BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

SEARCH FOR INVISIBLE DECAYS OF THE HIGGS BOSON PRODUCED VIA VECTOR BOSON FUSION AT LHC WITH THE CMS DETECTOR RUN-II DATA

by

ALP AKPINAR

B.Sc., Bogazici University, 2018

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2023

Approved by

First Reader

Zeynep Demiragli, PhD
Assistant Professor of Physics

Second Reader

James Rohlf, PhD
Professor of Physics

All things are difficult before they are easy.
- Dr. Thomas Fuller

Acknowledgments

My acknowledgments will go here.

SEARCH FOR INVISIBLE DECAYS OF THE HIGGS BOSON PRODUCED VIA VECTOR BOSON FUSION AT LHC WITH THE CMS DETECTOR RUN-II DATA

ALP AKPINAR

Boston University, Graduate School of Arts And Sciences, 2023

Major Professor: Zeynep Demiragli, PhD
Assistant Professor of Physics

ABSTRACT

There is multiple astrophysical evidence which support the presence of dark matter (DM), which stands out as the one of the open questions in the Standard Model (SM) of Particle Physics. One avenue to look for DM production is using particle colliders such as the Large Hadron Collider (LHC), where the production of DM can be detected as events with large missing transverse momentum (p_T^{miss}). This thesis documents a search for new DM particles using proton-proton collision data from LHC, which is recorded by the Compact Muon Solenoid (CMS) detector, at a center of mass energy of 13 TeV. In this search, the target signature is a Higgs boson, produced via Vector Boson Fusion (VBF) process, decaying into a pair of DM particles, resulting in two energetic jets and large p_T^{miss} in the final state. To estimate the background processes, multiple control regions are defined and a simultaneous fit to data over all regions is performed. The data for this search was collected in 2017 and 2018, during Run II of the LHC. A full result corresponding to an integrated luminosity of $137 fb^{-1}$ is also obtained by statistically combining this analysis result with the already published 2016 analysis. No excess of events is observed, compared

to the expected SM backgrounds. The results are interpreted in the context of Higgs-portal models, where upper bounds are set to $BR(H \rightarrow inv.)$, the branching ratio for the SM Higgs boson decaying to invisible DM particles.

Contents

List of Tables

List of Figures

Chapter 1

Theoretical Background

1.1 A brief history of particle physics

The understanding of the fundamental building blocks of the universe, which is known as particle physics, has undergone a remarkable journey over the course of human history. The idea that all different forms of matter is composed of elementary particles is believed to date back to 6th century BC. These elementary particles were termed as "atoms", which originates from the Greek word "atomos" meaning "indivisible". This view of matter is called *atomism*, and it was argued that if it was possible to divide matter into smaller blocks repeatedly, it would be then possible to reduce matter to nothing. Hence, fundamental building blocks of matter, atoms, were necessary. Experimental evidence for the atomic nature of matter started to emerge finally in 19th century, when in 1815, English chemist William Prout noticed that the atomic masses of many chemical elements were multiples of the mass of hydrogen, the lightest known element. Hence, Prout hypothesized that all matter was built from hydrogen, suggesting hydrogen is the fundamental building block of matter. Although not being very accurate in today's understanding of atoms, this was an important step (and the first significant experimental assertion) towards the understanding of atoms as fundamental building blocks of matter.

In the early 20th century however, it started to become apparent that atoms are not indivisible solid entities themselves. J.J. Thomson discovered electrons (1897), which were fundamental particles carrying negative electric charge [21]. Shortly after,

α-particle scattering experiments by E. Rutherford et. al. (1909) [22] revealed that at the center of each atom, a densely packed nucleus must be located, carrying positive electric charge. From the results of the scattering experiments, Rutherford was able to deduce that the size of the nucleus R_n must be $< 10^{-14}$ m. Surrounding this dense nucleus, orbits of negatively charged electrons are located. It soon became clear that the densely packed nucleus is not an indivisible object either, but is composed of positively charged protons [23] and electrically neutral neutrons [24].

This model of atoms in which the negatively charged electrons are orbiting the nucleus has been challenged once again in 1920s when quantum mechanics (QM) was developed, which completely changed the understanding of how particles behave in atomic scales. The development of QM started at the turn of the 20th century. In an attempt to explain the photoelectric effect experiment by Hertz, A. Einstein postulated that electromagnetic waves are composed of indivisible energy quanta, whose energy depends on the frequency of the wave, $E = h\nu$ (1905) [25]. Here, h refers to the Planck's constant (introduced in 1900 by M. Planck) and ν is the frequency of the electromagnetic wave. This understanding of the electromagnetic waves composed of identical particles (now called "photons") with small energy quanta led the way to the wave-particle duality, where each particle can be associated with a "matter wave"[1]. E. Schrodinger (1926) [27] came up with a wave equation that describes the time evolution of this matter wave, denoted as $\Psi(\vec{x}, t)$, which is shown in Eq. 1.1.

$$i\hbar\frac{\partial\Psi(\vec{x},t)}{\partial t} = \left(-\frac{\hbar^2}{2m}\nabla^2 + V(\vec{x},t)\right)\Psi(\vec{x},t) \tag{1.1}$$

where $\hbar = \frac{h}{2\pi}$ and $V(\vec{x}, t)$ describes the potential. It is worthy of note that Eq. 1.1 is a linear differential equation, meaning a linear combination of solutions $\{\Psi_i\}$ is itself also a solution. The physical meaning of $\Psi(\vec{x}, t)$ however, was not immediately

[1]The hypothesis that each particle has an associated wave with a wavelength inversely proportional to their momentum, $\lambda = h/p$, is attributed to De Broglie (1924) [26].

obvious from Eq. 1.1, it was eventually interpreted by M. Born (1926) [28] such that $|\Psi|^2$ is a probability density function that gives the probability to measure the particle at a volume V at time $t = t_0$:[2].

$$P(t = t_0) = \int_V d^3x \; |\Psi(\vec{x}, t = t_0)|^2 \tag{1.2}$$

Hence, with the advent of QM, determinism in the measurement of physical quantities of particles was lost. One can only predict probabilities of mesaurements using the wave function Ψ, which can be obtained from Eq. 1.1. It should be noted that this is a very significant change of perspective (a "paradigm shift") compared to classical physics where each particle can be described by a deterministic set of position and momenta coordinates $\{\vec{x}_i, \vec{p}_i\}$.

It soon became apparent that there was a problem with Eq. 1.1, namely that it was not compatible with special relativity, which was proposed by A. Einstein (1905) [29] and described how the laws of physics are the same in different frames of reference which move with constant velocities with respect to each other. The incompatibility can be immediately observed by noticing the second-order space derivatives and the first-order time derivative in Eq. 1.1, which shows that the differential equation treats space and time coordinates differently. P. Dirac (1928) [30] managed to write a differential equation with both first-order space and time derivatives, and is compatible with the energy-momentum relationship from special relativity, $E^2 = p^2c^2 + m^2c^4$. This led to the Dirac equation, shown in Eq. 1.3.

$$(i\hbar\gamma^\mu\partial_\mu - mc)\,\psi = 0 \tag{1.3}$$

[2]It should be noted that this interpretation of $|\Psi|^2$ is not trivial and imposes important constraints on possible forms of wave functions. $\Psi(x,t)$ must satisfy the condition $N(t) = \int_{-\infty}^{\infty} |\Psi(x,t)|^2 dx = 1$ for all t. It can be shown that if Ψ satisfies $\lim_{x\to\pm\infty} \Psi(x,t) = 0$, and $\lim_{x\to\pm\infty} \frac{\partial\Psi(x,t)}{\partial x} = K$ where $|K| < \infty$, $\frac{dN(t)}{dt} = 0$ under the time evolution of Ψ dictated by Eq. 1.1. Hence, if $N(t_0) = 1$, $N(t) = 1$ for all $t > t_0$. The argument above can be generalized to three space dimensions as well.

As with later equations in this thesis, Einstein summation convention is assumed in Eq. 1.3, where it is implied that the repeated indices (μ) are summed over. The γ-matrices in Eq. 1.3 are a set of four 4×4 matrices which satisfy the anti-commutation relation

$$\{\gamma^{\mu}, \gamma^{\nu}\} = 2g^{\mu\nu} \tag{1.4}$$

where $g^{\mu\nu}$ is the Minkowski metric tensor. Dirac equation is a remarkable result and it has very important implications. It predicts an internal property of particles called spin, which refers to the intrinsic angular momentum of a particle and has been experimentally confirmed [30]. In addition, Eq. 1.3 has solutions with negative energies, which was not immediately understood at the time[3]. Ultimately, it was understood that those solutions described antiparticles, which are counterparts of particles with the opposite electric charge [4]. In 1932, about four years later, positron (antiparticle of electron) was the first antiparticle to be discovered [33]. It is worthy of note that by combining principles of QM and special relativity, Eq. 1.3 was able to predict these very fundamental physics. Combination of QM with special relativity ultimately resulted in the development of Quantum Field Theory (QFT), which provides the mathematical framework of how we understand particles and their interactions to date.

By the beginning of 1960s, larger number of particles including pions and kaons were discovered. Initially, these particles were all thought to be fundamental in nature [34]. However, as the number of known particles grew larger and larger, it came into question whether there could be underyling elementary particles that make up all

[3]It is interesting to note that the negative energy electron was initially thought to be corresponding to a proton, for example see [31]. However, Dirac falsified this hypothesis in 1930 [32].

[4]The physical motivation behind this interpretation can be seen in this explanation by Dirac himself [32]: "... an electron with negative energy moves in an external field as though it carries a positive charge."

of these observed particles. This led Gell-Mann and Zweig to introduce quarks [35], which are proposed to be elementary particles making up the observed particles. Gell-Mann and Zweig proposed three types of quarks: Up (u), down (s) and strange (s) quarks, together with their antiquark counterparts. The particles made up by combining a quark and an antiquark are called mesons, while the particles made from combining three quarks are termed baryons. Using this approach, Gell-Mann and Zweig were able to explain all the known baryons and mesons, they even predicted new particles which would be discovered later [34]. Going further in time, three more types of quarks (and corresponding antiquarks) were measured in addition to u, d and s: The charm quark c in 1974 [36], the bottom (or beauty) quark b in 1977 [37] and finally the top quark t in 1995 [38].

The current understanding of these particles and their interactions are described by a group of quantum field theories, which is called the Standard Model (SM). SM describes the electromagnetic, weak and strong nuclear interactions between these particles. A brief overview of the SM will be given in the next section.

1.2 The Standard Model of particle physics

The Standard Model (SM) of particle physics describes different classes of known particles and their interactions with each other, using QFT as the mathematical framework. It was developed in the second half of the 20th century and has been very successful in predicting a wide variety of experimental observations. The development of the SM was led by Chen Ning Yang, Robert Mills, Sheldon Glasgow, Steven Weinberg, and Abdus Salam.

In the SM, the strong interaction is derived by requiring Eq. 1.3 to be invariant under SU(3) local phase transformations. The conserved quantum number of this symmetry is called color, and the quantum field theory describing the color interaction

of quarks and gluons is called Quantum Chronodynamics (QCD). The SU(3) local phase transformations on a spinor ψ can be written in the following form:

$$\psi(x) \rightarrow \psi'(x) = \exp\left[ig_S\vec{\alpha}(x)\cdot\hat{\mathbf{T}}\right]\psi(x) \tag{1.5}$$

where $\hat{\mathbf{T}} = \{T^a\}$ are the eight generators of the SU(3) symmetry group and $\alpha^a(x)$ are eight functions of the space-time coordinate x. g_S is the coupling constant for the QCD interaction, which describes the strength of the interaction. The invariance of Eq. 1.3 under local SU(3) transformations can be asserted by introducing eight new fields $G^a_\mu(x)$, the index $a = 1, \ldots, 8$ corresponding to one of the eight generators of the SU(3) symmetry, which physically correspond to the eight possible gluon states in QCD[5].

The QCD interaction between quarks and gluon fields G^a_μ can be written in terms of the Lagrangian formalism. The gluon field strength tensor, $G^i_{\mu\nu}$ can be defined as

$$G^i_{\mu\nu} = \partial_\mu G^i_\nu - \partial_\nu G^i_\mu + g_S f^{ijk} G^j_\mu G^k_\nu \tag{1.6}$$

where f^{ijk} refer to the fine structure constants of the SU(3) group, which can be defined by using the commutators of the generators of SU(3) group $\{T^i\}$ using the following relation: $[T^i, T^j] = if^{ijk}T^k$. Using the gluon field strength tensor defined in Eq. 1.6, the Lagrangian for the QCD interaction can be written as

$$\mathcal{L}_{QCD} = \sum_\psi \left[i\gamma^\mu\bar{\psi}\left(\partial_\mu - ig_S G^a_\mu T^a\right)\psi\right] - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a \tag{1.7}$$

$\mathcal{L}_{QCD}$ characterizes the interaction between the gluon fields G^a_μ and quark spinors ψ. The sum over ψ indicates the sum over the spinors of quarks from different flavors.

[5]The gauge invariance is asserted given that the gluon fields G^k_μ transform as $G^{k'}_\mu = G^k_\mu - \partial_\mu\alpha_k - g_S f_{ijk}\alpha_i G^j_\mu$, where f_{ijk} are the fine structure constants of the SU(3) group. The last term arises due to generators of SU(3) symmetry not commuting with each other, and gives rise to gluon self-interactions.

Adjoint spinor $\bar{\psi}$ for each type of quark is defined as $\bar{\psi} = \psi^\dagger \gamma^0$. γ^μ are the set of four 4×4 matrices from Eq. 1.3.

The electroweak sector of the SM describes the electromagnetic and weak interactions between particles in a unified manner. The electroweak interaction can be deduced by requiring the invariance of Eq. 1.3 under SU(2) x U(1) local phase transformations. This invariance can be asserted by introducing four new fields, W_μ^a ($a = 1, 2, 3$) and B_μ. The force carrier bosons of the weak interaction, W^+, W^-, Z and the force carrier boson for the electromagnetic interaction, γ, can be then expressed in terms of these four fields. Fields for the $W^\pm$ bosons can be expressed as

$$W_\mu^\pm = \frac{1}{\sqrt{2}} \left(W_\mu^1 \pm i W_\mu^2 \right) \tag{1.8}$$

In addition, fields for Z and γ bosons can be written in terms of a mixing between W_μ^3 and B_μ fields, as shown below in Eq. 1.9.

$$\begin{pmatrix} \gamma_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \tag{1.9}$$

where θ_W is the weak mixing angle. Experimental measurements provide $\sin^2\theta_W = 0.23101 \pm 0.00053$ [39]. Similar to gluon field strength tensor $G_{\mu\nu}^i$, one can write the field strength tensors for these four fields as $W_{\mu\nu}^i$ and $B_{\mu\nu}$ as

$$\begin{aligned} W_{\mu\nu}^i &= \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g_W \epsilon^{ijk} W_\mu^j W_\nu^k \\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu \end{aligned} \tag{1.10}$$

The extra term in $W_{\mu\nu}^i$ should be noted to be arising from the non-commuting nature of SU(2) generators. The Lagrangian describing the interactions of W_μ^a and B_μ fields can then be written as

$$\mathcal{L}_g = -\frac{1}{4} W^a_{\mu\nu} W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \tag{1.11}$$

The full electroweak Lagrangian describing the electroweak interactions of W^a_μ and B_μ fields with fermion fields can be in turn written as

$$\mathcal{L}_f = \sum_\psi \bar{\psi} \gamma^\mu \left(i\partial_\mu - g' \frac{1}{2} Y_W B_\mu - g \frac{1}{2} \sigma_a W^a_\mu \right) \psi \tag{1.12}$$

In Eq. 1.12, ψ refers to the SU(2) doublets of chirally left-handed particles and SU(2) singlets of right-handed particles, highlighting the unequal coupling of W^i_μ and B_μ fields to left-handed and right-handed fermions. This is also a crucial fact that is experimentally verified by C. S. Wu in 1956 [40]. Finally, the full electroweak Lagrangian $\mathcal{L}_{EW}$ can be written as the sum of Eqs. 1.11 and 1.12:

$$\mathcal{L}_{EW} = \sum_\psi \bar{\psi} \gamma^\mu \left(i\partial_\mu - g' \frac{1}{2} Y_W B_\mu - g \frac{1}{2} \sigma_a W^a_\mu \right) \psi - \frac{1}{4} W^a_{\mu\nu} W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \tag{1.13}$$

It should be noted that in Eq. 1.13, there are no mass terms for the W and Z bosons (in the form $W_\mu W^\mu$) which are known to be massive. However, inclusion of such terms in the Lagrangian would break the SU(2) x U(1) symmetry, hence the fermions and gauge bosons have to acquire mass through a different mechanism. In the SM, this mechanism is called the Higgs mechanism, which introduces a new scalar field called Higgs field, which can be written as a doublet as shown below:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{1.14}$$

With the Higgs doublet ϕ defined in Eq. 1.14, the Higgs-sector Lagrangian can be written as

$$\mathcal{L}_{Higgs} = \left[\left(\partial_\mu - i\frac{g}{2}W^a_\mu\sigma_a - ig'Y_\phi B_\mu\right)\phi\right]^2 + \mu^2\phi^\dagger\phi - \lambda\left(\phi^\dagger\phi\right)^2 \tag{1.15}$$

In the unitarity gauge, one can set $\phi^+ = 0$ and make ϕ^0 real in the doublet of Eq. 1.14, hence writing ϕ as $\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix}0, & v + h(x)\end{pmatrix}$. Then, $\langle\phi^0\rangle = v$ is the non-zero vacuum expectation value of the Higgs field, and the fluctuations around $\psi_0 = v$ (i.e. $h(x)$) describes a new boson (i.e. the Higgs boson). Writing Eq. 1.15 with ϕ in the unitary gauge, quadratic terms in W_μ and B_μ arise, giving mass to W and Z bosons:

$$M_W = \frac{1}{2}vg, \qquad M_Z = \frac{1}{2}v\sqrt{g^2 + g'^2} \tag{1.16}$$

The mass of the Higgs boson itself is then given by $M_H = \sqrt{2\mu^2} = \sqrt{2\lambda v^2}$. Through Yukawa couplings between fermions and the Higgs field, it can be shown that fermions also acquire mass due to the non-zero expectation value v of the Higgs field[6]. In this way, the observed masses of the weak bosons ($W^\pm, Z$) and fermions can be explained successfully with the SM.

In summary, the Standard Model (SM) can explain the strong and electroweak interactions between particles through the use of QFT as the mathematical framework. All the known particles in the SM and their interactions are summarized in Fig. 1·1. SM has been proven to be a very successful theory, predicting a variety of physics phenomena that has been observed in particle physics experiments over the last decades. The detection of $W^\pm$ and Z bosons [41] was a crucial test of the electroweak sector of the SM. In fact, the mass of these bosons as predicted by the SM, $M_W = 80357 \pm 6$ MeV and $M_Z = 91188 \pm 2$ MeV, are in very good agreement with the experimentally measured values [42, 43]. In addition, the Higgs boson (H) (predicted

[6]Writing the Higgs doublet as $\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix}0\\ v + h(x)\end{pmatrix}$, it can be shown that Lagrangian acquires terms $\mathcal{L}_f = -m_f\bar{f}f - \frac{m_f}{v}\bar{f}fh$. Here, the first term is the mass term, which represents the coupling of the fermion f to the Higgs field through the non-zero expectation value of the Higgs field. The second term represents the coupling of f to the Higgs boson itself.

by the Higgs mechanism) was observed in 2012 by ATLAS and CMS experiments [44], and the measured properties of the Higgs boson have so far been consistent with the SM predictions [45].

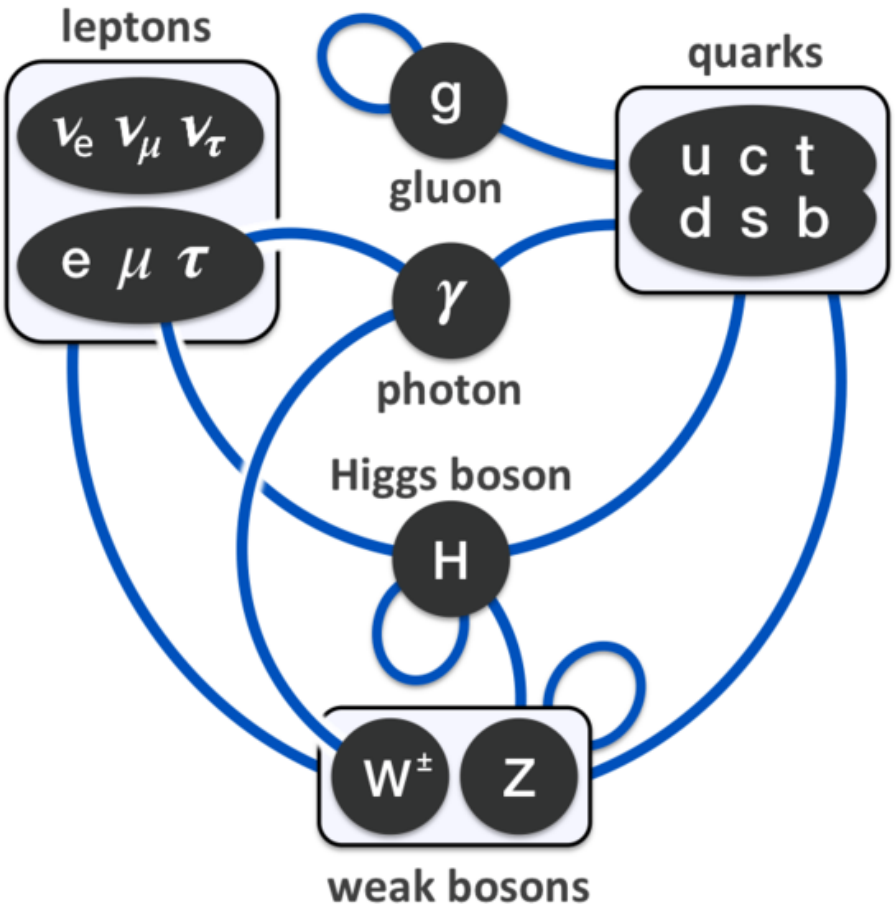

Figure 1·1: Schematic illustrating all the known particles in the SM and their interactions. Connections between different particles indicate that they interact. For the weak bosons ($W^{\pm}, Z$), gluons and the Higgs boson, self-interactions are also shown. Diagram is drawn by Eric Drexler.

1.3 Beyond the Standard Model: Dark matter

Despite its triumphs, Standard Model (SM) fails to explain many astrophysical observations [3]. One very important question that is closely related to this thesis is the existence of dark matter (DM), which refers to a form of matter that does not interact electromagnetically, and therefore neither emits or reflects light. As will be explained within this section, strong evidence exists for the existence of DM, however SM does not have any explanation about this potential form of matter. The first subsection below aims to give a brief overview of the evidence for the existence of DM. The following subsection covers searches of DM in particle collider experiments.

1.3.1 Experimental evidence for dark matter

Evidence for the existence of DM emerged from astrophysical observations within the 20th century. In 1933, F. Zwicky [46] observed that galaxies in the Coma cluster are moving faster than what is expected, requiring a larger amount of gravitational force to keep them in their orbits. Since then, the same phenomena has been observed within individual galaxies in 1970s [3]. This can be observed by looking at the *rotation curves* of the galaxies, namely the graph of circular velocities of stars and gas as a function of distance from the galactic center. One such rotation curve, for the NGC 6503 galaxy, is shown in Fig. 1·2.

As can be observed from Fig. 1·2, observed rotation curves usually exhibit a flat behavior at large distances from the galaxy center. As shown in Fig. 1·2, the contributions from the visible disk and gas alone cannot explain the observed rotation curves. From Newtonian dynamics, the circular velocity can be written as

$$v(r) = \sqrt{\frac{GM(r)}{r}} \tag{1.17}$$

where $M(r) = 4\pi \int \rho(r) r^2 dr$ and $\rho(r)$ is the mass density profile. Assumming

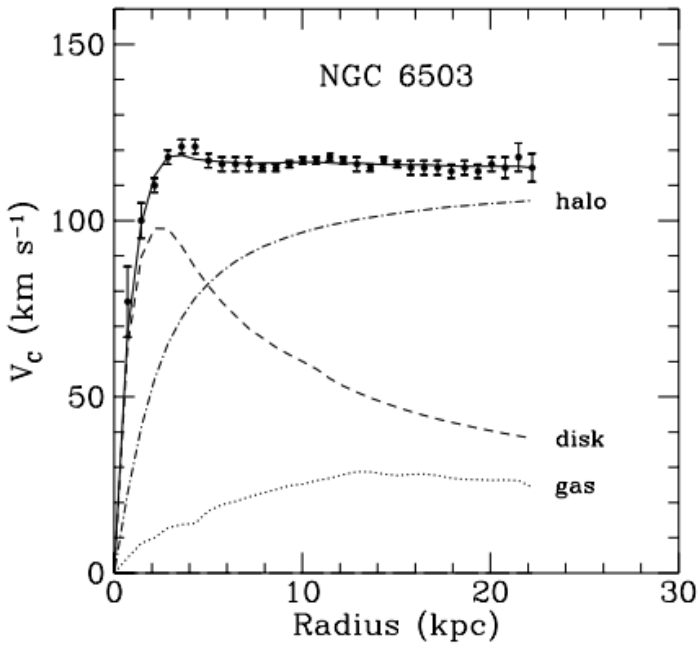

Figure 1·2: Rotation curve for NGC 6503 galaxy. The dashed and dotted lines represent the expected velocity profile from the disk (visible matter) and gas contributions, while the dash-dotted line represennts the halo (dark matter) contribution, required to explain the observed rotation curve. Taken from [3].

visible matter, $v(r) \propto \frac{1}{\sqrt{r}}$, hence the fact that the rotation curves are approximately flat implies the existence of DM with $M(r) \propto r$ [3].

Another piece of evidence for the existence of DM comes from the gravitational lensing effect [47], which refers to the phenomena where the trajectory of light bends in the vicinity of a massive object ("deflector"), due to the curvature in space-time near the deflector. The dependence of this bending on the mass density of the deflector implies that gravitational lensing can probe the mass of deflectors, making it a unique tool to probe the DM distribution in gravitational systems. In other words, the amount of deflector mass can be measured from the amount of deflection in the trajectory of light. The amount of bending can be deduced by measuring the angular shift of the image source $\Delta\vec{\theta}$. The amount of lensing effect can also be extracted from the distortion of the image source. As an example, if the image source is a circular galaxy, to first approximation it can be shown that the gravitational lensing effect transforms it into an ellipse [47]. Using the gravitational lensing technique, the presence of large non-visible matter clusters was inferred, for example see [48].

Evidence on the existence of DM is also present on the galactic scale, coming from the precise measurements of the Cosmic Microwave Background (CMB) radiation [49]. CMB is the electromagnetic radiation originating from the propagation of photons in the early Universe, when photons decoupled from baryons. Decades of experimental effort show that the CMB follows the spectrum of a blackbody radiation with $T = 2.726$ K, and it is nearly isotropic, with temperature fluctuations reaching only one part in 10^5 of the mean [3]. The residual anisotropies in the CMB can be studied to extract information about the abundance of baryons and matter in the Universe. This is achieved by fitting a given cosmological model with a fixed number of parameters, N, and finding the best-fit parameters from the peak of N-dimensional likelihood. Using the data from Wilkinson Microwave Anisotropy Probe (WMAP), one can obtain

the following values for the baryonic matter density $\Omega_b h^2$ and total matter density $\Omega_H h^2$ [3]:

$$\Omega_b h^2 = 0.024 \pm 0.01 \qquad \Omega_M h^2 = 0.14 \pm 0.02 \tag{1.18}$$

where h is Planck's constant. Therefore, it can be inferred that the DM relic density is $\Omega_{DM} h^2 \approx 0.12$, indicating that predictions based on WMAP data imply that DM is more abundant compared to baryonic matter by a factor of ≈ 6. In summary, multiple pieces of evidence, both at galactic and cosmological scales, suggest the existence of DM.

1.3.2 Searches for dark matter

In an effort to explain the DM observations outlined in Sec. 1.3.1, multiple hypotheses have been proposed, where the DM is made out of particles that are not described by the Standard Model (very commonly called "Beyond the Standard Model", or "BSM" for short). There are a wide range of models, where proposed types of particles include axions and weakly interacting massive particles (WIMPs) [50].

A WIMP refers to a new elementary particle which interacts via gravity[7]. Such particles are readily predicted by popular extension of the Standard Model, including supersymmetry [51], models with extra dimensions [52] and little Higgs [53]. In the scenario where DM is made up of WIMPs, it is possible to detect these new particles using a number of techniques in particle detectors.

One such technique is direct detection. The direct detection experiments look for signals from elastic scattering of DM particles with ordinary matter. These experiments typically use detectors made out of liquid noble gases or crystals, which are sensitive material which can generate detectable signals (such as scintillation light)

[7]It is possible that WIMPs interact with forces that are not described by the Standard Model, which is as weak as or weaker than the weak nuclear force.

in the event of a WIMP passing through. Such experiments include Xenon1T [54], CRESST-II [55], CDMSlite [56], LUX [57] and PandaX-II [58]. At the time of writing, no evidence has been observed of DM particles through direct detection experiments. However, very stringent constraints on the DM-nucleuon interaction cross section have been placed as a function of hypothesized WIMP mass. XENON1T was able to reach an upper bound of $\sigma_{DM-nucleon} < 7.7 \times 10^{47} cm^{-2}$ for a WIMP mass of $m_{DM} = 35$ GeV [54].

Another technique for WIMP detection is the indirect detection. Indirect detection experiments look for evidence of WIMP self-annihilation or decay products, such as high-energy photons or neutrinos. Hence, these experiments look for secondary products coming from the WIMP self-interactions, as opposed to attempting to directly observe signals from a WIMP itself. One such experiment is Fermi-LAT [59], which is a space-based instrument that detects gamma rays. It searches for an excess of gamma ray emission that might indicate WIMP self-annihilation or decay. Another such experiment is IceCube [60], which is a neutrino observatory in the South Pole. IceCube can detect high-energy neutrinos produced by cosmic-ray interactions, and observation of an excess of neutrinos can point to WIMP self-annihilation and decay processes. Similar to direct detection results, no evidence of WIMPs has been observed to date, but stringent limits are placed in WIMP self-annihilation cross sections.

Finally, another technique for WIMP searches is to produce them in particle detectors, such as the Large Hadron Collider (LHC)[8]. With a very large proton-proton collision center of mass energy, $\sqrt{s} = 13$ TeV[9], many WIMP models predict that pair production of WIMPs would be within the kinematic reach, provided that there is

[8]Please see Sec. 2.1 for the description of the LHC.

[9]The center of mass energy for the pp collision data (2016-2018) used in the analysis described in this thesis is $\sqrt{s} = 13$ TeV. However, it should be noted that starting from 2022 data-taking, this center of mass energy has been increased further to $\sqrt{s} = 13.6$ TeV.

some interaction between the SM sector and WIMPs. Hence, at particle colliders, one can expect to observe the process $pp \rightarrow \chi\bar{\chi} + X$, where χ refers to the WIMP (and $\bar{\chi}$ being it's antiparticle), and X is a placeholder for other physics objects in the final state, depending on the model being considered.

One interesting class of models that is relevant to the analysis described in this thesis is the so called Higgs portal models [5]. Such models hypothesize that there are Beyond the SM interactions of the Higgs boson with the WIMPs (commonly also called "dark sector"), and proton-proton colliders are excellent machines to probe such interactions of the Higgs boson through its decay products, i.e. $H \rightarrow \chi\bar{\chi}$, because of the large rates of Higgs bosons being produced. These models are discussed in more detail in the following section, where Sec. 1.4.1 gives an overview of the theory, and Sec. 1.4.2 gives an overview of the experimental signatures expected in proton-proton collisions.

1.4 Higgs portal models

The discovery of the Standard Model (SM) Higgs boson by ATLAS and CMS experiments [44] has paved the way for a new direction in the searches for DM, in which one can explore the potential connections between the Higgs boson and the dark sector. Dark sector refers to a Beyond the Standard Model (BSM) sector that contains the dark matter (DM) particle, and is almost decoupled from the SM. An interesting possibility is that the Higgs boson acting as a portal between the SM and dark sectors through couplings to both sectors. There are both experimental and theoretical arguments that make this an interesting possibility. Experimentally, the Higgs sector is far less explored and constrained, compared to the other sectors of the SM. Theoretically, Higgs field is the only field that allows to write down a renormalizable coupling to the dark sector, if DM is uncharged under the SM gauge group [5]. Therefore,

introducing DM-Higgs boson couplings is a well-motivated extension of the SM, and would allow the Higgs boson to decay into DM particle pairs, $\chi\bar{\chi}$. An example of this process is shown in Fig. 1·3, where the SM Higgs boson is produced via the Vector Boson Fusion (VBF) process, and decays into $\chi\bar{\chi}$ final state.

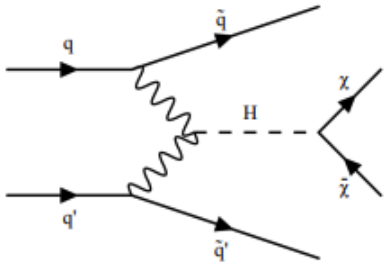

Figure 1·3: VBF production of a SM-like Higgs boson, where Higgs boson decays into a pair of DM particles, $\chi\bar{\chi}$, in the context of Higgs portal models. Diagram is taken from [4].

.

Given that the nature of the dark sector is mostly unknown, there are a multiple number of scenarios for the $H\chi\bar{\chi}$ coupling, where the DM particle χ can be a scalar, a vector or a Majorana fermion [61]. A brief overview of the theory of Higgs portal models under these different scenarios is provided in Sec. 1.4.1, and the experimental implications in particle colliders are discussed in Sec. 1.4.2.

1.4.1 Theory

This subsection aims to give a brief overview of the Higgs portal theories for different scenarios, following the footsteps of [61]. For the scenarios under which the DM particle is a scalar, vector or a Majorana fermion the corresponding Lagrangian terms are

$$\begin{aligned}
\mathcal{L}_S &= -\frac{1}{2}m_S^2 S^2 - \frac{1}{4}\lambda_S S^4 - \frac{1}{4}\lambda_{HSS} H^\dagger H S^2 \\
\mathcal{L}_V &= -\frac{1}{2}m_V^2 V_\mu V^\mu - \frac{1}{4}\lambda_V (V_\mu V^\mu)^2 + \frac{1}{4}\lambda_{HVV} H^\dagger H V_\mu V^\mu \\
\mathcal{L}_f &= -\frac{1}{2}m_f \bar{\chi}\chi - \frac{1}{4}\frac{\lambda_{Hff}}{\Lambda} H^\dagger H \bar{\chi}\chi
\end{aligned} \tag{1.19}$$

where the last terms represent the coupling between the DM-particle and SM-like Higgs boson, with $\lambda_{H\chi\bar{\chi}}$ being the coupling constants for each case. It should be noted that the $H\chi\bar{\chi}$ coupling is not renormalizable in the fermionic case. The S^4 and $(V_\mu V^\mu)^2$ are self-interaction terms for scalar and vector DM particles, respectively. The first terms in each Lagrangian correspond to the mass term for the DM particle.

After electroweak symmetry breaking, the doublet field H is written as $\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix}0, & v + h(x)\end{pmatrix}$ with $v = 246$ GeV being the vacuum expectation value of the Higgs field. In that case, the physical masses of the DM particles can be written as [61]

$$\begin{aligned}
M_S^2 &= m_S^2 + \frac{1}{2}\lambda_{hSS} v^2 \\
M_V^2 &= m_V^2 + \frac{1}{2}\lambda_{hVV} v^2 \\
M_f &= m_f + \frac{1}{2}\frac{\lambda_{hff}}{\Lambda} v^2
\end{aligned} \tag{1.20}$$

Provided that the DM particle χ is light enough for the $H \to \chi\bar{\chi}$ decay to occur (i.e. $M_\chi < M_H/2$), there will be a non-zero branching fraction for this decay of the Higgs boson, i.e. $B(H \to \chi\bar{\chi}) = \frac{\Gamma_{\chi\bar{\chi}}}{\Gamma_H} \neq 0$. The decay width for $H \to \chi\bar{\chi}$ can be written as follows for the different scenarios being considered [61]

$$\begin{aligned}\Gamma_{h\to SS} &= \frac{\lambda_{hSS}^2 v^2 \beta_S}{64\pi m_H} \\ \Gamma_{h\to VV} &= \frac{\lambda_{hVV}^2 v^2 m_h^3 \beta_V}{256\pi M_V^4}\left(1 - 4\frac{M_V^2}{m_h^2} + 12\frac{M_V^4}{m_h^4}\right) \\ \Gamma_{h\to ff} &= \frac{\lambda_{hff}^2 v^2 m_h \beta_f^3}{32\pi\Lambda^2}\end{aligned} \tag{1.21}$$

where $\beta_X = 1 - \sqrt{1 - 4M_X^2/m_h^2}$. Therefore, observing these decays predicted by the Higgs portal models in particle colliders would be an important leap towards understanding BSM physics. Observing these decays in particle colliders is the topic of the next subsection.

1.4.2 Experimental signatures in colliders

Under the assumption of the DM particle (χ) being a WIMP candidate, it is assumed that χ does not have electromagnetic interactions. That would imply that in particle detectors like ATLAS and CMS, these particles would pass through without interacting with the detector components. Therefore, the event of a $H \to \chi\bar{\chi}$ decay can be inferred by a large amount of transverse momentum imbalance in the event, i.e. large p_T^{miss} (provided that there is visible energy coming from other final state particles). Because of this, such decays are often referred to as "invisible decays of the Higgs boson", and very commonly denoted as $H \to inv$. This thesis will follow the same notation.

Such decays can be probed using proton-proton collisions by targeting events with the main production modes of the Higgs boson. Feynman diagrams for these production modes are shown in Fig. 1·4. A very nice overview of all the production modes, together with their cross sections and kinematics are given in [62].

At the Large Hadron Collider (LHC), with a proton-proton collision center of mass energy of $\sqrt{s} = 13$ TeV, the dominating production mode is the gluon-gluon

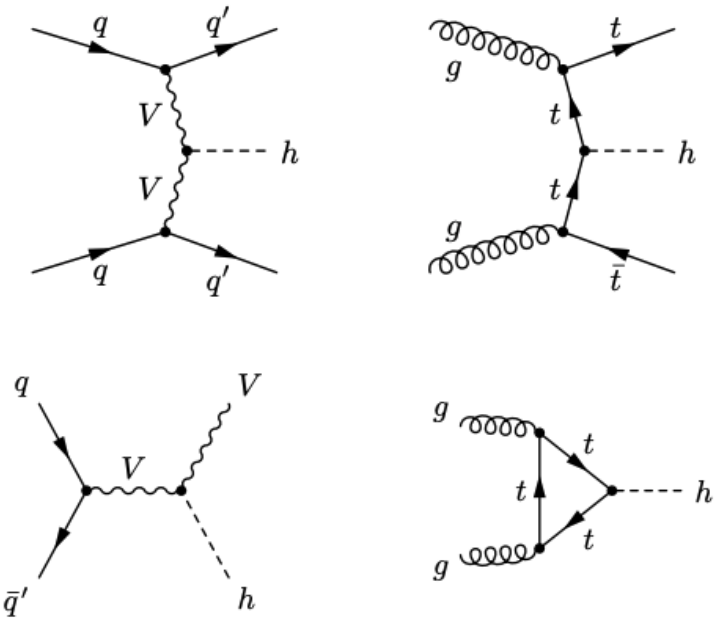

Figure 1·4: Main production modes of the Higgs boson at a proton-proton collider. The production modes are vector boson fusion (top left), production in association with a pair of top quarks (top right), in association with a vector boson (bottom left), and via gluon-gluon fusion (bottom right). Diagrams are taken from [5].

fusion. Events from this production mode can be probed by targeting final states with an energetic hadronic jet[10] coming from initial state radiation, together with large p_T^{miss} due to the invisible decay of the Higgs boson. This is the so called "monojet" search, while there was no evidence observed for such events, this channel provided important constraints on the invisible branching ratio of the Higgs boson, $B(H \rightarrow inv.) < 27.8\%$ [63]. The associated production with a vector boson or a top-antitop quark pair have smaller cross sections, but still are important final states to probe Higgs portal models, see for example [64].

Search for $H \rightarrow inv.$ decays through the remaining production mode, vector boson fusion (VBF), is the topic of this thesis. VBF has a lower production cross section, which is roughly an order of magnitude lower than the gluon-gluon fusion at $\sqrt{s} = 13$ TeV. But, it is still a very sensitive channel in the searches of new physics. This is due to the uniqueness of the experimental signature of the final state, which comprises of two energetic jets with large separation in the detector, and a large p_T^{miss} because of the invisible decay of the Higgs. These two jets typically have small scattering angles with respect to the proton-proton collision axis[11] , hence making them more forward in the detector and well-separated from each other. This unique final state signature of the VBF process can be used to reject a large amount of SM backgrounds, hence making this channel very sensitive to signal. In fact, the previous $H \rightarrow inv.$ searches done by collider experiments showed that VBF is the most sensitive channel, resulting

[10]Hadronic jets arise from the hadronization of the underlying final state quarks. Since individual quarks carry color charge, they cannot be observed individually due to the hypothesis of color confinement, which arises from the gluon-gluon self interactions. On a mathematical basis, this is explained by the $g_S f^{ijk} G^j_\mu G^k_\nu$ terms appearing in Eq. 1.6, which arise due to the non-commuting nature of SU(3) generators. Hence, jets are the corresponding experimental observables in particle colliders.

[11]A detailed explanation for this can be found in [62]. It can be shown that the matrix element $\mathcal{M}$ is bounded by $1/(p_T^2 + M_V^2)$, where p_T is the transverse momentum of the outgoing quarks, and M_V is the mass of the vector boson. Hence, while the final state quarks are relatively energetic, the VBF cross-section is suppressed for $p_T \gg M_V$, making the quarks (and hence the resulting final state jets from hadronization) more forward.

in the tightest constraints put on the $\mathcal{B}(H \to \text{inv})$, see for example [65]. Hence, it is experimentally very motivating to continue the VBF $H \to inv.$ search with the new data coming in from the Large Hadron Collider (LHC).

The rest of this thesis is structured as follows. Chapter 2 will give an overview of the experimental apparatus used to search for VBF $H \to inv.$ decays, namely the Large Hadron Collider (LHC) and the Compact Muon Solenoid (CMS) detector. Chapter 3 will then describe the analysis strategy, highlighting the analysis selections, the maximum likelihood fit procedure and systematic uncertainties considered in the analysis. Chapter 4 will describe the results of the analysis and it's physics interpretations. Finally, Chapter 5 will give perspective on the future direction of the CMS detector and the VBF $H \to inv.$ analysis, outlining upgrades on the detector and the triggering system, and the search for new analysis methodologies using novel machine learning techniques.

Chapter 2

Experimental Apparatus

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC), located near Geneva, Switzerland, is a superconducting circular particle accelerator, consisting of a 27-kilometer ring of superconducting magnets. For the majority of the time, protons are used as the accelerated particles, but in addition to protons, Xe and Pb ions are also used for heavy ion studies. For the context of this thesis, data from only the proton collisions are being considered.

Two sets of proton bunches travel along the circular tunnel in opposite directions, being accelerated to near the speed of light, then collide at the experiment sites. The four large experiments located at the LHC ring are CMS, ATLAS, LHCb and ALICE. CMS and ATLAS are multi-purpose detectors, while LHCb focuses on heavy hadron physics, and ALICE focuses on heavy ion physics. The center-of-mass energy for the proton-proton collision for Run II is 13 TeV. This energy is further increased to 13.6 TeV in more recent runs of the LHC, corresponding to Run III.

Protons are accelerated through a chain of accelerators before they reach LHC, which is shown in Fig. 2·1. Each machine boosts the energy of protons before injecting it into the next machine in the sequence. Linear Accelerator 4 (LINAC4) is the source of protons, accelerating H^- ions to 160 MeV to prepare them to enter the Proton Synchotron Booster (PSB). During the injection to PSB, the two ions are stripped from the H^- ions, leaving only the protons. These protons are accelerated to 2

GeV for injection to Proton Synchotron (PS), which increases the beam energy to 26 GeV. The proton beam is then sent to Super Proton Synchotron (SPS), where it is accelerated to an energy of 450 GeV. Finally, the proton beam is transferred to the two beam pipes of the LHC. Here, it takes 20 minutes for the protons to reach their maximum energy of 6.5 TeV.

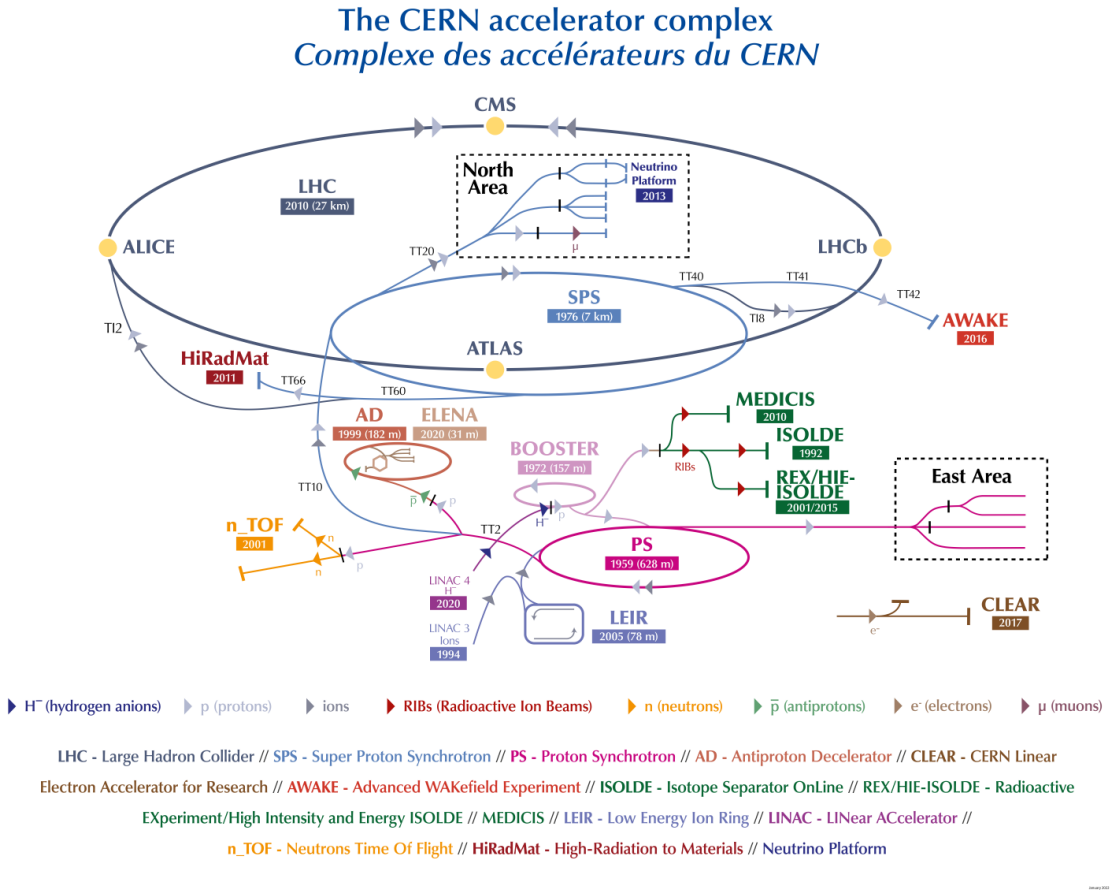

Figure 2·1: Diagram of the CERN accelerator complex, showing the LHC and the chain of other accelerator rings. Each ring boosts the energy of accelerated particles before injecting it into the next ring in the sequence. The diagram is taken from [6].

The LHC had a successful Run I between 2010 and 2013, where the center-of-mass energy for proton-proton collisions was 8 TeV. This first run was marked by the discovery of the Higgs boson by ATLAS and CMS experiments in 2012. After a long

shutdown, the center-of-mass energy was increased to 13 TeV, and the LHC operated from 2016 to 2018 for Run II data taking. During Run II, the instantenous luminosity was about $1 \times 10^{34}\ cm^{-2}s^{-1}$, corresponding to around 25 proton-proton collisions per bunch crossing. The integrated luminosity corresponding to Run II data taken by the CMS experiment is measured to be 137 fb^{-1}, with 1.6% overall uncertainty [66–68].

2.2 The CMS Detector

The CMS detector is a multi-purpose detector, designed trigger on [69, 70] and detect muons, electrons, photons, charged and neutral hadrons [71–73]. To be able to identify this wide range of particles, the CMS detector uses a combination of signals from different sub-detectors. These signals are then used in the Particle-Flow (PF) algorithm to reconstruct the particles in the event. The sub-detectors in the CMS detector are:

- Tracking system
- Electromagnetic calorimeter (ECAL)
- Hadron calorimeter (HCAL)
- Muon system

CMS detector also uses a two-level trigger system [69, 70], to make physics-based decisions on the data being saved to disk, therefore reducing the output rate. In the following sub-sections, the functionality of these subsystems and the overall design of the CMS detector are discussed.

2.2.1 Overall Layout

The CMS detector has an overall length of 28.7 meters, and weighs 14000 tonnes. The cylindrical detector has a diameter of 15 meters, as shown in Fig. 2·2. The

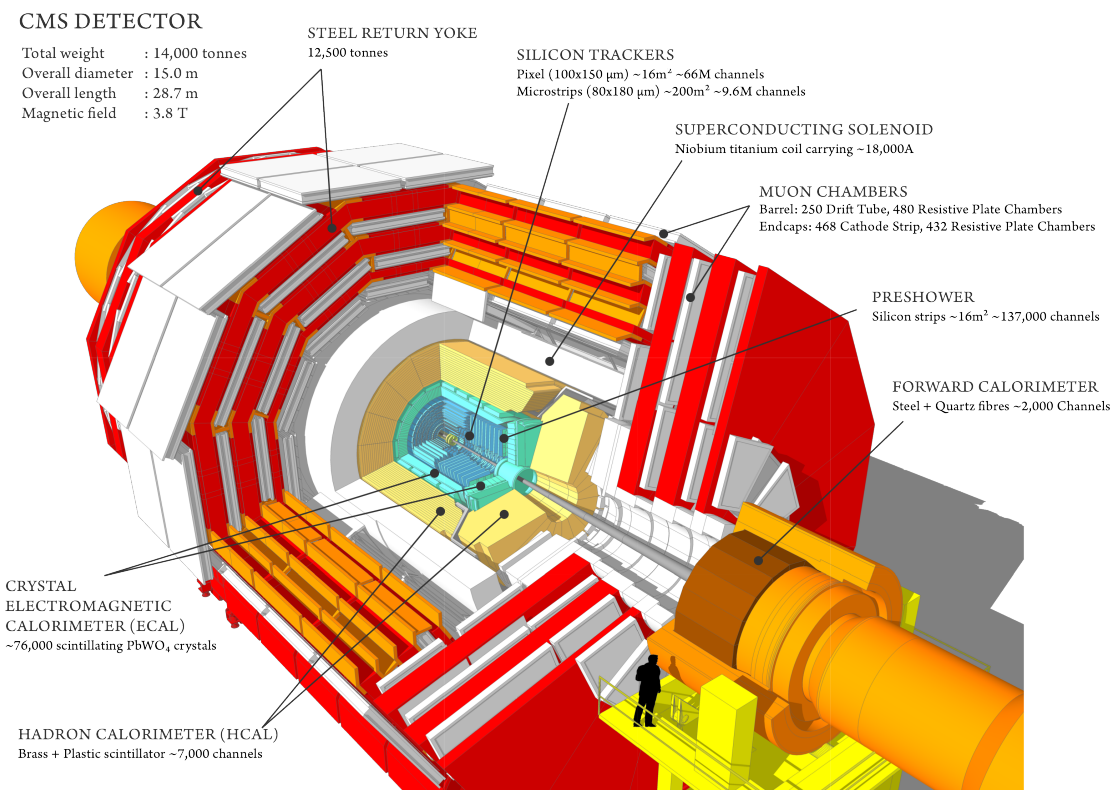

Figure 2·2: Sectional view of the CMS detector. The LHC beams travel in opposite directions along the central axis of the CMS cylinder colliding in the middle of the CMS detector [7].

CMS detector has a superconducting solenoid, which has an internal diameter of 7 meters, and creates a magnetic field of 3.8 T. This magnetic field curves the charged particles as they pass through the detector, hence enabling the measurement of their momentum from the radius of curvature of their tracks.

Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters are installed in the high psuedo-rapidity range to extend the coverage of the detector. Outside of the solenoid volume, a muon detection system is installed, where muons are measured in gas-ionization detectors embedded in the steel

flux-return yoke.

In the CMS detector, the right-handed coordinate system is used. The x-axis points to the center of the LHC ring, the y-axis points up, and the z-axis point along the beam in the counter-clockwise direction. Due to the cylindrical symmetry of the detector geometry, the angular coordinates of η and ϕ are often used to describe event kinematics. ϕ is the azimuthal angle, defined to be 0 along the x-axis, and $\frac{\pi}{2}$ along the y-axis. η is the psuedo-rapitiy, which depends on the angle from the z-axis, θ, with the following equation: $\eta = -\ln[\tan(\frac{1}{\theta})]$. This way, η equals to 0 along the z-axis, directly perpendicular to the beam-axis, and $|\eta| \to \infty$ as the $\theta \to \frac{\pi}{2}$, i.e. perpendicular to the beam axis. Most detector coverage extends to until $|\eta| = 5$.

2.2.2 Tracking System

The innermost layer of the CMS detector is the tracking system. This tracking system consists of a silicon pixel detector and a silicon strip detector, referred to as the inner tracker and the outer tracker, respectively. Both are used to reconstruct tracks of charged particles, under the influence of the 3.8 T magnetic field generated by the solenoid. The inner pixel detector has higher granularity, therefore it can reconstruct tracks more accurately near the beam pipe, where the track density is higher.

Tracks of charged particles are crucial objects for CMS, because they allow measuring the momenta and the electrical charges of particles. These particles include electrons, muons, taus and charged hadrons. The identification of tracks also allow to reconstruct the proton-proton collision vertices in the event. This enables the identification of the primary vertex in the event with the highest sum of $p^2_{T,track}$. Thus, the rest of the vertices can be identified as pile-up (PU), and charged particles from these PU vertices can be rejected in object reconstruction. Track reconstruction also allows to identify b-quarks by identifying displaced vertices.

The CMS inner tracker was upgraded during the year-end technical stop of the

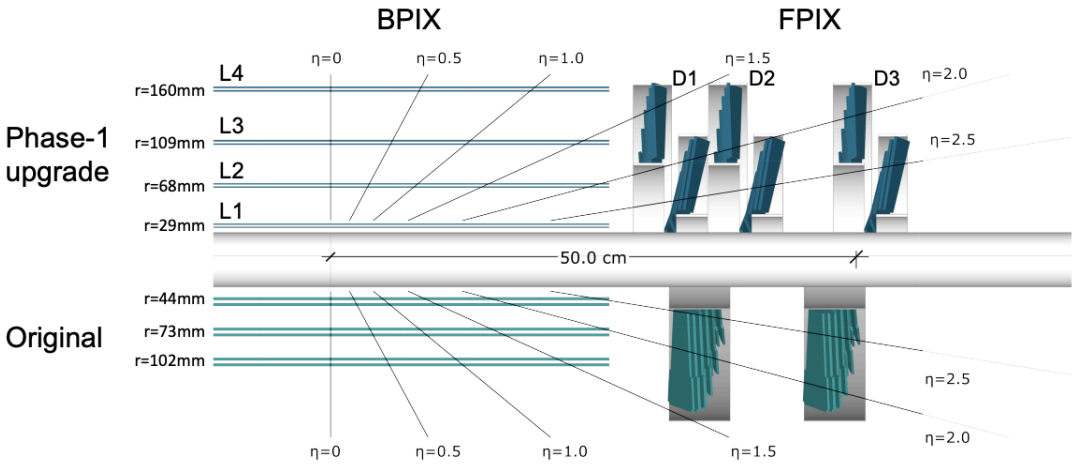

Figure 2·3: Comparison of the layouts between the original pixel detector (bottom) and the Phase-1 upgraded pixel detector (top) [8].

LHC in 2016/2017, referred to as the Phase-1 upgrade [8], in order to handle the instantaneous luminosity that exceeds the previously designed maximum value of $1 \times 10^{34}\ cm^{-2}s^{-1}$. Utilizing the new beam pipe with smaller radius installed during the LHC Long Shutdown 1 (LS1), the Phase-1 pixel detector sits closer to the beam center, with 4 barrel layers (L1-L4) at radii of 29, 68, 109, and 160 mm, compared to the original three layers at 44, 73, and 102 mm. There is also an additional endcap disk on each end closer to the collision point. Therefore, there are now three disks at each end with distances of 291, 396, and 516 mm from the center of the detector, compared to the original three layers at 345 and 465 mm. This new layout is optimized to have 4-hit coverage for tracks within $|\eta| < 2.5$. The original and the upgraded pixel detector layouts are shown in Fig. 2·3.

