

# **Searches for invisibly decaying Higgs bosons with the CMS detector**

Patrick James Dunne  
of Imperial College London

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

## Abstract

## Declaration

This dissertation is the result of my own work, except where explicit reference is made to the work of others, and has not been submitted for another qualification to this or any other university. This dissertation does not exceed the word limit for the respective Degree Committee.

Patrick Dunne

## Acknowledgements

# Contents

<b>1. Introduction and theory</b>	<b>2</b>
1.1. The standard model of particle physics . . . . .	2
1.1.1. Fundamental particles in nature . . . . .	3
1.1.2. Introduction to gauge theories . . . . .	5
1.2. The SM gauge group and fundamental particle representations . . . . .	6
1.2.1. Spontaneous symmetry breaking and the Higgs mechanism . . . . .	8
1.2.2. Higgs boson production and decay at the LHC . . . . .	10
1.2.3. Challenges for the SM . . . . .	10
1.3. Dark matter . . . . .	10
1.4. Some extensions of the standard model incorporating dark matter . . . . .	10
1.5. Simulation . . . . .	10
1.6. Statistics of exclusion limits . . . . .	10
<b>2. The LHC and the CMS experiment</b>	<b>12</b>
2.1. The LHC . . . . .	12
2.2. The CMS experiment . . . . .	15
2.2.1. Tracker . . . . .	19
2.2.2. Electromagnetic calorimeter . . . . .	20
2.2.3. Hadronic calorimeter . . . . .	22
2.2.4. Muon system . . . . .	23
2.2.5. Trigger system and data processing . . . . .	25
<b>3. Physics objects and event reconstruction</b>	<b>29</b>
3.1. Tracks . . . . .	29
3.2. Primary vertex . . . . .	30
3.3. Particle Flow . . . . .	31
3.4. Electrons . . . . .	32
3.5. Muons . . . . .	34

3.6. Jets . . . . .	35
3.6.1. Jet clustering . . . . .	35
3.6.2. Jet identification . . . . .	37
3.6.3. Jet energy corrections . . . . .	37
3.7. Missing transverse energy . . . . .	38
3.8. Taus . . . . .	40
<b>4. Search for invisibly decaying VBF produced Higgs bosons in Run 1 prompt data</b>	<b>42</b>
4.1. Event selection . . . . .	42
4.1.1. Trigger . . . . .	43
4.1.2. Offline selection . . . . .	46
4.2. Background estimation . . . . .	47
4.2.1. $W \rightarrow e\nu + \text{jets}$ . . . . .	47
4.2.2. $W \rightarrow \mu\nu + \text{jets}$ . . . . .	49
4.2.3. $W \rightarrow \tau\nu + \text{jets}$ . . . . .	51
4.2.4. $Z \rightarrow \nu\nu + \text{jets}$ . . . . .	53
4.2.5. QCD . . . . .	58
4.2.6. Minor backgrounds . . . . .	60
4.3. Systematic uncertainties . . . . .	60
4.3.1. Jet energy scale . . . . .	60
4.3.2. Jet energy resolution . . . . .	62
4.3.3. Unclustered energy scale . . . . .	63
4.3.4. Lepton identification and isolation efficiency . . . . .	63
4.3.5. Other uncertainties . . . . .	64
4.4. Results . . . . .	64
<b>5. Search for invisibly decaying Higgs bosons in Run 1 parked data</b>	<b>69</b>
5.1. Trigger . . . . .	69
5.2. Event selection . . . . .	73
5.2.1. Preselection . . . . .	73
5.2.2. Signal region selection . . . . .	77
5.3. Background estimation . . . . .	78
5.3.1. Top quarks . . . . .	79
5.3.2. $W \rightarrow e\nu + \text{jets}$ . . . . .	82
5.3.3. $W \rightarrow \mu\nu + \text{jets}$ . . . . .	82
5.3.4. $W \rightarrow \tau\nu + \text{jets}$ . . . . .	84

5.3.5.	Differences between the $W \rightarrow e\nu$ , $W \rightarrow \mu\nu$ and $W \rightarrow \tau\nu + \text{jets}$ backgrounds . . . . .	86
5.3.6.	$Z \rightarrow \nu\nu + \text{jets}$ . . . . .	89
5.3.7.	$V + \text{jets}$ closure tests . . . . .	94
5.3.8.	$V + \text{jets}$ scale factor investigations . . . . .	94
5.3.9.	QCD . . . . .	94
5.3.10.	Minor backgrounds . . . . .	94
5.4.	Systematic uncertainties . . . . .	94
5.5.	Results . . . . .	94
<b>6.</b>	<b>Combinations and interpretations of Run 1 searches for invisibly decaying Higgs bosons</b>	<b>95</b>
6.1.	Searches in other channels . . . . .	95
6.2.	Combination with prompt VBF search . . . . .	95
6.3.	Combination with the parked VBF search . . . . .	95
6.4.	Dark matter interpretations . . . . .	95
<b>7.</b>	<b>Search for invisibly decaying Higgs bosons in Run 2 data</b>	<b>96</b>
7.1.	Trigger efficiency measurements . . . . .	96
<b>A.</b>	<b>Parked data trigger efficiencies</b>	<b>97</b>
<b>Bibliography</b>		<b>104</b>
<b>List of figures</b>		<b>109</b>
<b>List of tables</b>		<b>115</b>
<b>Acronyms</b>		<b>117</b>

# Chapter 1.

## Introduction and theory

In order to describe the search for invisible decays of the Higgs boson (“Higgs to invisible”), it is necessary to describe the theory behind them and also the statistical techniques used in carrying out the search. This chapter will start with an introduction to the current best theory of particle physics, the standard model (SM), focussing on the Higgs mechanism, before outlining the motivations behind and some candidates for physics beyond the SM (BSM), then concluding with a discussion of the statistics of hypothesis testing. Natural units, where  $\hbar = c = 1$ , Einstein summation convention and Feynman slash notation are used throughout. Four vector indices are labelled using greek letters, and gauge group generators using roman letters.

### 1.1. The standard model of particle physics

The SM describes the interaction of the particles currently thought to be fundamental with the strong, weak and electromagnetic forces. Its predictions, which come from specifying the symmetries the theory respects and how they are broken, the particles in the theory, and 18 free parameters, have been tested in many different experiments, in some cases up to one part in a trillion [1]. However, it does face challenges, described in section Section 1.2.3, one example being that it does not describe dark matter (DM).

The SM is a gauge invariant quantum field theory (QFT). To construct a QFT the symmetries that are respected by the theory and the fields it describes must be specified. The symmetries are important because of Noether’s theorem, which states that for every continuously differentiable symmetry of the Lagrangian of a theory there is a corresponding conservation law [2, 3]. An example of this is Poincaré invariance, the

invariance of the laws of physics under translations and rotations in space and time, which leads through Noether's theorem to the conservation of energy, linear momentum and angular momentum. In addition to giving rise to conservation laws, some types of symmetry lead to additional fields being required to preserve invariance, this will be discussed further in Section 1.1.2 [4].

It is important to specify the fields described by the QFT as these are constrained by the fundamental particles seen in nature. This is because particles correspond to the quantised excitations of fields. Specifically, scalar fields correspond to spin zero bosons, spinor fields correspond to spin half fermions, and vector fields correspond to spin 1 bosons. In order to add a new field an explanation for why the corresponding particle has not yet been observed must, therefore, be provided. We will now go through the particles observed in nature and how they are represented in the SM.

### 1.1.1. Fundamental particles in nature

There are two types of fundamental particles in nature, fermions and bosons. The fermions observed in nature that are currently thought to be fundamental are then divided into those which interact via the strong nuclear force (the quarks), and those which don't (the leptons). Both the quarks and leptons have two further types: charged and neutral in the case of the leptons, and up type and down type in the case of the fermions. Another interesting feature of the fermions is that they are arranged in three generations. Each generation has one fermion of each type with the same quantum numbers as those in the other generations, except that the mass is different. Table 1.1 shows this structure.

The bosons in nature also have two types. The first type are vector bosons which mediate the three fundamental interactions described by the SM. The vector bosons are summarised in Table 1.2, where it can be seen that their masses are very different, the photon and the eight gluons being massless, while the  $W^\pm$  and Z bosons are very massive. As we will see in Section 1.2.1 explaining these masses requires the Higgs mechanism. The Higgs mechanism also gives rise to the other type of boson seen in nature, the scalar Higgs boson. In order to see how all of the above particles are represented in the SM an introduction to gauge theories is necessary.

**Table 1.1.:** The fundamental fermions observed in nature separated into their three generations. Each particle shown also has an antiparticle with opposite charge and identical mass. Values taken from [5]

Generation	Leptons			Hadrons		
	Particle	Mass	Charge	Particle	Mass	Charge
1	$e^-$	511 keV	-1	u	2.3 MeV	$+\frac{2}{3}$
	$\nu_e$	$\sim 0$	0	d	4.8 MeV	$-\frac{1}{3}$
2	$\mu^-$	105.7 MeV	-1	c	1.275 GeV	$+\frac{2}{3}$
	$\nu_\mu$	$\sim 0$	0	s	95 MeV	$-\frac{1}{3}$
3	$\tau^-$	1.777 GeV	-1	t	173.2 GeV	$+\frac{2}{3}$
	$\nu_\tau$	$\sim 0$	0	b	4.18 GeV	$-\frac{1}{3}$

**Table 1.2.:** The fundamental vector bosons observed in nature separated by the force which they mediate. Values taken from [5].

Force	Particle	Mass	Charge
Electromagnetism	$\gamma$	0	0
Weak	$W^\pm$	80.4 GeV	$\pm 1$
	Z	91.2 GeV	0
Strong	g	0	0

### 1.1.2. Introduction to gauge theories

Gauge symmetries are local transformations, i.e. the transformation can be different at different points in space and time, that form a symmetry group. To see the effect of imposing such a symmetry on a theory consider imposing local invariance under U(1) transformations on the Dirac Lagrangian for a massive fermion:

$$\mathcal{L} = i\bar{\psi}\not{\partial}\psi - m\bar{\psi}\psi \quad [6]. \quad (1.1)$$

This Lagrangian is invariant under a global  $U(1)$  transformation  $\psi \rightarrow e^{iq\theta}\psi$ , where  $q$  and  $\theta$  are constant. However, if the  $U(1)$  transformation is local i.e.  $\theta$  is a function of spacetime position the Lagrangian is no longer invariant and transforms as:

$$\mathcal{L} \rightarrow \mathcal{L} - q(\partial_\mu\theta)\bar{\psi}\gamma^\mu\psi. \quad (1.2)$$

In order to restore invariance a vector field,  $A_\mu$ , referred to as a gauge field or gauge boson, which transforms as  $A_\mu \rightarrow A_\mu + \partial_\mu\theta$  and has an interaction with the fermion field:

$$\mathcal{L}_{int} = q(\bar{\psi}\gamma^\mu\psi)A_\mu, \quad (1.3)$$

can be added to the theory. The interaction term of the new gauge field transforms as:

$$\mathcal{L}_{int} \rightarrow \mathcal{L}_{int} + q(\partial_\mu\theta)\bar{\psi}\gamma^\mu\psi, \quad (1.4)$$

which cancels out the non-gauge invariance seen in Equation 1.2.

Assuming the new gauge field to be massless the Lagrangian is now:

$$\mathcal{L} = i\bar{\psi}\not{\partial}\psi - m\bar{\psi}\psi + q(\bar{\psi}\gamma^\mu\psi)A_\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (1.5)$$

where  $F_{\mu\nu}$  is the field strength tensor of the vector field. For a gauge boson from a general gauge group is written as:

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f^{abc} A_\mu^b A_\nu^c, \quad (1.6)$$

where  $f^{abc}$  are the structure constants of the gauge group, which are a representation of the comutation relations between the group's generators. For  $U(1)$  which only has one self-commuting generator the single structure constant is 0. However, for non-Abelian

gauge groups (i.e. those with non-commuting generators) they can be non-zero causing the  $F_{\mu\nu}F^{\mu\nu}$  term in the Lagrangian to include self-interaction terms of the vector bosons.

It is also interesting to note that Equation 1.5 can be rewritten as:

$$\mathcal{L} = i\bar{\psi}\gamma^\mu \mathcal{D}_\mu \psi - m\bar{\psi}\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \quad (1.7)$$

where  $\mathcal{D}_\mu = \partial_\mu + iqA_\mu$  and is referred to as the covariant derivative. Comparing Equation 1.1 and Equation 1.7 it can be seen that to go from a globally invariant Lagrangian to a locally invariant one we have substituted the normal spacetime derivative for the covariant derivative and added the free term of the vector field.

$U(1)$  transformations have one degree of freedom and can be described by one parameter, in the above case  $\theta$ , and in order to make the Lagrangian locally invariant one interacting gauge boson had to be added. This correspondence between the number of degrees of freedom and the number of gauge bosons holds generally. For each degree of freedom of a group's transformations there exists a generator of the group, and for each generator one interacting gauge boson must be added to achieve local invariance.

## 1.2. The SM gauge group and fundamental particle representations

The SM is gauge invariant under the group  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ . Fermions in the SM are spin half spinor representations of these symmetry groups. These spinors can be split into chirally left and right handed components using the projection operators  $P_L = \frac{1}{2}(1 \mp \gamma^5)$ . Chirally left and right handed fermions transform differently under  $SU(2)_L$ . The right handed spinors are not charged under  $SU(2)_L$  and thus are represented as a singlet, while the left handed spinors transform as a doublet.

The first generation of leptons can, therefore, be written as:

$$\psi_1 = e_R, \psi_2 = L = \begin{pmatrix} \nu_e \\ e_L \end{pmatrix}. \quad (1.8)$$

The SM treats neutrinos as massless and has no right handed neutrino. Similarly the first generation of quarks can be written as:

$$\psi_3 = u_R, \psi_4 = d_R, \psi_5 = \begin{pmatrix} u_L \\ d_L \end{pmatrix}. \quad (1.9)$$

As we saw in Section 1.1.2 gauge symmetries in theories with fermions require the addition of an interacting vector boson per symmetry generator to preserve gauge invariance.  $SU(3)_C$  has eight generators whose eight vector bosons,  $G_{a\mu}$ , correspond to the eight physical gluons of Quantum Chromodynamics (QCD), which mediate the strong interaction.  $SU(2)_L$  has three generators whose three vector bosons,  $W_\mu^i$ , mix with the one vector boson from  $U(1)_Y$ ,  $B_\mu$  unifying the electromagnetic and weak forces into one electroweak force. The physical states that result are:

$$\begin{aligned} W^\pm_\mu &= \frac{1}{\sqrt{2}} (W_\mu^1 \mp i W_\mu^2) \\ Z_\mu &= \cos(\theta_W) W_\mu^3 - \sin(\theta_W) B_\mu \\ A_\mu &= \sin(\theta_W) W_\mu^3 + \cos(\theta_W) B_\mu, \end{aligned} \quad (1.10)$$

where  $\theta_W$  is the Weinberg angle and  $A_\mu$  is the photon field. Also, as described in Section 1.1.2 the interaction between these vector bosons and the fermion fields occurs through their presence in the covariant derivative, and interactions between the vector bosons occur because  $SU(3)_C$  and  $SU(2)_L$  are non-Abelian.

Now let us try to construct a Lagrangian for these fields. First ignoring the masses we find:

$$\mathcal{L} = i\bar{\psi}_i \not{D} \psi_i - \frac{1}{4} F_{\mu\nu j} F_j^{\mu\nu}, \quad (1.11)$$

where the sum over all  $\psi$  also includes the second and third generations,  $F_{\mu\nu j} F_j^{\mu\nu}$  is a sum of the free terms of all the SM gauge bosons and  $\not{D}$  is the SM covariant derivative:

$$\not{D}_\mu = \partial_\mu + ig_1 \frac{Y}{2} B_\mu + ig_2 \frac{\tau_i}{2} W_\mu^i + ig_3 \frac{\lambda_a}{2} G_\mu^a, \quad (1.12)$$

with  $Y$  being the constant generator of  $U(1)$ ,  $\tau_i$  the generators of  $SU(2)_L$ ,  $\lambda_a$  the generators of  $SU(3)_C$  and  $g_i$  the coupling constants of the fields. It should be noted that  $\frac{g_1}{g_2}$  is equal to  $\tan(\theta_W)$ .

When we try to include mass a problem occurs. We know that some of the fermions have mass, and consequently we should have fermion mass terms of the form:

$$\begin{aligned}\mathcal{L}_{m_f} &= -m_f \bar{f} f \\ &= -m_f \bar{f} \left[ \frac{1}{2} (1 - \gamma^5) + \frac{1}{2} (1 + \gamma^5) \right] f \\ &= -m_f (\bar{f}_R f_L + \bar{f}_L f_R),\end{aligned}\tag{1.13}$$

in our Lagrangian. However, as the left and right handed fields do not transform in the same way under  $SU(2)_L$  this term breaks the gauge symmetry of the Lagrangian and can't be present.

A similar problem occurs for vector fields. In Section 1.1.2 we didn't consider the mass term of these vector fields:

$$\mathcal{L}_{m_V} = \frac{1}{2} m_V^2 A_\mu A^\mu,\tag{1.14}$$

which is not gauge invariant, so massive vector bosons are not possible on their own in gauge invariant theories either. The additional piece of the SM required to allow particles to have mass is the Higgs mechanism.

### 1.2.1. Spontaneous symmetry breaking and the Higgs mechanism

The Higgs mechanism is a form of spontaneous symmetry breaking. A symmetry is said to be spontaneously broken when the Lagrangian remains invariant while the vacuum state, i.e. that with lowest energy, does not. Terms which are not gauge invariant can then be incorporated into the theory by adding a field which has a non-zero vacuum expectation value and coupling it to the other fields present in the term. For the Higgs mechanism this field is a complex scalar  $SU(2)_L$  doublet, called the Higgs field:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}.\tag{1.15}$$

The main part of the Higgs field Lagrangian is:

$$\mathcal{L} = T - V = (\mathcal{D}_\mu \phi)^\dagger (\mathcal{D}^\mu \phi) + \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2.\tag{1.16}$$

For  $\mu^2 > 0$  the minima of the potential are non-zero and form a circle in phase space of  $\phi$ . All of these vacua are equivalent and a particular vacuum can be chosen with no physical effect. By convention we choose the following vacuum:

$$\langle 0 | \phi | 0 \rangle = \begin{pmatrix} 0 \\ \sqrt{\frac{\mu^2}{2\lambda}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \quad (1.17)$$

Next we consider small perturbations around this vacuum, ignoring those that can be set to zero by gauge freedom gives:

$$\phi = \begin{pmatrix} 0 \\ v + H \end{pmatrix}. \quad (1.18)$$

Inserting this into Equation 1.16 and ignoring terms with more than one type of field gives at leading order:

$$\mathcal{L} = \frac{1}{2} \partial_\mu H \partial^\mu H - \frac{1}{2} \mu^2 H^2 + \frac{v^2}{8} [g_2^2 W_\mu^+ W^{+\mu} + g_2^2 W_\mu^- W^{-\mu} + (g_1^2 + g_2^2) Z_\mu Z^\mu]. \quad (1.19)$$

As expected, the weak vector bosons  $W_\mu^\pm$  and  $Z_\mu$  acquire masses  $\frac{gv}{2}$  and  $\frac{v}{2}\sqrt{g_1^2 + g_2^2}$  respectively. We also see an additional massive scalar  $H$ , which is the Higgs boson, which has mass  $\sqrt{2}\mu$ . The photon and gluons do not acquire masses as the particular choice of coupling constants and the structure of the group generators leads to the terms in  $A_\mu$  and  $G_{\mu a}$  being zero.

The final part of the Higgs field Lagrangian is that giving rise to the fermion masses. These are generated by a Yukawa term in the Lagrangian for each fermion as follows:

$$\mathcal{L}_{Yuk} = k_f (\bar{f}_L \phi f_R + \bar{f}_R \phi^\dagger f_L). \quad (1.20)$$

The fermion's mass is then  $\frac{k_f v}{\sqrt{2}}$ .

### 1.2.2. Higgs boson production and decay at the LHC

### 1.2.3. Challenges for the SM

## 1.3. Dark matter

## 1.4. Some extensions of the standard model incorporating dark matter

## 1.5. Simulation

## 1.6. Statistics of exclusion limits

Limits on the parameters of theoretical models are presented throughout this thesis. The limits are set by performing a hypothesis test to discriminate between a null, background physics model only, hypothesis,  $b$  and a test hypothesis, the signal process,  $s$ , plus background model. The particular procedure used is based on the  $CL_S$  statistic and was developed by the LHC Higgs Combination Group and is used by both the ATLAS and CMS experiments [7].

The procedure starts by defining a likelihood function,  $\mathcal{L}$ , which quantifies how likely a given observation is given the expectation under a given hypothesis.  $\mathcal{L}$  takes the form:

$$\mathcal{L} = \prod_i Poisson(n_i|\nu_i(\mu, \theta)) \cdot \prod_j Constraint(\theta_j, \bar{\theta}), \quad (1.21)$$

where the first term is the contribution from the Poisson probability to observe  $n_i$  events in each analysis category,  $i$ , considered given a predicted number of events from the hypothesis,  $\nu_i$ .  $\nu_i$  is a function of a signal strength parameter,  $\mu$ , which in the case of the signal hypothesis being an SM Higgs boson is 1 for the SM and 0 for the background only case, and the “nuisance parameters,”  $\theta$ , which account for the uncertainties on parameters of the signal and background models and any correlations between them. The second term in Equation 1.21 represents the constraints on the allowed values of these nuisance parameters, with  $\bar{\theta}$  being the best estimate of  $\theta$  obtained from external measurements. The shape of the constraint function varies depending on the nuisance

parameter it represents. For example, uncertainties on the event yield in a category are usually modelled with log-normal constraints, which exclude negative values of the event yield.

Profile likelihood ratios,  $q_\mu$ , are then calculated, which are defined as:

$$q_\mu = -2 \ln \frac{\mathcal{L}(obs|\mu \cdot s + b, \hat{\theta}_\mu)}{\mathcal{L}(obs|\hat{\mu} \cdot s + b, \hat{\theta})}, \quad (1.22)$$

where  $obs$  is the observation, and  $\hat{\mu}$  and  $\hat{\theta}$  are the values of  $\mu$  and the  $\theta$  where the likelihood is maximised given the constraint  $0 \geq \hat{\mu} \geq \mu$ .  $\hat{\theta}_\mu$  are the values of the nuisance parameters that maximise the likelihood for a given  $\mu$ . The profile likelihood ratio therefore describes how likely it is to observe a signal strength equal to or higher than  $\mu$  compared to the most likely signal strength.

Finally the  $CL_s$  statistic itself is defined as:

$$CL_s = \frac{P(q_\mu \geq q_\mu^{obs} | \mu \cdot s + b)}{P(q_\mu \geq q_\mu^{obs} | b)}, \quad (1.23)$$

Where the probability  $P$  of a given  $q_\mu$  is calculated using the asymptotic limit approximation [8]. The region in which a signal strength  $\mu \cdot s$  is excluded with  $1 - \alpha$  confidence level (CL) is then the region for which  $CL_s$  is less than or equal to  $\alpha$ , i.e. when the signal hypothesis is  $\alpha$  times less probably than the background.

# Chapter 2.

## The LHC and the CMS experiment

This chapter introduces the CMS [9] experiment and the LHC [10]. In Section 2.1 an overview of the LHC and the chain of accelerators which feed into it is given. This is then followed in Section 2.2 by a description of the CMS experiment focusing on the aspects most relevant to the search for invisibly decaying Higgs bosons.

### 2.1. The LHC

The LHC is situated 100m underground in a tunnel formerly built for the LEP accelerator [11] at CERN near Geneva, Switzerland. It is a 27km storage ring which accelerates both protons and heavy ions and collides them at the highest centre of mass energies of any collider built to date. The work contained in this thesis uses data from proton-proton collisions. These protons are obtained by taking hydrogen gas and stripping its atoms of their electrons with an electric field. The first accelerator in the chain of accelerators feeding into the LHC, Linac 2, accelerates the protons to 50 MeV. The protons are then accelerated to 1.4 GeV by the next accelerator, the Proton Synchrotron Booster (PSB), which is followed by the Proton Synchrotron (PS) where they reach 25 GeV. The beam energy is then increased to 450 GeV in the Super Proton Synchrotron (SPS). Finally, the protons are injected into the LHC where, at time of writing, the maximum energy the beams have been accelerated to is 6.5 TeV, close to the design maximum of 7 TeV.

When fully filled the LHC contains two counter-rotating beams which are formed of up to 2808 bunches spaced either 25 or 50 ns apart and each containing  $\mathcal{O}(10^{11})$  protons. The two beams are kept travelling in a circle by 1232 superconducting dipole magnets and steered to four collision points around the LHC. Detectors are situated at these



**Figure 2.1.:** The layout of the chain of accelerators feeding into the LHC, showing the position of the four main detectors.

collision points to observe the interactions, the main four being: ALICE [12], ATLAS [13], CMS [9] and LHCb [14]. A schematic of the chain of accelerators feeding into the LHC and the LHC detectors can be seen in Figure 2.1.

When studying a physical process occurring in particle collisions it is important to know how many times it will occur, this can be expressed as:

$$N = \mathcal{L}\sigma, \quad (2.1)$$

where  $\mathcal{L}$ , the integrated luminosity, depends only on the parameters of the collisions, and the cross-section,  $\sigma$ , depends only on the process. In order to observe rare (i.e. low cross-section) processes, such as those studied at the LHC, it is, necessary to use very high luminosity datasets. The integrated luminosity is obtained by integrating the instantaneous luminosity over time, so large luminosities can be obtained either by running the accelerator for a long time, or by operating at high instantaneous luminosity. For collisions at the LHC the instantaneous luminosity is given by:

$$\mathcal{L} = \frac{k_b N_b^2 f_{rev} \gamma}{4\pi \epsilon_n \beta} [15], \quad (2.2)$$



**Figure 2.2.:** A summary of the luminosity delivered to CMS during Run 1 of the LHC [16].

where  $k_b$  is the number of bunches per beam,  $N_b$  the number of protons per bunch,  $f_{rev}$  the revolution frequency,  $\epsilon_n$  the normalised transverse beam emittance,  $\beta^*$  the beta-function at the interaction point and  $\gamma$  the Lorentz factor. The design instantaneous luminosity of the LHC is  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$  with 25ns bunch spacing. The integrated luminosity is defined as  $\mathcal{L}_{int} = \int \mathcal{L} dt$ .

The LHC started physics runs in 2010, during which it operated at a centre of mass energy of 7 TeV and delivered an integrated luminosity of  $44.2 \text{ pb}^{-1}$  to CMS. In 2011 the LHC also operated at 7 TeV and delivered  $6.1 \text{ fb}^{-1}$  to CMS. The centre of mass energy was increased to 8 TeV in 2012 and  $23.3 \text{ fb}^{-1}$  of data were delivered to CMS. A summary of the luminosity delivered to CMS during the three periods of Run 1 can be seen in Figure 2.2. In Run 2 the centre of mass energy was further increased to 13 TeV and during 2015  $4.09 \text{ fb}^{-1}$  of data were delivered to CMS at this energy. In order to be used for physics analysis data must be certified. This certification ensures that the detector was fully operational when the data were recorded. In 2011  $5.1 \text{ fb}^{-1}$  were certified, in 2012  $19.7 \text{ fb}^{-1}$  were certified and in 2015  $2.2 \text{ fb}^{-1}$  were certified.

The cross-sections for several processes are shown in Figure 2.3 and it can be seen that the cross-section for VBF Higgs production is approximately 1.5 pb. Therefore, we expect approximately 30000 VBF produced Higgs bosons in the 2012 dataset. By contrast the vector boson production cross-section is approximately 100 nb and the total cross-section for any process is orders of magnitude higher still. The separation of the relatively small number of signal events from the large background is a major challenge for the search for invisibly decaying Higgs bosons.

The large total cross-section combined with the high instantaneous luminosities that the LHC operates at leads to the probability for multiple proton-proton interactions per bunch crossing being high. The distribution of the number of interactions per bunch crossing can be seen in Figure 2.4. The additional interactions on top of the process of interest in a bunch crossing are called pile-up (PU).

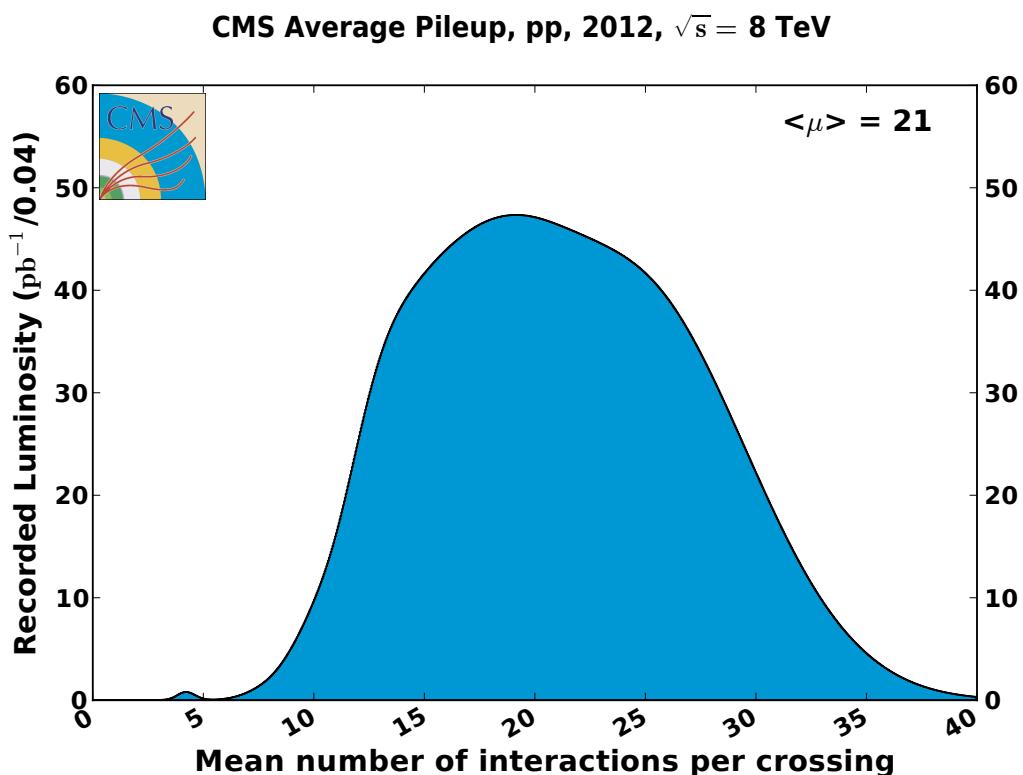
## 2.2. The CMS experiment

The CMS detector was designed to search for the SM Higgs and new physics at the TeV energy scale. Both because the nature of new physics is not known and the SM Higgs has a wide range of decays and production mechanisms CMS must be sensitive to many different types of final state particles and topologies. In order to achieve this it has a hermetic design comprising a barrel, endcaps and a forward calorimetry system, and is also composed of several layers of subdetectors each sensitive to different particles as shown in Figure 2.5. The hermiticity of the detector is particularly important for the VBF Higgs to invisible search, because, as described in Section 1.2.2, the VBF final state is highly likely to have jets in the forward regions of the detector. Further details on the CMS detector beyond those in this section can be found in Ref. [9].

A central design feature of CMS is the superconducting magnet, inside which is generated a 3.8T axial field. This field bends the path of charged particles travelling through it allowing their momentum to be measured. Not all particles are charged however, and the path of several types of particles through the CMS detector is shown in Figure 2.6. The first layer is the tracker which records the paths taken by charged particles, as well as providing a momentum measurement the tracks also allow the vertex from which the particle came to be identified. The next layer is the electromagnetic calorimeter (ECAL) where electrons and photons deposit energy through electromagnetic showers. This is followed by the hadron calorimeter (HCAL) where hadrons deposit most of their energy.



**Figure 2.3.:** Cross-sections for several processes in collisions of protons with protons or anti-protons as a function of centre of mass energy. The energies that the LHC and Tevatron ran at are highlighted [17].



**Figure 2.4.:** Distribution of the number of interactions per bunch crossing in CMS during 2012 running of the LHC [16].



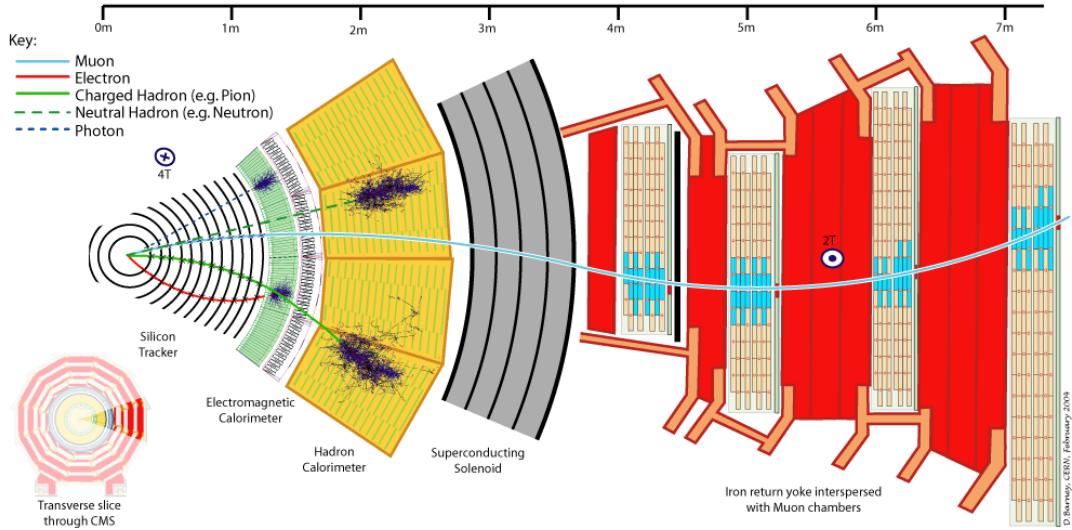
**Figure 2.5.:** A diagram of the subsystems making up the CMS detector, illustrating the hermeticity and layered structure of the experiment [18].

After the calorimetry systems is the superconducting magnet which is not instrumented. Outside the magnet are the muon detection systems, which are interspersed with iron plates which form the return yoke for the magnet. Due to their high mass compared to electrons, muons do not deposit much energy in the detector and often are not stopped, so the muon system is primarily a tracking detector.

