DISS. ETH NO. 30343

Precision measurements of the Higgs boson in the VH($\rightarrow \mathrm{b\bar{b}}$) channel with the CMS detector using LHC Run 2 data.

A thesis submitted to attain the degree of
DOCTOR OF SCIENCES
(Dr. sc. ETH Zurich)

presented by

Krunal Bipin Gedia

M.Sc. Physics, ETH Zurich and École Polytechnique

born on 24.01.1994

accepted on the recommendation of

Prof. Dr. Rainer Wallny
Prof. em. Dr. Christophorus Grab
Prof. Dr. Vuko Brigljević

2024

Abstract

Precision measurements of the Higgs boson predicted in the Standard Model (SM) are discussed in the first part of the thesis. The focus is on the dominant decay mode of the Higgs boson i.e. the decay of the Higgs boson to bottom quarks. The associated production of the Higgs boson with the vector boson decaying leptonically is considered as the leptonic decay mode providing useful trigger information to reduce the overwhelming multi-jet background. Using the full Run 2 data (2016-2018) collected in proton-proton collisions at the Large Hadron Collider (LHC) with the Compact Muon Solenoid (CMS) experiment, the cross section of $\mathrm{VH}(\rightarrow \mathrm{b\bar{b}})$ is presented in terms of mutually exclusive regions of phase space defined in bins of the transverse momentum of the vector boson, known as the simplified template cross sections (STXS) framework. Analysis categories targeting the high transverse momentum of the Higgs boson were studied for possible deviations from the SM. We measured the full Run 2 $\mathrm{VH}(\rightarrow \mathrm{b\bar{b}})$ signal strength of $1.15^{+0.22}_{-0.20}$, leading to the CMS Collaboration's first observation of the $\mathrm{VH}(\rightarrow \mathrm{b\bar{b}})$ process. No statistically significant deviations were observed in any of the STXS bins. In the second part of this thesis, we present a Graph Neural Network (GNN) based approach to obtain the efficiency of b-tagging classifiers. This approach overcomes the several limitations of the traditional approaches which calculates the efficiency using selection cuts.

Résumé

Une mesure précise du boson de Higgs prédit par le Modèle Standard (SM) est discutée dans la première partie de la thèse. L'accent est mis sur le mode de désintégration dominant du boson de Higgs, c'est-à-dire la désintégration du boson de Higgs en quarks bottom. La production associée du boson de Higgs dans laquelle le boson vecteur se désintègre en leptons est considérée car le mode de désintégration leptonique fournit des informations de sélection utiles pour réduire le fond multi-jet écrasant. En utilisant l'intégralité des données du Run 2 (2016-2018) collectées lors de collisions de protons-protons au Grand Collisionneur de Hadrons (LHC) avec l'expérience Compact Muon Solenoid (CMS), la section efficace de $\mathrm{VH}(\rightarrow \mathrm{b\bar{b}})$ est présentée dans des régions mutuellement exclusives de l'espace des phases et definies par le moment transversal du boson vecteur. Cette méthode est connue sous le nom de "simplified template cross sections" (STXS). Des catégories d'analyse ciblant un important moment transversal du boson de Higgs ont été étudiées pour mesurer d'éventuelles déviations par rapport au SM. Nous avons mesuré la force du signal $\mathrm{VH}(\rightarrow \mathrm{b\bar{b}})$ du Run 2, soit $1.15^{+0.22}_{-0.20}$, conduisant à la première observation du processus $\mathrm{VH}(\rightarrow \mathrm{b\bar{b}})$ par la Collaboration CMS. Aucune déviation statistiquement significative n'a été observée dans aucun des régions du modèle STXS. Dans la deuxième partie de cette thèse, nous présentons une approche basée sur les Graph Neural Networks (GNN) pour obtenir l'efficacité des classificateurs pour l'identification des quarks bottom. Cette approche surmonte plusieurs limitations des approches traditionnelles qui calculent l'efficacité en utilisant des coupures de sélection.

Contents

Part I
Introduction

1 The Standard Model of Physics

The Standard Model (SM) of particle physics is a theoretical framework that describes the fundamental particles and their interactions via three of the four known fundamental forces: electromagnetic, weak, and strong interactions. It is considered one of the most successful theories in physics due to its ability to predict and explain a wide range of experimental results.

With the discovery of the Higgs boson in 2012 [1][2], the last missing piece of the SM was also found. In the following Sections 1.1-1.5, we discuss the building blocks of the SM. Section 1.6 describes various production and decay modes of the Higgs boson. Finally in Section 1.7, the simulation of collision of events is described.

1.1 The SM Lagrangian

The principle of least action, also known as Hamilton's principle, is a fundamental concept in classical mechanics and field theory. The action S is a functional defined as:

$$S[q(t)] = \int_{t_1}^{t_2} L(q, \dot{q}, t)\, dt \tag{1.1}$$

where $q(t)$ represents the configuration of the system as a function of time t, L is the Lagrangian (which depends on the generalized coordinates q, their time derivatives $\dot{q}$, and time t), and the integral is taken over the time interval from t_1 to t_2. The principle of least action states that the path taken by a physical system between two points in time is the one for which the action functional is minimized ($\delta S = 0$). Here δS represents the variation of the action with respect to all possible variations of the trajectory $q(t)$ that preserves the boundary conditions at time t_1 and t_2.

The principle of least action can be extended from particle mechanics to field theories, where fields (such as scalar fields, vector fields, or tensor fields) vary continuously in space and time. In field theory, the action functional becomes a functional of the field variables and their derivatives with respect to space and time. For example, in classical field theory, the action functional takes the form:

$$S[\phi(x^\mu)] = \int \mathcal{L}(\phi, \partial_\mu \phi, x^\mu)\, d^4x \tag{1.2}$$

where $\phi(x^\mu)$ represents the field configuration as a function of space-time coordinates x^μ, $\mathcal{L}$ is the Lagrangian density, and the integral is taken over space-time. The principle of least action states that the physical field configuration is the one for which the action functional is minimized, subject to appropriate boundary conditions.

The principle of least action can be expressed as a differential equation in the form of Euler–Lagrange equation:

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \right) - \frac{\partial \mathcal{L}}{\partial \phi} = 0 \tag{1.3}$$

which for the simple case of a real scalar field $\phi(x)$ with a Lagrangian density

$$\mathcal{L}_{\text{Klein-Gordon}} = \frac{1}{2}(\partial^\mu \phi \partial_\mu \phi) - m^2\phi^2 \tag{1.4}$$

yields the Klein-Gordon equation (without interactions):

$$(\partial^\mu \partial_\mu + m^2)\phi = 0 \tag{1.5}$$

Field theories based on the principle of least action include classical field theories such as classical electromagnetism and general relativity, as well as quantum field theories such as quantum electrodynamics and the standard model of particle physics.

The Lagrangian density of the Standard Model (SM) ($\mathcal{L}_{\text{SM}}$) is given by:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{matter}} + \mathcal{L}_{\text{gauge}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}} \tag{1.6}$$

It needs to be renormalizable (renormalization is a procedure used to handle infinities that arise in quantum field theory calculations) and invariant under the local gauge symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$. According to Noether's theorem [3], for every discrete symmetry, there exists a corresponding conserved current and charge. The $SU(3)_C$ (color) symmetry corresponds to the strong force, which binds quarks together to form hadrons (such as protons and neutrons). For example, the Noether charge associated with $SU(3)_C$ symmetry is the color charge, which is conserved in strong interactions. The $SU(2)_L$ describes the weak isospin interactions which act only between left-handed fermions, and the $U(1)_Y$ describes the weak hypercharge interactions that differ between the left-handed and right-handed fermions. The corresponding conserved charges associated with the electroweak sector are the weak isospin and weak hypercharge respectively. The individual parts of Equation 1.6 are described in brief below.

1.2 Matter Lagrangian Density

The matter Lagrangian density describes the dynamics of fermions (quarks and leptons) in the Standard Model. It includes terms for the kinetic energy and interactions of fermion fields. Fermions are spin-1/2 particles and are arranged in weak isospin doublets. Weak isospin is a quantum number that characterizes the behavior of particles under the weak nuclear force. The weak isospin doublets are characterized by the same weak isospin quantum number but have different electric charges. They come in three different families and differ in mass.

Leptons

There are three lepton generations of isospin doublets as shown in Figure 1, each consisting of a left-handed neutrino (weak isospin up) and a left-handed charged lepton (weak isospin down). The charged leptons interact via both the electromagnetic and weak forces, whereas the neutral leptons only interact via weak interactions. Neutrinos are assumed to be massless in SM. However, neutrino oscillations, signaling non-zero neutrino masses, have been detected [5]. Given the negligible mass

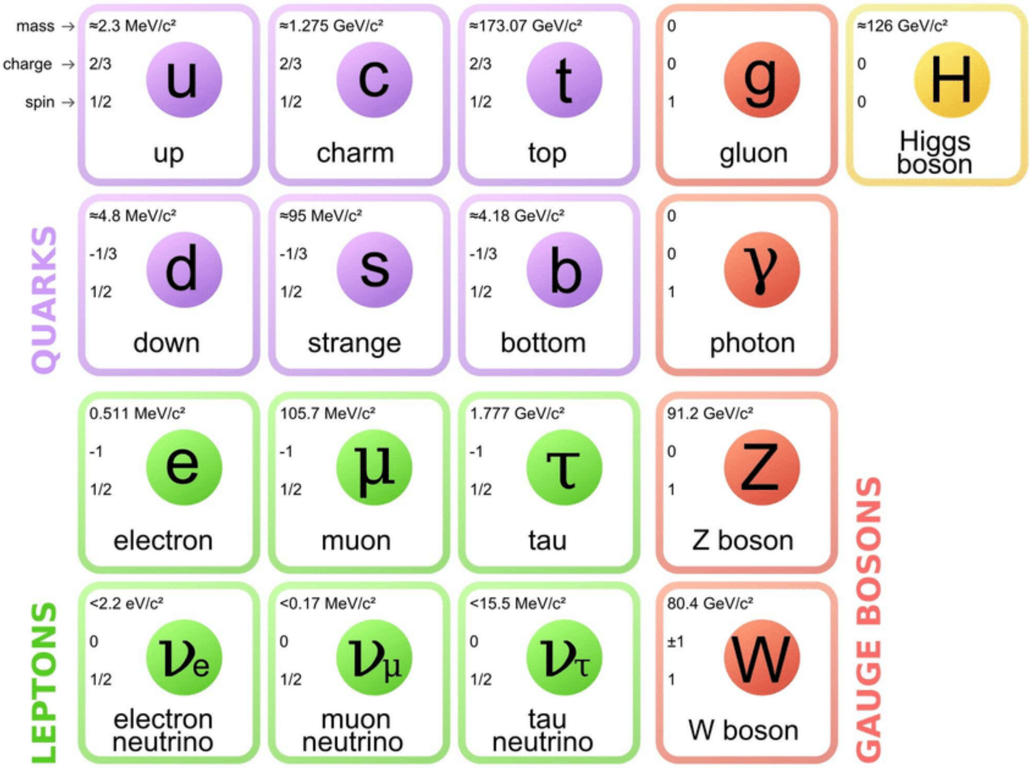


Figure 1: Fundamental particles of the SM with some of their properties [4].

of neutrinos relative to other particles considered in this study, they are treated as massless in all computations. Lepton flavor universality states that three generations of leptons interact with the same strength in an electroweak process. The lepton numbers before and after the interaction are conserved: the lepton number conservation. There are three different lepton numbers: the electron-lepton number, the muon-lepton number, and the tau-lepton number. The electron lepton number is 1 for the electron and the electron neutrino, and 0 for non-leptons. The muon and tau lepton numbers are computed similarly.

Quarks

As for leptons, there are three quark generations of isospin doublets, as shown in Figure 1, each consisting of a left-handed up-type quark (weak isospin up) and a left-handed down-type quark (weak isospin down). Quarks interact through all known forces with masses measured to be in a range from 2 MeV to 172 GeV. Quarks possess both electric and color charges. The confinement phenomenon within Quantum Chromodynamics (QCD) confines quarks solely within color-neutral bound states known as hadrons, ensuring color neutrality. These bound states take the form of mesons when composed of a quark and anti-quark, and baryons when composed of three (anti-)quarks. Experimental observations confirm the conservation of the baryon quantum number, similar to the lepton family quantum number. However, these quantum numbers are not based on a fundamental symmetry.

1.3 Gauge Lagrangian Density

The gauge Lagrangian density contains kinematic terms of gauge bosons (photon, gluons, and weak gauge bosons). The strength of the boson-fermion interaction is measured experimentally. Gauge bosons are spin-1 particles acting as mediators of the electromagnetic, weak and strong forces. The photon mediates the electromagnetic interaction while the charged W$^{\pm}$ and Z^{0} bosons mediate weak interaction. Unlike photons and gluons, W and Z bosons are massive and this leads to the limited range of weak force. Gluons mediate the strong interaction. They carry color (red, green, blue) and an anti-color (anti-red, anti-green, anti-blue) charge. Based on the SU(3) symmetry, the presence of both color and anti-color charges leads to the formation of octet and singlet states, similar to spin states. Singlet states, being color neutral, cannot engage in interactions, leaving the eight colored gluons observable.

1.4 Higgs Lagrangian Density

The introduction of the Higgs term in the SM Lagrangian allows to [6][7]:

- describe the particle masses by spontaneous symmetry breaking.
- make the SM a renormalizable theory.

The Higgs sector contains a SU(2) doublet of spin-0 complex scalar fields (ϕ) with hypercharge Y=1 which preserves $SU(2)_L \times U(1)_Y$ invariance. The scalar field ϕ adds four new real parameters (four degrees of freedom) to the SM:

$$\phi = \begin{pmatrix} \phi^{+} \\ \phi^{0} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.7}$$

where the superscripts denote the electric charge. $\phi^{+}$and ϕ^{0} are both complex fields. The Lagrangian density describing the field is:

$$\mathcal{L}_{\text{Higgs}} = \left(D_\mu \phi\right)^{\dagger} \left(D_\mu \phi\right) - V\left(\phi^{\dagger}\phi\right). \tag{1.8}$$

where

$$\left(D_\mu \phi\right)^{\dagger} \left(D_\mu \phi\right) \tag{1.9}$$

is the kinematic term and

$$V\left(\phi^{\dagger}\phi\right) = \mu^2 \phi^{\dagger}\phi + \lambda\left(\phi^{\dagger}\phi\right)^2. \tag{1.10}$$

is the potential term. The covariant derivative of ϕ is:

$$D_\mu \phi = \left(\partial_\mu + i g_W T^i W_\mu^i + i\frac{1}{2} g' Y B_\mu\right)\phi, \tag{1.11}$$

where W_μ^i and B_μ are, respectively, the $SU(2)_L$ and $U(1)_Y$ gauge bosons with $i = 1, 2, 3$. $T^i = \frac{\tau^i}{2}$ where τ^i are the three Pauli matrices. g_W and g' are the couplings of the W_μ^i and B_μ gauge bosons respectively.

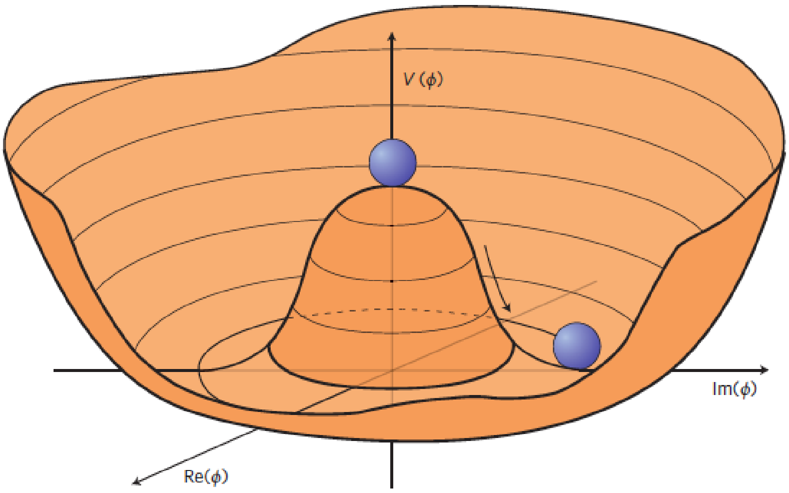


Figure 2: Potential of the complex scalar field ϕ from Equation 1.10 for $\lambda > 0$ and $\mu^2 < 0$.

The minima of the potential given in Equation 1.10 corresponds to the lowest energy state of the scalar filed ϕ known as the vacuum energy state. For such a minima to exist, λ is required to be positive. Further choosing $\mu^2 < 0$ leads to a 'Mexican hat-shaped' potential as shown in Figure 2. The expectation value of the vacuum energy state of the scalar potential ϕ is given by:

$$\langle 0|\phi|0\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.12}$$

The gauge symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$ is spontaneously broken into $SU(3)_C \times U(1)_{EM}$ by the introduction of a non-zero vacuum expectation energy state. The ground state of the SM Lagrangian is only symmetric with respect to $SU(3)_C \times U(1)_{EM}$ as opposed to $SU(3)_C \times SU(2)_L \times U(1)_Y$ symmetry of the SM Lagrangian.

Gauge boson masses

Using the Higgs field in a particular gauge, known as the unitary gauge,

$$\phi(x) = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.13}$$

in the $SU(2)_L \times U(1)_Y$ invariant $\mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{gauge}}$, we obtain terms in the Lagrangian describing the Higgs boson field, vector boson fields as well as the photon field. Applying the Euler-Lagrange equation 1.3 on the terms describing the Higgs boson field leads to the following Klein-Gordan equation:

$$\mathcal{L}_{\text{Higgs terms}} = \frac{1}{2}(\partial_\mu \partial^\mu h + 2\mu^2 h^2) \tag{1.14}$$

Identifying the mass term in Equation 1.14 by comparing it with Equation 1.5, we can show that the mass of the Higgs boson is:

$$m_{\text{H}} = \sqrt{2}|\mu| = \sqrt{2\lambda}v \tag{1.15}$$

Similarly, applying the Euler-Lagrange equation on the terms describing the vector boson and photon fields, we can obtain the mass of the W$^{\pm}$ boson:

$$m_{\mathrm{W}^{\pm}} = \frac{1}{2} g_W v \text{ where } W_\mu^{\pm} \equiv \frac{1}{\sqrt{2}}(W_\mu^1 \mp i W_\mu^2) \tag{1.16}$$

the Z boson:

$$m_Z = \frac{v}{2}\sqrt{g_W^2 + g'^2} \text{ where } Z_\mu \equiv \frac{1}{\sqrt{g_W^2 + g'^2}}\left(g_W W_\mu^3 - g' B_\mu\right) \tag{1.17}$$

and the photon:

$$m_A = 0 \text{ where } A_\mu \equiv \frac{1}{\sqrt{g_W^2 + g'^2}}\left(g' W_\mu^3 + g_W B_\mu\right) \tag{1.18}$$

Before the spontaneous symmetry breaking, we had four massless spin 1 gauge bosons (3 W and 1 B) and four spin 0 scalar fields from the Higgs doublet. This represents a total of 4×2 transverse polarizations$+4 = 12$ degrees of freedom. After the spontaneous symmetry breaking, we have three massive spin 1 bosons (W$^+$, W$^-$, Z), one massless spin 1 boson (the photon, A) and one massive scalar field, the Higgs field giving a total of 3×3 transverse and longitudinal polarizations of massive W bosons$+$ 1×2 transverse polarizations of massless photon $+ 1 = 12$. Thus, there is no loss of degrees of freedom. Three out of the four scalar degrees of freedom (the so-called 'Goldstone modes') of the Higgs doublet get absorbed by the unitary gauge of the Higgs field as longitudinal polarizations of the three vector bosons (W$^{\pm}$, Z) giving them mass while the remaining degree of freedom is the scalar Higgs boson.

1.5 Yukawa Lagrangian Density

Fermion masses

The Yukawa Lagrangian density describes the interactions between fermions and the Higgs field, which give rise to fermion masses. Unlike gauge boson masses, fermion masses are not generated via spontaneous symmetry breaking. Inserting the unitary gauge of the Higgs field (Equation 1.13) in the Yukawa Lagrangian leads to terms that describe the interactions between fermions and the Higgs field, and the fermion mass terms. The mass of fermion f in generation i is given by the Yukawa coupling λ_f^i as

$$m_f^i = \frac{1}{\sqrt{2}}\lambda_f^i v. \tag{1.19}$$

The Yukawa couplings (and the fermion masses) remain as free parameters in the SM Lagrangian and are determined from the experiments. The neutrino masses could also be added via Yukawa interactions but require the introduction of the right-handed neutrinos in the SM. Alternative ways like Majorana mass terms can also lead to massive neutrinos.

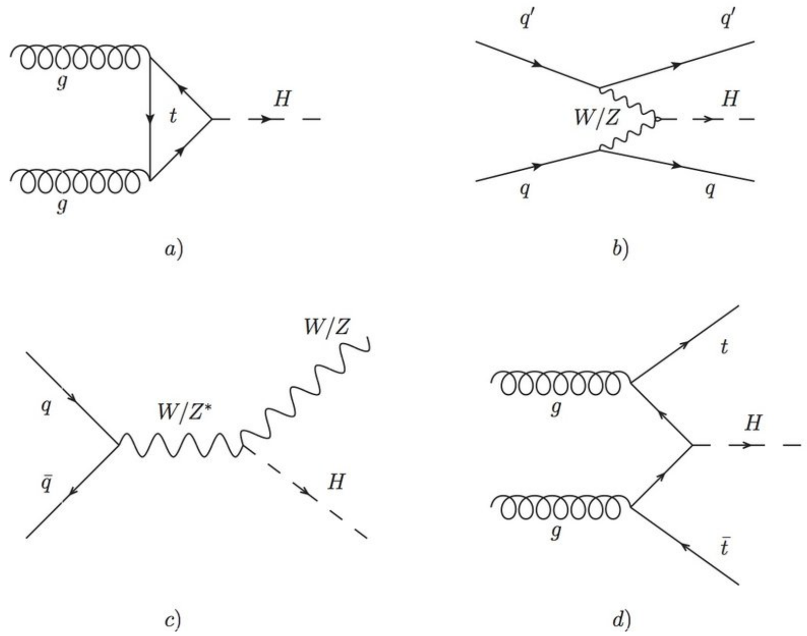


Figure 3: Feynman diagram for the different Higgs boson production modes. gluon-gluon fusion (a), vector boson fusion (b), quark-induced associated production (c), and associated production with top quarks (d).

1.6 Higgs boson production and decay channels

The coupling of the Higgs boson to fermions and vector bosons determines its production rate and branching ratios. As discussed in Section 1.4 and 1.5, the relationship between the Higgs boson coupling and the mass of fermions and vector bosons is given by:

$$\begin{aligned} g_F &= \sqrt{2}\frac{m_f}{v} \\ g_V &= 2\frac{m_V^2}{v} \end{aligned} \tag{1.20}$$

Thus the couplings are directly proportional to the mass of the fermions and to the square of the mass of the vector bosons.

The Feynman diagrams for the different Higgs boson production modes are shown in Figure 3. The cross section of the Higgs boson production mode as a function of the center of mass energies for proton-proton colliders is shown in Figure 4. Their cross section is calculated up to NNLO QCD and NLO electroweak corrections. The major Higgs boson production modes are:

- gluon-gluon fusion (ggF): The dominant production mechanism of the Higgs boson at LHC is the gluon-gluon initial state (ggH) coupling through a fermion loop to the Higgs boson dominated by top quark contributions (top-quark dominates since the coupling of the Higgs boson to fermions is proportional to their masses). When the Higgs boson is produced through this production mode, no additional particles are expected in the event apart from the Higgs boson decay products.

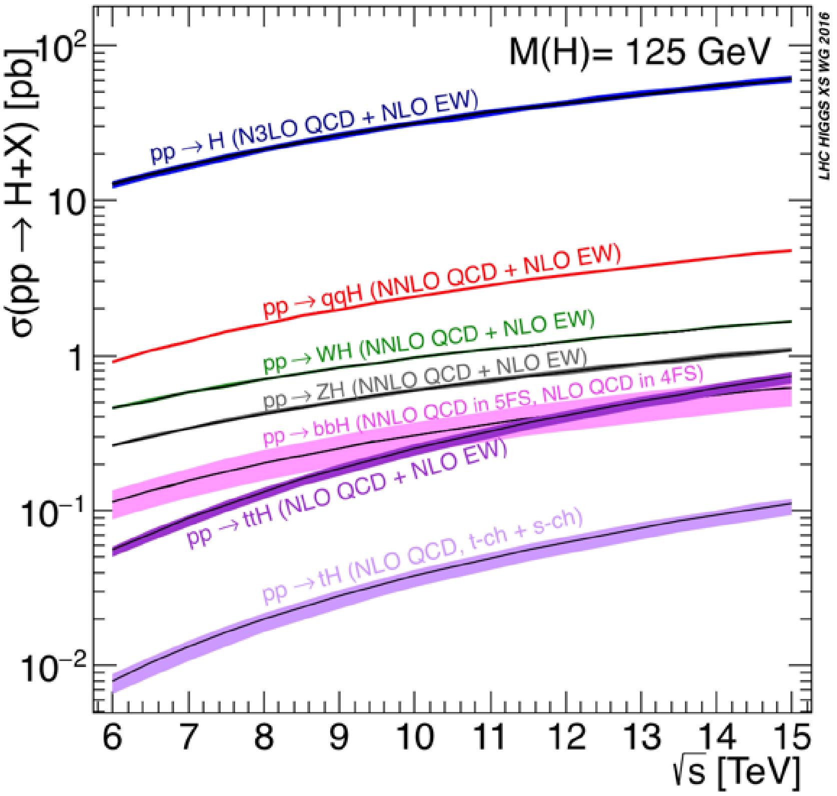


Figure 4: The cross section of the Higgs boson production mode as a function of the center of mass energies for proton-proton colliders [8]. The uncertainties on the calculations are represented by the bands around the curve.

- vector boson fusion (VBF): The sub-dominant production of Higgs boson initiated by the fusion of two electro-weak vector bosons into a Higgs boson, roughly an order of magnitude lower in cross section than ggF. This production mode predicts two energetic jets in addition to the Higgs boson products.

- associated production (VH): The Higgs boson production in association with a vector boson is dominated by the quark anti-quark induced process with some minor contribution from the gluon-induced process. This production mode predicts a vector boson (and its decay products) and the Higgs boson (and its decay products). The reconstruction of the vector boson in leptonic decay mode allows the suppression of the QCD multi-jet background (particularly important when the bottom quarks decay mode of the Higgs boson is targeted). The impact on measurements from the remaining dominant electroweak background such as the V+jets processes can be further reduced using kinematic cuts and multi-variate (MVA) methods. This production mode is the focus of this thesis.

- associated production with top quarks (ttH): This production channel has the lowest production cross section but is the only channel through which the top quark Yukawa coupling with the Higgs boson can be measured directly. This production mode predicts two top-quark along with the Higgs boson.

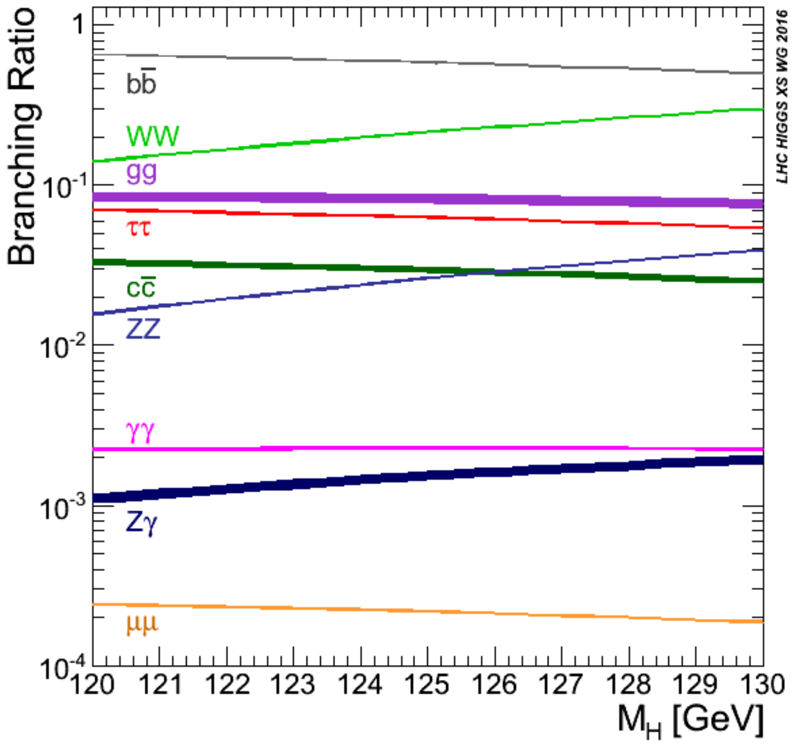


Figure 5: The branching ratios for the different Higgs boson decay channels for different Higgs masses M_H and their associated theoretical uncertainties (width of the band) [8].

Since the Higgs boson can couple to all massive particles, several decay channels are available. The branching ratios for the different Higgs boson decay channels are shown in Figure 5. The dominant fermion decay mode at the measured mass of about 125 GeV, is to a bottom quark pair. Yet, because of the overwhelming QCD background, it is measured with one of the lowest precision [9]. Increasing the precision of this measurement and sensitivity towards deviations from the SM by improving the analysis strategy and utilizing the full Run 2 data is the objective of this doctoral thesis.

Even though the branching fraction of the Higgs boson to photons ($H \rightarrow \gamma\gamma$) and Z bosons ($H \rightarrow ZZ^* \rightarrow 4l$) is low, because of the excellent resolution of the reconstructed energy and momentum of its final state particles (photons or leptons) in the CMS detector, enough background suppression can be achieved. This was one of the reasons why the Higgs boson was first observed in 2012 [1][2] in these golden decay channels. The $H \rightarrow W^{\pm}W^{\pm(*)} \rightarrow l^+\nu_l l'^- \bar{\nu}_{l'}$ channel has a relatively large branching fraction, however, since neutrinos in the final state are not reconstructed in the detector, the resolution obtained of the invariant mass of the Higgs candidate is poor impacting the signal by background ratio (S/B). The $H \rightarrow b\bar{b}$ and $H \rightarrow \tau^+\tau^-$ require the clustering of the b-jets and τ decays, and suffer from the large QCD multi-jet background. In this thesis, the focus is on $H \rightarrow b\bar{b}$ decay channel.

1.7 Event Simulation

The simulation of proton-proton collision is highly complex as shown in Figure 6. The proton is characterized by the parton (i.e. quarks and gluons) density functions. The parton density functions (PDFs) gives the probability of finding a parton carrying a momentum fraction x at a squared energy scale/momentum transfer Q^2. The initial and final state partons can emit QCD radiations. The colorless hadrons are then formed in the hadronization step. The unstable particles formed in the hadronization step undergo decay. Further, since this is a hadron-hadron collision, multiple partons interact in the same collision. This gives rise to additional hadron production called the underlying event. This is different from pile-up, which is the multiple proton-proton collisions, in the same bunch of colliding protons at LHC. These event simulation steps are described below.

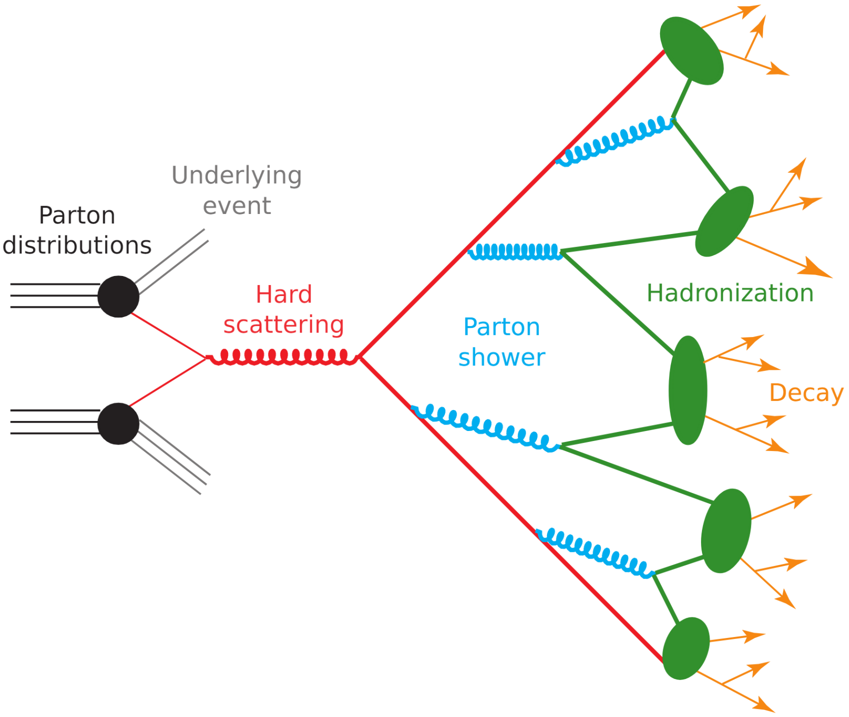


Figure 6: Pictorial representation of MC event generation process [10].

Hard scattering

The total production cross section σ of $p_1 p_2 \to X$ can be factorized into the hard scattering cross section $\hat{\sigma}_{ij}$ of the partonic process $(ij \to k)$ (where i, j and k are partons; i belongs to p_1 and j belongs to p_2) and the PDF f (see Figure 6) [11]

$$\sigma\left(p_1, p_2\right) = \sum_{i,j} \int \mathrm{d}x_1\, \mathrm{d}x_2 f_{i/p_1}\left(x_1, \mu_F^2\right) f_{j/p_2}\left(x_2, \mu_F^2\right) \times \hat{\sigma}_{ij}\left(p_1, p_2; \alpha_S\left(\mu_R\right); Q^2, \mu_R^2, \mu_F^2\right) \tag{1.21}$$

The PDFs can not be derived from theory and they rely on experimental measurements based on the deep inelastic scattering (DIS) cross sections. The DIS experiments [12] have shown that at low Q^2 the three valence quarks become more and more dominant in the nucleon. As Q^2 increases, more quark-antiquark pairs are created which carry a low momentum fraction x. They constitute the sea quarks. Also, the fraction carried by the gluons increases with increasing Q^2. The NNPDF3.0 set [13] used in this analysis is shown in Figure 7 for the energy scale $Q^2 = 10^4$ GeV^2.

The computation of the amplitude of the partonic process $ij \to k$ leads to divergences of two kinds:

- the ultra-violet (UV) divergence: they appear in the amplitude calculations of loops which can have large momentum transfer.
  - The UV divergences can be avoided by the renormalization. Renormalization is a procedure that allows to absorb the infinite quantities in the calculation by a redefinition of a finite number of parameters. For this, a renormalization scale μ_R is introduced to counter higher order terms.
- the infrared (IR) divergence: they can appear because (i) either a real or virtual particle can reach zero momentum (soft divergence), or because (ii) a particle can emit collinear radiation.
  - The IR divergences in case of (i) cancel out, (ii) are fixed by introducing a cut-off energy, the factorization scale μ_F. μ_F is the boundary above which the perturbation theory can be applied to QCD. The hard scattering (or the scale at which PDFs are evaluated) takes place above the factorization scale.

Parton shower

Since partons carry color charge, the incoming and outgoing partons emit soft (low transverse momentum) and collinear (small angle θ) quarks or gluons, as a part of the higher-order corrections to the hard process. These emissions can not be computed exactly and are simulated using Monte Carlo (MC) methods. The probability that no splitting occurs is given by the Sudakov form factor:

$$\Delta_i\left(Q^2, q^2\right) = \exp\left(-\int_{q^2}^{Q^2} \frac{\mathrm{d}k^2}{k^2}\frac{\alpha_s}{2\pi}\int_{q^2/k^2}^{1-q^2/k^2} \mathrm{d}z P_{ji}(z)\right) \tag{1.22}$$

for each parton at the energy scale $k^2 = Q^2$. In the Equation 1.22, α_s is the strong coupling constant. As long as $q^2 > Q_0^2$, where Q_0^2 is the hadronization cutoff scale (discussed in Section 1.7), radiation of a parton with momentum fraction z is generated according to the flavor-dependent splitting function $P_{ji}(z)$. The splitting function $P_{ji}(z)$ can be computed using perturbative QCD up to a fixed order and gives the distribution of energy fraction z. Below the cut-off scale, the evolution terminates. To avoid double counting of jets produced by the event generator (which generates the hard scattering process) and the parton shower generator, the so-called matching and merging procedures are applied (discussed further in Section 3.4).

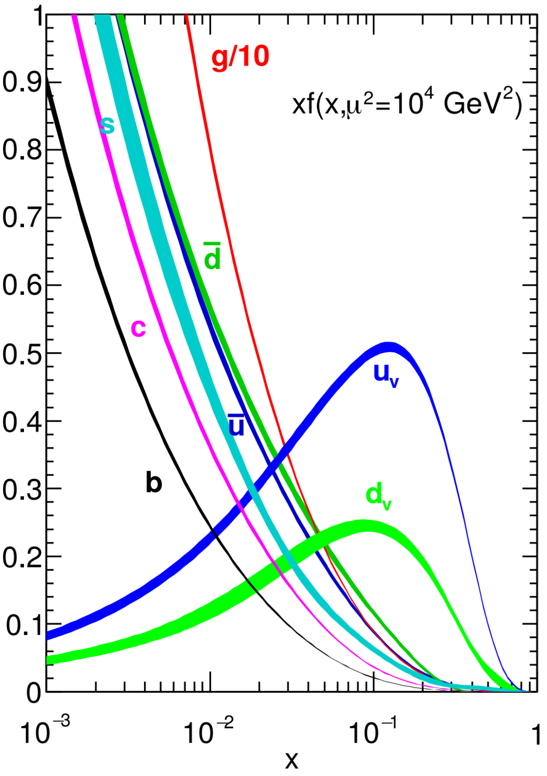


Figure 7: NNLO PDF distribution in the NNPDF3.0 set [13] for a fixed energy scale $Q^2 = 10^4$ GeV^2.

Hadronization

As the energy scale Q^2 in the parton shower decreases, the strong coupling constant α_s increases. At the hadronization cutoff scale, $Q_0^2 \approx 1$ GeV^2, $\alpha_s \approx 1$, the perturbation theory does not hold anymore and non-perturbative hadronization effects need to be taken into account. For the hadronization process, several models exist. The one used in PYTHIA8 [15] (the parton shower generator used in this analysis) is the so-called Lund string model (shown in Figure 8) and is tuned on experimental data [16]. It is based on the observation that the potential energy due to strong force between the color-connected objects increases linearly with the distance between them. If the constituents of a quark or anti-quark move apart, the potential energy increases until it is sufficient to produce a new quark/anti-quark pair. The procedure continues until the energy is insufficient to produce quarks from the vacuum.

Underlying event

The underlying event (UE) models the interaction among the colliding protons apart from the partons undergoing the hard interaction. The particles produced in the UE event typically have low momenta. The set of free parameters (such as the length of strings in hadronization, α_s, etc) is 'tuned' on experimental data. In this thesis, the CUETP8M1 tune [17] and the more recent CP5 [18] tunes are used for the MC event generation.

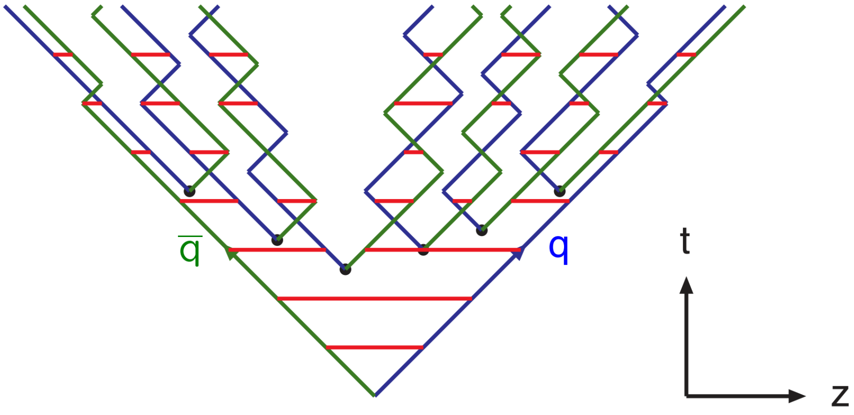


Figure 8: Schematic sketch of the Lund string model [14]. The quark and anti-quark connected via red string breaks creating new quark anti-quark pairs when they move apart.

Pile up

LHC accelerates protons in bunches, each containing about 10^{11} protons. The pile-up (PU) events are the additional proton-proton collisions apart from the triggered event collision in the same bunch crossing or different bunch crossing. The multiple proton-proton interactions in the same bunch are referred to as the 'in-time' pile-up. 'Out-of-time' pile-up comes from the proton-proton collisions in the earlier and later bunch crossings which leave signals in the detector. In-time PU event identification requires correct identification and association of particles with the primary interaction vertex.

2 CMS Detector

The Large Hadron Collider (LHC) is the world's largest and highest energy particle accelerator and is a part of the European Organization for Nuclear Research (CERN) located at the Swiss-French border. The LHC accelerates hadrons to the high energies needed for research in particle physics. The detectors at LHC record data through the interaction of particles with the detector material. The LHC is further discussed in Section 2.1 while the remaining Sections in the chapter describe the CMS detector which was used for taking the data used in this analysis.

2.1 LHC and beam operations

LHC situated at CERN is a hadron-hadron collider, currently operating at a center of mass energy of 13.6 TeV. It accommodates four experiments: the multi-purpose detectors ATLAS (A Toroidal LHC ApparatuS) [19] and CMS (Compact Muon Solenoid) [20], along with the ALICE (A Large Ion Collider Experiment) [21] detector, which focuses on heavy-ion collisions, and LHCb [22], dedicated to b quark physics. Additionally, numerous smaller experiments such as TOTEM [23] and LHCf [24] are positioned along the collider ring, often in proximity to one of the four primary experiments.

CERN operates a total of nine accelerators and two decelerators [26]. The accelerator complex as shown in Figure 9 consists of a series of machines that accelerate a beam of particles to increasingly higher energies before injecting it into the next machine in the sequence. The LHC is the last element in this chain. Since 2020, Linear accelerator 4 (Linac4) is the source of proton beams for the CERN accelerator complex. It accelerates negative hydrogen ions to 160 MeV which are then passed on to the Proton Synchrotron Booster (PSB). The ions undergo a process of electron stripping upon injection from Linac4 into the PSB, leaving protons. These protons are then accelerated to 2 GeV for injection into the Proton Synchrotron (PS), which further accelerates them to 26 GeV. Subsequently, the protons are directed to the Super Proton Synchrotron (SPS), where they undergo acceleration to reach the energy of 450 GeV.

The luminosity $\mathcal{L}$ of a collider is the number of collisions per second ($\frac{dR}{dt}$) per unit area (i.e. cross section, σ).

$$\frac{dR}{dt} = \mathcal{L} \times \sigma \tag{2.1}$$

It is expressed in units of cm^{-2}s^{-1}. This luminosity is sometimes referred to as the 'instantaneous' luminosity. A higher luminosity implies a greater likelihood of particles colliding. This is achieved by packing more particles in the beam and by focusing the beam more tightly. The 'integrated' luminosity $\mathcal{L}_{\text{int}}$ is the integral of the instantaneous luminosity over time:

$$\mathcal{L}_{\text{int}} = \int \mathcal{L}\text{dt} \tag{2.2}$$

It is expressed in units of fb^{-1} [1 barn = 10^{-24} cm^2] and measures the total number of events during a particular period of data taking. The Run 2 data used in this

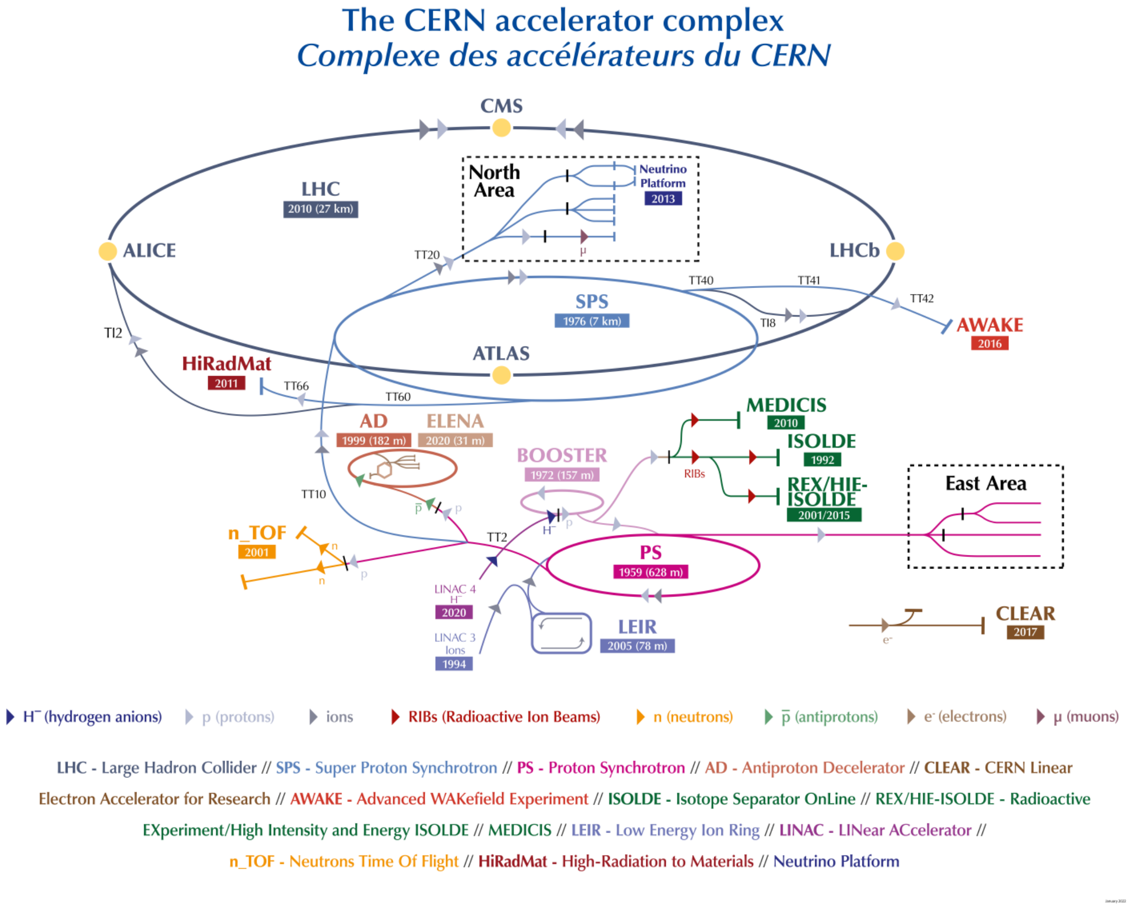

The CERN accelerator complex
Complexe des accélérateurs du CERN
CMS
North Area
Neutrino Platform
2013
LHC
2010 (27 km)
ALICE
LHCb
TT20
TT40
TT41
TT42
AWAKE
2016
SPS
1976 (7 km)
TI8
TI2
ATLAS
HiRadMat
2011
TT66
TT60
AD
1999 (182 m)
ELENA
2020 (31 m)
BOOSTER
1972 (157 m)
MEDICIS
2010
ISOLDE
1992
RIBs
REX/HIE-ISOLDE
2001/2015
East Area
TT10
n_TOF
2001
TT2
PS
1959 (628 m)
LINAC 4
H⁻
2020
LINAC 3
Ions
1994
LEIR
2005 (78 m)
CLEAR
2017
H⁻ (hydrogen anions)
p (protons)
ions
RIBs (Radioactive Ion Beams)
n (neutrons)
p̄ (antiprotons)
e⁻ (electrons)
μ (muons)
LHC - Large Hadron Collider // SPS - Super Proton Synchrotron // PS - Proton Synchrotron // AD - Antiproton Decelerator // CLEAR - CERN Linear Electron Accelerator for Research // AWAKE - Advanced WAKefield Experiment // ISOLDE - Isotope Separator OnLine // REX/HIE-ISOLDE - Radioactive EXperiment/High Intensity and Energy ISOLDE // MEDICIS // LEIR - Low Energy Ion Ring // LINAC - LINear ACcelerator // n_TOF - Neutrons Time Of Flight // HiRadMat - High-Radiation to Materials // Neutrino Platform

Figure 9: Schematic sketch of the CERN accelerator complex [25].

thesis corresponds to a total integrated luminosity of 138 fb^{-1} [27]. In Run 2, the LHC regularly achieved an instantaneous luminosity of 2×10^{34} $\text{cm}^{-2}\text{s}^{-1}$ [28].

The protons are directed into the two beam pipes of the LHC, where one beam circulates clockwise and the other anticlockwise. It requires approximately 4 minutes and 20 seconds to fully load each LHC ring (one 'fill'), and an additional 20 minutes for the protons to reach their maximum energy of 6.8 TeV [25]. Once operational, the fill is kept in the LHC for extended periods, typically lasting 8 hours. The collision occurs when the two beams cross within the four detectors (ALICE, ATLAS, CMS, and LHCb).

The protons within the LHC accelerated in two opposing beams collide with each other in bunches containing up to 1.15×10^{11} protons, each spaced apart by 25 nanoseconds. These protons traverse the entire 27-kilometer ring individually, with a revolution frequency of 11.2 kHz. Although there are 3564 potential spaces for bunches in a beam, only up to 2556 are occupied for operational reasons, leaving some small gaps due to the injection process, as well as a significant gap (abort gap) to allow for safe extraction and dumping of the beam.

Collision angles at the experiments' collision point can be adjusted to regulate collision rates to manage luminosity and control pile-up to a level compatible with

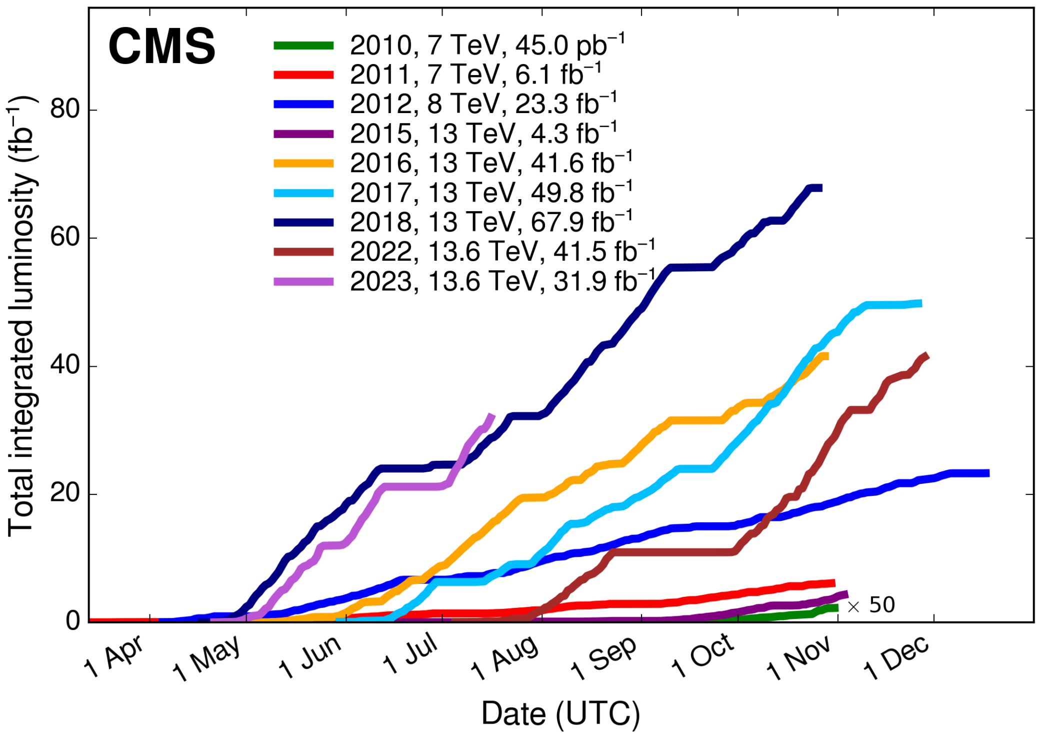


Figure 10: Integrated luminosity versus day delivered to CMS during stable beams for pp collisions at nominal center-of-mass energy [27].

detector capabilities (lumi-leveling). Figure 10 gives the integrated luminosity delivered to the CMS experiment versus time at center-of-mass energy [27].

The LHC accelerator schedule [29], spanning from the initial run of physics data collection in 2011 to the culmination of the upgraded high-luminosity LHC (HL-LHC), is depicted in Figure 11. Preceeding the HL-LHC era, the data-taking period is segmented into three distinct runs. The Run 1 occurred in 2011 and 2012 and operated at center-of-mass energies of 7 and 8 TeV, respectively while Run 2 operated at center-of-mass energies of 13.0 TeV in 2016-2018. Run 3 is currently running at 13.6 TeV. Throughout the LHC program, operations are periodically halted for long shutdowns (LS1-3), during which time technical maintenance and upgrades are implemented for both the accelerator and the detectors. At the end of each data-taking year, a year-end technical stop (YETS) is enforced, during which the accelerators and detectors are powered down to facilitate shorter maintenance and repair activities. An extended YETS (EYETS) was scheduled between the 2016 and 2017 data-taking periods to accommodate among several activities, a more extensive maintenance on the accelerator and, in CMS, the installation of new Phase 1 pixel detectors [30]. The data used in this analysis corresponds to the LHC Run 2 data collected during 2016, 2017 and 2018.

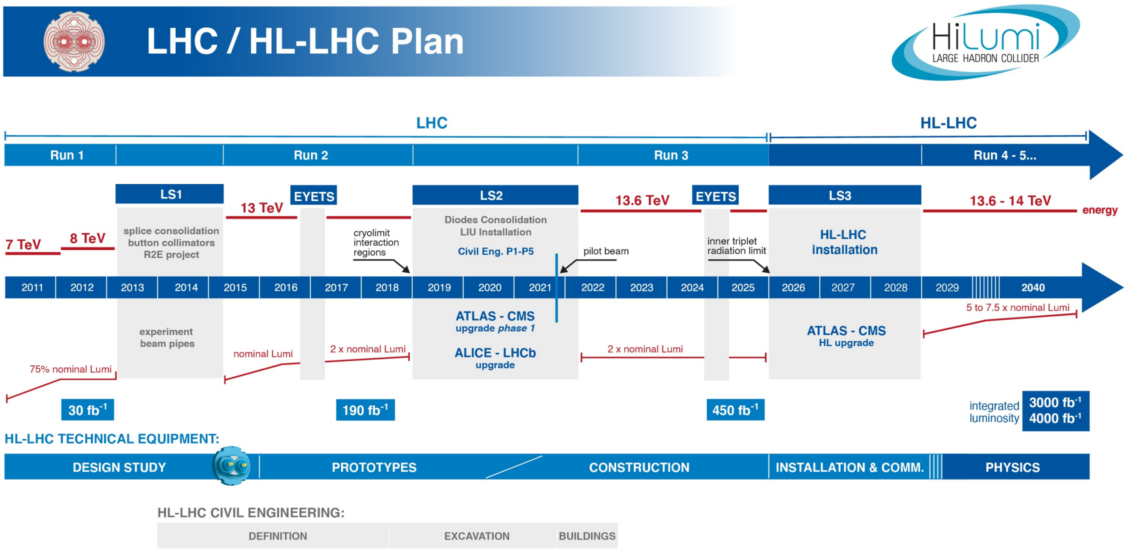


Figure 11: The LHC accelerator schedule [29], spanning from the initial run of physics data collection in 2011 to the culmination of the upgraded high-luminosity LHC (HL-LHC).

2.2 CMS Detector

The CMS detector [31] is a multi-purpose detector designed to detect particles produced in high energy collisions at the LHC. It is 15 meters high and 21 meters long. It can detect muons very accurately and has a 3.8 T solenoid magnet, and thus it is named as the Compact Muon Solenoid detector. The different sub-detectors are arranged cylindrically around the beam axis with the interaction point at the center of the barrel volume. Information from different sub-detectors is used to reconstruct the final states of signal and background processes of this analysis. A sketch illustrating the detector is shown in Figure 12. In the following Sections 2.2.1-2.2.4, we delineate all sub-detectors of the CMS experiment, moving outwards from the innermost detector adjacent to the beam pipe.

2.2.1 Silicon tracker

The silicon tracker is used to reconstruct the particle tracks so as to determine particle momentum and charge along with primary and secondary vertices. An accurate primary vertex reconstruction is necessary to identify PU events. The reconstruction of secondary or displaced vertices helps in the identification of heavy flavor decays.

The length of the silicon tracker is 5.8 meters with a diameter of 2.5 meters covering a pseudorapidity of $|\eta| < 2.5$. It is exposed to high particle flux of about 10^3 particles per bunch crossing, and is the closest detector to the beam pipe. Thus, the silicon tracker is required to be sufficiently radiation-hard and granular enough to detect individual charged particles crossing the tracker. Since the interaction of

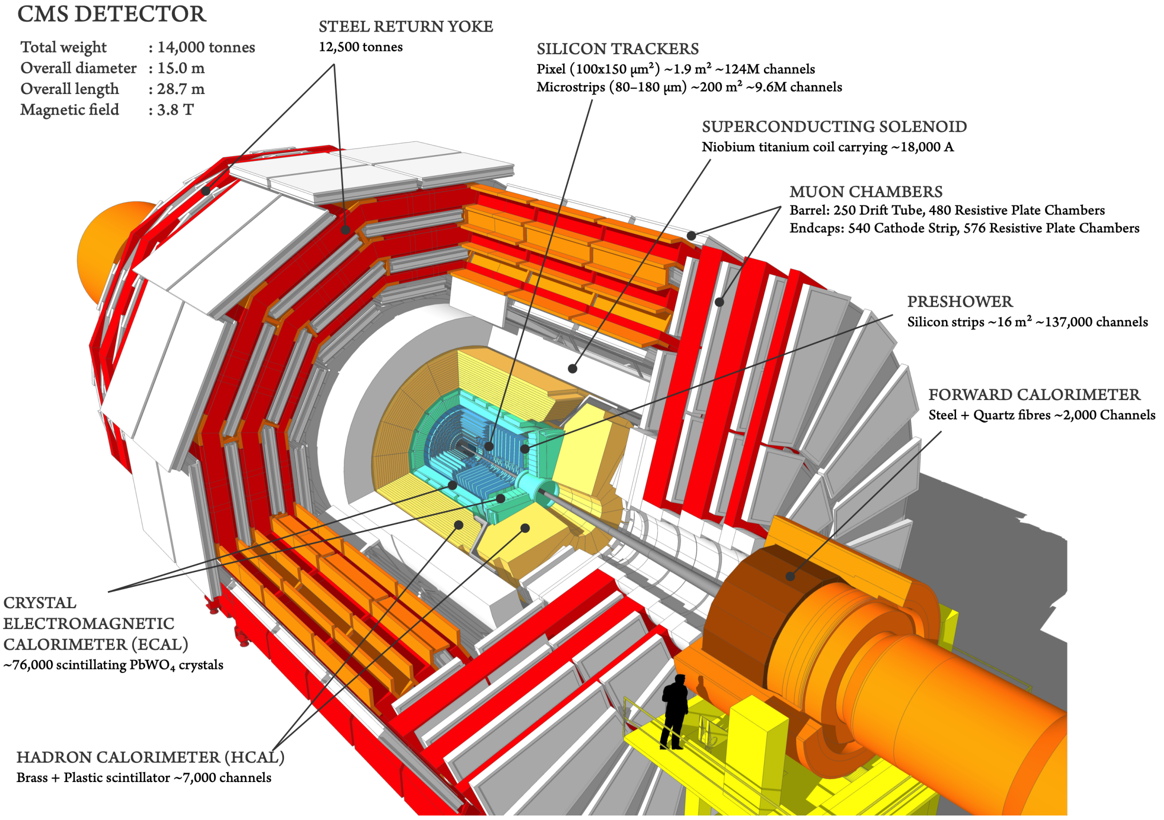

CMS DETECTOR
Total weight : 14,000 tonnes
Overall diameter : 15.0 m
Overall length : 28.7 m
Magnetic field : 3.8 T
STEEL RETURN YOKE
12,500 tonnes
SILICON TRACKERS
Pixel (100x150 μm²) ~1.9 m² ~124M channels
Microstrips (80–180 μm) ~200 m² ~9.6M channels
SUPERCONDUCTING SOLENOID
Niobium titanium coil carrying ~18,000 A
MUON CHAMBERS
Barrel: 250 Drift Tube, 480 Resistive Plate Chambers
Endcaps: 540 Cathode Strip, 576 Resistive Plate Chambers
PRESHOWER
Silicon strips ~16 m² ~137,000 channels
FORWARD CALORIMETER
Steel + Quartz fibres ~2,000 Channels
CRYSTAL
ELECTROMAGNETIC
CALORIMETER (ECAL)
~76,000 scintillating PbWO₄ crystals
HADRON CALORIMETER (HCAL)
Brass + Plastic scintillator ~7,000 channels

Figure 12: Schematic drawing of the CMS detector showing the different detector components [32].

particles with the material of the tracker adds unwanted effects to the measurement (multiple scattering), the tracker system is made of lightweight materials. The radiation length is the mean distance over which a high-energy electron loses all but 1/e of its energy by bremsstrahlung. For silicon, the radiation length is 9.36 cm [33]. The tracker thickness ranges from 0.4 to 1.8 X/X_0 (where X is the physical thickness of a material that a particle traverses and X_0 is the material radiation length) over the pseudorapidity from $-3.5 < \eta < 3.5$ [34]. The minimum thickness of the tracker is at $\eta = 0$. The tracker comprises silicon sensors that leverage the semi-conducting nature of silicon. When charged particles traverse the silicon sensors, they generate electron-hole pairs. Subsequently, an electric field is employed to guide both the electrons and the holes toward the respective electrodes, inducing a current. Within the CMS tracker, silicon sensors are organized in either a pixel or strip layout.

Pixel detector The pixel detector is divided into a barrel (BPix) and two endcaps (FPix). Initially, in phase 0, the barrel detector comprised three layers positioned at radial distances of 4.4 cm, 7.3 cm, and 10.2 cm from the beam pipe until the end of 2016. Subsequently, in the EYETS scheduled between 2016 and 2017, the pixel detector underwent the phase 1 upgrade to a four-layer configuration, with layers positioned at radial distances of 2.9 cm, 6.8 cm, 10.9 cm, and 16 cm from the nominal beam axis. The BPix design ensured a minimum of three pixel hits per track for optimal particle trajectory reconstruction. Further new disks were integrated in the phase-1 into the FPix detectors and the read-out electronics were optimized to accommodate for the increased instantaneous luminosity of $2 \times 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$ [35]. These upgrades resulted in approximately a 20% improvement in tracking efficiency (evaluated from $t\bar{t}$ events) and muon reconstruction efficiency at Run 2 pile up rates, with similar or lower fake rates compared to Run 1 [36]. The upgrade further lead to a 15% improvement in b-jet identification efficiency, calculated at the same charm and light jet misidentification probability of 1% [36].

The pixel detector comprises hybrid detector modules, which integrate silicon sensors bump-bonded to the front-end readout electronics. These modules are mounted on lightweight carbon fiber support ladders to minimize the material, powered by a DC-DC system (electronic circuit that converts a source of direct current i.e. DC from one voltage level to another), and cooled by a CO_2 bi-phase cooling system to a temperature of -23°C. The choice of a pixel size of $100 \times 150\ \mu\text{m}^2$ aims to achieve uniform resolution in all spatial directions, facilitating 3D vertex reconstruction. This capability is particularly crucial for secondary vertex (SV) reconstruction, essential for identifying heavy resonances decaying to heavy quarks. The displacement between the main vertex and the decay vertex of a b-hadron with $p_T = 50$ GeV is approximately 3 mm. With a spatial resolution ranging from 15 to 20 μm, the pixel detector enables the distinction between primary and secondary vertices. The hit efficiency, defined as the probability of finding a cluster within 1 mm of a hit from a $p_T > 1$ GeV particle, exceeds 99% for all layers, although it is slightly lower for the first layer [37].

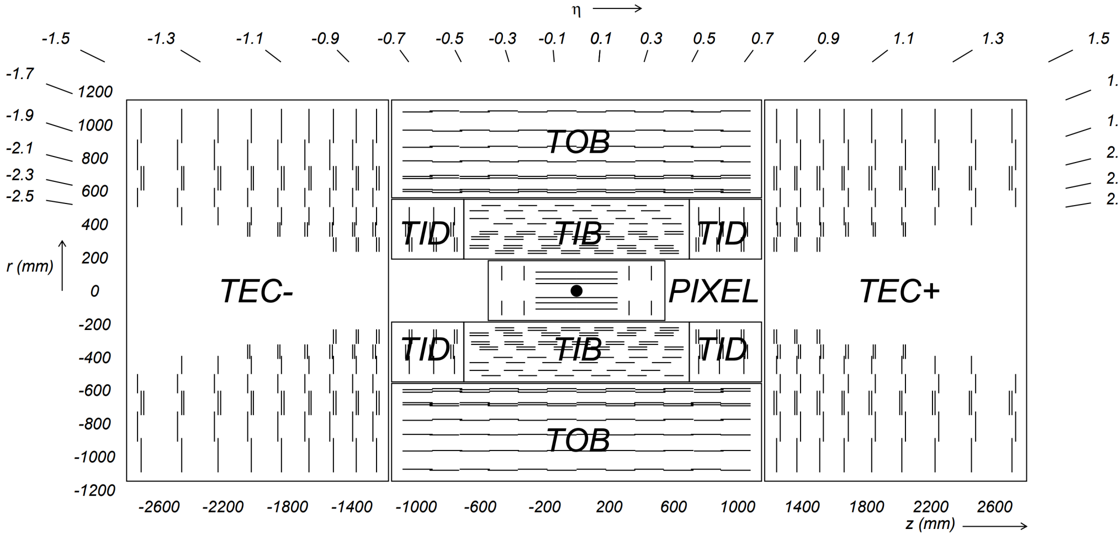


Figure 13: Schematic cross section through the CMS tracker [20]. Each line represents a detector module. Double lines indicate back-to-back modules.

Strip tracker With increasing distance from the beam pipe, the particle flux decreases, thus requiring lower granularity in the outer tracker. The larger-pitch silicon micro-strips, typically measuring 10 cm × 80 μm cover radial distances up to 1.1 m.

The barrel section consists of ten layers, while the endcap regions have nine layers. Wire bonds connect the strips to the read-out electronics at their ends. The outer tracker comprises the tracker inner barrel and disks (TIB/TID), the outer barrel (TOB), and the tracker endcaps (TEC), as shown in Fig. 13.

2.2.2 Calorimeters

Electrons, photons, and hadrons interact with the calorimeters, enabling accurate energy measurements. The Electromagnetic Calorimeter (ECAL), constituting the first calorimetry layer, is used to precisely measure the energies of electrons and photons. Hadrons, which are strongly interacting particles, predominantly deposit their energy in the second layer known as the Hadron Calorimeter (HCAL). Conversely, muons and tau leptons deposit only a minimal fraction of their energy in the calorimeters and are primarily detected through tracking and muon detector subsystems. Although neutrinos evade direct detection, their presence can be inferred from an apparent energy imbalance in the interaction, referred to as 'missing transverse energy' or 'MET'.

The relative energy resolution of a calorimeter (i.e. $\sigma(E)/E$) is described by the sum in quadrature of three terms:

- stochastic term $\propto \frac{1}{\sqrt{E}}$: originating from the statistical nature of the interaction of particles with matter leading to fluctuations in shower formation;
- noise term $\propto \frac{1}{E}$: originating from read-out electronics noise;
- constant term: originating from detector inhomogeneities, imperfections, etc.

ECAL The electromagnetic calorimeter (ECAL) measures the energy of electrons and photons by detecting electromagnetic showers created by successive pair production and Bremsstrahlung. The energy of the incoming particle affects the properties of the shower like its size and is directly correlated with the intensity of the scintillation light emitted. The ECAL consists of three detectors: ECAL Barrel, ECAL Endcap and Preshower detector at the endcap. Out of these, the ECAL Barrel and ECAL Endcap are scintillating calorimeters while the Preshower detector uses Silicon sensors.

The ECAL Barrel and Endcap calorimeters comprise nearly 76,000 lead tungstate $PbWO_4$ crystals (61,200 crystals in barrel, 14,648 crystals in endcaps), each measuring $2.2 \times 2.2 \times 23$ cm and $2.86 \times 2.86 \times 22$ cm respectively. $PbWO_4$ boasts high density ($8.3\,\mathrm{gcm}^{-1}$), a small radiation length (0.89 cm), and a compact Moliere radius (2.19 cm), which defines the transverse size of the electromagnetic shower [38]. These properties ensure minimal shower size and accurate energy measurements of electrons and photons.

The crystals convert the high energy electrons and photons energy into visible light that is then detected by silicon avalanche photodiodes (APDs) in the barrel region and vacuum phototriodes (VPTs) in the endcap region. Scintillation light is emitted at a wavelength of $420 - 430$ nm. The ECAL's performance is monitored using a laser system emitting 440 nm light, which measures crystal transparency and aids to mitigate radiation damage effects. The response of $PbWO_4$ crystals, with 80% of scintillation light emitted within the first 25 ns, is ideal for suppressing pileup at a 25 ns bunch crossing rate. Photons detected by APDs in the barrel and VPTs in the endcaps produce a light output of 4.5 photoelectrons per MeV. The ECAL is cooled by a water system to 18° C.

A preshower detector in front of the Endcap ECAL ($1.653 < |\eta| < 2.6$) composed of two planes of lead followed by a silicon strip sensor is used for π^0 rejection. When the photon passes through the lead layer it causes an electromagnetic shower, which is then detected by the silicon sensors. Due to the finer granularity of the preshower detector than the ECAL, it helps in detecting the pair of photons from a π^0 decay and distinguishes them from the single shower produced from a single photon interacting with the calorimeter material. It is composed of lead absorbers and silicon strip detectors.

The energy resolution for photons from Higgs boson decays ranges between 1.1% and 2.6% in the barrel and between 2.2% and 5% in the endcap [39]. For electrons, it is determined from Z decays to e^+/e^- and ranges from 1.7% to 4.5% depending on electron pseudorapidity and energy loss through bremsstrahlung in the detector material. [40].

HCAL The hadronic calorimeter (HCAL) plays a crucial role in identifying and quantifying quarks, gluons, and neutrinos by gauging the energy and direction of jets as well as the flow of missing transverse energy in events. The HCAL is made up of four parts: Hadron Barrel (HB) $|\eta| < 1.3$, Hadron Outer (HO) $|\eta| < 1.3$, Hadron Endcap (HE) $1.3 < |\eta| < 3.0$ and Hadron Forward (HF) $3.0 < |\eta| < 5.2$.

The HCAL is a sampling calorimeter, with alternating layers of plastic scintillators and absorber plates made from brass or steel. Scintillation light generated within the HCAL is converted into signals by wavelength-shifting (WLS) fibers embedded within the scintillator tiles, which then guide the light to photodetectors. These signals are captured by innovative photodetectors known as hybrid photodiodes (HPDs) and photomultipliers, capable of amplification and operation in high axial magnetic fields. While the majority of the HCAL is positioned inside the CMS magnet, there are supplementary layers located outside the magnet to detect particles emitted from high-energy showers.

The HCAL has a radial depth of approximately 1.2 m, whereas the ECAL's depth is only 23 cm. This is due to the longer interaction length of hadronically-interacting particles compared to electromagnetic interactions. Hadronic showers involve interactions such as nuclear excitations and spallation. The HCAL spans approximately 15 nuclear interaction lengths (the mean distance traveled by a hadronic particle before undergoing an inelastic nuclear interaction), with about 11 situated inside the solenoid. This depth helps HCAL to contain high-energy jets.

In HF HCAL modules, steel absorber plates are employed due to the more severe radiation environment, while hadronic showers are observed through radiation-resistant quartz fibers. The Cerenkov light produced within the quartz fibers is captured by conventional photomultiplier tubes. These forward calorimeters guarantee complete geometric coverage for transverse energy measurement of the event up to

$|\eta| < 5.2$.

The energy resolution of the HCAL is lower compared to the ECAL. For 20-300 GeV pion, the energy resolution as a function of incident energy is determined to be $\left(\frac{\sigma}{E}\right)^2 = \frac{115^2}{E} + 5.5^2$ [41].

2.2.3 Solenoid magnet

CMS detector has 13 m length, 5.9 m inner diameter and 3.8 T superconducting solenoid. The inner tracker and the calorimeters are situated inside the magnet, while the flux return yoke and muon detector are situated outside the magnet. The bending of the trajectory of a charged particle within a magnetic field provides information about particle charge and momentum. A titanium-niobium coil cooled at $-270°$ becomes super-conductive and generates the 3.8 T magnetic field with 20 kA current. A three-layered iron return yoke enveloping the coil serves to complete the magnetic field lines, ensuring that the muon chambers, situated within the return yoke, are adequately exposed to the magnetic field strength.

2.2.4 Muon Chambers

Muons are in the final states of many processes such as the Higgs boson decay to Z bosons and are crucial in the VH($\mathrm{b\bar{b}}$) analysis. Muons are the only particles that pass through the inner detectors and calorimeters and reach muon chambers. This is because, unlike electrons, muons do not lose most of their energy via bremsstrahlung. The energy loss by bremsstrahlung is inversely proportional to the squared of the mass of the incoming charged particle. Thus, electrons lose their energy mainly via bremsstrahlung. The dominant energy loss mechanism for muons is the ionization. Using the Bethe-Bloch formula [42], it can be shown that for a given material, (relativistic) muons have one of the least energy loss via the ionisation process while passing through the material [43]. Because of this, muons easily pass through the inner detectors.

The muon detection system comprises drift tubes (DT) in the barrel and cathode strip chambers (CSCs) in the endcaps, along with resistive plate chambers (RPCs). While DT and CSC detectors offer precise muon position measurement for momentum determination, RPC chambers provide fast information for the Level-1 trigger (discussed in Section 2.3). Drift chambers are filled with an $\mathrm{Ar/CO_2}$ gas mixture and they can provide cost-effective large area coverage. When muons pass through gas detectors, the gas gets ionized and electrons drift towards wires where electric signal is generated. Spatial resolution is approximately 250 µm to 600 µm. The CSCs in the endcap are more radiation hard and can operate in non-uniform magnetic field. Electrons and positrons from ionisation are collected in the horizontal and vertical grid of wires allowing a spatial resolution of 50 µm and 150 µm.

Muons are reconstructed using data from the muon system and the inner tracker, enhancing the momentum resolution. This combined approach achieves a muon reconstruction efficiency exceeding 98% for $|\eta| < 1.6$ and a p_T-resolution below 5% up to $p_T = 1000$ GeV [44]. Additionally, the timing resolution of each muon detection system is below 3 ns, enabling the identification of individual bunch crossings crucial for the trigger system.

2.3 Trigger

CMS has a two-level trigger system, the hardware-based L1 trigger and high-level trigger (HLT) to identify interesting events and suppress background. The L1 trigger operates at 40 MHz and consists of Application-specific Integrated Circuits (ASICs) and Field Programmable Gate Arrays (FPGAs) implemented on custom boards [45]. The custom hardware helps in the parallel and rapid computation of the 'calorimeter trigger' and 'muon trigger' algorithms. The calorimeter trigger reconstructs electron, photon, tau, and jet candidates based on the information from the calorimeters. The muon trigger reconstructs muon candidates based on the information from the muon chambers. The calorimeter and muon trigger system then decides about selecting the event within 4 µs. The events passing the L1 trigger are then further reduced to about 1 kHz by the HLT trigger at a processing time of 174 ms per event. The trigger paths of the HLT are designed to identify events with specific properties and are combined via 'OR' logic. In addition to the data from calorimeters and muon chambers, the HLT also utilizes data from the tracker which allows for some basic track and vertex reconstruction. In this thesis, the HLT paths filtering events with high-energetic muons and electrons or large MET are used (discussed in Section 3.5).

2.4 CMS coordinate system

CMS detector follows the right-handed coordinate system with the positive x-axis in the direction of the center of LHC, the z-axis along the counterclockwise beam direction, and the positive y-axis pointing upwards as shown in Figure 14. $\vec{p}$ is the momentum vector and points in the direction of particle momentum while p_T is its transverse component. ϕ is the azimuthal angle and θ is the polar angle.

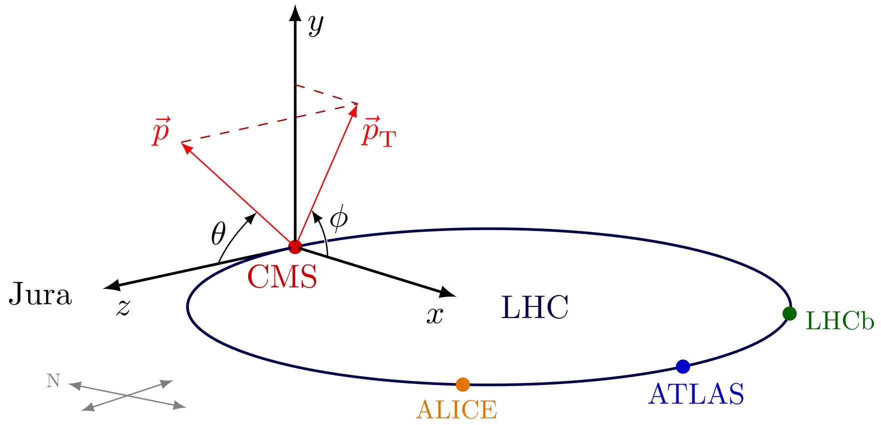


Figure 14: Right-handed coordinate system followed by the CMS detector [46]. The positive x-axis points towards the center of LHC, the z-axis along the counterclockwise beam direction, and the positive y-axis points upwards. $\vec{p}$ is the momentum vector and points in the direction of particle momentum while p_T is its transverse component. ϕ is the azimuthal angle and θ is the polar angle.

