Imperial College of Science, Technology and Medicine
Department of Physics

Search for invisible decays of the Higgs boson at $\sqrt{s} = 13$ TeV

Vukašin Milošević

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Physics of the Imperial College London, August 2020

Abstract

A search for invisibly decaying Higgs bosons using proton-proton collision data collected by the CMS experiment at $\sqrt{s} = 13$ TeV is presented. For the first time, this search is performed using the complete dataset collected during the Run 2 phase of operation, corresponding to total integrated luminosity of 137.1 fb^{-1}.

In 2017 and 2018, to supplement the traditional trigger strategy based on the presence of missing transverse energy, a new dedicated trigger based on the Vector Boson Fusion jet topology was introduced. Its motivation, implementation and subsequent usage is described in detail. This addition further enhanced the sensitivity to the Vector Boson Fusion production mechanism by allowing for a new analysis category focused around previously inaccessible phase space.

A combination with the study focusing on the 2016 data is performed. No deviations from the Standard Model have been observed and the final result is reported in the form of a 95 % Confidence Level upper limit on the Br(H→inv). The overall result of the measurement using the 2017 and 2018 data yields an observed (expected) Br(H→inv) $<$ 15 % (13 %), while the full Run 2 limit stands at Br(H→inv) $<$ 15 % (12 %). When interpreted in terms of dark matter searches, this result can be used to provide 90% Confidence Level upper limits on the spin-independent dark matter-nucleon scattering cross section. Lastly, a study of projected sensitivity of this study for the conditions provided by the future upgrades of the CMS experiment is also presented. The projected limit for the total integrated luminosity of 3000 fb^{-1} is placed, from the perspective of the CMS experiment, at Br(H→inv) $<$ 3.8 %, while a combined result from the ATLAS and CMS experiments is expected to yield Br(H→inv) $<$ 2.5 %.

Copyright Declaration

Acknowledgements

Na početku bih želeo da se zahvalim svojim roditeljima. Ova teza je vaš uspeh koliko i moj. Vaša ljubav i stalna podrška su mi pomogli da prebrodim čak i najteže trenutke. Neizmerno vam hvala!

I would like to express my deepest gratitude to my supervisor, Dr. Alexander Tapper, for guiding me through every step of this process. I will forever be thankful for the discussions and advice I received from you.

I also want to thank my good friend, Dr. Olivier Davignon, for all that you have done for me (the list would be too long to write here). I'm looking forward to working with you in the future. Also, I would like to thank the H→inv analysis team. Nick, Anne-Marie and Sam, thank you for all the useful chats.

Finally, I would like to thank the Imperial College HEP group and the President's Scholarship (former Imperial College Scholarship) for giving me the opportunity to be a part of this institution and to spend two years at CERN. This is where, besides working hard on my analysis, I made friendships which have been forged in the fires of PhD studies. Dwayne, Esh, Felix, Malte, Antoni and Julia, thank you for being there to lighten up the dark times and to enjoy the good ones too.

Dedication

To my loving parents.

“I do not fear invisible worlds.”
Ivo Andrić

Contents

List of Tables

List of Figures

Overview and declaration

The following document serves as the summary of my work during the past four years. All other statements and results are appropriately referenced. Based on the conventional thesis structure, Chapter 1 presents a description of the main pieces forming the most complete particle physics theory, the Standard Model. This is followed by a description of the idea for connecting the invisible final state of the Higgs Boson with the prospects for Dark Matter searches, formulated into Chapter 2. The two aforementioned chapters form Part I of this thesis, and serve as a theoretical/motivation basis, explained in my own words, allowing the reader to have a brief theory overview before moving forward with the details of the experimental approach.

Part II of this thesis serves as an introduction to the world of collider physics with the emphasis on the structure of the CMS experiment (Chapter 3) as well as a description of its data acquisition system (Chapter 4). Chapter 4 contains an overview of the Data Quality Monitoring system for the Level-1 trigger within the CMS experiment with an example being given in the form the Level-1 Trigger Calorimeter Layer 2. This section includes my work during the 2017-2018 period of data taking, when I was in charge of developing and maintaining that system. Chapter 4 also contains a detailed description of the implementation of High Level Trigger paths designed specifically for the purposes of the main study covered by this thesis. This part serves as a summary of my work on this topic, as I was responsible for the design, development, implementation and testing of those paths.

Part III presents the main results given in this thesis, the search for invisible decays of the Higgs boson at $\sqrt{s} = 13$ TeV. The focus of this thesis is the scenario in which the Higgs Boson is produced through the process of Vector Boson Fusion. Chapter 6 serves as an overview of algorithms deployed for the purpose of particle reconstruction followed with a discussion of respective object corrections (some of which were developed for this study specifically by the Imperial College analysis team). Chapters 7 and 8 introduce the main strategy behind this study. Having been its lead analyser (and publication contact), these chapters showcase my work, performed as a member of the Imperial College analysis team. The overview of the treatment of uncertainties and the signal extraction strategy begins the discussion presented in Chapter 9, which culminates with a summary of final results that came out of this search. This summary focuses on the benefits gained through the use of the full dataset collected by the CMS detector during the 2017-2018 period and a novel approach to the analysis strategy with the usage of new triggers. Combination with the previously published results based on the data collected during 2016 is also presented yielding a preliminary legacy result for this study.

Lastly, Chapter 10 serves as a conclusion of the journey summarised in this thesis. It is used to present final statements on the main analysis presented in previous chapters as well as to indicate what the future may hold. The prospects for the near future include the strategy for the combination of different searches for the invisible final state of the Higgs boson, where the previously described results focusing on the Vector Boson Fusion production mode are going to be combined with results from studies targeting other hadronic production modes. This section will summarise my work as a member of a UK wide collaboration of analysis teams forming the ”Combined Higgs to Invisible Project” (CHIP) working group. Before concluding, a simulation study of future prospects for these analyses is presented, covering part of my work done in order to test the sensitivity of these processes and the behaviour of the upgraded CMS detector under expected future operation conditions of the Large Hadron Collider.

* * *

Part I: The idea

"Ekser drži potkov, potkov konja, konj junaka, junak grad, a grad zemlju."

– Serbian proverb –

Chapter 1

The Standard Model of particle physics

> *"New ideas pass through three periods: 1) It can't be done. 2) It probably can be done, but it's not worth doing. 3) I knew it was a good idea all along!"*
>
> — Arthur C. Clarke

1.1 Introduction

THE CURRENT STATE OF PARTICLE PHYSICS allows us to unify three out of four interactions through which our universe exists. The most complete theory that summarises this knowledge into a mathematical formalism is the Standard Model (SM). The main focus of this chapter will be to describe the main constituents of matter and how they interact with each other.

In order to start the discussion, a short introduction on the importance of gauge theories is given before proceeding with describing details regarding each of the interactions relevant to the current state of the SM. Due to its importance to the main search presented with this thesis, special attention will be given to the Higgs mechanism and, slightly later, to the production modes of the Higgs boson itself. This chapter serves as a theoretical prelude enabling further discussion regarding the approach behind the main study, which is the focus of the next chapter.

1.2 Building blocks of the Universe

It is truly amazing to see that our entire physical realm can be explained through a finite set of particles. Fundamental (or elementary) particles, summarised in Tables 1.1 and 1.2, can be grouped into three generations of fermions and a set of gauge bosons. Each fermion generation consists of a set of two leptons, and two quarks. The first, being comprised of the up (u) and down (d) quarks, electron and (e^-) and electron neutrino (ν_e), is responsible for all visible matter in our universe. Moving away from the stable setting towards higher energies, achieved by collider experiments, helps to complete the picture by introducing the rest. The second generation is formed by the strange (s) and charm (c) quarks accompanied by a muon and muon neutrino lepton pair, while the third is made out of top (t) and bottom (b) quarks followed with a tau lepton and tau neutrino pair.

	Leptons				Quarks			
	Particle	(sym.)	Q	mass/GeV	Particle	(sym.)	Q	mass/GeV
First	electron	(e^-)	-1	0.0005	down	(d)	$-\frac{1}{3}$	0.003
Generation	neutrino	(ν_e)	0	$< 10^{-9}$	up	(u)	$+\frac{2}{3}$	0.005
Second	muon	(μ^-)	-1	0.106	strange	(s)	$-\frac{1}{3}$	0.1
Generation	neutrino	(ν_μ)	0	$< 10^{-9}$	charm	(c)	$+\frac{2}{3}$	1.3
Third	tau	(τ^-)	-1	1.78	bottom	(b)	$-\frac{1}{3}$	4.5
Generation	neutrino	(ν_τ)	0	$< 10^{-9}$	top	(t)	$+\frac{2}{3}$	174

Table 1.1: Summary of fundamental fermions grouped by their respective generations (with Q denoting the electromagnetic charge expressed in units of the elementary charge) [1].

The exchange of gauge bosons associated with each interaction (listed alongside their properties in Table 1.2) is the mechanism through which fermions communicate with each other. The most well known propagator, the photon (γ) serves as the voice of the electromagnetic interaction, with the $W^\pm/Z^0$ bosons and gluons (g) performing the same role for the weak and strong interaction respectively.

The "crown jewel" of particle physics, the Higgs boson, represents the final piece that completes the set of fundamental particles. Its mass currently stands at 125.18 ± 0.16 GeV [2] following the combined measurements performed by ATLAS and CMS collaborations.

The importance of the Higgs field and the mechanism through which it interacts with fundamental particles will be the main point of discussion in the following sections. Further topics regarding its production modes and subsequent decays are

Interaction	Boson		Spin	Mass/GeV
Weak	W/Z bosons	($W^{\pm}$/Z)	1	80.4/91.2
Electromagnetic	Photon	(γ)	1	0
Strong	Gluon	(g)	1	0

Table 1.2: Summary of gauge bosons grouped by their associated interactions [1].

introduced in Chapter 2, where the invisible final state will be explained. Currently no deviations from the SM have been observed, but, as it will be seen in the following chapters, uncertainties in these measurements leave the door open wide enough that it still motivates searches for physics beyond the SM.

1.3 Introduction to gauge theories

The quantum mechanical Lagrangian is widely used as an essential tool in High Energy Physics (HEP) when creating a blueprint of how particles interact with each other. Further focusing on its properties, gauge theories collect the information regarding local symmetries of the given Lagrangian allowing for the introduction of symmetry groups for the Lagrangian (theory) at hand. The associated transformations of gauges lead to the formation of symmetry groups [3, 4]. This as a consequence brings the fact that every group generator yields a corresponding gauge field already included in the Lagrangian due to the condition of invariance.

From the perspective of gauge theories, the formulation of the SM is seen as a non-Abelian theory associated with: $U(1)_Y \times SU(2)_L \times SU(3)_C$[1]. Breaking it down to core members, the $U(1)$ symmetry group is associated with the theory of Quantum Electrodynamics (QED) describing the electromagnetic interaction. The next item, the $SU(2)$ group, represents the symmetry group of the weak interaction. Finally, the Quantum Chromodynamics (QCD) theory, which focuses on the strong interaction, is connected with the $SU(3)$ symmetry group [5, 6]. The following sections introduce each of these theories in more detail.

Keeping with the introduction theme of this section, widely used mathematical apparatus should also be mentioned. Similarly to classical theories, the Euler-Lagrange equations can be used to obtain the equations of motion for a given theory [7]. They

---

[1] This chapter uses the notation and conventions introduced in Ref. [1]

are given as:

$$\frac{d}{dt}\left(\frac{\partial\mathcal{L}}{\partial\dot{q}_i}\right)-\frac{\partial\mathcal{L}}{\partial q_i}=0, \tag{1.1}$$

where q_i and $\dot{q}_i$ represent generalised coordinates and corresponding velocities for the given theory. The symbol $\mathcal{L}$, denoting the Lagrangian density, is going to be referred to in further text simply as the Lagrangian.

Another helping hand comes in the form of Noether's theorem [8]. The main statement originating from it, regarding the Lagrangian at hand, is the existence of a conserved current connected to the respective symmetry.

1.4 Quantum electrodynamics

Starting from the quantum mechanical generalisation of Einstein's energy-momentum relation, an expression denoting the Klein-Gordon equation can be written. Expressed in its Lorentz-invariant formulation, it takes the form of:

$$(\partial_\mu\partial^\mu+m^2)\psi=0, \tag{1.2}$$

where the ∂_μ denotes the partial derivative (with the summation of repeating indices being imposed as a convention) [1, 4]. The plain wave solutions of the aforementioned equation prove problematic when it comes to the interpretation of its negative energy solutions, which also arise from the original relation. The problematic nature is manifested in the fact that this scenario yields a probability density that has the possibility of being negative in value.

An alternative approach was taken by Dirac, leading to the formulation of the Dirac equation. Its covariant form, as well its Lagrangian, can be written as:

$$i\gamma^\mu\partial_\mu\psi-m\psi=0, \tag{1.3}$$

$$\mathcal{L}=i\overline{\psi}\gamma^\mu\partial_\mu\psi-m\overline{\psi}\psi, \tag{1.4}$$

The generalisation of the aforementioned Lagrangian to the QED leads to the following definition:

$$\mathcal{L}_{QED}=i\overline{\psi}\gamma^\mu\partial_\mu\psi-m\overline{\psi}\psi-\frac{1}{4}F^{\mu\nu}F_{\mu\nu}+e\overline{\psi}\gamma^\mu\psi A_\mu, \tag{1.5}$$

where A^μ represents the electromagnetic four-potential and $F^{\mu\nu}=\partial^\mu A^\nu-\partial^\nu A^\mu$ denotes the electromagnetic field strength tensor. The addition of two new terms (representing the kinetic term for A_μ and the interaction term respectively) allows for the fulfillment of Maxwell's equation, which can be obtained through the use of the Euler-Lagrange equations with respect to A^μ [9, 10].

This Lagrangian is invariant under the $U(1)$ local symmetry, or in other words, it remains unchanged with respect to the transformations: $\psi^{'}(x) = e^{-ie\lambda(x)}\psi(x)$ and $A^{'}_\mu = A_\mu + \partial_\mu \lambda$ (where the later corresponds to the choice of the Lorentz gauge). The redefinition of the partial derivative: $D_\mu = \partial_\mu - ieA^\mu$ leads to the following expression for the QED Lagrangian:

$$\mathcal{L}_{QED} = i\overline{\psi}\gamma^\mu D_\mu \psi - m\overline{\psi}\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}, \tag{1.6}$$

The previously written Lagrangian captures the basis of QED, whose formulation was awarded a Nobel prize [11].

1.5 Quantum chromodynamics

Similarly to the development of QED and its connection to the $U(1)$ symmetry group, the origin of QCD is strongly connected to the initial introduction of isospin and the formation of an isospin doublet comprised of a proton and neutron. Further development led to a better understanding of the non-elementary structure of protons and neutrons, paving way for the currently used quark model. QCD is associated with the $SU(3)$ symmetry group, whose generators give rise to eight gluon fields [1, 4, 12].

The local local phase transformations of the $SU(3)$ group can be written as: $\psi^{'} = e^{ig_s\lambda_a(x)T_a}\psi$, where the generators of the SU(3) group are defined as $T_a = \frac{1}{2}t_a$ (with the t_a being the Gell-Mann matrices). The QCD Lagrangian can now be formulated in the following way:

$$\mathcal{L}_{QCD} = i\overline{\psi}\gamma^\mu D_\mu \psi - m\overline{\psi}\psi - \frac{1}{4}G^{\mu\nu}_a G^a_{\mu\nu}, \tag{1.7}$$

where the covariant derivative takes the form of: $D_\mu = \partial_\mu + igT_aG^a_\mu$, with G^a_μ representing the gluon fields and $G^{\mu\nu}_a$ denoting field strength tensors. In the previous scenario the index a indicates the existence of eight fields being connected to the $SU(3)$ generators. Finally, these generators do not possess the commutation property of QED, making QCD a non-Abelian theory [13].

1.6 Electroweak interaction

In contrast to the previously discussed theories, a model describing the weak interaction is required to take an approach which accommodates the observed violation of parity. The aforementioned requirement led to the formulation of the V-A (vector and axial vector coupling) structure [6].

The initial theory of weak interactions was associated with the $SU(2)_L$ symmetry group and the weak isospin (I_W), where the L denotes the behavior of charged-current in which it can only be implicated with left handed particles [1, 4]. The transformations of the weak isospin doublet $\psi' = e^{ig_W \lambda_a(x) T^a}\psi$[2] (and the accompanying redefinition of the partial derivative as $D_\mu = \partial_\mu + ig_W T_a W^a_\mu$) leads to the appearance of three gauge fields W^a_μ, from which the $W^\pm$ fields can be formed as:

$$W^\pm = \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu). \tag{1.8}$$

The journey towards electroweak unification begins with the properties of QED. A massive undertaking by Glashow, Salam and Weinberg (GSW) [14–16] was formulated into the GSW model whose starting point begins with the introduction of the weak hypercharge (Y), through the use of the following relation:

$$Q = I^3_W + \frac{1}{2}Y, \tag{1.9}$$

where Q represents the electromagnetic charge and I^3_W denotes the third component of I_W. The, now $U(1)_Y$, symmetry group of QED brings the transformation $\psi' = e^{ig'\frac{Y}{2}\lambda(x)}\psi$ and a field $B_\mu(x)$ associated with it. The interpretation arising from the GSW model is that the Z boson and photon associated fields are linear combinations of the, previously introduced, W^3_μ and B_μ:

$$Z_\mu = -B_\mu \sin\theta_W + W^3_\mu \cos\theta_W, \tag{1.10}$$

$$A_\mu = B_\mu \cos\theta_W + W^3_\mu \sin\theta_W. \tag{1.11}$$

The summary of the GSW model is the conclusion that the unified electroweak interaction is represented by the $SU(2)_L \times U(1)_Y$ symmetry group. The corresponding Lagrangian can be written as:

$$\mathcal{L}_{EW} = i\bar{\psi}\gamma^\mu D_\mu\psi - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4}W^{\mu\nu}_a W^a_{\mu\nu}, \tag{1.12}$$

---

[2] With T_a being the generators of the $SU(2)$ group.

where $D_\mu = \partial_\mu + ig' B_\mu \frac{1}{2} + ig_W W^a_\mu T_a$ and $B_{\mu\nu}$ and $W^a_{\mu\nu}$ represent field tensors for B_μ and W^a_μ respectively.

1.7 Brout-Englert-Higgs mechanism

Following the discovery of the Higgs boson at the experiments within the Large Hadron Collider (LHC), the revolutionary nature of the Brout-Englert-Higgs (BEH) mechanism [17, 18] was recognised and awarded the Physics Nobel prize in 2013 [19]. Its base idea follows the spontaneous symmetry breaking approach which can be introduced by taking a look at a simple scenario of a Lagrangian of a scalar field defined as:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - \frac{1}{2}\mu^2\phi^2 - \frac{1}{4}\lambda\phi^4. \tag{1.13}$$

Omitting the kinetic term of the aforementioned Lagrangian and focusing only on the potential, a conclusion arises that there can be two distinct scenarios depending on the values of the μ^2 parameter[3] [1, 4]. Figure 1.1 shows the shapes of the given potential (denoted as $V(\phi)$) for both the $\mu^2 < 0$ (marked in blue) and $\mu^2 > 0$ (marked in red) scenarios.

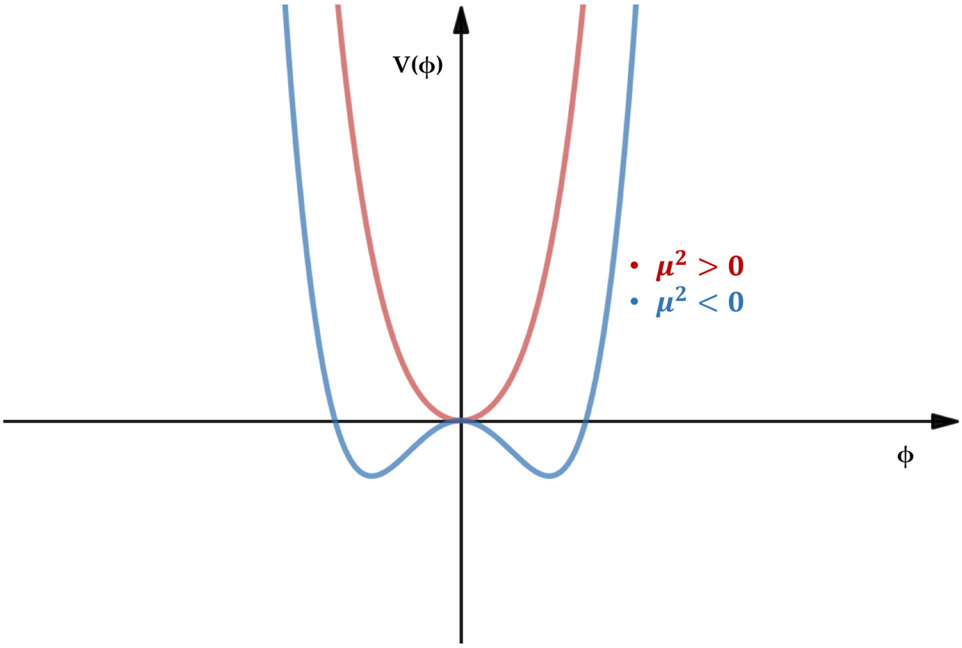


Figure 1.1: Graphical representation of the potential $V(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4$ for both the $\mu^2 < 0$ (blue) and $\mu^2 > 0$ (red) scenarios (made using Ref. [20]).

---

[3] The λ parameter is already constrained by having to be positive in order for the potential to have a global minimum.

Taking the look at the first derivative in order to find the local extrema:

$$\frac{\partial V}{\partial \phi} = \mu^2\phi + \lambda\phi^3 = 0, \tag{1.14}$$

leads to the scenarios of $\phi_a = 0$ (for the $\frac{\partial^2 V}{\partial \phi^2} = \mu^2 > 0$) and $\phi_b = \pm|\sqrt{-\frac{\mu^2}{\lambda}}|$ (for the $\frac{\partial^2 V}{\partial \phi^2} = -2\mu^2 > 0$ or the $\mu^2 < 0$ case). The symmetry breaking of the Lagrangian happens when a choice between the two values of ϕ_b are made (in further text renamed as $v = |\phi_b|$). Expressing ϕ in terms of its vacuum expectation state, one can introduce $H(x)$ and rewrite the Lagrangian as:

$$\phi(x) = v + H(x), \tag{1.15}$$

$$\mathcal{L} = \frac{1}{2}(\partial_\mu H)(\partial^\mu H) - \frac{1}{2}\mu^2(v+H)^2 - \frac{1}{4}\lambda(v+H)^4. \tag{1.16}$$

Grouping of terms quadratic in $H(x)$ leads to its mass term within the Lagrangian, or in other words: $m_H = \sqrt{-2\mu^2}$.

A natural extension leads to the inclusion of a complex version of the aforementioned potential with $\phi = \frac{1}{2}(\phi_1 + i\phi_2)$ and $V(\phi) = \mu^2\phi^*\phi + \lambda(\phi^*\phi)^2$. Repeating the search for global minimum leads to the possible choice of a vacuum state of $\phi_{vac} = (v, 0)$ and the re-composition of ϕ as: $\phi = \frac{1}{2}(v + H(x) + i\chi(x))$. As a second step in the process, a good way to evolve the simplified Lagrangian is to take a look at the U(1) gauge symmetry. This can be done by using the kinetic term with the appropriate re-definitions of the partial derivative (as done in Section 1.6 for B_μ) as well as using the choice of the unitary gauge in order to remove the χ dependency[4]. This new Lagrangian, when rewritten in terms of $H(x)$, yields a massive scalar field, the Higgs field, associated with m_H alongside a massive gauge boson (associated with B_μ) and the appropriate interaction and self interaction terms.

Lastly, this approach is to be applied to the $SU(2)_L \times U(1)_Y$ symmetry group, as previously defined in Section 1.6 closely associated with the unified electroweak interaction. This requires a re-definition of ϕ through a weak isospin doublet and the usage of the Lagrangian defined with Equation 1.12. Following the procedure of searching for a global minimum, applying the unitary gauge and expanding the Lagrangian in terms of $H(x)$ gives a similar, yet slightly more complex picture. Grouping the mass terms for the appropriate fields ultimately yields:

$$m_W = \frac{1}{2}g_W v,\ m_Z = \frac{v}{2}\sqrt{g'^2 + g_W^2} \text{ and } m_A = 0^5. \tag{1.17}$$

---

[4] This can be achieved by choosing $\lambda(x) = -\frac{1}{g'v}\chi(x)$

[5] Where the latter two are obtained through the diagonalization of the mass matrix as explained

Before concluding this part of the story, there is one more topic that needs to be addressed and that is the mechanism through which fermions acquire their mass. In order to add an item corresponding to the mass term of fermions within the electroweak Lagrangian, it has to be invariant under $SU(2)_L \times U(1)_Y$ transformations. For the $I_W^3 = -\frac{1}{2}$ fermions this term can be written as:

$$\mathcal{L} = -g_f(\overline{L}\phi R + \overline{R}\phi^\dagger L), \tag{1.18}$$

where g_f denotes the Yukawa coupling and L (R) represents the SU(2) doublet (singlet). Through the process of spontaneous symmetry breaking, upon rewriting the Lagrangian in terms of $H(x)$, it can be seen that the fermion masses can be associated with the Yukawa coupling as:

$$m_f = \frac{vg_f}{\sqrt{2}} \tag{1.19}$$

After the introduction to the SM presented here, the following chapter will connect the advancements made in collider physics with the potential for beyond the SM physics through the idea behind the invisible decays of the Higgs boson.

---

in great detail in Ref. [1].

Chapter 2

When Higgs boson met dark matter

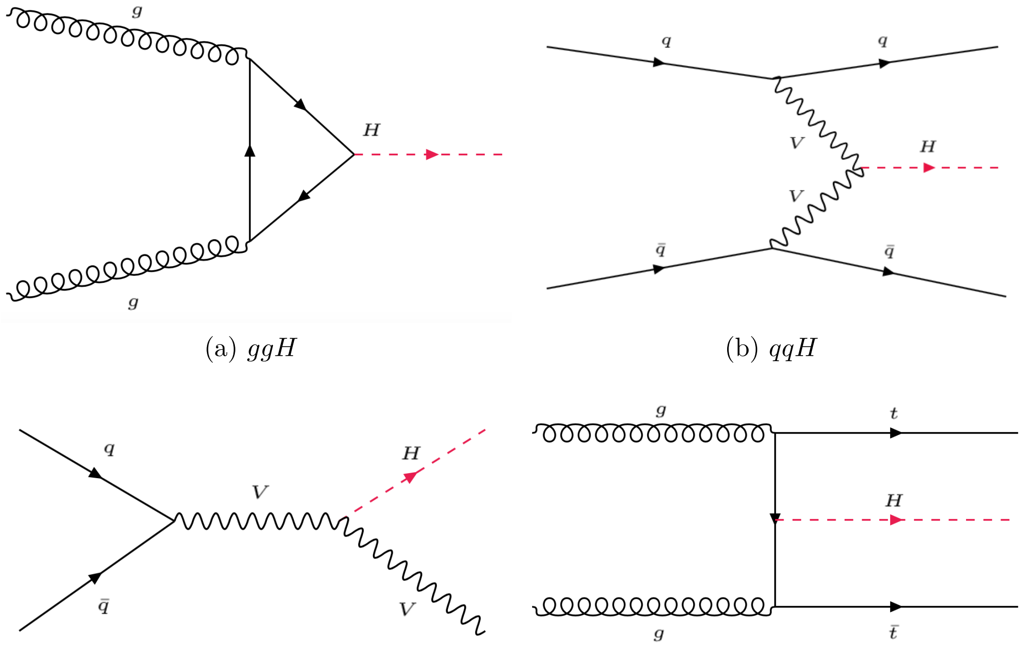
> *"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less."*
>
> — Marie Curie

2.1 Introduction

Being the main focus for studies covered by this thesis, special attention needs to be given to the Vector Boson Fusion production mode of the Higgs boson, while also discussing other topologies interesting for the invisible final state. Upon completing this discussion, a slight turn in focus is going to be taken in order to introduce the current understanding of the term dark matter (DM). It is presented alongside the investigation of a possible connection between DM searches and measurements of properties of the Higgs boson performed at hadron collider experiments. This will culminate with the illustration of the analysis idea for probing of the SM through the invisible final state of the Higgs boson decay. Finally, in order to complete the whole picture a brief overview of previously obtained results is presented.

2.2 Production modes of the Higgs boson

The main interest of the two general purpose experiments centered around the interaction points of the LHC [21] (consequently for this thesis as well) is the production and subsequent decays of the Higgs boson. Figure 2.1 shows diagrams for its various production mechanisms relevant for hadron colliders, in this case the proton-proton collisions (more details about the experimental setup are given in Chapter 3).



(a) ggH (b) qqH

(c) VH (d) ttH

Figure 2.1: Diagrams for the main production mechanisms of the Higgs boson: a) gluon gluon fusion, b) Vector Boson Fusion, c) Higgs-strahlung and d) associated production with top quarks (diagrams were made using Ref. [22]).

The production mode with the largest cross section is the gluon-gluon fusion (ggH) [23, 24], where jets[1] can arise due to initial state radiation (ISR)[2]. The production of a Higgs boson proceeds through, as the name suggests, a gluon fusion forming a virtual quark loop which ultimately yields the aforementioned boson. The production mode with the second largest cross section, but the highest importance for this thesis, is the Vector Boson Fusion (qqH or simply VBF) mechanism [23–26]. The collision enables an exchange of virtual vector bosons which fuse together to produce a Higgs boson. This leaves a Higgs+2 jets scenario, where the jets are located close to the beam line. There is no color exchange between these jets. From

---

[1] A collimated stream of hadrons.

[2] Leading to a reduction of the production cross section (subsequently its importance), but still being relevant to these studies.

the perspective of the plane perpendicular to the beam line, jets have a small separation due to the production of the Higgs boson. This distribution is mostly flat in shape as a direct reflection of the nature of Higgs boson's coupling to vector bosons [27]. Lastly, its main properties involve a large dijet invariant mass as a direct consequence of the large jet separation. Having defined these two production modes, it is good starting point for the introduction of the main focus of this thesis. The search for the invisible (from the experimental point of view) decays of the Higgs boson explores a possibility of physics beyond the SM being present in one of the decay modes of the Higgs boson. The main idea behind the study relies of the, previously introduced, distinct features of the VBF prodction mode, coupled with a strategy to quantify the invisible contribution from the perspective of the experiment at hand. As this strategy relies on the Higgs+2 jets scenarios, there will be a significant contribution originating from ggH processes, which also can produce a 2 jet final state, and populate the low dijet invariant mass ranges. For the analysis described in this thesis, the combination of both the VBF and ggH+2 jets scenarios is referred to as signal.

Continuing with other production modes, the Higgs strahlung or the associated VH production [28] happens when the colliding particles produce a virtual vector boson which can emit a Higgs boson. Lastly, a Higgs boson can be produced in association with top quarks [29, 30]. Figure 2.2 shows the production cross sections for these scenarios ranging over various proposed masses of the Higgs boson, serving as an illustrative example of the order of importance for the SM Higgs boson.

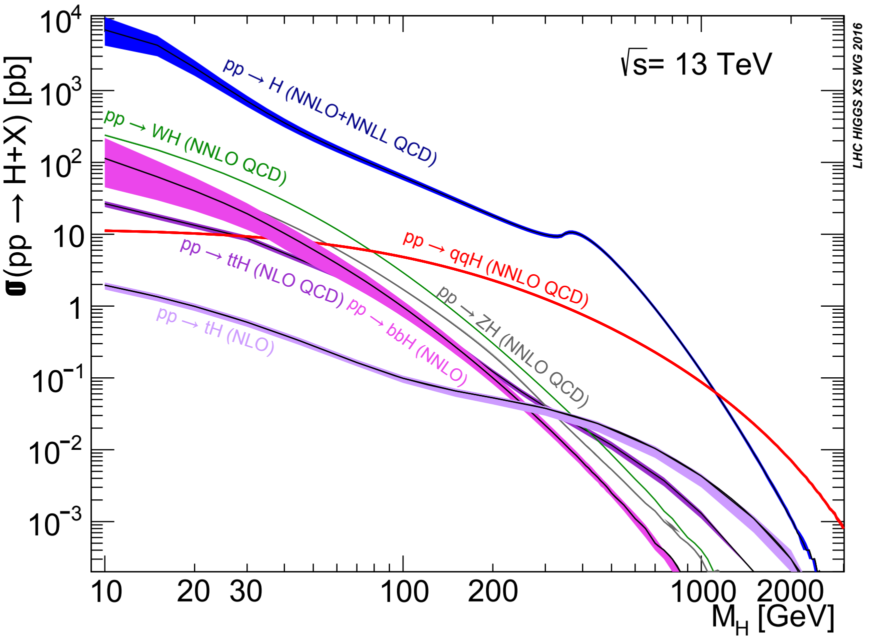


Figure 2.2: Cross sections for various production mechanisms of the Higgs boson originating from proton-proton collisions at the energies of $\sqrt{s} = 13$ TeV [31].

The measurements of properties of the Higgs boson, which followed its discovery by the Large Hadron Collider experiments [32, 33], show a good agreement with the SM [34–36], but the uncertainties on these measurement are not yet small enough to fully exclude the possibility of physics beyond the SM. The current generation of collider experiments is not able to reach the O(MeV) range of precision needed to test the Γ^{H}_{SM} (the total decay width of the Higgs boson predicted by the SM), leaving an indirect way of testing for beyond the SM physics[3] with only one other option - to make use of results arising from measurements of the visible final states in order to set an upper limit on the beyond the SM branching ratio. Using the measurements of the visible final states as well as assuming the Γ_{SM} leads to the result where this approach yields an 95% Confidence Level (CL) upper limit on the branching ratio of the Higgs boson to invisible final state[4] of Br(H→inv) ~ 0.34 [34]. From the SM point of view the only viable decay leaving the invisible signature is H→ZZ→4ν, which has a negligible probability of happening (with Br(H→ 4ν) $\sim 0.1\%$). This all calls for a more detailed approach through a dedicated VBF H→inv search which can further improve the given limit for the branching ratio for the invisible final state.

2.3 Invisible final state

The interest surrounding final states involving particles invisible to hadron collider detectors is due to the potential beyond the SM physics hiding within their cloak. From the perspective of studies involving the Higgs boson properties, the SM predicts that the fully invisible decay is highly suppressed with Br($H \rightarrow 4\nu$) $\sim 0.1\%$ [37], making it a good option for testing for beyond the SM physics. As mentioned in the previous section, the limiting factors arising from the indirect searches create motivation for this, more direct, approach [38–41]. Connection to DM searches can be found in Higgs portal theories [42–45], where the Higgs boson can take the role of a mediator between the particles from two sectors (SM and DM), but more about those in the following section. In order to be able to approach this invisible final state with current detectors, the Higgs boson is required to recoil against a visible system.

The production mode bringing the most sensitivity to this final state is the VBF where, as introduced in Section 2.2, the Higgs boson is associated with two jets. From the SM side, there are a few processes which can mimic the exact final state of interest and those are mostly originating from V+jets processes (where V denotes

---

[3] Within the area of physics concerned with Higgs boson decays.

[4] More details about this approach, alongside with a detailed introduction to the κ framework is given in Ref. [37]

W or a Z boson). Their contribution can be categorised into reducible or irreducible. The completely irreducible scenario is found when the Z boson decays to two neutrinos (the Z $\rightarrow \nu\nu$ decay) and for the W $\rightarrow l\nu$ case (where the charged lepton has been missed in the detection process). Example diagrams showing both the strong and electroweak modes of production (in further text referred to as QCD and EWK modes) of a Z boson decaying to neutrinos are shown in Figure 2.3.

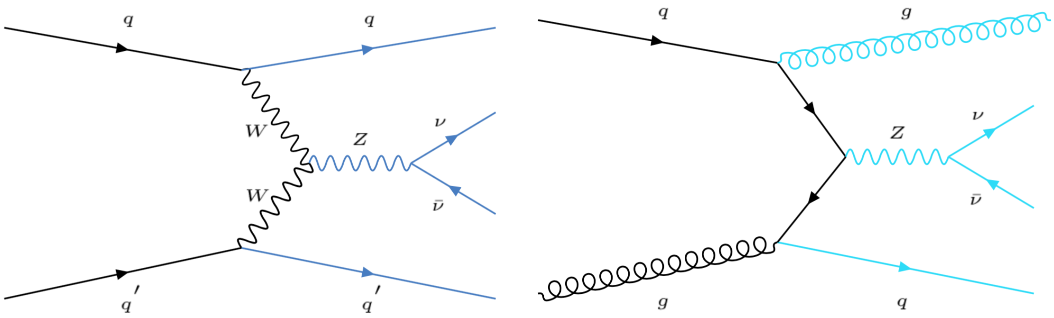


Figure 2.3: Feynman diagrams for two main $Z \rightarrow \nu\nu$ irreducible background SM processes shown for both the EWK (left) and QCD (right) production modes at leading order (colors assigned to final states follow the choice used to mark the corresponding processes in further chapters). The diagrams were made using Ref. [22].

An additional background contribution can arise from strong multijet processes, which due to their large production rate can produce a sizeable contribution which mimics the desired final state. The estimation of the contribution originating from these main irreducible background processes is the focus of Chapter 8 where data driven methods are deployed. Additional minor contributions, which can arise from diboson (VV) and top quark ($t\bar{t}$ and single top) processes, are estimated through the use of simulated samples of respective processes.

2.4 The dark connection

With all the great knowledge regarding particle physics currently written down as the SM, it still does not account for all of the matter in the universe. Many cosmological observations suggest that there exists another form of matter which does not posses an affinity towards the electromagnetic interaction - dark matter [46]. Following the current cosmological model, it is categorised that $\sim 5\%$ of the universe's mass-energy is visible matter, 27 % dark matter and 68% dark energy [47].

The most commonly used observation illustrating this is the appearance of a discrepancy in the distributions of rotation velocities of spiral galaxies when comparing the predicted result (which assumes they are comprised only from visible matter) to what is observed in reality. This is illustrated in Figure 2.4, where the observed velocity distribution $V_C(Radius)$ (with its corresponding fit), represented with points (solid line), is compared with the visible matter only approach (dashed lines denoting disk and gas contribution) alongside a DM halo contribution (dashed line with the point) accounting for the missing piece [48]. This can be expanded with a list of other studies such as the ones focusing on galaxy clusters [49] and gravitational lensing [50].

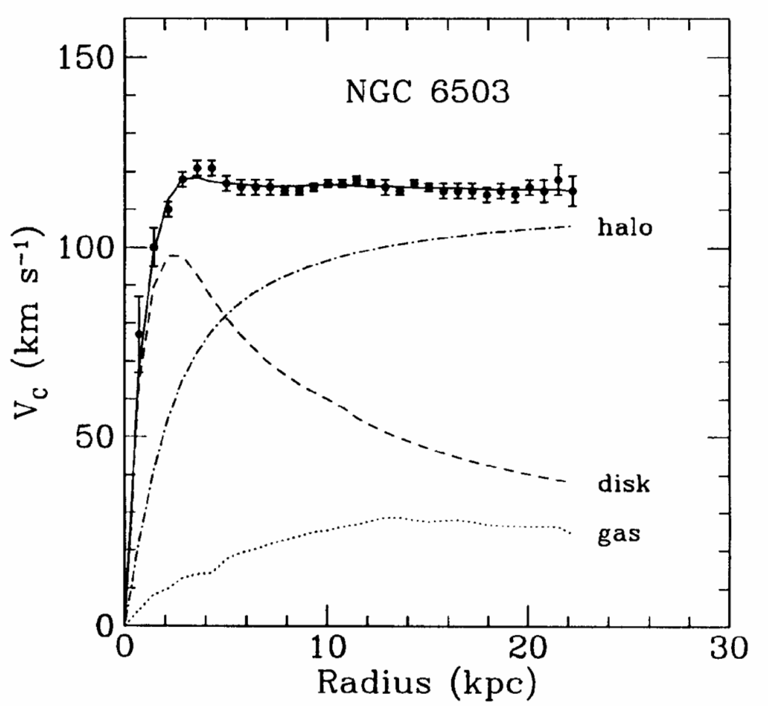


Figure 2.4: The distribution of the rotation velocity for the NGC 6503 is presented as an example of cosmological observation of DM [48].

Staying in the area relevant to the theme of this thesis, the potential candidates for particles comprising the invisible final state discussed above can be associated with the DM sector. This would require a modification of the SM Lagrangian to include terms which enable the coupling of the Higgs boson to the particles from the DM sector. These can include one of the following scenarios [38, 42, 51]:

$$\mathcal{L}_S = -\frac{1}{4} g_{HSS} H^\dagger H S^2, \; \mathcal{L}_V = \frac{1}{4} g_{HVV} H^\dagger H V_\mu^\dagger V^\mu \text{ and } \mathcal{L}_f = \frac{1}{4} g_{H\chi\chi} H^\dagger H \overline{\chi}\chi \tag{2.1}$$

where $\mathcal{L}_S$ shows the (quartic) interaction term connecting the SM Higgs doublet with the DM sector, this time being represented with a scalar type (S). The equivalent definitions stand for the latter two terms with the scalar DM field being replaced with a vector (V) and a fermion (f) type, respectively.

2.5 Current status

The most recent publication on this matter related to the CMS experiment represents a combination of efforts from both the entire Run 1 and early Run 2 phase of operation of the LHC[5] [39]. It yields a value of the 95% CL upper limit on the branching ratio for the invisible final state of Br(H $\rightarrow$ inv) = 0.19 (0.15) for the observed (expected) scenario, through the combination of various analyses targeting different production mechanisms of the Higgs boson. Figure 2.5 shows the 95% CL upper limits on the Br(H $\rightarrow$ inv) with respect to each production mode of interest (denoted as tag, eg. VBF-tag) for the 2016 era of data taking (representing the early stage of Run 2). The additional tags, besides the already discussed VBF, focus on the VH production, which is split into two separate tags depending on the decay of the vector boson being fully leptonic (Z$\rightarrow$ll) or hadronic (V$\rightarrow$qq), and the ggH production (with the requirement of having one jet from ISR associated with it).

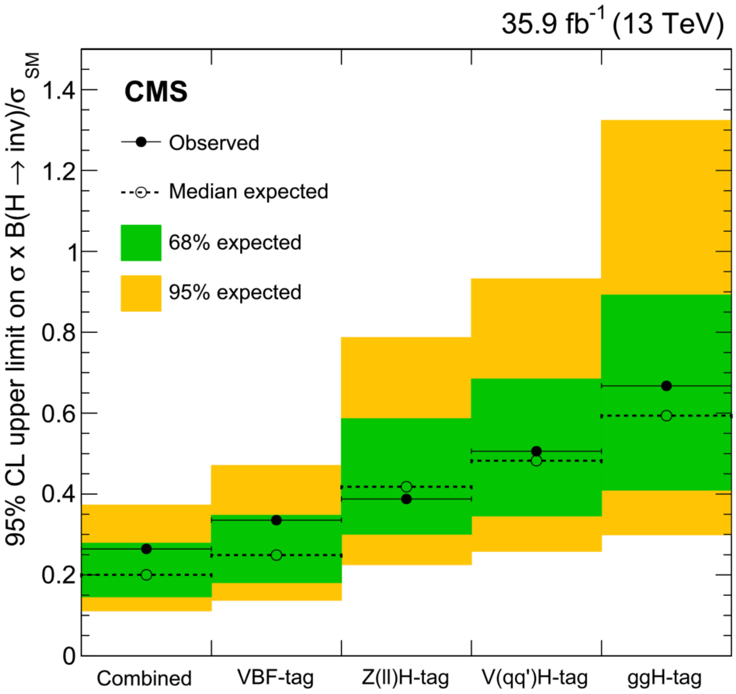


Figure 2.5: Summary of results from the CMS experiment for the 95% CL upper limit on the $\sigma\times$ Br(H$\rightarrow$ inv)$/\sigma_{SM}$ under the assumption of the SM Higgs boson, for the 2016 era of data taking. [39].

Continuing the story which began in Section 2.4, these results can be interpreted in terms of the upper limit on the spin-independent DM-nucleon interaction cross section, making it easier to compare with direct detection experiments. The conversion can be made using the following relations (with the assumption of a scalar or

---

[5] More details on the different phases are given in Chapter 3.

a fermion DM particle respectively) [38, 42]:

$$\mathrm{Br}(\mathrm{H} \to \mathrm{inv}) = \frac{\Gamma_{inv}}{\Gamma_{SM} + \Gamma_{inv}} \tag{2.2}$$

$$\sigma^{SI}_{Scalar-N} = \frac{4\Gamma_{inv}}{\beta m_H^3 v^2} \frac{f_N^2 m_N^4}{(m_N + m_{DM})^2}, \tag{2.3}$$

$$\sigma^{SI}_{Fermion-N} = \frac{8\Gamma_{inv} m_{DM}^2}{\beta^3 m_H^5 v^2} \frac{f_N^2 m_N^4}{(m_N + m_{DM})^2}, \tag{2.4}$$

where the values of the parameters are: $\Gamma_{SM} = 4.07$ MeV, $m_N = 0.939$ GeV (average mass of the proton and neutron), $v = 246$ GeV, the DM candidate's mass (m_{DM}), $\beta = \sqrt{1 - \frac{4m_{DM}^2}{m_H^2}}$, and the $f_N = 0.308$ (nuclear form-factor [39, 52]). One last important step is that the the main result, the Br(H→inv) is expressed in terms of the 90 % CL upper limit instead of the usual 95 % in order to be comparable with the results presented by the direct detection experiments. Figure 2.6 shows the comparison of the CMS experiment's results compared to results coming from direct detection experiments. These show that the approach taken by this analysis, covering $m_{DM} < \frac{m_H}{2}$, leads to a set of results complementing the majority of direct detection experiments.

This chapter concludes the first part of this thesis which served as an introduction to main mechanisms of the SM and the motivation driving the search for the invisible final state of the Higgs boson. The following chapters introduce the experimental setup - the CMS experiment. There it will be presented how these measurements are made possible on the detector and trigger level. The continuation of this idea from the perspective of the newly available data is the focus of Chapter 7 and beyond, where a combination of new trigger ideas and improved analysis strategy is discussed in great detail.

* * *

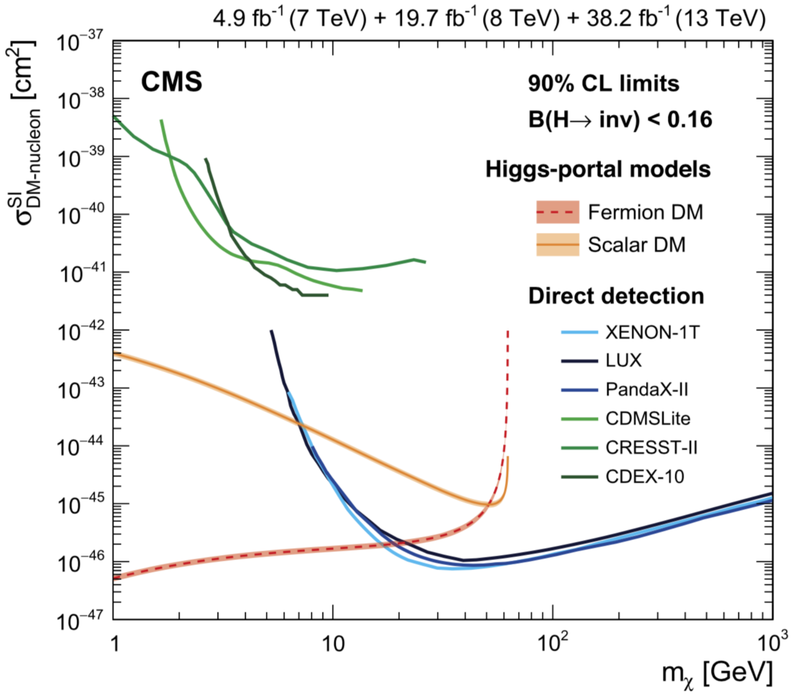


Figure 2.6: The reinterpretation of the CMS results in terms of the 90% CL upper limits on the spin-independent DM-nulcleon scattering cross section when assuming a fermion (red) or a scalar (orange) DM particle (presented as a function of m_{DM}, denoted here as m_χ). Limits are compared with results originating from direct DM detection experiments [39].

Part II: The experiment

“Bez alata nema zanata.”

– Serbian proverb –

Chapter 3

The LHC and the CMS experiment

"You cannot swim for new horizons until you have courage to lose sight of the shore."
— William Faulkner

3.1 Introduction

The European Organization for Nuclear Research, better known by its abbreviation "CERN", stands tall as a pillar of the human determination towards understanding nature. Paving the way for future collaborations by being one of the first examples of unity after the horrors of the decade preceding its establishment, CERN has been a home to experts coming from a multitude of fields ever since it opened its doors on the 29th of September 1954 [53].

3.1.1 A new dawn

Strategically positioned near Geneva (Switzerland), the newly formed institute had its eyes set on becoming the world's leading research facility for high energy physics. The following sections briefly summarize the overall achievements of this collaboration which lead to the fulfilment of this goal and explain how bold decisions have resulted in the state of the art experiments operating at the moment. The easiest way to describe it would be to categorize the main benefits of CERN into three groups: scientific results, industry application and education.

Starting with the scientific achievements, the discovery of weak neutral currents [54, 55] in 1973 made a huge breakthrough by confirming one of the basic ideas of the SM. Fast forward to ten years later when, after building on the previously gained experience, a new ground breaking result was reported from CERN's experiments. It was the experimental evidence of the existence of W and Z bosons [56]. Further technological advances led to the design and the creation of the Large Electron-Positron (LEP) collider [57]. What followed was a series of high profile results, some of which will be discussed in the next section, that culminated in the creation of the LHC and the subsequent discovery of the Higgs boson by the CMS and ATLAS experiments in 2012.

Looking at the aforementioned discoveries one could naively assume that, while they represent a giant leap in our understanding of nature, nothing coming out of CERN's doors has a direct impact on everyday life. Yet this couldn't be further from the truth. It can easily be seen by looking at something that is today considered an essential tool in lives of many, the World Wide Web standard. Born in the dark corridors of building 1 within the research campus of CERN, the basis of our digital life was created by Tim Berners-Lee [58] in the early 1990s for the purpose of easier data management within the experiments. This standard introduced the concept of associating Universal Resource Locators (URLs) to documents and other objects of interest as means of identifying and accessing them via the internet. After being used internally for two years, it was concluded that the possible applications far exceeded its original goal and thus it was released to the public, setting a basis for what we now call web browsing.

This general concern about the practicality of published results plagues not only high energy physics but also many of the fundamental branches of science. This constant questioning is even more baffling due to the fact that if only a minuscule investigative effort was made by those who present such claims, then it would be clear that many of technological leaps were enabled through fundamental research. Moving from the realm of the internet, additional examples of this can also be found in medicine where the contribution of accelerator sciences range from hadron therapies, which provide a less invasive way of treating cancer cells, to the development of super absorbent polymers for the purpose of creating thin but still multi layered baby diapers.

Finally, there is one additional way in which CERN impacts our society as a whole and that is through its education platform. Its main purpose is to give new generations of students the opportunity to participate in various programmes that range from high school projects to undergraduate internship positions. This approach has influenced a large number of young people from all over the world to get involved in this field from a fairly early age. Case in point, the author of this thesis began his journey into the field of particle physics with his high school graduation thesis by looking into a set of simulated samples for a four muon decay of the, then still theorised, Higgs boson. Even though it was a fairly simple project, it was still a harbinger of a, hopefully long, future research career.

3.1.2 Down the road less travelled

As a good rule of thumb, the HEP scientific community organizes meetings and workshops every few years in order to encourage the discussion regarding future experiments. This should come as no surprise as planning ahead and trying to see the bigger picture represent actions closely associated with almost every successful long term venture. The need for this approach became more evident in the decade that saw the experimental confirmation of the existence of the neutral current and the discovery of W and Z bosons, as it became clear that there was an urgency for a structure that could probe the, then freshly finalized, Standard Model.

The decision to move away from proton-antiproton machines, for which CERN had developed expertise in the previous decade, in favour of an electron-positron design had a much bigger impact on the institute (and the HEP community as a whole) than it was initially envisioned. From a strictly physics point of view, this opened the door to a much cleaner slate that enabled detailed probing of the electroweak sector. On the other hand, taking a look at the engineering and monetary side, a circular collider being built under Geneva had pretty much set in stone the future of Europe's particle physics strategy. The grandiose projects being discussed in the USA (the Superconducting Super Collider [59]) and Russia (the proton Accelerating and Storage Complex - UNK [60]), some of which were promising center of mass energies much larger than those achieved by the LHC today (although with much less total projected luminosity), were halted and ultimately cancelled due to budgetary and political constraints. It should come as no surprise that one of the main reasons for the success of the LHC project was the fact that it re-uses the same tunnel that was built years ago for LEP.

The construction of the tunnel turned out to be one of the biggest civil engineering achievement in Europe at the time. The completion of the, 27 km in circumference, home of the accelerator complex was done in 1988 after a five year effort. Time for proper celebration came one year later, when the first beams were collided in August of 1989. Following the planned scientific program, the first item on the agenda was to probe Z bosons, which is why the starting beams each bore the energy of 45 GeV, thus matching the boson's mass in the center of mass frame. Following in the same vein, the next interest was to take the total energy higher, leading it to the the WW production range. It required an upgrade of the accelerator, resulting in the increase in energy to 100 GeV per beam. The limit of beam energy of the machine reached 209 GeV near the end of its operation period. At the same time a possible appearance of a Higgs boson-like particle was looming over the planned end date of LEP and the beginning of LHC's construction.

At the time thought of as a possible swan song for LEP, a 91 % confidence level (or 1.7 σ) indication of an excess at around 115 GeV was reported by researchers in early 2000. Prompted by this information, an extension of a couple of months was approved in order to further explore this appearance. Unfortunately for researchers, this time didn't yield any relevant increase in significance and LEP was shut down by the end of the year, leaving a sense of a lost opportunity in the eyes of many [61]. All the subsequent arguments of biased decision making were confirmed to be void when the 125 GeV Higgs boson was discovered by the CMS and ATLAS experiments.

3.1.3 The Large Hadron Collider

The successor to LEP, the LHC [21] (schematically presented in Figure 3.1), originated with the idea of further testing the phase space of high energy physics. Commonly nicknamed the "Higgs discovery machine", it was designed with a much broader physics programme. Representing a natural evolution from the previous generation, this machine was built to collide protons at very high energies ranging from 7 to 14 TeV (in the centre of mass frame).

The diversity of research areas covered by the LHC can best be seen in by looking at the coverage achieved by main experiments residing at four beam intersection points. Two general purpose experiments, ATLAS and CMS, take the center stage, being the main reason for the LHC's nickname. Adding the heavy ion research to the list, the ALICE experiment takes the role of a primary experiment in within the field [62]. Finally, the LHCb experiment [63] proved crucial to the research ideas involving heavy flavour physics.

The journey of a proton bunch begins with a canister of hydrogen gas, whose atoms are stripped from electrons by using an electric filed while the remaining protons are being passed on to the first step of acceleration - CERN's Linear Accelerator 2 (LINAC 2). Bringing the energy of particles to 50 MeV, it passes them on to the second step, the Proton Synchrotron Booster (PSB). It further increases the energy of the particles reaching the values of 1.4 GeV before continuing towards the Proton

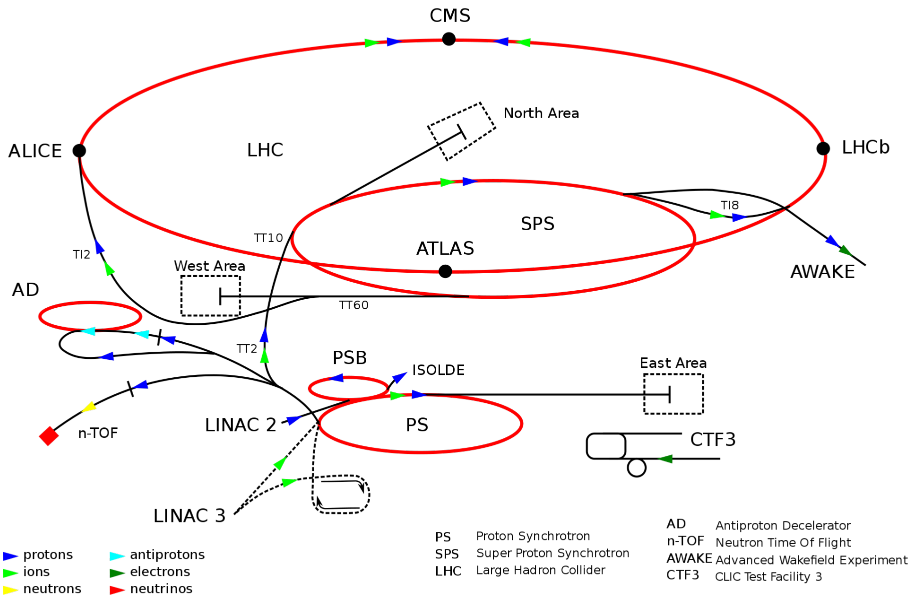


Figure 3.1: Schematic representation of the LHC complex.

Synchrotron (PS), where bunches reach the energies of 25 GeV. The final step in this process is the Super Proton Synchrotron (SPS) which brings the increase in energy to a total of 540 GeV. After this, particles finally reach the LHC, where their energy is further increased up to 7 TeV.

The status of the LHC operation is best described when separated into three parts. During the writing of this thesis these could've been defined as the past, present and future of collider physics. Ranging from the period of 2009 to 2013, the past is manifested in the form of the Run 1 phase of operation. Colliding beams at 7 (and later 8) TeV, the first cycle of operation brought in around 27 fb^{-1} of collected data, when looking from the CMS experiment's point of view. The beginning of this era was marked with a more than a year long delay, due to damage caused by an unexpected magnet quench. After the mandatory stop the journey was continued, first conservatively, by reaching energies of 1.18 TeV per beam. This was enough to beat the previous record held by Tevatron collider (Fermilab) [64] and to put a first check on LHC's goal list. Increases in operational energy continued starting with 3.5 TeV in 2010 and continuing in 2012, when the energy was set to 4 TeV per beam marking the maximum reached during this stage. The most important achievement of this era (or better to say until today) was the discovery of the Higgs boson in July of 2012. In order to be better prepared for the next stage, the LHC went into a shutdown period in 2013.

The first collision of 6.5 TeV beams achieved on the 20th May of 2015 sounded the beginning of the current, Run 2, era of operation. Starting with the summer of 2016, the LHC reached its designed luminosity of $1.0 \cdot 10^{34}$ cm^{2}s^{-1} (increasing it to twice the value before the shutdown). Bringing in a total of ~ 137 fb^{-1} of data, the Run 2 phase creates opportunities for precision measurements and probing of beyond the SM phase space (at the very least, excluding parts of it). Detector upgrades, some of which are going to be discussed in the following sections, brought a more efficient way of selecting relevant data by inserting complex, mode targeting selections early during the triggering process. Currently, LHC experiments are working towards publishing a multitude of ”legacy” results that will summarise and publicly release the conclusions made during this phase of operation.

The future begins with the Run 3 phase of operation, which is expected to bring new ways of selecting data. Applying industry standard Machine Learning (ML) algorithms for research purposes, newly created triggers are expected to bring a significant increase in sensitivity for many of the beyond the SM searches. Following this period, another upgrade stage is set to take place in order to prepare the machine for the next big step - the High Luminosity LHC. Figure 3.2 showcases the LHC operation timeline with the projected integrated luminosity[1].

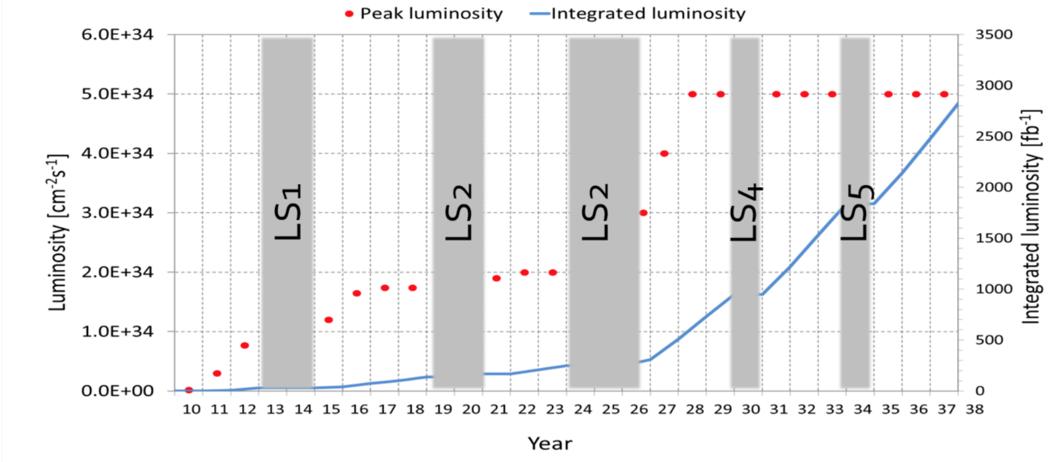


Figure 3.2: Performance of LHC, extended to include the luminosity projections for up until the end of the HL-LHC phase [65]

The contents of the chapter following the completion of the LHC era are yet to be set in stone. A set of different proposals exist, with each of them having benefits and constrains in both monetary and physics sense of speaking. One highly advocated proposal is the Future Circular Collider (FCC) project [66]. It discusses a similar approach to one used with LEP/LHC years ago. During the early years of the project, an electron-positron collider would be built (FCC-ee) in order to probe the electroweak sector even further, while the technology is being prepared for the hadron collider (FCC-hh) that would bring the increase of the operational energy to the record level of 100 TeV.

---

[1] As of early 2020, this timeline has been altered to accommodate for delays caused by the global pandemic.

3.2 The Compact Muon Solenoid

Taking the role of a leading general purpose detector[2], the CMS detector was designed to operate under proton-proton and ion collision beams with energies of 7 and 2.75 TeV respectively [67]. Its design philosophy revolves around a few items, with the main, discovery of the Higgs boson, already being marked as done early on during the initial Run 1 phase of LHC operation. As mentioned in the previous section, now the focus is to unravel what lies at TeV energies, either through precision measurements and studies of known particles or by trying to search for new physics using theoretical models whose ideas can be quantified in a collider experiments.

The CMS experiment itself is a cylindrical, layered detector comprised of several subsystems (as shown in Figure 3.3) working in harmony to perform particle detection and reconstruction. The subdetectors forming the CMS experiment are: the tracker, the electromagnetic calorimeter, the hadronic calorimeter, the superconducting solenoid, the muon chambers and the return yoke [68]. The general discussion of its operation can be split into two topics: the detection and data acquisition. Both of these will be covered in the following sections.

3.2.1 Geometry

The CMS experiment is positioned in the French village of Cessy, tucked in between the beautiful sights of the Jura mountains on one side and the city of Geneva from the other, allowing for a more descriptive approach when fixing the coordinate system [67]. The main collision point is chosen to be the point of origin, from which the y axis follows the vertical line leaping towards the sky. The x axis is chosen as the one which points towards the center of the LHC ring, leaving the z axis to be the one that is tangent to the beam line, while also enjoying the view of the aforementioned mountains (as illustrated in Figure 3.4). The definition of polar coordinates follows the standard procedure with the θ (polar) angle being defined from the z axis and the ϕ (azimuthal) angle being placed in the x-y plane starting from the x axis. In collider experiments it is usually convenient to define alternative geometrical variables used to describe an object's position. This leads to the definition of peseudorapidity as:

$$\eta = -ln\left[tan\left(\frac{\theta}{2}\right)\right] \tag{3.1}$$

---

[2] A completely unbiased observation by the author

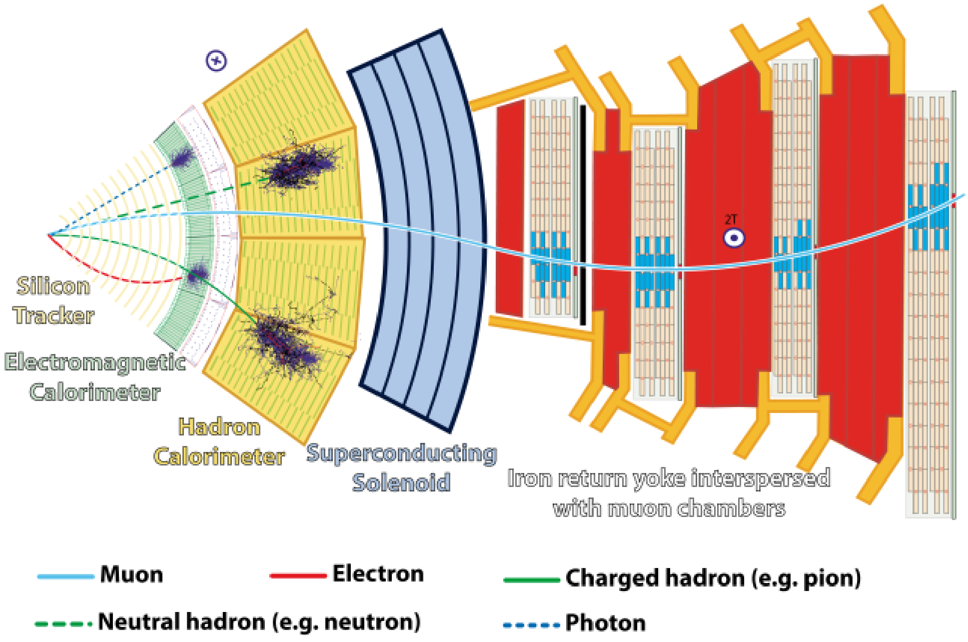


Figure 3.3: Transversal representation of different layers of the CMS experiment [69].

Another variable commonly used to quantify the geometrical distancing of the particles is ΔR. For two objects it is defined as:

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{3.2}$$

Due to the structure of the CMS detector itself, it is important to define additional variables in the x-y, in further text the transverse (or the r-ϕ), plane. Variables connected to it such as the transverse momentum and energy will be denoted as p_T and E_T respectively. The most important properties for this study are the missing transverse momentum and energy. They can be written as:

$$\begin{aligned} \vec{p}_{T,miss} &= -\textstyle\sum_i \vec{p}_{T,i}, \\ E_{T,miss} &= |\vec{p}_{T,miss}|, \end{aligned} \tag{3.3}$$

where the sum goes over all particles reconstructed by the detector. This imbalance in transverse momentum represents a valuable asset when quantifying the contribution of particles that are invisible to the detector. This will become more evident in the following chapters, where $E_{T,miss}$ will be used extensively.

The following sections will further dissect the structure of the CMS experiment using the terminology introduced above. Two terms that will be a recurring item are the barrel and the endcap region representing a literal way of describing two general structural parts when discussing subdetector layers.

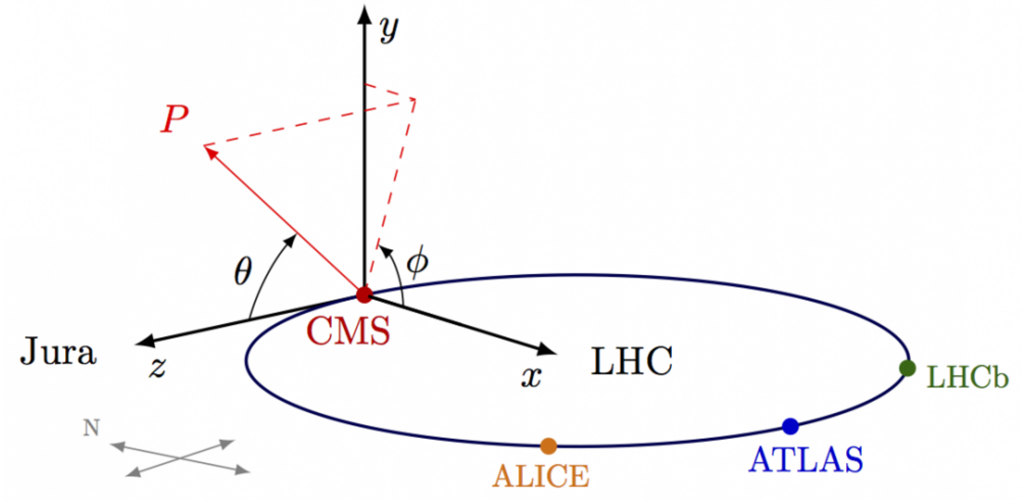


Figure 3.4: Graphical representation of the coordinate system fixed to the CMS experiment [70]

3.2.2 Tracker

The reconstruction of charged particle tracks and the measurement of the corresponding momenta is a crucial task for any collider experiment. The basis for tracking is formed around the interaction of charged particles with the atoms of the material they are travelling through. In a more precise way, the method revolves around exploring the trail of ionised atoms and free electrons left by particle moving in a magnetic field. If an electric field is introduced in order to separate those electron-hole pairs, a current pulse will be created allowing for a "hit" signal to be detected.

The, semiconductor based, tracking detector of the CMS experiment [67, 71, 72] is positioned around the beam line, fully surrounding the interaction region as shown in Figure 3.5. It's design philosophy follows that it has to sustain the high collision rate given by the LHC, while providing the output in a form of particle tracks originating from the primary and secondary interaction vertices for the pseudorapidity range of $|\eta|$ <2.5. The LHC input characteristics impose requirements that the tracker needs to have a highly granular structure, to provide longevity of the system and to have the ability to separate tracks originating from different bunch crossings (in other words, a fast enough response).

All of this leads to the tracker being a fully silicon detector. The system itself is made out of two, operationally independent, parts: the pixel detector (inner layer) and the the strip detector surrounding it, bringing the total dimensions to 5.8 m in length with a 2.5 m in diameter (following the cylindrical design of the whole CMS experiment).

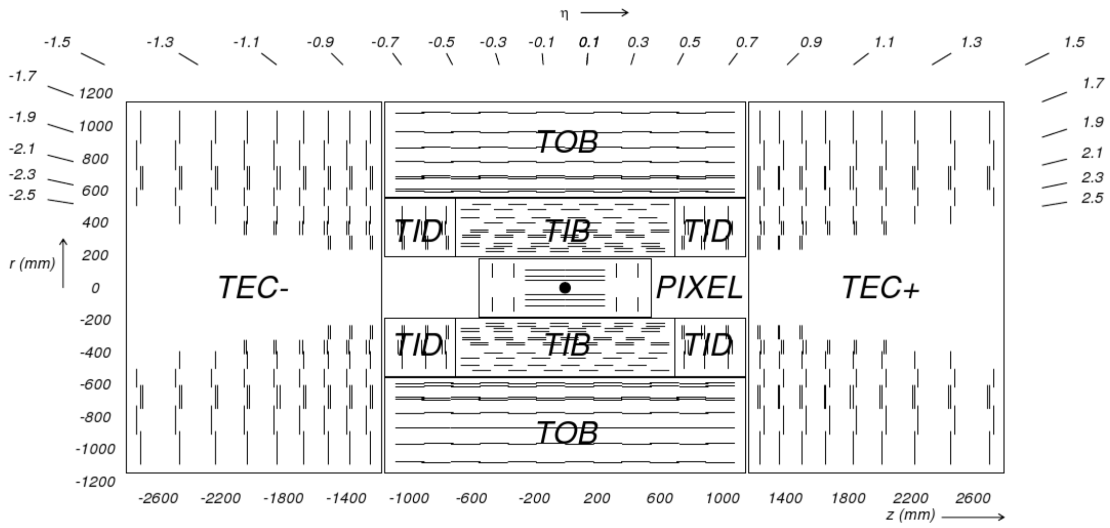


Figure 3.5: Structure of the CMS experiment's tracker system showcasing the subsystems in the barrel and endcap regions [72].

The pixel detector, covering the surface of 1.1 m^2, represents the first line of detection within the CMS experiment. Its barrel region (located within the $|\eta|$ <0.9 range) is formed out of three cylindrical layers of detector modules with the radius being 4.4, 7.4 and 10.2 cm respectively. Forming the endcap region, an additional two discs containing pixel modules are placed, completing the setup, and giving a total of 66 milion pixels. It outputs three spacial points of detection associated with each charged particle in the event. The pixel subsystem represents an essential tool in reconstruction process of secondary vertices which play an important role in detecting heavy flavor and tau decay products.

