A.D. 1308

unipg

UNIVERSITÀ DEGLI STUDI
DI PERUGIA

Dottorato di Ricerca in Scienza e Tecnologia per la Fisica e la Geologia
XXXV Ciclo – Curriculum Fisica

Search for New Physics in Vector Boson-mediated Electroweak processes with the CMS Experiment at LHC

Settore Scientifico Disciplinare FIS/01

Candidato:
Andrea Piccinelli

Relatori:
Prof. Livio Fanò (Università di Perugia)
Prof. Alberto Orso Maria Iorio (Università di Napoli "Federico II")

Coordinatore:
Prof. Giovanni Carlotti (Università di Perugia)

A.A. 2021/2022

“Nobody ever figures out what life is all about, and it doesn’t matter.
Explore the world. Nearly everything is really interesting
if you go into it deeply enough.”

(Richard P. Feynman)

Contents

Introduction

The Standard Model (SM) of particle physics is the current framework that successfully explains the electromagnetic, weak, and strong interactions. The SM has been tested to high precision for most of the processes it describes, and all of the particles it predicted have now been observed, following the discovery of the Higgs boson at the Large Hadron Collider in 2012 and the top quark at the Tevatron in 1995. Experimental evidence, both in particle physics and in astrophysical observations, suggests that the SM is not the ultimate fundamental theory. Many physics models have been proposed to extend it to include both the current SM predictions and as-yet unexplained phenomena in a more general picture. Several of these models predict new heavy particles that couple preferentially to the massive fermions. As a consequence, the top quark and the tau lepton could have a privileged relationship with New Physics particles and may play a crucial role in their discovery.

One of the most powerful tools currently available to precisely study the SM and obtain clues to New Physics at the TeV scale is the Large Hadron Collider (LHC), thanks to its high-energy and high-luminosity collisions. The LHC is a circular accelerator designed to produce proton-proton collisions with an instantaneous luminosity of 10^{34} cm^{-2} s^{-1} and center-of-mass energy of $\sqrt{s} = 14$ TeV, and has four major experiments: ALICE, ATLAS, CMS, and LHCb.

One way to search for New Physics emerging beyond the known processes of the SM is to directly search for new resonances, such as new heavy vector bosons. The present work presents a search for a new W' boson decaying into a tb quark pair in semileptonic final states [1]. Data from proton-proton collisions with a center-of-mass energy of $\sqrt{s} = 13$ TeV and corresponding to 138 fb^{-1}, collected by the CMS experiment at the LHC from 2016 to 2018, are analyzed. The study is designed to reduce and control as much as possible the source of uncertainties and inefficiencies that can affect the sensitive power of such a study, and therefore an event reconstruction strategy relying entirely on the kinematic properties of the final state is devised, as well as a data-driven estimation of the main SM background sources. This analysis was able to improve and extend the previous searches to hypothetical masses up to 6 TeV and lower limits up to 4.3 TeV, with the largest deviation from the SM expectation resulting in a mild local excess of 2.6 standard deviations (2.6σ) in correspondence with $m_{W'} = 3.8$ TeV.

Another feasible strategy to obtain clues to New Physics is the precise measurement of SM processes, where unitarity is guaranteed only by a fine cancellation of all contributing Feynman diagrams. For this reason, this method is very sensitive to possible deviations from SM expectations due to New Physics phenomena, even if the energy scale of the latter is not accessible with current experiments. This class of processes includes in particular the production of particles that play a crucial role in electroweak spontaneous symmetry breaking (EWSSB), namely the Higgs and the W, Z vector bosons. In particular, vector boson scattering (VBS) processes play a special role, since the longitudinal polarization scattering of vector bosons converges only thanks to the existence of the Higgs boson interacting with W and Z exactly as predicted within

the SM, and is significantly modified in the presence of additional loops induced by potential New Physics. Thus, it represents an excellent probe to study the EWSSB mechanism. Moreover, anomalies in boson-boson and boson-fermion couplings induced by New Physics can be searched and interpreted using the agnostic SM Effective Field Theory (SMEFT) approach.

In this work, a search for a same-charge W boson pair through VBS processes (ssWW VBS) and related anomalous couplings is carried out, introducing an unexplored final state by exploiting the decay of one of the W bosons to a tau lepton and its subsequent hadronic decay (τ_h or hadronic tau) in the final state $\ell\tau_h\nu\nu jj$. This new signature represents an important probe in the search for BSM physics, as the τ lepton is included for the first time in the study of VBS processes. The analysis of this final state has two goals: the first measurement of the VBS process in the final state under consideration and the evaluation of the sensitivity to indirect New Physics effects with the SMEFT approach. For the first time in this type of investigation, this study introduces SMEFT operators with different dimensions, six (dim-6 operators) and eight (dim-8 operators), in the same channel. Both the effect of one operator at a time and the combined effect of two operators with same and different dimensions are estimated. A preliminary analysis of the data from proton-proton collisions with a center-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to 138 fb^{-1}, collected by the CMS experiment at the LHC from 2016 to 2018, is performed. To increase the sensitivity of the study, three deep neural networks (DNNs) are designed and implemented to discriminate between the SM background processes and the three types of signals, ssWW VBS, EFT dim-6, and EFT dim-8, and the corresponding output probability is implemented as a discriminant variable to be used in the statistical inference procedure. This study also uses a data-driven method to obtain a reliable estimate of the main background source, consisting of processes where the final-state hadronic jets are misidentified as either hadronic taus or light leptons.

The preliminary results are promising and may lead to evidence of the ssWW VBS with a hadronic tau in the final state when considering only the SM contribution. Expected limits at 1σ and 2σ confidence levels are extracted for the Wilson coefficients of the SMEFT operators under study. In particular, for the first time, the effects of the dim-6 operators in VBS processes are considered. Finally, for all possible dim-6/dim-6 and dim-6/dim-8 operator pairs, two-dimensional constraints are obtained at 1σ and 2σ confidence levels.

The present thesis is divided into eight Chapters, as follows:

- Chapter 1 provides a brief introduction to the SM, with a focus on VBS processes, and concludes with the current limitations.
- Chapter 2 introduces promising models that have been proposed to solve some of the problems of SM, with particular attention to models that predict a W' boson, and the SMEFT framework as well.
- Chapter 3 describes the LHC accelerator and the CMS detector in detail.
- Chapter 4 treats the definition, selection, and reconstruction of the physics objects employed in the analyses.
- Chapter 5 gives a brief description of the matrix-element hard process, the showering, and the hadronization phases of the event simulations for the CMS physics studies as they are implemented in MadGraph and PYTHIA.
- Chapter 6 illustrates the search for a new heavy W' boson.
- Chapter 7 presents the analysis designed to search for ssWW VBS with one hadronic tau in the final state and the possible anomalous couplings that could arise in this channel.

Part I

The Standard Model and Beyond

Chapter 1

The Standard Model of particle physics

The Standard Model (SM) is a $\mathrm{SU}(3)_C \otimes \mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$[1] quantum field theory that comprehensively describes three of the four known fundamental interactions. Its foundation began in 1961 with the work of Sheldon Glashow [2] on the unification of the electromagnetic and weak interactions, continued in 1967 by Steven Weinberg [3] and Abdus Salam [4], who later incorporated the Higgs symmetry breaking mechanism [5–7]. Glashow, Weinberg, and Salam received the 1979 Nobel Prize in Physics for their incredible efforts.

SM predictions have been confirmed by numerous measurements over 60 years, in particular by the discovery of three gauge bosons (W^+, W^- and Z) at CERN in 1983 by UA1 and UA2 experiments [8–10], the discovery of the top quark at the Tevatron in 1995 by the D0 and CDF experiments [11], and finally the discovery of the Higgs boson at CERN in 2012 by the ATLAS and CMS experiments [12, 13]. As the main model for describing the fundamental interactions in particle physics, the SM is currently widely accepted by the scientific community. However, not only does it fail to explain some phenomena, but it also has several weaknesses that suggest that it is not the most fundamental theory, making it insufficient to provide a complete picture of the known universe.

1.1 The theoretical framework

SM is a quantum field theory that, besides the usual spacetime symmetries, exhibits invariance under transformations of the three constituent gauge groups, with each fundamental interaction associated with one. A gauge symmetry group is, for construction purposes, associated with a set of massless vector fields with a dimension equal to the number of parameters that are needed to uniquely distinguish one element of the group from any other. Because they obey the Bose-Einstein statistics, the particles associated with these fields are called bosons.

The gauge symmetry groups appearing in the SM are the following:

- the circle group $\mathrm{U}(1)_Y$, whose transformation can be represented by a unitary scalar complex operator multiplied by its quantum number (the value of which depends on the particle under study), which acts as the equivalent of the electromagnetic charge, called **weak hypercharge** Y. The associated boson is called B_μ;

[1] The subscripts C and Y refer to the conserved set (charges) of the groups, color, and hypercharge, respectively, and L refers to the left-handed nature of the weak interaction, described later in this chapter.

- the $n = 2$ special unitary group SU(2)$_L$, whose three fundamental transformations can be represented by the 2×2 Pauli matrices σ_i $(i = 1,2,3)$ multiplied by a quantum number called **weak isospin** I_3; the three associated vector fields are $W_\mu^{\pm,0}$;
- the $n = 3$ special unitary group SU(3)$_C$, whose eight operations can be represented by the 3×3 Gell-Mann matrices λ_j $(j = 1,\ldots,8)$ multiplied by a charge called **color** C, which takes the value r,g,b. The eight associated vector fields are $G_\mu^{1,\ldots,8}$.

Thus there are 12 vector fields associated with 3 gauge symmetries. They are summarized by the tensor product of groups, which defines the gauge symmetry of the Standard Model:

$$\mathrm{SU}(3)_C \otimes \mathrm{SU}(2)_I \otimes \mathrm{U}(1)_Y \tag{1.1}$$

The subgroup $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ defines the symmetry of the unified electroweak interaction. The bosons W_μ^0 and B_μ are then mixed into the observed Z_μ, γ_μ bosons by the electroweak symmetry breaking mechanism, as explained in Sec. 1.4. The bosons $\gamma, W^\pm, Z$, and $G^{1,\ldots,8}$ are thus the mediators of the known forces, and they regulate the interactions among the other spin-1/2 particles of the SM, called **fermions**, as follows:

- the electromagnetic force is carried by spin-1 photons γ and acts between electrically charged particles;
- the weak interaction is governed by the three massive gauge vector bosons, $W^\pm$ and Z, discovered in 1983 by the UA1 and UA2 experiments [8–10, 14] and is responsible for phenomena such as nuclear β decays;
- the strong interaction holds the colored fermions together to form a colored singlet, like the nuclei, with the mediation of the gluons g.

The fermions are the building blocks of what is generally called matter. The SM provides 12 fermionic fields with half-integer spin, thus obeying the Fermi-Dirac statistics. They are grouped into three generations or families, each of which is a distinct weak isospin doublet of particles. Depending on their behavior with respect to the fundamental interactions, fermions are classified as lepton or quark fields. Leptons do not interact via the strong force. Quarks, on the other hand, carry electromagnetic, weak, and color charges, and thus interact with every SM gauge boson. According to the Dirac equation, which governs the motion of fermions, and the CPT symmetry conservation, for every lepton and every quark, there exists a corresponding anti-particle, i.e. its copy with the same mass but opposite quantum numbers.

Fermions exhibit chirality, defined as the eigenvector of the operator:

$$\gamma^5 = i\prod_{k=0}^{4}\gamma^i \qquad \gamma^i \equiv \text{ Dirac matrices} \tag{1.2}$$

and they can be either left-chiral or right-chiral. In the relativistic speed regime, chirality can be considered as an observable equivalent of helicity, which is the projection of the spin of a particle onto its momentum. As a consequence, the labels “chiral” are sometimes also referred to as “left-handed” or “right-handed”. It has been shown experimentally that left-handed fermions transform differently from right-handed ones under the $\mathrm{SU}(2)_L$ gauge symmetry. This leads to an asymmetry in the properties of the fermions: only the left-handed and right-handed charged leptons and the left-handed neutrinos can be observed, but no right-handed neutrinos can be detected since they cannot interact via any of the fundamental forces according to the SM. This fact is considered a postulate of the theory, as it has no explanation from the fundamental principles.

A complex scalar doublet field φ, the Higgs, named after one of the theorists who predicted its existence in 1964, is the last field needed to complete the picture of the Standard Model. Such a field plays a key role in the Higgs mechanism, responsible for making three of the boson fields acquire mass and become the known weak bosons of physics: W^+, W^-, Z, as well as the masses of the fermions. The other nine fields of the SM, the eight gluons g for the strong force and the photon γ for the electromagnetic force, are without mass. All the bosons and fermions introduced in the SM are illustrated in Fig. 1.1.

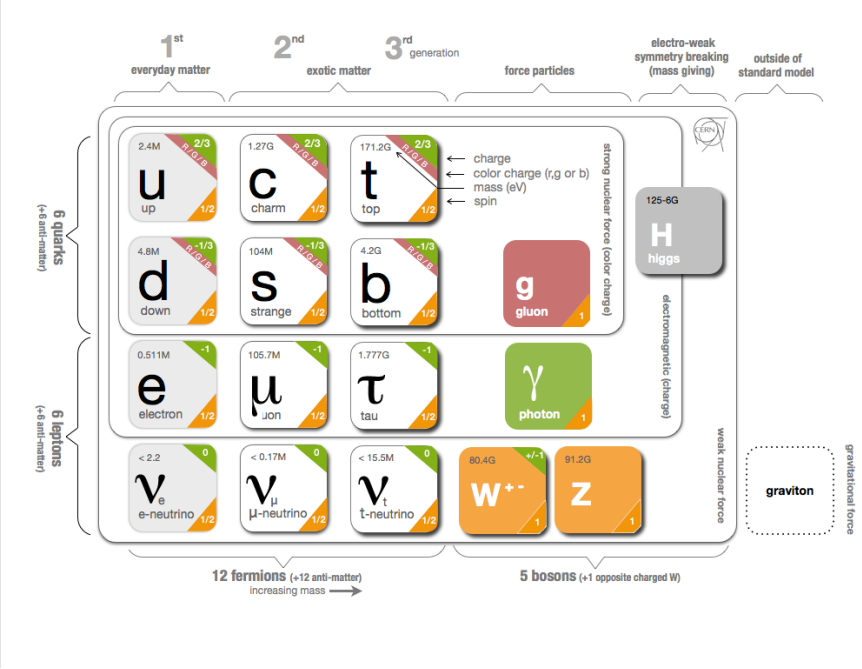


Figure 1.1: A pictorial scheme of the fields included in the Standard Model.

In the following, the three gauge theories of the SM interactions and the Higgs mechanism will be briefly described. Finally, the third generation of quarks and leptons, which are particularly relevant for the measurements presented here, will be treated separately.

1.2 Quantum Electrodynamics (QED)

Quantum electrodynamics (QED) is an Abelian gauge theory that describes the dynamics and the interactions of the fermions with the electromagnetic field. Starting from the free lagrangian density $\mathscr{L}_D$ of the Dirac field ψ with mass m, the lagrangian density for QED can be obtained:

$$\mathscr{L}_D = i\overline{\psi}\gamma^\mu \partial_\mu \psi - m\overline{\psi}\psi \tag{1.3}$$

where the first is the kinetic term and the last is the mass term, with γ^μ the Dirac matrices and ψ and $\overline{\psi}$ the 4-component spinor and its Dirac adjoint, respectively. Eq. (1.3) describes the kinematics of a free non-interacting fermion under a global gauge transformation of the symmetry group U(1). QED is associated with this symmetry group by defining $\mathrm{U}(1)_q$ transformations, where the subscript indicates the charge as the conserved quantum number. This number takes on the meaning of the charge of the particles rising as excitations of the fermionic fields.

To embed the interactions of the fermionic fields with the electromagnetic ones, Eq. (1.3) must satisfy the local gauge invariance principle. To satisfy this requirement, the partial deriva-

tive is replaced by the covariant derivative, which is introduced as follows:

$$\mathscr{D}_\mu \psi \equiv \partial_\mu + iqA_\mu \tag{1.4}$$

where Q is the charge of the fermion. At the end, the Eq. (1.3) takes the following locally invariant form:

$$\mathscr{L}_D = i\overline{\psi}\gamma^\mu \mathscr{D}_\mu \psi - m\overline{\psi}\psi \tag{1.5}$$

which can be explicitly written as follows:

$$\mathscr{L}_D = i\overline{\psi}\gamma^\mu \mathscr{D}_\mu \psi - m\overline{\psi}\psi - q\overline{\psi}\gamma^\mu A_\mu \psi = l_0 - J^\mu A_\mu \tag{1.6}$$

where the latter term includes the interaction between the Dirac particle and the electromagnetic field. The quantity J_μ is interpreted as the charge current, i.e. the probability current of the particle times its charge. Adding the following kinetic term for the field A_μ completes the lagrangian density of QED:

$$\mathscr{L}_\gamma = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{1.7}$$

where $F_{\mu\nu}$ is the field strength tensor written in terms of the 4-vector electromagnetic field A_μ:

$$F^{\mu\nu} = -F^{\nu\mu} = \partial^\mu A^\nu - \partial^\nu A^\mu \tag{1.8}$$

$\mathscr{L}_\gamma$ describes the propagation of free photons and it is invariant for local gauge transformation. Summing up Adding $\mathscr{L}_D$ and $\mathscr{D}_\gamma$, the complete QED lagrangian density is obtained:

$$\mathscr{L}_{QED} = \mathscr{L}_D + \mathscr{L}_\gamma = i\overline{\psi}\gamma^\mu \mathscr{D}_\mu \psi - m\overline{\psi}\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{1.9}$$

Since $\mathrm{U}(1)_q$ is an Abelian group, photons do not carry an electric charge. Therefore, boson self-interaction terms are not needed in QED.

1.3 Quantum Chromodynamics (QCD)

Quantum chromodynamics (QCD) is the gauge theory of strong interaction. The symmetry group at the basis of the theory is $\mathrm{SU}(3)_C$, where the subscript C stands for the color charge associated with this symmetry, which acts as the source of the chromodynamic field. Due to the non-Abelian structure of $\mathrm{SU}(3)_C$, the generators of the symmetry group do not commute, and the gauge fields, called gluons, are colored and interact with each other.

QCD is therefore invariant under local gauge transformations of the $\mathrm{SU}(3)_C$ group, and the conserved observable corresponding to this transformation is a new charge, called color. As a consequence, the strongly interacting fermionic ψ field, called a quark, takes on three possible states, which are labeled red, green, and blue. Local gauge invariance under $\mathrm{SU}(3)_C$ introduces 8 massless gauge fields, called gluons, and the covariant derivative, $\mathscr{D}_\mu$, given by:

$$\mathscr{D}_\mu = \partial_\mu + ig_s T_\alpha G_\mu^\alpha \tag{1.10}$$

where G_μ^α are the 8 gluon fields and T^α, with $\alpha = (1,\ldots,8)$, are the generators of the $\mathrm{SU}(3)_C$ group. The complete Lagrangian density for QCD is obtained by applying these considerations to the quark and gluon fields:

$$\mathscr{L}_{QCD} = \overline{\psi}\gamma^\mu \partial_\mu \psi - m\overline{\psi}\psi - ig_s\overline{\psi}\gamma^\mu \lambda_a \psi G_\mu^\alpha - \frac{1}{4}G_\alpha^{\mu\nu}G_{\mu\nu}^\alpha \tag{1.11}$$

where $G_\alpha^{\mu\nu}$ is the strong tensor field, defined as

$$G_\alpha^{\mu\nu} = \partial^\mu G_\alpha^\nu - \partial^\nu G_\alpha^\mu - g_s f_{\alpha\beta\gamma} G^{\beta,\mu} G^{\gamma,\nu} \tag{1.12}$$

Color confinement and asymptotic freedom are two relevant properties, based on experimental evidence, that significantly distinguish QCD from QED. The asymptotic freedom justifies the experimental fact that there are no colored hadrons observed, and thus states that the color degree of freedom of the quarks must be confined. The color confinement can be explained in terms of the form of the running coupling constant of the strong force:

$$\alpha_s(|q^2|) = \frac{\alpha_s(\mu^2)}{\left[1 + \alpha_s(\mu^2)\frac{33-2N_f}{12\pi}\ln\frac{|q^2|}{\mu^2}\right]} \tag{1.13}$$

where:

- q^2, the transferred 4-momentum;
- μ, scale parameter for the strength of the coupling;
- N_f, number of fermions capable of strong interactions at the scale considered.

As $|q^2|$ decreases, $\alpha_s(|q^2|)$ increases. The value of α_s is so large that no perturbative approach can be applied for $|q| \sim 200$ MeV. The calculations can be continued with the QCD lattice approach [15] in this region of the energy spectrum. For increasing values of $|q^2|$, the $\alpha_s(|q^2|)$ decreases, and a perturbative approach can be used to carry out the calculations with QCD.

1.4 Electroweak theory

The first attempt to model weak interactions was made in 1933 by the Italian physicist Enrico Fermi, who developed a theory to explain the β decay, leading to the first phenomenological description of the weak phenomena. It is determined by the current-current lagrangian density, known as the Fermi-like interaction:

$$\mathscr{L}_F = \frac{G_F}{\sqrt{2}} J^{\dagger\mu}(x) J_\mu(x) \tag{1.14}$$

where G is the Fermi constant:

$$G = 1.16638 \times 10^{-5}\ \text{GeV}^{-2} \tag{1.15}$$

and J_μ is the weak current, analogous to the electromagnetic one, and is the sum of a weak leptonic current $l_\mu(x)$ and a weak hadronic current $h_\mu(x)$. Although this interpretation successfully describes the short-range interaction approximation, it introduces divergences, manifested in the violation of the Fermi-like cross-section unitarity limit, mainly due to the dimensionality of the Fermi constant. This divergent behavior can be avoided by introducing an adimensional coupling constant and one or more intermediate vector bosons mediating the weak interaction.

The electroweak theory is thus the natural development of the Fermi theory and was proposed by S. Glashow, A. Salam, and S. Weinberg in the '60s and '70s [3, 16, 17]. It is also known as the GWS model of weak interactions, named after its creators.

1.4.1 The GWS Model

The GWS model is a quantum field theory based on the $SU(2)_L$ gauge symmetry group. The subscript L means that only the left-handed chiral components of the fields can participate in the weak interactions. For transformations of the $SU(2)_L$ group, the lagrangian density of this theory is locally invariant. The generators of this symmetry are the Pauli 2×2 matrices σ_i $(i=1,2,3)$. The particles, considered as eigenstates of the weak interaction, are arranged in six doublets of weak isospin with respect to the $SU(2)_L$ symmetry group. Calling I the weak isospin and I_3 its observed component on an isospin axis, the three doublets are for leptons:

$$I=\frac{1}{2} \qquad \begin{matrix} I_3=+1/2 \\ I_3=-1/2 \end{matrix} \qquad \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L \tag{1.16}$$

and similarly for the quarks:

$$I=\frac{1}{2} \qquad \begin{matrix} I_3=+1/2 \\ I_3=-1/2 \end{matrix} \qquad \begin{pmatrix} u \\ d' \end{pmatrix}_L \quad \begin{pmatrix} c \\ s' \end{pmatrix}_L \quad \begin{pmatrix} t \\ b' \end{pmatrix}_L \tag{1.17}$$

For the quark pairs, d',s',b' are the weak interaction eigenstates (also called flavors), which can be obtained as a linear combination of the strong interaction eigenstates (which are also mass eigenstates) d,s,b. Thus, the mixing of different flavors is given by

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = V \begin{pmatrix} d \\ s \\ b \end{pmatrix} \tag{1.18}$$

where V is the Cabibbo-Kobayashi-Maskawa (CKM) complex unitary matrix:

$$V = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \tag{1.19}$$

Although only the left-handed components can take part in the weak processes of the charged current, the right-handed components of the charged fermions are involved in the weak processes of the neutral current:

$$I=0 \qquad e_R^-, \mu_R^-, \tau_R^-, d_R, u_R, s_R, c_R, b_R, t_R \tag{1.20}$$

The requirement of local invariance under the $SU(2)_L$ group leads to the introduction of an isospin triplet of Yang-Mills fields: $W^\mu_{(i)}$, with $i=(1,2,3)$. These gauge fields are not the physical bosons. To obtain them, it is necessary to include the electromagnetic interaction in the weak sector. For this purpose, the isospin singlet B_μ, a boson field associated with a new $U(1)_Y$ local gauge symmetry, is introduced and combined with the $W^\mu_{(i)}$. The effect of this combination is to unify the electromagnetic interaction with the weak one, producing the physical boson states $W^\pm, Z, \gamma$. The quantum number associated with the extended electroweak symmetry group is the weak hypercharge Y, defined by the following relation:

$$Q = Y + \frac{I_3}{2} \tag{1.21}$$

which gives the electric charge Q, in units of e, of the I_3 member of a weak isospin multiplet, assigned weak hypercharge Y. The symmetry group of the transformation is $SU(2)_L \otimes U(1)_Y$ and leads to the four $W^\mu_{(i)}, B^\mu$ gauge fields mentioned above. The electroweak lagrangian density

can be obtained by requiring local gauge invariance under $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$, which leads to the introduction of the covariant derivative:

$$\mathscr{D}_\mu = \partial_\mu + ig\frac{\vec{\tau}}{2}W_\mu + ig'\frac{Y}{2}B_\mu \tag{1.22}$$

where g and g' are the two coupling constants for the two interactions. Neglecting the mass term for the time being and introducing the (1.22), the electroweak lagrangian density for fermions, completed with the kinematic term for the gauge, can be written as follows:

$$\mathscr{L}_{EW} = \sum_f \overline{\psi}_f \gamma^\mu \mathscr{D}_\mu \psi_f - \frac{1}{4}W^{\mu\nu}_{(i)}W^{(i)}_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} \tag{1.23}$$

where:

$$\begin{aligned} W^{\mu\nu}_{(i)} &= \partial^\mu W^\nu_{(i)} - \partial^\nu W^\mu_{(i)} \\ B^{\mu\nu} &= \partial^\mu B^\nu - \partial^\nu B^\mu \end{aligned}$$

Finally, the complete electroweak lagrangian density looks like this:

$$\begin{aligned} \mathscr{L}_{EW} = &- i\psi_L\gamma^\mu\left(\partial_\mu + ig\frac{\vec{\tau}}{2}\cdot\vec{W}_\mu + ig'\frac{Y}{2}B_\mu\right)\psi_L + \\ &- i\psi_R\gamma^\mu\left(\partial_\mu + ig'\frac{Y}{2}B_\mu\right)\psi_R + \\ &- \frac{1}{4}W^{\mu\nu}_i W^i_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} + \\ &+ \frac{1}{2}g\varepsilon_{ijk}W^{\mu\nu}_i W_{j\mu}W^{k\nu} + \frac{1}{4}g^2\varepsilon_{ijk}\varepsilon_{imn}W_{j\mu}W_{k\nu}W^\mu_m W^\nu_n \end{aligned} \tag{1.24}$$

where ψ_L and ψ_R are the left- and right-handed chiral components of the particle, respectively. The term in the last line describes the three- and four-point self-interactions of the vector bosons, which arise from the non-Abelian nature of the $\mathrm{SU}(2)_L$ group. The physical vector fields for the $W^\pm, Z$ bosons, and the photon are obtained by combining the four gauge fields:

$$\begin{aligned} W^\pm_\mu &= \frac{1}{\sqrt{2}}(W^1_\mu \mp iW^2_\mu) \\ \begin{pmatrix} Z^\mu \\ A^\mu \end{pmatrix} &= \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix}\begin{pmatrix} W^\mu \\ B^\mu \end{pmatrix} \end{aligned}$$

with:

$$\cos\theta_W = \frac{g}{\sqrt{g^2+g'^2}} \qquad \sin\theta_W = \frac{g'}{\sqrt{g^2+g'^2}} \tag{1.25}$$

The parameter θ_W W must be determined experimentally and is called the Weinberg angle or weak mixing angle. It is therefore the electromagnetic charge:

$$q = g'\cos\theta_W = g\sin\theta_W \tag{1.26}$$

In summary, the two vector bosons $W^\pm$ are electrically charged and can induce transitions between the members of the weak isospin doublets. The third gauge boson Z of the triplet is electrically neutral and is the source of the neutral currents.