The origin of the co-ordinate system used by CMS is at the nominal interaction point. It is a right handed cartesian system with the x axis pointing towards the centre of the LHC ring and the y-axis vertically upwards, the z axis then points along the beam line. The azimuthal angle  $\phi$  and the polar angle  $\theta$  are measured in radians from the x and z axes respectively. It is common to describe the direction of outgoing particles using  $\phi$  and their pseudo-rapidity,  $\eta$  which is defined as:

$$\eta = -\ln[\tan(\theta/2)]. \quad (2.3)$$

Distances in the  $\eta - \phi$  plane are given by  $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$ . Two other quantities often used at hadron colliders are the projections of a particle's momentum and energy in the transverse plane, these are denoted  $p_T$  and  $E_T$  respectively. The missing transverse



**Figure 2.6.:** A schematic cross-section of the CMS experiment showing the path taken by several types of particles [19].

energy, defined as the negative vector sum of the momentum of all particles in an event, is important in inferring the presence of invisible particles and is denoted  $\cancel{E}_T$ . Also, from here on where the terms leading and sub-leading are used they refer to the highest and second-highest  $p_T$  objects in an event respectively.

### 2.2.1. Tracker

The tracker is designed to precisely measure the paths of charged particles from LHC collisions which curve in CMS's magnetic field. The design transverse momentum resolution of the full tracking detector is 1-2% at 100 GeV. In order to precisely measure the particles' positions and ensure the occupancy of the tracker is low a high granularity is required. Due to the frequency of collisions at the LHC and the high instantaneous luminosity a radiation hard system with fast response is also necessary. This combination of requirements motivates the use of a silicon based system. When traversing silicon charged particles create electron-hole pairs, which are then separated by an applied electric field, causing a current pulse.

The tracker layout can be seen in Figure 2.7. In order to keep the sensor occupancy below 1% at design luminosity, the innermost component is a silicon pixel detector. This detector has three layers in the barrel, at radii of 4.7, 7.3 and 10.2 cm, and two in the endcap. Each pixel is 100 $\mu\text{m}$  x 150 $\mu\text{m}$  in size and in total there are 66 million of them. The resulting resolution of the pixel detector is approximately 10 $\mu\text{m}$  in the  $r - \phi$  plane



**Figure 2.7.:** A cross-section of the CMS tracker, indicating the subsystems that comprise it. Each line indicates a detector module [9].

and  $17\mu\text{m}$  in the  $r - z$  plane [20]. During run 1 the proportion of modules in the pixel (strip) tracker known to be defective was 2.4% (2.3%) [21].

Surrounding the pixel detector is a silicon strip detector with 10 layers in the barrel, at radii of 20 to 116 cm, and 12 pairs of disks in the endcap. The strips are typically 10-20 cm long and 80-180 $\mu\text{m}$  wide, with the strip size increasing with radius as the particle flux decreases. The strip detector's single point resolution is 230-530 $\mu\text{m}$  in the  $r - z$  plane and 23-52 $\mu\text{m}$  in the  $r - \phi$  plane. The better resolution in the  $r - \phi$  plane allows a better measurement of  $p_T$ , as this is the direction in which a particle's track bends in the CMS magnetic field. The barrel and endcap detectors together have an acceptance of  $|\eta| < 2.5$  for both the pixel and strip detectors. Further details on the position resolution of the tracking detector for vertex reconstruction will be given in Section 3.2.

### 2.2.2. Electromagnetic calorimeter

The ECAL is designed to provide accurate photon and electron reconstruction and precise measurement of the electromagnetic component of hadron jets. It is a homogeneous calorimeter made of lead tungstate ( $\text{PbWO}_4$ ) crystals, separated into a barrel (ECAL barrel (EB)) section, with 61200 crystals and two endcaps (ECAL endcap (EE)) each with 7234 crystals. These crystals are 25.8 radiation lengths in depth in the barrel and instrumented with photodetectors, avalanche photodiodes being used in the barrel and vacuum phototriodes in the endcap.



**Figure 2.8.:** A schematic of the CMS ECAL, indicating the subsystems that comprise it. The ECAL is 7.8m long by 3.5m wide [9].

The layout of the ECAL is shown in Figure 2.8. The EB crystals have a 170x360 arrangement in  $\eta - \phi$  space such that the gaps between crystals are offset by  $3^\circ$  from the vector to the detector origin, thus avoiding particles travelling through the gaps. The EB extends to  $|\eta| = 1.479$ , with higher values of  $\eta$  covered by the EE. The crystals in the EE are arranged in an  $x - y$  grid pointing at a focus 1.3m from the nominal interaction point, giving a  $2 - 8^\circ$  separation between the gaps between crystals and the vector to the detector origin. In addition to the main PbWO<sub>4</sub> detector the endcaps also have a preshower detector. This preshower is a lead silicon strip sampling calorimeter, which initiates the electromagnetic showers and provides sufficient position resolution to distinguish single photons from pairs produced in neutral pion decays. The total acceptance of the barrel and endcap detectors is  $|\eta| < 3.0$ .

On entering the ECAL high energy electrons or photons initiate an electromagnetic shower by undergoing Bremsstrahlung or pair production respectively. The resulting cascade of particles continues to lose energy by successive Bremsstrahlung and pair production until their energy is low enough that the photons no longer undergo pair-production and the electrons lose their energy mainly by ionisation. The excitation of the PbWO<sub>4</sub> crystals leads to the emission of scintillation light, proportional to the amount of energy deposited, which is collected by the photodetectors.

The choice of PbWO<sub>4</sub> is motivated by its high density (8.28 g/cm<sup>3</sup>), short radiation length (0.89 cm), small Moli  re radius (2.2cm) and radiation hardness this leads to the showers being contained in a small area and allows the calorimeter to be compact and have fine granularity. Another advantage of PbWO<sub>4</sub> is that 80% of the scintillation light is emitted within the LHC's 25ns design bunch crossing time, so particles can be properly associated with the bunch crossing from which they originate.

For particle energies below 500 GeV, where the resulting shower ceases to be contained in the full depth of the ECAL, the ECAL resolution can be parameterised as:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2. \quad (2.4)$$

Where  $S$  is the stochastic term,  $N$  the noise term and  $C$  the constant term. The stochastic term is comprised of fluctuations in the lateral containment of showers and also in the amount of scintillation light. The noise term is made up of electronic and digital noise, and signals from other bunch crossings which do not fully dissipate in time. The constant term comes from non-uniformity of light collection along the crystals, errors in the calibration of crystals against each other and leakage of energy from the back of the calorimeter. The energy resolution was measured without an applied magnetic field in an electron beam using particles with momenta between 20 and 250 GeV. The stochastic, noise and constant terms were found to be 0.028 GeV<sup>1/2</sup>, 0.12 GeV and 0.003 respectively.

As the ECAL is exposed to radiation the PbWO<sub>4</sub> crystals darken and as a result fewer photons are collected per unit energy deposited. The loss of response due to this darkening at the end of Run I varies from 6% for crystals in the most central region of the ECAL to 30% in the endcaps [22].

### 2.2.3. Hadronic calorimeter

The HCAL is designed to measure the energy of strongly interacting particles. This measurement is particularly important for neutral hadrons which do not leave tracks in the tracking system and deposit most of their energy in the HCAL, and for the determination of  $\cancel{E}_T$ . The main part of the HCAL consists of a brass and scintillator plus wavelength shifting fibre sampling calorimeter split into barrel (hadron barrel (HB)) and endcap (hadron endcaps (HE)) sections. The primary design consideration for the HCAL is that it must fit between the outer edge of the ECAL ( $r = 1.77$  m) and the

inner edge of the magnet ( $r = 2.95$  m). In order to satisfy this requirement and achieve satisfactory containment of hadronic showers the magnet coil is also used as an absorber, and there is a further layer of scintillator outside the magnet coil (hadron outer (HO)). The barrel and endcap detectors extend to  $|\eta| < 3$ .

Brass is chosen as the main HCAL absorber because it is not magnetic and has a relatively short nuclear interaction length of 16.42 cm. Once showers have been initiated in the absorber layers they then pass through the plastic scintillator tiles, where they create pulses of light. These pulses are transferred via wavelength shifting fibres to hybrid photodiodes. The segmentation of the scintillator is such that the  $\eta - \phi$  resolution in the HB (HE) is  $0.087 \times 0.087$  (between  $0.087 \times 0.087$  and  $0.17 \times 0.17$  depending on  $\eta$ ).

In addition to the barrel and endcap sections of the HCAL there is also a steel and quartz fibre Cherenkov forward calorimeter (hadron forward (HF)), which extends the calorimetry coverage of CMS to  $|\eta| < 5.2$ . The choice of this technology is driven by its ability to withstand the very high particle fluxes present so close to the beamline. Showers are initiated by the steel absorber and signals are generated in the quartz fibres by particles above the Cherenkov threshold generating Cherenkov light, which is collected by photomultiplier tubes. Due to the Cherenkov energy threshold increasing with particle mass the HF is primarily sensitive to the electromagnetic component of showers.

A diagram of the HCAL layout can be seen in Figure 2.9. In total the HCAL corresponds to 10-15 interaction lengths, depending on  $\eta$ . The resolution of the barrel and endcap sections of the HCAL as a function of the incident particle energy was measured in a pion beam and has been found to be well parameterised by:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{94.3\%}{\sqrt{E}}\right)^2 + (8.4\%)^2 \quad [23]. \quad (2.5)$$

#### 2.2.4. Muon system

As described above muons are highly penetrating, and thus are only rarely contained by the inner detector. Very few other charged particles are able to leave the calorimeters, so the presence of tracks in the muon system is sufficient to identify muons. The muon tracking system uses three types of gaseous particle detectors, located throughout the iron magnet return yoke. In all three types of detector when a charged particle travels through the gaseous detector it ionises the gas, the resulting free electrons then drift



**Figure 2.9.:** A schematic of a quadrant of the CMS HCAL in the  $r - z$  plane, indicating the subsystems that comprise it [9].

towards the detector's anode resulting in an electrical signal. The two primary types of detectors used are the drift tube (DT), which is used in the barrel section of the detector ( $|\eta| < 1.2$ ), and the cathode strip chamber (CSC), which is used in the endcap ( $0.9 < |\eta| < 2.4$ ). The DT and CSC systems identify muons and provide measurements of their momentum. These measurements can be combined with those from the tracker to improve the muon momentum resolution. This combined reconstruction and momentum measurement along with its resolution is described in Section 3.5. Additionally there is a resistive plate chamber (RPC) system in both the barrel and endcap regions ( $|\eta| < 1.6$ ), the primary purpose of which is to provide trigger and bunch crossing identification information. A diagram of the CMS muon system can be found in Figure 2.10.

Each system has its own particular advantages and disadvantages which make it best suited for use in the various parts of the muon system. DTs are inexpensive and reliable, but they are not usable in regions with high muon and neutron background rates, making them well suited to the barrel portion of the detector, where large areas must be instrumented and rates are low. Each DT is a 2.4m long wire in a  $13 \times 42\text{mm}^2$  tube. The length is limited by the segmentation of the iron return yoke, and the cross-section by the requirement that the occupancy and drift time are low enough to prevent multiple muon hits being read out at the same time. The DTs are organised in 4 stations, interspersed with return yoke iron plates. The first three stations have 8 chambers each, 4 to measure

the muon’s position in the  $r - \phi$  plane and 4 to measure the  $z$  co-ordinate. The final outermost layer does not have the  $z$ -measuring chambers. These chambers consist of 8-12 stacked DTs, with each layer offset from the previous one by half the width of a tube to avoid gaps.

Due to their fast response time, fine segmentation and radiation resistance CSCs are ideal for the endcap region where the muon and background rates are higher. Each CSC is a multiwire proportional chamber with 7 planes of cathode strips running radially outwards with 6 planes of anode wires, which run azimuthally, interleaved between them. Both the anode and cathode wires are read out to provide  $\eta$  and  $r - \phi$  co-ordinate measurements respectively. Similarly to the DT system the design number of CSC stations in each endcap is 4 interspersed with iron return yoke plates. During Run I only three of the CSC stations were present, the fourth station in each endcap was added during the long shutdown and is present for Run 2. The position resolution in the  $r - \phi$  plane of the CSCs varies from 75-80 $\mu\text{m}$ .

The RPCs are gas gaps surrounded by anode and cathode plates with read out strips between them. The advantage of RPCs is that their response is good at high rates, and they have very good time resolution, making them ideal for use in the trigger and assignment of muons to a bunch crossing. However, they have much poorer position resolution than the DTs or CSCs. There are 6 layers of RPCs in the barrel and 3 in the endcap.

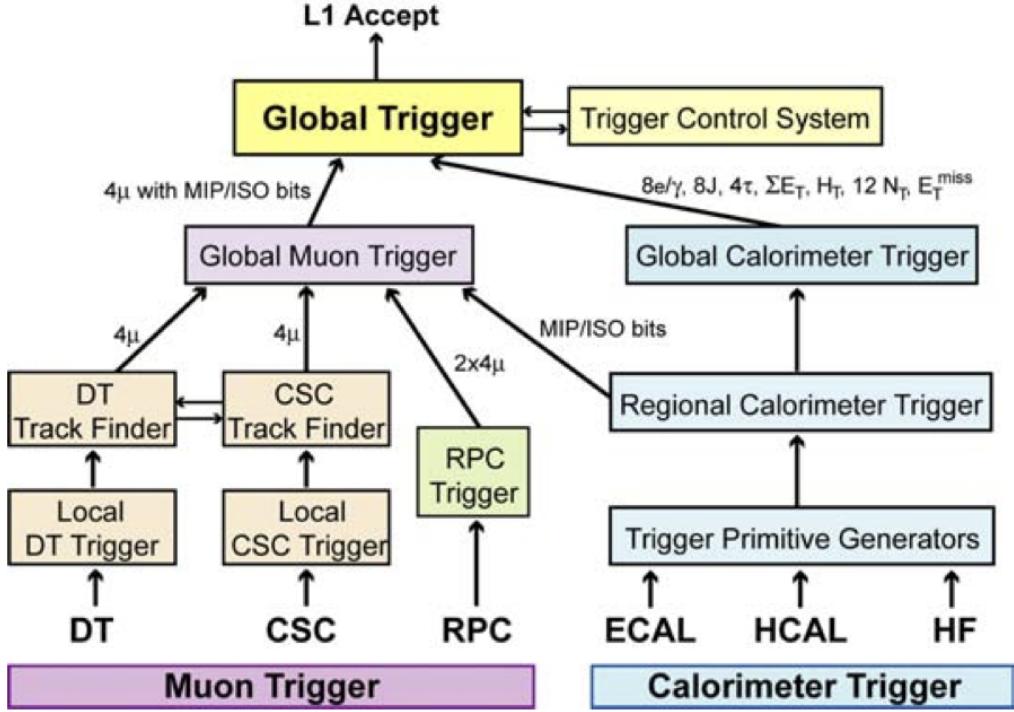
### 2.2.5. Trigger system and data processing

The design bunch crossing rate of the LHC is 40 MHz, and for the data used in this thesis it varied from 20-40 MHz. Since each event consists of approximately 1 MB of data, writing every event to tape would correspond to a data rate of 20-40 TB/s which is not feasible. It is also not feasible for the detector electronics to read out the detector at this frequency. It is therefore necessary to use a trigger system to perform an “online” reconstruction and reduce the event rate by selecting only the most interesting events.

The trigger is separated into two stages, the Level-1 (L1) trigger and the high-level trigger (HLT). First the L1 trigger, which is built of custom-designed electronics, reduces the rate to a maximum of 100 kHz. The decision to accept an event in the L1 trigger or not starts with local information on the energy deposits in the calorimeters and hits in the muon systems, which is stored for all events for 128 bunch crossings. A decision



**Figure 2.10.:** A schematic of a quadrant of the CMS muon system in the  $r - z$  plane, indicating the subsystems that comprise it [24].



**Figure 2.11.:** A schematic of the L1 trigger system. The arrows indicate the flow of data, the information transferred between systems is also indicated [9].

must therefore be made within 128 bunch crossings or the event is discarded. Due to the limited time available and the limited available bandwidth of the data acquisition system, this information is generally not available at the detector's full resolution. After it is collected from the detector the local information is then passed to the regional trigger systems, which generate lists of trigger candidates, such as electrons or jets, ranked by energy and quality. These ranked lists from each region are then passed to the global muon and calorimeter system triggers, which select the highest ranked candidates across the whole detector and give them to the global trigger, which makes a final decision. This process is shown in Figure 2.11.

If an event is accepted by the L1 trigger the full detector information is read out to the HLT farm on the surface, which reduces the rate further still to approximately 1 kHz. The HLT consists of several thousand commercially available CPUs. Despite having the full detector information, the time available does not allow for the full offline reconstruction to be performed. Never the less, the algorithms available at the HLT are much closer to those used offline than those available at L1, reducing the numbers of events lost due to differences in trigger level and offline quantities.

If they are accepted by the HLT, events are sent to be reconstructed using the Worldwide LHC Computing Grid (WLCG). The WLCG consists of several tiers. Data is first fully reconstructed at the Tier 0 centres. During Run I there was only one of these at CERN, for Run II there will also be a Tier 0 centre in Budapest. It is then sent to at least one Tier 1 centre, so that a full copy of the data is available at multiple sites in different geographic locations. Tier 2 and 3 centres then process this data according to the needs of specific analyses.

During 2012 running it was found that it was possible for data to be written to tape from the CMS detector at a higher rate than it could be reconstructed by the Tier 0. 30% of the output of CMS was therefore immediately sent for “prompt” reconstruction, while the remainder was “parked” to tape to be reconstructed during LHC shutdown periods when there is spare computing capacity available [25]. The extra events that could be stored through this parking allowed significantly lower trigger selection thresholds to be used for some of the analyses described in this thesis.

# Chapter 3.

## Physics objects and event reconstruction

The invisible Higgs analysis uses a wide range of objects from the jets and  $\cancel{E}_T$  that are present in the signal process, to charged leptons that are present in background processes. This range of objects means that information from all the CMS subdetectors must be used. The reconstruction of each physics object used from data collected by the CMS detector is described in this chapter, along with the overarching “particle flow” approach to data reconstruction used by CMS.

### 3.1. Tracks

The tracks reconstructed in the inner tracking detector of CMS are a key part of the reconstruction of most other objects used for physics analyses. For example the jet reconstruction algorithm combines information from the tracks and calorimeter energy deposits. The algorithm used by CMS is the Kalman filter based combinatorial track finder (CTF), which is described in Ref. [26].

The CTF starts with seeds generated from either two or three hits in the pixel tracker. Seeds with two hits use the nominal crossing point of the beams to constrain the initial momentum of the track. The layers of the tracker are then iterated through, from inside to outside. The most compatible hit in each layer is added to the track and the track is refitted before moving to the next layer. Once the outside of the detector is reached the algorithm checks for tracks which share more than 19% of their hits and discards the track with the fewest hits. In the case of the two tracks having an equal number of

hits the track with the best fit, i.e. that having the lowest  $\chi^2$ , is kept. This process of reconstructing tracks starting from seeds is repeated up to six times, with hits associated to a successfully reconstructed track removed for the next iteration.

After the full set of iterations is complete the tracks are refitted again using another Kalman filter, initialised with the innermost hit on the track and proceeding to iteratively add the hits on the track from inside to outside. This refitting aims to reduce biases from the track’s seed including those introduced for two hit seeds that include constraints from the beamspot. The refitted tracks are then smoothed by another Kalman filter, which is initialised with the current best fit track hypothesis and iterates from the outside of the detector inwards.

The smoothed tracks then have quality criteria, such as a requirement on the maximum number of layers the track traverses without leaving a hit, imposed to reject fake tracks. The efficiency of the CTF is estimated in data using tracks from muons from Z decays, and is found to be greater than 99% for muons with  $1 < p_T < 100$  GeV.

## 3.2. Primary vertex

The very high instantaneous luminosities present at the LHC lead to a large probability of multiple proton-proton interactions occurring in each bunch crossing. It is therefore essential to identify the Primary Vertex (PV), which relates to the highest energy interaction or “hard scatter”. It is also useful to identify the PV to distinguish “prompt” particles directly from the hard scatter from those resulting from processes which occur later such as heavy flavour hadron decay or photon conversion.

The CMS PV reconstruction algorithm has three steps, track selection, clustering of tracks into vertices and finally fitting the position of these vertices and is described in more detail in Ref. [26]. In the first step, track selection, the subset of tracks with non-significant transverse impact parameters is chosen. This selection removes tracks not coming from the primary interaction region.

The next step of clustering tracks into prototype vertices uses a “deterministic annealing” (DA) algorithm [27]. These prototype vertices then have their position determined by an adaptive vertex fitter [28]. This fitter starts by performing a fit to the position of the vertex, then assigning weights,  $w_i$  to each track according to the probability that it belongs to the vertex, before repeating the process iteratively. Both of these algorithms

also use the concept of “cooling,” where the algorithm is performed repeatedly as a temperature parameter, which controls the size of fluctuations around the current state of the system, is gradually reduced, to increase the chance of finding the global best fit solution.

The number of degrees of freedom of the resulting vertex is defined as:

$$n_{dof} = 2 \sum_{i=1}^{\# \text{tracks}} w_i - 3. \quad (3.1)$$

This variable is highly correlated with the number of tracks compatible with the vertex and can therefore be used to select vertices coming from true proton-proton interactions.

The PV is defined to be the vertex with the highest sum of the squared  $p_T$  of all the tracks contributing to it. If there is no reconstructed vertex the nominal beam crossing point is used. In the analyses described in this thesis events are required to have a real vertex, which has  $n_{dof} > 4$  and a maximum displacement in the  $z$ -direction ( $xy$ -plane) direction from the centre of the detector of 24 cm (2 cm).

The performance of the vertex reconstruction algorithm has been measured using events with at least one jet with  $p_T > 20$  GeV. The efficiency to reconstruct at least one primary vertex in these events is found to be greater than 99% for vertices with at least three tracks. The position resolution is found to vary as a function of the number of tracks associated to the vertex, being approximately 100 $\mu\text{m}$  for vertices with 5 tracks and approaching 10 $\mu\text{m}$  for vertices with greater than 50 tracks.

### 3.3. Particle Flow

Particle flow (PF) is an algorithm used by CMS to combine information from different sub-detectors into individual particles [29–31]. This approach is particularly beneficial for CMS as it allows the accurate momentum measurements of the inner tracker, and the excellent energy measurements and granularity of the ECAL to be combined and used to improve the energy measurement of objects seen in the HCAL. The PF approach also allows calibrations specific to charged and neutral hadrons to be applied. The PF algorithm classifies particles as charged hadrons, neutral hadrons, photons, muons and electrons. This set of particles, referred to as PF candidates, can then further be used

to calculate the  $\cancel{E}_T$ , as input to the jet reconstruction, for reconstructing taus and to calculate the isolation of leptons.

The PF algorithm starts with tracks, reconstructed as described in Section 3.1, and calorimeter clusters, which are reconstructed separately in each sub-detector of the calorimeter system. Clustering starts with seeds, which are the calorimeter cells which have the local maximum energy which is also more than twice the expected calorimeter noise, which is 80 (300) MeV in the EB (EE and 800 MeV in the HCAL. Cells adjacent to the cluster are added if they also have energy more than twice the expected calorimeter noise. Cluster-track pairs whose cluster position and track trajectory are compatible are then linked together to identify charged particles. Linking between tracks from the inner tracker and the muon system is also performed to identify muons. The information from tracks with associated ECAL clusters, i.e. those compatible with electrons, is further used to search for clusters compatible with having come from bremsstrahlung photons from the electron, this is described further in Section 3.4.

Once electrons, muons and charged hadrons have been identified, further calorimeter clusters are identified as neutral hadrons or photons if they are in the HCAL or ECAL respectively. Excess energy in a calorimeter cluster compared to that expected from the associated tracks also allows the presence of neutral particles that would otherwise not have been identified to be determined.

### 3.4. Electrons

As described in Section 3.3, electrons are reconstructed by matching ECAL deposits with tracks from the inner tracker. This process is complicated by the fact that electrons can lose significant amounts of energy, in the form of bremsstrahlung photons, as they traverse the inner tracker. Approximately 35% of electrons lose at least 70% of their initial energy in this way [32]. The bremsstrahlung photons often convert to electron-positron pairs which are then further spread in the  $\phi$  direction by CMS’s solenoidal magnetic field. The electron reconstruction, which is described in detail in Ref. [33], employs so-called “supercluster” algorithms to combine ECAL deposits from both the initial electron and the bremsstrahlung photons.

Due to their different geometries, different supercluster algorithms are used in the barrel and endcaps. In the barrel the “hybrid” clustering algorithm is used, this begins with a seed crystal which is the crystal with local maximum energy greater than 1 GeV. Arrays

of  $5 \times 1$  crystals in  $\eta \times \phi$  are then added around the seed crystal if they are within 17 crystals of it in either direction in  $\phi$  and have energy greater than 0.1 GeV. Contiguous arrays are grouped into clusters. The final supercluster consists of all clusters from a seed with cluster energy greater than 0.35 GeV.

In the endcap the “multi- $5 \times 5$ ” algorithm is used. This algorithm also starts with seed crystals, in this case those with energy higher than their four direct neighbours and also greater than 0.18 GeV. Clusters are then made up of the  $5 \times 5$  square of crystals centered on the seed. Individual clusters whose seeds are within 0.07 in  $\eta$  and 0.3 in  $\phi$  of each other are grouped and kept as a supercluster if their total energy is greater than 1 GeV. A reference position for the supercluster is taken to be the energy-weighted average position of all the clusters belonging to it, and the maximum difference in  $\phi$  between any cluster and the reference position is taken to be the size of the cluster in  $\phi$ . The individual clusters in a supercluster are then extrapolated to the preshower detector. Any preshower deposits within the supercluster’s  $\phi$  size plus 0.15 in  $\phi$  and within 0.15 in  $\eta$  of a cluster in the supercluster are added to it.

The energy-weighted average position and energy of the final supercluster are then used to extrapolate the electron’s track back to the innermost layers of the tracker for both electron charge hypotheses. This extrapolation is then matched to hits within a  $\phi - z$  window of it, whose size is determined by the uncertainties on the  $\phi$  position of the supercluster and the  $z$  position of the beamspot. This size was typically 5 cm in 2012. This matched hit is used to update the estimated electron trajectory so that a hit in the second layer of the inner tracker can be searched for in a much narrower window. Hits in both the first and second layers compatible with a supercluster are then used as seeds for dedicated electron track reconstruction, performed using a Gaussian sum filter (GSF) algorithm [34], which performs better than a Kalman Filter for tracks with significant energy loss.

Electron identification criteria are applied to reject fake electrons caused by other particles such as pions. The variables used include:

- $\Delta\eta_{in}$  and  $\Delta\phi_{in}$ , which are the  $\eta$  and  $\phi$  distances between the electron track position extrapolated to the ECAL and the supercluster position.
- $\sigma_{in\eta}$ , the energy-weighted  $\eta$  width of the cluster.
- $H/E$ , the ratio between the energy deposited in the HCAL and in the ECAL in the region of the electron’s seed cluster.

All of these variables are generally lower for real prompt electrons.

We also require the electrons to be isolated, i.e. have a low amount of other activity present around them in the detector. The variable used for this requirement is the effective area corrected PF isolation,  $I_{PF}$ . In Run 1 (Run 2) it was defined as the sum of the  $p_T$  of the PF candidates within a cone of  $\Delta R < 0.4(0.3)$  around the direction of the electron, minus the expected contribution from PU across the area of the electron.

In the vector boson fusion (VBF) invisible Higgs boson decay searches described later in this thesis two sets of requirements on the above variables are used to identify electrons, both of which require that  $|\eta| < 2.4$ . The “veto” set of identification criteria is looser and is used to veto events containing electrons. The other “tight” set of criteria is stricter and is used when we want to study events containing electrons. Tight electrons are required to be separated by more than 0.3 in  $\Delta R$  from any veto muons to remove fake electrons from muons. The veto (tight) criteria have an efficiency of 93% (85%) for reconstructing central electrons with  $p_T > 50$  GeV [35].

### 3.5. Muons

Due to their relatively high mass and lack of strong force interactions, most muons deposit very little energy in the CMS calorimeters and thus leave the detector after passing through the muon system. As described in Section 3.3, this means that muons can be reconstructed by searching for compatible tracks from the inner tracker and the muon system. The approach of requiring both inner tracker and muon system tracks greatly improves the discrimination between muons and hadronic activity and is referred to as “global” muon reconstruction.

The CMS global muon reconstruction algorithm starts with each track in the muon system and searches for compatible tracks in the inner tracker [36]. If a compatible inner tracker track is found, a track fit, similar to that described in section Section 3.1, is performed using the hits in both the inner tracker and muon system. The fit accounts for energy losses as the muon traverses the detector. It is found that for muons with  $p_T > 200$  GeV the global-muon fit is better than that from the tracker only. However, due to the increased hadron discrimination described above all muons used for analyses in this thesis are required to have both inner tracker and muon system tracks. As with electrons it is also required that muons are isolated. The same isolation variable,  $I_{PF}$ , as described in Section 3.4 is used for muon isolation. Global muon reconstruction

is sufficient for use in vetoing events containing muons, and muons passing the above reconstruction are referred to as “veto” muons.

Where we want to study events containing muons, further identification criteria are used. This is because whilst global muon reconstruction removes most hadrons, some so-called “punch through” hadrons, which are energetic enough to travel all the way through the CMS calorimeters, can still be reconstructed as muons. Furthermore, it is desirable to separate real but non-prompt muons from hadron decay, from prompt muons from the hard scatter or tau decay. The identification consists of requiring a high quality global muon track fit, that the muon’s track passes through at least at least 5 inner tracker layers, with at least one being a pixel layer, that the muon’s track includes at least two hits in the muon system, and that there is at least one muon system track segment present. Muons passing these additional requirements are referred to as “tight” muons. In addition to the above requirements both veto and tight muons are required to have  $|\eta| < 2.1$ .

The efficiency of veto (tight) muon reconstruction has been found to be 98-99% (96-98%) depending on the  $\eta$  of the muon, for muons with  $p_T > 10 \text{ GeV}$ . This efficiency measurement was performed using events with  $J/\psi$  or Z boson decays to muon pairs.

## 3.6. Jets

As it is a hadron collider, quarks and gluons are very common at the LHC. Furthermore, the presence of two final state quarks is one of the primary signatures of VBF Higgs production which is one of the main focuses of this thesis. Ascertaining the momentum of these strongly interacting particles is therefore very important. As discussed in Section 1.2.2, the hadronisation of strongly interacting particles results in highly collimated jets of particles. The momentum of the original parton which gave rise to the jet can be reconstructed by combining all of the particles in the resulting jet.

### 3.6.1. Jet clustering

Jet clustering algorithms take the many different types of particles that are expected to be present in the particle showers from hadronisation, and combine them into jets [37]. It is important that jet clustering algorithms do not produce different reconstructed jets if a jet undergoes soft gluon radiation (called infrared unsafety) or if a gluon in it splits

in two (called colinear unsafety). The algorithm used by CMS is a so-called sequential recombination algorithm. This class of algorithms requires a metric for calculating the distance between particles in the event,  $d_{ij}$ , and a metric for calculating the distance to a nominal beamline particle,  $d_{iB}$  to be defined. The algorithms then proceed as follows:

- 1 Calculate the distance between all pairs of particles in the event including the nominal beamline.
- 2 If the smallest distance is a  $d_{ij}$  combine  $i$  and  $j$  together into a single new particle and return to step 1.
- 3 If the smallest distance is a  $d_{iB}$ , consider  $i$  to be a final state jet and remove it from the list of particles. Return to step 1.
- 4 Stop when no particles remain.

The particular algorithm used by CMS is the infrared and colinear safe anti- $k_T$  algorithm [38], its distances are defined as:

$$d_{ij} = \min(p_{Ti}^{-2}, p_{Tj}^{-2}) \frac{\Delta R_{ij}^2}{R^2}, \quad (3.2)$$

$$d_{iB} = p_{Ti}^{-2}, \quad (3.3)$$

where  $\Delta R_{ij}$  is the distance in the  $\eta - \phi$  plane between particles  $i$  and  $j$  and  $R$  is a parameter of the algorithm analogous to the maximum radius of the jet. This algorithm starts by clustering around the hardest particle in a region and therefore usually produces circular jets, with easy to calculate areas.

The anti- $k_T$  algorithm is implemented using the FASTJET package [39] with the PF candidates, described in Section 3.3, used as input, the output jets are referred to as PF jets. For analyses using data from LHC Run 1 (Run 2)  $R$  of 0.5 (0.4) is used. In addition to these jets reconstructed from PF candidates, in Monte Carlo (MC) events “generator” jets are also reconstructed by applying the anti- $k_T$  algorithm, with the same radius as that used for the PF jets, to the final state particles produced by the generator before they are passed through the detector simulation.

### 3.6.2. Jet identification

In order to reject jets that are badly reconstructed or just due to detector noise, identification criteria are imposed on the jets reconstructed by the above algorithm. These requirements are that:

- The jet contains at least two PF candidates.
- The total jet energy contribution from neutral hadrons must be less than 99%.
- The total jet energy contribution from photons must be less than 99%.
- The jet has contributions from both the ECAL and HCAL.
- Jets with  $\eta$  such that tracking information is available must have at least one charged object which contributes to the jet’s energy and less than 99% of their energy from electrons.

Real jets from quarks or gluons pass these requirements with over 99% efficiency [40].

In addition to jets from detector noise, it is also possible for the jet reconstruction to include particles that are not from the PV, but instead come from PU vertices. This can lead either to an overestimation of the energy of a real jet from the PV, or to fake jets made up of energy from several vertices. The CMS pileup jet identification procedure [41] combines several variables sensitive to the pileup contribution in a jet, such as information on how the  $p_T$  of the jet is shared between its constituents and the constituents’ tracking information, into a boosted decision tree (BDT) [42]. Simulated real jets from quarks pass this identification with 88-99% efficiency depending on how central they are, while jets from pile-up are rejected with 40-87% efficiency [41].

