Measurement of Higgs boson production in association with a vector boson and decay to b-quarks in the Simplified Template Cross Section framework and interpretations with the Standard Model effective field theory

Dissertation

zur Erlangung des Doktorgrades

an der Fakultät für Mathematik, Informatik und Naturwissenschaften

Fachbereich Physik

der Universität Hamburg

vorgelegt von

Aliya Nigamova

Hamburg

2022

Eidesstattliche Erklärung

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Die Dissertation wurde in der vorgelegten oder einer ähnlichen Form nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

Hamburg, den December 5, 2022.

Aliya Nigamova

Gutachter der Dissertation:	Prof. Dr. Elisabetta Gallo
	Dr. Rainer Mankel
Zusammensetzung der Prüfungskommission:	Prof. Dr. Elisabetta Gallo
	Dr. Rainer Mankel
	Prof. Dr. Florian Gruener
	Prof. Dr. Johannes Haller
	Dr. Frank Tackmann
Vorsitzender des Prüfungsausschusses:	Prof. Dr. Florian Gruener
Datum der Disputation:	7 September 2022
Vorsitzender des Fach-Promotionsausschusses PHYSIK:	Prof. Dr. Wolfgang J. Parak
Leiter des Fachbereichs PHYSIK:	Prof. Dr. Günter H. W. Sigl
Dekan der MIN-Fakultät:	Prof. Dr.-Ing. Norbert Ritter

Zusammenfassung

Diese Dissertation fasst die Querschnittsmessung des Standardmodell-Higgs-Bosons zusammen, das in ein Bottom-Quark-Paar zerfällt und in Assoziation mit einem Vektorboson produziert wird. Die Messung wird unter Verwendung der während Run 2 gesammelten CMS-Daten mit der integrierten Luminositaet von 138 fb^{-1} durchgeführt. Der Wirkungsquerschnitt wird in den Simplified Template Cross Sections (STXS) Intervallen der Stufe 1.2 gemessen und in Bezug auf die Operatoren der effektiven Feldtheorie des Standardmodells interpretiert. Sowohl die inklusiven als auch die STXS-Messungen sind mit den Vorhersagen des Standardmodells kompatibel.

Abstract

This thesis summarises the cross-section measurement of a Higgs boson decaying into a bottom quark pair produced in association with a vector boson. The measurement is performed using the CMS data collected during Run 2 with the integrated luminosity of 138 fb^{-1}. The cross section is measured in Simplified Template Cross Section stage 1.2 bins and interpreted in terms of Standard Model Effective Field Theory operators. The inclusive and STXS measurements are both compatible with the Standard Model predictions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

Introduction

On a long quest of understanding the fundamental structure of the universe there have been many developments in the past decades. The Standard Model (SM) of elementary particle physics is the best description of the fundamental interactions. It was developed in the second part of the 20th century and supported by many experimental observations with remarkable precision, which exceeded all expectations.

Ten years ago on the 4th of July 2012, the ATLAS and CMS experiments operating at the Large Hadron Collider (LHC) announced the discovery of a particle compatible with the SM Higgs Boson. This particle plays a crucial role in the spontaneous symmetry breaking (SSB) in the SM. Through the SSB mechanism, the fundamental particles of the SM obtain their masses. The observation of the Higgs boson was followed by many detailed measurements of its properties. And as of now the scientific community is fairly confident that the observed scalar particle with the mass of 125 GeV is indeed the Higgs boson of the SM.

However, there are many questions that the SM does not provide answers for. For example, we still do not know the nature of the dark matter; gravity is not a part of the SM; the matter-antimatter asymmetry is not explained. These are strong indications that the current understanding of fundamental physics is not complete. Curious minds have been looking for a solution for many decades. The theorists have proposed many compelling theories such as supersymmetry (SUSY), which manages to address all of the issues discussed above and in addition predicts new particles. Direct searches for BSM particles have not been successful so far. This motivated a program that targets interpretations of precision measurements at the LHC using model-independent approaches such as Standard Model Effective Field Theory (SMEFT). It provides a possibility to search for BSM effects in a larger phase space in a model-independent way. The SMEFT interpretations can always be matched with specific BSM models. To succeed with the EFT program, the experiments should be able to provide differential measurements that can be easily combined and (re)interpreted. The Simplified Template Cross Section (STXS) framework was designed to provide this possibility.

In this thesis the STXS measurement of the Higgs boson production in association with the vector boson with the Higgs boson decaying to $b\bar{b}$ ($VH(\to b\bar{b})$) performed using Run 2 data collected by the CMS collaboration will be presented in detail. In Chapter 2 a theoretical introduction into the SM physics will be given and the EFT approach will be discussed. In Chapter 3 the CMS experiment and LHC are described and the features of the objects reconstruction at CMS are discussed. The next Chapter 4 is devoted to the reconstruction of the $VH(\to b\bar{b})$ final state. The STXS analysis strategy is formulated in Chapter 5, followed by the statistical inference procedure given in Chapter 6 and the results discussed in Chapter 7. The second to last Chapter 8 describes the interpretation analysis performed using the STXS $VH(\to b\bar{b})$ results. The thesis closes with a summary.

CHAPTER 2

Theoretical overview

2.1. Introduction

The Standard Model (SM) of elementary particle physics is a relativistic quantum field theory (QFT) that describes the particle content of the universe and the fundamental interactions. It was established in the 1960s [1, 2, 3], and provides essential predictions that are being supported by the experimental measurements with constantly increasing precision.

This chapter will lay down a theoretical basis for the measurements presented in this thesis. First, the particle content will be introduced, followed by the discussion of gauge symmetries and arising from them interactions. Gathering all pieces together, the SM Lagrangian will be introduced. We will go through the observations that can not be explained by the SM, and discuss the Effective Field Theory approach of exploring Beside Standard Model (BSM) physics. Lastly, we will focus on the Higgs boson production and decay processes.

2.2. The Standard Model of particle physics

2.2.1. Particle Content

The fields in the SM can be split into three categories: spin-1/2 particles (fermions), spin-1 particles (gauge bosons) and spin-0 (the Higgs boson). The first category encompasses the particles that form matter in the universe, including three generations of leptons: electron and electron neutrino, muon and muon-neutrino, and the tau and tau-neutrino; and three generations of quarks with the up and down, charm and strange, and top and bottom quarks. For each of these particles there is an anti-particle. The quarks carry the color charge: red, blue or green. The gauge bosons $W^{\pm}$ (Z), γ and gluon, mediate the weak, electromagnetic and strong interactions respectively. The only elementary spin-0 particle in the SM, the Higgs boson, plays a crucial role: the massive gauge bosons and fermions obtain their masses through the interaction with the Higgs field.

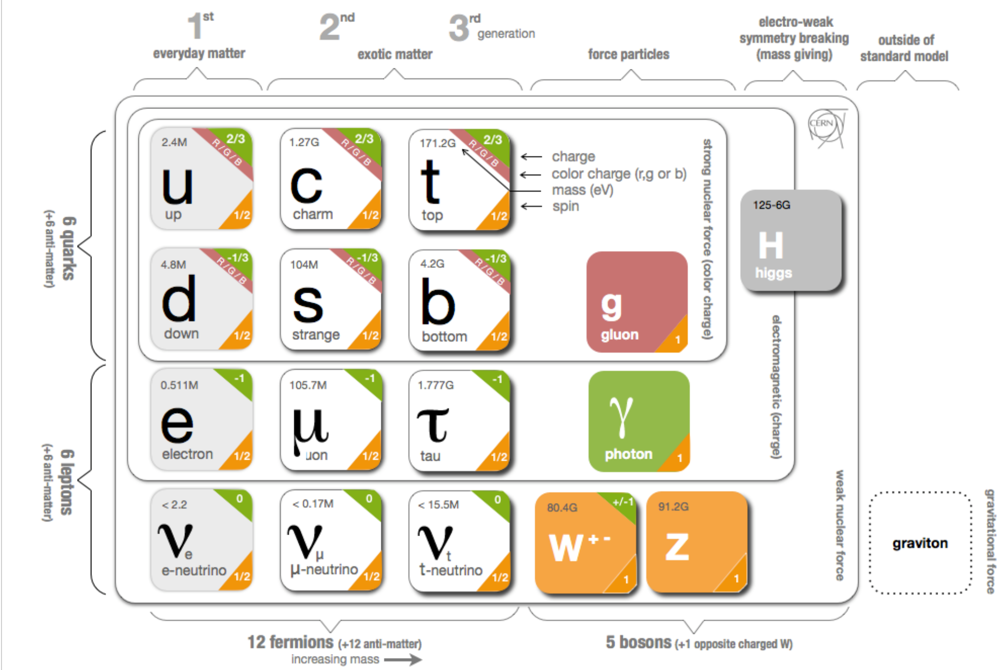


Figure 2.1: The SM particle content.

2.2.2. Symmetries and Interactions

Despite the enormous complexity and large number of degrees of freedom, the SM is so far the most elegant QFT. The SM is formulated in terms of the Lagrangian formalism. All of the interactions in the SM are governed by the principle of symmetries, which has been the driving mechanism in theoretical particle physics since the famous theorem by Emmy Noether [4], which stated that a symmetry gives rise to a quantity that should be conserved. If we invert the theorem, we find that every conserved quantity corresponds to a symmetry of Lagrangian ($\mathcal{L}$). The SM model is a local gauge theory, in other words the $\mathcal{L}_{SM}$ is constructed to be invariant under local gauge transformations. This assumption introduces the gauge bosons and links each force to the corresponding symmetry.

QED

To explicitly demonstrate it, we can start with the U(1) symmetry for quantum electrodynamics (QED) and write down the Lagrangian[1] for Dirac spin-1/2 fields:

$$\mathcal{L}_{QED} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m_{\psi}\bar{\psi}\psi \tag{2.1}$$

It is easy to see that it is invariant under U(1) global transformations, but under U(1) local transformations of the form

$$\psi \rightarrow \psi' = \psi e^{iq\phi(x)}, \tag{2.2}$$

the QED Lagrangian transforms as follows:

$$\mathcal{L}_{QED} \rightarrow \mathcal{L}'_{QED} = i\bar{\psi}e^{-iq\phi(x)}\gamma^{\mu}\partial_{\mu}e^{iq\phi(x)}\psi - m_{\psi}\bar{\psi}\psi. \tag{2.3}$$

This is resolved if we promote the simple derivative to the covariant derivative $\partial_{\mu} \rightarrow D_{\mu} = \partial_{\mu} + iqA_{\mu}$, and define the transformation rule for the field $A_{\mu} \rightarrow A'_{\mu} = A_{\mu} - iq\partial_{\mu}\phi(x)$.

After the manipulations mentioned above the $\mathcal{L}_{QED}$ takes the following form:

$$\mathcal{L}_{QED} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - q\bar{\psi}\gamma^{\mu}A_{\mu}\psi - m_{\psi}\bar{\psi}\psi, \tag{2.4}$$

where the new term $q\bar{\psi}\gamma^{\mu}A_{\mu}\psi$ is precisely the interactions term, with the coupling strength q and vector field A_{μ}, which represents the photon field. To allow the A_{μ} to propagate we need to add a kinetic term $-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ which is gauge invariant for the abelian group U(1). There is no mass term for the A_{μ} field, because it would break the gauge invariance of our

[1] In this thesis Einstein notation is followed, i.e. when an index variable appears twice in a single term, it implies summation of that term over all the values of the index.

theory, therefore the photon is massless. The QED Lagrangian takes its final form:

$$\mathcal{L}_{QED} = i\bar{\psi}\gamma^{\mu}D_{\mu}\psi - m_{\psi}\bar{\psi}\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \tag{2.5}$$

SU(N)

The same discussion can be expanded to SU(N) groups to define all other interactions in the SM, which can be done following the same procedure as in the case of the U(1) QED Lagrangian. As it has been shown above, the covariant derivative should be introduced to achieve the local gauge invariance, therefore it is useful to write it in a general form for an arbitrary SU(N) theory:

$$D_{\mu} = \partial_{\mu} + igA^{a}_{\mu}T^{a}, \tag{2.6}$$

where g is the coupling strength, A^{i}_{μ} is the vector gauge field, and the T^{i} are the $n^2 - 1$ group generators. In case of SU(2) we have three generators which are the Pauli matrices, while for SU(3) — these eight generators are the Gell-Mann matrices. It can also be shown that the $F_{\mu\nu}$ can be expressed in terms of covariant derivative as follows:

$$\begin{aligned} F_{\mu\nu} = \frac{i}{g}[D_{\mu}, D_{\nu}] = \partial_{\mu}A^{a}_{\nu} - \partial_{\nu}A^{a}_{\mu} + ig[A^{b}_{\mu}T^{b}, A^{c}_{\nu}T^{c}] = \\ T^{a}(\partial_{\mu}A^{a}_{\nu} - \partial_{\nu}A^{a}_{\mu} - gf^{abc}A^{b}_{\mu}A^{c}_{\nu}) = T^{a}F^{a}_{\mu\nu}, \end{aligned} \tag{2.7}$$

here the f^{abc} are the structure constants defined by SU(N) Lie algebra: $[T^{a}, T^{b}] = if^{abc}T^{c}$. The f^{abc} constants essentially define the interaction structure, e.g. for the abelian group U(1) $f^{abc} = 0$, while non-zero f^{abc} are responsible for the gluon-gluon self interactions in SU(3) and the vector boson interactions in SU(2) as it will be shown later.

QCD

The strong interaction is governed by the SU(3) colour local symmetry and affects the fields with colour charge (quarks). To keep the QCD Lagrangian gauge invariant the general SU(N) strategy can be directly applied for the SU(3). The general covariant derivative defined for

SU(N) Eq. 2.6 is rewritten — the coupling strength g, vector field A^a_μ and generators T^a are replaced with the strong coupling constant g_s, the gluon fields G^a_μ and the 3×3 Gell-Mann matrices $T^a = \frac{t^a}{2}$, respectively. It is interesting to look at the kinetic term for the gluon fields:

$$\begin{aligned}\frac{1}{4}G^a_{\mu\nu}G^{a\mu\nu} = (\partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g_s f^{abc}G^b_\mu G^c_\nu)(\partial^\mu G^{\nu a} - \partial^\nu G^{\mu a} - g_s f^{ade}G^{\mu d}G^{\nu e}) = \\ (\partial_\mu G^a_\mu - \partial_\nu G^a_\nu)(\partial_\mu G^a_\mu - \partial_\nu G^a_\nu) \\ -g_s f^{ade}G^d_\mu G^e_\nu(\partial^\mu G^{\nu a} - \partial^\nu G^a_\mu) - g_s f^{abc}G^{\mu b}G^{\nu c}(\partial_\mu G^a_\nu - G^a_\mu) \\ +g_s^2 f^{abc}f^{ade}G^b_\mu G^c_\nu G^{\mu d}G^{\nu e},\end{aligned} \tag{2.8}$$

where in addition to the abelian kinetic term we also get the term responsible for the interaction between color-charged gluons, as illustrated in Figure 2.2.

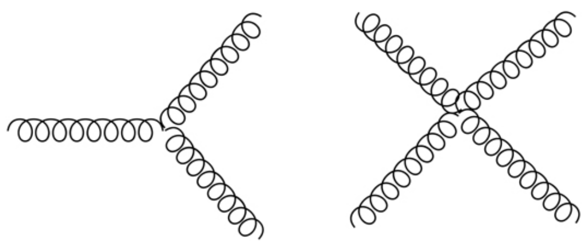


Figure 2.2: Gluon self interaction diagrams.

Electroweak unification

The electroweak (EWK) unification is one of the most important milestones of the SM development [3]. It generated crucial predictions to achieve the agreement with the experimental observations, such as the discovery of the weak neutral current [5], the parity violation in the weak interactions [6], and the observation of Z and $W^\pm$ bosons [7]. The EWK unification is achieved by requiring the $\mathcal{L}_{SM}$ to satisfy the $\mathrm{SU}(2)_L \times U(1)_y$ symmetry, where $\mathrm{SU}(2)_L$ implies that the gauge transformations of this symmetry affect only left particles. As indicated previously to keep the theory invariant under the assumed transformation the covariant derivative has to be defined:

$$D_\mu = \partial_\mu + igW_\mu^i \frac{\sigma^i}{2} + ig' B_\mu Y. \tag{2.9}$$

The $SU(2)_L \times U(1)_y$ group has 4 generators. In the non-abelian part $SU(2)_L$ we have three generators which are the Pauli spin-matrices σ^i with the corresponding conserved quantity which is the I_3 weak isospin. The abelian group $U(1)_y$ has one generator, with the conserved weak hypercharge Y. The weak isospin and hypercharge relate to the electric charge as

$$Q = I_3 + \frac{1}{2}Y, \tag{2.10}$$

which is the defining expression of electroweak unification.

The fields W_μ^i and B_μ mix to form the fields of gauge bosons' physical states as

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp W_\mu^2) \\ Z_\mu &= W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W \\ A_\mu &= W_\mu^3 \sin\theta_W + B_\mu \cos\theta_W, \end{aligned} \tag{2.11}$$

where the θ_W is the weak mixing angle, also called Weinberg angle, and the $W_\mu^\pm$, Z_μ, A_μ are the fields of the gauge bosons physical states $W^\pm$, Z, γ, respectively. Similar to the QCD gluon-gluon interaction, for the SU(2) due to its non-abelian nature we can also form the vertices with the bosons based on the interaction term of the EWK sector:

$$\mathcal{L}_{gauge}^{EW} = -\frac{1}{4} W^{i,\mu\nu} W_{\mu\nu}^i - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} \tag{2.12}$$

If we consider the physical states after applying the mixing we get interaction vertices as shown in Figure 2.3,

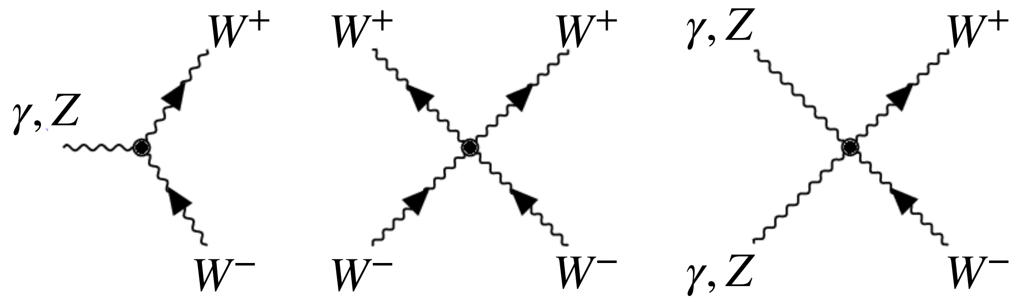


Figure 2.3: EWK self-interaction vertices.

The electroweak sector in the SM acts on all of the matter constituents (quarks and leptons). But there is one substantial feature of the weak force ($SU(2)_L$): it was experimentally observed that it only affects the left-handed fermions, therefore it is important to differentiate the left-handed and right-handed particles in the $\mathcal{L}_{SM}$. The fermion sector is constructed using the Weyl 4-component spinors which are projected into the right-handed (ψ_R) and left-handed (ψ_L) components using the $P_{L,R} = \frac{1}{2}(1 \pm \gamma_5)$ operators. Then electroweak interactions for the left-handed and right-handed fermions are represented by separate terms in the Lagrangian,

$$\begin{aligned} \mathcal{L}_{EW} &= \bar{\psi_L} i D^L_\mu \gamma^\mu \psi_L + \bar{\psi_R} i D^R_\mu \gamma^\mu \psi_R, \\ D^L_\mu &= \partial_\mu + igW^i_\mu \frac{\sigma^i}{2} + ig' B_\mu Y_L, \\ D^R_\mu &= \partial_\mu + ig' B_\mu Y_R. \end{aligned} \tag{2.13}$$

Therefore the left-handed fermions transform as doublets under the $SU(2)_L$, while the right-handed fermions transform as singlets. The separation of the Dirac fermions into the left and right component breaks the gauge symmetry of the Lagrangian. The next section addresses this issue through the spontaneous symmetry breaking described by Brout-Englert-Higgs (BEH) mechanism.

Brout-Englert-Higgs mechanism

Despite all of the success of electroweak unification there is still a major missing component in the definition of the SM theory at this point. As we assume the local gauge symmetry $SU(3)_c \times SU(2)_L \times U(1)_Y$, classical mass terms can not be introduced for the gauge bosons, which are massive, as observed experimentally. In this section we will discuss the BEH mechanism [8, 9, 10] that introduces the mass terms for EW bosons and fermions through spontaneous symmetry breaking (SSB). The first step is the introduction of a scalar $SU(2)_L$ doublet

$$\Phi = \begin{pmatrix} \phi^\dagger \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4, \end{pmatrix} \tag{2.14}$$

which adds four degrees of freedom to the theory

$$\mathcal{L}_{Higgs} = (D^\mu \Phi)^\dagger D_\mu \Phi - V(\Phi), \tag{2.15}$$

where D_μ is the covariant derivative defined for $SU(2)_L \times U(1)_Y$ and $V(\Phi)$ is the Higgs field potential, assumed to be of the form:

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \frac{1}{4}\lambda(\Phi^\dagger \Phi)^2. \tag{2.16}$$

To simplify the discussion it is useful to consider the behaviour of the Higgs potential for a scalar singlet $\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$, corresponding to a U(1) symmetry.

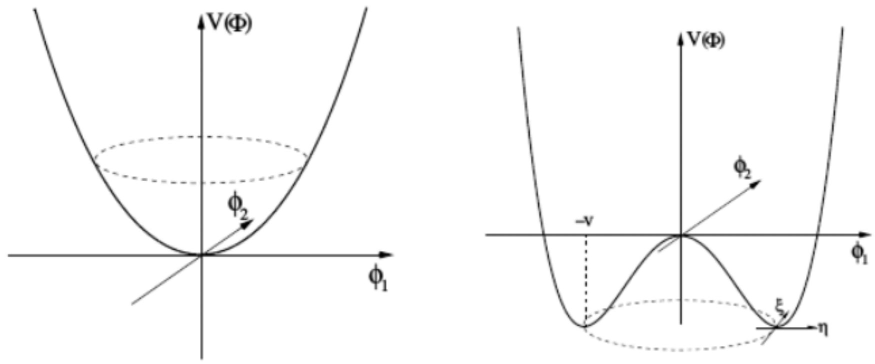


Figure 2.4: Higgs potential for $\mu^2 > 0$ (left) and $\mu^2 < 0$ (right)

To understand the shape of the Higgs potential we need to define its minima using the Euler-Lagrange equations of motion which essentially leads to the equations for the first and second derivative of the Higgs potential. The non-trivial solutions arise only if $\mu^2 < 0$, as it is demonstrated in Figure 2.4 on the right. We arrive at an infinite set of solutions satisfying $\phi_1^2 + \phi_2^2 = v^2 = \frac{-\mu^2}{\lambda}$. If we go back to the Higgs doublet for the SU(2) it is possible to derive a similar expression for the four components of the Higgs field, and we are free to choose one of the non-trivial solutions. Conventionally, the vacuum expectation value (vev) is chosen to be

$$< 0|\Phi|0 >= \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{2.17}$$

To consider the physical states we allow small perturbations of the Higgs field around the vev

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ v + H + i\phi_4 \end{pmatrix}, \tag{2.18}$$

If we consider the potential $V(\Phi)$, we find that the ϕ_1, ϕ_2, ϕ_4 are massless, while the H has a mass $m_H = \sqrt{2\lambda}v$. The Eq. 2.18 can be rewritten to emphasize the nature of massless Higgs field components,

$$\Phi = \frac{1}{\sqrt{2}} \exp \frac{i\xi^a \sigma^a}{v} \begin{pmatrix} 0 \\ v + H \end{pmatrix}, \tag{2.19}$$

where ξ^a are the fields (a=1,2,3) and σ^a are the Pauli matrices. Using the gauge transformation

$$SU(2)_L : \Phi \to e^{i\theta^a(x)\sigma^a} \Phi, \tag{2.20}$$

with $\theta^a(x) = -\frac{\xi^a(x)}{v}$ the ϕ_1, ϕ_2, ϕ_4 components disappear. This transformation is known as unitary gauge, and the Higgs field takes the form

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix}. \tag{2.21}$$

The ϕ_1, ϕ_2, ϕ_4 are massless nonphysical fields — Goldstone bosons, which are predicted to arise for each generator of broken symmetry ($SU(2)_L$) [8]. As shown above, they are eliminated and absorbed by the longitudinal components of the massive vector bosons if we choose the appropriate gauge.

After the Higgs field redefinition, the Lagrangian takes the form

$$\mathcal{L}_{Higgs} = \frac{1}{2}\partial_\mu H \partial^\mu H + \lambda v^2 H^2 - \lambda v H^3 - \frac{1}{4}\lambda H^4, \tag{2.22}$$

with the terms that describe the spin-0 field H with the mass $m_H = \sqrt{2\lambda}v$, and its self-interaction vertices, shown in Fig. 2.5.

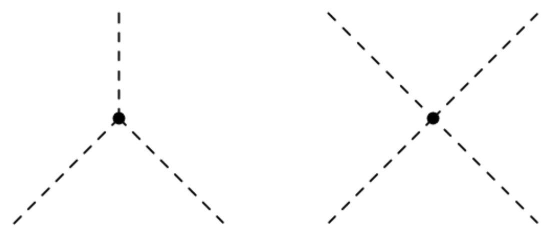


Figure 2.5: Higgs boson self-interaction diagrams

From the kinetic term $(D_\mu\Phi)^\dagger D^\mu\Phi$ we can obtain the masses for the vector bosons $W^\pm$, Z:

$$m_W = \frac{gv}{2},\ m_Z = \frac{v}{2}\sqrt{g^2 + g'^2}, \tag{2.23}$$

while the photon field A_μ stays massless. The BEH mechanism also helps to keep the SM theory renormalizable. When there are explicit mass terms in the Lagrangian, the number of terms for quantum corrections is not finite. Since the mass terms are introduced after the Higgs field obtains its vacuum expectation value at low energies, the general theory is not affected.

Yukawa interaction

The Yukawa interactions are introduced in the SM to describe the couplings of the Higgs field to the fermions, which are essential to describe the masses of fermions. Considering the dimension of the fermion fields and keeping in mind the $SU(2)_L$ symmetry one can derive the terms for leptons $L_L = (\nu_e, e)$ and e_R

$$\mathcal{L}_{\text{Yukawa}} \supset -[y_e \bar{e}_R \Phi^\dagger L_L + y_e^* \bar{L}_L \Phi^\dagger e_R]. \tag{2.24}$$

If we consider SSB and apply the unitary gauge for the Higgs field the leptons obtain their masses and the interaction vertex with the Higgs boson.

$$\mathcal{L}_{\text{Yukawa}} \supset -\frac{y_e v}{\sqrt{2}}\bar{e}e - \frac{y_e}{\sqrt{2}}H\bar{e}e, \tag{2.25}$$

where $m_e = \frac{y_e v}{\sqrt{2}}$, and the He^+e^- vertex is proportional to $\frac{y_e}{\sqrt{2}} = \frac{m_e}{v}$. The masses of leptons are measured experimentally, and they vary by orders of magnitude ($m_e = 511$ keV, $m_\tau = 1.78$ GeV). Therefore the lepton-Higgs couplings are very different for the three generations. The Yukawa interaction for quarks is introduced in the same way as for leptons. The three generations of quarks are subject to the weak flavour mixing, which is encoded into the Cabibbo-Kobayashi-Maskawa(CKM) matrix [11].

The SM Lagrangian

Combining all of the components discussed above we can write down the SM Lagrangian.

$$\begin{aligned} \mathcal{L}_{SM} = \mathcal{L}_{gauge} + \mathcal{L}_{EW} + \mathcal{L}_{QCD} + \mathcal{L}_H + \mathcal{L}_{\text{Yukawa}} &= \\ = -\frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} - \frac{1}{4} W^{i,\mu\nu} W^i_{\mu\nu} - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} & \\ + \sum_{\psi \in q,l} \bar{\psi}_L i D^L_\mu \gamma^\mu \psi_L + \sum_{\psi \in u,d,e} \bar{\psi}_R i D^R_\mu \gamma^\mu \psi_R & \\ + (D^\mu \Phi)^\dagger D_\mu \Phi - \mu^2 \Phi^\dagger \Phi + \frac{1}{4} \lambda (\Phi^\dagger \Phi)^2 & \\ - [y_u \bar{u}_R \Phi^\dagger q_L + y_l \bar{l}_L \Phi^\dagger l_R + h.c.], & \end{aligned} \tag{2.26}$$

where the covariant derivative $D_\mu = \partial_\mu + igW^i_\mu \frac{\sigma^i}{2} + ig' B_\mu Y + ig_s G^a_\mu \frac{t^a}{2}$ is defined to satisfy the $SU(3)_C \times SU(2)_L \times U(1)$ local symmetry.

2.2.3. Higgs boson production and decay at the LHC

At the LHC, during Run 2, protons were collided at the center-mass energy of $\sqrt{s} = 13\,\text{TeV}$. Fig. 2.6 shows the Feynman diagrams for the Higgs production processes at the LHC. As it can be seen from Fig. 2.7 the main production processes at this energy are: gluon-gluon fusion (ggF), vector-boson fusion (VBF), associated production with a vector boson (VH) and top-anti-top associated production ($\mathrm{t\bar{t}H}$). All of these production modes were observed at the LHC [12, 13]. The production in association with bottom quark pairs ($\mathrm{b\bar{b}H}$) and with top quark (tH) have lower cross section and more challenging signatures, so they have not been observed at the LHC yet.

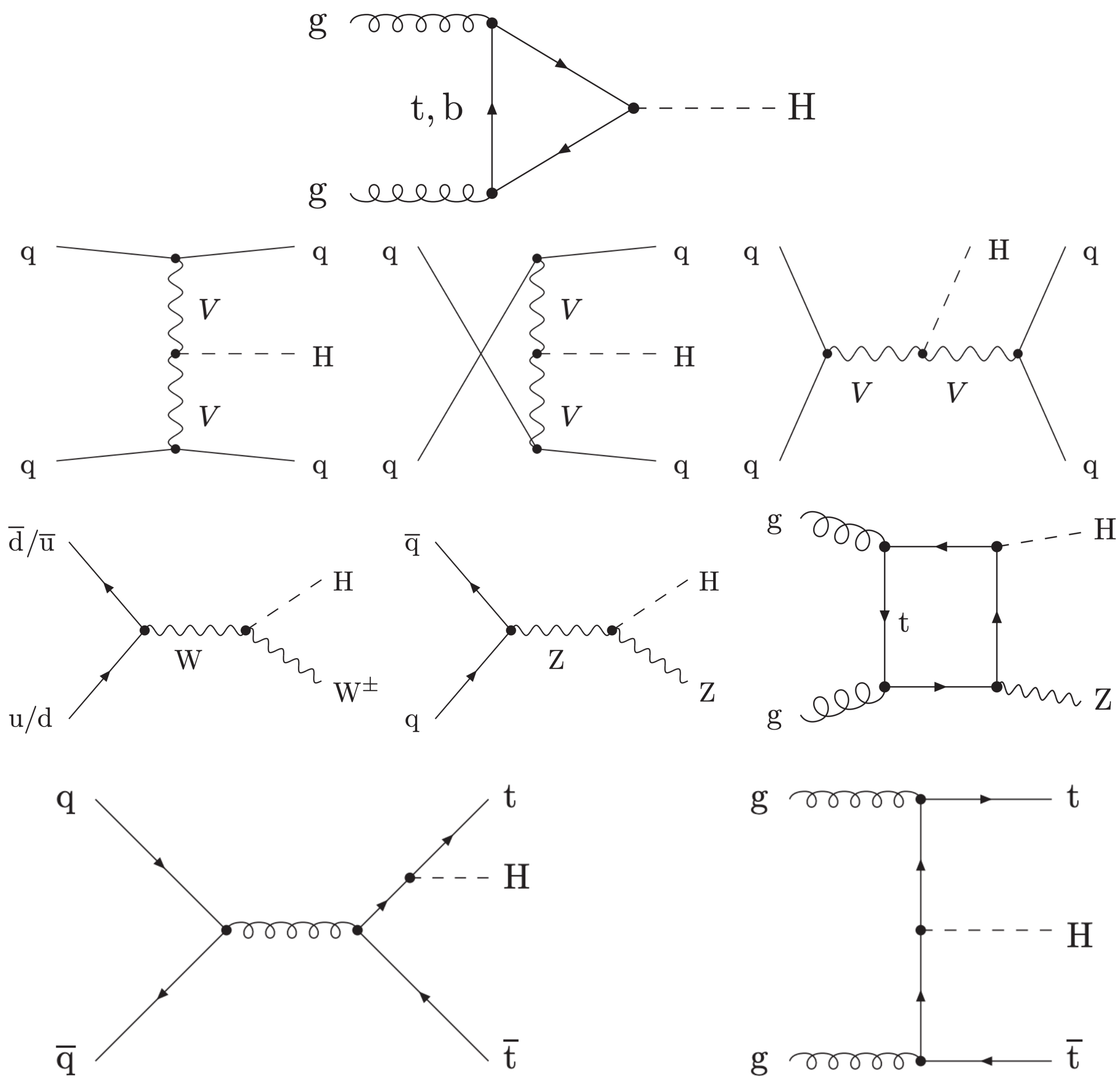


Figure 2.6: Feynman diagrams for the leading Higgs production modes at the LHC [14]. The ggH, VBF, VH and $t\bar{t}$H Higgs production processes are shown in the first, second, third and bottom rows respectively.

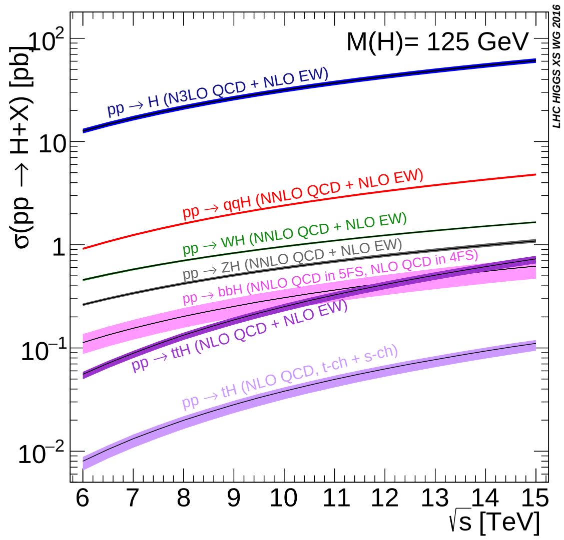


Figure 2.7: SM Higgs boson production predictions [15]

All of the mentioned production processes can be characterised by their final state. For example, the leading process ggF can be distinguished by an isolated Higgs boson in the event and is expected to have an increased sensitivity to BSM effects due to the top-quark loop, which can be modified by a BSM particle and affect the observed cross-section of this process. The VBF and V($\rightarrow qq$)H processes appear with two additional quarks, while VH ($V \rightarrow$ leptons) has leptons in the final state. The Higgs boson is an unstable particle with a short lifetime. The probability of the Higgs boson decay is defined by the Yukawa couplings (fermion decay channels) or weak coupling strength (vector bosons decay channels). The branching ratios (BRs) are shown in Fig. 2.8.

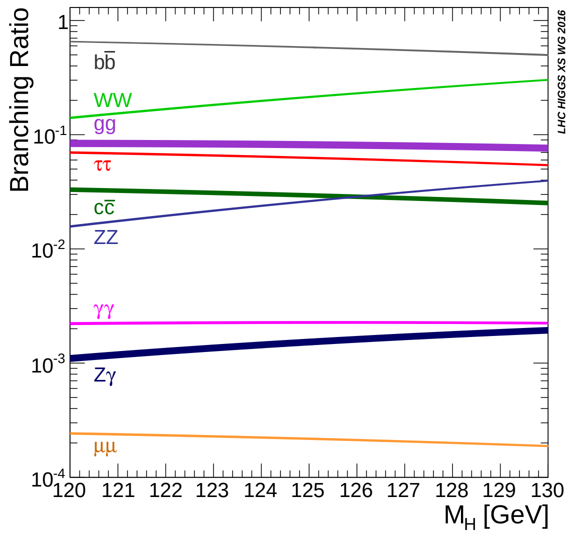


Figure 2.8: SM Higgs boson decay branching fraction predictions [15]

The leading decay channel is $H \to b\bar{b}$ with a BR of 0.58 which was observed by the ATLAS and CMS collaborations in 2018 [16, 17]. The observation of the dominant $H \to b\bar{b}$ decay was preceded by the discovery of the Higgs boson in the experimentally better accessible $H \to \gamma\gamma$ and $H \to ZZ^* \to 4l$ channels in 2012 [18, 19]. The $H \to \tau^+\tau^-$ and $H \to WW$ final states are observed as well [20, 21]. The couplings to the third generation fermions and vector bosons were observed and measured with high precision, as it can be seen in Fig. 2.9. Also the Higgs boson couplings to the second generation fermions are already being probed at the LHC through the $H \to \mu^+\mu^-$ and $H \to c\bar{c}$ channels. Evidence of the $H \to \mu^+\mu^-$ decay was recently published by the CMS collaboration [22], while the $H \to c\bar{c}$ [23] appears to be a more challenging final state and the required sensitivity might be achieved by the time of the high-luminosity phase of the LHC (HL-LHC).

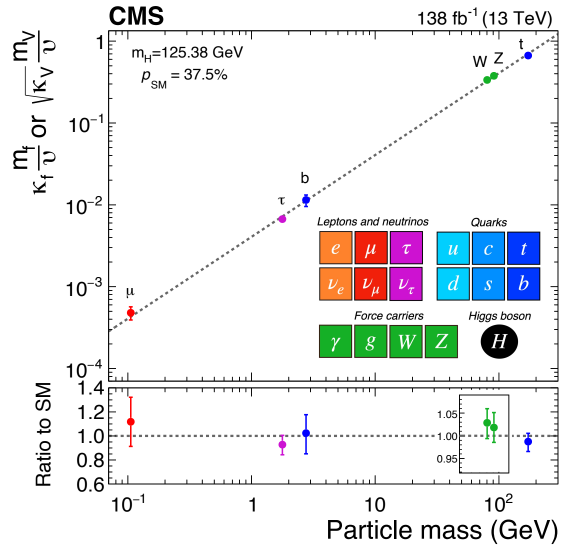


Figure 2.9: SM Higgs boson couplings to fermions and vector bosons measured in CMS [24].

While most production mechanisms and decay modes have been established at the LHC inclusively, it becomes more important to perform differential measurements, where regions with potential sensitivity to BSM physics can be isolated and explored. The Simplified Template Cross Section framework as one of the possibilities to perform such measurements will be discussed in Section 5.1.

2.2.4. Beyond Standard Model

While the SM provides a good description of the data collected at the LHC with high precision, there are many indications that it is not complete. It can be assumed that the SM is an effective theory that can describe the physics fairly well up to a certain energy scale Λ, where the more complete ultraviolet (UV) theory starts to play a role. The introduction of the new physics (NP) at some higher scale Λ generates a problem in the SM Higgs sector, due to the fact that the quantum corrections for the Higgs mass would be sensitive to any new particle of the UV theory $\Delta m_H^2\ O(\Lambda^2)$. A substantial fine-tuning is required to cancel out the corrections from the NP scale to retain the observed mass of the Higgs boson, which is unnatural and represents a hierarchy problem.

If designed with care the UV theory can be constructed is such a way that these corrections are naturally cancelled. One of the most prominent examples of such theories is the supersymmetry (SUSY). SUSY postulates an additional global symmetry between bosons and fermions. The corrections to the Higgs mass get cancelled because the contributions from the bosons and fermions are of the same size and of the opposite sign. SUSY also provides a dark matter candidate, a description of gravity and unification of EWK and strong couplings. There is a large program of direct searches for various manifestations of SUSY models, but none of them has been successful so far. Due to a major success of the precision measurements program at the LHC, there have been significant developments in the frameworks that allow model-independent BSM interpretations. The next Section will introduce the EFT approach and the Standard Model Effective Field Theory (SMEFT) framework, which is later used for the interpretations of the SM process in this thesis.

2.3. Effective Field Theory

Effective field theories (EFT) have been an important part of the SM development for a long time. The earliest and the most famous example is the contact interaction proposed by E. Fermi in 1933 [25] to describe the β decay. It was suggested that the decay proceeds through a 4-fermion interaction with the matrix element:

$$\frac{G_F}{\sqrt{2}}[\bar{u}(n)\gamma^{\mu}(1-\gamma^5)u(p)][\bar{v}(\nu_e)\gamma^{\nu}(1-\gamma^5)u(e)], \tag{2.27}$$

where G_F is Fermi constant, representing the coupling strength of the contact interaction. This approach delivered a good description of experimental data at low energies before the evidence of existence of $W^{\pm}$-boson and later the formulation of the SM. It was later discovered that this interaction is actually happening via the $W^{\pm}$-boson exchange and has an additional propagator term in the amplitude which affects the decay rate and is

$$[\frac{1}{2\sqrt{2}}\bar{u}(n)g\gamma^{\mu}(1-\gamma^5)u(p)]\frac{-i(\eta_{\mu\nu}-\frac{q_{\mu}q_{\nu}}{m_W^2})}{q^2-m_W^2}[\frac{1}{2\sqrt{2}}\bar{v}(\nu_e)g\gamma^{\nu}(1-\gamma^5)u(e)] \tag{2.28}$$

It is easy to see that for the low momenta ($q << m_W$) the contact interaction description Eq. 2.27 is valid, but for the higher energies the full underlying theory is needed Eq. 2.28. The EFT approach provides a method to describe the effects of UV theory at the infrared (IR) scale, without any assumption on UV theory until the energy scale of possible UV completion is reached, which makes the EFT approach almost model-independent. By measuring the parameters of the EFT we can get an access to the parameters of the full underlying theory. There are many frameworks in the EFT world, but the most relevant to the Higgs physics are the Standard Model Effective Field Theory (SMEFT) [26] and Higgs Effective Field Theory (HEFT) [27]. In HEFT the Higgs is not required to be a SU(2) doublet, therefore HEFT provides the most general description of possible Higgs coupling. But the matching to the concrete BSM models is complicated. The SMEFT has the same structure and the particle content as the SM, and the Higgs boson is an $SU(2)_L$ doublet. The measurements done within the SMEFT can be easily reinterpreted to match the complete BSM theory.

2.3.1. SMEFT

The SM predictions have been reinstated by the experimental measurements for many years now. The earlier measurements at LEP as well as the precision measurements at the LHC have been providing reassuring evidence that the new UV physics can have only small effect at the energy scale we can achieve so far. The SMEFT is a complete and powerful framework which is already widely used to interpret the LHC measurements in the search for BSM effects [28]. The SMEFT assumes that the $SU(3)_\mathrm{c} \times SU(2)_L \times U(1)_Y$ symmetry group and the SM particle content are correct, but suggests that it is incomplete by incorporating the contributions from higher-order momentum operators:

$$\mathcal{L}_{SMEFT} = \mathcal{L}_{SM} + \mathcal{L}^{(5)} + \mathcal{L}^{(6)} + \mathcal{L}^{(7)} + \mathcal{L}^{(8)} + ..., \tag{2.29}$$

where the $\mathcal{L}_{SM}$ contains dimension-4 operators while the higher-order terms are defined as follows:

$$\mathcal{L}^{(d)} = \sum_{i=1}^{N(d)} \frac{c_i^d}{\Lambda^{d-4}} O_i^d. \tag{2.30}$$

The O_i^d operators are constructed using the SM fields, the c_i^d are dimensionless parameters called Willson coefficients. To keep the action dimensionless the terms containing higher-order operators are normalised by a mass scale Λ, much higher than the energy scale of the SM. With this parametrisation the Willson coefficients c_i^d deviating from 0 would point to BSM effects. By measuring the c_i^d we can access not only the parameters of the UV theory but also its nature.

In the SMEFT analyses the order of BSM operators is usually limited. The $\mathcal{L}^{(5)}$ term violates the lepton number and all of the higher-order odd dimension operators violate the B-L number conservation. In this thesis the dimension-6 operators corresponding to the $1/\Lambda^2$ order will be considered. Further details and the Warsaw basis are introduced in Chapter 8.