Outside of the pixel detector, the silicon strip tracker is installed, which consists of three subsystems: the Tracker Inner Barrel and Disks (TIB/TID), the Tracker Outer Barrel (TOB), and the the Tracker EndCaps (TEC). Their radii range from 20 cm to 116 cm and are up to 282 cm away from the detector center in the z direction. The layout of the silicon strip tracker, together with the 2016 version of the pixel detector,

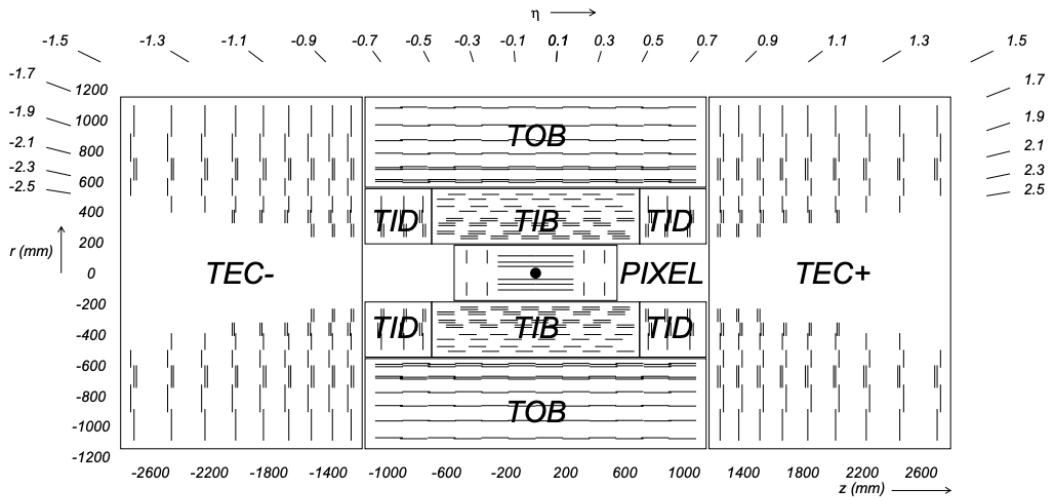

Figure 2·4: Schematic cross section through the CMS tracker. Notice that the pixel detector is shown as the 2016 version. The image is taken from [9].

is shown in Fig. 2·4.

2.2.3 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is installed on the second layer of the detector, outside the inner and outer tracking systems. It is made of 61200 lead tungstate ($PbWO_4$) crystals mounted in the barrel (EB), and 7324 of those crystals in each of the 2 endcaps (EE). The $PbWO_4$ crystals make a radiation-resistent high-granularity calorimeter, due to it's high density and short radiation length. The EB section covers the psuedo-rapidity range of $|\eta| < 1.479$, and the two EE sections cover $1.479 < |\eta| < 3.0$.

The energy of particles, primarily electrons and photons, are measured through the scintillation effect. Photons emitted from the scintillation are collected by the avalanche photodiodes (APD) in the barrel and vacuum phototriodes (VPT) in the EE. Two pre-shower detectors are placed on the inner side of the EEs, which helps to distinguish high-energy photons from pions, which can decay into a pair of closeby

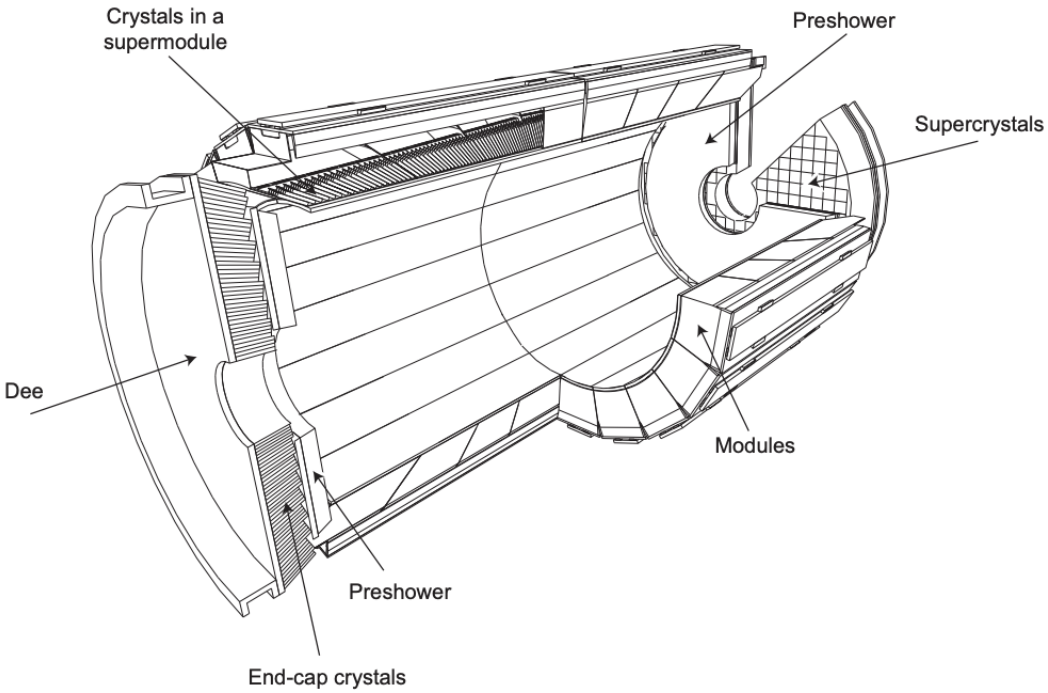

Figure 2·5: Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules, supermodules and endcaps, with the preshower in front [9].

low-energy photons. The preshower is a sampling calorimeter made of two planes on each end. Each plane has a layer of lead radiators to produce electromagnetic showers from incoming photons and electrons, and a silicon strip detector to measure the energy and shower profile. The silicon strips are placed with a width of 2 mm, and are oriented in perpendicular directions between two layers, giving a much finer granularity than EE and EB.

The crystals in the barrel section are contained by 1 mm wall of aluminum, which constitues a submodule. 400 or 500 submodules are assembled into a module, and every four of each constitute a supermodule. The crystals in the EE are organized in units of identically shaped 5×5 crystals called supercrystals. Each EE is divided into two halves (Dees). The structure of the ECAL is shown in Fig. 2·5.

2.2.4 Hadron Calorimeter

The hadron calorimeter (HCAL) is designed to measure the energy of hadrons, which is specifically important for reconstructing and measuring the energy of hadron jets. The HCAL is composed of several parts:

- A barrel part (HB), which sits between the ECAL and the magnet coil, covering the range of $|\eta| < 1.3$.
- An endcap part (HE), covering the range of $1.3 < |\eta| < 3$.
- Forward hadron calorimeters (HF), which are placed close to the beam pipe to capture outgoing particles with small angles, covering up to $|\eta| < 5$.

Due to the limited space between the ECAL and the magnet coils, additional hadron calorimeters (HO) are placed outside of the magnet coils to help contain hadron showers. The layout of the HCAL detector is shown in Fig. 2·6.

The majority of HCAL is a sampling calorimeter made of alternating absorbing layers of brass and plastic scintillating layers. The forward calorimeter HF utilizes a different design, which consists of a steel absorber and quartz fibers that generate Cherenkov light, in order to operate in a much harsher radiation environment, compared to other parts of the calorimeter.

2.2.5 Muon System

Muon detection is of great importance to the CMS, since it is a powerful tool for recognizing signatures of interesting physics processes. Muons also have the potential to be reconstructed with a larger resolution since they are less affected by radiative losses in the tracker material, compared to electrons.

The muon system is installed in the psuedo-rapidity range of $|\eta| < 2.4$, consisting of drift tubes (DTs), cathode strip chambers (CSCs) and resistive plate chambers

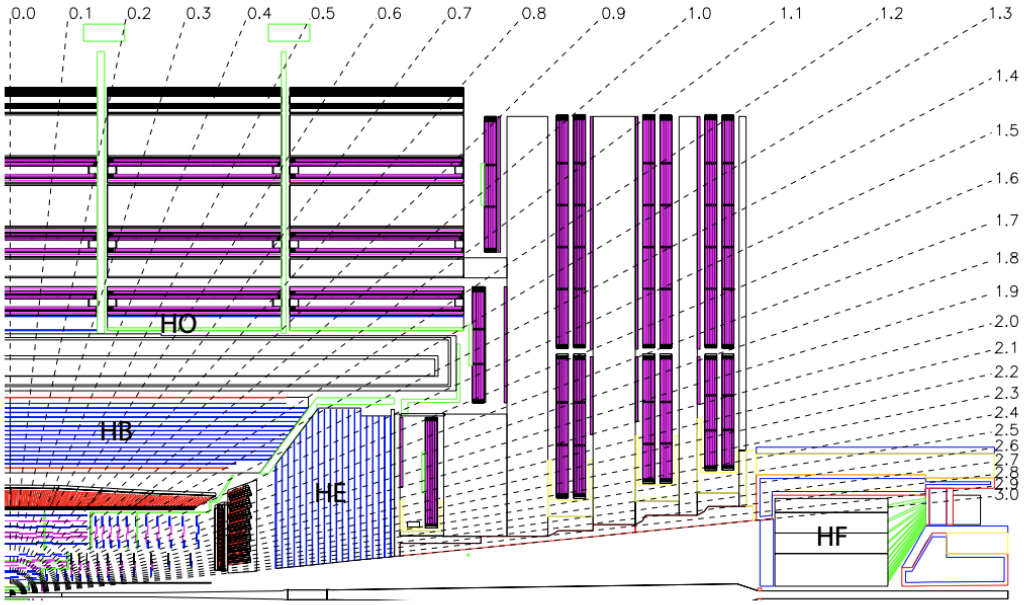

Figure 2·6: Longitudinal view of the CMS detector showing the locations of the hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters [9].

(RPCs). The DTs are placed in the barrel and cover the $|\eta| < 1.2$ range, where the particle rate and the local magnetic fill strength are low. CSCs, due to their fast response time, fine segmentation, and radiation resistance, are used in the endcap, covering $0.9 < |\eta| < 2.4$. RPCs are inserted into both barrel and endcaps to provide redundant trigger system with a time resolution of 1 ns, which is much shorter than the bunch crossing interval of 25 ns.

2.2.6 Trigger System

With a proton-proton bunch crossing interval of 25 ns, collecting and saving all collision events at a rate of 400 MHz in unsustainable, due to limitations in both storage space and data transfer rates. Storing all the events at this rate is also unnecessary, since interesting events with high-energy physics objects occur at a much lower rate.

To select the events of interest and reduce the rate of data being saved, a two-tiered trigger system is being used in the CMS experiment. The first-level trigger (L1), is composed of custom hardware processors, which uses event information from the calorimeters and the muon system to select events at a rate of 100 kHz within a fixed latency of about 4 μs [69]. The second level, known as the high-level trigger (HLT), consists of processors running a version of the full event reconstruction software optimized for fast processing, and reduces the event rate to around 1 kHz before data storage [70].

Chapter 3

Data Analysis Strategy

3.1 Physics objects

A global event reconstruction using the Particle Flow (PF) algorithm [74] is carried out in order to identify each particle in the event[1]. PF algorithm accomplishes this by combining detector hit information from all different sub-detectors within CMS. The specific information used to determine the energy and momentum of a particle depends on the identity of the particle: Whether it is a photon, muon, electron, neutral hadron or charged hadron. Photons are identified as ECAL energy clusters that are not linked to any charged particle tracks (following from the fact that they are electrically neutral). Electrons are identified as a primary charged particle track, potentially together with ECAL energy clusters that are consistent with either the extrapolation of this track or possible bremsstrahlung photons emitted when the electron passes through the tracker material. Muons are identified as consistent charged particle tracks in both the central tracker and the muon system, as well as potential energy clusters in the calorimeters that are consistent with the tracks. Tracks that are not already assigned to electrons or muons, together with calorimeter energy deposits, are identified as charged hadrons. Finally, neutral hadrons are identified using leftover HCAL and ECAL clusters after the identification of charged hadrons. A schematic summarizing the particle identification in CMS is shown in Fig. 3·1.

[1] An event typically refers to a single proton-proton bunch crossing, occuring every 25 nanoseconds in the LHC.

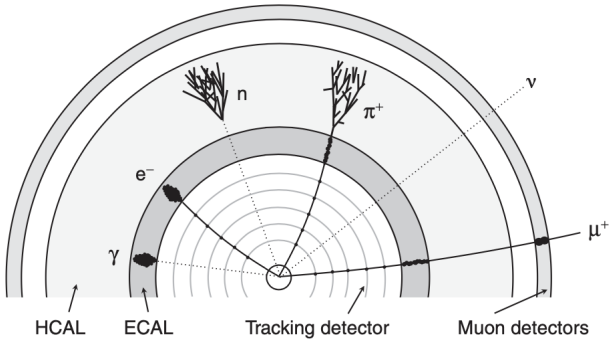

Figure 3·1: A schematic representing the particle detection at the CMS detector. Hadrons (e.g. n, $\pi^{\pm}$) are detected as hadronic showers of particles (jets) due to color confinement, and weakly interacting particles such as neutrinos (ν) pass through the detector without interaction. Schematic is taken from [10].

Since a bunch of protons are collided during each bunch crossing, multiple collision vertices arise due to multiple interactions within the bunch. Aim of the event reconstruction is to select a primary collision vertex, which is determined as the vertex containing the largest sum of p_T^2 from associated tracks [75]. Tracks and energy content from the rest of the vertices are very commonly referred as pileup (PU). Thanks to the assignment of the primary vertex, charged particles can be filtered out from the main event if their tracks are found to be originating from a non-primary vertex. This allows to reduce the PU contribution in the reconstructed event.

In the following subsections, details of how different physics objects are reconstructed will be discussed. To aid the discussion, signal region (SR) refers to the set of selections used to collect VBF $H \rightarrow inv.$ signal events, and control regions (CR) refer to the set of selections used to measure dominant background yields.

3.1.1 Jets

A typical jet consists of a shower of hadrons generated by a single quark or gluon. In this analysis, jets are reconstructed by clustering reconstructed particles using the

anti-k_T algorithm [76]. Jets are clustered with a distance parameter of 0.4 and are referred to as AK4 jets.

Jet momentum is determined as the vector sum of all particle momenta in the jet, and is found from simulation to be within 5 to 10% of the true momentum over the full p_T spectrum and detector acceptance. An offset correction is applied to jet energies to take into account the contribution from additional proton-proton interactions within the same or nearby bunch crossings (pileup). Jet energy corrections are derived from simulation, and are confirmed with measurements of the energy balance in dijet, multijet, γ+jet, and $Z(\ell\ell)$+jet events[2] [77]. For the analysis that will be explained in this thesis, the `Summer19UL17_V5` and `Summer19UL18_V5` versions of the jet energy corrections are used for the 2017 and 2018 datasets, respectively [78]. On top of the jet energy corrections, jet energy resolution (JER) corrections are also applied to correct for differences of jet resolution between data and simulation.

The AK4 jets used in this analysis are required to pass the loose identification criteria [79], and AK4 jets with $p_T < 50$ GeV must pass the pileup identification (ID) criteria [80]. Pileup ID criteria aims to reduce the contribution from pileup by removing jets from the event which contain a large amount of energy from secondary vertices within the bunch crossing.

In addition, jet-lepton overlap cleaning is done by removing any AK4 jet that is within a cone of radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.4$ surrounding a lepton or a photon. Only lepton and photon candidates satisfying the object criteria described below are considered during the cleaning. Lastly, if the leading or the subleading jet is found to be within the tracker coverage (i.e. $|\eta| < 2.5$), constraints on the charged hadron and neutral hadron energy fractions of the jet (R_{CHEF}, R_{NHEF} respectively) are placed such that $R_{CHEF} > 0.1$ and $R_{NHEF} < 0.8$. If a jet is found to violate these

[2]Jet energy corrections are applied as a function of the transverse momentum p_T and psuedo-rapidity η of the jet. They are composed of successive corrections due to PU effects, non-uniform detector response and residual differences in data and simulation.

requirements, it is removed from the event. These additional requirements aid in rejecting mis-identified jets with spuriously low charged hadron fractions[3].

b-tagged jets

Identifying jets that originate from a b-quark is important in this analysis, since a b-quark is not expected in the final state of the VBF $H \to inv.$ signal (for example, see Fig. 1·3), and vetoing events with b-jets in the final state allows to remove most of the backgrounds originating from top quark production.

In this analysis, jets with $p_T > 20$ GeV and $|\eta| < 2.4$ are identified as b-jets according to the DeepCSV algorithm [81]. A medium working point is adopted, corresponding to correctly identifying a b-jet with a probability of 80%, and mis-identifying a light-flavor jet with a probability of 10%. This working point corresponds to the value of DeepCSV tagger to be greater than 0.4506 in 2017, and 0.4168 in 2018.

3.1.2 Missing transverse momentum and recoil

The vector $\vec{p}_T^{miss}$ is defined as the imbalance in the transverse momentum of all particles that interact with the detectors. Due to momentum conservation in the plane transverse to the beam axis, $\vec{p}_T^{miss}$ corresponds to the transverse momentum that is carried by undetected particles such as neutrinos. Practically, $\vec{p}_T^{miss}$ is computed as the negative of the vectorial sum of transverse momenta of all Particle Flow (PF) candidates and is therefore also referred to as PF $\vec{p}_T^{miss}$. The magnitude of the $\vec{p}_T^{miss}$ is referred to as p_T^{miss}.

Minimum energy thresholds in the calorimeters, inefficiencies in the tracker, non-linearity of the response of the calorimeter for hadronic particles can lead to an over- or underestimation of p_T^{miss}. The bias on the p_T^{miss} measurement is reduced by prop-

[3]On average, one would expect $\approx 60\%$ of jet energy to be in the form of charged particles (e.g. from charged pions $\pi^{\pm}$). Hence a reconstructed jet with $< 10\%$ charged energy fraction is highly unexpected and potentially due to mis-identification of other particles or detector noise.

agating the effect of the jet energy corrections introduced in Sec. 3.1.1 according to

$$\vec{p}_T^{miss}(\text{corr}) = \vec{p}_T^{miss} - \sum_{\text{jets}} (\vec{p}_{\text{T,jet}}(\text{corr}) - \vec{p}_{\text{T,jet}}), \tag{3.1}$$

where the "corr" refers to the energy corrected measurements of the related objects. This correction for $\vec{p}_T^{miss}$ uses jet energy scale corrections for all corrected jets with $p_T > 15$ GeV that have less than 90% of their energy deposited in the ECAL. Furthermore, if a muon is found in a jet, its 4-momentum is subtracted from the 4-momentum of the jet when performing the correction and is added back to a corrected object.

Since $H \rightarrow inv.$ signal events in this analysis contain only jets and no other reconstructed candidates, p_T^{miss} is equivalent to the total hadronic momentum in the event that is transverse to the proton-proton collision axis. For the leading $Z(\nu\nu)$ + jets and $W(\ell\nu)$ + jets backgrounds (described in Sec. 3.4.1), this corresponds to the transverse momentum of the Z or W boson. To mimic this behavior in in the control regions of this analysis, the transverse momentum of the hadronic recoil $\vec{U}$, defined as the vectorial sum of the transverse momenta of all particles except the vector boson (or its decay products) is used. The variable is computed as

$$\vec{U} = \vec{p}_T^{miss} + \sum_{i \in \text{ leptons, photons}} \vec{p}_T^{\,i} \tag{3.2}$$

where the sum takes into account the leptons and photons used to define the respective control region. All control regions defined in this analysis are described in Sec. 3.4. The uncertainty of p_T^{miss} has a strong dependence on the event topology. Therefore, the uncertainty on p_T^{miss} is often factorized into its components of jets, leptons and unclustered energy. Each sub-component is then varied within its scale and resolution uncertainty.

In addition to the events with genuine p_T^{miss}, such as $Z(\nu\nu)$ + jets, anomalous

high-p_T^{miss} events can also appear due to various phenomena. In the ECAL, spurious deposits may appear due to particles striking sensors in the ECAL photodetectors, or from real showers with non-collision origins such as those caused by beam halo particles. ECAL dead cells can cause real energy to be missed, again leading to a spurious imbalance. In the HCAL, spurious energy can arise due to noise in the hybrid photodiode and readout box electronics, as well as direct particle interactions with the light guides and photomultiplier tubes of the forward calorimeter. A number of filters has been developed by the POG/DPG groups to identify and suppress anomalous high p_T^{miss} events [1]. The recommended filters are listed in Tab. 3.1 and are applied in the analysis.

Table 3.1: The p_T^{miss} filters recommended by the JME POG [1]. The recommendations apply to both 2017 and 2018. Except for the bad super cluster filter ("EE badSC"), all filters are applied both in data and simulation (MC).

Filter	Name in input dataset	Applied in data (MC)
Primary vertex filter	Flag_goodVertices	✓(✓)
Beam halo filter	Flag_globalSuperTightHalo2016Filter	✓(✓)
HBHE noise filter	Flag_HBHENoiseFilter	✓(✓)
HBHEiso noise filter	Flag_HBHENoiseIsoFilter	✓(✓)
ECAL TP filter	Flag_EcalDeadCellTriggerPrimitiveFilter	✓(✓)
Bad PF Muon Filter	Flag_BadPFMuonFilter	✓(✓)
EE badSC noise filter	Flag_eeBadScFilter	✓(×)
ECAL bad calibration filter update	Flag_ecalBadCalibFilter	✓(✓)

3.1.3 Leptons

Muons

Muons are reconstructed by combining information from silicon tracker detector and the muon system [72]. Requirements for a good-quality muon object include the fit quality of the muon track, and its consistency with the primary vertex in the event. All muons considered in this analysis have $p_T > 10$ GeV and $|\eta| < 2.4$ (i.e. they are reconstructed within the tracker range).

Two types of muon identification are used for this analysis, loose and tight muons [82, 83]. Loose muon identification is used to identify muons in the signal region and veto them[4]. Tight muon identification, on the other hand, has more strict quality conditions and are used to identify good quality muons in $W(\mu\nu)$ and $Z(\mu\mu)$ control regions.

Loose muon objects must pass the following requirements [82]:

- Candidate must be reconstructed as a PF muon
- Are either a Global or Tracker Muon

Tight muons pass all the requirements of loose muons, together with passing these additional muon quality requirements [83]:

- $p_T > 20$ GeV
- Normalized χ^2 of the muon track fit < 10
- At least one Muon Chamber hit must be included in the muon track fit
- Muon segments required in at least two muon stations

[4]This selection is due to the fact that final state muons are not expected in VBF $H \to inv.$ signal events. As explained later in Sec. 3.4.1, this veto helps suppressing background processes such as $\mathrm{W}(\mu\nu) + \mathrm{jets}$ and $\mathrm{Z}(\mu\mu) + \mathrm{jets}$.

- The muon track must have transverse impact parameter $d_{xy} < 2$ mm w.r.t. the primary vertex
- The longitudinal distance of the tracker track w.r.t. the primary vertex must be $d_z < 5$ mm
- Number of pixel hits > 0
- Cut on number of tracker layers with hits > 5

In addition to the requirements listed above, an energy isolation requirement is imposed on the muons. An isolation variable is computed based on the sum of energies of nearby PF candidates, within $\Delta R < 0.4$ of the muon. This isolation variable is then used to select muons that are relatively isolated from other physical objects. This isolation variable is required to be less than 0.25 for loose muons, and less than 0.15 for tight muons.

Electrons

Electrons are reconstructed using the information from the tracker and the ECAL detector [71]. All electrons considered in this analysis have $p_T > 10$ GeV and $|\eta| < 2.5$.

Similar to the case of muons, two types of identification are used for electrons, which are labeled as veto and tight electrons. Veto electrons are used to identify electrons in the signal region and veto events containing such electrons[5]. Tight electrons, which have tighter quality cuts, are used to identify electrons in $W(e\nu)$ and $Z(ee)$ control regions.

The quality requrements for veto and tight electron objects are different for electrons reconstructed in the barrel region ($|\eta| < 1.479$) and in the endcap region

[5]Similar to the case of muons, such a veto helps suppressing background processes such as W($e\nu$)+ jets and Z(ee) + jets.

($|\eta| > 1.479$). The details of the quality requirements are listed in Tables 3.2, 3.3, 3.4 and 3.5.

Quantity	Requirement
Full 5x5 $\sigma_{i\eta i\eta}$	< 0.0126
$\lvert\Delta\eta_{\text{seed}}\rvert$	< 0.00463
$\lvert\Delta\phi_{\text{In}}\rvert$	< 0.148
H/E	$< 0.05{+}1.16/E_{\text{SC}}{+}0.0324\rho/E_{\text{SC}}$
Relative Isolation With EA	$< 0.198{+}0.506/p_T$
|1/E-1/p|	< 0.209
Expected Missing Inner Hits	≤ 2
Pass conversion veto	yes

Table 3.2: Requirements used to define veto electrons in the barrel region ($|\eta| < 1.479$).

Quantity	Requirement
Full 5x5 $\sigma_{i\eta i\eta}$	< 0.0104
$\lvert\Delta\eta_{\text{seed}}\rvert$	< 0.00255
$\lvert\Delta\phi_{\text{In}}\rvert$	< 0.022
H/E	$< 0.026{+}1.15/E_{\text{SC}}{+}0.0324\rho/E_{\text{SC}}$
Relative Isolation With EA	$< 0.0287{+}0.506/p_T$
|1/E-1/p|	< 0.159
Expected Missing Inner Hits	≤ 1
Pass conversion veto	yes

Table 3.3: Requirements used to define tight electrons in the barrel region ($|\eta| < 1.479$).

In addition to the requirements listed in the tables above, electrons are cross-cleaned against the muons within $\Delta R < 0.3$.

Taus

Hadronically decaying τ leptons are required to pass identification criteria using the DeepTau algorithm [84]. In addition, τ candidates are required to be isolated from

Quantity	Requirement
Full 5x5 $\sigma_{i\eta i\eta}$	< 0.0457
$\lvert\Delta\eta_{\text{seed}}\rvert$	< 0.00814
$\lvert\Delta\phi_{\text{In}}\rvert$	< 0.19
H/E	$< 0.05{+}2.54/E_{SC}{+}0.183\rho/E_{SC}$
Relative Isolation With EA	$< 0.203{+}0.963/p_T$
|1/E-1/p|	< 0.132
Expected Missing Inner Hits	≤ 3
Pass conversion veto	yes

Table 3.4: Requirements used to define veto electrons in the endcap region ($|\eta| > 1.479$).

Quantity	Requirement
Full 5x5 $\sigma_{i\eta i\eta}$	< 0.0353
$\lvert\Delta\eta_{\text{seed}}\rvert$	< 0.00501
$\lvert\Delta\phi_{\text{In}}\rvert$	< 0.0236
H/E	$< 0.0188{+}2.06/E_{SC}{+}0.183\rho/E_{SC}$
Relative Isolation With EA	$< 0.0445{+}0.963/p_T$
|1/E-1/p|	< 0.0197
Expected Missing Inner Hits	≤ 1
Pass conversion veto	yes

Table 3.5: Requirements used to define tight electrons in the endcap region ($|\eta| > 1.479$).

other activity in the event. The isolation requirement is computed by summing the p_T of the charged PF candidates and PF photon candidates within an isolation cone of $\Delta R = 0.5$, around the tau candidate direction. The "VVLoose" isolation working point [85] is employed in this analysis for tau candidates with $p_T > 20$ GeV within $|\eta| < 2.3$. τ candidates within $\Delta R = 0.4$ of an identified veto electron or loose muon are rejected.

3.1.4 Photons

Photon candidates are reconstructed from energy deposits in the ECAL using algorithms that constrain the clusters to the size and shape expected from a photon [86]. The identification of the candidates is based on shower-shape and isolation variables. For isolated photons, scalar sums of the p_T of PF candidates within a cone of $\Delta R < 0.3$ around the photon candidate are required to be below the bounds defined. Only the PF candidates that do not overlap with the electromagnetic shower of the candidate photon are included in the isolation sums.

Similar to electrons and muons, two candidate definitions are employed for photons. Loose photons are used to reject events with reconstructed photons in the signal region. These photons are required to pass the EGamma POG loose identification criteria [87], have $p_T > 15$ GeV and be within $|\eta| < 2.5$. The exact identification criteria for loose photons in barrel and endcap are summarized in Tables. 3.6 and 3.7, respectively. Tight photons are used to identify photon objects in the dedicated photon control region. These photons are explicitly required to be in the barrel ($|\eta| < 1.479$), and have $p_T > 215$ GeV. The identification criteria for tight photons are summarized in Table 3.8.

Variable	Selection
	Barrel ($\|\eta\| < 1.479$)
Full 5x5 $\sigma_{i\eta i\eta}$	< 0.0106
H/E	< 0.04596
charged hadron isolation	< 1.694
neutral hadron isolation	$< 24.032 + 0.01512 \times p_T + 2.259 \times 10^{-5} \times {p_T}^2$
photon isolation	$< 2.876 + 0.004017 \times p_T$
Conversion safe electron veto	Yes

Table 3.6: Loose photon identification criteria for photons in the barrel. These photons must also have $p_T > 15$ GeV.

Variable	Selection
	Endcap ($1.479 < \|\eta\| < 2.5$)
Full 5x5 $\sigma_{i\eta i\eta}$	< 0.0272
H/E	< 0.0590
charged hadron isolation	< 2.089
neutral hadron isolation	$< 19.722 + 0.0117 \times p_T + 2.3 \times 10^{-5} \times {p_T}^2$
photon isolation	$< 4.162 + 0.0037 \times p_T$
Conversion safe electron veto	Yes

Table 3.7: Loose photon identification criteria for photons in the endcap. These photons must also have $p_T > 15$ GeV.

Variable	Selection
	Barrel
Full 5x5 $\sigma_{i\eta i\eta}$	< 0.01015
H/E	< 0.02197
charged hadron isolation	< 1.141
neutral hadron isolation	$< 1.189 + 0.01512 \times p_T + 2.259 \times 10^{-5} \times {p_T}^2$
photon isolation	$< 2.08 + 0.004017 \times p_T$
Conversion safe electron veto	Yes

Table 3.8: Tight photon identification criteria. The criteria are only given for the barrel region since endcap photons are not taken into account.

Photon purity

Photons are reconstructed from ECAL clusters, and can be discriminated from other sources of ECAL deposits due to the properties of the cluster, as well as their lack

of other associated signatures such as tracks or HCAL deposits. This discrimination is not perfect, however, and in some cases, non-photon objects will incorrectly be identified as photons, which in this text are also referred to as "fake photons". The leading source of fake photons is QCD production of multijet events, where a jet is misidentified as a photon. The QCD process is relevant mainly because of its large cross-section, which yields non-negligible contributions to the photon selection even if the per-jet probability of misreconstruction is small.

To estimate the fake contribution to the photon control region, a purity measurement is performed. The photon purity is defined as the fraction of reconstructed photons that is actually due to an isolated photon from the hard scattering event, as opposed to a fake. The purity is obtained from a template fit to the distribution of the $\sigma_{i\eta i\eta}$ variable in data, where $\sigma_{i\eta i\eta}$ variable represents the width of the ECAL shower in the η direction. Due to the different shower behavior of photons and hadrons, the $\sigma_{i\eta i\eta}$ distribution shows characteristically different behavior between these two classes of reconstructed photons: Real photons show a large peak with a cut-off at around $\sigma_{i\eta i\eta} \approx 0.01$, while fake photons will have a smaller peak (stemming from actual photons inside jets), and an additional non-peaking bulk region at larger $\sigma_{i\eta i\eta}$ (stemming from hadrons interacting in the ECAL). The inputs to the fit are defined as follows:

- Photons in data are selected by applying the same identification criteria as for the tight selection defined above, with the exception of the $\sigma_{i\eta i\eta}$ requirement. By removing this requirement, the full $\sigma_{i\eta i\eta}$ distribution can be observed.

- A real photon template is obtained from $\gamma+$ jets Monte Carlo (MC) simulation. The same identification criteria are applied as in data.

- A fake photon template is obtained from data. In this case, the identification criteria are modified by requiring that at least one of the isolation criteria is

not passed. Therefore, the photons in this template represent reconstructed photons inside jets, (i.e. "fake photons").

The templates are derived in separate bins of the photon p_T and the measurement is performed separately for 2017 and 2018. Events are selected using the following criteria:

- At least one jet with $p_T > 100$ GeV and $|\eta| < 2.4$, which is not overlapping with a photon of interest.
- The event must pass the `HLT_Photon200` trigger, which requires a photon with $p_T > 200$ GeV to be reconstructed at the HLT level.
- The event must pass the offline MET filters, and have $p_T^{miss} < 60$ GeV.

The fits are shown in Figs. 3·2 and 3·3. The resulting impurity values as a function of photon p_T are shown in Fig. 3·4. While the fits are performed in the range of $0.04 < \sigma_{i\eta i\eta} < 0.15$, the purity is evaluated only taking into account the contributions with $\sigma_{i\eta i\eta} < 0.01$, which is the requirement posed in the identification criteria used in the analysis (see Tab. 3.8). Overall, good fit performance is observed and the impurity is found to range between 1% and 4%, with a decreasing trend with increasing photon p_T. Residual differences between the best fit and the data appear because the template shapes do not match the data shapes perfectly. To take this effect into account, the measurement is repeated with four alternative binning schemes that range from being very fine to very coarse and are shown in Fig. 3·5. By varying the binning choice, the effect of shape differences can be amplified or mitigated, and its impact on the final impurity values can be tested. The spread of the resulting values per bin is found to be small and a 25% uncertainty is assigned to the impurity and thus the QCD background estimate. For application in the analysis, the nominal impurity values are interpolated using an exponential function fit.

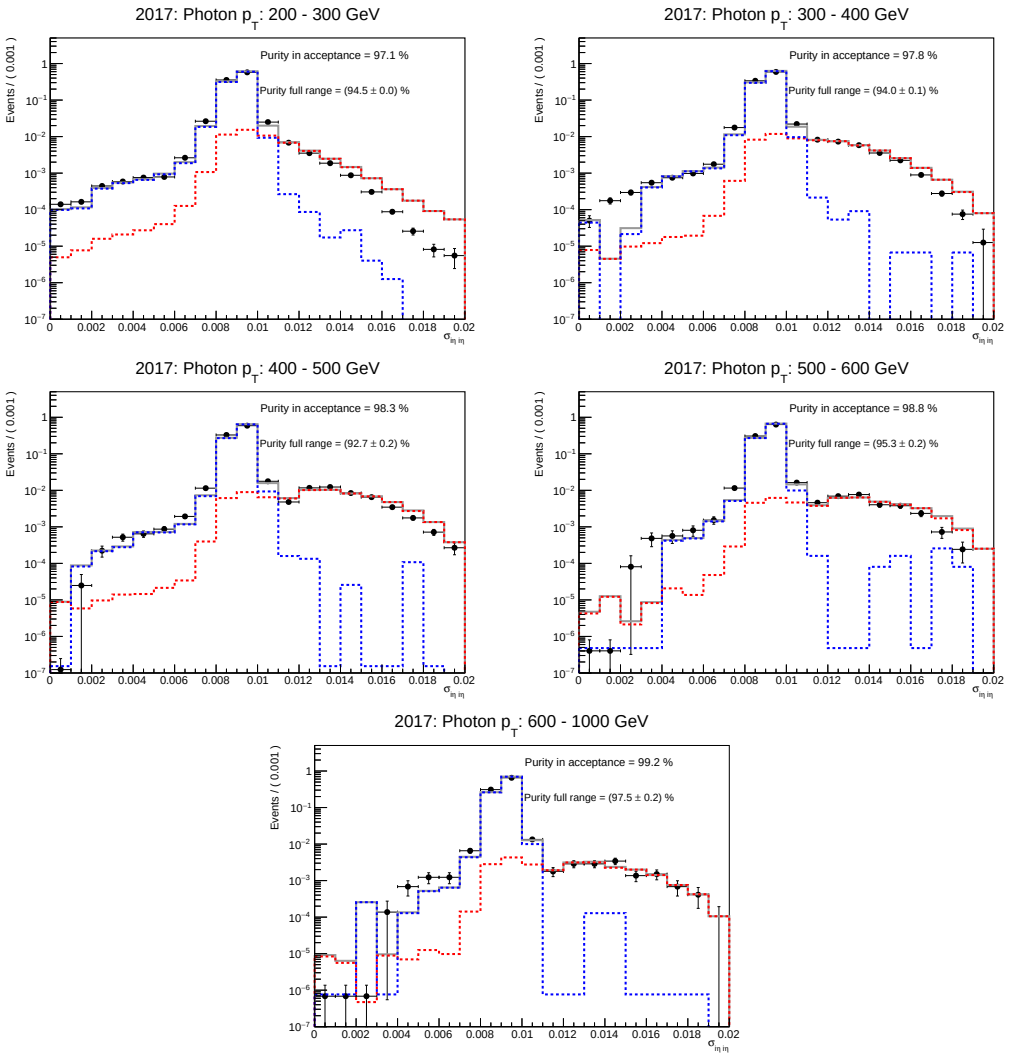

Figure 3·2: Template fits used to determine the photon purity in the 2017 dataset. The fits are shown in bins of photon p_T, increasing from left to right and top to bottom. The "nominal" binning scheme is used.

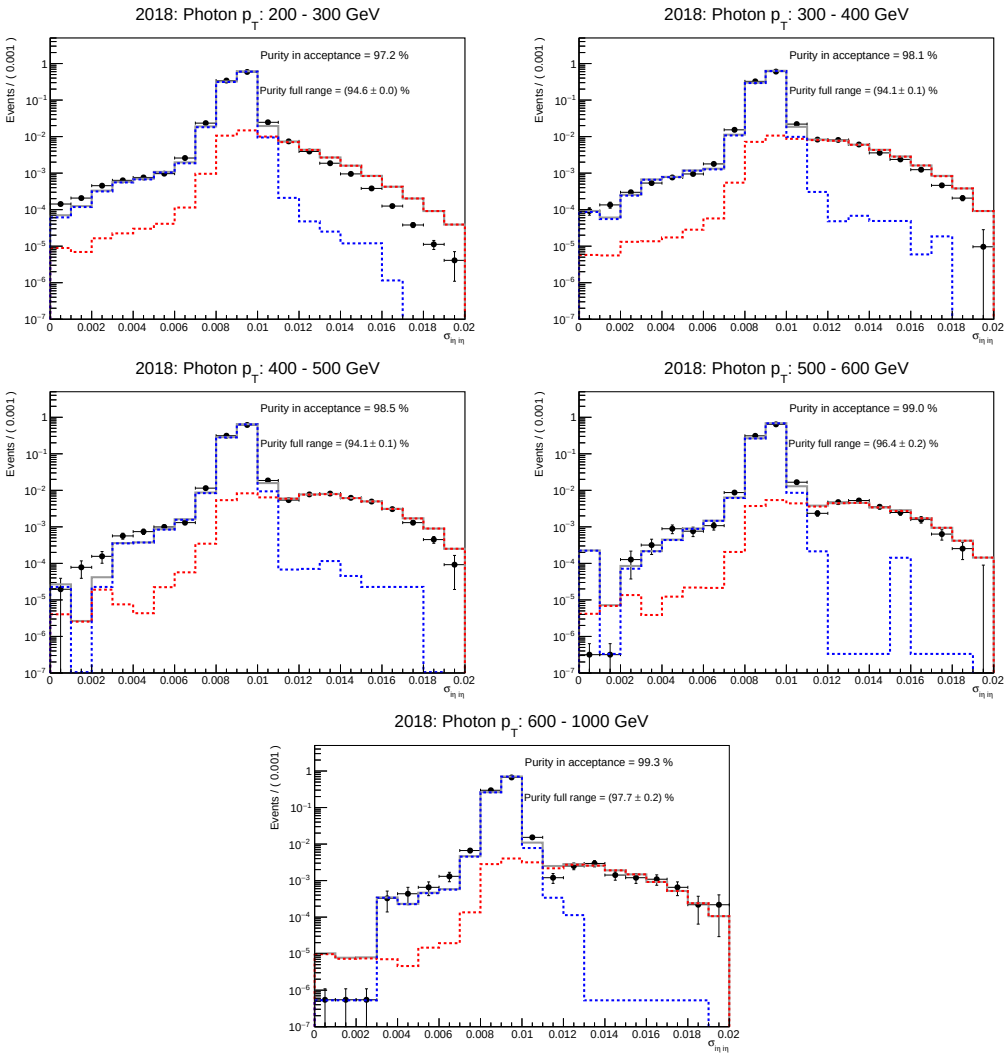

Figure 3·3: Template fits used to determine the photon purity in the 2018 dataset. The fits are shown in bins of photon p_T, increasing from left to right and top to bottom. The "nominal" binning scheme is used.

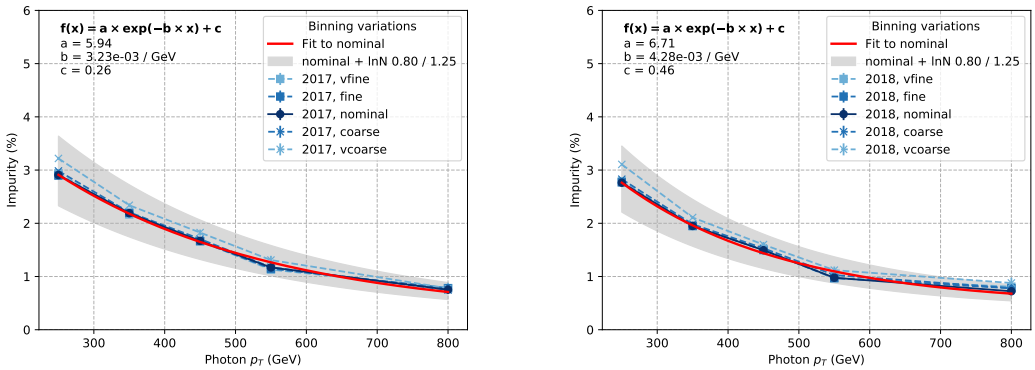

Figure 3·4: Photon impurity as a function of photon p_T for 2017 (left) and 2018 (right). The measured values for different binning choices are shown in the blue shaded lines and markers. The nominal result is interpolated using an exponential function fit, which is shown in the red solid line. The gray band represents a 25% uncertainty around the interpolated nominal result.

3.2 Reweighting of simulated events

Simulated signal and background samples can differ from collision data due to various effects. Therefore, reweighting of simulated events is necessary to correct simulated samples for these effects. The reweighting procedure for different types of effects are outlined in this section.

3.2.1 Trigger efficiency reweighting

$p_{T,no-\mu}^{miss}$ and $H_{T,no-\mu}^{miss}$ triggers

In this analysis, the collision data for signal region, $W(\mu\nu)$ and $Z(\mu\mu)$ control regions (see Sec. 3.4) are collected using high-level triggers that require $p_{T,no-\mu}^{miss} > 120$ GeV and $H_{T,no-\mu}^{miss} > 120$ GeV. The fact that muon momenta are subtracted from the p_T^{miss} and $H_{\mathrm{T}}^{\mathrm{miss}}$ calculations allow these triggers to be used both for collecting data in signal region (high p_T^{miss}), and muon control regions (high p_T muons). To simplify the discussion, this trigger will be referred to as the "MET trigger" in this text.

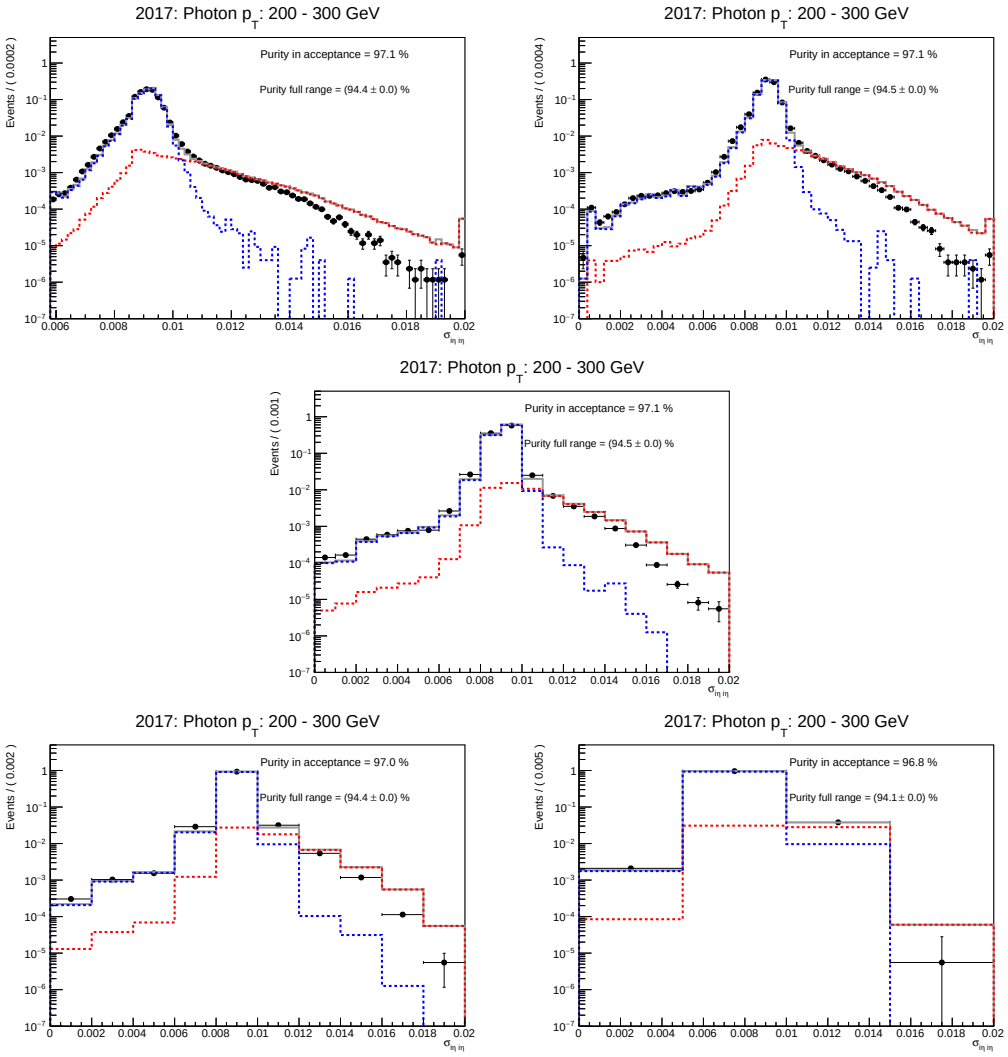

Figure 3·5: Comparison of binning schemes used to define a systematic uncertainty on the purity measurement. In all cases, the $200 < p_T < 300$ GeV bin of the 2017 data set is shown. The binning choices are very fine and fine (top row), nominal (middle row) and coarse and very coarse (bottom row).

The efficiency of MET trigger is calculated using W($\mu\nu$)+jets events, as a function of recoil. Here, recoil refers to the vectorial sum of the missing transverse momentum (MET) and the muon transverse momentum, as defined in Eq. 3.2. The W($\mu\nu$)+jets events are selected by requiring the event to pass `HLT_IsoMu27` trigger for 2017, and `HLT_IsoMu24` for 2018. In addition, the reconstructed muon is required to be tightly-identified as defined in Sec. 3.1.3, and have $p_T > 40$ GeV. The full set of selections is listed below:

1. Event must have exactly one tightly-identified muon with $p_T > 40$ GeV.
2. Veto on additional leptons, photons, b jets, τ_{had} candidates.
3. $\Delta\phi(jet, \vec{p}_T^{miss}) > 0.5$ for the four leading jets with $p_T > 30$ GeV.
4. (Calo p_T^{miss}- PF p_T^{miss}) / recoil < 0.5
5. Leading AK4 jet with $p_T > 80$ GeV, passing the tight jet ID.
6. Subleading AK4 jet with $p_T > 40$ GeV.
7. $\Delta\eta_{jj} > 1.0$
8. $\Delta\phi_{jj} < 1.5$

The efficiency of the MET trigger is computed separately for data and Monte Carlo simulation (MC), and the ratio between those two measurements are used to derive scale factors (SF) to correct the MC. The efficiency measurement procedure is described below.

To understand the dependence of the efficiencies on the jet kinematics, the efficiencies in data and MC are measured separately for two cases: Events where the two highest-p_T jets are both in the central region of the detector (i.e. $|\eta| < 2.5$), and events where at least one jet is in the forward region ($|\eta| > 2.5$).

To smooth out the fluctuations in efficiencies for each case, a sigmoid function is fit to both data and MC efficiency curves. The sigmoid function has three parameters and is written in the form $f(x, a, b, c) = \frac{c}{1+e^{-a(x-b)}}$. The factors for correcting the MC simulation are then calculated as the ratio of the best-fit sigmoid curves for data and MC. The efficiencies and resulting SFs are shown in Fig. 3·6. From Fig. 3·6, it can be seen that the scale factors are mostly $\simeq 1$ for the analysis phase space (i.e. Recoil > 250 GeV).

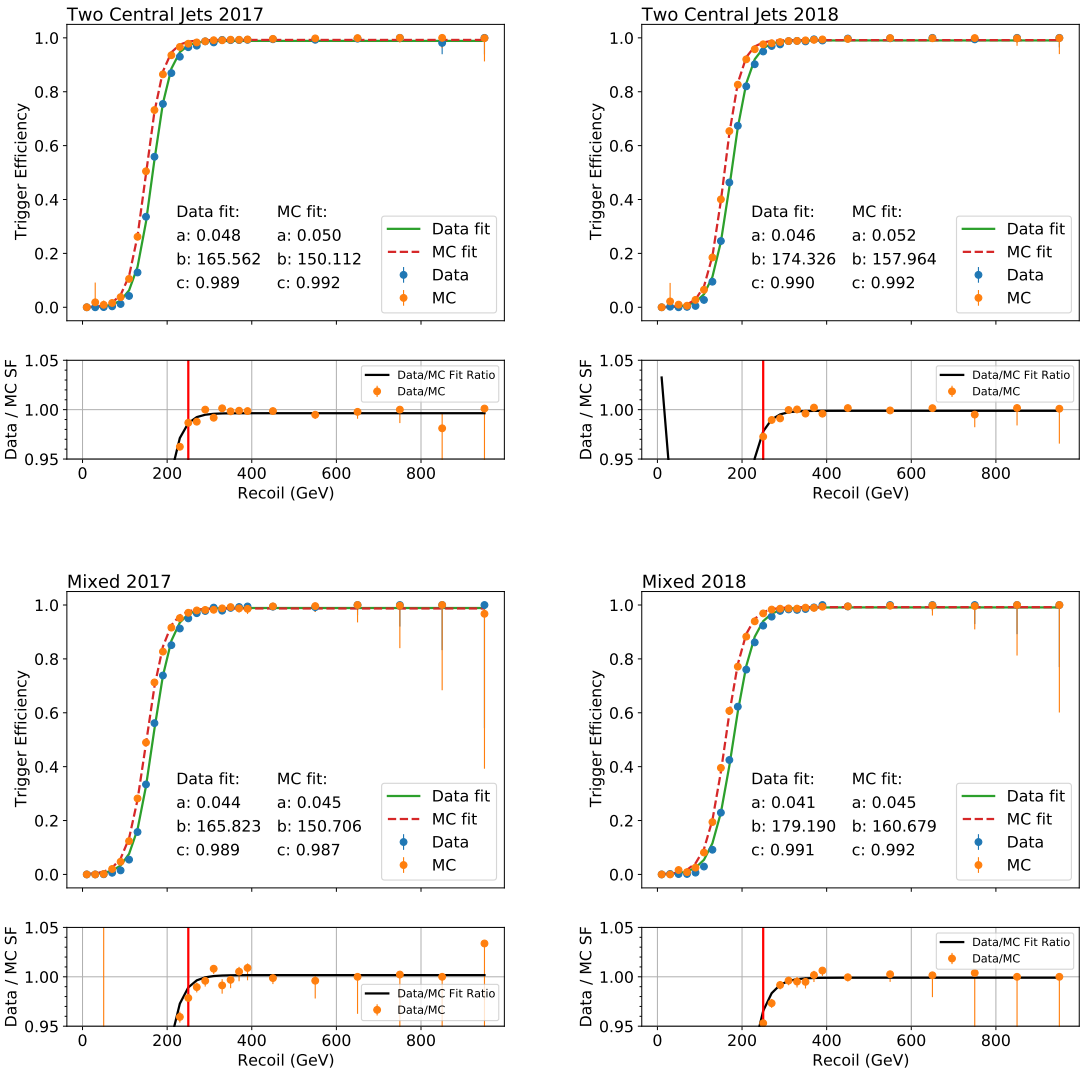

Figure 3·6: MET trigger efficiencies for data and MC, for events where the two VBF jets are both central (top) and where at least one jet is forward (bottom). Left column shows results with 2017 dataset, while the right column shows the results with 2018 dataset. To each efficiency curve, a sigmoid function with three parameters are fitted: $f(x, a, b, c) = \frac{c}{1+e^{-a(x-b)}}$. Resulting data/MC scale factors are shown in the bottom ratio pad for each case. The black line represents the ratio of two best-fit sigmoid functions, which is used as the correction factor to MC as a function of recoil.

Photon trigger

The photon trigger efficiency is measured using events collected with HLT_PFHT1050 trigger, which was fully unprescaled in 2017 and 2018, meaning that all the events passing the trigger were saved to the disk.

Events are selected in the same way as for the photon control region in the analysis (see Sec. 3.4.6) except for the photon p_T, recoil, H_{T} and trigger requirements. To ensure an unbiased measurement, an offline H_{T} of at least 1.5 TeV is required, where H_{T} is calculated as:

$$H_{\mathrm{T}} = \sum_{jet} p_T^{jet} \tag{3.3}$$

In Eq. 3.3, the sum runs over jets which pass the tight identification requirements and do not overlap with the selected photon within $\Delta R < 0.4$. The trigger efficiency ϵ is then determined as:

$$\epsilon(\texttt{HLT\_Photon200}) = \frac{\text{Offline selection \&\& } \texttt{HLT\_PFHT1050} \text{ \&\& } \texttt{HLT\_Photon200}}{\text{Offline selection \&\& } \texttt{HLT\_PFHT1050}}$$

The resulting efficiency in data and γ+jets H_{T}-binned simulation is shown in Fig. 3·7. In both data and simulation, the binned turn-on is fit using sigmoid functions, which are used to extract all further information. The trigger efficiencies in data and simulation are both found to be larger than 95% for photon p_T values of more than 230 GeV, which is used for the tight-identification requirement for photons reconstructed offline (see Sec. 3.1.4). The MC-to-data scale factor is evaluated as the ratio of the sigmoid functions in data and simulation and is found to be within 1% of unity consistent within an uncertainty of 1% with all individual points. In the analysis implementation, the scale factor is implemented as an event-by-event weight

based on the ratio of the sigmoid functions.

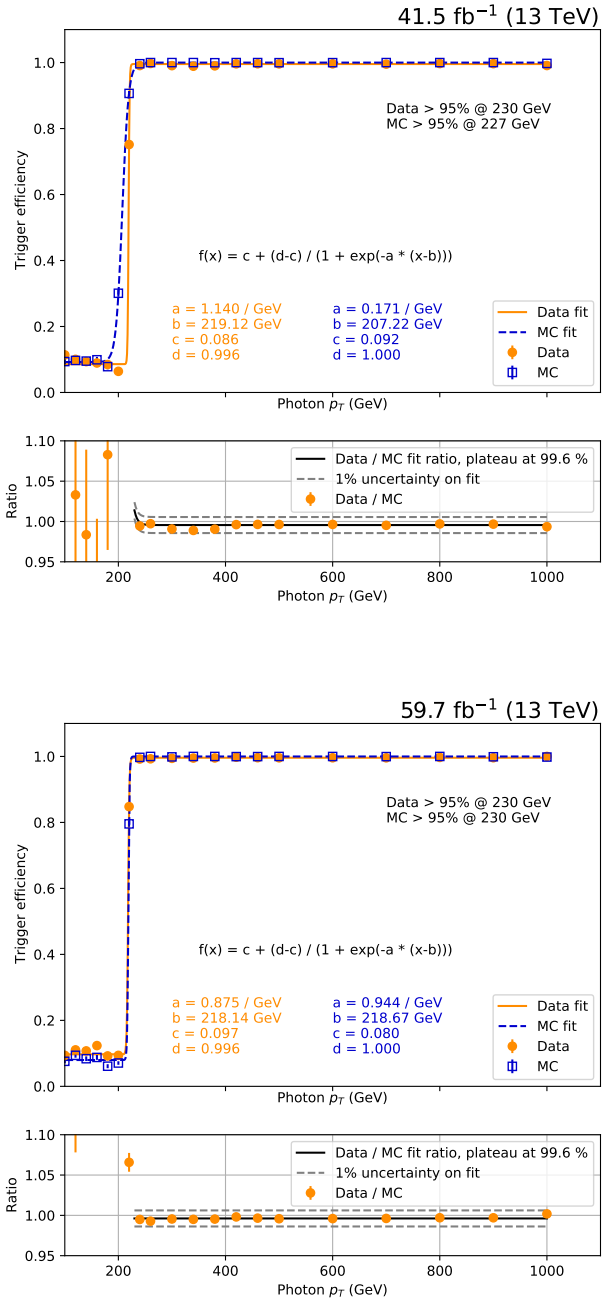

Figure 3·7: Efficiency of the `HLT_Photon200` trigger in data and H_{T}-binned γ+jets MC for 2017 (top) and 2018 (bottom) as a function of photon p_T. The orange and blue lines respectively represent sigmoid function fits to the turn-on in data and MC, with the fit function and best-fit parameter values given in the respectively colored labels. The bottom panel shows the ratio of the values measured in data over those in MC using orange markers. The solid black line corresponds to the ratio of the sigmoid fits to data and MC.