The second part of the tracker, the strip detector, is made from a 10 layered barrel region and a set of 9 discs forming each endcap side. The Inner Barrel (TIB) and Discs (TID) form the first out of three main subsystems, delivering up to four detection points in the r-ϕ plane. The Tracker Outer Barrel (TOB) coveres the TIB and TID systems, allowing for additional points to be detected. Finally, shifting to the endcap region, there are the Tracker Endcaps (TEC+/-) providing up to nine ϕ detection points. An additional micro strip module is placed in all subsystems in order to enable the estimation of the missing component (r for the endcaps and z-axis position for the barrel).

The importance of the tracking system is also seen in particle identification at the trigger decision level where seed tracks are being constructed and passed as input data for the trigger algorithms to make an online decision.

3.2.3 Electromagnetic Calorimeter

The next subsystem in particle's journey is the Electromagnetic Calorimeter (ECAL) [68, 73]. It is comprised of 75848 $PbWO_4$ (lead tungstate) crystals. The logic behind these types of detectors is based on the behaviour of high energy electrons and photons when passing through a material with a high atomic number. For electrons, the main type of interaction with matter (in this energy range) is bremsstrahlung, while the photons interact through electron-positron pair production. This, at the detector level, is cumulatively seen as an electromagnetic shower and is used to measure the energies of electrons and photons by amplifying the scintillation light produced by ECAL crystals as a response to particles moving through them. Additionally to these e/γ processes, there can be another contribution coming from π^0 decays to two photons originating from hardonic showers. The performance of the ECAL detector was crucial for the efforts towards the discovery of the Higgs boson. Its driving force and ultimately its most important role was associated with searches within the $H \to \gamma\gamma$ decay mode (as well as the rest of modes containing e/γ in the final state).

Scintillating crystals are chosen because of the compatibility of their properties with the expected performance of the LHC. The short radiation length ($X_0 =$ 0.89 cm) ensures a small Molière radius ($R_M = 2.2$ cm), allowing for compact dimensions of the detector. Being made out of a high density and optically clear crystal, ECAL can cope with the radiation impact as well as provide a compatibility with the bunch crossing time.

The detector is made out of a barrel section which is hermetically sealed with the endcap sections. The ECAL barrel (EB) section is responsible for energy deposits left within the $|\eta| <1.47$ range while the Endcap (EE) sections take over the responsibilities for the $1.68< |\eta| <3.0$ range. The EB/EE crystals are $\approx$25/26 X_0 in length respectively. This choice of dimensions for the side facing the particle's trajectory is motivated by the size of the EM shower left in the crystal. Having such frontward crystal dimension simplifies the particle identification as the shape of the shower can easily be added as a selection criteria. A more visual overview is given in Figure 3.6, which showcases the structure of the ECAL where the aforementioned crystals are arranged in 36 supermodules. In order to have another helping hand, a preshower detector, which aids especially when it comes to π^0/γ separation, is added to the setup. It is positioned in front of the EE rings covering the $1.64< |\eta| <2.5$ region with its set of lead absorbers and silicon strip sensors.

With all the benefits that this choice of crystal structure brings, there are bound to be certain downsides. One additional property of lead tungstate crystals is that

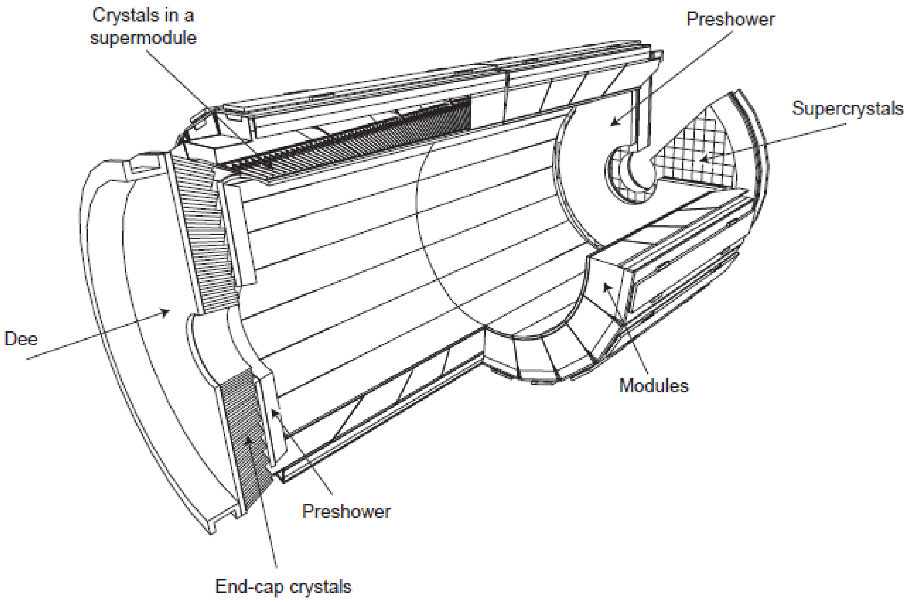


Figure 3.6: Schematic view of the CMS ECAL subdetector. The main subsystems (EB, EE and the preshower detector) as well as their substructure are presented [74].

they have a very low light yield. A set of signal amplifiers is needed in order to battle this disadvantage. For the EB region, a set of silicon avalanche photodiodes (APD) is used to boost the signal by a factor of ~50 compared to the original. A similar scenario is deployed for the EE region, where vacuum phototriodes (VPT) with amplification ability of ~10 times are used.

The energy resolution of the ECAL detector can be parametrized as:

$$\frac{\sigma^2}{E^2} = \left(\frac{A}{\sqrt{E}}\right)^2 + \left(\frac{B}{E}\right)^2 + C, \tag{3.4}$$

where A, B and C represent the stochastic, noise and constant term respectively. The stochastic term is measured to be 2.8 % . The noise term, comprised from the cumulative effects of pileup, electronics and digitisation noise, has the value of 12 %. Finally, the constant term of 0.3 % is assigned to account for any leakage of energy or inter calibration errors.

The story of the ECAL will continue in Chapter 4, where it will be shown how these outputs are used at the first level of selection and beyond.

3.2.4 Hadronic Calorimeter

Following a similar path derived for detecting the residuum of electromagnetic interactions of particles, a separate detector system has been added to measure the energies of hadronic showers - the Hadronic Calorimeter (HCAL) [68, 75]. When neutral and charged hadrons interact with the detector material through the strong interaction, a hadronic shower will be formed with associated parameter λ, the nuclear interaction length. Representing a mean distance between hadronic interactions within the shower, it takes much larger values than its electromagnetic equivalent, the radiation length. This would imply that the HCAL has to have larger dimensions in order to be efficient, but on the other hand the compactness of the CMS detector fixes its position (and maximum dimensions) in between the ECAL and the solenoid coil. This potential disadvantage is overcome with the structural design of the HCAL, which is made out of: barrel/endcap sections (HB/HE), a Hadron Forward (HF) subdetector and the outer barrel hadronic calorimeter (HO). Figure 3.7 shows the structure of the HCAL and all of its aforementioned parts.

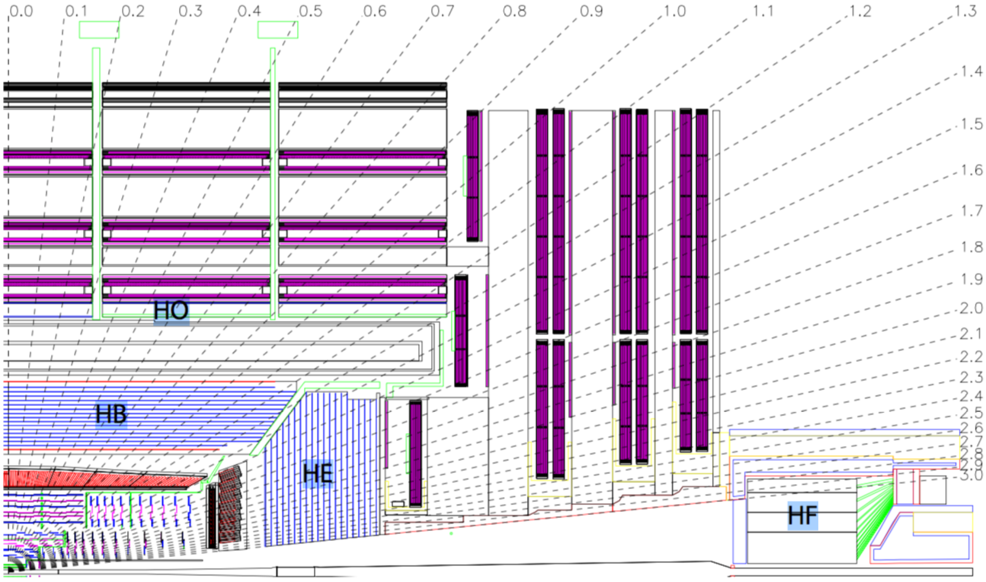


Figure 3.7: Schematic $(\mathrm{r}-\mathrm{z})$ view of the CMS HCAL detector highlighting different subsystems: the hadronic barrel/endap (HB/HE), forward (HF) and outer (HO) detectors [68]

The HCAL barrel is itself comprised out of two identical halves (HB$^+$/HB$^-$) covering the $|\eta| < 1.3$ region. The subsystem is made out of layers of brass/steel (absorbing material) and scintillator tiles. A total of 32 identical wedges (each being segmented into four ϕ regions) made out of absorber plates form each of the barrel regions, with the scintillator tiles being positioned into 16 η regions, bringing the total segmentation to $(\eta, \phi) = (0.087, 0.087)$. The absorber itself is formed out of a front and back set of steel plates, in between which resides a group of 11 thick brass plates. The total thickness varies with $|\eta|$ from 5.8 to 10.6 λ (for the edge values of $|\eta| = 0$ and 1.3 respectively). On top of the HB thickness, there is a contribution coming from the previous detector, where the ECAL brings an additional 1.1 λ to the total value. The scintialltor tiles are equipped with wavelength shifting fiber readout, which is responsible for the transfer of emitted optical light. In order to battle the, previously mentioned, constrain in barrel dimensions, another subsystem is added after the solenoid structure (covering the $|\eta| < 1.26$ region). The addition of the HO serves to effectively extend the size of the barrel section by moving the minimal thickness value up to 11 λ.

Moving on to the HE, it is put in place to be in charge of the $1.35 < |\eta| < 3.0$ region following the same brass/scintillator structure used in the HB sections. Covering the diameter region starting from 0.8 m until 6.0 m, the endcaps have the same (η, ϕ) segmentation as the HB sections for almost the entire η range (the edge case being the $|\eta| \sim 3.0$ region, where it doubles in value). A set of two HF calorimeters is placed behind the ends of the HE section. Each of them is constructed from a Cherenkov detector comprised out of quartz fibers placed into iron and are used due to their radiation durable properties to extend the coverage up to $|\eta| < 5.2$.

3.2.5 Magnet

The central component of the CMS experiment (and one third of its name) is the 4 T superconductive solenoid [68] used to bend the trajectories of charged particles. It is 12.5 m long cylindrical structure with a diameter of 6 m. A high number of ampere-turns needed in order to generate such a strong field (41.7 MA-turns), introduced a new, four layer winding design which was a clear departure from solutions used with previous experiments (maximum of two layers). During run periods, the value it operates on is bellow its design value and stands at 3.8 T in order to ensure the durability of the system as a whole.

Another important part of the magnet system is the return yoke. It is made out of 5 wheels in the barrel and 3 discs per endcap section interspersed alternately between layers of muon chambers (described in the following section). Contributing to the majority of the CMS experiments' weight, it stops almost all particles reaching it (except the muons and neutrinos).

3.2.6 Muon chambers

The final detector system discussed in this chapter are the muon chambers. Being a focal point for one of the main searches for the SM Higgs boson, the four muon final state represents the cleanest signature out of all other four lepton combinations due to the fact that muons are the least affected by energy loss in the previous sections (namely the tracker). This has influenced the way the CMS experiment was designed in a similar way how ECAL was connected to the diphoton final state.

In order to approach the detection of muons, a setup is created through the use of three subsystems [68, 76] based around gas chambers (shown in Figure 3.8).

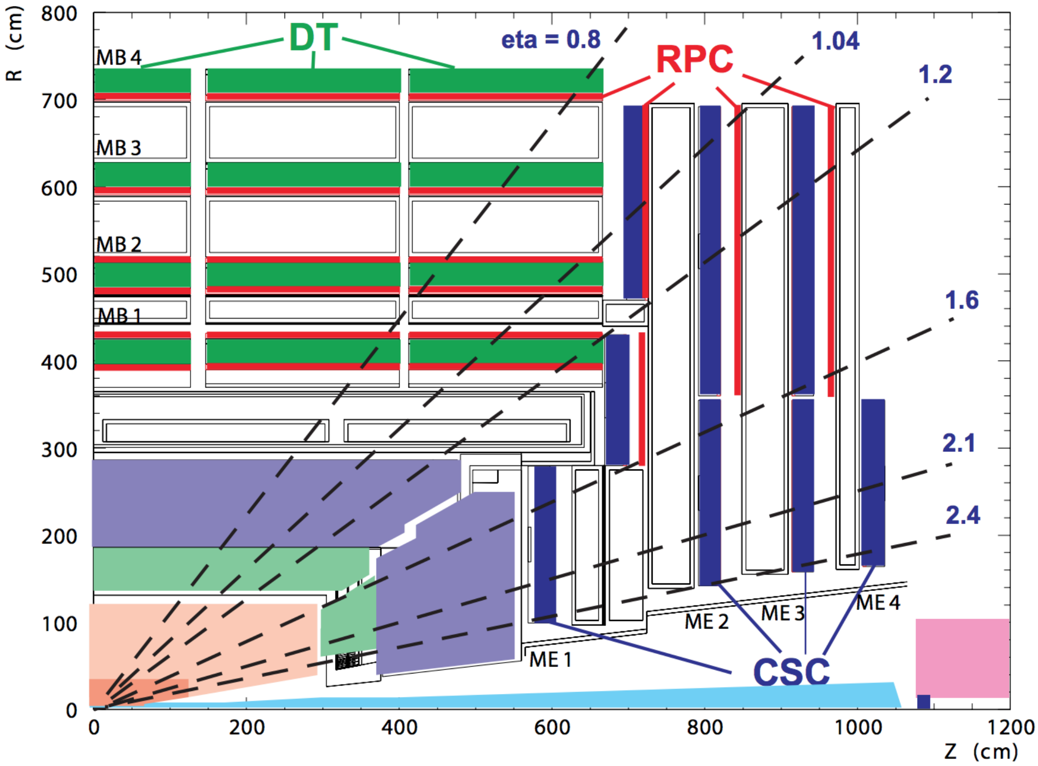


Figure 3.8: Structural design of a (r − z) view of the CMS Muon system highlighting different sections: the DT, CSC and RPC detectors [67]

Starting from the barrel section, the $|\eta| < 1.2$ region is covered through the usage of Drift Tube (DT) chambers. From the structural point of view, each DT is made out of a stretched wire placed in a gas volume. Comprised from four stations inserted within the yoke, the DT subsystem is responsible for measurement of muon coordinates for this η range. The first three inner stations (each containing 60 chambers) are responsible for mapping in both the r-ϕ plane as well as the z axis projection, while the last station (comprised out of 70 chambers) lacks the latter ability. For the encap region, a set of Cathode Strip chambers (CSC) is used instead of DTs. The change in the approach was introduced due to a much larger muon rate in this region, where radiation harder material and faster response given by the CSCs has proven to be a better option.

In order to battle the large amount of information, a dedicated trigger was introduced to both DTs and CSCs. Based on Resistive Plate chambers (RPC), this system provides fast triggering response over the majority of the η range ($|\eta| < 1.6$), helping with muon track reconstruction scenarios starting from multiple hits per chamber (in case of DTs and CSCs). They employ a double gap chamber structure which in return provides good performance during operation. The barrel section contains six layers of RPCs, while the endcap has areas covered with them for the first three sections, positioned in such way to complete the setup for an efficient (compact) muon detection system.

Chapter 4

Data acquisition

> *“It has long been an axiom of mine that the little things are infinitely the most important.”*
> — Sir Arthur Conan Doyle, The Memoirs of Sherlock Holmes

4.1 Introduction

The sheer amount of data generated by the LHC brings a lot of opportunities for physicists to explore at the TeV scale. As it is usually the case, one advantage also brings a plethora of difficulties that need to be overcome, usually requiring a more practical approach. When looked at the maximum collision rate at the LHC (40 MHz), it can be seen that the total stress on the data acquisition system is around 40 million collision events per second, which is far beyond the reach of any currently available data storage systems. In order for the CMS experiment to be able to perform as efficiently as planned, a two stage structure was deployed with a task of discarding events deemed ”unworthy” by its logic.

Facing this problem head on, the CMS triggering setup consists of a, hardware based, first level of selection - the Level-1 Trigger (L1T) and a software High Level Trigger (HLT). The initial selection process done using the L1T scales down the total input rate to, a more approachable, 100 kHz. These decisions are based around a limited set of information coming from a subset of subdetectors. The next stage in the triggering process, the HLT selection, is designed to further reduce the resulting output to ∼1 kHz of data, which is manageable by both the storage and offline reconstruction systems. Each of these triggering levels has a dedicated set of algorithms combined into a structure called the trigger menu. It is designed with the goal of allowing the passage of the appropriate rate of events expected from that system. The following sections are going to describe the structure of each of these triggering levels, while focusing on examples relevant to the main study presented in this thesis.

4.2 Level-1 trigger

The starting point of this discussion is the, as the name suggest, first level of selection within the CMS experiment [77]. It stands at the front line of data acquisition and is tasked with making a fast (in less than 4 μs) but sensible decision on whether an event should be passed further down the chain. The core of the entire setup are Field Programmable Gate Array (FPGA) custom processing boards. Their ability to perform faster parallel processing, re-programmability and the subsequent good price to performance/durability ratio, have lead to them being chosen instead of a CPU architecture. The many advantages that come with using the FPGAs in highly specialized tasks have lead to their wide spread usage, reaching even the video game industry[1]. The following pages focus on the structure of the Level-1 trigger, relying on a set of examples to explain its operation.

4.2.1 Overview

The L1T forms a decision relying on the information coming form two centres: Calorimeters and Muon chambers. This can be seen in Figure 4.1, which shows the structure of the L1T system during the Run 2 phase of operation. Named after the corresponding information pool, the substructure of the L1T can be split into the Calorimeter and Muon trigger. The Calorimeter (Calo) trigger is comprised of two layers and is responsible for the reconstruction, calibration and baseline identification of particles with ECAL/HCAL deposits: electrons/photons (e/γ), hadronically decaying taus (τ) and hadronic jets. The aforementioned jets represent a residuum of hadronic interaction located within a narrow geometrical cone. Due to the lack of tracker information at this stage, it is not possible to distinguish electrons from photons, which is why they are referred to as e/γ objects. The Global Muon Trigger (GMT) is used to summarise the information gained from the three muon subsystems previously defined in Section 3.2.6. The starting point for both of these

---

[1] FPGAs are being used in modern recreations of 16-bit custom chips making the core of the fourth video game console generation such as the ones found in the Super Nintendo Entertainment System and Sega Mega Drive systems.

triggers comes from the information provided by the trigger primitives (TPs) which are formed from basic detector units.

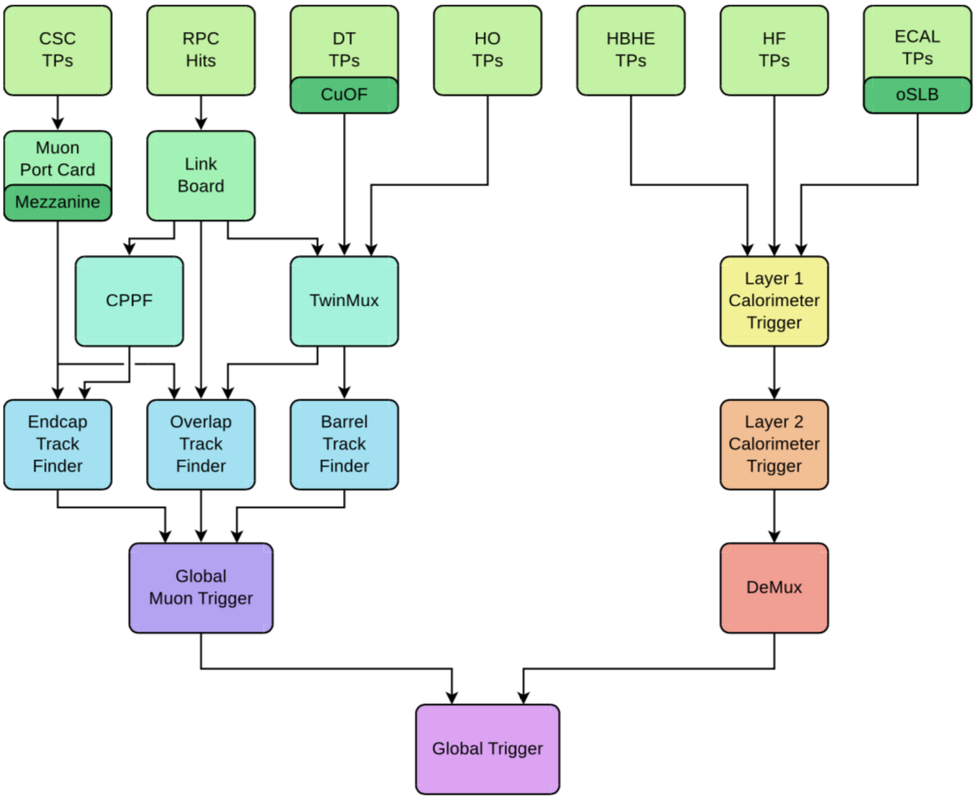


Figure 4.1: Structure of the CMS Level-1 system during its Run 2 operational phase. [77]

Figure 4.2 shows the structure of the Calo trigger. Detector inputs for this system are made out of Calo Trigger Towers (TTs). Each of these TT represents a 5x5 ECAL crystal structure combined with the HCAL tower positioned behind it [77]. The information (energy deposits) coming from TTs is passed to the Calo Layer-1. These deposits are then calibrated and corrected for additional detector effects (not accounted by the calorimeter electronics), preparing the input for the next step.

The Calo Layer-2 is comprised out of a set of 9 FPGA cards (MP7), connected together in a round robin scheduling sequence. This Time-Multiplexed Trigger (TMT) design helps to extend the processing time given to each of the Calo Layer-2 cards. As an example, if each of the boards received information regarding one event per bunch crossing using this TMT design, this would increase the available processing time to be nine times larger than the standard bunch crossing time without pipelining. At this stage, Calo objects are being reconstructed from the available information using dedicated algorithms. The resulting data stream is being passed onto the De-multiplexing node (Demux), another MP7 board, with the task of retrieving the original event ordering before sending this information to the Global Trigger (μGT).

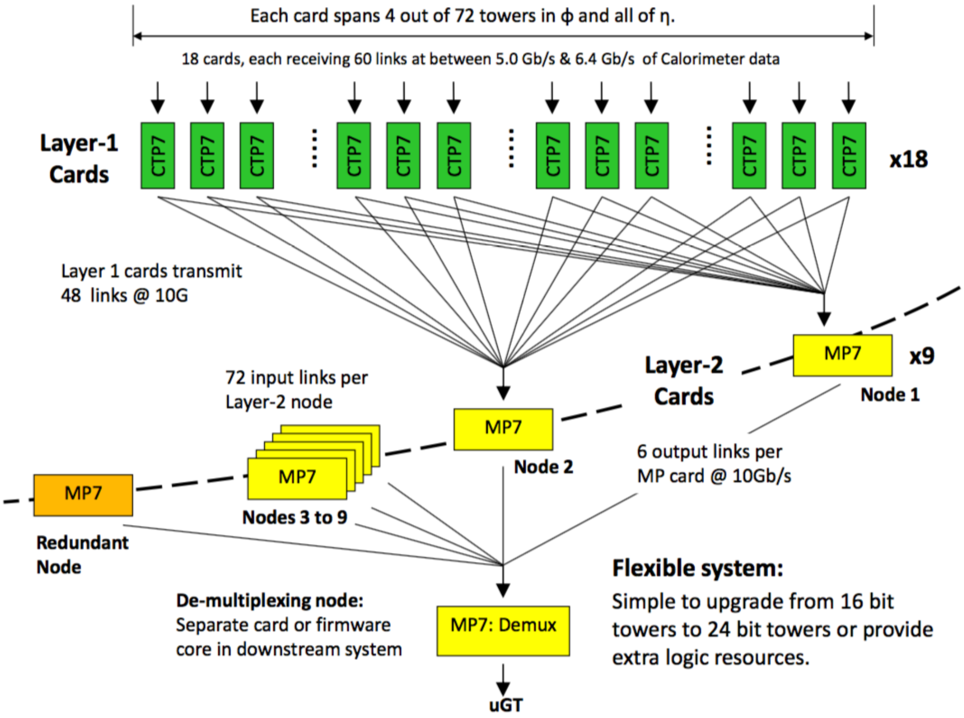


Figure 4.2: Structure of the Calorimeter Trigger and its Time-Multiplexed Trigger architecture. [77]

Due to their connection with the VBF production mode and the subsequent invisible final state of the Higgs boson, jets and energy sums will be used to illustrate L1T reconstruction algorithms [77, 78]. Reconstruction of jets at L1T is performed through the use of a dedicated algorithm. It focuses on a 9×9 TT range surrounding the TT with the maximal deposit within the area, which satisfies the $E_T > 4$ GeV requirement. This TT range is chosen to mimic the maximum geometrical (η, ϕ) area used with offline jet reconstruction, which revolves around the ΔR $= 0.4$ cone (the anti-k_T approach [79]). A simple comparison of energies with the TTs surrounding the central one are performed and the central TT is selected as a jet ”seed” candidate if no other TT in the 9x9 area has energy larger/larger or equal than it. This additional equality comparison was added in order to remove the possibility of the effect when TTs with same energies mutually veto each other. The chosen convention is shown in Figure 4.3. The total energy associated with the jet is then computed by summing the contribution from all TTs in the area, while the jet position is fixed to the central seed TT.

Subtraction of the pile-up[2] contribution is done by using a set of four 3x9 TT structures, positioned next to the edges of the original 9x9 structure. This combined set of TTs bears a similarity with a "chunky doughnut", which is the initial idea behind the name of this jet pile-up removal algorithm. In order to avoid the scenario where a contribution coming from another jet structure would be removed, pile-up E_T is computed using the sum of the three regions with the lowest energy which is then subtracted from the jet candidate energy.

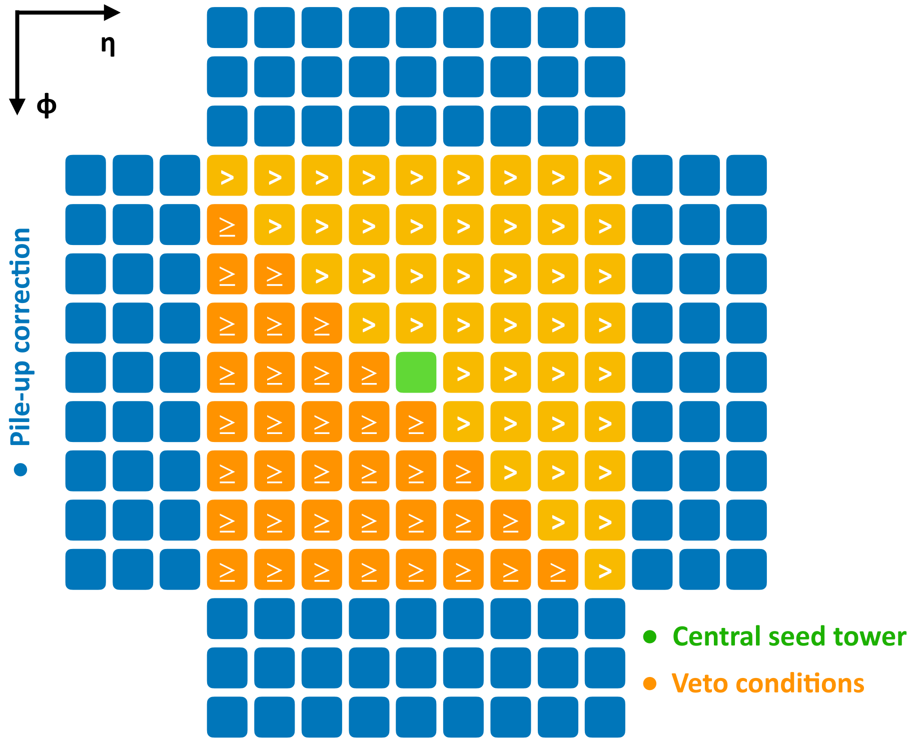


Figure 4.3: Graphical representation of 9x9 Calo TT formation used for L1T jet reconstruction, accompanied by 3x9 structures used for the pileup removal.

Using the energies of the TTs it is possible to reconstruct the $\vec{p}_{T,miss}$ vector following the formula:

$$\vec{p}_{T,miss} = -\sum_{i=1}^{n_\eta}\sum_{j=1}^{n_\phi}(E_T^i cos(\phi_j)\vec{e}_x + E_T^i sin(\phi_j)\vec{e}_y) \tag{4.1}$$

where (i, j) denote the sums over the (η, ϕ) coordinates of TTs. Upon obtaining this property, the L1T $E_{T,miss}$ variable, extensively used when creating trigger paths listed in Table 7.3 can be constructed as $E_{T,miss} = |\vec{p}_{T,miss}|$. Additionally, if the summation, when forming the previously defined property, goes over jet objects instead of all TTs, a new variable $\vec{H}_{T,miss}$ is created.

These physics objects reconstructed by the Calo Layer-2 are passed on to the

---

[2] Additional interactions whose resulting signatures are overlapping with the interaction of interest. It is coming from the fact that the LHC collides bunches of protons instead of a single pair.

final step, the μGT, where they are combined with the information coming from the Muon trigger. At this stage, a decision is made whether the event is going to be passed on to the HLT or discarded. In order to be able to make a decision, L1T uses a set of algorithms (seeds) which impose selection requirements on available physics objects. If the event passes all the requested qualities/thresholds and if the event is in accordance with the trigger rules, a Level-1 Accept (L1A) decision is made and the event is passed on to the next step.

4.2.2 Level-1 trigger menu

A set of aforementioned decision making algorithms, combined together through a logical ”or”, creates a structure called the L1T menu. The total rate of events selected by the menu is expected to be no larger than 100 kHz. In order to measure the collective rate given by the menu, an unbiased dataset (ZeroBias) is needed. The name ZeroBias is reflective of the fact that this data was collected using a selection that imposes only a single requirement: the event needs to contain a collision.

In order to control the total rate of events at the L1T, each seed within the menu has a ”prescale” option inserted in the workflow. It provides the ability to limit the number of events passing the algorithm by scaling it down using a given integer value. This results in a scenario where, for a prescale value of 10, every tenth event which satisfies conditions imposed by the seed, is allowed to advance to the next stage.

The implementation process for a new seed follows a standardised flow. Whether a simple or a complex algorithm, its efficiency needs to be measured using Monte Carlo (MC) simulation samples that have been produced with appropriate object reconstruction. The increase in rate of the trigger menu associated with the addition of a new seed is called its pure rate. It has to be low enough for the trigger to be added to the menu, without removing/prescaling other seeds. The majority of the trigger menu rate is dedicated to seeds that have wide-spread usage in the later stages. Good examples of this are single/double lepton and $E_{T,miss}$ triggers, whose logic is presented in Table 4.1.

For example, the lowest threshold, unprescaled single muon trigger is looking for a tight quality muon object that has $p_T > 22$ GeV. Muon objects that have been registered in at least three muon stations are marked down as tight quality objects. It is also possible to design triggers that handle complex object manipulation. One new seed taking this advantage is created for the purposes of selecting VBF events. More details about this seed are presented in Section 4.5.1. Finally, in order to safeguard the menu from any unforeseen rate problems, it is common practice to

Algorithm	Requirements
Single Muon	Tight Qualty & p_T >22 GeV
Double Muon	Tight Quality & p_T >8 GeV
Energy sums (E_T^{miss})	E_T >100 GeV

Table 4.1: A selected set of examples of most-used unprescaled trigger algorithms with their corresponding logics [77].

have a few backup options for each seed. They are implemented with the same logic as the original, but with slightly higher thresholds on selection requirements. They provide viable replacements in situations when the menu rate is too high and the primary seed needs to be prescaled until the issue is fixed.

4.2.3 Level-1 trigger pre-firing

Due to the radiation damage under which the ECAL is exposed during LHC operation, a loss in transparency in its crystals began to affect its performance during 2016. With time this effect gradually increased, with the high $|\eta|$ regions being the most affected. This introduced a timing offset in calibration of ECAL pulses, an effect that translated into resulting detector information which was turned into subsequent L1T Calo TTs. As a consequence of this, the Calo trigger information for the $|\eta|$ >2.5 could be wrongly assigned to the previous (BX = -1) bunch crossing instead of the correct one (BX = 0). This can lead to two major effects regarding the L1T decision.

Firstly, it can lead to the effect called the L1T pre-fire in which the system wrongly distributes the information that the BX = -1 event passes its logic to the HLT. As this effect is not present with the offline reconstruction, a simplified version of which the HLT uses when reconstructing objects, pre-firing will cause the total loss of an interesting event. This comes from the fact that BX = -1 event will most likely be rejected at the HLT level due to the low probability that both BX = 0 and -1 will contain interesting physics processes. On the other hand, accounting for the part of trigger rules which state that no more than one L1A decision can be made for three successive events (and no more than 2 L1As in a group of 25 successive events), the BX = 0 event in this case is lost.

Another scenario can happen due to a possibility of biased measurement of energy deposited in the Calo TTs. If the situation is reversed and the ”early” signal does not invoke a positive L1A decision, there is still a problem as the, potentially interesting, BX = -1 information will be wrongly lost causing a bias in the energy measurement for the BX = 0 event.

Unfortunately, a good example of a study affected by this issue is the VBF H→inv analysis, due to its dependence on forward jets[3]. The effect of this has been studied at the analysis level for the data collected in 2016 [39] and the results state that a correction ranging from 1-20 % needed to be applied (depending on the fit variable range), which resulted in a total loss of ∼17 % of the expected sensitivity towards the Br(H→inv). More details about the corrections of this issue for the data collection era of 2017 are given in Chapter 6.

4.3 Online data quality monitoring

In order to have constant surveillance on the operation of the CMS experiment, three daily shift crews are placed at the control cavern of the detector (positioned at the LHC Point 5). The constant monitoring of L1T menu rates is needed to quickly account for any unexpected increases, by prescaling problematic seeds. On the other hand, in order for the shifter to be able to monitor the performance of the L1T, a set of informational plots and alarms is displayed as a part of the data quality monitoring (DQM) package [78]. Taking into account that the knowledge possessed by the shifter is usually not at the level of a L1T expert, a separation of plots is being made into two groups: shifter and expert, where the latter contains much more detail (i.e. the full Calo granularity information) needed for debugging.

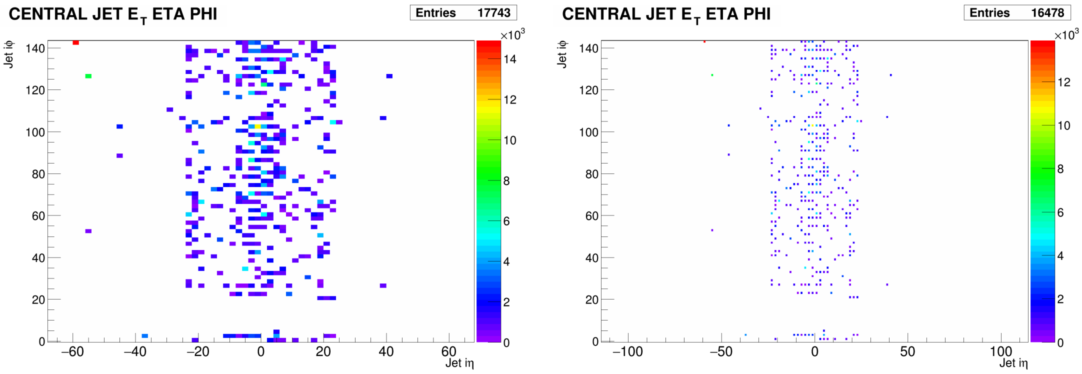
Taking the Calo trigger as the example, in order to have a separate control system, emulator software was written in C++ and optimised for a CPU architecture with the same logic as the one used on the trigger boards (both the MP and Demux nodes). This emulator was made a part of a standardised CMS Software Framework (CMSSW), which helps with the general ease of use and software compatibility. In order to monitor the performance of the Calo trigger during run periods an online DQM package was deployed containing several main checks needed for validation.

There are several types of plots needed to efficiently control the operation of the Calo Layer-2 during the data taking periods. One of the first checks would be to plot the distributions of main properties of physics objects in order to quickly spot any anomalies in the reconstruction. The "data to emulator" comparison plots sorted either by the overall agreement or disagreement between the firmware and the emulator represent a systematic way of ensuring early problem detection. A

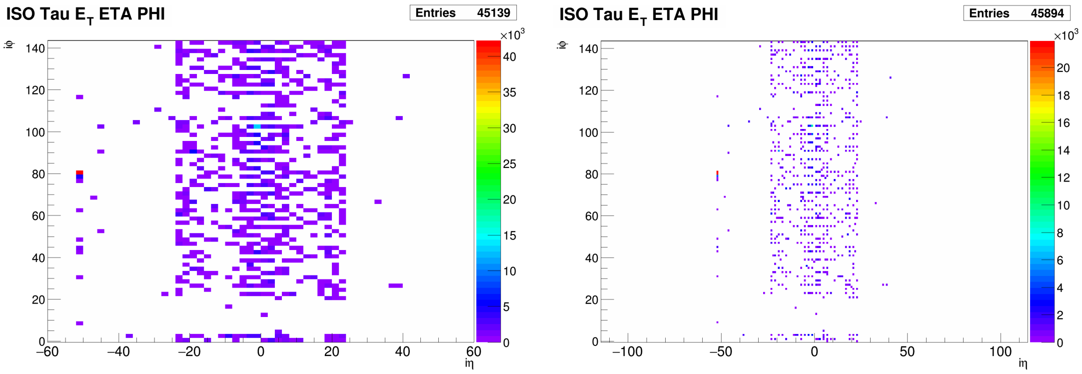
---

[3] Jets within the forward region ($|\eta_j| > 2.4$)

selected set of example plots summarising this setup are given in Figures 4.4 and 4.5 (with additional examples being given in Figure A.1).



(a) Central jet $E_T(\eta,\phi)$ - shifter (b) Central jet $E_T(\eta,\phi)$ - expert



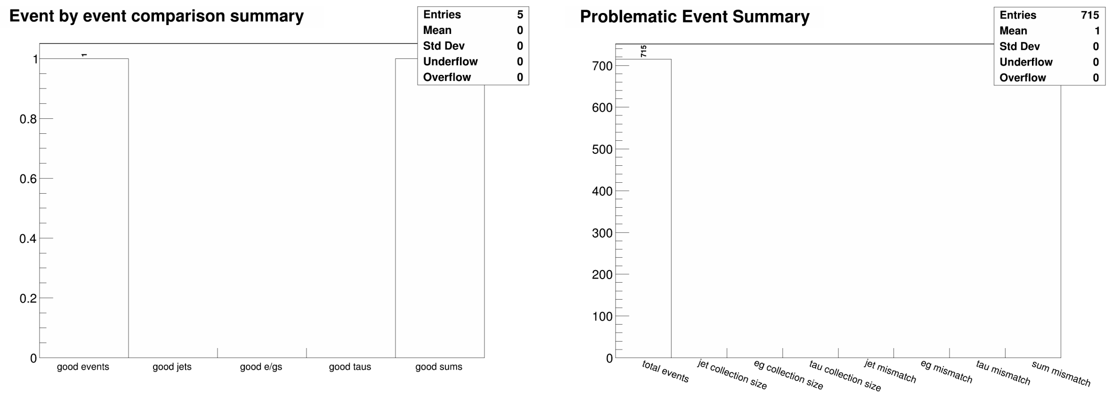
(c) Isolated tau $E_T(\eta,\phi)$ - shifter (d) Isolated tau $E_T(\eta,\phi)$ - expert

Figure 4.4: DQM example distributions. Comparison of information displayed to the shifter and expert is shown for jets with (a) and (b), and isolated taus with (c) and (d) respectively.

4.4 High Level Trigger

The second step of trigger selection operates under more manageable input conditions, allowing for a few hundred millisecond decision making time window. This has enabled the more traditional, high-level software, design of the High Level Trigger [80]. It is based on a set of algorithms (paths) written in C++, that are deployed on a server farm comprised of Intel Xeon CPU based machines.

This additional time allows for more complex object reconstruction, which is why the HLT uses a simplified version of the Particle Flow (PF) algorithm [81] in order to take the advantage of the full detector information. More details about the PF algorithm itself will be given in Chapter 6. The main feature of the HLT is its modular design. A collection of C++ modules allows experts to build more detailed



(a) High level comparison summary (b) Summary of problematic events

Figure 4.5: Two DQM summary plot examples focusing on the high level comparison (a) and problematic events (b).

selection requirements. A module can take the role of a producer of $E_{T,miss}$ variable from given PF candidates, it can impose a selection requirement on a given variable and many more. Each module is required to be made as a template structure and included in the CMSSW framework. A HLT path is created through the combination of different modules through a chain of intermediate decisions. This leads to another reduction in rate, resulting in the output of ∼1 kHz that gets passed on to the final step, the full reconstruction and storage of events.

A more user friendly approach to managing available HLT modules is enabled through the usage of the ”confDB” software [82]. Written in the Java language, its main purpose is to help with the creation of new HLT paths. It provides a GUI environment which can be used to easily import modules into newly created structures as well to set their default properties. As briefly mentioned before, an HLT path represents a flow of decision making modules starting from the information gained from desired input L1 seed. The process of making and deploying a HLT path will be the focus of the next section, using the implementation of VBF production mode triggers as an example.

In order to deploy a HLT path, an approval procedure, similar to the one used with L1 seeds, needs to be followed. The first item on the checklist is to prove the physics motivation for the trigger. This is usually done through a study of an efficiency gain compared to previous algorithms. Following that, a more technical part of the process needs to be done. Measurements of pure rate and timing need to be performed and checked before adding the path to the official menu. The addition of timing is important due the fact that this is a fully software trigger using a simplified version of the PF algorithm, which is the biggest offender.

4.5 Triggering of VBF events

Historically, the search for the invisible decays of the Higgs boson has been built around $E_{T,miss}$ and $H_{T,miss}$ based triggers. The similar statement holds true for many of searches for beyond the SM physics at the CMS experiment, which is why these triggers gained the status of general purpose algorithms. This was enabled by checking two main goals when it comes to defining a multi purpose trigger. Firstly, they were interesting to the analysers as they only imposed selection requirements on a very few variables during the triggering process. This, in return, allowed for any additional object manipulation to be done in the offline analysis[4], without having a constraint coming from events not being stored. The second point was extremely important when looked from the organisational perspective. As these triggers were used by more analysis groups studying different signals containing particles invisible to the detector, it became easier to justify the amount of rate allocated to these paths, which resulted in a set of generously tailored selection requirements.

The previous statement in no way sounds the end of the story, as even with a more relaxed set of requirements, the very idea of the search is dependent on having the lowest possible threshold on the ”invisible” contribution. There were several attempts at creating triggers based around a set of crude HLT level decisions following the VBF topology. The main idea behind these triggers was that, by adding kinematic requirements on the jets, the rate would be reduced enough to allow for an even lower selection threshold on $E_{T,miss}$. This provided mixed results and a more viable solution was continued to be sought after.

4.5.1 Overview of the VBF Level-1 trigger

The situation changed for the better, when it comes to analysis specific HLT algorithms, in recent years. Following the recent upgrade of the L1T [77], analysts were given the opportunity to impose a set of requirements including complex object manipulation even at this first stage of decision making. This new feature has led to the creation of dedicated L1T seeds specifically targeting the VBF production mode of the Higgs boson (in further text referred to as VBF L1 seed). This has opened a door for a new set of HLT paths, which will tailor the selection towards the invisible final state. Serving as an example of the HLT trigger implementation procedure, the following section is going to describe the idea behind a set of HLT paths deployed during the mid stage of the Run 2 phase.

---

[4] After the full object (offline) reconstruction and storing of events.

4.5.2 Implementation of VBF High Level Triggers

The base logic of the aforementioned L1 VBF seed, was created by transforming the physical properties of the VBF production mode into a set of decisions that can be interpreted at the first stage of the CMS triggering system. This mode targeting algorithm was built from the benefits given by the recent upgrade of the L1T, allowing for complex object manipulation. One of these variables interesting for the VBF production mode was the invariant mass of a dijet pair (dijet mass). It is defined as:

$$m_{j_1,j_2} = \sqrt{2 \cdot E_T^{j_1} \cdot E_T^{j_2} \cdot [cosh(\Delta\eta_{j_1,j_2}) - cos(\Delta\phi_{j_1,j_2})]}, \tag{4.2}$$

where $E_T^{j_{1/2}}$ denote transverse energies of L1 jets and $\Delta\eta_{j_1,j_2}/\Delta\phi_{j_1,j_2}$ measure their geometrical separation in the (η/ϕ) plane. It provided a valuable tool in reducing the rate of the trigger, while ensuring a minimal loss of desired signal sensitivity. Translated into the actual algorithm logic, this approach states that the event passes the L1T requirements if it contains both of these two scenarios [77]:

- The event contains at least two L1 jets that pass: $E_T > 110,\ 35\ (115,\ 40)$ GeV
- There exists a dijet pair, whose invariant mass $m_{jj} > 620$ GeV, where the jets forming it also pass $p_T > 35$ (40) GeV requirement

The first era of using this seed (2017) brought its fair share of problems (as previously discussed in Section 4.2.3) which increased the rate of L1 seeds targeting forward jets. This manifested itself in the fact that the L1 jet thresholds had to be much higher than expected in order to contain the rate within reasonable boundaries. This forced the experiment into activating additional backup seeds, listed in Table A.2 during this period. As the time passed, CMS trigger experts managed to implement a correction for this effect while also relaxing a part of its bandwidth, which allowed for the lower threshold seeds to be reinstated into the triggering menu during 2018. More details about the process of implementation and the performance of this seed can be found in [77, 83].

As a means to actually explore this newly available phase space at the analysis level, a set of second level decisions were combined to create new HLT paths. The main idea behind them was to take the events passing the VBF L1 seed and imposing a groups of kinematic requirements on PF reconstructed jets (mimicking what was done at the L1T stage), while matching the PF to the L1 jets. Second step was to further tailor this selection to the "invisible" final state by adding $E_{T,miss}$ requirements. The final goal was to have a reduction in rate coming from jet conditions being large enough to be able to justify a more relaxed set of $E_{T,miss}$ thresholds. The following points summarise the HLT requirements:

- $E_{T,miss}^{Calo,NC} > 66$ GeV requirement on $E_{T,miss}$ reconstructed from the Calorimeter information only, which has been cleaned from the HCAL noise;
- $E_{T,miss}^{PF} > 110$ GeV requirement on $E_{T,miss}$ reconstructed from all PF object candidates;
- Usage of a custom made module for the purpose of matching the L1 jet collection with the PF jet collection as well as for the application of additional kinematic requirements on the jets;
- Further separation into two and three jet categories based on the output of the previous step.

Starting with the jet part of the algorithm. First, events are pre-filtered based on a flag specifying whether the event passes a logical "or" of all L1 VBF seed variants. For the passing events, the L1 jet collection is matched to the PF jets that have $p_T > 35$ GeV, using the geometrical cone of $\Delta R < 0.5$ as its matching criteria. Upon obtaining this new collection of PF jets, they are required to fall under one of two options (otherwise the event fails):

- Option A: All possible dijet pair variations are created in order to find the combination which yields the largest dijet mass, with that value having to be greater than 650 GeV. After selecting the pair, it is required that the higher p_T jet passes the $p_{T,j} > 110$ GeV threshold.
- Option B: Jets forming the largest dijet mass in the event must pass the $p_{T,j} > 35$ GeV requirement, while the dijet mass must be larger than 650 GeV. Neither jet from the pair is allowed to have $p_T > 110$ GeV. Upon obtaining the dijet pair, the leading jet in the collection is looked at in terms of p_T. If it passes $p_T > 110$ GeV requirement, all three jets are selected and the event advances to the next stage.

After collecting the two possible scenarios, the final module of the path selects either the two jet (option A) or the three jet (option B) case, thus forming the two paths shown in Table 4.2. The structure and decision flow of these paths containing all of the modules is displayed in Figure 4.6.

Moving on to the $E_{T,miss}$ requirements, as mentioned above, certain requirements have been introduced in order to control issues of rate and timing. A certain threshold on the $E_{T,miss}^{PF}$ had to be introduced in order keep the rate within a reasonable

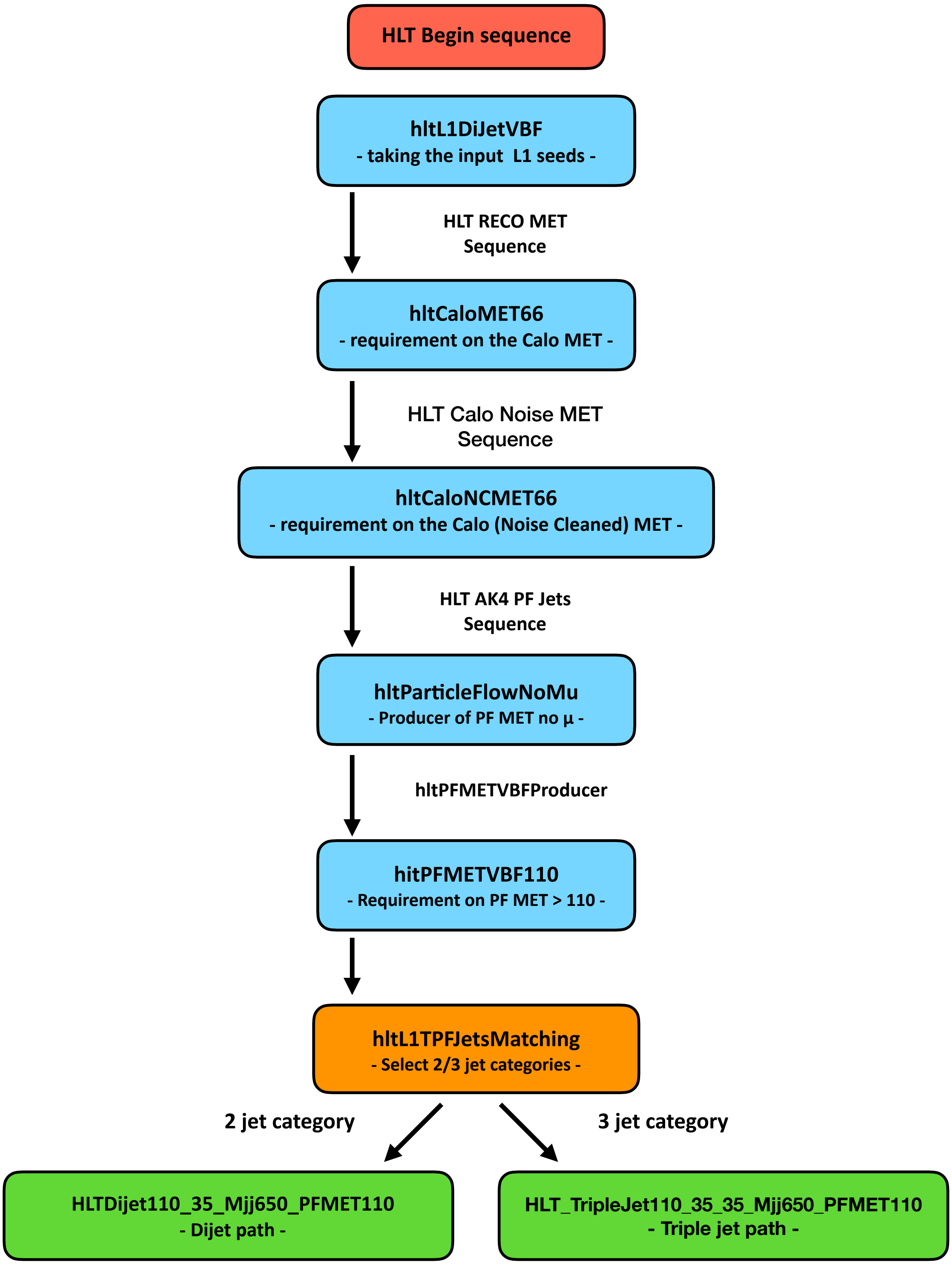


Figure 4.6: Modular structure of the VBF HLT paths. The common flow starting with the input decisions inherited form the VBF L1T seeds is shown for both the dijet and triple jet HLT paths.

range. Figure 4.7 shows the estimate of the total rate of the dijet path and highlights the final choice of $E_{T,miss}^{PF} > 110$ GeV as the minimal sustainable value. Potential issues that come with timing are directly related to the usage of the PF algorithm. As it requires the largest amount of time, a requirement was implemented in order to reduce the amount of events that would initialize it and then proceed to be rejected due to the $E_{T,miss}^{PF} > 110$ GeV requirement. This led to the inclusion of the $E_{T,miss}^{Calo,NC}$ selection listed above. A correlation study between these two $E_{T,miss}$ variables was performed (as shown in Figure 4.8). It can be seen that no significant signal loss is expected if a requirement on the $E_{T,miss}^{Calo,NC}$ is imposed to be ∼60 % of the one on the $E_{T,miss}^{PF}$. This, as a result, will stop a large amount of events from reaching modules which call the PF algorithm, instead stopping right after the $E_{T,miss}^{Calo,NC}$ producers, which take a lot less time.

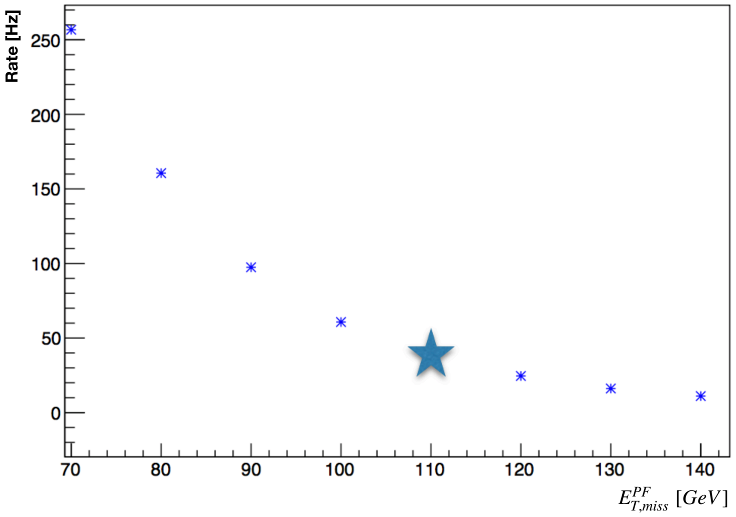


Figure 4.7: Estimate of the total rate of the dijet VBF HLT path versus the requirement on the $E_{T,miss}$ (for the instantaneous luminosity of $1.0 \cdot 10^{34}$ cm^{2}s^{-1}).

The triple jet path inherits most of the same workflow as its dijet counterpart, differing only in the final choice of jets. This triple jet category is more oriented towards analyses such as the VBF H→bb$\tau\tau$, where it might bring additional sensitivity. From the perspective of the H→ inv analysis, this triple jet scenario is valuable as a safeguard option for the subtle differences between the offline and HLT PF jets (which can also be seen by its significantly smaller rate).

These HLT paths were introduced during the 2017 data collection and continued to be used in their original state until the end of the Run 2 phase. The pure and total rate of these paths for the 2018 data taking period are given in Table 4.2. Their

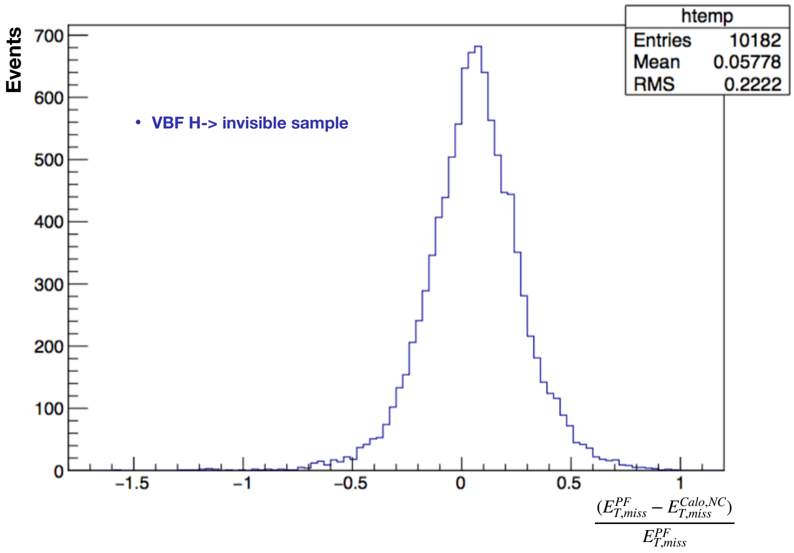


Figure 4.8: Correlation study between the $E^{PF}_{T,miss}$ and $E^{Calo,NC}_{T,miss}$ done using the VBF H→inv simulation sample.

performance is described in Section 7.2.2, where it is shown how they influenced the creation of a new analysis subcategory.

HLT path summary	Status	Total (pure) Rate [Hz]
Dijet		
$p_{T,j}$ > 110/35 GeV, M_{jj} >650 GeV, $E^{PF}_{T,miss}$ > 110 GeV	signal	36.88 (7.84)
$p_{T,j}$ > 110/35 GeV, M_{jj} >650 GeV, $E^{PF}_{T,miss}$ > 120 GeV	backup	25.81 (0)
$p_{T,j}$ > 110/35 GeV, M_{jj} >650 GeV, $E^{PF}_{T,miss}$ > 130 GeV	backup	18.99 (0)
Triple jet		
$p_{T,j}$ > 110/35 GeV, M_{jj} >650 GeV, $E^{PF}_{T,miss}$ > 110 GeV	signal	0.91 (0.44)
$p_{T,j}$ > 110/35 GeV, M_{jj} >650 GeV, $E^{PF}_{T,miss}$ > 120 GeV	backup	0.44 (0)
$p_{T,j}$ > 110/35 GeV, M_{jj} >650 GeV, $E^{PF}_{T,miss}$ > 130 GeV	backup	0.25 (0)

Table 4.2: Measurement of rates for main VBF paths and their backups during 2018 era of data taking.

As a final remark, it is important to note that even though the matching module was created for the purposes of the VBF H→inv analysis, it has been used in other studies of the Higgs boson within the CMS experiment, such as the aforementioned $H \to \tau\tau$/ $HH \to bb\tau\tau$ analyses, helping them to take the advantage of the L1 VBF seed by building their VBF HLT paths.

* * *

Part III: The study

“Triput meri, jednom seci.”

– Serbian proverb –

Chapter 5

Overview

> *"However beautiful the strategy, you should occasionally look at the results."*
>
> — Sir Winston Churchill

This chapter provides an overview of Part III of this thesis by summarising the core ideas behind the search for the invisible decays of the Higgs boson produced in a VBF event. As introduced in Chapter 2, this study is motivated as the invisible final state represents a highly suppressed scenario from the perspective of the SM, with Br($H \rightarrow 4\nu$) ~ 0.1 %, yielding a conclusion that any deviation from it would be a clear indication of physics beyond the SM.

The main region of interest or the signal region is defined following topological properties of the VBF Higgs boson production, namely its two jet topology and respective characteristics of said jets. One additional factor used to quantify the invisible contribution is the, previously defined, $E_{T,miss}$ variable. The signal region is formed from two analysis categories each built around a set of trigger algorithms. The low $E_{T,miss}$ category will be represented with the VBF triggers introduced in Chapter 4, while the high $E_{T,miss}$ category is connected to the, more generic, $E_{T,miss}$ based triggers. More details about the categorisation are given in Chapter 7, where a detailed discussion of selection requirements for each analysis category is added alongside the performance of the relevant trigger algorithms (whose performance also influences the selection). The structure of the analysis follows a standardised "blinded analysis" approach, where the data events falling under this region are being omitted from the study until the analysis strategy is finalised.

Contributions of main sources of backgrounds, in this case being the V+jets processes, are constrained through the introduction of four background control regions mimicking the dijet topology of the signal region, but being in a background dominated region with no signal contribution. An example of this would be regions which contain leptons whose invariant mass is found in a narrow, Z boson mass, resonance range (used to constrain Z+jets backgrounds). These regions require a slight redefinition of the $E_{T,miss}$ variable where the leptons forming the region are to be removed from the calculation of the $\vec{p}_{T,miss}$. This is done in order to have the equivalent selection requirement be as close to the signal region as possible. For these, well known lepton-enriched regions, a common conclusion is that the next-to-leading order computations of respective cross sections of SM processes are enough to describe the data. On the other hand, due to the amount of computational power (and time) needed to produce general purpose samples of that precision being extremely large, V+jet backgrounds are simulated using their leading order calculations. In order to account for the difference, a re-weighting procedure alongside with its associated uncertainties is introduced. These higher order corrections are the main focus of Section 6.3.6. Lastly, a special attention needs to be given to the estimation of the contribution of QCD multijet processes in the signal region. This procedure involves creation of a dedicated control region largely populated with multijet events by inverting a single requirement from the signal region definition. This approach is needed due to a lack of statistical precision in QCD multijet simulation samples. More details about these dedicated control regions and respective studies are given in Chapter 8.

This study focuses on data collected by the CMS experiment during 2017 and 2018 eras of data taking, with the final combination being performed with the study focusing on the 2016 era without re-analysing the data. These eras brought their share of detector problems which were affecting the quality of the collected data, unfortunately, in both years. A common problem is found in the appearance of a data excess in leading/subleading jet η distributions in the signal region, which is not well modeled in the simulation. Other important problem happened during the 2018 data taking period, when a part of the HCAL suddenly stopped working leaving a large portion of the data without any HCAL information in the affected region. Both of these problems are addressed in detail and accompanied with respective mitigation approaches in Section 7.4. The signal extraction strategy and the approach used to set the 95 % CL upper limit on the Br(H→inv), detailing the connection between the dedicated control regions and the signal region, is the main focus of Chapter 9. A discussion of major sources of uncertainties (both theoretical and experimental) is given, alongside the final results inclusive of the combination with the 2016 study. The conclusion, presented in Chapter 10, introduces an approach to combining the results from searches for the invisible state focusing on other production modes through the use of a novel software framework. Lastly, a discussion of the future stages of the LHC from the perspective of this analysis is presented, putting a fitting conclusion to the entire H→ inv story presented in this thesis.

Chapter 6

Object definitions

> *"All compromise is based on give and take, but there can be no give and take on fundamentals. Any compromise on mere fundamentals is a surrender. For it is all give and no take.*
>
> — Mahatma Gandhi

6.1 Introduction

THE FOUNDATION OF MOST EXPERIMENTAL STUDIES is based around the connection between the original idea and the actual reality presented in the form of technical possibilities, in practice more likely limitations, given by the apparatus at hand. The same statement is applicable for the main interest of this thesis. The idea is very appealing on paper (as previously discussed in Chapter 2), there is a production mode with a strong signature and a possible decay that is highly suppressed when looked at from the SM point of view.

From the perspective of the CMS experiment, each analysis needs to be built from the ground up using the same basis - reconstructed physics objects. Following the conclusion that a good way to describe the "invisible" part of an event is through the usage of $E_{T,miss}$ and through its definition in Equation 3.3, it can be seen that it takes the collective information from all parts of the detector to quantify the possible invisible contribution. Speaking in technical terms, all available objects will have a role to play in this analysis. This chapter will serves as a summary of the processes and algorithms used in order to reconstruct and define the base objects used in physics analyses.

6.2 Particle Flow reconstruction

With the limitations imposed by a large stream of events being removed through the usage of a two level triggering system, more detailed reconstruction can be used for the full (offline) reconstruction of physics objects. The PF algorithm [81, 84, 85] provides a valuable tool that connects information originating from all detector subsystems in order to provide the most detailed overview of the event possible. It relies on the features of the CMS detector to deliver exceptional tracking performance which is then combined with the information coming from calorimeters and muon detectors. The idea of combining calorimeter crystals into TTs replaced with the plan to identify, as precisely as possible, all stable particles originating from collision interactions, hence giving rise to PF candidates. The subsequent grouping of those PF candidates and performing identification techniques, as well as energy sum computations, provides analysers with a set of object collections on which to build their analysis on.

In order to efficiently present object collections vital to the main study of this thesis, each of the following sections will provide a brief overview of the reconstruction techniques used to define a particle collection. This description will be followed by a set of recommendations given by the corresponding CMS Physics Object Group (POG), which are then used to create separate collections for a particle type used further down the analysis chain (i.e. formation of the dedicated control regions etc.).

6.2.1 Tracks and primary vertex

The tracker information on charged particles is essential in their further identification and usage. The tracker also imposes itself as a better solution when measuring the momenta of charged hadrons than calorimeters (due to the energy loss in material before reaching them) with the added bonus of being able to pinpoint the original directions of particles before being affected by the magnetic field. This all indicates that the preferred course of action is to have the tracking efficiency being as high as possible [86]. In order to achieve that, an iterative approach is deployed.

A set of tightly tailored requirements is imposed in order to select a first set of tracks. This ensures the purity of track through the removal of fake contributions. The downside of this choice, the lack of high efficiency in track reconstruction, can be eliminated with next steps. The preparation for the second iteration sees the removal of high purity tracks selected in the first step while partially loosening the tight restrictions imposed on track candidates. This approach yields an increase in efficiency while keeping the high purity of selected tracks. A slight change in the approach is introduced after the third iteration. The last two iteration steps are responsible for covering particles originating from secondary vertices, which is being enabled through a modification of requirements regarding the track's origin.

Finally, in order to conclude this discussion, a choice of a primary vertex needs to be made. This is done by looking at the sum of p_T^2, where the sum goes over all reconstructed tracks. The vertex with the largest value is chosen to be the primary vertex for the event.

6.2.2 Muons

The easiest way to start discussing the definition of muons is to take a look at the criteria helping with the definition of loose muon objects (in further text referred to as Loose Muon ID). This categorisation is important in order to increase the purity of muon objects through the removal of contributions originating from charged hadrons. When applied to a PF object, the Loose Muon ID requires fulfilment of a few quality conditions. First, the object being looked at needs to be reconstructed as a PF Muon, accompanied by a supplementary condition that this PF Muon candidate needs to be defined either as a Global or a Tracker Muon [87, 88].

Both of the aforementioned requirements look for the scenario where the information about the muon candidate's track (originating from the muon subsystems) has been matched to its tracker counterpart. For the Global Muon criteria, tracks originating from two centres of information are extrapolated onto the same plane, upon which respective companions are being selected (one from each set). From there, a global track is extracted via a combined, Kalman-filter approach [89], fit using the information coming from, previously paired, tracks. The Tracker Muon criteria takes a slightly different approach. It considers all particles which pass very loose conditions[1] and extrapolates their tracks to the muon detectors while taking into account detector effects. If a hit in muon detectors can be associated with one of these tracks, the PF candidate is considered to be a Tracker Muon.

Following the blueprint instructed by the Muon POG [90], the definition of a tight muon objects imposes a stricter requirement on candidates by requiring the object to be a Global Muon with additional quality requirements (in further text referred to as Tight Muon ID). The first pair of quality requirements asks for a goodness of fit for the global muon track be expressed through $\chi^2/N_{DoF} < 10$ and the inclusion of at least one muon chamber hit (with there being at least two) in the aforementioned fit in order to suppress the fake contribution. Further suppression of these effects is

---

[1] p_T >0.5 GeV and p >2.5 GeV

enabled through the usage of $d_{xy} < 2$ mm, $d_z < 5$ mm [2] and $N^{hits}_{pixel} > 0$. Finally, in order to achieve an accurate measurement of muon p_T, a $N^{hits}_{tracker} > 5$ requirement is imposed.

Speaking in terms of the analysis level objects, previously defined Loose and Tight Muon ID criteria are combined with additional kinematic ($p_T > 10/20$ GeV), geometric ($|\eta| < 2.4$) and isolation ($\mathrm{I}^{Rel.}_{\Delta R<0.4} < 0.25/0.15$[3]) requirements in order to form the Loose and Tight Muon collections (respectively). Figure 6.1 shows the performance of this approach through a comparison of muon properties between data and simulated events. The information is presented after applying the selection requirements used to define the single muon control region for the VBF H→inv analysis. These are defined in Chapter 8 and are used to define a VBF-like region dominated by W+jets SM processes. The overall data to simulation comparison for the chosen muon variables shows a generally good agreement. The small discrepancies seen in the high muon p_T and $|\eta|$ ranges are covered by the associated uncertainties not shown in the comparison (uncertainty on the simulation samples shows the statistical uncertainty only). More general detector performance studies for the Run 2 phase of data collection can be found in Refs. [91, 92] showing the performance of the muon reconstruction and identification/isolation, respectively.

6.2.3 Electrons

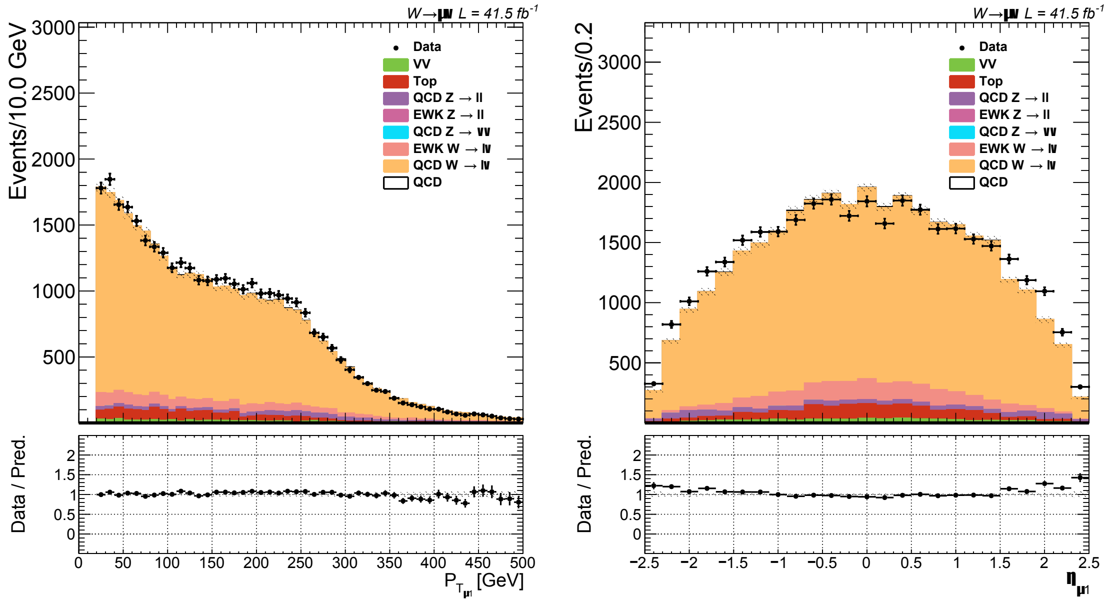
The interaction of electrons with the tracker material can lead to bremsstrahlung radiation manifesting itself in the form of emitted photons. Other detector effects such as the strong magnetic field cause electron energy deposits to be spread out in the ϕ range of ECAL [87, 93, 94].