CHAPTER 3

The LHC and CMS detector

3.1. LHC

The Large Hadron Collider (LHC) [29] is so far the most powerful machine constructed to explore the fundamental structure of the universe. The collider itself is a 27 km long underground ring located on the border of France and Switzerland near Geneva. It was constructed inside the LEP tunnel. The design of this machine assumes the collision of proton bunches at center mass energy ($\sqrt{s}$) up to 14 TeV and luminosity of $10^{34} cm^{-2} s^{-1}$, as well as heavy ions collisions with the energy of 5.5 TeV per nucleon and a peak luminosity of $10^{27} cm^{-2} s^{-1}$.

The protons are accelerated in several steps to an energy of 50 MeV in LINAC 2, to 1.4 GeV in the Proton Synchotron Booster (PSB), and to 25 GeV in the Proton Synchotron (PS). In the PS the protons are spaced into bunches and fed into the Super Proton Synchotron, where they get accelerated to an energy of 450 GeV. Then the bunches are released into two LHC pipes where they travel in the opposite directions until they reach the final energy. In Run 2 during the 2016-2018 data-taking the protons were collided every 25 ns with an energy of $\sqrt{s}$=13 TeV, which is slightly below the original design. The energy is increased in the new Run 3 which has started in July 2022 to $\sqrt{s}$=13.6 TeV, and it is expected that in the High-Luminosity era of the LHC the energy will be increased again to reach the designed limit of 14 TeV.

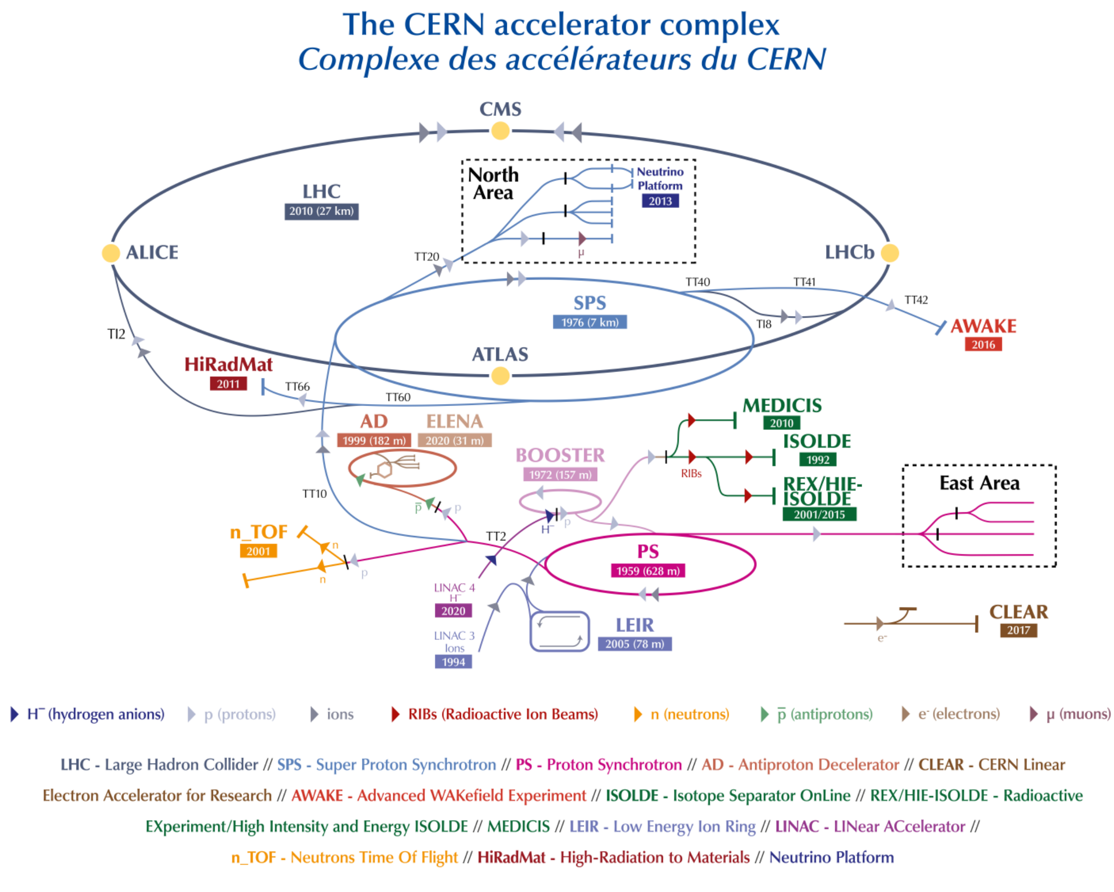


Figure 3.1: The LHC acceleration complex scheme.

The LHC host four major experiments, shown schematically shown in Figure 3.1. The first two are the ATLAS [30] and CMS [31], the general-purpose detectors, designed to study wide range of HEP topics. The ALICE [32] detector is focused on heavy-ion physics, and LHCb [33] has a dedicated design targeted at flavour physics. The results presented in this thesis are obtained using the data collected by the CMS experiment in Run 2 during the 2016-2018 data-taking period. The delivered luminosity by the LHC vs. time for each data-taking year is shown in Figure 3.2. Not all of what is recorded is used in the physics analyses. The total integrated luminosity certified to be used for physics in CMS is 138 fb^{-1}.

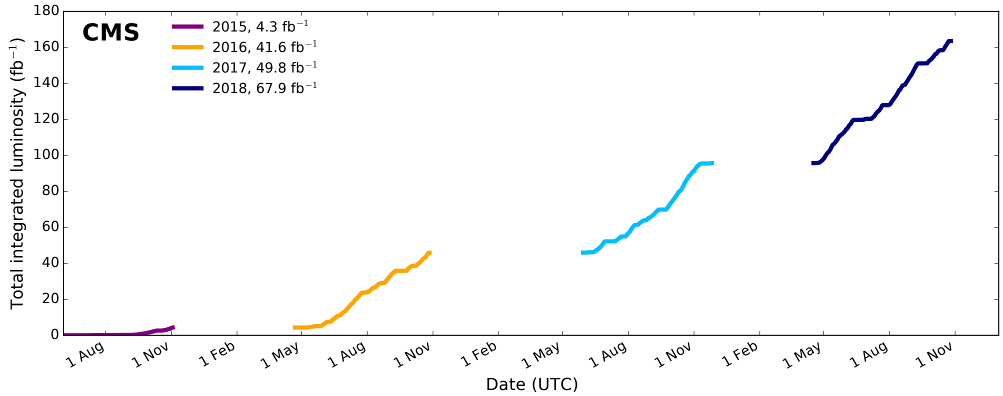


Figure 3.2: Delivered luminosity versus time for 2015-2018 (pp data only)

3.1.1. CMS

The CMS (Compact Muon Solenoid) detector [31] is designed to study the physics processes at TeV energy scale, where at the time of the design many expected to see hints of BSM effects. In comparison to the ATLAS detector [30], the CMS detector is indeed compact, it is 15 m in height and 29 m in length. The CMS detector has a cylindrical shape with the subdetectors situated around the beam pipe (Figure 3.3). The closest one is the silicon tracker, followed by the electromagnetic and hadron calorimeters, and the muon system embedded within the iron return yoke of the superconducting magnet. The powerful solenoid magnet, placed between the hadron calorimeter and the muon system, provides a magnetic field of 3.8 T, allowing the precise measurements of charged particles momenta.

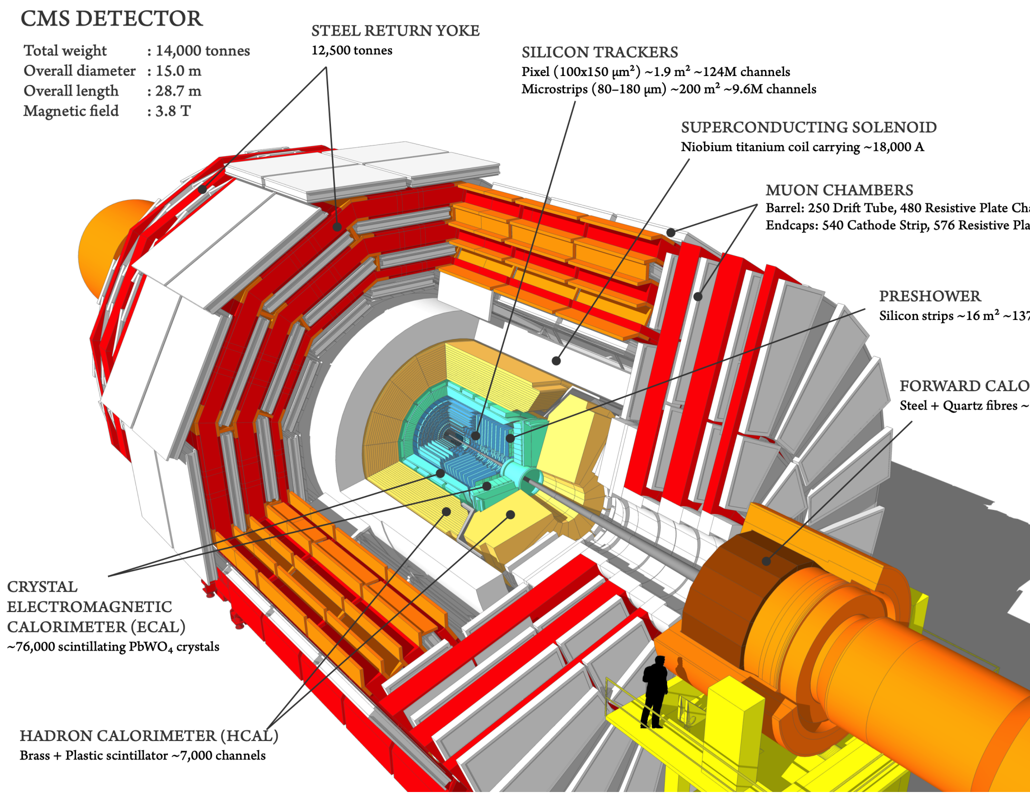


Figure 3.3: The CMS detector general view

The proton-proton collision point is situated in the center of the CMS surrounded by the subdetectors, designed to measure various physics objects. The coordinate system of CMS starts in the interaction point(IP), with the z-axis directed along the beam axis, and the called transverse plane (x,y) plane where x-axis is pointing to the center of the LHC ring and y-axis pointing in vertical direction. Within the CMS collaboration the cylindrical coordinates are used. Instead of the polar angle θ, the pseudo-rapidity variable is defined:

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

The angular separation between the particles is often characterised using the δR variable defined as:

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

where ϕ is the azimuthal angle. These variables are useful, because the differences are Lorentz invariant under the boost along the beam direction.

3.1.2. Tracker

The CMS tracking system [34] is the innermost part of the CMS detector, consisting of pixel, the closest to the beam pipe, and the silicon tracker. The main purpose of the CMS tracker is the charged particles tracks reconstruction and the consequent precise measurement of their momentum.

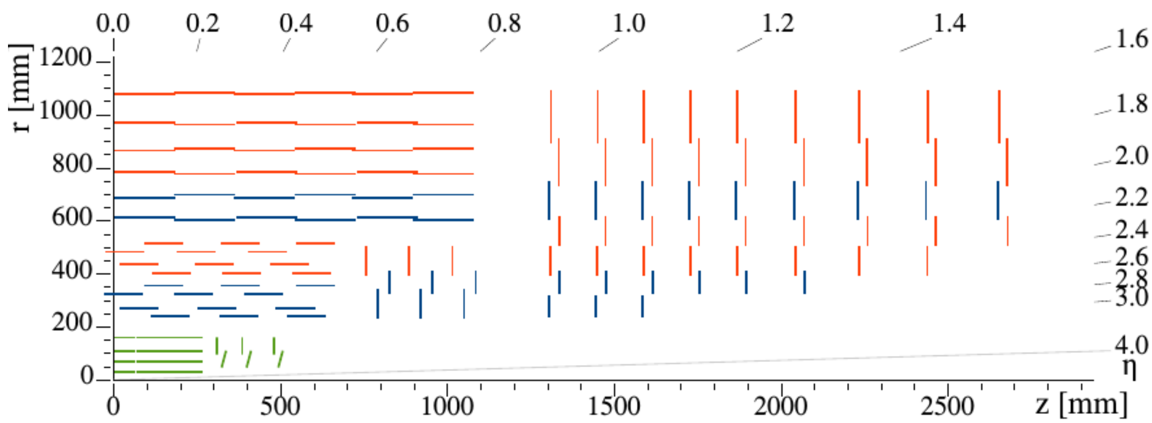


Figure 3.4: The CMS tracking system scheme.

The momentum of the charged particles is measured from the curvature of tracks inside the magnetic field of the CMS detector. The tracks are built from combination of hits in strip and pixel detectors. The momentum resolution of the tracks reconstructed in the barrel part ($\eta < 1.0$) of the tracker is 2% for $p_{\mathrm{T}} < 100$ GeV [35]. The tracker performance is crucial to the reconstruction of the primary and secondary vertices. The primary vertices (PV) are reconstructed by extrapolation of all reconstructed and selected tracks. The resolution of PV vertex resolution is measured to be 25 μm. The Secondary Vertices (SV) are reconstructed based on the tracker information, which is an important input to the successful b-tagging.

3.1.3. ECAL

The CMS electromagnetic calorimeter (ECAL) [36] is designed to reconstruct the energy of showers created by electrons and photons. It is situated next to the tracker and designed to have high reconstruction efficiency on a very large range of energies. The ECAL is a homogeneous calorimeter made of $PbWO_4$ scintilators. The schematic view of ECAL is shown in Figure 3.5. It is comprised of the ECAL barrel (EB) covering the region $|\eta| < 1.48$, and ECAL endcaps extending the coverage to $|\eta| < 3.0$.

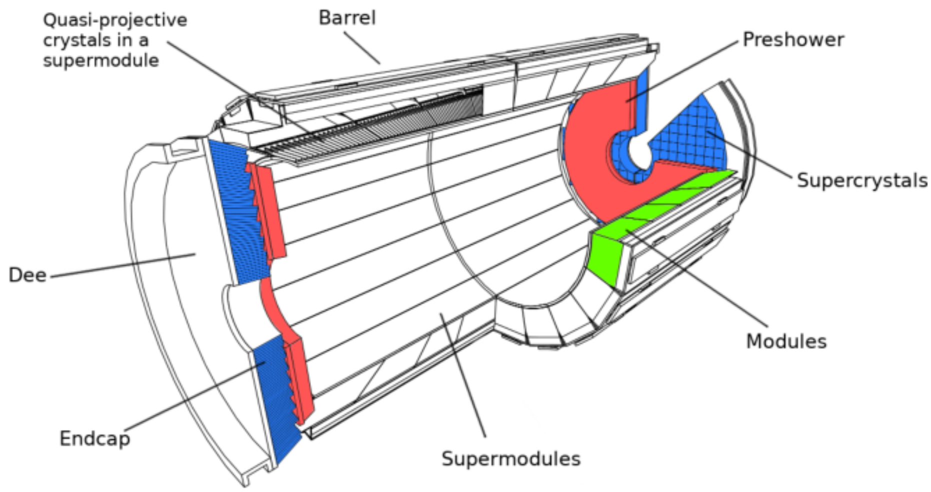


Figure 3.5: The CMS ECAL subdetector.

The energy resolution of the showers reconstructed in ECAL is given:

$$\frac{\sigma}{E} = \sqrt{(\frac{2.8\%}{\sqrt{E/GeV}})^2 + (\frac{12\%}{E/GeV})^2 + (0.3\%)^2}, \tag{3.3}$$

where the first term models the statistical fluctuations in the shower formation, the second term models the noise in electronics, and the last term accounts for the energy leakage.

3.1.4. HCAL

The hadron calorimeter (HCAL) [37] is designed to detect and reconstruct the showers generated by jets. The HCAL is a sampling calorimeter, consisting of active layers of plastic scintillator and brass absorber plates.

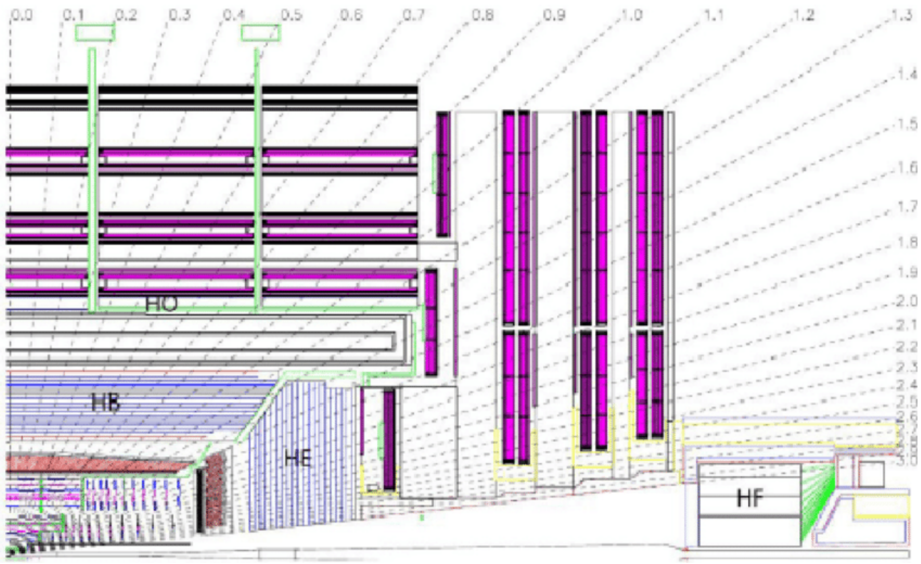


Figure 3.6: The CMS HCAL subdetector.

The HCAL is split up in four parts (Figure 3.6), which together ensure extensive coverage in $\eta < 5.2$. The barrel (HB) covers the $|\eta| < 1.3$ range, followed by the outer calorimeter (HO). The endcap (HE) covers the pseudo-rapidity range $1.3 < \eta < 3$, and the forward calorimeter (HF) extends to η<5.2. The HF is designed to be particularly radiation-resistant due to the very high particle flux in this eta region. For the reconstruction of the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ it is important to have a hermetic environment, and this is due to the extensive pseudorapidity coverage provided by the HCAL.

3.1.5. Muon

The purpose of the muon system [38] is to provide an accurate measurement of the muon momenta and good timing resolution to be used for triggering. The muon system detectors shown in Figure 3.8 are based on three different gaseous technologies: the drift tubes (DT) $|\eta| < 1.2$, the cathode strip chambers (CSC) $0.9 < |\eta| < 2.4$, and resistive plate chambers (RPC) $|\eta| < 1.6$. The choice of the detector technology is driven by the magnetic field configuration, as well as the density of muon tracks. The RPC and CSC detectors provide excellent timing resolution which is crucial for the triggering.

The muon tracks are reconstructed combining the information from strip detector, which ensures purity and high efficiency. The momenta resolution is below 6% even in the most forward region.

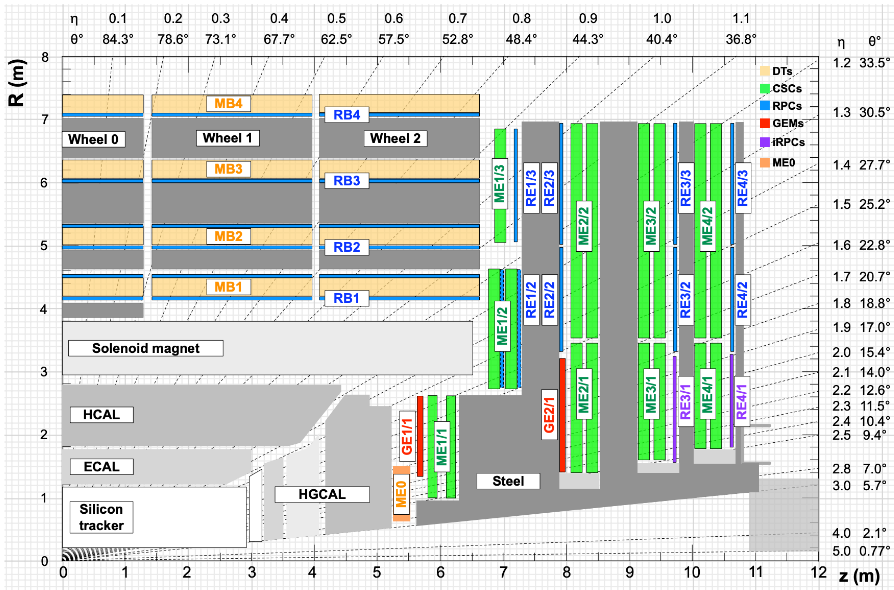


Figure 3.7: The CMS muon system.

3.1.6. Trigger

With the design luminosity of LHC the proton-proton bunch collisions are happening with the frequency of 40 MHz. It is not possible to contain and process such a large amount of data stream and most of these collisions are not interesting for the physics studies, therefore the triggering is essential. The CMS collaboration uses a two-tiered trigger system [39] to reduce the high collision rate. The Level-1 Trigger (L1T) is designed to reduce the rate to 100 kHz, and the High Level Trigger (HLT) reduces it further to 1 Hz, which is then saved to disk.

The L1T has to be fast in making decision whether the event can be accepted, therefore only low-level variables are used. In addition, it is not possible to include the track reconstruction at this level, because the event has to be processed within 4 μs.

Then the events are passed to the HLT, where a more detailed information is accessed to make a decision. The HLT selection relies on the information reconstructed from all detectors, but it is not as precise as the reconstruction performed at the offline level. The full HLT configuration is a combination of many different HLT paths, often created to target a specific final state.

3.2. Object reconstruction

In CMS the physics objects are reconstructed using the particle-flow (PF) method [40]. The information from all subdetectors is used to identify and measure the particles final states. The procedure starts with the reconstruction of the tracks in the silicon tracker and muon system, and the clusters of energy from the ECAL and HCAL. The tracks are built using the iterative tracking algorithm, with the Kalman fitter [41] to improve the accuracy. A clustering algorithm is used to convert the energy deposits in the ECAL and HCAL into the final physics objects.

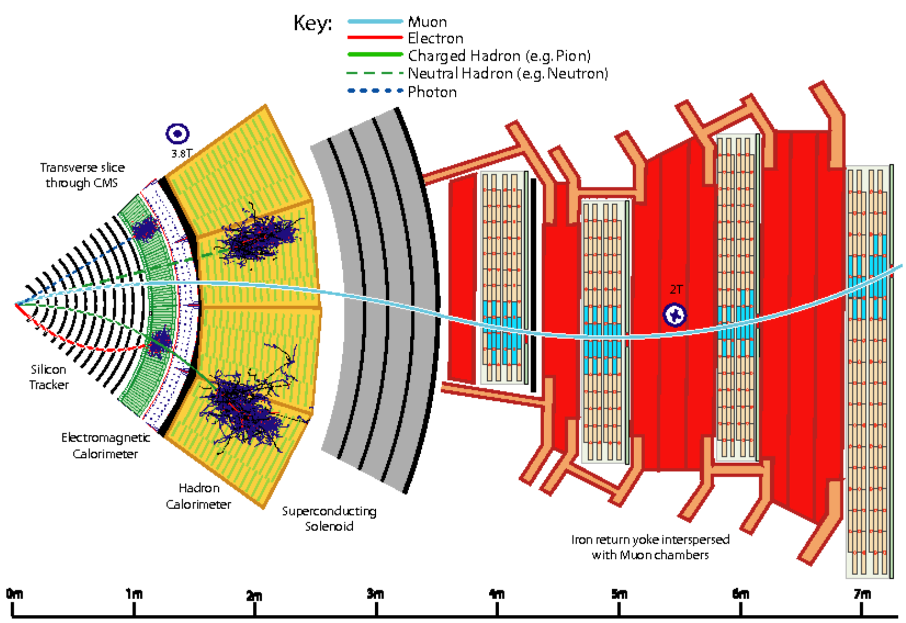


Figure 3.8: The CMS particle flow

The individual particles are reconstructed from the combination of various sub-detectors defined as follows:

- Muons: the track in the muon system is combined with the track in the tracker. The momentum is calculated from the curvature of the track.
- Electrons: identified as the cluster in the ECAL linked to a track from the tracker.
- Charged hadrons: identified as the associated track from the tracker with the clusters in ECAL and HCAL.
- Neutral hadrons: identified as the linked clusters in ECAL and HCAL, without the track in the tracker.
- Photons: the cluster in the ECAL without the track in the tracker.

3.2.1. Vertices reconstruction

The primary vertices (PV) reconstruction [35] relies on the tracks reconstructed in the tracker, especially in the pixel detector. It starts with the selection of tracks compatible with being originated from the IP. The selected tracks are clustered, and each cluster is fitted to find a vertex position. The leading vertex is selected as the vertex with the highest sum of squared transverse momentum of tracks. The other vertices are classified as pileup vertices.

The reconstruction of the secondary vertices (SV) is necessary for the b-tagging algorithms. First the seeding tracks are selected as the tracks with high impact parameter significance with respect to the PV. The seeding tracks are then used to cluster other tracks around them based on the angular and 3D distance measures. The clustered track are then fitted to a common vertex using an outlier-resistant fitter [42].

3.2.2. Leptons

The electrons and muons are required to be prompt (originate from a PV), therefore the impact parameter is required to be below 0.5 cm in the x-y plane, and below 1 cm along the z-axis. To select clean leptons the relative isolation variable defined below is used.

$$I_{PF,rel} = \frac{1}{p_{\mathrm{T}}^{l}}(\sum p_{\mathrm{T}}^{\mathrm{charged}} + max\left(0, \sum p_{\mathrm{T}}^{\mathrm{neutral}} + \sum p_{\mathrm{T}}^{\gamma} - p_{\mathrm{T}}^{\mathrm{PU}}\right)), \tag{3.4}$$

where the sum runs over all particles within $\Delta R = 0.3$ for electrons and $\Delta R = 0.4$ for muons. The perfectly isolated lepton would have $I_{PF,rel} = 0$. The isolation criteria used in this work varies with the multiplicity of leptons in the final state. In the event with one lepton the $I_{PF,rel}$ is required to be below 0.06, and below 0.15 for the events with two leptons. The contribution from fake electrons is reduced by employing the MVA based discriminator, for which the tight working point corresponding to 80% efficiency is used [43].

3.2.3. Jets

Jets clustering

The jets are the product of the parton hadronization. The main purpose of the jets reconstruction is to derive the energy and spatial properties of the original parton-level particle. Different clustering methods can be used, the most common one in the anti-k_T algorithm [44]. The particles are clustered together based on the separation variable defined as:

$$d_{ij} = min({p_{\mathrm{T}}^{-2}}_i, {p_{\mathrm{T}}^{-2}}_j)\frac{\Delta R_{ij}}{R}, \tag{3.5}$$

where R is usually chosen to be 0.4. The particles are sequentially combined until $d_{ij} < d_{iB}$, where $d_{iB} = {p_{\mathrm{T}}^{-2}}_i$. If this criteria is not satisfied, the jet is formed and removed from the collection of jets that are used for further clustering.

Jet size

The typical jet size used in many analyses in CMS is R=0.4 (AK4 jets). The jets produced in the boosted regime, e.g. the jets from $\mathrm{H} \to \mathrm{b\bar{b}}$ decay with the Higgs momenta $p_{\mathrm{T}} > 300$ GeV, are difficult to resolve. Therefore the boosted jets are reconstructed with a cone size of 0.8, and referred to as AK8 jets. In this thesis the AK8 jets are used to reconstructed the $\mathrm{H} \to \mathrm{b\bar{b}}$ decay in the boosted topology in the high momenta regime ($\mathrm{p_T^V} > 250$ GeV).

The AK8 jets are groomed with a soft drop declustering algorithm [45], which recursively removes soft wide-angle radiation from a jet. The invariant mass of such jet is referred to as soft-drop mass.

Pileup cleaning (CHS)

The particles associated with the tracks coming from the pileup vertices are subtracted. The resulting jets are called the CHS (Charged Hadron Subtracted) jets and contain only the tracks associated to the primary vertex.

Jet energy scale corrections

Jet energy scale corrections are applied in data and simulation. These corrections are derived to account for: the remaining pileup from neutral hadrons, derived from the QCD simulation; detector response, computed from the simulation in bins of p_T and η; and the corrections that account for the different detector response for various jet flavours. The typical size of jet energy scale corrections derived from simulation is shown in Figure 3.9 in different η regions and for the jets of different p_T.

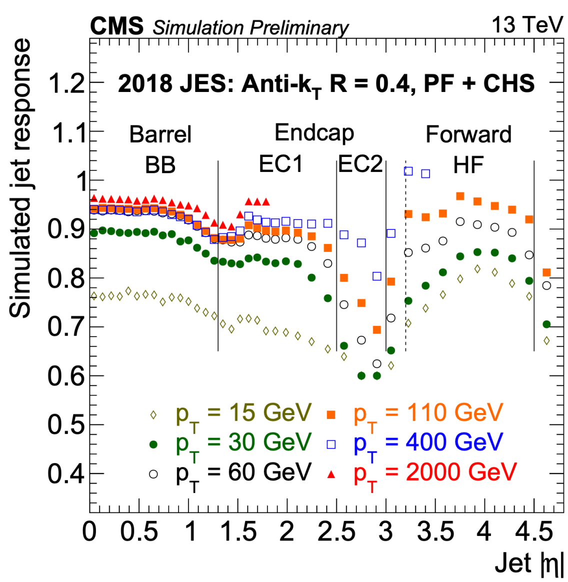


Figure 3.9: The CMS jet energy scale corrections measured in 2018 data-taking period [46].

Neutral hadron fraction	< 0.9
Neutral EM fraction	< 0.9
Charged hadron fraction	> 0
Charged EM fraction	< 0.8
Muon fraction	< 0.8
Charged multiplicity	> 0
Number of constituents	> 0

Table 3.1: The Jet ID selection definition

Jet ID

The fake jets are removed by applying the set of selection criteria on jet energy fractions and constituents listed in Table 3.1.

Jet flavour identification

The jets flavour identification is crucial for the analysis presented in this thesis. The identification of jets that originate from the hadronization of b quarks is performed with the algorithm based on deep neural network (DNN) Deep Combined Secondary Vertex (DeepCSV) [47]. As an input DeepCSV algorithm uses the information on secondary vertices and tracks and their impact parameters. This DeepCSV DNN has several probability outputs for different jet flavours: b – corresponding to 1 b-hadron, bb – corresponding to 2 b-hadrons, c – corresponding to 1 c-hadron, cc – corresponding to 2 c-hadrons and the *light* corresponding to no b and c-hadrons. The b and bb classes probabilities are used to construct the b-tagging output score used in this analysis.

The DeepCSV working points are defined for several tagging efficiency and mistag probabilities values. The corresponding values are shown in Table 3.2

WP Name	Mistag efficiency	DeepCSV
Loose (2018)	10%	0.1241
Medium(2018)	1%	0.4184
Tight (2018)	0.1%	0.7527
Loose (2017)	10%	0.1522
Medium(2017)	1%	0.4941
Tight (2017)	0.1%	0.8001
Loose (2016)	10%	0.2219
Medium(2016)	1%	0.6324
Tight (2016)	0.1%	0.8958

Table 3.2: Definition of b-tagging working points (WP) and the corresponding efficiencies.

The efficiency corrections of this working point are measured using the tag-and-probe method in p_{T} and η bins and applied to the MC simulation.

The AK8 jets are tagged using the DeepAK8 [48], tagger developed to select the boosted resonances. The AK8 jets are classified into several classes, for each of them the probability output is defined. This analysis uses the $Hb\bar{b}$ node, later referred to as *bbVsLight*. This analysis uses the jet-mass decorrelated version to get rid of the jet mass sculpting. The efficiency measurements are not available for the whole DeepAK8 output range, so the score is binned in 3 ranges: 0-0.8, 0.8-0.97, 0.97-1.0. As they are available only for signal topology, the efficiency corrections are applied to the $\mathrm{H} \rightarrow \mathrm{b}\bar{\mathrm{b}}$ and the V+bb processes. The efficiency for the other processes is measured within the analysis by the means of floating rate parameters as described in Section 4.2.2.

B-jet energy regression

The b-jet energy regression is applied for all b-jets to recover the energy lost due to the leptonic decays of b-hadrons with neutrino in the final state. The algorithm is based on a DNN, which was trained on the b-jets from QCD simulation, and applicable in many topologies with b-jets in the final state. The resolution improvement from the jet energy correction is estimated to be 13% for the event topology similar to the one used in this analysis [49].

3.2.4. Missing energy

The missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) refers to the energy that is not detected by any of the subdetectors, but is needed to ensure the momentum conservation in the collision events. It can be generated by neutrinos, new physics particles, and detector effects such as limited efficiency in some regions.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ can be only reconstructed in transverse plane, since it is difficult to estimate the longitudinal momentum of colliding partons. In general, $E_{\mathrm{T}}^{\mathrm{miss}}$ is defined as the negative sum of all particles momenta:

$$E_{\mathrm{T}}^{\mathrm{miss}} = -\sum_{\text{all particles}} \vec{p_{\mathrm{T}}} \tag{3.6}$$

Within the CMS analyses different types of $E_{\mathrm{T}}^{\mathrm{miss}}$ can be used depending on the analysis requirements.

The so called raw PF $E_{\mathrm{T}}^{\mathrm{miss}}$ uses all particles reconstructed with the particle-flow algorithm. The Type 1 PF $E_{\mathrm{T}}^{\mathrm{miss}}$, is calculated from PF jets after the application of jet energy corrections and used for the offline reconstruction within this thesis:

$$\mathrm{PF}\ \mathrm{E}_{\mathrm{T}}^{\mathrm{miss}} = -\left(\sum_{\text{jets}} \vec{\mathrm{p}_{\mathrm{T}}^{\mathrm{corr}}} + \sum_{\text{leptons}} \vec{\mathrm{p}_{\mathrm{T}}} + \sum_{\text{uncl}} \vec{\mathrm{p}_{\mathrm{T}}}\right). \tag{3.7}$$

3.2.5. Simulating the HEP collisions

The physics processes generated in the Monte Carlo (MC) simulation should be as close as possible to the real collisions observed in data. The proton-proton collisions final states are very complex, which means that there are many aspects to be taken into account when such a simulation is performed. In this section the main stages of the p-p collisions event simulation are discussed.

The hard scattering

The first steps is to generate the hard scattering process, with the differential cross-section that can be expressed:

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\Omega \mathrm{dx}_1 \mathrm{dx}_2} \propto \sum_{1,2} f_1(x_1, \mu_F) f_2(x_2, \mu_F) \times \sigma_{1,2\to f}(\mu_F, \mu_R), \tag{3.8}$$

where the matrix element cross-section $\sigma_{1,2\to f}(\mu_F, \mu_R)$ is separated from the parton distribution functions $f_1(x_1, \mu_F)$, $f_2(x_2, \mu_F)$. This is based on the assumption that the initial-state partons inside the protons are separated from the hard scattering process, this is called factorisation and parameterized by the factorisation scale μ_F.

The next important aspect of the hard scattering process is the renormalisation. It is a necessary process to resolve the divergences arising due to the unlimited integration over the momentum inside the loop diagrams. The renormalisation processes gives rise to the renormalisation scale μ_R which is a cutoff used in the momenta integration. The physics processes do not depend on the choice of μ_R and μ_F, but their variations are used to estimate the uncertainty arising from the missing higher order calculations. The higher order corrections are introduced into the matrix element cross-section, which is expanded in terms of α_s when the higher order QCD calculations are considered. The uncertainties from the μ_R and μ_F scale variations are decreased when the higher order QCD calculation is considered.

The functions $f_i(x_i, \mu_F)$ describe the parton density inside a proton. The partons are assumed to be decoupled within a proton. These PDFs are extracted from other collisions data and are implemented within the MC generators [50].

Parton shower and hadronization

The parton shower are used to approximate higher order QCD emissions that are challenging to predict analytically. This process is simulated by branching the partons outgoing from the hard scatter process into two. An example of a simulated event including the parton showering is shown in Figure 3.10, where the parton showers are shown in red as gluon emissions.

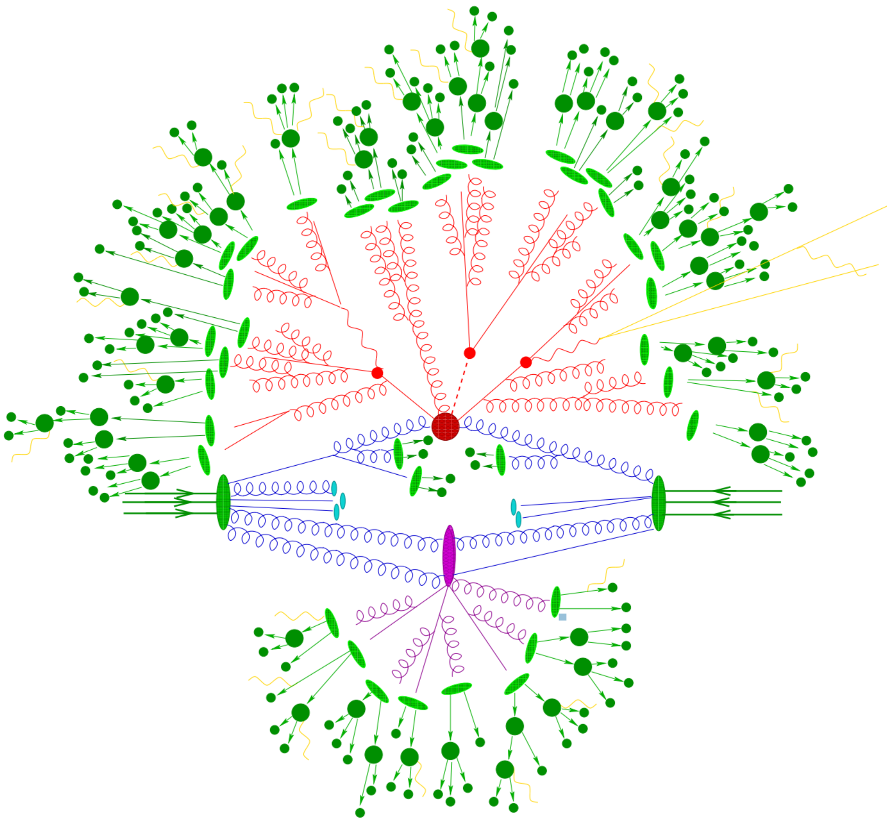


Figure 3.10: The proton-proton collision visualisation including the hard interaction, the parton shower and the hadronization processes [51].

The next step of event generation is the formation of hadrons, occurring right after the parton shower. To achieve a realistic event representation, the partons are confined into the colour-neutral final states. The factorisation model used in this thesis is the Lund string model [52] implemented in PYTHIA [53]. The interaction within quark-anti-quark pair is assumed to have a linear dependence on the distance and is modelled with a massless string which breaks apart to create new partons if the energy carried by quarks is large enough. At the end of the hadronization the colour-neutral final states are formed.

Underlying event (UE)

Since the initial state and the hard scattering processes are factorised, the exchange of the partons between the initial state particles is neglected. However, this additional activity generated by the multiple parton interactions or interactions between the partons that are not the part of the main hard scattering process plays a role. The partons are coloured and can have a non-negligible effect in the hard scattering process. The UE can not be accurately predicted within QCD and is modeled by the experiments by comparing UE activity in data and simulation. The input theory parameters are adjusted to describe the data and are combined into a set of parameters, which is then used in the MC simulation.

ME + PS matching

The partons produced at the matrix element level and the partons produced as a result of the parton shower can generate double-counting when these processes are combined. To remove the double-counting the ME + PS matching techniques are developed [54]. The matching is performed by separately generating the parton-level events for each jet multiplicity, and consequently showering them. Then the showered partons are clustered into jets, and each jet is matched to the particle-level parton. If all of the showered partons are matched with the ME partons the events are not discarded.

Detector effects

The detector response is modeled using the detailed CMS detector simulation with the GEANT4 package [55]. It allows to accurately model the interaction of the particles with each of the subdetectors and the magnetic field. The reconstruction of the simulated events follows exactly the same procedure as used for the data reconstruction.

CHAPTER 4

Analysis VH(H $\rightarrow b\bar{b}$) final state

4.1. Introduction

This chapter summarises the strategy of VH(H $\rightarrow b\bar{b}$) signal extraction. using the full CMS Run 2 proton-proton collision data.

The previous VH(H $\rightarrow b\bar{b}$) measurement is published by the CMS collaboration [16] which in combination with other production modes established the Higgs boson decay to the b-quark pair at the observation level. This previous VH(H $\rightarrow b\bar{b}$) analysis was based on the combination of Run 1 with partial Run 2 datasets and provided the inclusive signal strength measurement. The analysis strategy detailed in this thesis is built upon the ideas developed in the Run 1 + partial Run 2 measurement, that are refined to target the STXS measurement.

To extract the inclusive or STXS measurements the signal and background processes, introduced in Section 4.2, are fitted to data in orthogonal signal and control regions. Signal regions are defined through the selection criteria maximising the efficiency of signal events. To constrain the contributions of different backgrounds entering the signal region, the control regions are defined for each irreducible background (Section 4.2.2). The selection used for the control regions is defined to provide the enrichment by the corresponding backgrounds. The signal and background templates used in the fit are obtained from a Monte Carlo simulation.

The signal and background separation is improved by employing a multi-variate analysis (MVA) classifier as an observable in signal regions, as discussed in Section 4.7.

The analysis is further extended, by introducing the categories targeting the boosted Higgs decay topology in the $p_T^V > 250$ GeV. In this p_T^V region the boosted and resolved topologies are both considered.

Finally, the main highlight of the present analysis is the introduction of particle-level STXS categorisation for signal processes and the corresponding tagging of signal regions aligning with the particle-level STXS scheme. The details about the STXS categorisation and arising systematic uncertainties are given in Chapter 5.

The full Run 2 measurement is extracted from the simultaneous fit of 243 categories, where all of the measured signal and background processes are freely floated. The systematic uncertainties affecting the normalisation and the shape of considered processes are discussed in Section 6.4.

4.2. Signal and background processes

4.2.1. Signal

The VH production processes, shown as Feynman diagram in Figure 4.1, represent the most sensitive production mode for the reconstruction of $H \to b\bar{b}$ decay. The cross-section of these production modes are not the largest among the Higgs production processes, but due to the leptonic decay modes of vector bosons $Z \to \nu\nu$, $W \to l\nu$ and $Z \to ll$ the triggering and background rejection is more efficient.

The (gg) qqZH, Z$\to$ $\nu\nu$ final state can be characterised by the presence of E_T^{miss} in the final state with no additional leptons and is later referred to as *0-lepton channel*. The final state with $Z \to ll$ decay is reconstructed by requiring the presence of two isolated leptons (muons or electrons) of the same flavour and opposite charge and is defined as *2-lepton channel* in this analyses. The $W \to l\nu$ decay mode requires exactly one isolated lepton and is referred to as *1-lepton channel*. The decays of vector bosons to τ-leptons are not explicitly reconstructed. However, the τ leptons decaying leptonically enter the 2-lepton and 1-lepton channels.

The loop-induced vector boson associated Higgs production (ggZH) shown in Figure 4.2 and quark induced vector boson associated shown in Figure 4.1, have very similar final states and are very challenging to distinguish.

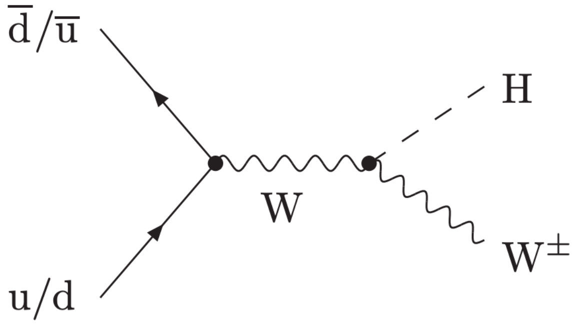


Figure 4.1: $qq' \to WH$ and $qq \to ZH$ production processes diagrams

Due to the destructive interference of box and triangle loop diagrams the ggZH process has significantly lower cross section and only populates around 10% of reconstructed events. Nevertheless, this process is quite interesting, due to a general BSM potential of SM loop processes. A separate measurement of the ggZH process requires improvements in theoretical predication and new ideas for experimental analysis [56].

