within the 95% CL best-fit region from LHC. The coupling of the Higgs boson to vector bosons has a significantly larger impact on $\frac{\sigma}{\sigma_{\text{SM}}} \times \mathcal{B}(\text{H} \rightarrow \text{inv.})$ in comparison with the coupling to fermions. ggH and $\text{gg} \rightarrow \text{ZH}$ are the only processes contributing to κ_{F} , whereas the VBF and VH modes drive the impact of κ_{V} on the final limit due to the very high sensitivity.

6.3.2 DM Interpretations

Higgs-portal models of DM interactions [82], that theorise a stable DM particle coupling to the SM Higgs boson, are used to interpret the results of Section 6.3.

The interaction between an atomic nucleus and a DM particle may occur through the exchange of a Higgs boson. The direct-detection experiments are the most sensitive to signatures characterised by nuclear recoil, and the mass of the DM particle (m_{χ}) governs the sensitivity of these searches. A DM-nucleon scattering cross-section interpretation

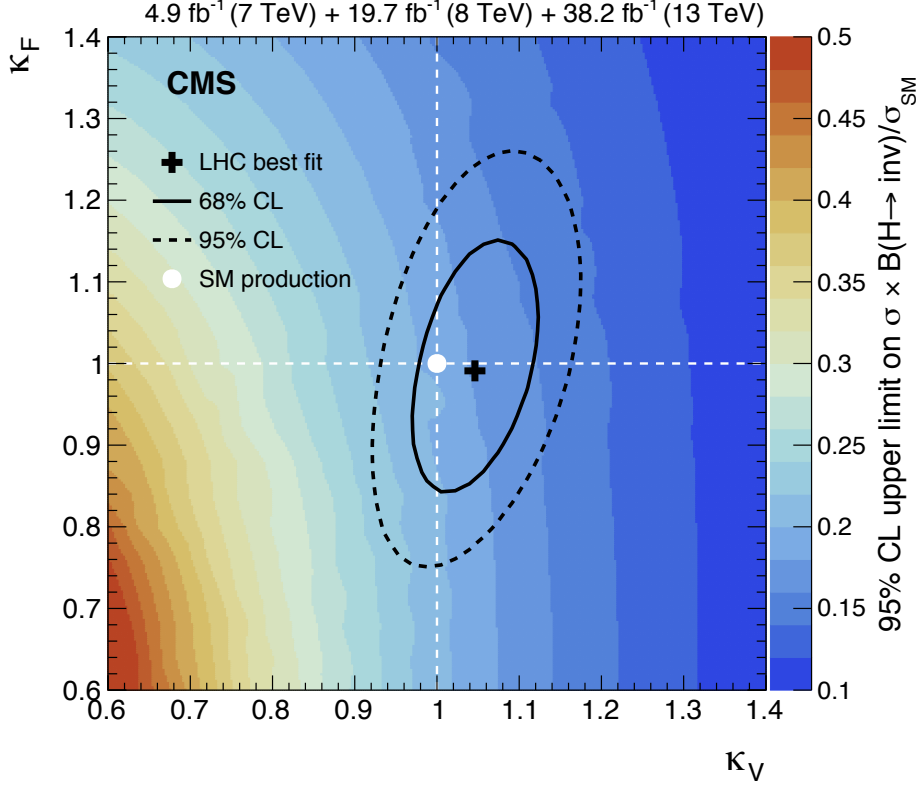


Figure 6.3: The observed 95% CL upper limits on $\frac{\sigma}{\sigma_{SM}} \times \mathcal{B}(H \rightarrow \text{inv.})$ as a function of κ_F and κ_V , assuming $m_H = 125.09 \text{ GeV}$. The best estimate from LHC and its 68% and 95% CL regions for the κ -model from Ref. [57] are reported. The SM production rates are obtained for $\kappa_F = \kappa_V = 1$ [1].

is used in this context to compare constraints from direct-detection experiments to the results of the combination of searches for invisibly decaying Higgs bosons.

The width of an invisibly decaying Higgs boson ($\Gamma_{\text{inv.}}$) is translated into a spin-independent DM-nucleon elastic scattering cross-section for $m_H > 2 m_\chi$ [82], using the relation [2]:

$$\mathcal{B}(H \rightarrow \text{inv.}) = \frac{\Gamma_{\text{inv.}}}{\Gamma_{SM} + \Gamma_{\text{inv.}}} , \quad (6.1)$$

with $\Gamma_{SM} = 4.07 \text{ MeV}$ [56]. This translation uses an effective field theory approach, and the cross-sections are calculated using the formula [2]:

$$\sigma_{DM-N}^{\text{SI}} = DM_c \cdot \frac{4\Gamma_{\text{inv.}}}{m_H^3 v^2 \beta} \frac{m_N^4 f_N^2}{(m_\chi + m_N)^2} . \quad (6.2)$$

For $\text{DM}_c = \frac{2m_\chi^2}{m_H^2\beta^2}$, a fermion DM candidate is assumed, whereas a scalar DM candidate is defined by $\text{DM}_c = 1$. The average between the proton and neutron masses is $m_N = 0.939 \text{ GeV}$, the Higgs vacuum expectation value is $v = 246 \text{ GeV}$, and $\beta = \sqrt{1 - \frac{4m_\chi^2}{m_H^2}}$. Finally, the nuclear form-factor $f_N = 0.308$ is taken from Ref. [88].

Figure 6.4 shows the 90% CL upper limit on the spin-independent DM-nucleon elastic scattering cross-section as a function of m_χ , for both fermion and scalar DM candidates. The limits from the XENON1T [89], PandaX-II [90], LUX [91], CDEX-10 [92], CDMSlite [93], and CRESST-II [94] direct-detection experiments are also reported. The observed 90% CL upper limit of $\mathcal{B}(\text{H} \rightarrow \text{inv.}) < 0.16$ represents the most stringent constraint for $m_\chi < 18(7) \text{ GeV}$ assuming a fermion (scalar) DM candidate. The combination in Section 6.3 is thus complementary to the sensitivity of XENON1T, PandaX-II, and LUX experiments in the low- m_χ region.

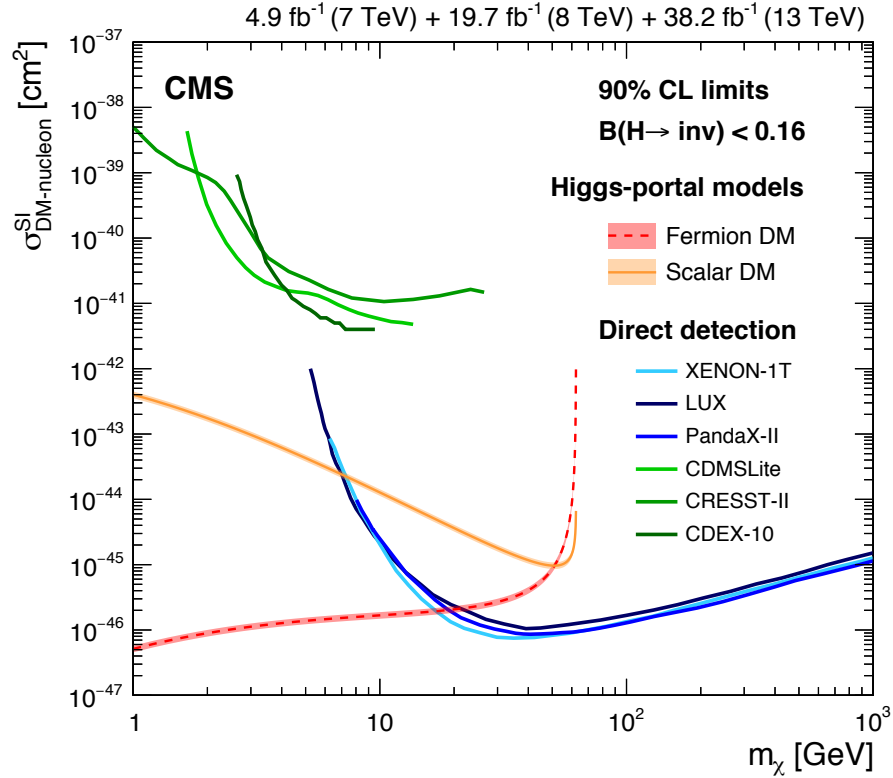


Figure 6.4: The 90% CL upper limit on the spin-independent DM-nucleon elastic scattering cross-section as a function of m_χ , for both fermion (dashed red) and scalar (solid orange) DM candidates in Higgs-portal models. The limits from the XENON1T [89], PandaX-II [90], LUX [91], CDEX-10 [92], CDMSlite [93], and CRESST-II [94] direct-detection experiments are also reported [1].

Chapter 7

Cloud Computing for High Energy Physics

This Chapter presents cloud computing within the CMS experiment and the effort to provide the dynamic provisioning of opportunistic resources. Section 7.1 gives an overview of the Grid computing at the LHC, while Section 7.2 describes the CMS computing model. Section 7.3 focuses on the dynamic provisioning of clusters and on the dynamic on demand analysis service (DODAS). The integrations with DODAS of the cloud resources at Imperial College London and as provided by the Amazon Web Services are detailed in Section 7.4 and Section 7.5, respectively. This work was performed by the author and is published in Ref. [3].

7.1 Grid Computing at the LHC

The data collected by the LHC experiments are analysed by physicists worldwide to look for new physics and perform precision measurements of the SM. An efficient data organisation, management, and access are required to experience low latencies which can affect the publication of physics results. Therefore, the success of the LHC programme relies on systems for data storage and management, and for workload management. However, numerous challenges are already posed by the future LHC upgrade, the High Luminosity phase (HL-LHC), that will cause the LHC experiments to enter the exabyte scale era.

The Worldwide LHC Computing Grid (WLCG) [33] is a collaboration with the purpose of building, managing, and maintaining the computing infrastructure required for the LHC operation. The data are handled by computing resources, services, and tools distributed worldwide through the ‘computing Grid’ paradigm. In fact, the Grid is a shared infrastructure designed to solve computing tasks. The computing centres work coherently, and end users interface to the whole distributed environment rather than the individual resources. Other Grid development projects, such as European Grid Infrastructure (EGI) [95] and Open Science Grid (OSG) [96], network providers, and commercial providers, such as Amazon Web Services (AWS) [97], collaborate with WLCG to support international research in many scientific disciplines across institutes, universities, and laboratories worldwide. However, the Grid computing infrastructure is not designed to cope with the large amount of data that will be produced by the HL-LHC, and new technologies need to be developed and integrated into the current computing model to overcome the upcoming challenges.

7.1.1 Services and Elements of the Grid

A customised computing model, that includes hardware and software components, operates 24/7 to fulfil the needs of the LHC experiments. The application layer of each experiment is hosted by a shared middleware in this distributed environment, significantly decreasing the costs of maintenance and upgrade of the resources. Several basic elements, each with a specific purpose, ensure an efficient operation of the Grid infrastructure. The logical elements for a generic Grid site are described in the following.

The generation or the processing of a certain amount of data defines a task. A job is defined as the unit resulting from splitting the task in equal parts, for easier and faster computation. The computing element (CE) service manages requests from users for computational power for their submitted jobs. It handles the load of job submissions that pile up in a queue, and interfaces to the relevant services in the Grid. A worker node (WN) is a single compute node of a Grid site, and is the unit that processes and computes the job. The storage element (SE) service stores data from processes or users on tapes for custodial storage, or on disk for quick access. The data can be accessed locally using site-dependent tools, and remotely via a common interface through the storage resource manager (SRM) protocol. A storage federation based on the XROOTD technology [98] is used in the specific case of CMS, and relies on ‘redirectors’ for the data location.

Central services are used to access and manage computing resources, and feature basic functions which the different experiments customise depending on the computing model implemented. The work presented in this Chapter necessitates the introduction of two of these services: the workload management system (WMS) [99, 100] and the virtual organisation management system (VOMS) [101]. WMS integrates tools and services to handle the processing of a workflow, including the submission and tracking of individual jobs. It receives updates on the status of the Grid and on job details to optimally distribute the workload across the sites. VOMS manages the authorisation of Grid users since it has information, roles, memberships, permissions, and privileges for each of them. A virtual organisation (VO) is defined as a group of individuals subject to rules and conditions determining the sharing of resources [102]. Finally, X.509 certificates [103] are used for secure authentication by both the users and services.

7.1.2 The Tiered Structure of the WLCG

A tiered structure is used to organise computing sites worldwide, with each ‘Tier’ providing different services [104]. Figure 7.1 gives an overview of the tiered structure of the WLCG within the CMS computing model [105, 106].

The Tier-0 (T0) is a unique logical computing centre that serves all LHC experiments. The CERN Data Centre and the Wigner Research Centre for Physics (Budapest) are the physical locations of the T0. The data collected by each experiment arrive at the T0 and are permanently stored on tape. An initial data processing is performed to provide feedback operations at the same rate at which data arrive from the online system of each experiment. The T0 also provides experiments with computing resources for simulation on an opportunistic basis.

The Tier-1 (T1) sites are large computing centres that provide additional storage for data arriving directly from the T0 and for simulated events produced at these tier-level sites. Data reprocessing is also performed and reliable services, such as databases and catalogues, are provided to the community through T1s.

The Tier-2 (T2) level represents the bulk of resources for analysis and simulation, and provides temporary data storage for analysis purposes.

A large number of smaller computing centres, referred to as Tier-3 (T3) sites, are hosted by universities and laboratories, fulfilling the requests of local users.