Jets are also required to have  $\eta < 4.7$  so that they are fully contained within the **CMS!** (**CMS!**) detector. Finally, jets which are within 0.5 in the  $\eta - \phi$  plane of any veto electron, defined in Section 3.4, or veto muon, defined in Section 3.5, are vetoed, to avoid using jets which are due to misreconstructed leptons.

### 3.6.3. Jet energy corrections

The energy of the jets clustered and identified by the CMS jet reconstruction often does not match the energy of the particle that initiated the jet. This can have many causes such as additional energy from PU, miscalibration of the energy response of the calorimeters or energy deposited in uninstrumented areas of the detector. To account

for these mismatches a correction to the jet energy is applied that has the following functional form and is described in detail in Ref. [43]:

$$p_\mu^{\text{cor}} = C_{\text{offset}}(p_T^{\text{raw}}) \cdot C_{\text{rel}}(\eta) \cdot C_{\text{abs}}(p'_T) \cdot C_{\text{res}}(p''_T, \eta) \cdot p_\mu^{\text{raw}}. \quad (3.4)$$

Each  $C$  in the equation represents a correction,  $p_\mu^{\text{cor}}$  is the corrected jet four-momentum,  $p_\mu^{\text{raw}}$  is the jet four-momentum before correction,  $p'_T$  is the  $p_T$  after the offset and relative correction,  $C_{\text{offset}}$  and  $p''_T$  is the  $p_T$  after all but the residual correction,  $C_{\text{abs}}$ .

The purpose of  $C_{\text{offset}}$  is to remove energy from the jet which is not due to activity from the PV such as detector noise and PU. The correction is calculated on a jet-by-jet basis by multiplying the median  $p_T$  density of the event in which the jet is by the jet's area.

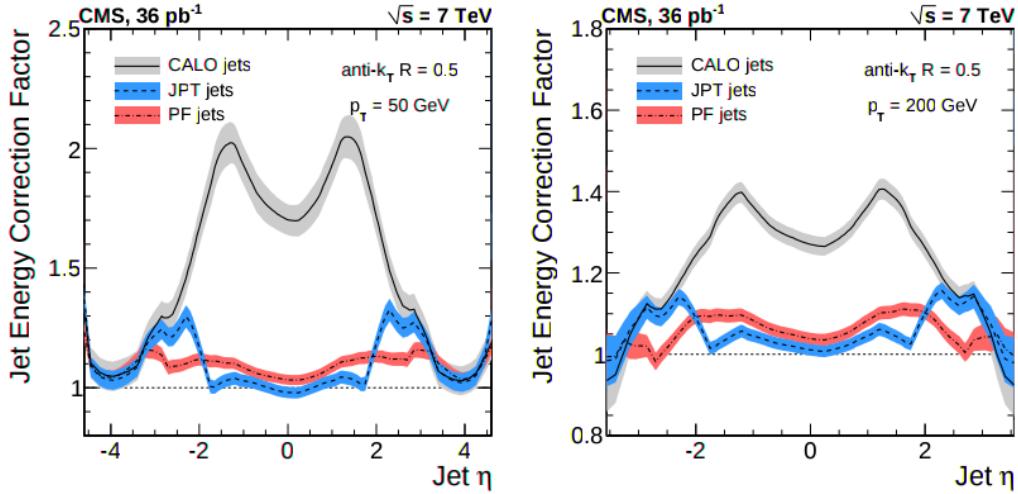
The relative correction,  $C_{\text{rel}}$ , serves to make the jet energy response uniform in  $\eta$ . MC truth information and the dijet  $p_T$  balance method, where the  $p_T$  of a well measured jet in the central region of the detector is compared to a second jet at a different  $\eta$  in events with only two jets, in data are used to calculate  $C_{\text{rel}}$ .

The absolute correction  $C_{\text{abs}}$ , makes the jet energy response uniform in  $p_T$ . As well as being calculated using MC truth information, the correction is also calculated by using  $Z/\gamma + \text{jets}$  events, where the transverse momentum of the jets should balance the  $Z/\gamma$ . Both  $Z$  bosons that decay leptonically and photons have very good energy resolution, so any imbalances can be assumed to be due to jet mismeasurement.

Finally  $C_{\text{res}}$ , which is applied only to data and not MC, corrects for residual differences seen in both  $p_T$  and  $\eta$  response between data and MC. The total uncertainty on the overall jet energy correction is taken to be the sum in quadrature of the uncertainties on the individual corrections. The correction and its uncertainty are shown in Figure 3.1, the other two types of jets (“CALO” and “JPT”) in the figure are not used in analyses described in this thesis and so are not discussed.

### 3.7. Missing transverse energy

Particles which interact only weakly with normal matter such as neutrinos and hypothetical DM particles will pass through the CMS detector without interacting. The only signature that they leave is a momentum imbalance between the visible particles in an event. The high hermeticity of the CMS detector allows this imbalance, the  $\cancel{E}_T$ , first



**Figure 3.1.:** Total jet-energy-correction factor as a function of jet  $\eta$  for jets with  $p_T = 50 \text{ GeV}$  (left) and  $p_T = 200 \text{ GeV}$  (right), for several types of jet reconstruction used at CMS. The bands indicate the corresponding uncertainty [43].

described in Section 1.2.2, to be measured accurately. As the analyses described in this thesis are searches for invisibly decaying Higgs bosons, the measurement of  $\cancel{E}_T$  is crucial.

The CMS  $\cancel{E}_T$  reconstruction algorithm defines the  $\cancel{E}_T$  as the negative vectorial sum of the  $p_T$  of all PF candidates [44]. For processes such as Z boson decays to muon pairs or  $\gamma$ +jets there should be no  $\cancel{E}_T$  as all the decay products are visible. However, as can be seen from Figure 3.2, these events often still appear to have  $\cancel{E}_T$  due to the resolution of the  $p_T$  measurements of the various objects making up the PF candidates, primarily the jets which are numerous and do not have as good resolution as other objects.

The jet energy corrections, described in Section 3.6.3, alter the energy of jets, and in doing so alter the total energy present in the event. These changes are propagated to the  $\cancel{E}_T$ . Furthermore, as charged particle flow candidates can be determined to be from the PV or a PU vertex it is also possible to correct the  $\cancel{E}_T$  for PU contributions. This correction uses the ratio of the energy response of CMS for charged and neutral particles to estimate the neutral PU contribution from the charged PU contribution.

In addition to the above corrections, filters are applied to reject events where detector or beam effects lead to a high probability of spurious  $\cancel{E}_T$ . Examples of the effects which are removed with these filters include particles directly hitting the photodetectors in the ECAL or significant energy deposits from the halo of particles surrounding the LHC beam.



**Figure 3.2.:** Distributions of the uncorrected  $\cancel{E}_T$  in  $Z \rightarrow \mu\mu$  events (left) and  $\gamma + \text{jets}$  events (right) in  $\sqrt{s} = 8$  TeV data and simulation. The shaded band corresponds to the systematic uncertainty [44].

In the VBF invisible Higgs boson decay searches described below events with W or Z boson decays to muons are used to estimate the rate of some background processes. As part of this estimation the muons from the W or Z boson decays are ignored when calculating the  $\cancel{E}_T$ . The variable  $\cancel{E}_T^{\text{no-}\mu}$ , which is the  $\cancel{E}_T$  calculated ignoring all tight muons, is used for this estimation.

### 3.8. Taus

Approximately 35% of taus decay to lighter charged leptons and neutrinos [45]. In this case, due to the short lifetime of the tau, the resulting charged leptons are reconstructed as prompt electrons or muons and the neutrinos cause  $\cancel{E}_T$ , therefore no specific tau reconstruction is necessary. However, the other  $\sim 65\%$  of tau decays are so-called hadronic tau decays, where the decay products are hadrons and a tau neutrino. This section will describe the reconstruction of these tau decays.

CMS uses a hadron plus strips (HPS) algorithm for reconstructing hadronic tau decays, described in detail in Ref. [46]. Almost all hadronic tau decay modes consist of one or three charged hadrons and up to two neutral pions [45]. The HPS algorithm aims to reconstruct both the charged hadrons and the photons which result from the neutral pion decays.

The HPS algorithm is seeded by a PF jet, described in Section 3.6, and starts by creating a strip with the four-momentum of the most energetic electromagnetic (EM) PF candidate, i.e. photon or electron, in the jet. Other EM candidates are then searched for within a window of 0.05 (0.2) in  $\eta$  ( $\phi$ ) of the strip’s centre. The most energetic particle that is found is added to the strip and the four-momentum is updated. This process is repeated until no more particles are found, and if the strip has  $p_T > 1 \text{ GeV}$  at this point it is kept. Combinations of charged hadrons and strips consistent with tau decay modes are then searched for, and if one is found the resulting combination is taken to be a hadronic tau.

Taus are required to be isolated. The isolation is calculated as the sum of all hadronic and photon PF candidates from the PV within a cone of size  $\Delta R = 0.5$  of the tau. PF candidates not compatible with the PV within 0.8 in  $R$  of the tau are used to estimate and correct for the contribution to the isolation from PU.

Electrons which emit bremsstrahlung photons can look very much like one charged hadron plus a neutral pion. A BDT is trained, using similar variables to those used for the electron identification in Section 3.4, to remove these particles. Taus that are consistent with being from a muon are also rejected. This rejection is performed by requiring that the tau is not reconstructed as a track compatible with hits in the muon system. The final efficiency of the CMS hadronic tau reconstruction for taus with  $p_T > 20 \text{ GeV}$  is found to be 55%, with a fake rate of 2 (3)% for non-hadronic tau objects to be reconstructed as hadronic taus in the barrel (endcap) region of the detector.

# Chapter 4.

## Search for invisibly decaying VBF produced Higgs bosons in Run 1 prompt data

As described in Chapter 1, searches for invisible Higgs boson decays are well motivated by their sensitivity to new physics, such as DM. Because  $\mathcal{B}(H \rightarrow \text{inv})$  of an SM 125 GeV Higgs boson is very small, any evidence for invisible Higgs boson decays at the LHC would be evidence for physics beyond the SM. This chapter describes the search for invisible Higgs boson decays using data taken by CMS in 2012 which was promptly reconstructed, a dedicated trigger was developed specifically for this analysis. The total integrated luminosity collected with this trigger that was certified for use in physics analyses was  $19.5\text{fb}^{-1}$ . The analysis was published in Ref. [47].

### 4.1. Event selection

In signal events it is expected that there will be two jets with a characteristic VBF topology and a large amount of  $\cancel{E}_T$ . Several background processes, with significantly higher cross-sections than the signal process, can also produce events with these objects present. It is therefore necessary to design selection criteria, known as “cuts”, to remove as many of these background events from the analysis as possible, whilst retaining the maximum number of signal events.

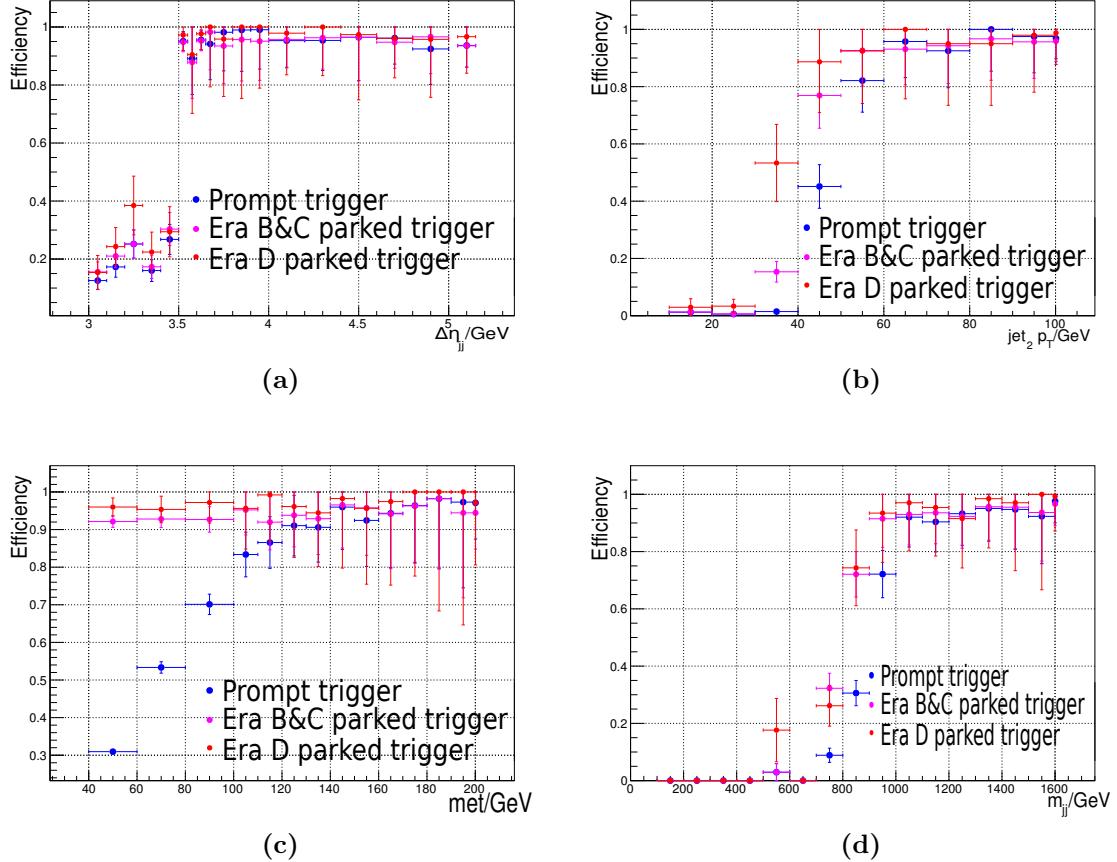
The most significant of these background processes is the production of a vector boson in association with jets, “V+jets”. Leptonic decays of W bosons and Z boson decays

to neutrinos both produce  $\cancel{E}_T$  and, due to the approximately 1000 times higher cross-section for vector boson production than Higgs boson production, there are many events where the associated jets have a VBF-like topology [48]. V+jets backgrounds with a W (Z) are referred to as “W (Z)+jets”. A further background process that can produce significant numbers of VBF-like jets due to its very large cross-section is QCD production of multiple jets (“QCD multijets”). Whilst these multijet events have very little  $\cancel{E}_T$  from real invisible particles, it is possible for significant “fake”  $\cancel{E}_T$  to be caused by mismeasurement of the jets. The production of two vector bosons or top quarks can also lead to two jets and real  $\cancel{E}_T$ , although they have much lower cross-sections than the other background processes and their contribution is not as significant.

#### 4.1.1. Trigger

The trigger requirements can be viewed as the first stage of the event selection. Their primary role is to reduce the rate of events that must be recorded by the detector, whilst retaining the maximum number of signal events. As described in Section 2.2.5 the decision whether to keep an event must be made very rapidly, and, as a result, the object reconstruction algorithms used are less sophisticated, and the granularity of the information available from the CMS subdetectors is worse, than those offline. The trigger criteria have therefore been chosen to be as loose as possible whilst achieving the required rate reduction.

As it is the key variable which indicates the presence of invisible particles all events passing the trigger are required to have significant  $\cancel{E}_T$ . To pass the L1 trigger selection events are required to have  $\cancel{E}_T > 40$  GeV. The HLT selection also requires that events have  $\cancel{E}_T^{\text{no-}\mu} > 65$  GeV. The use of  $\cancel{E}_T^{\text{no-}\mu}$  at trigger level ensures that events that are needed for the control regions used in the background estimation techniques described in Section 4.2 are not rejected. In addition to this  $\cancel{E}_T^{\text{no-}\mu}$  requirement events must have at least one pair of jets in the event which is VBF-like to pass the HLT selection. The VBF-like requirements on the jets consist of requiring their  $\eta$  separation,  $\Delta\eta_{jj}$ , be greater than 3.5, that they are in opposite forward/backward halves of the detector and that they have high invariant mass,  $M_{jj} > 800$  GeV. All of these jet requirements are motivated by the lack of colour connection between VBF events leading to large angular separations between the two jets, as described in Section 1.2.2. Not requiring that the VBF-like pair of jets also be the two highest  $p_T$  jets reduces inefficiencies caused by different  $p_T$  orderings in jets reconstructed by the trigger and by the offline reconstruction.

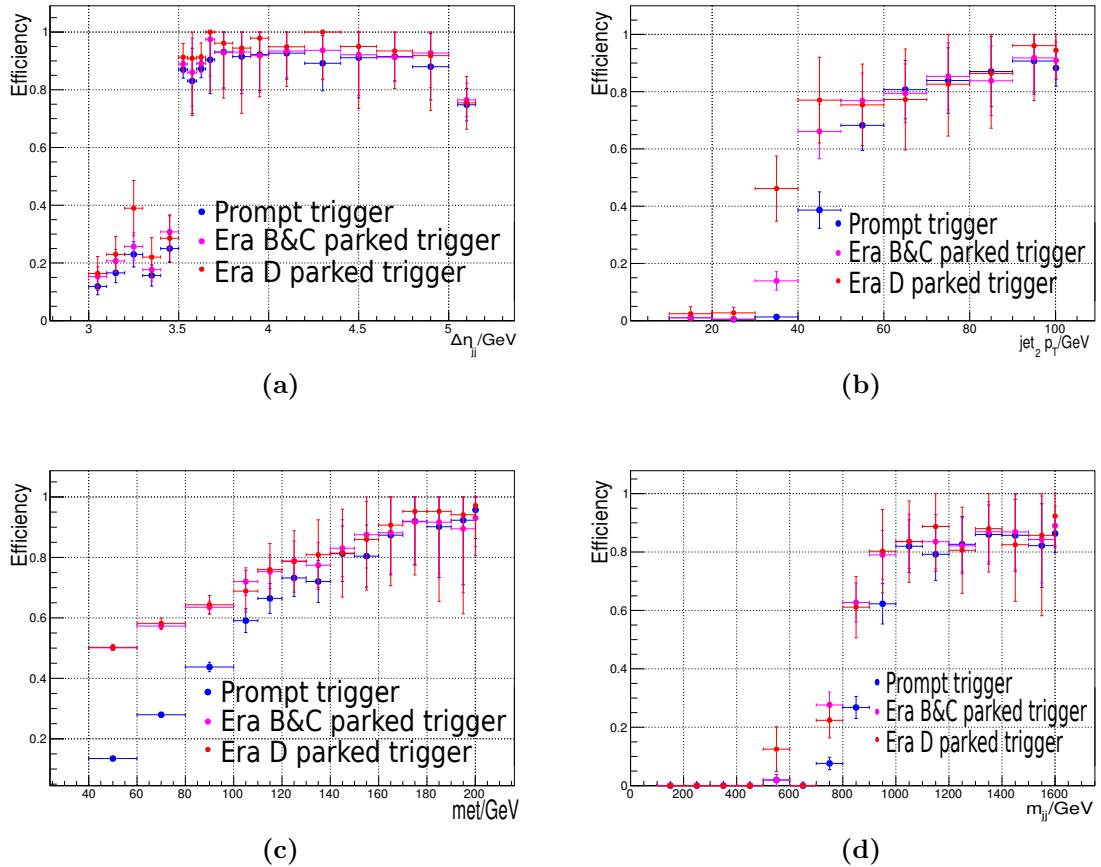


**Figure 4.1.:** The efficiency of the HLT requirements used in the prompt (blue) and parked (purple and red) data analyses as a function of the values of several offline variables, measured in a sample of events recorded on a single-muon trigger. (a) Efficiency as a function of offline  $\Delta\eta_{jj}$ , (b) efficiency as a function of sub-leading jet  $p_T$ , (c) efficiency as a function of offline  $\cancel{E}_T^{\text{no-}\mu}$ , (d) efficiency as a function of  $M_{jj}$ .

The efficiency for events to pass both the prompt analysis trigger and the two triggers used for the parked data analysis, described in Chapter 7 as a function of their values of several offline variables, measured in single muon data collected using an uncorrelated trigger, is shown in Figs. 4.1 and 4.2. The events used in the trigger efficiency measurement are required to pass the following cuts:

$$M_{jj} > 1100 \text{ GeV}, \cancel{E}_T^{\text{no-}\mu} > 130 \text{ GeV}, \text{leading 2 jets' } p_T > 50 \text{ GeV}, \Delta\eta_{jj} > 4.2, \eta_{j1} \cdot \eta_{j2} < 0. \quad (4.1)$$

In each measurement the cut on the variable being studied is removed. The measured trigger efficiency is applied as an event-by-event weight to all MC samples.



**Figure 4.2.:** The combined efficiency of the HLT and L1 trigger requirements used in the prompt (blue) and parked (purple and red) data analyses as a function of the values of several offline variables, measured in a sample of events recorded on a single-muon trigger. (a) Efficiency as a function of offline  $\Delta\eta_{jj}$ , (b) efficiency as a function of sub-leading jet  $p_T$ , (c) efficiency as a function of offline  $\cancel{E}_T^{\text{no-}\mu}$ , (d) efficiency as a function of  $M_{jj}$ .

### 4.1.2. Offline selection

The offline selection is chosen for three reasons. Firstly, the reconstruction algorithms for some objects are only well validated for certain values of  $p_T$  and  $\eta$ . This consideration decides the  $p_T$  thresholds for jets and leptons to be used. Secondly, as can be seen from Figure 4.1, the values of the offline variables where the trigger becomes fully efficient are in some cases much higher than the online cut. Because the variables used in the trigger are highly correlated, the offline cuts on all variables used in the trigger were chosen such that the trigger efficiency for the variable at that point is greater than 95%. Finally, some of the cuts imposed aim to reduce the contribution from background processes, which improves the signal to background ratio in the resulting region, and thus the expected limit on  $\mathcal{B}(H \rightarrow \text{inv})$ .

The specific set of offline selection cuts chosen begins by requiring that events have no veto muons or electrons, as defined in sections 3.4 and 3.5, with  $p_T > 10$  GeV. Signal events are not expected to contain leptons, while background events are, so this lepton veto reduces the background from W and Z boson decays and also from top quarks without removing signal events. The two highest  $p_T$  jets in the event are then identified as the VBF tag pair, and tighter versions of the trigger selection, motivated by the trigger efficiency considerations described above, are then applied. Specifically, the tag jets are required to be in opposite forward/backward halves of the detector, to both have  $p_T > 50$  GeV to have  $M_{jj} > 1100$  GeV and  $\Delta\eta_{jj} > 4.2$ . Again due to trigger efficiency considerations,  $\cancel{E}_T^{\text{no-}\mu}$  is required to be greater than 130 GeV. Because events with veto muons have been removed by the lepton veto,  $\cancel{E}_T^{\text{no-}\mu}$  in this region is identical to  $\cancel{E}_T$ . However, it is important for background estimation methods that  $\cancel{E}_T^{\text{no-}\mu}$  and not  $\cancel{E}_T$  is used.

As well as the trigger based selection, further cuts are made to reduce the QCD multijet background to a level much lower than the V+jets backgrounds. The two tag jets are required to have an azimuthal separation,  $\Delta\phi_{jj} < 1.0$ , since multijet events with  $\cancel{E}_T$  due to mismeasurement are most likely to have their jets back-to-back in the detector, i.e. with  $\Delta\phi_{jj} = \pi$ . Events where there are any jets with  $p_T > 30$  GeV between the two tag jets in  $\eta$  are also vetoed. This central jet veto (CJV) is motivated by the lack of colour connection, described in Section 1.2.2, between the quarks in VBF production that makes the presence of such jets unlikely in genuine signal events. The region of phase space remaining after all these cuts have been applied is called the signal region.

Finally, the values of the cuts are optimised to provide the best expected limit on  $\mathcal{B}(H \rightarrow \text{inv})$  for a 125 GeV Higgs boson, which is calculated using the method described in Section 1.6 using the same background estimation and systematic uncertainties as the final analysis as described in Sections 4.2 and 4.3 respectively. For the tag jet  $p_T$  and  $\cancel{E}_T^{\text{no-}\mu}$ , no improvement in the expected limit is seen by tightening the cut, so the requirement is at the 95% efficiency point of the trigger. The distributions and cut values for several of the other variables used are shown in Figure 4.3. The full selection gives an efficiency of  $(6.8 \pm 0.3) \times 10^{-3}$  for selecting events from invisible decays of a VBF produced 125 GeV Higgs boson.

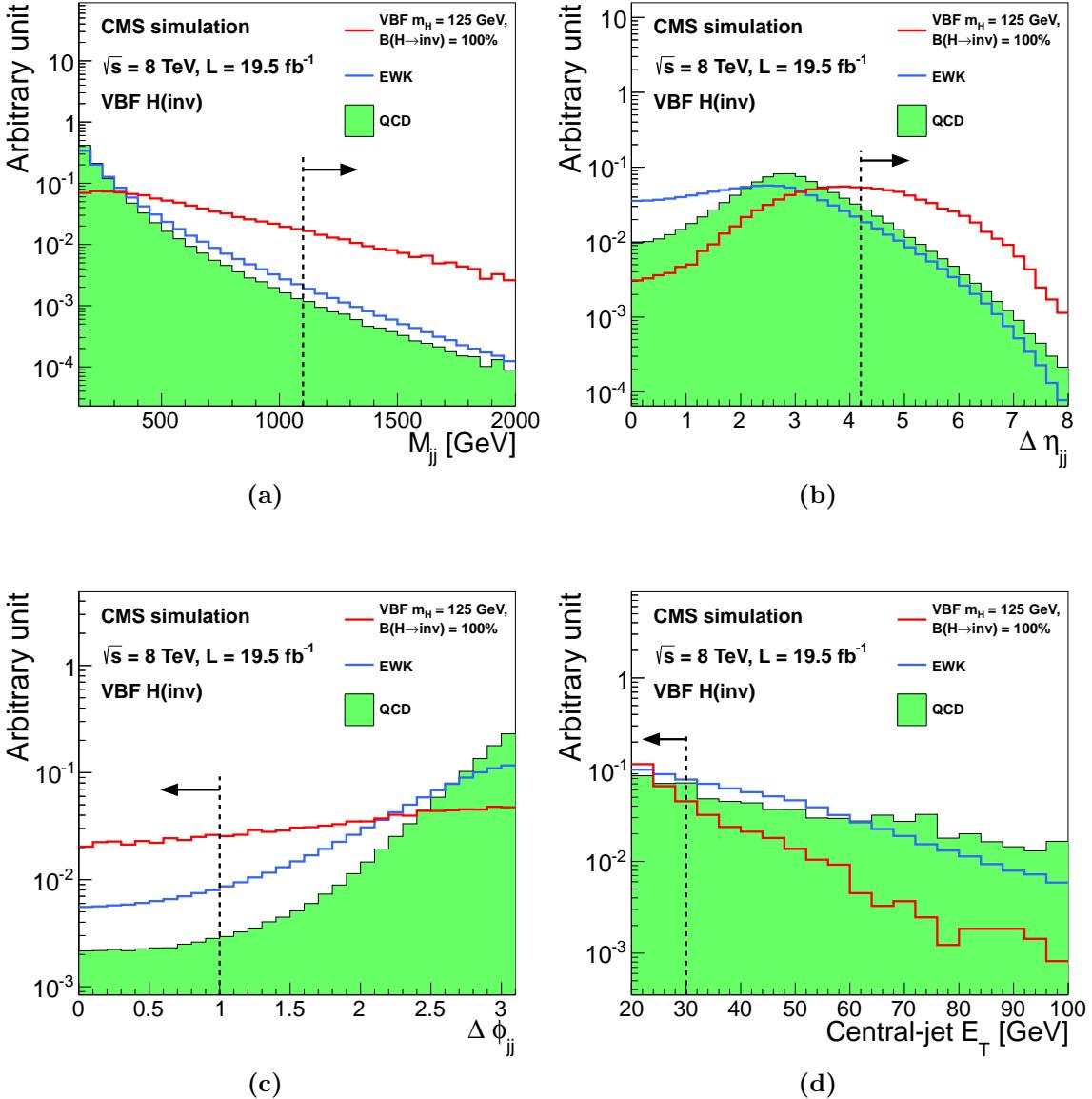
## 4.2. Background estimation

As discussed in Section 4.1 there are several background processes which are capable of producing VBF-like jets in association with  $\cancel{E}_T$ . The event selection removes most of these events, however a significant number still remain and it is important to estimate this number precisely. Data-driven methods, where data ‘‘control regions’’, which are similar to the signal region, are used to estimate the most significant backgrounds. This data-driven approach is particularly important due to the very stringent kinematic requirements placed on the tag jets, which lead to large uncertainties on estimates taken from MC alone. The particular method used to estimate each of the backgrounds will be described in this section.

### 4.2.1. $W \rightarrow e\nu + \text{jets}$

The  $W + \text{jets}$  background where the  $W$  boson decays to an electron and an electron neutrino,  $W \rightarrow e\nu$ , is estimated using single electron events. All aspects of the event selection are the same as those used in the signal region, except for the electron veto, which is replaced with the requirement that there is exactly one tight electron, with  $p_T > 20$  GeV, in the event and no other veto electrons. These requirements give a single electron control region composed of events with jets that have the same kinematics as those in the signal region, but which is dominated by  $W \rightarrow e\nu$  events.

The number of  $W \rightarrow e\nu$  events in the signal region is then estimated by using the ratio between the expected number of events in the signal and control regions from MC to extrapolate from the number of events seen in data in the single electron control region



**Figure 4.3.:** Distributions of (a)  $M_{jj}$ , (b)  $\Delta\eta_{jj}$ , (c)  $\Delta\phi_{jj}$  and (d) leading central jet  $p_T$  in background and signal MC events. The events shown are required to have two jets in opposite forward/backward halves of the detector with  $p_T > 50 \text{ GeV}$ ,  $M_{jj} > 150 \text{ GeV}$  and  $\cancel{E}_T^{\text{no-}\mu} > 130 \text{ GeV}$ . EWK refers to the V+jets and minor backgrounds described in Section 4.2 and QCD refers to the QCD multijet background. The dashed lines indicate the offline selection criteria applied to these variables, which are motivated in the text [47].

**Table 4.1.:** The inputs to, and results of, the  $W \rightarrow e\nu$  background estimation.  $N_{W \rightarrow e\nu}$  is, for the signal region the number of events expected from  $W \rightarrow e\nu$  backgrounds, and for the control region the number of events remaining in the region after the subtraction of other backgrounds.

	Signal region	Control region
$N_{Data}$	N/A	64
$N_{Bkg}$	N/A	$7.42 \pm 2.78$ (MC stat.)
$N_{MC}$	$105 \pm 10$ (MC stat.)	$86.6 \pm 7.1$ (MC stat.)
$N_{W \rightarrow e\nu}$	<b><math>68.7 \pm 10.3</math> (stat) <math>\pm 8.8</math>(MC stat.)</b>	$56.6 \pm 8.5$ (stat)

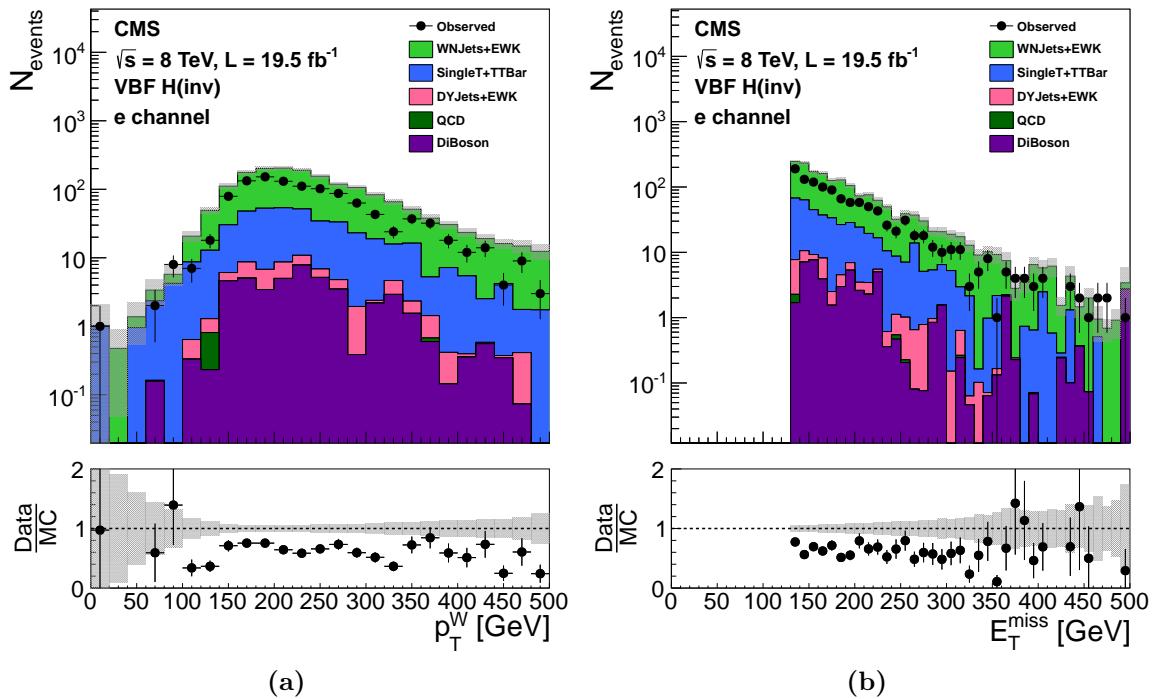
using the following formula:

$$N_{Exp}^S = (N_{Data}^C - N_{Bkg}^C) \cdot \frac{N_{MC}^S}{N_{MC}^C}, \quad (4.2)$$

where  $N_{Exp}^S$  is the number of expected events in the signal region from this background process,  $N_{Data}^C$  is the number of events seen in the control region in data,  $N_{Bkg}^C$  is the number of events from other backgrounds in the control region estimated using MC, which is expected to be small, and  $N_{MC}^S$  and  $N_{MC}^C$  are the numbers of events predicted by MC to be in the signal and control regions respectively. The fact that estimations from MC are only used in ratios, or where they are expected to be small, significantly reduces the dependence of the final background estimation on the overall rate of the process predicted by MC and instead allows the observed rate in data to be used. It is important that the shape of the variables which differ between the control and signal regions are well modelled by the MC. The modelling of the shape of two key variables, the  $E_T^{\text{no-}\mu}$  and the electron  $p_T$ , are shown in Figure 4.4. It can be seen that whilst the overall rate is significantly different between data and MC, the shape of the distribution is modelled well. The inputs to, and results of, the background estimation are shown in Table 4.1.