1.4.2 Spontaneous symmetry breaking and Higgs mechanism

In a gauge symmetry theory, the gauge fields are introduced as massless, since the introduction of an ad hoc mass term:

$$\frac{1}{2}m^2 B_\mu B^\mu \tag{1.27}$$

in Eq. (1.24), would break the local gauge symmetries. To make the mass terms of the gauge fields invariant to these transformations, the simplest and most elegant way is to introduce spontaneous symmetry breaking (SSB) and the Higgs mechanism. The key element is the Higgs boson field, namely a $\mathrm{SU}(2)_L \otimes \mathrm{U}(1)_Y$ doublet of complex scalar fields:

$$\begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.28}$$

The complex scalar field ϕ^+ destroys positively charged particles and creates negative particles, while ϕ^0 destroys neutral particles and creates neutral antiparticles. The lagrangian density for the ϕ field is:

$$\mathscr{L}_H = (\mathscr{D}^\mu \phi)^\dagger \mathscr{D}_\mu \phi - \frac{1}{2}\mu^2 \phi^\dagger \phi - \frac{1}{4}\lambda (\phi^\dagger \phi)^2 \tag{1.29}$$

with the last two terms usually indicated synthetically as Higgs potential $V(\phi)$:

$$V(\phi) = \frac{1}{2}\mu^2 \phi^\dagger \phi + \frac{1}{4}\lambda (\phi^\dagger \phi)^2 \tag{1.30}$$

This is responsible for the symmetry breaking, and the parameter λ is assumed to be positive. The ground state ϕ_0 results from the minimization of the potential. For $\mu^2 > 0$, $V(\phi)$ assumes a unique minimum at $\phi = 0$ and consequently the ground state is symmetric under $\mathrm{SU}(2)_L$. On the other hand, for $\mu^2 < 0$ the shape of the potential is modified, as can be seen in Fig. 1.2, and V assumes a non-trivial minimum:

$$|\phi_0|^2 = -\frac{\mu^2}{2\lambda} \equiv \frac{v^2}{2} \tag{1.31}$$

The vacuum expectation value, defined as the absolute value of the field at the minimum of the potential, is thus non-zero. A reference minimum can be chosen as follows:

$$\phi = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}}\big(v + H(x)\big) \end{pmatrix} \tag{1.32}$$

where:

$$v = \sqrt{-\frac{\mu^2}{\lambda}} \tag{1.33}$$

Adding to the lagrangian density of the electroweak interactions, Eq. (1.24), the term in Eq. (1.29) and carrying out the covariant derivative, one has:

$$\mathscr{L}_H = \big(\mathscr{D}^\mu \phi\big)^\dagger \mathscr{D}_\mu \phi - \frac{1}{2}\mu^2 \phi^\dagger \phi - \frac{\lambda}{4}(\phi^\dagger \phi)^2 - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} \tag{1.34}$$

with:

$$\begin{aligned} \mathscr{D}^\mu \phi &= \left(\partial^\mu + ig\frac{\vec{\tau}}{2}\cdot\vec{W}^\mu + ig'\frac{Y}{2}B^\mu\right)\phi \\ F^{\mu\nu} &= \partial^\mu W^\nu - \partial^\nu W^\mu - g\vec{W}^\mu \times \vec{W}^\nu \\ B^{\mu\nu} &= \partial^\mu B^\nu - \partial^\nu B^\mu \end{aligned}$$

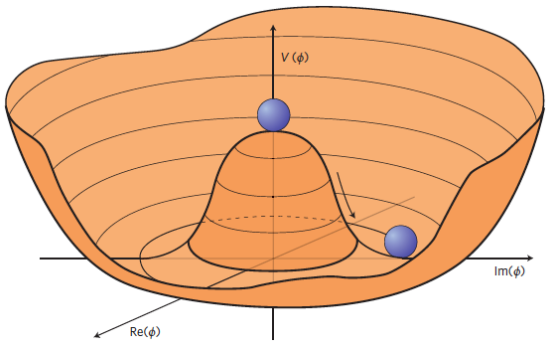


Figure 1.2: The shape of $V(\phi)$ when $\mu^2 < 0$.

Making the ϕ field fluctuate around its vacuum and finally substituting Eq. (1.32) in Eq. (1.34), one obtains:

$$\begin{aligned}\mathscr{L}_{G\phi} =& \frac{1}{2}\partial_\mu H \partial^\mu H - \mu^2 H^2 + \\ &- \frac{1}{4}\left(\partial_\mu W^1_\nu - \partial_\nu W^1_\mu\right)\left(\partial^\mu W^{1\nu} - \partial^\nu W^{1\mu}\right) + \frac{1}{8}g^2 v^2 W^1_\nu W^{1\nu} \\ &- \frac{1}{4}\left(\partial_\mu W^2_\nu - \partial_\nu W^2_\mu\right)\left(\partial^\mu W^{2\nu} - \partial^\nu W^{2\mu}\right) + \frac{1}{8}g^2 v^2 W^2_\nu W^{2\nu} \\ &- \frac{1}{4}\left(\partial_\mu Z_\nu - \partial_\nu Z_\mu\right)\left(\partial^\mu Z^\nu - \partial^\nu Z^\mu\right) + \frac{1}{8}(g^2 + g'^2)v^2 Z_\nu Z^\nu \\ &- \frac{1}{4}F^{\mu\nu}F_{\mu\nu}\end{aligned} \tag{1.35}$$

The first line of (1.35) is the lagrangian of a massive scalar field, the Higgs one, with mass equal to $\sqrt{2\mu}$. The next two lines show that the components W_1^μ and W_2^μ of the triplet $\vec{W}^\mu$ acquire mass:

$$M_1 = M_2 = \frac{1}{2}gv \equiv M_W \tag{1.36}$$

The fourth line shows that the field Z^μ acquires the following mass:

$$M_Z \equiv \frac{1}{2}v\sqrt{g^2 + g'^2} = \frac{M_W}{\cos\theta_W} \tag{1.37}$$

and the last line shows that the field A^μ has a null mass:

$$M_A = 0 \tag{1.38}$$

1.4.3 Masses of the leptons

The spontaneous breaking of the $SU(2)_L \otimes U(1)_Y$ gauge symmetry can also generate the masses of the fermions. In a theory where the symmetry is spontaneously broken, the mass term of the fermions can be obtained by coupling their fields to a scalar such as the Higgs. This is achieved by introducing a Yukawa coupling between the fermion field and the Higgs field:

$$\mathscr{L}_Y = g_f\left(\overline{\psi}_L \phi \psi_R - \overline{\psi}_R \phi^\dagger \psi_L\right) \tag{1.39}$$

where g_f is the Yukawa coupling constant. By substituting Eq. (1.32) in Eq. (1.39) one obtains:

$$\mathscr{L}_Y = \frac{g_f}{\sqrt{2}} v(l_L l_R + l_R l_L) + \frac{g_f}{\sqrt{2}} H(l_L l_R + l_R l_L) \tag{1.40}$$

The first term is a Dirac-like mass term and therefore permits to identify the constant coefficient of $(l_L l_R + l_R l_L)$ with the mass term for leptons:

$$m_f = \frac{v}{\sqrt{2}} g_f \tag{1.41}$$

This is the least satisfactory part of the model because even if this kind of Yukawa coupling solves the lepton mass problem, it does not follow a gauge principle. In fact, it is purely phenomenological and requires a specific coupling constant for each fermion-Higgs interaction according to the wide range of experimentally observed fermion masses.

1.4.4 Masses of quarks

The quarks can be given mass by the same mechanism as the leptons:

$$\mathscr{L}_Y = \frac{1}{\sqrt{2}}(v+H)\left[g^u_{ij}(u_{i,L}u_{j,R} + u_{j,R}u_{i,L}) + g^d_{ij}(d_{i,L}d_{j,R} + d_{j,R}d_{i,L} + h.c.\right]$$

with $u_i = (u,c,t)$ and $d_i = (d,s,b)$. The matrix of mass terms is not diagonal:

$$m^u_{ij} = -\frac{v}{\sqrt{2}} g^u_{ij} \qquad m^d_{ij} = -\frac{v}{\sqrt{2}} g^d_{ij} \tag{1.42}$$

and can be diagonalized in the mass eigenstate basis, leading to the introduction of the matrix:

$$V_{\alpha\beta} = \left[\mathscr{U}^u_L \mathscr{U}^{d^\dagger}_L\right]_{\alpha\beta} \tag{1.43}$$

V is the CKM matrix, and it is a 3×3 unitary matrix that quantifies the observed mismatch between the weak and mass eigenstates. It is also responsible for the transitions between quark generations through flavor-changing interactions. The definition of the CKM matrix implies that it is defined by three real parameters and an unavoidable phase factor, the latter inducing the violation of CP symmetry.

1.5 The third generation of fermions

As discussed in Sec. 1.6, the Higgs coupling to the fermions depends on the masses of the latter, and as such it is enhanced for the most massive quark and lepton, the top quark t and the tau lepton τ, respectively. The study of the properties of these two fermions allows for improving the experimental knowledge of the Higgs mechanism and the electroweak symmetry breaking. Moreover, since these two mechanisms are considered to be a preferential portal for the discovery of New Physics effects, in many Beyond the Standard Model theories the third generation of fermions presents a preferential relationship with the possible New Physics fields with respect to the lighter fermions. In the following, the main features of the t quark and the τ lepton are illustrated, with the perspective of their presence in the final states of the studies developed in Ch. 6 and Ch. 7.

1.5.1 The top quark (t)

The top quark was discovered in February 1995 by two experiments at the Tevatron proton-antiproton collider at Fermilab [11]. Its mass was measured to be $m_t = 176 \pm 13$ GeV, and the top quark finally completed the three-generation structure of the SM.

Many years after its discovery, the top quark continues to play a fundamental role in the particle physics program. Its properties have been studied extensively in high-energy hadron collisions, but some important questions remain unanswered.

The top quark is the heaviest fundamental particle known in nature, with a mass close to that of a gold atom. It is a fermion with $Q = 2/3\,e$ (where e is the elementary charge) and weak isospin $T_3 = +1/2$, as the partner of the bottom quark in the third generation quark doublet, and it carries color charge, so it participates in all fundamental interactions. It has a natural mass, which means that its Yukawa coupling to the vacuum expectation value is close to 1. For this reason, the top quark has a privileged connection with the Higgs boson and could also play a role in any model that would explain the EWSB. Due to its very short lifetime, on the order of 2×10^{-25} seconds, the top quark decays in the vast majority of cases before it has a chance to hadronize, i.e. produce hadrons. The top quark decays via the weak charged current interaction, almost exclusively into a W boson and a b quark. As a consequence, the top quark provides a unique opportunity to study the properties of a bare quark that are preserved in the decay chain and transferred to its decay products.

At hadron colliders such as the LHC at CERN, top quarks are usually produced in pairs by the strong interaction, or individually (single-top production) by the electroweak interaction. Depending on the production mode, the top quark allows different tests of the forces involved. The top quark is present in the higher-order diagrams for the calculation of the Higgs mass and, within the electroweak theory of particle interactions, it provides indirect constraints on the Higgs boson mass together with the W boson mass.

Top quark production also plays an important role in many scenarios in the search for New Physics beyond the SM, as it is present in models predicting new phenomena. Some examples include new particles decaying into top quark pairs, flavor-changing neutral currents, anomalous missing transverse energy, same-sign top pair production, and charged Higgs production.

Many extensions of the SM predict interactions with enhanced coupling to the top quarks, leading to resonances in the $t\bar{t}$ pairs. The absence of resonances in the first two generations is not significant, since the coupling could be small in this case. New particles could be spin-0 scalars or pseudoscalars, or spin-1 vector or axial vector particles, such as a Z' boson, a Kaluza$-$Klein gluon or axion, or even spin-2 particles [18]. Specific analysis tools are being developed for searches in the high-mass regions, where the top quarks are strongly boosted and the decay products tend to be collimated. Finally, new models proposed to explain the larger-than-expected forward-backward asymmetry measurement at the Tevatron require FCNC in the top sector mediated by the $t-$channel exchange of a new massive Z' boson. These mechanisms would produce top quark pairs with the same sign. However, the LHC results so far do not favor the region of parameter space consistent with the Tevatron measurement.

1.5.2 The tau lepton (τ)

In 1975 the Collaboration on the Mark I detector at the e^+e^- collider at SLAC found 24 events with a μ corresponding to an electron of opposite sign, i.e. $e^+e^- \to \mu^\pm e^\mp$ [19, 20]. These anomalous μe events presented a puzzle that could be explained by the creation and decay of a few heavy leptons, the tau leptons, namely $e^+e^- \to \tau^+\tau^-$ with the decays $\tau^+ \to e^+\nu_e\nu_\tau$ and $\tau^- \to \mu^-\nu_\mu\nu_\tau$, with a mass of $M_\tau \approx 2$ GeV and the hypothesis of the existence of a new tau neutrino ν_τ.

The tau lepton plays a crucial role in determining the nature of the Higgs boson, in the search for additional Higgs bosons, and in the search for other new particles. As the only known lepton heavy enough to decay into light hadrons, it provides a benchmark for studies of strong interactions at low energies, for the dynamics of hadronization, and for the precise determination of some SM parameters. Moreover, in the last decade, the tau lepton has been at the origin of some deviations from the leptonic universality rule in SM, which states that, for massless neutrinos, all leptons of equal electric charge have the same electroweak interactions, regardless of their flavor. Departures from this principle have been reported by the LHCb experiment at the LHC, in semileptonic decays of B mesons, although as of today there is no claimed discovery by New Physics [21].

The τ lepton has two properties that distinguish it from other charged leptons. One of them is its high mass compared to e and μ, $M_\tau = 1.77686(12)$ GeV [22]. The second property is related to the global symmetries of the SM lagrangian [23]. In the presence of Yukawa couplings, but with massless neutrinos, different leptons are characterized by a specific flavor that is conserved in all processes in the massless neutrinos SM, as well as the lepton number $L = N_e + N_\mu + N_\tau$ and the baryon number are conserved in all SM processes. The latter property gives more information about the hadron decays of the tau: although there is enough phase space to decay into baryons, such as the proton, Λ, Σ, there is not enough phase space for a baryon pair, as the B number conservation would require. This means that the tau lepton cannot decay into baryons, but only into mesons.

In SM, tau decays are driven by lepton currents:

$$\mathscr{L}_{cc} = \sum_{i=e,\mu,\tau} \frac{g_i}{2\sqrt{2}} \left[\overline{\nu}_i \gamma^\mu (1-\gamma_5) \ell_i W_\mu^\dagger \right] + h.c. \tag{1.44}$$

with $g_e = g_\mu = g_\tau = g$ are the $SU(2)_L$ couplings. This current drives the τ decays into leptons $\tau^- \to \nu_\tau \ell^- \nu_\ell$, for $\ell = e, \mu$, and those with final quarks $\tau^- \to \nu_\tau (\overline{u}d, \overline{u}s)$ and charge conjugates. Just by using this information and the corresponding one for hadrons, as shown in Fig. 1.3, it is possible to obtain good estimates for the exclusive branching ratios in leptons and the inclusive decays in hadrons. More precise measurements within the SM can correct these initial estimates and explain the experimental measurements reasonably well.

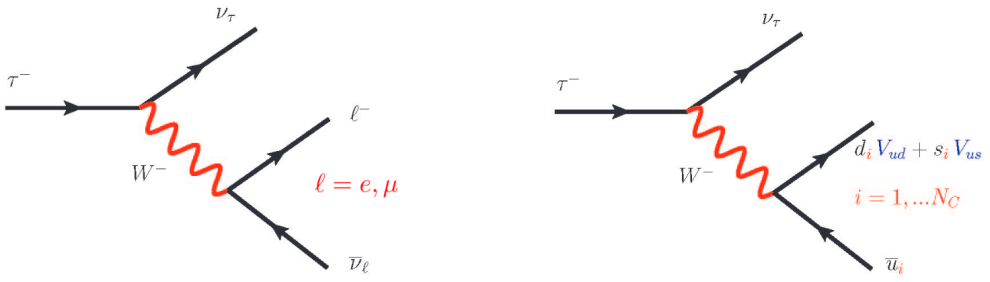


Figure 1.3: Leading tree-level Feynman diagrams describing the decays of the τ lepton in the SM, for lepton final states (left) and hadron final states (right). V_{ud} and V_{us} represents the CKM matrix elements involved in the process.

In the naturally clean lepton decays, two decay channels are possible, which allows learning more about the flavor aspects of τ leptons. The diagram on the left in Fig. 1.3 shows the leading contribution for the decays $\tau^- \to \nu_\tau \ell \overline{\nu}_\ell$, with $\ell = e, \mu$. When the dominant electroweak corrections are included, the width of the decay is [24]:

$$\Gamma(\tau^- \to \nu_\tau \ell^- \overline{\nu}_\ell) = \frac{G_F^2 M_\tau^5}{192\pi^3} f\left(\frac{M_\ell^2}{M_\tau^2}\right) r_{\text{EW}} \tag{1.45}$$

where $r_{\text{EW}} \simeq 0.9960$ represents the contribution from higher order electroweak corrections, and $f(M_\ell^2/M_\tau^2)$ the corrections due to the mass of the final lepton ℓ, which are very similar in magnitude for all leptons. As a consequence, the SM width is almost independent of the final lepton.

This scenario is a consequence of the equality of couplings in the lagrangian term reported in Eq. 1.44, known as lepton universality, which is however spoiled for massive neutrinos. The cleanest way to study this universality involves the decays of the gauge boson, i.e. $W^- \to \ell^- \overline{\nu}_\ell$. Looking at the measured ratios of widths that the SM predicts to be 1, they are [22]:

$$\mu/e = 0.996 \pm 0.008$$
$$\tau/e = 1.043 \pm 0.024$$
$$\tau/\mu = 1.070 \pm 0.026$$

These results come from the LEP data and show a tension related to the τ coupling. However, the ATLAS Collaboration has recently provided a new result [25]:

$$\tau/\mu = 0.992 \pm 0.013 \tag{1.46}$$

is in good agreement with the principle of universality, which shows that more precise experimental results are necessary for the clarification of this question.

In hadron decays, the produced hadrons decouple from the initial state and the whole process is driven by the hadronization of the charged current in low-energy QCD. The diagram on the right in Fig. 1.3 is the leading contribution to the process of production of hadrons through $\tau^- \to \nu_\tau(\overline{u}d, \overline{u}s)$. The amplitude for $\tau^- \to \nu_\tau H$, where H represents a compatible hadron final state, is given by:

$$\mathscr{A} = \frac{G_F}{\sqrt{2}} V_{\text{CKM}} \overline{\nu}_\tau \gamma_\mu (1-\gamma_5) \tau \langle H | \overline{q}_j \gamma_\mu (1-\gamma_5) q_i e^{i\mathscr{L}_{QCD}} | \Omega_H \rangle \qquad i,j = u,d,s \tag{1.47}$$

where V_{CKM} is the corresponding CKM matrix element and Ω_H is the hadronic vacuum state. The exponential of the QCD lagrangian $\mathscr{L}_{QCD}$ reminds us that the hadronization has to be done in the presence of the strong interaction. The determination of this matrix element is feasible for free quarks in the final state but fails to convey the information of a particular decay channel with mesons in the final state, namely exclusive decays, and then a hadronization process has to be carried out to determine a particular decay. The analysis of the total τ hadron width of the corresponding inclusive hadronic decay reaffirms the foundations of QCD and allows the determination of SM parameters [26]. The relevant observable is the full width normalized to one of the leptonic decays:

$$R_\tau = \frac{\Gamma(\tau^- \to \nu_\tau \text{ mesons})}{\Gamma(\tau^- \to \nu_\tau e^- \overline{\nu}_e)} \tag{1.48}$$

Experimentally, R_τ can be determined by calculating the sum of all possible tau decays into mesons, yielding [26]:

$$R_\tau = 3.6355 \pm 0.0081 \tag{1.49}$$

The phenomenological determination of the R_τ observable allows obtaining predictions for some SM observables such as the strong coupling $\alpha_S(M_\tau^2)$, the mass of the s quark, or the CKM matrix element V_{us}.

As mentioned above, the SM has a global symmetry that forbids the change of the lepton flavor or the number of leptons in any process. Since it is already known that this symmetry is violated by neutrino mixing, there is no obvious reason why processes with lepton flavor violation (LFV) in charged leptons should not occur, although it has not yet been observed and the best upper bound is $BR(\mu^+ \to e^+\gamma) < 4.2 \times 10^{-13}$ at 90% CL [27].

Some New Physics models, such as supersymmetry, Z', the little Higgs, and left-right symmetry, have been applied to the analysis of LFV τ decays, giving branching ratios in the range of the $B-$ factories, i.e. $\mathcal{O}(10^{-7} - 10^{-10})$. All of this depends on the existence of a higher energy scale, $\Lambda_{LFV} \gg \Lambda_{EW}$, where Λ_{EW} is the electroweak scale, so that the higher dimensional non-renormalizable operators violating the lepton global symmetry arise. Based on this idea, a more model-independent framework is given by the Standard Model Effective Theory (SMEFT) at the electroweak scale, which adds to the SM lagrangian new terms proportional to $D-$dimensional operators $\mathcal{O}_i^{(D)}$ that preserve the SM spectrum of particles and their fundamental symmetries, but violate global lepton flavor conservation. The lowest dimensional operators giving LFV but conserving B and L numbers have $D = 6$. Analyses in this framework have been carried out [28, 29], and it was found that the LFV τ decays constrain the dynamics more than the lepton conversion processes, although the latter can be used to distinguish the relative weights of the different contributing operators, and that the Wilson coefficient C_γ associated with the dipole operator $\mathcal{O}_\gamma = \cos\theta_W \mathcal{O}_{eB} - \sin\theta_W \mathcal{O}_{eW}$ [30] happens to be the more restricted one.

In conclusion, the physics of the τ lepton has many interesting aspects, and it offers a wide range of processes in the study of violations of SM global symmetries.

1.6 Vector Boson Scattering

Since some phenomena remain unexplained within the SM framework, extensions of the SM have been hypothesized that predict either new particles, parameter deviations from SM predictions, or both. Experimental searches and measurements, such as those performed by the ATLAS and CMS experiments, probe many different scattering processes to find any sign of physics beyond the SM (BSM). Among the various processes that have attracted the attention of the scientific community, vector boson scattering (VBS) is a prominent example. It simultaneously probes two key aspects of the SM:

- the gauge interactions, which are one of the few processes with tree-level sensitivity to the quartic gauge couplings;
- the couplings between the Higgs and the gauge bosons, which are probed at energy scales that can be meaningfully different from the Higgs mass.

In the SM, the three classes of Feynman diagrams that contribute to this process are diagrams involving triple gauge couplings, those involving quartic couplings, and those involving the Higgs boson. It can be shown that each of these classes violates unitarity when considered separately, but that unitarity is fully restored when all are considered together [31]. The typical VBS process searched for at the LHC involves two gauge bosons radiating from two scattering separate quarks, with examples of representative Feynman diagrams shown in the top row of Fig. 1.4. In the case of W and Z bosons, the VBS process is thus defined at $\mathcal{O}(\alpha^6)$ as its leading order (LO) when their decay products are included. It presents at least two quarks and up to four leptons in the final state. This results in three possible VBS experimental signatures that can be identified depending on the final state particles:

1. 2 jets and 4 leptons (fully leptonic);
2. 4 jets and 2 leptons (semi-leptonic/semi-hadronic);
3. 6 jets (fully hadronic).

This definition is gauge invariant and is valid for all non-resonant, off-shell, and interference effects. It also implies that many diagrams other than the VBS ones are included, such as

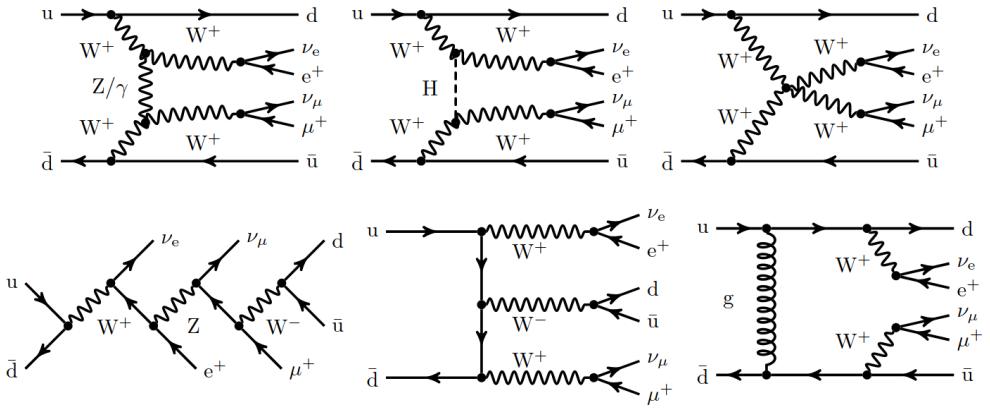


Figure 1.4: Typical Feynman diagrams for VBS contributions (top), as well as non-VBS contributions (bottom left and center) and QCD VBS background (bottom right), contributing to the process $pp \to \mu^+\nu_\mu e^+\nu_e jj$ [32].

contributions from the production of three vector bosons (tribosons), as shown in the bottom row of Fig. 1.4. To select only VBS diagrams, approximations to the full process such as the effective vector boson [33–37] or the vector boson scattering [38, 39] have to be used.

The measurement of a VBS signature necessarily implies the measurement of non-VBS contributions, since it is not possible, for example, to distinguish a VBS event from a triboson from information related only to a specific final state. Therefore, experimental analyses aiming at VBS measurements often include specific requirements to suppress unwanted contributions in the definition of fiducial regions for the measurements. Alternatively, such contributions can be subtracted from the measurements using predictions derived from dedicated Monte-Carlo simulations.

In addition to the purely electroweak contributions of order $\mathcal{O}(\alpha^6)$, VBS signatures also contain irreducible contributions of order $\mathcal{O}(\alpha_s\alpha^5)$ and $\mathcal{O}(\alpha_s^2\alpha^4)$. They are usually referred to as EW (or VBS) signal, interference, and QCD background, with the EW and QCD contributions behaving rather differently. In particular, due to their rather different QCD structures, they tend to be maximal in different phase-space regions [38]. The EW component is typically characterized by the identification or "tagging" of jet pairs with large total invariant masses m_{jj} and rapidity differences Δy_{jj}, leaving the central part of the scattering free of QCD activity, while the opposite is true for the QCD component. While this procedure may be reasonable at LO, it does not make sense at NLO and beyond. At NLO, the number of contributions increases, and some of them represent interference between two different types of amplitudes, EW or QCD, and therefore cannot be separated [32]. The introduction of multiple fiducial regions in experimental analyses thus allows the study of different physical effects that represent multiple sides of the same physical process.

1.6.1 Polarized Vector Boson Scattering

Among the various features of the SM that are particularly relevant to VBS, the different polarization states of the vector bosons involved is one of the most interesting. The study of the different polarization states makes it possible to validate the SM or to find deviations from its predictions. BSM physics may perturb the unitarization of longitudinally polarized vector bosons [40–43], or alter the relative effects of different polarization states.

Several aspects have to be considered when trying to define the production cross-section for a specific polarization state of a vector boson involved in a VBS process, as reported in [44].

The first aspect is that vector bosons are unstable particles that decay, and then any information about their polarization must be preserved by the decay processes. It can be shown that, in a VBS process, the polarization states of a pair of vector bosons interfere with each other. In principle, this endangers the definition of a polarized cross-section. However, the interference terms integrate to zero over the whole range of the decay azimuth angle θ, and this makes it possible to have a well-defined polarized cross-section.

In general, observables that show good agreement between the incoherent sum of the different polarization states and the full result can be used to extract the polarization fractions for the different states. Applying constraints to the leptons can in principle break a such agreement, but in practice, their effect is generally negligible for most observables.

When defining a polarized cross-section, there are other aspects to consider. An important one is that the polarization vectors are not Lorentz-covariant, so a particular reference frame must be chosen. Typical choices are the partonic center-of-mass frame, the laboratory frame, or the diboson center-of-mass frame. The choice of a particular frame is usually dictated by practical reasons, such as the experimental ability to reconstruct the frame. However, the studies available so far show that no frame has a particular advantage over the others [45].

In [46], the effect of a composite Higgs model in a W^+W^-jj VBS process is compared to the SM case. In this class of models, the interaction of the Higgs boson and the weak bosons is rescaled by a common factor a and can be described by the following effective lagrangian [47, 48]:

$$\mathscr{L}_{eff} = \left(\frac{m_Z^2}{2} Z^\mu Z_\mu + M_W^2 W^\mu W_\mu \right) \left(1 + 2a\frac{h}{v} + \dots \right) \tag{1.50}$$

If the scattering of the longitudinal vector bosons is unitarized, corresponding to $a = 1$, the SM case is recovered. When varying the value of a, the longitudinal-longitudinal scattering is the most affected component, as expected from its interplay with the Higgs field.

1.6.2 Vector Boson Scattering at the LHC

Taking into account the decay of heavy bosons, VBS cross-sections in proton-proton (pp) collisions at a center-of-mass energy of 13 TeV are typical of the order of femtobarns (fb). For this reason, the CMS and ATLAS experiments at the LHC are sensitive to these processes. Both Collaborations have analyzed the 13 TeV LHC dataset collected between 2016 and 2018, referred to as Run-II of the accelerator, which corresponds to about $135 - 140$ fb^{-1} per experiment, being able to claim evidence or observation for all the main VBS processes [31]. Since in this thesis the search for same-charge W-pair VBS with a final state consisting of two jets, two neutrinos, an electron/muon, and a tau lepton decaying into hadrons is presented, the focus in the following will be on the VBS studies with leptonic final states.