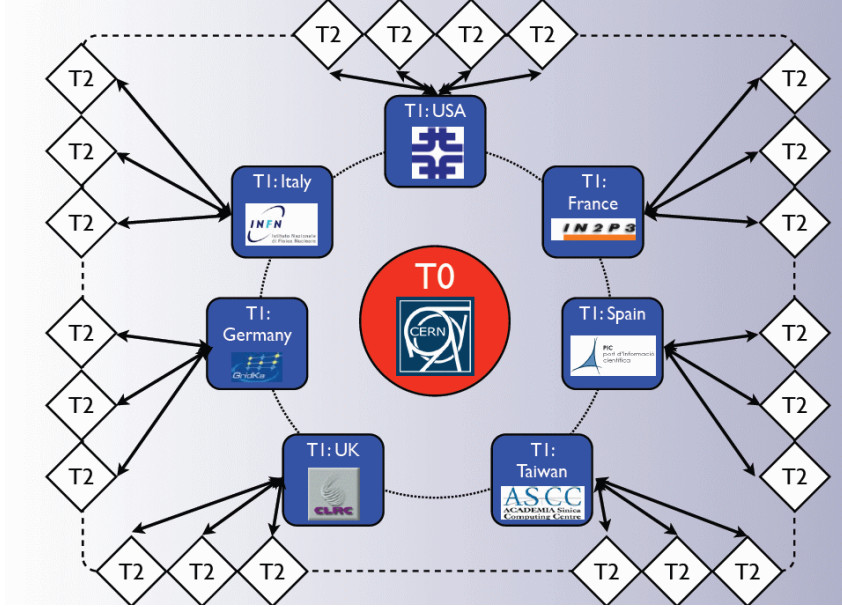


Figure 7.1: A schematic view of the tiered structure of the WLCG within the CMS computing model [107].

7.2 Computing in CMS

The CMS computing model [105, 106] is characterised by tasks and jobs running at sites where data are located. The CMS data management system [108] and workload management system [109] are designed to handle both automated and user requests across the Grid, and are detailed in the following Sections. The computing model of CMS generally follows the tiered structure within the WLCG with the caveat of the CMS-CERN Analysis Facility (CMS-CAF) [110].

7.2.1 The CMS Workload Management System

The CMS WMS is designed to distribute the workload on the Grid, minimising the job execution time and maximising the execution efficiency. The system is updated on the status and features of each site to efficiently distribute the workload across the Grid. WMS manages the submission to the queue of a target site using a placeholder-based mechanism. The system creates ‘pilot jobs’, or glideins (from here ‘glidein-WMS’ [100]), based on the demands of the submitted jobs, and submits them to the CE of the site. A HTCondor [99] overlay batch system is created when a pilot job starts running, with a ‘startd’ daemon that establishes a connection to the distributed global pool. As a result,

the WN is able to accept jobs from the global pool [111], whose HTCondor structure remains hidden to the experiment software framework.

The CMS Remote Analysis Builder (CRAB) tool [112, 113] is designed to manage the entire lifecycle of a job submitted on the Grid by a CMS user, allowing a physics analysis to be ‘distributed’ for fast processing [114]. The user does not necessarily know or understand the structure of the Grid as CRAB manages the creation, submission, monitoring, and deletion of the job. The connection with the logical Grid components is also handled by CRAB to access data worldwide regardless of their physical location.

7.2.2 The CMS Data Management System

The data management system includes catalogues and tools to locate, transfer, and access data, and relies on services from both the experiment and the Grid. Several requirements in terms of data security, reliability, and scalability are implemented to ensure effective monitoring and management of the data flow. The complexity of data management is reduced by defining high-level objects or formats: the data-tiers. These are used to group equivalent event information for both data and simulation with respect to the stage of processing in the reconstruction and simulation chain, respectively.

For the data workflow, the CMS TriDAS, detailed in Section 2.3.6, temporarily stores the events selected, transferring them to the T0. The RAW data contain low-level information from each sub-detector, and are grouped (repacking procedure) into primary datasets based on trigger information. A custodial copy of these data is made on tape libraries at the T0 and in at least one T1 to ensure that data are not lost if damaged or erased. A prompt calibration is performed at the T0 to get calibration constants for the reconstruction stage. Further calibration studies are carried out at a later stage using T2s. A prompt processing of RAW data at the T0 and T1s produces RECO data that can be used for analysis. Algorithms for tracking, identification, filtering, reconstruction, and compression are used in this step, producing physics objects such as leptons, photons, jets, and MET. The RECO datasets are distributed from the T0 to T1s that also host the corresponding RAW dataset, and elsewhere (T2s). The RECO data are typically updated with the latest calibration constants and software versions when these are available, using T1s. Further compression and skimming of this data format are performed at the T1-level to produce the analysis object data (AOD) data-tier. These data are stored at T1s and then distributed to T2s. However, users tend to use the highest-level data-tier, designed to be efficiently copied across the Grid: the mini-AOD [115]. These data contain

physics objects with several corrections applied, e.g. on MET and jets, and can be used to perform a physics analysis on a full data-taking run using T2 sites. A new data-tier, referred to as nano-AOD, is a reduced size format with respect to mini-AOD designed to deal with the possible data storage issues during the HL-LHC phase, starting after LS3 in Figure 2.3.

For the simulation workflow, the events are generated through Monte Carlo (MC) simulation techniques and collected in the GEN data-tier using T2s. The custodial storage of these data is offered at T1-level. The MC events are passed to GEANT4 [75] to simulate particle energy deposits in the detector, and are stored in the SIM data-tier. The response of the detector is simulated through a digitalisation step, leading to a DIGI output whose data-tier is similar to the RAW data from the detector. The generated MC samples are transferred from the T2 to the associated T1 site for distribution within the CMS Collaboration.

7.2.3 Cloud Computing and Opportunistic Resources

The computing infrastructure is designed to cope with the large amount of data produced by the LHC. However, breakneck use cases or large backlogs generated by peaks in usage can saturate the available resources and overload the infrastructure causing delays in the scheduled physics activity. Moreover, the HL-LHC phase will cause LHC experiments to enter the exabyte scale era. CMS plans to cope with the inevitable upcoming challenges by relying on the integration of heterogeneous resources and revolutionising the computing model as the optimisation of the current infrastructure will not be sufficient to meet the demands of the experiment. As an example, resources based on high performance computing (HPC) systems are integrated in CMS for this purpose [116, 117]. These are designed to overcome several limitations from physics problems that could be caused by the implementation of machine learning (ML) techniques or the requirement of high computing power.

Another possible solution to implement heterogeneous and dynamic resources is provided by cloud techniques. These can be used to provision configurable computing resources, characterised by high scalability and flexibility and purchased on-demand. Different services, such as network, storage, and applications, are easily managed without interacting with the provider. The characteristics of these services and resources fix the cost.

The services are provided implementing one of the following fundamental models. An infrastructure-as-a-service (IaaS) is typically made of bare virtual machines (VMs) and has basic functionalities customisable with operating systems and applications depending on the needs. A platform-as-a-service (PaaS) is a complex infrastructure that includes high-level services, such as databases and web servers, and is ready to run a user software framework or environment. A software-as-a-service (SaaS) provides users with on-demand services, such as software, applications, and databases, without any requirement on the management of the underlying infrastructure.

Clouds can be categorised into four groups: private, public, community, and hybrid. A private cloud infrastructure is operated and managed by a single organisation, typically a university or a laboratory. A public cloud infrastructure, which is usually commercial, provides services through an open network accessible by the worldwide community. A community cloud infrastructure shares several requirements, such as security, compliance, and jurisdiction, between organisations of the same community. A hybrid cloud infrastructure implements and offers services from the private, public, and community deployment models. The hybrid solution extends the capacity and capability of a service through customisation, aggregation, and integration.

The CMS experiment has explored several possibilities for the use of cloud techniques. A generic user is not aware of the type of infrastructure fulfilling the requests, i.e. cloud or Grid, since the target site offers resources via the Grid interface in both cases. The usage of the CMS HLT farm during LHC technical stops is an example of provisioning of on-demand resources for offline computing tasks using cloud technologies [118]. Among several pioneering activities on cloud solutions, the CMS-Bologna cloud working group designed and implemented two prototype computing services in 2014, referred to as ‘cloud bursting’ and ‘HEP computing-as-a-service’ [119, 120]. The first was designed to dynamically allocate resources within a cluster and on external and commercial clouds, whereas the second demonstrated a suitable use of direct access to external cloud resources. During this study, the very first CMS-wide CRAB (version 3) submission through glidein-WMS towards a cloud site was performed, considered an important milestone in the CMS computing activity and the precursor of following R&D activities.

The following Sections focus on a possible approach and effort in the design and operation of CMS computing solutions. The dynamic resource provisioning aims to face the challenges of the current LHC era and beyond.

7.3 Dynamic Resource Provisioning and Dynamic Clusters

CMS plans to use cloud technologies to provide dynamic opportunistic resources. The flexible provisioning of additional resources and sites to the existent pledged infrastructure is vital for all HEP experiments whose needs vary during the data-taking period.

The success and sustainability of opportunistic cloud solutions is ensured by decreasing the cost and time of management and maintenance for experiment-specific services in a Grid environment. The aim is to increase the efficiency of the setup and operation of a computing site. This can be achieved, for example, by provisioning on-demand clusters that use software applications to potentially benefit from any cloud provider. The aim of the work in this thesis is to implement this approach, as detailed in the following Sections.

7.3.1 Dynamic On-Demand Analysis Service

The dynamic on demand analysis service (DODAS) [121] is an open source PaaS tool and builds on European Open Science Cloud (EOCS) hub services [122] within the INDIGO-DataCloud project [123]. DODAS is developed and maintained by the Italian National Institute for Nuclear Physics (INFN), and deploys services to hybrid and heterogeneous clouds through on-demand container-based clusters. These clusters are instantiated using Apache MESOS [124] and potentially exploit any cloud software system, such as OpenStack [125] and AWS. This approach has a high level of abstraction and does not necessarily require one to know and understand the underlying technologies. The Alpha Magnetic Spectrometer (AMS-02, International Space Station) [126] and CMS integrate DODAS into the data analysis workflow and submission infrastructure (SI), respectively.

DODAS automates the provisioning procedure of a virtual hardware site. In fact, it creates, accesses, and manages heterogeneous resources, reducing the operational cost. DODAS also offers a high-level of customisation and scalability to implement experiment-specific requirements and services for the deployment of two cluster platforms: a HTCondor-based batch system-as-a-service, and a Big Data platform based on Apache Spark [127] for ML-as-a-service. As a result, the DODAS approach can be adopted for the exploitation of opportunistic resources, extension of existing computing sites, and deployment of on-demand batch systems for data processing.

Architecture

DODAS is characterised by a highly modular architecture whose main features are summarised in: multi-cloud support, automation, abstraction, and flexible authentication and authorisation infrastructure (AAI). These built on services developed by the INDIGO-DataCloud project, i.e. the identity and access management (IAM), token translation service (TTS), PaaS orchestrator, and infrastructure manager (IM), which together are the DODAS PaaS core service.

Figure 7.2 gives a schematic view of the DODAS architecture. The IAM server authenticates a user and generates an access token for the authorisation of the services to be requested. The user submits requests using a TOSCA template [128], which are accounted for by the PaaS orchestrator and IM [129]. These services create a cluster of containers over the IaaS using MESOS for the resource management. The single containers are orchestrated through MARATHON [130] that executes services and applications requested by the user. This configuration allows workflows to be highly customisable within DODAS, which consequently is able to integrate any external service and manage user-tailored code.

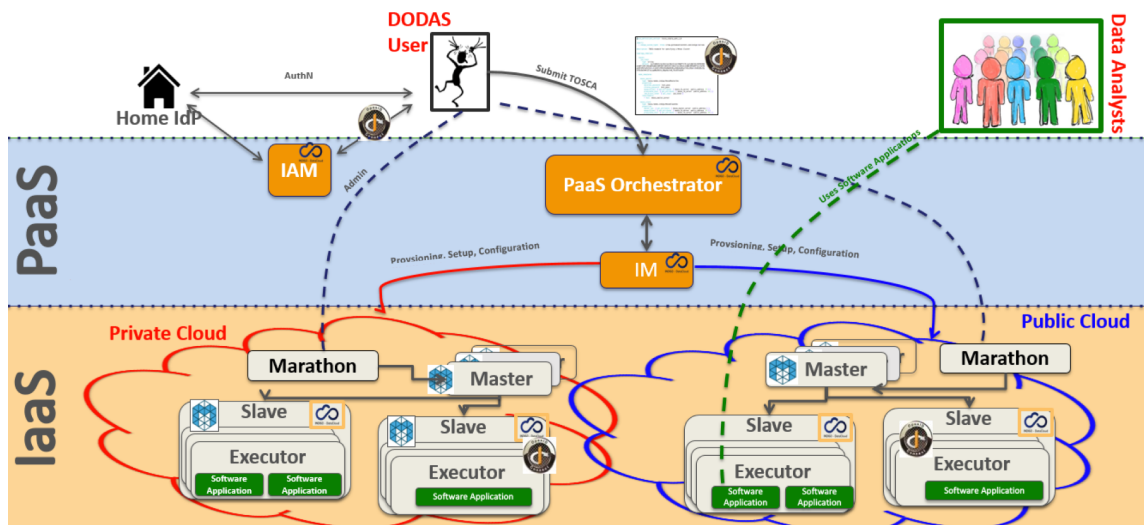


Figure 7.2: A schematic view of the DODAS architecture, where the IaaS resource (orange), PaaS core service (blue), and user domain (white) layers are shown [3].