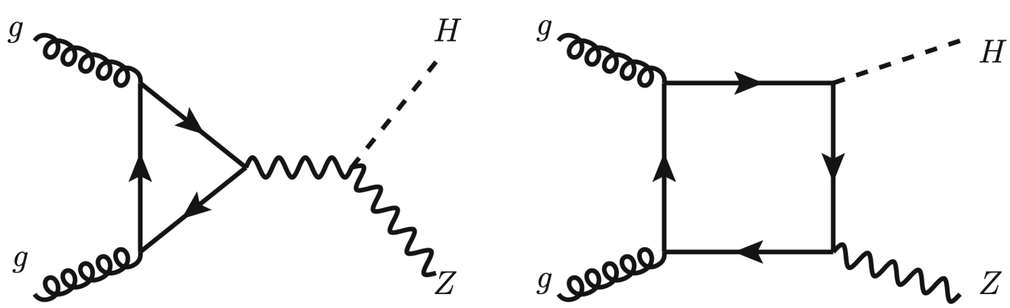


Figure 4.2: ggZH production processes diagram

4.2.2. Background processes

The leading background contributions are defined by the final state of the considered signal processes, discussed in the previous section. Therefore, for the VH(H $\rightarrow b\bar{b}$) final state the main irreducible SM backgrounds are the following: V+jets, $t\bar{t}$, diboson, and the single-top-quark production.

V+jets

The V+jets process illustrated in Figure 4.3 can produce a final state very similar to the signal. The jets in the V+jets process final state can be classified according to their flavour. The light flavour jets background can be largely reduced by employing the b-tagging algorithms. The background induced by the V+b jets is mostly irreducible, due to a large production cross-section. The application of the analysis selection such as choosing the jets with the invariant mass close to SM Higgs mass and large p_{T} helps to reduce their contribution.

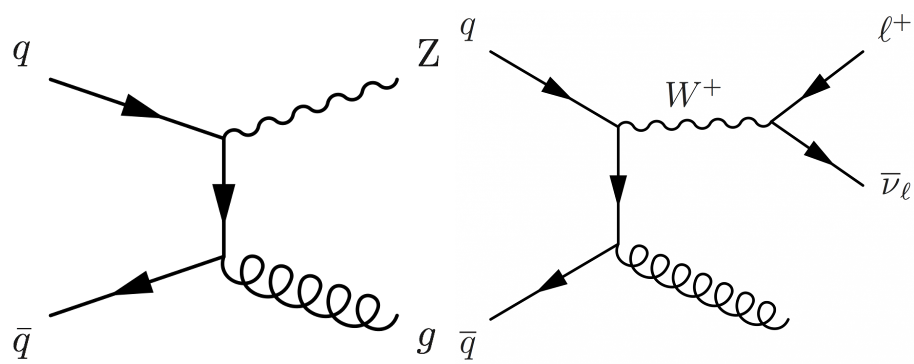


Figure 4.3: V+jets production processes diagrams

In the analysis the contributions from V+c jets, V+b jets and V+udsg jets are modelled separately. The processes are classified at the MC level by the presence of B/D-hadrons with the $p_{\mathrm{T}} > 25$ GeV and $\eta < 2.6$:

- V+udsg: 0 B-hadrons, 0 D-hadrons
- V+c: 0 B-hadrons, >0 D-hadrons

- V+b: >0 B-hadrons

Dedicated control regions are introduced to model this background as detailed in Section 6.1. The V+b-jets processes create the largest background contribution in the 2-lepton channel.

Top quark production

The top quarks are primarily produced in pairs at the LHC and due to the large cross-section contribute in many analyses as the leading background. The Feynman diagrams for the top-quark pairs ($t\bar{t}$) production are shown in Figure 4.4. The leading mechanisms at the LHC are the gluon induced.

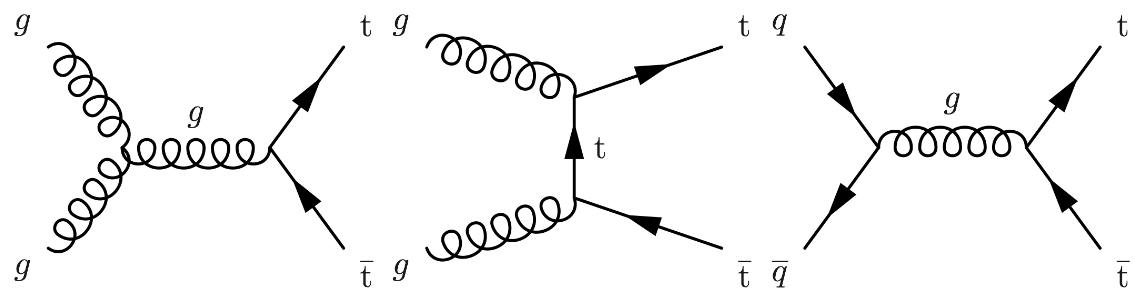


Figure 4.4: $t\bar{t}$ production processes diagrams

Most of the times a top quark decays to a W-boson and a b-quark. The consequent decay mode of the W boson defines how this background contributes. The $t\bar{t}$ production with hadronically decaying W bosons contributes to the 0-lepton channel, but the additional jets activity is higher than in VH production. If one of the W-bosons decays leptonically the final state is similar to 1-lepton channel. The final states with both of the W-bosons decaying to leptons contribute to the 2-lepton channel. The invariant mass of the leptons from W-bosons decay does not form a resonance. This background is dominant in 0-lepton and 1-lepton channels. The $t\bar{t}$ processes are constrained with the use of control regions as detailed in Section 4.5.

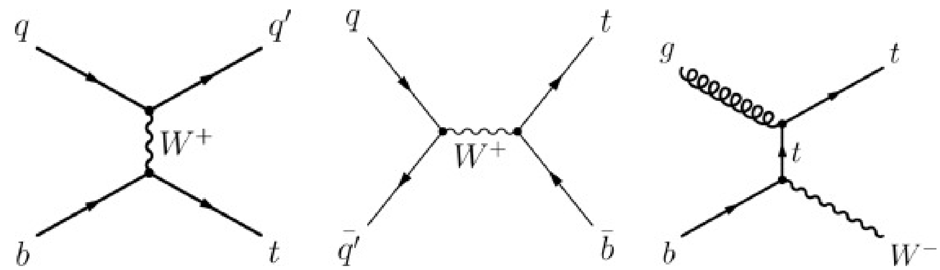


Figure 4.5: Single-top-quark production processes diagrams. From left to right: t-channel, s-channel and tW process.

The single-top quark electroweak production diagrams for t-channel, s-channel and tW process are shown in 4.5. The single-top processes are manifested similarly to the $t\bar{t}$ production, but the kinematics is closer to the signal process, which makes it harder to suppress despite the relatively low production cross-section.

Diboson production

The diboson processes WZ and ZZ, shown in Figure 4.6, can produce the same final state as the VH process, when a Z-boson decays to $b\bar{b}$ and the other vector boson follows the leptonic decay mode. The main observable that helps reducing this background is the invariant mass of b-quark pairs, which is peaked around the Z-boson mass.

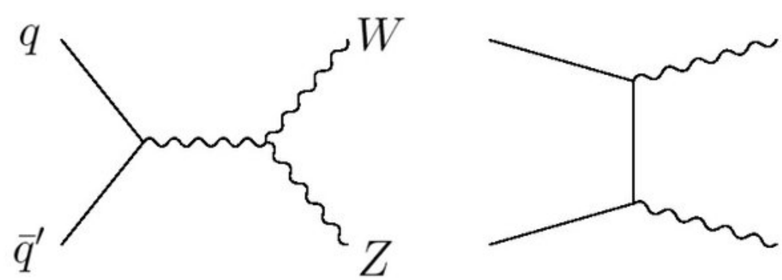


Figure 4.6: Di-boson production Feynman diagrams.

QCD

The QCD events are abundant at the LHC and the b-quark pairs can be easily produced from the QCD interaction. If other particles in the event are mis-reconstructed, the QCD processes can contribute to all channels in the analysis. Anti-QCD selection criteria are used in this analysis, to minimise the multi-jet background contribution as detailed in Section 4.5.

4.3. Samples

4.3.1. Data

This analysis is performed using the full Run 2 CMS data with the combined luminosity of 138 fb^{-1}. In the CMS experiment data is collected using a two-level trigger system, described in Section 3.1.6. For each data taking year a set of un-prescaled High Level Trigger (HLT) paths with the lowest threshold is selected. The paths are summarised in Table 4.1.

In the 0-lepton channel, events are selected with the trigger that requires the presence of MET, defined in Eq. 3.7 and MHT [1] with thresholds 110 GeV in 2016 and 120 GeV in 2017 and 2018. The MET and MHT are constructed with the jets passing tight identification criteria. In the 1-lepton channel, the presence of an isolated lepton is required. The p_{T} threshold for the HLT paths used to trigger the isolated electron are 27 GeV in 2016 and 32 GeV in 2017 and 2018. For the muon paths the thresholds are 27 GeV in 2017 data-taking period and 24 GeV in 2016 and 2018.

For the 2-lepton channel, the double-muon and double-electron triggers are used. The thresholds for the leading p_{T} muon is 17 GeV and 8 GeV for the sub-leading muon. An online requirement on the dimuon invariant mass is applied to remove the contribution from low-mass resonances. For the electron 2-lepton channel, the trigger thresholds are 23 GeV and 12 GeV for the leading and sub-leading electrons, respectively.

In general the offline selection applied in the analysis is more tight than the online selection criteria of the considered HLT paths, therefore the efficiency effects from HLT are not

[1] $\mathrm{MHT} = |-\sum_{\mathrm{jets}} \vec{p_{\mathrm{T}}}|$, where jets are required to satisfy $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 5$.

significant. The central CMS efficiency corrections are used, when possible.

Channel	HLT path
	2016
0-lepton	HLT_PFMET110_PFMHT110_IDTight OR HLT_PFMET120_PFMHT120_IDTight OR HLT_PFMET170_NoiseCleaned OR HLT_PFMET170_BeamHaloCleaned OR HLT_PFMET170_HBHECleaned
1-lepton (e) 1-lepton (μ)	HLT_Ele27_WPTight_Gsf HLT_IsoMu24 OR HLT_IsoTkMu24
2-lepton (e) 2-lepton (μ)	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL OR HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL OR HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ OR HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ
	2017
0-lepton	HLT_PFMET120_PFMHT120_IDTight OR HLT_PFMET120_PFMHT120_IDTight_PFHT60
1-lepton (e) 1-lepton (μ)	HLT_Ele32_WPTight_Gsf_L1DoubleEG HLT_IsoMu27
2-lepton (e) 2-lepton (μ)	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8 OR HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_Mass8
	2018
0-lepton	HLT_PFMET120_PFMHT120_IDTight
1-lepton (e) 1-lepton (μ)	HLT_Ele32_WPTight_Gsf HLT_IsoMu24
2-lepton (e) 2-lepton (μ)	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8

Table 4.1: Triggers used to collect the data samples in the 2016, 2017 and 2018 data-taking periods, for each channel

4.3.2. Simulation

The general procedure for MC simulation is given in Section 3.2.5. This analysis relies on the accurate MC prediction with QCD and EWK corrections. The summary of signal and background samples for the processes, discussed in Section 4.2, is given in Table 4.2.

The signal samples for the WH and qqZH production processes are generated using the full differential prediction computed at NLO in QCD [57, 58, 59] using POWHEG-BOX V2 with the MiNLO procedure [60, 61]. The loop-induced ggZH process is generated at LO but due to a loop present in a tree-level diagram it corresponds to the same order of α_s. The signal prediction is further improved by the application of differential p_T^V NLO electroweak and NNLO QCD corrections calculated in VH@NLO [62, 63, 64] and HAWK v2.0 [65].

The di-boson samples WZ, ZZ and WW are produced at NLO with the MadGraph5_aMC@NLO [66] using FxFx merging scheme [67]. The $t\bar{t}$ and single-top in t-channel are simulated with POWHEG-Box V2, and for single-top in s-channel and tW production the POWHEG-Box V1 is used.

The proper modelling of V+jets background is extremely important in this analysis, hence using the most accurate available prediction is crucial. In the 2017 and 2018 analyses the NLO simulation with MadGraph5_aMC@NLO is used for all V+jets samples. For the 2016 analysis the LO V+jets simulation is used, because the available NLO prediction samples are low in statistics. To improve the statistics of the V+jets MC sample, the simulation is done in exclusive bins of H_T (for LO samples) or p_T^V and number of additional jets (NLO samples). It helps to isolate the phase space regions at the first steps of generation, and significantly increases the number of generated events in comparison to the inclusive simulation, especially in the high p_T^V region. The samples generated in different kinematic regions are merged with the use of the stitching technique [68] in case of phase space overlap to further boost the statistical power.

For all of the LO simulations the NNPDF3.0LO PDF set is used, while for the NLO simulation NNPDF3.0NLO [69]. The hadronization and showering is performed with PYTHIA8 [53], where the underlying event is configured with the CP5 tune [70] for 2018 and 2017 analysis, while for 2016 the CUETP8M1 tune is used.

All particle-level events are passed through a detailed simulation of the CMS detector implemented in Geant4 [55]. For each analysis year a separated set of samples was produced to account for different CMS detector conditions as well as LHC parameters. The generated samples were reweighted to take into account different pileup conditions.

Samples	Generator	QCD and EWK order
gg→ZH, H → $b\bar{b}$,Z	POWHEG-Box V2	LO
qq→ZH(WH) H → $b\bar{b}$	POWHEG-Box V2	NNLO NLO
qq'→WH, H → $b\bar{b}$	POWHEG-Box V2	NNLO NLO
WZ, ZZ	MadGraph5_aMC@NLO	NLO
$t\bar{t}$, single-top t-channel	POWHEG-Box V2	NNLO NLO
Single-top s-channel, tW	POWHEG-Box V1	NNLO NLO
V+jets in 2016 analysis	MadGraph5_aMC@NLO	LO
V+jets in 2017 and 2018 analyses	MadGraph5_aMC@NLO	NLO

Table 4.2: MC simulation summary.

4.4. Event reconstruction

The events passing the HLT selection and after the reconstruction of basic objects such as leptons and jets are then passed through the offline reconstruction procedure detailed in this section.

4.4.1. Higgs candidate

In the resolved analysis the Higgs candidate is reconstructed by combining two b-tagged AK4 jets with the highest b-tag scores. The Higgs candidate jets are required to pass tight pileup rejection cuts, tight jet ID as well as a lepton filter. To account for the energy loss due to the final state radiation (FSR), a recovery algorithm is applied. The jets around ($\Delta R < 0.8$) the two selected b-jets, with $p_{\mathrm{T}} > 20$ GeV, within tracker acceptance, with tight jet ID, lepton filter and pileup rejection are added to the selected b-jets, which are then combined to form a Higgs candidate.

The boosted Higgs candidate is reconstructed from one large radius AK8 jet passing the DeepAK8 b-tagger. The boosted selected jet is required to have $p_{\mathrm{T}} > 250$ GeV. The leptons within a $\Delta\phi = 1.57$ cone are removed to ensure that the leptons from vector boson decay are not mis-identified as the lepton from b-hadron decays.

4.4.2. Vector boson candidate

The vector boson reconstruction is based on the leptons and $E_{\mathrm{T}}^{\mathrm{miss}}$ passing the trigger criteria defined above. In the 0-lepton channel the neutrinos from Z-boson decay are only reconstructed in the transverse plane as $E_{\mathrm{T}}^{\mathrm{miss}}$, and the $\mathrm{p_T^V}$ is therefore defined as $E_{\mathrm{T}}^{\mathrm{miss}}$. The leptonic decay of W-boson in 1-lepton channel is reconstructed in the transverse plane only by combining the isolated lepton with the $p_{\mathrm{T}}^{\mathrm{miss}}$ vector. The Z-boson in the 2-lepton channel is formed from two opposite sign same flavour isolated leptons. The best resolution among all of the channels in $\mathrm{p_T^V}$ due to the full event information is therefore ensured for the 2-lepton channel. The dilepton invariant mass helps to reduce the $\mathrm{t\bar{t}}$ background when the cut around the Z-peak is applied.

4.4.3. Kinematic fit

A good resolution of lepton momenta from $\mathrm{Z}\rightarrow ll$ decay allows to further constrain the 2-lepton channel kinematics to improve the resolution of the dijet system. The kinematics of the whole event including the selected Higgs b-jets, selected leptons, FSR jets and recoil jets[2], was used as an input in the kinematic fit procedure. The following constraints are applied in the fit: the dilepton invariant mass is constrained at 91 GeV with the Gaussian uncertainty of 5 GeV; the transverse momentum of all fitted particles is constrained at 0.

[2] additional jets that are not FSR jets and Higgs candidate jets

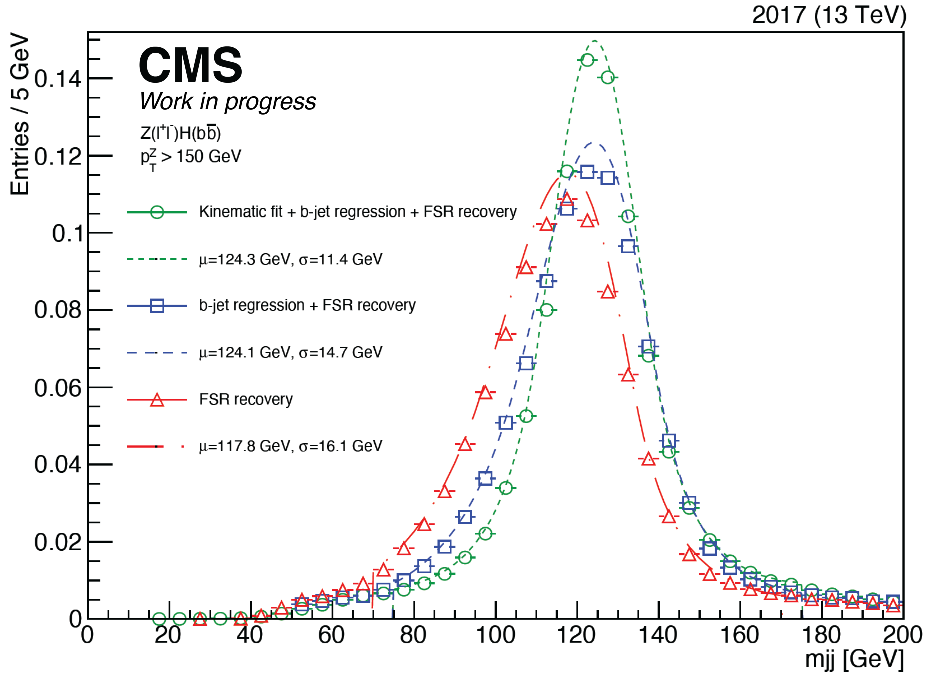


Figure 4.7: The impact of the kinematic fit in the 2-lepton channel shown for signal processes. The dijet invariant mass distribution without kinematic fit is shown in blue, with the kinematic fit in green.

The impact of the kinematic fit on the m_{jj} resolution is shown in Figure 4.7. The dijet invariant mass resolution is compared for the signal process before and after the kinematic fit is applied. The improvement is quite significant: from 14.7 GeV to 11.4 GeV.

4.4.4. Additional event information

After the main event topology is reconstructed additional objects are defined and the related variables are used in the MVA training.

Top quark

The $t\bar{t}$ in semi-leptonic final state can easily fake signal events. To improve the handle on the $t\bar{t}$ background a top-quark mass is defined in every event. The top-quark candidates are reconstructed by combining in 1-lepton channel the selected lepton, the $E_{\mathrm{T}}^{\mathrm{miss}}$ and the closest in angular distance b-jet. To improve the top-quark resolution the neutrino momentum is calculated by constraining the kinematics with the W mass. In Figure 4.8 the reconstructed top quark mass is shown in one of the VH(H $\to b\bar{b}$) 1-lepton signal region where the peak around 172 GeV is visible. This variable is used as an input in the MVA training to help separating the signal from $\mathrm{t\bar{t}}$ background.

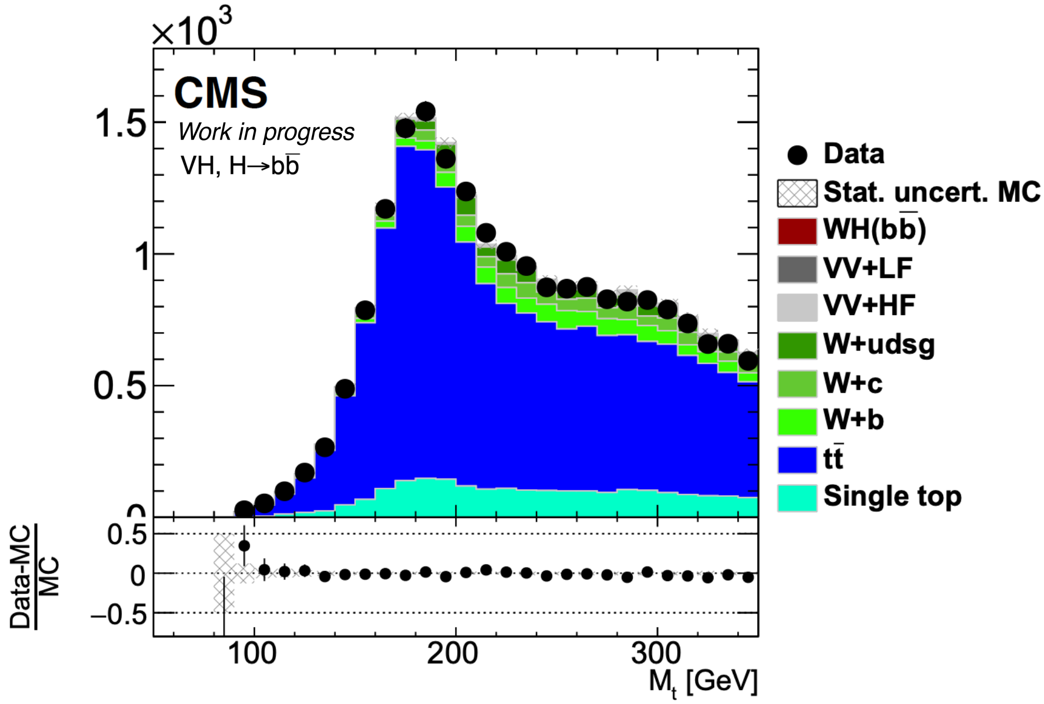


Figure 4.8: Top-quark mass distribution reconstructed in 1-lepton signal region

Additional jets

The additional jets activity in the event is used in many ways in this analysis. It is crucial to differentiate signal region events from $\mathrm{t\bar{t}}$ background events in 0-lepton and 1-lepton channels. It is also used for the definition of the signal regions splitting to align with the STXS scheme as discussed in Section 5. The number of additional jets N_{aj} is also used in the MVA training. The additional jets are defined as the non-Higgs candidate jets with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.5$.

4.5. Resolved analysis selection

Expanding on the general event reconstruction used in the VH(H $\rightarrow b\bar{b}$) analysis given in Section 4.4, in this section the selection will be given in more detail.

4.5.1. $Z \rightarrow \nu\nu$ channel

The events passing the HLT criteria are further filtered before the actual reconstruction starts. This is done to reduce the sample size and keep the events with jets and leptons that can be used in the analysis.

The 0-lepton channel is selected by requiring the MET>170 GeV and no isolated leptons with $p_{\mathrm{T}} > 15$ GeV. The selection on min(MET,MHT)>100 GeV is applied in addition to the cut on MET to mimic the HLT online selection, for which the efficiency is calculated. The Higgs candidate jet with higher momentum (leading jet) is required to have regressed [3] transverse momentum $p_{\mathrm{T}} > 60$ GeV, and for the subleading jet $p_{\mathrm{T}} > 35$ GeV. The Higgs candidate p_{T} is required to exceed 120 GeV, and $m_{jj} < 500$ GeV. The QCD multi-jet background is reduced by requiring the MET to be isolated from jets with $p_{\mathrm{T}} > 30$ GeV: $\Delta\phi(MET, \mathrm{jet}) > 0.5$.

[3] The b-jet energy regression is applied for all b-tagged jets in the analysis (see Section 3.2.3).

Variable	SR	Z + b-jets	Z + light-jets	$t\bar{t}$
Common selection:				
min(MET,MHT)	> 100	-//-	-//-	-//-
$E_{\mathrm{T}}^{\mathrm{miss}}$	> 170	-//-	-//-	-//-
p_{T}^{j1}	> 60	-//-	-//-	-//-
p_{T}^{j2}	> 35	-//-	-//-	-//-
$p_{\mathrm{T}}(\mathrm{jj})$	> 120	-//-	-//-	-//-
$\Delta\phi$(Z,H)	> 2.0	-//-	-//-	-//-
$M(\mathrm{jj})$	> 50, < 500	-//-	-//-	-//-
N_{al}	< 1	-//-	-//-	-//-
Njets close to MET	0	-//-	-//-	-//-
SR and CRs:				
N_{aj}	≤ 1	≤ 1	≤ 1	≥ 2
$M(\mathrm{jj})$	[90-150]	$\notin [90 - 150]$	-	-
$btag_{max}$	> medium	> medium	< medium	> medium
$btag_{min}$	> loose	> loose	> loose	> loose
$\Delta\phi$(pfMET,trkMET)	< 0.5	< 0.5	< 0.5	-
$\min\Delta\phi$(pfMET,J)	-	-	-	$< \pi/2$

Table 4.3: 0-lepton channel selection for control and signal regions

At the next step the events passing 0-lepton channel are sub-categorised into events from signal regions (SR), defined as the region with enhanced signal efficiency, and the control regions, defined to constrain the leading background and enriched in the corresponding background process. The summary of the selection for signal and control regions is given in Table 4.3. The selection procedure is illustrated in Fig 4.9.

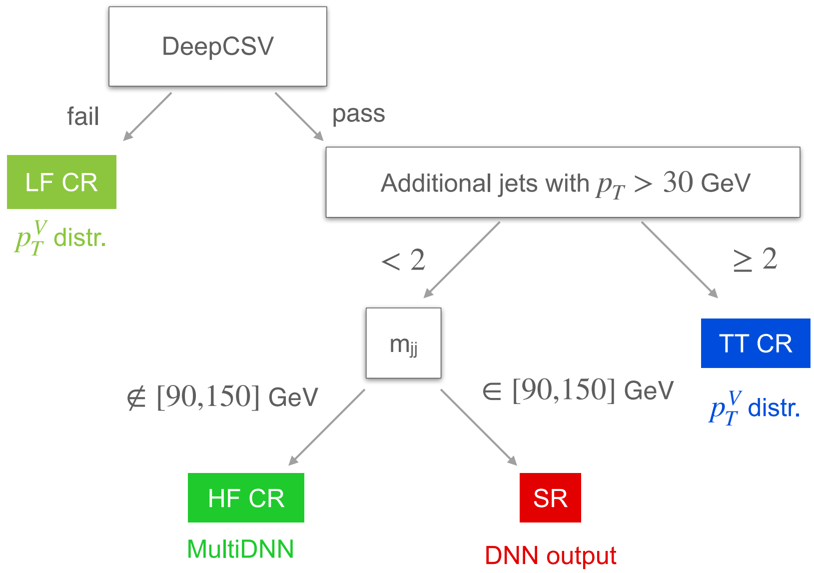


Figure 4.9: The 0-lepton and 1-lepton channels selection scheme.

The signal region selection starts with a requirement on the DeepCSV scores, the leading in b-tagging probability pre-selected b-jet is required to pass the medium DeepCSV working point (see Table 3.2) which correspond to 1% mistag efficiency. The subleading in b-tagging probability jet is required to pass the loose DeepCSV WP. If the cuts on the b-tagging WP are not passed the event is assigned to the V+light flavour jets control region (LF CR). The number of additional jets in the signal region is required to be below 2 $N_{\mathrm{aj}} < 2$, otherwise the event is assigned to the $\mathrm{t\bar{t}}$ control region. Next, the events in the 90-150 GeV dijet invariant mass window are assigned to the signal region, and outside of this window to the V+heavy flavour jets control region (HF CR).

4.5.2. $W \to l\nu$ channel

After passing HLT criteria, the 1-lepton channel events with one isolated lepton are selected by requiring the p_T^V>150 GeV. The explicit filtering of events with additional isolated leptons with $p_T > 15$ GeV is applied. The lepton and MET originating from W-boson decay are expected to be close in the transverse plane, therefore the angular distance between their momenta is required to be $\Delta\phi(\text{lep}, \text{pfMET}) < 2$. The Higgs candidate jets are required to have regressed transverse momentum $p_T > 25$ GeV, and the Higgs candidate p_T is required to exceed 100 GeV.

Variable	SR	W + b-jets	W + light-jets	$t\bar{t}$
Common selection:				
$p_T(\text{jj})$	> 100	-//-	-//-	-//-
$p_T(V)$	> 150	-//-	-//-	-//-
N_{lep}	< 1	-//-	-//-	-//-
p_T^{j1}	> 25	-//-	-//-	-//-
p_T^{j2}	> 25	-//-	-//-	-//-
$\Delta\phi(\text{lep}, \text{pfMET})$	< 2	-//-	-//-	-//-
SR and CRs:				
$btag_{max}$	>medium	>medium	[loose-medium]	>tight
$btag_{min}$	> *loose*	-	-	-
$M(\text{jj})$	[90,150]	[150,250] and <90	<250	< 250
N_{aj}	< 2	< 2	-	>1
$\sigma(\text{MET})$	-	> 2	> 2	-
$\Delta\phi(H, V)$	> 2.5	-	-	-

Table 4.4: Definition of the SR and CR for the 1-lepton channel resolved selection.

The following categorisation of events into signal and control regions is very similar to the procedure defined for 0-lepton channel illustrated in Fig 4.9. The complete set of selection criteria is detailed in Table 4.4.

The signal region is defined to have the leading and sub-leading jets to be above the medium and loose working points. If the cuts on the b-tagging WP are not passed the event is assigned to V+light flavour jets control region (LF CR). The number of additional jets in the signal region is required to be below 2 $N_{\text{aj}} < 2$, otherwise the event is assigned to the

$t\bar{t}$ control region. Next, the events in 90-150 GeV dijet invariant mass window are assigned to the signal region, and outside of this window to the V+heavy flavour jets control region (HF CR). In addition, the Higgs boson and W boson candidates are expected to be recoiling against each other, so the SR purity is improved with a criteria on the angular separation between them: $\Delta\phi(H, V) > 2.5$.

4.5.3. $Z \rightarrow ll$ channel

In the 2-lepton channel the presence of exactly two isolated leptons passing HLT criteria is required. The kinematic fit is performed to improve the resolution in the dijet system. The Z-boson momentum $\mathrm{p_T^V}$ is required to exceed 75 GeV. Then the Higgs candidate jets are selected with $p_\mathrm{T} > 20$ GeV, and the Higgs candidate p_T is required to exceed 100 GeV.

Variable	SR	Z + b-jets	Z + light-jets	$\mathrm{t\bar{t}}$
Common selection:				
p_T^{j1}	> 20	-//-	-//-	-//-
p_T^{j2}	> 20	-//-	-//-	-//-
$p_\mathrm{T}(V)$	> 75	-//-	-//-	-//-
M(jj)	$[50, 250]$	-//-	-//-	-//-
SR and CRs				
$btag_{max}$	>medium	>medium	<loose	>tight
$btag_{min}$	>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 4.5: Definition of the SR and CR for the 2-lepton channel resolved selection.

The following categorisation of events into signal and control regions is illustrated in Fig 4.10. The complete set of selection criteria is detailed in Table 4.5.

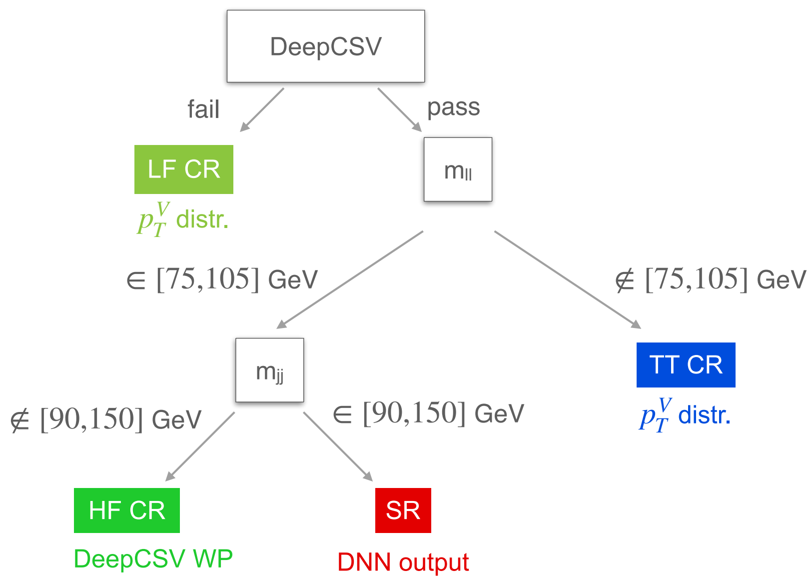


Figure 4.10: The 2-lepton channel selection scheme.

The signal region is defined to have the leading and sub-leading jets to be above the medium and loose working points. If the cuts on the b-tagging WP are not passed the event is assigned to V+light flavour jets control region (LF CR). To select $Z \to ll$ decays the invariant mass of dilepton system is constrained between 75 and 105 GeV, otherwise the events are assigned to $t\bar{t}$ control region. Next, the events in 90-150 GeV dijet invariant mass window are assigned to the signal region, and outside of this window to the V+heavy flavour jets control region (HF CR). In addition, the Higgs boson and Z boson candidates are expected to be recoiling against each other, so the purity of SR and V+jets CRs is improved with a criteria on the angular separation between the Higgs and Z: $\Delta\phi(H, Z) > 2.5$.

4.6. Boosted analysis selection

The vector boson selection in the boosted analysis follows exactly the same procedure as for the resolved analysis described in Section 4.5. The Higgs boosted decay topology is considered for the vector boson momentum range of $p_T^V > 250$ GeV in all analysis channels. In this section the special features of the boosted selection will be summarised.

The boosted Higgs boson candidate is reconstructed using the AK8 jets with $p_T > 250$ GeV, $\eta < 2.5$ and soft-drop mass m_{SD} (see Section 3.2.3) above 50 GeV. The boosted jet tagging is performed using DeepAK8 algorithm [71], described in Section 3.2. The *bbVsLight* DeepAK8 node is used designed to tag the b-jets against the light-flavor jets. The same strategy for the signal and control region definition is used as for the resolved analysis. The cut on DeepAK8 discriminant at 0.8 is used to differentiate LF CR from the rest of the regions with heavy flavour jets, i.e. SR, HF CR and $t\bar{t}$ CR.

In 0-lepton and 1-lepton channels the additional jets activity is also used to separate the $t\bar{t}$ CR events from SR and HF CR, but the variable is adapted for the boosted Higgs decay topology. The additional jets are defined as b-jets, with $p_T > 25$ GeV and $\eta < 2.5$, passing medium DeepCSV working point and not in vicinity of AK8 jet ($\Delta R(\mathrm{AK8jet}, \mathrm{add.\ jet}) > 0.8$). The same Higgs mass window is applied on the m_{SD} variable to select the signal events and define the side-band region for the HF CR. The selection criteria are summarised in Table 4.6.

Variable	SR	V + HF	V + LF	$t\bar{t}$
$p_T(V)$	> 250	> 250	> 250	> 250
$p_T(H)$	> 250	> 250	> 250	> 250
DeepAK8(bbVsLight)	> 0.8	> 0.8	< 0.8	> 0.8
m_{SD}	$\in [90, 150]$	$\notin [90, 150]$	> 50	> 50
N_{aj}	$= 0$	$= 0$	$= 0$	> 1

Table 4.6: Boosted 0-lepton and 1-lepton channels selection for the control regions and the signal regions.

In the 2-lepton channel, as in the resolved analysis, instead of the additional jets the Z mass window is applied for dilepton system as detailed in Table 4.7.

Variable	SR	Z + HF	Z + LF	$t\bar{t}$
$p_T(V)$	> 250	> 250	> 250	> 250
$p_T(H)$	> 250	> 250	> 250	> 250
DeepAK8(bbVsLight)	> 0.8	> 0.8	< 0.8	> 0.8
$\mathrm{m_{SD}}$	$\in [90, 150]$	$\notin [90, 150]$	> 50	> 50
m_{ll}	$\in [75, 105]$	$\in [75, 105]$	-	$\notin [75, 105]$

Table 4.7: Boosted 2-lepton channel selection for the control regions and the signal regions.

The events passing both the resolved and boosted analyses selection were studied in detail, and by comparing the overall analysis sensitivity under different options it was decided that the best choice is to assign the overlap events to the resolved categories unless they enter the boosted signal region selection.

4.7. MVA

The signal region selection enriches in signal the phase space. The multivariate analysis techniques allow to further improve the signal versus background discrimination power. Three multivariate methods are used in this analysis: a deep neural networks (DNN) binary classifier for the resolved signal region, a multi-class DNN in the V+HF control regions to improve the separation of different backgrounds, and a boosted decision tree (BDT) technique for the binary classification in the boosted signal region.

4.7.1. DNN

For the resolved Higgs decay topology, a signal vs. background DNN classifier is trained for each channel separately. The output in the signal region is used in the fit for all the channels. For the 0-lepton and 1-leptons channel $V + HF$ control region, a multi-class DNN classifier is used.

The tensorflow framework [72] was used to train a 6 hidden layer DNN classifier, with each layer having 512, 256, 128, 64, 64 and 64 nodes. Convergence improving features such as dropout [73] and stochastic optimization [74] are used. As the last layer the softmax function is added to interpret the output as a probability.

For 2-classes DNN all signal processes were grouped into signal class, and all of the background processes into a background class. In the multi-class DNN instead of signal and background output nodes, the classification is performed according to the 5 leading background processes listed in Table 4.8. A background class is assigned to each event, if the corresponding class probability is the largest.

0	V+udsg
1	V+c
2	V+b
3	Single top
4	$t\bar{t}$

Table 4.8: Classes used for the 0/1-lepton multi-DNN classifier.

In both 2-class DNN and multi-class DNN the same architecture and the same set of input features are used (Table 4.9). The agreement of data and simulation for all of the MVA input variables is studied and found to be sufficient (Figures 4.12, 4.11).

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 vector boson and higgs boson		✓	✓	✓
$\Delta\phi(V, H)$	azimuthal angle between vector boson and dijet directions	✓	✓	✓	✓
$\mathrm{btag}_{\mathrm{max}}$ WP	1,2,3 if b-tagging discriminant (DeepCSV) score of leading jet is above T, M, L WP resp.	✓	✓	✓	✓
$\mathrm{btag}_{\mathrm{min}}$ 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	✓	✓		
$\mathrm{btag}_{\mathrm{max}}(add)$	maximum btagging discriminant score among additional jets	✓			
$\mathrm{pT}_{\mathrm{max}}$(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			✓	
$\Delta R(V, H)$	angular separation between vector boson and Higgs boson			✓	✓
$\Delta R(V, H)$	angular separation between leading and sub-leading jets			✓	
$\sigma(m(jj))$	resolution of dijet invariant mass				✓
N_{rec}	number of recoil jets				✓

Table 4.9: List of input variables used in the training.

4.7.2. BDT

The boosted decision trees classifier was used for the signal vs. background classification in the boosted signal regions. Both boosted and overlap events were used in the training, and for the overlap events the boosted variables: soft-drop mass of the AK8 jet, transverse momentum of the AK8 jet, $\mathrm{p_T^V}$, DeepAK8 score; the resolved variables listed in Table 4.9 were included. Further details on the MVA training and performance for this analysis are given in [75].

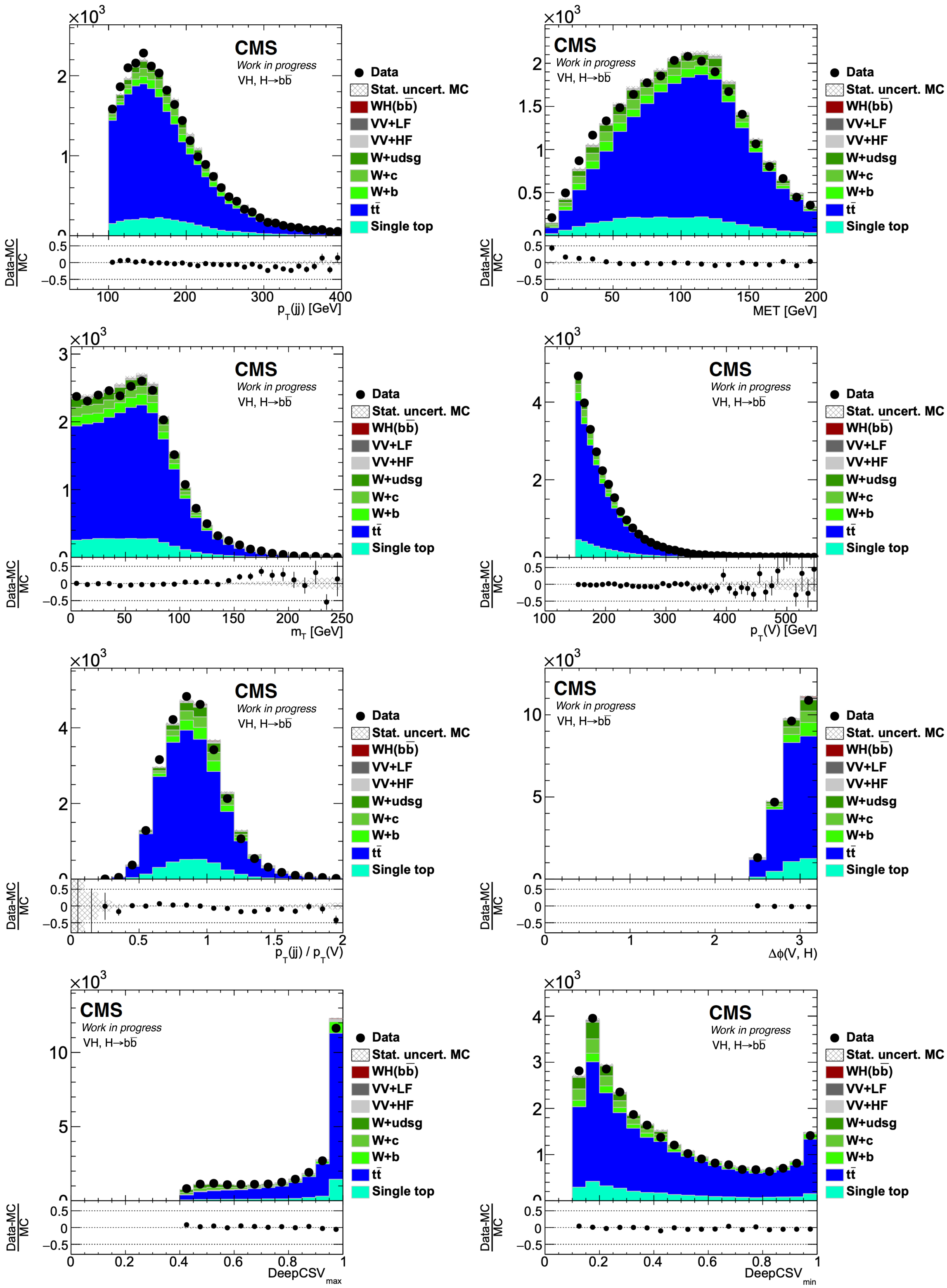


Figure 4.11: 1-lepton signal region pre-fit distribution of MVA input variables

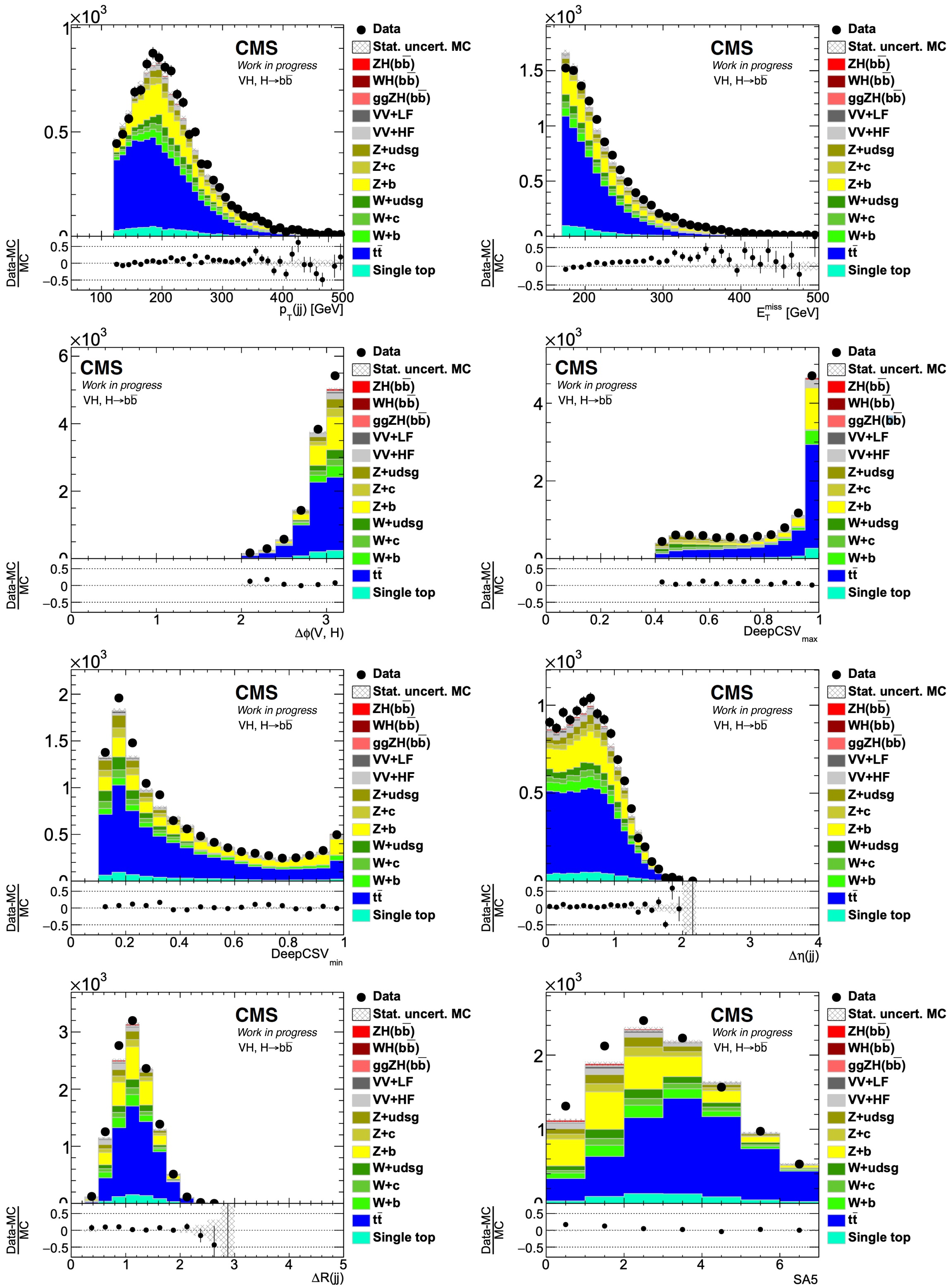


Figure 4.12: 0-lepton signal region pre-fit distribution of MVA input variables

4.8. Control region observables

The DNN and BDT score distributions evaluated on data and simulation were used as the templates in the fit for the signal regions. For each control region the observable that gives the best separation for the background processes is chosen. For the resolved V+HF control regions in 0-lepton and 1-lepton channel the multi-class DNN output was used, the output classes are shown in Table 4.8.