The analysis techniques for the reconstruction and selection of vector bosons in VBS processes depend on the final state under study. Focusing on fully leptonic channels, the heavy boson decays involved are $W^\pm \to \ell\nu$ and $Z \to \ell^+\ell^-$ (where ℓ usually denotes either an electron or a muon, these final states are the cleanest in terms of background contamination from SM processes and can cover the phase spaces of all VBS processes.The decays of the W and Z bosons into τ leptons have not been considered in previous analyses because, although they have identical branching fractions as electrons and muons, they are much more difficult to reconstruct due to the presence of the missing neutrinos in the τ secondary decays. Nevertheless, the events in which the τ decays leptonically can enter the fully leptonic channels. In the $W^\pm \to \ell \pm \nu$ case, due to the presence of neutrinos, the process cannot be fully reconstructed. As opposed to the $Z \to \ell^+\ell^-$ case, it implies that non-resonant contributions also enter the selected sample, and therefore must also be included in the simulation.

Evidence for the VBS process using semileptonic decays was also obtained with the Run II dataset, targeting events where one of the bosons decays to a boosted hadronic jet [49]. A summary of the results in the hadronic channel can be found in [50] and references therein.

1.6.3 The $W^{\pm}W^{\pm}jj \to \ell^{\pm}\nu\ell^{\pm}\nu jj$ final state

The $W^{\pm}W^{\pm}jj$ process is considered the golden channel in the study of VBS. The cross-section ratio of the EW VBS signal production to the QCD irreducible background is very large, of the order of 4-6 in typical fiducial regions, while it is typically 1 for other processes. This is due to charge conservation, which prevents gluon-initiated processes in the QCD background as opposed to $W^{\pm}Zjj, ZZjj$, or $W^{\mp}W^{\pm}jj$. In addition, the EW component of the cross-section can be further enhanced by applying special event selections. For this reason, the $W\pm W\pm jj$ channel is the most sensitive to potential New Physics effects, including those involving polarization and anomalous quartic gauge couplings (aQGCs).

Both ATLAS and CMS have observed the EW $W^{\pm}W^{\pm}jj$ production using a partial 13 TeV dataset [51, 52]. CMS has already published the same search on the full Run-II dataset, in combination with the $W^{\pm}Zjj$ final state [53].

Both ATLAS and CMS rely on Monte-Carlo simulations to evaluate the signal and various background contributions to the selected data samples. The CMS Collaboration uses the MG5_aMC generator [54] to simulate the electroweak, strong, and interference components separately at the leading order. All samples have no extra partons beyond the two quarks in the simulated process and are therefore included in the number of extra jets. The interference is estimated to be about 4% of the signal cross-section and is included in the signal yield. The CMS analysis benefits from the application of NLO QCD+EW corrections [32], which reduce the cross-sections by $10-15\%$, and similar settings are used to simulate the $W\pm Zjj$ component, which is analyzed together in a single study. Other minor backgrounds, including tribosons, processes with at least one top quark, and other processes with two bosons, are simulated with either POWHEG [55–57] or MG5_aMC, mostly with inclusive NLO QCD accuracy.

In both analyses, the event selection takes full advantage of the features of the VBS processes designed above. In particular, the Zeppenfeld variable z^*_ℓ is introduced in CMS for this purpose, as follows [58]:

$$z^*_\ell = \left| \frac{\eta(\ell) - [\eta(j_1) + \eta(j_2)/2]}{\Delta\eta_{jj}} \right| \tag{1.51}$$

The main backgrounds are estimated either by simulation or by special data-driven methods, with the statistical fits to the data performed exploiting the sensitivity to the signal of the di-jet m_{jj} and di-lepton m_{ll} invariant masses. The ATLAS and CMS data with superimposed signal and background components are shown in Fig. 1.5.

ATLAS reports a measured VBS fiducial cross-section of $\sigma_{EW} = 2.89^{+0.59}_{-0.55}$ fb in agreement with the NLO QCD estimate of the SM cross-section, with the uncertainty dominated by the statistical one, and corresponding to a rejection of the background only hypothesis with a significance of 6.5σ. Similarly, CMS reports $\sigma_{EW} = 3.98 \pm 0.45$ fb, also statistically dominated and in agreement with the NLO QCD+EW estimate in the respective fiducial region, with a significance much larger than 5σ. The total $W^{\pm}W^{\pm}jj$ cross-section including EW and strong components is also measured to be $\sigma_{tot} = 4.42 \pm 0.47$ fb. Differential cross-sections in bins of $m_{jj}, m_{\ell\ell}$ and the leading lepton p_T are obtained by simultaneously fitting the corresponding regions of phase space. All results are in agreement with SM expectations, although the experimental uncertainties are of the order of 20% due to the limited statistics. On the left side of Fig. 1.6 are the EW differential cross-sections as a function of $m_{\ell\ell}$. Limits are also placed on aQGC, with no BSM excess observed and some of the world's best limits extracted.

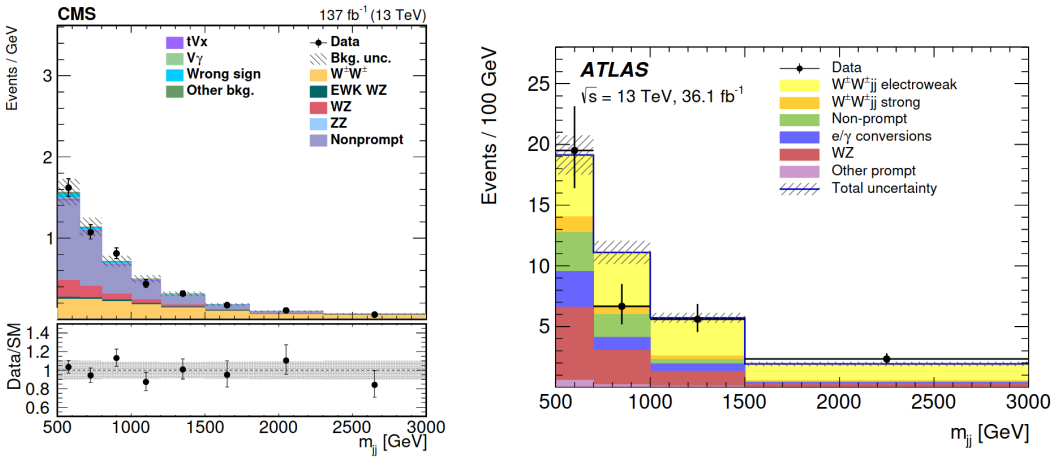


Figure 1.5: Post-fit m_{jj} distributions in CMS (left) and ATLAS (right) in the $W^{\pm}W^{\pm}jj$ analysis.

In a separate analysis, but with a very similar strategy, CMS examines the same selected dataset to measure the polarization of W bosons in $W^{\pm}W^{\pm}jj$ events [59]. Polarized signal for the three possible combinations $W_L^{\pm}W_L^{\pm}, W_L^{\pm}W_T^{\pm}, W_T^{\pm}W_T^{\pm}$, are generated using a new version of MG5_aMC [46], and the analysis uses Boosted Decision Tree (BDT) algorithms. The resulting cross-sections of $\sigma_{fid} = 1.2^{+0.6}_{-0.5}$ fb for the $W_L^{\pm}W_X^{\pm}$ process and $\sigma_{fid} < 1.17$ fb at the 95% confidence level (CL) for the $W_L^{\pm}W_L^{\pm}$ process are in agreement with the SM. There is no evidence yet even for a single-boson polarization state, the significance of the $W_L^{\pm}W_X^{\pm}$ background hypothesis rejection is only 2.3σ. On the right side of Fig. 1.6 (right) is the likelihood scan as a function of the cross-section of $W_L^{\pm}.W_L^{\pm}$ events. Results are also extracted by considering polarization eigenstates not in the standard helicity frame, but in the colliding parton frame. Results in both frames are compatible.

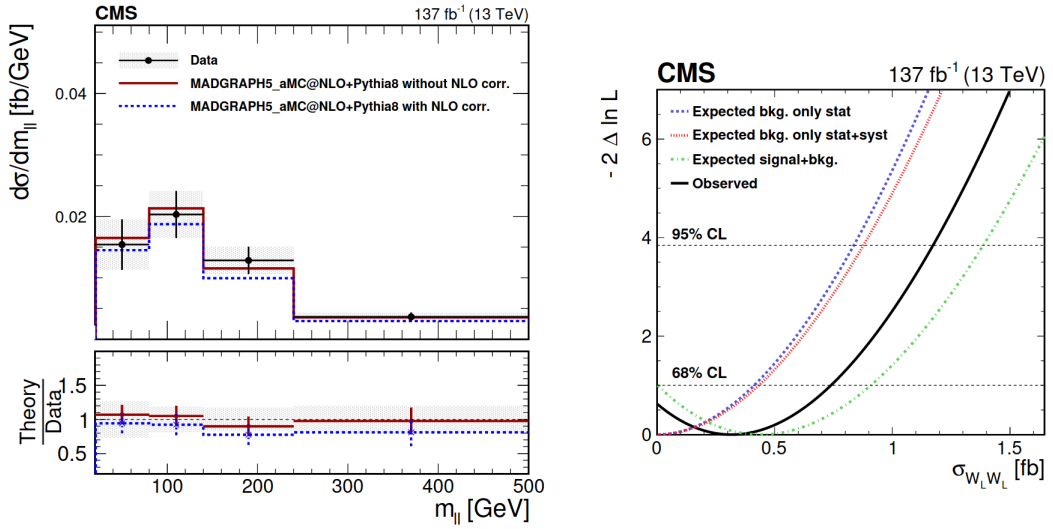


Figure 1.6: CMS $W^{\pm}W^{\pm}jj$ analysis. Left: EW differential cross-sections as a function of $m_{\ell\ell}$ measured in data, in LO and NLO-corrected simulation. Right: likelihood scan as a function of the cross-section of $W_L^{\pm}W_L^{\pm}$ events.

1.6.4 Future perspectives for VBS studies

Measurements of the VBS processes are expected to benefit from the huge amount of data that will be collected with the upcoming Run III data-taking phase of the LHC (see Sec 3.2), possibly strengthening the compelling results already obtained with the data set collected from 2016 to 2018 by the ATLAS and CMS experiments. However, a huge boost in this field is expected with the planned upgrades of the ATLAS and CMS detectors in Phase-2 in view of the high luminosity phase of the LHC collider (HiLumi-LHC, more details will be given in Sec 3.2). Including, for example, extended geometric coverage and finer detector resolution, the sensitivity to VBS processes will be significantly improved. VBS analyses, which currently have statistically dominated uncertainties, will benefit from the larger amount of data collected and the higher center-of-mass energy. In addition, detector upgrades such as the extension of tracker coverage and the addition of timing detectors will help to reject additional leptons and pileup jets, thus reducing background contamination. Finally, an increase in statistics will allow a better calibration, which is also expected to reduce experimental uncertainties.

In particular, the prospects of measuring the VBS production of $W^{\pm}W^{\pm}jj$ at the HiLumi LHC have been studied by the CMS Collaboration [60]. The results are obtained in the fully leptonic final state, using a full simulation of the upgraded Phase-2 CMS detector and considering an average number of 200 proton-proton interactions per bunch crossing. In the same study, projections of polarization studies are also performed. The longitudinal component of the same sign WW scattering is only 6%-7% of the total cross-section. Thus, even at the HL-LHC, isolation of the longitudinal component in this analysis will be severely limited by statistics. The expected significance for an integrated luminosity of 3000 fb^{-1} (i.e. the planned luminosity for the HiLumi LHC program) is expected to reach 2.7 standard deviations and exceed 3σ when CMS and ATLAS results are combined. A similar study for the same VBS channel has been performed by the ATLAS Collaboration in [61].

1.7 Unsolved issues in Standard Model

Over the past 30 years, experiments have tested the Standard Model in many ways. Every predicted particle has been found, features of the electromagnetic and strong interactions have been accurately described, and in many cases, the predictions of the SM have been verified to be within 12 orders of magnitude. Nevertheless, the SM cannot explain other phenomena and observations in nature, and therefore it cannot be considered a comprehensive and fully satisfactory theory of the known universe. Some of the most important theoretical questions and unexplained observations for which the SM cannot provide a satisfactory answer are listed below:

- **Electroweak unification**
  The GWS model cannot be considered a real unification theory, since the symmetry group is the product of two different groups, each with its own constant, g and g', which are not connected by the theory. The quotient:

$$\frac{g'}{g} = \tan\theta_W \tag{1.52}$$

  must be determined experimentally.

- **Large set of free parameters**
  In the model, many parameters are not postulated by the theory and must be obtained by measurements:
  - 3 coupling constants, g, g', g_s;
  - the fermion masses (or the corresponding Yukawa coupling to the Higgs field);
  - Higgs boson mass;
  - CKM matrix elements;
  - PMNS matrix elements.

- **Fermion masses**
  Since there is no explanation or prediction of fermion masses, a problem arises in this sector because of the differences in magnitude, up to 5 orders of magnitude, between the top quark and the electron.

- **Hierarchy problem**
  The problem, also known as the naturalness problem, is related to the radiative corrections that the Higgs mass receives through boson and fermion loops, as can be seen in Fig. 1.7. Among these corrections are self-interaction terms, such as

$$m_H^2 \approx M_{H,0}^2 + \frac{\lambda}{4\pi^2}\Lambda^2 \tag{1.53}$$

  where $M_{H,0}$ is the bare mass of the Higgs, λ is the strength of the coupling, and Λ^2 is the energy scale of the possible New Physics. Since the Higgs boson must couple to every massive particle, the one-loop corrections to the Higgs mass can be recalculated as

$$m_H^2 \approx M_{H,0}^2 + \frac{g_f}{4\pi^2}\left(\Lambda^2 + m_f^2\right) + \frac{g_s}{4\pi^2}\left(\Lambda^2 + m_S^2\right) \tag{1.54}$$

  where g_f and g_S are the coupling constants of the fermion and scalar particles to the Higgs and m_f and m_S are the associated mass terms. These corrections are quadratically divergent with the cutoff Λ. Usually, a typical cutoff is the Planck scale, where the corrections are $\sim 10^{30}\ m_H^2$, many orders of magnitude larger than the tree-level Higgs mass. A precise balance between fermions and scalars, known as fine-tuning, is required to reduce or eliminate these divergences. This fine-tuning should also respect the tight constraints on the Higgs mass, as shown in Fig. 1.8.

- **Flavor Changing Neutral Currents (FCNC)**
  Another observation not fully explained in the SM is the presence of only three families of quarks and leptons, with each family identified by the flavor quantum number. The suppression of flavor-changing neutral currents at the tree level, as expected by the Glashow, Iliopoulos, and Maiani mechanism, GIM, is also entered into the full theory by hand, without explanations from first principles.

- **Symmetries and conservations**
  Some experimental facts have to be inserted ad hoc into the theory, such as the quantization of the electric charge, the stability of the proton, and the conservation of the baryonic number, the latter not coming from a symmetry. On the other hand, there are other charges, such as the colored and electric ones, which are related to exact symmetries, since they come from the $\mathrm{U}(1)_Y$ and $\mathrm{SU}(3)_C$ groups.

- **Neutrino sector**
  SM assumes that neutrinos are massless, while observations of flavor oscillations can only

be explained if neutrinos have mass, via a mixing of the electroweak eigenstates. Indeed, the seesaw mechanism introduces neutrino masses into the SM by introducing heavy Majorana neutrinos, whose masses are inversely coupled to the light SM neutrino masses, thus motivating their small values of $O(1$ eV). Despite extensive searches, however, the experimental proof is still missing.

- **Cosmological problem**
  SM is not able to justify the actual matter-antimatter unbalance measured in the Universe by the SM CP violation in the quark sector, nor to predict and incorporate some results from astroparticle physics.

- **Gravity**
  One of the most glaring shortcomings of the Standard Model is that it lacks a description of gravity. Gravity was the first force to be fully understood at large distances, but it is likely to be the last to be fully understood at very short distances. This is because the coupling strength of gravity is very weak compared to other interactions, with a coupling constant that is 10^{34} times smaller than the electromagnetic coupling α. Gravitational effects would not be observable in particle collisions below center-of-mass energy close to the Planck scale $(10^{19}$ GeV), well beyond the energy limits of current particle colliders. Most theoretical models that could describe the Standard Model and gravity, sometimes called theories of everything or ultraviolet completion models, manifest new phenomena only above a large energy called Λ_{UV}, roughly near the Planck scale.

- **Dark Matter and Dark Energy**
  Astronomical evidence shows that the Universe consists of only the 5% of ordinary matter, while the rest does not correspond to known matter. Indeed, cosmological observations of galaxy rotation profiles provide evidence for a large amount of non-SM matter that interacts very weakly with SM particles. This matter is called dark matter and is thought to make up 24% of the universe. The remaining 71% is attributed to constant vacuum energy, called dark energy. The existence of dark energy would explain the accelerating expansion of the universe.

In the next Chapter, the possible scenarios that can be encountered when trying to extend the knowledge of particle physics beyond the Standard Model will be discussed.

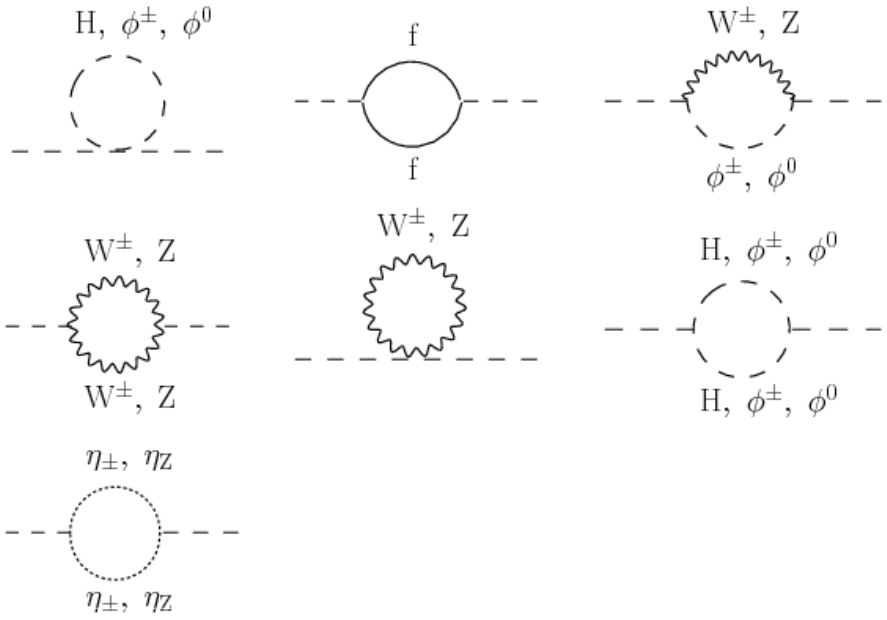


Figure 1.7: One-loop self-energy corrections to the Higgs mass.

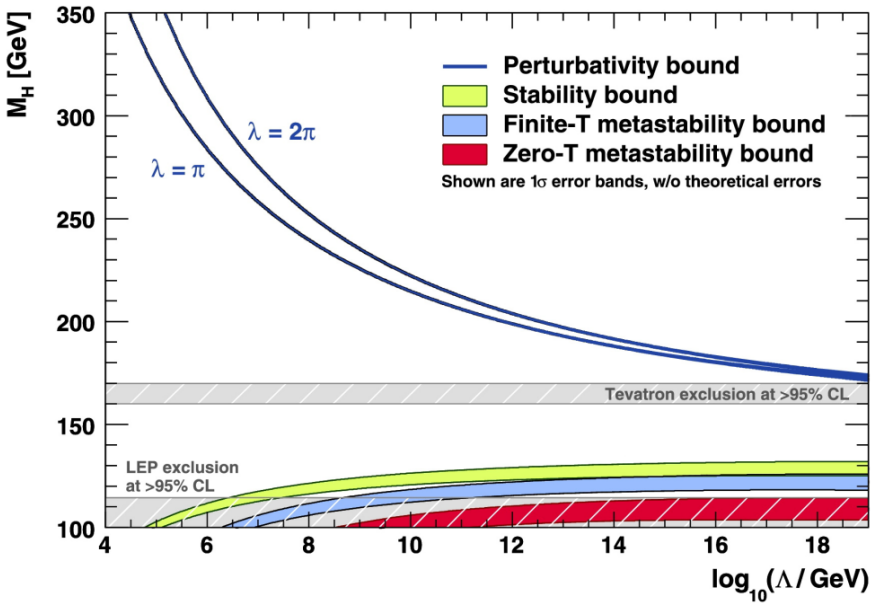


Figure 1.8: The scale Λ at which the two-loop renormalization group equations (RGEs) drive the quartic SM Higgs coupling non-perturbatively, and the scale Λ at which the RGEs create instability in the electroweak vacuum ($\lambda < 0$). The width of the bands indicates the errors introduced by the uncertainties. The perturbativity upper bound (sometimes called the "triviality" bound) is given for $\lambda = \pi$ (the blue lower bold line) and $\lambda = 2\pi$ (the blue upper bold line). Their difference indicates the size of the theoretical uncertainty in this bound. The absolute vacuum stability bound is shown by the light-shaded green band, while the less restrictive finite-temperature and zero-temperature metastability bounds are shown in the middle, blue, and dark-shaded red, respectively. The gray shaded areas indicate the LEP and Tevatron exclusion regions.

Chapter 2

Beyond the Standard Model

Much thought has been given by the particle physics community over the past decades to how the Standard Model might be modified to understand the puzzles illustrated in Sec. 1.7. The vast majority of possible answers to the above questions fall into one or more of three broad classes of proposals:

1. new large extra dimensions, meaning that there could be more than the standard three spatial dimensions, which is a likely possibility in the string theory [62], and that some of these are large enough to be seen in high-energy accelerators;

2. the known fundamental fields can be kept as they are, but new interactions are introduced, and this path leads to great ynification [63], supersymmetry [64], string theory;

3. new fundamental fields and interactions can be added, leading to theories such as compositeness, i.e. that some of the known particles could be composites of even smaller objects, such as condensed fermion-antidermal, technicolor, extended technicolor, preons [65].

Many theories that extend the SM predict additionally charged gauge bosons, often called W' bosons. In anticipation of the search presented in Ch. 6, the main theories introducing this hypothetical boson are reviewed below, as well as the theoretical properties of W' and the experimental direct searches for it performed so far.

Besides the direct search for new particles predicted by a specific BSM theory, ananother approach is to search for novel interactions for the known particles of the SM. This idea aims at a model-independent approach to the physics of non-standard interactions. A model-independent approach is useful in two ways. First, it allows one to search for New Physics without committing to a particular extension of the SM. Second, in the case that New Physics does not appear, it allows one to quantify the accuracy with which New Physics is excluded. In this context, the Effective Field Theory (EFT) framework is implemented in the study presented in Ch. 7, and it is illustrated together with its application to the VBS processes in the conclusion of this Chapter.

2.1 New resonance: W'

2.1.1 Models including W'

Extra Dimensions

The first ideas about a space-time with more than three spatial dimensions date back to the 1920s, most notably through the work of Kaluza and Klein [66]. In the context of string theory, it was again proposed to increase the number of dimensions of space as a prerequisite for describing a consistent theory of quantum gravity. The extra dimensions should be compactified at a scale close to the Planck scale, and thus not testable experimentally in the near future.

A different approach was taken by Arkani-Hamed, Dimopoulos, and Dvali (ADD for short) [67]. In 1998, they showed that the weakness of gravity could be explained by postulating two or more extra dimensions in which only gravity could propagate. The size of these extra dimensions should be between about a millimeter and $\sim 1/\text{TeV}$, leading to possible observable consequences in current and future experiments. Models with warped extra dimensions are related to four-dimensional strongly interacting theories, allowing to understand the properties of five-dimensional fields as those of four-dimensional composite states [68]. This approach has opened new avenues for addressing outstanding questions in particle physics, such as the flavor problem, grand unification, and the origin of electroweak symmetry breaking or supersymmetry breaking. Many models have observable implications that could be detected by current experiments. Stringent constraints are already imposed by cosmological and astronomical observations. Deviations from the SM due to extra dimensions are also searched for at the colliders. The ATLAS and CMS Collaborations, among others, have produced many results that find no evidence to support these models [69].

Composite Higgs Model

Within the SM, the electroweak symmetry breaking has no dynamical origin [66]. Moreover, the Higgs boson appears to be unnaturally light. In the context of weakly coupled models of electroweak symmetry breaking, multiple Higgs SU(2)$_L$ doublets, as well as additional Higgs singlets, triplets, or even more complicated multiplet structures, with or without low energy supersymmetry, can be considered [70]. One scenario that remedies these two pitfalls is to consider the Higgs boson as a bound state of new dynamics that becomes strong around the weak scale. The idea that the Higgs boson itself might be a composite bound state emerging from a new strongly coupled sector has recently regained some interest.The Higgs boson can be made significantly lighter than the other resonances of the strong sector if it appears as a pseudo-Nambu-Goldstone boson. The main prediction of the so-called composite Higgs model is the existence of new particles with masses $\tilde{1}$ TeV, which are excitations of the composite Higgs. In the most convincing scenarios, each SM particle has a partner with the same quantum number but a higher mass. For example, the γ, W, and Z bosons have heavy replicas with masses determined by the compositeness scale, expected around TeV. Indeed, composite Higgs models typically require a larger global symmetry of the underlying theory and hence additional relatively light scalar particles, additional electroweak vector bosons, e.g. an additional SU$(2)\times$U(1) gauge group, and vector-like partners of the top quark.

Left-Right symmetry

Spatial parity (P) is explicitly broken in SM by the asymmetry between left- and right-handed multiplets. Thus, the breaking of P is not on the same footing as the breaking of the gauge group

$SU(2)_L \times U(1)_Y$, which is spontaneous in the SM. This could be a starting point for extensions of the Standard Model itself [71, 72].

A recent proposal has suggested that the observed $V-A$ structure of the weak interactions may only be a low-energy phenomenon, which should disappear when we reach energies of the order of 10^3 GeV. In this contest, all interactions above these energies are assumed to be parity-conserving and, in addition, describable by a single gauge coupling constant g. Within the framework of unified gauge theories, a left-right symmetric model has been developed that not only exhibits the above properties but also provides a natural basis for the CP violating interactions observed, for example, in K decays. Another relevant feature of this model is the close connection between the strength of the CP violation and the departure from the exact left-right symmetry observed in nature. Thanks to this, we may have a model for CP violation where its magnitude isn't entirely arbitrary. The requirement that the Higgs mass term is not invariant under the latter transformation guarantees that the right-handed gauge bosons are heavier than the left-handed gauge bosons, as required by the non-observation of the right-handed current interactions at energies we have reached experimentally so far. However, this requirement does not induce a spontaneous breaking of parity, but it is possible to obtain a genuine spontaneous breaking of the left-right discrete symmetry dynamically, without requiring the Higgs mass terms from the beginning. All distinctions between left- and right-handed sectors could then arise purely as a result of the spontaneous breaking of the local symmetry.

Sequential Standard Model (SSM)

The Sequential Standard Model [73] predicts the existence of a new massive W_{SSM} boson with couplings to fermions identical to those of the SM W boson, i.e. with final states consisting of either a charged lepton and neutrino or a quark pair. It also assumes that the neutrino coming from the W_{SSM} decay is light and stable. This model provides a good benchmark, as the results can be interpreted in the context of other models of New Physics, and is useful for comparing the sensitivity of different experiments. This approach is used to interpret the results of the search reported in Ch. 6.

Top-Flavor model

Another way to extend the SM is to consider a larger gauge group, such as $SU(2)_1 \times SU(2)_2 \times U(1)_Y$ [74]. Here, in contrast to the LR model discussed earlier, the first and second generations of fermions couple to $SU(2)_1$ and the third generation couples to $SU(2)_2$. In this model, the weak interaction mediated by new bosons has a privileged coupling to the third family of fermions. The extra $SU(2)$ group could also be used to dynamically break the weak interaction via the Nambu-Jona-Lasinio mechanism [75, 76]. Predicted values for the mass of the new boson range from 800 GeV to the TeV scale [77].