Implementation

The CMS SI, namely the HTCondor global pool detailed in Section 7.2.1, integrates DODAS to generate ex-novo ephemeral Grid sites which potentially fulfil requirements ranging from a T3 to the T0 level.

A ‘vacuum’ model, characterised by the spontaneous instantiation of WNs, is implemented. The WN is a HTCondor startd process created as a Docker container [131] and configured as a MARATHON application. A service named PROXYCACHE integrates TTS to obtain the X.509 proxy certificate from the access token of the user. The certificate is cached and transferred to all WNs instantiated for the authentication of the ‘startd’ processes. This setup allows a WN to automatically join the CMS global pool. An additional authorisation layer is centrally managed by the global pool to map a distinguished name (DN) certificate to authorised workflows at a site. A proxy service named SQUID [132] caches the mapping of database queries to HTTP, and the CVMFS client [133] is mounted on the host. DODAS automatically installs all these services as containers and configures them as MARATHON applications.

On-going Improvements

The cloud provider used to implement the IaaS and PaaS solutions does not typically host the data to be processed. These data must be read from outside the computing site, possibly causing latency. This source of inefficiency is addressed in DODAS by proxy services, called XCACHE, between the remote storage and the site. This data ingestion technology is based on the XROOTD software within the CMS Anydata, Anytime, Anywhere (AAA) project [134], and uses a common namespace to federate the cluster of proxies to be implemented. The server managing the federation is named ‘redirector’. This handles requests to each target XCACHE server establishing a direct connection between the client and the proxy service that caches the file requested. In case none of the XCACHE servers host the file, an available server is chosen randomly to fulfil the request from the client using a ‘proxying while caching’ function. This approach is considered a dynamic caching of data between a bare central processing unit (CPU) and a remote storage system.

7.4 DODAS at Imperial College London, UK

The DODAS approach needed to be tested outside the development environment of INFN for validation. The aim was thus to demonstrate that it is a fully exportable solution and can benefit from any cloud provider. The cloud resources of the T-System Open Telekom Cloud (OTC), Instituto de Fisica de Cantabria (IFCA), and Imperial College London (ICL) were used to implement DODAS. The contribution from these experiences provided important feedback to validate and improve the tool, develop new features, and benchmark high energy physics use-cases.

This thesis details the ICL experience for both private and public cloud scenarios. A new opportunistic CMS Grid T3 site is created using DODAS on private resources hosted at ICL. The public resources of AWS are federated to the same Grid site to demonstrate the full exportability and to perform a scale test, as presented in Section 7.5.

7.4.1 Underlying Infrastructure and Implementation

The cloud infrastructure at ICL uses OpenStack as the software system, and experienced a change in capacity and configuration during the performing of the work presented, with the latest setup having: 500 instances, 768 virtual-CPU's, and 2.2 TB random access memory (RAM). The first version of the infrastructure was characterised by a shared file system hosting the VMs instantiated. The data ingestion and the connection to each VM was via a private virtual local area network (VLAN) at the backend. This configuration caused important inefficiencies in input/output (I/O) when more than 100 single-core WNs were created. The last version of the cloud infrastructure is not affected by this problem since each VM has its own volume. The credentials to access the infrastructure are referred to as USER and PASSWD.

The only requirement from DODAS is the presence of an UBUNTU-16.04 image, whose ID must be in the TOSCA template. The user sends a request to the IAM server as:

```
curl -s -L \
  -d client_id      = ${IAM_CLIENT_ID} \
  -d client_secret = ${IAM_CLIENT_SECRET} \
  -d grant_type     = password \
  -d username       = ${IAM_USER} \
  -d password       = ${IAM_PASSWORD} \
  -d scope          = "openid profile email offline_access" \
  ${IAM_ENDPOINT:-https://dodas-iam.cloud.cnaf.infn.it/token}
```

to retrieve an access token needed for the authorisation of the services to be requested. The token and the IAM-related configuration are passed through the TOSCA template to enable the translation from the authorisation server to the X.509 certificate. This approach, referred to as ‘password flow’, is exclusively used during the ‘development’ cycle. It is viable but limited since it must be enabled for each new client, authenticates the user with the local IAM credentials, and exposes the user credentials to the client application. A ‘device code flow’ approach, not described in this thesis, accounts for these limitations by using an external browser for authentication, ensuring the security of the system in a production environment.

DODAS defines several custom types imported in the TOSCA template to facilitate the formulation of requests, whose types are either integer or string. The experiment-specific configuration:

```
tosca_definitions_version: tosca_simple_yaml_1_0
...
topology_template:
  inputs:
# CMS specific configurations
    cms_local_site:
      type: string
      default: "T3_UK_Opportunistic"
    cms_stageoutsite:
      type: string
      default: "T2_UK_London_IC"
    cms_stageoutprotocol:
      type: string
      default: "srmv2"
    cms_stageoutcommand:
      type: string
      default: "gfal2"
    cms_phedexnode:
      type: string
      default: "T2_UK_London_IC"
...
    cms_services:
      type: tosca.nodes.indigo.CmsServices
      properties:
        marathon_password: { get_input: marathon_password }
        marathon_username: { get_input: marathon_username }
        mysquid_host: { get_attribute: [ mesos-lb-server, private_address, 0 ] }
        proxycache_host: { get_attribute: [ mesos-lb-server, private_address, 0 ] }
        iam_access_token: { get_input: iam_token }
        iam_client_id: { get_input: iam_client_id }
        iam_client_secret: { get_input: iam_client_secret }
        master_ips: { get_attribute: [ mesos-master-server, private_address ] }
        cms_local_site: { get_input: cms_local_site }
      requirements:
        - host: mesos_master
```

defines the new CMS Grid T3 site, that relies on the stageout site for storage running at ICL, called T2_UK_London_IC. This configuration requires the definition of the basic services needed by DODAS, as detailed in the following.

A master server:

```
mesos_master:
  type: tosca.nodes.indigo.MesosMaster
  properties:
    mesos_username: { get_input: mesos_username }
    mesos_password: { get_input: mesos_password }
    marathon_username: { get_input: marathon_username }
    marathon_password: { get_input: marathon_password }
    mesos_masters_list: { get_attribute: [ mesos-master-server, private_address ] }
  requirements:
    - host: mesos-master-server
...
mesos-master-server:
  type: tosca.nodes.indigo.Compute
  capabilities:
    endpoint:
      properties:
        network_name: PUBLIC
        dns_name: mesosserverpublic
      ports:
        mesos_port:
          protocol: tcp
          source: 5050
        marathon_port:
          protocol: tcp
          source: 8443
  scalable:
    properties:
      count: { get_input: number_of_masters }
  host:
    properties:
      num_cpus: { get_input: num_cpus_master }
      mem_size: { get_input: mem_size_master }
  os:
    properties:
      image: { get_input: server_image }
```

governs the cluster of containers to be created over the IaaS, and is the only service deployed on a public network. Multiple master servers ensure redundancy and reliability of the service, especially when a site must feature characteristics of a T2 or above. A T3 site such as T3_UK_Opportunistic does not require a high level of complexity. However, a container with a suitable amount of CPU and RAM is vital to manage a large number ($\mathcal{O}(10^3)$) of dependent services. For this reason, a single master server with 16 CPUs and 46 GB RAM was the minimum required for the T3_UK_Opportunistic site.

The master server controls two additional servers that are deployed on a private network. The first server balances the load (lb-server) and has a configuration analogous to the master. The second server, named ‘slave’, is necessary to implement the basic services required to build a CMS Grid T3 site, i.e. WNs, PROXYCACHE, SQUID, and CVMFS. The ICL cloud infrastructure is dimensioned to host a maximum of 45 slave-containers with 16 CPUs and 46 GB RAM, which are orchestrated by MARATHON to run the services as applications. The CMS-specific configuration attributes are defined in the TOSCA template to set the related dependencies with the T2_UK_London_IC stageout site.

```
cms_wn:
  type: toska.nodes.indigo.CmsWnConfig
  properties:
    mysquid_host: { get_attribute: [ mesos-lb-server, private_address, 0 ] }
    proxycache_host: { get_attribute: [ mesos-lb-server, private_address, 0 ] }
    cms_local_site: { get_input: cms_local_site }
    cms_stageoutsite: { get_input: cms_stageoutsite }
    cms_stageoutprotocol: { get_input: cms_stageoutprotocol }
    cms_stageoutcommand: { get_input: cms_stageoutcommand }
    cms_phedexnode: { get_input: cms_phedexnode }
  requirements:
    - host: mesos_slave
```

The monitoring of the MESOS cluster and the orchestration through MARATHON are performed using two web pages, whose attributes are defined in the TOSCA template:

```
outputs:
  mesos_lb_ip:
    value: { get_attribute: [ mesos-lb-server, public_address ] }
  mesos_endpoint:
    value: { concat: [ 'http://', get_attribute: [ mesos-master-server,
public_address, 0 ], ':5050' ] }
  marathon_endpoint:
    value: { concat: [ 'https://', get_attribute: [ mesos-master-server,
public_address, 0 ], ':8443' ] }
```

where the public addresses are set and retrieved during the deployment of the site. The TOSCA template is used during the deployment request by the site manager as:

```
curl -v -k \
-H 'Content-type: text/yaml' \
-H 'Authorization: id = os; \
type = OpenStack; \
host = https://wk00.grid.hep.ph.ic.ac.uk:5000/; \
username = '${USER}'; password = '${PASSWD}'; \
tenant = cmstest; \
auth_version = 3.x_password; \
domain = default;\nid = im; \
type = InfrastructureManager; \
token = '${TOKEN}'' \
-i -X POST https://im.cloud.cnaf.infn.it:443/infrastructures \
--data-binary "@ToscaICL_v3.YAML"
```

that also uses the infrastructure credentials, token, and end-points of the IM and IaaS. It is important to note that this approach is exclusively used by a developer since the security of the entire system could be compromised by such an authorisation method. The output of the request is a unique ID created for the infrastructure that is used to check the status of the cluster and of each container, retrieve internet protocol (IP) addresses for the DODAS services, and terminate the deployment.

The first deployments experienced a failure due to the configuration of the cloud infrastructure at ICL. The IM required IPs with a 128-bit precision (IPv6), whereas the infrastructure set IPs with a 32-bit precision (IPv4). The ICL experience highlighted this problem and supported the debugging work performed by the DODAS group. The IM was updated to support both configurations, leading to a successful deployment as experienced in the development environment of INFN.

7.4.2 CMS Workflows

The main objective of the activity presented in this thesis was to perform a functional test of DODAS outside the INFN development environment. This includes running requirement-specific workflows, detailed in the following and named WF-N, that a generic CMS user would need when carrying out a physics analysis.

WF-1 produces CMS data in nano-AOD format from mini-AOD, and about 5×10^3 jobs of this type were submitted while testing the T3_UK_Opportunistic site. This workflow is typically used within standard model physics analyses at ICL, and includes a slimming stage. WF-2 creates custom derivatives from mini-AOD, and is used to test the infrastructure with more than 2×10^4 jobs submitted. These files are an ICL skimmed version of the standard CMS format, and are used within the $H \rightarrow \tau^+ \tau^-$ analysis. Both workflows require 1 CPU and 2 GB RAM, representing the standard configuration used by a CMS user.

WF-3 performs the step from DIGI to AOD-SIM during the offline reconstruction stage of gluon-fusion and $ggH + 2$ jets events ($\mathcal{O}(10^3)$ jobs submitted). These simulated events were previously generated using the MADGRAPH generator [72, 73]. WF-4 produces CMS simulation in GEN-SIM format from LHE files [135], and less than 1×10^3 jobs were submitted. This workflow performs several steps, such as PYTHIA8 [74] parton shower, hadronisation and particle decays, and GEANT4 [75] detector simulation. Both workflows require 8 CPU and 16 GB RAM, and are typically managed by CMS centrally.

WF-5 creates LHE files for MADGRAPH $V \rightarrow jj$ processes, and $\mathcal{O}(10^2)$ jobs were submitted to test the infrastructure. The computation of this workflow is highly demanding, and so requires 15 CPU and 32 GB RAM.

7.4.3 Hammer Cloud Workflow

CERN IT developed a system for ATLAS to test Grid sites with custom workflows, referred to as Hammer Cloud (HC). The CMS and LHCb experiments also implement HC, adapted to their different computing models.

CRAB integrates HC to run jobs at all CMS sites up to the T2-level with the following objectives: heavily test a site infrastructure, continuously generate functional tests, and make performance studies of the CMS software. Specific properties, such as the input dataset, parameters of job splitting, and version and configuration of the CMS software, characterise a test. The HC workflow is submitted as a CRAB task to a set of sites by defining certain throttling parameters, e.g. maximum number of queued and running jobs. If the test is not completed within 24 hours, HC terminates it and submits another identical test. The CMS dashboard [136] is used to extract the success rate of HC jobs. This rate is inserted in the site status board and used as a quality metric.

HC workflows are typically used within CMS for quality checking of sites running glidein-WMS based workflows. They also comprise the first test required to validate new sites before entering the centralised production environment.

7.4.4 Results

The DODAS approach allows the instantaneous deployment of a Grid site to process requests on-demand. The new T3 Grid site, T3_UK_Opportunistic, requires a maximum of 15 minutes to be fully operative, depending on the size and number of services requested. Specifically, this deployment time is reached when services are mounted using the full ICL cloud infrastructure, i.e. 768 virtual-CPU's and 2.2 TB RAM. The instantiation of a Grid site in such a short time is considered an important achievement since it can serve several purposes on-demand, as detailed in the following.