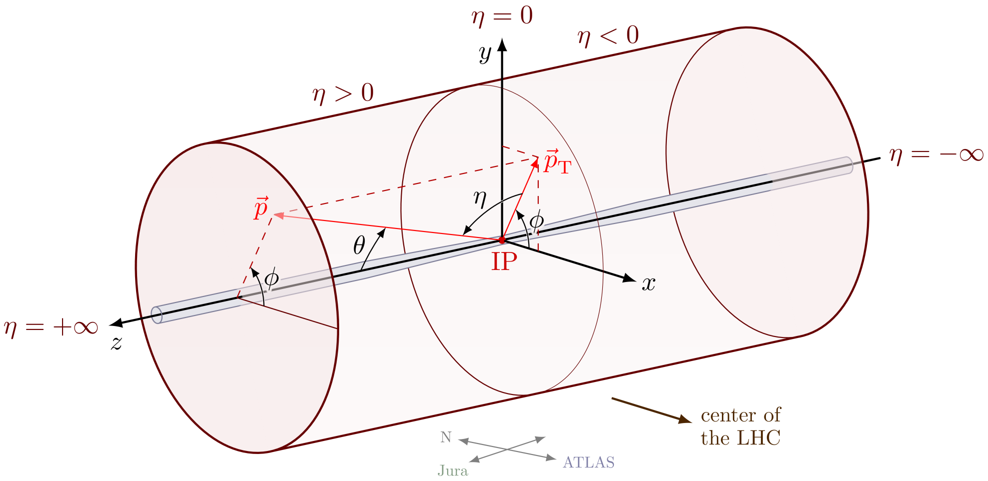


Figure 15: CMS coordinate system in the cylindrical system [46] where IP is the interaction point of the particles in the beam and η is the pseudo-rapidity as defined in the Equation 2.3

Instead of the polar angle θ, the pseudo-rapidity (η) is used and is defined by

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) = \frac{1}{2}\ln\left(\frac{|\vec{p}| + p_z}{|\vec{p}| - p_z}\right) \tag{2.3}$$

This is valid when the mass of the particles is negligible compared to their energy. For massive particles, rapidity (y) is used:

$$y = \frac{1}{2}\ln\left(\frac{E + P_z}{E - p_z}\right) \tag{2.4}$$

Both $\Delta\eta$ and $\Delta\phi$ are Lorentz invariant under boost along the beam direction. The angular distance between two objects is defined as:

$$\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} \tag{2.5}$$

The distance ΔR between two particles is also directly correlated with the invariant mass of the mother particle decaying to two products (1 → 2 decay) as given by:

$$\begin{aligned} m &= \sqrt{(E_1 + E_2)^2 - \|\mathbf{p}_1 + \mathbf{p}_2\|^2} \\ &= \sqrt{2p_{\mathrm{T},1}p_{\mathrm{T},2}(\cosh(\Delta\eta) - \cos(\Delta\phi))} \\ &\approx \sqrt{p_{\mathrm{T},1}p_{\mathrm{T},2}\left(\Delta\eta^2 + \Delta\phi^2\right)} \\ &= \sqrt{p_{\mathrm{T},1}p_{\mathrm{T},2}} \cdot \Delta R \end{aligned} \tag{2.6}$$

Figure 15 shows the CMS coordinate system in the cylindrical system where IP is the interaction point of the particles in the beam and η is the pseudo-rapidity as defined in the Equation 2.3.

Part II
VH($\rightarrow b\bar{b}$) STXS measurements

3 STXS measurements of the Higgs boson in the VH($\rightarrow b\bar{b}$) channel

Even though H→ $b\bar{b}$ is the dominant decay mode of the Higgs boson, its properties have been measured with low precision at LHC as discussed in Section 1.6. This is due to the overwhelming multi-jet backgrounds making the identification of H→ $b\bar{b}$ events difficult.

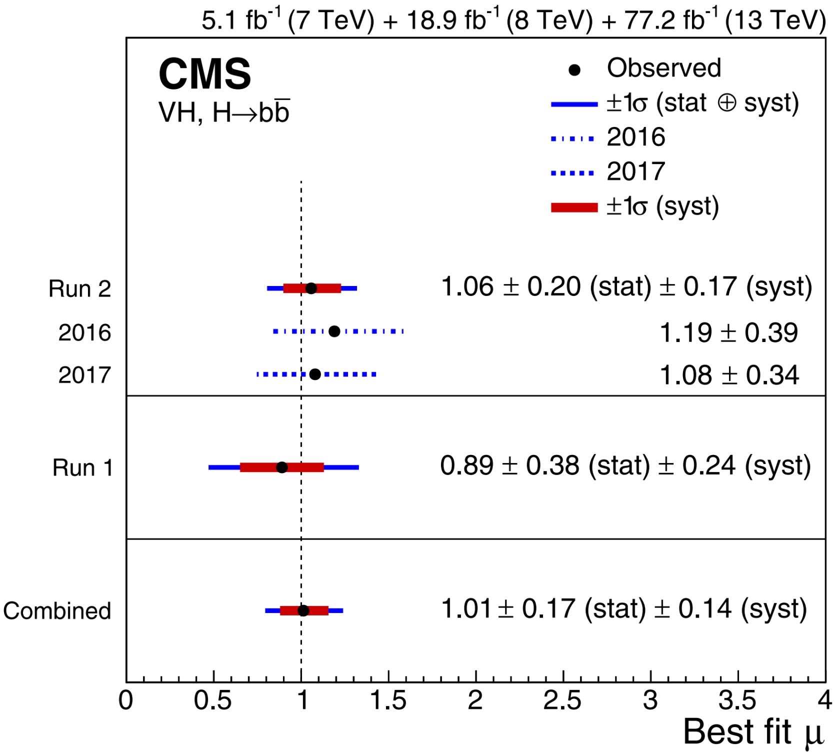


Figure 16: VH($\rightarrow b\bar{b}$) measurement by the CMS experiment using the Run 1, 2016 and 2017 data [47]. The Run 2 label in this Figure corresponds to the partial Run 2 (2016-2017).

The current measurement led to single-experiment observation of the H→ $b\bar{b}$ channel in the associate production mode of the Higgs boson where the vector boson decays leptonically (VH→ $b\bar{b}$). The advantage of targeting only the leptonic decay is that it provides a trigger signature in the form of isolated leptons or MET to identify VH→ $b\bar{b}$ events and reduce contamination from the overwhelming multijet backgrounds. The CMS Collaboration discovered VH→ $b\bar{b}$ in 2018 [47] using the combination of data from Run 1 (2010-2013), 2016 and 2017. The results of the measurement are shown in Figure 16.

The current measurements of the Higgs boson in VH→ $b\bar{b}$ channel are presented in terms of mutually exclusive regions of phase space defined in bins of the transverse momentum of the vector boson and number of additional jets, known as the simplified template cross sections (STXS) framework. Section 3.1 gives an overview of the analysis strategy used in this thesis followed by an in-depth explanation of the analysis. The focus of this thesis is mainly on the analysis of data collected in 2017 and 2018, though the final result with the combination of data collected in 2016 is discussed in Section 3.14.

3.1 Analysis overview

The analysis is optimized for measurements in the mutually exclusive regions of phase space provided by the STXS framework (described in Section 3.9) for the Higgs boson decay to a pair of bottom quarks with the associated vector boson decaying leptonically. Based on the number and flavor of leptons in the final state, the analysis is divided into 3 channels:

- 0-lepton channel: Z($\rightarrow \nu\bar{\nu}$)H
- 1-lepton channel: W($\rightarrow \mu^{\pm}\nu$)H, W($\rightarrow$ e$^{\pm}\nu$)H
- 2-lepton channel: Z($\rightarrow \mu^{\pm}\mu^{\mp}$)H, Z($\rightarrow$ e$^{\pm}$e$^{\mp}$)H

Only the leptonic (electron and muon) decay modes of the vector boson are analyzed. Tau leptons are not explicitly reconstructed, but as they decay to electron and muon with 0.35 branching fraction, a fraction of such events are reconstructed in the electron or muon channels of this analysis.

The analysis phase space in each of the channels is divided into Signal Regions (SR) and Control Regions (CRs) using kinematic cuts. Signal regions are defined to have high signal efficiency. Control regions are defined to be enriched in a particular background process. We have three control regions: TT CR (enriched in t$\bar{\text{t}}$ process), V+HF CR (enriched in V+b/bb/c jets process), and V+LF CR (enriched in V+udsg jets process). The flavor categorization of V+jets events is discussed in Section 3.3. The signal and major background processes are described in Section 3.2 and the analysis phase space (SR and CRs) is further described in Section 3.8.

The parameter estimation technique, namely the maximum likelihood fit of the model to data (discussed in Section 3.13) is performed simultaneously on all the SR and CRs to determine the 8 POIs (parameter of interest) or signal strength modifiers, each describing the ratio of cross-section of the signal process observed in data divided by the SM expectations in different STXS bins. The 8 POIs are implemented in the fit as freely floating parameters. Along with those parameters, additional nuisance parameters (parameters in the fit model other than the POIs) are used to constrain background processes in data through the likelihood fit of the model to data. The fit model and parameters used in the fit are further described in Section 3.13.

The Section 3.2 describes the signal and the major background processes in this analysis.

3.2 Signal

The VH process can be initiated by both quark and gluon-induced production modes. For quark-induced production mode, the vector boson can be both W or Z boson. For gluon-induced production mode, only the Z boson can be produced via

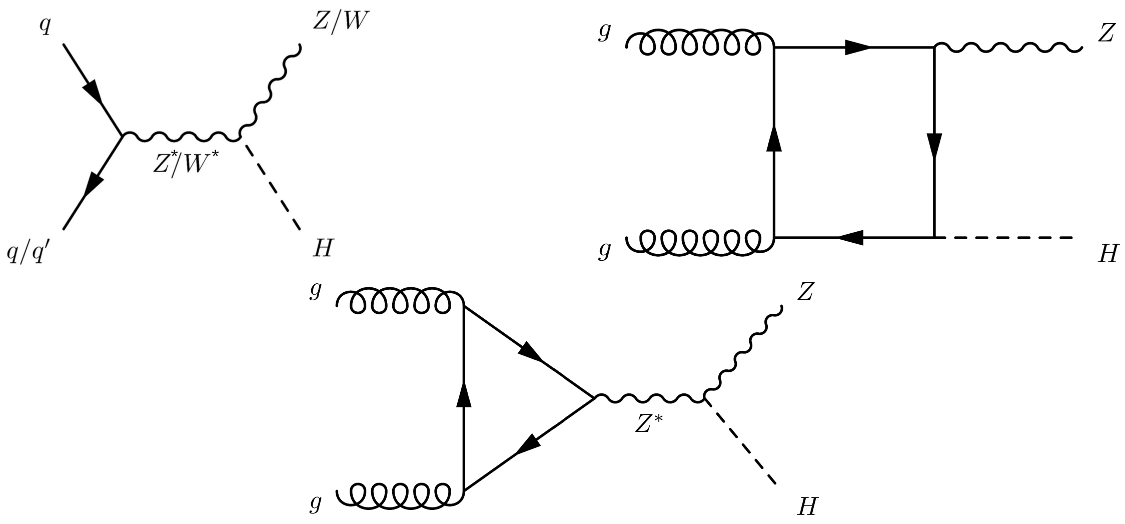


Figure 17: Feynman diagram for the signal VH process. qqVH production process (top left) and ggZH production process (top right and bottom).

a top-quark loop as shown in the box type diagram in Figure 17. This process is sensitive to Higgs-top quark coupling due to the fermionic top loop on the production side.

3.3 Backgrounds

V+jets

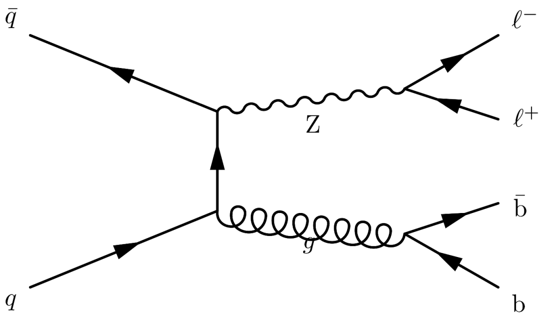


Figure 18: Feynman diagram for Drell-Yan + 2 jets.

A quark-antiquark pair can produce a vector boson plus a gluon which resembles the final state of the signal process when the gluon decays to a pair of bottom quarks (gluon split) and the vector boson decays leptonically. However, the kinematics of this process is different than that of the signal. It has a faster falling dijet mass distribution and lower transverse momentum distribution for both the dijet system and the vector boson. To reduce this background in the SR, the b-tagging score (an MVA score which can be interpreted as the probability of a jet to come from a b-quark, described further in Section 3.6.6) of the two jets is used coupled with cuts on the dijet mass and the transverse momentum of the vector boson. The remaining V+jets background in the signal region (along with the V+jets background in CRs) is modeled using MC simulation model in the fit.

The V+jets events are categorized into four flavors based on the number of B hadrons and D hadrons within the detector acceptance region of $|\eta| < 2.6$ and having $p_T > 25$ GeV. They are required to match a generator level jet within a cone of $\Delta R = 0.8$. Based on the counting of B and D hadrons, the following V+jets flavor categories are defined:

- V+udsg (light partons): no B or D hadrons
- V+c: no B hadrons, but at least one D hadron
- V+1b: exactly one B hadron
- V+2b: more than one B hadron

Figure 18 shows an example DY+jets process having two B hadrons. Thus it is classified as a V+2b process. In this thesis, processes V+1b, V+2b, and V+c are collectively known as V+HF (V + heavy flavor) process while the V+udsg is known as V+LF (V + light flavor) process.

Top quarks

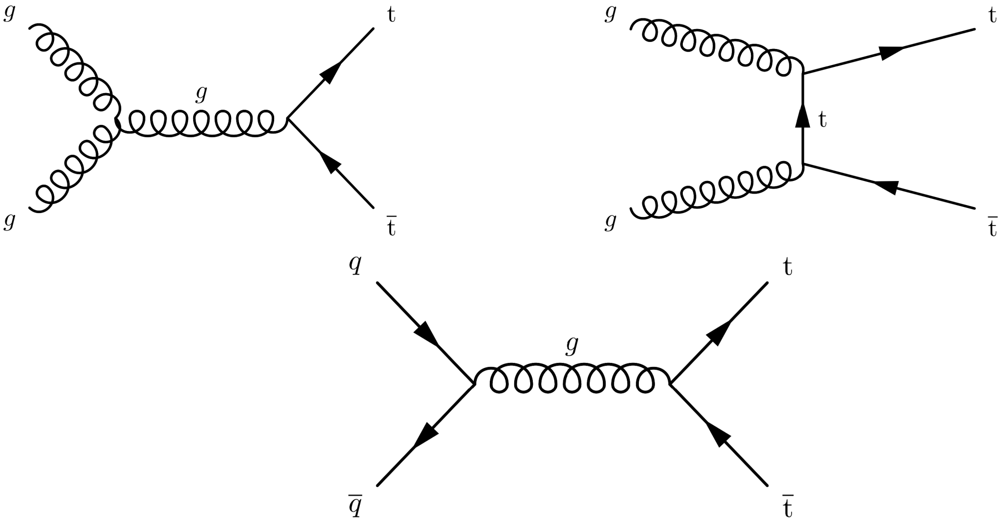


Figure 19: Feynman diagrams for $\mathrm{t\bar{t}}$ processes.

Top quarks form a dominant source of background in the 0- and 1-lepton channels. Top quarks can be produced through strong interaction in $\mathrm{t\bar{t}}$ production (mostly through gluons) or through single top production in tW (s and t channel). The Feynman diagrams for $\mathrm{t\bar{t}}$ and single top production are shown in Figure 19 and 20 respectively.

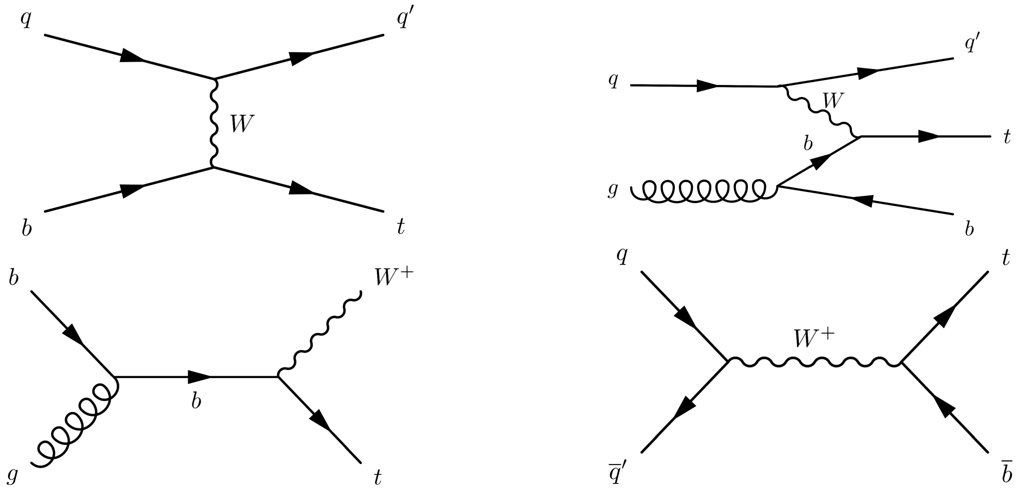


Figure 20: Feynman diagrams for single top t-channel (top left and right), s-channel (bottom left), tW (bottom right).

When both the top quarks decay leptonically, (i.e. t → Wb with W → lν) the final state is similar to the signal of 2-lepton channel Z(ll)H(bb). The contribution of this process in 2-lepton SR is reduced by requiring the mass of the dilepton system to be not equal to the mass of the Z boson. When one of the top quarks decays leptonically and the decay products of the vector boson from the other top quark are outside of the detector acceptance, the final state is similar to the signal of the 1-lepton channel W(lν)H(bb). The contribution of this process in the 1-lepton SR is reduced by requiring back-to-back vector boson and dijet pair and by reconstructing the top quark mass (discussed in Section 3.7.4). When both the top quark decays hadronically (i.e. t → Wb with W → q$\bar{\text{q}}$), the final state is similar to the 0-lepton channel Z(νν)H(bb) process. We reduce the contribution of this process in SR of 0-lepton by limiting the number of additional jets in the event.

Even though the single-top background has a very low cross-section, it contributes $10 - 20\%$ of the total top quark background. This is because it is produced via electroweak interactions and it is difficult to isolate it from the signal. The t$\bar{\text{t}}$ background is produced via QCD interactions and can be isolated from the signal process using its properties like the number of jets or dijet mass which cannot be exploited in the case of single-top due to its topology as shown in Figure 20. In 0- and 1-lepton channels, t$\bar{\text{t}}$ and single-top are estimated using a multiclass classifier in HF CR (discussed in Section 3.11).

Diboson

The diboson background produces a signal-like final state when one of the vector bosons decays to a pair of bottom quarks and the other decays leptonically. This background process can be separated from the signal using selections on the dijet pair invariant mass since for the signal, the mass of the dijet is close to the mass of the Higgs boson which is not the case for the dijet mass reconstructed from the

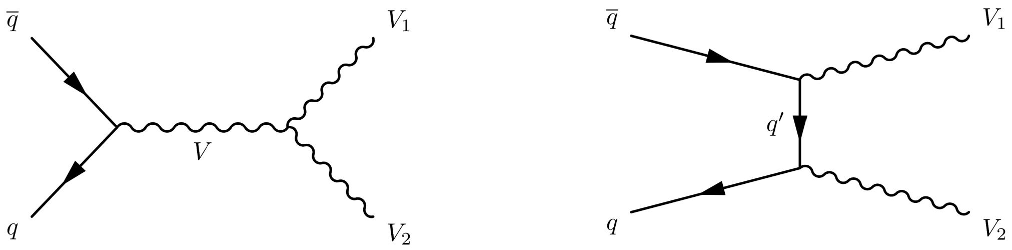


Figure 21: Diboson production in the s-channel (left) and t-channel (right). u-channel can be derived by interchanging V_1 and V_2 in t-channel.

diboson process. The diboson production in s and t-channel is shown in Figure 21.

For flavor categorization of the VV process, we consider the flavor of the generator level jets in the VV event with $p_T > 20$ GeV and $|\eta| < 2.4$. If at least one generator level b-jet is present, we categorize the VV event to be VVHF, i.e. one of the vector boson from VV decays into heavy-flavored jets. Otherwise, we categorize the VV event as VVLF, i.e. one of the vector boson decays into light-flavored jets.

QCD

The gluon splitting due to strong interactions in QCD multi-jet events can produce a pair of bottom quarks leading to background events for this analysis. When the energy of a jet is mismeasured and is opposite to a bottom quark pair in the transverse plane, it can lead to unphysical MET and a topology similar to the signal in the 0-lepton channel. Due to the presence of non-prompt leptons from hadron decays or other object mis-identification, QCD can contribute to the background of the 1- and 2-lepton channels as well.

Even though the QCD processes have a very high cross section compared to other background processes, the fraction of QCD events passing the selection cuts of the analysis is negligible (discussed in Section 3.8.1). This implies that large QCD samples must be simulated in order to have sufficient number of events passing the final selection. Since such large simulated samples of QCD are not available for this analysis, the analysis selection cuts are optimized such that the contribution of QCD events in the analysis regions is negligible. Since MET in QCD (MET due to mis-measured jet energy) is less back-to-back (opposite) to the pair of b quarks, selection cuts on the difference between the azimuthal angle of MET and the Higgs candidate b-jets can be used to reduce QCD contribution in 0-lepton (described further in Section 3.8.1). In 1 and 2-lepton channels, QCD contribution is minimized by using lepton isolation and identification cuts since the leptons in QCD are always non-prompt while leptons from signal events are prompt (discussed in Section 3.6.4).

3.4 Simulation of signal and background processes

The simulation of MC events follows the procedure described in Section 1.7. The simulated events are generated in a factorized approach. Event generators like MadGraph [48][49], POWHEG [50], and aMC@NLO (or MadGraph5_aMC@NLO)[51] simulate the hard scattering process (matrix elements, ME). Different generators differ in the procedure for calculating real and virtual process amplitudes [52]. Further, having effective luminosity of event generators larger than data helps in having effective event weight < 1, thus increasing statistical precision. The output of the event generator is then fed to parton shower (PS) models such as PYTHIA8 [15] for the PS and hadronization. Merging and Matching schemes are further implemented to avoid double counting of jets between the hard scattering simulation and the parton shower (for example, the extra jet radiation can be originated from the ME or the PS). In the matching scheme, ME and/or PS are modified to fit them together [53] while in the merging scheme a so-called merging scale is defined above which we generate with exact matrix elements and then add parton shower below [54].

These events are then passed through the CMS detector simulation implemented in GEANT4 [55] which outputs energy deposits in various parts of the detector using the 4-vectors of the particles from the generation step. The digitization step then uses this information to simulate detector hits. This is followed by a mixing step where minimum-bias events (inelastic proton-proton collisions) are added to account for the pile-up. Simulated events (as well as data) are reconstructed using the Particle flow algorithm (described in Section 3.6). For all the simulated samples, the NNLO NNPDF3.1 [56] set is used to model the parton density functions (PDF). In this methodology, the parameters that define the shape of PDF are determined through global fits using experimental data, higher-order perturbative calculations in both QCD and QED/electroweak theory, and a statistical framework dealing with aspects such as the PDF parametrization and their uncertainty estimate and propagation. Further details can be found in [56]. Parton showering and hadronization are modeled with Pythia 8.212 [57]. CUETP8M1 tune [17] is used for 2016 and CP5 tune [18] for 2017 and 2018 for UE description.

The full list of simulated samples used in 2017 and 2018 are given in Appendix Section A.

Signal samples

Differential cross sections for WH and ZH signal processes have been computed in [58]. The quark-induced signal samples used in the analysis are generated using POWHEGv2 event generator up to NLO in QCD using the MiNLO procedure [59] which merges the matrix-element calculation and the parton shower at NLO. The gluon-induced signal samples are generated using POWHEGv2 at LO. The contribution of the gluon-induced ZH process to the cross section of the ZH process is around 6% [60]. In all the simulation samples, the mass of the Higgs boson is assumed to be 125 GeV. The cross sections are then rescaled to an inclusive cross section up to NNLO [61] in QCD and NLO in EWK. Since up to NLO, electroweak corrections can be factorized, they are derived and applied differentially in p_T of the vector boson. The electroweak corrections are described further in Section 3.10.1.

Table 1 lists the simulated signal samples generated with POWHEG+PYTHIA8 and their cross-section. As shown, the dominant signal in terms of cross-section is WH.

Sample	Cross-section (pb)
$\mathrm{W^-H, H \to b\bar{b}, W \to l\nu}$	0.17202
$\mathrm{W^+H, H \to b\bar{b}, W \to l\nu}$	0.10899
$\mathrm{q\bar{q} \to ZH, H \to b\bar{b}, Z \to \nu\nu}$	0.04718
$\mathrm{gg \to ZH, H \to b\bar{b}, Z \to \nu\nu}$	0.01437
$\mathrm{q\bar{q} \to ZH, H \to b\bar{b}, Z \to ll}$	0.00720
$\mathrm{gg \to ZH, H \to b\bar{b}, Z \to ll}$	0.09322

Table 1: Simulated signal samples generated with POWHEG and PYTHIA8 and their cross-section.

V+jets samples

V+jets samples for 2017 and 2018 are generated using the aMC@NLO event generator [51] at NLO, PYTHIA8 for PS and FxFx [54] merging. aMC@NLO event generator incorporates NLO fixed-order QCD calculation. The FxFx merging technique combines multiple NLO and PS samples with different final state jet multiplicities. Around 1/3 of the NLO events of V+jets samples used in 2017 and 2018 have negative weights. Since the statistical power is reduced by a factor of $(1-2f)^2$ (where f is the fraction of events with negative weight) in these cases, we use larger datasets for NLO V+jets samples to cover up for the loss in statistical precision compared to the available LO samples.

For 2016, V+jets are generated using MadGraph [48][49] at LO accuracy with MLM matching [53]. MLM matching scheme is implemented in PYTHIA8 to avoid double counting of jets as discussed before. For 2017 and 2018, LO V+jets samples were centrally produced with bugged PDF weights settings in MadGraph [62]. The effect of this bug was observed in both vector boson and jet kinematics, most importantly the p_T of the jets. The bug was reported to cause an increase in cross-section of V+jets samples by 15%-50% [62]. The p_T spectrum of jets was found to be harder than the correct LO V+jets samples and was not covered by the QCD scale uncertainties [62]. In light of the issue affecting the LO V+jets samples in the 2017 and 2018 era productions, the NLO V+jets samples were used in 2017 and 2018. Moreover, due to lower statistics of the NLO V+jets samples, the LO V+jets samples were used for training MVA variables used in the fit model (discussed in Section 3.11). The cross sections are further reweighted to NLO in EWK (discussed further in Section 3.10.1). For 2016, LO V+jets samples' cross section is scaled to NLO using the $\Delta\eta$(bb) LO to NLO reweighting discussed in Section 3.10.1 and further to NNLO in QCD [63] derived by comparing the inclusive cross section at NNLO and NLO.

2016 LO V+jets samples are generated in exclusive bins of the scalar sum of the transverse parton momentum (H_T) in addition to two sets of b-enriched samples (one set requiring at least one b-quark from hard scattering and the other requiring at least one b quark in parton shower irrespective of the presence of b-quarks from the hard scattering process). The b-enriched samples are produced in bins of generator level vector boson transverse momentum. For 2017 and 2018, NLO V+jets samples are generated in bins of the transverse momentum of the vector boson and/or number of b-quarks from hard scattering.

The set of V+jets samples in the individual years are then stitched together by weighing them by additional weights (known as the stitching weights) to have the same effective luminosity. For example, the phase space region $A \cap B$ is the largest possible subset common between sample A and sample B, having m and n unweighted events respectively in the common phase space. In this case, the events in the common phase space of sample A are reweighted by m/m+n and those of the sample B by n/m+n. This stitching procedure is applied to the V+jets samples of all three years. Figure 22 shows the distribution for generated H_T distribution for Z+jets in Z+HF CR after applying the stitching weights in 2016. The smoothness and continuity of the distribution is a measure of the correctness of the stitching procedure.

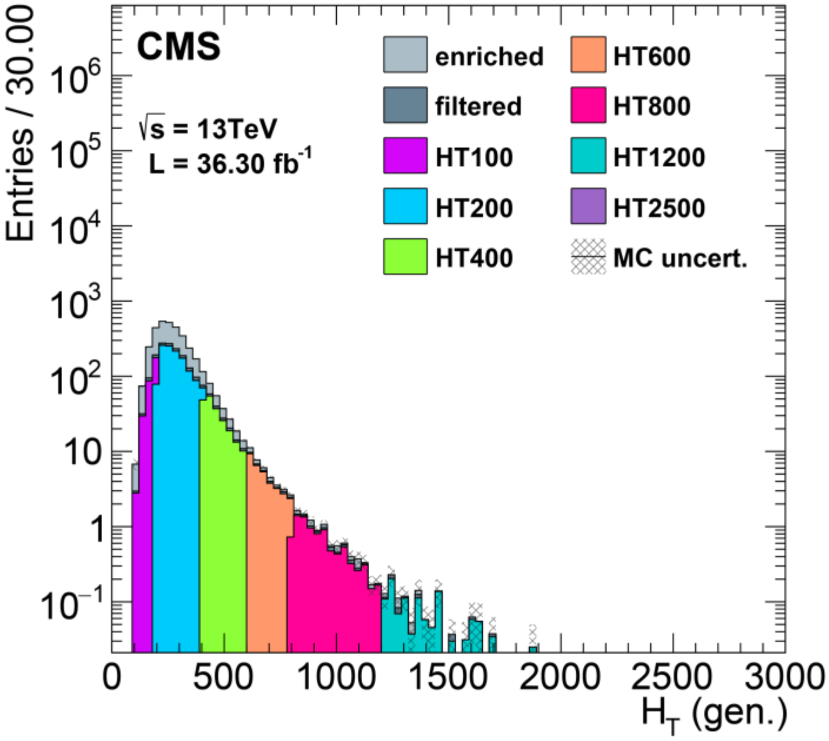


Figure 22: Distribution for generated H_T distribution for Z+jets in Z+HF CR after applying stitching weights in the 2016 dataset. The samples in the legend represent various H_T binned samples, for example, HT100 represents the 100-200 GeV H_T binned sample. The label 'enriched' represents the sample generated with the requirement of at least one b-quark from hard scattering while 'filtered' represents the sample generated with the requirement of at least one b-quark in the parton shower irrespective of hard scattering.

Table 2 shows the cross sections for some of the V+jets samples used in 2018. The W+jets samples in the Table are split by the number of b-quarks from hard scattering while Z+jets samples are split across the transverse momentum of vector boson and number of b-quarks from hard scattering.

Sample	Cross-section (pb)
$\mathrm{Drell-Yan, 50GeV < m_{ll}}$	0.17202
$\mathrm{W+Jets}$, 0 B hadron	54500.0
$\mathrm{W+Jets}$, 1 B hadron	8750.0
$\mathrm{W+Jets}$, 2 B hadrons	3010.0
$\mathrm{Z+Jets}$, 1 B hadron, $\mathrm{50GeV < m_{ll}, 50GeV < p_T(Z) < 150GeV}$	596.3
$\mathrm{Z+Jets}$, 2 B hadrons, $\mathrm{50GeV < m_{ll}, 50GeV < p_T(Z) < 150GeV}$	325.7
$\mathrm{Z+Jets}$, 1 B hadron, $\mathrm{50GeV < m_{ll}, 150GeV < p_T(Z) < 250GeV}$	17.98
$\mathrm{Z+Jets}$, 2 B hadrons, $\mathrm{50GeV < m_{ll}, 150GeV < p_T(Z) < 250GeV}$	29.76
$\mathrm{Z+Jets}$, 1 B hadron, $\mathrm{50GeV < m_{ll}, 250GeV < p_T(Z) < 400GeV}$	2.045
$\mathrm{Z+Jets}$, 2 B hadrons, $\mathrm{50GeV < m_{ll}, 250GeV < p_T(Z) < 400GeV}$	5.166
$\mathrm{Z+Jets}$, 1 B hadron, $\mathrm{50GeV < m_{ll}, 400GeV < p_T(Z)}$	0.2243
$\mathrm{Z+Jets}$, 2 B hadrons, $\mathrm{50GeV < m_{ll}, 400GeV < p_T(Z)}$	0.8457

Table 2: Simulated V+jets samples with their cross sections for 2018. W+jets samples are split by the number of b-quarks from hard scattering while Z+jets samples are split across the transverse momentum of vector boson and number of b-quarks from hard scattering.

Sample	Cross-section (pb)
$\mathrm{WW, W \rightarrow l\nu, W \rightarrow q\bar{q}}$	51.87
$\mathrm{WZ, W \rightarrow l\nu, Z \rightarrow q\bar{q}}$	10.875
$\mathrm{WZ, W \rightarrow q\bar{q}, Z \rightarrow ll}$	6.331
$\mathrm{ZZ, Z \rightarrow q\bar{q}, Z \rightarrow ll}$	2.387
$\mathrm{ZZ, Z \rightarrow q\bar{q}, Z \rightarrow \nu\nu}$	4.726

Table 3: Diboson samples with their cross sections for 2018.

Diboson samples

Diboson samples are generated using the aMC@NLO generator at NLO with FxFx merging scheme except for the 0-lepton ZZ process and WW process which are generated at LO using POWHEGv2. In the previous analysis [47], all diboson samples were simulated only at LO. The usage of NLO simulation improves the modelling of kinematic observables of the diboson process. Measured cross sections are used for diboson events [64] and are given in Table 3.

Single top and $\mathrm{t\bar{t}}$ samples

aMC@NLO generator is also used for single-top s-channel while $t\bar{t}, tW$ and single-top t-channel is generated using POWHEGv2. Table 4 gives the cross section for simulated $\mathrm{t\bar{t}}$ and single top samples for 2018.

Sample	Cross-section (pb)
$\mathrm{t\bar{t}, t \to bl\nu, t \to bl\nu}$	88.29
$\mathrm{t\bar{t}, t \to bl\nu, t \to bq\bar{q}}$	365.34
$\mathrm{t\bar{t}, t \to bq\bar{q}, t \to bq\bar{q}}$	377.96
Wt	35.85
$\mathrm{W\bar{t}}$	35.85
$\mathrm{t(s-channel)}$	3.692
$\mathrm{t(t-channel)}$	136.02
$\mathrm{\bar{t}(t-channel)}$	80.95

Table 4: Simulated $\mathrm{t\bar{t}}$ and single top samples with their cross sections for 2018.

QCD samples

H_T binned QCD samples $(100-200, 200-300, 300-500, 500-700, 700-1000, 1000-1500, 1500-2000, > 2000$ GeV) are generated at LO using aMC@NLO generator with the MLM matching scheme.

3.5 Trigger

The triggers used in this analysis focus on the identification of events with one or two highly energetic electrons or muons (1-lepton channel and 2-lepton channel) or large MET (0-lepton channel).

For the 2-lepton channel, the trigger threshold on p_T of the leading/sub-leading electrons is 23/12 GeV and muons are 17/8 GeV. Since only one hard lepton per event is required in the 1-lepton channel, the trigger threshold on p_T of lepton is higher for the 1-lepton channel than the 2-lepton channel. For the 1-lepton channel, the trigger threshold on p_T of the lepton is 32 GeV for electron and 24 GeV for muon. 0-lepton channel events are triggered using dedicated MET and MHT (missing hadronic transverse momentum, discussed in Section 3.6.8) triggers. The list of HLT triggers as well the datasets used in this analysis for the 2017 and the 2018 datasets is given in Table 5 and 6 respectively.

Channel	dataset	HLT paths
Z(ee)H	/EGamma	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL
Z(μμ)	/DoubleMuon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8
W(eν)	/EGamma	HLT_Ele32_WPTight_Gsf
W(μν)	/SingleMuon	HLT_IsoMu24
Z(νν)	/MET	HLT_PFMET120_PFMHT120_IDTight

Table 5: Datasets and corresponding Triggers used for the 2018 analysis. The different decisions are combined through a logic OR.

Channel	dataset	HLT paths
Z(ee)H	/DoubleEG	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL
Z(μμ)	/DoubleMuon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8
		HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8*
W(eν)	/SingleElectron	HLT_Ele32_WPTight_Gsf_L1DoubleEG
W(μν)	/SingleMuon	HLT_IsoMu27
Z(νν)	/MET	HLT_PFMET120_PFMHT120_IDTight
		HLT_PFMET120_PFMHT120_IDTight$_PFHT60$

Table 6: Datasets and corresponding Triggers used for the 2017 analysis. The different decisions are combined through a logic OR.

3.6 Object Reconstruction

The information from various sub-detectors is combined using the Particle Flow (PF) algorithm [65] to identify particles like photons, charged hadrons, neutral hadrons, muons, and electrons. In the following, each PF object will be described:

3.6.1 Tracks

Charged particles (leptons and hadrons) follow a curved trajectory when passing through the detector due to the solenoid magnet. The track is characterized by 5 parameters: transverse and longitudinal impact parameters (d_0 and z_0), azimuthal angle (ϕ_0), polar angle (θ_0), and transverse momentum of the track (p_T) at the point of closest approach. The transverse impact parameter at the primary vertex is the distance in the xy-plane between the interaction point and the track at the point of closest approach.

Following is a brief description of the steps for track reconstruction in the CMS detector:

1. Track seeding: Based on the clusters of hits in the pixel detector and constraints (such as the position of the centre) related to the reconstructed beam-spot [66], track seeds are identified. The track seeds allow to estimate the track parameters mentioned above.

2. Track fitting: The Kalman filter algorithm [67] is then used to fit the track seeds to the charged particle's curved path adding new hits on the path. Multiple scattering, inhomogeneous magnetic field, and energy loss of the particle are taken into account in the fit. This re-fitting of the track helps to refine the track parameters from step 1 and to reduce any possible bias from the beam-spot constraint in the track seeding step.

3. Track selection: Tracks are then filtered based on their quality, goodness of fit, and number of hits in the detector.

4. Track association: Using information from other detector components, tracks are then identified and associated with the possible reconstructed objects. Tracks

are then used to reconstruct primary and secondary vertices.

3.6.2 Primary vertices

After track reconstruction, the primary vertices are reconstructed in the following steps:

1. Track selection: reconstructed tracks compatible with the beam spot are selected.

2. Track clustering: based on the proximity of the hits, tracks are clustered using a simulated annealing algorithm [68].

3. Vertex fitting: a fit is performed to the clusters to determine the position of the primary vertex. The vertex with the highest sum of squares of the transverse momenta is labeled as the primary vertex (PV) while other reconstructed vertices are referred to as PU vertex.

The efficiency of correctly identifying PV is approximately 70%. The particles associated with the PV are called prompt.

3.6.3 Secondary vertices

First seed tracks are identified as tracks having high impact parameter significance. For every seed track, tracks compatible to be originating from the same vertex as the seed track are clustered. The compatibility is computed using metrics like separation distance in three dimensions. The track clusters are then fitted using the adaptive vertex fitter (AVF) algorithm [69]. The AVF algorithm is a non-linear least square fit algorithm developed by CMS and relies on the Kalman filter algorithm. Usually, a B hadron has three or more tracks originating from the secondary vertex.

3.6.4 Leptons

To select prompt leptons (leptons from primary interactions) and reject those from electroweak decay of bosons, from the heavy flavor jets, or the decay in flight of charged pions and kaons, threshold on the isolation of electrons (muons) in the (η, ϕ) plane is used. The lepton relative isolation is defined as the total p_T of particles within ΔR cone of 0.4 (0.3) for electrons (muons) as

$$I_{\mathrm{PF,\,rel.}} \equiv \frac{1}{p_{\mathrm{T}}^{\ell}} \left(\sum p_{\mathrm{T}}^{\mathrm{charged}} + \max\left[0, \sum p_{\mathrm{T}}^{\mathrm{neutral}} + \sum p_{\mathrm{T}}^{\gamma} - p_{\mathrm{T}}^{\mathrm{PU}}(\ell)\right]\right) \tag{3.1}$$

$\sum p_{\mathrm{T}}^{\mathrm{charged}}$ is the scalar sum of the transverse momentum of the charged hadrons originating from the same primary vertex of the event, $\sum p_{\mathrm{T}}^{\mathrm{neutral}}$ and $\sum p_{\mathrm{T}}^{\gamma}$ are the transverse momentum sum of neutral hadrons and photons respectively. The last term in the equation, $p_{\mathrm{T}}^{\mathrm{PU}}(\ell)$ accounts for the energy deposits of the pileup. This analysis uses a relative isolation threshold of 0.06 for both electrons and muons in the 1-lepton channel whereas it is 0.15 (0.25) for electrons (muons) in the 2-lepton channel. Since only one prompt lepton is expected in the 1-lepton channel compared to two in the 2-lepton channel, tighter isolation cuts are used in the 1-lepton channel. Further, usage of these isolation cuts also helps in reducing the QCD process in analysis regions to a negligible level.

Electrons Electron tracks are reconstructed using the Gaussian Sum Filter (GSF) algorithm [70] which uses hits in the silicon tracker and the pixel detector. The energy of the electrons is reconstructed using ECAL clustering algorithms [71] which reconstruct the energy of the electron by aggregating the energy deposits in clusters in super-clusters (groups of clusters in broad η and narrow ϕ windows) while mitigating the noise and pile-up contributions. Electrons are preselected by requiring them to be compatible with the primary vertex of an event ($d_{xy} < 0.05$ cm, $d_z < 0.2$ cm). To reduce the effect of fake electrons, electrons are required to pass an MVA discriminant threshold (working point), the lepton identification (ID) threshold [72]. The Loose working point (WP) corresponding to the 90% expected selection efficiency is used for the global event selection, for counting of additional leptons for veto requirement, and to count the electron multiplicity in the 2-lepton electron channel. For the 1-lepton electron channel, a Tight WP corresponding to the 80% expected selection efficiency is used to reduce background from fake leptons. The lepton ID and the isolation cuts are loosened in the 2-lepton with respect to the 1-lepton channel since requiring two leptons, already eliminates most of the QCD background. The p_T threshold on the electron in the 1-lepton channel is 30 GeV while for the 2-lepton channel, the p_T threshold for the leading (sub-leading) electrons is 25 GeV (17 GeV). Efficiency scale factors (the ratio of efficiency of data and MC) corresponding to the various MVA WPs, isolation selection, and trigger account for the differences in the efficiency of data and MC and are applied to the MC simulation. As discussed in Section 3.6.6, the tag and probe approach is used to measure efficiency SF, where Z→ ee process is used requiring one of the electron (tag electron) to pass tight selection criteria, while the efficiency is measured using the other unbiased probe electron.

Muons Muons are reconstructed using hit information from muon chambers. They are then preselected by requiring them to be compatible with the primary vertex of an event ($d_{xy} < 0.5$ cm, $d_z < 1.0$ cm). To reduce the number of fake muons, a cut-based identification (lepton ID) is used. The algorithm is based on the cuts on the χ^2 of global tracks, the impact parameter of the track and the number of muon-chamber, tracker, and pixel hits. Similar to the electron selection, a Tight WP lepton ID cut is used in the 1-lepton channel and a Loose WP lepton ID threshold is used in the 2-lepton channel. The p_T threshold on the muon in the 1-lepton channel is 25 GeV while for the 2-lepton channel, the p_T threshold for the leading (sub-leading) electrons is 25 GeV (15 GeV). As for the electrons, efficiency scale factors corresponding to the various MVA working points, isolation selection, and trigger are applied to the MC simulation. These efficiency scale factors are calculated with the tag and probe approach using Z→ $\mu\mu$ events.

3.6.5 Jets

The stable (mean lifetime $> 0.3 \times 10^{-10}$s) colorless particles from hadronization of quarks and gluons are clustered together to form jets. The energy fraction of jets is mainly dominated by charged hadrons and photons along with smaller fractions of neutral hadrons, charged particles from pile-up, and leptons. This analysis uses jets reconstructed with the anti-kt algorithm [73]. This algorithm is infrared and collinear safe. An infrared safe algorithm yields the same set of jets after modifying an event to add a soft radiation. An algorithm is a collinear safe algorithm if the final set of jets is not changed after introducing a collinear splitting of one of the inputs. Tracks in a jet that cannot be associated with the primary vertex are subtracted to account for in-time pile-up. This is known as charged hadron subtraction (CHS) algorithm. CHS jets are used in the analysis.

Jet energies do not match their corresponding parton energies as the measurement is strongly impacted by the PU and the detector response. The jet energy corrections (JECs) are applied sequentially to data and simulation to account for this mismatch. There are two levels of corrections for jets in MC simulation and data: a correction for pileup and electronic noise; and a correction for the response of the detector as a function of jet p_T and η. An additional residual correction is applied only to jets in data, to account for differences between data and MC [74]. These JECs are derived from simulation studies so that the average measured energy of jets becomes closer to that of particle level jets. In-situ measurements of the momentum balance in dijet, photon+jet, Z+jet, and multijet events are used to determine any residual differences between the jet energy scale (JES) in data and in simulation, and corrections are made [75]. The corrected energy is used in the measured and simulated jet p_T distributions of this thesis.

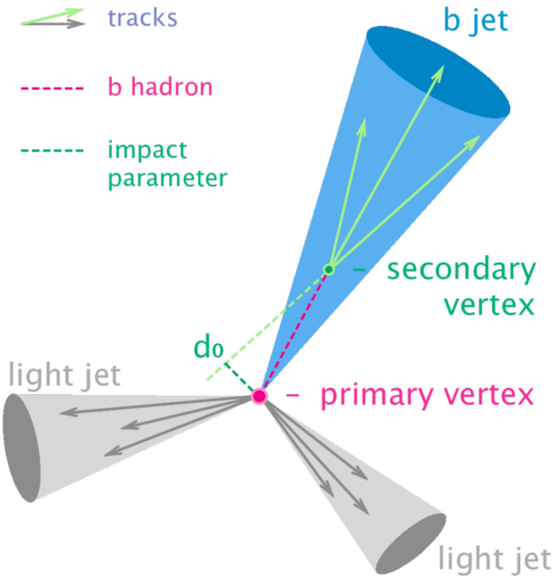


Figure 23: Primary vertex (PV), secondary vertex (SV), and the impact parameter d_0 for a b-jet [76].

The b-jets because of the large bottom quark mass are different from the light jets and have higher transverse momentum hadrons relative to the jet axis. This causes the b-jets to be relatively wider (larger ΔR) than the light jets. Also, they have higher daughter multiplicity (about five charged tracks per decay) and may contain soft leptons (low p_T leptons compared to p_T of the jet) with momentum perpendicular to the jet. Further, B hadrons have a relatively long lifetime. The lifetime of a B hadron is about 1.5 ps and thus it travels a measurable distance before decaying. The tracks of its charged daughter particles create a displaced vertex (secondary vertex). The transverse impact parameter d_0 represents the closest distance between a particle track and the primary interaction vertex. The measured uncertainty of d_0 is denoted as $\sigma(d_0)$, and the significance is expressed as $S(d_0) = d_0/\sigma(d_0)$. The sign of d_0 is positive (negative) when the track intersects the jet axis ahead of (behind) the primary vertex. Generally, the sign of d_0 is positive for most tracks associated with b-jets due to the B hadron lifetime. The primary vertex (PV), secondary vertex (SV), and the impact parameter for a b-jet are shown in Figure 23. The silicon tracking detectors allow to reconstruct these secondary vertices.

The bottom quarks generally decay weakly to a c-quark or a u-quark along with a virtual W^{*-} boson which can decay to a lepton-antineutrino pair (semi-leptonically) or to quark-anti quark pair which further undergoes hadronization. As for the semi-leptonic decays, the branching fraction of $b \rightarrow Xl\nu$ is about 11% while for $b \rightarrow Xc$ is $> 80\%$, the c quark then decays to $Xl\nu$ with a branching fraction of about 30%.

3.6.6 Jet flavor tagging

For MC events, the jet flavor is determined by matching generated and reconstructed jets within $\Delta R < 0.30$. If the generator jet has a B hadron, it is labeled as a b-jet. In case it has a D hadron, it is labeled as c-jet and the remaining jets are labeled as light (udsg) jets.

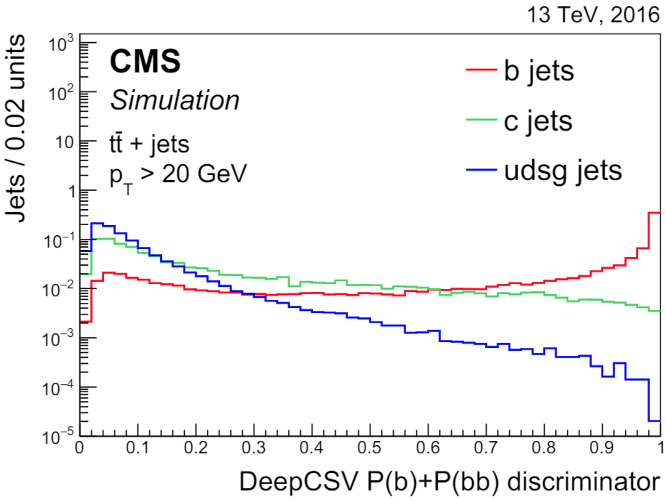


Figure 24: Normalized discriminator distribution of P(b)+P(bb) output nodes of DeepCSV for light (udsg),c and b-jets in $t\bar{t}$ samples [77].

To separate b-jets from other jets, we used a DNN-based multiclass tagger called

DeepCSV [78]. Its architecture consists of 4 layers of a feed-forward network. It uses tracks, secondary vertex observables, and global variables as input features and provides the probability for a jet to belong to each of the 5 classes: one B hadron, at least two B hadrons, one D hadron and no B hadron, at least two D hadrons and no B hadron, and no D and B hadrons. The probabilities for the two classes, one B hadron and at least two B hadron are summed together (i.e. P(b)+P(bb)). They represent the combined b-tag score of a jet which is used in the analysis. Figure 24 shows the normalized discriminator distribution for the sum of P(b) and P(bb) output nodes of DeepCSV for the light-jets, c-jets, and b-jets in a $t\bar{t}$ sample [77]. As shown, the heavy flavor jets have a higher probability of having a score close to 1 while light flavor jets have a higher probability of having a score close to 0.

Different thresholds (working points s_{wp}) on the classifier score (s) are used to quantify the performance of the classifier.

$$\frac{N_{udsg}\left(s > s_{\mathrm{wp}}\right)}{N_{udsg}} \in [0,1] \tag{3.2}$$

where N_{udsg} is the number of light jets. Three working points for the combined b-tag are used in CMS, based on the light-flavor mistag rate as shown in Table 7.

Working point	Light-flavour jet mistag rate
Loose	10%
Medium	1%
Tight	0.1%

Table 7: DeepCSV WP labeled based on the light-flavor mistag rate.

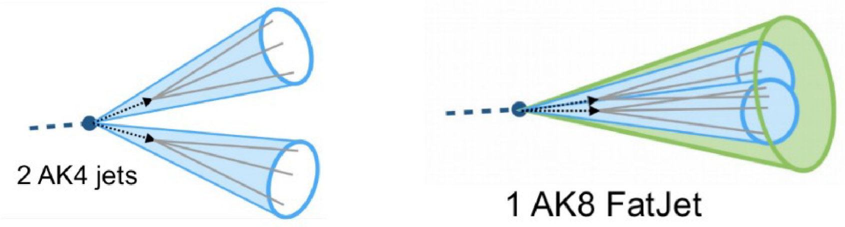


Figure 25: Resolved topology (left) and boosted topology (right).

The events in this analysis are categorized into two categories: resolved and boosted. When the leading and subleading jets can be reconstructed using the anti-kt algorithm as two separate jets each having $R = 0.4$, the event is categorized as resolved event and the jets as resolved jets (AK4 jets). Also, the corresponding event topology is referred to as resolved topology and is shown in Figure 25 (left).

In the high p_T range, the two back-to-back b-jets in the rest frame of the high momentum Higgs boson can no longer be separated into AK4 jets. It can be shown for two body decay events that this effect occurs when $p_T > m_H/R \approx 310$ GeV.

These events are categorized as boosted events, and the leading and sub-leading jets are reconstructed as one 'fat' jet with a radius parameter $R = 0.8$ (AK8 jet). The corresponding event topology is referred to as boosted topology and is shown in Figure 25 (right).

To categorize the boosted jets, we use the DeepAK8 tagger [79]. This is a multiclass DNN classifier that classifies AK8 jets based on possible hadronic decays of H/Z/W/t, and other categories depending on their decay channel. We will be interested in particular in the Hbb output node corresponding to the Higgs boson to bottom quark pair decay. DeepAK8 tagger uses up to 100 particles per jet with each jet having 42 low-level features such as p_T, charge, energy deposits, angular observables, and 7 secondary vertex features. The network architecture consists of 14 1-dimensional CNN (Convolution Neural Network) layers followed by feed-forward layers. Furthermore, to avoid mass sculpting, an adversarial loss function is used to prevent the model from learning the AK8 jet mass.

Scale factors for the AK4 DeepCSV tagger are available in p_T and η bins for different working points of each jet flavor. They are derived using an iterative fit procedure which calculates data/MC SF using a tag and probe approach. The tag and probe approach is a data-driven technique to measure efficiencies, where the semi-leptonic $\mathrm{t\bar{t}}$ events are identified in data (and MC) by requiring tight selection criteria on one of b-jets (known as tag) and leaving the other unbiased b-jet (known as probe) to be used to measure the efficiency of the selection criteria in data and MC. The ratio of the data and MC efficiencies gives the calibration SF for the tagger.

For the DeepAK8 Hbb output node, custom SFs are made available for the DeepAK8 score binning of [0, 0.8, 0.97, 1.0] differentially in bins of jet momentum. ($p_T : [200-300, 300-400, 400-500, 500-600, > 600]$). This specific DeepAK8 score binning was chosen to correspond to the selection cut boundaries for the boosted analysis regions used in this analysis optimized for signal sensitivity (discussed in Section 3.8.4). The gluon split to a bottom quark pair decay process was used as a proxy for the Higgs boson to AK8 b-jet decay. These custom SFs were used only for AK8 b-jet for boosted topology of the signal. For AK8 jets of background processes in the boosted topology, the efficiency SF was estimated in the fit model using freely-floating rate parameters associated to the background processes (discussed in Section 3.13.5).

3.6.7 Jet ID

To reduce fake jets, a cut-based jet ID [80] Tight cut is used in the analysis. The cuts are based on the expected jet constituents fractions and help to remove 98% of noise jets from the calorimeter and have a jet efficiency of 99% in the central rapidity region.

3.6.8 Missing transverse energy

The amount of energy expected from energy conservation that is not detected in the detector is referred to as the missing energy. This can be due to the so-called true missing energy accounted by the undetectable particles (i.e. neutrinos) or due to the so-called instrumental missing energy accounted by limited detector acceptance/efficiency, detector defect, or mis-reconstruction. As the longitudinal information of the colliding particles is not available at hadron colliders, only the transverse component of the energy of particles is used to compute the total transverse missing energy/momentum. The missing transverse energy is a vector defined as:

$$\vec{E}_T^{miss} = - \sum_{\text{particles}} \vec{p}_T \tag{3.3}$$

and its modulus:

$$\left| \vec{E}_T^{miss} \right| = E_T^{miss} \tag{3.4}$$

is what we will refer to as the missing transverse energy in this analysis. Two transverse missing energy reconstructions are used in this analysis:

1. Type 1 PF MET: First, raw PF MET is computed using the particles reconstructed using the particle flow reconstruction:

$$\begin{aligned} E_T^{miss} &= \left| - \left(\sum_{\text{PF candidates}} \vec{p}_T \right) \right| \\ &= \left| - \left(\sum_{\text{jets}} \vec{p}_T^{\text{ corr}} + \sum_{\text{e, } \mu} \vec{p}_T + \sum_{\text{unclustered PF cand.}} \vec{p}_T \right) \right| \end{aligned} \tag{3.5}$$

   where $\sum_{\text{jets}} \vec{p}_T^{\text{ corr}}$ is the sum of JEC corrected p_T of all the PF reconstructed jets, $\sum_{\text{e, } \mu} \vec{p}_T$ is the sum of p_T of all PF electrons and muons, and $\sum_{\text{unclustered PF cand.}} \vec{p}_T$ is the sum of p_T of all the other PF candidates in the event. Raw PF MET with Type 1 MET corrections [81] (in Type 1 MET corrections, the JEC corrections are propagated to MET) applied is then labeled as Type 1 PF MET and is referred in this thesis as 'MET'.

2. Tracker MET: In the so-called CaloMET, MET is calculated as the negative vector sum of the transverse energies deposited in the calorimeter towers. It is further corrected for the presence of identified muons and the mis-measurement of the hadronic energy in the calorimeters [82]. If the transverse momentum of each reconstructed charged particle track is added to the CaloMET, from which the corresponding transverse energy expected to be deposited in the calorimeters is subtracted, track-corrected missing transverse energy called Tracker MET (trkMET) is obtained. TrkMET corrects the imperfect response of the calorimeter to charged hadrons since it relies on the silicon tracker which has excellent linearity and angular resolution.

The significance of the measured missing energy ($\sigma\left(E_T^{miss}\right)$) is computed [83] as:

$$\sigma\left(E_T^{miss}\right) = \frac{E_T^{miss}}{\sqrt{\sum_{\text{non-PU jets, } p_T > 30\text{GeV}} |\vec{p}_T|}} \tag{3.6}$$

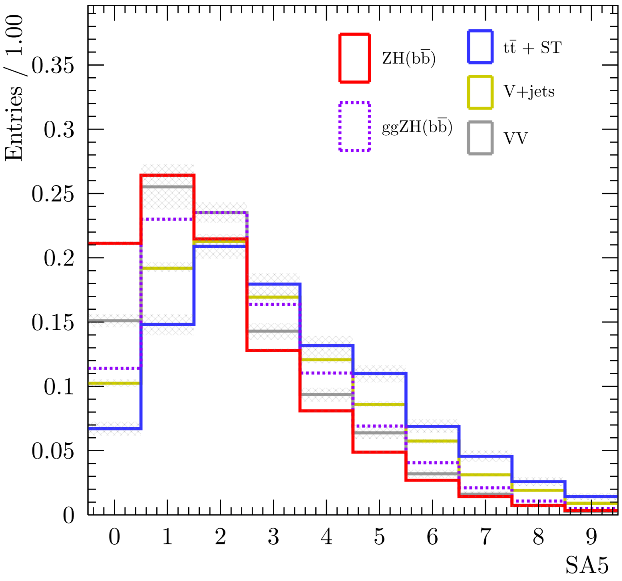


Figure 26: Normalized plot of SA5 with respect to each process. As shown, the signal and background processes have discriminatory shapes.

and used in the 0-lepton channel as a discriminating quantity to select events.

In the 0-lepton channel, since we expect jets but no prompt leptons in the final state, we use pfMHT as well. It is defined as:

$$\text{pfMHT} = \left| - \sum_{\text{non-PU jets, } p_T > 30, |\eta| < 2.4\text{GeV}} \vec{p}_T \right| \tag{3.7}$$

and is referred in the thesis as 'MHT'.

3.6.9 Additional soft hadronic activity

Signal events generally do not have additional hadronic activity after excluding the Higgs boson and vector boson decay products. The amount of additional hadronic activity can thus provide a useful quantity for separating signal from background, for example, as shown in Figure 26. The total additional hadronic activity of an event is expected to be soft and is constructed using only charged tracks compatible with the primary vertex of the event ($(|d_z(PV)| < 2$ mm), with $p_T > 300$ MeV. Potential tracks associated to the leptons of the two b-jets are not counted. The final collection of these 'soft-tracks' are clustered using the anti-kt [73] algorithm with a distance parameter $\Delta R = 0.4$. Soft activity is referred to as SA5 ('5' in 'SA5' corresponds to the $p_T > 5$ GeV threshold on the selected soft jets). Further discussion of data/MC modelling on the SA5 variable is given in Section 3.11.1.

3.6.10 b-jet energy regression

Compared to jets originating from the light-flavor quarks or gluons, b-jets have different characteristics like displaced secondary vertex, or possible presence of leptons from the B hadron decay. The neutrinos, which interact only via weak interaction, escape detection leading to an underestimation of the b-jet energy. The possibility of further mis-estimation of energy due to energy leakage outside the reconstructed jet increases the degradation of the energy resolution.

To correct the mismeasured energy of the b-jet in both data and MC, a multi-dimensional correction is derived using a neural network known as the b-jet energy regression [84]. Since the top quark decays to the bottom quark via over 99% branching fraction [85] via t → W^+b, a large sample of simulated top quark (100 million b-jets from simulated t$\bar{t}$ events) is used for the training. The target of the regression is $p_T^{\text{gen}}/p_T^{\text{reco}}$: the ratio of p_T of generator jet with reconstructed jet. Jet kinematics, PU information, jet shape, jet energy fraction, tracking information, and secondary vertex variables are used as input to the b-jet regression. Figure 27 shows how the peak of the invariant mass of the reconstructed Higgs boson shifts towards the mass of the Higgs boson (125 GeV) along with an improvement of about 8% in the dijet mass resolution.

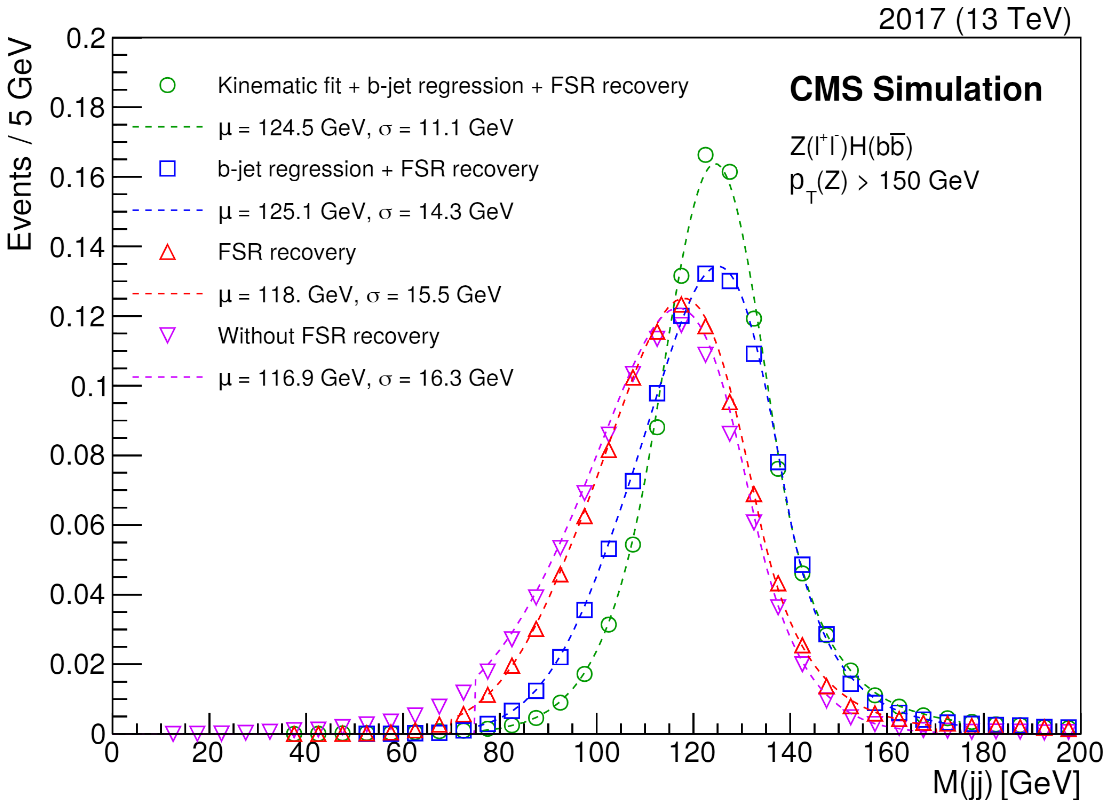


Figure 27: The invariant mass of the reconstructed Higgs boson in the 2-lepton channel fitted with the Bukin fit function [86] in different configurations: without Final State Radiation (FSR) recovery, with FSR recovery, with additional b-jet energy regression correction and with kinematic fit.

3.6.11 Smearing of b-jets after b-jet regression

To match the jet energy resolution in MC (where the b-jet regression is trained) to the one measured in the data, the energy resolution of the b-jet in MC is smeared. To measure jet energy resolution, we consider an event where a jet is recoiling against a Z boson that decays into leptons. The transverse momentum of the Z boson is balanced against the transverse momentum of the jet. Since the lepton energies in the CMS detector have relatively precise measurement, the ratio of p_T of the reconstructed jet to the Z boson can be used to measure the jet energy resolution. Because of the flavor conservation (g → b$\bar{\rm b}$) such an event giving a Z boson and one jet is suppressed. Thus, we consider events with two jets where one jet has relatively low p_T. A fit is then performed to estimate the expected resolution, extrapolating the second jet to $p_T = 0$.

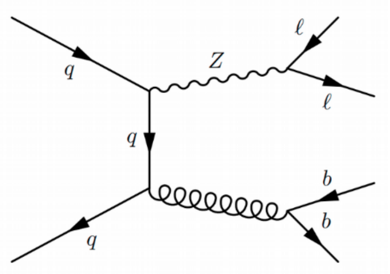


Figure 28: Feynman diagram for Z(l$\bar{\rm l}$)g(b$\bar{\rm b}$) process.

For this, we consider the Z(l$\bar{\rm l}$)g(b$\bar{\rm b}$) topology as shown in Figure 28. DeepCSV Tight WP selection is applied on the leading jet and loose on the sub-leading jet to reduce the contamination of light jets in the measurement process. Additionally, the two leptons are required to pass the selection for the 2-lepton channel. The di-lepton system is required to have $p_T > 100$ GeV with 71 GeV< $\rm m_{ll}$ < 111 GeV.

The selected events are then divided into four bins of $\alpha = p_{T,j2}/p_{T,ll}$, i.e. the ratio of p_T of the subleading jet with respect to the dilepton system, with bin boundaries (0, 0.155, 0.185, 0.23, 0.3). When $\alpha = 0$, the Z(l$\bar{\rm l}$)g(b$\bar{\rm b}$) topology corresponds to the 1-jet system recoiling against the Z boson since the other jet will have $p_T = 0$ GeV. The jet response in each of the four bins, together with the dominant uncertainties, namely the renormalization and factorization scale uncertainties, is shown in Figure 29. From each of these plots, the scale (μ) and resolution (σ) are extracted and fitted as a function of α (as shown in Figure 30 (left)) as:

$$f(\alpha) = (m \times \alpha) \oplus b \times (1 + c_k \times \alpha) \tag{3.8}$$

where, c_k is fixed by a linear fit to the MC's intrinsic jet resolution (i.e. $p_T^{\rm reco}\,/p_T^{\rm gen}$).

Smearing for MC jets is performed by scaling the difference between $p_T^{\rm gen}$ and $p_T^{\rm reco}$ by $b_{\rm data}/b_{\rm MC}$. This leads to a better agreement of the jet energy resolution as shown by better modelling after the fit at $\alpha = 0$ in Figure 30 (right). On the other hand, as the scaling is compatible with 0 within uncertainties, we apply no additional scaling factor for all three years. The smearing and scaling factors with their fit uncertainties are given in Table 8.

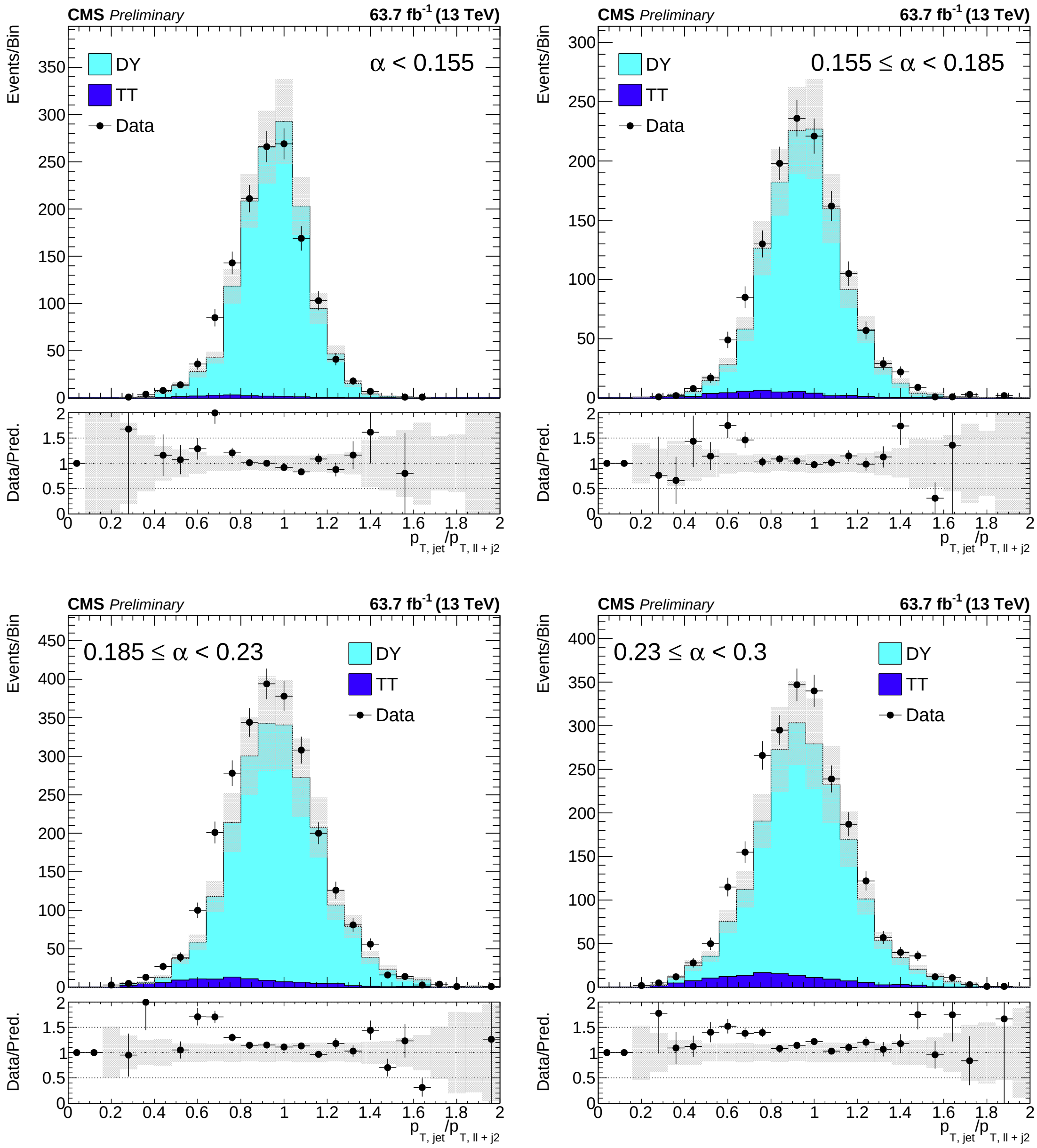


Figure 29: Jet response in the four bins of $\alpha = p_{T,j2}/p_{T,ll}$ for the Z($\mathrm{l\bar{l}}$)g($\mathrm{b\bar{b}}$) topology using the 2018 dataset.

Year	Scaling	Smearing
2016	$+0.4 \pm 1.8\%$	$-4.4 \pm 6.1\%$
2017	$+1.1 \pm 2.2\%$	$+5.1 \pm 6.8\%$
2018	$-1.8 \pm 1.9\%$	$+5.0 \pm 7.9\%$

Table 8: Scaling and smearing needed for each year of data as a percent of the jet transverse momentum.

3.6.12 Pile-up jet identification

Pile-up affects several aspects of this analysis such as the jet momentum resolution, Higgs boson reconstruction, lepton isolation, MET reconstruction and b-tagging. Two

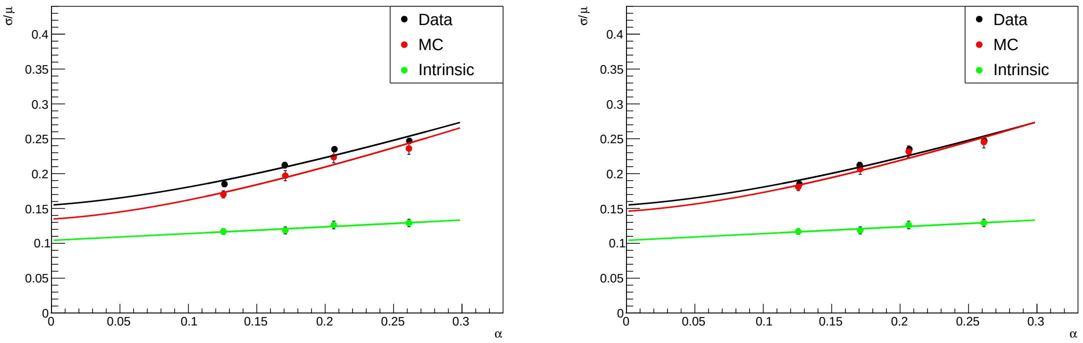


Figure 30: Data, MC, and intrinsic resolution before (left) and after (right) the fit. The x-axis is $\alpha = p_{T,j2}/p_{T,\ell\ell}$ i.e. the ratio of p_T of subleading jet with the dilepton system; and the y-axis is the ratio of resolution σ and scale μ.

observables are used in the analysis to describe the pile-up: N_{PV}, the number of primary vertices in the luminous region with at least 4 tracks and ρ, the median energy density in the calorimeters over a fixed grid of cells in (η, ϕ).

Pile-up jets generally have $p_T < 50$ GeV. However, they can overlap and get reconstructed as a single jet with high p_T. It is estimated for events with 35 PV to have around 2 PU jets per event with $p_T > 25$ GeV and $\eta < 2.5$. A BDT algorithm (gradient-boosted decision tree) is used to separate PU jets from hard scatter jets. The algorithm has an efficiency of 80-90% for non-PU jets with $p_T < 50$ GeV passing the Tight working point of PU ID. The difference in PU distribution in data and MC is further corrected using dedicated weights and is discussed in Section 3.10.1.

3.7 Event Reconstruction

The reconstruction of the Higgs boson in resolved and boosted topology is discussed in Section 3.7.1 and 3.7.2 respectively. The reconstruction of the associated vector boson is discussed in Section 3.7.3. As the top quark is an important background in the 0- and 1-lepton channel, the mass of the top quark is reconstructed and is discussed in Section 3.7.4. Since the 2-lepton channel has no prompt neutrinos, the kinematics of the event can be fully reconstructed. The dijet mass resolution can thus be improved by the kinematic fit in the 2-lepton channel and is discussed in Section 3.7.5.

3.7.1 Higgs Reconstruction with FSR recovery

The resolved Higgs boson candidates are reconstructed by adding the four vectors of the two highest b-tagged resolved jets. They are required to pass a channel-dependent transverse momentum threshold, Tight pile-up ID, Tight jet ID, lepton filter, and should be within the acceptance of the tracker. Before reconstructing the Higgs boson, additional jets (the Final State Radiation i.e. FSR jets) with $p_T > 20$ GeV and $|\eta| < 3.0$ and within $\Delta R = 0.8$ of the selected jet are added to the selected jet. The FSR jets are also required to pass the Tight pile-up ID, Tight jet ID, and lepton filter. This procedure is known as the FSR recovery. The FSR-recovered b-jets are then used to reconstruct the Higgs boson candidate.

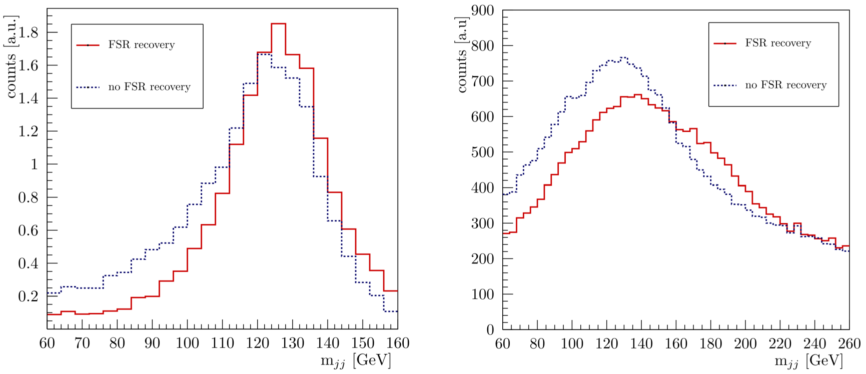


Figure 31: Dijet mass with and without FSR recovery algorithm applied for WH signal (left) and TT background (right) in the ≥ 1 jet signal region in the 1-lepton channel

According to the definition of the FSR jets described above, around 15% of signal events have one FSR jet while 1% have more than one FSR jet. The effect of FSR recovery on the dijet mass resolution of the signal is around 5% (as shown in Figure 27), helping significantly in the signal background separation. The dijet mass for events with more than one additional jet for the signal process (left) and $t\bar{t}$ background (right) in the SR of the 1-lepton channel is shown in Figure 31. As observed, the dijet mass peak for $t\bar{t}$ background shifts to the right reducing the background contribution in the signal region (90 < m(jj) < 150 GeV).

3.7.2 Boosted Higgs boson reconstruction

The AK8 FatJets are reconstructed using the anti-kt clustering algorithm [73] with the distance parameter R= 0.8. Due to the contamination from ISR, FSR, UE, and PU, the measured FatJet mass is shifted to larger values with respect to the underlying parton mass. The soft drop declustering algorithm [87] is applied to the FatJet masses which recursively removes the soft and wide radiation. This helps to recover the agreement between the measured mass and the true parton mass. The mass of the FatJet after applying the soft-drop algorithm is known as the soft-drop mass, m_{SD}.

The boosted Higgs boson candidate is reconstructed using the AK8 jet having the highest DeepAK8bbVSlight score, passing the lepton filter and having $p_T > 150$ GeV, $|\eta| < 2.5$ and $m_{\mathrm{SD}} > 50$. The lepton filter removes all jets within a $\Delta\phi$ =1.57 distance around the ϕ of the selected vector boson. This avoids misidentification of leptons from vector boson decay as leptons from b-decays inside the jet.