Electron trigger

This analysis uses the logical OR of three triggers for the selection of events with electrons in the final state: `HLT_Ele35*` (`HLT_Ele32*`), `HLT_Ele115*` and `HLT_Photon200` for 2017 (2018) data. The OR of these three triggers is henceforth referred to as "the electron trigger", to simplify the discussion. The higher-threshold triggers are advantageous in that they either do not contain isolation requirements (`HLT_Ele115`) or do not require a well-reconstructed track (`HLT_Photon200`), both of which enhance the selection efficiency at large electron p_T.

The efficiency of the electron trigger is measured in data and simulation using a "tag and probe" method. Tag electrons are required to pass a logical OR of all triggers considered here. Both the tag electron and the probe electron are required to pass the tight identification criteria used for the analysis selection (see Tables 3.3 and 3.5). In data, events are separated based on whether the probe electron passes the same trigger criteria as the tag. In both of these categories, separate fits to the distribution of the invariant mass of the tag-probe system are performed to extract the number of signal-like $Z(ee)$ events in each category, and the efficiency is defined as the ratio of the number of passing signal events and the number of all signal events. The efficiency in simulation is measured in a Drell-Yan sample, and the signal event counts are defined by simply counting all events in the passing and failing categories. The MC-to-data scale factor is then defined as the ratio of the efficiency in data and that in simulation. The efficiency in data, as well as the scale factors are shown in Fig. 3·8. Note the appearance of steps in the efficiency at electron momenta of 115 and 200 GeV, which are a result of the addition of the high-momentum triggers to the logical OR expression.

As shown in the bottom panel of Fig. 3·8, the data to MC scale factors are derived from the two efficiency measurements from data and MC. Those correction factors

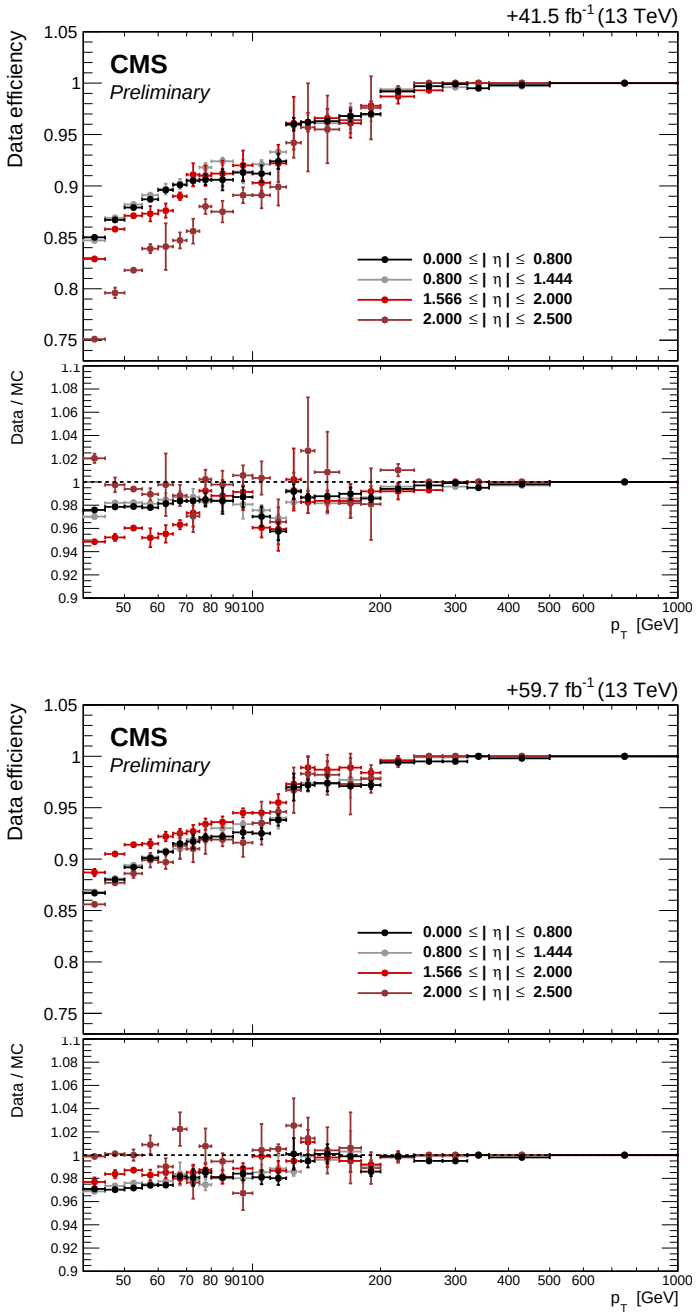

Figure 3·8: Efficiency of the OR of the three triggers used to select electron events for 2017 (top) and 2018 (bottom) as a function of the electron transverse momentum. The efficiency is shown for multiple regions of absolute electron pseudorapidity. In each plot, the upper panel shows the efficiency in data, while the lower panel shows the ratio of the efficiency in data and that in simulation.

are applied per-electron in the analysis to correct MC.

3.2.2 Pileup reweighting

The pileup (PU) conditions[6] in the simulated samples are not identical to the ones observed in data, and a per-event reweighting is applied to remove the difference. The reweighting is performed by matching the true pileup distribution of each simulated sample with the pileup distribution in data. The pileup distribution in data is obtained through the pileupCalc tool, assuming a minimum bias cross section of 69.2±4.6% mb, following the recommendations in [88]. The true pileup distributions in data and simulation are shown in Fig. 3·9, which show the normalized number of events as a function of the number of reconstructed collision vertices. The distribution of the number of reconstructed vertices for $W \to \mu\nu$ events before and after PU reweighting is shown in Fig. 3·10. The distribution of the event energy density ρ is shown in Fig. 3·11, again before and after PU reweighting. In terms of these two variables, it is observed that PU reweighting improves the agreement between data and MC.

[6]The term "pileup conditions" refers to the distribution of the number of reconstructed p-p collision vertices in a given p-p bunch crossing.

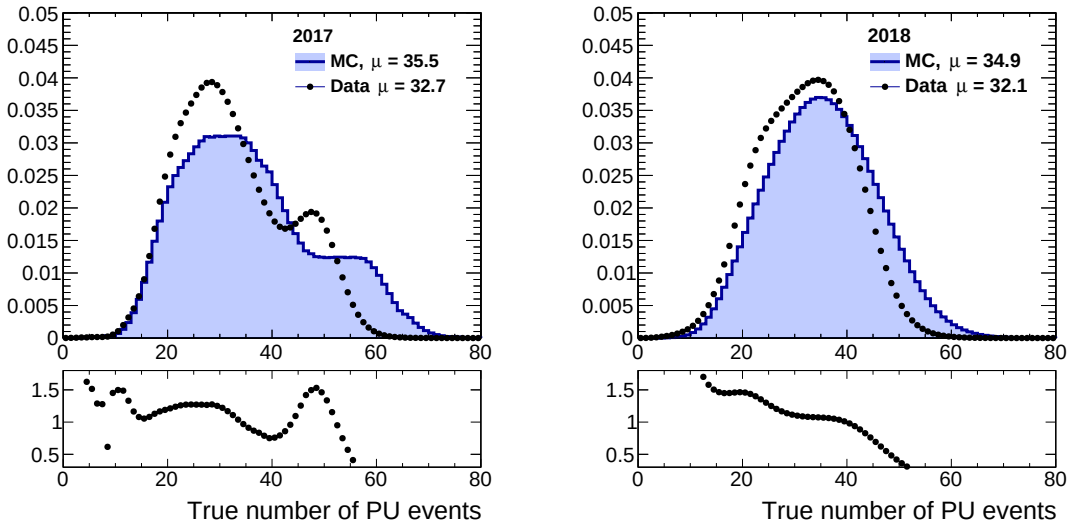

Figure 3·9: Distribution of the true number of PU events in data and simulation for 2017 (left) and 2018 (right). The distributions for data are extracted assuming a minimum bias cross section of 69.2 mb.

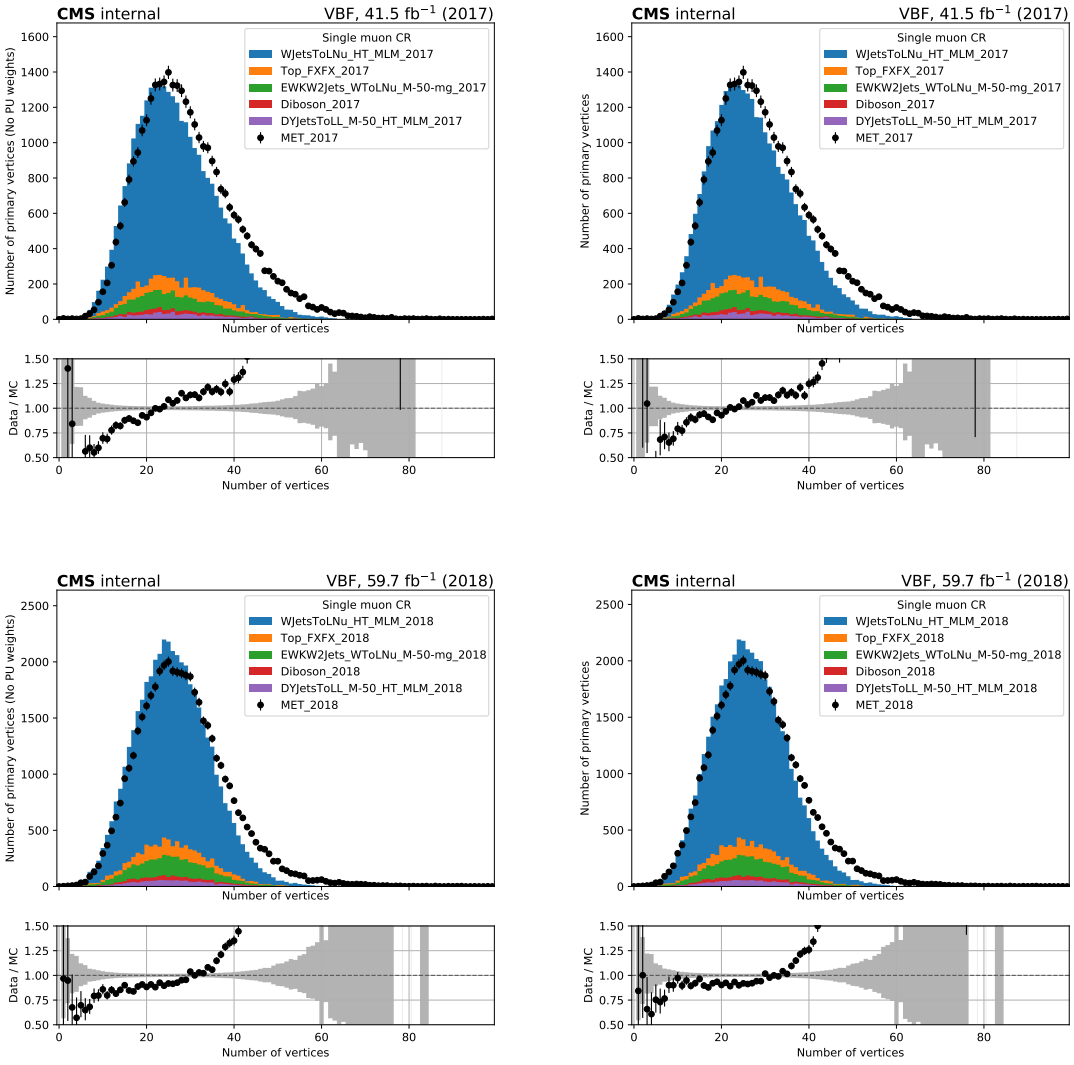

Figure 3·10: Distribution of the number of vertices in W $\to \mu\nu$ events in data and simulation before pileup re-weighting (left) and after pileup reweighting (right). The Monte Carlo is normalized to the luminosity of 41.5 and 59.7 fb^{-1}, respectively for 2017 and 2018.

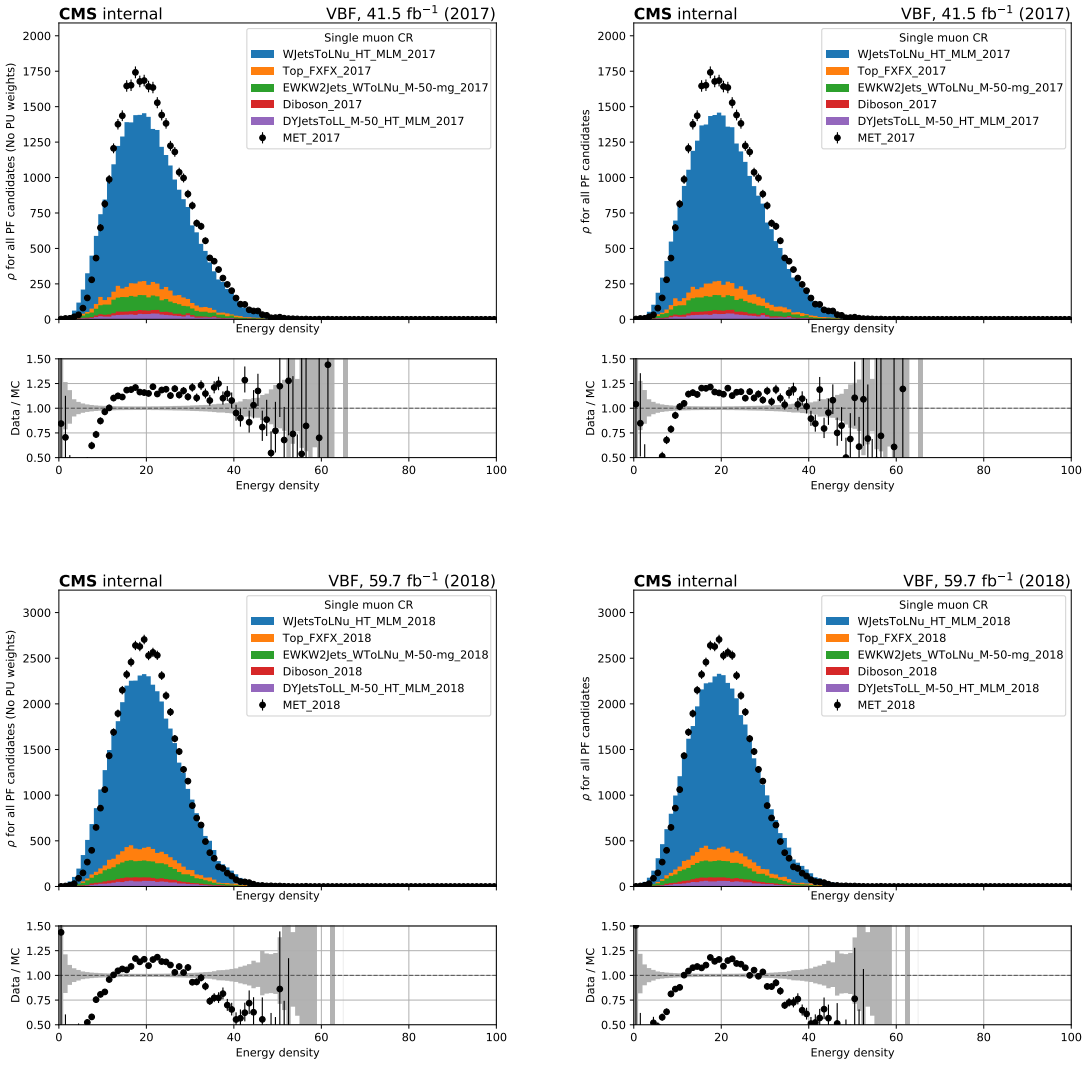

Figure 3·11: Distribution of the event energy density ρ in W $\rightarrow$ $\mu\nu$ events in data and simulation before pileup re-weighting (left) and after pileup reweighting (right). The Monte Carlo is normalized to the luminosity of 41.5 and 59.7 fb^{-1}, respectively for 2017 and 2018.

3.2.3 Lepton identification efficiency reweighting

Data-to-MC scale factors are applied to events in the control regions to account for differences in the reconstruction, identification and isolation of leptons between data and MC. These data-to-MC scale factors are derived from the efficiencies that are measured for the electron and muon selections in bins of p_T and η in both data and MC. The muon reconstruction and identification scale factors are provided by the relevant POGs. Electron identification scale factors are measured using a tag-and-probe method, and the results are reviewed and approved by EGamma POG.

The reconstruction scale factors for electrons are shown in Fig. 3·12. The corresponding identification scale factors for veto and tight electrons are shown in Fig. 3·13, and include the effect of the isolation efficiency.

The identification scale factors for muons are shown in Fig. 3·14. Here, isolation scale factors are applied separately and are shown in Fig. 3·15. The corresponding corrections for muons are deemed negligible [89].

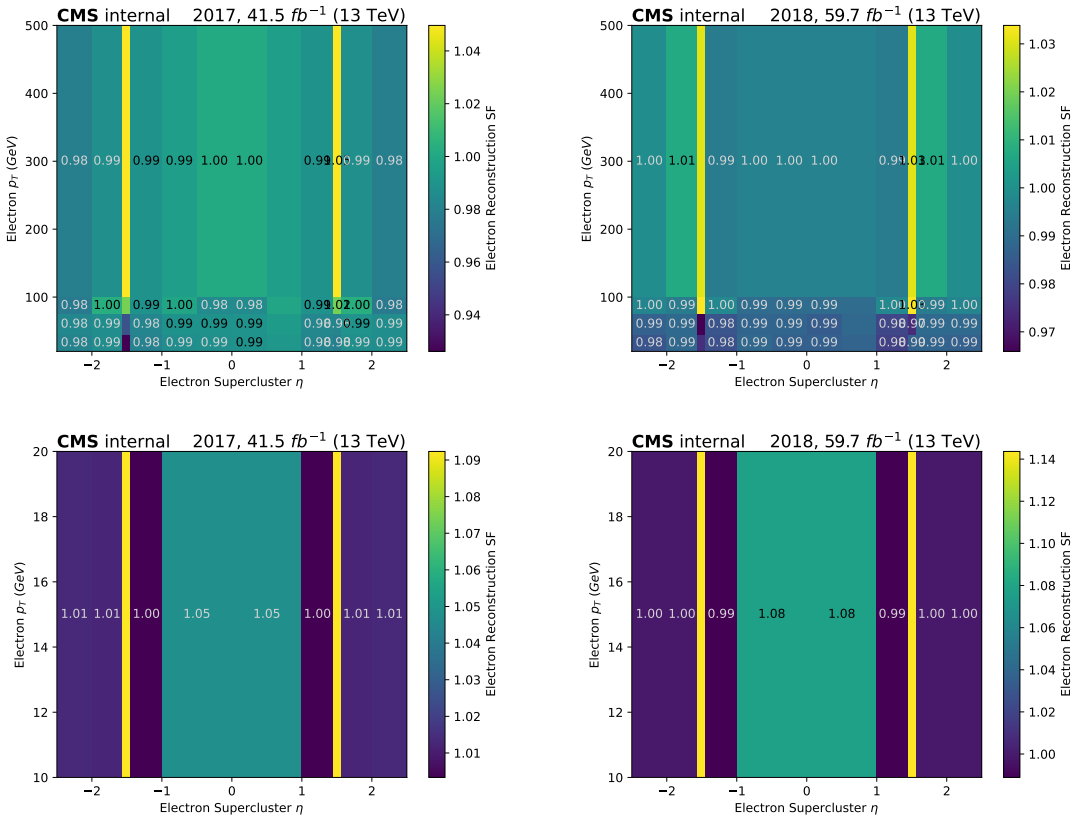

Figure 3·12: Scale factors for the reconstruction efficiency of electrons starting from a super cluster for 2017 (left) and 2018 (right), for electrons with $p_T > 20$ GeV (top) and $p_T < 20$ GeV (bottom).

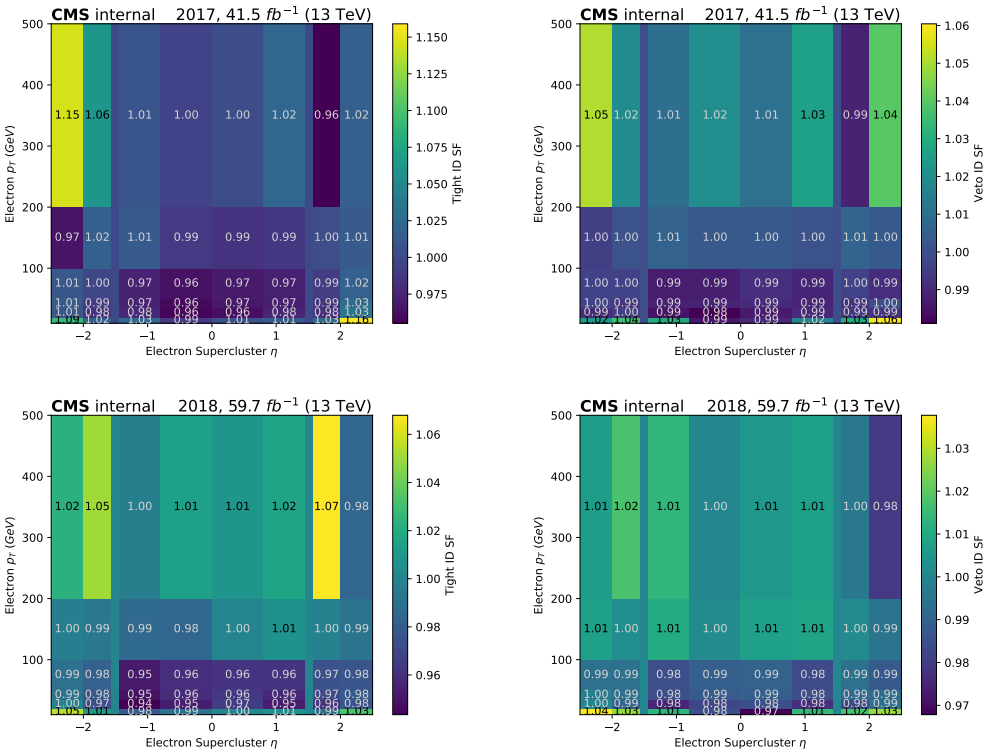

Figure 3·13: Scale factors for the identification efficiency of tight (left) and veto (right) electrons are shown for 2017 (top) and 2018 (bottom). The scale factors are provided in bins of electron p_T and η.

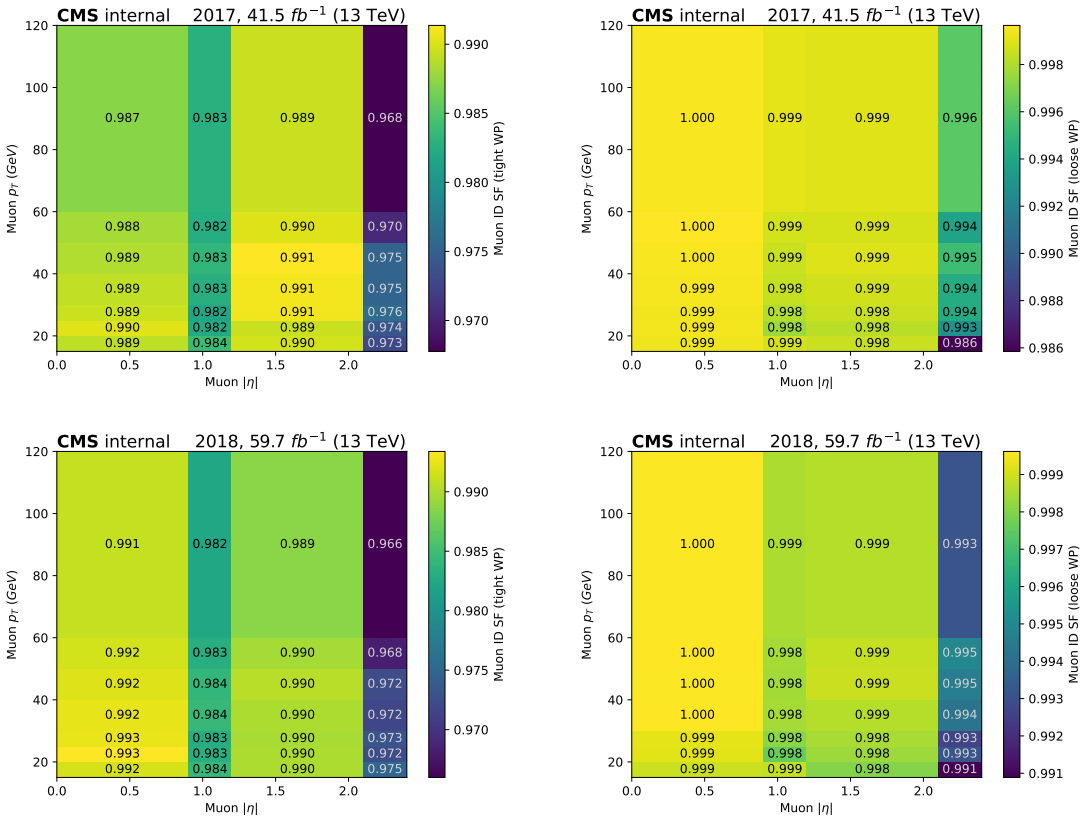

Figure 3·14: Scale factors for tight (left) and veto (right) muon identification are shown for 2017 (top) and 2018 (bottom). The scale factors are provided in bins of muon p_T and η.

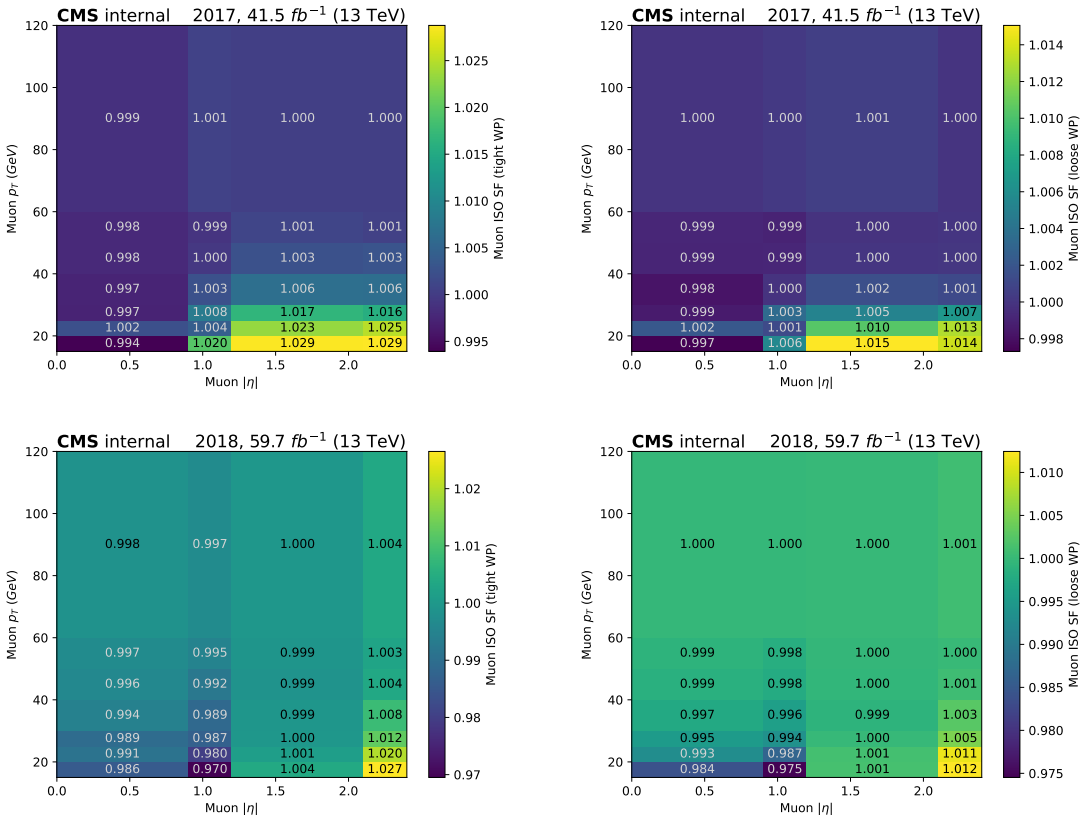

Figure 3·15: Scale factors for tight (left) and veto (right) muon isolation are shown for 2017 (top) and 2018 (bottom). The scale factors are provided in bins of muon p_T and η.

3.2.4 Lepton veto efficiency reweighting

To reduce contribution from $W(\ell\nu)$+ jet(s) events in the analysis signal region (SR), events are rejected if they contain any identified charged lepton. However, due to finite efficiencies of the reconstruction and identification of these candidates, $W(\ell\nu)$+ jet(s) events still contribute significantly. Furthermore, the differences in veto efficiency between data and MC need to be taken into account. For these purposes, the "veto-weight" method is used. In this method, simulated events in the SR are allocated a veto weight instead of being removed completely if they have charged lepton(s) in the final state. Such events with identified leptons can enter the SR with a veto weight dependent on the identification, isolation and reconstruction efficiency scale factors for the identified objects in the event. The veto weight ω is calculated as:

$$\omega = \prod_{i \ \epsilon \ leptons} (1 - SF_i) \tag{3.4}$$

where the product above runs over all identified veto leptons in the event (i.e. electrons, muons and taus). SF_i represents the total scale factor for the lepton i, which is the product of identification and isolation scale factors as described in Sec. 3.2.3. It can be observed that this method is equivalent to the ordinary method with hard-veto on events with leptons, as $SF \approx 1$ for events with well identified leptons, making the event weight, $\omega \approx 0$, hence suppressing the contribution of the event in the SR. Comparing the simulated event yields with hard lepton veto applied (instead of lepton veto weights), it is observed that the lepton veto weights in the signal region introduces a small correction $\ll 1\%$.

The uncertainty on these veto weights are computed by varying the veto weights within their uncertainties and propagating the variation into the M_{jj} distribution in the SR. Since the lepton kinematics are not largely correlated with the kinematics of two outgoing VBF jets, the uncertainties are observed to be flat as a function of M_{jj}.

3.2.5 Reweighting For HF Noise Cuts

As will be explained in Sec. 3.3.1, several noise-cleaning cuts are implemented to reduce the contribution from mis-reconstruction of forward jets, reconstructed in the Forward HCAL (HF) detector (i.e. $|\eta| > 3.0$). To account for the difference in the impact of these cuts in data and MC, a per-jet scale factor is applied to the events in MC. These scale factors are derived by calculating the efficiency of these cuts on $\mathrm{Z}(\mu\mu) + \text{jets}$ and γ + jets events in data and MC, as a function of jet p_T and η, and taking the ratio of those two efficiencies.

The HF noise scale factors are applied for each HF jet in the event which would be taken into account for the HF cuts (i.e. $p_T > 100$ GeV and $\Delta\phi(j, p_T^{miss}) > 2.5$), as a function of p_T and η of the jet. The event weight is calculated by multiplying the individual jet weights:

$$\omega_{event} = \prod_{jet \in HF} \omega_{jet}(p_T, \eta) \tag{3.5}$$

where the product runs over each jet with $p_T > 100$ GeV and $\Delta\phi(j, p_T^{miss}) > 2.5$[7]. The per-jet scale factors ω_{jet} are shown in Fig. 3·16 for 2017 and 2018, together with their statistical uncertainties. For the statistically limited phase space of very high p_T and η (where the scale factor is measured to be 0), the scale factor in the closest $[p_T, \eta]$ bin is applied to the HF jet.

[7]These cuts are required to be compatible with the application of HF cuts described in Sec. 3.3.1. $\Delta\phi > 2.5$ requirement on the jets being considered is utilized to pick jets that are reconstructed back-to-back (i.e. high $\Delta\phi$) with the missing transverse momentum

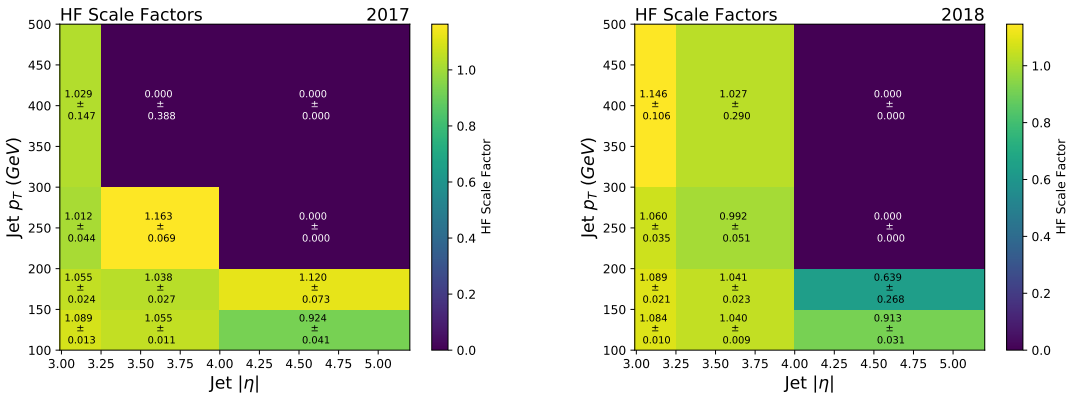

Figure 3·16: HF scale factors and their statistical uncertainties as a function of jet p_T and η, for 2017 (left) and 2018 (right). For the phase space with statistical limitations (i.e. $SF = 0$), the scale factor in the closest $[p_T, \eta]$ bin is applied to the HF jet.

3.2.6 Reweighting for ECAL pre-firing

Prefiring is a problem with the 2017 dataset that results from L1 trigger primitives in the ECAL endcap being incorrectly assigned to an earlier bunch crossing because of a timing issue. In most cases, the event to which the trigger primitives are now assigned to fail to pass the HLT selection and is discarded. This makes the event for which the trigger primitives were originally belonged to be discarded as well, which results in a loss of trigger efficiency. A solution has been developed by parametrizing the probability of a jet or photon in the event causing prefiring in terms of their p_T and η. In this analysis, the prefiring weights are applied based on the parametrization provided by the JME POG [90], and a central implementation provided in the NanoAOD-tools software is used [91]. In this implementation, a per-event prefiring weight is computed as:

$$\omega = 1 - P(prefiring) = \prod_{i=photons,jets} (1 - \epsilon_i^{pref}(p_T, \eta)) \tag{3.6}$$

where ϵ_i^{pref} is the prefiring probability caused by a single photon or a jet, and the product runs over the photons and jets reconstructed in the event. The M_{jj} distributions with the nominal prefiring weight applied to simulation, together with the up and down variations of the prefiring weight are shown in Figs. 3·17 and 3·18. It can be observed that the uncertainty due to the prefiring weight increases with higher M_{jj}. It can be argued that this is an expected effect because of the origins of the prefiring problem, which is much more sizable for jets in the ECAL endcap region, $2.5 < |\eta| < 3.0$. Therefore, events with at least one forward jet is much likely to get impacted by prefiring. Since M_{jj} is correlated with $\Delta\eta_{jj}$ (i.e. the η separation between the two VBF jets), events with higher M_{jj} are also more impacted, and hence the relevant uncertainty is more pronounced at high M_{jj}.

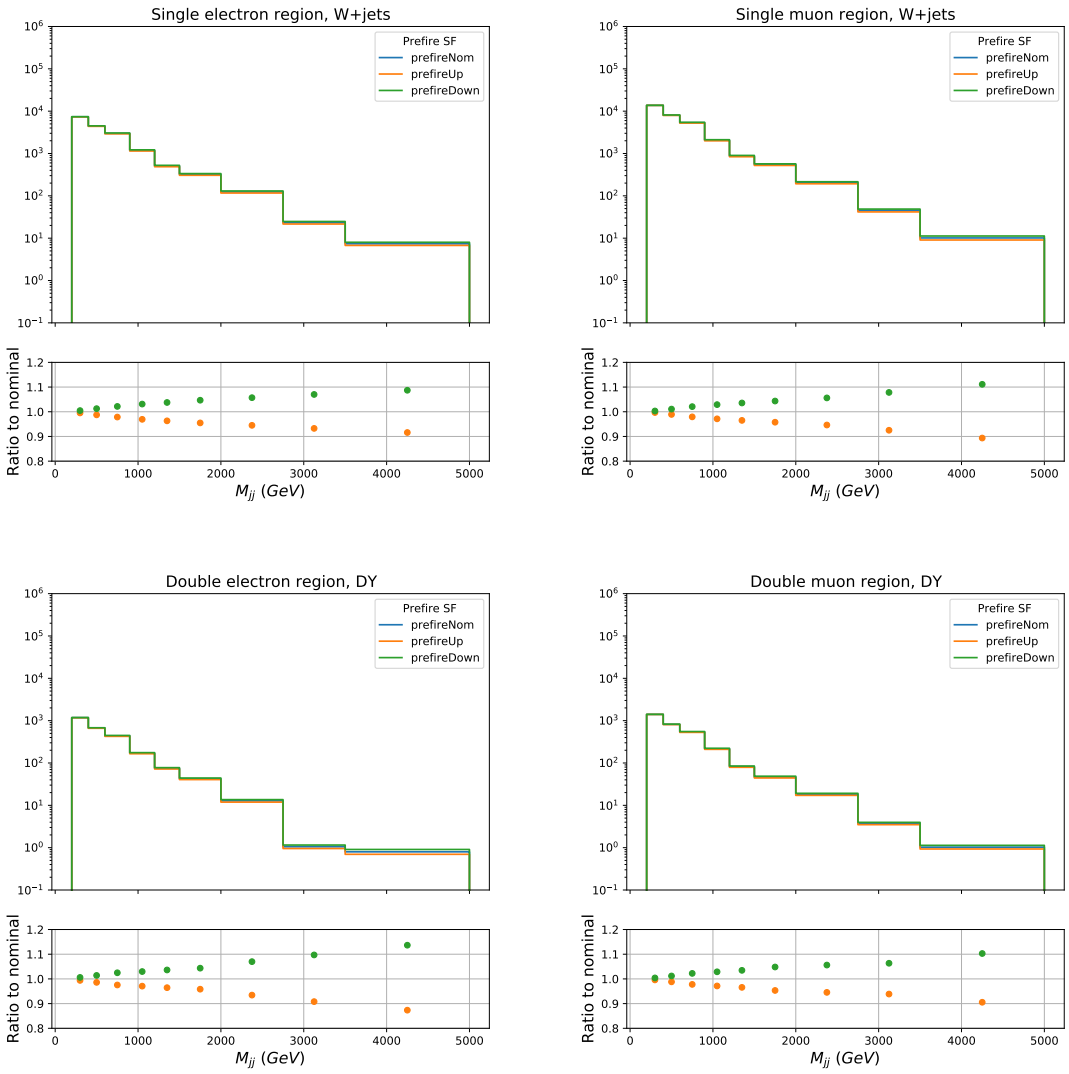

Figure 3·17: The M_{jj} distributions with the nominal prefiring weights and its variations are shown for W + jets simulation in single lepton CRs (top), and Z + jets simulation in double lepton CRs (bottom).

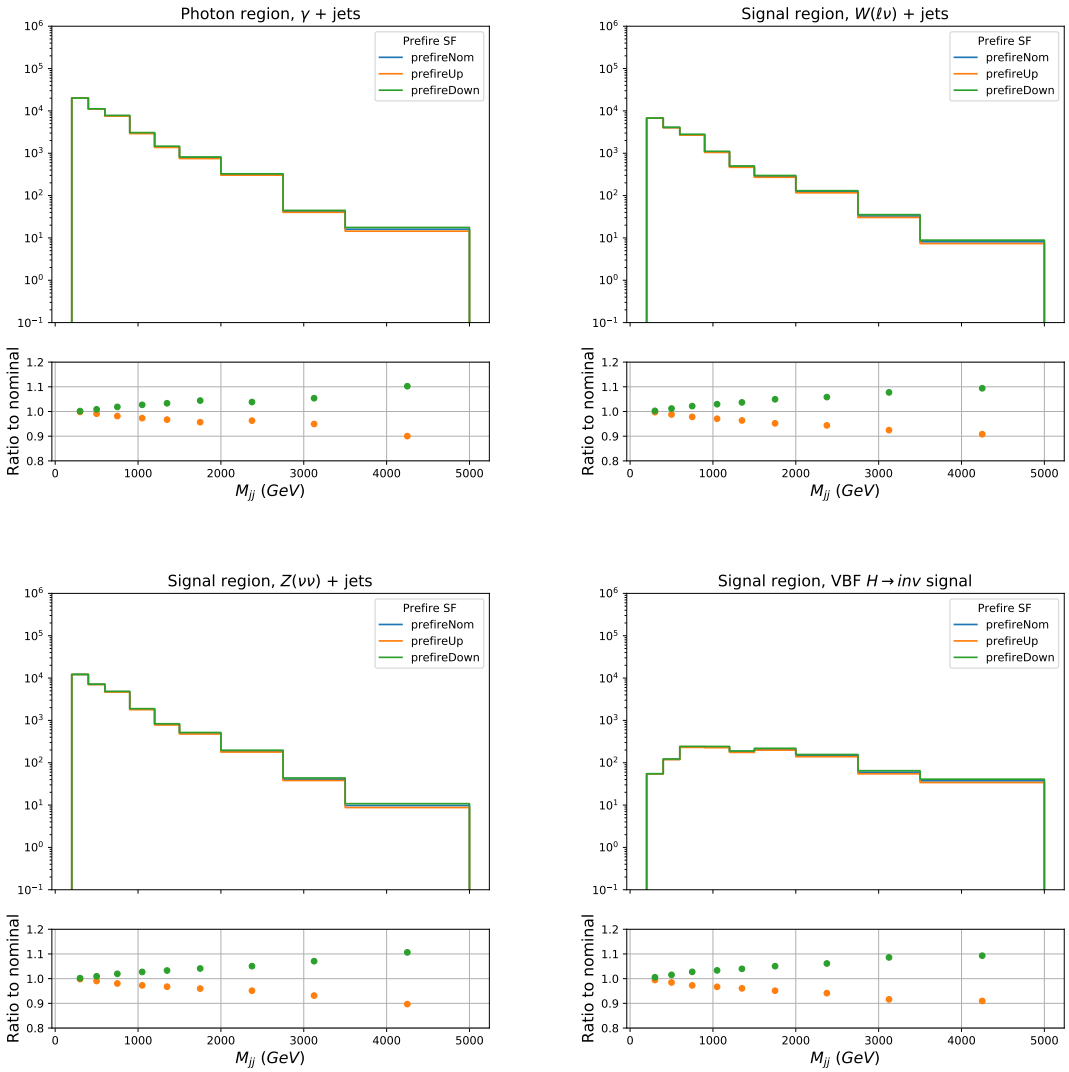

Figure 3·18: The M_{jj} distributions with the nominal prefiring weights and its variations are shown for γ + jets simulation in the photon CR (top left), $W(\ell\nu)$ + jets (top right), $Z(\nu\nu)$ + jets (bottom left) and VBF $H \to inv.$ (bottom right) in the signal region.

3.2.7 Higher-order reweighting

As will be detailed in Sec. 3.5.1, this analysis uses the ratios of V+jets distributions in signal and control regions to constrain the final background estimate for $\mathrm{Z}(\nu\nu) + \mathrm{jets}$ and $\mathrm{W}(\ell\nu) + \mathrm{jets}$ processes. As signal and control regions both have large statistical power, precise predictions of these ratios are necessary. To achieve this goal, next-to-leading-order (NLO) in QCD simulation samples are used for $\mathrm{W}(\ell\nu) + \mathrm{jets}$ and $\mathrm{Z}(\ell\ell) + \mathrm{jets}$ backgrounds, and corrected with higher-order EWK corrections. To model the γ + jets background in photon control region, leading-order (LO) in QCD simulation samples are used instead, which are then reweighted using higher-order QCD and EWK corrections. These corrections are described in more detail in the following subsections. A concise overview of which corrections are applied to which processes is given in Tab. 3.9.

Table 3.9: Summary of higher-order corrections applied to simulated samples. For each boson production process, separate samples and corrections are available for the EWK and QCD production modes. "MC order" reflects the perturbative order used in the generation of the simulation sample, while the further columns represent corrections applied on a per-event level in the analysis process.

Boson	Production mode	MC order	NLO QCD	NLO EWK
Z	QCD	NLO	–	✓
	EWK	LO	✓	–
W	QCD	NLO	–	✓
	EWK	LO	✓	–
γ	QCD	LO	✓	✓
	EWK	LO	–	–

Generator-level boson construction

All theory-based corrections of the W($\ell\nu$)+jets, Z($\ell\ell$)+jets, and γ + jets backgrounds are parametrized as a function of the generator-level transverse momentum of the respective boson, $p_{\mathrm{T},V}$. For W($\ell\nu$) + jets and Z($\ell\ell$) + jets events, this quantity is calculated as follows:

1. If a boson is found in generator-level collection with `status = 62`, it's $p_{\mathrm{T},V}$ is taken as the generator level $p_{\mathrm{T},V}$. (If multiple such entries are found, highest $p_{\mathrm{T},V}$ is chosen.)

2. In the rare cases where the boson is not found in the generator-level collection, then the boson is defined as the four-vector sum of the selected leptons. Leptons are selected as described in the next two steps:

3. If there are electrons or muons that are not coming from a tau decay, these leptons are selected to compute $p_{\mathrm{T},V}$.

4. If the electrons or muons are found to be coming from a tau decay, or if there are no leptons in the event, the generator-level taus are chosen to compute $p_{\mathrm{T},V}$.

For γ + jets events, the photon at generator-level is selected by requiring the following:

- Photon needs to have *status = 1*, indicating that it is a final state particle (i.e. a particle that is not decayed further by the generator).

- Photon needs to have a prompt status flag.

- $|\eta| < 1.46$, so that the photon is in the barrel region.

If there are multiple such photons, the one with highest p_T is selected.

Photon scale factors

The photon scale factors are derived with a two-dimensional dependence on generator-level $p_{\mathrm{T},V}$ and M_{jj}, and are shown for the MTR category in Fig. 3·19. These scale factors are applied to γ + jets events as a NLO correction, and the correction depends on the kinematics of the event.

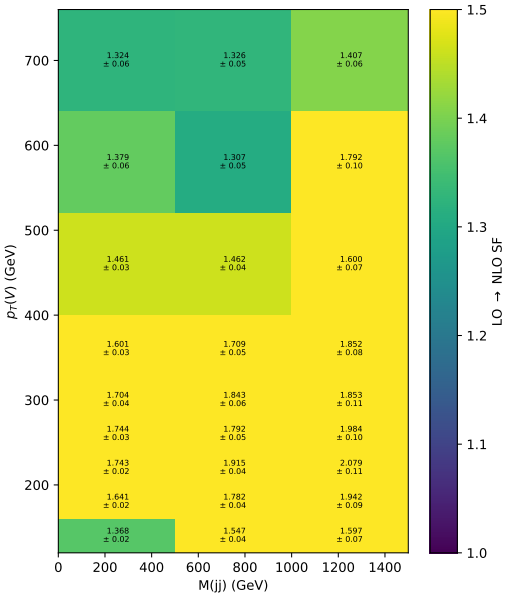

Figure 3·19: LO-to-NLO theory scale factors binned in the generator-level boson p_T and M_{jj}, shown photon production. The k factors are derived within the generator-level VBF selection described in the text. The uncertainties quoted in each bin are the statistical uncertainties due to the finite size of the simulated samples.

EWK NLO corrections to QCD V processes

Scale factors corresponding to NLO EWK corrections are obtained from Ref. [92] and applied as a function of the generator-level boson p_T to each event. The scale factors are shown in Fig. 3·20.

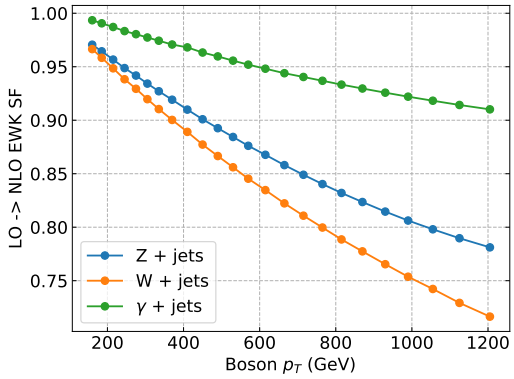

Figure 3·20: EWK NLO scale factors for DY, W and photon production as a function of $p_{\mathrm{T},V}$.

QCD NLO corrections to EWK V processes

The QCD NLO corrections to VBF W + jets and Z + jets production have been calculated in Ref. [93] using the VBF@NLO program. They are parametrized in $p_{\mathrm{T},V}$ and M_{jj} and are shown in Fig. 3·21.

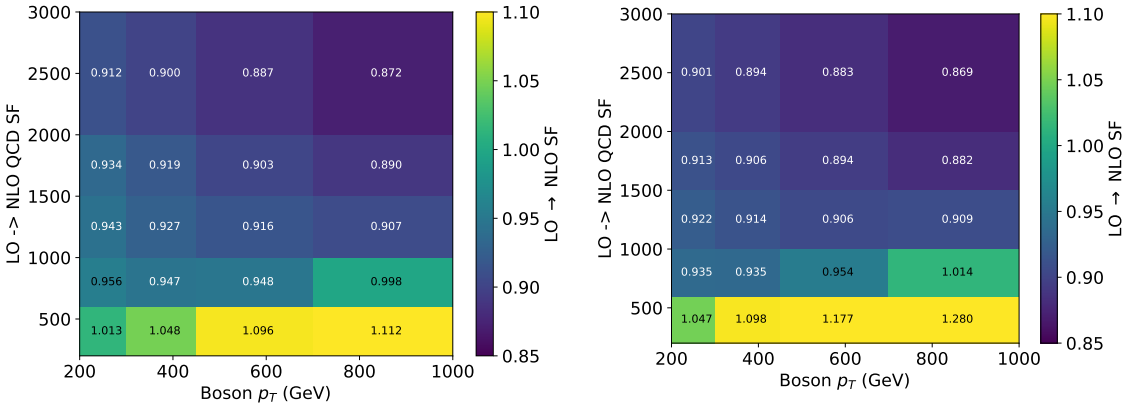

Figure 3·21: QCD NLO scale factors for EWK DY, W production of $p_{\mathrm{T},V}$ and M_{jj}.

NLO EWK Corrections on VBF Signal

Next to leading order (NLO) EWK corrections on VBF signal are computed using the HAWK generator [94]. The NLO corrections are computed in the VBF phase space, where events are selected based on the kinematics of the two final state VBF jets. These kinematic selections are described in Sec. 3.4. The correction is calculated as a function of the p_T of the Higgs boson, and is parametrized using the following fit function:

$$\epsilon_{EWK}(p_T^H) = (1 - 0.000372 * p_T^H - 0.0304)/0.95 \tag{3.7}$$

For each event in the VBF signal sample, this correction is applied based on the p_T of the Higgs boson found in the generator-level collection.

3.3 Data-quality issues

3.3.1 Forward HCAL (HF) Noise

The forward HCAL (HF) detector covers the most-forward rapidity range of HCAL ($3 < |\eta| < 5$), as described in Sec. 2.2.4. Due to its very forward location and close proximity to the beam pipe, HF detector operates in a much harsher radiation environment. It should also be noted that the tracking detector, covering $|\eta| < 2.5$ range, does not extend to this high rapidity phase space. Combination of these different factors makes it harder to reconstruct and cluster hadronic jets, and measure their energies accurately at the HF detector. Therefore, there is a sizable probability that the p_T of a jet reconstructed at high $|\eta|$ being highly mismeasured. Such mismeasurements can lead to spurious p_T^{miss} in QCD multijet events (e.g. $q\bar{q} \to q\bar{q}$), due to p_T of one of the final state jets being severely mismeasured. It should be noted that while the probability of such severe mismeasurement is relatively small, the very large cross section of QCD multijet processes in a p-p collider compensates for this, and the

number of events with spurious p_T^{miss} cannot be ignored. This becomes problematic for analyses that target p_T^{miss} (such as analyses targeting $H \to inv.$ signature), since mismeasured QCD multijet events can pass the high p_T^{miss} requirement imposed on the analysis signal region. Hence, this effect needs to taken into account.

It was observed that jets resulting from such mismeasurement present a distinguishable shape with respect to well-identified jets: Their shower shape is spread in η and narrow in ϕ. Shower shape variables of HF jets ($\sigma_{i\eta i\eta}$, $\sigma_{i\phi i\phi}$, HF central strip size) have therefore been introduced and added to the event content in the input datasets. $\sigma_{i\eta i\eta}$ represents the shower width of the HF jet in η direction, and similarly, $\sigma_{i\phi i\phi}$ represents the shower width in ϕ direction. Their definitions can be found below:

$$\begin{aligned} \sigma_{i\eta i\eta} &= \sqrt{\frac{\sum_i \Delta\eta(i,jet)^2\omega_i}{\sum_i \omega_i}} \\ \sigma_{i\phi i\phi} &= \sqrt{\frac{\sum_i \Delta\phi(i,jet)^2\omega_i}{\sum_i \omega_i}} \end{aligned} \tag{3.8}$$

In Eq. 3.8, the sums run over all PF candidates with $p_T > 3$ GeV, and ω_i is the weight applied to PF candidate i within the jet, which is computed as follows:

$$\omega_i = \frac{p_T^i - p_T^{\text{PU offset}}}{\sum_i p_T^i - p_T^{\text{PU offset}}} \tag{3.9}$$

In Eq. 3.9, $p_T^{\text{PU offset}}$ refers to the offset correction on the p_T of the particle due to pileup (PU). It is defined as follows:

$$p_T^{\text{PU offset}} = \frac{\text{jet L1 offset / gen vtx}}{\pi \times 0.4^2} \times \frac{N_{\text{reco vertices}}}{\epsilon^{reco}(\text{PU vtx})} \times S_{\text{HF tower}} \tag{3.10}$$

The other distinguishing variable, HF central strip size (CSS_{HF}), is computed as follows: N_{cands} with $p_T > 10$ GeV & $|\Delta\phi(cand, jet)| < 0.05$. It is a measure of the number of highly energetic particles reconstructed close to the jet centroid (i.e.

$|\Delta\phi(cand, jet)| < 0.05$). In the event of spurious large energy deposits from a jet, one can expect larger number of energetic particles being reconstructed closer to the jet centroid, as a result of those energy deposits.

These variables are used to define a series of cuts rejecting most of the HF-related noise, while keeping a high efficiency on well-identified jets. The efficiency of the cuts on physics jets in data and simulation is measured and their ratio is used to correct the simulation. This procedure is described in Sec. 3.2.5. After the HF noise cuts are applied, the residual noise in the signal region is estimated using a data driven technique. The study and the definition of the HF noise cuts, and the residual HF noise estimation are discussed in the subsections below.

Definition of HF Noise Cuts

In order to study the discriminating power of these variables directly in data, two regions enriched in "physics" jets and "noise" jets are defined. "Physics" jets refer to HF jets that are recoiling against a well-reconstructed physics object (in this study, a photon), and "noise" jets refer to HF jets that recoil against no visible physics object, hence creating p_T^{miss} in the event. The region where the "physics" jets are tagged is termed as the physics-enriched region, and the region where the "noise" jets are tagged is termed as the noise-enriched region.