#### 4.2.2. $W \rightarrow \mu\nu + \text{jets}$

The method used to estimate the background from  $W + \text{jets}$  where the  $W$  boson decays to a muon and a muon neutrino,  $W \rightarrow \mu\nu$ , is very similar to that used for  $W \rightarrow e\nu$ . A single muon control region is used which replaces the muon veto of the signal region with a requirement that there is exactly one tight muon, with  $p_T > 20$  GeV, and no other



**Figure 4.4.:** Distributions of the visible  $W$  boson  $p_T$  (i.e. the electron  $p_T$ ) (a) and  $E_T^{\text{miss}}$  (b) in the single electron control region. WNJets+EWK indicates the  $W+jets$  contribution to this region, SingleT+TTBar indicates the contribution from top quark related processes, DYJets+EWK indicates the  $Z+jets$  contribution, QCD indicates the QCD multijet contribution and DiBoson indicates the two vector boson contribution. All of these contributions are estimated from MC. The hatched region illustrates the systematic uncertainty [49].

**Table 4.2.:** The inputs to, and results of, the  $W \rightarrow \mu\nu$  background estimation.  $N_{W \rightarrow \mu\nu}$  is, for the signal region the number of events expected from  $W \rightarrow \mu\nu$  backgrounds, and for the control region the number of events remaining in the region after the subtraction of other backgrounds.

	Signal region	Control region
$N_{Data}$	N/A	216
$N_{Bkg}$	N/A	$30.1 \pm 4.5$ (MC stat.)
$N_{MC}$	$108 \pm 10$ (MC stat.)	$306 \pm 15$ (MC stat.)
$N_{W \rightarrow \mu\nu}$	<b><math>65.8 \pm 5.4</math>(stat) <math>\pm 6.7</math>(MC stat.)</b>	$186 \pm 15$ (stat)

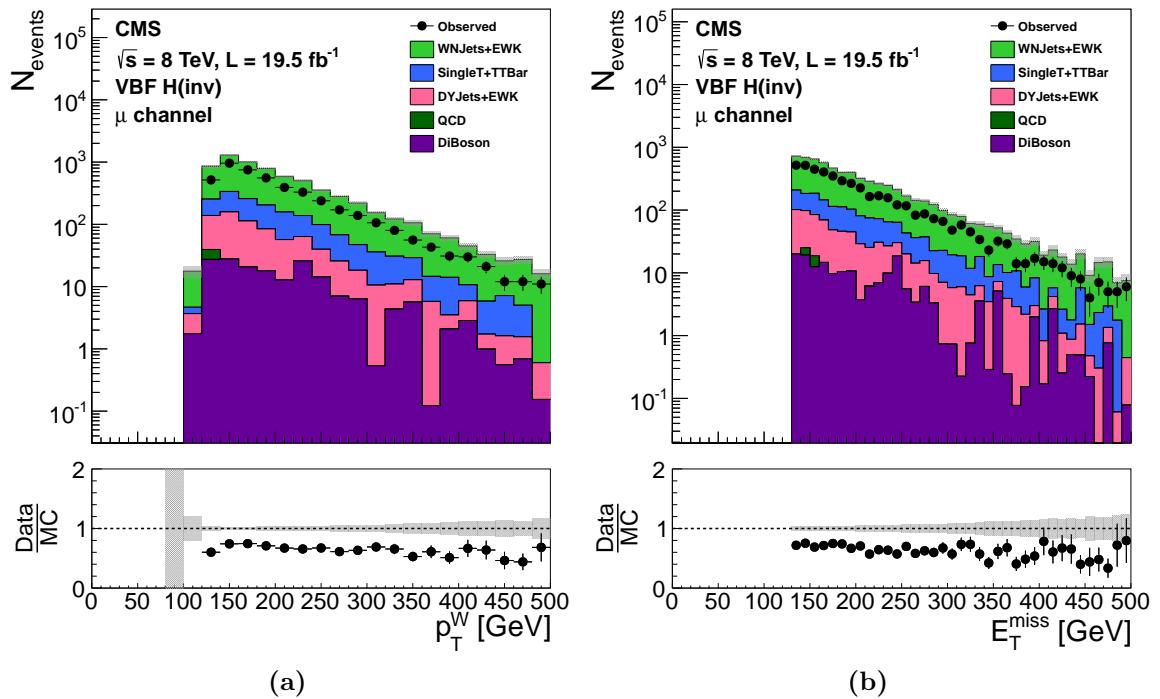
veto muons. All other signal region cuts remain unchanged. Equation 4.2 is then used, with the control region now being the single muon control region, to estimate the number of events from  $W \rightarrow \mu\nu$  expected in the signal region. The inputs to, and results of, the background estimation are shown in Table 4.2, and distributions of the muon  $p_T$  and the  $\cancel{E}_T^{\text{no-}\mu}$  in the single muon control region are shown in Figure 4.5.

### 4.2.3. $W \rightarrow \tau\nu + \text{jets}$

The background from  $W+\text{jets}$  where the  $W$  boson decays to a tau and a tau neutrino,  $W \rightarrow \tau\nu$ , is estimated using a single tau control region data-driven method. However, in this case the control region used has more differences from the signal region than those used above. The reason for these increased differences is that the reconstruction efficiency for tau leptons is significantly lower than that for electrons or muons, and they are also more likely to be misreconstructed as jets, causing the event to be vetoed by the CJV. There are therefore only  $3.76 \pm 1.27$ (stat)  $W \rightarrow \tau\nu$  events with identified taus with  $p_T > 20$  GeV expected in the signal region from MC.

To increase the number of events in the single tau control region, the CJV has been removed. The resulting control region has  $29.2 \pm 3.61$ (stat)  $W + \text{jets}$  events expected and thus a much lower statistical uncertainty. As there is no veto of tau leptons in the signal region the tau control region and the signal region are not mutually exclusive. However, as stated above the number of events in the signal region with identified taus is expected to be small, so the overlap is considered negligible.

In addition to the tau identification algorithm described in Section 3.8, alternative algorithms were studied to check for better performance in terms of identification



**Figure 4.5.:** Distributions of the visible W boson  $p_T$  (i.e. the muon  $p_T$ ) (a) and  $E_T^{\text{miss}}$  (b) in the single muon control region. WNJets+EWK indicates the W+jets contribution to this region, SingleT+TTBar indicates the contribution from top quark related processes, DYJets+EWK indicates the Z+jets contribution, QCD indicates the QCD multijet contribution and DiBoson indicates the two vector boson contribution. All of these contributions are estimated from MC. The hatched region illustrates the systematic uncertainty [49].

**Table 4.3.:** The inputs to, and results of, the  $W \rightarrow \tau\nu$  background estimation.  $N_{W \rightarrow \tau\nu}$  is, for the signal region the number of events expected from  $W \rightarrow \tau\nu$  backgrounds, and in the control region the number of events remaining in the region after the subtraction of other backgrounds.

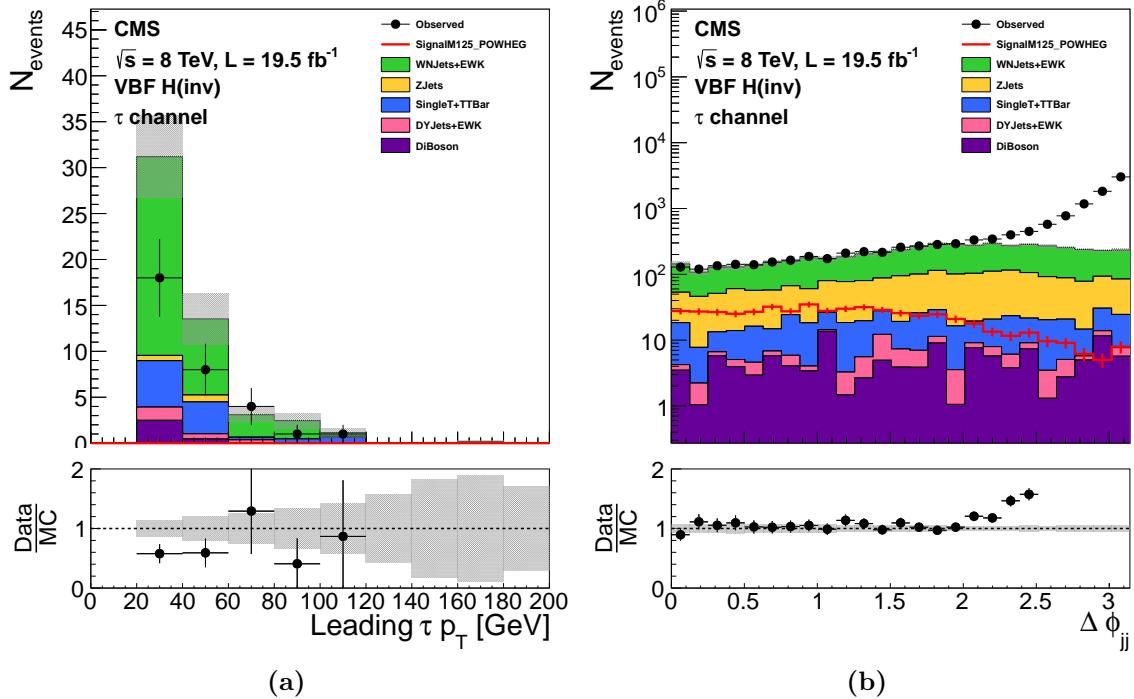
	Signal region	Control region
$N_{Data}$	N/A	32
$N_{Bkg}$	N/A	$14.7 \pm 3.4$ (MC stat.)
$N_{MC}$	$95.6 \pm 8.5$ (MC stat.)	$29.2 \pm 3.6$ (MC stat.)
$N_{W \rightarrow \tau\nu}$	<b><math>56.5 \pm 21.5</math>(stat) <math>\pm 8.6</math>(MC stat.)</b>	$17.3 \pm 3.9$ (stat)

efficiency and fake rate. Specifically an alternative isolation algorithm was investigated which used a multi-variate analysis (MVA) approach to estimate the isolation sum, as well as different working points for the anti-electron and anti-muon discriminators [46]. The tau identification efficiency was found to be higher for both the alternative isolation algorithm and different working points for the anti-lepton discriminators, being twice as large if both were used compared to the standard tau identification. However, the rate of  $W \rightarrow e\nu$  events being identified as  $W \rightarrow \tau\nu$  was also significantly increased, going from 2% for the standard identification to 15% when the alternative isolation and anti-lepton discriminators were used. It was therefore decided to use the tau identification described in Section 3.8.

The final estimation of the background from  $W \rightarrow \tau\nu$  is carried out using Equation 4.2, with the single tau control region with no CJV being used as the control region. The inputs to, and results of, the background estimation are shown in Table 4.3. Distributions of the tau  $p_T$  and  $\Delta\phi_{jj}$  in the single tau control region are shown in Figure 4.6, it can be seen that the shape of the two distributions in data and MC agree well with the exception of the high  $\Delta\phi_{jj}$  region which is not part of either the signal or tau control regions.

#### 4.2.4. $Z \rightarrow \nu\nu + \text{jets}$

The background from  $Z + \text{jets}$  where the  $Z$  decays to neutrinos,  $Z \rightarrow \nu\nu$ , is different from the  $W + \text{jets}$  backgrounds described above, in that nothing is required to be misidentified in order for these events to contribute to the signal region. The method used to estimate the  $Z \rightarrow \nu\nu$  background therefore differs slightly from that used above. The method



**Figure 4.6.:** Distributions of the the tau  $p_T$  (a) and  $\Delta\phi_{jj}$  (b) in the single tau control region. WNJets+EWK indicates the W+jets contribution to this region, ZJets indicates the contribution from Z+jets, SingleT+TTBar indicates the contribution from top quark related processes, QCD indicates the QCD multijet contribution and DiBoson indicates the two vector boson contribution. All of these contributions are estimated from MC. The hatched region illustrates the systematic uncertainty [49].

uses a dimuon control region which is populated by events from the process  $Z/\gamma^* \rightarrow \mu\mu$ . As this process can be mediated by a photon, the kinematics of the jets in  $Z/\gamma^* \rightarrow \mu\mu$  events can be different to those from  $Z \rightarrow \nu\nu$ . The dimuon control region that is defined therefore has a requirement that the invariant mass of the dimuons be between 60 and 120 GeV. The control region is otherwise identical to the signal region, except that the muon veto is replaced with a requirement that there are exactly two tight muons with  $p_T > 20$  GeV and no other veto muons.

As well as the possibility of different kinematics,  $Z/\gamma^* \rightarrow \mu\mu$  and  $Z \rightarrow \nu\nu$  also have different cross-sections. The formula used to estimate the  $Z \rightarrow \nu\nu$  background takes this into account as follows:

$$N_{Exp}^S = N_{Data}^C - N_{Bkg}^C \cdot \frac{\sigma(Z \rightarrow \nu\nu)}{\sigma(Z/\gamma^* \rightarrow \mu\mu)} \cdot \frac{\epsilon_{VBF}^S}{\epsilon_{VBF}^C}, \quad (4.3)$$

where  $\sigma(Z \rightarrow \nu\nu)$  is the cross-section for  $Z \rightarrow \nu\nu$  and  $\sigma(Z/\gamma^* \rightarrow \mu\mu)$  is the cross-section for  $Z/\gamma^* \rightarrow \mu\mu$ .  $\epsilon_{VBF}^S$  and  $\epsilon_{VBF}^C$  are the efficiencies for  $Z \rightarrow \nu\nu$  events to pass the signal region selection and  $Z/\gamma^* \rightarrow \mu\mu$  events to pass the control region selection respectively. As  $Z$  bosons can be created via either QCD or electroweak processes, which both have different cross-sections and efficiencies,  $\epsilon_{VBF}^S$  and  $\epsilon_{VBF}^C$ , are a cross-section weighted average of the efficiency for both types of production, calculated as:

$$\epsilon_{VBF}^S = \frac{\sigma(Z \rightarrow \nu\nu, EWK) \frac{N_{MC}^S(EWK)}{N_{gen}(Zmass,EWK)} + \sigma(Z \rightarrow \nu\nu, QCD) \frac{N_{MC}^S(QCD)}{N_{gen}(Zmass,QCD)}}{\sigma(Z \rightarrow \nu\nu, EWK) + \sigma(Z \rightarrow \nu\nu, QCD)}, \quad (4.4)$$

$$\epsilon_{VBF}^C = \frac{\sigma(Z/\gamma^* \rightarrow \mu\mu, EWK) \frac{N_{MC}^C(EWK)}{N_{gen}(EWK)} + \sigma(Z/\gamma^* \rightarrow \mu\mu, QCD) \frac{N_{MC}^C(QCD)}{N_{gen}(QCD)}}{\sigma(Z/\gamma^* \rightarrow \mu\mu, EWK) + \sigma(Z/\gamma^* \rightarrow \mu\mu, QCD)}, \quad (4.5)$$

where  $EWK$  and  $QCD$  denote where cross-sections or numbers of events are for electroweak or QCD production of a  $Z$  boson.  $N_{gen}$  is the number of events in the  $Z+jets$  MC sample at generator level. Due to the limited size of the available  $Z \rightarrow \nu\nu$  MC samples, the same  $Z/\gamma^* \rightarrow \mu\mu$  samples used for the MC estimate of the number of events in the control region are used to obtain an estimate from MC of the number of events from the  $Z \rightarrow \nu\nu$  process in the signal region. For this estimate the leptons in the  $Z/\gamma^* \rightarrow \mu\mu$  samples are ignored, the production cross-section is scaled to the appropriate  $Z \rightarrow \nu\nu$  value and it is required that there is a generator level dimuon in the event with invariant mass between 80 and 100 GeV. The generated dimuon mass for this sample was required to be greater than 50 GeV, so the cross-sections used in Equations 4.3, 4.4 and 4.5 are also calculated with this constraint. For the control region  $N_{gen}$  is calculated

**Table 4.4.:** The input variables for the calculation of  $\epsilon_{VBF}^S$  and  $\epsilon_{VBF}^C$  using Equations 4.4 and 4.5 respectively.

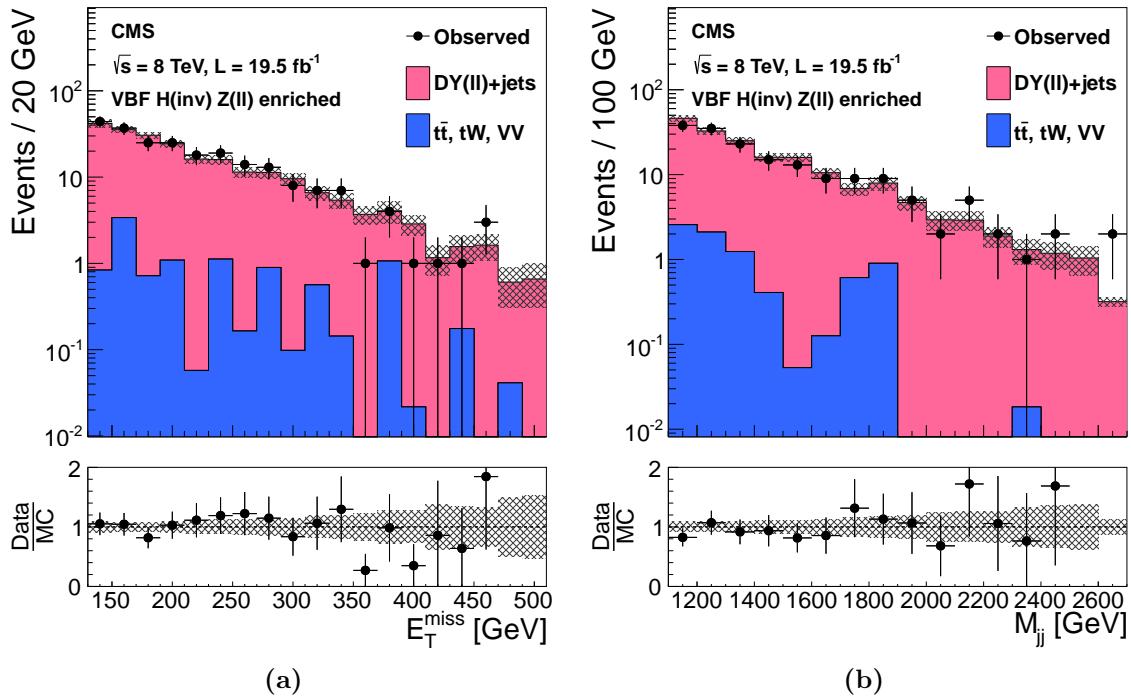
Variable	Value
$\sigma(Z \rightarrow \nu\nu, EWK)$	1.380 pb
$\sigma(Z \rightarrow \nu\nu, QCD)$	6600 pb
$\sigma(Z/\gamma^* \rightarrow \mu\mu, EWK)$	0.303 pb
$\sigma(Z/\gamma^* \rightarrow \mu\mu, QCD)$	1168 pb
$\frac{N_{MC}^S(EWK)}{N_{gen}(Zmass,EWK)}$	$(1.3 \pm 0.1) \cdot 10^{-3}$
$\frac{N_{MC}^S(QCD)}{N_{gen}(Zmass,QCD)}$	$(1.4 \pm 0.2) \cdot 10^{-6}$
$\frac{N_{MC}^C(EWK)}{N_{gen}^S(EWK)}$	$(7.5 \pm 0.3) \cdot 10^{-4}$
$\frac{N_{MC}^C(QCD)}{N_{gen}^S(QCD)}$	$(9.2 \pm 1.2) \cdot 10^{-7}$

after requiring that the mass of the generator level dimuon is between 60 and 120 GeV, denoted by the label Zmass in Equations 4.4 and 4.5.

The inputs to equations 4.4 and 4.5 are given in Table 4.4. The cross-sections for electroweak Z boson production were calculated at NLO using VBFNLO which specialises in vector boson production. The cross-section for QCD production of  $Z/\gamma^* \rightarrow \mu\mu$  is calculated using FEWZ inclusively for all leptons and then divided by three to obtain the figure for muons only. This cross-section is then multiplied by the ratio between the cross-section for  $Z \rightarrow \nu\nu$  and  $Z/\gamma^* \rightarrow \mu\mu$ , which was calculated to be 5.651 at NLO using MCFM, to obtain the QCD production cross-section for  $Z \rightarrow \nu\nu$ . The inputs to Equation 4.3 are given in Table 4.5, with the exception of the ratio between the total production cross-sections for  $Z \rightarrow \nu\nu$  and  $Z/\gamma^* \rightarrow \mu\mu$  which is taken to be the same 5.651 that it is found to be for QCD production. This approximation is used because the electroweak contribution to the ratio is smaller than that from QCD by more than a factor of one thousand and is therefore negligible. The distributions of  $E_T^{\text{no-}\mu}$  and  $M_{jj}$  for a Z control region with relaxed selection, to ensure sufficient numbers of events, are shown in Figure 4.7, demonstrating that the MC samples model the data distribution well.

**Table 4.5.:** The inputs to, and results of, the  $Z \rightarrow \nu\nu$  background estimation using Equation 4.3.  $\epsilon_{VBF}$  in the signal (control) region is calculated using Equation 4.4 (4.5).  $N_{Z \rightarrow \nu\nu}/N_{Z/\gamma^* \rightarrow \mu\mu}$  is in the signal region the number of events expected from  $Z \rightarrow \nu\nu$  backgrounds, and for the control region the number of events remaining in the region after the subtraction of other backgrounds. The systematic uncertainties are calculated as described in Section 4.3.

	Signal region	Control region
$N_{Data}$	N/A	12
$N_{Bkg}$	N/A	$0.3 \pm 0.1$ (MC stat.)
$\epsilon_{VBF}$	$(1.65 \pm 0.15 \text{ (stat)} \pm 0.22 \text{ (syst)}) \cdot 10^{-6}$	$(1.11 \pm 0.12 \text{ (stat)} \pm 0.12 \text{ (syst)}) \cdot 10^{-6}$
$N_{Z \rightarrow \nu\nu}/N_{Z/\gamma^* \rightarrow \mu\mu}$	<b><math>99 \pm 29 \text{ (stat)} \pm 25 \text{ (syst)}</math></b>	$11.7 \pm 0.1$ (MC stat.)



**Figure 4.7.:** Distributions of the  $E_T^{\text{no-}\mu}$  (a) and  $M_{jj}$  (b) in a relaxed  $Z$  control region, with no requirement on  $\Delta\phi_{jj}$ , the CJV removed, and the requirements on  $M_{jj}$  and  $\Delta\eta_{jj}$  relaxed to 1000 GeV and 3.5 respectively. DY(II)+jets indicates the contribution from  $Z+jets$  processes and  $t\bar{t}$ ,  $tW$ ,  $VV$  indicates the contribution from minor backgrounds. The hatched region illustrates the systematic uncertainty [47].

	Fail $\not{E}_T^{\text{no-}\mu}$		Pass $\not{E}_T^{\text{no-}\mu}$
Fail CJV	A	B	
	C	D	

**Figure 4.8.:** A diagram of the regions used in the QCD multijet ABCD background estimation method. Region D is the signal region and regions A, B and C are mutually exclusive control regions.

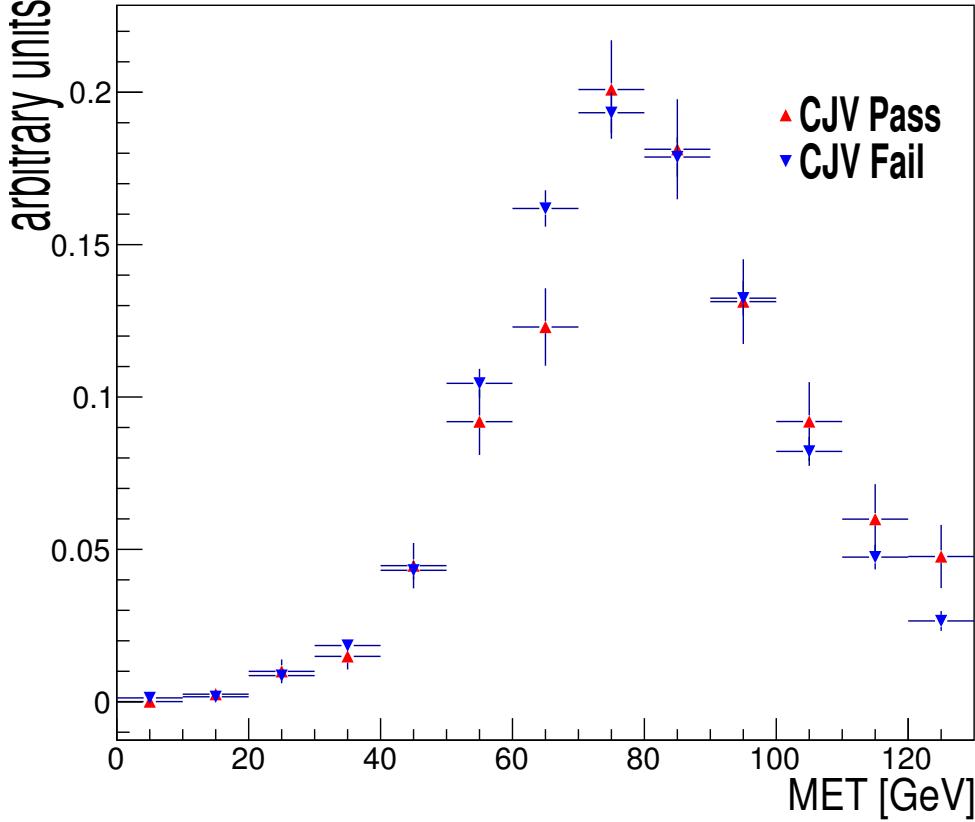
#### 4.2.5. QCD

The QCD multijet background remaining after the full event selection is mostly from events where jets are mismeasured. The size of the MC samples available for studying this process is not sufficient for them to be relied upon for extrapolation from a control region to the signal region. The remaining QCD multijet background is therefore estimated using a so-called “ABCD” method. In this method four regions, A, B, C and D, are defined according to whether events pass or fail the  $\not{E}_T^{\text{no-}\mu}$  and CJV cuts as shown in Figure 4.8. Region D is the signal region and regions A, B and C are three mutually exclusive control regions.

The efficiency to pass the  $\not{E}_T^{\text{no-}\mu}$  and CJV cuts can be determined from the ratios between regions A and B, and A and C respectively. The number of events expected in the signal region is then:

$$\begin{aligned} N_D &= N_A \cdot \frac{N_B}{N_A} \cdot \frac{N_C}{N_A} \\ &= \frac{N_B \cdot N_C}{N_A}, \end{aligned} \tag{4.6}$$

where  $N_{A,B,C}$  is the number of events observed in region  $A, B, C$  in data minus the number expected from V+jets or other minor backgrounds, i.e. the number of events in the region believed to be from QCD. This method relies on the probability of an event passing the CJV being uncorrelated with the  $\not{E}_T^{\text{no-}\mu}$  of the event. This has been checked by comparing the  $\not{E}_T^{\text{no-}\mu}$  distribution, below 130 GeV, for events which pass and fail the CJV. The maximum fractional difference observed between bins of these two distributions is 40%, so this is added as a systematic to the QCD multijet background yield. The method was also tested in a region orthogonal to the signal region with all requirements the same as those of the signal region except  $\Delta\phi_{jj}$  which was required to be greater than 2.6. In this test region, which is expected to be QCD dominated, the prediction agreed with the expectation within 15%, which is within the systematic



**Figure 4.9.:** Distribution of the  $\cancel{E}_T^{\text{no-}\mu}$  for events passing and failing the CJV. The largest difference seen in any bin is 40%. This difference is used to assign as a systematic uncertainty on the QCD multijet background estimation.

**Table 4.6.:** Numbers of events from data and MC in each region used in the QCD multijet background estimate and the final estimated number of events.

Region	Data	Background	Data-Background
$N_A$	5118	$222 \pm 14$	$4896 \pm 73$
$N_B$	773	$586 \pm 17$	$184 \pm 33$
$N_C$	896	$76.9 \pm 8.3$	$819 \pm 31$
$N_D$	-	-	$30.9 \pm 1.6$

uncertainty assigned to the method. The results of using this method to estimate the number of QCD events in the signal region are shown in Table 4.6.

#### 4.2.6. Minor backgrounds

In addition to the V+jets and QCD multijet backgrounds above which calculated, as described in this section, account for 94% of the expected events in the signal region, there is also a small number of events expected from other minor backgrounds (single top quark production, top quark pair production and diboson production). Due to its small contribution this number of events is taken directly from MC, the diboson backgrounds being simulated using PYTHIA 6, the single top quark background using POWHEG and the top quark pair production and  $Z/\gamma^* \rightarrow \ell\ell$  backgrounds using MADGRAPH. The cross-sections used to normalise these MC samples were taken from the most up to date CMS published results at the time of the analysis [50–54]. The final estimate of the number of events from minor backgrounds in the signal region is  $20 \pm 8.2$ (MC stat), with 70% of these expected to be from diboson production.

### 4.3. Systematic uncertainties

The dominant uncertainties in the analysis are the statistical uncertainties on the V+jets backgrounds due to the number of data events observed in the control regions. In addition, as well as those mentioned in Section 4.2, there are several further systematic uncertainties on the expected numbers of signal and background events. These uncertainties are described individually below and the fractional uncertainty on the total expected number of signal and background events from all sources of uncertainty are summarised in Table 4.7.

#### 4.3.1. Jet energy scale

The true energy of a jet reconstructed by CMS is not necessarily the same as the true energy of all the particles that make it up. As described in Section 3.6.3, jet corrections are applied to remedy this. The correction for the ratio between reconstructed and true jet energy is referred to as the jet energy scale (JES). Uncertainties on the JES come from several sources. The JES obtained from the dijet  $p_T$  balance method for instance has an uncertainty from the jet resolution bias [43]. This bias arises because the jet  $p_T$  spectrum sharply falls with increasing  $p_T$ . Such a spectrum leads to the well measured jet being used as the base for the balance method to be more likely to have fluctuated up in  $p_T$  than down. The main uncertainties in the photon/Z balance methods come



**Figure 4.10.:** The distribution of the sub-leading jet’s  $p_T$  in W+Jets MC events with the nominal JES and for JESUP and JESDOWN.

from the limited number of events in the samples used. The JES obtained in MC is also different when measured with different MC generators, which leads to an uncertainty.

These uncertainties on the JES give rise to an uncertainty on the energy of all jets in CMS events. The impact of this uncertainty on the event yields expected and observed in this analysis was estimated by altering the JES correction up by one standard deviation, “JESUP”, and down by one standard deviation “JESDOWN”, and recalculating the energy and momentum of all jets in each event. The  $\cancel{E}_T$  is recalculated taking into account the updated jet energies. Furthermore, as the jet energy scale uncertainty varies with  $p_T$  and  $\eta$ , it is possible for the  $p_T$  ordering of the jets to change when the JES is changed. The VBF tag pair is therefore rechosen to be the new highest  $p_T$  pair of jets.

After modifying the JES, the analysis is reperformed and the resulting change in the expected signal and background yields is taken to be the uncertainty due to the JES. The sub-leading jet’s  $p_T$  in a W+jets MC sample is shown for the nominal JES, JESUP and JESDOWN in Figure 4.10. It can be seen that altering the JES results in a smooth change in the jet  $p_T$ , indicating that the difference in the number of events passing the analysis cuts is not due to individual events with large weights migrating in and out of the signal region.

### 4.3.2. Jet energy resolution

The width of the jet energy distribution, the jet energy resolution (JER), differs between MC and data. This is partly due to MC samples being generated before and during data taking, and thus before a measurement of the exact resolution in data can be performed. Measuring the JER in data is done in a very similar way to the measurement of the JES using dijet and photon/Z-jet balance techniques and by comparing the reconstructed energy to that generated in MC, and thus has the same sources of uncertainty.

To correct the MC resolution to match that in the data, the  $p_T$  of all jets in MC is “smeared”. The smearing is carried out using two methods. The first method is used for jets in MC that are within 0.5 in the  $\eta - \phi$  plane of a generator jet. In this method the difference between the reconstructed and generator jet  $p_T$  is scaled by a correction factor,  $c$ , chosen to be the ratio between the resolution of the data and MC and calculated as a function of the jet’s  $p_T$  and  $\eta$ . The resulting jet  $p_T$  is given by:

$$p'_T = \max [0, p_{T\text{gen}} + c(p_T - p_{T\text{gen}})], \quad (4.7)$$

where  $p_T$  is the initial transverse momentum,  $p'_T$  is the transverse momentum after smearing and  $p_{T\text{gen}}$  is the matched generator jet’s  $p_T$ . This procedure has the advantage that the smearing is not reliant on random factors and is therefore reproducible, making synchronisation between analysis implementations easier.

The second method is used when a jet has no matching generator jet. In this case a random correction is necessary. The technique used is to add a fluctuation to the jet’s  $p_T$  with a size obtained by sampling a gaussian with width:

$$\sqrt{(c^2 - 1) \sigma_{MC}}, \quad (4.8)$$

where  $c$  is the same correction factor from the method above and  $\sigma_{MC}$  is the initial MC resolution as a function of  $p_T$  and  $\eta$ . As well as being random and thus difficult to reproduce, this method has the disadvantage that it can only be used to worsen the resolution.

The analysis is performed three times with three different smearings, one where the MC resolution is smeared to match the nominal data resolution, which is used for the main signal and background estimations, and two where the MC resolution is made to match the improvement, “JERBETTER” and worsening, “JERWORSE” of the jet energy resolution by one standard deviation. For the nominal and JERWORSE smearings the

limitation that the unmatched jet smearing method can only worsen the resolution is not a problem, as the initial MC resolution is better than that in data for all values of jet  $p_T$  and  $\eta$ . However, for the JERBETTER smearing it is necessary to improve the resolution for some jets. Fortunately, the differences between the generated resolution and the JERBETTER resolutions where improved resolution is required are small, so in these cases no smearing is applied. For all smearings the resulting changes in jet  $p_T$  are propagated through to the  $\cancel{E}_T$ . The differences between the signal and background yields obtained with the nominal JER, JERBETTER and JERWORSE are used to assign an uncertainty due to the JER.