2.1.2 Standard Model anomalies and W'

W' and anomalies of the meson sector

Recently, the BaBar, Belle, and LHCb Collaborations have measured the semi-leptonic decays of B mesons to D and D^* and found a discrepancy from the value predicted by SM [78]. Specifically, the anomaly is manifest in the observables:

$$R(D^*) = \frac{BR(\overline{B} \to D^* \tau^- \overline{\nu}_\tau)}{BR(\overline{B} \to D^* l^- \overline{\nu}_l)}$$

where $l = e, \mu$. The combined experimental values of these two observables are:

$$R(D^*)_{Exp} = 0.295 \pm 0.010 \pm 0.010$$

while the theory values are:

$$R(D^*)_{SM} = 0.255 \pm 0.004$$

The measurement is found to disagree with the SM predictions at about 4σ. The discrepancy suggests a possible violation of lepton flavor universality and has led to much work to explain the anomaly in the context of New Physics models. An easy way to get a New Physics contribution to this charged current is to use a W' gauge boson that couples to the second and third-generation fermions. To explain the anomaly, the W' does not need to couple to the first generation, and this fact seems to make the model more difficult to study at colliders like the LHC. If we take into account the contribution of New Physics phenomena to these processes, including the W' boson, the $R(D^*)$ observed above are modified as follows:

$$R(D^*)_{NP} = \left(1 + \frac{g'_q g'_\tau}{m^2_{W'}} \frac{\sqrt{2}}{4 G_F V_{cb}}\right) R(D^*)_{SM} \tag{2.1}$$

where $m_{W'}$ is the W' mass, g'_q, g'_τ are the W' couplings to the q flavor quark and to the τ lepton, G_F is the Fermi constant, and V_{cb} is the cb entry of the CKM matrix. Assuming that both new couplings are positive, the central values of $R(D)$ and $R(D^*)$ require the factor $g'_q g'_\tau / m^2_{W'}$ to be $0.002 \times (100\ \text{GeV}/m_{W'})^2$ and $0.001 \times (100\ \text{GeV}/m_{W'})^2$, respectively. The limits in the $m_{W'} - g'_q$ plane for several representative values of g'_τ as constrained by the experimental results are reported in [79].

W' and the anomaly of the baryon asymmetry

It is well known that the leptogenesis mechanism offers a very good possibility to explain the baryon asymmetry of the universe. The recent evidence for neutrino masses has made leptogenesis a very attractive mechanism to explain this asymmetry of the universe. According to this mechanism, the baryon asymmetry of the universe is explained by the same interactions that can explain the neutrino masses. However, it will be very difficult, if not impossible, to test experimentally the peculiarity of this mechanism in the foreseeable future, since most of the mechanism typically takes place at an energy scale that is not accessible today, or results from suppressed interactions. However, it is possible to investigate by asking whether it can be falsified. Possible discoveries at current and future colliders, in particular the discovery of right-handed gauge interactions, would indeed rule out at least the leptogenesis mechanisms based on right-handed neutrino decay. For high-mass right-handed neutrinos and W_R[1], around 10^{10} GeV or higher, the W_R can have suppressive effects on leptogenesis through dilution and scattering of the right-handed neutrino (the latter reported as N_R in Fig. 2.1).

2.1.3 Interaction Lagrangian

The W' boson is a massive hypothetical particle with spin 1 and electric charge ± 1, which is a color singlet [66]. The lowest dimension effective lagrangian of the interactions of W' bosons

[1] W_R is the name for the right-handed form of the hypothetical W' boson.

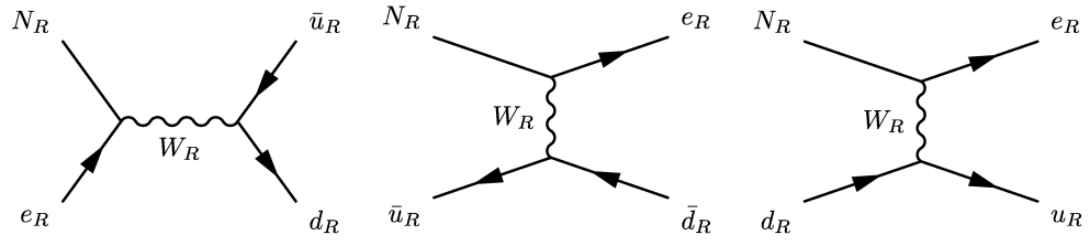


Figure 2.1: Scatterings involving the W_R. N_R is the hypothetical right-handed neutrino, L a not-SM lepton.

with fermions can be written in the most general form as:

$$\mathscr{L} = \frac{V'_{q_i q_j}}{2\sqrt{2}} g' \bar{q}_i \gamma_\mu [\, \alpha^R_{q_i q_j} (\, 1+\gamma^5 \,) + \alpha^L_{q_i q_j} (\, 1-\gamma^5 \,) \,] W' q_j + \\ \frac{g'}{2\sqrt{2}} \bar{\nu}_\ell \gamma_\mu [\, \alpha^R_{\bar{\nu}_\ell \ell} (\, 1+\gamma^5 \,) + \alpha^L_{\bar{\nu}_\ell \ell} (\, 1-\gamma^5 \,) \,] W' \ell + h.c. \tag{2.2}$$

where $\alpha^R_{q_i q_j}$ and $\alpha^L_{q_i q_j}$ are the left and right couplings of W' to quarks, $\alpha^R_{\bar{\nu}_\ell \ell}$ and $\alpha^L_{\bar{\nu}_\ell \ell}$ are the left and right couplings of W' to leptons, g' is the coupling constant of the interaction and $V'_{q_i q_j}$ is the element of a quark mixing matrix defined analogously to the SM CKM matrix. The parameters α^R and α^L regulate the chirality fractions of LH and RH W' bosons, respectively, and depend on the specific gauge model represented in the New Physics scenario. All these parameters are free, and different models have different assumptions and corresponding implications. If the model also includes a Z' boson, then the ratio of W' to Z' is another free parameter.

A tree-level mass mixing may be induced between the electrically-charged gauge bosons. Upon diagonalization of their mass matrix, the W-to-Z mass ratio and the couplings of the observed W boson are shifted from the SM values. The measurements of these quantities available up to now imply that the mixing angle θ^+ between the gauge eigenstates must be smaller than about 10^{-2} [22]. In certain theories the mixing is negligible, for example due to a new parity [80], even when the W' mass is close to the electroweak scale. It is important, by the way, to take into account that the SU(2) gauge invariance suppresses the kinetic mixing between the W and W' bosons, in contrast to the case of a Z' boson. In models based on the left-right symmetry, the SM fermions that couple to the W boson transform as doublets under SU(2)$_L$ while the other fermions transform as doublets under SU(2)$_R$. Consequently, the W' boson couples primarily to right-handed fermions, and its coupling to left-handed fermions arises due to the θ^+ mixing, so that C^L_q is proportional to the CKM matrix and its elements are much smaller than the diagonal elements of C^R_q. In the "alternate left-right" model [81], all the couplings shown in Eq. 2.2 vanish, but there are some new fermions such that the W' boson couples to pairs involving a SM fermion and a new fermion. In the "un-unified SM" [82], the left-handed quarks are doublets under one SU(2) group, and the left-handed leptons are doublets under a different SU(2), leading to a mostly leptophobic W' boson. Fermions of different generations may also transform as doublets under different SU(2) gauge groups [83]. In particular, the couplings to third generation quarks may be enhanced [74]. It is also possible that the W' couplings to SM fermions are highly suppressed. For example, if the quarks and leptons are singlets under one SU(2) [84], then the couplings are proportional to the tiny mixing angle θ^+. Similar suppressions may arise if some vector-like fermions mix with the SM fermions [85].

In this work, more precisely in Ch. 6, the assumption considered is that of the SSM, i.e. g' is identified with the SM weak interaction coupling $g_w = e/\sin\theta_W$, and the interaction of the W' boson with quarks has been studied, with the quark mixing matrix $V'_{q_i q_j}$ identified with the

SM CKM matrix. With reference to Eq. 2.2 the following cases are considered: $\alpha^R_{q_iq_j} = 1$ and $\alpha^L_{q_iq_j} = 0$ for the pure right-handed (RH) scenario, $\alpha^R_{q_iq_j} = 0$ and $\alpha^L_{q_iq_j} = 1$ for the pure left-handed (LH) scenario, and $\alpha^R_{q_iq_j} = 1/\sqrt{2}$ and $\alpha^L_{q_iq_j} = 1/\sqrt{2}$ for the equally mixed coupling scenario.

The pure RH and LH cases are usually considered to set limits on the mass hypotheses in direct searches. The main difference between the RH and LH cases is that in the latter case, the SM perturbation must be taken into account. The processes that could contribute to the W' boson production in the $s-$channel are the SM single top quark production modes. For large values of the W' boson mass, the only process that has a significant contribution is the SM single top quark $s-$channel, while the $t-$channel and tW associated production can be neglected [86].

For the leading $s-$channel subprocess $u\bar{d} \to t\bar{b}$, the cross-section can be written as:

$$\begin{aligned}\hat{\sigma}(\hat{s}) = &\frac{\pi\alpha^2_{\mathrm{W}}}{6} V^2_{\mathrm{tb}} V^2_{\mathrm{ud}} \frac{(\hat{s}-m_t^2)^2(2\hat{s}+m_t^2)}{\hat{s}^2}\left[\frac{1}{(\hat{s}-m_W^2)^2+\Gamma^2_W m^2_{\mathrm{W}}} + \right.\\ &+ 2\alpha^L_{\mathrm{ud}}\alpha^L_{\mathrm{tb}} \frac{(\hat{s}-m^2_{\mathrm{W}})(\hat{s}-m_{W'}{}^2)+\gamma^2_W\Gamma^2_{W'}}{((\hat{s}-m^2_{\mathrm{W}})^2+\Gamma^2_W m^2_W)((\hat{s}-m_{W'}{}^2)^2+\Gamma^2_{W'} m_{W'}{}^2)} + \\ &\left. + \frac{(\alpha^L_{\mathrm{ud}}\alpha^L_{\mathrm{tb}})^2+(\alpha^R_{\mathrm{ud}}\alpha^R_{\mathrm{tb}})^2+(\alpha^L_{\mathrm{ud}}\alpha^R_{\mathrm{tb}})^2+(\alpha^R_{\mathrm{ud}}\alpha^L_{\mathrm{tb}})^2}{(\hat{s}-m_{W'}{}^2)^2+\Gamma^2_{W'} m_{W'}{}^2}\right] \end{aligned} \tag{2.3}$$

where $\hat{s} = x_{\mathrm{u}} x_{\bar{\mathrm{d}}} s$, where s is the energy scale of the process. The first term of Eq. 2.3 is the pure SM single top quark production in the $s-$channel, the second is the interference term and the last is the pure BSM production.

From now on, $\alpha^R_{q_iq_j}$ and $\alpha^L_{q_iq_j}$ will be indicated as α_R and α_L, respectively, as they are the right and left coupling of interest in this work.

2.1.4 Searches at colliders

At hadron colliders, a W' boson can be produced as a resonance from a quark-antiquark pair or electroweak boson processes. For values of the total decay width much smaller than its mass ($\Gamma_{W'}/m_{W'} \sim 7\%$), the $s-$channel process $pp \to f\bar{f}'X$, where X is any final state, can be written as the branching fraction $\mathrm{BR}(W' \to f\bar{f})$ times the production cross-section $\sigma(pp \to W'X)$, given by:

$$\sigma(pp \to W'X) = \frac{\pi}{6s}\sum_{i,j}\left[\left(C^L_{q_{ij}}\right)^2 + \left(C^R_{q_{ij}}\right)^2\right]\omega_{ij}\left(\frac{M^2_{W'}}{s}, m_{W'}\right). \tag{2.4}$$

The functions ω_{ij} contain the information about the proton structure and are given by the leading order in α_S:

$$\omega_{ij}(z,\mu) = \int_1^z \frac{dx}{x}\left[u_i(x,\mu)\bar{d}_j\left(\frac{z}{x},\mu\right) + \bar{u}_i(x,\mu)d_j\left(\frac{z}{x},\mu\right)\right], \tag{2.5}$$

where $u_i(x,\mu)$ and $d_j(\frac{z}{x},\mu)$ are the parton distribution functions inside the proton at the factorization scale μ and the parton momentum fractions x and z/x for the up- and down-type quarks of the $i-$th generation, respectively. QCD corrections to the W' production are large since they include quark-gluon initial states, but preserve the above factorization of the couplings at the next-to-leading order [87].

The easiest searches to perform in a hadronic environment are those that require a high-momentum lepton and a large amount of missing transverse momentum due to the escaping neutrino, representing the decay chain $W' \to \ell\nu$. The invariant mass of the lepton-neutrino system is reconstructed in the transverse plane, yielding a Jacobian peak ending at $m_{W'}$ [88]. Since the branching fractions in $\mu\nu$ and $e\nu$ are not guaranteed to be equal, the results are given separately for these two decay modes. The current results assume negligible interference

between the LH W' and EW W bosons, while for the RH coupling an RH neutrino is assumed to be much lighter than the W' boson. Usually, the SSM is used as a benchmark model, and lepton universality is also assumed to be preserved. Both the ATLAS and CMS Collaborations set limits on the W' production cross-section times branching fraction in $\mu\nu$, $e\nu$, and in the combined $\ell\nu$ final states [88, 89]. Similarly, the CMS Collaboration has searched for new high-mass resonances in proton-proton collisions with final states involving an electron or muon and missing transverse momentum. The analysis uses proton-proton collision data collected in 2016 with the CMS detector at the LHC at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 35.9 fb^{-1} [90]. The $\tau\nu$ channel is treated separately because of the different decay channels of the τ leptons, which could have both leptonic and hadronic final states, and because some models predict a privileged coupling of the W' boson to the third families of quarks and leptons. The ATLAS Collaboration searched for excesses above the SM expectation in the $W' \to \tau\nu$ channel in pp collision data corresponding to 36.1 fb^{-1}at a center-of-mass energy of $\sqrt{s} = 13$ TeV. Only τ lepton decays with hadrons in the final state are considered. Heavy W' bosons with masses below 3.7 TeV in the SSM and masses below 2.2-3.8 TeV depending on the coupling in the non-universal $G(221)$ model are excluded at the 95% confidence level. Similarly, the CMS Collaboration excluded $0.4 < m_{W'} < 4.0$ TeV at 95% CL in the SSM hypothesis [91].

Although more challenging at hadron colliders, both ATLAS and the CMS Collaboration have performed searches for the W' boson in hadronic final states [92–94]. The large amount of data collected at the LHC has also allowed the testing of rarer decay channels, such as $W' \to WH$, by both the ATLAS [95, 96] and CMS Collaborations [97, 98].

In the following, more details are given for the W' searches in the decay channel that were investigated in the study performed in this thesis in Ch. 6.

$W' \to tb$ searches

The channel is particularly important because a W' boson that couples only to right-handed fermions cannot decay to leptons if the right-handed neutrinos are heavier than $m_{W'}$. Additional motivations are provided by a W' boson with enhanced coupling to the third generation and by a leptophobic W' boson mentioned above. Experimentally, such channels have a more pronounced signature compared to other hadronic decays due to the presence of two b quarks in the final state, one from the W' decay and the other from the top quark decay, together with the additional decay products of the W boson. The ATLAS Collaboration searched for this new boson with proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 36.1 fb^{-1}[99]. This analysis focuses on final states with one electron or muon plus jets. The search covers resonance masses between 0.5 and 5.0 TeV and considers right-handed W' bosons. For right-handed W' bosons with coupling to the SM particles equal to the SM weak coupling constant, masses below 3.15 TeV are excluded at the 95% confidence level. This search also performs the combination with the fully hadronic final state [100], excluding RH W' boson masses below 3.25 TeV at the 95% confidence level, as reported in Fig. 2.2. The CMS Collaboration also presented a search in this channel with proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV and corresponding to an integrated luminosity of 35.9 fb^{-1}. Both RH and LH W' hypotheses are tested. For the LH case, interference with the SM single top quark $s-$channel is taken into account. Both cases for $M_{\nu_R} \ll m_{W'}$ and $M_{\nu_R} > m_{W'}$ are considered for the expected cross-sections and the resulting exclusion plot is shown in Fig. 2.2. The CMS Collaboration also used the cross-section measurement to extract bounds varying the values of the α_L and α_R couplings. The obtained results are shown in Fig. 2.3 as a function of the left-handed (α_L) and right-handed (α_R) couplings for different W' mass hypotheses. The measurement with proton-proton collision data collected in 2016-2018 with the CMS detector

at the LHC at a center-of-mass energy of 13 TeV, corresponding to an integrated luminosity of 137 fb^{-1}, is one of the topics of this work and is described in detail in Ch. 6.

The $W' \to tb$ channel can also be studied in the full hadronic final state. The ATLAS Collaboration set the upper limit at 95% CL at 3.0 (2.9) TeV [100] for RH (LH) W' using proton-proton collisions at $\sqrt{s} = 13$ TeV, corresponding to 36.1 fb^{-1}. The CMS Collaboration instead searched for LH W' production excluding masses up to 3.4 TeV [101] using proton-proton collisions at $\sqrt{s} = 13$ TeV corresponding to 137 fb^{-1}.

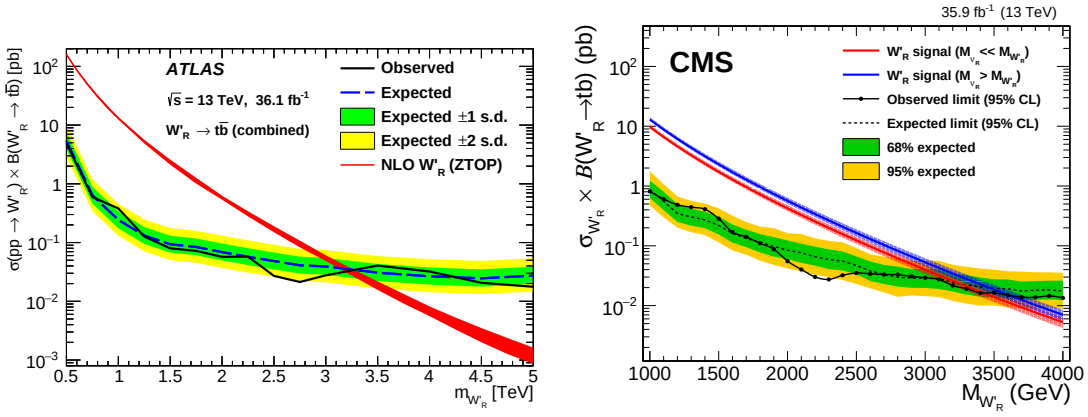


Figure 2.2: Upper limit at 95% CL on the RH W' boson production cross-section for the combined electron and muon channels. Signal masses for which the theoretical cross-section (in red and blue for $M_{\nu_R} \ll m_{W'}$ and $M_{\nu_R} > m_{W'}$) exceeding the observed upper limit (in solid black) are excluded at 95% CL. The green and yellow bands represent the ± 1 and 2 standard deviation uncertainties in the expected limit, respectively [102].

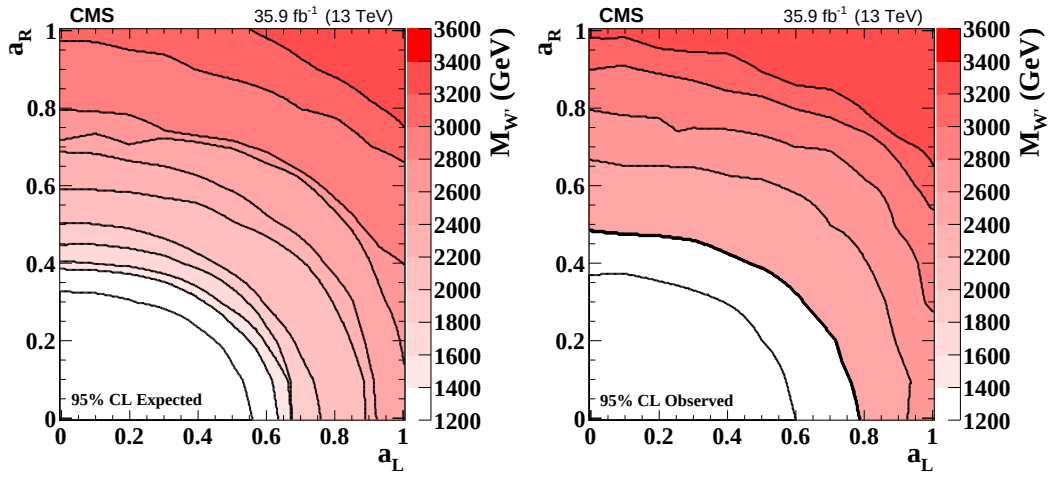


Figure 2.3: Expected (left) and observed (right) bounds on the W' boson mass as a function of the left-handed (α_L) and right-handed (α_R) couplings. Black lines represent contours of the same W' boson mass [102].

2.2 Indirect approach: Effective Field Theories (EFTs)

When considering a model-independent approach to non-standard interactions, there are several desirable features to be taken into account:

- any extension of the SM should satisfy the $S-$matrix axioms of unitarity, analyticity, etc;
- the symmetries of the SM, namely Lorentz invariance and $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry, should be respected;
- should be possible to recover the SM in an appropriate limit;
- the extended theory should be general enough to capture all BSM physics, but should give some guidance as to where the effects of the New Physics are most likely to be seen.

An effective quantum field theory can incorporate all these features, as shown in [103, 104]. In the SM framework described in Ch. 1, all operators constructed as products of fields in the lagrangian density are restricted to be of mass dimension four or less, and thus new operators of higher dimension are needed to extend the theory to construct an EFT. By dimensional analysis, these operators have coefficients that are inverse powers of the mass and are therefore suppressed when the mass is large compared to the experimentally accessible energies. As a consequence, the dominant extended operators will be those of the lowest dimensions. There is only one operator of mass dimension five, and it is responsible for generating Majorana masses for neutrinos [105]. There are many dimension six operators, but typically only a few contribute to a given physical process [106, 107], and the same reasoning can be applied to dimension eight operators [108]. The following discussion will focus on dimension six (dim-6) and dimension eight (dim-8) EFT operators.

Two different types of EFTs can be distinguished, top-down and bottom-up. The first approach starts with a full theory and then effective operators are calculated in certain energy limits. For example, a heavy degree of freedom can be integrated out and the full theory adapted to a low energy EFT [109]. This can be useful to get better insight into the model by separating long-range effects from short-range effects, using only the necessary degrees of freedom. The bottom-up approach can be used when the underlying New Physics is unknown. If we denote the mass scale characterizing the coefficients of the higher dimensional operators as Λ_{NP}, i.e. the scale of New Physics, the underlying assumption in the bottom-up approach of an EFT is that this scale is large compared to the experimentally accessible energies. Thus, an EFT is a low-energy approximation to New Physics, where "low" means much smaller than Λ_{NP}.The generality of the framework allows for studying the low-energy effects of any possible BSM model, including new particles, extra spacetime dimensions, or even physics not described by ordinary quantum field theory, such as string theory. The starting point for constructing an EFT is, in the end, the SM Lagrangian $\mathscr{L}_{SM}$ extended by additional effective operators suppressed by a power of Λ:

$$\mathscr{L}_{EFT} = \mathscr{L}_{SM} + \sum_i \frac{c_i}{\Lambda_{NP}^{d-4}} \mathscr{O}_i \tag{2.6}$$

with an unknown coupling c_i. In the EFT approach, the particle content of the theory must be fully specified, i.e. the fields from which the operators $\mathscr{O}_i$ are constructed. These are the fields present at low energies $E \ll \Lambda_{NP}$. These are assumed to be the SM fields and nothing else. An EFT is only useful up to energies of the order of Λ_{NP}. At energies beyond that, there is no justification for neglecting operators of arbitrary dimension, since they are not suppressed, and since there are an infinite number of them, the usefulness of this approach is lost.

In this thesis, a bottom-up approach is used to construct and implement SMEFT operators and both dim-6 and dim-8 operators that affect same-sign W VBS with a hadronic tau in the

final state, and are fully illustrated in the remainder of this Section. The experimental investigation of these classes of SMEFT operators is fully illustrated in Ch. 7. Thus, the effective extended Lagrangian considered in this thesis to investigate indirect New Physics effects with the SMEFT approach is the following:

$$\mathscr{L}_{EFT} = \mathscr{L}_{SM} + \sum_{\alpha} \frac{c_\alpha}{\Lambda^2} Q_\alpha + \sum_{i} \frac{f_i}{\Lambda^2} \mathscr{O}_i \tag{2.7}$$

where c_i and f_j are the Wilson coefficients measuring the coupling to each dim-6 Q_α and dim-8 $\mathscr{O}_i$ operator, respectively. The implementation of all of them is discussed in Sec. 5.1 for what concerns simulated events.

The dim-6 and dim-8 SMEFT operators that affect the ssWW VBS process investigated in this study are presented below, along with an explanation of their impact on the number of observable events. They are defined according to the Warsaw and Eboli bases, respectively. Historically, in the search for anomalous couplings in VBS processes, only the effects of the dim-8 operators are included, considering only the anomalous quartic gauge couplings. This is possible when the focus of the study is accessing the self-coupling of the gauge boson in the quartic boson vertex, which is the main feature of the VBS processes, and the effects of the dim-6 operators are neglected. The study presented in Ch. 7 proposes to consider, for the first time in the same study, all the possible anomalous couplings in the ssWW VBS with a hadronic tau channel, both the triple and the quartic ones and therefore to implement the usual dim-8 operators together with the dim-6 operator, which affects both the gauge boson self-couplings and the boson-fermion ones.

2.2.1 dim-6 operators: the Warsaw basis

As mentioned above, the first contributions from New Physics could come from dim-6 operators within the EFT approach, which arise mainly from the integration of loop corrections of heavy new particles. An analysis of the complete dim-6 basis with all available SM operators has been done in [107]. The basis consists of 16 operators contributing to anomalous gauge and Higgs couplings, 29 operators consisting of four fermion fields, and 35 mixed operators having only two fermion fields. In total, there are 80 operators. A more relevant study showed that many of the two-fermion-field operators are redundant, and the full basis can be reduced to 59 independent operators [30], named the "Warsaw basis" after the institution to which the authors are affiliated. Among these operators, which require that the lepton and baryon numbers are conserved as well as the CP symmetry, 15 operators significantly affect the VBS process of a same-sign W boson pair VBS studied in this work, modifying all vertices involving at least three fermionic or bosonic currents [110].

Specifically, there are:

- 6 pure fermionic operators,

$$\begin{aligned}
Q_{ll} &= \delta_{pr}\delta_{st}\left(\bar{l}_p\gamma_\mu l_r\right)\left(\bar{l}_s\gamma^\mu l_t\right),\\
Q_{ll}^{(1)} &= \delta_{pt}\delta_{sr}\left(\bar{l}_p\gamma_\mu l_r\right)\left(\bar{l}_s\gamma^\mu l_t\right),\\
Q_{qq}^{(1)} &= \delta_{pr}\delta_{st}\left(\bar{q}_p\gamma_\mu q_r\right)\left(\bar{q}_s\gamma^\mu q_t\right),\\
Q_{qq}^{(1,1)} &= \delta_{pt}\delta_{sr}\left(\bar{q}_p\gamma_\mu q_r\right)\left(\bar{q}_s\gamma^\mu q_t\right),\\
Q_{qq}^{(3)} &= \delta_{pr}\delta_{st}\left(\bar{q}_p\gamma_\mu \sigma^i q_r\right)\left(\bar{q}_s\gamma^\mu \sigma^i q_t\right),\\
Q_{qq}^{(3,1)} &= \delta_{pt}\delta_{sr}\left(\bar{q}_p\gamma_\mu \sigma^i q_r\right)\left(\bar{q}_s\gamma^\mu \sigma^i q_t\right);
\end{aligned}$$

- 5 pure bosonic operators,

$$
\begin{aligned}
Q_W &= \varepsilon^{ijk} W_\mu^{\nu i} W_\nu^{\rho j} W_\rho^{\mu k}, \\
Q_{H\Box} &= \left(H^\dagger H\right) \Box \left(H^\dagger H\right), \\
Q_{HD} &= \left(H^\dagger D^\mu H\right) \left(H^\dagger D_\mu H\right), \\
Q_{HW} &= H^\dagger H W_{\mu\nu}^i W^{\mu\nu i}, \\
Q_{HWB} &= H^\dagger \sigma^i H W_{\mu\nu}^i B^{\mu\nu};
\end{aligned}
$$

- 4 mixed operators,

$$
\begin{aligned}
Q_{Hl}^{(1)} &= \left(H^\dagger i \overleftrightarrow{D}_\mu H\right) \left(\bar{l}_p \gamma^\mu l_p\right), \\
Q_{Hl}^{(3)} &= \left(H^\dagger i \overleftrightarrow{D}_\mu^j H\right) \left(\bar{l}_p \sigma^j \gamma^\mu l_p\right), \\
Q_{Hq}^{(1)} &= \left(H^\dagger i \overleftrightarrow{D}_\mu H\right) \left(\bar{q}_p \gamma^\mu q_p\right), \\
Q_{Hq}^{(3)} &= \left(H^\dagger i \overleftrightarrow{D}_\mu^j H\right) \left(\bar{q}_p \sigma^j \gamma^\mu q_p\right).
\end{aligned}
\tag{2.8}
$$

with more details about the notation given in [30]. Among them, $Q_{Hl}^{(3)}, Q_{ll}, Q_{ll}^{(1)}, Q_{HD}, Q_{HWB}$ enter via modifications of the EW inputs, $Q_{Hl}^{(1)}$, $Q_{Hl}^{(3)}, Q_{Hq}^{(1)}, Q_{Hq}^{(3)}$ induce modifications of the Vff couplings, Q_W modifies the gauge coupling Q_{HD}, $Q_{HW}, Q_{HWB}, Q_{H\Box}$ modify HVV couplings, and finally $Q_{qq}^{(1)}, Q_{qq}^{(3)}, Q_{qq}^{(1,1)}, Q_{qq}^{(3,1)}$ enter the four-quark contact terms. Moreover, when present, the first superscript indicates if the charged current in the operator is a SU(2) triplet (3) or a singlet (1), and the second distinguishes among the flavor content of the currents mentioned above. Fig. 2.4 gives a pictorial description of what has just been illustrated.