For the 2-lepton channel the observable is defined based on the $\text{DeepCSV}_{\text{min}}$ and $\text{DeepCSV}_{\text{max}}$ values as described in Table 4.10 and shown in Figure 4.13. In the V+LF control regions

value	DeepCSV max	DeepCSV min
0	$<$ Tight	$<$ Medium
1	$<$ Tight	$>$ Medium
2	$>$ Tight	$<$ Medium
3	$>$ Tight	$>$ Medium, $<$ Tight
4	$>$ Tight	$>$ Tight

Table 4.10: Variable used for template fit in 2-lepton HF control region.

the p_T^V distributions are employed, while in the $t\bar{t}$ CR only the inclusive processes yields are considered. Examples of resolved control regions distributions are shown in Figure 4.13.

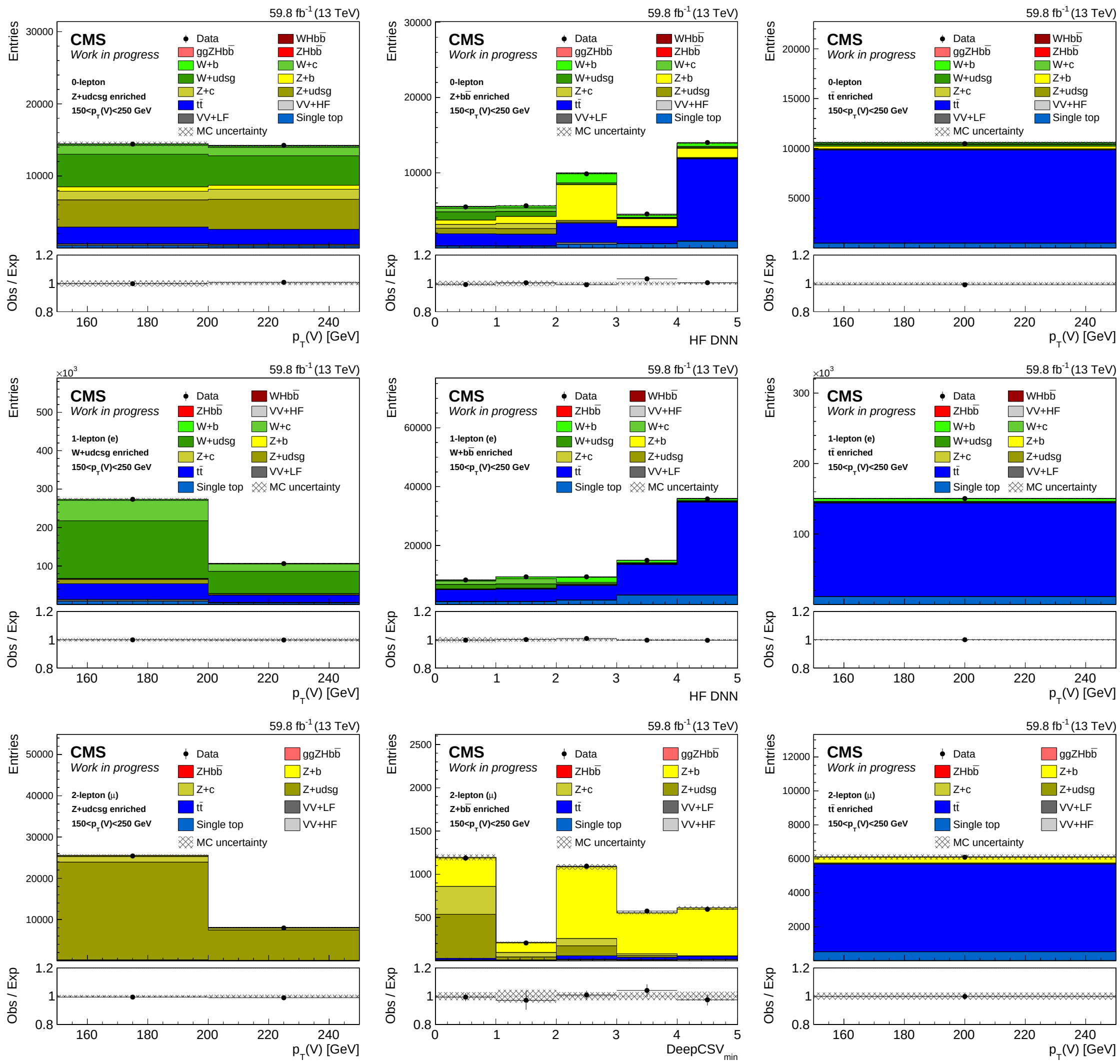


Figure 4.13: Control region post-fit distributions in the resolved analysis. Top, middle and bottom rows corresponds to the 0-lepton, 1-lepton and 2-lepton channels respectively. The left, middle and right columns corresponds to the V+LF, V+HF and $t\bar{t}$ control regions respectively.

In the boosted analysis for all control regions the DeepAK8 score is used as shown in Figure 4.14.

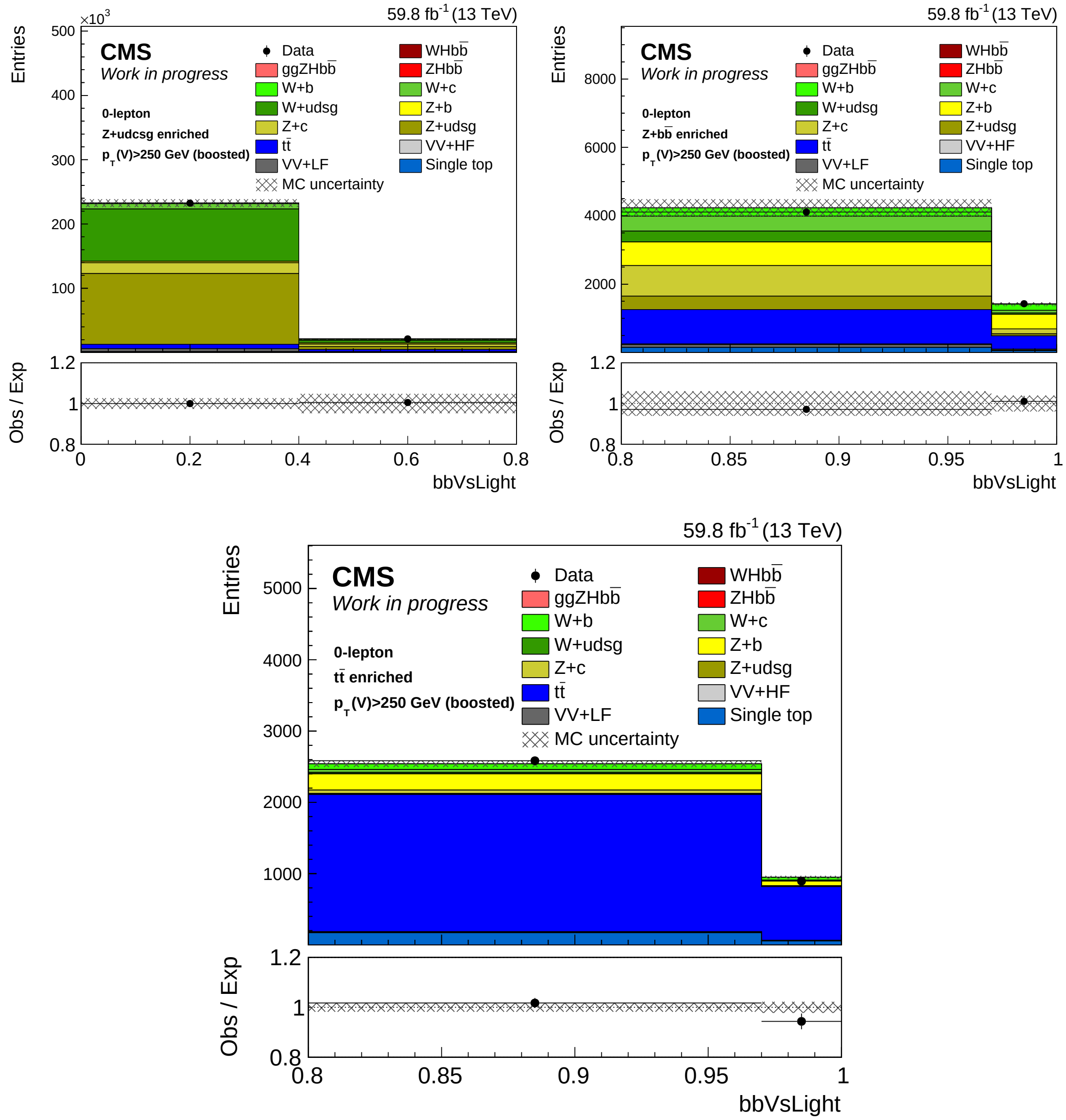


Figure 4.14: Control region post-fit distributions in the boosted analysis in 0-lepton channel. The left, middle and right columns corresponds to the V+LF, V+HF and $t\bar{t}$ control regions respectively.

CHAPTER 5

Simplified Template Cross Section framework

5.1. Introduction

The Simplified Template Cross Section framework was developed after the first measurements of the Higgs couplings with the Run 1 data at the LHC by the theoretical and experimental communities including the ATLAS and CMS collaborations [15]. The main objectives are the high granularity of future measurements and the possibility to combine and reinterpret the published results.

The Run 1 SM Higgs measurements were usually either fiducial, signal strength ($\mu = \frac{[\sigma B]_{obs}}{[\sigma B]_{SM}}$) or coupling modifiers ($\kappa^2 = \frac{\sigma^{obs}}{\sigma^{SM}}$) measurements. The μ and κ measurements allow the usage of multivariate methods to improve the sensitivity, but by definition they are highly dependent on the theoretical predictions and uncertainties. This of course complicates the reinterpretations of these measurements if the theoretical predictions are improved, for example. The fiducial measurements are less model-dependent, but the usage of sophisticated MVA techniques or complicated analysis categorisation are limited.

The STXS framework combines the best features of both fiducial and signal strength measurements and also provides new features such as identification of a BSM-specific phase space. The STXS kinematic regions, also called bins, are defined at the generator level for each production process. The ultimate goals are to maximise the experimental sensitivity, to minimize the dependence on theoretical uncertainties, and to isolate BSM effects. The results are presented as cross section measurements which allows disentangling theoretical uncertainties from the measurement, leaving only small dependence due to acceptance effects. The fiducial selection for the Higgs decay channels is not applied, hence the combination of cross section measurements in different Higgs final states is straightforward. The evolution of the STXS bins is predefined for each production process to accommodate for increasing statistics. It is developed in so-called stages, with the stage-0 corresponding to a single bin for each production process, and the consequent stages with increased granularity in defined kinematic variables. The variables chosen for each production process individually are shown in Fig. 5.1.

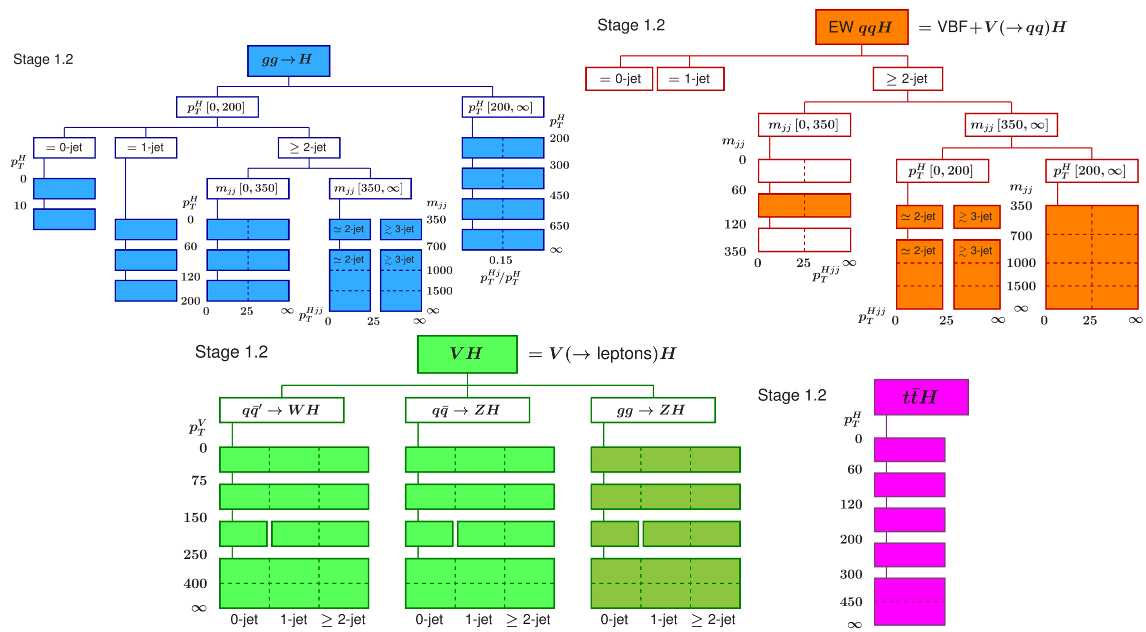


Figure 5.1: STXS categorisation for the leading Higgs production processes.

The most relevant STXS category for this thesis is the VH leptonic process. For this production mode the largest contribution comes from $H \to b\bar{b}$ decay, the sensitivity of which has a large impact on the bins definition. In Fig. 5.2 the latest recommended stage 1.2 categorisation for VH mode is shown. The STXS bins are defined using the transverse momenta of vector boson $\mathrm{p_T^V}$ and the number of additional jets. The VH process is split into three channels $qq \to ZH$, $qq \to WH$ and $gg \to ZH$. Each of them is consequently separated into four $\mathrm{p_T^V}$ regions: 0–75, 75–150, 150–250, >250 GeV. The 150–250 GeV bin is additionally split by the number of additional jets (n_{jet}) with $p_{\text{T}} > 30$ GeV: 0 jets, and at least one additional jet. The $\text{p}_{\text{T}}^{\text{V}}$>250 GeV bin represents the region sensitive to the BSM effects. The dashed boundaries are defined to consider further splitting if possible experimentally. The STXS bins are supposed to be merged by the experiments if a lack of sensitivity for the proposed binning is observed.

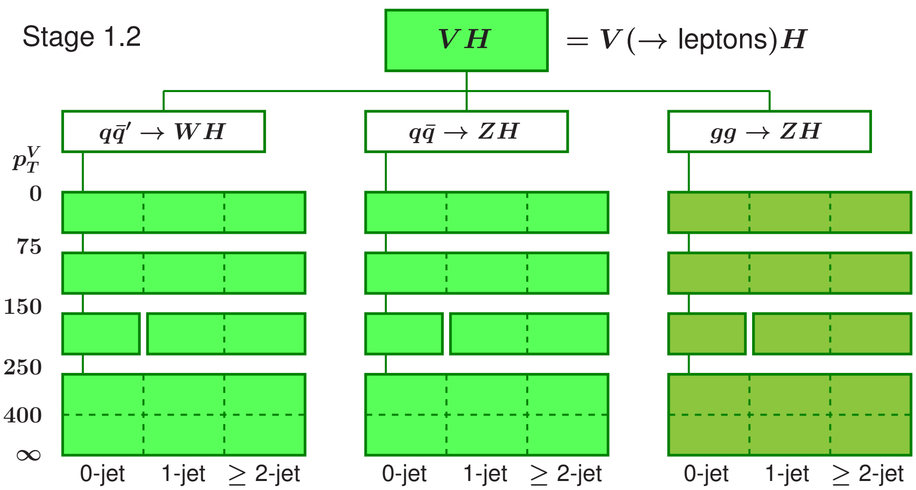


Figure 5.2: STXS bins stage 1.2 VH leptonic

To apply the STXS categorisation, the particle-level events hadronised by `PYTHIA8` are passed through the `Rivet` package [76], using the methods implemented within the `HiggsTemplateCrossSections` class. The `HiggsTemplateCrossSections` routine was designed to provide a tool to perform generator particle-level Higgs analysis. Within this routine the jets are formed with the anti-k_T algorithm implemented in the FastJet package [77]. The high level observables are calculated to proceed further and apply the process classification. The Higgs boson decay products are ignored and do not affect the STXS categorisation.

5.2. STXS scheme and categories in VH(H $\to b\bar{b}$) analysis

The general STXS stage 1.2 for VH process is discussed in the previous section and summarised in Figure 5.2. Due to the sensitivity limitations some of the STXS bins are merged in the region $\text{p}_{\text{T}}^{\text{V}} < 250$ GeV. However, the combination of resolved and boosted analyses allows to perform more granular measurement in the $p_T^V > 250$ GeV region.

The measured STXS processes are shown in Figure 5.3. The WH STXS process in $p_T^V < 150$ GeV is not measured and fixed at the SM value. In the $p_T^V < 150$ GeV region the QCD multi-jet background starts playing a major role so it is not considered in the analysis. Then the $150 < p_T^V < 250$ GeV STXS bin is measured inclusively, since at the current level of precision the separation of the WH process in number of additional jets is not feasible.

The ZH process is not measured in the region $p_T^V < 75$ GeV, and in the $75 < p_T^V < 150$ GeV only $Z \to ll$ channel is contributing. In addition the quark-induced (qqZH) and gluon-induced (ggZH) ZH processes are merged[1]. Overall, we obtain three STXS bins for the WH process: $150 < p_T^V < 250$ GeV, $250 < p_T^V < 400$ GeV and $p_T^V > 400$ GeV; and five bins for ZH[2] process: $75 < p_T^V < 150$ GeV, $150 < p_T^V < 250$ GeV 0 jets, $150 < p_T^V < 250$ GeV ≥ 1 jet, $250 < p_T^V < 400$ GeV and $p_T^V > 400$ GeV.

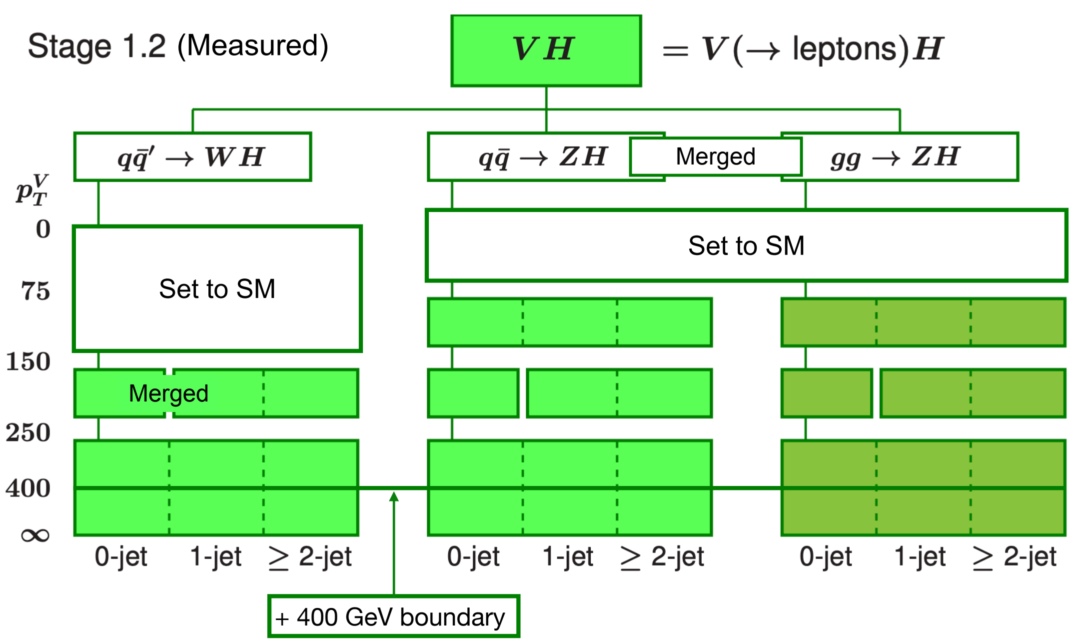


Figure 5.3: VH STXS stage 1.2 scheme adapted to the VH($H \to b\bar{b}$) analysis sensitivity limitations

[1]A single POI for ggZH and qqZH in each STXS bin
[2]From now on the combination of qqZH and ggZH will be referred to as ZH.

The signal regions defined in Section 4 were partitioned using the reconstructed p_T^V and number of additional jets observables to align with the STXS bins. In Figure 5.4 the confusion matrix between reconstructed signal regions and particle-level STXS categories is shown. The ultimate goal of STXS analysis is to achieve the diagonality of this matrix, which then improves the sensitivity to the individual STXS cross-section. The current level of reco- vs particle-level categories matching is a result of the good p_T^V resolution.

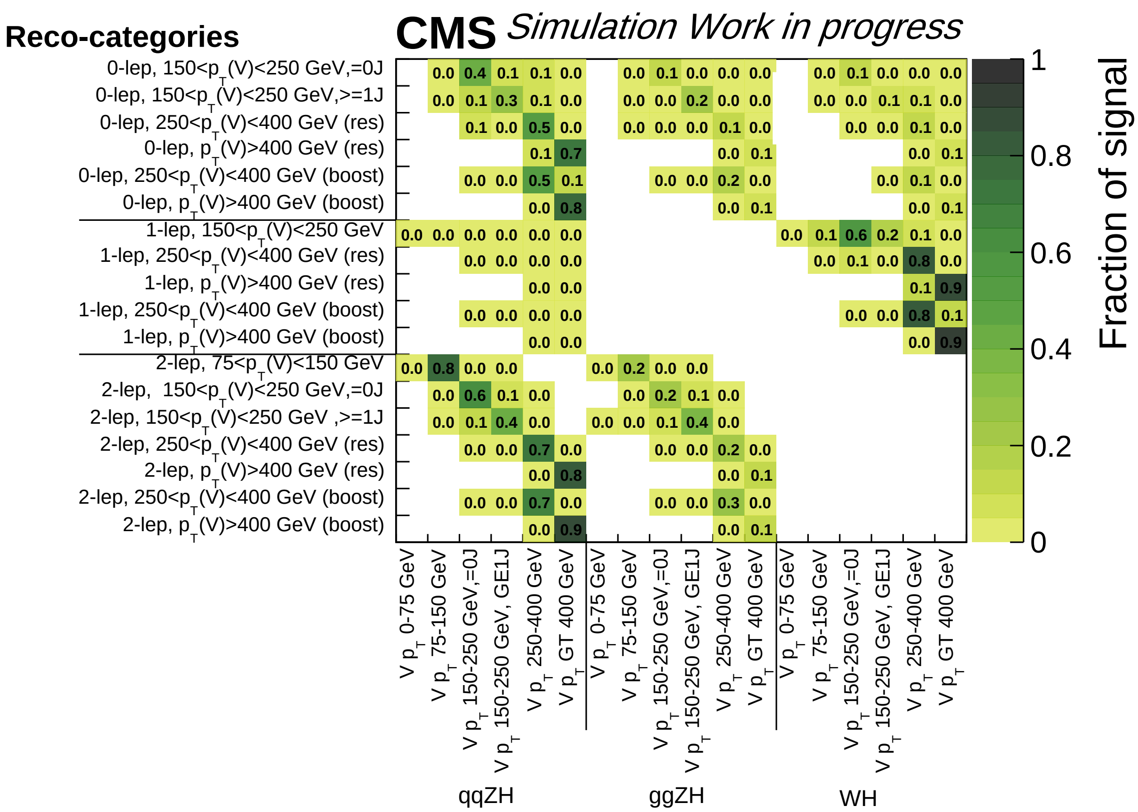


Figure 5.4: STXS signal processes fraction inside reconstruction level analysis signal categories.

5.3. STXS uncertainties

In this section the theoretical uncertainties derived for the STXS VH($H \to b\bar{b}$) measurement are discussed. The considered components are the STXS migration uncertainties, applied as normalisation uncertainties for each STXS signal process, and the acceptance uncertainty, which is introduced as a shape variation and acts as a residual theoretical uncertainty for the merged STXS bins.

5.3.1. Migration uncertainties from QCD scale variations.

The migration uncertainties were derived following the strategy defined in [78] from the `POWHEG` VH signal prediction and cross checked with the uncertainties from the fixed order (FO) and resummation scale variations available in NNLO `GENEVA` prediction.

Considering both solid and dashed bin boundaries in the STXS 1.2 scheme for VH process (Figure 5.2) six uncertainty components can be defined to model migrations across the bins: four of them Δ_X (X = 75, 150, 250, 400 GeV) are induced by the $\mathrm{p_T^V}$ bin boundaries, while the other two, Δ_1 and Δ_2, by the n_{jet} bins boundaries.

To estimate the uncertainties the renormalization (μ_R) and factorisation (μ_F) scales are varied by the factor of 2 with respect to the nominal values:

$$(\frac{\mu_R}{\mu_R^{nom}}, \frac{\mu_F}{\mu_F^{nom}}) = [(1,2);(2,1);(2,2);(1,0.5);(0.5,1);(0.5,0.5)] \tag{5.1}$$

The scale variations and the nominal yields are shown in Figure 5.6 (Figure 5.5) for all WH and qqZH (ggZH) processes. It is evident that the scale variations for ggZH process introduce significantly larger uncertainties if compared with qqZH and WH processes, which is expected due to a loop present in a tree level ggZH diagram.

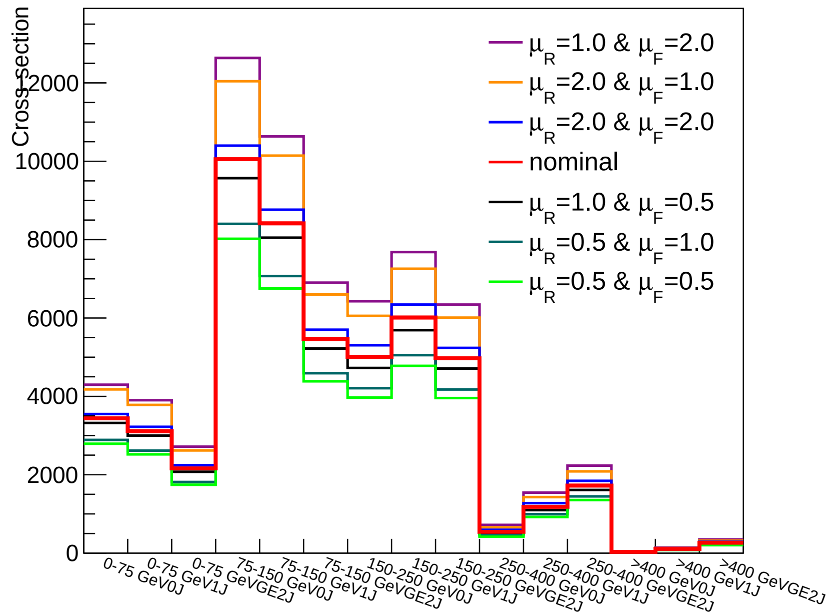


Figure 5.5: Scale variations for ggZH process.

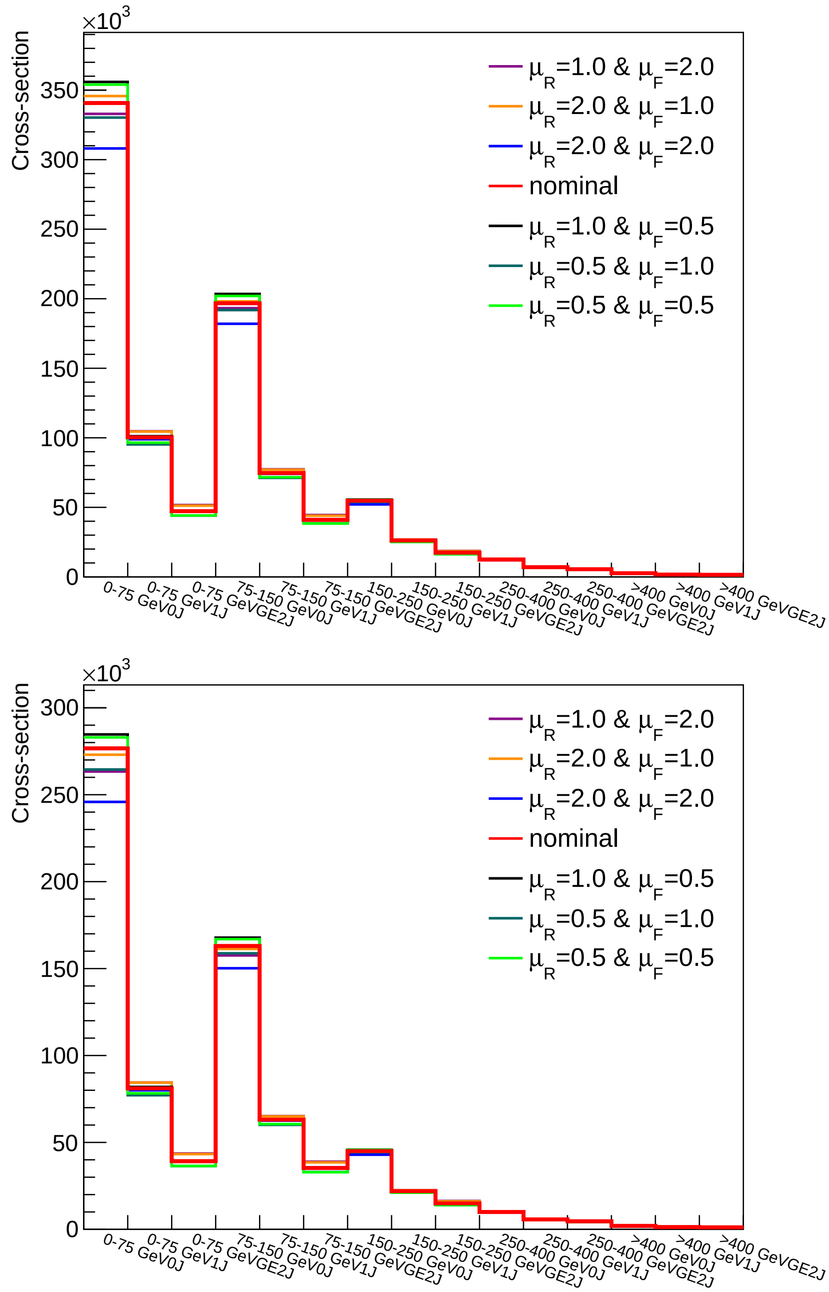


Figure 5.6: Scale variations for WH, qqZH processes.

The absolute p_T^V migration uncertainties Δ_X (X = 75, 150, 250, 400 GeV) are defined as the maximum deviations from the nominal yields under the scale variations listed in Eq. 5.1 in the corresponding inclusive phase space of $p_T^V > X$ GeV.

The Δ_{150}, Δ_{250} and Δ_{400}, were multiplied by 0.5 scale factor, to ensure that they are of the same order as the uncertainty on the total cross section.

The relative impact of each uncertainty component is calculated assuming the short-range correlation scheme, i.e. the absolute Δ_X variation is normalised by the total cross section above the considered boundary, shown in Table 5.1.

$\mathrm{p_T^V}$ bin (GeV)	Δ_{75}	Δ_{150}	Δ_{250}	Δ_{400}
[0, 75[	$-\Delta_{75}/\sigma_{[0,75[}$	0	0	0
[75, 150[	$\Delta_{75}/\sigma_{[75,\infty[}$	$-\Delta_{150}/\sigma_{[75,150[}$	0	0
[150, 250[	$\Delta_{75}/\sigma_{[75,\infty[}$	$\Delta_{150}/\sigma_{[150,\infty[}$	$-\Delta_{250}/\sigma_{[150,250[}$	0
[250, 400[	$\Delta_{75}/\sigma_{[75,\infty[}$	$\Delta_{150}/\sigma_{[150,\infty[}$	$\Delta_{250}/\sigma_{[250,\infty[}$	$-\Delta_{400}/\sigma_{[250,400[}$
[400, ∞[	$\Delta_{75}/\sigma_{[75,\infty[}$	$\Delta_{150}/\sigma_{[150,\infty[}$	$\Delta_{250}/\sigma_{[250,\infty[}$	$\Delta_{400}/\sigma_{[400,\infty[}$

Table 5.1: Short-range correlation scheme (scheme-2) uncertainty parametrization for STXS $\mathrm{p_T^V}$ bins.

The uncertainties generated by n_{jet} bin boundaries Δ_X (X = 1, 2) are evaluated using the same approach as described above for $\mathrm{p_T^V}$ bin boundaries. The $\mathrm{p_T^V}$ bin boundaries and n_{jet} bin boundaries are considered as independent components and the corresponding uncertainties are uncorrelated. The calculation starts with the μ_R and μ_F scale variations to obtain the maximum absolute deviations from the nominal, the Δ_1 and Δ_2. Then the impacts of Δ_1 and Δ_2 on the n_{jet} bins are evaluated using the scheme-2 correlation which is shown in Table 5.2. The n_{jet} bin uncertainties are calculated in each $\mathrm{p_T^V}$ bin.

n_{jets} bin	Δ_1	Δ_2
0 jets	$-\Delta_1/\sigma_{n_{\mathrm{jets}}=0}$	0
1 jet	$\Delta_1/\sigma_{n_{\mathrm{jets}}\geq 1}$	$-\Delta_2/\sigma_{n_{\mathrm{jets}}=1}$
$\geq$ 2 jets	$\Delta_1/\sigma_{n_{\mathrm{jets}}\geq 1}$	$\Delta_2/\sigma_{n_{\mathrm{jets}}\geq 2}$

Table 5.2: Short-range correlation scheme (scheme-2) uncertainty parametrization for STXS n_{jet} bins.

In Figure 5.14 the distributions of migration uncertainties in STXS bins are shown. The total migration uncertainty for $\mathrm{p_T^V}$ bin boundaries is calculated as a quadrature sum of the individual Δ_X components, which is then correlated among the n_{jet} bins within a $\mathrm{p_T^V}$

bin, and summed in quadrature with Δ_1 and Δ_2 components to obtain the total migration uncertainty for each STXS bin. The numbers are summarised in Table 5.3 for the ZH production process. The uncertainties for WH and ggZH are also evaluated and shown in Figure 5.8. The WH uncertainties are of the same order as those obtained for qqZH production mode. The ggZH uncertainties are significantly larger, especially in high p_T^V region, the numbers are summarised in Table 5.4

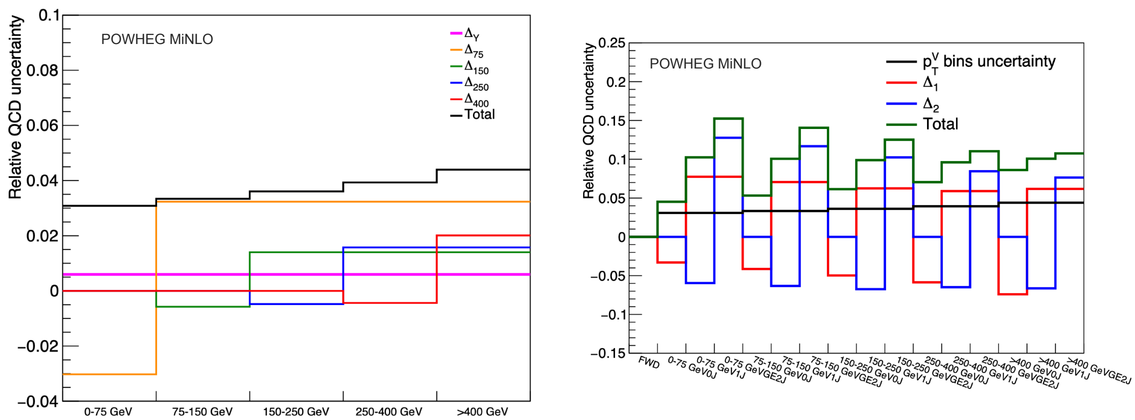


Figure 5.7: Relative QCD scale uncertainties in STXS bins for ZH process (POWHEG).

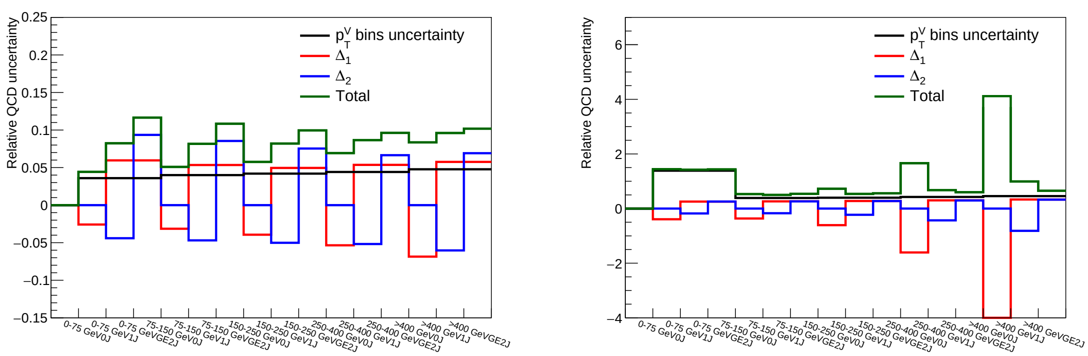


Figure 5.8: Relative QCD scale uncertainties in STXS bins for WH (left) and ggZH (right) processes (POWHEG).

STXS bin	Δ_{75}	Δ_{150}	Δ_{250}	Δ_{400}	Δ_1	Δ_2
[0, 75[GeV 0 jets	-0.037	0	0	0	-0.028	0
[0, 75[GeV 1 jets	-0.037	0	0	0	0.065	-0.054
[0, 75[GeV $\geq$ 2 jets	-0.037	0	0	0	0.065	0.112
[75, 150[GeV 0 jets	0.04	-0.005	0	0	-0.035	0
[75, 150[GeV 1 jets	0.04	-0.005	0	0	0.058	-0.058
[75, 150[GeV $\geq$ 2 jets	0.04	-0.005	0	0	0.058	0.103
[150, 250[GeV 0 jets	0.04	0.013	-0.0042	0	-0.044	0
[150, 250[GeV 1 jets	0.04	0.013	-0.0042	0	0.053	-0.062
[150, 250[GeV $\geq$ 2 jets	0.04	0.013	-0.0042	0	0.053	0.091
[250, 400[GeV 0 jets	0.04	0.013	0.014	-0.004	-0.057	0
[250, 400[GeV 1 jets	0.04	0.013	0.014	-0.004	0.055	-0.062
[250, 400[GeV $\geq$ 2 jets	0.04	0.013	0.014	-0.004	0.055	0.077
[400, ∞[GeV 0 jets	0.04	0.013	0.014	0.0196	-0.072	0
[400, ∞[GeV 1 jets	0.04	0.013	0.014	0.0196	0.059	-0.066
[400, ∞[GeV $\geq$ 2 jets	0.04	0.013	0.014	0.0196	0.059	0.073

Table 5.3: Migration uncertainty values for qqZH process in fine STXS scheme (stage 1.2 including dashed boundaries).

STXS bin	Δ_{75}	Δ_{150}	Δ_{250}	Δ_{400}	Δ_1	Δ_2
[0, 75[GeV 0 jets	-1.36	0	0	0	-0.39	0
[0, 75[GeV 1 jets	-1.36	0	0	0	0.26	-0.18
[0, 75[GeV $\geq$ 2 jets	-1.36	0	0	0	0.26	0.26
[75, 150[GeV 0 jets	0.27	-0.12	0	0	-0.36	0
[75, 150[GeV 1 jets	0.27	-0.12	0	0	0.26	-0.17
[75, 150[GeV $\geq$ 2 jets	0.27	-0.12	0	0	0.26	0.26
[150, 250[GeV 0 jets	0.27	0.14	-0.04	0	-0.61	0
[150, 250[GeV 1 jets	0.27	0.14	-0.04	0	0.28	-0.23
[150, 250[GeV $\geq$ 2 jets	0.27	0.14	-0.04	0	0.28	0.28
[250, 400[GeV 0 jets	0.27	0.14	0.15	-0.02	-1.61	0
[250, 400[GeV 1 jets	0.27	0.14	0.15	-0.02	0.30	-0.43
[250, 400[GeV $\geq$ 2 jets	0.27	0.14	0.15	-0.02	0.30	0.30
[400, ∞[GeV 0 jets	0.27	0.14	0.15	0.17	-4.1	0
[400, ∞[GeV 1 jets	0.27	0.14	0.15	0.17	0.33	-0.82
[400, ∞[GeV $\geq$ 2 jets	0.27	0.14	0.15	0.17	0.33	0.33

Table 5.4: Migration uncertainty values for ggZH process in fine STXS scheme (stage 1.2 including dashed boundaries).

The uncertainties shown in Fig. 5.7 are evaluated using the events generated in `POWHEG` `MiNLO` and then showered and hadronized with `PYTHIA8`. Recently, there have been many developments for VH production and several calculations of the NNLO order became available, one of them the GENEVA NNLL'+NNLO [79]. The uncertainties obtained from the μ_R and μ_F variations in `POWHEG MiNLO` can be compared with the corresponding uncertainties produced with GENEVA NNLO simulation with the usage of appropriate scale variations.