### 4.3.3. Unclustered energy scale

In addition to the uncertainties on the  $\cancel{E}_T$  from the propagation of JES and JER uncertainties there are also uncertainties from the other elements contributing to the  $\cancel{E}_T$ . Electrons and muons contribute to the  $\cancel{E}_T$ , but have very good resolution and small scale uncertainties compared to jets so their contribution to the  $\cancel{E}_T$  uncertainty is considered negligible [44]. Unclustered energy, which is made up of all of the energy deposits in the calorimeters not identified as part of an object, such as a jet or lepton, still contributes to the  $\cancel{E}_T$ , and has a non-negligible scale uncertainty [44].

The unclustered energy scale is measured using photon and Z events with jets present in them, where it can be assumed that the  $\cancel{E}_T$  should be zero[44]. After jet energy corrections, the distribution of the remaining difference between the photon or Z momentum and the jets is therefore centered around zero, and its width can be taken as the uncertainty on the unclustered energy scale. The unclustered energy scale in all events is modified up and down by this uncertainty and the  $\cancel{E}_T$  recalculated. The differences in the obtained signal and background event yields obtained through this process are used as the uncertainty from this source.

### 4.3.4. Lepton identification and isolation efficiency

As described in Section 1.5 MC events are reweighted using scale factors to account for differences in the electron and muon identification and isolation efficiency. The weights due to these efficiencies are varied up and down by the uncertainties from the tag and probe method used to measure them, and the difference in the resulting signal and background yield used as the uncertainty from this source.

An uncertainty of 8% is added to the estimate of the  $W \rightarrow \tau\nu$  background to account for the uncertainty in the tau identification efficiency, which is measured using  $Z \rightarrow \tau\tau$  events where one tau decays to a muon and the other hadronically [55]. 5% of events in the  $W \rightarrow \tau\nu$  control region also appear to be due to  $W \rightarrow e\nu$  events where the electron or a jet has been misreconstructed as a tau, so a further 5% systematic is assigned to the  $W \rightarrow \tau\nu$  estimate.

#### 4.3.5. Other uncertainties

Additional uncertainties arise from several sources. For instance, the  $Z \rightarrow \nu\nu$  background estimate is reliant on the ratio of the cross-sections for the  $Z \rightarrow \nu\nu$  and  $Z/\gamma^* \rightarrow \mu\mu$  processes. A 20% uncertainty on this ratio was applied to cover the difference between the values of the ratio calculated using MADGRAPH and using MCFM [49]. Further uncertainties come from the measurement of the distribution of the number of primary vertices used in the pileup weights, described in Section 1.5, the probability density function (pdf)s and QCD scale used in the signal cross-section measurements, differences in the gluon fusion ( $ggH$ )  $\Delta\phi_{jj}$  spectrum depending on the MC generator used [49], the cross-sections used to normalise the minor backgrounds, and the measurement of the total integrated luminosity.

### 4.4. Results

The final results of all the background estimation methods and systematic uncertainty studies are summarised in Table 4.8. The total number of events expected from background processes in the signal region is  $332 \pm 46$  (stat)  $\pm 45$  (syst). Assuming the presence of a Higgs boson with a mass of 125 GeV, SM production and  $\mathcal{B}(H \rightarrow \text{inv}) = 100\%$ ,  $224 \pm 31$  signal events would be expected, with 6% of these from  $ggH$  and the remainder from VBF production. 390 events are observed, which is within one standard deviation of the background only prediction. Figure 4.11 shows the  $\cancel{E}_T^{\text{no-}\mu}$  and  $M_{jj}$  of the background and signal events expected, and the data observed, in the signal region.

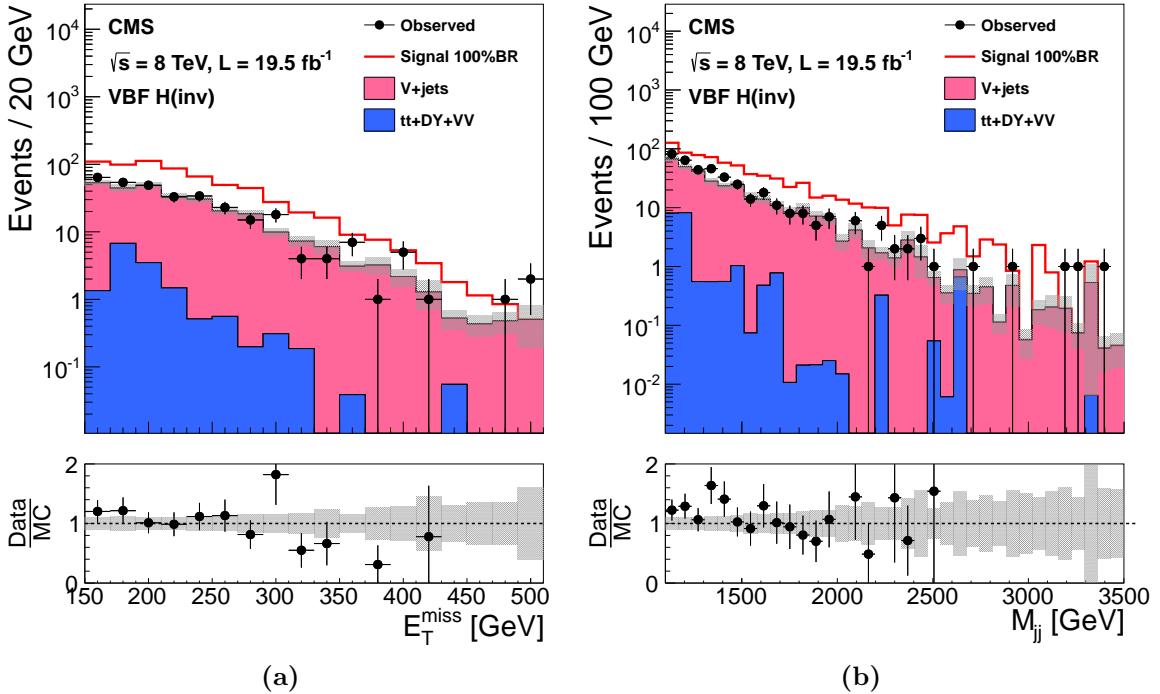
As no excess is observed, the asymptotic  $CL_S$  technique described in Section 1.6 is used to place an upper limit on the production cross-section times branching fraction,  $\sigma \times \mathcal{B}$ , at 95% CL. Under the assumption of SM production this limit can be interpreted

**Table 4.7.:** A summary of the uncertainties in the total expected signal and background yields. All uncertainties are quoted as the percentage change in the yield when each effect is varied up and down according to its uncertainty. The signal yields assume a Higgs boson mass of 125 GeV.

Uncertainty source	Total background	Signal
Control region statistics	11%	—
MC statistics and $Z \rightarrow \nu\nu$ cross-section ratio	11%	4%
Jet/ $\cancel{E}_T$ energy scale/resolution	7%	13%
QCD background estimation	4%	—
Lepton efficiency	2%	—
Luminosity	0.2%	2.6%
Cross-sections	0.5–1%	—
Pdfs	—	5%
QCD scale	—	4%
ggH $\Delta\phi_{jj}$ spectrum	—	4%
Total	18%	14%

**Table 4.8.:** The estimated numbers of background and signal events, together with the observed yield, in the signal region. The signal yield assumes a Higgs boson mass of 125 GeV and  $\mathcal{B}(H \rightarrow \text{inv.}) = 100\%$ .

Process	Event yield
$Z \rightarrow \nu\nu$	$99 \pm 29 \text{ (stat)} \pm 25 \text{ (syst)}$
$W \rightarrow e\nu$	$67 \pm 5 \text{ (stat)} \pm 16 \text{ (syst)}$
$W \rightarrow \mu\nu$	$63 \pm 9 \text{ (stat)} \pm 18 \text{ (syst)}$
$W \rightarrow \tau\nu$	$53 \pm 18 \text{ (stat)} \pm 18 \text{ (syst)}$
QCD multijet	$31 \pm 5 \text{ (stat)} \pm 23 \text{ (syst)}$
Minor backgrounds	$20 \pm 8 \text{ (syst)}$
Total background	$332 \pm 36 \text{ (stat)} \pm 45 \text{ (syst)}$
VBF H(inv.)	$210 \pm 29 \text{ (syst)}$
ggH(inv.)	$14 \pm 10 \text{ (syst)}$
Observed data	390



**Figure 4.11.:** Distributions of the  $\cancel{E}_T^{\text{no-}\mu}$  (a) and  $M_{jj}$  (b) of events observed in data and expected from the background estimation methods described in Section 4.2 in the signal region. tt+DY+VV indicates the contribution from minor backgrounds. The hatched region illustrates the systematic uncertainty. The QCD multijet background is not shown due to the very low number of events in the MC samples. The cumulative effect of a signal from a Higgs boson with mass of 125 GeV, decaying 100% to invisible final states is also shown [47].

as a limit on  $\mathcal{B}(H \rightarrow \text{inv})$ . All systematic uncertainties are modelled as log-normally distributed nuisance parameters, with the exception of the statistical uncertainty on the  $Z \rightarrow \nu\nu$  background, which is modelled as a gamma-normally distributed nuisance due to the low number of events in the dimuon control region. The resulting upper limits are shown as a function of Higgs boson mass in Figure 4.12, with the 95% CL observed (expected) limit on  $\mathcal{B}(H \rightarrow \text{inv})$  for a 125 GeV Higgs boson being 65% (49%). The green and yellow bands shown in Figure 4.12 denote the one and two sigma uncertainty bands respectively of the expected limit, also calculated using the asymptotic technique. The one (two) sigma band represents the region that the observation is expected to lie in 68% (95%) of the time if the background only hypothesis is true.

For all values of Higgs boson mass investigated the observed limit is approximately one sigma above the expected limit. If the measurements of the limit at each Higgs boson mass were not correlated, this could be seen as evidence for an excess. However, as this analysis has only a single bin, and no information on the shape of the event variable distributions is used, the measurements for the different Higgs boson masses are 100% correlated with each other, so there is no significant evidence of non-SM behaviour.



**Figure 4.12.:** Expected and observed 95% CL upper limits on the VBF  $\sigma \times \mathcal{B}$  in pb (a) and normalised to the SM VBF Higgs boson production cross-section (b) [47]. The green and yellow bands are the one and two sigma uncertainty bands of the expected limit.

# Chapter 5.

## Search for invisibly decaying Higgs bosons in Run 1 parked data

The parked data, described in Section 2.2.5, used for this analysis was collected using a range of triggers with similar but looser requirements than that used for the prompt data analysis described in the previous chapter. These looser requirements allow areas of phase space which were previously removed by the prompt trigger to be used. However, these areas also have very high levels of QCD multijet backgrounds, and require the analysis selection and some background estimation methods to be redesigned compared to the prompt analysis. As it was reconstructed later the parked data also uses different, better, detector calibrations, such as jet energy calibrations, calculated with the full Run 1 LHC dataset.

### 5.1. Trigger

The parked data trigger varied throughout LHC Run 1 due to the estimates of both the trigger bandwidth available and the rate of the triggers used in the particular LHC conditions that occurred being updated. Run 1 was split into 4 “eras”, A, B, C and D. During era A data was not being parked, so the prompt data is used. The two other triggers used, one for eras B and C, and one for era D, differed from the prompt trigger in that there was no requirement on the  $\cancel{E}_T$  present at HLT level in each event and the jet  $p_T$  and  $M_{jj}$  requirements were looser. These looser requirements allow the trigger driven selection applied in the prompt data analysis to be relaxed, and more optimal signal and background control regions to be used. As the region accessible with the parked data includes the prompt data signal region as a subset, no improvement would be possible

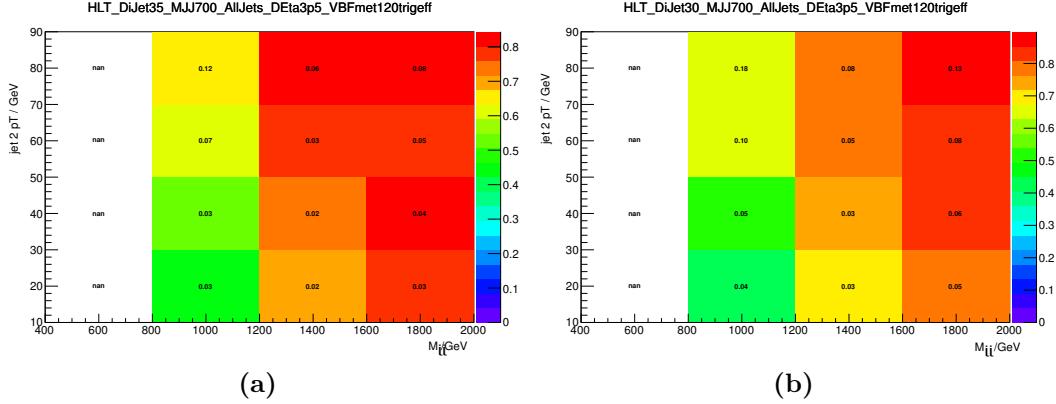
**Table 5.1.:** A summary of the requirements of the triggers used for this analysis in each of LHC Run 1’s eras. For the jet requirements all triggers require that there is at least one pair of jets in the event satisfying all of the jet requirements listed in this table. All requirements are on HLT variables unless stated otherwise.

Variable	Cut in era				
	A	B & C	D		
L1 $\cancel{E}_T$	$> 40 \text{ GeV}$				
$\cancel{E}_T^{\text{no-}\mu}$	$> 65 \text{ GeV}$	No requirement			
jet $p_T$ of both jets	$> 40 \text{ GeV}$	$> 35 \text{ GeV}$	$> 30 \text{ GeV}$		
$M_{jj}$	$> 800 \text{ GeV}$	$> 700 \text{ GeV}$			
$\Delta\eta_{jj}$	$> 3.5$				
$\eta_{j1} \cdot \eta_{j2}$	$> 0$				

from applying the analysis designed for the prompt data to the parked data without modification. The exact values of the trigger selection cuts are summarised in Table 5.1.

In order to get the biggest improvement in sensitivity from the parked data, it is important to measure the trigger efficiency accurately, so as to take advantage of as much of the increase in accessible phase space as possible. As three different triggers are used the measurement of trigger efficiency must be performed separately for each one. Also, the variables used in the trigger are highly correlated with each other. These correlations mean it is important to either only use regions of phase space where the trigger is fully efficient, as was done in the prompt analysis, or to measure the trigger efficiency in a way that accurately models the effect of these correlations. The cuts required to ensure that each trigger is fully efficient throughout the region selected can be ascertained from Figure 4.1.

As the trigger used in era A was the same as that used for the prompt analysis, no relaxation would be possible if full trigger efficiency is required and the data from era A is to be used. Era A only accounts for 5% of the total data, so one possibility is not to use the era A data and to relax the selection to the point of full efficiency of the next tightest trigger. However, it would still be necessary to discard data in the trigger turn on region which is expected to contain signal events. For these reasons several approaches to measuring the trigger efficiency as a function of the values of all variables used in it were investigated.



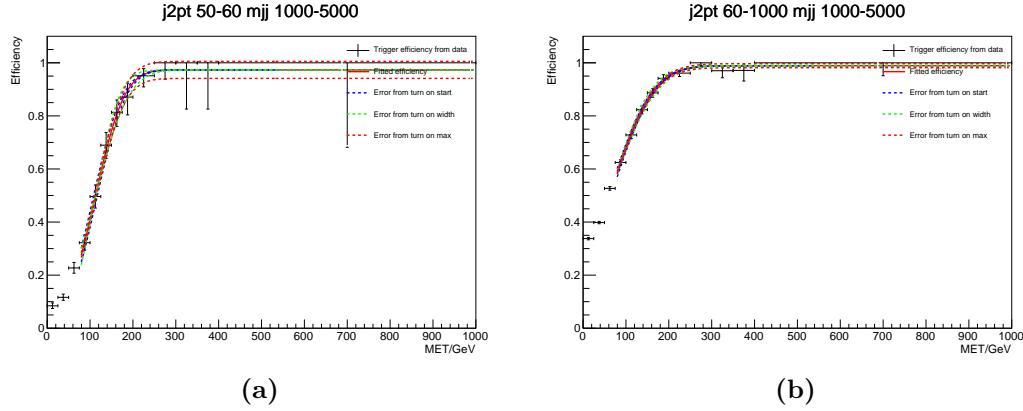
**Figure 5.1.:** The efficiency for the trigger (color-scale) used in eras B and C (a) and era D (b), as a function of  $M_{jj}$  and sub-leading jet  $p_T$  for events with  $\cancel{E}_T^{\text{no-}\mu}$  between 60 and 120 GeV. The efficiency was measured using a single muon dataset collected with an orthogonal trigger.

First, the trigger efficiency was measured three dimensionally as a function of  $\cancel{E}_T^{\text{no-}\mu}$ ,  $M_{jj}$  and sub-leading jet's  $p_T$ . An example of one of the results of these measurements in one of the bins in  $\cancel{E}_T^{\text{no-}\mu}$  for the era B and C, and the era D triggers can be seen in Figure 5.1. The three variables used were chosen because the trigger becomes fully efficient very quickly as a function of the  $\eta$  related variables, so no parameterisation of this efficiency is necessary. The number and size of the bins was chosen to ensure that sufficient events are present in each bin to prevent the statistical error on the efficiency measurement being larger than the differences between bins. As can be seen from the figure, this leads to very large differences in efficiency between bins, which leads to discontinuities in the  $\cancel{E}_T^{\text{no-}\mu}$ ,  $M_{jj}$  and sub-leading jet  $p_T$  distributions when the measured efficiency was applied to MC events as a weight. This method was therefore not suitable for use in the final analysis.

In order to achieve a smoother trigger efficiency parameterisation, coarse bins in  $M_{jj}$  and sub-leading jet  $p_T$  were chosen and a fit to the  $\cancel{E}_T^{\text{no-}\mu}$  efficiency distribution in each bin was performed, using the following function:

$$f(x) = \frac{A}{2} \cdot \left( 1 + \frac{2}{\sqrt{\pi}} \int_0^{\frac{x-B}{\sqrt{C}}} e^{-t^2} dt \right), \quad (5.1)$$

which has a maximum value of A, and is derived from the error function with centre B and width C. The width, maximum and centre of the function are all allowed to float in the fit. The events used in this study were required to have leading jet  $p_T > 50$  GeV,



**Figure 5.2.:** The results of performing a fit of the function in Equation 5.1 to the efficiency of the trigger used in era D, measured in a sample of single muon events collected with an orthogonal trigger. The dashed bands show the uncertainty on the fit due to each of the three parameters of the fit. The two bins of dijet mass ( $m_{jj}$ ) and sub-leading jet's  $p_T$  ( $j_2 p_T$ ) shown are those with the two highest numbers of events from the final signal region described in Section 5.2.

$\eta_{j1} \cdot \eta_{j2} < 0$  and  $\Delta\eta_{jj} > 3.6$  to ensure that there are no inefficiencies due to these variables. The results of these fits for the two  $M_{jj}$  and sub-leading jet  $p_T$  bins containing the most events entering the final analysis selection are shown in Figure 5.2, and the results for the remaining bins are shown in Appendix A. Some of the plots in the appendix indicate that the parameters of the fit have taken extreme values, or have very large uncertainties. These extreme values and poor fits are mostly due to low numbers of events in the bin. The analysis selection described in Section 5.2, ensures that no events in these bins are used in the analysis.

Each MC event was weighted by the average of the efficiency found for each of the three triggers weighted by the amount of integrated luminosity recorded using each trigger as shown in the following equation:

$$w(p_{Tj2}, M_{jj}, \not{E}_T^{\text{no-}\mu}) = \frac{\sum_i \mathcal{L}_i \epsilon_i(p_{Tj2}, M_{jj}, \not{E}_T^{\text{no-}\mu})}{\sum_i \mathcal{L}_i}, \quad (5.2)$$

Where  $i$  are the three triggers,  $\epsilon_i(p_{\mathrm{T}j2}, M_{jj}, \cancel{E}_{\mathrm{T}}^{\mathrm{no-}\mu})$  is the measured efficiency for trigger  $i$  as a function of the event's sub-leading jet  $p_{\mathrm{T}}$ ,  $M_{jj}$  and  $\cancel{E}_{\mathrm{T}}^{\mathrm{no-}\mu}$ , and  $\mathcal{L}_i$  is the integrated luminosity collected using trigger  $i$ . The resulting trigger efficiency varies smoothly and leads to no unphysical discontinuities in the distributions of event variables as can be seen from the figures in the remainder of this chapter.

## 5.2. Event selection

As mentioned above a significant challenge of the parked data analysis is that the areas of phase space collected by the parked data triggers but not by the prompt data triggers have very large contributions from QCD multijet backgrounds. The QCD multijet contribution to VBF analyses is very hard to model because whilst the cross-sections for these processes are very high, the probability of any individual event being VBF-like is very low. The number of MC events that must be generated to make a representative sample is therefore prohibitively large. As a result of these difficulties, the parked data selection is separated into two stages. The first “preselection” stage selects a region of phase space which is not expected to be dominated by QCD processes. After this preselection has been made the background processes expected to make contributions are the same as in the prompt data analysis, and studies were undertaken into which background estimation methods and final signal region selection led to the best expected limit.

### 5.2.1. Preselection

The first element of the preselection was motivated by the trigger. The following selection was applied to ensure that the values of all event variables are above the trigger thresholds of all triggers used:

$$\begin{aligned} \eta_{j1} \cdot \eta_{j2} &< 0, \text{ leading jet } p_T > 50 \text{ GeV}, \Delta\eta_{jj} > 3.6, \\ \text{subleading jet } p_T &> 40 \text{ GeV}, M_{jj} > 800 \text{ GeV}, \cancel{E}_T^{\text{no-}\mu} > 90 \text{ GeV}. \end{aligned} \quad (5.3)$$

Where  $j1$  and  $j2$  are the leading and sub-leading  $p_T$  jets in the event and are chosen as the VBF tag jets. We also require that for the “signal-like” selection there are no veto electrons or muons in the event. The W+jets and Z+jets control regions used in the background estimation methods described in Section 5.3 impose different lepton requirements. QCD multijet processes still dominate the region defined by this selection, as can be seen in Figure 5.3a, where there are a lot more data events than expected from the background MC prediction. This difference is due to mismeasured QCD multijets events not being adequately modelled by the available MC samples, which are described in further detail in Section 5.3.9. The first variable that was used to achieve this reduction is the  $\cancel{E}_T$  significance,  $\mathcal{S}$ , which is defined as the ratio between  $\cancel{E}_T^{\text{no-}\mu}$  and the square root of the sum of the transverse energy of all particles in the event which is an estimate of

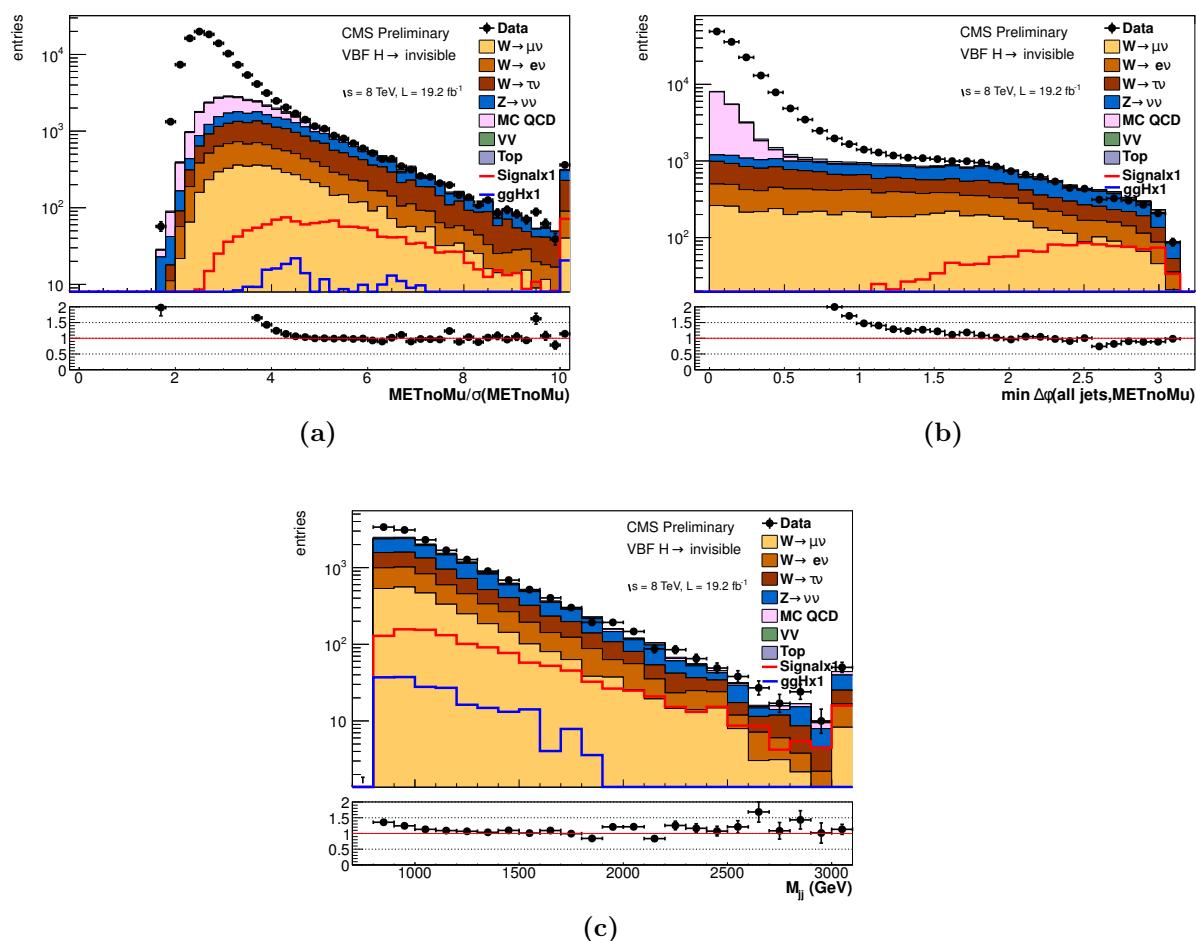
the statistical error on the  $\cancel{E}_T$ . The intention of the  $\mathcal{S}$  cut is to remove events which have a large amount of  $\cancel{E}_T$ , but also have an even larger amount of visible energy, meaning that the  $\cancel{E}_T$  is likely to be from mismeasurement of the visible particles. The preselection requires that this variable be greater than 3. The value of this cut was chosen by looking at Figure 5.3a and removing the region with the most disagreement between data and MC. The resulting region shown in Figure 5.3b still does not display good agreement between data and the MC prediction, however the disagreement is smaller.

After the cut on  $\mathcal{S}$ , a requirement that the  $\cancel{E}_T^{\text{no-}\mu}$  is not too close to any jets in  $\phi$  was made. This requirement was motivated by the high probability for the  $\cancel{E}_T$  to be aligned with a jet in the case where the jet has been mismeasured. Two variables were investigated, the first was the minimum azimuthal angle difference between either of the two tag jets and the  $\cancel{E}_T^{\text{no-}\mu}$ ,  $\min\Delta\phi(j_{1/2}, \cancel{E}_T)$ , and the second was the minimum azimuthal angle difference between any jet with  $p_T$  greater than 30 GeV and the  $\cancel{E}_T^{\text{no-}\mu}$ ,  $\min\Delta\phi(j, \cancel{E}_T)$ . At a similar signal efficiency the difference between the observed number of events and the MC background prediction, which is an indication of the remaining QCD multijet background, was found to be 80% smaller for a cut on  $\min\Delta\phi(j, \cancel{E}_T)$  than a cut on  $\min\Delta\phi(j_{1/2}, \cancel{E}_T)$ . The same cut on  $\min\Delta\phi(j, \cancel{E}_T)$  was also found to reduce top quark related backgrounds by a factor of two compared to a cut on  $\min\Delta\phi(j_{1/2}, \cancel{E}_T)$ . We therefore require that  $\min\Delta\phi(j, \cancel{E}_T) > 1.0$  for events to pass the preselection.  $\min\Delta\phi(j, \cancel{E}_T)$  was found to give significantly better signal efficiency than  $\Delta\phi_{jj}$  for the same background rejection, so no cut was made on  $\Delta\phi_{jj}$ .

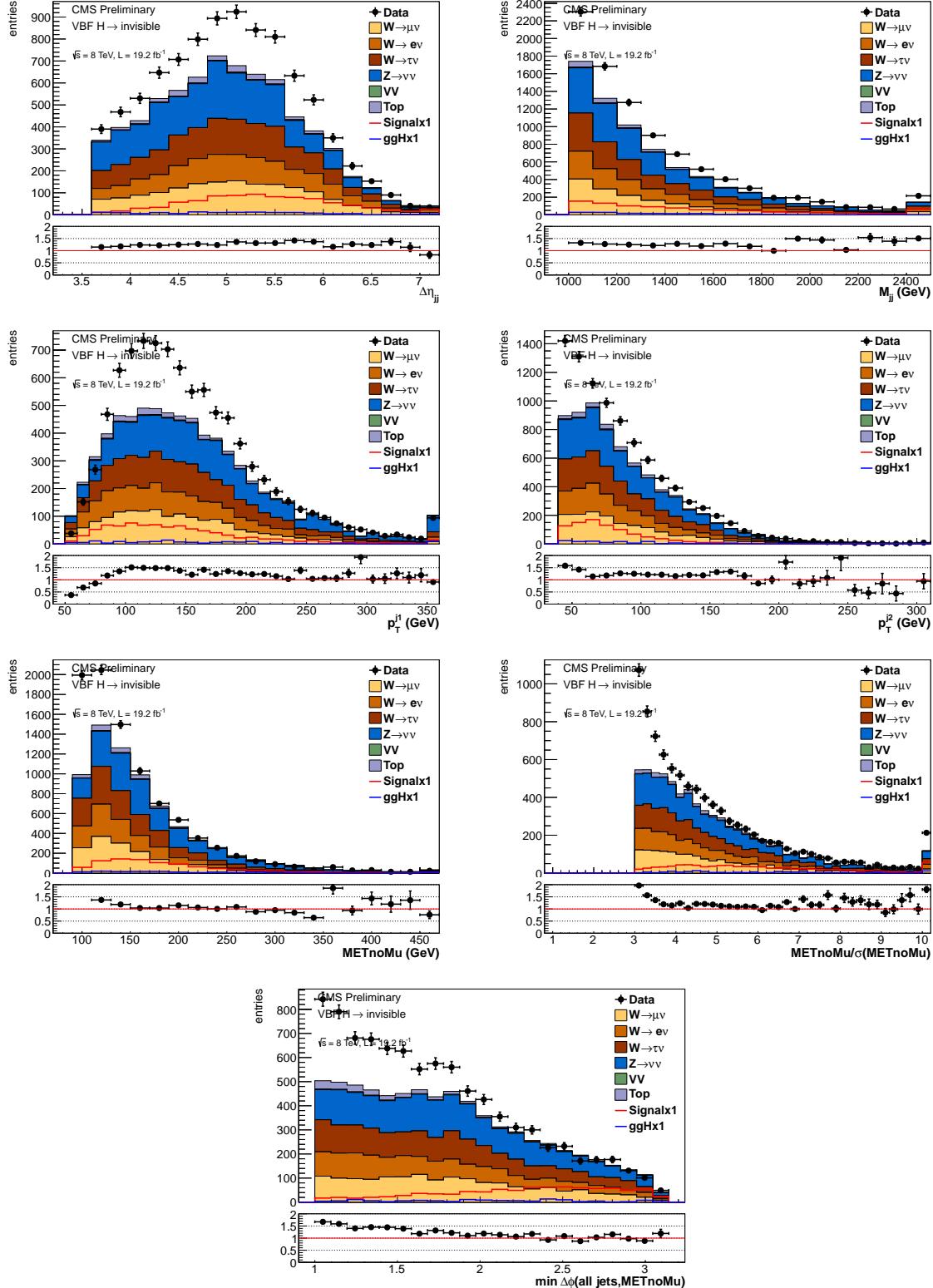
The  $M_{jj}$  distribution after the  $\min\Delta\phi(j, \cancel{E}_T)$  cut is shown in Figure 5.3c. Whilst the agreement for large  $M_{jj}$  is good, it can be seen that the first bin of the distribution, where mismeasured QCD multijet events would be expected, due to their not recoiling against another object, shows a significant disagreement. The final cut of the preselection is therefore to require that  $M_{jj} > 1000$  GeV. This cut also ensures that none of the bins of the trigger efficiency which have extreme values due to low numbers of events are used. In summary the full preselection is as follows:

$$\begin{aligned} \eta_{j1} \cdot \eta_{j2} &< 0, \text{ leading jet } p_T > 50 \text{ GeV}, \Delta\eta_{jj} > 3.6, \\ \text{subleading jet } p_T &> 40 \text{ GeV}, M_{jj} > 1000 \text{ GeV}, \cancel{E}_T^{\text{no-}\mu} > 90 \text{ GeV}, \\ \min\Delta\phi(j, \cancel{E}_T) &> 1.0, \mathcal{S} > 3.0. \end{aligned} \quad (5.4)$$

Distributions of several variables after the full preselection are shown in Figure 5.4.



**Figure 5.3.:** (a)  $\mathcal{S}$  after the trigger driven selection described in Equation 5.3. (b)  $\min\Delta\phi(j, \cancel{E}_T)$  after the trigger driven selection and requiring  $\mathcal{S} > 3$ . (c)  $M_{jj}$  after the trigger driven selection and requiring  $\mathcal{S} > 3$  and  $\min\Delta\phi(j, \cancel{E}_T) > 1$ . All three plots are of the signal-like region with the MC scaled using the background estimation methods described in Section 5.3. The disagreement between data and the predictions from background MC samples is believed to be due to mismeasured QCD multijet events which are not well modelled by the available MC samples.



**Figure 5.4.:** From top to bottom left to right distributions of  $\Delta\eta_{jj}$ ,  $M_{jj}$ , leading jet  $p_T$ , sub-leading jet  $p_T$ ,  $\cancel{E}_T^{\text{no-}\mu}$ ,  $\mathcal{S}$  and  $\min\Delta\phi(j, \cancel{E}_T)$  for events passing the full preselection. No QCD multijet contribution is shown, which accounts for the difference between the data observation and background prediction.