Taking a look at the information given by the calorimeter, it can be seen that ~ 97 % of electron's energy (the same statement stands for photons) is deposited in a 5x5 ECAL crystal structure named the supercluster [95]. This allows for a matching procedure to be applied, pairing the supercluster to an electron track (obtained through a fit strategy which takes detector effects into account). As the transverse momenta goes down in value, it becomes more difficult to use the aforementioned approach due to the fact that the radii of the curvature of the particle's trajectory gets smaller. This introduces a problem for the supercluser formation as now the photon contribution (originating from bremhsstrahlung) can be much further in ϕ

---

[2] Representing the transverse impact parameter and the z-axis distance of the track when taking the primary vertex as the point of origin

[3] The relative isolation variable is defined as the ratio of the sum of E_T of photons and p_T of charged hadrons with respect to the muon candidate's p_T (where the sum covers particle candidates over the area of $\Delta R < 0.4$).



(a) $p_{T,\mu}$ (b) η_{μ}

Figure 6.1: Data to simulation comparison of leading tight muon p_T and η variables in a muon enriched (single muon) control region for 2017 data.

than before, asking for a more careful approach using a multivariate estimator in order to discover pure electron tracks (being relevant for values of $p_T < 10$ GeV).

Following recommendations given by the E/Gamma POG [94, 96], definitions of two main electron collections used in this analysis (Veto and Tight) are summarized in Tables 6.1 and 6.2 (in further text referred to as Cut Based ID). Similarly to the previous section, which dealt with the definition of muon collections, when approaching the definition of analysis level electron collections, the POG recommended Veto and Tight ID criteria are combined with kinematic ($p_T > 10/40$ GeV), geometrical ($|\eta| < 2.5$) and impact parameter requirements[4] in order to acquire the final Veto/Tight Electron collection (respectively). Figure 6.2 shows the performance of this approach through a comparison of electron properties between data and simulated events. Similarly to the muon discussion, these distributions display the data to simulation agreement for the purposes of another control region within the VBF H→inv analysis (also introduced in more detail in Chapter 8). This region focuses on the electron final state of the W boson decay and makes the single electron region. The data to simulation comparison shows a good agreement in general, with the discrepancies seen in $|\eta| > 1.5$ being covered by the associated uncertainties not shown in these plots (simulation samples are only presented with their statistical uncertainty). A more general summary of the performance of electron reconstruction

---

[4] For the barrel section a requirement of $|d_{xy}| < 0.05$ and $d_z < 0.1$ is imposed, while the endcap requirement asks for $|d_{xy}| < 0.1$ and $d_z < 0.2$

and identification is presented in Ref. [97].

Requirement	Veto	Tight
full 5x5 $\sigma_{i\eta i\eta}$	< 0.0126	< 0.0104
$\|\Delta\eta_{\mathrm{In,seed}}\|$	< 0.00463	< 0.00255
$\|\Delta\phi_{\mathrm{In,seed}}\|$	< 0.148	< 0.022
H/E	$< 0.05+1.16/E_{\mathrm{SC}}+0.0324\rho/E_{\mathrm{SC}}$	$< 0.026+1.15/E_{\mathrm{SC}}+0.0324\rho/E_{\mathrm{SC}}$
Rel. Isolation With EA	$< 0.198+0.506/p_T$	$< 0.0287+0.506/p_T$
$\|1/\mathrm{E} - 1/\mathrm{p}\|$	< 0.209	< 0.159
Exp. Missing Inner Hits	≤ 2	≤ 1
Pass conversion veto	yes	yes

Table 6.1: Summary of E/Gamma POG recommendations used to define Veto and Tight electrons in the barrel region ($|i\eta| \leq 1.479$) [94, 96, 98]. The conventional names $|\Delta\eta_{\mathrm{In,seed}}|$ and $|\Delta\phi_{\mathrm{In,seed}}|$ represent the geometrical distance between the extrapolated electron track and the selected supercluster. The $\sigma_{i\eta i\eta}$ variable is used to quantify the η dimension of the supercluser (weighted by its energy). Finally, the H/E variable controls the ratio of HCAL over ECAL contribution.

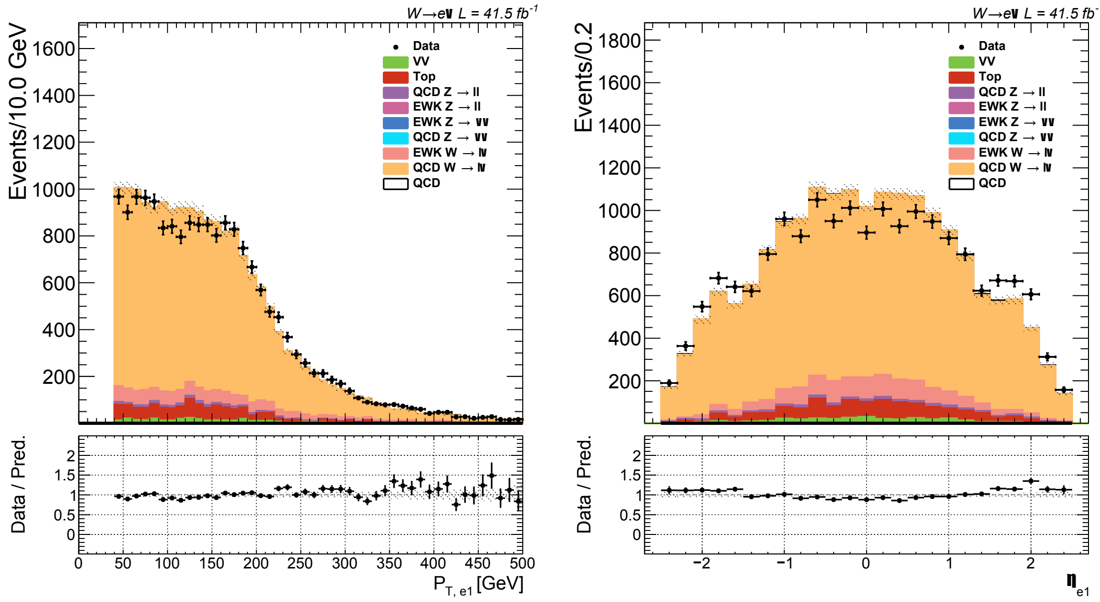
Requirement	Veto	Tight
full 5x5 $\sigma_{i\eta i\eta}$	< 0.0457	< 0.0353
$\|\Delta\eta_{seed}\|$	< 0.00814	< 0.00501
$\|\Delta\phi_{in}\|$	< 0.19	< 0.0236
H/E	$< 0.05+2.54/E_{SC}+0.183\rho/E_{SC}$	$< 0.0188+2.06/E_{SC}+0.183\rho/E_{SC}$
Rel. Isolation With EA	$< 0.203+0.963/p_T$	$< 0.0445+0.963/p_T$
|1/E-1/p|	< 0.132	< 0.0197
Exp. Missing Inner Hits	≤ 3	≤ 1
Pass conversion veto	yes	yes

Table 6.2: Summary of E/Gamma POG recommendations used to define Veto and Tight electrons in the endcap region ($|i\eta| > 1.479$) [94, 96, 98]. The naming convention used for control variables follows definitions introduced with Table 6.1.

6.2.4 Photons

Staying within the ECAL area of authority, the next item of discussion is the definition of the photon object collection. Upon completing definitions of collections revolving around charged particles and removing their contributions from the ECAL summary, the resulting clusters are used to form photon candidate objects. Further identification of candidates involves using algorithms which vary supercluster dimensions by using a set of predefined shapes associated with a photon deposit [98, 99] as well as relying on isolation variables. The definition of isolation requirements follows the idea that the scalar sum of transverse momenta of PF candidates (not being associated with the photon candidate's EM shower) is located around a certain geometrical distance from the tested object (in this case $\Delta R < 0.3$).

Being used for vetoing in the process of reducible background rejection, the definition of photons for this analysis involves using objects which pass the Loose photon



(a) $p_{T,e}$ (b) η_e

Figure 6.2: Data to simulation comparison of leading tight electron p_T and η variables in a electron enriched region for 2017 data.

criteria provided by the E/Gamma POG [100] (summarised in Table 6.3). Additional kinematic ($p_T > 15$ GeV) and geometric ($|\eta| < 2.5$) requirements are introduced alongside the Photon ID when defining the analysis level collection.

Variable	Requirement: Barrel (Endcap)
Full 5x5 $\sigma_{i\eta i\eta}$	< 0.0106 (< 0.0272)
H/E	< 0.04596 (< 0.0590)
charged hadron isolation	< 1.694 (< 2.089)
neutral hadron isolation	$< 24.032(19.722) + 0.01512(0.0117) \cdot p_T + 2.259(2.3) \times 10^{-5} \cdot {p_T}^2$
photon isolation	$< 2.876(4.162) + 0.004017(0.0037) \cdot p_T$
Conversion safe electron veto	Yes (Yes)

Table 6.3: Requirements used to define loose photon objects [98, 100].

6.2.5 Jets

Identification of jets is enabled through the use of the anti-k_T algorithm [79]. It produces PF jet candidates which are then used as the basis for creating analysis level jet collections. The aforementioned algorithm, relies on the following properties when defining a jet:

$$d_{i,B} = \frac{1}{p_{T_i}^2} \tag{6.1}$$

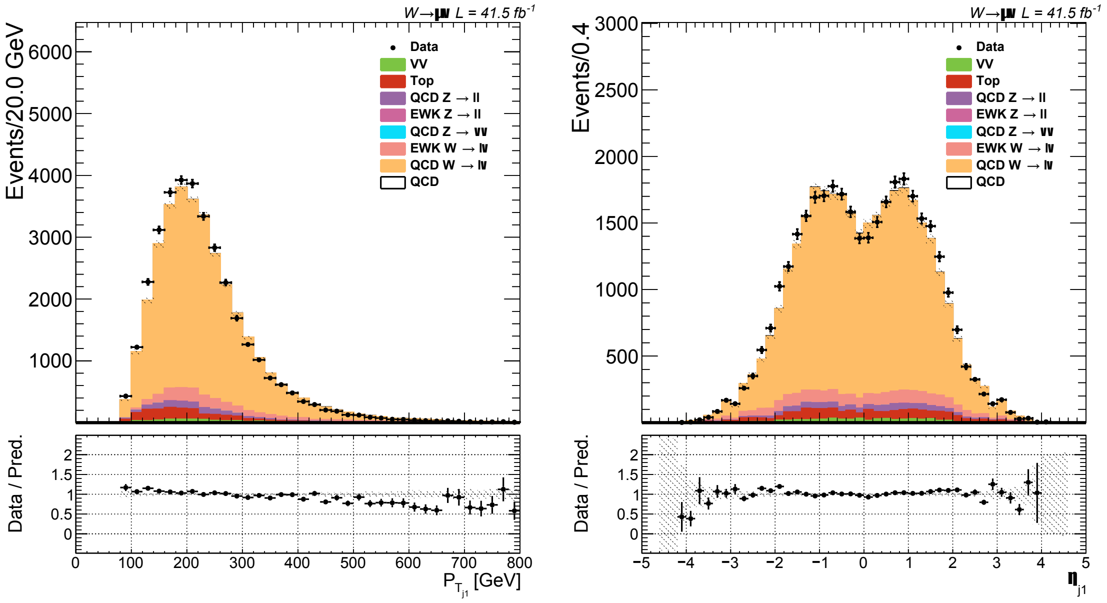
$$d_{i,j} = min\left(\frac{1}{p_{T_i}^2}, \frac{1}{p_{T_j}^2}\right)\frac{\Delta R_{i,j}^2}{R^2} \tag{6.2}$$

where $p_{T_{i/j}}$ are transverse momenta of particles i and j, $\Delta R_{i,j}$ is the geometrical distance between those particles. The R parameter (taking the value of 0.4 in this scenario) is used as a benchmark jet cone size (similar to the choice of TTs when defining Level-1 jets in Section 4.2.1) [98].

The original idea is, similarly to the Calo jet reconstruction in Chapter 4, to group softer particle candidates around the one which has the largest p_T within the area of preference (in this case R= 0.4). For a hard particle h, the algorithm computes both $d_{h,j}$ and $d_{h,B}$ for all soft particles j. The soft particle yielding a smallest $d_{h,j}$ is then merged with the hard particle to form a new particle candidate and the process is then re-started from the beginning. The iteration ends when the minimal $d_{h,j} = d_{h,B}$. This leads to the particle h, now a combination of the original hard and all soft particles chosen from previous iterations, being defined as a reconstructed jet. Upon removing the newly defined jet from the computation, the algorithms again resets and repeats the procedure until all particle candidates have been assigned to a jet [40, 79]. This process leads to a set of mostly conically shaped jet objects, with the edge case being represented with a scenario when there are two hard particle candidates within the 2R range. This leads to an overlap (and a slight change in shape) of the reconstructed jet cones created from those candidates.

Comparing the reconstructed jet p_T values between data and simulation leads to the conclusion that the resulting p_T value differs by $\sim 5 - 10$ % from the true momenta (where the comparison is inclusive of the full detector acceptance and p_T spectra) [98]. Jet objects are corrected for the contribution originating from pile-up through the introduction of an offset in their respective energies. These jet energy corrections are obtained from simulation [98, 101, 102].

Following recommendations given by the Jet/MET POG [103] a set of quality criteria (Jet ID) are added on top of PF jet collection in order to create analysis level objects [104]. These involve using a dedicated threshold on the fractions of neutral particles from ECAL and HCAL contributions as well as the muon fraction, number of constituents in a jet object, and the number of neutral particles. This study used the tight Jet ID working point, ensuring identification efficiency of $> 99/98$ % for 2017/2018 eras. Jet ID requirements are supplemented with a requirement of using a medium point of Jet Pile-up ID in order to reject pile-up contributions [105]. For the 2017 era of data taking, an additional veto requirement was added for jets within $p_T < 50$ GeV and $2.65 < |\eta| < 3.139$ range in order to suppress the contribution from jets originating from detector noise [98]. This final collection is cleaned from overlap with the lepton and photon collection using a $\Delta R <\ 0.4$ condition. Figure 6.3 shows the data to simulation agreement for properties relevant to the leading jet from the perspective of the single muon control region for the H→inv analysis. The general agreement seems to be good across all control regions, with the discrepancies seen in the high p_T region of the aforementioned distribution being covered by the associated uncertainty, not shown in this figure. The signal region has a set of jet quality issues mostly plaguing the high jet $|\eta|$ region. Details about these problems and the respective mitigation techniques are summarised in Chapter 7. More general detector performance studies focusing on the jet reconstruction can be found in Ref. [106].



(a) $p_{T,j}$ (b) η_j

Figure 6.3: Data to simulation comparison of leading jet p_T and η variables in a muon enriched region for 2017 data.

6.2.6 B jets

The definition of b jets[5] is important for the control of reducible SM background processes as these objects are used to veto events. This action is closely connected with the contributions originating from top quark processes [98]. The POG recommended quality criteria advises the usage of the DeepCSV (Combined Secondary Vertex) tagging algorithm [107] with a working point of 0.4941 and 0.4184 for 2017 and 2018 era respectively [108]. These numbers correspond to a medium working point of DeepCSV algorithm ensuring an 80 % efficiency of identifying a b jet. Additional kinematic ($p_T > 20$ GeV) and geometric ($|\eta| < 2.4$) requirements are applied when forming the analysis level object collection.

---

[5] The b jets or beauty quark jets represent, as the alternative name suggests, jets originating from b quarks.

6.2.7 Tau leptons

Similarly to the previous section, tau objects are important for vetoing events, thus reducing the contribution of V+jets SM backgrounds. A special algorithm is deployed in order to select the hadronically decaying taus[6]. The idea behind the algorithm is to check if the jet object is comprised from objects associated with a tau decay. The selected tau candidates are requested to be completely isolated from other objects (the comparison point for the isolation is $\Delta R < 0.5/0.3$ for 2017/2018 era). Finally, a set of kinematic ($p_T > 20$ GeV) and geometric ($|\eta| < 2.3$) requirements is imposed when creating the analysis level collection [109].

6.2.8 Missing transverse energy

Defined with Equation 3.3, the $E_{T,miss}$ variable provides an important view of the transverse contribution of particles invisible to the detector. From the reconstruction point of view, it is defined through the use of all PF particle candidates by taking a negative vector sum of their corresponding transverse momenta. Following additional corrections changing the jet p_T (previously mentioned in Section 6.2.5), a recalculation of the $\vec{p}_{T,miss}$ is performed in order to reflect this change:

$$\vec{p}_{T,miss}(\text{corrected}) = \vec{p}_{T,miss} - \sum_{\text{j}} (\vec{p}_{T,j}(\text{corrected}) - \vec{p}_{T,j}), \tag{6.3}$$

where the sum runs over all jet objects [98]. The $p_{T,j}$(corrected) tends to form a connection between the PF jet, with momenta $p_{T,j}$, and its real transverse momentum. This corrected transverse momentum is obtained through a set of successive operations, which combined can be illustrated as: $p_{T,j}(\text{corrected}) = \mathcal{C}_{all} \cdot p_{T,j}$ (as presented in more detail in Ref. [110]). The first correction forming the $\mathcal{C}_{all}$ is being applied to the PF jet p_T and performs offset corrections which include removal of pile up effects and electronic noise. The newly obtained, offset corrected, momentum is being put through another procedure designed to perform a simulation calibration. It uses the information provided by the simulated samples of well known processes to correct the non-uniformity and non-linearity in jet η and p_T respectively. Lastly, the final set of corrections tends to the absolute and relative energy scale calibration, yielding the $p_{T,j}$(corrected) used in order to perform the $p_{T,miss}$ recalculation introduced with Equation 6.3.

Additionally, a set of dedicated filters, listed in Table 6.4, has been implemented by the Jet/MET POG [98, 111] in order to mitigate issues of high $E_{T,miss}$ originating

---

[6] The final state particles originating from lepton decays of taus are already included in respective muon/electron collections.

from detector problems. They are used to account for contributions arising from detector effects (HCAL/ECAL noise, ECAL calibration, etc.), beam-halo particles and cosmic rays. The procedure follows a simple path, if a filter associates the reconstructed $E_{T,miss}$ to be connected to one of the aforementioned sources, it is marked as being ”fake” and the evenr is discarded. These filters are applied as selection requirements at the analysis level.

The performance of this approach is presented in Chapters 7 and 8 for all, VBF H→inv analysis relevant, regions. A summary of the performance of $E_{T,miss}$ reconstruction is given in Ref.[112, 113] [7].

Filter description	Applied in data (simulation)
Primary vertex filter	✓(✓)
Beam halo filter	✓(✓)
HBHE noise filter	✓(✓)
HBHEiso noise filter	✓(✓)
ECAL TP filter	✓(✓)
Bad PF Muon filter	✓(✓)
EE badSC noise filter	✓(×)
ECAL bad calibration filter update	✓(✓)

Table 6.4: The list of $E_{T,miss}$ filters recommended by the JME POG [98, 111] applied both in 2017 and 2018. Almost all filters are applied both in data and simulation with the exception being the bad super cluster (EE badSC) filter.

6.3 Data and simulation samples

This study focuses on data collected by the CMS experiment during 2017 and 2018 eras of data taking, resulting with total integrated luminosity values of 41.5 and 59.8 fb^{-1} respectively [114, 115]. The main focus of this section is the summary of details regarding these datasets as well as the introduction of the approach taken with simulation samples of SM processes. These will include additional corrections which are applied to simulation samples in order to accurately account for the real performance of the experiment already reflected in data.

6.3.1 Overview

Starting first with data, a strategy following similarities between trigger algorithms is applied when storing the data (grouping algorithms targeting similar phase

---

[7] These studies are basing their measurements around processes which are well known and do not contain real $E_{T,miss}$. These includes final states such as Z→ e^+e^- and Z→ $\mu^+\mu^-$.

space). This analysis relies on a few of these groups, with the main one being the ”MET” dataset. It combines all events which have triggered an logical OR of algorithms based on the $E_{T,miss}$ variable, which included the main triggers used in the formation of the signal region for this analysis (summarised in Table A.1). Additionally, ”SingleElectron” (”EGamma” for 2018) and ”SingleMuon” datasets are used when forming dedicated control regions (being inclusive of trigger algorithms used to form these region).

In order to compare the observed results with predictions associated with the SM, a set of simulated samples covering the main sources of SM backgrounds are used in the analysis. The main production details about these samples, accompanied with the relevant signal samples, are summarised in Table 6.5. The general workflow used when generating these samples follows the procedure where the initial production is performed using the *POWHEG* [116] or *MADGRAPH5_aMC@NLO* [117] generators which are then interfaced with *PYTHIA* [118] (through the usage of the *CP5* tune)[8].

In order to recreate the conditions of the CMS experiment for the corresponding era, the final state particles are passed through a framework based on the *GEANT 4* package [119]. Finally, simulation samples for signal processes, in this case VBF and ggH production topologies, are produced at NLO using the *POWHEG* generator. All simulation samples are weighted to their respective cross sections as listed in Ref. [98].

6.3.2 Trigger re-weighting

An event by event based re-weighting procedure is applied to simulation samples in order to match the trigger performance in data. Trigger efficiencies are measured both in data and in simulation from which a scale factor is derived and used as the final weight. Detailed description of efficiency studies and the final estimation of trigger scale factors are given in Chapter 7.

6.3.3 Pile-up re-weighting

When looking at the pile-up conditions in data and simulation samples, it can be seen (similarly to the previously described trigger performance) that there is a discrepancy between the two. A re-weighting procedure is applied in order to mitigate this effect. The approach taken here follows the standard recipe presented in Refs. [98, 120], which involves matching the pileup distribution of simulated samples with the actual distribution obtained from data.

---

[8] Using the terms LO and NLO to denote the leading and next to leading order, respectively.

SM background process	Details
QCD/EWK Z($\nu\nu$)+jets	LO - QCD (bins of H_T)/EWK *MADGRAPH* generator
QCD/EWK W(lν)+jets	LO - QCD (bins of H_T)/EWK *MADGRAPH* generator
QCD/EWK Z(ll)+jets	LO - QCD (bins of H_T)/EWK *MADGRAPH* generator
Top	NLO - *POWHEG* generator (single top) NLO - *MADGRAPH@aMC@NLO* generator ($t\bar{t}$)
VV (dibosons: WW, WZ and ZZ)	LO - *PYTHIA8* generator
Signal process	**Details**
ggH→inv	N3LO - *POWHEG/PYTHIA8* generator
VBF H→inv	NLO - *POWHEG/PYTHIA8* generator

Table 6.5: List of main simulation samples originating from SM processes, with the corresponding production details [98].

6.3.4 Level-1 pre-fire effect

During the Run 2 phase of data taking, ECAL crystals located in the high $|\eta|$ regions suffered from a loss of transparency, due to radiation damage. This has led to an effect called the Level-1 pre-firing (addressed in Section 4.2.3). In order to mitigate this effect, which unfortunately affected this analysis due to its dependence on forward jets, another re-weighting procedure had to be applied. To account for the lack of this issue in simulation samples, there was a need to compute how probable would it be for an event not to pre-fire [121]. This probability and the final weight can be expressed as:

$$w_{\text{pre-fire}} = 1 - P(\text{pre-firing}) = \prod_{i=\gamma,\ j} (1 - \epsilon_i^{pref}(\eta, p_T)), \tag{6.4}$$

where the product runs over all offline photon and jet objects, and the ϵ_i^{pref} represent two-dimensional (p_T, η) pre-fire maps derived separately for jets and photon objects.

6.3.5 Lepton and b jet related weights

As this analysis uses leptons for two purposes, to select or veto a region, two different approaches are taken when looking at weights associated with their behaviour. The starting point for both of these scenarios is the discrepancy between data and simulation when it comes to reconstruction processes (including identification and isolation) of leptons. A set of data to simulation scale factors (expressed in terms of lepton p_T and η) is provided by the corresponding POGs [109, 122, 123]. They are computed through the use of selection efficiencies coming from special, lepton enriched regions.

For the formation of dedicated lepton control regions for the purposes of this study, one of the main requirements is the existence of at least one lepton (e or μ flavour) in the event. For these scenarios, events are re-weighted as:

$$w_{selection} = \prod_l \frac{\epsilon^l_{data}}{\epsilon^l_{simulation}}, \tag{6.5}$$

where the product runs over all elements of a given lepton collection, and $\epsilon_{data/simulation}$ represent the aforementioned efficiencies measured from data and simulation respectively. A similar approach can be taken when vetoing the events where, instead of asking for a hard $N_{lepton} = 0$ requirement, simulated events are weighted with a veto weight defined as:

$$w_{veto} = \prod_l \left(1 - \frac{\epsilon^l_{data}}{\epsilon^l_{simulation}}\right), \tag{6.6}$$

where in this scenario l represents the product of b jet collection as well as the lepton ones. The corresponding b jet weights are computed through the usage of POG recommended scale factors [124].

6.3.6 Higher order corrections

This step was introduced to further help with the understanding of the agreement between data and simulation in respective regions of interest for this study. It originated as causal effect of choosing to produce LO samples for the main V+jets backgrounds (which ensured easier production of a large number of simulated events). As a result it was necessary to apply higher order QCD and EWK corrections to the corresponding V+jets production modes in order to have a better understanding of their contribution. Table 6.6 summarises these corrections and their association to different V+jets production scenarios. The following paragraphs introduce each of the corrections used in this study.

V+jets process	Production	Perturbation order	NLO QCD	NLO EWK
Z→ll/$\nu\nu$	QCD	LO	✓	✓
	EWK	LO	✓	–
W→ $l\nu$	QCD	LO	✓	✓
	EWK	LO	✓	–

Table 6.6: Summary of higher order correction applied to main V+jet background processes [98].

A common thread for both types of corrections is that their derivation and subsequent application relies on a generator level property, the boson transverse momenta (p_T^V). For V+jets processes it is computed from generator level leptons combined into a dilepton object (the procedure of defining the object takes place before the final-state radiation) [98, 125].

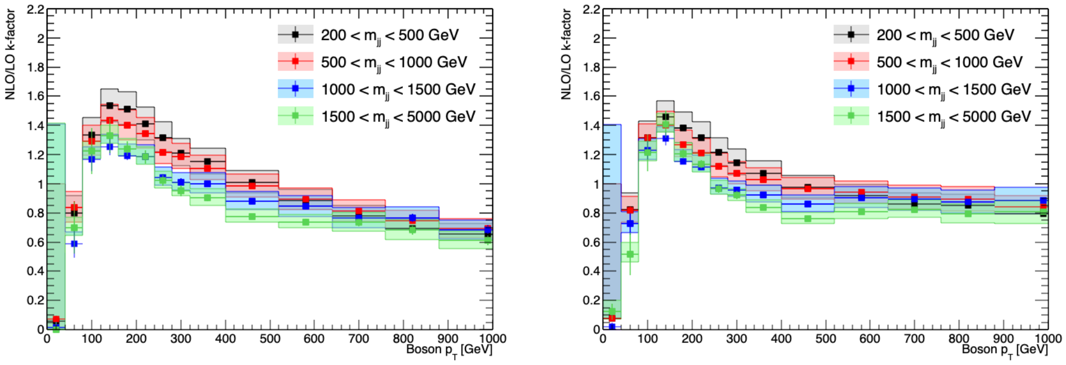
Starting first with the EWK production of V+jets processes. These SM backgrounds yield a significant contribution in the high m_{jj} spectrum (as seen in Figures 9.4 - 9.8) motivating the further investigation of higher order corrections. The QCD corrections (NLO k-factors) for these processes are derived in the form of a two-dimensional (p_T^V, generator m_{jj}) weight as explained in Refs. [98, 125].

The QCD NLO correction on the QCD V+jet processes are derived specifically with two main analysis categories in mind (a detailed description of each of them is given in Chapter 7). The NLO simulation samples of V+jet processes were generated using the *MADGRAPH_aMC@NLO* framework with up to two additional partons included when forming the matrix element [98]. The selection used at generator level objects closely mimics the offline selection requirements used for categories formed around the $E_{T,miss}$ and VBF based triggers.

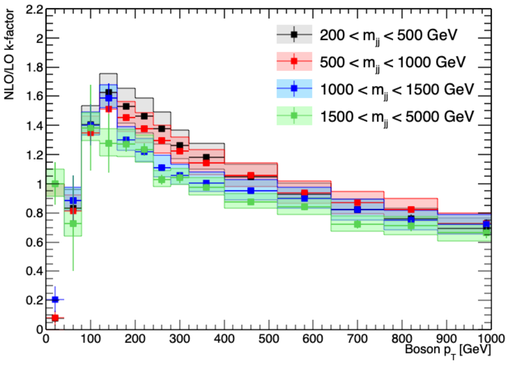
This states that an event with at least two generator level jets will be asked to have the leading pair pass equivalent selection requirements as the ones defined in Sections 7.2.1 and 7.2.2 (this time being applied on generator level jets[9]). An additional requirement on the boson mass ($60 < m_{V=Z} < 120$ GeV) is applied for Z+jets processes. The corresponding scale factor used as the event weight for LO samples is derived, again as a function of (p_T^V, generator m_{jj}), as SF = NLO/LO (where NLO and LO represent the contributions of events passing aforementioned selection requirements). Figures 6.4 and 6.5 show the resulting scale factors for both of these categories. A similar approach is taken when applying the EWK corrections for QCD V+jets processes. A special (p_T^V, generator m_{jj}) weight map is derived through the application of an equivalent generator level selection as the one used above. Finally, these weights are all applied on an event by event basis [98].

---

[9] The generator jet collection has leptons/photons removed.

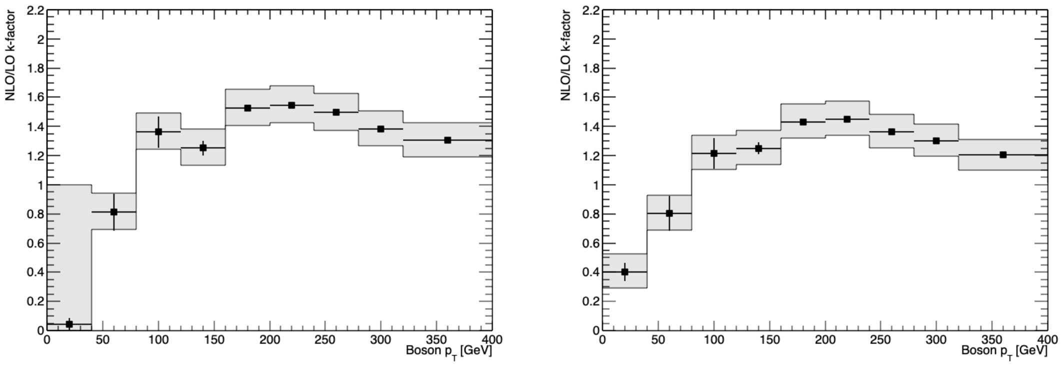


(a) Z → ll+jets (b) W → $l\nu$+jets

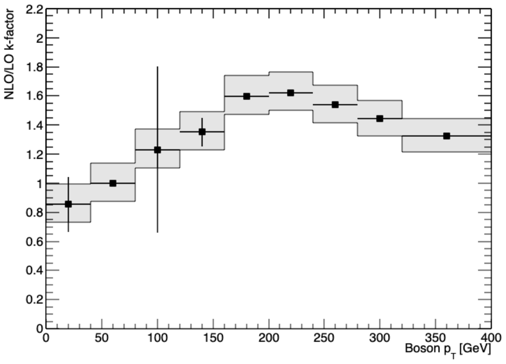


(c) Z → $\nu\nu$+jets

Figure 6.4: The LO-to-NLO theory scale factors binned in the generator level p_T^V and m_{jj}, shown for QCD V+jets processes. The scale factors are derived within the generator level selection requirements equivalent to the ones used to form the analysis category defined in Section 7.2.1. The error bars reflect the statistical uncertainty on the bin, while the bands represent the total systematic uncertainty [98].



(a) Z → ll+jets (b) W → $l\nu$+jets



(c) Z → $\nu\nu$+jets

Figure 6.5: The LO-to-NLO theory scale factors binned in generator level p_T^V, shown for QCD V+jets processes. The scale factors are derived within the generator level selection requirements equivalent to the ones used to form the analysis category defined in Section 7.2.2. The error bars reflect the statistical uncertainty on the bin, while the bands represent the total systematic uncertainty [98].

Chapter 7

Analysis strategy

> *"All men can see these tactics whereby I conquer, but what none can see is the strategy out of which victory is evolved."*
>
> — Sun Tzu

7.1 Introduction

THIS CHAPTER SERVES TO PRESENT a detailed documentation of the analysis strategy used for the search for invisibly decaying Higgs bosons, where the Higgs boson is produced via the Vector Boson Fusion production mechanism. The study at hand has had a long history within the CMS experiment starting all the way back in the early days of Run 1 [39]. This thesis tends to build on conclusions and methods achieved with previous efforts by improving them where possible, while also taking the advantage of the full Run 2 dataset.

Each of the following sections summarizes motivations and definitions that came into fruition while forming two main analysis categories. Special attention will be given to the analysis category based around new, production mode targeting triggers introduced in Section 4.5, which allowed for further exploration of the signal sensitive phase space. Being a large part of the motivation for the selection requirements, performance studies of analysis related trigger algorithms are going to be presented at this stage. Finally, this chapter concludes with a discussion regarding additional data quality issues plaguing the 2017 and 2018 eras of data taking and studies performed in order to mitigate their effects on the final result.

In order to be consistent with the notation used in Chapter 10, which presents a strategy for a combination of all hadronic production modes of the Higgs boson, a simple naming convention is going to be used based on the two main focuses in terms of the preferred Higgs boson production mode. All studies focusing on the Higgs boson produced via the Vector Boson Fusion will be grouped under one roof named the “VBF analysis”, while the remaining modes of interest such as the ttH, ggH and the VH production will represent the “non-VBF analysis”.

7.2 Selection requirements

This section serves as a summary of two main analysis categories. In order to simplify the way of addressing different categories within the VBF analysis, the following notation is used in future text. All studies built around the $E_{T,miss}$ and $H_{T,miss}$ triggers shown in Table 7.3 are part of the Missing Energy Trigger (MTR) category. On the other hand, the new category being formed through the use of the VBF production mode triggers (listed in Table 4.2) is named the VBF Trigger (VTR) category. A complete set of information regarding the L1 seeds used as inputs to the HLT algorithms forming these analysis categories is given in Table A.1.

The common ground for both categories is the approach to rejecting the contributions from major sources of SM background when forming the signal region (SR) through the implementation of object vetos. In order to remove the reducible contribution coming from main V+jets backgrounds $\mu/e/\tau$ vetos are imposed. A similar approach is taken in order to contain the γ+jets processes with a veto on photon objects being put in place. Finally, a b-jet object veto requirement reduces the background originating from top-quark SM processes. The implementation of the vetos follows the strategy described in Section 6.3, where a veto weight is applied for simulation samples as opposed to a $N_{object} = 0$ condition, which is used for data. The following pages introduce the main selection criteria for each of the categories.

7.2.1 Missing Energy Trigger category

This category follows the strategy published with the results originating from data collected in 2016 [39]. It is represented with the requirements shown in Table 7.1. With the set of object vetos already covered in the introduction, the discussion regarding the rest of the requirements can be split into two categories. Their origin can be traced to be either related to the topological properties expected from VBF jets, introduced in order to reduce a major contribution from SM processes, or they are purely motivated by the performance of the HLT algorithms used for data collection.

Variable	Selection	Target background
μ (e) veto	$p_T > 10$ GeV, $\|\eta\| < 2.4(2.5)$	$Z(ll)$ +jets, $W(l\nu)$ +jets
τ lepton veto	$p_T > 20$ GeV, $\|\eta\| < 2.3$	$Z(ll)$ +jets, $W(l\nu)$ +jets
γ veto	$p_T > 15$ GeV, $\|\eta\| < 2.5$	γ +jets
b jet veto	$p_T > 20$ GeV, $\|\eta\| < 2.4$	Top quark
$E_{T,miss}$	>250 GeV	QCD, top quark, $Z(ll)$ +jets
$min\Delta\phi(j, E_{T,miss})$	>0.5 radians	QCD
$\|1-E^{Calo}_{T,miss}/E_{T,miss}\|$	<0.5	QCD
p_{T,j_1} and η_{j1}	>80 GeV and $\|\eta\| < 4.7$	All
p_{T,j_2} and η_{j2}	>40 GeV and $\|\eta\| < 4.7$	All
$\eta_{j1} \cdot \eta_{j2}$	< 0	All
m_{jj}	>200 GeV	All
$\Delta\eta_{jj}$	>1.0	All
$\Delta\phi_{jj}$	<1.5	All

Table 7.1: Summary of the MTR selection requirements, accompanied with the target background processes affected by them [98]

Starting with the topological information, the VBF signature is characterised by a jet pair which has a large geometrical separation and therefore a large dijet mass. When interpreted in terms of the detector geometry (introduced in Section 3.2.1), this leads to conditions that jets, chosen as the two leading in p_T, have to be separated by a large value of $\Delta\eta$, a small $\Delta\phi$ and to have $\eta_{j1} \cdot \eta_{j2} < 0$ by being in the opposite halves of the detector.

Moving on from the purely topological properties, the next step is to try and determine which variables represent a good basis for additional lowering of the contribution coming from various sources of SM background. Variables such as the $min\Delta\phi(j, E_T^{miss})$[1] allow for control of the QCD multijet background. Figure 7.1[2] shows its separation power using the N-1 selection[3] approach. With it showing combined background composition originating from simulated samples of SM processes overlaid with the distribution of the signal, it can be seen that the requirement $min\Delta\phi(j, E_{T,miss}) > 0.5$ yields a significant reduction of QCD multijet backgrounds, without a large loss of signal sensitivity. Another useful requirement comes from the comparison of the offline (after the full reconstruction) and calorimeter only $E_{T,miss}$.

---

[1] The minimal $\Delta\phi$ between one of the four leading jets and the $\vec{p}_{T,miss}$.

[2] Signal yield has been scaled by a factor of 5 in order to have a better visual comparison against the main SM backgrounds (same statement holds true for the rest of this chapter).

[3] All of selection requirements are being applied but the one involving the variable of interest.

A requirement that the relative difference is less than 50% (when taking the offline $E_{T,miss}$ as reference) provides another way of rejecting events which contain energetic mismeasured jets originating from multijet processes.

As indicated at the beginning of this section, geometrical properties of the selected dijet pair play a large role in the recognition of signal-like events, as illustrated in Figure 7.1. This comparison of $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ distributions between main backgrounds and signal shows a clear opening for a set of requirements which improve the selection. An optimisation procedure, taking into account the entire setup for the analysis (including the contributions from dedicated control regions), was performed [39] in order to obtain the thresholds presented in Table 7.1. It involved scanning for the best expected sensitivity towards the Br(H→inv) based on different variations of m_{jj}, $\Delta\phi_{jj}$ and $\Delta\eta_{jj}$ selection requirements. Testing was performed in two instances, first being optimised only for the VBF H→inv signal in order to have a VBF topology overlook and second for the combined VBF and ggH→inv signal (providing a better look into the contamination of VBF topology introduced by the addition of the ggH production mode).

Finally, there are requirements that arise strictly from trigger limitations. This can be seen in the choice of the $E_{T,miss}$ threshold, where the requirement on it being larger than 250 GeV was imposed in order to stay above the 95 % efficiency for the category-forming triggers, when measured in data. This efficiency improves to 99 % for the values of $E_{T,miss} > 300$ GeV. A more detailed description of this measurement is given in Section 7.3.1.

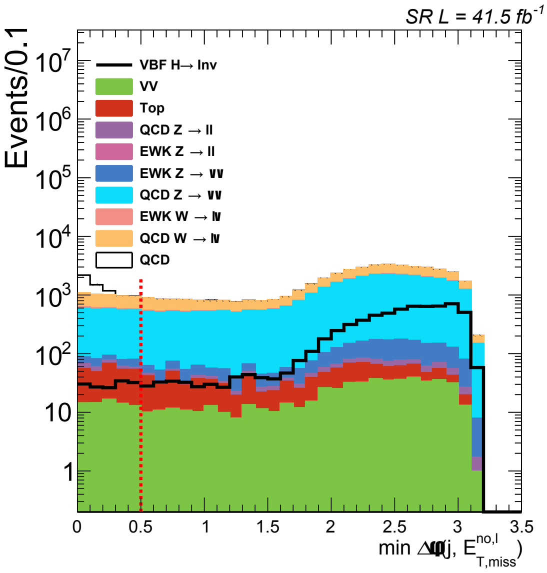
Lastly, Figure 7.2 shows data to simulation agreement for a selected set of main analysis variables after the application of the complete MTR selection for the 2017 era of data taking[4]. Corresponding information regarding the SR for the 2018 era is given in Appendix A.3. The MTR category is constructed to represent the main analysis category, covering a large piece of the phase space of interest. In order to further improve on it, the following section describes a complementary category formed around the set of VBF triggers introduced in Section 4.5.

7.2.2 VBF Trigger category

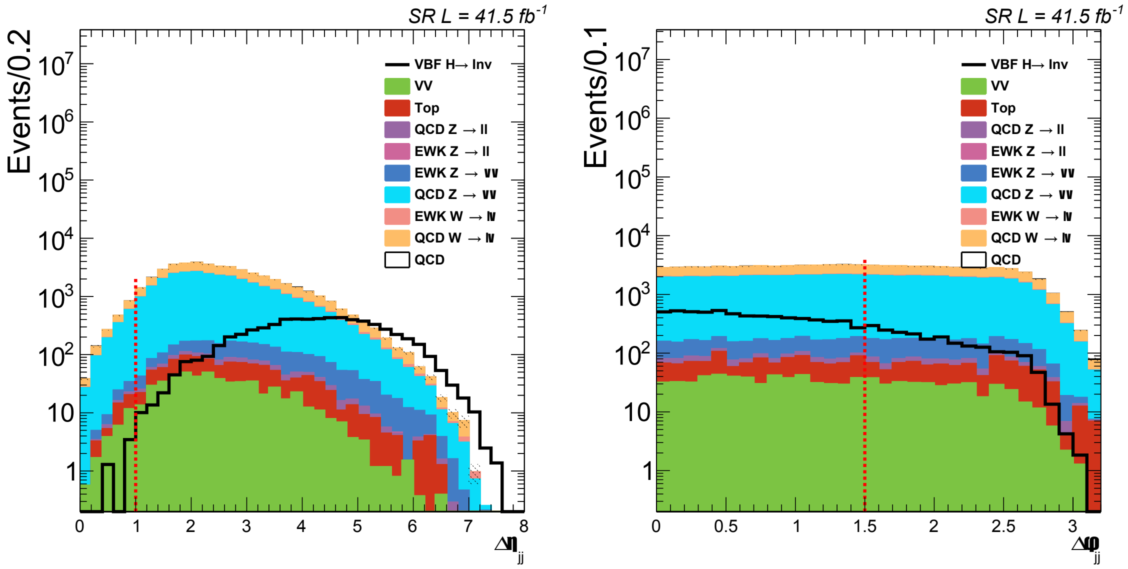
Continuing the narrative started in Section 4.5, the main focus of this section is going to be the creation of a new analysis category designed to select the phase space of interest that was inaccessible to the MTR category. Following the, previously discussed, blueprint for the MTR category, there is an opening to form a category

---

[4] These selection requirements are inclusive of the additional data quality requirements introduced in Section 7.4 (this holds true for both categories across all eras).



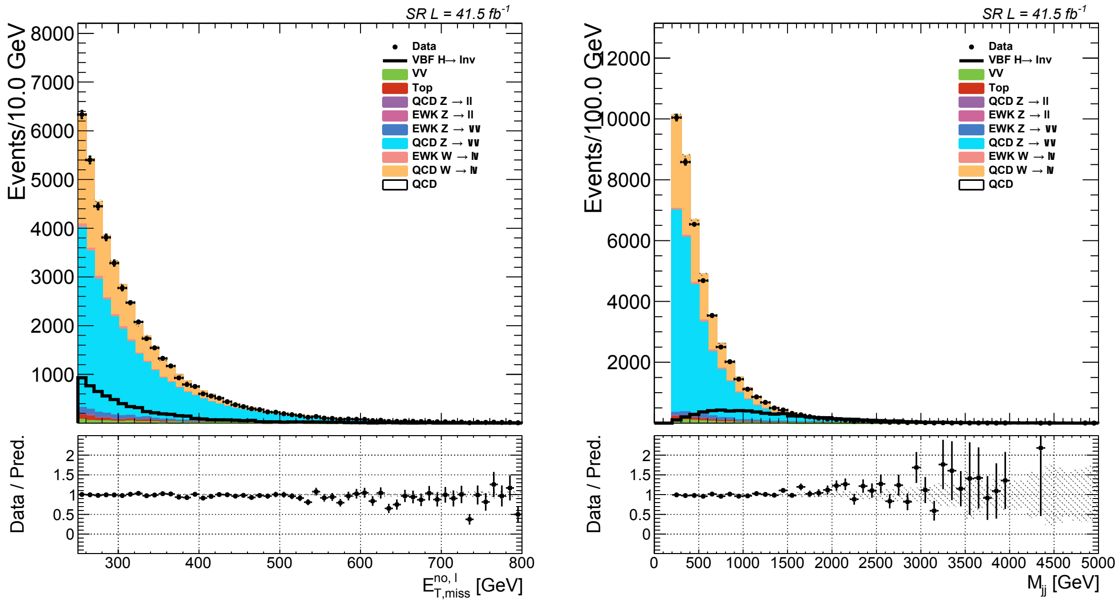
(a) $min\Delta\phi(j, E_T^{miss})$



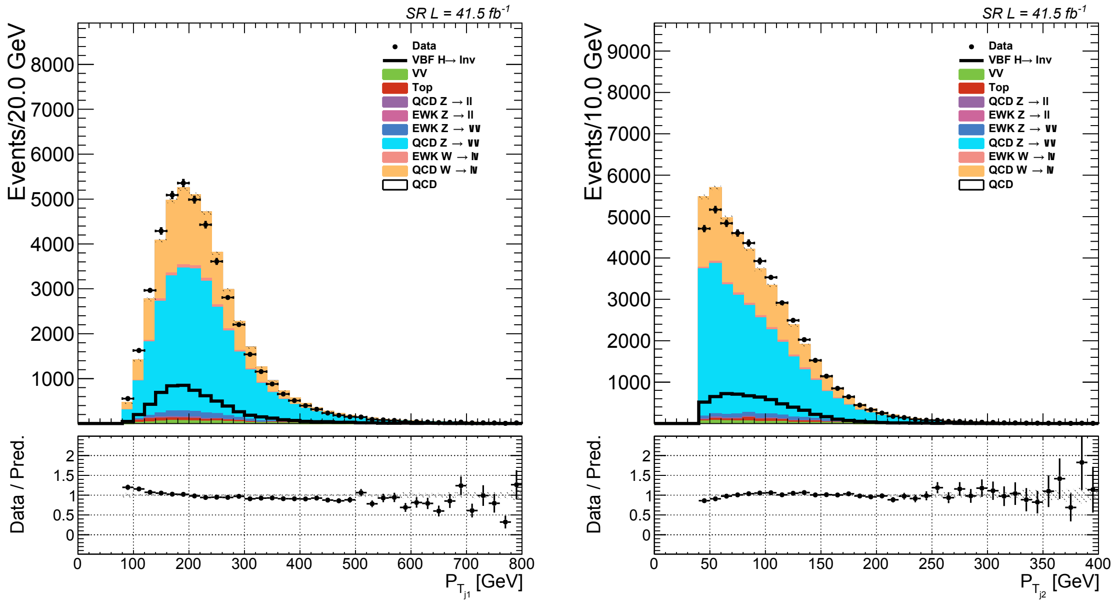
(b) $\Delta\eta_{jj}$

(c) $\Delta\phi_{jj}$

Figure 7.1: Distributions of $min\Delta\phi(j, E_{T,miss})$, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ variables in the SR, for the MTR category after the N-1 selection, for the 2017 data.



(a) $E_{T,miss}$ (b) m_{jj}



(c) $p_{T,j1}$ (d) $p_{T,j2}$

Figure 7.2: Distributions of $E_{T,miss}$, m_{jj}, $p_{T,j1}$ and $p_{T,j2}$ variables in the SR after the full MTR selection, for the 2017 data.

that will be orthogonal to it. One way to look for its basis is to start from the comparison of performances of both trigger groups and picking an $E_{T,miss}$ range which is outside the MTR threshold, in which VBF triggers perform better than the MTR forming ones.

As will be explained in Section 7.3.2, this study results in a category formed within the $[160, 250)$ GeV range of the $E_{T,miss}$ variable. In order to follow the logic deployed at the trigger level, the choice of jets at the analysis level is again based on the dijet pair which yields the largest invariant mass in the event. In retrospect, this choice follows the equivalent procedure to the one presented in Section 4.5.2. The jets entering the computation are required to have $p_T > 40$ GeV in order to mimic the trigger logic, while the slightly larger threshold is used to account for subtle differences between the HLT and offline jets. Upon constructing the dijet object and optimising the selection based on the trigger performance, thresholds on the values of the dijet mass and transverse momentum of the two jets were set. These selection requirements are then used to form the VTR category and are summarised in Table 7.2.

Variable	Selection	Target background
μ (e) veto	$p_T > 10$ GeV, $\lvert\eta\rvert < 2.4(2.5)$	$Z(ll)$ +jets, $W(l\nu)$ +jets
τ lepton veto	$p_T > 20$ GeV, $\lvert\eta\rvert < 2.3$	$Z(ll)$ +jets, $W(l\nu)$ +jets
γ veto	$p_T > 15$ GeV, $\lvert\eta\rvert < 2.5$	γ +jets
b jet veto	$p_T > 20$ GeV, $\lvert\eta\rvert < 2.4$	Top quark
$E_{T,miss}$	$[160, 250)$ GeV	QCD, top quark, $Z(ll)$ +jets
$min\Delta\phi(j, E_{T,miss})$	>1.8 radians	QCD
Largest m_{jj}	> 900 GeV	
p_{T,j_1} and η_{j1}	>140 GeV and $\lvert\eta\rvert < 4.7$	All
p_{T,j_2} and η_{j2}	>70 GeV and $\lvert\eta\rvert < 4.7$	All
$\eta_{j1} \cdot \eta_{j2}$	< 0	All
$\Delta\eta_{jj}$	>1.0	All
$\Delta\phi_{jj}$	<1.8	All

Table 7.2: Summary of the VTR selection requirements, accompanied with the target background processes affected by them [98].

This approach ensures orthogonality between two analysis categories. The high thresholds for jet p_T values and the large dijet mass requirement reduce the difference between two choices of the jet pair between categories[5]. The requirements on the $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ are built on this, allowing for the usage of similar requirements

---

[5] With the MTR category looking for the leading and subleading jet, and the VTR taking the jet pair forming the largest dijet mass.

(slightly relaxed for the $\Delta\phi_{jj}$ case) to those introduced for the MTR category.

As seen in the previous section, a good handle for dealing with QCD multijet processes is given in the form of the $min\Delta\phi(j, E_{T,miss})$ variable. Figure 7.3 shows the distribution of this variable in the VTR SR after the application of N-1 type selection. It can be seen that values bellow 1.8 are largely background dominated, with the QCD multijet processes leading the way, while being populated with a low yield of signal events, thus motivating a requirement which would remove this region without risking a significant loss in signal acceptance.

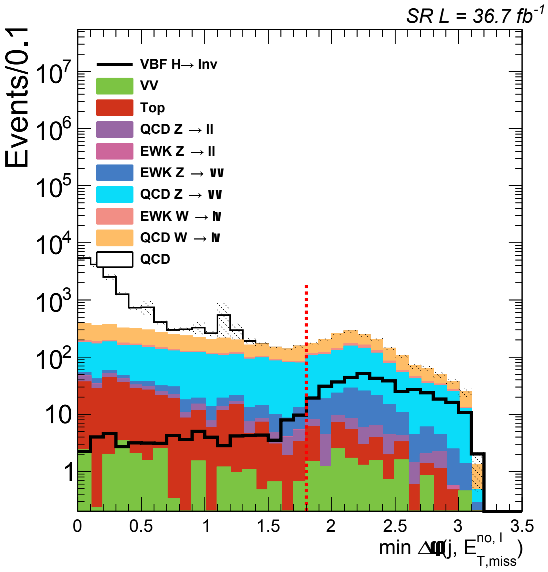
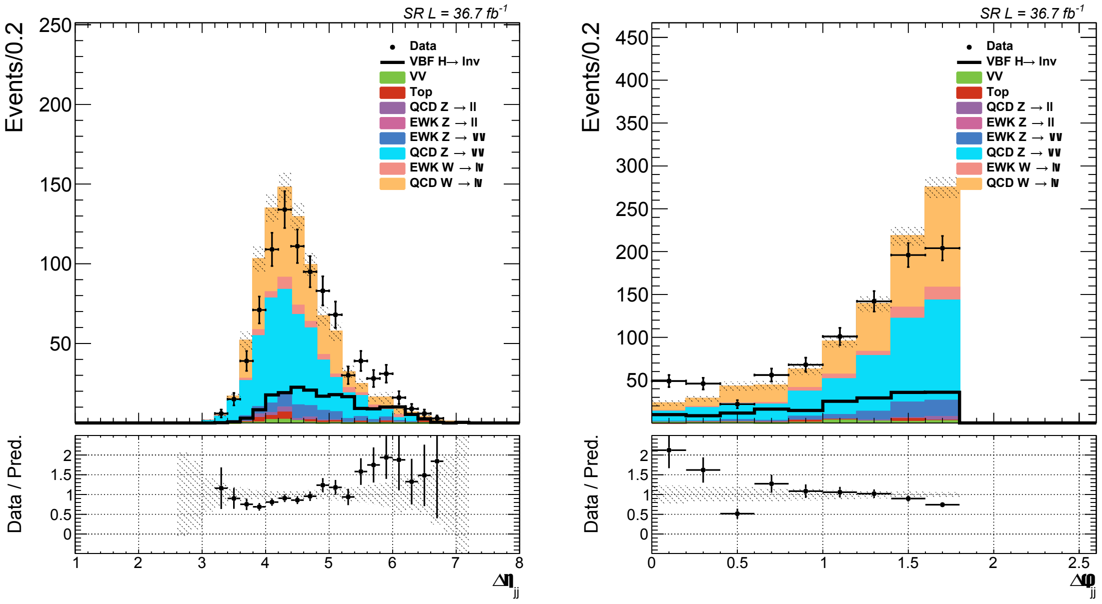


Figure 7.3: Distribution of the $min\Delta\phi(j, E_{T,miss})$ variable in the SR after the N-1 selection.

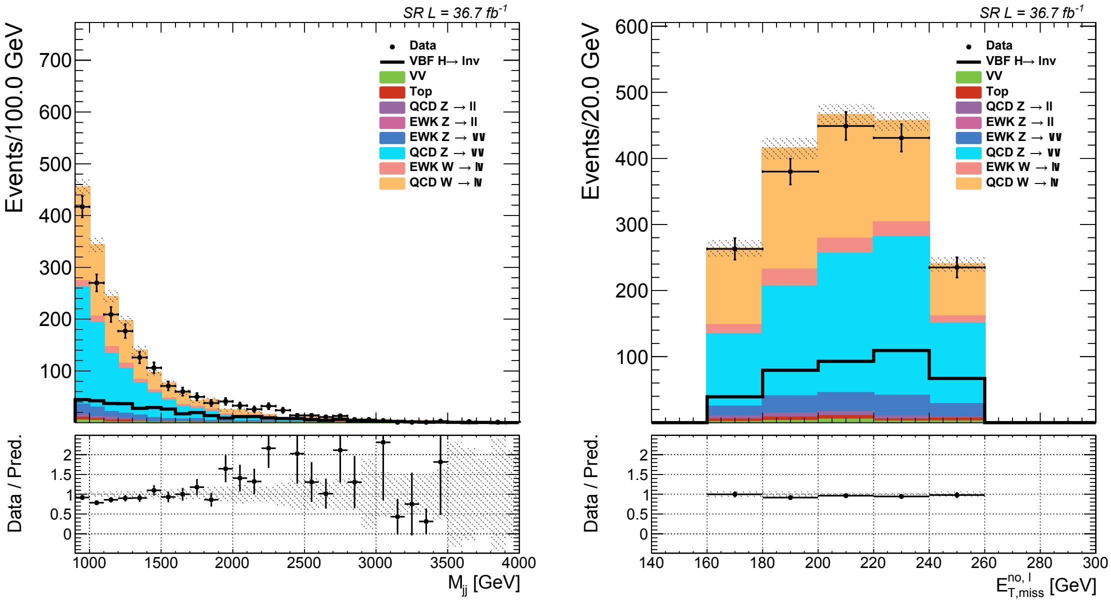
Figure 7.4 shows the data to simulation comparisons for a few selected variables after the full VTR selection. A set of dedicated control regions will be created for the VTR category, analogous to the MTR approach, in order to improve the overall data to simulation agreement. Additional set of distributions for the VTR category representing the 2018 era is given in Appendix A.3 alongside their MTR counterparts.

7.3 Trigger Performance

The following sections focus on the performance of the HLT algorithms used in this analysis. These paths can be split into two groups based on their usage: the signal/muon region and the electron region forming triggers. The added benefit of the signal triggers (listed in Tables 4.2 and 7.3) comes from the fact that they are built by adding conditions on $E_{T,miss}^{no,\mu}$ (as seen in the diagram presented in Figure 4.6).



(a) $\Delta\eta_{jj}$ (b) $\Delta\phi_{jj}$



(c) m_{jj} (d) $E_{T,miss}$

Figure 7.4: Distributions of $\Delta\eta_{jj}$, $\Delta\phi_{jj}$, m_{jj} and $E_{T,miss}$ variables in the signal region after the full VTR selection, for the 2017 data.

The $\vec{p}^{\,no,\mu}_{T,miss}$ vector, whose magnitude is represented by $E^{no,\mu}_{T,miss}$, is defined by adding the transverse momentum of muon objects back into the $\vec{p}_{T,miss}$ computation. This can be expressed as:

$$\vec{p}^{\,no,\mu}_{T,miss} = \vec{p}_{T,miss} + \sum_i \vec{p}_{T,\mu_i}, \tag{7.1}$$

where the sum goes over all muon objects. Such definition of the missing energy is equivalent to the standard $E_{T,miss}$ for the SR, due to the use of the muon veto. This re-definition extends the usability of the aforementioned trigger algorithms, which can now be used to select events entering muon regions (more details about these specialised regions are given in Chapter 8). For the regions containing electron objects, a combination of electron and photon triggers is used. The addition of photon triggers was performed in order to achieve better statistical precision in these regions (which will benefit the double electron region in particular, due to its tight requirements). The complete list of HLT paths used in this analysis (and the corresponding L1 seeds used as input) is given in Table A.1.

7.3.1 Performance of $E_{T,miss}$ and $H_{T,miss}$ based triggers

Following the ordering used in the previous section, the first topic of discussion is the performance of triggers used to form the MTR category. Listed in Table 7.3, they are represented by a set of selection requirements on $E^{no,\mu}_{T,miss}$ and $H^{no,\mu}_{T,miss}$ variables at the HLT stage. This section summarizes the study of their performance during the 2017 and 2018 eras of data collection. It is based around the analysis selection requirements for the MTR category (as shown in Table 7.1), presenting the motivation behind the $E_{T,miss}$ threshold (which is set with respect to the overall performance). More details about the choice of the selection itself can be found in Section 7.2.1.

Era	Trigger logic
2017 & 2018	$E^{no,\mu}_{T,miss} > 120$ GeV & $H^{no,\mu}_{T,miss} > 120$ GeV & Tight ID (& $H_T > 60$ Gev)

Table 7.3: A set of $E_{T,miss}$ and $H_{T,miss}$ triggers used in the analysis. The backup path contains an additional $H_T > 60$ GeV requirement.

During the 2017 era of data taking a special, control, version of main $E_{T,miss}$ path had to be introduced, as the main one was prescaled due to problems which led to the drastic increase of its rate. The origin of this problem is connected to the same issues affecting the L1 VBF seed as explained in Section 4.5. This has resulted in the inclusion of the additional $H_T > 60$ GeV requirement at the trigger level. This issue was quickly mitigated in 2017 by applying slightly higher thresholds on the input L1 seeds, thus reducing the amount of rate brought by the trigger. This backup path was kept for the entirety of the 2018 era, but with certain modifications applied to its input. As it was noticed that it brought an efficient reduction in rate when used as a backup, it was redeployed with lower requirements applied to the input L1 seeds, allowing for it to be used alongside the main path in the analysis.

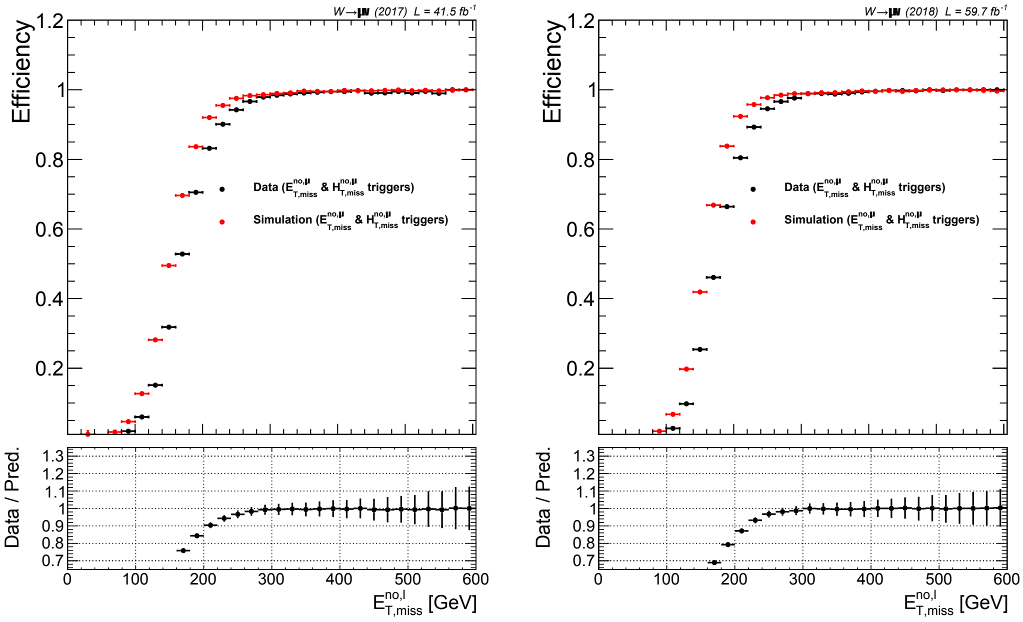
The general approach when measuring efficiencies for both eras is to form dedicated regions in which the actual measurement will be performed. For this study it is done by using a slightly modified definition of a single muon control region. As described in Section 8.2.2, this modified the muon veto condition by asking for the existence of exactly one muon object in the event which passes the tight conditions. For the measurement itself, a control HLT path is being used, which represents a single muon trigger looking for muon objects with $p_T > 27$ GeV which pass the isolation requirement at the HLT level. In order to ensure that the study is performed in a region of high control trigger efficiency (on the trigger ”plateau”), a requirement $p_{T,\mu} > 30$ GeV is imposed on the selected tight muon. The efficiency measurement is presented in terms of $E_{T,miss}^{no,\mu}$ bins and defined as:

$$\text{Efficiency} = \frac{\text{(Passing analysis selection requirements) } \textit{and} \text{ (HLT triggered)}}{\text{Passing analysis selection requirements}}, \quad (7.2)$$

The efficiency of the logical OR of the two HLT paths is measured both in data and simulation and the resulting difference is used as a scale factor which is applied in the main analysis to simulated samples on an event-by-event basis (as described in Section 6.3). Inputs for these measurements are formed from the muon enriched dataset and the QCD W+jets simulaton samples, representing the dominant SM process in this region. Figure 7.5 presents the resulting efficiencies, with the corresponding scale factors, for both 2017 and 2018 eras. An additional set of efficiencies was produced by separating jets into three categories by looking at the geometric properties of the two leading jets. This resulted into three categories: two central jets (CC), one central, one forward jet (CF) and two forward jets (FF)[6]. This was implemented in order to have a confirmation that no significant deviations appear between two regions with larger population of events (CC and CF). Figure 7.6 summarises these results by presenting the efficiencies expressed in $E_{T,miss}^{no,\mu}$ bins.

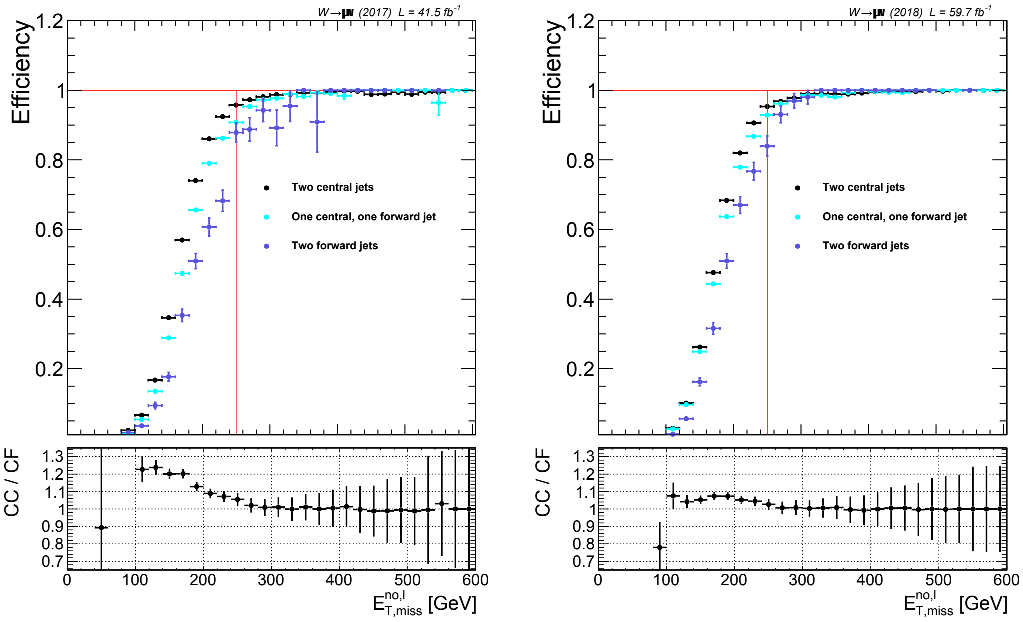
---

[6] A jet is categorised as central if $|\eta| < 2.5$, while it is a forward jet if $|\eta| \geq 2.5$.



(a) 2017 (b) 2018

Figure 7.5: Trigger efficiencies for the MTR forming algorithms presented in $E_{T.miss}^{no,\mu}$ bins for both eras. Separate efficiencies were measured for data and simulation with the resulting scale factor also being shown.



(a) E_T^{miss} - 2017 (data) (b) E_T^{miss} - 2018 (data)

Figure 7.6: Trigger efficiencies for the MTR category algorithms presented in $E_{T.miss}^{no,\mu}$ bins for both eras measured in data. Separation into three different jet η regions (CC, CF and FF) is performed. The resulting comparison between the CC and CF is also presented with the ratio plot.

7.3.2 Performance of VBF triggers

Similarly to previously described efficiency study regarding the MTR category triggers, the first step when approaching the VBF triggers is the formation of the single muon region. From this point, the general approach takes a slightly different turn. As the MTR category relied on already proven, optimised selection, the VBF triggers brought a new part of the phase space that is yet to be explored. Starting from the information about the building blocks of VBF paths (discussed in Section 4.5), the next step was the optimisation the of the selection requirements for three main variables: m_{jj}, $p_{T,j1}$ and $p_{T,j2}$. In order to follow the trigger logic, the selection of the dijet pair is based around the largest m_{jj} logic.

Figure 7.7 shows the resulting efficiency of the logical OR of both VBF triggers presented in terms of m_{jj} bins, where the thresholds for the other two main variables are kept high enough to ensure that an unbiased decision can be made. The resulting requirement of $m_{jj} > 900$ GeV is motivated by the desire to stay above the 95 % trigger efficiency in data. Figure 7.8 presents results of equivalent optimisation studies performed for the purposes of tailoring selection requirements for the jet p_T variables.

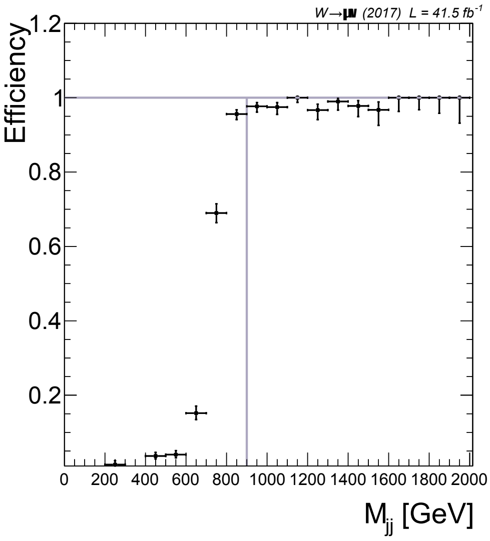
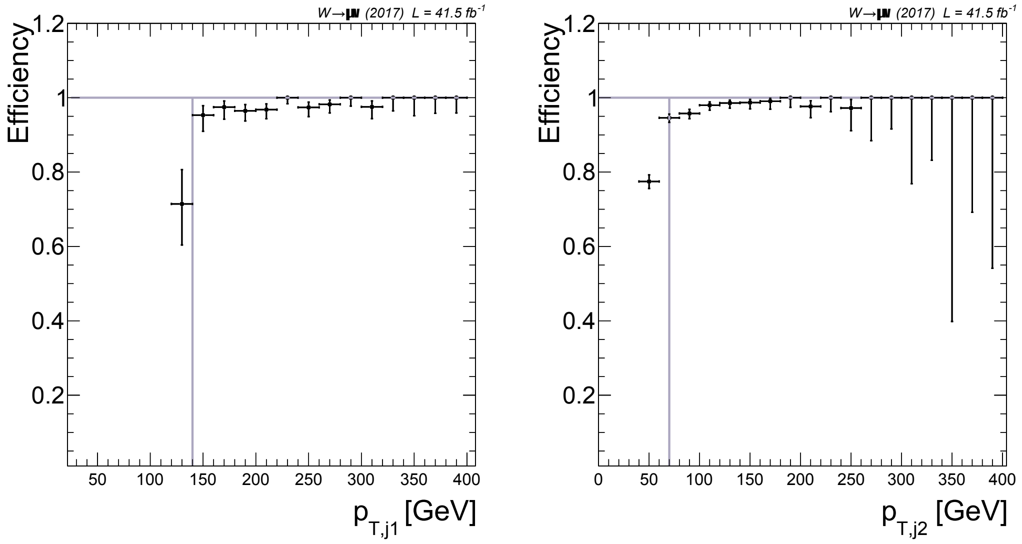


Figure 7.7: Trigger efficiency of the logical OR of both VBF triggers (performed in data), presented in m_{jj} bins, used to motivate a selection requirement for this variable.

Following the high thresholds on jet properties, the difference between selecting the two leading p_T jets and the ones forming the largest m_{jj} is small enough that it brings the option of using the rest of the already optimised MTR strategy into the VTR selection. At this stage, these requirements form almost the final VTR category, differing only with the $min\Delta\phi(j, E_{T,miss}) > 0.5$ requirement. Figure 7.9



(a) $p_{T,j1}$ efficiency

(b) $p_{T,j2}$ efficiency

Figure 7.8: Trigger efficiency of the logical OR of both VBF triggers (performed in data), presented in p_T bins for the leading (a) and subleading (b) jet, used to motivate selection requirements for these variable.