The GENEVA simulation is NNLO accurate with the additional improvement from next-to-next-to-leading logarithmic resummation of the 0-jettines resolution variable. The kinematic distribution of events produced with POWHEG and GENEVA are compared in Figure 5.9. The $\mathrm{p_T^V}$ and p_T^H distributions show a good agreement, while the n_jet distributions features significant difference. It can be attributed to the discrepancies in the leading jet p_T distributions, especially in the low p_T region, which is expected due to the different resummation sensitivity at low jet-p_T.

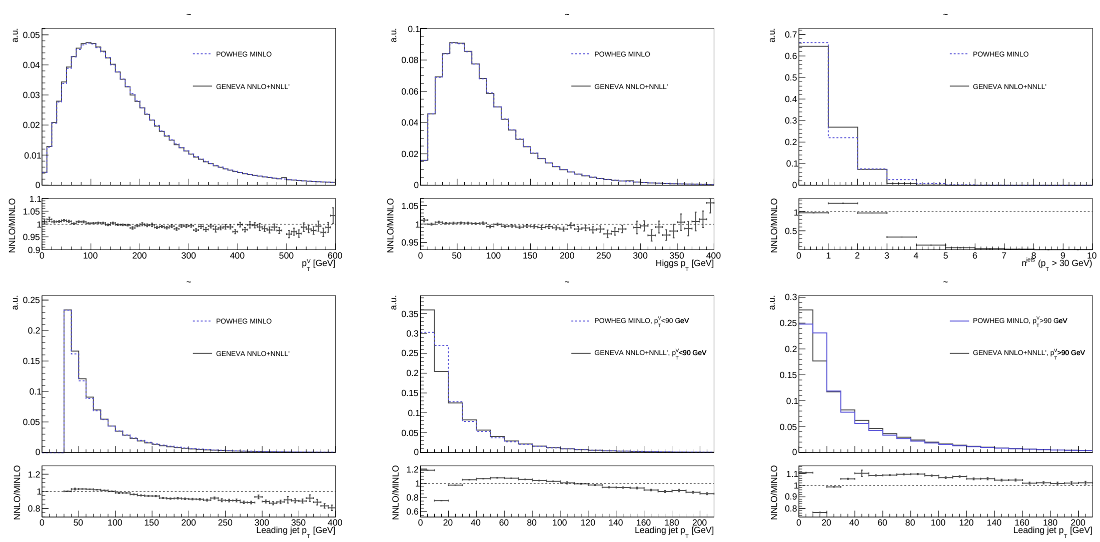


Figure 5.9: Comparison of kinematic distributions of events produced with GENEVA and POWHEG generators. Top row: vector boson p_T distribution on the left, the Higgs boson p_T distribution in the middle and n_jet distribution on the right; bottom row: p_T distribution of the leading jet on the left; p_T distribution of the leading jet for events with $\mathrm{p_T^V} < 90$ GeV in the middle, p_T distribution of the leading jet for events with $\mathrm{p_T^V} > 90$ GeV on the right.

To estimate the corresponding uncertainties, the variations of fixed order and resummation scales are considered in each STXS bin, as shown in Fig. 5.10. Here the leading differences with respect to the nominal cross-section come from FO scale variations μ_FO and μ_overall, which is defined as the FO scale variation corrected to keep the inclusive cross-section unchanged. The other 6 variations correspond to the resummation scales.

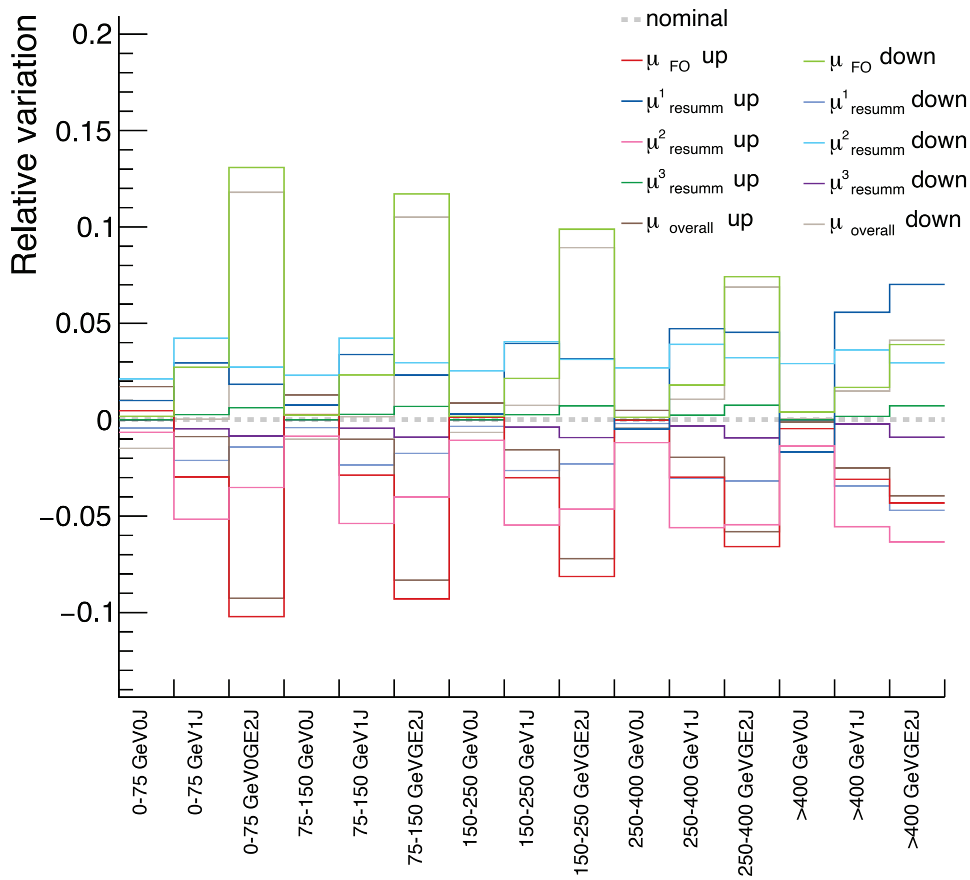


Figure 5.10: Scale variations in `GENEVA NNLL'+NNLO` simultaion

The total perturbative uncertainty is calculated as the quadrature sum of the resummation and FO uncertainties, the results are shown in Fig. 5.14.

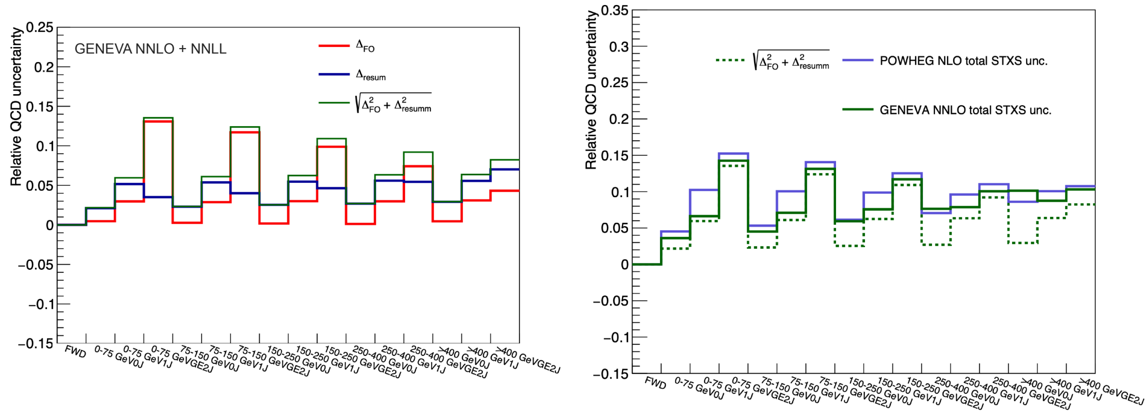


Figure 5.11: Relative FO and resummation uncertainties (left) from GENEVA samples and total uncertainties in comparison with the POWHEG QCD scale uncertainties (right) in STXS bins for VH process.

The resummation and FO scale variations were used to estimate the STXS migration uncertainties using the GENEVA prediction following the same strategy as for the results produced with POWHEG MiNLO. In Fig. 5.14 the total STXS uncertainties, defined as the quadrature sum of n_{jet} and $\text{p}_{\text{T}}^{\text{V}}$ migration uncertainties, and calculated using GENEVA and POWHEG events, are compared with the total perturbative uncertainty from GENEVA simulation.

In each STXS bin the migration uncertainties calculated with POWHEG events are of the same order as the STXS migration uncertainties estimated with GENEVA sample, and both are of the same order as the full perturbative uncertainties derived using the FO and resummation scale variations available in GENEVA simulation. This provides assurance that the uncertainties from renormalisation and factorisation scale variations used in this analysis provide a good estimate for the theoretical uncertainties of the signal.

5.3.2. Acceptance uncertainties

In the previous inclusive VH(H $\rightarrow b\bar{b}$) measurement the acceptance uncertainties were estimated by simultaneously varying μ_R and μ_F scales by a factor of two, which were then propagated to the final observables as shape variations. This approach provided conservative

STXS bin	Δ_{total} (POWHEG)	Δ_{total} (GENEVA)	$\Delta_{\text{pert.}}$ (GENEVA)
[0, 75[GeV 0 jets	0.045	0.036	0.021
[0, 75[GeV 1 jets	0.102	0.066	0.059
[0, 75[GeV $\geq$ 2 jets	0.152	0.143	0.136
[75, 150[GeV 0 jets	0.053	0.045	0.023
[75, 150[GeV 1 jets	0.1	0.07	0.06
[75, 150[GeV $\geq$ 2 jets	0.14	0.129	0.122
[150, 250[GeV 0 jets	0.061	0.059	0.026
[150, 250[GeV 1 jets	0.099	0.076	0.062
[150, 250[GeV $\geq$ 2 jets	0.125	0.012	0.011
[250, 400[GeV 0 jets	0.071	0.078	0.043
[250, 400[GeV 1 jets	0.096	0.079	0.064
[250, 400[GeV $\geq$ 2 jets	0.11	0.099	0.092
[400, ∞[GeV 0 jets	0.086	0.112	0.059
[400, ∞[GeV 1 jets	0.1	0.087	0.063
[400, ∞[GeV $\geq$ 2 jets	0.11	0.1	0.08

Table 5.5: Migration uncertainties for qqZH process calculated with POWHEG, GENEVA, the total perturbative uncertainties in GENEVA simulation.

estimate of acceptance effects for the inclusive measurement. In the present analysis this strategy was adapted to the STXS measurement.

The acceptance uncertainties are evaluated to account for the difference in acceptance in the measured STXS bins with respect to the fine STXS bins where all dashed boundaries are included and acceptance effects should be negligible. In other words, the *dashed* boundaries between the merged STXS bins (Figure 5.2) are considered to be the source of the acceptance uncertainty. If a measured bin is not merged the simultaneous μ_R and μ_F scale variations are used to estimate the acceptance uncertainties. By construction, the acceptance uncertainties do not affect the normalisation of a given process and only alter the shape within a measured STXS bin. The migration uncertainties Δ_X[3] given in the section above are used as input in the derivation of the acceptance uncertainties.

[3] X is a dashed boundary inside a measured STXS bin

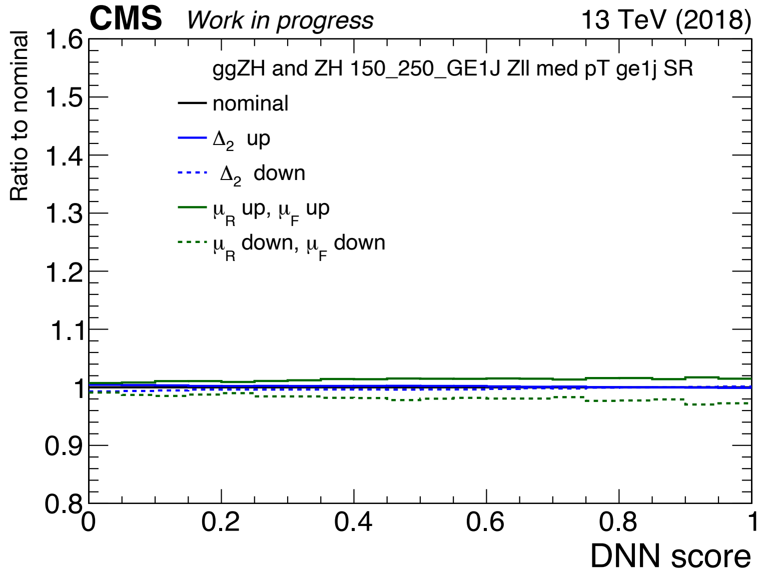


Figure 5.12: Acceptance uncertainties for ZH $150 < p_T^V < 250$ GeV ≥ 1 jet process.

As example, we can consider the ZH STXS bins within the $150 < p_T^V < 250$ GeV region. In the fine STXS binning used to calculate migration uncertainties this region is split according to the number of jets into the bins with 0 jets, 1 jets and at least 2 jets. In the STXS 1.2 scheme used in this measurement only the bins with 0 jets and at least 1 jet are used. Therefore, within ZH $150 < p_T^V < 250$ GeV ≥ 1 jet STXS bin, according to the discussion given above, we can consider the Δ_2 as a source of acceptance uncertainty. In Figure 5.12 the variations imposed by the Δ_2 migrations for the ZH $150 < p_T^V < 250$ GeV ≥ 1 in the corresponding signal region are shown.

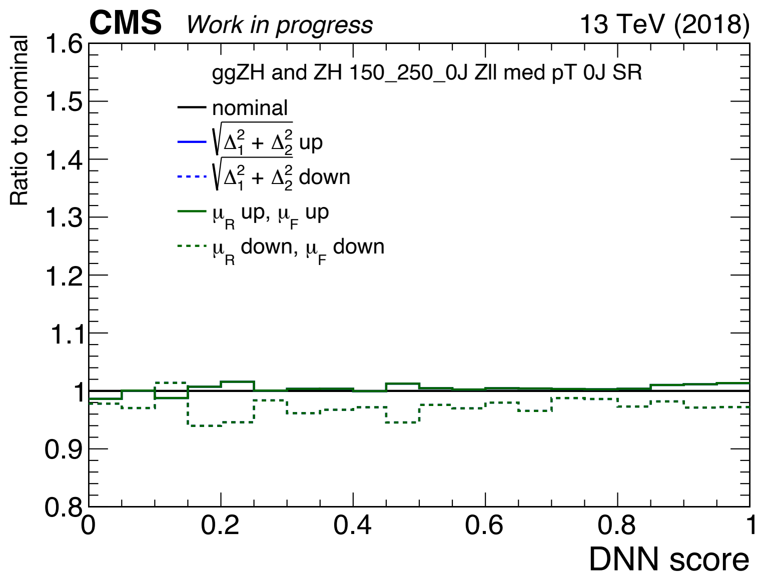


Figure 5.13: Acceptance uncertainties for ZH $150 < p_T^V < 250$ GeV 0 jets process.

The STXS bin ZH $150 < p_T^V < 250$ GeV 0 jets does not have a dashed boundary within, so to estimate the acceptance uncertainty for this STXS process the diagonal scale variations are applied.

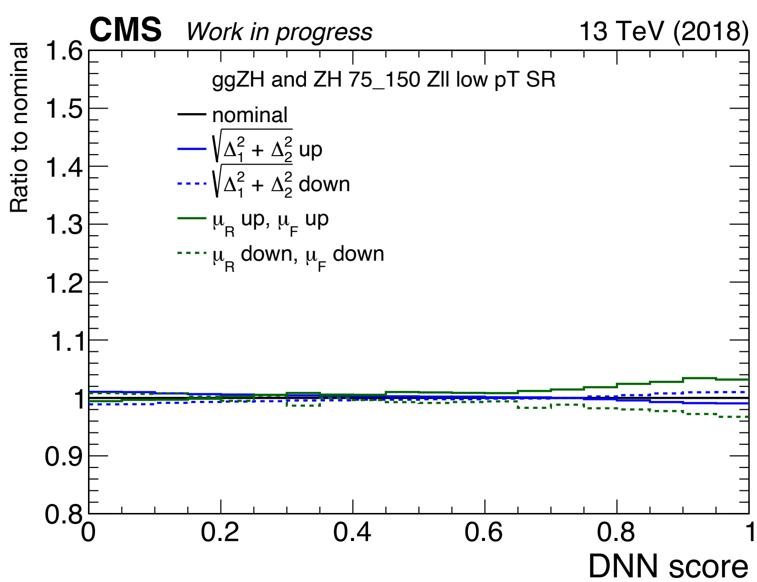


Figure 5.14: Acceptance uncertainties for ZH $75 < p_T^V < 150$

As another example the calculated acceptance uncertainties are shown for ZH 75 $< \mathrm{p_T^V} <$ 150 GeV STXS process within the corresponding signal region. In this STXS bin the contributions from Δ_1 and Δ_2 are considered.

For all of the given examples the shape variations from migration uncertainties are at the order of 1%, while the uncertainties coming from diagonal scale variations are 2-3%. This difference is expected since for the migration uncertainties only the maximal scale variation is considered, while the diagonal scale variations imply that the μ_F and μ_R scales are varied simultaneously, which can provide overly conservative results.

CHAPTER 6

Statistical model

The VH(H $\rightarrow b\bar{b}$) signal is extracted by performing a simultaneous maximum-likelihood fit including all signal and control regions. The control and signal regions were split in the reconstructed p_T^V regions to align with the particle-level STXS splitting (see Section 5.1). In this section the details on the statistical inference procedure are given. In the Section 6.1 all analysis categories are summarised. Then the likelihood function and the systematic uncertainties are described.

6.1. Categories counting

With the STXS-driven categorisation of the signal regions introduced in Chapter 5 the following 5 signal regions for the 2-lepton channel are defined: $75 < p_T^V < 150$ GeV, $150 < p_T^V < 250$ GeV $N_{aj} = 0$, $150 < p_T^V < 250$ GeV $N_{aj} \geq 1$, $250 < p_T^V < 400$ GeV, $p_T^V > 400$.

Since the electron and muon categories are reconstructed separately, we end up with 10 categories coming from the resolved 2-lepton channel signal regions. To match the signal region kinematics and have a better control on p_T^V shape of the background processes, the control regions are also split in p_T^V regions: 75-150 GeV, 150-250 GeV, and $\geq$ 250 GeV. In total, the 2-lepton control regions sum up into 18 categories, if added with the signal regions we get **28 categories from the 2-lepton channel**.

In 1-lepton channel the categories counting is similar, but the p_T^V regions start from 150 GeV. The signal regions form 6 categories, and the control regions form 12 categories, and in total **18 categories form the 1-lepton channel**.

The 0-lepton channel $\mathrm{p_T^V}$ starts from 170 GeV. The signal regions are categorised as follows: $\mathrm{p_T^V} < 250$ GeV $N_{\mathrm{aj}} = 0$, $\mathrm{p_T^V} < 250$ GeV $N_{\mathrm{aj}} \geq 1$, $250 < \mathrm{p_T^V} < 400$ GeV, $\mathrm{p_T^V} > 400$ GeV. Following the same counting as for the other channels, we have 4 categories from signal regions and 6 categories from control regions, summing up in total to **10 categories from the 0-lepton channel**.

In the boosted analysis all signal regions are separated in 2 vector boson transverse momenta regions: $250 < \mathrm{p_T^V} < 400$ GeV and $\mathrm{p_T^V} > 400$ GeV. Therefore 10 categories[1] from all boosted signal regions are defined. For each vector boson decay channel ($Z \to \nu\nu$, $W \to e\nu$, $W \to \mu\nu$, $Z \to \mu\mu$, $Z \to ee$) 3 control regions (V+HF, V+LF, $\mathrm{t\bar{t}}$) are defined, summing up to 15 categories for boosted control regions. In total the boosted analysis uses **25 categories**.

Therefore, for each data-taking year 81 categories is defined, which for the three years 2016, 2017 and 2018 results in 243 categories.

6.2. Likelihood construction

The details of statistics inference procedures commonly used in the ATLAS and CMS collaborations are given in [80]. In this section a short summary will be given.

For each category, k, the likelihood function L_k can be defined as follows:

$$L_k(\mathrm{obs}, \mu, \vec{\theta}) = \prod_{\mathrm{b=1}}^{\mathrm{nb(k)}} \mathrm{P}\left(\mathrm{obs}, \sum_{\mathrm{i}} (\mathrm{S^i_{kb}}(\mu^{\mathrm{i}}, \vec{\theta}_{\mathrm{sig}})) + \sum_{\mathrm{j}}^{\mathrm{CR}} \beta^{\mathrm{j}} \mathrm{B^j_{kb}}(\vec{\theta}_{\mathrm{bkg}}) + \sum_{\mathrm{j}}^{\mathrm{pred}} \mathrm{B^j_{kb}}(\vec{\theta}_{\mathrm{bkg}})\right), \quad (6.1)$$

where $P(\mathrm{obs}, \lambda^{\mathrm{i}}(\mu^{\mathrm{i}}, \vec{\theta}))$ is the Poisson function, defined as $P(n, \lambda) = \frac{\lambda^n}{n!} e^{-\lambda}$. The index b runs over the nb(k) bins in each category k. The S^i_{kb} is the expected number of signal events for STXS process i. B_{kb} is the expected background yield in bin b of category k.

The $\vec{\theta} = \{\vec{\theta}_{bkg}, \vec{\theta}_{sig}\}$ term denotes the nuisance parameters modelling the systematic uncertainties. The $\vec{\theta}_{sig}$ term includes the following components: $\vec{\theta}_{sig} = \{\vec{\theta}^{th}_{sig}, \vec{\theta}^{\epsilon}_{sig}, \vec{\theta}^{exp}_{sig}\}$. The $\vec{\theta}^{th}_{sig}$ is responsible for purely theoretical uncertainties, the $\vec{\theta}^{\epsilon}_{sig}$ corresponds to the acceptance

[1] 2*dim{$Z \to \nu\nu$, $W \to e\nu$, $W \to \mu\nu$, $Z \to \mu\mu$, $Z \to ee$})

uncertainties for signal processes and $\vec{\theta}^{exp}_{sig}$ term denotes the experimental uncertainties. The $\vec{\theta}^{\epsilon}_{sig}$ affects only the shape, and $\vec{\theta}^{exp}_{sig}$ can be either shape or normalisation uncertainty. The $\vec{\theta}^{th}_{sig}$ term is defined for each STXS process and affects the overall yield of each STXS process i.

The $\vec{\theta}_{bkg} = \{\vec{\theta}^{th}_{bkg}, \vec{\theta}^{exp}_{bkg}\}$ term is responsible for the systematic uncertainties modelling background processes, generated by imperfections of theoretical predictions or experimental effects and can be either shape or normalisation uncertainties. The systematic uncertainties are modelled with a Gaussian distribution $G(\theta) = \frac{1}{\sqrt{2\pi}} e^{\frac{(\tilde{\theta}-\theta)^2}{2}}$. The sources of systematic uncertainties considered in this analysis are listed in Section 6.4.

In the context of STXS measurement and the SMEFT interpretation studies reported in this work (see Chapter 8), it is useful to look at the expected signal yield in detail. For the STXS measurements it is parameterised by the signal strength modifier μ^i for each STXS process i:

$$S^i(\mu^i, \vec{\theta}_{sig}) = \mu^i \times [\sigma \times B]^i_{SM}(\vec{\theta}^{th}_{sig}) \epsilon^i_k(\vec{\theta}^{\epsilon}_{sig}, \vec{\theta}^{exp}_{sig}) L(\theta_{lumi}), \tag{6.2}$$

where the dependence on purely theoretical uncertainties is only present in the SM prediction $[\sigma \times B]^i_{SM}(\vec{\theta}^{th}_{sig})$. If instead of a signal strength measurement the cross-section measurement is performed, i.e. if the signal strength is expressed as:

$$\mu^i = \frac{[\sigma \times B]^i_{obs}}{[\sigma \times B]^i_{SM}(\vec{\theta}^{th}_{sig})}, \tag{6.3}$$

the explicit dependence on theory uncertainties is removed from the measurement, which is the main feature of the STXS measurements. The only theory-dependent contribution left is the residual $\vec{\theta}^{\epsilon}_{sig}$ acceptance uncertainty. The acceptance uncertainties can be minimised by introducing finer STXS splitting.

There are two types of background processes in this analysis: those that are constrained from the control regions with the rate parameters; and the processes for which the predicted yields are used. This is denoted in the background term in the likelihood function

$$\sum_{j}^{\mathrm{CR}} \beta^j B^j_{kb}(\vec{\theta}_{bkg}) + \sum_{j}^{\mathrm{pred}} B^j_{kb}(\vec{\theta}_{bkg}), \tag{6.4}$$

where the first terms corresponds to the background processes constrained from control regions with the rate parameters β^j, these processes are the V+c, V+udsg, V+b jets and $\mathrm{t\bar{t}}$. The second term corresponds to the single-top and VZ backgrounds, their yields are only modified by the theoretical uncertainties on their cross section $\vec{\theta}^{th}_{bkg}$ and the experimental uncertainties $\vec{\theta}^{exp}_{bkg}$.

In addition to the likelihood term shown in Eq. 6.1, the term modelling the MC statistical uncertainties in each bin of category k is added as a Poisson term. The Barlow-Beeston lite method described in [81] allows to reduce the number of parameters and consider only one nuisance in each bin of category k, instead of introducing a nuisance parameter for each process within a bin. This simplification is only valid if the number of MC events in each bin is sufficient, which is ensured with the choice of appropriate binning in each signal category.

The total likelihood function is constructed by multiplying the terms L_k from each category k including the Poisson terms to model the MC statistical uncertainties. The signal is then extracted by minimising the $-2\ln L(\mathrm{obs}, \alpha, \vec{\theta})$, where α is the generalised parameter of interest (POI), which is convenient to introduce at this point, since within this thesis the signal is extracted in various configuration of signal parametrisation. To extract the confidence intervals the test statistics $q(\alpha)$ is used:

$$q(\alpha) = -2\ln \frac{L(\mathrm{obs}, \alpha, \vec{\theta}^{\mu})}{L(\mathrm{obs}, \hat{\alpha}, \hat{\vec{\theta}})}, \tag{6.5}$$

where $\hat{\alpha}$ and $\hat{\vec{\theta}}$ correspond to the global minimum of the $-2\ln L(\mathrm{obs}, \alpha, \vec{\theta})$, and the $\vec{\theta}^{\alpha}$ corresponds to the conditional estimates of $\vec{\theta}$ for a value α of POI.

6.3. Background modelling

In this section the modelling of V+udsg, V+c, V+b, $t\bar{t}$ processes will be described. In this analysis the background yields and p_T^V shape are controlled in the fit.

First, as it is shown in Eq. 6.4 the predicted background yields are modified with the rate parameters common for the signal and control regions, therefore the background normalisation in signal region is corrected by the yields that are observed in control regions enriched in specific background processes. This analysis uses an extremely conservative rate parameters (RP) scheme, where all of the four CR-constrained processes, (V+udsg, V+c, V+b, $t\bar{t}$) are measured separately in the 0-lepton, 1-lepton and 2-lepton channels, and uncorrelated in lepton flavours. So for each background process 5 RP are defined inclusively in all p_T^V ranges. The post-fit values of background rate parameters are reported in Table 6.1.

To constrain the p_T^V shape of the CR-constrained background processes a linear shape variation is added as additional nuisance parameters in the fit. These shape variations act as migration uncertainties that allow linear variations in the shape of p_T^V. They are considered for each p_T^V boundary defined for the CR 6.1. This means that for 2-lepton channel 2 p_T^V background migration uncertainties are defined (150 GeV, 250 GeV). In 0-lepton and 1-lepton channels, only one p_T^V migration nuisance is considered.

Lastly, an additional set of rate parameters is added for the boosted categories to model the DeepAK8 tagger efficiency measurements that are not available for jet flavours. These parameters were correlated in some of the lepton channels due to the low statistics of the corresponding background processes.

Background process	$Z \to \nu\nu$	$W \to \mu\nu$	$W \to e\nu$	$Z \to e^+e^-$	$Z \to \mu^+\mu^-$
$t\bar{t}$	0.96 ± 0.07	0.89 ± 0.07	0.86 ± 0.07	0.82 ± 0.10	1.07 ± 0.20
V + udsg	1.18 ± 0.11	0.90 ± 0.08	0.91 ± 0.08	1.01 ± 0.08	0.99 ± 0.07
V + c	0.79 ± 0.20	0.90 ± 0.18	0.97 ± 0.16	0.67 ± 0.15	0.74 ± 0.14
V + b	1.04 ± 0.06	1.19 ± 0.14	0.99 ± 0.13	1.10 ± 0.10	1.02 ± 0.09
Rate parameters modelling the DeepAK8 efficiency					
V + b	0.93 ± 0.08	0.88 ± 0.11	0.91 ± 0.12	0.91 ± 0.12	0.88 ± 0.11
$t\bar{t}$	0.92 ± 0.02	0.92 ± 0.02	0.92 ± 0.02	0.92 ± 0.02	0.92 ± 0.02
V + udsg	0.95 ± 0.05	0.96 ± 0.04	1.01 ± 0.04	1.01 ± 0.04	0.96 ± 0.04
V + c in b-tagged	2.26 ± 0.52	2.26 ± 0.52	2.26 ± 0.52	2.26 ± 0.52	2.26 ± 0.52
V + c fail b-tag	1.27 ± 0.26	1.27 ± 0.26	1.27 ± 0.26	1.27 ± 0.26	1.27 ± 0.26

Table 6.1: Post-fit values of background rate parameters for the 2018 analysis

6.4. Systematic uncertainties

In this section the considered systematic uncertainties are described. In the full Run 2 combination, when all data-taking year categories are fitted simultaneously, the theoretical uncertainties are fully correlated, while the experimental uncertainties are uncorrelated.

6.4.1. Theory uncertainties.

The uncertainties entering the $\vec{\theta}^{th}_{sig}$ and $\vec{\theta}^{th}_{bkg}$ term are described in this section.

- **Signal theory:** The uncertainty on the total cross-section as well as the STXS migrations uncertainties are applied for all signal processes. The total migration uncertainties for qqZH and WH range from 5% to 13%, for ggZH they are substantially larger and exceed 100% in some of the bins, in these cases they are limited at 90%. These uncertainties are derived from factorisation and renormalisation scale variations, the details are given in Section 5.3.

- **Parton density function and α_s:** These uncertainties are derived from NNLOPDF3.0 set following the recommendations [69], and applied as uniform uncertainties for all STXS processes. The uncertainties amount to 1.6% for ZH processes and 1.9% for WH processes. The uncertainties for background processes vary from 0.5% for $t\bar{t}$ process to 5% for V+c jets process.

- p_T^V **spectrum of signal processes:** The uncertainties from the NLO electroweak and NNLO QCD corrections for signal processes are 2% and 5%, respectively.

- **Background cross-section:** For the backgrounds not measured in control regions, the single-top and diboson, a 15% uncertainty is assigned. This corresponds to the uncertainty on the measured cross-sections.

- $\mathbf{H \rightarrow b\bar{b}}$ **branching ratio:** The uncertainty in the $H \rightarrow b\bar{b}$ branching ratio is 0.5% [15].

- **Signal acceptance:** In general the STXS measurement provides an opportunity to reduce the acceptance uncertainties with respect to the inclusive measurements. Nevertheless, the acceptance effects are still present due to the STXS bin merging. The corresponding uncertainties are derived and included as shape variations within the STXS bin, explicitly removing any normalisation effects. The details can be found in Section 5.3.

- **Background QCD scale uncertainties:** For each background process the uncertainties derived from factorisation and renormalisation scales are applied as shape variations.

6.4.2. Experimental uncertainties

- **Luminosity:** The uncertainty on the integrated luminosity measurement is 2.5% in 2016 and 2018, and 2.3% in 2017 [82, 83, 84]. These uncertainties are partially correlated between the different data-taking years following the recommendations of the CMS group responsible for the luminosity measurements.

- **Lepton efficiencies:** Uncertainties in the electron and muon ID, isolation, and trigger scale factors amount to 2% [85], [43].

- **MET trigger efficiencies:** Uncertainties in the MET trigger efficiency measurement amount to 1% [86].

- **Jet energy resolution:** For all b-tagged jets in this analysis the DNN-based energy regression is applied to recover the energy loss due to the leptonic decays of b-hadrons. The regression algorithm evaluated on MC and data events results in different resolution. Therefore the jet energy resolution corrections are extracted for each data-taking year and summarised in Table 6.2.

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 6.2: The extracted scaling and smearing needed for each year of data as a percent of the jet's p_T.

- **Jet energy scale:** The jet energy scale uncertainties generated by various sources are dependent on η and p_{T} of a jet. These uncertainties shift the jet energies and therefore the kinematics of events.

- **B-tagging:** The b-tagging efficiency measurements are applied to the simulation to account for the difference in performance observed in data and MC. The uncertainties on b-tagging efficiency measurements are derived in different bins of p_{T} and η. These uncertainties are derived and applied for all jet flavours.

- **DeepAK8 double-b-tagging:** The uncertainties for the DeepAK8 double b-tagger efficiency are derived only for the 2 b-jets topologies. The uncertainties on DeepAK8 are uncorrelated in working points and in boosted jet momentum bins (200-300 GeV, 300-400 GeV, 400-500 GeV, 500-600 GeV, >600 GeV). For other jet flavours freely floating rate parameters were added to account for efficiency effects (see Table 6.1).

- **Pileup corrections:** The simulated samples are reweighted to account for differences in pileup profiles observed in MC and data. The corresponding uncertainty is applied as a shape variation.

- **Uncertainties on V+jets reweighting:** The 2016 LO V+jets samples were reweighted to NLO accuracy in $\Delta\eta(bb)$ variable of the Higgs jet candidate. The corresponding uncertainties are included as shape variations. The same method was used in the previous publication VH(H $\to b\bar{b}$) measurement and detailed in the dissertation [75].

  The 2017 and 2018 V+jets samples are already produced at NLO accuracy, the additional reweighting of $\Delta R(bb)$ variable was found to be necessary to improve the agreement of data and simulation in the range $\Delta R(bb) < 1$. The systematic uncertainties and the methods are detailed in the dissertation [87].

6.4.3. Shape uncertainties template smoothing

The jet energy scale (JES) and jet energy resolution (JER) uncertainty templates were additionally tuned after observing unrealistic constraints caused by large fluctuations of up and down variation templates. The up/nominal and down/nominal histogram ratios were smoothed using the methods implemented in ROOT software and described in [88]. Among the tested smoothing methods are `TH1F::Smooth(n)`, where n is the number of smoothing iterations, where $n = \{2, 3, 4, 10\}$ are tested; `TH1F::SmoothLowess()`. The results are shown in Figure 6.1. It was ensured that the smoothed templates repeat the trends of the initial variations by comparing the χ^2 values. Since all of the compared methods yield similar results, the simplest one was used (`TH1F::Smooth(2)`).

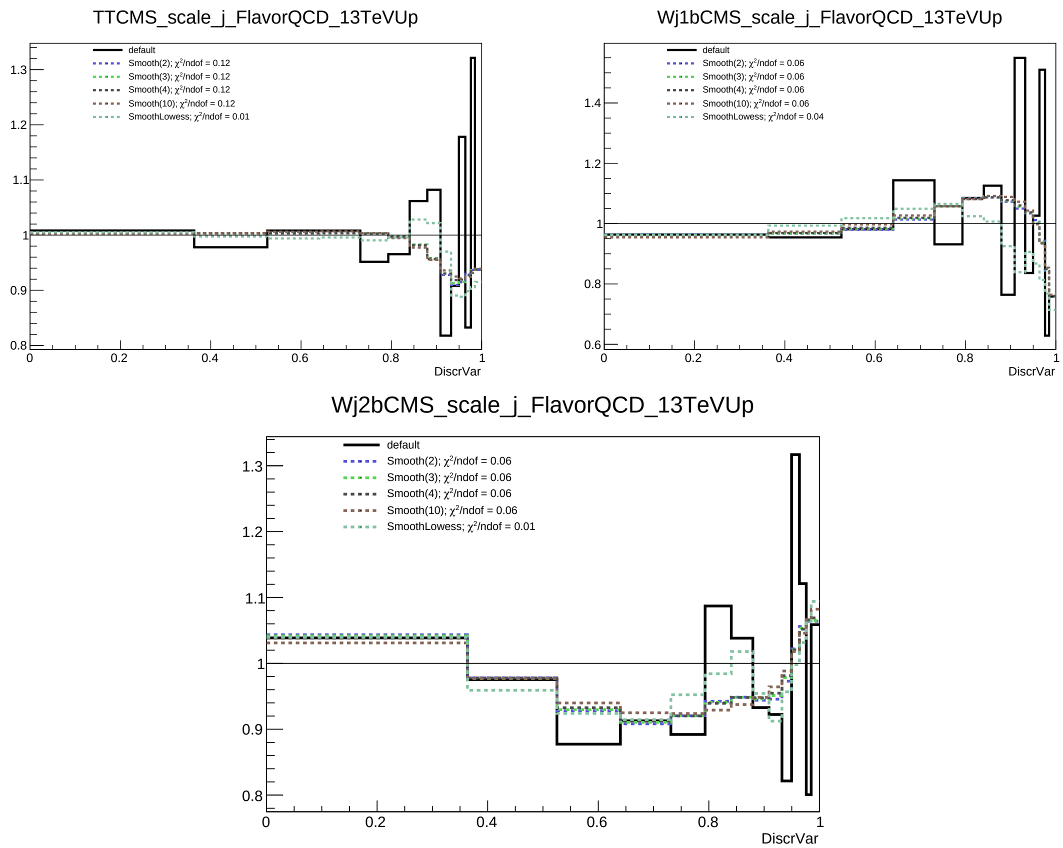


Figure 6.1: Smoothing of up(down) variations for JES systematic uncertainty templates.

CHAPTER 7

STXS VH(H $\to b\bar{b}$) MEASUREMENT RESULTS

7.1. Introduction

This chapter summarizes the measurement of the VH(H $\to b\bar{b}$) process performed with the full Run 2 CMS data. The results are presented in different configurations of signal parameters of interest (POI). The signal strength modifier (μ) measurements are discussed in Section 7.2, where the inclusive and per-production mode results are presented. Next, the STXS cross-section measurement results, in the configuration discussed in Chapter 5, are detailed in Section 7.3.

7.2. Signal strength results

The inclusive signal strength measurement is performed by assigning a single parameter of interest for all signal processes. The maximum-likelihood fit of 243 categories (see Section 6.1) in total was performed, including all of the analysis control and signal regions discussed. In the signal regions the MVA scores are used as fit templates. The control regions observables are defined in Section 4.8. The expected and observed likelihood curves are shown in Figure 7.1 from where the best fit value of μ and the confidence intervals are extracted to be

$$\mu = 0.57^{+0.14}_{-0.13}\,(\text{stat.})\,^{+0.13}_{-0.12}\,(\text{syst.}), \tag{7.1}$$

corresponding to observed (expected) significance of 3.3 (5.2) standard deviation with respect to the background only hypothesis.

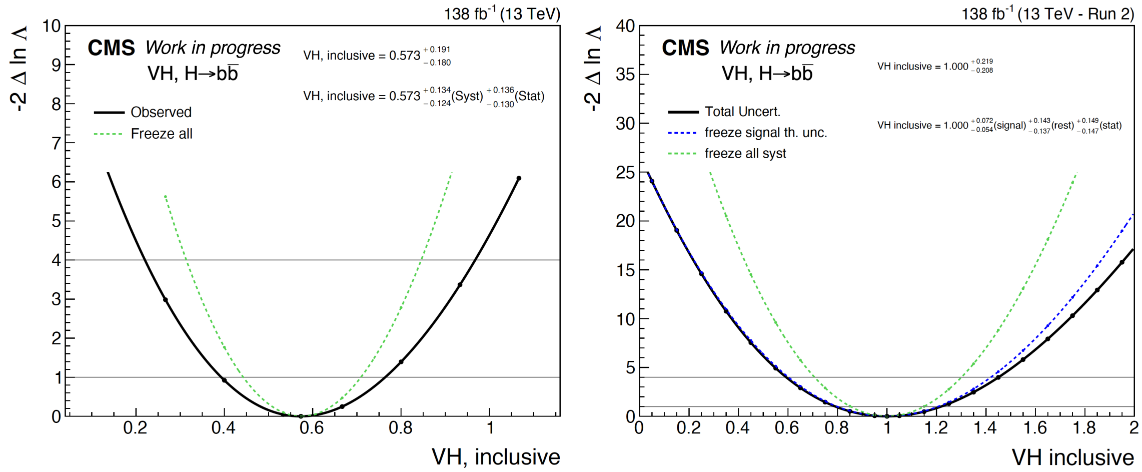


Figure 7.1: Test statistics distributions for inclusive single strength μ extracted from the full Run 2 fit. The curve where all of the uncertainties are included are shown in black, without theoretical signal uncertainty in violet, and the curve where only statistical uncertainties are included in green.

Due to the accumulated data in full Run 2 dataset, the current inclusive measurement is not statistically limited. Therefore it is important to discuss the largest contributions to the systematic uncertainty.

The most impactful nuisances in the full Run 2 inclusive fit are shown in Figure 7.2. The impacts are calculated by shifting each nuisance parameter by 1σ and checking the impact on μ. The nuisance parameters impacts demonstrate the level of correlation with the POI. The leading nuisance parameters shown in Figure 7.2 are the signal and background theoretical uncertainties as well as the nuisance parameters modelling the MC statistical uncertainties. To illustrate the contributions better, all of the nuisances were categorised, and the contribution of each group to the total uncertainty is estimated and reported in Table 7.1. The MC statistical uncertainty is the single leading source, followed by relatively large signal and background theory uncertainties. The next important source is the uncertainty attributed to background modelling.

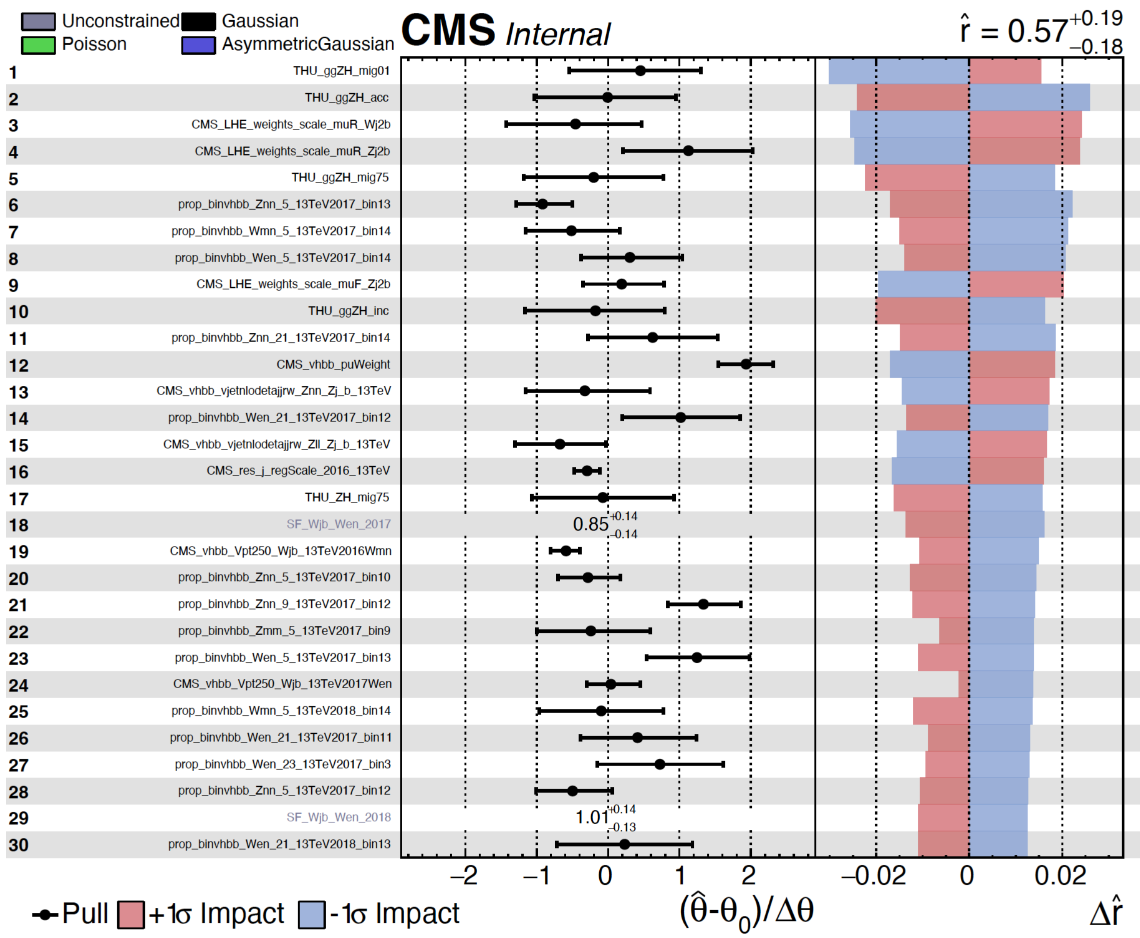


Figure 7.2: Pulls and constraints for the most impactful nuisances in the full Run 2 inclusive fit.

source	$\Delta\mu$
Background (theory)	+0.067 -0.064
Signal (theory)	+0.082 -0.060
MC stats.	+0.092 -0.093
Sim. modelling	+0.070 -0.066
b-tagging	+0.059 -0.041
JER	+0.045 -0.057
Luminosity	+0.041 -0.034
Jet energy scale	+0.029 -0.036
LeptonID	+0.016 -0.002
Trigger(MET)	+0.001 -0.001

Table 7.1: Impacts of different nuisance groups on inclusive single strength.