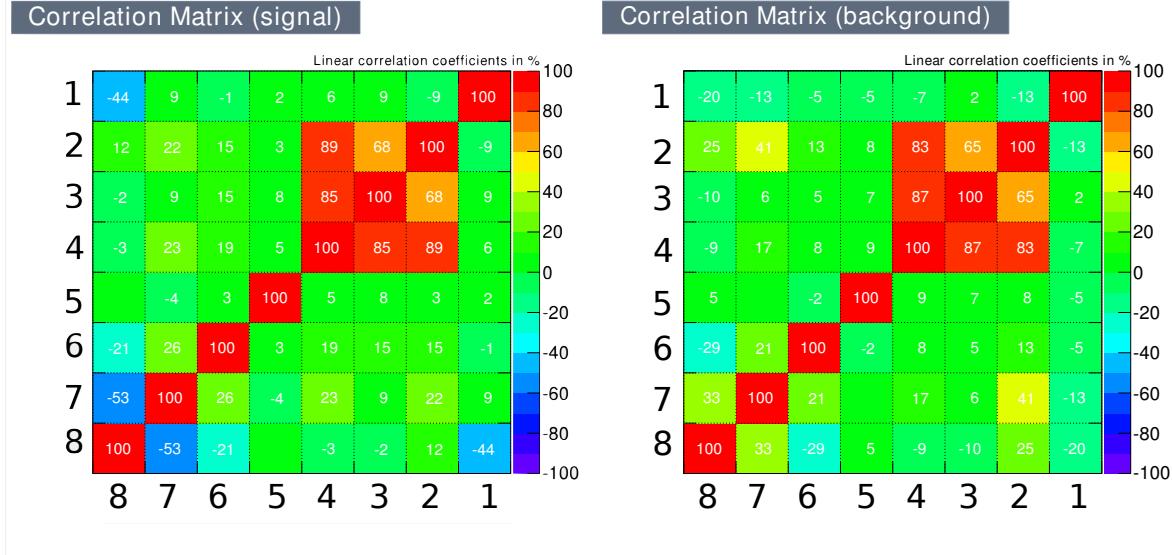
### 5.2.2. Signal region selection

As can be seen from Figure 5.4, there is still a significant difference between the data and the MC background prediction. It is also evident that the main areas where disagreement occurs are where contributions from QCD multijet backgrounds (which are not modelled in the figure) would be expected to contribute, i.e. at low  $\mathcal{S}$  and low  $\min\Delta\phi(j, \cancel{E}_T)$ , i.e. with jets close to the  $\cancel{E}_T^{\text{no-}\mu}$ . Outside these QCD-like regions good agreement between data and MC is seen. The approach taken was to place tight requirements on these two variables to reduce the QCD multijet background to sub-leading levels. The large uncertainties on any estimates of the number of events from QCD multijet processes therefore also become negligible. A requirement that events have  $\min\Delta\phi(j, \cancel{E}_T) > 2$  and  $\mathcal{S} > 4$ , was therefore imposed. The resulting relatively QCD-free “optimisation” region was blinded (i.e. the data were not looked at) to use for studies to determine the final signal region selection.

Two methods of signal region selection were investigated. The first method was a cut based selection. Starting from the optimisation region the cuts on  $\mathcal{S}$ ,  $\min\Delta\phi(j, \cancel{E}_T)$ ,  $\Delta\eta_{jj}$ , sub-leading jet  $p_T$  and  $M_{jj}$  were varied one at a time and the expected limit for each combination of cuts was calculated. The method described in Section 1.6 was used, with the background estimation techniques and systematic uncertainties described in Sections 5.3 and 5.4 respectively, to calculate the expected limit. In the case of the QCD multijet background, which has only a very small contribution to the signal region, the background estimation was performed once for the optimisation selection and used for all cut values. The estimations for all other background processes were repeated for each set of cuts. After each variable was varied the selection was updated to use the cut value that gave the best expected limit. After all the variables had been varied the process was repeated until no improvement in the expected limit was seen so as to avoid ignoring other better sets of cuts. The cut values that gave the best expected limit define the signal region and are as follows:

$$\begin{aligned} \eta_{j1} \cdot \eta_{j2} < 0, \Delta\eta_{jj} > 3.6, \text{ leading jet } p_T > 50 \text{ GeV}, \\ \text{subleading jet } p_T > 45 \text{ GeV}, M_{jj} > 1200 \text{ GeV}, \\ \cancel{E}_T^{\text{no-}\mu} > 90 \text{ GeV}, \mathcal{S} > 4.0, \min\Delta\phi(j, \cancel{E}_T) > 2.3. \end{aligned} \quad (5.5)$$

After this selection was defined a second MVA based method of optimising the selection was investigated to see if could improve on the cut based selection. BDT and fisher discriminants were trained using signal and background events passing the signal region



**Figure 5.5.:** Matrices of correlation coefficients for several variables in signal (left) and V+jets background (right) events passing the signal region selection. The variables are 1) the azimuthal angle difference between the  $\vec{E}_T^{\text{no-}\mu}$  and the unclustered energy in the event, 2) the square root of the hadronic energy in the event, 3)  $S$ , 4)  $\vec{E}_T^{\text{no-}\mu}$ , 5)  $M_{jj}$ , 6) the number of jets with  $p_T > 30$  GeV between the two tag jets in  $\eta$ , 7) the vectorial sum of the tag jets  $p_T$  and the  $\vec{E}_T^{\text{no-}\mu}$ , 8) the ratio between the magnitude of the vectorial sum of the tag jet's  $p_T$  and the  $\vec{E}_T^{\text{no-}\mu}$ .

selection. The signal region selection was used as the basis for this training so as to ensure that the number of events from QCD multijet backgrounds in the studied region was small. The optimisation procedure defined above was then repeated with the value of the discriminant considered as an additional variable. The correlation coefficients between the variables used as inputs to the MVA are shown for signal and V+jets background events in Figure 5.5. These variables were chosen as they showed the most difference between signal and background distributions and correlations out of a wide range of variables investigated. Without the addition of any additional systematic uncertainties from the understanding of the variables which were input to the MVA the best improvement in the expected limit over that of the signal region was less than 1%. It was therefore decided to use the cut based selection as the final event selection.

### 5.3. Background estimation

After selection the V+jets backgrounds, as in the prompt analysis, dominate. Also, as in the prompt data analysis contributions are expected from top quark and diboson related

processes. Finally, whilst it is reduced significantly by the selection described above, it is necessary to at least place an upper bound on the expected contribution from QCD multijet events.

The methods used to estimate the V+jets backgrounds are based on those used in the prompt data analysis. Equation 4.2 is used in several of these methods in this section, and it is repeated here:

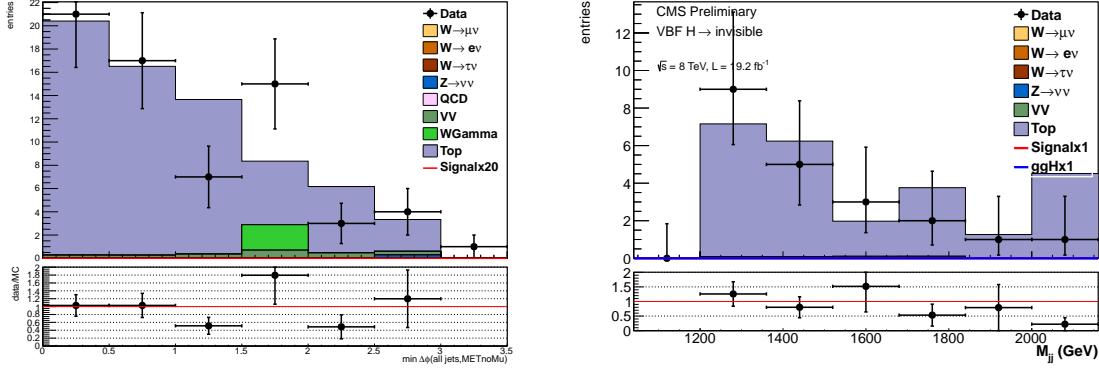
$$N_{Exp}^S = (N_{Data}^C - N_{Bkg}^C) \cdot \frac{N_{MC}^S}{N_{MC}^C}. \quad (5.6)$$

The terms on the right hand side of this equation which multiply the estimation from MC of the number of events due to a particular background process are often collectively referred to as the data driven scale factor. The changes in event selection for this analysis necessitated several changes from the methods for the prompt analysis. The use of this data driven method to investigate the top quark related background was also investigated. Furthermore, among other improvements, the systematic uncertainty on the  $Z \rightarrow \nu\nu$  background was reduced. All of these changes and improvements are described in this section.

### 5.3.1. Top quarks

Almost all top quarks decay to a W boson and a b quark [45]. Top quarks are either created in pairs, or via “single top” production where only one top quark is created in association with other quarks or a W boson. Top pair production results in two W bosons and two b quarks. Single top production results in some combination of W bosons and quarks. Where at least one of the W bosons decays leptonically and the lepton is misreconstructed, either of these processes can result in the appearance of  $\cancel{E}_T$  and jets that can coincidentally have VBF-like topology. Whilst the contribution from these processes is expected to be small in the signal region, making up around 1% of events there, the presence of W bosons and jets makes these processes very likely to contribute to the control regions used to estimate the W +jets background contribution. In the  $W \rightarrow \tau\nu$  control region approximately 15% of events are estimated to come from top quark processes. Data driven methods for estimating the top quark background and its uncertainties were therefore investigated.

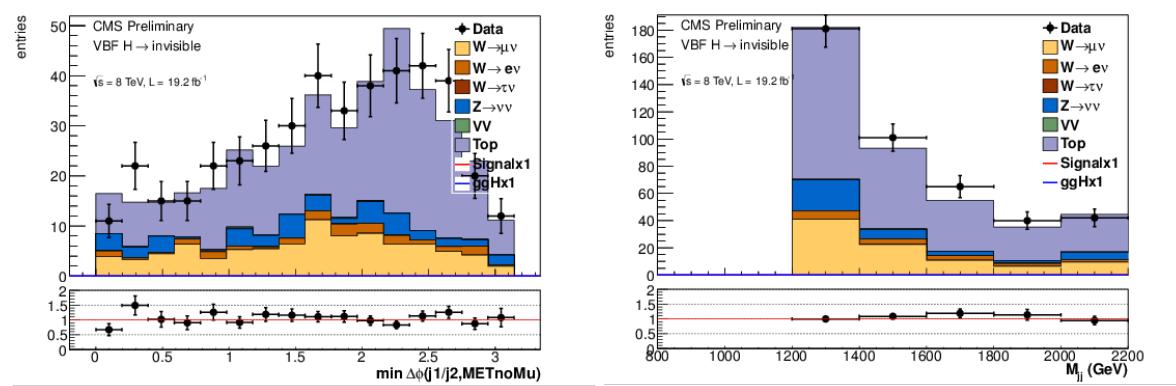
Initially a dilepton control region was investigated. This had the same cuts on the jet and  $\cancel{E}_T$  related variables as the signal region, but the lepton veto was replaced with a



**Figure 5.6.:** The distribution of  $\min\Delta\phi(j, \cancel{E}_T)$  (left) and  $M_{jj}$  (right) in the top control region with one tight electron and one tight muon.

requirement that there is exactly one tight electron and one tight muon. This final state would be expected in the case of top quark pair production or single top production with a W boson, where both the resulting W bosons decay leptonically to different leptons. This region had only single figure numbers of events expected, so the cut on  $\min\Delta\phi(j, \cancel{E}_T)$  was loosened to 0. It can be seen from Figure 5.6 that the ratio between data and MC in this region does not depend significantly on  $\min\Delta\phi(j, \cancel{E}_T)$ . The data driven scale factor obtained from this dilepton region was, within the statistics available, consistent with 1 and a good agreement between data and MC was seen in all variables studied.

A modification of this control region where events with either two tight electrons or two tight muons, and no other leptons was also studied. This final state would also be expected where two W bosons from top quark production decayed leptonically, except this time to the same flavour of lepton. In order to avoid Z boson contributions the lepton's invariant mass was required to be incompatible with that of a Z boson, i.e. outside of the range from 60 to 120 GeV. This control region also yielded good data-MC agreement and a scale factor compatible with 1. An issue with both of these control regions is that the top quark pair production to single top production ratio is very different to both the signal region and the  $W \rightarrow \tau\nu$  control region. MC estimations indicate that these top control regions have a negligible single top contribution, while the top background in the signal region has almost no top quark pair contribution. The  $W \rightarrow \tau\nu$  region is expected to be a mixture, its top quark background being 30% from single top events and 70% for top quark pairs, again estimated from MC. A single top control region was therefore also investigated.



**Figure 5.7.:** The distribution of  $\min\Delta\phi(j_{1/2}, \cancel{E}_T)$  (left) and  $M_{jj}$  (right) in the single top control region.

The single top region differed from the signal region in that the  $\min\Delta\phi(j, \cancel{E}_T)$  cut was removed, exactly one tight electron or muon was required, further leptons were vetoed and one of the tag jets was required to be compatible with being a b-jet. The restriction to a single lepton significantly reduces the top quark pair production contribution where both resulting W bosons decay leptonically, and the requirement of one b-jet reduces the W + jets contribution.

Identification of the b-jet was done using the combined secondary vertex (CSV) discriminant. b quarks are both heavier and longer lived than many other particles created at the LHC, meaning that their secondary decay vertex can be distinguished from the PV. CSV is an MVA based discriminant which uses information on secondary vertices and the lifetime of the particle to discriminate between jets from b quarks and those from other lighter quarks. The medium working point used for this control region correctly has an efficiency of approximately 85% for b-quarks and mis-identifies lighter quark jets as b jets approximately 1% of the time [56].

MC estimates indicate the single top region has a 17% single top contribution. This region again showed good data-MC agreement (as can be seen in Figure 5.7 and a scale factor compatible with 1 within uncertainties. Because good agreement between data and MC and scale factors compatible with 1 are seen in all investigated control regions, it was decided to use the MC prediction for the top background in all regions with no additional scale factor. A 20% systematic uncertainty was applied to this prediction which covered the largest deviation from 1 seen in the scale factors from the various control regions.

**Table 5.2.:** The inputs to and results of Equation 5.6, when used to estimate the  $W \rightarrow e\nu$  estimate in the signal region.

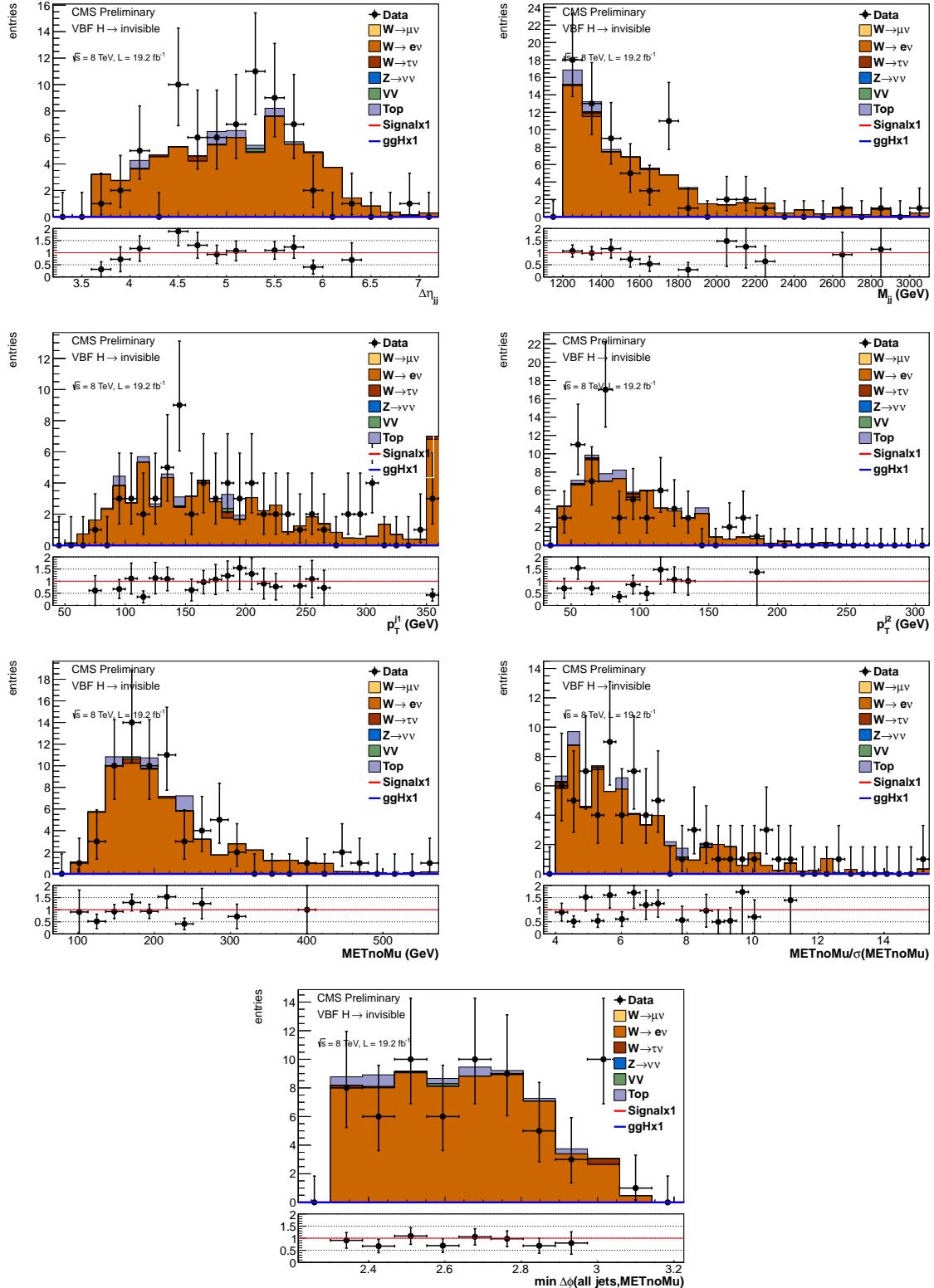
	Signal region	Control region
$N_{Data}$	N/A	$68 \pm 8.2 (\text{stat})$
$N_{Bkg}$	N/A	$3.5 \pm 1.2 (MC\text{stat})$
$N_{MC}$	$114.9 \pm 8.9 (MC\text{stat})$	$128.0 \pm 8.0 (MC\text{stat})$
$\frac{N_{data} - N^{bkg}}{N_C^{WMC}}$	$0.50 \pm 0.06 (\text{stat}) \pm 0.03 (\text{MC stat.})$	
$N_{W \rightarrow e\nu}$	$57.9 \pm 7.4 (\text{stat}) \pm 7.7 (\text{syst})$	N/A

### 5.3.2. $W \rightarrow e\nu + \text{jets}$

The  $W \rightarrow e\nu$  background in the parked data analysis is estimated using the same method as that used for the prompt data analysis being based on Equation 5.6. The control region used has the same requirements as the signal region, except that the electron veto is replaced with a requirement that there is one tight electron and no other electrons present in the event. The requirement of an electron removes signal events and enriches the region in  $W \rightarrow e\nu$  events. The distributions of several variables in data and MC (which has been scaled by the data driven scale factor extracted from this control region) are shown in Figure 5.8, where good agreement can be seen. A table of the inputs to and results of Equation 5.6 can be seen in Table 5.2. The table also shows that the expected contribution to this region from other background processes is small, being approximately 5%. The scale factor obtained for this background is 0.5, which is significantly different from 1, this difference is further investigated in Section 5.3.8.

### 5.3.3. $W \rightarrow \mu\nu + \text{jets}$

As for the  $W \rightarrow e\nu$  background the  $W \rightarrow \mu\nu$  background is estimated using Equation 5.6 with a control region enriched in  $W \rightarrow \mu\nu$  events through a change in lepton requirements. The control region used has the same requirements as the signal region, but with the muon veto replaced with a requirement that there is one tight muon and no other muons present in the event. The distributions of several variables in data and MC (which has been scaled by the data driven scale factor extracted from this control region) are shown in Figure 5.9, where good agreement can be seen. A table of the inputs to Equation 5.6 can be seen in Table 5.3. The contribution from other backgrounds in the  $W \rightarrow \mu\nu$



**Figure 5.8.:** Distributions of variables in data and MC events in the  $W \rightarrow e\nu$  control region. MC events from  $V+jets$  backgrounds are scaled by their data-driven scale factors. The variables shown are from top to bottom and left to right  $\Delta\eta_{jj}$ ,  $M_{jj}$ , the leading and subleading jet's  $p_T$ ,  $\cancel{E}_T^{\text{no-}\mu}$ ,  $\mathcal{S}$  and  $\min\Delta\phi(j, \cancel{E}_T)$ .

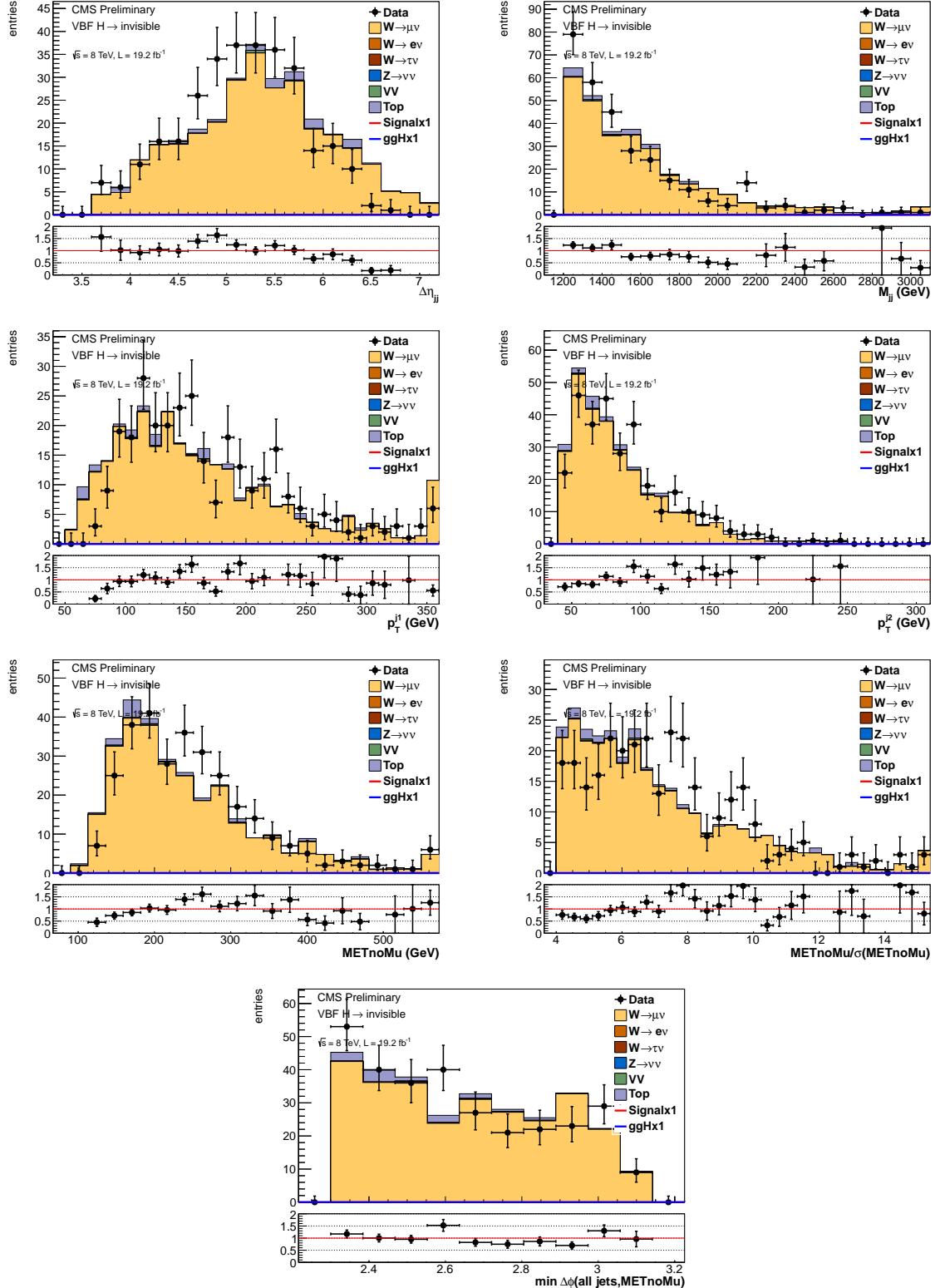
control region is approximately 5%. Again the scale factor obtained for this background is significantly different from 1, being 0.71, and further investigation of this is detailed in Section 5.3.8. Furthermore, the estimated contribution from this background is very different to that expected from  $W \rightarrow e\nu$ , an investigation of this difference is described in Section 5.3.5.

**Table 5.3.:** The inputs to and results of Equation 5.6, when used to estimate the  $W \rightarrow \mu\nu$  estimate in the signal region.

	Signal region	Control region
$N_{Data}$	N/A	$300 \pm 17.3 (\text{stat})$
$N_{Bkg}$	N/A	$14.8 \pm 2.5 (\text{MC stat})$
$N_{MC}$	$143.7 \pm 10.2 (\text{MC stat})$	$399.9 \pm 14.9 (\text{MC stat})$
$\frac{N_{Data} - N^{bkg}}{N_C^{MC}}$	$0.71 \pm 0.04 (\text{stat}) \pm 0.03 (\text{MC stat.})$	
$N_{W \rightarrow \mu\nu}$	<b><math>102.5 \pm 6.2 (\text{stat}) \pm 11.7 (\text{syst})</math></b>	N/A

### 5.3.4. $W \rightarrow \tau\nu + \text{jets}$

The signal region requirements do not include a veto of hadronic taus, due to the low identification efficiency and relatively high probability for a jet to be identified as a fake tau. Requiring that there is an identified hadronic tau in addition to the signal region selection results in a region containing only 2 data events. In order to increase the number of events in the tau control region the  $\min\Delta\phi(j, \cancel{E}_T)$  cut was removed. Whilst the requirement that there is an identified tau reduces the QCD multijet contribution in the low  $\min\Delta\phi(j, \cancel{E}_T)$  region significantly compared to what was seen during the choice of the preselection. However, poor data-MC agreement was still observed in the  $\min\Delta\phi(j, \cancel{E}_T) > 1$  region, which is evidence that some events from multijet processes are still present. To remove these QCD multijet events, whilst keeping a reasonable number of events in the resulting control region, a requirement that  $\min\Delta\phi(j_{1/2}, \cancel{E}_T)$  is greater than 1 and that the transverse mass of the hadronic tau and  $\cancel{E}_T$  system is greater than 20 GeV. The  $\min\Delta\phi(j_{1/2}, \cancel{E}_T)$  requirement reduces multijet backgrounds for the same reason that  $\min\Delta\phi(j, \cancel{E}_T)$  does, but it is a slightly looser requirement as it only considers the two leading jets. The transverse mass of the tau- $\cancel{E}_T$  system is a good variable to reject QCD where a lepton is present as in real W boson events the tau and  $\cancel{E}_T$  are expected to originate from the same object and therefore have significant invariant mass, which is not the case for QCD multijet events.



**Figure 5.9.:** Distributions of variables in data and MC events in the  $W \rightarrow \mu\nu$  control region. MC events from V+jets backgrounds are scaled by their data-driven scale factors. The variables shown are from top to bottom and left to right  $\Delta\eta_{jj}$ ,  $M_{jj}$ , the leading and subleading jet's  $p_T$ ,  $\cancel{E}_T^{\text{no-}\mu}$ ,  $\mathcal{S}$  and  $\min\Delta\phi(j, \cancel{E}_T)$ .

After the anti-QCD cuts the agreement between data and MC is good as can be seen in Figure 5.10 To account for the different  $\min\Delta\phi(j, \not{E}_T)$  selection in this region and the signal region the data driven scale factor was calculated both in the  $W \rightarrow \mu\nu$  control region, which has the signal region  $\min\Delta\phi(j, \not{E}_T)$  selection, and in a modified single muon control region with the  $W \rightarrow \tau\nu$  control region  $\min\Delta\phi(j, \not{E}_T)$  selection. The difference between these two scale factors was found to be 20%, so a 20% systematic uncertainty was added to the estimate of the  $W \rightarrow \tau\nu$  background.

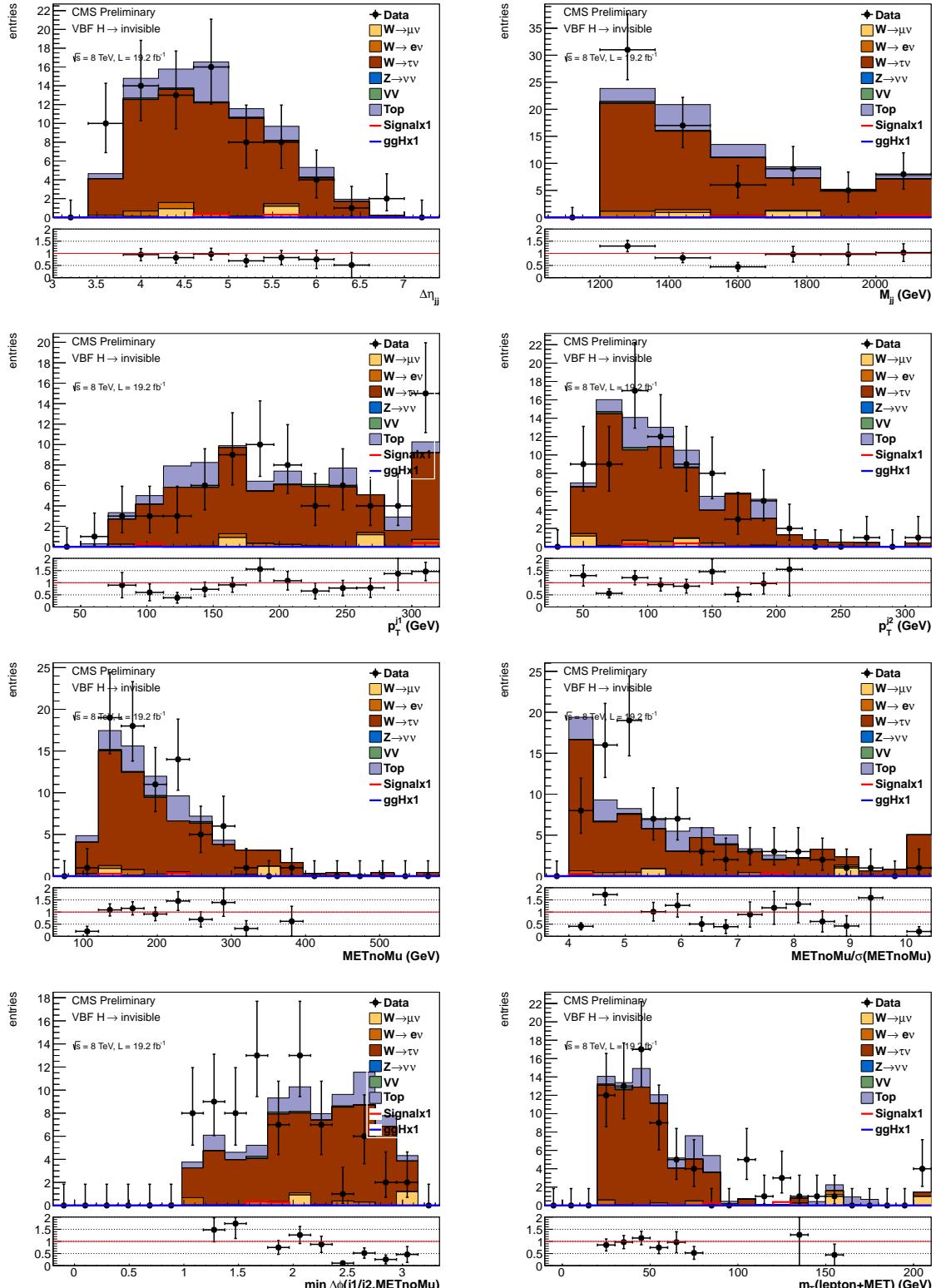
The single tau control region with a  $\min\Delta\phi(j_{1/2}, \not{E}_T)$  cut and no  $\min\Delta\phi(j, \not{E}_T)$  cut was used with Equation 5.6 to estimate the number of  $W \rightarrow \tau\nu$  events in the signal region. A table of the inputs to Equation 5.6 can be seen in Table 5.4. The contribution from other backgrounds in the  $W \rightarrow \tau\nu$  control region is approximately 15%, with most of these other background events being due to top quark related processes. The scale factor obtained for this background is 0.78, which is closer to 1 than those seen in the other  $W + \text{jets}$  backgrounds, however it also has the largest uncertainty. Further investigation of the  $V + \text{jets}$  scale factors is detailed in Section 5.3.8.

**Table 5.4.:** The inputs to and results of Equation 5.6, when used to estimate the  $W \rightarrow \tau\nu$  estimate in the signal region.

	Signal region	Control region
$N_{Data}$	N/A	$76 \pm 8.7 \text{ (stat)}$
$N_{Bkg}$	N/A	$13.3 \pm 2.8 \text{ (MC stat)}$
$N_{MC}$	$121.9 \pm 8.7 \text{ (MC stat)}$	$80.8 \pm 6.4 \text{ (MC stat)}$
$\frac{N_{data} - N_{bkg}}{N_C^{MC}}$	$0.78 \pm 0.11 \text{ (stat)} \pm 0.07 \text{ (MC stat.)}$	
$N_{W \rightarrow \mu\nu}$	$94.6 \pm 13.1 \text{ (stat)} \pm 23.8 \text{ (syst)}$	N/A

### 5.3.5. Differences between the $W \rightarrow e\nu$ , $W \rightarrow \mu\nu$ and $W \rightarrow \tau\nu + \text{jets}$ backgrounds

The number of  $W + \text{jets}$  events decaying to a particular flavour of lepton with VBF-like jet kinematics should be the same for all three flavours of lepton through lepton universality [45]. Differences between the numbers of background events from  $W \rightarrow e\nu$  and  $W \rightarrow \mu\nu$  and  $W \rightarrow \tau\nu$  must therefore be due to differences in the identification of the leptons. Hadronic taus have much lower identification efficiencies than the other two flavours of leptons, so might naively be expected to give rise to a much larger



**Figure 5.10.:** Distributions of variables in data and MC events in the  $W \rightarrow e\nu$  control region. MC events from V+jets backgrounds are scaled by their data-driven scale factors. The variables shown are from top to bottom and left to right  $\Delta\eta_{jj}$ ,  $M_{jj}$ , the leading and subleading jet's  $p_T$ ,  $\cancel{E}_T^{\text{no-}\mu}$ ,  $\mathcal{S}$ ,  $\min\Delta\phi(j_{1/2}, \cancel{E}_T)$  and the tau- $\cancel{E}_T$  system's transverse mass.

number of background events. However, due to the similarities between hadronic taus and jets, unidentified taus often lead to additional jets in the event and therefore increase the probability that an event will fail the analysis  $\min\Delta\phi(j, \cancel{E}_T)$  cut [46]. These two competing effects mean that the number of background events from  $W \rightarrow \tau\nu$  passing the signal region selection are not necessarily expected to be the same as those from  $W \rightarrow e\nu$  and  $W \rightarrow \mu\nu$ .