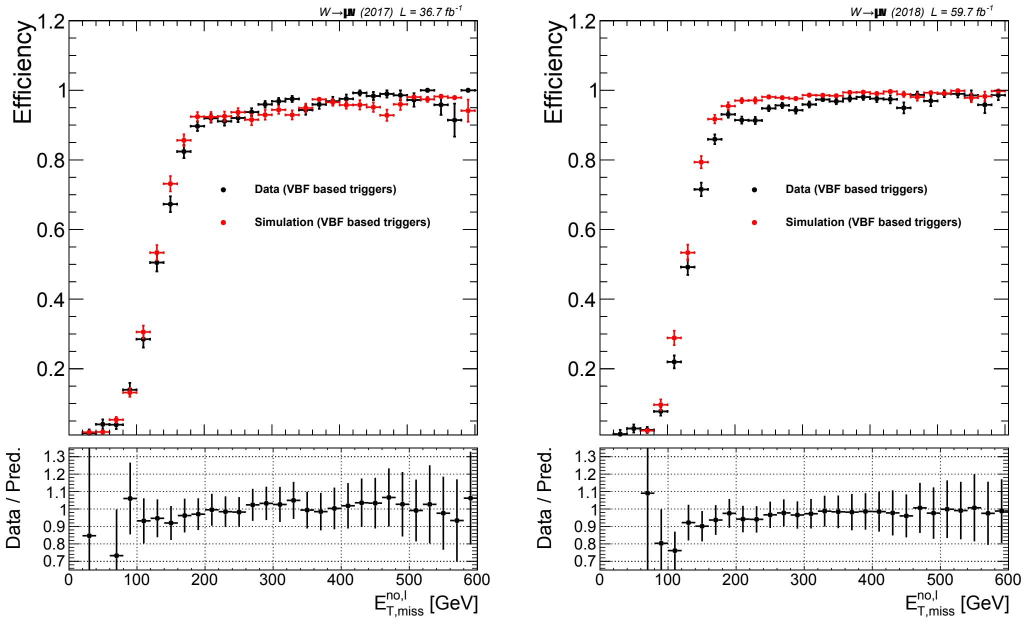
shows the data to simulation comparison of the performance of VBF triggers for both eras (for the current, proto-VTR, selection requirements).

The next step is the decision on the exact $E_{T,miss}$ range used to define the VTR category. In order to have a better overview, a comparison of performance for both sets of triggers needed to be made. Figure 7.10 presents the comparison between two trigger groups. Better performance for VBF triggers in the lower $E_{T,miss}^{no,\mu}$ range motivated the choice of the range for VTR to be [160, 250) GeV. The overlayed shape of the VBF H→inv simulation sample illustrates the potential signal gain achieved by this choice.

Upon forming the preliminary VTR selection and observing the background composition given with Figure 7.3, a tighter selection requirement was imposed on the $min\Delta\phi(j, E_{T,miss})$ variable in order to suppress the contribution originating from QCD multijet processes. Data to simulation efficiency comparison accompanied with the final scale factors used in the analysis are given in Figure 7.11.

7.3.3 Performance of the electron and photon triggers

For the purposes of creating dedicated control regions for V+jets background processes, in a similar vein to muon regions, a set of dedicated triggers is needed. This set is formed from three HLT paths. The first algorithm employed here is a single electron path requiring $p_T > 32$ GeV threshold at the HLT level (35 GeV for



(a) $E_{T,miss}$ - 2017

(b) $E_{T,miss}$ - 2018

Figure 7.9: Trigger efficiency of the logical OR of both VBF triggers presented in $E_{T,miss}$ bins, measured using the proto-VTR selection requirements.

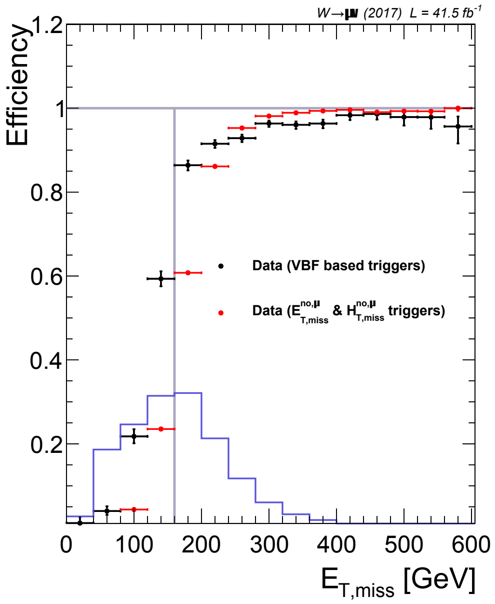
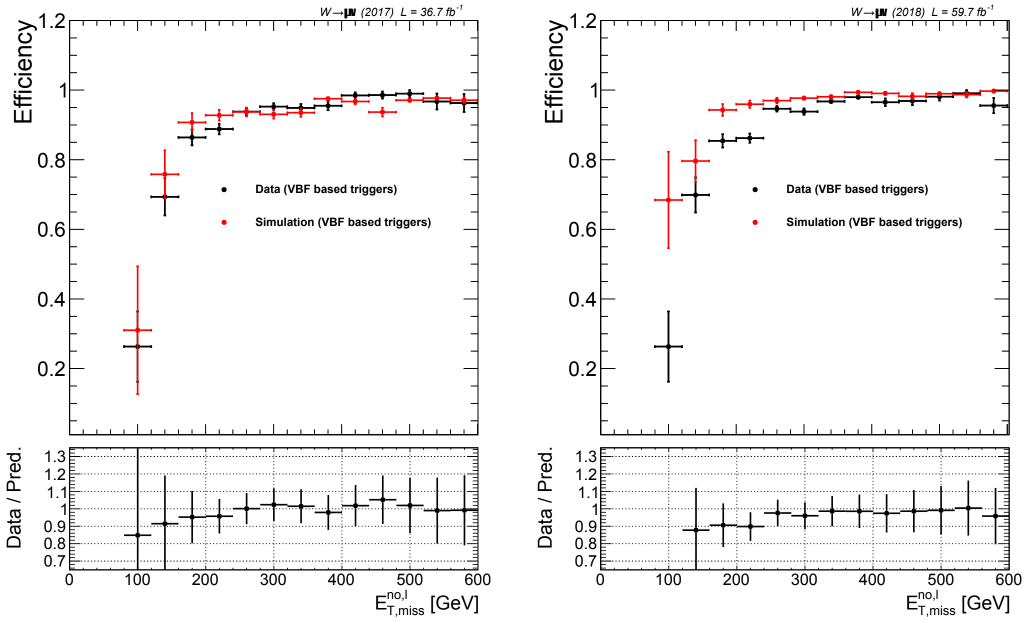


Figure 7.10: Comparison of efficiencies for both the MTR and VTR constructing trigger groups, presented in $E_{T,miss}$ bins. Study was performed for the 2017 era of data taking.



(a) $E_{T,miss}$ - 2017

(b) $E_{T,miss}$ - 2018

Figure 7.11: Trigger efficiency of the logical OR of both VBF triggers presented in $E_{T,miss}$ bins, measured using the full VTR selection requirements for both 2017 (a) and 2018 (b) eras.

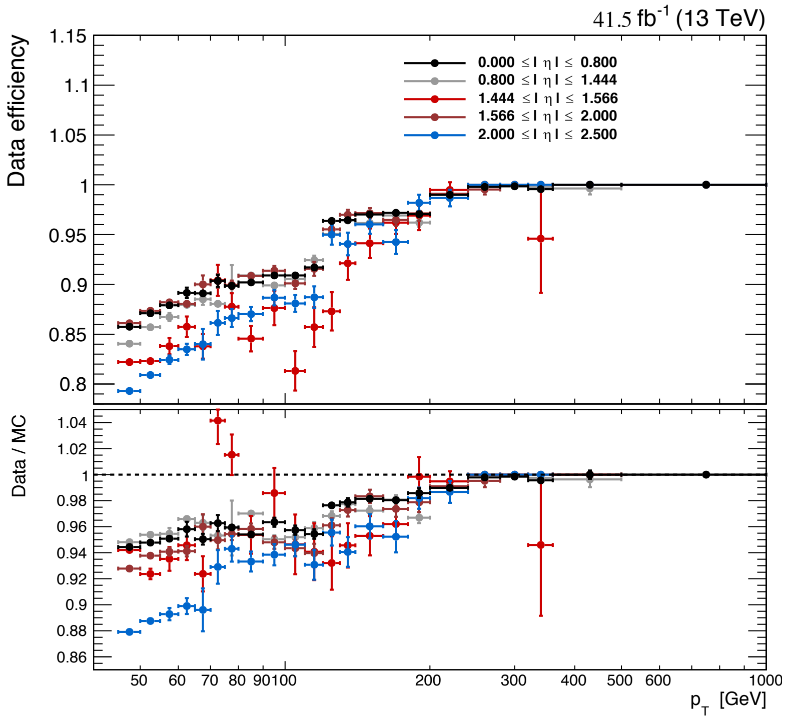
2018). This lower threshold is enabled through the implementation of the isolation requirement on the electron objects.

Continuing with the set, a higher p_T threshold electron trigger ($p_{T,e} > 115$ GeV) is used to increase the number of selected events, bringing the benefits of not being constrained by the isolation requirement. Finally, in order to further boost the statistical precision within these regions, a photon trigger (requiring $p_{T,\gamma} > 200$ GeV) is brought into the setup. A logical OR of these triggers is taken as the resulting algorithm used to select events. The efficiency measurement for this set of triggers is performed using the standard "tag and probe" algorithm recommended by the E/Gamma POG [126].

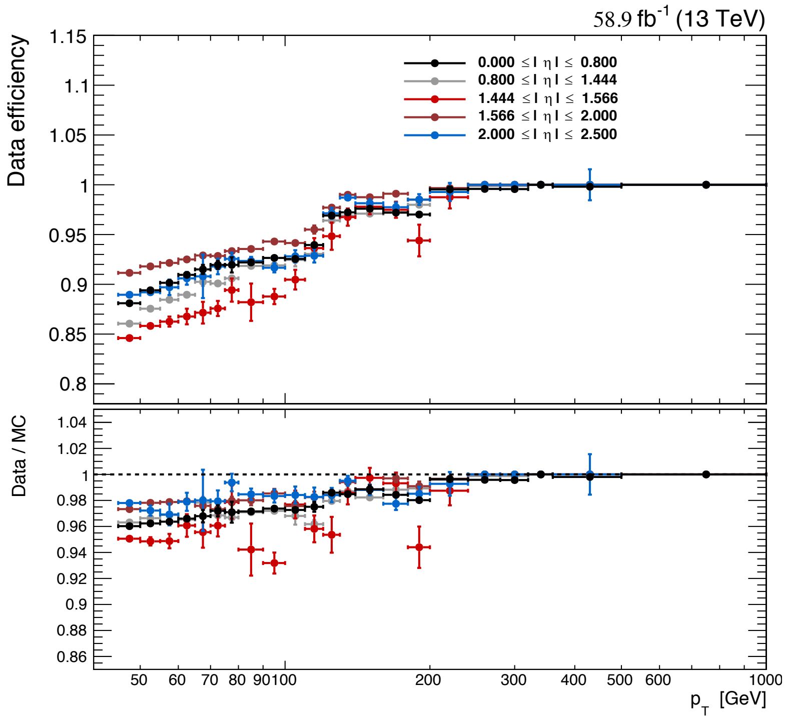
Figure 7.12 shows results of these efficiency measurements for data collected during 2017 and 2018. The usage of the aforementioned information requires special attention when forming the scale factors (SFs). Depending on the number of electrons they are defined as:

$$SF = \frac{1 - \prod_i (1 - \text{efficiency}_{data})}{1 - \prod_i (1 - \text{efficiency}_{simulation})}, \tag{7.3}$$

where the iteration goes over each electron in the collection, allowing for easier usage when defining both electron regions.



(a) 2017



(b) 2018

Figure 7.12: Efficiencies of the logical OR of the three aforementioned triggers used to select electron events for 2017 (a) and 2018 (b) as a function of the electron transverse momentum (separated into several categories in electron $|\eta|$) [98].

7.4 Data Quality Issues

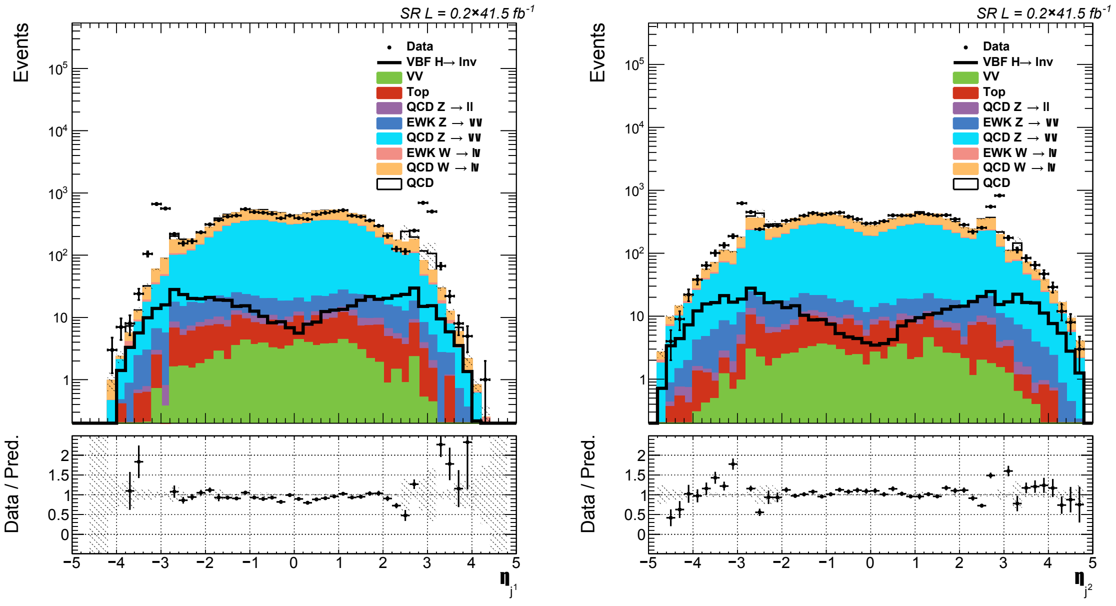
Midway during the 2018 era of data taking, an incident occurred at the CMS experiment, leaving a part of the HE unresponsive. In terms of detector geometry, this meant that no HCAL information was available in the $-90° < \phi < -50°$ and $-3.0 < \eta < -1.4$ region (HE sectors 15/16). This has affected ~ 65 % of the collected data during this era and has caused major concerns about its impact on the analysis. This effect was observed and treated in one of the control regions (more in Section 8.2.4), but the main concern was the SR. In order to be able to derive a proper mitigation for this effect, a blueprint involving unblinding 1/5th of the data was approved. It allowed for a look into "safe" variables that would show the effect of this problem, while not being one of the main dijet variables (removing the possibility of a biased strategy).

Additional motivation for this approach was also fueled by the desire to check for the appearance of the effect called the jet "horns". It is represented with a large number of events with jets being reconstructed in the $2.8 < |\eta| < 3.2$ region for data only (not being well modelled in the simulation). The following paragraphs summarise the strategies that were the result of this unblinding.

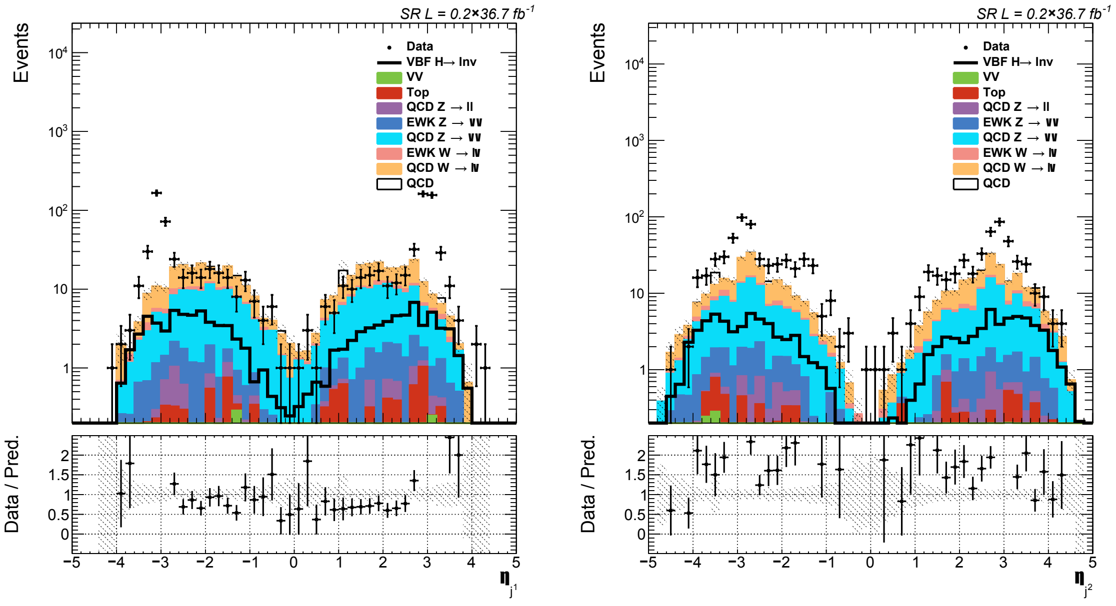
7.4.1 Jet "horns" mitigation

One of the problems associated with the studies focused on jets is the appearance of a large number of low quality jets in data for a certain $|\eta|$ range, not properly represented in the simulation. The range of this effect, covering the $2.8 < |\eta| < 3.2$ region, creates issues for this analysis as well, unfortunately for both eras. Figure 7.13 shows that even the PU jet ID, imposed when creating analysis level jets, was not enough to diminish this effect (even though this was enough to eliminate this occurrence in the dedicated control regions). The effect is more prominently seen for the the leading jet compared to the subleading jet, which is a clear consequence of the high $E_{T,miss}$ selection requirement (being computed from a higher p_T jet in the former case).

In order to look for a proper mitigation approach, various techniques for computing the $E_{T,miss}$ were checked (based on the set of particles entering the computation). As explained in Section 6.2.8, the offline $E_{T,miss}$ is computed using all PF objects. A restriction, allowing only objects which are registered by the tracker into the calculation of $E_{T,miss}$, would give a better control of the problem, as the difference between the two includes the neutrally charged PF candidates. This difference is well accounted for in the simulation, leading to the conclusion that any significant



(a) η_{j1} - MTR (b) η_{j2} - MTR

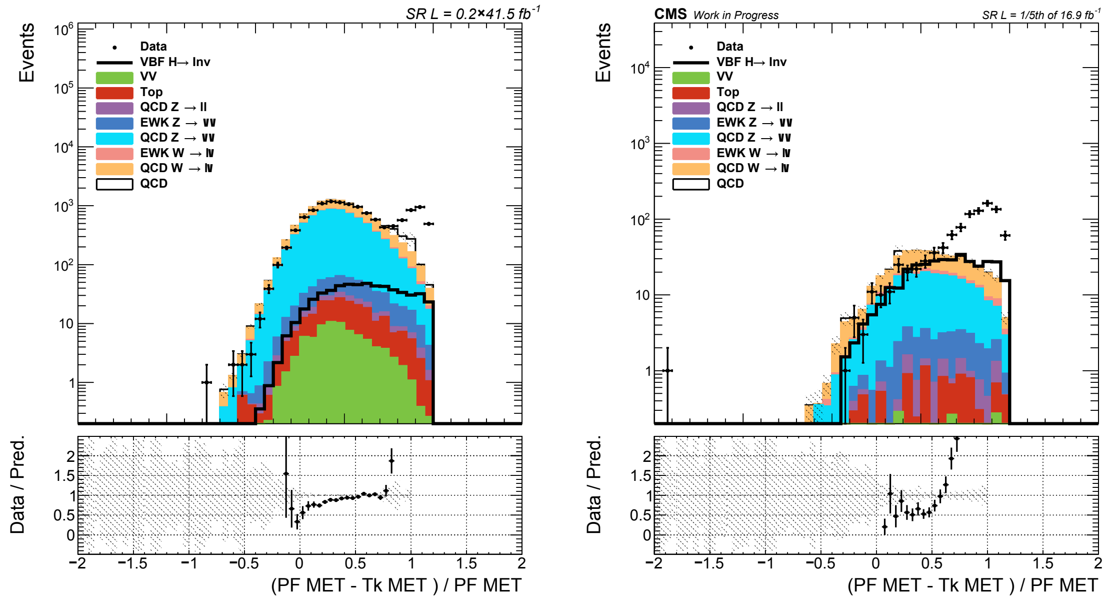


(c) η_{j1} - VTR (d) η_{j2} - VTR

Figure 7.13: Distributions of η_{j1} (left) and η_{j2} (right) variables in the signal region after the unbinding of 1/5th of the 2017 data. Both MTR (top) and VTR (bottom) categories are presented.

deviations from the expected behaviour can be associated largely to these low quality jets.

Figure 7.14 shows distributions of the relative difference between the tracker ($E_{T,miss}^{track}$) and the standard $E_{T,miss}$. It can be seen that a large deviation from the expected simulation behaviour is observed in the region where this difference is larger than 0.8. Using the aforementioned region in order to contain the jet behavior, a requirement that a jet within the $2.8 < |\eta| < 3.2$ region has to have value of 1-$E_{T,miss}^{track}/E_{T,miss} < 0.8$ was introduced into the existing set of analysis requirements for both categories.



(a) 1-$E_{T,miss}^{track}/E_{T,miss}$ - MTR (b) 1-$E_{T,miss}^{track}/E_{T,miss}$ - VTR

Figure 7.14: Distributions of the 1-$E_{T,miss}^{track}/E_{T,miss}$ variable in the signal region after the unbinding of 1/5th of the 2017 data. Both MTR (a) and VTR (b) categories are represented.

Figures 7.15 and 7.16 show the distribution of jet eta (for both the leading and subleading jets) and the 1-$E_{T,miss}^{track}/E_{T,miss}$ relation after the inclusion of the aforementioned requirement, showcasing the mitigation power of this choice which led to a significantly diminished effect of jet ”horns”. All of the presented distributions used to illustrate this effect are related to the 2017 era, while the equivalent set representing the 2018 era is given in Appendix A.3. This mitigation process led to a small loss of signal efficiency. Computed using the simulated samples of signal processes, it impacted the VBF production by ∼8 % and gluon-fusion by ∼1% of signal efficiency loss, leading to a combined value of 4%.

Lastly, the unblinding process for this analysis has brought an observation of an excess in the high m_{jj} values. In order to investigate this behaviour a set studies was performed leading to the conclusion that this excess was connected to two main sources: detector noise in the endcap region (mostly affecting the 2017 data) and a localised excess for the scenario where both of the jets are within the HF η area[7]. As the treatment of the HF-HF jet scenario requires more studies, it has been removed from the analysis for the purposes of this thesis. In order to battle the resulting detector noise, an additional set of quality criteria was added. Similarly to the previously introduced jet horns treatment, they revolve around the $E_{T,miss}$ variables:

$$\Delta\phi_{E_{T,miss}} = \Delta\phi(E_{T,miss}^{track}, E_{T,miss}), \tag{7.4}$$

$$B = \frac{|\vec{H}_{T,miss} - \vec{p}_{T,miss}|}{|\vec{H}_{T,miss} + \vec{p}_{T,miss}|}, \tag{7.5}$$

$$\Delta\phi B = \sqrt{(X \cdot B)^2 + \Delta\phi^2_{E_{T,miss}}}, \tag{7.6}$$

where the X takes the value od 3.3 (6.6) for the MTR (VTR) category[8]. The final addition to the selection requirements is the $B < 0.2$ (for the scenarios where one jet is in the HF region) and $\Delta\phi B < 1$ (1.3) for the rest.

7.4.2 Missing HE sectors

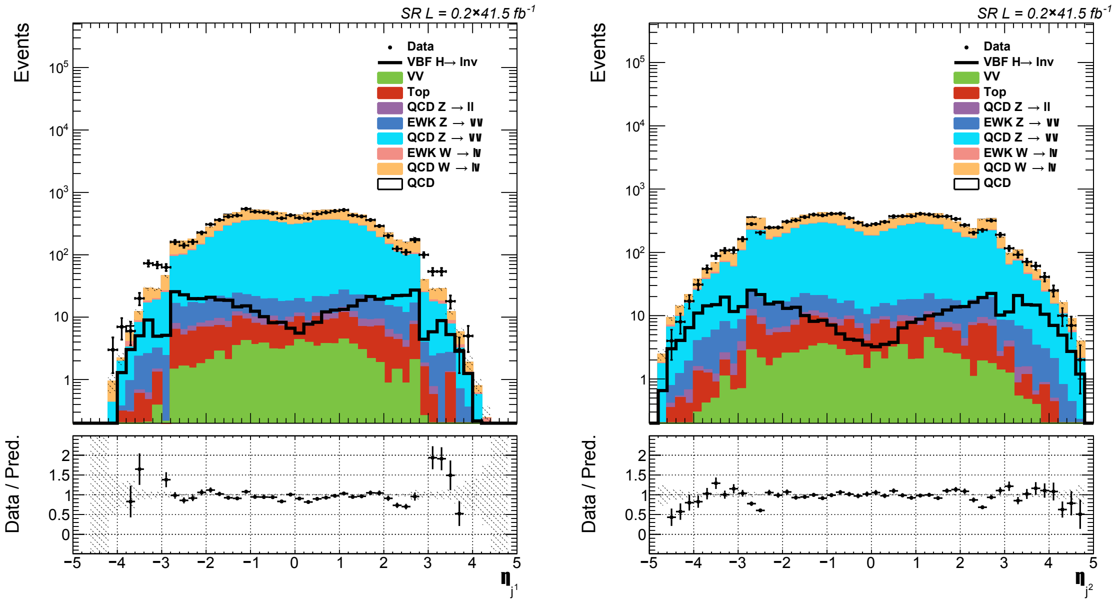
The incident leaving a large amount of data collected during 2018 without any information about HCAL sectors 15/16 was the main motivation driving these, partial unblinding, studies. Being interested in forward jets, this analysis was especially affected in two regions, the signal and the electron regions. The latter is discussed in more detail in Section 8.2.4. The reason why this is problematic is connected with the position of the subdetectors, The affected region, observed from the point of view of the jet η is -2.5< η <-1.4, thus marking an area which has a higher probability of jets being mistaken for electrons.

Another potential problem occurs if the PF candidate in the problematic range fails to be associated with the ECAL cluster of tracks. Following that there is no HCAL information in this HE-minus (HEM) region, this can be seen in the appearance of mismeasured $E_{T,miss}$ populating the ϕ area covered by the HEM region. In order to test the SR for problems associated with the HEM region, the distribution of the ϕ variable associated with the $\vec{p}_{T,miss}$ has been studied. Figure 7.17 shows the data to simulation comparison of this variable for both the MTR and VTR categories, indicating a significant discrepancy in the -1.8 < ϕ < -0.6 region. Figure 7.18 shows the appearance of a similar discrepancy for the leading/subleading jet ϕ in

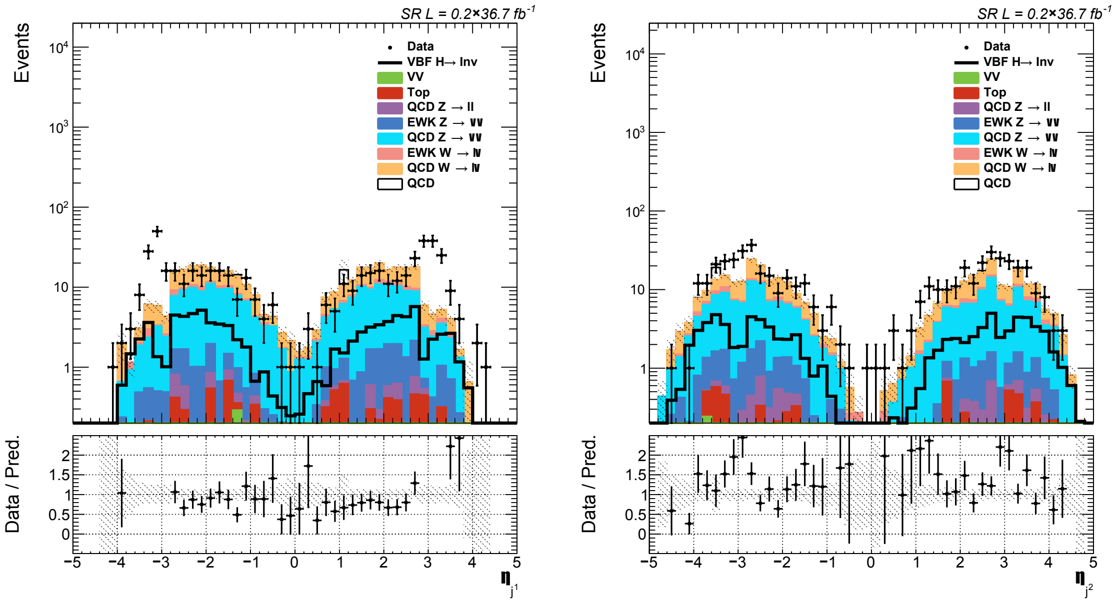
---

[7] Translating into jet $|\eta| > 3.0$.

[8] This constant term was introduced in order to have the option of a circular requirement by controlling the $\Delta\phi B$. Additionally, a set of jet quality cuts is also applied on the leading pair.

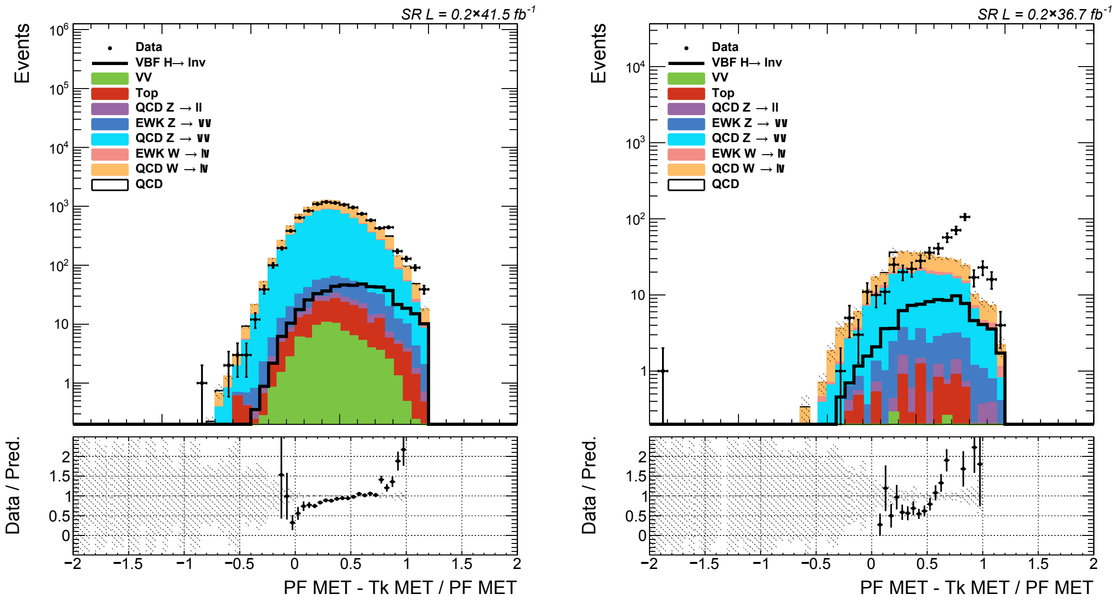


(a) η_{j1} - MTR (b) η_{j2} - MTR



(c) η_{j1} - VTR (d) η_{j2} - VTR

Figure 7.15: Distributions of η_{j1} (left) and η_{j2} (right) variables in the signal region after the unbinding of 1/5th of the 2017 data and the mitigation of the jet ”horns” effect. Both MTR (top) and VTR (bottom) categories are presented.

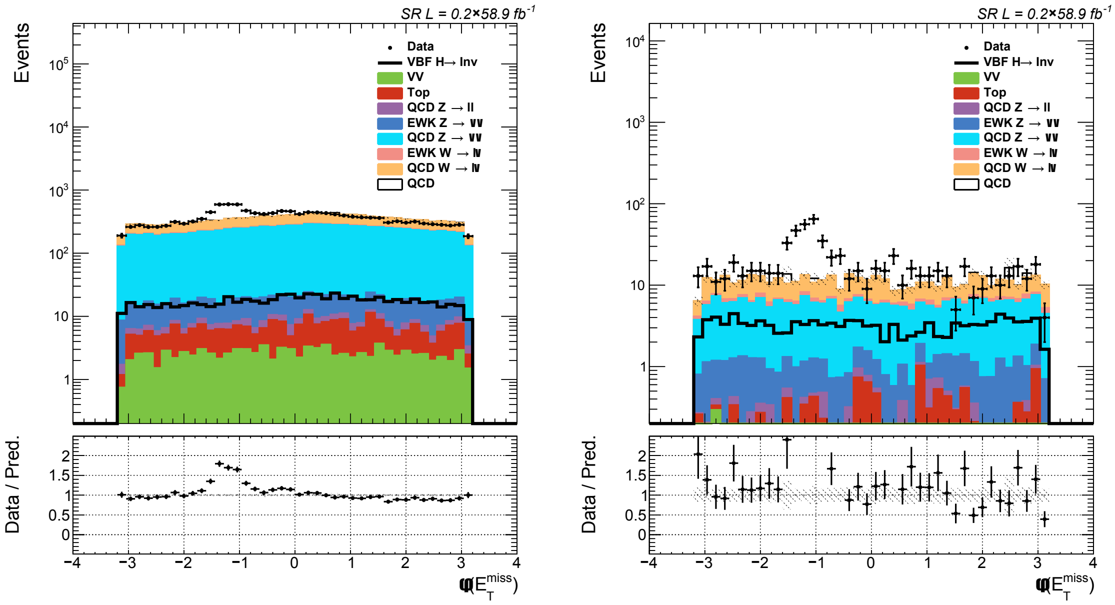


(a) 1-$E_{T,miss}^{track}/E_{T,miss}$ - MTR (b) 1-$E_{T,miss}^{track}/E_{T,miss}$ - VTR

Figure 7.16: Distributions of the 1-$E_{T,miss}^{track}/E_{T,miss}$ variable in the signal region after the unbinding of 1/5th of the 2017 data and the mitigation of the jet "horns" effect. Both MTR (a) and VTR (b) categories are represented.

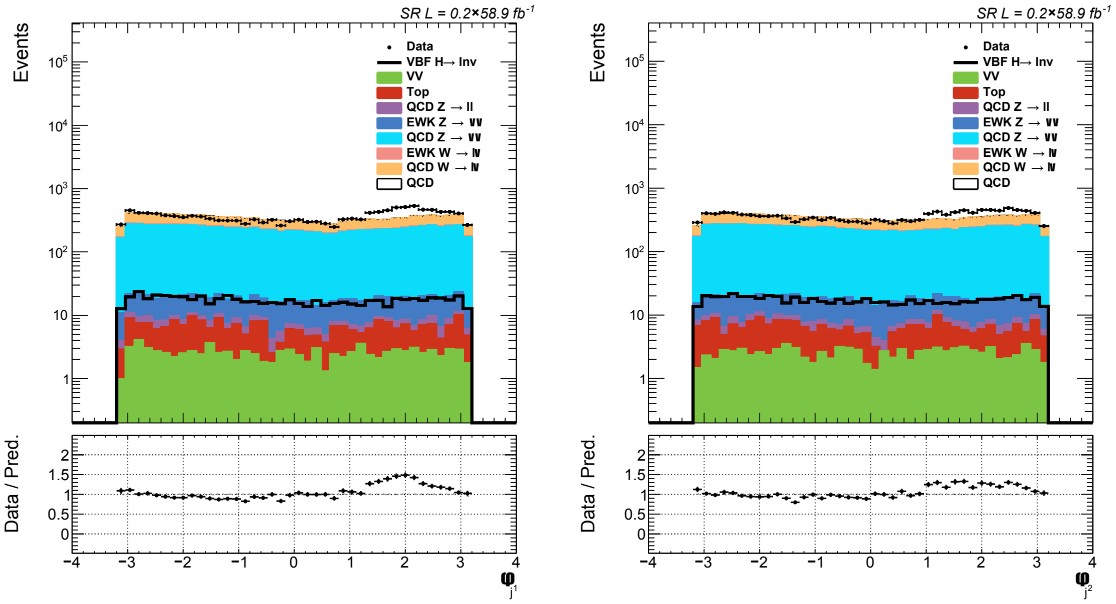
the opposite side in ϕ, as influenced by the signal selection.

Upon performing a selection of studies with different jet quality requirements [98], the simplest solution was chosen for the end result. The whole ϕ region of the $\vec{p}_{T,miss}$ affected by the HEM problem was rejected from the selection requirements for the SR. Figure 7.19 shows distributions of leading and subleading jet ϕ after the application of the aforementioned veto. This has helped to return the overall data to simulation agreement to the expected ranges, but it also brought in a loss of sensitivity with respect to the Br(H→inv) of ∼11 %. The resulting veto was added to the selection requirements for the 2018 era for both categories.

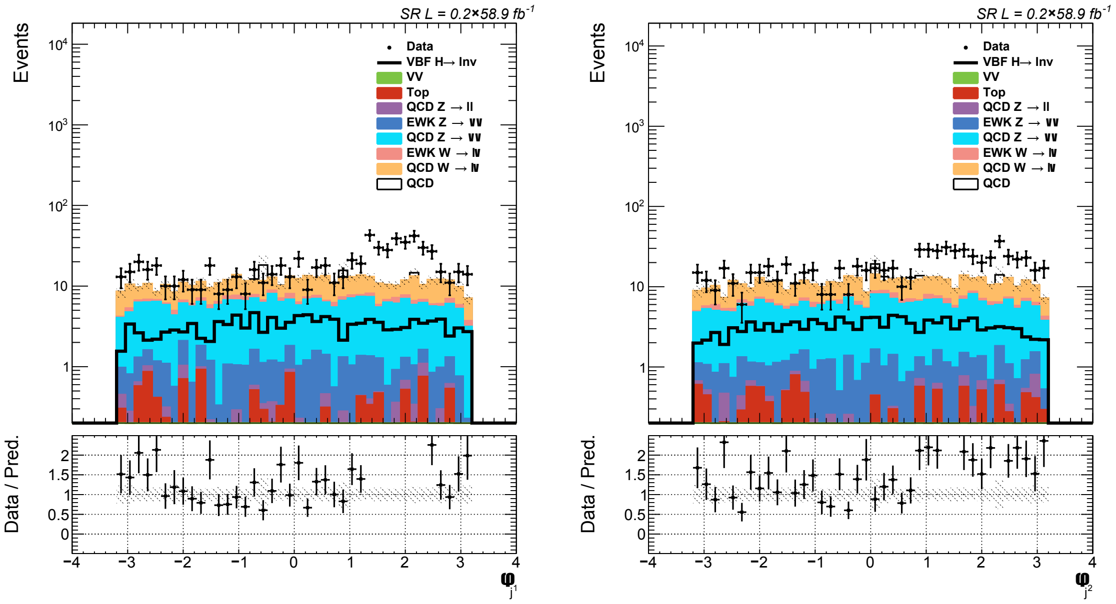


(a) $E_{T,miss}$ - MTR (b) $E_{T,miss}$ - VTR

Figure 7.17: Distributions of the $E_{T,miss}$ variable for the signal region presenting effects of the HEM problem for MTR (a) and VTR (b) categories.

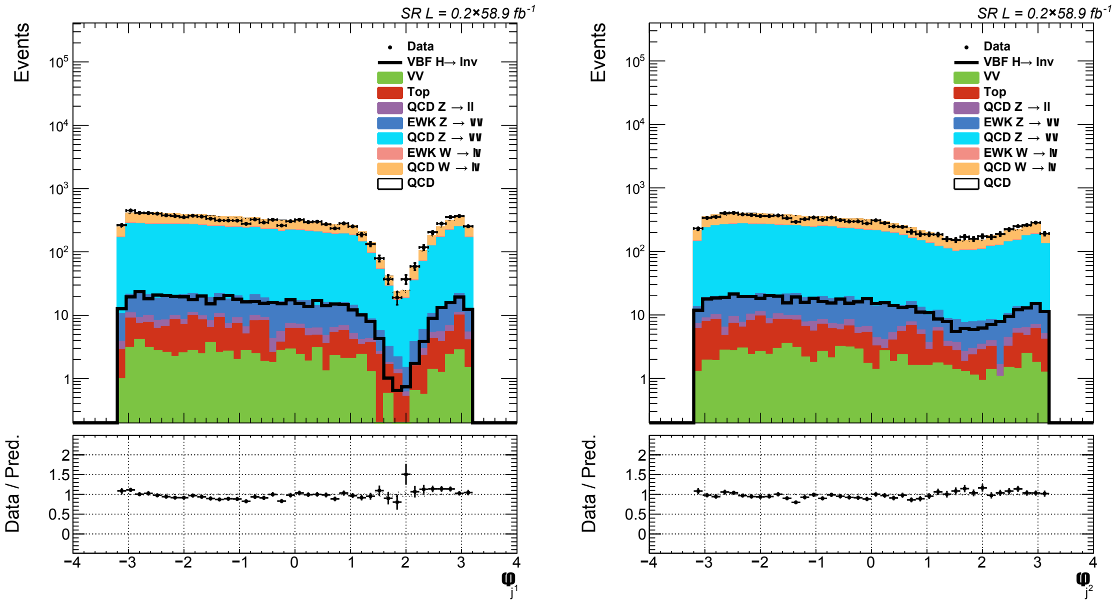


(a) ϕ_{j1} - MTR (b) ϕ_{j2} - MTR

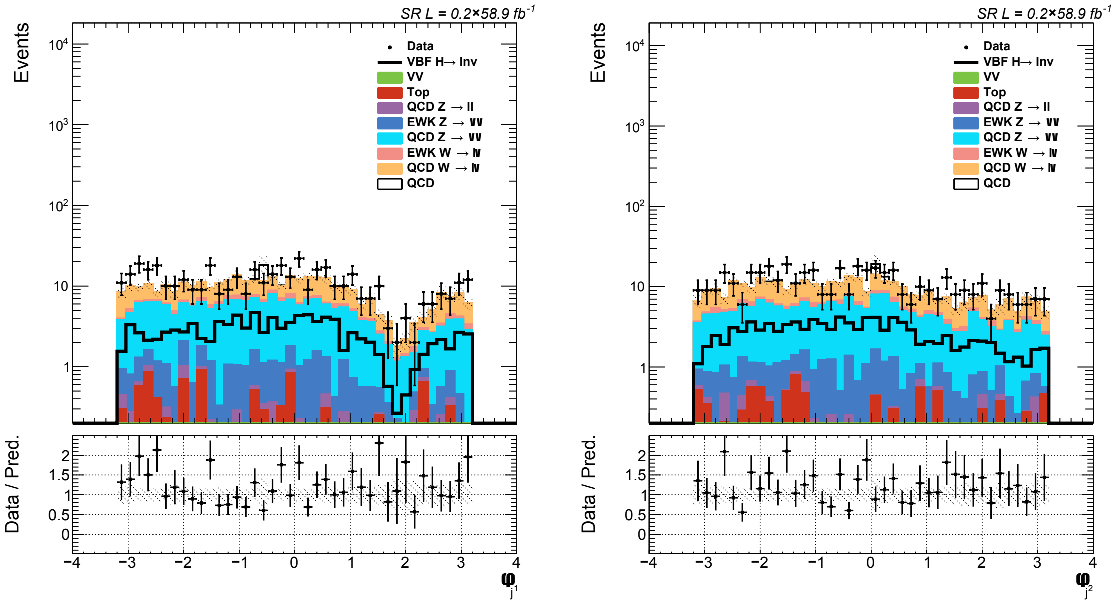


(c) ϕ_{j1} - VTR (d) ϕ_{j2} - VTR

Figure 7.18: Distributions of ϕ_{j1} (left) and ϕ_{j2} (right) variables in the signal region, after the unbinding of 1/5th of the 2018 data, showing the effect of the HEM problem. Both MTR (top) and VTR (bottom) categories are presented.



(a) ϕ_{j1} - MTR (b) ϕ_{j2} - MTR



(c) ϕ_{j1} - VTR (d) ϕ_{j2} - VTR

Figure 7.19: Distributions of ϕ_{j1} (left) and ϕ_{j2} (right) variables in the signal region, after the unbinding of 1/5th of the 2018 data and following the mitigation of the HEM problem. Both MTR (top) and VTR (bottom) categories are presented.

Chapter 8

Dedicated control regions

> *"Noise proves nothing. Often a hen who has merely laid an egg cackles as if she laid an asteroid."*
>
> — Mark Twain

8.1 Introduction

THE APPROACH TAKEN for the estimation of the main sources of backgrounds, originating from V+jets SM processes, is to use a set of dedicated control regions (CR) dominated by each of these processes. The irreducible contribution of these Z($\nu\nu$)+jets and W($l\nu$)+jets (where the charged lepton is unidentified by the detector) SM backgrounds is estimated using a set of well identified lepton control regions that are associated with the same dijet properties as the ones used for the definition of both SR categories.

Regions are selected to be orthogonal to each other, bearing similarities to the SR in order to ensure a smooth transition between the information obtained from the CR and the final estimation in the SR. The advantage of this approach is the usage of $E_{T,miss}$ where the leptons have been removed from the computation (as previously defined in Section 7.3.1), which ensures the preservation of the VBF-like selection. Following sections further explain the definition of each of these regions.

In addition to the main V+jets backgrounds, there is one more SM source of background that requires special attention. The QCD multijet processes present a problem due to the lack of statistical precision in their corresponding simulation samples, requiring an estimation from regions enriched with contributions from these processes. The formation of the of these dedicated QCD enriched regions and the methodology behind their usage in the final extrapolation to the SR are also the focus of few of the following pages.

8.2 Lepton regions

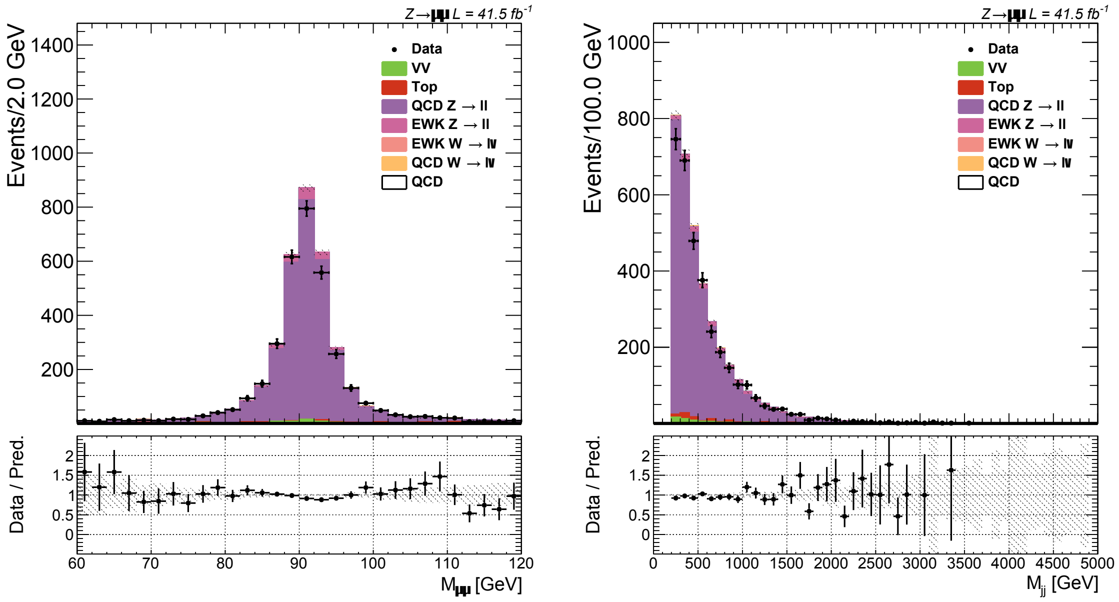
Focusing on regions dedicated to the estimation of V+jets influence on the SR, the upcoming sections are going to introduce the four main lepton CRs, which can be grouped into two categories based on their targeted processes: the double and single lepton CRs. The idea behind the double lepton regions is to provide a good tool when tackling the Z($\nu\nu$)+jets processes, while the single lepton regions take over the responsibilities associated with the W($l\nu$)+jets processes. Their respective definitions, focusing on further separation using lepton flavour, are going to be summarised in the following sections.

8.2.1 Double muon CR

As described in the introduction, the double muon CR is used to estimate the Z($\nu\nu$)+jets background. This region is formed by replacing the muon veto selection requirement from VTR and MTR selections (Tables 7.1 and 7.2) with a requirement that there are exactly two muon objects in the event. At least one of these two objects must pass the tight requirements defined in Section 6.2.2. Additional selection requirements are imposed on muon p_T with thresholds being set at p_T >20/10 GeV for the leading/subleading muon respectively (connected to the tight muon requirements). The final item is related to the formation of a dilepton mass region around the value of the Z boson mass, imposing a $60 < m_{ll} < 120$ GeV requirement.

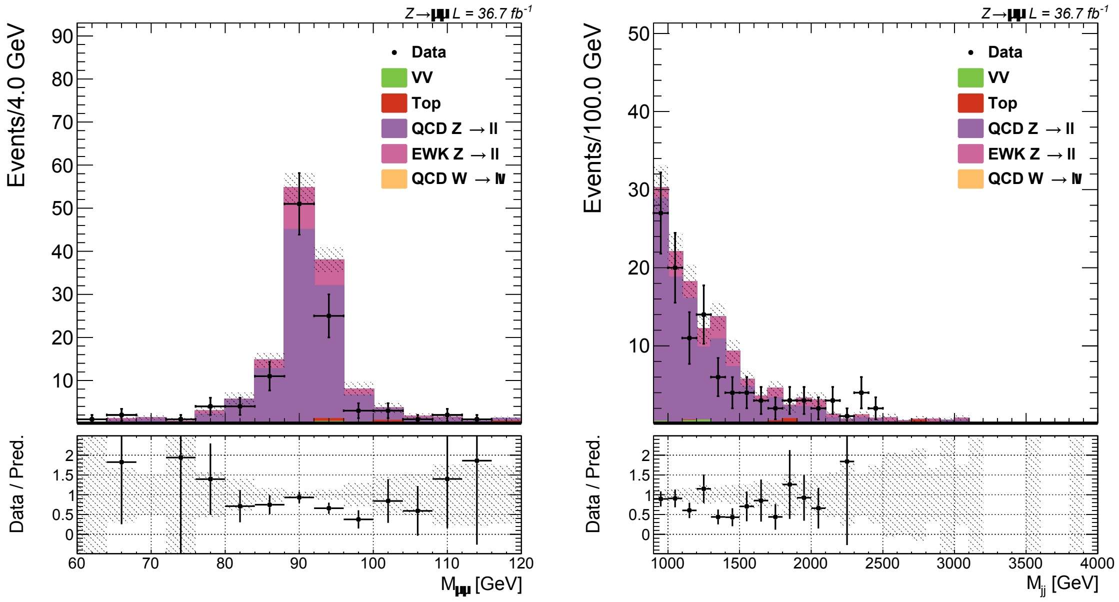
Figures 8.1 and 8.2 show the distributions of invariant masses of dilepton and dijet objects for both the MTR and VTR selection for both 2017 and 2018 eras. Overall good data to simulation agreement is observed for both variables in all categories, with the agreement being within the values controlled by the final fit. The much lower number of events in the VTR category originates from a set of higher p_T thresholds included in VTR selection requirements.

Figures 8.3 and 8.4 are showing the $min\Delta\phi(j, E_{T,miss}^{no,\mu})$ and $E_{T,miss}^{no,\mu}$ variables for both categories for the 2017 and 2018 eras of data taking, respectively. The inclusion of tight muon requirements as well as the constraint applied to the dimuon mass has helped with removing contributions from additional processes.



(a) m_{ll} - MTR

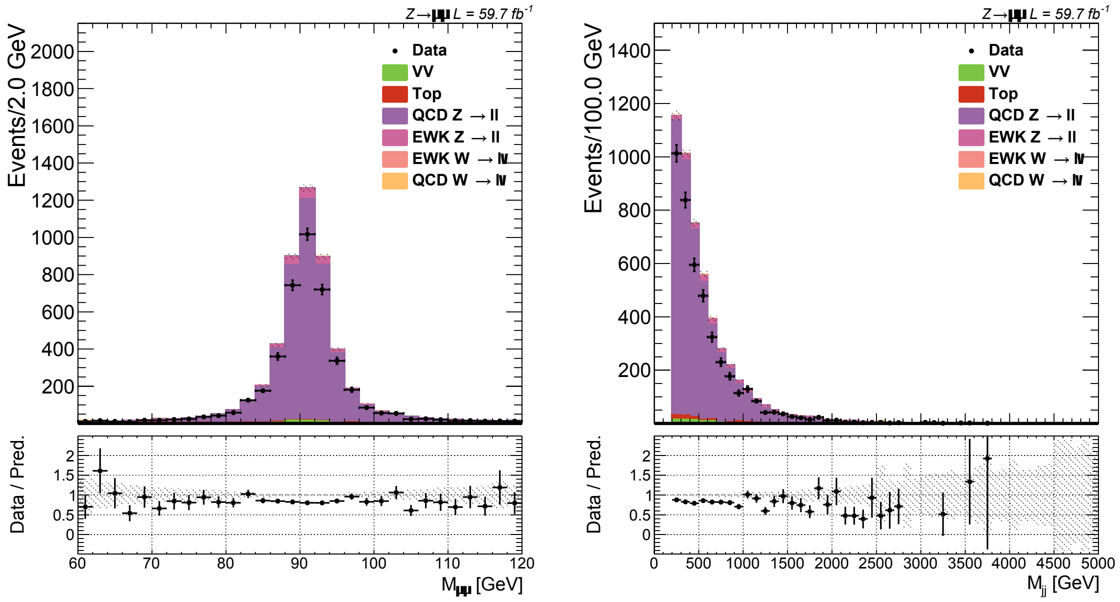
(b) m_{jj} - MTR



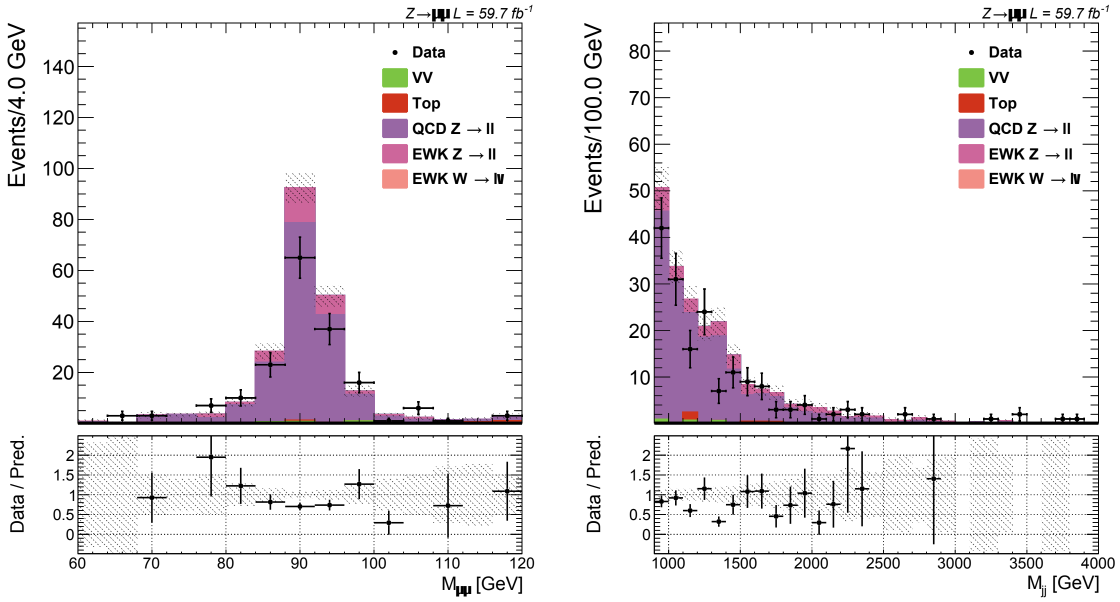
(c) m_{ll} - VTR

(d) m_{jj} - VTR

Figure 8.1: Distributions of m_{ll} and m_{jj} variables in the double muon region for MTR (top) and VTR (bottom) categories for the 2017 era of data taking.

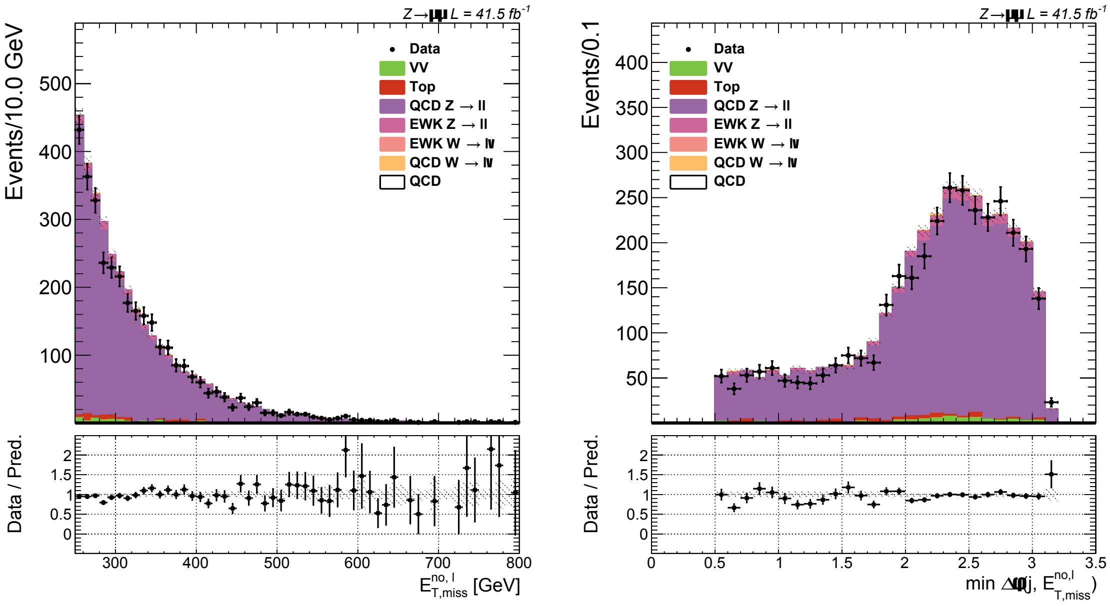


(a) m_{ll} - MTR (b) m_{jj} - MTR

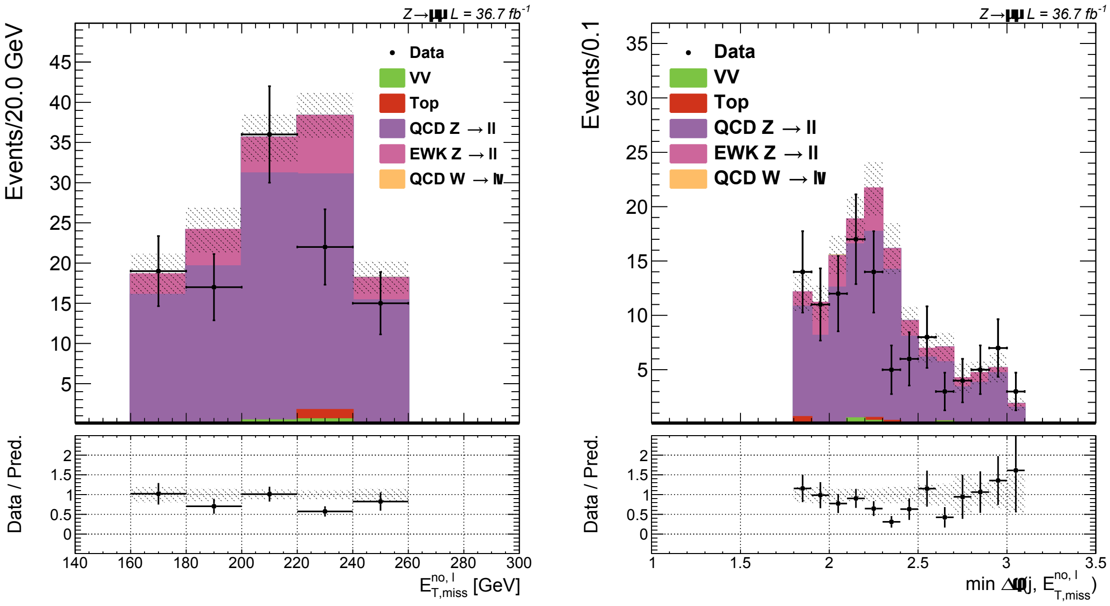


(c) m_{ll} - VTR (d) m_{jj} - VTR

Figure 8.2: Distributions of m_{ll} and m_{jj} variables in the double muon region for MTR (top) and VTR (bottom) categories for the 2018 era of data taking.

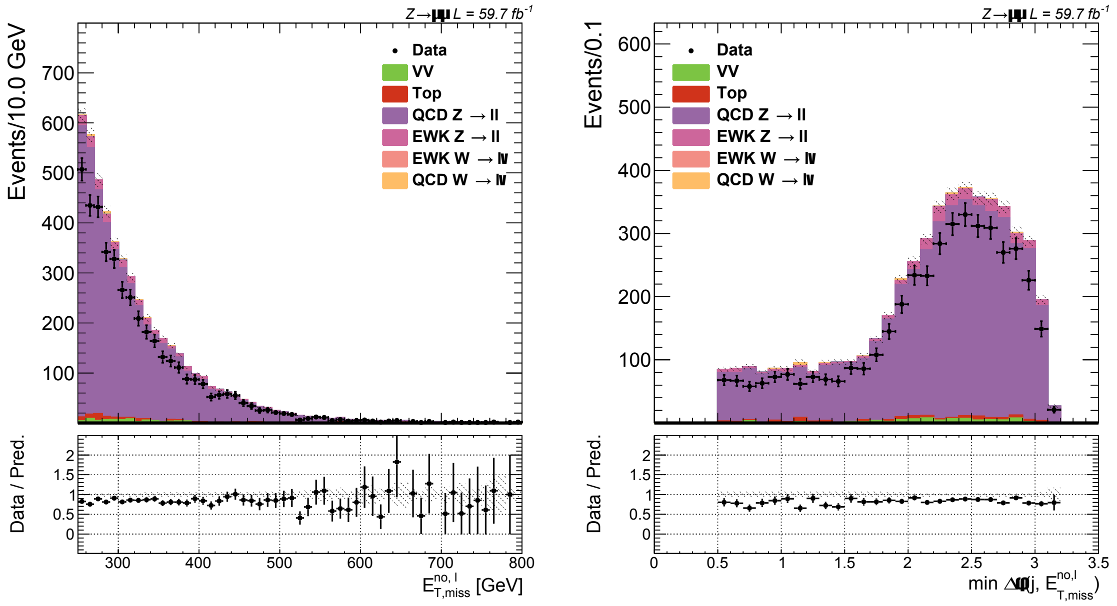


(a) $E_{T,miss}^{no,\mu}$ - MTR (b) $min\Delta\phi(j, E_{T,miss}^{no,\mu})$ - MTR

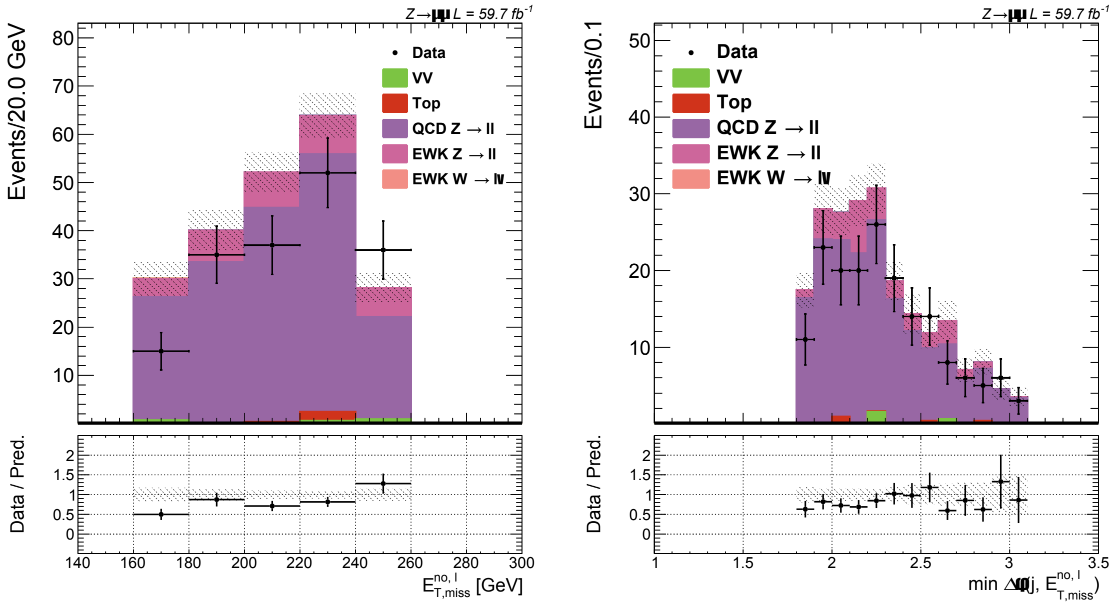


(c) $E_{T,miss}^{no,\mu}$ - VTR (d) $min\Delta\phi(j, E_{T,miss}^{no,\mu})$ - VTR

Figure 8.3: Distributions of the $E_{T,miss}^{no,\mu}$ and $min\Delta\phi(j, E_{T,miss}^{no,\mu})$ variables in the double muon region for MTR (top) and VTR (bottom) categories for the 2017 era of data taking.



(a) $E_{T,miss}^{no,\mu}$ - MTR (b) $min\Delta\phi(j, E_{T,miss}^{no,\mu})$ - MTR



(c) $E_{T,miss}^{no,\mu}$ - VTR (d) $min\Delta\phi(j, E_{T,miss}^{no,\mu})$ - VTR

Figure 8.4: Distributions of the $E_{T,miss}^{no,\mu}$ and $min\Delta\phi(j, E_{T,miss}^{no,\mu})$ variables in the double muon region for MTR (top) and VTR (bottom) categories for the 2018 era of data taking.

This is evident in 2018, where there is a complete absence of any excess originating from the HEM problem, which can be tested by taking a look at ϕ variables for $\vec{p}_{T,miss}$ (shown in Figure A.6 for the MTR category). The influence of QCD multijet processes is also diminished by these conditions for both eras. Additional distributions showing leading and muon (p_T/η) properties are presented in Appendix A.4.

8.2.2 Single muon CR

Continuing with the muon structures, the next item is the single muon region. The CR is formed from VBF-like events, similarly to double muon regions, by modifying the muon veto requirement for both MTR/VTR selections. The new requirement states that the event needs to contain exactly one muon with $p_T > 20$ GeV, which also satisfies tight muon requirements.

Figures 8.5 and 8.6 show distributions of the m_{jj} and $E_{T,miss}^{no\ \mu}$ in this region for both categories and both eras. The data to prediction agreement for the main variable considered in the fit is very good for the low m_{jj} bins, with the disagreement in the higher bins being significantly reduced with a choice of wider bins when performing the fit (more details about the fit procedure are given in Chapter 9).

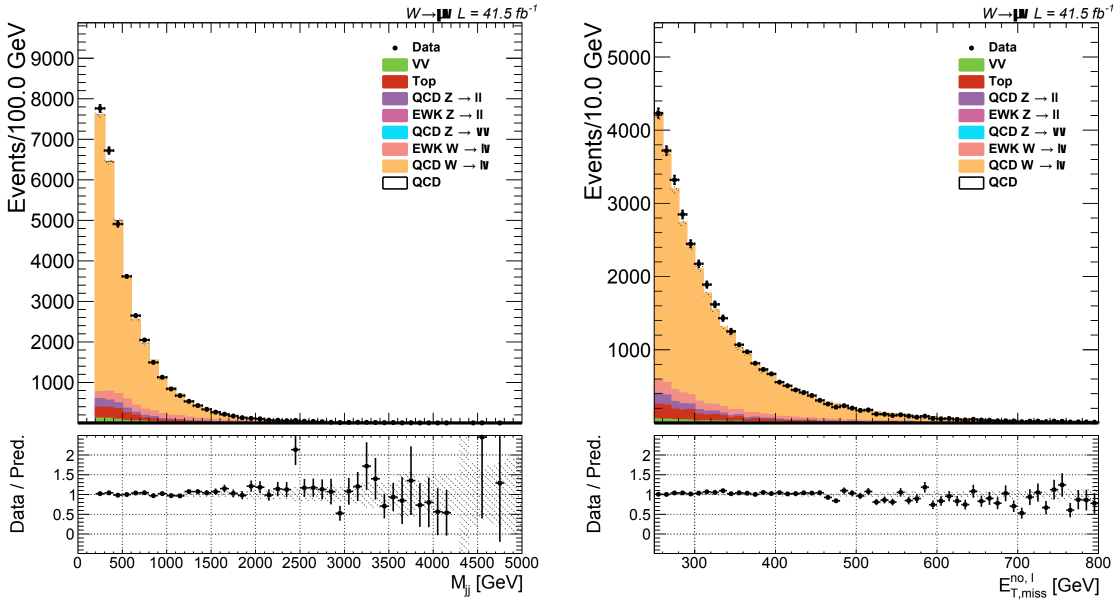
For these single lepton regions, a new variable of interest is introduced. The transverse mass of a two object system is defined as:

$$M_T = \sqrt{m_1^2 + m_2^2 + 2 \cdot (E_{T_1} E_{T_2} - \vec{p}_{T_1} \vec{p}_{T_2})}, \tag{8.1}$$

where the m_i, $E_{T,i}$ and p_{T_i} denote the mass, the transverse energy and the transverse momentum of a physics object. For the scenarios where $m_i \to 0$, the previous formula can be rewritten using the following approximation:

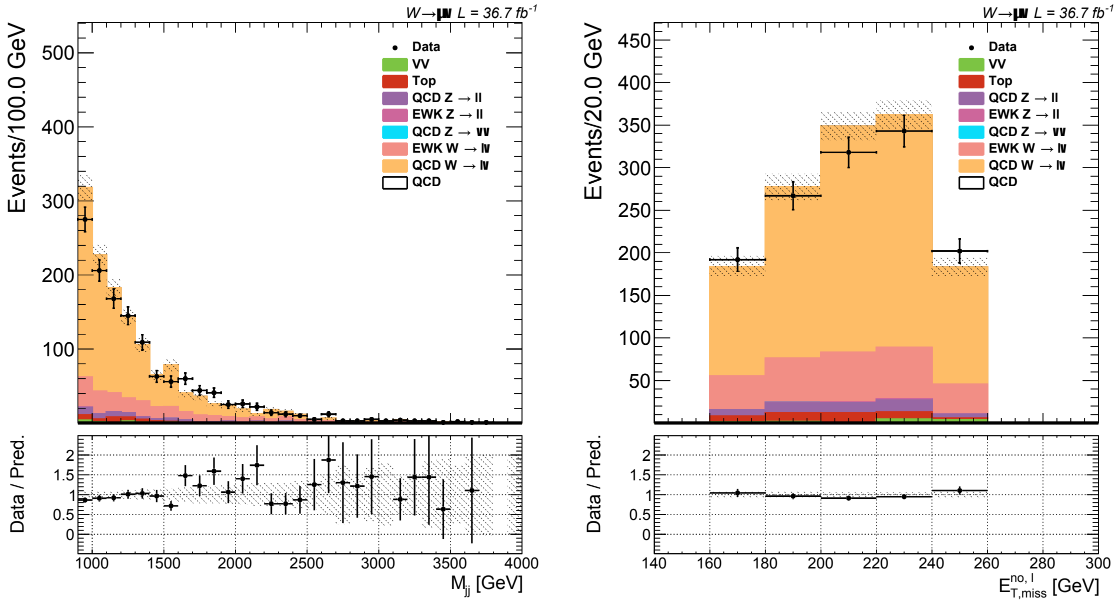
$$M_T = \sqrt{2 \cdot (E_{T_1} E_{T_2} (1 - cos\theta))}, \tag{8.2}$$

where θ represents the angle between two transverse momentum vectors. For the purposes of this region, two physics objects considered in the aforementioned calculation are going to be the $\vec{p}_{T,miss}^{no,\mu}$ and the transverse momentum of the selected muon. In the past, this variable (in further text referred to as $M_{T,\mu}$) proved useful, in W boson studies, for the control of the contribution originating from the QCD multijet processes. Presented alongside the $min\Delta\phi(j, E_{T,miss}^{no,\mu})$ variable in Figures 8.7 and 8.8, for 2017 and 2018 eras respectively, it shows good agreement between data and simulation and significantly reduced QCD multijet contribution (requiring no additional requirement for this region).