The next set of results summarised in Figure 7.3 was produced by introducing a separate POI for each of the 0-lepton, 1-lepton and 2-lepton channels. The compatibility of the per-channel fit results with the inclusive fit results is estimated to be 1.9 σ.

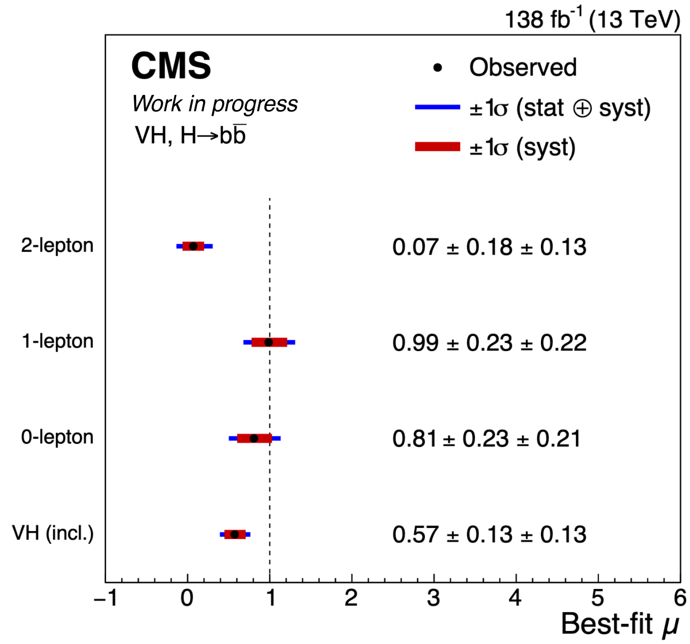


Figure 7.3: Observed signal strengths for the 0-lepton, 1-lepton and 2-lepton channels, as well as the combined signal strength.

The signal strength modifiers extracted for the ZH and WH production processes are shown in Figure 7.4. The per-production mode fit results are compatible with the inclusive fit results at the level of 2 standard deviations.

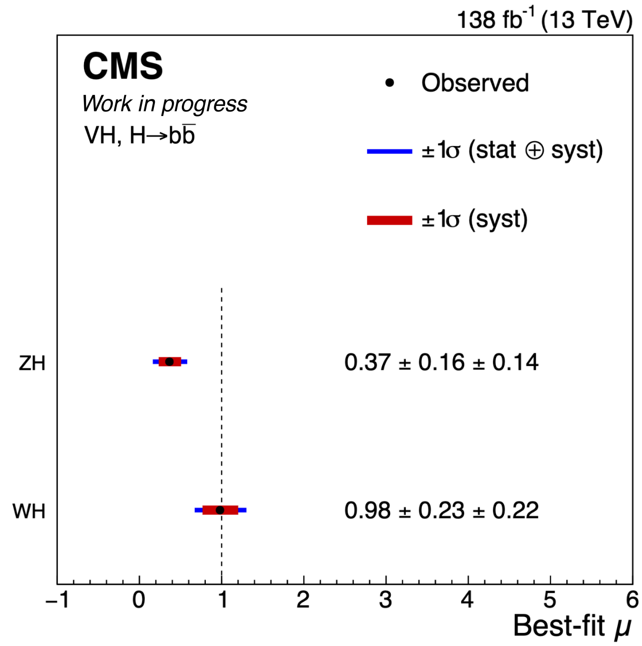


Figure 7.4: Signal strengths for the ZH and WH production modes.

It is evident, that there is some level of deviation from the SM predictions in the fit results presented above. The p-values were calculated, and it was found that all of the observed deviations are 3.9σ in case of the STXS fit, and 3σ for the per-channel and per-process fit.

7.3. STXS results

In this section the STXS measurements are presented. The strategy for the STXS measurement is discussed in Chapter 5 in details. The STXS cross sections are extracted by assigning a parameter of interest for each STXS process. In Figure 7.5 the measured cross sections are shown for eight STXS bins and the values are summarised in Table 7.2. The theoretical uncertainties for the signal (Section 5.3) are decoupled from the measurements and reported in the gray bands. The predicted STXS cross-sections are calculated using the inclusive values reported in [15].

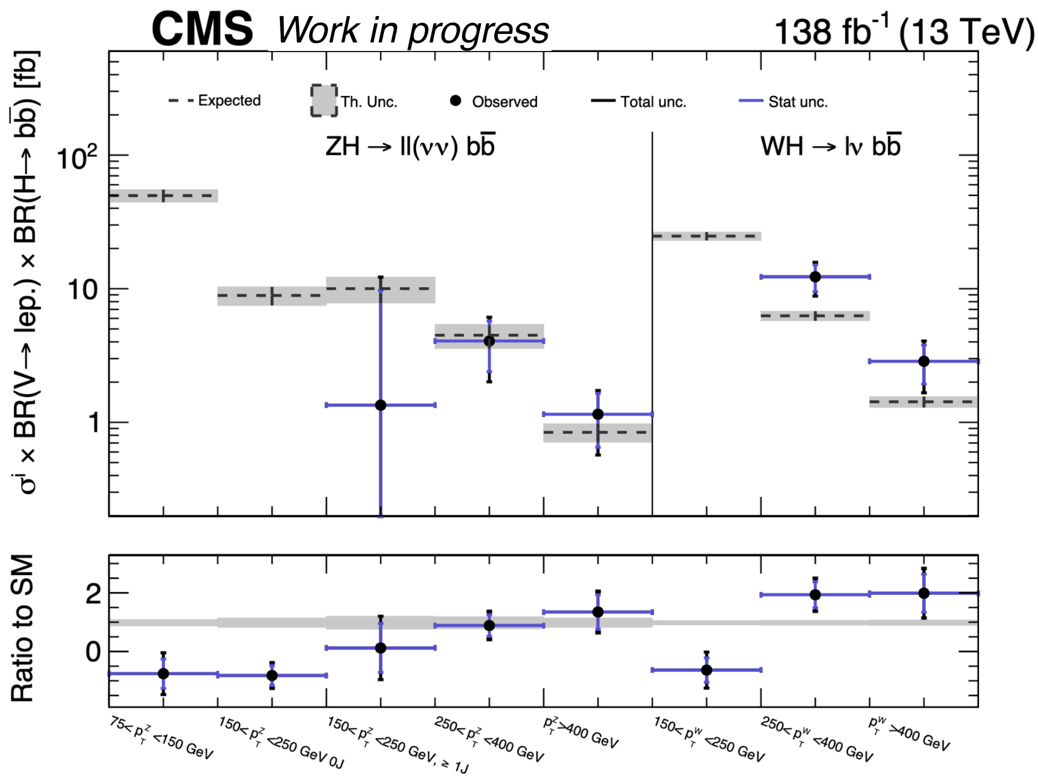


Figure 7.5: Measured values of $\sigma\mathcal{B}$ in STXS bins, combining all years. In the bottom panel, the ratio of the observed results with associated uncertainties to the SM expectations is shown. For the bins where the negative signal strength is measured the observed cross-section values are not reported.

STXS bin	Expected $\sigma\times\ \mathcal{B}$ [fb]	Observed $\sigma\times\ \mathcal{B}$ [fb]	σ/σ^{SM}
ZH $75 < p_T(Z) < 150$ GeV	50.0 ± 5.3	< 0	-0.7 ± 0.7
ZH $150 < p_T(Z) < 250$ GeV 0 jets	9.0 ± 1.4	< 0	-0.8 ± 0.4
ZH $150 < p_T(Z) < 250$ GeV ≥ 1 jets	10.1 ± 2.2	1.4 ± 10.9	0.1 ± 1.1
ZH $250 < p_T(Z) < 400$ GeV	4.5 ± 0.9	4.1 ± 2.1	0.9 ± 0.5
ZH $p_T(Z) > 400$ GeV	0.9 ± 0.1	1.2 ± 0.6	1.4 ± 0.7
WH $150 < p_T(W) < 250$ GeV	24.9 ± 1.8	< 0	-0.6 ± 0.6
WH $250 < p_T(W) < 400$ GeV	6.3 ± 0.5	12.4 ± 3.5	2.0 ± 0.6
WH $p_T(W) > 400$ GeV	1.4 ± 0.1	2.9 ± 1.2	2.2 ± 0.8

Table 7.2: The cross section values for VH process in STXS 1.2 scheme multiplied by the branching fraction of V → leptons and $H \to b\bar{b}$. The SM predictions for each bin are calculated using the inclusive values reported in YR4.

The correlation matrix for the STXS parameters of interest is shown in Figure 7.6. The observed correlations are low for almost all of the STXS parameters, except ZH $150 < p_T^V < 250$ GeV 0 jets and ZH $150 < p_T^V < 250$ GeV ≥ 1 jet reaching 20%, resulting in the low constraints as reported in 7.2.

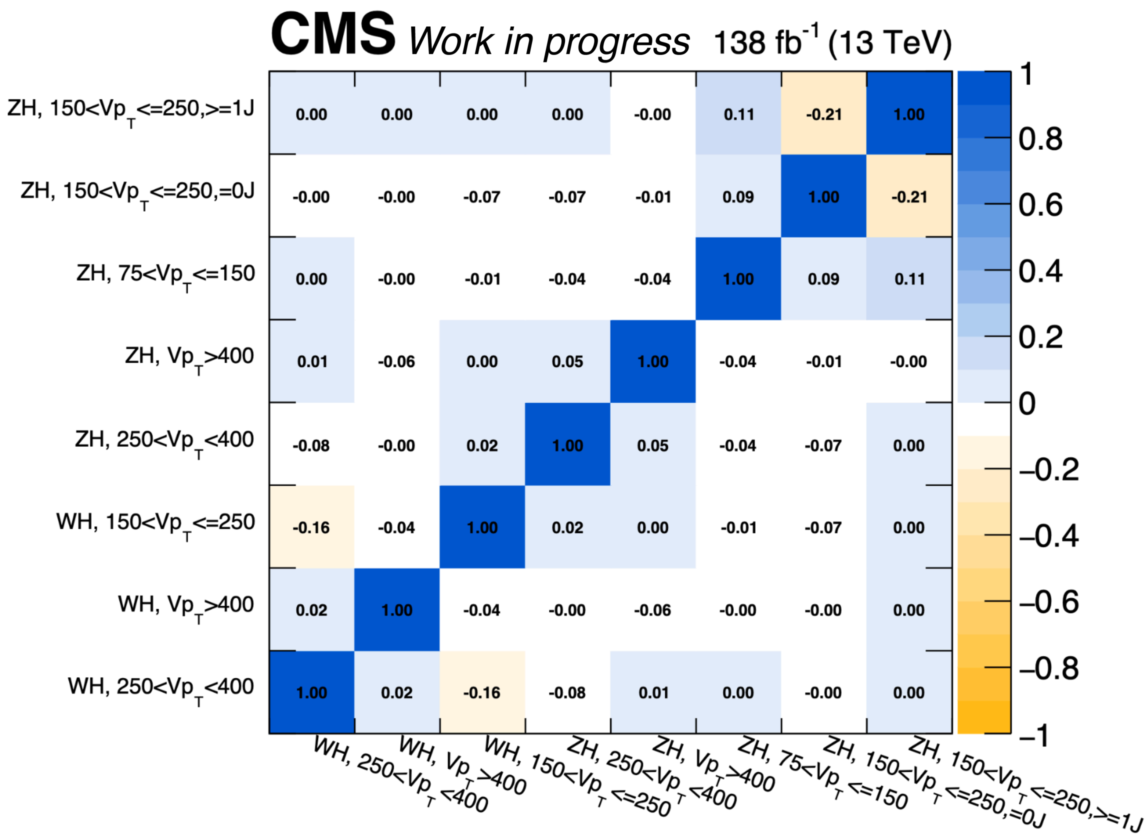


Figure 7.6: Observed correlations between the STXS parameters of interest.

The fitted boosted and resolved signal regions' observables from the 2018 analysis are shown in Figures 7.7, 7.8, 7.9. Where a good post-fit agreement can be observed.

7.4. Summary

This chapter summarises the measurement of VH($H \to b\bar{b}$) process performed using the full Run 2 dataset corresponding to an integrated luminosity of 138 fb^{-1} collected by the CMS collaboration. The VH($H \to b\bar{b}$) analysis presented in this thesis includes the categories where the Higgs boson is reconstructed from two resolved jets, and the categories with the boosted Higgs decay topology. The results of inclusive signal strength and STXS measurements are reported.

In the inclusive VH(H $\rightarrow b\bar{b}$) measurement, the systematic uncertainties are dominated by the background MC statistical uncertainties as well as theoretical uncertainties from both signal and background predictions. The STXS measurement is performed using the stage 1.2 VH scheme, with the inclusion of additional boundary at $\mathrm{p_T^V} = 400$ GeV. Additional granularity in high $\mathrm{p_T^V}$ is particularly valuable for BSM interpretations, as it will be shown in the next chapter focused on SMEFT interpretations of the STXS VH(H $\rightarrow b\bar{b}$) measurement.

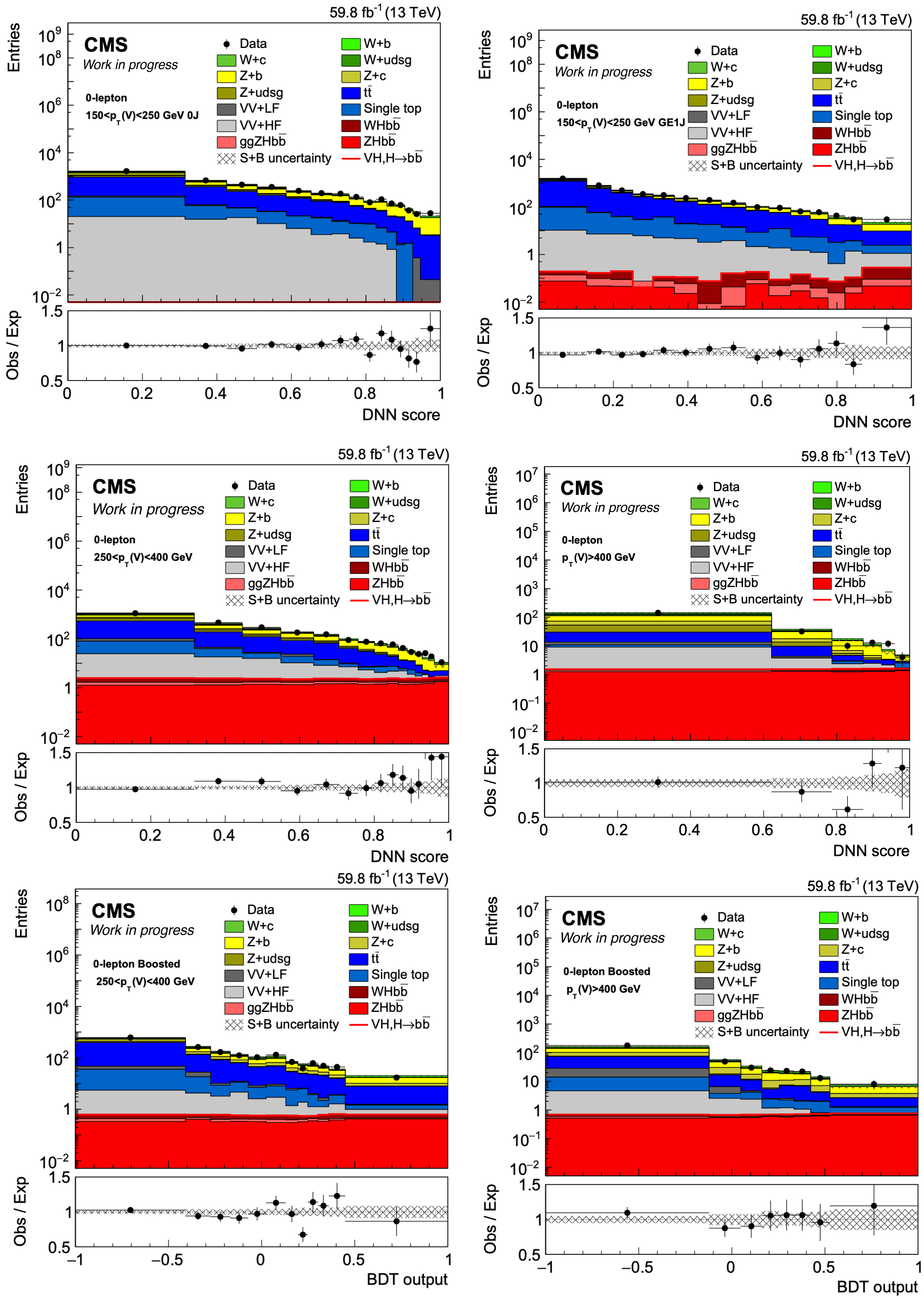


Figure 7.7: Signal regions post-fit distributions for 0-lepton channel signal regions in the 2018 analysis.

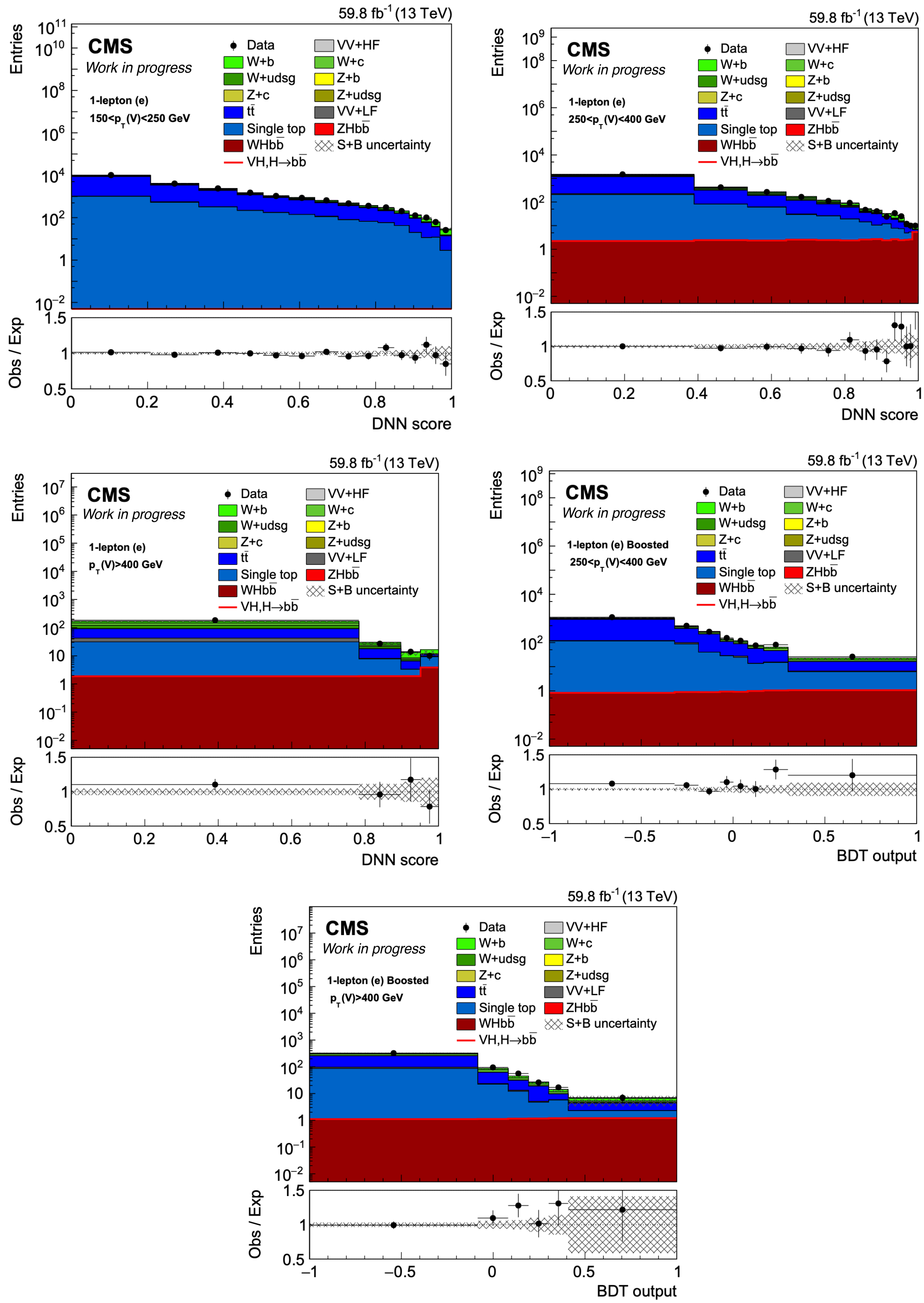


Figure 7.8: Signal regions post-fit distributions for 1-lepton channel signal regions in the 2018 analysis.

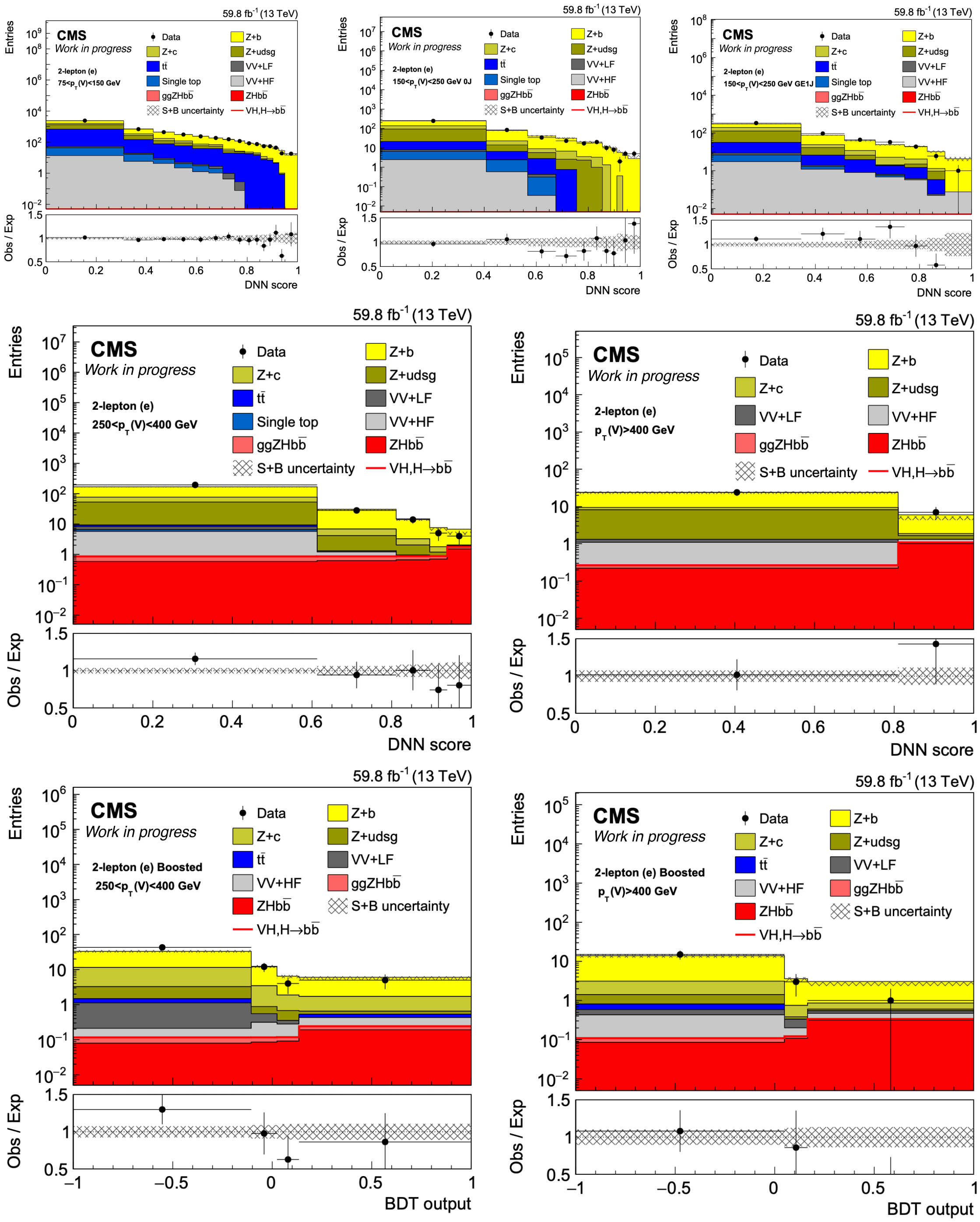


Figure 7.9: Signal regions post-fit distributions for 2-lepton channel signal regions in the 2018 analysis.

CHAPTER 8

SMEFT INTERPRETATION OF STXS VH(H $\to b\bar{b}$) MEASUREMENT

Effective field theories and the SMEFT in particular have been introduced in Section 2.3. The SMEFT framework is very well motivated by the large number of high precision measurements showing a good agreement with the SM predictions. This indicates that the BSM effects at the low energies are at most small, which supports the validity of SMEFT interpretations. The SMEFT framework provides a general methodology for the BSM searches that can be applied in many HEP sectors and combined in the global EFT fits. Moreover, the constraints of SMEFT operators can be matched to any BSM model defined at the UV scale.

The objective of any SMEFT interpretation analysis is to improve constraints on a relevant set of Wilson coefficients. While in global EFT analyses it is possible to constrain a large number of SMEFT operators, within one final state it is only possible to consider the contributions from a limited set of operators directly altering the considered production and decay mechanisms.

Nevertheless, the interpretations within experiments are still important. The additional experimental information allows to study and incorporate the effects that are often assumed to be negligible. For example, the impact of SMEFT operators on the experimental acceptance and the shape of analysis observables are not taken into account usually. In this work the SMEFT analysis is improved by considering the acceptance effects and the shape variations of analysis observables contributing to the multivariate discriminants distributions.

The SMEFT interpretation for the VH(H $\rightarrow b\bar{b}$) analysis is performed in the STXS framework, which allows to include the kinematic information and target the BSM regions. As it will be shown in this chapter, it is important to isolate the BSM specific regions as much as experimentally possible to improve the constraints on Wilson coefficients.

This chapter starts with a discussion of the SMEFT Warsaw basis — the full set of non redundant dimension-6 (dim-6) operators in Section 8.1. Then a set of operators and corresponding Willson coefficients relevant for the VH(H $\rightarrow b\bar{b}$) process is defined. The reweighting technique used to incorporate acceptance effects in the derived parametrisation is introduced in Section 8.3. The choices for the parametrisation strategy are discussed in Section 8.4 and supported by numerous studies presented in Section 8.5. The final parametrisation is then used to build a fit model and consequently derive the confidence intervals for all considered Wilson coefficients summarised at the end of Section 8.4.

8.1. SMEFT and the Warsaw basis.

The SMEFT dimension-6 Lagrangian consists of SM fields and satisfies the $SU(3)_c \times SU(2)_L \times U(1)_Y$ symmetry. The most general SMEFT dim-6 Lagrangian includes 2499 operators.

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{SM} + \sum_{i=1}^{N(6)} \frac{c_i^6}{\Lambda^2} O_i^6. \tag{8.1}$$

A substantial fraction of these operators is redundant and/or does not affect the matrix elements. It took many years before the complete set of non-redundant dim-6 operators was defined and became widely known as the Warsaw basis [89]. The authors of this work managed to reduce the number of operators to 59, retaining the completeness of the basis. It is highly non-trivial to measure the contributions from all of the 59 SMEFT operators, because they target a wide range of physics processes. With the accumulation of LHC data and combination of different final states in Higgs, EWK and top physics it becomes feasible to include a significant fraction of SMEFT operators in global EFT fits [28]. For a given Higgs production mode we can consider only a limited set of operators modifying the amplitudes of a considered process. In particular, as it is shown in [26], the VH production

SMEFT vertices include the following set of operators:

$$O_{Hq}^{(3)}, O_{Hq}^{(1)}, O_{Hu}, O_{Hd}, O_{HW}, O_{HWB} \tag{8.2}$$

The explicit form of these operators, corresponding Wilson coefficients and the interaction vertices are listed in Table 8.1. The Warsaw basis implementation is available within the `SMEFTsim` package [90] for various flavour symmetry assumptions. In this thesis a $U(3)^5$ flavour symmetry[1] is considered. In addition, to configure the SMEFTsim model one needs to fix observables based on the theoretical predictions with the set parameters (m_W, m_Z, G_F) or (α_{em}, m_Z, G_F), where the first one is usually referred to as m_W-scheme, and the second as α_{em}-scheme. In this work the m_W-scheme is used. By default, the Wilson coefficients are defined for the new physics scale $\Lambda = 1$ TeV.

[1] The maximal flavour symmetry unbroken by the SM Lagrangian kinetic terms $U(3)^5 = U(3)_q \times U(3)_u \times U(3)_d \times U(3)_l \times U(3)_e$.

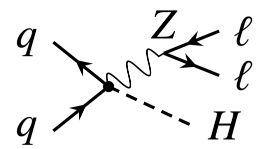
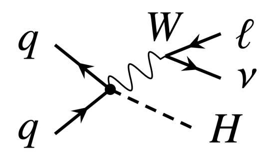
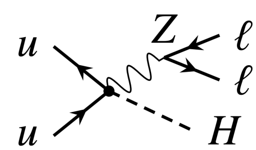
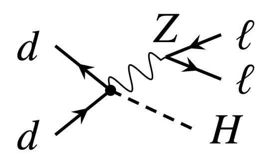
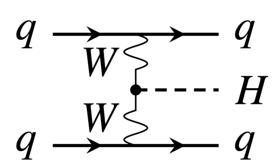
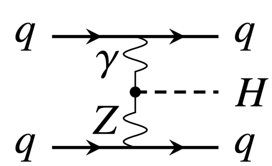
Parameter	Operator definition	Example diagram
$c_{Hq}^{(1)}$	$(H^\dagger i \overleftrightarrow{D}_\mu H)(\bar{q}_L \gamma^\mu q_L)$	
$c_{Hq}^{(3)}$	$(H^\dagger i \overleftrightarrow{D}^i_\mu H)(\bar{q}_L \sigma^i \gamma^\mu q_L)$	
c_{Hu}	$(H^\dagger i \overleftrightarrow{D}_\mu H)(\bar{u}_R \gamma^\mu u_R)$	
c_{Hd}	$(H^\dagger i \overleftrightarrow{D}_\mu H)(\bar{d}_R \gamma^\mu d_R)$	
c_{HW}	$(H^\dagger H)(W^i_{\mu\nu} W^{i,\mu\nu})$	
c_{HWB}	$(H^\dagger \sigma^i H)(W^i_{\mu\nu} B^{\mu\nu})$	

Table 8.1: Warsaw basis operators relevant for VH production.

8.2. Deriving the STXS parametrisation

The EFT effects are included through the parametrisation of the signal STXS bin, modifying only the normalisation of each STXS bin.

The parametrisation is derived for all ZH and WH STXS processes using the `SMEFTsim` package [90]. For the ggZH process it is not possible to derive the parametrisation using the LO `SMEFTsim`, because the corrections for this process are only introduced starting with the dimension 8 EFT operators. In the future studies the EFT modifications for ggZH process will be extracted using `SMEFT@NLO` [91]. In this work only the LO EFT corrections are considered, therefore the ggZH process is fixed at its SM prediction.

The signal yield in each STXS region expression already introduced in Section 6 is modified to included the dependence on the EFT operators as follows:

$$S^i = \mu^i(\vec{c})\epsilon^i(\vec{c})[\sigma \times B]^i_{SM} L. \tag{8.3}$$

Introducing an explicit dependence on the Wilson coefficients in acceptance term $\epsilon^i(\vec{c})$ ensures that the EFT acceptance effects are not neglected. The signal yields are modified by EFT effects with the term $\mu^i(\vec{c})$. The acceptance term is convoluted into the parametrisation equations, as discussed in detail in Section 8.5.1.

If we consider the expression for the EFT Lagrangian in Eq. 8.1 the matrix element can be written as follows

$$\mathrm{M}_{\mathrm{SMEFT}} = \mathrm{M}_{\mathrm{SM}} + \mathrm{M}_{\mathrm{BSM}}, \tag{8.4}$$

where the BSM term has a linear dependence on the Wilson coefficients ($\mathrm{M}_{\mathrm{BSM}} \propto \mathrm{c_j}$). Therefore, the total cross-section can be written as a sum of a SM-only term, a SM-BSM interference term and a pure BSM term:

$$\sigma^i_{\mathrm{SMEFT}} = \sigma^i_{\mathrm{SM}} + \sigma^i_{\mathrm{Int}} + \sigma^i_{\mathrm{BSM}}, \tag{8.5}$$

where the interference term is linearly dependent on the Wilson coefficient $\sigma^i_{\mathrm{Int}} \propto c_j$ and is suppressed as Λ^{-2}, and the pure BSM term, quadratically dependent on the Wilson coefficient $\sigma^i_{\mathrm{BSM}} \propto c_j c_k$ and is suppressed as Λ^{-4}. It is important to note here that the linear terms from dim-8 operators are suppressed by the same order of Λ as quadratic terms from dim-8. However, the dim-8 SMEFT expansion is not yet available. Therefore, providing the results obtained with linear-only terms in addition to those derived with the full parametrisation can help to estimate the effects from the dim-8 operators.

The signal strength modifier due to the BSM effects in a STXS bin i can be defined as the ratio of the total cross-section $\sigma^i_{\mathrm{SMEFT}}$ with respect to the SM cross-section, which is a quadratic function of Wilson coefficients:

$$\mu^i_{\mathrm{prod}}(\vec{C}) = \frac{\sigma^i_{\mathrm{SMEFT}}}{\sigma^i_{\mathrm{SM}}} = 1 + \sum_j A^i_j \, c_j + \sum_{jk} B^i_{jk} \, c_j \, c_k, \tag{8.6}$$

where A^i_j, B^i_{jk} are the constant factors defining the EFT scaling functions in each STXS bin.

The same discussion can be applied to the EFT effects for the Higgs boson decay, which is decoupled from the Higgs boson production process, due to the narrow Higgs boson width approximation. It also means that the decay scaling functions can be derived inclusively for all STXS bins. For the branching fraction parametrisation it is important to consider the EFT effects Higgs boson total width:

$$\mu_{decay} = \frac{\mathcal{B}_{\mathrm{SMEFT}}}{\mathcal{B}_{\mathrm{SM}}} = \frac{\Gamma^{b\bar{b}}_{\mathrm{SMEFT}}/\Gamma^{b\bar{b}}_{SM}}{\Gamma^{\mathrm{tot}}_{\mathrm{SMEFT}}/\Gamma^{\mathrm{tot}}_{\mathrm{SM}}} = \frac{1 + \sum_j A^{b\bar{b}}_j \, C_j + \sum_{jk} B^{b\bar{b}}_{jk} \, C_j \, C_k}{1 + \sum_j A^{\mathrm{tot}}_j \, C_j + \sum_{jk} B^{\mathrm{tot}}_{jk} \, C_j \, C_k}, \tag{8.7}$$

where Γ^{tot}, $\Gamma^{b\bar{b}}$ are the total and partial Higgs widths. The parametrisation for $\Gamma^{b\bar{b}}_{\mathrm{SMEFT}}$ was derived in this work following the same approach as for the cross-sections parametrisation, while the equations for $\Gamma^{\mathrm{tot}}_{\mathrm{SMEFT}}$ are taken from the results provided in [92].

For the total parametrisation we obtain the following:

$$\mu^{\text{total}}(\vec{C}) = \mu^{i}_{\text{prod}}(\vec{C}) \cdot \mu_{\text{decay}}(\vec{C}). \tag{8.8}$$

For the VH(H $\rightarrow b\bar{b}$) process the EFT operators contributing to production and decay processes do not overlap. In this work only the contributions of the production EFT operators are considered.

8.3. Standalone reweighting

The EFT effects in the `SMEFTsim` model can be simulated using the `SMEFTsim_UFO` model available in `Madgraph` framework. Using the Madgraph reweighting feature, events generated at the SM point ($\vec{c} = 0$) can be reweighted according to different assumptions in the `SMEFTsim` model parameter space. The weights are defined as the ratio of matrix elements:

$$W_{c_j>0} = \frac{\mathcal{M}_{\text{SMEFT}}(c_j > 0)}{\mathcal{M}_{\text{SM}}} \cdot W_{\text{SM}}, \tag{8.9}$$

where $\mathcal{M}_{\text{SMEFT}}$ is the matrix element for BSM hypothesis($c_j > 0$), $\mathcal{M}_{\text{SM}}$ is the matrix element for the SM hypothesis($c_j = 0$). To maintain the validity of such a reweighting it is important to note that the reweighting points should be close to the SM so that the reweighted phase space does not significantly differ from the SM phase space.

To derive the scaling functions from Eq. 8.6, the EFT cross-section is evaluated for arbitrary values of Wilson coefficients $c_j > 0$. The cross-sections are calculated by reweighting the SM MC ($c_j = 0$) as shown in Eq. 8.9. To extract the linear term coefficients A_j and quadratic term coefficients B_{jj} the considered Wilson coefficient c_j is set to the values c, $2c$[2], resulting in $2N$ EFT points for linear and quadratic terms if N Wilson coefficients are considered. To extract the coefficients for the cross-terms $B_{jk}(j \neq k)$, the cross-section is evaluated for two different Wilson coefficients set to arbitrary $c > 0$ values at the same time. Using the already derived A_j, B_{jj}, the $B_{jk}(j \neq k)$ can be estimated, resulting in additional $\frac{N(N-1)}{2}$

[2]The values are chosen to be close to SM to maintain the validity of the reweighting procedure.

evaluated BSM points.

The number of reweighting points is defined by the number of the probed points in BSM parameter space. As discussed above, to derive the parametrisation for linear and quadratic terms one needs to consider $2N$ EFT points, and $\frac{N(N-1)}{2}$ for the coefficients entering the cross-terms. In total, including the SM point ($\vec{c} = 0$), we obtain:

$$N_{\text{weights}} = 1 + 2N + \frac{N(N-1)}{2}, \tag{8.10}$$

where N is the number of EFT operators. In this work six operators are considered, which amounts to 28 reweighting points.

Using the **Madgraph** reweighting functionality embedded within the EFT2Obs package [93] and NanoAOD reweighting tool [94][3] it is possible to create a standalone reweighting module implemented in `python`, that can be applied to a SM MC sample. It is a convenient alternative to the costly and time-consuming approach of generating the new MC samples for each EFT hypothesis.

For the VH(H $\rightarrow b\bar{b}$) analysis EFT effects are modelled by reweighting the default SM samples generated at `MiNLO` with the `POWHEG-BOX` framework. Naturally, one should question the validity of using a LO **Madgraph** matrix element to reweight a `POWHEG` `MiNLO` prediction. The agreement between the SM `POWHEG` `MiNLO` and **Madgraph** was studied and found to be sufficient, as can be seen for example in the p_T^V distribution comparison shown in Figure 8.1.

[3]Developed by CMS collaborators and publicly available

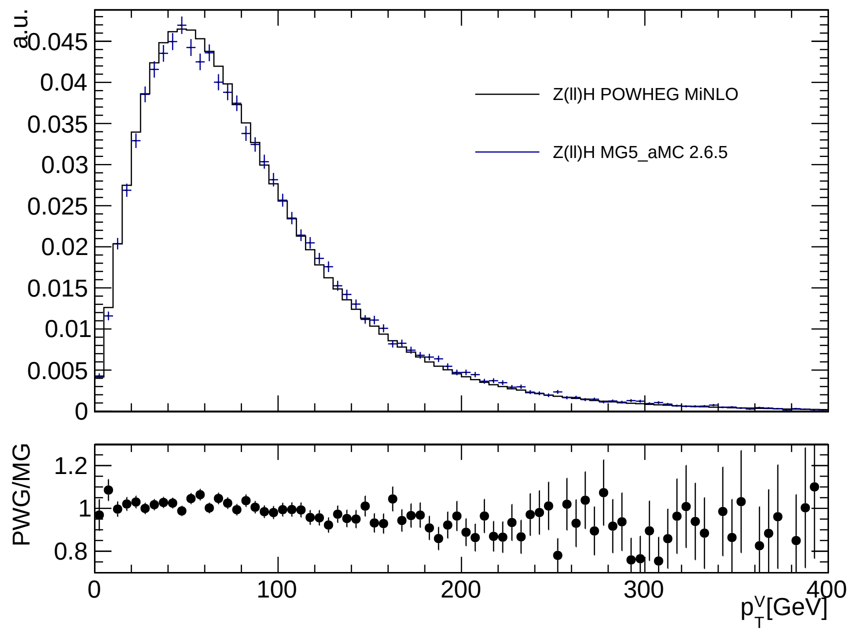


Figure 8.1: p_T^V distribution for ZH(ll) `Madgraph` and `POWHEG` MiNLO events. The uncertainties shown for each point are statistical.

8.4. Final parametrisation and expected limits on SMEFT Wilson coefficients

In this section the EFT parametrisation derived for VH(H $\to b\bar{b}$) process is summarised, followed by the expected SMEFT interpretation results based on the full Run 2 VH(H $\to b\bar{b}$) STXS measurement.

The scaling functions are extracted in STXS bins with the additional boundary at $p_T^V = 400$ GeV (modified STXS stage 1.2 bins), which improves the sensitivity to SMEFT effects with respect to the default STXS stage 1.2 bins as shown in Section 8.5. The full parametrisation includes linear and quadratic terms. The results derived with linear only terms are also provided to isolate the constraints from SM-BSM interference term in Eq. 8.5. The acceptance effects are detailed in Section 8.5.1 and fully incorporated into the final results.

The derived parametrisation is summarised in Table 8.2 and visualised as SMEFT signal strength dependence on Wilson coefficients shown in Figures 8.2, 8.3. As expected, there is a strong p_T^V dependence for most of the considered operators, but especially for $O_{Hq}^{(3)}$ and $O_{Hq}^{(1)}$. The $O_{HW}^{(1)}$ and $O_{HWB}^{(1)}$ operators create overall modification of the VH cross-section.