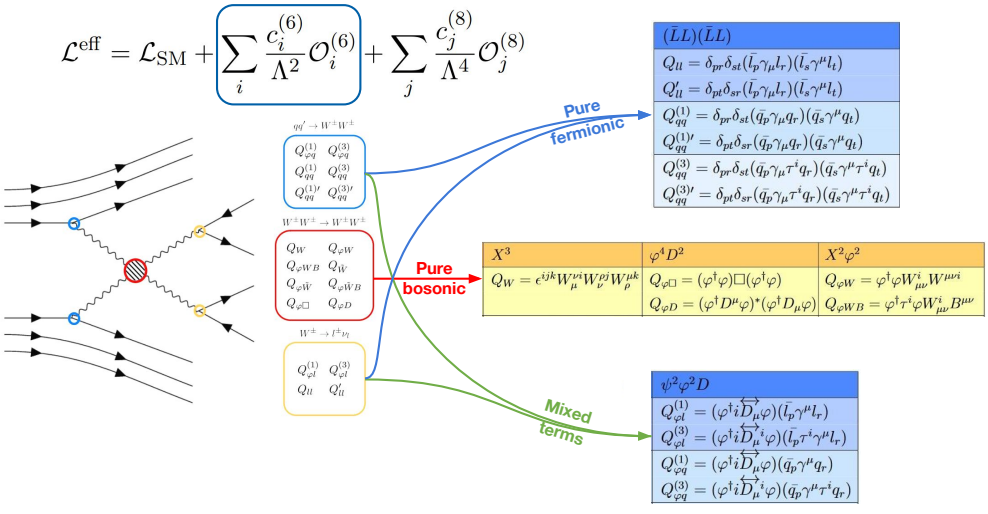


Figure 2.4: Pictorial representation of the dim-6 SMEFT operators affecting the electroweak vertices in same sign W pair scattering.

2.2.2 dim-8 operators: the Eboli basis

Some New Physics contributions can induce anomalies in the quartic gauge couplings without perturbing the triple gauge couplings [111]. These cannot be modeled by dim-6 operators, so dim-8 operators are needed.

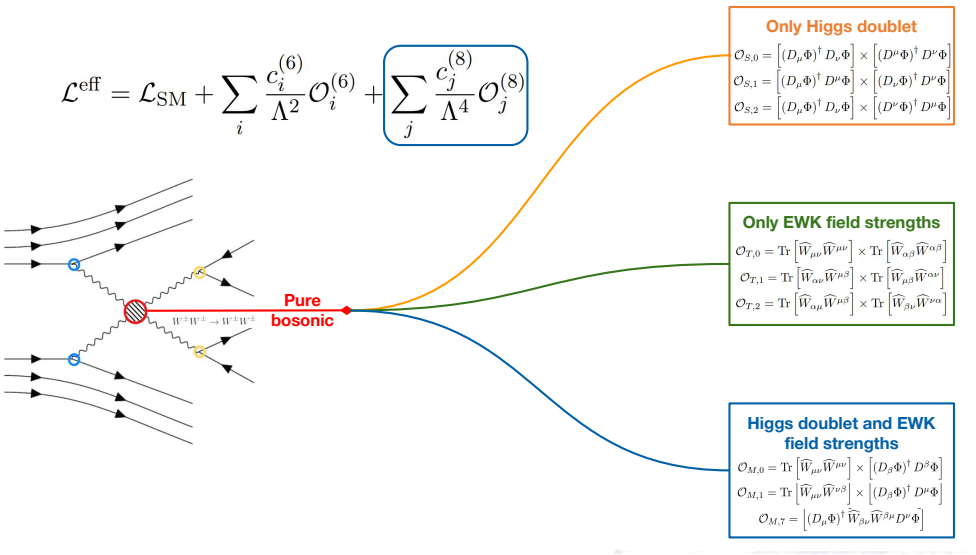


Figure 2.5: Pictorial representation of the dim-8 SMEFT operators affecting the electroweak same-sign W pair scattering.

A comprehensive set of dim-8 operators is introduced in [112], then perfected in [108], and is based on all possible combinations of SM fields. The basis is named "Eboli" after the main author who proposed it, and it consists of 18 operators. Among them, the following ones contribute to the same-sign W pair scattering process [113] studied in Ch. 7:

- 3 longitudinal operators, containing only Higgs doublet fields:

$$\begin{aligned}
\mathcal{O}_{S0} &= \left[\left(D_\mu H\right)^\dagger D_\nu H\right] \times \left[\left(D^\mu H\right)^\dagger D^\nu H\right], \\
\mathcal{O}_{S1} &= \left[\left(D_\mu H\right)^\dagger D^\mu H\right] \times \left[\left(D_\nu H\right)^\dagger D^\nu H\right], \\
\mathcal{O}_{S2} &= \left[\left(D_\mu H\right)^\dagger D_\nu H\right] \times \left[\left(D^\nu H\right)^\dagger D^\mu H\right];
\end{aligned}$$

- 3 purely transverse operators, containing only EW field strengths:

$$\begin{aligned}
\mathcal{O}_{T0} &= \text{Tr}\left[\hat{W}_{\mu\nu}\hat{W}^{\mu\nu}\right] \times \text{Tr}\left[\hat{W}_{\alpha\beta}\hat{W}^{\alpha\beta}\right], \\
\mathcal{O}_{T1} &= \text{Tr}\left[\hat{W}_{\alpha\nu}\hat{W}^{\mu\beta}\right] \times \text{Tr}\left[\hat{W}_{\mu\beta}\hat{W}^{\alpha\nu}\right], \\
\mathcal{O}_{T0} &= \text{Tr}\left[\hat{W}_{\alpha\mu}\hat{W}^{\mu\beta}\right] \times \text{Tr}\left[\hat{W}_{\beta\nu}\hat{W}^{\nu\alpha}\right];
\end{aligned}$$

- 3 mixed operators, containing both Higgs doublet and EW field strengths:

$$\begin{aligned}
\mathscr{O}_{M0} &= \mathrm{Tr}\left[\hat{W}_{\mu\nu}\hat{W}^{\mu\nu}\right] \times \left[\left(D_\beta H\right)^\dagger D^\beta H\right], \\
\mathscr{O}_{M1} &= \mathrm{Tr}\left[\hat{W}_{\mu\nu}\hat{W}^{\mu\nu}\right] \times \left[\left(D_\beta H\right)^\dagger D^\beta H\right], \\
\mathscr{O}_{M7} &= \mathrm{Tr}\left[\left(D_\mu H\right)^\dagger \hat{W}_{\beta\nu}\hat{W}^{\beta\mu} D^\nu H\right].
\end{aligned} \tag{2.9}$$

A pictorial description of what was previously illustrated is reported in Fig. 2.5.

2.2.3 Effects of SMEFT operators with different dimensionality

Starting from the effective lagrangian $\mathscr{L}_{EFT}$ written in Eq. 2.7, the scattering amplitude $|\mathscr{A}|^2$ can be calculated to deduce the observable effects of implementing SMEFT operators in known SM processes, and finally to predict the expected number of events N:

$$\begin{aligned}
N \propto |\mathscr{A}|^2 = |\mathscr{A}_{\mathrm{SM}}|^2 &+ \sum_\alpha \frac{c_\alpha}{\Lambda^2} \cdot 2\Re\left(\mathscr{A}_{\mathrm{SM}}\mathscr{A}^\dagger_{Q_\alpha}\right) + \sum_{\alpha,\beta} \frac{c_\alpha c_\beta}{\Lambda^4} \cdot \left(\mathscr{A}_{Q_\alpha}\mathscr{A}^\dagger_{Q_j}\right) && \text{only EFT dim}-6 \text{ contribution} \\
&+ \sum_i \frac{f_i}{\Lambda^4} \cdot 2\Re\left(\mathscr{A}_{\mathrm{SM}}\mathscr{A}^\dagger_{\mathscr{O}_i}\right) + \sum_{i,j} \frac{f_i f_j}{\Lambda^8} \cdot \left(\mathscr{A}_{\mathscr{O}_i}\mathscr{A}^\dagger_{\mathscr{O}_j}\right) && \text{only EFT dim}-8 \text{ contribution} \\
&+ \sum_{\gamma,k} \frac{c_\gamma f_k}{\Lambda^6} \cdot 2\Re\left(\mathscr{A}_{Q_\gamma}\mathscr{A}^\dagger_{\mathscr{O}_k}\right) && \text{mixed EFT contribution}
\end{aligned} \tag{2.10}$$

at the end:

$$\begin{aligned}
N = N_{SM} &+ \sum_\alpha \left[\frac{c_\alpha}{\Lambda^2} N^{int}_\alpha + \frac{c^2_\alpha}{\Lambda^4} N^{pure}_\alpha\right] + \sum_{\alpha\neq\beta} \frac{c_\alpha c_\beta}{\Lambda^4} N^{mix}_{\alpha\beta} && \text{only EFT dim}-6 \text{ contribution} \\
&+ \sum_i \left[\frac{f_i}{\Lambda^4} N^{int}_i + \frac{f^2_i}{\Lambda^8} N^{pure}_i\right] + \sum_{i\neq j} \frac{f_i f_j}{\Lambda^8} N^{mix}_{ij} && \text{only EFT dim}-8 \text{ contribution} \\
&+ \sum_{\alpha,k} \frac{c_\gamma f_k}{\Lambda^6} N^{mixed}_{\gamma k} && \text{mixed EFT contribution}
\end{aligned} \tag{2.11}$$

Usually, in collider searches, the first and second row contributions are treated separately, both in simulation studies and in data studies, and then independent constraints on the dim-6 c_α and dim-8 f_i Wilson coefficients are extracted.

However, dim-6 and dim-8 can simultaneously act on the same vertex, i.e. W pair scattering, and if only one dim-8 operator contribution to the expected number of events N is considered, the contribution of any dim-6 operator of equal or smaller order in terms of the power of Λ is discarded. Since anomalous Triple Gauge Coupling (aTGC) effects induced by dim-6 operators are competitive with anomalous Quartic Gauge Coupling (aQGC) effects induced by dim-8 operators, ignoring the former may lead to an overestimation of the Wilson coefficients associated with the latter in dedicated experimental studies. Even if only EFT operators of the same dimension are considered, they are not uncorrelated, and a combination of the effects of more than one operator at a time has to be performed to properly estimate and eventually constrain the corresponding Wilson coefficients. It is, therefore, necessary to properly study the combination of more than one SMEFT operator, even with different dimensions. In this way, it would be possible to correctly understand the impact of each of them in the possible anomalous

W coupling and to give more reliable interpretations in terms of the most promising directions where to find hints of New Physics.

Vector boson scattering is an ideal process to study dim-6 and dim-8 operators simultaneously, both separately and considering their interaction since both classes of SMEFT operators have a consistent effect on this peculiar SM process. In this thesis, the first study of the indirect BSM effect modeled by combining dim-6 and dim-8 SMEFT operators in a VBS process is presented in Ch. 7, including all the operators described in this Section. The impact of each of them is examined separately, considering the contribution in terms of the deviation of the expected number of events with respect to the SM process as reported in Eq. 2.11. A combination of all possible SMEFT operator pairs constructed with the chosen set of dim-6 and dim-8 operators is also performed, considering all contributions in the series expansion Eq. 2.10 up to Λ^4. The presence of a τ lepton decaying in hadrons is expected to enhance the sensitivity to the operators, including the Higgs field, being the heaviest known charged lepton.

2.3 Impact of New Physics in Vector Boson Scattering

VBS studies can constrain the existence of Beyond Standard Model (BSM) physics in several ways. Searches for New Physics in VBS channels can be divided into those based on an explicit New Physics model and general model-independent searches, usually parameterized with the effective field theory (EFT) approach [114].

EFT constraints obtained in experimental analyses in ATLAS and CMS use the parameterization reported in [108], where dim-8 operators are considered. As introduced in Sec. 2.2, at dim-8 one can assume the presence of anomalous quartic gauge couplings (aQGC) and no anomalous triple gauge couplings (aTGC). There are 18 independent bosonic dim-8 operators relevant to 2-to-2 scattering processes involving Higgs or gauge bosons at the tree level. Several CMS analyses constrain the physics of non-zero dimension 8 operators, with the latest results published in [115, 116], while ATLAS Run-II measurements mostly focus on SM VBS observations and do not provide explicit constraints on BSM physics, but ATLAS results using the same model are available in 7 and 8 TeV analyses. A complete collection for both Collaborations is reported in [117] and the available constraints are shown from Fig. 2.6 to Fig. 2.7. ATLAS has instead investigated the effect of this class of operators in the vector boson fusion process with a Z boson [118]. Since these operators induce modifications of the high-energy tail of the scattering process distributions, the experimental analyses first select events in VBS-enhanced regions and then use a distribution sensitive to this modification to constrain the couplings themselves. At the LHC, such distributions include the invariant mass of the diboson system (or approximations if it is not accessible) or the transverse momentum of one of the scattered gauge bosons.

It is important to note that the presence of non-zero aQGCs would violate unitarity at sufficiently high energies. Limits that account for this effect can be set by stopping the EFT integration at the unitarity limit and considering only the expected SM contribution for generated events with diboson-invariant masses above the unitarity limit. The unitarity limits for each aQGC parameter, typically around $1.5-2.5$ TeV, are usually calculated using the VBFNLO program [119] or taken from [108]. These limits are typically less stringent than those obtained without considering the unitarity violation.

If it exists, the BSM physics is unlikely to be restricted to the VBS processes. The choice of selecting operators that generate aQGC in the absence of aTGC is then not obvious, since it breaks the EFT paradigm where operators with lower powers of Λ should be constrained first, while the data do not yet exclude all possible aTGC. Moreover, in most VBS processes, aTGC effects enter directly, for example through special $s-$channel diagrams. In a dimension-6 real-

ization of the EFT approach, operators affecting VBS analyses are also relevant for non-VBS VV production, triple-gauge boson, and Higgs boson production. As a consequence, they should eventually be constrained together in a larger-scale fit. Studies of dim-6 effects on specific VBS final states have been done [120, 121], and in this direction, a very recent phenomenological work [122] attempts a parameterization of many existing VBS results. The goal is to compare the bounds on several Wilson coefficients considering only the inclusive VV production versus the inclusion of VBS, finding $15-50\%$ weaker bounds when VBS is not included, depending on the specific operator. As mentioned above, the investigation of anomalous couplings in the VBS process presented in Ch. 7 takes a step forward in this direction, combining for the first time dim-6 and dim-8 SMEFT operator effects in the same channel.

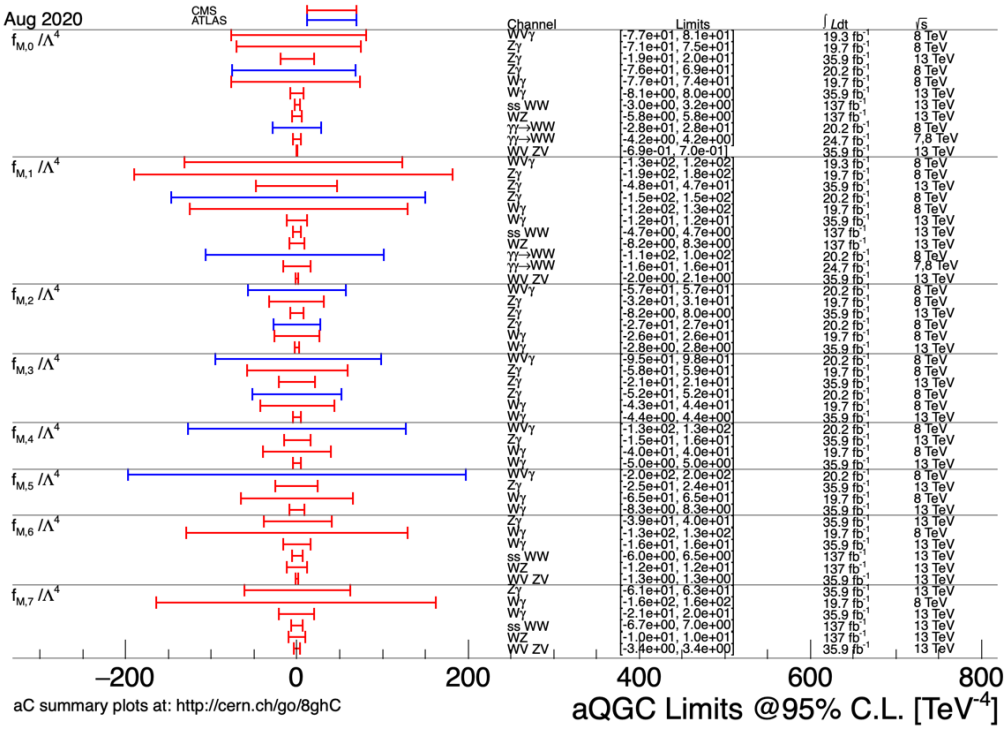
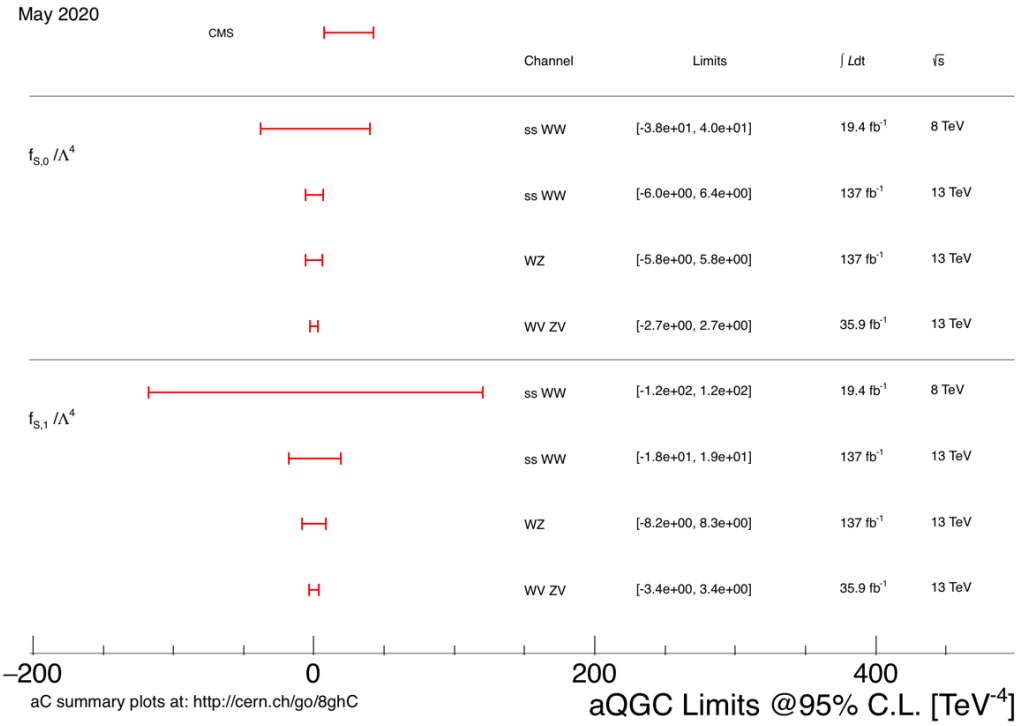


Figure 2.6: Current constraints on mixed dim-8 operator couplings from various ATLAS (blue) and CMS (red) analyses, with corresponding integrated luminosities.



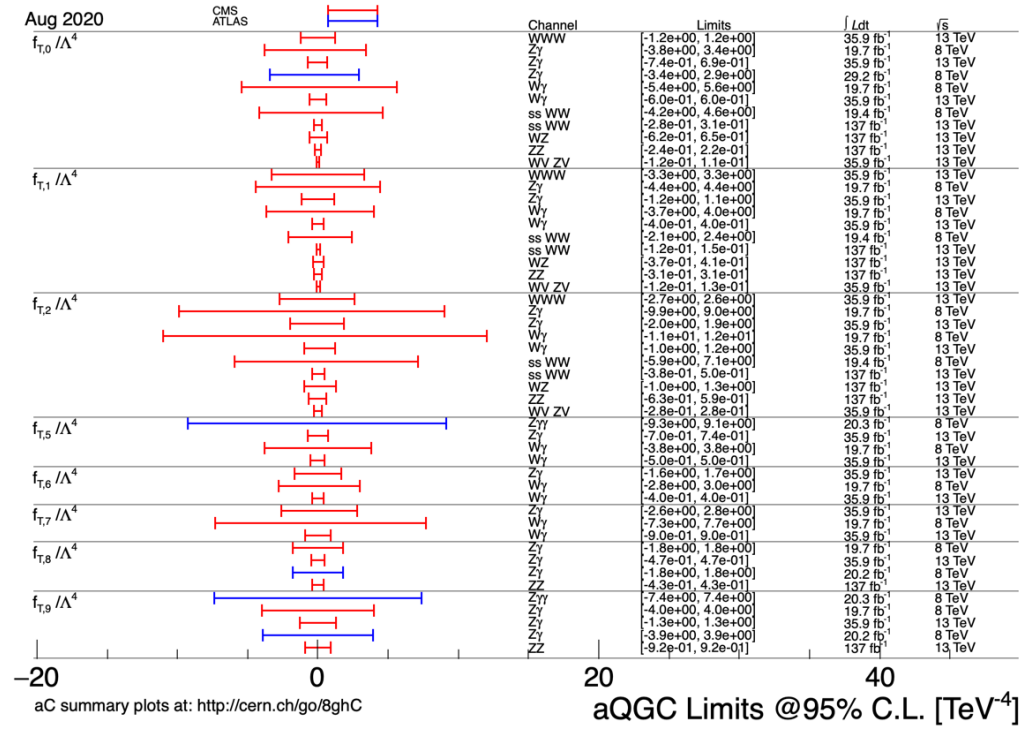


Figure 2.7: Current constraints on scalar (top) and transverse (bottom) dim-8 operator couplings from various ATLAS (blue) and CMS (red) analyses, with corresponding integrated luminosities.

Part II

The experimental setup and techniques

Chapter 3

The CMS experiment at the LHC

The prime objective of the Large Hadron Collider (LHC) [123] is to explain the nature of the electroweak symmetry breaking responsible for the Higgs mechanism. The experimental study of the Higgs mechanism can also shed light on the mathematical consistency of the SM at energy scales in the multi-TeV range. Various alternatives to the SM invoke new symmetries, forces, or constituents. As discussed in Ch. 2, there are many compelling reasons to investigate the TeV energy scale and beyond.

The LHC also provides high-energy heavy-ion beams with energies more than 30 times higher than those of previous accelerators, allowing the study of QCD matter under extreme conditions of temperature, density, and parton momentum fraction. Hadron colliders are well suited to the task of exploring new energy regimes, and the 1 TeV center-of-mass energy regime can be explored if the proton energy and luminosity are high enough. Access to the TeV and multi-TeV scale could provide clues to New Physics beyond the SM, as well as answers to current open questions in fundamental particle physics. The beam energy and design luminosity of the LHC has been chosen to study physics at the TeV energy scale: this corresponds to a sevenfold increase in energy and a hundredfold increase in integrated luminosity compared to previous hadron collider experiments, which means that a wide range of physics processes are potentially within reach of the LHC. To exploit this immense physics potential, four major experiments have been designed. This Chapter provides an overview of the LHC machine and its operation and a detailed description of the Compact Muon Solenoid (CMS) experiment.

3.1 The Large Hadron Collider

The LHC at CERN (Conseil Européen pour la Recherche Nucléaire) near Geneva is the largest circular accelerator to date and one of the most powerful tools for fundamental particle physics research. It is designed to collide proton beams with a center-of-mass energy of 14 TeV and a luminosity of 10^{34} cm^{-2} s^{-1}. It can also collide heavy (Pb) ions with an energy of 2.8 TeV per nucleon and a peak luminosity of 10^{27} cm^{-2} s^{-1}. The goal of the LHC is to unravel physics beyond the SM with center-of-mass collision energies up to 14 TeV.

The number of events per second generated in the LHC collisions is given by:

$$N_{event} = L\sigma_{event} \tag{3.1}$$

where σ_{event} is the cross-section for the event under study and L is the machine luminosity. The machine luminosity depends only on the beam parameters and can be written, for circular

proton-proton accelerators and assuming a Gaussian beam distribution, as:

$$L = \frac{N_b^2 \gamma_r f_{rev} n_b}{4\pi\varepsilon_n \beta^*} F, \tag{3.2}$$

where N_b is the number of particles per bunch, n_b is the number of bunches per beam, f_{rev} is the revolution frequency, γ_r is the relativistic gamma factor, ε_n is the normalized transverse beam emittance, β^* is the beta function at the point of collision, and F is the geometric luminosity reduction factor due to the crossing angle at the point of interaction.

The LHC is located in a 26.7 km-long underground tunnel originally built to house CERN's Large Electron Positron (LEP) accelerator. The tunnel lies between 45 to 170 meters below the surface, crossing the border between France and Switzerland and is connected to the CERN accelerator complex by two tunnels.

The CERN accelerating complex is shown in Fig. 3.1 and produces proton beams of energy 450 GeV: the first step of the acceleration process is performed by LINAC 2 (LINear particle ACcelerator), which produces proton beams of energy 50 MeV; the second step is performed by PSB (Proton Synchrotron Booster), which accelerates the beams to 1.4 GeV; then the beams are injected in PS (Proton Synchrotron) which produces proton beams of energy of 26 GeV; finally, the last step is SPS (Super Proton Synchrotron) and the beams reach the energy of 450 GeV and can be injected in the LHC where they reach their maximum energy of 7 TeV. To achieve this goal, the LHC consists of a superconducting cavity system of radio-frequencies

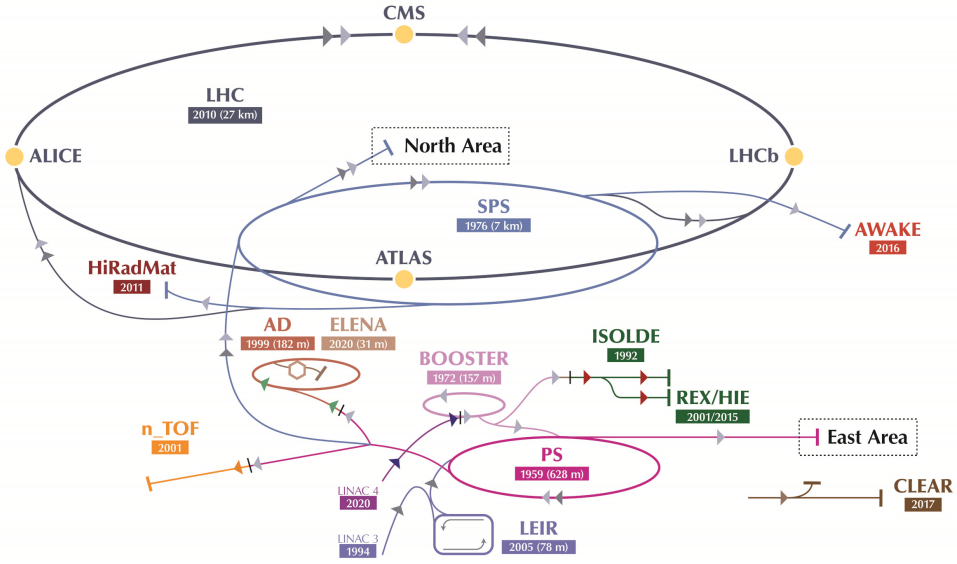


Figure 3.1: The accelerator chain at the CERN.

(RF) that accelerate the beams, 1232 dipole magnets needed to hold the beams in the circular ring, 392 quadrupole magnets that focus the beams, and other magnets for spool piece correction magnets (sextupole and octupole/decapole). The LHC magnet system uses proven technology based on NbTi Rutherford cables and cools the magnets to a temperature below 2 K using superfluid helium and operates at fields above 8.33 T.

The beam lifetime and beam-induced background requirements for the experiments determine the vacuum pressure conditions in the magnet tube. The LHC has three vacuum systems: the cryo-magnet isolation vacuum, the helium distribution isolation vacuum, and the beam vacuum. The isolation vacuum before cooling does not need to be better than 10^{-1} mbar, but at cryogenic temperatures, in the absence of significant leakage, the pressure will stabilize around 10^{-6} mbar. In the interaction regions around the experiments, the requirements for the beam

vacuum are much more stringent in order to minimize the background of the experiments. In the room temperature parts of the beam vacuum system, the pressure should be in the range of 10^{-10} to 10^{-11} mbar.