3.7.3 Vector boson Reconstruction

The vector boson candidate is reconstructed from the missing transverse energy and/or the isolated leptons in the event. In the case of a 2-lepton channel, the vector boson kinematics can be fully reconstructed using the four vectors of the two charged leptons. In the case of 1- and 0-lepton channels, the vector boson kinematics can be reconstructed only in the transverse plane. For the 1-lepton channel, the vector boson transverse mass and the azimuthal angle are reconstructed using MET and isolated leptons. The two (one) leptons used to reconstruct the vector boson in the 2-lepton channel (1-lepton channel) are chosen as the two (one) highest p_T leptons in the event, passing the p_T, isolation, and lepton ID cuts discussed in Section 3.6.4. The transverse mass in the 1-lepton channel is defined as

$$m_T(V) = \sqrt{2p_T(l)\mathrm{MET}\left(1 - \cos\left(\phi_{\mathrm{lep,MET}}\right)\right)} \tag{3.9}$$

where $m_T(V)$ is the vector boson transverse mass, $p_T(l)$ is the p_T of the isolated lepton, and $\phi_{\mathrm{lep,MET}}$ is the azimuthal angle between MET and isolated lepton. The transverse mass of the vector boson defined above is used as a discriminant variable in MVA to separate signal and background and is discussed in Section 3.11.1. In the 0-lepton channel, only MET can be used for the reconstruction of the vector boson.

3.7.4 Top quark reconstruction

Since the top quark is a dominant background in the 1-lepton channel, the top quark reconstruction is performed in the 1-lepton channel and only for top quarks from leptonically decaying W bosons in the semi-leptonic $\mathrm{t\bar{t}}$ decay. Reconstructed top mass can then be used as a discriminating variable in the MVA of the HF CR (discussed in Section 3.11). The top quark is reconstructed using the selected lepton, MET, and the closest b-jet passing the Medium working point of the DeepCSV classifier. The longitudinal momentum of the neutrino is estimated by imposing the on-shell mass constraint on the W boson in the kinematic decay equation leading to a quadratic equation. In the case of two real solutions, the smaller solution is taken. The top mass reconstruction distribution in the 1-lepton channel is shown in Figure 32. The long tail to the right of the top mass peak (at 172 GeV) is due to the incorrect selection of the b-jet in the event.

3.7.5 Kinematic Fit

Finally, kinematic fit is applied to improve the resolution of the kinematic variables of the final state particles in the 2-lepton channel. It is an event-by-event chi-square fit by which the resolution can be improved by using kinematic constraints in the event. The kinematic constraints are implemented as Lagrange multipliers [88]. The least-square minimization is then performed to obtain the fit parameters such as p_T of the jets and leptons.

For the 2-lepton channel, the kinematics of the events can be fully reconstructed due to the absence of prompt neutrino in the event kinematics. The kinematic fit can thus be performed to use the excellent momentum resolution of leptons recoiling against the Higgs boson candidate to constrain the transverse momentum of the jets, improving the dijet resolution, and in turn, improving the mass of the reconstructed Higgs boson. The energy of the neutrinos from the b-jet decay is partially recovered

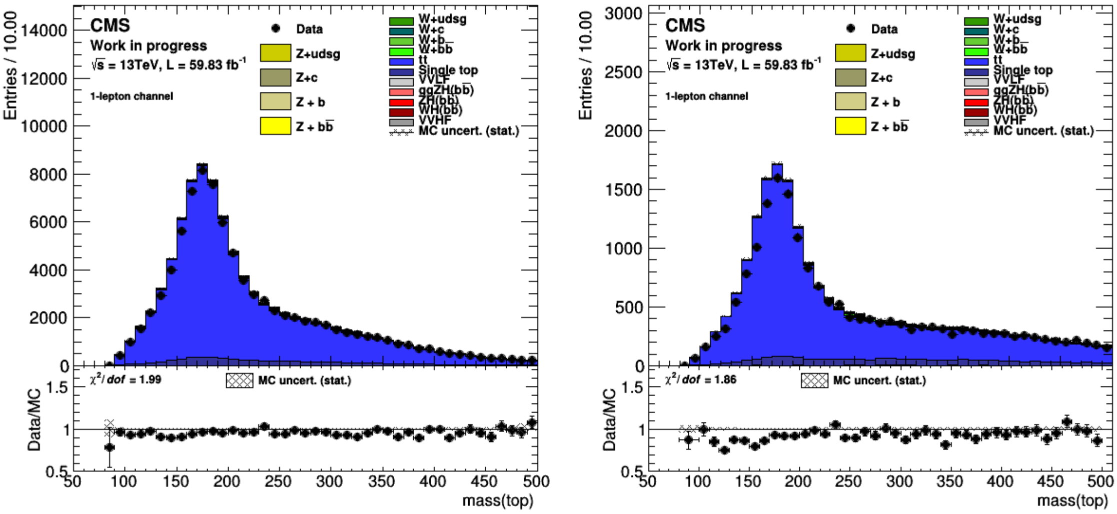


Figure 32: Reconstructed top mass in the 1-lepton channel for $150 < p_T(\mathrm{V}) < 250$ GeV and $p_T(\mathrm{V}) > 250$ GeV.

using the b-jet energy regression. The jet direction in η, ϕ is not introduced in the fit since its resolution is already good.

The following objects, in the 2-lepton channel events, are balanced in the kinematic fit:

- two jet candidates for the Higgs boson candidate after b-jet energy regression: j_1, j_2

- FSR jets: The FSR jet candidates are selected from all the additional jets within ΔR <0.8 to the closest Higgs candidate jets having p_T > 20 GeV, pass loose pileup ID, Jet ID, and the lepton filter.

- One recoil jet: The vectorial sum of non-FSR or Higgs candidate jets is identified as an ISR (Initial State Radiation) jet. To reduce the effect of pileup, a Tight pile-up ID is used. Also, the entire detector coverage, $|\eta| < 5$ is used.

- Two lepton candidates (l_1, l_2): the two lepton candidates used in the reconstruction of the vector boson (discussed in Section 3.7.3).

The following constraints are used in the least squared fit:

- Z boson mass: The sum of the mass of the two candidate leptons is constrained to the Z boson mass.

- Transverse momentum: As the signal process does not have an intrinsic MET, the vector sum of all fitted particles is constrained to zero in the transverse plane.

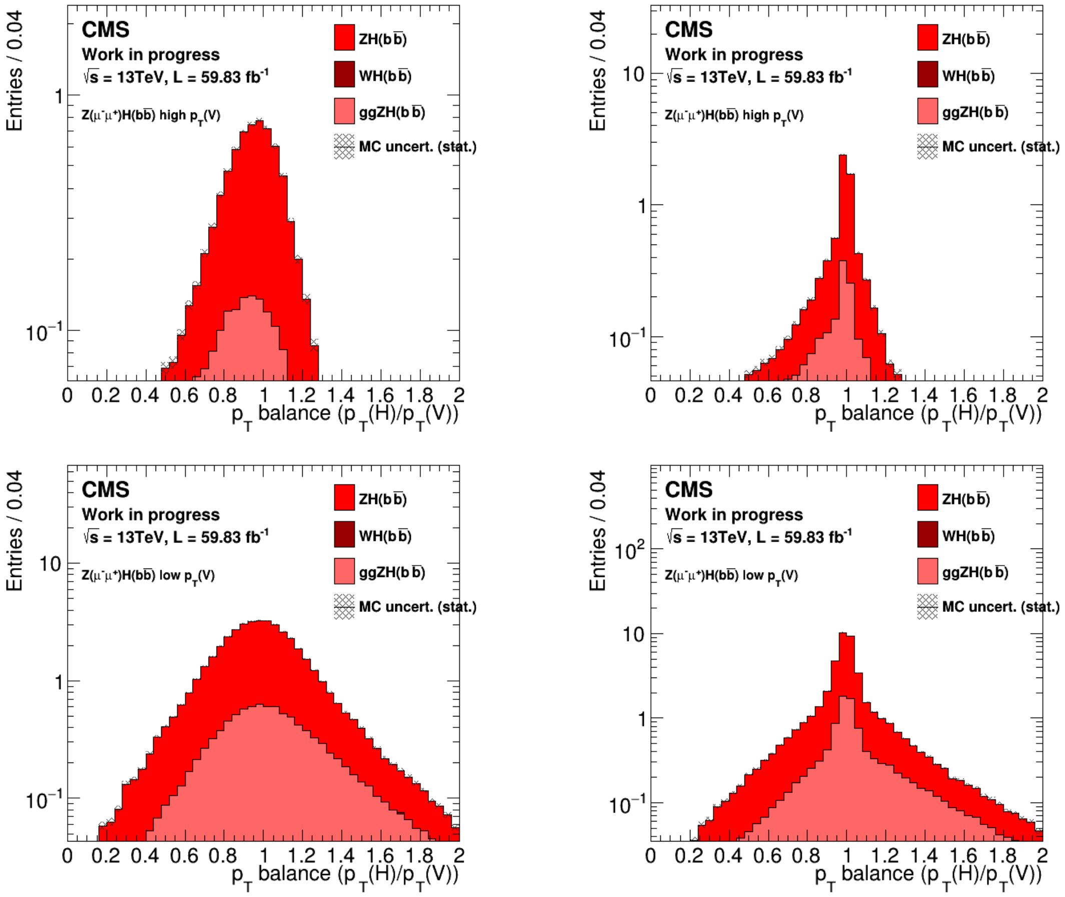


Figure 33: The balance of the reconstructed transverse momentum of the jets and leptons in the signal processes before (left) and after (right) the kinematic fit for p_T(V) > 250 GeV (top) and 150 < p_T(V) < 250 GeV (bottom).

The four vectors of the Higgs candidate jets post kinematic fit along with the FSR jets are then summed together to form the Higgs boson candidate four vectors while the vector boson is formed by summing the two leptons. As a closure test, the balance of reconstructed transverse momentum of the jets and leptons, before and after the kinematic fit, is shown in Figure 33. As shown, the resolution of the distribution is reduced post-kinematic fit.

The invariant mass of the reconstructed Higgs boson fitted with the Bukin fit function [86] (a function for fitting asymmetric peaks) is shown in different configurations in Figure 27. The dijet mass resolution after kinematic fit reduces by almost 22%.

3.8 Analysis Selection

All needed datasets are centrally available for our analysis in NanoAOD format [89]. The VH(bb) analysis-specific very loose selection, as well as the jet energy corrections, are first applied to these datasets in the so-called post-processing step. After that, channel-dependent basic cuts are applied to reduce the amount of data to process in the later steps. These include requiring at least two jets with $p_T > 20$ GeV, isolated leptons with $p_T > 20$ GeV or MET > 170 GeV and applying the HLT trigger selection. These cuts ensure the presence of the required objects in the initial selected events. The object reconstruction is then applied only to these selected events.

The analysis phase space of each of the three channels is then further divided into orthogonal regions enriched in the signal process (Signal Region/SR), $t\bar{t}$ process (TT CR), V+HF process (HF CR), and V+LF (LF CR). CRs are used to extract the normalization and shape of the background processes, extrapolate background predictions (shapes and normalization) to SR, and to verify or correct the data/MC agreement of analysis observables. All CRs are optimized to have a phase space with high efficiency of a given background process. The background process in CRs help constrain the scale factor (SF) (discussed in Section 3.13).

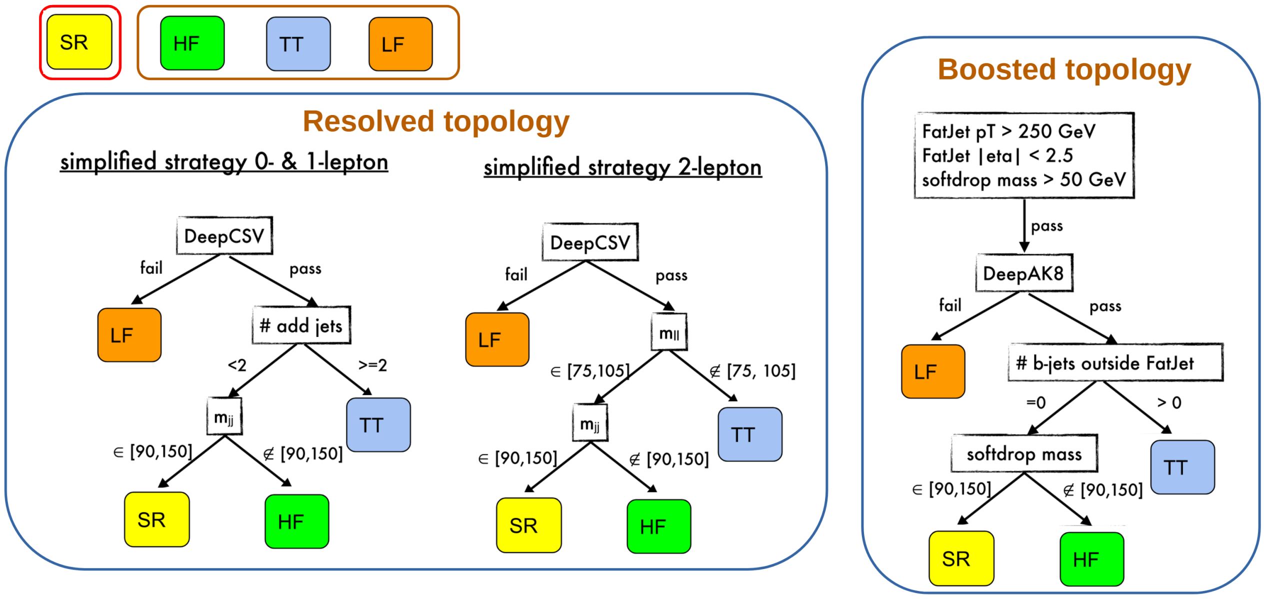


Figure 34: Channel-dependent analysis selections in simplified form for the resolved and boosted topology.

The dijet mass window around the Higgs boson mass for the SR is optimized by checking the sensitivity of the SR as a function of the dijet mass cut, while ensuring enough events in the orthogonal HF CR. This leads to a mass window from 90 to 150 GeV. The DeepCSV cuts on the leading and subleading jets are also similarly optimized by checking the sensitivity of the SR as a function of the DeepCSV working point. The optimization converges on the Medium WP for the leading jet and Loose WP for the subleading jet. Similar b-tagging cuts are applied for the HF CR and SR to have a similar phase space and V+jets flavor composition, and to reduce the extrapolation uncertainties (from CR to SR). The orthogonality in HF and SR is obtained by selecting events in the sidebands for the HF CR.

Figure 34 shows the channel-dependent analysis selections in a simplified form for the resolved and boosted topology. They are further described in detail in the following Sections.

HEM 15/16 issue in 2018 data

The endcaps modules of the hadron calorimeter, namely HEM15 and HEM16, were switched off in the region $\phi \in [-1.57; -0.87], \eta \in [-3.0; -1.3]$ during the 2018 data taking period due to a malfunctioning power supply starting from run 319077. This issue is referred to as the hadronic calorimeter endcaps minus (HEM 15/16) issue. It affected jet detection, and thus MET reconstruction, in the 0-lepton channel while the mis-reconstruction of jets as leptons affected the 1-lepton channel.

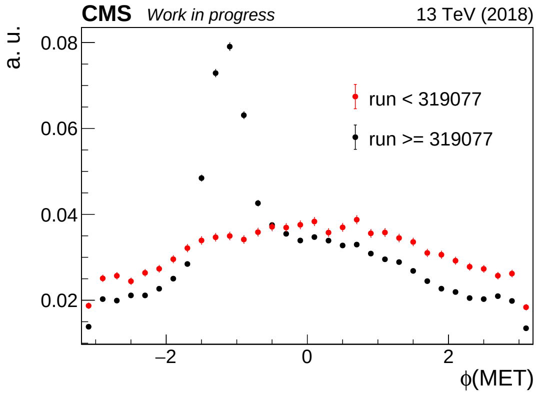


Figure 35: ϕ(MET) distributions before and after the run 319077 in the 0-lepton channel for 2018 data.

Figure 35 shows the distribution of ϕ(MET) before and after the run 319077 in the 0-lepton channel for 2018 data. As shown, the spike corresponds to the region where the mis-reconstruction of jets lead to fake MET due to the HEM 15/16 issue.

Similarly, the mis-reconstruction of jets to leptons was observed in the 1-lepton channel as shown by the excess in the number of the reconstructed electrons in the 2D (ϕ, η) distribution of the reconstructed electrons in the 1-lepton (electron) channel in Figure 36.

To account for this mis-reconstruction, additional selection cuts are applied. In the 0-lepton channel, events in data with $-1.86 < \phi < -0.7$ region are removed. This window is obtained by computing the fractional change in the number of events before and after run 319077 as a function of ϕ(MET). In the 1-lepton channel, events in data with electrons in $-1.5 < \phi < -0.75$ and $\eta < -1.5$, corresponding to the affected region in Figure 36 (right), are removed. MC samples in the affected regions are reweighted by about 65% to account for the loss in luminosity in the region. Figure 37 shows the ϕ(MET) distribution in the HF (left) and LF CR (right) in the 0-lepton channel after the additional treatment due to the HEM 15/16 issue on ϕ(MET).

3.8.1 0-lepton channel

The 0-lepton signal event is characterized by the presence of large MET due to the decay of the Z boson to neutrinos which recoils against the Higgs boson decaying to a pair of bottom quarks. Additionally, it has low additional lepton activity and no additional high p_T leptons. MET is required to be larger than 170 GeV, along

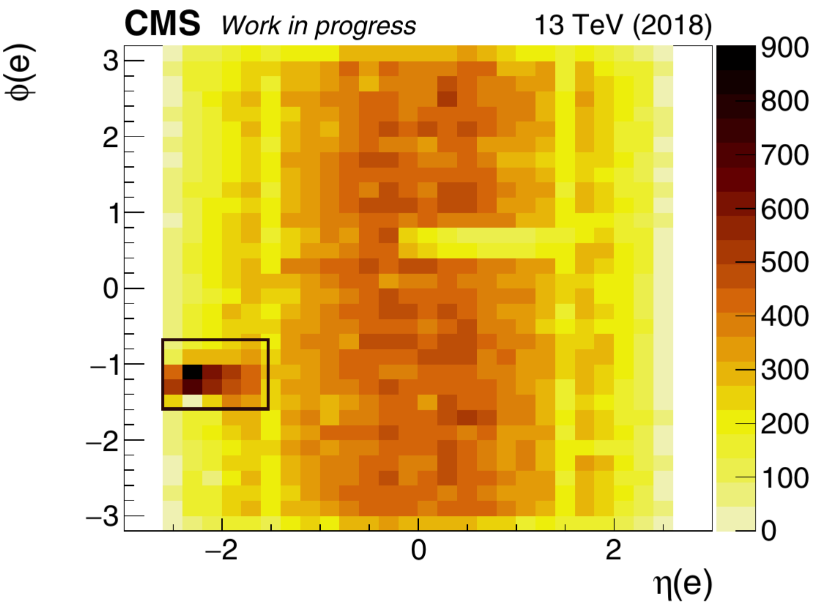


Figure 36: 2D (ϕ, η) distributions of reconstructed electrons in the 1-lepton (electron) channel for 2018 data from run 319077 onwards.

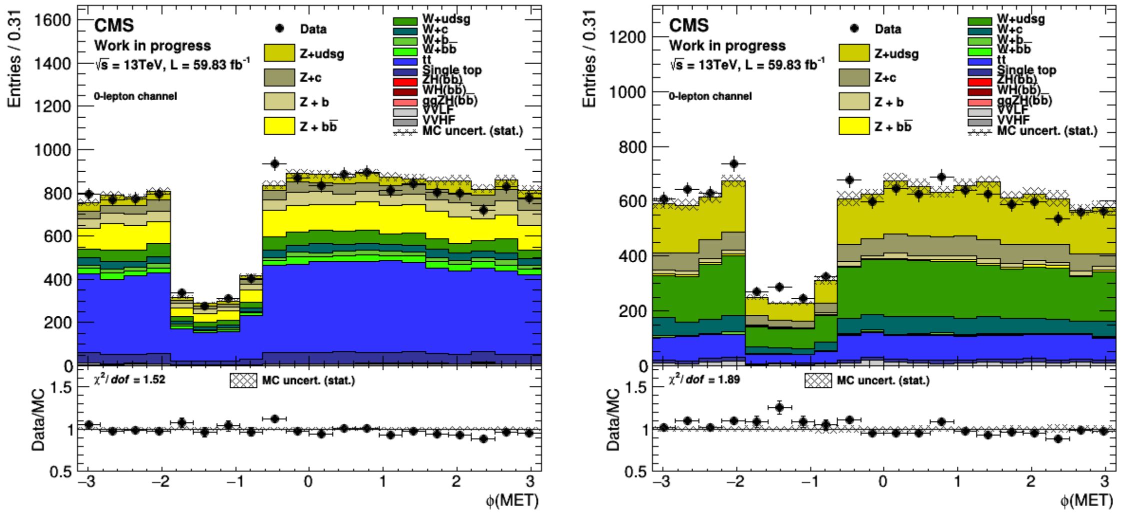


Figure 37: ϕ(MET) distribution in HF (left) and LF CR (right) in the 0-lepton channel in 2018 after the additional treatment due to the HEM 15/16 issue on ϕ(MET).

with min(MET, MHT) to exceed 100 GeV. The Higgs boson candidate is constructed using the two highest b-tagged jets with leading and subleading jets having regressed $p_T > 60$ GeV and $p_T > 35$ GeV respectively.

Signal region selection In addition to the mass window and b-tagging requirement, due to low expected additional jet activity, the number of additional jets ($\mathrm{N_{aj}}$) with $|\eta| < 2.5$ and $p_T > 30$ GeV is required to be 0 or 1. No leptons with $p_T > 30$ GeV should be present. No jets are expected to be close to the direction of MET i.e. $\Delta\phi$(MET, jet) > 0.5 for jets with $p_T > 30$ GeV, passing Tight jet ID and pileup ID cut. These cuts are referred to as the 'anti-QCD cuts' and help in reducing the background from multi-jet QCD. The complete list of cuts for SR is given in Table 9.

Control region selection For the LF CR, the b-tagging requirement on the leading jet is inverted and the dijet mass window requirement is removed. For the TT CR, two additional jets ($\mathrm{N_{aj}}$) with $p_T > 30$ GeV are required. The complete list of cuts for CR is given in Table 9.

Variable	SR	HF CR	LF CR	TT CR
Common selection between SR and CRs:				
min(MET, MHT)	> 100	-//-	-//-	-//-
MET	> 170	-//-	-//-	-//-
$\mathrm{p}_T^{j_1}$	> 60	-//-	-//-	-//-
$\mathrm{p}_T^{j_2}$	> 60	-//-	-//-	-//-
p_T(jj)	-//-	-//-	-//-	
$\Delta\phi$(Z,H)	> 2.0	-//-	-//-	-//-
m(jj)	$[50 - 500]$	-//-	-//-	-//-
$\mathrm{N_{al}}$	< 1	-//-	-//-	-//-
N(jets) close to MET	0	-//-	-//-	-//-
Different selection between SR and/or CRs:				
$\mathrm{N_{aj}}$	≤ 1	≤ 1	≤ 1	≥ 2
m(jj)	$[90 - 150]$	$\notin [90 - 150]$	-	-
DeepCSV(max)	> medium	> medium	< medium	> medium
DeepCSV(min)	> loose	> loose	> loose	> loose
$\Delta\phi$(MET, trkMET)	< 0.5	< 0.5	< 0.5	-
min$\Delta\phi$(MET, jet)	-	-	-	$< \pi/2$

Table 9: Definition of the SR and CR for the 0-lepton channel resolved selection. The symbol '-//- ' represents the same selection cut for CRs and SR. The symbol '-' refers to no selection. Mass and momentum have units of GeV. 'Loose', 'medium', and 'tight' refers to the DeepCSV WP. j_1 and j_2 refer to the leading and sub-leading jet in p_T.

The distribution of selected observables in the resolved SRs and CRs of the 0-lepton channel is shown in Figures 38-41.

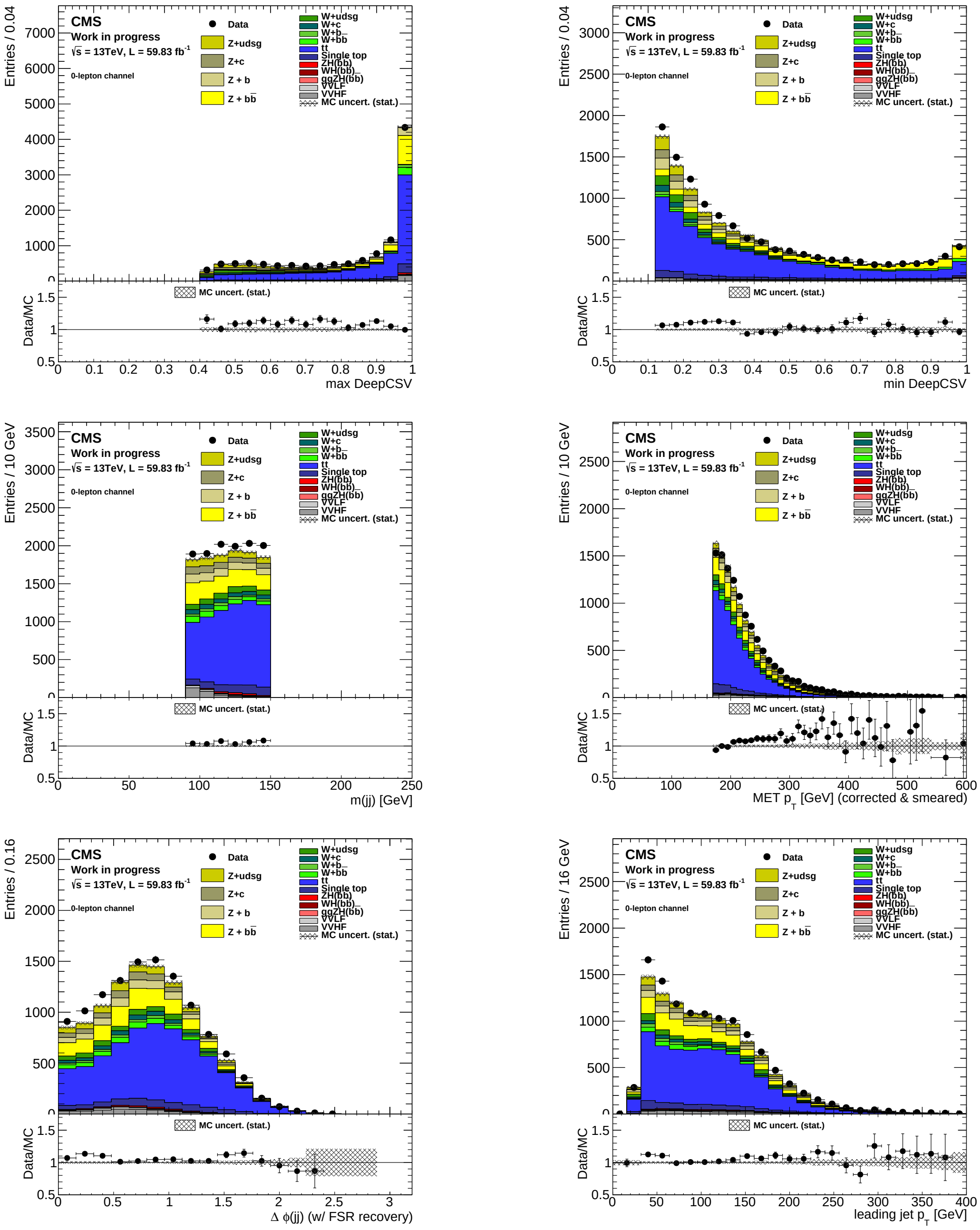


Figure 38: Selected kinematic observables in the resolved 0-lepton SR: DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), MET (middle right), $\Delta\phi$ between the dijets (bottom left) and p_T of the leading jet (bottom right).

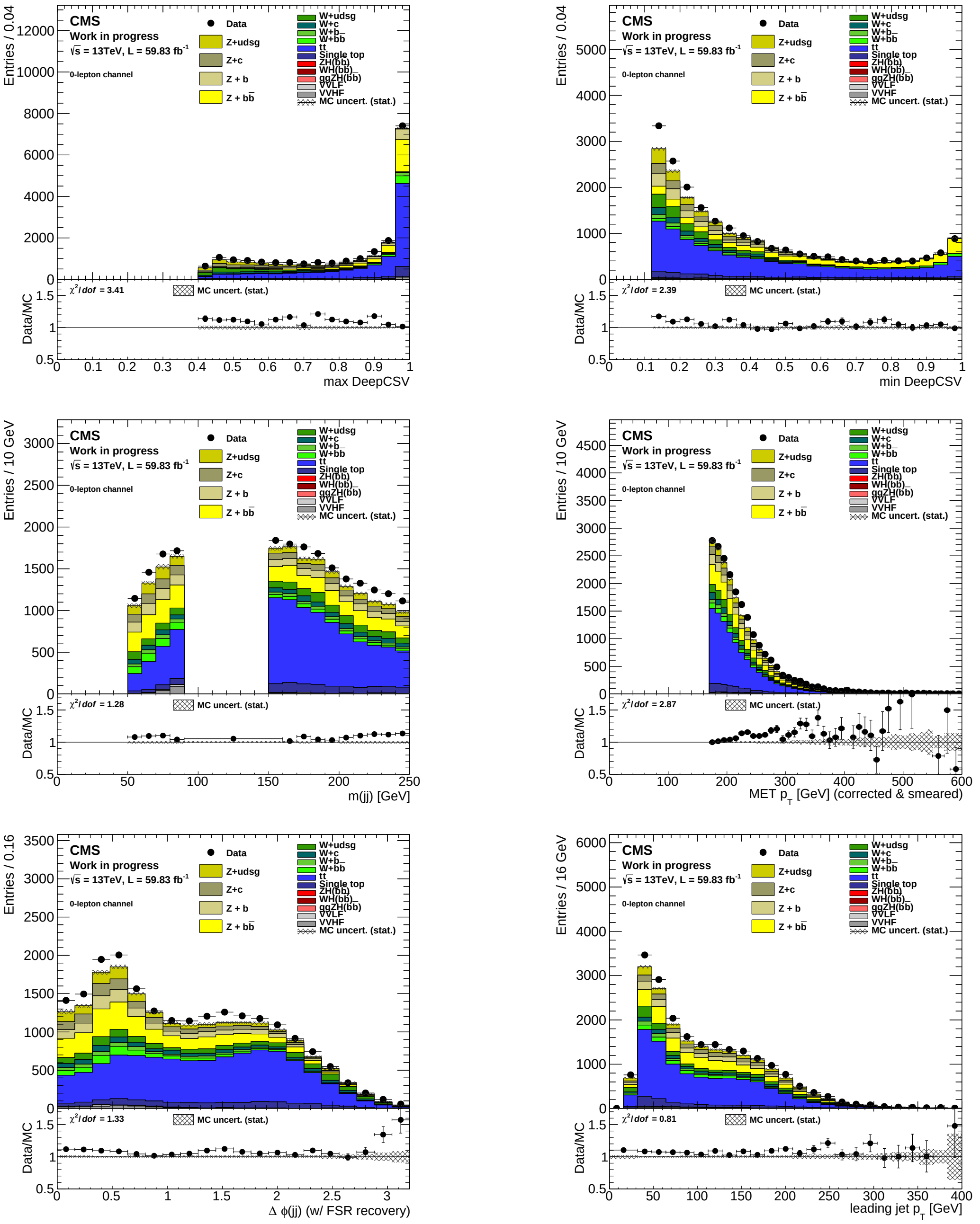


Figure 39: Selected kinematic observables in the resolved 0-lepton HF CR: DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), MET (middle right), $\Delta\phi$ between the dijets (bottom left) and p_T of the leading jet (bottom right).

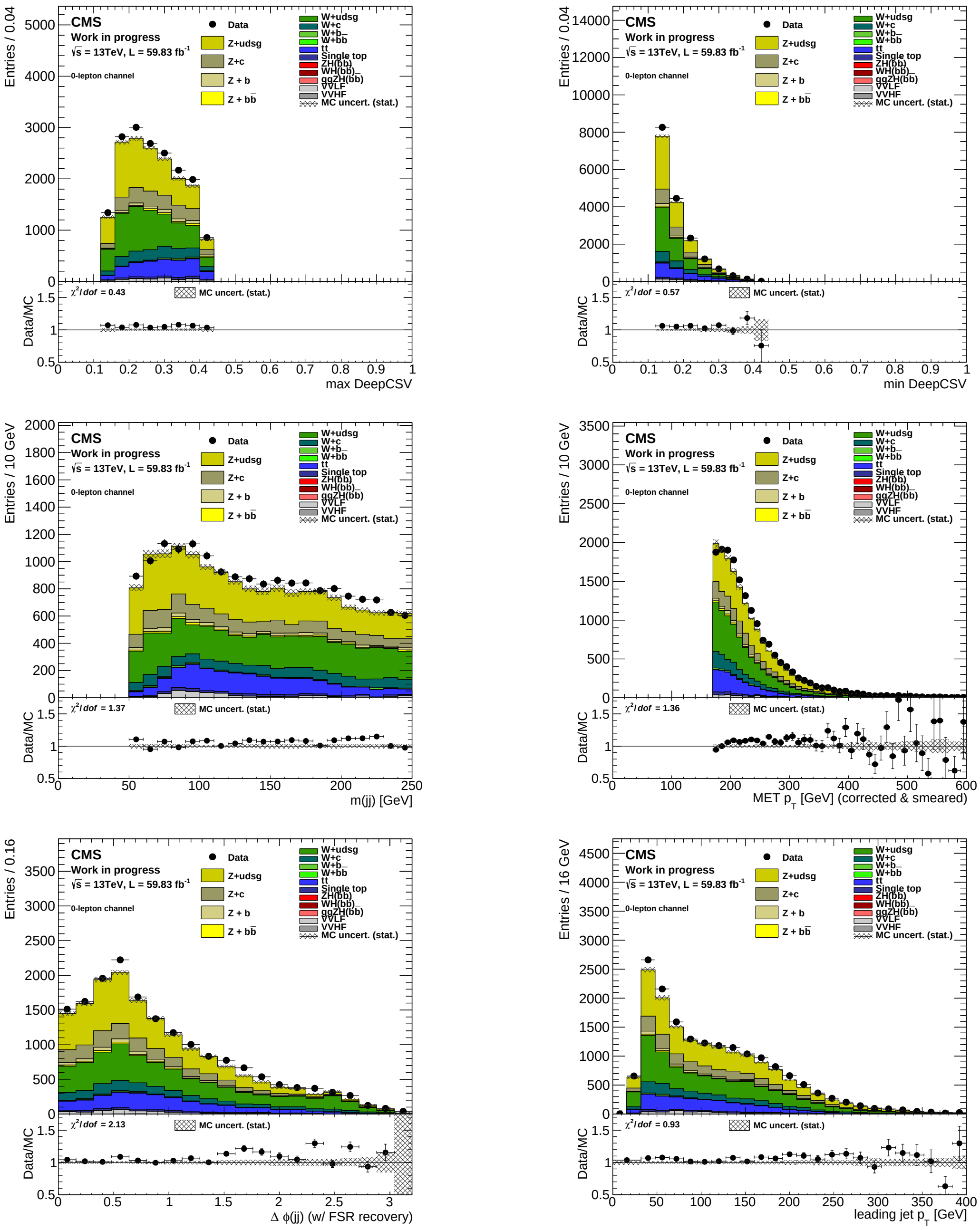


Figure 40: Selected kinematic observables in the resolved 0-lepton LF CR: DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), MET (middle right), $\Delta\phi$ between the dijets (bottom left) and p_T of the leading jet (bottom right).

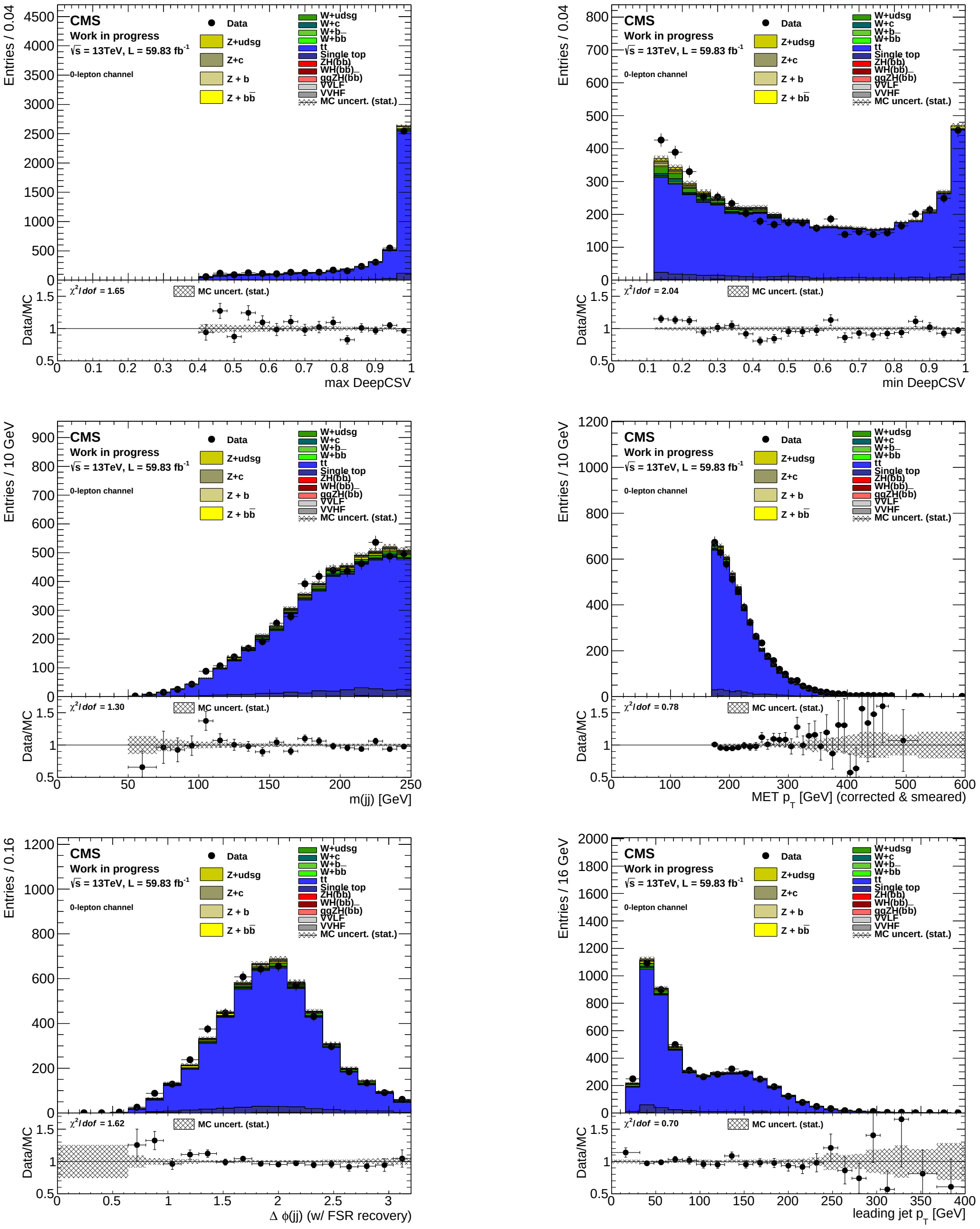


Figure 41: Selected kinematic observables in the resolved 0-lepton $t\bar{t}$ CR: DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), MET (middle right), $\Delta\phi$ between the dijets (bottom left) and p_T of the leading jet (bottom right).

3.8.2 1-lepton channel

Events in the 1-lepton channel are characterized by the presence of a single isolated lepton from the leptonic decay of the W boson which recoils against the Higgs boson decaying to a pair of bottom quarks. The presence of a single isolated lepton provides a trigger signature for this channel. The Higgs candidate is reconstructed using the two highest b-tagged jets with leading and subleading jets having regressed $p_T > 25$ GeV each. The transverse momentum of the reconstructed Higgs boson is further required to be more than 100 GeV. Events having additional leptons (N_{al}) with $|\eta| < 2.5$ and $p_T > 25$ GeV are not selected.

Signal region selection In addition to the dijet mass window and b-tagging requirement, the number of additional jets ($\mathrm{N_{aj}}$) with $|\eta| < 2.5$ and $p_T > 30$ GeV are restricted to 0 or 1. The complete list of cuts for SR is given in Table 10.

Control region selection For the LF CR, the b-tagging requirement on the leading jet is inverted and the dijet mass window requirement is removed. For the TT CR, at least one additional jet ($\mathrm{N_{aj}}$) with $p_T > 30$ GeV is required. The complete list of cuts for CR is given in Table 10.

Variable	SR	HF CR	LF CR	TT CR
Common selection between SR and CRs:				
p_T(jj)	> 100	-//-	-//-	-//-
p_V	> 150	-//-	-//-	-//-
$\mathrm{N_{al}}$	≥ 1	-//-	-//-	-//-
$\mathrm{p}_T^{j_1}$	> 25	-//-	-//-	-//-
$\mathrm{p}_T^{j_2}$	> 25	-//-	-//-	-//-
$\Delta\phi$(MET, l)	< 2	-//-	-//-	-//-
Different selection between SR and/or CRs:				
DeepCSV(max)	>medium	>medium	[loose-medium]	>tight
DeepCSV(min)	>loose	-	-	-
m(jj)	$[90 - 150]$	$[150 - 250]$ or <90	<250	< 250
$\mathrm{N_{aj}}$	< 2	< 2	-	>1
σ(MET)	-	> 2	> 2	-
$\Delta\phi$(H,V)	> 2.5	-	-	-

Table 10: Definition of the SR and CR for the 1-lepton channel resolved selection. The symbol '-//- ' represents the same selection cut for CRs and SR. The symbol '-' refers to no selection. Mass and momentum have units of GeV. 'Loose', 'medium', and 'tight' refers to the DeepCSV WP. j_1 and j_2 refer to the leading and sub-leading jet in p_T.

The distribution of selected observables in the resolved SRs and CRs of the 1-lepton channel is shown in Figures 42-49.

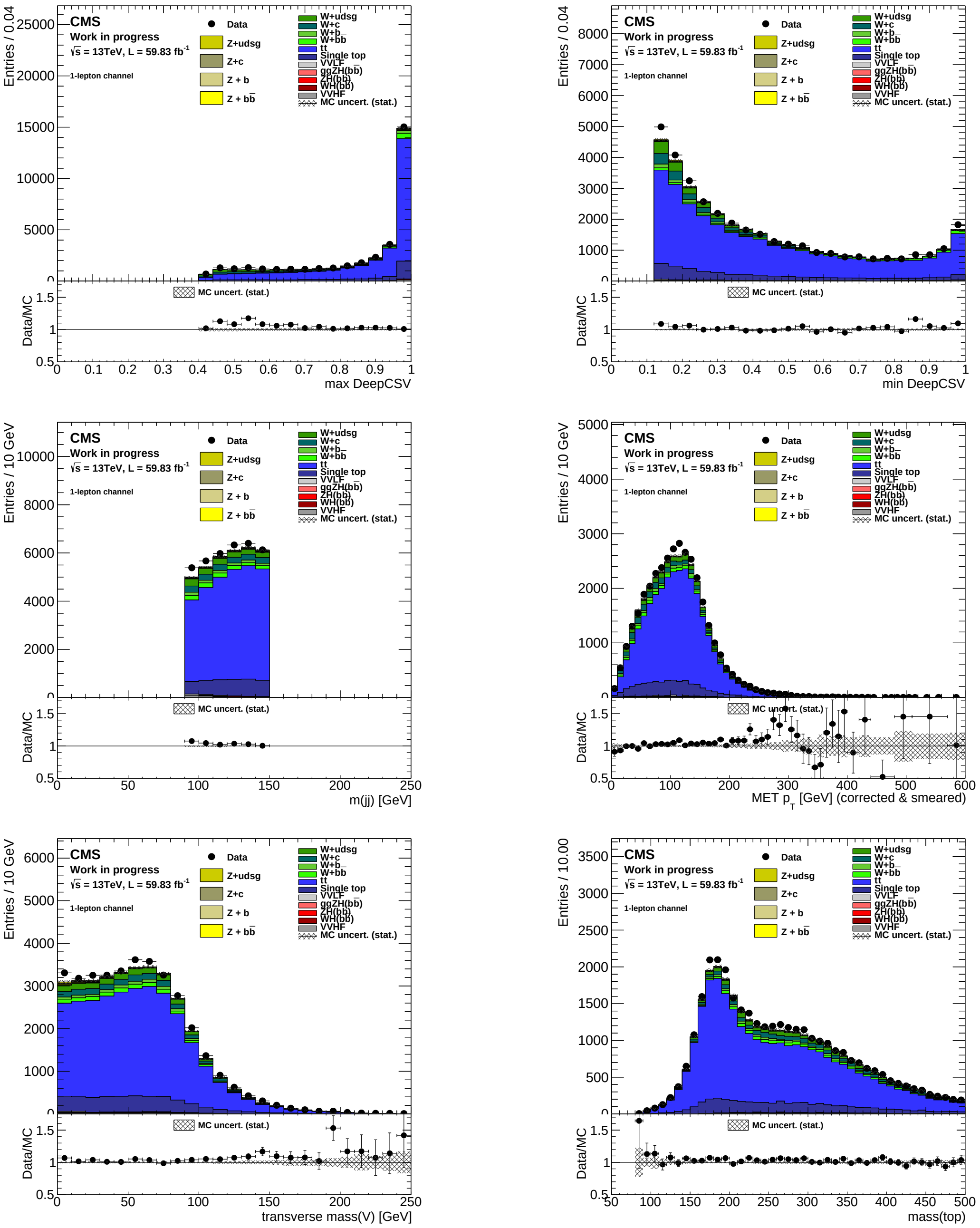


Figure 42: Selected kinematic observables in the resolved 1-lepton SR (muon channel): DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), MET (middle right), the transverse mass of the reconstructed vector boson (bottom left) and the reconstructed top mass (bottom right) after the kinematic fit and FSR recovery procedure.

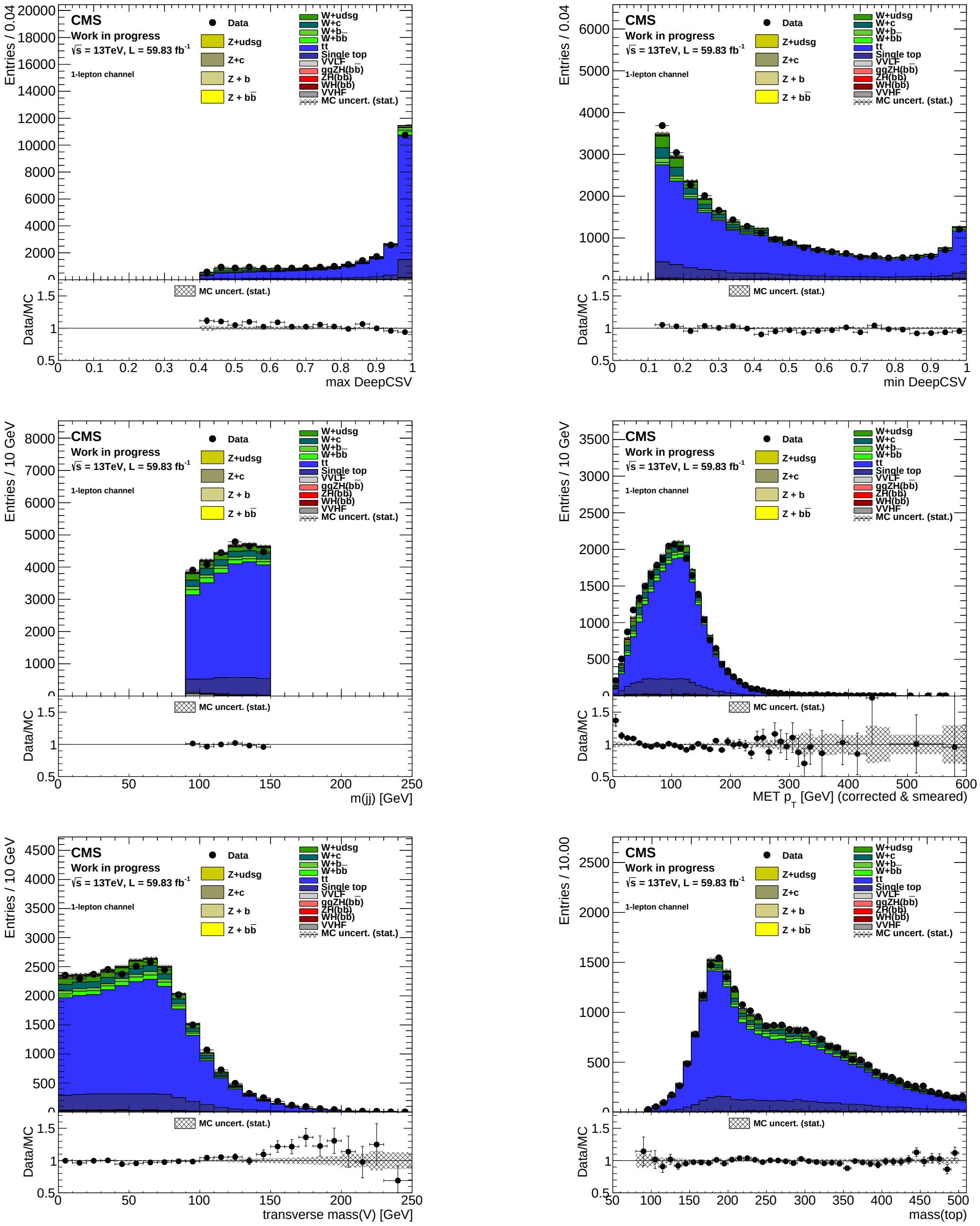


Figure 43: Selected kinematic observables in the resolved 1-lepton SR (electron channel): DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), MET (middle right), the transverse mass of the reconstructed vector boson (bottom left) and the reconstructed top mass (bottom right) after the kinematic fit and FSR recovery procedure.

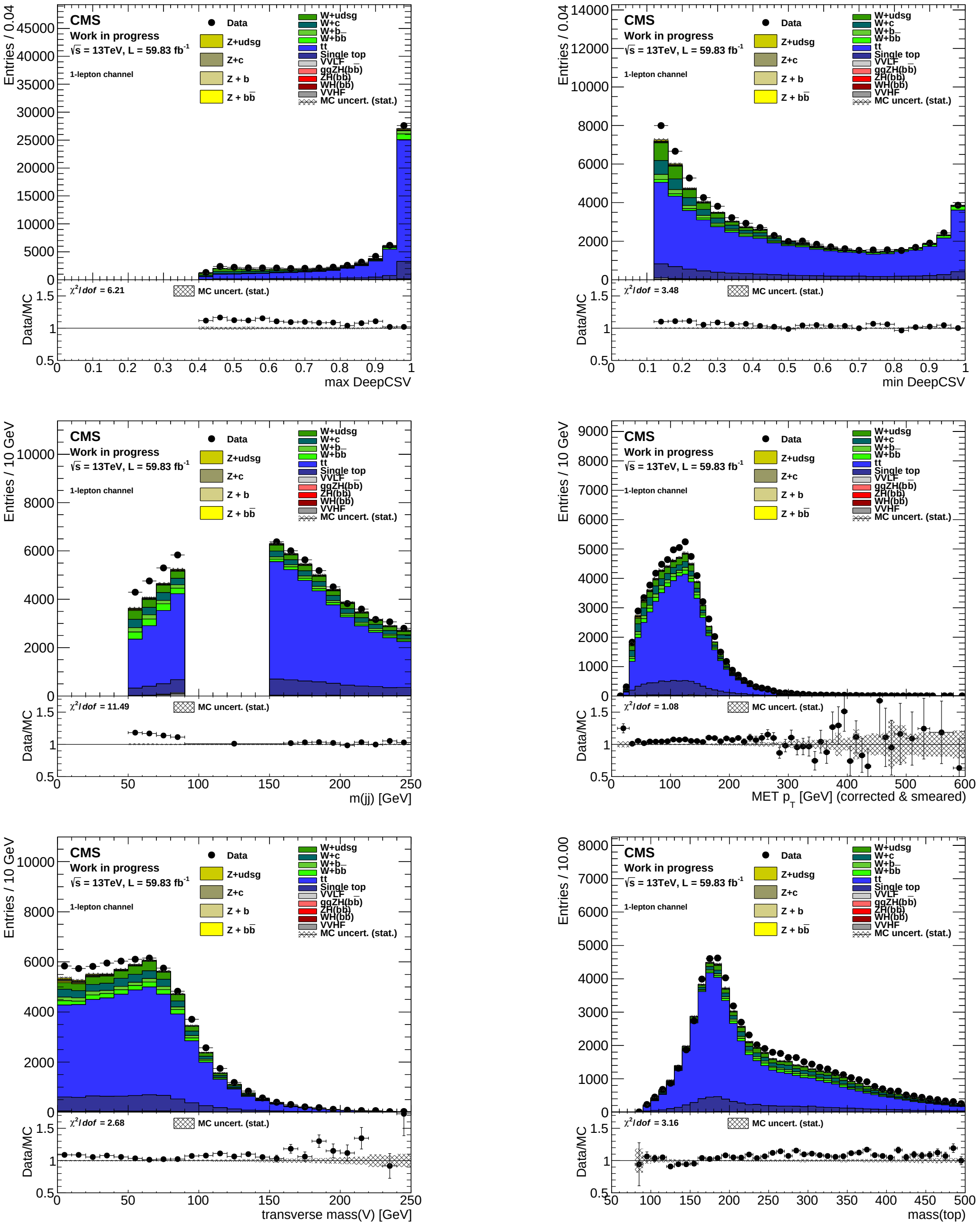


Figure 44: Selected kinematic observables in the resolved 1-lepton HF CR (muon channel): DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), MET (middle right), the transverse mass of the reconstructed vector boson (bottom left) and the reconstructed top mass (bottom right) after the kinematic fit and FSR recovery procedure.

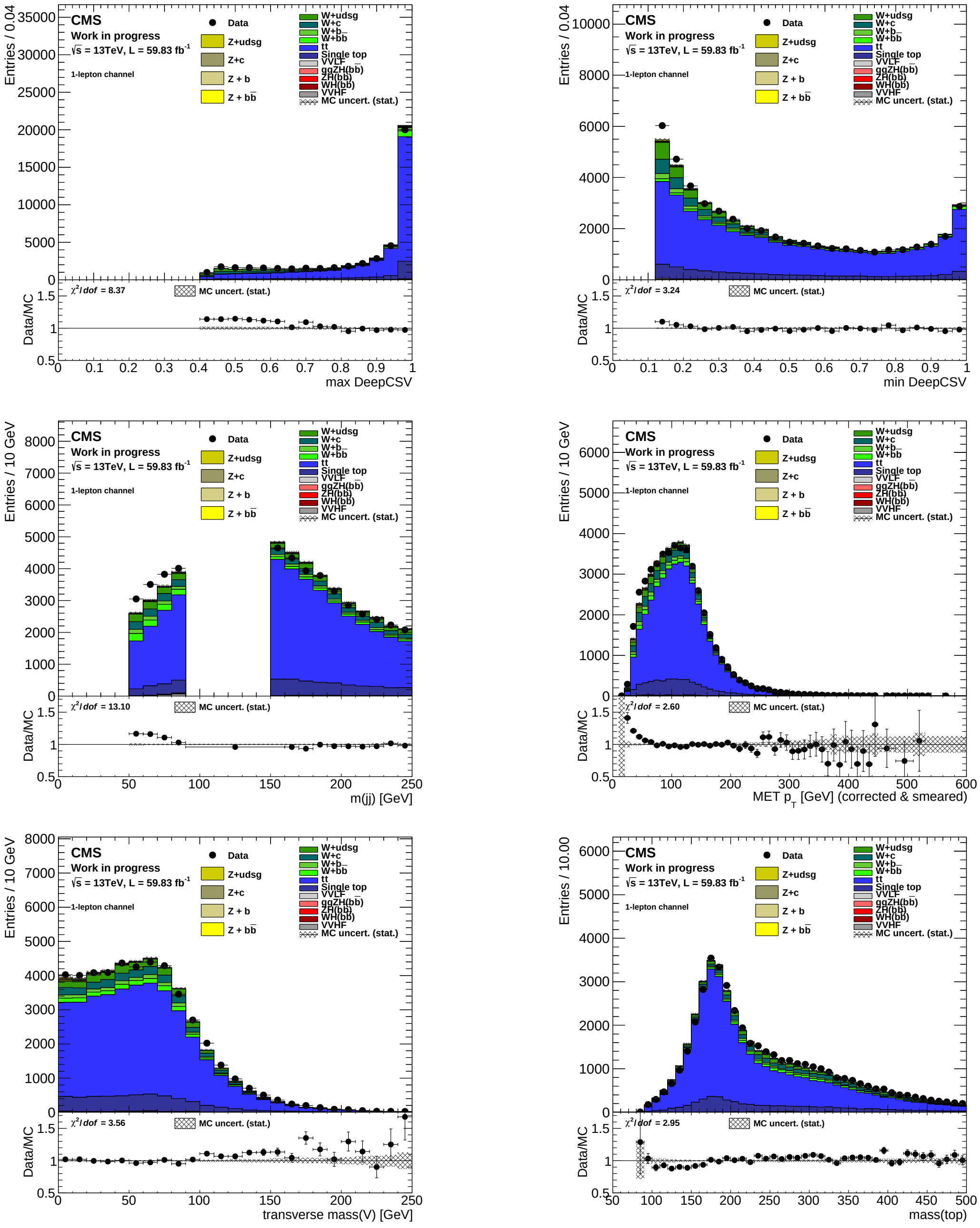


Figure 45: Selected kinematic observables in the resolved 1-lepton HF CR (electron channel): DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), MET (middle right), the transverse mass of the reconstructed vector boson (bottom left) and the reconstructed top mass (bottom right) after the kinematic fit and FSR recovery procedure.

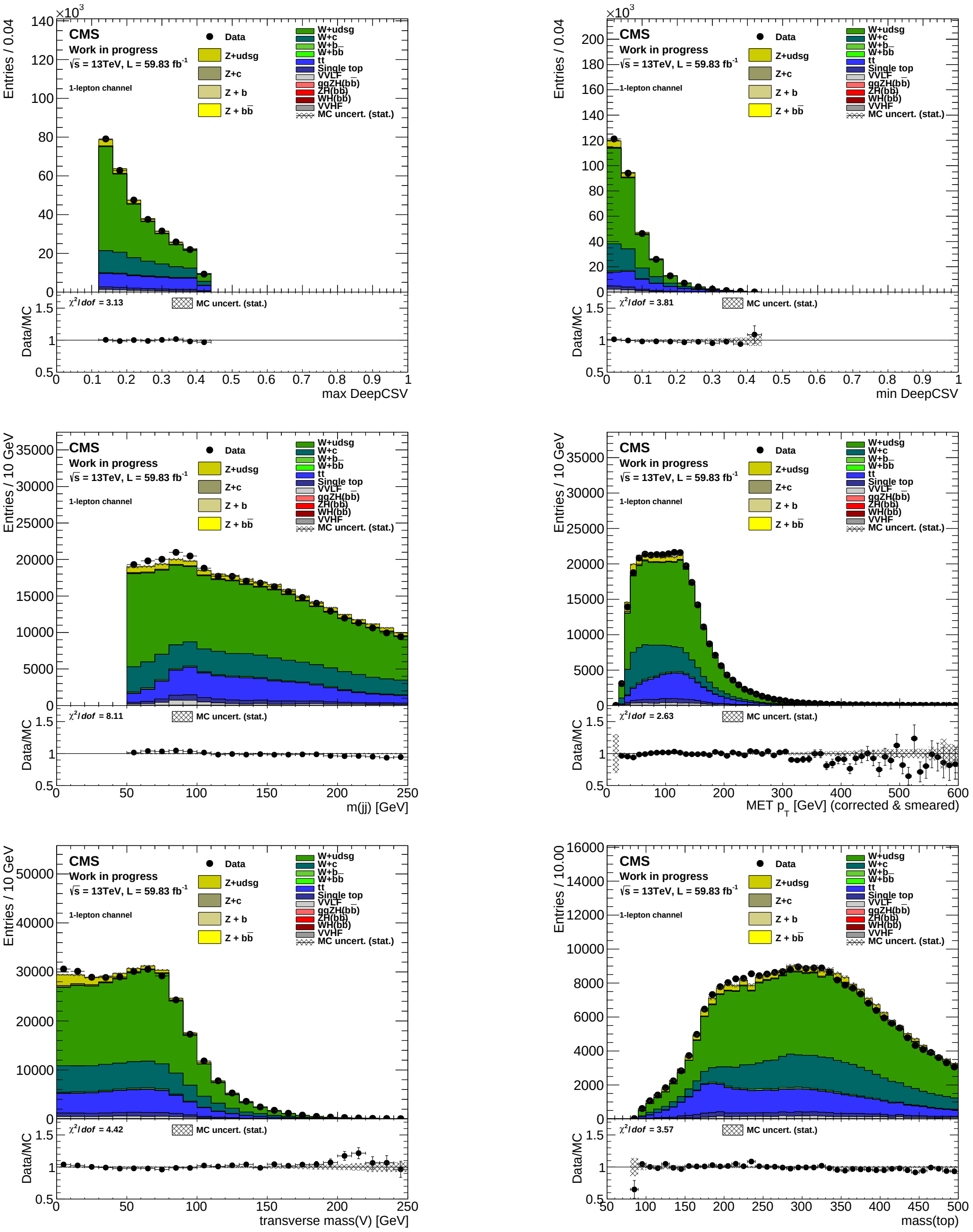


Figure 46: Selected kinematic observables in the resolved 1-lepton LF CR (muon channel): DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), MET (middle right), the transverse mass of the reconstructed vector boson (bottom left) and the reconstructed top mass (bottom right) after the kinematic fit and FSR recovery procedure.

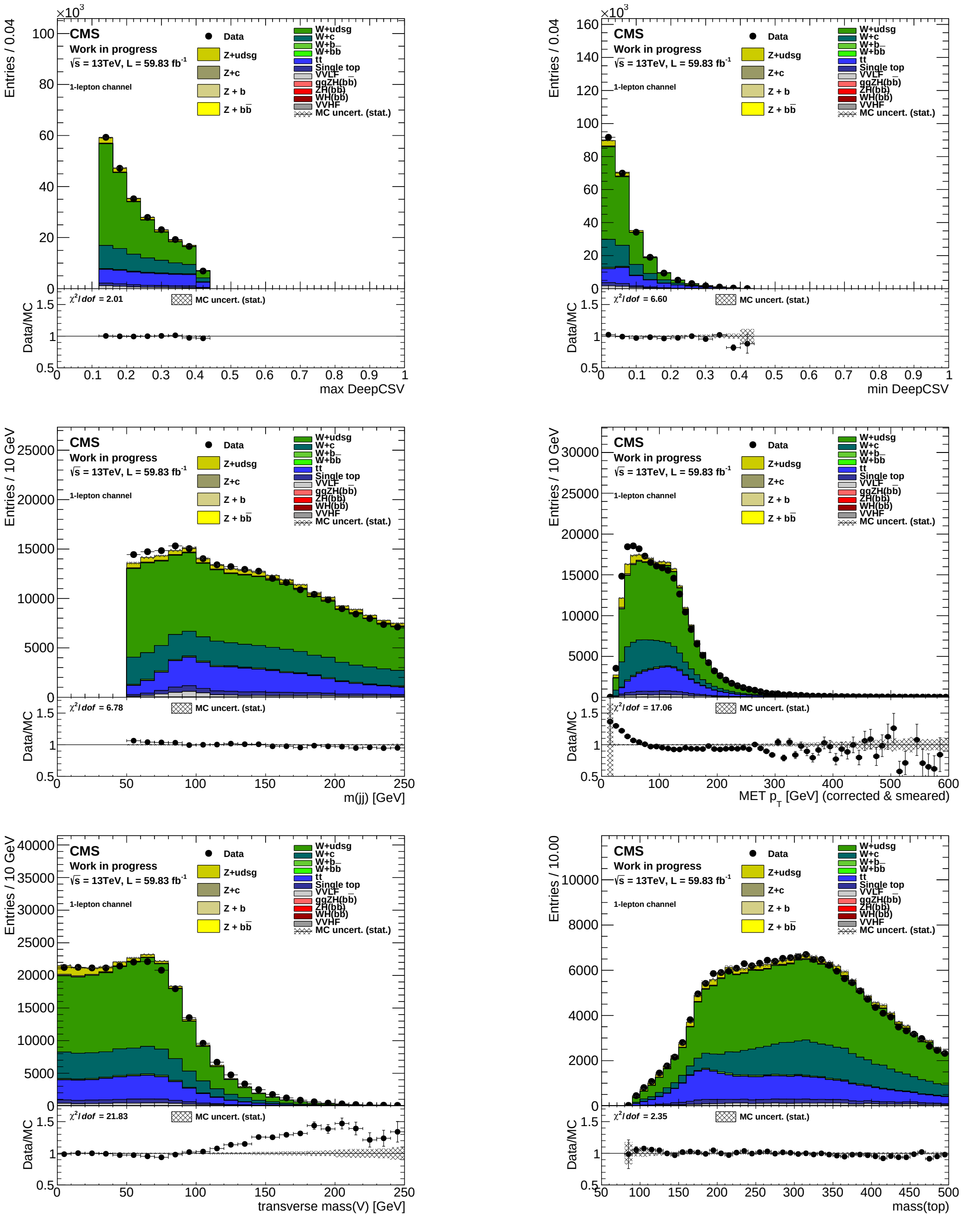


Figure 47: Selected kinematic observables in the resolved 1-lepton LF CR (electron channel): DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), MET (middle right), the transverse mass of the reconstructed vector boson (bottom left) and the reconstructed top mass (bottom right) after the kinematic fit and FSR recovery procedure.

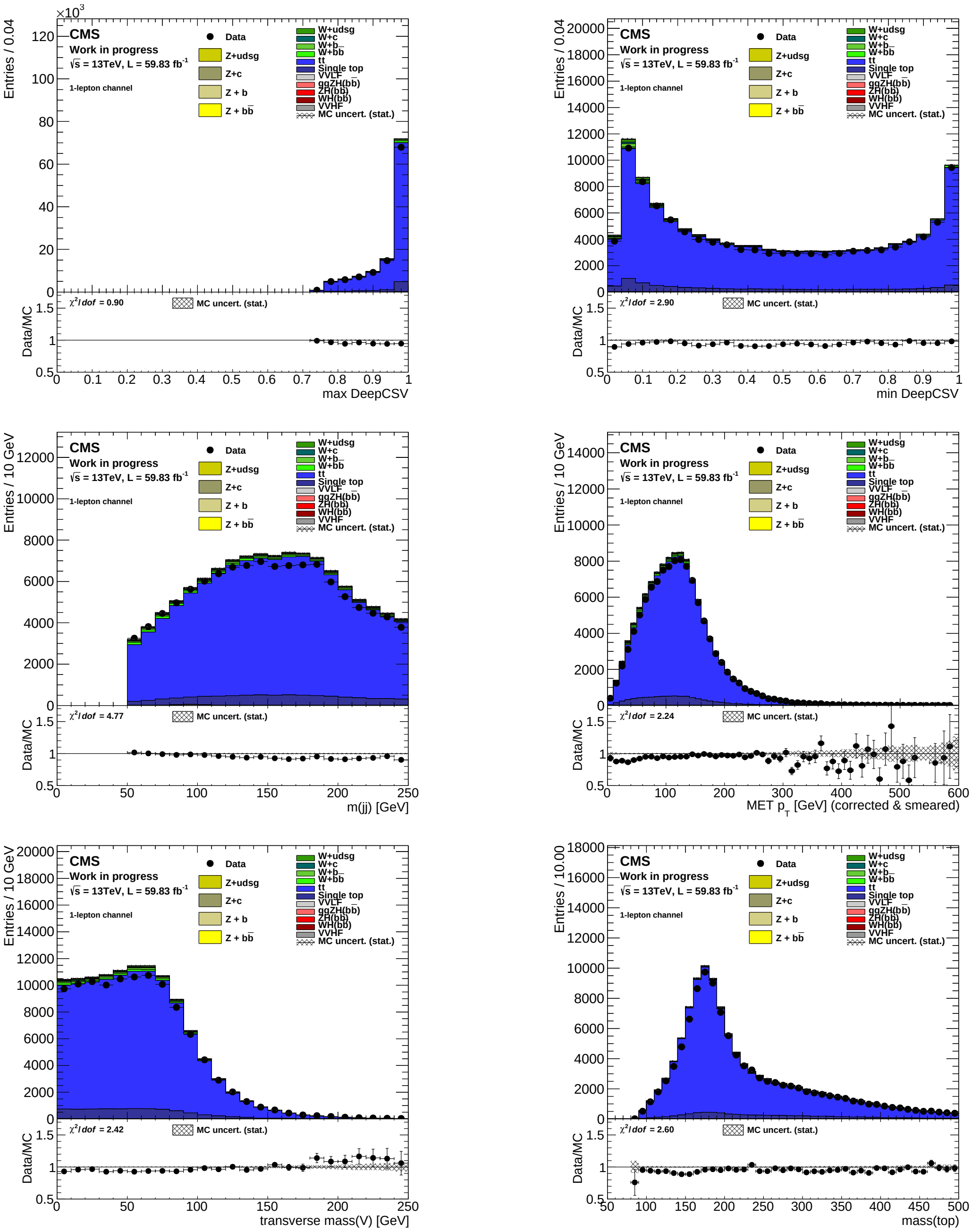


Figure 48: Selected kinematic observables in the resolved 1-lepton $t\bar{t}$ CR (muon channel): DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), MET (middle right), the transverse mass of the reconstructed vector boson (bottom left) and the reconstructed top mass (bottom right) after the kinematic fit and FSR recovery procedure.

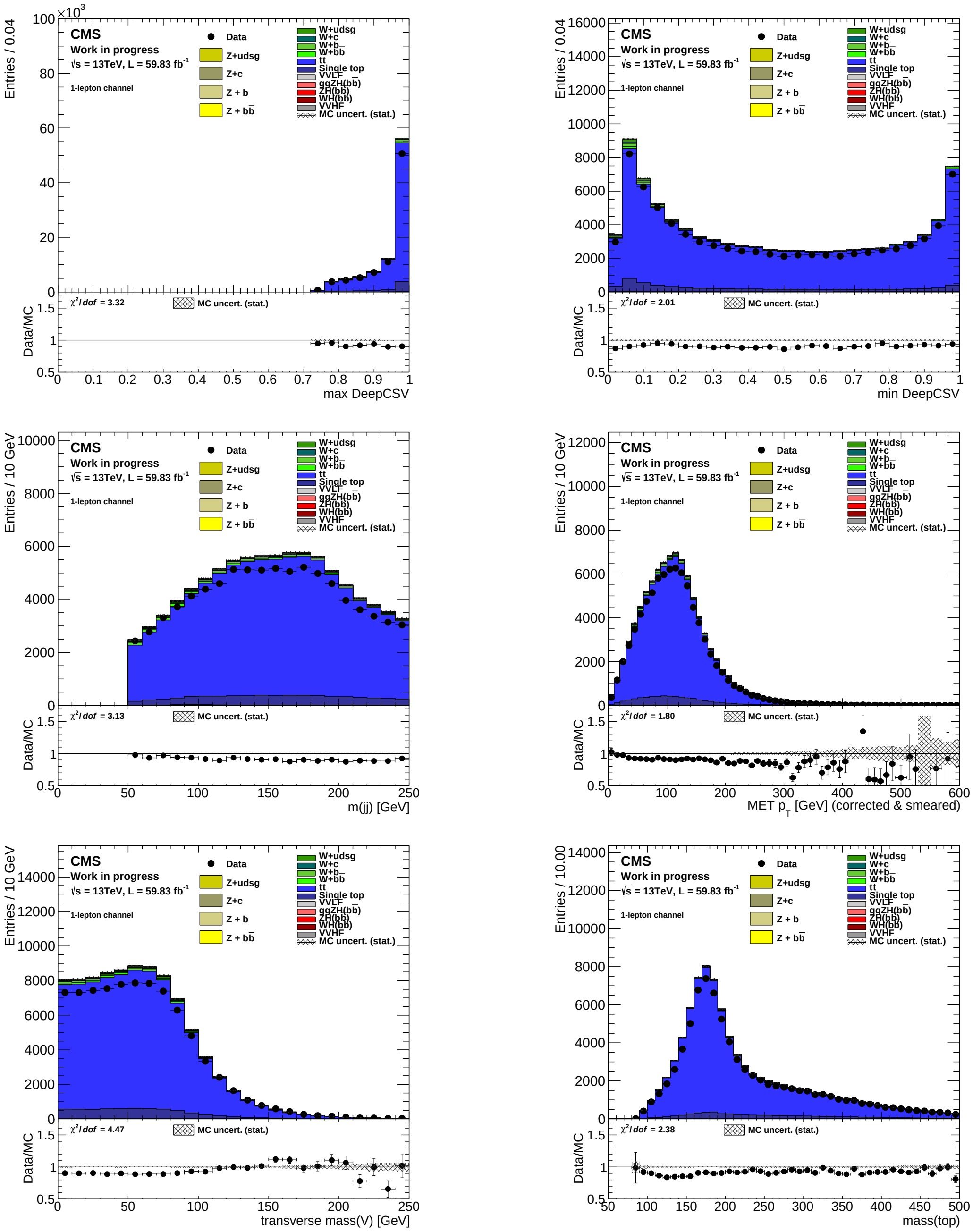


Figure 49: Selected kinematic observables in the resolved 1-lepton t$\bar{t}$ CR (electron channel): DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), MET (middle right), the transverse mass of the reconstructed vector boson (bottom left) and the reconstructed top mass (bottom right) after the kinematic fit and FSR recovery procedure.

3.8.3 2-lepton channel

Events in the 2-lepton channel are characterized by the presence of two isolated leptons from the decay of the Z boson which recoil against the Higgs boson decaying to a pair of bottom quarks. The presence of the two isolated leptons provides a trigger signature for this channel. The Higgs candidate is reconstructed using the two highest b-tagged jets with leading and subleading jets having regressed $p_T > 20$ GeV each. The transverse momentum of the reconstructed vector boson is further required to be more than 75 GeV.

Signal region selection A cut on the dijet mass window and b-tagging requirement is used as in other channels. The complete list of cuts for SR is given in Table 11.

Control region selection The TT CR is obtained by applying a veto to the dilepton mass from the mass window of the Z boson. The complete list of cuts for CR is given in Table 11.

Variable	SR	HF CR	LF CR	TT CR
Common selection between SR and CRs:				
$\text{p}_T^{j_1}$	> 20	-//-	-//-	-//-
$\text{p}_T^{j_2}$	> 20	-//-	-//-	-//-
p_V	> 75	-//-	-//-	-//-
m(jj)	$[50 - 250]$	-//-	-//-	-//-
Different selection between SR and/or CRs:				
DeepCSV(max)	>medium	>medium	<loose	>tight
DeepCSV(min)	>loose	>loose	<loose	>loose
m(V)	$[75 - 105]$	$[85 - 97]$	$[75 - 105]$	$[10 - 75]$ and >120
m(jj)	$[90 - 150]$	$\notin [90 - 150]$	$[90 - 150]$	-
$\Delta\phi$(H,V)	> 2.5	> 2.5	> 2.5	-

Table 11: Definition of the SR and CR for the 2-lepton channel resolved selection. The symbol '-//- ' represents the same selection cut for CRs and SR. The symbol '-' refers to no selection. Mass and momentum have units of GeV. 'Loose', 'medium', and 'tight' refers to the DeepCSV WP. j_1 and j_2 refer to the leading and sub-leading jet in p_T.

The distribution of selected observables in the resolved SRs and CRs of the 2-lepton channel is shown in Figures 50-57.

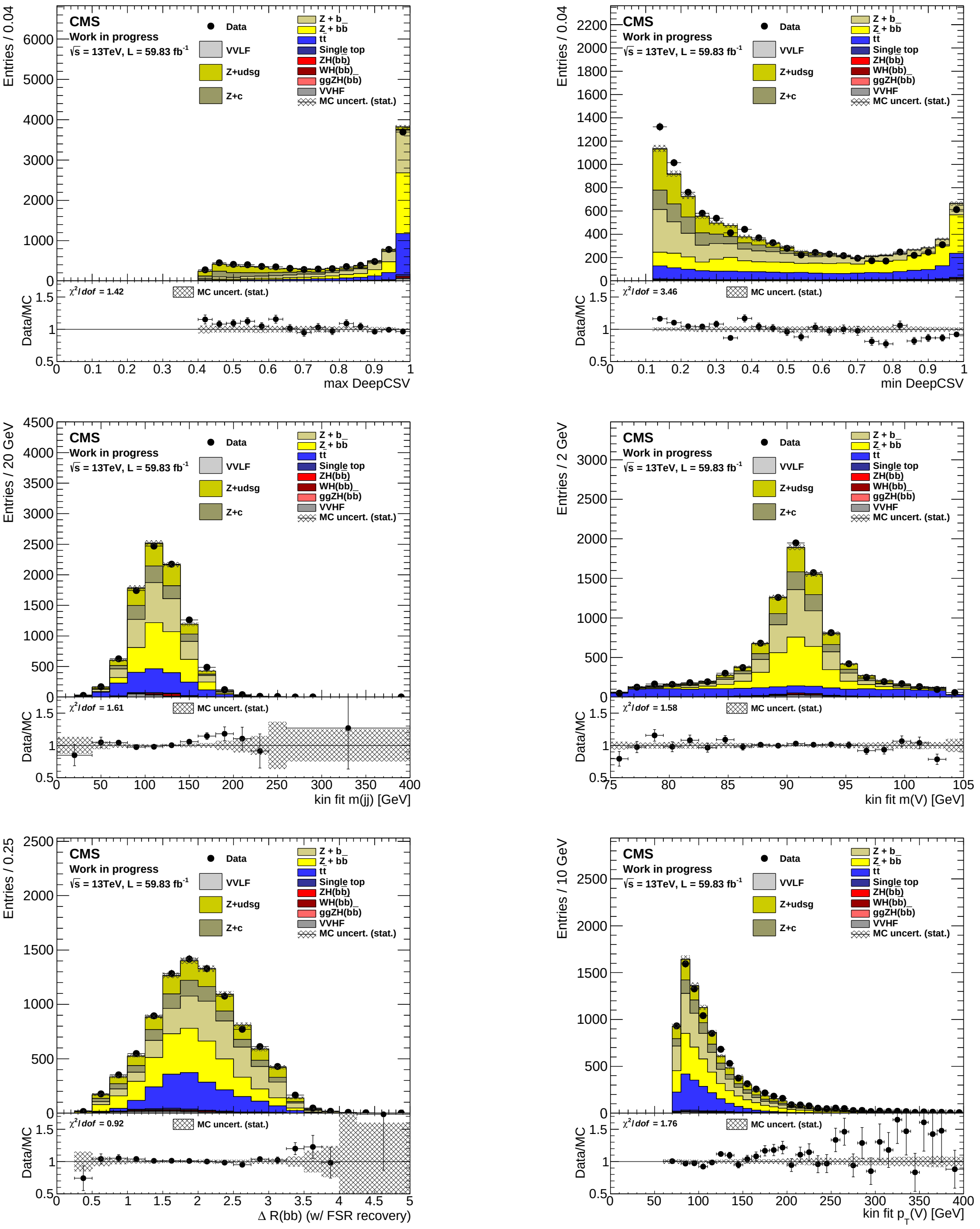


Figure 50: Selected kinematic observables in the resolved 2-lepton SR (muon): DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), reconstructed vector boson mass (middle right), ΔR between the dijets (bottom left) and p_T of the reconstructed vector boson (bottom right) after the kinematic fit and FSR recovery procedure.