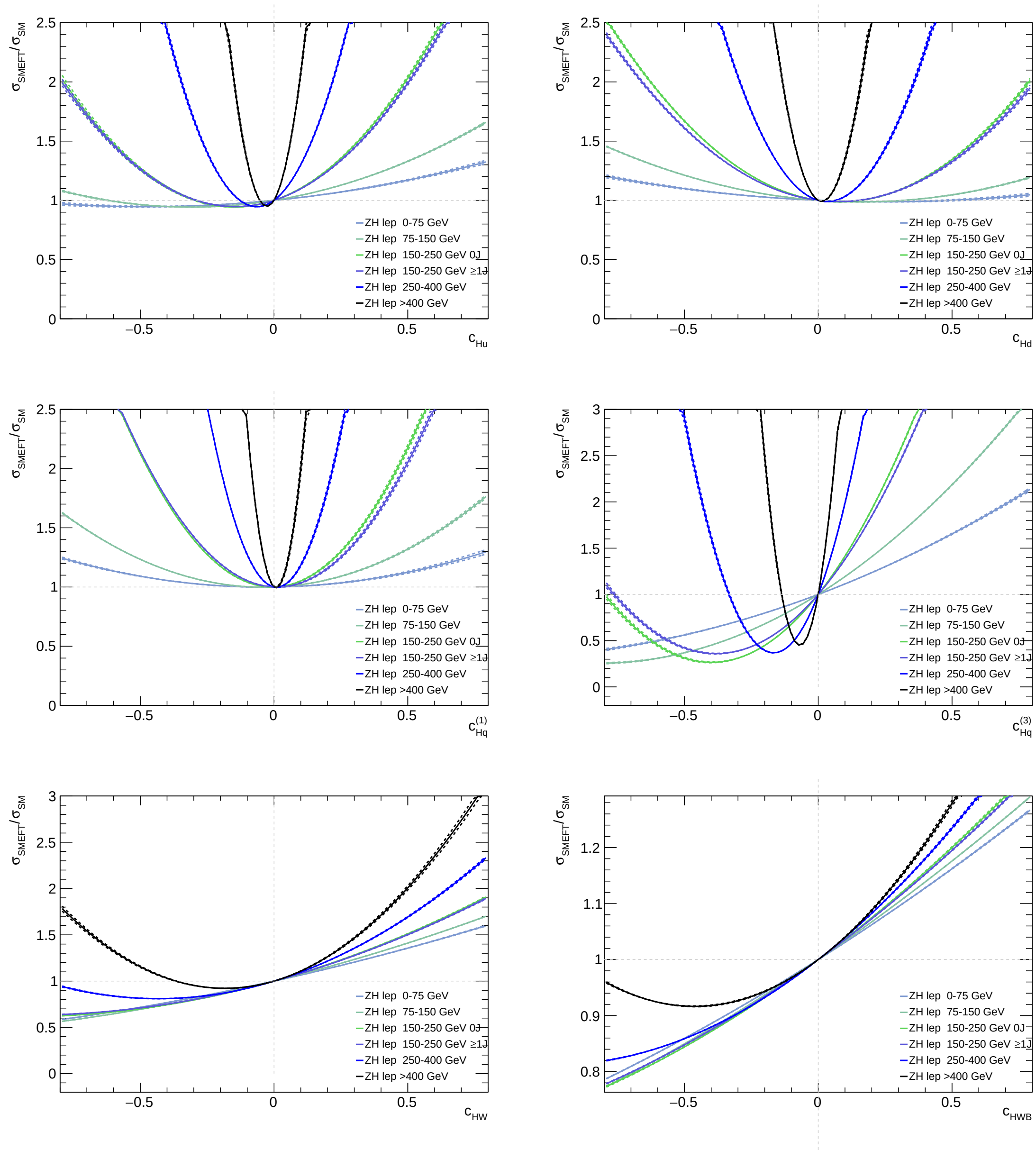


Figure 8.2: Cross-section scaling functions for the modified ZH STXS stage 1.2 bins for each of the $O_{Hq}^{(3)}, O_{Hq}^{(1)}, O_{Hu}, O_{Hd}, O_{HW}, O_{HWB}$ operators. The equations are listed in Table 8.2

STXS bin	Scaling function
ZH lep 0-75 GeV	$1 + 1.09\, c_{Hq}^{(3)} + 0.22\, c_{Hu} + 0.64\, c_{HW} + 0.30\, c_{HWB} + 0.43\, {c_{Hq}^{(1)}}^2 + 0.43\, {c_{Hq}^{(3)}}^2 + 0.23\, {c_{Hu}}^2 + 0.20\, {c_{Hd}}^2 + 0.15\, {c_{HW}}^2 + 0.44\, c_{Hq}^{(3)} c_{HW} + 0.17\, c_{Hq}^{(3)} c_{HWB}$
ZH lep 75-150 GeV	$1 + 1.81\, c_{Hq}^{(3)} + 0.37\, c_{Hu} - 0.17\, c_{Hd} + 0.72\, c_{HW} + 0.33\, c_{HWB} + 1.11\, {c_{Hq}^{(1)}}^2 + 1.11\, {c_{Hq}^{(3)}}^2 + 0.59\, {c_{Hu}}^2 + 0.52\, {c_{Hd}}^2 + 0.21\, {c_{HW}}^2 - 0.14\, c_{Hq}^{(1)} c_{Hq}^{(3)} + 0.80\, c_{Hq}^{(3)} c_{HW} + 0.30\, c_{Hq}^{(3)} c_{HWB} + 0.11\, c_{Hu} c_{HWB} + 0.14\, c_{HW} c_{HWB}$
ZH lep 150-250 GeV 0J	$1 + 3.68\, c_{Hq}^{(3)} + 0.77\, c_{Hu} - 0.33\, c_{Hd} + 0.81\, c_{HW} + 0.36\, c_{HWB} + 4.60\, {c_{Hq}^{(1)}}^2 + 4.60\, {c_{Hq}^{(3)}}^2 + 2.60\, {c_{Hu}}^2 + 2.03\, {c_{Hd}}^2 + 0.43\, {c_{HW}}^2 - 1.10\, c_{Hq}^{(1)} c_{Hq}^{(3)} - 0.15\, c_{Hq}^{(1)} c_{HW} + 1.81\, c_{Hq}^{(3)} c_{HW} + 0.65\, c_{Hq}^{(3)} c_{HWB} + 0.27\, c_{Hu} c_{HWB} - 0.11\, c_{Hd} c_{HWB} + 0.25\, c_{HW} c_{HWB}$
ZH lep 150-250 GeV >= $1J$	$1 + 3.35\, c_{Hq}^{(3)} + 0.73\, c_{Hu} - 0.28\, c_{Hd} + 0.79\, c_{HW} + 0.35\, c_{HWB} + 4.40\, {c_{Hq}^{(1)}}^2 + 4.39\, {c_{Hq}^{(3)}}^2 + 2.50\, {c_{Hu}}^2 + 1.87\, {c_{Hd}}^2 + 0.42\, {c_{HW}}^2 - 1.26\, c_{Hq}^{(1)} c_{Hq}^{(3)} - 0.24\, c_{Hq}^{(1)} c_{HW} + 1.80\, c_{Hq}^{(3)} c_{HW} + 0.62\, c_{Hq}^{(3)} c_{HWB} + 0.28\, c_{Hu} c_{HWB} - 0.11\, c_{Hd} c_{HWB} + 0.25\, c_{HW} c_{HWB}$
ZH lep 250-400 GeV	$1 - 0.44\, c_{Hq}^{(1)} + 7.58\, c_{Hq}^{(3)} + 1.69\, c_{Hu} - 0.63\, c_{Hd} + 0.88\, c_{HW} + 0.38\, c_{HWB} + 22.75\, {c_{Hq}^{(1)}}^2 + 22.75\, {c_{Hq}^{(3)}}^2 + 13.22\, {c_{Hu}}^2 + 9.52\, {c_{Hd}}^2 + 1.02\, {c_{HW}}^2 + 0.19\, {c_{HWB}}^2 - 7.68\, c_{Hq}^{(1)} c_{Hq}^{(3)} - 0.74\, c_{Hq}^{(1)} c_{HW} + 4.55\, c_{Hq}^{(3)} c_{HW} + 1.51\, c_{Hq}^{(3)} c_{HWB} + 0.75\, c_{Hu} c_{HWB} - 0.28\, c_{Hd} c_{HWB} + 0.54\, c_{HW} c_{HWB}$
ZH lep >400 GeV	$1 - 1.76\, c_{Hq}^{(1)} + 16.05\, c_{Hq}^{(3)} + 3.75\, c_{Hu} - 1.24\, c_{Hd} + 0.85\, c_{HW} + 0.36\, c_{HWB} + 117.30\, {c_{Hq}^{(1)}}^2 + 117.31\, {c_{Hq}^{(3)}}^2 + 70.00\, {c_{Hu}}^2 + 46.94\, {c_{Hd}}^2 + 2.33\, {c_{HW}}^2 + 0.39\, {c_{HWB}}^2 - 46.57\, c_{Hq}^{(1)} c_{Hq}^{(3)} - 2.33\, c_{Hq}^{(1)} c_{HW} + 10.58\, c_{Hq}^{(3)} c_{HW} + 3.37\, c_{Hq}^{(3)} c_{HWB} + 1.79\, c_{Hu} c_{HWB} - 0.60\, c_{Hd} c_{HWB} + 1.17\, c_{HW} c_{HWB}$
WH lep 0-75 GeV	$1 + 1.12\, c_{Hq}^{(3)} + 0.80\, c_{HW} + 0.48\, {c_{Hq}^{(3)}}^2 + 0.20\, {c_{HW}}^2 + 0.50\, c_{Hq}^{(3)} c_{HW}$
WH lep 75-150 GeV	$1 + 1.93\, c_{Hq}^{(3)} + 0.92\, c_{HW} + 1.42\, {c_{Hq}^{(3)}}^2 + 0.35\, {c_{HW}}^2 + 1.01\, c_{Hq}^{(3)} c_{HW}$
WH lep 150-250 GeV 0J	$1 + 4.14\, c_{Hq}^{(3)} + 0.94\, c_{HW} + 5.52\, {c_{Hq}^{(3)}}^2 + 0.72\, {c_{HW}}^2 + 2.18\, c_{Hq}^{(3)} c_{HW}$
WH lep 150-250 GeV >= $1J$	$1 + 3.71\, c_{Hq}^{(3)} + 0.97\, c_{HW} + 5.45\, {c_{Hq}^{(3)}}^2 + 0.75\, {c_{HW}}^2 + 2.59\, c_{Hq}^{(3)} c_{HW}$
WH lep 250-400 GeV	$1 + 8.23\, c_{Hq}^{(3)} + 0.96\, c_{HW} + 25.35\, {c_{Hq}^{(3)}}^2 + 1.43\, {c_{HW}}^2 + 4.85\, c_{Hq}^{(3)} c_{HW}$
WH lep >400 GeV	$1 + 18.16\, c_{Hq}^{(3)} + 0.92\, c_{HW} + 178.67\, {c_{Hq}^{(3)}}^2 + 3.32\, {c_{HW}}^2 + 14.48\, c_{Hq}^{(3)} c_{HW}$

Table 8.2: Scaling functions for the VH STXS stage 1.2 bins.

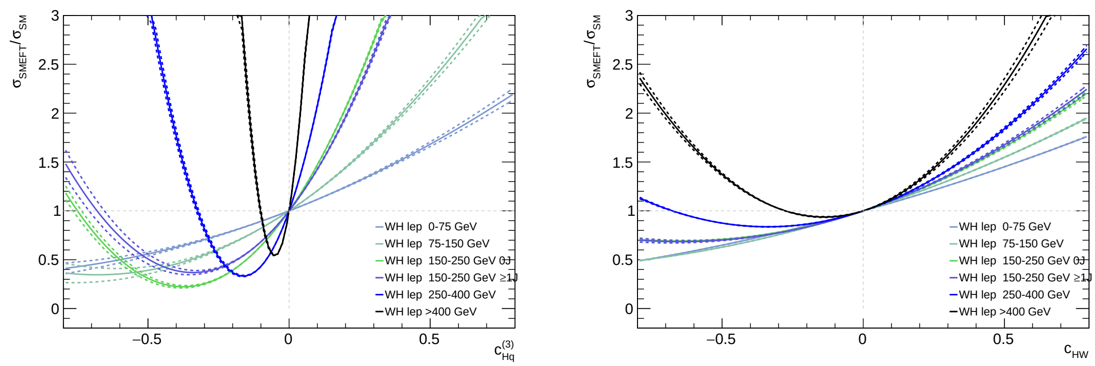


Figure 8.3: Cross-section scaling functions for the modified WH STXS stage 1.2 bins for the $O_{Hq}^{(3)}, O_{HW}$ operators. The functions are listed in Table 8.2

Functions listed in Table 8.2 are included into the full Run 2 VH(H $\rightarrow b\bar{b}$) likelihood, by scaling the signal strength in each STXS bin, as described in Section 8.2. The likelihood-scans derived in the full Run 2 VH(H $\rightarrow b\bar{b}$) fit with all systematical uncertainties included are shown in Fig. 8.4. The likelihood curves derived with the linear parametrisation are also shown. The expected confidence intervals are summarised in Fig. 8.5. It is evident that with the full parametrisation the constraints are improved for all Wilson coefficients, most significantly for $c_{Hq}^{(1)}$, c_{Hu} and c_{Hd}. The best sensitivity with the full parametrisation is obtained for the operators with a strong $\mathrm{p_T^V}$ dependence such as $O_{Hq}^{(3)}$ and $O_{Hq}^{(1)}$. The high granularity measurements in BSM-specific regions is essential to further improve these measurements.

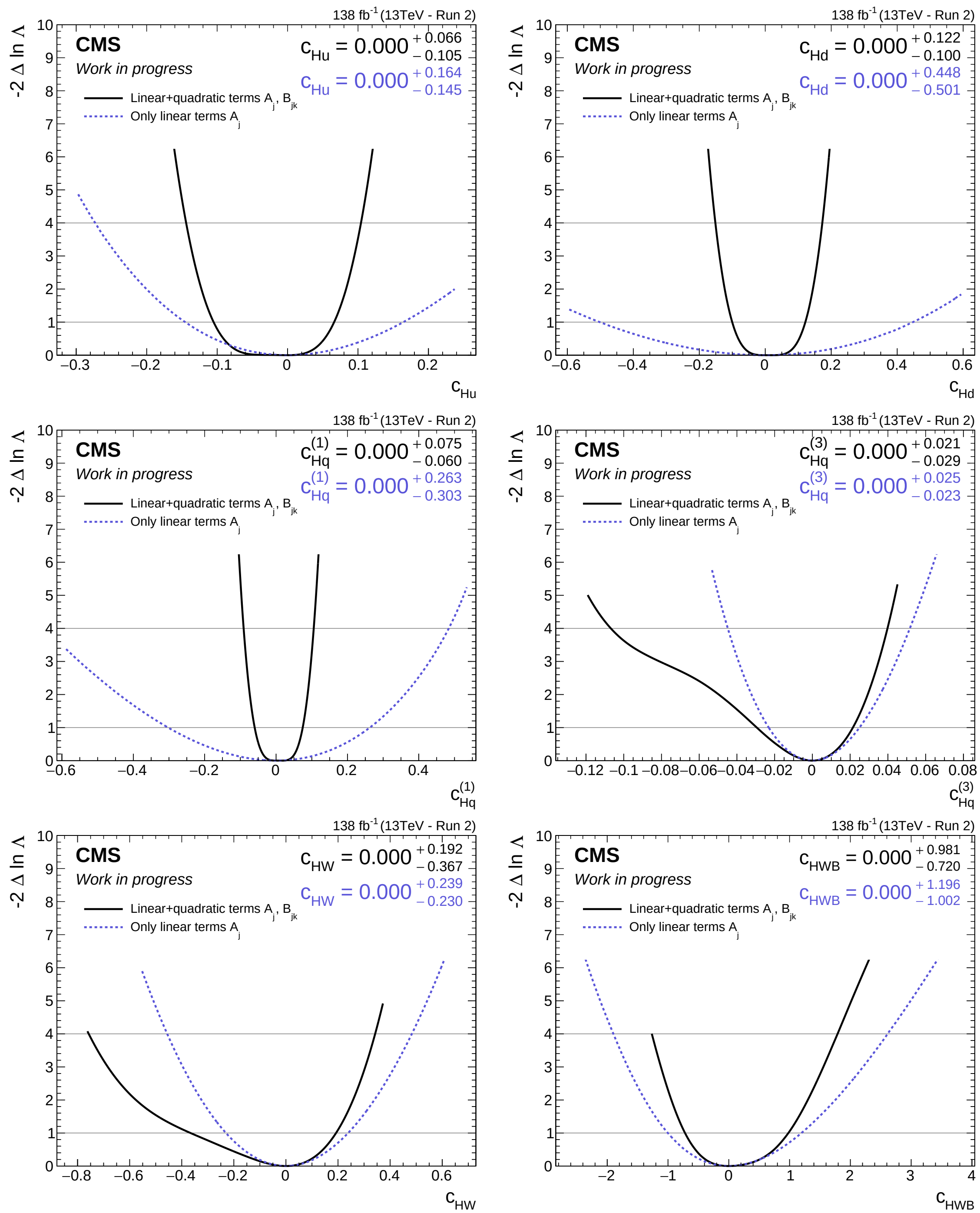


Figure 8.4: The expected likelihood curves for the six dim-6 SMEFT operators considered in this work extracted with linear only (black) and linear-plus-quadratic (violet) parametrisation. The fits are performed by considering EFT effects for a single operator.

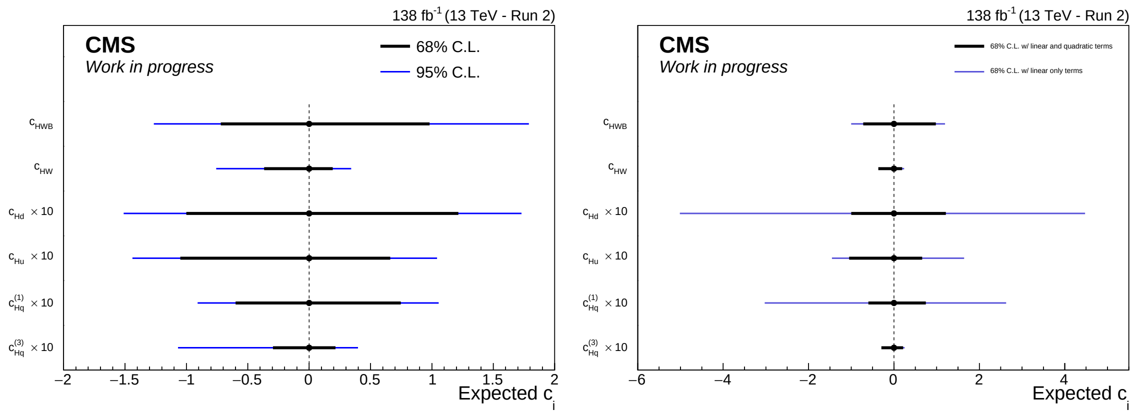


Figure 8.5: Expected 68% and 95% CL intervals obtained with linear-plus-quadratic parametrisation are shown in the left plot. The comparison of 68% CL derived with the full parametrisation and linear only is shown in the right plot.

In Fig. 8.6 the effect of EFT effects on cross-sections in STXS bins is shown. The Wilson coefficients are set to the values corresponding to 1σ boundaries extracted with the full parametrisation and summarised in Figure 8.5. The WH production is affected by $c_{Hq}^{(3)}$ and c_{HW}, while the ZH is altered by all six operators, and in Fig. 8.6 the variations from operators affecting only ZH production are shown. This figure illustrates the $\mathrm{p_T^V}$-dependency of all six Wilson coefficients, in particular highlighting the importance $\mathrm{p_T^V} > 250$ GeV region for the $c_{Hq}^{(1)}$, $c_{Hq}^{(3)}$, c_{Hu} and c_{Hd} Wilson coefficients.

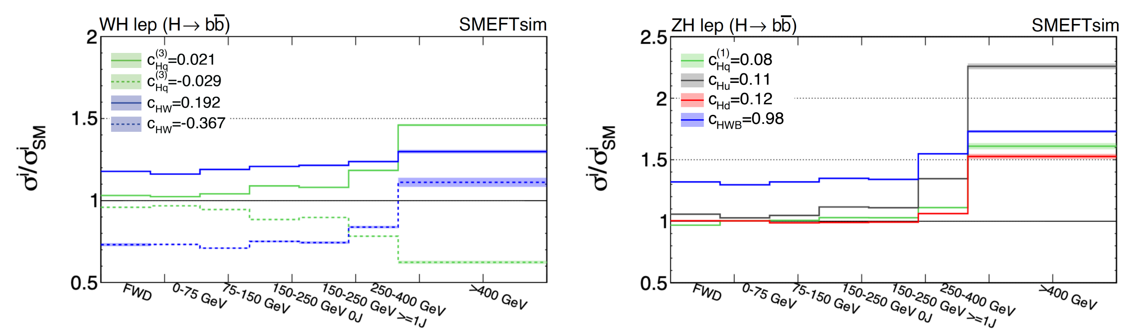


Figure 8.6: The modified STXS stage 1.2 for WH (left) and ZH (right) processes with the SMEFT parameters set at upper 68% CL boundaries.

The results summarised in Figures 8.4,8.5 are obtained by considering EFT effects from a single SMEFT operator. This is a limitation of performing SMEFT interpretation of an individual measurement where only one production mechanism is considered. Attempting to perform the simultaneous fit is problematic due to large correlations between the SMEFT parameters, as shown in Figure 8.7.

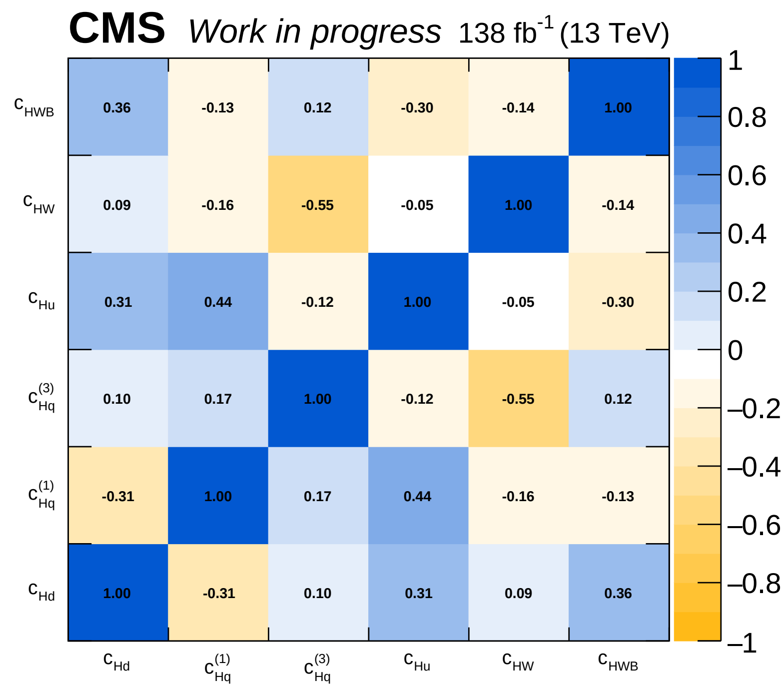


Figure 8.7: Expected correlation matrix for the Wilson coefficients.

In the future combination with other production and final states it will be possible to float several Wilson coefficients simultaneously. In this analysis two dimensional fits were also performed. The results are shown in Figure 8.8, where the comparison of the contours obtained with full parametrisation and linear-only parametrisation is provided. The quadratic terms entering the full parametrisation help to reduce the correlation between the considered pairs $O^{(3)}_{Hq}$-vs-O_{Hu} and $O^{(3)}_{Hq}$-vs-O_{HW}, and the constraints are significantly reduced especially for the O_{Hu} operator.

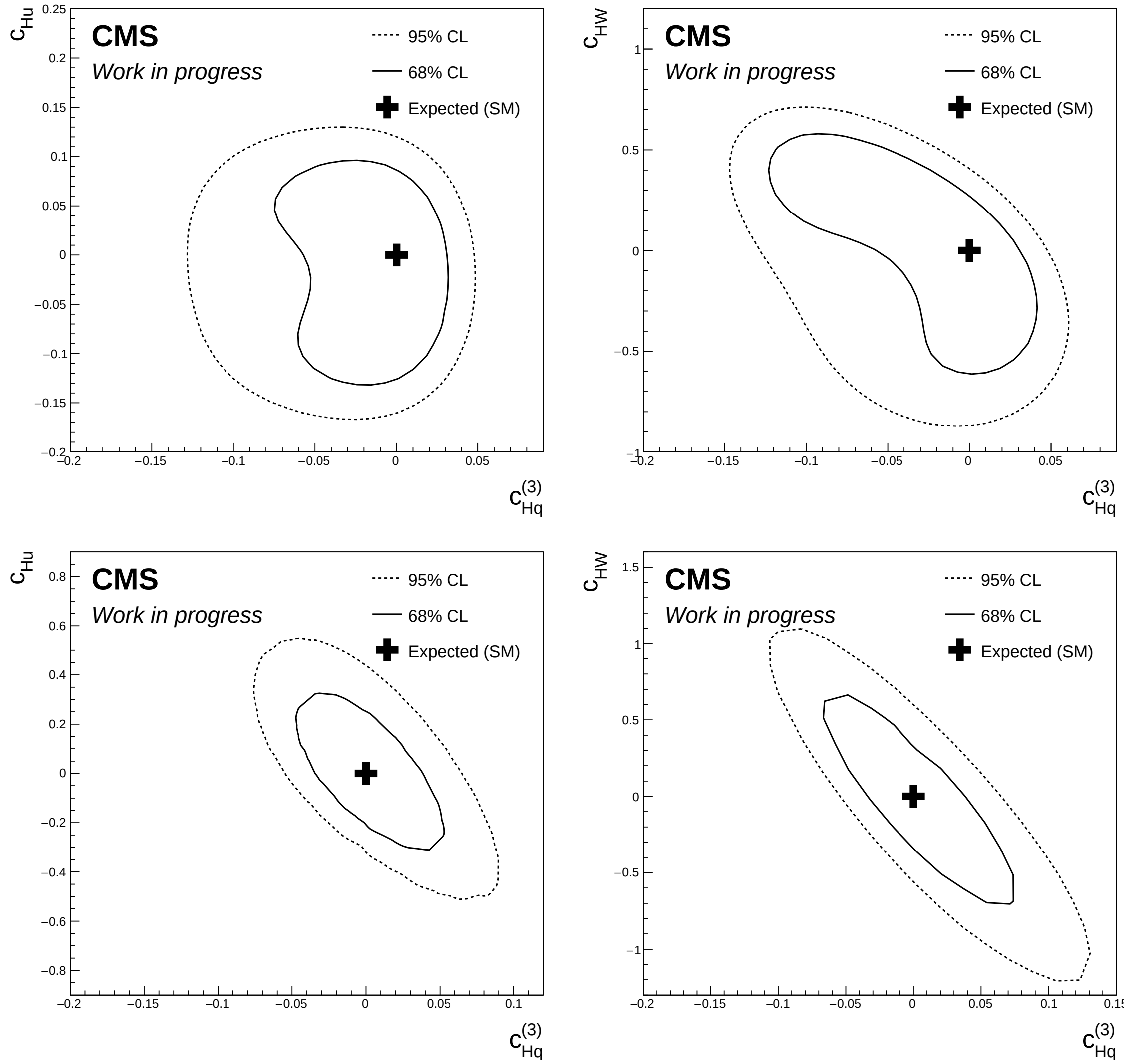


Figure 8.8: The expected 68% and 95% CL contours obtained with the full parametrisation (top row) and linear only (bottom row) for the $O_{Hq}^{(3)}$ vs O_{Hu} (left) and $O_{Hq}^{(3)}$ vs O_{HW} (right) operators.

8.4.1. Sensitivity in resolved analysis only

The sensitivity from the VH(H $\rightarrow b\bar{b}$) resolved analysis was compared with the baseline fit where the resolved and boosted analyses are combined in Figure 8.9. While the resolved analysis alone is driving the sensitivity for all operators, the boosted categories slightly help to improve the constraints.

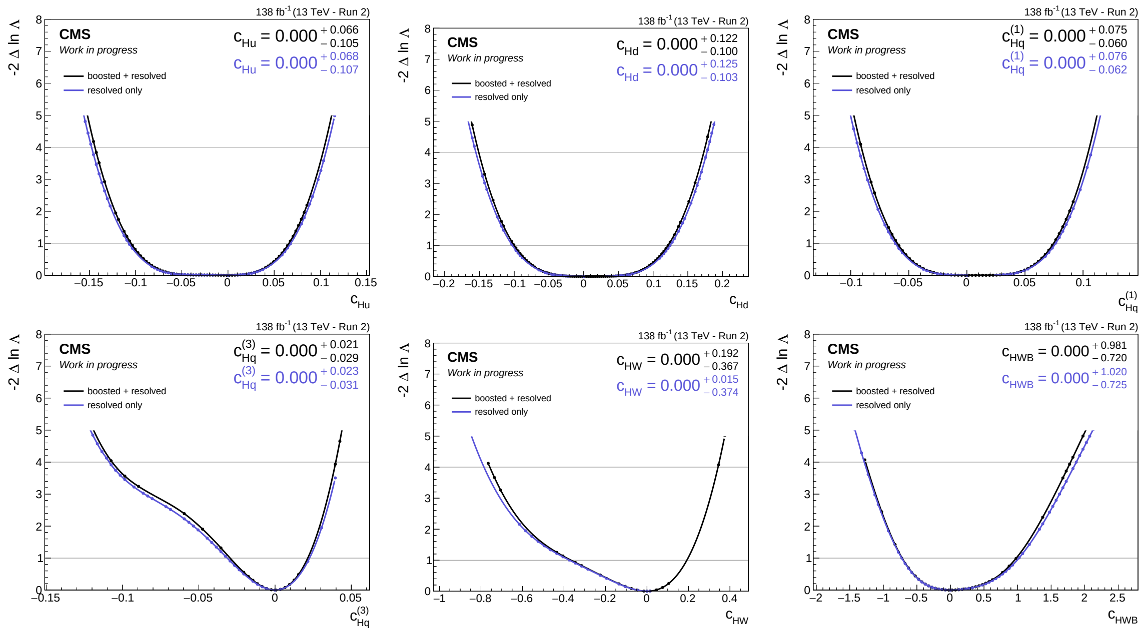


Figure 8.9: The likelihood curves for the six dim-6 SMEFT operators considered in this work. The violet line corresponds to the limits extracted from the resolved analysis only and the black line shows the combined analysis results. The SMEFT parametrisation includes both linear and quadratic terms. The fits are performed by considering EFT effects for a single operator.

8.4.2. Comparison to ATLAS results

The ATLAS collaboration released a conference note summarising the SMEFT interpretation of full Run 2 STXS VH($H \to b\bar{b}$) measurement [95]. In this document they report the combination of their already published boosted and resolved analyses, and interpret the results using the Warsaw basis implemented in `SMEFTsim` model. One should keep in mind the following differences when comparing the extracted confidence intervals listed in Table 8.3:

- In the ATLAS analysis the $p_T^V > 400$ GeV is populated by boosted events, while in this work the boosted and resolved Higgs decay topologies both enter full p_T^V region
- The acceptance effects on SMEFT cross-section are not included in the ATLAS results and reported to be below 10%.

Overall the constraints derived in this work are of a similar sensitivity as the results reported by ATLAS, as shown in Table 8.3.

Wilson coefficient	ATLAS result	CMS result
$c_{Hq}^{(3)} \times 100$	${}^{+2.2}_{-2.0}$	${}^{+2.1}_{-2.9}$
$c_{Hq}^{(1)} \times 100$	-	${}^{+7.5}_{-6.0}$
$c_{Hu} \times 100$	${}^{+4.6}_{-10.0}$	${}^{+6.3}_{-10.5}$
$c_{Hd} \times 10$	-	${}^{+1.2}_{-1.0}$
$c_{HWB} \times 10$	${}^{+11.0}_{-4.4}$	${}^{+9.81}_{-7.2}$
$c_{HW} \times 10$	${}^{+1.7}_{-3.2}$	${}^{+1.9}_{-3.7}$

Table 8.3: The expected 68% CL for the SMEFT Wilson coefficients obtained in this work compared with ATLAS results [95]. The results were extracted with linear and quadratic terms included in the parametrisation.

8.5. Parametrisation studies

8.5.1. Acceptance effects

The acceptance effects were estimated by comparing the parametrisation derived before the application of analysis selection with the parametrisation derived on selected set of events. In Figure 8.10 these equations are compared for the $c_{Hq}^{(3)}, c_{Hq}^{(1)}, c_{Hu}$ operators in all considered STXS bins. The acceptance effects can be estimated as the relative difference in the parametrisation derived before and after applying the analysis selection. It is evident that the acceptance has a larger effect in low $\mathrm{p_T^V}$ region, where the selection modifies the phase space due to the reconstruction level $\mathrm{p_T^V}$ selection applied in the VH(H $\rightarrow b\bar{b}$) analysis. The acceptance effects can be as high as 10% in some of the STXS bins, therefore for the final parametrisation the acceptance will be fully incorporated in the scaling functions.

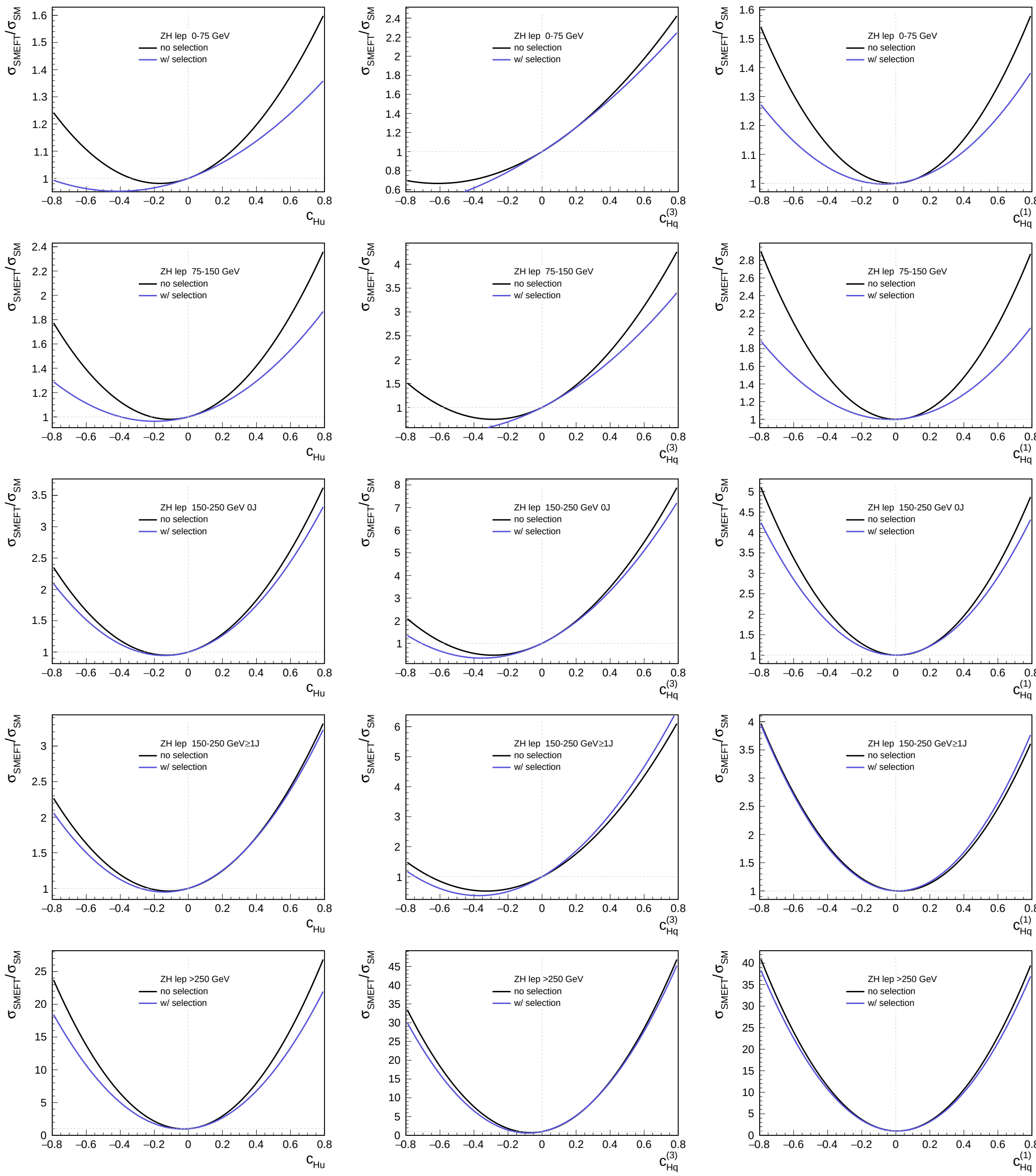


Figure 8.10: Parametrisation derived in STXS stage 1.2 bins with and without acceptance effects taken into account for the $c_{Hq}^{(3)}, c_{Hq}^{(1)}, c_{Hu}$ operators.

8.5.2. EFT shape effects within STXS bins

The STXS SMEFT parametrisation assumes that the BSM variations only change the total cross-section inside each STXS bin. At the same time, depending on the analysis observable, the EFT reweighting can introduce shape variations inside the STXS categories. In the VH(H $\rightarrow b\bar{b}$) analysis, the main observables are the DNN score (resolved) and the BDT discriminant (boosted). In Figure 8.11 the DNN and BDT distributions are shown. Clearly, the shape effect is significant for the boosted observable, while the resolved observable is not affected in shape.

A different behaviour of resolved and boosted observables under EFT variations can be explained by the choice of input features that were used for MVA training. The main difference is the presence and importance of the $\mathrm{p_T^V}$ variable in the set of boosted BDT features, in the resolved DNN the leading features are the Higgs candidate kinematic observables.

The EFT production vertices introduce a very noticeable $\mathrm{p_T^V}$ dependence, which reflects on the BDT output due to the high level of correlation with the vector boson transverse momentum $\mathrm{p_T^V}$. The BDT dependence on EFT variations can be taken into account by introducing the shape templates for various EFT scenarios into the final SMEFT fit; by a finer granularity of the STXS bins; or one can derive a separate parametrisation for various BDT regions. While the first approach can be quite complicated and also goes beyond the SMEFT interpretation of STXS measurements, the last two options were studied and summarised in the following sections.

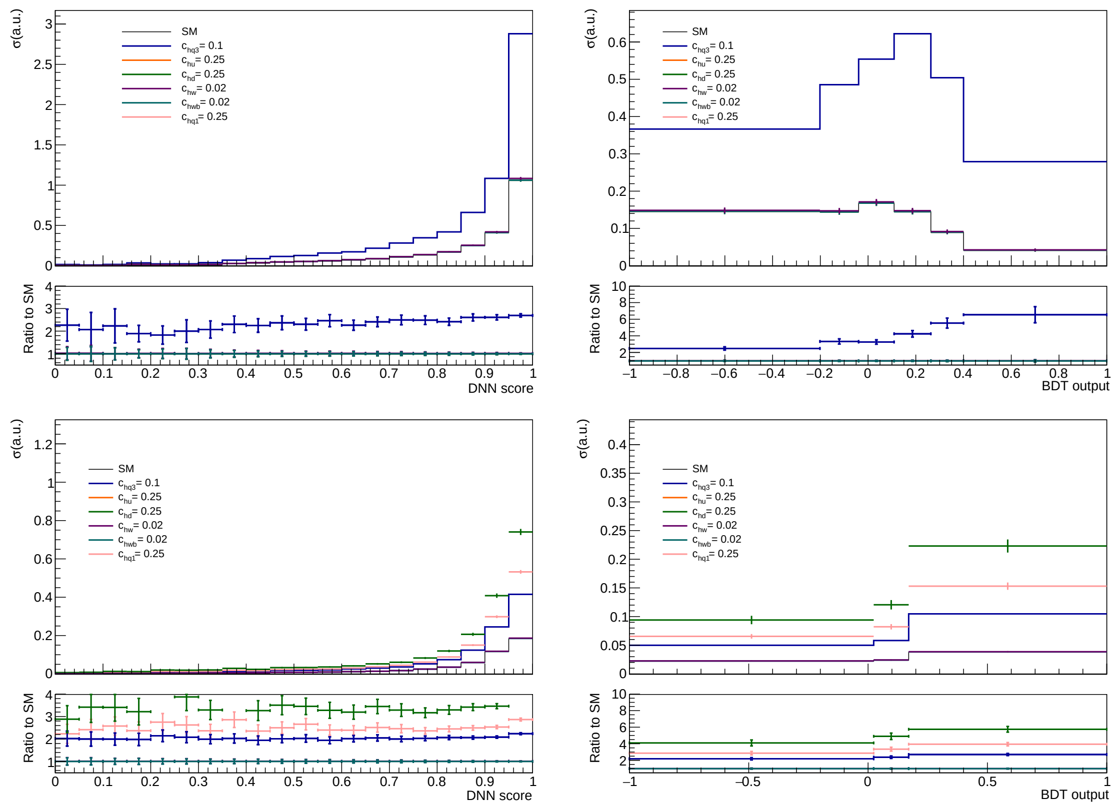


Figure 8.11: DNN output distribution for resolved analysis (left) and boosted (right) under EFT variations for WH (upper row) and ZH (lower row) p_T^V>250 GeV STXS bins.

Scaling functions and DNN distributions in STXS stage 1.2 with p_T^V boundary at 400 GeV

In Fig. 8.12 the BDT output distributions in boosted categories are shown for the modified STXS 1.2 binning, i.e. with an additional p_T^V boundary at 400 GeV. Clearly, the shape effect is very minor for the VH (ZH and WH) 250<p_T^V<400 GeV bin and ZH p_T^V>400 GeV, but it is still noticeable for the WH p_T^V>400 GeV bin.

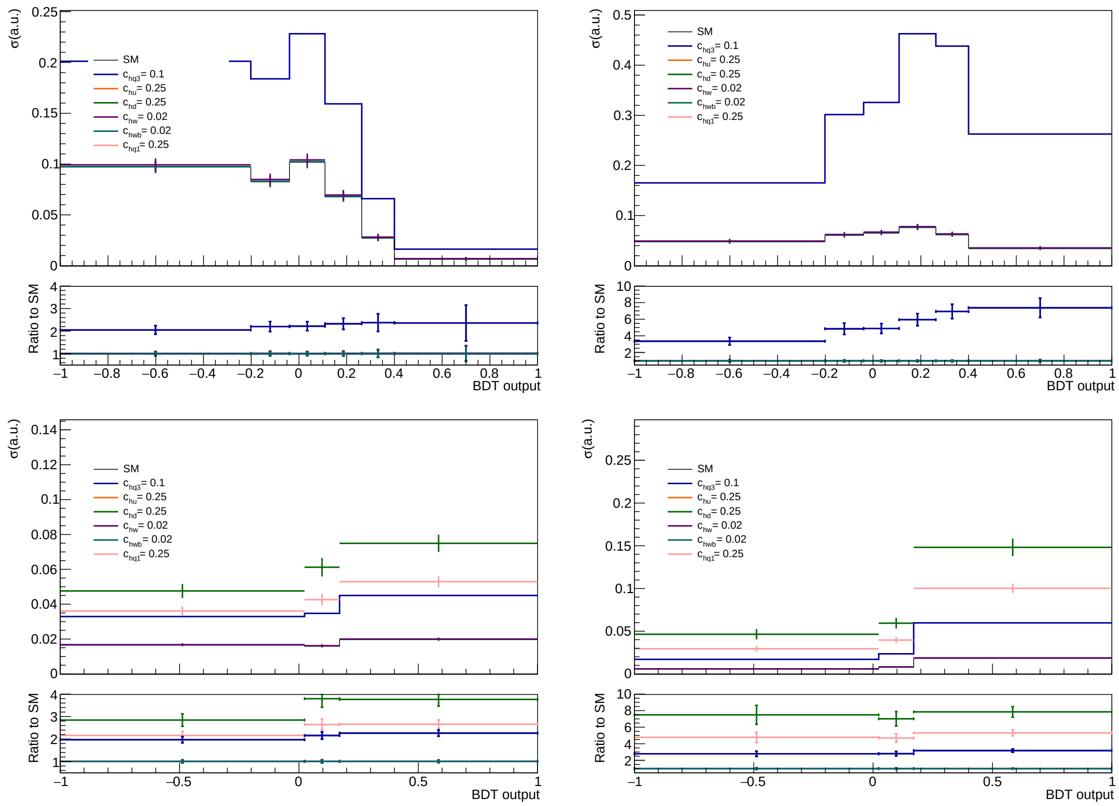


Figure 8.12: BDT output distribution for boosted category under EFT variations for WH (upper row) and ZH (lower row), in 250<$\mathrm{p_T^V}$<400 (left) and $\mathrm{p_T^V}$>400 GeV (right) GeV STXS bins.

