FACULTÉ DES SCIENCES

UNIVERSITÉ LIBRE DE BRUXELLES

Study of the couplings between the Higgs boson and the electroweak gauge bosons in off-shell processes with the CMS experiment

Thesis presented by Mostafa Mahdavikhorrami

in fulfilment of the requirements of the PhD Degree in Science ("Docteur en Sciences")

Academic year 2022-2023

Supervisors: Prof. Pascal Vanlaer, Prof. Nick van Remortel

Thesis jury:

Prof. Gilles De Lentdecker (Université libre de Bruxelles, Chair)
Prof. Pierre van Mechelen (University of Antwerp, Secretary)
Prof. Pascal Vanlaer (Université libre de Bruxelles, Supervisor)
Prof. Nick van Remortel (University of Antwerp, Supervisor)
Prof. Giacomo Artoni (Sapienza Università di Roma, External member 1)
Prof. Majid Hashemi (Shiraz University, Iran, External member 2)

FREEDOM TO RESEARCH

Acknowledgement

I would like to begin by expressing my sincere gratitude to the FNRS for their generous financial support through the Excellence of Science (EoS) programme run together with the FWO, which has made this research possible.

I am also deeply grateful to the individuals who have played a critical role in the successful completion of my PhD thesis. Firstly, I am truly thankful to my supervisors, Pascal Vanlaer and Nick van Remortel for their support in getting my PhD grant and their constant support, guidance, and expert feedback throughout my research journey.

The members of my thesis committee also deserve special mention. I am grateful for their time, dedication, and valuable feedback, which have greatly contributed to the refinement of my thesis.

My wife, Shiva, has been a constant source of inspiration, motivation, and unwavering support. Her love, encouragement, and patience have been instrumental in keeping me focused and motivated throughout my academic journey.

I would also like to extend my sincere appreciation to my friends and teammates, Ulascan Sarica and Andrey Popov, for their significant guidance and advice throughout my PhD journey. Their support and insights have been invaluable in shaping my research.

In addition, my furry companion, Kishmish, provided me with unconditional love, comfort, and companionship throughout my research, keeping me grounded and focused.

Finally, I would like to thank Nicolas and Hanwen for their help and support with my projects whenever I needed it. Their guidance has been essential in the successful completion of my research. Last but not least, I am grateful for one of my best friend, Ali, for his encouragement and humor.

Contents

Introduction

The Standard Model (SM) is a theory that explains the behavior of elementary particles and their interactions. It was developed in the 1970s and has been successful in explaining many observations since then. Recently, an important missing piece of the puzzle, the Brout-Englert-Higgs boson (a.k.a H boson), was discovered in 2012 by two scientific collaborations, ATLAS and CMS, after nearly 50 years of searching. Now that the H boson has been found, it is important to study its properties to determine if it is the same as the one predicted by the Standard Model or if there are differences, which would imply new physics beyond the Standard Model.

By introducing an interaction between the weak force carriers, the W and Z bosons, and the Higgs field, the electroweak symmetry breaking mechanism allows mass to be attributed to these force carriers. The mathematical form of this interaction is therefore an important prediction of the SM. This form has been already studied to some precision using events where the H boson is produced on-shell, and decays into a VV pair, where V denotes a W or a Z boson. Complementary information is obtained by studying on-shell production through the vector-boson fusion mechanism, a mechanism that is also due to the HVV interaction. Assuming the SM for the mathematical form of the interaction, the strength of the coupling of the H boson with W and Z bosons was measured to about 10% precision. Relaxing this assumption, constraints have also been set on deviations from the mathematical form predicted by the SM, parametrized by anomalous couplings.

In the present thesis, we study events where a virtual H boson is produced with a mass much larger than its pole mass ($m^*_{\mathrm{H}} > 220$ GeV as compared to $m_{\mathrm{H}} = 125$ GeV), and subsequently decays into a pair of Z bosons. H boson production in this mass range is called off-shell production. There are at least three reasons why off-shell production is interesting to measure. First, the SM predicts destructive interference between the Higgs-boson mediated contributions and the non-Higgs-mediated contributions to the pp $\to$ ZZ spectrum. The off-shell contribution of the H boson is indispensable for the SM predictions to satisfy unitarity. Evidence for this contribution was missing until the analysis described in this thesis started. Second, since new physics effects can be expected to manifest themselves at high energies, off-shell events could be sensitive to such effects, and bring additional information that is independent from the on-shell events. Studies of beyond-the-standard-model (BSM) anomalous couplings have already been performed in CMS in the H $\to$ ZZ $\to 4\ell$ final state (ℓ = electron or muon) with the data collected in the 2016-2018 period (so-called LHC run 2, corresponding to $138 fb^{-1}$ of integrated luminosity). In the present thesis, we study similar effects using the run 2 dataset in the H $\to$ ZZ $\to 2\ell2\nu$ final state. Third, under certain hypotheses, a measurement of the H boson decay width can be deduced from the measurement of the ratio of off-shell to on-shell production cross sections. Once the mass of the H boson is known, the H boson width is accurately predicted in the SM. Therefore any deviation from this prediction would indicate new physics. The method employing the ratio of off-shell to on-shell event cross sections is the only method known so far that can reach the precision needed in order to test the SM value of the width, around 4.1 MeV. Until the analysis described in this thesis started, the best measurement was not very precise. The extra sensitivity brought by off-shell events in the $2\ell2\nu$ final state and the use of the large datasets collected by CMS during LHC run 2 was expected to significantly improve the sensitivity, which further motivated our work.

The main advantage of using the $2\ell2\nu$ final state is its favourable branching fraction, about 6 times larger than the 4ℓ final state. To study this final state, events with 2 charged leptons compatible with the decay of a Z boson and with a large imbalance in transverse momentum ($p_{\mathrm{T}}^{\mathrm{miss}}$) attributed to the neutrinos are selected. The large branching fraction is particularly interesting when testing the vector-boson fusion (VBF) production of the H boson. The VBF production mode is about 10 times less frequent than the gluon-fusion (GF) mode. Using $2\ell2\nu$ events significantly improves the statistical precision of the measurements of the VBF process, in a measurement in which statistical uncertainties are an important source of uncertainty on the final results. We therefore categorize our events according to the jet multiplicity, and make use of observables sensitive to the kinematical characteristics of vector-boson fusion production. Such a categorization requires good modelling of the jet multiplicity. The modelling of the jet multiplicity in the signals and in the interfering backgrounds is one of the aspects on which the author of this thesis has worked.

One of the main disadvantages of the $2\ell2\nu$ channel is that it is affected by instrumental background. This background is due to the pp $\to \mathrm{Z}(\to 2\ell) + $jet process where the jets are mismeasured, leading to instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$. This instrumental background has two major consequences. First, it makes the on-shell production of the H boson unobservable in the $2\ell2\nu$ final state. Therefore, in order to extract a measurement of the H boson width, the off-shell measurements in the $2\ell2\nu$ final state have to be combined with on-shell measurements in the 4ℓ channel. Second, since the background is due to tails in the measurement of all the particles in the event, it cannot be simulated reliably. It must be measured from data. Unfortunately the control samples do not provide large statistics in the signal region, which causes extra uncertainties. The author of this thesis has not been directly involved in the measurement of the instrumental background. However, we have contributed to the design of event filters for CMS in order to reduce the contribution from large instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$.

The presentation of this thesis is organized in the following way. In chapter 1, a general overview of the standard model is provided, with a specific emphasis on the H boson and its interactions with the massive vector bosons, the Z and W bosons. The production and decay of off-shell H bosons, specifically in the $\mathrm{H} \to \mathrm{ZZ}$ channel are discussed, and recent experimental results pertaining to this research are highlighted. Chapter 2 provides an overview of the Large Hadron Collider (LHC) and of the CMS detector. In chapter 3, we provide an explanation of the methods used to reconstruct individual particles from detector signals - muons, electrons, photons, jets and missing transverse momentum - as well as the particle identification criteria applied in this search. The $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction and calibration, and also the performance of $p_{\mathrm{T}}^{\mathrm{miss}}$ algorithms and $p_{\mathrm{T}}^{\mathrm{miss}}$ filters (a.k.a MET filters) are presented in chapter 4. Finally, chapter 5 describes the analysis strategy, the modelling of the signals and of the interfering backgrounds, the estimation of non-interfering backgrounds, the observables used to extract the off-shell signal contributions, the statistical method to fit those contributions, and the relevant systematic uncertainties. We also discuss the results and their interpretations. Summary and perspectives (chapter 6) conclude the thesis.

Chapter 1

Review of Theory and Experimental Results

In this chapter, we summarize the theoretical framework describing elementary particles and their interactions, i.e. the Standard Model of particle physics (SM). We introduce the mechanism of electroweak symmetry breaking known as the Brout-Englert-Higgs mechanism, and one of its observable consequences, the existence of a scalar particle, the H boson. The H boson was discovered in 2012 at CERN by the ATLAS and CMS experiments [1–3]. We then focus on the subject of this thesis, namely the production of off-shell H bosons. More precisely, we focus on what can be learned about the electroweak symmetry breaking and about the H boson properties by studying off-shell H boson production. We also summarize the status of experimental results on the questions that we have studied in the present thesis.

1.1 Standard Model of Particle Physics

The Standard Model of particle physics is a theory describing all known fundamental interactions apart from gravity, i.e. the electromagnetic, weak and strong interactions. The SM is a quantum field theory in which the interactions between particles are described by gauge symmetries.

Matter particles are described by elementary fermions, i.e. leptons, quarks, and their antiparticles. Interaction messengers are described by elementary bosons, namely the photon, the Z and W bosons, which are spin-1 bosons, and the Higgs (H) boson, which is

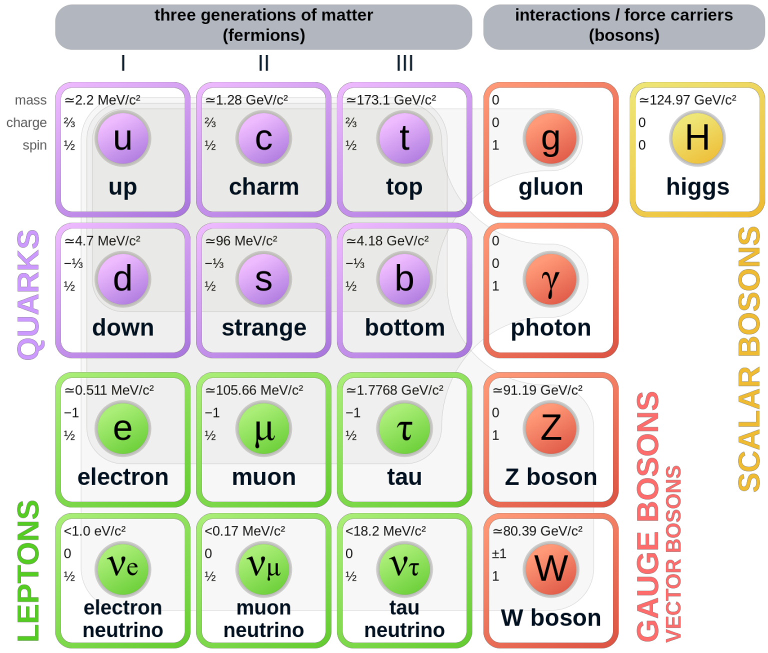

Figure 1.1: Standard Model of elementary particles.

a spin-0 boson. By the discovery of the H boson, strong evidence for the Brout-Englert-Higgs mechanism of electroweak symmetry breaking is established, which makes computations in the SM consistent up to very high energy scales, where quantum effects of gravity would have to be accounted for by a more complete theory than the SM.

The particle content of the SM is shown in Fig.1.1. In this figure, the fermions are categorized in 3 generations (leftmost three columns), and the shaded loops connect those fermions to the gauge bosons with which they can interact.

1.1.1 The Standard Model as a Gauge Theory

From a mathematical point of view, the SM is a kind of gauge theory with the local symmetry group $G_{SM} = U(1)_Y \times SU(2)_L \times SU(3)_C$, in which the $U(1)_Y \times SU(2)_L$ subgroup corresponds to the electromagnetic and weak interactions unified in the electroweak interaction, and $SU(3)_C$ describes the strong interaction. According to Noether's theorem, due to these symmetries, there will be conserved charges, such as the hypercharge denoted by the index Y, the weak isospin, and color (index C). After electroweak symmetry breaking, $U(1)_Y \times SU(2)_L$ is broken into the electromagnetic interaction

$U(1)_{\text{EM}}$, that conserves electric charge (denoted Q further in the text).

Based on these symmetries the Lagrangian of the SM can be summarized as

$$\mathcal{L}_{SM} = \mathcal{L}_{QCD} + \mathcal{L}_{EW} \tag{1.1}$$

where $\mathcal{L}_{QCD}$ is the Lagrangian of the strong interaction, described by quantum chromodynamics (QCD), and $\mathcal{L}_{EW}$ is the Lagrangian of the electroweak interaction.

Quarks are the only fermions participating in the strong interaction. Their field is described in the fundamental representation of $SU(3)_C$ group in which each quark is introduced by three components that correspond to the three color charges. Since $SU(3)_C$ group has eight generators, known as Gell-Mann matrices, there are eight corresponding gluon fields (gauge bosons) G^a, $a = 1, 2, ..., 8$, described in the adjoint representation of $SU(3)_C$.

The QCD part of the Lagrangian can be written as

$$\mathcal{L}_{QCD} = \overline{\Psi}_i \left(i\not{D}_{ij} - \delta_{ij} m_q \right) \Psi_j - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a \tag{1.2}$$

where the first term corresponds to the interaction and propagation of quarks, and the second term corresponds to the propagation and interaction of gluons. Summation over all repeated indices is implied, Ψ_i is the quark field indexed by i and j running from 1 to 3, $\not{D} \equiv \gamma^\mu D_\mu$ in which γ^μ are Dirac matrices (a.k.a gamma matrices) the D_μ is the gauge covariant derivative and $G^a_{\mu\nu} G^{\mu\nu}_a$ represent the gauge invariant gluon field strength tensors. The gauge covariant derivative is defined as

$$D_\mu = \partial_\mu - i g_s \frac{\lambda_a}{2} G^a_\mu \tag{1.3}$$

where g_s is the coupling constant of the strong interaction, λ_a are the Gell-Mann matrices. The gauge invariant gluon field reads as

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu \tag{1.4}$$

where f^{abc} are the structure constants of $SU(3)$.

Fermions are considered as two-component objects known as left-handed and right-handed chiral states,

$$\psi_L = \frac{1-\gamma_5}{2}\psi, \quad \psi_R = \frac{1+\gamma_5}{2}\psi \tag{1.5}$$

The left-handed component of fermions are in the structure of $SU(2)_L$ doublets, in the other word, they are described in the fundamental representation of the subgroup that are,

$$Q_L^m = \begin{pmatrix} u_L^m \\ d_L^m \end{pmatrix}, \quad L_L^m = \begin{pmatrix} \nu_L^m \\ e_L^m \end{pmatrix} \tag{1.6}$$

in which $m = 1,2,3$ index indicates generation. On the other hand, the right-handed component of fermions such as u_R^m, d_R^m, e_R^m are described in the trivial representation of the $SU(2)_L$ thus they are invariant under transformation corresponding to $SU(2)_L$ subgroup. All fermions are transformed under $U(1)_Y$ subgroup. Therefore, these two chiral states are transformed differently according to $SU(2)_L \times U(1)_Y$ symmetry group. Electroweak and electric charges of different components of fermions are listed in Table 1.1 where Q is the electric charge, T_3 is the third component of weak isospin (the $SU(2)_L$ component) and Y_W is the weak hypercharge satisfying the relation

$$Q = T_3 + \tfrac{1}{2} Y_\mathrm{W}. \tag{1.7}$$

Table 1.1: Fermion charges under $SU(2)_L$, $U(1)_Y$, and $U(1)_Q$ groups: electric charge Q, weak isospin T_3 , and weak hypercharge Y_W , respectively. Fermions are split into left- and right- handed chiral states. Each line refers to all three generations.

Fermions	Q	T_3	Y_W
ν_L	0	$+1/2$	-1
e_L	-1	$-1/2$	-1
u_L	$+2/3$	$+1/2$	$+1/3$
d_L	$-1/3$	$-1/2$	$+1/3$
ν_R	0	0	0
e_R	-1	0	-2
u_R	$+2/3$	0	$+4/3$
d_R	$-1/3$	0	$-2/3$

The electroweak part of the SM Lagrangian can be organized in

$$\mathcal{L}_{EW} = \mathcal{L}_\mathrm{gauge} + \mathcal{L}_\mathrm{fermions} + \mathcal{L}_\mathrm{Higgs} + \mathcal{L}_\mathrm{Yukawa}, \tag{1.8}$$

in which the $\mathcal{L}_\mathrm{gauge}$ term determines the interactions between gauge bosons i.e. the W_μ^a ($a = 1,2,3$ denoting three weak gauge bosons corresponding to the generators of $SU(2)_L$) and the B_μ which are described in the adjoint representation of $SU(2)_L$ and

$U(1)_Y$ symmetry groups respectively. The $\mathcal{L}_{\text{gauge}}$ term is specified as

$$\mathcal{L}_{gauge} = -\tfrac{1}{4}W_a^{\mu\nu}W_{\mu\nu}^a - \tfrac{1}{4}B^{\mu\nu}B_{\mu\nu}, \tag{1.9}$$

where $W_a^{\mu\nu}$, $a = 1,2,3$ and $B^{\mu\nu}$ are the field strength tensors for the weak isospin and weak hypercharge gauge fields, given by

$$\begin{aligned} W_{\mu\nu}^a &= \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g\epsilon^{abc}W_\mu^b W_\nu^c & (1.10)\\ B_{\mu\nu} &= \partial_\mu B_\nu - \partial_\nu B_\mu & (1.11)\end{aligned}$$

Here g and ϵ^{abc} the Levi-Civita symbol, are the coupling constant and the structure constant of the $SU(2)$ group respectively.
$\mathcal{L}_{\text{fermions}}$ is the kinematic term for fermions which is described through the interaction between fermions and vector bosons by the gauge covariant derivative,

$$\mathcal{L}_{fermions} = \overline{Q}_j\, i\not{D}\, Q_j + \overline{u}_j\, i\not{D}\, u_j + \overline{d}_j\, i\not{D}\, d_j + \overline{L}_j\, i\not{D}\, L_j + \overline{e}_j\, i\not{D}\, e_j, \tag{1.12}$$

where sums over the three generations of fermions are implied through j index. Q, u, and d are the left-handed doublet, right-handed singlet up, and right handed singlet down quark fields respectively, and L and e are the left-handed doublet and right-handed singlet charged lepton fields respectively. D is the gauge covariant derivative defined as

$$D_\mu \equiv \partial_\mu - i\frac{g'}{2}Y\,B_\mu - i\frac{g}{2}\sigma_j\,W_\mu^j, \tag{1.13}$$

where Y is the weak hypercharge, the σ_j are the components of the weak isospin that are the Pauli matrices and g and g' are the coupling constants for groups $SU(2)_L$ and $U(1)_Y$. The $\mathcal{L}_{\text{Higgs}}$ indicates the interaction between Higgs field, vector bosons and itself, given by

$$\mathcal{L}_{\text{Higgs}} = (D_\mu\Phi)^\dagger D^\mu\Phi - V\left(\Phi^\dagger\Phi\right), \tag{1.14}$$

where Φ is an $SU(2)_L$ doublet of complex scalar fields with a weak hypercharge $Y_H = 1$ and V is the potential of the Higgs field chosen to be in the form of

$$V\left(\Phi^\dagger\Phi\right) = \mu^2\Phi^\dagger\Phi + \lambda\left(\Phi^\dagger\Phi\right)^2, \tag{1.15}$$

with μ and $\lambda > 0$ as its parameters. The $\mathcal{L}_{\text{Yukawa}}$ term describes the way the Higgs field generates fermion masses through Yukawa interactions and it reads as

$$\mathcal{L}_{\text{Yukawa}} = -Y_{ij}^u\overline{Q}_{Li}\tilde{\Phi}u_{Rj} - Y_{ij}^d\overline{Q}_{Li}\Phi d_{Rj} - Y_{ij}^e\overline{L}_{Li}\Phi e_{Rj} + \text{h.c.}, \tag{1.16}$$

here $\tilde{\Phi} = i\sigma_2\Phi^*$ and $Y_{ij}^{u,d,e}$ are matrices of Yukawa couplings.

1.1.2 The Brout-Englert-Higgs Mechanism

So far, all SM interactions can be explained by Lagrangians in equations (1.2) and (1.8).On the other hand, from experimental point of view the $\mathrm{W}^{\pm}$ and Z gauge bosons are observed to be massive and their masses are measured to be around 80 GeV and 91 GeV respectively, however, they seem to be massless according to the aforementioned Lagrangians. Moreover, fermions are also observed to have mass which is another missing part of the Lagrangians in equations (1.2) and (1.8). Although, one could add terms like $-m_f(\bar{f_L}f_R + \bar{f_R}f_L)$ to these Lagrangians which assign m_f to fermion f mass, these terms are not invariant under the $SU(2)_L$ transformations and thus can not be added explicitly. These apparent contradictions are resolved by introducing a spin-0 boson (H boson) through the Brout-Englert-Higgs mechanism. In this mechanism the $SU(2)_L \times U(1)_Y$ symmetry is spontaneously broken to $U(1)_Q$ and consequently weak bosons and fermions mass are generated.

The shape of the Higgs potential (1.15) has a key role in determining the ground state of the theory. If $\mu^2 > 0$, the vacuum expectation $\langle 0|\Phi|0\rangle$ is zero. Otherwise, if $\mu^2 < 0$ then $|\Phi| = 0$ is the point of an unstable local maximum of the potential. In this case the minimum is computed as,

$$\Phi^\dagger\Phi = \frac{v^2}{2}, \tag{1.17}$$

where $v = \sqrt{-\frac{\mu^2}{2}}$. Without loss of generality we can choose the ground state to be,

$$\langle 0|\Phi|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.18}$$

Despite the fact that the Lagrangian (1.14) preserves the $U(1)_Y \times SU(2)_L$ symmetry, the chosen ground state (1.18) does not contain such symmetry and it breaks all generators of the group indeed. However, since the ground state (1.18) respects the combination of the group generators $\sigma^3 + Y_H/2 \cdot \mathbb{1}$, which corresponds to the electric charge, it is invariant under $U(1)_Q$ transformation.
In perturbation point of view, with unitary gauge, the scalar field can be parametrized as

$$\Phi = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + H \end{pmatrix} \tag{1.19}$$

where H is a Hermitian field. By substituting the scalar field doublet (1.19) into Lagrangian (1.14) and ignoring terms containing gauge bosons, we will get

$$\mathcal{L}_{\text{Higgs}} \supset \frac{1}{2}\partial_\mu H \partial^\mu H - \lambda v^2 H^2 - \lambda v H^3 - \frac{\lambda}{4} H^4 + \frac{\lambda}{4} v^4, \tag{1.20}$$

in which the H boson mass is determined as $m_{\text{H}} = \sqrt{2\lambda v^2}$. The Brout-Englert-Higgs mechanism also determines the gauge bosons masses. These masses are provided with terms quadratic in v in the Higgs Lagrangian,

$$\begin{aligned}\mathcal{L}_{\text{Higgs}} \supset \frac{v^2}{8}\left(g^2|W_\mu^1 - iW_\mu^2|^2 + \left(gW_\mu^3 - g'B_\mu\right)^2\right) = \\ \frac{g^2v^2}{8}W_\mu^+ W^{+\mu} + \frac{g^2v^2}{8}W_\mu^- W^{-\mu} + \frac{\left(g^2+g'^2\right)v^2}{8} Z_\mu Z^\mu + 0\cdot A_\mu A^\mu\end{aligned} \tag{1.21}$$

where the mass eigenstates $W_\mu^\pm, Z_\mu$ and A_μ are defined as

$$W_\mu^\pm \equiv \frac{W_\mu^1 \mp iW_\mu^2}{\sqrt{2}} \tag{1.22}$$

$$A_\mu \equiv \frac{g'W_\mu^3 + gB_\mu}{\sqrt{g^2+g'^2}}, \quad Z_\mu \equiv \frac{gW_\mu^3 - g'B_\mu}{\sqrt{g^2+g'^2}} \tag{1.23}$$

These eigenstates $W_\mu^\pm, Z_\mu$ and A_μ correspond to the physical $W^\pm$ and Z bosons and the photon (γ) respectively with masses $m_W = \frac{gv}{2}, m_Z = \frac{v\sqrt{g^2+g'^2}}{2}$ and $m_\gamma = 0$. Similarly, the Yukawa Lagrangian (1.16) with the parametrization (1.19) translates into,

$$\mathcal{L}_{\text{Yukawa}} = -Y_{mn}^u \bar{u}_{Lm} u_{Rn} \frac{v+H}{\sqrt{2}} - Y_{mn}^d \bar{d}_{Lm} d_{Rn} \frac{v+H}{\sqrt{2}} - Y_{mn}^e \bar{e}_{Lm} e_{Rn} \frac{v+H}{\sqrt{2}} + \text{h.c.} \tag{1.24}$$

where the terms which are proportional to v are responsible to give masses to fermions. Generally, the mass matrices $v/\sqrt{2}\cdot Y$ are not diagonal and need to be diagonalized so that the fermions fields will be mass eigenstates. Therefore, the mass matrices can be diagonalized using unitary matrices $V_{L,R}^u$:

$$\begin{aligned}-\frac{v}{\sqrt{2}}\bar{u}_L Y^u u_R + \text{ h. c. } = -\frac{v}{\sqrt{2}}\bar{u}_L V_L^{u\dagger} \left(.V_L^u Y^u V_R^{u\dagger}\right) V_R^u u_R + \text{ h. c. } = \\ -\frac{v}{\sqrt{2}}\bar{u}_L V_L^{u\dagger} Y_D^u V_R^u u_R + \text{ h.c.}\end{aligned} \tag{1.25}$$

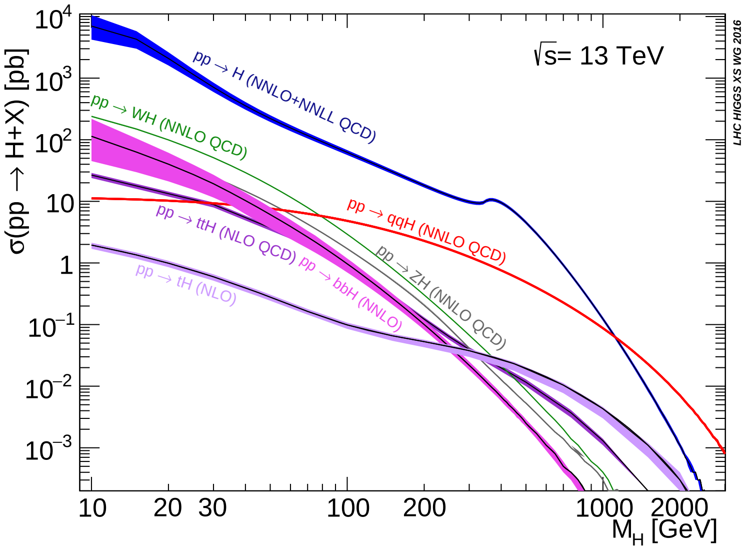

Figure 1.2: The SM H boson production cross section as a function of the H boson mass (SM-like coupling, narrow width approximation, no electroweak corrections) at $\sqrt{s} = 13$ TeV [7].

where $Y_D^u = \mathrm{diag}(m_u, m_c, m_t)$. For other fermions such as down-type quarks and charged leptons a similar operation can be done. As a result, the mass of fermion f is given by

$$m_f = \frac{y_f v}{\sqrt{2}} \tag{1.26}$$

in which y_f is the corresponding component of the diagonalized Yukawa matrix Y_D.

1.2 Standard Model H Boson Physics at the LHC

By the discovery of Higgs particle with a mass of $125.09 \pm 0.21(\mathrm{stat}) \pm 0.11(\mathrm{sys})$ GeV by ATLAS and CMS experiments, its properties are being studied. So far the measured value of the cross sections and the partial decay widths of the H boson are all in agreement with the SM predictions. However, there is still room for BSM effects in the Higgs sector which are being actively searched for [4–6]. The cross section of H boson production through different processes and its branching ratios (BR) to different particles (SM-like coupling, narrow width approximation, no electroweak corrections) are shown in Figs. 1.2 and 1.3 respectively [7].

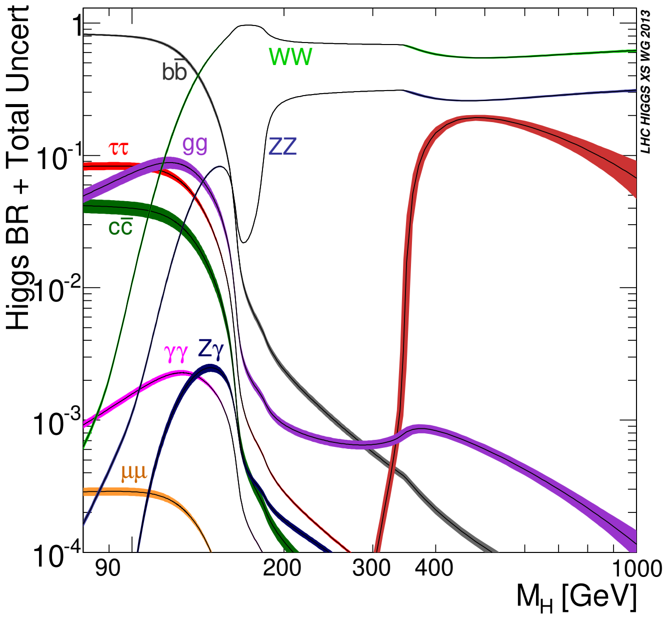

Figure 1.3: H boson branching ratios and their uncertainties as a function of the H boson mass [7].

1.2.1 Production and Decay of a SM-like H Boson

The H boson is an unstable particle. The distribution of mass values follows a relativistic Breit-Wigner function in the narrow width approximation, expressed as:

$$f(E) = \frac{1}{\left(E^2 - m_H^2\right)^2 + m_H^2\Gamma_{\mathrm{H}}^2} \tag{1.27}$$

where m_H is the pole mass and Γ_{H} is its width. By Heisenberg's uncertainty particle, the particle width Γ is related to its lifetime τ by

$$\Gamma\tau = \frac{h}{2\pi} \tag{1.28}$$

where h is Planck's constant. Once the H boson mass is known, all the couplings of the H boson to other particles are defined in the SM. The H boson width Γ_{H} can thus be computed and is predicted to be 4.14 ± 0.02 MeV [7].

Higgs particles with a mass value close to their pole mass, $|m - m_H| <$ a few times Γ_{H}, will be referred to as 'on-shell' H bosons further in this thesis.

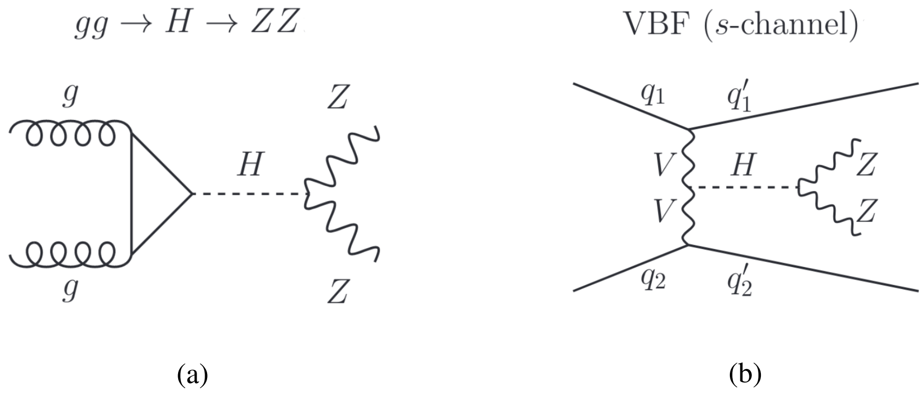

Figure 1.4: The tree-level Feynman diagrams for (a) GF and (b) VBF Higgs production processes.

1.2.2 Off-shell H boson Production

In certain processes, there is a significant probability for the H boson to appear with a mass very different from its pole mass. For instance, at the LHC, in the H $\to$ ZZ decay channel, about 10% of the events involve H bosons with masses above 200 GeV. Such events will be referred to as events with an "off-shell" H boson. These events provide data by which some of the H boson properties can be tested, in a way that complements the information from on-shell H boson production. The analysis of off-shell H boson production in CMS is the focus of the present thesis.

In the following section, the physics of off-shell H boson contributions is first summarized. The measurements of H boson properties that are sensitive to off-shell H boson data are described afterwards. Eventually, the state-of-the-art of off-shell measurements from ATLAS and CMS are described.

Following the discussion in [8], the differential cross section of a process i $\to$ H $\to$ f mediated by a H boson and with initial and final states i and f can be written as:

$$d\sigma = \frac{dq^2 d\phi_d d\phi_q}{4\pi s} \left(\left| \mathcal{M}_p \left(q^2\right) \right|^2 \, \left| D\left(q^2\right) \right|^2 \, \left| \mathcal{M}_d \left(q^2\right) \right|^2 \right), \tag{1.29}$$

where $\sqrt{s}$ is the center-of-mass energy of the initial states particles, q is the four-momentum of the H particle hence $q^2 = m^2_{\mathrm{ZZ}}$, $\mathcal{M}_{p,d}$ are the H boson production and decay Matrix element respectively, and $D(q)$ is the H boson propagator. In the H $\to$ ZZ decay channel considered, q^2 is equal to the square of the invariant mass of the Z boson pair, further denoted m_{ZZ}.

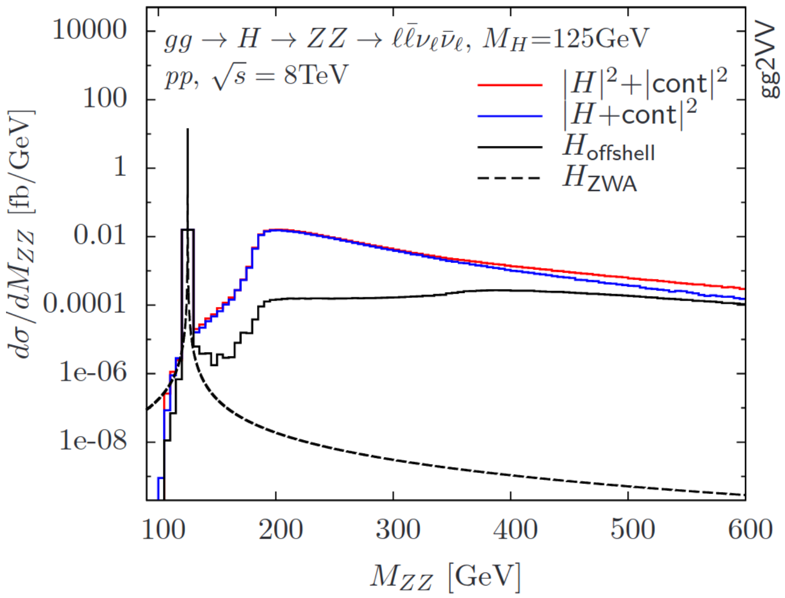

Figure 1.5: Contributions to m_{ZZ} distributions for $\mathrm{gg} \to \mathrm{H} \to \mathrm{ZZ} \to 2\ell 2\nu$ for $m_{\mathrm{H}} =$ 125 GeV at $\sqrt{s} = 8$ TeV [8].

In the fixed-width Breit-Wigner scheme, the H boson propagator reads:

$$D\left(q^2, m_{\mathrm{H}}, \Gamma_{\mathrm{H}}\right) = \frac{1}{\left(q^2 - m_{\mathrm{H}}^2\right)^2 + \left(m_{\mathrm{H}}\Gamma_{\mathrm{H}}\right)^2} \tag{1.30}$$

Therefore, when H boson is off-shell ($q^2 > m_{\mathrm{H}}^2$), the propagator term is proportional to $1/m_{\mathrm{ZZ}}^4$ which causes a rapid decrease in the cross section. However, for $\sqrt{(q^2)}$ similar or greater than $2m_{\mathrm{Z}}$, the H decay matrix element $\mathcal{M}_d(q^2)$ behaves as $(q^2)^2$, compensating the q^2 dependence of the H propagator. Thus we expect a remarkable enhancement in the contribution of Higgs to the ZZ spectrum in $\sqrt{q^2} \gtrapprox 2m_{\mathrm{Z}} \simeq 182$ GeV. Distributions of m_{ZZ} in the GF process at center of mass energy of 8 TeV for different contributions to the off-shell cross section are shown in Fig.1.5 [8].

However, processes with identical initial- and final-state particles and identical quantum numbers as the processes involving the H boson will interfere with it. Representative diagrams are shown in Fig.1.4 left for the $g \to$ ZZ process, and in Fig.1.4 right for the electroweak process. These processes lead to a smoothly-decreasing continuum in the m_{ZZ} spectrum. As illustrated in Fig.1.5, the summation of the H and continuum contributions is larger than once their interference is taken into account. This means that the interference is negative which is shown in Fig.1.6 [9]. This is required in the

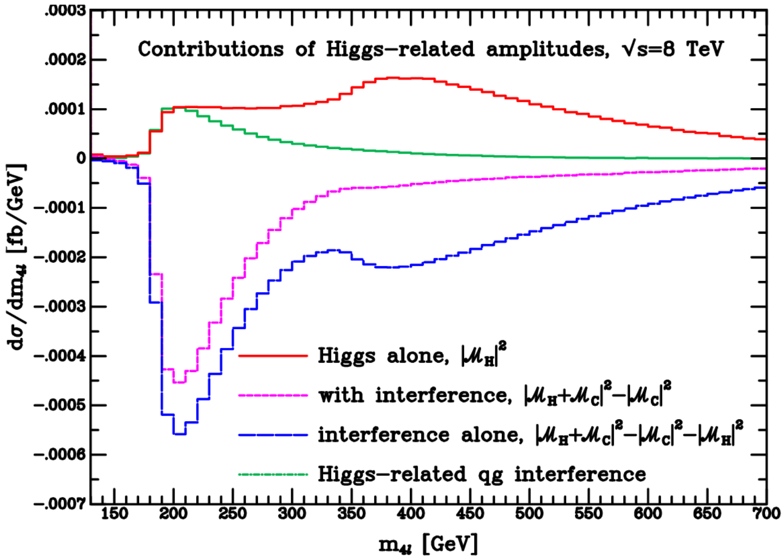

Figure 1.6: Higgs related contributions in the high $m_{4\ell}$ region [9].

SM to preserve unitarity.

1.2.3 H boson Decay Width

As pointed out in [10], the off-shell H boson production offers a unique way to constrain the H boson decay width. Experimentally, the precision on the width of the on-shell H boson production peak is limited by detector resolution, and is around 1 GeV [11–13]. However, a combination of off-shell and on-shell measurements offers a way to extract the width, as follows. From the differential cross section equation (1.29) we have

$$\frac{d\sigma}{dq^2} \propto \frac{g_i(m_{\mathrm{ZZ}})^2 g_f(m_{\mathrm{ZZ}})^2}{\left(q^2 - m_{\mathrm{H}}^2\right)^2 + (m_{\mathrm{H}}\Gamma_{\mathrm{H}})^2}, \tag{1.31}$$

where g_i and g_f are the Higgs coupling constants with initial and final states respectively. By integrating over two different regions such as the on-shell region $q^2 = m_{\mathrm{H}}^2 \pm \Gamma_{\mathrm{H}}^2/2$ and the very off-shell region $q^2 \gg m_{\mathrm{H}}^2$, one can approximate the total cross section in aforementioned regions as

$$\sigma^{\text{on-shell}} \sim \frac{g_i^2 g_f^2}{m_{\mathrm{H}}\Gamma_{\mathrm{H}}} \tag{1.32}$$

$$\sigma^{\text{off-shell}} \sim \frac{g_i^2 g_f^2}{q^2}, \tag{1.33}$$

as a result, by measuring the relative production in on- and off-shell regions, a direct information on Γ_{H} can be obtained. At the time of the start of this Ph.D. thesis the results on H boson decay width constrained by ATLAS and CMS experiments showed acceptable agreement between the SM predictions and observed values. According to the analysis done by the ATLAS experiment in the $ZZ \to 4\ell$ and $ZZ \to 2\ell 2\nu$ final states, the 95% confidence level (CL) upper limit on the $\frac{\Gamma_{\mathrm{H}}}{\Gamma_{\mathrm{H}}^{SM}}$ was observed (expected) to be 3.5 (3.7) [14]. In a similar analysis done by the CMS experiment, the H boson width was measured to be $3.2^{+2.8}_{-2.2}$ MeV at 68% CL intervals [15].

1.2.4 Anomalous Couplings in HVV Interaction

The SM of particle physics is incomplete. For instance there are at least three experimental evidences that can not be described by the SM: the non-zero value of neutrinos mass , dark matter and the higher density of matter v.s. anti-matter in the universe. There are also theoretical questions which are open, such as the description of the gravity by a quantum field theory. However, we don't have any experimental evidence supporting one of the many BSM models. Therefore, one way to look for beyond the standard model physics is to model such effects in an Effective Field Theory (EFT). EFT is an approximation of the underlying quantum field theory, and allows to parametrize small deviations from the SM in a consistent way by introducing higher order operators in the Lagrangian, that are suppressed at low energy scales. As pointed out in [7] any EFT Lagrangian has generally the form reading as,

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \sum_i \frac{c_i^{(5)}}{\Lambda}\mathcal{O}_i^{(5)} + \sum_i \frac{c_i^{(6)}}{\Lambda^2}\mathcal{O}_i^{(6)} + \sum_i \frac{c_i^{(7)}}{\Lambda^3}\mathcal{O}_i^{(7)} + \sum_i \frac{c_i^{(8)}}{\Lambda^4}\mathcal{O}_i^{(8)} + \cdots$$

where $c_i^{(d)}$ are called Wilson coefficient of degree of d and Λ is mass scale of new heavy particles or interactions which $\Lambda \gg$ SM masses and LHC energy scale, one such parametrization is the Standard Model Effective Field Theory (SMEFT).

Alternatively, a similar study of BSM can be done more specifically to the H boson with two spin-one gauge bosons a.k.a vector bosons. In this method the HVV interaction is parameterized with a scattering amplitude that includes additional terms describing beyond the standard model interactions. According to [7, 16] the extended HVV amplitude is given by,

$$\begin{aligned} A(\text{HVV}) \sim & \left[a_1^{\text{VV}} + e^{i\Phi_{\Lambda_1}} \frac{q_1^2 + q_2^2}{\Lambda_1^2} + \ldots \right] m_{\text{V}}^2 \epsilon_{\text{V}1}^* \epsilon_{\text{V}2}^* \\ & + |a_2| e^{i\Phi_{a_2}} f_{\mu\nu}^{*(1)} f^{*(2),\mu v} - |a_3| e^{i\Phi_{a_3}} f_{\mu\nu}^{*(1)} \tilde{f}^{*(2),\mu v} \end{aligned} \tag{1.34}$$

where ϵ_i and q_i^{μ} are polarization vector and 4-momentum of gauge boson V_i respectively, $f^{(i)\mu\nu} = \epsilon_i^{\mu} q_i^{v} - \epsilon_i^{v} q_i^{\mu}$, $\tilde{f}_{\mu\nu}^{(i)} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma} f^{(i)\rho\sigma}$ are field and dual field strength tensors, a_1^{VV} are SM leading tree-level contribution which only $a_1^{ZZ,WW} \neq 0$ and from custodial symmetry we have $a_1^{ZZ} = a_1^{WW}$. The rest of the couplings are considered anomalous contributions arising either from SM loop corrections or new BSM contributions. The $\frac{q_1^2 + q_2^2}{\Lambda_1^2} m_{\text{V}}^2 \epsilon_{\text{V}1}^* \epsilon_{\text{V}2}^*$ term is the leading order of the dipole structure assumption for Higgs boson and the Λ_1 is the correspondent anomalous coupling, and anomalous couplings $a_{2,3}$ are the CP-conserving (CP-event) and CP-violation (CP-odd) BSM contributions in HVV interactions respectively. These anomalous couplings may change the m_{VV} and other kinematic variables which help us to investigate the possible deviations from the standard model. Figure 1.7 illustrates the $m_{4\ell}$ distribution for different hypotheses such as the standard model and other anomalous couplings individually [16]. These two methods of studying BSM anomalous couplings are equivalent methods such that one can find the relations between the a_i and Λ_1 couplings in (1.34) and Wilson coefficients $c_i^{(d)}$ in (1.2.4) by referring to [17].

Instead of measuring a_i directly, the measurements of a_i relative to the dominant SM-like contribution a_1 are the preferred approach. This is due to the fact that most of the systematic uncertainties cancel when taking ratios to the total cross section. For this purpose, the effective fractional Higgs to ZZ cross sections f_{ai} and phases ϕ_{ai} are defined as

$$\begin{aligned} f_{ai} &= \frac{|a_i|^2 \sigma_i}{\sum_{j=1,2,3\ldots} |a_j|^2 \sigma_j}, \\ \phi_{ai} &= \arg\left(\frac{a_i}{a_1}\right), \end{aligned} \tag{1.35}$$

where σ_i is the cross section for the process corresponding to $a_i = 1$, $a_{j \neq i} = 0$, while $\tilde{\sigma}_{\Lambda 1}$ is the effective cross section for the process corresponding to $\Lambda_1 = 1$ TeV, given in units of fb$\times$ TeV^4. The cross section ratios are quoted in Table 1.2 [18]. The a_i / a_1 ratios can be obtained from the ratio f_{ai} / f_{a1}, the cross section ratios, and the phase

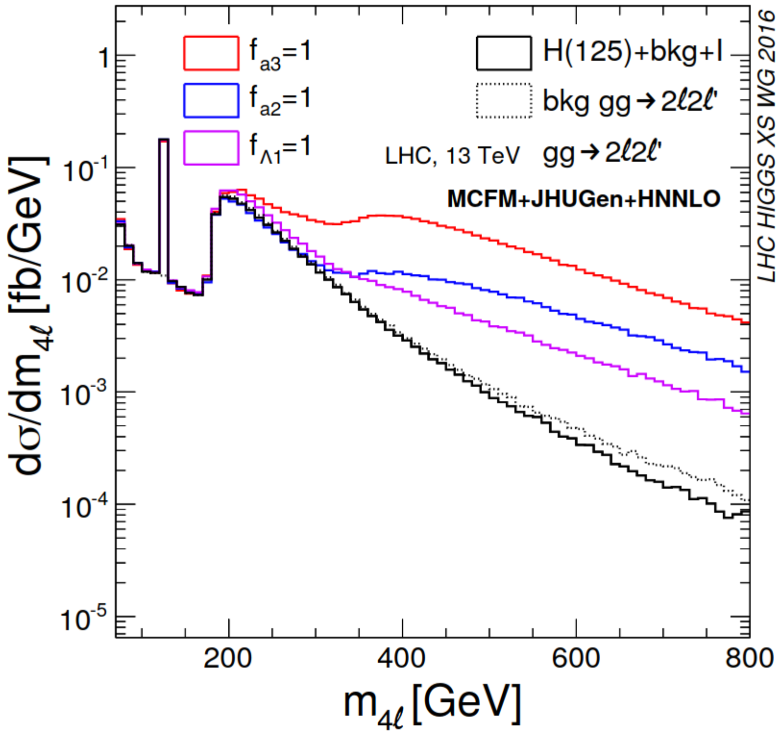

Figure 1.7: Several scenarios of H(125) anomalous couplings to two vector bosons with enhancement in the off-shell region with the a_3, a_2 and Λ_1 terms as coloured histograms, as well as the a_1 term (SM), as the solid black histogram [16].

ϕ_{ai} as

$$\frac{a_i}{a_1} = \sqrt{\frac{f_{ai}}{f_{a1}} \frac{\sigma_1}{\sigma_i}} e^{i\phi_{ai}} \,. \tag{1.36}$$

The effective fractions f_{ai} are bound between 0 and 1 and do not depend on the coupling convention.

In the SM, we expect all these anomalous couplings such as a_2, a_3 and $\frac{1}{\Lambda_1}$ to have 0 value. The latest results achieved by both ATLAS and CMS experiments show that the observed constraints on the measurable parameters of theses couplings are consistent with 0 i.e. no significant deviation from the SM prediction is observed [15, 17, 19]. For instance, the summary of constraints on the HVV coupling parameters in the Warsaw basis of SMEFT are shown in Tab. 1.3 [17].

Table 1.2: Listed are the anomalous HVV couplings considered in our analysis assuming a spin-zero H boson. The translation constants are the cross section ratios corresponding to the process $H \to 2\ell 2\nu$ (or equivalently, $H \to 2e\,2\mu$) with the H boson mass $m_H = 125$ GeV [18].

Anomalous Coupling	Coupling Phase	Effective Fraction	Translation Constant
a_2	ϕ_{a2}	f_{a2}	$\sigma_1 / \sigma_2 = 2.77$
a_3	ϕ_{a3}	f_{a3}	$\sigma_1 / \sigma_3 = 6.53$
Λ_1	$\phi_{\Lambda 1}$	$f_{\Lambda 1}$	$\sigma_1 / \tilde{\sigma}_{\Lambda 1} = 1.47 \times 10^4 \text{TeV}^{-4}$

Table 1.3: Summary of constraints on the HVV coupling parameters in the Warsaw basis of SMEFT [17]. For each coupling constraint reported, three other independent operators are left unconstrained, where only one of the three operators c_{HW} , c_{HWB}, and c_{HB} is independent, and only one of $c_{H\tilde{W}}$, $c_{H\tilde{W}B}$, and $c_{H\tilde{B}}$ is independent.

Channels	Coupling	Observed	Expected
VBF & VH & H$\to 4\ell$	$c_{H\square}$	$0.04^{+0.43}_{-0.45}$	$0.00^{+0.75}_{-0.93}$
	c_{HD}	$-0.73^{+0.97}_{-4.21}$	$0.00^{+1.06}_{-4.60}$
	c_{HW}	$0.01^{+0.18}_{-0.17}$	$0.00^{+0.39}_{-0.28}$
	c_{HWB}	$0.01^{+0.20}_{-0.18}$	$0.00^{+0.42}_{-0.31}$
	c_{HB}	$0.00^{+0.05}_{-0.05}$	$0.00^{+0.03}_{-0.08}$
	$c_{H\tilde{W}}$	$-0.23^{+0.51}_{-0.52}$	$0.00^{+1.11}_{-1.11}$
	$c_{H\tilde{W}B}$	$-0.25^{+0.56}_{-0.57}$	$0.00^{+1.21}_{-1.21}$
	$c_{H\tilde{B}}$	$-0.06^{+0.15}_{-0.16}$	$0.00^{+0.33}_{-0.33}$

Chapter 2

The CMS Experiment at the LHC

This analysis is performed with experimental data taken by the Compact Muon Solenoid (CMS) detector [20]. The CMS detector is installed at the Large Hadron Collider (LHC) [21, 22], close to Geneva, Switzerland. The LHC was constructed by the European organization for nuclear research (CERN) between 1998 and 2008.

2.1 The Large Hadron Collider

The LHC is a circular proton-proton collider which is currently the largest and the most energetic particle collider built. It is located in a 26.7 km circumference tunnel which is between 45 to 170 m underground. The CERN accelerator complex is made of several accelerators that gradually increase the energy of the proton beams up to 6.5 TeV which results in a 13 TeV center of mass energy. Fig. 2.1 is illustrating a schematic view of the accelerators at CERN.

2.2 The CMS detector

The Compact Muon Solenoid (CMS) detector is a general-purpose detector operating at high-luminosity. It is located at LHC Point 5 in Cessy, France. The detector is large and cylindrical with 21.6 m length and a diameter of 14.6 m. Its weight is approximately 12500 t. The CMS detector is described in detail in Ref. [20]. The performance of CMS for reconstructing stable particles is described in Chapter 3.

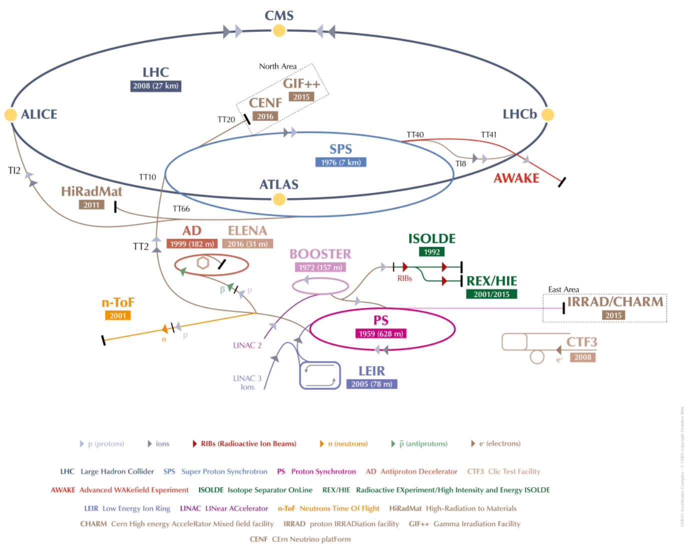

Figure 2.1: Various accelerators are piped to each other to deliver the proton beams to the last ring, the LHC, to reach an energy of 6.5 TeV for the colliding protons.

The CMS detector is constructed in different layers and elements which are shown in Fig. 2.2. The subdetector closest layer to the interaction point is the inner tracking system which is responsible for for the detection and the measurment of the trajectory of electrically charged particles. The next subdetectors surrounding the inner tracker are calorimeters, the electron and hadron calorimeters for the measurement of the energy of electrons, photons and hadrons by absorbing such particles. The last system in CMS detector is the muon system which consists of gaseous detectors responsible to detect muons, as they are the only detectable particles traversing the calorimeters and the coil of the CMS magnet.

2.2.1 The Inner Tracker

The inner tracker system measures the trajectories of electrically charged particles coming from the interaction point or within a few tens of centimeters from it. This information provides the measument of particles momentum and also their charge signature based on the curvature of the trajectory. Moreover, is utilized to reconstruct primary and secondary vertices.

The inner tracker has two subsystems, the pixel tracker which is the CMS component

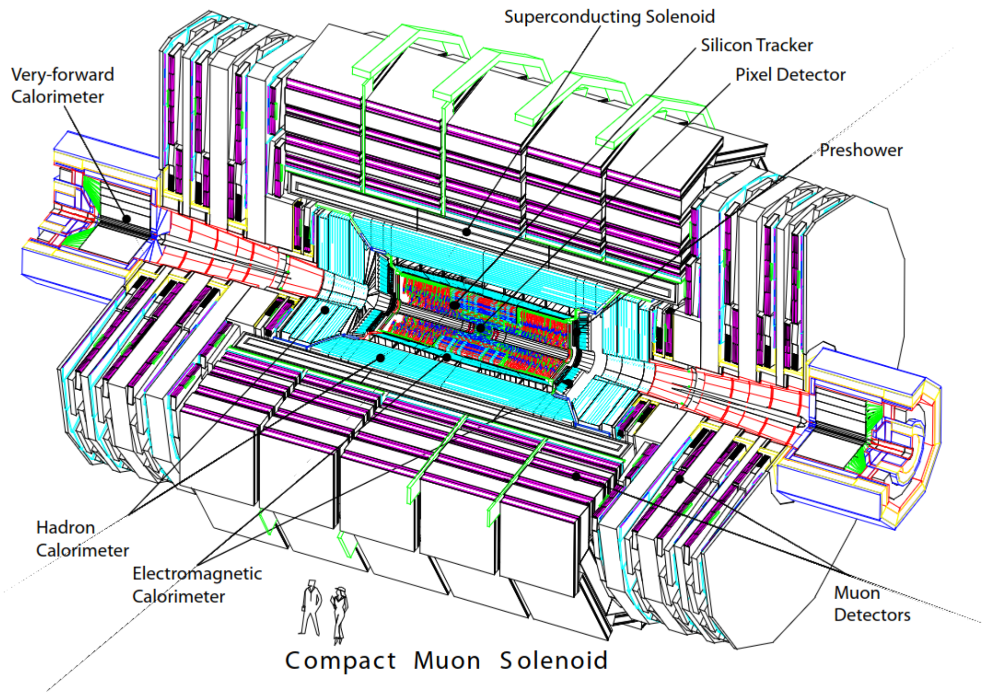

Figure 2.2: View of the CMS detector and its main elements [20].

closest to the interaction point, and the silicon strip tracker. Before the pixel Phase-1 upgrade, originally it consisted of three cylindrical layers in the barrel and two disks in each endcap. The schematic view of the CMS tracker indicated in Fig. 2.3. After the pixel Phase-1 upgrade in the technical stop 2016-2017, the pixel tracker now has four cylindrical layers in the barrel and three disks in each endcap. The mean distance of barrel layers from the interaction point are 2.9, 6.8, 10.9, and 16.0 cm with the length of 54 cm. The endcap disks are located on each endcap at distances of 29.1, 39.6, and 51.6 cm with different inner and outer radius for each half disks. Fig. 2.4 illustrates a comparison between the layout of the CMS Phase-1 pixel detector and the original pixel detector.

The silicon microstrip modules are located at a distance of 20 cm up to 110 cm from the interaction point. The strip tracker has different parts in the barrel and endcap regions. In the barrel, the strip detector is divided into an inner (TIB) and an outer barrel (TOB). The TIB consists of four layers and covers up to $|z| < 65$ cm and the TOB is made of six layers with a half-length of $|z| < 110cm$. In the endcap, the strip tracker is separated into the TEC (Tracker End Cap) and TID (Tracker Inner Disks). Each TEC contains 9 disks covering the region $120 < |z| < 280$ cm, and each TID is made of 3 small disks which fill the gap between the TIB and the TEC.

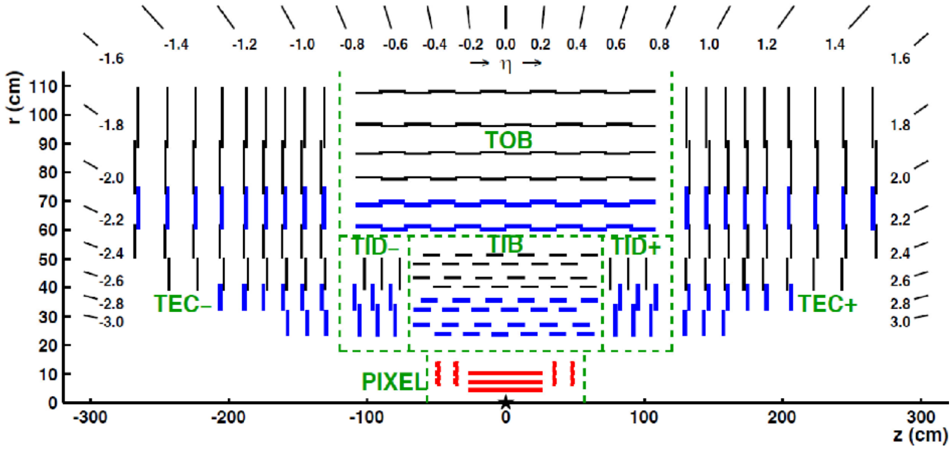

Figure 2.3: Schematic view of the CMS tracker in the r-z plane. The original pixel detector (before the CMS Phase-1 pixel detector upgrade) are shown in red which has three layers (horizontal lines) in the barrel and four (vertical lines) disks in the endcap region. The black and blue lines correspond to Strip tracker modules (stereo modules) [23].

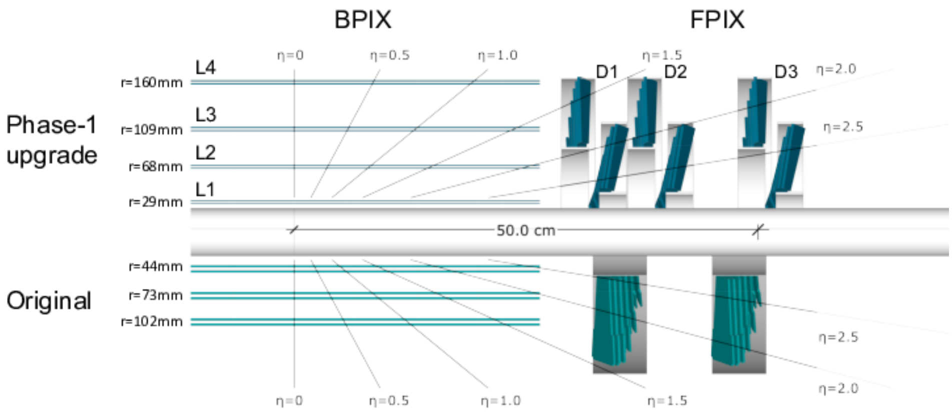

Figure 2.4: Schematic view of the CMS tracker in the r-z plane. The upper side shows the layout of the CMS Phase-1 pixel detector while the lower side shows the original pixel detecor layout [23].

2.2.2 The Electromagnetic Calorimeter

Electromagnetic calorimeter (ECAL) which surrounds the inner tracker system, is the other element of the CMS detector. The goal of the electromagnetic calorimeter is to absorb electrons and photons and precisely measure their energy.

The ECAL is composed of lead-tungstate ($PbWO_4$) scintillating crystals which are very dense (8.3 g/cm^3) [20, 24]. Moreover, the lead-tungstate crystals have short radiation length (0.89 cm), small Molière radius (2.2 cm), and are fast (80% of the light is emitted within 25 ns). However, the short scintillation decay time of $PbWO_4$ is temperature-dependent, therefore, the calorimeter is conserved at a temperature of $(18.00 \pm 0.05)^{\circ}$C. These properties of $PbWO_4$ have thus made it achievable to design a compact calorimeter with a high granularity inside the solenoid.

The ECAL has two separate sections in the barrel (EB) and in the endcaps (EE). The EB has an inner radius of 129 cm and consists of 61200 lead-tungstate crystals in quasi-projective geometry i.e the axes are tilted at 3° with respect to the line from the nominal vertex position. The pseudorapidity coverage of EB is in the range of $|\eta| < 1.479$. Each crystal in EB covers 0.0174 (i.e. 1°) in $\Delta\phi$ and $\Delta\eta$ which has a front face cross section of $\sim 22 \times 22$ mm^2 and a length of 230 mm. The EE partition is located at a distance of 314 cm from the interaction point and 7324 crystals are mounted in each endcaps which are arranged in x-y grid. Each crystal in EE has a front face cross section of 28.6×28.6 mm^2 and a length of 220 mm. The EE covers the pseudorapidity range of $1.479 < |\eta| < 3.0$.

Moreover, a preshower device covers the endcap pseudorapidity range of $1.653 < |\eta| < 2.6$. Each preshower device consists 2 planes of silicon strip detectors with a pitch of 1.9 mm. These devices mainly aim to discriminate between photons and neutral pions which decays into pairs of close photons. In addition, they enhance the accuracy of position measurement for electrons and photons. The schematic view of the EE and EB partitions, and the preshower device are shown in Fig. 2.5.

2.2.3 The Hadron Calorimeters

The Hadron calorimeter (HCAL) is behind the inner tracker and ECAL from the interaction point of view. The main goal of the HCAL is for hadron jets measurement and has an important role in the estimation of missing transverse energy. The HCAL is divided into 4 partitions i.e. hadron barrel (HB), hadron outer (HO), hadron endcap (HE), and hadron forward (HF). The locations of these four partitions are schematically illustrated in a longitudinal view of the CMS detector in Fig. 2.6.

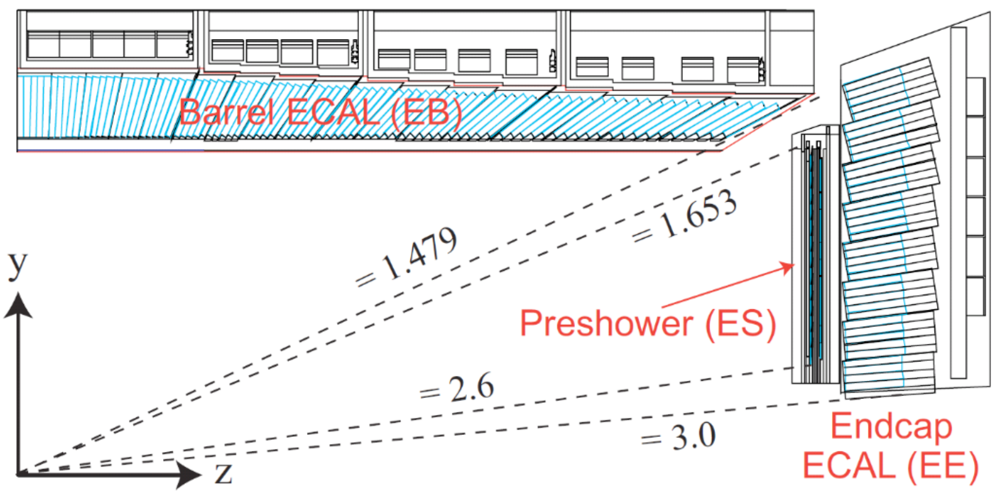

Figure 2.5: The EE and EB partitions and preshower device in x-y plane [24].

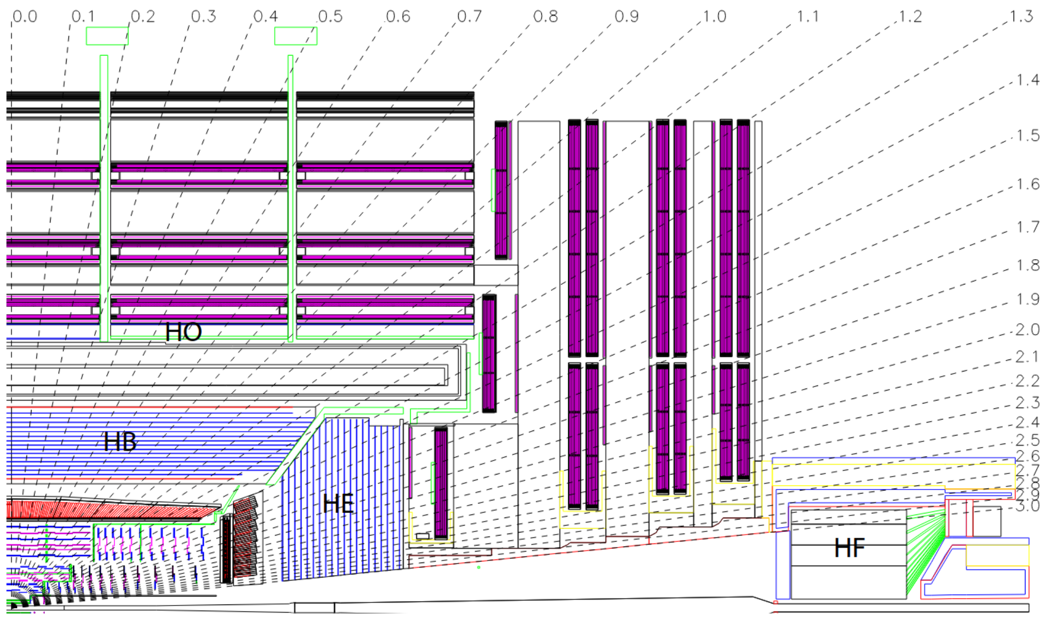

Figure 2.6: The HB, HO, HE, and HF partitions of hadron calorimeters in longitudinal view of CMS detector [20].

The HB is radially placed between the outer limit of the ECAL (R = 1.77 m) and the inner radius of the magnet coil (R = 2.95 m) which covers the pseudorapidity range of $|\eta| < 1.3$. It comprises 36 identical azimuthal wedges, which are made of flat absorber plates parallel to the beam axis. The innermost and outermost absorber plates are constructed out of stainless steel and have thickness of 40 and 75 mm respectively. These two kinds of plates are set up to bring a supplementary structural support. Between them, there are 14 intermediate absorber plates made of brass and each plate is 55.5 or 55.6 mm thick. Between absorber plates there are in total 70000 plastic scintillating tiles inserted. The HO is also mounted in the barrel and covers the pseudorapidity range of $|\eta| < 1.3$ but is mounted around the magnet coil. The HO partition's goal is to capture potential tails of hadronic showers so that prevents energy leakage from the HCAL subsystem.

The HE partition consists of 14 η towers with $5°\phi$ segmentation. It covers the pseudorapidity range of $1.3 < |\eta| < 3.0$ which contains approximately 34% of the particles produced in the final state Each HE partition is made of 17 absorber plates layers in which each layer has 79 mm thickness. There are 18 layers of plastic scintillator which are constructed out of 10458 tiles in each HE partition.

The furtherest partition of the HCAL with respect to the interaction point is the forward hadron calorimeters (HF) which is mounted at 11.2 m from the interaction point. The HF is extended from $|\eta| = 3$ down to $|\eta| = 5.2$ in which a Cherenkov-based, radiation-hard technology is utilized. Due to the location of the HF, they collect a large flux of particles. As a result, the two HF partitions collects in average 760 GeV per inelastic proton-proton collision which is significant compared to only 100 GeV for the rest of the HCAL partitions. Therefore, the HF partitions are specifically designed such that they are able to operate under tremendously intense radiation conditions. Each HF is a cylinder-shape which has an outer radius of 130 cm and a hole with a radius of 12.5 cm for the beam pipe. The absorber layer utilized in the HF is made of steel and has a 165 cm thickness. Moreover, quartz fibres parallel to the beam axis are designed inside the absorber in order to detect the showers developing inside the absorber through Cherenkov light coming from particles emission while the shower are passing through the fibres.

2.2.4 The Muon System

The Muon system has three important goal: muon identification, momentum measurement, and triggering. The muon chambers are interleaved with the successive layers of the return yoke of the magnet, and is the outermost part of the CMS detector.