The physics-enriched region consists of events where a HF jet recoils against a photon. Jet and photon being back-to-back is ensured by a series of cuts on $\Delta\phi(\gamma, jet)$, ratio of transverse momenta of the photon and the jet, and also the requirement of low p_T^{miss} in the event. The selections are listed below:

- Exactly one well identified photon (tight identification, as discussed in Sec. 3.1.4)
- Events passing a single photon trigger requiring $p_T^{\gamma} > 110$ GeV
- $\Delta\phi(\gamma, jet) > 2.7$

- $0.5 < \frac{p_T(\gamma)}{p_T(jet)} < 1.5$
- $p_T^{miss} < 50$ GeV

The noise-enriched region includes events with exactly one high p_T HF jet and high p_T^{miss}. The aim here is to collect events with a mis-measured HF jet, such that it has very high p_T, but there is no other high p_T physics object in the event, giving rise to p_T^{miss} in the event. The complete requirements for this region are listed below:

- Events passing the MET trigger described in Sec. 3.2.1, requiring $p_T^{miss} > 120$ GeV and $H_{\mathrm{T}}^{\mathrm{miss}} > 120$ GeV (as reconstructed by the trigger algorithm).
- $p_T^{miss} > 100$ GeV
- Exactly one HF jet with $p_T > 100$ GeV
- No extra jet with $p_T > 30$ GeV
- No lepton with $p_T > 10$ GeV

The two regions outlined above are used to study the distribution of HF shower shape variables for well-identified jets, as opposed to events containing a mis-measured ("noisy") jet. Mainly, two sets of cuts are studied and adopted in the analysis:

- Cuts on $\sigma_{i\eta i\eta}$ and $\sigma_{i\phi i\phi}$
- Cut on HF central strip size

Since the HF-related noise strongly peaks in the $3 < |\eta| < 3.25$ region and the size of the HF towers is $|\eta|$ dependent, possibly affecting the variables under study, three $|\eta|$ regions are considered. Two dimensional distributions of $\sigma_{i\phi i\phi}$ vs $\sigma_{i\eta i\eta}$ are shown in Fig. 3·22 and Fig. 3·23, for the noise and physics enriched regions respectively. From Fig. 3·22, the skewness of the 2D distribution towards high $\sigma_{i\eta i\eta}$ and low $\sigma_{i\phi i\phi}$ can be

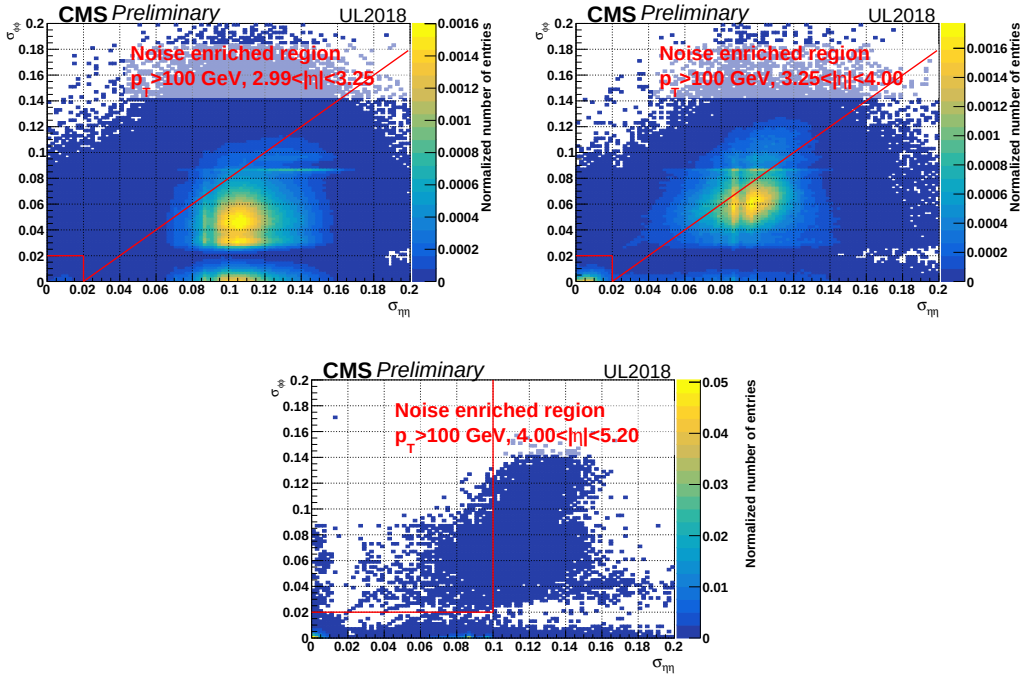

Figure 3·22: Two dimensional distribution of $\sigma_{i\eta i\eta}$ and $\sigma_{i\phi i\phi}$ in the noise-enriched region, split by the $|\eta|$ of the jet. The first plot shows $2.99 < |\eta| < 3.25$ interval, the second one shows $3.25 < |\eta| < 4$ and the last one shows $4 < |\eta| < 5.2$. The red lines on the plots indicate the cuts applied on these variables.

observed, specifically for $|\eta| < 4$. This motivates the diagonal cut of $\sigma_{i\eta i\eta} - \sigma_{i\phi i\phi} < 0.02$, cutting away a sizable number of noisy events while having much less impact in the physics-enriched region.

In the noise enriched region, there are also a large number of events observed within the region defined by $\sigma_{i\eta i\eta} < 0.02$ and $\sigma_{i\phi i\phi} < 0.02$. These are thought to be interactions of halo muons in the HF detector and this region with very low $\sigma_{i\eta i\eta}$ and $\sigma_{i\phi i\phi}$ is therefore also vetoed.

For jets with $|\eta| > 4$, the requirement becomes $\sigma_{i\eta i\eta} < 0.1$ and $\sigma_{i\phi i\phi} > 0.02$, as large numbers of noise events are observed at very low $\sigma_{i\phi i\phi}$ and high $\sigma_{i\eta i\eta}$.

The second cut is applied on the HF central strip size (CSS_{HF}) of the jet. The two-dimensional distributions of central and adjacent strip sizes for the noise enriched

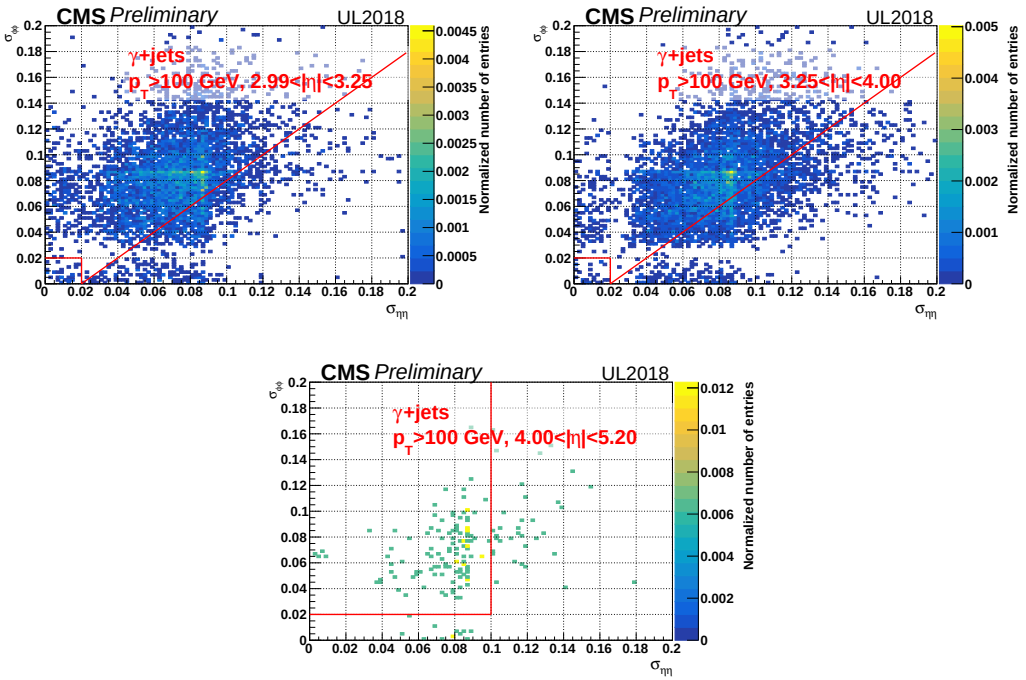

Figure 3·23: Two dimensional distribution of $\sigma_{i\eta i\eta}$ and $\sigma_{i\phi i\phi}$ in the physics-enriched region, split by the $|\eta|$ of the jet. The first plot shows $2.99 < |\eta| < 3.25$ interval, the second one shows $3.25 < |\eta| < 4$ and the last one shows $4 < |\eta| < 5.2$. The red lines on the plots indicate the cuts applied on these variables.

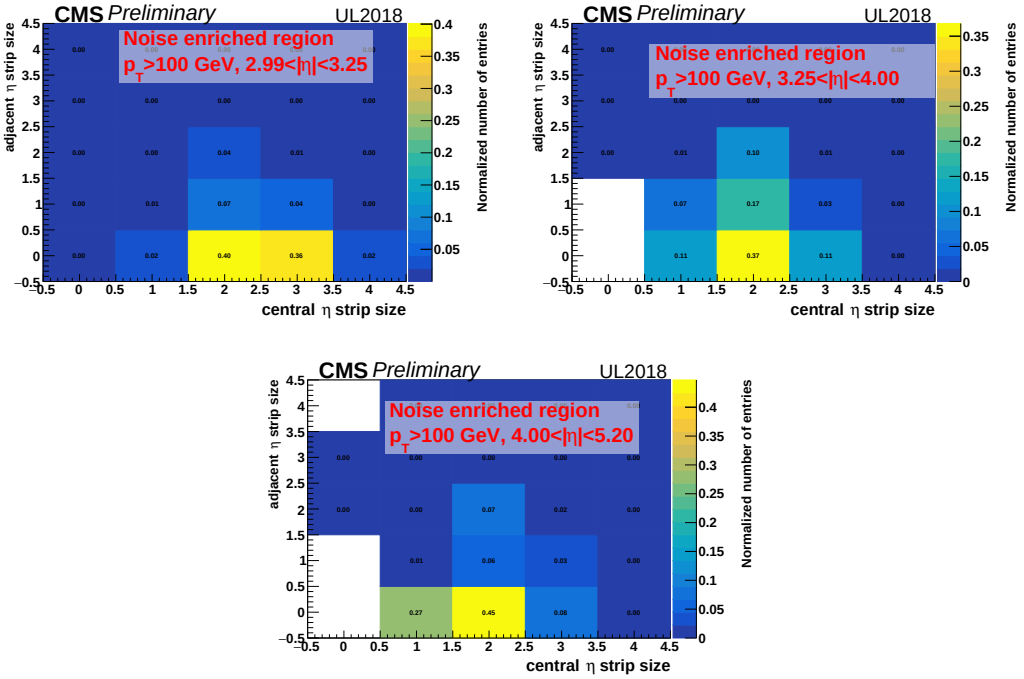

Figure 3·24: Two dimensional distribution of central and adjacent strip sizes in the noise enriched region, split by the $|\eta|$ of the jet. The first plot shows $2.99 < |\eta| < 3.25$ interval, the second one shows $3.25 < |\eta| < 4$ and the last one shows $4 < |\eta| < 5.2$.

region are shown in Fig. 3·24, while the same distributions in the physics enriched region are shown in Fig. 3·25. For the jets in noise enriched region with $2.99 < |\eta| < 3.25$ (where the noise contribution is the highest), it can be observed that around 40% of jets have $CSS_{HF} \geq 3$, while in the physics enriched region, this fraction is very small, $\mathcal{O}(1\%)$. Therefore, a cut on the CSS_{HF} variable is adopted in the analysis, requiring $CSS_{HF} < 3$ for HF jets that are back to back with p_T^{miss}.

In summary, the cleaning cuts applied to HF jets are:

- $2.99 < |\eta| < 4.0$: $\sigma_{i\eta i\eta} - \sigma_{i\phi i\phi} < 0.02$ condition is required. In addition, the corner region, defined as $\sigma_{i\eta i\eta} < 0.02$ & $\sigma_{i\phi i\phi} < 0.02$, is also removed from the analysis.

- $|\eta| > 4.0$: $\sigma_{i\eta i\eta} < 0.1$ & $\sigma_{i\phi i\phi} > 0.2$ condition is required.

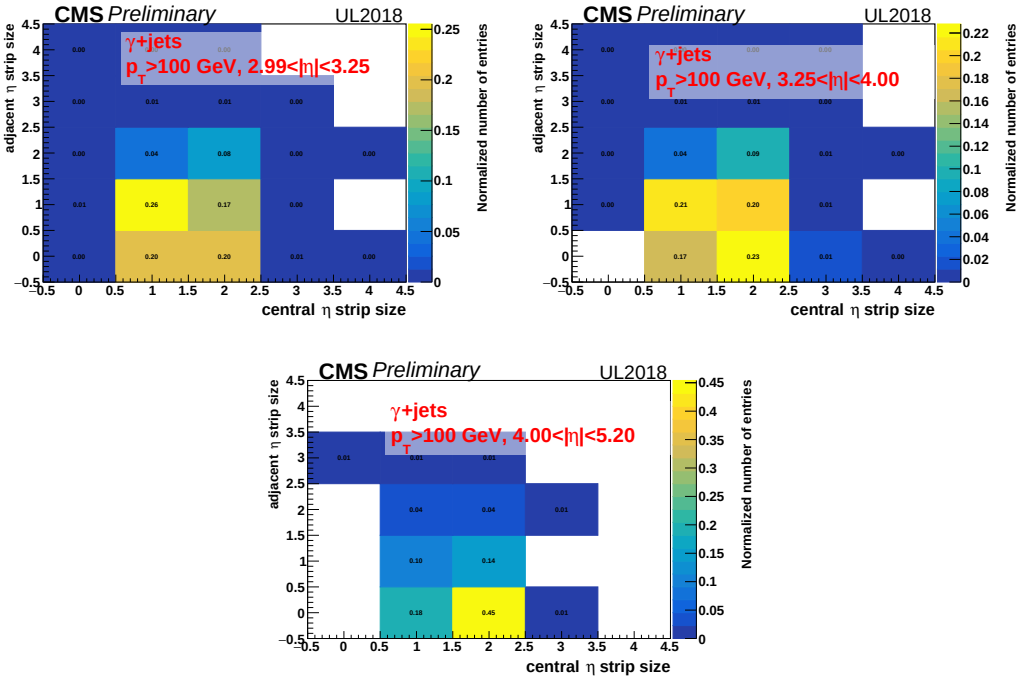

Figure 3·25: Two dimensional distribution of central and adjacent strip sizes in the physics enriched region, split by the $|\eta|$ of the jet. The first plot shows $2.99 < |\eta| < 3.25$ interval, the second one shows $3.25 < |\eta| < 4$ and the last one shows $4 < |\eta| < 5.2$.

- $|\eta| > 2.99$: $CSS_{HF} < 3$ is required.

They are applied to all jets with $p_T > 100$ GeV and $|\eta| > 2.99$ that are back-to-back with p_T^{miss} (i.e. $\Delta\phi(jet, p_T^{miss}) > 2.5$). If any such jet fails the cuts, the event is rejected.

Residual HF Noise Estimation

In addition to the HF cleaning cuts explained in the previous subsection, a noise estimation is done to estimate the leftover noise contribution in the analysis signal region. For the noise estimation, a new control region is defined in which the HF shape cuts are inverted. Other than the HF shape cuts, the same cuts are required in this control region as in the VBF signal region, which are discussed in Sec. 3.4.1. For the remainder of this text, this control region will be referred to as HF control region.

With the addition of this HF control region (CR), the estimation is done in three steps. First, data and MC yields in this CR are computed (N_{data}, N_{MC}). The excess amount of data over MC in this control region gives the number of excess events in data with HF noise, not modeled by the simulation: $N_{noise}^{CR} = N_{data} - N_{MC}$. Finally, the noise estimate in analysis signal region is calculated from N_{noise}^{CR} by using $[p_T, \eta]$ dependent transfer factors, defined as the relative probability for a noise event to pass the HF shape cuts, relative to failing them. This probability is derived from events in a noise-enriched region, where the probability of a jet in such an event to pass the HF cuts can be computed as a function of its pseudorapidity and transverse momentum, $P(\text{pass}|\text{noise})[p_T, \eta]$. Using these probabilities, the transfer factor R_{jet} for a single jet can be then written as:

$$R_{jet}(p_T, \eta) = \frac{P(\text{pass}|\text{noise})}{P(\text{fail}|\text{noise})}(p_T, \eta) = \frac{P(\text{pass}|\text{noise})}{1 - P(\text{pass}|\text{noise})}(p_T, \eta) \tag{3.11}$$

For a given event, the total transfer factor (taking all HF jets into account in the event) can be then computed as

$$R_{event} = \prod_{jet} R_{jet}(p_T, \eta) \tag{3.12}$$

where the product runs over the jets in the event with $p_T > 100$ GeV and $|eta| > 2.99$ (i.e. HF jets). If there are no jets in the event that satisfy this criteria, R_{event} is defined to be 0, signifying the fact that there will be no contribution to the signal region noise estimate template from this event. If there is such a jet however, $R_{event} \neq 0$, hence signifying a non-zero contribution to the noise estimation in the signal region.

In this way, the contribution of a single event in the noise-enriched CR to the HF noise template in the signal region can be written as

$$N^{SR}_{noise} = R_{event} \times (N^{CR}_{data} - N^{CR}_{MC}) \tag{3.13}$$

Eq. 3.13 shows the contribution from a single event in the HF control region. As the final step, summing up the contributions from all events in the HF control region gives the full M_{jj} noise estimate template for the HF noise, which are shown in Fig. 3·26. Plots on the left column show the data and total MC yields in the HF control region, each event being already scaled by the transfer factor R_{event} as defined above. Plots on the right column show the resulting HF noise estimation in the analysis signal region, which correspond to the difference of data and MC yields in the left-hand side plots. Plots on the top show the results with 2017 data, and plots on the bottom show results with 2018 data.

From Fig. 3·26, it can be observed that the HF noise estimate in the signal region is mostly flat as a function of M_{jj}, and can make a sizable contribution at higher M_{jj} values where most of the other background processes are suppressed.

To make sure that this noise estimation provides an accurate modeling of the

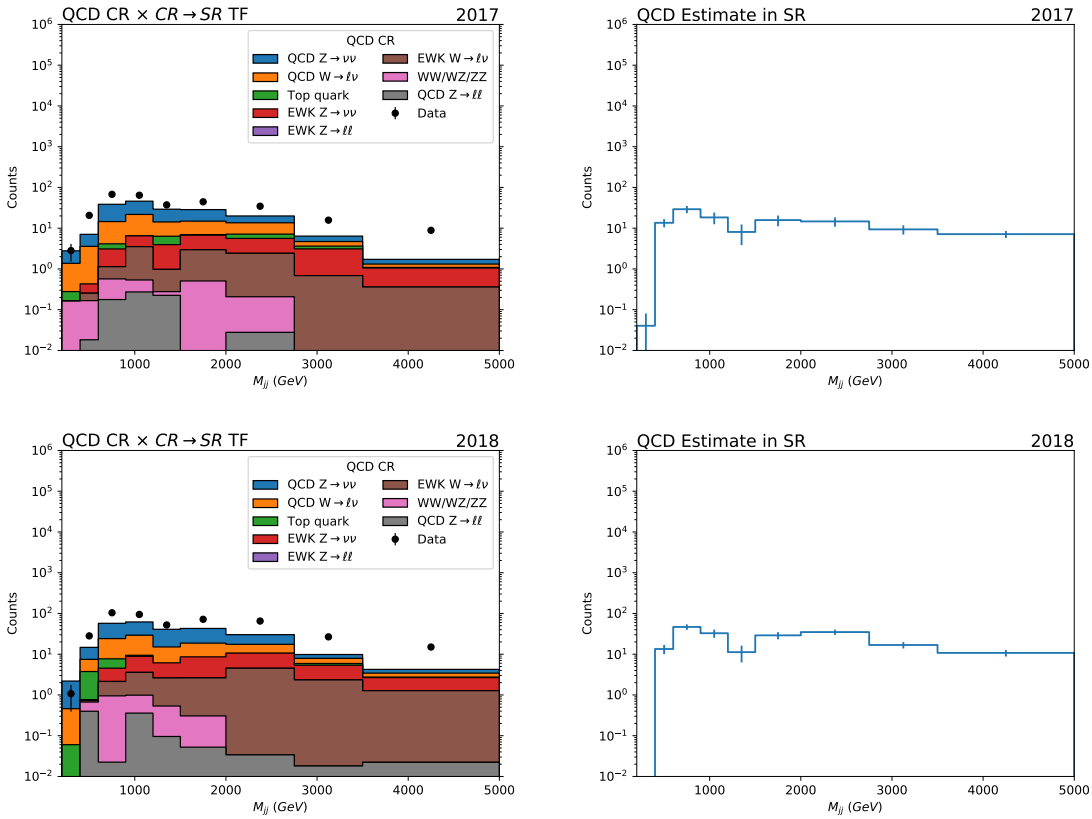

Figure 3·26: HF noise estimate in 2017 (top) and 2018 (bottom) data. The plots on the left column show the data and total MC yields in the HF noise control region, where each event is scaled by a transfer factor R_{event}, as defined in the text. Plots on the right show the resulting noise estimation in the signal region, which correspond to the difference of data and MC yields in the left-hand side plot.

data in the phase space where HF noise contribution is highest, $3 < |\eta| < 3.25$, a closure test was performed. A subset of events in VBF signal region were picked, which have the leading jet in the $3 < |\eta| < 3.25$ range, and the data yields are compared to the expected background yields in this region, including the HF noise estimation. For these events, Fig. 3·27 shows the $\Delta\phi(p_{T,trk}^{miss}, p_T^{miss})$ distribution, where $p_{T,trk}^{miss}$ refers to the p_T^{miss} computed only relying on the tracks reconstructed by the tracker. The HF noise contribution to the background estimation is expected to be especially significant at high $\Delta\phi$ values, since such events have a forward jet with

$|\eta| > 3$, which is well outside the tracker range. Due to the presence of such jets, large $\Delta\phi$ between the tracker-only p_T^{miss} and full p_T^{miss} is expected. It should also be noted that for this high $\Delta\phi$ region, the VBF $H \rightarrow inv.$ signal contribution is expected to be small, hence it would be expected that the data and background yields agree within a reasonable extent.

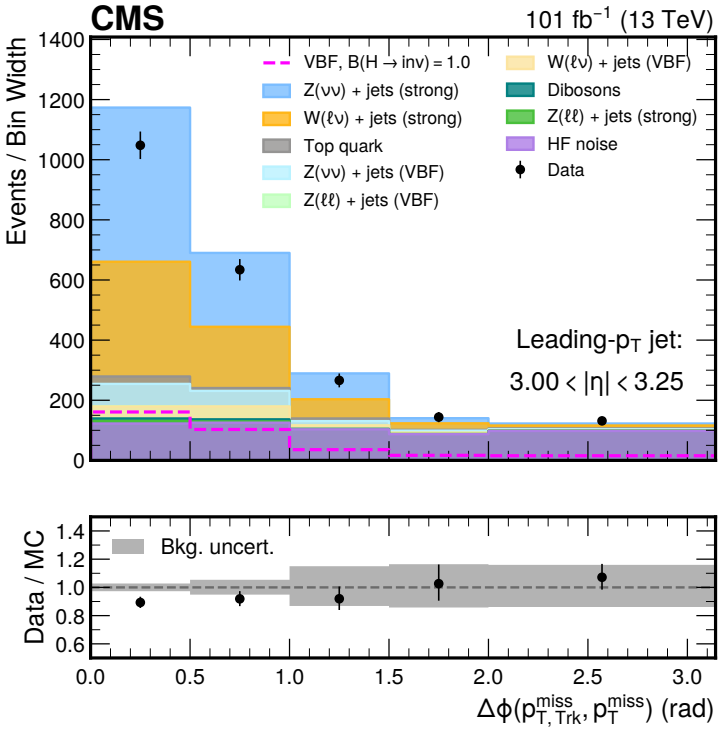

Figure 3·27: $\Delta\phi(p_{T,trk}^{miss}, p_T^{miss})$ distribution for the closure test on the HF noise estimate. Events plotted here pass the VBF signal region selection, and have the leading jet in $3 < |\eta| < 3.25$ range, where the HF noise contribution is highest. It can be observed, especially at high $\Delta\phi(p_{T,trk}^{miss}, p_T^{miss})$, the agreement between data and background yields is good after the residual HF noise estimation is taken into account. To account for residual differences, a 20% flat uncertainty is assigned to the HF noise template.

From Fig. 3·27, it can be observed, especially at high $\Delta\phi(p_{T,trk}^{miss}, p_T^{miss})$, the agreement between data and background yields is good after the residual HF noise estima-

tion is taken into account. To account for residual disagreements between data and background estimation, a 20% flat uncertainty is assigned to the HF noise template.

3.3.2 Missing HE-15/16 sectors in 2018

During a large part of data-taking in 2018, a submodule of the HCAL was not functioning. As a result, no HCAL deposits have been recorded in the region of $-3.0 < \eta < -1.3$ and $-1.57 < \phi < -0.87$ for the affected portion of the recorded dataset. These missing HCAL deposits have the following consequences for this analysis:

- The jets that are affected will often be wrongly reconstructed as an electron. These mis-reconstructed electrons can pass the identification criteria imposed in the electron control regions, due to the lack of HCAL energy deposits. Therefore, such events with mis-reconstructed jets in this $[\eta, \phi]$ region can contaminate the electron control regions.

- Affected jets will not be calibrated correctly, because the energy of the reconstructed jet is dependent on all the HCAL deposits. Such mis-calibration will result in anomalous p_T^{miss} in such events.

Therefore, it was found that the analysis regions that are most impacted by this problem are the signal region and single electron control region. Additional requirements were imposed on these two regions to reject mis-reconstructed events arising due to the missing HCAL deposits. The studies done to identify the impact on the analysis, and the requirements imposed are outlined below.

Figure 3·28 shows the impact of a veto on events having electrons reconstructed in the impacted $[\eta, \phi]$ region. The alternative of "ignoring" electrons in that region, i.e. demoting them to jets and recalculating the leading jet pair and other relevant selection variables, leads to almost identical results: The source of these jets is QCD

multijet events, hence very few of these events have another good electron to pass the selection for $W \to e\nu$ events. Therefore, an event veto is imposed on the single electron control region, if the reconstructed electron is found to be within the impacted phase space.

The muon control regions are not affected from the missing HCAL deposits, thanks to the tight muon criteria imposed both in Z and W muon control regions.

The other impacted analysis region is the signal region. Figs. 3·29 and 3·30 show the distributions, in data and MC, of the leading and subleading jet ϕ and the ϕ of the p_T^{miss} vector ($\phi_{p_T^{miss}}$), in the signal region in in 2018. From Figs. 3·29 and 3·30, it can be observed that the agreement between data and MC is good except for the region where $-1.8 < \phi_{p_T^{miss}} < -0.6$. This is co-incident with the missing HCAL submodule. The same disagreement can be seen in the jet ϕ distributions, on the opposite side in ϕ, which arises due to the signal selection requirement that the jets and missing energy are not aligned. To mitigate this problem in the signal region, the following additional requirement is imposed on 2018 data:

- For events in data where the run number is smaller than 319077 (i.e. before the HCAL submodule malfunctioning) no additional requirement is imposed, since those events are not impacted by the problem. Out of the full 59.7 fb^{-1} of 2018 data, these events correspond to 21.1 fb^{-1} of integrated luminosity.

- For events in data where the run number is larger or equal to 319077, the event is discarded if it has $-1.8 < \phi_{p_T^{miss}} < -0.6$. These events correspond to a 38.6 fb^{-1} of integrated luminosity.

The usage of run number to identify if the event is impacted by the missing HCAL submodule allows to keep 100% of the data before the problem occurred. This effect is modeled in simulation by reweighting the set of events with $-1.8 < \phi_{p_T^{miss}} < -0.6$ with

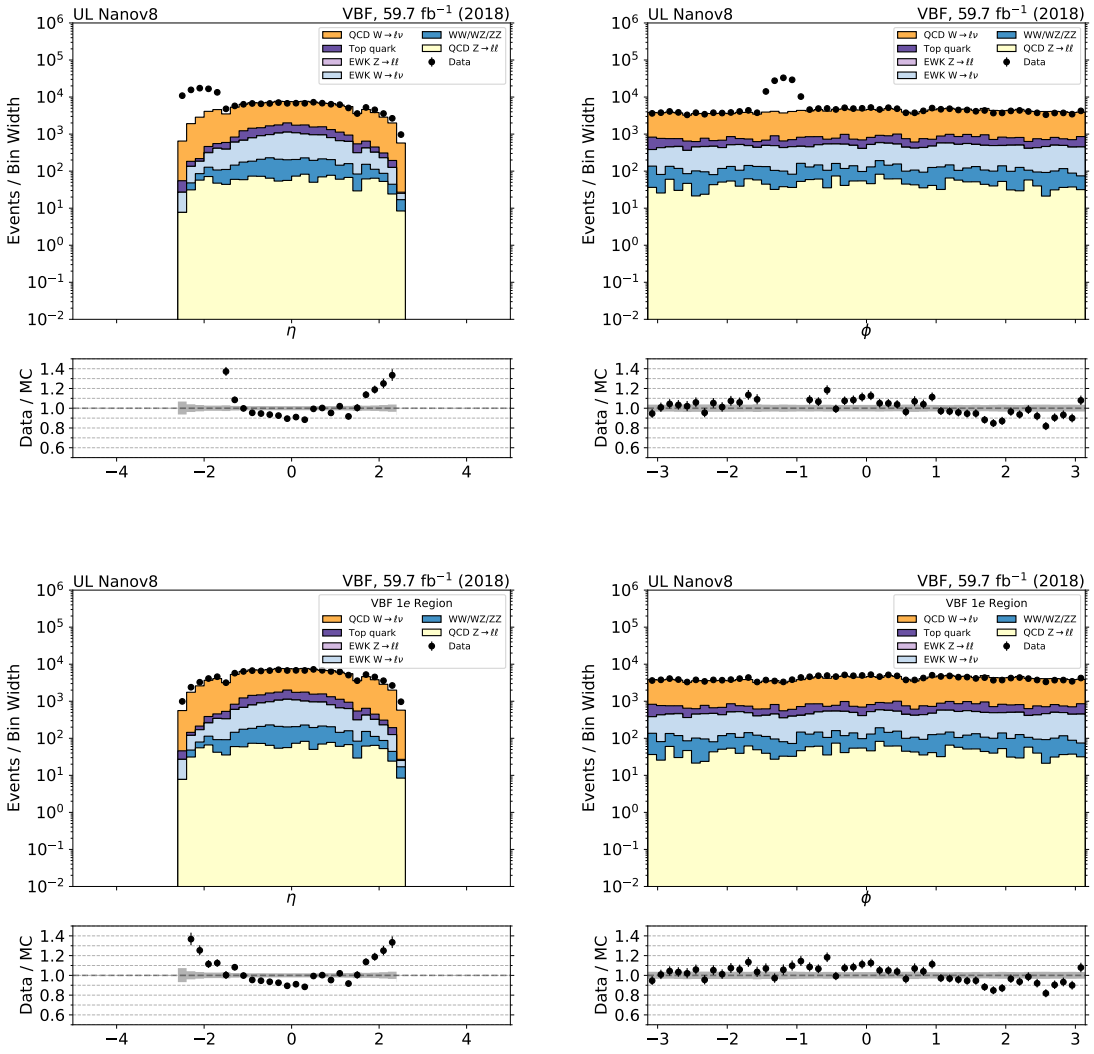

Figure 3·28: Effects of an electron veto in the affected detector region (single electron control region 2018). Plots on the top show the η and ϕ of the highest p_T electron in the event without any veto, while plots on the bottom show the η and ϕ of the electron with the veto.

the fraction of luminosity without impact. This weight corresponds to 21.1/59.7 = 0.35. Therefore, about 35% of the events with $-1.8 < \phi_{p_T^{miss}} < -0.6$ in simulation are kept in the analysis.

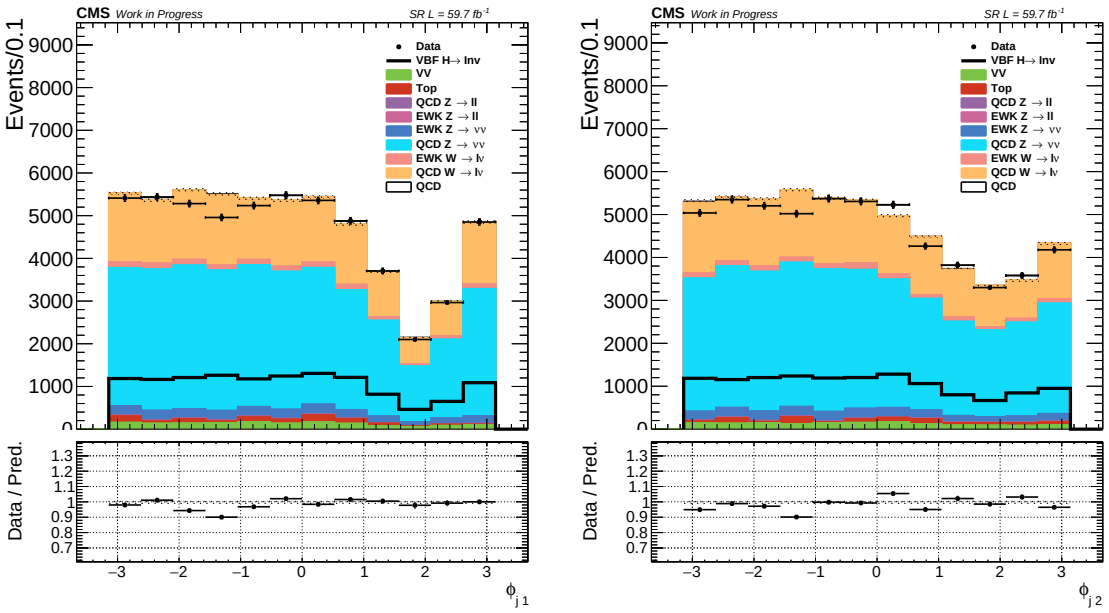

Figure 3·29: Leading (left) and subleading (right) jet ϕ distributions in the signal region from the 2018 dataset. The bands in the ratio plots represent the statistical uncertainties in the MC.

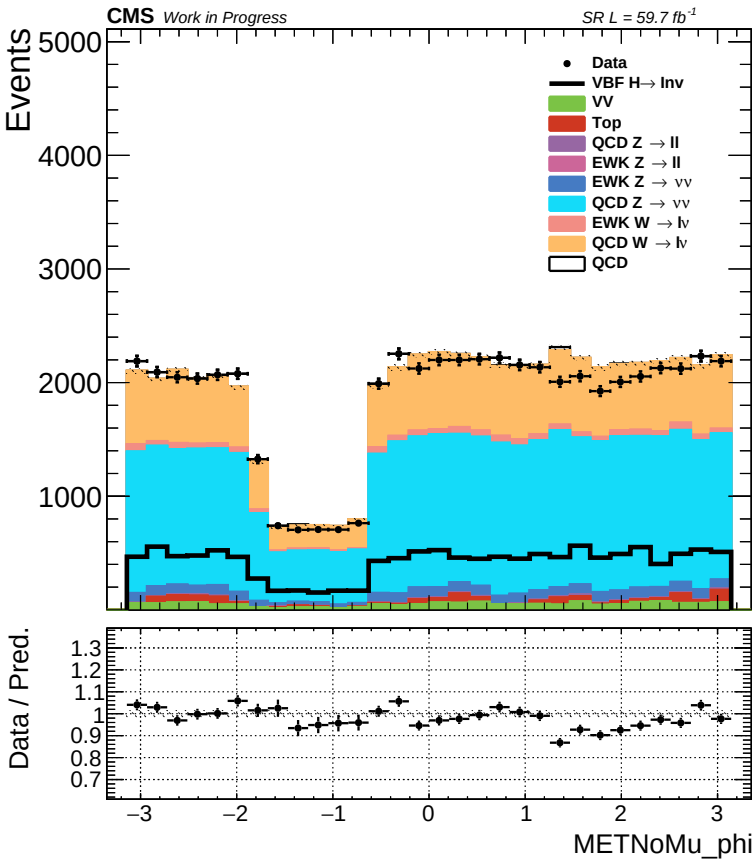

Figure 3·30: $\phi_{p_T^{miss}}$ distribution in the signal region from the 2018 dataset.

3.4 Event selection

3.4.1 Signal region selection

Events in the VBF signal region are selected by using triggers which require $p_{T,no-\mu}^{miss} >$ 120 GeV and $H_{T,no-\mu}^{miss} > 120$ GeV at HLT-level (described in Sec. 3.2.1). The $p_{T,no-\mu}^{miss}$ corresponds to the magnitude of the vector $p_{\mathrm{T},V}$ sum of all the particles reconstructed by the trigger algorithm, while the $H_{T,no-\mu}^{miss}$ is computed as the magnitude of the vector $p_{\mathrm{T},V}$ sum of jets with $p_T > 20$ GeV and $|\eta| < 5.0$. The energy fraction attributed to neutral hadrons in these jets is required to be smaller than 90%. This requirement suppresses anomalous events with jets originating from detector noise. To be able to use the same triggers for selecting events in the muon control regions used for background prediction, muon candidates are not included in the $p_{T,no-\mu}^{miss}$ and $H_{T,no-\mu}^{miss}$ computation. The trigger efficiency is measured to be 96% for events passing the analysis selection for $p_T^{miss} > 250$ GeV and becomes more than 99% efficient for events with $p_T^{miss} > 350$ GeV. The performance of these triggers are further discussed in Sec. 3.2.1.

Candidate events are required to have $p_T^{miss} > 250$ GeV. The leading AK4 jet in the signal event is required to have $p_T > 80$ GeV and $|\eta| < 4.7$, and the subleading AK4 jet is required to have $p_T > 40$ GeV and $|\eta| < 4.7$. Both leading jets are required to have neutral hadron fraction values of less than 80% and charged hadron fraction values of at least 10% to further reduce the possibility of noise contamination. In addition, the analysis employs various event filters to reduce events with large misreconstructed p_T^{miss} [95] originating from non-collision backgrounds. The filters are complemented by a requirement on the relative difference between the default PF p_T^{miss} value used in the analysis and the p_T^{miss} value reconstructed just from calorimeter deposits and muons. In an event with genuine p_T^{miss}, one would expect the two p_T^{miss} calculations to result in reasonably close values. Following this constraint, the following cut is

imposed on events:

$$\frac{\Delta p_T^{miss}(\mathrm{PF},\mathrm{Calo})}{p_T^{miss}} < 0.5 \tag{3.14}$$

In Eq. 3.14, $\Delta p_T^{miss}(\mathrm{PF},\mathrm{Calo})$ refers to the absolute difference of p_T^{miss} values calculated using the full event information (PF) and calorimeter information only (Calo).

For the VBF $H \to inv.$ signal events, two leading jets in opposite hemispheres are expected in the final state, with large dijet mass. Furthermore, these jets are expected to have large rapidity separation ($\Delta\eta_{jj}$) and small azimuthal separation ($\Delta\phi_{jj}$). Therefore, this analysis employs several requirements on M_{jj}, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$, which can be found in Table 3.10.

The main background processes in this search are the $\mathrm{Z}(\nu\nu)$+jets and $\mathrm{W}(\ell\nu)$+jets processes. The $\mathrm{Z}(\nu\nu)+\mathrm{jets}$ process is an irreducible background and constitutes the largest background in the search. In contrast, the background from $\mathrm{W}(\ell\nu)+\mathrm{jets}$ is suppressed by imposing a veto on events containing one or more loose muons or electrons with $p_T > 10$ GeV, or hadronically decaying τ leptons with $p_T > 20$ GeV. Events that contain a loose, isolated photon with $p_T > 15$ GeV and $|\eta| < 2.5$ are also rejected. This helps to suppress electroweak (EWK) backgrounds in which a photon is radiated from the initial state. To reduce the contamination from top quark backgrounds, events are rejected if they contain a b tagged jet with $p_T > 20$ GeV and $|\eta| < 2.4$. These jets are identified using the DeepCSV algorithm [81, 96], adopting the "medium" working point, which corresponds to correctly identifying a jet originating from a bottom quark with a probability of 80% and misidentifying a jet originating from a charm quark (light-flavor jet) with a probability of 12 (2)%. Lastly, QCD multijet background with p_T^{miss} arising from mismeasurements of the jet momenta is suppressed by requiring the minimum azimuthal angle between the $\vec{p}_T^{miss}$

direction and each of the first four leading jets with $p_T > 30$ GeV and $|\eta| < 4.7$ to be larger than 0.5 radians.

The selection requirements for the signal region in this analysis are summarized in Table 3.10.

Table 3.10: Summary of the signal region selection requirements. Leading and subleading jets refer to the highest and second-highest p_T jets in the event.

Variable	Selection	Target background
Muon (electron) veto	$p_T > 10$ GeV, $\|\eta\| < 2.4(2.5)$	$Z(\ell\ell) + \text{jets}$, $W(\ell\nu) + \text{jets}$
τ lepton veto	$p_T > 20$ GeV, $\|\eta\| < 2.3$	$Z(\ell\ell) + \text{jets}$, $W(\ell\nu) + \text{jets}$
Photon veto	$p_T > 15$ GeV, $\|\eta\| < 2.5$	$\gamma + \text{jets}$
Bottom jet veto	DeepCSV medium $< 0.4941/0.4184$ (2017 / 2018) for all jets with $p_T > 20$ GeV, $\|\eta\| < 2.4$	Top quark
p_T^{miss}	>250 GeV	QCD, top quark, $Z(\ell\ell) + \text{jets}$
$\Delta\phi(\vec{p}_T^{\text{jet}}, \vec{p}_T^{miss})$	>0.5 radians	QCD
Two leading jet IDs	POG tight ID, CHF> 0.1, NHF< 0.8	Noise
HF-noise rejection	Cuts on HF shape variables, $\sigma_{i\eta i\eta}, \sigma_{i\phi i\phi}$ and HF central strip size	HF noise
$\Delta p_T^{miss}(\text{PF}, \text{Calo})$ / recoil	< 0.5	Noise
Leading AK4 jet p_T and η	>80 GeV and $\|\eta\| < 4.7$	All
Subleading AK4 jet p_T and η	>40 GeV and $\|\eta\| < 4.7$	All
M_{jj}	>200 GeV	
$\Delta\eta_{jj}$	>1.0	
$\Delta\phi_{jj}$	<1.5	

It should be noted that the selections listed in Table. 3.10 refer to one of the two event selection categories in the analysis, the MET-triggered (MTR) category. In addition to the MTR category, there is another orthogonal event selection category called the VBF-triggered (VTR) category [4]. Events in this category are collected using different HLT paths that require the presence of two energetic jets in the final state, and have $160 < p_T^{miss} < 250$ GeV, hence the orthogonality to the MTR category being considered here ($p_T^{miss} > 250$ GeV). For the purposes of this thesis, the VTR category won't be discussed in detail, but the combined results with the VTR category will be mentioned in Sec. 4.

Figs. 3·31, 3·32, and 3·33 show the distribution of the p_T^{miss}, M_{jj}, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ of the VBF jet pair (two highest-p_T jets), the p_T and η distributions of the leading and subleading jets, and the η distributions of the most central and most forward jets, re-

spectively, for events passing the VBF signal region selection. The same distributions are shown for the 2018 samples in Figs. 3·34, 3·35, and 3·36.

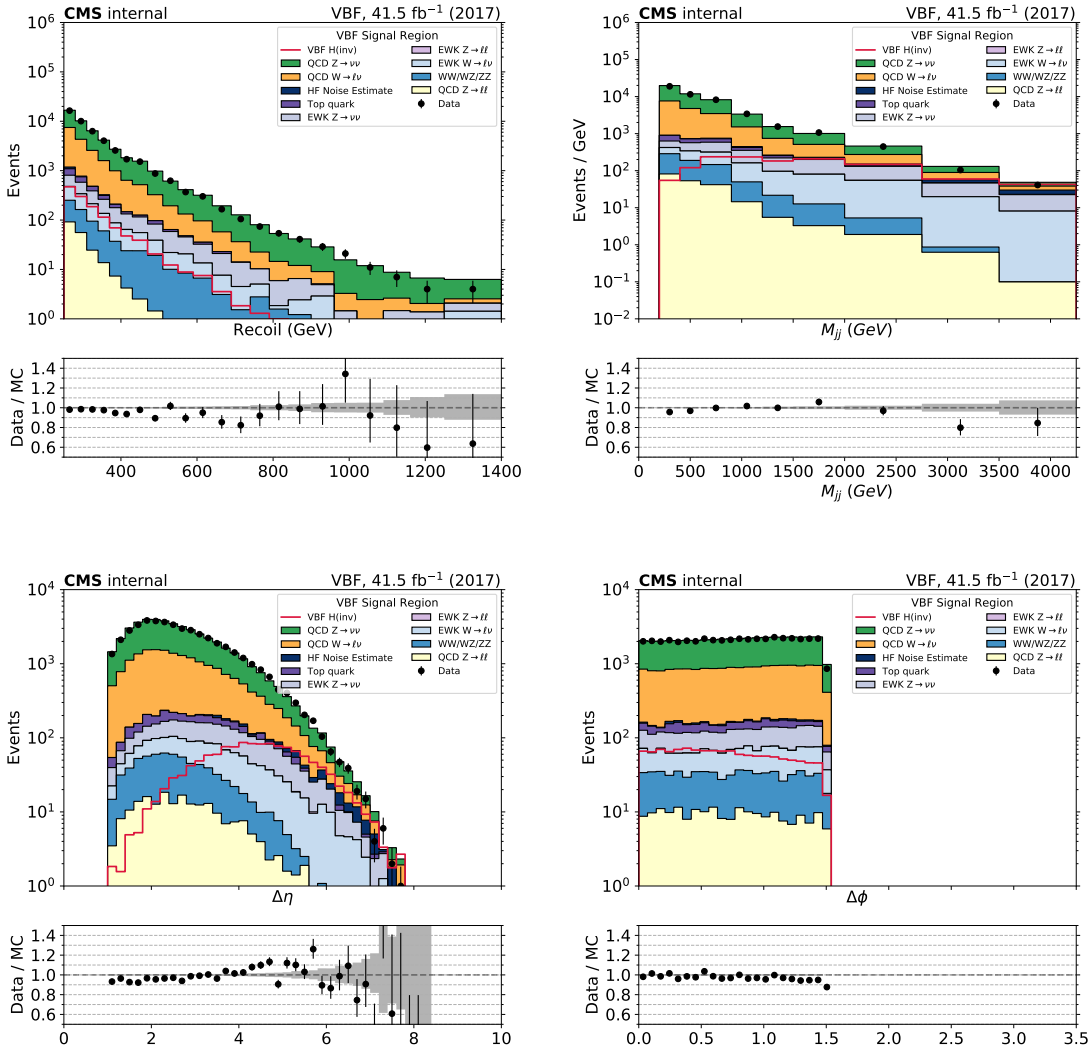

Figure 3·31: Recoil distribution, M_{jj} distribution, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ distribution in the VBF signal region, using 2017 dataset.

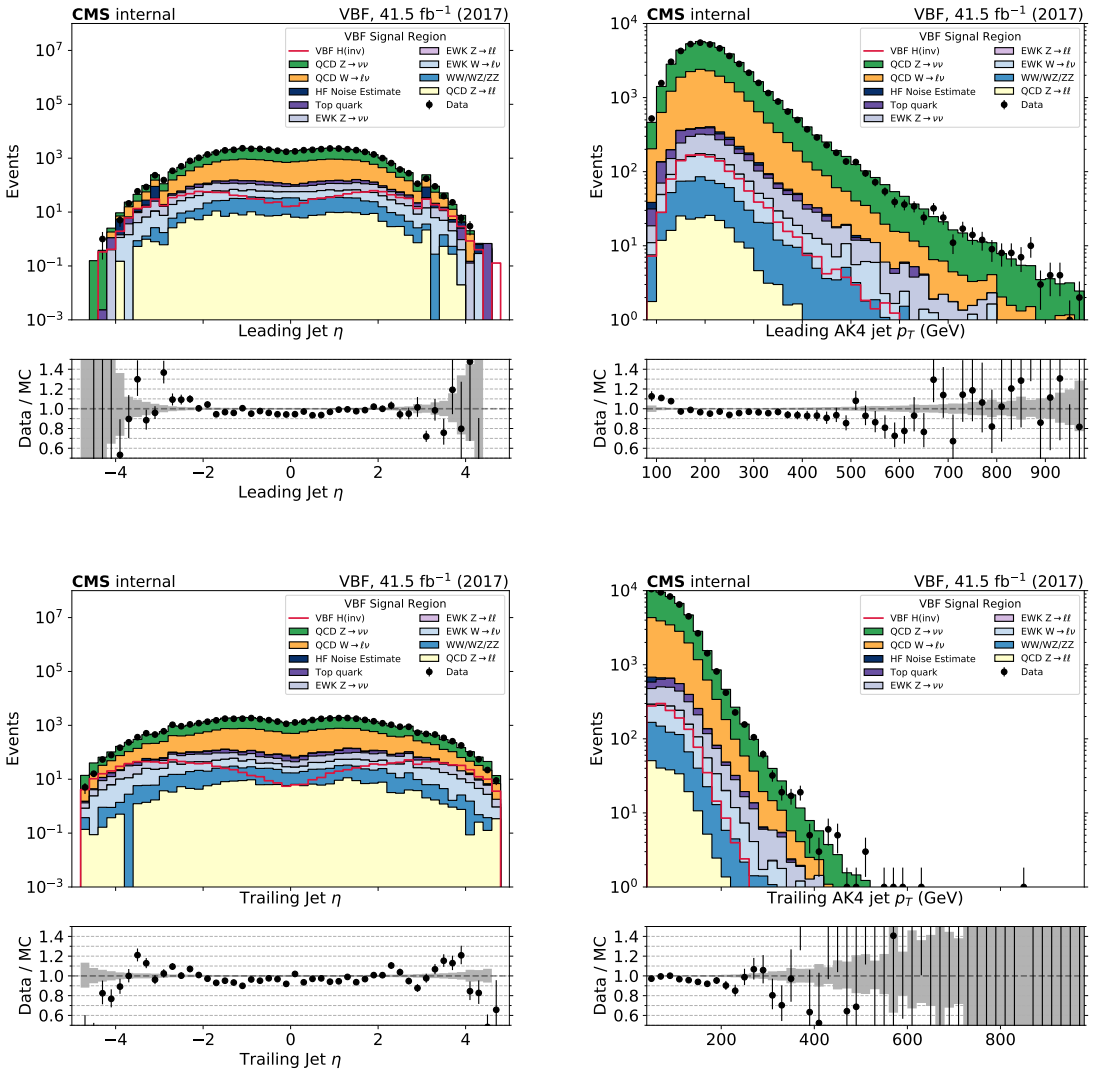

Figure 3·32: Leading and subleading jets p_T and η in the VBF signal region, using 2017 dataset.

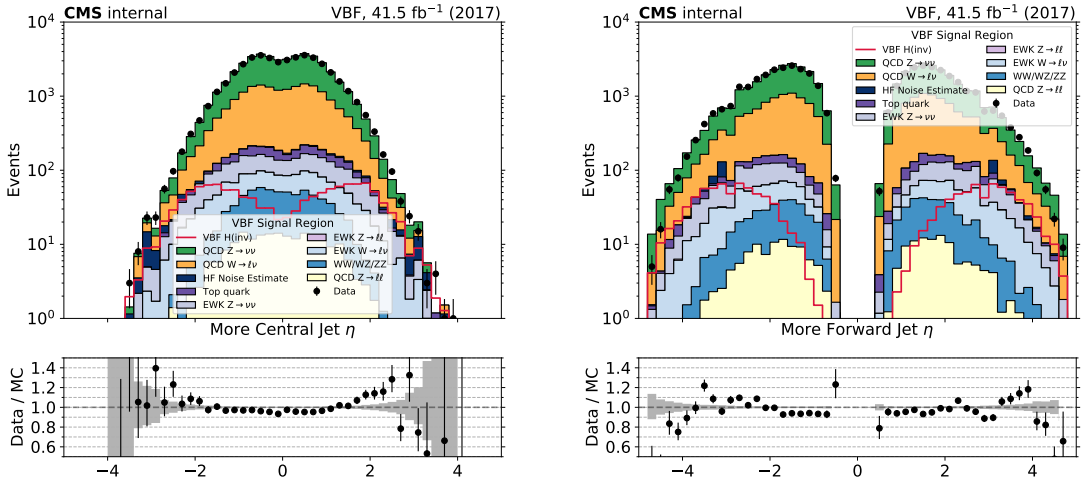

Figure 3·33: Most central (left) and most forward jets (right) η of the VBF pair in the VBF signal region, using 2017 dataset.

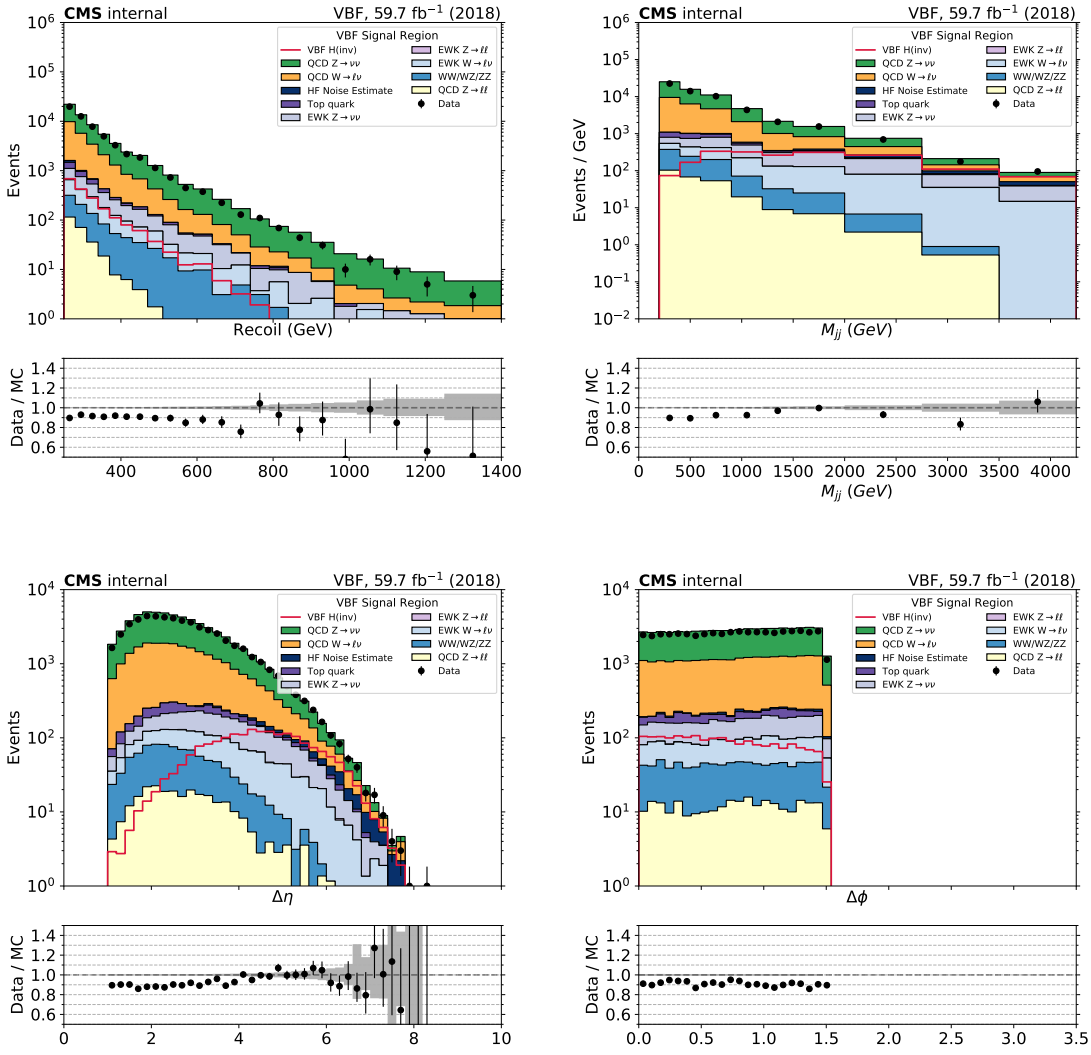

Figure 3·34: Recoil distribution, M_{jj} distribution, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ distribution in the VBF signal region, using 2018 dataset.