(a) m_{jj} - MTR

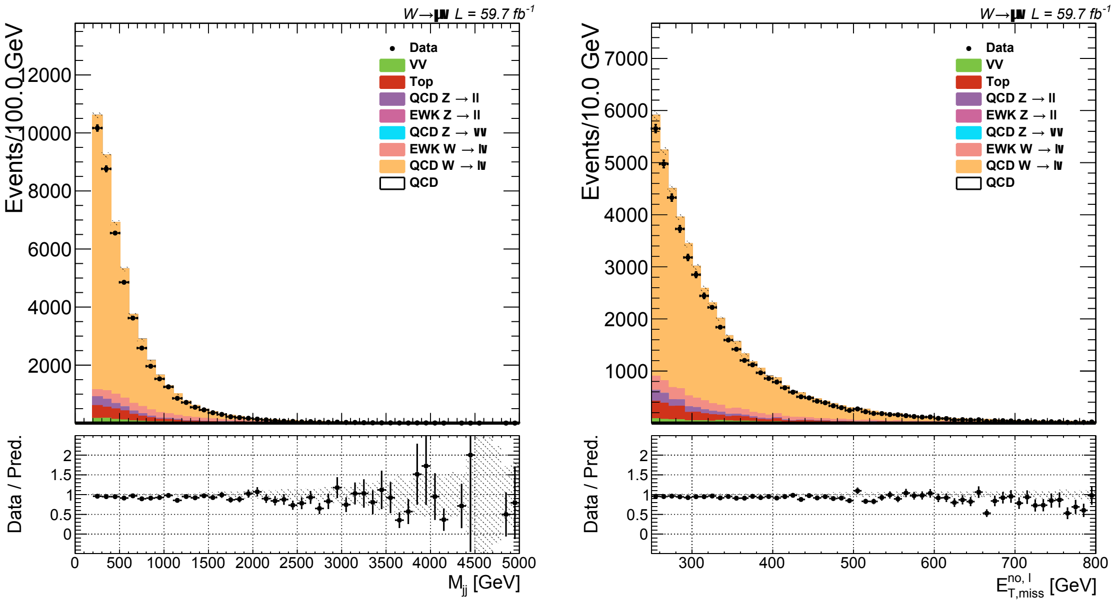
(b) $E_{T,miss}$ - MTR



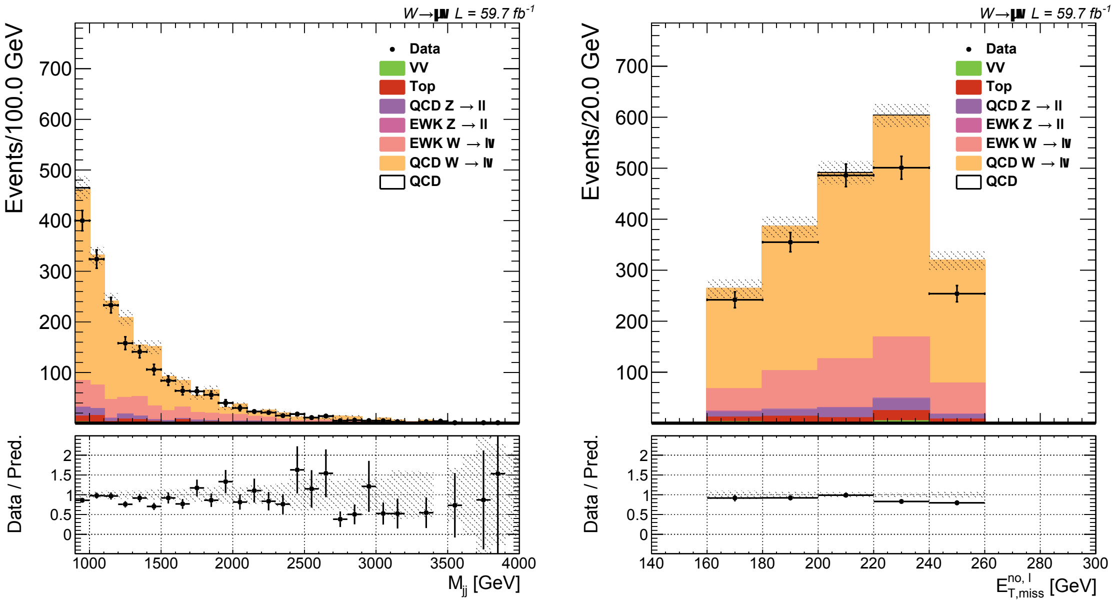
(c) m_{jj} - VTR

(d) $E_{T,miss}$ - VTR

Figure 8.5: Distributions of mjj and $E_{T,miss}$ variables in the single muon region for MTR (top) and VTR (bottom) categories for the 2017 era of data taking.

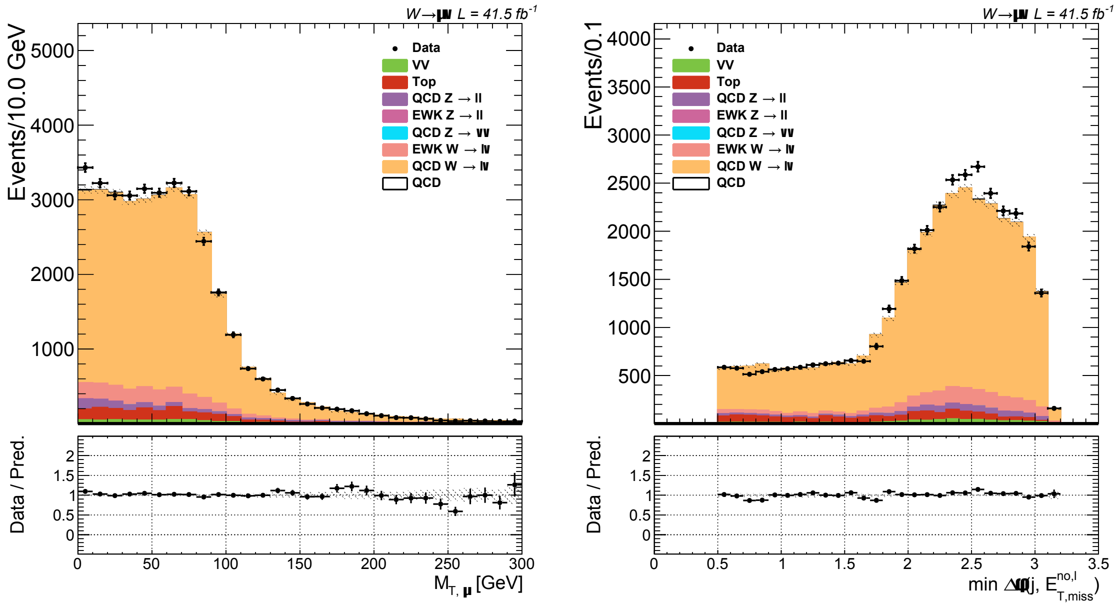


(a) m_{jj} - MTR (b) $E_{T,miss}$ - MTR

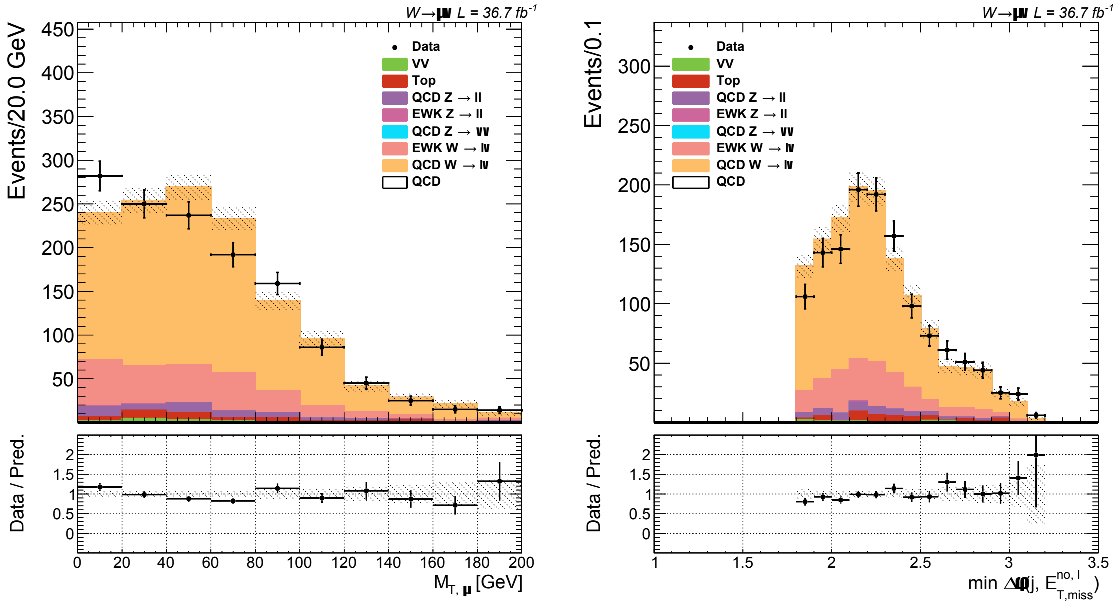


(c) m_{jj} - VTR (d) $E_{T,miss}$ - VTR

Figure 8.6: Distributions of mjj and $E_{T,miss}$ variables in the single muon region for MTR (top) and VTR (bottom) categories for the 2018 era of data taking.

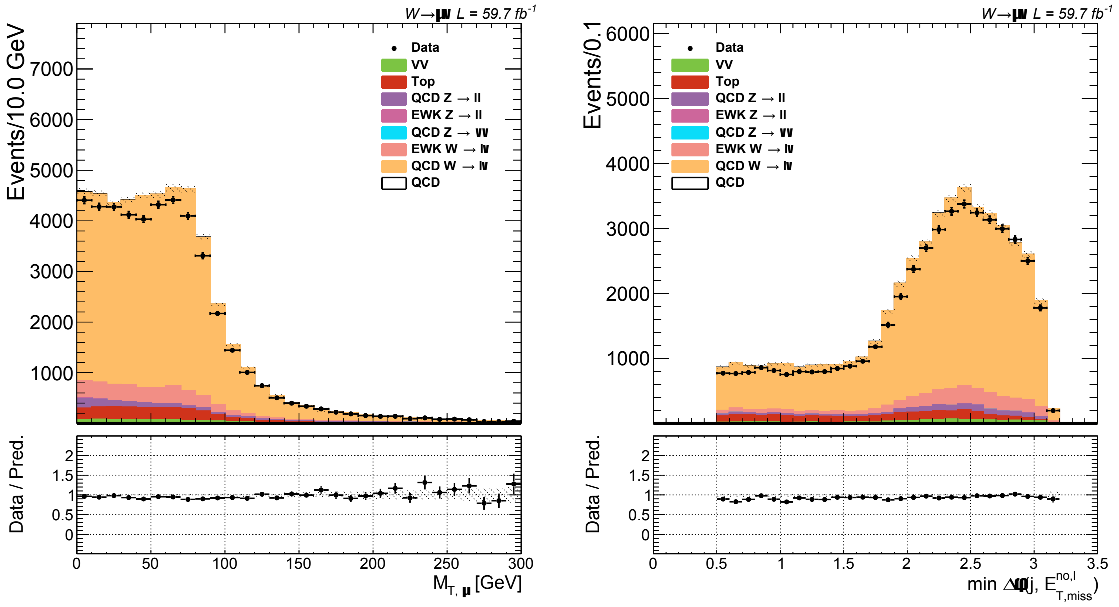


(a) $M_{T,\mu}$ - MTR (b) $min\Delta\phi(j, E_{T,miss}^{no,\mu})$- MTR



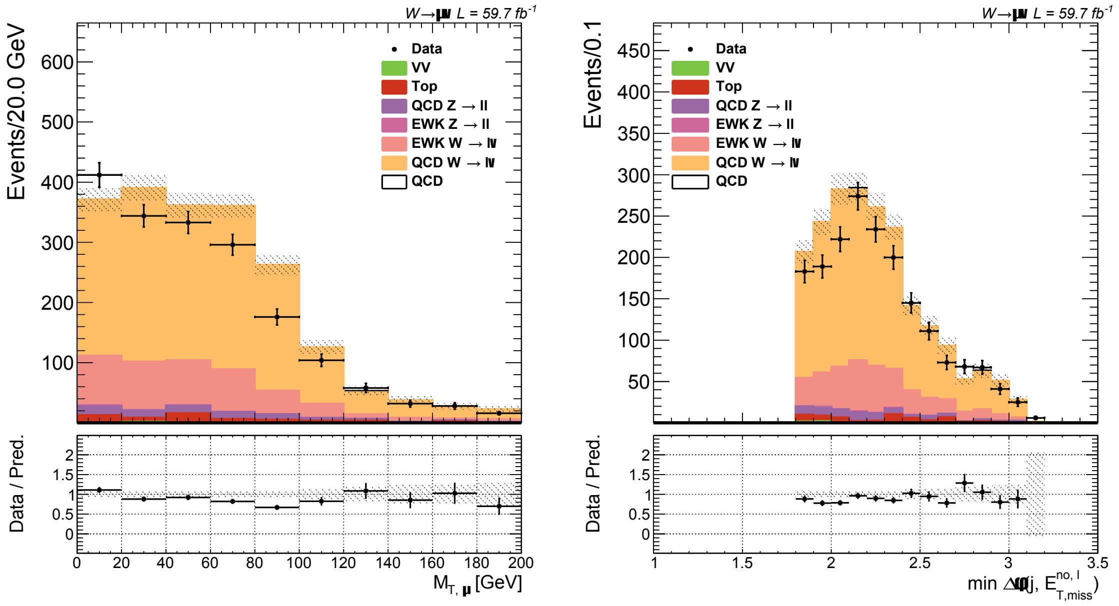
(c) $M_{T,\mu}$ - VTR (d) $min\Delta\phi(j, E_{T,miss}^{no,\mu})$ - VTR

Figure 8.7: Distributions of $M_{T,\mu}$ and $min\Delta\phi(j, E_{T,miss}^{no,\mu})$ variables in the single muon region for MTR (top) and VTR (bottom) categories for the 2017 era of data taking.



(a) $M_{T,\mu}$ - MTR

(b) $min\Delta\phi(j, E_{T,miss}^{no,\mu})$- MTR



(c) $M_{T,\mu}$ - VTR

(d) $min\Delta\phi(j, E_{T,miss}^{no,\mu})$ - VTR

Figure 8.8: Distributions of $M_{T,\mu}$ and $min\Delta\phi(j, E_{T,miss}^{no,\mu})$ variables in the single muon region for MTR (top) and VTR (bottom) categories for the 2018 era of data taking.

Similarly to the double muon region, the single muon region is not affected by the HEM issue in 2018 due to the tight muon requirement. Figure A.7 confirms this, showing no significant excess in the affected (η, ϕ) range for the MTR category. The following sections introduce electron regions where, for the single electron case, this effect will require special attention.

8.2.3 Double electron CR

The first step, when adapting the double lepton region structure for the purposes of the electron case, is to modify the electron veto from the SR selection requirements. It asks for exactly two electrons in the event, at least one of which needs to satisfy tight requirements, while both of them have to follow the $p_T > 40/10$ GeV thresholds for the leading/subleading electron, respectively. Upon selecting the objects, the dilepton mass requirement of $60 \ < m_{ll} < \ 120$ GeV is applied.

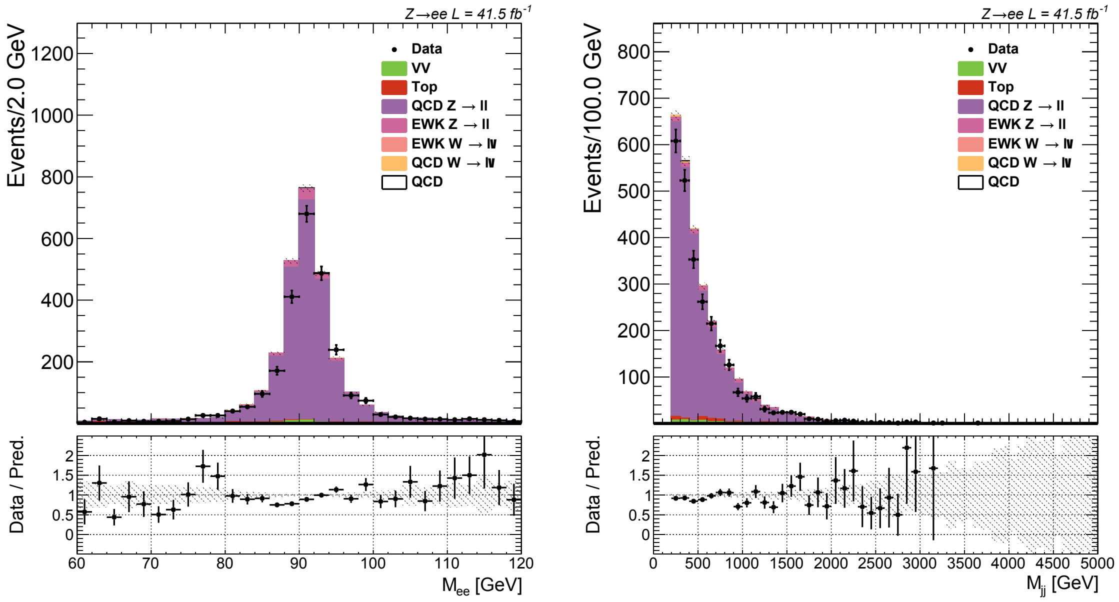
Figures 8.9 and 8.11 show the m_{ll} and m_{jj} distributions for both categories and both eras of data taking. The m_{jj} variable shows a good level of agreement between data and simulation. Additionally, Figures 8.11 and 8.12 show the data to prediction agreement for the $min\Delta\phi(j, E_{T,miss}^{no,e})$ and $E_{T,miss}^{no,e}$ (with the rest of the distributions being shown in Appendix A.4). The latter is defined in the same way as its muon counterpart, by eliminating the contribution from the electron objects when computing it.

As was the case for the double muon CR, this region is unaffected by the HEM problem in 2018 due to very tight requirements in the form of the electron identification and the dilepton mass. The previous reasoning also accounts for the lack of contribution originating from QCD multijet processes.

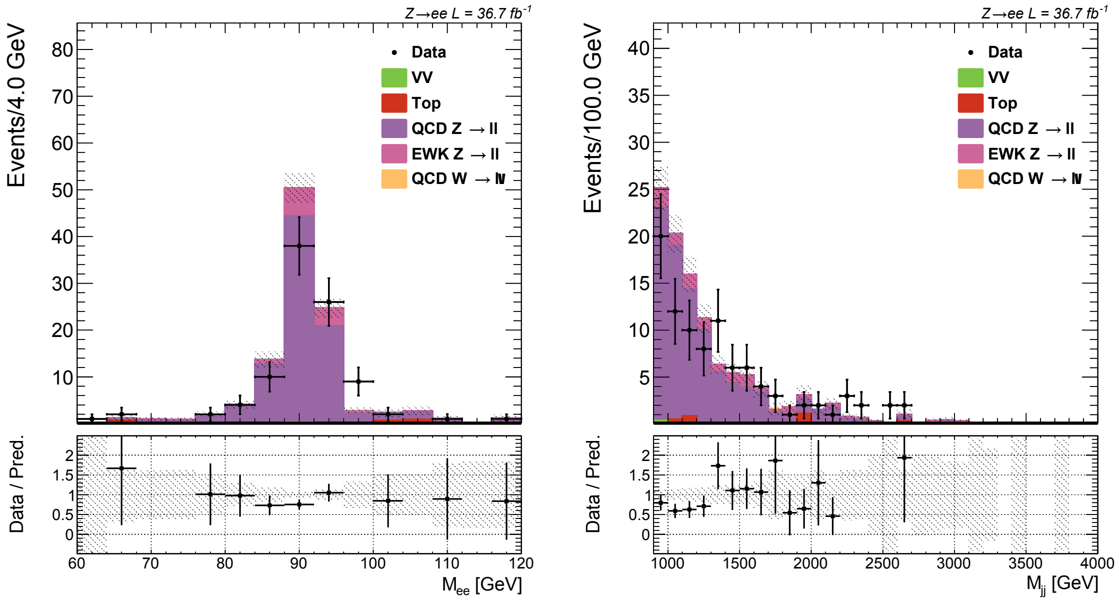
8.2.4 Single electron

There is one final lepton CR left to define, the single electron region, which proves to be the most interesting one. It requires the modification of the electron veto, in order to select events with only one electron object that passes the tight requirements and passes the p_T threshold of 40 GeV. In order to fight the contribution originating from QCD multijet processes a selection requirement of $E_{T,miss} > 50$ GeV was imposed.

Looking at the 2018 era, the HEM problem started affecting this CR as well as the SR. With the lack of a tight m_{ll} requirement, such as the Z mass window used for the dilepton CRs, an effect was expected to show in the electron η/ϕ distributions.

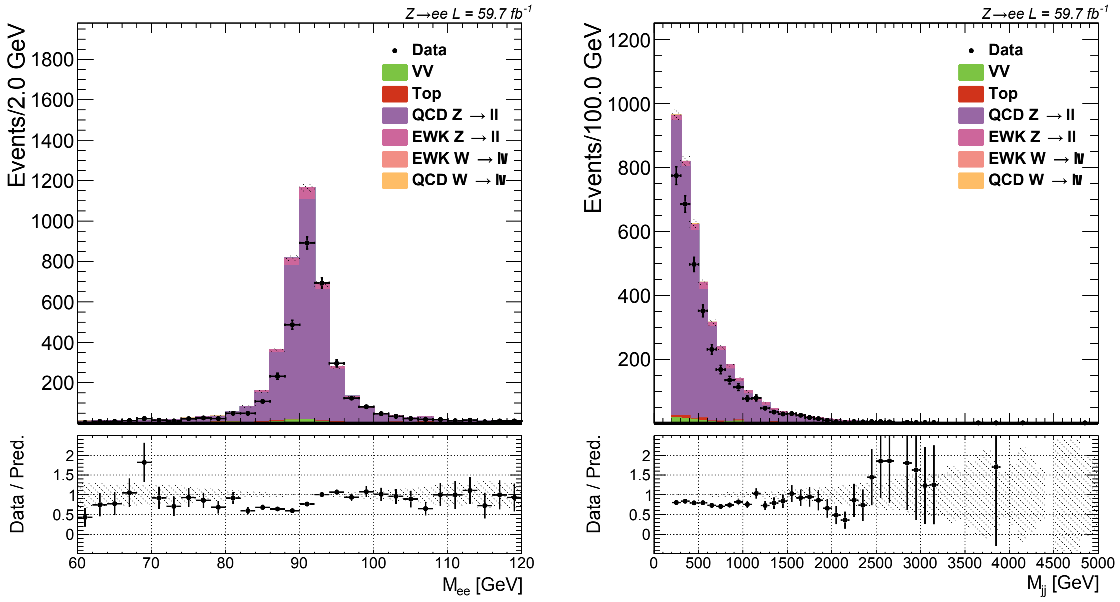


(a) m_{ll} - MTR (b) m_{jj} - MTR

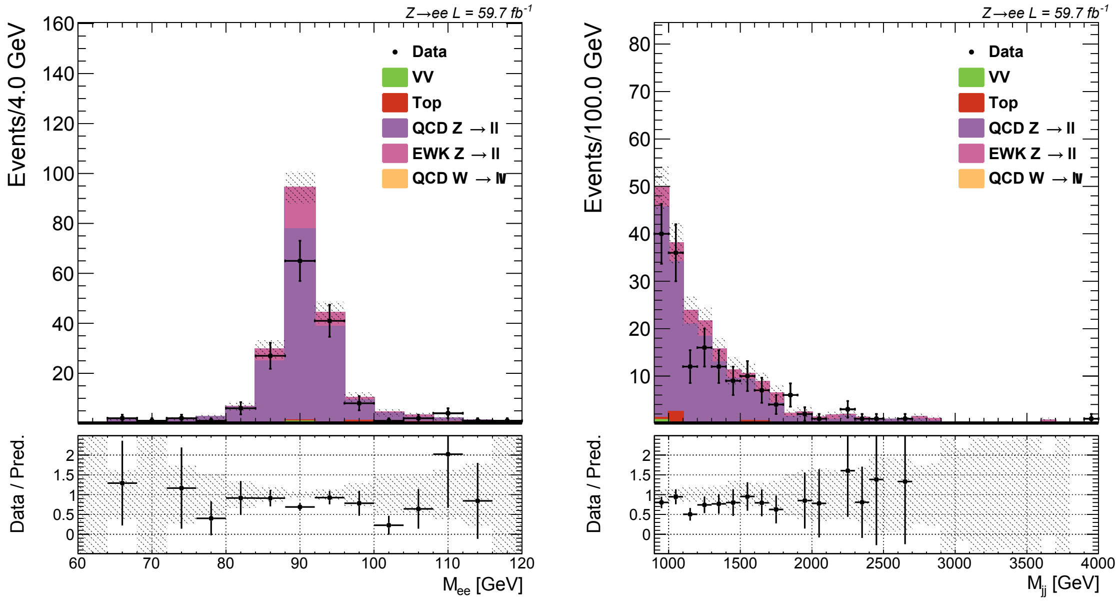


(c) m_{ll} - VTR (d) m_{jj} - VTR

Figure 8.9: Distributions of m_{ll} and m_{jj} variables in the double muon region for MTR (top) and VTR (bottom) categories for the 2017 era of data taking.

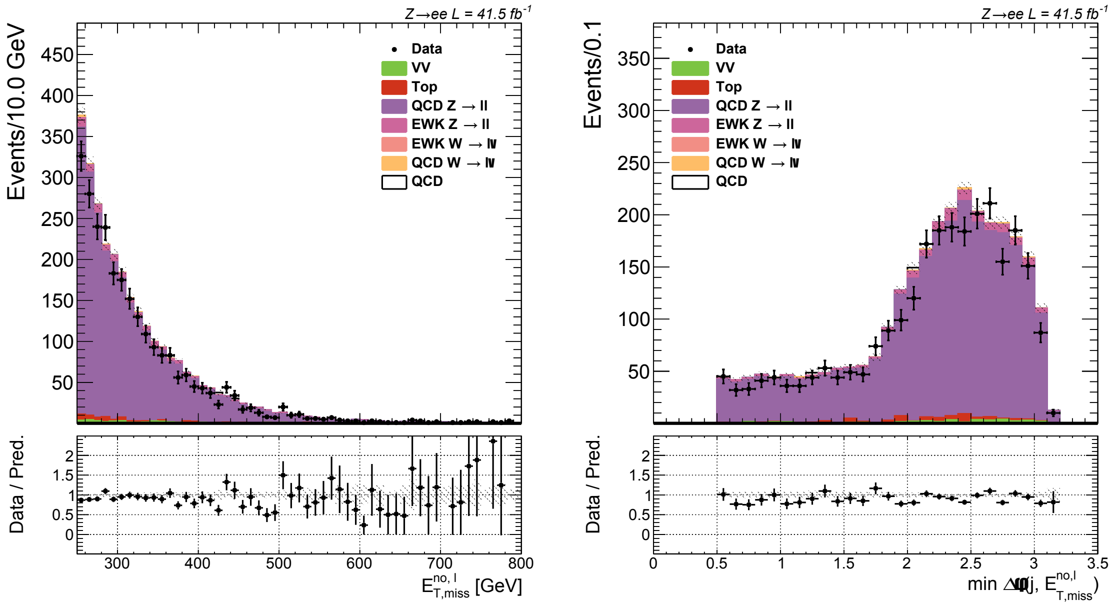


(a) m_{ll} - MTR (b) m_{jj} - MTR

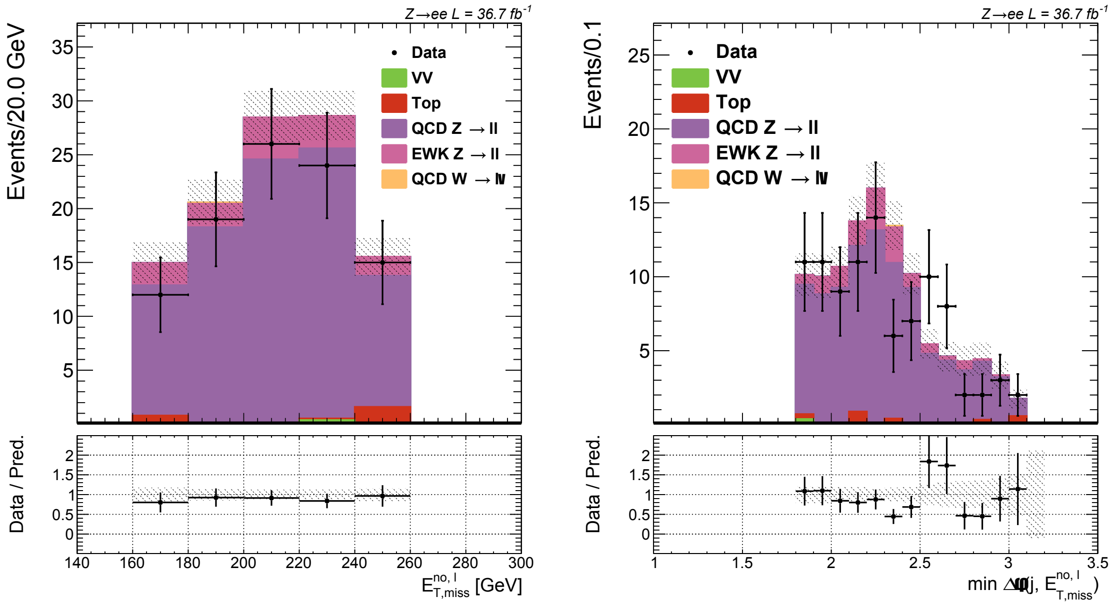


(c) m_{ll} - VTR (d) m_{jj} - VTR

Figure 8.10: Distributions of m_{ll} and m_{jj} variables in the double muon region for MTR (top) and VTR (bottom) categories for the 2018 era of data taking.

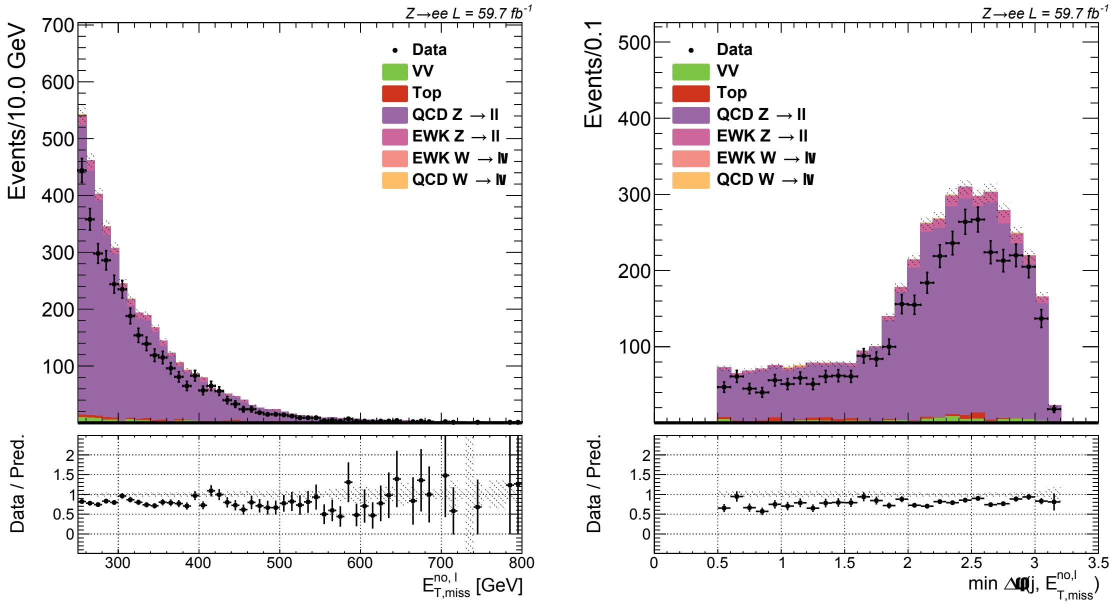


(a) $E_{T,miss}^{no,e}$ - MTR (b) $min\Delta\phi(j, E_{T,miss}^{no,e})$ - MTR

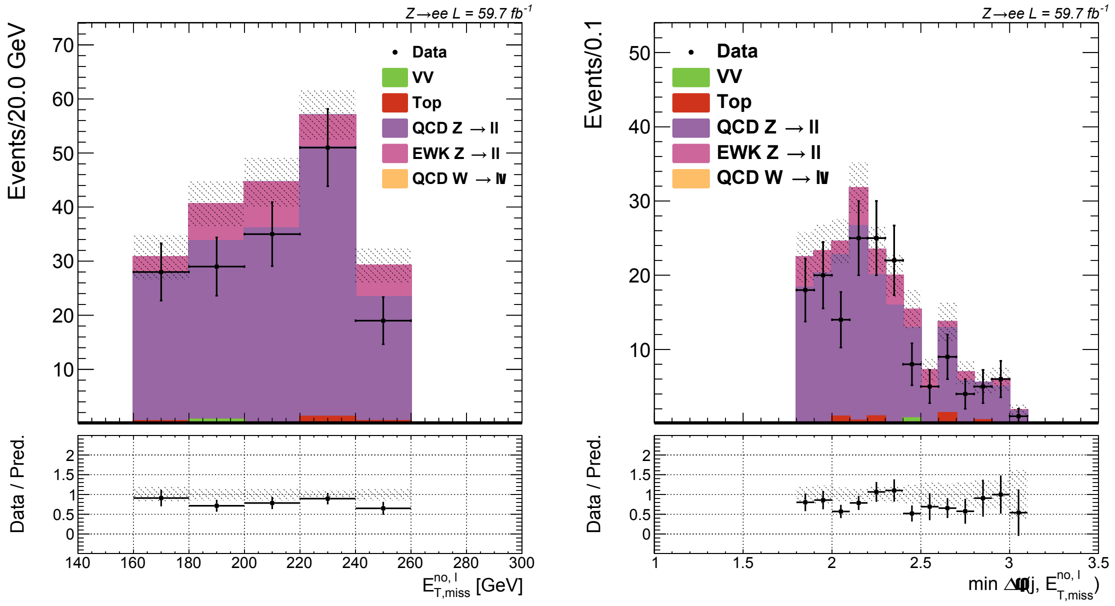


(c) $E_{T,miss}^{no,e}$ - VTR (d) $min\Delta\phi(j, E_{T,miss}^{no,e})$ - VTR

Figure 8.11: Distributions of $E_{T,miss}^{no,e}$ and $min\Delta\phi(j, E_{T,miss}^{no,e})$ variables in the double muon region for MTR (top) and VTR (bottom) categories for the 2017 era of data taking.



(a) $E_{T,miss}^{no,e}$ - MTR (b) $min\Delta\phi(j, E_{T,miss}^{no,e})$ - MTR



(c) $E_{T,miss}^{no,e}$ - VTR (d) $min\Delta\phi(j, E_{T,miss}^{no,e})$ - VTR

Figure 8.12: Distributions of $E_{T,miss}^{no,e}$ and $min\Delta\phi(j, E_{T,miss}^{no,e})$ variables in the double muon region for MTR (top) and VTR (bottom) categories for the 2018 era of data taking.

This is confirmed by Figure 8.13, which shows both of these variables. It can be seen that there is a large excess in data coming from the affected region which is not found in the simulation. This is a direct result of a lack of the HE information in the HEM region. Objects that would have been reconstructed as jets have a larger probability to be misidentified as an electron, creating the large spikes seen in η/ϕ distributions.

There are two approaches that can be taken at this stage. The first one, also the simpler one, is to effectively veto events that have an electron in the HEM region. This translates into a veto range of $\eta <$ -1.3 and -1.6 $\leq \phi \leq$ -0.9. An alternative approach would be to redefine the electron requirements by removing the objects which are found in the affected region. This option provided no improvements over the first approach, leading to the simpler, veto approach being chosen. This was due to the fact that events gained mostly originated from QCD multijet processes, hence being unable to successfully satisfy the tight requirements of MTR/VTR jet selection.

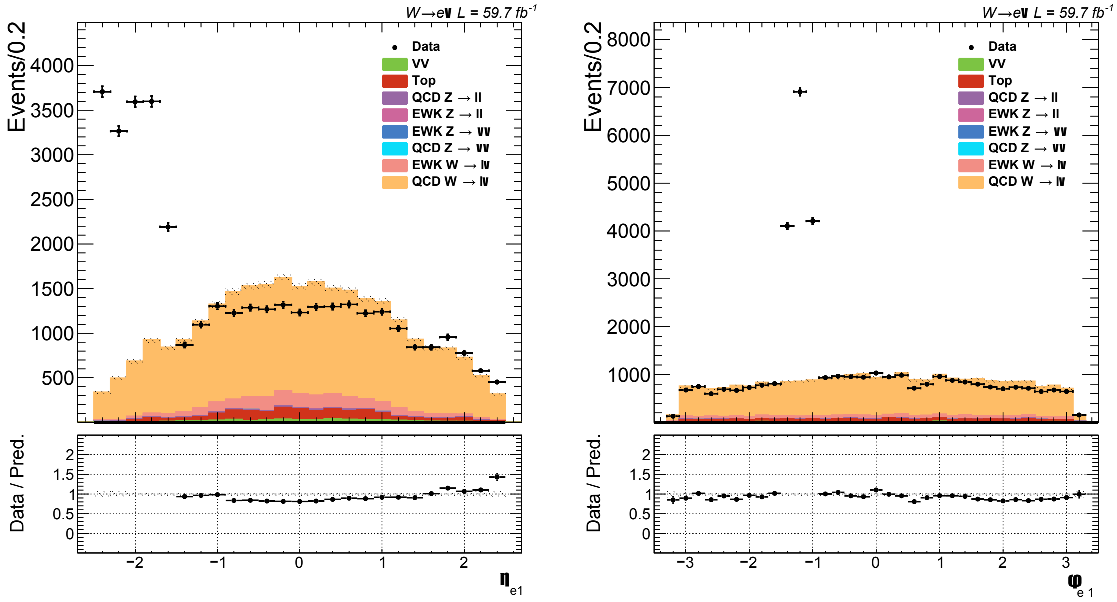
Figures 8.14 and 8.15 show the m_{jj} and $E_{T,miss}^{no,e}$ distributions for 2017 and 2018 era respectively, with the m_{jj} showing good agreement between data and simulation for all eras. Additionally, the $M_{T,e}$ (computed from $\vec{p}_{T.miss}$ and the electron p_T) and $min\Delta\phi(j, E_{T,miss}^{no,e})$ data to simulation comparisons are shown in Figures 8.16 and 8.17 for both categories and eras. Additional distributions related to this region are given in Appendix A.4.

The story of the background extraction using these four lepton regions is one of the main focuses of Chapter 9, where they will be included in the final fit through the use of transfer factors connecting CRs with the final SR background estimation.

8.3 Dedicated QCD multijet region

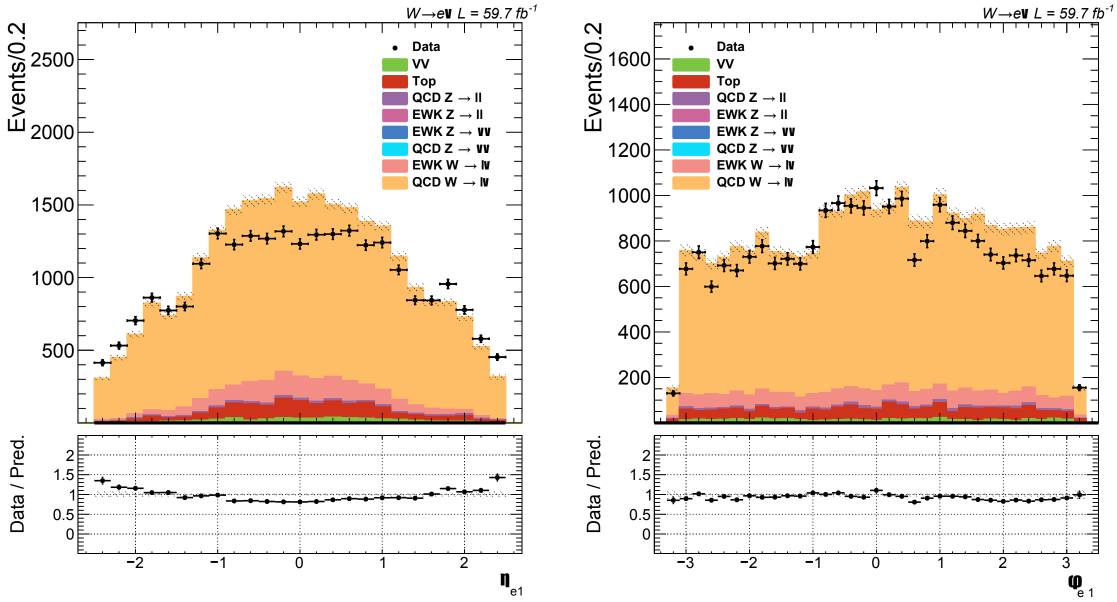
The main reason behind the problematic nature of QCD multijet processes is their high production rate at the LHC. They are not expected to contain a lot of events with the ability to pass the MTR/VTR selection requirements (especially due to the existence of the $min\Delta\phi(j, E_{T,miss})$ threshold), as they usually produce events which are well balanced in the transverse plane. Unfortunately, any mismeasurement of jet energy, poorly functioning detector regions or neutrinos from semileptonic decays of heavy-flavour mesons, coupled with large cross section value, result in a non negligible amount of them containing VBF-like signature (energetic leading/subleading jets accompanied with a large $E_{T,miss}$ originating from these edge cases).

Their contribution is significantly reduced in the lepton CRs through the use



(a) η_e - pre veto

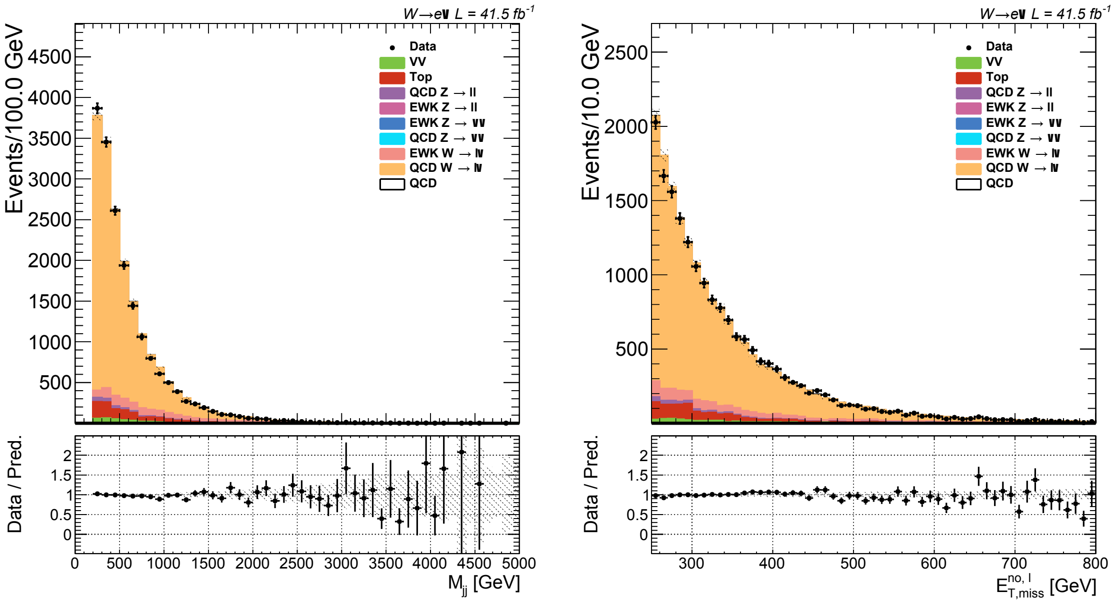
(b) ϕ_e - pre veto



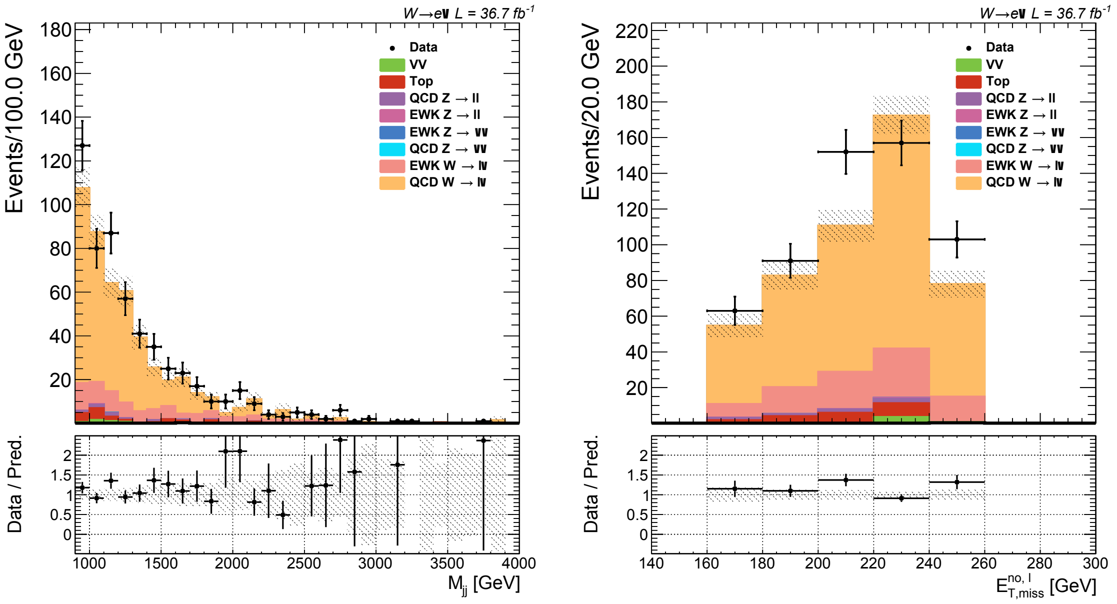
(c) η_e - post veto

(d) ϕ_e - post veto

Figure 8.13: Distribution of electron η and ϕ variables showing the pre (top) and post (bottom) veto mitigation results for the single electron region (presented for the MTR category).

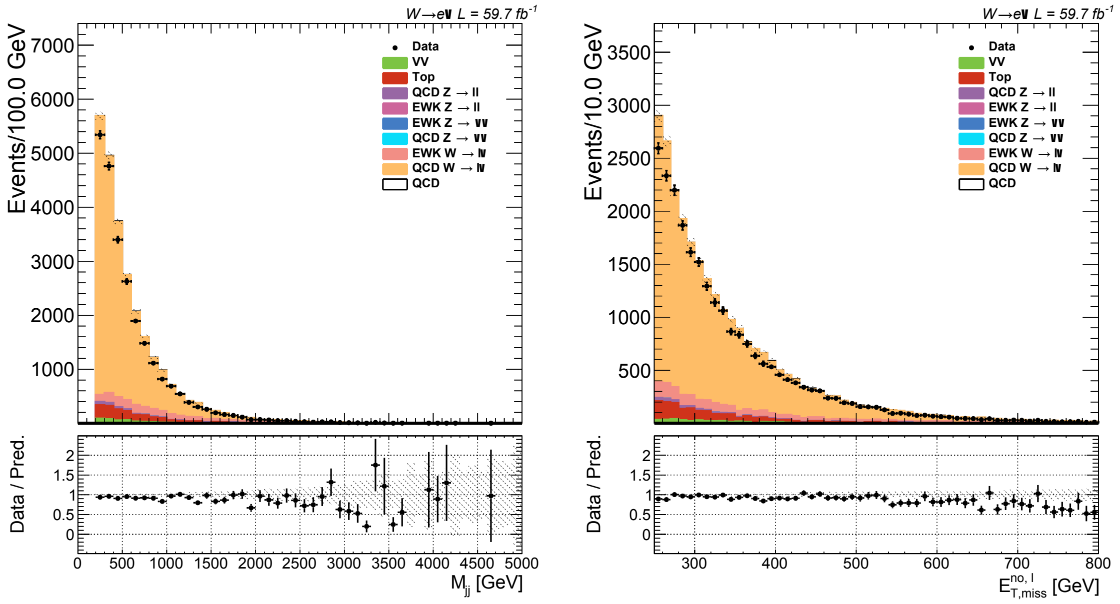


(a) m_{jj} - MTR (b) $E_{T,miss}^{no,e}$ - MTR

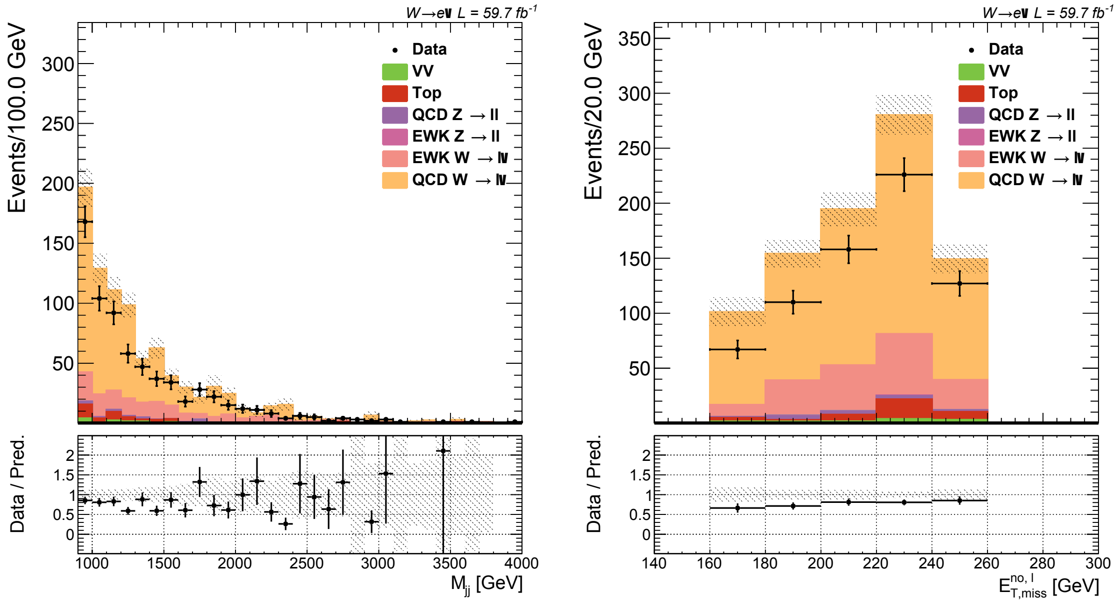


(c) m_{jj} - VTR (d) $E_{T,miss}^{no,e}$ - VTR

Figure 8.14: Distributions of mjj and $E_{T,miss}^{no,e}$ variables in single electron region for MTR (top) and VTR (bottom) categories for the 2017 era of data taking.

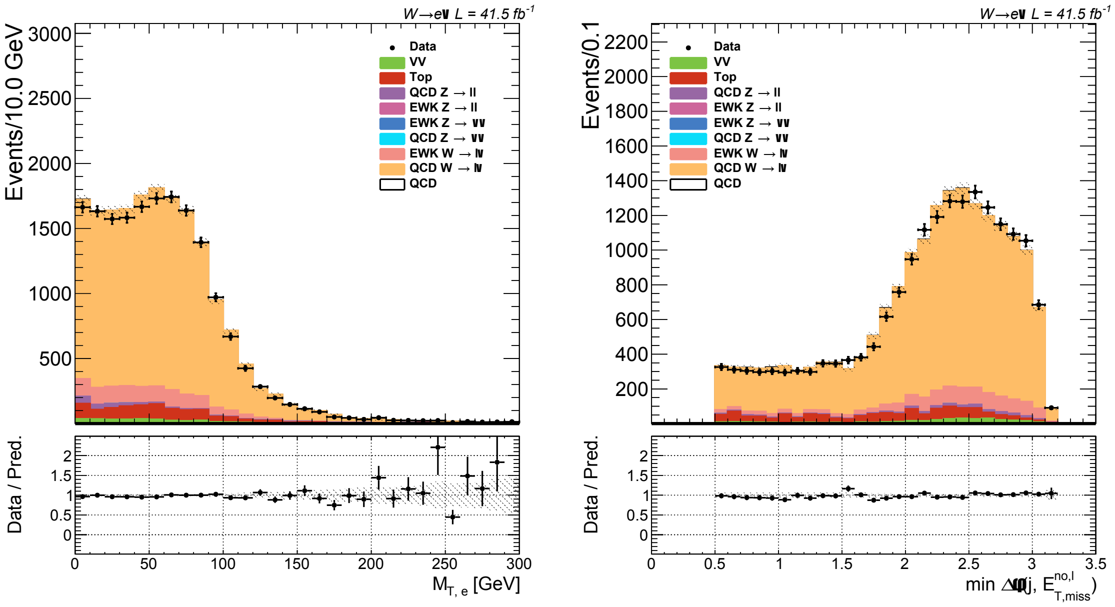


(a) m_{jj} - MTR (b) $E^{no,e}_{T,miss}$ - MTR



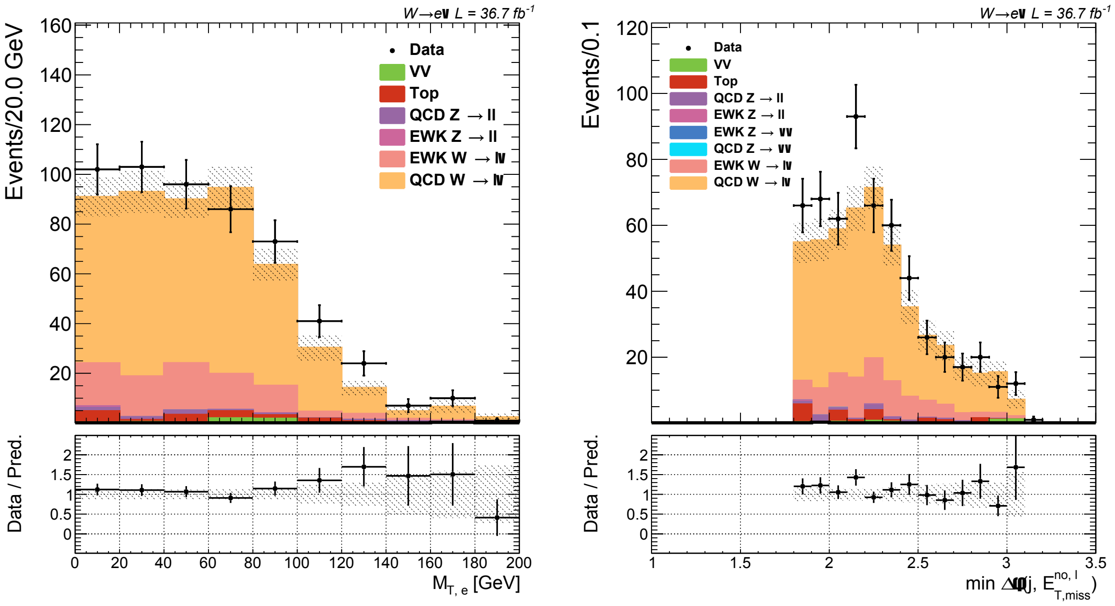
(c) m_{jj} - VTR (d) $E^{no,e}_{T,miss}$ - VTR

Figure 8.15: Distributions of mjj and $E^{no,e}_{T,miss}$ variables in single electron region for MTR (top) and VTR (bottom) categories for the 2018 era of data taking.



(a) $M_{T,e}$ - MTR

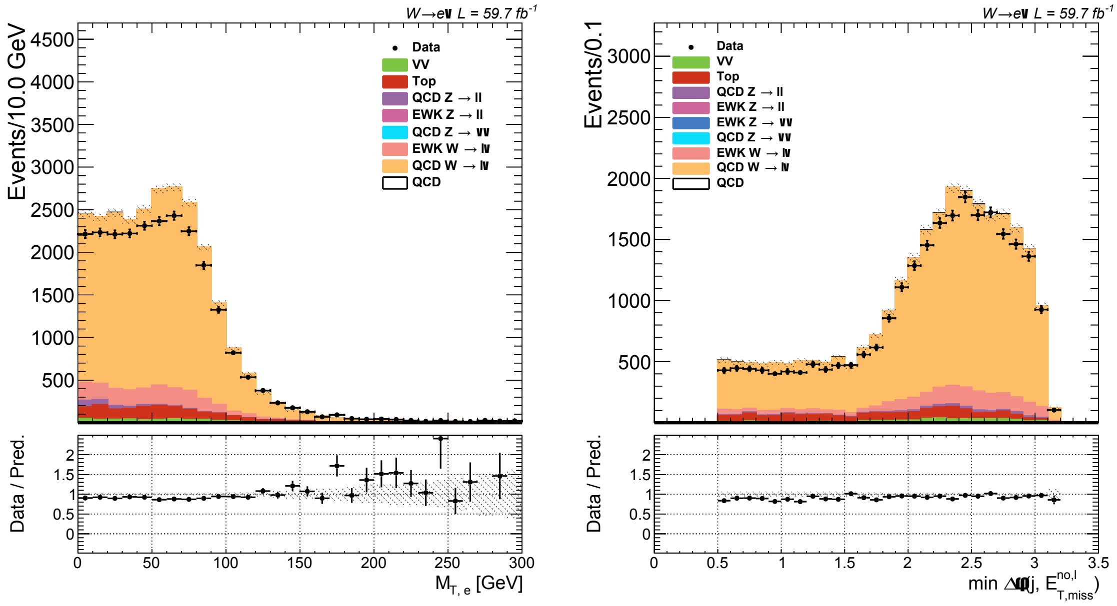
(b) $min\Delta\phi(j, E_{T,miss}^{no,e})$ - MTR



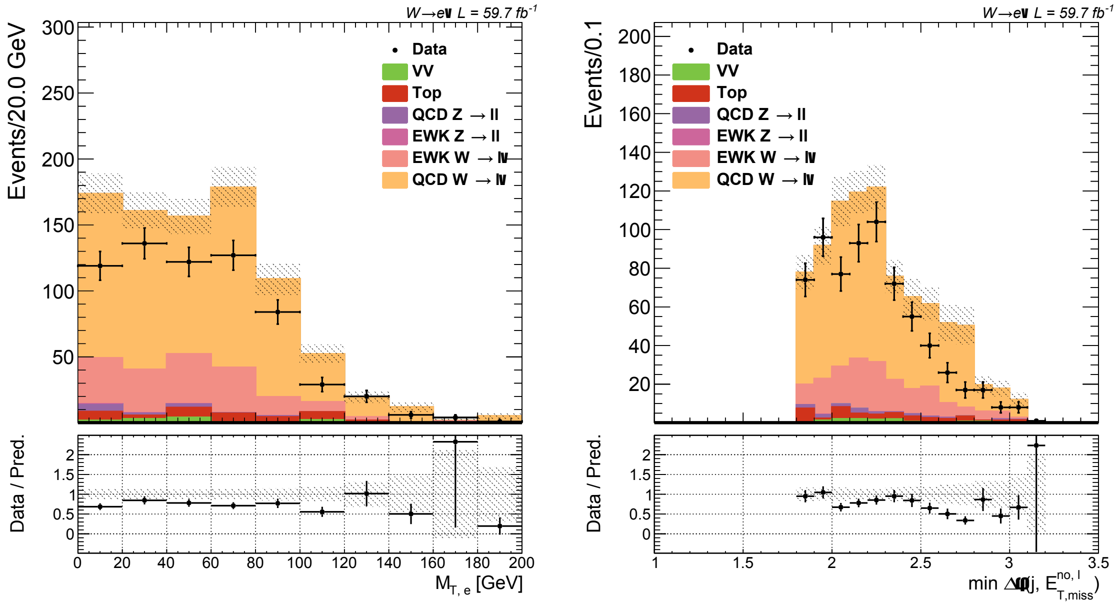
(c) $M_{T,e}$ - VTR

(d) $min\Delta\phi(j, E_{T,miss}^{no,e})$ - VTR

Figure 8.16: Distributions of $M_{T,e}$ and $min\Delta\phi(j, E_{T,miss}^{no,e})$ variables in single electron region for MTR (top) and VTR (bottom) categories for the 2017 era of data taking.



(a) $M_{T,e}$ - MTR (b) $min\Delta\phi(j, E_{T,miss}^{no,e})$ - MTR



(c) $M_{T,e}$ - VTR (d) $min\Delta\phi(j, E_{T,miss}^{no,e})$ - VTR

Figure 8.17: Distributions of $M_{T,e}$ and $min\Delta\phi(j, E_{T,miss}^{no,e})$ variables in single electron region for MTR (top) and VTR (bottom) categories for the 2018 era of data taking.

of the $min\Delta\phi(j, E_{T,miss})$ variable. In order to properly estimate this contribution in the SR (also bypassing the low statistical precision of simulation samples, by not relying simply on them), an extrapolation approach was deployed through the use of a region which is considered "enriched" in these processes. The following sections define this region and present two approaches to determine the background contribution, one of which is chosen for the final estimation, while the other is used as a cross check.

8.3.1 Definitions

The importance of the $min\Delta\phi(j, E_{T,miss})$ variable can be seen in its connection to QCD multijet processes. If there is a jet energy mismeasurement for a process which is expected to be a well balanced one, the $\vec{p}_{T,miss}$ will be influenced by the direction of those missmeasured jets. The low $min\Delta\phi(j, E_{T,miss})$ range represents a good way of creating a region largely populated by multijet processes, which can, in return, be used to estimate their contribution in the SR. This difference between two regions (low and high $min\Delta\phi(j, E_{T,miss})$) can be expressed as:

$$r = \frac{min\Delta\phi(j, E_{T,miss}) > X}{min\Delta\phi(j, E_{T,miss}) < X}, \tag{8.3}$$

where X denotes the threshold corresponding to each category, taking the value of 0.5 (1.8) for MTR (VTR) category respectively.

In order to obtain an estimation of the SR-related contribution, as well as to perform a set of closure tests, four orthogonal regions are introduced. They are defined with slight modifications of the MTR (VTR) selection by inverting the $min\Delta\phi(j, E_{T,miss})$ and $E_{T,miss}$ requirements, as shown in Table 8.1.

Name	Definition
Region A	$min\Delta\phi(j, E_{T,miss})$ <0.5 (1.8) and 100< $E_{T,miss}$ <160 GeV
Region B	$min\Delta\phi(j, E_{T,miss})$ >0.5 (1.8) and 100< $E_{T,miss}$ <160 GeV
QCD CR	$min\Delta\phi(j, E_{T,miss})$ <0.5 (1.8) and $E_{T,miss}$ >250 GeV (∈[160, 250) GeV)
Signal Region	$min\Delta\phi(j, E_{T,miss})$ >0.5 (1.8) and $E_{T,miss}$ >250 GeV (∈[160, 250) GeV)

Table 8.1: Definition of four regions used for the estimation of the total contribution of QCD multijet processes in the SR. Thresholds are presented for MTR (VTR) category respectively.

For the purposes of this analysis two methods were tested. Both of them are summarised in the following sections. The first depends both on the simulation samples and data, while the second one has no dependence on the simulated QCD multijet processes (making it the preferred option, due to the lack of statistical precision in the simulation samples).

8.3.2 Method A

The basis for this method is the, previously defined, factor r. Since the analysis relies on the m_{jj} variable for the final fit, its definition will be expanded to appropriately reflect dependence on the dijet mass. It can be expressed in terms of m_{jj} bins as:

$$r(m_{jj}) = \frac{F^{MC}_{SR}(m_{jj})}{F^{MC}_{QCD\ CR}(m_{jj})}, \tag{8.4}$$

where the F^{MC}_{i} denotes the m_{jj} distributions of QCD multijet simulation samples in the QCD CR and SR.

Due to the lack of statistical precision in respective simulation samples, some thresholds in the VBF selection had to be loosened in order for this method to produce sensible output. This resulted in a relaxed, $\Delta\phi_{jj} < 2.5$, requirement being introduced when computing $r(m_{jj})$. The final multijet contribution in the SR is obtained, by translating the data-driven estimation in the CR using the $r(m_{jj})$, through the usage of the following formula:

$$N^{QCD}_{SR}(m_{jj}) = \left(N^{Data}_{CR}(m_{jj}) - \sum_{i} N^{i}_{CR}(m_{jj})\right) \cdot r(m_{jj}), \tag{8.5}$$

where the sum goes over other backgrounds (V+jets, now a minor background in this region, $t\bar{t}$ and diboson processes). Following the strong dependence on the simulated sample (and due to its the lack of statistical precision), a more favourable method was searched for that would be mostly data driven. The following section defines this method and its usage in further analysis steps.

8.3.3 Method B

An alternative approach to method A is to remove the dependence on the QCD multijet simulation and instead rely only on a mostly data driven estimation[1]. For the purposes of this method, a fit procedure is deployed on the $min\Delta\phi(j, E_{T,miss})$ in its low region. The resulting fit function enables the extrapolation into the SR.

---

[1] The simulated samples of other SM processes besides QCD multijet, which present a set of minor backgrounds in the QCD CR, have sufficient statistical precision to be used for the purposes of this method. This removes the constraint found in Method A where the main worry was assigned to the low statistical precision of multijet simulated samples.

Firstly, in order to have a good way of estimating the contribution of other backgrounds in the, QCD enriched, low $min\Delta\phi(j, E_{T,miss})$ region, a fit is performed on the combined contribution of V+jets, diboson and $t\bar{t}$ simulation samples. This fit is performed using a function defined as:

$$F_B(x) = Q_0 e^{-Q_1 x}(1 + Q_2 x + Q_3 x^2), \tag{8.6}$$

where Q_i denotes fit parameters, while x represents the fit variable (in this situation $min\Delta\phi(j, E_{T,miss})$). The next step is to fit the data using a function defined as:

$$F(x) = F_B(x) + F_{QCD}(x) = Q_0 e^{-Q_1 x}(1 + Q_2 x + Q_3 x^2) + P_0 e^{-P_1 x} + P_2, \tag{8.7}$$

where the P_i represent the new fit parameters. Through the usage of the information gained by fitting the smaller backgrounds [2], the parameters Q_i are fixed when performing the fit in data, thus allowing for the definition of the $F_{QCD}(x)$ through the remaining P_i parameters. In order to estimate the total contribution coming from QCD multijet processes in the SR, this newly obtained function $F_{QCD}(x)$ can be integrated over the SR, or in other words:

$$N^{SR}_{QCD} = \int_X^{\pi} F_{QCD}(x)dx, \tag{8.8}$$

where the lower bound X takes the value of 0.5 (1.8) for MTR (VTR) category respectively. The actual implementation requires a separate fit procedure per category for each era. For the MTR category, a fit range of $0 \leq min\Delta\phi(j, E_{T,miss}) \leq 1.0$ is chosen, while the $0.5 \leq min\Delta\phi(j, E_{T,miss}) \leq 1.0$ range covers the VTR category fit strategy. The extended fit range used for the MTR category is enabled by the partial unblinding strategy defined in Section 7.4 and supported by the fact that this range is not expected to have any significant signal contribution. One important remark regarding this option is the statement that it is built on the fact that the dijet mass is not strongly correlated with with $min\Delta\phi(j, E_{T,miss})$ for the QCD multijet processes. This is clearly shown in the distribution $r(m_{jj})$ (from the implementation process of method A) shown in Figure 9.2. It displays the flat behaviour of the aforementioned ratio across the m_{jj} range with sufficient statistical precision.

The final step is to translate this information into the final SR contribution in the m_{jj} variable. The overall normalisation of the QCD multijet SR contribution is there, the only thing which remains is to determine the shape of the distribution.

---

[2] Smaller backgrounds only from the current, QCD CR, perspective, as V+jets are main irreducible backgrounds contributing to the SR.

Due to the aforementioned small correlation between the two variables, this can be achieved by looking into the QCD CR and subtracting the estimated contribution of other backgrounds from data. Upon obtaining the resulting m_{jj} distribution, the only thing left to do is to scale it to the proper normalisation, which in this case is represented by the value of N^{SR}_{QCD}. Figure 8.18 summarises this method in a form of a diagram.

Figure 8.19 shows the first step of this process - the fitting procedure of the $\min\Delta\phi(j, E_{T,miss})$ variable used to extract the total yield of multijet processes in the SR (presented for the MTR category for both 2017 and 2018 data). The fit range is defined to reach $\min\Delta\phi(j, E_{T,miss}) < 1$ (1.5) for MTR (VTR) categories. Similarly, a validation region is introduced in order to estimate the quality of the fit. It is starts from the end of the fit range until $\min\Delta\phi(j, E_{T,miss}) < 1.2$ (1.8) for MTR (VTR) categories, respectively. This was made possible as a result of studies following the unblinding strategy, which was put in place in order to mitigate the effects of jet horns and HEM problems (as described in Chapter 7.4), with the extended range (both for the fit and validation) being supported with the assumption that no significant signal contribution is expected bellow those values.

These distributions present data (black points), sum of minor backgrounds (blue points), multijet simulated events (red points - only for comparison, not being used anywhere) and the fit functions for the data ($F(x)$ - dashed black line), minor backgrounds ($F_B(x)$ - dashed blue line) and the multijet estimation (F_{QCD} - dashed red line). Additionally, a sum of the contributions from minor backgrounds combined with F_{QCD} is presented (solid magenta line). The bottom panel focuses on the ratios between the fit functions and their corresponding processes. This includes a comparison of $F(x)$ versus data (expressed with black color), which shows a good agreement above the fit range. Lastly, comparisons are made between the F_B and the sum of other (non-QCD multijet) background processes (blue) as well as the F_{QCD} versus multijet simulated events (red - but, similar to above, only for illustration purposes). The difference between the QCD multijet simulated events (accompanied with their large uncertainties) and the resulting F_{QCD} arises due to the low statistical precision of the simulated samples [3] and ultimately represents yet another indication of the necessity for this approach.

As previously mentioned, the following chapter is going to focus on the practical implementation of these CRs into the analysis flow and their associated uncertainties, serving as a companion by concluding the story of background processes from the point of view of this study.

---

[3] Due to the existence of a few events originating from low statistical precision samples but having a large production cross section value.