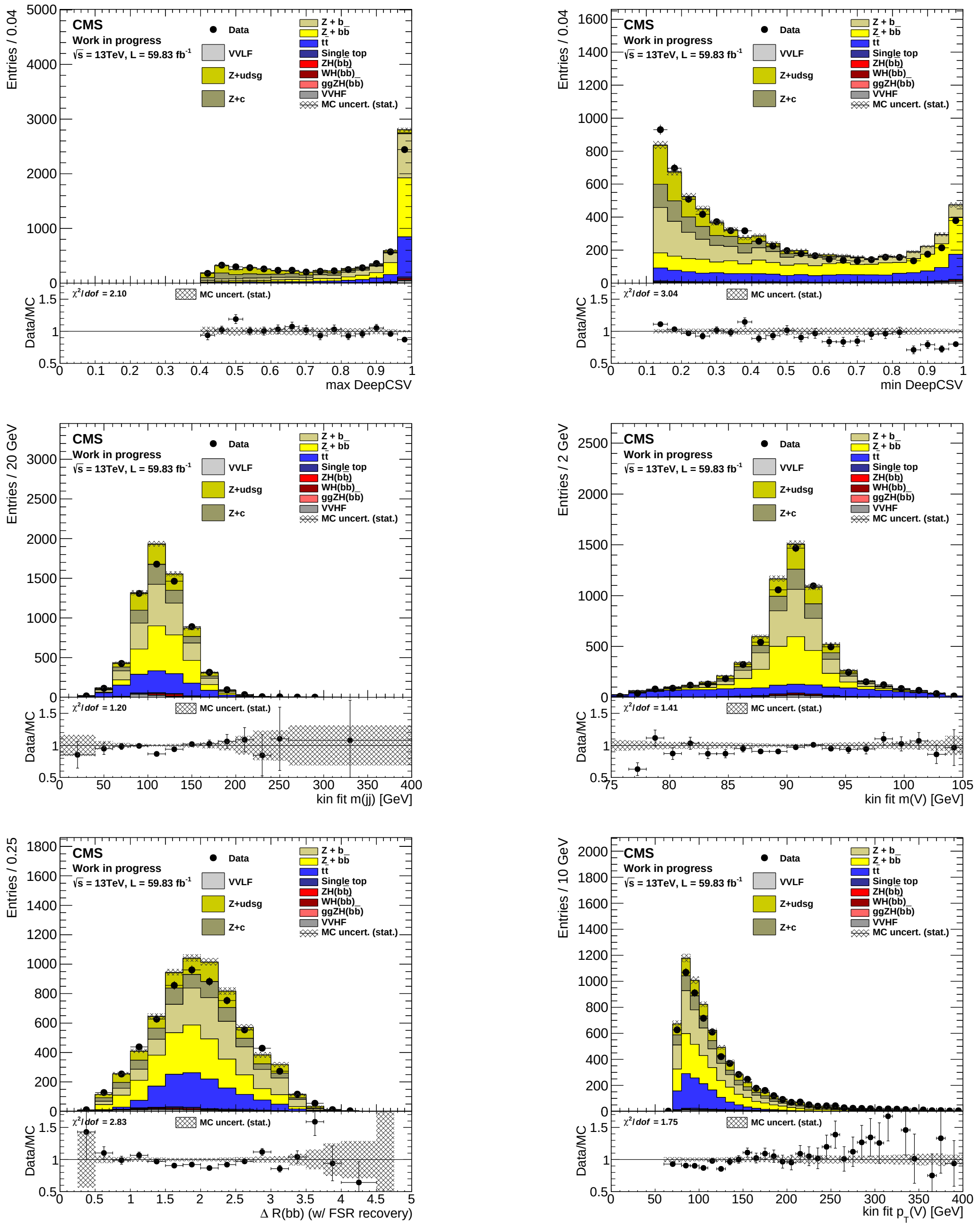


Figure 51: Selected kinematic observables in the resolved 2-lepton SR (electron): DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), reconstructed vector boson mass (middle right), ΔR between the dijets (bottom left) and p_T of the reconstructed vector boson (bottom right) after the kinematic fit and FSR recovery procedure.

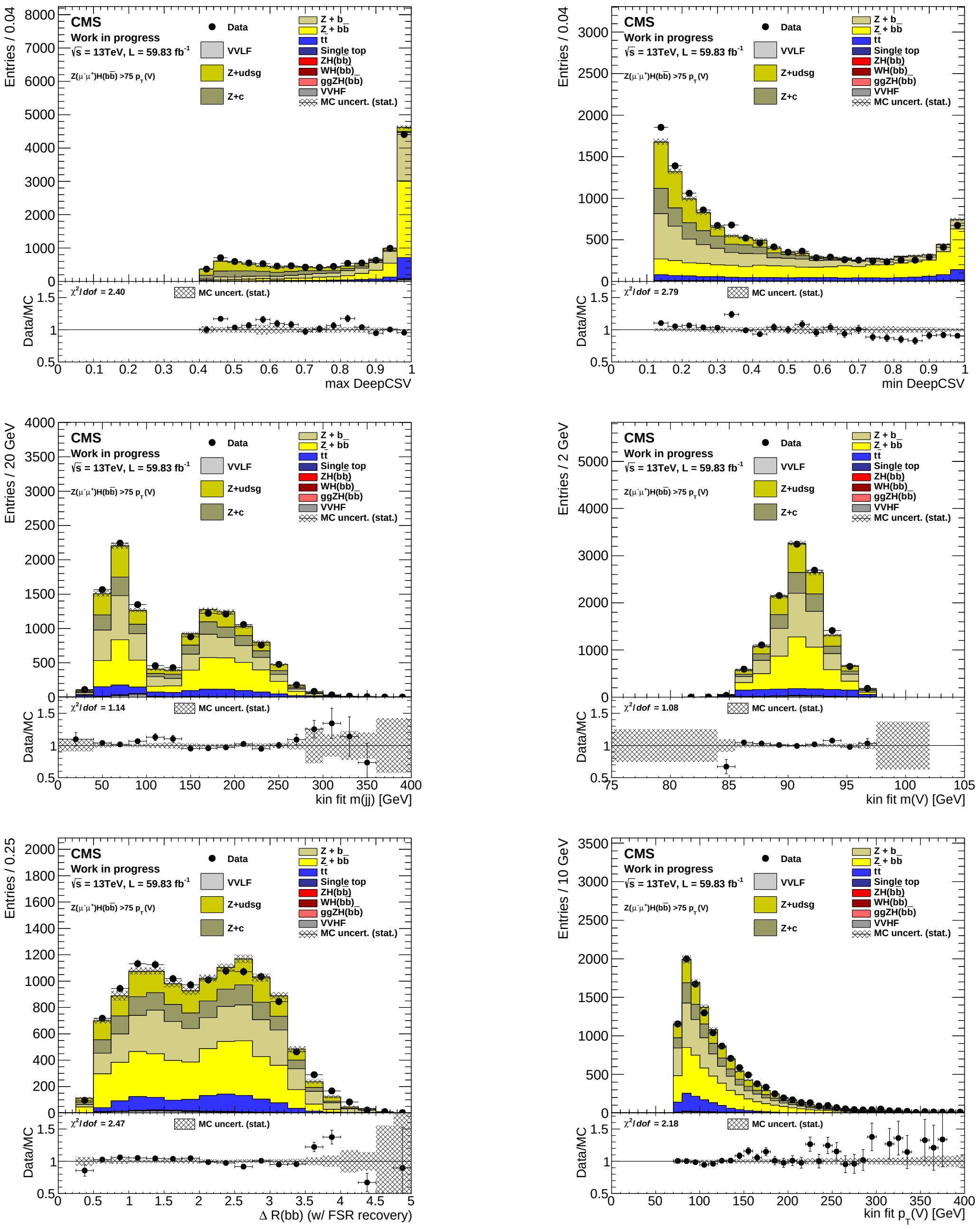


Figure 52: Selected kinematic observables in the resolved 2-lepton HF CR (muon): DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), reconstructed vector boson mass (middle right), ΔR between the dijets (bottom left) and p_T of the reconstructed vector boson (bottom right) after the kinematic fit and FSR recovery procedure.

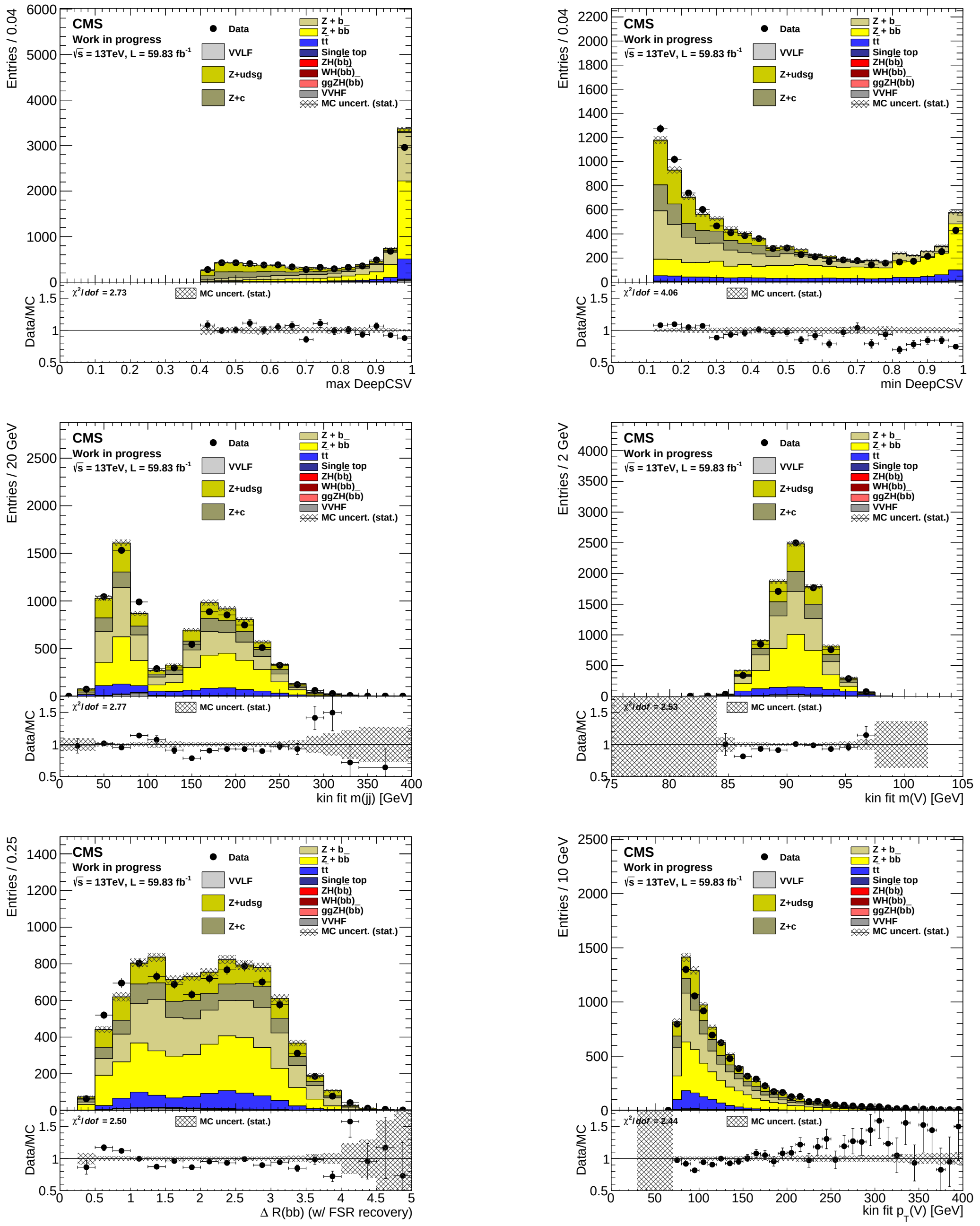


Figure 53: Selected kinematic observables in the resolved 2-lepton HF CR (electron): DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), reconstructed vector boson mass (middle right), ΔR between the dijets (bottom left) and p_T of the reconstructed vector boson (bottom right) after the kinematic fit and FSR recovery procedure.

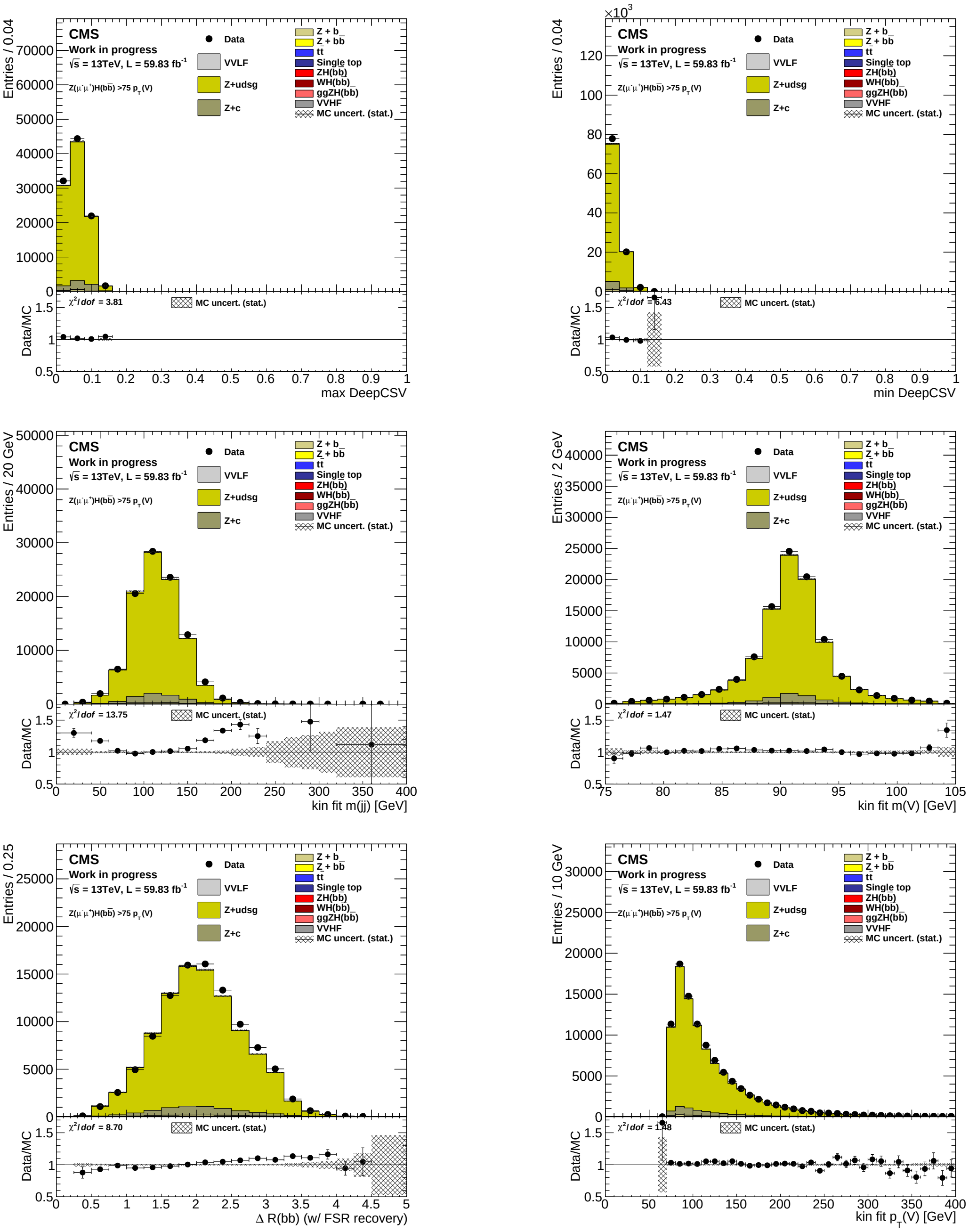


Figure 54: Selected kinematic observables in the resolved 2-lepton LF CR (muon): DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), reconstructed vector boson mass (middle right), ΔR between the dijets (bottom left) and p_T of the reconstructed vector boson (bottom right) after the kinematic fit and FSR recovery procedure.

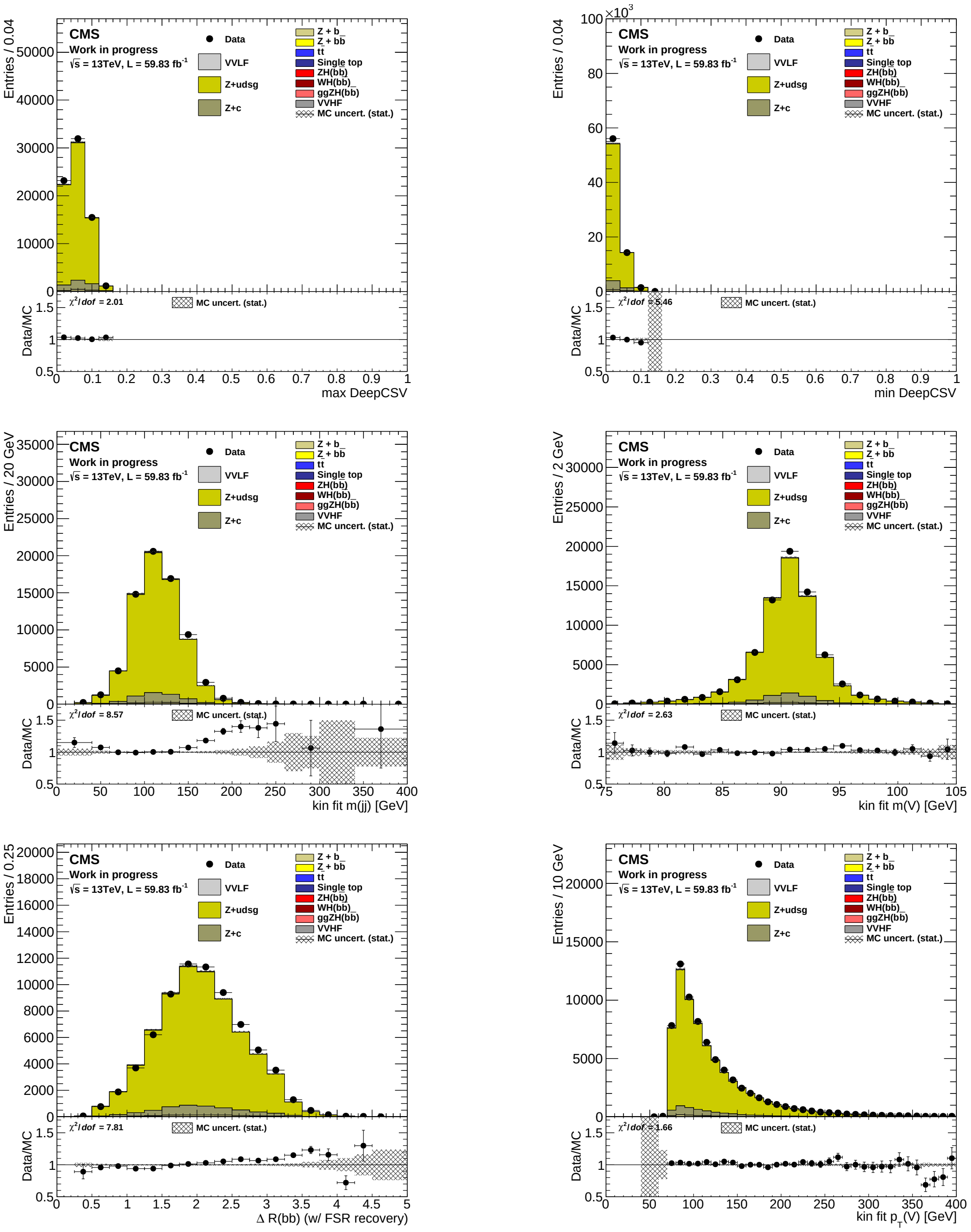


Figure 55: Selected kinematic observables in the resolved 2-lepton LF CR (electron): DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), reconstructed vector boson mass (middle right), ΔR between the dijets (bottom left) and p_T of the reconstructed vector boson (bottom right) after the kinematic fit and FSR recovery procedure.

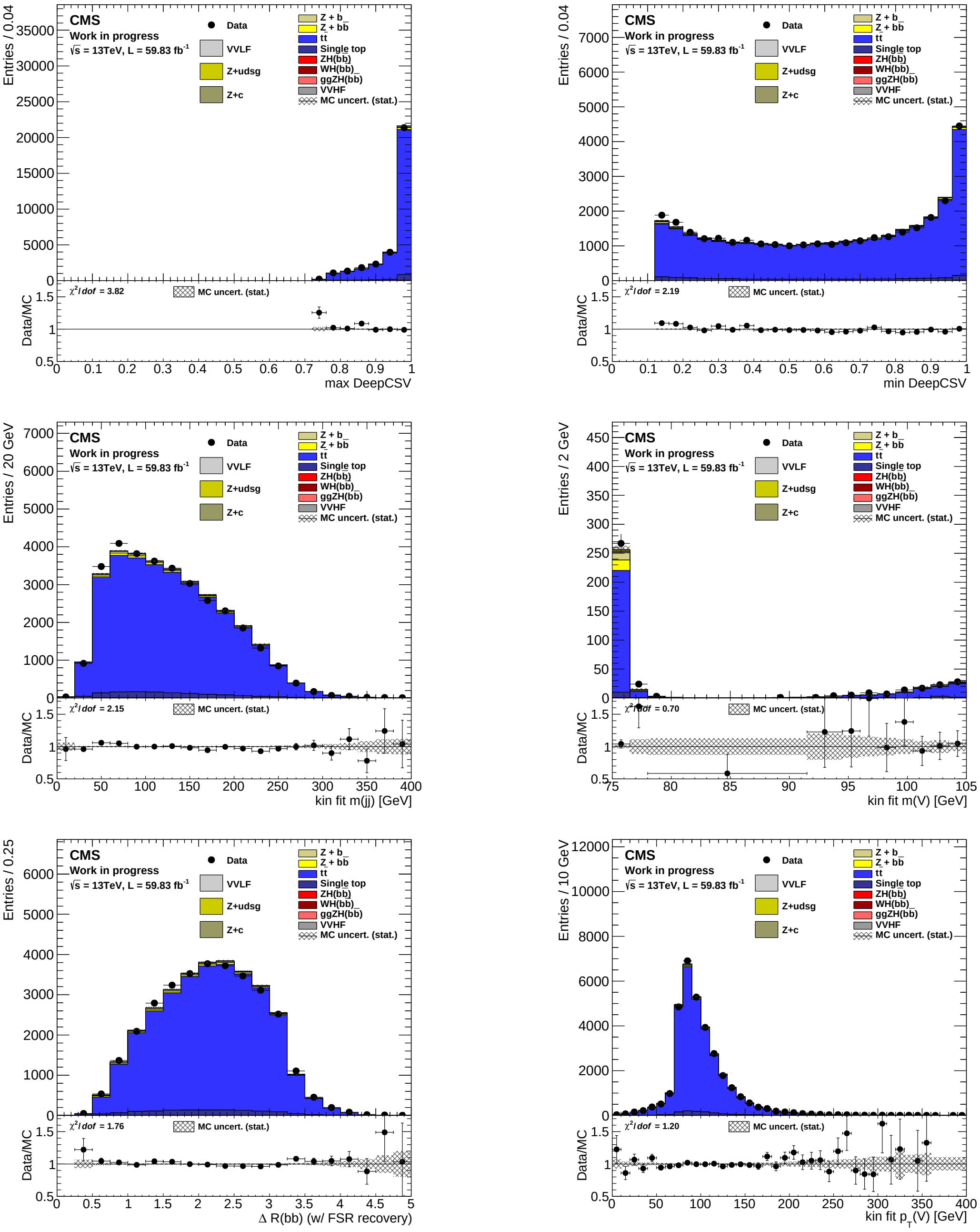


Figure 56: Selected kinematic observables in the resolved 2-lepton $t\bar{t}$ CR (muon): DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), reconstructed vector boson mass (middle right), ΔR between the dijets (bottom left) and p_T of the reconstructed vector boson (bottom right) after the kinematic fit and FSR recovery procedure.

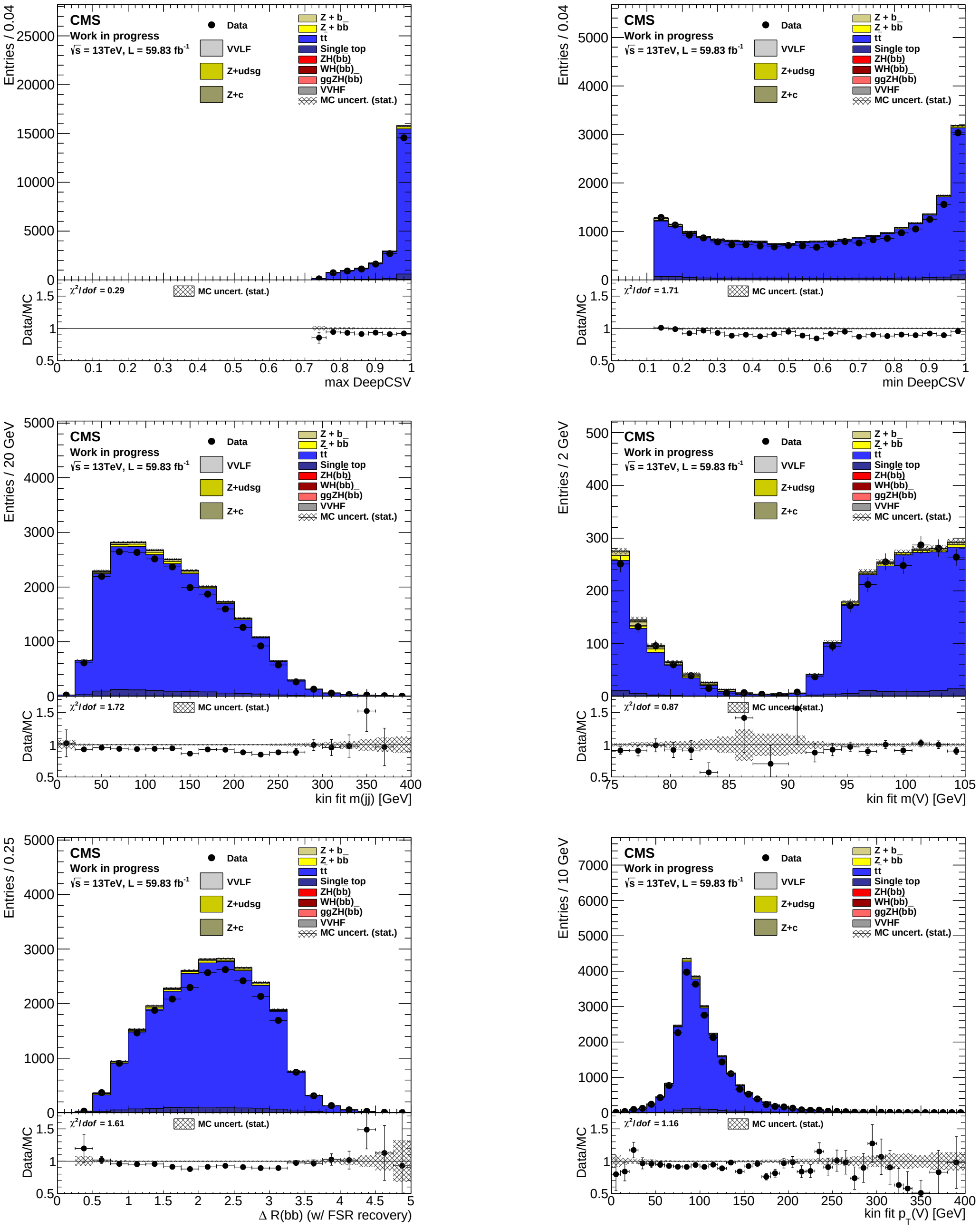


Figure 57: Selected kinematic observables in the resolved 2-lepton $t\bar{t}$ CR (electron): DeepCSV score of the leading jet (top left) and sub-leading jet (top right), dijet mass (middle left), reconstructed vector boson mass (middle right), ΔR between the dijets (bottom left) and p_T of the reconstructed vector boson (bottom right) after the kinematic fit and FSR recovery procedure.

3.8.4 Boosted analysis selection

The boosted topology targets Higgs boson signals with $p_T > 250$ GeV of the recoiling vector boson. The Higgs boson is reconstructed using the AK8 jet (also known as the FatJet) tagged with the highest DeepAK8 score.

As the signal is expected to have a vector boson p_T spectrum harder than the background, the preselection cuts include $p_T(\mathrm{V}) > 250$ GeV, $p_T^J > 250$ GeV (J refers to the FatJet) and soft-drop mass $m_{SD} > 50$ GeV. The harder vector boson p_T spectrum in the boosted signal than the background implies the sensitivity of the boosted topology is promising and therefore is being studied in this analysis.

Since the DeepAK8 algorithm does not have an explicit veto on prompt muons in the event, the presence of an additional muon gets interpreted as a muon from semi-leptonic B decay, thus classifying the event as bb instead of b. To reduce this bias, an additional preselection cut is applied requiring $(\mathrm{FatJet}, \mathrm{V}) > 1.57$. This can avoid cases when the muon from the vector boson can get interpreted as muon from semi-leptonic b-decay leading to incorrect classification of the AK8 FatJet.

The number of additional AK4 b-jets ($\mathrm{N_{aj}}$) in the event is defined by the number of b-tagged jets passing the medium DeepCSV WP, with $p_T > 25$ GeV, the jet-lepton filter and $\Delta\mathrm{R}(\mathrm{FatJet},\ \mathrm{jet}) > 0.8$ in $\eta \in [-2.5, 2.5]$.

The signal region in the boosted analysis is defined by selecting events with a FatJet with DeepAK8bbVSlight score greater than 0.8, soft-drop mass in the Higgs mass window of 90 to 150 GeV, and no additional b-tagged AK4 jets outside the AK8 jet cone or additional isolated leptons. The optimized threshold of 0.8 is chosen to retain most of the signal and thus the DeepAK8 calibration SF (discussed in Section 3.6.6) was computed for that threshold.

The HF CR is obtained by inverting the soft-drop mass window cut, LF CR by inverting the DeepAK8 cut, and TT CR by inverting the number of additional jets requirement cut. Only for the 2-lepton channel, TT CR is obtained by inverting the mass window around the mass of the Z boson. In addition to these cuts, anti-QCD cuts (discussed in Section 3.8.1) are used in the 0-lepton boosted channel. The full list of selection cuts used for the boosted analysis in the 0-lepton, 1-lepton and 2-lepton channels are given in Table 12, 13 and 14 respectively.

Soft-drop mass distribution in 2018 data-taking era A large mis-modelling in the soft-drop mass observable starting around the top mass peak was found in the 2018 datasets as shown in Figure 58 (left). The pileup per particle identification (PUPPI) algorithm [90] was incorrectly used for soft-drop mass calculations in the production of inputs for the 2018 dataset used in the analysis. PUPPI is an alternative algorithm to the CHS algorithm discussed in Section 3.6.5 to correct for in-time pile-up. Due to this, the mass distribution of the top quark is shifted as shown in Figure 58 (left). To correct for this shift, the top mass in MC was shifted to get the well-modelled m_{SD} distribution in MC. The value of the shift was calculated using the χ^2/ndf (number of degrees of freedom) between data and MC in the top mass distribution as a figure of merit for various values of the shift in MC. Based on the χ^2/ndf, the optimal value of the shift was found to be 0.9. The incorrect soft-drop mass was thus multiplied by a factor of 0.9 to obtain the corrected soft-drop mass distribution.

Variable	SR	HF CR	LF CR	TT CR
Common selection between SR and CRs:				
p_T^{miss}	> 250	-//-	-//-	-//-
p_T^{J}	> 250	-//-	-//-	-//-
$\lvert\eta(\text{J})\rvert$	< 2.5	-//-	-//-	-//-
m_{SD}	> 50	-//-	-//-	-//-
H_T^{miss}	> 100	-//-	-//-	-//-
N_{al}	$= 0$	-//-	-//-	-//-
anti-QCD cuts	✓	-//-	-//-	-//-
Different selection between SR and/or CRs:				
DeepAK8bbVsLight	> 0.8	> 0.8	< 0.8	> 0.8
m_{SD}	$\in [90 - 150]$	$\notin [90 - 150]$ or > 250	> 50	> 50
N_{aj}	$= 0$	$= 0$	$= 0$	> 0

Table 12: Definition of the SR and CR for the 0-lepton channel boosted selection. Symbol '-//- ' represents the same selection cut in CRs as mentioned in the SR. Symbol '-' refers to no selection. Mass and momentum have units of GeV. J refers to the FatJet. The subscript J refers to the FatJet. 'anti-QCD cuts' refer to the cuts described in Section 3.8.1.

Variable	SR	HF CR	LF CR	TT CR
Common selection between SR and CRs:				
$p_T(\text{V})$	> 250	-//-	-//-	-//-
p_T^{J}	> 250	-//-	-//-	-//-
$\lvert\eta(\text{J})\rvert$	< 2.5	-//-	-//-	-//-
N_{al}	$= 0$	-//-	-//-	-//-
$\Delta\phi\left(\text{lep}, p_T^{\text{miss}}\right)$	< 2.0	-//-	-//-	-//-
m_{SD}	> 50	-//-	-//-	-//-
Different selection between SR and/or CRs:				
DeepAK8bbVsLight	> 0.8	> 0.8	< 0.8	> 0.8
m_{SD}	$\in [90 - 150]$	$\notin [90 - 150]$ or > 250	> 50	> 50
N_{aj}	$= 0$	$= 0$	$= 0$	> 0

Table 13: Definition of the SR and CR for the 1-lepton channel boosted selection. Symbol '-//- ' represents the same selection cut in CRs as mentioned in the SR. Symbol '-' refers to no selection. Mass and momentum have units of GeV. J refers to the FatJet. The subscript J refers to the FatJet.

Since PUPPI corrections were derived in bins of η and p_T of the FatJet, while the corrections to remove the PUPPI issue were derived inclusively in FatJet kinematics (η and p_T), checks were performed to observe the data/MC modelling of the top mass distribution. This was done in η and p_T bins (250 GeV $< p_T(\text{J}) <$ 400 GeV,

Variable	SR	HF CR	LF CR	TT CR
Common selection between SR and CRs:				
p_T(V)	> 250	-//-	-//-	-//-
p_T^{J}	> 250	-//-	-//-	-//-
$\|\eta(\mathrm{J})\|$	< 2.5	-//-	-//-	-//-
Different selection between SR and/or CRs:				
DeepAK8bbVsLight	> 0.8	> 0.8	< 0.8	> 0.8
m_{SD}	$\in [90 - 150]$	$\in [50 - 90]$ or $[150 - 250]$	$\in [90 - 150]$	> 50
m(V)	$\in [75 - 105]$	$\in [75 - 105]$	$\in [75 - 105]$	$\notin [75 - 105]$

Table 14: Definition of the SR and CR for the 2-lepton channel boosted selection. Symbol '-//- ' represents the same selection cut in CRs as mentioned in the SR. Symbol '-' refers to no selection. Mass and momentum have units of GeV. J refers to the FatJet. The subscript J refers to the FatJet.

p_T(J) > 400 GeV, $|\eta(\mathrm{J})| < 1.3$, $|\eta(\mathrm{J})| > 1.3$) where the top mass was scaled using the inclusively derived 0.9 scaling factor. It was found that a maximum of 0.03 scaling of the soft-drop mass was required in addition to the scaling factor of 0.9 to obtain the data/MC modelling of the top mass distribution in the η and p_T bins. Thus, for the 2018 datasets, the soft-drop mass was scaled by a factor of 0.9 with a constant uncertainty of 3% implemented as a log-normal prior in the fit model discussed in Section 3.13.

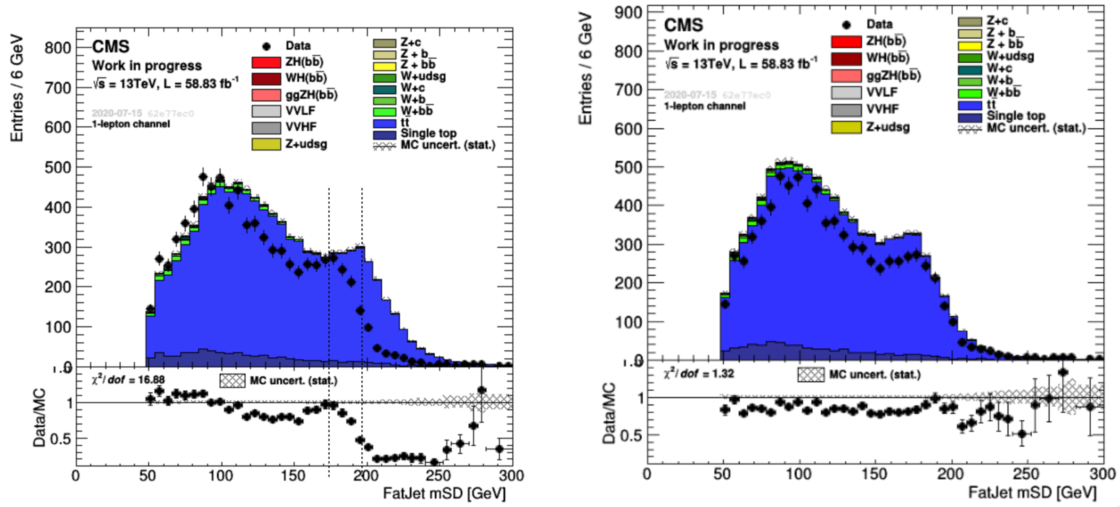


Figure 58: Soft-drop mass in the TT CR of the 1-lepton channel before (left) and after (right) correcting the shift in the top mass peak. In the left plot, the first dashed line refers to the position of the reconstructed mass of the top quark in data while the second dashed line refers to the peak of the distribution in MC corresponding to the top mass.

Figure 58 shows the soft-drop mass observable before (left) and after (right) this scaling factor of 0.9 in the TT CR of the 1-lepton channel.

The distribution of selected observables in the boosted SRs and CRs of different channels is shown in Figures 59-55.

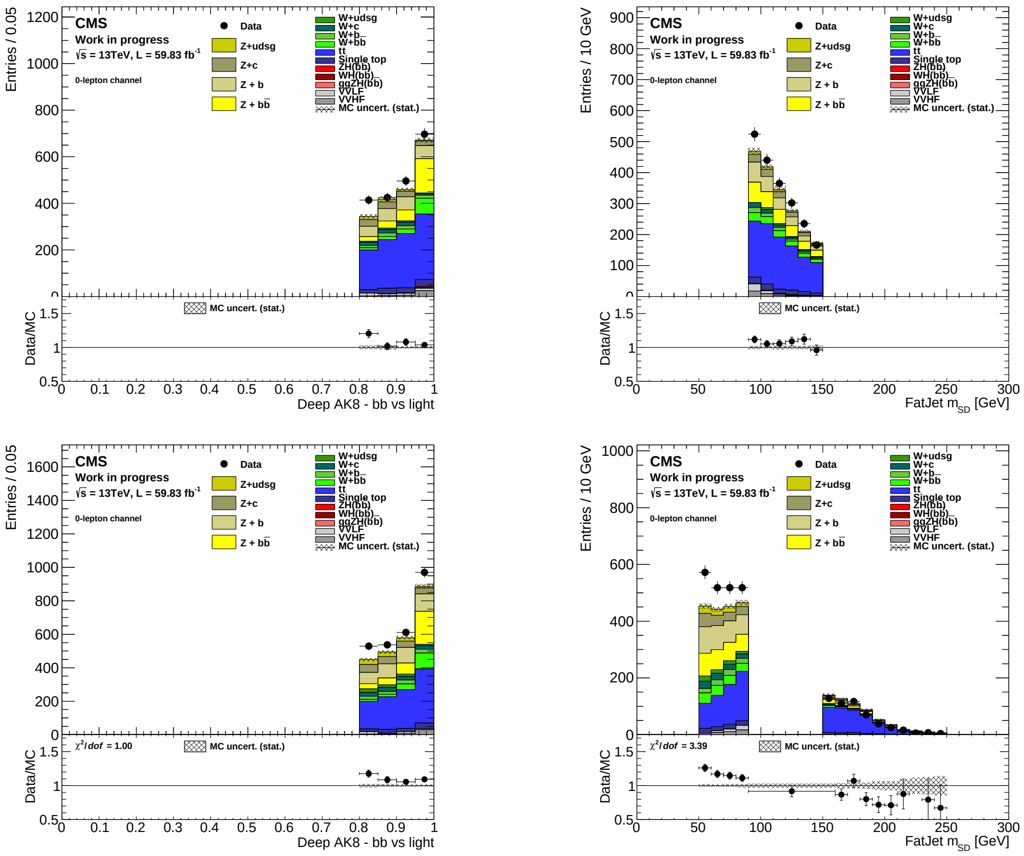


Figure 59: DeepAK8bbVSlight score (left) and the soft-drop mass (right) of the FatJet in the boosted 0-lepton SR (top) and HF CR (bottom).

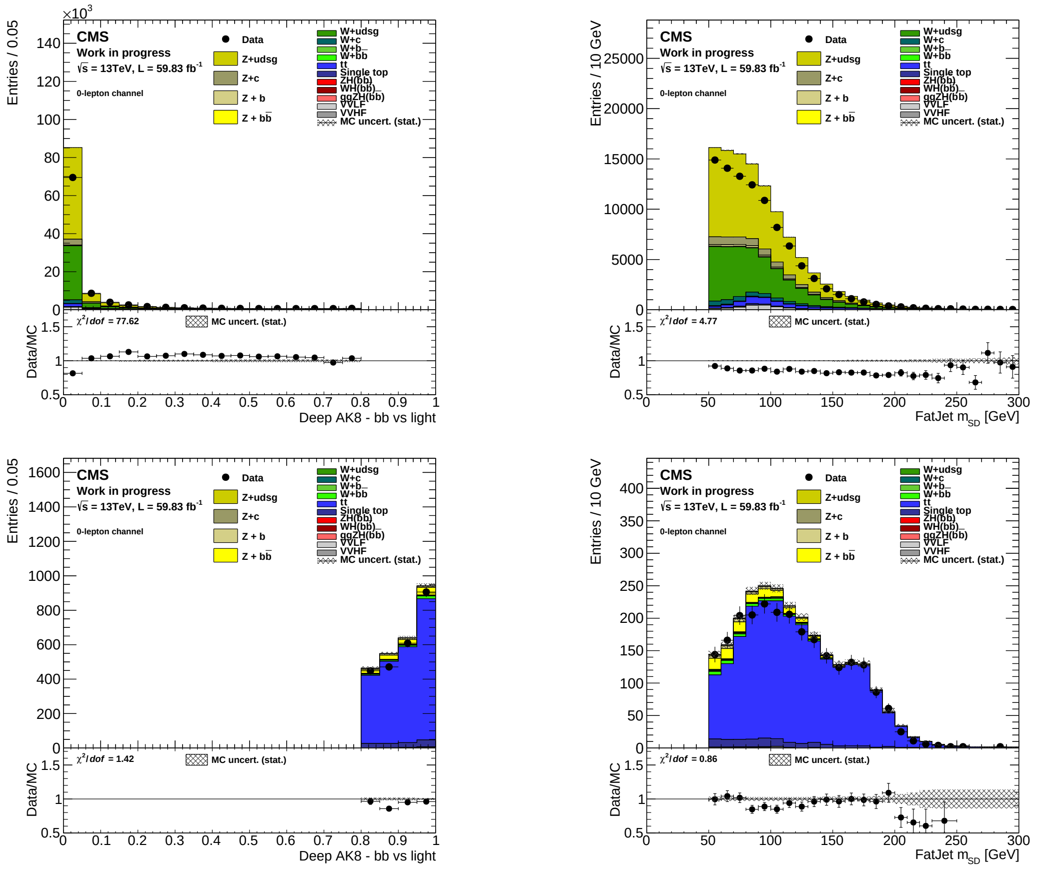


Figure 60: DeepAK8bbVSlight score (left) and the soft-drop mass (right) of the FatJet in the boosted 0-lepton LF CR (top) and $t\bar{t}$ CR (bottom)

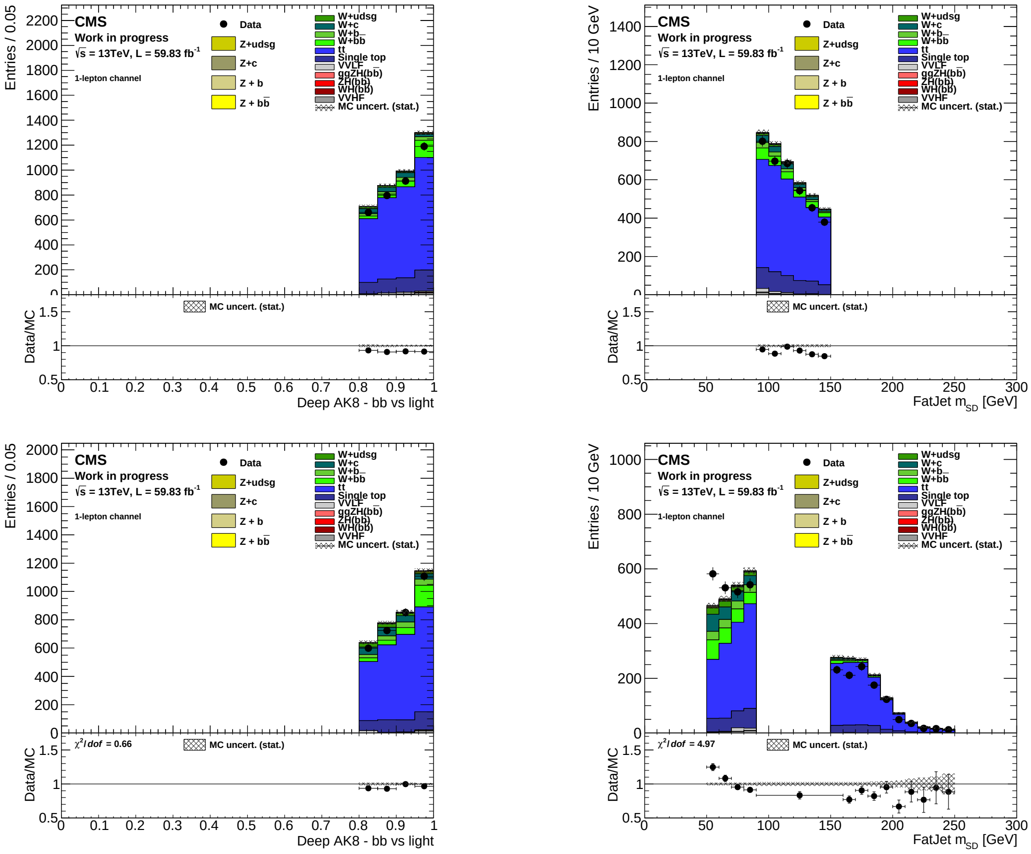


Figure 61: DeepAK8bbVSlight score (left) and the soft-drop mass (right) of the FatJet in the boosted 1-lepton SR (top) and HF CR (bottom) in the muon channel.

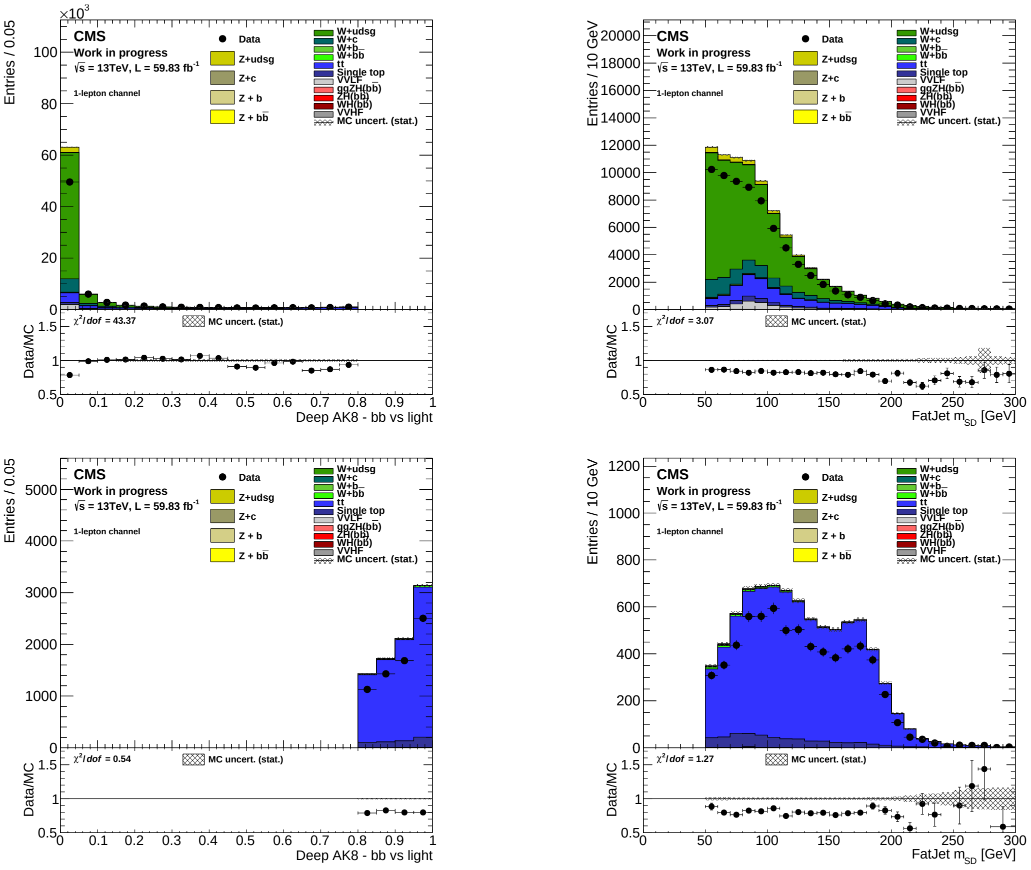


Figure 62: DeepAK8bbVSlight score (left) and the soft-drop mass (right) of the FatJet in the boosted 1-lepton LF CR (top) and $t\bar{t}$ CR (bottom) in the muon channel.

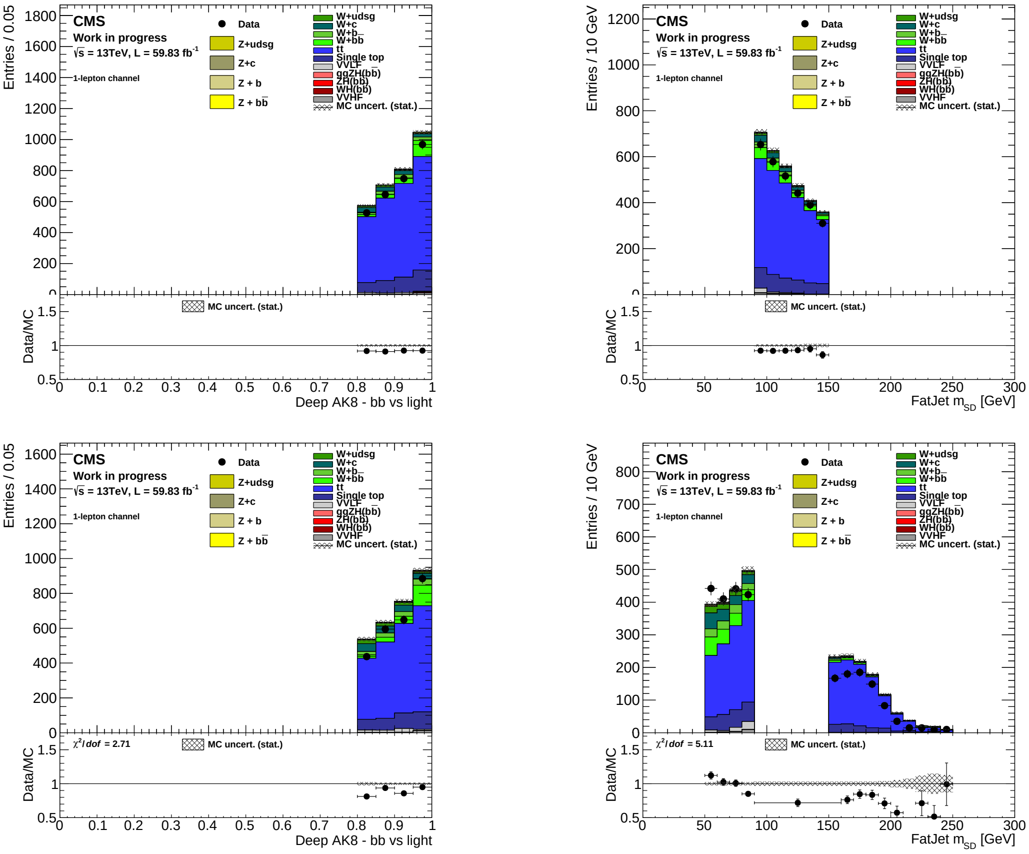


Figure 63: DeepAK8bbVSlight score (left) and the soft-drop mass (right) of the FatJet in the boosted 1-lepton SR (top) and HF CR (bottom) in the electron channel.

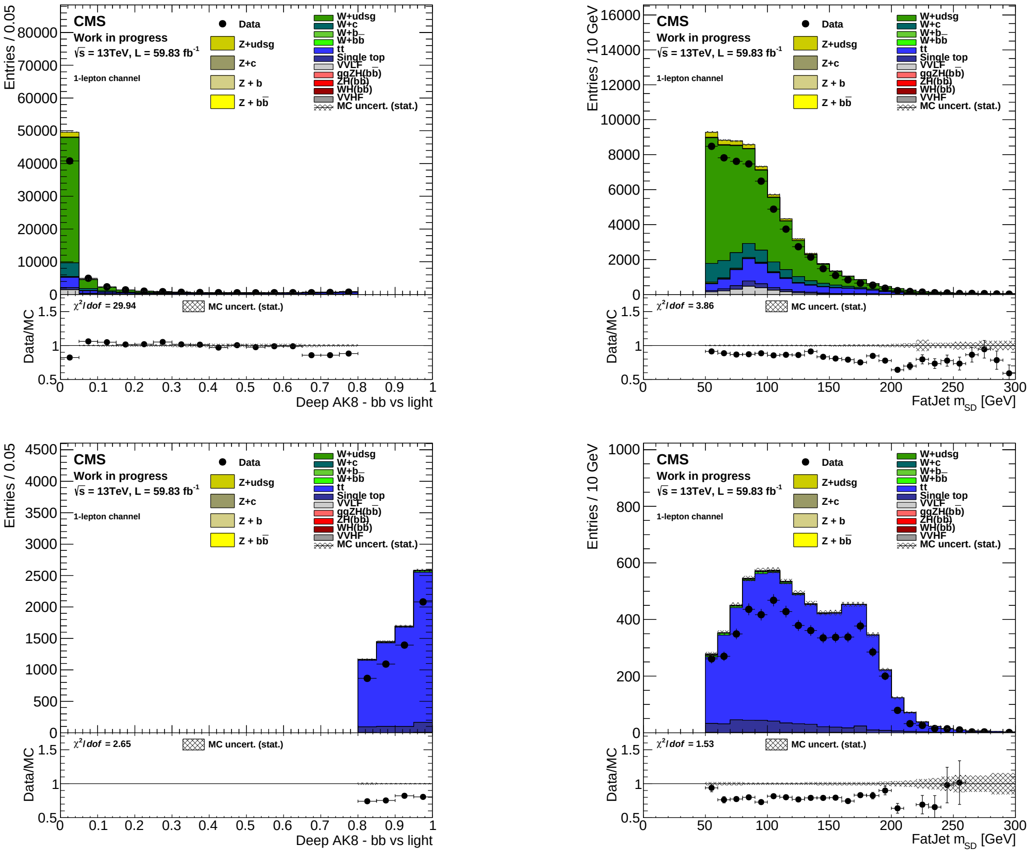


Figure 64: DeepAK8bbVSlight score (left) and the soft-drop mass (right) of the FatJet in the boosted 1-lepton LF CR (top) and $t\bar{t}$ CR (bottom) in the electron channel.

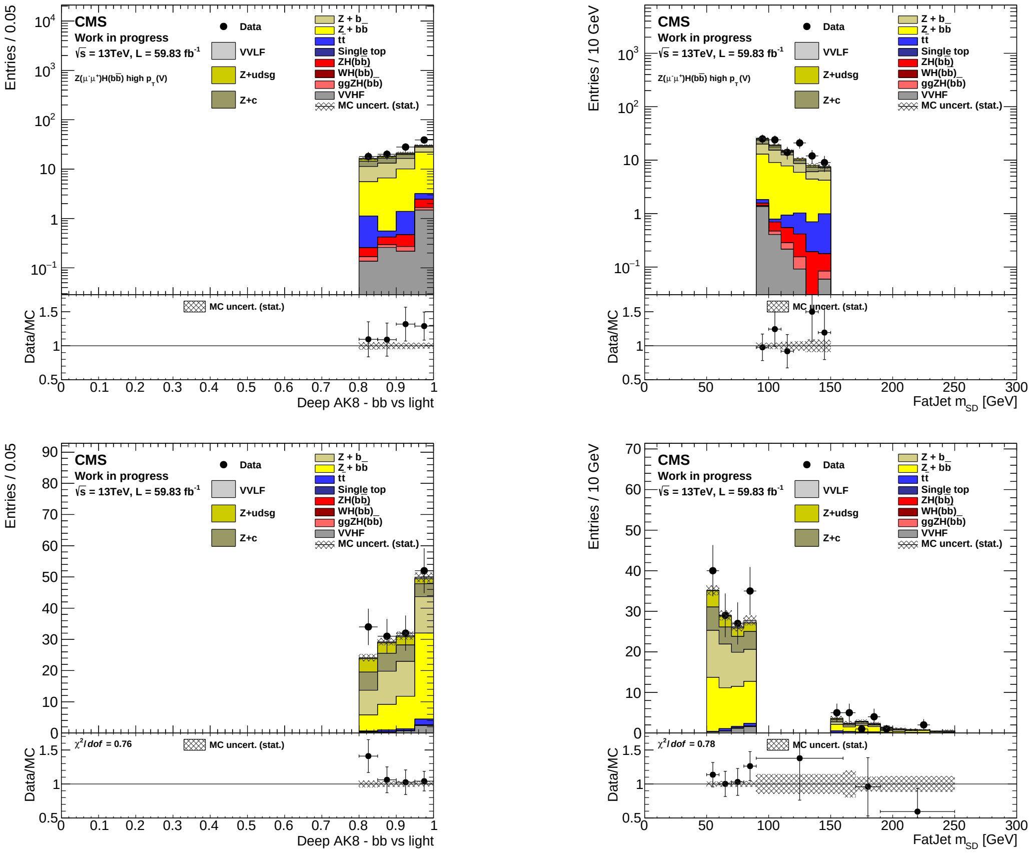


Figure 65: DeepAK8bbVSlight score (left) and the soft-drop mass (right) of the FatJet in the boosted 2-lepton SR (top) and HF CR (bottom) in the muon channel.

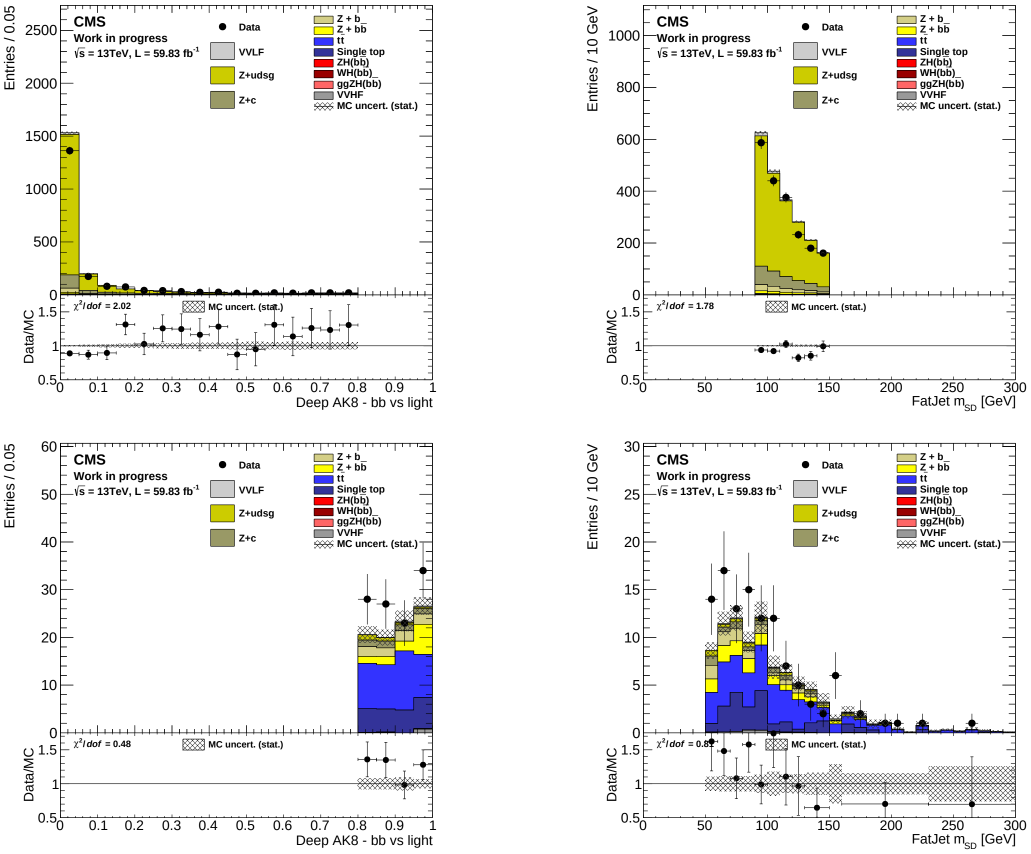


Figure 66: DeepAK8bbVSlight score (left) and the soft-drop mass (right) of the FatJet in the boosted 2-lepton LF CR (top) and $t\bar{t}$ CR (bottom) in the muon channel.

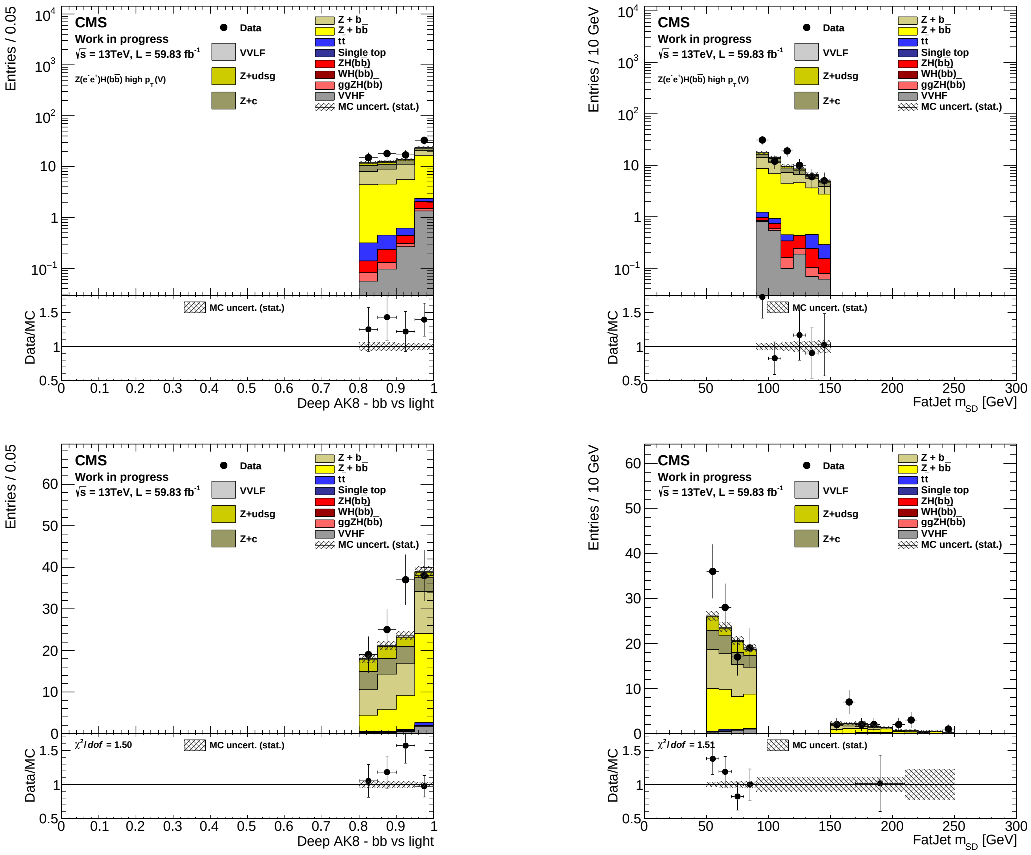


Figure 67: DeepAK8bbVSlight score (left) and the soft-drop mass (right) of the FatJet in the boosted 2-lepton SR (top) and HF CR (bottom) in the electron channel.

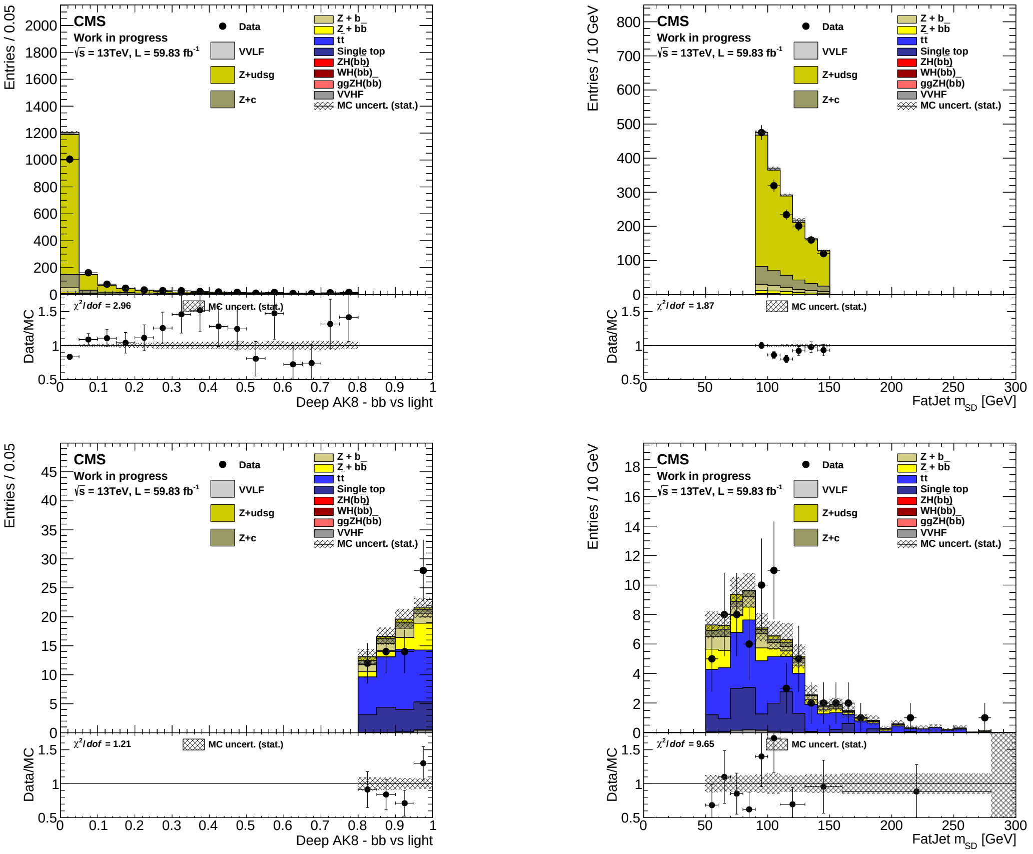
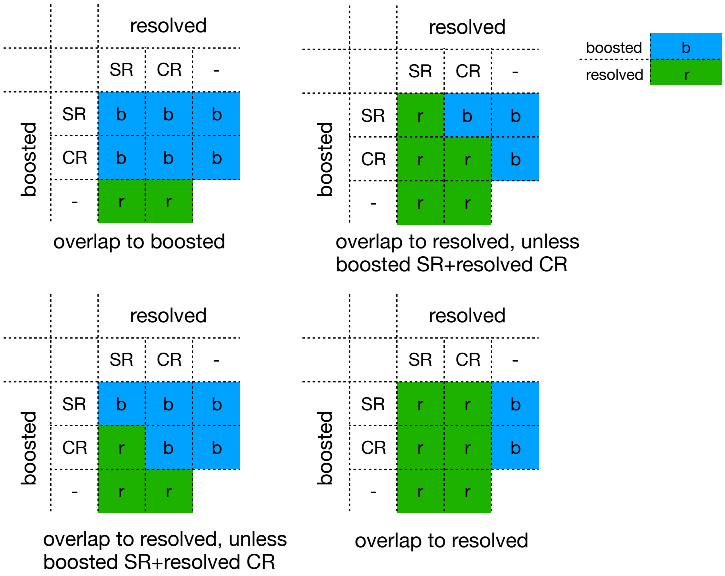


Figure 68: DeepAK8bbVSlight score (left) and the soft-drop mass (right) of the FatJet in the boosted 2-lepton LF CR (top) and $t\bar{t}$ CR (bottom) in the electron channel.

3.8.5 Overlap region between resolved and boosted selections

Some events with $p_T(V) > 250$ GeV pass both the resolved and boosted selections. This overlap needs to be solved to avoid double counting. For this, four schemes of event overlap reassignment were compared. The expected (Asimov) uncertainty on the inclusive signal strength was used as a figure of merit to decide the best scheme.



overlap to boosted

overlap to resolved, unless boosted SR+resolved CR

overlap to resolved, unless boosted SR+resolved CR

overlap to resolved

Figure 69: Different schemes of assigning overlapping events from resolved and boosted topology. The symbol '-' next to CR refers to the events that have only resolved or boosted topology (i.e. the non-overlapping events). Image from [91].

In scheme 1 (top-left of Figure 69), the overlapping events were assigned to the boosted regions; in scheme 2 (top-right of Figure 69) the overlapping events were assigned to the resolved regions unless the event was categorized in the boosted SR; in scheme 3 (bottom-left of Figure 69) the overlapping events were assigned to the boosted regions unless the event was categorized in the resolved SR; and in the scheme 4 (bottom-right of Figure 69) the overlapping events were assigned to the resolved regions. Table 15 shows the expected uncertainty on the signal strength for the STXS bin targeting the $p_T(V) > 250$ GeV for each of the four schemes. Scheme 2 is used for overlap events treatment in this analysis since it has the smallest uncertainty on the signal strength.

	scheme 1	scheme2	scheme 3	scheme 4
ZH p_{T}(V) > 250	0.59	0.45	0.57	0.47
WHp_{T}(V) > 250	0.69	0.47	0.67	0.49

Table 15: Expected uncertainty on the signal strength for the STXS bin targeting the $p_T(V) > 250$ GeV for schemes 1-4 as discussed in Section 3.8.5.

3.9 Simplified Template Cross Section (STXS) framework

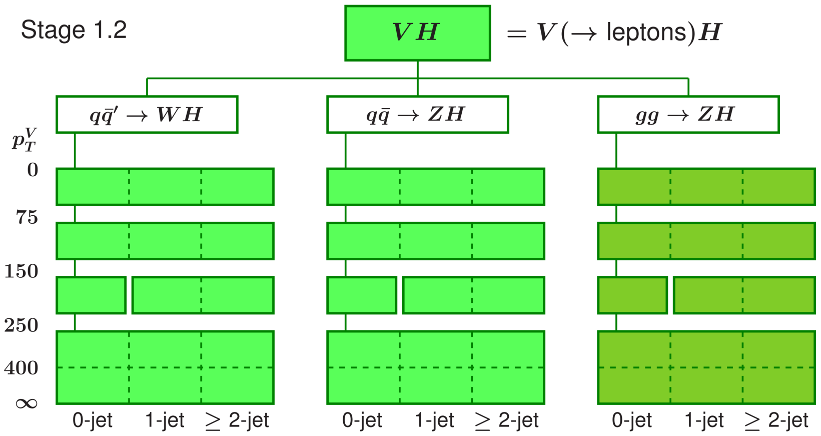

Stage 1.2
VH
$= V(\rightarrow \text{leptons})H$
$q\bar{q}' \rightarrow WH$
$q\bar{q} \rightarrow ZH$
$gg \rightarrow ZH$
p_T^V
0
75
150
250
400
∞
0-jet
1-jet
$\geq$ 2-jet

Figure 70: STXS scheme version 1.2 for Higgs boson production in association with leptonically decaying vector boson [92].

The (inclusive) Higgs boson measurements using the κ-framework [93] have been used to test deviations of the inclusive measurements from the SM. In this framework, a set of coupling modifiers κ_i are introduced scaling the coupling strength between the particle i and the Higgs boson. However, such a measurement has a high theoretical dependence and for every new/updated theory, the analysis is required to be reproduced. The theoretical dependence can be originated either by the model (in the above example, the SM) or by the theoretical uncertainties associated with the measurements.

When measurements of the properties of the Higgs boson are performed at the reconstruction level, the selected phase space includes detector effects making it difficult to compare the result to theories and other experiments. Unfolding the detector effects leads to the particle level information. This is different from the generator level which is the information at the hard scattering level. The particle level information includes parton shower and hadronization effects in addition to the hard scattering level.

To have a minimal theoretical dependence, the fiducial differential cross section measurement framework is proposed. In these measurements, the cross section of the Higgs boson is presented differentially in observable(s) at particle level (for example, the p_T of the dijet) using the reconstruction level information and the response matrix used to unfold the detector effects. The detector effects are calculated by the efficiency (i.e. the ratio of the reconstructed events by the total number of events). The theory dependence is estimated by the acceptance (i.e. the ratio of the generator level events to the total number of events). The usage of ML/MVA methods in these measurements is avoided since their training introduces model dependence but this reduces the sensitivity of the differential cross section measurements.

To increase the sensitivity and to study the deviations from a particular model (in this thesis, the SM) in mutually exclusive bins of the phase space, the Simplified Template Cross Section (STXS) framework is proposed. The STXS scheme

- measures cross sections instead of the signal strength, thus reducing theoretical dependence.

- the measurements are performed in 'simplified fiducial volumes' (STXS bins).

- allows the use of MVA/ML for signal extraction in the STXS bins, thus increasing the signal sensitivity.

- is defined for each production channel of the Higgs boson but is common for its decay modes.

The bins in the STXS 1.2 framework (used in this thesis) are split according to the reconstruction level vector boson p_T (see Figure 70), which was used in previous $\mathrm{H}(\rightarrow \mathrm{b\bar{b}})$ measurements as it provides a clean proxy for the Higgs p_T definition. In addition to this, they are also split by the number of additional jets (jets above 30 GeV with $|\eta| < 2.4$) as shown in Figure 70.

The high p_T bins allow for the addition of the boosted topology in the analysis which could probe BSM physics at high energies. The cross section measurement of the STXS bins can be interpreted in terms of the signal strength or constrains on effective field theories (EFT) parameters. The STXS stages are defined for each production channel of the Higgs boson but are common for its decay modes. Thus, the STXS measurements from different decay modes and different experiments can be combined, thus maximizing experimental sensitivity.

Due to limited statistics to constrain the signal strength, this analysis is not sensitive to all the bins of the STXS scheme shown in Figure 70. The sensitivity is checked using the Asimov dataset (discussed in Section 3.12.4). STXS bins that have limited sensitivity are fixed to their SM expectation, in the fit model to avoid unconstrained parameters.

- qqZH and ggZH signal contributions are merged: qqZH and ggZH have very similar topologies; thus it is difficult to separate them. Using kinematic differences like the number of ISR jets (expected to be higher in the ggZH process), p_T of the Higgs boson (higher resolution expected in the ggZH process), and other derived quantities can be used. Such separation can provide some sensitivity to top coupling in the ggZH process (through top quark loop as shown in Figure 17). However, for this analysis, the two topologies are merged due to limited sensitivity on ggZH bins. Further, the contribution of ggZH to ZH NNLO cross section is about 6% [60]. Thus constraints on the ZH STXS bins are dominated by quark-induced ZH production. The contribution of gluon-induced ZH production increases with increasing p_T of the Higgs boson.

- For WH, 0 and ≥ 1 additional jet categories are merged.

- The 0-75 GeV p_T(V) bins are not used.

- The 75-150 GeV p_T(V) bin is used only for ZH.

In the following Sections, the STXS bin corresponding to $75 < p_T(\mathrm{V}) < 150$ GeV is labeled as 'low p_T' bin, $150 < p_T(\mathrm{V}) < 250$ GeV as 'med p_T' bin, $250 < p_T(\mathrm{V}) < 400$

GeV as 'high1 (high1 BOOST) p_T' bin for resolved (boosted topology), p_T(V) > 400 GeV as 'high2 (high2 BOOST) p_T' bin for resolved (boosted topology) and p_T(V) > 250 GeV as 'high (high BOOST) p_T' bin for resolved (boosted topology).