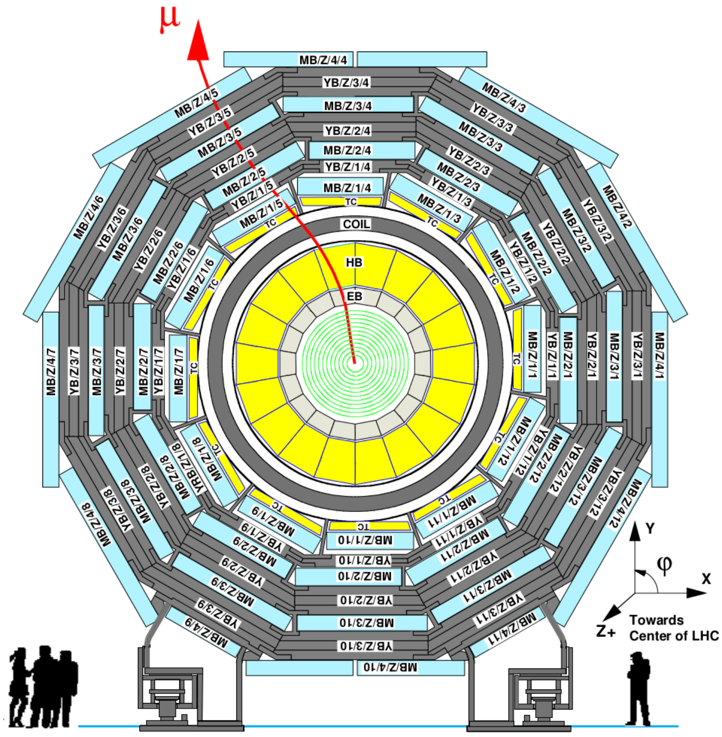

Figure 2.7: Layout of the CMS barrel muon DT chambers in one of the 5 wheels [20].

Similar to other subsystems, the muon system also has a barrel section covering pseudorapidity range of $|\eta| < 1.2$, and two planar sections in both endcap regions which cover the pseudorapidity up to $|\eta| = 2.4$. In the barrel regions the neutron-induced background is small, the muon rate is also low. Moreover, in this region the magnetic field is uniform and mostly contains in the steel yoke. These conditions led designers to utilize drift tube (DT) chambers in the barrel section of muon system. In the barrel section, muon system comprises 4 station in which each station consists of drift tube (DT) systems and resistive plate chamber (RPC) trigger systems. There are a total of 6 layers of RPCs and 250 drift chambers mounted in the barrel muon system. There are almost 172000 sensitive wires in DT barrel. Fig. 2.7 shows a layout of the CMS barrel muon DT chambers in one of the 5 wheels.

Similar to the barrel section, two complementary muon detector systems are situated in the endcap regions for a robust muon identification, i.e. the cathode strip chambers (CSC) and RPCs. There are 4 CSC stations (ME1-4) and 3 RPC stations (RE1-3) mounted in the endcap section as illustrated in Fig. 2.8. In the endcaps, muon system consists of 542 CSCs laid out in 9 groups as shown in Fig. 2.8. The ME2/2, ME3/2 and ME4/2 chambers are the largest ones which size almost $3.4 \times 1.5\text{m}^2$. The sensitive planes of the all chambers cover approximately the area of 5000 m^2 which contains

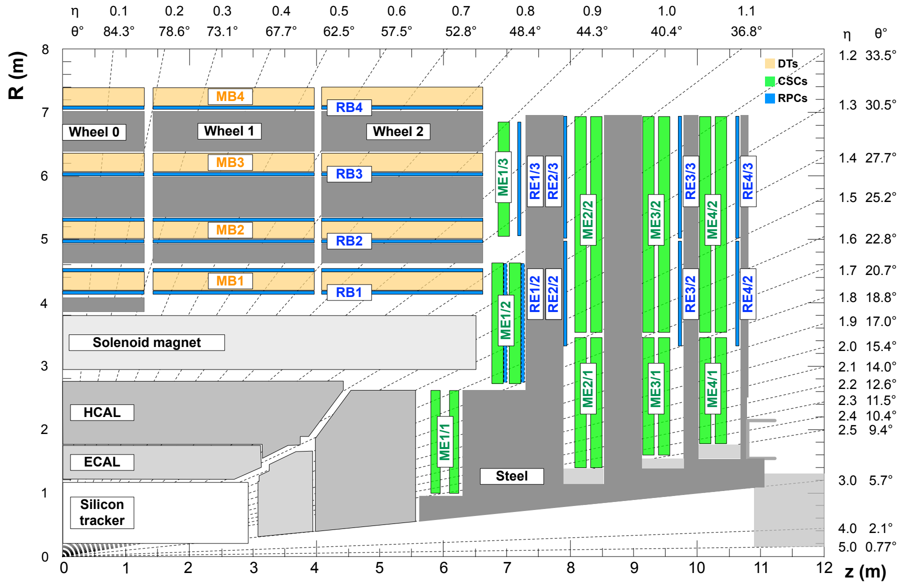

Figure 2.8: One quadrant of the CMS detector in which the DT, CSC and RPC systems are highlighted in yellow, green and blue respectively [25].

more than 50 m^3 gas. The number of wires utilized in CSC system is about 2 million. The CSCs have special properties, they are capable of precision muon measurement and muon trigger in one device. Moreover, they are able to function at high rates and in large and non-uniform magnetic field and operate properly without precise gas, temperature, or pressure control.

2.2.5 The Trigger System

In the proton-proton collision at LHC, the beam crossing interval is 25 ns which corresponds to a collision rate of 40 MHz. Such amount of data is not possible to be stored and processed, therefore an expeditious procedure is demanded in order to decide whether to store a particular event for further analyses or to discard the event. This task is executed through the trigger system which is the first step toward the process of physics event selection. The trigger system performs such a drastic rate reduction in two steps known as the level 1 trigger (L1T) [26] and the high-level trigger (HLT) [27].

The Level-1 Trigger is based on custom-designed and largely programmable electronics (like FPGA and ASIC technologies). The L1T is able to analyse every bunch crossing and decides whether to accept the event within 3.2 μs. The L1T's decision is based on the information achieved from calorimeters and muon system, however the inner tracker information is not utilized in L1T system. The L1T system constraint the rate down to 100 kHz.

The L1 Trigger has various components such as local, regional and global components. The architecture of the Level-1 Trigger is presented in Fig. 2.9. The muon system trigger contains the DT and CSC local triggers, regional muon triggers which are based on the precision tracking chambers such as DT Track Finder (DTTF) in the barrel [28] and the CSC Track Finder (CSCTF) in the endcaps [29], and a Global Muon Trigger (GMT) [30]. The calorimeter trigger comprises calorimeter trigger primitive generator, regional calorimeter trigger and the global calorimeter trigger (GCT). Finally, the Global Trigger (GT) [30] takes the decision whether to accept or reject an event at L1 based on trigger objects delivered by the GCT and GMT.

The HLT consists of a software-based system carried through a filter farm of order(1000) commercial processors. The HLT runs essentially the same event reconstruction code as the offline event reconstruction, but it is tuned for processing speed, and by definition cannot use the most precise detector calibrations.

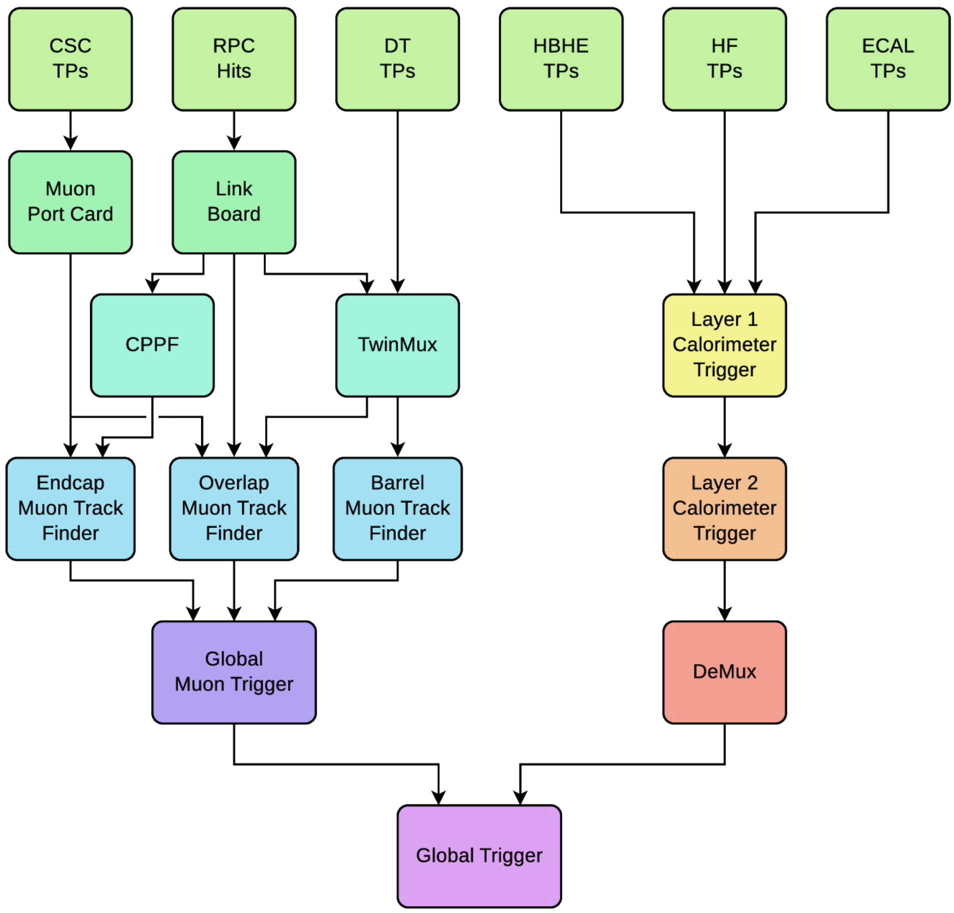

Figure 2.9: Architecture of the Level-1 Trigger [26].

Chapter 3

Event and Physics Objects Reconstruction

In high-energy proton-proton collisions almost all types of elementary particles and hadrons may be produced, but only few of them live long enough to reach the detector. These particles can be electrons, photons, muons, neutrinos, and some hadrons such as pions, kaons, protons, and neutrons. Although neutrinos cannot be directly detected by the CMS detector due to their weak interaction with matter, there are algorithms that can give us some information on them. In addition to the neutrinos, other particles behave differently with respect to CMS subdetectors. Charged particles like electrons, muons and etc. produce signals in the inner tracker. Electrons and photons are absorbed in the ECAL, while hadrons interacts with both ECAL and HCAL and deposit most of their energy in the HCAL. Since muons have much larger mass than electrons, they traverse the solenoid of the CMS magnet, the ECAL and the HCAL almost without interaction, and finally cross the muon chambers Finally particles and their kinematics are reconstructed by gathering and analyzing all signals obtained from different detector systems. After physics object reconstruction, the next step is the identification of such objects. Generally, the identification conditions for different physics objects may be different for various analysis. In the following sections, the reconstruction process and identification requirements of the main physics objects used in this analysis are discussed briefly.

3.1 Particle flow

A particle-flow (PF) candidate is an object which may represent any final-state particle detected individually by CMS detector systems in each event. To obtain a PF candidate, the measurements from all detector layers (i.e., tracks and clusters) need to be combined. This approach which is known as particle-flow reconstruction, is a part of the CMS event reconstruction based on the particle-flow (PF) algorithm [31]. Hence, the particle-flow algorithm aims at reconstructing and identifying all final-state particles in the event, i.e., electrons, muons, photons and hadrons, by thoroughly combining all signals from CMS detector systems towards optimal determination of particles properties such as their direction, momentum, energy and type.

3.2 Primary vertices and pileup collisions

In each proton bunch crossing, a number of proton collisions may occur. The average number of proton collisions in each bunch crossing is around 20. The points at which the proton collisions occur, are known as "vertices" [23]. In the proton-proton collisions, a collision with a large momentum transfer rarely occurs. Among all collisions the one with the highest energy collision identified by largest sum of particle p_{T} is considered as the "signal vertex" whereas, the remaining low energy collisions are referred to as "pileup".

In order to measure the location, and the associated uncertainty, of all vertices, primary-vertex reconstruction using the available reconstructed tracks, [32] is performed. The primary-vertex reconstruction consists of three steps [23]:

- Selection of the tracks
- Clustering of the tracks that appear to originate from the same interaction vertex
- Fitting for the position of each vertex using its associated tracks

Moreover, the primary vertex is supposed to satisfy $|z| < 24$ cm and $|\rho| < 2$ cm in cylindrical geometry. It is also required to be marked as valid and not fake vertex fits by the vertex reconstruction algorithm.

In the simulation, the pileup profile is reweighted based on the instantaneous luminosity per bunch crossing per luminosity section as a function of the number of true pileup vertices. Because the pileup profile may affect object performance, most notably the multiplicity of jets and the resolution of missing transverse momentum (discussed in the section 3.7), we derive the reweighting factors separately for each data era. The reweighting is done in a correlated way with other run-dependent considerations. The inelastic pp collision cross section is taken to be 69200 mb with a 2.6% uncertainty [33–35].

3.3 Muons

3.3.1 Muon reconstruction

Muon reconstruction is based on the information from both the inner tracker and the muon system which is implemented in 3 stages:

- local reconstruction
- standalone reconstruction
- global reconstruction

In local reconstruction, input information comes from only a single muon station (i.e. RPC, CSC, or DT) to specify the trajectory of a muon through the chamber [36]. The muons or any other charged particles traversing a muon sub-detector ionize the gas in the chambers, which eventually result in electric signals production. These signals are read out by electronics and have a well-defined locations called "hits", in the detector. Depending on the detector technology, different algorithms are used to reconstruct the precise location of each hit from the electronic signals. In standalone muon reconstruction, only data from muon chamber detectors are used. In the other word, no input from the silicon tracker is used in standalone muon reconstruction.

To reconstruct the standalone muons, data from individual chambers are combined to build the muons tracks using a Kalman filter (KF) [37]. Finally, at the global reconstruction level, the muon trajectory is extrapolated inwards towards the tracker, accounting for energy loss and multiple scattering, and it is then combined with the hits measured in the Silicon tracker in order to reconstruct so-called global muons. Moreover, the tracker muons, which are tracks reconstructed in the tracker, that are then extrapolated outwards to the muon system and matched to at least one local muon segment, are taken into account for global muon reconstruction.

In addition, the muon momentum is corrected in both real and simulated data according to the Rochester calibration method using $Z \to \mu\mu$ events [38].

3.3.2 Muon identification

In this analysis, the baseline muon identification criteria which corresponds to the cut-based "medium" identification requirements are used [39]. The list of requirements for muon identification are outlined below,

- The muon must be a PF muon, and also must be either a tracker or global muon.
- The fraction of valid inner tracker hits must be greater than 0.8.
- If the muon has to be a global muon, the global fit must have a χ^2 per degree of freedom less than 3, the position match between the tracker muon and standalone muon must have $\chi^2 < 12$, and the maximum χ^2 computed by the kink-finding algorithm must be less than 20, the muon segment compatibility should be greater than 0.303. Otherwise, the segment compatibility should be greater than 0.451.
- The best track of the muon should satisfy the longitudinal closest approach requirement $|d_z| < 0.1$, and the transverse closest approach requirement $|d_{xy}| < 0.02$. The distance values are computed with respect to the primary vertex of the event.

Furthermore, muons, that are produced as a result of hadrons decays, must be suppressed. Therefore the muons are also required to be isolated. The isolation of the muon candidates is estimated from the flux of particle-flow candidates found nearby the muon direction within a cone of $\Delta R < 0.3$. The flux of particles is calculated independently for the charged hadrons ($\mathcal{I}_{\text{ch}}$), neutral hadrons ($\mathcal{I}_{\text{nh}}$), and photon candidates ($\mathcal{I}_{\gamma}$). Therefore, the muon isolation is defined as

$$\mathcal{I}^{\mu}_{\text{rel}} = \frac{\mathcal{I}_{\text{ch}} + \max(\mathcal{I}_{\text{nh}} + \mathcal{I}_{\gamma} - 0.5 \times \mathcal{I}^{\text{PU}}_{\text{ch}}, 0)}{p^{\mu}_{\text{T}}}, \tag{3.1}$$

where the p^{μ}_{T} in the denominator is the muon traverse momentum after momentum corrections. The 0.5 factor behind the $\mathcal{I}^{\text{PU}}_{\text{ch}}$ is a correction to the flux of neutral hadrons $\mathcal{I}_{\text{nh}}$ due to pileup. For this purpose, half of the scalar p_{T} sum over the charged particles within the cone of interest but not originating from the primary vertex ($\mathcal{I}^{\text{PU}}_{\text{ch}}$), is subtracted. In this analysis, muons are required to satisfy $\mathcal{I}^{\mu}_{\text{rel}} < 0.15$, $p^{\mu}_{\text{T}} \geq 5$ GeV , and $|\eta| < 2.4$ in addition to the identification conditions.

3.4 Electrons

3.4.1 Electron reconstruction

In order to reconstruct electrons, the hit measurements obtained from the inner tracking system and the clusters of energy depositions in the ECAL are associated. Bremsstrahlung radiation, which electrons experience while passing through the material of the tracker, is the major challenge of the electron reconstruction process. Due to the bremsstrahlung effect, electrons radiate from about 30% to 90% of their total energy in average, depending on the pseudorapidity, before reaching the ECAL [40]. To ensure a successful reconstruction, a particular algorithm has been implemented [41] considering the bremsstrahlung effects. Similar to muons, the electron tracks are reconstructed from hits in the tracker by a KF algorithm. Additionally, a specific tracking technique for electrons is utilized to estimate the track parameters. This approach is based on the Gaussian sum filter (GSF) [42]. The initial energy of the electrons are accurately measured by collecting the energy of the radiated photons mainly spreading along the azimuthal angle (ϕ) direction. In different parts of the ECAL, barrel and endcaps, different algorithms, "hybrid" and "multi-5 $\times$ 5" are implemented respectively, for energy clustering [40].

3.4.2 Electron identification

In order to identify the reconstructed electrons, an XGBoost [43] boosted decision tree (BDT) algorithm is used. In this algorithm, observables from the electromagnetic clusters, electron tracking, and track-cluster matching are utilized to determine prompt electrons.

The full list of used observables in the training of this BDT can be found in Table 3.1. The working points defined as the threshold value on the output BDT score is adjusted for 6 categories of p_{T}^{ℓ} and electron supercluster pseudorapidity (η_{SC}) ranges, summarized in Table 3.2. The same working points are used for all three years of data used in the present analysis. These identification BDT score working points correspond to the "Fall17 (no iso.) WP90" criteria provided by the Egamma Physics Object Group (POG).

The concept of electron isolation is similar to that of muon isolation defined in Eq. 3.1, also with the same cone radius used for muons. However, in the electron case, the neutral hadron flux $\mathcal{I}_{\mathrm{nh}}$ is corrected by using the average energy density (ρ) due to pileup and underlying event in the central region of the detector, accounting for an effective area of the electron isolation cone correction ($A_{\mathrm{eff}}^{\mathrm{e}}$) to normalize this estimator. This

Table 3.1: The observables used in the training of the BDT in order to identify prompt electrons. The training is done using the 2017 simulation, and this same training is used for 2016 and 2018.

Cluster variables
RMS of the energy-crystal numbering along η and ϕ, $\sigma_{\mathrm{i}\eta\mathrm{i}\eta}$ and $\sigma_{\mathrm{i}\phi\mathrm{i}\phi}$
Supercluster width along η and ϕ
Ratio of the hadronic energy behind the electron supercluster to the supercluster energy, H/E
Circularity, $(E_{5\times5} - E_{5\times1})/E_{5\times5}$
Sum of the seed and adjacent crystal over the supercluster energy, R_9
For endcap electrons: Energy fraction in preshower, $E_{\mathrm{PS}}/E_{\mathrm{raw}}$
Tracking variables
Fractional momentum loss, $f_{brem} = 1 - p_{\mathrm{out}}/p_{\mathrm{in}}$
Number of hits of the KF and GSF tracks
Reduced χ^2 of the KF and GSF tracks
Number of expected but missing inner hits
Probability transform of conversion vertex fit χ^2
Track-cluster matching variables
Energy-momentum agreement: $E_{\mathrm{tot}}/p_{\mathrm{in}}$, $E_{\mathrm{e}}\ /p_{\mathrm{out}}$, $1/E_{\mathrm{tot}} - 1/p_{\mathrm{in}}$
Position matching: $\Delta\eta_{\mathrm{in}}$, $\Delta\varphi_{\mathrm{in}}$, $\Delta\eta_{\mathrm{seed}}$

Table 3.2: The working points of the electron identification BDT, defined as a function of the p_{T}^{ℓ} before any residual energy scale and smear corrections. The same working points are used in all three years.

p_{T}^{ℓ} range (GeV)	$\lvert\eta_{\mathrm{SC}}\rvert$ range	Working point definition
< 10	< 0.800	$2.771 - \exp\left(-p_{\mathrm{T}}^{\ell}/3.815\right) \times 8.163$
	$[0.800, 1.479)$	$1.856 - \exp\left(-p_{\mathrm{T}}^{\ell}/2.187\right) \times 11.856$
	≥ 1.479	$1.735 - \exp\left(-p_{\mathrm{T}}^{\ell}/2.016\right) \times 17.014$
≥ 10	< 0.800	$5.918 - \exp\left(-p_{\mathrm{T}}^{\ell}/13.481\right) \times 9.320$
	$[0.800, 1.479)$	$5.016 - \exp\left(-p_{\mathrm{T}}^{\ell}/13.128\right) \times 8.794$
	≥ 1.479	$4.169 - \exp\left(-p_{\mathrm{T}}^{\ell}/13.202\right) \times 9.007$

correction is performed in such a way that the isolation is independent of the number of pileup interactions. The values of $A_{\text{eff}}^{\text{e}}$ depend the $|\eta_{\text{SC}}|$ range which are listed in Table 3.3. Therefore, using these quantities, the electron isolation is defined as

$$\mathcal{I}_{\text{rel}}^{\text{e}} = \frac{\mathcal{I}_{\text{ch}} + \max(\mathcal{I}_{\text{nh}} + \mathcal{I}_{\gamma} - A_{\text{eff}}^{\text{e}} \times \rho, 0)}{p_{\text{T}}^{\ell}}, \tag{3.2}$$

where the p_{T}^{ℓ} in the denominator is the electron transverse momentum after electron energy corrections. The electrons used in this analysis are required to satisfy $\mathcal{I}_{\text{rel}}^{\text{e}} < 0.1$, $p_{\text{T}}^{\ell} \geq 5$ GeV , and $|\eta| < 2.5$.

Table 3.3: The effective area $A_{\text{eff}}^{\text{e}}$ values used in each $|\eta_{\text{SC}}|$ range to mitigate the dependence of the isolation requirement on pileup. The same values are used in all three years.

$\lvert\eta_{\text{SC}}\rvert$ range	$A_{\text{eff}}^{\text{e}}$
< 1	0.1440
$[1, 1.479)$	0.1562
$[1.479, 2)$	0.1032
$[2, 2.2)$	0.0859
$[2.2, 2.3)$	0.1116
$[2.3, 2.4)$	0.1321
≥ 2.4	0.1654

3.5 Photons

3.5.1 Photon reconstruction

Since photons are neutral, they either traverse the tracker without interaction, or convert into an electron-positron pair in the tracker material. They are caught by ECAL and deposit their energy there. Therefore, the ECAL energy deposits need to be collected to reconstruct photons employing techniques that constrain the clusters to the size and shape expected for electrons and photons with $p_{\text{T}} > 15$ GeV [44]. Several stages are taken in the photon reconstruction,

- **Clustering**: Due to the magnetic field of CMS, the radiating electrons and converted photons are spread in the Φ direction. Hence, in order to perform the energy clustering, some algorithms are developed [40] to collect the energy depositions from these electrons and photons.

- **Correction of cluster energy**: Correction of the initial integrated energy deposits forming the supercluster, has significant impact on improvements of energy resolution. A first correction to the photon energy scale is computed from simulation using a regression multivariate analysis, the inputs of which are variables describing the energy spread in the ECAL, data from the preshower detector, and pile-up sensitive observables. Then, scale factor corrections as a function of photon eta, ECAL energy spread and p_{T} are derived from data and simulations, using $\mathrm{Z} \rightarrow \mathrm{e}^{+}\mathrm{e}^{-}$ events where the electron signals in the calorimeter are reconstructed with the photon algorithms. These corrections are also validated using photons from final state radiation in dimuon decays of Z bosons [45].

3.5.2 Photon selection

Selection of the reconstructed photons uses a cut-based selection flow. The baseline requirements are listed in Table 3.4 and are kept the same for all three years. The isolation requirements are different for the charged and neutral particle-flow hadron, or the particle-flow photon fluxes within a cone of $\Delta R < 0.3$. In order to mitigate the dependence of the selection efficiency on pileup, various independent effective area corrections $A^{\gamma}_{\text{eff,ch}}$, $A^{\gamma}_{\text{eff,nh}}$, and $A^{\gamma}_{\text{eff},\gamma}$ are applied for the charged and neutral particle-flow hadron, or the particle-flow photon fluxes respectively. Values of effective areas are outlined in Table 3.5 which are same for the three years, compatible with the requirements on the baseline selection. These baseline selection requirements correspond to the ‘Fall17 tight’ cut-based selection criteria provided by the Egamma POG.

Moreover, to enrich the purity of photons in the single-photon CR which will be discussed in Sec. 5.6.1, a few additional selection requirements are applied in a similar way as in Ref. [46]. These additional requirements are outlined in the following,

- Particle-flow photon requirement:

- Pixel seed and electron vetoes: These two vetoes reduce contamination from $\mathrm{e} \rightarrow \gamma$ fake photons.

- $\sigma_{\mathrm{i}\eta\mathrm{i}\eta} > 0.001$ and $\sigma_{\mathrm{i}\phi\mathrm{i}\phi} > 0.001$: These requirements remove the ECAL noise in real data characterized as a single-cell spike. These upper values are determined according to ones described in section 5.6 of Ref. [47].

Table 3.4: The baseline selection requirements on the photons are listed. The requirements are kept the same among the three data taking periods.

Requirement	Value for $\|\eta_{\mathrm{SC}}\| < 1.479$	Value for $\|\eta_{\mathrm{SC}}\| \geq 1.479$
$H/E <$	0.02148	0.0321
$\sigma_{\mathrm{i}\eta\mathrm{i}\eta} <$	0.00996	0.0271
$\mathcal{I}_{\mathrm{ch}} <$	0.65	0.517
$\mathcal{I}_{\mathrm{nh}} <$	$0.317 + 0.01512 \times p_{\mathrm{T}}^{\gamma} + 2.259 \cdot 10^{-5} \times p_{\mathrm{T}}^{\gamma\,2}$	$2.716 + 0.0117 \times p_{\mathrm{T}}^{\gamma} + 2.3 \cdot 10^{-5} \times p_{\mathrm{T}}^{\gamma\,2}$
$\mathcal{I}_{\gamma} <$	$2.044 + 0.004017 \times p_{\mathrm{T}}^{\gamma}$	$3.032 + 0.0037 \times p_{\mathrm{T}}^{\gamma}$

Table 3.5: The values of the effective areas $A^{\gamma}_{\mathrm{eff,ch}}$, $A^{\gamma}_{\mathrm{eff,nh}}$, and $A^{\gamma}_{\mathrm{eff},\gamma}$ used in each $|\eta_{\mathrm{SC}}|$ range to mitigate the dependence of the isolation requirement on pileup. The same values are used in all three years.

$\|\eta_{\mathrm{SC}}\|$ range	$A^{\gamma}_{\mathrm{eff,ch}}$	$A^{\gamma}_{\mathrm{eff,nh}}$	$A^{\gamma}_{\mathrm{eff},\gamma}$
< 1	0.0112	0.0668	0.1113
$[1, 1.479)$	0.0108	0.1054	0.0953
$[1.479, 2)$	0.0106	0.0786	0.0619
$[2, 2.2)$	0.01002	0.0233	0.0837
$[2.2, 2.3)$	0.0098	0.0078	0.1070
$[2.3, 2.4)$	0.0089	0.0028	0.1212
≥ 2.4	0.0087	0.0137	0.1466

- $E_{\mathrm{MIP}} < 4.9$ GeV : This quantity is the total minimum-ionizing particle energy in the photon cluster, deposited in the ECAL by a beam halo muon that might leave a trail of low-energy clusters along its trajectory. The upper value for E_{MIP} is chosen as described in section 5.6 of Ref. [47].
- $|t_{\mathrm{seed}}| < 2\,\mathrm{ns}$ ($-2\,\mathrm{ns} < t_{\mathrm{seed}} < 1\,\mathrm{ns}$ in 2018): This is a further rejection of electromagnetic showers coming from a beam halo muon with slightly earlier hits than the prompt collision products by comparing the timing readout of the ECAL seed relative to the estimated collision time.

3.6 Jets

3.6.1 Jet reconstruction

The direct observation of particles like quarks and gluons is not possible. However, as a result of the parton shower and hadronization processes, numerous particles are produced, and quarks and gluons manifest themselves as collimated sprays of hadrons, which are known as jets. Jets are the experimental evidence of quarks and gluons produced in high energy collisions like proton-proton collisions at CMS.

Jets are reconstructed from PF candidates after removing charged hadrons identified by the "charged hadron subtraction" (CHS) procedure as being due to pile-up interactions [48], and loose muons and electrons. Afterward, the remaining particles are clustered using the anti-k_T algorithm [49–51] with a distance parameter of 0.4.

3.6.2 Jet identification and corrections

In this analysis jets are required to satisfy $p_{\mathrm{T}} \geq 30$ GeV and $|\eta| < 4.7$, unless otherwise specified. They also must be separated from all leptons and photons that pass the selection criteria of this analysis. For this purpose, the distance parameter $\Delta R > 0.4$, where $(\Delta R)^2 = (\Delta\phi)^2 + (\Delta\eta)^2$ is the distance between the two objects in the $\eta - \phi$ plane. A multivariate technique is used to suppress jets from pileup interactions [33, 52], and the technique also helps reduce detector noise. The tight JetMET POG working point is applied for this pileup jet identification requirement [53].

The energy of the jets are corrected for the calorimetric energy scale (JES corrections), and the p_{T} of the simulated jets are smeared further for the differences in resolution between the real and simulated data (JER corrections). Furthermore, due to malfunctioning of the HEM 15/16 detectors in the period of data taking in 2018, the events in runs number after and including 319077 in 2018, are rejected if they contain a jet inside the span of the HEM 15/16 detectors, i.e. $-3 < \eta < -1.4$, $-1.6 < \phi < -0.8$ in the $\eta - \phi$ plane.

3.6.3 b-tagging

Jets are tagged as b-tagged jets using the DeepJet algorithm [54], which provides performance improvements over the DeepCSV algorithm [55, 56] by using approximately 650 input variables related to PF candidates, vertexing and jet constituents, and improved neural network training. The b-tagging can be considered for all jets with $|\eta| < 2.5$ ($|\eta| < 2.4$ in 2016 due to different tracker geometry). The loose (tight) working point is defined in this analysis based on the "loose" ("medium") working point prescription of the "JetMET" POG in order to veto (accept) events with b-tagged jets in the signal region (the control region for the nonresonant background estimation).

3.7 Missing Transverse Momentum

In principle the energy and momentum is conserved in pp collision, however, some particles may escape CMS detectors due to their weak interactions such as neutrinos and possible BSM particles, or as a matter of detectors inefficiency. Therefore, almost in all events there is a potential imbalance in energy and momentum. The imbalance in energy and momentum can be estimated in transverse plane as following,

$$\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\sum_{i} \vec{p}_{\mathrm{T}}(i), \tag{3.3}$$

where the vector $\vec{p}_{\mathrm{T}}(i)$ is the transverse momentum of PF candidate i and the index i runs over all PF candidates including the ones identified by the CHS procedure. In order to better approximate the true missing momentum, multiplicative corrections are applied to this quantity [57, 58]. One of the important corrections is the JES correction applied on jets (as discussed in the Section 3.6.2) that is propagated to $\vec{p}_{\mathrm{T}}^{\,miss}$ which is called a "Type-1" correction. Despite of the JES correction, the JER correction in simulation are not propagated to $p_{\mathrm{T}}^{\mathrm{miss}}$ as it is found to over-smear the $p_{\mathrm{T}}^{\mathrm{miss}}$. Moreover, $p_{\mathrm{T}}^{\mathrm{miss}}$ filters are applied during event selection in both real and simulated data to reject the events containing potential fake $p_{\mathrm{T}}^{\mathrm{miss}}$. The list of these filters are outlined for each year below which are described in the Section 4.3.1,

- **2016**: goodVertices, HBHENoiseFilter, HBHENoiseIsoFilter, EcalDeadCellTriggerPrimitiveFilter, globalSuperTightHalo2016Filter, BadPFMuonFilter (common); eeBadScFilter (real data only)
- **2017, 2018**: goodVertices, HBHENoiseFilter, HBHENoiseIsoFilter, EcalDeadCellTriggerPrimitiveFilter, globalSuperTightHalo2016Filter, BadPFMuonFilter, ecalBadCalibFilter (common); eeBadScFilter (real data only).

In the next chapter (Chapter 4), a specific focus is dedicated to $p_{\mathrm{T}}^{\mathrm{miss}}$ in which corrections, the $p_{\mathrm{T}}^{\mathrm{miss}}$ filters and the performance of missing transverse momentum reconstruction, will be discussed in detail.

Chapter 4

Missing Transverse Momentum Reconstruction and Scanning

At the LHC, proton-proton (pp) collisions can produce weakly interacting neutral particles that go unobserved as they pass through the collider detectors. However, the presence of these particles can be inferred when they are produced alongside strong or electromagnetically interacting particles, due to the resulting imbalance in momentum measured in the plane perpendicular to the beam direction, known as missing transverse momentum ($\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$).

The $p_{\mathrm{T}}^{\mathrm{miss}}$ is a signature that is very specific to the process that we are studying. Instrumental MET from the Z+jet process is the main source of experimental uncertainty in our analysis (see section 5.8). Large $p_{\mathrm{T}}^{\mathrm{miss}}$ (> 125 GeV) is required to suppress the Z+jet background, which reduces the acceptance for off-shell $\mathrm{H} \to \mathrm{ZZ} \to 2\ell 2\nu$ events (see section 5.3). The measurement of the remaining instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$ background is difficult, and is affected by large uncertainties (see section 5.6). In addition, the author of this thesis contributed to the study of event filters deployed by the CMS collaboration to suppress the contribution of events with anomalous $p_{\mathrm{T}}^{\mathrm{miss}}$. For these reasons, the reconstruction and performance of CMS for $p_{\mathrm{T}}^{\mathrm{miss}}$ is detailed in this chapter.

The magnitude of this missing transverse momentum is given by $p_{\mathrm{T}}^{\mathrm{miss}}$. The precise determination of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is essential for standard model measurements that involve final states with neutrinos, such as those featuring leptonic decays of the W and Z boson. In addition, $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is an important observable in searches for physics beyond the standard model that aim to discover new weakly interacting particles. The $p_{\mathrm{T}}^{\mathrm{miss}}$ arising from weakly interacting particles will be referred to as 'genuine $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ ' in the following. However, reconstructing $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ is prone to experimental resolutions, mismeasurement of reconstructed particles, and detector artifacts.

4.1 Missing Transverse Momentum Reconstruction and Calibration

The CMS event reconstruction utilizes two different $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction algorithms, the first is known as PF $p_{\mathrm{T}}^{\mathrm{miss}}$ and the second one is "pileup per particle identification" (PUPPI) $p_{\mathrm{T}}^{\mathrm{miss}}$. Although in this chapter the PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ is also a part of the study, in our analysis we do not use PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ as in the versions of PF and PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ available by the time of our analysis, the tails of instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$ were slightly larger for PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ than for PF $p_{\mathrm{T}}^{\mathrm{miss}}$.

4.1.1 The PF and PUPPI estimators of $\mathbf{p}_{\mathrm{T}}^{\mathrm{miss}}$

The PF $p_{\mathrm{T}}^{\mathrm{miss}}$ is based on a complete event interpretation by the particle-flow technique [59]. It is defined as the negative vector p_{T} sum of all PF candidates recorded by CMS in a particular triggered event, as depicted in Eq. (3.3) [57, 58].

To address the dependence of reconstructed $p_{\mathrm{T}}^{\mathrm{miss}}$ on pileup, the second algorithm has been designed which utilizes the PUPPI method [60]. This algorithm incorporates spatial p_{T} distributions of the particle p_{T} around each PF candidate in the event as well as pileup properties within the event, and tracking data to alleviate the pileup dependence of various jet and $p_{\mathrm{T}}^{\mathrm{miss}}$ variables.

More precisely, PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ uses a variable called α to distinguish between the collinear structure of QCD and the soft, diffuse radiation present in pileup. This α variable is computed for each particle in the event, and a weight is derived for each particle from the comparison of its α value to the distribution of α computed from the nearby particles that are associated to pileup events.

The α variable for a given particle i is defined as [60],

$$\alpha_i = \log \sum_{j \neq i} \left(\frac{p_{\mathrm{T},j}}{\Delta R_{ij}} \right)^2 \Theta(R_0 - \Delta R_{ij}) \tag{4.1}$$

where Θ is the step function, i refers to the candidate particle and j to the neighboring particles inside a cone of radius R_0 which is set to 0.4. The ΔR_{ij} is the distance in

$\eta - \phi$ space between the i and j particles. For $|\eta_i| < 2.5$, j corresponds to any charged PF candidate from the primary vertex of the hard event, whereas, for $|\eta_i| > 2.5$, j runs over all kinds of reconstructed PF candidates including charged and neutral ones. In addition, charged PF candidates which are not associated with the primary vertex, are taken into account if they satisfy $d_z < 0.3$ cm, where d_z is the distance in z between the track and the primary vertex.

The PUPPI algorithm uses a measure called the χ^2 approximation to estimate the likelihood that a particular PF candidate is associated with pileup. The χ^2 approximation reads as,

$$\chi_i^2 = \frac{(\alpha_i - \bar{\alpha}_{PU})^2}{\mathrm{RMS}_{PU}^2}, \tag{4.2}$$

where the variable $\bar{\alpha}_{PU}$ represents the median value of the alpha variable for all PF candidates. The RMS_{PU} variable is the root-mean-square (RMS) of the alpha variable for these pileup-associated PF candidates. In the tracker region, $\bar{\alpha}_{PU}$ and RMS_{PU} are calculated using all charged pileup particles (i.e. all charged particles associated to the pileup vertices) while in the forward region they are calculated using all the particles in the event. Particles are then assigned a weight given by

$$w_i = F_{\chi^2,\mathrm{NDF}=1}\left(\chi_i^2\right), \tag{4.3}$$

where F_{χ^2} is the cumulative distribution function of the χ^2 distribution. Eventually, the momentum of each particle is multiplied by w_i when computing PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$, and particles with either a very small weight, or a very small weighted transverse momentum are not considered in the computation.

4.1.2 $p_{\mathrm{T}}^{\mathrm{miss}}$ Corrections and Uncertainties

Type-1 correction

There are several factors that can contribute to inaccurate estimation of the $p_{\mathrm{T}}^{\mathrm{miss}}$, in the CMS experiment. These include non-linearities in the response of the calorimeters to hadronic particles, energy thresholds in the calorimeters, and inefficiencies in the tracker. One way to improve the estimation of $p_{\mathrm{T}}^{\mathrm{miss}}$ is to correct the transverse momentum of the jets to the particle level using jet energy corrections, as described in reference

[61]. These corrections can then be propagated to $p_{\mathrm{T}}^{\mathrm{miss}}$ in the following way,

$$\text{Type-I } p_{\mathrm{T}}^{\mathrm{miss}} = p_{\mathrm{T}}^{\mathrm{miss}} - \sum_{\text{jets}} \left(\vec{p}_{\mathrm{T,jet}}^{\,\mathrm{corr}} - \vec{p}_{\mathrm{T,jet}} \right) \tag{4.4}$$

where the term "Type-I $p_{\mathrm{T}}^{\mathrm{miss}}$ " refers to values of the $p_{\mathrm{T}}^{\mathrm{miss}}$ corrected to account for the corrections to the energy of the jets reconstructed in the event. All of the results presented further in the text, use the corrected values of $p_{\mathrm{T}}^{\mathrm{miss}}$ and therefore the prefix "Type-I" is omitted in order to improve clarity.

In order to accurately correct the $p_{\mathrm{T}}^{\mathrm{miss}}$, it is necessary to carefully select the transverse momentum thresholds of the jets used in the calculation. The thresholds should be chosen in such a way as to minimize the contributions of jets coming from pileup, while still maintaining a good overall scale for $p_{\mathrm{T}}^{\mathrm{miss}}$. To achieve this balance, the $p_{\mathrm{T}}^{\mathrm{miss}}$ algorithms considers all jets with corrected transverse momentum above 15 GeV in the $p_{\mathrm{T}}^{\mathrm{miss}}$ correction. Additionally, jets that are matched to electrons, photons, or muons are handled separately in the calculation of corrected $p_{\mathrm{T}}^{\mathrm{miss}}$. To remove overlap between jets and electrons or photons, jets with more than 90% of their energy deposited in the ECAL are excluded from the $p_{\mathrm{T}}^{\mathrm{miss}}$ correction. If a muon reconstructed using both the inner and outer tracking system or only the outer tracking system overlaps with a jet, the algorithm subtracts the muon's 4-momentum from the jet's 4-momentum and uses a modified jet energy scale correction in the $p_{\mathrm{T}}^{\mathrm{miss}}$ correction. These steps help to ensure that the corrected $p_{\mathrm{T}}^{\mathrm{miss}}$ values accurately reflect the true missing transverse momentum in the event.

ϕ correction

In the LHC collisions, particles are produced uniformly in the azimuthal angle ϕ, due to the rotational symmetry of the collisions around the beam axis. This leads to the expectation that the $p_{\mathrm{T}}^{\mathrm{miss}}$ quantity should be independent of ϕ. However, an asymmetry in ϕ has been observed in the sums of transverse momenta ($\vec{p}_{\mathrm{T}}$) of calorimeter energy deposits, tracks, and reconstructed particles. This asymmetry also results in an asymmetry in $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ that exhibits a sinusoidal shape with a period of 2π. This modulation is present in both Monte Carlo simulated events and collision data and may be caused by inactive calorimeter cells, anisotropic detector response leading to some ϕ-dependence, and possible misalignment or displacements of the detector or beam spot. It has also been observed that the amplitude of the modulation approximately grows linearly with the number of pileup interactions. This asymmetry can be represented as a shift in the x and y components of $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ i.e. $\not{E}_x$ and $\not{E}_y$, respectively, which increases roughly linearly with the multiplicity of PF candidates. Separate corrections are proposed for simulated and data events in order to account for this modulation in $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$.

Some functions are used to describe the relationship between $\not{E}_x$ and $\not{E}_y$ and the number of PF candidates in various η bins. These functions are obtained by fitting a line to the correlation between these variables. Then the correction on $\vec{p}_{\mathrm{T}}^{\mathrm{miss}}$ components are performed as,

$$\not{E}_x^{\mathrm{corr}} = \not{E}_x - \langle \not{E}_x \rangle = \not{E}_x - \left(c_{x_0} \cdot n + c_{x_s} \cdot n^2\right) \tag{4.5}$$

$$\not{E}_y^{\mathrm{corr}} = \not{E}_y - \langle \not{E}_y \rangle = \not{E}_y - \left(c_{y_0} \cdot n + c_{y_s} \cdot n^2\right), \tag{4.6}$$

where the n indicates the multiplicity of PF candidates and coefficients c_{x_0}, c_{x_s}, c_{y_0}, and c_{y_s} are obtained separately from Drell-Yan (DY) + jet, W + jet and $\mathrm{t\bar{t}}$ + jet candidate events in both data and simulation samples.

Uncertainties

Since the $p_{\mathrm{T}}^{\mathrm{miss}}$ is not an observable that is directly measured, its accuracy strongly relies on the accurate measurement of other reconstructed physics objects in the event. Therefore, the uncertainties of $p_{\mathrm{T}}^{\mathrm{miss}}$ measurement are highly dependent on the topology of the event. The typical uncertainties are outlined below,

- Jet energy scale uncertainty; 1-12%
- Jet energy resolution uncertainty; 1-7%
- Muon energy scale uncertainty; 0.2%
- Electron and photon energy scale uncertainties; 0.6% in the barrel and 1.5% in the endcap
- Tau energy scale uncertainty; 1.2%
- Unclustered energy uncertainty

where the "unclustered energy" refers to the total energy of the PF candidates that are not included in any of the physics objects described in Chapter 3, nor in reconstructed hadronic tau decays.

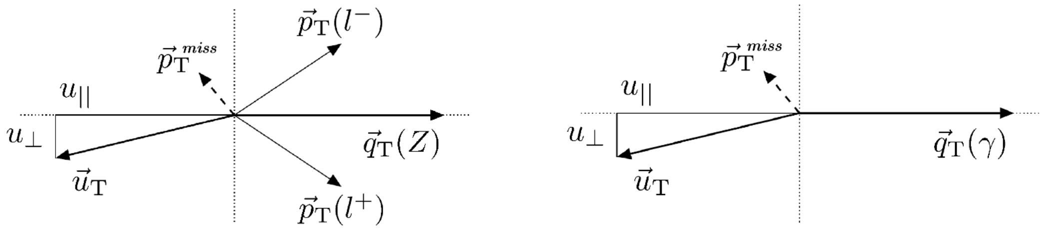

Figure 4.1: The kinematics of Z boson (left) and photon (right) events in the transverse plane. The $\vec{u}_{\mathrm{T}}$ represents the sum of all particles reconstructed in the event except for the two leptons from the Z decay (left), or the photon (right) [57].

4.2 Performance of $p_{\mathrm{T}}^{\mathrm{miss}}$ Algorithms

Samples with a well-measured isolated photon, or Z boson decaying to two electrons or muons can be used to investigate the response and resolution of $p_{\mathrm{T}}^{\mathrm{miss}}$. These events have little or no genuine $p_{\mathrm{T}}^{\mathrm{miss}}$ and can be selected with little background. The performance of $p_{\mathrm{T}}^{\mathrm{miss}}$ can be evaluated by comparing the momenta of the vector boson or photon to that of the hadronic recoil system, which is defined as the vector p_{T} sum of all PF candidates except for the vector boson or photon and its decay products. Any differences between the measured momenta can be attributed to the performance of $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction. The transverse momentum of the vector boson and the hadronic recoil are denoted as $\vec{q}_{\mathrm{T}}$ and $\vec{u}_{\mathrm{T}}$ respectively. Fig. 4.1 shows the kinematic illustration of $\vec{q}_{\mathrm{T}}$ and $\vec{u}_{\mathrm{T}}$.

The hadronic recoil can be broken down into the components parallel ($u_{\|}$) and perpendicular ($u_{\perp}$) to the boson transverse momentum. These are used to evaluate the performance of $p_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction. The root mean square (RMS) of the $u_{\|}+q_{\mathrm{T}}$ and $u_{\perp}$ distributions (denoted by $\sigma(u_{\|})$ and $\sigma(u_{\perp})$, respectively) are used to assess the resolution of $u_{\|}$ and $u_{\perp}$. The response of $p_{\mathrm{T}}^{\mathrm{miss}}$ is calculated as $-\frac{\langle u_{\|}\rangle}{\langle q_{\mathrm{T}}\rangle}$, where $\langle\rangle$ represents the mean of the distributions. Fig. 4.2 represents the plots of The PF and PUPPI responses, and also the PF resolutions as a function of q_{T} in the Z $\rightarrow \mu^+\mu^-$, Z $\rightarrow e^+e^-$, and the $\gamma+$jets events. The PF $p_{\mathrm{T}}^{\mathrm{miss}}$ has at better response at lower q_{T} values compare to PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$. In the analysis described in this thesis, we use PF $p_{\mathrm{T}}^{\mathrm{miss}}$ rather than PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ as it hass less tails in the EOY (end-of-year) datasets.

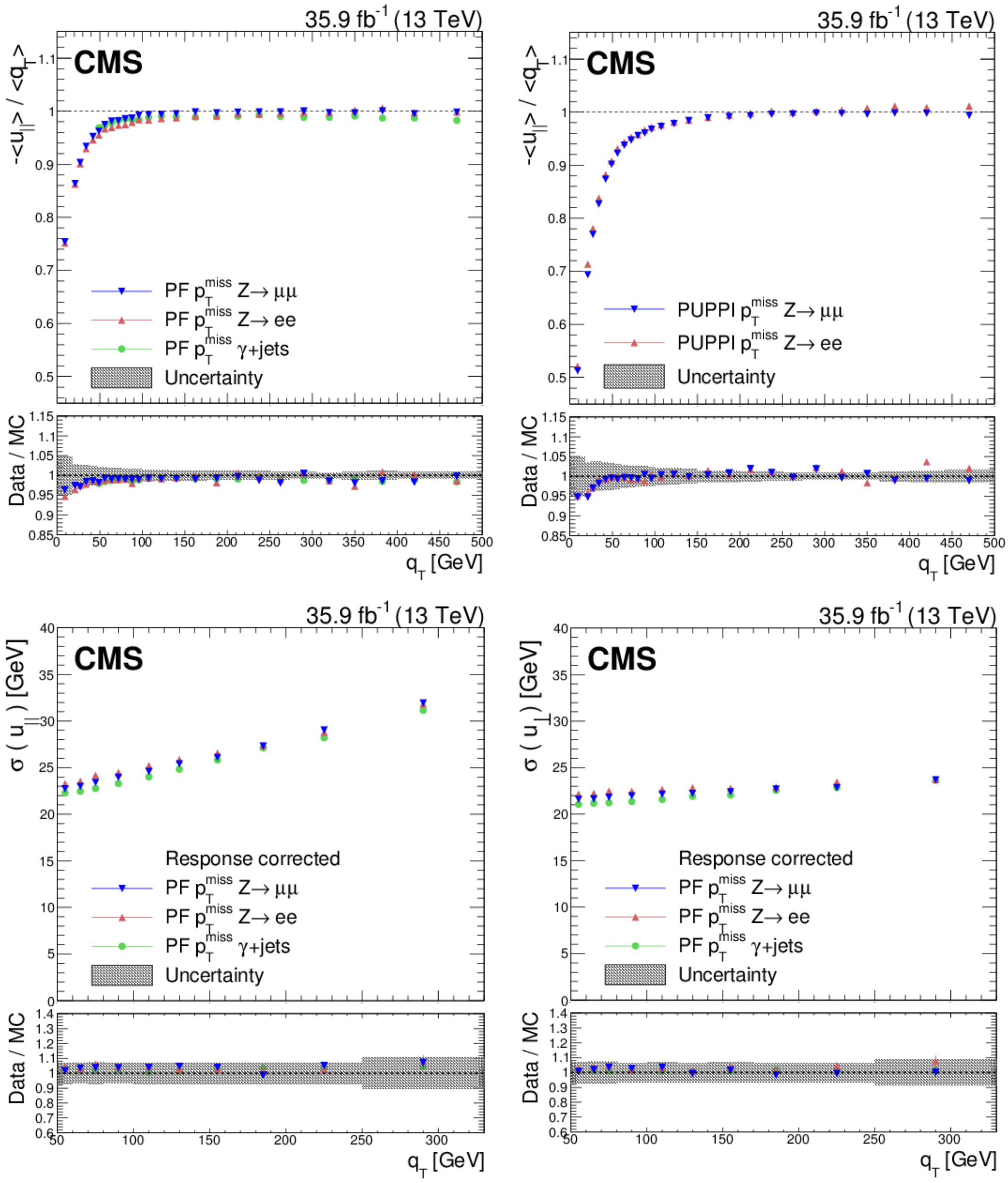

Figure 4.2: The PF (upper left) and PUPPI (upper right) $p_{\mathrm{T}}^{\mathrm{miss}}$ responses i.e. $-\dfrac{\langle u_{\parallel}\rangle}{\langle q_{\mathrm{T}}\rangle}$ are plotted. The resolution of the $u_{\parallel}$ (lower left) and $u_{\perp}$ (lower right) components of the hadronic recoil as a function of q_{T} are presented. In each plot, the blue, red, and green points refer to data in the $\mathrm{Z} \to \mu^+\mu^-$, $\mathrm{Z} \to e^+e^-$, and $\gamma+$jets events. The lower panel of each plot indicates the ratio of the $p_{\mathrm{T}}^{\mathrm{miss}}$ response in data and simulation. The systematic uncertainties (estimated from the $\mathrm{Z} \to e^+e^-$ sample) are represented as a gray band in the ratio plot [57].

4.3 Anomalous $p_{\mathrm{T}}^{\mathrm{miss}}$ Events

There are several reasons why anomalous high $p_{\mathrm{T}}^{\mathrm{miss}}$ events might occur, such as problems with the event reconstruction or detectors, issues with energy estimation due to faulty sensors or dead cells, or background events like muon production during beam collisions (beam halo). Typical problems are, in the ECAL, dead cells, noisy photodetectors or beam halo muons producing signals in the photodetectors, in the HCAL, noise in the hybrid photodiodes (HPD) and readout box (RBX) electronics, and in the HF, direct particle interactions with the light guides and photomultiplier tubes. To address these issues, various filters known as “MET filters”, were developed and implemented during LHC Run 1 [58, 62] and modified for use in LHC Run 2 [63] in order to take into account the upgraded detectors and different data-taking conditions. These filters aim to identify and eliminate events with spurious high $p_{\mathrm{T}}^{\mathrm{miss}}$. In addition, a new set of filters was developed specifically for run 2 to address any previously unknown sources of artificial $p_{\mathrm{T}}^{\mathrm{miss}}$.

4.3.1 MET filters

The following provides additional information about the various filters that have been implemented [64, 65]:

- **Beam Halo filters**: Beam halos are machine induced particles flying with the beam, at large radius (up to 5m) which are produced through either beam-gas or beam-pipe interaction. There are also high energy halo muons which have a non negligible probability to interact in the calorimeters. Usually beam halos leave energy deposits peaking around $\phi = 0$ and $\phi = \pi$. Additionally, the CSC subdetector, which has a high ability to accurately reconstruct both collision and non-collision muons, often experiences interactions that align with energy deposits in the calorimeter. Accordingly several filters are developed for LHC run 1 and redesigned for run 2 utilizing information from both the CSC and the calorimeters.
- **HCAL filters:**
  - **HBHE Noise filters**: This filter aims to identify and eliminate noise signals from sources like single-ion feedback, magnetic field dependent collective HPD noise and RBX noise pulses in the HB/HE detector by utilizing geometrical patterns of HPD or RBX channels, as well as information about the pulse shape and timing. The algorithm uses these features to flag and control the noise i.e. whether “noise filtering” or “event filtering”.

- **HBHEiso Noise filters**: The isolation-based noise filter is used to identify isolated instances of noise activity in the HBHE detector by combining and comparing energy deposits in the HCAL and ECAL detectors with tracking measurements through a topological algorithm. Isolated HB/HE Noise filter rejects a large fraction of noise not caught by the standard HBHE Noise filter.

- **ECAL filters:**
  - **ECAL Dead Cell Trigger Primitive (ECAL TP) filter**: This filter targets events where ECAL energy recovery is performed based on L1 ECAL trigger primitives that are saturated. When the energy based on the trigger primitives is near the saturation energy, the recovered energy is likely underestimated so that leads to high artificial $p_{\mathrm{T}}^{\mathrm{miss}}$ events.
  - **EE Bad SC Noise filter**: There are four supercrystals (SC) in the EE (endcap) region that have been observed to produce anomalously high energy pulses in the TeV range. These pulses can appear in multiple channels simultaneously and are not caught by the standard noise flagging processes in the EE. This has led to the development of a filter that is designed to identify and reject events that exhibit this phenomenon.
  - **ECAL Bad Calibration filter**: This filter is designed to flag events where the $p_{\mathrm{T}}^{\mathrm{miss}}$ is attributed to a noisy EE crystal (based on a crystal list provided by ECAL). With increasing ECAL ageing, more noisy crystals have been added to the list. For example, particularly 2018 data in era D are affected.

- **Reconstruction filters:**
  - **Bad PF Muon and Bad PF Muon Dz filters**: During the LHC Run 2, there were instances of anomalously high $p_{\mathrm{T}}^{\mathrm{miss}}$ events that were caused by poor reconstruction of muons during the muon tracking step [25]. These events were caused by poorly reconstructed muons, which are identified by their relative p_{T} uncertainty of the muon best track or muon inner track. These filters were specifically designed to flag such events.
  - **Primary Vertex filter**: This filter simply requires at least one good reconstructed vertex in the event.

4.3.2 Performance of MET filters

This section discusses the performance of $p_{\mathrm{T}}^{\mathrm{miss}}$ filters during Run 2. The performance of such filters is investigated based on the Run 2 data, both the EOY (end-of-year) and UL (ultra legacy) datasets where the latter has a better calibaration. Moreover, as the artificial $p_{\mathrm{T}}^{\mathrm{miss}}$ effects have been simulated to a certain extent in the MC samples, this study is extended to MC samples as well. The performances of the MET filters are studied according to two properties of the filters such as their efficiencies and noise rejection fraction. The efficiency of a filter indicates how much the filter is good at not rejecting physical events. The noise rejection fraction of a filter reveals how much the filter is good at rejecting unphysical events.

Strategy

To evaluate the filters efficiencies and noise rejection fractions two perpendicular phase spaces are defined respectively: the noise free region (NFR) and the background enriched region (BER). The NFR phase space is specific to each dataset regardless of which filter is under study. However, the BER phase space is different for each filter under study. The filters that have less than 98% efficiency in NFR are targeted for further investigations and finally if they could not reach the 98% efficiency they are excluded from recommended filters. The efficiency and noise rejection fraction definitions read as,

$$\text{efficiency} = \frac{\mathrm{h}^{\text{ pass}}}{\mathrm{h}^{\text{ pass}-\text{except}}}, \tag{4.7}$$

$$\text{noise rejection fraction} = 1 - \frac{\mathrm{h}^{\text{ pass}}}{\mathrm{h}^{\text{ pass}-\text{except}}} \tag{4.8}$$

where $\mathrm{h}^{\text{ pass}}$ ($\mathrm{h}^{\text{ pass}-\text{except}}$) is the number of events passing all filters (respectively all but the filter under study), binned in four important observables (PF $p_{\mathrm{T}}^{\mathrm{miss}}$, PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$, number of primary vertices and leading jet p_{T}).

Noise Free Region

In order to remove noise contaminations from NFR as much as possible, recommended identification criteria for jets and muons are applied [39, 66],

- Jet identification

  - 2016: Loose working point (WP)
  - 2017 and 2018: TightLepVeto WP
- Muon identification: Medium WP
- Muon PF isolation: Tight WP

The event selection conditions in NFR are outlined as,

- "Jet HT" dataset
  - Number of jets (with $p_{\mathrm{T}} > 200$ GeV) ≥ 2
  - $\Delta\phi(j_1, j_2) > 2.9$
  - $0.8 < p_{\mathrm{T}}^{j_1}/p_{\mathrm{T}}^{j_2} < 1.2$
  - PF $p_{\mathrm{T}}^{\mathrm{miss}} < 100$ GeV
- "Single Muon" dataset
  - Number of muons (with $p_{\mathrm{T}} > 30$ GeV) $==1$
  - Transvers mass of (muon + PF $p_{\mathrm{T}}^{\mathrm{miss}}$) system < 120 GeV
  - PF $p_{\mathrm{T}}^{\mathrm{miss}} < 100$ GeV

Background Enriched Region

To estimate how much a filter is good at rejecting noises, we need to have a phase space highly populated by noisy events. In the following the event selection for BER of recommended filters [64] are listed,

- Beam halo filters
  - PF $p_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV
  - Number of jets (with $p_{\mathrm{T}} > 200$ GeV) $==1$
  - $|\eta_{jet}| < 2.4$ (central jets)
  - Jet CHEF (Charged Hadron Energy Fraction) < 0.01
  - $|\phi_{p_{\mathrm{T}}^{\mathrm{miss}}}| < 0.2$ or $|\phi_{p_{\mathrm{T}}^{\mathrm{miss}}}| > 2.9$

  - No additional jet with $p_{\mathrm{T}} > 25$ GeV
- Bad PF Muon filters
  - PF $p_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV
  - Number of muons (with $p_{\mathrm{T}} > 200$ GeV) $== 1$
  - No additional muons with $p_{\mathrm{T}} > 10$ GeV
  - $\Delta\phi(p_{\mathrm{T}}^{\mathrm{miss}}, \mathrm{muon}) > 3.0$
  - No additional jet with $p_{\mathrm{T}} > 25$ GeV
- HBHE noise filters
  - PF $p_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV
  - Number of jets (with $p_{\mathrm{T}} > 200$ GeV) $== 1$
  - $|\eta_{jet}| < 3.0$
  - Jet NHEF (Neutral Hadron Energy Fraction) > 0.9
  - $\Delta\phi(p_{\mathrm{T}}^{\mathrm{miss}}, \mathrm{jet}) > 3.0$
  - No additional jet with $pt > 25$ GeV
- ECAL bad calibration filter
  - PF $p_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV
  - Number of jets (with $p_{\mathrm{T}} > 200$ GeV) $== 1$
  - $2.5 < |\eta_{jet}| < 3.0$
  - Jet NEMF (Neutral EM Fraction) > 0.9
  - $\Delta\phi(p_{\mathrm{T}}^{\mathrm{miss}}, \mathrm{jet}) > 3.0$
  - No additional jet with $pt > 25$ GeV
- Other filters
  - PF $p_{\mathrm{T}}^{\mathrm{miss}} > 200$ GeV

These investigations resulted in updating some filters such as Bad PF Muon (considering a threshold for d_{xy}) and ECAL Bad Calibration filters, and also excluding "Bad

Charged Candidate" filter from recommended MET filters list that is still under investigation. Figure 4.3 shows recommended MET filters performances in NFR and BER phase spaces based on 2018 data.

The efficiency plots of the recommended filters show that they are not rejecting physical events. A decreasing trend for efficiency is seen as the $p_{\mathrm{T}}^{\mathrm{miss}}$ is increasing. This is normal because the chance of having a noisy event included in the NFR defined above increases at higher $p_{\mathrm{T}}^{\mathrm{miss}}$ values. In the BER we see that most of the filters are good at rejecting the noisy events. Although some filters such as "ECAL TP" and "EE Bad SC" filters have small rejection fraction, we still keep them as their efficiencies are almost 100%.

Performance of $p_{\mathrm{T}}^{\mathrm{miss}}$ tail cleaning

In order to see the effect of cleaning spurious $p_{\mathrm{T}}^{\mathrm{miss}}$, we compare between MC and data, before and after cleaning fake $p_{\mathrm{T}}^{\mathrm{miss}}$. For this purpose, in MC we use samples having large cross section and fake $p_{\mathrm{T}}^{\mathrm{miss}}$ like QCD with multi-jet production processes; Processes contributing to significant real $p_{\mathrm{T}}^{\mathrm{miss}}$ are included as well: $\mathrm{Z} \to \nu\nu$ + jets, $\mathrm{W}^{\pm} \to \ell^{\pm}\nu$ + jets, $\mathrm{t\bar{t}}$. For data, the EOY and UL MINIAOD samples used for this purpose are outlined in Table D.1 in the Appendix D. The list of MC samples and the correspondent cross sections are given in Tables D.2 and D.3 in the Appendix D.

Four sample categories are considered: "cleaned data", "cleaned MC", "uncleaned data" and "uncleaned MC". There are some common selection cuts performed on events to select very high $p_{\mathrm{T}}^{\mathrm{miss}}$ and high $\mathrm{H_T}$ events where $\mathrm{H_T} = \sum_i p_{\mathrm{T}}^{i}$ and i is runs over all jets with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.4$. Such selection cuts applied on all sample categories are listed below,

- Jets: $p_{\mathrm{T}} > 30$ GeV ; $\mathrm{N_{jets}} \geq 2$,
- PF $p_{\mathrm{T}}^{\mathrm{miss}} > 300$ GeV,
- $\mathrm{H_T} > 300$ GeV,
- $\Delta\phi(p_{\mathrm{T}}^{\mathrm{miss}}$,leading jet$) > 0.3, \Delta\phi(p_{\mathrm{T}}^{\mathrm{miss}}$,2$^{\mathrm{nd}}$ leading jet$) > 0.3$.

In addition to the selections above, the "cleaned data" and "cleaned MC" samples are required to satisfy the following conditions,

- Jets must pass identification criteria known as "JetID" as described in [66] with the following working points (WPs),
  - 2016: "Loose" WP,

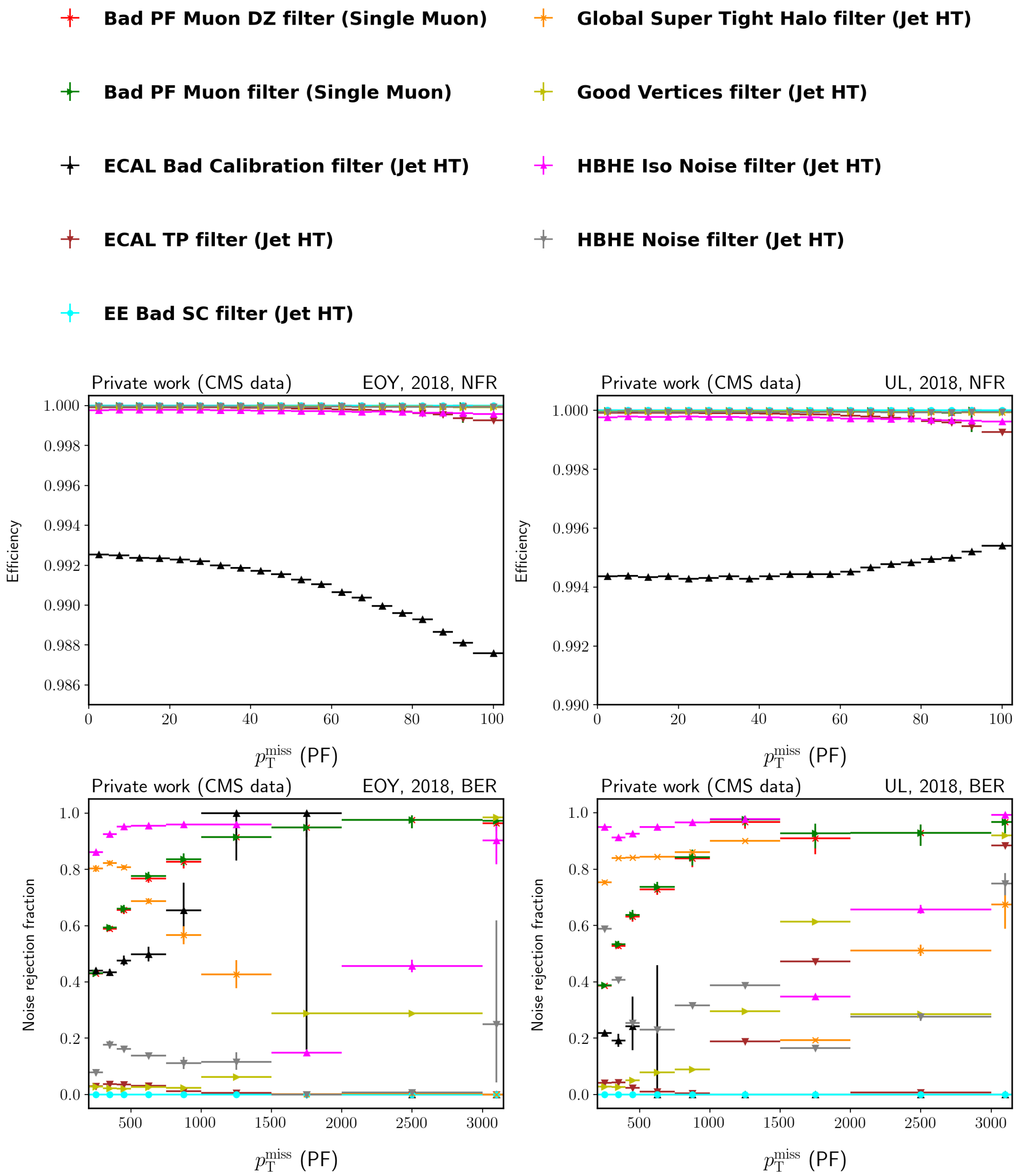

Figure 4.3: Recommended MET filters efficiencies for EOY (top left) and UL (top right), and their noise rejection fractions for EOY (bottom left) and UL (bottom right) data w.r.t PF $p_{\mathrm{T}}^{\mathrm{miss}}$ are shown. The plots are based on 2018 data. The list of filters and the dataset used for each filter are illustrated in the legend. The vertical error bars indicate the statistical error of each bin.

  - 2017: "TightLepVeto" WP,
  - 2018: "TightLepVeto" WP.

- Events must pass all recommended MET filters according to reference [64].
- **only for data**: Events must pass at least one of the following triggers `HLT_PFMET{X}_PFMHT{X}_IDTight_v* (where X = 90, 100, 110, 120, 130, 140).`

Figure 4.4 compares the PF and PUPPI $p_{\mathrm{T}}^{\mathrm{miss}}$ (version 15) distributions in both the data and MC samples, both before and after applying the recommended filters. Without using these filters, the $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution has a long tail in both the data and MC samples. The effect is much more pronounced in the data. However, after applying the filters, there is improved agreement between the data and simulated samples.

A large fraction of the remaining difference can be attributed to the use of LO samples for some of the MC processes. However, as we will see later in this thesis, other sources of discrepancies remain especially in the instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$. This is mostly due to the difficulty of perfectly simulating all the details of the detector response. In the analysis described in this thesis, a data-driven method is used to further improve the modelling of the instrumental MET. The figure also demonstrates that the uncleaned samples have peaks in the leading jet ϕ distribution near 0 and π, which is an indication of anomalous events. Many of these events are eliminated by the beam halo filter.