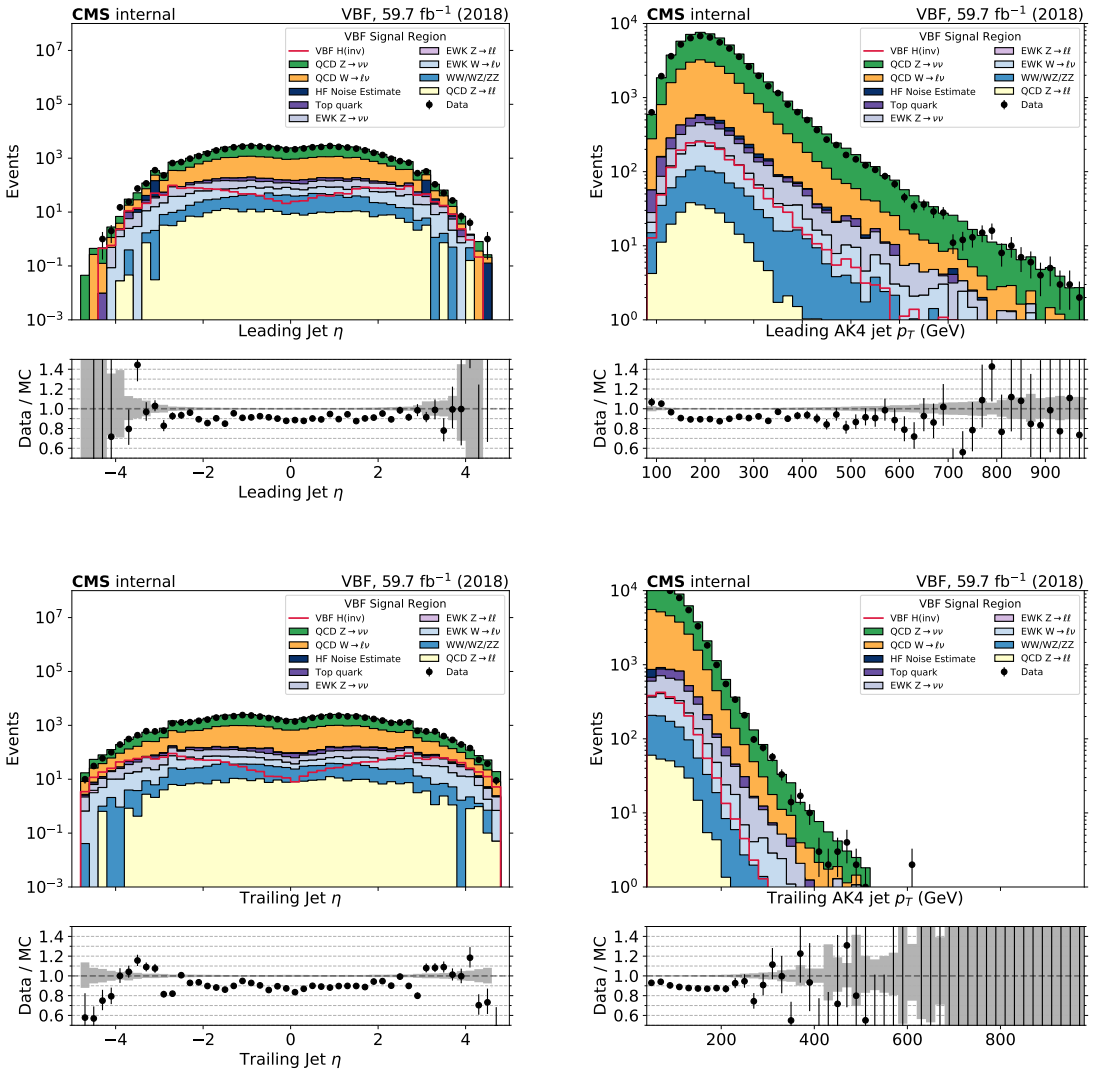

Figure 3·35: Leading and subleading jets p_T and η in the VBF signal region, using 2018 dataset.

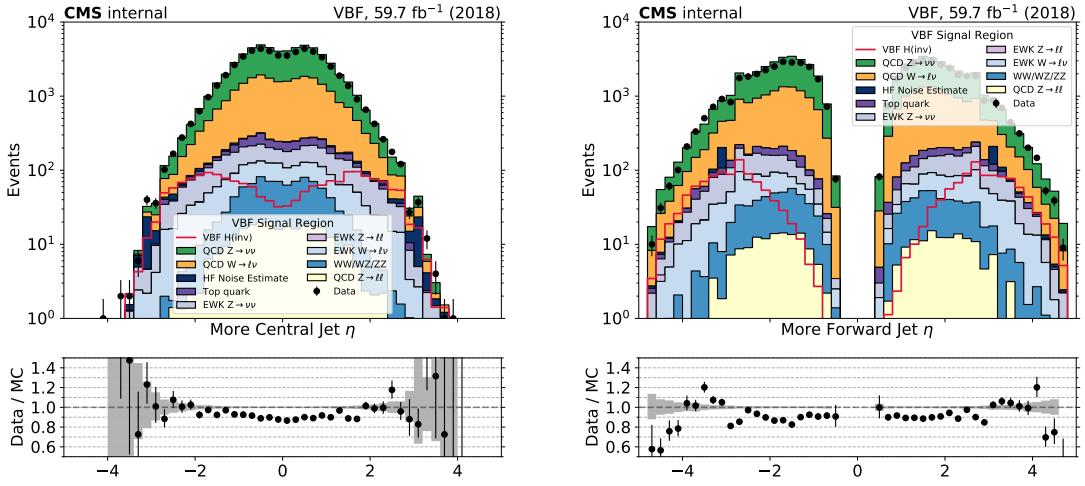

Figure 3·36: Most central (left) and most forward jets (right) η of the VBF pair in the VBF signal region, using 2018 dataset.

3.4.2 Single Muon Control Region

Single-muon control region events are selected using full signal region criteria of VBF selection with the exception of the muon veto. The p_T^{miss} requirement is replaced by an identical requirement on the hadronic recoil, which is defined as the sum of $\vec{p}_T^{miss}$ and the muon $\vec{p}_{\mathrm{T}}$ (Eq. 3.2), and thus corresponds to the p_T of the W boson. In the single-muon control region, exactly one tightly identified, isolated muon with $p_T > 20$ GeV is required. No additional loose muons or electrons with $p_T > 10$ GeV are allowed in the event.

Figs. 3·37 and 3·39 show the distributions of the recoil, M_{jj}, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ of the two leading AK4 jets for events in the single muon control region for the VBF category in 2017 and 2018 datasets, respectively. Figs. 3·38 and 3·40 show the distributions of the leading muon p_T and η, for 2017 and 2018, respectively.

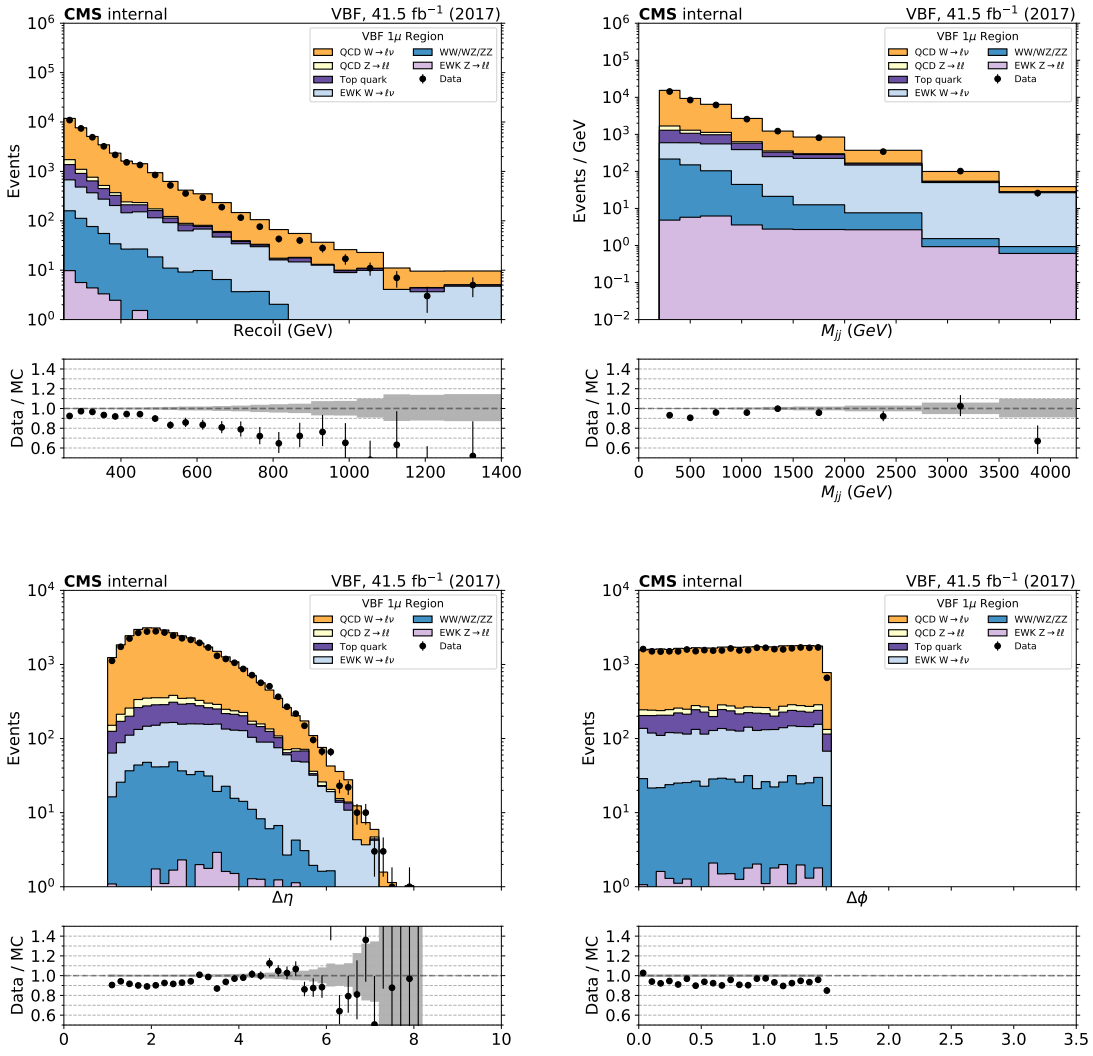

Figure 3·37: Comparison between 2017 data and Monte Carlo simulation in the single muon control region for the recoil, M_{jj} distribution, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ distributions for the two leading AK4 jets.

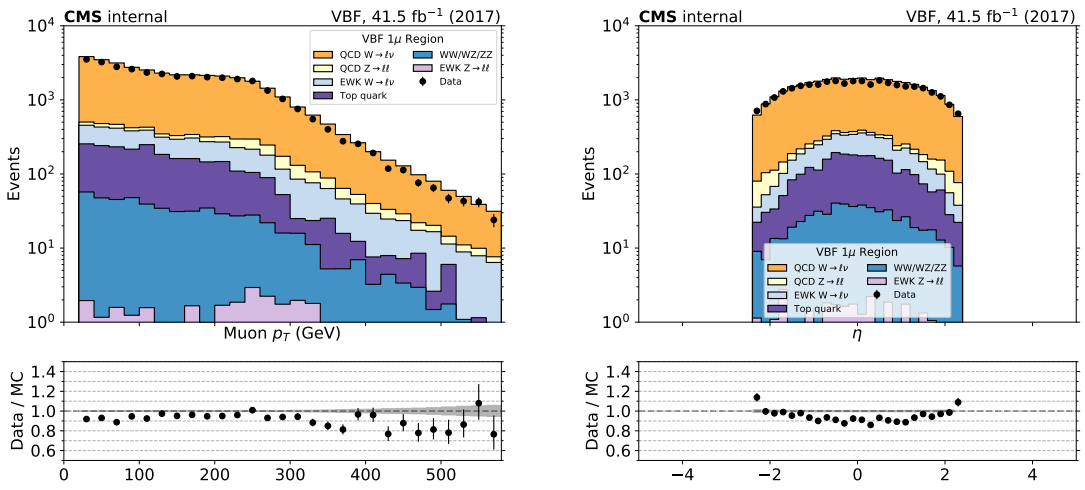

Figure 3·38: Comparison between 2017 data and Monte Carlo simulation in the single muon control region for the p_T and η distribution of the reconstructed muon.

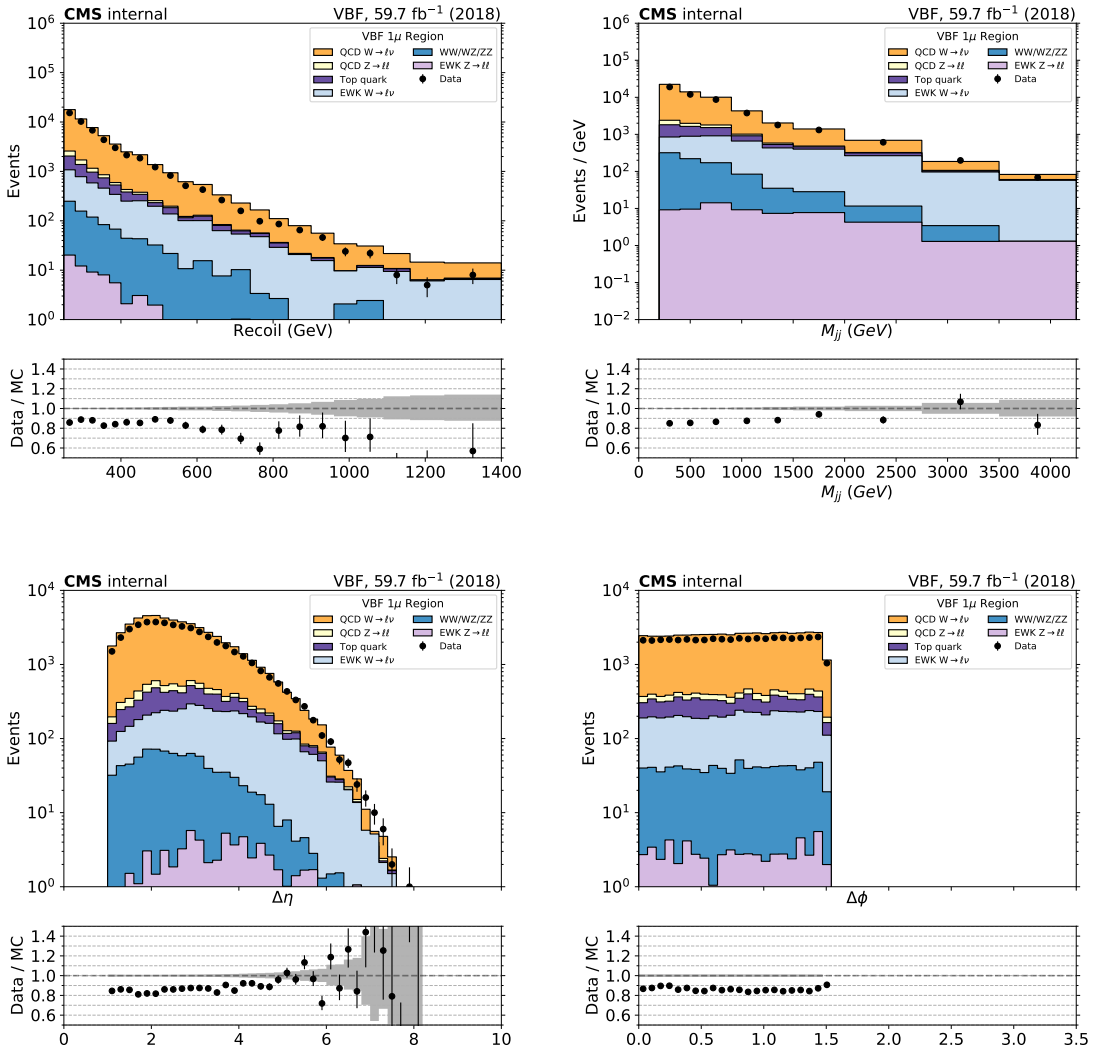

Figure 3·39: Comparison between 2018 data and Monte Carlo simulation in the single muon control region for the recoil, M_{jj} distribution, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ distributions for the two leading AK4 jets.

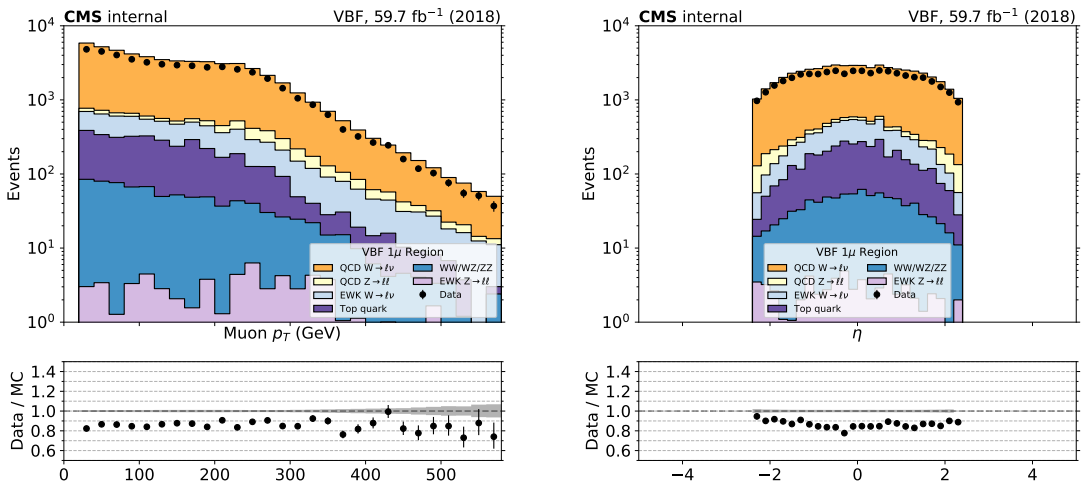

Figure 3·40: Comparison between 2018 data and Monte Carlo simulation in the single muon control region for the p_T and η distribution of the reconstructed muon.

3.4.3 Single Electron Control Region

Events for the single-electron control region are collected with the single-electron and photon triggers described in Sec. 3.2.1. Similar to the muon control regions, the p_T^{miss} requirement is replaced with an identical requirement on the hadronic recoil, which is defined as the sum of $\vec{p}_T^{miss}$ and the electron $\vec{p}_{\mathrm{T}}$ (Eq. 3.2), and thus corresponds to the p_T of the W boson. The events in the single-electron control region are required to contain exactly one tightly identified and isolated electron with $p_T > 40$ GeV. In addition, the contamination from QCD multijet events in this control region is suppressed by requiring $p_T^{miss} > 80$ GeV.

Figs. 3·41 and 3·43 show the distributions of the recoil, M_{jj}, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ of the two leading AK4 jets for events in the single-electron control region for the VBF category in 2017 and 2018 datasets, respectively. Figs. 3·42 and 3·44 show the distributions of the leading electron p_T and η, for 2017 and 2018, respectively.

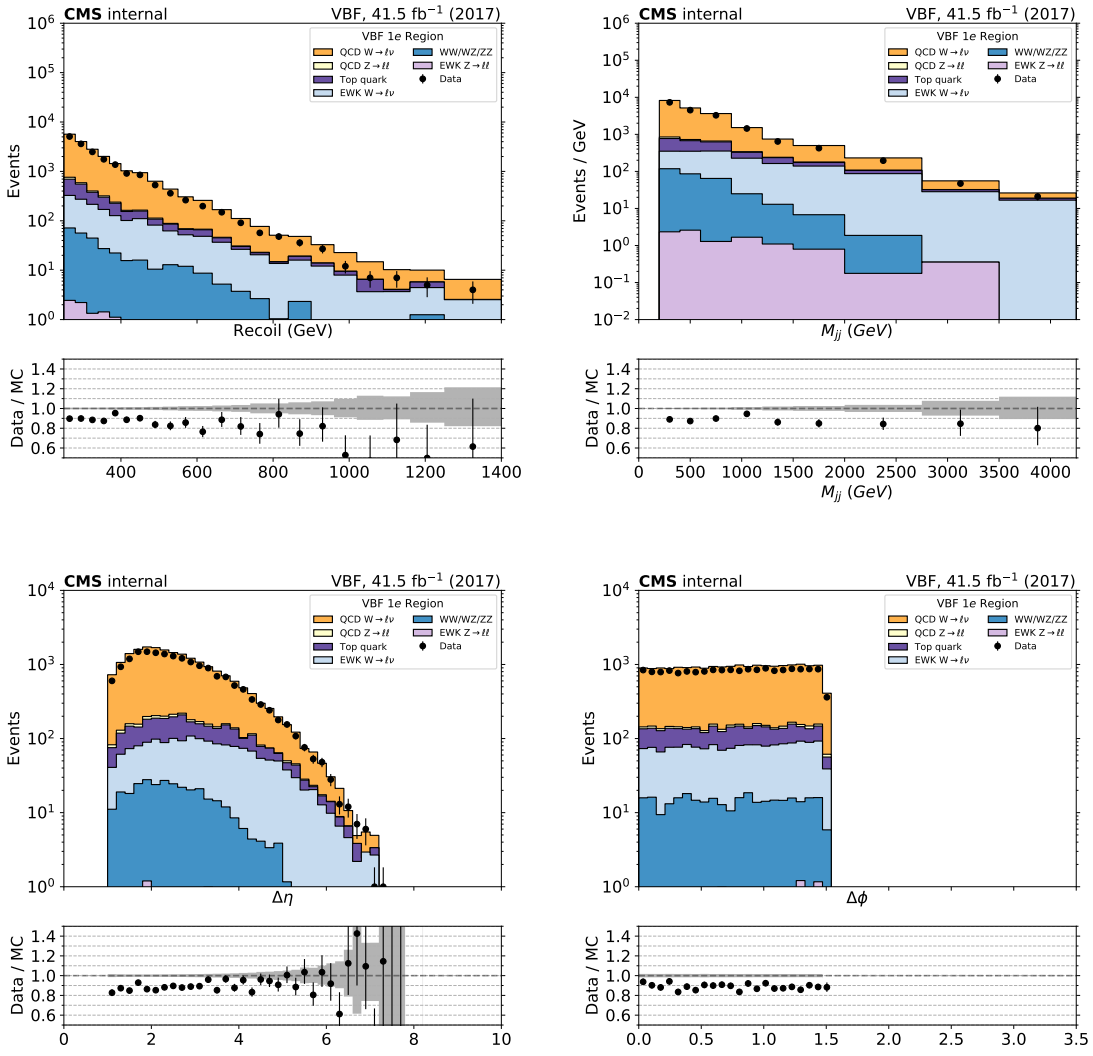

Figure 3·41: Comparison between 2017 data and Monte Carlo simulation in the single electron control region for the recoil, M_{jj} distribution, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ distributions for the two leading AK4 jets.

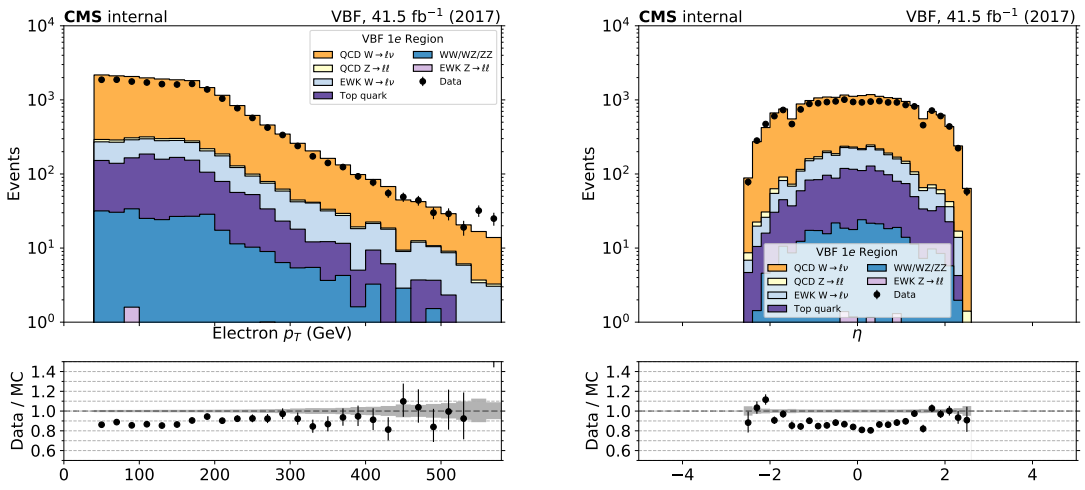

Figure 3·42: Comparison between 2017 data and Monte Carlo simulation in the single electron control region for the p_T and η distribution of the reconstructed electron.

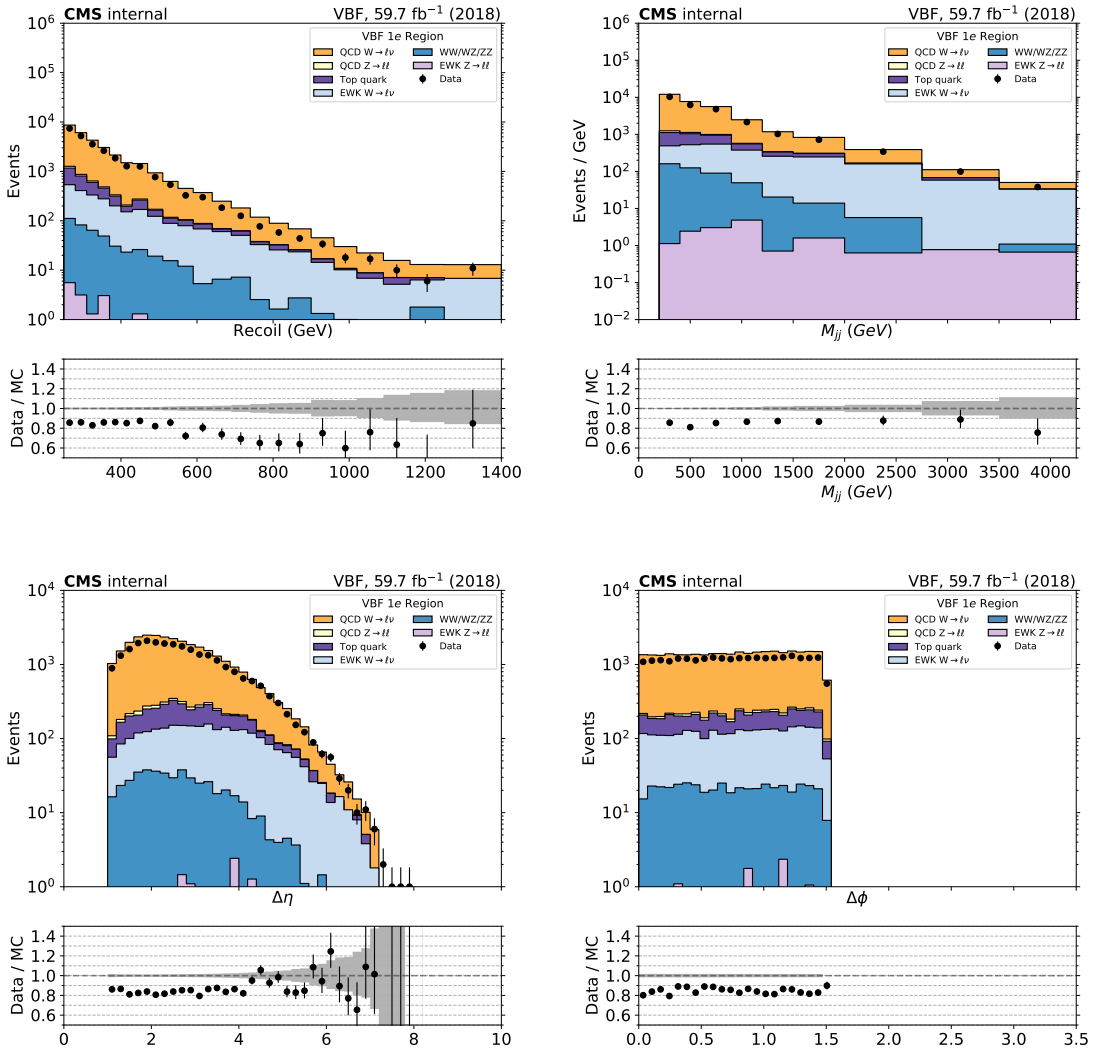

Figure 3·43: Comparison between 2018 data and Monte Carlo simulation in the single electron control region for the recoil, M_{jj} distribution, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ distributions for the two leading AK4 jets.

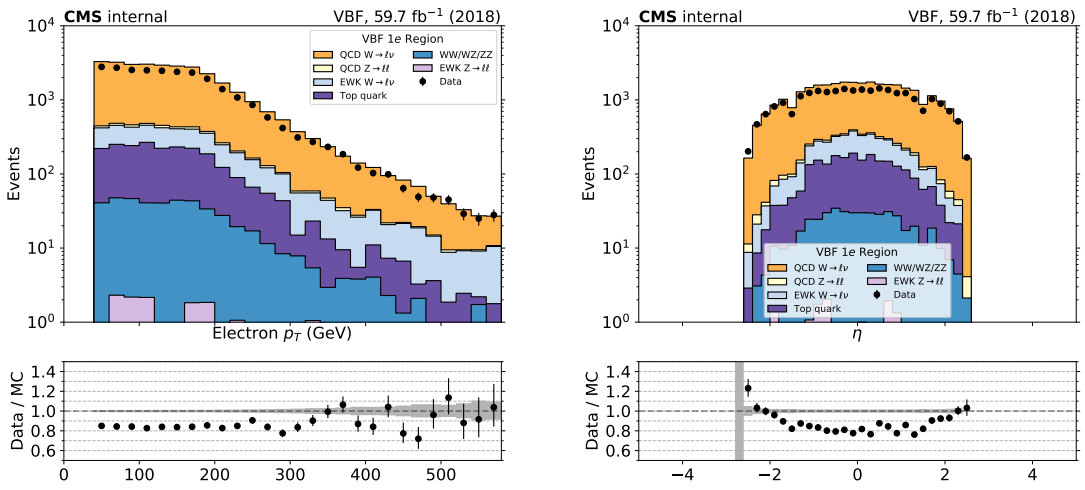

Figure 3·44: Comparison between 2018 data and Monte Carlo simulation in the single electron control region for the p_T and η distribution of the reconstructed electron.

3.4.4 Double Muon Control Region

Double-muon control region events are selected using full signal region criteria of VBF category with the exception of the muon veto. In the double-muon control region, events are selected by requiring leading (subleading) muon p_T greater than 20 (10) GeV and an invariant mass of the two muons is required to be in the range of 60 to 120 GeV, compatible with a $Z \to \mu\mu$ decay. At least one of the two muons is required to pass the tight candidate definition, as defined in Sec. 3.1.3. Events are rejected if there is an additional loose muon or electron with $p_T > 10$ GeV. The SR p_T^{miss} requirement is replaced with an identical requirement on the hadronic recoil, which is defined as the sum of $\vec{p}_T^{miss}$ and the muon $\vec{p}_{\mathrm{T}}$ (Eq. 3.2), and thus corresponds to the distribution of the Z p_T smeared with the p_T^{miss} resolution of the detector.

Figs. 3·45 and 3·47 show the distributions of the recoil, M_{jj}, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ of the two leading AK4 jets for events in the double-muon control region for the VBF category in 2017 and 2018 datasets, respectively. Figs. 3·46 and 3·48 show the distributions of the leading muon p_T and η, as well as the dimuon mass and minimum $\Delta\phi$ between 4 leading jets and recoil, for 2017 and 2018, respectively.

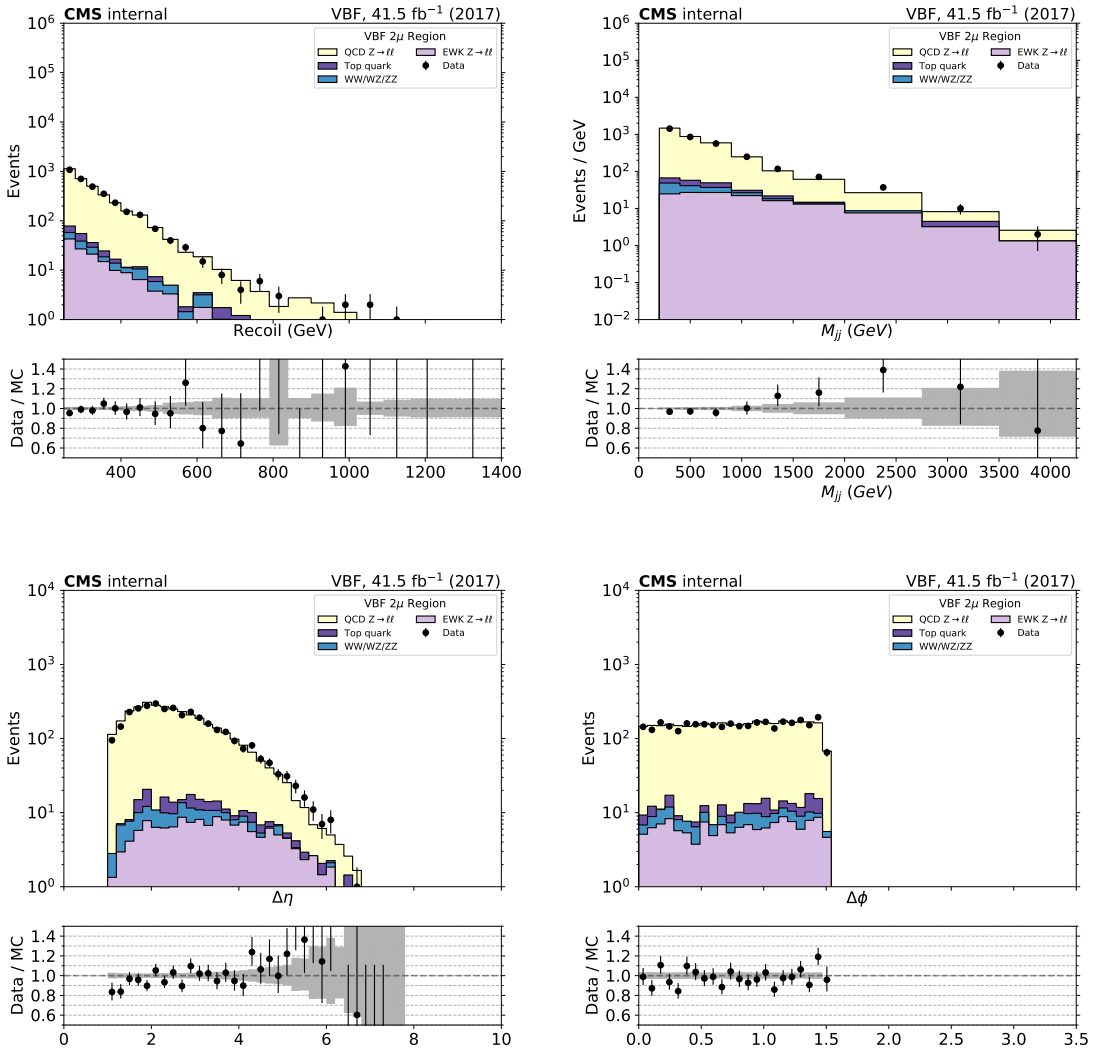

Figure 3·45: Comparison between 2017 data and Monte Carlo simulation in the double muon control region for the recoil, M_{jj} distribution, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ distributions for the two leading AK4 jets.

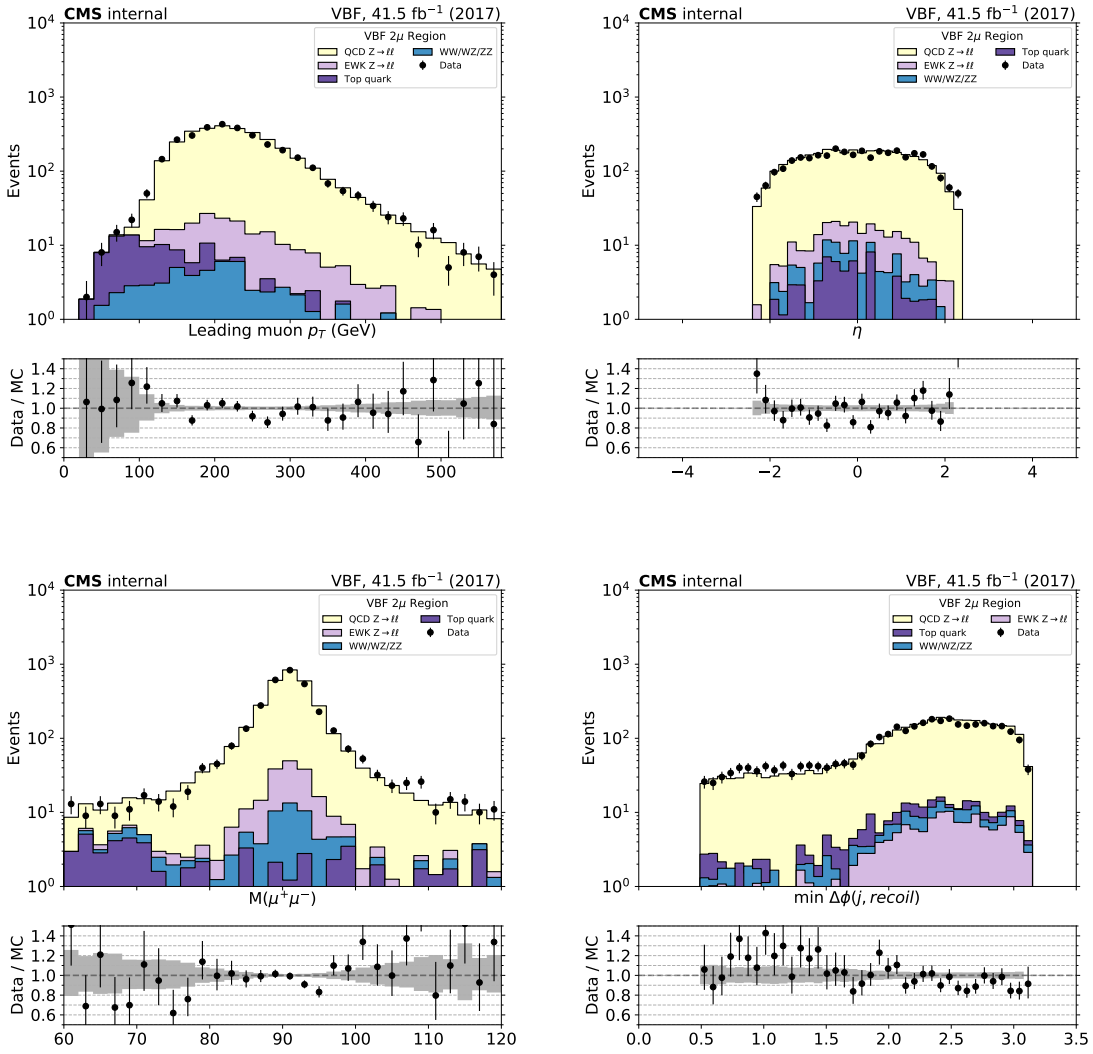

Figure 3·46: Comparison between 2017 data and Monte Carlo simulation in the double muon control region for the p_T and η of the leading muon (top), dimuon mass and minimum $\Delta\phi$ between 4 leading jets and recoil (bottom).

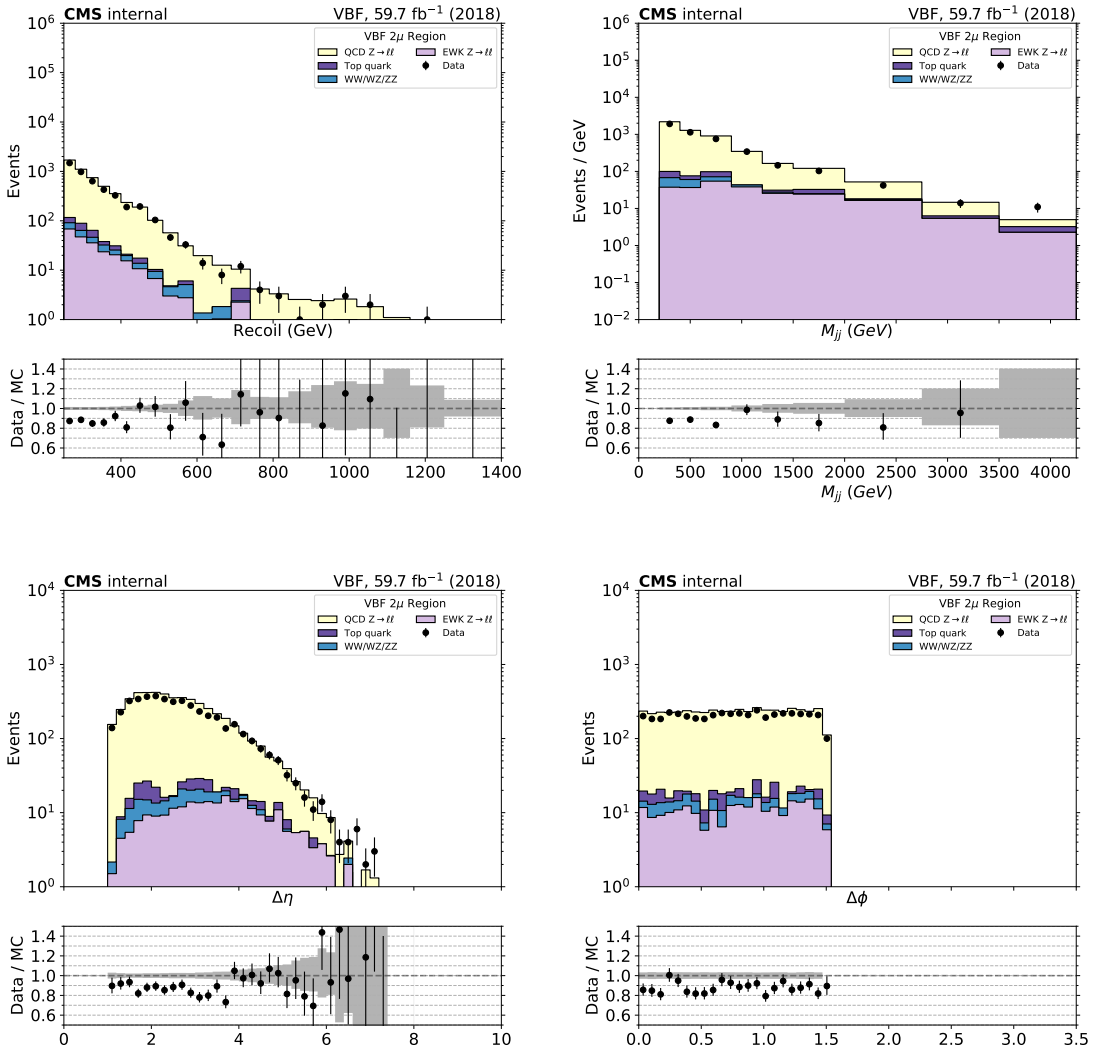

Figure 3·47: Comparison between 2018 data and Monte Carlo simulation in the double muon control region for the recoil, M_{jj} distribution, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ distributions for the two leading AK4 jets.

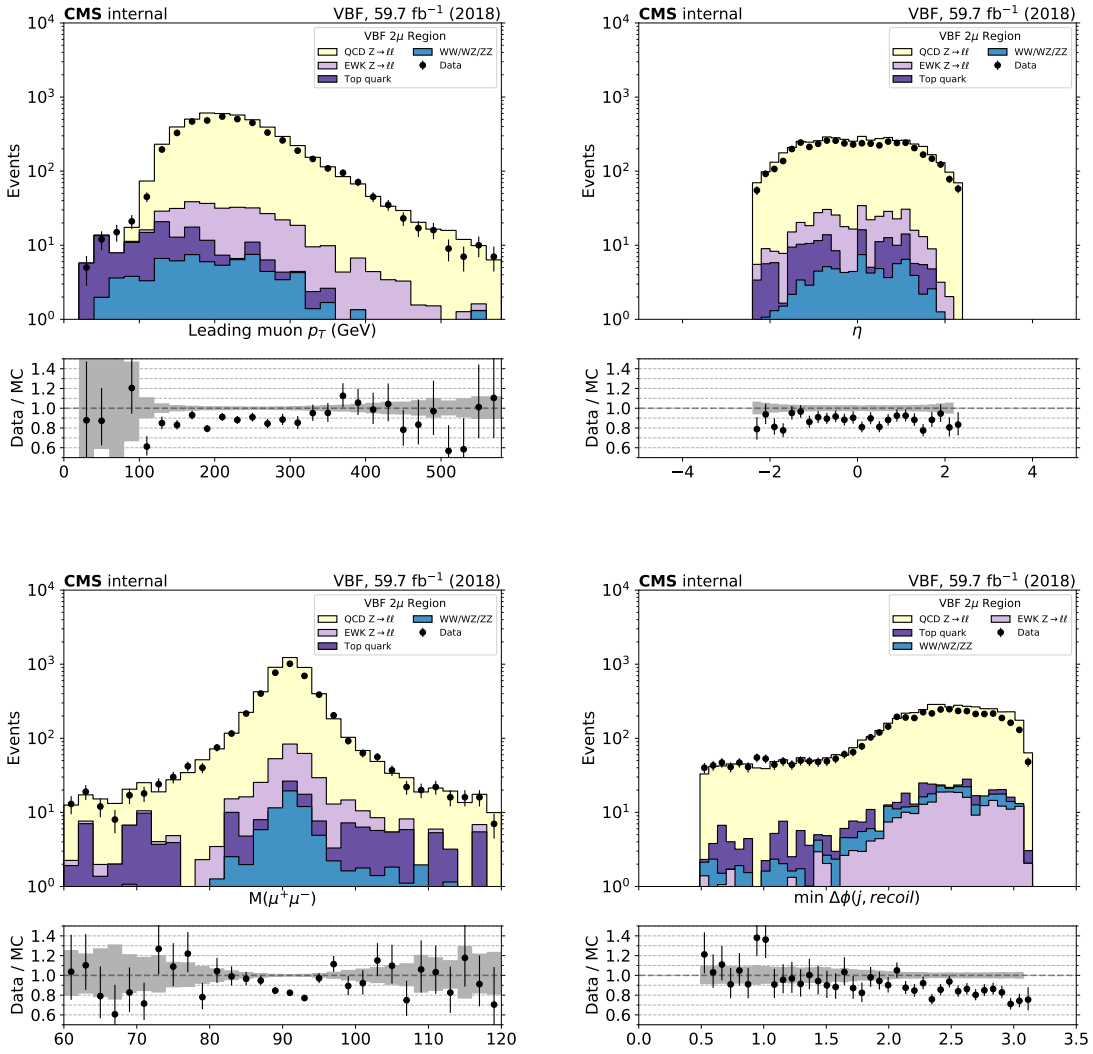

Figure 3·48: Comparison between 2018 data and Monte Carlo simulation in the double muon control region for the p_T and η of the leading muon (top), dimuon mass and minimum $\Delta\phi$ between 4 leading jets and recoil (bottom).

3.4.5 Double Electron Control Region

Events for the double-electron control region are collected with the single-electron and photon triggers described in Sec. 3.2.1. In the offline analysis, events in the dielectron control region are required to contain exactly two oppositely charged electrons with leading (trailing) electron p_T greater than 40 (10) GeV, with at least one of the two passing the tight candidate definition, as defined in Sec. 3.1.3. Similar to the double muon region, the invariant mass of the electron-positron pair is required to be in the range of 60 to 120 GeV, thus being compatible with a $Z \to ee$ decay. The SR p_T^{miss} requirement is replaced with an identical requirement on the hadronic recoil, which is defined as the sum of $\vec{p}_T^{miss}$ and the muon $\vec{p}_{\mathrm{T}}$ (Eq. 3.2), and thus corresponds to the distribution of the Z p_T smeared with the p_T^{miss} resolution.

Figs. 3·49 and 3·51 show the distributions of the recoil, M_{jj}, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ for the two leading AK4 jets for events in the double-electron control region for the VBF category in 2017 and 2018 datasets, respectively. Figs. 3·50 and 3·52 show the distributions of the leading electron p_T and η, as well as the dielectron mass and minimum $\Delta\phi$ between 4 leading jets and recoil, for 2017 and 2018, respectively.

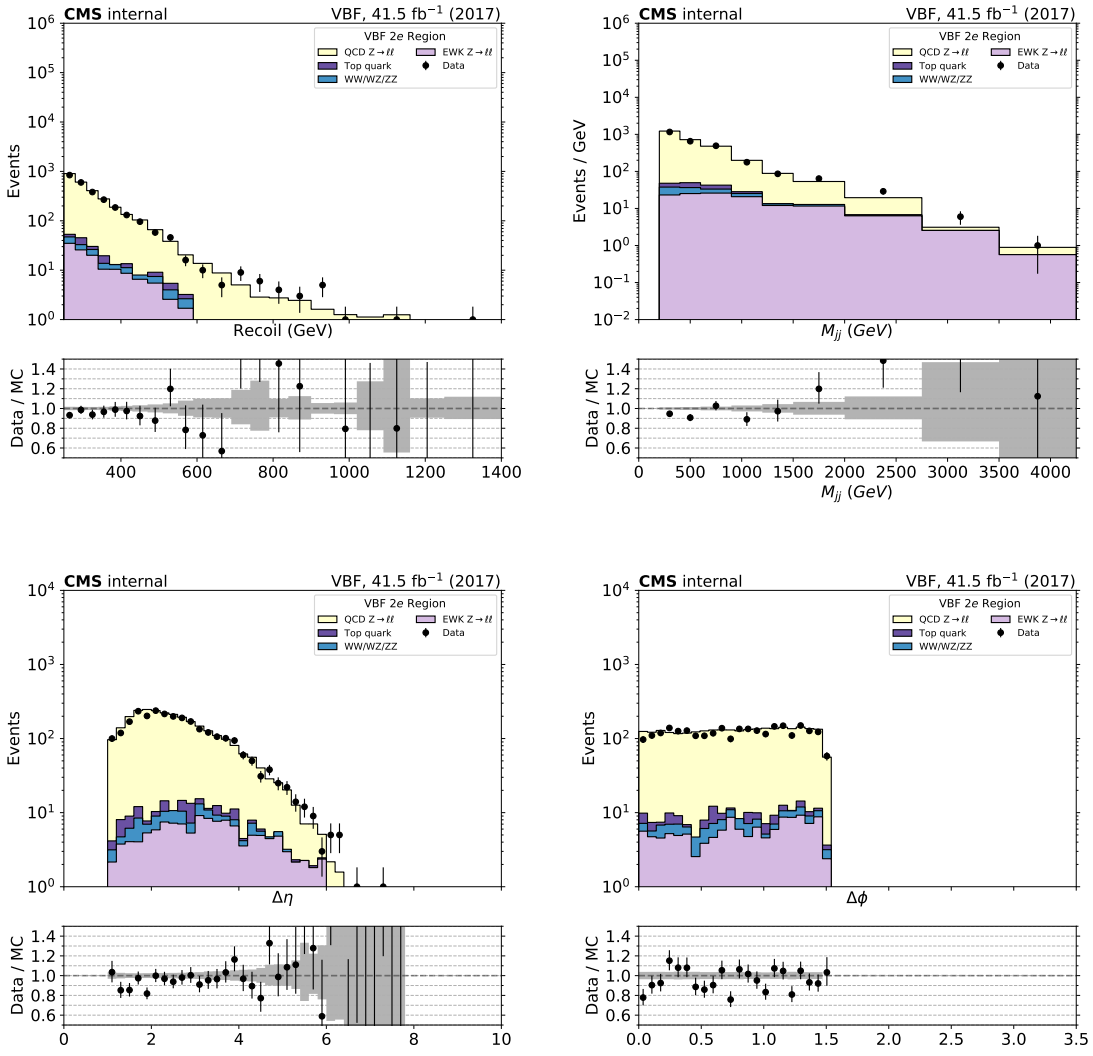

Figure 3·49: Comparison between 2017 data and Monte Carlo simulation in the double electron control region for the recoil, M_{jj} distribution, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ distributions for the two leading AK4 jets.

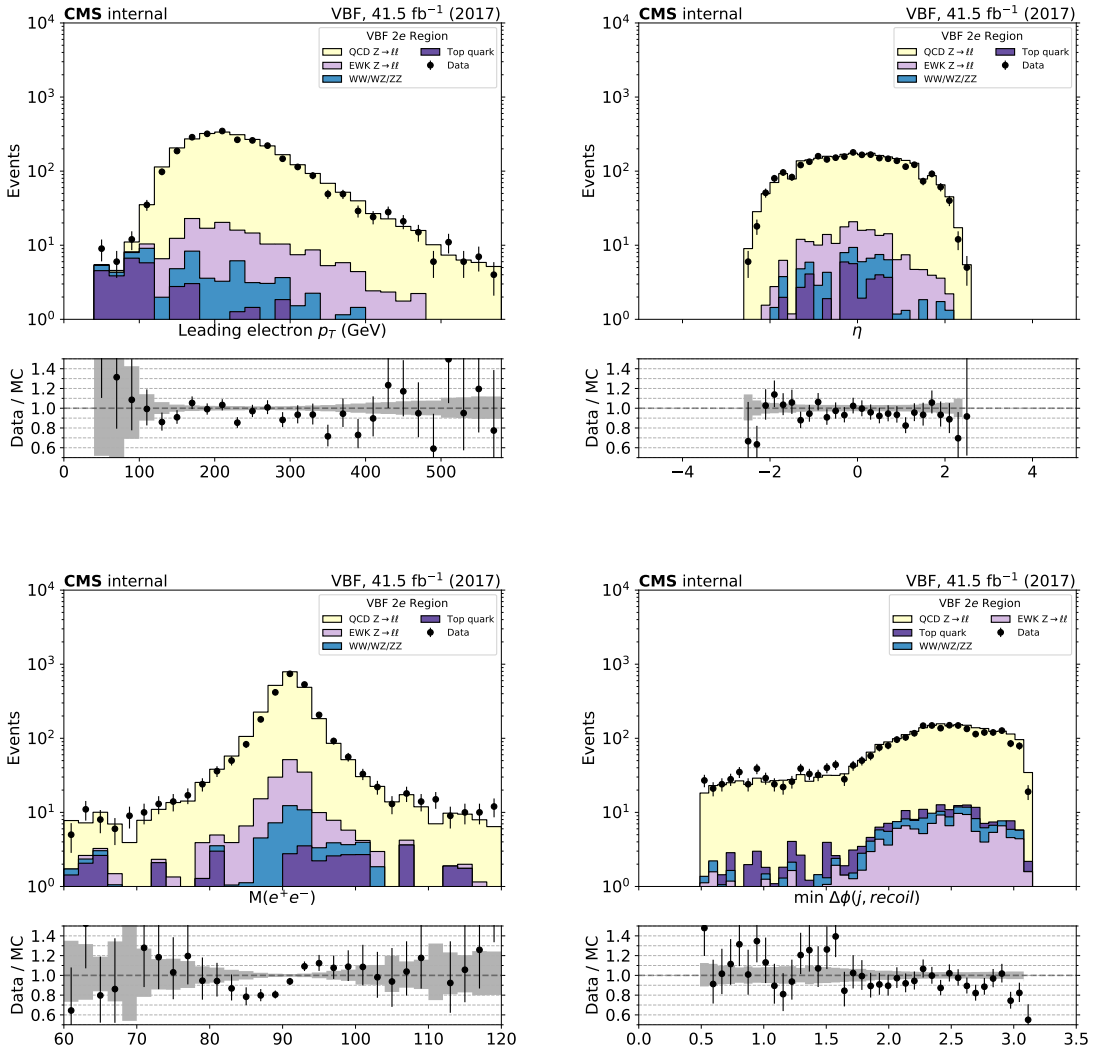

Figure 3·50: Comparison between 2017 data and Monte Carlo simulation in the double electron control region for the p_T and η of the leading electron (top), dielectron mass and minimum $\Delta\phi$ between 4 leading jets and recoil (bottom).