3.1.1 Experiments and data-taking activity

The LHC has four main experiments:

- ALICE: *A Large Ion Collider Experiment* [124] is a dedicated ion experiment; it works with $\sqrt{s} = 2.67$ TeV lead-lead ion collisions, aiming for a peak luminosity of $L = 10^{27}$ cm^{-2} s^{-1}.
- ATLAS: *A Toroidal LHC ApparatuS* [125] is a general-purpose detector whose goals are precision measurements of the SM, the search and study of the Higgs boson, and mechanisms due to New Physics. It is 46 m long and 25 m in diameter and is the largest experiment at the LHC.
- CMS: *Compact Muon Solenoid* [126] is described in Sec. 3.2.
- LHCb: *LHC beauty* [127] is an experiment designed and optimized to study the properties of the b quark and its production mechanism.

The LHC activity started in 2009 with a pivot run and ramped up over the year from a center-of-mass energy of 0.9 TeV, corresponding to the LHC beam injection energy, to 2.56 TeV at $\sqrt{s} = 0.9$ and $\sqrt{s} = 2.56$ TeV. In 2010, p-p collisions with center-of-mass energy $\sqrt{s} = 7$ TeV were produced until 2011; in 2012 and 2013 the p-p collisions were at $\sqrt{s} = 8$ TeV. The 2010-2013 activity is called LHC Run I. In 2013, the LHC stopped for the upgrade of the detectors in view of the $\sqrt{s} = 13$ TeV p-p collisions. In 2015 the LHC activity started again with $\sqrt{s} = 13$ TeV p-p collisions until 2018. This last period is called LHC Run II. The data sets collected during the years are listed in Tab. 3.1, as well as the part of those data sets where the whole CMS detector was active.

The data collected in a year by the CMS experiment are divided into different so-called primary data sets, depending on the trigger algorithms used (see also sec trigger), and into data-taking periods, and are labeled according to the convention “RunXY”, where X is the year and Y is a letter identifying the period of the year in which the experimental conditions, both for the beam and the detector, remained unchanged.

Period	Year	$\sqrt{s}$ [TeV]	LHC delivered [fb^{-1}]	CMS Recorded [fb^{-1}]	CMS Validated [fb^{-1}]
Run I	2010	7	40.22 $\times$ 10^{-2}	40.76$\times$ 10^{-2}	34.68 $\times$ 10^{-2}
	2011	7	6.13	5.55	5.09
	2012	8	23.30	21.79	19.79
Run II	2015	13	4.22	3.81	2.39
	2016	13	40.82	37.76	35.92
	2017	13	49.79	44.98	41.53
	2018	13	67.86	63.67	59.7

Table 3.1: The cumulative luminosity delivered by LHC, recorded by CMS and certified as Good for physics analysis, i.e. where all detectors were active and taking data, during each period of activity.

The plans for the LHC are shown in Fig. 3.2. The data-taking period of the LHC, called Run III, began in 2022 and is expected to double the integrated luminosity of Run II. The center-of-mass energy for pp collisions will be increased to 14 TeV. To reach the highest luminosity

in the 3 years of Run III, a high pile-up of interactions in a single bunch crossing is expected. The maximum instantaneous pile-up in Run 3 will be similar to that of the 2018 data period, but larger on average. All four major experiments have been upgraded to handle such challenging conditions.

The future High Luminosity (HL) phase of the LHC, scheduled to be operational in early 2029, will collide protons at $\sqrt{s} = 14$ TeV, with a projected peak instantaneous luminosity of about $5 \cdot 10^{34}$ cm^{-2} s^{-1}, which will be increased to about $7.5 \cdot 10^{34}$ cm^{-2} s^{-1} [128]. These rates are about three to four times higher than the peak luminosity of Run II and will accumulate to a total integrated luminosity of about 3 ab^{-1}. All four large detectors will be upgraded to cope with the new HL-LHC operating conditions.

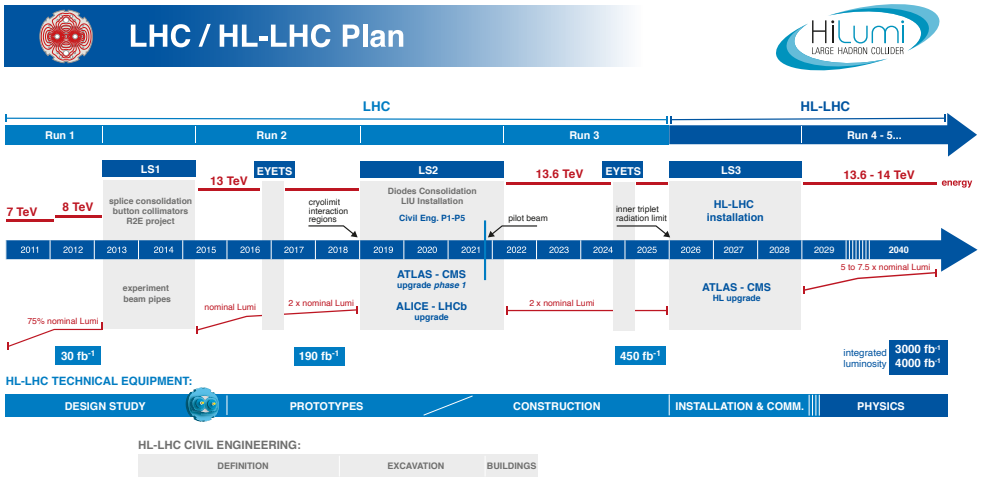


Figure 3.2: Plans for the LHC collider as updated in February 2022. “LS” stands for Long Shutdown, indicating periods when the LHC is not in operation and the detectors and collider are undergoing major upgrade programs.

3.2 The Compact Muon Solenoid

The CMS experiment is installed at one of the four interaction points in the LHC tunnel, near the French village of Cessy, between Lake Geneva and the Jura mountains. As the name suggests, the main feature of CMS is a 3.8 T superconducting magnet. With a length of 13 m and a diameter of 6 m, it provides a high bending force to precisely measure the momentum of charged particles. Both the tracking system and most of the calorimeter systems are located inside the solenoid. The magnetic field lines of the solenoid are parallel to the beam direction in the central region. The flux is closed by a return yoke consisting of a 12-sided iron structure interspersed with four layers of muon stations. The CMS itself has a diameter of 15 m and a length of 28.7 m. The different sub-detectors exploit the fact that different types of particles interact differently and can therefore be distinguished by the signals they leave in different detector components. A sectional view of the CMS detector is shown in Fig. 3.3.

The detector has been designed to operate in a high-radiation environment, to maintain good performance over several years of data acquisition, and to be optimized to distinguish processes of interest from the background. A major experimental challenge is the so-called pile-up, i.e. the overlapping of different particles from the same event or from interactions in the same bunch crossing, which produces irreducible background. To reduce the impact of the

pile-up on the quality of the recorded data, the detector must have a high granularity, i.e. a high spatial resolution. In addition, a good time resolution is required to avoid the response of the detector, i.e. the readout electronics, is slower than the nominal time interval between two bunch crossings. One of the most important requirements of a detector such as CMS is hermeticity,

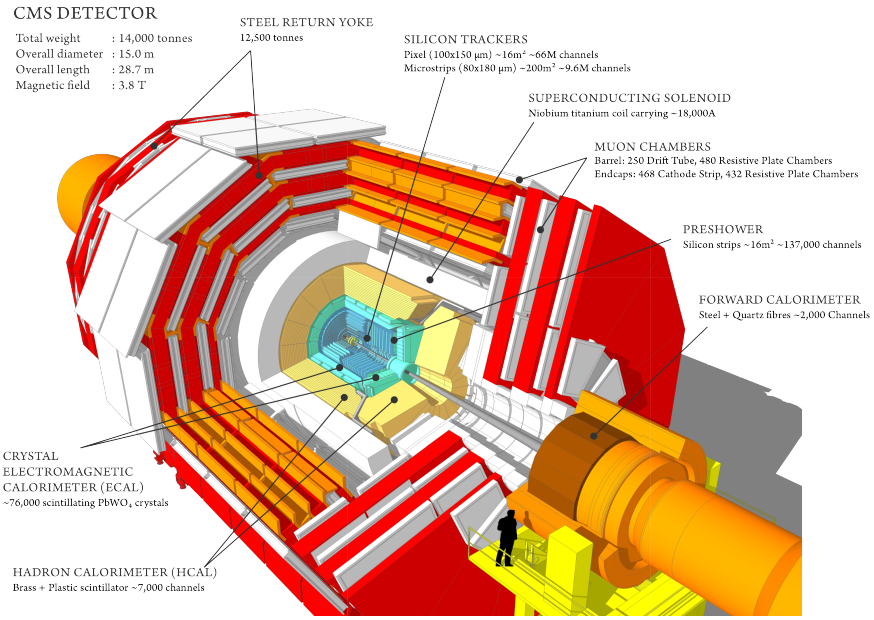


Figure 3.3: Global view of CMS detector.

and therefore CMS has a cylindrical structure (called a barrel) that covers the central region and two caps (endcaps) that cover the regions closest to the beams. In this way, it is possible to maintain cylindrical symmetry around the beam axis.

The reference frame of the CMS experiment is a right-handed Cartesian system with the origin centered at the nominal collision point. The Fig. 3.3 shows the Cartesian reference frame: the $x-$axis points radially towards the center of the LHC, the $y-$axis points vertically upwards and the $z-$axis is parallel to the beam direction, so the coordinate system is right-handed. The $x-y$ plane is also called the transverse plane. Given the cylindrical symmetry of the detector, the triplet (r,η,φ) can also be used as a coordinate system, where r is the radial distance from the $z-$axis, η is the so-called pseudorapidity, and φ is the azimuthal angle in the transverse plane. The polar angle θ measured from the $z-$axis is not used directly because the differences $\Delta\theta$ are not Lorentz invariant. It is related to pseudorapidity in the following way:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{3.3}$$

This is a good approximation of the particle rapidity y for high energy particles, defined as:

$$y = \frac{1}{2}\ln\left(\frac{E+p_z c}{E-p_z c}\right) \tag{3.4}$$

where E is the energy and p_z is the longitudinal momentum of the particle. The pseudorapidity results from taking the rapidity to the limit of massless particles. In this relativistic approximation, η transforms linearly under boosts along the $z-$ axis, implying that differences in this quantity between relativistic particles along the $z-$ axis are Lorentz invariant.

In addition to pseudorapidity, there are other useful quantities used to describe particle properties that will be mentioned in this document. An often used quantity is the distance ΔR in the

$\eta - \varphi$ plane, defined as:

$$\Delta R = \sqrt{(\Delta\varphi)^2 + (\Delta\eta)^2} \tag{3.5}$$

Two other very important quantities are the transverse momentum p_T and the transverse energy E_T, which are calculated from the components of the momentum vector in the transverse plane as follows:

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} \tag{3.6}$$

$$E_T = E \sin\theta \tag{3.7}$$

The CMS subdetector system from the interaction point outward consists of the inner tracking system, the electromagnetic calorimeter, the hadron calorimeter, the magnet, and finally the iron return yoke with muon chambers (Fig. 3.3). In the following, each of the subdetectors will be described in detail.

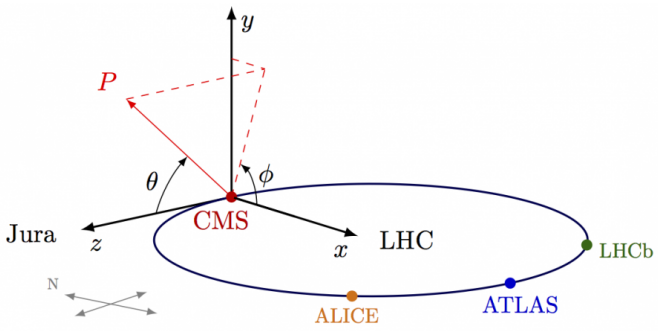


Figure 3.4: The CMS coordinate system.

3.2.1 Tracking system

Accurate reconstruction of the particle momentum is crucial for studying events in proton-proton collisions. The trajectory of a particle with an electric charge $e = \pm 1$ in a magnetic field B is bent by the Lorentz force. In a homogeneous magnetic field, the particle moves along a path described by a helix. Its transverse momentum p_T can be determined in the non-relativistic limit by:

$$p_T = e \int BdL \simeq eBR \tag{3.8}$$

where R is the radius of the curved track projected onto the transverse plane.

Due to the challenging environment in the LHC, the tracking system is required to have high granularity and a fast response given the high track multiplicity produced at each bunch crossing, high radiation hardness to face the intense particle flux in the region close to the proton beam and to be the least invasive possible and then limit processes that can mislead the trajectory reconstruction. A silicon-based tracking system meets all these requirements.

The CMS tracking system [129, 130] is composed of two parts, the Pixel Tracker and the Strip Tracker, and covers a pseudorapidity range up to $|\eta| = 2.5$.

The Pixel Tracker is the innermost subsystem of the CMS detector. It is located at a radial distance of less than 10 cm from the beamline, where the particle rate is about 10 million particles per cm^{-2} s^{-1}. It consists of 66 million silicon pixels with dimensions of 100 μm $\times$ 250 μm $\times$ 250 μm, covering a total area of about 1 m^2. The modules are arranged in three barrel layers (PXB) with mean radii of 4.4 cm, 7.3 cm, and 10.2 cm, and in four disks (PXF) with $z = \pm 34.5$ and $z = \pm 46.5$ cm. In the barrel layers, the magnetic field induces a Lorentz angle,

which increases the charge sharing between neighboring pixels. The pixel detectors allow a position resolution of 10 μm for the (r,φ) coordinate and 15 μm in the z direction, and in the forward direction a resolution of 15 μm and 20 μm respectively. This resolution is necessary not only for a precise track reconstruction but also for the determination of both the vertices produced in the primary interaction and the decay vertices of short-lived particles.

The Strip Tracker constitutes the outer layers of the tracking system. The basic building blocks are silicon strip modules, with each one of them equipped with one or two silicon sensors. The modules are arranged in ten layers in the barrel region extending up to a radius of 1.1 m. It is separated into an inner and outer part, called Tracker Inner Barrel (TIB) and Tracker Outer Barrel (TOB), respectively. Within one layer, there is a sub−structure of modules grouped called strings in TIB and rods in TOB. In the endcap region, the silicon Strip Tracker consists of two blocks of disks, three disks belonging to the Tracker Inner Disks (TID), and nine disks to the Tracker End Caps (TEC). In one disk of the TEC, 16 sub-structures of modules are installed. In total, the CMS silicon strip tracker has 9.3 million strips and covers 198 m^2 of active silicon area. The expected resolutions grow to 30 μm in (r,φ) and 300 μm along the $z-$coordinate.

The pixel readout electronics have been designed and optimized for the data rates and occupancies expected at the original LHC design luminosity, which is expected to operate at 1×10^{34} cm^{-2} s^{-1} with 25 ns bunch spacing. For these beam parameters, the dynamic inefficiency is about 4% at the nominal CMS readout rate. With the LHC operating parameters of the Run II period, the luminosity and the pileup are more than double the nominal ones, so this inefficiency in the inner layer increases to 16% in 2016. In order to maintain excellent performance, the Pixel Tracker was replaced during the technical shutdown at the end of 2016/2017. The Pixel Tracker that operated in 2017 and 2018 is a high-efficiency detector consisting of four barrel layers and three forward disks, providing four-hit pixel coverage up to $|\eta| = 2.5$. This "Phase-1" upgraded version of the CMS Tracker is shown in Fig. 3.5.

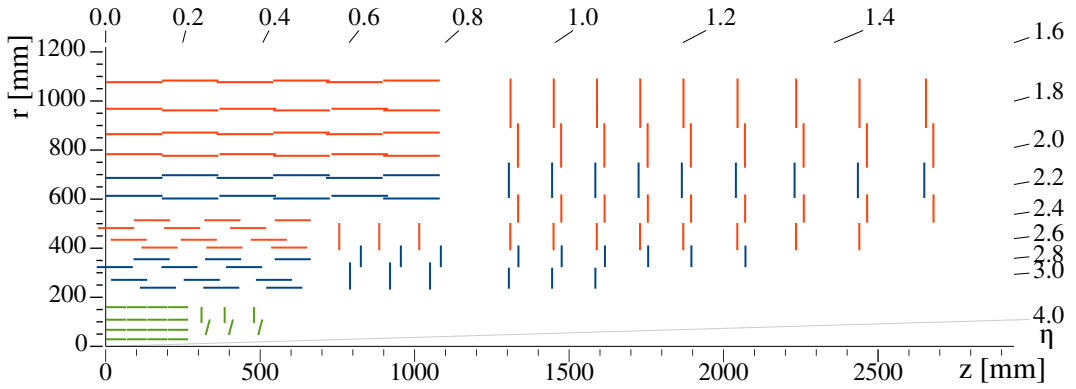


Figure 3.5: Sketch of a quarter of the Phase-1 CMS tracking system in $r-z$ view. The pixel detector is shown in green, while single-sided and double-sided strip modules are shown as red and blue segments, respectively.

The Iterative Tracking CMS track reconstruction algorithm [131] is based on the Combinatorial Kalman Filter (CKF) [132]. The first set of iterations looks for tracks that are easier to find, such as tracks with relatively high p_T. Track reconstruction starts with constructing seeds, which provide the initial estimate of the trajectory parameters. These proto-candidates are then propagated through the entire tracker, with the Kalman filter finding the compatible hit at each layer and updating the track candidate and its parameters. The tracks are then fitted after combining all associated hits and marked with quality flags. The track reconstruction has been updated to take advantage of the improved pixel detector capabilities in early 2017. More details about the track and vertex reconstruction algorithms used in CMS are described in Sec. 4.1.

3.2.2 The electromagnetic calorimeter

The CMS Electromagnetic Calorimeter (ECAL) [133] is located between the tracking system and the Hadron Calorimeter, inside the superconducting solenoid. Its main task is to allow the reconstruction of electrons and photons. It is a hermetic, homogeneous scintillating detector to obtain excellent energy resolution, good angular resolution, and accurate pion/photon separation. High-energy electrons and photons interact in the calorimeter to produce a cascade of lower-energy electrons and photons, called an electromagnetic shower. The process evolves due to subsequent pair production by photons and bremsstrahlung by electrons.

ECAL consists of 75848 lead tungstate ($PbWO_4$) crystals organized in structures of modules, 61200 crystals in the electromagnetic calorimeter barrel (EB), and 7324 crystals in the electromagnetic calorimeter endcap (EE). The EB has an inner radius of 1.29 m and its modules are structured in so-called supermodules. There are 36 supermodules in total, 18 for each side of the interaction point. The supermodules are organized in such a way that the crystal axes form a 3° angle with respect to a straight line from the interaction point, both in η and in φ directions. This arrangement is used to avoid holes in geometric acceptance. The two EE are placed at a distance of 3.17 m from the interaction point. They cover the pseudorapidity range of $1.479 < |\eta| < 3.0$. The choice of lead tungstate as the active detector material is dictated by some special properties, such as the short radiation length $X_0 = 0.89$ cm and the small Moliere radius $r_M = 21.9$ mm, the high radiation resistance, and the fast response. The result is a compact and granular calorimeter that can be placed inside the magnet and collects most of the electromagnetic shower in a few crystals. The energy resolution of the CMS ECAL is:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{0.12}{E}\right)^2 + (0.30\%)^2 \tag{3.9}$$

where the three different represent stochastic, noise, and constant terms, respectively.

3.2.3 The hadron calorimeter

The Hadron Calorimeter (HCAL) [134] measures the energy of hadronic jets and is particularly important for the detection of neutrinos or exotic particles, which manifest themselves in the detector as an apparent missing transverse energy. It has a large hermetic coverage down to $\eta < 5$ as well as a fine lateral segmentation. HCAL is a sampling calorimeter, i.e. the beam position, energy, and arrival time are reconstructed by alternating layers of absorber material, in the case of CMS brass or steel, and a fluorescent scintillator. When hadrons interact with the absorber material, they produce a significant amount of secondary particles. These secondary particles interact with successive layers of the absorber, creating a cascade of particles called a hadronic shower. As the shower evolves, its constituent particles pass through the scintillator material.

The layers of the HCAL are staggered so that there are no gaps in the direct lines through which a particle could escape. The HCAL is organized into Hadron Calorimeter Barrel (HB) and Hadron Calorimeter Outer (HO) sections in the barrel, Hadron Calorimeter Endcap (HE), and Hadron Calorimeter Forward (HF) sections. The hadron calorimeter barrel is confined between the outer radius of the electromagnetic calorimeter, $r = 1.77$ m, and the inner radius of the magnetic coil, $r = 2.95$ m. There are 36 barrel wedges composed of flat brass alloy absorber plates parallel to the beam axis with very high granularity. The innermost and outermost absorber layers are made of stainless steel for structural reasons. There are 17 active plastic scintillator tiles interspersed between the stainless steel and brass absorber plates. Similarly, 36 endcap wedges in the HE measures particle energies with high granularity. The HO and HF were added to improve energy resolution and extend coverage. The HO is placed outside

the solenoid and complements the barrel calorimetry to ensure that no energy in the $|\eta| < 1.26$ range escapes undetected from the back of the HB. The HF is located 11.2 m from the interaction point and covers the region $3 < \eta < 5$. It consists of a steel absorber structure composed of 5 mm thick grooved plates interspersed with quartz fibers that form the active medium of the calorimeter. These fibers detect the energy emitted by the particles via the Cherenkov radiation and, as a consequence, HF is practically insensitive to neutral hadrons but very sensitive to the electromagnetic component of showers.

The expected energy resolution for single-pion interactions in the central part of the calorimeter is, according to the measurements:

$$\frac{\sigma_E}{E} = \frac{94\%}{\sqrt{E}} + 4.5\% \tag{3.10}$$

The expected energy resolution for the very forward calorimeter is worse since it is for the hadronic component and the electromagnetic component, respectively:

$$\frac{\sigma_E}{E_{had}} = \frac{172\%}{\sqrt{E_{had}}} + 9\% \qquad \frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E_{em}}} + 5\% \tag{3.11}$$

3.2.4 The superconducting magnet

The superconducting magnet for CMS [135] is designed to achieve a field of 3.8 T in a free bore of 6 m diameter and 12.5 m length with a stored energy of 2.6 GJ at full current. The flux is returned through a 10000-ton yoke consisting of 5 wheels and 2 end caps, each consisting of three disks. The iron return yoke allows a constant 1.8 T field even outside the magnet.

3.2.5 The muon system

The muon system [136] is central to the physics goals of the CMS. The main goals are a robust muon identification, a precise momentum resolution, a good double muon mass resolution, and the possibility to determine the charge of muons with p_T up to 1 TeV.

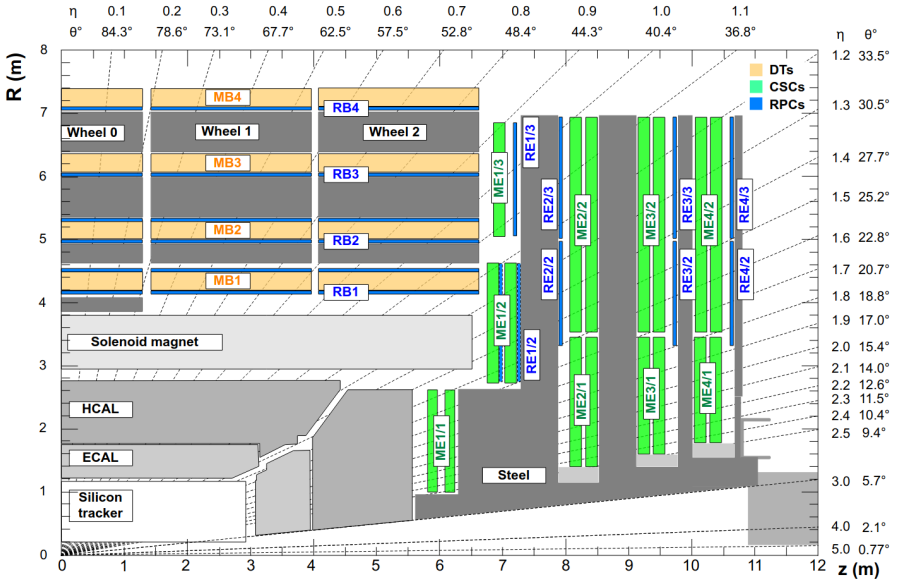


Figure 3.6: View in (r, z) of a quarter of the operational CMS muon system layout during the Run II.

To meet these requirements, the muon system is located inside the magnet return yoke and consists of three types of gas detectors to account for different radiation environments: Drift Tube (DT) chambers in the barrel and Cathode Strip Chambers (CSC) in the end caps, both complemented by a Resistive Plate Chamber (RPC) system (Fig. 3.6). The muon system consists of four layers of muon chambers in the barrel and four in each endcap region, each providing track segments reconstructed from distributed hits. These tracks are combined with information from the tracker to form a complete muon track. This geometric scheme allows covering a pseudorapidity range of up to 2.4. The high return yoke field guarantees good momentum resolution and charge measurement even without the inner tracker information.

Due to the neutron-induced background, the low muon rate, and the uniform magnetic field, the DT system is used in the barrel region within $|\eta| < 1.2$. DTs are long aluminum cells operated with an Ar/CO_2 mixture and with an anode wire in the center to collect ionization charges.

A high number of background events, a high muon flux, and a high non-uniform magnetic field are the factors that determined the choice of CSCs as muon detectors in the two end caps. They cover the pseudorapidity range between 0.9 and 2.4. These chambers are contained in a disk and overlap in φ to avoid holes in the geometric acceptance. The CSCs are trapezoidal multi-wire proportional chambers filled with an $Ar/CO_2/CF_4$ gas mixture. A charged particle passing through the gas mixture creates an avalanche of electrons on a wire, which then generates an electrical signal. Each individual hit has a resolution of about 100 μm in r and 10 mrad in φ.

The RPC system provides a complementary muon system in CMS up to $|\varphi| < 1.6$. The spatial resolution of a single chamber is poor due to the large width of the stripes in which each plane is segmented, but it has excellent time resolution, about 1 ns, which allows unambiguous identification of bundle crossings, essential for triggering purposes. Inside the barrel are six layers of RPCs. Two layers in each of the two inner stations and one layer in each of the two outer stations. The RPCs consist of double-gap bakelite chambers, 2 mm apart, filled with a mixture of $C_2H_2F_4/C_4H_10/SF_6$.

3.2.6 The trigger system

The LHC provides proton-proton and heavy ion collisions at high interaction rates. For protons the beam crossing interval is 25 ns, corresponding to a crossing frequency of 40 MHz. Depending on the luminosity, several collisions occur at each crossing of the proton bunches. Since it is impossible to store and process the large amount of data associated with the resulting large number of events, a drastic reduction in the rate must be achieved.

This task is performed by the trigger system, which is the beginning of the physical event selection process. The rate reduction is done in two steps called the Level-1 (L1) Trigger [137] and the High-Level Trigger (HLT) [138]. The Level-1 Trigger consists of custom-designed, highly programmable electronics, while the HLT is a software system implemented in a filter farm of about a thousand commercial processors. The rate reduction capability is designed to be at least a factor of 10^6 for the combined L1 trigger and HLT. The design output rate limit of the L1 Trigger is 100 kHz, which in practice translates to a calculated maximum output rate of 30 kHz, assuming an approximate safety factor of three.

The L1 trigger uses coarsely segmented data from the calorimeters and the muon system, while the high-resolution data is held in pipelined memory in the front-end electronics.

The HLT has access to the complete readout data and can therefore perform complex calculations, similar to those performed by the offline analysis software if required for particularly interesting events.

For flexibility, the L1 trigger hardware is implemented in FPGA technology where possible, but ASICs and programmable memory lookup tables are also widely used where speed, density,

and radiation resistance requirements are important. A software system, the Trigger Supervisor, controls the configuration and operation of the trigger components. The L1 trigger has local, regional, and global components. At the lower end, the Local Triggers, also called Trigger Primitive Generators (TPG), are based on energy deposits in calorimeter trigger towers and track segments or hit patterns in muon chambers. Regional Triggers combine their information and use pattern logic to determine ranked and sorted trigger objects, such as an electron or muon candidate, in limited spatial regions. The rank is determined as a function of energy or momentum and quality, which reflects the level of confidence assigned to the L1 parameter measurements based on detailed knowledge of the detectors and trigger electronics and the amount of information available. The Global Calorimeter and Global Muon Triggers determine the highest-ranked calorimeter and muon objects in the entire experiment and report them to the Global Trigger, the top entity in the Level 1 hierarchy. The Global Trigger decides whether to reject or accept an event for further evaluation by the HLT. The decision is based on algorithmic calculations and the readiness of the sub-detectors and the DAQ as determined by the Trigger Control System (TCS). The Level 1 Accept (L1A) decision is communicated to the sub-detectors by the Timing, Trigger, and Control (TTC) system. The architecture of the L1 trigger is shown in Fig. 3.7.