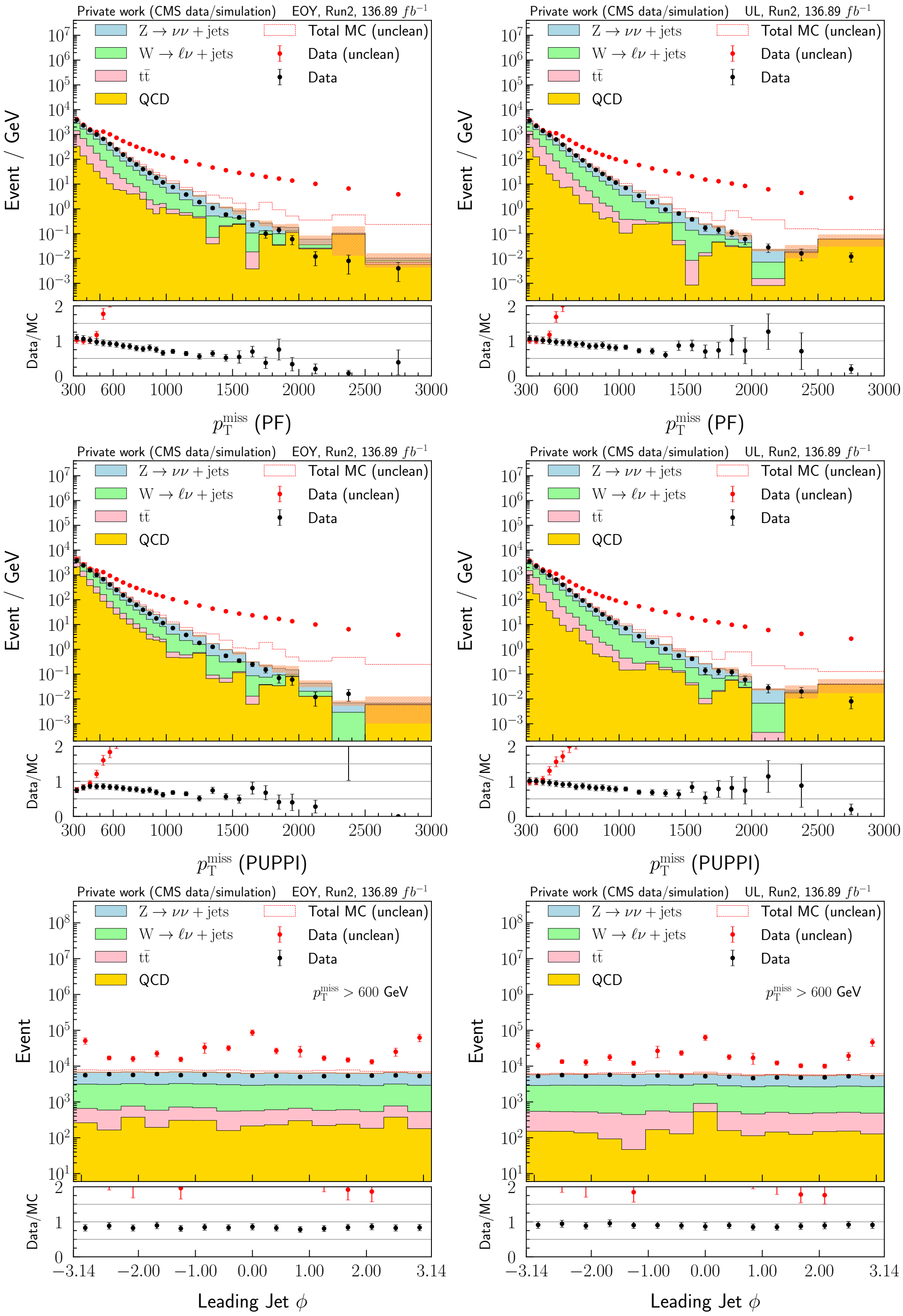

Figure 4.4: Stacked distributions of PF (top) and PUPPI version 15 (middle) p_T^{miss} , and stacked leading jet ϕ distributions, for the EOY (left) and UL (right) datasets. The lower panel of each plot is showing the ratio of data and MC distributions. The data before and after cleaning are indicated as red and black bullet points respectively.

Chapter 5

Analysis of the Off-shell H Boson Production in the $\mathrm{H} \to \mathrm{ZZ} \to \ell^+\ell^-\nu\bar{\nu}$ Final State

5.1 Analysis strategy

In this chapter the studies of off-shell H boson production in the $\mathrm{H} \to \mathrm{ZZ} \to 2\ell 2\nu$ final state will be presented. In this analysis we use data from the CMS experiment at the LHC corresponding to an integrated luminosity of around 138 fb^{-1} at a center-of-mass energy of 13 TeV. The signal searched for is the off-shell contribution of the H boson, where off-shell in this analysis corresponds to the mass range $m_{\mathrm{H}} > 200$ GeV. In this range, characteristic enhancements are expected, both in the Standard Model and in the presence of anomalous couplings (see Chapter 1). We focus on the final state where the H boson decays into a pair of on-shell Z bosons, one of which decaying into a e^+e^- pair or a $\mu^+\mu^-$ pair, and the other decaying into a neutrino-antineutrino pair giving rise to missing transverse momentum. Previous studies [67, 68] have shown that the sensitivity of this channel to the production of a scalar particle is good when compared to the decay channels with 4 charged fermions, in the mass range 300-1000 GeV, where most of the SM off-shell contribution is expected. The gluon fusion (GF) and vector boson fusion (VBF) production mechanisms are the main production mechanisms of this analysis. The associated production with a vector boson (VH) typically plays a small role in this analysis, but can be important when considering anomalous HVV couplings. The events are classified into 6 categories, according to the charged-lepton flavour (e^+e^- and $\mu^+\mu^-$) and to the number of jets with $p_{\mathrm{T}} \geq 30$ GeV (0, 1 and ≥ 2 jets).

The signal distribution is estimated by Monte Carlo simulations. Since some production processes of Z boson pairs have large interference with the H-boson induced production, signal simulation must account for this interference. The analysis makes use of jet categories, therefore an effort is made to simulate the signal and the interfering background at the highest order available in QCD. This is achieved by a matrix-element based reweighting technique that is one of the author's main contributions to the analysis. The reweighting technique will be described in section 5.2.2.

In this analysis the main backgrounds are:

1) non-interfering diboson production (ZZ, WZ),
2) backgrounds without a true Z boson, called non-resonant, mainly composed of $t\bar{t}$ and W^+W^- processes,
3) Z+jet production where the missing transverse momentum is instrumental ('fake').

These backgrounds have comparable impacts on the results and must be estimated carefully. Diboson production is estimated by Monte Carlo simulations. The other backgrounds are not well modeled by simulations. Non-resonant background is therefore estimated from data by selecting $e\mu$ events satisfying the same analysis kinematic selections and using the flavour universality of W decays into charged leptons. The Z+jet background with fake missing transverse momentum is also estimated from data, using a photon+jet control sample. In the photon+jet sample, the source and distribution of fake missing transverse momentum are similar to those in Z+jet events.

The signal strength is extracted by a likelihood fit of the expected signal and background contributions to the histograms of a few kinematic observables. The free fit parameters are 1) the signal strength of the off-shell GF contribution, and 2) the signal strength of the off-shell VBF and VH contributions grouped together into what we call the electroweak (EW) contribution. The observables are chosen such that they are expected to be distributed differently in the signal and in the backgrounds. Some of these observables target the EW contribution specifically. These are defined only for the event categories which have ≥ 2 jets. Systematic uncertainties are treated as nuisance parameters that are profiled in the fit. The $2\ell 2\nu$ off-shell data are interpreted in terms of the off-shell signal strengths of the SM GF and VBF production modes. The $2\ell 2\nu$ analysis results are then combined with previously-published CMS results in the 4ℓ final state to improve the measured off-shell signal strengths, derive a measurement of the H decay width, and constrain anomalous HVV couplings.

The simulated samples are described in Section 5.2 In particular, the reweighting technique applied in order to model the signal and the interfering background at NLO in QCD is described in detail. The event selection and categorization are described in

Sections 5.3 and 5.4. The kinematic observables are defined in Section 5.5. The data-driven estimations of the non-resonant and Z+jet backgrounds are discussed in Section 5.6. The likelihood parametrization is described in detail in Section 5.7. Systematic uncertainties are described in Section 5.8. Results are presented and discussed in Section 5.9.

5.2 Monte Carlo Simulation

Event simulation and correction techniques are important in this analysis in order to ascertain the various kinematic properties of the H boson signal with associated jets and the interference of the different signal processes and moreover to obtain the best precision in the dominant irreducible background processes, $q\bar{q} \to ZZ$ and WZ, with the backgrounds that involve the same initial and final states. All events are simulated using MC sampling methods, and the events are interfaced to PYTHIA 8 [69] for parton-showering, using version 8.212 for the simulation of the 2016 data period and 8.230 for the simulation of the 2017 and 2018 data periods. In almost all simulation samples, the CMS tune CUETP8M1 [70] is used for the 2016 data period, and tune CP5 [71] for the 2017 and 2018 data periods

5.2.1 Simulation of Non-interfering Backgrounds

Various simulation samples are used in the analysis to understand the different background components in each of the signal or control regions. They are grouped in different tables based on whether they are used in the analysis of the signal region or dilepton control regions (Tables D.4, D.5, and D.6 in the Appendix D for the 2016, 2017, and 2018 data periods, respectively), or the analysis of the single photon control region (Tables D.7, D.8, and D.9 in the Appendix D for the 2016, 2017, and 2018 data periods, respectively).

The nominal PDF set used in generating these samples is NNPDF 3.0 [72, 73] for the 2016 data set samples or NNPDF 3.1 [74] for the 2017 and 2018 data set samples with various LO, NLO, or NNLO QCD choices. However, in the determination of the $q\bar{q} \to VV$ (VV = WZ or ZZ) background contributions to the signal region, the events are reweighted to the NNPDF 3.0 NLO QCD PDF, suitable also to the QCD order of the simulation itself, in order to have a uniform cross section prediction across the data periods. This detail also applies to the simulation of the signal samples, discussed in Sec. 5.2.2. The simulated events for the $q\bar{q} \to ZZ$ contribution are further reweighted for the NLO EW virtual correction (corresponding to loop diagrams) for two on-shell Z bosons as a function of the Mandelstam variables $\hat{s}$ and $\hat{t}$ for each quark flavor contribution in the initial states [68, 75, 76]. For $q\bar{q} \to WZ$, two contributions exist to the NLO EW corrections: there is a negative correction due to virtual effects, applied in the same way as for $q\bar{q} \to ZZ$, and a positive contribution due to photon-induced processes, parametrized as a function of $\hat{s}$. These two contributions are found to nearly cancel each other. Both $q\bar{q} \to ZZ$ and $q\bar{q} \to WZ$ are reweighted for NNLO QCD effects as a function of m_{VV} [13, 77, 78]. The NLO EW correction reaches a value up to -15% at $m_{VV} = 1$ TeV with a comparable uncertainty driven by cross-contamination with NLO QCD corrections, and the NNLO QCD correction is an approximately uniform $+15\%$ correction at high m_{ZZ} values in the ZZ case.

5.2.2 Signal and Interfering Backgrounds Modeling

Since the analysis is categorizing events according to the number of jets, the simulation of the signal and interfering backgrounds should be made at the highest order available in QCD. No generator is available for the simulation of the off-shell tail beyond LO. Therefore, samples generated at NLO with different values of the H boson pole mass in the range 125-3000 GeV are reweighted and combined in order to obtain the spectra corresponding to the different signal hypotheses, to the SM background, and to their interference. The simulation of the samples with fixed H pole masses is described in section 5.2.2.1. The reweighting technique is detailed in section 5.2.2.3 and its validity is discussed in section 5.2.2.4. Eventually, the differences between direct simulation at LO and reweighted NLO samples are shown for the GF and EW processes in section 5.2.2.5.

5.2.2.1 Event Simulation

As pointed in Sec. 5.1, in this analysis we consider two main H boson production processes i.e. GF and VBF. In part of the phase space, the VBF process interferes with the VH process in which the V boson in which V stands for vector bosons such as $W^{\pm}$ and Z bosons, decays into a quark-antiquark pair. The $ZH \to ZZZ \to 2\ell 2\nu + 2$jets can also contribute in the off-shell signal region even when the H boson is on-shell. We therefore also need to include the VH processes in order to obtain the full topology.

These events are simulated in two steps. The first step involves producing the events with a stable H boson in GF, VBF, ZH, or $W^{+}H$ and $W^{-}H$ production modes using the POWHEG 2 [79–83] event generator. The GF and VBF production modes are generated at NLO in QCD, and the VH production mode is generated by including the MiNLO HVJ add-on [84, 85], which brings the precision of event generation up to NNLO in QCD.

POWHEG does not allow the simulation of the off-shell spectrum. However samples are produced with wide range of H boson pole masses m_{H} i.e. 125, 160, 170, 180, 190, 200, 210, 230, 250, 270, 300, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1500, 2000, 2500, and 3000 GeV that can be reweighted and combined into a continuous spectrum ranging from low mass values (around 125 GeV) up to 3 TeV. As the H boson pole mass can have a very high value like 1 TeV and etc., in order to allow for the presence of a stable and broad resonance generation in POWHEG, the H boson propagator is defined in the complex-pole scheme (CPS) [86]. Even for the low pole mass values, the same H boson propagator i.e CPS is used consistently to allow the application of a uniform reweighting procedure described later in this section (Sec. 5.2.2.3). In the reweighting procedure the CPS propagator, which is part of the samples generated by POWHEG, is reweighted to a regular Breit-Wigner (BW) propagator based on the SM H boson properties i.e. $m_{\mathrm{H}} = 125$ GeV and $\Gamma_{\mathrm{H}} = 4.07$ MeV . All distributions shown from these samples use the NNPDF 3.0 NLO nominal PDF set with $\alpha_S = 0.118$. The list of cross sections of all samples in different production modes are shown in Table 5.1 as well.

The second step of the simulation involves decaying the H boson to $H \to 2\ell 2\nu$ final states through intermediate Z or W bosons using the JHUGen [16, 87–90] program versions between 6.9.8 and 7.4.0. Afterwards, the events are interfaced to PYTHIA 8.230 [69] for parton showering and underlying event simulation using the CMS tune CP5 [71].

5.2.2.2 Precision Corrections

In addition to what is presented in table 5.1 for cross sections, we apply the ratio of the gluon-fusion cross section computed at NNLO in QCD to the one computed at NLO. This ratio is further referred to as NNLO/NLO ggF k-factor. This factor is computed and applied as a function of m_{VV} [91] and inclusively in jet bins. The N^{3}LO / NNLO k-factor value i.e. a uniform k-factor of 1.1, computed at $m_{\mathrm{VV}} = 125$ GeV, is also applied on the GF process, uniformly across the m_{VV} spectrum and for all jet bins [91].

Table 5.1: The cross section values for the different H boson production modes are listed for the samples produced at different values of the H boson pole mass m_H using the POWHEG V2 event generator. The MiNLO HVJ add-on is used in the VH modes. The cross sections are obtained with the NNPDF 3.0 NLO nominal PDF set. The cross sections are reported in units of pb.

m_{H}	σ_{GF}	σ_{VBF}	$\sigma_{\mathrm{W^+H}}$	$\sigma_{\mathrm{W^-H}}$	σ_{ZH}
125	30.0	3.77	8.50×10^{-1}	5.34×10^{-1}	7.53×10^{-1}
160	20.0	3.00	3.89×10^{-1}	2.37×10^{-1}	3.47×10^{-1}
170	17.9	2.79	3.15×10^{-1}	1.90×10^{-1}	2.80×10^{-1}
180	16.3	2.62	2.61×10^{-1}	1.56×10^{-1}	2.31×10^{-1}
190	14.8	2.44	2.15×10^{-1}	1.28×10^{-1}	1.90×10^{-1}
200	13.6	2.29	1.80×10^{-1}	1.06×10^{-1}	1.59×10^{-1}
210	12.7	2.18	1.55×10^{-1}	9.00×10^{-2}	1.36×10^{-1}
230	11.2	1.97	1.14×10^{-1}	6.56×10^{-2}	1.01×10^{-1}
250	9.88	1.75	8.49×10^{-2}	4.77×10^{-2}	7.37×10^{-2}
270	8.86	1.56	6.36×10^{-2}	3.51×10^{-2}	5.49×10^{-2}
300	7.89	1.33	4.29×10^{-2}	2.32×10^{-2}	3.66×10^{-2}
350	8.03	1.04	2.39×10^{-2}	1.24×10^{-2}	2.02×10^{-2}
400	7.14	8.47×10^{-1}	1.47×10^{-2}	7.33×10^{-3}	1.21×10^{-2}
450	5.06	6.90×10^{-1}	9.31×10^{-3}	4.51×10^{-3}	7.60×10^{-3}
500	3.36	5.61×10^{-1}	6.07×10^{-3}	2.90×10^{-3}	4.90×10^{-3}
550	2.21	4.59×10^{-1}	5.11×10^{-3}	1.89×10^{-3}	3.23×10^{-3}
600	1.47	3.80×10^{-1}	2.82×10^{-3}	1.27×10^{-3}	2.21×10^{-3}
700	6.87×10^{-1}	2.71×10^{-1}	1.46×10^{-3}	6.31×10^{-4}	1.12×10^{-3}
800	3.49×10^{-1}	2.02×10^{-1}	8.24×10^{-4}	3.48×10^{-4}	6.22×10^{-4}
900	1.92×10^{-1}	1.56×10^{-1}	5.01×10^{-4}	2.01×10^{-4}	3.72×10^{-4}
1000	1.13×10^{-1}	1.24×10^{-1}	3.18×10^{-4}	1.27×10^{-4}	2.36×10^{-4}
1500	1.46×10^{-2}	4.42×10^{-2}	5.49×10^{-5}	2.04×10^{-5}	3.92×10^{-5}
2000	6.23×10^{-3}	3.45×10^{-2}	2.76×10^{-5}	1.00×10^{-5}	2.00×10^{-5}
2500	2.51×10^{-3}	2.14×10^{-2}	1.18×10^{-5}	4.35×10^{-6}	7.95×10^{-6}
3000	1.19×10^{-3}	1.41×10^{-2}	5.56×10^{-6}	2.12×10^{-6}	4.24×10^{-6}

5.2.2.3 Reweighting Technique

After the event simulation, we follow a reweighting procedure to obtain the H boson off-shell signal (for the SM and aforementioned anomalous couplings), the background with the same initial and final states, and their interference. The simulation of the different tensor structures (see Eq. (1.34)) for different anomalous couplings follows the methods developed for the analyses described in Ref. [92]. The reweighting procedure is done for each production process (i.e. GF, VBF and VH) separately for all aforementioned hypotheses such as SM signal, SM background and etc.

The main concept of reweighting is to compute the matrix elements probability densities one time for the available simulated sample i.e. the fixed pole mass POWHEG H boson signal, and one time for the target hypothesis. We utilize the four momenta of the incoming quarks and gluons as well as the four momenta of all outgoing final state particles and the intermediated H boson. Then the event in the sample is reweighted by the ratio of the square of the computed matrix elements $\frac{\mathcal{P}_{\mathrm{h}}}{\mathcal{P}_{\mathrm{s}}}$. At the end, in each process, for a given hypothesis, all samples with different m_H values, are combined such that the full mass spectrum is obtained after the combination step. The details of matrix elements calculations and the rest of reweighting procedure will be discussed in the following paragraphs.

The core of the procedure is the reweighting of the hard process using MELA [16, 87–90] and MELAAnalytics [93] packages. The MELA package provides a C++ interface to the matrix elements used in the JHUGen and MCFM simulation programs. The H boson signal or VV continuum amplitudes are taken at LO in QCD from the MCFM program with the signal amplitudes being modified for BSM HVV couplings using the amplitude formalism in JHUGen. The MELA package also provides the reweighting functionality between CPS and BW propagators as the POWHEG samples are only using one of the propagator types, the CPS one.

It should be note that in matrix element computations, the MELA package uses the four-momenta of the initial and final state particles configured according to the LO QCD topology. In the case of the VBF process at NLO or the VH process at MiNLO ($\sim$ NNLO), there is one extra gluon or two extra partons respectively. Therefore, we need a tool to derive the approximated LO topology expected from the generated N(N)LO QCD final state configurations. This kind of tool is implemented as a part of the MELA-Analytics package which tries to merge the four-momentum of any initial or final state gluon in the hard process (which is prior to parton shower) into that of the incoming or outgoing quark with the smallest four-vector dot product. This approach corresponds to the assumption that gluon radiation from the quarks is soft or collinear, which is found to give sufficient approximation in previous studies [93, 94] compared to the relevant statistical precision of the data analyses. The rules of the merging procedure for EW processes are described below.

- An incoming gluon is never merged into an incoming quark. This rule is invoked implicitly as the q^2 of the pair of incoming partons is always the largest compared to that of any other pair of partons in the event.
- Gluons are never merged into the decay products of the H boson from the JHUGen step as they are produced during the production of the H boson with no prior knowledge about the boson's decay.
- Gluons are also never merged into the decay products of the associated W or Z boson in the VH samples. Doing so distorts the Breit-Wigner nature of these resonances significantly.
- All merging is done in the convention of outgoing particles. This means the four-momenta and charge of incoming particles are reversed in the intermediate steps when those of the two merged particles are summed.
- When an incoming gluon is merged into an outgoing quark, the charge (i.e., PDG id), and the four-momentum of the quark are reversed in the final step of the LO topology construction. This reversion is done so that the event topology ensures having exactly two incoming quarks as expected by the LO matrix elements.
- In VH samples, when extra gluons are encountered, the merging of individual gluons and that of a combined gluon (i.e., from a g $\rightarrow$ gg process) are all considered separately.
- In VH samples, it is also possible to encounter two extra quarks instead of gluons. These extra quarks are merged into a gluon substitute first, as they are from a g $\rightarrow$ q$\bar{\text{q}}$ branching process, before the merging of this gluon substitute is considered.
- Every merging case is considered, and those that do not produce an incoming-outgoing parton composition that is compatible with the LO physics process of the sample (i.e., VBF, ZH, or WH) are skipped.
- A momentum redistribution procedure is applied on the incoming and outgoing particles associated with H boson production so that the resultant topology features massless particles, which is what is required by the use of massless spinors in matrix elements. Denoting the momenta of the two final incoming or outgoing partons as p_1 and p_2, an intermediate four-momentum k is added to p_1 and subtracted from p_2 such that $|p_1+k|^2 = |p_2-k|^2 = 0$. This step is common to

any matrix element computed using the MELA package. Because event-by-event reweighting is done through a ratio of matrix elements, which are invariant under any arbitrary boost of the event topology, and because factors coming from PDFs cancel in the ratio, the common boost of all particles does not affect reweighting as long as momentum conservation is maintained strictly, and is therefore adjusted arbitrarily.

In contrast with the EW processes, the NLO QCD corrections on the GF process are already known to be large [7], but such corrections are not factorizable due to the fact that the GF process is induced by QCD loops. Therefore, the full matrix element at NLO in QCD would have to be used ideally, but those matrix elements are not known for the continuum contribution of the full $\mathrm{gg} \to \mathrm{VV}$ amplitude yet. Hence, the matrix element calculations for the GF process is done using the decay products of the VV system kinematics and boosting all VV decay products to the $p_T^{\mathrm{VV}} = 0$ frame. Then the value of p_z^{VV} is used, together with m_{VV} to assign the four-momenta to the two incoming gluons expected in the tree level amplitude.

It should be noted that, since in the two-stage event simulation (i.e. the POWHEG production and JHUGen decay), only the s-channel diagrams are considered, the denominator of the reweighting factors uses only the s-channel diagram as well. The numerator of the weights, instead, includes the full set of Feynman diagrams which are appropriate for gauge invariance. So in the case of H boson signal amplitudes, in the numerator of GF process only s-channel diagrams are considered, however, for the EW processes the combined $s + t + u$-channel is considered, where subdominant contributions come from the t and u channels.

Furthermore, the POWHEG program generates events for the decay of the H boson without any specific assumptions about the decay process. Therefore, the program approximates the dependence of the H boson decay on the H boson mass (m_{VV}) using a factor that includes the full total width of the H boson, denoted as $m_{\mathrm{VV}} \times \Gamma_{\mathrm{H}}(m_{\mathrm{H}})$. The values of $\Gamma_{\mathrm{H}}(m_{\mathrm{H}})$ are taken from Ref. [95]. However, this approximation is not accurate enough for the SM decay of $\mathrm{H} \to \mathrm{VV}$. To improve the accuracy, we need to correct the partial decay width of each sample. For this we reweight the samples with $m_{\mathrm{H}} \geq 200$ GeV based on the expected evolution of the $\mathrm{H} \to \mathrm{VV}$ partial decay width as a function of m_{VV}. This correction is calculated relative to the pole mass of each sample. Events with $m_{\mathrm{H}} < 200$ GeV are not corrected because they have very small total decay widths and the corrections are negligible.

The reweighting procedure is split into several steps. Since reweighting is performed over a broad m_{VV} spectrum, and as the kinematics of samples at different m_H values differ only over this quantity, the m_{VV} spectrum is first divided into bins that enclose the different m_H values. The bin boundaries are determined as the average of two consecutive m_H values, except when there is a relatively large interval between m_H values like for $m_H = 125$ GeV and 160 GeV, in which the interval is split at $m_{\mathrm{VV}} = 136.7$ GeV and 148.3 GeV. We classify the weights as native sample weights and reweighting weights. Native sample weights include any weight coming from the POWHEG simulation or the partial decay width correction, and also include a constant normalization factor for the cross section and the branching ratio of the sample. Regarding the samples BR, it should be noted that all samples BRs are corrected back to the 125 GeV sample. The reweighting weights are defined as the product of the ratios of matrix elements corresponding to different target hypotheses (SM or BSM signal, background, interference) and of the CPS to BW factors. Events with artificial large weights, which show up because of statistical fluctuations in the samples, are removed according to an outlier rejection algorithm detailed in Appendix A.

Once events with large reweighting weights in any hypothesis are removed from all hypotheses, the events in each m_{VV} bin need to be combined according to the number of effective events in that bin for that sample. The number of effective events N_{eff}^{si}, defined as:

$$N_{\mathrm{eff}}^{si} = \frac{{S_{si}^{h0}}^2}{V_{si}^{h0}}, \tag{5.1}$$

where

$$\begin{aligned} S_{si}^{h0} &= \sum_j w_{sj} \cdot r_{sj}^{h0} \\ V_{si}^{h0} &= \sum_j w_{sj}^2 \cdot {r_{sj}^{h0}}^2, \end{aligned} \tag{5.2}$$

where the index s denotes the sample, i denotes the m_{VV} bin, $h0$ denotes a reference target hypothesis, the weight w denotes the product of native sample weights, the weight r denotes the reweighting weight, and the sum over index j runs over the events in bin i of sample s. Regarding the reference target hypotheses, one may choose any hypothesis without loss of generality, since all would provide statistically consistent results. In this study the SM signal hypothesis is chosen as the reference target hypothesis $h0$. The weight of the sample s in the m_{VV} bin i is determined as the fractional size of the number of effective events from sample s in bin i,

$$W_{si} = \frac{N_{\mathrm{eff}}^{si}}{\sum_k N_{\mathrm{eff}}^{ki}}. \tag{5.3}$$

The W_{si} weight is determined in two iterations. In the second iteration, it removes events of sample s in the bins with a relative contribution smaller than 0.01 (0.005 in the case of the GF production mode) unless the bin in question is immediately adjacent to the bin of the sample that encloses the value of its m_H. If the bin to be removed is positioned at the left (right) of this central bin (bin i), all preceding (succeeding) bins are also removed. As the last step, for each sample with $m_H > 200$ GeV the cross section normalization factor need to be calculated back to the sample with $m_H = 200$ GeV taken as a reference sample. This cross section normalization factor is based on the sum of S_{si}^{h0} over common bins for a sample s and sample $s-1$ with non-zero W_{si}. Finally, the cross section normalization factor for sample s, X_s is computed as,

$$\begin{aligned} X_s &= \prod_{k=s_0+1}^{s} R_k \\ R_s &= \frac{\sum\limits_{j_{s-1}} S_{(s-1)j}^{h0} \cdot \mathcal{H}(W_{sj}) \cdot \mathcal{H}\left(W_{(s-1)j}\right)}{\sum\limits_{j_s} S_{sj}^{h0} \cdot \mathcal{H}(W_{sj}) \cdot \mathcal{H}\left(W_{(s-1)j}\right)}, \end{aligned} \tag{5.4}$$

where s_0 denotes the sample with $m_H = 200$ GeV, and $\mathcal{H}$ is the Heaviside theta function. The X_s for samples with $m_H \leq 200$ GeV is set to be 1. Now, the final weights for any target hypothesis *h* is determined to be the product of $r_{sj}^h, w_{sj}, W_{si},$ and X_s.

5.2.2.4 Quality of the LO Approximation for the NLO EW Event Topology

The idea of computing ME weights for NLO samples on the basis of a corresponding LO topology, obtained by the merging procedure described earlier, can be validated by checking the agreement with characteristic distributions simulated at LO for a few processes. In this section, we illustrate this validation for EW processes. By lack of centrally-produced LO samples for the qq $\to 2\ell 2\nu + (\mathrm{qq})$ final state, the comparison is done with qq->4l+qq final-state samples to which we have applied the merging procedure. We also include a comparison with the unmerged NLO distributions, but we will defer the discussion of the differences to next section, where a more relevant comparison is made after parton showering. The LO samples are generated via JHUGen/MCFM in which the samples are produced with the same PDF set as that used for the POWHEG samples in order to minimize differences in parton distributions when comparisons are examined. Besides JHUGen/MCFM samples, the PHANTOM event generator and the JHUGen on-shell-only generation mode samples are also studied as additional references pertaining to the off-shell region and the on-shell region respectively. These two sets of samples are produced with the NNPDF 3.0 LO QCD PDF set with $\alpha_S = 0.130$. For unmerged NLO samples, the two leading jets are taken as the

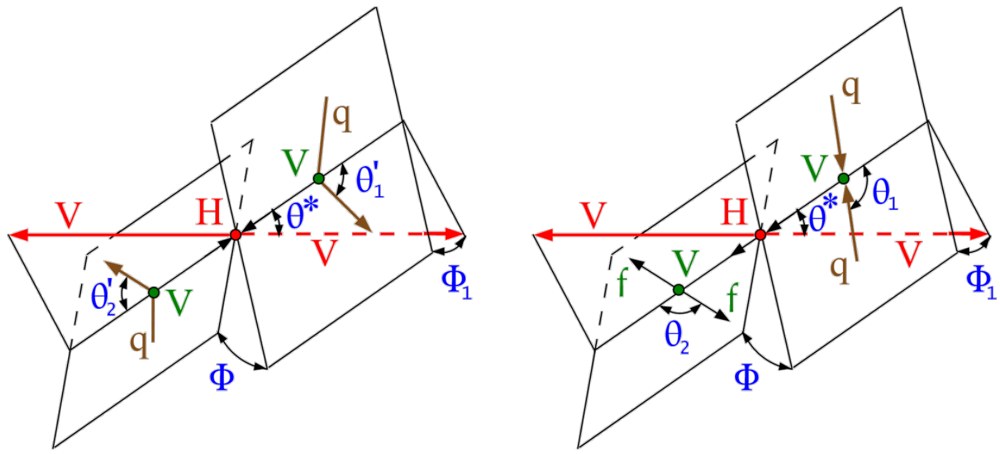

Figure 5.1: Two topologies of the H boson production and decay: vector boson fusion $\mathrm{qq} \rightarrow \mathrm{VV(qq)} \rightarrow \mathrm{H(qq)} \rightarrow \mathrm{VV(qq)}$ (left); and the VH i.e. the associated H boson production $\mathrm{qq} \rightarrow \mathrm{V} \rightarrow \mathrm{VH} \rightarrow f\bar{f}\mathrm{H} \rightarrow f\bar{f}\mathrm{VV}$ (right). The incoming particles are shown in brown, the intermediate vector bosons and their fermion daughters are shown in green, the H boson and its vector boson daughters are shown in red, and angles are shown in blue. The angles are defined in either the H or V boson rest frames [87, 89].

two outgoing partons leading in p_T. Before doing the comparisons, some selection cuts are applied on the events which are tighter than those placed at generator level on any of these examined samples. Leptons (i.e. electrons and muons) are required to satisfy $p_\mathrm{T} \geq 7$ GeV and $|\eta| < 2.4$, and the first two leading-p_T outgoing partons are required to satisfy $p_\mathrm{T} \geq 30$ GeV and $|\eta| < 4.7$. Moreover, the requirements $m_{\ell\ell} \geq 4$ GeV on any dilepton pair with same flavor and opposite charge, $m_\mathrm{jj} \geq 30$ GeV over the two leading-p_T partons, and $|\Delta R_{j,\ell}| > 0.4$ over any parton-lepton pair are imposed.

As described in [87–89] a generic $qq \rightarrow X \rightarrow V_1(q_1)V_2(q_2) \rightarrow 4f$ interaction (X is determined as H boson in our case) can be fully characterized by three invariant masses $m_{\mathrm{V_1V_2}}, m_{\mathrm{V_1}}$, and $m_{\mathrm{V_2}}$, and the five angles defined in Fig. 5.1. Due to the fact that the $\mathrm{H} \rightarrow f_1 f_2 f_3 f_4$ (H boson decay), $f_1 f_2 \rightarrow f_3 f_4 \mathrm{VV}, \mathrm{VV} \rightarrow \mathrm{H}$ (VBF), and $f_1 f_2 \rightarrow \mathrm{V}^* \rightarrow \mathrm{H} + f_3 f_4$ (VH) diagrams can be taken as space-time transformations of each other, these kinematic variables fully describes the EW processes (i.e. VBF and VH) in our study. In the following discussion, we will examine the angular quantities $\cos\theta_1$, $\cos\theta_2$, $\cos\theta^*$, Φ_1, and Φ demonstrated in Fig. 5.1, and the squares of the four-momenta of the V bosons, q_1^2 and q_2^2. Further details can be found in Ref. [88].

In the case of VBF-like definitions, the q^2 variables correspond to the virtual Z or W exchange, so they are negative. For this reason, $\cos\theta_1$ and $\cos\theta_2$ for this topology are defined in the H boson rest frame instead of the rest frame of V bosons. In VH-like definitions, V_1 in the computation of $\cos\theta_1$ is taken to be the incoming off-shell V^*

boson.

The first set of comparison is displayed for off-shell H boson production in Figs. 5.2 and 5.3 for the VBF ($m_{\mathrm{jj}} > 130$ GeV), Figs. 5.4 and 5.5 for the ZH (80 GeV $< m_{\mathrm{jj}} <$ 100 GeV with compatible initial and final partons), and Figs. 5.6 and 5.7 for the WH (70 GeV $< m_{\mathrm{jj}} <$ 90 GeV with compatible initial and final partons). All kinematic variables are observed to agree well among the LO samples and with the reweighted NLO samples when the LO approximation method (merged scenario) is used.

The most challenging EW contribution to either simulate or reweight from the NLO POWHEG samples is the continuum-only ZZ production with two associated partons: many more Feynman diagrams are involved in this contribution than the number of H boson-related diagrams, and when reweighting is done from generated events with only s-channel H boson diagrams, significant changes in the tensor structure are expected, which makes reweighting more susceptible to statistical fluctuations from event generation. Comparisons are displayed in Figs. 5.8 and 5.9, Figs. 5.10 and 5.11, and Figs. 5.12 and 5.13 for ZZ production through vector boson scattering (VBS), ZZZ, and WZZ topologies, respectively, where the 4ℓ system is produced at high invariant mass. The definition of these topologies is kept the same as for the aforementioned VBF, ZH, and WH counterparts, respectively.

In the next section the comparison of LO samples i.e. JHUGen/MCFM samples, and reweighted samples i.e. the POWHEG samples, will be made and discussed in detail.

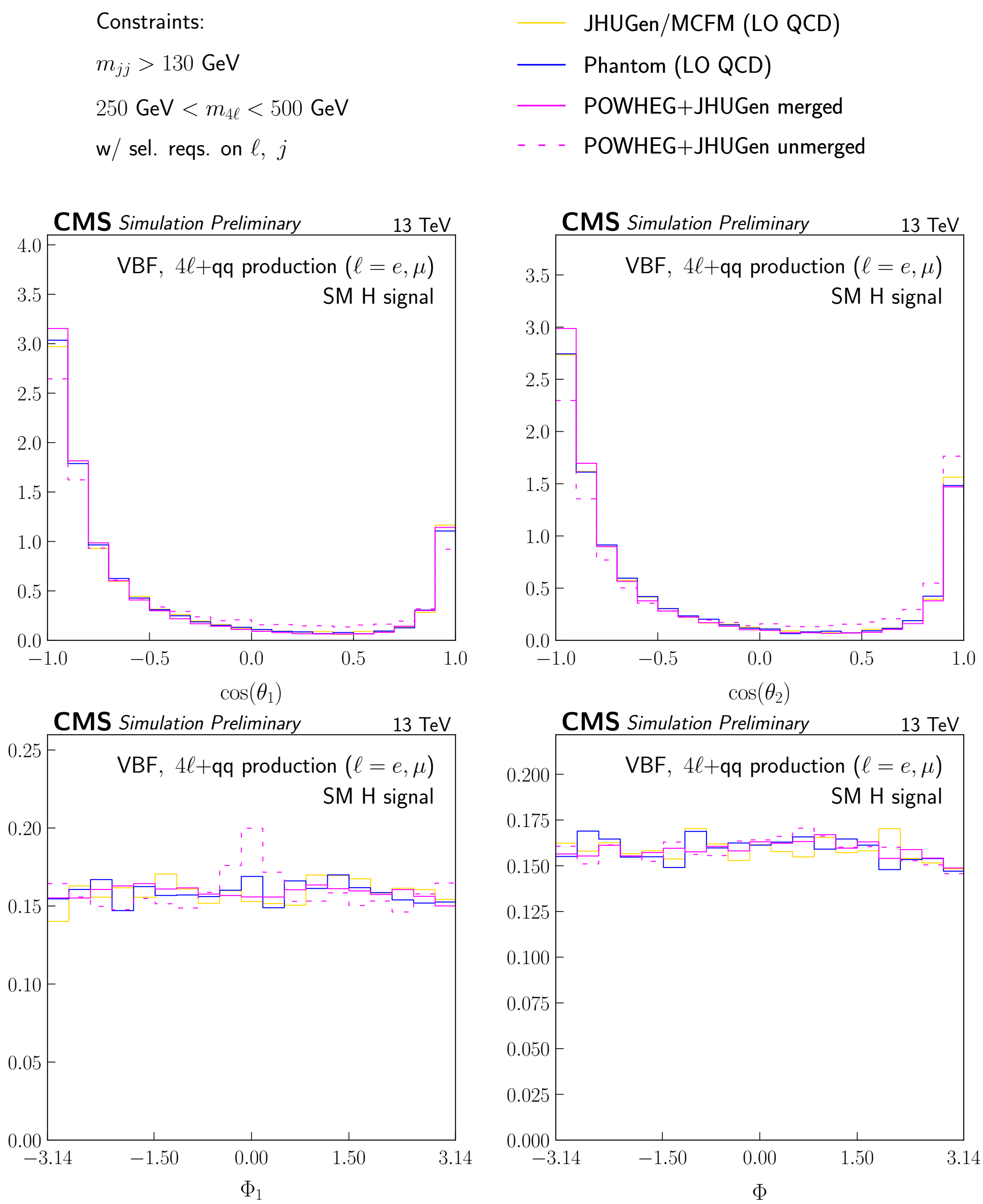

Figure 5.2: The distributions of $\cos\theta_1$, $\cos\theta_2$, Φ_1, Φ for off-shell H boson production in the SM EW signal process. The requirement $m_{\mathrm{jj}} > 130$ GeV is applied on all distributions to emphasize the VBF-like topology. All distributions are normalized to unit area.

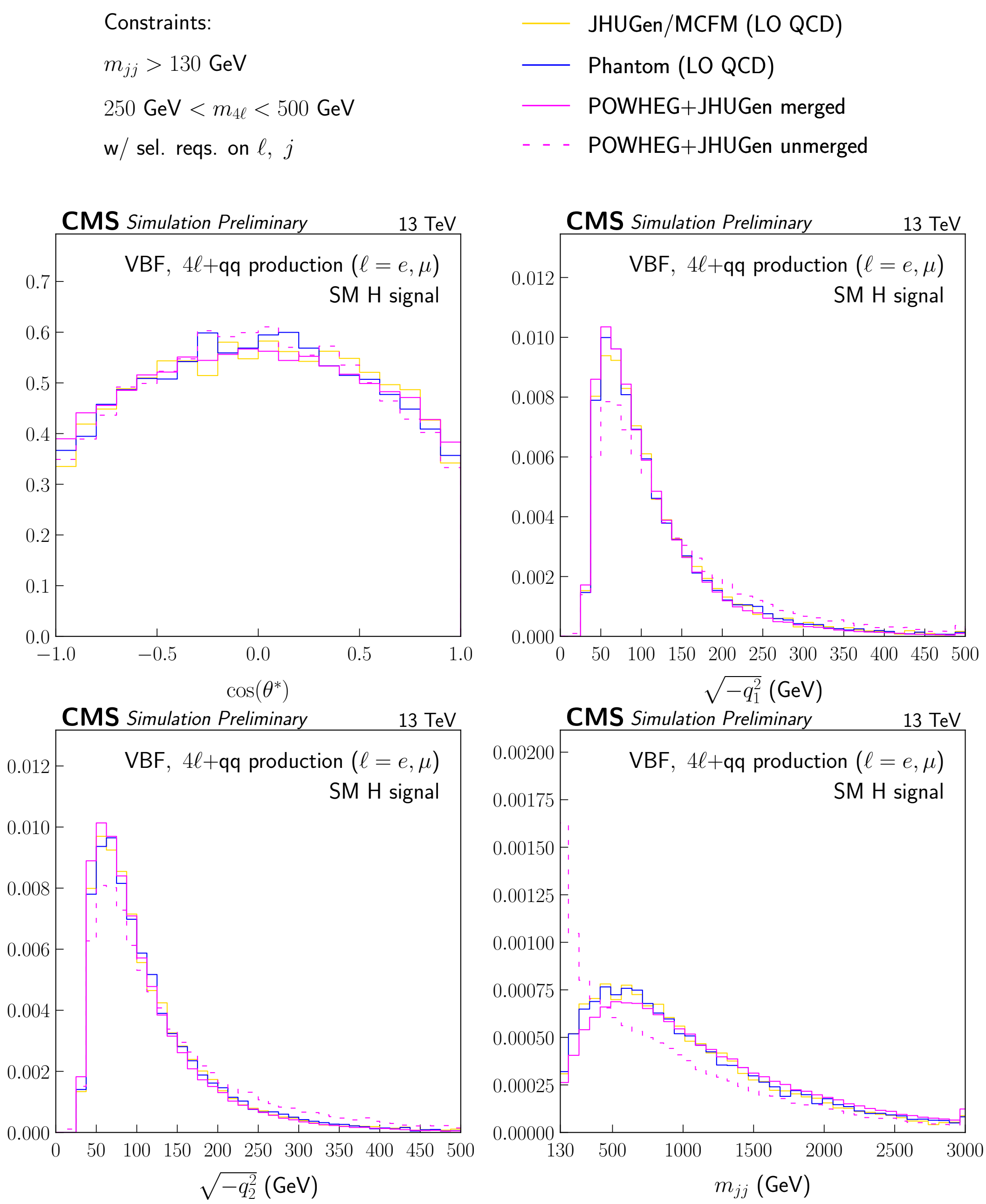

Figure 5.3: The distributions of $\cos\theta^*$, $\sqrt{-q_1^2}$, $\sqrt{-q_2^2}$, and m_{jj} for off-shell H boson production in the SM EW signal process. The requirement $m_{\mathrm{jj}} > 130$ GeV is applied on all distributions to emphasize the VBF-like topology. All distributions are normalized to unit area.

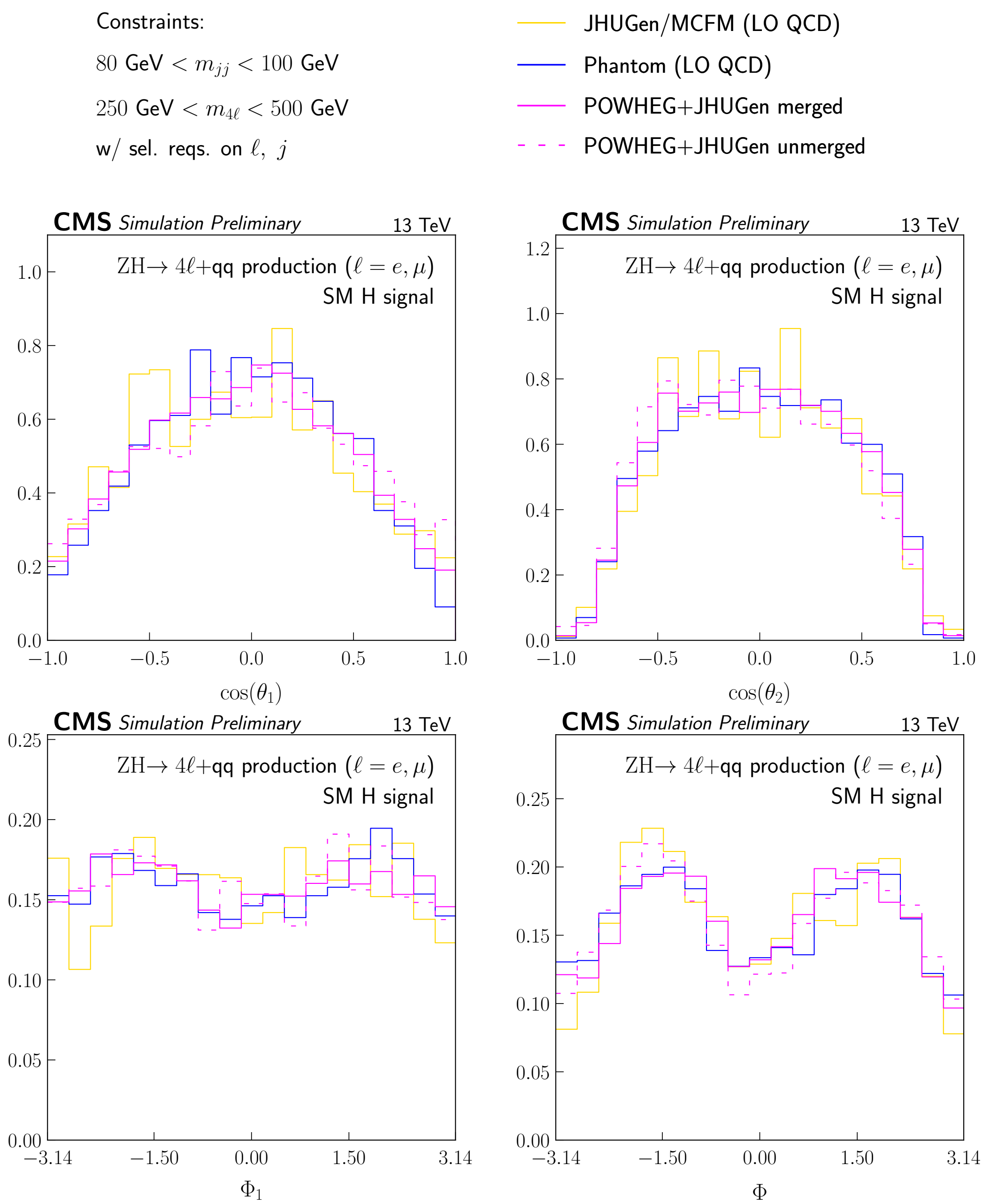

Figure 5.4: The distributions of $\cos\theta_1$, $\cos\theta_2$, Φ_1 and Φ for off-shell H boson production through the SM EW signal process. The requirement 80 GeV $< m_{\mathrm{jj}} <$ 100 GeV is applied along with requirements on the initial and final state composition to emphasize the ZH-like topology. All distributions are normalized to unit area.

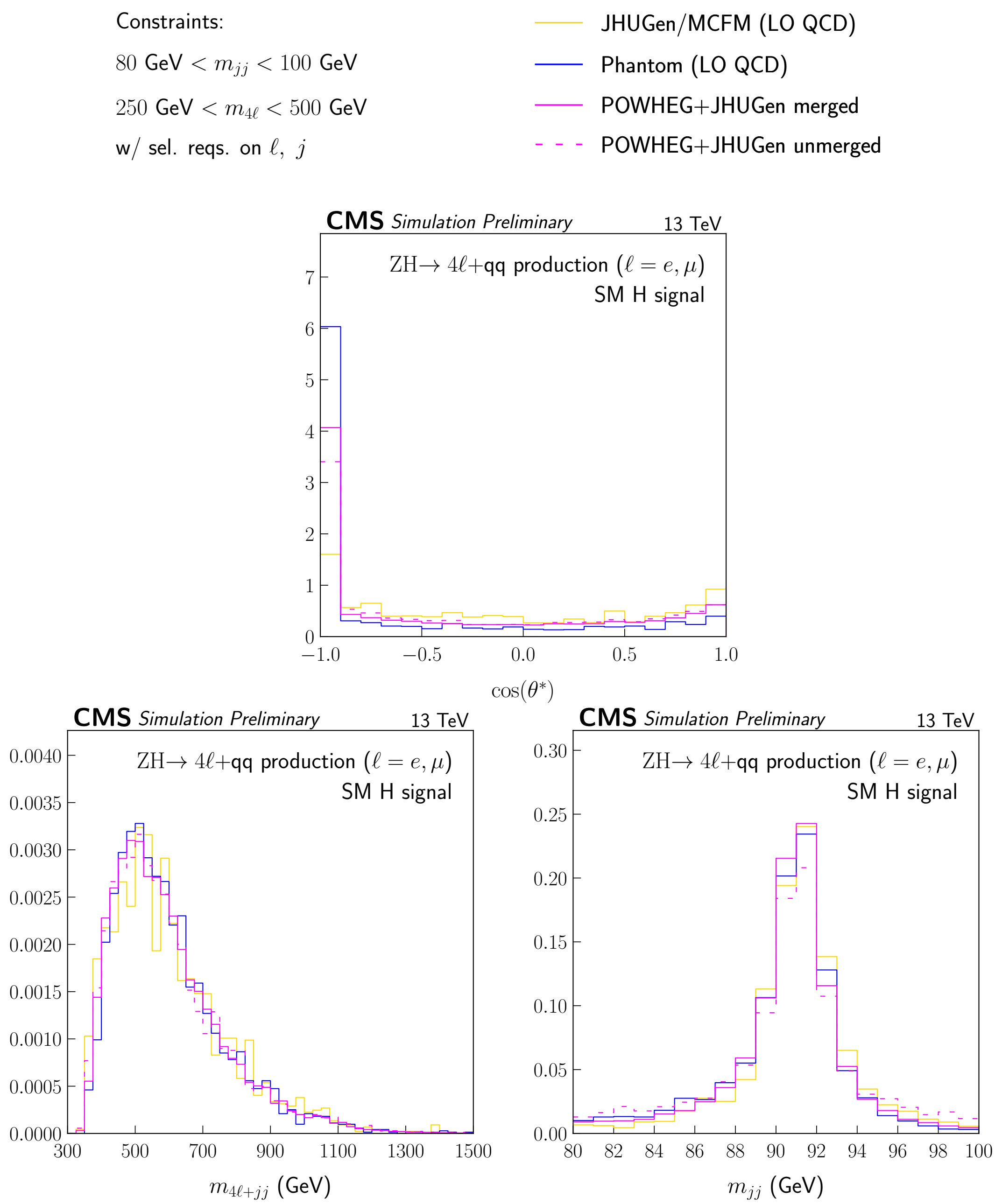

Figure 5.5: The distributions of $\cos\theta^*$, $m_{4\ell+jj}$ and m_{jj} for off-shell H boson production through the SM EW signal process. The requirement 80 GeV $< m_{\mathrm{jj}} <$ 100 GeV is applied along with requirements on the initial and final state composition to emphasize the ZH-like topology. All distributions are normalized to unit area.

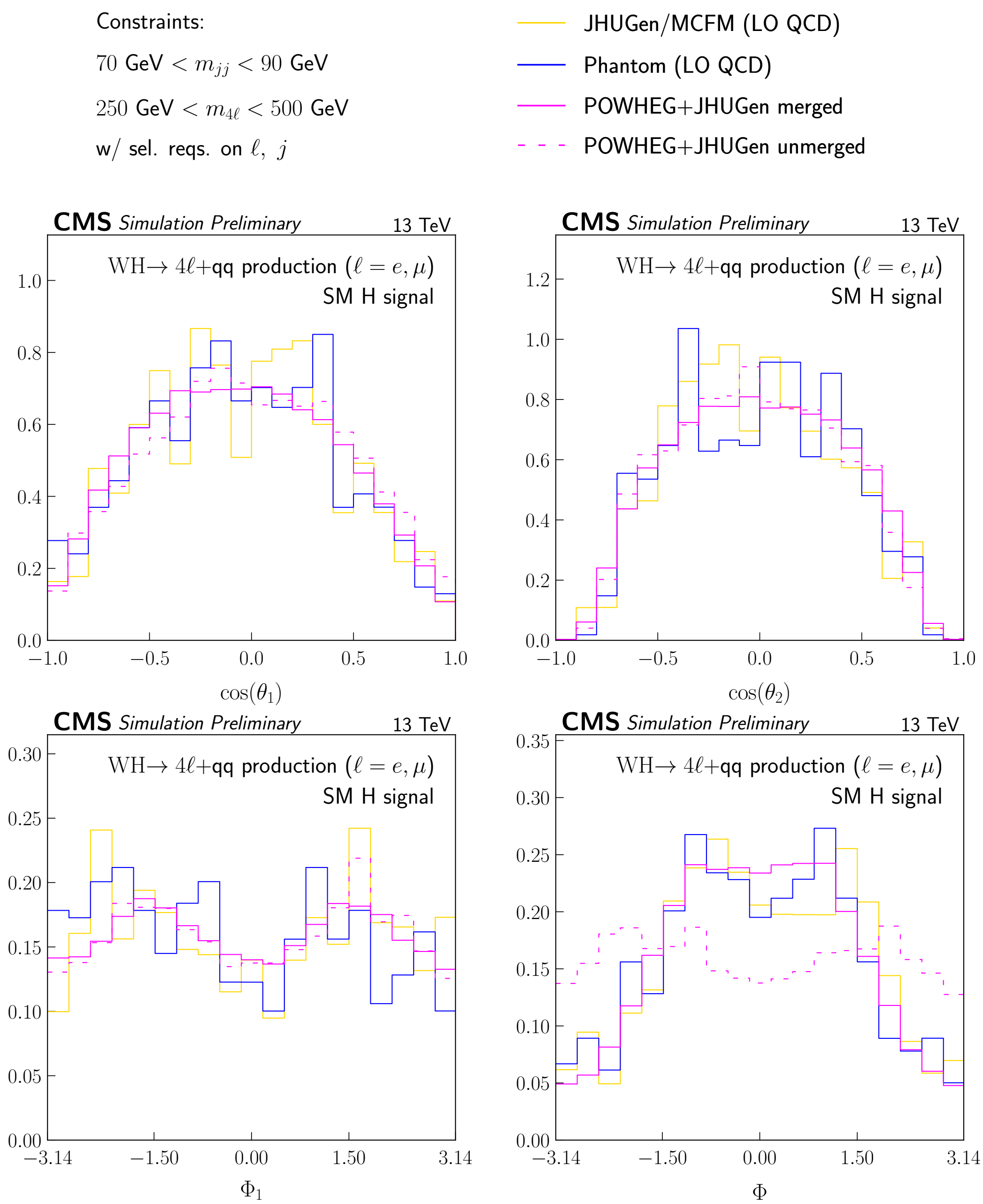

Figure 5.6: The distributions of $\cos\theta_1$, $\cos\theta_2$, Φ_1 and Φ for off-shell H boson production in the SM EW signal process. The requirement 70 GeV $< m_{\text{jj}} <$ 90 GeV is applied along with requirements on the initial and final state composition to emphasize the WH-like topology. All distributions are normalized to unit area.

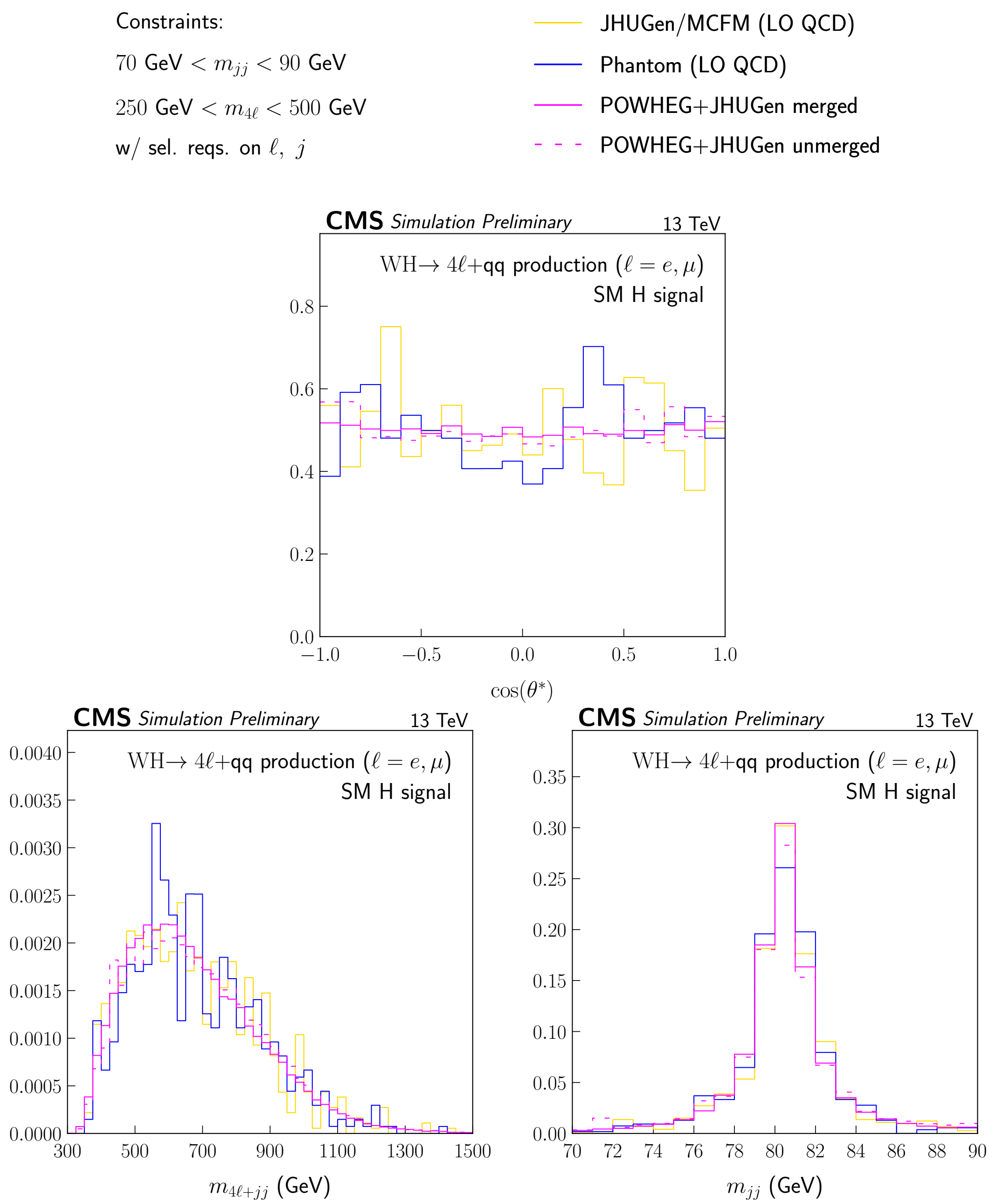

Figure 5.7: The distributions of $\cos\theta^*$, $m_{4\ell+jj}$ and m_{jj} for off-shell H boson production through the SM EW signal process. The requirement 70 GeV $< m_{\mathrm{jj}} <$ 90 GeV is applied along with requirements on the initial and final state composition to emphasize the WH-like topology. All distributions are normalized to unit area.

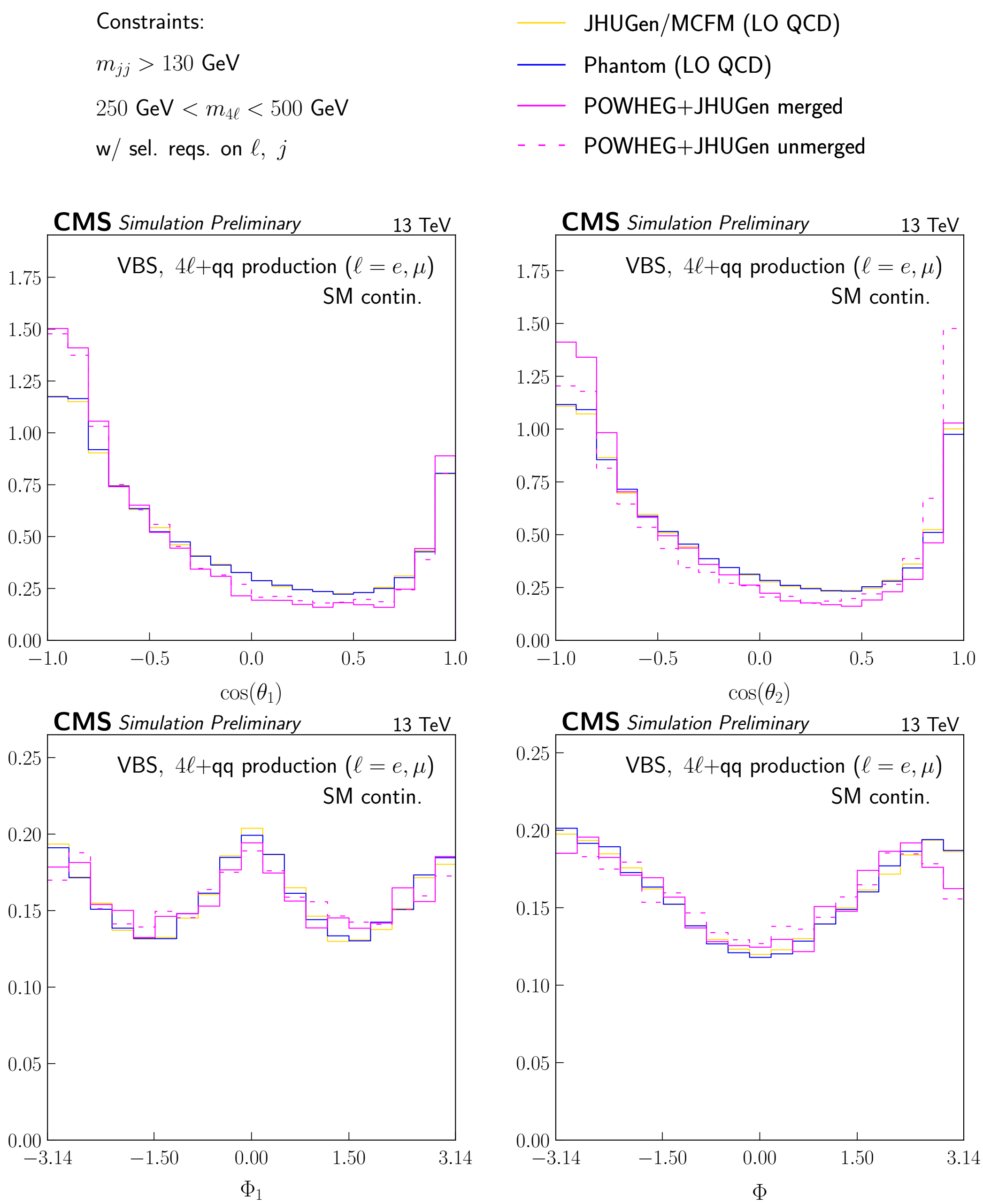

Figure 5.8: The distributions of $\cos\theta_1$, $\cos\theta_2$, Φ_1 and Φ for the SM EW continuum-only ZZ production process with 4ℓ production in the off-shell region. The requirement $m_{jj} > 130$ GeV is applied on all distributions to emphasize the VBF-like topology. All distributions are normalized to unit area.

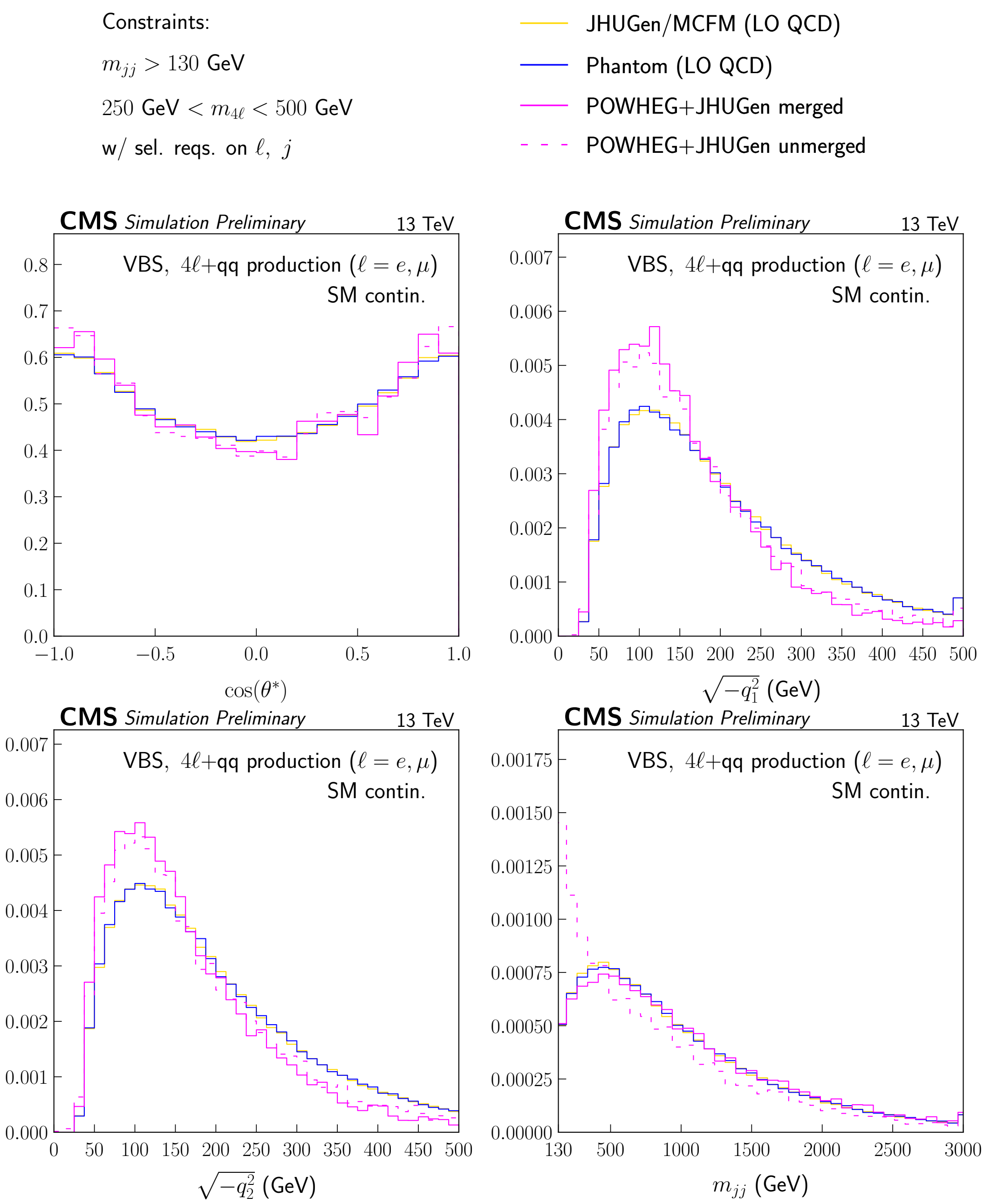

Figure 5.9: The distributions of $\cos\theta^*$, $\sqrt{-q_1^2}$, $\sqrt{-q_2^2}$, and m_{jj} (from top left to bottom right) for the SM The requirement $m_{\mathrm{jj}} > 130$ GeV is applied on all distributions to emphasize the VBF-like topology. All distributions are normalized to unit area.