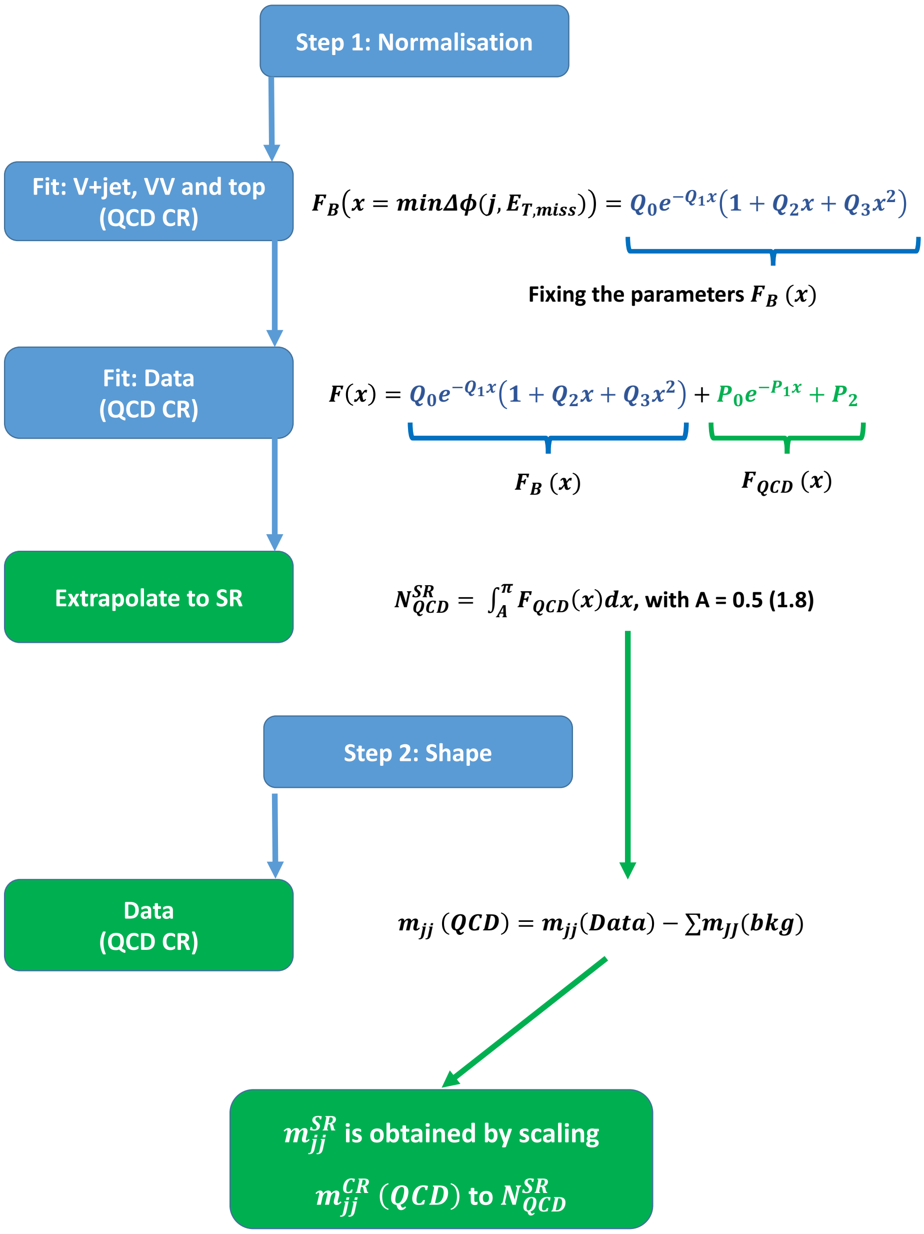
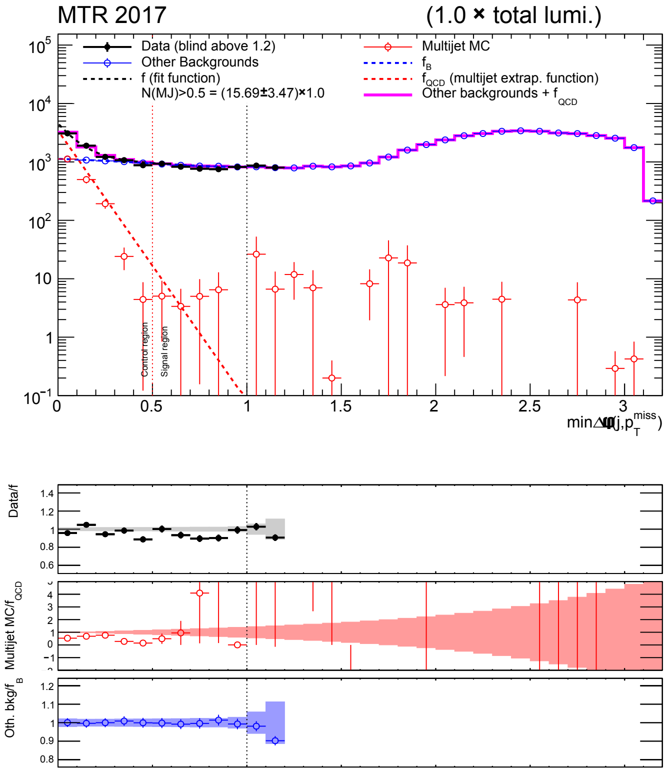
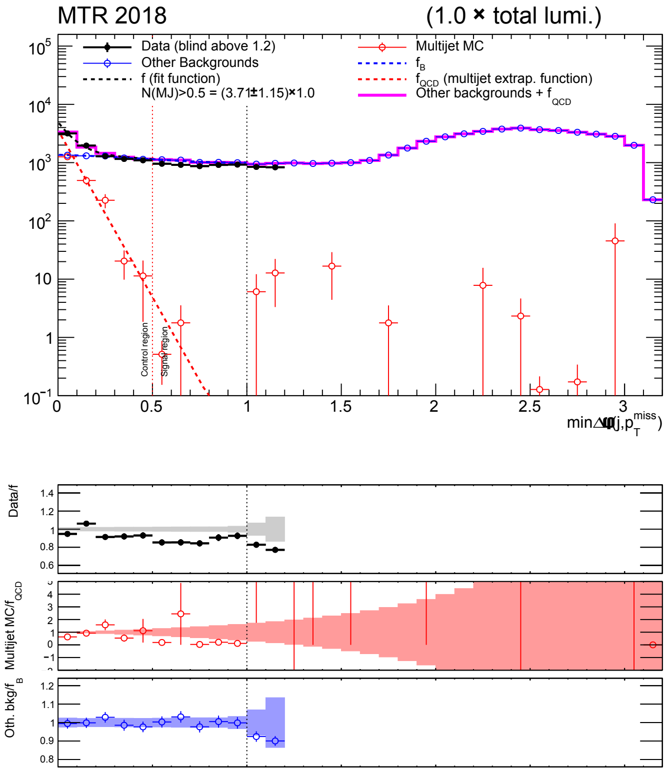


Figure 8.18: Diagram showing main steps of the Method B approach to the estimation of the contribution from QCD multjet processes in the SR.



(a) MTR 2017



(b) MTR 2018

Figure 8.19: Extraction of the overal normalisation of the SR contribution originating from QCD multijet processes obtained through the use of Method B for the MTR category for (a) 2017 and (b) 2018 data.

Chapter 9

Fit structure and results

> *"What could I say to you that would be of value, except that perhaps you seek too much, that as a result of your seeking you cannot find. "*
>
> — Hermann Hesse

9.1 Introduction

Before proceeding with the discussion of the final result of this study, an overview of all analysis inputs needs to be made in order to summarise the constituents entering the final measurement. The following sections introduce the fit strategy, presented in terms of the signal extraction approach, paired with a general introduction to the statistical apparatus. Building on that, the next item for discussion is the formation of the likelihood function for the searches of the invisible state. It will be shown how it includes the information from dedicated control regions as well as from the signal region.

Following the summary of contributions from each region, the focus is placed on the overview of systematic uncertainties. This discussion is split into two distinct parts. The first one focuses on theoretical uncertainties and encapsulates details regarding their origin from the NLO corrections on V+jets simulation samples. The second part is comprised of a discussion of experimental uncertainties covering their various sources.

Lastly, the third act of this chapter presents results arising from measurements defined by the previously introduced strategy. This section is dedicated to measurements through the use of data collected by the CMS experiment during the 2017 and 2018 eras of data taking. Th final result from the VBF H→inv search using the full Run 2 dataset is obtained through a combination with the, previously published, study detailing the analysis of 2016 data [39].

9.2 The CLs approach

The characteristics of the VBF topology are manifested through the existence of two jets. Through an optimisation technique it was shown that the largest signal versus background shape separation is gained when deploying the dijet invariant mass as the main analysis tool[1] [39, 41]. For a measurement of the aforemnetioned property, the resulting data can be represented with a binned histogram (with d_i denoting a certain mass bin). Due to the nature of collider experiments, the use of Poisson statistics is applicable to studies of this kind. It allows for the introduction of the binned likelihood function as:

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_i \frac{(\mu S_i(\boldsymbol{\theta}) + B_i(\boldsymbol{\theta}))^{d_i} e^{-(\mu S_i(\boldsymbol{\theta}) + B_i(\boldsymbol{\theta}))}}{d_i!} = \prod_i \mathrm{Pois}(d_i | (\mu S_i(\boldsymbol{\theta}) + B_i(\boldsymbol{\theta})), \qquad (9.1)$$

where the terms comprising the product can be interpreted as the probabilities that d_i occurrences of the dijet mass, confined to the bin range of i, has been observed given the expected valued of events being: $\mu S_i(\boldsymbol{\theta}) + B_i(\boldsymbol{\theta})$ [127, 128]. The bin values associated with the signal ($S_i(\boldsymbol{\theta})$) and background ($B_i(\boldsymbol{\theta})$) processes are obtained from the simulation of SM processes and the dependency on a set of nuisance parameters $\boldsymbol{\theta}$. Lastly, μ is also a free parameter in the fit and in this scenario it represents the desired branching ratio. The test statistic may be formed as: $q_\mu = -2ln\frac{\mathcal{L}(\mu,\boldsymbol{\theta}(\boldsymbol{\mu}))}{\mathcal{L}(\mu_m,\boldsymbol{\theta}_{\boldsymbol{m}})}$, where μ_m and θ_m represent the values of the parameters yielding the largest value of the likelihood function, and $\theta(\mu)$ denotes the value of a parameter θ which maximises the likelihood function for a given choice of μ.

When approaching the task of setting a limit on the probability of the Higgs boson decaying invisibly, one must propose a way of thinking opposite to the case when there is a hunt for a discovery. In these scenarios, the null hypothesis (H_0) is represented by the signal + background scenario which is compared to the alternative (H_1) denoted as background only scenario (as introduced in Ref. [127]). For the purposes of the VBF H→inv search this involves introducing the SM Higgs by fixing the values of Br(H→inv) to be 1 or 0 respectively. Following from the previous definitions, the final comparison can be made by following the CLs criterion [127,

---

[1] More details about the jet properties from the perspective of the VBF production mode are given in Chapter 7.

128], through which the value of the 95% CL upper limit on Br(H→ inv) is obtained. This procedure requires a definition of the comparison criteria (CL_S) in order to exclude one of two hypotheses (signal+background and background only). For the purposes of VBF H→inv search, this is done in the following way [40, 41]:

$$CL_S = \frac{p\left(q_\mu \geq q_\mu^{obs} | \mu \cdot S(\boldsymbol{\theta}) + B(\boldsymbol{\theta})\right)}{p\left(q_\mu \geq q_\mu^{obs} | 0 \cdot S(\boldsymbol{\theta}) + B(\boldsymbol{\theta})\right)} \leq 0.05 = 1 - \alpha, \tag{9.2}$$

where the q_μ^{obs} represents the observed value of q_μ. The numerator denotes the p value of the signal+background model and is defined as:

$$p\left(q_\mu \geq q_\mu^{obs} | \mu \cdot S(\boldsymbol{\theta}) + B(\boldsymbol{\theta})\right) = \int_{q_\mu^{obs}}^{\infty} F\left(q_\mu | \mu \cdot S(\boldsymbol{\theta}) + B(\boldsymbol{\theta})\right) dq_\mu, \tag{9.3}$$

where F represents the probability density function of q_μ. A similar definition is used to introduce the background only model represented with the denominator by replacing $\mu = 0$. Finally, in order to get the desired α CL upper limit on μ (within this study $\alpha = 95$ % and μ is the desired Br(H→inv)) an inversion of the previously introduced criteria is performed to obtain that the 95 % CL upper limit on Br(H→inv) = $CL_S^{-1}(1-\alpha)$.

This simplified method of having only one region represented with $\mu S_i + B_i$ is used to illustrate the entire process without the pressure of multiple additional background enriched regions. The details of their inclusion into the signal extraction procedure are the focal point of the following section.

9.3 Signal extraction strategy

The information gathered from four lepton control regions (described in Chapter 8) and the signal region (with both analysis categories being defined in Chapter 7) is used when forming of the, m_{jj} binned, likelihood function. It is formed in such a way that it allows the fit to simultaneously access all available information, with the end result of having a final estimate of the contribution originating from Z→ $\nu\nu$+jets and W→ $l\nu$+jets irreducible backgrounds. The formation of the likelihood function can be split in the following way:

$$\mathcal{L}(\boldsymbol{\mu}^{Z\rightarrow\nu\nu}, \boldsymbol{\mu}, \boldsymbol{\theta}) = \mathcal{L}_{SR} \times \prod_{j=\mu,e} \mathcal{L}^{i}_{CR,Z} \times \prod_{k=\mu,e} \mathcal{L}^{j}_{CR,W} \times \prod_{l} \mathrm{P}(\theta_l), \tag{9.4}$$

where $\boldsymbol{\mu}^{Z\rightarrow\nu\nu} = (\mu_i^{Z\rightarrow\nu\nu})$ summarises the binned contribution of QCD $Z \rightarrow \nu\nu$ SM background processes and $\boldsymbol{\mu}$ represents the signal strength parameter. Both of these are free parameters withing the fit. Additionally, $\boldsymbol{\theta} = (\theta_l)$ symbol is used to summarise systematic uncertainties included in the likelihood in the form of constrained nuisance parameters ($P(\theta_l)$ terms). These terms are represented with log-normal functions, where the logarithm of the parameter behaves as normally distributed property [39]. This choice of the form describing the behaviour of various uncertainties is best suited for scenarios involving a range of small multiplicative uncertainties, which is suitable for this analysis. This is crucial as this approach bounds the parameters to be positive in value, which would not always be the case with normal distribution) [129–131].

Starting with the first member of the likelihood function, the item representing the information given by the SR, akin to the one previously introduced with Equation 9.1, can be written as:

$$\mathcal{L}_{SR} = \prod_i \mathrm{Pois}\left(d_i | B_i(\boldsymbol{\theta}) + (1 + f_i(\boldsymbol{\theta})_{\mathrm{QCD}})\mu_i^{Z\to\nu\nu} + R_i^{\frac{EWK}{QCD}}(1 + f_i(\boldsymbol{\theta})_{\mathrm{EWK}})\mu_i^{Z\to\nu\nu} + \boldsymbol{\mu} S_i(\boldsymbol{\theta})\right), \tag{9.5}$$

where d_i represents the number of events observed in data for a given m_{jj} bin i, and B_i and S_i define the total minor background and nominal signal yields[2] within the same bin range. With the $\mu_i^{Z\to\nu\nu}$ being free parameters, a parameter $R_i^{\frac{EWK}{QCD}}$ is introduced in order to connect the SR contribution from EWK and QCD productions of $Z \to \nu\nu$+jets processes (estimated from simulation[3]). Lastly, a connection with the W→ $l\nu$+jets processes is introduced through the addition of $f_i(\theta)_{QCD} = \frac{SR_{QCD}^{W\to l\nu}}{SR_{QCD}^{Z\to\nu\nu}}$, which represents a ratio of simulated contributions of these two backgrounds in the SR. It serves as a connection between these two sources of backgrounds (the equivalent definition follows for the EWK production).

The term summarising the contribution from each of the dilepton CRs, using the dielectron region as an example, can be introduced as:

$$\mathcal{L}_{CR,Z}^{e} = \prod_i \mathrm{Pois}\left(d_i^Z | B_i^Z(\boldsymbol{\theta}) + \frac{\mu_i^{Z\to\nu\nu}}{R_i^Z(\boldsymbol{\theta})_{\mathrm{QCD}}} + R_i^{\frac{EWK}{QCD}} \cdot \frac{\mu_i^{Z\to\nu\nu}}{R_i^Z(\boldsymbol{\theta})_{\mathrm{EWK}}}\right). \tag{9.6}$$

where d_i^Z represents the number of events observed in data for a given m_{jj} bin i, and the ratios $R_i^Z(\boldsymbol{\theta})_{QCD}$ and $R_i^Z(\boldsymbol{\theta})_{EWK}$ represent transfer factors which connect the overall CR yields of QCD (EWK) Z→ ll+jets processes with their QCD (EWK)

---

[2] The S_i is comprised from both the VBF and ggH+X productions, where the X represents scenarios with two or more additional jets, thus having the possibility of passing the VBF selection requirements.

[3] Assuming it does not posses any additional uncertainty.

Z→ $\nu\nu$+jets counterparts in the SR (again within the given m_{jj} bin range). The single lepton contribution follows a similar idea and can be defined as [4]:

$$\mathcal{L}^{e}_{CR,W} = \prod_{i} \text{Pois}\left(d_i^W | B_i^W(\boldsymbol{\theta}) + f_i(\boldsymbol{\theta})_{QCD} \cdot \frac{\mu_i^{Z\to\nu\nu}}{R_i^W(\boldsymbol{\theta})_{\text{QCD}}} + R_i^{\frac{EWK}{QCD}} \cdot f_i(\boldsymbol{\theta})_{EWK} \cdot \frac{\mu_i^{Z\to\nu\nu}}{R_i^W(\boldsymbol{\theta})_{\text{EWK}}} \right), \tag{9.7}$$

where the definitions of transfer factors follow their dilepton counterparts. These, followed with their muon variants, construct the likelihood function introduced with Equation 9.4. For the MTR category there is one more constituent added to this definition, a product of Poissonian terms focusing on the photon region. Following a similar blueprint (and motivation), the photon CR is added in order to aid with the constrain of the $Z \to \nu\nu$+jets SM background. Its topology, for large values of photon transverse momenta, follows a similar trend as $Z \to \nu\nu$+jets. The formation of this region requires exactly one tight photon in the event, having no additional photons or leptons, with the dijet topology requirements being placed on top of it. The likelihood contribution of the photon region is equivalent to the one definied for dilepton regions in Equation 9.6[5].

9.4 Treatment of uncertainties

Systematic uncertainties play a significant role in these measurements and can be either a result of higher order theory corrections applied to the LO simulation samples (as explained in Chapter 6) or they can be associated with one of multiple experimental sources. The following sections detail each of these groups.

9.4.1 Theoretical uncertainties

Simulated samples of the main signal processes, VBF and ggH production, have associated uncertainties originating from factorization and renormalization (covering the relevant higher order terms in the calculation of respective cross sections), and parton density function (PDF) variations (reflective both on the choice of the PDF and the coupling constant)[6]. Starting with the factorization/renormalization variations, they introduce the "QCD scale" uncertainty of $^{+4.6\ \%}_{-6.7\ \%}$ on ggH and $^{+0.4\ \%}_{-0.3\ \%}$

---

[4] Again using the electron region as the example.

[5] With the redefintion of the transfer factors in order to include the information coming from the photon region.

[6] These uncertainties were derived for the 2016 analysis using the recipes from Refs. [132] (as explained in more detail in Refs. [39, 41]).

on VBF production modes. Additionally, an uncertainty covering the change in acceptance of the VBF production (with respect to this analysis) is set to be 2 %.

Similar is done for the PDF variation, where an uncertainty of 2.1 % (3.2 %) arises for the VBF (ggH) production mode, with an accompanying acceptance uncertainty of 1 % for both cases. These uncertainties associated with the signal acceptance are treated separately per process. The ggH production has additional sources of uncertainty. The first one originates from the limited information about the ggH+X cross sections. The second contribution arises from the uncertainty related to the estimation of the ggH cross section for large values of p_T^{Higgs}[7]. This totals to an additional ∼45 % uncertainty [39, 41].

Theoretical uncertainties also have an effect on the transfer factors $f(\boldsymbol{\theta})$[8]. The core of these uncertainties is represented with the effects connected to the EWK and QCD higher order corrections coupled with the uncertainty associated with the PDF modeling. The first item taken into consideration arises again due to variations of the choice of the central renormalization and factorization scale, this time with respect to the QCD NLO corrections on QCD V+jets processes. These variations involve changing both scales by increasing or decreasing them by a factor of two with respect to their nominal value, continuing with the convention used for the 2016 analysis [39]. This is reflected in the final weights being used in the analysis by creating scale ”Down” and ”Up” alternatives. The second item related to the choice of PDFs is also following the previously introduced path. This uncertainty follows the recipe presented in Ref. [133] and it is inclusive of both the PDF uncertainty and the uncertainty arising from a particular choice of the coupling constant[9].

These uncertainties are assumed to be partially correlated between the W and Z samples with a more conservative approach when estimating their effect on the $f(\boldsymbol{\theta})$ ratios. This option considers only the variation concerning the W+jets processes (being the larger contribution in the W/Z ratio). This set of uncertainty values is used to cover the effect of the QCD NLO corrections of the transfer factor $f(\boldsymbol{\theta})$ related to the EWK V+jets processes as well (again being a conservative choice as the respective uncertainties for the EWK production are expected to be smaller), but are being kept uncorrelated to their QCD production counterparts in the final fit. Figure 9.1 shows an example graph focusing on the uncertainties on the $f(\boldsymbol{\theta})_{QCD}$ transfer factor for the MTR category for the 2017 data (the corresponding summary for the VTR category is shown in Figure A.16) [10]. The effect of each of the previously discussed items translates into uncertainties of: 2-4 % for the PDF,

---

[7] $p_T^{Higgs} > 250$ GeV.

[8] For the rest of the Z and W transfer factors they are expected to cancel out due to similar jet topological properties between the SR and respective CRs

[9] With them being added in quadrature when forming the resulting uncertainty.

[10] Each of these uncertainties is implemented to be correlated across the m_{jj} bins for final fit.

1-9% for the factorization and 1-2 % for the renormalization variation. The EWK NLO corrections on the QCD V+jets production are implemented as one additional uncertainty (with them being a considered a full correction) [11].

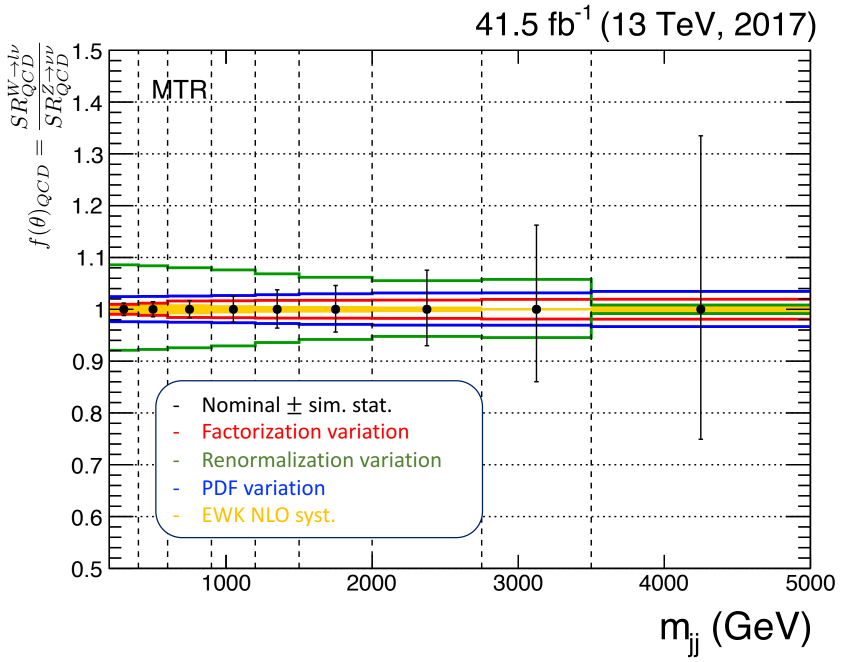


Figure 9.1: Theoretical uncertainties on $f(\boldsymbol{\theta})$ ratios for the MTR category and 2017 data, presented as a function of m_{jj} for the QCD production modes.

9.4.2 Experimental uncertainties

The experimental uncertainties trace their origin to the effects of the performance of the detector and the object reconstruction. Sources comprising this group include pileup re-weighting, trigger performance, jet energy corrections (both the scale and resolution) and object reconstruction and isolation criteria (expressed through the usage of selection/veto weights as explained in Chapter 6).

The uncertainties on the V+jets transfer factors are summarised in Table 9.1. Following the odering presented in the aforementioned table, this discussion can start with the trigger efficiencies. The uncertainty of the electron trigger efficiencies has been measured to have an 1% effect on the relevant electron transfer factors. The VBF and $E_{T,miss}$ triggers, being used to form muon CRs and SR, add 2 % and 10% uncertainties on the SR and muon transfer factors, respectively.

---

[11] The correction as treated as correlated between the numerator and denominator, which leads to a significant correlation [98]. These uncertainties (as well as the ones arising from the statistical precision of simulated samples) are treated as uncorrelated between m_{jj} bins.

The next group of uncertainties is associated with the lepton reconstruction, identification and isolation efficiencies taking into account uncertainties of the (p_T, η) scale factors introduced with Equation 6.5, resulting in per-lepton uncertainties for the muons (0.5 % for the identification and 0.1 % for the isolation efficiency) and electrons (0.5 % for the reconstruction and 3 % for the combined identification/isolation efficiency)[12]. Equivalent approach is taken for the veto weights introduced for the simulated background samples when defining the SR (as introduced with Equation 6.6). These include a set of uncertainties arising from tau (1%), muon (0.5%) and electron (3% from identification/isolation and up to 1.5 % from the reconstruction) vetos.

The last group of uncertainties relevant to the V+jets transfer factors is related to the uncertainty of the simulation modelling of $E_{T,miss}$ [112]. The dominating effects arise form the jet energy scale and resolution. Their effect is estimated by taking into account each of these corrections, which translates into variations of momenta within the jet collection and are, as a consequence, propagated to the $E_{T,miss}$ computation. Tests from the perspective of the analysis[13] led to the conclusion that these effects mostly cancel in the transfer factors leaving a residuum summarised in Table 9.1 where these uncertainties are separated by its source and the affecting transfer factor.

Lastly, minor contributions form other SM background such as the top, QCD multijet and diboson processes enter the likelihood function as shown in Equations 9.5-9.7. Their associated uncertainties are summarised in Table 9.2 denoting the affected SM process and the region in which the uncertainty becomes relevant.

9.5 QCD estimation

The commonly used method when estimating the contribution of QCD multijet processes in the SR is the Method A, introduced in Section 8.3.2. Its reliance on simulated multijet events and their lack of statistical precision led to the method's limited usability, especially when it comes to the VTR category.

This is a consequence of the core idea of establishing the connection between the QCD CR and SR trhough the use of the $r(m_{jj})$ factor, which is estimated using simulated events. This is illustrated in Figure 9.2, which shows an attempt to salvage this method by relaxing the $\Delta\phi_{jj}$ requirement in steps of 0.2 until 2.5 from the perspective of the MTR category.

---

[12] Being relevant for the corresponding SR to lepton CR transfer factors.

[13] Important as these changes can lead to a different set of jets being chosen as the desirable pair.

Source	Process	Uncertainty
Electron trigger	$W_{SR}/W_{e\nu}$, $Z_{\nu\nu}/Z_{ee}$	1%
$E_{T,miss}$triggers (MTR)	W_{SR}/W_{CR}, $Z_{\nu\nu}/Z_{CR}$, Z/W, signal	2%
VBF triggers (VTR)	W_{SR}/W_{CR}, $Z_{\nu\nu}/Z_{CR}$, Z/W, signal	10%
Muon-ID efficiency	$W_{SR}/W_{\mu\nu}$, $Z_{\nu\nu}/Z_{\mu\mu}$	0.5% (per muon)
Muon-Iso efficiency	$W_{SR}/W_{\mu\nu}$, $Z_{\nu\nu}/Z_{\mu\mu}$	0.1% (per muon)
Electron-reco efficiency	$W_{SR}/W_{e\nu}$, $Z_{\nu\nu}/Z_{ee}$	0.5% (per electron)
Electron-IDiso efficiency	$W_{SR}/W_{e\nu}$, $Z_{\nu\nu}/Z_{ee}$	3% (per electron)
Electron veto from reco	W_{SR}/WCR, Z/W	1% (QCD), 1.5% (EW)
Electron veto from idiso	W_{SR}/WCR, Z/W	3%
Muon veto	W_{SR}/WCR, Z/W	0.5%
Tau veto	W_{SR}/WCR, Z/W	1%
Jet energy scale	Z/W	1–2%
	W_{CR}/W_{SR}	1.0–1.5%
	$Z_{CR}/Z_{\nu\nu}$	1%
Jet energy resolution	Z/W	1.0–2.5%
	W_{CR}/W_{SR}	1.0–1.5%
	Z_{CR}/Z_{SR}	1%

Table 9.1: Summary of experimental uncertainties on the transfer factors for main V+jets backgrounds. Where specified in the form of a range of values, the values vary with era of data taking (or transfer factors).

Source	Process	Uncertainty
Luminosity	All	$\approx$ 2.5 %
Pile-up	All	up to 3 %
Electron trigger	All in Z(ee) and W(eν)	1 %
$E_{T,miss}$ triggers	All in Z($\mu\mu$),W($\mu\nu$) and SR (MTR)	2 %
VBF triggers	All in Z($\mu\mu$),W($\mu\nu$) and SR (VTR)	10 %
Level-1 pre-fire	All (2017 only)	3 %
Muon-ID efficiency	All in Z($\mu\mu$),W($\mu\nu$)	up to 2 %
Muon-Iso efficiency	All in Z($\mu\mu$),W($\mu\nu$)	up to 0.5 %
Electron-reco efficiency	All in Z(ee) and W(eν)	up to 1 %
Electron-IDiso efficiency	All in Z(ee) and W(eν)	up to 3 %
Electron veto from reco	All in SR	up to 1.5 %
Electron veto from IDiso	All in SR	up to 5 %
Muon veto	All in SR	up to 0.5 %
Tau veto	All in SR and CR	up to 1 %
Jet energy scale	All	10%
Jet energy resolution	All	0.5–5%
QCD estimation	QCD MTR (VTR)	40 (10)%

Table 9.2: Summary of experimental uncertainties affecting smaller backgrounds.

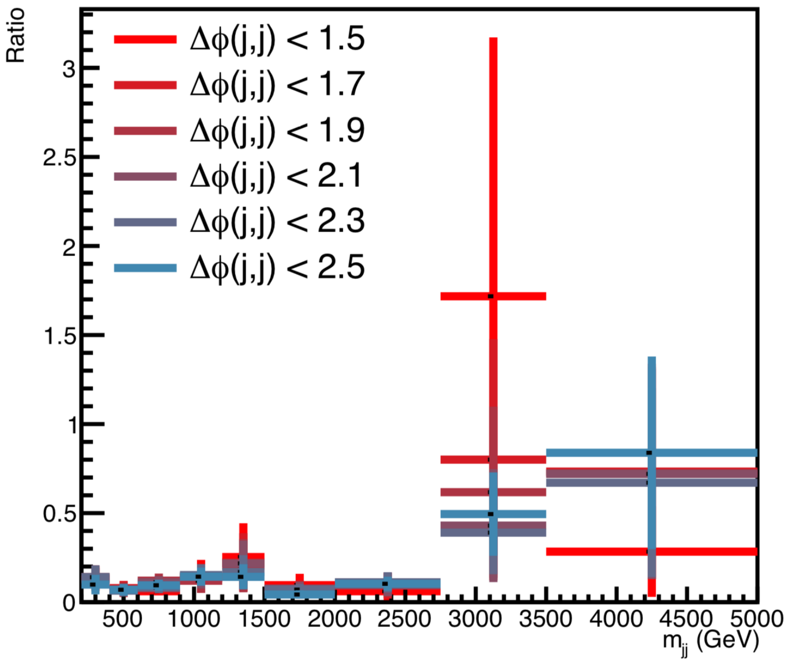


Figure 9.2: Various distributions of the ratio $r(m_{jj})$ shown for a range of $\Delta\phi_{jj}$ thresholds.

The alternative method, method B, was introduced to mitigate the issues which arise due to the dependency on simulation samples. It instead, as defined in Section 8.3.3, relies on a mostly data driven approach when estimating the final multijet contribution. This is achieved by using a data driven method of estimating the normalisation of the multijet processes in the SR by fitting the $min\Delta\phi(j, E_{miss})$ variable for data and minor backgrounds. This in return gives a function describing the behaviour of the multijet processes in QCD CR: $F_{QCD}(x) = F(x) - F_B(x)$[14].

This function can be extrapolated to the SR and integrated within the SR defined region of $min\Delta\phi(j, E_{miss})$, giving an estimate on the normalisation. The final step is to extract the m_{jj} shape of multijet processes from the QCD CR and scale it to the correct SR normalisation. This is performed under the assumption that the QCD CR defining variable ($min\Delta\phi(j, E_{miss})$) and m_{jj} are not strongly correlated[15].

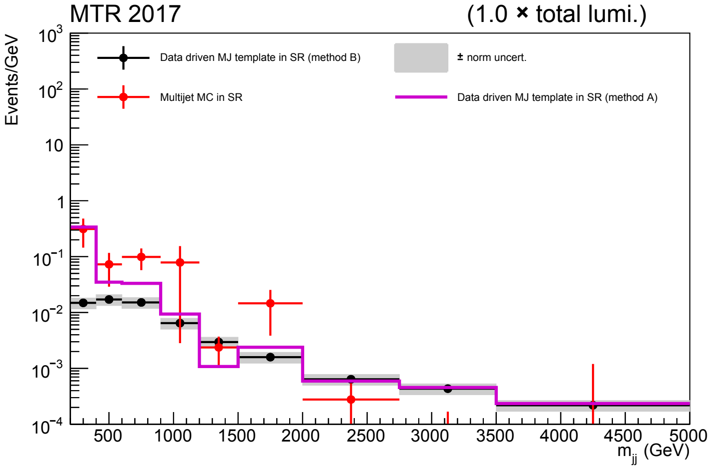
The final prediction from this method, obtained by scaling the estimation of the multijet m_{jj} shape from the QCD CR to the SR normalisation, is shown in Figure 9.3 for the MTR category for both 2017 and 2018 data. A comparison with the results originating from Method A is also presented (magenta line) as well as the prediction from QCD multijet simulated events (red). As the method relies on the extended CR range enabled though the unblinding of 20 % of data, the final shape needed to be further scaled up by a factor of 5 until the analysis was fully unblinded. Lastly,

---

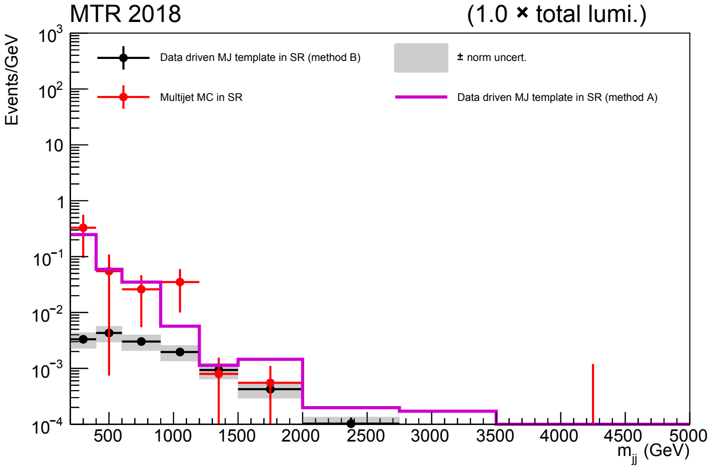
[14] Using the notation introduced in Section 8.3.3, where x = $min\Delta\phi(j, E_{miss})$.

[15] This assumption is further supported by the small variation of the value of $r(m_{jj})$ for bins with higher statistical precision (as indicated with Figure 9.2).

the uncertainty of this prediction arises from the normalization fit error and stands at 40 % (10 %) for MTR (VTR) category. This is propagated as a flat uncertainty on this prediciton when it is passed on as an input for the signal extraction fit.



(a) MTR 2017



(b) MTR 2018

Figure 9.3: Final estimation of the SR contribution originating from QCD multijet processes obtained through the use of Method B for the MTR category for both (a) 2017 and (b) 2018 data.

9.6 Results

The signal extraction procedure introduced in previous sections is applied to all regions simultaneously, allowing the fit to access all available data taking periods and categories. This section summarises final results, expressed in terms of the 95 % CL upper limit on Br(H→ inv)[16] for all analysis categories. In order to formulate a preliminary look at the combined Run 2 (or the ”legacy”) result, a combination is performed with the studies targeting the 2016 data without re-analysing the data (explained in great detail in Refs. [39, 41]).

Treatment of the most important nuisance parameters, such as the ones related to lepton efficiencies, have been left uncorrelated between the three years with the uncertainties assigned to the jet energy scale and resolution also following the uncorrelated path reflecting the different operational conditions of the CMS experiment. A correlation between theory uncertainty has been established between 2017 and 2018 (although they differ between MTR and VTR categories). Lastly, the addition of the photon region has been performed for the MTR category for the 2017 and 2018 eras of data taking.

Starting with the main analysis category, Figures 9.4 and 9.5 show post-fit distributions for the MTR CRs. These distributions show the data versus simulated background composition for the lepton/photon regions (with the background contributions being presented with their post-fit predictions). The ratio panel is used to show data to simulation ratio for both the pre-fit (red) and post-fit (black) scenarios. The corresponding distributions for the VTR category are shown in Figures 9.6 and 9.7. Concluding this set, post-fit distributions of the SR for all categories and data taking periods are shown in Figure 9.8[17]. Tables A.2-A.3 (A.4-A.5) present the final yields in the SR for the MTR (VTR) category. The summary of the main nuisance parameter impacts for the combined fit (depicting 2017 and 2018 data) are shown in Figures 9.9-9.12. The leading nuisance parameters (for both MTR and VTR categories), ordered by their influence over the Br(H→inv), are the ones related to statistical precision of the simulation samples, theory corrections and trigger scale factors. Looking into the per-category summaries shown in the aforementioned figures, it can be seen that the parameter that is most significantly pulled by the fit for both the MTR 2017 and 2018 categories (”qcd_photon_ewk_vbf_*”)[18] is related to the theory uncertainties associated to the photon CR. The necessity for the pull came from the fit procedure adjusting the prediction to the reality seen in data,

---

[16] As no significant deviations from the SM have been observed.

[17] The post-fit distributions from the fit procedure which uses the ”CR-only” approach (by using the data from CRs only) are shown in Figures A.17-A.21.

[18] With the pull being defined as the difference between the post and the pre-fit value of a nuisance parameter divided by its uncertainty.

but it is still low enough (∼1 σ) to raise any concerns regarding the photon region. For the MTR 2018 category, there is one more parameter that is largely pulled by the fit - ”vbf_2018_stat_error_qcd_photonCR_*”, related to the statistical precision of the photon region, but its importance to the Br(H→inv) is small enough that its large pull isn't an indication of a problem. For the VTR category, there aren't any highly pulled nuisance parameters. One difference for these categories (as compared to the MTR) is that the nuisance parameters related to the trigger algorithms have a much larger role from the Br(H→inv) point of view.

The fit procedure, following the CLs technique, yiled a 95 % CL upper limit on the Br(H→ inv), under the assumption of a SM Higgs boson with m_H =125.09 GeV. For the analysis focusing on the 2017 and 2018 data taking periods (with a total luminosity of 101.2 fb^{-1}), it places an observed (expected) 95 % CL upper limit on Br(H→inv) to be 0.15 (0.13). The combination with the study focusing on the 2016 data states a limit of Br(H→inv) < 0.15 (0.12). The per category limits and subsequent combinations (for 2017 and 2018 data) are summarised in Table 9.3.

A discrepancy between the observed and the expected limit can be seen in 2017 data results where it appears in both the independent, MTR and VTR, categories (for VTR the difference is slightly larger than 1σ). The reason behind this could be an additional source of background that was not properly accounted for (possibly seen in the excess in relevant SR m_{jj} distributions for m_{jj} > 2000 GeV)[19]. The supporting fact that this is not an appearance of signal is that the ”observed to estimated” limit agreement for the 2018 era is much better across both categories[20].

Category	Observed	Expected	1-σ interval	2-σ interval
MTR 2017	0.26	0.22	[0.16 – 0.31]	[0.12 – 0.43]
VTR 2017	0.81	0.53	[0.38 – 0.76]	[0.29 – 1.04]
MTR 2018	0.17	0.17	[0.12 – 0.24]	[0.09 – 0.32]
VTR 2018	0.33	0.33	[0.23 – 0.47]	[0.18 – 0.64]
MTR 2017 2018	0.15	0.14	[0.10 – 0.20]	[0.08 – 0.27]
all 2017	0.28	0.21	[0.15 – 0.30]	[0.12 – 0.41]
all 2018	0.15	0.15	[0.11 – 0.22]	[0.08 – 0.30]
all 2017 2018	0.15	0.13	[0.10 – 0.19]	[0.07 – 0.25]
Run2	0.15	0.12	[0.08 – 0.16]	[0.06 – 0.22]

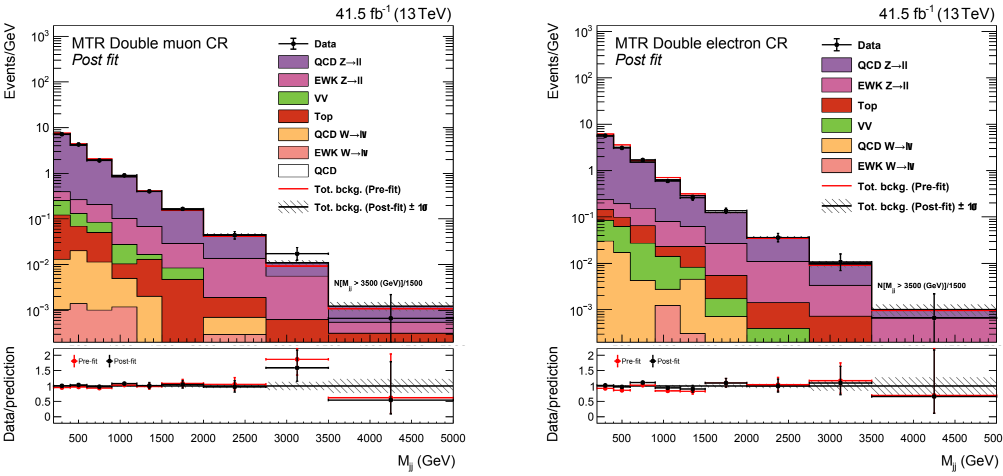
Table 9.3: Summary of results expressed as 95 % CL upper limit on the Br(H→inv). Contributions from each category and their subsequent combinations are presented, culminating with the result combining all Run 2 studies.

The size of the contribution from different sources of uncertainties can be obtained by freezing groups of nuisance parameters (being grouped by their common origin):

---

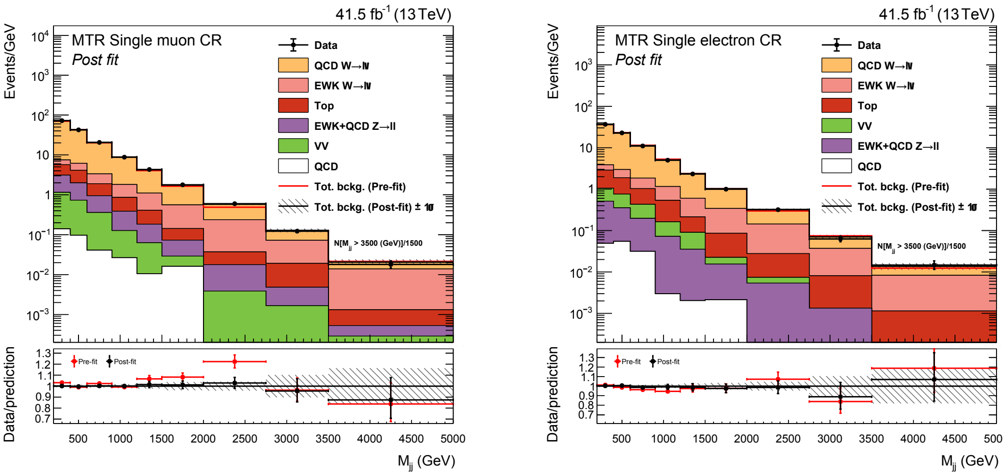
[19] Similar discrepancy was observed in the study focusing on the 2016 data [39].

[20] Detailed study of the 2017 era is currently ongoing within the CMS collaboration.



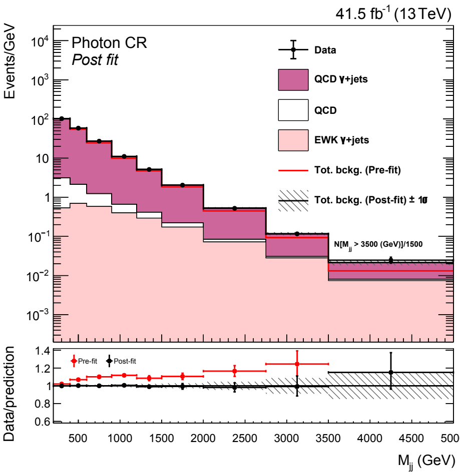
(a) Dimuon CR

(b) Dielectron CR



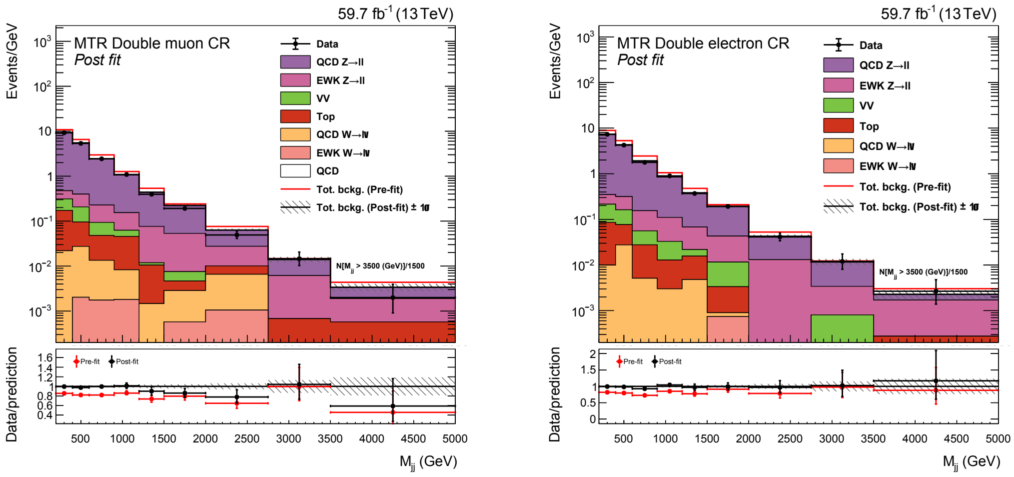
(c) Single muon CR

(d) Single electron CR

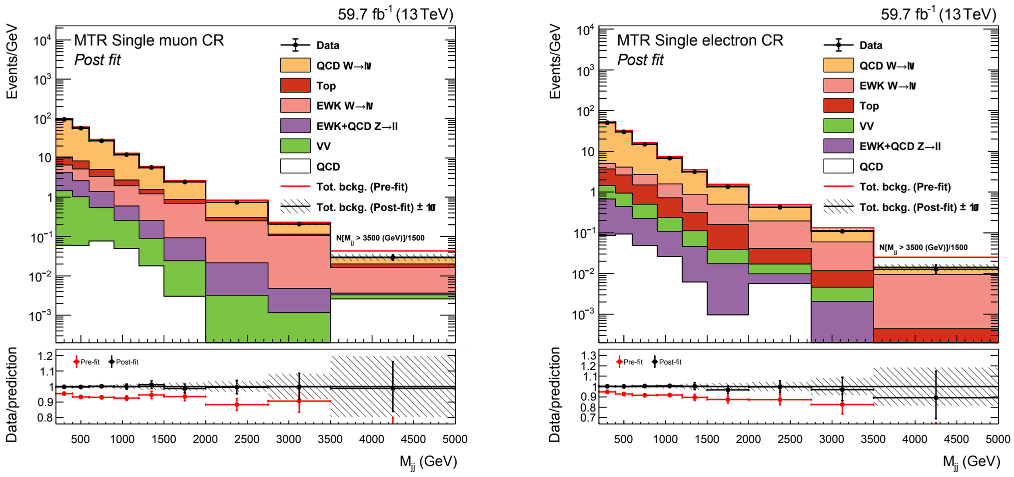


(e) Photon CR

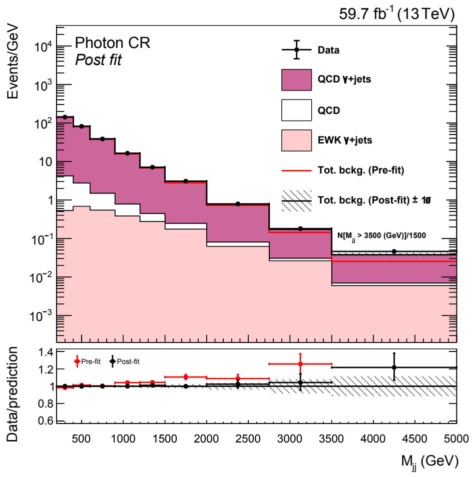
Figure 9.4: Post-fit distributions for 2017 data, showing the: (a) dimuon, (b) dielectron, (c) single muon, (d) single electron and (e) photon CR region.



(a) Dimuon CR (b) Dielectron CR

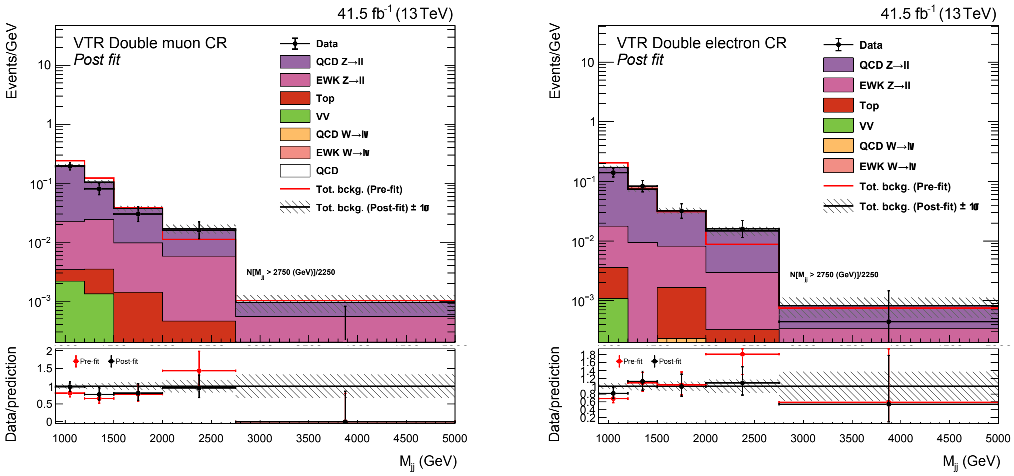


(c) Single muon CR (d) Single electron CR



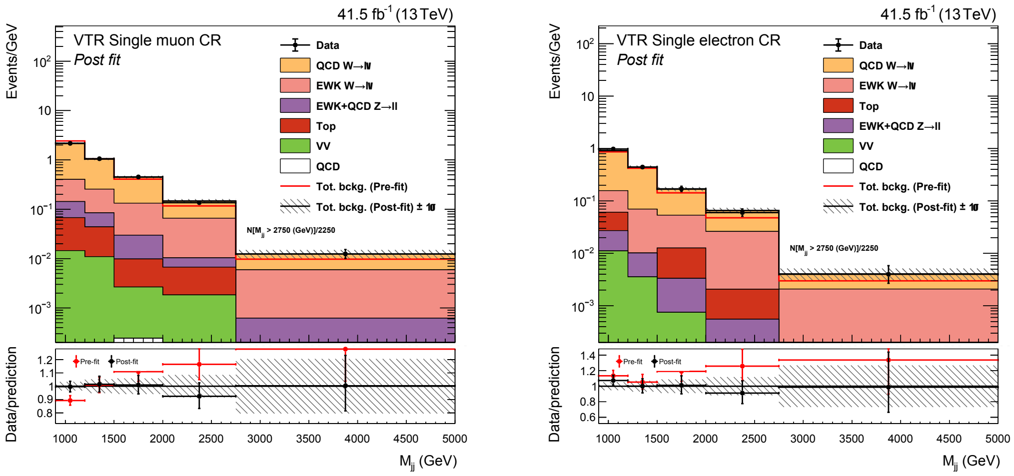
(e) Photon CR

Figure 9.5: Post-fit distributions for 2018 data, showing the: (a) dimuon, (b) dielectron, (c) single muon, (d) single electron and (e) photon CR region.



(a) Dimuon CR

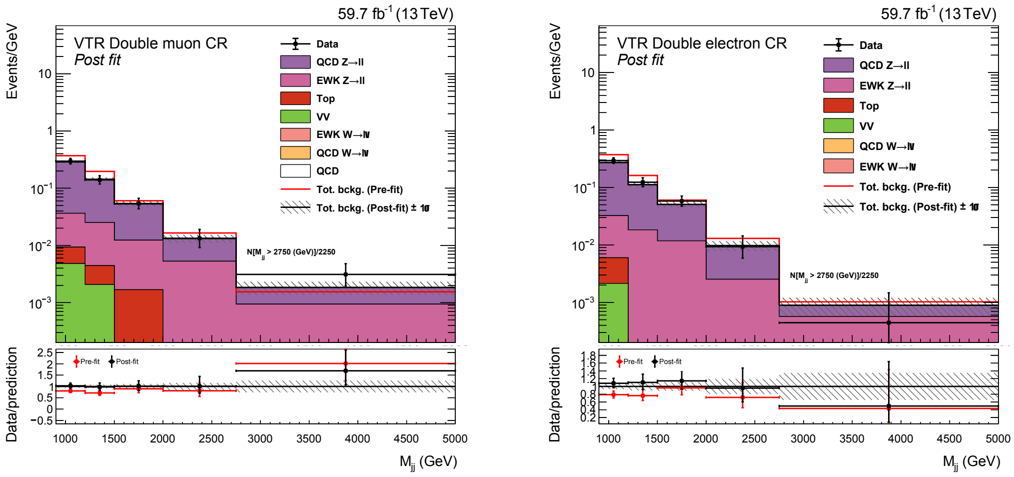
(b) Dielectron CR



(c) Single muon CR

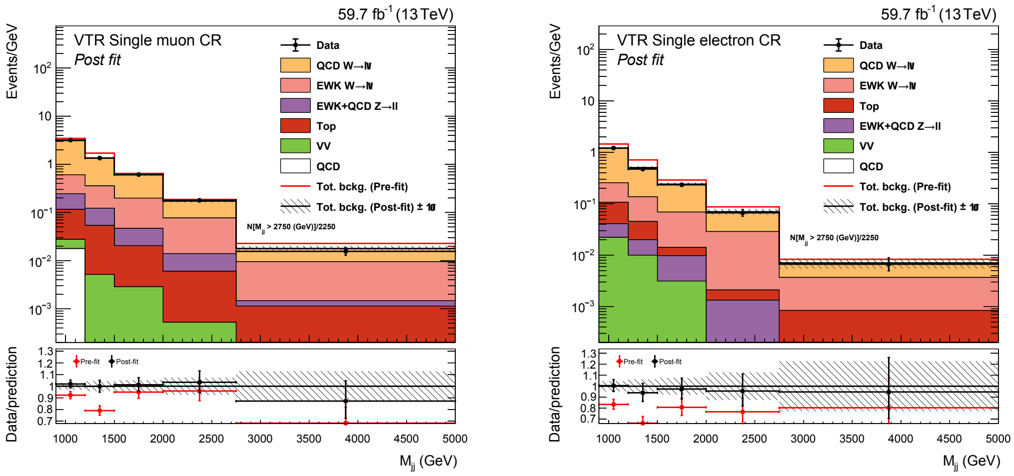
(d) Single electron CR

Figure 9.6: Post-fit distributions for 2017 data, showing the: (a) dimuon, (b) dielectron, (c) single muon an (d) single electron region.



(a) Dimuon CR

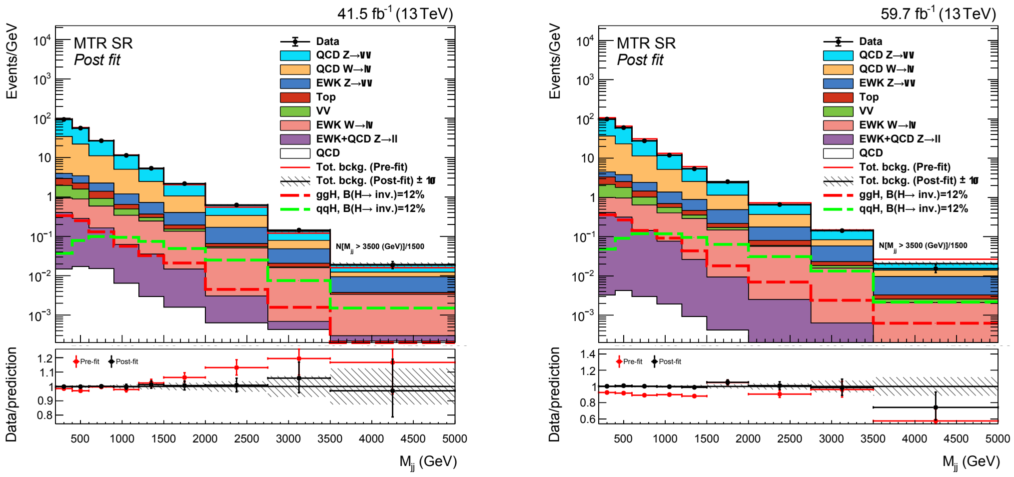
(b) Dielectron CR



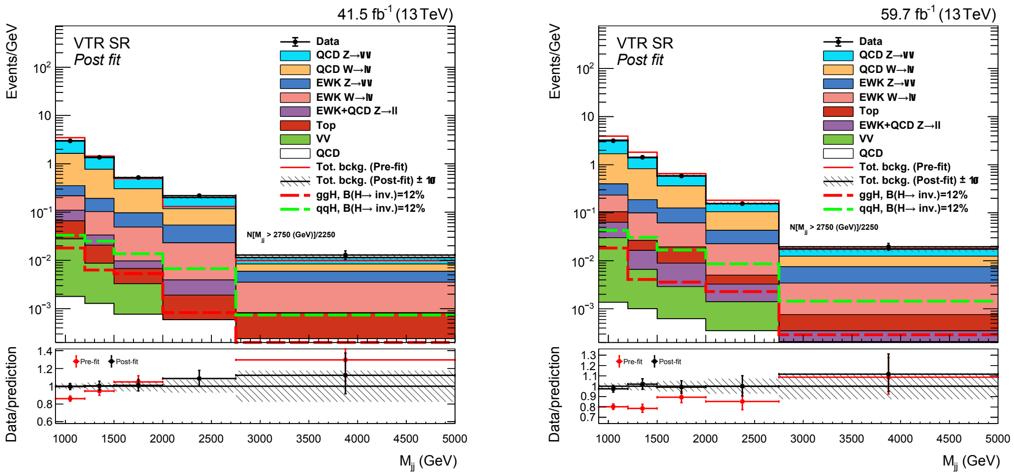
(c) Single muon CR

(d) Single electron CR

Figure 9.7: Post-fit distributions for 2018 data, showing the: (a) dimuon, (b) dielectron, (c) single muon an (d) single electron region.



(a) MTR 2017 (b) MTR 2018



(c) VTR 2017 (d) VTR 2018

Figure 9.8: Post-fit distributions for the SR, showing the: (a) MTR 2017, (b) MTR 2018, (c) VTR 2017 an (d) VTR 2018.

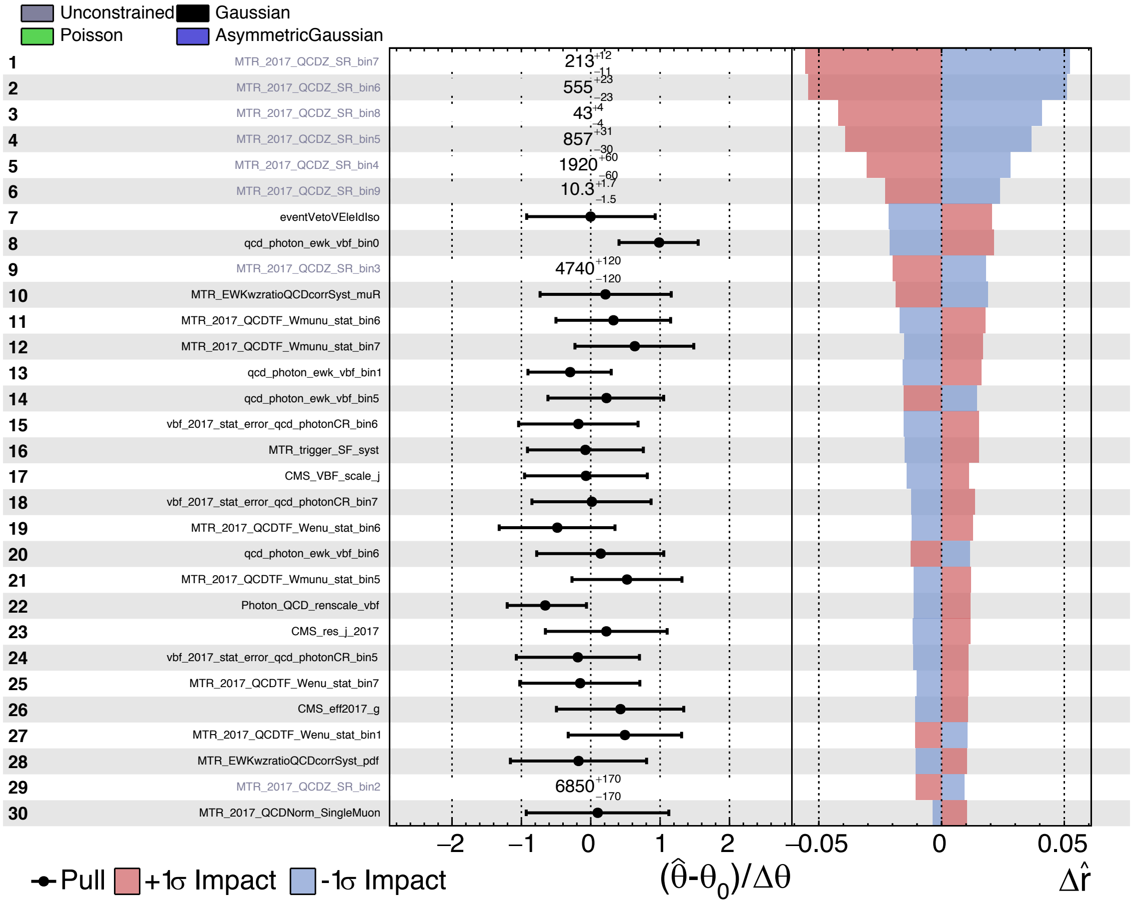


Figure 9.9: Impacts of the nuisance parameters from the final fit for the MTR category for 2017 data. The left panel shows the difference between the post and pre-fit value of the nuisance parameter divided by its pre-fit uncertainty. The parameter Δr in the right panel shows the difference between the the best fit value of Br(H→inv) after setting the given nuisance parameter at $\pm 1\sigma$ of its nominal value.

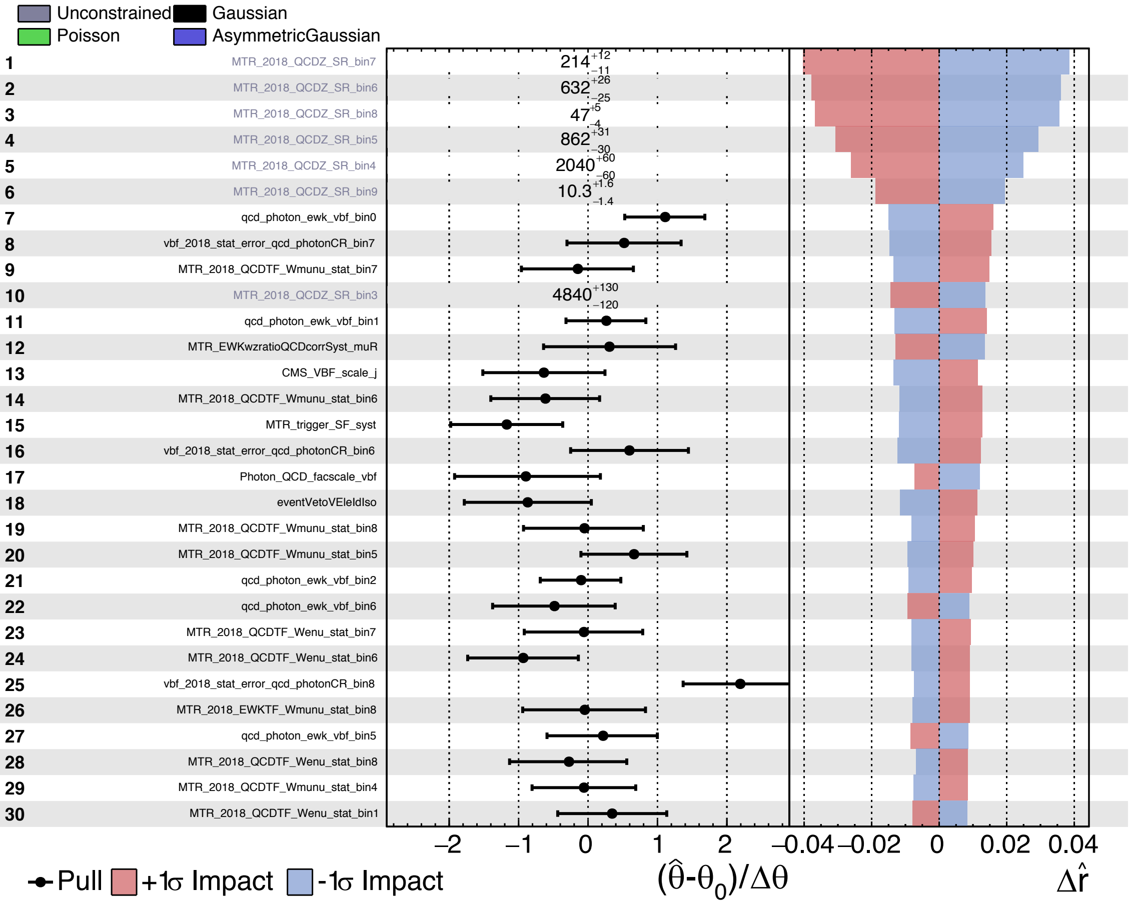


Figure 9.10: Impacts of the nuisance parameters from the final fit for the MTR category for 2018 data. The panel details follow the convention introduced in 9.9.

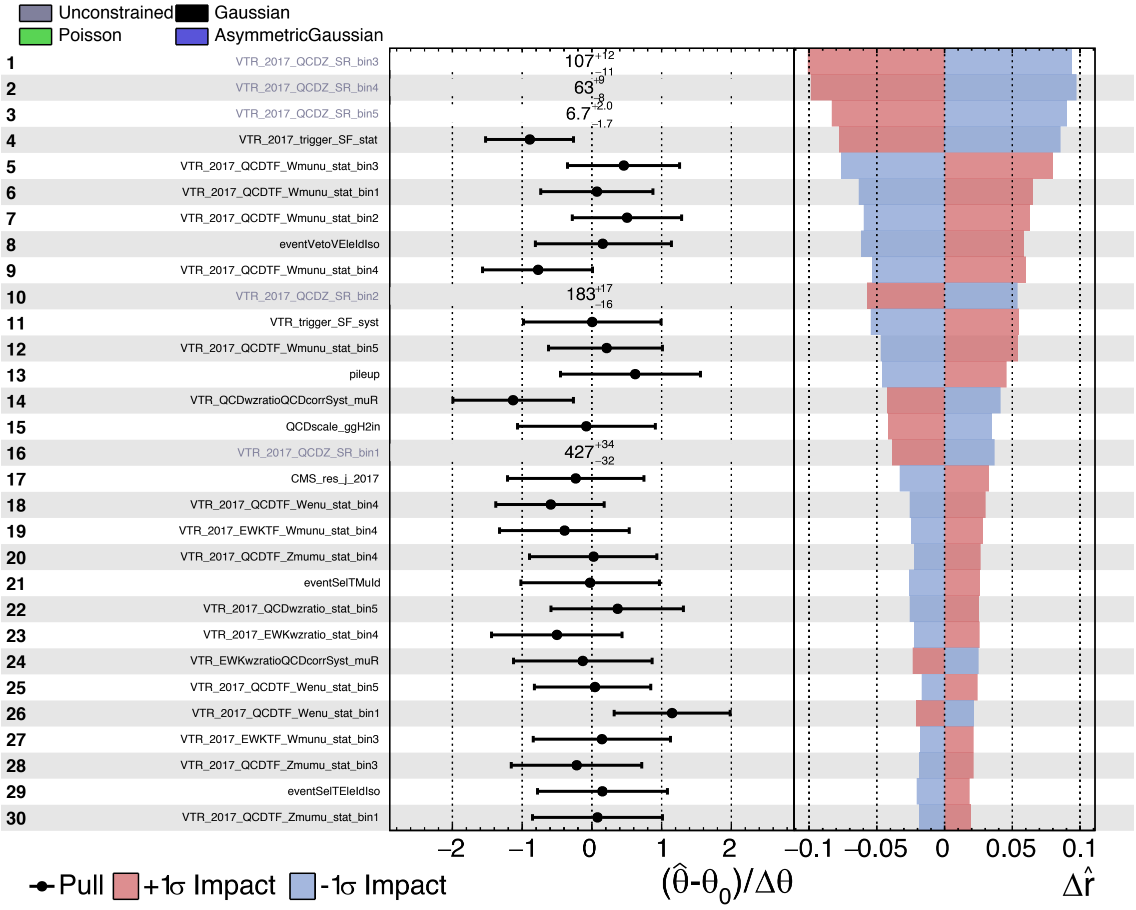


Figure 9.11: Impacts of the nuisance parameters from the final fit for the VTR category for 2017 data. The panel details follow the convention introduced in 9.9.

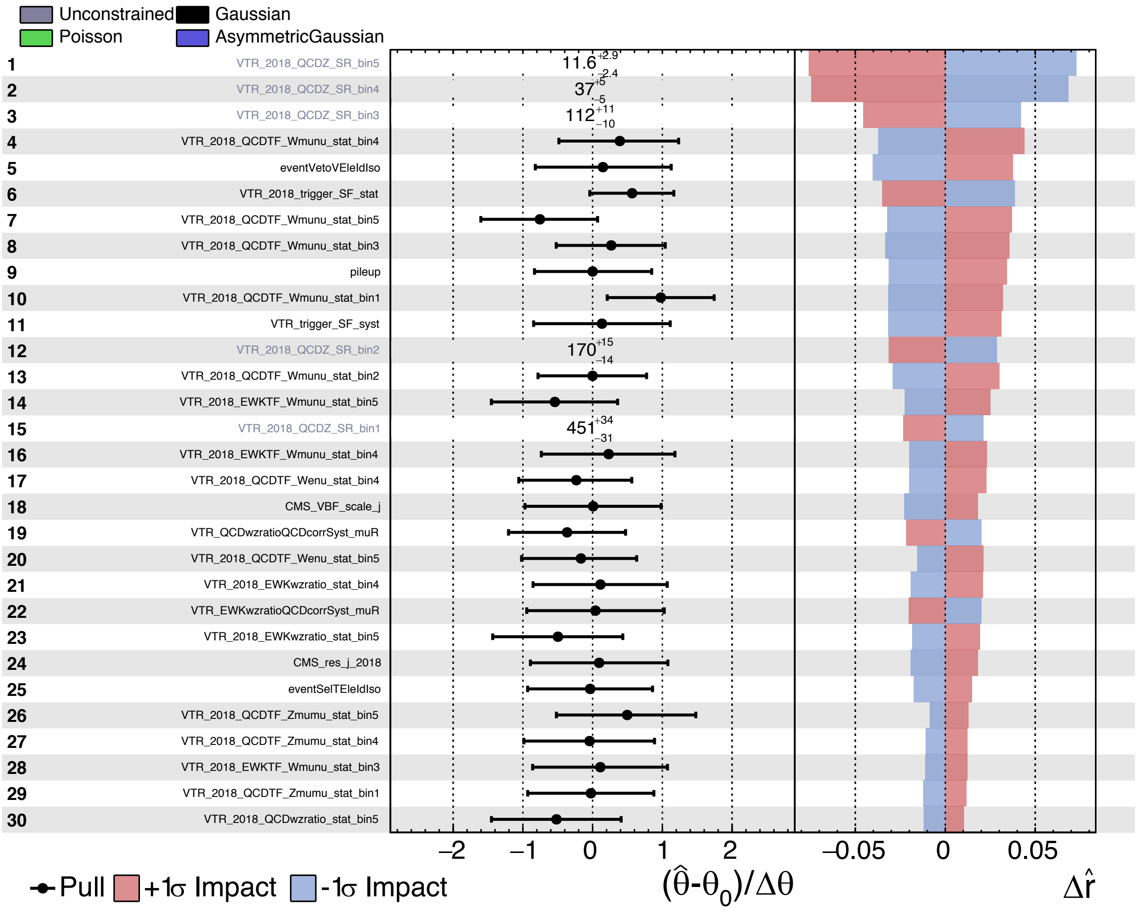


Figure 9.12: Impacts of the nuisance parameters from the final fit for the VTR category for 2018 data. The panel details follow the convention introduced in 9.9.

theory (as introduced in Section 9.4.1), statistical precision of simulation samples, lepton/photon efficiencies, jet calibration (jet energy scale and resolution), trigger and other (covering the multijet estimate, effects on smaller backgrounds, pile-up, etc). This leads to the following result expressed for the best fit[21]:

$$\begin{aligned}\mathrm{Br}(\mathrm{H}\to\mathrm{inv}) &= 0.045^{+0.060}_{-0.061}\\ &= 0.045^{+0.029}_{-0.030}(\mathrm{theory})^{+0.028}_{-0.028}(\mathrm{sim.stat.})^{+0.014}_{-0.15}(\mathrm{lepton/photon\ eff})\\ &\pm 0.004(\mathrm{jet\ calib.})^{+0.021}_{-0.022}(\mathrm{trigger})^{+0.017}_{-0.015}(\mathrm{other}) \pm 0.032(\mathrm{stat.}),\end{aligned} \tag{9.8}$$

where it can be seen that this analysis is systematically limited with the main contribution originating from theoretical uncertainties and the statistical precision of simulation samples. This is graphically summarised in Figure 9.13 (while a corresponding breakdown for the best "blinded" fit can be found in Figure A.22).