The final SR and CR used in the analysis, split by the reconstructed p_T of the vector boson as per the STXS scheme for resolved topology and boosted topology, are shown in Figure 71 and 72 respectively. As shown, for the high1 and high2 regions, we did not split the CR to get better constraining power for the major background processes described in the fit model in Section 3.13.

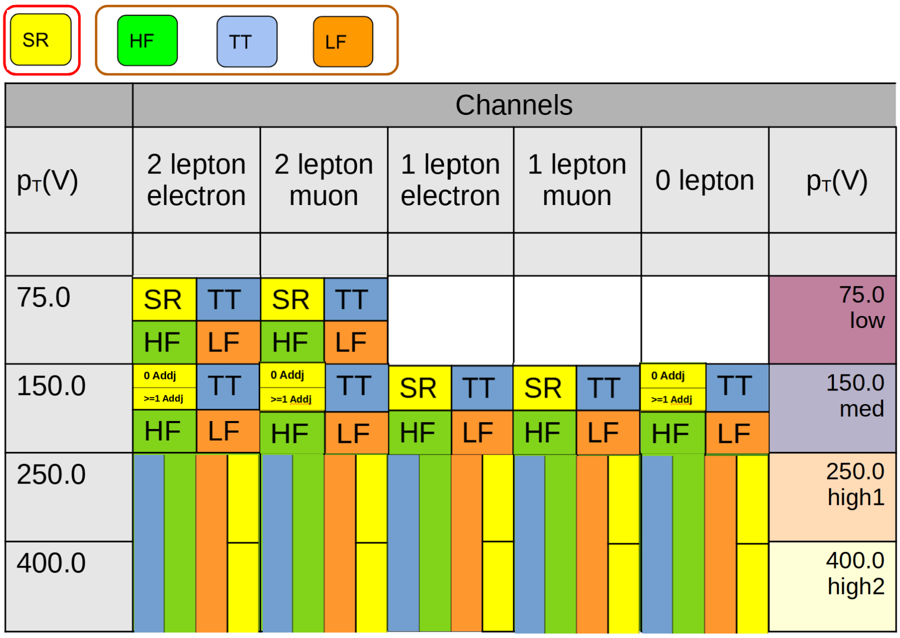
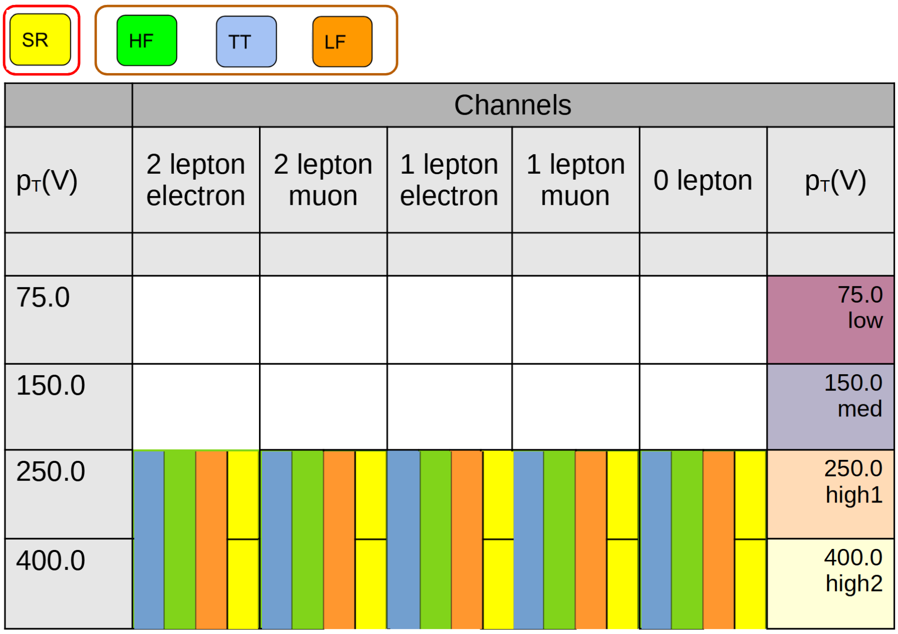


Figure 71: Distribution used in the fit model of SR and CR in the STXS bins in each channel for resolved topology. p_T(V) is in units of GeV.



	Channels					
pT(V)	2 lepton electron	2 lepton muon	1 lepton electron	1 lepton muon	0 lepton	pT(V)
75.0						75.0 low
150.0						150.0 med
250.0	🟦🟩🟧🟨	🟦🟩🟧🟨	🟦🟩🟧🟨	🟦🟩🟧🟨	🟦🟩🟧🟨	250.0 high1
400.0	🟨	🟨	🟨	🟨	🟨	400.0 high2

Figure 72: Distribution used in the fit model of SR and CR in the STXS bins in each channel for boosted topology. p_T(V) is in units of GeV.

3.10 Corrections applied on simulated events

Corrections on simulated events can be classified as: MC modelling corrections to account for data/MC mis-modelling (effectively also used to remove the effect of bugs in MC production); object SF to account for the difference between data and MC efficiencies; and corrections due to detector response. The uncertainties on these corrections are included in the fit model and are described in Section 3.13.2.

3.10.1 MC modelling corrections

Pile up corrections The amount of pileup in the CMS detector depends on the instantaneous luminosity during the run. The pile-up profile (mean number of interactions per bunch crossing) for the 2018 run is shown in Figure 73, assuming a minimum bias cross section of 69.2 mb [94].

A per-event weight was derived for simulation to match the pile-up profile in data. Figures 74 and 75 show the data/MC agreement before (left) and after (right) the corrections in TT CR for the number of vertices and ρ respectively. As shown, the simulated distribution, after the application of the corrections, does not model the data satisfactorily. To deal with this mis-modelling, we use the uncertainty on the pile-up weight as a nuisance parameter in the fit model (discussed further in Section 3.13).

Corrections on LO V+jets in the 2016 analysis As mentioned in Section 3.4, for 2016, LO V+jets samples are used. To correct the sample observables to the

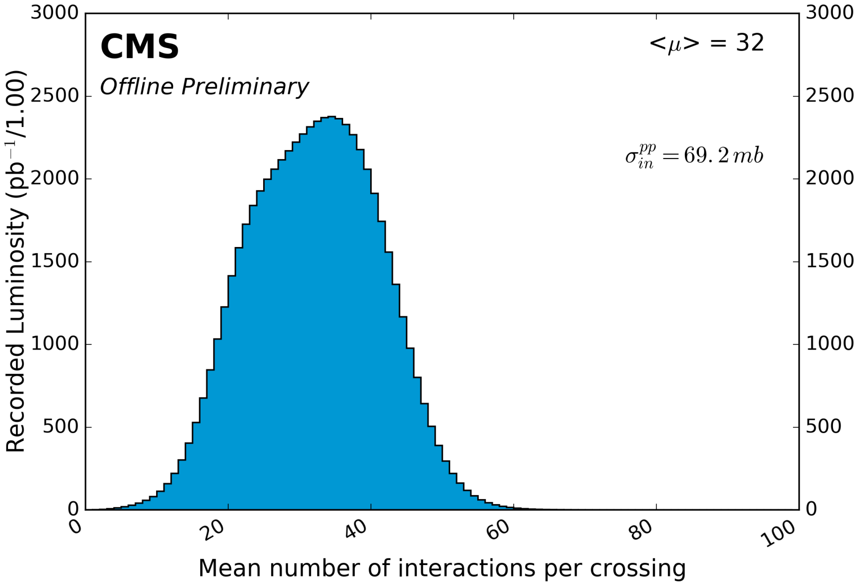


Figure 73: Mean number of interactions per bunch crossing for proton-proton collisions at the LHC in 2018. The minimum bias cross section is assumed to be 69.2 mb and is found to agree with the data [94].

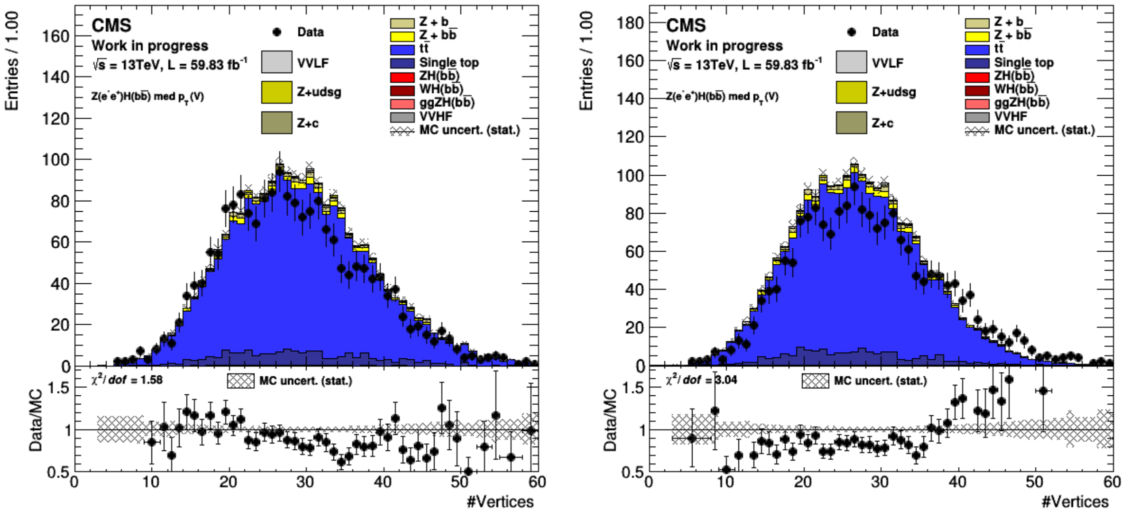


Figure 74: Number of primary vertices in 2-lepton TT CR before (left) and after (right) pile-up reweighting.

corresponding NLO ones, LO to NLO correction factors are derived in bins of $\Delta\eta$ of the two b jets. Ideally, a multi-dimensional correction should have been used, but due to limited NLO statistics, only $\Delta\eta$(jj) was used for reweighting.

Figure 76 (left) shows the comparison between data and MC for LO and NLO V+jets samples in bins of dijet mass. As shown, the agreement for the NLO samples

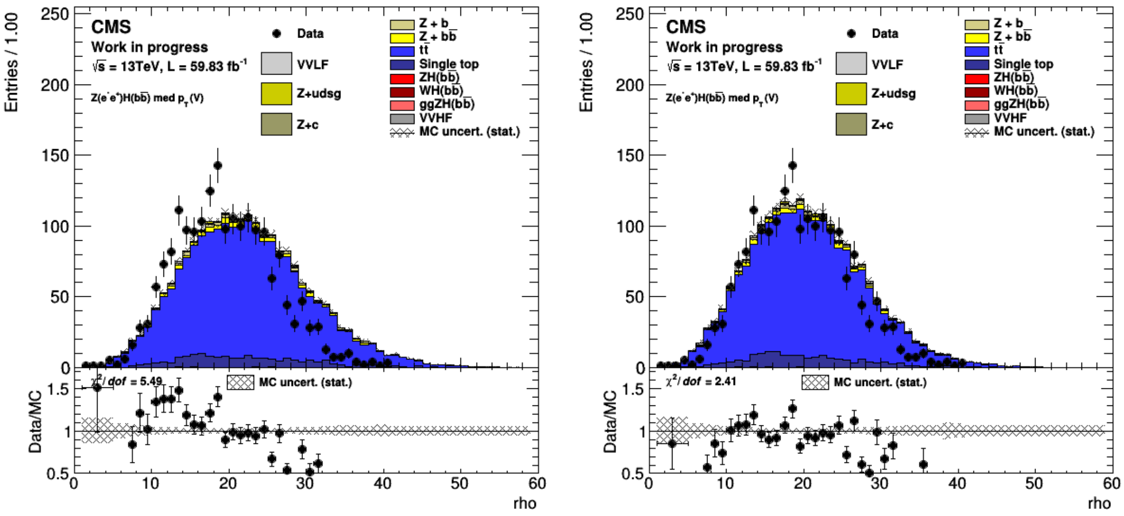


Figure 75: ρ in 2-lepton TT CR before (left) and after (right) pile-up reweighting.

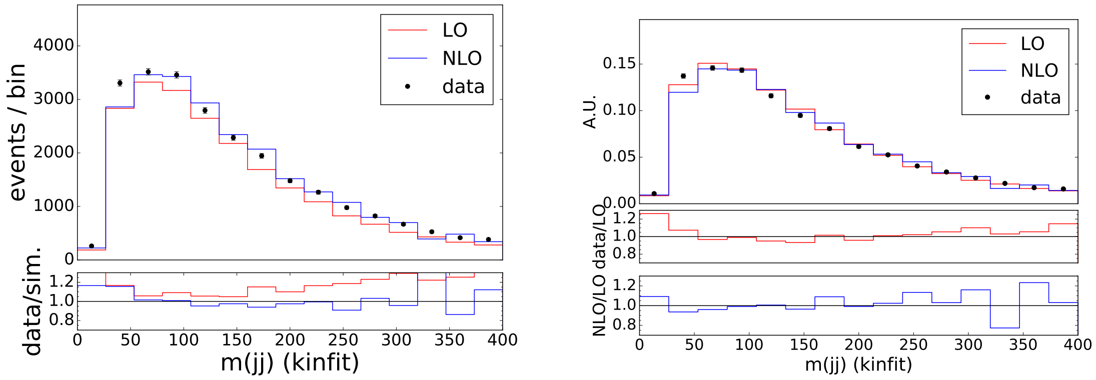


Figure 76: Data/MC agreement for LO V+jets samples and NLO V+jets with data before (left) and after $\Delta\eta$(bb) corrections (right).

is better compared to LO. Figure 76 (right) shows the dijet mass for LO after the LO to NLO corrections are applied. The usage of uncertainties on $\Delta\eta$(jj) weights are further described in Section 3.13.2.

Corrections on NLO p_T split V+jets sample in the 2018 analysis As mentioned in Section 3.4, the 2018 DY+jets and W+jets samples are binned in generator level p_T (LHE p_T) of the vector boson. However, on validating these samples, non-physical spikes were observed at the sample boundaries. This is a known issue [95] due to a bug in Madgraph settings for LHE p_T binned V+jets samples. The following procedure was used to correct this mis-modelling:

- remove the events outside the sample LHE p_T range;

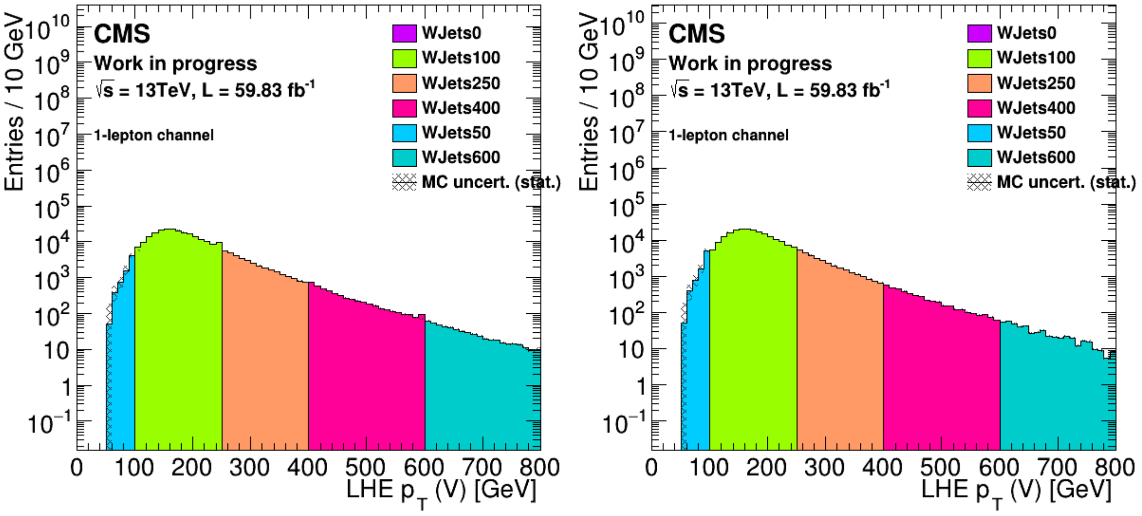


Figure 77: Generator level $p_T(V)$ before (left) and after (right) the application of correction to LHE p_T binned W+jets samples in 2018. WJets0 corresponds to the sample with $0 <$ LHE p_T(V) < 50 GeV. Similarly, WJets50 $\implies$ $<$ LHE p_T(V) $<$ 100 GeV sample, WJets100 $\implies$ $100 <$ LHE p_T(V) < 250 GeV sample, WJets250 $\implies$ $250 <$ LHE p_T(V) < 400 GeV sample, WJets400 $\implies$ $400 <$ LHE p_T(V) $<$ 600 GeV sample, WJets600 $\implies$ p_T^{LHE}(V) > 600 GeV sample.

- reweight events based on ratio of LHE p_T histogram of remaining set of V+jets samples to LHE p_T-binned V+jets sample;
- the uncertainty on the ratio is used as the uncertainty on the correction.

Figure 77 shows the LHE p_T(V) distribution before (left) and after (right) applying the correction to the LHE p_T binned W+jets samples. As seen, the correction removes the non-physical spikes in the LHE p_T distribution.

Corrections on NLO V+jets in 2017/2018 analysis For the NLO V+jets samples generated with the MadGraph generator, a mis-modelling in the $\Delta R < 1.0$ of the two b-jets Higgs candidates was observed. In addition to this, the mis-modelling was found to be independent of the jet flavor. Due to a larger amount of V+jets statistics in LF CR compared to other CRs, the LF CR is used for the derivation of the additional V+jets corrections. These are extrapolated to the V+jets in HF CR and SR. This mis-modelling was reported by other analyses in the CMS Collaboration as well [96].

Before deriving the correction, it was found that the LF CR suffers from mis-modelling in the DeepCSV score of the jets below the Loose WP. This is due to the lack of calibration SF for jets with DeepCSV scores less than Loose WP. For this reason, the modelling of the DeepCSV score of the two leading jets in the event is first corrected if their score is less than the Loose WP and then for the mis-modelling in the $\Delta R < 1.0$.

For the 2-lepton channel, both the leading and subleading jets have a DeepCSV score less than the Loose WP. For the 1-lepton channel, only the sub-leading jet has a DeepCSV score less than the Loose WP while for the 0-lepton channel, both leading and sub-leading jets have a DeepCSV score greater than the Loose WP. These V+jets DeepCSV SF are derived by first subtracting the MC background template for all processes (other than V+jets) from the data in the LF region. A data-to-MC ratio for the V+jets MC templates to the non-V+jets background subtracted data template is then extracted. This is done in two dimensions using the DeepCSV score of leading and sub-leading jets. The result is then used as a correction factor to account for the mis-modelling of the DeepCSV score in LF CR.

Figures 78 and 79 show the DeepCSV distribution of the leading and sub-leading jets before and after the application of the V+jets correction described above. No mis-modelling is observed in Figure 78 (middle left), since the leading jet in the 1-lepton channel has a DeepCSV score greater than the Loose WP.

After the application of DeepCSV correction factors, the additional V+jets correction is derived for ΔR $<$ 1.0 for V+jets in LF CR, using the bin-by-bin ratio of data and simulation. The same factors are used to correct the V+jets distribution in ΔR $<$ 1.0 for the SR and HF CR as well, since the mis-modelling is found to be independent of the jet flavor [96]. Figures 80-85 show ΔR(bb) distributions before (left) and after (right) corrections in ΔR $<$ 1.0 region in SR, V+HF and V+LF CR. The statistical uncertainty of this bin-by-bin correction is used as a nuisance parameter in the fit model discussed in Section 3.13.3.

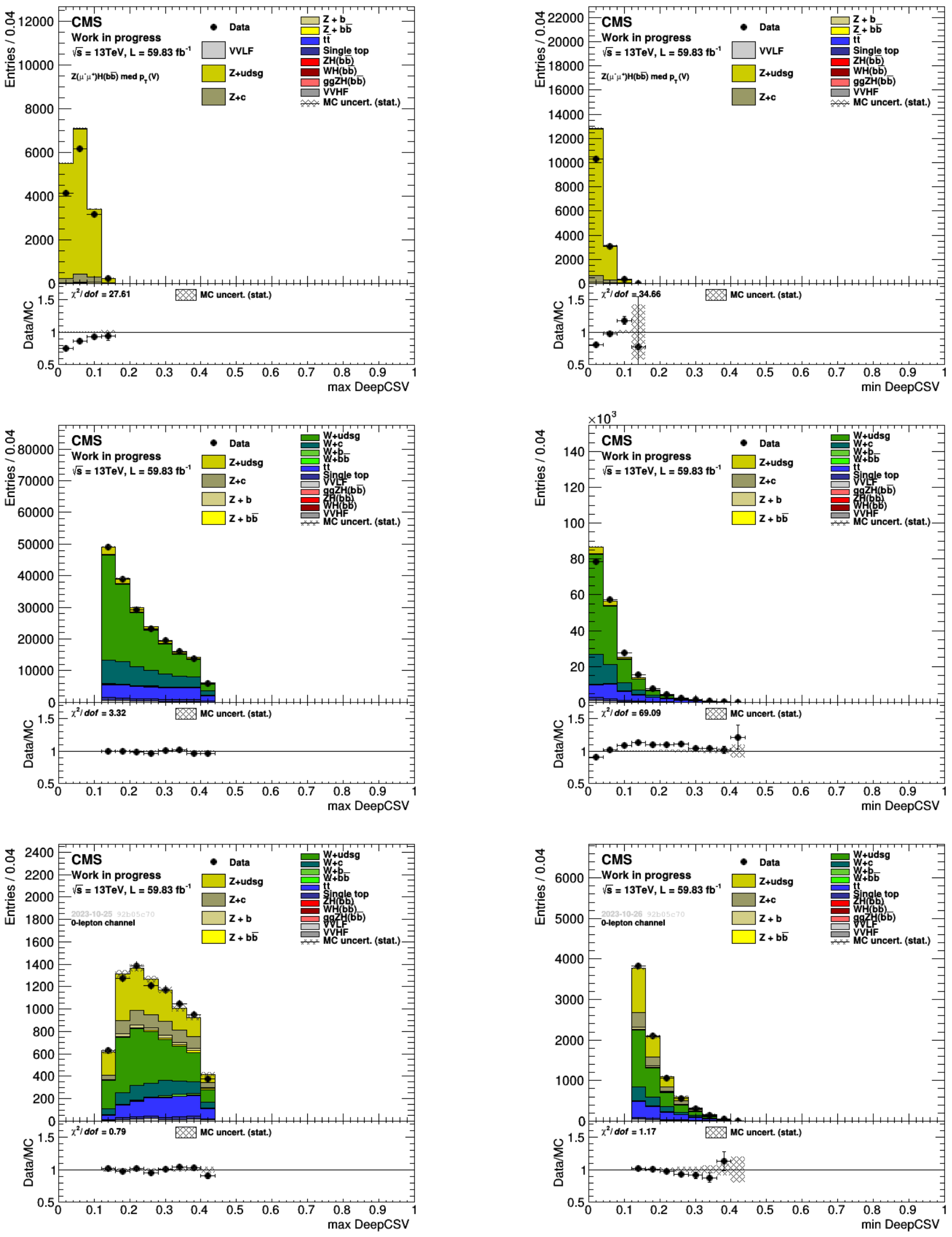


Figure 78: DeepCSV score of the leading (left) and sub-leading (right) jet before application of 2D btag corrections in the LF CR of the 2-lepton channel (top), 1-lepton channel (middle), and 0-lepton channel (bottom)

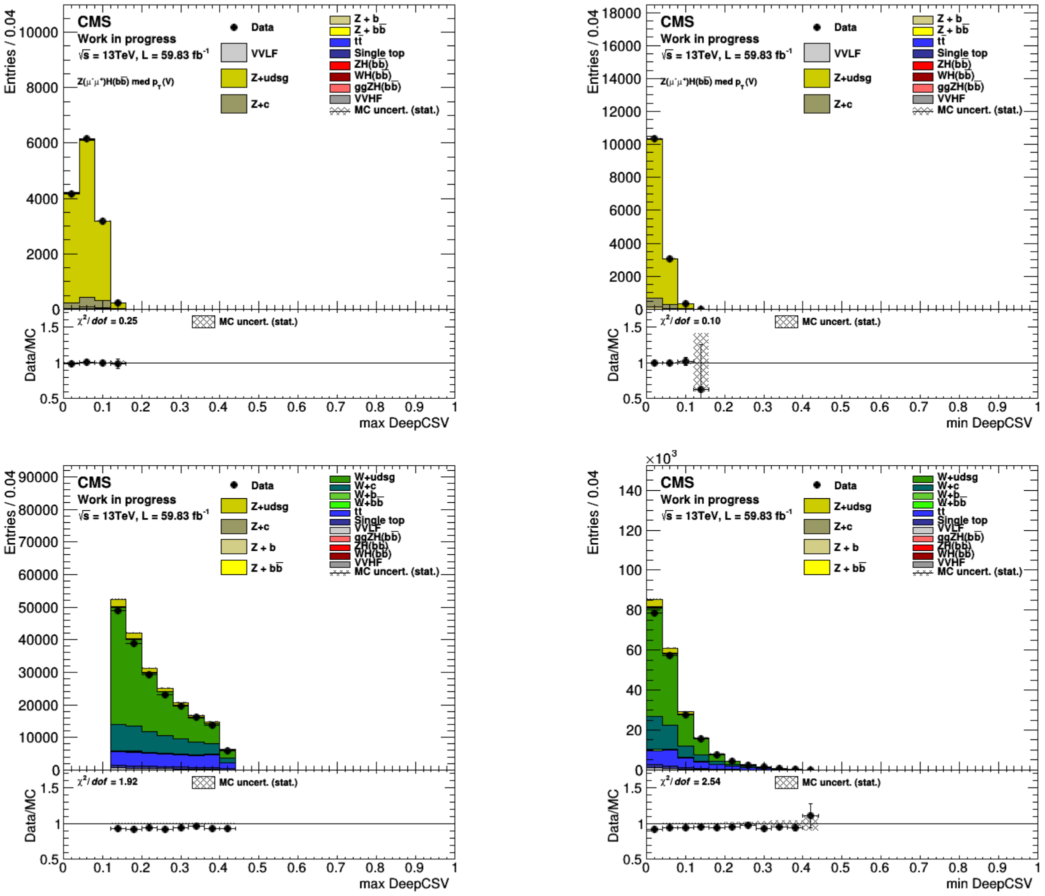


Figure 79: DeepCSV score of the leading (left) and sub-leading (right) jet after application of 2D btag corrections in the LF CR of the 2-lepton channel (top) and 1-lepton channel (bottom).

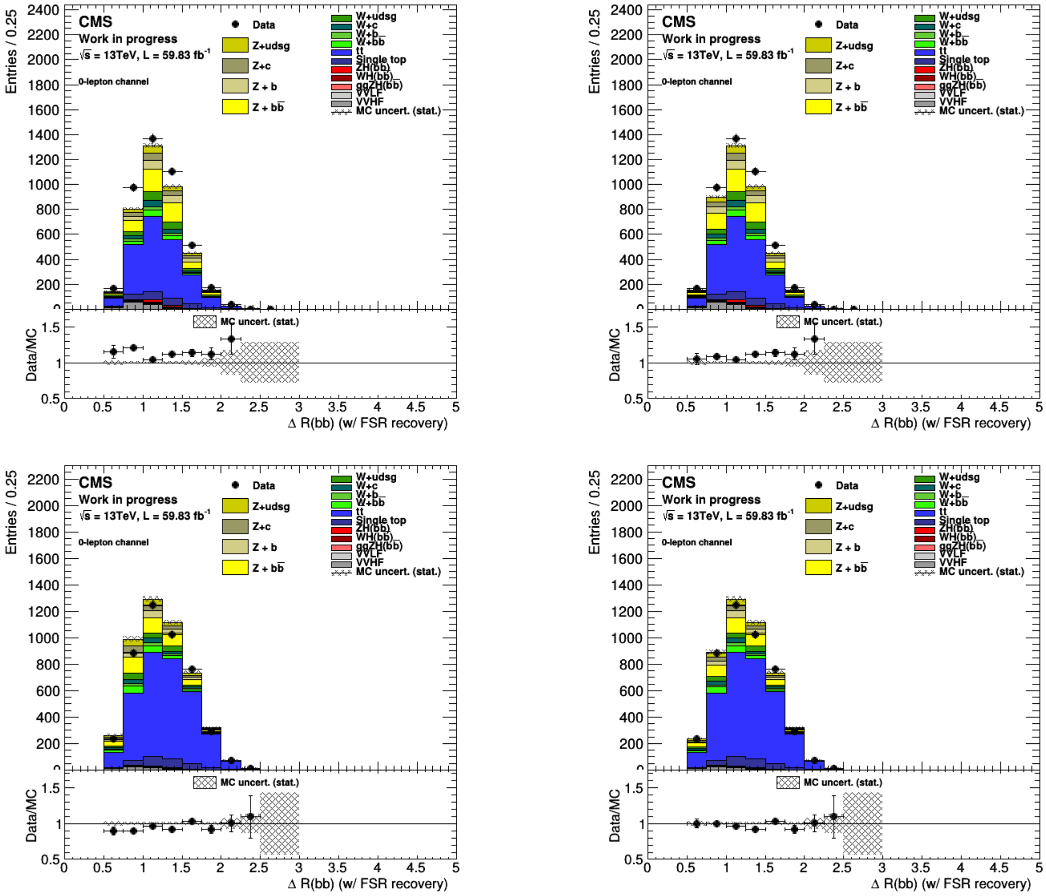


Figure 80: ΔR(bb) distribution in 0 (top) and greater than or equal to 1 additional jet (bottom) in med p_T SR bin of the 0-lepton channel before (left) and after (right) the application of corrections for ΔR(bb)< 1.

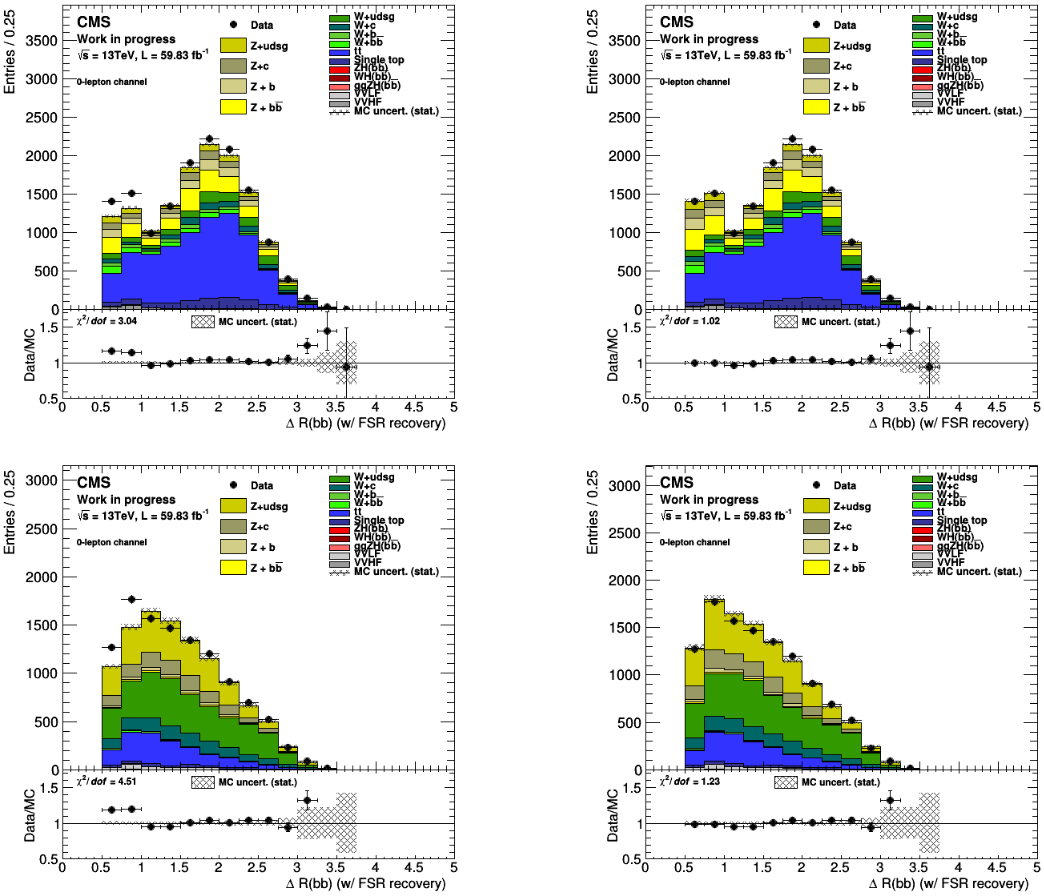


Figure 81: ΔR(bb) distribution in HF CR (top) and LF CR (bottom) in med p_T SR bin of the 0-lepton channel before (left) and after (right) the application of corrections for ΔR(bb)< 1.

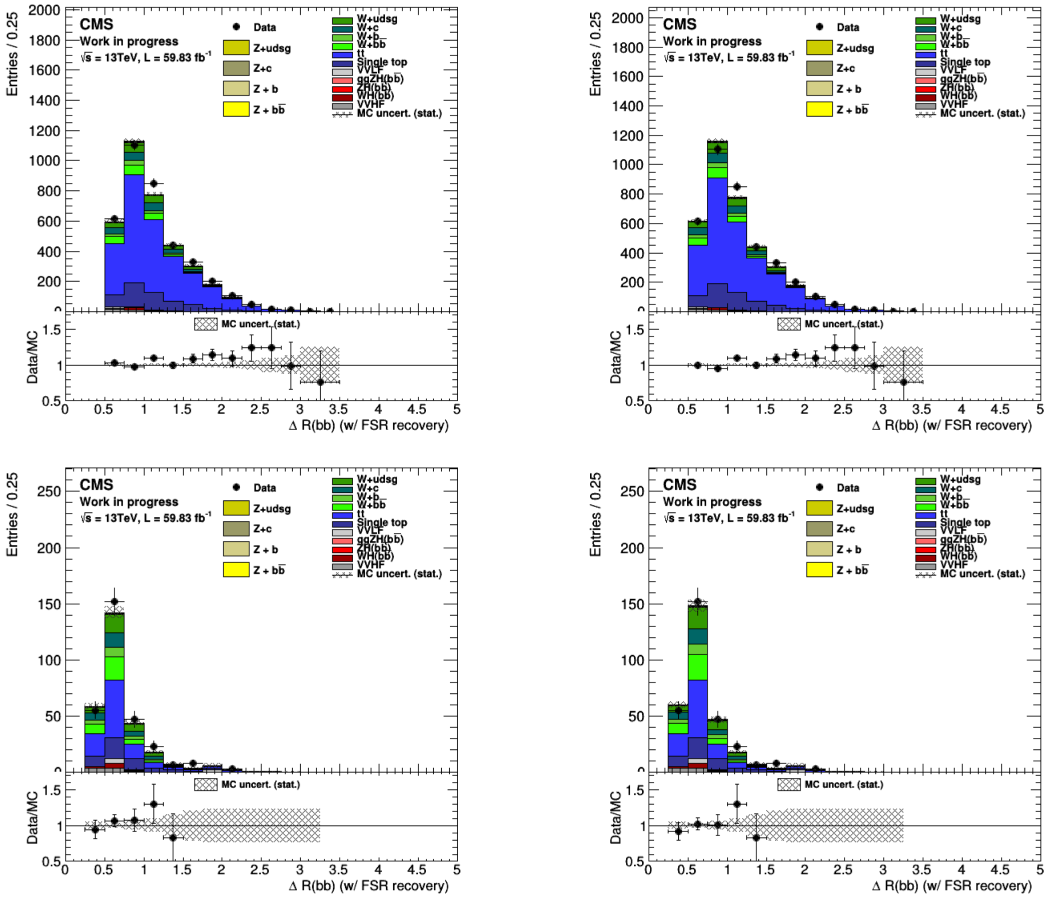


Figure 82: ΔR(bb) distribution in high1 (top) and high2 (bottom) p_T SR bin in the 1-lepton channel before (left) and after (right) the application of corrections for ΔR(bb)< 1.

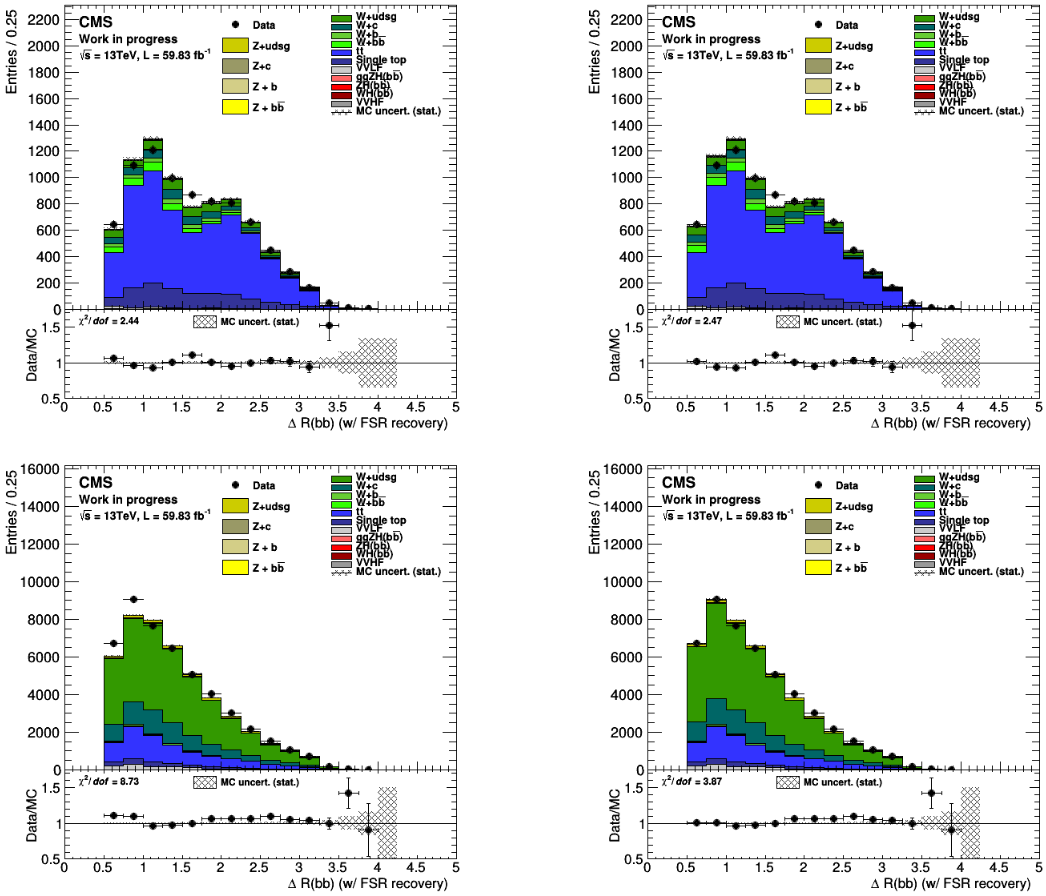


Figure 83: ΔR(bb) distribution in V+HF CR (top) and V+LF CR (bottom) high p_T SR bin in the 1-lepton channel before (left) and after (right) the application of corrections for ΔR(bb)< 1.

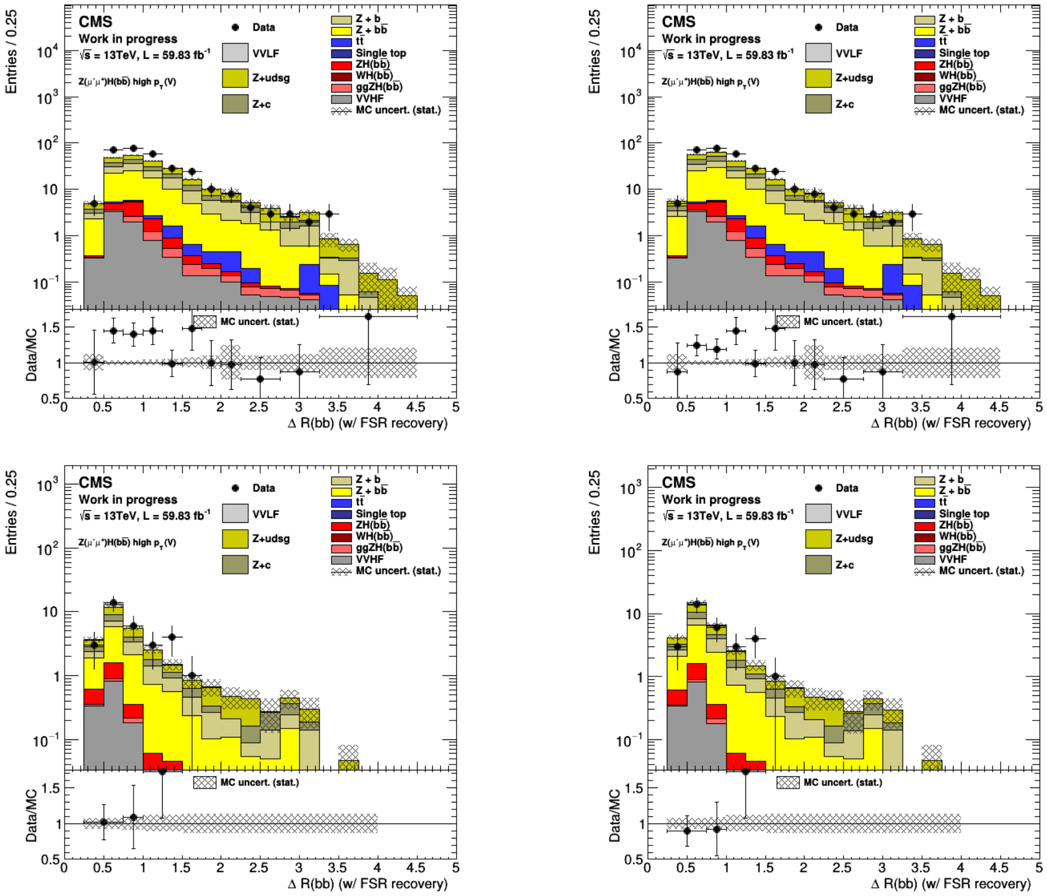


Figure 84: ΔR(bb) distribution in high1 (top) and high2 (bottom) p_T SR bin in the 2-lepton channel before (left) and after (right) the application of corrections for ΔR(bb)< 1.

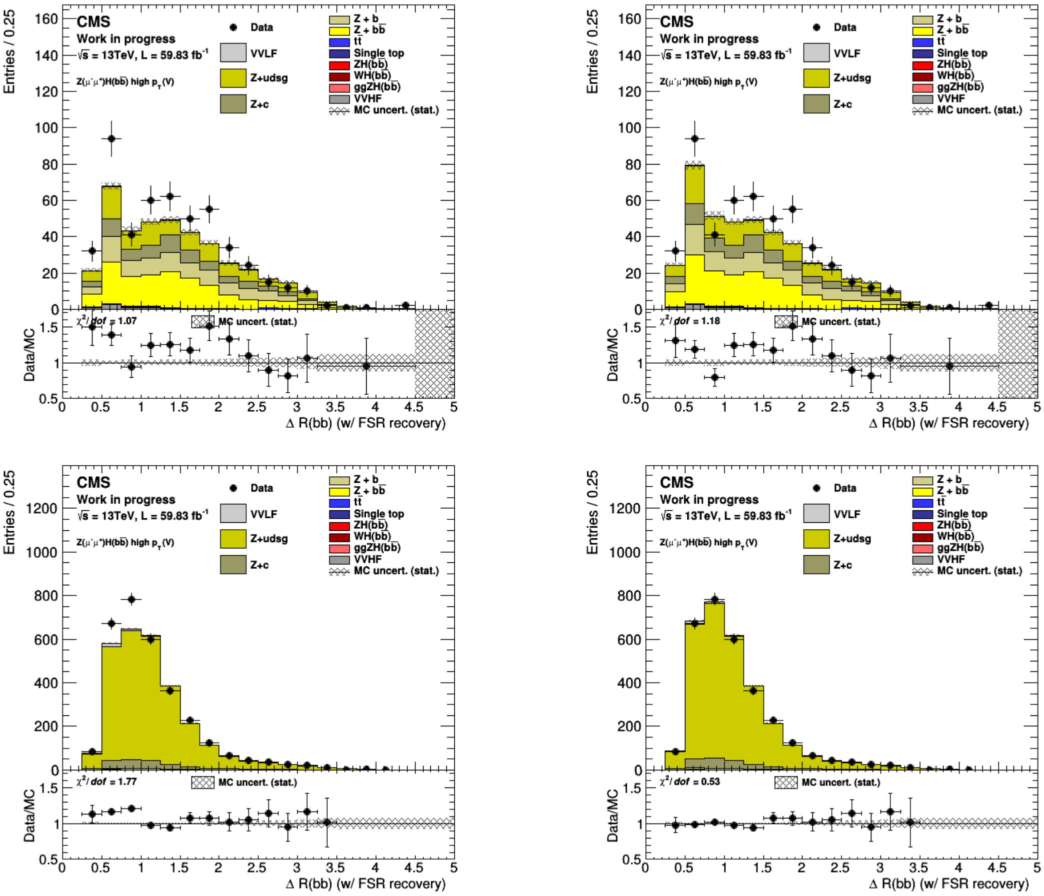


Figure 85: ΔR(bb) distribution in V+HF CR (top) and V+LF CR (bottom) high p_T SR bin in the 2-lepton channel before (left) and after (right) the application of corrections for ΔR(bb)< 1.

Correction on the number of additional jets on V+jets samples As shown in the left column of Figures 86, 87 and 88 a mis-modelling was observed in the number of additional jets variables associated with the NLO V+jets sample in the 2017 and 2018 analyses. This mis-modelling in the number of additional jets variables was limited to parton shower-related variables like SA5 (discussed in Section 3.6.9) and the number of additional jets, correlating to the matrix element and parton shower matching scheme. The effect was observed only for the 2017 and 2018 simulations, since the 2016 analysis relies on LO V+jets which uses a different matching scheme (MLM) instead of the NLO (FxFx). The SA5 mis-modelling is further discussed in Section 3.11.1 while the number of additional jets mis-modelling is discussed below.

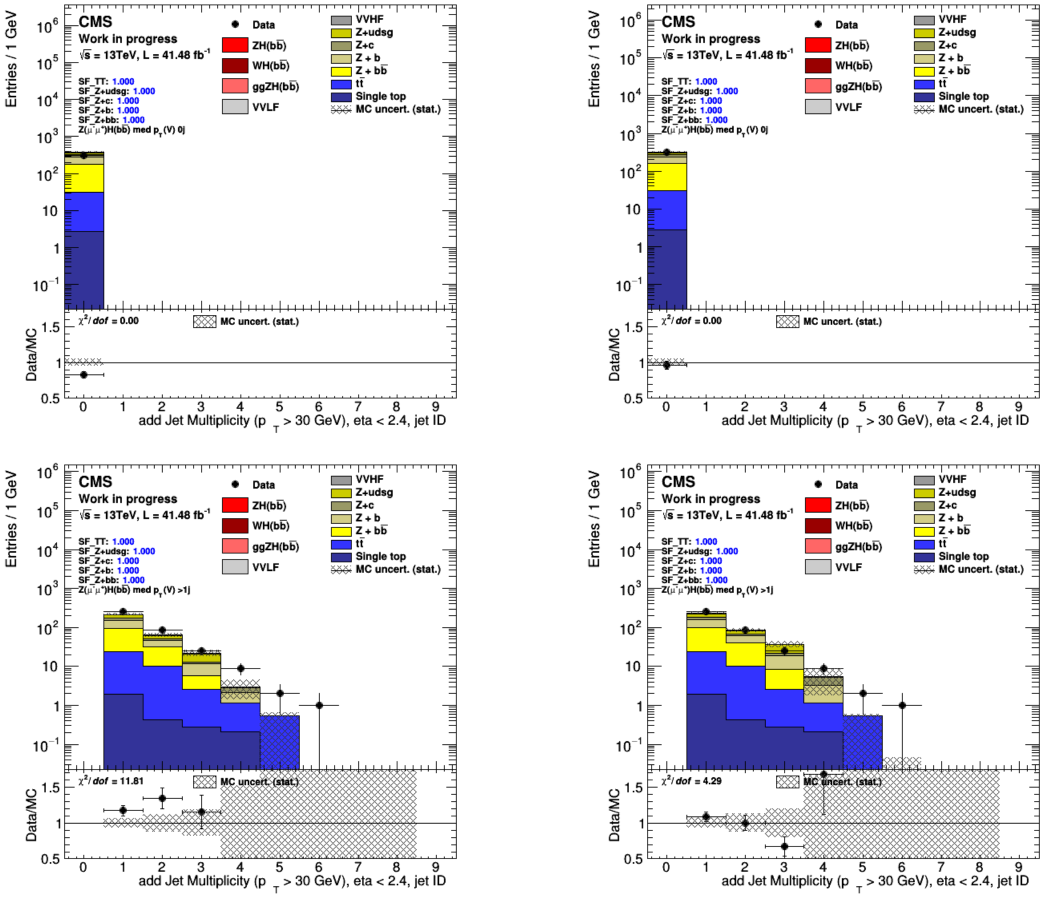


Figure 86: Number of additional jets distribution before (left) and (right) after corrections in SR 0 (top) and greater than 0 (bottom) additional jet STXS bin in the 2-lepton channel.

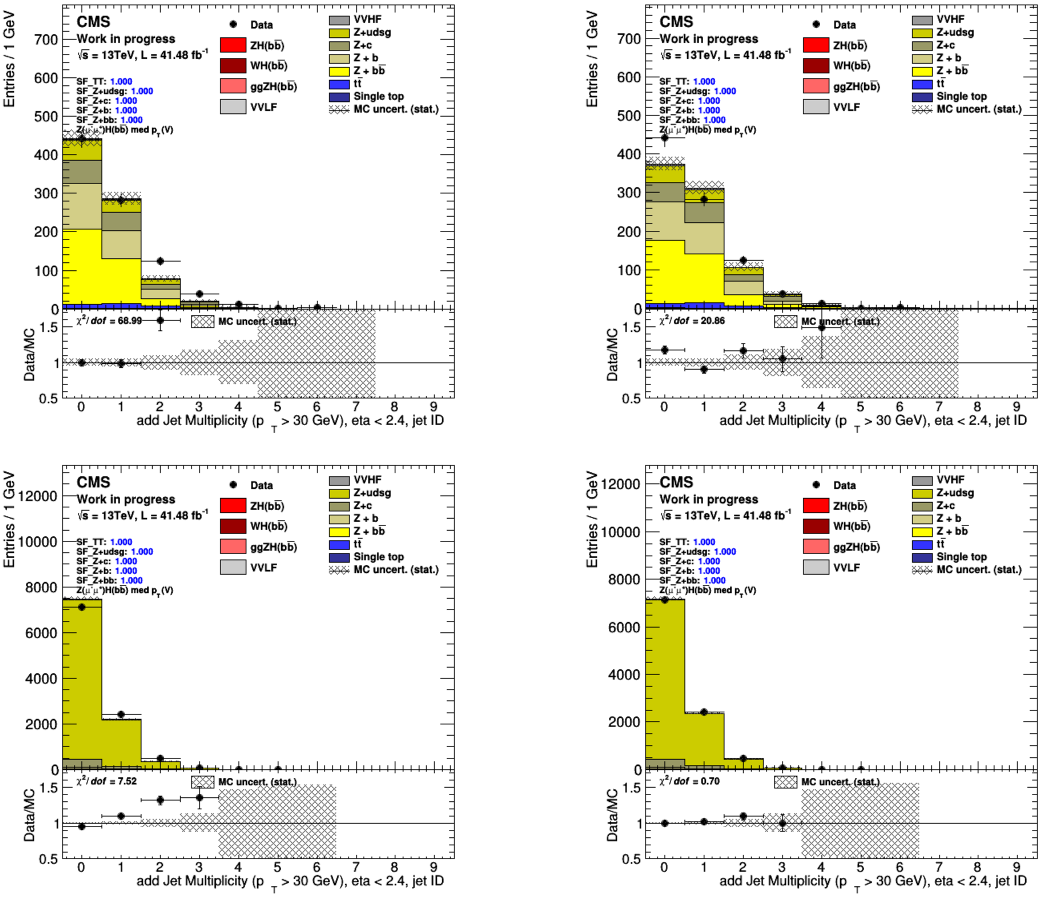


Figure 87: Number of additional jets distribution before (left) and (right) after corrections in HF (top) and LF CR (bottom) med p_T STXS bin in the 2-lepton channel.

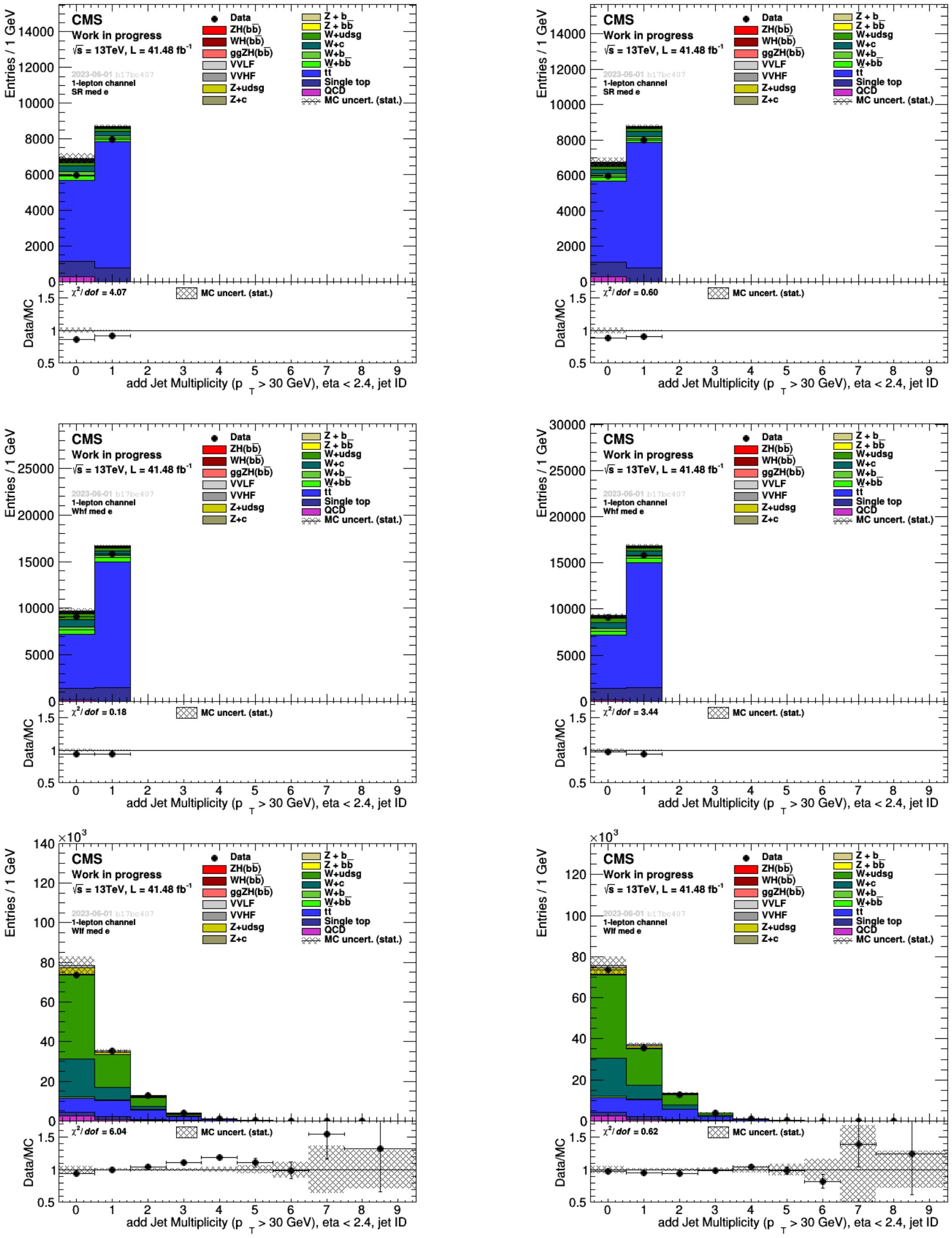


Figure 88: Number of additional jets distribution before (left) and (right) after corrections in SR (top), HF CR, and LF CR (bottom) in med p_T STXS bin in the 1-lepton channel.

The number of additional jets is used for the STXS definition of the VH production mode (as discussed in Section 3.9). To correct the mis-modelling in the number of additional jets, a dedicated reweighting was derived as a function of p_T of the vector boson. Figure 89 shows the number of additional jets distribution in MC in bins of p_T of the vector boson. Since a Poisson modelling of these distributions was found satisfactory, as shown in Figure 89, the number of additional jet distributions for MC and data was fitted with a Poisson distribution for HF and LF CR.

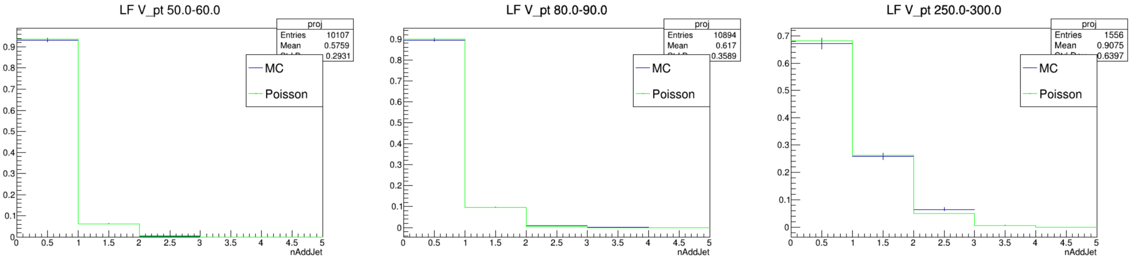


Figure 89: The distribution of the number of additional jet observables in bins of p_T of the vector boson. A Poisson fit is overlayed to the distributions [97][98].

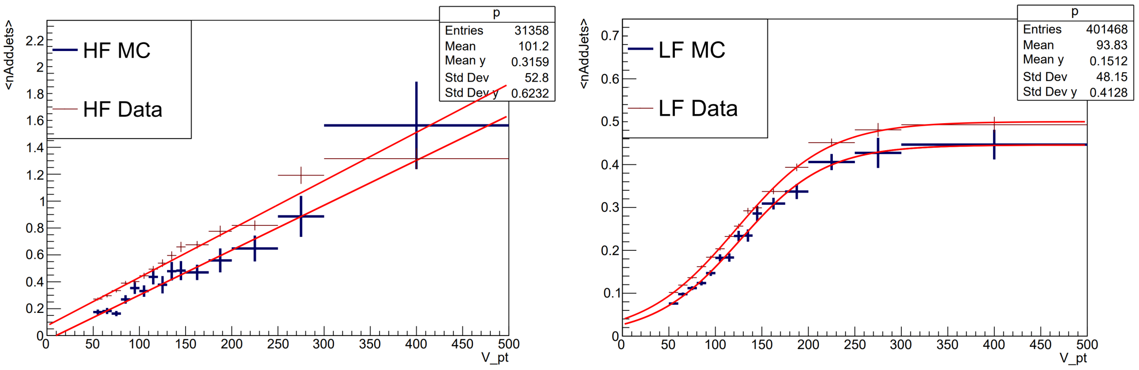


Figure 90: The rate parameter (λ) of the Poisson distribution fitted to the number of additional jets as a function of p_T of the vector boson in HF CR (left) and LF CR (right) [97][98]. Since the Poisson rate parameter is same as the mean of the Poisson distribution, the y-axis is labeled as a mean of the number of additional jets.

The rate parameter (λ) of the Poisson distribution was extracted from these fits and plotted as a function of the p_T of the vector boson as shown in Figure 90 for HF CR (left) and LF CR (right). The mean parameter of the fitted Poisson distribution as a function of p_T of the vector boson is modeled by a linear function for HF CR and a logistic function for the LF CR. The correction factors for an event in MC are derived as the ratio of the value of fitted Poisson distribution of data and MC. That is, for an event with a particular value of p_T(V), the rate parameter for data and MC is obtained from the fitted linear function (for HF CR) or logistic function (for LF CR). Since the phase space of SR and HF CR is similar, the same corrections are used for both HF CR and SR. The corrected distributions of the number of additional jets in SR and CR for 2- and 1-lepton channels are shown in the right columns of Figures 86, 87 and 88.

EWK corrections Electroweak corrections up to NLO can be factorized at parton level and thus are calculated using the parton-level generator HAWK [99] and applied differentially in p_T of the vector boson as a multiplicative weight to both (quark induced) signal and V+jets backgrounds. The EWK weights for W+jets and Z+jets are shown in Figure 91.

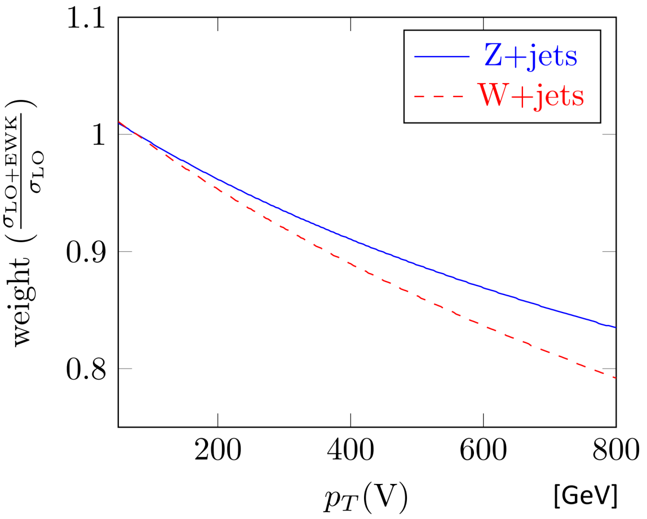


Figure 91: EWK weights as a function of p_T of vector boson for Z+jets and W+jets processes.

The signal MC sample for the quark-induced process is generated at NLO using POWHEG + MiNLO and then rescaled to the cross-section to NNLO of QCD and NLO at EWK (using EWK corrections). The total NNLO cross-section σ^{VH} for the signal is given by

$$\begin{aligned}\sigma^{\text{WH}} &= \sigma^{\text{WH,DY}}_{NNLOQCD}\,(1+\delta_{\text{EW}}) + \sigma_{\text{t loop}} + \sigma_{\gamma}, \\ \sigma^{\text{ZH}} &= \sigma^{\text{ZH,DY}}_{NNLOQCD}\,(1+\delta_{\text{EW}}) + \sigma_{\text{t loop}} + \sigma_{\gamma} + \sigma^{\text{ggZH}}\end{aligned} \tag{3.10}$$

where $\sigma^{\text{Z/WH}}_{NNLOQCD}$ is the contribution from DY-like processes. σ^{ggZH} is the cross section of gluon-induced contribution and is computed using VHNNLO generator [100]. The EWK component $(1+\delta_{\text{EW}})$ in the equation is calculated by the parton-level generator HAWK [99] up to NLO and is shown in Figure 93.

3.10.2 Efficiency corrections for physics objects

Lepton SF As discussed in Section 3.6.4, lepton ID and relative isolation selections are used to reduce the fake muons and electrons in this analysis. The tag and probe method on $Z \to \mu\mu$ (for muons) or $Z \to ee$ (for electrons) is used, to calculate the

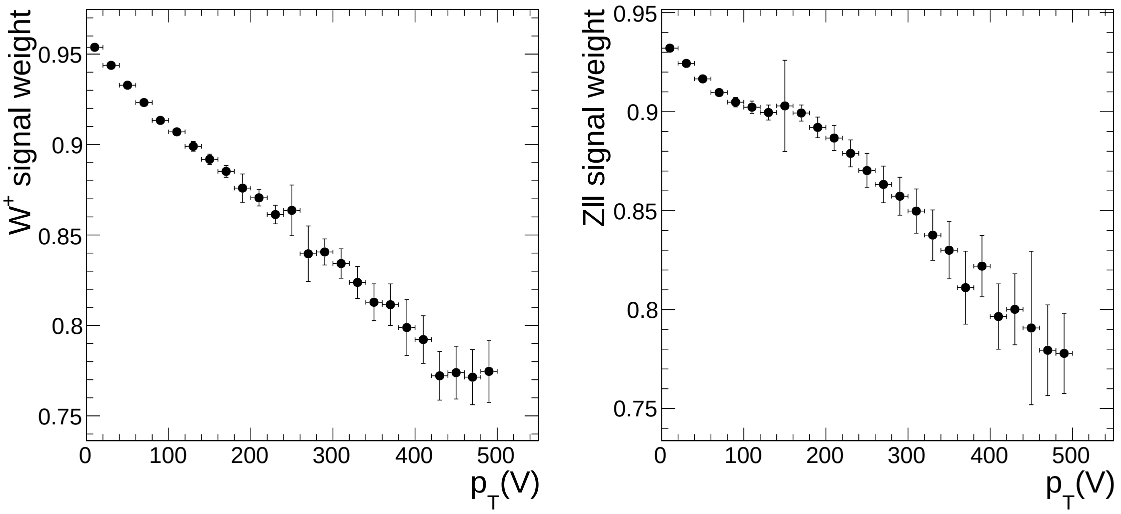


Figure 92: Multiplicative factor $(1 + \delta_{\text{EW}})$ for signal WH($\rightarrow$ b$\bar{\text{b}}$) (left) and ZH($\rightarrow$ b$\bar{\text{b}}$) (right).

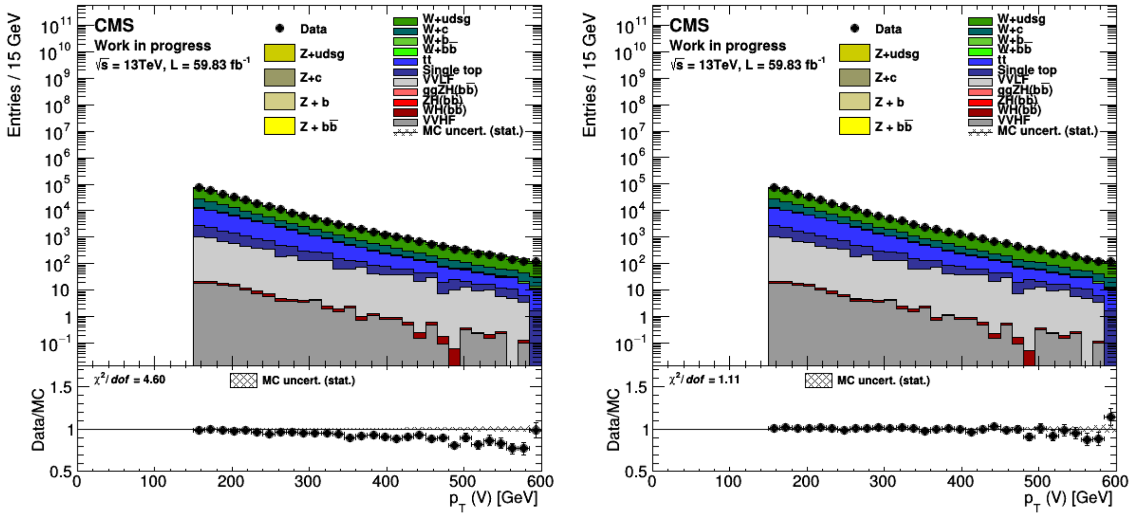


Figure 93: p_T of the vector boson before (left) and after (right) application of EWK corrections in the 1-lepton channel.

efficiency of the lepton ID, relative isolation WP along with trigger selections in data, and MC. To account for differences in the efficiency of lepton ID, relative isolation, and trigger selection between data and MC, SFs are used. Thus, the corresponding combined lepton SF is

$$\text{SF}^{l}_{ID} \times \text{SF}^{l}_{ISO|ID} \times \text{SF}^{l}_{\text{Trigger}|ISO+ID} \tag{3.11}$$

MET SF As mentioned in Section 3.5, events in the 0-lepton channel are selected using combined MET and MHT triggers as in Table 16.

Year	HLT path for analysis	Tag HLT path
2016	HLT_PFMET110_PFMHT110_IDTight HLT_PFMET120_PFMHT120_IDTight HLT_PFMET170_NoiseCleaned HLT_PFMET170_BeamHaloCleaned HLT_PFMET170_HBHECleaned	HLT_Ele27_WPTight_Gsf
2017	HLT_PFMET120_PFMHT120_IDTight_OR HLT_PFMET120_PFMHT120_IDTight_PFHT60	HLT_Ele35_WPTight_Gsf
2018	HLT_PFMET120_PFMHT120_IDTight	HLT_Ele32_WPTight_Gsf

Table 16: Triggers used to collect the data samples for measuring the MET trigger efficiencies, for each channel and year.

Unbiased single electron events with large MET in W+jets and $\mathrm{t\bar{t}}$ simulation events passing the tag HLT path (mentioned in Table 16) and MET filters (to reject fake MET due to detector noise, cosmic rays, etc) [81] are used to calculate the trigger efficiency. In addition, to have similar phase space for measurement as in this analysis, events are also required to have at least 2 jets with $p_T > 20$ GeV and $|\eta| < 2.5$ with $|\Delta\phi(\mathrm{Electron, MET})| < 2.5$ so that the lepton and reconstructed MET are not back-to-back. The efficiency of MET triggers is defined as the fraction of events passing the selections.

The differences in the efficiency of the MET trigger in data and MC are accounted for by using the MET trigger SF. The central value of the scale factor is derived using the efficiency measurement in the $\mathrm{t\bar{t}}$ sample. The difference in scale factor observed using the $\mathrm{t\bar{t}}$ and W+Jets samples is used as a systematic uncertainty in the fit model (discussed in Section 3.13.2).

3.10.3 Corrections due to detector response

MET xy corrections The true MET distribution is independent of the azimuthal angle due to the rotational symmetry of the collisions around the beam axis, while ϕ of the reconstructed MET has a sinusoidal distribution with period 2π. This is due to the modulation caused by anisotropic detector response, inactive calorimeter cells or tracking regions, the detector misalignment, and the displacement of the beam spot. To account for this, corrections are applied to MET in the x and y coordinate frame as prescribed in reference [101]. Figure 94 shows ϕ(MET) before (left) and after (right) application of MET xy corrections in the 0-lepton channel in the 2016 dataset. Compared to 2016, smaller sinusoidal modulations in ϕ(MET) were observed in the 2017 and 2018 data.

3.11 MVA classifier in signal and control regions

To extract the cross section in the STXS bins, we use multivariate analysis tools like DNN for the resolved topology and BDT for the boosted topology. Further details about these MVA variables are given in the following sections.

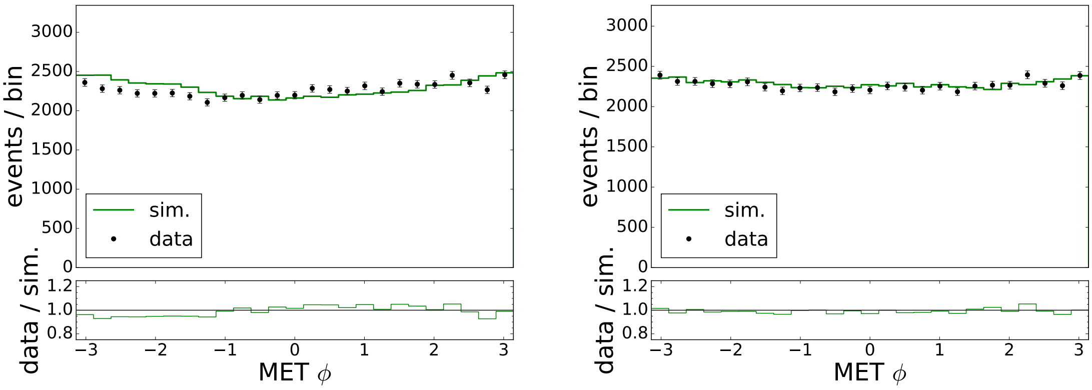


Figure 94: ϕ(MET) before (left) and after (right) application of MET xy corrections in the 0-lepton channel in the 2016 analysis.

3.11.1 MVA in the resolved 0- and 1-lepton channels

We use a binary signal/background DNN classifier in SR and a multiclass DNN classifier in HF CR. The reason for using a multiclass DNN classifier in HF CR is to classify different background processes because these CRs have a high contribution from $\mathrm{t\bar{t}}$ and V+udsg processes in addition to V+ b/bb process.

The DNN architecture consists of a feed-forward network with 6 hidden layers having 512, 256, 128, 64, 64, and 64 nodes respectively. For the signal-background classifier, only one output node is present, which through the sigmoid activation function, gives the probability of the event being signal-like. For the multiclass classifier, 6 output nodes are present which through the final softmax layer gives a probability of an event to be V+light jet (V+udsg), V+c jet, V+b jet, V+bb jet, single top (ST), and $\mathrm{t\bar{t}}$ (TT) process. The ReLU activation [102] is used for each of the inner layers. The classifier weights are trained using the Adam optimizer [103] to minimize the weighted cross-entropy loss function with a batch size of 32. Multiclass cross-entropy loss is defined as:

$$\text{Loss} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} c_i(c)\log(y_i(c)) \tag{3.12}$$

where N is the number of data points, C is the number of classes, $c_i(c) = 1$ if class c is the true label for ith data point, 0 otherwise (i.e. one-hot encoding) and $y_i(c)$ is the predicted probability for class c in data point i. This function measures the dissimilarity between true labels (c_i) and predicted probabilities (y_i) for each class (c) and data point (i).

For both, the binary and multiclass classifier, we use a 50:50 train:test split of the input samples. The loss of the classifier per epoch is monitored for both the train and test samples during the training. In the fit model, only the test samples are used.

To avoid overfitting, regularization techniques such as batch normalization and drop-out [104] (of 0.2) for each intermediate layer is used. Skip connections for even numbered layers were further added to improve the information flow. The bias was initialized to 0.01 and weights using the Glorot Normal initialization [105]. The DNNs in the SR are trained inclusively in the transverse momentum of vector boson, while the multi-class DNNs in the HF CR are trained separately for med and high p_T STXS bins to get more accurate classification on backgrounds in the different phase spaces. The QCD multijet process is not used in the training due to its negligible contribution to analysis regions.

For the 2016, 2017, and 2018 analyses, we used LO V+jets samples for training. Even though LO V+jets samples of 2017 and 2018 are affected by the bug described in Section 3.4. Its effect in the training process can only lead to a sub-optimal classifier. This choice allows us to use the full NLO V+jets samples in the analysis leading to an improvement of the systematic uncertainty associated with the limited MC statistics by up to 20% in some STXS bins.

As discussed in Section 3.10.1, variables related to the parton shower, namely SA5 and the number of additional jets were found to be mis-modeled in the NLO V+jets samples for 2017 and 2018. The number of additional jet variables is corrected as presented in Section 3.10.1. As SA5 was found to be mis-modeled in low and med regions of the 2-lepton channel and 1-lepton channel, as shown in Figure 99 and Figure 100 respectively, it was decided not to use it in the DNN training. For the 0-lepton channel, the MC modelling of the data of the SA5 variable was found to be satisfactorily modeled in simulation as shown in Figure 101. The effect of not using SA5 only in 2017 in the low and med p_T regions (labeled as in 'Custom SA5 model'), instead of all the regions reduces the effect on inclusive sensitivity for 2017 data from around 13% to around 6%. Table 17 provides the Asimov fit uncertainties on the STXS POIs for all configurations of the fit model. The highest impact of the Custom SA5 model is on the most sensitive POI: ZH (250-400 GeV) and ZH (> 400 GeV) STXS bin.

	With SA5 model	No SA5 model	Custom SA5 model
ZH (75-150)	1.95	1.95	1.95
ZH (150-250, 0j)	1.22	1.28	1.31
ZH (150-250, ge1j)	2.22	2.35	2.33
ZH (250-400)	0.72	0.81	0.72
ZH (> 400)	1.17	1.26	1.17
WH (150-250)	1.23	1.41	1.41
WH (250-400)	0.92	1.06	1.06
WH (> 400)	1.32	1.48	1.48
inclusive	0.365	0.413	0.391

Table 17: Expected uncertainties on STXS POIs for various fit models: including SA5 variable as an input to MVA (With SA5 model); excluding SA5 from input to MVA (No SA5 model); and the custom input to MVA of SA5 as described in the Section 3.11.1 (Custom SA5). The p_T(V) is in units of GeV.

To reduce the effect of the mis-modelling of the DeepCSV variable, the binned DeepCSV WPs are used as an input to the DNN. The impact on the signal sensitivity of using such a binned DeepCSV variable instead of an un-binned DeepCSV variable as an input to the DNN was found to be negligible.

The normalized input variables used for the DNN training are shown in Figure 95 and 96 respectively. A complete list of variables used in the DNN training is given in Table 18. The same set of inputs is used for training the multiclass DNN for HF CR.