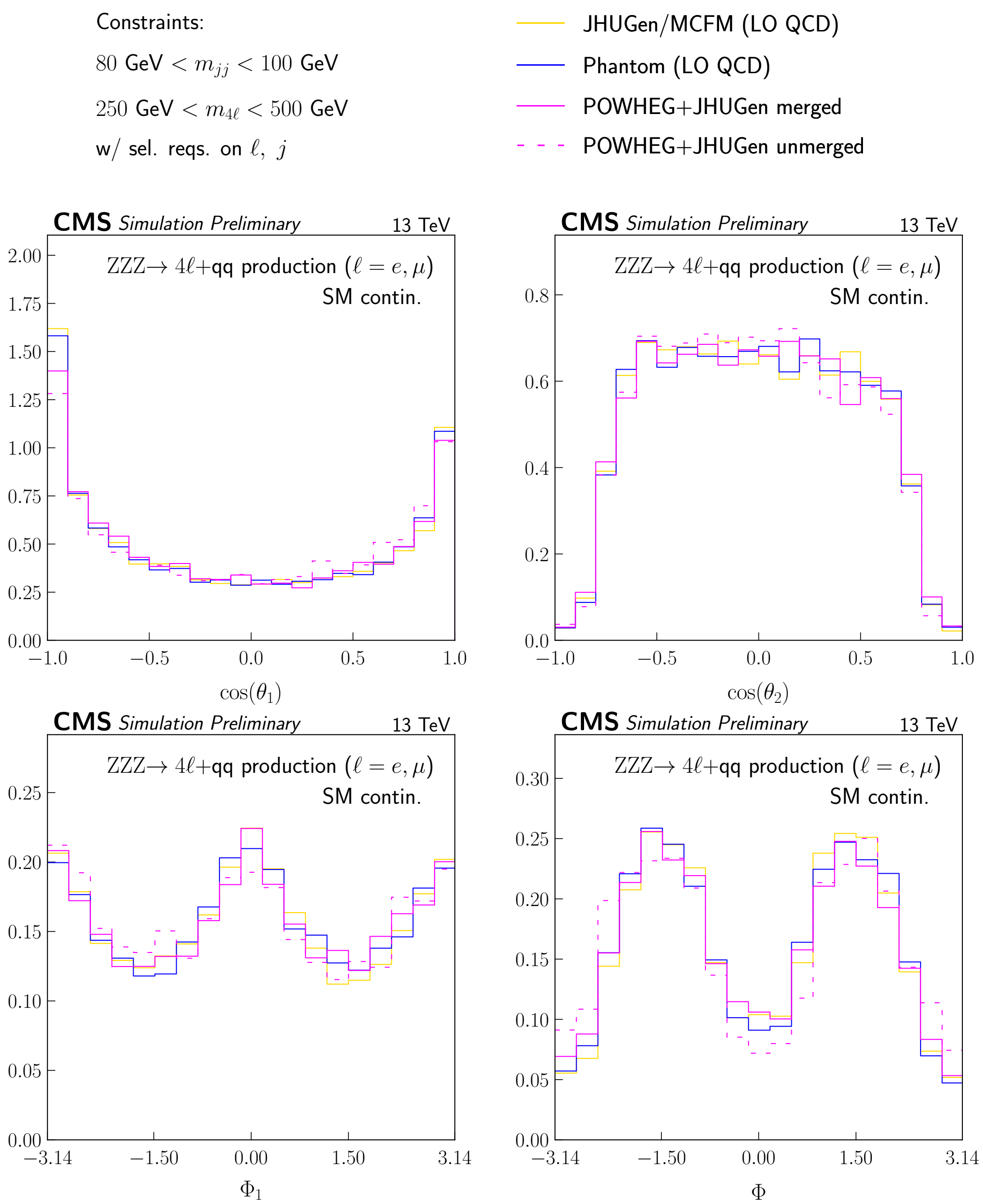

Figure 5.10: The distributions of $\cos\theta_1$, $\cos\theta_2$, Φ_1 and Φ for the SM EW continuum-only ZZ production process with 4ℓ production in the off-shell region. The requirement 80 GeV $< m_{\mathrm{jj}} <$ 100 GeV is applied along with requirements on the initial and final state composition to emphasize the ZZZ-like topology. All distributions are normalized to unit area.

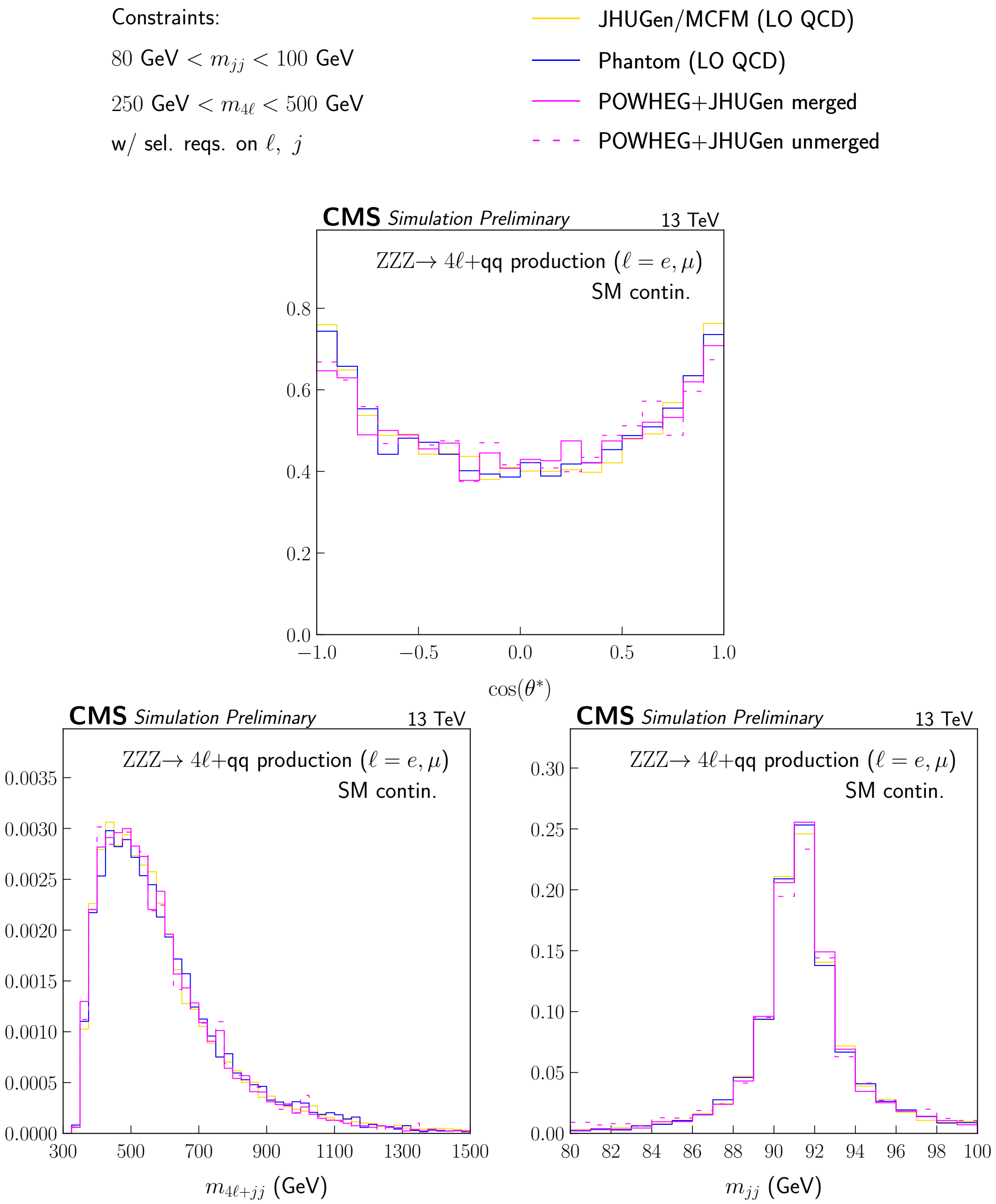

Figure 5.11: The distributions of $\cos\theta^*$, $m_{4\ell+jj}$ and m_{jj} for the SM EW continuum-only ZZ production process with 4ℓ production in the off-shell region. The requirement 80 GeV $< m_{\mathrm{jj}} <$ 100 GeV is applied along with requirements on the initial and final state composition to emphasize the ZZZ-like topology. All distributions are normalized to unit area.

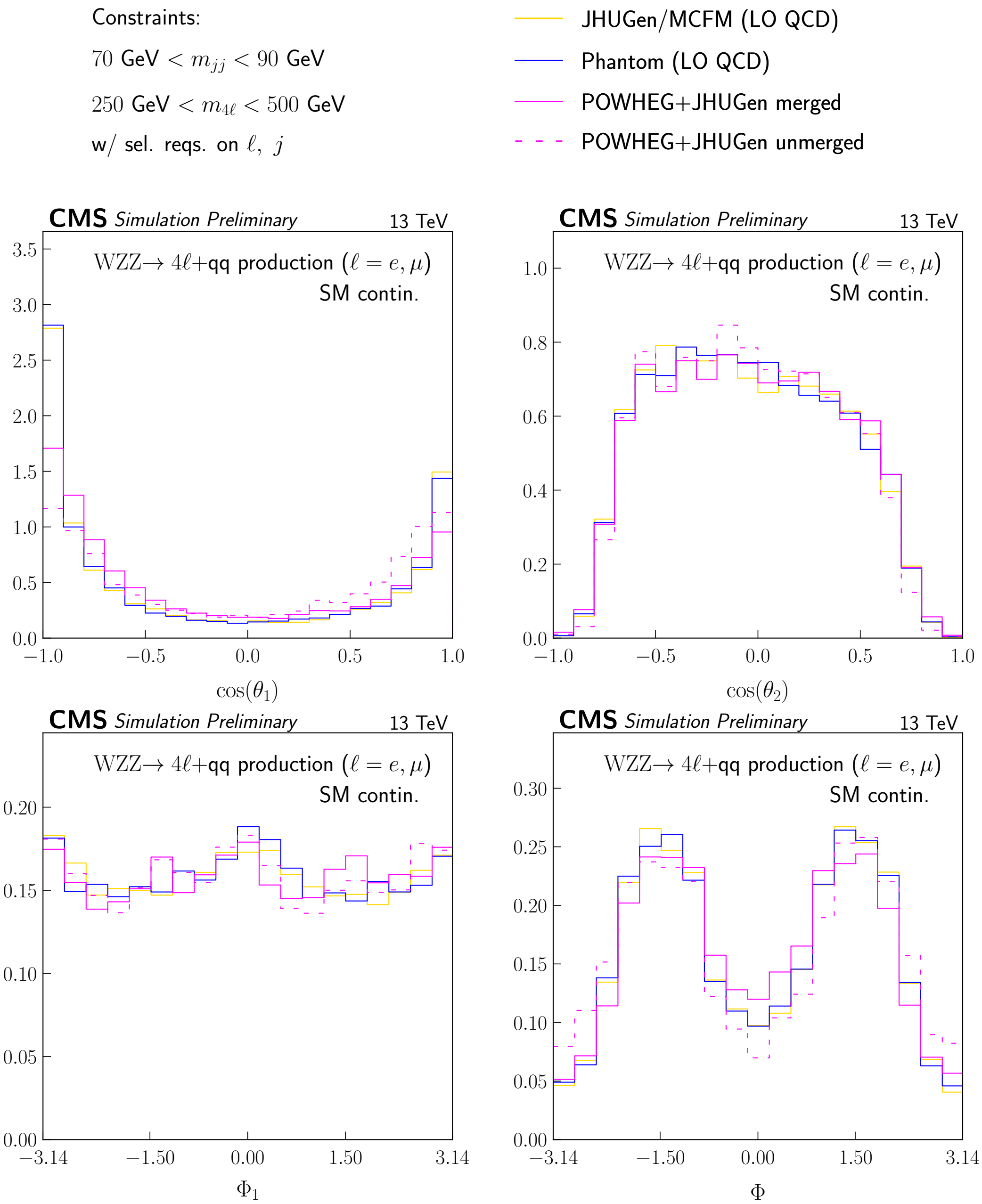

Figure 5.12: The distributions of $\cos\theta_1$, $\cos\theta_2$, Φ_1 and Φ for the SM EW continuum-only ZZ production process with 4ℓ production in the off-shell region. The requirement 70 GeV $< m_{\mathrm{jj}} <$ 90 GeV is applied along with requirements on the initial and final state composition to emphasize the WZZ-like topology. All distributions are normalized to unit area.

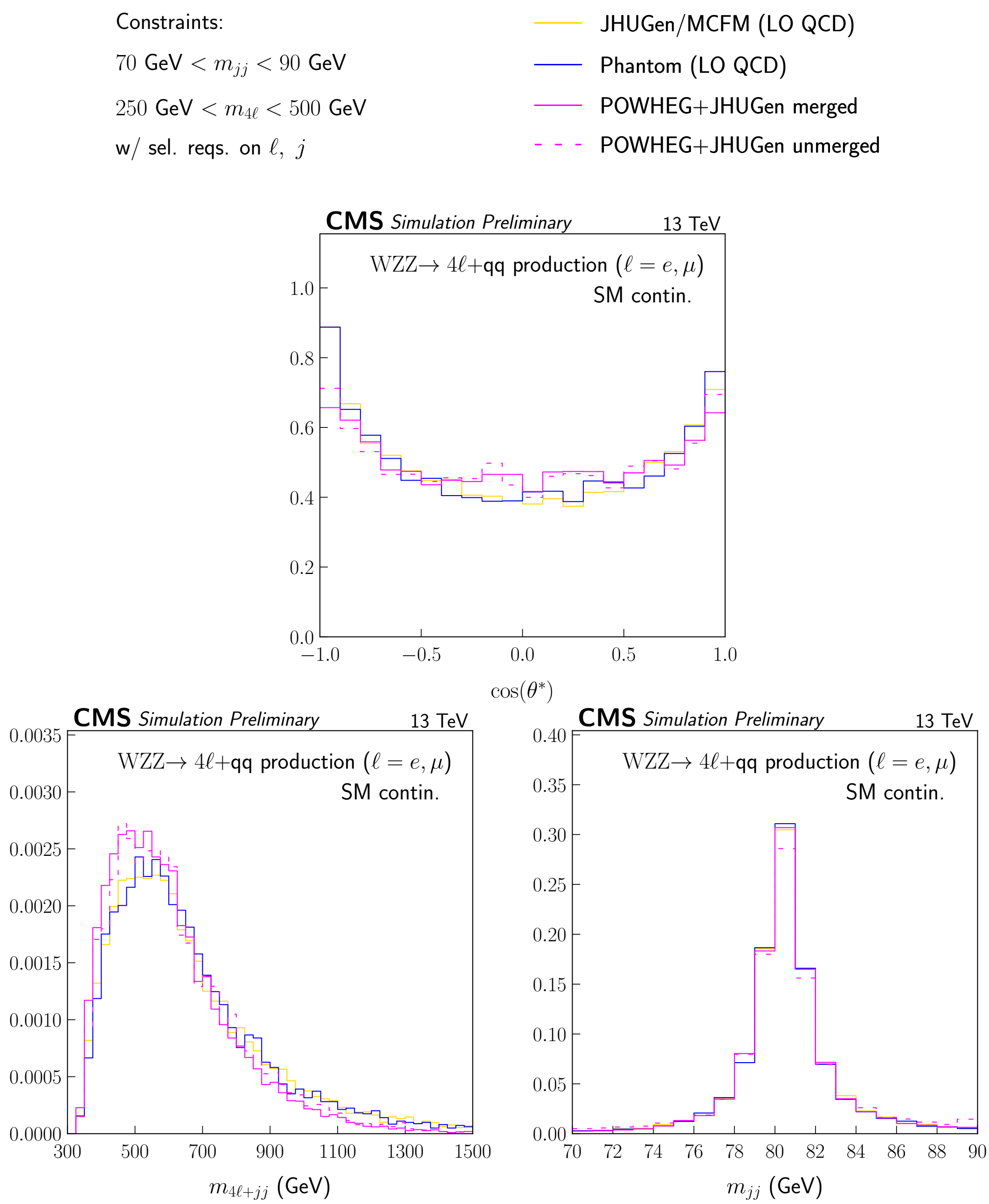

Figure 5.13: The distributions of $\cos\theta^*$, $m_{4\ell+jj}$ and m_{jj} for the SM EW continuum-only ZZ production process with 4ℓ production in the off-shell region. The requirement 70 GeV $< m_{\mathrm{jj}} <$ 90 GeV is applied along with requirements on the initial and final state composition to emphasize the WZZ-like topology. All distributions are normalized to unit area.

5.2.2.5 Comparisons after Parton Shower and Event Reconstruction

A fair comparison between LO and NLO samples should be made after parton showering (PS) since the PS is meant to compensate for the lack of higher-order terms in the event generation. This is the focus of the present section. The distributions of kinematic quantities are compared in different fractions of the phase space. Comparisons are made inclusive in jet multiplicity, as well as in the 0, 1 and ≥ 2 jet multiplicity bins. In the latter bin, particular attention is given to the off-shell region, and to regions with jet topologies that are characteristic of EW processes. To bring the distributions closer to the analysis conditions, kinematic observables are studied at reconstruction level, after applying an event preselection meant to emulate the preselection applied in the analysis of real data. The leptons (respectively the jets clustered according to the anti-k_T algorithm [49–51] with a $\Delta R = 0.4$ parameter) are required to satisfy $p_T \geq 7$ (respectively 30) GeV and $|\eta| < 2.4$ (respectively 4.7). An extra cut $m_{\ell\ell} > 4$ GeV on any charged lepton pair with same flavor and opposite charge is applied to reject events with low-mass resonances. Comparisons in the gg process also feature the relevant k-factors mentioned in Sec.5.2.2.2, to scale the contributions to N^3LO in QCD inclusively.

In order to match the predicted cross sections by POWHEG/JHUGen and JHUGen/MCFM at $|m_{VV} - 125 \text{ GeV}| < 0.05$ GeV, the JHUGen/MCFM samples are scaled by overall normalization factors. These overall normalization factors are computed to be 1.0 and 1.09 for GF and EW processes respectively. In GF case, this factor is obtained from the ratio of the SM signal cross sections after the quoted requirements on charged leptons; And for EW processes the ratio of the sums of the SM signal cross sections for the VBF, ZH, and WH processes are used in which the lepton and quark selection requirements are the same as what is used at hard process level described in Sec. 5.2.2.4.

In the comparisons shown below, we consider the following set of theoretical uncertainties in the same ways they are considered in Refs. [93, 96]:

- Renormalization scale: The renormalization scale is varied by a factor of 0.5 or 2, and the effect is included through per-event weights. In the case of the GF process, the weights are applied such that the relative variations resulting from these weights are scaled as a function of true m_{VV} to those for the inclusive NNLO k-factor variation [91].
- Factorization scale: This source is considered to be uncorrelated from other uncertainty sources, and the scale is varied in the same way as the renormalization scale. The GF process is adjusted to match the inclusive NNLO k-factor variations in the same way as above.
- $\alpha_s(m_Z)$: A variation of $\alpha_s(m_Z) = 0.118 \pm 0.0015$ is considered. The variations

for the GF process are adjusted to match those for the inclusive k-factor in the same way as above.

- PDF variations: This variation is taken as a conservative, envelope-type variation evaluated on a per-event basis over 100 variations from NNPDF [73] using the Hessian method. The gg process is adjusted to match the inclusive k-factor variations in the same way as above.
- Scale variations in parton shower: The scale variations are taken from additional per-event weights calculated from PYTHIA. The variation multiplicative factors considered are 0.25 and 4. The weights for the variations of initial- (ISR) and final-state radiation (FSR) are calculated separately, and the "down" and "up" variations are obtained by multiplying the ISR and FSR weights computed with the factors of 0.25 and 4, respectively.
- Simulation of the second jet in the GF samples: The uncertainty is evaluated as the difference of the nominal POWHEG samples generated at the pole masses $m_{\mathrm{H}} = 125$ GeV and $m_{\mathrm{H}} = 300$ GeV from the simulation with the MiNLO HJJ program [97], generated at the same pole masses. The reweighting factors are parametrized in three dimensions, in bins of m_{VV} below or above 150 GeV, in bins of $p_{\mathrm{T}}^{\mathrm{VV}}/m_{\mathrm{VV}}$, evaluated for the hard process, and in bins of 0, 1, and ≥ 2 in the number of jets ($\mathrm{N_j}$), with the jet definition as aforementioned.
- Approximation of the NNLO QCD corrections to the continuum gg $\rightarrow$ ZZ contribution: Because the signal k-factor is applied on all contributions in the GF process, the continuum ZZ contribution is multiplied by a factor $k_{\mathrm{gg}} = 1 \pm 0.1$. The interference between the continuum amplitude, and the SM or BSM signal amplitudes is multiplied by $\sqrt{k_{\mathrm{gg}}}$ accordingly. The uncertainty of 0.1 is determined based on the level of disagreement observed between the amounts of corrections the SM signal and continuum ZZ contributions receive at approximately NLO in QCD [98].

It should be noted that in the error bands shown in the subsequent comparisons does not include PYTHIA tune uncertainties and the theoretical uncertainties in the JHUGen/MCFM samples are also not considered. On top of the aforementioned theoretical uncertainties the statistical uncertainties, computed simply from the sums of squared weights, are also added in quadrature for illustration in the displayed uncertainty bands. Furthermore, in the following comparisons the H boson is either a SM particle or a pure pseudoscalar (PS) which corresponds to anomalous coupling $a_3 = 1$.

Figures 5.15 and 5.16 show the comparison of the m_{VV} spectra inclusive in jet kinematics, and either inclusive or split in different $\mathrm{N_j}$ categories. No significant disagreement is observed in the inclusive m_{VV} line shape of the GF process, and the discrepancies in the transition region between 130 GeV and 200 GeV are understood to arise from the lack of statistics in the reweighted sample, which could be improved if samples with a finer binning in m_H were used. In contrast, the sum of the reweighted POWHEG EW production samples predicts more events than its JHUGen/MCFM counterpart in the off-shell region, especially in the mass interval 200-500 GeV. This observation is made even in the SM signal hypothesis, shown in red in the figure. For this hypothesis, apart the from different propagator model, the reweighting does not affect the kinematics of the events simulated at NLO by POWHEG. The difference indicates that the LO samples with parton shower do not approximate the EW processes very well, in particular in the low jet multiplicities, and also inclusively. When split in exclusive $\mathrm{N_j}$ categories, disagreements from event migration across jet multiplicities, effects are observed in both production mechanisms.

We also examine the comparisons in the $\mathrm{N_j} \geq 2$ category more extensively by splitting the events further into VBF-like topologies i.e. the events with $m_\mathrm{jj} \geq 130$ GeV and $|\Delta\eta_\mathrm{jj}| \geq 3$, and VH-like topologies i.e. the events with 60 GeV $\leq m_\mathrm{jj} < 130$ GeV and $|\Delta\eta_\mathrm{jj}| < 3$. It is important to note that the GF production mode is one of the dominant modes contaminating the VH processes, so it is useful to look at the behavior of this process just like the target EW production mode. As can be seen from Fig. 5.17, in the LO samples shows less (respectively more) events as compared to the NLO samples in the $\mathrm{N_j} \geq 2$, VH- (VBF-) like topology below (above) 350 GeV . For most of these regions, the disagreements are beyond the uncertainties predicted from the POWHEG simulation.

In the EW process, while the VBF-like topology shows good agreement, the LO samples slightly over-predict the contribution from the continuum below 600 GeV , lying beyond the uncertainties in the POWHEG prediction. In the VH-like topology, however, the region below 600 GeV is severely under-predicted by the LO samples. The source of this under-prediction can be partially attributed to the requirements placed at the generation of the LO samples on the outgoing quarks from the hard process $p_\mathrm{T}^\mathrm{q} > 15$ GeV , $|\eta_\mathrm{q}| < 6.5$, $\Delta R_\mathrm{qq} > 0.3$, and $m_\mathrm{qq} > 30$ GeV . In order to examine how much the disagreement improves, we produce an alternative set of comparisons in Figs. 5.18 and 5.19 by placing the tighter requirements of $p_\mathrm{T}^\mathrm{q} > 30$ GeV , $|\eta_\mathrm{q}| < 4.7$, $\Delta R_\mathrm{qq} > 0.4$, and $m_\mathrm{qq} > 50$ GeV . The kinematics of quarks in the case of the POWHEG samples are corrected based on the gluon merging scheme to emulate the LO topology, as described in Sec. 5.2.2.3. We observe that even after these tighter requirements, disagreements in the $\mathrm{N_j} = 0$ and 1 categories remain even though they are reduced. The disagreement in the VH-like topology within the $\mathrm{N_j} \geq 2$ category improves only slightly, and the distributions in the VBF-like topology remain very similar. Therefore, for the rest of the discussion, we do not apply the tighter requirements on the quarks in order to observe the full extent of where disagreements could also arise in other kinematic quantities in the $\mathrm{N_j} \geq 2$ category, observing also that these additional requirements do not make substantial difference in the distributions.

It is instructive to examine the distributions of m_{jj} in different bins of m_{VV} in order to understand the evolution of these variables in the off-shell region. From Figs. 5.20 and Figs. 5.21, one might notice that the m_{jj} spectra for the GF production regions an unphysical evolution going from low to high $\Delta\eta_{\mathrm{jj}}$ regions is visible for the LO samples. This kind of unphysical evolution is also evident by comparing the m_{jj} spectra from low to high m_{VV} for the LO samples. while the NLO (POWHEG) spectra show more stable evolution. and relying on parton shower alone in the LO samples does not produce stable results across different m_{VV} bins. For the EW process, one can see from Figs. 5.22 and 5.23 that the shape of m_{jj} spectra are different between the two samples, most apparently above 100 GeV in $|\Delta\eta_{\mathrm{jj}}| < 3$, and in the region between 130 GeV and 300 GeV in $|\Delta\eta_{\mathrm{jj}}| \geq 3$. The differences also evolve as a function of m_{VV}. We note especially the over-prediction of the m_{jj} tail in the LO simulation in the $|\Delta\eta_{\mathrm{jj}}| < 3$ region, resulting from the QCD jets after parton shower.

One can also examine the distributions of $|\Delta\eta_{\mathrm{jj}}|$ in different bins of m_{VV} to understand the source of these discrepancies. These distributions are shown for the GF production in Figs. 5.24 and 5.25, and for the EW production in Figs. 5.26 and 5.27. In the case of GF production, LO samples seem to show a deficit of small-$|\Delta\eta_{\mathrm{jj}}|$ dijet events at low m_{VV} values in both VH or VBF-like m_{jj} regions, which disappears at higher m_{VV} values. In the case of the EW production, the LO samples seem to agree better in shape at lower m_{VV} values for the VH-like topology even though there is a disagreement in the overall normalizations. At higher m_{VV} values, the LO samples seem to predict more events above 0.5 considering the shape features. For the VBF-like topology, there is an overall disagreement on the amount of central dijet events.

The disagreement in central jets can also be observed by examining the Zeppenfeld z_3 variable that quantifies the centrality of the third jet in the event, which is defined as

$$z_3 = \frac{y_3 - (y_1 + y_2)/2}{|\Delta y_{12}|}. \tag{5.5}$$

In this definition, the indices 1, 2 and 3 refer to the index of the jets ordered in descending order in p_{T}, and the variable y is the rapidity of the jet. The value of z_3 is closer to 0 for central jets. The distribution of this variable in the $m_{\mathrm{jj}} \geq 130$ GeV region are shown for the EW production in Fig. 5.28. While it is known that the prediction of POWHEG with default PYTHIA shower settings also overshoots at around $z_3 \sim 0$ and that using the dipole recoil setting improves the performance [99], the level of difference in the case of POWHEG is expected to be small. While CMS analyses have not been using the third jet explicitly, it is possible to obtain even better performance for the POWHEG samples by switching to the dipole recoil scheme. The JHUGen/MCFM samples show poor performance on this quantity; this performance could be improved potentially by deploying the dipole recoil scheme in PYTHIA shower settings.

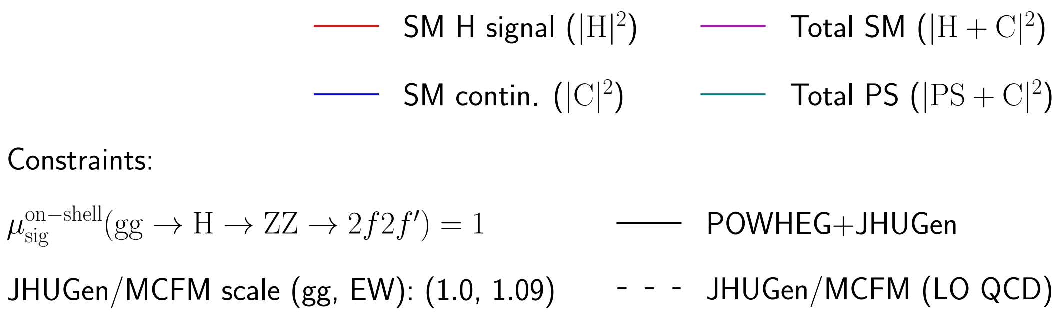

Figure 5.14: The different hypotheses are listed in the legend which holds for all plots in this section. The green color line corresponds to the total distribution of a pure pseudoscalar (PS) which corresponds to anomalous coupling $a_3 = 1$.

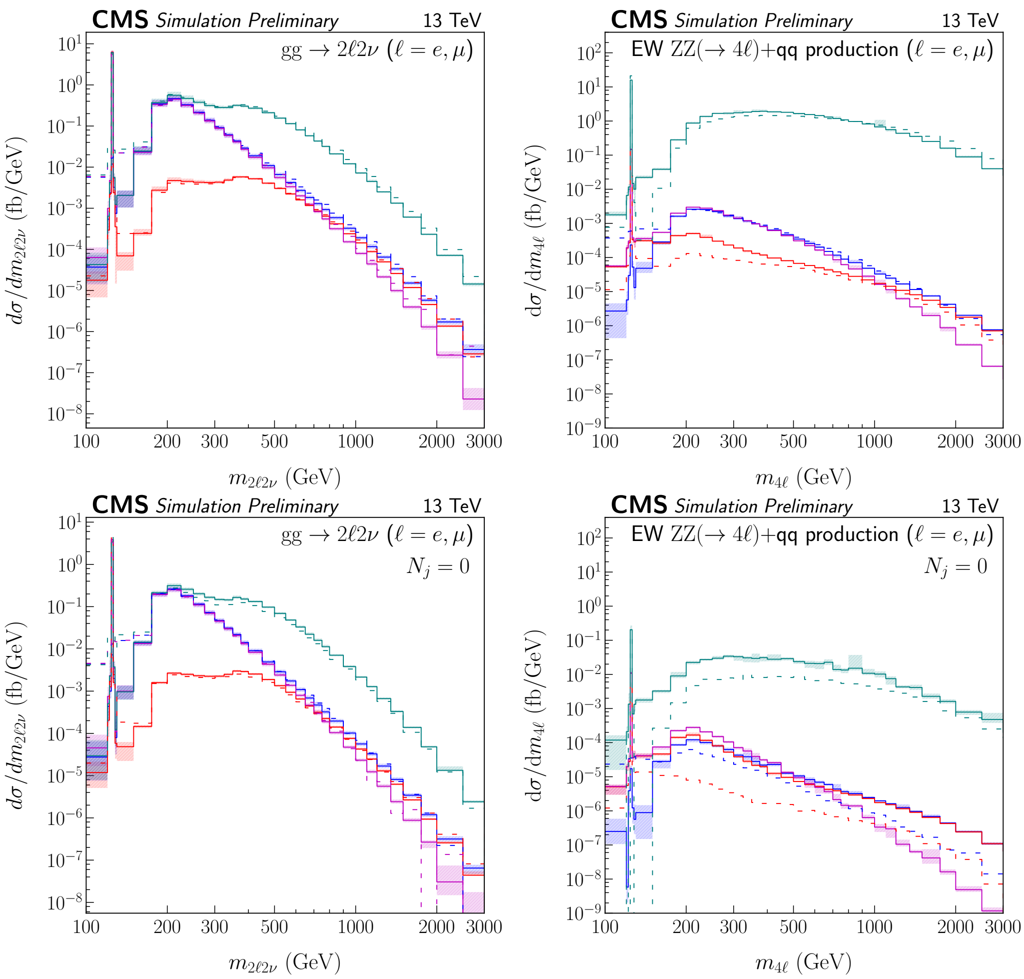

Figure 5.15: The distributions of $m_{2\ell 2\nu}$ in the gg (GF) and $m_{4\ell}$ in the EW production modes are shown from top to bottom as inclusive in jet bins and $\mathrm{N_j} = 0$ respectively. The different hypotheses are listed in the legend showed in Fig. 5.14.

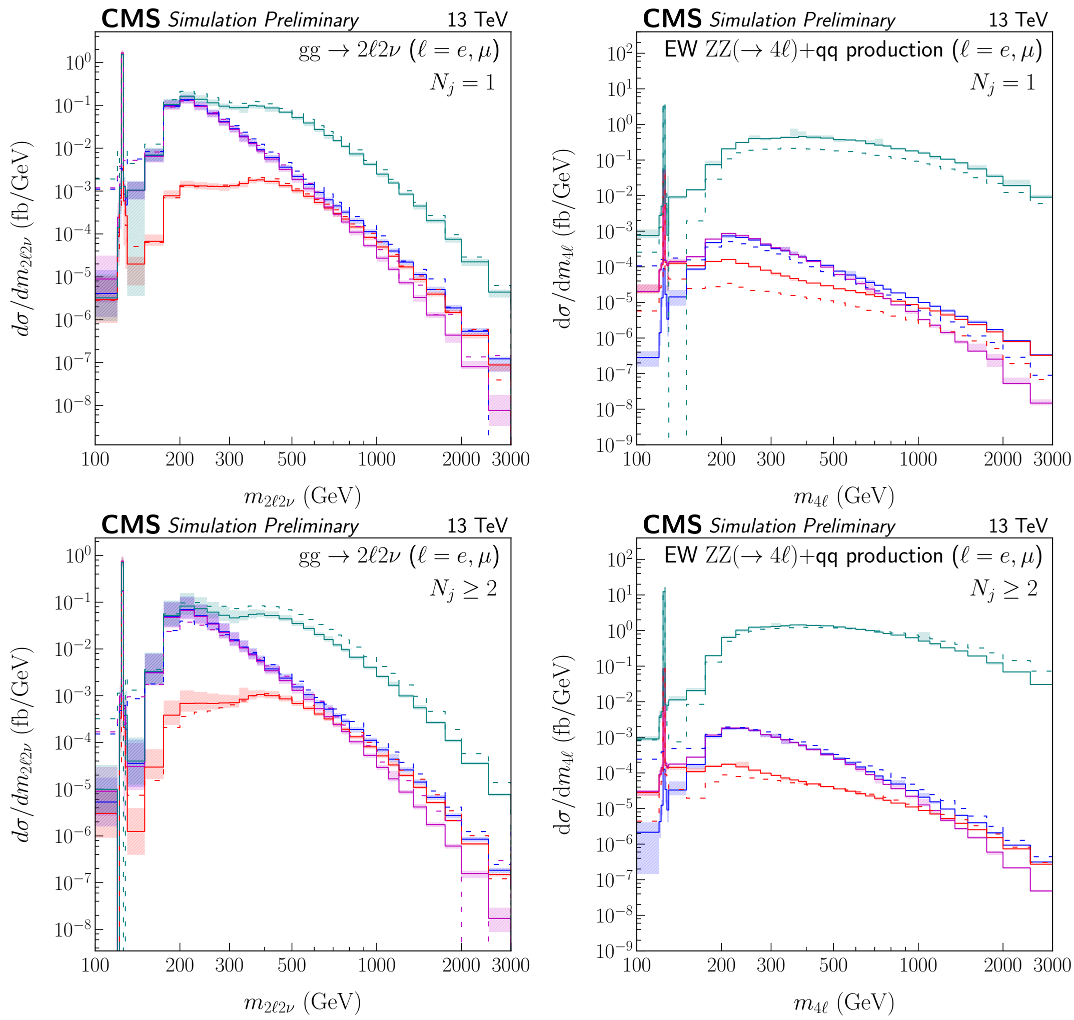

Figure 5.16: The distributions of $m_{2\ell 2\nu}$ in the gg (GF) and $m_{4\ell}$ in the EW production modes are shown from top to bottom as $\mathrm{N_j} = 1$ and ≥ 2 respectively. The different hypotheses are listed in the legend showed in Fig. 5.14.

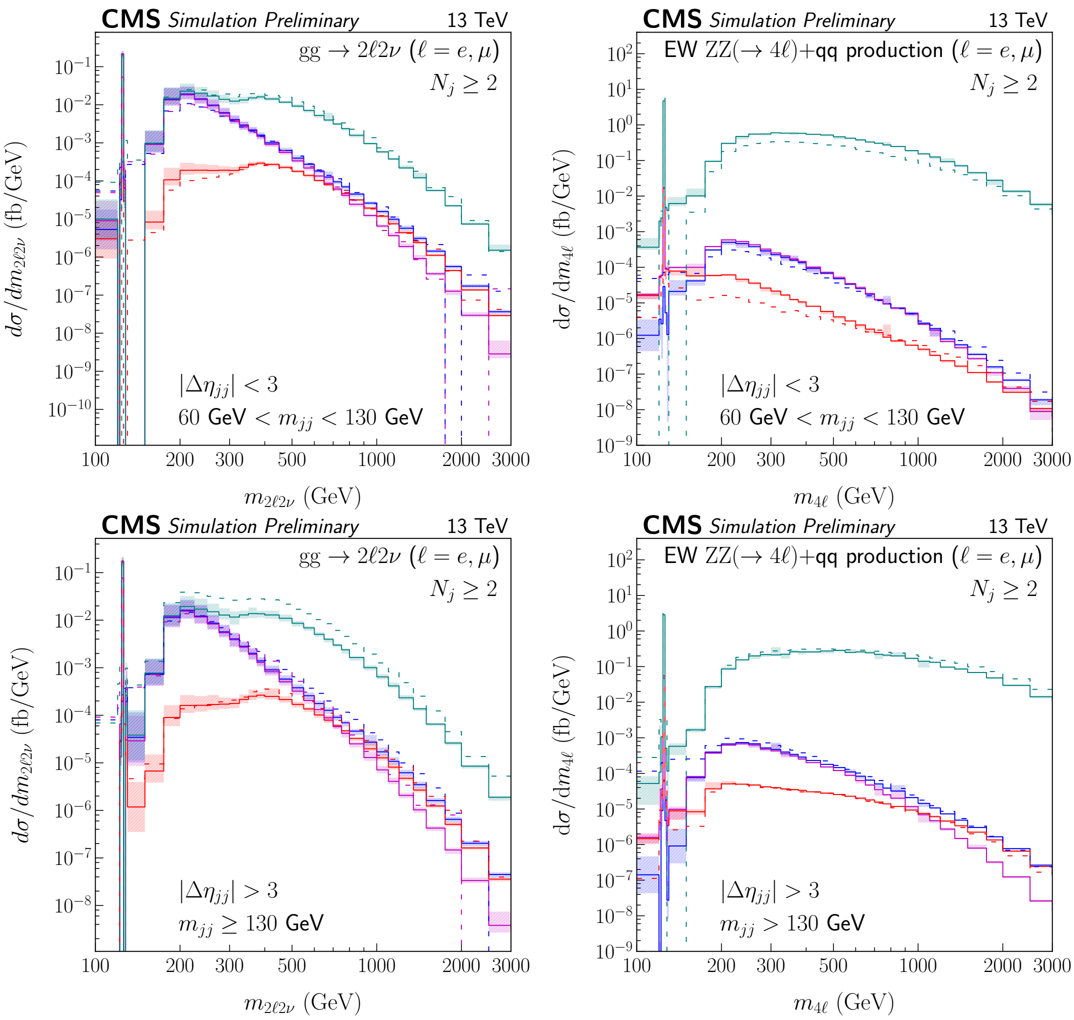

Figure 5.17: The distributions of $m_{2\ell2\nu}$ in the gg and $m_{4\ell}$ in the EW production modes are shown for $\mathrm{N_j} \geq 2$. The different hypotheses are listed in the legend showed in Fig. 5.14. The top panels are shown for events that are in a VH-like topology with 60 GeV $\leq m_{\mathrm{jj}} < 130$ GeV and $|\Delta\eta_{\mathrm{jj}}| < 3$, and the bottom panels are shown in a VBF-like topology with $m_{\mathrm{jj}} \geq 130$ GeV and $|\Delta\eta_{\mathrm{jj}}| \geq 3$.

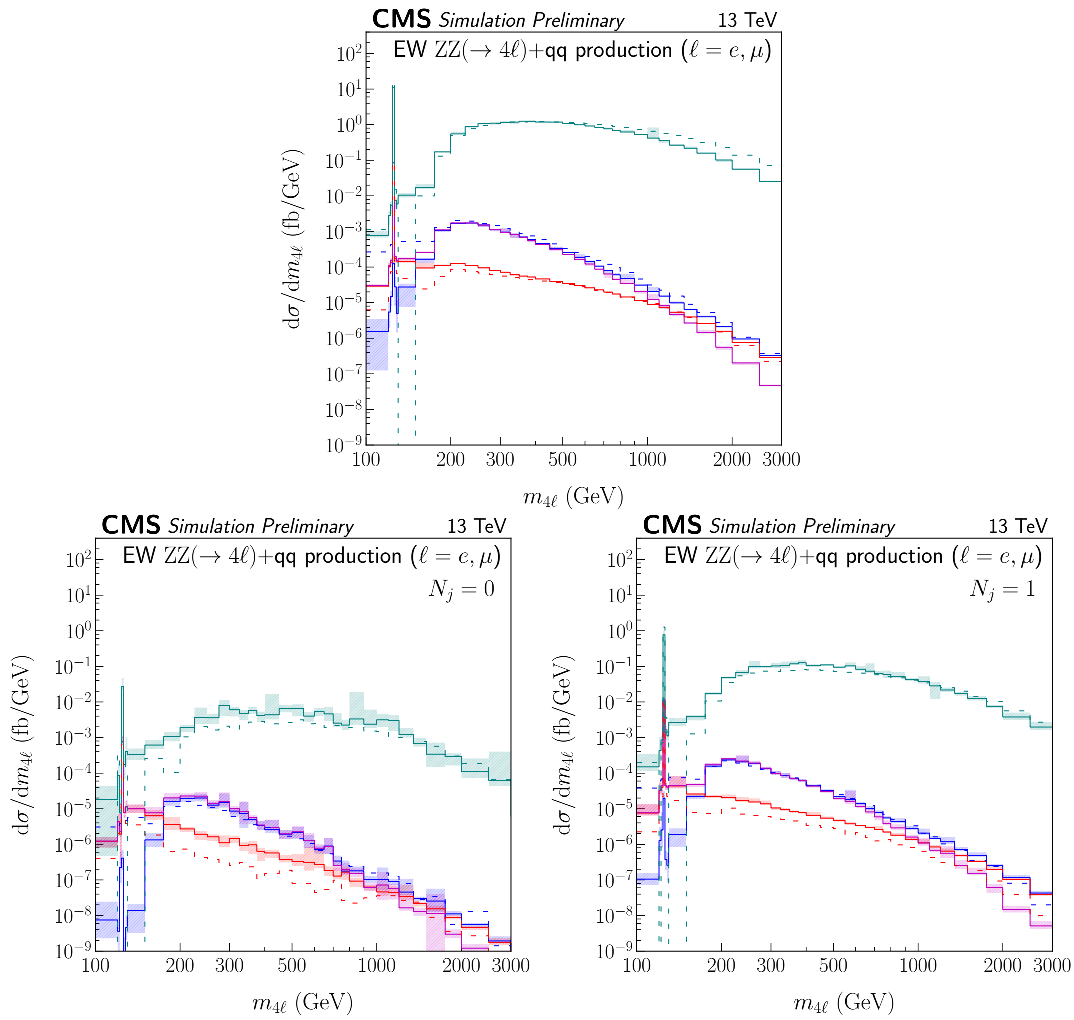

Figure 5.18: The distributions of $m_{4\ell}$ are shown in the EW production mode after the tighter selection requirements on the quarks from the LO topology. The different hypotheses are listed in the legend showed in Fig. 5.14. The top plot shows the distributions inclusive in $\mathrm{N_j}$, and the bottom plots show the distributions in the $\mathrm{N_j} = 0$ and 1 categories from left to right respectively.

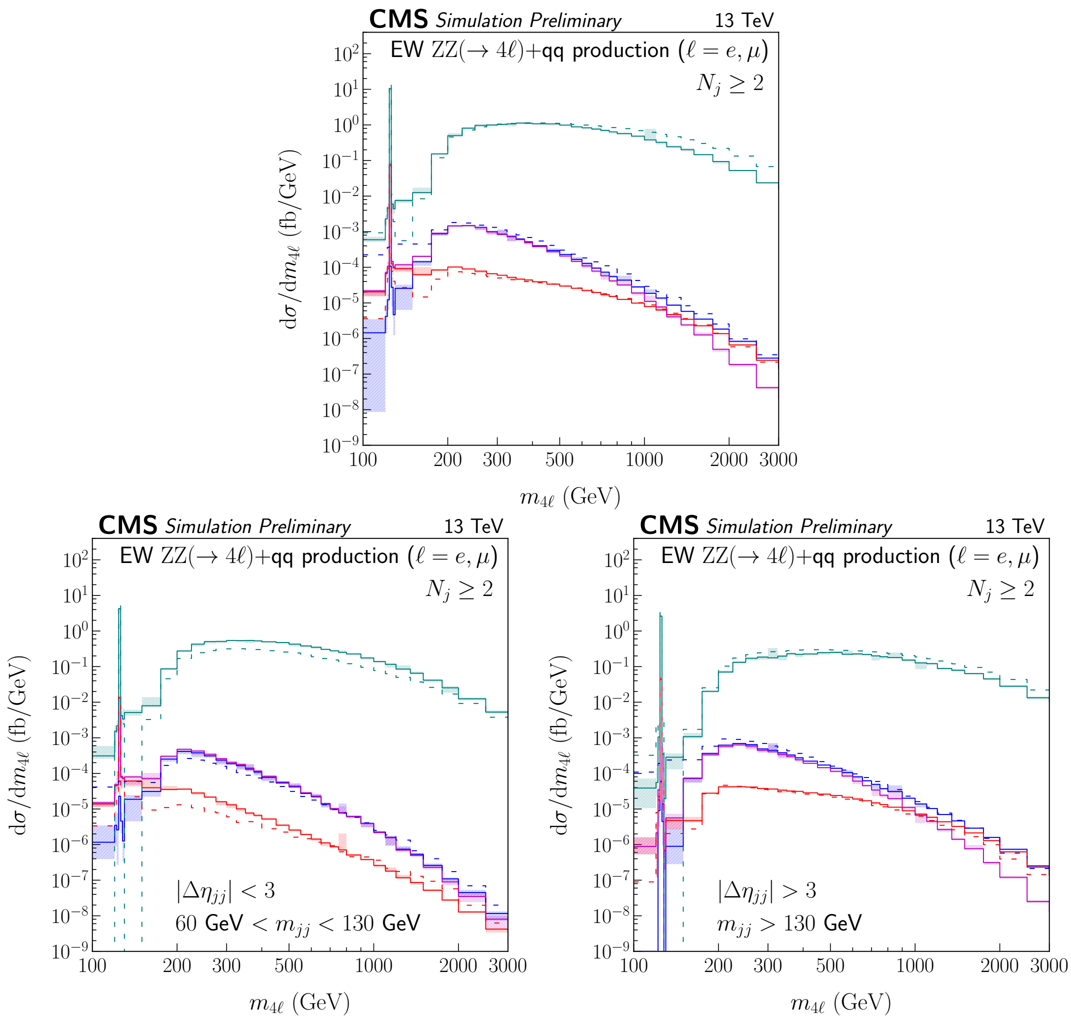

Figure 5.19: The distributions of $m_{4\ell}$ are shown in the EW production mode after the tighter selection requirements on the quarks from the LO topology. The different hypotheses are listed in the legend showed in Fig. 5.14. The top plot shows the total predictions for the $\mathrm{N_j} \geq 2$ category, and the bottom plots show the VH- and VBF-like topologies from left to right respectively within the $\mathrm{N_j} \geq 2$ category.

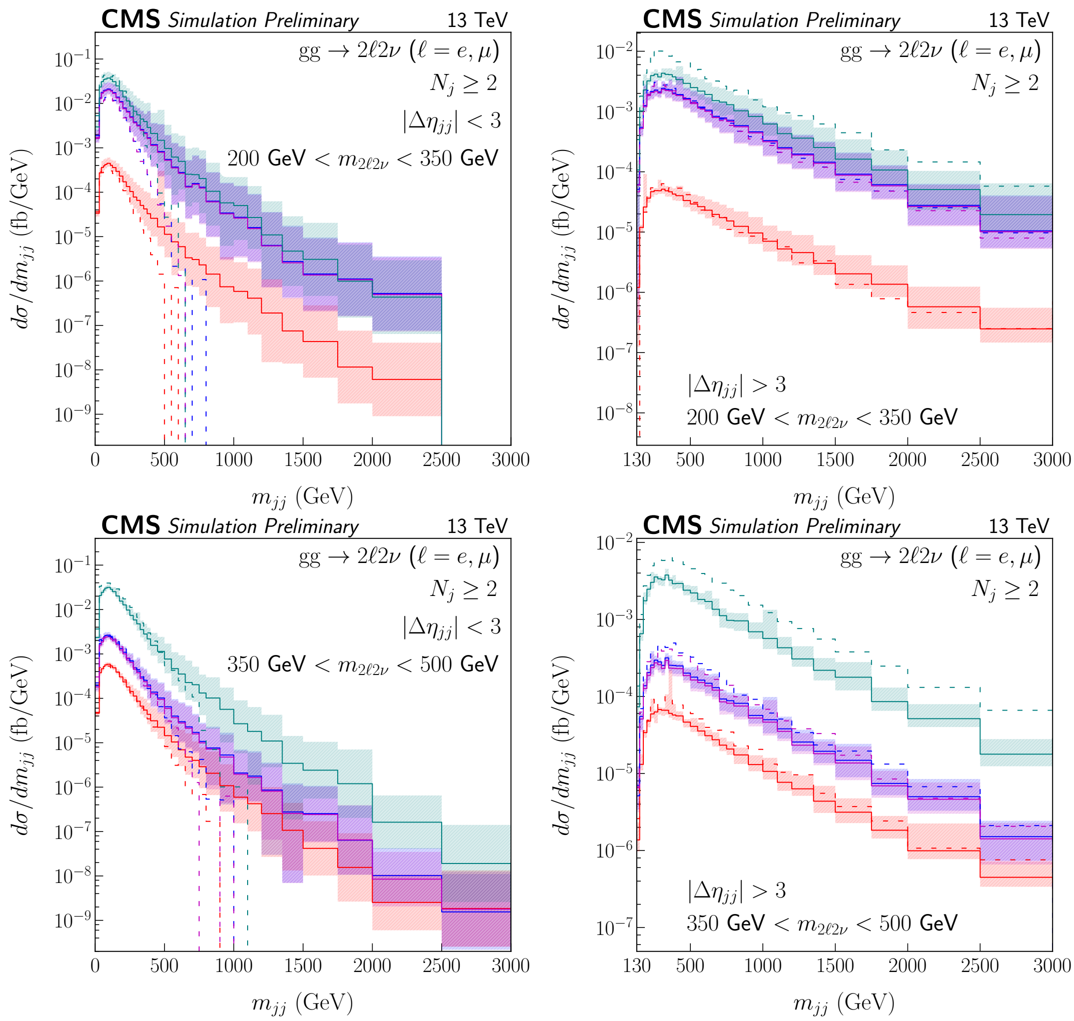

Figure 5.20: The distributions of m_{jj} are shown in the gg production in the $\mathrm{N_j} \geq 2$ category. The left panels show the spectra with $|\Delta\eta_{\mathrm{jj}}| < 3$, and the right panels show the spectra with $|\Delta\eta_{\mathrm{jj}}| \geq 3$. The panels are ordered from top to bottom in the $m_{2\ell 2\nu}$ bins of 200–350 GeV and 350–500 GeV . The different hypotheses are listed in the legend showed in Fig. 5.14.

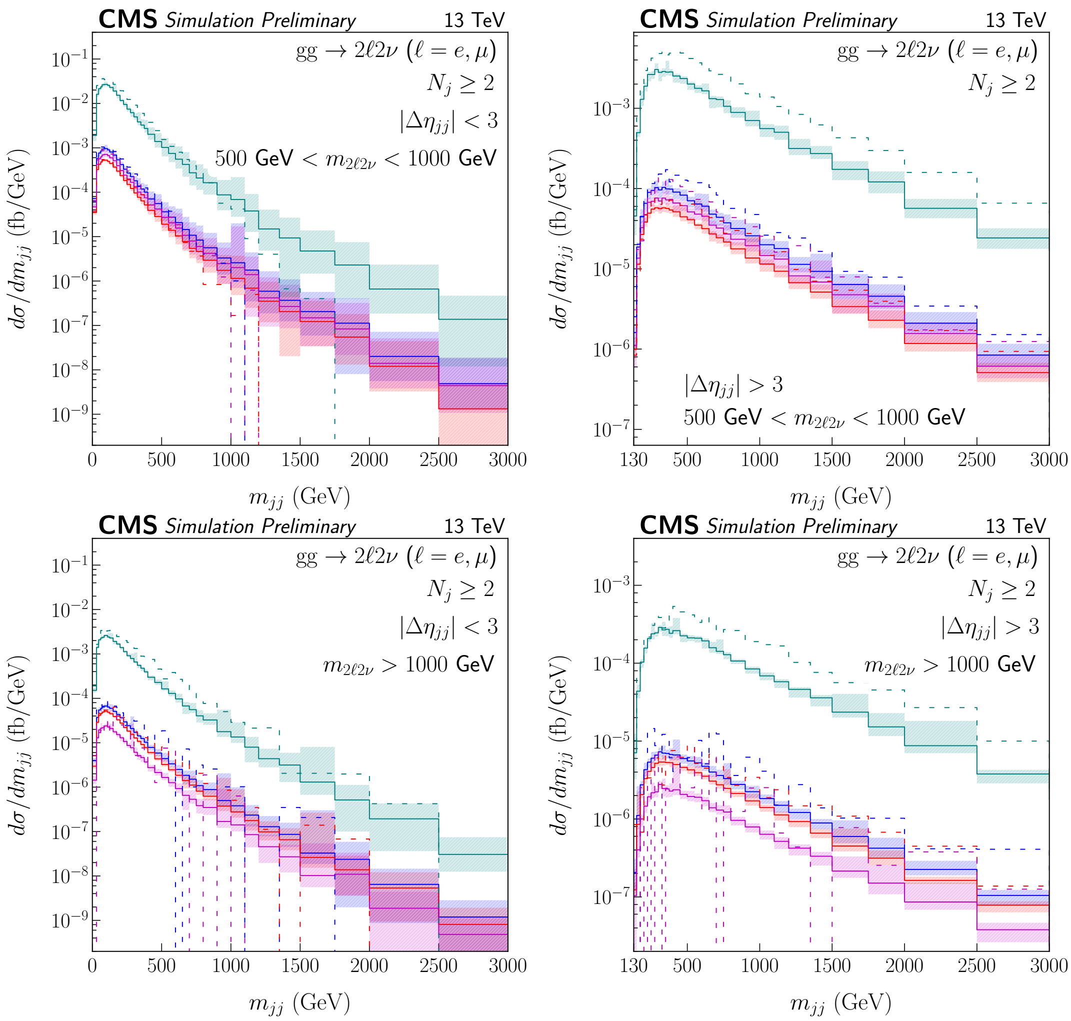

Figure 5.21: The distributions of m_{jj} are shown in the gg production in the $\mathrm{N_j} \geq 2$ category. The left panels show the spectra with $|\Delta\eta_{\mathrm{jj}}| < 3$, and the right panels show the spectra with $|\Delta\eta_{\mathrm{jj}}| \geq 3$. The panels are ordered from top to bottom in the $m_{2\ell 2\nu}$ bins of 500–1000 GeV , and beyond 1000 GeV . The different hypotheses are listed in the legend showed in Fig. 5.14.

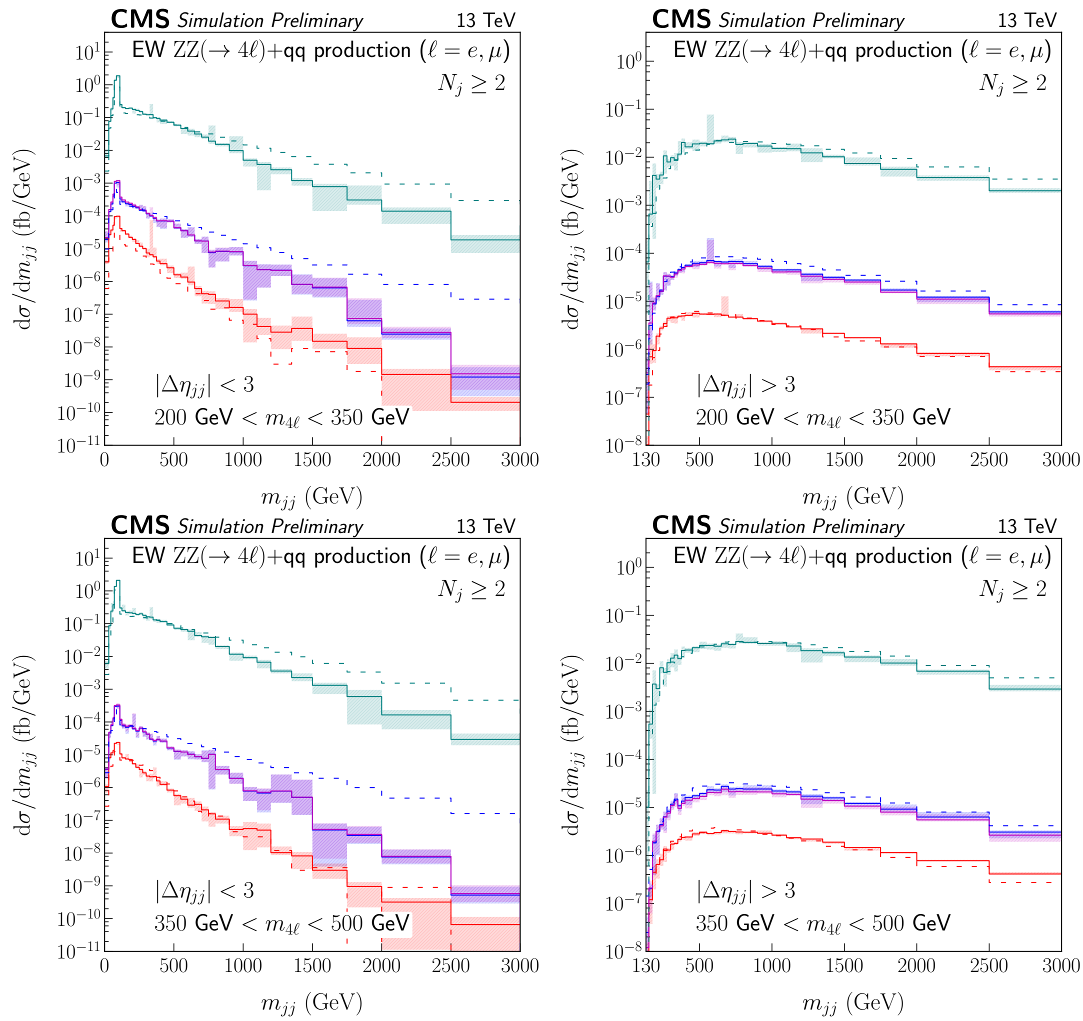

Figure 5.22: The distributions of m_{jj} are shown in the EW production in the $\mathrm{N_j} \geq 2$ category. The left panels show the spectra with $|\Delta\eta_{\mathrm{jj}}| < 3$, and the right panels show the spectra with $|\Delta\eta_{\mathrm{jj}}| \geq 3$. The panels are ordered from top to bottom in the $m_{4\ell}$ bins of 200–350 GeV and 350–500 GeV . The different hypotheses are listed in the legend showed in Fig. 5.14.

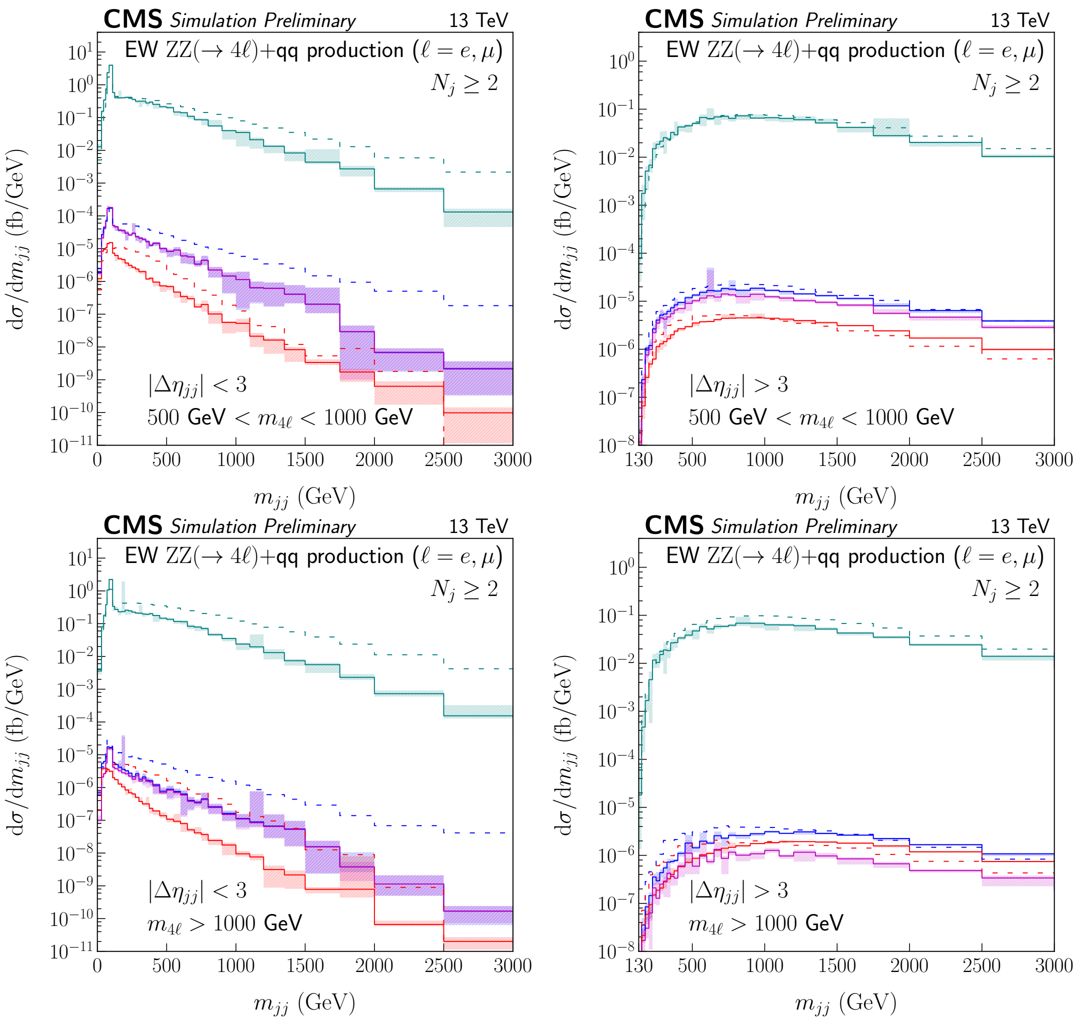

Figure 5.23: The distributions of m_{jj} are shown in the EW production in the $\mathrm{N_j} \geq 2$ category. The left panels show the spectra with $|\Delta\eta_{\mathrm{jj}}| < 3$, and the right panels show the spectra with $|\Delta\eta_{\mathrm{jj}}| \geq 3$. The panels are ordered from top to bottom in the $m_{4\ell}$ bins of 500–1000 GeV , and beyond 1000 GeV . The different hypotheses are listed in the legend showed in Fig. 5.14.

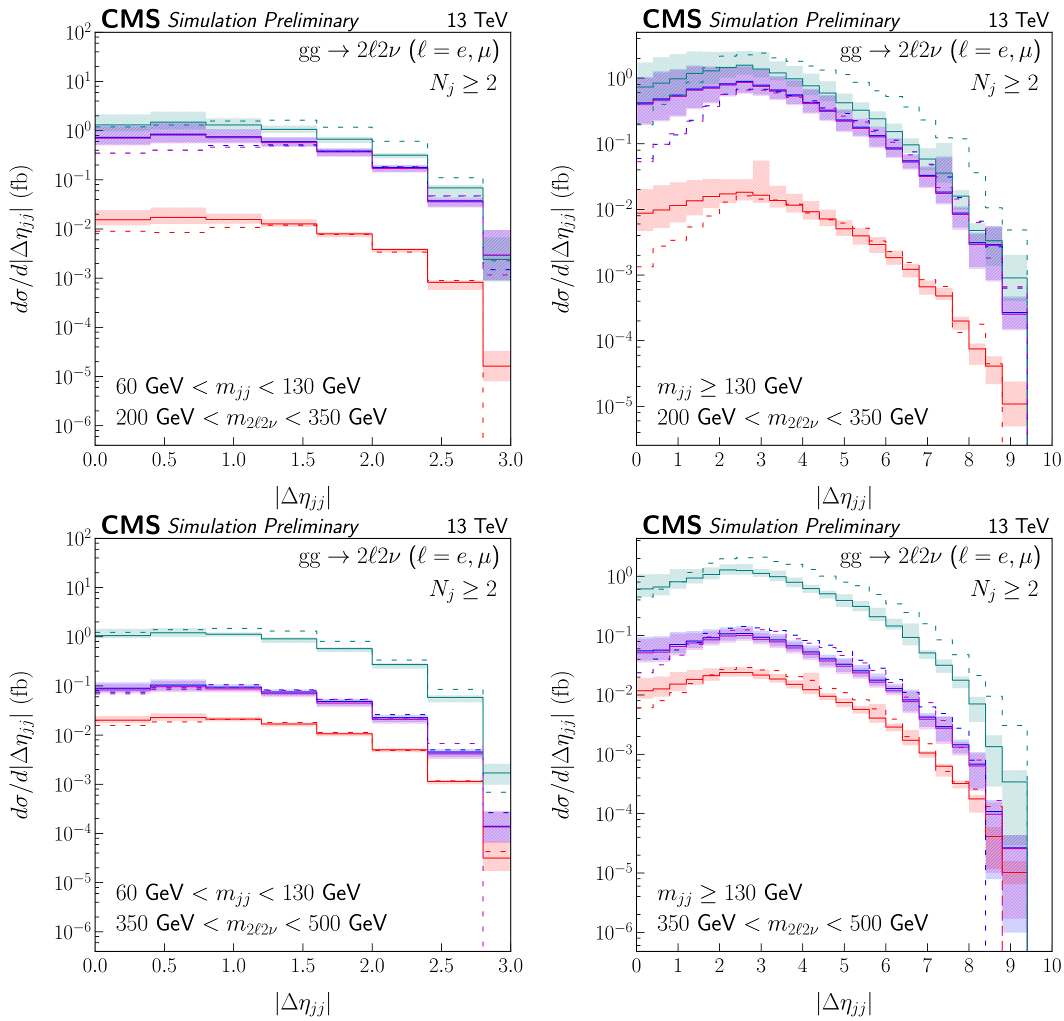

Figure 5.24: The distributions of $|\Delta\eta_{\mathrm{jj}}|$ are shown in the gg production and the $\mathrm{N_j} \geq 2$ category. The left panels show the spectra with 60 GeV $\leq m_{\mathrm{jj}} <$ 130 GeV , and the right panels show the spectra with $m_{\mathrm{jj}} \geq$ 130 GeV . The panels are ordered from top to bottom in the $m_{2\ell 2\nu}$ bins of 200–350 GeV and 350–500 GeV . The different hypotheses are listed in the legend showed in Fig. 5.14.

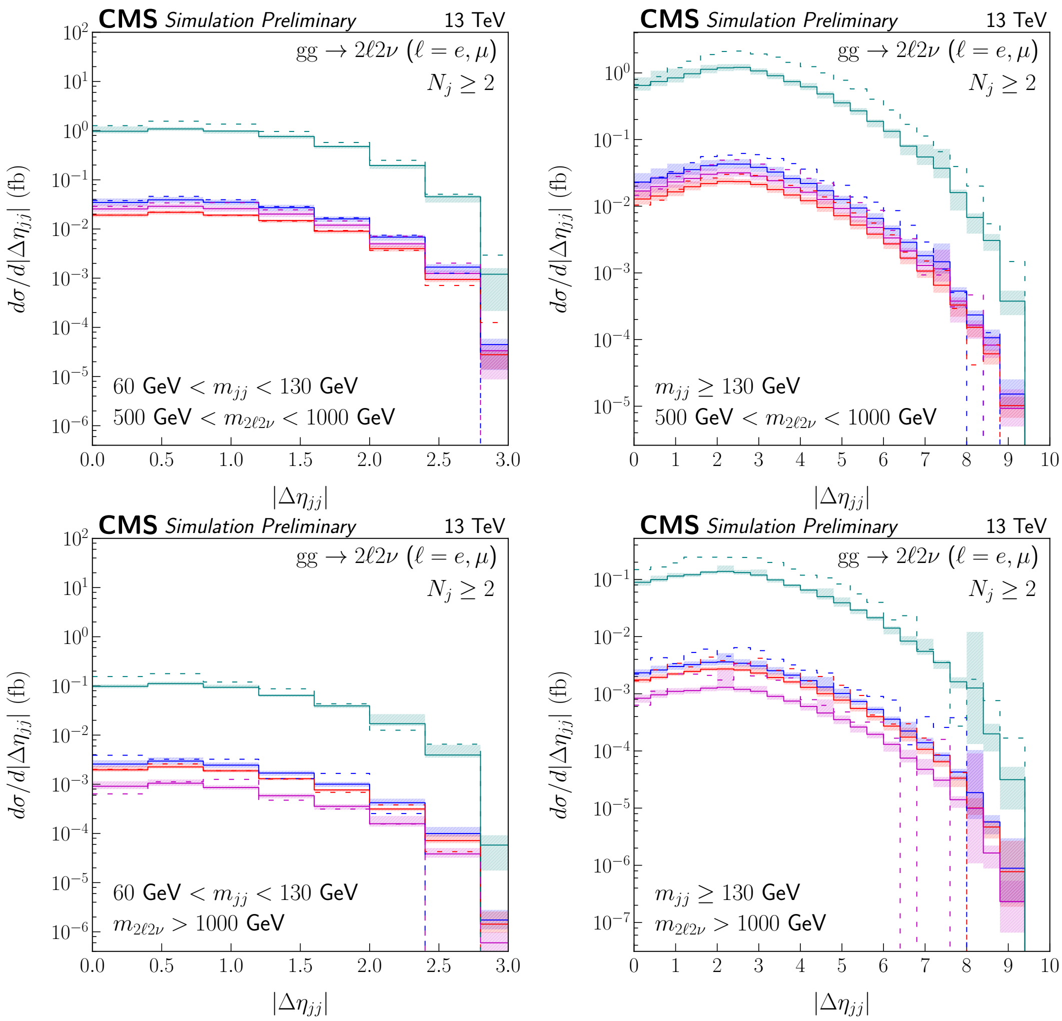

Figure 5.25: The distributions of $|\Delta\eta_{\mathrm{jj}}|$ are shown in the gg production and the $\mathrm{N_j} \geq 2$ category. The left panels show the spectra with 60 GeV $\leq m_{\mathrm{jj}} <$ 130 GeV , and the right panels show the spectra with $m_{\mathrm{jj}} \geq$ 130 GeV . The panels are ordered from top to bottom in the $m_{2\ell 2\nu}$ bins of 500–1000 GeV and beyond 1000 GeV . The different hypotheses are listed in the legend showed in Fig. 5.14.