Generic workflows requested by any CMS user on a daily basis, such as WF-1 and WF-2, are the core of CRAB submissions within the CMS experiment. The request to run 'non-standard' workflows, such as WF-3, WF-4, and WF-5 by generic CMS users is not

supported within the standard Grid computing model implemented by CMS. WF-3 and WF-4 must be managed centrally since they could affect the balance of the load on the whole Grid infrastructure. However, the T3_UK_Opportunistic site demonstrates the full integration of these workflows using DODAS. As a result, a DODAS site can be used for the full range of CMS workflows, from physics analyses to validation studies.

WF-5 requires a configuration of the WN not implemented by CMS. Similar WF-5 workflows, requiring a smaller WN configuration, are also generally discouraged by CMS since several users could end up requesting the same workflow, whose output is then created in multiple copies. However, physicists are often driven by conference or publishing constraints that necessitate the submission of such workflows, whose load on the computing infrastructure is not precisely predicted and accounted for. T3_UK_Opportunistic fulfils these requests by providing opportunistic resources without impacting the CMS computing infrastructure. Moreover, the WN configuration required by WF-5 is flexibly addressed by scaling the container according to the needs.

The T3_UK_Opportunistic site successfully ran and completed all workflows detailed in Section 7.4.2, experiencing no failure related to the site. For this reason, an additional effort was attempted towards the integration of CMS production workflows, i.e. glidein-WMS based workflows. This pioneering activity necessitated a collaboration with the CMS computing operation group since standard CMS policies are not applied in R&D prototyping. A feasibility study has started from this work to implement improved and customisable tools for quality checking of the resources, as discussed in the following.

As detailed in Section 7.4.3, HC workflows are largely used within CMS to test Grid sites and check their reliability. The T3_UK_Opportunistic site successfully ran and completed all HC workflows automatically submitted over a 6-month period. After terminating the deployment and instantiating a new one, the site started to receive jobs without any interaction. This behaviour is vital to ensure the opportunistic and dynamic provisioning of resources in a centralised production environment. Figure 7.3 shows the jobs running at T3_UK_Opportunistic between May 1st and 4th, 2019. During this period, only HC workflows were submitted to the site, and the visible characteristic pattern is caused by the intrinsic nature of HC, as detailed in Section 7.4.3.

A site within the centralised production environment must also run tests, referred to as SAM tests (site availability metrics), for reliability and availability checking, as required by CMS. These tests are centrally provided by CMS, but the ‘vacuum’ model implemented by DODAS was not initially designed to run such tests. Production jobs can be run

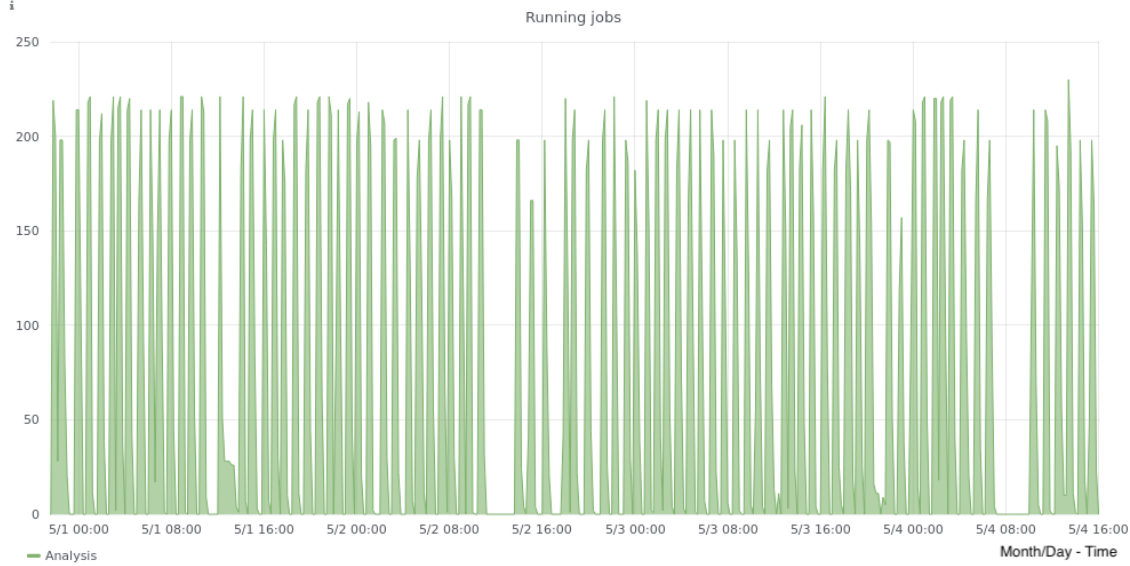


Figure 7.3: The jobs running at the T3_UK_Opportunistic site between May 1st and 4th, 2019.

without having a resource covered by SAM tests. However, these tests are a useful tool for the resource provider since they are designed to detect site problems causing production job failures. Since the ‘site failure mode’ does not necessarily highlight the specific issue, the absence of SAM tests would complicate the support of CMS operations on opportunistic resources. The work presented in this thesis led to customised SAM prototypes being integrated into DODAS, that are currently in the final stages of production implementation.

The DODAS approach is characterised by ‘startd’ processes launched within the Docker container, as detailed in Section 7.3.1. There is no policy for cloud-instantiated startds joining the CMS global pool, and the opportunistic site is completely independent on the choice of the number of processes to instantiate. The work in this thesis contributed to the implementation of a generalised solution for all cloud-slots, that also provides validation of opportunistic resources. The new method uses the so-called ‘glidein-wrapper’ to enclose the HTCondor together with the customised environment needed to efficiently run within CMS. As in the case of the SAM tests, the glidein-wrapper solution is currently in the final stage of integration within CMS, that requires a scale test for reliability checking of the resources before entering the production.

Finally, the full exportability of the T3_UK_Opportunistic site to cloud infrastructures different than the development one, i.e. ICL cloud, was demonstrated. The deployment of the site is independent of the physical location of the opportunistic resources, allowing

a ‘copy’ of T3_UK_Opportunistic to be instantiated on the cloud resources hosted at INFN-CNAF¹. The federation of ICL and INFN-CNAF clouds is transparent for both the developer and final user, and uses an overlay HTCondor where the site continues to rely on the T2_UK_London_IC stageout site. This result was necessary before validating the integration of heterogeneous opportunistic resources from an external commercial cloud provider, i.e. AWS.

7.5 Amazon Web Services for DODAS

The challenge for the high energy physics community is to access heterogeneous resources on-demand, instead of exploiting computing power from the static allocation of VMs as it happens in a standard commercial cloud. A collaboration between ICL and AWS allowed the exploitation of public cloud resources to prove the full exportability of the DODAS-based solution implemented in Section 7.4. The integration of these resources and the results from this work are described in the following Sections.

7.5.1 Implementation

The request of on-demand instances through the TOSCA template does not feature any change with respect to the procedure detailed in Section 7.4.1. The ID of the image to be requested, i.e. ‘aws://eu-west-1/ami-0773391ae604c49a4’, points towards a standard UBUNTU-16.04 image, and is provided in the AWS website. The string ‘eu-west-1’ refers to the region and endpoint hosting the physical resources, and the Ireland AWS region was chosen for this work as it provides the greatest resources in terms of quantity, size, and typology.

AWS offers an additional type of resources, which are referred to as ‘Spot instances’ and whose collection is indicated as ‘Spot fleet’. The user specifies the maximum price and the target capacity for a generic Spot fleet, whose request is fulfilled if it does not exceed the current price and available capacity of the Spot instances requested. For any change in price or availability after the deployment, the fleet attempts to maintain the target requested. A Spot instance is thus characterised by a significant lower cost compared to an on-demand instance. The work presented in this thesis used on-demand instances for

¹INFN-CNAF is the Italian National Centre for research and development in the disciplines of information technology and telematics.

the master and load balancer servers to always ensure availability and reliability of the resources. Spot instances are used for the slave server, which requires the addition of:

```

policies:
- set_availability_zone:
  type: tosca.policies.Placement
  properties: { availability_zone: 'eu-west-1c' }
  targets: [ mesos-slave-server ]

```

in the TOSCA template. The configuration of all servers and the CMS-specific requirement are detailed in Section 7.4.1

7.5.2 Results

The T3_UK_Opportunistic site was federated with cloud resources at either ICL or INFN-CNAF, validating the integration of heterogeneous opportunistic resources from the AWS commercial provider. The site successfully completed the workflows detailed in Section 7.4.2, experiencing no failure related to the site. As for the cloud federation discussed in Section 7.4.4, this result is fully transparent to the final user, and T3_UK_Opportunistic continues to rely on T2_UK_London_IC as the stageout site.

The result of the integration of AWS resources in CMS is vital for possible scenarios in which the CMS computing infrastructure must not be impacted. The cost to pledge additional resources for a short period of time, which absorb possible peaks in usage overloading the infrastructure, could be higher than exploiting opportunistic resources from a commercial provider. The DODAS-AWS approach of this work demonstrated the feasibility to create, federate, and use a Grid site on-demand, without any constraint on the physical location of the resources.

Figure 7.4 shows the jobs completed at T3_UK_Opportunistic between March 1st and May 4th, 2019. Different patterns are related to several types of workflows and tests submitted. These include results from both Section 7.4.4 and this Section.

Figure 7.5 shows the average CPU efficiency at T3_UK_Opportunistic between April 20th and May 10th, 2019. The amount of jobs during this period of time is not sufficient to perform an accurate and exhaustive study with respect to CPU efficiency, and the nature of the workflows and tests submitted significantly affect this calculation. However, T3_UK_Opportunistic was characterised by an efficiency between 60% and 80%, that is compatible with the average CPU efficiency of a generic CMS T2 site, demonstrating the validity of the DODAS-based solution.

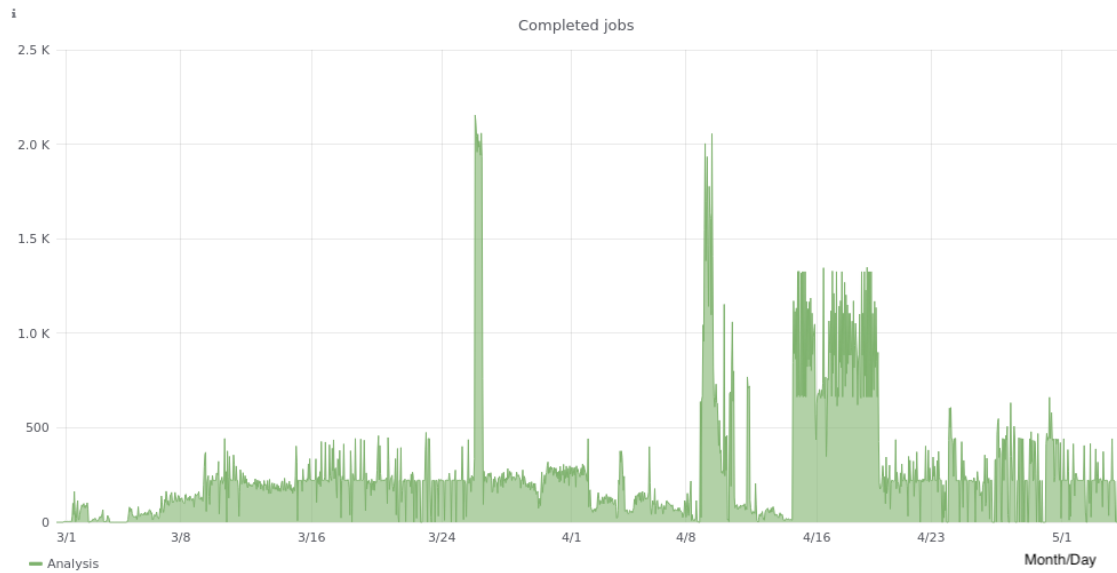


Figure 7.4: The jobs completed at the T3_UK_Opportunistic site between March 1st and May 4th, 2019.

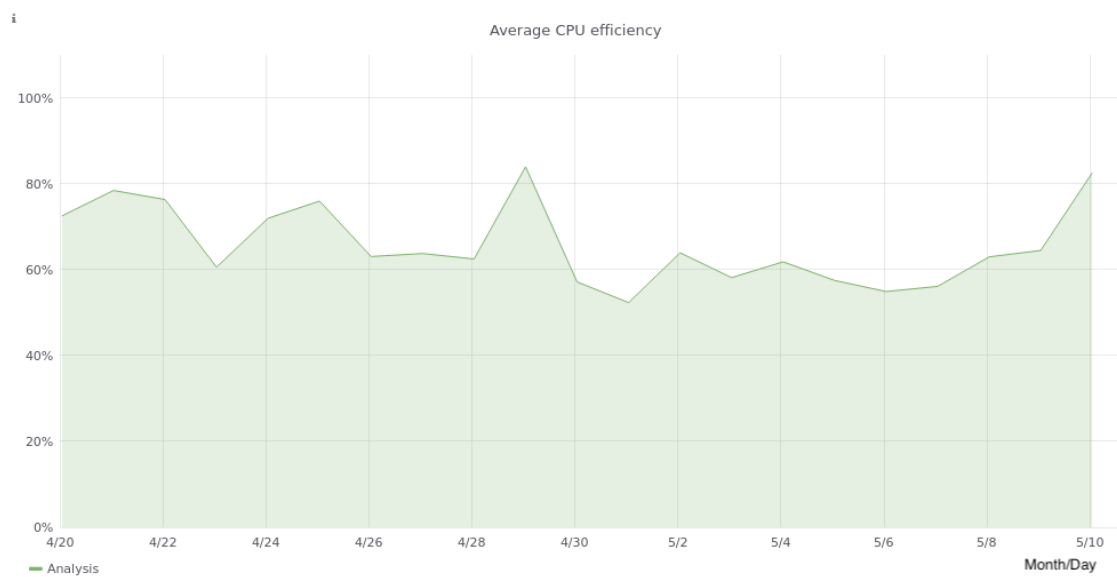


Figure 7.5: The average CPU efficiency at the T3_UK_Opportunistic site between April 20th and May 10th, 2019.