By contrast, electrons and muons have much more similar identification efficiencies to each other (see Sections 3.4 and 3.5). The 45% difference in the number of expected background events from these two processes was therefore unexpected. This difference was also not seen in the prompt data analysis, as can be seen from Table 4.8. The prediction of the  $W \rightarrow e/\mu\nu$  backgrounds is made up of a data driven scale factor and an estimate from MC of the number of events from the process expected in the signal region,  $N_{MC}^S$ . Both these elements were studied to try to understand the observed differences.

Firstly, the data driven scale factors for the electron and muon backgrounds differ by 30%, which is not sufficient to explain the full difference between the electron and muon background estimates. Furthermore, when systematic errors are taken into account this difference is only approximately one standard deviation. On the other hand,  $N_{MC}^S$  does show a significant difference.

To study the difference in  $N_{MC}^S$  two subregions of the signal region were studied, that with a generator level lepton inside the detector acceptance for both electrons and muons ( $|\eta| < 2.1$ ), and that with a generator lepton outside the detector acceptance for both electrons and muons ( $|\eta| > 2.4$ ). The number of events in these two subregions from both  $W \rightarrow e\nu$  and  $W \rightarrow \mu\nu$  MC can be seen in Table 5.5. The numbers of events with a generator level lepton inside acceptance are slightly but not much higher for electrons. This small difference is expected due to the lower identification efficiency for veto leptons making them less likely to cause an event to fail the lepton veto. Distributions of several variables were also studied for the inside acceptance events and found to be very similar for both  $W \rightarrow e\nu$  and  $W \rightarrow \mu\nu$  events.

Outside acceptance there are a lot more muon events than electron events. In this region neither flavour of lepton can be reconstructed and therefore cannot lead to an event failing the lepton veto, which implies that any difference is due to one or both flavours of lepton being reconstructed as a different object and affecting the jet or  $\cancel{E}_T$  related variables in the signal region selection. To study this effect further the  $\min\Delta\phi(j, \cancel{E}_T)$  requirement was relaxed to 1 and the distributions of several variables plotted for outside acceptance

**Table 5.5.:** The numbers of events predicted by MC in the two subregions of the signal region with a generator level lepton that is inside/outside the detector acceptance for both electrons and muons from  $W \rightarrow e\nu$  and  $W \rightarrow \mu\nu$  processes. The errors shown are from MC statistics.

Process	Inside acceptance	Outside acceptance
$W \rightarrow e\nu$	$73.7 \pm 6.8$	$30.2 \pm 4.9$
$W \rightarrow \mu\nu$	$61.5 \pm 6.8$	$74.4 \pm 7.3$

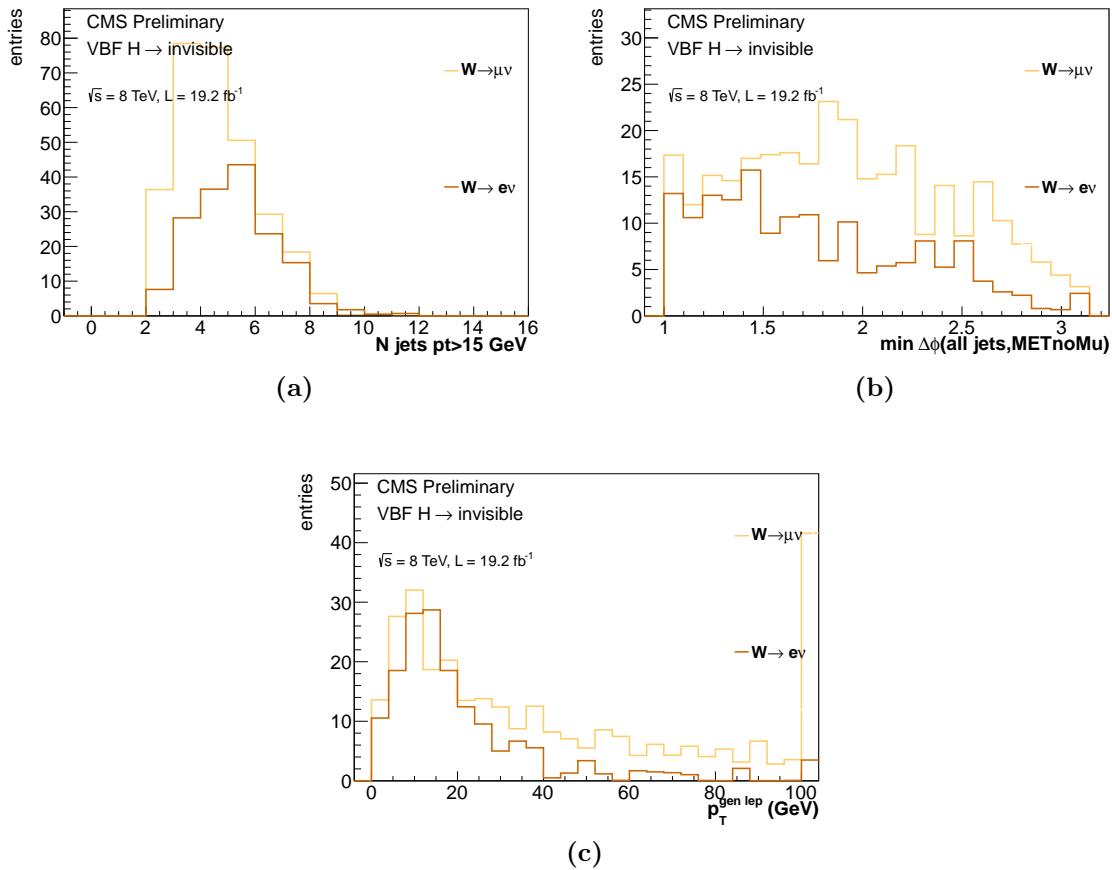
electron and muon events. Three of these distributions are shown in Figure 5.11. It can be seen from Figure 5.11a that electron events generally have more events than muon events. Figure 5.11b indicates that the electron events also have much lower values of  $\min\Delta\phi(j, \cancel{E}_T)$ . Finally, Figure 5.11c shows that there are very few events passing this region’s selection requirements which have a generator level electron with  $p_T > 30$  GeV, the threshold for jets to be included in the  $\min\Delta\phi(j, \cancel{E}_T)$  cut. These three pieces of information suggest that electrons are being reconstructed as jets when outside acceptance significantly more often than muons are. These misreconstructed jets then cause the electron events to fail the  $\min\Delta\phi(j, \cancel{E}_T)$  requirement.

It is to be expected that electrons outside of the detector acceptance will be reconstructed as jets, as these electrons will only be seen as deposits in the forward HCAL and therefore be indistinguishable from jets. By contrast, muons deposit very little energy in the calorimetry systems, so will simply not be identified if they are outside of the muon system acceptance. No requirements were made on jets which were further forward in  $\eta$  than the tag jets in the prompt data analysis, explaining why the difference in numbers of  $W \rightarrow e\nu$  and  $W \rightarrow \mu\nu$  events were not seen there.

### 5.3.6. $Z \rightarrow \nu\nu + \text{jets}$

The irreducible  $Z \rightarrow \nu\nu$  background is estimated using a very similar method to that used in the prompt data analysis, and a detailed description can be found in Section 4.2.4. A reminder of the method highlighting the differences from the prompt analysis is given here.

The method starts by defining a dimuon control region by taking the signal region requirements and replacing the muon veto with the requirement that there are two tight muons with invariant mass compatible with a  $Z$  boson, i.e. between 60 and 120 GeV,



**Figure 5.11.:** Distributions of the number of jets with  $p_T > 15$  GeV (a),  $\min\Delta\phi(j, \cancel{E}_T)$  (b) and the  $p_T$  of the leading generator level electron/muon in  $W \rightarrow e/\mu\nu$  events in a region with cuts which are the same as those of the signal region except that the  $\min\Delta\phi(j, \cancel{E}_T)$  requirement has been loosened to 1 and a generator level lepton outside the detector acceptance for both electrons and muons is required.

and no other muons in the event. The number of events in this dimuon control region is then extrapolated to the signal region using efficiencies and cross-section ratios calculated using MC events. As in the prompt data analysis  $Z/\gamma^* \rightarrow \mu\mu$  MC with the leptons removed and a requirement that there is a generator level dimuon system with invariant mass between 80 and 100 GeV (i.e. compatible with a Z boson) is used to estimate the contribution in the signal region from  $Z \rightarrow \nu\nu$  processes.  $Z \rightarrow \nu\nu$  MC events are not used for this estimation due to the limited size of the available  $Z \rightarrow \nu\nu$  MC samples. Whilst the number of events in the dimuon control region is small the agreement between data and MC is good as can be seen from Figure 5.12

The formulas used to carry out the extrapolation from the control region to the signal region are given in Equation 4.3 which is repeated here for reference:

$$N_{Exp}^S = N_{Data}^C - N_{Bkg}^C \cdot \frac{\sigma(Z \rightarrow \nu\nu)}{\sigma(Z/\gamma^* \rightarrow \mu\mu)} \cdot \frac{\epsilon_{VBF}^S}{\epsilon_{VBF}^C}, \quad (5.7)$$

where  $\epsilon^S$  and  $\epsilon^C$  are calculated as in Equations 4.4 and 4.5 repeated here:

$$\epsilon_{VBF}^S = \frac{\sigma(Z \rightarrow \nu\nu, EWK) \frac{N_{MC}^S(EWK)}{N_{gen}(Zmass,EWK)} + \sigma(Z \rightarrow \nu\nu, QCD) \frac{N_{MC}^S(QCD)}{N_{gen}(Zmass,QCD)}}{\sigma(Z \rightarrow \nu\nu, EWK) + \sigma(Z \rightarrow \nu\nu, QCD)}, \quad (5.8)$$

$$\epsilon_{VBF}^C = \frac{\sigma(Z/\gamma^* \rightarrow \mu\mu, EWK) \frac{N_{MC}^C(EWK)}{N_{gen}(EWK)} + \sigma(Z/\gamma^* \rightarrow \mu\mu, QCD) \frac{N_{MC}^C(QCD)}{N_{gen}(QCD)}}{\sigma(Z/\gamma^* \rightarrow \mu\mu, EWK) + \sigma(Z/\gamma^* \rightarrow \mu\mu, QCD)}. \quad (5.9)$$

As the  $Z/\gamma^* \rightarrow \mu\mu$  and  $Z \rightarrow \nu\nu$  cross-sections are calculated before analysis selection and do not depend on detector calibration, the same values are used as in the prompt data analysis shown in Table 4.4. The other inputs to the above equations and the results of the estimation are shown in Table 5.6.

In the prompt data analysis one of the largest uncertainties, accounting for

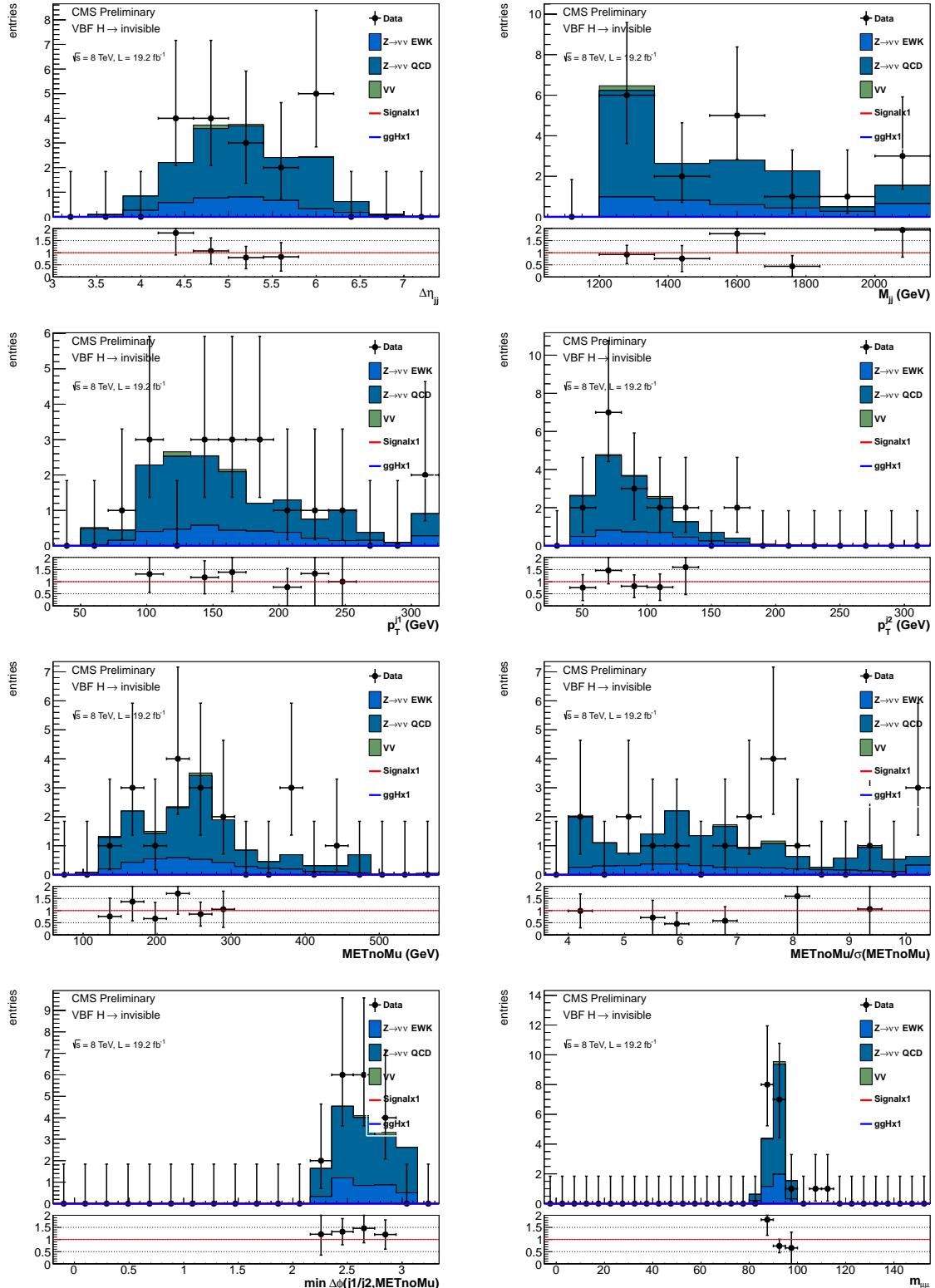


Figure 5.12.

**Table 5.6.:** The inputs to Equations 5.7, 5.8 and 5.9 and the final estimate of the  $Z \rightarrow \nu\nu$  background in the parked data analysis signal region. Also shown for comparison is the ratio between the MC prediction of the number of  $Z/\gamma^* \rightarrow \mu\mu$  events in the control region and the number of data events in the control region. This ratio is analogous to the data driven scale factor used in the  $W + \text{jets}$  background estimation, however due to the different efficiencies for the *EWK* and *QCD* components, it is not directly used in the  $Z \rightarrow \nu\nu$  estimation. The systematic uncertainties quoted include MC statistical uncertainties and all sources listed in Section 5.4.

$N_{gen} (\text{EWK})$	5781.9	
$N_{gen} (\text{Zmass, EWK})$	4226.5	
$N_{gen} (\text{QCD})$	22789000	
$N_{gen} (\text{Zmass, QCD})$	20334000	
	Signal region	Control region
$N_{Data}$	N/A	$18 \pm 4.2 \text{ (stat)}$
$N_{Bkg}$	N/A	$0.2 \pm 0.1 \text{ (MCstat)}$
$N_{MC} (\text{EWK})$	$7.9 \pm 0.2 \text{ (MCstat)}$	$6.0 \pm 0.2 \text{ (MCstat)}$
$N_{MC} (\text{QCD})$	$29.5 \pm 3.0 \text{ (MCstat)}$	$20.5 \pm 2.5 \text{ (MCstat)}$
$\frac{N_{Data}^C - N_{Bkg}^C}{N_{MC}^C (\text{EWK}) + N_{MC}^C (\text{QCD})}$	$0.67 \pm 0.16 \text{ (stat)} \pm 0.06 \text{ (MCstat)}$	
$N_{Z \rightarrow \nu\nu / Z \rightarrow \gamma^* \rightarrow \mu\mu}$	$158.1 \pm 37.8 \text{ (stat)} \pm 21.2 \text{ (syst)}$	$17.8 \pm 4.2 \text{ (stat)} \pm 0.1 \text{ (MCstat)}$

### 5.3.7. V+jets closure tests

### 5.3.8. V+jets scale factor investigations

### 5.3.9. QCD

### 5.3.10. Minor backgrounds

## 5.4. Systematic uncertainties

The uncertainty due to the reweighting process used to account for trigger inefficiencies cancels in all data driven background estimates, as a ratio of MC event yields is taken. To estimate the size of the uncertainty that should be applied to processes not taken from data driven background, the bin with the largest uncertainties on its fit for each era was chosen. It was then assumed that all bins had this worst-case uncertainty. This resulted in a 2.3% overall uncertainty on these processes from this effect. Given that this error is smaller than many of the other errors considered, and that the uncertainty on the efficiency in most of the fit bins is significantly lower than this worst case this uncertainty is considered negligible.

## 5.5. Results

# **Chapter 6.**

## **Combinations and interpretations of Run 1 searches for invisibly decaying Higgs bosons**

**6.1. Searches in other channels**

**6.2. Combination with prompt VBF search**

**6.3. Combination with the parked VBF search**

**6.4. Dark matter interpretations**

# **Chapter 7.**

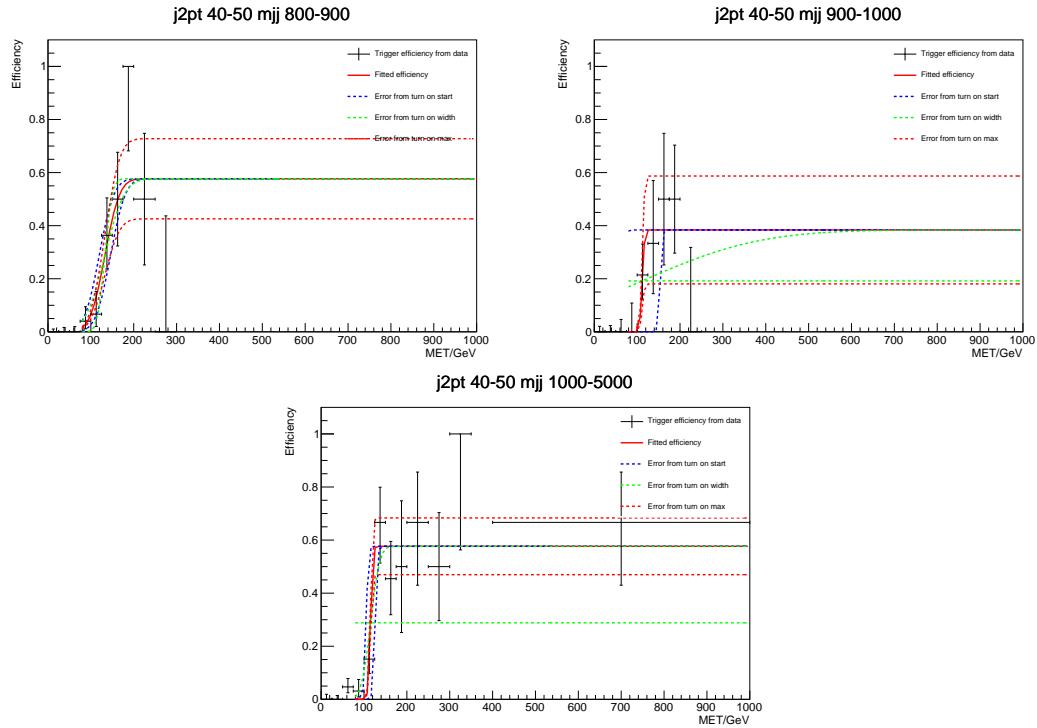
## **Search for invisibly decaying Higgs bosons in Run 2 data**

### **7.1. Trigger efficiency measurements**

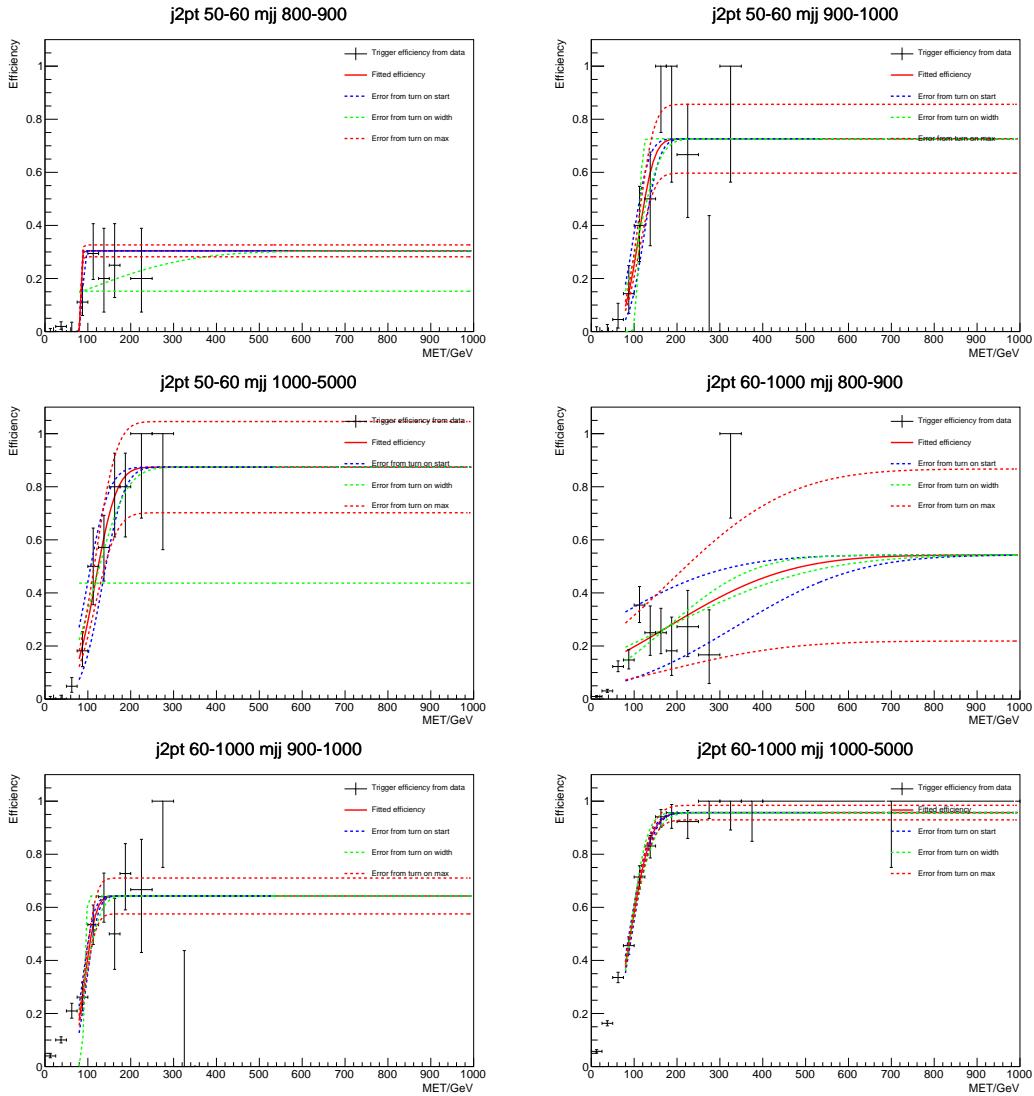
# Appendix A.

## Parked data trigger efficiencies

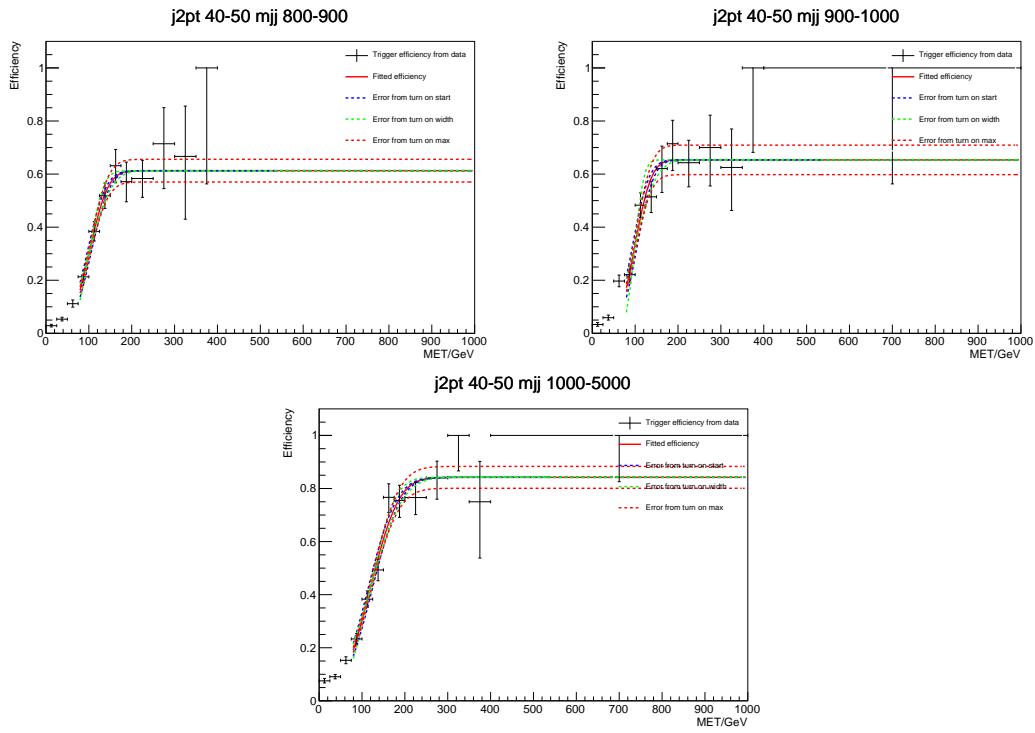
This appendix contains the trigger efficiency curves with overlaid error function fits and their errors as described in Section 5.1. Due to the event selection applied in the parked data analysis only the highest bin in  $M_{jj}$  is used.



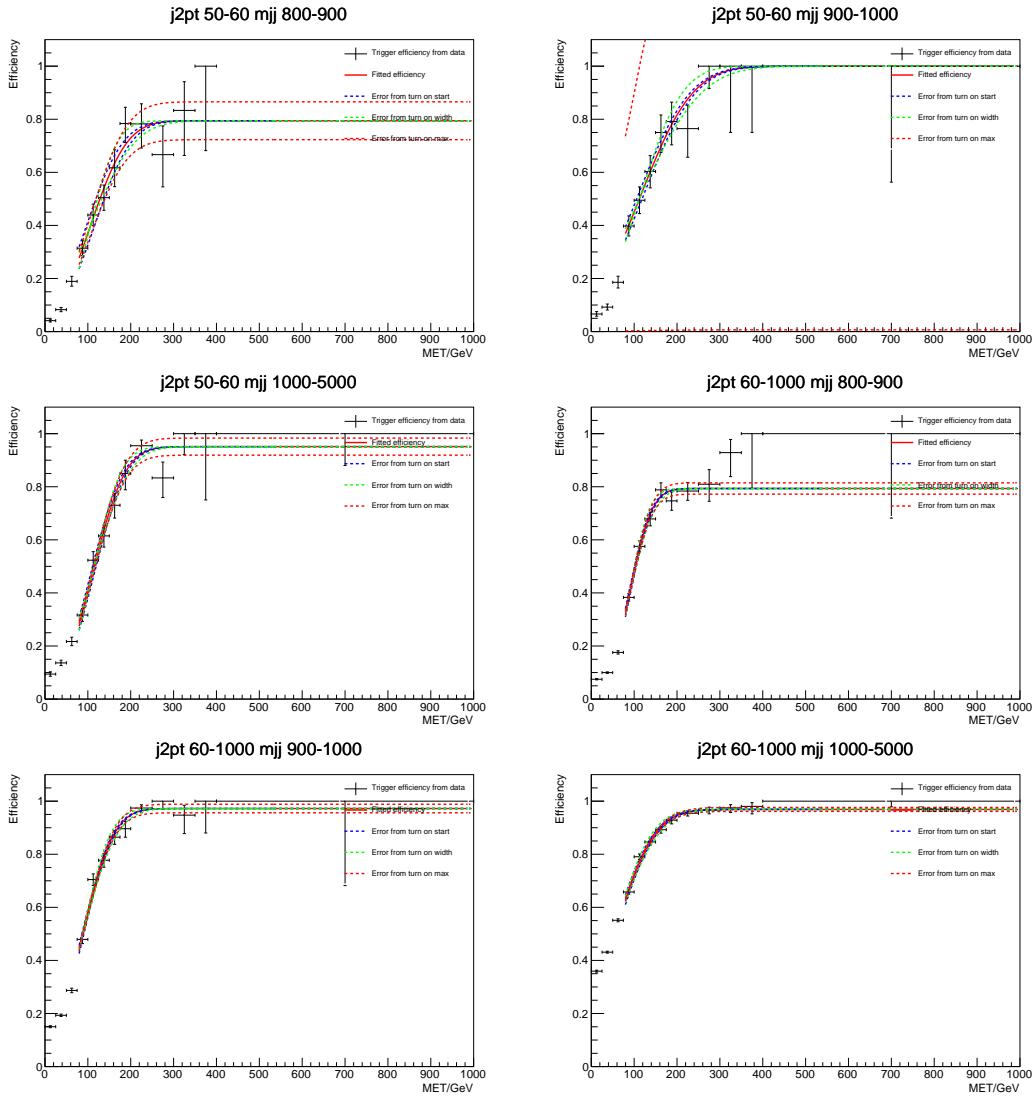
**Figure A.1.:** The measured efficiency of the trigger used in run A as a function of MET in bins of dijet mass ( $m_{jj}$ ) and sub-leading jet  $p_T$  (j2pt). The bin that each plot corresponds to is displayed at the top of the plot



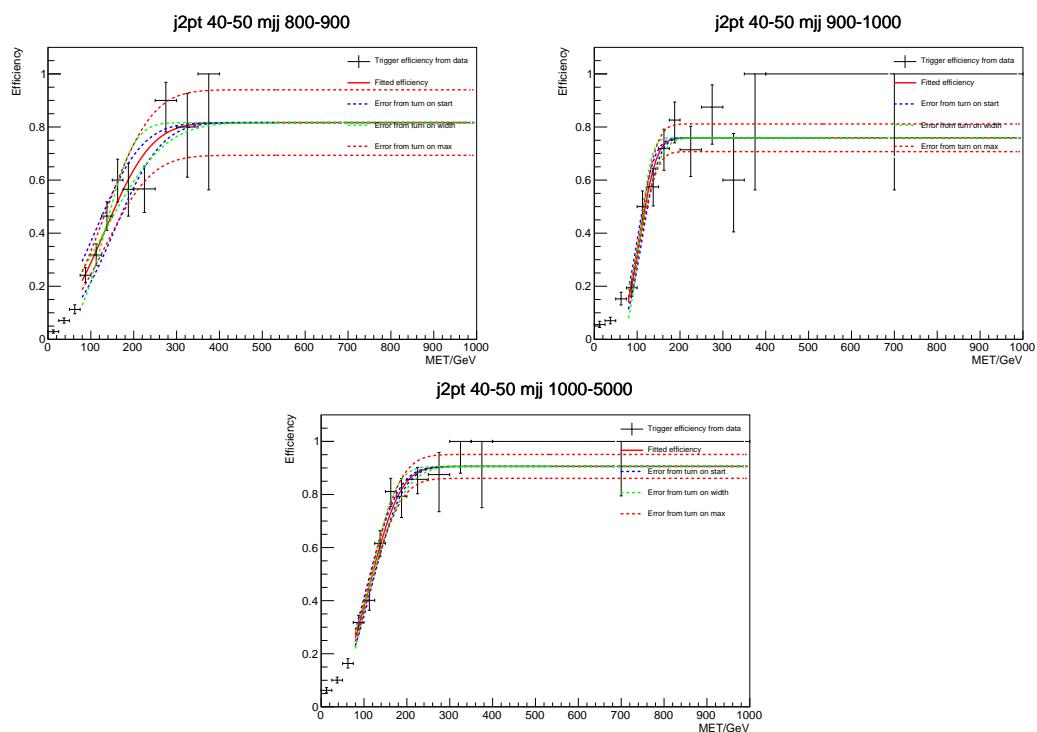
**Figure A.2.:** The measured efficiency of the trigger used in run A as a function of MET in bins of dijet mass (mjj) and sub-leading jet  $p_T$  (j2pt). The bin that each plot corresponds to is displayed at the top of the plot