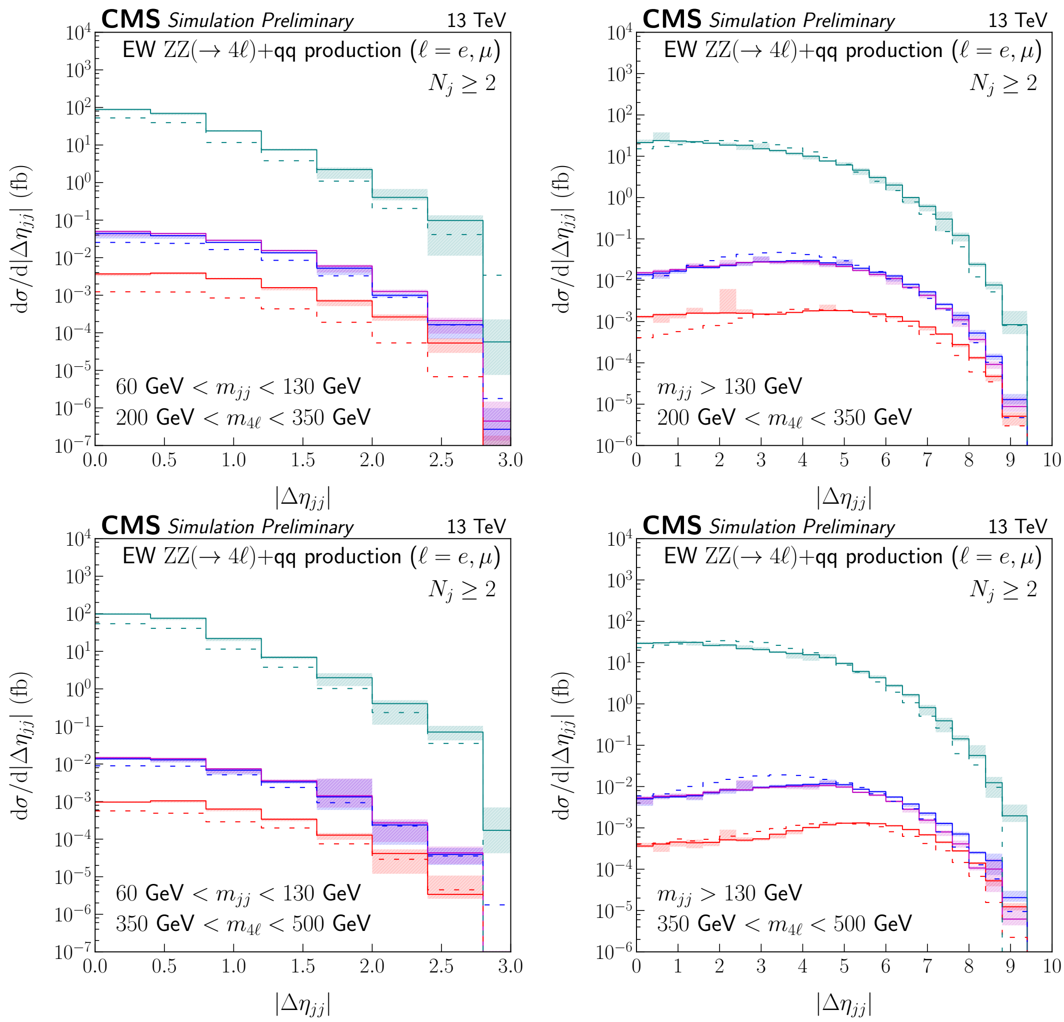

Figure 5.26: The distributions of $|\Delta\eta_{\mathrm{jj}}|$ are shown in the EW production and the $\mathrm{N_j} \geq 2$ category. The left panels show the spectra with 60 GeV $\leq m_{\mathrm{jj}} <$ 130 GeV , and the right panels show the spectra with $m_{\mathrm{jj}} \geq$ 130 GeV . The panels are ordered from top to bottom in the $m_{4\ell}$ bins of 200–350 GeV and 350–500 GeV . The different hypotheses are listed in the legend showed in Fig. 5.14.

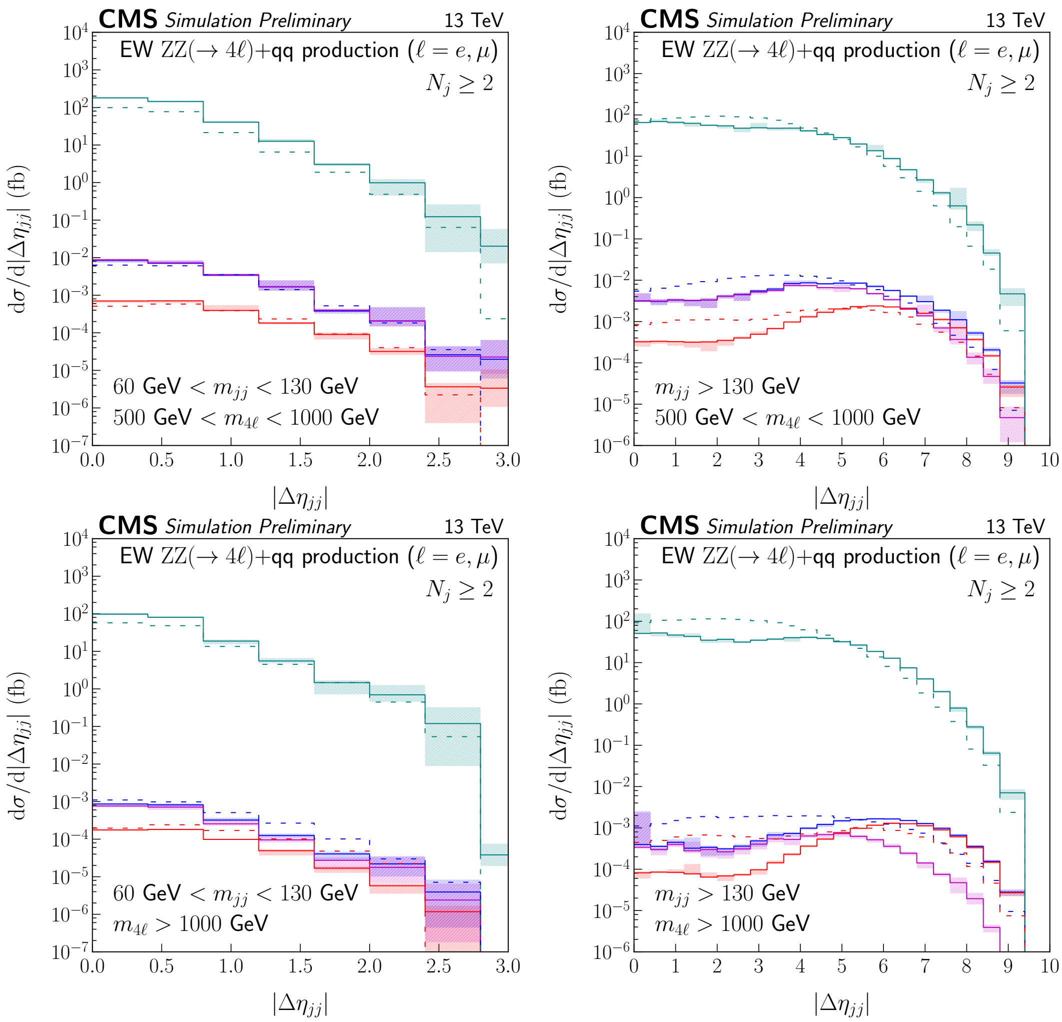

Figure 5.27: The distributions of $|\Delta\eta_{\mathrm{jj}}|$ are shown in the EW production and the $\mathrm{N_j} \geq 2$ category. The left panels show the spectra with 60 GeV $\leq m_{\mathrm{jj}} <$ 130 GeV , and the right panels show the spectra with $m_{\mathrm{jj}} \geq$ 130 GeV . The panels are ordered from top to bottom in the $m_{4\ell}$ bins of 500–1000 GeV and beyond 1000 GeV . The different hypotheses are listed in the legend showed in Fig. 5.14.

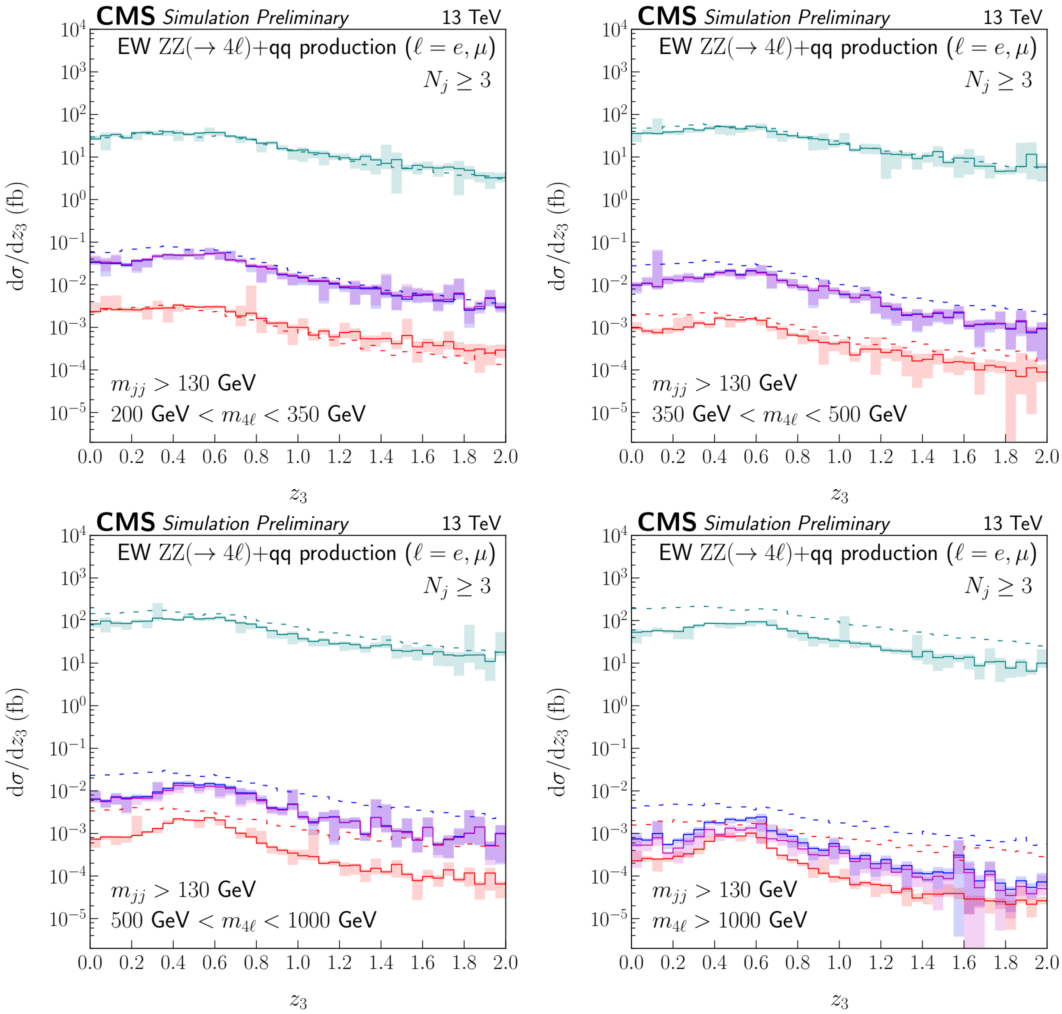

Figure 5.28: The distributions of the Zeppenfeld z_3 variable are shown in the EW production and the $\mathrm{N_j} \geq 3$ category with $m_{\mathrm{jj}} \geq 130$ GeV for the leading and subleading jets. The $m_{4\ell}$ bins shown are 200–350 GeV (top middle), 350–500 GeV (top right), 500–1000 GeV (bottom left), and beyond 1000 GeV (bottom right). The different hypotheses are listed in the legend showed in Fig. 5.14.

5.3 Event Selection

Before defining signal and control regions (SR and CR), we define the cuts that are meant to suppress processes that do not feature one Z boson decaying into a e^+e^- pair or a $\mu^+\mu^-$ pair and one Z boson decaying into a neutrino-antineutrino pair. These processes are referred to as reducible backgrounds. The main reducible backgrounds in this analysis are the Drell-Yan process with instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$, and fully leptonic $\mathrm{t\bar{t}}$ decays. The following selection criteria are applied:

- having no b-tagged jets based on the loose working point defined for this analysis,
- $p_{\mathrm{T}}^{\mathrm{boson}} \geq 55$ GeV,
- $p_{\mathrm{T}}^{\mathrm{miss}} \geq 125$ GeV if $N_j < 2$, or ≥ 140 GeV otherwise,
- $|\Delta\phi_{\mathrm{miss}}^{\mathrm{boson+jets}}| > 2.5$,
- $|\Delta\phi_{\mathrm{miss}}^{\mathrm{boson}}| > 1.0$,
- $\min\Delta\phi_{\mathrm{miss}}^{\mathrm{j}} > 0.25$ if $N_j = 1$, or > 0.5 if $N_j \geq 2$,

where the 'boson' indicates the $\ell^+\ell^-$ pair in the SR.

The quantity N_j denotes the multiplicity of any jet passing analysis criteria described in Sec. 3.6. The quantity $\Delta\phi_{\mathrm{miss}}^{\mathrm{boson}}$ denotes the difference in ϕ between the momentum of the boson and the missing transverse momentum; the quantity $\Delta\phi_{\mathrm{miss}}^{\mathrm{boson+jets}}$ denotes the difference in ϕ between the total transverse momentum vector composed of the boson and jets, and the missing transverse momentum; and the quantity $\min\Delta\phi_{\mathrm{miss}}^{\mathrm{j}}$ denotes the minimum unsigned difference in ϕ between any jet and the missing transverse momentum. Requiring these different $\Delta\phi$ selection cuts help to reduce the instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$ contribution from the DY process by vetoing events with jets that have a large, misreconstructed energy, or events with large unclustered energy. In addition to these selection cuts, all events are also required to satisfy the event veto requirements discussed in Chapter 3. The efficiencies of the different cuts are shown in Fig. 5.29 for the gluon fusion signal, in Fig. 5.30 for the VBF signal, and in 5.31 for the Drell-Yan and $\mathrm{t\bar{t}}$ processes.

When the signal region is analyzed, a pair of leptons with same flavor and opposite charge are required to be present with an invariant mass, $m_{\ell\ell}$, satisfying $|m_{\ell\ell} - m_{\mathrm{Z}}| \leq$ 15 GeV , where m_{Z} taken to be 91.2 GeV . Both leptons are required to have $p_{\mathrm{T}}^{\ell} \geq$ 25 GeV . The signal selection criteria are summarized in Table 5.2.

Table 5.2: Signal selection requirements.

Quantity	Requirement
p_{T}^{ℓ}	p_{T}^{ℓ} ≥ 25 GeV on both leptons
$\lvert\eta_{\ell}\rvert$	< 2.4 on μ , < 2.5 on e
N_{ℓ}	Exactly two leptons with tight isolation, no extra leptons with loose isolation and $p_{\mathrm{T}} \geq 5$ GeV
N_{trk}	No isolated tracks satisfying the selection requirements
N_{γ}	No photons with $p_{\mathrm{T}} \geq 20$ GeV , $\lvert\eta\rvert < 2.5$ satisfying the baseline selection requirements
p_{T}^{j}	≥ 30 GeV , used in counting N_j
$\lvert\eta_j\rvert$	< 4.7, used in counting N_j
$m_{\ell\ell}$	$\lvert m_{\ell\ell} - 91.2\rvert < 15$ GeV
N_b	No b-tagged jets based on the loose working point
$p_{\mathrm{T}}^{\ell\ell}$	≥ 55 GeV
$p_{\mathrm{T}}^{\mathrm{miss}}$	≥ 125 GeV if $N_j < 2$, ≥ 140 GeV otherwise
$\Delta\phi_{\mathrm{miss}}^{\mathrm{boson+jets}}$	> 2.5
$\Delta\phi_{\mathrm{miss}}^{\mathrm{boson}}$	> 1.0
$\min\Delta\phi_{\mathrm{miss}}^{\mathrm{j}}$	> 0.25 if $N_j = 1$, > 0.5 if $N_j \geq 2$

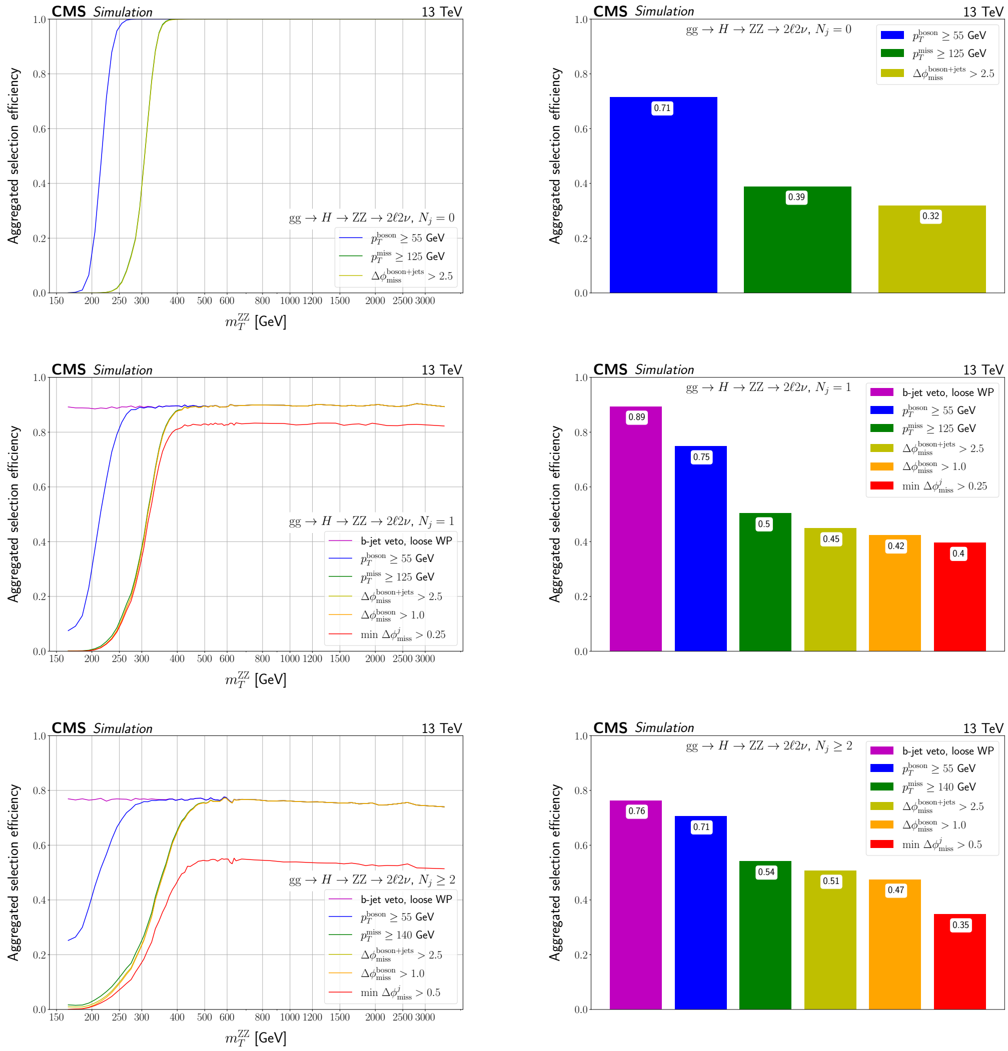

Figure 5.29: Efficiencies of the different selection requirements are shown for the SM gg $\to$ H signal process samples (i.e. samples after reweighting) in bins of $m_{\mathrm{T}}^{\mathrm{ZZ}}$(left) or inclusively (right) for $N_j = 0$ (top panels), $N_j = 1$ (middle panels), or $N_j \geq 2$ (bottom panels). The efficiencies shown in each line or bar are after applying the previous set of cuts listed in the legend and also after the set of cuts listed in 5.2 up to $|m_{\ell\ell} - 91.2| < 15$ GeV cut.

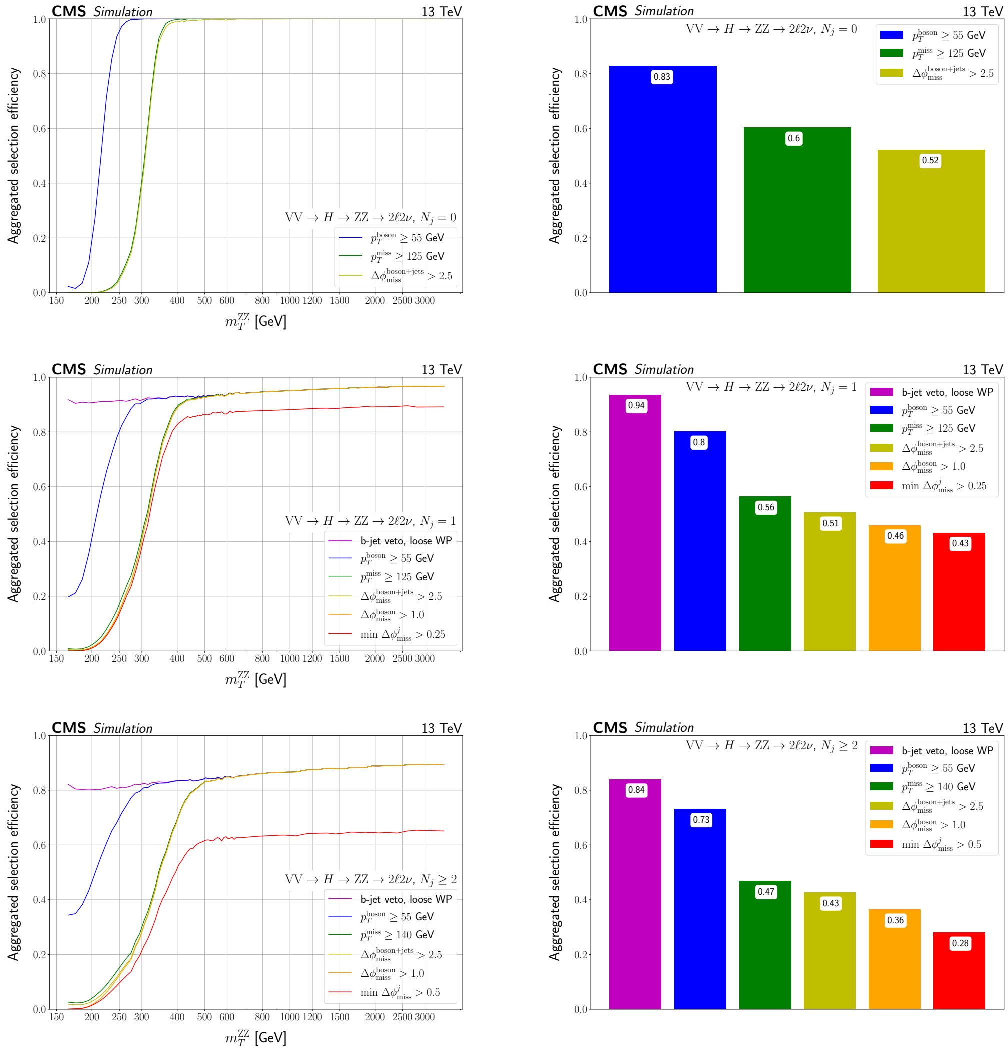

Figure 5.30: Efficiencies of the different selection requirements are shown for the SM VBF signal process samples (i.e. samples after reweighting) in bins of $m_{\mathrm{T}}^{\mathrm{ZZ}}$(left) or inclusively (right) for $N_j = 0$ (top panels), $N_j = 1$ (middle panels), or $N_j \geq 2$ (bottom panels). The efficiencies shown in each line or bar are after applying the previous set of cuts listed in the legend and also after the set of cuts listed in 5.2 up to $|m_{\ell\ell} - 91.2| < 15$ GeV cut.

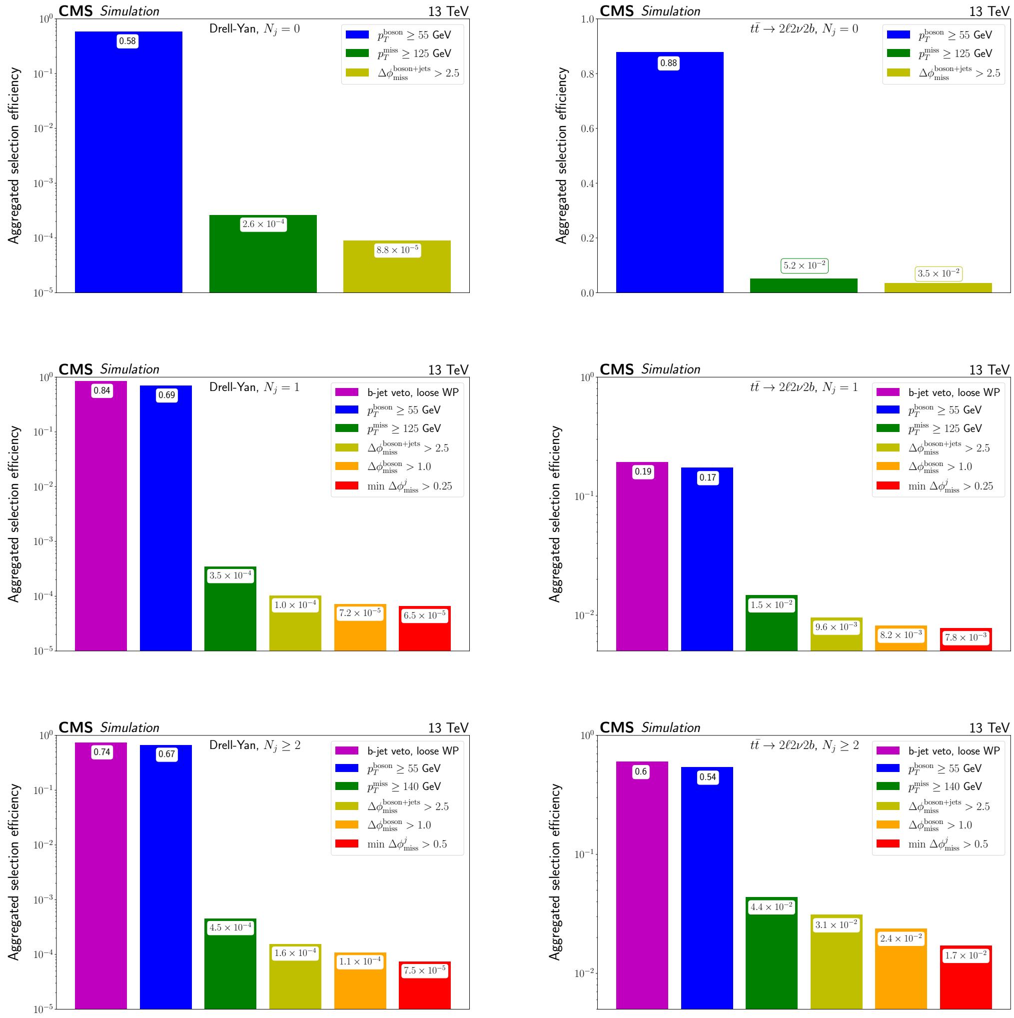

Figure 5.31: Efficiencies of the different selection requirements are shown for the DY (left) and $t\bar{t} \to 2\ell 2\nu 2b$ processes (right) over the inclusive m_T^{ZZ}range for $N_j = 0$ (top panels), $N_j = 1$ (middle panels), or $N_j \geq 2$ (bottom panels). The efficiencies shown in each bar are after applying the previous set of cuts listed in the legend and also after the set of cuts listed in 5.2 up to $|m_{\ell\ell} \; - 91.2| < 15$ GeV cut.

5.4 Event Categorization

Events are categorized based on the jet jet multiplicity into the $\mathrm{N_j} = 0$, $\mathrm{N_j} = 1$, and $\mathrm{N_j} \geq 2$ categories, and based on the flavor of the dilepton pairs, into the $\mu\mu$ and ee categories. The $\mathrm{N_j} = 0$ and $\mathrm{N_j} = 1$ categories are expected to be more sensitive to the ggsignal process, while the $\mathrm{N_j} \geq 2$ category is expected to be more sensitive to the VBF signal.

5.5 Kinematic Observables

In this analysis the invariant mass of ZZ system, m_{ZZ} plays an important role in charactering the events. However, due to the presence of 2 neutrinos in the final state, emerging as a missing transverse momentum, $p_{\mathrm{T}}^{\mathrm{miss}}$, the ZZ system invariant mass can not be reconstructed. Thus, we need another observable playing the role of m_{ZZ} based on the available information, which is known as transverse mass of the ZZ system defined as

$$m_{\mathrm{T}}^{\mathrm{ZZ}^2} = \left(\sqrt{p_{\mathrm{T}}^{\ell\ell^2} + m_{\ell\ell}^2} + \sqrt{p_{\mathrm{T}}^{\mathrm{miss}^2} + m_{\mathrm{Z}}^2}\right)^2 - \left|\vec{p}_T^{\,\ell\ell} + \vec{p}_T^{\,\mathrm{miss}}\right|^2 \tag{5.6}$$

where m_{Z} is the Z boson resonance pole mass, taken to be 91.2 GeV. The other important observable in this analysis is the $p_{\mathrm{T}}^{\mathrm{miss}}$ itself. There are different aspect which makes this observable significant,

- The distribution of $p_{\mathrm{T}}^{\mathrm{miss}}$ is sensitive to the p_{T} of the H boson, which boosts the ZZ system and therefore also increases the $p_{\mathrm{T}}^{\mathrm{miss}}$.
- The distribution of $p_{\mathrm{T}}^{\mathrm{miss}}$ is affected by the presence of anomalous couplings a_i at production, i.e. in the VBF and VH processes.
- Tails of the $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution is purer in signal contributions point of view, as illustrated in Fig. 5.34.

In addition to $m_{\mathrm{T}}^{\mathrm{ZZ}}$ and $p_{\mathrm{T}}^{\mathrm{miss}}$, in this analysis we use kinematic discriminants for events containing two or more jets, using matrix elements provided by the MELA package [16, 87–89]. In order to discriminate between the different production processes, the matrix elements are computed for the production part of the Feynman diagram. Examples of the use of such discriminants in off-shell analyses can be found in Refs. [93, 100]. The general form for the types of discriminants considered in this analysis reads as,

$$\mathcal{D}_{\text{2jet}}^{\text{VBF},h} = \frac{\mathcal{P}_{\text{VBF}}^{h}}{\mathcal{P}_{\text{VBF}}^{h} + \mathcal{P}_{\text{GF+2jet}}^{\text{SM}}}, \tag{5.7}$$

where $\mathcal{P}_{\text{VBF}}^{h}$ and $\mathcal{P}_{\text{GF+2jet}}^{\text{SM}}$ are the probability densities calculated based on the VBF and the GF H boson production mode which contains at least 2 jets, respectively; and the index h denotes the SM or either one of the BSM hypotheses with couplings listed in Table 1.2 ($h = a_1$ i.e. SM, a_2, a_3 and λ_1). The index in the notation is dropped for the discriminant with the SM hypothesis. No mixtures between the SM and BSM hypotheses are considered in these discriminants. In order to compute the kinematic discriminants, the H boson momentum and the two leading-p_{T} jets are needed. In addition to the four-momenta of the two leading-p_{T} jets, the four-momentum of the H boson candidate is needed as an input to the computation of the matrix element.

In order to approximate the H boson momentum, the four-momentum of $\text{Z} \to 2\ell$ and $\text{Z} \to 2\nu$ systems must be available or at least be properly approximated. The only kinematic which can not be properly identified is the $\eta_{\nu\nu}$, however, we need to consider a value for that (for example $\eta_{\nu\nu} = \eta_{\ell\ell}$) to approximate the H boson momentum. Therefore with this consideration, and the four-momenta of 2ℓ system and the $\vec{p}_{\text{T}}^{\text{miss}}$, the H boson momentum can be approximated. In the off-shell regime, it is observed from the simulation that the longitudinal momentum of the 2ν system usually has the similar value as the longitudinal momentum of the 2ℓ in the $\text{H} \to 2\ell 2\nu$ decay. Therefore by using this approximation $\eta_{\nu\nu} = \eta_{\ell\ell}$, and the four-momenta of 2ℓ system and the $\vec{p}_{\text{T}}^{\text{miss}}$, the H boson momentum can be approximated.

In the following pages, the m_{T}^{ZZ} distributions in different jet categories and different cuts on the $p_{\text{T}}^{\text{miss}}$ are shown in Figs. 5.33 and 5.34. The $p_{\text{T}}^{\text{miss}}$ in jet = 1 and 2 categories are illustrated in Fig. 5.35. The $\mathcal{D}_{\text{2jet}}^{\text{VBF},h}$ distributions with $p_{\text{T}}^{\text{miss}} \geq 200$ and < 200 GeV and $m_{\text{T}}^{\text{ZZ}} \geq 450$ GeV cuts are also demonstrated in Figs. 5.36 to 5.39. Despite of the GF process, in the EW process there are still some contributions from on-shell H boson production at high m_{T}^{ZZ} values. This is due to the fact that in the ZH process (which is a part of the EW process), in which the extra Z boson may decay to neutrinos resulting in high $p_{\text{T}}^{\text{miss}}$ and therefore high m_{T}^{ZZ} values. Moreover, similarly in the WH process the W may decay to a charged lepton and neutrino which again increases the $p_{\text{T}}^{\text{miss}}$ of the event. For this, in the above mentioned figures, we split the distributions of the EW process into on-shell and off-shell parts to address this phenomena. In general, the high tails of m_{T}^{ZZ}, $p_{\text{T}}^{\text{miss}}$ and $\mathcal{D}_{\text{2jet}}^{\text{VBF},h}$ show a better signal-to-noise ratio than the lower part of the distributions. This is particularly true for the EW part of the signal. The GF signal is concentrated in the intermediate range of values of the discriminating observables.

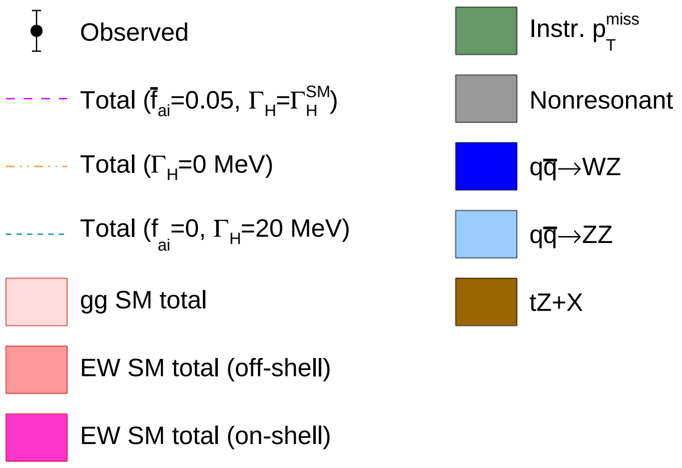

Figure 5.32: Shown are the legend of the different figures shown in this section. The different BSM HVV hypotheses are collectively abbreviated as f_{ai} . The filled histograms correspond to the SM expectations for the backgrounds, or the signal processes with the interfering backgrounds included. In the following figures, the middle panel shows the ratio of the different model predictions and of the data to the SM prediction, and the lower panel shows the expected composition of the event sample in the SM hypothesis. The pink dashed line shows the expectation for a BSM model with the a_i coefficient mentioned in the figure legend set to 0.05, and the H boson width set to the SM value.

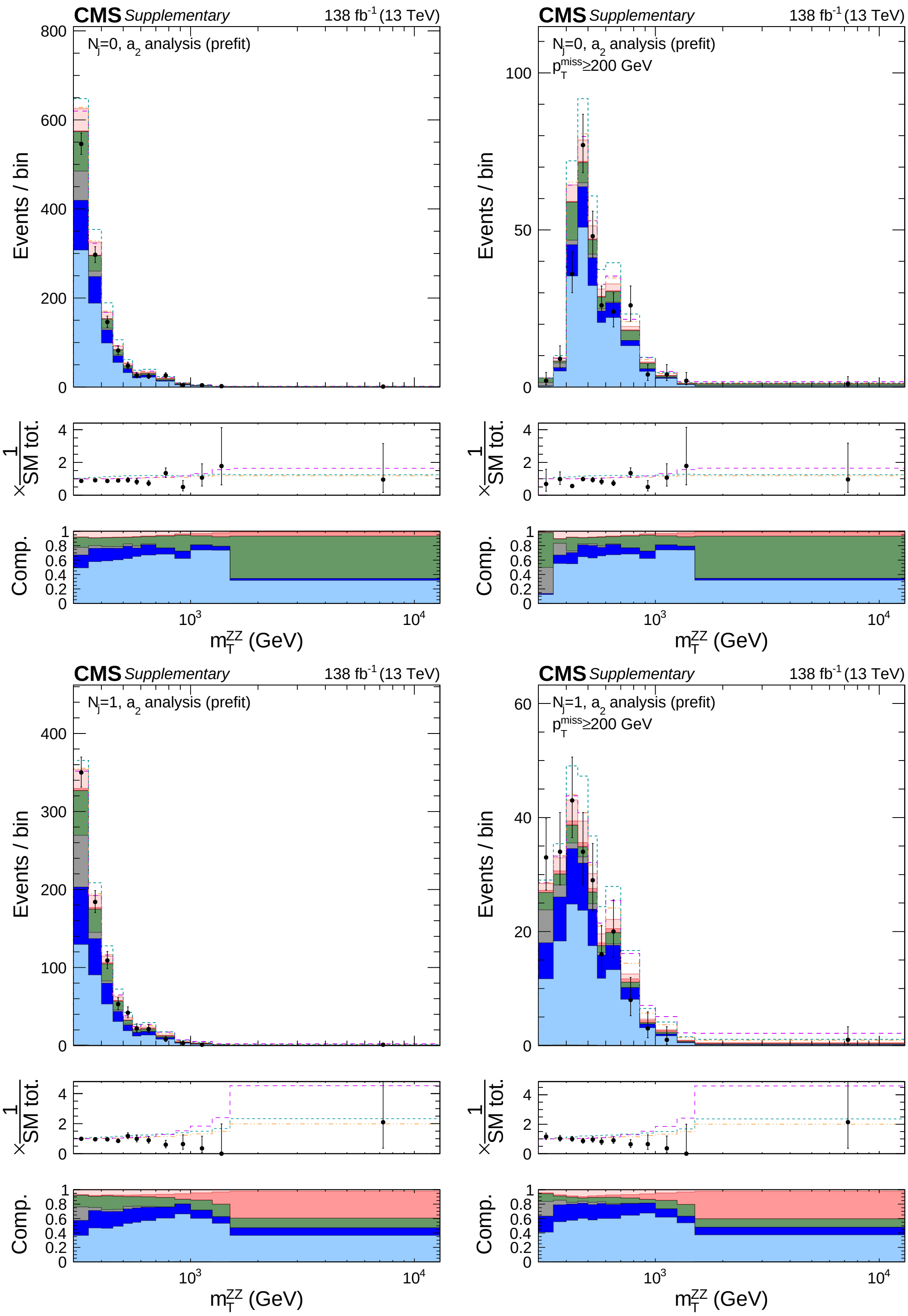

Figure 5.33: Distributions of $m_{\mathrm{T}}^{\mathrm{ZZ}}$in the $N_j = 0$ (top) and $N_j = 1$ (bottom) categories with (left) or without (right) a $p_{\mathrm{T}}^{\mathrm{miss}} \geq 200$ GeV requirement in order to reduce the contribution of the nonresonant and instrumental backgrounds. The various contributions are defined in Fig. 5.44.

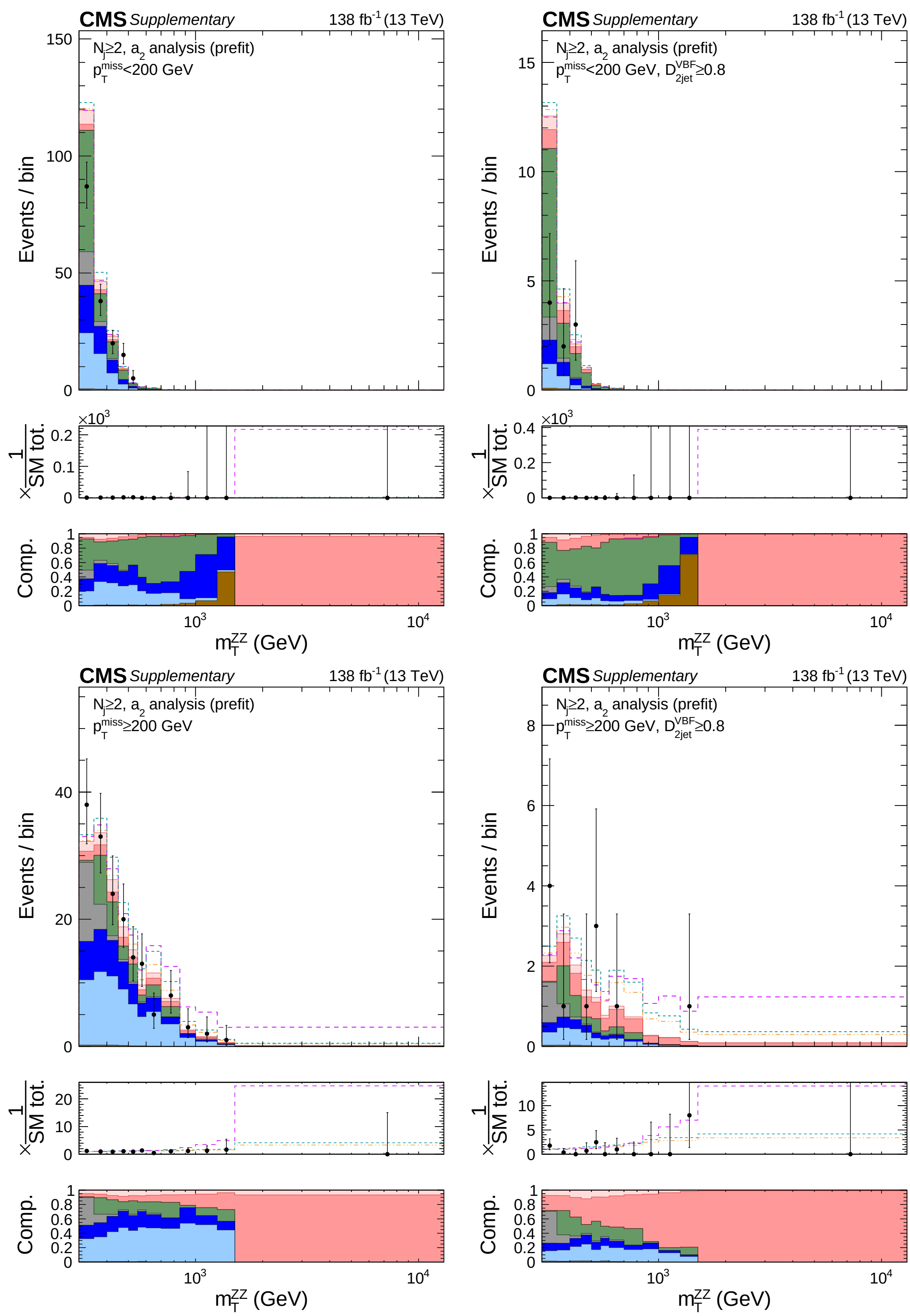

Figure 5.34: Distributions of $m_{\mathrm{T}}^{\mathrm{ZZ}}$in the $N_j \geq 2$ category for the $p_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV (top) and $p_{\mathrm{T}}^{\mathrm{miss}} \geq 200$ GeV (bottom) bins with (left) or without (right) a $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF}} \geq 0.8$ requirement to enhance the contribution of VBF-like events. The various contributions are defined in Fig. 5.44.

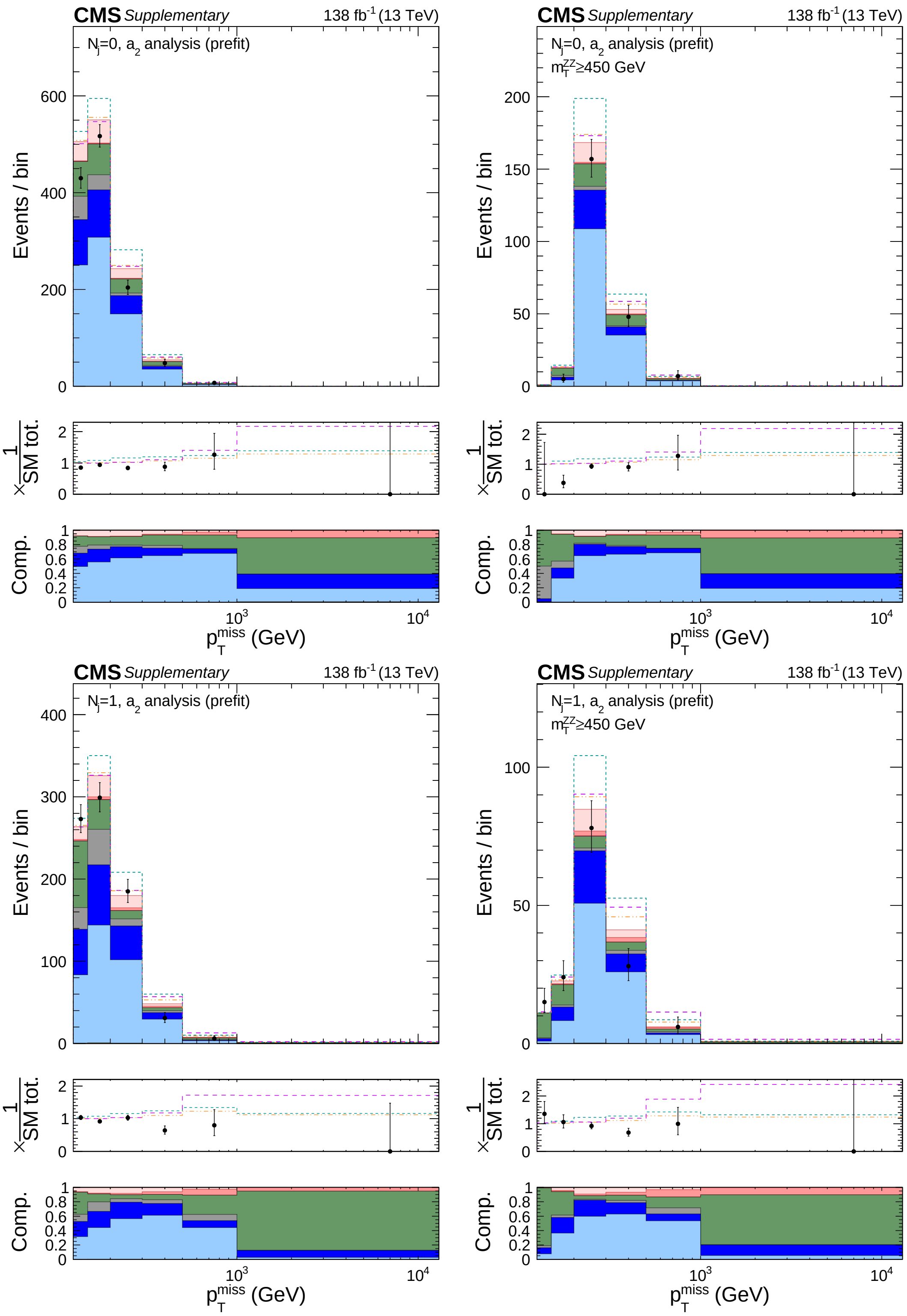

Figure 5.35: Shown are the distributions of $p_{\mathrm{T}}^{\mathrm{miss}}$ in the $N_j = 0$ (top) and $N_j = 1$ (bottom) categories with (left) or without (right) an $m_{\mathrm{T}}^{\mathrm{ZZ}} \geq 450$ GeV requirement in order to reduce the contribution of the nonresonant and instrumental backgrounds. The various contributions are defined in Fig. 5.44.

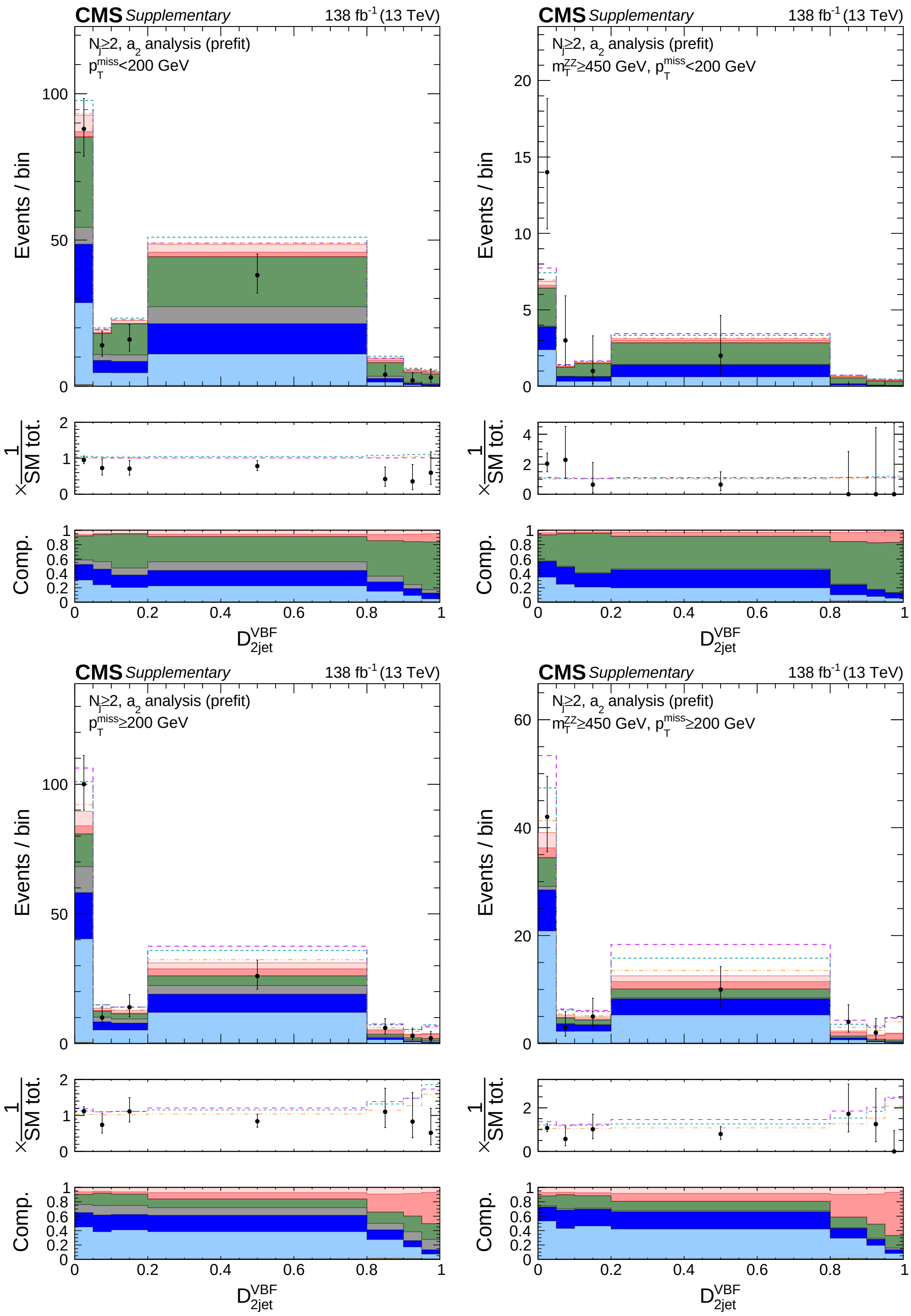

Figure 5.36: Shown are the distributions of $\mathcal{D}_{\text{2jet}}^{\text{VBF}}$ in the $N_j \geq 2$ category for the $p_{\text{T}}^{\text{miss}} < 200$ GeV (top) and $p_{\text{T}}^{\text{miss}} \geq 200$ GeV (bottom) bins with (left) or without (right) a $m_{\text{T}}^{ZZ} \geq 450$ GeV requirement in order to reduce the contribution of the nonresonant and instrumental backgrounds. The various contributions are defined in Fig. 5.44.

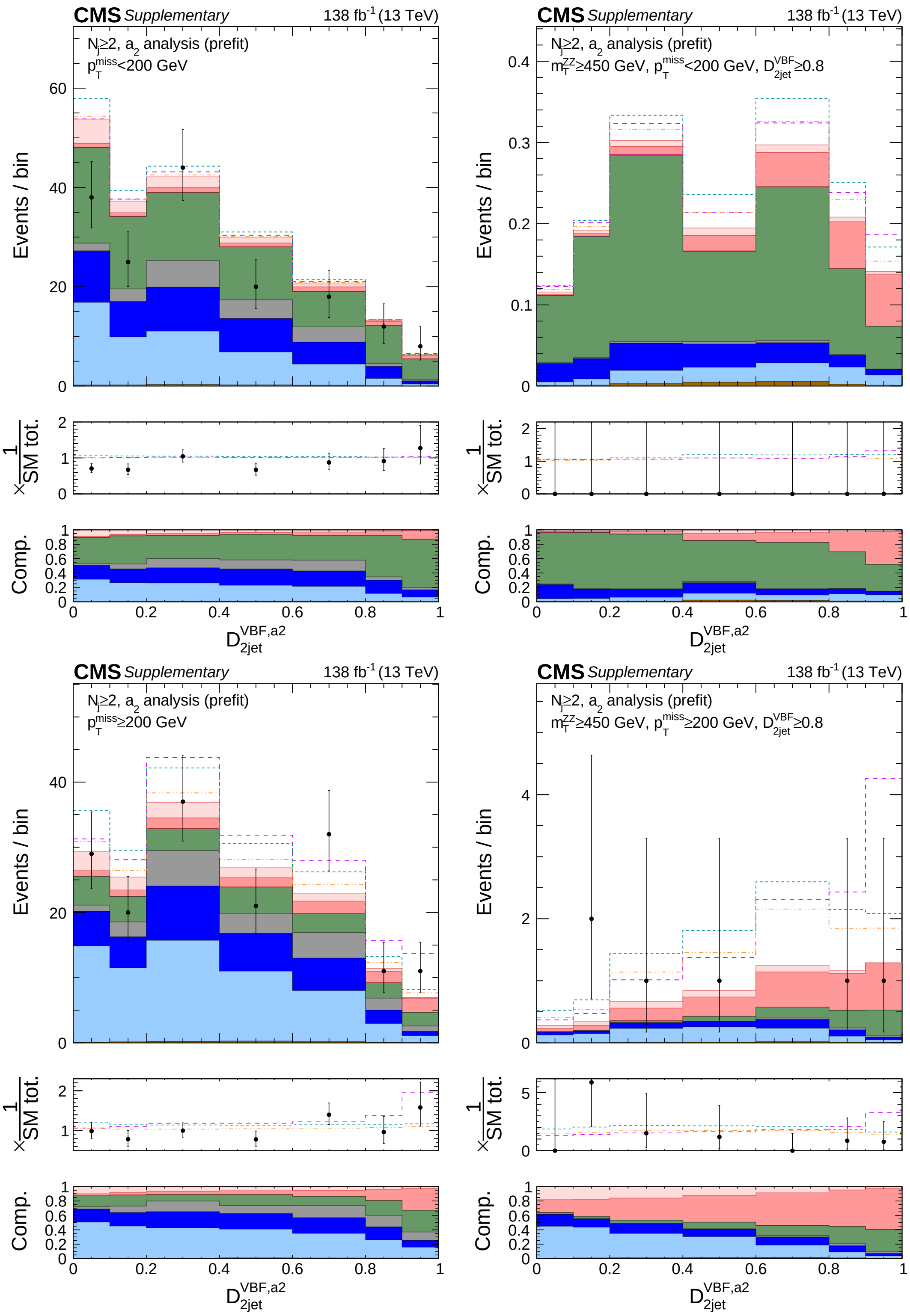

Figure 5.37: Shown are the distributions of $\mathcal{D}_{\text{2jet}}^{\text{VBF},a2}$ in the $N_j \geq 2$ category for the $p_{\text{T}}^{\text{miss}} < 200$ GeV (top) and $p_{\text{T}}^{\text{miss}} \geq 200$ GeV (bottom) bins with (left) or without (right) the $m_{\text{T}}^{ZZ} \geq 450$ GeV and $\mathcal{D}_{\text{2jet}}^{\text{VBF}} \geq 0.8$ requirements to enhance the contribution from SM VBF-like signal events. The various contributions are defined in Fig. 5.44.

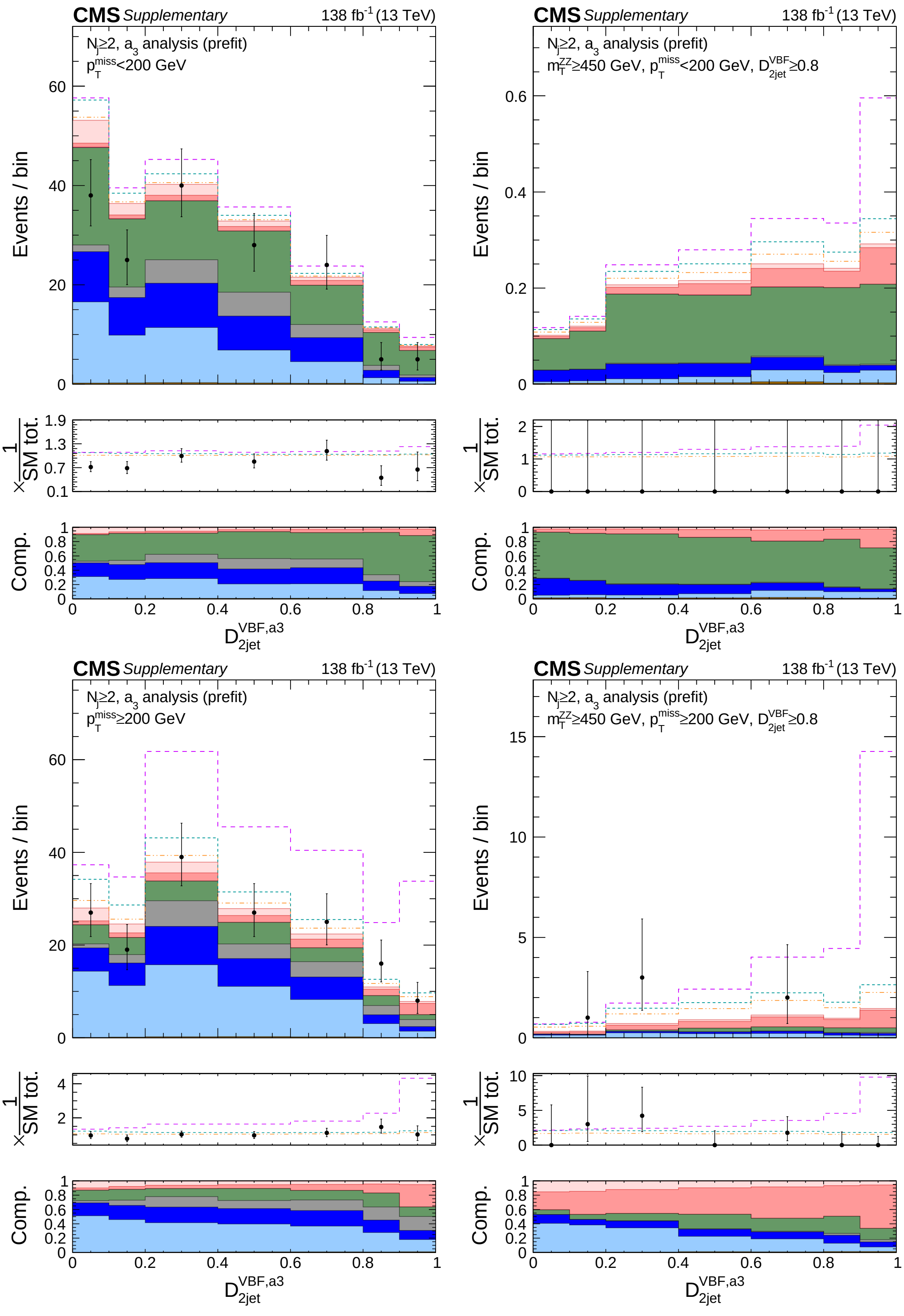

Figure 5.38: Shown are the distributions of $\mathcal{D}_{\text{2jet}}^{\text{VBF},a3}$ in the $N_j \geq 2$ category for the $p_{\text{T}}^{\text{miss}} < 200$ GeV (top) and $p_{\text{T}}^{\text{miss}} \geq 200$ GeV (bottom) bins with (left) or without (right) the $m_{\text{T}}^{ZZ} \geq 450$ GeV and $\mathcal{D}_{\text{2jet}}^{\text{VBF}} \geq 0.8$ requirements to enhance the contribution from SM VBF-like signal events. The various contributions are defined in Fig. 5.44.

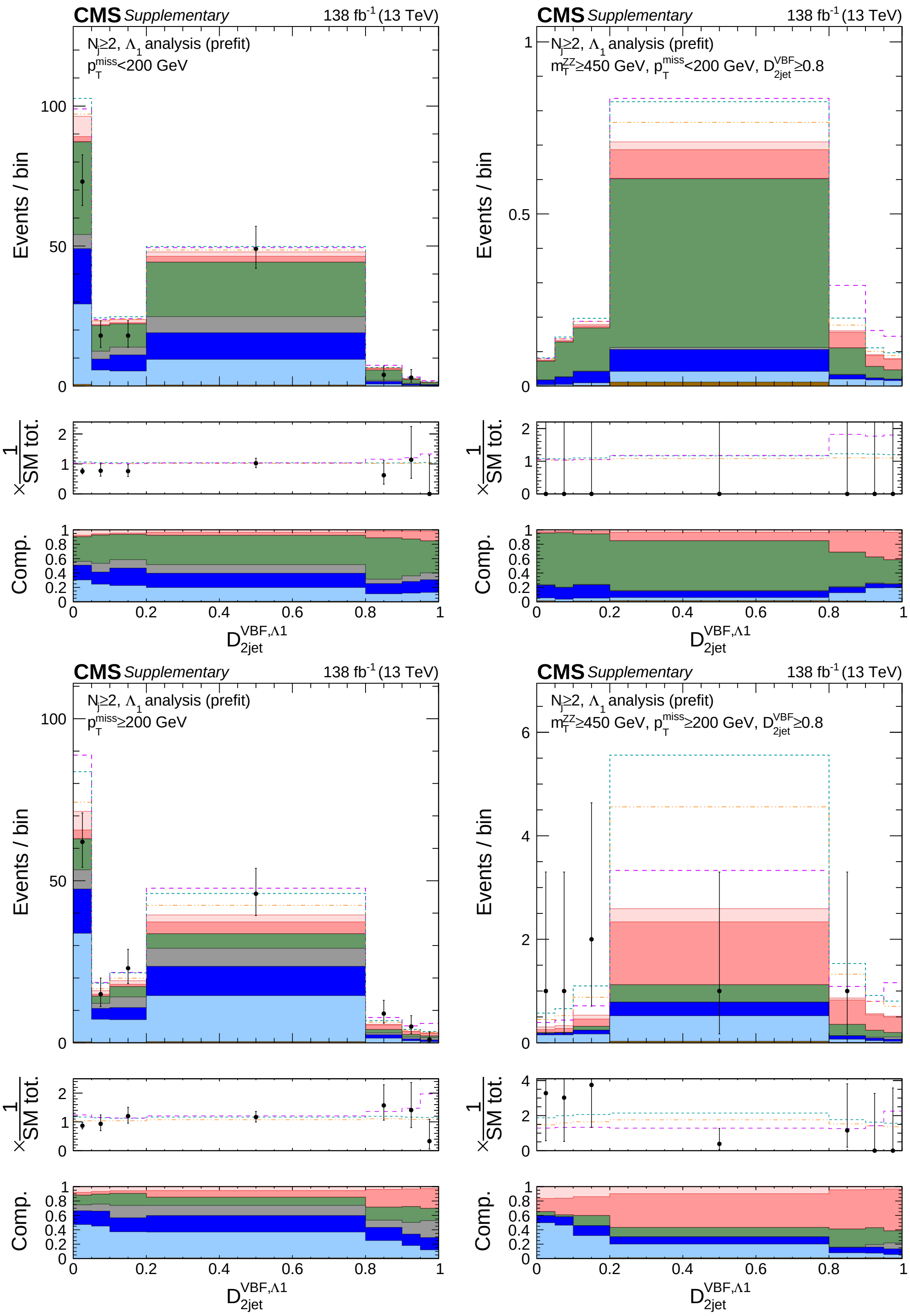

Figure 5.39: Shown are the distributions of $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF},\Lambda 1}$ in the $N_j \geq 2$ category for the $p_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV (top) and $p_{\mathrm{T}}^{\mathrm{miss}} \geq 200$ GeV (bottom) bins with (left) or without (right) the $m_{\mathrm{T}}^{\mathrm{ZZ}} \geq 450$ GeV and $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF}} \geq 0.8$ requirements to enhance the contribution from SM VBF-like signal events. The various contributions are defined in Fig. 5.44.

5.6 Data-driven Methods for Non-interfering Backgrounds Estimations

Although the MC samples of Z+jets is used to study the efficiency and optimization of selection requirements, they are not reliable to estimate the correct contamination from this process because simulations are not expected to describe the tails of $p_{\mathrm{T}}^{\mathrm{miss}}$ accurately. Furthermore, the number of events present in the simulation is not sufficient to make an accurate prediction of this contamination. Similarly, for the non-resonant backgrounds such as $\mathrm{t\bar{t}}$ and WW processes, MC samples can not be used to estimated the contamination of such processes with enough accuracy. This is due to the fact that in this analysis we apply veto requirements based on b-tagging in order to reduce the $\mathrm{t\bar{t}}$ contribution in signal region, and this veto is sensitive to inaccuracies in b-tagging efficiencies and scale factors. Since the analysis is not sensitive to the $\mathrm{H} \rightarrow \mathrm{WW}$ contribution to the dilepton sample, we prefer to measure this contribution together with the other non-resonant dilepton contributions, and treat it as a background, rather than include it in the signal model. Therefore, we use data-driven techniques to estimate these two types of backgrounds i.e. the instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$ and non-resonant background processes.

5.6.1 Estimation of the Z+jets Background

The Z+jets background is particularly important in this analysis because the $p_{\mathrm{T}}^{\mathrm{miss}}$ for this process can arise from instrumental sources, such as detector energy resolution, jet energy mismeasurement, fluctuations in pileup energy, or instrumental noise, which can contaminate the signal region. Although the fraction of Z+jets events that have a large enough instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$ to pass the analysis selection is small, the large cross section of the process itself can still contribute significantly in the signal region. Unfortunately, the current simulations are not capable of modeling the detector and pileup effects accurately enough in the tails of the distributions used for our analysis selection. Moreover, the phase space for such selections is so small that too few events are simulated which results in large statistical uncertainties from the simulation.

Since the author of the present thesis was not involved in the estimation of this background, that is very complex, only the main features are described below. In order to estimate the Z+jets background, the idea is to use photons as a proxy for Z bosons. In the photon + jet sample, the $p_{\mathrm{T}}^{\mathrm{miss}}$ has the same instrumental cause as in the Z+jet process i.e. the mismeasurement of jet energy, while the photon energy is well measured, as the Z boson momentum in Z+jet events. In addition to this analogy, the single-photon control region (CR) provides a sample that is independent from the signal region, and also, a large sample that allows a good estimation in each analysis category in particular in the ≥ 2 jet category.

This can compensate the lack of statistics compared to the Z+jets sample. For this purpose, a single-photon CR is used, where exactly one photon is required to pass the identification criteria described in Sec. 3.5. The events are required to pass the CR-γ triggers. It should be noted that in the single-photon CR, at low $p_{\mathrm{T}}^{\mathrm{miss}}$, it is dominated by γ+jets processes; however, at high $p_{\mathrm{T}}^{\mathrm{miss}}$, contributions from processes containing genuine $p_{\mathrm{T}}^{\mathrm{miss}}$ become important. Such processes are Zγ (with Z $\rightarrow \nu\nu$), Wγ (with W $\rightarrow \ell\nu$, where the ℓ is lost or not identified) and W+jets (with W $\rightarrow$ e ν, where an e is misidentified as a photon).

To have a enough accurate estimation, we correct the photon + jet event kinematics in order to properly resemble the kinematics of the Z+jet process. This is done by applying weights to the photon + jet events. These weights are computed as the ratio of data event yields in photon + jet and Z + jet events with an inverted $p_{\mathrm{T}}^{\mathrm{miss}}$ cut (therefore from samples perpendicular to the signal region). However, since the γ-CR also contains a considerable amount of aforementioned processes (Zγ, Wγ, W+jets and Z $\rightarrow \nu\nu$) with genuine $p_{\mathrm{T}}^{\mathrm{miss}}$ besides γ+jets events, the reweighted yield of the CR cannot be simply used as an estimation for Z+jets in the SR, thus a subtraction of these processes is required. To estimate such contributions correctly, the trigger efficiency for photons was measured, and a control sample from Z $\rightarrow \ell\ell\gamma$ was studied in order to constrain the Zγ cross section that is poorly-known in our phase space and jet categories.

For illustration purposes, Fig. 5.40 shows the reweighting factors for $\ell\ell = ee + \mu\mu$ in bins of boson p_{T} and Fig. 5.41 shows the full $p_{\mathrm{T}}^{\mathrm{miss}}$ distribution in the photon CR. In these distributions, some discrepancy between data and MC are expected as the γ+jets distributions coming from simulations are not expected to be fully reliable.