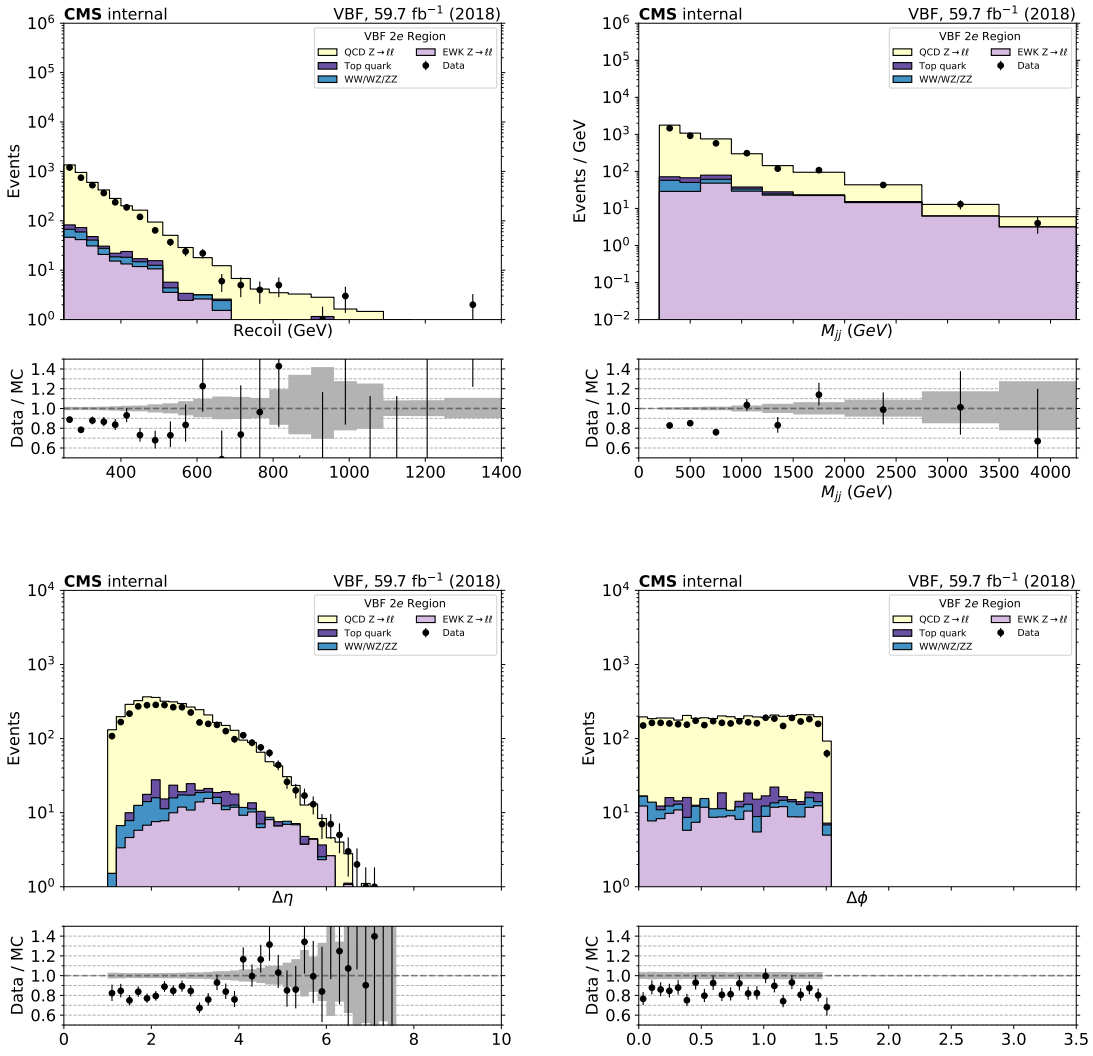

Figure 3·51: Comparison between 2018 data and Monte Carlo simulation in the double electron control region for the recoil, M_{jj} distribution, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ distributions for the two leading AK4 jets.

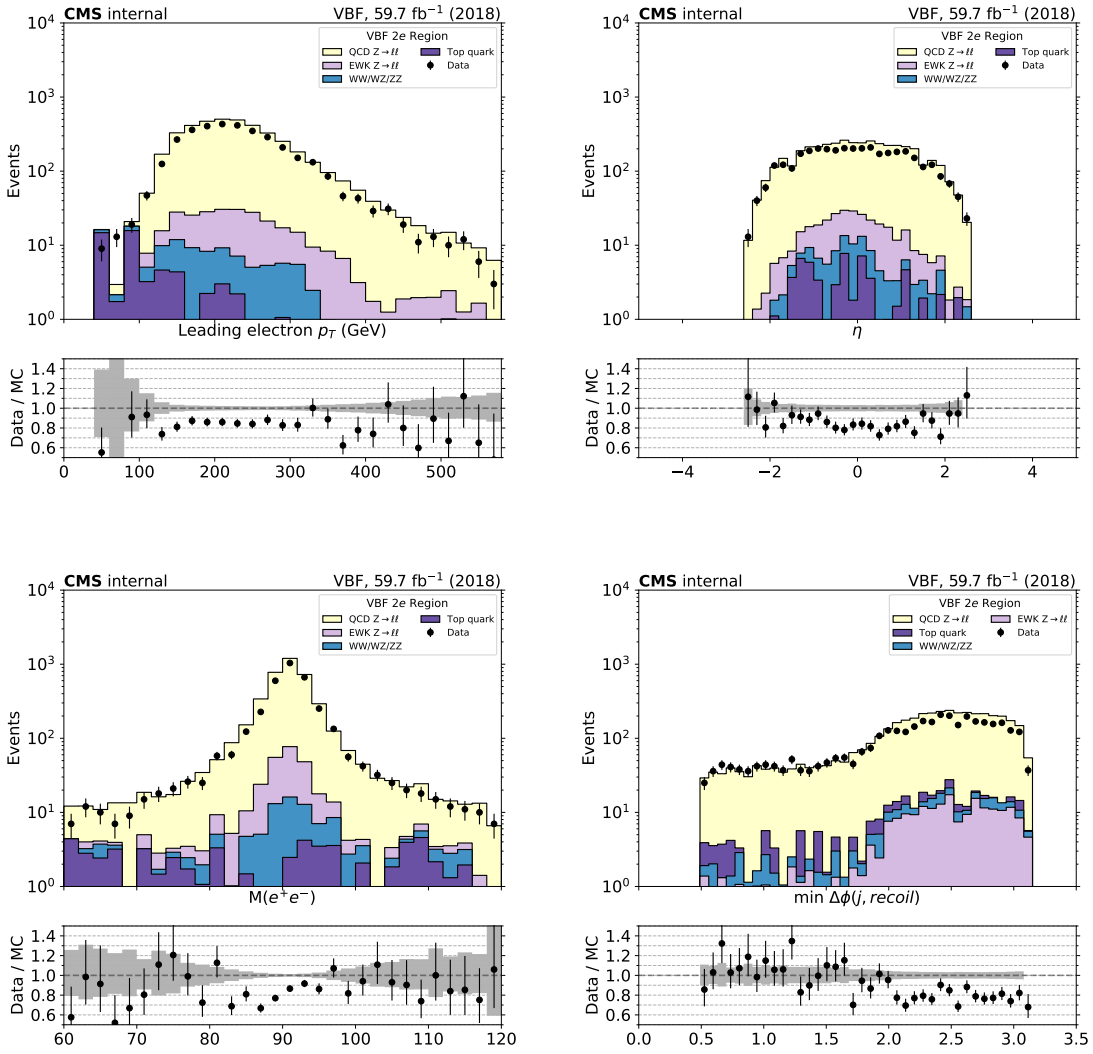

Figure 3·52: Comparison between 2018 data and Monte Carlo simulation in the double electron control region for the p_T and η of the leading electron (top), dielectron mass and minimum $\Delta\phi$ between 4 leading jets and recoil (bottom).

3.4.6 Photon Control Region

To further constrain the $Z \to \nu\nu$ background in the signal region, a photon control region is used. At large transverse momenta, the kinematic properties of photon production become similar to those of the $Z \to \nu\nu$ process, and can therefore be used to estimate the latter. Events for the control region are selected using a trigger requiring an online photon p_T of at least 200 GeV. In the offline analysis, photons are required to be located in the barrel part of the detector ($|\eta| < 1.4442$), have transverse momenta of at least 230 GeV to ensure full trigger efficiency, and pass additional identification criteria based on the properties of the associated supercluster in the ECAL, as well as the isolation of the photon relative to nearby energy objects. To be considered for the control region, events must have exactly one such photon, with no additional photons or leptons passing the loose criteria described above. Jets and hadronic recoil are required to pass criteria identical to those imposed in the signal region, where the recoil is defined as as the vectorial sum of $\vec{p}_T^{miss}$ and the photon transverse momentum (Eq. 3.2).

Figs. 3·53 and 3·55 show the distributions of the recoil, M_{jj}, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ distributions of the two leading AK4 jets for events in the photon control region for the VBF category in the 2017 and 2018 datasets, respectively. Similarly, Figs. 3·54 and 3·56 show the distributions of the photon p_T and η.

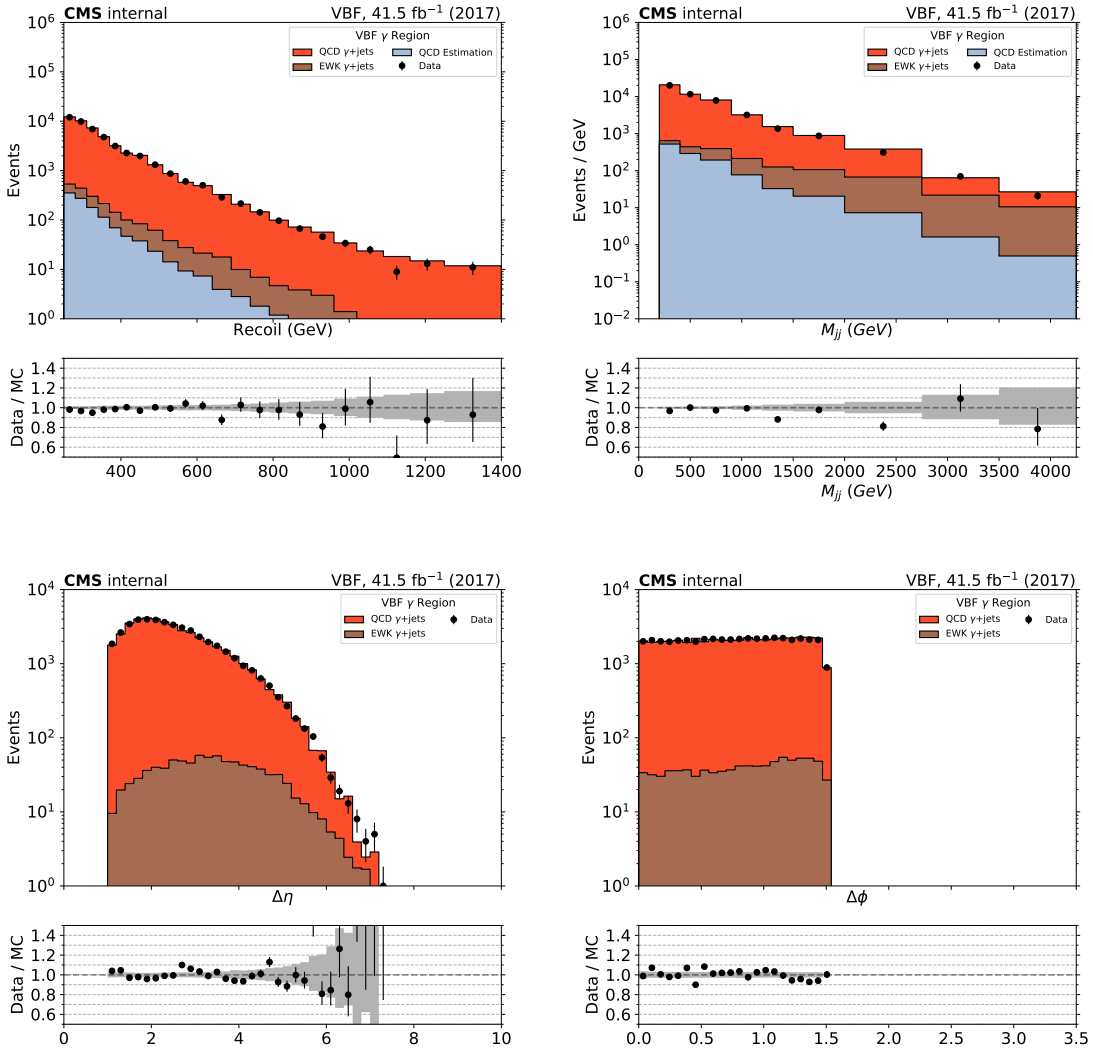

Figure 3·53: Comparison between 2017 data and Monte Carlo simulation in the photon control region for the recoil distribution, the M_{jj} distribution, $\Delta\eta_{jj}$ distribution and $\Delta\phi_{jj}$ distribution for the two leading AK4 jets. M_{jj} and recoil distributions include the QCD estimate calculated from the photon purity measurement, as described in Sec. 3.1.4.

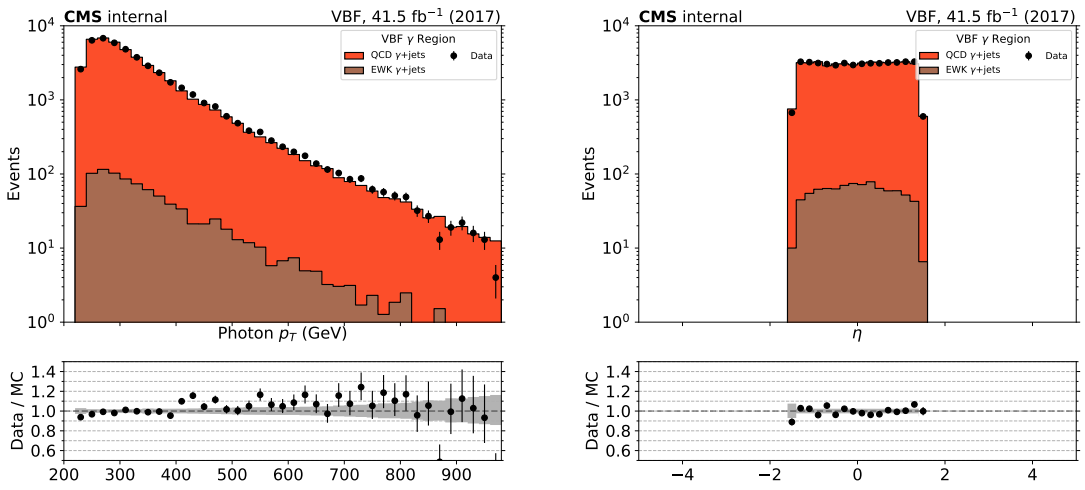

Figure 3·54: Comparison between 2017 data and Monte Carlo simulation in the photon control region for the p_T and η of the leading photon.

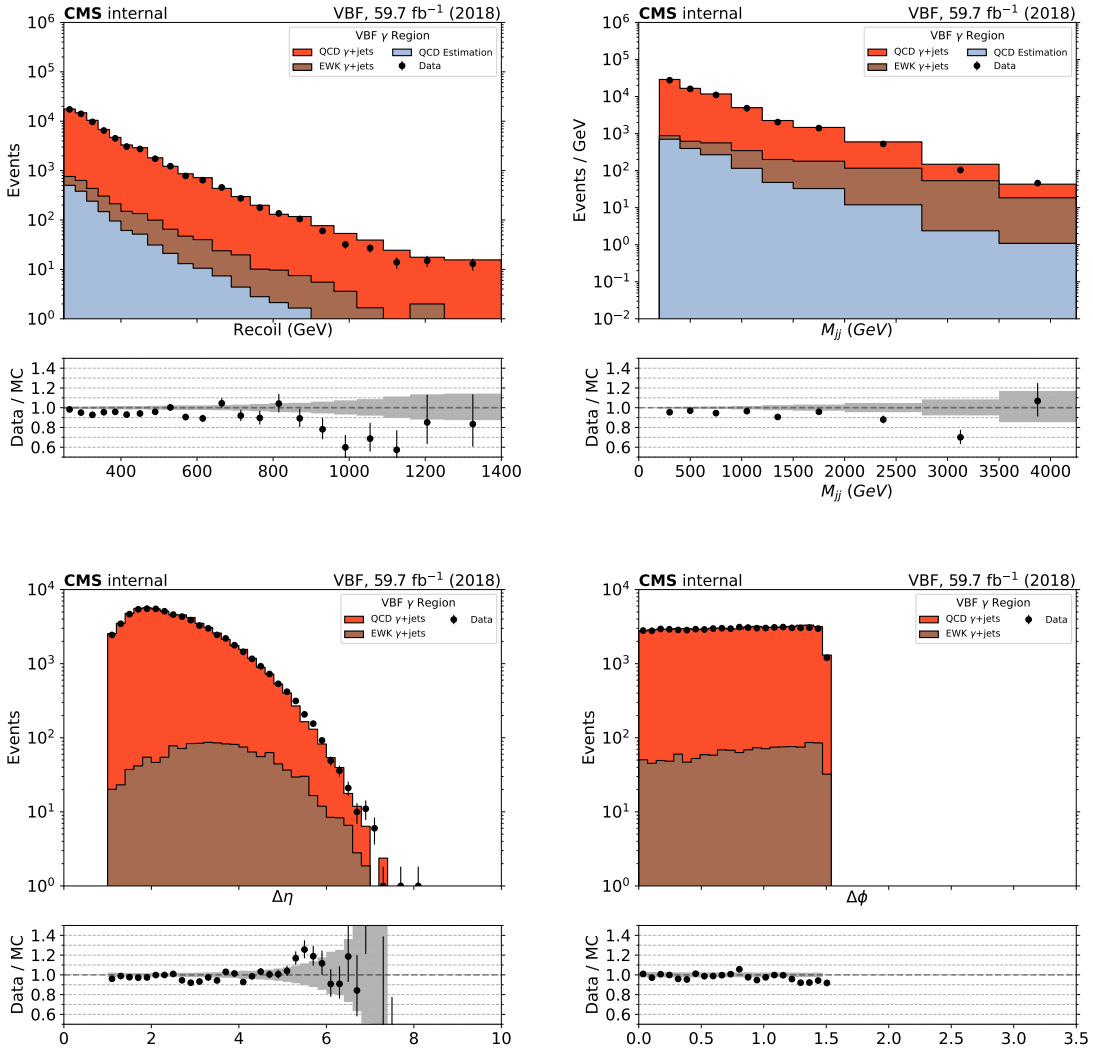

Figure 3·55: Comparison between 2018 data and Monte Carlo simulation in the photon control region for the recoil distribution, the M_{jj} distribution, $\Delta\eta_{jj}$ distribution and $\Delta\phi_{jj}$ distribution for the two leading AK4 jets. M_{jj} and recoil distributions include the QCD estimate calculated from the photon purtiy measurement, as described in Sec. 3.1.4.

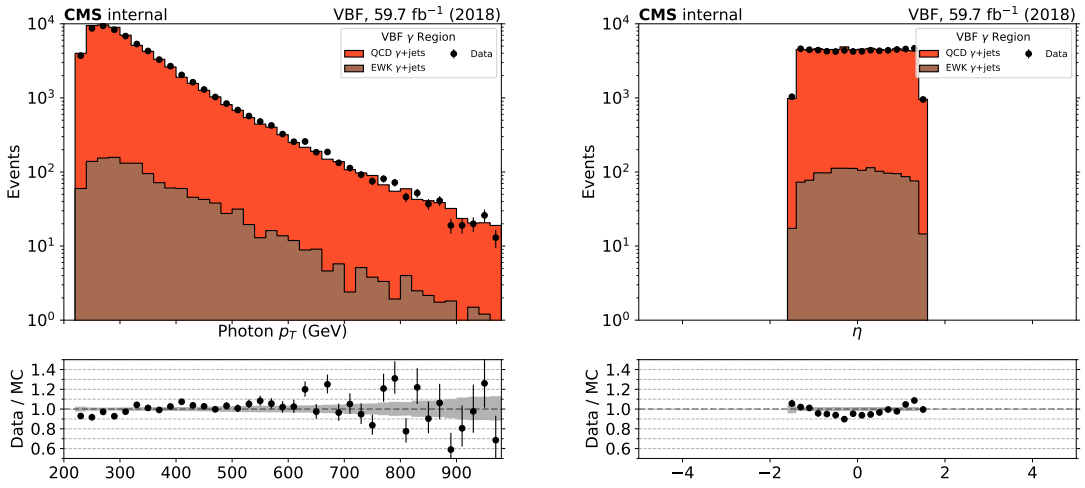

Figure 3·56: Comparison between 2018 data and Monte Carlo simulation in the photon control region for the p_T and η of the leading photon.

3.5 Signal extraction strategy

The largest background contributions from $\mathrm{Z}(\nu\nu)+\mathrm{jets}$ and $\mathrm{W}(\ell\nu)+\mathrm{jets}$ processes are estimated using data from five mutually exclusive control regions (CR), which are described in Sec. 3.4. These control regions consist of:

- $\mathrm{W}(\mu\nu)+\mathrm{jets}$ region (Sec. 3.4.2)
- $\mathrm{W}(e\nu)+\mathrm{jets}$ region (Sec. 3.4.3)
- $\mathrm{Z}(\mu\mu)+\mathrm{jets}$ region (Sec. 3.4.4)
- $\mathrm{Z}(ee)+\mathrm{jets}$ region (Sec. 3.4.5)
- γ + jets region (Sec. 3.4.6)

V+jets in the signal region is estimated by performing a simultaenous fit to the data in all signal and control regions, as will be explained in detail in Sec. 3.5.1.

The remaining backgrounds that contribute to the total event yield in the signal region are much smaller than those from $\mathrm{Z}(\nu\nu)+\mathrm{jets}$ and $\mathrm{W}(\ell\nu)+\mathrm{jets}$ processes. These backgrounds include the following:

- QCD multijet events, measured from data using a $\Delta\phi$ extrapolation method [4].
- Events from forward HCAL (HF) noise, measured from an orthogonal CR as explained in Sec. 3.3.1.
- Top and diboson processes, which are estimated from simulation.

3.5.1 Binned likelihood fit

A binned maximum likelihood fit to the observed data is performed simultaneously across the signal and control regions. For each M_{jj} bin i in every region, the expected

yield of events is parametrized as a function of expected strong $\mathrm{Z}(\nu\nu)+\mathrm{jets}$ yields in the signal region, labeled as $\kappa_i^{\nu\bar{\nu}}$ in the following. This parametrization is done using transfer factors between different V+jets processes, as will be explained later in this section. To scale the expected amount of VBF $H \to inv.$ signal (in case of the signal region), a signal strength parameter μ is also introduced to the fit, and left freely floating. The resulting likelihood function is shown in Eq. 3.15.

$$
\begin{aligned}
\mathcal{L}(\mu, \kappa^{\nu\bar{\nu}}, \boldsymbol{\theta}) =& \prod_i \mathrm{P}\left(d_i \middle| B_i(\boldsymbol{\theta}) + Z_i(\kappa_i^{\nu\bar{\nu}}) + W_i(\kappa_i^{\nu\bar{\nu}}, \boldsymbol{\theta}) + \mu S_i(\boldsymbol{\theta})\right) \\
& \prod_{\mathrm{CR}} \left(\prod_i \mathrm{P}\left(d_i^{\mathrm{CR}} \middle| B_i^{\mathrm{CR}}(\boldsymbol{\theta}) + V_i^{\mathrm{CR,strong}}(\kappa_i^{\nu\bar{\nu}}, \boldsymbol{\theta}) + V_i^{\mathrm{CR,VBF}}(\kappa_i^{\nu\bar{\nu}}, \boldsymbol{\theta})\right)\right) \\
& \prod_j \mathrm{P}(\theta_j), \\
Z_i(\kappa_i^{\nu\bar{\nu}}) =& (1 + Z_i^{\frac{\mathrm{VBF}}{\mathrm{strong}}})\kappa_i^{\nu\bar{\nu}}, \\
W_i(\kappa_i^{\nu\bar{\nu}}, \boldsymbol{\theta}) =& (f_i^{\mathrm{W/Z,strong}}(\boldsymbol{\theta}) + Z_i^{\frac{\mathrm{VBF}}{\mathrm{strong}}} f_i^{\mathrm{W/Z,VBF}}(\boldsymbol{\theta}))\kappa_i^{\nu\bar{\nu}}, \\
V_i^{\mathrm{CR,strong}}(\kappa_i^{\nu\bar{\nu}}, \boldsymbol{\theta}) =& C_i^{\mathrm{CR,strong}}(\boldsymbol{\theta}) R_i^{\mathrm{CR}}(\boldsymbol{\theta})\kappa_i^{\nu\bar{\nu}}, \\
V_i^{\mathrm{CR,VBF}}(\kappa_i^{\nu\bar{\nu}}, \boldsymbol{\theta}) =& C_i^{\mathrm{CR,VBF}}(\boldsymbol{\theta}) Z_i^{\frac{\mathrm{VBF}}{\mathrm{strong}}} R_i^{\mathrm{CR}}(\boldsymbol{\theta})\kappa_i^{\nu\bar{\nu}},
\end{aligned}
\tag{3.15}
$$

where $\mathrm{P}(x|y) = y^x e^{-y}/x!$. d_i^{CR} and d_i are the observed number of events in each bin i of the M_{jj} distribution in the CRs and SR, respectively. The index i runs over the M_{jj} bins in all regions and data taking years (i.e. 2017 and 2018).

In a given bin, V+jets background yields expected in the SR are obtained from transfer factors relating the yields in different CRs to the yields in the SR, denoted as $R_i^{\mathrm{CR}}(\boldsymbol{\theta})$. These transfer factors are obtained from simulation. For the single-lepton (dilepton) CRs, the factor $R_i^{\mathrm{CR}}(\boldsymbol{\theta})$ refers to the ratio of W+jets (Z+jets) yields from the corresponding CR to the SR. In the photon CR, it refers to the ratio of $\gamma+\mathrm{jets}$ to $\mathrm{Z}(\nu\nu)+\mathrm{jets}$ yields.

In addition, transfer factors are defined between the W (γ) and the Z processes, separately for the VBF and strong productions, denoted as $f_i^{\mathrm{W/Z,proc}}(\boldsymbol{\theta})$ ($f_i^{\gamma/\mathrm{Z,proc}}(\boldsymbol{\theta})$), with proc denoting the production mode (strong or VBF). Lastly, a transfer factor allows the relation of the VBF production to the strong production for $\mathrm{Z}(\nu\nu)+\mathrm{jets}$, denoted as $Z_i^{\frac{\mathrm{VBF}}{\mathrm{strong}}}$. The factors $C_i^{\mathrm{CR,strong}}(\boldsymbol{\theta})$ and $C_i^{\mathrm{CR,VBF}}(\boldsymbol{\theta})$ are dependent on the CR, with: $C_i^{(ee,\mu\mu),\mathrm{proc}}=1$, $C_i^{(e,\mu),\mathrm{proc}}=f_i^{\mathrm{W/Z,proc}}(\boldsymbol{\theta})$, $C_i^{\gamma,\mathrm{proc}}=f_i^{\gamma/\mathrm{Z,proc}}(\boldsymbol{\theta})$.

The contributions from subleading backgrounds in each region are estimated directly from simulation and they are denoted by $B_i^{\mathrm{CR}}(\boldsymbol{\theta})$ in the CRs, and $B_i(\boldsymbol{\theta})$ in the SR. Finally, the likelihood also includes a signal term in which S_i represents the expected signal prediction from the sum of the main Higgs production mechanisms (gg$H$, qq$H$, VH, tt$H$) assuming SM cross sections, while $\mu=(\sigma_H/\sigma_H^{\mathrm{SM}})\mathcal{B}(H\to\mathrm{inv})$ denotes the signal strength parameter, also left freely floating.

Systematic uncertainties are modeled as constrained nuisance parameters ($\boldsymbol{\theta}$), with a log-normal distribution for those which affect the overall normalisation of a given process, and Gaussian priors for those which directly affect the transfer factors, indicated by $\mathrm{P}(\theta_j)$ in Eq. 3.15.

Transfer factors

Transfer factors, derived from simulation, are used to link the yields of the $\mathrm{Z}(\ell\ell)$+jets, $\mathrm{W}(\ell\nu)+\mathrm{jets}$ and $\gamma+\mathrm{jets}$ processes in the control regions with the $\mathrm{Z}(\nu\nu)+\mathrm{jets}$ and $\mathrm{W}(\ell\nu)+\mathrm{jets}$ background estimates in the signal region. These transfer factors are defined as the ratio of expected yields of the target process in the signal region and the process being measured in the control region. As an example:

$$R_i^{Z(\mu\mu),\mathrm{strong}}(\boldsymbol{\theta})=\frac{N_{i,MC}^{\mathrm{Z}\to\nu\nu}}{N_{i,MC}^{\mathrm{Z}\to\mu^+\mu^-}}(\boldsymbol{\theta}) \tag{3.16}$$

where N_i is the number of events in bin i of the M_{jj} distribution, $R_i^{Z(\mu\mu),\mathrm{strong}}(\boldsymbol{\theta})$

is the transfer factor between the strong $\mathrm{Z}(\mu\mu)+\mathrm{jets}$ process yields in dimuon control region and strong $\mathrm{Z}(\nu\nu)+\mathrm{jets}$ background in the signal region. The transfer factor in Eq. 3.16 allows us to write the strong $\mathrm{Z}(\mu\mu)+\mathrm{jets}$ yields in dimuon control region as a function of strong $\mathrm{Z}(\nu\nu)+\mathrm{jets}$ yields in the signal region (i.e. $\kappa_i^{\nu\bar{\nu}}$ in Eq. 3.15):

$$V_i^{Z(\mu\mu),strong}(\kappa_i^{\nu\bar{\nu}},\boldsymbol{\theta}) = R_i^{Z(\mu\mu),\mathrm{strong}}(\boldsymbol{\theta}) \times \kappa_i^{\nu\bar{\nu}} \tag{3.17}$$

Other transfer factors are constructed in a similar manner, so that all $V_i^{\mathrm{CR,strong}}$ and $V_i^{\mathrm{CR,VBF}}$ can be written as a function of $\kappa_i^{\nu\bar{\nu}}$ and the nuisance parameters, $\boldsymbol{\theta}$.

Using this transfer factor formalism, $\mathrm{Z}(\nu\nu)+\mathrm{jets}$ background prediction in the signal region is connected to the yields of $\mathrm{Z}(\mu\mu)+\mathrm{jets}$ and $\mathrm{Z}(ee)+\mathrm{jets}$ events in the dilepton control regions. The associated transfer factors account for the differences in the branching ratio of Z bosons to charged leptons relative to neutrinos and the impact of lepton acceptance and selection efficiencies. In the case of dielectron events, the transfer factor also takes into account the difference in the trigger efficiencies. The resulting constraint on the $\mathrm{Z}(\nu\nu)+\mathrm{jets}$ background from the dilepton control regions is limited by the statistical uncertainty in the dilepton control regions due to the large branching fraction difference of the Z boson decays to muons and electrons compared to that to neutrinos.

Similarly, $\mathrm{W}(\ell\nu)+\mathrm{jets}$ background prediction in the signal region is connected to the yields of $\mathrm{W}(\mu\nu)+\mathrm{jets}$ and $\mathrm{W}(e\nu)+\mathrm{jets}$ event yields in single-lepton control regions. These transfer factors take into account the impact of lepton acceptances and efficiencies, lepton veto efficiencies, and the difference in the trigger efficiencies in the case of the single-electron control region.

The transfer factors linking Z and W processes in control regions are validated by comparing the ratio of data in different control regions to the predicted ratio in simulation. These transfer factors are shown as a function of M_{jj} in Fig. 3·57, where

the left column shows the results with 2017 data, and the right column shows the 2018 data.

To further constrain the $Z(\nu\nu) + \text{jets}$ background and to avoid statistical limitations, the $Z(\nu\nu) + \text{jets}$ process is also linked to the $\gamma + \text{jets}$ process in the photon CR, using the same transfer factor scheme as for the control regions with two charged leptons. The transfer factor accounts for all differences in triggering and identification between these two regions. Those transfer factors are shown in Fig. 3·58.

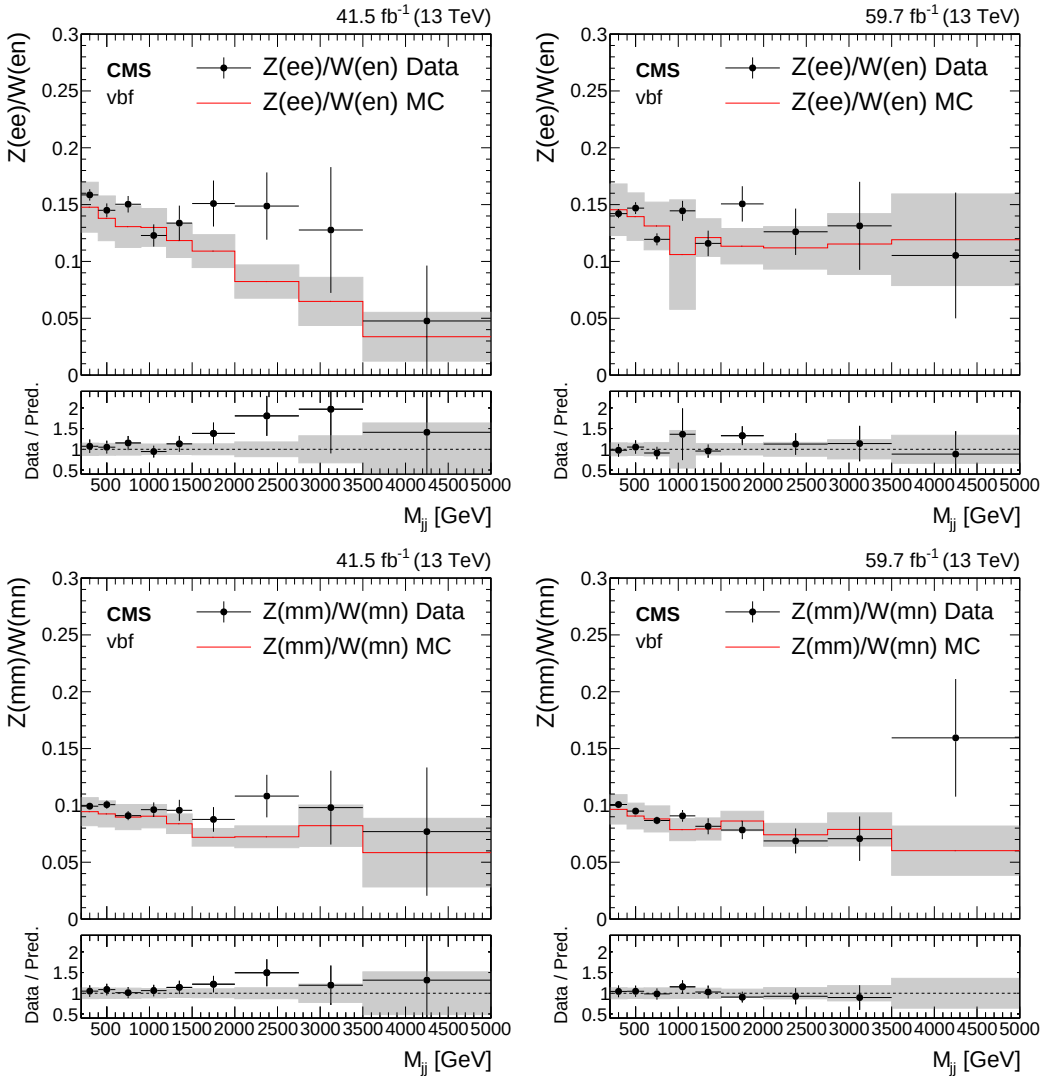

Figure 3·57: Transfer factors between Z and W control regions as a function of the dijet mass. Results for both 2017 (left) and 2018 (right) datasets are shown. Plots on the top row show the ratios between the electron regions, while the plots on the bottom show the ratios between muon regions. The bands show the theoretical and experimental systematic uncertainties on the ratios.

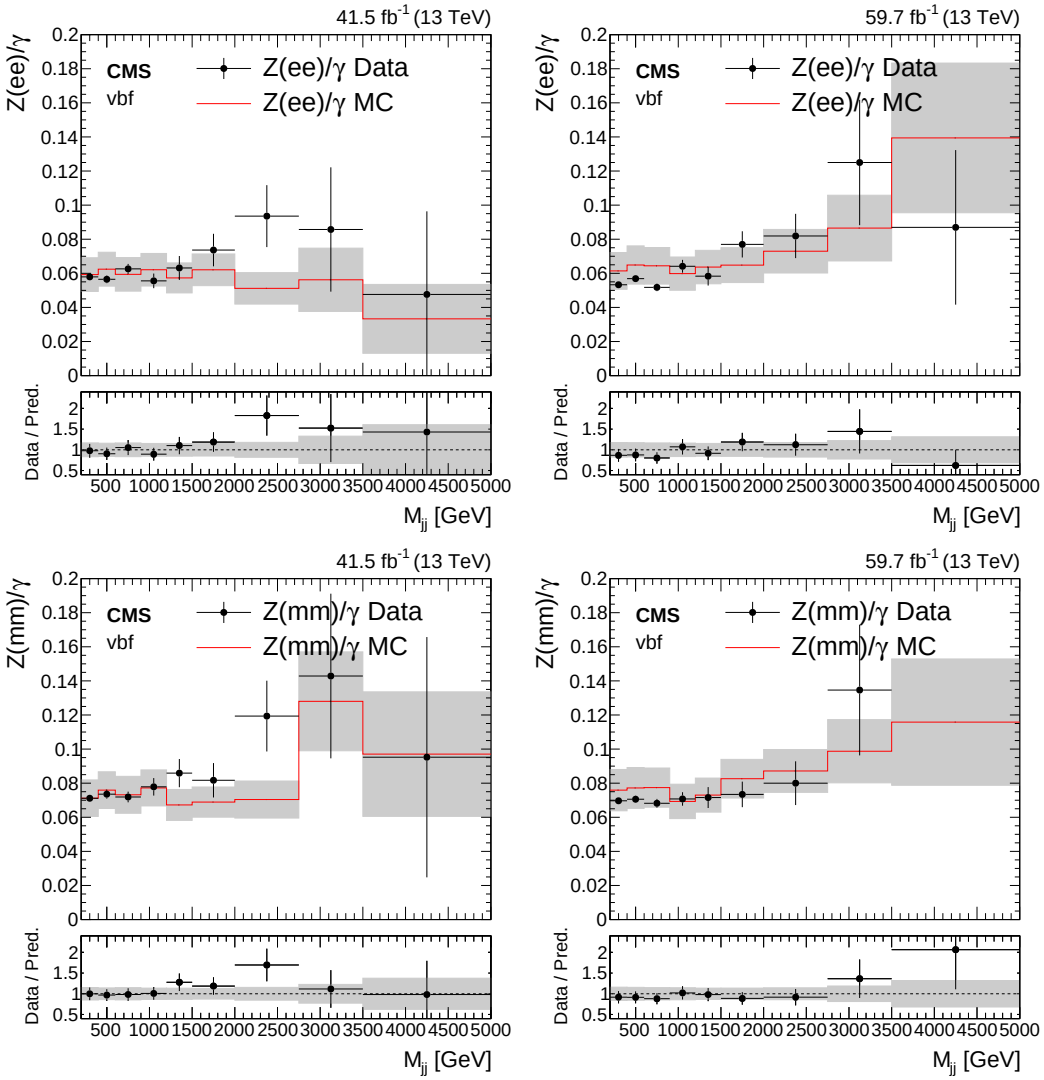

Figure 3·58: Transfer factors for $Z(ee)/\gamma + jet$ (top) and $Z(\mu\mu)/\gamma + jets$ (bottom) processes. Results with 2017 datasets are shown on the left column, and 2018 results are shown on the right column. The bands show the experimental systematic uncertainties on the ratios.

3.5.2 Systematic uncertainties

Systematic uncertainties in the transfer factors are modeled as constrained nuisance parameters and include both experimental uncertainties and theoretical uncertainties in W + jets to Z + jets and γ + jets to Z + jets cross section ratios. Theoretical and experimental uncertainties considered in the analysis are discussed in the following two subsections.

Theoretical uncertainties

Theoretical uncertainties in W + jets, Z + jets and γ + jets processes include effects from QCD and EWK higher-order corrections along with the parton distribution function (PDF) modeling uncertainty. One of the uncertainties considered comes from the variations around the central renormalization and factorization scale choice. It is evaluated by taking the differences in the NLO cross section two-dimensionally as a function of boson p_T and M_{jj} after changing the renormalization and factorization scales by a factor of two and a factor of one-half with respect to the default value. These constant scale variations mainly affect the overall normalization of the boson p_T distributions. The PDF uncertainty on the k-factors from the p_T/jet binned sample is evaluated using the recommendation from the PDF4LHC authors [97]. This is added in quadrature with uncertainty due to the choice of α_s. For the photon transfer factor, the procedure is identical with the exception that the uncertainties are estimated one-dimensionally versus boson p_T.

The scale uncertainties are treated as partially correlated between the Z+jets and W+jets processes in the following fashion. For a certain scale uncertainty component (e.g. the factorization scale) the W/Z ratio is evaluated from the Z + jets and W + jets processes separately and the difference to the nominal calculated. An envelope of the largest uncertainty out of the two processes is taken as the uncertainty on

the ratio. This is done for all of the theoretical uncertainty components on the ratios. It is observed that the contribution from varying the W + jets process is the larger uncertainty source, hence taking the envelope equates to taking the W + jets uncertainty contribution only and ignoring that from the Z.

The PDF uncertainties are treated as fully correlated between the Z + jets and W + jets processes. For up and down variations of the PDF, W/Z ratio is evaluated by varying the W + jets and Z + jets processes simultaneously, and the varied W/Z ratio is calculated accordingly. Both for the scale and PDF uncertainties, a similar correlation scheme is applied while computing the uncertainties for γ / Z + jets ratio.

The full set of theory uncertainties un the W/Z and γ/Z ratios are shown in Fig. 3·59. On the top row, the uncertainties for QCD W/Z ratio (left) and VBF W/Z ratio (right) are shown. And similarly, on the bottom row, the uncertainties for QCD γ/Z ratio (left) and VBF γ/Z ratio (right) are shown. All uncertainties are shown as a function of the dijet invariant mass, M_{jj}.

From Fig. 3·59, it can be observed that the dominating theory uncertainty, is the renormalization scale uncertainty, which reaches to around 10% at lower M_{jj}. It can also be observed that the PDF uncertainties are typically smaller compared to other theoretical uncertainty sources. This is mainly due to them being fully correlated between different V+jets processes, resulting in large cancellations in the ratios.

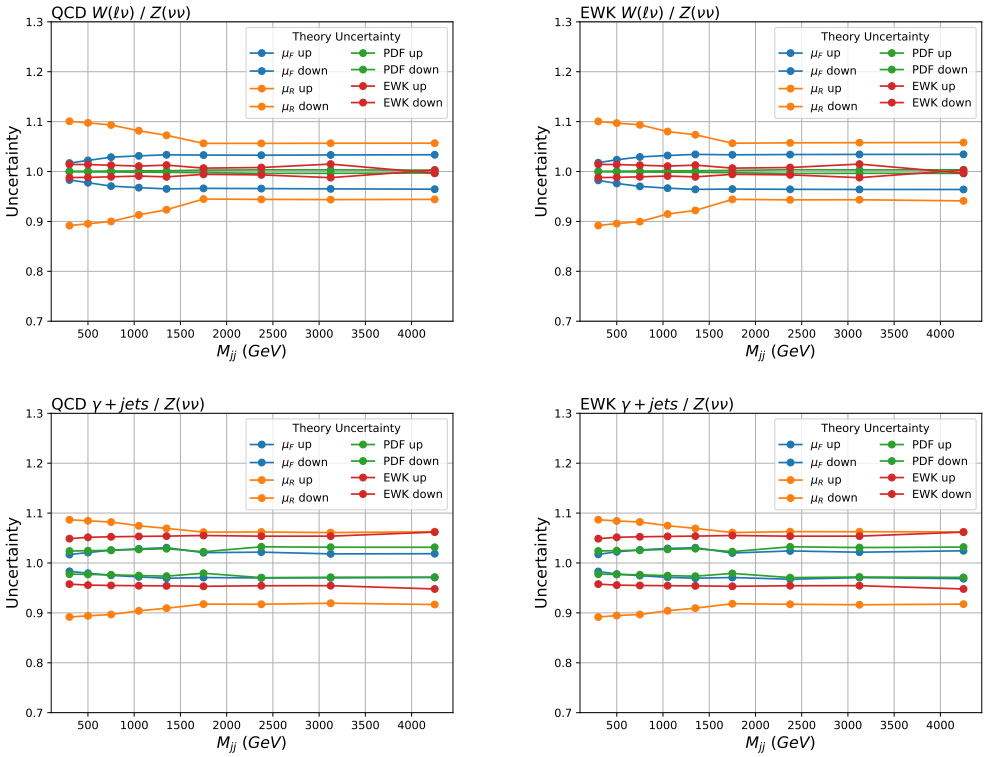

Figure 3·59: Theoretical uncertainties on W/Z (top) and γ/Z (bottom) transfer factors. Uncertainties are calculated as a function of M_{jj}. Uncertainties for QCD ratios are shown on the left column, while uncertainties for EWK ratios are shown on the right column.

Experimental uncertainties

Experimental uncertainties include uncertainties on the lepton reconstruction and identification criteria, jet energy scale and resolution, pileup reweighting and prefire reweighting.

Uncertainties on veto weights are applied for electrons, muons, taus and b-jets. For the case of electrons and muons, the uncertainties are split into identification and isolation uncertainties, in accordance with the weight definitions for these objects, as defined in Sec. 3.2.3. These uncertainties are typically not correlated with the kinematics of the two VBF jets, and hence M_{jj}. Therefore, no significant shape as a function of M_{jj} is observed for these uncertainties, and flat uncertainties are applied. These uncertainties are summarized in Table. 3.11.

The uncertainty on the prefire reweighting, as explained in Sec. 3.2.6, is computed by varying the prefire weight within its uncertainties, and computing the impact on the M_{jj} distribution. The variations in the M_{jj} shape are computed from the VBF $H \rightarrow inv.$ signal sample, and the resulting shapes are applied to all $H \rightarrow inv.$ samples as a shape uncertainty. For minor backgrounds, a 3% flat uncertainty is applied instead, which is observed to be a good approximation of the uncertainty. Since prefire reweighting is done only on 2017 data, this uncertainty is only applied to 2017 data. The uncertainty on the prefiring weights as a function of M_{jj} is shown in Fig. 3·60. Similar to the prefire reweighting uncertainty, the uncertainty on pileup reweighting (explained in Sec. 3.2.2) is done by varying the pileup weight within its uncertainties, and computing the impact on the M_{jj} distribution per process. The uncretainty due to the pileup reweighting is found to be $O(1\%)$.

The uncertainty in the modeling of p_T^{miss} in simulation [98] is dominated by the uncertainty on the jet energy scale (JES) and resolution (JER). The effect is estimated by varying the p_T of VBF jets within their uncertainty, propagating the effect to

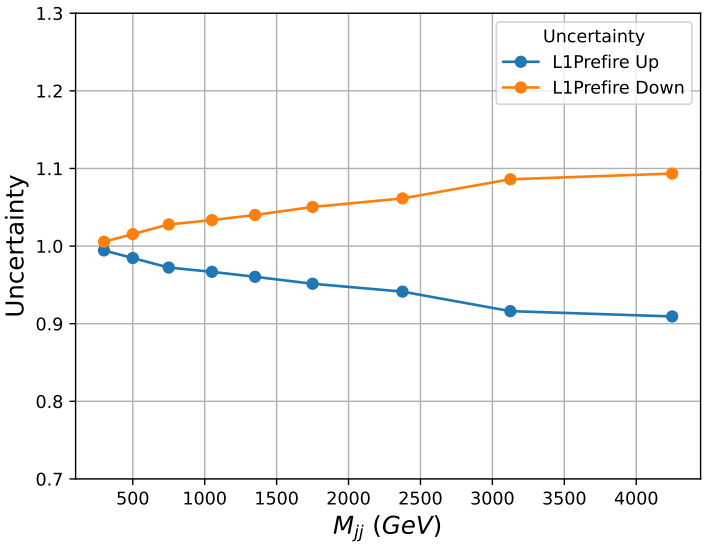

Figure 3·60: Uncertainty due to the L1 prefire reweighting as a function of M_{jj}. The uncertainty is computed from VBF $H \to inv.$ signal sample by varying the prefiring weight within its uncertainty. This uncertainty is applied as a shape uncertainty to all $H \to inv.$ signals considered in the analysis.

p_T^{miss}, and then performing the full analysis selection based on the varied inputs. For JES uncertainties, this variation is done for 11 sub-sources, where each sub-source is defined in accordance with the correlation scheme defined by JetMET POG, which can be found in [99]. The uncertainties for JES and JER are applied both to the transfer factors, and to the individual processes such as signal processes and minor backgrounds.

In the transfer factors, the majority of the JES and JER uncertainties cancel, but residual non-cancellation is observed. The residual uncertainty remaining in the ratios is found to be within approximately 2%. For the transfer factors, a flat single-bin uncertainty is assigned for each JES and JER uncertainty source for each ratio. All the JES and JER uncertainties for the Z(SR)/Z(CR) (Z($\mu\mu$) + jets and Z(ee) + jets channels combined for Z(CR)) ratio are shown in Fig. 3·61. The uncertainties for Z(SR)/γ are shown in Fig 3·62.

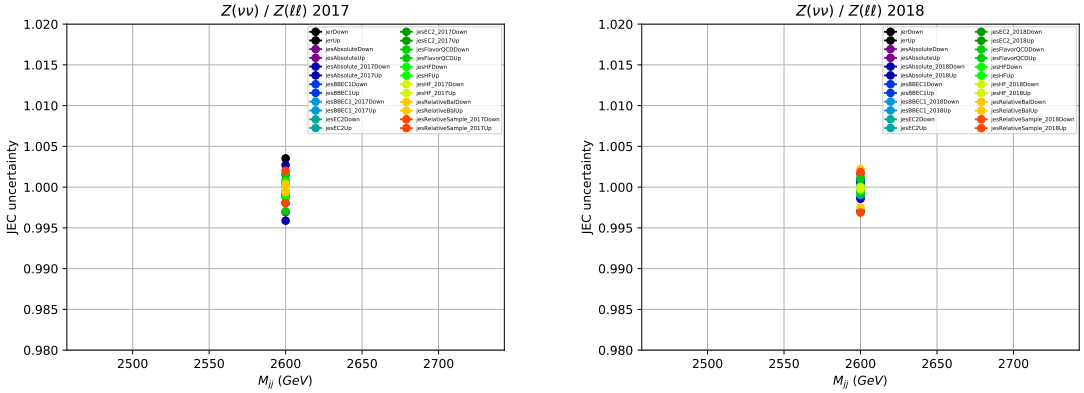

Figure 3·61: All single-bin JES/JER uncertainties on QCD Z(SR)/Z(CR) ratio for 2017 (left) and 2018 (right). The black dot shows the JER uncertainty, and the others show up and down variation from all 11 JES sub-sources. All jet energy scale and resolution uncertainties cancel to within less than 1%.

JES and JER uncertainties on minor backgrounds and signals are calculated in a very similar way as for the transfer factors. For these, the uncertainties are derived as

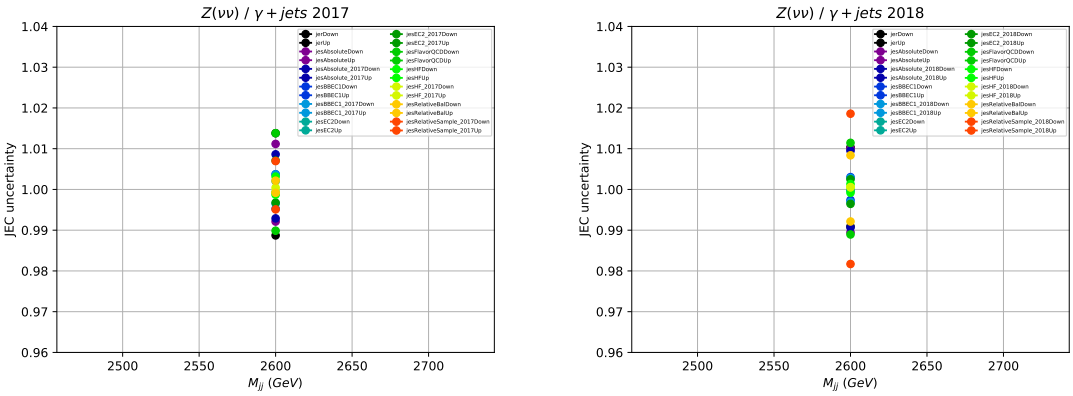

Figure 3·62: All single-bin JES/JER uncertainties on QCD Z(SR)/γ ratio for 2017 (left) and 2018 (right). The black dot shows the JER uncertainty, and the others show up and down variation from all 11 JES sub-sources. All jet energy scale and resolution uncertainties cancel within up to 2%.

a function of M_{jj} for each jet energy uncertainty source. The uncertainties for minor backgrounds (top, diboson) are derived using the strong Z($\nu\nu$) + jets sample, due to the limited statistics from the minor backgrounds. The uncertainties for $H \to inv.$ signal samples are derived from VBF $H \to inv.$ sample. Most jet energy uncertainty sources are found to be on the order of $\mathcal{O}(1\%)$, while the uncertainty coming from the relative corrections are typically found to be the dominating ones, also displaying an increasing shape as a function of M_{jj}, reaching to $\mathcal{O}(10\%)$ uncertainty levels at high M_{jj}.

The most dominating uncertainty source is the "Relative Sample" uncertainty, which is the jet η-dependent uncertainty due to different residual jet energy corrections obtained by measurement from different channels, such as dijet events, $\gamma + jet$ events and $Z + jet$ events. For these larger uncertainty sources, a second-degree polynomial is fit to smooth out the uncertainty shape. For smaller jet energy uncertainty sources, a line fit is performed instead. For most jet energy uncertainty sources, no significant shape is observed as a function of M_{jj}. The dominating jet energy uncer-

tainty source, "Relative Sample", is plotted in Fig. 3·63 for QCD Z($\nu\nu$) + jets and VBF $H \to inv.$ samples.

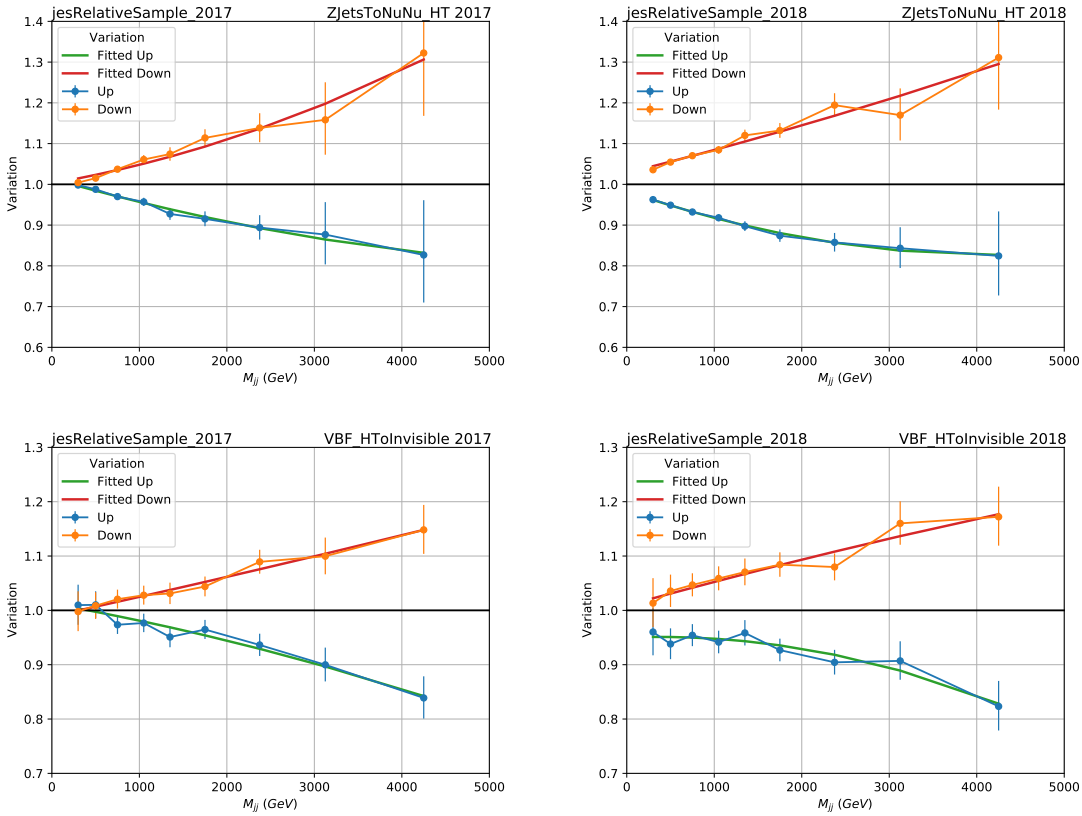

Figure 3·63: Relative sample JEC uncertainties calculated with strong Z($\nu\nu$) + jets (top) and VBF $H \to inv.$ signal (bottom). Left column shows the 2017 results, while the right column shows the 2018 results. The fitted uncertainties, shown in solid lines, are used as the final uncertainty shapes.

As discussed in Sec. 3.3.1, a 20% flat uncertainty is applied to the data-driven HF noise estimate, to take residual differences between data and expected background yields into account.

Uncertainties on trigger efficiency reweighting, as discussed in Sec. 3.2.1, are also applied in the analysis. For the electron, photon and p_T^{miss} trigger reweightings, flat uncertainties are applied on the transfer factors. The magnitudes of the flat

uncertainties are 1%, 1% and 2% respectively.

List of all theoretical and experimental uncertainties on transfer factors are shown in Tab. 3.11, together with the transfer factors they are applied on, and their magnitudes.

Table 3.11: Experimental and theoretical sources of systematic uncertainties in the V+jets transfer factors. The second column highlights on which ratio specifically a given source of uncertainty acts. The subscript SR (CR) refers to the process yield in the SR (corresponding CRs). The impact on M_{jj} is given in the 3rd column, either as a single value (if no dependence on M_{jj} is observed) or as a range of impact on low to high M_{jj} values.