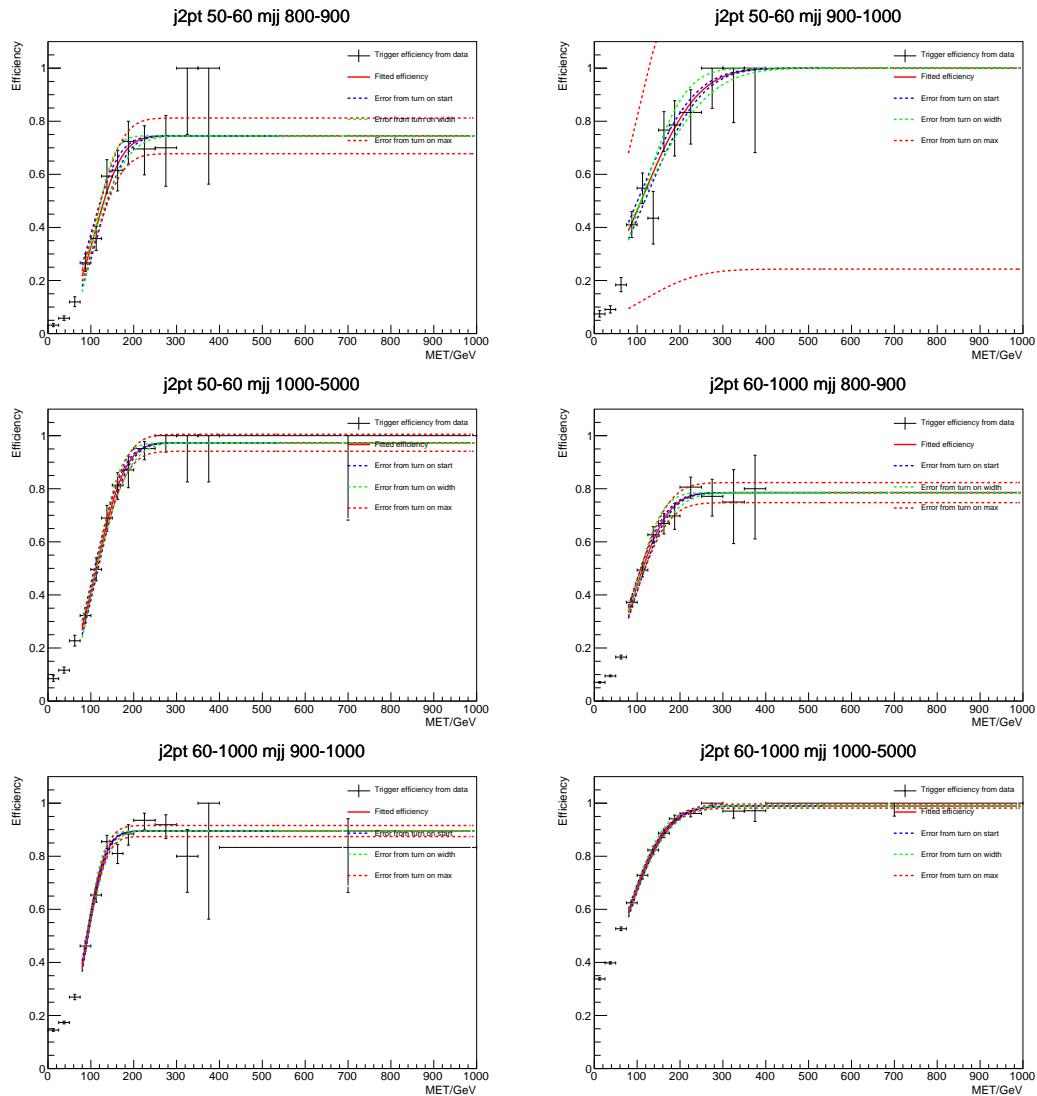
**Figure A.3.:** The measured efficiency of the trigger used in runs B and C as a function of MET in bins of dijet mass (mjj) and sub-leading jet  $p_T$  (j2pt). The bin that each plot corresponds to is displayed at the top of the plot



**Figure A.4.:** The measured efficiency of the trigger used in runs B and C as a function of MET in bins of dijet mass (mjj) and sub-leading jet  $p_T$  (j2pt). The bin that each plot corresponds to is displayed at the top of the plot



**Figure A.5.:** The measured efficiency of the trigger used in run D as a function of MET in bins of dijet mass (mjj) and sub-leading jet  $p_T$  (j2pt). The bin that each plot corresponds to is displayed at the top of the plot



**Figure A.6.:** The measured efficiency of the trigger used in run D as a function of MET in bins of dijet mass (mjj) and sub-leading jet  $p_T$  (j2pt). The bin that each plot corresponds to is displayed at the top of the plot



# Bibliography

- [1] D. Hanneke, S. Fogwell, and G. Gabrielse, “New Measurement of the Electron Magnetic Moment and the Fine Structure Constant”, *Phys. Rev. Lett.* **100** (Mar, 2008) 120801, [doi:10.1103/PhysRevLett.100.120801](https://doi.org/10.1103/PhysRevLett.100.120801).
- [2] E. Noether, “Invariant Variation Problems”, *Gott. Nachr.* **1918** (1918) 235–257, [doi:10.1080/00411457108231446](https://doi.org/10.1080/00411457108231446), [arXiv:physics/0503066](https://arxiv.org/abs/physics/0503066). [Transp. Theory Statist. Phys.1,186(1971)].
- [3] E. Noether, “Invariant variation problems”, *Transport Theory and Statistical Physics* **1** (1971), no. 3, 186–207, [doi:10.1080/00411457108231446](https://doi.org/10.1080/00411457108231446).
- [4] C. N. Yang and R. L. Mills, “Conservation of Isotopic Spin and Isotopic Gauge Invariance”, *Phys. Rev.* **96** (Oct, 1954) 191–195, [doi:10.1103/PhysRev.96.191](https://doi.org/10.1103/PhysRev.96.191).
- [5] Particle Data Group Collaboration, “Review of Particle Physics”, *Chin. Phys. C* **38** (2014) 090001, [doi:10.1088/1674-1137/38/9/090001](https://doi.org/10.1088/1674-1137/38/9/090001).
- [6] D. Griffiths, “Introduction to Elementary Particles”. Physics textbook. Wiley, 2008.
- [7] “Procedure for the LHC Higgs boson search combination in summer 2011”, Technical Report ATL-PHYS-PUB-2011-011, CMS NOTE 2011/005, CERN, Geneva, (Aug, 2011).
- [8] G. Cowan, K. Cranmer, E. Gross et al., “Asymptotic formulae for likelihood-based tests of new physics”, *Eur. Phys. J. C* **71** (2011) 1554, [doi:10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0), [10.1140/epjc/s10052-013-2501-z](https://doi.org/10.1140/epjc/s10052-013-2501-z), [arXiv:1007.1727](https://arxiv.org/abs/1007.1727). [Erratum: Eur. Phys. J.C73,2501(2013)].
- [9] C. Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004, [doi:10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [10] L. Evans and P. Bryant, “LHC Machine”, *JINST* **3** (2008), no. 08, S08001, [doi:10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001).

- [11] LEP Injector Study Group, “LEP design report, volume I: The LEP injector chain; LEP design report, volume II: The LEP Main Ring”. CERN, Geneva, 1983.
- [12] A. Collaboration, “The ALICE experiment at the CERN LHC”, *JINST* **3** (2008) S08002, [doi:10.1088/1748-0221/3/08/S08002](https://doi.org/10.1088/1748-0221/3/08/S08002).
- [13] A. Collaboration, “The ATLAS Experiment at the CERN Large Hadron Collider”, *JINST* **3** (2008) S08003, [doi:10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- [14] LHCb Collaboration, “The LHCb Detector at the LHC”, *JINST* **3** (2008) S08005, [doi:10.1088/1748-0221/3/08/S08005](https://doi.org/10.1088/1748-0221/3/08/S08005).
- [15] M. Benedikt, P. Collier, V. Mertens et al., “LHC Design Report”. CERN, Geneva, 2004.
- [16] CMS Collaboration, “CMS Luminosity - Public Results”,.
- [17] W. J. Stirling, “Private communication”,.
- [18] T. Sakuma and T. McCauley, “Detector and Event Visualization with SketchUp at the CMS Experiment”, *J. Phys.: Conf. Ser.* **513** 022032, [doi:10.1088/1742-6596/513/2/022032](https://doi.org/10.1088/1742-6596/513/2/022032).
- [19] CMS Collaboration, “Interactive Slice of the CMS detector”, *CMS Document 4172-v2*.
- [20] Tricomi, Alessia, “Performances of the ATLAS and CMS silicon tracker”, *Eur Phys J C* **33** (2004) s1023–s1025, [doi:10.1140/epjcd/s2004-03-1801-1](https://doi.org/10.1140/epjcd/s2004-03-1801-1).
- [21] CMS Collaboration Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *J. Instrum.* **9** (May, 2014) P10009. 80 p. Comments: Replaced with published version. Added journal reference and DOI.
- [22] CMS Collaboration Collaboration, “History of ECAL response with laser data”,.
- [23] USCMS, ECAL/HCAL Collaboration, “The CMS barrel calorimeter response to particle beams from 2-GeV/c to 350-GeV/c”, *Eur. Phys. J.* **C60** (2009) 359–373, [doi:10.1140/epjc/s10052-009-0959-5](https://doi.org/10.1140/epjc/s10052-009-0959-5), [10.1140/epjc/s10052-009-1024-0](https://doi.org/10.1140/epjc/s10052-009-1024-0). [Erratum: Eur. Phys. J.C61,353(2009)].
- [24] CMS Collaboration Collaboration, “CMS Physics: Technical Design Report Volume 1: Detector Performance and Software”. Technical Design Report CMS. CERN, Geneva, 2006. There is an error on cover due to a technical problem for some items.

- [25] CMS Collaboration Collaboration, “Data Parking and Data Scouting at the CMS Experiment”,.
- [26] T. C. Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *Journal of Instrumentation* **9** (2014), no. 10, P10009.
- [27] K. Rose, “Deterministic Annealing for Clustering, Compression, Classification, Regression, and related Optimization Problems”, *Proceedings of the IEEE* **86** (1998) 2210.
- [28] R. Frühwirth, W. Waltenberger, and P. Vanlaer, “Adaptive vertex fitting”, *J. Phys. G* **34** (2007) N343, [doi:10.1088/0954-3899/34/12/N01](https://doi.org/10.1088/0954-3899/34/12/N01).
- [29] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”, *CMS Physics Analysis Summary CMS-PAS-PFT-09-001* (2009).
- [30] CMS Collaboration, “Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector”, *CMS Physics Analysis Summary CMS-PAS-PFT-10-001* (2010).
- [31] CMS Collaboration, “Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV”, *CMS Physics Analysis Summary CMS-PAS-PFT-10-002* (2010).
- [32] S. Baffioni, C. Charlot, F. Ferri et al., “Electron reconstruction in CMS”, *Eur. Phys. J. C* **49** (2007) 1099–1116, [doi:10.1140/epjc/s10052-006-0175-5](https://doi.org/10.1140/epjc/s10052-006-0175-5).
- [33] C. collaboration, “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at s=8TeV”, *Journal of Instrumentation* **10** (2015), no. 06, P06005.
- [34] W. Adam, R. Frühwirth, A. Strandlie et al., “Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC”, *CMS Note CMS-NOTE-2005-001* (2005).
- [35] L. Sani, “Private communication”,.
- [36] CMS Collaboration, “Performance of CMS muon reconstruction in  $pp$  collision events at  $\sqrt{s} = 7$  TeV”, *JINST* **7** (2012) P10002, [doi:10.1088/1748-0221/7/10/P10002](https://doi.org/10.1088/1748-0221/7/10/P10002), [arXiv:1206.4071](https://arxiv.org/abs/1206.4071).

- [37] G. P. Salam, “Towards Jetography”, *Eur. Phys. J.* **C67** (2010) 637–686, [doi:10.1140/epjc/s10052-010-1314-6](https://doi.org/10.1140/epjc/s10052-010-1314-6), [arXiv:0906.1833](https://arxiv.org/abs/0906.1833).
- [38] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-  $k_t$  jet clustering algorithm”, *Journal of High Energy Physics* **2008** (2008), no. 04, 063.
- [39] M. Cacciari, G. P. Salam, and G. Soyez, “FastJet User Manual”, *Eur. Phys. J. C* **72** (2012) 1896, [doi:10.1140/epjc/s10052-012-1896-2](https://doi.org/10.1140/epjc/s10052-012-1896-2), [arXiv:1111.6097](https://arxiv.org/abs/1111.6097).
- [40] N. Saoulidou and E. Tziaferi, “Performance of the Particle-Flow jet identification criteria using proton-proton collisions at 8 TeV”, *CMS Analysis Note: AN-14-227* (2014).
- [41] CMS Collaboration, “Pileup Jet Identification”, *CMS Physics Analysis Summary CMS-PAS-JME-13-005* (2013).
- [42] A. Hocker, J. Stelzer, F. Tegenfeldt et al., “TMVA - Toolkit for Multivariate Data Analysis”, *PoS ACAT* (2007) 040, [arXiv:physics/0703039](https://arxiv.org/abs/physics/0703039).
- [43] CMS Collaboration, “Determination of jet energy calibration and transverse momentum resolution in CMS”, *JINST* **6** (2011) 11002, [doi:10.1088/1748-0221/6/11/P11002](https://doi.org/10.1088/1748-0221/6/11/P11002), [arXiv:1107.4277](https://arxiv.org/abs/1107.4277).
- [44] CMS Collaboration, “Performance of Missing Transverse Momentum Reconstruction Algorithms in Proton-Proton Collisions at  $\sqrt{s} = 8$  TeV with the CMS Detector”, *CMS Physics Analysis Summary CMS-PAS-JME-12-002* (2013).
- [45] Particle Data Group Collaboration, “Review of Particle Physics”, *Chin. Phys. C* **38** (2014) 090001, [doi:10.1088/1674-1137/38/9/090001](https://doi.org/10.1088/1674-1137/38/9/090001).
- [46] CMS Collaboration, “Performance of tau-lepton reconstruction and identification in CMS”, *JINST* **7** (2012) P01001, [doi:10.1088/1748-0221/7/01/P01001](https://doi.org/10.1088/1748-0221/7/01/P01001), [arXiv:1109.6034](https://arxiv.org/abs/1109.6034).
- [47] CMS Collaboration, “Search for invisible decays of Higgs bosons in the vector boson fusion and associated ZH production modes”, *Eur. Phys. J. C* **74** (2014) 2980, [doi:10.1140/epjc/s10052-014-2980-6](https://doi.org/10.1140/epjc/s10052-014-2980-6), [arXiv:1404.1344](https://arxiv.org/abs/1404.1344).
- [48] CMS Collaboration, “Summaries of CMS cross section measurements”, .
- [49] R. Aggleton, C. Asawatangtrakuldee, J. Brooke, O. Buchmueller, D. Colling, G. Davies, P. Dunne, S. Kumar, Q. Li, A. Magnan, K. Mazumdar, A. Nikitenko, J.

- Pela, P. Srimanobhas, D. Wang, “Search for an Invisible Higgs Boson”, *CMS Analysis Note: AN-12-403* (2012).
- [50] CMS Collaboration, “Top pair cross section in dileptons”,.
- [51] CMS Collaboration, “Measurement of the single-top t-channel cross section in pp collisions at centre-of-mass energy of 8 TeV”,.
- [52] CMS Collaboration, “Measurement of WW production rate”,.
- [53] CMS Collaboration, “Measurement of WZ production rate”,.
- [54] CMS Collaboration, “Measurement of the ZZ production cross section and anomalous trilinear gauge couplings in lll'l' decays at  $\sqrt{s} = 8$  TeV at the LHC”,.
- [55] CMS Collaboration, “Performance of  $\tau$ -lepton reconstruction and identification in CMS”, *J. Instrum.* **7** (Sep, 2011) P01001. 33 p.
- [56] CMS Collaboration, “Identification of b-quark jets with the CMS experiment”, *JINST* **8** (2013) P04013, [doi:10.1088/1748-0221/8/04/P04013](https://doi.org/10.1088/1748-0221/8/04/P04013),  
[arXiv:1211.4462](https://arxiv.org/abs/1211.4462).

# List of figures

2.1.	The layout of the chain of accelerators feeding into the LHC, showing the position of the four main detectors. . . . .	13
2.2.	A summary of the luminosity delivered to CMS during Run 1 of the LHC[16].	14
2.3.	Cross-sections for several processes in collisions of protons with protons or anti-protons as a function of centre of mass energy. The energies that the LHC and Tevatron ran at are highlighted[17]. . . . .	16
2.4.	Distribution of the number of interactions per bunch crossing in CMS during 2012 running of the LHC[16]. . . . .	17
2.5.	A diagram of the subsystems making up the CMS detector, illustrating the hermeticity and layered structure of the experiment[18]. . . . .	18
2.6.	A schematic cross-section of the CMS experiment showing the path taken by several types of particles[19]. . . . .	19
2.7.	A cross-section of the CMS tracker, indicating the subsystems that comprise it. Each line indicates a detector module[9]. . . . .	20
2.8.	A schematic of the CMS ECAL, indicating the subsystems that comprise it. The ECAL is 7.8m long by 3.5m wide[9]. . . . .	21
2.9.	A schematic of a quadrant of the CMS HCAL in the $r - z$ plane, indicating the subsystems that comprise it[9]. . . . .	24
2.10.	A schematic of a quadrant of the CMS muon system in the $r - z$ plane, indicating the subsystems that comprise it[24]. . . . .	26
2.11.	A schematic of the L1 trigger system. The arrows indicate the flow of data, the information transferred between systems is also indicated[9]. . .	27
3.1.	Total jet-energy-correction factor as a function of jet $\eta$ for jets with $p_T = 50\text{ GeV}$ (left) and $p_T = 200\text{ GeV}$ (right), for several types of jet reconstruction used at CMS. The bands indicate the corresponding uncertainty[43]. . . . .	39

3.2. Distributions of the uncorrected $\cancel{E}_T$ in $Z \rightarrow \mu\mu$ events (left) and $\gamma$ +jets events (right) in $\sqrt{s} = 8\text{TeV}$ data and simulation. The shaded band corresponds to the systematic uncertainty[44]. . . . .	40
4.1. The efficiency of the HLT requirements used in the prompt (blue) and parked (purple and red) data analyses as a function of the values of several offline variables, measured in a sample of events recorded on a single-muon trigger. (a) Efficiency as a function of offline $\Delta\eta_{jj}$ , (b) efficiency as a function of sub-leading jet $p_T$ , (c) efficiency as a function of offline $\cancel{E}_T^{\text{no-}\mu}$ , (d) efficiency as a function of $M_{jj}$ . . . . .	44
4.2. The combined efficiency of the HLT and L1 trigger requirements used in the prompt (blue) and parked (purple and red) data analyses as a function of the values of several offline variables, measured in a sample of events recorded on a single-muon trigger. (a) Efficiency as a function of offline $\Delta\eta_{jj}$ , (b) efficiency as a function of sub-leading jet $p_T$ , (c) efficiency as a function of offline $\cancel{E}_T^{\text{no-}\mu}$ , (d) efficiency as a function of $M_{jj}$ . . . . .	45
4.3. Distributions of (a) $M_{jj}$ , (b) $\Delta\eta_{jj}$ , (c) $\Delta\phi_{jj}$ and (d) leading central jet $p_T$ in background and signal MC events. The events shown are required to have two jets in opposite forward/backward halves of the detector with $p_T > 50$ GeV, $M_{jj} > 150$ GeV and $\cancel{E}_T^{\text{no-}\mu} > 130$ GeV. EWK refers to the V+jets and minor backgrounds described in Section 4.2 and QCD refers to the QCD multijet background. The dashed lines indicate the offline selection criteria applied to these variables, which are motivated in the text[47]. . . . .	48
4.4. Distributions of the visible W boson $p_T$ (i.e. the electron $p_T$ ) (a) and $\cancel{E}_T^{\text{no-}\mu}$ (b) in the single electron control region. WNJets+EWK indicates the W+jets contribution to this region, SingleT+TTBar indicates the contribution from top quark related processes, DYJets+EWK indicates the Z+jets contribution, QCD indicates the QCD multijet contribution and DiBoson indicates the two vector boson contribution. All of these contributions are estimated from MC. The hatched region illustrates the systematic uncertainty[49]. . . . .	50

4.5. Distributions of the visible W boson $p_T$ (i.e. the muon $p_T$ ) (a) and $\cancel{E}_T^{\text{no-}\mu}$ (b) in the single muon control region. WNJets+EWK indicates the W+jets contribution to this region, SingleT+TTBar indicates the contribution from top quark related processes, DYJets+EWK indicates the Z+jets contribution, QCD indicates the QCD multijet contribution and DiBoson indicates the two vector boson contribution. All of these contributions are estimated from MC. The hatched region illustrates the systematic uncertainty[49]. . . . .	52
4.6. Distributions of the the tau $p_T$ (a) and $\Delta\phi_{jj}$ (b) in the single tau control region. WNJets+EWK indicates the W+jets contribution to this region, ZJets indicates the contribution from Z+jets, SingleT+TTBar indicates the contribution from top quark related processes, QCD indicates the QCD multijet contribution and DiBoson indicates the two vector boson contribution. All of these contributions are estimated from MC. The hatched region illustrates the systematic uncertainty[49]. . . . .	54
4.7. Distributions of the $\cancel{E}_T^{\text{no-}\mu}$ (a) and $M_{jj}$ (b) in a relaxed Z control region, with no requirement on $\Delta\phi_{jj}$ , the CJV removed, and the requirements on $M_{jj}$ and $\Delta\eta_{jj}$ relaxed to 1000 GeV and 3.5 respectively. DY(ll)+jets indicates the contribution from Z+jets processes and $t\bar{t}$ , $tW$ , VV indicates the contribution from minor backgrounds. The hatched region illustrates the systematic uncertainty[47]. . . . .	57
4.8. A diagram of the regions used in the QCD multijet ABCD background estimation method. Region D is the signal region and regions A, B and C are mutually exclusive control regions. . . . .	58
4.9. Distribution of the $\cancel{E}_T^{\text{no-}\mu}$ for events passing and failing the CJV. The largest difference seen in any bin is 40%. This difference is used to assign as a systematic uncertainty on the QCD multijet background estimation. . . . .	59
4.10. The distribution of the sub-leading jet's $p_T$ in W+Jets MC events with the nominal JES and for JESUP and JESDOWN. . . . .	61

4.11. Distributions of the $\cancel{E}_T^{\text{no-}\mu}$ (a) and $M_{jj}$ (b) of events observed in data and expected from the background estimation methods described in Section 4.2 in the signal region. tt+DY+VV indicates the contribution from minor backgrounds. The hatched region illustrates the systematic uncertainty. The QCD multijet background is not shown due to the very low number of events in the MC samples. The cumulative effect of a signal from a Higgs boson with mass of 125 GeV, decaying 100% to invisible final states is also shown[47]. . . . .	66
4.12. Expected and observed 95% CL upper limits on the VBF $\sigma \times \mathcal{B}$ in pb (a) and normalised to the SM VBF Higgs boson production cross-section (b)[47]. The green and yellow bands are the one and two sigma uncertainty bands of the expected limit. . . . .	68
5.1. The efficiency for the trigger (color-scale) used in eras B and C (a) and era D (b), as a function of $M_{jj}$ and sub-leading jet $p_T$ for events with $\cancel{E}_T^{\text{no-}\mu}$ between 60 and 120 GeV. The efficiency was measured using a single muon dataset collected with an orthogonal trigger. . . . .	71
5.2. The results of performing a fit of the function in Equation 5.1 to the efficiency of the trigger used in era D, measured in a sample of single muon events collected with an orthogonal trigger. The dashed bands show the uncertainty on the fit due to each of the three parameters of the fit. The two bins of dijet mass ( $m_{jj}$ ) and sub-leading jet's $p_T$ ( $j2pt$ ) shown are those with the two highest numbers of events from the final signal region described in Section 5.2. . . . .	72
5.3. (a) $\mathcal{S}$ after the trigger driven selection described in Equation 5.3. (b) $\min\Delta\phi(j, \cancel{E}_T)$ after the trigger driven selection and requiring $\mathcal{S} > 3$ . (c) $M_{jj}$ after the trigger driven selection and requiring $\mathcal{S} > 3$ and $\min\Delta\phi(j, \cancel{E}_T) > 1$ . All three plots are of the signal-like region with the MC scaled using the background estimation methods described in Section 5.3. The disagreement between data and the predictions from background MC samples is believed to be due to mismeasured QCD multijet events which are not well modelled by the available MC samples. . . . .	75
5.4. From top to bottom left to right distributions of $\Delta\eta_{jj}$ , $M_{jj}$ , leading jet $p_T$ , sub-leading jet $p_T$ , $\cancel{E}_T^{\text{no-}\mu}$ , $\mathcal{S}$ and $\min\Delta\phi(j, \cancel{E}_T)$ for events passing the full preselection. No QCD multijet contribution is shown, which accounts for the difference between the data observation and background prediction. . . . .	76

5.5. Matrices of correlation coefficients for several variables in signal (left) and V+jets background (right) events passing the signal region selection. The variables are 1) the azimuthal angle difference between the $\cancel{E}_T^{\text{no-}\mu}$ and the unclustered energy in the event, 2) the square root of the hadronic energy in the event, 3) $\mathcal{S}$ , 4) $\cancel{E}_T^{\text{no-}\mu}$ , 5) $M_{jj}$ , 6) the number of jets with $p_T > 30$ GeV between the two tag jets in $\eta$ , 7) the vectorial sum of the tag jets $p_T$ and the $\cancel{E}_T^{\text{no-}\mu}$ , 8) the ratio between the magnitude of the vectorial sum of the tag jet's $p_T$ and the $\cancel{E}_T^{\text{no-}\mu}$ . . . . .	78
5.6. The distribution of $\min\Delta\phi(j, \cancel{E}_T)$ (left) and $M_{jj}$ (right) in the top control region with one tight electron and one tight muon. . . . .	80
5.7. The distribution of $\min\Delta\phi(j_{1/2}, \cancel{E}_T)$ (left) and $M_{jj}$ (right) in the single top control region. . . . .	81
5.8. Distributions of variables in data and MC events in the $W \rightarrow e\nu$ control region. MC events from V+jets backgrounds are scaled by their data-driven scale factors. The variables shown are from top to bottom and left to right $\Delta\eta_{jj}$ , $M_{jj}$ , the leading and subleading jet's $p_T$ , $\cancel{E}_T^{\text{no-}\mu}$ , $\mathcal{S}$ and $\min\Delta\phi(j, \cancel{E}_T)$ . . . . .	83
5.9. Distributions of variables in data and MC events in the $W \rightarrow \mu\nu$ control region. MC events from V+jets backgrounds are scaled by their data-driven scale factors. The variables shown are from top to bottom and left to right $\Delta\eta_{jj}$ , $M_{jj}$ , the leading and subleading jet's $p_T$ , $\cancel{E}_T^{\text{no-}\mu}$ , $\mathcal{S}$ and $\min\Delta\phi(j, \cancel{E}_T)$ . . . . .	85
5.10. Distributions of variables in data and MC events in the $W \rightarrow e\nu$ control region. MC events from V+jets backgrounds are scaled by their data-driven scale factors. The variables shown are from top to bottom and left to right $\Delta\eta_{jj}$ , $M_{jj}$ , the leading and subleading jet's $p_T$ , $\cancel{E}_T^{\text{no-}\mu}$ , $\mathcal{S}$ , $\min\Delta\phi(j_{1/2}, \cancel{E}_T)$ and the tau- $\cancel{E}_T$ system's transverse mass. . . . .	87
5.11. Distributions of the number of jets with $p_T > 15$ GeV (a), $\min\Delta\phi(j, \cancel{E}_T)$ (b) and the $p_T$ of the leading generator level electron/muon in $W \rightarrow e/\mu\nu$ events in a region with cuts which are the same as those of the signal region except that the $\min\Delta\phi(j, \cancel{E}_T)$ requirement has been loosened to 1 and a generator level lepton outside the detector acceptance for both electrons and muons is required. . . . .	90
5.12. . . . .	92

A.1. The measured efficiency of the trigger used in run A as a function of MET in bins of dijet mass ( $m_{jj}$ ) and sub-leading jet $p_T$ ( $j2pt$ ). The bin that each plot corresponds to is displayed at the top of the plot . . . . .	97
A.2. The measured efficiency of the trigger used in run A as a function of MET in bins of dijet mass ( $m_{jj}$ ) and sub-leading jet $p_T$ ( $j2pt$ ). The bin that each plot corresponds to is displayed at the top of the plot . . . . .	98
A.3. The measured efficiency of the trigger used in runs B and C as a function of MET in bins of dijet mass ( $m_{jj}$ ) and sub-leading jet $p_T$ ( $j2pt$ ). The bin that each plot corresponds to is displayed at the top of the plot . . . . .	99
A.4. The measured efficiency of the trigger used in runs B and C as a function of MET in bins of dijet mass ( $m_{jj}$ ) and sub-leading jet $p_T$ ( $j2pt$ ). The bin that each plot corresponds to is displayed at the top of the plot . . . . .	100
A.5. The measured efficiency of the trigger used in run D as a function of MET in bins of dijet mass ( $m_{jj}$ ) and sub-leading jet $p_T$ ( $j2pt$ ). The bin that each plot corresponds to is displayed at the top of the plot . . . . .	101
A.6. The measured efficiency of the trigger used in run D as a function of MET in bins of dijet mass ( $m_{jj}$ ) and sub-leading jet $p_T$ ( $j2pt$ ). The bin that each plot corresponds to is displayed at the top of the plot . . . . .	102

# List of tables

1.1.	The fundamental fermions observed in nature separated into their three generations. Each particle shown also has an antiparticle with opposite charge and identical mass. Values taken from[5] . . . . .	4
1.2.	The fundamental vector bosons observed in nature separated by the force which they mediate. Values taken from[5]. . . . .	4
4.1.	The inputs to, and results of, the $W \rightarrow e\nu$ background estimation. $N_{W \rightarrow e\nu}$ is, for the signal region the number of events expected from $W \rightarrow e\nu$ backgrounds, and for the control region the number of events remaining in the region after the subtraction of other backgrounds. . . . .	49
4.2.	The inputs to, and results of, the $W \rightarrow \mu\nu$ background estimation. $N_{W \rightarrow \mu\nu}$ is, for the signal region the number of events expected from $W \rightarrow \mu\nu$ backgrounds, and for the control region the number of events remaining in the region after the subtraction of other backgrounds. . . . .	51
4.3.	The inputs to, and results of, the $W \rightarrow \tau\nu$ background estimation. $N_{W \rightarrow \tau\nu}$ is, for the signal region the number of events expected from $W \rightarrow \tau\nu$ backgrounds, and in the control region the number of events remaining in the region after the subtraction of other backgrounds. . . . .	53
4.4.	The input variables for the calculation of $\epsilon_{VBF}^S$ and $\epsilon_{VBF}^C$ using Equations 4.4 and 4.5 respectively. . . . .	56
4.5.	The inputs to, and results of, the $Z \rightarrow \nu\nu$ background estimation using Equation 4.3. $\epsilon_{VBF}$ in the signal (control) region is calculated using Equation 4.4 (4.5). $N_{Z \rightarrow \nu\nu}/N_{Z/\gamma^* \rightarrow \mu\mu}$ is in the signal region the number of events expected from $Z \rightarrow \nu\nu$ backgrounds, and for the control region the number of events remaining in the region after the subtraction of other backgrounds. The systematic uncertainties are calculated as described in Section 4.3. . . . .	57
4.6.	Numbers of events from data and MC in each region used in the QCD multijet background estimate and the final estimated number of events. .	59

4.7.	A summary of the uncertainties in the total expected signal and background yields. All uncertainties are quoted as the percentage change in the yield when each effect is varied up and down according to its uncertainty. The signal yields assume a Higgs boson mass of 125 GeV . . . . .	65
4.8.	The estimated numbers of background and signal events, together with the observed yield, in the signal region. The signal yield assumes a Higgs boson mass of 125 GeV and $\mathcal{B}(H \rightarrow \text{inv}) = 100\%$ . . . . .	65
5.1.	A summary of the requirements of the triggers used for this analysis in each of LHC Run 1's eras. For the jet requirements all triggers require that there is at least one pair of jets in the event satisfying all of the jet requirements listed in this table. All requirements are on HLT variables unless stated otherwise. . . . .	70
5.2.	The inputs to and results of Equation 5.6, when used to estimate the $W \rightarrow e\nu$ estimate in the signal region. . . . .	82
5.3.	The inputs to and results of Equation 5.6, when used to estimate the $W \rightarrow \mu\nu$ estimate in the signal region. . . . .	84
5.4.	The inputs to and results of Equation 5.6, when used to estimate the $W \rightarrow \tau\nu$ estimate in the signal region. . . . .	86
5.5.	The numbers of events predicted by MC in the two subregions of the signal region with a generator level lepton that is inside/outside the detector acceptance for both electrons and muons from $W \rightarrow e\nu$ and $W \rightarrow \mu\nu$ processes. The errors shown are from MC statistics. . . . .	89
5.6.	The inputs to Equations 5.7, 5.8 and 5.9 and the final estimate of the $Z \rightarrow \nu\nu$ background in the parked data analysis signal region. Also shown for comparison is the ratio between the MC prediction of the number of $Z/\gamma^* \rightarrow \mu\mu$ events in the control region and the number of data events in the control region. This ratio is analogous to the data driven scale factor used in the $W + \text{jets}$ background estimation, however due to the different efficiencies for the $EWK$ and $QCD$ components, it is not directly used in the $Z \rightarrow \nu\nu$ estimation. The systematic uncertainties quoted include MC statistical uncertainties and all sources listed in Section 5.4. . . . .	93

# List of Acronyms

**VBF** vector boson fusion

**ggH** gluon fusion

**DM** dark matter

**CL** confidence level

**PSB** Proton Synchrotron Booster

**PS** Proton Synchrotron

**SPS** Super Proton Synchrotron

**PU** pile-up

**SM** standard model

**BSM** beyond the SM

**QFT** quantum field theory

**ECAL** electromagnetic calorimeter

**HCAL** hadron calorimeter

**EB** ECAL barrel

**EE** ECAL endcap

**HB** hadron barrel

**HE** hadron endcaps

**HF** hadron forward

**HO** hadron outer

**L1** Level-1

**HLT** high-level trigger

**CSC** cathode strip chamber

**DT** drift tube

**RPC** resistive plate chamber

**WLCG** Worldwide LHC Computing Grid

**PV** Primary Vertex

**CTF** combinatorial track finder

**DA** “deterministic annealing”

**PF** Particle flow

**GSF** Gaussian sum filter

**BDT** boosted decision tree

**MC** Monte Carlo

**HPS** hadron plus strips

**JES** jet energy scale

**EM** electromagnetic

**CJV** central jet veto

**MVA** multi-variate analysis

**JER** jet energy resolution

**CSV** combined secondary vertex

**pdf** probability density function

**QCD** Quantum Chromodynamics