3.11.2 MVA in the resolved 2-lepton channel

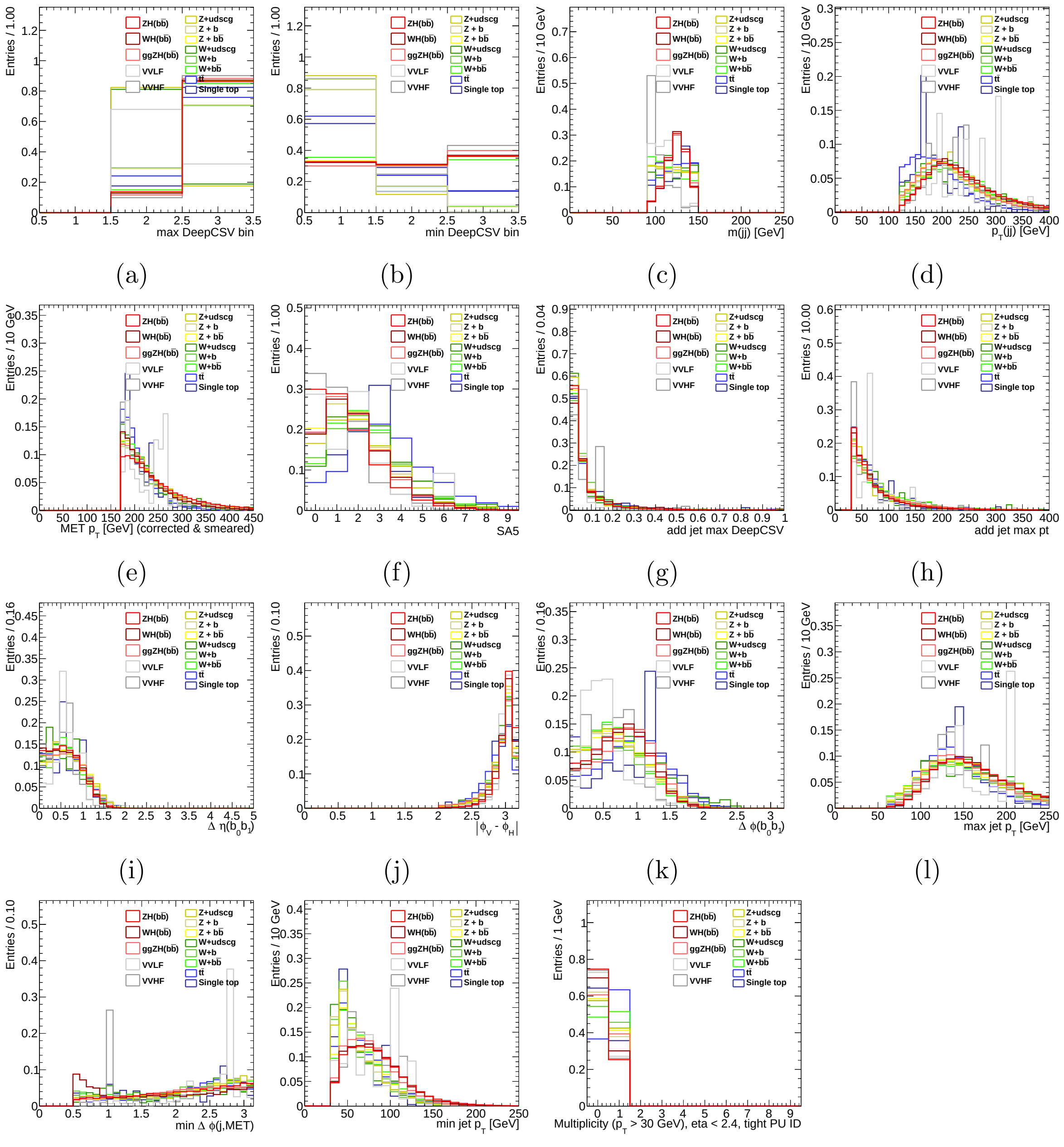
For the 2-lepton channel, a signal-background DNN classifier is trained separately for the low, med, and high (i.e. p_T(V) >250 GeV) STXS bins. Since the HF CR of 2-lepton is free from $\mathrm{t\bar{t}}$, a multiclass DNN is not required. Instead, we use bins of DeepCSV WP. This is discussed in Section 3.13.6. The architecture of the DNN is similar to the one used for the DNN training in the 0- and 1-lepton channels (discussed in Section 3.11.1). Similar to the DNN training in 0- and 1-lepton channels, QCD simulation samples are not used in the training. Also, LO V+jets samples are used for training.

The kinematic fit affected variables are input to the SR DNN. The use of variables with kinematic fit along with the variables without the kinematic fit leads to an Asimov significance (discussed in Section 3.12.4) improvement of almost 6%, with respect to using only variables without kinematic fit. The Higgs candidate jet variables are input to the DNN after applying the b-jet energy regression and the FSR recovery. The SA5 variable is not used in the training of the DNN targeting the low and med STXS bins due to the mis-modelling discussed in Section 3.11.1. Similar to the DNNs of 0- and 1-lepton channels, the WP-binned DeepCSV is used in the training. The complete list of variables used in the DNN training is given in Table 18.

The input variables used for the DNN training of the 2-lepton channel, in the resolved signal regions in low and med STXS, bins are shown in Figure 97 and 98 respectively.

Variable	explanation	0-lepton	1-lepton	2-lepton	
					kin fitted
m(jj)	dijet invariant mass	✓	✓	✓	✓
pT(jj)	dijet transverse momentum	✓	✓	✓	✓
pT(MET)	transverse momentum of MET	✓	✓	✓	✓
V(mt)	transverse mass of vector boson		✓		
V(pt)	transverse momentum of vector boson		✓	✓	✓
pT(jj)/pT(V)	ratio of transverse momentum of higgs boson and vector boson		✓	✓	✓
$\Delta\phi(V,jj)$	azimuthal angle between vector boson and dijet directions	✓	✓	✓	✓
btag_{max} WP	1,2,3 if b-tagging discriminant (DeepCSV) score of leading jet is above T, M, L WP resp.	✓	✓	✓	✓
btag_{min} WP	1,2,3 if b-tagging discriminant (DeepCSV) score of sub-leading jet is above T, M, L WP resp.	✓	✓	✓	✓
$\Delta\eta(jj)$	pseudorapidity difference between leading and sub-leading jet	✓	✓	✓	
$\Delta\phi(jj)$	azimuthal angle between leading and sub-leading jet	✓	✓		
$pT_{max}(j_1,j_2)$	maximum transverse momentum of jet between leading and sub-leading jet	✓	✓		
pT(j_2)	maximum transverse momentum of jet between leading and sub-leading jet	✓	✓		
SA5	number of soft-track jets with pT $>$ 5GeV	✓		✓	
N_{aj}	number of additional jets	✓	✓		
$\text{btag}_{\text{max}}(add)$	maximum btagging discriminant score among additional jets	✓			
pT_{max}(add)	maximum transverse momentum among additional jets	✓			
$\Delta\phi(jet,MET)$	azimuthal angle between additional jet and MET	✓			
$\Delta\phi(lep,MET)$	azimuthal angle between lepton and MET		✓		
M_t	Reconstructed top quark mass		✓		
$pT(j_1)$	transverse momentum of leading jet			✓	✓
M_t	transverse momentum of sub-leading jet			✓	✓
$m(V)$	Reconstructed vector boson mass			✓	
$\Delta R(V,jj)$	angular separation between vector boson and Higgs boson			✓	✓
$\sigma(m(jj))$	resolution of dijet invariant mass				✓
N_{rec}	number of recoil jets				✓

Table 18: List of input variables used in the DNN training. The kin fitted column refers to the variables after the kinematic fit (discussed in Section 3.7.5) is applied.



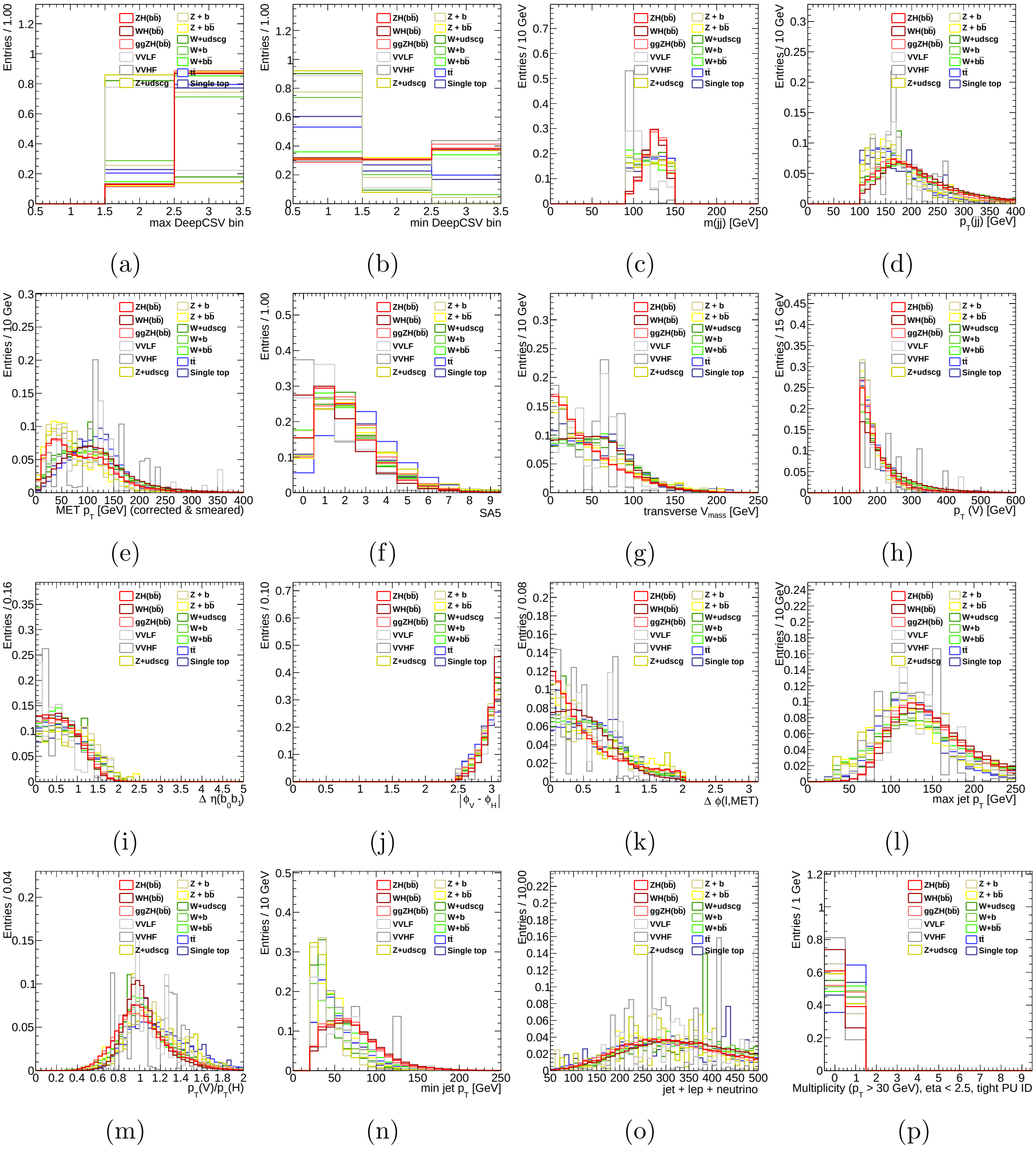
(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

(m) (n) (o)

Figure 95: Normalized shapes of the variables used for the training of the DNNs in the analysis SR and V+HF CR for the 0-lepton channel.



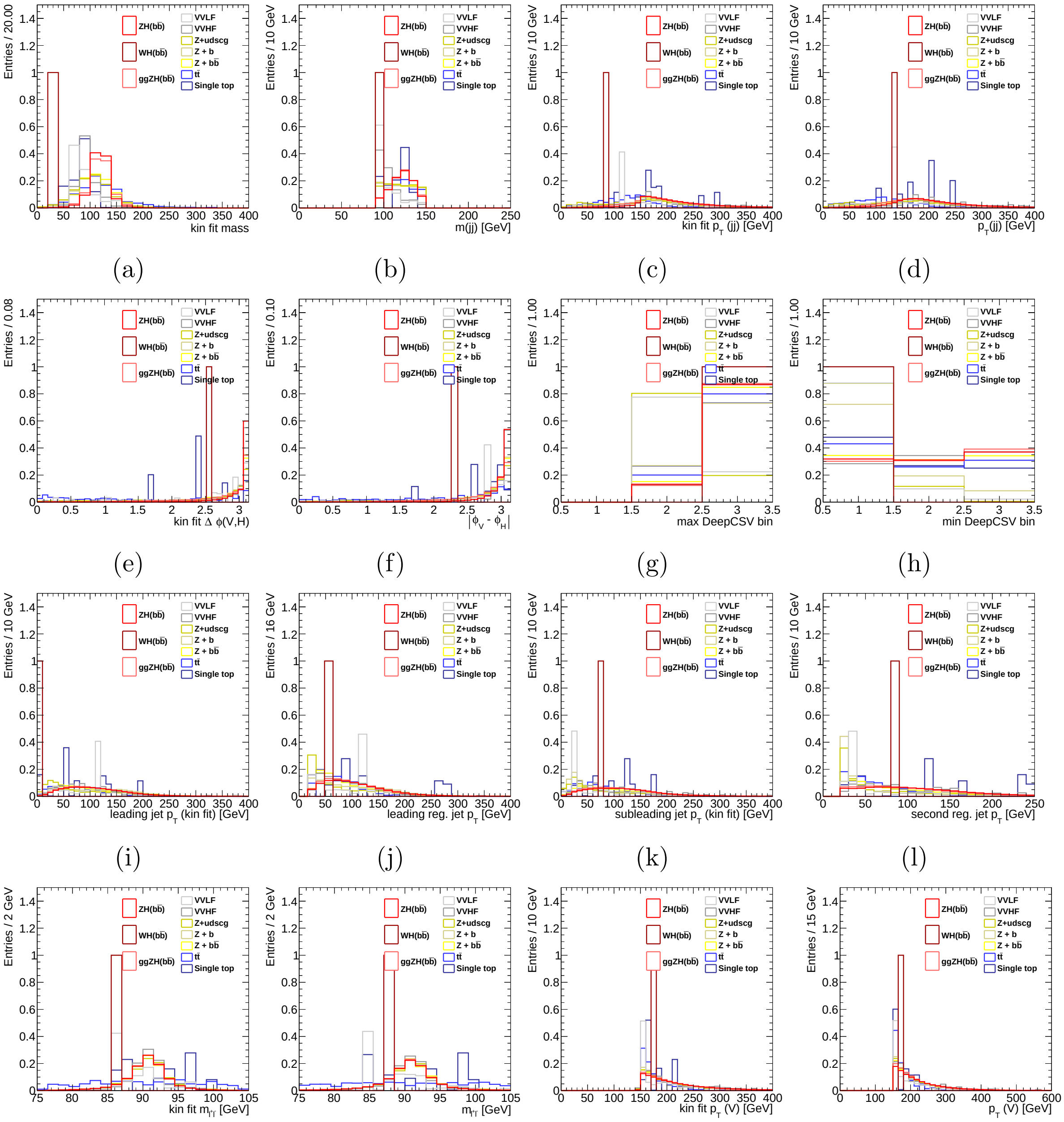
(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

(m) (n) (o) (p)

Figure 96: Normalized shapes of the variables used for the training of the DNNs in the analysis SR and V+HF CR for the 1-lepton channel.



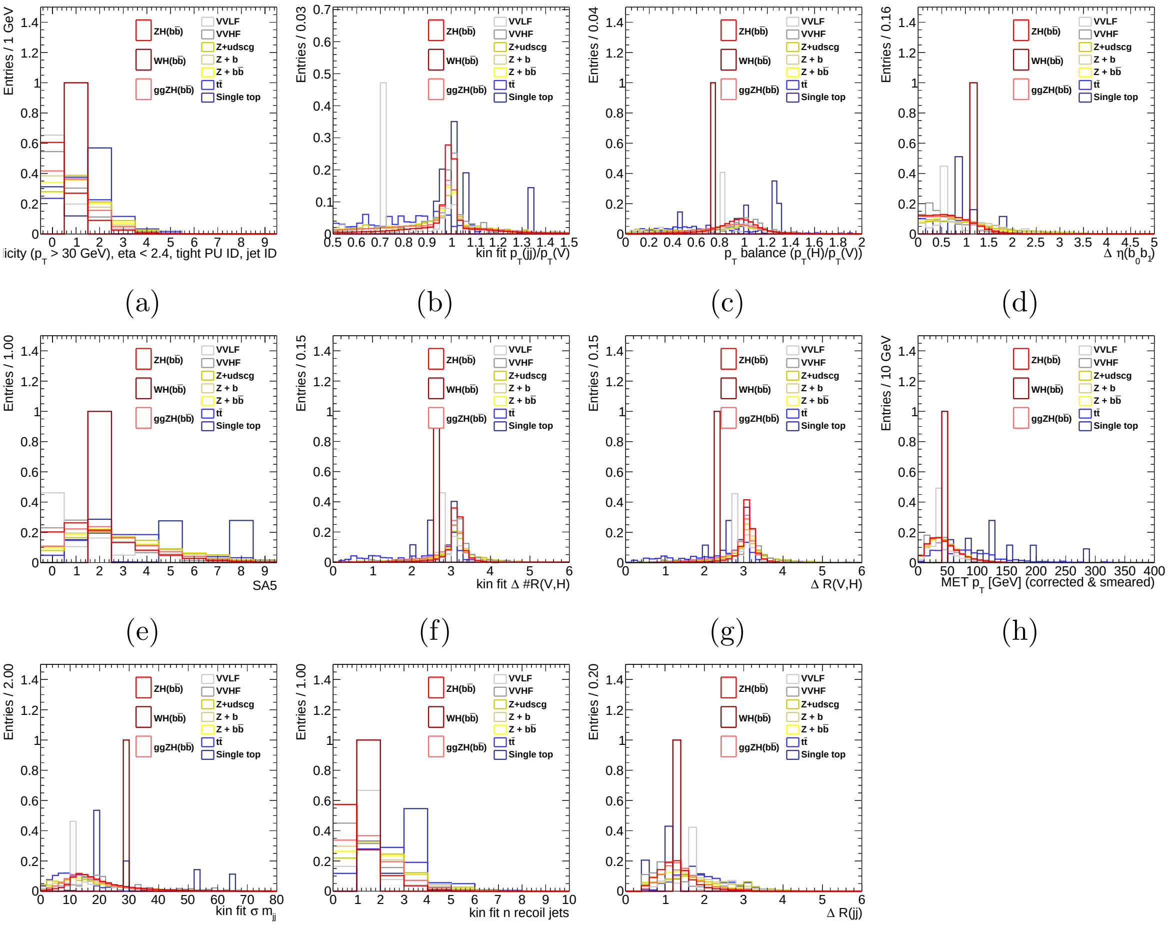
(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

(m) (n) (o) (p)

Figure 97: Normalized shapes of the variables used for the training of the DNNs in the analysis SR and V+HF CR for the 2-lepton channel.



(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k)

Figure 98: Normalized shapes of the variables used for the training of the DNNs in the analysis SR and V+HF CR for the 2-lepton channel.

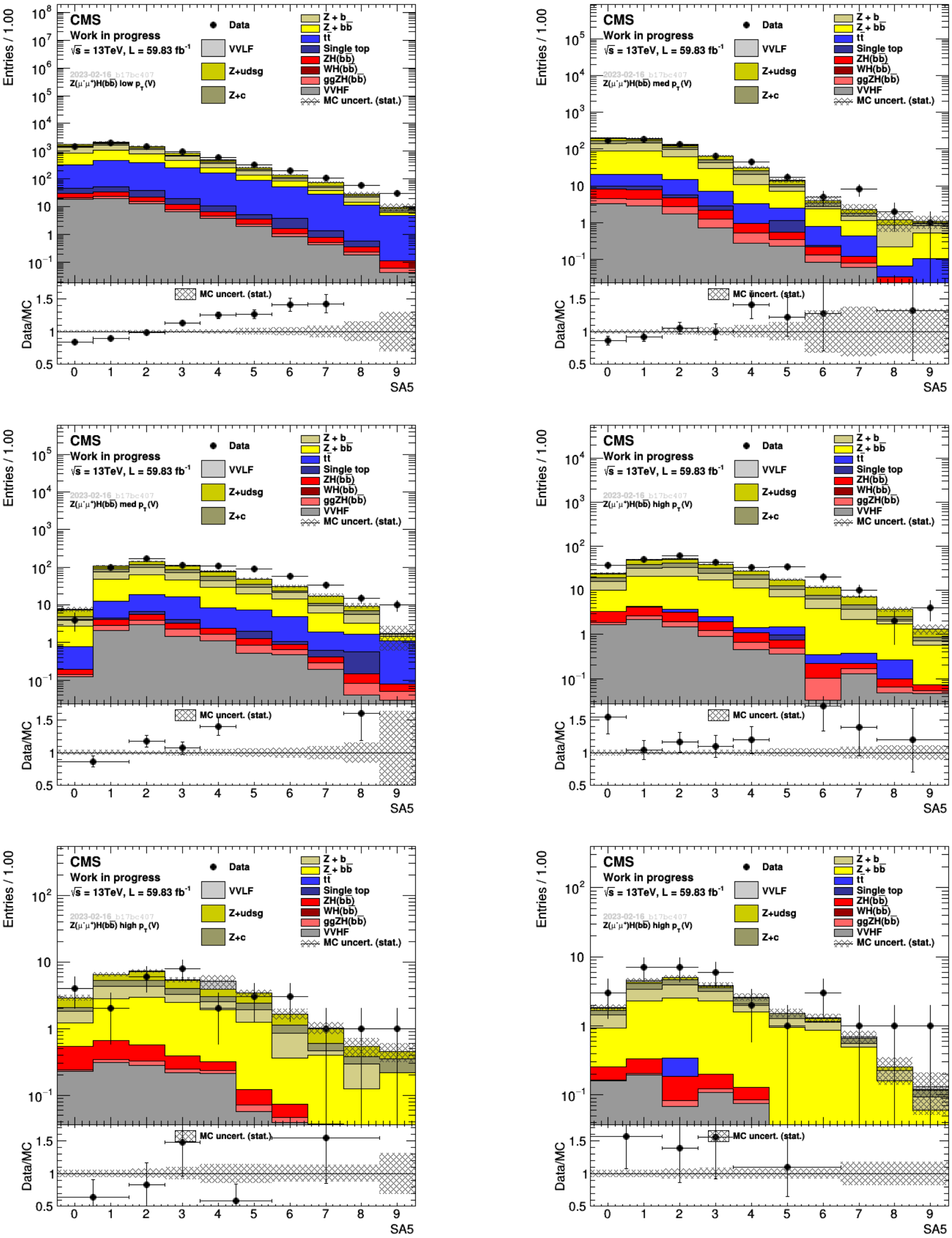


Figure 99: SA5 modelling in low (top left), med 0 additional jet (top right), greater than 0 additional jet (middle left), high1 (middle right), high2 (bottom left) and high2 BOOST (bottom right) STXS SR of 2-lepton channel.

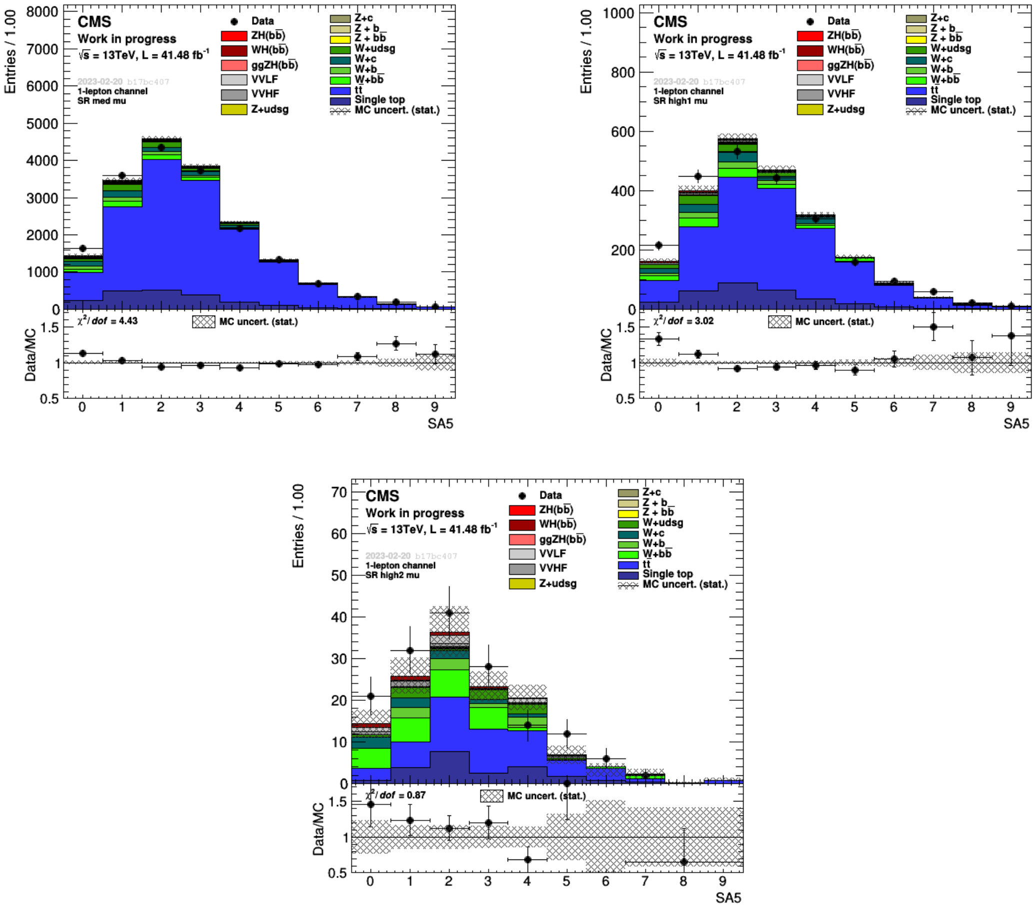


Figure 100: SA5 modelling in med (top left), high1 (top right), and high2 (bottom) STXS SR of 1-lepton channel.

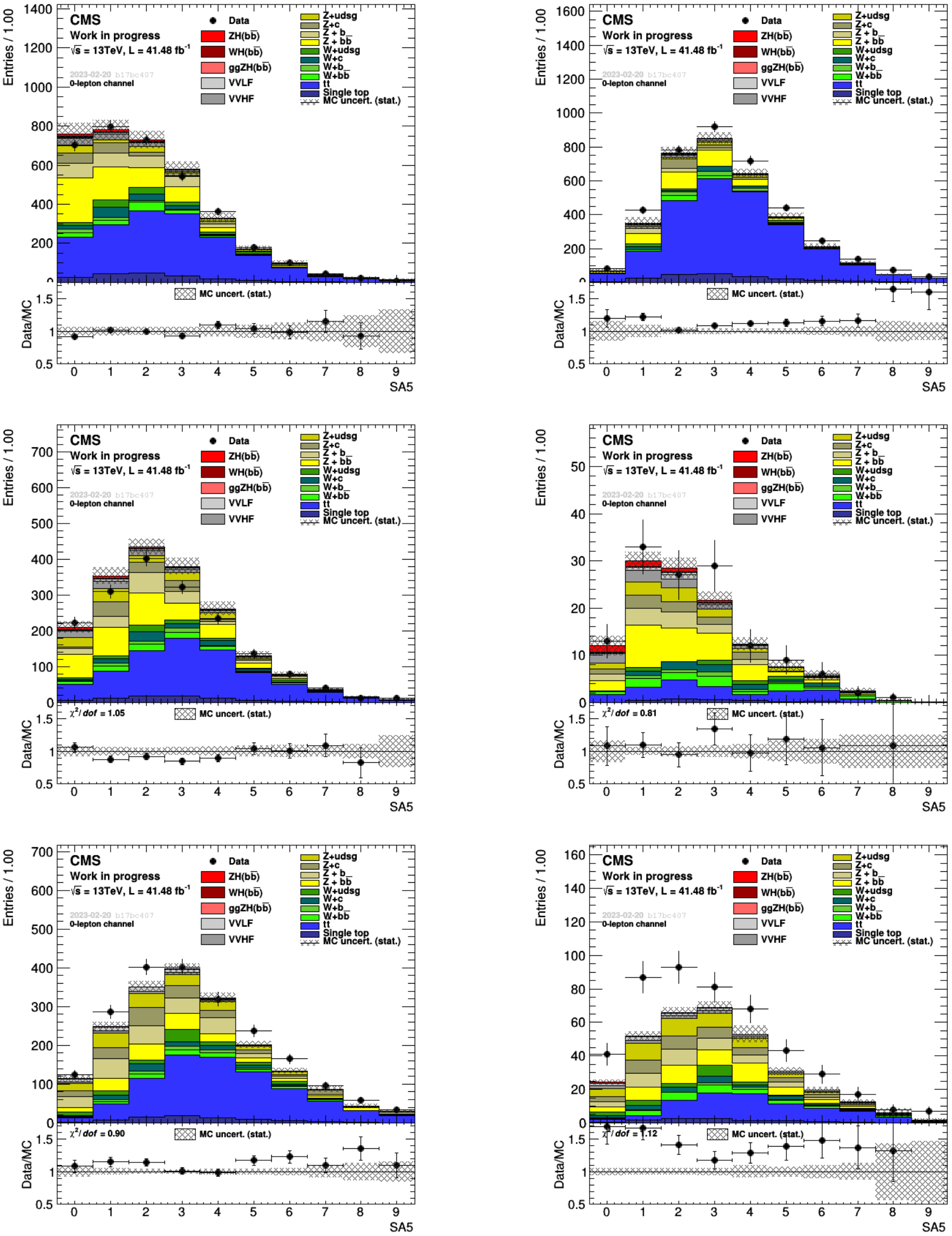


Figure 101: SA5 modelling in med 0 additional jet (top left), greater than 0 additional jet (top right), high1 (middle left), high2 (middle right) and high1 BOOST (bottom left) and BOOST high2 (bottom right) STXS SR of 0-lepton channel

3.11.3 DNN application

The data/MC distribution of DNN in the STXS SR bins for 0-lepton, 1-lepton and 2-lepton channels are shown in Figure 102, 103 and 104 respectively. The multiclass DNNs evaluated for HF CR in all three channels are shown in Figure 105.

3.11.4 BDT for boosted SR

For the boosted topology the signal is extracted from a fit to the output of BDTs in various SRs. The BDTs are trained separately for each channel. The BDT architecture (implemented in ROOT [106]) consists of 100 trees, using 20 cuts with a minimum node size of 0.05. The BDTs were checked for overtraining and no overtraining was observed. As in the resolved topology, the QCD process is not used due to its negligible contribution in the analysis regions.

3.11.5 Input variables for BDT

The following four features are used for purely boosted events:

- Soft-drop mass of the FatJet candidate;
- Transverse momentum of the FatJet candidate;
- Transverse momentum of the vector boson;
- DeepAK8bbVSlight FatJet output node.

Events in the boosted category can also have additional resolved AK4 jets. In cases of events classified as boosted but having two AK4 resolved jets as well, we use both the input features from resolved DNN and the ones of pure boosted events. This leads to an improvement of about 15% in sensitivity.

Because the calibration factors for DeepAK8 output nodes were not available for full shape, due to a lack of statistics, similar to the resolved analysis, instead of the DeepAK8 full shape, we use only DeepAK8 WP at 0.97. Similarly, only calibration factors of the DeepAK8bbVSlight node were available and so only the binned bbVSlight output node of DeepAK8 was used as input.

The loss in signal sensitivity due to the usage of the binned version of DeepAK8bbVSlight as opposed to full shape is about 7%. The loss in signal sensitivity due to the usage of only bbVslight output node as opposed to all output nodes is about 10%.

3.11.6 BDT application

The data/MC distribution of BDT in the STXS SR high p_T bins for 0-lepton, 1-lepton, and 2-lepton channels are shown in Figure 102, 103 and 104 respectively.

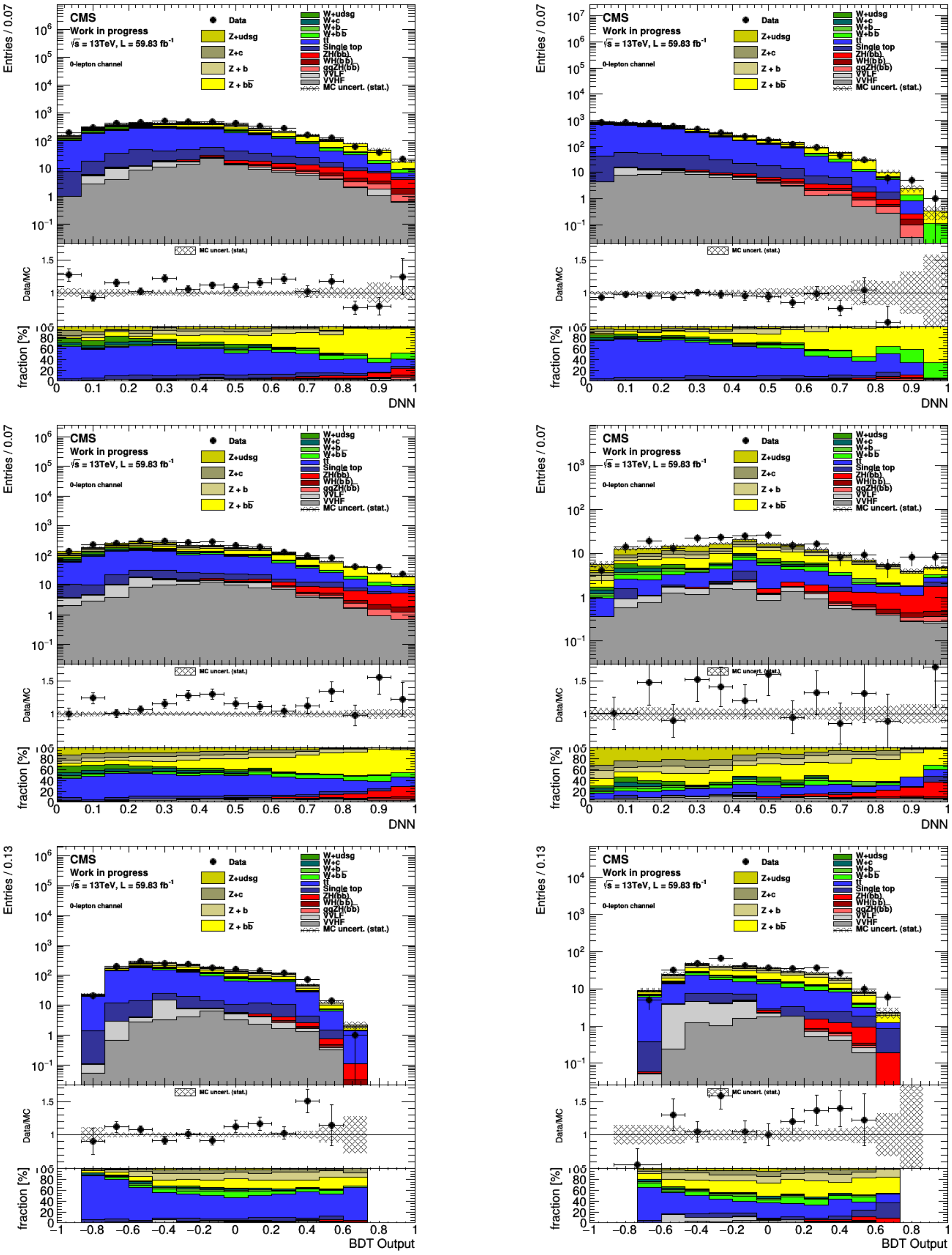


Figure 102: Data/MC comparison of the DNN in SR med 0 additional jet (top left), greater than 0 additional jet (top right), high1 (middle left), high2 (middle right) and the BDT output in high1 BOOST (bottom left) and high2 BOOST (bottom right) STXS bin of 0-lepton channel. The bottom plot gives the fractional component of different processes in a bin.

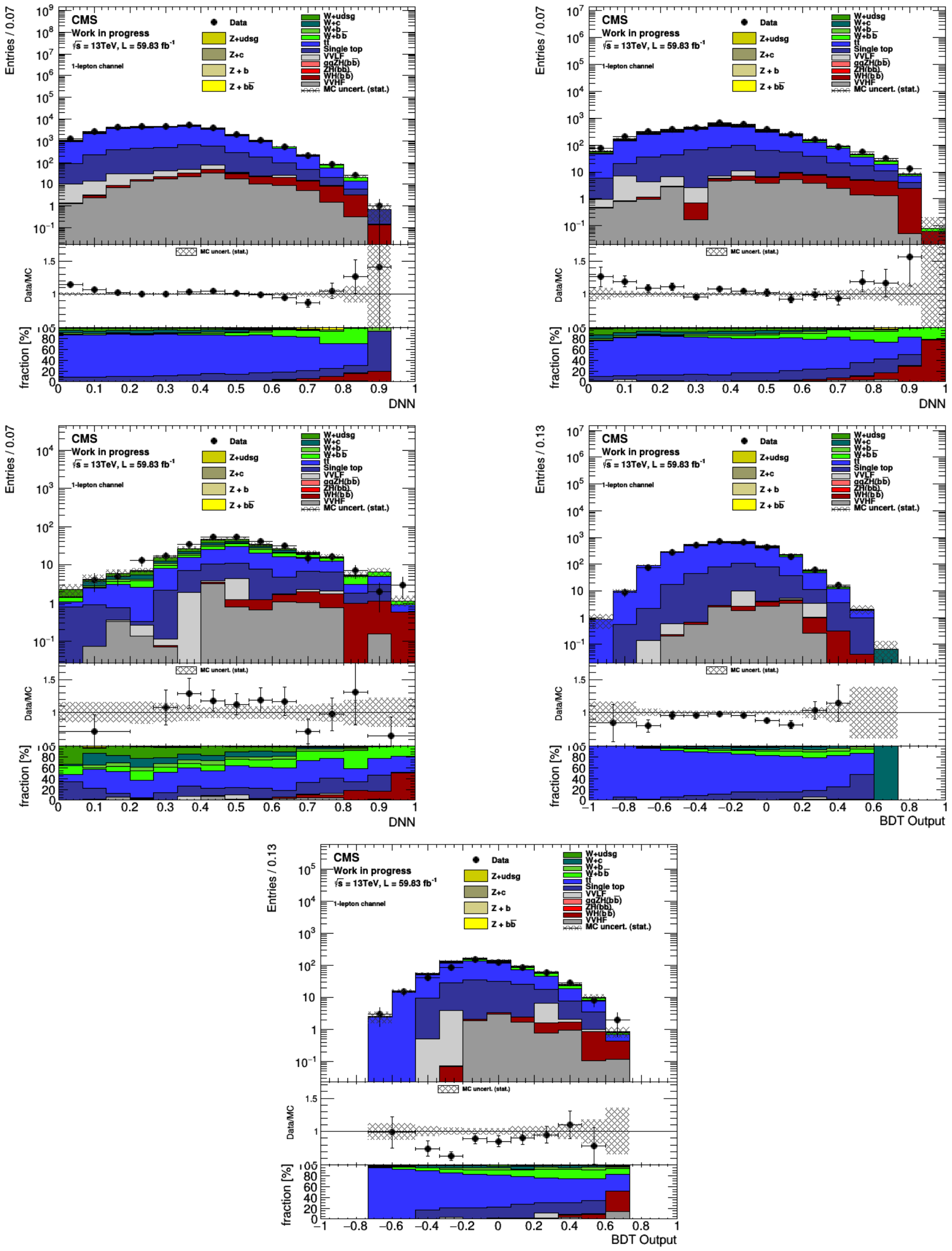


Figure 103: Data/MC comparison of the DNN in SR med (top left), high1 (top right), high2 (middle left) and the BDT output in high1 BOOST (middle right) and high2 BOOST (bottom left) STXS bin of 1-lepton channel. The bottom plot gives the fractional component of different processes in a bin.

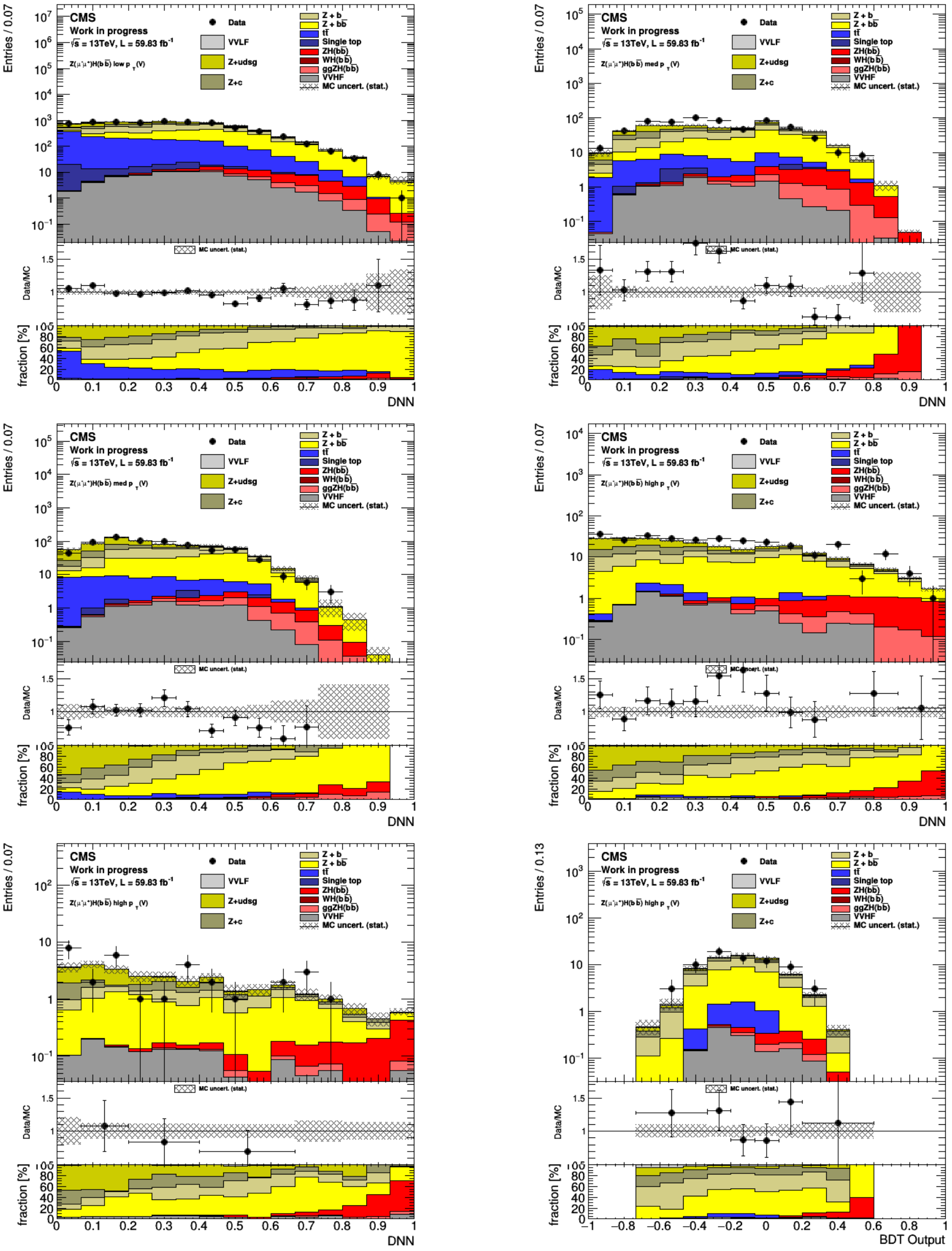


Figure 104: Data/MC comparison of the DNN in SR low (top left), med 0 additional jet (top right), greater than 0 additional jet (middle left), high1 (middle right), high2 (bottom left) and the BDT output in high1 BOOST (bottom right) STXS bin of 2-lepton channel. The bottom plot gives the fractional component of different processes in a bin.

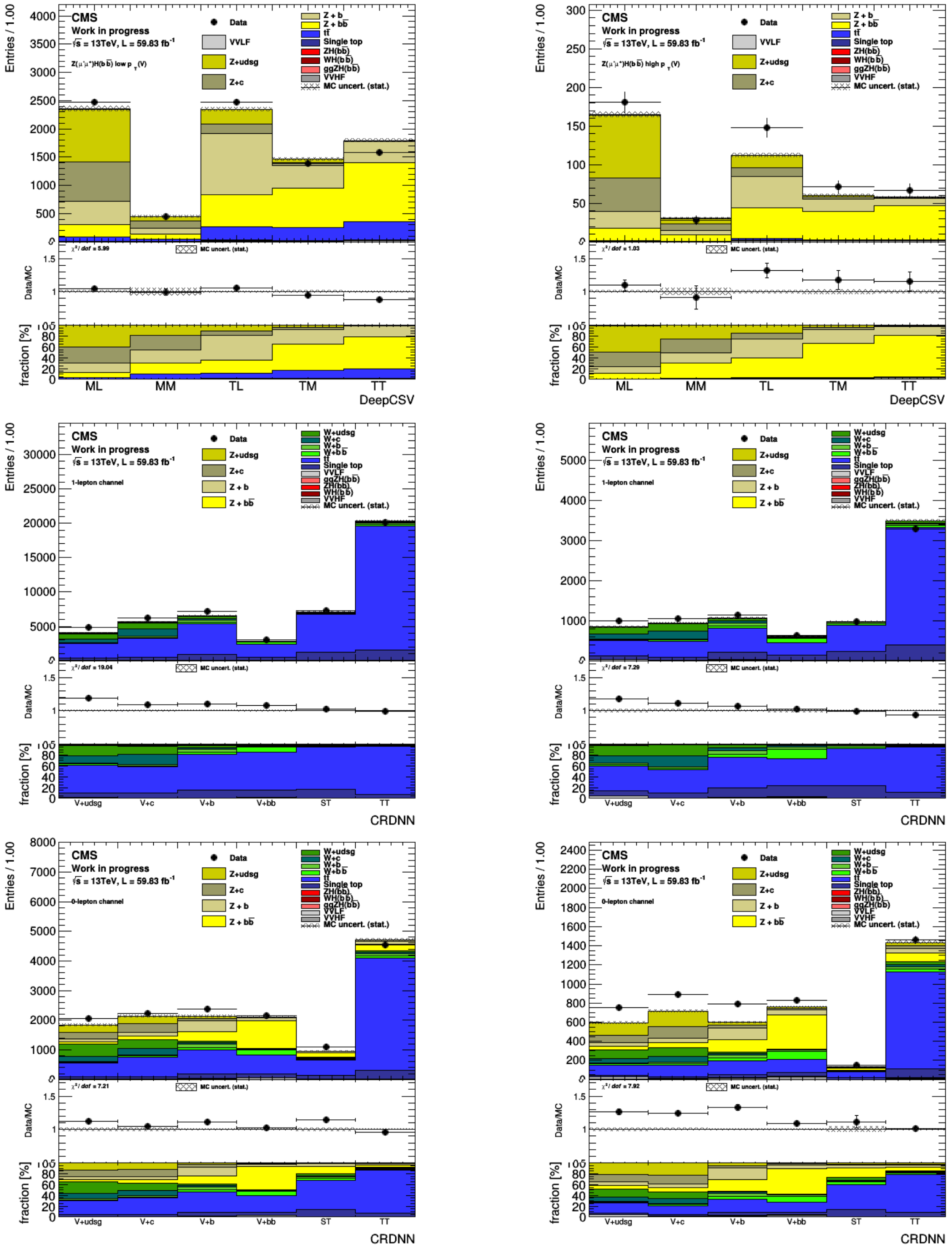


Figure 105: Data/MC comparison of the DeepCSV binned input for V+HF CR for the 2-lepton channel in the fit for low (top left) and high (top right) STXS bin. Data/MC comparison of the Multiclass DNN in V+HF CR in med (middle left) and high (middle right) in the 1-lepton channel and in the bottom row for the 0-lepton channel. The gives the fractional component of different processes in a bin.

3.12 Statistical Analysis

The probability of observing n signal events in N proton-proton collision when λ signal events are expected is given by the binomial distribution:

$$p(n \mid \lambda, N) = \binom{N}{n} \left(\frac{\lambda}{N}\right)^n \left(1 - \frac{\lambda}{N}\right)^{N-n} \tag{3.13}$$

In the limit $N \to \infty$, the probability distribution can be described by a Poisson distribution where λ is the rate of expected signal events:

$$p(n \mid \lambda) = \frac{\lambda^n}{n!} e^{-\lambda} \tag{3.14}$$

If n_i is the number of events in the ith bin having s_i signal events and b_i background events, Equation 3.14 can be expressed as:

$$p(n \mid \lambda) = \prod_i \frac{\left(\mu s_i(\vec{\theta}) + b_i(\vec{\theta})\right)^{n_i}}{n_i!} e^{-\left(\mu s_i(\vec{\theta}) + b_i(\vec{\theta})\right)} \tag{3.15}$$

where μ, known as the signal strength, is the ratio of the observed by the SM signal cross-section:

$$\mu = \frac{\sigma}{\sigma_{\mathrm{SM}}} \tag{3.16}$$

The estimated number of signal and background events (s_i and b_i respectively) are functions of nuisance parameters ($\vec{\theta}$) (parameters in the model apart from the POI). Nuisance parameters that have positive and negative values are generally modeled as Gaussian distribution, while parameters like cross-section, which are strictly positive, have associated nuisances described as log-normal distribution. Some other nuisances like process scale factors are modeled with flat priors for positive yields with an associated uncertainty. In the case of Gaussian distributed nuisances, the likelihood function is given by

$$L(\mu, \vec{\theta}) = \prod_i \frac{\left(\mu s_i(\vec{\theta}) + b_i(\vec{\theta})\right)^{n_i}}{n_i!} e^{-\left(\mu s_i(\vec{\theta}) + b_i(\vec{\theta})\right)} \prod_\kappa \frac{{u_\kappa}^{m_\kappa}}{m_\kappa!} e^{-u_\kappa} \tag{3.17}$$

The second term in the likelihood function accounts for the control region or auxiliary histograms ($\mathbf{m} = (m_1, \ldots, m_M)$ with $E\left[m_i\right] = u_i(\vec{\theta})$) used for constraining the background in the analysis and is similarly described by a Poison distribution.

For hypothesis testing of two simple hypotheses $H_0 : (\mu, \vec{\theta}) = (\mu_0, \vec{\theta_0})$ (null hypotheses) and $H_1 : (\mu, \vec{\theta}) = (\mu_1, \vec{\theta_1})$ (alternative hypotheses), the Neyman-Pearson lemma [107] states that the likelihood ratio is the most powerful test. For composite or complex hypotheses involving more than one nuisance, the Neymann-Pearson lemma is not applicable. However, when nuisances are well-constrained, the complex hypothesis can be assumed to be simple, and thus the likelihood ratio can be considered as an optimal test.

3.12.1 Profile Likelihood Ratio and observed significance

The test statistic used for hypothesis testing at LHC is a profile likelihood ratio defined as:

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\vec{\theta}}})}{L(\hat{\mu}, \hat{\vec{\theta}})} \tag{3.18}$$

where $\hat{\vec{\mu}}$ and $\hat{\theta}$ are the parameter estimates obtained through maximum likelihood (ML) fit and $\hat{\hat{\vec{\theta}}}$ is the ML estimate at a fixed μ value.

Since we are searching for an excess of events above the background, the number of signal events has to be positive and the test statistic from Equation 3.18 can be written as

$$\tilde{q}_0 = \begin{cases} -2\ln\lambda(\mu) & \text{for } \hat{\mu} > 0 \\ 0 & \text{for } \hat{\mu} \leq 0 \end{cases} \tag{3.19}$$

Using the observed value of the test statistic, the p-value of the background-only hypothesis ($\mu = 0$) can be obtained as

$$p_0 = \int_{\tilde{q}_{0,\text{ obs.}}}^{\infty} f(\tilde{q}_0 \mid \mu = 0)\mathrm{d}\tilde{q}_0 \tag{3.20}$$

alternatively, the corresponding z-statistic value can be given as:

$$Z = \phi^{-1}(1 - p_0) \tag{3.21}$$

where ϕ^{-1} is the inverse cumulative distribution function.

Discovery corresponds to a p-value of 2.87×10^{-7} or a z-statistic value of 5. The p-value can be obtained through the distribution of the test statistics (obtained from multiple toy distribution which can be computationally expensive) or directly by using the asymptotic formula given by the Wilks theorem (explained in Section 3.12.3).

3.12.2 Wald's theorem

The Wald's theorem states that for a sufficiently large data sample, the test statistic from Equation 3.19 can be approximated as:

$$-2\ln\lambda(\mu) = \frac{(\mu - \hat{\mu})^2}{\sigma^2} \tag{3.22}$$

This test statistic is distributed as a 'non-central χ^2' distribution [108] with one degree of freedom given as:

$$f(q;\Lambda) = \frac{1}{2\sqrt{q}}\frac{1}{\sqrt{2\pi}}\left[\exp\left(-\frac{1}{2}(\sqrt{q}+\sqrt{\Lambda})^2\right) + \exp\left(-\frac{1}{2}(\sqrt{q}-\sqrt{\Lambda})^2\right)\right] \tag{3.23}$$

with non-centrality parameter Λ given as:

$$\Lambda = (\mu - \mu')^2/\sigma^2 \tag{3.24}$$

Here μ' is the mean of the Gaussian distribution of $\hat{\mu}$. The importance of Wald's theorem is that we can write the distribution of the test statistics in a closed analytic form.

3.12.3 Wilk's theorem and observed significance

The wilk's theorem describes the special case when $\mu' = \mu$ and $\Lambda = 0$, the test statistic distribution from Equation 3.23 approaches a χ^2 distribution with one degree of freedom.

For the background-only hypothesis ($\mu = 0$), the test statistic from Equation (3.19) in the large sample approximation (using equation 3.22) can thus be written as:

$$q_0 = \begin{cases} \hat{\mu}^2/\sigma^2 & \text{for } \hat{\mu} \geq 0 \\ 0 & \text{for } \hat{\mu} < 0 \end{cases} \tag{3.25}$$

The corresponding distribution of test statistic simplifies to:

$$f(q_0 \mid 0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}e^{-q_0/2} \tag{3.26}$$

and the significance/p-value can be computed easily from the observed test statistic as:

$$Z_0 = \sqrt{q_0} \tag{3.27}$$

3.12.4 Asimov dataset and expected significance

The Asimov dataset is a dataset that is constructed such that each histogram bin has a number of events equal to its expected value i.e. $\mu = 1$ and $n = s + b$. The idea of having an Asimov dataset is that the corresponding expected significance can be obtained by calculating the test statistic on such an expected dataset.

Plugging the Poisson-distributed likelihood function

$$L(\mu) = \frac{(\mu s + b)^n}{n!}e^{-(\mu s + b)} \tag{3.28}$$

into the expression for the likelihood ratio, the significance can be computed to be:

$$Z_0 = \sqrt{q_0} = \begin{cases} \sqrt{2\left(n \ln \frac{n}{b} + b - n\right)} & \text{for } n \geq b \\ 0 & \text{for } n < b \end{cases} \tag{3.29}$$

The median value of the expected significance to reject the background-only hypothesis can be written as:

$$\text{median}(Z_0) = \sqrt{q_0^{\mathrm{A}}} = \sqrt{2\left((s+b)\ln\left(1 + \frac{s}{b}\right) - s\right)} \tag{3.30}$$

In the limit of $s \ll b$, Equation 3.30 simplifies to:

$$\sqrt{q_0^{\mathrm{A}}} = s/\sqrt{b} \tag{3.31}$$

In case we include uncertainty of the background prediction, the significance is given as:

$$\sqrt{q_0^{\mathrm{A}}} = \sqrt{2\left((s+b)\ln\left[\frac{(s+b)\left(b+\sigma_b^2\right)}{b^2+(s+b)\sigma_b^2}\right] - \frac{b^2}{\sigma_b^2}\ln\left[1 + \frac{\sigma_b^2 s}{b\left(b+\sigma_b^2\right)}\right]\right)} \tag{3.32}$$

3.13 Fit model for STXS signal strength extraction

To extract the signal strength in the STXS bins discussed in Section 3.9, we use the likelihood ratio from Equation 3.18. In the following Sections 3.13.1-3.13.6, we describe the input variables and associated systematic uncertainties used in the fit model.

3.13.1 Input templates of the fit model

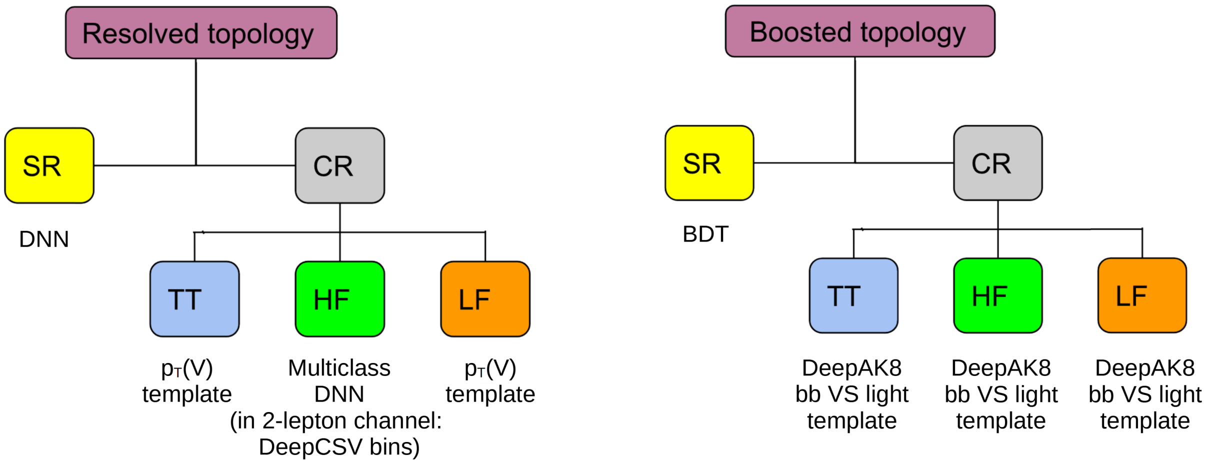

Resolved topology
SR
DNN
CR
TT
HF
LF
p_T(V) template
Multiclass DNN (in 2-lepton channel: DeepCSV bins)
p_T(V) template
Boosted topology
SR
BDT
CR
TT
HF
LF
DeepAK8 bb VS light template
DeepAK8 bb VS light template
DeepAK8 bb VS light template

Figure 106: Variables used in the fit model for resolved topology (left) and boosted topology (right).

The observables in the SR and CR used for signal extraction are shown in Figure 106. MVA variables (DNN or BDT) are used in the SR of all channels and HF CR of 1- and 0-lepton channels. For 2-lepton HF CR, we use the DeepCSV score of leading and subleading jet binned in 5 bins based on the WPs as shown in Figure 105. The first bin 'ML' refers to those events having the DeepCSV score of the leading jet between the Medium WP and Tight WP and that of the sub-leading jet between the Loose WP and Medium WP. These bins are designed to have the maximal separation between different flavors of V+jets events. For TT CR, we use the p_T of the recoiling vector boson in resolved topology and DeepAK8bbVSlight score in the boosted topology (further discussed in Section 3.13.5). For LF CR, we use p_T of the vector boson in resolved topology and DeepAK8bbVSlight output score in boosted topology. For HF CR in boosted topology, we use the DeepAK8bbVSlight output node of DeepAK8. These inputs in the fit model are used separately for the electron and muon channels in 1 and 2-lepton channels.

Further along with these nominal templates, several corresponding systematic variations are also used in the fit model as described in the following section.

3.13.2 Systematic uncertainties in the fit model

Several sources of systematic uncertainties are considered in this analysis. They can affect either the normalization of the process or both the normalization and shape of the process. The following Sections summarize the systematic uncertainties affecting the MC samples used to build the fit model.

3.13.3 Theoretical uncertainties

Theoretical uncertainties on the signal The STXS framework introduces sensitivity to higher-order QCD effects through mutually exclusive bins leading to additional sources of uncertainty compared to analyses targeting inclusive phase spaces. The dominant theory uncertainties come from perturbative QCD scale variations which can affect both

- the overall cross section
  - The normalization uncertainties on the total cross section of qqZH, ggZH, and qqWH are 0.5%, 22.0%, and 0.6% respectively [109][110][111].
- and the cross section of the individual STXS bins.
  - Since the reconstructed observables of particles are affected by resolution effects, events at the particle level in a given STXS bin can get reconstructed in its neighboring STXS. Since such migration can happen at any of the STXS bin boundaries, the corresponding migration uncertainties are considered at every bin boundary. These migration uncertainties are normalization uncertainties. Migration of events within a given 'merged' STXS bin (for example, for WH, the 0 and ≥ 1 additional jet categories are merged in Figure 70), do not change the overall cross section of the sum of those bins since migration happens at the boundary of the bins which are merged. The corresponding acceptance uncertainties are shape uncertainties.

Theoretical uncertainty on the total cross section and the migration uncertainties introduce changes in the normalization of the cross section and are considered only for signal strength measurement. Acceptance uncertainties affect the shape of the input templates and are considered for both the signal strength and cross section measurements.

The migration and acceptance uncertainties are calculated using the maximum splitting scheme [112] where each of the bins from Figure 70 contributes. In this scheme, for migration uncertainty, every bin boundary is considered as a source of uncertainty. This leads to 6 sources of uncertainty corresponding to p_T(V) boundary of 75, 150, 250, and 400 GeV, and the number of additional jet boundaries at 0 and > 0 additional jets. In each of the bins (i), six combinations of renormalization and factorization scale variations

$$[\mu_{\mathrm{R}}/\mu_{\mathrm{R}}^{\mathrm{nom}}, \mu_{\mathrm{F}}/\mu_{\mathrm{F}}^{\mathrm{nom}}] = [1/2, 1], [1, 1/2], [1, 2], [2, 1], [1/2, 1/2], [2, 2] \tag{3.33}$$

are computed and the maximal one is used as the absolute uncertainty Δ_i. The individual Δ_i is then varied in an anti-correlated manner (across the p_T(V) or the number of additional jets bins on the left and right side of the bin boundary under consideration) to compute relative migration uncertainty Δ_t at the bin boundary t. Short-range anti-correlation scheme-2 [112] is used wherein the migration effects are assumed to occur only to the first neighboring bin below the bin boundary t.

However, since the analysis is not sensitive to each of the bins as discussed in Section 3.9, we compute the relative migration uncertainty of the merged bin as the mean of the relative migration uncertainty of individual bins. The acceptance uncertainty is then given by the residual difference between the merged relative bin uncertainty and individual relative bin uncertainty. The relative uncertainties on WH and qq→ZH are similar while gg→ZH is larger due to the smaller cross section [113]. Also, the uncertainties related to the number of additional jet bins dominate over the uncertainties between the p_T(V) bins.

Apart from theory uncertainties from perturbative QCD scale variations, other sub-dominant theoretical uncertainties on the signal consist of PDF uncertainties, electroweak uncertainties, and uncertainty on branching ratio. These uncertainties are implemented with log-normal priors.

- The analysis uses the nominal NNPDF 3.0 PDF [114] and the PDF uncertainties are derived using alternative PDFs from the same set. For signal processes, the PDF acceptance uncertainty is 1% for all processes while the PDF uncertainty on the cross section is 1.6% for qqZH, 1.9% for qqWH, and 2.4% for ggZH.

- Due to the presence of electroweak bosons in the signal process the theoretical computations are affected by electroweak corrections. A 2% log-normal prior is assumed to account for the corresponding uncertainty for all the signal processes [115].

- The uncertainty on the branching ratio of decay of the Higgs boson to bottom quark pairs is 0.5% [115].

Theoretical uncertainties on the background

- Since the normalization for single top and diboson is not modelled using freely-floating SF, as done with other major background processes, a normalization uncertainty of 15% is assumed on their measured cross-section. This is larger than the uncertainty of the recent cross section measurement [64]. Since the phase space for the cross section measurement and the current STXS VH(→ b$\bar{\text{b}}$) analysis are different, we allow for these additional phase space differences to be accounted for. The potential bias from the pre-fit normalization of the diboson was studied by generating toys, assuming the normalization of diboson was within 1σ of the assumed uncertainty (i.e. ±15%), and fitting each of these toys with the model with nominal prefit normalization. Figure 107 shows the distribution of the inclusive signal strength (μ) extracted fitting the model with nominal diboson prefit normalization to the model with prefit diboson normalization up to $+1\sigma$ (left) and -1σ (right). This shows that the model with nominal diboson prefit normalization has a negligible bias with respect to to a model with the prefit diboson normalization scaled within 1σ.

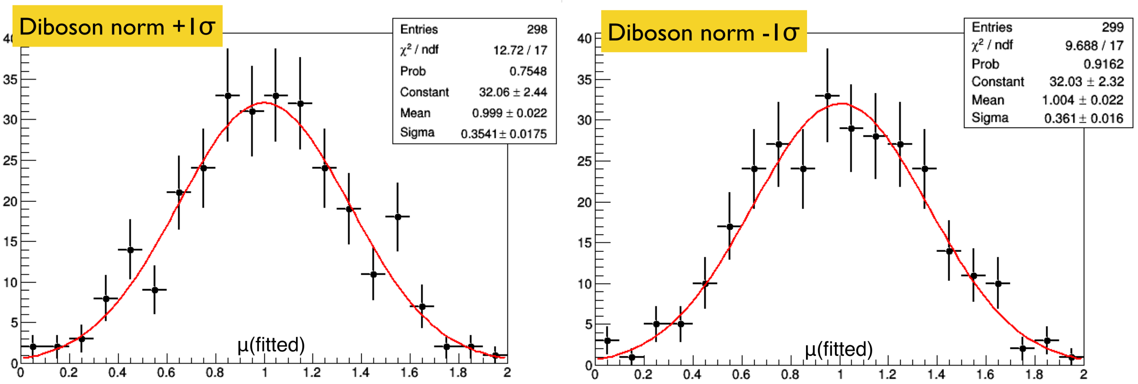


Figure 107: Distributions of the inclusive signal strength (μ) extracted by fitting the model with nominal diboson prefit normalization to toys generated with a model with diboson prefit normalization scaled by $+1\sigma$ (left) and -1σ (right) [116].

- The effect of higher-order QCD corrections which are not accounted for the simulation is estimated by varying the renormalization and factorization scales up and down by a factor of two, excluding the extreme variations where one scale is varied up by factor two and the other down by factor 1/2. These shape uncertainties are decorrelated among the background processes in the fit since the effect of higher-order QCD corrections can be different for different processes.

- The QCD renormalization and factorization scale uncertainties for NLO V+jets samples in 2017 and 2018 were found to be very large compared to LO samples, which led to them being highly constrained in the fit. For NLO samples, these uncertainties are expected to be equivalent to or smaller than the corresponding LO uncertainties. It was verified that scaling these uncertainties by a factor of 1/2, 1/4, 1/5, 1/6, 1/7, and 1/10 had a negligible impact on the signal strength extraction. An arbitrary scaling of a factor of 6 was found to provide postfit constraints equivalent to prefit constraints. Thus, both QCD renormalization and factorization scale uncertainties for NLO V+jets samples in 2017 and 2018 were scaled by a factor of 1/6.

- Since NLO and LO samples are generated with different event generators, a check was conducted by decorrelating the QCD renormalization and factorization scale nuisance parameters for the V+jets processes in 2017 and 2018 analyses with the V+jets processes in 2016 analysis in the fit model. This is because the 2017 and 2018 V+jets samples are generated at NLO in QCD while 2016 V+jets samples are generated at LO in QCD. This test was found to have a negligible impact on the signal strength. Thus, these uncertainties were kept correlated among the years for a particular process in the Run2 combined fit.

- As discussed in Section 3.13.3, PDF uncertainties affecting the analysis acceptance are estimated using the alternative PDFs from the nominal NNPDF 3.0 PDF set. They are implemented as log-normal uncertainties and are decorrelated among different flavors and processes. They are listed in Table 19.

Process	PDF acceptance uncertainties
Z+udsg	5%
Z+c	5%
Z+b	3%
Z+bb	2%
W+udsg	5%
W+c	5%
W+b	3%
W+bb	2%
$t\bar{t}$	0.5%
VVHF	2%
WZLF	2%
ZZLF	3%

Table 19: PDF acceptance uncertainties for background processes.

MC V+jets uncertainties

- An uncertainty on the $\Delta\eta$(bb) weight was introduced in the form of the square of the LO to NLO $\Delta\eta$(bb) correction weight in the 2016 analysis. Only the up variation to the weight was applied. Figure 108 shows the $\Delta\eta$(bb) corrections versus its relative uncertainty assigned in the analysis. As shown, the larger the correction factor the larger the uncertainty associated with it, and thus the corresponding nuisance parameter gets constrained in the fit. These uncertainties are decorrelated among V+jet flavors to avoid artificially over-constraining the nuisance parameter in the fit.

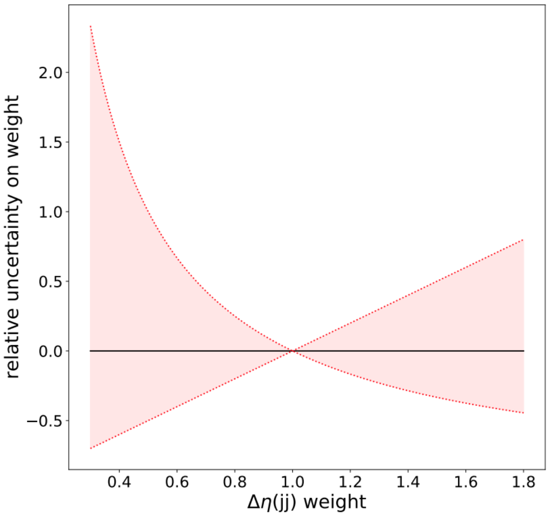


Figure 108: Relative uncertainty versus its $\Delta\eta$(bb) weights assigned in the 2016 analysis.

- For 2017 and 2018 analyses, the statistical uncertainty associated with the

ΔR(bb) weights (Section 3.10.1) is used in the fit model as a shape uncertainty. These uncertainties are also correlated among the years and the jet flavor.

- For the 2018 p_T(V) split samples corrections discussed in Section 3.10.1, the uncertainties on the fit parameters are decorrelated for Z+jets and W+jets processes and each p_T slice of the samples.
- For the number of additional jet corrections for 2017 and 2018, discussed in Section 3.10.1, the uncertainties from the fit parameters are uncorrelated for Z+jets and W+jets processes. A bias test was performed by generating Asimov toys with the model having inputs corrected for the number of additional jet mis-modelling, and fitted with the model having mis-modelled inputs. An additional nuisance parameter was introduced in the text as a systematic uncertainty to cover for the mis-modelling. It was found that the nuisance parameter gets pulled by around 1σ and the injected signal strength was recovered in each STXS bin. This test further assures the necessity of the correction given its non-negligible impact on the signal strength.

3.13.4 Experimental uncertainties

Luminosity and Pile-up Several detectors like the Pixel Luminosity Telescope and the pixel detector are used to measure the instantaneous luminosity in CMS [117]. The total uncertainty on the measurement of the integrated luminosity is 2.5% in 2016 [118], and 2018 [119], and 2.3% in 2017 [120]. These uncertainties are partially correlated between the different data-taking periods ($2016 - 2017, 2017 - 2018$) following the recommended scheme from CMS, and are implemented as log-normal uncertainties.

Pile-up weights discussed in Section 3.10.1 affect both the shape and normalization of all processes. Uncertainties on the pile-up weights are obtained by varying the minimum bias cross section up and down by 4.6% [121], and are implemented as shape uncertainties. Due to the mis-modelling in the prefit data/MC in pile-up distribution discussed in Section 3.10.1, it is observed that the corresponding nuisances get pulled in the fit to correct, as shown in Figure 130.

Lepton and Trigger efficiencies The uncertainties on the ID, isolation, and triggers of electron and muon SF are assumed to be 2% for the 1-lepton channel and 4% for the 2-lepton channel. The uncertainties for electron and muon are implemented as uncorrelated. The uncertainties on the MET trigger efficiency measurement amount to 0.5%. These nuisances are implemented as log-normal uncertainties as the normalization uncertainty dominates on uncertainties due to shape effects.

MC simulation statistical uncertainties The uncertainty on MC statistics per bin is calculated using the Beeston-Barlow lite procedure [122]. This procedure assigns one additional nuisance to each bin for each process. However, to reduce the computational complexity, as an approximation, the Beeston-Barlow lite procedure is used which assigns one additional nuisance per bin assuming them process-independent.

The contribution of this uncertainty to the log likelihood of the fit can be minimized analytically thus simplifying the minimization process of the rest of the likelihood function.

Since the Beeston-Barlow lite procedure uses Gaussian distributions to describe the statistical uncertainties of individual bins, it may not be valid for bins that are sparsely populated. Thus, bin-by-bin uncertainty calculated using the Poisson distribution was also tested. Since the impact on signal extraction was not significant, the Beeston-Barlow lite procedure was used since it was easier to implement it.

Jet energy scale The jet energy corrections (JECs) discussed in Section 3.6.5 are derived with both data-driven and simulation-based methods. Variation of these corrections within their uncertainties affects the jet momentum. These changes in normalization and shape of the jet momentum and MET are then propagated as uncertainties on the jet energy scale in the analysis. The central CMS implementation splits these uncertainties into 25 different sources which are then grouped into 12 groups (also known as the Reduced JES scheme) by taking the square root of the sum of jet uncertainties within the group in quadrature. This Reduced JES scheme was not available in the samples used in this analysis i.e. the uncertainties available in the samples used in this analysis corresponded to the 25 different sources but were not available in the grouped form of the Reduced scheme. Thus, the Reduced scheme uncertainty belonging to a particular group was calculated by summing in quadrature its associated group uncertainties. Checks were performed to ensure the calculation of the Reduced JES scheme uncertainties implemented in this analysis matched with the calculation implemented by the CMS Collaboration. The various sources of uncertainties are listed below:

- sources affecting the absolute difference in JES: These sources include the uncertainties that cover the impact of contributions from FSR and ISR, extrapolation uncertainties of high transverse momentum jets, and uncertainties related to the combined photon and Z boson decays to electron and muon pair reference scales. Additionally, the uncertainty accounts for the potential effects of pile-up.

- Differences obtained in JES when applying JECs to data and simulation with different jet flavor mixture, specifically to QCD processes.

- The variation in residual corrections derived from datasets enriched with dijet events and Z+jets events.

- Difference between the corrections obtained with the missing transverse energy projection fraction (MPF) method and the momentum balance methods.

- The uncertainties, dependent on transverse momentum (p_T) and pseudorapidity (η) combining the effects from pile-up, statistical uncertainties, and discrepancies resulting from various methods used for determining residual corrections, along with uncertainties in Jet Energy Resolution (JER). Distinct nuisance parameters are introduced for different detector regions: barrel and endcap 1 (BBEC1) with $|\eta| < 2.5$, endcap 2 (EC2) with $2.5 < |\eta| < 3$, and hadronic forward (HF) with $|\eta| > 3$.

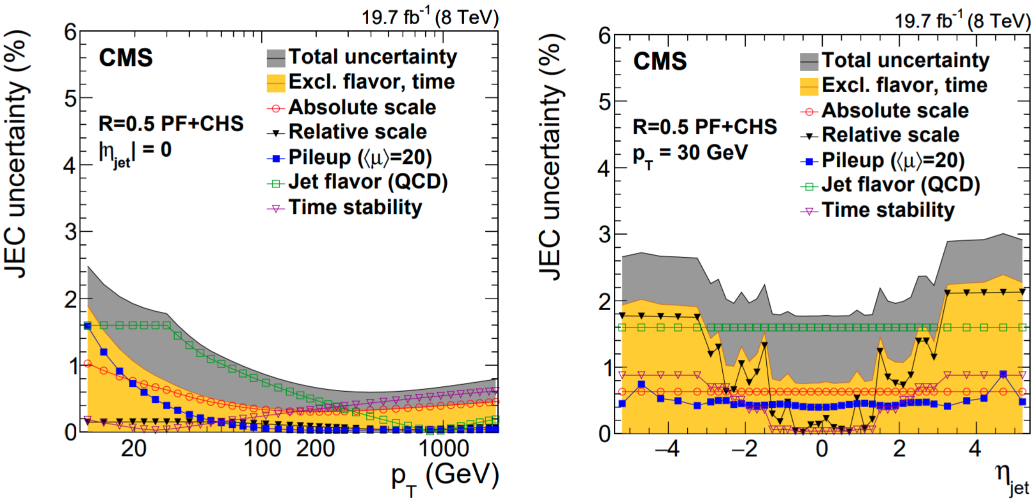


Figure 109: JES systematic uncertainties as a function of p_T and η of jet for 2016 analysis. Similar variation in uncertainty exists for 2017 and 2018 analyses [123].

Generally, the variations decrease with an increase in transverse momentum and in the central region in pseudorapidity where tracking is available (see Figure 109).

To account for similarities in the derivation of JECs across different datasets, the uncertainties pertaining to the absolute scale, BBEC1, EC2, and HF are subdivided into a collective uncertainty shared across all years and another uncertainty specific to each year. This division facilitates the partial correlation of JES uncertainties for Run 2 measurements. Due to the grouping of uncertainties, higher postfit constraints are expected (and observed) than the individual uncertainty scheme.

b-jet energy regression uncertainties The fit uncertainties on jet smearing discussed in Section 3.6.11 are propagated through the MVA to the fit model for the transverse momentum of both jet and dijet. These nuisance parameters are uncorrelated for signal and backgrounds since it was observed that the post-fit constraints on these nuisances were driven mainly by backgrounds due to their statistical dominance in CRs.

b tagging uncertainties A total of 9 sources contribute to the data/MC calibration of the DeepCSV b-tagger. They can be categorized into 4 groups:

- LF contamination in HF CR.
- HF contamination in LF CR.
- Uncertainty on the jet energy scale.
- Statistical uncertainties of udsg/c/b jets used for heavy and light flavor efficiency determination.

Since the DeepCSV calibration factors are p_T and η dependent, we decorrelate them in this analysis. We split the b-tagging nuisances into 5 bins of p_T and 3 bins of η.

This leads to a total of 135 nuisance parameters to account for the separate degrees of freedom in the fit for various p_T and η bins.

For the DeepAK8 tagger used in the boosted topology, the calibration uncertainties discussed in Section 3.6.6 for the signal are implemented as shape variation decorrelated for tagger score of 0.8-0.97 and 0.97-1.0 in 5 bins of p_T of the FatJet. The uncertainties are around 10%. For the background processes in boosted topology, the in-situ SF is measured in the fit, so no prior constraints are used (see Section 3.13.5).

3.13.5 Fit model studies

The fit model consists of a total of 8 POI corresponding to the 8 STXS bins or one POI in case of an inclusive fit. These POIs are the ratio of the observed cross section and the standard model expectation as mentioned in Equation 3.16. Along with this, we have several nuisance parameters comprised of the freely-floating scale factors (SF) (nuisances with no prior uncertainty), nuisances with log-normal prior uncertainty, and nuisances on the shape and normalization on the template of the input variable to the fit model as discussed in Section 3.13.2.