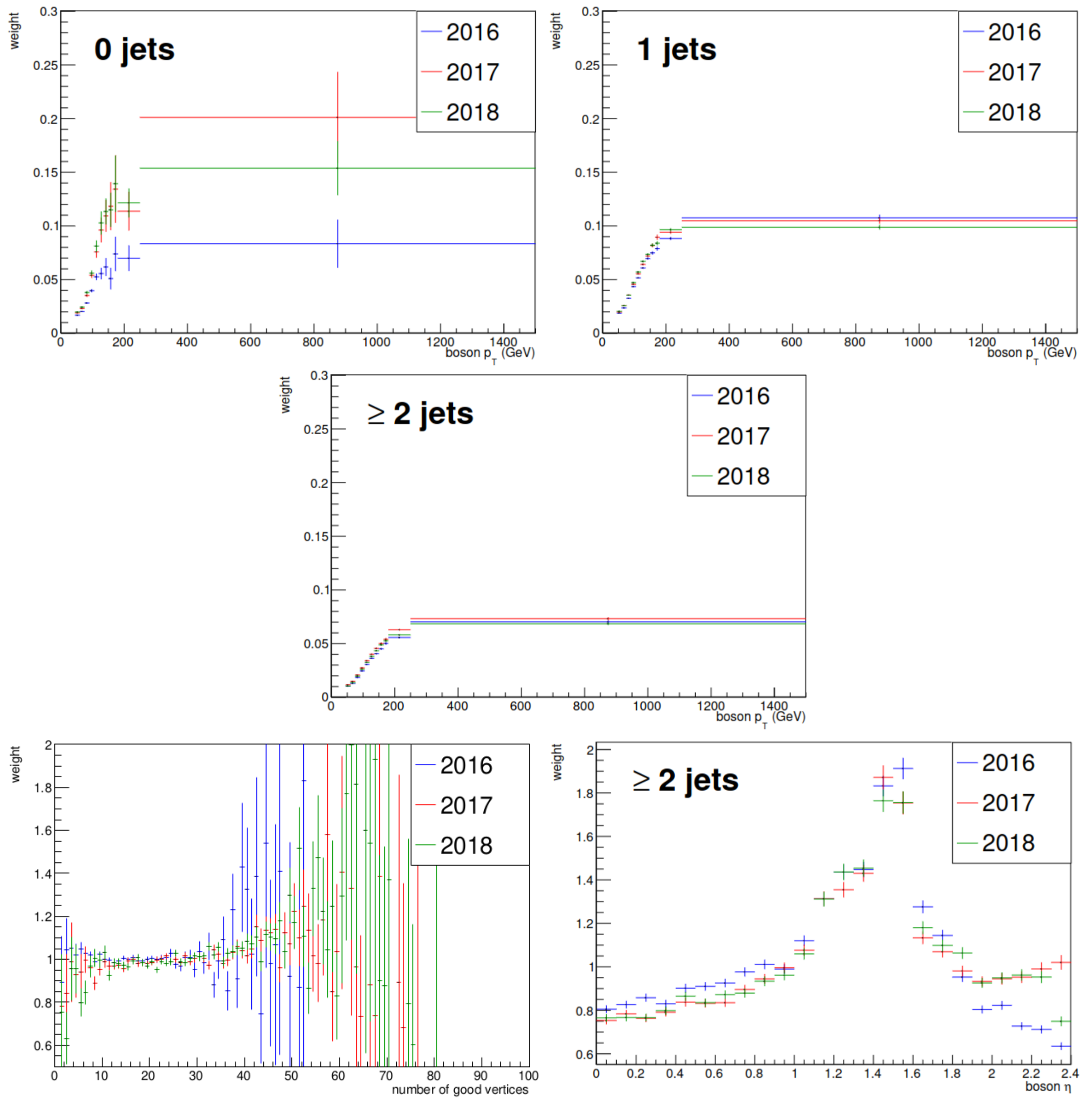

Figure 5.40: Transfer factors used in instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$ estimates for the 3 years. Reweighting in p_{T}, for 0 jet (top left), 1 jet (top right), and ≥ 2 jets (middle), and in the number of good vertices (bottom left), and in η for ≥ 2 jets (bottom right).

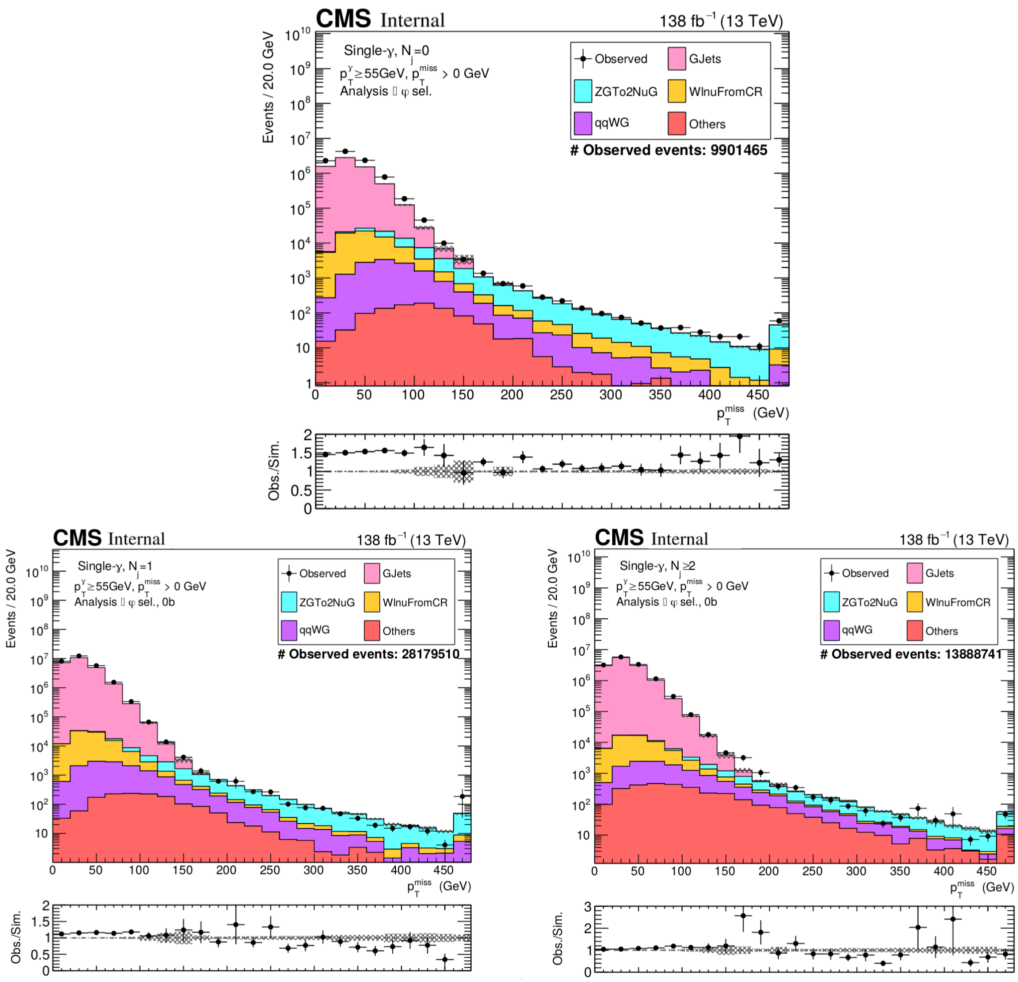

Figure 5.41: The data vs MC distributions of $p_{\mathrm{T}}^{\mathrm{miss}}$ in the photon + jet sample, in 0 jet (top), 1 jet (bottom left) and ≥2 jets (bottom right) categories are illustrated. The gray error bands only correspond to the statistical uncertainty in the simulation.

5.6.2 Estimation of the Non-resonant Backgrounds

The non-resonant background in this analysis is referred to any background process that results in a flat $m_{\ell\ell}$ spectrum. The main processes contributing to it are $\mathrm{t\bar{t}} \to 2\ell 2\nu 2\mathrm{b}$ and $\mathrm{WW} \to 2\ell 2\nu$ with real, flavor-symmetric leptonic decay. We utilize the flavor-symmetric nature of this background component, and the lepton and trigger efficiencies derived in this analysis by examining $e\mu$ events satisfying exactly the same analysis kinematic selections and reweighting these events for the differences in their efficiencies with those of ee and $\mu\mu$ events. The reweighting factor is defined as

$$w_{e\mu}^{\ell 1\ell 2} = \frac{1}{2} \times \frac{\varepsilon_{\ell 1} \cdot \varepsilon_{\ell 2}}{\varepsilon_{\mathrm{e}} \cdot \varepsilon_{\mu}} \times \frac{\varepsilon_{\ell 1\ell 2}^{\mathrm{trigger}}}{\varepsilon_{\mathrm{e}\,\mu}^{\mathrm{trigger}}} \times f_{corr}^{\ell\ell}\left(p_{\mathrm{T}}^{\mathrm{miss}}\right), \tag{5.8}$$

where ε_i is the lepton identification and isolation efficiency of lepton i, and $\varepsilon_{ij}^{\mathrm{trigger}}$ is the combined trigger efficiency on the ij lepton pair. The factor $1/2$ accounts for the SM ratio of ee or $\mu\mu$ events to $e\mu$ events. The last factor, $f_{corr}^{\ell\ell}$, is a correction factor on the spectrum of $p_{\mathrm{T}}^{\mathrm{miss}}$ due to slightly different resolution of ee , $\mu\mu$, and $e\mu$ events. All other kinematic requirements on this control region are the same as those on the signal region except for loosening the $p_{\mathrm{T}}^{\ell\ell}$ requirement to ≥ 25 GeV to gain more events.

Figure 5.42 compares the $m_{\mathrm{T}}^{\mathrm{ZZ}}$ distributions between the predicted and expected non-resonant background component in the signal region for the 2018 data sets. The distributions of various kinematic discriminants are also compared in Fig. 5.43 for the SM $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF}}$ discriminant used in the analysis for the 2018 data period. The corresponding distributions and also other kinematic distributions for other data sets periods (i.e. 2016 and 2017) are shown in Appendix C as well. In general by comparing the distributions in these figures we can see an acceptable agreement within the given uncertainties, between the data and the estimation in the control regions.

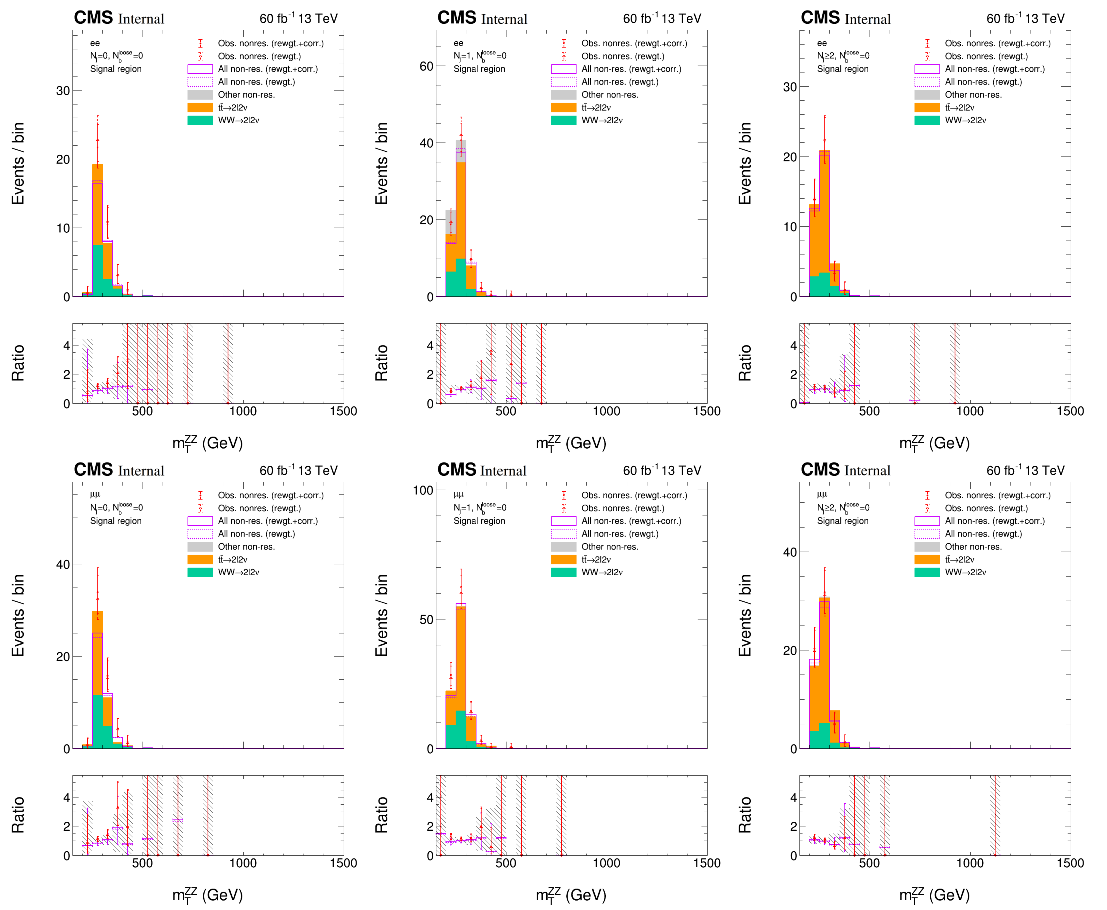

Figure 5.42: The $m_{\mathrm{T}}^{\mathrm{ZZ}}$ distributions in e e (top) and $\mu\ \mu$ (bottom) channels in the signal region are shown for the 2018 data set in each N_j category ($= 0$, $= 1$, and ≥ 2 from left to right).

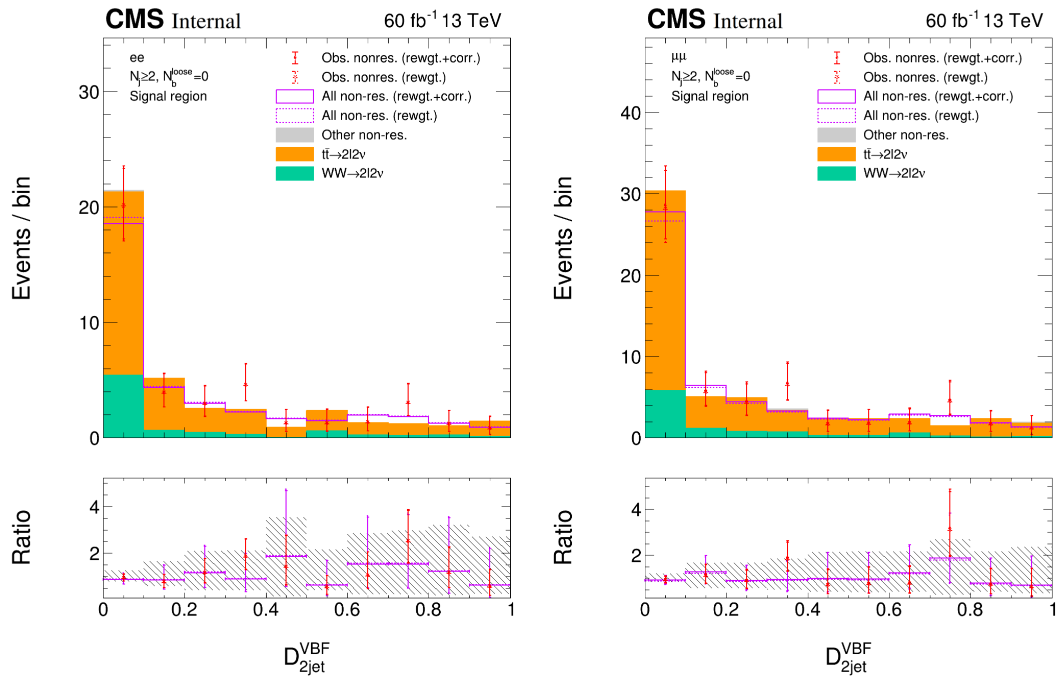

Figure 5.43: The distributions of SM $\mathcal{D}_{\text{2jet}}^{\text{VBF}}$ kinematic discriminants in the signal region are shown for the 2018 data period in the $N_j \geq 2$ category. The decay channel e e is shown on the left plot, and $\mu\ \mu$ is shown on the right one.

5.7 Likelihood Parametrization

The results of this analysis are extracted using binned extended maximum likelihood fits [101] applied on the events in different categories mentioned in Sec.5.4. We construct four independent joint fits in order to determine the total width of the H boson concerning different hypotheses, i.e. either with SM-like or the three anomalous couplings a_2, a_3, or Λ_1 assumption. These fits also constrain the three corresponding anomalous coupling parameters $\bar{f}_{ai} = f_{ai} \cdot \cos(\Phi_{ai})$, where f_{ai} is defined in Eq. (1.36). It should be noted that when a particular anomalous coupling is tested, all other anomalous couplings are assumed to be zero, and only the real couplings in Eq. (1.34) are considered i.e. with $a_i \geq 0$ and $\cos(\phi_{ai}) = \pm 1$ assumption.

Before going forward, let us recall the Eqs. (1.32) and (1.33) and rewrite them in a more appropriate format reading as,

$$\sigma^{\text{on-shell}} \propto \mu \tag{5.9}$$

$$\sigma^{\text{off-shell}} \propto \mu\, \Gamma_{\text{H}} = \mu^{\text{off-shell}} \tag{5.10}$$

where μ ($\mu^{\text{off-shell}}$) is defined as the on-shell (off-shell) signal strength, that is the ratio of the observed on-shell (off-shell) H boson cross section relative to the SM expectation. This ratio is defined as either μ_{F} (and respectively $\mu_{\text{F}}^{\text{off-shell}}$ for off-shell region) for H boson production via GF (ggH), or μ_{V} (and respectively $\mu_{\text{V}}^{\text{off-shell}}$ for off-shell region) for H boson production through VBF or in association with an EW vector boson W or Z (VH). In the analysis of the $2\ell 2\nu$ final state, the on-shell signal strength is not accessible due to the instrumental $p_{\text{T}}^{\text{miss}}$ background which is overwhelming in the on-shell region. However, constraints on this parameter can be obtained by combining with other final states with similar decay properties, such as $\text{H} \to 4\ell$. In the present analysis, constraints on the on-shell signal strengths are obtained by using $\text{H} \to \text{ZZ} \to 4\ell$ data [93, 102, 103]. We also make use of the off-shell 4ℓ data [93, 102, 103] in combination with the $2\ell 2\nu$ data, in order to compute the best constraints on the off-shell signal strengths and H boson width. Since the 4ℓ channel involves the same couplings at H boson production and decay as our $2\ell 2\nu$ channel, the same off-shell signal strength parameters enter in the combined fit of $2\ell 2\nu$ and 4ℓ data as in the fit of $2\ell 2\nu$ data alone.

The generic probability density for any considered process with potential contribution of interference between the signal and background amplitudes can be written in a form similar to that in Ref. [15] as,

$$\begin{aligned}\mathcal{P}_{jk}(\vec{x};\vec{\xi}_{jk},\vec{\zeta}) &= \tilde{\mu}_j\, \mathcal{P}_{jk}^{\text{sig}}\left(\vec{x};\vec{\xi}_{jk},f_{ai},\phi_{ai}\right) + \sqrt{\tilde{\mu}_j}\, \mathcal{P}_{jk}^{\text{int}}\left(\vec{x};\vec{\xi}_{jk},f_{ai},\phi_{ai}\right) \\ &\quad + \mathcal{P}_{jk}^{\text{bkg}}\left(\vec{x};\vec{\xi}_{jk}\right),\end{aligned} \tag{5.11}$$

where the "sig", "int" and "bkg" abbreviations in the superscript of $\mathcal{P}$ stand for signal, interference and background terms respectively. The $\vec{x}$ are the observables examined in the analysis described in Sec.sec:observables and the binning for these observables in each jet category is outlined in Table 5.3. The $\vec{\zeta}$ are the unconstrained parameters of interest such as $\mu_{\mathrm{F}}, \mu_{\mathrm{V}}, \Gamma_{\mathrm{H}}, \bar{f}_{ai}$ or any other re-parametrization based on the interpretation. The set of observables used in each interpretation are also summarized in Table 5.4. The j is the index of the process, which can be GF (on-shell or off-shell); VBF, ZH, or WH (on-shell); EW (VBF, ZH, WH combined, off-shell); or any of the non-

Table 5.3: The binning used in each of the observables is listed. Binning for $m_{\mathrm{T}}^{\mathrm{ZZ}}$is identical in each N_j category, and the binning for $p_{\mathrm{T}}^{\mathrm{miss}}$ is different between $N_j <$ and ≥ 2.

Observable	Bin boundaries
$m_{\mathrm{T}}^{\mathrm{ZZ}}$(GeV)	(300, 350, 400, 450, 500, 550, 600, 700, 850, 1000, 1250, 1500, 13000)
$p_{\mathrm{T}}^{\mathrm{miss}}$ (GeV , $N_j < 2$)	(125, 150, 200, 300, 500, 1000, 13000)
$p_{\mathrm{T}}^{\mathrm{miss}}$ (GeV , $N_j \geq 2$)	(140, 200, 13000)
$\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF}}$, $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF},\Lambda 1}$	(0, 0.05, 0.1, 0.2, 0.8, 0.9, 0.95, 1)
$\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF},a2}$, $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF},a3}$	(0, 0.1, 0.2, 0.4, 0.6, 0.8, 0.9, 1)

Table 5.4: The set of observables used in each interpretation are summarized for the $2\ell 2\nu$ final state. The interpretations are grouped by the parameters constrained. The observables for $N_j < 2$ are identical in any interpretation scenario, and the BSM $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF}}$ discriminant in the $N_j \geq 2$ category changes based on the anomalous HVV coupling constrained. The BSM discriminant $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF},a2}$ is added to the SM-like HVV coupling scenarios as well in order to gain equivalent signal separation strength. Only two bins in $p_{\mathrm{T}}^{\mathrm{miss}}$, < and $\geq$ 200 GeV , are considered for $N_j \geq 2$ as outlined in Table 5.3.

Interpretation parameters	$N_j < 2$	$N_j \geq 2$
$\mu_{\mathrm{F}}^{\text{off-shell}}$, $\mu_{\mathrm{V}}^{\text{off-shell}}$, $\mu^{\text{off-shell}}$	$m_{\mathrm{T}}^{\mathrm{ZZ}}$, $p_{\mathrm{T}}^{\mathrm{miss}}$	$m_{\mathrm{T}}^{\mathrm{ZZ}}$, $p_{\mathrm{T}}^{\mathrm{miss}}$, $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF}}$, $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF},a2}$
Γ_{H} ($f_{ai} = 0$)	$m_{\mathrm{T}}^{\mathrm{ZZ}}$, $p_{\mathrm{T}}^{\mathrm{miss}}$	$m_{\mathrm{T}}^{\mathrm{ZZ}}$, $p_{\mathrm{T}}^{\mathrm{miss}}$, $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF}}$, $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF},a2}$
Γ_{H} , $\bar{f}_{a2}$	$m_{\mathrm{T}}^{\mathrm{ZZ}}$, $p_{\mathrm{T}}^{\mathrm{miss}}$	$m_{\mathrm{T}}^{\mathrm{ZZ}}$, $p_{\mathrm{T}}^{\mathrm{miss}}$, $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF}}$, $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF},a2}$
Γ_{H} , $\bar{f}_{a3}$	$m_{\mathrm{T}}^{\mathrm{ZZ}}$, $p_{\mathrm{T}}^{\mathrm{miss}}$	$m_{\mathrm{T}}^{\mathrm{ZZ}}$, $p_{\mathrm{T}}^{\mathrm{miss}}$, $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF}}$, $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF},a3}$
Γ_{H} , $\bar{f}_{\Lambda 1}$	$m_{\mathrm{T}}^{\mathrm{ZZ}}$, $p_{\mathrm{T}}^{\mathrm{miss}}$	$m_{\mathrm{T}}^{\mathrm{ZZ}}$, $p_{\mathrm{T}}^{\mathrm{miss}}$, $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF}}$, $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF},\Lambda 1}$

interfering backgrounds. The index k refers to different event categories described in Sec. 5.4. The vector $\vec{\xi}_{jk}$ denotes the constrained nuisance parameters pertaining to the different processes in each category. The parameter $\tilde{\mu}_j$ equals either the on-shell signal strength μ_j for on-shell processes, or $\mu_j \cdot \dfrac{\Gamma_\mathrm{H}}{\Gamma_0}$ for the off-shell processes, with the reference value Γ_0 taken to be 4.07 MeV. There are two HVV vertices in the EW production mechanisms with the subsequent $\mathrm{H} \to \mathrm{VV} \to 2\ell 2\nu$ decay while there is only one HVV decay vertex in the GF case. In addition, there is interference with the background in the off-shell region. This leads to the following general expressions for the signal (sig) or interference (int) contributions appearing in Eq. (5.11) [15]:

$$\mathcal{P}_{jk}^{\mathrm{sig/int}}\left(\vec{x};\vec{\xi}_{jk},f_{ai},\phi_{ai}\right) = \sum_{m=0}^{M} \mathcal{P}_{jk,m}^{\mathrm{sig/int}}\left(\vec{x};\vec{\xi}_{jk}\right) f_{ai}^{\frac{m}{2}} \left(1-f_{ai}\right)^{\frac{M-m}{2}} \cos^{m}(\phi_{ai}). \quad (5.12)$$

In this expression, the index m corresponds to the exponent of a_i coupling in the squared scattering amplitude from Eq. (1.34), which may contain contributions from production and/or decay. The factor $\cos(\phi_{ai}) = \pm 1$ affects only the sign of the terms that scale with an odd power of a_i. The sum over the index m runs up to $M = 4$ in the case of the EW signal process; $M = 2$ in the case of the GF process, or the interference between the signal and background in the EW process, and $M = 1$ in the case of the interference between the signal and background in the GF process.

The final constraints on $\vec{\zeta}$ are placed using the profile likelihood method using the “RooFit” toolkit [104] within the “ROOT” [105] framework. The extended likelihood function is constructed using the probability densities as in (5.11) with each event characterized by the discrete category k and the set of observables $\vec{x}$. The likelihood $\mathcal{L}$ is maximized with respect to the nuisance parameters $\vec{\xi}_{jk}$ describing the systematic uncertainties discussed in the next section and the parameters of interest. The allowed 68% (95%) confidence-level (CL) interval is defined using the profile likelihood function, $-2\Delta\ln\mathcal{L} = 1.00$ (3.84) for one-parameter constraints, and $-2\Delta\ln\mathcal{L} = 2.30$ (5.99) for two-parameter constraints, for which exact coverage is expected in the asymptotic limit [106].

5.8 Systematic Uncertainties

Several systematic uncertainties are featured in the vectors of constrained parameters $\vec{\xi}_{jk}$. The template shapes describing the probability distributions in (5.11) and (5.12) are varied separately within theoretical, experimental, or statistical uncertainties. Most of the uncertainties affect both the shape of the observables and the normalization of the different contributions.

The following are the theoretical uncertainties considered in this analysis:

- **Renormalization scale**: This source is considered separately for $\mathrm{q\bar{q}} \to \mathrm{V_1V_2}$ ($\mathrm{V} = \mathrm{W}$ or Z), $\mathrm{q\bar{q}} \to \mathrm{V}\gamma$, processes with a t quark and a Z boson, and the QCD background sample used in the single-photon CR. The $\mathrm{gg} \to \mathrm{VV}$, and the EW $\mathrm{VV} + \mathrm{jets}$ production mechanisms with H boson contributions are also treated separately. The relative variations for the gg process obtained from the weight variations in the NLO simulation are normalized to the prediction of the variation of the inclusive NNLO k-factor [7].

- **Factorization scale**: The correlation scheme with the different processes is as mentioned for the renormalization scale.

- **Strong coupling** $\alpha_s(m_\mathrm{Z})$: A variation of $\alpha_s(m_\mathrm{Z}) = 0.118 \pm 0.0015$ is considered. The correlation scheme with the different processes is as mentioned above, and the GF process is adjusted in the same way.

- **Parton distribution function variations**: Ths variation is taken as a conservative, envelope-type variation evaluated on a per-event basis. The correlation scheme with the different processes is as mentioned above, and the GF process is adjusted in the same way.

- **Simulation of the second jet in GF samples**: The uncertainty is evaluated as the difference of the nominal POWHEG samples for $m_\mathrm{H} = 125$ GeV and $m_\mathrm{H} = 300$ GeV from the simulation with the MiNLO HJJ program [97] applied. The reweighting factors are extracted in three dimensions, in bins of m_VV below or above 150 GeV , in bins of $p_\mathrm{T}^\mathrm{VV}/m_\mathrm{VV}$, evaluated for the hard process, and in bins of N_j bins of 0, 1, and ≥ 2 with jets taken to be clustered using anti-k_T algorithm with a $\Delta R = 0.4$ parameter after hadronization.

- **Scale and tune variations of the hadronizer PYTHIA**: The scale variations are taken from the embedded variations of weights in the simulations

- **NLO EW correction on** $\mathrm{q\bar{q}} \to \mathrm{V_1V_2}$: This uncertainty is specific to the WZ and

ZZ processes, and they are treated as correlated [67, 107].

In addition to the theoretical uncertainties, the instrumental uncertainties on the simulation as a part of the systematic uncertainties are mentioned below:

- **Luminosity**: This source is a normalization uncertainty applied only to the simulation and is uncorrelated across the three data periods. The values of these uncertainties are provided in Refs. [108–111] and summarized in Ref. [112].
- **L1 prefiring scale**: This source of uncertainty is applied only in 2016 and 2017, where the prefiring weight needs to be applied. They are also treated as correlated.
- **Pile-up, JES, JER, and $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution correction**: These sources are uncorrelated across the years. The $p_{\mathrm{T}}^{\mathrm{miss}}$ resolution is evaluated for each of the pile-up, JES, and JER variation separately.
- **Lepton, trigger, pile-up jet identification, and b-tagging efficiencies**: The lepton efficiencies are measured with a tag-and-probe method using $\mathrm{Z} \to \ell\ell$ events with selection requirements that are orthogonal to the event preselection. The details of this method is fully described in the reference [18].

In the estimation of instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$ background, the $\gamma \to \ell\ell$ transfer factors in the estimation of the contamination from genuine-$p_{\mathrm{T}}^{\mathrm{miss}}$ contributions also account for the theoretical and instrumental uncertainties in the simulation in a correlated manner. In the estimation of the nonresonant background, the uncertainty on $f_{\mathrm{corr}}^{\ell\ell}$ from the statistics of the sideband control region is also taken into account. All components that enter into the statistical analysis also take into account at least two uncertainties to account for the shape and normalization of the different histograms.

5.9 Results

5.9.1 Event yields and distributions

The event distributions are shown for $m_{\mathrm{T}}^{\mathrm{ZZ}}$ in Fig. 5.45 for $N_j = 0$ and 1 categories. Distributions for $N_j \geq 2$ are shown in Fig. 5.46 which are split into two $p_{\mathrm{T}}^{\mathrm{miss}}$ bins, $<$ and ≥ 200 GeV . The distributions of $p_{\mathrm{T}}^{\mathrm{miss}}$ are explicitly demonstrated in Fig. 5.47 for the $N_j = 0$ and 1 jet categories. Moreover, distributions of the $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF}}$ discriminant corresponding to the SM-like HVV couplings, and those of the dedicated discriminants concerning BSM HVV couplings i.e. the $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF},a2}$, $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF},a3}$, and $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF},\Lambda1}$ (used in the a_2 , a_3 , and Λ_1 analyses) are illustrated in Figs. 5.48 and 5.49 for $p_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV , and Figs. 5.50 and 5.51 for $p_{\mathrm{T}}^{\mathrm{miss}} \geq 200$ GeV . The histograms in each of these figures correspond to the same contributions, so their description is collected in a common legend in Fig. 5.44. In these figures, the a_2 analysis is taken to exemplify the effect of BSM HVV couplings in the different distributions, unless other couplings are explicitly mentioned. Distributions are shown after a joint fit using all events used in the analysis, including on-shell and off-shell 4ℓ events. The fit is performed assuming the SM couplings and total width. The binning used for all of these observables is outlined in Table 5.3 and is used in the same way when interpretation scenarios are considered. The observed and postfit expected number of events are listed for the full analysis region $m_{\mathrm{T}}^{\mathrm{ZZ}} \geq 300$ GeV and a more signal-enriched $m_{\mathrm{T}}^{\mathrm{ZZ}} \geq 450$ GeV region in Table 5.5.

Table 5.5: The numbers of postfit expected and observed events are listed with the expected number of events split into its different signal (s), background (contin.) or signal-background interference (i) components. Postfit refers to the fit performed with the couplings and total width of the H boson assuming the SM couplings and total width. The expectations from the different signal contributions and the relevant signal-background interference terms are listed for each of the different SM or BSM hypotheses analyzed, but and the total number of expected events is only shown as illustration for the SM scenario. The signal, background, and interference contributions are shown separately for the gluon fusion (gg) and EW processes under the $\Gamma_{\mathrm{H}} = \Gamma_{\mathrm{H}}^{\mathrm{SM}}$ assumption. All data periods, and the $\mu\,\mu$ and e e decay channels are combined. Event yields in the $N_j \geq 2$ category are split into $p_{\mathrm{T}}^{\mathrm{miss}}$ < and $\geq$ 200 GeV . The vertical bars separate the event counts for $m_{\mathrm{T}}^{ZZ} \geq 300$ GeV (left) and $m_{\mathrm{T}}^{ZZ} \geq 450$ GeV (right). The abbreviations 'off. ' and 'on. ' stand for 'off-shell ' and 'on-shell ', respectively.

Contribution	$N_j = 0$	$N_j = 1$	$N_j \geq 2$ $p_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV	$N_j \geq 2$ $p_{\mathrm{T}}^{\mathrm{miss}} \geq 200$ GeV
EW contin.	5.85 | 3.13	16.6 | 9.24	6.66 | 0.687	15 | 9.9
EW s. (SM, off.)	2.18 | 1.54	6.77 | 4.8	1.14 | 0.148	4.71 | 3.92
EW i. (SM, off.)	−4.15 | −3.07	−13.4 | −9.77	−2.11 | −0.262	−9.19 | −7.85
EW s. ($\bar{f}_{a2} = 1$, off.)	6900 | 5170	15200 | 11600	3920 | 1170	25800 | 21800
EW i. ($\bar{f}_{a2} = 1$, off.)	−45.9 | −24.1	−118 | −65.4	−74.7 | −13.4	−231 | −171
EW s. ($\bar{f}_{a3} = 1$, off.)	29700 | 23600	70800 | 56100	17100 | 5520	130000 | 111000
EW i. ($\bar{f}_{a3} = 1$, off.)	−65.1 | −42.1	−214 | −131	−115 | −25.4	−501 | −366
EW s. ($\bar{f}_{\Lambda 1} = 1$, off.)	7670 | 6490	18800 | 16500	2440 | 1060	22800 | 20200
EW i. ($\bar{f}_{\Lambda 1} = 1$, off.)	12.2 | 6.68	66.2 | 46.4	17.4 | 2.58	86.1 | 72.2
EW s. (SM, on.)	1.53 | 0.289	0.976 | 0.265	0.198 | 0.0288	0.105 | 0.0529
EW s. ($\bar{f}_{a2} = 1$, on.)	546 | 170	633 | 384	113 | 34.9	54.6 | 34.7
EW s. ($\bar{f}_{a3} = 1$, on.)	1030 | 333	1320 | 924	262 | 56.7	112 | 72.5
EW s. ($\bar{f}_{\Lambda 1} = 1$, on.)	1920 | 1190	1820 | 1290	229 | 122	304 | 248
gg contin.	112 | 23.5	64.4 | 16.9	11.8 | 0.616	14.2 | 7.88
gg s. (SM, off.)	43.7 | 17.3	27.2 | 12.5	3.59 | 0.395	8.23 | 5.92
gg i. (SM, off.)	−58.6 | −25.1	−37.2 | −18.5	−4.63 | −0.538	−11.9 | −8.94
gg s. ($\bar{f}_{a2} = 1$, off.)	851 | 348	478 | 229	59.4 | 6.57	138 | 104
gg i. ($\bar{f}_{a2} = 1$, off.)	−246 | −51.8	−126 | −31.8	−21.1 | −1.29	−24.1 | −12.9
gg s. ($\bar{f}_{a3} = 1$, off.)	1920 | 795	1110 | 533	138 | 15.6	326 | 246
gg i. ($\bar{f}_{a3} = 1$, off.)	0 | 0	0 | 0	0 | 0	0 | 0
gg s. ($\bar{f}_{\Lambda 1} = 1$, off.)	183 | 72.6	113 | 52.2	14.8 | 1.62	34 | 24.7
gg i. ($\bar{f}_{\Lambda 1} = 1$, off.)	120 | 51.5	75.4 | 37.5	9.38 | 1.07	23.9 | 18.1
Instr. $p_{\mathrm{T}}^{\mathrm{miss}}$	99.7 | 14.2	130 | 24.5	59.6 | 4.92	17.6 | 7.67
Nonresonant	78.2 | 4.59	84.5 | 4.06	16.1 | 0.131	27.6 | 1.11
$\mathrm{q\bar{q}} \to$ ZZ	760 | 152	323 | 77.9	50.6 | 3.66	63.8 | 30.9
$\mathrm{q\bar{q}} \to$ WZ	238 | 34.9	161 | 27.9	41.6 | 3.03	32.1 | 13.5
tZ + X	0.434 | 0.0793	1.79 | 0.33	1.25 | 0.121	1.05 | 0.41
Total SM expected	1279 | 223	765 | 150	186 | 12.9	163 | 64.4
Total exp., no off.	1296 | 232	781 | 161	188 | 13.2	171 | 71.3
Total observed	1206 | 217	794 | 151	165 | 20	161 | 66

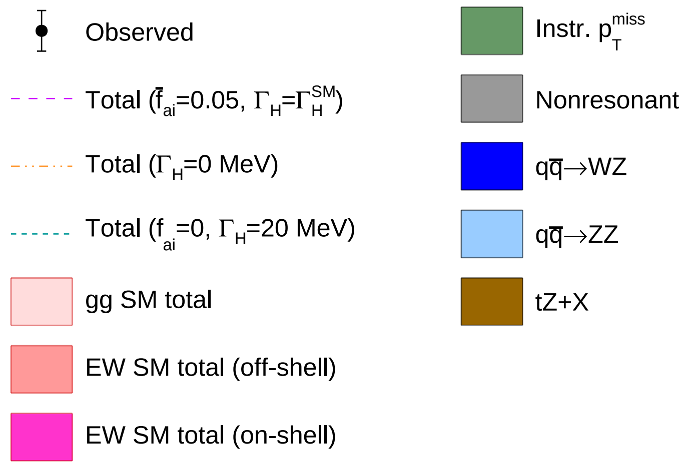

Figure 5.44: Shown are the legend of the different figures shown in this section. The different BSM HVV hypotheses are collectively abbreviated as f_{ai} . The filled histograms correspond to the SM expectations for the backgrounds, or the signal processes with the interfering backgrounds included. In the following figures, the middle panel shows the ratio of the different model predictions and of the data to the SM prediction, and the lower panel shows the expected composition of the event sample in the SM hypothesis. The pink dashed line shows the expectation for a BSM model with the a_i coefficient mentioned in the figure legend set to 0.05, and the H boson width set to the SM value.

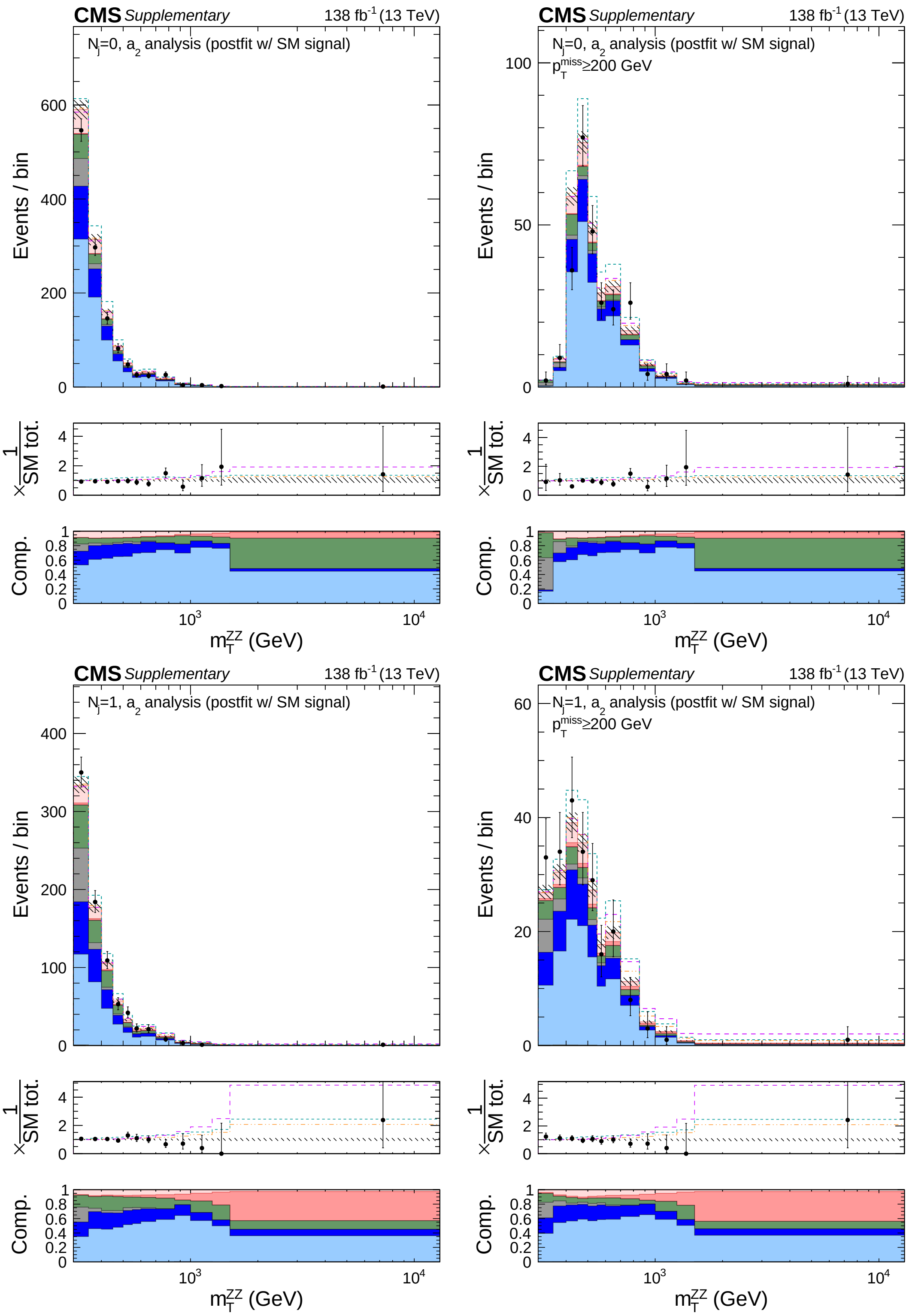

Figure 5.45: Shown are the distributions of $m_{\mathrm{T}}^{\mathrm{ZZ}}$in the $N_j = 0$ (top) and $N_j = 1$ (bottom) categories with (left) or without (right) a $p_{\mathrm{T}}^{\mathrm{miss}} \geq 200$ GeV requirement in order to reduce the contribution of the nonresonant and instrumental backgrounds. The various contributions are defined in Fig. 5.44.

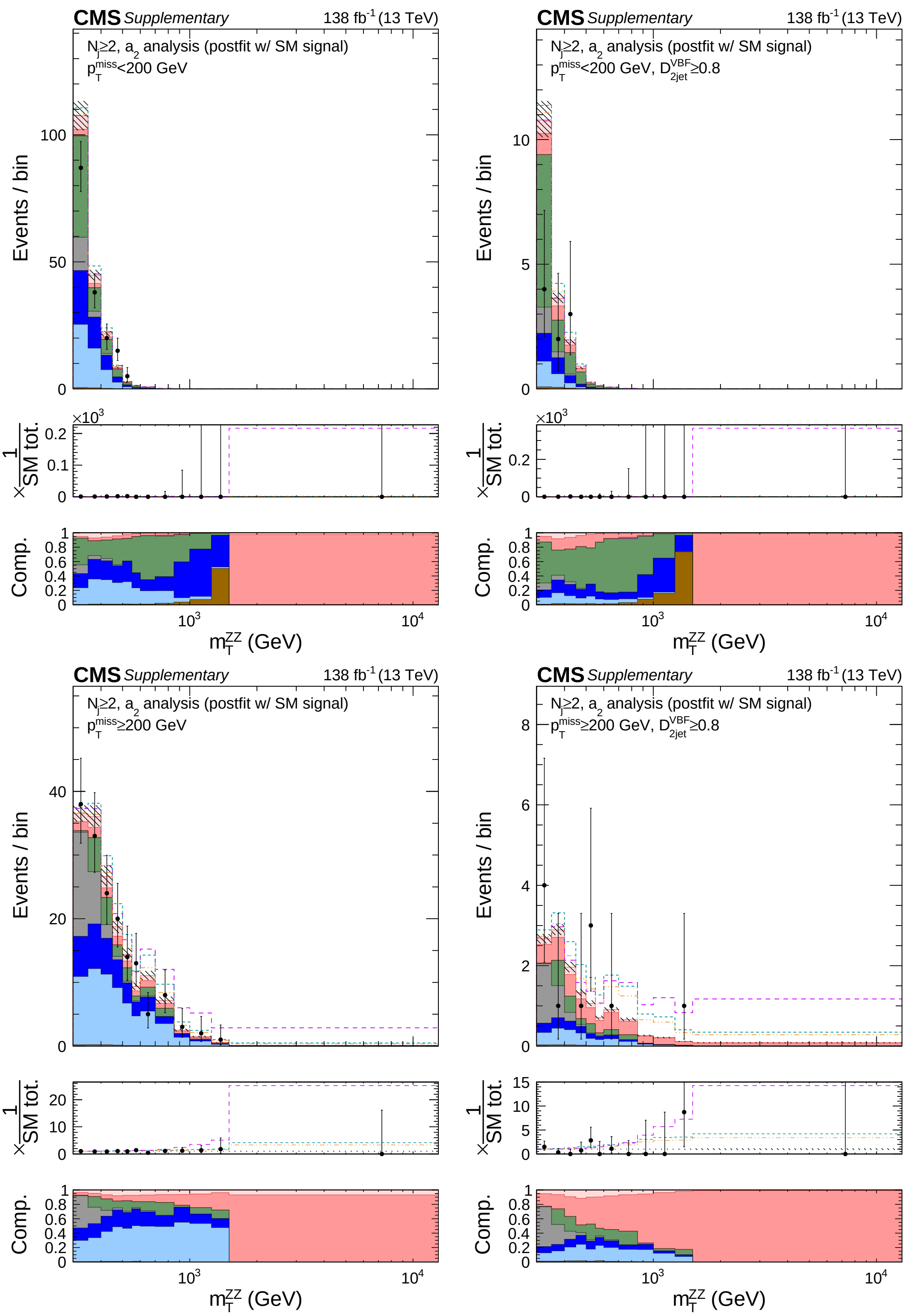

Figure 5.46: Shown are the distributions of $m_{\mathrm{T}}^{\mathrm{ZZ}}$in the $N_j \geq 2$ category for the $p_{\mathrm{T}}^{\mathrm{miss}} <$ 200 GeV (top) and $p_{\mathrm{T}}^{\mathrm{miss}} \geq$ 200 GeV (bottom) bins with (left) or without (right) a $\mathcal{D}_{\text{2jet}}^{\text{VBF}} \geq 0.8$ requirement to enhance the contribution of VBF-like events. The various contributions are defined in Fig. 5.44.

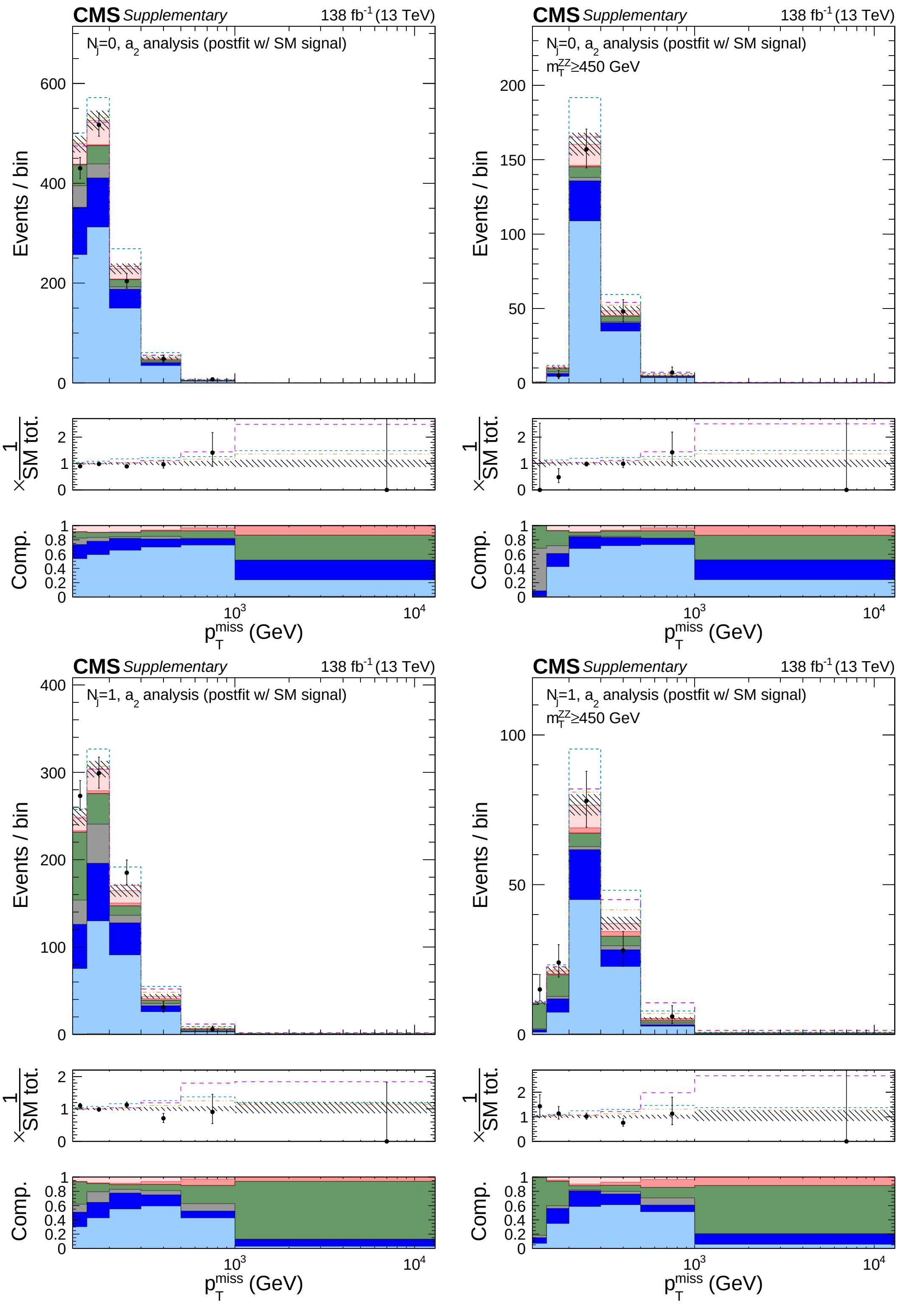

Figure 5.47: Shown are the distributions of $p_{\mathrm{T}}^{\mathrm{miss}}$ in the $N_j = 0$ (top) and $N_j = 1$ categories with (left) or without (right) an $m_{\mathrm{T}}^{ZZ} \geq 450$ GeV requirement in order to reduce the contribution of the nonresonant and instrumental backgrounds. The various contributions are defined in Fig. 5.44.

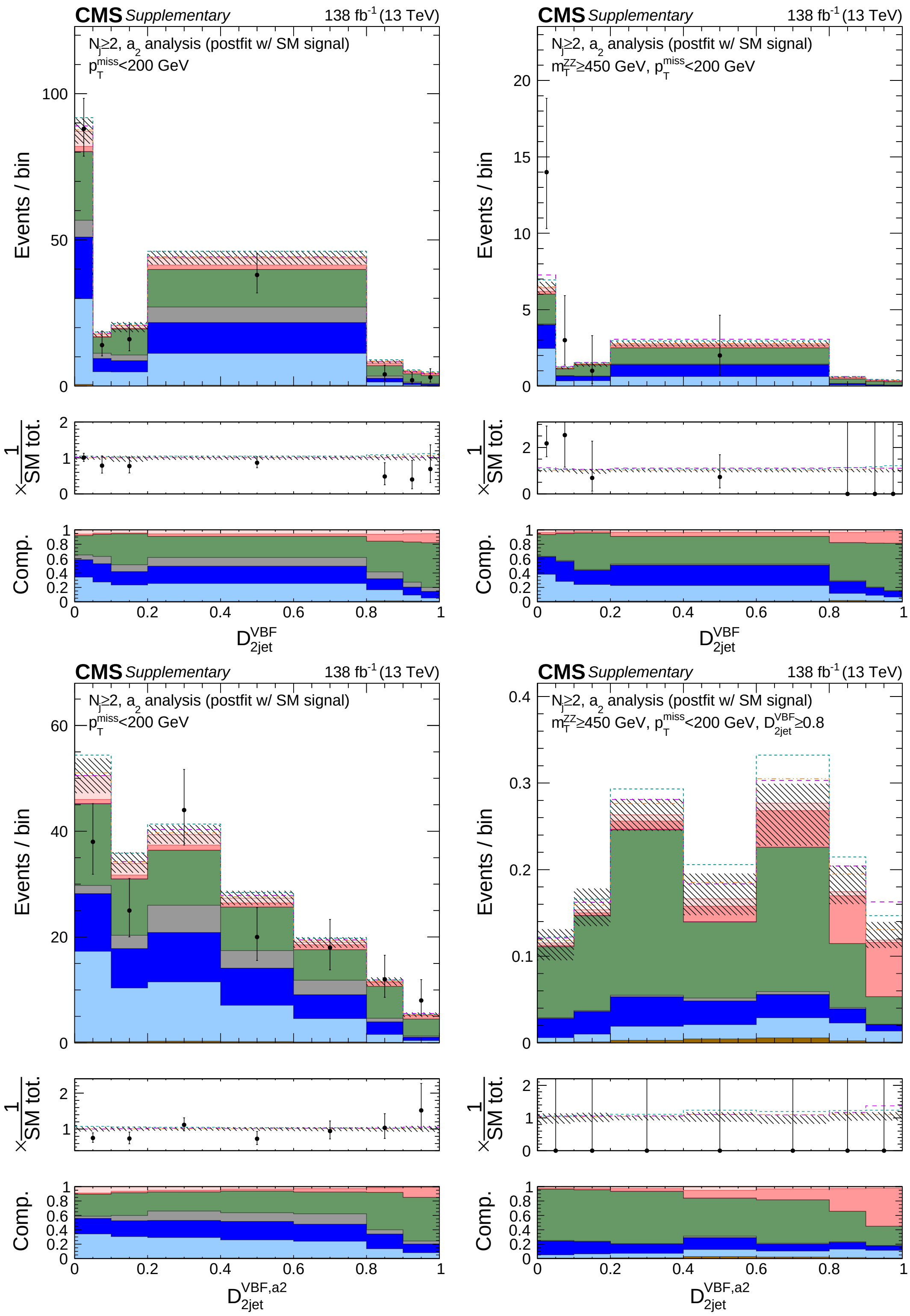

Figure 5.48: Shown are the distributions of $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF}}$ (top) and $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF},a2}$ (bottom) in the $N_j \geq 2$ category for the $p_{\mathrm{T}}^{\mathrm{miss}} < 200$ GeV bin with (left) or without (right) the $m_{\mathrm{T}}^{\mathrm{ZZ}} \geq 450$ GeV requirement in order to reduce the contribution of the nonresonant and instrumental backgrounds. The various contributions are defined in Fig. 5.44.

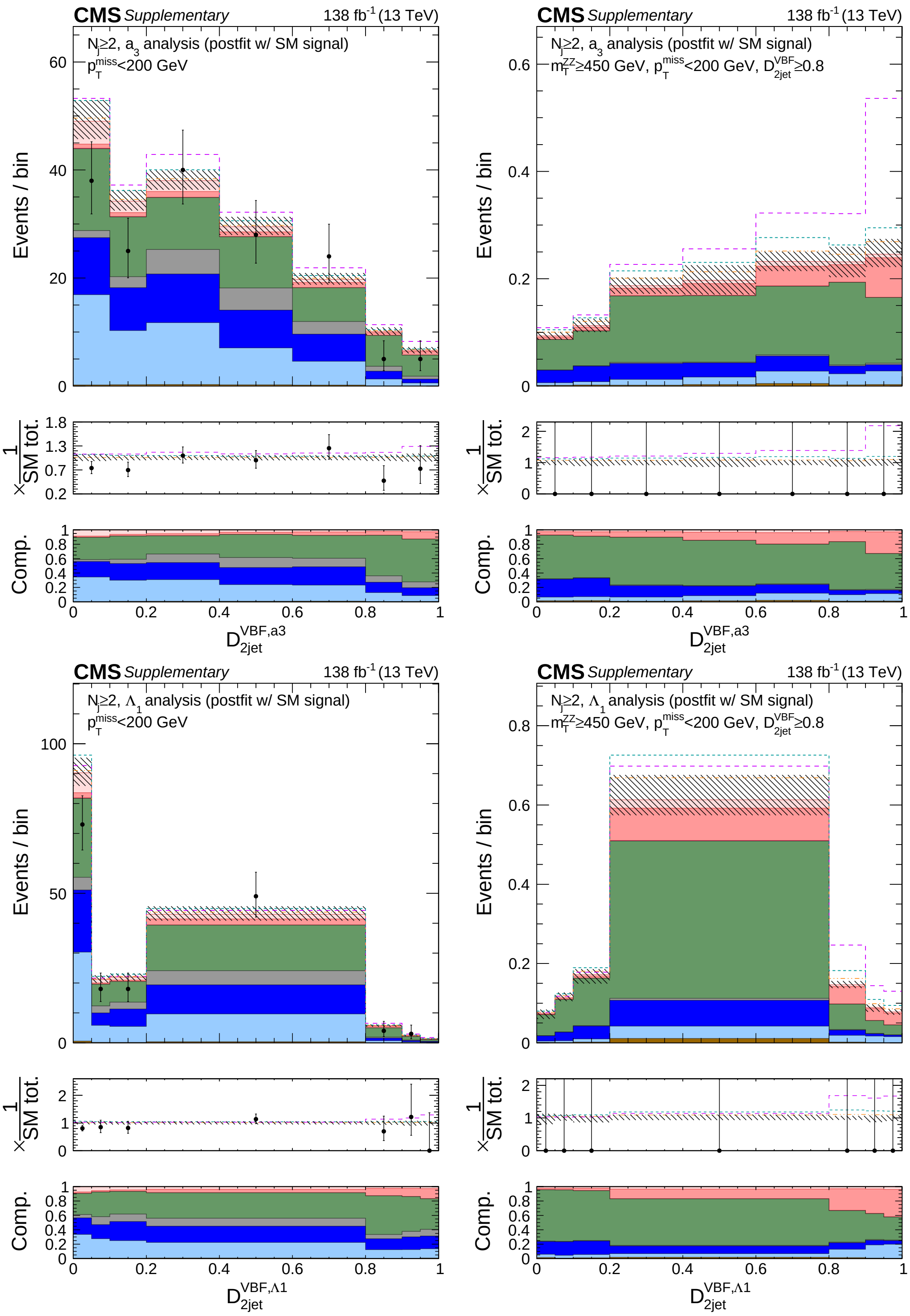

Figure 5.49: Shown are the distributions of $\mathcal{D}_{\text{2jet}}^{\text{VBF},a3}$ (top) and $\mathcal{D}_{\text{2jet}}^{\text{VBF},\Lambda1}$ (bottom) in the $N_j \geq 2$ category for the $p_{\text{T}}^{\text{miss}} < 200$ GeV bin with (left) or without (right) the $m_{\text{T}}^{ZZ} \geq 450$ GeV and $\mathcal{D}_{\text{2jet}}^{\text{VBF}} \geq 0.8$ requirements to enhance the contribution from SM VBF-like signal events. The various contributions are defined in Fig. 5.44.

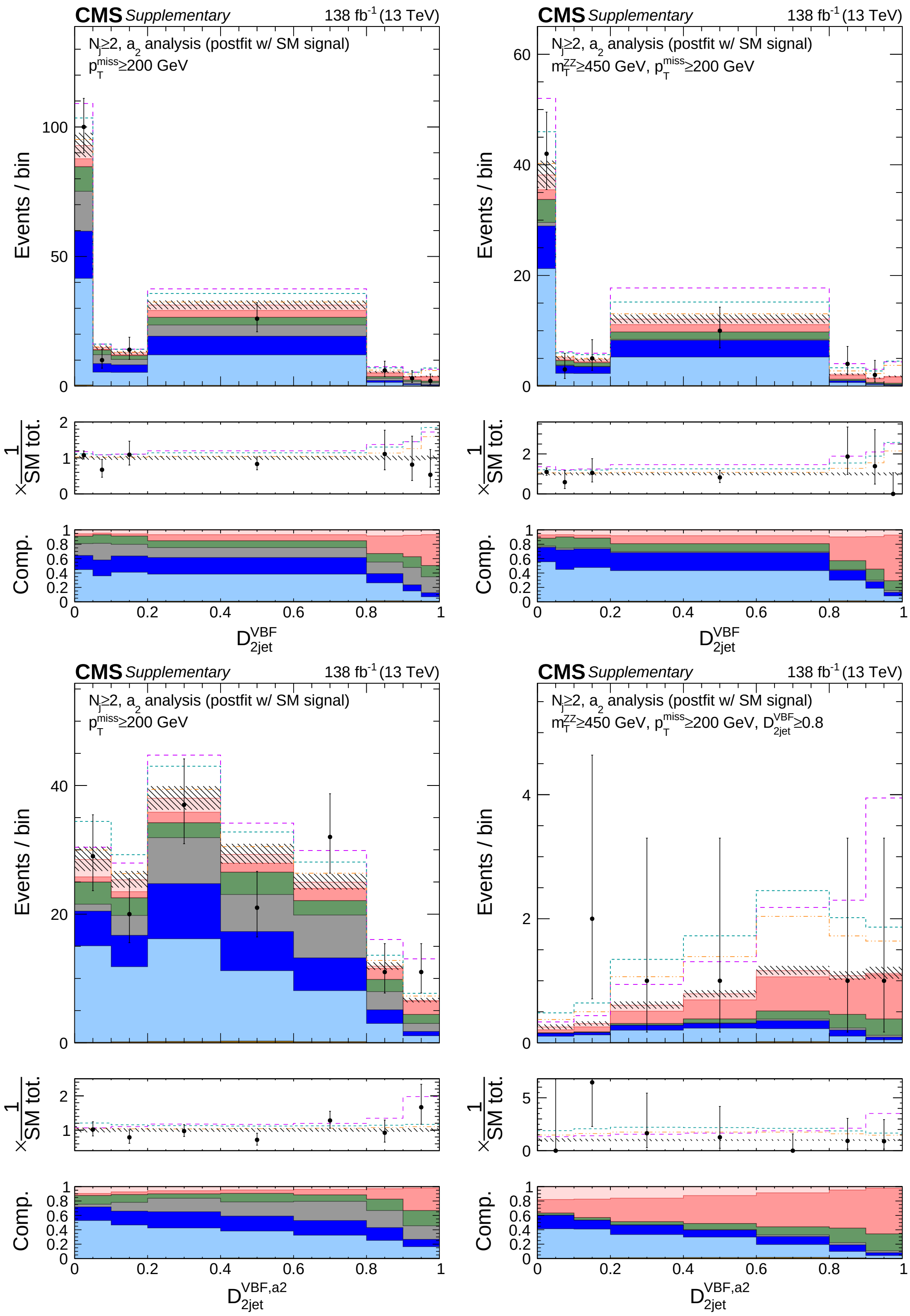

Figure 5.50: Shown are the distributions of $\mathcal{D}_{\text{2jet}}^{\text{VBF}}$ (top) and $\mathcal{D}_{\text{2jet}}^{\text{VBF},a2}$ (bottom) in the $N_j \geq 2$ category for the $p_{\text{T}}^{\text{miss}} \geq 200$ GeV bin with (left) or without (right) the $m_{\text{T}}^{\text{ZZ}} \geq 450$ GeV requirement in order to reduce the contribution of the nonresonant and instrumental backgrounds. The various contributions are defined in Fig. 5.44.

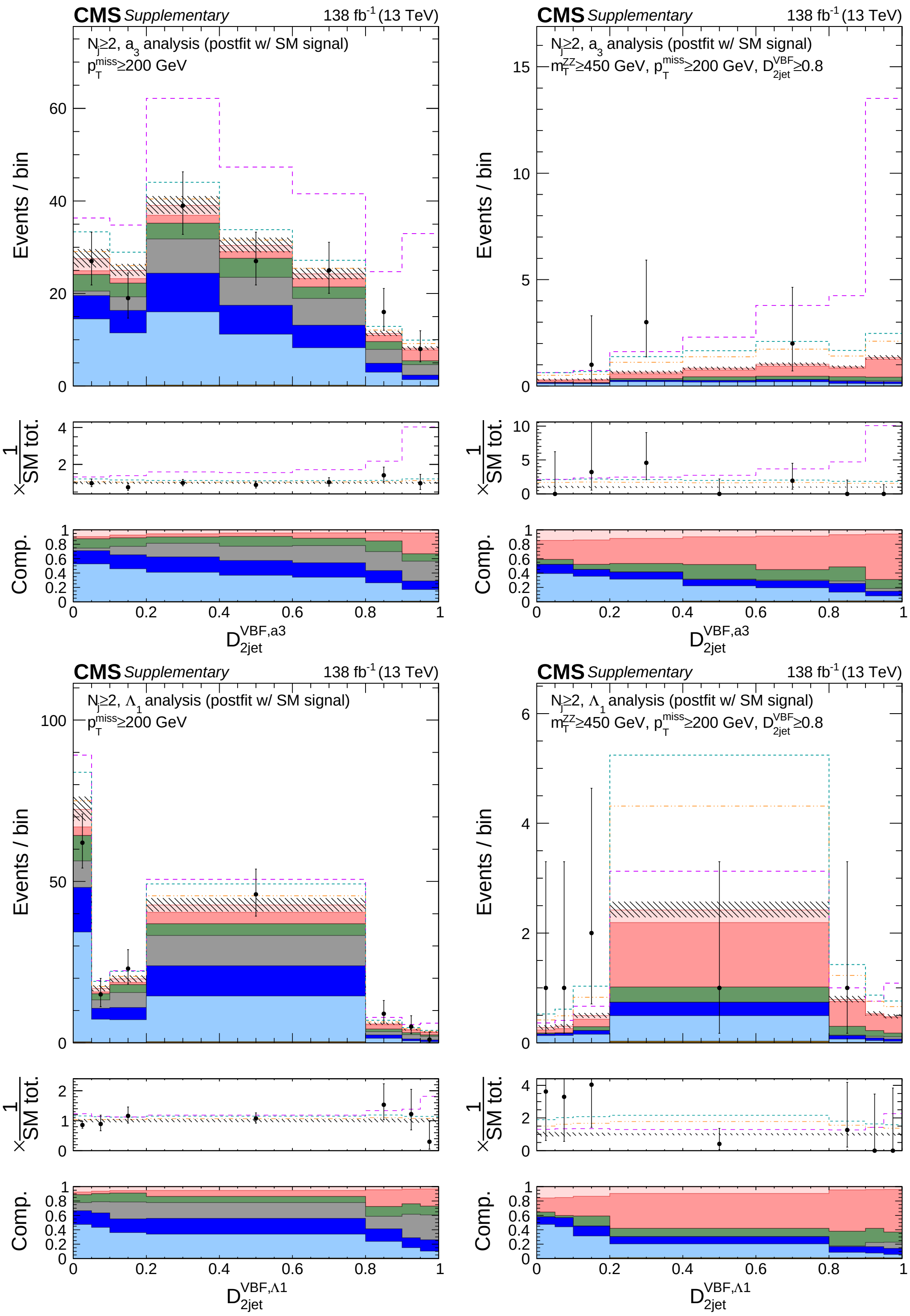

Figure 5.51: Shown are the distributions of $\mathcal{D}_{\text{2jet}}^{\text{VBF},a3}$ (top) and $\mathcal{D}_{\text{2jet}}^{\text{VBF},\Lambda1}$ (bottom) in the $N_j \geq 2$ category for the $p_{\text{T}}^{\text{miss}} \geq 200$ GeV bin with (left) or without (right) the $m_{\text{T}}^{\text{ZZ}} \geq 450$ GeV and $\mathcal{D}_{\text{2jet}}^{\text{VBF}} \geq 0.8$ requirements to enhance the contribution from SM VBF-like signal events. The various contributions are defined in Fig. 5.44.

5.9.2 Interpretations

The event distributions demonstrated in Sec. 5.9.1 are interpreted in several ways:

- Constraints on the off-shell signal strengths i.e. $\mu_{\mathrm{F}}^{\text{off-shell}}$, $\mu_{\mathrm{V}}^{\text{off-shell}}$, or $\mu^{\text{off-shell}}$.
- Constraints on the H boson total width, Γ_{H}, with or without the three BSM HVV coupling cross section fractions $\bar{f}_{a2}$, $\bar{f}_{a3}$, or $\bar{f}_{\Lambda 1}$.

Different observables are used according to different jet categories and also different BSM scenarios. The summary of observables used n different jet categories are outlined in Table 5.4 for each interpretation. In each case, the constraints are obtained using profile likelihood scans with the construction of the likelihood as pointed in Sec. 5.7. The coverage of the confidence intervals is tested using explicit p-value calculations at the reported 95% CL boundaries of Γ_{H} and at $\Gamma_{\mathrm{H}} = 0$ MeV according to the prescription of Feldman-Cousins confidence intervals [113].