Source of uncertainty	Ratios	Uncertainty vs. M_{jj}
Theoretical uncertainties		
Ren. scale V+jets (VBF)	$Z_{\mathrm{SR}}/W_{\mathrm{SR}}$	5–10%
Ren. scale V+jets (strong)	$Z_{\mathrm{SR}}/W_{\mathrm{SR}}$	5–10%
Fac. scale V+jets (VBF)	$Z_{\mathrm{SR}}/W_{\mathrm{SR}}$	1.5%
Fac. scale V+jets (strong)	$Z_{\mathrm{SR}}/W_{\mathrm{SR}}$	1.3%
PDF V+jets (VBF)	$Z_{\mathrm{SR}}/W_{\mathrm{SR}}$	0%
PDF V+jets (strong)	$Z_{\mathrm{SR}}/W_{\mathrm{SR}}$	0%
NLO EW corr. V+jets (strong)	$Z_{\mathrm{SR}}/W_{\mathrm{SR}}$	0.5%
Ren. scale $\gamma+$jets (VBF)	$Z_{\mathrm{SR}}/\gamma_{\mathrm{CR}}$	6–10%
Ren. scale $\gamma+$jets (strong)	$Z_{\mathrm{SR}}/\gamma_{\mathrm{CR}}$	6–10%
Fac. scale $\gamma+$jets (VBF)	$Z_{\mathrm{SR}}/\gamma_{\mathrm{CR}}$	2.5%
Fac. scale $\gamma+$jets (strong)	$Z_{\mathrm{SR}}/\gamma_{\mathrm{CR}}$	2.5%
PDF $\gamma+$jets (VBF)	$Z_{\mathrm{SR}}/\gamma_{\mathrm{CR}}$	2.5%
PDF $\gamma+$jets (strong)	$Z_{\mathrm{SR}}/\gamma_{\mathrm{CR}}$	2.5%
NLO EW corr. $\gamma+$jets	$Z_{\mathrm{SR}}/\gamma_{\mathrm{CR}}$	3%
Experimental uncertainties		
Electron reco. eff.	$Z_{\mathrm{CR}}/Z_{\mathrm{SR}}$, $W_{\mathrm{CR}}/W_{\mathrm{SR}}$	$\approx 0.5\%$ (per lepton)
Electron id. eff.	$Z_{\mathrm{CR}}/Z_{\mathrm{SR}}$, $W_{\mathrm{CR}}/W_{\mathrm{SR}}$	$\approx 1\%$ (per lepton)
Muon id. eff.	$Z_{\mathrm{CR}}/Z_{\mathrm{SR}}$, $W_{\mathrm{CR}}/W_{\mathrm{SR}}$	$\approx 0.5\%$ (per lepton)
Muon iso. eff.	$Z_{\mathrm{CR}}/Z_{\mathrm{SR}}$, $W_{\mathrm{CR}}/W_{\mathrm{SR}}$	$\approx 0.1\%$ (per lepton)
Photon id. eff.	Z_{SR}/γ	5%
Electron veto (reco)	$Z_{\mathrm{SR}}/W_{\mathrm{SR}}$, $W_{\mathrm{CR}}/W_{\mathrm{SR}}$	≈ 1.5 (1)% for VBF (strong)
Electron veto (id)	$Z_{\mathrm{SR}}/W_{\mathrm{SR}}$, $W_{\mathrm{CR}}/W_{\mathrm{SR}}$	≈ 2.5 (2)% for VBF (strong)
Muon veto	$Z_{\mathrm{SR}}/W_{\mathrm{SR}}$, $W_{\mathrm{CR}}/W_{\mathrm{SR}}$	$\approx 0.5\%$
τ_h veto	$Z_{\mathrm{SR}}/W_{\mathrm{SR}}$, $W_{\mathrm{CR}}/W_{\mathrm{SR}}$	$\approx 1\%$
Electron trigger	$Z_{\mathrm{CR}}/Z_{\mathrm{SR}}$, $W_{\mathrm{CR}}/W_{\mathrm{SR}}$	$\approx 1\%$
p_T^{miss} trigger	$Z_{\mathrm{CR}}/Z_{\mathrm{SR}}$, $W_{\mathrm{CR}}/W_{\mathrm{SR}}$	$\approx 2\%$
Photon trigger	Z_{SR}/γ	1%
JES	$Z_{\mathrm{SR}}/W_{\mathrm{SR}}$	1–2%
	$W_{\mathrm{CR}}/W_{\mathrm{SR}}$	1.0–1.5%
	$Z_{\mathrm{CR}}/Z_{\nu\nu}$	1%
	Z_{SR}/γ	3%
JER	$Z_{\mathrm{SR}}/W_{\mathrm{SR}}$	1.0–2.5%
	$W_{\mathrm{CR}}/W_{\mathrm{SR}}$	1.0–1.5%
	$Z_{\mathrm{CR}}/Z_{\mathrm{SR}}$	1%
	Z_{SR}/γ	1–4%

Chapter 4

Results

4.1 Combined likelihood fit

To get the best-fit value of the signal strength $\mu = (\sigma_H/\sigma_H^{\mathrm{SM}})\mathcal{B}(H \to \mathrm{inv})$, a maximum-likelihood fit is performed across all different signal and control regions in the analysis, using the likelihood function given in Eq. 3.15. In this fit, data from 2017 and 2018 are combined. Theoretical uncertainties are treated as correlated between different years, while some experimental uncertainties, such as the jet energy scale uncertainties, are partially correlated between different years, depending on the uncertainty source. The partial correlation scheme is in accordance with the recommendations from the relevant POGs. The theoretical and experimental uncertainties are discussed in depth in Sec. 3.5.2.

Figs. 4·1, 4·2, 4·3 show the results of the 2017+2018 combined fit in all control regions and the signal region of the analysis. Data observed in each region is compared to the background predictions from simulation and data-driven estimates (pre-fit) and background predictions after the fit (post-fit). Overall, it can be observed that post-fit background estimates are in very good agreement with the observed data, and no significant pulls ($data - prediction/uncertainty > 2\sigma$) are observed when fitting the data.

For the signal region plots in Fig. 4·1 (top row), two data-driven multijet background estimates have been shown, which represent the events where p_T^{miss} arises from mismeasured jets. The first multijet estimate, labeled as "HF Estimate", represents

the noise estimation related to events with forward jets in Forward HCAL (HF) detector, where the mismeasured jet is balanced with p_T^{miss}, therefore $\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{jet}},\vec{p}_T^{\,miss})$ is large. The derivation of this estimate is explained in Sec. 3.3.1.

The second data-driven estimate is labeled as "QCD", which is an independent estimate of events where the mismeasured jet is aligned with p_T^{miss}, therefore $\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{jet}},\vec{p}_T^{\,miss})$ is small. This background is estimated from another control region, where the $\Delta\phi(\vec{p}_{\mathrm{T}}^{\,\mathrm{jet}},\vec{p}_T^{\,miss}) > 0.5$ selection (listed in Table. 3.10) is inverted, as explained in [4].

The observed event yields for each M_{jj} bin, and the corresponding expected event yields from each background process are summarized in Tables 4.1 and 4.2.

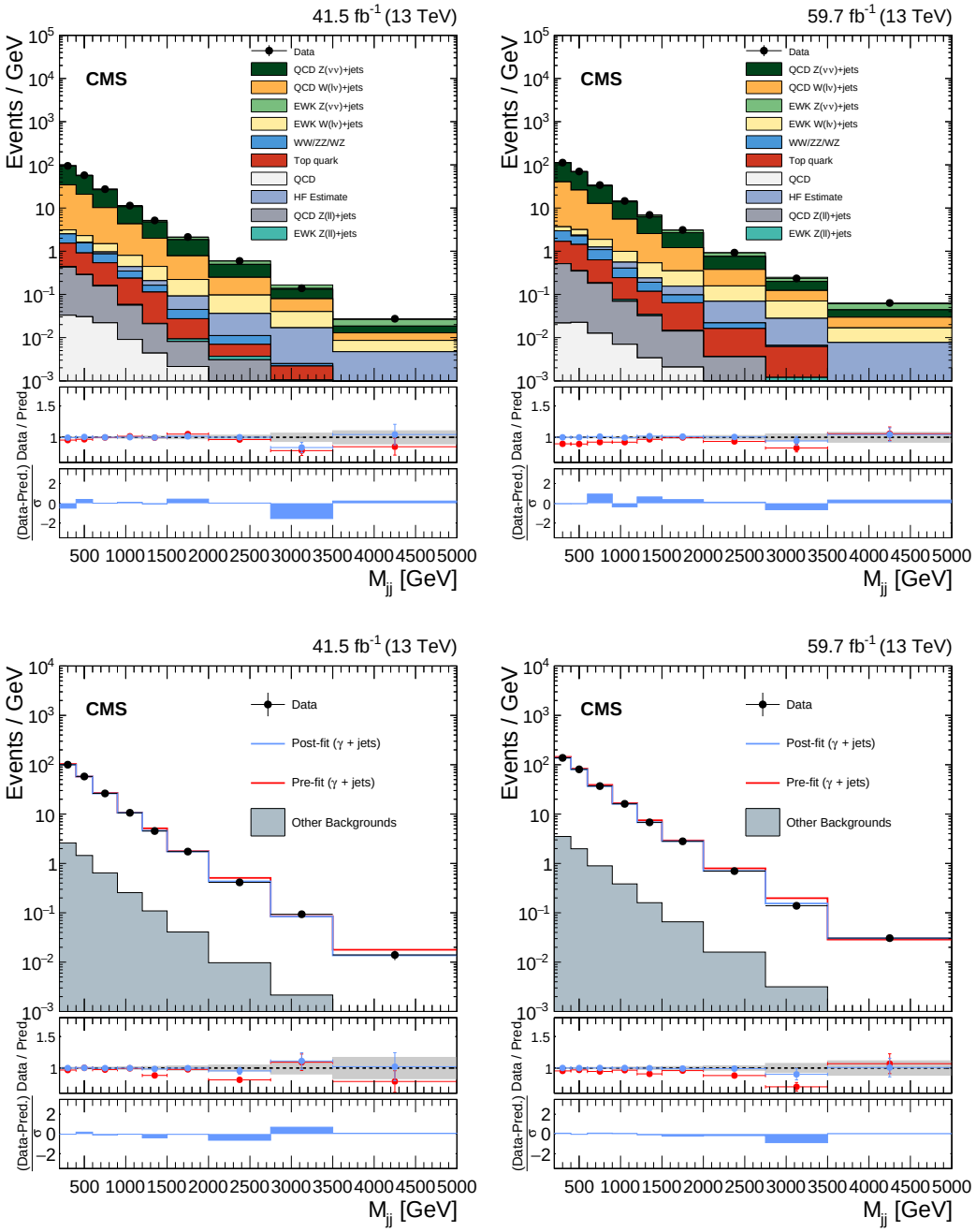

Figure 4·1: Comparison between data and background estimation in VBF SR (top) and γ + jets CR (bottom), before and after the simultaneous fit. The fit includes the data in all CRs and the signal region. The resulting distributions are shown separately for 2017 (left) and 2018 (right). In the ratio pads, ratios of data and background estimation are shown before (red) and after (blue) the fit. The gray band indicates the post-fit uncertainty. Finally, the distribution of the difference between data and post-fit background prediction relative to the quadrature sum of the uncertainties in the prediction and in data is shown in the lowest panel.

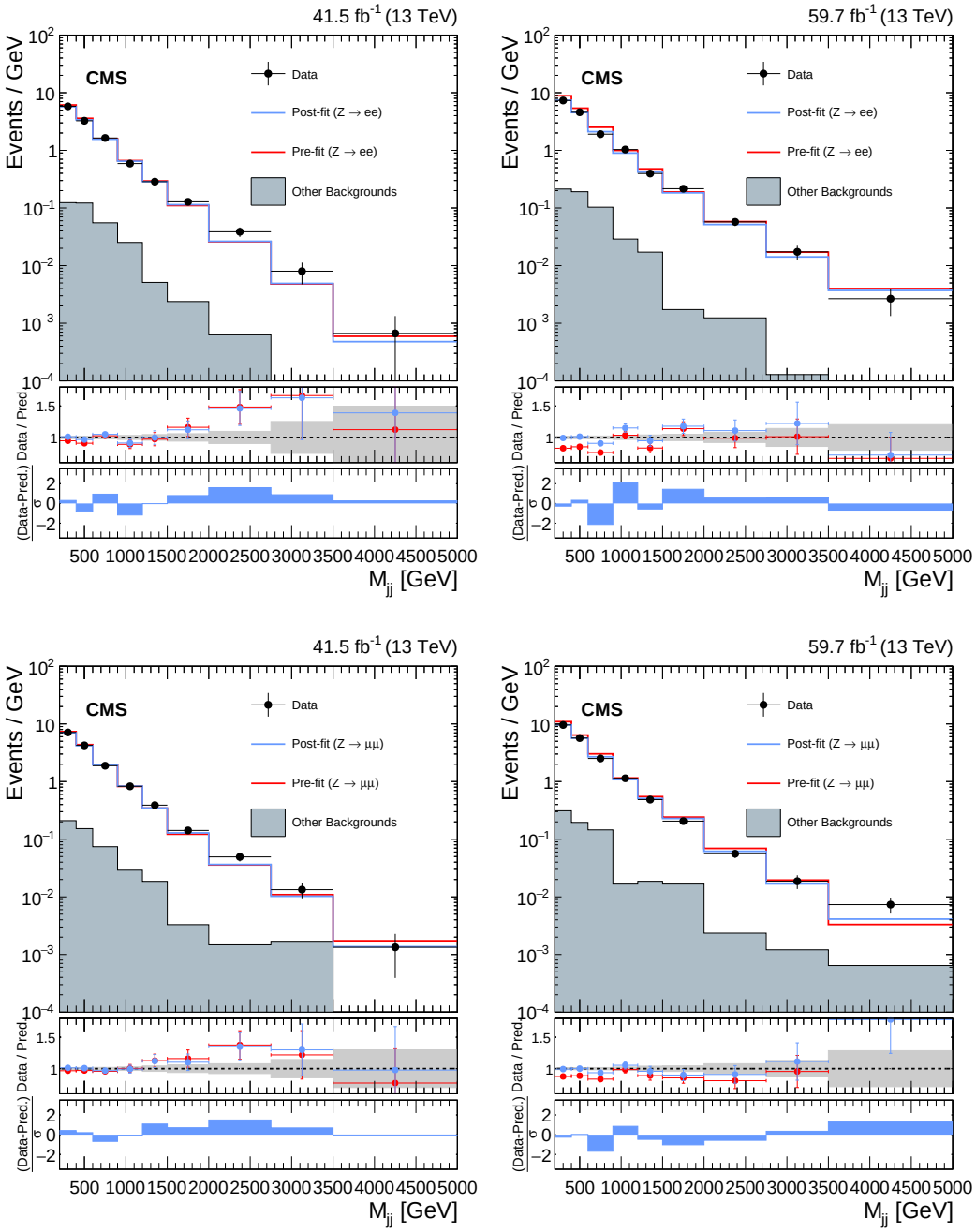

Figure 4·2: Same as Fig. 4·1 but with dilepton CRs. The other backgrounds include top quark and diboson processes.

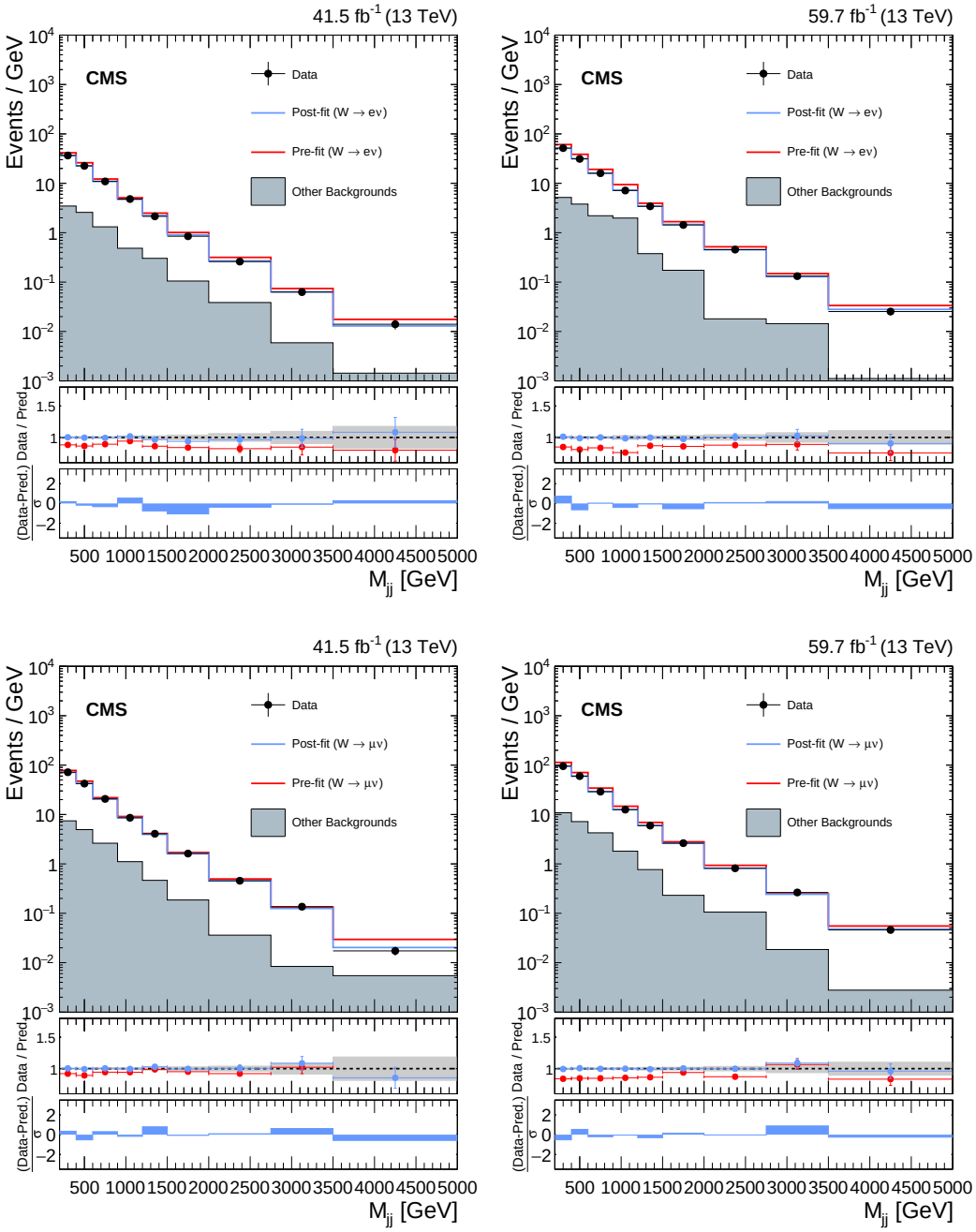

Figure 4·3: Same as Fig. 4·1 but with single lepton CRs. The other backgrounds include top quark and diboson processes.

Table 4.1: Expected event yields in each M_{jj} bin for the different background processes in the VBF signal region, with the 2017 samples. The background yields and the corresponding uncertainties are obtained after performing a combined fit across all of the CRs and SR. The expected signal contributions for a Higgs boson, produced in the non-VBF and VBF modes, decaying to invisible particles with a branching fraction of $\mathcal{B}(H \to \text{inv}) = 1$, and the observed event yields are also reported.

M_{jj}bin range (GeV)	200-400	400-600	600-900	900-1200	1200-1500	1500-2000	2000-2750	2750-3500	>3500
$Z(\nu\nu)+$jets (strong)	11957.1 ± 55.5	7022.4 ± 42.2	4855.8 ± 34.6	1914.1 ± 17.6	826.8 ± 11.4	531.3 ± 8.5	183.5 ± 4.7	39.6 ± 4.1	8.3 ± 0.9
$Z(\nu\nu)+$jets (VBF)	202.5 ± 4.1	222.2 ± 4.1	272.3 ± 4.3	197.6 ± 3.8	127.2 ± 3.2	126.4 ± 3.6	74.0 ± 2.9	25.3 ± 2.9	11.5 ± 1.4
$W(\ell\nu)+$jets (strong)	6247.9 ± 57.1	3727.1 ± 36.6	2624.7 ± 31.6	1052.3 ± 15.7	450.0 ± 11.7	285.5 ± 7.1	116.5 ± 4.9	27.1 ± 2.7	5.1 ± 1.0
$W(\ell\nu)+$jets (VBF)	122.6 ± 7.2	137.9 ± 7.5	161.9 ± 8.1	109.4 ± 5.3	72.3 ± 3.5	65.8 ± 3.1	45.7 ± 2.9	17.5 ± 1.8	5.9 ± 0.8
$t\bar{t}$ + single-top	237.6 ± 16.0	135.8 ± 9.1	124.0 ± 8.0	60.1 ± 3.7	30.7 ± 2.0	9.7 ± 0.8	2.7 ± 0.3	0.9 ± 0.2	0.4 ± 0.1
Diboson	201.0 ± 24.8	132.9 ± 16.0	101.7 ± 12.2	34.4 ± 4.2	15.8 ± 1.9	9.2 ± 1.2	3.3 ± 0.5	0.3 ± 0.0	0.0 ± 0.0
$Z/\gamma^*(\ell^+\ell^-)+$jets	86.6 ± 3.3	54.9 ± 2.1	44.7 ± 1.6	15.7 ± 0.6	6.0 ± 0.4	4.3 ± 0.3	2.6 ± 0.2	0.5 ± 0.1	0.1 ± 0.0
Multijet	6.6 ± 1.5	6.1 ± 1.4	6.6 ± 1.5	2.7 ± 0.6	1.3 ± 0.3	1.1 ± 0.2	0.4 ± 0.1	0.2 ± 0.0	0.1 ± 0.0
HF Noise	0.8 ± 0.1	16.6 ± 2.1	28.2 ± 3.6	25.1 ± 3.2	9.3 ± 1.2	18.4 ± 2.3	18.2 ± 2.3	10.7 ± 1.4	7.4 ± 0.9
gg$H(\to$ inv.)	570.5	411.5	338.0	162.8	82.5	61.8	30.4	8.1	3.6
qq$H(\to$ inv.)	56.2	125.7	245.8	244.0	191.2	217.9	156.1	62.6	45.6
W$H(\to$ inv.)	29.7	17.0	11.5	3.9	1.9	0.8	0.5	0.1	0.0
qqZ$H(\to$ inv.)	14.0	6.8	4.0	1.2	0.7	0.4	0.4	0.0	0.0
ggZ$H(\to$ inv.)	14.0	8.6	5.5	2.3	1.0	0.6	0.3	0.1	0.0
tt$H(\to$ inv.)	3.6	2.6	1.8	0.7	0.3	0.2	0.1	0.0	0.0
Total bkg.	19062.6 ± 85.4	11455.8 ± 59.5	8220.0 ± 50.1	3411.5 ± 25.4	1539.4 ± 17.3	1051.6 ± 12.4	446.9 ± 8.3	122.0 ± 6.2	38.9 ± 2.3
Observed	18945	11500	8218	3419	1549	1068	447	104	41

Table 4.2: Expected event yields in each M_{jj} bin for the different background processes in the VBF signal region, with the 2018 samples. The background yields and the corresponding uncertainties are obtained after performing a combined fit across all of the CRs and SR. The expected signal contributions for a Higgs boson, produced in the non-VBF and VBF modes, decaying to invisible particles with a branching fraction of $\mathcal{B}(H \to \text{inv}) = 1$, and the observed event yields are also reported.

M_{jj}bin range (GeV)	200-400	400-600	600-900	900-1200	1200-1500	1500-2000	2000-2750	2750-3500	>3500
$Z(\nu\nu)$ + jets (strong)	14150.5 ± 61.3	8498.6 ± 45.5	5891.5 ± 34.4	2490.3 ± 18.1	1096.6 ± 12.2	730.4 ± 9.4	278.9 ± 5.8	55.9 ± 2.3	20.5 ± 1.2
$Z(\nu\nu)$ + jets (VBF)	228.8 ± 4.5	276.0 ± 5.1	348.3 ± 5.6	254.4 ± 4.9	167.7 ± 4.3	191.5 ± 4.8	123.3 ± 4.4	36.8 ± 2.1	24.3 ± 1.9
$W(\ell\nu)$ + jets (strong)	7323.4 ± 51.4	4566.3 ± 38.3	3243.7 ± 29.5	1357.2 ± 17.3	603.5 ± 11.1	418.9 ± 8.3	160.2 ± 5.8	38.3 ± 3.0	18.4 ± 2.2
$W(\ell\nu)$ + jets (VBF)	145.4 ± 7.7	163.6 ± 8.4	191.7 ± 9.7	133.5 ± 6.6	90.7 ± 4.7	98.1 ± 5.2	66.2 ± 3.7	31.9 ± 2.6	13.3 ± 1.4
$t\bar{t}$ + single-top	261.3 ± 13.9	234.8 ± 12.6	151.5 ± 8.7	55.2 ± 3.5	28.9 ± 1.9	29.0 ± 2.0	12.2 ± 1.2	4.4 ± 0.4	1.4 ± 0.2
Diboson	264.0 ± 31.4	172.2 ± 20.7	144.5 ± 17.5	51.0 ± 6.2	23.6 ± 3.0	18.2 ± 2.4	4.6 ± 0.6	0.4 ± 0.1	0.0 ± 0.0
$Z/\gamma^*(\ell^+\ell^-)$ + jets	105.7 ± 2.8	71.4 ± 2.0	57.3 ± 1.9	22.6 ± 0.8	10.1 ± 0.4	7.6 ± 0.5	2.3 ± 0.2	0.9 ± 0.1	0.2 ± 0.0
Multijet	4.4 ± 1.3	4.5 ± 1.3	3.8 ± 1.1	2.1 ± 0.6	1.0 ± 0.3	1.0 ± 0.3	0.5 ± 0.2	0.2 ± 0.0	0.1 ± 0.0
HF Noise	0.0 ± 0.0	18.5 ± 2.2	54.4 ± 6.4	45.2 ± 5.3	18.8 ± 2.2	38.0 ± 4.4	44.0 ± 5.1	19.8 ± 2.3	13.4 ± 1.6
ggH($\to$ inv.)	719.3	534.7	461.5	232.2	119.0	95.1	52.2	15.7	7.2
qqH($\to$ inv.)	74.3	171.3	340.3	327.8	269.2	321.6	271.1	115.3	72.4
WH($\to$ inv.)	39.6	22.9	12.6	5.2	1.9	1.1	0.3	0.2	0.1
qqZH($\to$ inv.)	18.4	8.5	3.9	1.9	0.4	0.6	0.2	0.0	0.0
ggZH($\to$ inv.)	17.2	11.2	7.6	3.1	1.2	1.0	0.4	0.1	0.0
ttH($\to$ inv.)	4.7	3.4	2.6	1.1	0.5	0.4	0.1	0.0	0.0
Total bkg.	22483.4 ± 87.5	14005.9 ± 65.1	10086.6 ± 51.1	4411.4 ± 27.8	2040.8 ± 18.1	1532.7 ± 15.4	692.3 ± 11.4	188.6 ± 5.6	91.7 ± 3.7
Observed	22505	14036	10220	4374	2080	1555	695	176	95

4.2 Combinations

This section describes the combination of the results presented in Sec. 4.1 with different analysis categories. Sec. 4.2.1 describes the combination with an orthogonal VBF analysis category called VBF-triggered region (VTR). Sec. 4.2.2 describes the combination with the VBF $H \rightarrow inv.$ analysis by CMS using 2016 dataset [100].

4.2.1 Combination with VTR category

As discussed in Sec. 3.4.1, an orthogonal set of selection categories (compared to selection categories listed in Sec. 3.4) are included in the analysis as well. Events in this category are collected using a different set of VBF triggers (instead of the $p_{T,no-\mu}^{miss}$ triggers described in Sec. 3.2.1), and target events at a lower p_T^{miss} range of $[160, 250]$ GeV, hence making the event category orthogonal to what has been discussed so far, with $p_T^{miss} > 250$ GeV. This category is called the VBF-triggered region (VTR). The VTR category comprises of the signal region, Z and W control regions, but it does not have the $\gamma + jets$ control region due to the lower recoil range it targets (note that $p_T^{\gamma} > 200$ GeV is required for the photon triggers in use, as explained in Sec. 3.2.1).

When combining the VTR category with this analysis to do a combined fit to data, likelihood terms from each region in the VTR category are added to the likelihood function shown in Eq. 3.15. The details of VTR event selection, results and the combination are discussed in [4].

4.2.2 Combination with 2016 dataset

The results from this analysis are also statistically combined with the results from the CMS experiment with 2016 dataset [100]. Data from 2016 analysis is considered as different event categories, similar to the way which is done when combining 2017 and 2018 datasets.

During the combination, theoretical uncertainties are treated as correlated between different years, while most experimental uncertainties are treated as uncorrelated between 2016 and 2017+2018 categories. The integrated luminosity has been updated for the 2016 dataset to 36.3 fb^{-1} to reflect the latest improvements [101]. In addition, to be consistent with the treatment of the VBF $H \to inv.$ signal with 2017 and 2018 analyses, the Higgs boson p_T dependent NLO EWK corrections (described in Sec. 3.2.7) are also applied to the 2016 signal shape.

Fig. 4·4 shows the fit results in the VBF signal region when data from all three years (2016-2018) are combined. Total background estimated from the fit to the data (as described in Sec. 3.5.1) is shown (S+B fit), together with a background estimate from a fit assuming $\mathcal{B}(H \to \text{inv}) = 0$ (B-only fit) are shown. In the S+B fit, the best-fit signal strength is computed to be $\mathcal{B}(H \to \text{inv}) = 0.086^{+0.054}_{-0.052}$. Expected event yields from different Higgs production modes are also plotted in Fig. 4·4, each scaled to the best-fit signal strength value of $\mathcal{B}(H \to \text{inv}) = 0.086$.

The impact of each uncertainty source to the total uncertainty on $\mathcal{B}(H \to \text{inv})$ is shown in Tab. 4.3. It can be observed that the largest contributions to the uncertainty on determining $\mathcal{B}(H \to \text{inv})$ come from the theory uncertainties on V+jets transfer factors (e.g. uncertainties on μ_R), together with the statistical uncertainties on the collected data and simulated events.

From Fig. 4·4, it can be observed that no significant excess of data is observed compared to the Standard Model background. Therefore, the results from this analysis are interpreted as an upper bound on $\mu = (\sigma_H/\sigma_H^{\text{SM}})\mathcal{B}(H \to \text{inv})$, the methodology and resulting upper bounds are described in the next section.

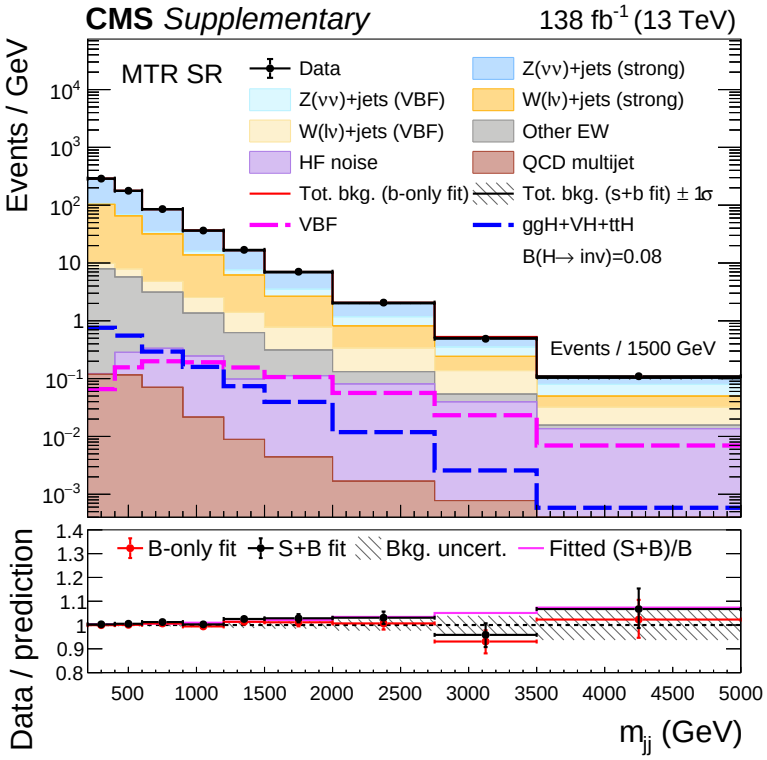

Figure 4·4: The observed M_{jj} distribution in the VBF signal region compared to the postfit backgrounds, with the 2016, 2017, and 2018 datasets. The signal processes are scaled by the fitted value of $\mathcal{B}(H \to \text{inv})$, shown in the legend. The background contributions are estimated from the fit to the data described in Sec. 3.5.1 (S+B fit). Total background estimated from a fit assuming $\mathcal{B}(H \to \text{inv}) = 0$ (B-only fit) is also shown. The last bin of each distribution integrates events above the bin threshold divided by the bin width. Figure taken from [4].

Table 4.3: Uncertainty breakdown in $\mathcal{B}(H \to \text{inv})$. The sources of uncertainty are separated into different groups. Observed and expected results are quoted for the full combination of 2016, 2017, and 2018 data. The expected results are obtained using an Asimov dataset [2] with $\mathcal{B}(H \to \text{inv}) = 0$.

Group of systematic uncertainties	Observed impact on $\mathcal{B}(H \to \text{inv})$	Expected impact on $\mathcal{B}(H \to \text{inv})$
Theory	$^{+0.026}_{-0.025}$	± 0.024
MC event count	$^{+0.024}_{-0.023}$	$^{+0.023}_{-0.024}$
Triggers	$^{+0.021}_{-0.022}$	± 0.021
Leptons/photons/b	$^{+0.012}_{-0.011}$	$^{+0.010}_{-0.011}$
QCD multijet mismodelling	± 0.013	± 0.014
Jet calibration	$^{+0.010}_{-0.007}$	± 0.007
Int. luminosity/pileup	± 0.005	$^{+0.004}_{-0.005}$
Other systematic uncertainties	$^{+0.013}_{-0.010}$	± 0.010
Stat.	± 0.029	± 0.030

4.3 Interpretations

4.3.1 Invisible branching fraction of SM Higgs boson

Since no statistically significant excess is observed in data compared to estimated SM background, the results are interpreted as upper bounds to $\mu = (\sigma_H/\sigma_H^{\mathrm{SM}})\mathcal{B}(H \to \mathrm{inv})$. Assuming a Standard Model (SM) Higgs boson, μ can be interpreted as $\mathcal{B}(H \to \mathrm{inv})$. Observed and expected 95% CL upper limits are computed using an asymptotic approximation of the CL_s method, which is detailed in Refs. [102, 103].

Observed and expected upper limits on $\mu = (\sigma_H/\sigma_H^{\mathrm{SM}})\mathcal{B}(H \to \mathrm{inv})$ at 95% CL are presented in Fig. 4·5. A more detailed breakdown of upper limits coming from different years and categories is also provided in Tab. 3.11.

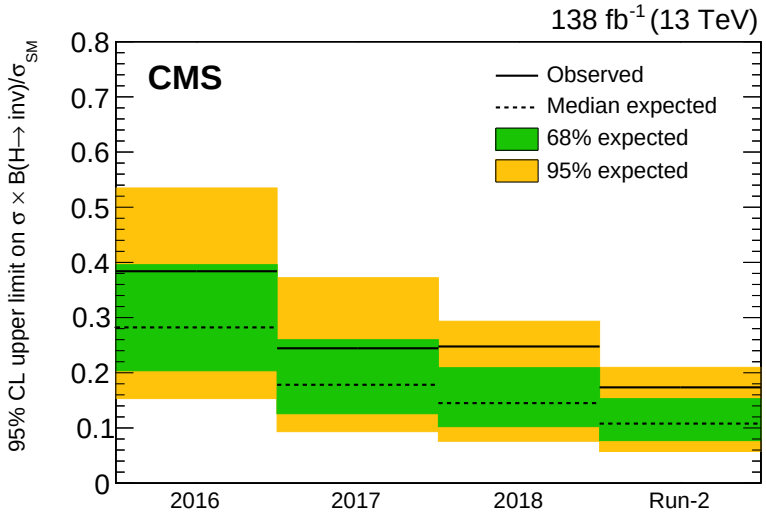

Figure 4·5: Observed and expected 95% CL upper limits on $(\sigma_H/\sigma_H^{\mathrm{SM}})\mathcal{B}(H \to \mathrm{inv})$ for all three data-taking years, together with the 1σ (green) and 2σ (yellow) uncertainty bands on the expected upper limits. The combination of 2016-2018 is also shown. These results assume a SM Higgs boson with a mass of 125.38 GeV. Figure taken from [4].

Table 4.4: The 95% CL upper limits on $(\sigma_H/\sigma_H^{\mathrm{SM}})\mathcal{B}(H \to \mathrm{inv})$, assuming an SM Higgs boson with a mass of 125.38 GeV. The observed and median expected results are shown, along with the 68% and 95% interquartile ranges for each category and for the combinations.

Category	Observed	Median expected	65% expected	95% expected
2012-2016	0.33	0.21	[0.15,0.29]	[0.11,0.39]
VTR 2017	0.57	0.45	[0.32,0.66]	[0.24,0.94]
VTR 2018	0.44	0.34	[0.24,0.49]	[0.18,0.69]
VTR 2017 2018	0.40	0.28	[0.20,0.40]	[0.15,0.56]
MTR 2017	0.25	0.19	[0.14,0.28]	[0.10,0.40]
MTR 2018	0.24	0.15	[0.11,0.22]	[0.08,0.31]
MTR 2017 2018	0.17	0.13	[0.09,0.18]	[0.07,0.25]
all 2017	0.24	0.18	[0.13,0.26]	[0.09,0.37]
all 2018	0.25	0.15	[0.10,0.21]	[0.08,0.29]
all 2017 2018	0.18	0.12	[0.08,0.17]	[0.06,0.23]
Run2	0.18	0.10	[0.07,0.14]	[0.05,0.20]

4.3.2 Upper bound on DM-nucleon interactions

The upper limit on $\mathcal{B}(H \to \mathrm{inv})$, obtained from combining the data taken between 2012 and 2018, is interpreted in the context of Higgs-portal models of DM interactions, in which a stable DM particle couples to the SM Higgs boson. The interaction between a DM particle and an atomic nucleus may be mediated by the exchange of a Higgs boson, producing nuclear recoil signatures, such as those investigated by direct-detection experiments.

If the mass of the DM particle, m_{DM}, is smaller than half of the mass of the Higgs boson, the Higgs boson invisible width (Γ_{inv}) can be translated, within an effective field theory approach, into a spin-independent DM-nucleon elastic scattering cross section, as outlined in Ref. [61]. This translation is performed assuming that the DM candidate is either a scalar or a Majorana fermion, and both the central value

and the uncertainty of the dimensionless nuclear form-factor f_N are taken from the recommendations of Ref. [104]. The conversion from $\mathcal{B}(H \to \text{inv})$ to Γ_{inv} uses the relation $\mathcal{B}(H \to \text{inv}) = \Gamma_{\text{inv}}/(\Gamma_{\text{SM}} + \Gamma_{\text{inv}})$, where Γ_{SM} is set to 4.07 MeV [105]. The assumption of a vector DM candidate is not provided in the context of this analysis, since it requires an extended dark Higgs sector, which may lead to modifications of kinematic distributions assumed for the invisible Higgs boson signal in this analysis.

Fig. 4·6 shows the 90% CL upper limits on the spin-independent DM-nucleon scattering cross section as a function of m_{DM}, for both the scalar and the fermion DM scenarios. These limits are computed at 90% CL so that they can be compared with those from direct detection experiments such as Xenon1T [11], Cresst-II [12], CDMSlite [13], LUX [14], Panda-X 4T [15], and DarkSide-50 [16], which provide the strongest constraints in the m_{DM} range probed by this search. The collider-based results complement the direct-detection experiments in the range m_{DM} smaller than 12 (6) GeV, assuming a fermion (scalar) DM candidate.

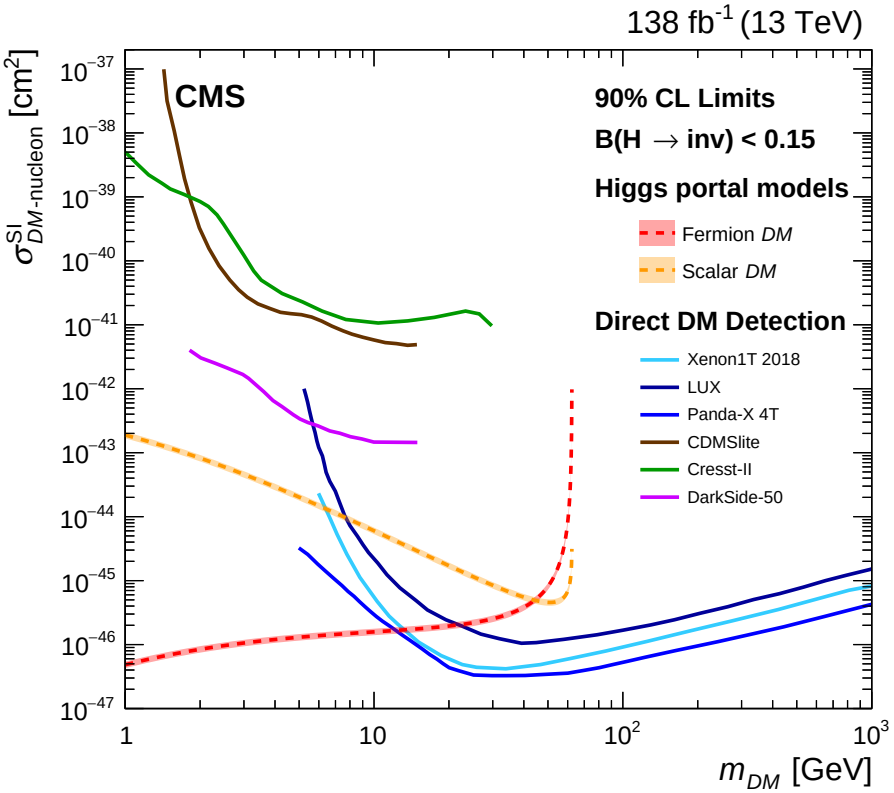

Figure 4·6: The 90% CL upper limits on the spin-independent DM-nucleon scattering cross section in Higgs-portal models, assuming a scalar (dashed orange) or fermion (dashed red) DM candidate. Limits are computed as functions of m_{DM} and are compared to those from the Xenon1T [11], Cresst-II [12], CDMSlite [13], LUX [14], Panda-X 4T [15], and DarkSide-50 [16] experiments, which are shown as solid lines. Figure taken from [4].

Chapter 5

Outlook

5.1 Improvements in the high-level trigger system

As described in Sec. 2.2.6, high-level trigger (HLT) in CMS is the last step of event selection before accepted events are saved for storage and offline analysis. The event selection at this step is done by reconstructing physics objects within the event, such as electrons, muons, jets and missing transverse momentum (each defined in Sec. 3.1). Several identification criteria are applied on these reconstructed physics objects to select events of potential interest. The reconstruction of each type of particle is described in [70].

The event selection logic at HLT is structured around the concept of a *HLT path*. Each HLT path is a sequence of processes running in a pre-defined order, which reconstructs physics objects and applies selections on them. Typically every HLT path is defined to reconstruct a certain type of physics object (e.g. jets), and apply selections on that object to decide whether the event should be kept or not. For the purposes of the VBF $H \rightarrow inv.$ analysis described in earlier sections of this thesis, HLT paths that select events based on p_T^{miss} are very important, since the $H \rightarrow inv.$ decays are expected to produce p_T^{miss} in the final state. A discussion of how p_T^{miss} based triggers are used to collect data for the analysis is already given in Sec. 3.2.1. It should be noted that such p_T^{miss}-based triggers are not only important for analyses targeting $H \rightarrow inv.$ signal, but they are commonly used for many analyses that search for Beyond the Standard Model (BSM) physics. Therefore, improvements in the event

selection of such HLT paths can play a crucial role for many analyses within the CMS experiment.

5.1.1 Revisiting HF noise mitigation

Extending the discussion in Sec. 3.3.1 to HLT paths which select events based on p_T^{miss}, it was observed that a large number of events with mismeasured HF jets are getting accepted by such HLT paths. As described in Sec. 3.3.1, this is due to the creation of large p_T^{miss} when the energy of one final state jet in QCD multijet production events is mismeasured. This impacts a number of things:

- Events with fake p_T^{miss} are accepted and saved to offline storage. Such events would then require offline noise-cleaning treatment to be used in an analysis.
- The output event rate of p_T^{miss} triggers are increased, due to events with mismeasured HF jets being accepted together with events with true p_T^{miss}.

Therefore, discrimination between a mismeasured HF jet and a well-identified jet at these HLT paths can play a crucial role to reject events that are not of interest. This will also help decreasing the output rate coming from these paths, freeing up rates that could be used by other HLT paths in CMS. Expanding on the ideas introduced in Sec. 3.3.1, the jet shower shape variables are studied again at HLT reconstruction, and a jet-based filter is developed to increase the rate of rejection of events with mismeasured HF jets. The studies for the filter are outlined below, and the performance checks with these new p_T^{miss} based paths using Run3 data are shown in Sec. 5.1.2.

The studies conducted to develop the filter are identical to the studies described in Sec. 3.3.1. The same jet shower shape variables are used to distinguish mismeasured jets from well-identified jets: $\sigma_{i\eta i\eta}$, $\sigma_{i\phi i\phi}$ and CSS_{HF}. Jets are measured in "physics-enriched" and "noise-enriched" regions, where "noise-enriched" region refers to the

events with a high-p_T jet recoiling against high p_T^{miss}, and "physics-enriched" region refers to the events with a high-p_T jet recoiling against a well-identified photon. One difference compared to the study described in Sec. 3.3.1 is that this study is restricted to jets which are reconstructed where the noise is observed to be peaking, $2.99 < |\eta| < 3.25$. Since the studies were done before the start of Run3 data taking (2022), data taken in 2018 is used. The jet shower shape variables in these two regions are shown in Fig. 5·1.

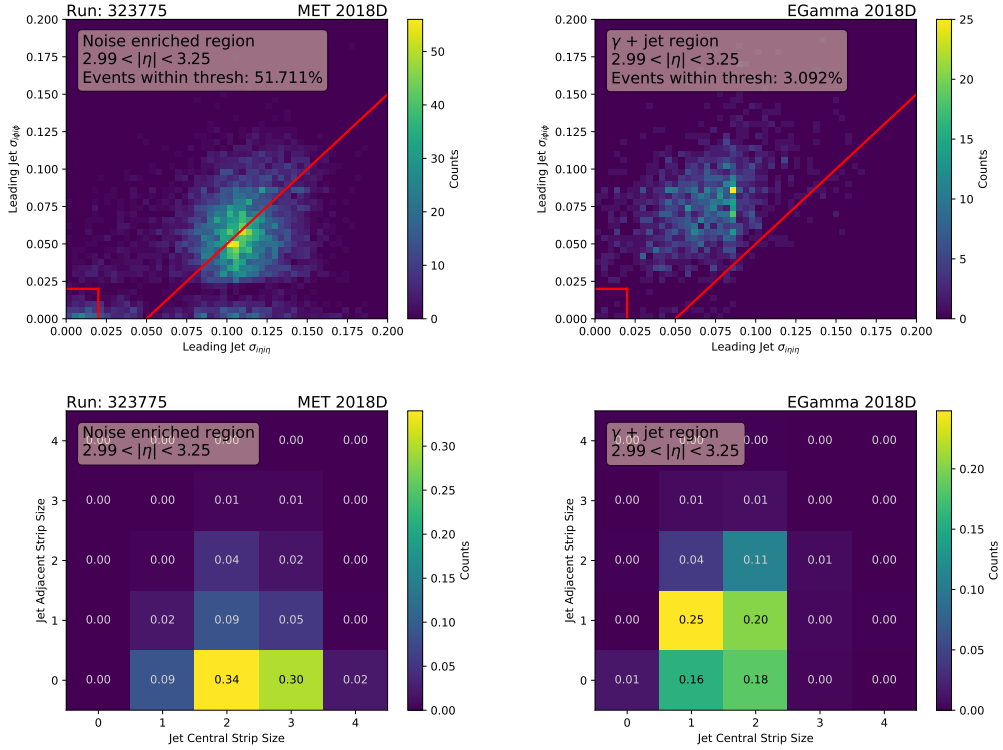

Figure 5·1: The HF jet shower shape variables for noise-enriched (left) and physics-enriched regions (right). The top row shows the two-dimensional distributions of η and ϕ width, $\sigma_{i\eta i\eta}$ and $\sigma_{i\phi i\phi}$. The red lines on these figures correspond to the filter applied to reject the jets, and the legend on top left quotes the percentage of jets that fail the requirements. The bottom row shows the distributions of central and adjacent strip sizes. Higher values of central strip size (i.e. $= 3$) are observed for mismeasured jets, and hence the this variable is required to be < 3 as a part of the filter.

From Fig. 5·1, similar trends are observed compared to Figs. 3·22, 3·23, 3·24 and 3·25. Therefore, the following cuts were introduced to the HLT paths to identify a jet:

- Jet should have $\sigma_{i\eta i\eta} - \sigma_{i\phi i\phi} < 0.05$
- Jet should not lie in the region where $\sigma_{i\eta i\eta} < 0.02$ & $\sigma_{i\phi i\phi} < 0.02$
- Central strip size of the jet, $CSS_{HF} < 3$

From Fig. 5·1, it can be observed that applying these cuts to the jets within $2.99 < |\eta| < 3.25$ can reject $\sim 50\%$ of mismeasured jets, and the impact on well-reconstructed events (here, γ + jet) is on the order of 1%.

For the Run3 data taking at CMS experiment, a new set of HLT paths were included that utilize this jet filtering technique. The implementation is done on HLT paths which cut on p_T^{miss} and $H_{\mathrm{T}}^{\mathrm{miss}}$, where $H_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the imbalance in total jet momenta, as shown in Eq. 5.1.

$$H_{\mathrm{T}}^{\mathrm{miss}} = \left| \sum_{jet} \vec{p_T}^{jet} \right| \tag{5.1}$$

where the sum goes over jets with $p_T > 20$ GeV. The HF cuts are included in the HLT path such that if a jet fails the requirements, it will not be considered in the $H_{\mathrm{T}}^{\mathrm{miss}}$ calculation. This way, $H_{\mathrm{T}}^{\mathrm{miss}}$ will be corrected from the impact of jets with mismeasured p_T values, and the probability of such events making it through the $H_{\mathrm{T}}^{\mathrm{miss}}$ filter is lower.

5.1.2 Performance checking with Run3 data

Using the most recent proton-proton collision data taken during Run3, the performance of the p_T^{miss} and $H_{\mathrm{T}}^{\mathrm{miss}}$ based paths with and without this filter is monitored. Ideally, one would like to have the following performance constraints:

1. For events with well identified physics objects, the loss of events caused by adding the filter should be minimal.

2. For events with mismeasured jets, the rejection rate should be much larger, which would then translate to a reduction in rate.

These performance considerations are explicitly checked using Run3 data. Two HLT paths are considered for this study, both paths require $p_{T,no-\mu}^{miss} > 120$ GeV and $H_{T,no-\mu}^{miss} > 120$ GeV to decide whether an event should be kept or not. The only difference between those two paths is the presence of the HF-filter derived in Sec. 5.1.1. These HLT paths are summarized in Tab. 5.1.

Table 5.1: HLT paths used in this study. The left column shows the full name of the path and the right column shows whether the path has the HF-jet filter included as described in Sec. 5.1.1. Both paths require $p_{T,no-\mu}^{miss} > 120$ GeV and $H_{T,no-\mu}^{miss} > 120$ GeV to make a decision on whether to keep an event.

HLT path name	Has HF jet-filter
HLT_PFMETNoMu120_PFMHTNoMu120_IDTight	No
HLT_PFMETNoMu120_PFMHTNoMu120_IDTight_FilterHF	Yes

Impact on well-reconstructed events

To check the loss of events due to the newly added HF-jet filter, efficiencies of the two paths in Tab. 5.1 are compared on $\mathrm{W}(\mu\nu) + \text{jets}$ events. The events to consider in the efficiency measurement are selected by the following criteria:

- Leading jet $p_T > 30$ GeV, leading jet must be tightly identified, according to the definition in Sec. 3.1.1.

- Events must pass a muon HLT path *HLT_IsoMu27*, which requires a muon to be reconstructed with $p_T > 27$ GeV at HLT.

- Events must have a well-identified muon (reconstructed offline) with $p_T > 30$ GeV, which is tightly identified, as defined in Sec. 3.1.3.

The efficiency of the $p_{T,no-\mu}^{miss}$-based paths are measured as a function of recoil, which corresponds to p_T of the W boson (see Sec. 3.1.2). The efficiencies are shown in Fig. 5·2, where the left-hand side plot shows the efficiency of the HLT path without the HF-jet filter, and the right-hand side plot shows the efficiency of the other path with the filter. To model the turn-on behavior of the efficiencies, a sigmoid function is fit to both curves, which is defined as follows:

$$f(x) = \frac{1}{1 + e^{-(x-\mu)/\sigma}} \tag{5.2}$$

where μ and σ are determined from the best-fit to the data. The best-fit μ and σ parameters are also shown in the legend for each plot. It can be observed that for the two paths, the efficiencies are found to be almost identical, hence supporting the fact that the loss of well-reconstructed events from the HF-filter is very minimal.

Impact on output rate

To check the impact of the HF-jet filter on the output rate, events that are accepted by the two paths in Tab. 5.1 are examined. Fig. 5·3 shows the number of events passing the two paths, as a function of η of their leading jet. Note that events with leading jet $p_T > 30$ GeV are considered. The blue curve corresponds to events passing the HLT path without the filter, and the orange curve corresponds to the path with the filter. Finally, the ratio pad at the bottom gives the ratio between the orange and blue curves. It can observed that, close to an additional 50% of events near $3 < |\eta| < 3.25$ are rejected with the HF-filter. Note that the fraction of event rejection in this phase space roughly agree with what was estimated from 2018 data in Fig. 5·1. This reduction of events is found to be corresonding to about 10% overall

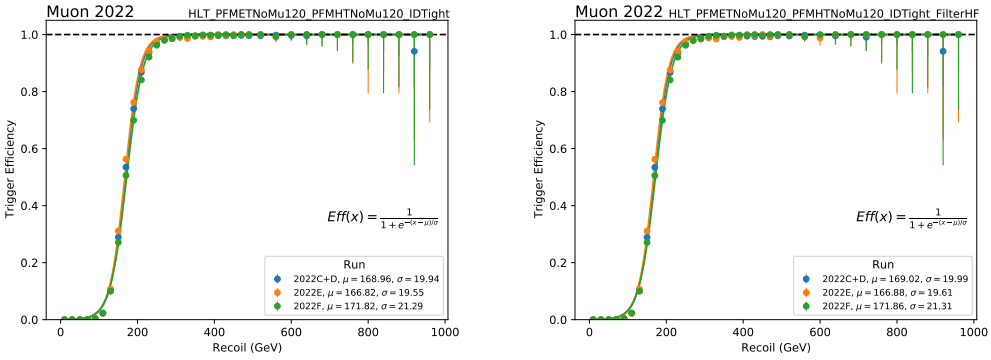

Figure 5·2: Efficiency of the $p_{T,no-\mu}^{miss}$ based HLT paths without (left) and with (right) the HF-filter. To each efficiency curve, a sigmoid function is fit to describe the turn-on behavior. The best-fit parameters μ and σ are shown in the legend, located on the bottom right of each plot. Different curves represent different data taking periods within 2022. It can be observed that the two efficiencies are almost identical, hence supporting the fact that the loss of well-reconstructed events from the HF-filter is very minimal.

reduction in rate of the HLT paths where the filter is introduced. It should also be noted that outside the region where the leading jet is within $3 < |\eta| < 3.25$, the impact due to the HF-jet filter is either 0 or very small. The small impact in this case comes from events where another jet with high p_T exists in the forward region and fails the HF cuts, resulting in the event failing the trigger.