Chapter 8

Conclusions and Outlook

8.1 Physics Analyses

The standard model of particle physics is a highly successful theory explaining most of the phenomena observed in nature. However, the lack of a dark matter candidate is one of its important weaknesses, which the collider-based search programme at the LHC aims to address. Dark matter can arise from the decay of the Higgs boson and manifest itself as missing energy in the final state, which is ‘invisible’ for the CMS detector. The event is thus tagged by requiring additional physical objects in the final state, such as a VBF-jet pair. The decay of the Higgs boson to dark matter or non-SM particles contributes to $\mathcal{B}(H \rightarrow \text{inv.})$, which is about 10^{-3} in the SM. Thus, searching for invisibly decaying Higgs bosons is a powerful method to probe for dark matter.

In this thesis, a search for invisibly decaying Higgs bosons produced in the VBF mode, in proton-proton collisions recorded with the CMS detector during 2016 at $\sqrt{s} = 13$ TeV has been described. The VBF mode has the second highest production cross-section, but is the most sensitive channel for invisibly decaying Higgs boson searches. The analysis features two methods: the shape-based approach, which improves the sensitivity of the search by fully exploiting the VBF topology, and the cut-based (cut-and-count) approach, which features easy re-interpretations of the results using different phenomenological models. The cut-based approach is the main work conducted by the author, who also significantly contributed to the shape-based approach by performing synchronisation studies and background modelling. A reweighting procedure is performed to match simulation to data, and a data-driven method is used for the background prediction in the signal region for both approaches. Various improvements in comparison with the previous VBF Higgs to invisible analysis are made. New control samples in data were

introduced thereby further improving the precision of the background estimation. A simultaneous fit across the signal and control regions is performed for the extraction of the final limit. The data are compatible with the SM prediction, and the result is the most sensitive reported to date for either the VBF or any other individual production mode. The observed (expected) upper limits on $\mathcal{B}(H \rightarrow \text{inv.})$ are 0.58(0.30) and 0.33(0.25) at 95% CL for the cut-and-count and shape-based approaches, respectively. This result is published in Ref. [1] and was presented by the author on behalf of the CMS Collaboration at the conference ‘LHCP 2018’, as detailed in Ref. [137].

Combinations with other CMS analyses have also been described, which improve the sensitivity to invisibly decaying Higgs bosons. The combination of searches in Run-1 and in the first part of Run-2 (2015) was conducted by the author. The observed (expected) upper limit on $\mathcal{B}(H \rightarrow \text{inv.})$ is 0.24(0.23) at 95% CL, and the result is published in Ref. [2] and was presented by the author on behalf of the CMS Collaboration in a plenary talk at the conference ‘Higgs Couplings 2017’. The result of this combination is further combined with searches performed with the Run-2 (2016) dataset, including the search detailed in Chapter 5. The author was responsible for several aspects of this combination, such as studies of the correlations of the systematic uncertainties. The observed (expected) upper limit on $\mathcal{B}(H \rightarrow \text{inv.})$ is 0.19(0.15) at 95% CL, and the combination provides the most sensitive limit to date, as published in Ref. [1]. Interpretations under non-SM production assumptions and Higgs-portal models were also provided. The expected and observed 95% CL upper limits on $\frac{\sigma}{\sigma_{\text{SM}}} \times \mathcal{B}(H \rightarrow \text{inv.})$ vary from 0.11 to 0.19 and from 0.14 to 0.24, respectively, assuming non-SM production cross-sections within the κ -framework. The observed 90% CL upper limit of $\mathcal{B}(H \rightarrow \text{inv.}) < 0.16$ is translated into an upper limit on the spin-independent DM-nucleon elastic scattering cross-section. This limit is the most stringent constraint for $m_\chi < 18$ or < 7 GeV assuming a fermion or a scalar DM candidate, respectively.

Despite numerous searches, both direct and indirect, dark matter still remains undetected through non-gravitational interactions. Missing energy searches are among the most interesting to probe the existence of dark matter in light of the next LHC upgrades, where many of the existing physics analyses have to be re-thought to exploit the potential of the LHC to the maximum. Moreover, these analyses typically cover complementary phase-space to the direct detection experiments. However, searches for invisibly decaying Higgs bosons are heavily driven by the selections at the trigger level and are particularly challenging, requiring the use of cutting-edge techniques to fully exploit the LHC data. Designing and implementing ad-hoc optimised machine learning techniques at the trigger

level of CMS could minimise the thresholds, allowing the system to be more inclusive and flexible when triggering an event. Lowered thresholds would also improve background modelling and all additional ‘shape’ information to be used. As a consequence, the missing energy searches would be the first to benefit from this approach since an accurate modelling of the background is a key aspect to gain sensitivity. The designing and implementation of these machine learning techniques would require the exploitation of new computing solutions, such as the DODAS approach presented in this thesis, to avoid the CMS computing infrastructure being significantly affected.

Currently, data from the 2017 and 2018 data-taking periods are being analysed. In addition, a search for invisibly decaying Higgs boson, produced in the VBF mode, has been investigated using an event selection optimised for the HL-LHC conditions. The expected upper limit on $\mathcal{B}(H \rightarrow \text{inv.})$ for an integrated luminosity of 3000 fb^{-1} is 0.038 at 95% CL, requiring $m_{jj} > 2500 \text{ GeV}$ and $E_T^{\text{miss}} > 190 \text{ GeV}$, as published in Ref. [138]. This prediction illustrates the potential of Higgs to invisible searches for DM or new physics more generally. However, new phenomenological models need to be developed to realise the full potential for DM searches.

8.2 Cloud Computing

High energy physics at the LHC relies on the optimal handling of such a vast quantity of data, as collected by CMS, and the computing activities are the core to efficiently perform physics analyses. Although the computing infrastructure is designed to cope with the data produced by the LHC, peaks in usage can saturate the available resources, causing delays in the scheduled physics activity. CMS implements cloud technologies for the dynamic provisioning of opportunistic resources, to which DODAS significantly contributes by using a ‘vacuum’ model.

In this thesis, R&D activities performed by the author within the Computing & Offline Project at CMS have been described. The pioneering implementation of DODAS in different cloud environments and the benchmarking of high energy physics use-cases were carried out by the author, and the results are published in Ref. [3]. The cloud resources at ICL were used to validate DODAS outside the INFN development environment. This highlighted different problems, dealt with by the author and the DODAS working group. A newly deployed Grid site, i.e. T3_UK_Opportunistic, successfully ran and completed all user and HC workflows without reporting failures related to the site. This important

milestone in CMS was the precursor of feasibility studies for the integration of CMS production workflows. The work in this thesis contributed to design prototypes for the glidein-wrapper and SAM tests for the reliability and availability of opportunistic resources. These solutions are currently in the final stages of implementation in CMS, and will require a scale test before entering production.

Finally, the full exportability of the DODAS-based solution, which exploits the collaboration between ICL and UK's Amazon Web Services, has been described. The integration of DODAS with commercial public resources was conducted by the author. T3_UK_Opportunistic successfully ran and completed all user workflows without reporting failures related to the site, and the federation of AWS with ICL and INFN-CNAF resources demonstrated the transparency of the result to the end user. The outcome of this work outlines the feasibility for the dynamic provisioning of cloud resources in the Grid on-demand and without any constraint on their physical location.

With HL-LHC, high energy physics will enter the exabyte scale era. A significant contribution to the access, management, and organisation of data is necessary to cope with the upcoming challenges, and cutting-edge computing solutions must be devised to face network, CPU, and data co-location problems. This data revolution will require extended use of the cloud, and either new strategies or tools to easily find, access, and reuse the data, and will encourage the whole scientific community to push forward the boundaries of fundamental science.

Appendix A

Z Control Regions

The double-lepton CRs, i.e. $Z \rightarrow \mu^+\mu^-$ and $Z \rightarrow e^+e^-$, are used to estimate the background from $Z(\nu\bar{\nu})$ +jets processes in the SR using a data-driven method. However, the event yields in these regions are considerably smaller than the $Z(\nu\bar{\nu})$ +jets contribution in the SR, as shown in Table 5.5. For this reason, the single-lepton CRs are jointly used to predict the $Z(\nu\bar{\nu})$ +jets contribution in the SR.

Figure A.1 shows the distribution of m_{jj} for both data and simulated signal and backgrounds in the double-lepton CRs. Both regions show a reasonable agreement between data and MC given the available number of events passing the selection. The shape of the m_{jj} distribution is also used in a simultaneous fit with the SR and CRs to constrain the MC to data for the extraction of the final result in the shape-based approach.

Figure A.2 (Figure A.3) shows the distributions of p_T and η for the leading and sub-leading muons (electrons) for both data and simulated signal and backgrounds in the $Z \rightarrow \mu^+\mu^-$ ($Z \rightarrow e^+e^-$) CR. These distributions provide topological information on Z +jets processes, and overall the data are reasonably consistent with simulation given the finite statistical power of the double-lepton CRs.

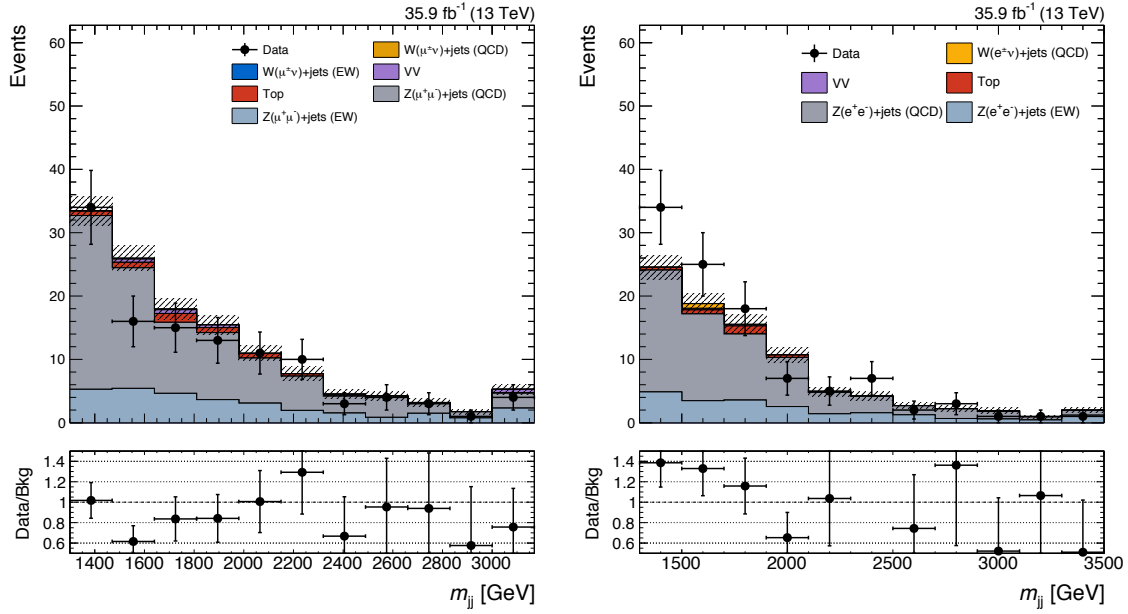


Figure A.1: The distribution of m_{jj} for data and simulated signal and backgrounds in the $Z \rightarrow \mu^+\mu^-$ (left) and $Z \rightarrow e^+e^-$ (right) CRs.