The interpretations for $\mu_{\mathrm{F}}^{\text{off-shell}}$and $\mu_{\mathrm{V}}^{\text{off-shell}}$are considered with the other parameter kept unconstrained. When we are extracting the total off-shell signal strength $\mu^{\text{off-shell}}$, the ratio of the $\mu_{\mathrm{V}}^{\text{off-shell}}$ to $\mu_{\mathrm{F}}^{\text{off-shell}}$ i.e. $R_{\mathrm{V,F}}^{\text{off-shell}}$, is either profiled, or set to 1. These constraints are summarized in Table 5.6 for the combination of $2\ell 2\nu$ and high-mass 4ℓ events, or using $2\ell 2\nu$ events alone, and Fig. 5.52 and 5.53 show the corresponding observed and expected likelihood scans in these parameters. The two-parameter likelihood scan over $\mu_{\mathrm{F}}^{\text{off-shell}}$and $\mu_{\mathrm{V}}^{\text{off-shell}}$is also presented in this figure. The expected value of $-2\Delta\ln\mathcal{L}$ at $\mu^{\text{off-shell}} = 0$ exceeds the 95% CL threshold in both of the scenarios considered with events from the $2\ell 2\nu$ final state used alone, or when the events from the two final states are combined. The scenario with $\mu^{\text{off-shell}} = 0$ is excluded at a p-value of 0.0003 (3.6 standard deviations) in the observed result. This is the first evidence for off-shell contributions of the H boson, showing up as a destructive interference with the ZZ continuum as predicted by the standard model. This result strongly supports the standard model description of the interactions between the H boson and the Z bosons.

Constraints on the H boson total width, Γ_{H} , and joint constraints of the three BSM HVV coupling cross section fractions $\bar{f}_{a2}$, $\bar{f}_{a3}$, and $\bar{f}_{\Lambda 1}$ with Γ_{H} are obtained by a combined fit to the distribution of $2\ell 2\nu$ and 4ℓ events accounting for correlation between the uncertainties. Whereas the 4ℓ events were analyzed in Refs. [93, 102, 103] for the 2015–2018 data periods. The parameters of the fit are the on-shell signal strengths μ_{F} and μ_{V}, and the total width and in all cases the on-shell signal strength μ_{V} is left unconstrained. The ratio of the off-shell signal strengths is fixed to what the model considered predicts. The different constraints on the Γ_{H} either under the SM-like assumption or with one of the three f_{ai} parameters unconstrained are summarized in Table 5.7, and the corresponding likelihood scans are shown in Fig. 5.54. The observed results are similar between the

Table 5.6: Results on the off-shell signal strengths and Γ_{H} are summarized. The various fit conditions are indicated in the column labeled "Cond.": Results on $\mu^{\text{off-shell}}$ are presented with $R_{\mathrm{V,F}}^{\text{off-shell}} = \mu_{\mathrm{V}}^{\text{off-shell}}/\mu_{\mathrm{F}}^{\text{off-shell}}$ either unconstrained (u) or = 1, and constraints on $\mu_{\mathrm{F}}^{\text{off-shell}}$ and $\mu_{\mathrm{V}}^{\text{off-shell}}$ are shown with the other signal strength unconstrained. Results on Γ_{H} (in units of MeV) are obtained with the on-shell signal strengths unconstrained, and the different conditions listed for this quantity reflect which off-shell final states are combined with on-shell 4ℓ data. The expected central values, not quoted explicitly in this table, are either unity for $\mu^{\text{off-shell}}$, $\mu_{\mathrm{F}}^{\text{off-shell}}$, and $\mu_{\mathrm{V}}^{\text{off-shell}}$, or $\Gamma_{\mathrm{H}} = 4.1$ MeV.

Param.	Cond.	Observed		Expected	
		68 % CL	95 % CL	68 % CL	95 % CL
$\mu_{\mathrm{F}}^{\text{off-shell}}$	$\mu_{\mathrm{V}}^{\text{off-shell}}$(u)	$0.62^{+0.68}_{-0.45}$	${}^{+1.38}_{-0.614}$	${}^{+1.1}_{-0.99998}$	< 3.0
$\mu_{\mathrm{V}}^{\text{off-shell}}$	$\mu_{\mathrm{F}}^{\text{off-shell}}$ (u)	$0.90^{+0.9}_{-0.59}$	${}^{+2.0}_{-0.849}$	${}^{+2.0}_{-0.89}$	< 4.5
$\mu^{\text{off-shell}}$	$R_{\mathrm{V,F}}^{\text{off-shell}} = 1$	$0.74^{+0.56}_{-0.38}$	${}^{+1.06}_{-0.61}$	${}^{+1.0}_{-0.84}$	${}^{+1.7}_{-0.9914}$
	$R_{\mathrm{V,F}}^{\text{off-shell}}$ (u)	$0.62^{+0.68}_{-0.45}$	${}^{+1.38}_{-0.6139}$	${}^{+1.1}_{-0.99996}$	${}^{+2.0}_{-0.99999}$
Γ_{H}	$2\ell 2\nu + 4\ell$	$3.2^{+2.4}_{-1.7}$	${}^{+5.3}_{-2.7}$	${}^{+4.0}_{-3.5}$	${}^{+7.2}_{-4.07}$
Γ_{H}	$2\ell 2\nu$	$3.1^{+3.4}_{-2.1}$	${}^{+7.3}_{-2.9}$	${}^{+5.1}_{-3.7}$	${}^{+9.1}_{-4.099}$
Γ_{H}	4ℓ	$3.8^{+3.8}_{-2.7}$	${}^{+8.0}_{-3.73}$	${}^{+5.1}_{-4.05}$	< 13.8

SM-like scenario and with $\bar{f}_{a2}$ unconstrained, and between $\bar{f}_{a3}$ and $\bar{f}_{\Lambda1}$ unconstrained. Any of the constraint scenarios exclude $\Gamma_{\mathrm{H}} = 0$ MeV with more than 99.7% confidence in the observed result. The range of postfit sensitivity on the $\Gamma_{\mathrm{H}} = 0$ MeV hypothesis coming from each bin in the $2\ell2\nu$ and 4ℓ off-shell signal regions is visualized in Fig. 5.55.

Table 5.7: The constraints on the H boson total width Γ_{H} at 68% and 95% CL are summarized. They are reported for the analysis of the SM-like scenario, or an anomalous coupling parameter of interest unconstrained. The designation 'c.v.' stands for the central value obtained in the likelihood scan. The expected central value is always 4.1 MeV, so it is not quoted explicitly.

Parameter	Condition	Observed			Expected	
		Best fit	68%CL	95%CL	68%CL	95%CL
Γ_{H}(MeV)	SM-like	3.2	$[1.5, 5.6]$	$[0.5, 8.5]$	$[0.6, 8.1]$	$[0.03, 11.3]$
	f_{a2}(u)	3.4	$[1.6, 5.7]$	$[0.6, 8.4]$	$[0.5, 8.0]$	$[0.02, 11.3]$
	f_{a3}(u)	2.7	$[1.3, 4.8]$	$[0.5, 7.3]$	$[0.5, 8.0]$	$[0.02, 11.3]$
	$f_{\Lambda1}$(u)	2.7	$[1.3, 4.8]$	$[0.5, 7.3]$	$[0.6, 8.1]$	$[0.02, 11.3]$
$f_{a2}\,(\times10^5)$	$\Gamma_{\mathrm{H}} = \Gamma_{\mathrm{H}}^{\mathrm{SM}}$	79	$[6.6, 225]$	$[-32, 514]$	$[-78, 70]$	$[-359, 311]$
	Γ_{H}(u)	72	$[2.7, 216]$	$[-38, 503]$	$[-82, 73]$	$[-413, 364]$
$f_{a3}\,(\times10^5)$	$\Gamma_{\mathrm{H}} = \Gamma_{\mathrm{H}}^{\mathrm{SM}}$	2.2	$[-6.4, 32]$	$[-46, 107]$	$[-55, 55]$	$[-198, 198]$
	Γ_{H}(u)	2.4	$[-6.2, 33]$	$[-46, 110]$	$[-58, 58]$	$[-225, 225]$
$f_{\Lambda1}\,(\times10^5)$	$\Gamma_{\mathrm{H}} = \Gamma_{\mathrm{H}}^{\mathrm{SM}}$	2.9	$[-0.62, 17]$	$[-11, 46]$	$[-11, 20]$	$[-47, 68]$
	Γ_{H}(u)	3.1	$[-0.56, 18]$	$[-10, 47]$	$[-11, 21]$	$[-48, 75]$

The joint constraints can also be interpreted along the different $\bar{f}_{ai}$ variables for the $\Gamma_{\mathrm{H}} = \Gamma_{\mathrm{H}}^{\mathrm{SM}}$ and unconstrained Γ_{H} scenarios. Off-shell data from the $2\ell2\nu$ and 4ℓ channels do improve the constraints on the anomalous coupling parameters, but the improvement is only of order 20%. They are likewise summarized in Table 5.7 with the corresponding likelihood scans shown in Fig. 5.56.

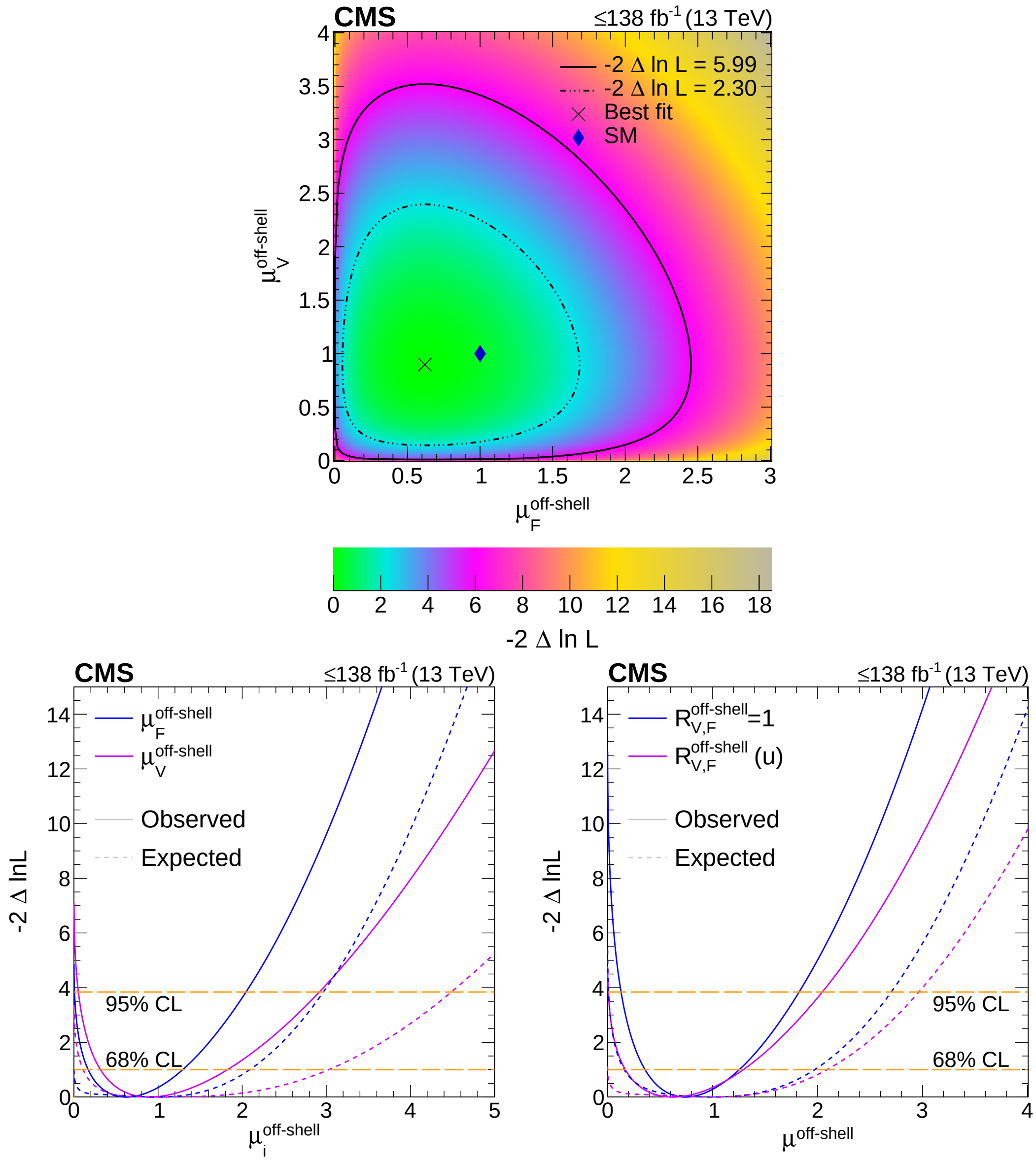

Figure 5.52: Top panel: Shown is the observed two-parameter likelihood scan of $\mu_{\mathrm{F}}^{\text{off-shell}}$ and $\mu_{\mathrm{V}}^{\text{off-shell}}$. The dot-dashed and dashed contours enclose the two-parameter 68% and 95% CL regions defined by $-2\Delta\ln\mathcal{L} = 2.30$ and $-2\Delta\ln\mathcal{L} = 5.99$, respectively. The cross marks the observed minimum, and the blue rhombus marks the SM values. Bottom panels: Shown are the observed (solid) and expected (dashed) likelihood scans of $\mu_{\mathrm{F}}^{\text{off-shell}}$, $\mu_{\mathrm{V}}^{\text{off-shell}}$ (left), and $\mu^{\text{off-shell}}$ (right). The likelihood scans for $\mu_{\mathrm{F}}^{\text{off-shell}}$ (blue) and $\mu_{\mathrm{V}}^{\text{off-shell}}$ (magenta) are obtained with the other parameter unconstrained, and in the case of $\mu^{\text{off-shell}}$, the likelihood scans are shown with (blue) and without (magenta) the constraint $R_{\mathrm{V,F}}^{\text{off-shell}} = 1$. The solid curves show the scans for the observed data whereas the dashed are for the expectation. The long-dashed horizontal lines show the one-parameter 68% ($-2\Delta\ln\mathcal{L} = 1.0$) and 95% ($-2\Delta\ln\mathcal{L} = 3.84$) CL regions.

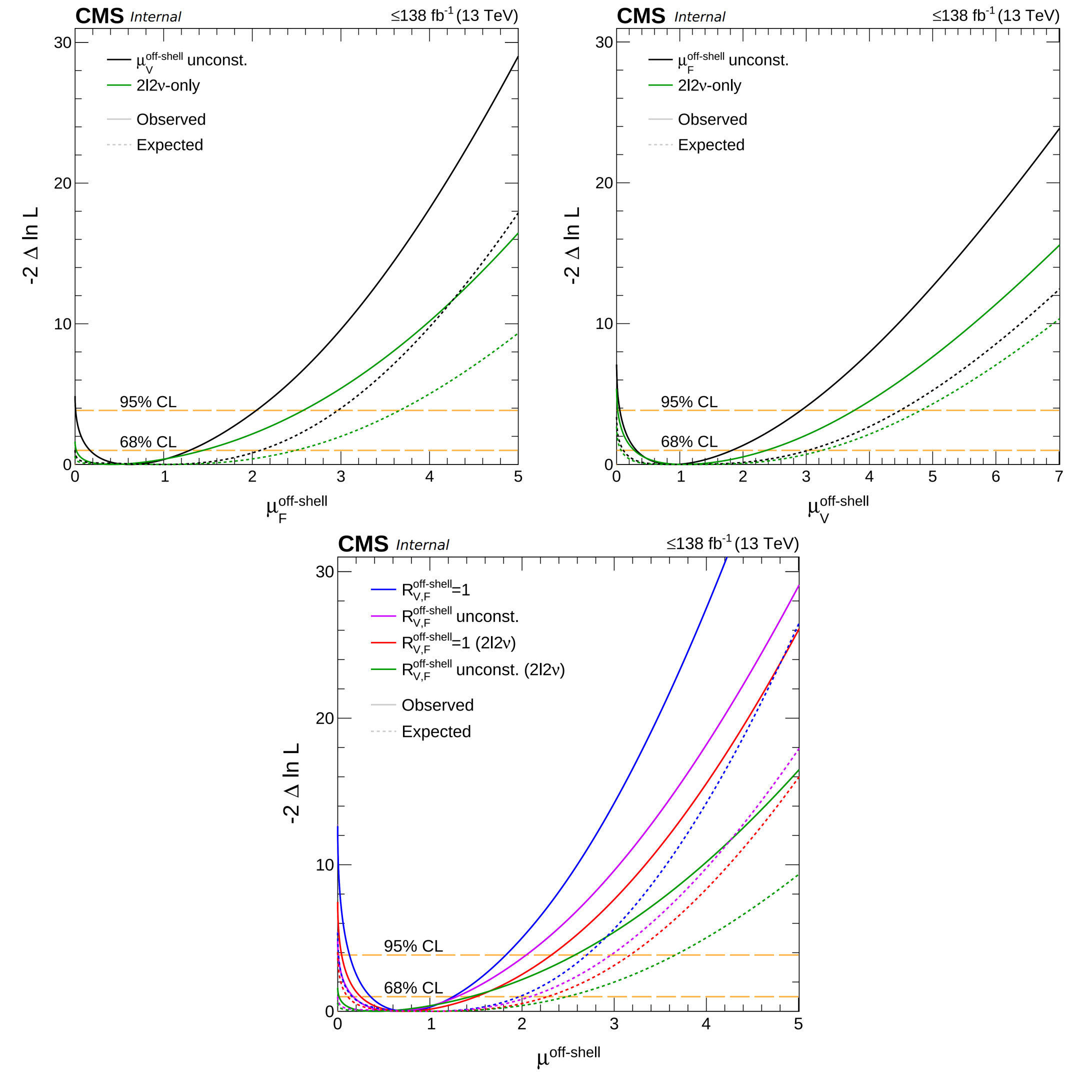

Figure 5.53: These likelihood scans are showing the observed (solid) and expected (dashed) likelihood scans of $\mu_{\mathrm{F}}^{\text{off-shell}}$, $\mu_{\mathrm{V}}^{\text{off-shell}}$, and $\mu^{\text{off-shell}}$ from left to right. The green or black curves in the first two panels show the results with the $2\ell 2\nu$ final state or the combination with 4ℓ events, respectively. The likelihood scans for $\mu_{\mathrm{F}}^{\text{off-shell}}$ and $\mu_{\mathrm{V}}^{\text{off-shell}}$ are obtained with the other parameter unconstrained, and in the case of $\mu^{\text{off-shell}}$, the interpretations with the constraint $R_{\mathrm{V,F}}^{\text{off-shell}} = 1$ (red for the $2\ell 2\nu$ final state only, and blue with the inclusion of 4ℓ events) or $R_{\mathrm{V,F}}^{\text{off-shell}}$ unconstrained (green for the $2\ell 2\nu$ final state only, and violet with the inclusion of 4ℓ events) are shown. In all cases, μ_{V} is constrained in the $2\ell 2\nu$ parametrization of the small on-shell EW H boson contribution. The solid curves show the scans for the observed data whereas the dashed are for the expectation. The long-dashed horizontal lines show the one-parameter 68% and 95% CL regions.

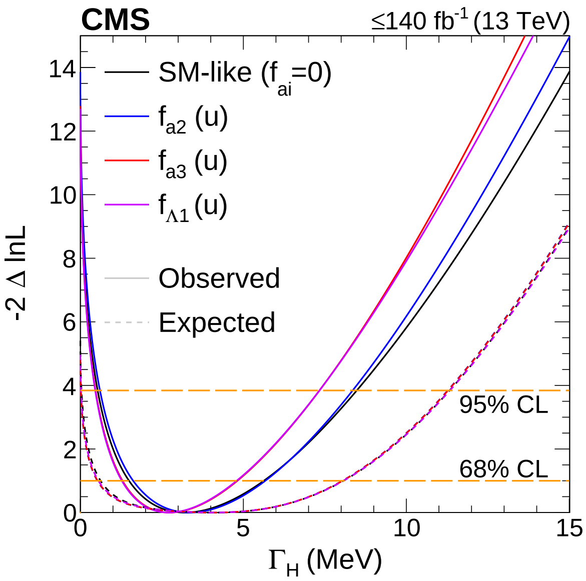

Figure 5.54: The observed (solid) and expected (dashed) likelihood scans of Γ_{H} are shown for the SM-like couplings analysis (black) or with one of the three anomalous coupling parameters, $\bar{f}_{a2}$ (blue), $\bar{f}_{a3}$ (red), or $\bar{f}_{\Lambda1}$ (magenta) unconstrained. The solid curves show the scans for the observed data whereas the dashed are for the expectation. The long-dashed horizontal lines show the one-parameter 68% ($-2\Delta\ln\mathcal{L} = 1.0$) and 95% ($-2\Delta\ln\mathcal{L} = 3.84$) CL regions.

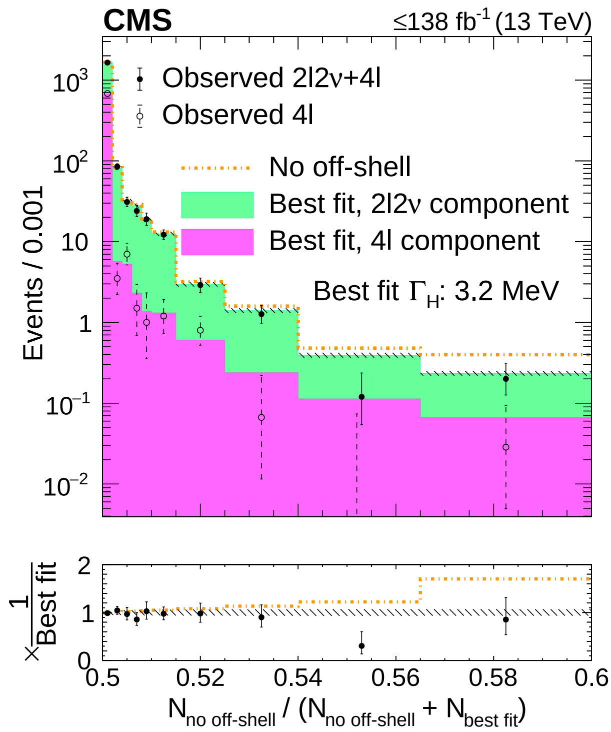

Figure 5.55: Distributions of ratios of the numbers of events in each off-shell signal region bin. The ratios are taken after separate fits to the no off-shell hypothesis ($N_{\text{no off-shell}}$) and the best overall fit ($N_{\text{best fit}}$) with the observed Γ_H value of 3.2 MeV in the SM-like HVV couplings scenario. The stacked histogram displays the predicted contributions (pink from the 4ℓ off-shell and green from the $2\ell2\nu$ off-shell signal regions) after the best fit, with the hashed band representing the total postfit uncertainty at 68% CL, and the gold dot-dashed line shows the predicted distribution of these ratios for a fit to the no off-shell hypothesis. The black solid (hollow) points, with error bars as uncertainties at 68% CL, represent the observed $2\ell2\nu$ and 4ℓ (4ℓ-only) data. The first and last bins contain the underflow and the overflow, respectively. The bottom panel displays the ratio of the various displayed hypotheses or observed data to the prediction from the best fit. The integrated luminosity reaches only up to 138 fb^{-1} since on-shell 4ℓ events are not displayed.

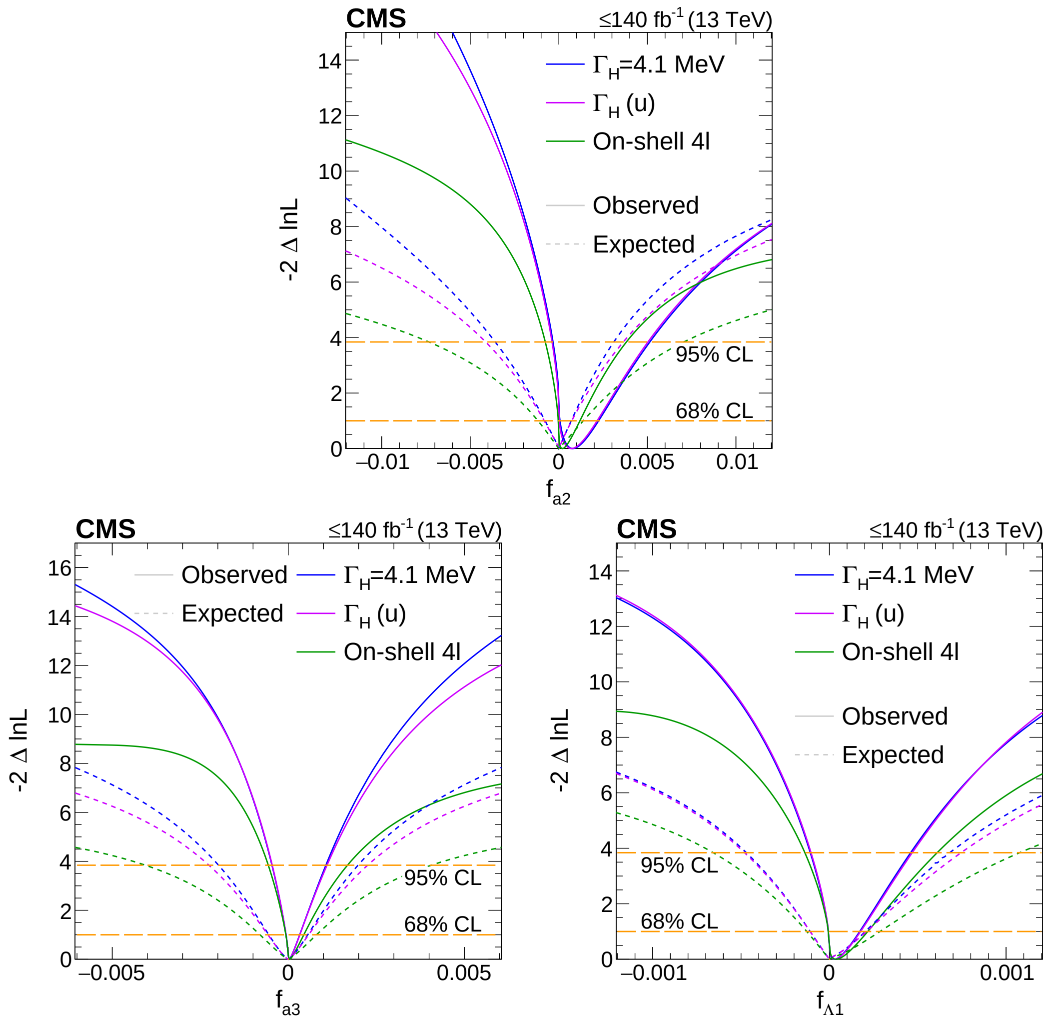

Figure 5.56: The likelihood scans of $\bar{f}_{a2}$ (top), $\bar{f}_{a3}$ (bottom left), and $\bar{f}_{\Lambda 1}$ (bottom right) are shown with the constraint $\Gamma_{\mathrm{H}} = \Gamma_{\mathrm{H}}^{\mathrm{SM}}$ (blue) or Γ_{H} unconstrained (magenta), or with only the on-shell 4ℓ data used (green). The solid curves show the scans for the observed data whereas the dashed are for the expectation. The dashed horizontal lines show the 68% and 95% CL regions.

Chapter 6

Conclusion and Perspectives

In this thesis the main focus was on H boson off-shell production, its decay width measurement and the couplings between the H boson and the electroweak gauge bosons in the off-shell region considering the SM and BSM hypotheses. For this purpose, we studied the $\mathrm{H} \to \mathrm{ZZ} \to 2\ell 2\nu$ processes in the main production modes i.e. the gluon fusion (GF) and vector boson fusion (VBF) processes.

From the analysis described in this thesis, evidence for off-shell contributions of the H boson to the production of Z boson pairs could be published for the first time, when combining our analysis of the $2\ell 2\nu$ final state with previously published results from the analysis of the 4ℓ channel. The off-shell contributions appear as a destructive interference in the ZZ spectrum at high m_{ZZ}. This interference is detected with a significance of 3.6 sigma when assuming that the ratio of the gluon-fusion and vector-boson fusion processes is as predicted by the standard model. Two-dimensional constraints on the gluon-fusion and vector-boson fusion signal strengths are also derived. Our findings support the standard-model description of the interaction between the H boson and the massive vector bosons, which, by unitarity constraints, should feature such a destructive interference. The sensitivity of the $2\ell 2\nu$ sample is found to be comparable to the one of the 4ℓ sample, and is even superior for the vector-boson fusion production process. This can be seen by comparing the likelihood scans of $\mu_{\mathrm{V}}^{\text{off-shell}}$ in Fig. 5.53 between $2\ell 2\nu$-only and the combination with 4ℓ events.

Combining the $2\ell 2\nu$ and the 4ℓ final states and the on-shell and off-shell event samples, the most precise measurement of the width of the Higgs boson was performed, $\Gamma_{\mathrm{H}} = 3.2^{+2.4}_{-1.7}$ MeV. This result is extracted under the assumption of the SM tensor structure for the HVV interaction. The picture does not change much when allowing the tensor structure of the HVV interaction to deviate from the standard model, as parametrized by anomalous couplings. Two-parameter fits of one anomalous coupling and the width show that the precision of the width measurement is not degraded significantly when allowing the presence of anomalous couplings.

In contrast with what was expected from the higher mass scale tested by off-shell events, we observed that the constraints on anomalous couplings improve little when including off-shell events, as compared to what can be learned from on-shell 4ℓ events. The improvements are of the order of 10%.

A similar study was later conducted by the ATLAS collaboration [114], but with differences in experimental methods and data analysis techniques. Specifically, while both experiments aimed to achieve the same goals, they differed in their event selection criteria, categorization, and observables used in the analysis. The result of ATLAS on the significance of the off-shell signal strength is 3.3 and the H boson decay width is measured to be $4.5^{+3.3}_{-2.5}$ MeV.

The results of the ATLAS and the CMS experiments firmly establish the presence of the off-shell H boson production. Furthermore, the uncertainty on the measurement of the off-shell production cross section is dominated by statistical uncertainties, as can be seen when comparing the results from the $2\ell 2\nu$ channel alone with the $2\ell 2\nu$ and 4ℓ combined results. This suggests clear ways of improving this measurement: adding the 2018 data to the CMS 4l off-shell analysis, combining the CMS result with the recently-published ATLAS result, and repeating the analysis with the run 3 dataset.

The leading systematic uncertainties are of theoretical nature. The uncertainty with the largest impact is the one that affects the NLO electroweak corrections to the irreducible ZZ and WZ backgrounds. In the present analysis, these corrections are applied as differential k-factors on top of simulations performed at NLO in QCD, and a conservative uncertainty is applied. The uncertainty is constrained using a 3-lepton control region targeting the WZ process, which is assumed to be affected by an uncertainty that is 100% correlated with the one affecting the ZZ background. In the future, our k-factors could be compared to the full NLO results that will be available at some point and a less conservative uncertainty could be used. The other theory uncertainties with a large impact are the ones due to the simulation of the signal and of the interfering background. In this respect, our analysis should already be close to optimal, as it relies on fixed-mass NLO samples reweighted to model the off-shell spectrum, while generators of the off-shell process are only available at LO. This is one of the main contributions of the author of the present thesis to this analysis.

The instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$ background is difficult to measure, therefore reducing it as much as possible is a good strategy. By investigating and improving the performance of filters showing anomalous $p_{\mathrm{T}}^{\mathrm{miss}}$ contributions, the author of this thesis contributed to make the analysis of the off-shell $2\ell 2\nu$ events more robust. Furthermore, the author of this thesis contributed to the optimization of the analysis selection cuts rejecting the instrumental $p_{\mathrm{T}}^{\mathrm{miss}}$ backgrounds.

Publications and Presentations

The publications and presentations related to this research as the main contributions of the author of this thesis are listed below.

Publications:

- A. Tumasyan et al., "Measurement of the Higgs boson width and evidence of its off-shell contributions to ZZ production", Nature Phys., vol. 18, no. 11, pp. 1329-1334, 2022. DOI : 10 . 1038 / s41567 - 022 - 01682 - 0. arXiv: 2202.06923 [hep-ex].
- CMS Collaboration, "Methods for off-shell Higgs boson production simulation used in CMS analyses", CMS-NOTE-2022-010, CERN-CMS-NOTE-2022-010, CERN, Geneva, Tech. Rep., 2022. [Online]. Available: https://cds.cern.ch/record/2826782.
- Performance of missing transverse momentum in pp collision at 13 TeV with the full Run2 dataset, CMS PAS JME-22-001, Analysis Note number: AN-21-167 (in CMS internal review)

Presentations:

- Analysis Approval Talk, " Off-shell Higgs production with $\mathrm{H} \rightarrow \mathrm{ZZ} \rightarrow 2\ell 2\nu$ Run 2", 13 October 2021, Special Higgs PAG Meeting.
- Higgs 2021 Conference, "Evidence for off-shell Higgs boson production and the measurement of its width", 18-22 October 2021, Stony Brook University.

Bibliography

[1] G. Aad *et al.*, 'Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC', *Phys. Lett. B*, vol. 716, pp. 1–29, 2012. DOI: `10.1016/j.physletb.2012.08.020`. arXiv: `1207.7214 [hep-ex]`.

[2] S. Chatrchyan *et al.*, 'Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC', *Phys. Lett. B*, vol. 716, pp. 30–61, 2012. DOI: `10.1016/j.physletb.2012.08.021`. arXiv: `1207.7235 [hep-ex]`.

[3] S. Chatrchyan *et al.*, 'Observation of a New Boson with Mass Near 125 GeV in pp Collisions at $\sqrt{s} = 7$ and 8 TeV', *JHEP*, vol. 06, p. 081, 2013. DOI: `10.1007/JHEP06(2013)081`. arXiv: `1303.4571 [hep-ex]`.

[4] G. Aad *et al.*, 'Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV', *JHEP*, vol. 08, p. 045, 2016. DOI: `10.1007/JHEP08(2016)045`. arXiv: `1606.02266 [hep-ex]`.

[5] 'A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery', *Nature*, vol. 607, no. 7917, pp. 52–59, 2022, [Erratum: Nature 612, E24 (2022)]. DOI: `10.1038/s41586-022-04893-w`. arXiv: `2207.00092 [hep-ex]`.

[6] 'A portrait of the Higgs boson by the CMS experiment ten years after the discovery', *Nature*, vol. 607, no. 7917, pp. 60–68, 2022. DOI: `10.1038/s41586-022-04892-x`. arXiv: `2207.00043 [hep-ex]`.

[7] D. de Florian *et al.*, 'Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector', vol. 2/2017, Oct. 2016. DOI: `10.23731/CYRM-2017-002`. arXiv: `1610.07922 [hep-ph]`.

[8] N. Kauer and G. Passarino, 'Inadequacy of zero-width approximation for a light Higgs boson signal', *JHEP*, vol. 08, p. 116, 2012. DOI: 10.1007/JHEP08(2012)116. arXiv: 1206.4803 [hep-ph].

[9] J. M. Campbell, R. K. Ellis, and C. Williams, 'Bounding the Higgs Width at the LHC Using Full Analytic Results for $gg \to e^- e^+ \mu^- \mu^+$', *JHEP*, vol. 04, p. 060, 2014. DOI: 10.1007/JHEP04(2014)060. arXiv: 1311.3589 [hep-ph].

[10] F. Caola and K. Melnikov, 'Constraining the Higgs boson width with ZZ production at the LHC', *Phys. Rev. D*, vol. 88, p. 054 024, 2013. DOI: 10.1103/PhysRevD.88.054024. arXiv: 1307.4935 [hep-ph].

[11] V. Khachatryan *et al.*, 'Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV', *Eur. Phys. J. C*, vol. 75, no. 5, p. 212, 2015. DOI: 10.1140/epjc/s10052-015-3351-7. arXiv: 1412.8662 [hep-ex].

[12] G. Aad *et al.*, 'Measurement of the Higgs boson mass from the $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$ channels with the ATLAS detector using 25 fb^{-1} of pp collision data', *Phys. Rev. D*, vol. 90, no. 5, p. 052 004, 2014. DOI: 10.1103/PhysRevD.90.052004. arXiv: 1406.3827 [hep-ex].

[13] A. M. Sirunyan *et al.*, 'Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV', *JHEP*, vol. 11, p. 047, 2017. DOI: 10.1007/JHEP11(2017)047. arXiv: 1706.09936 [hep-ex].

[14] M. Aaboud *et al.*, 'Constraints on off-shell Higgs boson production and the Higgs boson total width in $ZZ \to 4\ell$ and $ZZ \to 2\ell 2\nu$ final states with the ATLAS detector', *Phys. Lett. B*, vol. 786, pp. 223–244, 2018. DOI: 10.1016/j.physletb.2018.09.048. arXiv: 1808.01191 [hep-ex].

[15] A. M. Sirunyan *et al.*, 'Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state', *Phys. Rev. D*, vol. 99, no. 11, p. 112 003, 2019. DOI: 10.1103/PhysRevD.99.112003. arXiv: 1901.00174 [hep-ex].

[16] A. V. Gritsan, J. Roskes, U. Sarica, M. Schulze, M. Xiao, and Y. Zhou, 'New features in the JHU generator framework: constraining Higgs boson properties from on-shell and off-shell production', *Phys. Rev. D*, vol. 102, no. 5, p. 056 022, 2020. DOI: 10.1103/PhysRevD.102.056022. arXiv: 2002.09888 [hep-ph].

[17] A. M. Sirunyan *et al.*, 'Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state', *Phys. Rev. D*, vol. 104, no. 5, p. 052 004, 2021. DOI: 10.1103/PhysRevD.104.052004. arXiv: 2104.12152 [hep-ex].

[18] Campagnari, Claudio and Cheng, Yang and Ling, Jerry and Mahdavikhorrami, Mostafa and Popov, Andrey and Postiau, Nicolas and Sarica, Ulascan and Vanlaer, Pascal and Wang, Hanwen and Wang, Sicheng and Yuan, Li, 'Measurements of the Higgs boson properties from off-shell production in the $ZZ \to 2\ell 2\nu$ final state at 13 TeV', CMS Note 2018/081, 2018.

[19] M. Aaboud *et al.*, 'Measurements of Higgs boson properties in the diphoton decay channel with 36 fb^{-1} of pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector', *Phys. Rev. D*, vol. 98, p. 052 005, 2018. DOI: 10.1103/PhysRevD.98.052005. arXiv: 1802.04146 [hep-ex].

[20] S. Chatrchyan *et al.*, 'The CMS Experiment at the CERN LHC', *JINST*, vol. 3, S08004, 2008. DOI: 10.1088/1748-0221/3/08/S08004.

[21] A. Alici, M. Bomben, I. Dawson, and J. Sonneveld, 'The LHC machine and experiments', *CERN Yellow Rep. Monogr.*, vol. 1, I. Dawson, Ed., pp. 23–34, 2021. DOI: 10.23731/CYRM-2021-001.23.

[22] 'LHC Machine', *JINST*, vol. 3, L. Evans and P. Bryant, Eds., S08001, 2008. DOI: 10.1088/1748-0221/3/08/S08001.

[23] S. Chatrchyan *et al.*, 'Description and performance of track and primary-vertex reconstruction with the CMS tracker', *JINST*, vol. 9, no. 10, P10009, 2014. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[24] G. L. Bayatian *et al.*, 'CMS Physics: Technical Design Report Volume 1: Detector Performance and Software', 2006.

[25] A. M. Sirunyan *et al.*, 'Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV', *JINST*, vol. 13, no. 06, P06015, 2018. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det].

[26] A. M. Sirunyan *et al.*, 'Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV', *JINST*, vol. 15, no. 10, P10017, 2020. DOI: 10.1088/1748-0221/15/10/P10017. arXiv: 2006.10165 [hep-ex].

[27] 'CMS: The TriDAS project. Technical design report, Vol. 2: Data acquisition and high-level trigger', P. Sphicas, Ed., Dec. 2002.

[28] J. Ero *et al.*, 'The CMS drift tube trigger track finder', *JINST*, vol. 3, P08006, 2008. DOI: 10.1088/1748-0221/3/08/P08006.

[29] S. Claude, Ed., *Tenth workshop on electronics for LHC and future experiments, Boston, USA, 13-17 Sep 2004: Proceedings*, CERN Yellow Reports: Conference Proceedings, Nov. 2004, pp.353–357. DOI: 10.5170/CERN-2004-010.353.

[30] S. Claude, Ed., *Proceedings, eleventh Workshop on Electronics for LHC and Future Experiments, Heidelberg, Germany, 12-16 September 2005*, CERN Yellow Reports: Conference Proceedings, 2005, pp.279–283. DOI: 10.5170/CERN-2005-011.279.

[31] A. M. Sirunyan *et al.*, 'Particle-flow reconstruction and global event description with the CMS detector', *JINST*, vol. 12, no. 10, P10003, 2017. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[32] T. Speer, K. Prokofiev, R. Fruhwirth, W. Waltenberger, and P. Vanlaer, 'Vertex fitting in the CMS tracker', Feb. 2006.

[33] A. M. Sirunyan *et al.*, 'Pileup mitigation at CMS in 13 TeV data', *JINST*, vol. 15, no. 09, P09018, 2020. DOI: 10.1088/1748-0221/15/09/P09018. arXiv: 2003.00503 [hep-ex].

[34] S. Chatrchyan *et al.*, 'Measurement of the inclusive production cross sections for forward jets and for dijet events with one forward and one central jet in pp collisions at $\sqrt{s} = 7$ TeV', *JHEP*, vol. 06, p. 036, 2012. DOI: 10.1007/JHEP06(2012)036. arXiv: 1202.0704 [hep-ex].

[35] 'Measurement of the inelastic pp cross section at $\sqrt{s} = 7$ TeV', CERN, Geneva, Tech. Rep., 2011. [Online]. Available: https://cds.cern.ch/record/1433413.

[36] S. Chatrchyan *et al.*, 'The Performance of the CMS Muon Detector in Proton-Proton Collisions at $\sqrt{s}$ = 7 TeV at the LHC', *JINST*, vol. 8, P11002, 2013. DOI: 10.1088/1748-0221/8/11/P11002. arXiv: 1306.6905 [physics.ins-det].

[37] R. Frühwirth, 'Application of Kalman filtering to track and vertex fitting', *Nucl. Instrum. Meth. A*, vol. 262, p. 444, 1987. DOI: 10.1016/0168-9002(87)90887-4.

[38] A. Bodek, A. van Dyne, J. Y. Han, W. Sakumoto, and A. Strelnikov, 'Extracting Muon Momentum Scale Corrections for Hadron Collider Experiments', *Eur. Phys. J. C*, vol. 72, p. 2194, 2012. DOI: 10.1140/epjc/s10052-012-2194-8. arXiv: 1208.3710 [hep-ex].

[39] CMS Collaboration. (2022). Baseline muon selections for run-ii, [Online]. Available: https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideMuonIdRun2 (visited on 01/19/2022).

[40] V. Khachatryan *et al.*, 'Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV', *JINST*, vol. 10, no. 06, P06005, 2015. DOI: 10.1088/1748-0221/10/06/P06005. arXiv: 1502.02701 [physics.ins-det].

[41] S. Baffioni *et al.*, 'Electron reconstruction in CMS', *Eur. Phys. J. C*, vol. 49, p. 1099, 2007. DOI: 10.1140/epjc/s10052-006-0175-5.

[42] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, 'Reconstruction of electrons with the gaussian-sum filter in the CMS tracker at the LHC', *Journal of Physics G: Nuclear and Particle Physics*, vol. 31, no. 9, N9–N20, Jul. 2005. DOI: 10.1088/0954-3899/31/9/n01. arXiv: physics/0306087. [Online]. Available: https://doi.org/10.1088%2F0954-3899%2F31%2F9%2Fn01.

[43] T. Chen and C. Guestrin, 'XGBoost: A scalable tree boosting system', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16, San Francisco, California, USA: ACM, 2016, pp. 785–794, ISBN: 978-1-4503-4232-2. DOI: 10.1145/2939672.2939785. [Online]. Available: http://doi.acm.org/10.1145/2939672.2939785.

[44] V. Khachatryan *et al.*, 'Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV', *JINST*, vol. 10, no. 08, P08010, 2015. DOI: 10.1088/1748-0221/10/08/P08010. arXiv: 1502.02702 [physics.ins-det].

[45] S. Chatrchyan *et al.*, 'Measurement of the $W\gamma$ and $Z\gamma$ Inclusive Cross Sections in pp Collisions at $\sqrt{s} = 7$ TeV and Limits on Anomalous Triple Gauge Boson Couplings', *Phys. Rev. D*, vol. 89, no. 9, p. 092 005, 2014. DOI: 10.1103/PhysRevD.89.092005. arXiv: 1308.6832 [hep-ex].

[46] A. M. Sirunyan *et al.*, 'Search for new physics in final states with a single photon and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV', *JHEP*, vol. 02, p. 074, 2019. DOI: 10.1007/JHEP02(2019)074. arXiv: 1810.00196 [hep-ex].

[47] CMS Collaboration, 'Search for dark matter and large extra dimensions in monophoton final state using 2016 data (Common AN)', CMS Note 2016/224, 2016.

[48] 'Pileup Removal Algorithms', 2014.

[49] M. Cacciari and G. P. Salam, 'Dispelling the N^3 myth for the k_t jet-finder', *Phys. Lett. B*, vol. 641, pp. 57–61, 2006. DOI: 10.1016/j.physletb.2006.08.037. arXiv: hep-ph/0512210.

[50] M. Cacciari, G. P. Salam, and G. Soyez, 'The anti-k_t jet clustering algorithm', *JHEP*, vol. 04, p. 063, 2008. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[51] M. Cacciari, G. P. Salam, and G. Soyez, 'FastJet User Manual', *Eur. Phys. J. C*, vol. 72, p. 1896, 2012. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[52] 'Jet algorithms performance in 13 TeV data', CERN, Geneva, Tech. Rep. CMS-PAS-JME-16-003, 2017. [Online]. Available: http://cds.cern.ch/record/2256875.

[53] CMS Collaboration. (2020). Jet identification in high pile-up environment (pileupjetid), [Online]. Available: https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupJetID (visited on 09/17/2020).

[54] E. Bols, J. Kieseler, M. Verzetti, M. Stoye, and A. Stakia, 'Jet Flavour Classification Using DeepJet', Aug. 2020. arXiv: 2008.10519 [hep-ex].

[55] S. Chatrchyan *et al.*, 'Identification of b-quark jets with the CMS experiment', *JINST*, vol. 8, P04013, 2013. DOI: 10.1088/1748-0221/8/04/P04013. arXiv: 1211.4462 [hep-ex].

[56] A. M. Sirunyan *et al.*, 'Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV', *JINST*, vol. 13, P05011, 2018. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det].

[57] A. M. Sirunyan *et al.*, 'Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13$ TeV using the CMS detector', *JINST*, vol. 14, no. 07, P07004, 2019. DOI: 10.1088/1748-0221/14/07/P07004. arXiv: 1903.06078 [hep-ex].

[58] V. Khachatryan *et al.*, 'Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s}$ = 8 TeV', *JINST*, vol. 10, no. 02, P02006, 2015. DOI: 10.1088/1748-0221/10/02/P02006. arXiv: 1411.0511 [physics.ins-det].

[59] 'Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET', Apr. 2009.

[60] D. Bertolini, P. Harris, M. Low, and N. Tran, 'Pileup Per Particle Identification', *JHEP*, vol. 10, p. 059, 2014. DOI: 10.1007/JHEP10(2014)059. arXiv: 1407.6013 [hep-ph].

[61] V. Khachatryan *et al.*, 'Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV', *JINST*, vol. 12, no. 02, P02014, 2017. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].

[62] C. Collaboration, 'Missing transverse energy performance of the cms detector', *JINST*, vol. 6, P09001, 2011. DOI: 10.1088/1748-0221/6/09/P09001.

[63] C. Collaboration, 'Technical proposal for the upgrade of the cms detector through 2020', CERN-LHCC-2011-006 ; CMS-UG-TP-1 ; LHCC-P-004, 2011.

[64] CMS Collaboration. (2022). Met filter recommendations for run ii, [Online]. Available: https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2 (visited on 09/08/2022).

[65] Illia Babaunikau and Mariarosaria D'Alfonso and Zeynep Demiragli and Milos Dordevic and Loukas Gouskos and Isabell Melzer-Pellmann and Seema Sharma and Daniel Spitzbart and David Vannerom and Leonora Vesterbacka Olsson, 'Performance of missing transverse momentum in pp collision at 13 TeV using the full 2016 data', CMS Note 2018/003, 2018.

[66] CMS Collaboration. (2022). Jet identification, [Online]. Available: https://twiki.cern.ch/twiki/bin/view/CMS/JetID (visited on 10/26/2022).

[67] V. Khachatryan *et al.*, 'Constraints on the Higgs boson width from off-shell production and decay to Z-boson pairs', *Phys. Lett. B*, vol. 736, pp. 64–85, 2014. DOI: 10.1016/j.physletb.2014.06.077. arXiv: 1405.3455 [hep-ex].

[68] A. M. Sirunyan *et al.*, 'Search for a new scalar resonance decaying to a pair of Z bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV', *JHEP*, vol. 06, p. 127, 2018, [Erratum: JHEP 03, 128 (2019)]. DOI: 10.1007/JHEP06(2018)127. arXiv: 1804.01939 [hep-ex].

[69] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, 'An introduction to PYTHIA 8.2', *Comput. Phys. Commun.*, vol. 191, pp. 159–177, 2015. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[70] V. Khachatryan *et al.*, 'Event generator tunes obtained from underlying event and multiparton scattering measurements', *Eur. Phys. J. C*, vol. 76, no. 3, p. 155, 2016. DOI: 10.1140/epjc/s10052-016-3988-x. arXiv: 1512.00815 [hep-ex].

[71] A. M. Sirunyan *et al.*, 'Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements', *Eur. Phys. J. C*, vol. 80, no. 1, p. 4, 2020. DOI: 10.1140/epjc/s10052-019-7499-4. arXiv: 1903.12179 [hep-ex].

[72] R. D. Ball, V. Bertone, F. Cerutti, L. Del Debbio, S. Forte, A. Guffanti, J. I. Latorre, J. Rojo, and M. Ubiali, 'Unbiased global determination of parton distributions and their uncertainties at NNLO and at LO', *Nucl. Phys. B*, vol. 855, p. 153, 2012. DOI: 10.1016/j.nuclphysb.2011.09.024. arXiv: 1107.2652 [hep-ph].

[73] R. D. Ball *et al.*, 'Parton distributions for the LHC Run II', *JHEP*, vol. 04, p. 040, 2015. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[74] R. D. Ball *et al.*, 'Parton distributions from high-precision collider data', *Eur. Phys. J. C*, vol. 77, no. 10, p. 663, 2017. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph].

[75] A. Bierweiler, T. Kasprzik, and J. H. Kühn, 'Vector-boson pair production at the LHC to $\mathcal{O}(\alpha^3)$ accuracy', *JHEP*, vol. 12, p. 071, 2013. DOI: 10.1007/JHEP12(2013)071. arXiv: 1305.5402 [hep-ph].

[76] A. Manohar, P. Nason, G. P. Salam, and G. Zanderighi, 'How bright is the proton? A precise determination of the photon parton distribution function', *Phys. Rev. Lett.*, vol. 117, p. 242 002, 2016. DOI: 10.1103/PhysRevLett.117.242002. arXiv: 1607.04266 [hep-ph].

[77] M. Grazzini, S. Kallweit, and D. Rathlev, 'ZZ production at the LHC: fiducial cross sections and distributions in NNLO QCD', *Phys. Lett. B*, vol. 750, pp. 407–410, 2015. DOI: 10.1016/j.physletb.2015.09.055. arXiv: 1507.06257 [hep-ph].

[78] M. Grazzini, S. Kallweit, and M. Wiesemann, 'Fully differential NNLO computations with MATRIX', *Eur. Phys. J. C*, vol. 78, no. 7, p. 537, 2018. DOI: 10.1140/epjc/s10052-018-5771-7. arXiv: 1711.06631 [hep-ph].

[79] S. Frixione, P. Nason, and C. Oleari, 'Matching NLO QCD computations with parton shower simulations: the POWHEG method', *JHEP*, vol. 11, p. 070, 2007. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[80] P. Nason and C. Oleari, 'NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG', *JHEP*, vol. 02, p. 037, 2010. DOI: 10.1007/JHEP02(2010)037. arXiv: 0911.5299 [hep-ph].

[81] E. Bagnaschi, G. Degrassi, P. Slavich, and A. Vicini, 'Higgs production via gluon fusion in the POWHEG approach in the SM and in the MSSM', *JHEP*, vol. 02, p. 088, 2012. DOI: 10.1007/JHEP02(2012)088. arXiv: 1111.2854 [hep-ph].

[82] G. Luisoni, P. Nason, C. Oleari, and F. Tramontano, '$HW^{\pm}$/HZ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO', *JHEP*, vol. 10, p. 083, 2013. DOI: 10.1007/JHEP10(2013)083. arXiv: 1306.2542 [hep-ph].

[83] H. B. Hartanto, B. Jager, L. Reina, and D. Wackeroth, 'Higgs boson production in association with top quarks in the POWHEG BOX', *Phys. Rev. D*, vol. 91, p. 094 003, 2015. DOI: 10.1103/PhysRevD.91.094003. arXiv: 1501.04498 [hep-ph].

[84] G. Ferrera, M. Grazzini, and F. Tramontano, 'Associated WH production at hadron colliders: a fully exclusive QCD calculation at NNLO', *Phys. Rev. Lett.*, vol. 107, p. 152 003, 2011. DOI: 10.1103/PhysRevLett.107.152003. arXiv: 1107.1164 [hep-ph].

[85] K. Hamilton, P. Nason, C. Oleari, and G. Zanderighi, 'Merging H/W/Z + 0 and 1 jet at NLO with no merging scale: a path to parton shower + NNLO matching', *JHEP*, vol. 05, p. 082, 2013. DOI: 10.1007/JHEP05(2013)082. arXiv: 1212.4504 [hep-ph].

[86] S. Goria, G. Passarino, and D. Rosco, 'The Higgs Boson Lineshape', *Nucl. Phys. B*, vol. 864, p. 530, 2012. DOI: 10.1016/j.nuclphysb.2012.07.006. arXiv: 1112.5517 [hep-ph].

[87] Y. Gao, A. V. Gritsan, Z. Guo, K. Melnikov, M. Schulze, and N. V. Tran, 'Spin determination of single-produced resonances at hadron colliders', *Phys. Rev. D*, vol. 81, p. 075 022, 2010. DOI: 10.1103/PhysRevD.81.075022. arXiv: 1001.3396 [hep-ph].

[88] S. Bolognesi, Y. Gao, A. V. Gritsan, K. Melnikov, M. Schulze, N. V. Tran, and A. Whitbeck, 'Spin and parity of a single-produced resonance at the LHC', *Phys. Rev. D*, vol. 86, p. 095 031, 2012. DOI: 10.1103/PhysRevD.86.095031. arXiv: 1208.4018 [hep-ph].

[89] I. Anderson, S. Bolognesi, F. Caola, Y. Gao, A. V. Gritsan, C. B. Martin, K. Melnikov, M. Schulze, N. V. Tran, A. Whitbeck, and Y. Zhou, 'Constraining anomalous HVV interactions at proton and lepton colliders', *Phys. Rev. D*, vol. 89, p. 035 007, 2014. DOI: 10.1103/PhysRevD.89.035007. arXiv: 1309.4819 [hep-ph].

[90] A. V. Gritsan, R. Röntsch, M. Schulze, and M. Xiao, 'Constraining anomalous Higgs boson couplings to the heavy flavor fermions using matrix element techniques', *Phys. Rev. D*, vol. 94, p. 055 023, 2016. DOI: 10.1103/PhysRevD.94.055023. arXiv: 1606.03107 [hep-ph].

[91] D. de Florian *et al.*, 'Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs Sector', vol. 2/2017, Oct. 2016. DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].

[92] 'Measurements of Higgs boson properties from on-shell and off-shell production in the four-lepton final state', 2018. [Online]. Available: https://cds.cern.ch/record/2638605.

[93] A. M. Sirunyan *et al.*, 'Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state', *Phys. Rev. D*, vol. 99, no. 11, p. 112 003, 2019. DOI: 10.1103/PhysRevD.99.112003. arXiv: 1901.00174 [hep-ex].

[94] A. M. Sirunyan *et al.*, 'Search for a heavy Higgs boson decaying to a pair of W bosons in proton-proton collisions at $\sqrt{s} = 13\text{TeV}$ ', *JHEP*, vol. 03, p. 034, 2020. DOI: 10.1007/JHEP03(2020)034. arXiv: 1912.01594 [hep-ex].

[95] S. Dittmaier *et al.*, 'Handbook of LHC Higgs Cross Sections: 2. Differential Distributions', Jan. 2012. DOI: 10.5170/CERN-2012-002. arXiv: 1201.3084 [hep-ph].

[96] A. Tumasyan *et al.*, 'Measurement of the Higgs boson width and evidence of its off-shell contributions to ZZ production', *Nature Phys.*, vol. 18, no. 11, pp. 1329–1334, 2022. DOI: 10.1038/s41567-022-01682-0. arXiv: 2202.06923 [hep-ex].

[97] K. Hamilton, P. Nason, and G. Zanderighi, 'MINLO: multi-scale improved NLO', *JHEP*, vol. 10, p. 155, 2012. DOI: 10.1007/JHEP10(2012)155. arXiv: 1206.3572 [hep-ph].

[98] F. Caola, M. Dowling, K. Melnikov, R. Röntsch, and L. Tancredi, 'QCD corrections to vector boson pair production in gluon fusion including interference effects with off-shell Higgs at the LHC', *JHEP*, vol. 07, p. 087, 2016. DOI: 10.1007/JHEP07(2016)087. arXiv: 1605.04610 [hep-ph].

[99] A. Ballestrero *et al.*, 'Precise predictions for same-sign W-boson scattering at the LHC', *Eur. Phys. J. C*, vol. 78, no. 8, p. 671, 2018. DOI: 10.1140/epjc/s10052-018-6136-y. arXiv: 1803.07943 [hep-ph].

[100] V. Khachatryan *et al.*, 'Limits on the Higgs boson lifetime and width from its decay to four charged leptons', *Phys. Rev. D*, vol. 92, no. 7, p. 072 010, 2015. DOI: 10.1103/PhysRevD.92.072010. arXiv: 1507.06656 [hep-ex].

[101] R. J. Barlow, 'Extended maximum likelihood', *Nucl. Instrum. Meth. A*, vol. 297, p. 496, 1990. DOI: 10.1016/0168-9002(90)91334-8.

[102] A. M. Sirunyan *et al.*, 'Constraints on anomalous Higgs boson couplings using production and decay information in the four-lepton final state', *Phys. Lett. B*, vol. 775, p. 1, 2017. DOI: 10.1016/j.physletb.2017.10.021. arXiv: 1707.00541 [hep-ex].

[103] A. M. Sirunyan *et al.*, 'Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state', Apr. 2021. arXiv: 2104.12152 [hep-ex].

[104] W. Verkerke and D. P. Kirkby, 'The RooFit toolkit for data modeling', in *13th International Conference for Computing in High-Energy and Nuclear Physics (CHEP03)*, CHEP-2003-MOLT007, 2003. arXiv: physics/0306116 [physics].

[105] R. Brun and F. Rademakers, 'ROOT: An object oriented data analysis framework', *Nucl. Instrum. Meth. A*, vol. 389, p. 81, 1997. DOI: 10.1016/S0168-9002(97)00048-X.

[106] S. S. Wilks, 'The large-sample distribution of the likelihood ratio for testing composite hypotheses', *Annals Math. Statist.*, vol. 9, p. 60, 1938. DOI: 10.1214/aoms/1177732360.

[107] Postiau, Nicolas and Vanlaer, Pascal, 'Photon-induced electroweak corrections to the $pp \to W^{\pm}Z + X$ process at $\sqrt{s} = 13$ TeV, and associated uncertainties', CMS Note 2016/431, 2016.

[108] 'CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13\text{TeV}$ ', CERN, Geneva, Tech. Rep. CMS-PAS-LUM-18-002, 2019. [Online]. Available: https://cds.cern.ch/record/2676164.

[109] CMS Collaboration, 'CMS luminosity measurement for the 2017 data taking period at $\sqrt{s} = 13\text{TeV}$ ', CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018. [Online]. Available: https://cds.cern.ch/record/2621960.

[110] A. M. Sirunyan *et al.*, 'Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13\text{TeV}$ in 2015 and 2016 at CMS', *Submitted to EPJ C*, 2021. arXiv: 2104.01927 [hep-ex].

[111] CMS Collaboration, 'CMS luminosity measurements for the 2016 data taking period', CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017. [Online]. Available: https://cds.cern.ch/record/2257069.

[112] CMS Collaboration. (2020). Luminosity physics object group (lumi pog), [Online]. Available: https://twiki.cern.ch/twiki/bin/viewauth/CMS/TWikiLUM (visited on 12/16/2020).

[113] G. J. Feldman and R. D. Cousins, 'A Unified approach to the classical statistical analysis of small signals', *Phys. Rev. D*, vol. 57, pp. 3873–3889, 1998. DOI: 10.1103/PhysRevD.57.3873. arXiv: physics/9711021.

[114] G. Aad *et al.*, 'Evidence of off-shell Higgs boson production from ZZ leptonic decay channels and constraints on its total width with the ATLAS detector', Apr. 2023. arXiv: 2304.01532 [hep-ex].

[115] C. Collaboration, 'Methods for off-shell Higgs boson production simulation used in CMS analyses', CERN, Geneva, Tech. Rep., 2022. [Online]. Available: https://cds.cern.ch/record/2826782.

[116] CMS Collaboration, 'Methods for off-shell Higgs boson production simulation used in CMS analyses', CMS DP Note 2022/010, 2022.

Appendix A

Removing outliers in the reweighting procedure

The reweighting weights for a given sample and hypothesis may occasionally be too large a.k.a outliers which can artificially create irregular line shapes. The outliers are due to the statistical effect from generating a finite number of events in each sample. Therefore we need to avoid such effect by removing outliers larger than a threshold. For this, weight thresholds are determined individually for each m_{VV} bin in each sample by sorting the reweighting weights from largest to smallest and examining the first quantile of events according to a fraction per reweighting hypothesis, determined approximately according to the distribution of weights. This fraction is taken to be 0.001 for hypotheses that involve the continuum VV or BSM H boson contributions in the VBF samples, or 0.0005 for any other hypothesis or GF production mode samples. If the ratio of the first and the last weight of this quantile is larger than 5 for any of the hypothesis, the contribution from these events are removed for all other hypotheses too. After removing the events with too large reweighted wights considered as outliers, the final weights sum are no longer normalized to the actual sample cross section at the end. In order to compensate the sample cross section normalization, the rest of the events in this bin for this sample are scaled by the lost sum of native sample weights. This loss compensation procedure is also applied in the same way if all matrix element ratio weights happen to be 0 for any event, which only happens when the initial-final state configuration in the VBF or WH topologies correspond solely to $\mathrm{Wff'}$ couplings that scale with off-diagonal CKM terms (i.e. the MCFM matrix elements assume a diagonal CKM matrix).

Appendix B

Additional plots for validation of merging procedure in the reweighting process

These are the comparison displayed for on-shell H boson production in Figs. B.1, B.2, and B.3 for the VBF ($m_{\mathrm{jj}} > 130$ GeV), ZH (80 GeV $< m_{\mathrm{jj}} < 100$ GeV with compatible initial and final partons), and WH (70 GeV $< m_{\mathrm{jj}} < 90$ GeV with compatible initial and final partons) topologies, respectively.

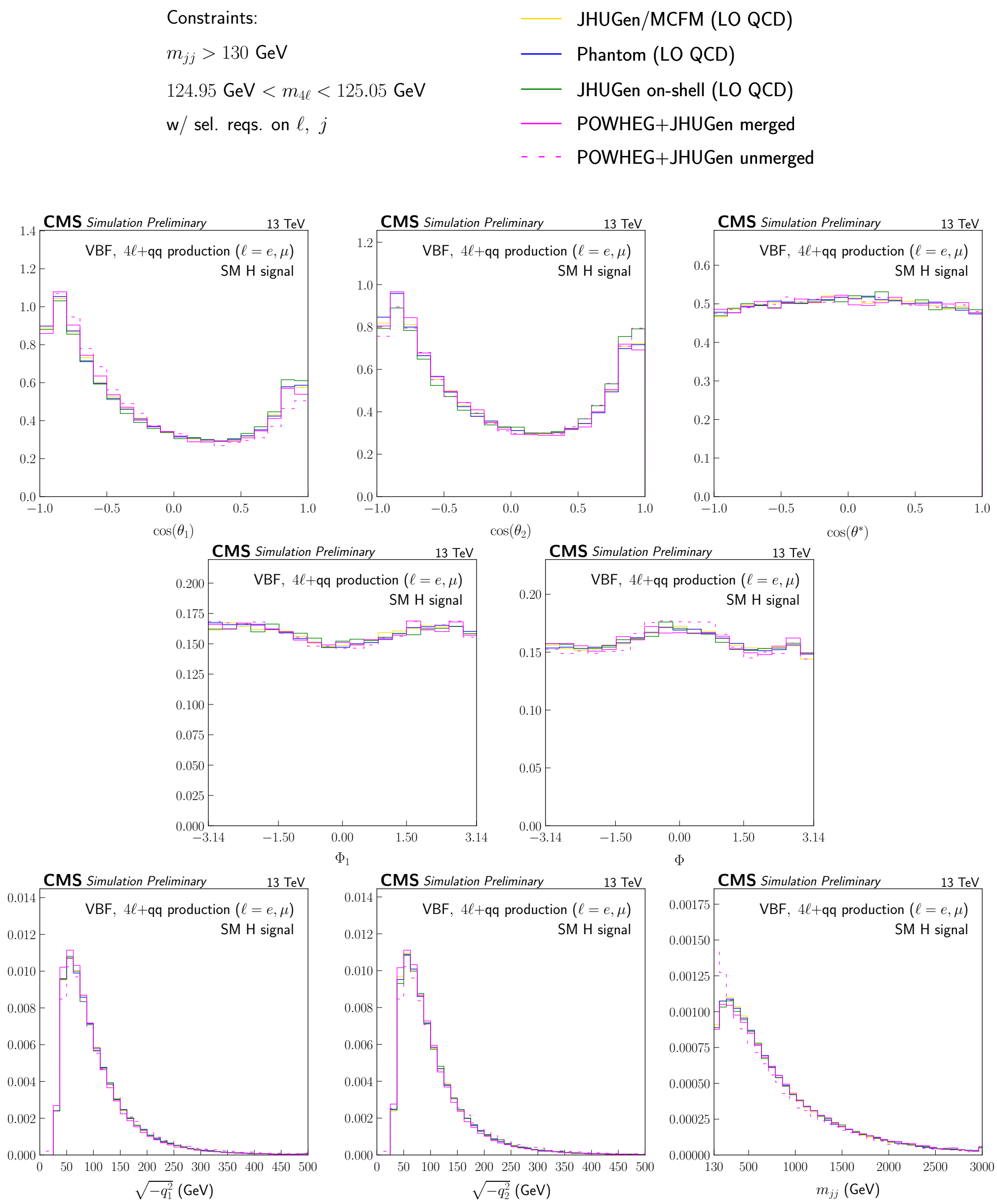

Figure B.1: The distributions of $\cos\theta_1$, $\cos\theta_2$, $\cos\theta^*$, Φ_1, Φ, $\sqrt{-q_1^2}$, $\sqrt{-q_2^2}$, and m_{jj} (from top left to bottom right) for on-shell H boson production through the SM EW signal process. The requirement $m_{\mathrm{jj}} > 130$ GeV is applied on all distributions to emphasize the VBF-like topology. The different predictions are shown on the legend at the top, and all distributions are normalized to unit area. The unmerged POWHEG scenario uses the two outgoing partons leading in p_{T} to compute the kinematic quantities.