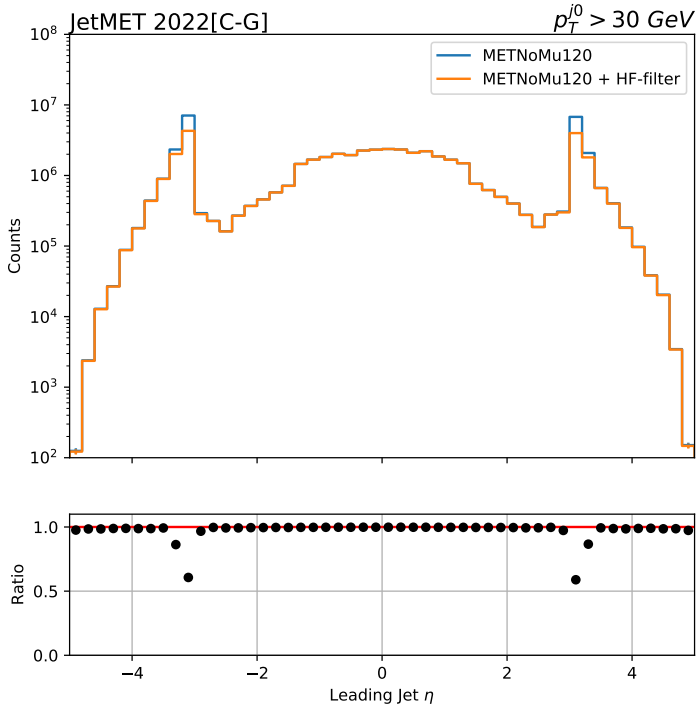

Figure 5·3: Number of events passing the $p_{T,no-\mu}^{miss}$-based path with (orange) and without (blue) the HF-filter, as a function of η of the highest-p_T jet in the event. Note that events with leading jet $p_T > 30$ GeV are considered. The ratio pad at the bottom shows the ratio between orange and blue curves, hence effectively showing the additional rejection of events due to the new HF-jet filter.

5.2 Upgrade of the CMS tracker system

The Large Hadron Collider (LHC) will be upgraded to the High-Luminosity LHC (HL-LHC) configuration. Currently, this upgrade is planned to take place in Long Shutdown 3 era between 2026 and 2029, and HL-LHC is expected to start running starting from 2029. With HL-LHC, the instantaneous luminosity is expected to reach an unprecedented peak of 7.5×10^{34} $\mathrm{cm}^{-2}\mathrm{s}^{-1}$, with an average number of pileup interactions up to 200. To cope with the increased instantaneous luminosity, the CMS detector will be upgraded, which is known as the Phase-2 upgrade. One crucial upgrade will be in the tracker detector, which is the closest detector to the proton-proton collision point. Hence, to cope with the demanding operating conditions, the CMS tracker detector will be replaced. The aim of the new tracker detector is to provide robust tracking under increased instantaneous luminosity, and also provide inputs to the Level-1 (L1) trigger [17], which is also a crucial physics goal since having tracking information at L1 trigger will improve the data selection performance.

The Phase-2 tracker of CMS will be composed of two parts: Inner tracker (IT) and outer tracker (OT), as shown in Fig. 5·4. IT is made of silicon pixel detectors, and OT is made of silicon micro-strips and macro-pixel detectors. Both IT and OT are designed to have better radiation hardness and higher granularity compared to the current tracker detector. Furthermore, the tracking acceptance will be extended in the forward region, with the IT covering a range up to $|\eta| < 4$, as can be seen in Fig. 5·4.

The IT will cover a total area of 4.9 m^2 with 3892 silicon modules. There will be two types of silicon modules deployed: Double-chip (1x2) modules and quad-chip (2x2) modules. The arrangement of modules in the detector is shown in Fig. 5·4, where double-chip modules are denoted with green, and quad-chip modules are denoted with orange. A detailed overview of IT and OT features and the upgrade planning is given

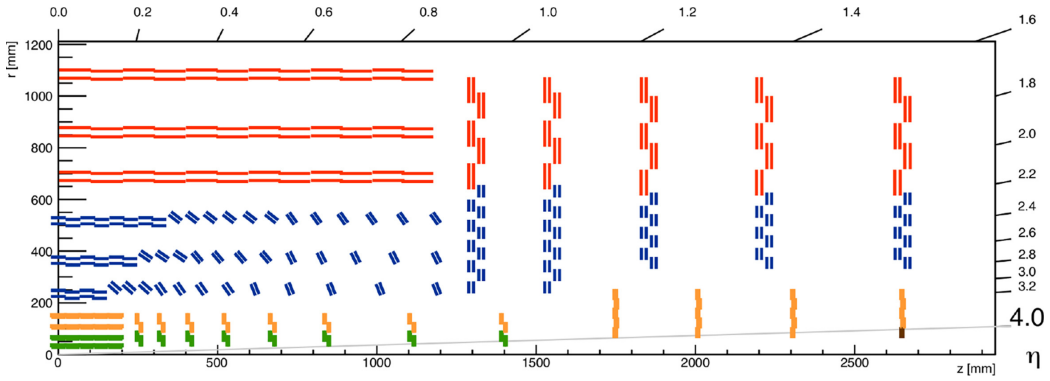

Figure 5·4: Layout of the CMS Phase2 tracker detector, each colored line indicates a detector module. Pixel modules, in orange (quad-chip modules) and green (double-chip modules), form the inner tracker system. The outer tracker is composed of two different types of modules indicated with blue (PS modules) and red (2S modules) lines. One quarter of the detector is shown. Figure taken from [17].

in [17].

The OT consists of dual-sensor modules, sometimes also called "p_T-modules", which are two closely spaced silicon sensors read out by the same electronics. A charged particle traversing the module will leave a signal in both sensors, and readout electronics can compare the two hit positions. Once the hit position in the first sensor is identified, the hit position in the second sensor will depend on the track curvature in the magnetic field, which is then related to the p_T of the charged particle. If the hit in the second sensor is within the correlation window relative to the hit in the first sensor, a track segment (called stub) is generated and the tracks will be reconstructed. It should be noted that this procedure is equivalent to the requirement of a minimum p_T of the charged particle, since the particles with lower p_T will experience more bending, and are more likely to fall out of the correlation window. The p_T threshold is set by the spacing between the two sensors in the module, and can be tuned in order to have a threshold of between 2 and 3 GeV [17]. A schematic of such dual-sensors is shown in Fig. 5·5.

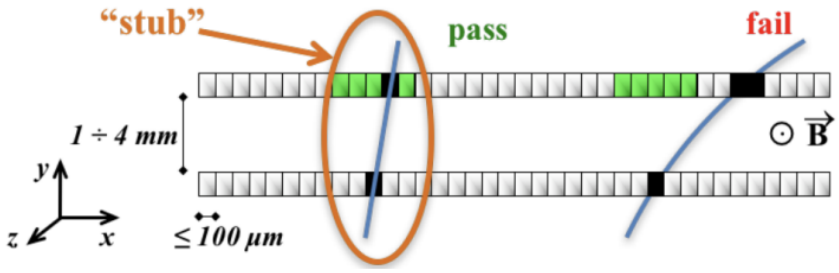

Figure 5·5: Schematic showing the functioning of dual-sensor OT modules. The high-p_T track on the left experiences smaller bending due to the magnetic field and falls into the green correlation window, thus producing a stub. The low-p_T track on the right however, experiences larger bending and falls out of the correlation window, and does not generate a stub. Figure taken from [18].

All silicon modules, both in the IT and OT, communicate with CMS back-end Data Acquision System (DAQ) optically through Low Power GigaBit Transciever (lpGBT) links, which run at speeds of either 5 or 10 Gbps in uplink (i.e. from front-end electronics to DAQ), and 2.56 Gbps in downlink [18]. In the service cavern, sets of Data Trigger and Control (DTC) cards are placed, which are located at the opposite end of the lpGBT optical links coming from the IT and OT modules. The DTCs are high bandwidth processors based on commercial FPGAs. Their task is to receive data from the IT and OT modules, which represent the hits pertaining to the triggered events, and forward them to the DAQ system. The schematic of this workflow is shown in Fig. 5·6.

In the case of OT, the DTC boards are also tasked with forwarding hit information in the detector to the Track Finder Processors (TFP). When track segments (stubs) are generated from the dual-sensors in the OT, DTCs are tasked with forwarding the stub coordinates to TFPs, after converting the coordinates to the global reference frame [18].

Two prototypes of DTC boards are produced: The Apollo board targeting the

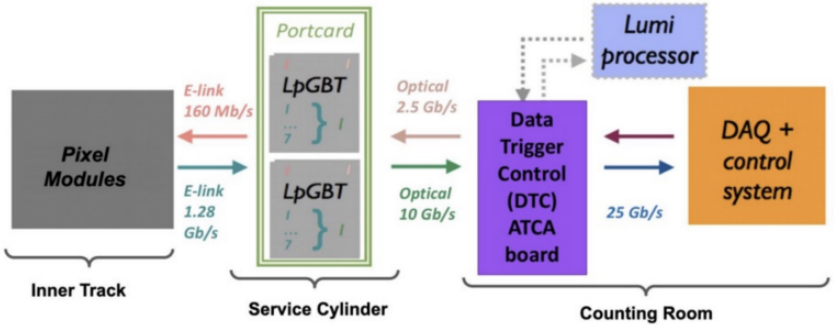

Figure 5·6: Schematic representation of the data acquisition system (DAQ) for the CMS Phase2 tracker. Figure is taken from [17].

role of DTC for the IT [19], and the Serenity board targeting the role of DTC for the OT [106]. In order to meet the design criteria with high processing and bandwidth requirements, both DTCs use high performance Xilinx Ultrascale+ Virtex FPGAs, and 12-channel, 25 Gbps-per-channel optical transcievers to implement their functions. [18]. Both DTC boards are designed to comply with the Advanced Telecommunications Computing Architecture (ATCA) design specifications [107]. The following subsection will go into more detail about the Apollo board architecture and the hardware components.

5.2.1 Apollo DTC Board: The Design

The Apollo DTC board is designed to provide a reusable hardware interface for different types of applications within the CMS (and ATLAS) detector. It is composed of two main modules: A Service Module (SM) and a Command Module (CM). The Apollo SM is an ATCA-compliant front board which provides power, communications and control infrastructure. On the other hand, Apollo CM is a module with two Xilinx Ultrascale+ Virtex FPGAs, which can be programmed with application specific firmware, hence making the Apollo hardware reusable across different applications.

The CM is connected to the SM using 2 board-to-board connectors, which provide

electrical and mechanical connectivity between the SM and CM. Apollo SM provides multiple interfaces to communicate with the CM, providing I2C, UART, JTAG and AXI chip-to-chip links to the CM. A block diagram of Apollo SM and CM is shown in Fig. 5·7 (left), together with a schematic of the Apollo SM (right).

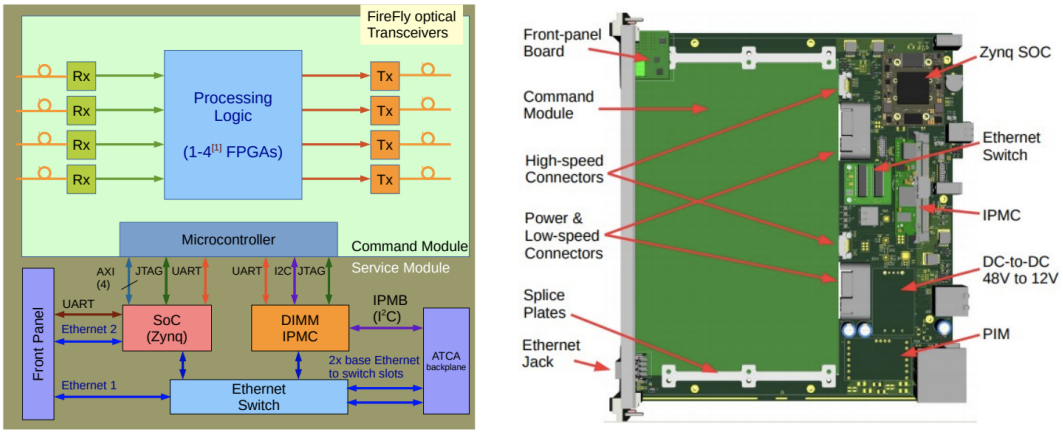

Figure 5·7: Block diagram of Apollo SM and CM (left) and a schematic of the Apollo SM (right). The SM provides multiple connection interfaces to the CM (I2C, AXI chip-to-chip, JTAG, UART), and the CM includes FPGAs with application specific firmware. Each SM is connected to a front-end panel as specified by the ATCA standards. Schematics are taken from [19].

Apollo Service Module

The Apollo SM hosts a Xilinx Zynq Ultrascale+ system-on-chip (SoC), which provides a 64-bit ARM processor core, together with an FPGA and memory interfaces. A userspace I/O (UIO) driver is used to map the register addresses from different devices on the SM and CM, to the memory space of the Zynq SoC. This way, also using the AXI chip-to-chip interface to the CM, monitoring data from registers on the CM can be read out, including current and temperature values in the CM. The SoC also provides a JTAG interface to the CM, which makes it possible to program the FPGAs located in the CM through the SoC on the SM.

Another important component of the SM is the Intelligent Platform Management Controller (IPMC), also specified within the ATCA design standards. The essential functions of the IPMC can be summarized as follows:

- Manage the power up and power down sequences of the Apollo board.
- Provide communication with the ATCA shelf manager by reading a set of monitoring values (temperature, current and voltage) from on-board sensors and deliver them to the shelf manager.

For the Apollo SM, the OpenIPMC hardware is used [108]. OpenIPMC is a Cortex-M7 ARM-core based microcontroller, which provides flash memory and multiple I/O peripheral interfaces to communicate with the other devices on Apollo SM and CM. OpenIPMC firmware is written in C language, and it is based on the FreeRTOS real-time operating system, allowing the microcontroller to process multiple tasks in parallel using the FreeRTOS scheduler. Most of these tasks are a part of the core-IPMC functionality, including:

- Communication with the shelf manager through the backplane I2C bus.
- Monitoring the state of the front-panel handle, and triggering state transitions accordingly. As an example, if the front-panel handles are released, the OpenIPMC will trigger a shut down of the Apollo SM.
- Provide a Telnet command-line interface for remote users to access OpenIPMC functionality.

A detailed explanation of the core tasks handled by the OpenIPMC firmware is given in [108].

The OpenIPMC firmware also provides custom software tasks to be developed based on the board-specific needs. This way, the power-up and power-down sequences

and which sensors to read out from can be customized for an Apollo board. This flexibility of the OpenIPMC allows it to be used in other flavors of DTC boards such as Serenity [106].

In the Apollo boards, the OpenIPMC firmware is customized to read out sensor values, such as temperature, current and voltage, from the SM and CM. This is made possible by the use of multiple I2C buses within the Apollo board, which interfaces with the multiple I2C peripherals of the OpenIPMC microcontroller. OpenIPMC then reports these values to the ATCA shelf manager using the backplane I2C bus for each board. It also can report sensor values to the Zynq SoC on the SM, communicating through another I2C bus connecting the OpenIPMC to the Zynq SoC. All of these features allow for continuous monitoring of the entire shelf. As an example, thanks to the interface between the shelf manager and OpenIPMC, shelf manager can adjust fan speeds based on the temperature values being read out. Sensor values forwarded to the SoC via I2C are also forwarded to third-party monitoring tools like Grafana, which allows for easy monitoring of each board in the shelf.

5.2.2 Apollo DTC Board: The Applications

The Apollo DTC board is intended to be used for different applications in Phase-2 CMS and ATLAS detectors. For each application, the Apollo CM are to be customized accordingly. Each application and the anticipated hardware needs are briefly presented in this subsection, following [19].

One of the applications of Apollo DTC is the CMS track finder algorithm. This algorithm uses pattern-matching to identify coincidences between "tracklets" transmitted from the readout modules and then uses a Kalman filter to establish precise track parameters [19]. This application requires substantial FPGA resources, and is expected to require two Virtex Ultrascale+ FPGAs and 60 optical links running at 25 Gbps. A total of between 126 and 180 track finder boards are anticipated to be

required for CMS.

Another application of Apollo DTCs within CMS is the pixel data acquisition (DAQ) and timing. The DTCs used here are expected to receive data from up to 512 front-end links, which transmit data in a compressed format. The received compressed data needs to be decoded in real time to build events. For this application, it is currently foreseen that two XCVU7P or similar Virtex Ultrascale+ class FPGAs will be required, together with 72 optical links at 10 Gbps and 16 optical links at 25 Gbps. A total of 28 pixel DTCs are required in CMS for this application [19].

Apollo DTC boards also have applications in the ATLAS experiment. They are being planned to use in the ATLAS Monitored Drift-Tube Trigger Processor (MDTTP), which performs a similar function with the DTC and track finder combined for CMS, but for the muon-drift tubes in ATLAS. Data from drift-tube hits are received on about 60 fiber optic links and a sophisticated twodimensional fit is used to identify track segments. These segments are joined to form tracks, and the Zynq processor is used to calculate transverse momentum for the identified tracks. In addition, drift-tube hits are buffered and stored until a trigger is received after which they are built into an event and sent to the DAQ. A total of 64 MDTTP boards are required for ATLAS [19].

5.2.3 Online monitoring software development

In parallel with the development of ATCA boards for the Phase-2 CMS tracker, an online software system is being developed which allows an easy-to-use user interface (UI) to access large number of Apollo and Serenity boards deployed in the detectors. The online software system is designed to allow users to perform a wide range of operations on the boards over the local area network (LAN), including:

- Access the list of available boards and their IP addresses over the LAN.

- From each available board, read monitoring data such as temperature and current values.
- Execute commands for each board over the LAN, such as powering the CM up and programming an FPGA hosted on the CM.

The online software stack is mainly composed of two parts. First part is the SHEP UI, which is a front-end web-application running on a separate computer in the LAN. SHEP provides an easy-to-use UI for users to register new boards to the system, read monitoring data and access board resources. SHEP software registers each board into a SQL-database for storage, and provides communication with every registered Apollo board using a back-end software running on the Zynq SoC in the Apollo SM. This back-end software is called HERD, which sets up an API server on the Zynq SoC. It is tasked with listening for messages coming from the SHEP front-end UI, and handling the message and returning a response back to the SHEP UI. This communication between SHEP UI and HERD API server allows easy remote access to the Apollo boards. A schematic of this workflow is shown in Fig. 5·8.

To make the deployment of the online software stack more standardized, SHEP and HERD software are currently designed to be ran as Docker containers that are communicating over the LAN. This way, every dependency of each piece of software is installed within the Docker container, and the dependence on the underlying SoC platform is minimized. Continuous integration (CI) pipelines in GitLab are developed such that Docker images for each software in the stack can be automatically built within the CI jobs, and be stored in the GitLab container registry.

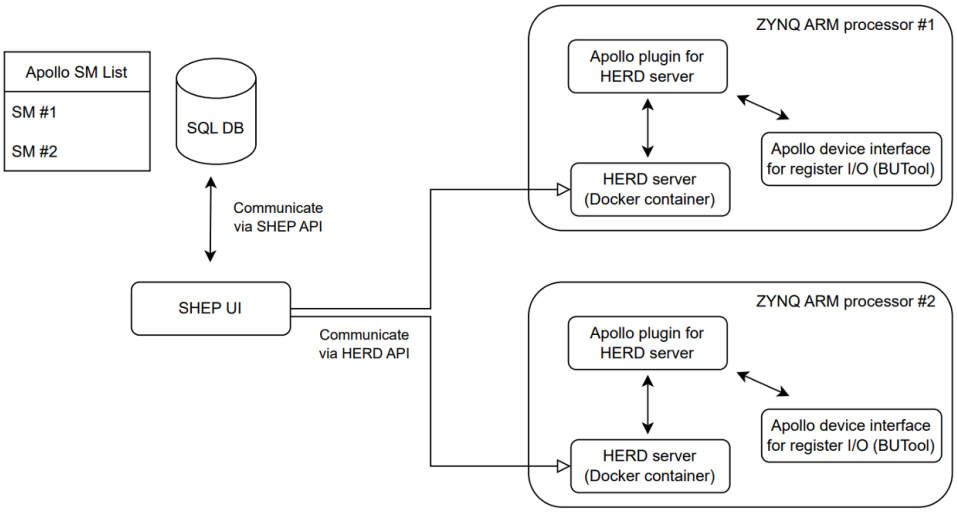

Figure 5·8: The schematic showing the communication between the SHEP UI and HERD API server, running on the Zynq SoC of each Apollo board. SHEP UI stores a list of boards in a SQL database, and communicates with each registered Apollo board by connecting to the HERD API server. The HERD application running on the SoC, in turn interfaces with Apollo-specific software to fulfill it's tasks.

5.3 Using new techniques for VBF H(inv) analysis

In addition to the detector-related upgrades mentioned in the previous sections, search for novel techniques to improve the sensitivity of the VBF $H \to inv.$ analysis are also underway. In the analysis methodology described in this thesis, kinematic variables of two outgoing VBF jets were used to distinguish VBF $H \to inv.$ signal from most of the SM backgrounds. An interesting line of development related to the goal of increasing the sensitivity of the analysis is the use of new machine learning (ML) techniques, in an attempt to develop classifiers to more efficiently classify signal from SM background. One important example of such a classifier is called ParticleNet [20].

ParticleNet is a graph neural network which takes lower-level point-like objects as input, for example the set of particles reconstructed within the event. It treats all particles as a "point cloud" data structure, which is a permutationally invariant set of points each carrying a feature vector, composed of features such as p_T, η, ϕ and so on. ParticleNet then applies the edge convolution (EdgeConv) operation [109] on the point-cloud data structure. This operation can be understood as a convolution-like operation for point clouds, where the point cloud data structure is treated as a graph, where each point is a vertex, and connections between points represent edges. For each vertex, k nearest neighbors can be identified. The EdgeConv operation for each point x_i then has the form

$$\mathbf{x'_i} = \frac{1}{k} \sum_{j=1}^{k} h_\Theta(\mathbf{x_i}, \mathbf{x_{j_i}}) \tag{5.3}$$

where $\mathbf{x_i}$ denotes the feature vector of the point x_i and $i_1, ..., i_k$ are the indices of the k-nearest neighbors of x_i. The edge function h_Θ is some function that is parametrized by a set of learnable parameters, Θ. It should also be noted that this EdgeConv operation is stackable, just as a regular convolution operation. This is

because EdgeConv is essentially a mapping between an input point cloud to another point cloud with the same number of points with updated feature vectors. This allows to build a deep neural network architecture using stacked EdgeConv operations, which can learn features of point clouds hierarchically [20]. This feature is exploited in the ParticleNet model architecture, which is shown in Fig. 5·9.

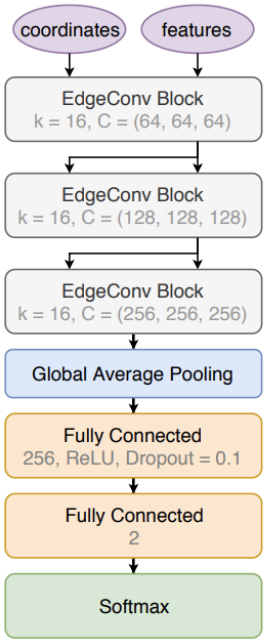

Figure 5·9: Architecture of the ParticleNet model. Three EdgeConv operations are stacked together with the number of nearest neighbors is taken to be $k = 16$. Afterwards, a global average pooling is applied to aggregate the learned features over all particles in the cloud, which is followed with a densely connected network. A softmax function is used to compute the output for the binary classification task. Figure is taken from [20].

5.3.1 Applying ParticleNet to develop a VBF-tagger

ParticleNet architecture can be used to develop a classifier with the goal of distinguishing VBF $H \to inv.$ signal events from other Higgs production mechanism modes, such as gluon-gluon fusion (gg$H$). To classify a given event as such, the set of par-

ticles reconstructed in the event can be fed into the ParticleNet model, with the list of $[\eta, \phi]$ coordinates for each particle, and their feature vectors specifying their energy, p_T, electrical charge and so on. The aim is to achieve this classification using the hadronic activity in the events (i.e. the two energetic VBF jets, and potentially softer jets coming from NLO effects), therefore any reconstructed lepton or photon particle within the event is discarded from the set of input particles. This has profound importance when the tagger is applied to SM processes such as $\mathrm{W}(\ell\nu) + \text{jets}$ to predict whether it is kinematically "gluon-fusion like" or "VBF-like".

Such a ParticleNet-based VBF classifier, trained on events with gluon-fusion and VBF Higgs production, is tested by applying the model to classify simulated events within the VBF signal region phase space, as described in Sec. 3.4. A distribution of score values for different types of Higgs production events is shown in Fig. 5·10. It can be observed that VBF $H \to inv.$ events are accumulated at higher score values, while all the other Higgs production events have a flat distribution as a function of the ParticleNet score.

To quantify the performance of the ParticleNet-based classification, and compare it with predictions made by different M_{jj} cuts, the receiver-operator characteristic (ROC) curve can be calculated for both scenarios and those can be compared. These ROC curves are shown in Fig. 5·11. Here, the blue curve corresponds to the ParticleNet-based VBF classifier, and the orange curve corresponds to a purely M_{jj}-based classifier, which corresponds to the true positive and false positive event fractions when different M_{jj} cuts are applied. It can be observed from the area under the ROC curves (AUC), the ParticleNet-based classifier performs better than a purely M_{jj}-based classifier.

It is also interesting and instructive to study the event features which are correlated with the ParticleNet prediction. Fig. 5·12, shows kinematic distributions of the

two final-state jets for the gg$H$ events, where the events are categorized into two: Events correctly classified as being "gluon-fusion like" (hence, DNN score < 0.5), and events classified as "VBF-like". The first plot in Fig. 5·12 shows the $\Delta\eta_{jj}$ between the two final state jets, and the second plot shows the η of the second (trailing in p_T) jet. It can be observed that the presence of a forward jet in the event (with higher η) increases the probability of that event being classified as "VBF-like". From the $\Delta\eta_{jj}$ plot on Fig. 5·12, it can be observed that almost all gg$H$ events with $\Delta\eta_{jj} > 5$ are classified as "VBF-like". This is however not an unexpected feature, because of the event topology of VBF $H \to inv.$ events with forward final-state jets, which is learned by the model during the training phase. This however, makes it more challenging to classify SM backgrounds that originate from VBF processes, such as the electroweak production of $\mathrm{Z}(\nu\nu) + \mathrm{jets}$ or $\mathrm{W}(\ell\nu) + \mathrm{jets}$, where forward jets also appear in the final state.

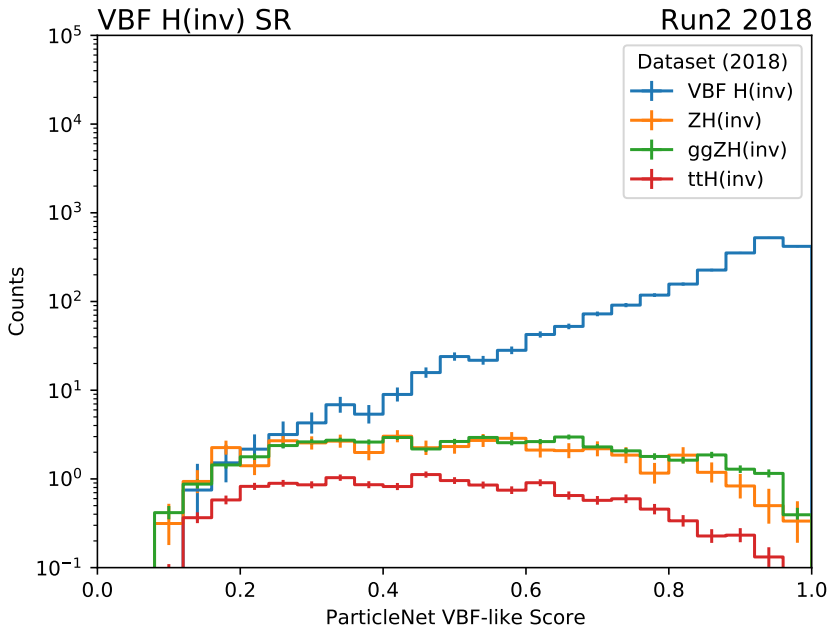

Figure 5·10: VBF-tagger score distribution for different types of Higgs production events that pass the VBF signal region selection. It can be observed that the VBF $H \to inv.$ shape is accumulated at higher score values, while all the other Higgs production events have a flat distribution.

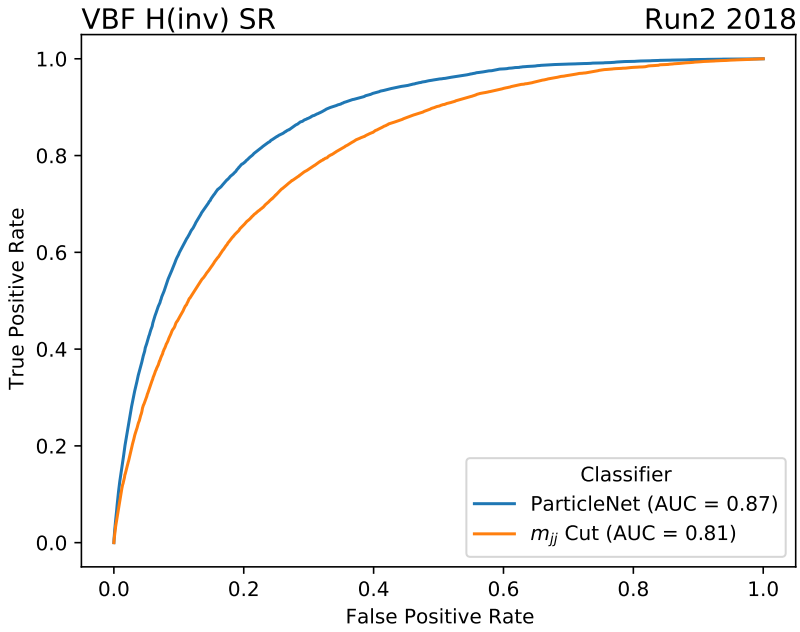

Figure 5·11: Receiver-operator characteristic (ROC) curve for the classification of gg$H$ and VBF $H \to inv.$ events that pass the VBF signal region selection. The blue curve shows the ROC curve for the ParticleNet-based classifier, while the orange curve shows the ROC curve for the case of different M_{jj} cuts being applied to label events. It can be observed that the ParticleNet classifier performs better than purely M_{jj}- based event discirmination. Area under the ROC curve (AUC) is also provided in the legend for the two ROC curves.

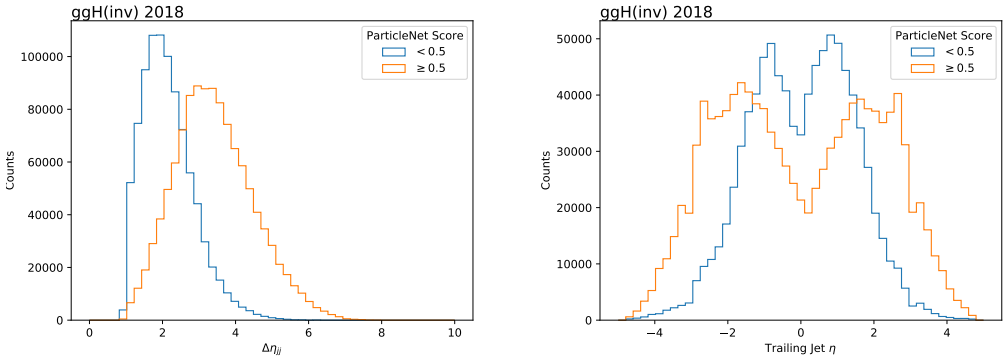

Figure 5·12: $\Delta\eta_{jj}$ between the two final state jets (left) and η of the trailing jet (right) for gg$H$ simulation events. The events are categorized into two, where the first group (blue) is classified as "gluon-fusion-like" and the second group (orange) is classified as "VBF-like". It can be observed that the presence of a forward jet is correlated with the score output of the VBF classifier, making it more probable that the event will be classified as VBF-like.

5.3.2 Preliminary results

At the time of writing of this thesis, the application of this ParticleNet based VBF-tagger to the VBF $H \to inv.$ analysis is preliminary, and final results with the fit to full Run2 dataset has not yet been obtained. However, some preliminary results will be shown in this section in an attempt to display the work done so far, and provide a basis of discussion for the potential future work.

One aim of the studies done so far is to apply the VBF-tagger to classify events in proton-proton collision data, and see if the tagger can provide similar performance of event classification in data and simulation. For this purpose, 2018 dataset is used, and the regular analysis selection (described in Sec. 3.4) is applied. To test the agreement between data and simulation as a function of the VBF-tagger score, data in control regions are used. For the purposes of this study, minor backgrounds such as top quark and diboson production are ignored. Fig. 5·13 shows the data-to-simulation agreement as a function of the VBF-tagger score for the Z and W control regions.

From Fig. 5·13, a few interesting observations can be made:

- Total expected (and observed) event yields do not have a significant shape as a function of VBF-tagger score. This is expected because the event yields are dominated by the strong production of V+jets, due to their larger cross section compared to the electroweak V+jets production modes.

- The electroweak V+jets contributions have an increasing shape with increasing VBF-tagger scores, which is also expected due to this process originating from VBF production of the corresponding vector boson.

- There is a trend in data to simulation ratio where the disagreement reaches to $\approx 20\%$. While this is not clearly understood, it is important to note that this trend appears consistently across different control regions. This allows

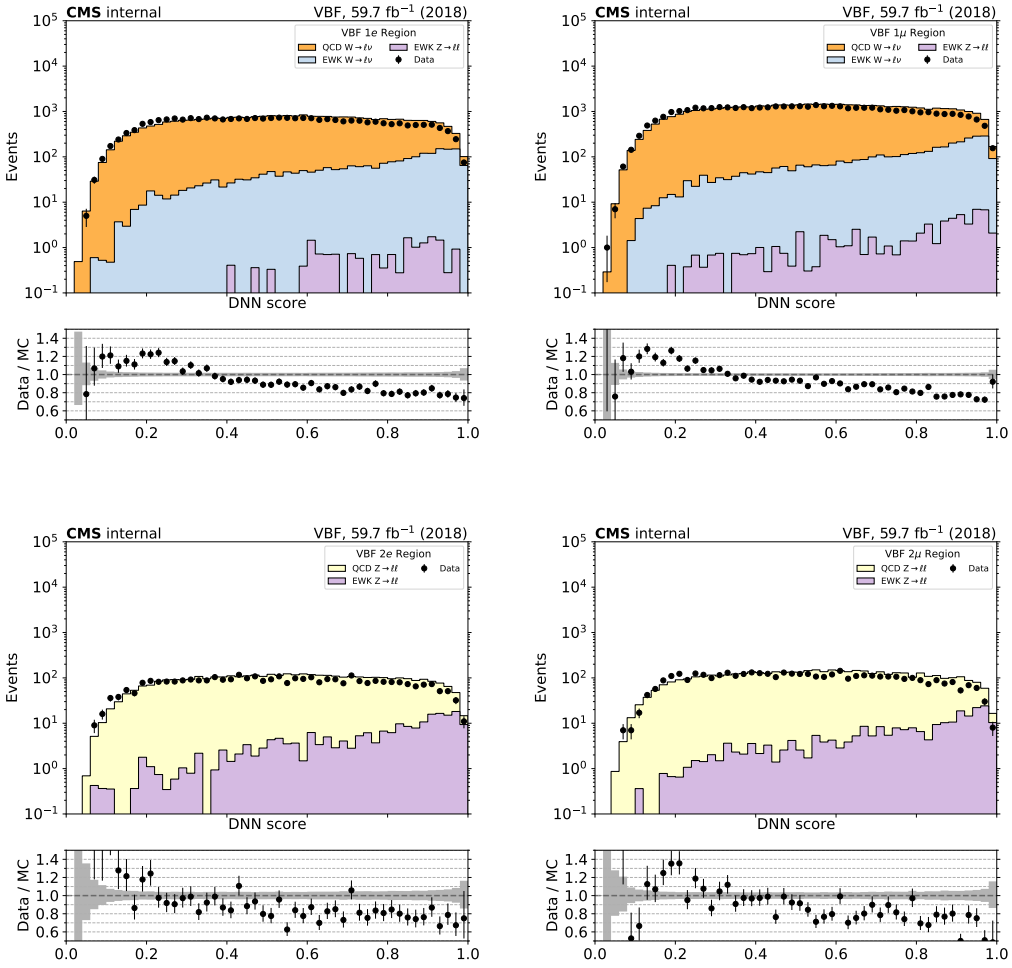

Figure 5·13: Single electron (top left), single muon (top right), double electron (bottom left) and double muon (bottom right) control regions as a function of the VBF-tagger score. The black markers in the top pad represent the observed data in each bin, and the stacks represent expected contributions from each process, obtained from simulation. The ratio pad at the bottom shows the ratio between the observed data and expected yields from simulation for each bin, and the grey band accounts for the statistical uncertainties in the ratio.

this effect to approximately cancel out when the ratio of processes (transfer factors) are taken into account. It also supports the fact that the VBF-tagger predictions are not correlated with the reconstructed lepton, and only based on the hadronic activity within the event.

To observe the partial cancellation mentioned in the last observation above, it is instructive to look at ratios of V+jets processes (i.e. transfer factors) in different control regions. These transfer factors as a function of VBF tagger score are shown in Fig. 5·14. The left-hand side plot shows the $\mathrm{Z}(\ell\ell)+\mathrm{jets}$ to $\mathrm{W}(\ell\nu)+\mathrm{jets}$ ratio, where electron and muon channels are combined. It can be observed that there is reasonable agreement between the ratios observed in data and predicted by simulation, within the statistical and systematical uncertainties. The right-hand side plot shows the $\mathrm{Z}(\ell\ell)+$jets to $\gamma+$ jets ratio, where some left-over disagreement in ratios are observed. At the time of writing, this effect is not fully understood, but it should be noted that the simulation samples for $\mathrm{Z}(\ell\ell)+$ jets and $\mathrm{W}(\ell\nu)+$ jets are simulated at NLO in QCD, while the γ + jets are simulated in LO, with NLO corrections applied as a function of $p_{\mathrm{T},V}$ and M_{jj}, as explained in Sec. 3.2.7. Considering that all identified lepton and photon particles are removed from the input collection to the VBF tagger, it is plausible that this difference can be playing a role in the shape of the predicted ratio in simulation.

5.3.3 Outlook

Introducing powerful classifiers such as ParticleNet to enhance the sensitivity of the VBF $H \to inv.$ analysis is definitely an interesting prospect. As mentioned in the previous sections, it is understood that ParticleNet can perform better than a pure-M_{jj} based classifier (i.e. Fig. 5·11). However, the question of whether this classifier can improve the statistical bounds on the $\mathcal{B}(H \to \mathrm{inv})$, hence enhancing the sensitivity

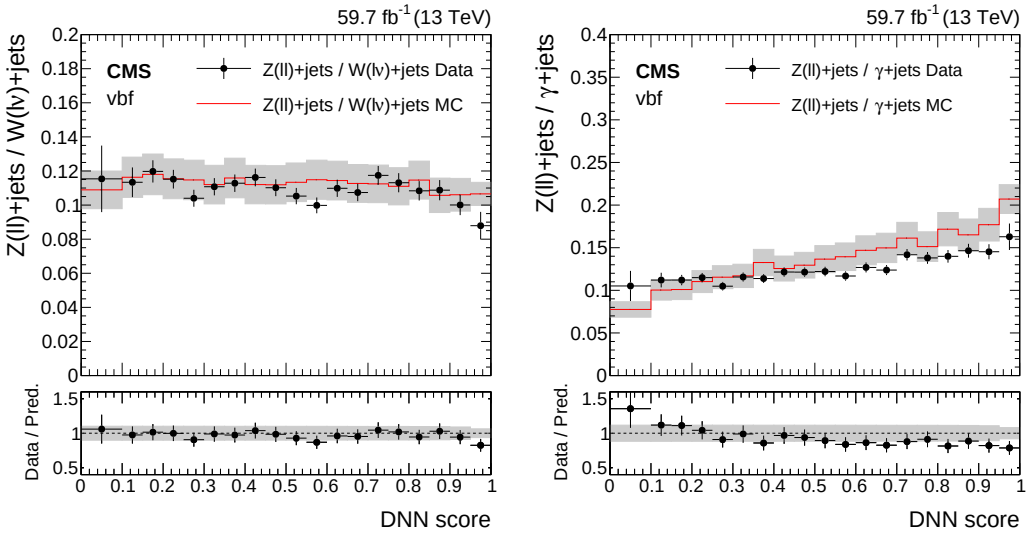

Figure 5·14: Ratios of data and total simulation yields in $\mathrm{Z}(\ell\ell)+\mathrm{jets}$ to $\mathrm{W}(\ell\nu)+\mathrm{jets}$ ratio (left) and $\mathrm{Z}(\ell\ell)+\mathrm{jets}$ to $\gamma+\mathrm{jets}$ ratio (right), as a function of the VBF-tagger score. The black markers show the ratio in data observed in control regions, while the red line shows the ratio in simulated V+jets processes. The gray band represents the statistical and systematical uncertainties on the ratio.

of the analysis, is still open.

Application of the ParticleNet classifier to this analysis can be grouped into two different ways:

- Perform the fit on the VBF-like score distribution, where the likelihood is described by Eq. 3.15.
- Perform a cut on the VBF-like score to reduce background (i.e. score $> X$), and perform the fit on M_{jj} as before.

In both cases, the systematic uncertainties described in Sec. 3.5.2 have to be recomputed. In the first case, they need to be recomputed as a function of the ParticleNet score, while in the second case, they have to be recomputed for the high VBF-like score phase space. Through further reduction of background using such a sophisticated classifier can in turn enhance the signal to background ratio, resulting

in an overall increase in the sensitivity of the analysis. With the Run3 data taking ongoing with Large Hadron Collider (LHC) and the high luminosity stage (HL-LHC) on the horizon, much larger volumes of data will be able to analyze to probe $H \to inv.$ decays, and the application of sophisticated classifiers such as ParticleNet remain an interesting variation of the analysis in an attempt to enhance the sensitivity of the analysis.

Bibliography

[1] CMS JME POG. https://twiki.cern.ch/twiki/bin/view/CMS/MissingETOptionalFiltersRun2.

[2] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011.

[3] Gianfranco Bertone, Dan Hooper, and Joseph Silk. Particle dark matter: Evidence, candidates and constraints. *Phys. Rept.*, 405:279–390, 2005.

[4] CMS Collaboration. Search for invisible decays of the Higgs boson produced via vector boson fusion in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. D*, 105:092007, May 2022.

[5] Spyros Argyropoulos, Oleg Brandt, and Ulrich Haisch. Collider Searches for Dark Matter through the Higgs Lens. *Symmetry*, 13(12):2406, 2021.

[6] CERN. Cern accelerator complex. https://home.cern/science/accelerators/accelerator-complex.

[7] CMS Collaboration. Cms detector design. https://cms.cern/news/cms-detector-design.

[8] CMS Collaboration. The CMS phase-1 pixel detector upgrade. *Journal of Instrumentation*, 16(P02027), 2021.

[9] CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(S08004), 2008.

[10] Mark Thomson. *Modern particle physics*. Cambridge University Press, New York, 2013.

[11] E. Aprile et al. Dark matter search results from a one ton-year exposure of XENON1T. *Phys. Rev. Lett.*, 121:111302, 2018.

[12] G. Angloher et al. Results on light dark matter particles with a low-threshold CRESST-II detector. *Eur. Phys. J. C*, 76:25, 2016.

[13] R. Agnese et al. New results from the search for low-mass weakly interacting massive particles with the CDMS low ionization threshold experiment. *Phys. Rev. Lett.*, 116:071301, 2016.

[14] D. S. Akerib et al. Results from a search for dark matter in the complete LUX exposure. *Phys. Rev. Lett.*, 118:021303, 2017.

[15] Yue Meng et al. Dark Matter Search Results from the PandaX-4T Commissioning Run. 7 2021.

[16] P. Agnes et al. Low-mass dark matter search with the DarkSide-50 Experiment. *Phys. Rev. Lett.*, 121:081307, 2018.

[17] CMS Collaboration. The cms tracker for the high luminosity lhc. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1048, 2023.

[18] Luigi C. CMS Collaboration. Status of the Phase-2 Tracker Upgrade of the CMS experiment at the HL-LHC. *J. Phys.: Conf. Ser.*, (1690), 2020.

[19] A. Albert et. al. The Apollo ATCA Platform. *arXiv:1911.06452*, 2020.

[20] Loukas Gouskos Huilin Qu. ParticleNet: Jet Tagging via Particle Clouds. *arXiv:1902.08570*, 2020.

[21] J. J. Thomson. Cathode rays. *Phil. Mag. Ser. 5*, 44:293–316, 1897.

[22] E. Rutherford. The scattering of alpha and beta particles by matter and the structure of the atom. *Phil. Mag. Ser. 6*, 21:669–688, 1911.

[23] E. Rutherford. Collision of α particles with light atoms. IV. An anomalous effect in nitrogen. *Phil. Mag. Ser. 6*, 37:581–587, 1919.

[24] J. Chadwick. Possible Existence of a Neutron. *Nature*, 129:312, 1932.

[25] Albert Einstein. Concerning an heuristic point of view toward the emission and transformation of light. *Annalen Phys.*, 17:132–148, 1905.

[26] Louis Victor Pierre Raymond de Broglie. Recherches sur la théorie des quanta. *Annals Phys.*, 2:22–128, 1925.

[27] E. Schrödinger. Quantisierung als Eigenwertproblem. *Annalen Phys.*, 384(4):361–376, 1926.

[28] M. Born, W. Heisenberg, and P. Jordan. Zur Quantenmechanik. II. *Z. Phys.*, 35(8-9):557–615, 1926.

[29] Albert Einstein. On the electrodynamics of moving bodies. *Annalen Phys.*, 17:891–921, 1905.

[30] Paul A. M. Dirac. The quantum theory of the electron. *Proc. Roy. Soc. Lond. A*, 117:610–624, 1928.

[31] H. Weyl. Electron and Gravitation. 1. (In German). *Z. Phys.*, 56:330–352, 1929.

[32] P. A. M. Dirac. A Theory of Electrons and Protons. *Proc. Roy. Soc. Lond. A*, 126(801):360–365, 1930.

[33] Carl D. Anderson. The Apparent Existence of Easily Deflectable Positives. *Science*, 76:238–239, 1932.

[34] E. M. Riordan. The Discovery of quarks. *Science*, 256:1287–1293, 1992.

[35] Murray Gell-Mann. A Schematic Model of Baryons and Mesons. *Phys. Lett.*, 8:214–215, 1964.

[36] J. J. Aubert et al. Experimental Observation of a Heavy Particle J. *Phys. Rev. Lett.*, 33:1404–1406, 1974.

[37] S. W. Herb et al. Observation of a Dimuon Resonance at 9.5-GeV in 400-GeV Proton-Nucleus Collisions. *Phys. Rev. Lett.*, 39:252–255, 1977.

[38] S. Abachi et al. Observation of the top quark. *Phys. Rev. Lett.*, 74:2632–2637, 1995.

[39] CMS Collaboration. Measurement of the weak mixing angle using the forward-backward asymmetry of Drell-Yan events in pp collisions at 8 TeV. *arXiv:1806.00863*, 2018.

[40] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental Test of Parity Conservation in β Decay. *Phys. Rev.*, 105:1413–1414, 1957.

[41] UA1 Collaboration. Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV. *Physics Letters B*, 122, 1983.

[42] Improved W boson Mass Measurement using 7 TeV Proton-Proton Collisions with the ATLAS Detector. 2023.

[43] W-like measurement of the Z boson mass using dimuon events collected in pp collisions at $\sqrt{s} = 7$ TeV. 2016.

[44] A. Nisati and G. Tonelli. The discovery of the Higgs boson at the Large Hadron Collider. *Riv. Nuovo Cim.*, 38(11):507–573, 2015.

[45] Armen Tumasyan et al. A portrait of the Higgs boson by the CMS experiment ten years after the discovery. *Nature*, 607(7917):60–68, 2022.

[46] F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helv. Phys. Acta*, 6:110–127, 1933.

[47] Yannick Mellier. Probing the universe with weak lensing. *Ann. Rev. Astron. Astrophys.*, 37:127–189, 1999.

[48] D. Narasimha and S. M. Chitre. Giant Luminous Arcs in Galaxy Clusters. *The Astrophysical Journal*, 332:75, September 1988.

[49] Wayne Hu, Naoshi Sugiyama, and Joseph Silk. The Physics of microwave background anisotropies. *Nature*, 386:37–43, 1997.

[50] Stefano Profumo. *An Introduction to Particle Dark Matter*. World Scientific, 2017.

[51] Gerard Jungman, Marc Kamionkowski, and Kim Griest. Supersymmetric dark matter. *Phys. Rept.*, 267:195–373, 1996.

[52] Keith R. Dienes, Emilian Dudas, and Tony Gherghetta. Extra space-time dimensions and unification. *Phys. Lett. B*, 436:55–65, 1998.

[53] N. Arkani-Hamed, A. G. Cohen, E. Katz, and A. E. Nelson. The Littlest Higgs. *JHEP*, 07:034, 2002.

[54] E. Aprile et al. First Dark Matter Search Results from the XENON1T Experiment. *Phys. Rev. Lett.*, 119(18):181301, 2017.

[55] G. Angloher et al. Results from 730 kg days of the CRESST-II Dark Matter Search. *Eur. Phys. J. C*, 72:1971, 2012.

[56] R. Agnese et al. Low-mass dark matter search with CDMSlite. *Phys. Rev. D*, 97(2):022002, 2018.

[57] D. S. Akerib et al. Improved Limits on Scattering of Weakly Interacting Massive Particles from Reanalysis of 2013 LUX Data. *Phys. Rev. Lett.*, 116(16):161301, 2016.

[58] Andi Tan et al. Dark Matter Results from First 98.7 Days of Data from the PandaX-II Experiment. *Phys. Rev. Lett.*, 117(12):121303, 2016.

[59] M. Ackermann et al. Constraining Dark Matter Models from a Combined Analysis of Milky Way Satellites with the Fermi Large Area Telescope. *Phys. Rev. Lett.*, 107:241302, 2011.

[60] Nadège Iovine. Dark matter searches in the centre of the Milky Way with IceCube. *PoS*, ICHEP2022:311, 2022.

[61] Abdelhak Djouadi, Oleg Lebedev, Yann Mambrini, and Jeremie Quevillon. Implications of LHC searches for Higgs–portal dark matter. *Phys. Lett. B*, 709:65–69, 2012.

[62] Abdelhak Djouadi. The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model. *Phys. Rept.*, 457:1–216, 2008.

[63] Armen Tumasyan et al. Search for new particles in events with energetic jets and large missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 11:153, 2021.

[64] A search for decays of the Higgs boson to invisible particles in events with a top-antitop quark pair or a vector boson in proton-proton collisions at $\sqrt{s} = 13$ TeV. 3 2023.

[65] Albert M Sirunyan et al. Search for invisible decays of a Higgs boson produced through vector boson fusion in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 793:520–551, 2019.

[66] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13\ TeV$. `https://cds.cern.ch/record/2676164`, 2019.

[67] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13\ TeV$. `https://cds.cern.ch/record/2621960`, 2018.

[68] CMS Collaboration. Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS. *European Physical Journal C*, 81(800 (2021)), 2021.

[69] CMS Collaboration. Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13\ TeV$. *Journal of Instrumentation*, 15(P10017), 2020.

[70] CMS Collaboration. The CMS trigger system. *Journal of Instrumentation*, 12(P01020), 2017.

[71] CMS Collaboration. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *Journal of Instrumentation*, 16(P05014), 2021.

[72] CMS Collaboration. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13\ TeV$. *Journal of Instrumentation*, 13(P06015), 2018.

[73] CMS Collaboration. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *Journal of Instrumentation*, 9(P10009), 2014.

[74] CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12(P10003), 2017.

[75] CMS Collaboration. Technical proposal for the phase-II upgrade of the CMS detector. http://cds.cern.ch/record/2020886, 2015.

[76] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[77] Vardan Khachatryan et al. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *JINST*, 12:P02014, 2017.

[78] CMS JetMET POG. https://twiki.cern.ch/twiki/bin/view/CMS/JECDataMC.

[79] CMS JME POG. Jet identification for the 13 TeV UL data. https://twiki.cern.ch/twiki/bin/view/CMS/JetID13TeVUL.

[80] CMS JME POG. Jet identification in high pile-up environment. https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJetID.

[81] A. M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13(05):P05011, 2018.

[82] CMS Muon POG. https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2#Loose_Muon.

[83] CMS Muon POG. https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2#Tight_Muon.

[84] CMS Collaboration. Performance of the DeepTau algorithm for the discrimination of taus against jets, electron, and muons. CMS Physics Analysis Summary, CMS-DP-2019-033, 2019.

[85] CMS Tau POG. https://twiki.cern.ch/twiki/bin/view/CMS/TauIDRecommendation13TeV#Measurement_in_Z_tautau_events.

[86] Vardan Khachatryan et al. Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *JINST*, 10:P08010, 2015.

[87] CMS EGamma POG. https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedPhotonIdentificationRun2#SPRING15_selections_bunch_crossi.

[88] CMS Lumi POG. https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData#Pileup_JSON_Files_For_Run_II.

[89] CMS Muon POG. https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceSelectionAndCalibrationsRun2.

[90] CMS JME POG. Reweighting recipe to emulate Level 1 ECAL prefiring. https://twiki.cern.ch/twiki/bin/viewauth/CMS/L1ECALPrefiringWeightRecipe.

[91] Prefiring correction implementation in NanoAOD-Tools. https://github.com/cms-nanoAOD/nanoAOD-tools/blob/master/python/postprocessing/modules/common/PrefireCorr.py.

[92] J. M. Lindert et al. Precise predictions for V+jets dark matter backgrounds. 2017.

[93] Search for VBF Higgs boson decays to invisible particles with 2016 dataset, 2017. CMS Analysis Note : AN-17-267.

[94] A Monte Carlo generator for the production of Higgs bosons attached to weak bosons (HAWK).

[95] Albert M Sirunyan et al. Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector. *Submitted to: JINST*, 2019.

[96] Pierguilio Lenzi, Christopher A. Palmer, Joshuha Thomas-Wilsker, et al. Performance of flavour tagging algorithms at 13 tev with 2018 data. CMS Note 2018/323, 2019.

[97] Jon Butterworth et al. PDF4LHC recommendations for LHC Run II. *Journal of Physics G: Nuclear and Particle Physics*, 43(023001), 2016.

[98] Vardan Khachatryan et al. Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV. *JINST*, 10:P02006, 2015.

[99] CMS JME. https://twiki.cern.ch/twiki/bin/view/CMS/JECUncertaintySources#Run_2_reduced_set_of_uncertainty.

[100] CMS Collaboration. Search for invisible decays of a Higgs boson produced through vector boson fusion in proton-proton collisions at $\sqrt{s} = 13\ TeV$. *Physics Letters B*, 793, 2019.

[101] Albert M Sirunyan et al. Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS. *Eur. Phys. J. C*, 81:800, 2021.

[102] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth. A*, 434:435, 1999.

[103] A. L. Read. Presentation of search results: the $\mathrm{cl_s}$ technique. *J. Phys. G*, 28:2693, 2002.

[104] Martin Hoferichter, Philipp Klos, Javier Menéndez, and Achim Schwenk. Improved limits for Higgs-portal dark matter from LHC searches. *Phys. Rev. Lett.*, 119:181803, 2017.

[105] LHC Higgs Cross Section Working Group. Handbook of LHC Higgs cross sections: 3. Higgs properties. 2013.

[106] A. Rose et. al. Serenity: An ATCA prototyping platform for CMS Phase-2. *PoS Proc. Sci.*, 2019.

[107] PICMG. *PICMG 3.0 - AdvancedTCA Base Specification.*

[108] L. Calligaris et. al. OpenIPMC: A Free and Open-Source Intelligent Platform Management Controller Software. *IEEE Trans. Nucl. Sci.*, 68(8):2105–2112, 2021.

[109] Yue Wang et. al. Dynamic Graph CNN for Learning on Point Clouds. *arXiv:1801.07829*, 2019.

CURRICULUM VITAE

ALP AKPINAR

19 Fairbanks Street, Brookline, MA 02446
aakpinar@bu.edu

Education

- **PhD, Physics (anticipated)** **08/2023**
  Boston University Boston, MA, USA
- **BSc, Civil Engineering & Physics** **06/2018**
  Bogazici University Istanbul, Turkey

Doctoral Research

Title: Search for invisible decays of the Higgs boson produced via vector boson fusion at LHC with the CMS detector Run-II data
Thesis advisor: Zeynep Demiragli, PhD
Defense date: TBD
Summary: A search for invisible decays of the Higgs boson is performed using CMS Run-II data. No statistically significant excess has been observed, however stringent limits are placed on $\mathcal{B}(H \to \text{inv})$, the branching ratio corresponding to invisible decays of the Higgs boson.

Publications

- The CMS Collaboration, A. Tumasyan *et al.* Search for invisible decays of the Higgs boson produced via vector boson fusion in proton-proton collisions at $\sqrt{s} = 13$ TeV. Phys. Rev. D **105**, No.9, 092007 (2022)