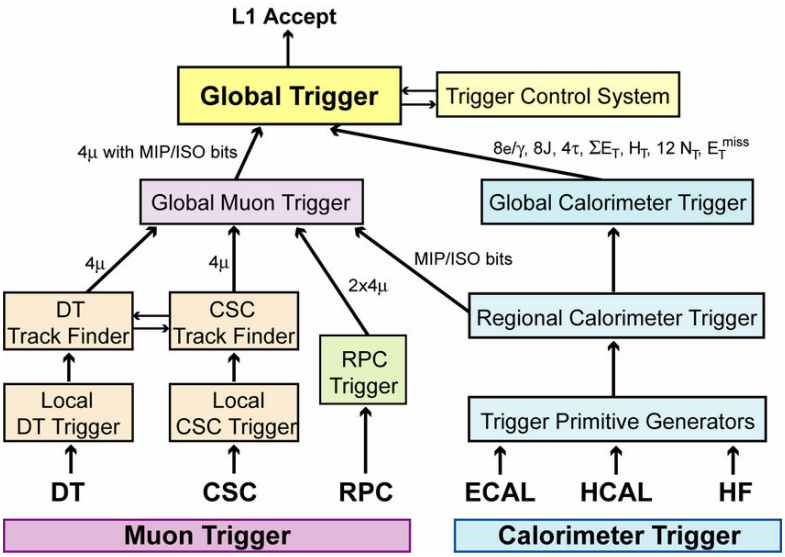


Figure 3.7: Architecture of the Level-1 Trigger.

The L1 trigger must analyze each bundle crossing. The allowed L1 trigger latency between a given bundle crossing and the distribution of the trigger decision to the detector front-end electronics is 3.2 μs. Therefore, the processing must be pipelined to provide virtually dead-time-free operation. The L1 trigger electronics are located partly on the detectors and partly in the underground control room, which is located about 90 m from the experimental cavern.

Chapter 4

Physics object reconstruction, classification and identification

As described in Ch. 3, the CMS detector has cylindrical symmetry along the beam axis, and a particle emerging from the nominal interaction point crosses several subsystems that are sensitive to its passage and can provide information about its nature and properties.The reconstruction strategy exploits the strengths of each sub-detector to obtain a complete and accurate reconstruction and the identification of all particles. The compact structure of the CMS detector layout and the fine granularity of its subsystems are well suited to an event description that correlates the basic elements from all detector layers (tracks and clusters) to identify each final state particle. The combination of the measurements from the tracker, the calorimeters, and the muon system allows a global fit to the physics objects, which can be determined with superior efficiency and resolution. This holistic approach is called particle-flow (PF) reconstruction [139] and benefits from the fine spatial granularity of the CMS detector and its compact nature, reducing uncertainties due to inter-detector propagation.

4.1 Tracks

One way to determine the momentum and charge of a particle is to immerse it in a magnetic field and measure the deviation from a straight trajectory with a position-sensitive detector. To determine the charge and momentum of a particle, the trajectory must be expressed as a function of a finite number of parameters. The set of solutions to the equation of motion is the trajectory model [140] with a dedicated fit. The track fit requires knowledge of the geometrical layout of the detector and the resolution of the sensors that provide the track measurements and their associated errors, a stochastic model of material effects such as multiple scattering and energy loss, and a track model that depends on the magnetic field.

To obtain an accurate and efficient measurement of the charged particle momentum in this environment, CMS uses fast and robust track reconstruction software.

The first step of the reconstruction process is a local reconstruction, which is mainly performed in each sensor. In this step, the signals above specified thresholds are clustered into hits in both pixel and strip channels. A first estimate of the hit position and its uncertainty is given by summing the charge collected in neighboring sensors, corrected for the Lorentz drift due to the CMS magnetic field and a possible noise contribution.

After that, the CMS software reconstructs the tracks by using an adaptation of the combinatorial Kalman filter introduced in Sec. 3.2 to allow pattern recognition and track fitting in the same

framework. It is called the Combinatorial Track Finder (CTF). The collection of reconstructed tracks is produced in several iterations of the CTF. The whole track reconstruction process is called "iterative tracking". The basic idea is to search for the simplest tracks first, discard the hits found in previous iterations, and move on to the next iteration. In this way, the combinatorial complexity of the tracking problem is reduced and simplified for more difficult classes of tracks.

Each iteration can be divided into four distinct steps:

1. The seed generation step provides the initial trajectory parameters and associated uncertainties with a constraint on the trajectory origin. The seed is generated using two main parameters, the seeding layers, which correspond to the detector layers where the hits are searched, and the tracking regions, which limit the possible number of seeds by specifying the bounds on the acceptable seed parameters. Seeds composed of hit pairs are found by combining hits in different layers of the detector. In the tracking step, the KF algorithm is run for each seed. It starts with the coarse estimate of the track parameters provided by the trajectory seed to build track candidates by searching for other hits in adjacent layers of the detector. If the found hits are considered compatible with the trajectory, the algorithm incorporates them into the track and updates the corresponding trajectory parameters. To avoid a rapid increase in the number of candidates, only a limited number of them are retained at each step, with the best candidates being selected. The track-finding step for a given track candidate continues until either there are no more compatible layers, more than one missing hit is found, or its p_T falls below a certain value.

2. After the hits are collected and a first estimate of the track is given, a track fitting step is performed using a filter to obtain the final estimate of the track parameters at the interaction point using the full trajectory information. The fit is performed on the full list of hits, iteratively updating the estimate of the track trajectory, including the hits from the innermost to the outermost. Once this first filter has been run, a second filter is initialized with the results obtained and run back towards the beamline. The weighted average of the track parameters of these two filters is used to find the optimal track.

3. The last step is the track selection. By introducing quality cuts on several track variables, the number of tracks produced by a random combination of hits is reduced. such as the compatibility of the track with respect to the signal peak. The quality cuts are used to set the track quality flags, with the "High Purity" one being used in most of the physics analysis.

The selected tracks are then clustered based on their position relative to the closest approach to the center of the beam spot. This clustering allows the reconstruction of any number of proton-proton interactions in the same LHC bunch crossing. During Run II, the track clustering is performed using a deterministic annealing (DA) algorithm [141], which finds the global minimum for a problem with many degrees of freedom. Among the candidate vertices identified by DA clustering, those containing at least two tracks are then fitted using an adaptive vertex fitter [142] to compute the best estimate of the vertex parameters, the so-called primary vertex, such as its x,y,z position and the indicators of the success of the fit, such as the number of degrees of freedom for the vertex. Each trace in the vertex is assigned a weight between 0 and 1, such that traces that match the position of the reconstructed vertex have a weight close to 1, while those that are far from the vertex have small weights. The number of degrees of freedom in the fit is defined as

$$n_{\text{dof}} = -3 + 2 + \sum_{i=1}^{N_{\text{tracks}}} w_i \tag{4.1}$$

where w_i is the weight of the $i-$th track, and the sum runs over all tracks associated with the vertex. The value of n_{dof} is therefore strongly correlated with the number of tracks compatible

with the origin of the interaction region. For this reason, n_{dof} can also be used to select true proton-proton interactions.

The tracking performance is measured in the data collected during the Run II [143] and it is about 99.9% for tracks associated with muons in the whole muon pseudorapidity acceptance with a very mild dependence on the pile-up conditions (see Fig. 4.1).

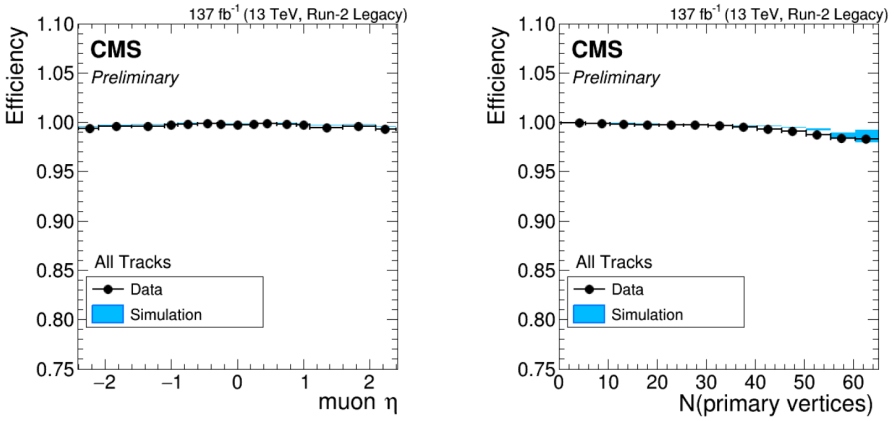


Figure 4.1: The tracking efficiency in Run II dataset as a function of pseudo-rapidity η (left) and the number of primary vertices (right) [143].

In the analyses shown in Ch. 6 and Ch. 7, according to the CMS recommendations, if more than one primary vertex is identified, the one with the largest sum of the squared transverse momenta of the associated tracks is taken as the hard scattering vertex. The main primary vertex must also have at least 4 degrees of freedom and its distance from the nominal interaction point must be less than 2 cm in the $x-y$ plane and 24 cm in the $z-$ direction.

4.2 Muons

The muon reconstruction is not PF specific, since the muon spectrometer allows to identify and reconstruct muon tracks with very high efficiency over the whole detector acceptance [144]. This is mainly because the calorimeters absorb almost all particles except muons and neutrinos. Three different types of muon candidates can be defined, depending on how they are reconstructed:

- standalone muons, whose tracks are seeded by hits from DT and CSC and reconstructed using only the muon spectrometer, gathering all hits from all muon subsystems through a Kalman filter procedure;
- tracker muons, where tracks with $p_{\mathrm{T}} > 0.5$ GeV and a total momentum > 2.5 GeV are propagated to the muon system and required to match with the muon system hits;
- global muons, obtained by matching a standalone muon with information from the tracker and performing a global fit to the resulting track.

Global muons with momentum above 10 GeV, i.e. with hits in at least two muon stations, have the highest reconstruction efficiency. Below this threshold, there is a higher probability that the muons suffer from multiple scattering in the iron of the return yoke. Thus, the tracker muon turns out to be the one with higher efficiency. Charged hadrons can also be reconstructed as muons if their shower is not completely blocked in the calorimeters (punch-through). In the PF algorithm, muons are reconstructed by also using information from the muon energy deposits in ECAL and HCAL, further improving the identification performance.

Additional quality criteria are applied to reconstructed muon tracks to be used for analyses, with multiple working points defined depending on their purpose. In the analyses presented in this work, the following WPs are used:

- Loose ID selection aims to identify prompt muons from the primary vertex and muons from light and heavy flavor decays, as well as to maintain a low rate of misidentification of charged hadrons as muons; a loose muon is a muon selected by the PF algorithm that is also either a tracker or a global muon. The

- Tight ID selection aims to suppress muons from decay in flight and hadronic punch-through. A tight muon is a loose muon with a tracker track that uses hits from at least six layers of the inner tracker, including at least one-pixel hit. The muon must be reconstructed as both a tracker muon and a global muon. The tracker muon must have a segment match in at least two of the muon stations. The global muon fit must have χ^2/dof < 10 and include at least one hit from the muon system. A close muon must be compatible with the primary vertex and have a transverse impact parameter $|d_{xy}| < 0.2$ cm and a longitudinal impact parameter $|d_z| < 0.5$ cm.

Variable	Tight muon requirement	Loose muon requirement
isGlobalMuon & isTrackerMuon	true	true
isPFMuon	true	true
χ^2/dof	< 10	-
Number of valid muon hits	> 0	-
Number of valid pixel hits	> 0	-
Number of matched stations	> 1	-
Number of tracker layers	> 5	-
$\lvert d_{xy}\rvert$	< 0.2 cm	-
$\lvert d_z\rvert$	< 0.5 cm	-

Table 4.1: Muon identification requirements for the Tight and Loose WP. A dash "-" indicates the requirement is not applied [144].

More details about the requirements for the Tight and Loose WP are summarized in Table 4.1.

To select prompt leptons produced in the decays of massive particles such as W and Z bosons, and to reject the large number of leptons produced in jets by the decay of heavy hadrons or the decay in the flight of charged pions and kaons, one must define lepton isolation. The isolation is quantified by estimating the total p_{T} of particles emitted around the direction of the lepton. The particle-based isolation relative to the lepton p_{T} is defined as

$$I_{PF} = \frac{1}{p_{\mathrm{T}}} \left(\sum_{h^{\pm}} p_{\mathrm{T}}^{h^{\pm}} + \sum_{\gamma} p_{\mathrm{T}}^{\gamma} + \sum_{h^0} p_{\mathrm{T}}^{h^0} \right), \tag{4.2}$$

where the sums run over the charged hadrons ($h^{\pm}$), photons (γ), and neutral hadrons (h^0) with a distance ΔR to the lepton less than 0.4 for muons in the (η, ϕ) plane.

To mitigate the degradation of the isolation efficiency due to pile-up (PU), the isolation as defined in Eq. 4.2 is complemented in two ways. First, only charged hadrons associated with the hard scattering vertex are considered. Second, the expected contributions from pile-up are subtracted from the p_{T} sums of neutral hadrons and photons, and the relative isolation is then defined as

$$I_{\mathrm{rel}} = \frac{I^{\mathrm{ch.h}} + \max((I^{\gamma} + I^{\mathrm{n.h}} - I^{\mathrm{PU}}), 0)}{p_{\mathrm{T}}}, \tag{4.3}$$

where $I^{\mathrm{ch.h}}$, I^{γ}, and $I^{\mathrm{n.h}}$ are the sum of the transverse energies deposited by stable particles such as charged hadrons, photons, and neutral hadrons, respectively, in a cone of size $\Delta R = 0.4$ around the muon direction. In this case, the contribution of a pile-up in the isolation cone is removed with the so-called $\Delta\beta$ correction. With this method, $I^{\mathrm{PU}} \equiv \Delta\beta \sum p_T^{PU} \equiv 0.5 \sum p_T^{PU}$ is estimated as the sum of the transverse momenta of the tracks associated with non-leading vertices (i.e., pile-up), which is used to estimate the contribution of neutral particles from pile-up events by applying a multiplicative factor of 0.5, which takes into account the ratio of neutral to charged particles expected from isospin invariance. Thus, the $\Delta\beta$ factor represents the expected neutral contribution in the isolation cone from the observed pile-up charged contribution.

4.3 Electrons

Electrons in the CMS detector are reconstructed by the PF algorithm using measurements from both the tracker and ECAL. The electron passing through the tracker is likely to emit *Bremsstrahlung* photons, which in turn are transformed into electron-positron pairs. For this reason, the basic properties and technical issues to be solved for tracking and the energy deposition patterns of electrons and photons are similar. If the fraction of energy radiated by the electron is small, its track can be reconstructed by the tracker with a good χ^2 and propagated to the inner surface of the ECAL, where it can be accurately matched to the closest cluster. If the radiated energy is large, the same task is much more difficult and the reconstructed track might have very few hits and a large value of χ^2 from the track fit. A preselection based on the number of hits and the χ^2 fit is therefore applied, and the selected tracks are re-fitted with a Gaussian Sum Filter (GSF) [145]. The GSF fit is more suitable for electrons than the Kalman filter used in iterative tracking, as it allows for sudden and substantial energy losses along the trajectory. An electron trajectory candidate is seeded by a GSF trajectory and matched to a cluster in ECAL. After the reconstruction of the electrons, all used tracks and clusters are masked for the reconstruction of other objects.

Electron candidates must meet additional quality criteria to be used in physics analyses. These criteria can be established by examining some critical variables and applying a set of requirements to them, or by using a boosted decision tree (BDT) [146].

The second approach is the one implemented in the studies reported in Ch. 6 and Ch. 7. In this approach, fourteen variables, including those of the cut-based definition, are combined in BDTs that are trained separately in the ECAL barrel and endcap acceptance, and for isolated and non-isolated electrons. The electron BDT is trained on Drell-Yan (DY) plus jets Monte-Carlo simulated samples, with prompt electrons as the signal and unmatched plus non-prompt electrons as the background. Reconstructed electrons from τ decays are ignored. There are two different Multivariate Analysis (MVA) electron ID versions: one with the 3 PF isolation components as input variables (Iso), and one without (NoIso). Both MVA IDs have three categories: the inner barrel ($|\eta| < 0.8$), the outer barrel ($0.8 < |\eta| < 1.44$), and the endcap ($1.57 < |\eta| < 2.5$). For each region, there are low-pt ($5 < p_{\mathrm{T}} < 10$ GeV) and high-pt ($p_{\mathrm{T}} > 10$ GeV) categories, for a total of 6 categories. There are three working points:

- WPLoose with high overall efficiency (98%) for vetoing or multilepton analyses requiring high efficiency,
- WP90 with average efficiency $\sim 90\%$ in each category,
- WP80 with average efficiency $\sim 80\%$ in each category.

The relative isolation for electrons is defined as in Eq. 4.3, but a $\Delta R = 0.3$ cone is used to define it. The contribution from the pile-up in the isolation cone is subtracted by assuming

$I^{\mathrm{PU}} = \rho A_{eff}$, where ρ is the event-specific average pile-up energy density per unit area in the $\phi - \eta$ plane and the A_{eff} is the effective area specific to the given type of isolation.

4.4 Hadronic jets

Once electrons and muons have been reconstructed, the corresponding tracks and calorimetric clusters are removed from the PF object collection. The remaining objects are instead used to reconstruct hadrons from the fragmentation of quarks and gluons, and non-isolated photons. Photons are reconstructed with ECAL clusters not associated with hits in the tracker, if in the tracker acceptance, or not associated with HCAL clusters for higher values of pseudorapidity η. HCAL deposits are not associated with tracks in the tracker or ECAL clusters in the endcaps, instead, they are reconstructed as neutral hadrons. All remaining HCAL clusters associated with one or more tracks in the tracker are reconstructed as charged hadrons.

The energy of the hadrons is not simply the energy of the HCAL cluster, but must be calibrated. The calibrated energy is given by

$$E_{calib} = a + b(E) f(\eta) E_{ECAL} + c(E) g(\eta) E_{HCAL} \tag{4.4}$$

where E_{ECAL} and E_{HCAL} are the measured energies in the calorimeters, and E and η are the true energy and pseudorapidity of the hadron. The coefficient a, measured in GeV, accounts for the energy losses due to the energy thresholds of the clustering algorithm and is treated as independent of E. A large sample of simulated single neutral hadrons (in particular K^0_L) is used to determine the calibration coefficients a, b, and c as well as the functions f and g. The barrel and endcap regions are treated separately to account for different thresholds and cell sizes. In each region, the determination of b and c is performed separately for hadrons leaving energy only in the HCAL (in which case only c is determined) and for those leaving energy in both the ECAL and the HCAL. The calibration is also repeated with real data from dedicated runs.

Some hadrons decay by interacting with the tracker material, producing multiple charged and neutral particles from a secondary interaction vertex. This is a typical situation that occurs in top quark decay, where the b hadrons decay in flight, producing secondary vertices that help in the reconstruction of such hadrons. In these cases, the direction of the primary particle is obtained from the vectorial sum of the momenta of the secondary charged particles, its energy is given by the sum of their energies (denoted E_{sec}), and its mass is set to the charged particle mass. If, in addition, the track associated with the primary particle allows a good measure of the four-momentum, p_{prim}, the energy of the undetected secondary particles can be estimated. The energy of the primarily charged hadron is then estimated as

$$E = E_{sec} + f(\eta, p_{prim}) p_{prim}. \tag{4.5}$$

The small fraction of undetected energy $f(\eta, p_{prim})$ is obtained by simulating single-charged hadron events.

4.4.1 Jet clustering and calibration

The tracks and calorimetric deposits are clustered into objects called jets. Many algorithms can be used to reconstruct the jets [147], but to avoid differences between the data and the Monte-Carlo simulation, the algorithm should be infrared and collinear safe. In the CMS detector, the jets are reconstructed from the PF particles using the anti-k_{T} algorithm [148] as implemented in the FastJet package [149]. In this algorithm two distances are defined: d_{ij} as the distance

between the physical objects (particles, jet candidates) i and j,

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} \tag{4.6}$$

and d_{iB} as the one between object i and the beam (B),

$$d_{iB} = \frac{1}{k_{ti}^{2p}} \tag{4.7}$$

In the previous equations, $\Delta_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ and k_{ti}, η_i and ϕ_i are the transverse momentum, the rapidity, and the azimuth of the particle i, respectively. The radius of the parameter R, which defines the angular opening of the jet cone, can vary, and in CMS the values 0.4 and 0.8 are chosen for jet reconstruction, referred to as AK4 or narrow jets and AK8 or fat jets, respectively. The jet clustering algorithm proceeds iteratively by searching for the smallest of the distances between d_{ij} and d_{iB}:

- if $d_{ij} < d_{iB}$ then the physics objects i and j are combined together,
- if $d_{iB} < d_{ij}$ then the object i is called jet.

The used physics objects are removed, the distances are recalculated, and the process is repeated until no more physics objects are left. The resulting jets are usually called "detector jets".

The pile-up interactions come from other tracks and clusters reconstructed by the PF. After the primary vertex is selected, all charged hadrons whose tracks are associated with a pile-up vertex are removed from the list of particles to be used in the jet reconstruction for that event. This procedure is called pile-up charged hadron subtraction and is referred to as CHS [150]. Other pile-up mitigation techniques are also used, such as the Pile-Up Per Particle Identification (PUPPI) [151].

Further requirements are applied to nozzles used in physics analyses [152]. In particular, for the AK4 CHS jets, a tight ID can be defined with an efficiency of $98 - 99\%$ in all η regions. The details of the tight ID requirements are reported in Tab. 4.2

To further improve jet identification in high pile-up environments, a technique known as Pileup JetID [153] is developed and used in CMS publications to reduce the amount of pile-up jets in the AK4 CHS jet collections, especially for jets with $p_{\mathrm{T}} < 50$ GeV. The identification of pile-up jets relies on three types of jet properties:

- within the tracker assumption, the trajectories of the tracks associated with the jets can be used to determine the compatibility of the jet with the primary interaction vertex;
- the topology of the jet shape can be used to disentangle jets resulting from the overlap of multiple interactions from truly hard jets;
- the multiplicity of the object can be used as an additional handle.

Using all the variables related to the above properties, a BDT is trained and three different WPs are defined: Loose, Medium, and Tight, with 99%, 90%, and 80% efficiency for prompt jets, i.e. not induced by pile-up.

The measured energy of the jets must be calibrated to relate it to the true energy of the particles originating in the jet. A true "particle level" jet results from the clustering of all stable particles from the fragmenting parton and the particles from the underlying event (UE) activity, using the same algorithm just illustrated for the detector jets, but performed with Monte-Carlo simulated jets. The correction is applied as a multiplicative factor to each particle in the jet:

$$p_{\mathrm{T}}^{corr} = \mathscr{C} \cdot p_{\mathrm{T}}^{raw}, \tag{4.8}$$

Variable	Requirement in 2016	Requirement in 2017	Requirement in 2018
Requirements for jet $\lvert\eta\rvert \leq 2.6$			
Neutral hadron fraction	< 0.90	< 0.90	< 0.90
Neutral EM fraction	< 0.90	< 0.90	< 0.90
Number of constituent	> 1	> 1	> 1
In addition for jet $\lvert\eta\rvert \leq 2.4$			
Charged hadron fraction	> 0	> 0	> 0 (up to $\lvert\eta\rvert = 2.6$)
Charged multiplicity	> 0	> 0	> 0 (up to $\lvert\eta\rvert = 2.6$)
Charged EM fraction	< 0.99	-	-
Requirements for jet $2.6 < \lvert\eta\rvert \leq 2.7$			
Neutral hadron fraction	-	-	< 0.90
Neutral EM fraction	-	-	< 0.99
Charged multiplicity	-	-	> 0
Requirements for jet $2.7 < \lvert\eta\rvert \leq 3.0$			
Neutral EM fraction	> 0.01	$[0.02, 0.99]$	$[0.02, 0.99]$
Neutral hadron fraction	< 0.98	-	-
Neutral multiplicity	> 2	> 2	> 2
Requirements for jet $\lvert\eta\rvert > 3.0$			
Neutral EM fraction	< 0.90	< 0.90	< 0.90
Neutral hadron fraction	-	> 0.02	> 0.02
Neutral multiplicity	> 10	> 10	> 10

Table 4.2: Requirements needed in the different years of the Run II data taking for the tight jet ID definition.

where

$$\mathscr{C} = \mathscr{C}_{offset}(p_{\mathrm{T}}^{raw})\mathscr{C}_{MC}(p'_{\mathrm{T}}, \eta)\mathscr{C}_{rel}(\eta)\mathscr{C}_{abs}(p_{\mathrm{T}}\text{"}). \tag{4.9}$$

The offset correction $\mathscr{C}_{offset}$ removes the extra energy due to noise and pile-up, and the Monte-Carlo correction $\mathscr{C}_{MC}$ removes most of the non-uniformity in η and the non-linearity in p_{T}. Finally, the residual corrections $\mathscr{C}_{rel}$ and $\mathscr{C}_{abs}$ provide corrections for the relative and absolute energy scales, accounting for the small differences between data and simulation. The corrections are applied sequentially, and p'_{T} is the transverse momentum of the jet after the offset correction is applied, and p_{T}" is the p_{T} of the jet after all previous corrections. These corrections are commonly referred to as Jet Energy Corrections and Jet Energy Resolutions, or JECs and JERs for short [154].

4.4.2 $b-$jets

A large part of the LHC physics program relies on the identification of jets containing the decay of a b hadron, also called $b-$jets. The $b-$jets can be distinguished from jets produced by the hadronization of light quarks and gluons by the following properties of b hadrons:

- long lifetime, $\tau \approx 1.5$ ps, corresponding to a distance $c\tau \approx 450$ μm, or equal to $\gamma\beta c\tau \approx 1.8$ mm at 20 GeV in the $b-$hadron rest frame;
- large mass ~ 5 GeV;
- high track multiplicity;
- large semileptonic branching fraction, being $\sim 20\%$ each for electrons and muons;

- hard fragmentation function, since a large fraction of the original b quark momentum is carried by the b hadron.

The CMS detector, with its excellent tracking system, robust lepton identification, and finely segmented calorimeters, is well suited to the task of identifying b jets.

The so-called $b-$tagging is a reconstruction technique that takes advantage of the above b hadron properties and assigns to each jet a probability that it contains a b hadron. The $b-$tagging can be based on combined information about the track, the secondary vertex, and the soft lepton produced by the b hadron decay.

Several $b-$tagging algorithms of different complexity and based on different input information have been developed by CMS [155, 156]. A common feature of all these algorithms is that they produce as output a numerical discriminator, and the higher the discriminator value, the more likely it is that the jet is a true $b-$jet.

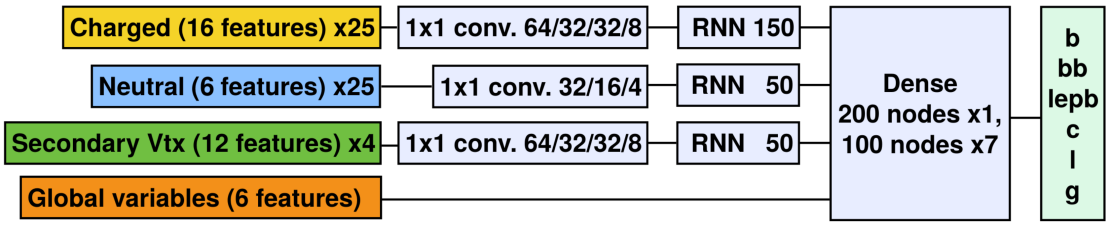


Figure 4.2: An illustration of the DeepJet architecture [157].

In this work, the DeepFlavour algorithm is used among the several $b-$tagging algorithms developed in the CMS Collaboration [157]. It exploits a large number of tracks within each jet and, for the first time, includes neutral candidates. The algorithm directly uses the PF jet constituents as clustered by the Jet algorithm, sorted by decreasing importance of impact parameters, distance from the secondary vertex, and transverse momentum. The algorithm uses a deep neural network architecture with a set of convolutional layers to learn a compressed representation for the input charged and neutral particles and secondary vertices separately, followed by three recurrent layers that summarize the input from each set of candidates. Such compressed information is passed to a set of fully connected layers that ultimately provide the a-posteriori probability that the jet contains one b hadron, multiple b hadrons, at least one c hadron, or a hadron derived from a light quark or gluon (see Fig. 4.2).

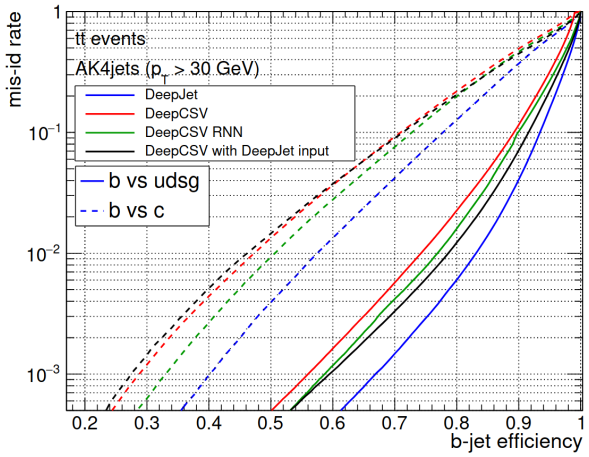


Figure 4.3: Misidentification probability for c and light-flavour jets versus $b-$jet identification efficiency for DeepFlavour and previous $b-$tagging algorithms in $t\bar{t}$ events [157].