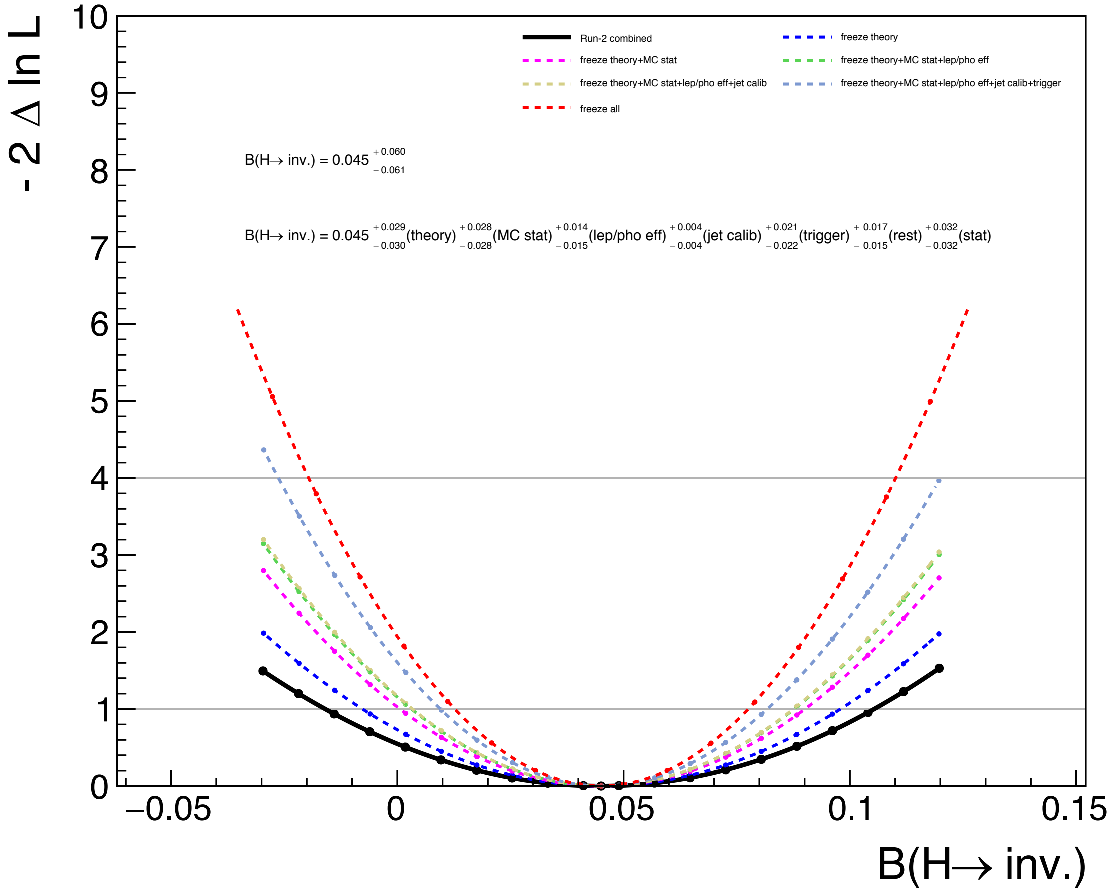


Figure 9.13: Likelihood scan for the Run 2 combination for the best fit Br(H→inv) = 0.045, with scans obtained by sequentially freezing the groups of nuisance parameters.

---

[21] The SM prediction for the Br(H→inv) stands at ∼ 0.1 %.

Finally, a comparison with direct DM detection experiments using the procedure introduced in Chapter 2 is made. The observed limit for the combined measurement describing the Run 2 phase is presented as a 90% CL upper limit in order to be comparable with the results from direct DM searches following the previously introduced strategy. This result is shown in Figure 9.14 comparing the measurement form the CMS experiment with those coming from LUX [134], CDMS [135], XENON [136, 137], CRESST [138] and PandaX-II [139] collaborations. Following the introduction of Equations 2.3-2.4 for the scalar and fermion scenario, it can be seen that the limitation of the used method, seen in both lines for $m_{DM} \approx \frac{m_H}{2}$, is arising due to the factor $\beta = \sqrt{1 - \frac{4m_{DM}^2}{m_H^2}}$. On the other end of m_{DM}, the mass range stops at 1 GeV due to the assumed Higgs portal model not being sustainable for DM particles located within the mass order of quark masses. Lastly, the difference between the two curves (red and orange) is arising from the different dependencies on the m_{DM} assumed by the model, being best seen in Equations 2.3 and 2.4.

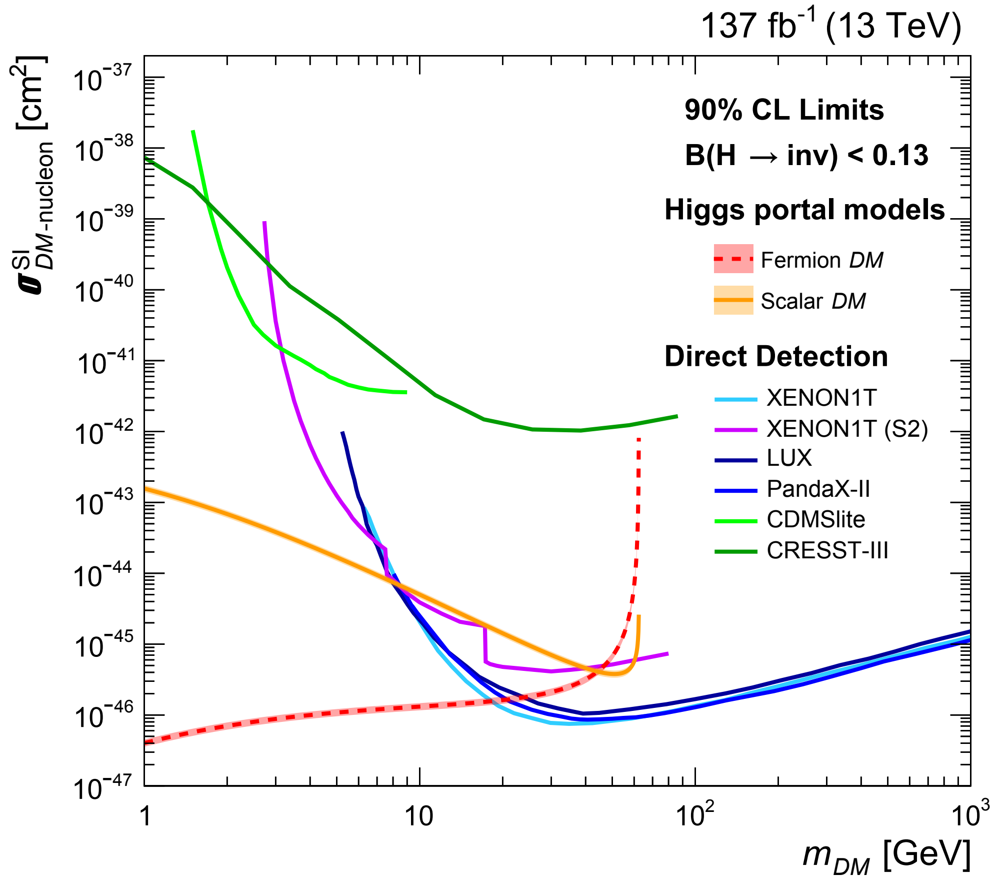


Figure 9.14: The reinterpretation of the CMS results in terms of the 90% CL upper limits on the spin-independent DM-nucleon scattering cross section when assuming a fermion (red) or a scalar (orange) DM particle (presented as a function of m_{DM}). Limits are compared with results originating from direct DM detection experiments.

9.7 Summary

The presented study covered the search for the invisible decays of the Higgs boson, where the production mode in question is VBF. The study was performed using the 101.2 fb^{-1} of data collected by the CMS experiment corresponding to the 2017 and 2018 years of operation. Exploration of additional phase space found in lower $E_{T,miss}$ range was enabled through the introduction of a new analysis category based on new trigger algorithms tailored to look for the VBF characteristics in events. No deviations from the SM have been observed. The result is interpreted as the observed (expected) 95% CL upper limit on the branching ratio of the Higgs boson decaying invisibly and it stands at: Br(H→inv) = 0.15 (0.13). A combination with previous measurements targeting the VBF topology during the Run 2 phase is presented, bringing the total integrated luminosity to 137.1 fb^{-1}. The observed (expected) value of the Br(H→inv) for the entire Run 2 phase is found to be 0.15 (0.12). Figure 9.15 summarises the results of the individual measurements and the subsequent combination of categories.

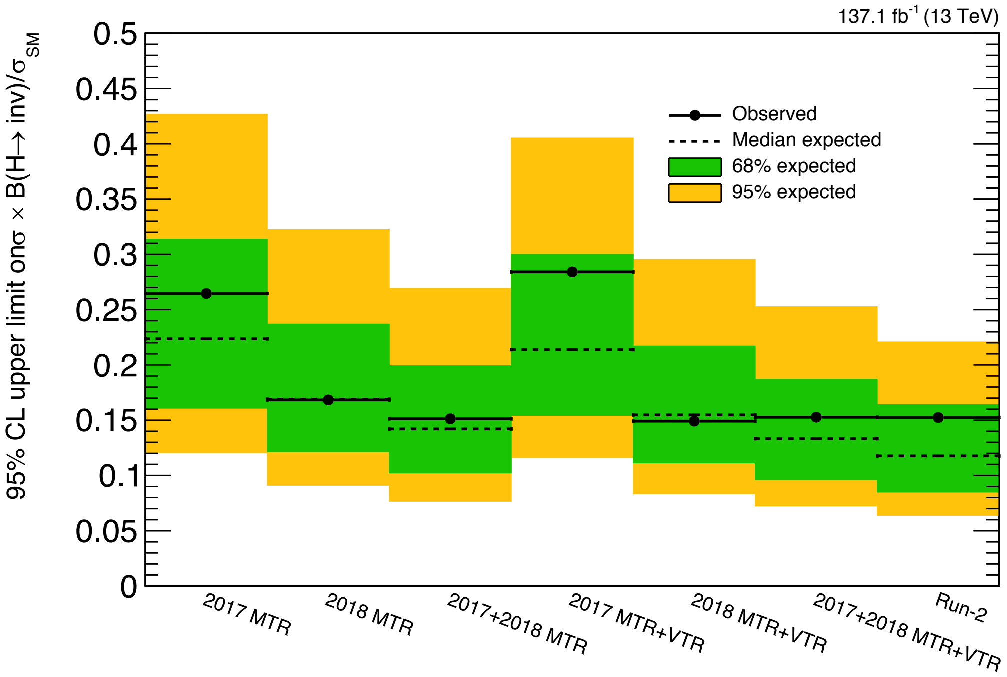


Figure 9.15: Final limits for the Run 2 phase showing the contribution from each of the categories.

Chapter 10

Conclusion

> *“Of everything that man erects and builds in his urge for living, nothing is in my eyes better and more valuable than bridges. They are more important than houses, more sacred than shrines. Belonging to everyone and being equal to everyone, useful, always built with a sense, on the spot where most human needs are crossing, they are more durable than other buildings and they do not serve for anything secret or bad.”*
>
> — Ivo Andrić

10.1 Introduction

The following sections serve as a brief introduction to the ongoing work focused on combining the efforts from all analyses searching for the invisible final state of the Higgs boson[1]. Finally, in order to have a complete narrative when describing the H→inv studies, there is one more era that needs covering - the future. The main idea and the previously obtained results have been the focus of Part 1 of this thesis, while the present has been represented with all the Run 2 efforts covered in Parts 2 and 3. In order to come to a proper conclusion, this chapter is going to contain a discussion about studies of future prospects for this channel with respect to the later phases of the LHC and corresponding upgrades of the CMS detector.

---

[1] Summarising the Run 2 phase of operation. The previous combination summarising early Run 2 (2015 and 2016) and Run 1 is detailed in Ref. [39].

10.2 The grand combination

The VBF production mode represents the leading channel in terms of sensitivity towards the invisible final state of the Higgs boson. As introduced in Chapter 2, the other production modes have been explored in order to have the best possible coverage of interesting phase space. The approach taken early on in the development of these analyses for the 2017-18 data taking period was to create exclusive channels which will be sorted by the importance towards the final state. This has led to the VBF taking the prime spot within the event processing cycle, and if an event didn't satisfy the conditions of the VBF channel it would begin its journey towards the non-VBF clusters of categories.

The non-VBF categorisation follows a similar strategy to the older approaches, but modifies the traditional channel ordering based on the production mode in question (namely ggH, VH, and ttH) in favour of subcategories whose expected sensitivity is estimated through an optimisation technique based around the $\mathrm{S}/\sqrt{\mathrm{B}}$ criteria[2]. Having designed the event categorisation in such an orthogonal way there is one more important aspect which, if properly approached, would make for a smooth combination. This aspect is seen in the way these categories share the treatment of uncertainties. Sharing of as many input parameters which form these analyses allows for a by design level of consistency usually seen only after a hard work of connecting various parameters and inputs through the scan of inputs from different analysis teams using different software packages. This is where the more technical aspect comes to light - nothing is stopping the analysers from taking this combined analysis strategy and implementing it in a novel analysis framework. One larger effort could also help overcome general problems affecting HEP analyses as a whole (large time intervals needed when processing datasets, different output files used by various teams, rigid software having a steep learning curve, etc).

This is where the Faster Analysis Software Taskforce (FAST) [140] framework comes into focus. It represents the main data processing software used for the purposes of these studies. It is based on standard python libraries which allow for the usage of dataframes, array techniques and easy to read/write configuration files[3]. This approach bridges a connection between industry standards and science. One of the benefits arising from this sort of data processing is the lightweight output in

---

[2] The S and B parameters denote the overall signal and total background yields, where the expected sensitivity term connects to the fact that this is an estimation relying purely on simulated samples.

[3] The inclusion of different array techniques allows for the removal of an event loop, instead relying on a "chunk" of data being loaded and operated on at the time. This overcomes the main issue which comes in mind of many when choosing more traditional C++ based processing tools - processing speed

the form of standard dataframe formats and its support for modular designs.

This embrace of modular design philosophy is especially important when it comes to previously discussed combination efforts as the main analysis framework was built to explore all the benefits of this data processing approach. All channels would be dependent on a set of core software modules creating base analysis object collections, while all other specifics can be implemented through a set of custom modules, which can simply be added in the chain without much effort. Finally, this approach brings the usage of configuration files which are based on the easy to read data-serialisation language - YAML [141]. They are basically used to summarise which modules are deployed in the analysis, how the important regions are defined and what output is needed. This reduces the debugging time significantly by keeping all core analysis inputs defined in a single location.

The previously described VBF H→inv study represents the first step in the combined analysis. The sensitivity given by the VBF analysis (and its comparison with other channels presented in Chapter 2) can serve as a good figure of merit regarding what can be expected from other channels, which due to a more detailed categorisation are expected to yield a better result compared to what would be achieved by keeping the old strategy. These non-VBF studies are currently in progress and are the main focus of Ref. [142], where this approach is presented in much more detail.

10.3 A look into the future

Due to its strong dependence on forward jets and $E_{T,miss}$, the VBF H→inv analysis represents a good way to test the potential sensitivity gains and modified reconstruction algorithms arising with upgrades of the CMS detector for various operational conditions of the LHC which are expected to be reached in the future (such as the HL-LHC phase [143]). This is the purpose of studies published in Ref. [51] where, on the altruistic side, the performance of reconstruction algorithms designed to incorporate upcoming upgrades (such as the HGCal upgrade [144]) is tested. On the more analysis oriented side, they can be used to test how the expected sensitivity of the current approach scales with the size of the dataset and to see if these strategies are still viable for upcoming phases.

A simulation study of H→inv analysis prospects was performed for three different values of total integrated luminosity: $L = 300$, 1000 and 3000 fb^{-1}. The simulation of detector effects was performed using the Delphes software package [145] (which mimics the upgraded, Phase 2, state of the CMS experiment[4]). The analysis strategy

---

[4] This includes the increase in the number of pile-up interactions to 200 and the expected increase

approached here was to use the main MTR analysis category (defined similarly to what is described in Chapter 7). The assumption made at the beginning was that the $E_{T,miss}$ triggers would perform in such a rate-controlled way that they would allow for even lower $E_{T,miss}$ thresholds than what was used during the Run 2 phase. Figure 10.1 shows the background composition in the signal region for two main variables of interest: the $E_{T,miss}$ and the dijet mass.

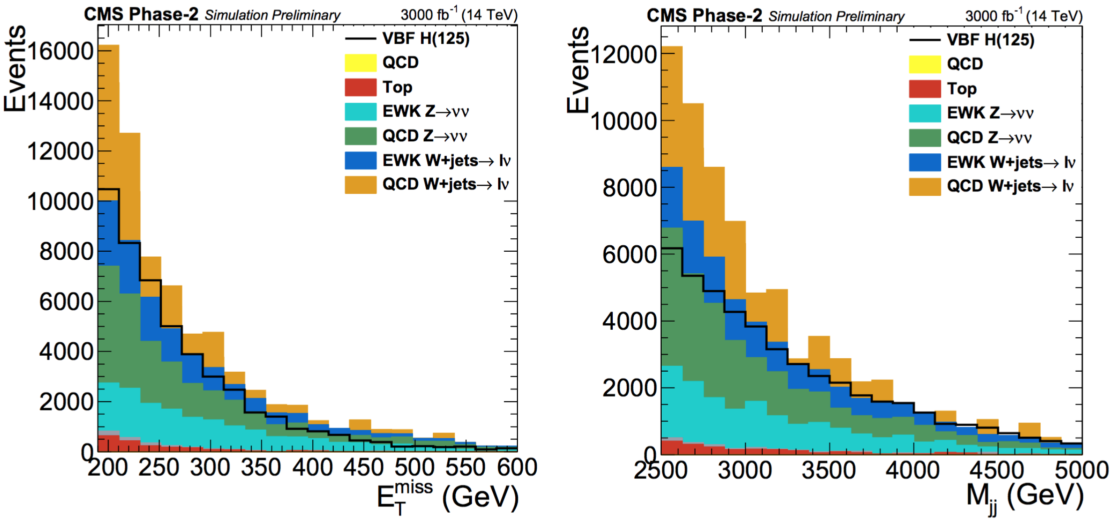


Figure 10.1: Composition of background processes overlayed with the signal simulation in the SR for $E_{T,miss}$ (left) and m_{jj} (right) [51].

An optimisation was performed varying the ranges of these two variables with the purpose of obtaining the best constraint on the invisible final state. The resulting 95% CL upper limits expressed as a function the the respective $E_{T,miss}$ thresholds are shown in Figure 10.2 for three scenarios of total integrated luminosity[5]. By focusing on the minima for each of three scenarios it can be seen that the expected limit value does not decrease significantly simply due to an increase in total integrated luminosity. This indicates that, besides a constant improvement of the theoretical uncertainty treatment, this analysis needs to develop a better approach (akin to the one taken in 2017-18 period with a dedicated VBF trigger) in order to further increase the sensitivity towards the invisible final state. The best reported constrain on the Br(H→inv) is obtained to be 3.8 % (expected for the scenario with $L = 3000$ fb^{-1}).

A similar simulation study has been performed by the ATLAS collaboration, this time targeting the the second most sensitive channel - the VH production mode.

---

in the operational energy to 14 TeV.

[5] For a m_{jj} threshold which resulted from optimisation of each of these scenarios. The treatment of uncertainties follows the Run 2 approach albeit with small differences when it comes to better expected performance of the CMS experiment (the same can be said for the definitions of object collections) and is discussed in more details in Ref. [51].

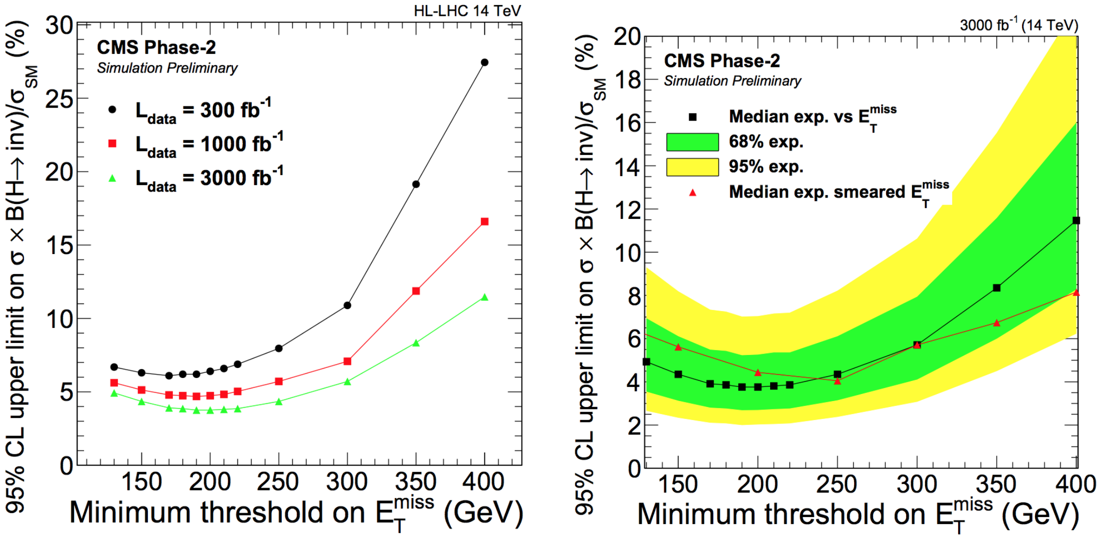


Figure 10.2: Estimation of the 95% CL upper limits presented for different values of the $E_{T,miss}$ threshold for three different scenarios based on the values of total integrate luminosities (left), and the detailed look at the behaviour of the upper limit bands for the most sensitive scenario of $L = 3000\ fb^{-1}$ (right) [51].

Their projection gives an 95% CL upper limit value of 8%, which can be further combined with the previously presented study. Making one more assumption, that both experiments are expected to perform similarly in all channels, the final combination provides a total projected constrain on the B(H→inv) $\leq$ 2.5%[6] for the HL-LHC phase of operation.

10.4 Final words

A search for the invisible decays of Higgs bosons was presented. Specific triggers have been shown following the entire study process from the trigger design until the final result. Discussion of the results has been presented in three eras: past, present and future. The overview of the past era introduced the method and corresponding results in a chronological manner, showing the motives and ideas behind these studies and their realisation. The present was used as an example of how these studies can mature, take advantage of the technical advancements of the detection process and showed a first look at the full Run 2 result. A peak behind the curtain showed that a more analysis focused approach taken with the Run 2 VBF trigger is going to be of even greater importance for the next phases, as it will be crucial to have trigger strategies that will efficiently target interesting topologies.

The upcoming period leaves a lot of opportunities for young researchers to start

---

[6] When assigning the same VBF result to the ATLAS measurement and vice versa.

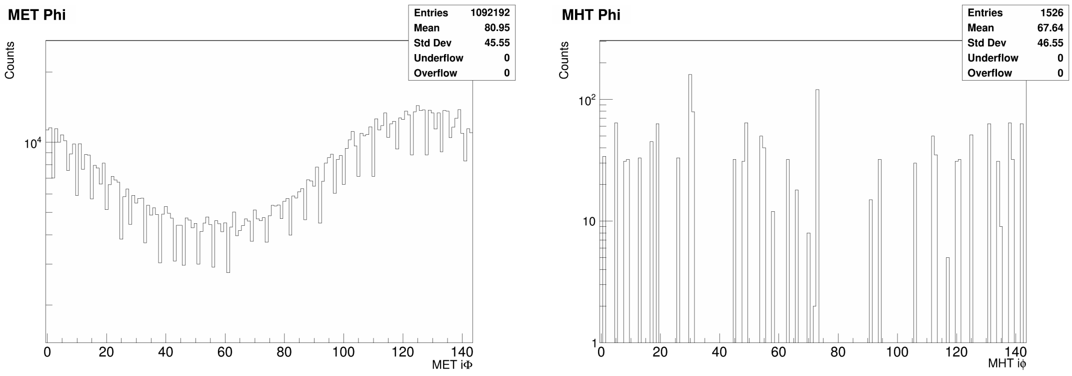
their journey. A prospect of being able to build the entire analysis process from the first, trigger level until and follow it trough until the end result is, from the author's point of view, the best possible reward one can get from PhD studies. The amount of experience gained and the ability to appreciate how every small piece forms the bigger picture is a gift very few endeavours provide as a result.

The final result of this thesis summarises that there was no observed deviation from the SM with respect to the process of interest. An 95% CL upper limit has been set on the branching ratio of the VBF H→ inv decay and it currently stands at 0.15 (0.13) observed (expected) value for the 2017-18 data taking period, while the combination effort with the study focusing on the 2016 era yields a Br(VBF H→inv) = 0.15 (0.12). These results present a preliminary status, which is expected to be improved on when the final result is published in the near future.

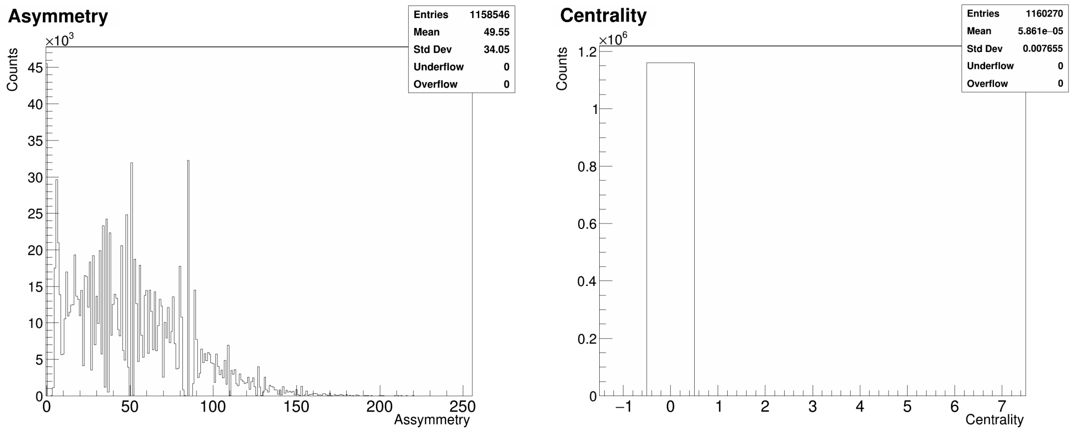
◆

Appendix A

A.1 Supplementary DQM example plots



(a) $\phi_{E_{T,miss}}$ (b) $\phi_{H_{T,miss}}$



(c) Asymmetry (d) Centrality

Figure A.1: DQM example distributions showing the: (a) ϕ of the $\vec{p}_{T,miss}$, (b) ϕ of the $\vec{H}_{T,miss}$, (c) asymmetry and (d) centrality variables.

A.2 List of triggers used in the analysis

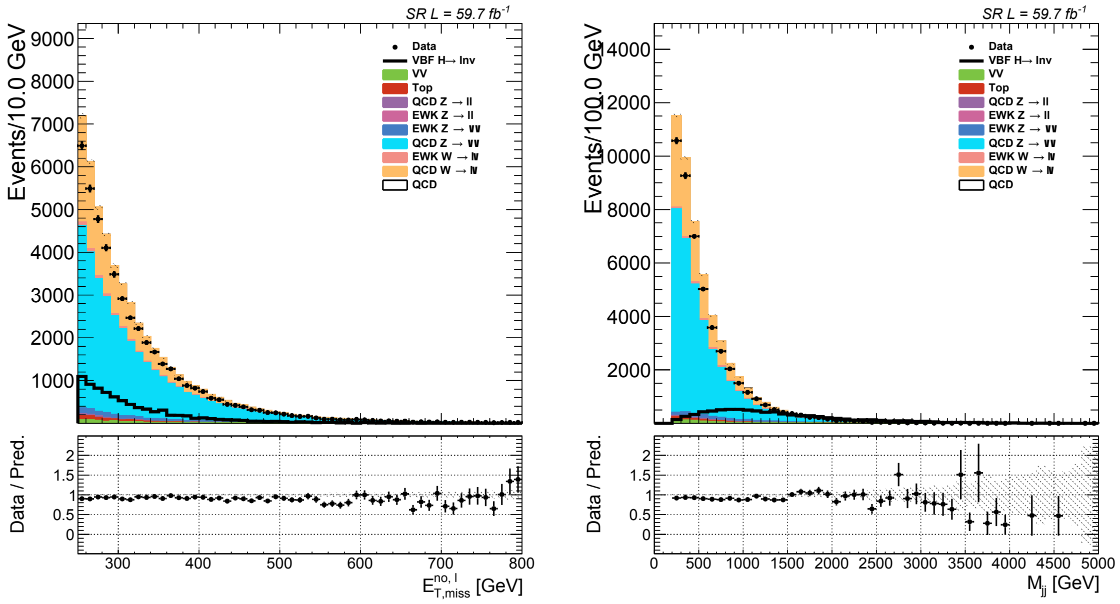
Table A.1 shows the list of HLT algorithms used in the analysis. The corresponding input L1 seeds are listed for each HLT path, with the separation being made for each era of data taking in order to present different thresholds for each year.

Year	HLT path	L1 seed
2017	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight	L1_ETMHF70
	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight_PFHT60	L1_ETMHF80_HTT60er
	HLT_Ele35_WPTight_Gsf	L1_SingleEG24
	HLT_Photon200	L1_SingleEG30
		L1_SingleJet170
		L1_SingleTau100er2p1
	HLT_DiJet110_35_Mjj650_PFMET110	L1_DoubleJet_*_*_DoubleJet*_Mass_Min620
	HLT_TripleJet110_35_35_Mjj650_PFMET110	L1_DoubleJet_*_*_DoubleJet*_Mass_Min620
2018	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight	L1_ETMHF100
		L1_ETM150
	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight_PFHT60	L1_ETMHF90_HTT60er
	HLT_Ele32_WPTight_Gsf	L1_SingleIsoEG24er2p1
		L1_SingleEG26er2p5
		L1_SingleEG60
	HLT_Photon200	L1_SingleEG34er2p5
		L1_SingleJet160er2p5
		L1_SingleJet180
		L1_SingleTau120er2p1
		L1_SingleEG60
	HLT_DiJet110_35_Mjj650_PFMET110	L1_DoubleJet_*_*_DoubleJet*_Mass_Min620
		L1_DoubleJet_*_*_DoubleJet*_Mass_Min620_Jet60TT28
	HLT_TripleJet110_35_35_Mjj650_PFMET110	L1_DoubleJet_*_*_DoubleJet*_Mass_Min620
		L1_DoubleJet_*_*_DoubleJet*_Mass_Min620_Jet60TT28

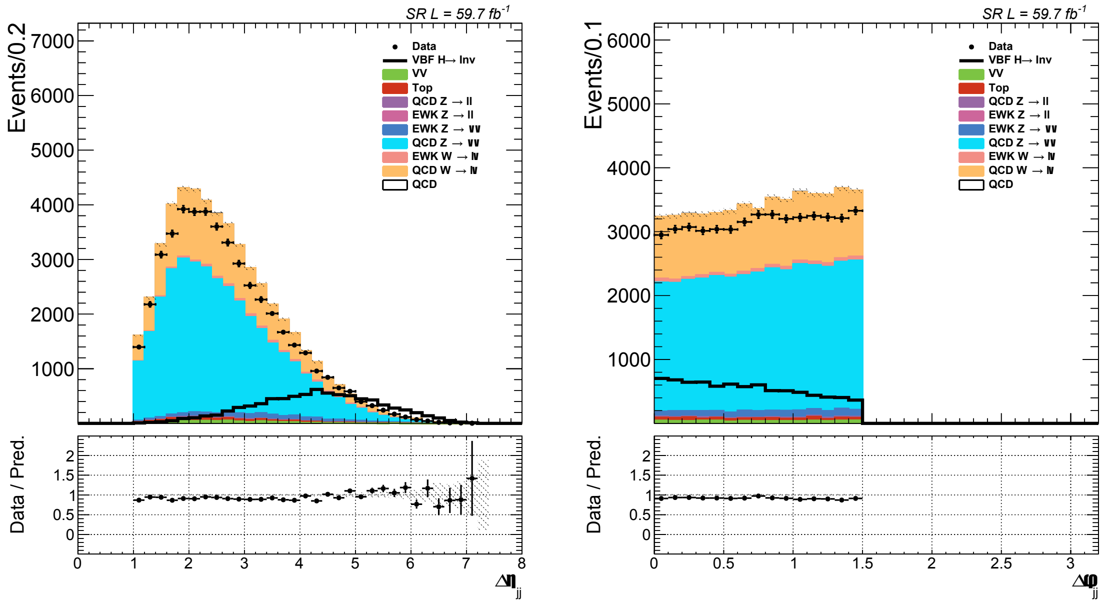
Table A.1: List of HLT paths accompanied by the corresponding L1 seeds used as input [98]. During the 2017 era the L1_DoubleJet seeds imposed thresholds for leading jet p_T threshold ranging from 90 to 115 GeV, while the subleading jet threshold took values from 30 to 40 GeV. Similarly for the 2018 data taking period, the L1_DoubleJet seeds required the leading jet p_T threshold range from 90 to 120 GeV and the subleading jet p_T minimum value ranging from 30 to 45 GeV.

A.3 Analysis strategy for the 2018 era

Figure A.2 shows the distributions of main analysis variables in the SR after the full MTR selection being applied, for the 2018 era. The corresponding distributions for the VTR category are given in Figure A.3.

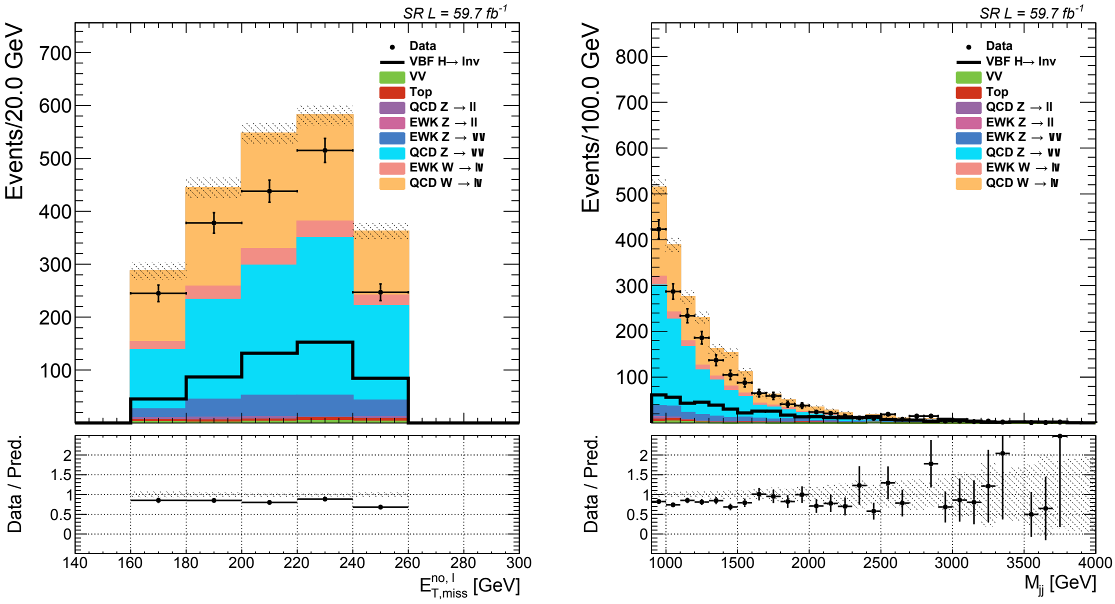


(a) $E_{T,miss}$ (b) m_{jj}

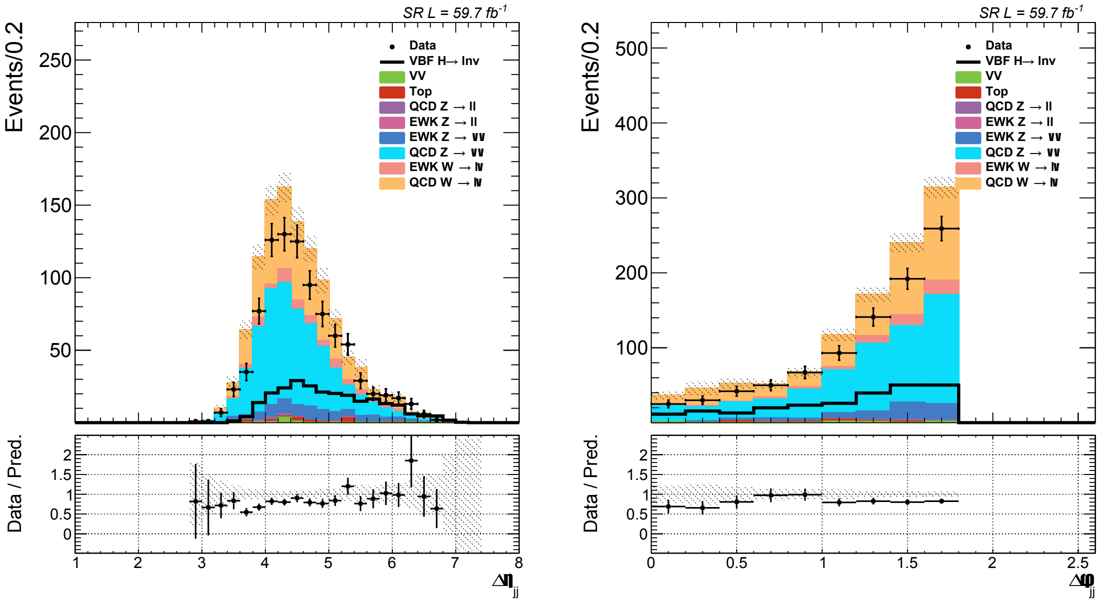


(c) $\Delta\eta_{jj}$ (d) $\Delta\phi_{jj}$

Figure A.2: Distributions of $E_{T,miss}$, m_{jj}, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ variables in the SR after the full MTR selection, for 2018 data.



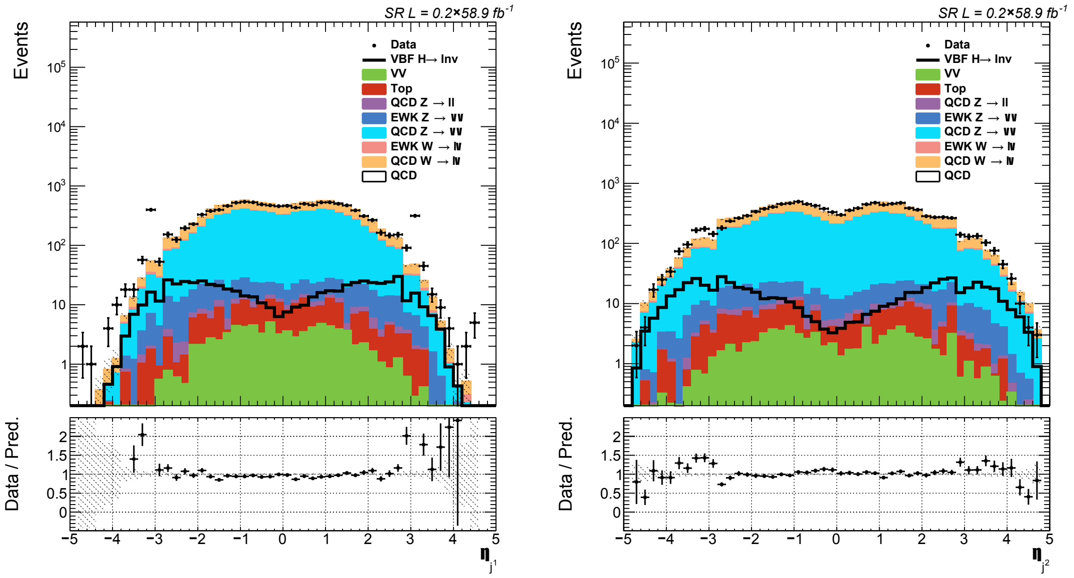
(a) $E_{T,miss}$ (b) m_{jj}



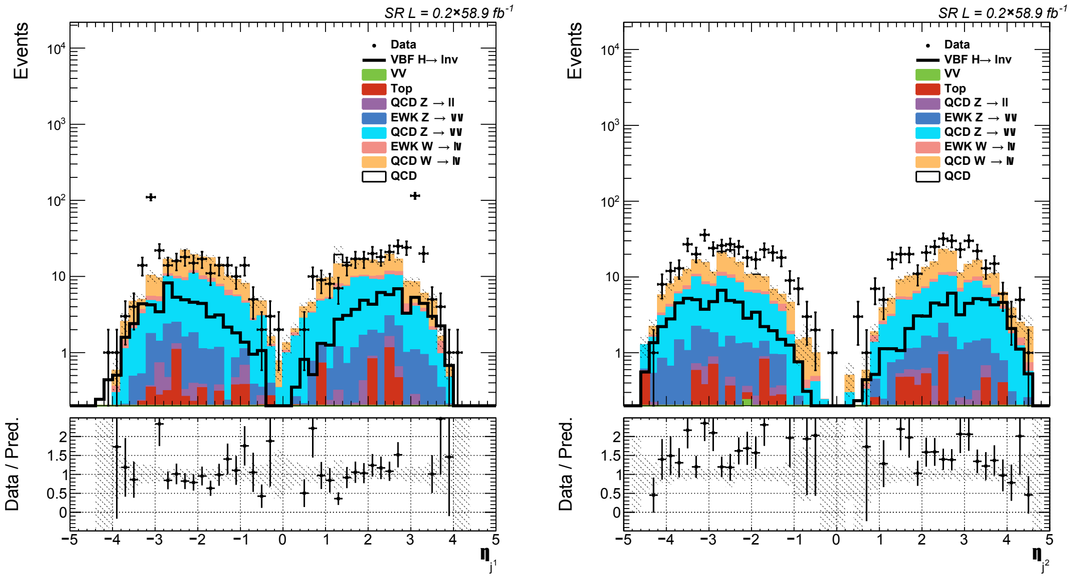
(c) $\Delta\eta_{jj}$ (d) $\Delta\phi_{jj}$

Figure A.3: Distributions of $E_{T,miss}$, m_{jj}, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ variables in the SR after the full VTR selection, for 2018 data.

Figures A.4 and A.5 show distributions of jet η for the leading jet pair, pre and post mitigation veto being applied, respectively (for both categories).

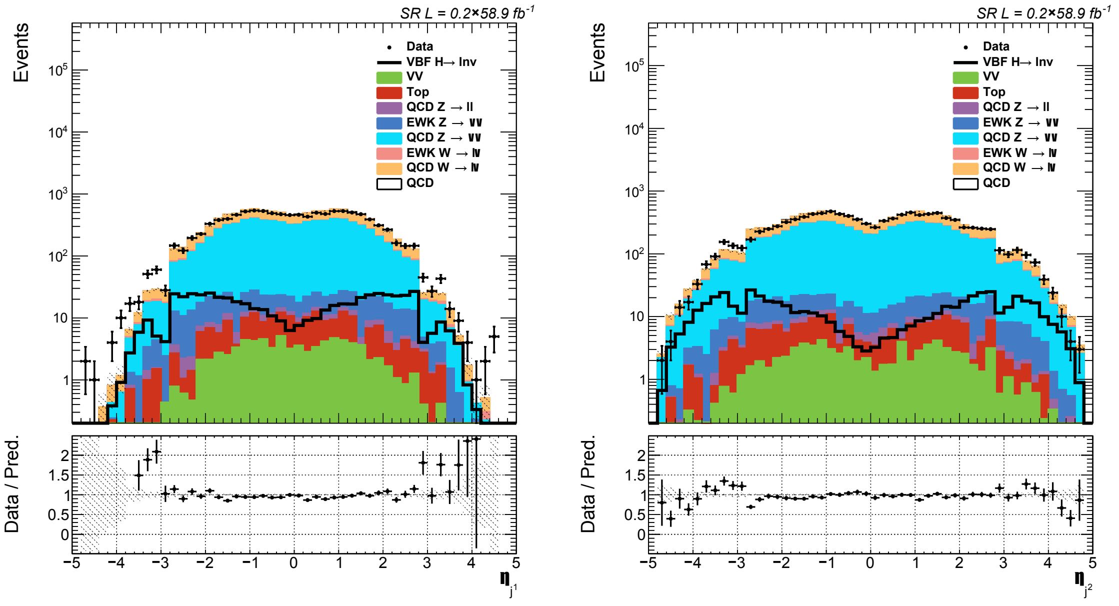


(a) $\eta_{j,1}$ - MTR (b) $\eta_{j,2}$ - MTR

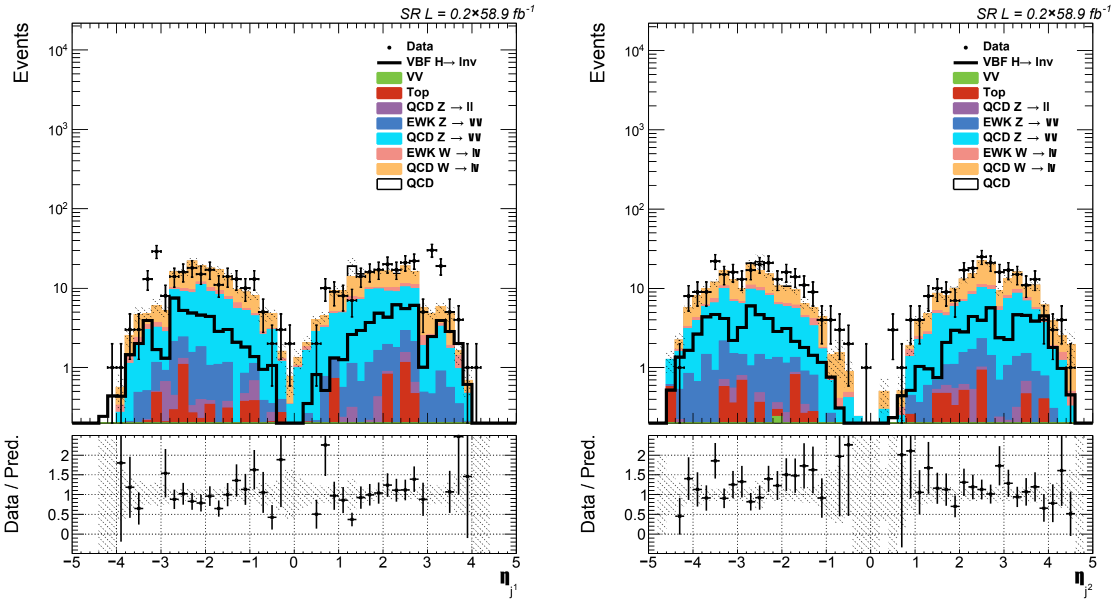


(c) $\eta_{j,1}$ - VTR (d) $\eta_{j,2}$ - VTR

Figure A.4: Distributions of $\eta_{j,1}$ (left) and $\eta_{j,2}$ (right) variables in the signal region after the unbinding of 1/5th of the 2018 data. Both MTR (top) and VTR (bottom) categories are presented.



(a) $\eta_{j,1}$ - MTR (b) $\eta_{j,2}$ - MTR

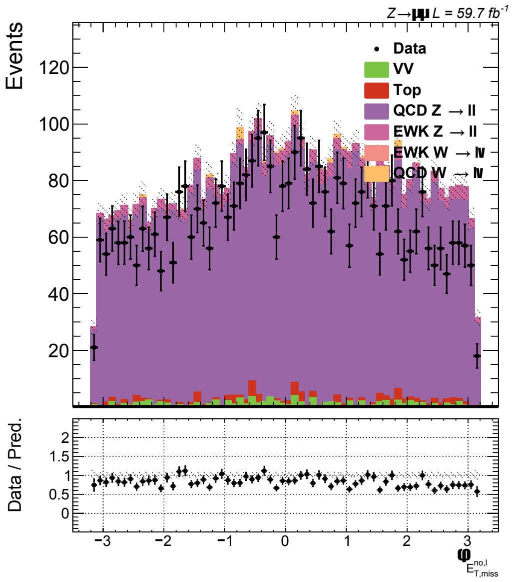


(c) $\eta_{j,1}$ - VTR (d) $\eta_{j,2}$ - VTR

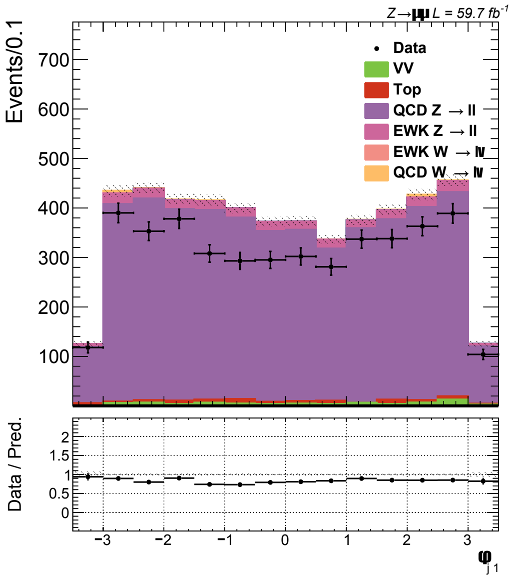
Figure A.5: Distributions of $\eta_{j,1}$ (left) and $\eta_{j,2}$ (right) variables in the signal region after the unbinding of 1/5th of the 2018 data and the mitigation of the jet ”horns” effect. Both MTR (top) and VTR (bottom) categories are presented.

A.4 Dedicated CRs - supplementary material

.

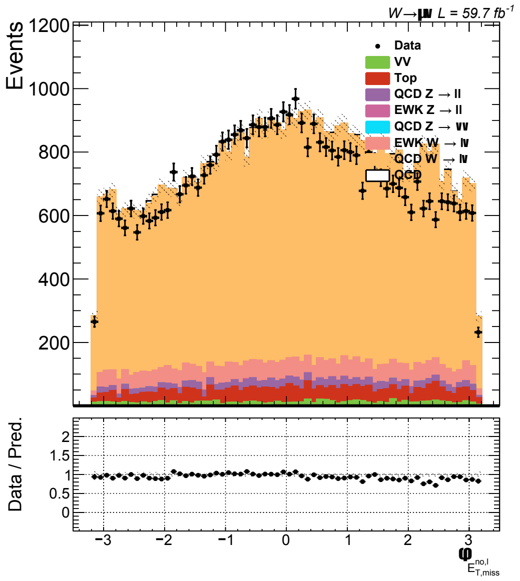


(a) $\phi_{E_{T,miss}^{no,l}}$ - MTR

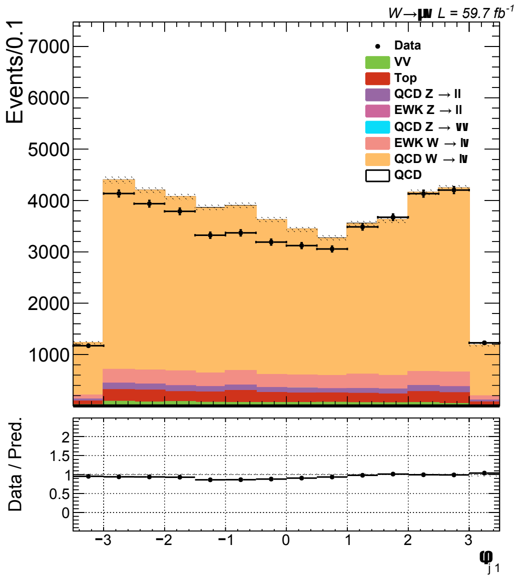


(b) ϕ_{j1} - MTR

Figure A.6: Distributions of $\phi_{E_{T,miss}^{no,l}}$ (a) and ϕ_{j1} (b) variables in the double muon CR for the MTR 2018 category, showing the absence of effects related to the HEM problem.

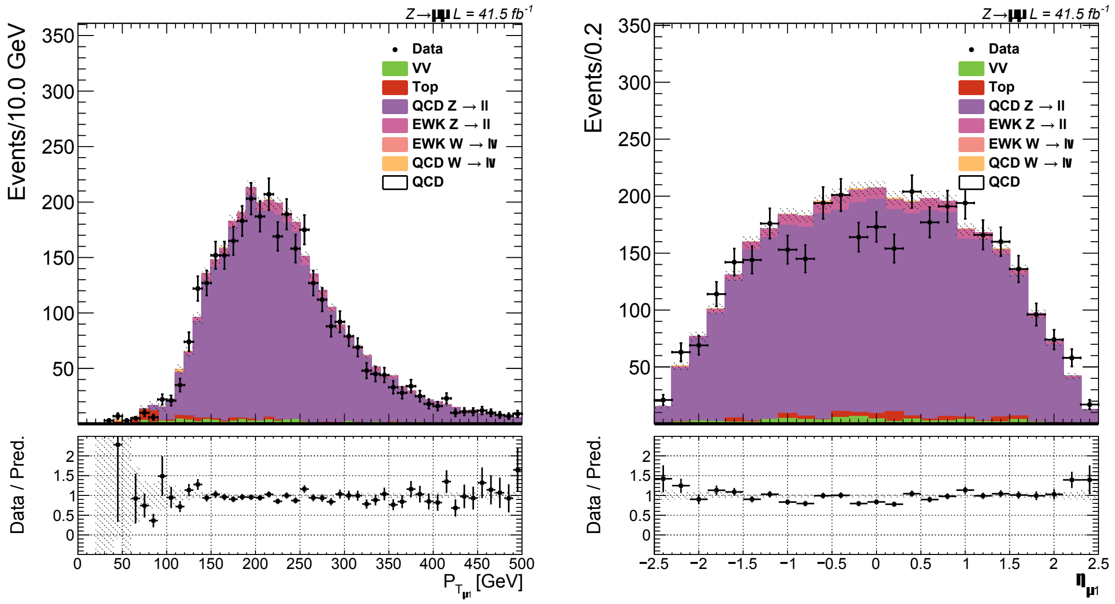


(a) $E_{T,miss}$ - MTR

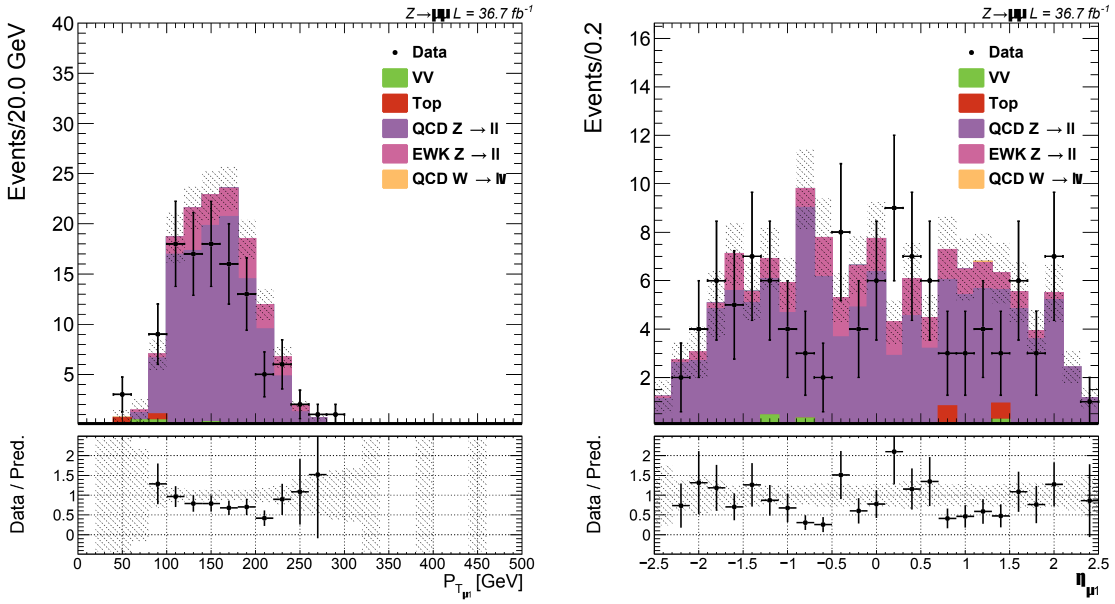


(b) ϕ_{j1} - MTR

Figure A.7: Distributions of $E_{T,miss}$ (a) and ϕ_{j1} (b) variables in the single muon CR for the MTR 2018 category, showing the absence of effects related to the HEM problem.

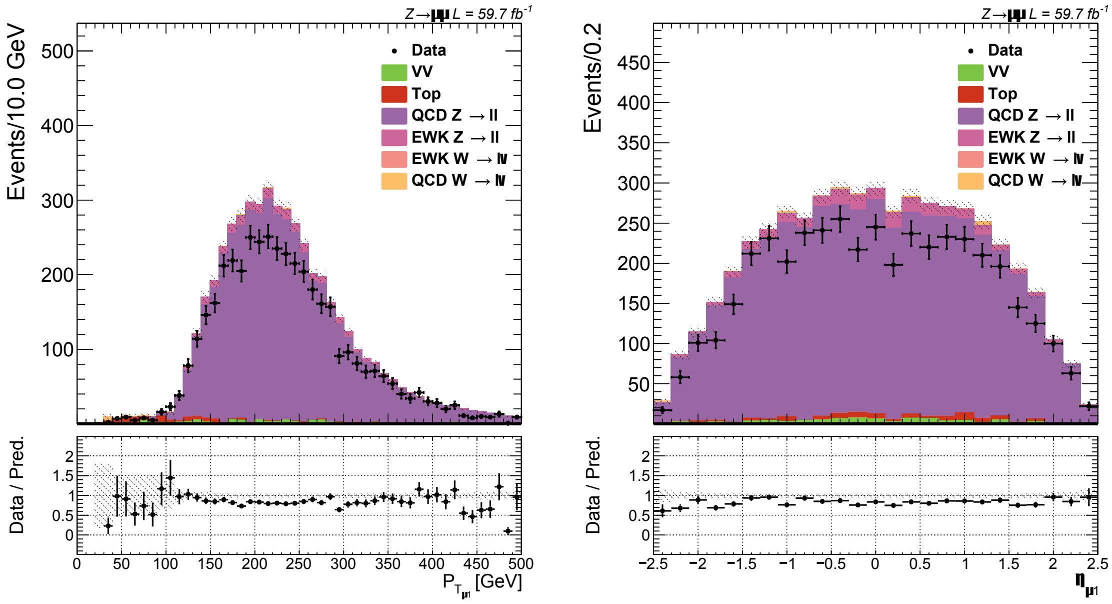


(a) $p_{T_{\mu,1}}$ - MTR (b) $\eta_{\mu,1}$ - MTR

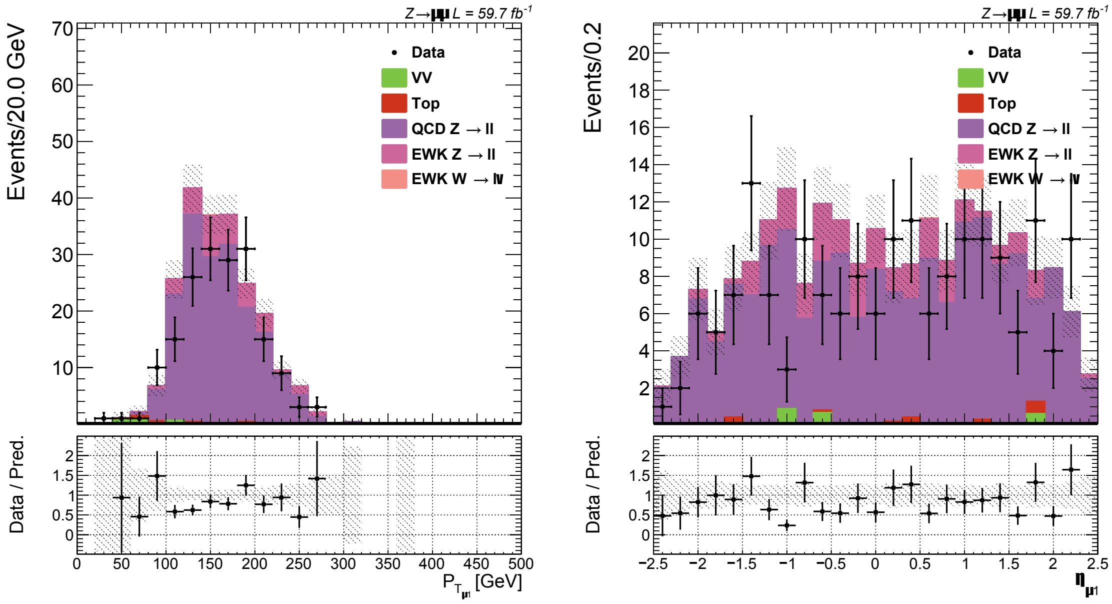


(c) $p_{T_{\mu,1}}$ - VTR (d) $\eta_{\mu,1}$ - VTR

Figure A.8: Distributions of $p_{T_{\mu,1}}$ and $\eta_{\mu,1}$ variables in the double muon region for MTR (top) and VTR (bottom) categories for the 2017 era of data taking.

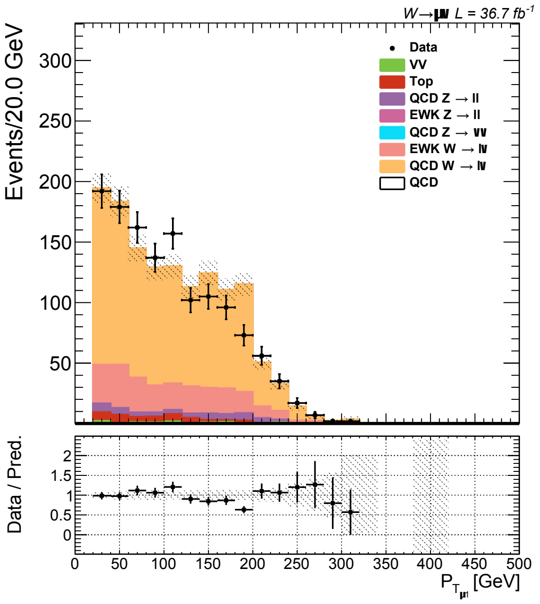


(a) $p_{T_{\mu,1}}$ - MTR (b) $\eta_{\mu,1}$ - MTR

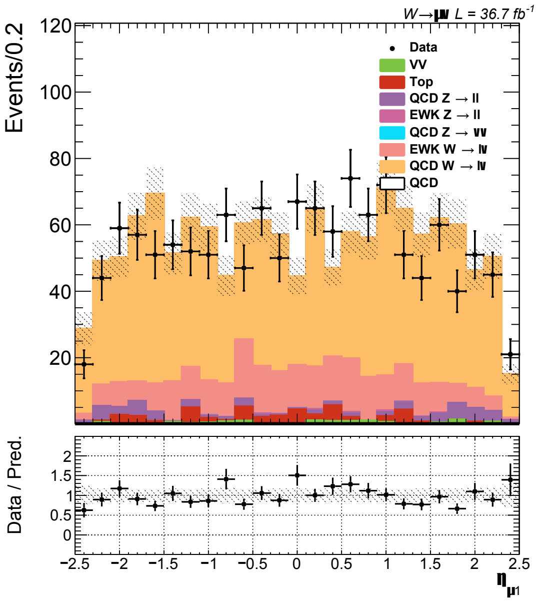


(c) $p_{T_{\mu,1}}$ - VTR (d) $\eta_{\mu,1}$ - VTR

Figure A.9: Distributions of $p_{T_{\mu,1}}$and $\eta_{\mu,1}$ variables in the double muon region for MTR (top) and VTR (bottom) categories for the 2018 era of data taking.

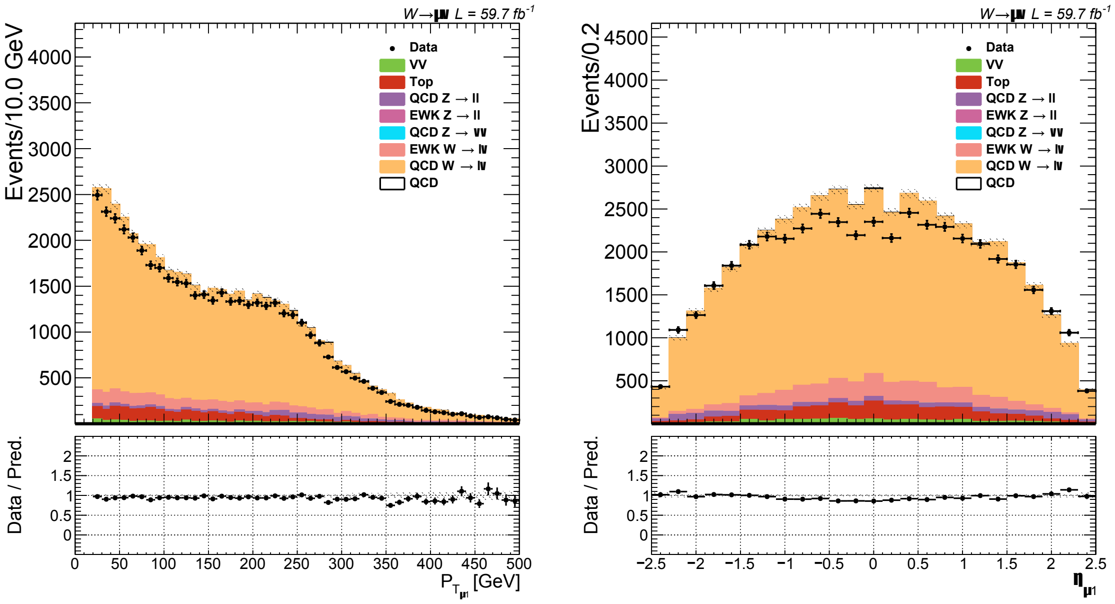


(a) $p_{T_{\mu,1}}$ - VTR

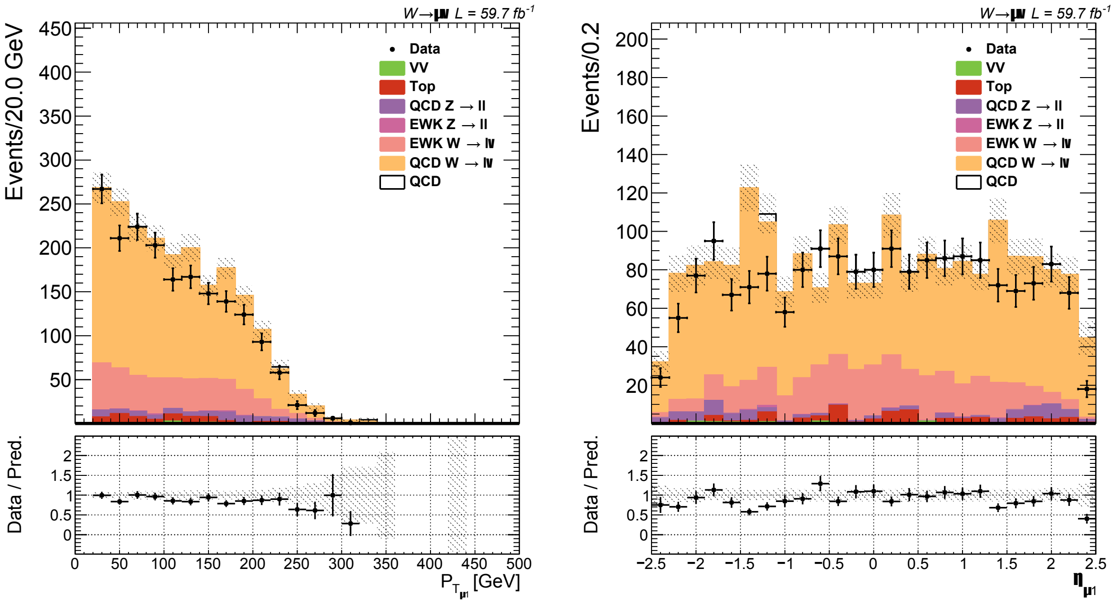


(b) $\eta_{\mu,1}$ - VTR

Figure A.10: Distributions of $p_{T_{\mu,1}}$and $\eta_{\mu,1}$ variables in the single muon region for the VTR category for the 2017 era of data taking.

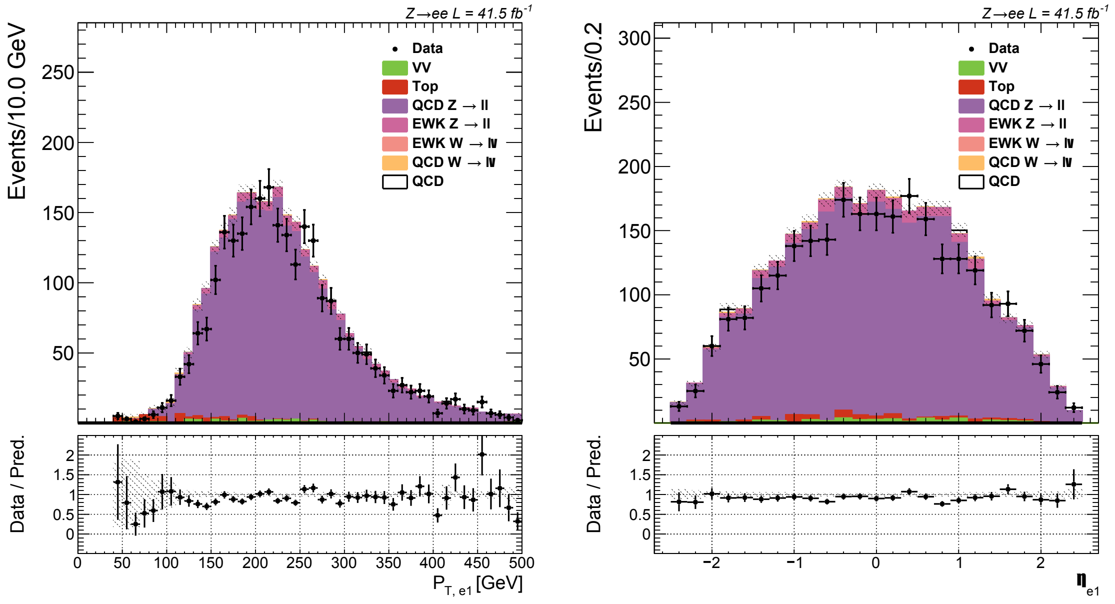


(a) $p_{T_{\mu,1}}$ - MTR (b) $\eta_{\mu,1}$ - MTR

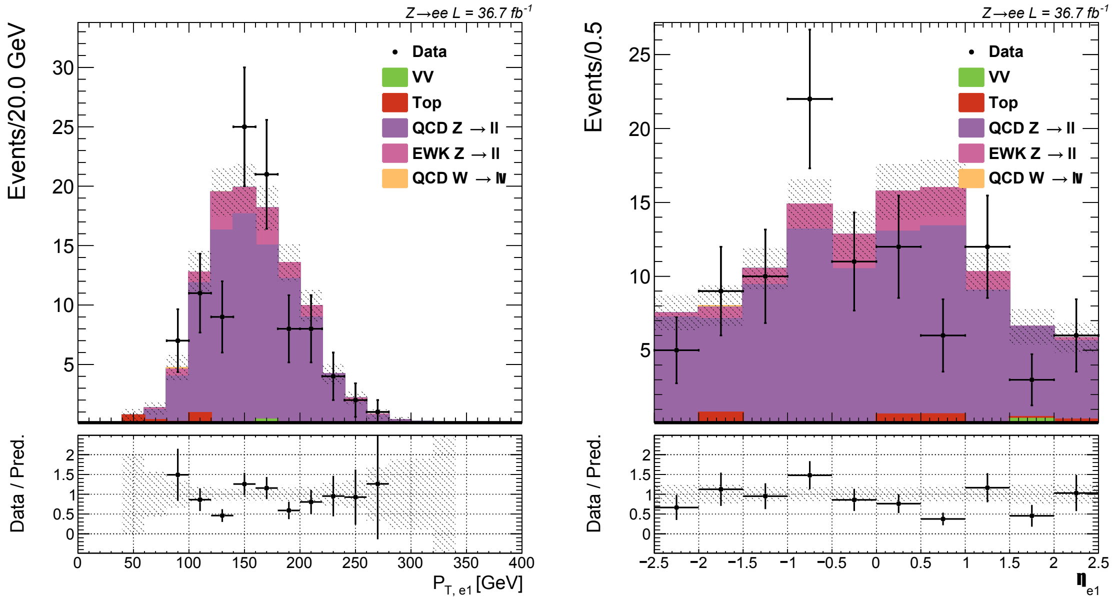


(c) $p_{T_{\mu,1}}$ - VTR (d) $\eta_{\mu,1}$ - VTR

Figure A.11: Distributions of $p_{T_{\mu,1}}$and $\eta_{\mu,1}$ variables in the single muon region for MTR (top) and VTR (bottom) categories for the 2018 era of data taking.