In the next section the parametrisation for WH $\mathrm{p_T^V}$>400 GeV in BDT bins of the boosted categories will be further addressed, this section is dedicated to exploring the parametrisation for modified STXS 1.2 VH binning. In Figure 8.13 the full VH(H $\to b\bar{b}$) likelihood scans are shown for the full set of Wilson coefficients considered in this work. These plots demonstrate the sensitivity improvement due to the finer STXS categorisation in the high $\mathrm{p_T^V}$ region. The effect is significant for all operators, but especially $c_{Hq}^{(3)}$ for which the shape degeneracy is removed, and 95% C.L. intervals improved significantly. Therefore, the final parametrisation is extracted using the modified STXS 1.2 VH binning and summarised in Table 8.2. For completeness the parametrisation extracted in the default STXS stage 1.2 bins is summarised in Section 8.5.2.

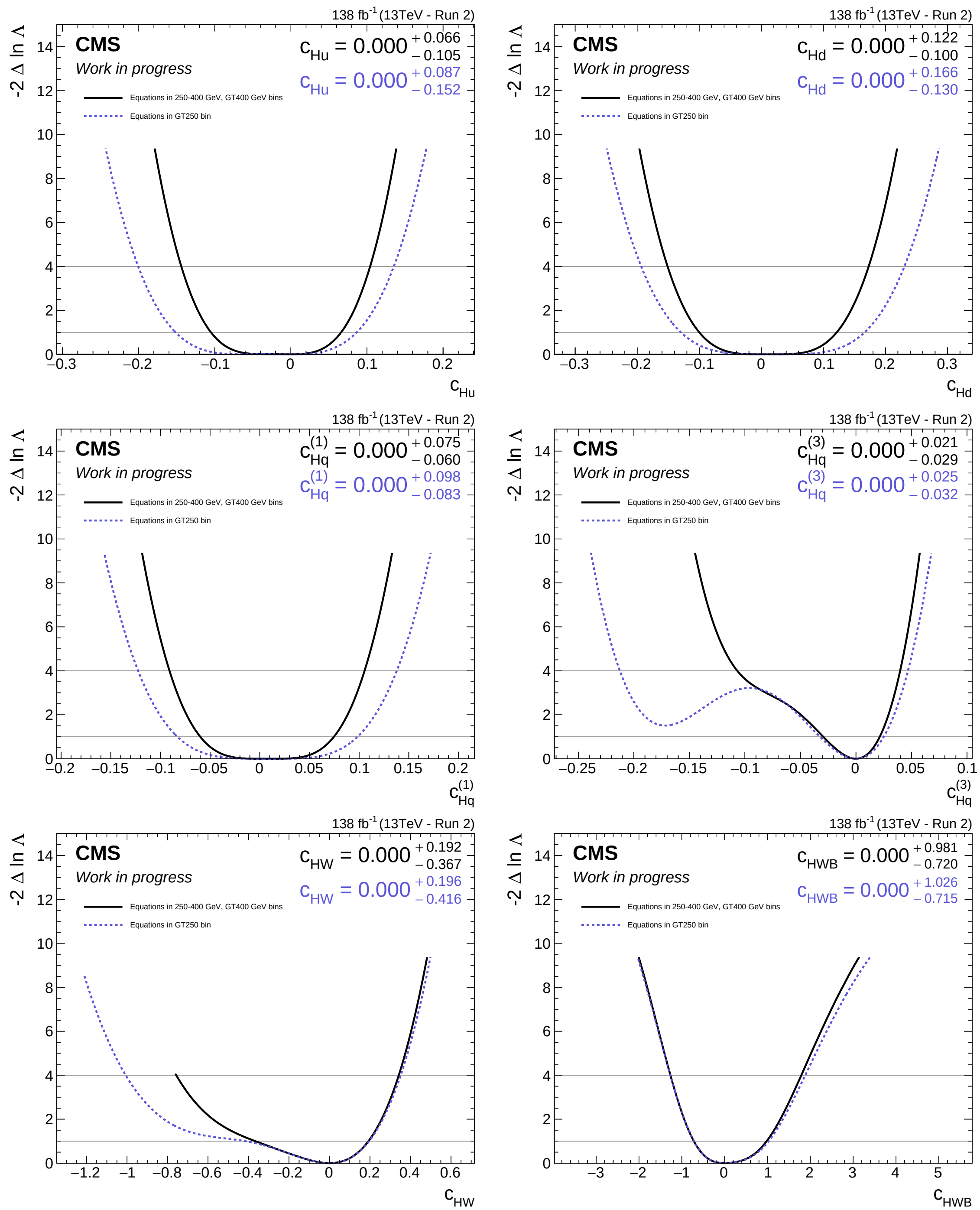


Figure 8.13: The likelihood curves for the full set of operators with the parametrisation derived for the default STXS 1.2 scheme in violet and the modified scheme in black.

Scaling functions in default STXS stage 1.2 scheme.

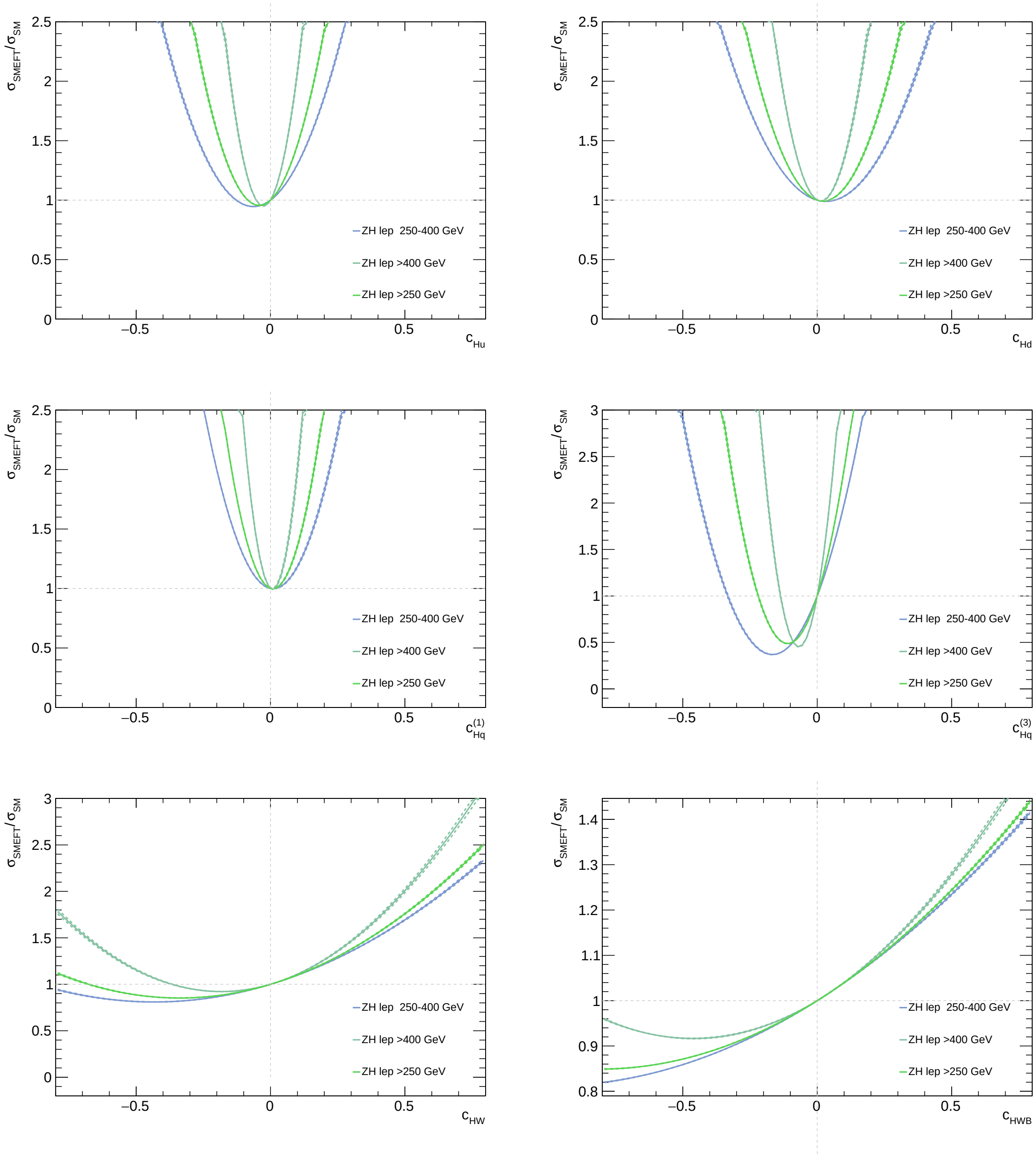


Figure 8.14: Cross-section scaling functions for the ZH STXS stage 1.2 bins

STXS bin	Scaling function
ZH lep 0-75 GeV	$1+1.09\ c_{Hq}^{(3)}+0.22\ c_{Hu}+0.64\ c_{HW}+0.30\ c_{HWB}+0.43\ {c_{Hq}^{(1)}}^2+0.43\ {c_{Hq}^{(3)}}^2+0.23\ {c_{Hu}}^2+0.20\ {c_{Hd}}^2+0.15\ {c_{HW}}^2+0.44\ c_{Hq}^{(3)}c_{HW}+0.17\ c_{Hq}^{(3)}c_{HWB}$
ZH lep 75-150 GeV	$1+1.81\ c_{Hq}^{(3)}+0.37\ c_{Hu}-0.17\ c_{Hd}+0.72\ c_{HW}+0.33\ c_{HWB}+1.11\ {c_{Hq}^{(1)}}^2+1.11\ {c_{Hq}^{(3)}}^2+0.59\ {c_{Hu}}^2+0.52\ {c_{Hd}}^2+0.21\ {c_{HW}}^2-0.14\ c_{Hq}^{(1)}c_{Hq}^{(3)}+0.80\ c_{Hq}^{(3)}c_{HW}+0.30\ c_{Hq}^{(3)}c_{HWB}+0.11\ c_{Hu}c_{HWB}+0.14\ c_{HW}c_{HWB}$
ZH lep 150-250 GeV 0J	$1+3.68\ c_{Hq}^{(3)}+0.77\ c_{Hu}-0.33\ c_{Hd}+0.81\ c_{HW}+0.36\ c_{HWB}+4.60\ {c_{Hq}^{(1)}}^2+4.60\ {c_{Hq}^{(3)}}^2+2.60\ {c_{Hu}}^2+2.03\ {c_{Hd}}^2+0.43\ {c_{HW}}^2-1.10\ c_{Hq}^{(1)}c_{Hq}^{(3)}-0.15\ c_{Hq}^{(1)}c_{HW}+1.81\ c_{Hq}^{(3)}c_{HW}+0.65\ c_{Hq}^{(3)}c_{HWB}+0.27\ c_{Hu}c_{HWB}-0.11\ c_{Hd}c_{HWB}+0.25\ c_{HW}c_{HWB}$
ZH lep 150-250 GeV >= $1J$	$1+3.35\ c_{Hq}^{(3)}+0.73\ c_{Hu}-0.28\ c_{Hd}+0.79\ c_{HW}+0.35\ c_{HWB}+4.40\ {c_{Hq}^{(1)}}^2+4.39\ {c_{Hq}^{(3)}}^2+2.50\ {c_{Hu}}^2+1.87\ {c_{Hd}}^2+0.42\ {c_{HW}}^2-1.26\ c_{Hq}^{(1)}c_{Hq}^{(3)}-0.24\ c_{Hq}^{(1)}c_{HW}+1.80\ c_{Hq}^{(3)}c_{HW}+0.62\ c_{Hq}^{(3)}c_{HWB}+0.28\ c_{Hu}c_{HWB}-0.11\ c_{Hd}c_{HWB}+0.25\ c_{HW}c_{HWB}$
ZH lep >250 GeV	$1-0.72\ c_{Hq}^{(1)}+9.36\ c_{Hq}^{(3)}+2.12\ c_{Hu}-0.76\ c_{Hd}+0.87\ c_{HW}+0.37\ c_{HWB}+42.59\ {c_{Hq}^{(1)}}^2+42.59\ {c_{Hq}^{(3)}}^2+25.13\ {c_{Hu}}^2+17.37\ {c_{Hd}}^2+1.29\ {c_{HW}}^2+0.23\ {c_{HWB}}^2-15.84\ c_{Hq}^{(1)}c_{Hq}^{(3)}-1.07\ c_{Hq}^{(1)}c_{HW}+5.81\ c_{Hq}^{(3)}c_{HW}+1.90\ c_{Hq}^{(3)}c_{HWB}+0.97\ c_{Hu}c_{HWB}-0.35\ c_{Hd}c_{HWB}+0.67\ c_{HW}c_{HWB}$

Table 8.4: Scaling functions for the ZH STXS stage 1.2 bins.

STXS bin	Scaling function
WH lep 0-75 GeV	$1+1.12\ c_{Hq}^{(3)}+0.80\ c_{HW}+0.48\ {c_{Hq}^{(3)}}^2+0.20\ {c_{HW}}^2+0.50\ c_{Hq}^{(3)}c_{HW}$
WH lep 75-150 GeV	$1+1.93\ c_{Hq}^{(3)}+0.92\ c_{HW}+1.42\ {c_{Hq}^{(3)}}^2+0.35\ {c_{HW}}^2+1.01\ c_{Hq}^{(3)}c_{HW}$
WH lep 150-250 GeV 0J	$1+4.14\ c_{Hq}^{(3)}+0.94\ c_{HW}+5.52\ {c_{Hq}^{(3)}}^2+0.72\ {c_{HW}}^2+2.18\ c_{Hq}^{(3)}c_{HW}$
WH lep 150-250 GeV >= $1J$	$1+3.71\ c_{Hq}^{(3)}+0.97\ c_{HW}+5.45\ {c_{Hq}^{(3)}}^2+0.75\ {c_{HW}}^2+2.59\ c_{Hq}^{(3)}c_{HW}$
WH lep >250 GeV	$1+10.44\ c_{Hq}^{(3)}+0.95\ c_{HW}+59.47\ {c_{Hq}^{(3)}}^2+1.85\ {c_{HW}}^2+7.00\ c_{Hq}^{(3)}c_{HW}$

Table 8.5: Scaling functions for the WH STXS stage 1.2 bins.

Extracting the parametrisation in DNN bins

As it is shown in Figure 8.12, the BDT discriminant of the boosted categories has quite significant dependence on EFT variations even after separating the WH $p_T^V > 250$ GeV STXS bin into two bins: WH $250 < p_T^V < 400$ GeV and WH $p_T^V > 400$ GeV. It was checked if the additional splitting of the WH $p_T^V > 400$ GeV phase space will produce different parametrisation equations, and result in improved sensitivity for the Wilson coefficients modifying the WH production diagrams ($c_{Hq}^{(3)}$ and c_{HW}).

The WH $p_T^V > 400$ GeV bin can be split in different ways, for example adding additional p_T^V boundary, but instead it was decided to target the BDT output directly. The boosted BDT region was split into three intervals: [-1,0), [0,0.3), [0.3,1], so that the shape information is retained, but the regions are still populated and the parametrisation uncertainties are not too large. The WH $p_T^V > 400$ events were split according to these sub-regions, and a separate parametrisation was derived. The curves for the derived equations are shown in Figure 8.15.

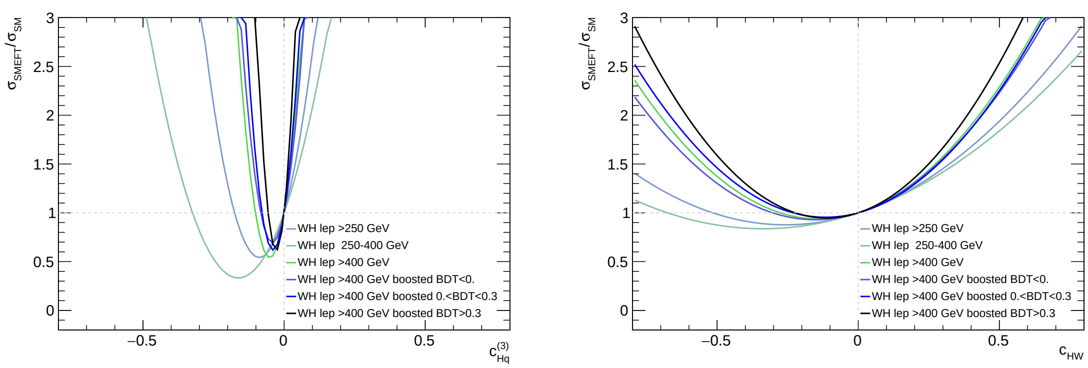


Figure 8.15: Scaling functions for the WH production in $p_T^V > 250$ GeV, $250 < p_T^V < 400$ GeV and $p_T^V > 400$ GeV bins; in addition the scaling functions are also produced in 3 BDT regions: [-1,0), [0,0.3), [0.3,1]

The BDT shape effects can be seen in the derived parametrisation, the highest BDT range has a stronger dependence on both the $c_{Hq}^{(3)}$ and c_{HW} Wilson coefficients. But at the same time the parametrisation derived differentially in BDT output is close to the inclusive equations derived for WH $p_T^V > 400$ events. The difference is also not as significant as between the equations derived inclusively in WH $p_T^V > 250$ GeV and WH $250 < p_T^V < 400$ GeV, WH $p_T^V > 400$ GeV bins.

The effect on the overall sensitivity was also checked, and is shown in Figure 8.16 for the affected operators. It can be seen that the effect is minor and does only appear at 2σ level. So it was decided that with a good approximation the parametrisation derived inclusively in the WH $p_T^V > 400$ GeV STXS bin is sufficient.

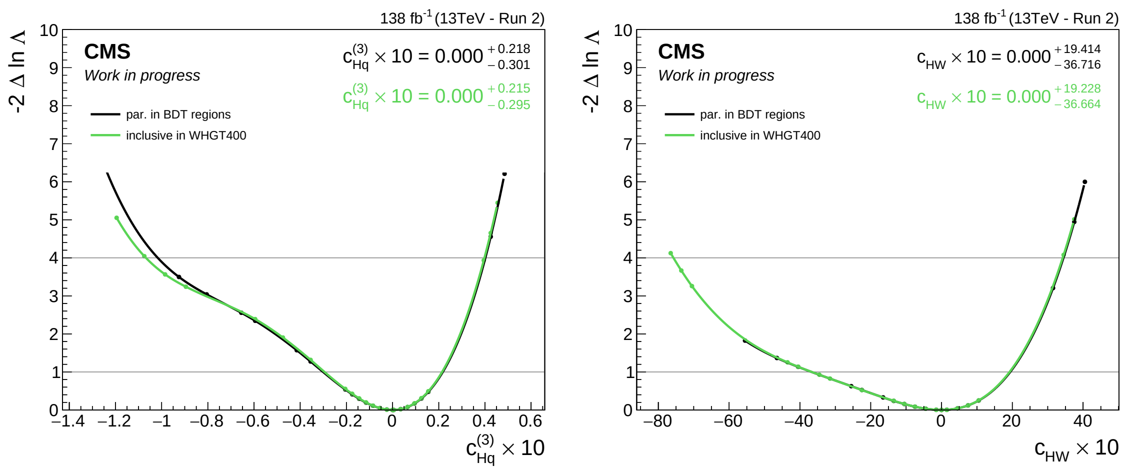


Figure 8.16: Likelihood-scans shapes obtained with the parametrisation derived inclusively for $p_T^V > 400$ GeV bin and with additional separation based on BDT output region.

8.6. Summary

This chapter has discussed the SMEFT interpretation of the full Run 2 VH(H $\to b\bar{b}$) STXS measurement with the CMS data. The SMEFT analysis uses the STXS full Run 2 VH(H $\to b\bar{b}$) measurement described in Chapter 4. The Warsaw basis implemented in the SMEFTsim model was used to include a subset of dimension-6 operators, parameterising the signal cross-section with the inclusion of linear and quadratic terms in the Wilson coefficients equations. The parametrisation is derived by employing the Madgraph reweighting techniques and including the acceptance effects. The results are summarised in Section 8.4 and compared with the previously published VH(H $\to b\bar{b}$) interpretation results by ATLAS.

A few examples of such effects are studied in Section 8.5, which will be used and further expanded for the future CMS STXS full Run 2 combination for other Higgs channels, to fully benefit from performing the interpretation within the CMS experiment. It is also important to stress that an appropriate STXS categorisation at the parametrisation step is necessary to improve the constraints on EFT operators.

CHAPTER 9

Summary

This thesis detailed the STXS measurements of the VH(H $\rightarrow b\bar{b}$) process performed with the CMS Run 2 data. The results are based on 138 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV.

The thesis describes the VH(H $\rightarrow b\bar{b}$) analysis strategy particularly focusing on the STXS measurement. The selection used in this analysis is developed to minimise the contribution of the QCD backgrounds, while the other irreducible backgrounds are modelled with the control regions method. The selected signal region events are categorised to match the STXS 1.2 stage scheme for VH production. Multivariate analysis methods are employed to improve the analysis sensitivity to the VH(H $\rightarrow b\bar{b}$). The results are extracted by performing a maximum-likelihood fit of all analysis categories. This analysis reports the combined measurement of the boosted and resolved $\mathrm{H} \rightarrow \mathrm{b\bar{b}}$ topologies, which improves the sensitivity in the BSM phase space $\mathrm{p_T^V} > 250$ GeV. The full Run 2 STXS measurement is summarised in Figure 9.1 and the inclusive signal strength is measured to be $\mu = 0.57^{+0.19}_{-0.18}$.

The second part of the thesis focuses on the EFT interpretations of the STXS VH(H $\rightarrow b\bar{b}$) measurement. The sensitivity to the BSM effects within the SMEFT model is measured for a set of dimension-6 Warsaw basis operators (Figure 9.2). The EFT effects on the analysis acceptance and the analysis observable shape are studied and included in the measurement. The work reported in this chapter highlights the importance of SMEFT interpretation studies within the experiments where additional information about the analysis final states is available. The global EFT fit are invaluable but it is difficult to incorporate various experimental effects in the global EFT parametrisation. These studies will serve as an input to the CMS Higgs SMEFT combination.

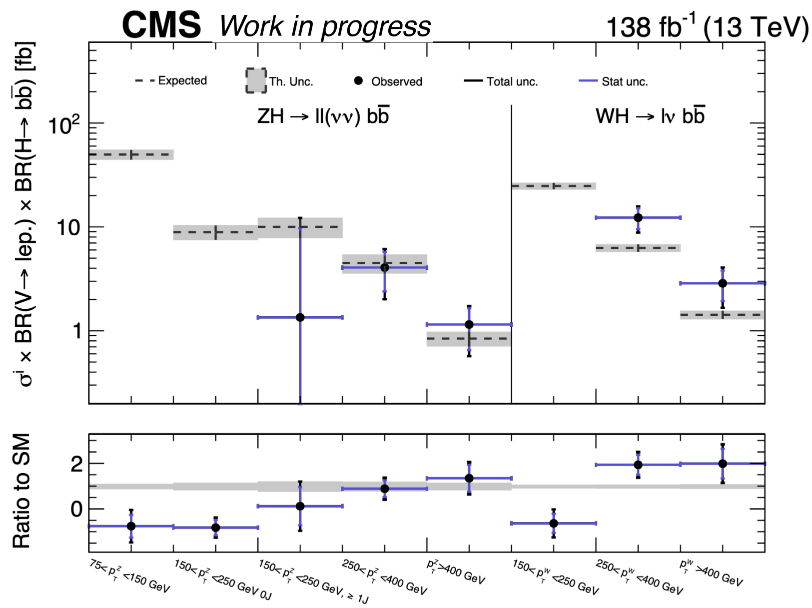


Figure 9.1: Measured values of $\sigma\mathcal{B}$ in STXS bins from the full Run 2 VH(H $\rightarrow b\bar{b}$) analysis.

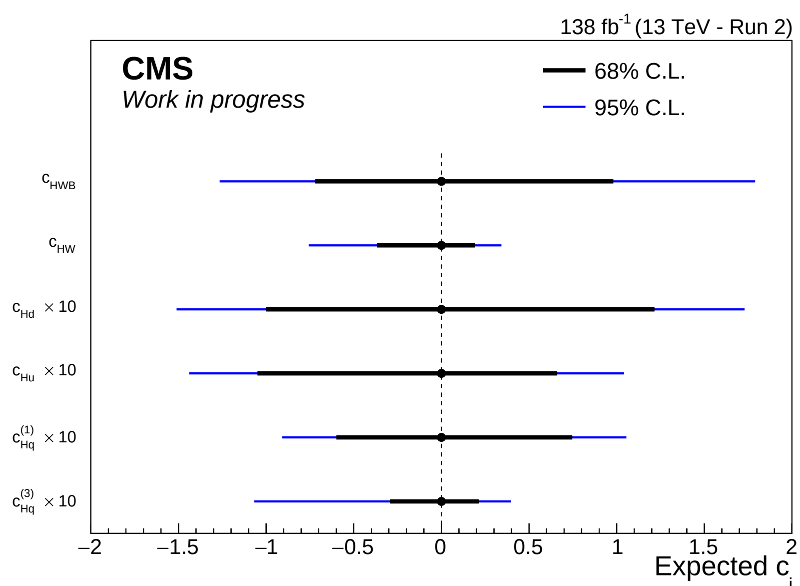


Figure 9.2: Expected 68% and 95% CL intervals obtained with linear-plus-quadratic parametrisation.

ACKNOWLEDGEMENT

First and foremost I would like to thank my supervisor Elisabetta Gallo who invited me to work at DESY in the CMS Higgs group for the opportunity to contribute to the challenging and exciting analysis summarised in this thesis. Her guidance and general advice have been very impactful. I am very grateful to my co-supervisor Rainer Mankel for his questions and comments which pointed me into the right direction many times.

This analysis summarised in this thesis is a result of collaboration with many people I've had a chance to learn from and discuss the results within CMS. Thank you Adinda for sharing your expertise, answering my numerous questions, guiding and inspiring me. I would like to also thank the members of the CMS VHbb team.

Thanks to my fellow colleagues and friends I found at DESY Sam, Antonio, Paul and Andrea for the times we spent together, the meals we shared and the long conversations we had.

Finally and most importantly, I would like to thank my family, the love and support I received from them have been invaluable. I am immensely grateful to my father who always encouraged me to be curios and in many ways shaped my interest to science, and Nadya for her support, warmth and kindness. I am very lucky and grateful to have Sasha in my life who made the toughest moments much more bearable, thank you.

Bibliography

[1] Sheldon L. Glashow. "Partial-symmetries of weak interactions." In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2.

[2] Steven Weinberg. "A Model of Leptons." In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[3] Abdus Salam. "Weak and Electromagnetic Interactions." In: *Conf. Proc. C* 680519 (1968), pp. 367–377. DOI: 10.1142/9789812795915_0034.

[4] Emmy Noether. "Invariant variational problems." German. In: *Nachr. Ges. Wiss. Göttingen, Math.-Phys. Kl.* 1918 (1918), pp. 235–257.

[5] F.J. Hasert et. all. "Observation of neutrino-like interactions without muon or electron in the Gargamelle neutrino experiment." In: *Nuclear Physics B* 73.1 (1974), pp. 1–22. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(74)90038-8.

[6] C. S. Wu et al. "Experimental Test of Parity Conservation in Beta Decay." In: *Phys. Rev.* 105 (4 Feb. 1957), pp. 1413–1415. DOI: 10.1103/PhysRev.105.1413.

[7] G. Arnison et. all. "Experimental observation of lepton pairs of invariant mass around 95 GeV/c2 at the CERN SPS collider." In: *Physics Letters B* 126.5 (1983), pp. 398–410. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(83)90188-0.

[8] F. Englert and R. Brout. "Broken symmetry and the mass of gauge vector mesons." In: *Phys. Rev. Lett.* 13 (1964), p. 321. DOI: 10.1103/PhysRevLett.13.321.

[9] P. W. Higgs. "Broken symmetries, massless particles and gauge fields." In: *Phys. Lett.* 12 (1964), p. 132. DOI: 10.1016/0031-9163(64)91136-9.

[10] P. W. Higgs. "Broken symmetries and the masses of gauge bosons." In: *Phys. Rev. Lett.* 13 (1964), p. 508. DOI: 10.1103/PhysRevLett.13.508.

[11] Makoto Kobayashi and Toshihide Maskawa. "CP-Violation in the Renormalizable Theory of Weak Interaction." In: *Progress of Theoretical Physics* 49.2 (Feb. 1973), pp. 652–657. ISSN: 0033-068X. DOI: 10.1143/PTP.49.652.

[12] CMS Collaboration. *Combined Higgs boson production and decay measurements with up to 137 fb-1 of proton-proton collision data at sqrts = 13 TeV*. Tech. rep. Geneva: CERN, 2020.

[13] ATLAS Collaboration. *Combined measurements of Higgs boson production and decay using up to 139 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment*. Tech. rep. Geneva: CERN, Nov. 2021.

[14] S. Dittmaier et al. "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables." In: (Jan. 2011). DOI: 10.5170/CERN-2011-002. arXiv: 1101.0593 [hep-ph].

[15] D. de Florian et al. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector." In: 2/2017 (Oct. 2016). DOI: 10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph].

[16] A. M. Sirunyan et al. "Observation of Higgs boson decay to bottom quarks." In: *Phys. Rev. Lett.* 121.12 (2018), p. 121801. DOI: 10.1103/PhysRevLett.121.121801. arXiv: 1808.08242 [hep-ex].

[17] Morad Aaboud et al. "Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector." In: *Phys. Lett.* B786 (2018), pp. 59–86. DOI: 10.1016/j.physletb.2018.09.013. arXiv: 1808.08238 [hep-ex].

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

[19] S. Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC." In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.021.

[20] Albert M Sirunyan et al. "Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector." In: *Phys. Lett. B* 779 (2018), pp. 283–316. DOI: 10.1016/j.physletb.2018.02.004. arXiv: 1708.00373 [hep-ex].

[21] Georges Aad et al. "Observation and measurement of Higgs boson decays to WW* with the ATLAS detector." In: *Phys. Rev. D* 92.1 (2015), p. 012006. DOI: 10.1103/PhysRevD.92.012006. arXiv: 1412.2641 [hep-ex].

[22] Albert M Sirunyan et al. "Evidence for Higgs boson decay to a pair of muons." In: *JHEP* 01 (2021), p. 148. DOI: 10.1007/JHEP01(2021)148. arXiv: 2009.04363 [hep-ex].

[23] *Direct search for the standard model Higgs boson decaying to a charm quark-antiquark pair*. Tech. rep. Geneva: CERN, 2022.

[24] "A portrait of the Higgs boson by the CMS experiment ten years after the discovery." In: *Nature* 607.7917 (2022), pp. 60–68. DOI: 10.1038/s41586-022-04892-x. arXiv: 2207.00043 [hep-ex].

[25] Enrico Fermi. "Tentativo di una teoria dell'emissione dei raggi beta." In: *Ric. Sci.* 4 (1933), pp. 491–495.

[26] Ilaria Brivio and Michael Trott. "The Standard Model as an Effective Field Theory." In: *Phys. Rept.* 793 (2019), pp. 1–98. DOI: 10.1016/j.physrep.2018.11.002. arXiv: 1706.08945 [hep-ph].

[27] I. Brivio et al. "The complete HEFT Lagrangian after the LHC Run I." In: *Eur. Phys. J. C* 76.7 (2016), p. 416. DOI: 10.1140/epjc/s10052-016-4211-9. arXiv: 1604.06801 [hep-ph].

[28] John Ellis et al. "Top, Higgs, Diboson and Electroweak Fit to the Standard Model Effective Field Theory." In: *JHEP* 04 (2021), p. 279. DOI: 10.1007/JHEP04(2021)279. arXiv: 2012.02779 [hep-ph].

[29] Lyndon Evans and Philip Bryant. "LHC Machine." In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001.

[30] G Aad and Bentvelsen et. all. "The ATLAS Experiment at the CERN Large Hadron Collider." In: *JINST* 3 (2008), S08003. 437 p. DOI: 10.1088/1748-0221/3/08/S08003. URL: https://cds.cern.ch/record/1129811.

[31] S. Chatrchyan et al. "The CMS experiment at the CERN LHC." In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[32] K. Aamodt et al. "The ALICE experiment at the CERN LHC." In: *JINST* 3 (2008), S08002. DOI: 10.1088/1748-0221/3/08/S08002.

[33] A. Augusto Alves Jr. et al. "The LHCb Detector at the LHC." In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[34] *The CMS tracker: addendum to the Technical Design Report*. Technical design report. CMS. Geneva: CERN, 2000. URL: https://cds.cern.ch/record/490194.

[35] CMS Collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker." In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 `[physics.ins-det]`.

[36] CMS Collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Tech. rep. CERN-LHCC-97-033, CMS-TDR-4. (1997).

[37] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Tech. rep. CERN-LHCC-97-031, CMS-TDR-2. (1997).

[38] CMS Collaboration. "The Performance of the CMS Muon Detector in Proton-Proton Collisions at $\sqrt{s} = 7$ TeV at the LHC." In: *JINST* 8 (2013), P11002. DOI: 10.1088/1748-0221/8/11/P11002. arXiv: 1306.6905 `[physics.ins-det]`.

[39] CMS Collaboration. "The CMS trigger system." In: *JINST* 12.01 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020. arXiv: 1609.02366 `[physics.ins-det]`.

[40] A. M. Sirunyan et al. "Particle-flow reconstruction and global event description with the CMS detector." In: *JINST* 12 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 `[physics.ins-det]`.

[41] P. Billoir and S. Qian. "Simultaneous pattern recognition and track fitting by the Kalman filtering method." In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 294.1 (1990), pp. 219–228. DOI: https://doi.org/10.1016/0168-9002(90)91835-Y. URL: https://www.sciencedirect.com/science/article/pii/016890029091835Y.

[42] Wolfgang Waltenberger, Rudolf Frühwirth, and Pascal Vanlaer. "Adaptive vertex fitting." In: *Journal of Physics G: Nuclear and Particle Physics* 34.12 (Nov. 2007), N343–

N356. DOI: 10.1088/0954-3899/34/12/n01. URL: https://doi.org/10.1088/0954-3899/34/12/n01.

[43] CMS Collaboration. "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s}$=8 TeV." In: *JINST* P06005 (2015), p. 30. DOI: 10.48550/arXiv.1502.02701.

[44] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm." In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[45] Andrew J. Larkoski et al. "Soft Drop." In: *JHEP* 05 (2014), p. 146. DOI: 10.1007/JHEP05(2014)146. arXiv: 1402.2657 [hep-ph].

[46] "Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016-2018." In: (Apr. 2020). URL: https://cds.cern.ch/record/2715872.

[47] A.M. Sirunyan. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV." In: *Journal of Instrumentation* 13.05 (May 2018), P05011–P05011. DOI: 10.1088/1748-0221/13/05/p05011. URL: https://doi.org/10.1088/1748-0221/13/05/p05011.

[48] *Machine learning-based identification of highly Lorentz-boosted hadronically decaying particles at the CMS experiment*. Tech. rep. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2683870.

[49] Albert M Sirunyan et al. "A Deep Neural Network for Simultaneous Estimation of b Jet Energy and Resolution." In: *Comput. Softw. Big Sci.* 4.1 (2020), p. 10. DOI: 10.1007/s41781-020-00041-z. arXiv: 1912.06046 [hep-ex].

[50] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II." In: *J. Phys. G* 43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph].

[51] Stefan Höche. "Introduction to parton-shower event generators." In: *Theoretical Advanced Study Institute in Elementary Particle Physics: Journeys Through the Precision*

Frontier: Amplitudes for Colliders. 2015, pp. 235–295. DOI: 10.1142/9789814678766_0005. arXiv: 1411.4085 [hep-ph].

[52] Bo Andersson. *The Lund Model.* Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 1998. DOI: 10.1017/CBO9780511524363.

[53] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "A Brief Introduction to PYTHIA 8.1." In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[54] Michelangelo L. Mangano et al. "Matching matrix elements and shower evolution for top-quark production in hadronic collisions." In: *JHEP* 01 (2007), p. 013. DOI: 10.1088/1126-6708/2007/01/013. arXiv: hep-ph/0611129.

[55] S. Agostinelli et al. "GEANT4–a simulation toolkit." In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[56] R. V. Harlander et al. "Exploiting the WH/ZH symmetry in the search for New Physics." In: *Eur. Phys. J. C* 78.9 (2018), p. 760. DOI: 10.1140/epjc/s10052-018-6234-x. arXiv: 1804.02299 [hep-ph].

[57] Paolo Nason. "A New method for combining NLO QCD with shower Monte Carlo algorithms." In: *JHEP* 11 (2004), p. 040. DOI: 10.1088/1126-6708/2004/11/040. arXiv: hep-ph/0409146.

[58] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method." In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].

[59] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX." In: *JHEP* 06 (2010), p. 043. DOI: 10.1007/JHEP06(2010)043. arXiv: 1002.2581 [hep-ph].

[60] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. "MINLO: Multi-Scale Improved NLO." In: *JHEP* 10 (2012), p. 155. DOI: 10.1007/JHEP10(2012)155. arXiv: 1206.3572 [hep-ph].

[61] Gionata Luisoni et al. "$HW^{\pm}$/HZ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO." In: *JHEP* 10 (2013), p. 083. DOI: 10.1007/JHEP10(2013)083. arXiv: 1306.2542 [hep-ph].

[62] Giancarlo Ferrera, Massimiliano Grazzini, and Francesco Tramontano. "Higher-order QCD effects for associated WH production and decay at the LHC." In: *JHEP* 04 (2014), p. 039. DOI: 10.1007/JHEP04(2014)039. arXiv: 1312.1669 [hep-ph].

[63] Giancarlo Ferrera, Massimiliano Grazzini, and Francesco Tramontano. "Associated ZH production at hadron colliders: the fully differential NNLO QCD calculation." In: *Phys. Lett. B* 740 (2015), pp. 51–55. DOI: 10.1016/j.physletb.2014.11.040. arXiv: 1407.4747 [hep-ph].

[64] Oliver Brein, Robert V. Harlander, and Tom J. E. Zirke. "vh@nnlo - Higgs Strahlung at hadron colliders." In: *Comput. Phys. Commun.* 184 (2013), pp. 998–1003. DOI: 10.1016/j.cpc.2012.11.002. arXiv: 1210.5347 [hep-ph].

[65] Ansgar Denner et al. "HAWK 2.0: A Monte Carlo program for Higgs production in vector-boson fusion and Higgs Strahlung at hadron colliders." In: *Comput. Phys. Commun.* 195 (2015), pp. 161–171. DOI: 10.1016/j.cpc.2015.04.021. arXiv: 1412.5390 [hep-ph].

[66] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations." In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[67] Rikkert Frederix and Stefano Frixione. "Merging meets matching in MC@NLO." In: *JHEP* 12 (2012), p. 061. DOI: 10.1007/JHEP12(2012)061. arXiv: 1209.6215 [hep-ph].

[68] Karl Ehatäht and Christian Veelken. "Stitching Monte Carlo samples." In: *Eur. Phys. J. C* 82.5 (2022), p. 484. DOI: 10.1140/epjc/s10052-022-10407-9. arXiv: 2106.04360 [physics.data-an].

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

[70] Vardan Khachatryan et al. "Event generator tunes obtained from underlying event and multiparton scattering measurements." In: *Eur. Phys. J. C* 76.3 (2016), p. 155. DOI: 10.1140/epjc/s10052-016-3988-x. arXiv: 1512.00815 [hep-ex].

[71] CMS Collaboration. *W and top tagging scale factors for Run 2 data.* CMS Detector Performance Note CMS-DP-2020-025. 2020. URL: https://cds.cern.ch/record/2718978.

[72] Martín Abadi et. all. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.* 2015. URL: https://www.tensorflow.org/.

[73] Nitish Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting." In: *Journal of Machine Learning Research* 15.56 (2014), pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

[74] D. P. Kingma and J. Ba. "Adam: A method for stochastic optimization." In: (2014). eprint: 1412.6980.

[75] 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 (main), 2021. DOI: 10.3929/ethz-b-000491182.

[76] Andy Buckley et al. "Rivet user manual." In: *Comput. Phys. Commun.* 184 (2013), pp. 2803–2819. DOI: 10.1016/j.cpc.2013.05.021. arXiv: 1003.0694 [hep-ph].

[77] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet User Manual." In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[78] *Evaluation of theoretical uncertainties for simplified template cross section measurements of V-associated production of the Higgs boson.* Tech. rep. Geneva: CERN, Nov. 2018. URL: https://cds.cern.ch/record/2649241.

[79] Simone Alioli et al. "Higgsstrahlung at $\mathrm{NNLL}' + \mathrm{NNLO}$ matched to parton showers in GENEVA." In: *Phys. Rev. D* 100 (9 Nov. 2019), p. 096016. DOI: 10.1103/PhysRevD.100.096016. URL: https://link.aps.org/doi/10.1103/PhysRevD.100.096016.

[80] *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. Geneva: CERN, Aug. 2011. URL: http://cds.cern.ch/record/1379837.

[81] R.J. Barlow. *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*. Manchester Physics Series, 1989.

[82] CMS Collaboration. *CMS luminosity measurements for the 2016 data taking period*. CMS Physics Analysis Summary CMS-PAS-LUM-17-001. 2017. URL: https://cds.cern.ch/record/2257069.

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

[84] CMS Collaboration. *CMS luminosity measurement for the 2018 data-taking period at* $\sqrt{s} = 13$ *TeV*. CMS Physics Analysis Summary CMS-PAS-LUM-18-002. 2019. URL: https://cds.cern.ch/record/2676164.

[85] CMS Collaboration. "Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$=13 TeV." In: *JINST* P06015 (2018), p. 30. DOI: 10.48550/arXiv.1804.04528.

[86] Serguei Chatrchyan et al. "Performance of the CMS missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s}$=13 TeV using the CMS detector." In: *JINST* 14 (2019), p. 081. DOI: 10.1088/1748-0221/14/07/P07004. arXiv: 1903.06078 [hep-ex].

[87] Hessamoddin Kaveh. "Simplified template cross-section measurement for Higgs boson decay to b-quarks in association with a vector boson with the full Run 2 CMS dataset." PhD thesis. Hamburg University, 2022. URL: https://ediss.sub.uni-hamburg.de/handle/ediss/9672.

[88] J H Friedman. "Data analysis techniques for high energy particle physics." In: (Oct. 1974), 96 p. DOI: 10.5170/CERN-1974-023.271. URL: http://cds.cern.ch/record/695770.

[89] B. Grzadkowski et al. "Dimension-Six Terms in the Standard Model Lagrangian." In: *JHEP* 10 (2010), p. 085. DOI: 10.1007/JHEP10(2010)085. arXiv: 1008.4884 [hep-ph].

[90] Ilaria Brivio. "SMEFTsim 3.0 — a practical guide." In: *JHEP* 04 (2021), p. 073. DOI: 10.1007/JHEP04(2021)073. arXiv: 2012.11343 [hep-ph].

[91] Céline Degrande et al. "Automated one-loop computations in the standard model effective field theory." In: *Phys. Rev. D* 103.9 (2021), p. 096024. DOI: 10.1103/PhysRevD.103.096024. arXiv: 2008.11743 [hep-ph].

[92] Ilaria Brivio, Tyler Corbett, and Michael Trott. "The Higgs width in the SMEFT." In: *JHEP* 10 (2019), p. 056. DOI: 10.1007/JHEP10(2019)056. arXiv: 1906.06949 [hep-ph].

[93] *EFT2Obs. A tool to automatically parametrize the effect of EFT coefficients on arbitrary observables.* URL: https://github.com/ajgilbert/EFT2Obs.

[94] *NanoAOD reweighting tool.* URL: https://github.com/MatthewDKnight/nanoAOD-tools.

[95] ATLAS Collaboration. *Combination of measurements of Higgs boson production in association with a W or Z boson in the $b\bar{b}$ decay channel with the ATLAS experiment at $\sqrt{s} = 13$ TeV.* Tech. rep. Geneva: CERN, Sept. 2021. URL: http://cds.cern.ch/record/2782535.