The following Section discusses how these SFs are used in the fit model.

Freely-floating SF with category migration Additional degrees of freedom are added to the fit model to ensure the total normalization of MC and data is the same within uncertainties. This approach was already used in the previous VHb$\bar{\text{b}}$ analysis [47] which was an inclusive measurement.

Extending this idea to the current STXS analysis corresponds to using 1 free floating SF per major background process for each STXS bin. However, using only the normalization factor per STXS bin in the fit model can lead to an unphysical shape for the p_T of the vector boson due to the discontinuities at the STXS bin boundaries. As an example, Figure 110 (left) shows the p_T distribution of the vector boson in TT CR for high p_T STXS bin before the CR-only fit while Figure 110 (right) shows the distribution after using 1 normalization SF per STXS bin per major process. Even though the overall normalization in MC is found to be almost correct in Figure 110 (right), the normalization per bin of the histogram is not correct. An incorrect estimate could affect the background estimate in SR since the MVA score used in SR is correlated with the p_T of the vector boson.

Such issues in the postfit p_T(V) template in TT CR can be avoided by using the distribution of a variable as an input instead of just an overall normalization, for example, in the above case, p_T distribution of the vector boson for TT CR in each STXS bin. This will also ensure continuity in the p_T(V) template at the STXS bin boundary.

Another way to address this problem is the usage of freely-floating SF and category migrations. Here, instead of using N free floating SF per process (N is the number of CR STXS bins), we use 1 SF and N-1 category migration uncertainties at each STXS bin boundary. Since only 1 SF is used per major background process, the additional category migrations per bin ensure the correct background normalization per STXS bin since the normalization of the background processes is phase-space dependent and not the same for every STXS bin. These category migration nuisance parameters

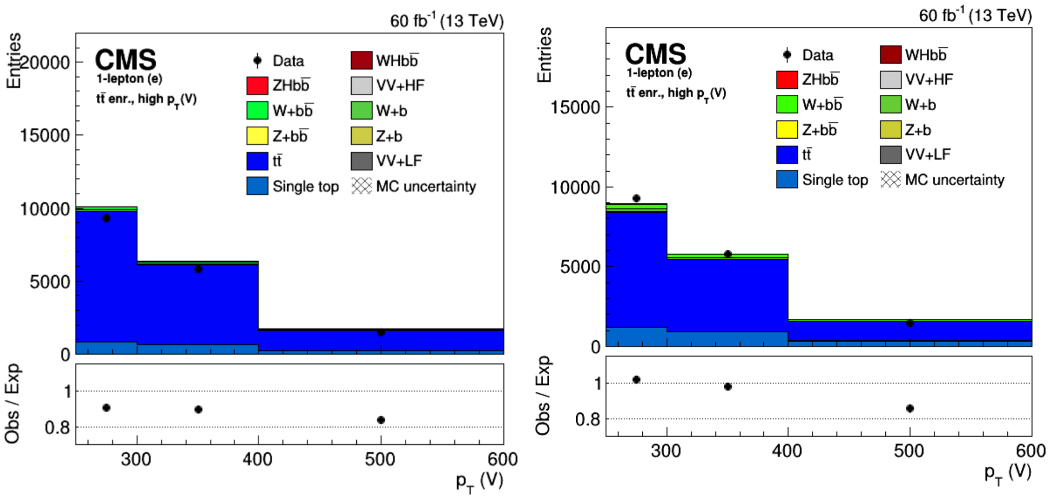


Figure 110: p_T distribution of the vector boson in TT CR for high p_T STXS bin before the CR-only fit (left),p_T distribution after using 1 normalization SF per STXS bin per major process (right) [124].

are designed to have additional constraints to ensure continuity at the bin boundary and are designed to be large enough (so as to behave similarly to a freely-floating SF) without having the possibility of going negative. Figures 111 shows the p_T of the vector boson after applying N SF per major background process in the TT CR for CR-only fit and 111 (right) after using 1 SF and N-1 migration uncertainty per major background process in the fit. Thus, the approach of 1 SF with N-1 category migration models the data better. Though this example demonstrates the use of p_T of vector boson as an input instead of a normalization factor in the fit model, the idea of 1 SF and N-1 category migration can be extended to other analysis regions. In the current VH(→ b$\bar{\text{b}}$) STXS analysis, since we have CRs in p_T(V) in range of 75−150 GeV, 150−250 GeV, and > 250 GeV (even though we have STXS bins from 250−400 GeV and > 400 GeV, as discussed in Section 3.9, the corresponding CRs are merged), we consider $N - 1 = 3 - 1 = 2$ category migrations: mig150 at p_T(V) =250 GeV boundary (only for the 2-lepton channel) and mig250 at p_T(V) >250 GeV boundary (for all the channels).

The linear migration uncertainties corresponding to the migration uncertainty at the p_T(V) =150 GeV boundary is defined as:

$$\text{mig150} = \text{event\_weight} \cdot (1.0 \pm 0.005 \cdot (\min(250.0, \max(p_T(\text{V}), 75.0)) - 150.0)) \tag{3.34}$$

thus ensuring anti-correlation and continuity at the p_T(V) =150 GeV boundary of 2-lepton channel.

The linear migration uncertainties corresponding to the migration uncertainty at the p_T(V) =250 GeV boundary is defined as:

$$\text{mig250} = \text{event\_weight} \cdot (1.0 \pm 0.005 \cdot (\min(400.0, \max(p_T(\text{V}), 150.0)) - 250.0)) \tag{3.35}$$

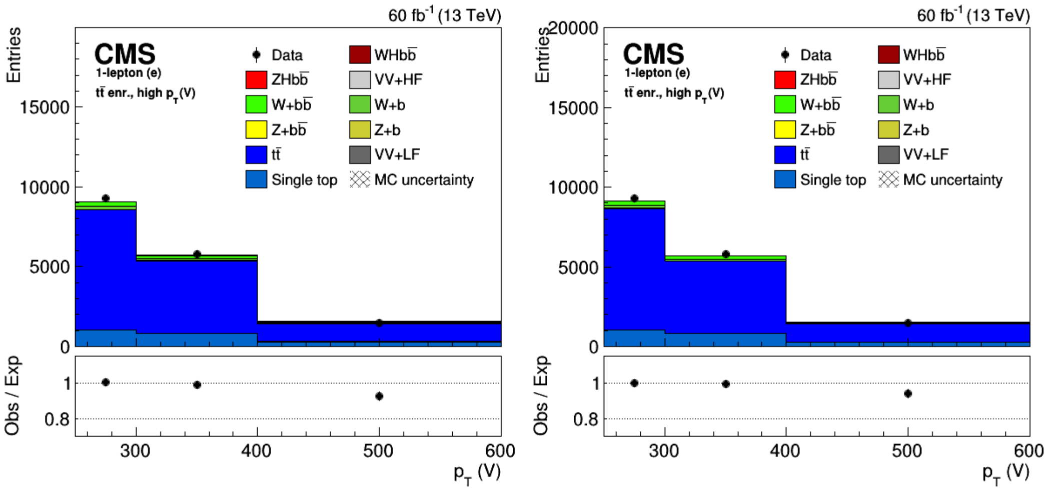


Figure 111: p_T distribution of the vector boson in TT CR for high p_T STXS bin (left Figure) after using N SF per major background process in the CR-only fit (N is the number of CR STXS bins) while the right Figure shows the distribution after using 1 SF with N-1 category migration nuisances per major process in the fit [124].

thus ensuring anti-correlation and continuity at the p_T(V) =250 GeV boundary for all channels.

These migration uncertainties are presented in Figure 112. As shown, the up and down variations intersect at the bin boundary to ensure continuity and anti-correlation among the migration on either side of the boundary.

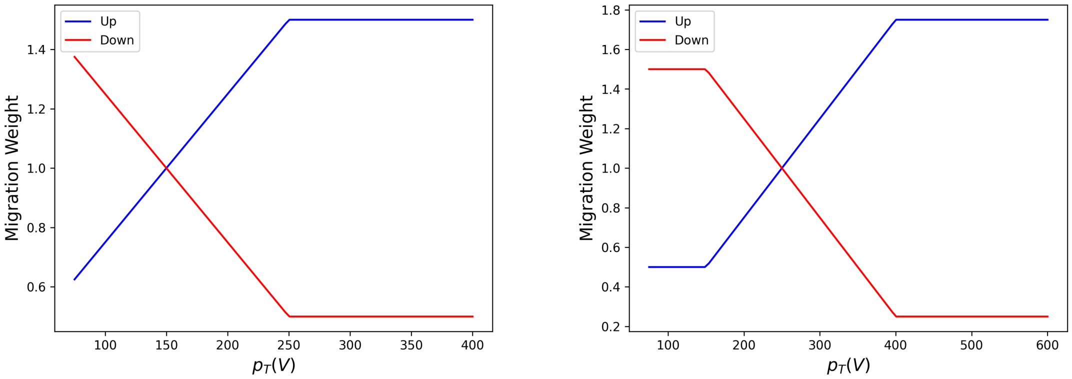


Figure 112: Migration uncertainties corresponding to Equation 3.34 and Equation 3.35

Finally, we thus use the following freely-floating process SF in the fit:

- 1 freely-floating SF (TT SF) for $t\bar{t}$ process in each of the channels.
- 1 freely-floating SF (Z+udsg SF) for Z+udsg process in 0- and 2-lepton separately for electron and muon channels.

- 1 freely-floating SF (Z+c SF) for the Z+c jets process in 0- and 2-lepton separately for electron and muon channels. They behave as in-situ SF as there are no external DeepCSV calibration SF available for Z+c jets.

- 1 freely-floating SF (Z+b SF) for the Z+b jet process in 0- and 2-lepton separately for electron and muon channels.

- 1 freely-floating SF (Z+bb SF) for the Z+bb jets process in 0- and 2-lepton separately for electron and muon channels.

- 1 freely-floating SF (W+udsg SF) for W+udsg process in 0- and 1-lepton separately for electron and muon channels.

- 1 freely-floating SF (W+c SF) for the W+c jets process in 0- and 1-lepton separately for electron and muon channels.

- 1 freely-floating SF (W+2b SF) for the W+2b jets process in 0- and 1-lepton separately for electron and muon channels. For the W+1b process, a log-normal with 30% uncertainty on the ratio of normalization of W+2b by the W+1b process is used.

The reason for having separate electron and muon SFs is discussed in Section 3.13.5. The split of SF in V+1b and V+2b processes is further discussed in Section 3.13.5.

Freely-floating object SF for b-tagging of FatJets In addition to the process SFs (described in the previous Section) which are defined for both resolved and boosted topology together, another set of SF is freely floated to account for the b-tagging efficiency measurements of the AK8 jets in background processes. This is necessary as external SFs for DeepAK8 are unavailable as discussed in Section 3.6.6.

- 1 freely-floating efficiency SF for AK8 b jets in $t\bar{t}$ process correlated across all channels.

- A freely-floating efficiency SF for AK8 light-flavored jets in V+udsg process correlated in the electron channels of 1 and 2-lepton channel. Similarly, a freely-floating efficiency SF is correlated for muon channels of 1- and 2-lepton channels. Another freely-floating efficiency SF is used for the 0-lepton channel.

- A freely-floating efficiency SF for AK8 b-jet in V+1b and V+2b processes correlated in the electron channels of 1 and 2-lepton channel. Similarly, a freely-floating efficiency SF is correlated for muon channels of 1- and 2-lepton channels. Another freely-floating efficiency SF is used for the 0-lepton channel.

- A freely-floating efficiency SF for V+c jets correlated in Boosted V+LF CR of all channels. It measures the failing efficiency of the V+c process with respect to the 0.97 DeepAK8bbVSlight threshold defining the HF and LF CR.

- A freely-floating efficiency SF for V+c jets correlated in Boosted V+HF, Boosted TT CR, and Boosted SR of all channels. It measures the passing efficiency of the V+c process with respect to the 0.97 DeepAK8bbVSlight threshold defining the HF and LF CR. The passing and failing efficiency SF for V+c processes were introduced to improve the fit stability.

Electron and muon SF The separation of electron and muon SF in the 2-lepton and the 1-lepton channel is necessary due to their different prefit mis-modelling. Figure 113 shows the transverse mass distribution for the muon (left) and electron (right) channels in HF CR of the 1-lepton channel. As shown, the modelling of the data at prefit level is different. Given the negligible contribution of QCD in the analysis, the QCD background templates are not included in the fit model. Still tests were performed to ensure the QCD is not the cause of mis-modelling in the muon CR. For example, the mis-modelled regions of the transverse mass of the muon CRs were excluded by additional selection on the transverse mass. However, the SFs were found to be compatible with those obtained without the additional cuts within uncertainty. Further tests were performed by using tighter thresholds on the lepton isolation but no improvement was observed. Independent SFs for electron and muon help to mitigate the effects of different prefit data/MC modelling and has also been found to have a significant impact on Goodness of the fit [125] of the model discussed in Section 3.14.1.

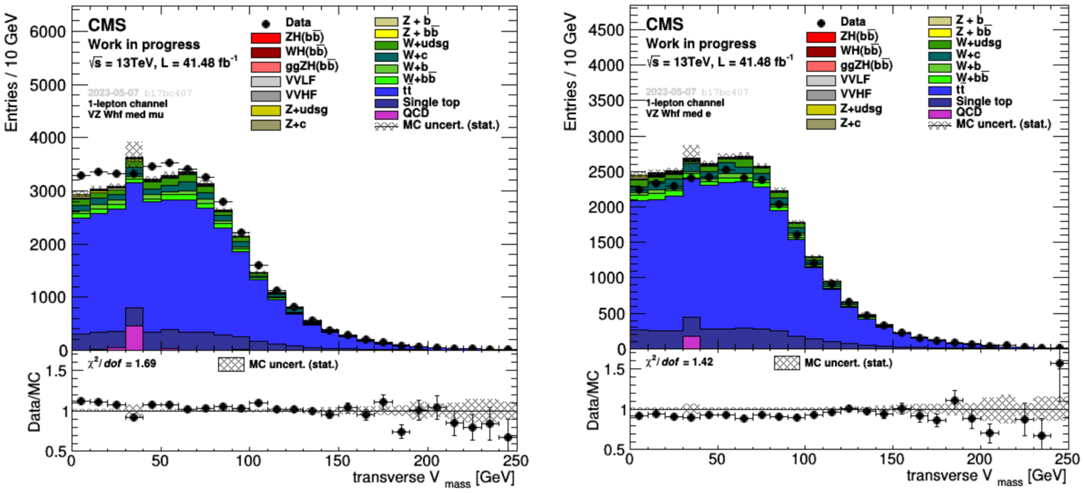


Figure 113: Transverse mass distribution for the muon (left) and electron (right) channels in HF CR of the 1-lepton channel.

V+1b and V+2b processes in the fit model The classification of an event as V+1b or V+2b depends on the total number of B hadrons (which originates either from matrix element or parton shower) in a V+jets event within the detector acceptance region. Figure 114 shows the distribution of different flavors of V+jets in the SR of 2-lepton (left) and 1-lepton channel (right). The kinematics of V+1b and

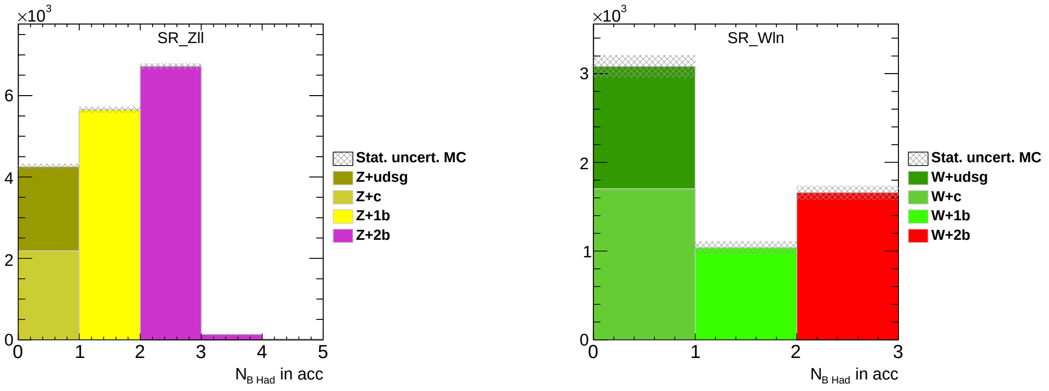


Figure 114: Distribution of different jet flavors of V+jets in SR of 2-lepton (left) and 1-lepton channel (right) [126].

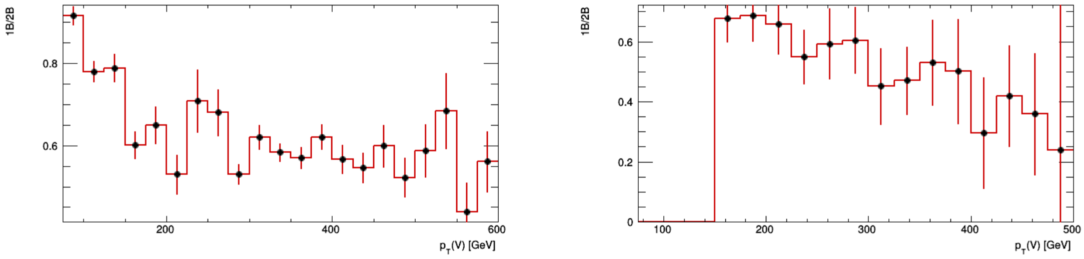


Figure 115: Ratio of p_T distribution of the vector boson for V+1b and V+2b events in the 2-lepton channel (left) and 1-lepton channel (right) [126].

V+2b events is different. For example, the ratio of p_T distribution of the vector boson for V+1b and V+2b events in the 2-lepton channel (left) and 1-lepton channel (right) is shown in Figure 115. Similarly, Figure 116 shows the ratio of p_T distribution of the leading (left) and subleading (right) jet in 2-lepton HF CR for V+1b and V+2b events. These plots show p_T dependence of the ratio. Thus, separate freely-floating SF for V+1b and V+2b processes were tested in the fit model along with the associated p_T(V) category migration nuisances. However, due to the Cabibbo suppressed

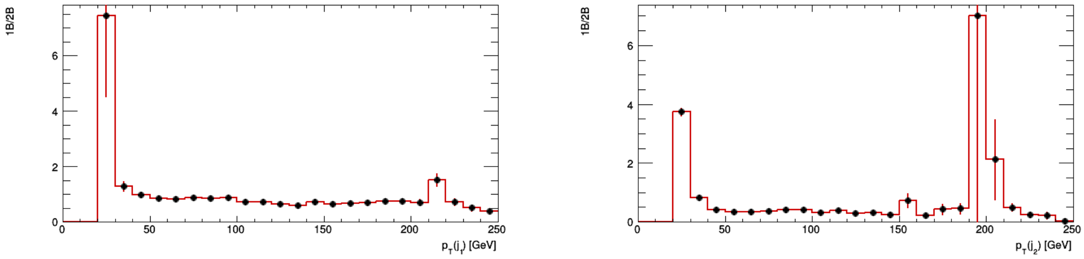


Figure 116: Ratio of p_T distribution of the leading (left) and subleading (right) jet in 2-lepton HF CR for V+1b and V+2b events [126].

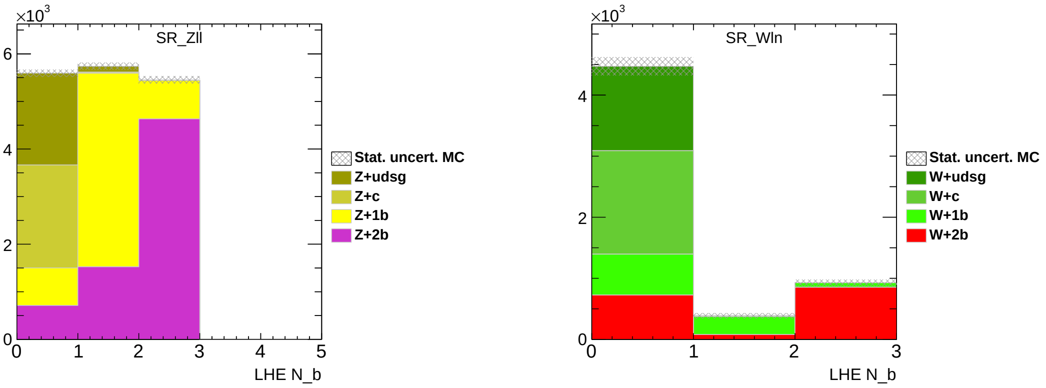


Figure 117: Distribution of the number of matrix element b-quarks of V+jets events in the SR of the 2-lepton channel (left) and 1-lepton channel (right) [126].

(see Figure 118) matrix element of W+1b events compared to Z+1b (see Figure 117), the total number of W+1b events is significantly reduced compared to W+2b events.

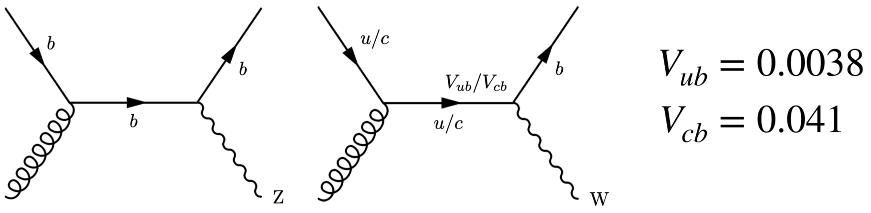


Figure 118: LO feynman diagram for Z+1b (left) and W+1b (right) events.

This W+1b suppression in turn leads to weak constraints on the W+1b SF. Further, due to the anti-correlation of V+1b SF with V+2b SF, the W+1b SF having larger uncertainty gets pushed to unphysical values. This issue is not observed with Z+1b and Z+2b SF since the total number of Z+1b and Z+2b events in the 2-lepton channel is comparable. Thus, instead of freely-floating W+1b and W+2b SF, a log-normal prior uncertainty of 30% on the ratio of W+1b and W+2b normalization (with W+2b SF freely-floated) is used. Studies were performed to ensure the choice of prior does not significantly impact the signal strength extraction.

3.13.6 Binning of input variables to the fit model

Binning of CRs variables As discussed in Section 3.13.1, variables like p_T(V) and the DeepAK8bbVSlight score are used as input to the fit model in CRs. For the HF CR of 0- and 1-lepton channels, a 6-bin multiclass classifier (as discussed in Section 3.11) is used to classify V+udsg, V+c, V+1b, V+2b, single top (ST), and $t\bar{t}$ processes where the contamination of $t\bar{t}$ is very high. For the 2-lepton HF CR, we use the DeepCSV score of leading and subleading jet binned in 5 bins as discussed in Section 3.13.1 (see Figure 105). These bins are designed to have maximal separation between different flavors of V+jets events. The higher score bins have more V+HF jets while the lower score bins have more V+LF jets. For resolved TT and LF CR, we use the p_T distribution of the vector boson binned coarsely. Finer binning improves the shape information of the histogram, but it exposes possible mis-modelling, resulting in poor goodness of fit. Thus, we had to find a balance between the goodness of fit and the additional information added to the fit model by finer binning. A similar approach is used for the DeepAK8bbVSlight output score templates in HF and LF Boosted CR. These DeepAK8bbVSlight input templates are binned at [0, 0.8, 0.97, 1.0] as discussed in Section 3.6.6.

Binning of SRs variables The binning of MVA variables is important for the analysis sensitivity. The signal event binning was performed transforming the signal to a flat distribution, leading to 1 signal event/bin. In some low p_T signal regions (such as in the 2-lepton channel) with a large number of signal events, this procedure would lead to a too large number of bins, leading to unphysical constraints on the JEC/JES uncertainties in the fit caused by the shape fluctuations. By limiting the number of bins to 15 with a flat signal distribution, the over-constraints on JEC/JES were found to be mitigated. On the contrary, some of the high p_T regions have a limited number of signal events. Thus to avoid the collapse of the MVA variable shape into too few bins, we decided to keep a minimum of 5 bins. This procedure results in a flat distribution of signal events with no more than a total of 15 bins in any given MVA template and no less than 5 bins. This binning approach for the signal templates was used in all three years.

For the 2017 and 2018 datasets, since we use the NLO V+jets sample which can have negative generator weights, we choose an additional threshold based on the total weight to ensure a positive number of NLO V+jets events in each bin. The 2016 analysis is not affected by this issue since we use LO V+jets sample and thus the generator weights are positive by construction. Table 20 and 21 give the Asimov systematic and statistical uncertainties in STXS bins for the V+jets events threshold of at least 1, 3 and 5 weighted events per bin for the 2018 dataset. The uncertainties are stable across all STXS bins for 1 and 3 weighted events. We thus choose a threshold of at least 3 weighted V+jets events per bin. Scaling this threshold by the luminosity difference between 2017 VS 2018, we use a threshold of 2 weighted events per bin for 2017.

Further, it was observed that the criteria of 2 weighted events per bin led to the merging of all the bins for some high p_T STXS bin templates. Figure 119 (left) shows one such case of high1 SR of the electron channel of the 2017 2-lepton channel. The LO in the plot refers to the default binning choice used with LO V+jets samples. The choice of 2 weighted events was too conservative. Thus, for high p_T templates, we relaxed the threshold to 1 weighted V+jets event per bin. Figure 119 (right) compares the NLO template binning after applying the threshold of 1 weighted V+jets event per bin, with that of the LO template along with their corresponding MC statistical uncertainties. For 2018, we again perform an approximate luminosity scaling to ensure at least 2 weighted events per bin for high p_T STXS templates.

The shape of the systematic uncertainties related to the jet energy scale, res-

	#1	#3	#5
ZH (> 400)	0.60	0.72	0.85
ZH (250-400)	0.37	0.40	0.43
ZH (150-250, ge1j)	2.18	2.18	2.17
ZH (150-250, 0j)	0.91	0.92	0.92
ZH (75-150)	2.23	2.23	2.26
WH (> 400)	1.05	1.10	1.19
WH (250-400)	0.91	0.92	1.19
WH (150-250)	1.05	1.06	1.05

Table 20: Asimov systematic uncertainties in the STXS bins for the V+jets events threshold of at least 1,3 and 5 weighted events per bin for the 2018 dataset.

	#1	#3	#5
ZH (> 400)	0.93	1.00	1.09
ZH (250-400)	0.50	0.52	0.53
ZH (150-250, ge1j)	1.11	1.11	1.11
ZH (150-250, 0j)	0.61	0.61	0.61
ZH (75-150)	0.74	0.74	0.76
WH (> 400)	1.08	1.10	1.17
WH (250-400)	0.63	0.63	0.69
WH (150-250)	0.66	0.66	0.66

Table 21: Asimov statistical uncertainties in the STXS bins for the V+jets events threshold of at least 1, 3, and 5 weighted events per bin for the 2018 dataset.

olution, or b-tagging is affected by large statistical fluctuations. These systematic uncertainties are derived from MC event migrations (for example, a shift in p_T in the case of JES). Similar shape variations were observed also for the pile-up weight. Figure 120 (left) shows the up (blue) and down (green) variations of pile-up weight systematic in the DNN distributions for $t\bar{t}$ process in 1-lepton SR. Such fluctuations lead to unphysical asymmetric impacts on the signal strength.

To address this issue, we apply a smoothing procedure and force symmetry on up and down variations of these systematics. Figure 121 shows two types of smoothing functions we tested, namely, median smoothing (Smooth(X)) and LOWESS smoothing [128]. The Smooth(X) label stands for median smoothing with the number of neighboring bins used in smoothing denoted by X. The other smoothing procedure is the LOWESS (Locally Weighted Scatterplot Smoothing). The impact on inclusive signal strength was found to be similar for Smooth with X=2,3,4 while for X=10 and LOWESS, it was different and large. Thus, Smooth(4) was applied to all templates describing the systematics uncertainty related to JEC, JER, b-tagging, and Pile-up. Figure 120 (right) shows up (blue) and down (green) variations of pile-up weight systematic for $t\bar{t}$ process after application of smoothing and ensuring both up and down variations symmetric.

This procedure helps to remove the spurious impacts of these nuisance parameters on the signal strength. Figure 122 shows how the spurious asymmetric im-

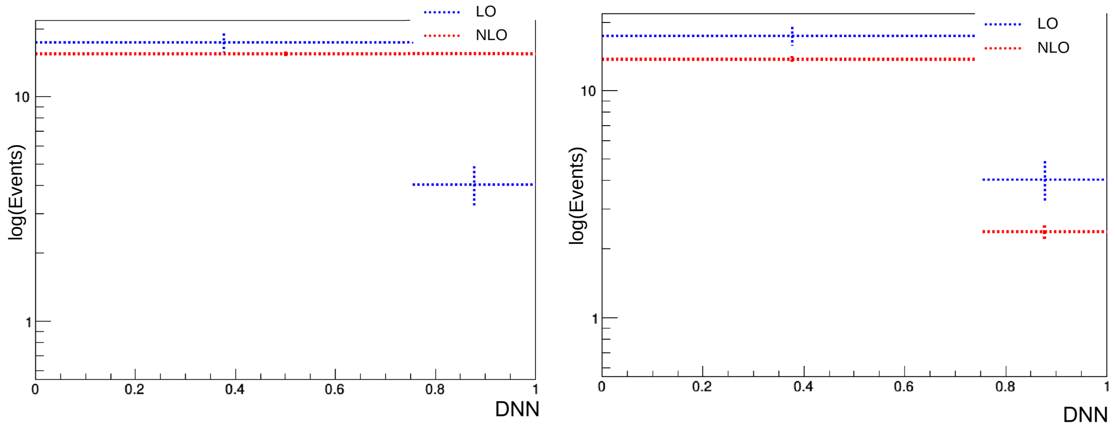


Figure 119: DNN template of V+jets events in high1 SR of the 2-lepton muon channel in 2017. LO refers to the template corresponding to the LO V+jets samples and NLO refers to the template corresponding to the NLO V+jets samples. The NLO template in the left plot has at least two weighted events per bin while the right plot has at least one weighted event per bin. In both plots, the LO templates are binned only according to the signal region binning criteria discussed in Section 3.13.6.

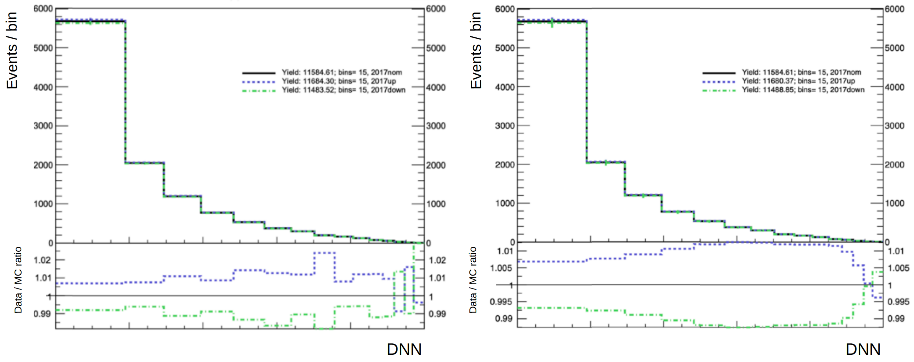


Figure 120: DNN distributions showing the up (blue) and down (green) variations of pile-up weight systematic for $t\bar{t}$ process before (left) and after application of smoothing and making both up and down variations symmetric as described in Section 3.13.6 [127].

pacts of nuisances related to the Jet energy scale and quark flavor/gluon response of Herwig/Pythia (QCD Flavor) on signal strength are removed after smoothing and symmetrization of the up and down variations.

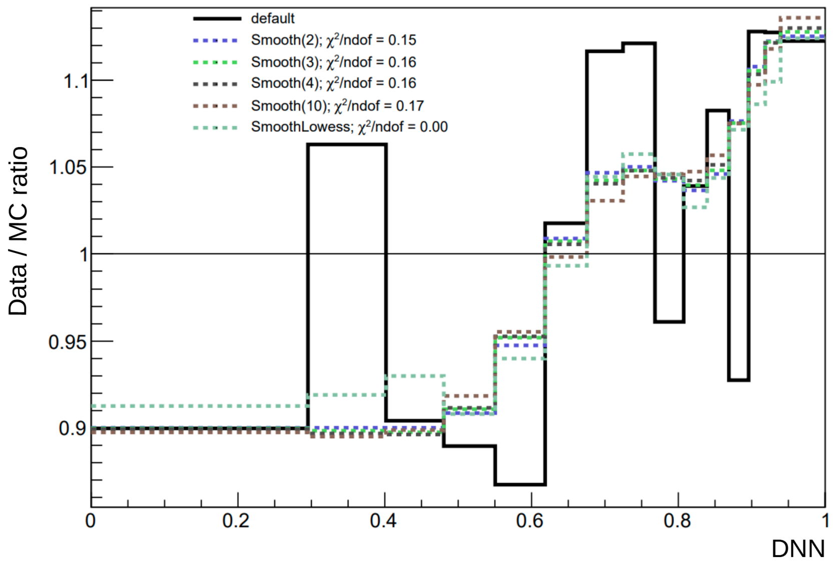


Figure 121: DNN distribution showing the effect of median smoothing (Smooth(X) where X refers to the number of bins used in the smoothing) and LOWESS smoothing on default shape (without smoothing) for the Up variation of the ggZH signal process in DNN for the 0-lepton channel [127].

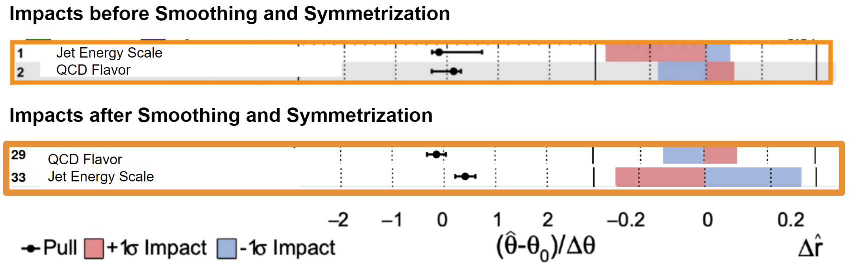


Figure 122: Impact of nuisance parameters related to the Jet energy scale and quark flavor/gluon response of Herwig/Pythia (QCD Flavor) on signal strength before (top) and after (bottom) smoothing and symmetrization of the up and down variations [127].

3.14 Results

3.14.1 Goodness of fit tests of the model

In order to quantitatively test the robustness of the fit model, goodness of fit (GoF) tests are performed. In general, the strength of a GoF test depends on the alternative hypothesis. Thus there is no single best choice of GoF test. In this analysis, we perform two GoF tests to access our fit model described in Section 3.13.

The first GoF test is the saturated model GoF test. This test is based on the likelihood ratio test and thus the test statistic asymptotically follows a χ^2 distribution and can be interpreted as a generalization of the χ^2 GoF test for binned data [125]. The numerator is the likelihood of the fit of the model to data while the denominator is the likelihood of fitting a saturated model (i.e. a model that fits the data perfectly because it has as many estimated parameters as there are values to be fitted) to data. The result of the saturated GoF test using the fit model and 2017 and 2018 datasets is given in Figure 123. The p-value giving the compatibility of the observed result and the fit model is 0.84 and 0.48 respectively. p-values close to 0.5 are considered to be ideal, and thus we find the current fit model describing data satisfactorily. The major improvement to the fit model was found to come from the split of the SFs in the lepton flavor described in Section 3.13.5.

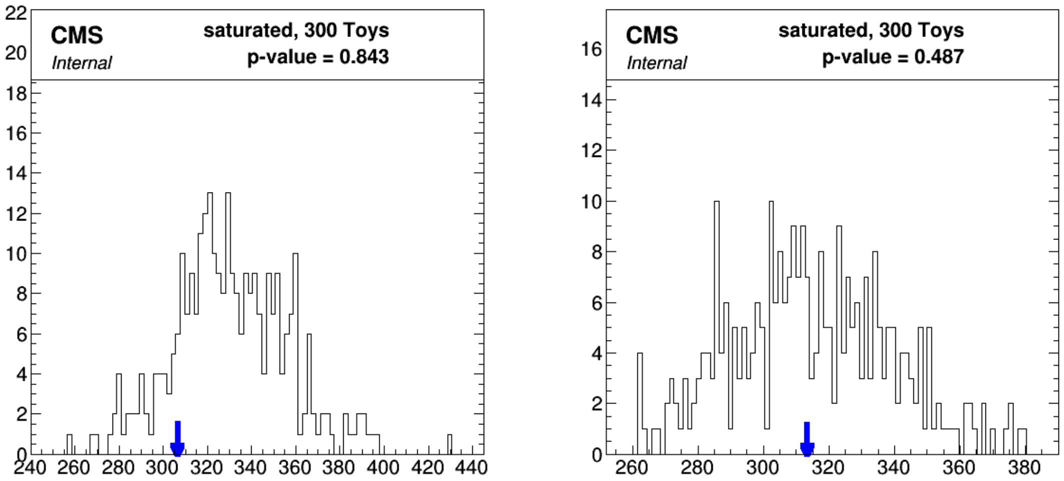


Figure 123: Distribution of the saturated GoF test statistic of the model (obtained through the fit of 300 toys on data) for the 2017 analysis (left) and 2018 analysis (right). The value of the test statistic obtained through the fit of the MC simulation model to data is given by the blue arrow.

Along with the saturated GoF test, the Kolmogorov-Smirnov (KS) test [129] was also performed. In this test, the cumulative distributions of the histogram of data and postfit MC are compared. In order to have a reasonable outcome of the KS test, a sufficient number of histogram bins are required. Thus, this test is performed only for specific regions where the number of bins is large enough i.e. in low and med p_T(V) STXS bins, and the p-values were again found to be satisfactory.

3.14.2 VZ(bb) cross check analysis

In order to validate the analysis strategy, we perform a cross check analysis targeting the associated production VZ($\rightarrow b\bar{b}$) where the vector boson decays leptonically and the Z boson decays to a pair of bottom quarks. The VZ($\rightarrow b\bar{b}$) has a similar final state as VH($\rightarrow b\bar{b}$) but the two processes differ in the following points:

- The Z boson mass is measured to be 91.2 GeV while the Higgs boson mass is measured to be 125.25 GeV.

- The production cross section of VZ(→ $b\bar{b}$) multiplied by the branching ratio of V(→ leptons) and Z(→ $b\bar{b}$) is about 30 times larger than the cross section of VH(→ $b\bar{b}$) multiplied by the corresponding branching ratio.
- The Z boson is a spin-1 particle while the Higgs boson has spin-0.

Given that the mass of the Z boson is different from the H boson, we define the SR as the dijet mass window as 60-120 GeV and corresponding orthogonal cuts on dijet mass to define the HF CR. Due to changes in HF and SR definition along with different signal processes, both DNNs and BDTs used in the fit model are re-trained using the VZ(→ $b\bar{b}$) process as the signal. Furthermore, unlike the Higgs boson analysis, VZ(→ $b\bar{b}$) is not split in generator-level STXS bins.

The overall signal strength for VZ(→ $b\bar{b}$) analysis for full Run 2 data was measured to be

$$\mu = 1.25 \pm 0.14. \tag{3.36}$$

The p-value of the likelihood ratio test of this measurement with the SM expectation was found to be 6%. The expected and observed significance were found to be over 5σ. The signal strength for WZ(→ $b\bar{b}$) and ZZ(→ $b\bar{b}$) using full Run 2 data is shown in Figure 124

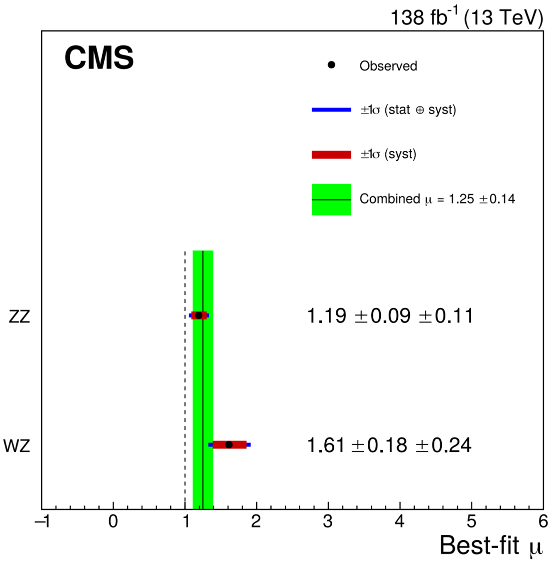


Figure 124: The signal strength for WZ(→ $b\bar{b}$) and ZZ(→ $b\bar{b}$) with the associated vector boson decaying leptonically using Run 2 data.

3.14.3 STXS Run 2 results

The inclusive and STXS POIs were extracted by a simultaneous likelihood fit on the full Run 2 (2016-2018) data using the fit model described in Section 3.13, and are given in Table 22 and Figure 125 respectively. The overall inclusive signal strength for the Run 2 dataset is

$$\mu = 1.15 \pm 0.14(\text{stat.})^{+0.16}_{-0.15}(\text{syst.}) = 1.15^{+0.22}_{-0.20}\ (\text{tot.}) \tag{3.37}$$

Year	Inclusive μ
2016	1.43 ± 0.37
2017	0.68 ± 0.36
2018	1.23 ± 0.30

Table 22: Inclusive signal strength μ extracted for individual years of Run 2 in a combined Run 2 fit.

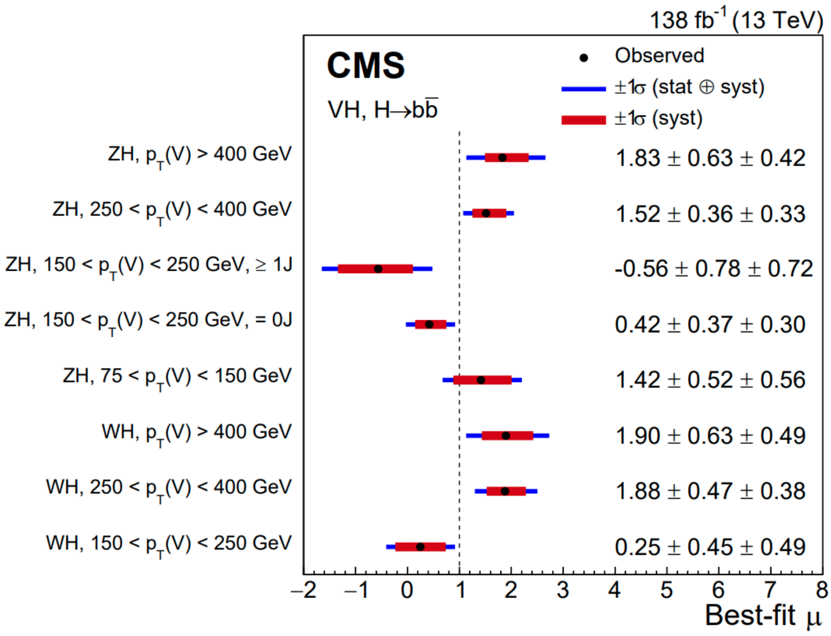


Figure 125: STXS signal strengths for Run 2 data (2016-2018). The dashed line at 1.0 corresponds to the standard model expected value of signal strength. The first and the second uncertainty of the signal strength values correspond to the statistical and systematic uncertainties respectively.

The observed (expected) VH signal significance is 6.3(5.6) σ. This significance leads to the observation of the VH(→ b$\bar{\text{b}}$) process at CMS with Run 2 data. Figure 126 shows signal (under SM hypothesis), postfit background, and data yields for all three years sorted by the logarithm of the signal-to-background ratio. The bins with a high data-to-background ratio indicate an excess in data in agreement with the signal+background hypothesis over the background-only hypothesis.

Table 23 shows the absolute impacts $\Delta\mu$ of the individual uncertainty groups on the inclusive signal strength measurement in the Run 2 fit. The largest impact on the inclusive signal strength measurement is driven by the limited size of MC simulation samples followed by signal theory uncertainties (discussed in Section 3.13.3) and V+jets modelling uncertainties ('simulation modelling' in Table) discussed in Section 3.13.3.

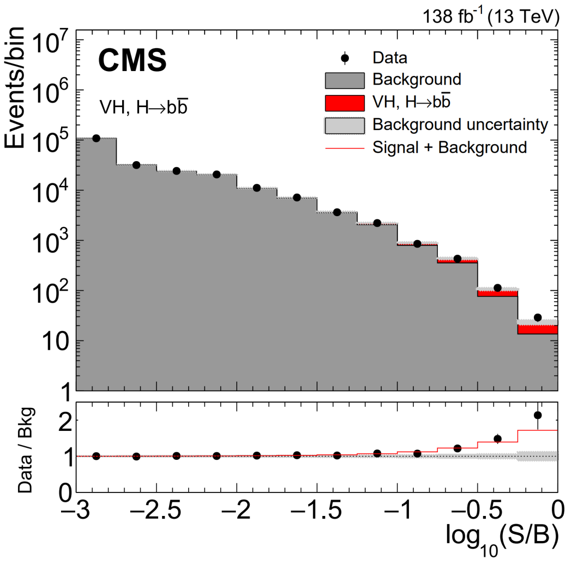


Figure 126: Signal (under SM hypothesis), postfit background, and data yields for all three years sorted by the logarithm of the signal-to-background. The red histogram in the top plot corresponds to the signal under SM expectation, the grey histogram corresponds to the observed background and the hashed region corresponds to background uncertainty. In the ratio plot, the red histogram corresponds to the expected signal + observed background/observed background while the black marker dots correspond to data/observed background.

Uncertainty group	$\Delta\mu$
background (theory)	+0.043 -0.043
signal (theory)	+0.088 -0.059
MC sample size	+0.078 -0.078
simulation modelling	+0.059 -0.059
b tagging	+0.050 -0.046
JER	+0.036 -0.028
integrated luminosity	+0.032 -0.027
JES	+0.025 -0.025
lepton identification	+0.008 -0.007
trigger (0-lepton)	+0.002 -0.001

Table 23: Absolute impacts $\Delta\mu$ of the individual uncertainty groups on the inclusive signal strength measurement in the Run 2 fit.

Figure 127 shows the signal strengths measured per channel (left) and Higgs boson production mode (right) for full Run 2 fit. The p-value for compatibility of the per channel (per production mode) with SM expectation is 64% (34%). The p-value for compatibility of the per channel (per production mode) with inclusive measurement is 84% (56%).

The observed signal strength is also interpreted in terms of cross section times

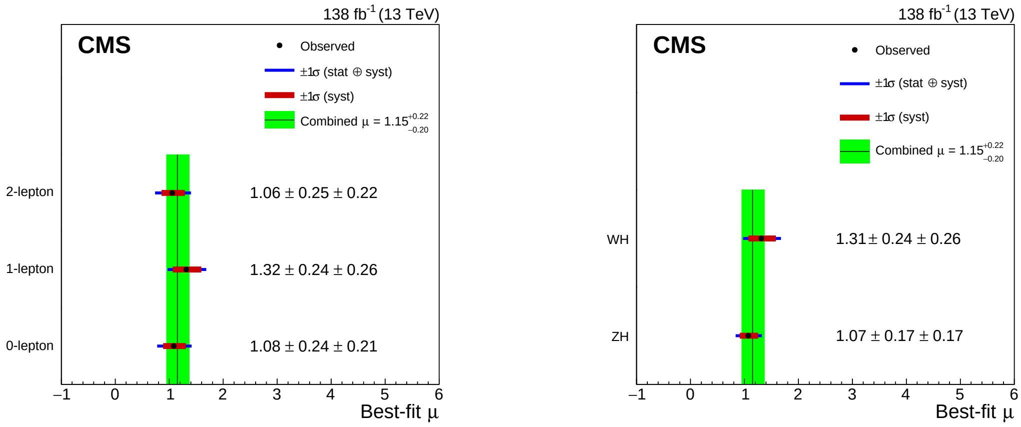


Figure 127: Signal strengths measured per channel (left) and Higgs boson production mode (right) for Run 2 fit

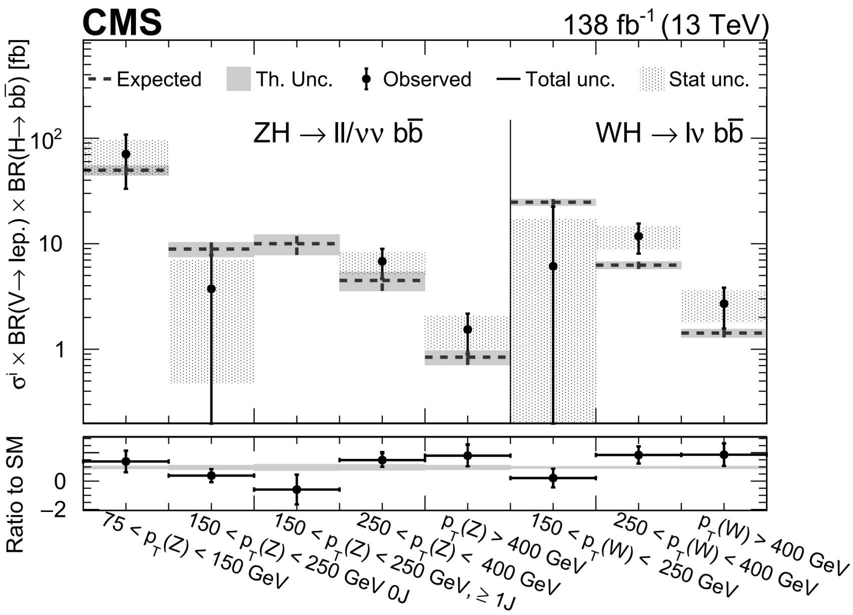


Figure 128: The product of observed cross section (σ) and branching fraction of V → leptons and H → bb. Here, only STXS bins with cross sections having positive observed signal strengths have been shown.

branching ratios as is shown in Figure 128. In this Figure, only cross sections for positive signal strengths have been shown.

The postfit correlation for the different signal strength POIs of the STXS measurement is shown in Figure 129. The largest correlation of 21% exists for the ZH production mode in 150-250 p_T(V) range between 0 and > 0 additional jet bins. This is expected since the two STXS bins cover the entire 150-250 p_T(V) range simultaneously.

The process and in-situ efficiency SFs discussed in Section 3.13.5 corresponding

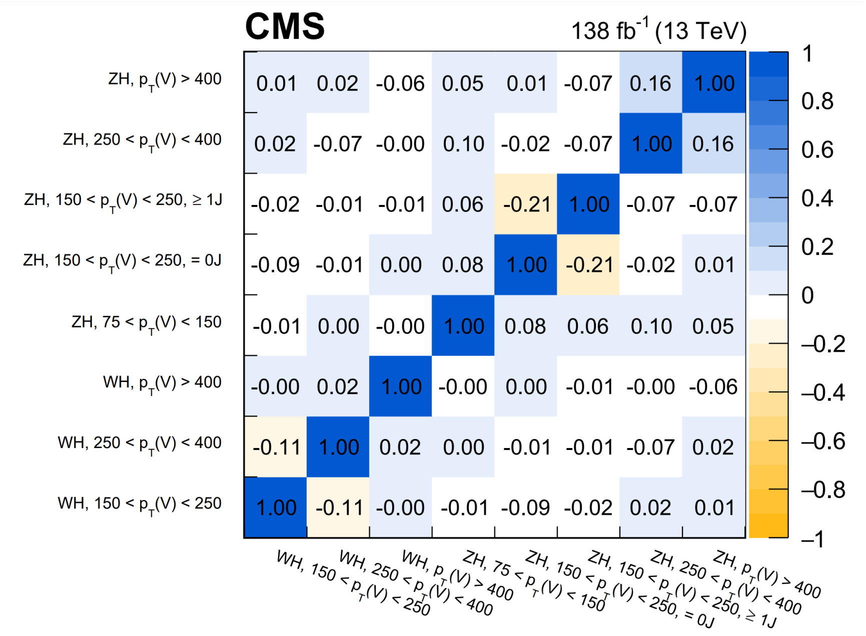


Figure 129: Postfit correlation for the different signal strength POI of the STXS measurement.

to 2017 data extracted from the full Run 2 STXS fit are given in Tables 24, 25 and 26. Most of the process SFs are close to unity within one standard deviation as expected. The process SFs for V+jets background in muon channels are larger than the ones in electron channels, mainly in the 1-lepton channel, due to data/MC mis-modelling discussed in Section 3.13.5. The in-situ efficiency SF has some deviations from one along with large uncertainty. However, since they measure the efficiency of DeepAK8bbVSlight (since external efficiency SFs were not available for DeepAAK8 unlike the DeepCSV scores), they are not in general expected to have a fixed value close to 1.0 (i.e. efficiency of data and MC is not expected to match prior to a prefit calibration).

Process	Z($\nu\nu$)	W(eν)	W($\mu\nu$)	Z(ee)	Z($\mu\mu$)
t$\bar{\text{t}}$	1.08 ± 0.08	0.89 ± 0.07	0.97 ± 0.08	0.84 ± 0.11	0.98 ± 0.15
V+usdg	0.69 ± 0.17	0.89 ± 0.09	0.91 ± 0.10	0.98 ± 0.07	0.95 ± 0.07
V+c	1.39 ± 0.39	1.10 ± 0.24	1.13 ± 0.29	0.95 ± 0.34	1.12 ± 0.33
V+1b	1.52 ± 0.39	1.12 ± 0.26	1.32 ± 0.26	1.64 ± 0.38	1.48 ± 0.53
V+2b	1.20 ± 0.14	1.06 ± 0.20	1.46 ± 0.24	0.84 ± 0.20	0.78 ± 0.11

Table 24: Background process SF (described in Section 3.13.5) for 2017 obtained using STXS-based fit to full Run 2 data. The values reported for the W+1b process are not the process SF but the pull and constraint of a log normal prior on the ratio of the W+1b to W+2b backgrounds.

Figure 130 lists the top 15 most impactful nuisance parameters in an inclusive signal strength extraction fit to full Run 2 data. The impact of a nuisance parameter

Process	Z(ee) and W(eν)	Z($\mu\mu$) and W($\mu\nu$)	Z($\nu\nu$)
V+usdg	1.09 ± 0.08	1.08 ± 0.07	1.27 ± 0.13
V+1b/2b	0.67 ± 0.14	0.68 ± 0.14	0.84 ± 0.10

Table 25: Background efficiency in-situ SF (described in Section 3.13.5) for the 2017 analysis obtained using STXS fit to full Run 2 data.

Process	all channels
$t\bar{t}$	0.87 ± 0.02
V+c (pass)	1.72 ± 0.39
V+c (fail)	1.47 ± 0.31

Table 26: Background efficiency in-situ SF (described in Section 3.13.5) for the 2017 analysis obtained using STXS fit to full Run 2 data.

is computed by extracting the signal strength by varying the nuisance parameter by $\pm 1\sigma$ where σ refers to the observed/postfit constraint on that nuisance keeping all other nuisances frozen. As observed, the theoretical uncertainties due to QCD scale variations on signal gg→ZH discussed in Section 3.13.3 are the most impactful. The 'Pull' label in the Figure refers to the pull taking into account the prefit constraint on the nuisance parameter. For a given nuisance parameter, θ, with prefit(postfit) value $\theta_1(\hat{\theta})$ and prefit(postfit) constraint $\sigma_1(\sigma)$ is defined as:

$$\text{pull}(\theta) = \frac{\hat{\theta} - \theta_1}{\sqrt{\sigma_1^2 - \sigma^2}} \tag{3.38}$$

The pull defined in Equation 3.38 is denoted by blue cross in the Figure 130. However, if the fit model is not sensitive to a given nuisance parameter, the fit does not improve the measurement of that nuisance parameter. In that case, the prefit measurement of the nuisance parameter is equal to the postfit measurement of the nuisance parameter, $\theta_1 \pm \sigma_1 = \hat{\theta} \pm \sigma$, and the Equation 3.38 is not defined. In those cases, we rely on the other definition of pull which takes only the prefit constraint into account:

$$\text{pull}(\theta) = \frac{\hat{\theta} - \theta_1}{\sigma_1} \tag{3.39}$$

The pull defined in the Equation 3.39 is labelled as 'Fit' in the Figure 130.

The nuisance parameter associated with the pile-up weight (see Figure 130) is pulled to almost 2σ as expected from the discussion in Section 3.10.1 to account for the prefit mis-modelling. Similar trends in nuisances are observed in the STXS POI extraction, except for nuisance parameters associated with the number of additional jet corrections (discussed in Section 3.10.1). These nuisance parameters were found to be impactful for STXS POI ZH 150-250 0 jet and > 0 jet bins.

Figure 131 presents the most constrained nuisance parameters in an inclusive fit to the Run 2 dataset. Due to large linear priors on V+jets migration nuisances (discussed in Section 3.13.5) approximating a flat prior constrain, overconstraining of these nuisance parameters is expected. They get constrained even more significantly for processes like $t\bar{t}$ and V+udsg jets which have large number of events in their respective CR.

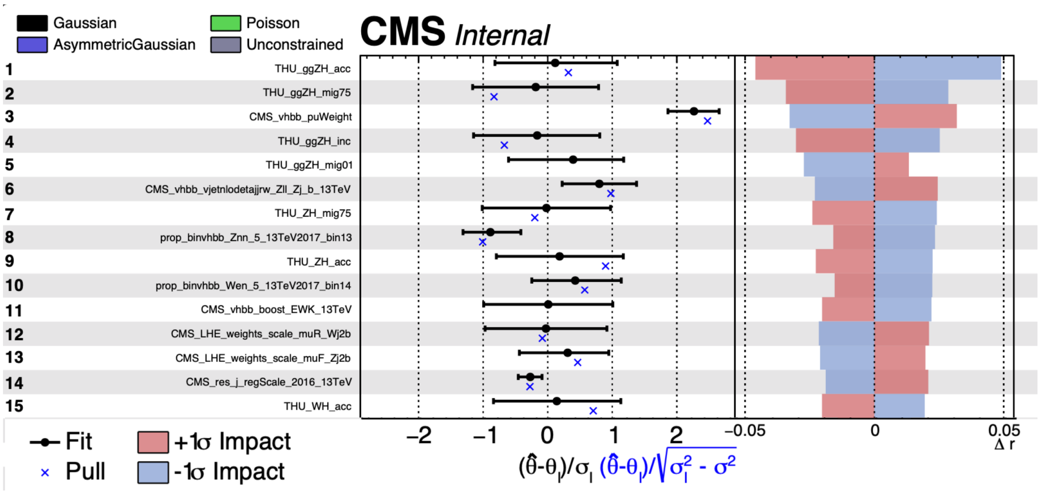
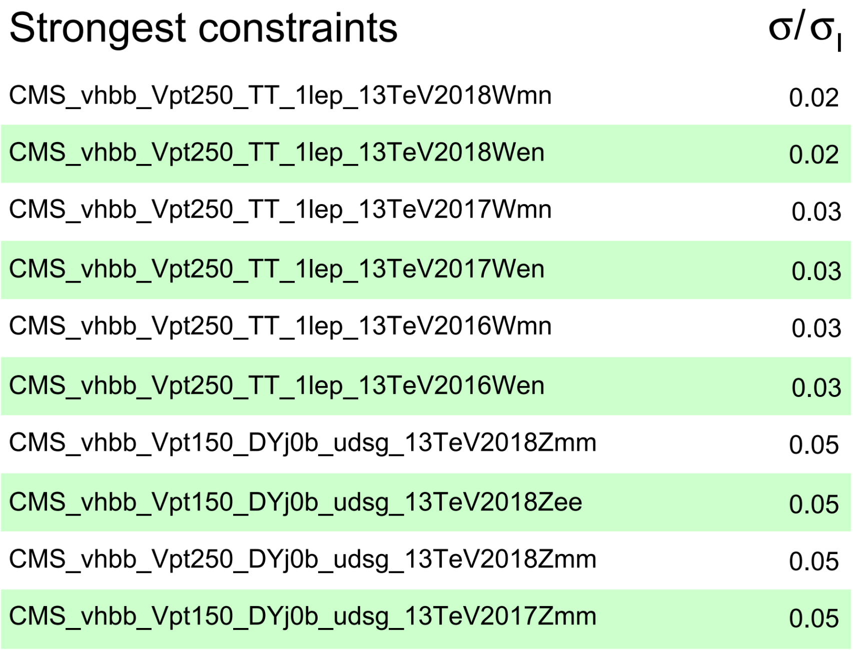


Figure 130: Top 15 most impactful nuisance parameters in an inclusive fit to full Run 2 dataset. $\sigma(\sigma_1)$ refers to postfit(prefit) constraint. $\hat{\theta}(\theta_1)$ refers to postfit(prefit) central value of the nuisance parameter. Δr refers to the change in the central value of extracted signal strength on changing the value of a given nuisance parameter within $\pm\sigma$ of its postfit constraints.



Strongest constraints	σ/σ_I
CMS_vhbb_Vpt250_TT_1lep_13TeV2018Wmn	0.02
CMS_vhbb_Vpt250_TT_1lep_13TeV2018Wen	0.02
CMS_vhbb_Vpt250_TT_1lep_13TeV2017Wmn	0.03
CMS_vhbb_Vpt250_TT_1lep_13TeV2017Wen	0.03
CMS_vhbb_Vpt250_TT_1lep_13TeV2016Wmn	0.03
CMS_vhbb_Vpt250_TT_1lep_13TeV2016Wen	0.03
CMS_vhbb_Vpt150_DYj0b_udsg_13TeV2018Zmm	0.05
CMS_vhbb_Vpt150_DYj0b_udsg_13TeV2018Zee	0.05
CMS_vhbb_Vpt250_DYj0b_udsg_13TeV2018Zmm	0.05
CMS_vhbb_Vpt150_DYj0b_udsg_13TeV2017Zmm	0.05

Figure 131: Most constrained nuisance parameters in an inclusive fit to full Run 2 dataset. Here 'VptX' with X=250,150 refers to migration nuisances (described in Section 3.13.5) at p_T(V) bin boundary of 250, 150 respectively. 'TT' and 'DY0b_udsg' refers to the $t\bar{t}$ and DY+udsg jets background. $\sigma(\sigma_1)$ refers to the postfit (prefit) constraint.

3.15 Comparison of current 2017 measurement with previous 2017 measurement

Since there were changes in the analysis strategy used for the previous measurements [130] and the current analysis strategy, several checks were performed to ensure the robustness of the analysis. A selection of these checks are discussed below:

- Analysis Categorization: The current analysis follows the STXS categorization discussed in Section 3.9 while the previous analysis [130] followed an inclusive categorization (i.e. no split based on reconstruction level variables like in the STXS framework) except for the 2-lepton channel where the events were categorized in two bin: $75 < p_T^V < 150$ and $p_T^V > 250$ with only resolved topology. In order to check the robustness of the current measurement against changes in the analysis categorization, events in the present analysis were similarly categorized along with the removal of boosted topology events. The postfit SFs as well as the pulls and constraints of the most impactful nuisance were found to be similar. Moreover, the extracted signal strength was also compatible within uncertainties indicating the STXS categorization does not have an impact on the inclusive VH($\rightarrow$ b$\bar{\text{b}}$) signal strength.

- Fit model: Another important change in the fit model was that the previous analysis used only process SFs for each bin and no category migration nuisances. Similar SFs and nuisance parameter fit model were used to fit the event categorization similar to the previous analysis. It was found that the postfit SFs as well as the pulls and constraints of this analysis are compatible within the uncertainties of the current STXS analysis.

- Event Selection: The previous analysis [130] used Tight WP of DeepCSV tagger for leading AK4 b-jet selection while the current STXS analysis uses Medium WP so as to retain b-jets in the STXS categories. The analysis selection was thus replicated and it was found that apart from the expected changes in the constraints on the V+b SFs and other correlated nuisances, the fit results (process SF and other impactful nuisance parameters) were consistent within uncertainties.

- MVA inputs: Potential bias from usage of DNN in SR of resolved topology in the fit model was also tested. DNNs were replaced with BDTs in the resolved topology. As expected, replacing DNNs with BDTs led to an increase in uncertainty, however the fit parameters were found to be consistent. Along with this test, DNNs were trained using similar inputs as the previous analysis [130] with the major change being the use of full shape of DeepCSV score instead of WP bins and not using kinematic fit observables in the 2-lepton channel inputs. The fit results were found to be compatible with the current analysis.

- Boosted analysis: Several checks were performed to ensure no bias in the STXS measurements was due to the addition of boosted topology. Following are some of the checks:

  - removal of the boosted and resolved SR and CR above $p_T^V > 250$ GeV.
  - decorrelation of background process SF between resolved and boosted topology.
  - removal of boosted SR and CR.

  The fit results were found to be consistent in all the checks discussed above.

- Extrapolation SF: Since in the current fit models, no additional extrapolation SFs between CR and SR are used, SF for SR+CR fit was compared with CR-only fit. SFs were found to be consistent within uncertainties in both cases. Furthermore, the HF CR was defined to include either the left or the right dijet invariant mass sideband only to check if the V+jets SF were consistent given the absence of any extrapolation uncertainty from CR to SR. These checks also gave consistent fit results ruling out any bias on the signal extraction from the absence of extrapolation SF from CR to SR.

Part III
GNN-based efficiency parameterization of b-tagging classifier

4 Efficiency parametrization of a b-tagging classifier using Graph Neural Networks

As shown in Table 23, the dominant source of systematic uncertainty on the signal strength extraction of the VHb$\bar{\text{b}}$ analysis is the sample size of the MC simulation. This mainly comes from the limited MC sample size of the dominant backgrounds (such as V+jets in high p_T(V) bins) and the usage of the NLO samples having significant events with negative generator weights. Also, the initial MC samples are already very large ($\sim 10^3$ events) and it is practically not possible to produce a larger statistics sample. To address this source of uncertainty, a new methodology of weighting events by their probability of passing a selection cut is proposed. These weights are referred to as the efficiency weights in the text.

For example, for the signal region, we select events with the leading jet passing the Tight WP and sub-leading jet passing the Medium WP of their DeepCSV b-tagging score. Instead of selecting only such events, this method retains all the events (hence increasing the statistical power of the MC simulation) but weighs them by the probability of each event to have the leading jet passing the Tight WP and the sub-leading jet passing the Medium WP of their respective DeepCSV b-tagging scores.

4.1 Jet classifier efficiency measurement algorithms

In the following sub-sections we discuss various methods to measure the efficiency (ε) of the jet classifiers, namely the direct tagging (Section 4.1.1), efficiency maps (4.1.2), and Graph Neural Network i.e GNN-based approach (Section 4.1.2) [131]. The GNN-based approach is discussed in detail in Sections 4.2 and 4.3.

4.1.1 Efficiency using selection cuts

This method, also known as direct tagging, involves calculating the efficiency by selecting the events passing a given selection cut. Here, the selection cut is the WP of the DeeepCSV score:

$$\varepsilon = \frac{N\left(s > s_{\text{wp}}\right)}{N} \in [0, 1] \tag{4.1}$$

where $N(s > s_{\text{wp}})$ is the number of events having a given jet passing a particular WP score (s_{wp}) of DeepCSV and s is the DeepCSV score of that particular jet. The uncertainty of the efficiency measurement associated with this approach depends on the number of events that can be simulated in a given region of the selection cuts. This approach thus suffers from the limited statistical precision of the simulation. For VH(b$\bar{\text{b}}$) STXS measurements, the region corresponding to > DeepCSV Tight WP of the leading and sub-leading Higgs boson candidate jets is significantly correlated with the events in the high DNN score region, where the S/B ratio is large and thus relevant for signal extraction. This region is also the one that suffers the most due to limited MC statistical precision.

4.1.2 Efficiency weighting

In this method, also known as 'truth tagging', a weight is applied to each jet. The event weight is then calculated as weight(event) = weight(leading jet) × weight(subleading jet). This is different from the cut-based approach where an event is either accepted or rejected.

In order to achieve a parameterization of the classifier efficiency weights, a set of low-level observables (θ) are chosen which capture the dependency of the classifier score ($s(x)$ where x are the inputs of the classifier) on the jet characteristics and the event environment (such as neighboring jets).

$$\varepsilon(\theta) = \frac{N\left(s(x) > s_{\mathrm{wp}} \mid \theta\right)}{N(\theta)} \tag{4.2}$$

Example choices for θ are variables like the jet transverse momentum (p_T) and pseudorapidity (η) which are correlated to the secondary vertex reconstruction, track reconstruction, and efficiency, etc. There are several ways to estimate efficiency weights. We will discuss:

- Efficiency maps
- GNN (multidimensional parameterization)

Efficiency Maps The efficiency of each flavor of the jet (f) is parameterized in bins of p_T and η. For this particular flavor of jet required in the analysis:

$$\varepsilon_{\mathrm{f},i,j} = \frac{N_{\mathrm{f}}\left((s > s_{\mathrm{wp}}) \wedge (p_{\mathrm{T}_i} \leq p_{\mathrm{T}} < p_{\mathrm{T}_{i+1}}) \wedge (\eta_j \leq \eta < \eta_{j+1})\right)}{N_{\mathrm{f}}\left((p_{\mathrm{T}_i} \leq p_{\mathrm{T}} < p_{\mathrm{T}_{i+1}}) \wedge (\eta_j \leq \eta < \eta_{j+1})\right)} \tag{4.3}$$

where $\varepsilon_{\mathrm{f},i,j}$ is the efficiency weight of the f flavoured jet in the ith p_T and jth η bin.

Figure 132 shows the efficiency map for Tight WP of the DeepCSV score for different flavors of the jet in QCD multijet sample as a function of the jet p_T and η.

The main limitations of this method are:

- The efficiency is parameterized in a small number of dimensions based on the available MC sample size (i.e. it suffers from the curse of dimensionality, higher dimensionality requires larger sample size). For example, in the above case, only the efficiency dependence on p_T and η is captured.
- correlations between neighboring jets (also referred as the 'environment effects') are neglected.

GNN approach This approach was first proposed by researchers in ATLAS [132]. As in standard GNN architecture (see Section 4.3.1), nodes are used as entities, and edges are used to establish a relationship between the entities. In this analysis, we take the full event as input to the GNN (jets as nodes and ΔR between the jets as edges) which then provides efficiency weights for each jet flavor and for each of the standard working points of the jet classifier. This approach also captures higher-order

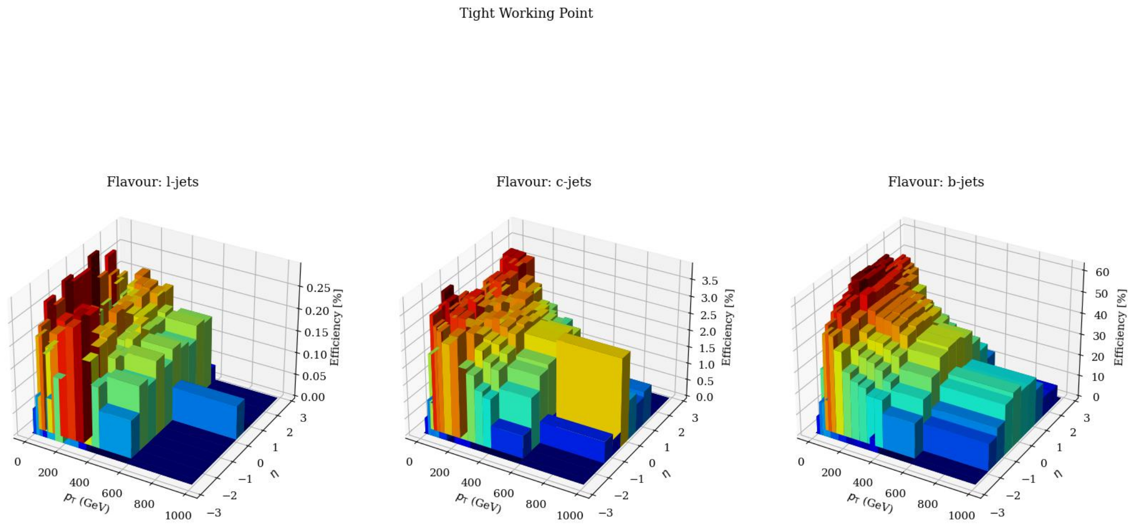


Figure 132: Efficiency map for Tight WP of DeepCSV score for different flavors of the jet in the QCD multijet sample.

correlations among its inputs and the output (i.e. efficiency weight), and correlations among the jets. Moreover, no binning of the parameterization variables is required (as the efficiency weights are calculated per event), thus providing unbinned efficiency weights.

4.2 Gain in statistical uncertainty due to efficiency weighting technique over direct tagging

For a given bin if w_i is the weight of the ith event, the total bin weight and the uncertainty on the total bin weight are [133]:

$$b = \sum_{i=1}^{N} w_i \quad \sigma_b = \sqrt{\sum_{i=1}^{N} w_i^2} \tag{4.4}$$

Since the number of unweighted events is the same in a bin for the efficiency weighted histogram (ew) and the direct tagging histogram (dt), we can write:

$$b_{\text{ew}} = N_{\text{total}} \cdot \varepsilon = N_{\text{selected}} = b_{\text{dt}} \tag{4.5}$$

where ε is the efficiency weight of an event in the efficiency histogram. Assuming a weight of 1 for direct tagging histogram,

$$\sigma_{b_{\text{dt}}} = \sqrt{b_{\text{dt}}}, \tag{4.6}$$

Assuming ε as the weight of each event in a bin of efficiency weighted histogram,

$$\sigma_{b_{\text{ew}}} = \sqrt{N_{\text{total}} \cdot \varepsilon^2} = \sqrt{\varepsilon} \cdot \sqrt{N_{\text{selected}}} = \sqrt{\varepsilon} \cdot \sigma_{b_{\text{dt}}}, \tag{4.7}$$

Thus, the statistical uncertainty in efficiency weight based approach is reduced by a factor of $\sqrt{\varepsilon}$ over the cut-based approach.

4.3 GNN approach in detail

4.3.1 GNN architecture

GNNs are used in particle physics [134] since this architecture describes an event better than a feed-forward neural network [135]. This is because its architecture takes into account not only the features of a given particle but also the relationship of that particle with its neighboring particles. A graph consists of nodes and edges. In the context of this thesis, the graph is an event, in which jets of the event are represented as nodes and the relationship between the jets as edges. Thus an event having N jets has N nodes and N − 1 edges. Both nodes (sometimes referred to as vertex as well) and edges of a graph have features. The nodes, i.e. jets, have features such as $p_{\mathrm{T}}, \eta, \phi$ while edges have features such as the ΔR distance between two jets, invariant mass of the two jets, etc. In this thesis, we perform the so-called node classification. In the node classification, we estimate the probability of a node belonging to some predefined classes. In this analysis, the classes chosen represent the different bins of the DeepCSV score. For example, exclusive classes such as 'above Tight WP', 'between Medium and Tight WP', etc.

A sketch of the GNN architecture used in this thesis is shown in Figure 133.

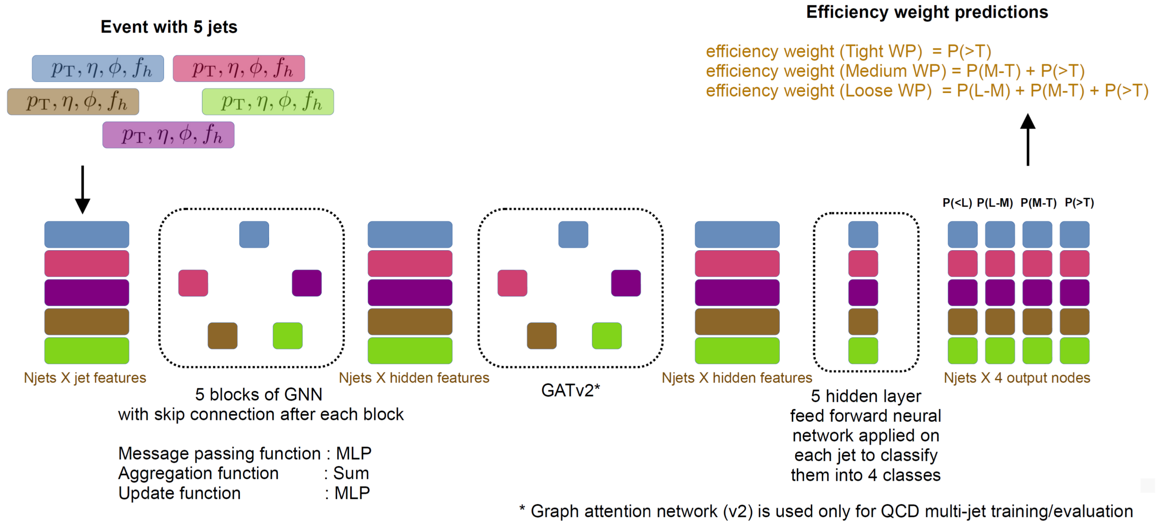

Event with 5 jets
$p_{\mathrm{T}}, \eta, \phi, f_h$
$p_{\mathrm{T}}, \eta, \phi, f_h$
$p_{\mathrm{T}}, \eta, \phi, f_h$
$p_{\mathrm{T}}, \eta, \phi, f_h$
$p_{\mathrm{T}}, \eta, \phi, f_h$
Efficiency weight predictions
efficiency weight (Tight WP) = P(>T)
efficiency weight (Medium WP) = P(M-T) + P(>T)
efficiency weight (Loose WP) = P(L-M) + P(M-T) + P(>T)
P(<L) P(L-M) P(M-T) P(>T)
Njets X jet features
5 blocks of GNN
with skip connection after each block
Message passing function : MLP
Aggregation function : Sum
Update function : MLP
Njets X hidden features
GATv2*
Njets X hidden features
5 hidden layer
feed forward neural
network applied on
each jet to classify
them into 4 classes
Njets X 4 output nodes
* Graph attention network (v2) is used only for QCD multi-jet training/evaluation

Figure 133: Overview of the GNN architecture. A detailed explanation is given in Section 4.3.1.