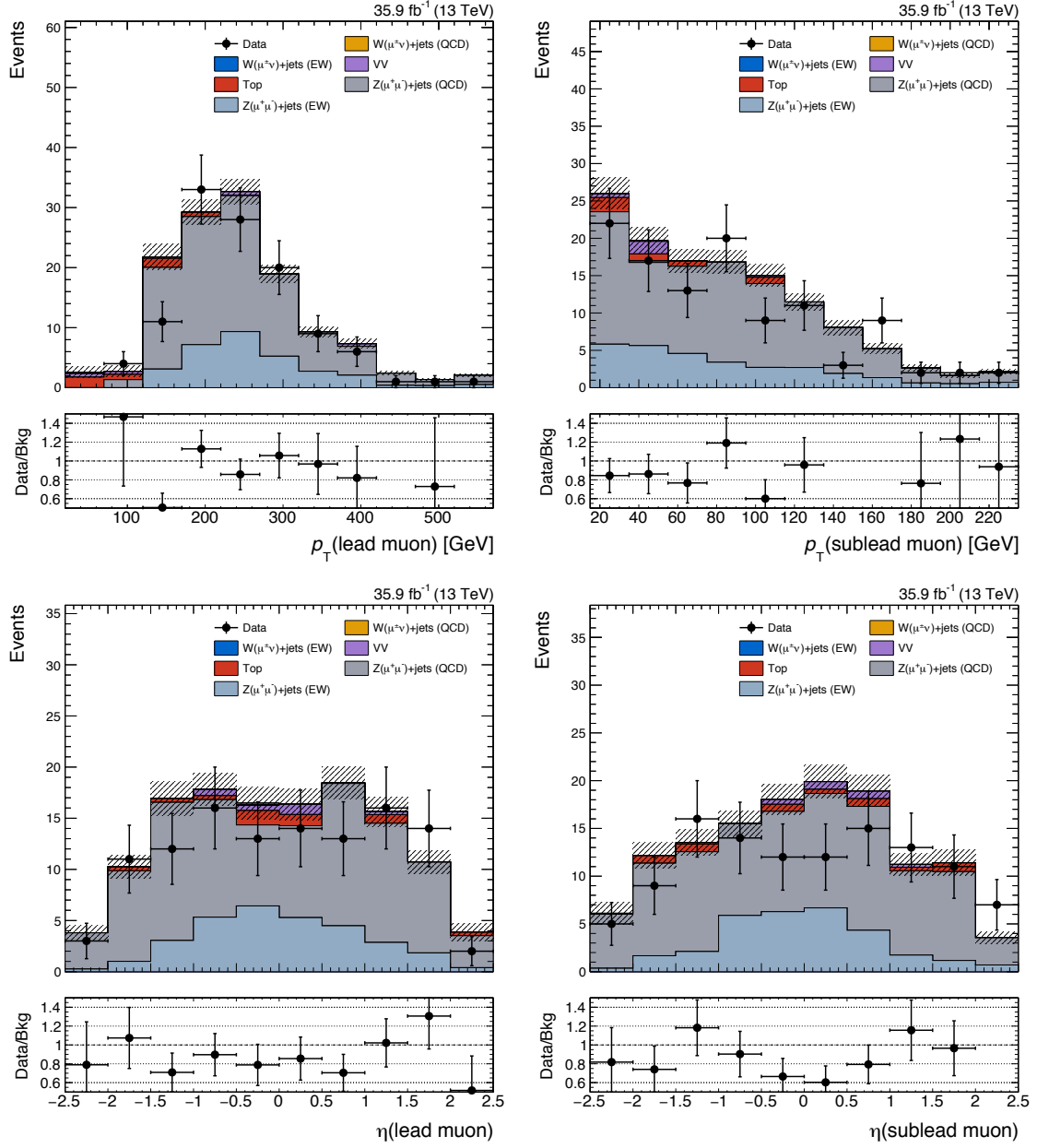


Figure A.2: The distributions of p_T (top) and η (bottom) for the leading (left) and sub-leading (right) muons for data and simulated signal and backgrounds in the $Z \rightarrow \mu^+ \mu^-$ CR.

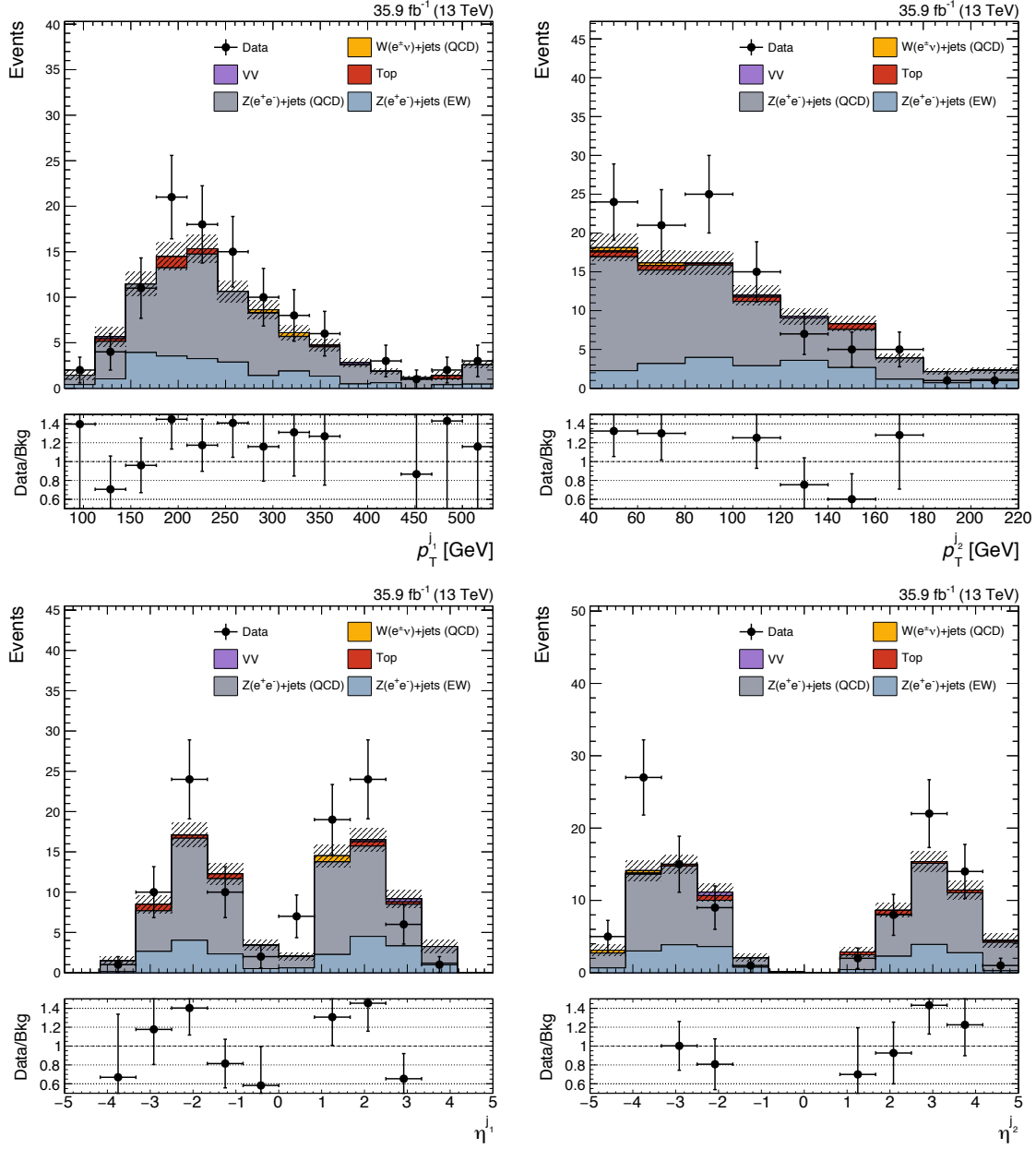


Figure A.3: The distributions of p_T (top) and η (bottom) for the leading (left) and sub-leading (right) electrons for data and simulated signal and backgrounds in the $Z \rightarrow e^+e^-$ CR.

Appendix B

W Control Regions

The single-lepton CRs, i.e. $W \rightarrow \mu^\pm \nu$ and $W \rightarrow e^\pm \nu$, are used to estimate the background from $W(\ell^\pm \nu)$ +jets processes in the SR using a data-driven method. These regions have high statistical power since a large number of events pass the selection, as shown in Table 5.5. Thus, they are also used to predict the $Z(\nu\bar{\nu})$ +jets contribution in the SR together with the double-lepton CRs.

Figure B.1 shows the distribution of m_{jj} for both data and simulated signal and backgrounds in the single-lepton CRs. The agreement between data and MC is significantly improved in comparison with the double-lepton CRs. The simultaneous fit across the SR and CRs uses information from the shapes of this distribution in the two regions to constrain the MC to data for the extraction of the final result in the shape-based approach.

Figure B.2 (Figure B.3) shows the distributions of p_T and η for the leading muon (electron) for both data and simulated signal and backgrounds in the $W \rightarrow \mu^\pm \nu$ ($W \rightarrow e^\pm \nu$) CR. These distributions provide topological information on W +jets processes, and overall the data are consistent with simulation.

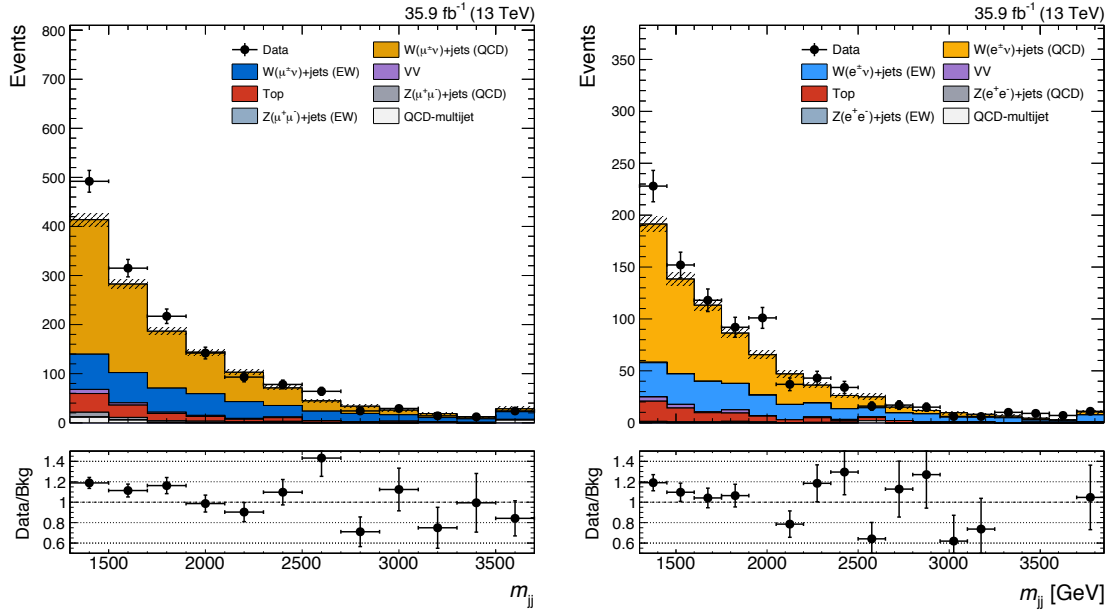


Figure B.1: The distribution of m_{jj} for data and simulated signal and backgrounds in the $W \rightarrow \mu^\pm \nu$ (left) and $W \rightarrow e^\pm \nu$ (right) CRs.

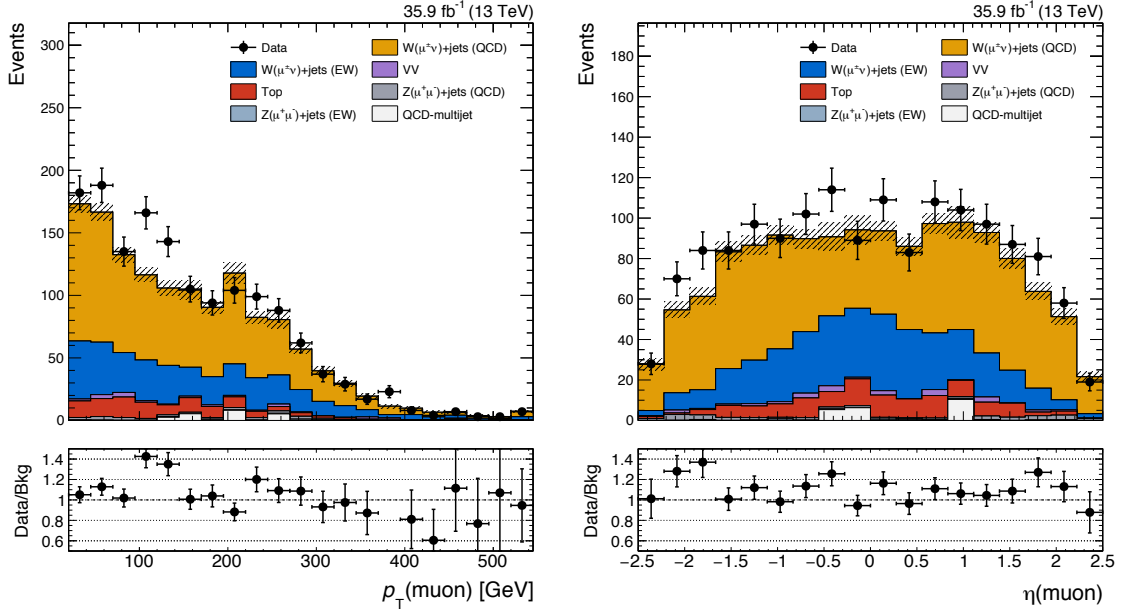


Figure B.2: The distributions of p_T (left) and η (right) for the leading muon for data and simulated signal and backgrounds in the $W \rightarrow \mu^\pm \nu$ CR.

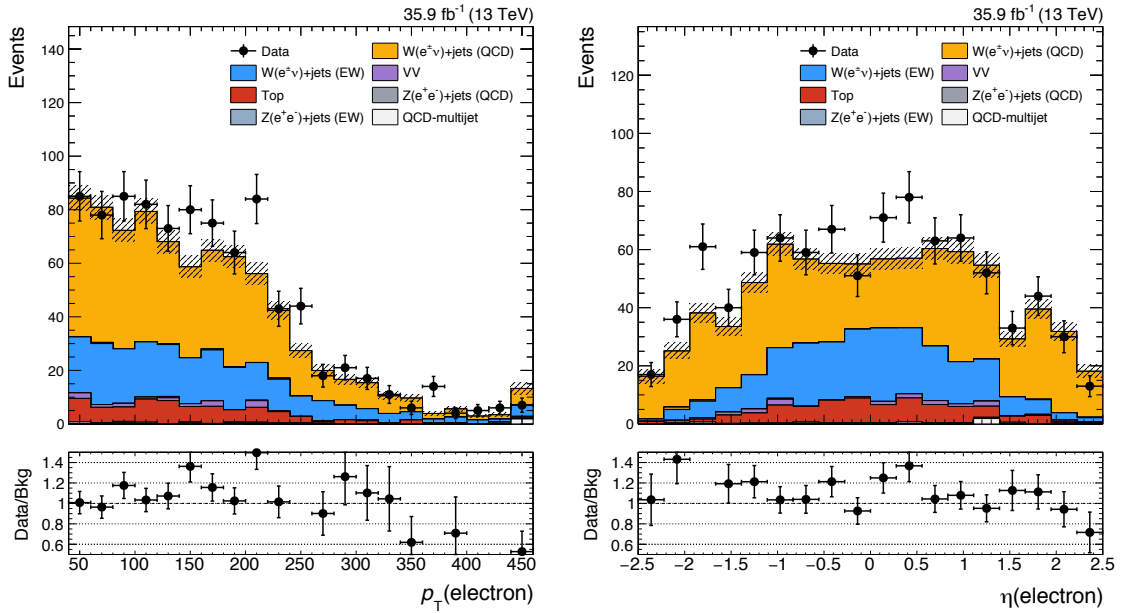


Figure B.3: The distributions of p_T (left) and η (right) for the leading electron for data and simulated signal and backgrounds in the $W \rightarrow e^\pm \nu$ CR.