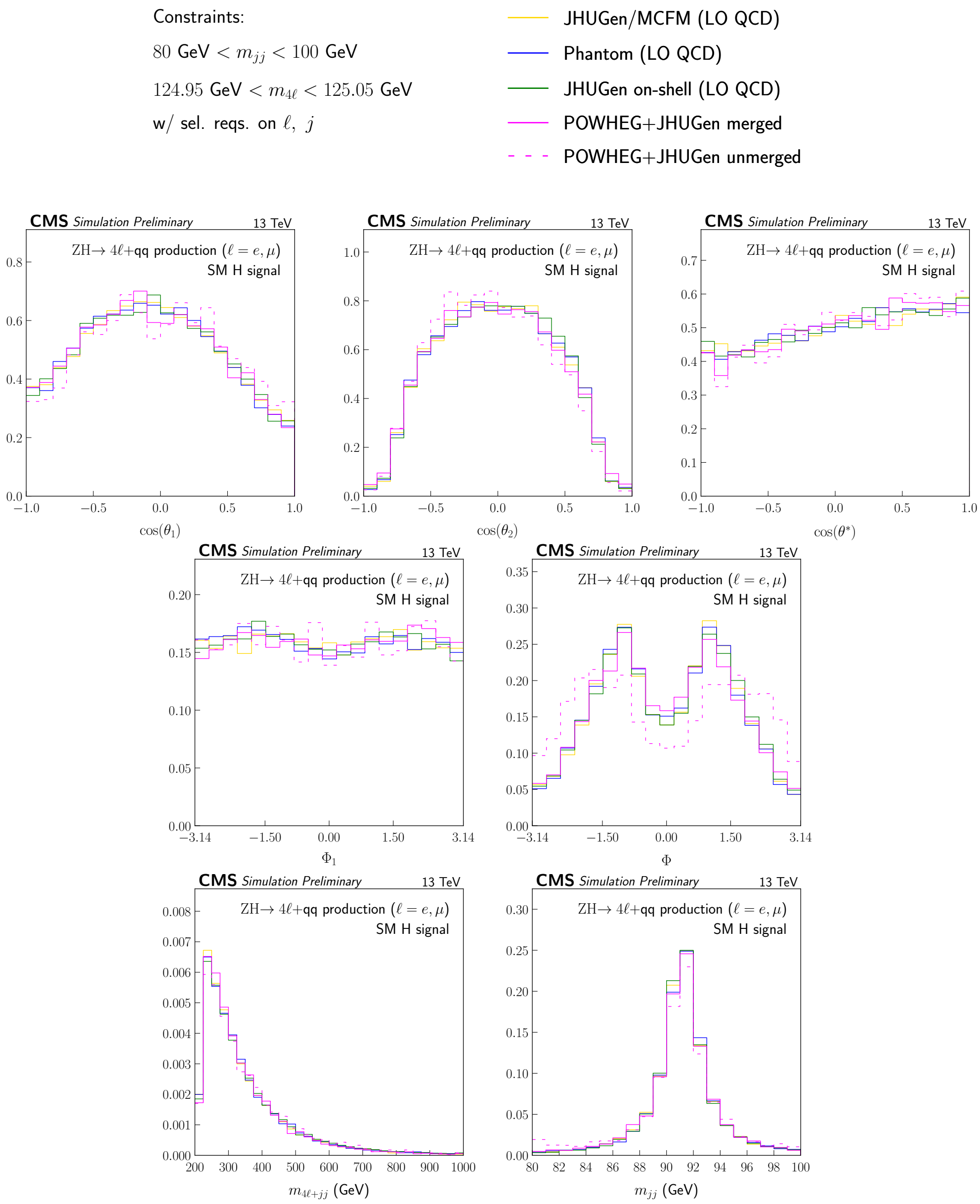

Figure B.2: The distributions of $\cos\theta_1$, $\cos\theta_2$, $\cos\theta^*$, Φ_1, Φ, $m_{4\ell+jj}$ and $m_{\rm jj}$ (from top left to bottom right) for on-shell H boson production through the SM EW signal process. The requirement 80 GeV $< m_{\rm jj} <$ 100 GeV is applied along with requirements on the initial and final state composition to emphasize the ZH-like topology. The different predictions are shown on the legend at the top, and all distributions are normalized to unit area. The unmerged POWHEG scenario uses the two outgoing partons leading in $p_{\rm T}$ to compute the kinematic quantities.

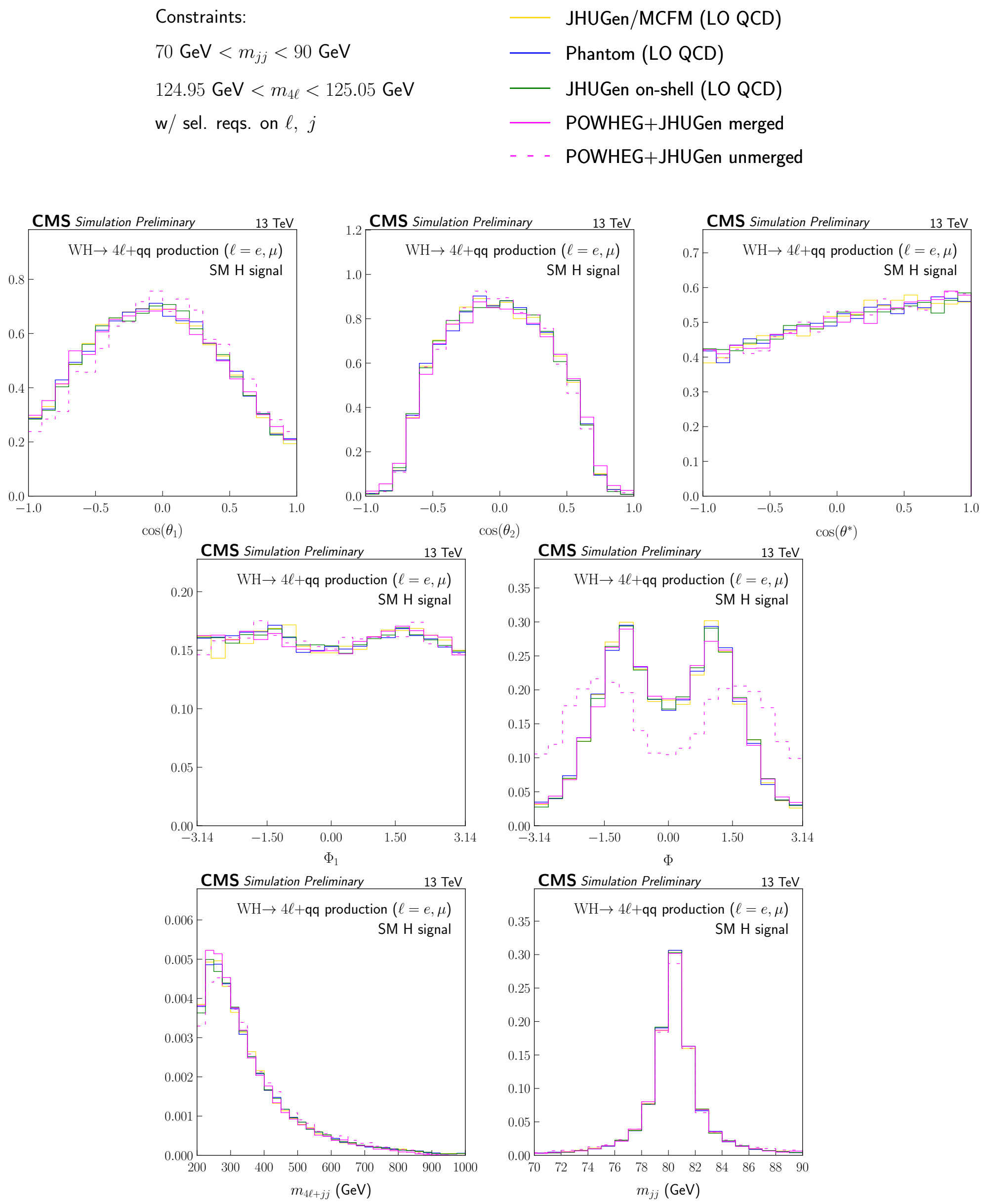

Figure B.3: The distributions of $\cos\theta_1$, $\cos\theta_2$, $\cos\theta^*$, Φ_1, Φ, $m_{4\ell+jj}$ and m_{jj} (from top left to bottom right) for on-shell H boson production through the SM EW signal process. The requirement 70 GeV $< m_{\mathrm{jj}} <$ 90 GeV is applied along with requirements on the initial and final state composition to emphasize the WH-like topology. The different predictions are shown on the legend at the top, and all distributions are normalized to unit area. The unmerged POWHEG scenario uses the two outgoing partons leading in p_{T} to compute the kinematic quantities.

Appendix C

Additional plots for non-resonant background estimation

Figures C.1 and C.2 illustrate the agreement between the predicted and expected non-resonant background component on the $m_{\ell\ell}$ distributions in the signal region for the 2016 and 2017 data sets, respectively, and also compare the simulation to the prediction from observed $e\mu$ events. Likewise, Figs. C.4 and C.5 compare the $p_{\mathrm{T}}^{\mathrm{miss}}$ distributions in the signal region for the 2016 and 2017 data sets, respectively, and Figs. C.6 and C.7 correspond to the $m_{\mathrm{T}}^{\mathrm{ZZ}}$ distributions over these different data periods. The distributions of various kinematic discriminants defined in Eq. (5.7) and used in the $N_j \geq 2$ category are also compared in Fig. C.8 for the SM $\mathcal{D}_{\mathrm{2jet}}^{\mathrm{VBF}}$ discriminant used in the analysis for the 2016, 2017, 2018 data periods separately. Figures C.9 and C.10 likewise show comparisons for the remaining three different BSM discriminants for the 2016 and 2017 data sets, respectively. The distributions of N_j itself are shown in Fig. C.11.

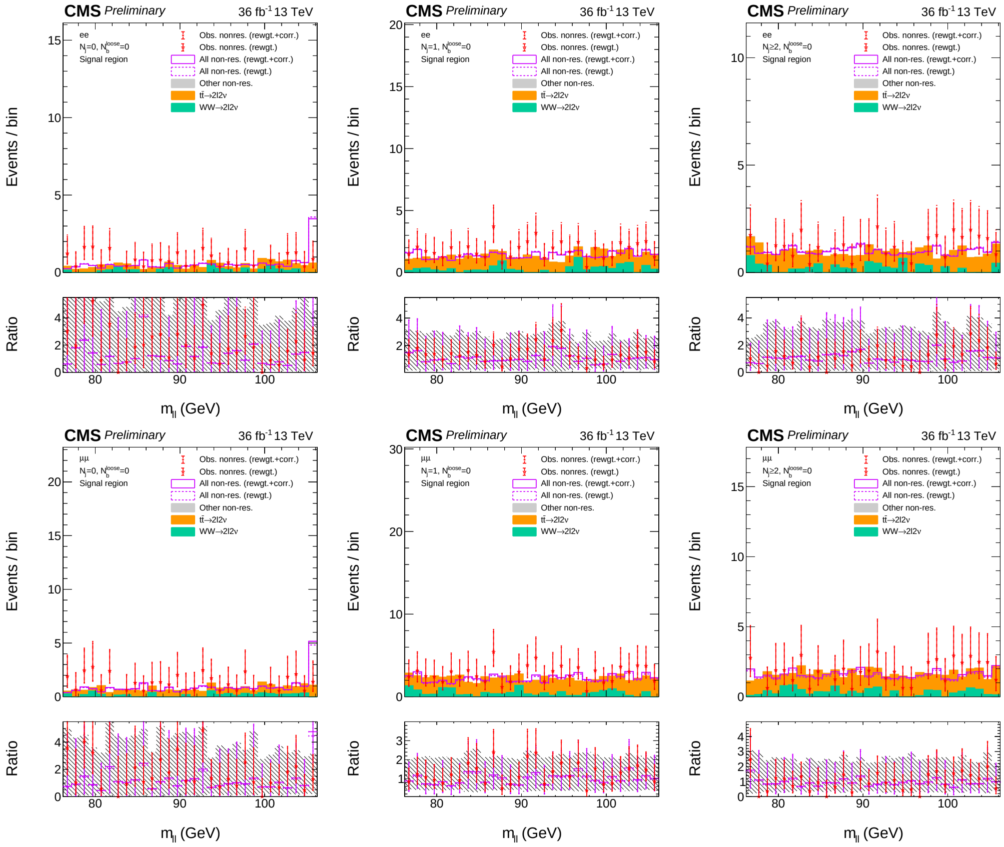

Figure C.1: The $m_{\ell\ell}$ distributions (e e on the top, $\mu\,\mu$ on the bottom panels) in the signal region are shown for the 2016 data set in each N_j category ($= 0$, $= 1$, and ≥ 2 from left to right). The red points with error bars show the prediction from the observed data, either with $f^{\ell\ell}_{corr}$ factors applied (filled stars with solid error bars) or without (hollow stars with dashed error bars). The predictions of each simulated non-resonant background component are shown in histograms filled with different colors, stacked on top of each other. The hollow, solid and dashed purple histograms show the prediction from simulated $e\mu$ events, reweighted according to 5.8 with the $f^{\ell\ell}_{corr}$ factor included or omitted, respectively. The hollow, solid histograms agree with the prediction of the sum of light green (WW continuum production), orange (WW production through the tt̄ process), and gray (other small nonresonant contributions) histograms within the statistical uncertainties of the simulation. The statistical uncertainties on the expected and predicted distributions from simulation are not shown in the top event distribution panels to avoid cluttering, but they are shown in the bottom ratio panels, which are shown relative to the stacked histograms.

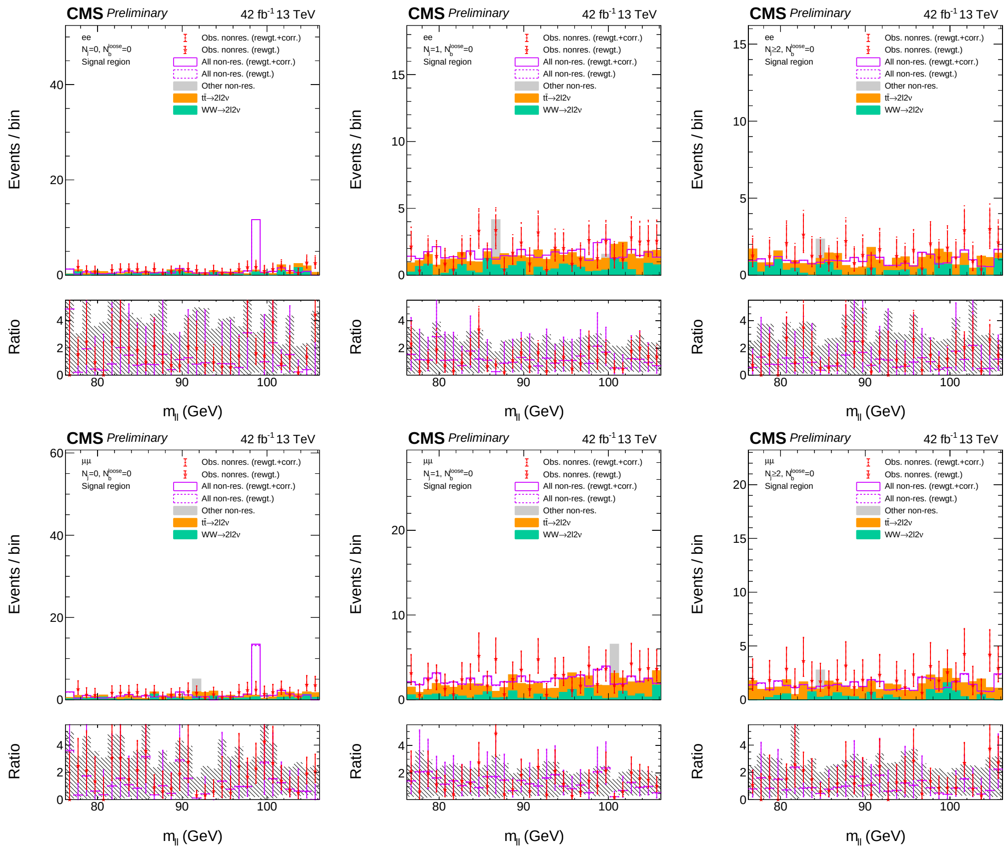

Figure C.2: The $m_{\ell\ell}$ distributions (e e on the top, μ μ on the bottom panels) in the signal region are shown for the 2017 data set in each N_j category ($=0$, $=1$, and ≥ 2 from left to right). The red points with error bars show the prediction from the observed data, either with $f^{\ell\ell}_{corr}$ factors applied (filled stars with solid error bars) or without (hollow stars with dashed error bars). The predictions of each simulated non-resonant background component are shown in histograms filled with different colors, stacked on top of each other. The hollow, solid histograms agree with the prediction of the sum of light green (WW continuum production), orange (WW production through the t$\bar{\text{t}}$ process), and gray (other small nonresonant contributions) histograms within the statistical uncertainties of the simulation. The statistical uncertainties on the expected and predicted distributions from simulation are not shown in the top event distribution panels to avoid cluttering, but they are shown in the bottom ratio panels, which are shown relative to the stacked histograms.

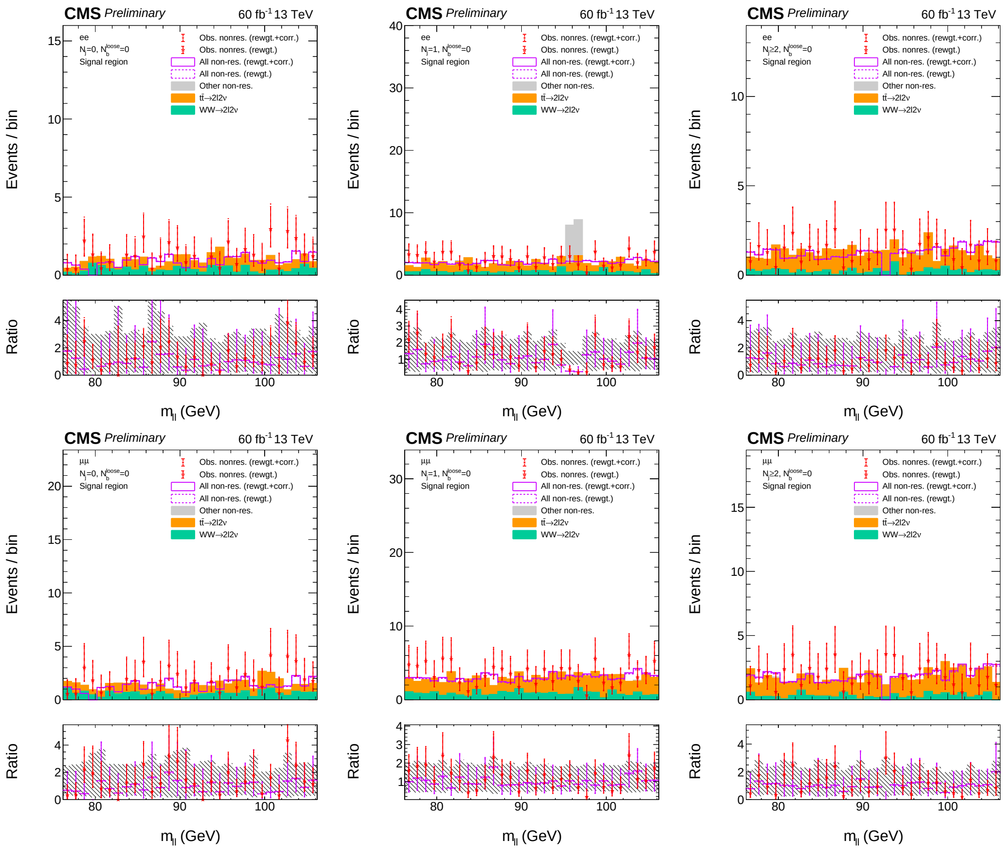

Figure C.3: The $m_{\ell\ell}$ distributions (e e on the top, $\mu\,\mu$ on the bottom panels) in the signal region are shown for the 2018 data set in each N_j category ($=0$, $=1$, and ≥ 2 from left to right). The red points with error bars show the prediction from the observed data, either with $f^{\ell\ell}_{corr}$ factors applied (filled stars with solid error bars) or without (hollow stars with dashed error bars). The predictions of each simulated non-resonant background component are shown in histograms filled with different colors, stacked on top of each other. The hollow, solid histograms agree with the prediction of the sum of light green (WW continuum production), orange (WW production through the $t\bar{t}$ process), and gray (other small nonresonant contributions) histograms within the statistical uncertainties of the simulation. The statistical uncertainties on the expected and predicted distributions from simulation are not shown in the top event distribution panels to avoid cluttering, but they are shown in the bottom ratio panels, which are shown relative to the stacked histograms.

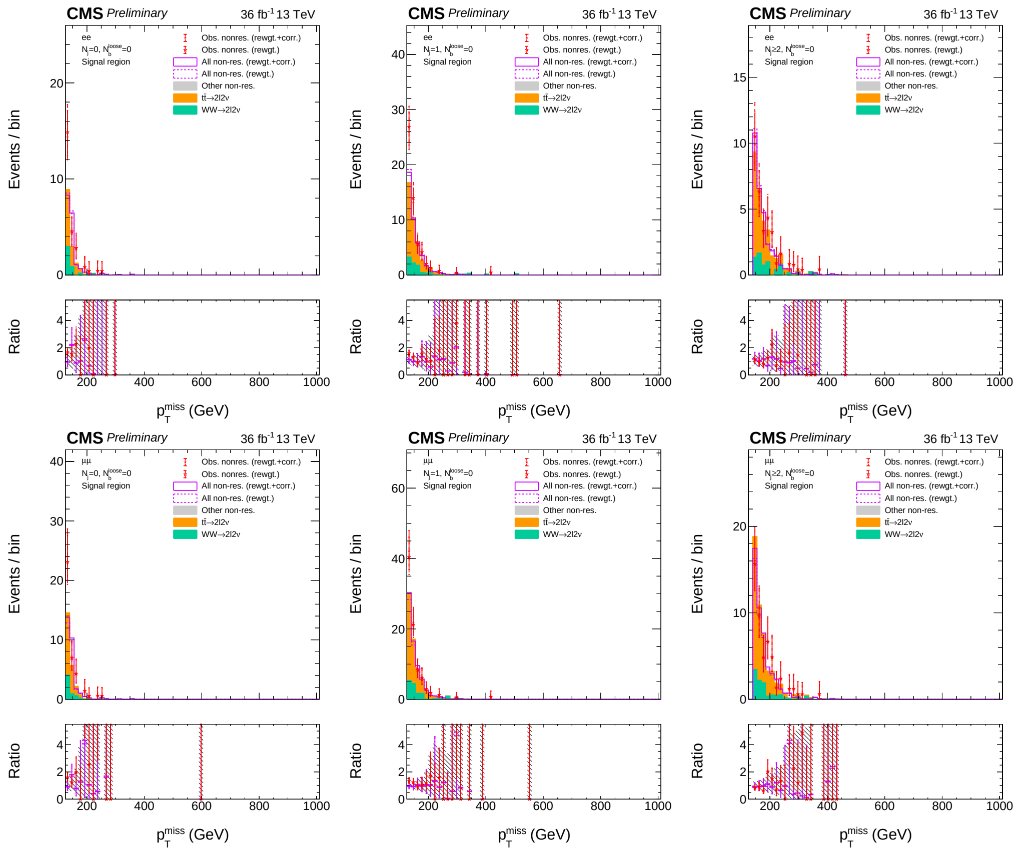

Figure C.4: The $p_{\mathrm{T}}^{\mathrm{miss}}$ distributions (e e on the top, $\mu\,\mu$ on the bottom panels) in the signal region are shown for the 2016 data set in each N_j category ($= 0$, $= 1$, and ≥ 2 from left to right). The red points with error bars show the prediction from the observed data, either with $f^{\ell\ell}_{corr}$ factors applied (filled stars with solid error bars) or without (hollow stars with dashed error bars). The predictions of each simulated non-resonant background component are shown in histograms filled with different colors, stacked on top of each other. The hollow, solid and dashed purple histograms show the prediction from simulated $e\mu$ events, reweighted according to 5.8 with the $f^{\ell\ell}_{corr}$ factor included or omitted, respectively. The hollow, solid histograms agree with the prediction of the sum of light green (WW continuum production), orange (WW production through the tt̄ process), and gray (other small nonresonant contributions) histograms within the statistical uncertainties of the simulation. The statistical uncertainties on the expected and predicted distributions from simulation are not shown in the top event distribution panels to avoid cluttering, but they are shown in the bottom ratio panels, which are shown relative to the stacked histograms.

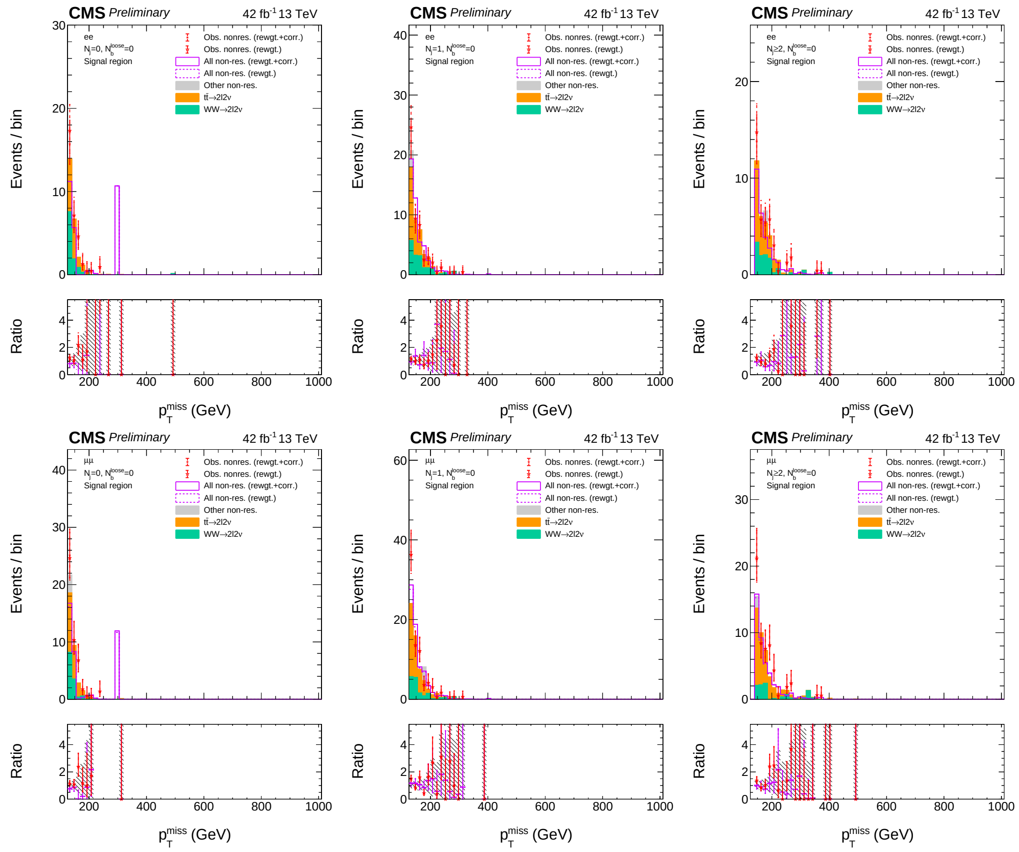

Figure C.5: The $p_{\mathrm{T}}^{\mathrm{miss}}$ distributions (e e on the top, $\mu\,\mu$ on the bottom panels) in the signal region are shown for the 2017 data set in each N_j category ($= 0$, $= 1$, and ≥ 2 from left to right). The red points with error bars show the prediction from the observed data, either with $f^{\ell\ell}_{corr}$ factors applied (filled stars with solid error bars) or without (hollow stars with dashed error bars). The predictions of each simulated non-resonant background component are shown in histograms filled with different colors, stacked on top of each other. The hollow, solid histograms agree with the prediction of the sum of light green (WW continuum production), orange (WW production through the $\mathrm{t\bar{t}}$ process), and gray (other small nonresonant contributions) histograms within the statistical uncertainties of the simulation. The statistical uncertainties on the expected and predicted distributions from simulation are not shown in the top event distribution panels to avoid cluttering, but they are shown in the bottom ratio panels, which are shown relative to the stacked histograms.

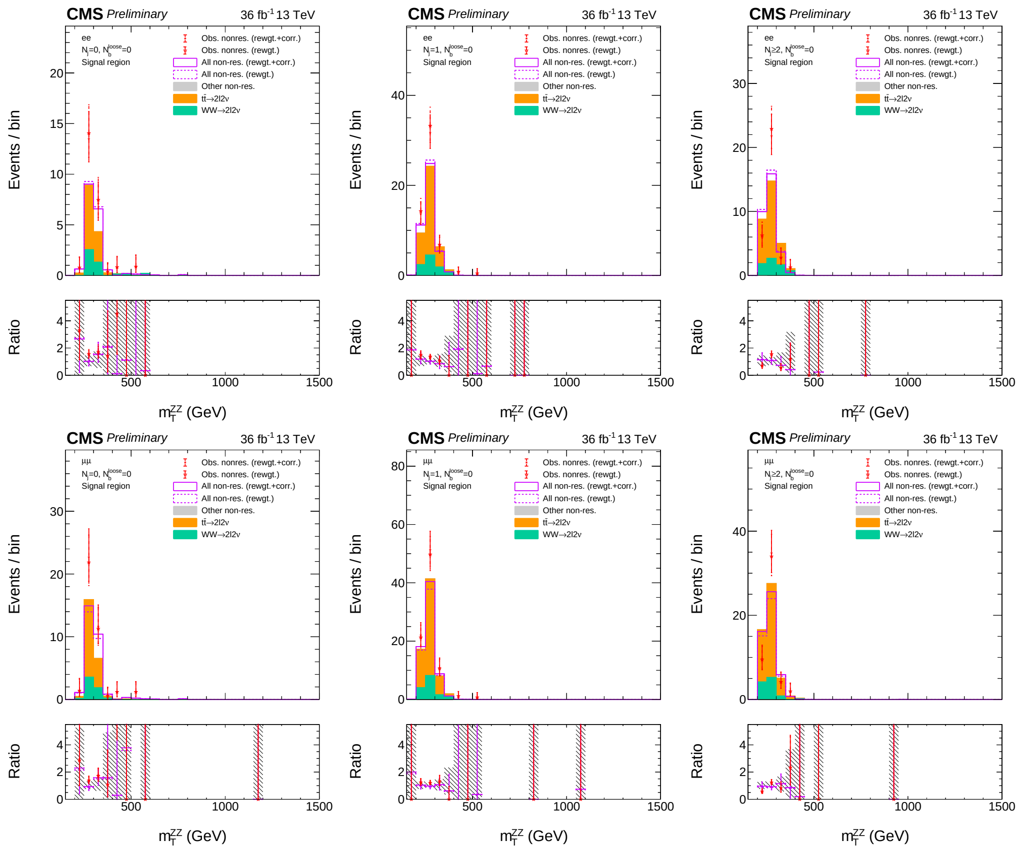

Figure C.6: The $m_{\mathrm{T}}^{\mathrm{ZZ}}$ distributions (e e on the top, $\mu\,\mu$ on the bottom panels) in the signal region are shown for the 2016 data set in each N_j category ($= 0$, $= 1$, and ≥ 2 from left to right). The different components and style features shown match those in Fig. C.1.

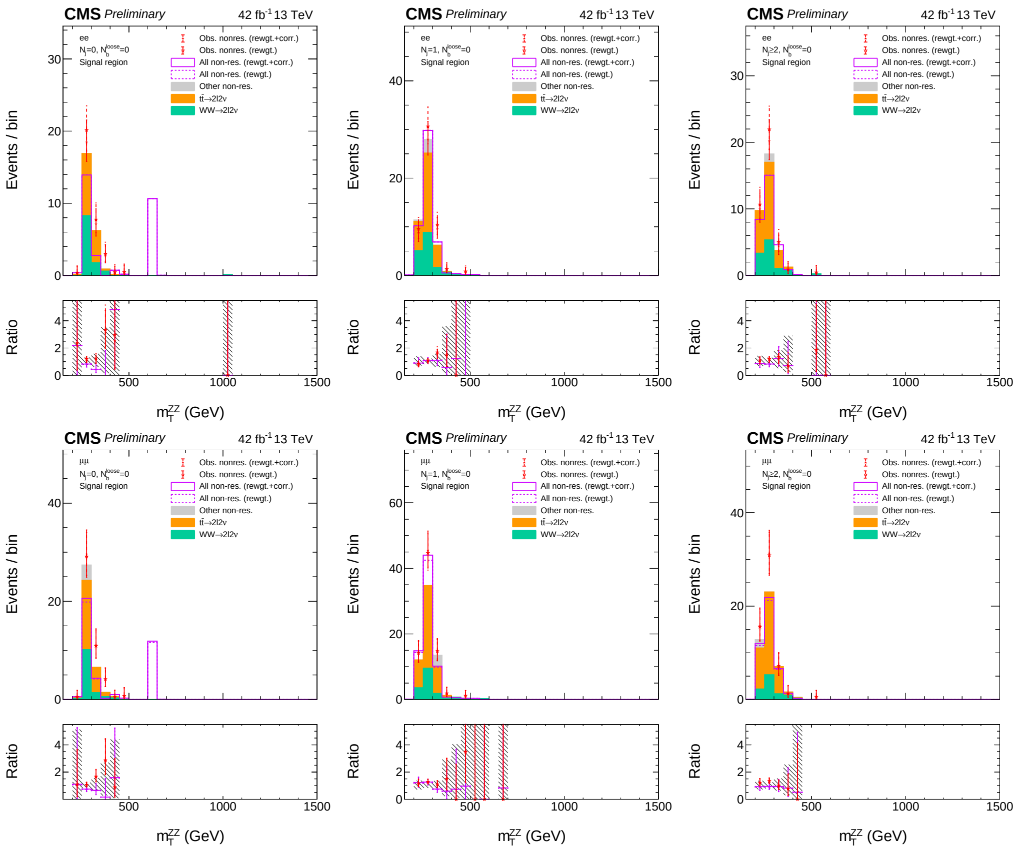

Figure C.7: The $m_{\mathrm{T}}^{\mathrm{ZZ}}$ distributions (e e on the top, μ μ on the bottom panels) in the signal region are shown for the 2017 data set in each N_j category ($= 0$, $= 1$, and ≥ 2 from left to right). The different components and style features shown match those in Fig. C.2.

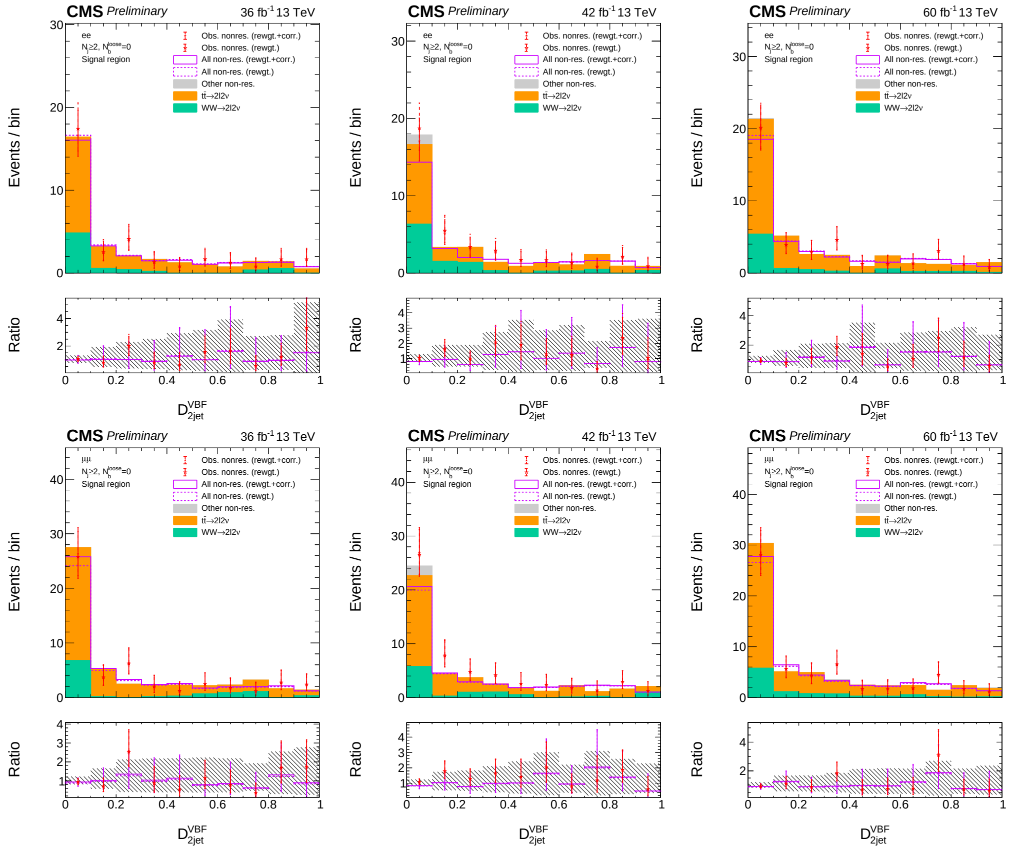

Figure C.8: The distributions of SM $\mathcal{D}_{\text{2jet}}^{\text{VBF}}$ kinematic discriminants in the signal region are shown for the 2016, 2017, and 2018 data periods from left to right in the $N_j \geq 2$ category. The decay channel e e is shown on the top panels, and $\mu\,\mu$ is shown on the bottom ones. The components shown and the style conventions shown are the same as the corresponding figures for the $m_{\ell\ell}$ distributions.

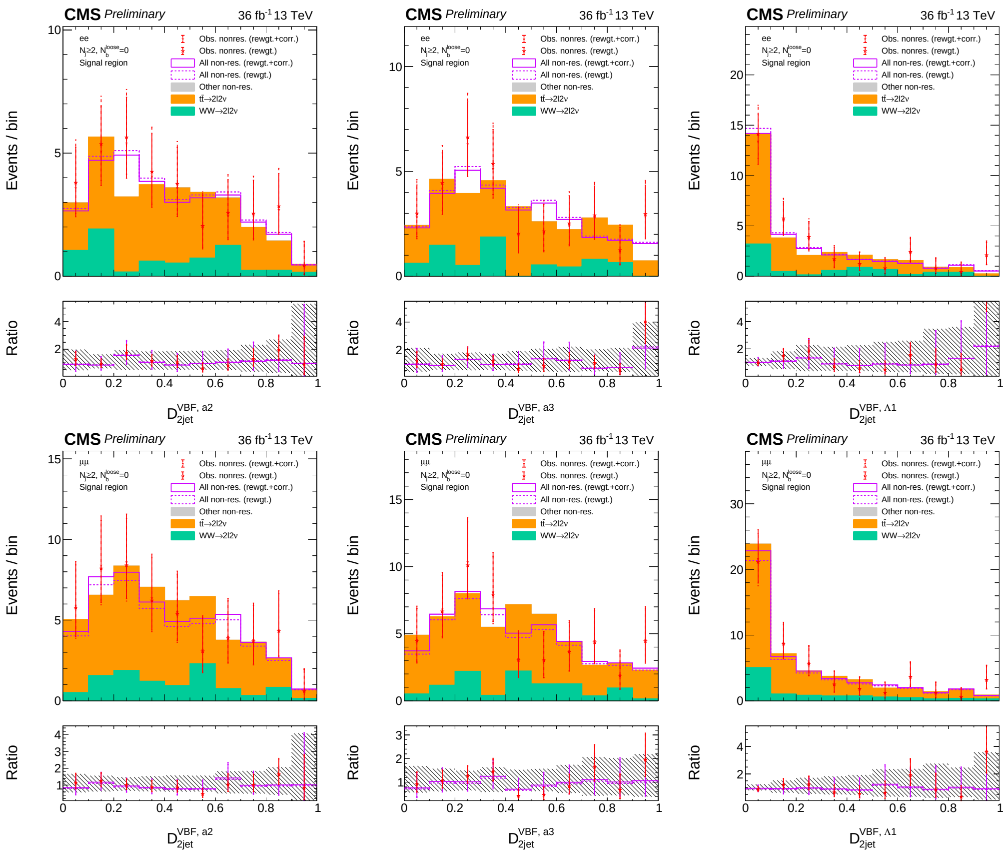

Figure C.9: The distributions of different BSM $\mathcal{D}_{\text{2jet}}^{\text{VBF}}$ kinematic discriminants in the signal region are shown for the 2016 data set in the $N_j \geq 2$ category. The decay channel e e is shown on the top panels, and $\mu\,\mu$ is shown on the bottom ones. From left to right the $\mathcal{D}_{\text{2jet}}^{\text{VBF}}$ discriminants shown are for the $f_{a2} = 1$, $f_{a3} = 1$ and $f_{\Lambda 1} = 1$ VBF hypotheses. The components shown and the style conventions shown are the same as the corresponding figures for the $m_{\ell\ell}$ distributions.

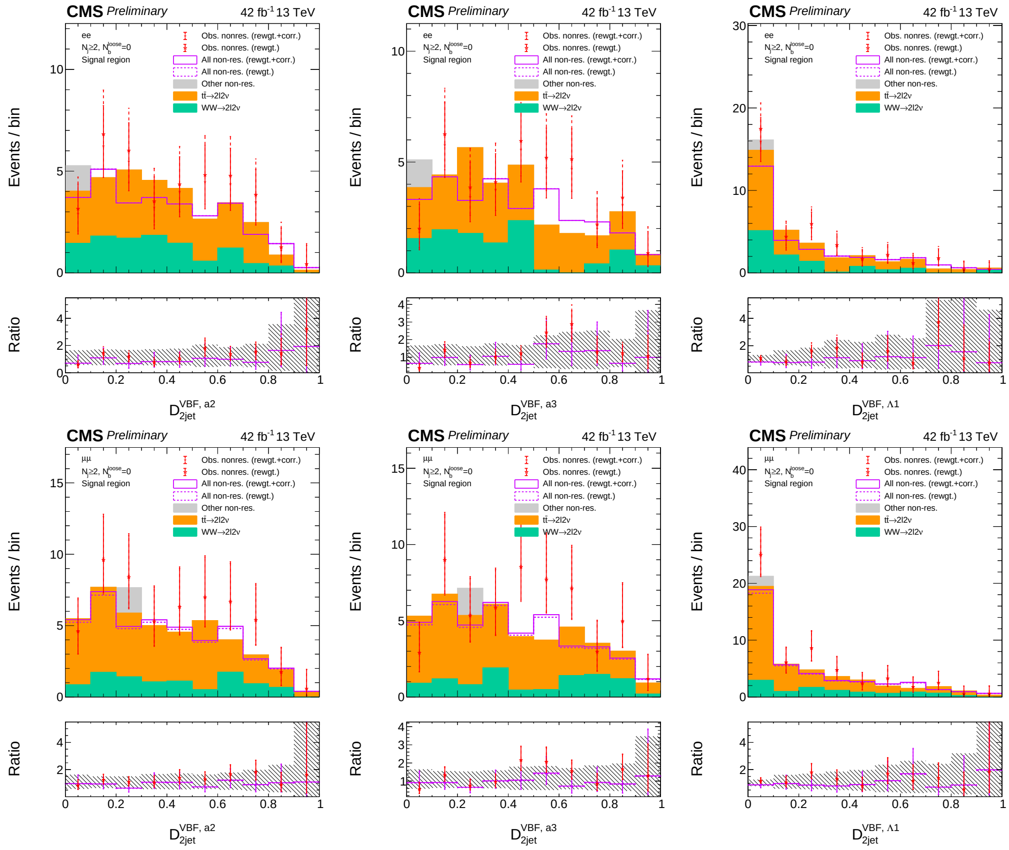

Figure C.10: The distributions of different BSM $\mathcal{D}_{\text{2jet}}^{\text{VBF}}$ kinematic discriminants in the signal region are shown for the 2017 data set in the $N_j \geq 2$ category. The decay channel e e is shown on the top panels, and $\mu\,\mu$ is shown on the bottom ones. From left to right the $\mathcal{D}_{\text{2jet}}^{\text{VBF}}$ discriminants shown are for the $f_{a2} = 1$, $f_{a3} = 1$ and $f_{\Lambda 1} = 1$ VBF hypotheses. The components shown and the style conventions shown are the same as the corresponding figures for the $m_{\ell\ell}$ distributions.

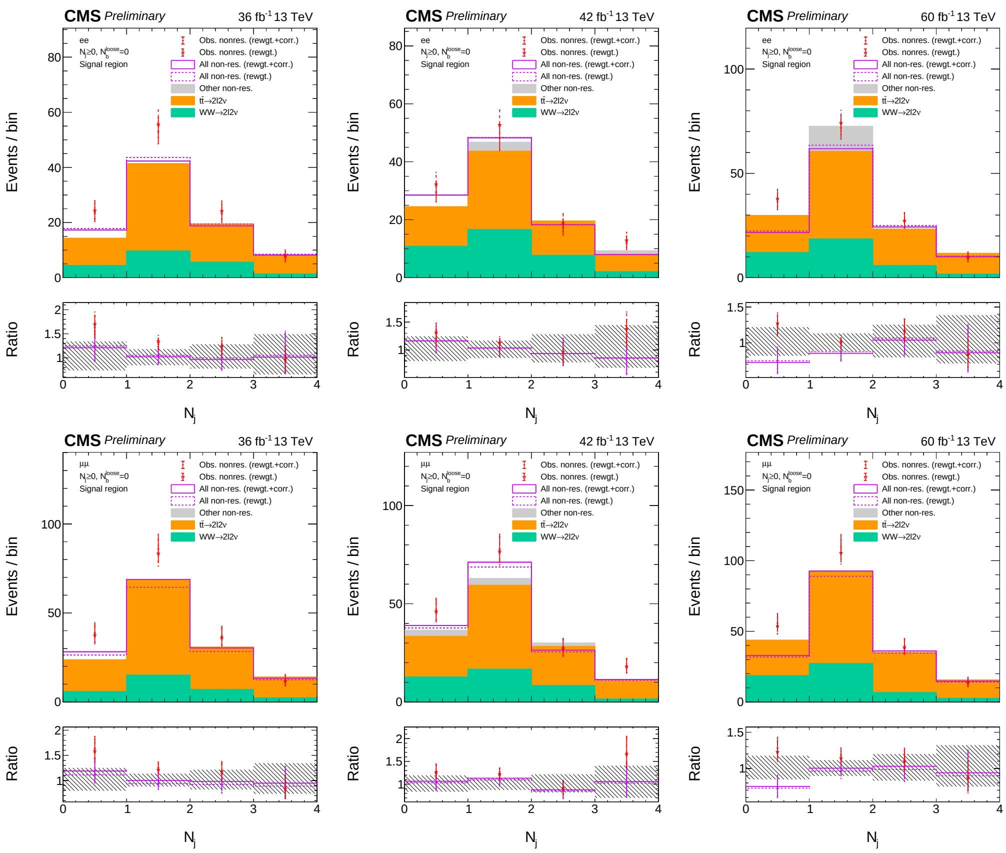

Figure C.11: The distributions of N_j are shown from the signal region for the data periods 2016–2018 from left to right. The top and bottom panels display e e and μ μ channels, respectively. The last bins contain the overflow.

Appendix D

List of various samples used in the thesis

Table D.1: List of data samples used for EOY and UL datasets for MET filter studies.

EOY datasets	UL datasets
MET/Run2016*-17Jul2018*	MET/Run2016*-21Feb2020*UL2016_HIPM-v1
	MET/Run2016*-21Feb2020_UL2016-v1
MET/Run2017*-31Mar2018-v1	MET/Run2017*-UL2017_MiniAODv2-v1
MET/Run2018*-17Sep2018-v1	MET/Run2018*-UL2018_MiniAODv2-v1

Table D.2: MC production campaigns for EOY and UL datasets used for MET filter studies.

Dataset	Campaign
2016 EOY	Summer16MiniAODv3
2016 UL	Summer20UL16MiniAOD-106X, Summer20UL16MiniAODAPV
2017 EOY	Fall17MiniAODv2
2017 UL	Summer19UL17MiniAODv2, Summer20UL17MiniAOD(v2)
2018 EOY	Autumn18MiniAOD
2018 UL	Summer19UL18MiniAOD, Summer20UL18MiniAOD

Table D.3: List of EOY and UL MC samples as well as their cross section (i.e. σ) used for MET filter studies.

EOY 2016	UL 2016	σ (pb)
QCD_HT500to700_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	QCD_HT500to700_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	32060
QCD_HT700to1000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	6829
QCD_HT1000to1500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	1207
QCD_HT1500to2000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	120
QCD_HT2000toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	25.25
TTToHadronic_TuneCP5_PSweights_13TeV-powheg-pythia8	TTToHadronic_TuneCP5_13TeV-powheg-pythia8	377.96
TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	88.29
TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	365.34
WJetsToLNu_Pt-100To250_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	WJetsToLNu_Pt-100To250_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	689.75
WJetsToLNu_Pt-250To400_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	WJetsToLNu_Pt-250To400_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	24.51
WJetsToLNu_Pt-400To600_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	WJetsToLNu_Pt-400To600_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	3.11
WJetsToLNu_Pt-600ToInf_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	WJetsToLNu_Pt-600ToInf_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	0.468
ZJetsToNuNu_HT-100To200_13TeV-madgraph	ZJetsToNuNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	93.35
ZJetsToNuNu_HT-200To400_13TeV-madgraph	ZJetsToNuNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	25.85
ZJetsToNuNu_HT-400To600_13TeV-madgraph	ZJetsToNuNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	3.584
ZJetsToNuNu_HT-600To800_13TeV-madgraph	ZJetsToNuNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8	0.853
ZJetsToNuNu_HT-800To1200_13TeV-madgraph	ZJetsToNuNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8	0.3934
ZJetsToNuNu_HT-1200To2500_13TeV-madgraph	ZJetsToNuNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8	0.0954
ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph	ZJetsToNuNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.0023
EOY 2017	UL 2017	σ (pb)
QCD_HT500to700_TuneCP5_13TeV-madgraph-pythia8	QCD_HT500to700_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	29980
QCD_HT700to1000_TuneCP5_13TeV-madgraph-pythia8	QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	6334
QCD_HT1000to1500_TuneCP5_13TeV-madgraph-pythia8	QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	1088
QCD_HT1500to2000_TuneCP5_13TeV-madgraph-pythia8	QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	99.11
QCD_HT2000toInf_TuneCP5_13TeV-madgraph-pythia8	QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	20.23
TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	88.29
TTToHadronic_TuneCP5_PSweights_13TeV-powheg-pythia8	TTToHadronic_TuneCP5_13TeV-powheg-pythia8	377.96
TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	365.34
WJetsToLNu_Pt-100To250_TuneCP5_13TeV-amcatnloFXFX-pythia8	WJetsToLNu_Pt-100To250_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	689.750
WJetsToLNu_Pt-250To400_TuneCP5_13TeV-amcatnloFXFX-pythia8	WJetsToLNu_Pt-250To400_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	24.507
WJetsToLNu_Pt-400To600_TuneCP5_13TeV-amcatnloFXFX-pythia8	WJetsToLNu_Pt-400To600_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	3.110
WJetsToLNu_Pt-600ToInf_TuneCP5_13TeV-amcatnloFXFX-pythia8	WJetsToLNu_Pt-600ToInf_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	0.468
ZJetsToNuNu_HT-100To200_13TeV-madgraph	ZJetsToNuNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	342.338
ZJetsToNuNu_HT-200To400_13TeV-madgraph	ZJetsToNuNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	103.311
ZJetsToNuNu_HT-400To600_13TeV-madgraph	ZJetsToNuNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	14.946
ZJetsToNuNu_HT-600To800_13TeV-madgraph	ZJetsToNuNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8	3.612
ZJetsToNuNu_HT-800To1200_13TeV-madgraph	ZJetsToNuNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8	1.552
ZJetsToNuNu_HT-1200To2500_13TeV-madgraph	ZJetsToNuNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8	0.496
ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph	ZJetsToNuNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.00515
EOY 2018	UL 2018	σ (pb)
QCD_HT500to700_TuneCP5_13TeV-madgraphMLM-pythia8	QCD_HT500to700_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	30140
QCD_HT700to1000_TuneCP5_13TeV-madgraphMLM-pythia8	QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	6310
QCD_HT1000to1500_TuneCP5_13TeV-madgraphMLM-pythia8	QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	1094
QCD_HT1500to2000_TuneCP5_13TeV-madgraphMLM-pythia8	QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	99.38
QCD_HT2000toInf_TuneCP5_13TeV-madgraphMLM-pythia8	QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraphMLM-pythia8	20.2
TTToHadronic_TuneCP5_13TeV-powheg-pythia8	TTToHadronic_TuneCP5_13TeV-powheg-pythia8	377.96
TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	88.29
TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8	365.34
WJetsToLNu_Pt-100To250_TuneCP5_13TeV-amcatnloFXFX-pythia8	WJetsToLNu_Pt-100To250_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	723.784
WJetsToLNu_Pt-250To400_TuneCP5_13TeV-amcatnloFXFX-pythia8	WJetsToLNu_Pt-250To400_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	25.993
WJetsToLNu_Pt-400To600_TuneCP5_13TeV-amcatnloFXFX-pythia8	WJetsToLNu_Pt-400To600_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	3.348
WJetsToLNu_Pt-600ToInf_TuneCP5_13TeV-amcatnloFXFX-pythia8	WJetsToLNu_Pt-600ToInf_MatchEWPDG20_TuneCP5_13TeV-amcatnloFXFX-pythia8	0.515
ZJetsToNuNu_HT-100To200_13TeV-madgraph	ZJetsToNuNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	345.12
ZJetsToNuNu_HT-200To400_13TeV-madgraph	ZJetsToNuNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	102.72
ZJetsToNuNu_HT-400To600_13TeV-madgraph	ZJetsToNuNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	14.63
ZJetsToNuNu_HT-600To800_13TeV-madgraph	ZJetsToNuNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8	3.61
ZJetsToNuNu_HT-800To1200_13TeV-madgraph	ZJetsToNuNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8	1.566
ZJetsToNuNu_HT-1200To2500_13TeV-madgraph	ZJetsToNuNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8	0.344
ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph	ZJetsToNuNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.00767

Table D.4: The miniAOD tags of the simulation samples used in the analysis of dilepton signal or control regions for the 2016 data set. The cross sections quoted include any branching ratio corrections when necessary. The nanoAOD tags used derive from these tags based on v7 of the nanoAOD data format. The miniAOD version tag corresponds to RunIISummer16MiniAODv3 - PUMoriond17_94X_mcRun2 _asymptotic_v3* in these samples.

Sample tag	Purpose	σ (pb)
DYJetsToLL_M-10to50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	DY / Instr. $p_{\mathrm{T}}^{\mathrm{miss}}$	18780
DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8		5931
ST_t-channel_top_4f_inclusiveDecays_13TeV*pythia8_TuneCUETP8M1	Single top / NRB	123.308811
ST_t-channel_antitop*_13TeV-powhegV2*pythia8_TuneCUETP8M1		74.41085
ST_s-channel_4f_leptonDecays_13TeV-amcatnlo-pythia8_TuneCUETP8M1		3.365
ST_tW_top_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1		38.09
ST_tW_antitop_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1		38.06
TTTo2L2Nu_TuneCUETP8M2_ttHtranche3_13TeV-powheg-pythia8	$\mathrm{t\bar{t}}$ / NRB	87.3348
TTZToQQ_TuneCUETP8M1_13TeV-amcatnlo-pythia8		0.5297
TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8		0.2529
TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8		0.2005
WWToLNuQQ_13TeV-powheg	W decays / NRB	43.53
WWTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8		45.67
WWTo2L2Nu_13TeV-powheg		10.48
WWTo2L2Nu_13TeV-powheg-CUETP8M1Down		10.48
WWTo2L2Nu_13TeV-powheg-CUETP8M1Up		10.48
WWW_4F_TuneCUETP8M1_13TeV-amcatnlo-pythia8		0.208569
tZq_ll_4f_13TeV-amcatnlo-pythia8	Irreducible BKG	0.0758
ZZTo2L2Nu_13TeV_powheg_pythia8*		0.5644
ZZTo4L_13TeV_powheg_pythia8*		1.256
ZZTo2Q2Nu_13TeV_amcatnloFXFX_madspin_pythia8		4.041
ZZTo2L2Q_13TeV_powheg_pythia8		3.521
WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8		10.74
WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8		5.605
WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8		3.067
WZTo3LNu_mllmin01_13TeV-powheg-pythia8* (*)		35.1526
WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8		4.42965
WZTo3LNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8		4.699
WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8		0.1651
WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8		0.055646
ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8		0.01398

(*) The reported cross section is scaled by 0.6 with respect to the POWHEG value.

Table D.5: The miniAOD tags of the simulation samples used in the analysis of dilepton signal or control regions for the 2017 data set. The cross sections quoted include any branching ratio corrections when necessary. The nanoAOD tags used derive from these tags based on v7 of the nanoAOD data format. The miniAOD version tag corresponds to RunIIFall17MiniAODv2-PU2017_12Apr2018_*94X_mc2017 _realistic_v14* in these samples, where almost all are corrected for the pileup premixing software bug.

Sample tag	Purpose	σ (pb)
DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8	DY / Instr. $p_{\mathrm{T}}^{\mathrm{miss}}$	15800
DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8		6225.4
ST_t-channel_top_5f_TuneCP5_13TeV-powheg-pythia8	Single top / NRB	138
ST_t-channel_antitop_5f_TuneCP5_PSweights_13TeV-powheg-pythia8		82.5
ST_s-channel_top_leptonDecays_13TeV-PSweights_powheg-pythia		5.756
ST_s-channel_antitop_leptonDecays_13TeV-PSweights_powheg-pythia		3.58
ST_s-channel_4f_leptonDecays_TuneCP5_PSweights_13TeV-amcatnlo-pythia8		3.74
ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8		20.248
ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8		18.502
TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8	$\mathrm{t\bar{t}}$ / NRB	87.3348
TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8		0.2432
TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8		0.2181
WWToLNuQQ_NNPDF31_TuneCP5_*13TeV-powheg-pythia8	W decays / NRB	45.99
WWTo2L2Nu_NNPDF31_TuneCP5_*13TeV-powheg-pythia8		11.08
WWTo2L2Nu_NNPDF31_TuneCP5Down_*13TeV-powheg-pythia8		11.08
WWTo2L2Nu_NNPDF31_TuneCP5Up_*13TeV-powheg-pythia8		11.08
WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8		0.2154
tZq_ll_4f_ckm_NLO_TuneCP5_PSweights_13TeV-amcatnlo-pythia8	Irreducible BKG	0.0758
ZZTo2L2Nu_13TeV_powheg_pythia8		0.6008
ZZTo2L2Nu_mZMin-18_TuneCP5_13TeV-powheg-pythia8 (*)		0.64819
ZZTo4L_13TeV_powheg_pythia8		1.325
ZZTo2Q2Nu_TuneCP5_13TeV_amcatnloFXFX_madspin_pythia8		4.325
ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8		3.691
WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8		11.74
WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8		6.284
WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8_v2		3.325
WZTo3LNu_mllmin01_NNPDF31_TuneCP5_13TeV_powheg_pythia8 (**)		37.3008
WZTo3LNu_13TeV-powheg-pythia8		4.658
WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8		5.087
WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8		0.1675
WZZ_TuneCP5_13TeV-amcatnlo-pythia8		0.0571
ZZZ_TuneCP5_13TeV-amcatnlo-pythia8		0.01473

(*) The reported cross section is scaled by 1.078882 with respect to the POWHEG value.
(**) The reported cross section is scaled by 0.6 with respect to the POWHEG value.

Table D.6: The miniAOD tags of the simulation samples used in the analysis of dilepton signal or control regions for the 2018 data set. The cross sections quoted include any branching ratio corrections when necessary. The nanoAOD tags used derive from these tags based on v7 of the nanoAOD data format. The miniAOD version tag corresponds to RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15* in these samples.

Sample tag	Purpose	σ (pb)
DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8	DY / Instr. $p_{\mathrm{T}}^{\mathrm{miss}}$	15800
DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8		6225.4
ST_t-channel_top_5f_TuneCP5_13TeV-powheg-pythia8	Single top / NRB	138
ST_t-channel_antitop_5f_TuneCP5_13TeV-powheg-pythia8		82.5
ST_s-channel_top_leptonDecays_13TeV-PSweights_powheg-pythia		5.756
ST_s-channel_antitop_leptonDecays_13TeV-PSweights_powheg-pythia		3.58
ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8		20.248
ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8		18.502
TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8	$\mathrm{t\bar{t}}$	87.3348
TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8		0.2432
TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8		0.2181
WWToLNuQQ_NNPDF31_TuneCP5_13TeV-powheg-pythia8	W decays / NRB	45.99
WWTo2L2Nu_NNPDF31_TuneCP5_13TeV-powheg-pythia8		11.08
WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8		0.2154
tZq_ll_4f_ckm_NLO_TuneCP5_13TeV-madgraph-pythia8	Irreducible BKG	0.0758
ZZTo2L2Nu_TuneCP5_13TeV_powheg_pythia8		0.6008
ZZTo2L2Nu_mZMin-18_TuneCP5_13TeV-powheg-pythia8 (*)		0.64819
ZZTo4L_TuneCP5_13TeV_powheg_pythia8		1.325
ZZTo2Q2Nu_TuneCP5_13TeV_amcatnloFXFX_madspin_pythia8		4.325
ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8		3.691
WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8		11.74
WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8		6.284
WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8		3.325
WZTo3LNu_mllmin01_NNPDF31_TuneCP5_13TeV_powheg_pythia8 (**)		37.3008
WZTo3LNu_TuneCP5_13TeV-powheg-pythia8		4.658
WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8		5.087
WWZ_TuneCP5_13TeV-amcatnlo-pythia8		0.1675
WZZ_TuneCP5_13TeV-amcatnlo-pythia8		0.0571
ZZZ_TuneCP5_13TeV-amcatnlo-pythia8		0.01473

(*) The reported cross section is scaled by 1.078882 with respect to the POWHEG value.
(**) The reported cross section is scaled by 0.6 with respect to the POWHEG value.

Table D.7: The miniAOD tags of the simulation samples used in the analysis of single photon control region for the 2016 data set. The nanoAOD tags used derive from these tags based on v7 of the nanoAOD data format. The miniAOD version tag corresponds to RunIISummer16MiniAODv3-PUMoriond17_94X_mcRun2_asymptotic_v3* in these samples.

Sample tag	σ (pb)
WJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	60490
GJets_HT-40To100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	20820
GJets_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	9249
GJets_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2305
GJets_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	275.3
GJets_HT-600ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	93.41
QCD_HT50to100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	246300000
QCD_HT100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	27970000
QCD_HT200to300_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1711000
QCD_HT300to500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	347800
QCD_HT500to700_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	32100
QCD_HT700to1000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	6835
QCD_HT1000to1500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1209
QCD_HT1500to2000_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	119.8
QCD_HT2000toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	25.36
TTJets_TuneCUETP8M2T4_13TeV-amcatnloFXFX-pythia8	747.8
TGJets_leptonDecays_13TeV_amcatnlo_madspin_pythia8	2.97
TTGJets_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8	3.761
WGToLNuG_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	489
WGToLNuG_01J_5f_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	178.2
WGToLNuG_01J_5f_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	178.2
WplusG_WtoLNu_TuneCUETP8M1_13TeV-powheg-pythia8	33420
WminusG_WtoLNu_TuneCUETP8M1_13TeV-powheg-pythia8	24780
WZG_TuneCUETP8M1_13TeV-amcatnlo-pythia8	0.04123
ZJetsToNuNu_HT-100To200_13TeV-madgraph	344.83
ZJetsToNuNu_HT-200To400_13TeV-madgraph	95.534
ZJetsToNuNu_HT-400To600_13TeV-madgraph	13.198
ZJetsToNuNu_HT-600To800_13TeV-madgraph	3.221
ZJetsToNuNu_HT-800To1200_13TeV-madgraph	1.474
ZJetsToNuNu_HT-1200To2500_13TeV-madgraph	0.3586
ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph	0.008203
ZGTo2NuG_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	27.99
ZGTo2NuG_PtG-130_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	0.278
ZNuNuGJets_MonoPhoton_PtG-40to130_TuneCUETP8M1_13TeV-madgraph	2.786
ZGTo2LG_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	99.1
ZGTo2LG_PtG-130_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8	0.158

Table D.8: The miniAOD tags of the simulation samples used in the analysis of single photon control region for the 2017 data set. The nanoAOD tags used derive from these tags based on v7 of the nanoAOD data format. The miniAOD version tag corresponds to RunIIFall17MiniAODv2-PU2017_12Apr2018_*94X_mc2017_realistic_v14* in these samples, where almost all are corrected for the pileup premixing software bug.

Sample tag	σ (pb)
WJetsToLNu_0J_TuneCP5_13TeV-amcatnloFXFX-pythia8	49141
WJetsToLNu_1J_TuneCP5_13TeV-amcatnloFXFX-pythia8	8045.1
WJetsToLNu_2J_TuneCP5_13TeV-amcatnloFXFX-pythia8	3159.9
GJets_HT-40To100_TuneCP5_13TeV-madgraphMLM-pythia8	20790
GJets_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	9238
GJets_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	2305
GJets_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	274.4
GJets_HT-600ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	93.46
QCD_HT50to100_TuneCP5_13TeV-madgraphMLM-pythia8	185800000
QCD_HT100to200_TuneCP5_13TeV-madgraph-pythia8	23660000
QCD_HT200to300_TuneCP5_13TeV-madgraph-pythia8	1559000
QCD_HT300to500_TuneCP5_13TeV-madgraph-pythia8	323300
QCD_HT500to700_TuneCP5_13TeV-madgraph-pythia8	30000
QCD_HT700to1000_TuneCP5_13TeV-madgraph-pythia8	6330
QCD_HT1000to1500_TuneCP5_13TeV-madgraph-pythia8	1098
QCD_HT1500to2000_TuneCP5_13TeV-madgraph-pythia8	99.8
QCD_HT2000toInf_TuneCP5_13TeV-madgraph-pythia8	20.35
TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8	748.8
TGJets_leptonDecays_TuneCP5_PSweights_13TeV-amcatnlo-pythia8	2.872
TTGJets_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8	3.746
WGToLNuG_TuneCP5_13TeV-madgraphMLM-pythia8	444.6
WGToLNuG_01J_5f_*TuneCP5_13TeV-amcatnloFXFX-pythia8	191.4
WplusG_WtoLNu_TuneCP5_13TeV-powheg-pythia8	34220
WminusG_WtoLNu_TuneCP5_13TeV-powheg-pythia8	25350
WZG_TuneCP5_13TeV-amcatnlo-pythia8	0.04123
ZJetsToNuNu_HT-100To200_13TeV-madgraph	303.9
ZJetsToNuNu_HT-200To400_13TeV-madgraph	91.03
ZJetsToNuNu_HT-400To600_13TeV-madgraph	13.07
ZJetsToNuNu_HT-600To800_13TeV-madgraph	3.26
ZJetsToNuNu_HT-800To1200_13TeV-madgraph	1.509
ZJetsToNuNu_HT-1200To2500_13TeV-madgraph	0.3401
ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph	0.00527
ZGTo2NuG_TuneCP5_13TeV-amcatnloFXFX-pythia8	30.05
ZGTo2NuG_PtG-130_TuneCP5_13TeV-amcatnloFXFX-pythia8	0.2828
ZNuNuGJets_MonoPhoton_PtG-40to130_TuneCP5_13TeV-madgraph-pythia8	3.003
ZNuNuGJets_MonoPhoton_PtG-130_TuneCP5_13TeV-madgraph-pythia8	0.1926
ZGTo2LG_PtG-130_TuneCP5_13TeV-amcatnloFXFX-pythia8	0.1595
ZLLGJets_MonoPhoton_PtG-40to130_TuneCP5_13TeV-madgraph-pythia8	5.485
ZLLGJets_MonoPhoton_PtG-130_TuneCP5_13TeV-madgraph-pythia8	0.1472

Table D.9: The miniAOD tags of the simulation samples used in the analysis of single photon control region for the 2018 data set. The nanoAOD tags used derive from these tags based on v7 of the nanoAOD data format. The miniAOD version tag corresponds to RunIIAutumn18MiniAOD-102X_upgrade2018_realistic_v15* in these samples.

Sample tag	σ (pb)
WJetsToLNu_0J_TuneCP5_13TeV-amcatnloFXFX-pythia8	49141
WJetsToLNu_1J_TuneCP5_13TeV-amcatnloFXFX-pythia8	8045.1
WJetsToLNu_2J_TuneCP5_13TeV-amcatnloFXFX-pythia8	3159.9
GJets_HT-40To100_TuneCP5_13TeV-madgraphMLM-pythia8	20790
GJets_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8	9238
GJets_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8	2305
GJets_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8	274.4
GJets_HT-600ToInf_TuneCP5_13TeV-madgraphMLM-pythia8	93.46
QCD_HT50to100_TuneCP5_13TeV-madgraphMLM-pythia8	185800000
QCD_HT100to200_TuneCP5_13TeV-madgraphMLM-pythia8	23660000
QCD_HT200to300_TuneCP5_13TeV-madgraphMLM-pythia8	1559000
QCD_HT300to500_TuneCP5_13TeV-madgraphMLM-pythia8	323300
QCD_HT500to700_TuneCP5_13TeV-madgraphMLM-pythia8	30000
QCD_HT700to1000_TuneCP5_13TeV-madgraphMLM-pythia8	6330
QCD_HT1000to1500_TuneCP5_13TeV-madgraphMLM-pythia8	1098
QCD_HT1500to2000_TuneCP5_13TeV-madgraphMLM-pythia8	99.8
QCD_HT2000toInf_TuneCP5_13TeV-madgraphMLM-pythia8	20.35
TTJets_TuneCP5_13TeV-amcatnloFXFX-pythia8	748.8
TGJets_leptonDecays_TuneCP5_13TeV-madgraph-pythia8	2.872
TTGJets_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8	3.746
WGToLNuG_TuneCP5_13TeV-madgraphMLM-pythia8	444.6
WGToLNuG_01J_5f_*TuneCP5_13TeV-amcatnloFXFX-pythia8	191.4
WplusG_WtoLNu_TuneCP5_13TeV-powheg-pythia8	34220
WminusG_WtoLNu_TuneCP5_13TeV-powheg-pythia8	25350
WZG_TuneCP5_13TeV-amcatnlo-pythia8	0.04123
ZJetsToNuNu_HT-100To200_13TeV-madgraph	303.9
ZJetsToNuNu_HT-200To400_13TeV-madgraph	91.03
ZJetsToNuNu_HT-400To600_13TeV-madgraph	13.07
ZJetsToNuNu_HT-600To800_13TeV-madgraph	3.26
ZJetsToNuNu_HT-800To1200_13TeV-madgraph	1.509
ZJetsToNuNu_HT-1200To2500_13TeV-madgraph	0.3401
ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph	0.00527
ZGTo2NuG_TuneCP5_13TeV-amcatnloFXFX-pythia8	30.05
ZGTo2NuG_PtG-130_TuneCP5_13TeV-amcatnloFXFX-pythia8	0.2828
ZNuNuGJets_MonoPhoton_PtG-40to130_TuneCP5_13TeV-madgraph-pythia8	3.003
ZNuNuGJets_MonoPhoton_PtG-130_TuneCP5_13TeV-madgraph-pythia8	0.1926
ZGTo2LG_PtG-130_TuneCP5_13TeV-amcatnloFXFX-pythia8	0.1595
ZLLGJets_MonoPhoton_PtG-40to130_TuneCP5_13TeV-madgraph-pythia8	5.485
ZLLGJets_MonoPhoton_PtG-130_TuneCP5_13TeV-madgraph-pythia8	0.1472