Consider an event with N^v nodes and $N^e = N^v - 1$ edges. Consider a particular node, the so-called source node s with features $n_{f,s}$. The source node is connected to all so-called destination nodes $d_i \in D$ (d_i is a particular destination node in the set D of destination nodes corresponding to the source node s) in the graph each with features $n_{f,d}$ via an edge having feature $e_{f,sd}$ (see Figure 134). For a given source node in a GNN, the following steps are performed:

- Message passing: The features of the source node, a particular destination node, and the edge between them are passed through a feed-forward neural network (or

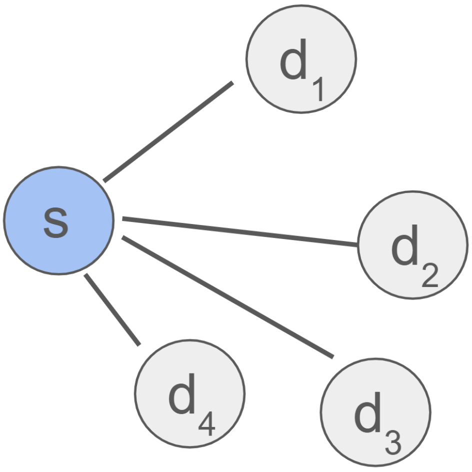


Figure 134: Representation of an event with 5 jets as a graph. The 's' labelled node represents the source node while destination nodes, 'd$_{\mathrm{i}}$', are connected to the source node via edges.

multi-layer perception i.e. MLP) to obtain latent/abstract representation $e_{l,sde}$ of its inputs. The MLP used in this step is referred to as the 'edge network' in this thesis. The edge network used in this thesis is a single hidden layer neural network having 256 dimensions. The output dimension of the MLP which captures the latent representation of the inputs was set to 256.

- Aggregation: For a given source node, the message passing is performed separately with all other $d_i \in D$ destination nodes to obtain a total of N-1 latent representations representing the relationship between the source node and N-1 destination nodes. For the given source node, all the N-1 latent representations are then summed to represent the latent source node features $n_{fl,s}$. $n_{fl,s}$ features now contain the information not only about the source node's initial features $n_{f,s}$, but also about all other nodes and edges/relationships between them.

- Update source node: The source node features $n_{f,s}$ are now updated by using a feed-forward network with inputs as the initial node feature $n_{f,s}$ and the latent source node features $n_{fl.s}$ to obtain the updated features of the source node as $n_{fnew,s}$. The MLP used in this step is referred to as the 'node network' in this thesis. The node network used in this thesis is a single hidden layer neural network having 512 dimensions. The output dimension of the MLP was set to 512.

The above steps are simultaneously performed for each of the nodes in the event i.e. each node is considered a source node with all other nodes as destination nodes for that source node. These steps together form one block of the GNN shown in Figure 133. By repeating the block five times, we perform this process five successive times for each node in the event. This forms the first part of the architecture.

The second part of the architecture is the attention block, the GATv2 [136] block. In this block, instead of performing summation in the aggregation step, we perform a weighted summation where the weight denotes the importance of a given edge and a destination node to a source node. This weighted summation thus helps to capture additional environmental information about the importance of particular neighboring nodes over other neighboring nodes to the source node. The un-normalized attention weight $we_{f,sd}$ for a given source node s and destination node d can be given as:

$$we_{f,sd} = \mathbf{W}_a \text{LeakyReLU}\Big(\mathbf{W}_l \times n_{fnew,s} + \mathbf{W}_r \times n_{fnew,d}\Big) \tag{4.8}$$

Here, $\mathbf{W}_a, \mathbf{W}_l, \mathbf{W}_r$ are weight matrices corresponding to a linear trainable layer (a single dense layer with bias set to 0) with 512 nodes each. $\mathbf{W}_l \times n_{fnew,s}$ gives the latent representation of the source node, and $\mathbf{W}_r \times n_{fnew,d}$ gives the latent representation of the destination node. The non-linear LeakyReLU function [137] defined as $f(x) = \max(0.01x, x)$ is used as the activation function. Using the softmax layer over the un-normalized weights of a given source node, we can obtain normalized attention weights of a given source node s and destination node d as:

$$\alpha_{sd} = \text{softmax}_j(we_{f,sd}) = \frac{\exp(we_{f,sd})}{\sum_{d' \in \mathcal{D}} \exp(we_{f,sd'})} \tag{4.9}$$

where D represents the set of all the destination nodes. Once the attention weights are available, the aggregation is performed using weighted sum, after which the source node is updated as discussed in the step 'Update source node'.

The third part of the network is a 5-layer feed-forward neural network having 512, 256, 128, 50, and 4 nodes respectively. It takes as input the updated source node features $n_{fnew,s}$ and the initial source node features $n_{f,s}$ and classifies the node in four exclusive DeepCSV WP classes.

We use $p_{\mathrm{T}}, \eta, \phi$, f_h (hadron flavor of the jet) of the jet as input node features and ΔR between the two jets as input edge feature. A two-dimensional embedding vector is used to embed the categorical variable f_h. Four output classes are defined to classify nodes/jets into the four exclusive DeepCSV WP categories: 'below Loose WP', 'between Loose WP and Medium WP', 'between Medium WP and Tight WP', 'above Tight WP'. The efficiency weight corresponding to each working point is then calculated by summing all the probabilities for the jet to be classified above that working point class. For example, the efficiency weight corresponding to the Medium WP of a jet (to be used when the jet has the DeepCSV score above Medium WP) is the sum of the probability of the jet to be in class 'between Medium WP and Tight WP' and 'above Tight WP'. The MC dataset is split into the ratio of 75:25 for the train and test sets. The training set is further split in the ratio of 95:5 for the training and validation set. The model overtraining was checked by monitoring the cross entropy loss on the train and validation set. The unbiased test set was then used to present the results in this thesis. The hyperparameters corresponding to the first and the third part of the network are the same as the ones used by the ATLAS [132]. The hyperparameters of the second part of the model, the GATv2 block (i.e. the dimension of the three linear trainable layers) were set to 512 since the third part of the model, the feed-forward network expects an input of 512 dimensions.

4.3.2 GNN training and evaluation

We use $t\bar{t}$ dilepton (3 million events) and QCD samples enriched in muon having leading jet in 300-470 GeV p_T slice (600 thousand events) for training. The dilepton $t\bar{t}$ process is the $t\bar{t}$ process where $t \to bW^+$, $W^+ \to l^+\nu$, thus giving rise to two leptons and two b-jets in the final state. The presence of muon in the QCD multi-jet event indicates the semi-leptonic decay of b-jets leading to muon production. Thus, with muon-enriched QCD multi-jet events, we obtain multi-jet events enriched in b-jets.

Bootstrap aggregation, also known as 'bagging' [138], is performed to obtain the central value and uncertainty of an observable having events weighted by the GNN predicted efficiency weights. In bagging, a model is trained multiple times on different subsets of the training data, sampled with replacement. The sampling with replacement technique involves selecting events from the sample where each event selected is returned to the sample before the next event, allowing for the possibility of selecting the same event multiple times. The predictions on a common test set of these multiple trainings are then aggregated to obtain the central value (median of the aggregation) and uncertainty (standard deviation of the aggregation). The central value of an observable in a given histogram bin is the median of the value of that bin computed on all the bootstrap samples. Its uncertainty is the corresponding statistical uncertainty of the central values of that bin of the bootstrap samples.

Figure 135 shows the distribution of the transverse momentum of the jet (p_T(j)) using the QCD sample. The top pad shows the predictions for the 7 bootstrap samples and the ensemble obtained by using the median of the central values of the bootstrap samples. The middle pad shows the ratio of the central values of each of the bootstrap samples with the ensemble. The bottom pad shows the ratio of statistical uncertainty of an individual bootstrap training with the ensemble. As shown in the middle plot, the prediction of the central values of the individual bootstrap samples is within 5% of the median central value.

We use the χ^2 as a distance metric to quantify the modelling of histograms of an observable obtained by efficiency weight-based methods (H_{ew}) (efficiency map and GNN approach) with respect to ('the ground truth') the direct tagging (H_{dt}). This metric is used to quantify the modelling of the central values predicted by the efficiency weight-based approaches.

$$\mathrm{D}_{\chi^2}\left(H_{\text{ew}}, H_{\text{dt}}\right) = \sum_{i=1}^{N_{\text{bins}}} \frac{\left(O_i - E_i\right)^2}{E_i} = \sum_{i=1}^{N_{\text{bins}}} \frac{\left(b_{\text{ew},i} - b_{\text{dt},i}\right)^2}{b_{\text{dt},i}} \tag{4.10}$$

where

- O_i and E_i refer to the observed and expected number of events in the ith bin of the histogram obtained using direct tagging and efficiency weight-based methods respectively

- $b_{\text{dt},i}$ and $b_{\text{ew},i}$ are the number of events in the ith bin of histogram obtained by direct tagging and efficiency weight-based methods respectively.

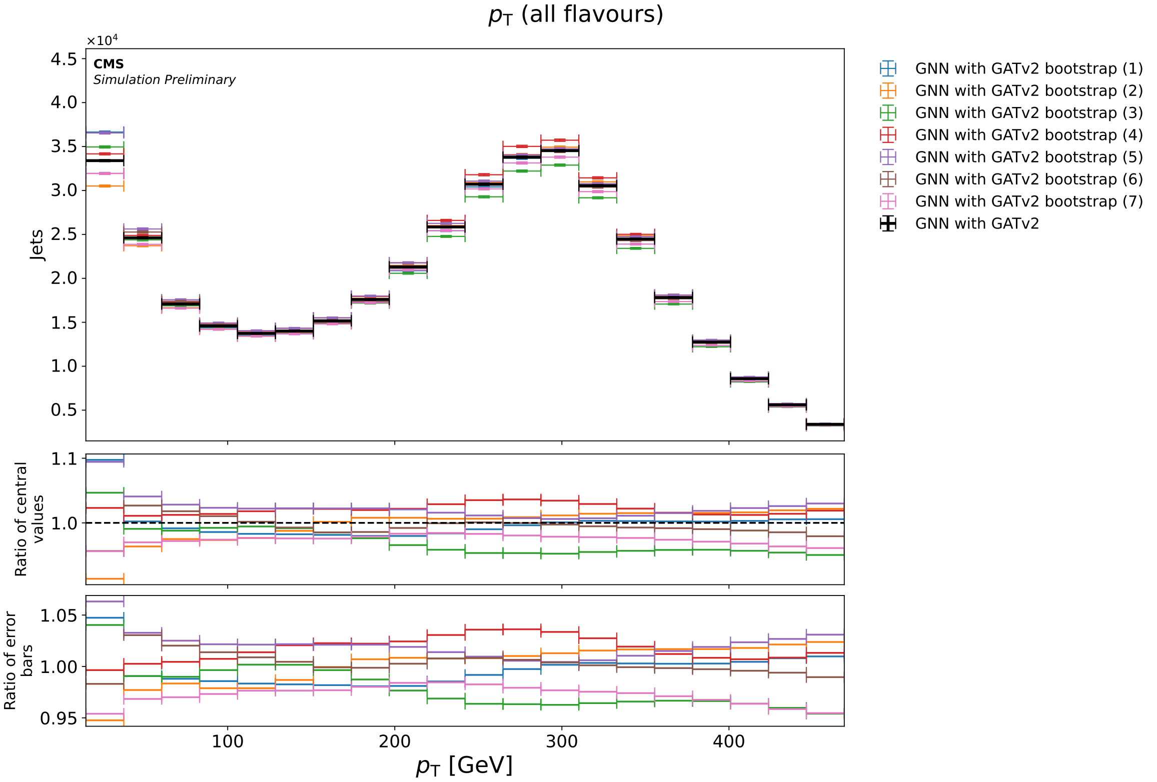


Figure 135: Distribution of transverse momentum of the jet (p_T(j)) for QCD multijet sample. The top pad shows the predictions for the bootstrap sample (colorful markers) and the ensemble (black marker) obtained using the median of the central values of the bootstrap samples. In this Figure, seven individual bootstrap samples are listed in the legend. The middle pad shows the ratio of the central values of each of the bootstrap samples with the ensemble. The bottom pad shows the ratio of statistical uncertainty of individual bootstrap training with the ensemble.

4.4 Results

Table 27 and 28 show the comparison of the modelling of the central values of the efficiency weight-based methods (efficiency map and GNN approach) compared to direct tagging for Tight and Medium working points respectively. The single jet observables listed are the transverse momentum (p_T(j)), pseudo-rapidity (η(j)), azimuthal angle (ϕ(j)), area (area(j)) of the jets (πR^2) in the event while the dijet observables listed are mass (m(j)), and invariant mass (m(jj)) and ΔR(jj) of the leading and sub-leading jets in the event. Since both the single jet variables and dijet variables are used in physics analyses such as VH($\rightarrow$ b$\bar{\text{b}}$), the performance of the efficiency weight-based methods is checked for both the sets of variables. Table 27 gives the predictions corresponding to the $t\bar{t}$ sample while Table 28 for the QCD sample. The GNN with attention (GATv2) model is used only for QCD training/evaluation. As GNN capturing higher-order correlations between its inputs of a jet, and also of its neighboring jets in the event, based on the χ^2 metric, outperforms the efficiency map approach for both Tight and Medium working point predictions. The usage of attention models as tested for QCD samples leads to further improvement in variables of a single jet compared to the dijet variables. This is attributed to the additional environment information for each jet added through the attention mechanism. For Loose WP, the modelling of the central values of the predictions using the GNN-based approach with the direct tagging, was found to be comparable to that obtained using the efficiency map based approach.

	Efficiency map	GNN
p_T(j)	203.86	113.78
η(j)	350.01	103.53
ϕ(j)	145.34	61.81
m(j)	232.11	186.12
area(j)	105.30	85.79
m(jj)	22.16	11.71
ΔR(jj)	25.85	11.00

(a) Predictions corresponding to the $t\bar{t}$ sample.

	Efficiency map	GNN	GNN with GATv2
p_T(j)	20.02	30.36	14.66
η(j)	50.11	36.28	13.31
ϕ(j)	24.67	28.37	18.12
m(j)	25.84	32.17	14.96
area(j)	15.67	23.15	7.15
m(jj)	22.55	8.44	6.81
ΔR(jj)	23.83	9.59	8.18

(b) Predictions corresponding to the QCD sample.

Table 27: Comparison of the modelling of the central values of the efficiency weight-based methods (Efficiency map and GNN approach) with respect to direct tagging for the Tight working point using the χ^2 metric. The GNN with attention (GATv2) model is used only for QCD training/evaluation.

	Efficiency map	GNN
p_T(j)	388.09	303.01
η(j)	5997.29	2441.21
ϕ(j)	192.82	153.69
m(j)	358.32	314.00
area(j)	199.80	174.52
m(jj)	48.52	26.17
ΔR(jj)	48.02	26.89

(a) Predictions corresponding to the $t\bar{t}$ sample.

	Efficiency map	GNN	GNN with GATv2
p_T(j)	71.18	74.92	37.04
η(j)	731.08	450.10	67.73
ϕ(j)	34.24	35.60	20.29
m(j)	81.17	81.67	44.84
area(j)	36.52	34.66	17.12
m(jj)	24.72	8.64	6.56
ΔR(jj)	24.81	9.66	7.33

(b) Predictions corresponding to the QCD sample.

Table 28: Comparison of the modelling of the central values of the efficiency weight-based methods (Efficiency map and GNN approach) with respect to direct tagging for the Medium working point using the χ^2 metric. The GNN with attention (GATv2) model is used only for QCD training/evaluation.

Figure 136 and 137 show the dijet mass distribution and the ΔR of the leading and sub-leading jet pair for the Tight working point for QCD and $t\bar{t}$ samples respectively. The results are inclusive in the flavor of the jet. The top pad shows the histograms resulting from the direct tagging, efficiency map, and the GNN approach. The middle pad shows the ratio of the central value of the efficiency map and the GNN approach

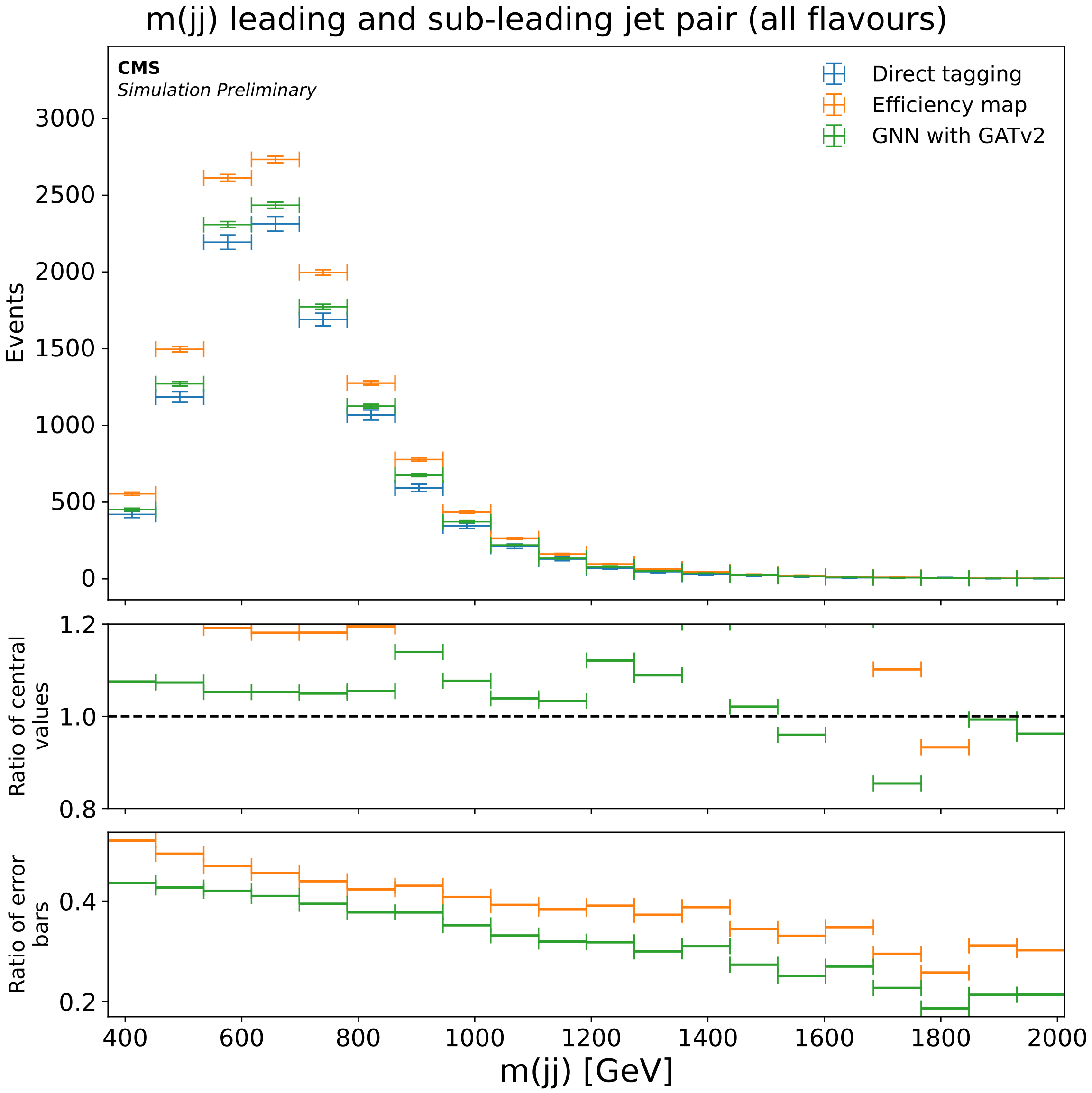


Figure 136: Dijet invariant mass distribution of the leading and sub-leading jet pair for the Tight working point. The top pad shows the histograms resulting from the direct tagging, efficiency map, and the GNN approach. The middle pad shows the ratio of the central value of the efficiency map and the GNN approach to direct tagging. The bottom pad shows the ratio of the statistical uncertainty of the bin values resulting from the efficiency map and the GNN approach to direct tagging.

to direct tagging. The bottom pad shows the ratio of the statistical uncertainty of the bin values resulting from the efficiency map and the GNN approach to direct tagging. The central values of the GNN-based approach with respect to the direct tagging approach show better modelling than the 2D efficiency map approach.

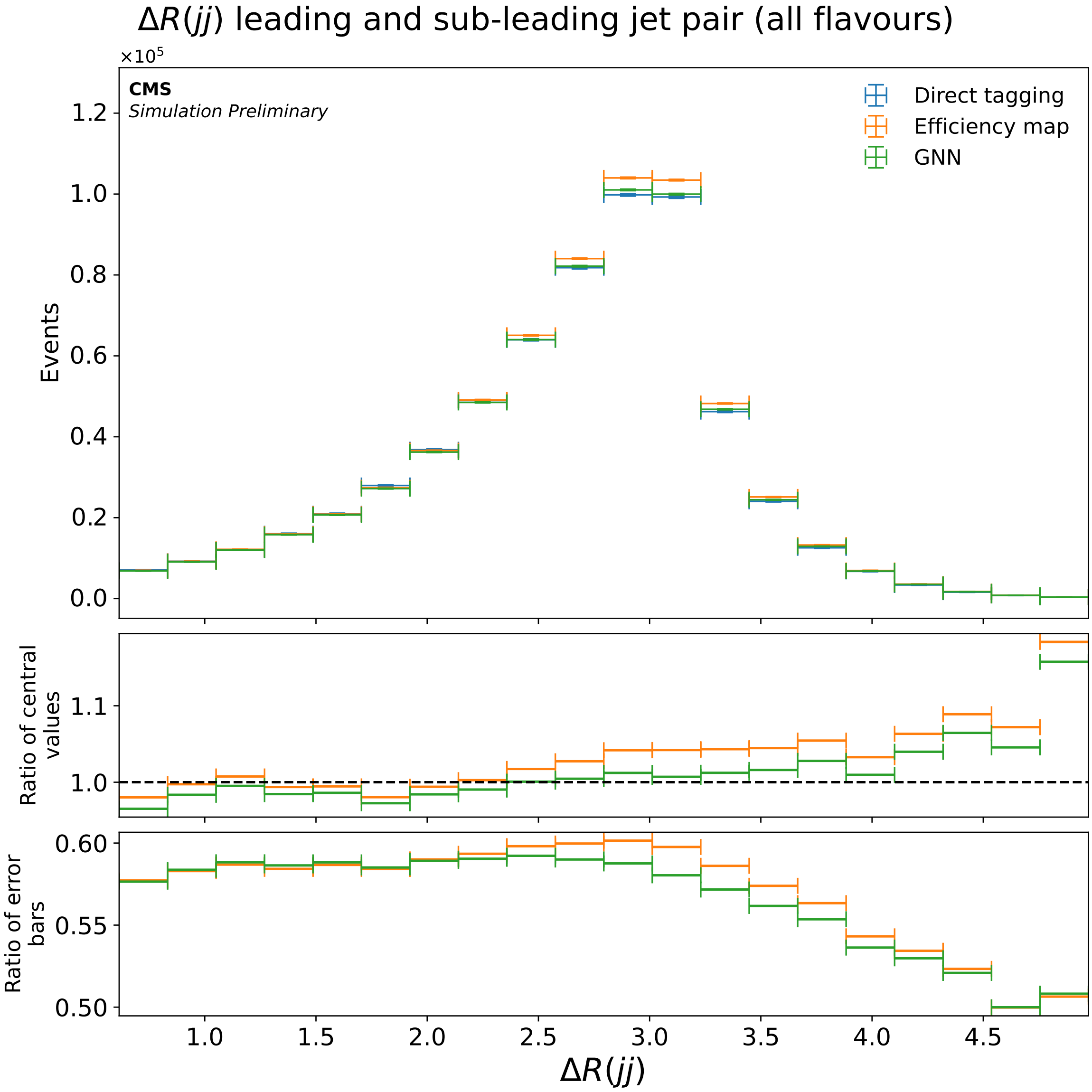


Figure 137: ΔR distribution of the leading and sub-leading jet pair for the Tight working point in $t\bar{t}$ sample. The top pad shows the histograms resulting from the direct tagging, efficiency map, and the GNN approach. The middle pad shows the ratio of the central value of the efficiency map and the GNN approach to direct tagging. The bottom pad shows the ratio of the statistical uncertainty of the bin values resulting from the efficiency map and the GNN approach to direct tagging.

4.5 Summary and outlook

The estimation of the efficiency of a jet classifier using selection cuts is statistically limited as it depends on the number of events that can be simulated in a given phase space. Efficiency weighting approaches help to mitigate this issue. The traditional approach of efficiency map parameterized in p_T and η do not capture the efficiency dependence from other possibly important variables describing correlations among neighboring jets in an event. The GNN-based approach presented in the thesis helps to solve these issues and outperforms the traditional efficiency map approach for both Tight and Medium working point predictions. Also, as expected, the efficiency weights approach leads to significant gains in statistical uncertainty with respect to the direct tagging approach. The outlooks for this analysis are the following: improvement of the performance of the GNN-based approach with respect to Loose WP of DeepCSV; inclusion of the new parameters such as the mass of dijets (in addition to the ΔR between the dijets) as edge features could also be explored.

Part IV
Conclusion and Outlook

5 Conclusion and Outlook

The precision measurement of the VH($\rightarrow$ b$\bar{\text{b}}$) process has been presented in the inclusive fiducial phase space and in the STXS framework. The leptonic decay mode of the vector boson is targeted since the presence of prompt leptons and MET in the final state provides a trigger path to identify the signal events and suppresses the overwhelming QCD multijet background. The measurements in the STXS framework are performed in mutually exclusive bins of p_T of the vector boson and the number of jets additional to the Higgs candidate b-jets. Depending on the vector boson involved (W, Z), the analysis is split into three channels, 0-, 1- and 2-lepton channels, corresponding to the number of leptons in the final state. Events are categorized into SR or CRs. SRs are defined to have a high signal efficiency while the three CRs (TT CR, V+HF CR, V+LF CR) are defined to be enriched in one of the major background processes. Both resolved as well as boosted topologies are included in the analysis. The cross section and signal strength measurements are obtained by a simultaneous maximum likelihood fit to the MVA templates (DNN for resolved topology and BDT for boosted topology) in the SR and V+HF CR, and kinematic and flavor observables in TT and Z+LF CR.

The LHC Run 2 dataset collected in 2016-2018 by the CMS detector with an integrated luminosity of 138 fb^{-1} is used in this analysis. The analysis strategy is validated using a cross check analysis targeting the VZ($\rightarrow$ b$\bar{\text{b}}$) process where the vector boson decays leptonically and the Z boson decays to a pair of bottom quarks. The inclusive signal strength for this cross check analysis using the Run 2 data was measured to be $\mu = 1.25 \pm 0.14$ with both, the observed and the expected significance well over 5σ.

The inclusive signal strength of the VH($\rightarrow$ b$\bar{\text{b}}$) process for the Run 2 data was measured to be $\mu = 1.15 \pm 0.14(\text{stat.})^{+0.16}_{-0.15}(\text{syst.}) = 1.15^{+0.22}_{-0.20}$ (tot.) with an observed (expected) significance of 6.3(5.6) σ. This measurement corresponds to the first observation of the VH($\rightarrow$ b$\bar{\text{b}}$) process with the CMS detector. The signal strength as well as the cross section times branching ratios measurements in the STXS bins are shown in Figures 125 and 128 respectively. No statistically significant deviations from the SM expectations were observed in any of the STXS bins or in the inclusive, per-channel and per-process measurements. Future steps in the exploration of the VH($\rightarrow$ b$\bar{\text{b}}$) analysis apart from the analysis of larger datasets, target the reduction of systematic uncertainties.

The dominant source of systematic uncertainty on the inclusive signal strength extraction is the sample size of the MC simulation, dominated by the MC sample size of major backgrounds such as V+jets. To reduce these uncertainties, larger MC samples could be produced. However, this would require extensive resources and time. Also, since NLO samples are plagued by negative weighted events, samples larger than the corresponding LO samples would be required to reach similar MC statistical precision. Another way to reduce the statistical uncertainty without increasing the MC sample size is by using the efficiency weight approach extensively discussed in Section 4.3. The V+jets modeling uncertainty is another dominant source of uncertainty. The usage of more precise V+jets samples, such as NLO in QCD samples for

2016 datasets could reduce these uncertainties.

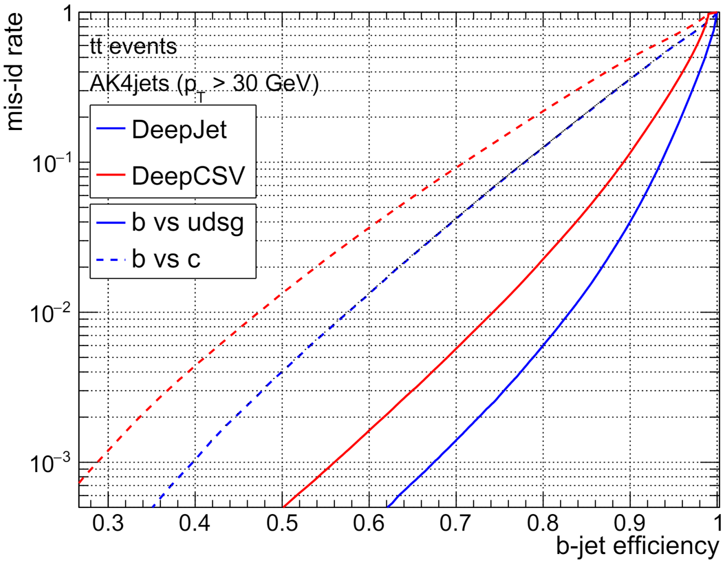


Figure 138: Performance of the DeepCSV and DeepJet algorithms for identifying AK4 b-jets [139]. The plot shows the probability of misidentification of non b-jets as b-jets with respect to the efficiency of identifying b-jets.

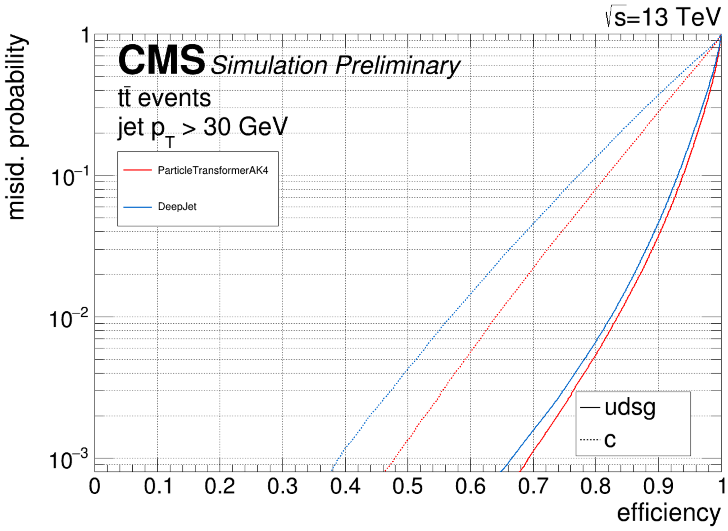


Figure 139: Performance of the DeepJet and ParticleTransformerAK4 algorithms for identifying AK4 b-jets [140].

The third dominant uncertainty is related to the calibration of the b-tagging algorithms used in this analysis. For resolved topology, the DeepCSV b-tagging algorithm was used. However, a more advanced algorithm, DeepJet [139] was also studied. It uses Convolution and Recurrent Neural Networks [141][142] in its architecture. Along with the change in the architecture, it also has more input features such as those of the neutral particles in the jet. It was found that the Asimov sensitivity of the STXS measurement improved by a maximum of 10% with respect to the usage of the DeepCSV algorithm due to better performance of the tagger (as shown in Figure 138). However, even after the usage of the external DeepJet calibration factors, significant differences were found in the data/MC modelling in the VH(→ $\mathrm{b\bar{b}}$) phase space, explaining why the DeepCSV algorithm was used in the current analysis. Recently, another b-tagging algorithm was also developed, called the Particle Transformer [143] which is expected to further improve the analysis sensitivity (as shown in Figure 139). The usage of this tagger requires further studies about data/MC modelling in the VH(→ $\mathrm{b\bar{b}}$) phase space as well as the availability of the dedicated calibration factors. For the boosted topology, the DeepAK8 b-tagging algorithm is used in this analysis. However, only the DeepAK8bbVSlight output node was used in the analysis due to the non-availability of external calibration factors for the other output nodes. This led to a loss of about 10% in the signal sensitivity with respect to the usage of all the output nodes. In the boosted regime, further improvements are expected by the usage of a more recent ParticleNet b-tagging algorithm [144] as shown in Figure 140. This algorithm is based on graph neural network architecture. However, as discussed, the usage of such algorithms in this analysis requires the availability of dedicated calibration factors and extensive validation of data/MC modelling in the analysis SR and CRs.

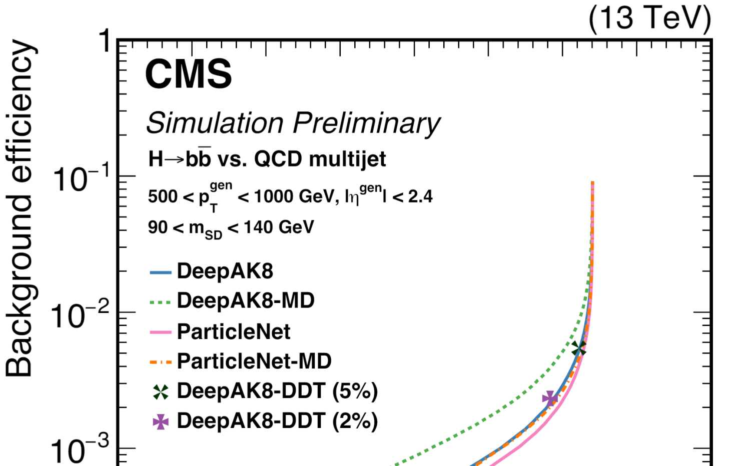


Figure 140: Performance of the DeepAK8 and ParticleNet algorithms for identifying FatJets from Higgs bosons and QCD multijet events [145]. MD (Adversarial training) and DDT (Designing Decorrelated Taggers) are the different approaches to decorrelate the taggers from the jet masses [146].

With the expected ×2 higher integrated luminosity in Run 3 than Run 2, the statistical uncertainty of the Run 3 STXS measurements is expected to reduce by 30% compared to the Run 2 measurements. While with the expected integrated luminosity of ×10 of the Run 3, the statistical uncertainty of the HL-LHC measurements is expected to reduce by 70% compared to Run 3 measurements. Thus, more fine-grained STXS bins could also be targeted, for example, the number of additional jets split in the WH($\rightarrow \mathrm{b\bar{b}}$) production mode. Finally, this analysis also serves as a baseline for further VH($\rightarrow \mathrm{b\bar{b}}$) measurements which constrains the Effective Field Theory (EFT) parameters and differential measurements.

Appendices

A Simulation samples and their cross section

Sample Name	Xsec (pb)
DYBJetsToLL_M-50_Zpt-100to200_TuneCP5_13TeV-madgraphMLM-pythia8_newgridpack/	3.224
DYBJetsToLL_M-50_Zpt-200toInf_TuneCP5_13TeV-madgraphMLM-pythia8_newgridpack/	0.3298
DYJetsToLL_BGenFilter_Zpt-100to200_M-50_TuneCP5_13TeV-madgraphMLM-pythia8_newgridpack/	2.671
DYJetsToLL_BGenFilter_Zpt-200toInf_M-50_TuneCP5_13TeV-madgraphMLM-pythia8_newgridpack/	0.3934
DYJetsToLL_M-50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8/	161.1
DYJetsToLL_M-50_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8/	0.1933
DYJetsToLL_M-50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8/	48.66
DYJetsToLL_M-50_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8/	0.003468
DYJetsToLL_M-50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8/	6.968
DYJetsToLL_M-50_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8/	1.743
DYJetsToLL_M-50_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8/	0.8052
DYJetsToLL_M-4to50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8/	204.0
DYJetsToLL_M-4to50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8/	54.39
DYJetsToLL_M-4to50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8/	5.697
DYJetsToLL_M-4to50_HT-600toInf_TuneCP5_13TeV-madgraphMLM-pythia8/	1.85
DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/	5343.0
DY1JetsToLL_M-50_LHEZpT_50-150_TuneCP5_13TeV-amcnloFXFX-pythia8	316.6
DY2JetsToLL_M-50_LHEZpT_50-150_TuneCP5_13TeV-amcnloFXFX-pythia8	169.6
DY1JetsToLL_M-50_LHEZpT_150-250_TuneCP5_13TeV-amcnloFXFX-pythia8	9.543
DY2JetsToLL_M-50_LHEZpT_150-250_TuneCP5_13TeV-amcnloFXFX-pythia8	15.65
DY1JetsToLL_M-50_LHEZpT_250-400_TuneCP5_13TeV-amcnloFXFX-pythia8	1.098
DY2JetsToLL_M-50_LHEZpT_250-400_TuneCP5_13TeV-amcnloFXFX-pythia8	2.737
DY1JetsToLL_M-50_LHEZpT_400-inf_TuneCP5_13TeV-amcnloFXFX-pythia8	0.1193
DY2JetsToLL_M-50_LHEZpT_400-inf_TuneCP5_13TeV-amcnloFXFX-pythia8	0.4477
DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8	6508.0
DYJetsToLL_0J_TuneCP5_13TeV-amcatnloFXFX-pythia8	5333.0
DYJetsToLL_1J_TuneCP5_13TeV-amcatnloFXFX-pythia8	962.8
DYJetsToLL_2J_TuneCP5_13TeV-amcatnloFXFX-pythia8	362.0
QCD_HT1000to1500_TuneCP5_13TeV-madgraph-pythia8/	1088.0
QCD_HT1500to2000_TuneCP5_13TeV-madgraph-pythia8/	99.11
QCD_HT2000toInf_TuneCP5_13TeV-madgraph-pythia8/	20.23
QCD_HT200to300_TuneCP5_13TeV-madgraph-pythia8/	1547000.0
QCD_HT300to500_TuneCP5_13TeV-madgraph-pythia8/	322600.0
QCD_HT500to700_TuneCP5_13TeV-madgraph-pythia8/	29980.0
QCD_HT700to1000_TuneCP5_13TeV-madgraph-pythia8/	6334.0
ST_s-channel_4f_leptonDecays_TuneCP5_PSweights_13TeV-amcatnlo-pythia8/	3.74
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8/	80.95
ST_t-channel_top_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8/	136.02
ST_tW_antitop_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8/	35.85
ST_tW_top_5f_inclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8/	35.85
TTToHadronic_TuneCP5_PSweights_13TeV-powheg-pythia8/	377.96
TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8/	88.29
TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8/	365.34
WBJetsToLNu_Wpt-100to200_TuneCP5_13TeV-madgraphMLM-pythia8_newgridpack/	5.542
WBJetsToLNu_Wpt-200toInf_TuneCP5_13TeV-madgraphMLM-pythia8_newgridpack/	0.801

Sample Name	Xsec (pb)
WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8/	1395.0
WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8/	1.074
WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8/	407.9
WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8/	0.03216
WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8/	57.48
WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8/	12.87
WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8/	5.366
WJetsToLNu_BGenFilter_Wpt-100to200_TuneCP5_13TeV-madgraphMLM-pythia8_newgridpack/	20.56
WJetsToLNu_BGenFilter_Wpt-200toInf_TuneCP5_13TeV-madgraphMLM-pythia8_newgridpack/	2.936
WJetsToLNu_TuneCP5_13TeV-madgraphMLM-pythia8/	52940.0
W1JetsToLNu_LHEWpT_50-150_TuneCP5_13TeV-amcnloFXFX-pythia8	2661.0
W2JetsToLNu_LHEWpT_50-150_TuneCP5_13TeV-amcnloFXFX-pythia8	1331.0
W1JetsToLNu_LHEWpT_100-150_TuneCP5_13TeV-amcnloFXFX-pythia8	286.1
W2JetsToLNu_LHEWpT_100-150_TuneCP5_13TeV-amcnloFXFX-pythia8	277.7
W1JetsToLNu_LHEWpT_150-250_TuneCP5_13TeV-amcnloFXFX-pythia8	71.9
W2JetsToLNu_LHEWpT_150-250_TuneCP5_13TeV-amcnloFXFX-pythia8	105.9
W1JetsToLNu_LHEWpT_250-400_TuneCP5_13TeV-amcnloFXFX-pythia8	8.05
W2JetsToLNu_LHEWpT_250-400_TuneCP5_13TeV-amcnloFXFX-pythia8	18.67
W1JetsToLNu_LHEWpT_400-inf_TuneCP5_13TeV-amcnloFXFX-pythia8	0.885
W2JetsToLNu_LHEWpT_400-inf_TuneCP5_13TeV-amcnloFXFX-pythia8	3.037
WJetsToLNu_0J_TuneCP5_13TeV-amcatnloFXFX-pythia8	54500.0
WJetsToLNu_1J_TuneCP5_13TeV-amcatnloFXFX-pythia8	8750.0
WJetsToLNu_2J_TuneCP5_13TeV-amcatnloFXFX-pythia8	3010.0
WW_TuneCP5_13TeV-pythia8/	115.3
WWTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8/	50.858
WWToLNuQQ_NNPDF31_TuneCP5_PSweights_13TeV-powheg-pythia8/	50.858
WZ_TuneCP5_13TeV-pythia8/	48.1
WminusH_HToBB_WToLNu_M125_13TeV_powheg_pythia8/	0.10899
WplusH_HToBB_WToLNu_M125_13TeV_powheg_pythia8/	0.17202
ZBJetsToNuNu_Zpt-100to200_TuneCP5_13TeV-madgraphMLM-pythia8_newgridpack/	6.209
ZBJetsToNuNu_Zpt-200toInf_TuneCP5_13TeV-madgraphMLM-pythia8_newgridpack/	0.6286
ZH_HToBB_ZToNuNu_M125_13TeV_powheg_pythia8/	0.09322
ZJetsToNuNu_BGenFilter_Zpt-100to200_TuneCP5_13TeV-madgraphMLM-pythia8_newgridpack/	1.689
ZJetsToNuNu_BGenFilter_Zpt-200toInf_TuneCP5_13TeV-madgraphMLM-pythia8_newgridpack/	0.2476
ZJetsToNuNu_HT-100To200_13TeV-madgraph/	302.8
ZJetsToNuNu_HT-200To400_13TeV-madgraph/	92.59
ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph/	0.005146
ZJetsToNuNu_HT-400To600_13TeV-madgraph/	13.18
ZJetsToNuNu_HT-600To800_13TeV-madgraph/	3.257
ZJetsToNuNu_HT-800To1200_13TeV-madgraph/	1.496
Z1JetsToNuNu_M-50_LHEZpT_50-150_TuneCP5_13TeV-amcnloFXFX-pythia8	596.4
Z2JetsToNuNu_M-50_LHEZpT_50-150_TuneCP5_13TeV-amcnloFXFX-pythia8	325.7
Z1JetsToNuNu_M-50_LHEZpT_150-250_TuneCP5_13TeV-amcnloFXFX-pythia8	17.98
Z2JetsToNuNu_M-50_LHEZpT_150-250_TuneCP5_13TeV-amcnloFXFX-pythia8	29.76
Z1JetsToNuNu_M-50_LHEZpT_250-400_TuneCP5_13TeV-amcnloFXFX-pythia8	2.057
Z2JetsToNuNu_M-50_LHEZpT_250-400_TuneCP5_13TeV-amcnloFXFX-pythia8	5.166
Z1JetsToNuNu_M-50_LHEZpT_400-inf_TuneCP5_13TeV-amcnloFXFX-pythia8	0.224
Z2JetsToNuNU_M-50_LHEZpT_400-inf_TuneCP5_13TeV-amcnloFXFX-pythia8	0.8457
ZZ_TuneCP5_13TeV-pythia8/	14.6
ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/	10.88

Sample Name	Xsec (pb)
ZZTo4L_13TeV_powheg_pythia8/	2.038
ggZH_HToBB_ZToLL_M125_13TeV_powheg_pythia8/	0.0072
ggZH_HToBB_ZToNuNu_M125_13TeV_powheg_pythia8/	0.01437

Table 29: Simulation samples used in 2017 with their cross section.

Sample Name	Xsec (pb)
DYBJetsToLL_M-50_Zpt-100to200_TuneCP5_13TeV-madgraphMLM-pythia8/	3.206
DYBJetsToLL_M-50_Zpt-200toInf_TuneCP5_13TeV-madgraphMLM-pythia8/	0.3304
DYJetsToLL_BGenFilter_Zpt-100to200_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/	2.662
DYJetsToLL_BGenFilter_Zpt-200toInf_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/	0.3949
DYJetsToLL_M-50_HT-100to200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/	160.8
DYJetsToLL_M-50_HT-1200to2500_TuneCP5_PSweights_13TeV-madgraphMLM-pythia/	0.1931
DYJetsToLL_M-50_HT-200to400_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/	48.63
DYJetsToLL_M-50_HT-2500toInf_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/	0.003513
DYJetsToLL_M-50_HT-400to600_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/	6.982
DYJetsToLL_M-50_HT-600to800_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/	1.756
DYJetsToLL_M-50_HT-800to1200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/	0.8094
DYJetsToLL_M-50_TuneCP5_13TeV-madgraphMLM-pythia8/	5343.0
DY1JetsToLL_M-50_LHEZpT_50-150_TuneCP5_13TeV-amcnloFXFX-pythia8	316.6
DY2JetsToLL_M-50_LHEZpT_50-150_TuneCP5_13TeV-amcnloFXFX-pythia8	169.6
DY1JetsToLL_M-50_LHEZpT_150-250_TuneCP5_13TeV-amcnloFXFX-pythia8	9.543
DY2JetsToLL_M-50_LHEZpT_150-250_TuneCP5_13TeV-amcnloFXFX-pythia8	15.65
DY1JetsToLL_M-50_LHEZpT_250-400_TuneCP5_13TeV-amcnloFXFX-pythia8	1.098
DY2JetsToLL_M-50_LHEZpT_250-400_TuneCP5_13TeV-amcnloFXFX-pythia8	2.737
DY1JetsToLL_M-50_LHEZpT_400-inf_TuneCP5_13TeV-amcnloFXFX-pythia8	0.1193
DY2JetsToLL_M-50_LHEZpT_400-inf_TuneCP5_13TeV-amcnloFXFX-pythia8	0.4477
DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8	6508.0
QCD_HT1000to1500_TuneCP5_13TeV-madgraphMLM-pythia8/	1088.0
QCD_HT1500to2000_TuneCP5_13TeV-madgraphMLM-pythia8/	99.11
QCD_HT2000toInf_TuneCP5_13TeV-madgraphMLM-pythia8/	20.23
QCD_HT200to300_TuneCP5_13TeV-madgraphMLM-pythia8/	1547000.0
QCD_HT300to500_TuneCP5_13TeV-madgraphMLM-pythia8/	322600.0
QCD_HT500to700_TuneCP5_13TeV-madgraphMLM-pythia8/	29980.0
QCD_HT700to1000_TuneCP5_13TeV-madgraphMLM-pythia8/	6334.0
ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-madgraph-pythia8/	3.74
ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/	80.95
ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/	136.02
ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/	35.85
ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8/	19.56
ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/	35.85
ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8/	19.56
TTToHadronic_TuneCP5_13TeV-powheg-pythia8/	377.96
TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/	88.29
TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/	365.34
WBJetsToLNu_Wpt-100to200_TuneCP5_13TeV-madgraphMLM-pythia8/	5.527
WBJetsToLNu_Wpt-200toInf_TuneCP5_13TeV-madgraphMLM-pythia8/	0.7996
WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8/	1392.0
WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8/	1.084
WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8/	410.3
WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8/	0.008067
WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8/	57.85
WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8/	12.95
WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8/	5.45
WJetsToLNu_BGenFilter_Wpt-100to200_TuneCP5_13TeV-madgraphMLM-pythia8/	20.49
WJetsToLNu_BGenFilter_Wpt-200toInf_TuneCP5_13TeV-madgraphMLM-pythia8/	2.935
WJetsToLNu_HT-70To100_TuneCP5_13TeV-madgraphMLM-pythia8/	1353.0
WJetsToLNu_Pt-50To100_TuneCP5_13TeV-amcatnloFXFX-pythia8	3570.0
JetsToLNu_Pt-100To250_TuneCP5_13TeV-amcatnloFXFX-pythia8	770.8
WJetsToLNu_Pt-250To400_TuneCP5_13TeV-amcatnloFXFX-pythia8	28.06
WJetsToLNu_Pt-400To600_TuneCP5_13TeV-amcatnloFXFX-pythia8	3.591

Sample Name	Xsec (pb)
WJetsToLNu_Pt-600ToInf_TuneCP5_13TeV-amcatnloFXFX-pythia8	0.5495
WJetsToLNu_0J_TuneCP5_13TeV-amcatnloFXFX-pythia8	54500.0
WJetsToLNu_1J_TuneCP5_13TeV-amcatnloFXFX-pythia8	8750.0
WJetsToLNu_2J_TuneCP5_13TeV-amcatnloFXFX-pythia8	3010.0
WminusH_HToBB_WToLNu_M125_13TeV_powheg_pythia8/	0.10899
WplusH_HToBB_WToLNu_M125_13TeV_powheg_pythia8/	0.17202
ZBJetsToNuNu_Zpt-100to200_TuneCP5_13TeV-madgraphMLM-pythia8	6.195
ZBJetsToNuNu_Zpt-200toInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.6293
ZH_HToBB_ZToLL_M125_13TeV_powheg_pythia8/	0.04718
ZH_HToBB_ZToNuNu_M125_13TeV_powheg_pythia8/	0.09322
ZJetsToNuNu_BGenFilter_Zpt-100to200_TuneCP5_13TeV-madgraphMLM-pythia8	1.679
ZJetsToNuNu_BGenFilter_Zpt-200toInf_TuneCP5_13TeV-madgraphMLM-pythia8	0.2468
ZJetsToNuNu_HT-100To200_13TeV-madgraph	303.4
ZJetsToNuNu_HT-1200To2500_13TeV-madgraph	0.3425
ZJetsToNuNu_HT-200To400_13TeV-madgraph	91.71
ZJetsToNuNu_HT-2500ToInf_13TeV-madgraph	0.005263
ZJetsToNuNu_HT-400To600_13TeV-madgraph	13.1
ZJetsToNuNu_HT-600To800_13TeV-madgraph	3.248
ZJetsToNuNu_HT-800To1200_13TeV-madgraph	1.496
Z1JetsToNuNu_M-50_LHEZpT_50-150_TuneCP5_13TeV-amcnloFXFX-pythia8	596.3
Z2JetsToNuNu_M-50_LHEZpT_50-150_TuneCP5_13TeV-amcnloFXFX-pythia8	325.7
Z1JetsToNuNu_M-50_LHEZpT_150-250_TuneCP5_13TeV-amcnloFXFX-pythia8	17.98
Z2JetsToNuNu_M-50_LHEZpT_150-250_TuneCP5_13TeV-amcnloFXFX-pythia8	29.76
Z1JetsToNuNu_M-50_LHEZpT_250-400_TuneCP5_13TeV-amcnloFXFX-pythia8	2.045
Z2JetsToNuNu_M-50_LHEZpT_250-400_TuneCP5_13TeV-amcnloFXFX-pythia8	5.166
Z1JetsToNuNu_M-50_LHEZpT_400-inf_TuneCP5_13TeV-amcnloFXFX-pythia8	0.2243
Z2JetsToNuNu_M-50_LHEZpT_400-inf_TuneCP5_13TeV-amcnloFXFX-pythia8	0.8457
ggZH_HToBB_ZToLL_M125_13TeV_powheg_pythia8/	0.01437
ggZH_HToBB_ZToNuNu_M125_13TeV_powheg_pythia8/	0.01437

Table 30: Simulation samples used in 2018 with their cross section.

B Declaration of personal contribution

For the Simplified Template Cross Section (STXS) measurements of the VH($\rightarrow$ b$\bar{\text{b}}$) channel, I was one of the lead researchers. My contribution includes the end-to-end analysis of the 2017 and 2018 datasets and the combination with the 2016 analysis to achieve the results for the Run 2 (2016-2018) datasets. I have extensively contributed to all the steps of analysis: end-to-end processing of simulation and data samples, checking the modelling of the MC, improving the modelling by introducing corrections, training and evaluation of MVAs, designing the fit model, and extraction of the results. These steps were performed for both the VH($\rightarrow$ b$\bar{\text{b}}$) analysis as well as the cross-check VZ($\rightarrow$ b$\bar{\text{b}}$) analysis. I want to emphasize my studies on the fit model improvement by introducing additional degrees of freedom in the model to cover mis-modelling of electron and muon analysis regions which led to significant improvement in the Goodness of Fit of the model. This was one of the major road-block for the team to get the go-ahead with the unblinding from the CMS Collaboration. After having noticed the bug in the LO V+jets samples of 2017 and 2018 after the preapproval, I was one of the early researchers of the team who started exploring the usage of the NLO V+jets samples. The move from the LO to NLO came with its own expected issues such as MC modelling, and limited statistics. I played a major role in tackling these issues as well as laying the foundation of the NLO V+jets samples as the new baseline simulation of the analysis on which the final results are based. I have also performed an extensive number of tests to answer the questions asked during pre-approval to the way to the final analysis publication. These tests further helped to establish the robustness of the analysis as well as introduced newer features. In summary, my contributions were essential for the CMS Collaboration's first STXS measurement and observation of the VH($\rightarrow$ b$\bar{\text{b}}$) process. The correspoding results are accepted by the Physics Review D journal (arXiv:2312.07562).

For the efficiency parameterization of the b-tagging classifier (DeepCSV in this thesis) using the Graph Neural Network, I played a major role in the improvement of the existing model architecture based on the physics principle which led to significant improvement in the prediction of the per-event efficiency weights. The results were published in the CMS Detector Performance Note (CMS-DP-2022-051).

All text in this thesis is my own. The presented Figures contain my own research output unless specified in the caption.

References

[1] CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30–61, September 2012.

[2] ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, September 2012.

[3] E. Noether. Invariante variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1918:235–257, 1918.

[4] Gianluca Bianco. Study of the quantum interference between singly and doubly resonant top-quark production in proton-proton collisions at the LHC with the ATLAS detector. December 2020.

[5] Y. et al. Fukuda. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, August 1998.

[6] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508–509, October 1964.

[7] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, August 1964.

[8] LHC Higgs Working Group. Higgs cross sections and decay branching ratios. `https://twiki.cern.ch/twiki/bin/view/LHCPhysics/HiggsXSBR`.

[9] CMS Collaboration. A portrait of the Higgs boson by the CMS experiment ten years after the discovery. *Nature*, 607(7917):60–68, July 2022.

[10] M. Renz. *Erste Messung des Wirkungsquerschnitts der Top-Quark-Paarproduktion bei $\sqrt{s} = 7$ TeV im Elektron+Jets Kanal mit dem CMS-Experiment.* Phd thesis, Karlsruhe Institute of Technology, 2011. CERN-THESIS-2011-192.

[11] John C. Collins and Davison E. Soper and George Sterman. Factorization of Hard Processes in QCD, 2004.

[12] W.M. Alberico, S.M. Bilenky, and C. Maieron. Strangeness in the nucleon: neutrino–nucleon and polarized electron–nucleon scattering. *Physics Reports*, 358(4):227–308, March 2002.

[13] Particle Data Group. Review of Particle Physics. *Phys. Rev. D*, 98:030001, August 2018.

[14] Andy et al. Buckley. General-purpose event generators for LHC physics. *Physics Reports*, 504(5):145–233, July 2011.

[15] Christian Bierlich, Smita Chakraborty, Nishita Desai, Leif Gellersen, Ilkka Helenius, Philip Ilten, Leif Lönnblad, Stephen Mrenna, Stefan Prestel, Christian T. Preuss, Torbjörn Sjöstrand, Peter Skands, Marius Utheim, and Rob Verheyen. A comprehensive guide to the physics and usage of PYTHIA 8.3, 2022.

[16] Francesco Knechtli. Lattice Quantum Chromodynamics. *arXiv:1706.00282*, 2017.

[17] CMS Collaboration. Event generator tunes obtained from underlying event and multiparton scattering measurements. *The European Physical Journal C*, 76(3), March 2016.

[18] CMS Collaboration. Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements. *The European Physical Journal C*, 80(1), January 2020.

[19] ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.

[20] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[21] ALICE Collaboration. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008.

[22] LHCb Collaboration. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008.

[23] TOTEM Collaboration. The TOTEM experiment at the CERN Large Hadron Collider. *JINST*, 3:S08007, 2008.

[24] LHCf Collaboration. The LHCf detector at the CERN Large Hadron Collider. *Journal of Instrumentation*, 3(08):S08006, August 2008.

[25] CERN. The accelerator complex. `https://home.cern/science/accelerator-complex`. Accessed: 2-05-2024.

[26] CERN. Accelerators. `https://home.cern/science/accelerators`, Accessed: 2-05-2024.

[27] CMS Collaboration. Public CMS Luminosity Information. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`.

[28] Oliver Brüning et al. The scientific potential and technological challenges of the High Luminosity Large Hadron Collider program, Rep. Prog. Phys. 85:046201, 2022.

[29] CERN. High Luminosity LHC Project. `https://hilumilhc.web.cern.ch/content/hl-lhc-project`.

[30] Adam W. et al. The CMS Phase-1 Pixel Detector Upgrade. *JINST*, 16(02):P02027, 2021.

[31] CMS Collaboration. CMS Physics Technical Design Report. https://cds.cern.ch/record/922757/files/lhcc-2006-001.pdf, February 2006.

[32] CMS Collaboration. Development of the CMS detector for the CERN LHC Run 3. *arXiv:2309.05466*, 2023.

[33] B. P. Solano. *Introduction to Silicon Trackers at LHC.* Universitat Politècnica de Catalunya · Barcelona Tech.

[34] R. Brenner. Status of development of technology for wireless data transfer in future tracking detectors. *PoS(Vertex 2016)*, August 2017.

[35] W. et al. Adam. The CMS Phase-1 Pixel Detector Upgrade. *JINST*, 16(02):P02027, 2021.

[36] CMS Collaboration. Technical proposal for the upgrade of the CMS detector through 2020, CERN-LHCC-2011-006, CMS-UG-TP-1, LHCC-P-004. Technical report, 2011.

[37] Maren Tabea Meinhard. *Performance of detector modules for the CMS Pixel Phase 1 upgrade and search for ttH, H→bb events in the boosted regime.* PhD thesis, ETH Zurich, 2021.

[38] CMS Collaboration. The Electromagnetic Calorimeter Technical Design Report. https://cds.cern.ch/record/349375/files/ECAL_TDR.pdf, December 1997.

[39] CMS Collaboration. Energy calibration and resolution of the CMS electromagnetic calorimeter in pp collisions at $\sqrt{s} = 7$ TeV. *Journal of Instrumentation*, 8(09):P09009–P09009, September 2013.

[40] CMS Collaboration. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *Journal of Instrumentation*, 10(06):P06005–P06005, June 2015.

[41] Victor Daniel Elvira. Measurement of the Pion Energy Response and Resolution in the CMS HCAL Test Beam 2002 Experiment, CMS-NOTE-2004-020. Technical report, CERN, Geneva, 2004.

[42] John Douglas Cockcroft. Experimental Nuclear Physics. *Nature*, 175:53–54, 1955.

[43] Univeristy of Freiburg. Interaction of Charged Particles with Matter. https://www.particles.uni-freiburg.de/dateien/vorlesungsdateien/expmethfiles/kap2. Accessed: 3-05-2024.

[44] A.M. Sirunyan et al. Performance of the reconstruction and identification of high-momentum muons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Journal of Instrumentation*, 15(02):P02027, February 2020.

[45] CMS Collaboration. The Phase-2 Upgrade of the CMS Level-1 Trigger. `https://cds.cern.ch/record/2283192/files/CMS-TDR-017.pdf`, September 2017.

[46] Izaak Neutelings. CMS coordinate system. `https://tikz.net/axis3d_cms/`. Accessed: 09-02-2024.

[47] A.M. et al. Sirunyan. Observation of Higgs Boson Decay to Bottom Quarks. *Physical Review Letters*, 121(12), September 2018.

[48] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7), July 2014.

[49] R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H.-S. Shao, and M. Zaro. The automation of next-to-leading order electroweak calculations. *Journal of High Energy Physics*, 2018(7), July 2018.

[50] Andrea Banfi, Silvia Ferrario Ravasio, Barbara Jäger, Alexander Karlberg, Felix Reichenbach, and Giulia Zanderighi. A POWHEG generator for deep inelastic scattering. *Journal of High Energy Physics*, 2024(2), February 2024.

[51] Stefano Frixione, Paolo Nason, and Bryan R Webber. Matching NLO QCD and parton showers in heavy flavour production. *Journal of High Energy Physics*, 2003(08):007–007, August 2003.

[52] Fabio Maltoni, Center for Cosmology, Particle Physics and Phenomenology (CP3) Université Catholique de Louvain. Basics of QCD - AEPSHEP. `https://indico.cern.ch/event/492098/contributions/1169736/attachments/1356309/2050123/QCD2.pdf`, October 2016.

[53] Stefan Hoeche, Frank Krauss, Nils Lavesson, Leif Lonnblad, Michelangelo Mangano, Andreas Schaelicke, and Steffen Schumann. Matching Parton Showers and Matrix Elements. *arXiv:hep-ph/0602031*, 2006.

[54] Rikkert Frederix and Stefano Frixione. Merging meets matching in MC@NLO. *Journal of High Energy Physics*, 2012(12), December 2012.

[55] S. Agostinelli et al. Geant4: a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.

[56] Richard D. Ball, Stefano Carrazza, Luigi Del Debbio, Stefano Forte, Zahari Kassabov, Juan Rojo, Emma Slade, and Maria Ubiali. Precision determination of the strong coupling constant within a global PDF analysis: NNPDF Collaboration. *The European Physical Journal C*, 78(5), May 2018.

[57] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Computer Physics Communications*, 191:159–177, June 2015.

[58] Giancarlo Ferrera, Massimiliano Grazzini, and Francesco Tramontano. Associated Higgs-W-Boson Production at Hadron Colliders: A Fully Exclusive QCD Calculation at NNLO. *Phys. Rev. Lett.*, 107:152003, October 2011.

[59] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. MINLO: Multi-scale improved NLO. *JHEP*, 10:155, 2012.

[60] Long Chen, Joshua Davies, Gudrun Heinrich, Stephen P. Jones, Matthias Kerner, Go Mishima, Johannes Schlenk, and Matthias Steinhauser. ZH production in gluon fusion at NLO in QCD. *Journal of High Energy Physics*, 2022(8), August 2022.

[61] LHC Higgs Working Group. CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector, 2017.

[62] Andreas Albert and Jing Peng Kevin Stenson Siqi Yuan Material by: AA, Kenneth Long. The pdfwgt setting in MG 2.4.2 and its effect on 2017/18 LO V samples. `https://indico.cern.ch/event/841559/contributions/3547525/attachments/1901143/3139399/2019-08-28_vht_comparison_forhn.pdf`, 2019. Material by: AA, Kenneth Long, Jing Peng, Kevin Stenson, Siqi Yuan.

[63] Tomislav Seva and Henry Yee-Shian Tong. Standard Model Cross Sections for CMS at 13 TeV. `https://twiki.cern.ch/twiki/bin/viewauth/CMS/StandardModelCrossSectionsat13TeV?rev=28`, 2022.

[64] Chris Palmer. Di-boson Cross Sections. `https://indico.cern.ch/event/904979/contributions/3856716/attachments/2035384/3409110/DibosonXSections_Run2Legacy_May8.pdf`, 2022. Association: Princeton University/FNAL.

[65] A.M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12(10):P10003, October 2017.

[66] CMS Collaboration. Description and performance of track and primary-vertex iggs cross sections and decay branching ratios with the CMS tracker. *Journal of Instrumentation*, 9(10):P10009, October 2014.

[67] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, March 1960.

[68] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by Simulated Annealing. *Science*, 220(4598):671–680, 1983.

[69] Jonathan Shlomi, Sanmay Ganguly, Eilam Gross, Kyle Cranmer, Yaron Lipman, Hadar Serviansky, Haggai Maron, and Nimrod Segol. Secondary vertex finding in jets with neural networks. *The European Physical Journal C*, 81(6), June 2021.

[70] J.H. Kotecha and P.M. Djuric. Gaussian sum particle filtering. *IEEE Transactions on Signal Processing*, 51(10):2602–2612, 2003.

[71] CMS Collaboration. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *Journal of Instrumentation*, 16(05):P05014, May 2021.

[72] CMS Electron/gamma Physics Object Group. Recipes for Run 2 ID Criteria. Technical report, CERN.

[73] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-kt jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063, April 2008.

[74] CMS Collaboration. Determination of jet energy calibration and transverse momentum resolution in CMS. *Journal of Instrumentation*, 6(11):P11002–P11002, November 2011.

[75] Khachatryan V. a. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *Journal of Instrumentation*, 12(02):P02014–P02014, February 2017.

[76] PARTICLE BITES. Jets aren't just a game of tag anymore. `https://www.particlebites.com/?cat=61`, August 2016. Accessed: 2-05-2024.

[77] Antimo Cagnotta, Francesco Carnevali, and Agostino De Iorio. Machine Learning Applications for Jet Tagging in the CMS Experiment. *Applied Sciences*, 12(20), 2022.

[78] A.M. et al. Sirunyan. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV, CMS-BTV-16-002, CERN-EP-2017-326, CMS-BTV-16-002-004. *JINST*, 13(05):P05011, 2018.

[79] CMS Collaboration. Machine learning-based identification of highly Lorentz-boosted hadronically decaying particles at the CMS experiment, CMS-PAS-JME-18-002. Technical report, CERN, Geneva, 2019.

[80] CMS Collaboration. Jet algorithms performance in 13 TeV data, CMS-PAS-JME-16-003. Technical report, CERN, Geneva, 2017.

[81] CMS Collaboration. CMS Workbook: MET Analysis. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMetAnalysis`. Accessed: 3-05-2024.

[82] CMS Collaboration. Missing E_T Performance in CMS. *CMS Physics Analysis Summary*, (CMS PAS JME-07-001).

[83] CMS Collaboration. Missing transverse energy performance of the CMS detector. *Journal of Instrumentation*, 6(09):P09001, September 2011.

[84] Albert M Sirunyan et al. A Deep Neural Network for Simultaneous Estimation of b Jet Energy and Resolution. *Comput. Softw. Big Sci.*, 4(1):10, 2020.

[85] P. A. Zyla and et al. (Particle Data Group). QUARKS. *Progress of Theoretical and Experimental Physics*, 2020.

[86] RooBukinPdf class reference. `https://root.cern.ch/doc/master/classRooBukinPdf.html`. Accessed: 09-02-2024.

[87] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler. Soft drop. *Journal of High Energy Physics*, 2014(5), May 2014.

[88] Lagrange, Joseph Louis. *M'ecanique Analytique.* Cambridge University Press, Cambridge, England, February 2015.

[89] The CMS NanoAOD data tier. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookNanoAOD`. Accessed: 09-02-2024.

[90] Daniele Bertolini, Philip Harris, Matthew Low, and Nhan Tran. Pileup per particle identification. *Journal of High Energy Physics*, 2014(10), October 2014.

[91] Pirmin Berger. *Measurement of the standard model Higgs Boson decay to b-quarks in association with a vector boson decaying to leptons, and module qualification for the CMS Phase-1 barrel pixel detector.* PhD thesis, ETH Zurich, 2021.

[92] LHC Higgs Working Group. Fiducial and STXS. Technical report, 2020. rev. 20, 202009-11.

[93] C T Potter et al. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group. `http://cds.cern.ch/record/1559921`, 2013.

[94] CMS Collaboration. Public CMS Luminosity Information. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`.

[95] Kenneth Long David Morse, Seth Cooper. V+jets MadGraph5_aMC@NLO Comparisions. `https://indico.cern.ch/event/658253/contributions/2683676/attachments/1504862/2344638/20170807_VJets_GenMeeting.pdf`, 2017.

[96] L. Mastrolorenzo. VH(cc) Run-2 analysis: Investigation on $dR(jj)$ feature in NLO V+Jets. `https://indico.cern.ch/event/1039177/contributions/4453165/attachments/2284588/3882800/VHcc_20_07_21_dRjjIssueSummary_compressed.pdf`. SMP General Meeting, 20-07-2021.

[97] J. Olsen, N. Haubrich, C. Palmer, Y. Lai. nAddJet Reweighting Fits. `https://nihaubri.web.cern.ch/vhbb/VHbb_nAddJet_Reweighting_Fits.pdf`. Accessed: 2-05-2024.

[98] J. Olsen, N. Haubrich, C. Palmer, Y. Lai. nAddJet Reweighting in 2lep. `https://nihaubri.web.cern.ch/vhbb/VHbb_nAddJet_Reweighting_in_2lep.pdf`. Accessed: 2-05-2024.

[99] Ansgar Denner, Stefan Dittmaier, Stefan Kallweit, and Alexander Mück. Electroweak corrections to Higgs-strahlung off W/Z bosons at the Tevatron and the LHC with Hawk. *Journal of High Energy Physics*, 2012(3), March 2012.

[100] Oliver Brein, Robert V. Harlander, and Tom J.E. Zirke. vh@nnlo—Higgs Strahlung at hadron colliders. *Computer Physics Communications*, 184(3):998–1003, March 2013.

[101] CMS Jet and Missing Energy Results. `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsJME`. Accessed: 09-02-2024.

[102] Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv:1803.08375*, 2018.

[103] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980*, 2017.

[104] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958, January 2014.

[105] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *International Conference on Artificial Intelligence and Statistics*. 2010.

[106] ROOT Reference Guide. TMVA::MethodBDT Class Reference. `https://root.cern.ch/doc/master/classTMVA_1_1MethodBDT.html`. ROOT Reference guide.

[107] J. Neyman and E. S. Pearson. On the Problem of the Most Efficient Tests of Statistical Hypotheses. In Samuel Kotz and Norman L. Johnson, editors, *Breakthroughs in Statistics: Foundations and Basic Theory*, pages 73–108. Springer New York, New York, NY, 1992.

[108] P. B. Patnaik. The Non-Central χ^2- and F-Distribution and their Applications. *Biometrika*, 36(1/2):202–232, 1949.

[109] LHC Higgs Working Group. *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*. CERN Yellow Reports: Monographs. CERN, Geneva, 2012.

[110] LHC Higgs Working Group. *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*. CERN Yellow Reports: Monographs. CERN, Geneva, 2011.

[111] LHC Higgs Working Group. *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*. CERN Yellow Reports: Monographs. 2013.

[112] ATLAS Collaboration. Evaluation of theoretical uncertainties for simplified template cross section measurements of $V-$associated production of the Higgs boson (ATL-PHYS-PUB-2018-035). Technical report, CERN, Geneva, 2018.

[113] A. de Wit and A. Nigamova. STXS migration uncertainties for VH. `https://indico.cern.ch/event/904970/contributions/3843849/attachments/2029052/3395392/STXSUnc2904.pdf`, April 2020.

[114] Richard D. et al. Parton distributions for the LHC run II. *Journal of High Energy Physics*, 2015(4), April 2015.

[115] LHC Higgs Working Group. CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector. `https://e-publishing.cern.ch/index.php/CYRM/issue/view/32`, 2017.

[116] A. Calandri. Diboson impact. `https://indico.cern.ch/event/1233721/contributions/5290376/attachments/2600948/4494551/VVnorm_check6.pdf`. Accessed: 2-05-2024.

[117] Ayala E. et al. The Pixel Luminosity Telescope: a detector for luminosity measurement at CMS using silicon pixel sensors. *The European Physical Journal C*, 83(7), July 2023.

[118] CMS Collaboration. CMS Luminosity Measurements for the 2016 Data Taking Period, CMS-PAS-LUM-17-001. Technical report, CERN, Geneva, 2017.

[119] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s}$ = 13 TeV, CMS-PAS-LUM-18-002. Technical report, CERN, Geneva, 2019.

[120] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$TeV. Technical report, CERN, Geneva, 2018.

[121] CMS Collaboration. Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV. *Journal of High Energy Physics*, 2018(7), July 2018.

[122] J. S. Conway. Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra. *arXiv:1103.0354*, 2011.

[123] Khachatryan V. et al. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *Journal of Instrumentation*, 12(02):P02014–P02014, February 2017.

[124] C.Palmer. Single Process SFs with VPT Category Migration (in TT). https://indico.cern.ch/event/922781/contributions/3957069/attachments/2078751/3491209/SFwithShape_TT_July23.pdf. Accessed: 2-05-2024.

[125] Robert D. Cousins. Generalization of Chisquare Goodness-of-Fit Test for Binned Data Using Saturated Models, with Application to Histograms. *arXiv:0802.0041*, 2008.

[126] C.Palmer. V+Jets (+b/bb) in 1- and 2-lepton channels. https://indico.cern.ch/event/1233719/contributions/5270876/attachments/2592831/4474953/VHbb_VJetsPlusBs_13Feb2023.pdf, February 2023. Accessed: 2-05-2024.

[127] R.Mankel, A.Nigamova, H.Kaveh. Systematic unc. study. https://indico.cern.ch/event/1109111/contributions/4727029/attachments/2386225/4078479/20-001vs18-016_update_07022022.pdf, July 2023. Accessed: 2-05-2024.

[128] Patrick Royston. Lowess Smoothing. *Stata Technical Bulletin*, 1, February 1992.

[129] Frank J. Massey. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, 46(253):68–78, 1951.

[130] CMS Collaboration. Observation of Higgs boson decay to bottom quarks, CMS-PAS-HIG-18-016. Technical report, CERN, Geneva, 2018.

[131] CMS Collaboration. Efficiency parametrization of b-tagging classifier using Graph Neural Networks, CMS-DP-2022-051. https://cds.cern.ch/record/2839921, 2022.

[132] Francesco Bello, Jonathan Shlomi, Chiara Badiali, Guglielmo Frattari, Eilam Gross, Valerio Ippolito, and Marumi Kado. Efficiency Parameterization with Neural Networks. *Computing and Software for Big Science*, 5, December 2021.

[133] Ralf Wischnewski. Errors in weighted histograms. https://www.zeuthen.desy.de/~wischnew/amanda/discussion/wgterror/working.html, August 2000. Accessed: 27-04-2024.

[134] Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant. Graph neural networks in particle physics. *Machine Learning: Science and Technology*, 2(2):021001, December 2020.

[135] Bengio, Yoshua and Ducharme, R'ejean and Vincent, Pascal and Janvin, Christian. A neural probabilistic language model. *J. Mach. Learn. Res.*, 3(null):1137–1155, mar 2003.

[136] Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks? In *International Conference on Learning Representations*. 2022.

[137] Andrew L. Maas. Rectifier Nonlinearities Improve Neural Network Acoustic Models. https://api.semanticscholar.org/CorpusID:16489696, 2013.

[138] Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, August 1996.

[139] E. Bols and J. Kieseler and M. Verzetti and M. Stoye and A. Stakia. Jet flavour classification using DeepJet. 15(12):P12012, December 2020.

[140] CMS Collaboration. Transformer models for heavy flavor jet identification, CMS DP -2022/050. https://cds.cern.ch/record/2839920/files/DP2022_050.pdf, 2022.

[141] LeCun, Yann and Bengio, Yoshua and Hinton, Geoffrey. Deep learning. *Nature*, 521(7553):436–444, May 2015.

[142] David E. Rumelhart and Geoffrey E. Hinton and Ronald J. Williams. Learning internal representations by error propagation. 1986.

[143] Huilin Qu and Congqiao Li and Sitian Qian. Particle Transformer for Jet Tagging. 2024.

[144] Qu, Huilin and Gouskos, Loukas. Jet tagging via particle clouds. *Physical Review D*, 101(5), March 2020.

[145] CMS Collaboration. Identification of highly Lorentz-boosted heavy particles using graph neural networks and new mass decorrelation techniques, CMS DP -2020/002. https://cds.cern.ch/record/2707946/files/DP2020_002.pdf, 2020.

[146] Dolen, James and Harris, Philip and Marzani, Simone and Rappoccio, Salvatore and Tran, Nhan. Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure. *Journal of High Energy Physics, publisher=Springer Science and Business Media LLC*, 2016(5), May 2016.

Acknowledgements

Firstly, I would like to thank Rainer Wallny for accepting me as his PhD student. Apart from ensuring a respectful work environment and a healthy work-life balance, he ensured students could do research without major financial constraints. For example, my efficiency parametrization study was a new addition outside my Research Plan for which I got his full support. Further, his motivation for his students to do well both in academia and industries is very appreciable. Secondly, I would like to thank Christoph Grab for accepting the role of co-supervisor. This humble nature and ever-lasting love for physics are very inspiring for young students like me. I enjoyed the meetings and insights from him, especially when the in-person meetings were scheduled before the Covid-19 lockdown.

I am thankful to Günther Dissertori for accepting me as his master student where I was exposed to the excellent research done by the ETH group which motivated me to continue further with my doctoral studies. I am also grateful to Mauro Donegà for being an excellent mentor, and teacher with a down-to-earth personality. I am inspired by his all-round skills, and being available for every student. I am thankful to him for having spotted me during my master's degree program, guiding me in an exciting master's thesis, and further recommending me for doctoral studies in the ETH group. He played a key role in my scientific career by connecting the dots. I am also profoundly thankful to Alessandro Calandri, one of the calmest people I have worked with. Apart from his soft personality and mentorship, I am inspired by the endless dedication and commitment by which he works in the team. Our $\mathrm{VH} \rightarrow \mathrm{b\bar{b}}$ journey tested patience to high levels and I never saw him losing patience at any stage, working with the same focus all along the way.

I am also very thankful to Christina Reissel, my fellow PhD student with whom I collaborated the most in my doctoral journey. Apart from being an incredible physicist, she has a very joyful personality adding positivity to the team atmosphere. I am confident she would be an invaluable asset to any team she joins. It was a pleasure working with her. I would also like to thank Kaustav Dutta, another fellow PhD student for being a wonderful colleague and being there for me whenever I wanted to discuss anything, from physics to day-to-day life. He is one of the sweetest people I have come across. I would also like to thank Pirmin Berger, another fellow PhD student with whom I collaborated. An incredibly hard-working, knowledgeable, and grounded individual with whom I have worked. I learned a lot from him during the initial days of my PhD. I am also thankful to Maren Meinhard, from whom I learned more about detectors during my short collaboration with her. I would also like to thank Vasilije Perovic, Matteo Marchegiani, and Franz Glessgen, my neighboring office mates, for all the regular fun chats I had with them.

Lastly, I would like to thank my family and friends for their unwavering support without whom I wouldn't have reached up to this stage of my life. I remain thankful to my lovely parents and brother who always encouraged me to follow my passion, took pride in my success, and were there every time I needed them.