Bibliography

- [1] CMS Collaboration. Search for Invisible Decays of a Higgs Boson Produced through Vector Boson Fusion in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett.*, B793:520–551, 2019.
- [2] CMS Collaboration. Searches for Invisible Decays of the Higgs Boson in pp Collisions at $\sqrt{s} = 7, 8$, and 13 TeV. *JHEP*, 02:135, 2017.
- [3] D. Spiga. Exploiting Private and Commercial Clouds to Generate On-Demand CMS Computing Facilities with DODAS. Technical Report CMS-CR-2018-314, CERN, Geneva, 2018.
- [4] O. S. Brüning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004. CERN-2004-003-V-1.
- [5] *LEP Design Report*. CERN, Geneva, 1984. CERN-LEP-84-01.
- [6] R. Steerenberg. LHC Report: Full House for the LHC. CERN, Geneva, 2017.
- [7] E. Halkiadakis. Introduction to the LHC Experiments. In *Physics of the Large and the Small, TASI 09, Proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics, Boulder, Colorado, USA, 1-26 June 2009*, pages 489–518, 2011.
- [8] ALICE Collaboration. The ALICE Experiment at the CERN LHC. *JINST*, 3:S08002, 2008.
- [9] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.
- [10] CMS Collaboration. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.
- [11] LHCb Collaboration. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.

- [12] H. T. Edwards. The Tevatron Energy Doubler: A Superconducting Accelerator. *Ann. Rev. Nucl. Part. Sci.*, 35:605–660, 1985.
- [13] ATLAS Collaboration. Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC. *Physics Letters B*, 716(1):1 – 29, 2012.
- [14] CMS Collaboration. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Physics Letters B*, 716(1):30 – 61, 2012.
- [15] CMS Collaboration. CMS Luminosity - Public Results. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#2016_proton_proton_collisions.
- [16] G. Apollinari, O. Brüning, T. Nakamoto, and Lucio Rossi. High Luminosity Large Hadron Collider HL-LHC. *CERN Yellow Report*, (5):1–19, 2015. FERMILAB-DESIGN-2017-02.
- [17] A. Hoecker. Physics at the LHC Run-2 and Beyond. In *Proceedings, 2016 European School of High-Energy Physics (ESHEP2016): Skeikampen, Norway, June 15-28 2016*, 2016.
- [18] CMS Collaboration. CMS Physics: Technical Design Report Volume 1: Detector Performance and Software. CERN-LHCC-2006-001, CMS-TDR-8-1.
- [19] CMS Collaboration. CMS Physics : Technical Design Report Volume 2: Physics Performance. *J. Phys.*, G34(6):995–1579, 2007.
- [20] CMS Collaboration. Performance of the CMS Drift-Tube Local Trigger with Cosmic Rays. *JINST*, 5:T03003, 2010.
- [21] CMS Collaboration. Interactive Slice of the CMS detector. CERN, Geneva, 2016. CMS-OUTREACH-2016-027.
- [22] CMS Collaboration. The CMS Tracker System Project : Technical Design Report. CERN, Geneva, 1997. CERN-LHCC-98-006, CMS-TDR-5.
- [23] CMS Collaboration. Description and Performance of Track and Primary-Vertex Reconstruction with the CMS Tracker. *JINST*, 9(10):P10009, 2014.
- [24] CMS Collaboration. The CMS Electromagnetic Calorimeter Project : Technical Design Report. CERN, Geneva, 1997. CERN-LHCC-97-033, CMS-TDR-4.

- [25] CMS Collaboration. Energy Resolution of the Barrel of the CMS Electromagnetic Calorimeter. *JINST*, 2:P04004, 2007.
- [26] CMS Collaboration. The CMS Hadron Calorimeter Project : Technical Design Report. CERN, Geneva, 1997. CERN-LHCC-97-031, CMS-TDR-2.
- [27] CMS HCAL Collaboration. Design, Performance, and Calibration of CMS Hadron-Barrel Calorimeter Wedges. *Eur. Phys. J.*, C55(1):159–171, 2008.
- [28] CMS Collaboration. The CMS Magnet Project : Technical Design Report. CERN, Geneva, 1997. CERN-LHCC-97-010, CMS-TDR-1.
- [29] CMS Collaboration. The CMS Muon Project : Technical Design Report. CERN, Geneva, 1997. CERN-LHCC-97-032, CMS-TDR-3.
- [30] CMS Collaboration. The CMS TriDAS Project : Technical Design Report, Volume 1: The Trigger Systems. CERN, Geneva, 2000. CERN-LHCC-2000-038, CMS-TDR-6-1.
- [31] CMS Collaboration. The CMS TriDAS Project : Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CERN, Geneva, 2002. CERN-LHCC-2002-026, CMS-TDR-6.
- [32] A. Tapper. The CMS Level-1 Trigger for LHC Run II. *Journal of Instrumentation*, 13(02):C02040–C02040, 2018.
- [33] I. Bird, K. Bos, N. Brook, D. Duellmann, C. Eck, I. Fisk, D. Foster, B. Gibbard, C. Grandi, F. Grey, et al. LHC Computing Grid: Technical Design Report. CERN, Geneva, 2005. CERN-LHCC-2005-024, LCG-TDR-001.
- [34] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579, 1961.
- [35] S. Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.
- [36] A. Salam. Weak and Electromagnetic Interactions. In *Elementary Particle Physics: Relativistic Groups and Analyticity, Proceedings of the Eighth Nobel Symposium*, page 367, Stockholm, 1968. Almquist & Wiskell.
- [37] M. Gell-Mann and Y. Ne’eman. *The Eightfold Way: a Review, with a Collection of Reprints*. Frontiers in Physics. W.A. Benjamin, 1964.
- [38] E. Noether. Invariant Variation Problems. *Gott. Nachr.*, 1918:235–257, 1918. [Transp. Theory Statist. Phys.1,186(1971)].

-
- [39] C. N. Yang and R. L. Mills. Conservation of Isotopic Spin and Isotopic Gauge Invariance. *Phys. Rev.*, 96:191–195, 1954.
- [40] Particle Data Group. Review of Particle Physics. *Chin. Phys.*, C40(10):100001, 2016.
- [41] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321, 1964.
- [42] P. W. Higgs. Broken Symmetries, Massless Particles and Gauge Fields. *Phys. Lett.*, 12:132, 1964.
- [43] P. W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508, 1964.
- [44] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global Conservation Laws and Massless Particles. *Phys. Rev. Lett.*, 13:585, 1964.
- [45] P. W. Higgs. Spontaneous Symmetry Breakdown without Massless Bosons. *Phys. Rev.*, 145:1156, 1966.
- [46] T. W. B. Kibble. Symmetry Breaking in Non-Abelian Gauge Theories. *Phys. Rev.*, 155:1554, 1967.
- [47] C. Quigg. Gauge Theories of the Strong, Weak and Electromagnetic Interactions. *Front. Phys.*, 56:1–334, 1983.
- [48] J. Ellis. Higgs Physics. In *Proceedings, 2013 European School of High-Energy Physics (ESHEP 2013): Paradfurdo, Hungary, June 5-18, 2013*, pages 117–168, 2015.
- [49] K. Freese. Status of Dark Matter in the Universe. *Int. J. Mod. Phys.*, 1(06):325–355, 2017.
- [50] Super-Kamiokande Collaboration. Evidence for Oscillation of Atmospheric Neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998.
- [51] D. Griffiths. Introduction to Elementary Particles. *Weinheim, Germany: Wiley-VCH (2008) 454 p*, 2008.
- [52] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky. A Direct Empirical Proof of the Existence of Dark Matter. *Astrophys. J.*, 648:L109–L113, 2006.

- [53] LHC Higgs Cross Section Working Group. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. *CERN Yellow Report*, CERN, Geneva, 2016. FERMILAB-FN-1025-T, CERN-2017-002-M.
- [54] CMS Collaboration. Measurements of Higgs Boson Production via Gluon Fusion and Vector Boson Fusion in the Diphoton Decay Channel at $\sqrt{s} = 13$ TeV. CERN, Geneva, 2019. CMS-PAS-HIG-18-029.
- [55] D. J. Gross and F. Wilczek. Ultraviolet Behavior of Non-Abelian Gauge Theories. *Phys. Rev. Lett.*, 30:1343–1346, Jun 1973.
- [56] LHC Higgs Cross Section Working Group. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. *CERN Yellow Report*, CERN, Geneva, 2013. CERN-2013-004, FERMILAB-CONF-13-667-T.
- [57] ATLAS and CMS Collaborations. Measurements of the Higgs Boson Production and Decay Rates and Constraints on its Couplings from a Combined ATLAS and CMS Analysis of the LHC pp Collision Data at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 08:045, 2016.
- [58] CMS Collaboration. Particle-Flow Reconstruction and Global Event Description with the CMS Detector. *JINST*, 12(10):P10003, 2017.
- [59] R. Fruhwirth, W. Waltenberger, and P. Vanlaer. Adaptive Vertex Fitting. *J. Phys.*, G34:N343, 2007.
- [60] S. Baffioni, C. Charlot, F. Ferri, D. Futyan, P. Meridiani, I. Puljak, C. Rovelli, R. Salerno, and Y. Sirois. Electron Reconstruction in CMS. *Eur. Phys. J.*, C49:1099–1116, 2007.
- [61] CMS Collaboration. Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV. *JINST*, 10(06):P06005, 2015.
- [62] CMS Collaboration. Performance of CMS Muon Reconstruction in pp Collision Events at $\sqrt{s} = 7$ TeV. *JINST*, 7:P10002, 2012.
- [63] CMS Collaboration. Performance of Reconstruction and Identification of τ Leptons Decaying to Hadrons and ν_τ in pp Collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(10):P10005, 2018.
- [64] M. Cacciari, G. P. Salam, and G. Soyez. The Anti- k_t Jet Clustering Algorithm. *JHEP*, 04:063, 2008.

-
- [65] CMS Collaboration. Identification of b-Quark Jets with the CMS Experiment. *JINST*, 8:P04013, 2013.
- [66] CMS Collaboration. Identification of Heavy-Flavour Jets with the CMS Detector in pp Collisions at 13 TeV. *JINST*, 13(05):P05011, 2018.
- [67] S. Alioli, P. Nason, C. Oleari, and E. Re. NLO Higgs Boson Production via Gluon Fusion Matched with Shower in POWHEG. *JHEP*, 04:002, 2009.
- [68] P. Nason and C. Oleari. NLO Higgs Boson Production via Vector-Boson Fusion Matched with Shower in POWHEG. *JHEP*, 02:037, 2010.
- [69] P. Nason. A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms. *JHEP*, 11:040, 2004.
- [70] S. Frixione, P. Nason, and C. Oleari. Matching NLO QCD Computations with Parton Shower Simulations: the POWHEG Method. *JHEP*, 11:070, 2007.
- [71] S. Alioli, P. Nason, C. Oleari, and E. Re. A General Framework for Implementing NLO Calculations in Shower Monte Carlo Programs: the POWHEG BOX. *JHEP*, 06:043, 2010.
- [72] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The Automated Computation of Tree-Level and Next-to-Leading Order Differential Cross Sections, and their Matching to Parton Shower Simulations. *JHEP*, 07:079, 2014.
- [73] R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H. S. Shao, and M. Zaro. The Automation of Next-to-Leading Order Electroweak Calculations. *JHEP*, 07:185, 2018.
- [74] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.
- [75] GEANT4 Collaboration. GEANT4: A Simulation Toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003.
- [76] CMS Collaboration. Measurement of the Inelastic Proton-Proton Cross Section at $\sqrt{s} = 13$ TeV. *JHEP*, 07:161, 2018.