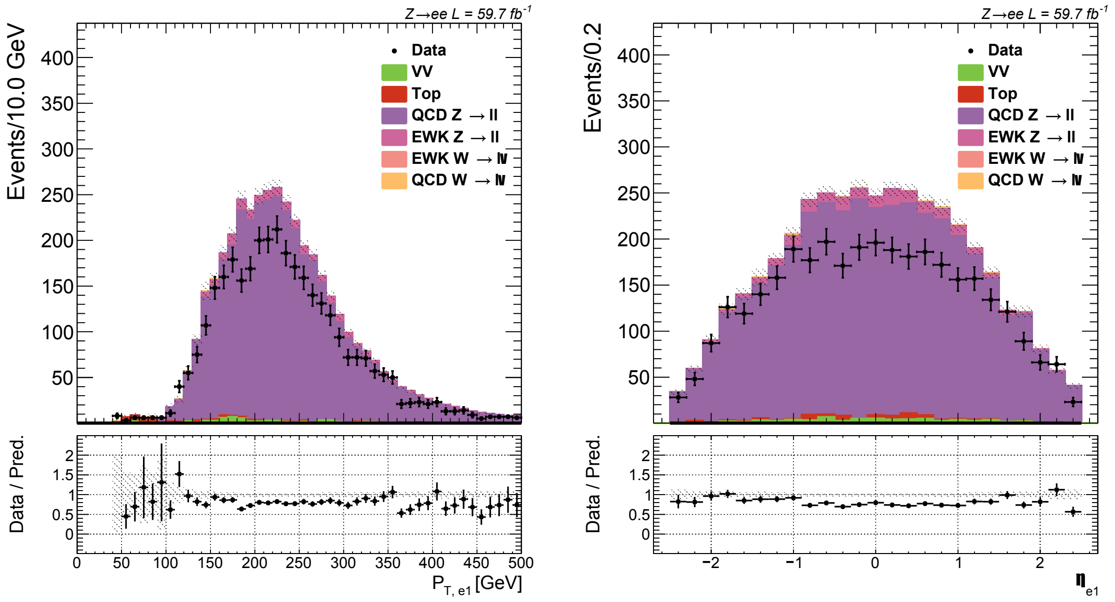


(a) $p_{T_{e,1}}$ - MTR (b) $\eta_{e,1}$ - MTR

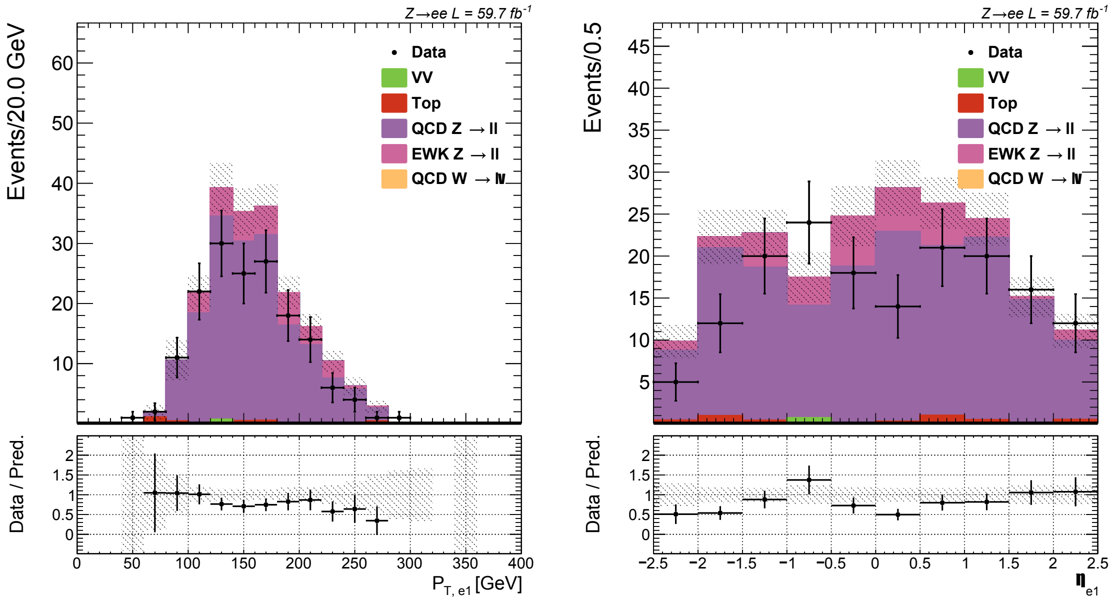


(c) $p_{T_{e,1}}$ - VTR (d) $\eta_{e,1}$ - VTR

Figure A.12: Distributions of $p_{T_{e,1}}$ and $\eta_{e,1}$ variables in the double electron region for MTR (top) and VTR (bottom) categories for the 2017 era of data taking.

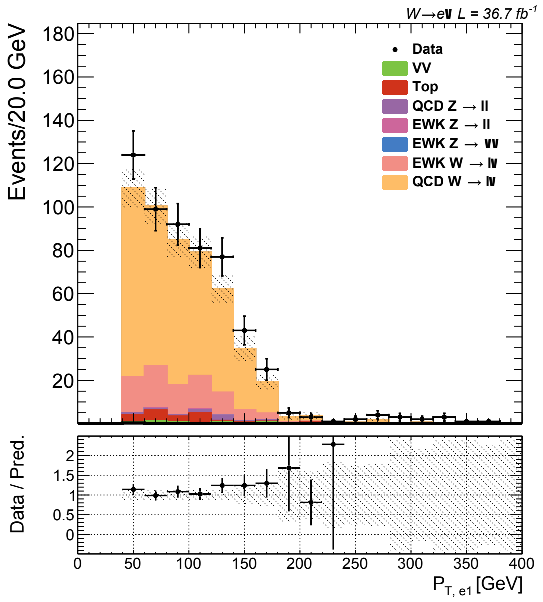


(a) $p_{T_{e,1}}$ - MTR (b) $\eta_{e,1}$ - MTR

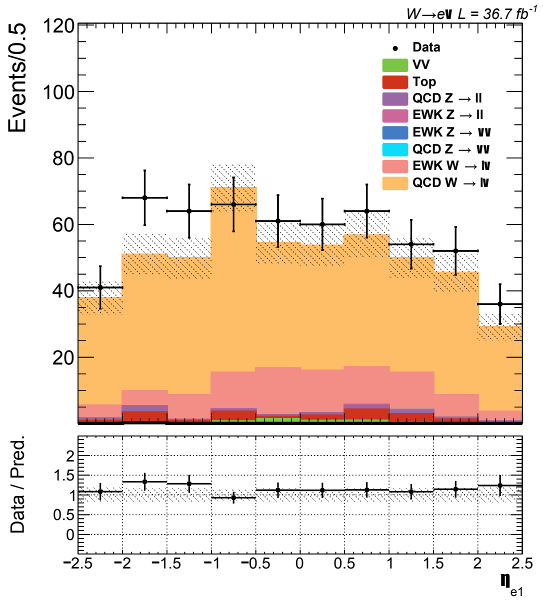


(c) $p_{T_{e,1}}$ - VTR (d) $\eta_{e,1}$ - VTR

Figure A.13: Distributions of $p_{T_{e,1}}$ and $\eta_{e,1}$ variables in the double electron region for MTR (top) and VTR (bottom) categories for the 2018 era of data taking.

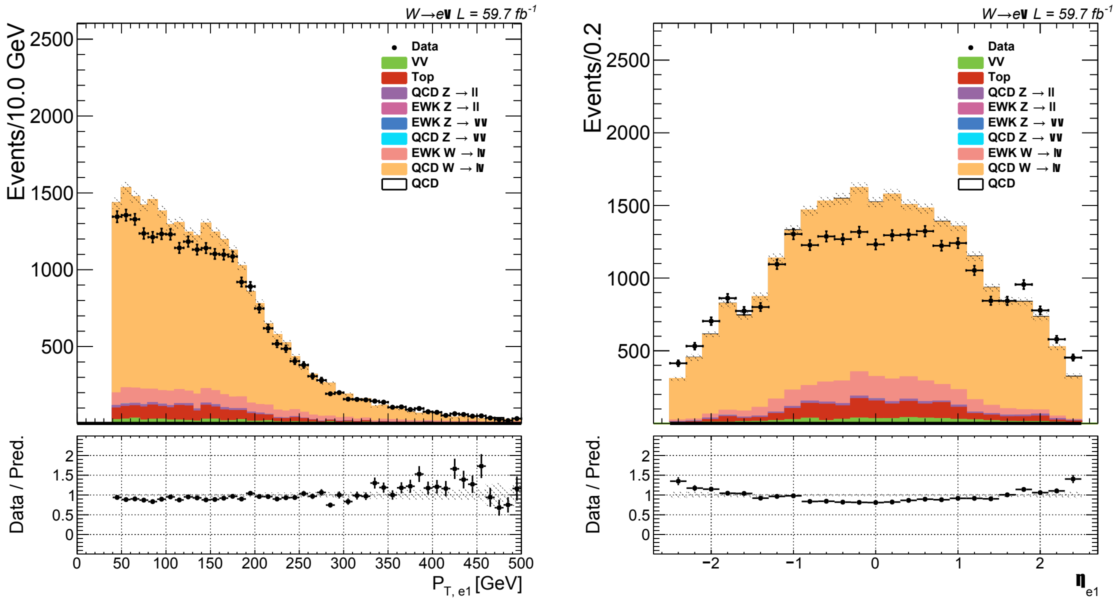


(a) $p_{T_{e,1}}$ - VTR

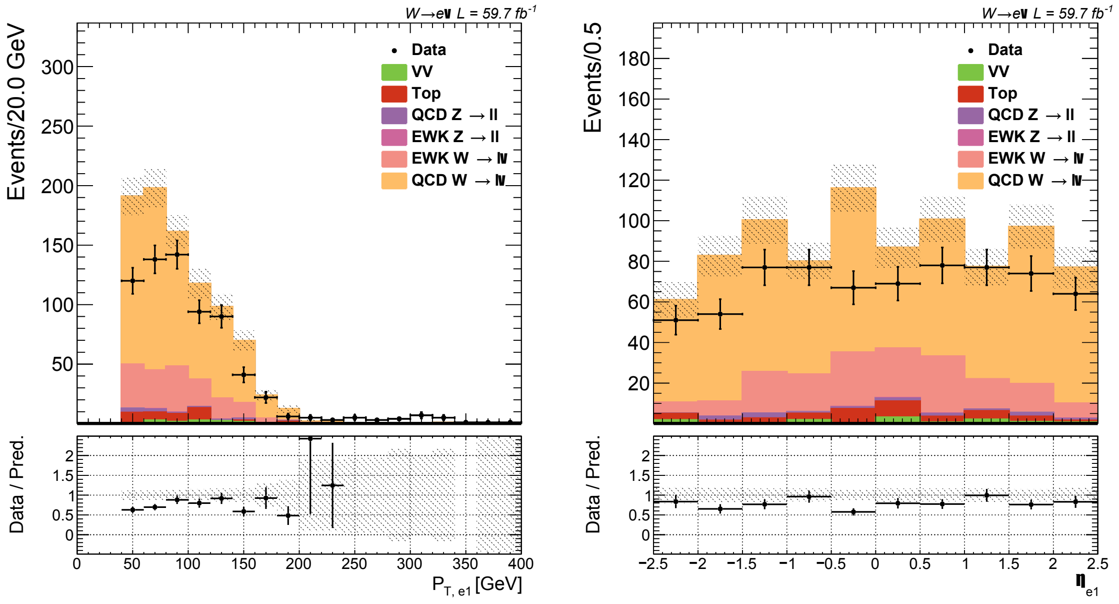


(b) $\eta_{e,1}$ - VTR

Figure A.14: Distributions of $p_{T_{e,1}}$ and $\eta_{e,1}$ variables in the single electron region for the VTR category for the 2017 era of data taking.



(a) $p_{T_{e,1}}$ - MTR (b) $\eta_{e,1}$ - MTR



(c) $p_{T_{e,1}}$ - VTR (d) $\eta_{e,1}$ - VTR

Figure A.15: Distributions of $p_{T_{e,1}}$ and $\eta_{e,1}$ variables in the single electron region for MTR (top) and VTR (bottom) categories for the 2018 era of data taking.

A.5 Fit structure and results - supplementary material

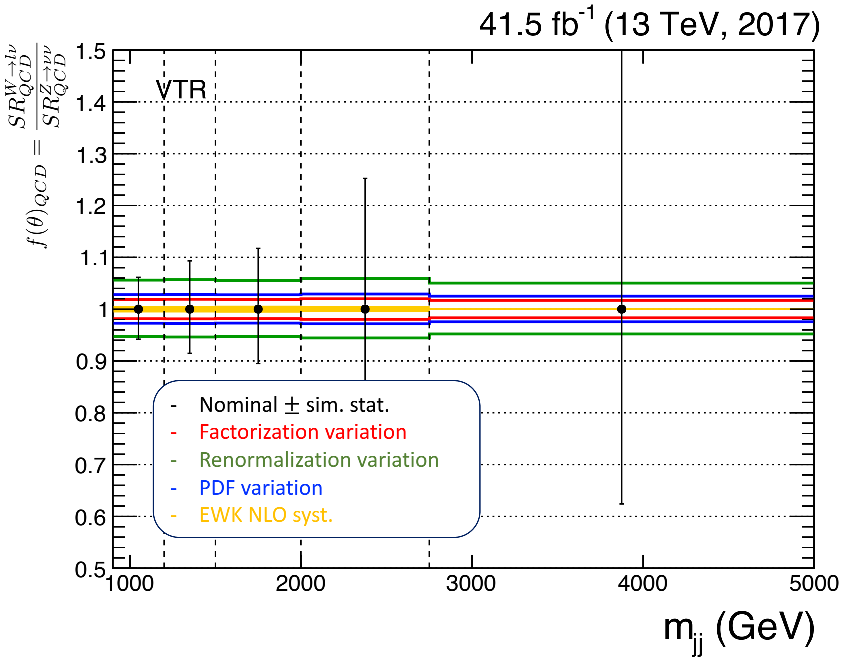
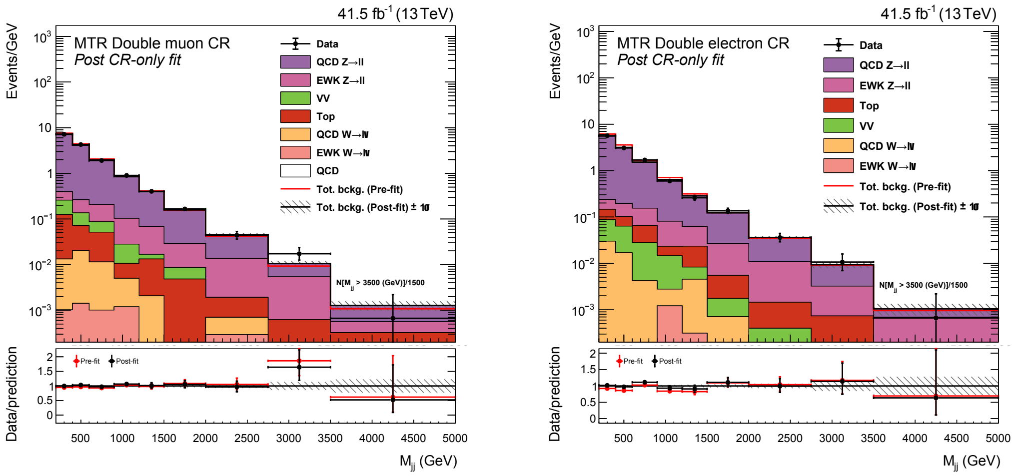
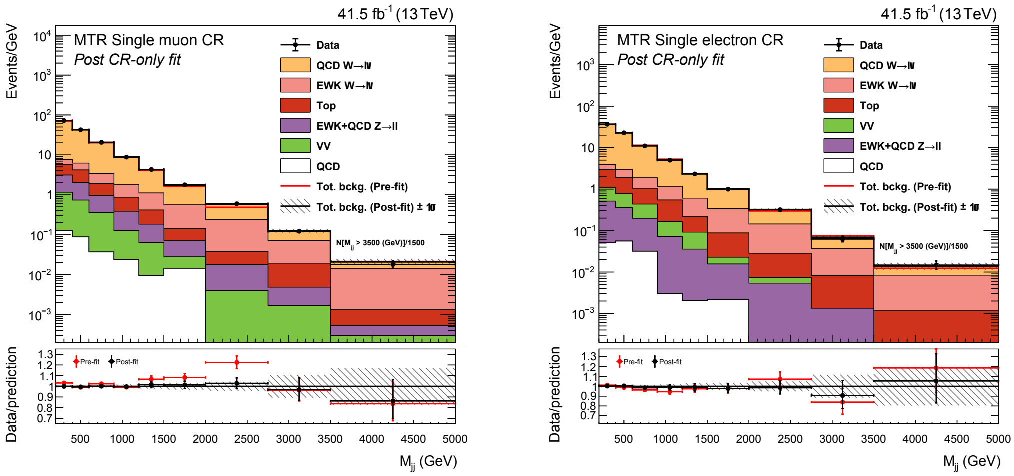


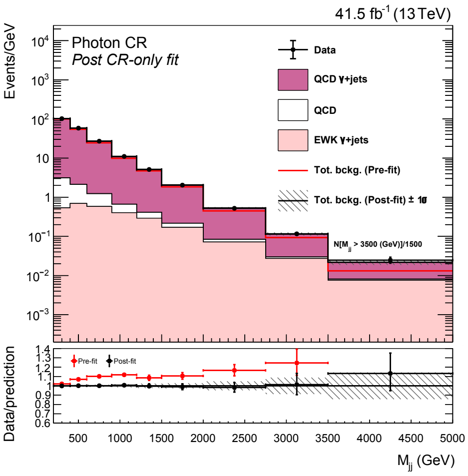
Figure A.16: Theoretical uncertainties on $f(\boldsymbol{\theta})$ ratios for the VTR category and 2017 data, presented as a function of m_{jj} for the QCD production modes.



(a) Dimuon CR (b) Dielectron CR

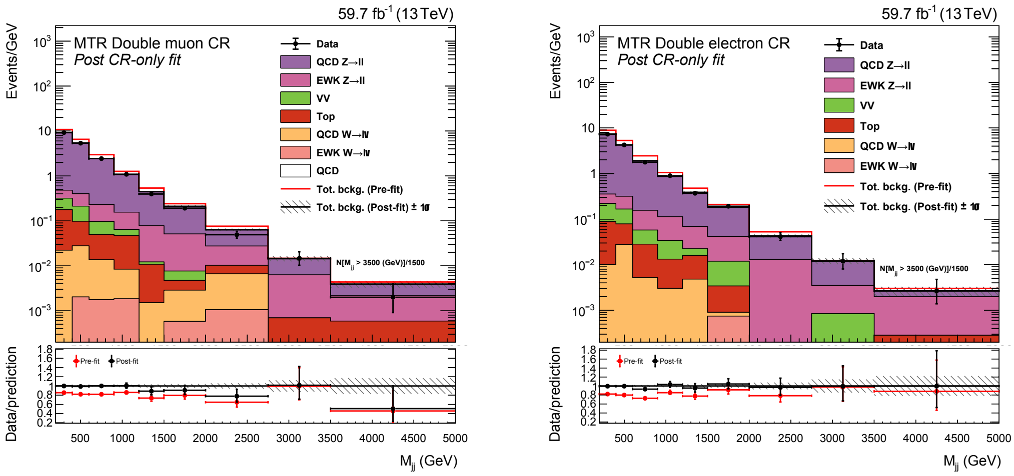


(c) Single muon CR (d) Single electron CR



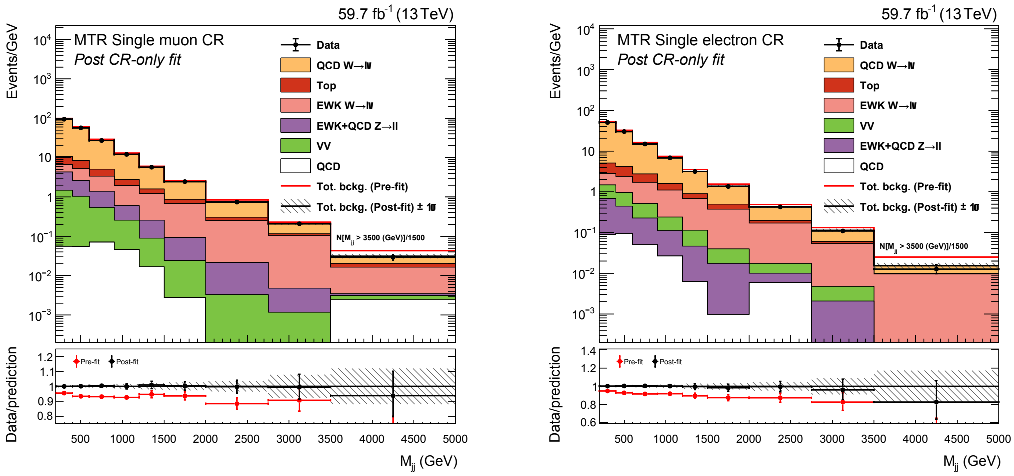
(e) Photon CR

Figure A.17: Post-fit distributions for 2017 data, showing the: (a) dimuon, (b) dielectron, (c) single muon, (d) single electron and (e) photon CR.



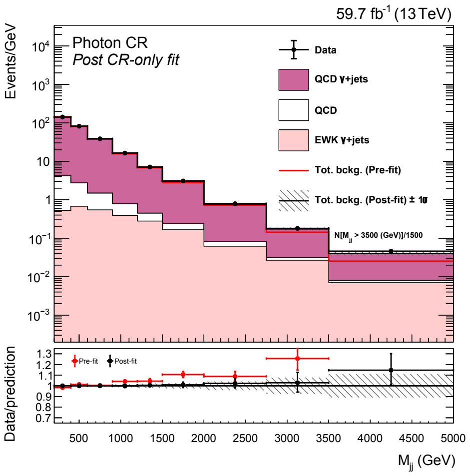
(a) Dimuon CR

(b) Dielectron CR



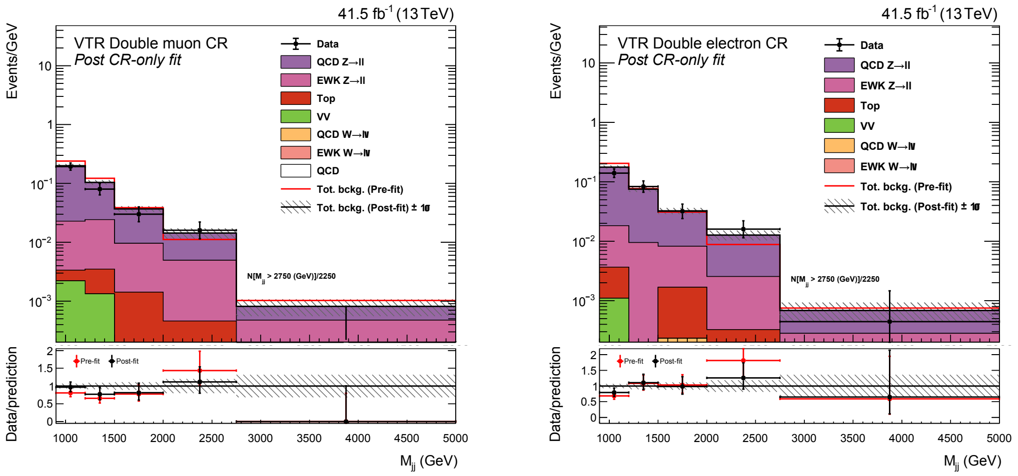
(c) Single muon CR

(d) Single electron CR



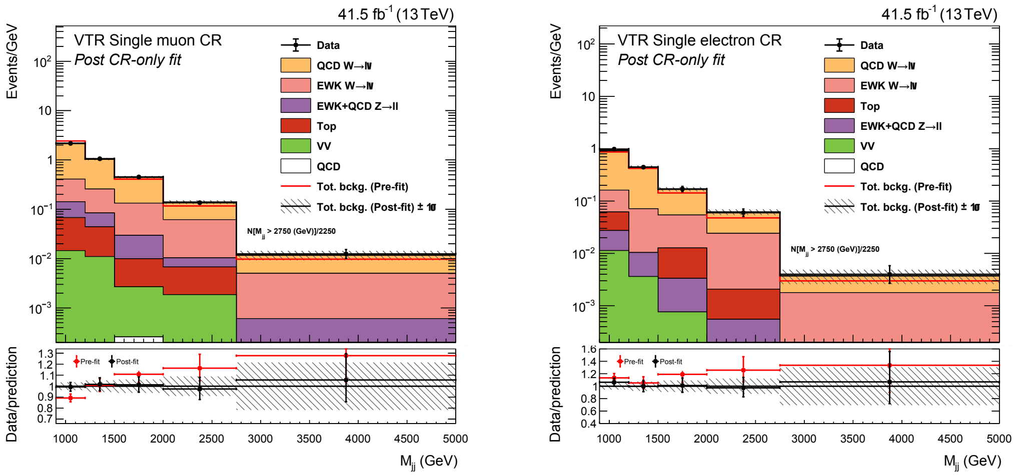
(e) Photon CR

Figure A.18: Post-fit distributions for 2018 data, showing the: (a) dimuon, (b) dielectron, (c) single muon, (d) single electron and (e) photon CR.



(a) Dimuon CR

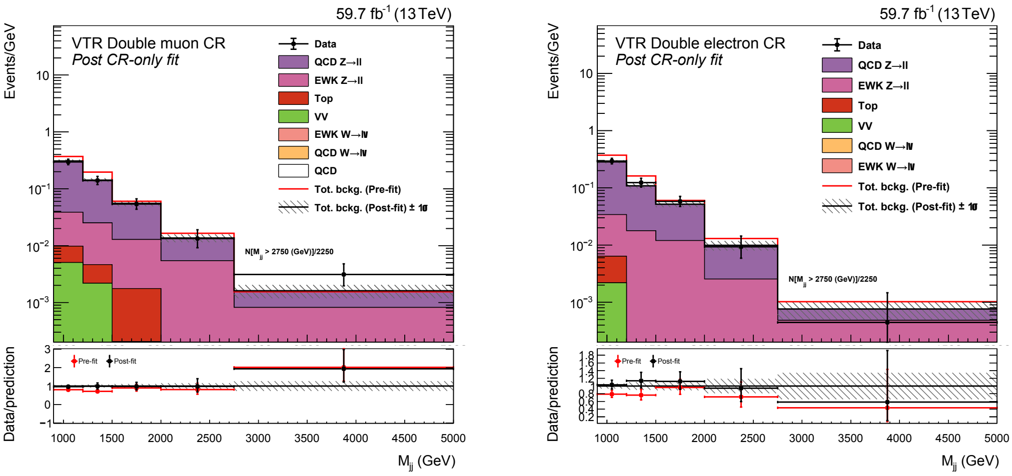
(b) Dielectron CR



(c) Single muon CR

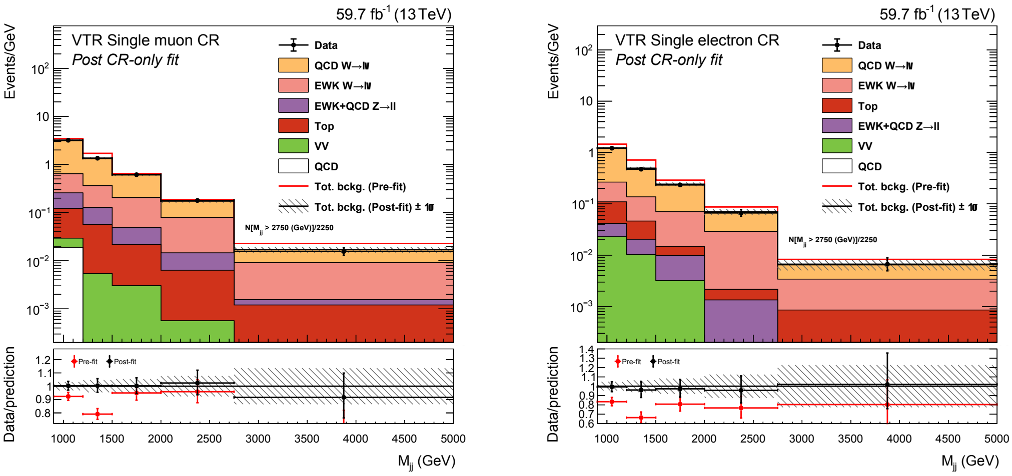
(d) Single electron CR

Figure A.19: Pos-tfit distributions for 2017 data, showing the: (a) dimuon, (b) dielectron, (c) single muon an (d) single electron.



(a) Dimuon CR

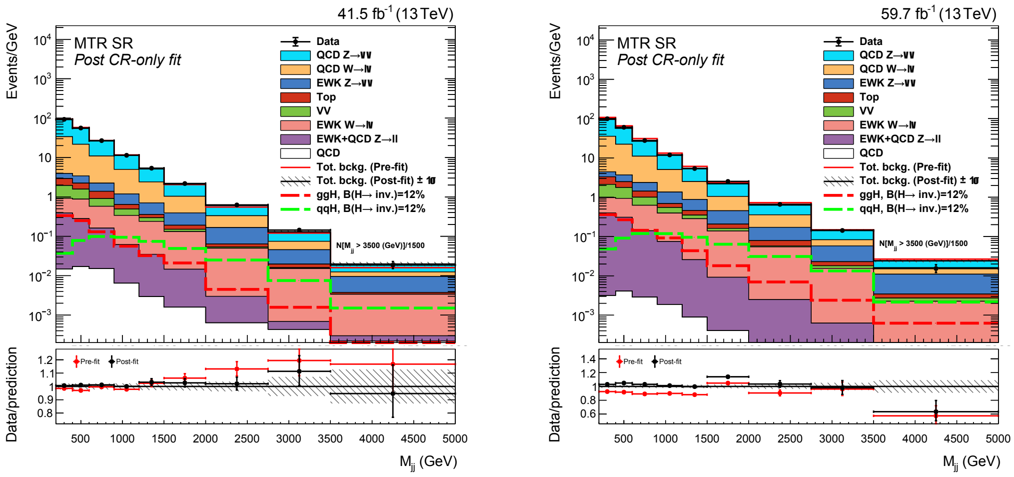
(b) Dielectron CR



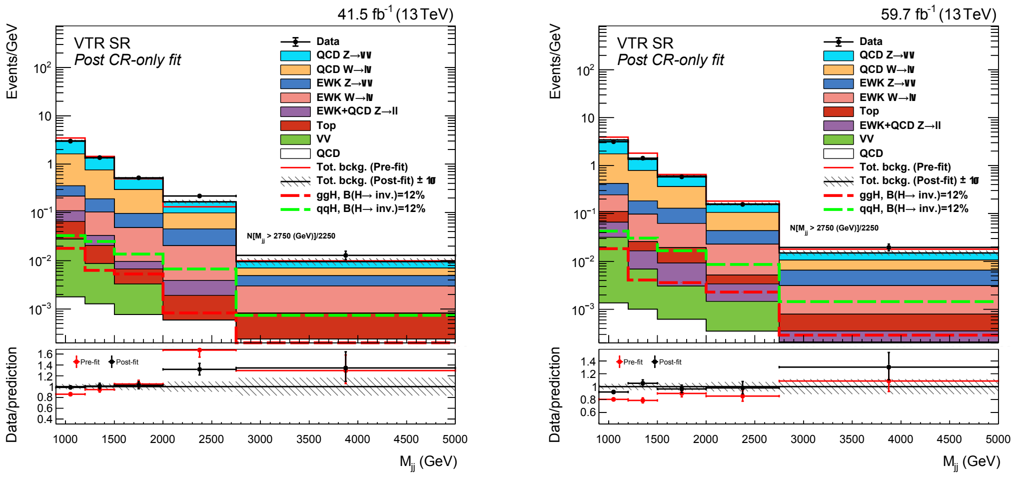
(c) Single muon CR

(d) Single electron CR

Figure A.20: Post-fit distributions for 2018 data, showing the: (a) dimuon, (b) dielectron, (c) single muon an (d) single electron.



(a) MTR 2017 (b) MTR 2018



(c) VTR 2017 (d) VTR 2018

Figure A.21: Post-fit distributions for the SR, showing the: (a) MTR 2017, (b) MTR 2018, (c) VTR 2017 an (d) VTR 2018.

Process	200-400	400-600	600-900	900-1200	1200-1500	1500-2000	2000-2750	2750-3500	>3500
QCD Z($\nu\nu$)+jets	11835.6 ± 163.0	6852.9 ± 99.7	4726.4 ± 79.1	1918.0 ± 38.4	849.5 ± 23.1	556.7 ± 19.3	212.2 ± 9.4	43.3 ± 4.0	10.6 ± 1.6
EWK Z($\nu\nu$)+jets	205.4 ± 5.5	240.1 ± 7.1	257.0 ± 7.1	164.7 ± 5.7	106.1 ± 4.7	103.2 ± 4.8	77.6 ± 4.7	20.5 ± 2.2	8.4 ± 1.4
QCD W(lν)+jets	6027.0 ± 138.5	3696.8 ± 95.4	2645.2 ± 65.6	1148.7 ± 34.9	521.5 ± 20.9	328.7 ± 15.2	126.3 ± 9.2	22.9 ± 3.5	4.3 ± 1.3
EWK W(lν)+jets	107.2 ± 10.5	121.4 ± 11.8	126.9 ± 12.2	84.2 ± 7.8	62.7 ± 6.2	59.8 ± 5.7	35.6 ± 3.7	11.4 ± 1.8	4.6 ± 1.1
$t\bar{t}$ + single-top	194.5 ± 20.1	132.4 ± 12.8	143.3 ± 13.6	44.7 ± 4.9	19.7 ± 4.4	20.1 ± 2.0	8.1 ± 1.2	2.9 ± 0.5	0.6 ± 0.2
VV	204.6 ± 18.4	138.6 ± 12.4	99.1 ± 9.2	46.1 ± 4.4	16.9 ± 1.5	9.4 ± 0.8	2.8 ± 0.3	0.6 ± 0.1	0.0 ± 0.0
QCD Z(ll)+jets	77.0 ± 4.2	53.7 ± 2.8	44.6 ± 2.2	16.7 ± 0.8	9.7 ± 0.4	6.6 ± 0.3	1.8 ± 0.1	0.2 ± 0.0	0.1 ± 0.0
QCD	3.0 ± 1.1	3.4 ± 1.2	4.5 ± 1.7	1.9 ± 0.7	0.9 ± 0.3	0.8 ± 0.3	0.5 ± 0.2	0.3 ± 0.1	0.3 ± 0.1
gg$H(\rightarrow$ inv)	558.2	416.4	325.4	140.1	80.8	87.7	28.1	9.9	0.0
VBF$H(\rightarrow$ inv)	61.9	130.5	247.2	235.8	185.8	204.7	156.3	47.1	18.9
Observed	18628	11221	8055	3418	1607	1092	469	108	28

Table A.2: Post-fit yields of processes in the SR for the MTR category for 2017 data.

Process	200-400	400-600	600-900	900-1200	1200-1500	1500-2000	2000-2750	2750-3500	>3500
QCD Z($\nu\nu$)+jets	12490.8 ± 168.8	7286.1 ± 114.8	4878.6 ± 84.2	2042.2 ± 40.3	869.4 ± 21.9	635.1 ± 20.0	214.7 ± 10.0	46.7 ± 4.1	10.2 ± 1.4
EWK Z($\nu\nu$)+jets	230.3 ± 7.5	251.7 ± 8.0	287.2 ± 8.7	201.6 ± 7.1	126.0 ± 5.9	133.6 ± 6.8	70.9 ± 4.9	25.9 ± 3.3	9.4 ± 1.7
QCD W(lν)+jets	6413.8 ± 153.0	3873.2 ± 98.4	2694.8 ± 72.6	1136.7 ± 37.5	495.6 ± 20.4	323.6 ± 17.1	143.2 ± 11.3	18.8 ± 3.2	6.4 ± 1.6
EWK W(lν)+jets	118.8 ± 12.9	131.0 ± 13.4	150.1 ± 14.9	98.7 ± 10.9	78.9 ± 7.9	69.3 ± 7.3	40.3 ± 4.9	11.3 ± 1.9	3.0 ± 0.8
$t\bar{t}$ + single-top	236.9 ± 20.9	145.2 ± 14.6	129.5 ± 13.1	41.4 ± 3.6	32.7 ± 3.0	23.5 ± 2.5	15.7 ± 1.2	3.6 ± 0.4	1.1 ± 0.1
VV	228.0 ± 18.9	165.7 ± 13.8	112.4 ± 9.3	45.4 ± 3.7	19.8 ± 1.6	11.3 ± 0.9	2.3 ± 0.2	1.9 ± 0.2	0.7 ± 0.1
QCD Z(ll)+jets	81.4 ± 3.8	61.9 ± 2.1	42.0 ± 1.7	22.3 ± 0.9	7.6 ± 0.3	4.5 ± 0.1	1.8 ± 0.1	0.4 ± 0.0	0.0 ± 0.0
QCD	0.7 ± 0.2	0.9 ± 0.3	0.9 ± 0.3	0.6 ± 0.2	0.3 ± 0.1	0.2 ± 0.1	0.1 ± 0.0	0.0 ± 0.0	0.1 ± 0.0
gg$H(\rightarrow$ inv)	602.8	445.5	359.6	230.2	108.9	75.4	43.6	15.0	7.8
VBF$H(\rightarrow$ inv)	80.4	151.7	297.6	296.6	239.3	265.3	194.2	83.4	26.8
Observed	19848	12025	8320	3578	1612	1262	494	107	23

Table A.3: Post-fit yields of processes in the SR for the MTR category for 2018 data.

Process	900-1200	1200-1500	1500-2000	2000-2750	>2750
QCD Z(uu)+jets	410.4 ± 19.6	177.0 ± 10.8	105.9 ± 8.2	63.2 ± 6.6	7.0 ± 1.9
EWK Z($\nu\nu$)+jets	40.7 ± 2.8	26.6 ± 2.5	23.5 ± 2.5	22.9 ± 2.9	5.3 ± 1.4
QCD W(lν)+jets	383.9 ± 19.1	173.2 ± 11.4	103.8 ± 9.1	47.6 ± 6.9	5.6 ± 1.8
EWK W(lν)+jets	32.1 ± 4.5	20.6 ± 3.5	19.5 ± 3.4	14.3 ± 2.7	6.0 ± 2.0
$t\bar{t}$ + single-top	11.3 ± 1.4	3.6 ± 0.4	1.7 ± 0.2	1.0 ± 0.1	1.3 ± 0.1
VV	7.8 ± 0.8	2.2 ± 0.3	1.3 ± 0.1	0.0 ± 0.0	0.0 ± 0.0
QCD Z(ll)+jets	13.1 ± 0.9	3.8 ± 0.3	1.5 ± 0.1	1.5 ± 0.1	0.0 ± 0.0
QCD	0.5 ± 0.1	0.4 ± 0.0	0.4 ± 0.0	0.4 ± 0.0	0.5 ± 0.1
gg$H(\rightarrow$ inv)	45.1	15.8	22.2	5.2	0.0
VBF$H(\rightarrow$ inv)	82.7	62.5	57.2	42.2	13.9
Observed	896	409	260	164	29

Table A.4: Post-fit yields of processes in the SR for the VTR category for 2017 data.

Process	900-1200	1200-1500	1500-2000	2000-2750	>2750
QCD Z($\nu\nu$)+jets	463.4 ± 20.0	173.5 ± 10.5	114.4 ± 8.4	37.9 ± 4.0	11.7 ± 2.1
EWK Z($\nu\nu$)+jets	49.2 ± 3.7	26.4 ± 2.5	31.8 ± 3.2	15.1 ± 2.1	9.1 ± 2.0
QCD W(lν)+jets	385.9 ± 17.7	190.2 ± 12.1	117.0 ± 10.2	46.2 ± 6.5	11.0 ± 2.9
EWK W(lν)+jets	38.3 ± 6.1	21.6 ± 3.8	20.9 ± 3.7	13.1 ± 3.0	6.0 ± 1.6
$t\bar{t}$ + single-top	12.4 ± 1.2	3.0 ± 0.9	5.2 ± 0.9	1.3 ± 0.2	0.9 ± 0.2
VV	8.7 ± 0.9	1.7 ± 0.2	1.1 ± 0.3	0.8 ± 0.1	0.0 ± 0.0
QCD Z(ll)+jets	9.9 ± 0.7	2.9 ± 0.3	3.0 ± 0.2	1.4 ± 0.1	0.3 ± 0.0
QCD	0.4 ± 0.0	0.3 ± 0.0	0.3 ± 0.0	0.3 ± 0.0	0.5 ± 0.0
gg$H(\rightarrow$ inv)	46.0	10.2	15.0	14.3	5.4
VBF$H(\rightarrow$ inv)	107.1	76.5	69.7	54.0	27.1
Observed	944	428	291	116	44

Table A.5: Post-fit yields of processes in the SR for the VTR category for 2018 data.

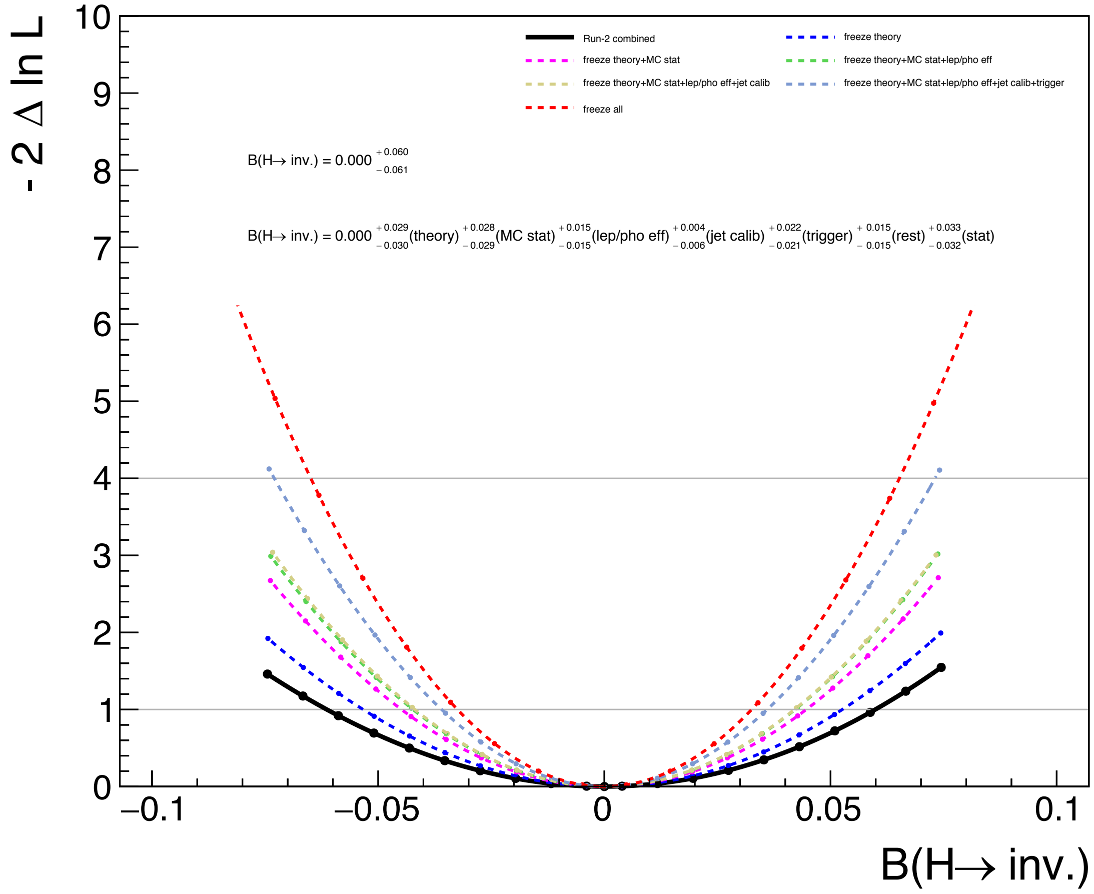


Figure A.22: Likelihood scan for the Run 2 combination for the estimated best fit Br(H→inv) = 0, with scans obtained by sequentially freezing the groups of nuisance parameters.

Bibliography

[1] Mark Thomson, *Modern Particle Physics.* Cambridge University Press, 2013. DOI: `10.1017/CBO9781139525367`.

[2] Particle Data Group, “Review of Particle Physics,” *Phys. Rev. D*, vol. 98, 2018. DOI: `10.1103/PhysRevD.98.030001`.

[3] C. N. Yang and R. L. Mills, “Conservation of isotopic spin and isotopic gauge invariance,” *Phys. Rev.*, vol. 96, 1954. DOI: `10.1103/PhysRev.96.191`.

[4] M. Schwartz, *Quantum Field Theory and the Standard Model.* Cambridge University Press, 2014. DOI: `10.1080/00107514.2014.970232`.

[5] H. Georgi, *Relativistic Quantum Mechanics.* McGraw-Hill Book Company, 1964.

[6] E. Leader and E. Predazzi, *An introduction to gauge theories and modern particle physics: Volumes I and II.* Cambridge University Press, 1996. DOI: `10.1017/CBO9780511622595`.

[7] C. Fox, *An Introduction to the Calculus of Variations.* Dover Publications Inc., 2010.

[8] E. Noether, “Invariante Variationsprobleme (in German),” *Nachrichten von der Gesellschaft der Wissenschaften zu Gottingen, Mathematisch-Physikalische Klasse*, 1918. [Online]. Available: `http://eudml.org/doc/59024`.

[9] J.D. Bjorken and S.D. Drell, *Lie algrbras is particle physics (second edition).* Westview Press, 1999.

[10] J.D. Bjorken and S.D. Drell, *Relativistic Quantum Fields.* McGraw-Hill Book Company, 1965.

[11] Nobel Media AB, “The Nobel Prize in Physics 1965,” [Online]. Available: `https://www.nobelprize.org/prizes/physics/1965/summary/`.

[12] M. Gell-Mann, “The Eightfold Way: A Theory of strong interaction symmetry,” Mar. 1961. DOI: `10.2172/4008239`.

[13] G. Dissertori et al., *Quantum Chromodynamics: high energy experiments and theory.* Oxford University Press, 2002. DOI: `10.1093/acprof:oso/9780199566419.001.0001`.

[14] S. L. Glashow, "Partial Symmetries of Weak Interactions," *Nucl. Phys.*, vol. 22, 1961. DOI: `10.1016/0029-5582(61)90469-2`.

[15] A. Salam, "Weak and Electromagnetic Interactions," *Conf. Proc. C*, vol. 680519, 1968. DOI: `10.1142/9789812795915_0034`.

[16] S. Weinberg, "A model of leptons," *Phys. Rev. Lett.*, vol. 19, 1967. DOI: `10.1103/PhysRevLett.19.1264`.

[17] F. Englert and R. Brout, "Broken symmetry and the mass of gauge vector mesons," *Phys. Rev. Lett.*, vol. 13, 1964. DOI: `10.1103/PhysRevLett.13.321`.

[18] P. W. Higgs, "Broken symmetries and the masses of gauge bosons," *Phys. Rev. Lett.*, vol. 13, 1964. DOI: `10.1103/PhysRevLett.13.508`.

[19] Nobel Media AB, "The nobel prize in physics 2013," [Online]. Available: `https://www.nobelprize.org/prizes/physics/2013/summary/`.

[20] *Desmos graphing calculator*, 2015. [Online]. Available: `https://www.desmos.com/calculator`.

[21] O. S. Bruning et al, *LHC Design Report.* CERN, 2004. DOI: `10.5170/CERN-2004-003-V-1`.

[22] A. Aivazis, *Feynman: A javascript application for creating Feynman diagrams.* [Online]. Available: `https://feynman.aivazis.com`.

[23] 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, Tech. Rep. CMS-PAS-HIG-18-029, 2019. [Online]. Available: `https://cds.cern.ch/record/2667225`.

[24] ATLAS Collaboration, "Measurements of gluon-gluon fusion and vector-boson fusion Higgs boson production cross-sections in the $H \to WW^* \to e\nu\mu\nu$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector," *Phys. Lett. B*, vol. 789, 2019. DOI: `10.1016/j.physletb.2018.11.064`.

[25] D. Green, "Vector Boson Fusion Higgs Production at the LHC - Mass Variables," 2005. [Online]. Available: `https://cds.cern.ch/record/815121`.

[26] M. Rauch, "Vector-Boson Fusion and Vector-Boson Scattering," 2016. [Online]. Available: `https://arxiv.org/abs/1610.08420`.

[27] V. del Duca et all, "Monte carlo studies of the jet activity in higgs + 2 jet events," *JHEP*, vol. 2006, 2006. DOI: `10.1088/1126-6708/2006/10/016`.

[28] L. Mastrolorenzo, "Higgs production in the VH mode at ATLAS and CMS," *PoS*, vol. LHCP2019, 2019. DOI: `10.22323/1.350.0126`.

[29] CMS Collaboration, “Observation of $\mathrm{t\bar{t}H}$ production,” *Phys. Rev. Lett.*, vol. 120, 2018. DOI: `10.1103/PhysRevLett.120.231801`.

[30] ATLAS Collaboration, “Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector,” *Phys. Lett. B*, vol. 784, 2018. DOI: `10.1016/j.physletb.2018.07.035`.

[31] LHC Higgs Cross Section Working Group, *Summary of results.* [Online]. Available: `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG`.

[32] CMS Collaboration, “Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC,” *Phys. Lett. B*, vol. 716, 2012. DOI: `10.1016/j.physletb.2012.08.021`.

[33] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC,” *Phys. Lett. B*, vol. 716, 2012. DOI: `10.1016/j.physletb.2012.08.020`.

[34] 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*, vol. 08, 2016. DOI: `10.1007/JHEP08(2016)045`.

[35] CMS Collaboration, “Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$,” *Eur. Phys. J. C*, vol. 79, 2019. DOI: `10.1140/epjc/s10052-019-6909-y`.

[36] ATLAS Collaboration, “Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s}$ = 13 TeV collected with the ATLAS experiment,” *Phys. Rev. D*, vol. 101, 2020. DOI: `10.1103/PhysRevD.101.012002`.

[37] S. Heinemeyer et al, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group.* CERN, 2013. DOI: `10.5170/CERN-2013-004`.

[38] CMS Collaboration, “Searches for invisible decays of the Higgs boson in pp collisions at $\sqrt{s}$ = 7, 8, and 13 TeV,” *JHEP*, vol. 2, 2016. DOI: `10.1007/JHEP02(2017)135`.

[39] CMS Collaboration, “Search for invisible decays of a Higgs boson produced through vector boson fusion in proton-proton collisions at $\sqrt{s}$ =13 TeV,” *Physics Letters B*, vol. 793, 2019. DOI: `10.1016/j.physletb.2019.04.025`.

[40] P. Dunne, “Searches for invisibly decaying Higgs bosons with the CMS detector,” Presented in 2016. DOI: `10.25560/41978`.

[41] R. Di Maria, "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," Presented in 2019. [Online]. Available: `http://cds.cern.ch/record/2690018`.

[42] S. Baek et al, "Higgs Portal Vector Dark Matter: Revisited," *JHEP*, vol. 5, 2013. DOI: `10.1007/JHEP05(2013)036`.

[43] A. Goudelis et al, "Dark matter in the Inert Doublet Model after the discovery of a Higgs-like boson at the LHC," *JHEP*, vol. 9, 2013. DOI: `10.1007/JHEP09(2013)106`.

[44] A. Berlin et al, "Pseudoscalar Portal Dark Matter," *Phys. Rev. D*, vol. 92, 2015. DOI: `10.1103/PhysRevD.92.015005`.

[45] C. Englert et al, "Exploring the Higgs portal," [Online]. Available: `https://arxiv.org/abs/1106.3097`.

[46] V. Trimble, "Existence and Nature of Dark Matter in the Universe," *Ann. Rev. Astron. Astrophys.*, vol. 25, 1987. DOI: `10.1146/annurev.aa.25.090187.002233`.

[47] Planck Collaboration, "Planck 2013 results. I. Overview of products and scientific results," *Astron. Astrophys.*, vol. 571, 2014. DOI: `10.1051/0004-6361/201321529`.

[48] K. Freese, "Review of Observational Evidence for Dark Matter in the Universe and in upcoming searches for Dark Stars," *EAS Publ. Ser.*, vol. 36, 2009. DOI: `10.1051/eas/0936016`.

[49] S. W. Allen et al, "Cosmological parameters from observations of galaxy clusters," *Ann. Rev. Astron. Astrophys.*, vol. 49, 2011. DOI: `10.1146/annurev-astro-081710-102514`.

[50] R. Massey et al, "The dark matter of gravitational lensing," *Reports on Progress in Physics*, vol. 73, 2010. DOI: `10.1088/0034-4885/73/8/086901`.

[51] M. Cepeda et al, "Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC," DOI: `10.23731/CYRM-2019-007.221`.

[52] M. Hoferichter et al, "Improved Limits for Higgs-Portal Dark Matter from LHC Searches," *Phys. Rev. Lett.*, vol. 119, 2017. DOI: `10.1103/PhysRevLett.119.181803`.

[53] J. Dowell, "History of CERN vol 1," *Physics Bulletin*, vol. 39, 1988. DOI: `10.1088/0031-9112/39/6/028`.

[54] F. Hasert et al, "Observation of neutrino-like interactions without muon or electron in the gargamelle neutrino experiment," *Phys. Lett. B*, vol. 46, 1973. DOI: `10.1016/0370-2693(73)90499-1`.

[55] F. Hasert et al, “Search for elastic muon-neutrino electron scattering,” *Phys. Lett. B*, vol. 46, 1973. DOI: `10.1016/0370-2693(73)90494-2`.

[56] P.M. Watkins, “Discovery of the W and Z bosons,” *Contemp. Phys.*, vol. 27, 1986. DOI: `10.1080/00107518608211015`.

[57] *LEP design report.* Geneva: CERN. [Online]. Available: `https://cds.cern.ch/record/102083`.

[58] A. Gentil-Beccot, “An image of the first page of Tim Berners-Lee’s proposal for the World Wide Web in March 1989,” [Online]. Available: `https://cds.cern.ch/record/2665088`.

[59] J. D. Jackson et al, “Conceptual Design of the Superconducting Super Collider,” 1986. [Online]. Available: `https://lss.fnal.gov/archive/other/ssc/ssc-sr-2020.pdf`.

[60] G. Gurov, “UNK status and plans,” 1996. [Online]. Available: `https://accelconf.web.cern.ch/p95/ARTICLES/MPG/MPG06.PDF`.

[61] LEP Working Group for Higgs Boson Searches et al, “Search for the Standard Model Higgs boson at LEP,” *Phys. Lett. B*, vol. 565, 2003. DOI: `10.1016/S0370-2693(03)00614-2`.

[62] ALICE Collaboration, “The ALICE experiment at the CERN LHC,” *JINST*, vol. 3, 2008. DOI: `10.1088/1748-0221/3/08/s08002`.

[63] LHCb Collaboration, “The LHCb Detector at the LHC,” *JINST*, vol. 3, 2008. DOI: `10.1088/1748-0221/3/08/S08005`.

[64] S. Holmes et al, “Overview of the Tevatron collider complex: goals, operations and performance,” *JINST*, vol. 6, 2011. DOI: `10.1088/1748-0221/6/08/t08001`.

[65] B. Schmidt, “The high-luminosity upgrade of the LHC: Physics and technology challenges for the accelerator and the experiments,” *Journal of Physics: Conference Series*, vol. 706, 2016. DOI: `10.1088/1742-6596/706/2/022002`.

[66] A. Abada et al, “FCC Physics Opportunities: Future Circular Collider Conceptual Design Report Volume 1,” *Eur. Phys. J. C*, vol. 79, 2019. DOI: `10.1140/epjc/s10052-019-6904-3`.

[67] G. L. Bayatian et. al, *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software.* CERN, 2006. [Online]. Available: `http://cds.cern.ch/record/922757`.

[68] CMS Collaboration, “The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment,” *JINST*, vol. 3, 2008. DOI: `10.1088/1748-0221/3/08/S08004`.

[69] D. Barney, “CMS Detector Slice,” 2016. [Online]. Available: `https://cds.cern.ch/record/2120661`.

[70] University of Zurich, “3D axis with spherical coordinates & CMS coordinate system,” [Online]. Available: `https://wiki.physik.uzh.ch/cms/latex:tikz`.

[71] V. Veszpremi, “Operation and performance of the CMS tracker,” *JINST*, vol. 9, 2014. DOI: `10.1088/1748-0221/9/03/c03005`.

[72] D. Sprenger. et. al, “Validation of Kalman Filter alignment algorithm with cosmic-ray data using a CMS silicon strip tracker endcap,” *JINST*, vol. 5, 2010. DOI: `10.1088/1748-0221/5/06/P06007`.

[73] A. Benaglia, “The CMS ECAL performance with examples,” *JINST*, vol. 9, 2014. DOI: `10.1088/1748-0221/9/02/c02008`.

[74] A. Bartoloni et. al, “The CMS ECAL barrel HV system,” *JINST*, 2013. DOI: `10.1088/1748-0221/8/02/C02039`.

[75] CMS Collaboration, “Performance of CMS Hadron Calorimeter Timing and Synchronization using Test Beam, Cosmic Ray, and LHC Beam Data,” *JINST*, vol. 5, 2010. DOI: `10.1088/1748-0221/5/03/T03013`.

[76] P. Paolucci, “The CMS Muon system,” in *Astroparticle, particle and space physics, detectors and medical physics applications.*, 2006. DOI: `10.1142/9789812773678_0096`.

[77] CMS Collaboration, “Performance of the CMS Level-1 Trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV,” *JINST*, 2020. DOI: `10.3204/PUBDB-2020-02629`.

[78] Antoni Shtipliyski, “Systems and algorithms for low-latency event reconstruction for upgrades of the Level-1 trigger of the CMS experiment at CERN,” Presented 2019. [Online]. Available: `http://cds.cern.ch/record/2708430`.

[79] M. Cacciari et al, “The anti-kt jet clustering algorithm,” *JHEP*, vol. 2008, 2008. DOI: `10.1088/1126-6708/2008/04/063`.

[80] S. Donato, “CMS trigger performance,” *EPJ Web Conf.*, vol. 182, 2018. DOI: `10.1051/epjconf/201818202037`.

[81] F. Beaudette, “The CMS Particle Flow Algorithm,” in *International Conference on Calorimetry for the High Energy Frontier*, 2013. arXiv: `1401.8155`.

[82] V. Daponte and A. Bocci, “CMS - HLT configuration management system,” *Journal of Physics: Conference Series*, vol. 664, 2015. DOI: `10.1088/1742-6596/664/8/082008`.

[83] C. Amendola, “Vector Boson Fusion trigger and search for Higgs boson pair production at the LHC in the bb$\tau\tau$ channel with the CMS detector,” Presented 2019.

[84] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector,” *JINST*, vol. 12, 2017. DOI: `10.1088/1748-0221/12/10/P10003`.

[85] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET,” CERN, Tech. Rep., 2009. [Online]. Available: `https://cds.cern.ch/record/1194487`.

[86] W. Adam et. al, “Track Reconstruction in the CMS tracker,” CERN, Tech. Rep., 2006. [Online]. Available: `https://cds.cern.ch/record/934067`.

[87] CMS collaboration, “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV,” *JINST*, vol. 7, 2012. DOI: `10.1088/1748-0221/7/10/p10002`.

[88] CMS Collaboration, “Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV,” CERN, Tech. Rep. CMS-PAS-PFT-10-003, 2010. [Online]. Available: `https://cds.cern.ch/record/1279347`.

[89] R. Fruhwirth, “Application of Kalman filtering to track and vertex fitting,” *Nucl. Instrum. Methods Phys. Res Section A*, vol. 262, 1987. DOI: `10.1016/0168-9002(87)90887-4`.

[90] Muon POG, *Muon Identification for Run 2 phase (restricted)*. [Online]. Available: `https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2`.

[91] CMS Collaboration, “Muon identification and isolation efficiencies with 2017 and 2018 data,” 2018. [Online]. Available: `https://cds.cern.ch/record/2629364`.

[92] CMS Collaboration, “Muon reconstruction performance during Run II,” 2019. [Online]. Available: `https://cds.cern.ch/record/2682902`.

[93] S. Baffioni et al, “Electron reconstruction in CMS,” CERN, Tech. Rep., 2006. [Online]. Available: `http://cds.cern.ch/record/934070`.

[94] E/Gamma POG, *E/Gamma POG summary page (restricted)*. [Online]. Available: `https://twiki.cern.ch/twiki/bin/view/CMS/EgammaPOG`.

[95] E/Gamma POG, *ECAL Clustering (restricted)*. [Online]. Available: `https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideEcalRecoClustering`.

[96] E/Gamma POG, *Electron identification for Run 2 (restricted)*. [Online]. Available: `https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2`.

[97] C. Collaboration, “Performance of electron and photon reconstruction in Run 2 with the CMS experiment,” 2020. [Online]. Available: `https://cds.cern.ch/record/2725004`.

[98] CMS Collaboration, “Search for VBF Higgs bosons decaying to invisible particles at 13 TeV with full Run 2 data,” *CMS Analysis Note (internal)*, 2020. [Online]. Available: `http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%5C%20AN-2019/257`.

[99] CMS Collaboration, “Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV,” *JINST*, vol. 10, 2015. DOI: `10.1088/1748-0221/10/08/p08010`.

[100] E/Gamma POG, *Photon identification for Run 2 (restricted)*. [Online]. Available: `https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedPhotonIdentificationRun2`.

[101] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV,” *JINST*, vol. 12, 2017. DOI: `10.1088/1748-0221/12/02/p02014`.

[102] Jet/MET POG, *Jet energy Resolution and Corrections summary page (restricted)*. [Online]. Available: `https://twiki.cern.ch/twiki/bin/view/CMS/JetEnergyScale`.

[103] Jet/MET POG, *Jet information summary page (restricted)*. [Online]. Available: `https://twiki.cern.ch/twiki/bin/viewauth/CMS/JetMET`.

[104] Jet/MET POG, *et identification for Run 2 (restricted)*. [Online]. Available: `https://twiki.cern.ch/twiki/bin/view/CMS/JetID`.

[105] Jet/MET POG, *Jet pileup identification for Run 2 (restricted)*. [Online]. Available: `https://twiki.cern.ch/twiki/bin/view/CMS/PileupJetID`.

[106] J. Strologas, “Performance of Jet reconstruction in CMS at 13 TeV,” *PoS*, vol. ICHEP2016, 2016. DOI: `10.22323/1.282.0736`.

[107] C. Beluffi, “B jet Identification in CMS,” *Nuclear and Particle Physics Proceedings*, vol. 273-275, 2016. DOI: `10.1016/j.nuclphysbps.2015.09.435`.

[108] B-tag POG, *Recommendation for Using b-tag Objects in Physics Analyses (restricted)*. [Online]. Available: `https://twiki.cern.ch/twiki/bin/viewauth/CMS/BtagRecommendation`.

[109] Tau POG, *Tau ID recommendations for Run-2: 2016, 2017, and 2018 (restricted)*. [Online]. Available: `https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauIDRecommendationForRun2`.

[110] CMS Collaboration, "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS," *JINST*, 2011. DOI: `10.1088/1748-0221/6/11/P11002`.

[111] Jet/MET POG, *Mitigation of anomalous missing transverse energy (restricted).* [Online]. Available: `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsDP20018`.

[112] CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV," *JINST*, vol. 10, 2015. DOI: `10.1088/1748-0221/10/02/P02006`.

[113] A. A. and, "Performance of MET reconstruction in CMS," *JoP: Conference Series*, vol. 587, 2015. DOI: `10.1088/1742-6596/587/1/012006`.

[114] CMS Collaboration, "CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV," CERN, Tech. Rep. CMS-PAS-LUM-17-004, 2018. [Online]. Available: `https://cds.cern.ch/record/2621960`.

[115] CMS Collaboration, "CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV," CERN, Tech. Rep. CMS-PAS-LUM-18-002, 2019. [Online]. Available: `https://cds.cern.ch/record/2676164`.

[116] C. Oleari, "The POWHEG BOX," *Nuclear Physics B - Proceedings Supplements*, vol. 205-206, 2010. DOI: `10.1016/j.nuclphysbps.2010.08.016`.

[117] J. Alwall et al, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations," *JHEP*, vol. 2014, 2014. DOI: `10.1007/jhep07(2014)079`.

[118] T. Sjostrand et al, "An introduction to PYTHIA 8.2," *Computer Physics Communications*, vol. 191, 2015. DOI: `10.1016/j.cpc.2015.01.024`.

[119] S. Agostinelli et al, "Geant4 - a simulation toolkit," *Nucl. Instrum. Methods Phys. Res Section A*, vol. 506, 2003. DOI: `10.1016/S0168-9002(03)01368-8`.

[120] CMS Lumi POG, *Pileup recommendations for Run 2 (restricted).* [Online]. Available: `https://twiki.cern.ch/twiki/bin/view/CMS/%20PileupJSONFileforData#Pileup_JSON_Files_For_Run_II`.

[121] L. Thomas, *Reweighting recipe to emulate Level-1 ECAL prefiring (restricted).* [Online]. Available: `https://twiki.cern.ch/twiki/bin/view/CMS/L1ECALPrefiringWeightRecipe`.

[122] E/Gamma POG, *Electron efficiencies and scale factors (restricted).* [Online]. Available: `https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2`.

[123] Muon POG, *Reference efficiencies for muon ID and isolation with Run 2 data (restricted).* [Online]. Available: `https://twiki.cern.ch/twiki/bin/view/CMS/MuonPOG`.

[124] B tag POG, *Methods to apply b-tagging efficiency scale factors (restricted).* [Online]. Available: `https://twiki.cern.ch/twiki/bin/viewauth/CMS/BTagSFMethods`.

[125] CMS Collaboration, "Search for VBF Higgs bosons decaying to invisible particles at 13 TeV with 2016 data," *CMS Analysis Note (restricted)*, 2018. [Online]. Available: `http://cms.cern.ch/iCMS/jsp/db_notes/noteInfo.jsp?cmsnoteid=CMS%5C%20AN-2016/4187`.

[126] E/Gamma POG, *Electron Tag-and-Probe recipe (restricted).* [Online]. Available: `https://twiki.cern.ch/twiki/bin/view/CMSPublic/ElectronTagAndProbe`.

[127] G. Cowan et al, "Asymptotic formulae for likelihood-based tests of new physics," *Eur. Phys. J. C*, vol. 71, 2011. DOI: `10.1140/epjc/s10052-011-1554-0`.

[128] A. L. Read, "Presentation of search results: The CL(s) technique," *J. Phys. G*, vol. 28, 2002. DOI: `10.1088/0954-3899/28/10/313`.

[129] B. Cousins, *Probability Density Functions for Positive Nuisance Parameters.* 2010. [Online]. Available: `http://www.physics.ucla.edu/~cousins/stats/cousins_lognormal_prior.pdf`.

[130] W. E. et al., "Statistical methods in experimental physics," *Journal of the American Statistical Association*, 2013. DOI: `10.2307/2284113`.

[131] F.James, *Statistical Methods in Experimental Physics; 2nd ed.* World Scientific, 2006. [Online]. Available: `https://cds.cern.ch/record/1019859`.

[132] D. de Florian et al, "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector," 2016. DOI: `10.23731/CYRM-2017-002`.

[133] J. Butterworth et al, "PDF4LHC recommendations for LHC Run II," *J. Phys. G*, vol. 43, 2016. DOI: `10.1088/0954-3899/43/2/023001`.

[134] LUX Collaboration, "Results from a search for dark matter in the complete LUX exposure," *Phys. Rev. Lett.*, vol. 118, 2017. DOI: `10.1103/PhysRevLett.118.021303`.

[135] SuperCDMS Collaboration, "Search for Low-Mass Dark Matter with CDMSlite Using a Profile Likelihood Fit," *Phys. Rev. D*, vol. 99, 2019. DOI: `10.1103/PhysRevD.99.062001`.

[136] XENON Collaboration, “Dark Matter Search Results from a One Ton-Year Exposure of XENON1T,” *Phys. Rev. Lett.*, vol. 121, 2018. DOI: `10.1103/PhysRevLett.121.111302`.

[137] XENON Collaboration, “Light Dark Matter Search with Ionization Signals in XENON1T,” *Phys. Rev. Lett.*, vol. 123, 2019. DOI: `10.1103/PhysRevLett.123.251801`.

[138] CRESST Collaboration, “First results from the CRESST-III low-mass dark matter program,” *Phys. Rev. D*, vol. 100, 2019. DOI: `10.1103/PhysRevD.100.102002`.

[139] Panda-II Collaboration, “Dark Matter Results From 54-Ton-Day Exposure of PandaX-II Experiment,” *Phys. Rev. Lett.*, vol. 119, 2017. DOI: `10.1103/PhysRevLett.119.181302`.

[140] Faster Analysis Software Taskforce, *FAST framework*. [Online]. Available: `http://fast-hep.web.cern.ch/fast-hep/`.

[141] Brian Ingerson et al, *Yet Another Markup Language (YAML) 1.0*. [Online]. Available: `https://yaml.org/spec/history/2001-12-10.html`.

[142] Eshwen Bhal, “Searches for dark matter with a focus on invisibly decaying Higgs bosons using the full Run-2 dataset of the CMS experiment at the LHC,” Presented 2020.

[143] G. Apollinari et al, “Chapter 1: High Luminosity Large Hadron Collider HL-LHC. High Luminosity Large Hadron Collider HL-LHC,” *CERN Yellow Report*, 2017. DOI: `10.5170/CERN-2015-005.1`.

[144] CMS Collaboration, “The Phase-2 Upgrade of the CMS Endcap Calorimeter,” CERN, Tech. Rep. CERN-LHCC-2017-023, 2017. [Online]. Available: `https://cds.cern.ch/record/2293646`.

[145] J. de Favereau et al, “DELPHES 3, A modular framework for fast simulation of a generic collider experiment,” *JHEP*, vol. 2, 2014. DOI: `10.1007/JHEP02(2014)057`.