- [77] S. Kallweit, J. M. Lindert, P. Maierhöfer, S. Pozzorini, and M. Schönherr. NLO Electroweak Automation and Precise Predictions for W+Multijet Production at the LHC. *JHEP*, 04:012, 2015.
- [78] S. Kallweit, J. M. Lindert, S. Pozzorini, M. Schönherr, and P. Maierhöfer. NLO QCD+EW Automation and Precise Predictions for V+Multijet Production. In *Proceedings, 50th Rencontres de Moriond, QCD and High Energy Interactions: La Thuile, Italy, March 21-28, 2015*, pages 121–124, 2015.
- [79] J. Baglio et al. Release Note - VBFNLO 2.7.0. FTUV-14-2903, IFIC-14-26, KA-TP-10-2014, LPN14-062, MAN-HEP-2014-03.
- [80] A. L. Read. Presentation of Search Results: the CL(s) Technique. *J. Phys.*, G28:2693–2704, 2002.
- [81] G. Cowan, K. Cranmer, E. Gross, and O. Vitells. Asymptotic Formulae for Likelihood-Based Tests of New Physics. *Eur. Phys. J.*, C71:1554, 2011. [Erratum: *Eur. Phys. J.*C73,2501(2013)].
- [82] A. Djouadi, O. Lebedev, Y. Mambrini, and J. Quevillon. Implications of LHC Searches for Higgs-Portal Dark Matter. *Phys. Lett.*, B709:65–69, 2012.
- [83] CMS Collaboration. Search for Invisible Decays of Higgs Bosons in the Vector Boson Fusion and Associated ZH Production Modes. *Eur. Phys. J.*, C74:2980, 2014.
- [84] NNPDF Collaboration. Parton Distributions for the LHC Run II. *JHEP*, 04:040, 2015.
- [85] CMS Collaboration. Search for the Standard Model Higgs Boson Produced in Association with a W or a Z Boson and Decaying to Bottom Quarks. *Phys. Rev. D*, 89:012003, 2014.
- [86] CMS Collaboration. Search for New Physics in Final States with an Energetic Jet or a Hadronically Decaying W or Z Boson and Transverse Momentum Imbalance at $\sqrt{s} = 13$ TeV. *Phys. Rev. D*, 97:092005, 2018.
- [87] CMS Collaboration. Search for New Physics in Events with a Leptonically Decaying Z Boson and a Large Transverse Momentum Imbalance in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J.*, C78(4):291, 2018.
- [88] M. Hoferichter, P. Klos, J. Menéndez, and A. Schwenk. Improved Limits for Higgs-Portal Dark Matter from LHC Searches. *Phys. Rev. Lett.*, 119:181803, 2017.

-
- [89] XENON Collaboration. Dark Matter Search Results from a One Ton-Year Exposure of XENON1T. *Phys. Rev. Lett.*, 121:111302, 2018.
 - [90] PandaX-II Collaboration. Dark Matter Results from First 98.7 Days of Data from the PandaX-II Experiment. *Phys. Rev. Lett.*, 117:121303, 2016.
 - [91] LUX Collaboration. Results from a Search for Dark Matter in the Complete LUX Exposure. *Phys. Rev. Lett.*, 118:021303, 2017.
 - [92] CDEX Collaboration. Limits on Light Weakly Interacting Massive Particles from the First 102.8 kg $\times$ day Data of the CDEX-10 Experiment. *Phys. Rev. Lett.*, 120:241301, 2018.
 - [93] SuperCDMS Collaboration. New Results from the Search for Low-Mass Weakly Interacting Massive Particles with the CDMS Low Ionization Threshold Experiment. *Phys. Rev. Lett.*, 116:071301, 2016.
 - [94] CRESST Collaboration. Results on Light Dark Matter Particles with a Low-Threshold CRESST-II Detector. *Eur. Phys. J. C*, 76:25, 2016.
 - [95] D. Kranzlmüller. The Future European Grid Infrastructure – Roadmap and Challenges. In *Proceedings of the ITI 2009 31st International Conference on Information Technology Interfaces*, pages 17–20, 2009.
 - [96] R. Pordes et al. The Open Science Grid. *J. Phys. Conf. Ser.*, 78:012057, 2007.
 - [97] B. Holzman et al. HEPCloud, a New Paradigm for HEP Facilities: CMS Amazon Web Services Investigation. *Comput. Softw. Big Sci.*, 1(1):1, 2017.
 - [98] A. Dorigo, P. Elmer, F. Furano, and A. Hanushevsky. XROOTD - A Highly Scalable Architecture for Data Access. *WSEAS Transactions on Computers*, 4:348–353, 04 2005.
 - [99] D. Thain, T. Tannenbaum, and M. Livny. Distributed Computing in Practice: the Condor Experience. *Concurrency - Practice and Experience*, 17:323–356, 02 2005.
 - [100] I. Sfiligoi. GlideinWMS - a Generic Pilot-Based Workload Management System. *J. Phys. Conf. Ser.*, 119:062044, 2008.
 - [101] R. Alfieri et al. Managing Dynamic User Communities in a Grid of Autonomous Resources. *CoRR*, cs.DC/0306004, 2003.

- [102] I. Foster, C. Kesselman, and S. Tuecke. The Anatomy of the Grid: Enabling Scalable Virtual Organizations. *The International Journal of High Performance Computing Applications*, 15(3):200–222, 2001.
- [103] A. Lenstra, X. Wang, and B. de Weger. Colliding X.509 Certificates. Cryptology ePrint Archive, Report 2005/067, 2005. <https://eprint.iacr.org/2005/067>.
- [104] D. Bonacorsi and T. Ferrari. WLCG Service Challenges and Tiered Architecture in the LHC Era. In *IFAE 2006*, pages 365–368, Milano, 2007. Springer Milan.
- [105] CMS Collaboration. The CMS Computing Model : the ‘CMS Computing Model RTAG’. CERN, Geneva, 2004. CERN-LHCC-2004-035, CERN-CMS-NOTE-2004-031, CMS-NOTE-2004-031, LHCC-G-083.
- [106] CMS Collaboration. The CMS Computing Project: Technical Design Report. CERN, Geneva, 2005. CERN-LHCC-2005-023, CMS-TDR-7.
- [107] D. Bonacorsi. The CMS Computing Model. *Nuclear Physics B - Proceedings Supplements*, 172:53 – 56, 2007. Proceedings of the 10th Topical Seminar on Innovative Particle and Radiation Detectors.
- [108] M. Giffels, Y. Guo, V. Kuznetsov, N. Magini, and T. Wildish. The CMS Data Management System. *J. Phys. Conf. Ser.*, 513:042052, 2014. Proceedings, 20th International Conference on Computing in High Energy and Nuclear Physics (CHEP 2013): Amsterdam, The Netherlands, October 14-18, 2013.
- [109] M. Cinquilli et al. The CMS Workload Management System. *Journal of Physics: Conference Series*, 396(3):032113, 2012.
- [110] O. Buchmüller et al. The CMS CERN Analysis Facility (CAF). *Journal of Physics: Conference Series*, 219(5):052022, 2010.
- [111] J. Balcas et al. Using the GlideinWMS System as a Common Resource Provisioning Layer in CMS. *J. Phys. Conf. Ser.*, 664(6):062031, 2015.
- [112] D. Spiga et al. The CMS Remote Analysis Builder (CRAB). *Lect. Notes Comput. Sci.*, 4873:580–586, 2007.
- [113] G. Codispoti et al. CRAB: A CMS Application for Distributed Analysis. *IEEE Transactions on Nuclear Science*, 56(5):2850–2858, 2009.
- [114] A. Fanfani et al. Distributed Analysis in CMS. *J. Grid Comput.*, 8:159–179, 2010.

-
- [115] G. Petrucciani et al. Mini-AOD: A New Analysis Data Format for CMS. *J. Phys. Conf. Ser.*, 664(7):072052, 2015.
- [116] D. Hufnagel. CMS Use of Allocation Based HPC Resources. *J. Phys. Conf. Ser.*, 898(9):092050, 2017.
- [117] D. Hufnagel. HPC Resource Integration into CMS Computing via HEPCloud. Technical Report CMS-CR-2018-283, CERN, Geneva, 2018.
- [118] CMS Collaboration. Experience with Dynamic Resource Provisioning of the CMS Online Cluster Using a Cloud Overlay. Technical Report CMS-CR-2018-398, CERN, Geneva, 2018.
- [119] G. Codispoti, R. Di Maria, et al. Elastic Extension of a CMS Computing Centre Resources on External Clouds. *J. Phys. Conf. Ser.*, 762(1):012013, 2016.
- [120] V. Ciaschini, G. Codispoti, L. Rinaldi, D. C. Aiftimiei, D. Bonacorsi, P. Calligola, S. Dal Pra, D. De Girolamo, R. Di Maria, C. Grandi, D. Michelotto, M. Panella, S. Taneja, and F. Semeria. Elastic Extension of a Local Analysis Facility on External Clouds for the LHC Experiments. *J. Phys. Conf. Ser.*, 898(5):052024, 2017.
- [121] DODAS Project. <https://dodas-ts.github.io/dodas-doc/>.
- [122] EOSC-hub Project. <https://eosc-hub.eu/>.
- [123] INDIGO-DataCloud Collaboration. INDIGO-DataCloud: a Data and Computing Platform to Facilitate Seamless Access to E-Infrastructures. *Journal of Grid Computing*, 16(3):381–408, 2018.
- [124] Apache Software Foundation: Apache MESOS. <http://mesos.apache.org/>.
- [125] OpenStack Foundation: OpenStack. <https://www.openstack.org>.
- [126] S. Ting. The Alpha Magnetic Spectrometer on the International Space Station. *Nucl. Phys. Proc. Suppl.*, 243-244:12–24, 2013.
- [127] M. Zaharia et al. Apache Spark: a Unified Engine for Big Data Processing. *Communications of the ACM*, 59:56–65, 11 2016.
- [128] OASIS Standard: TOSCA Simple Profile in YAML Version 1.1. <http://docs.oasis-open.org/tosca/TOSCA-Simple-Profile-YAML/v1.1/os/TOSCA-Simple-Profile-YAML-v1.1-os.html>.

- [129] M. Caballer et al. Orchestrating Complex Application Architectures in Heterogeneous Clouds. *Journal of Grid Computing*, 16(1):3–18, 2018.
- [130] Marathon. <https://mesosphere.github.io/marathon/>.
- [131] D. Merkel. Docker: Lightweight Linux Containers for Consistent Development and Deployment. *Linux Journal*, 2014, 03 2014.
- [132] D. Dykstra and L. Lueking. Greatly Improved Cache Update Times for Conditions Data with Frontier/Squid. *Journal of Physics: Conference Series*, 219(7):072034, 2010.
- [133] P. Buncic et al. CernVM - a Virtual Software Appliance for LHC Applications. *Journal of Physics: Conference Series*, 219(4):042003, 2010.
- [134] K. Bloom et al. Any Data, Any Time, Anywhere: Global Data Access for Science. In *Proceedings of the 2nd IEEE/ACM International Symposium on Big Data Computing (BDC)*, 2015.
- [135] J. Alwall et al. A Standard Format for Les Houches Event Files. *Comput. Phys. Commun.*, 176:300–304, 2007.
- [136] E. Karavakis et al. CMS Dashboard Task Monitoring: a User-Centric Monitoring View. *J. Phys. Conf. Ser.*, 219:072038, 2010.
- [137] R. Di Maria. Higgs-to-Invisible Searches with the CMS Experiment at the LHC. In *Proceedings of the Sixth Annual Conference on Large Hadron Collider Physics*, page 057, 10 2018.
- [138] CMS Collaboration. Search for Invisible Decays of a Higgs Boson Produced through Vector Boson Fusion at the High-Luminosity LHC. CERN, Geneva, 2018. CMS-PAS-FTR-18-016.

List of Acronyms

AAA Anydata, Anytime, Anywhere

AAI authentication and authorisation infrastructure

ALICE A Large Ion Collider Experiment

AOD analysis object data

ATLAS A Toroidal LHC ApparatuS

AWS Amazon Web Services

BR branching ratio

BSM beyond the standard model

CE computing element

CL confidence level

CMS Compact Muon Solenoid

CPU central processing unit

CR control region

CRAB CMS Remote Analysis Builder

CSVv2 combined secondary vertex version-2

CTF combinatorial track finder

DM dark matter

DN distinguished name

DODAS dynamic on demand analysis service

ECAL electromagnetic calorimeter

EGI European Grid Infrastructure
EM electromagnetic
EOCS European Open Science Cloud
EW electroweak
ggH gluon fusion
HB hadronic barrel calorimeters
HC Hammer Cloud
HCAL hadron calorimeter
HE hadronic end-cap calorimeters
HF hadronic forward calorimeters
HLT High Level Trigger
HO hadronic outer barrel calorimeters
HPC high performance computing
HPS hadron plus strips
I/O input/output
IaaS infrastructure-as-a-service
IAM identity and access management
ICL Imperial College London
ID identification
IM infrastructure manager
INFN National Institute for Nuclear Physics
IP internet protocol
ISO isolation
ISR initial state radiation
JER jet energy resolution
JES jet energy scale

L1	Level-1
LEP	Large Electron-Positron Collider
LHC	Large Hadron Collider
LHCb	LHC-beauty
LO	leading order
LS	long shutdown
MC	Monte Carlo
MET	missing transverse energy
ML	machine learning
NLO	next-to-leading order
OSG	Open Science Grid
PaaS	platform-as-a-service
PDF	parton distribution function
PF	particle-flow
POGs	physics object groups
pp	proton-proton
PV	primary vertex
QCD	quantum chromodynamics
QED	quantum electrodynamics
QFT	quantum field theory
RAM	random access memory
SaaS	software-as-a-service
SE	storage element
SF	scale factor
SI	submission infrastructure
SM	standard model of particle physics

SR signal region

SRM storage resource manager

SSB spontaneous symmetry breaking

SUSY supersymmetry

T0 Tier-0

T1 Tier-1

T2 Tier-2

T3 Tier-3

TriDAS trigger and data-acquisition system

ttH top quark associated production

TTS token translation service

VBF vector boson fusion

VH vector boson associated production

VLAN virtual local area network

VM virtual machine

VO virtual organisation

VOMS virtual organisation management system

VV diboson pair

WLCG Worldwide LHC Computing Grid

WMS workload management system

WN worker node