The jet is $b-$tagged if the discriminator value is above a certain threshold, often called the cut value, and the efficiency for a given flavor is defined as the number of jets of that flavor that have a discriminator value above that cut divided by the total number of jets of the same flavor. Various $b-$tagging algorithms developed by the CMS experiment can identify $b-$jets with typical efficiencies between 40% and 85% while keeping the rate of misidentified light flavor jets between 0.1% and 10%. A comparison of $b-$tagging efficiency and misidentification as a light jet between different algorithms used in CMS is shown in Fig. 4.3.

4.5 Hadronic taus

Since tau leptons decay most frequently to hadrons and neutrinos, all analyses interested in the final states of τ require the efficient reconstruction and identification of hadronic tau lepton decays, collectively named in the following τ_h or hadronic tau.

Hadronic taus are reconstructed in one of their decay modes with the "hadrons-plus-strips" (HPS) algorithm [158], which reconstructs their decay mode and the visible four-momentum, i.e. the four-momentum of all decay products except the neutrinos. The τ_h candidates are reconstructed in one-prong decays with and without π^0s, three-prong decays with and without π^0s, and three-prong τ_h decays in which one of the charged hadrons is either not reconstructed as a charged hadron or not even its track is recovered.

The reconstruction is done in four steps as follows. First, seed regions are defined by reconstructing one τ_h candidate per seed region. Each seed region is defined by a reconstructed hadronic jet. The jets used for seeding are the narrow ones defined in Sec. 4.4, and all particles in a cone of radius $\Delta R = 0.5$ around the jet are considered for the next steps of τ_h reconstruction. Second, π^0 candidates are reconstructed using "strips" in $\eta - \varphi$ space where the four moments of electrons and photons are added, and charged hadron $h^{\pm}$ candidates are selected using charged particles from the PF algorithm as input. Third, all possible τ_h candidates in different decay modes are reconstructed from the reconstructed charged hadrons and striations. The τ_h four-momentum is obtained by summing the four-momenta of the charged hadrons and strips used to reconstruct the τ_h candidate in a given decay mode. There are seven different reconstructed τ_h decay modes. An overview of the τ decay modes and their branching fractions is given in Tab. 4.3.

Decay mode		Resonance	Branching fraction (%)
Leptonic decays	$\tau^- \to e^- \overline{\nu}_e \nu_\tau$		17.8
	$\tau^- \to \mu^- \overline{\nu}_\mu \nu_\tau$		17.4
Hadronic decays	$\tau^- \to h^- \nu_\tau$		11.5
	$\tau^- \to h^- \pi^0 \nu_\tau$	$\rho(770)$	25.9
	$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	$a_1(1260)$	9.5
	$\tau^- \to h^- h^+ h^- \nu_\tau$	$a_1(1260)$	9.8
	$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$		4.8
Other			3.3

Table 4.3: Decays of τ leptons and their branching fractions in %. The known intermediate resonances of all listed hadrons are given where appropriate. Although only τ^- decays are shown, the decays and values of the branching fractions are identical for charge-conjugate decays.

Fourth, a single τ_h candidate is selected from all possible reconstructed τ_h candidates within a seed region. Reconstructed τ_h candidates are subject to the following constraints:

- the mass of the reconstructed τ_h must be compatible with the collected hadronic decay products;
- the τ_h charge must be equal to ± 1 unless the τ_h candidate is reconstructed in a mode with a missed charged hadron, in which case the charge is set to be equal to the charge of the charged hadron with the highest p_T; all reconstructed $h^{\pm}$ and π^0 must lie in the signal cone defined with radius $\Delta R = 3.0/p_T$, with $0.05 \leq \Delta R \leq 0.1$ with respect to the τ_h momentum.

Among the selected τ_h candidates, the one with the highest p_T is chosen.The overall reconstruction efficiency is mostly limited by the ability to reconstruct tracks from charged hadrons of about 90%. For the decay modes without missing charged hadrons, the charge assignment is 99% correct for an average $Z \to \tau\tau$ event sample, 98% for τ_h with $p_T \approx 200$ GeV, and 92% for τ_h with $p_T \approx 1$ TeV. Finally, similar to what is done for hadronic jets, the reconstructed hadronic τs in Monte-Carlo simulations require additional energy corrections to be applied before they can be used in physics studies. These corrections are called Tau Energy Scale (TES) when the reconstructed τs are generated as real hadronic τs, and Fake Energy Scale (FES) when they are generated as real electrons.

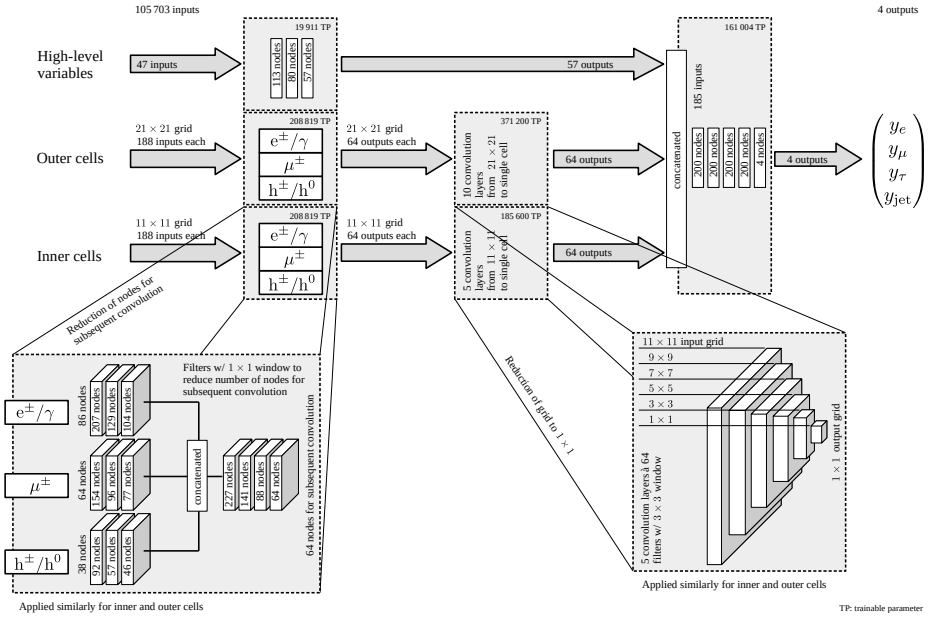


Figure 4.4: The DNN architecture of the DeepTau algorithm. The sets of input variables are first processed separately through different subnetworks, whose outputs are then concatenated and processed through five fully connected layers before the output is computed, which gives the probabilities for a candidate to be either a τ_h, an electron, a muon, or a hadronic.

The identification of τ_h is performed with the "DeepTau" algorithm [158], which is based on a multi-classifier DNN that simultaneously discriminates against jets, electrons, and muons. The DNN uses a combination of high-level inputs and information from all reconstructed particles in the vicinity of the τ_h candidate. The information from all reconstructed particles near the τ_h axis is processed with convolutional layers in $\eta - \varphi$ space,because the superior performance of such a deep learning algorithm leads to an increase in signal efficiency for a given target background rate of jets, electrons, or muons misidentified as τ_h, which typically translates directly into improved sensitivity or precision of the analysis.

The overall DNN architecture is shown in Fig. 4.4. The goal of the DNN is ultimately to use and process the inputs to optimally classify the τ_h candidate as belonging to a target class, which is equivalent to determining whether the reconstructed τ_h comes from a real τ_h, a muon,

an electron, or a hadronic jet. Working points for the DeepTau discriminators are defined to guide their use in physics analyses and to derive appropriate data-to-simulation corrections. The target τ_h identification efficiencies range from 40% to 98% for the jet discriminator, from 99.5% to 99.95% for the muon discriminator, and from 60% to 99.5% for the electron discriminator. In Fig. 4.5, the performance of the discriminator against jets is evaluated by studying the efficiency for hadronic jets and real τ_h to pass the discriminator and comparing it with the classifier that was used in previous CMS analyses[158], which was based the output of a multivariate (MVA) algorithm as well. The efficiency is evaluated separately for the $W+$jets and $t\bar{t}$ event samples since the $W+$jets sample is enriched in quark jets, while the $t\bar{t}$ event sample has a larger fraction of b quark and gluon jets and a busier event topology. The performance of the discriminator against electrons is shown in Fig. 4.6, compared to the previous MVA discriminators. Finally, the performances of the discrimination of τ_h candidates against muons are shown in Fig. 4.7.

In conclusion, the DeepTau discriminators against electrons, muons, and hadronic jets are recommended for all CMS analyses, since they lead to large gains in discrimination performance against various backgrounds when real τ_h are included in the desired final state.

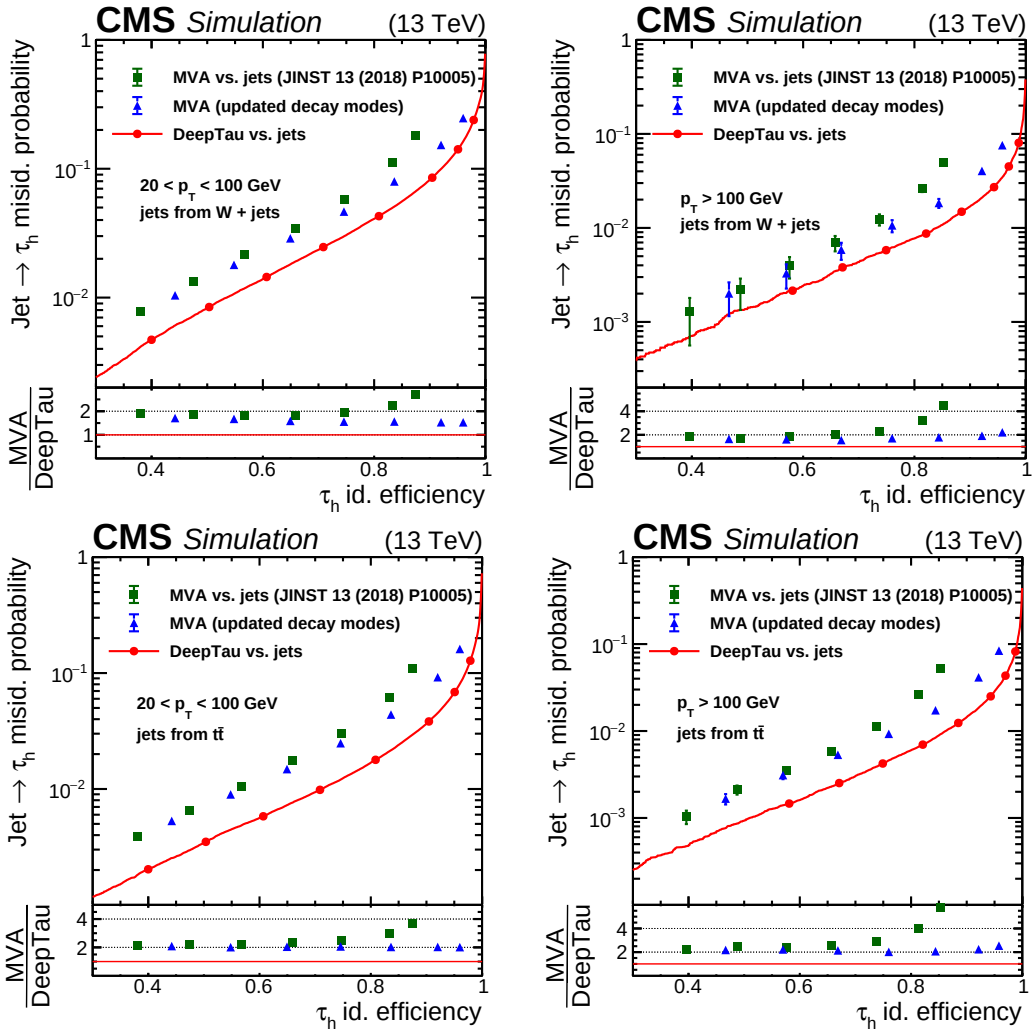


Figure 4.5: Efficiency for hadronic jets to pass various τ_h identification discriminators compared to the efficiency for genuine τ_h. The top two plots are obtained with jets from the $W+$jets simulated sample and the bottom two plots with jets from the $t\bar{t}$ sample. The left two plots include jets and real τ_h with $p_T < 100$ GeV, while the right two plots include those with $p_T > 100$ GeV. Working points are shown as full circles.

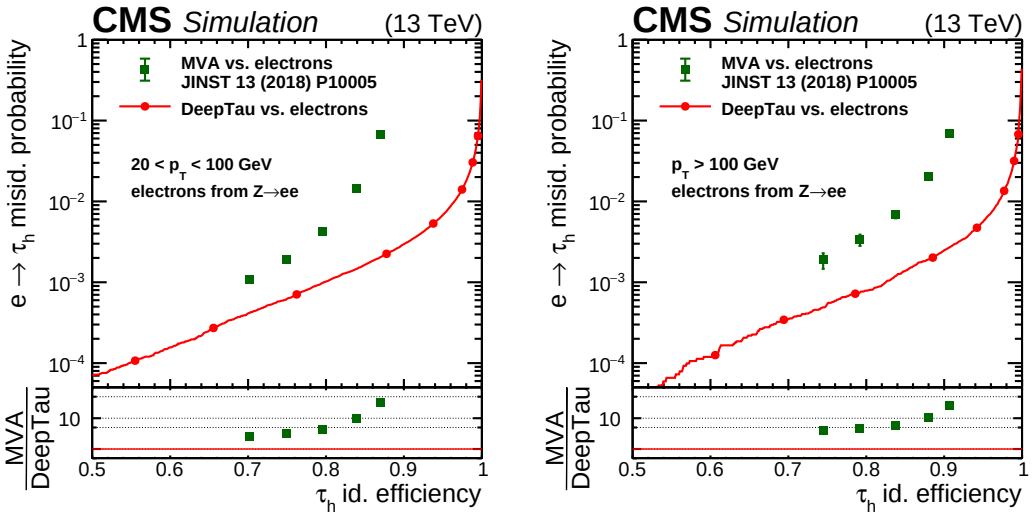


Figure 4.6: Efficiency for electrons versus efficiency for real τ_h to pass the MVA and DeepTau versus electrons discriminators, separately for electrons and τ_h with $20 < p_T < 100$ GeV (left) and $p_T > 100$ GeV (right). The vertical bars correspond to the statistical uncertainties. The τ_h candidates are reconstructed in one of the τ_h decay modes without missing charged hadrons.

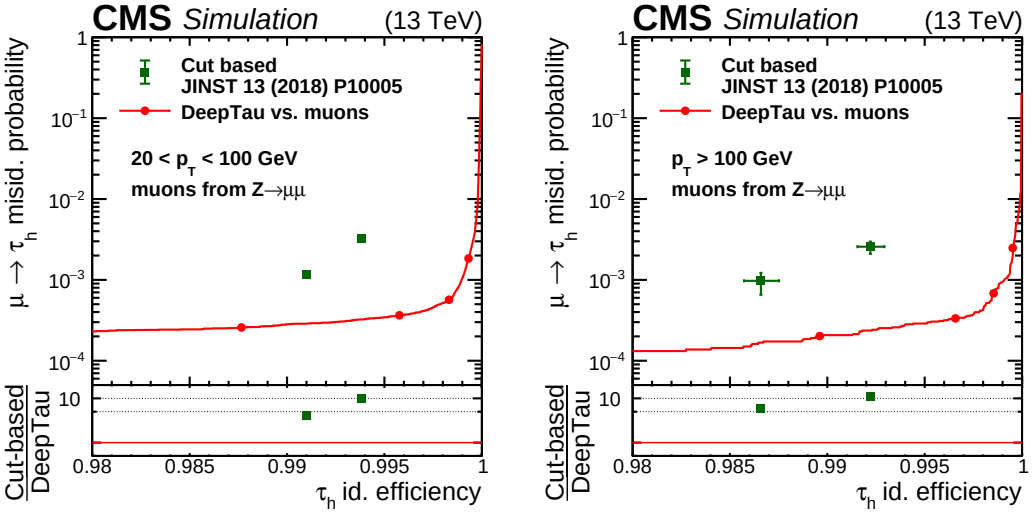


Figure 4.7: Efficiency for muons versus efficiency for simulated τ_h to pass the cutoff-based and DeepTau vs. muons discriminators, separately for muons and τ_h with $20 < p_T < 100$ GeV (left) and $p_T > 100$ GeV (right). The four working points are shown as full circles.

4.6 Missing transverse momentum

The particles that do not interact with the detector, i.e. the neutrinos, are reconstructed from the momentum imbalance in the transverse plane. The PF-based $\vec{p}_{\mathrm{T}}^{miss}$ is the opposite of the vectorial sum of the transverse momenta of the identified PF particles:

$$\vec{p}_{\mathrm{T}}^{miss}(raw) = -\sum_{i}^{\text{all PF}} \vec{p}_i. \tag{4.10}$$

To be validated as good for physics analysis, each event must pass the following noise filters:

- Beam halo filter: machine-induced particles flying with the beam, with a large radius (up to 5 m), produced by beam-gas/beam-pipe interactions are occasionally detected; high-energy halo muons have a non-negligible probability of interacting in the calorimeters, producing clusters of up to several hundred GeV.
- HBHE and HBHEIso Noise Filter: The HCAL in both Barrel and Endcaps (HBHE) is known to record sporadic anomalous signals (noise) at a fixed rate independent of beam conditions, which are targeted by this filter.
- ECAL TP filter: removes events where an ECAL dead cell trigger primitive is hit.
- Bad PF muon filter: this filter removes events where the muon is reconstructed as a PF muon, but its reconstruction quality is low and it has a large p_{T}, leading to an artificially large missing transverse momentum.

The $\vec{p}_{\mathrm{T}}^{miss}$ receives corrections by propagating the vectorial difference between uncalibrated and calibrated jets to it. The correction can be expressed as:

$$\vec{p}_{\mathrm{T}}^{miss}(corr.) = -\sum_{i}^{\text{all PF}} \vec{p}_i - \sum_{j}^{\text{all jets}} (\vec{p}_j^{\,corr} - \vec{p}_j), \tag{4.11}$$

which yields an improved $\vec{p}_{\mathrm{T}}^{miss}$ energy scale and resolution [159].

Chapter 5

Event simulation and systematic effects

In particle physics, the result of the occurrence of a fundamental interaction, like a proton-proton collision in the case of LHC, or the isolated decay of a particle, is called an “event”. At the most basic level, an event thus consists of multiple outgoing particles, as might be recorded in a snapshot taken by an idealized detector, with conservation laws implying that the total summed energies and momenta of the final-state particles should match those of the initial state, as should any discrete quantum numbers conserved by the physics processes in question.

Due to the randomness of quantum processes, the number of outgoing particles and their properties varies from event to event. The probability distributions for these properties can be inferred by studying an ensemble of events in the data. Conversely, given a set of theoretically calculated probability distributions, it is possible to generate ensembles of simulated events for comparison with data. A numerical algorithm that can produce or “generate” random sequences of such simulated events, one after another, is called an “event generator”. The simulations can be based on known or hypothetical laws of nature. This allows the exploration and comparison of competing paradigms, and studies of the sensitivity of proposed physical observables.

A high-energy physics event can be thought of as a list of the subatomic particles produced in a collision, along with a measure of the probability that this event will occur, representing a kind of event record, including history information showing intermediate stages of the event modeling. The measure of the relative probability of a given event within a sample is given by the weight of that event relative to the total sum of weights. The total cross-section for the sample is also computed, allowing probabilities to be converted to cross-sections.

Although the starting point is often a relatively simple cross-section computed at first order in perturbation theory, the total probability distribution for simulated events, fully differential in all relevant phase-space variables and quantum numbers of the produced set of final-state particles, typically cannot be expressed analytically. Instead, it is evaluated directly using numerical methods, with Markov Chain Monte-Carlo (MCMC) algorithms based on pseudo-random number generators as the main ingredient [160, 161].

Event generator predictions are crucial because they serve not only to predict the observable final state at the detector level but also to infer the modeling of each involved process before any interaction with the detector so that the generated events effectively act as an ensemble theory prediction. As such, it can be used to study the consequences of new and old phenomena and to investigate the loss, incorrect measurement, and misidentification of particles in experiments. It is thus an important tool for the interpretation of collider data. Event generators are automated via algorithms implemented in dedicated software packages, and nowadays most of the larger

projects are developed in the programming languages C++ and Python.

Simulating particle production in high-energy collisions over the full range of experimentally accessible energy scales in as much detail as possible, taking into account hadron collisions and hadron production, is extremely complex, and there is currently no comprehensive theory capable of predicting event properties over this full range. For practical purposes, the wide range of phenomena is broken down into several components. A natural division for these components is a time or, equivalently, energy or transverse momentum ordering, with the best-understood physics computed at the shortest time scales and highest energies (hard processes), and the least understood physics modeled at the longest time scales and lowest energies (soft processes). This division is well-motivated and often supported by factorization theorems, but it is not completely unambiguous and sometimes open to correction.

An explicit scheme of a fully simulated $pp \to t\bar{t}$ event is given in Fig. 5.1. The radial coordinate illustrates hardness scales, starting with the hardest subprocess near the center, labeled $\mathrm{d}\hat{\sigma}^0$, and ending with stable final state particles and the incoming beam particles at the periphery. The components of an energetic proton-proton collision are:

1. A hard scattering of two partons, one from each incoming hadron, into a few outgoing particles. The initial partons are selected using parton distribution functions (PDFs) for the incoming hadrons, and the kinematics of the outgoing particles are based on matrix elements calculated in perturbation theory. Such calculations introduce the factorization and renormalization scales. Partons with momenta below these scales are not included in hard scattering but are introduced by other stages of event generation. Hard scattering predictions depend on a few universal input parameters determined from data, such as the value of the strong coupling α_S at the Z boson mass and PDFs.

2. The hard process may produce short-lived resonances, such as Z or W gauge bosons or top quarks, whose decay to normal particles must be considered in close connection with the hard process itself.

3. Fixed-order radiative corrections can be incorporated by a combination of matrix-element corrections, matching, and/or merging strategies. In Fig. 5.1, the purple-shaded region surrounding the hard process represents the range of scales covered by a generic matrix-element merging strategy that is active above a given minimum $p_\perp$ scale.

4. Initial-State Radiation (ISR) of additional particles from the scattering initiators using numerical resummation of soft and collinear gluon emission. This, together with its final-state equivalent reported below, is commonly referred to as the parton shower.

5. Final State Radiation (FSR) of additional particles from the hard scattering itself and also from any resonance decays.

6. In competition with ISR and FSR, further scattering processes between additional partons from the incoming beams may take place in a phenomenon known as Multiple Parton Interactions (MPI), not to be confused with the pile-up in hadron-hadron collisions.

7. At some point after the MPIs, and perhaps before the resonance decays, strings begin to form as the non-perturbative limit of color dipoles. However, these dipoles are typically defined by color bonds that are assigned in the $N_c \to \infty$ limit and are not unique for $N_c = 3$. The associated color space ambiguities can be modeled by color reconnection (CR). It is also possible that long-range dynamical interactions could physically alter the color flow and/or change the configuration of the expanding strings before they fragment.

8. The strong interaction now leads to the confinement of QCD partons into color-singlet subsystems known as strings or, in the case of small masses, clusters. What is left of

the incoming hadronic constituents is combined into beam remnants. In Fig. 5.1, the transition between the partonic and hadronic stages of event generation is highlighted by the concentric rings shaded in blue.

9. The strings fragment into hadrons based on the Lund string model [162, 163]. Optionally, the effects of overlapping strings can be taken into account, e.g. by collecting them in so-called “ropes” and/or allowing interactions between them.

10. Identical particles that are close in phase space may exhibit Bose-Einstein enhancements (Fermi-Dirac suppressions) for integer-spin particles (for half-integer-spin particles).

11. Unstable hadrons produced in the fragmentation process decay into other particles until only stable particles remain, with some flexibility for the user to define what is stable.

12. In densely populated regions of phase space, the produced particles may rescatter, reannihilate, and/or recombine.

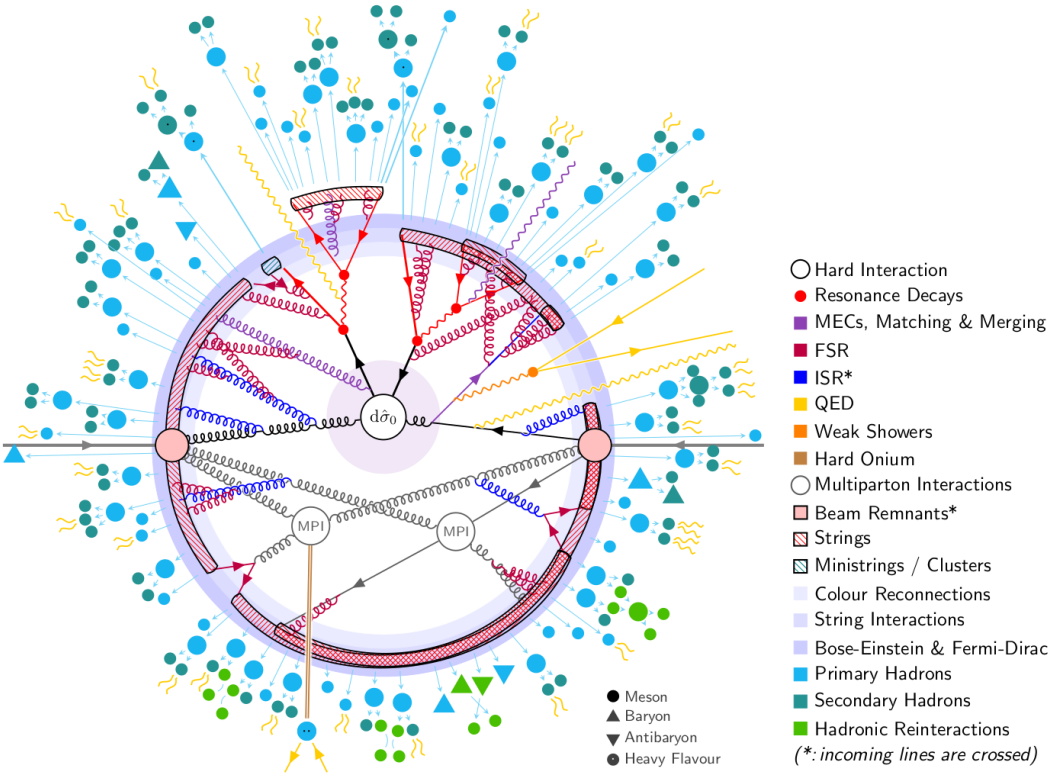


Figure 5.1: Schematic of the structure of a $pp \to t\bar{t}$ event, as modeled by the currently most used event generators.

Particular attention was paid to the preparation, production, and validation of the Monte-Carlo simulated samples necessary for both studies. Part of the work of this thesis was therefore dedicated to giving a direct contribution to the efforts of the CMS Collaboration in performing the simulations of the processes with two vector bosons and the addition of the SMEFT operators effect in them has been offered. This as well as other contributions to the CMS Monte-Carlo production were realized in the context of an official responsibility within the group as the so-called “Monte-Carlo Contact person”, in charge of realizing simulations for the CMS Physics Analysis Group for Standard Model Physics. For this task, the MADGRAPH [54] package for step 1 of the above list and the PYTHIA [164] package for the later steps were used. For this reason, the rest of this Chapter will describe these two event generators in more detail. These

two generators, together with POWHEG [56] for step 1, represent the current state of the art in event generators and are the most widely used in the CMS Collaboration.

The limited knowledge of the theoretical and experimental parameters given as input to the simulation, the assumptions of each of the steps, and the changing data-taking conditions can result in discrepancies between prediction and data. Such differences can manifest in and be relevant for the efficiency of the reconstruction and identification of the physical objects. As a consequence, appropriate corrections are applied to them to reliably reproduce the behavior in the data and to use the Monte-Carlo simulations in physical analyses. In Sec. 5.5 all data-to-MC corrections implemented in both the analyses reported in Ch. 6 and Ch. 7 are presented. In addition, as mentioned before the overall modeling based on event generations suffers from irreducible uncertainties arising from all the theoretical parameters used, which affect the overall uncertainty of the physical measurements. These uncertainties are referred to as "systematics", and the ones relevant to this work are collected in Sec. 5.6.

5.1 Hard scattering process simulation: Madgraph

MADGRAPH [165] is a tool for automatically generating matrix elements for high-energy physics processes such as decays and $2 \to n$ scattering. The central idea of MADGRAPH is that the structure of a cross-section, regardless of the theory under consideration and of the perturbative order, is essentially independent of the process, and as such, it can be written in a computer code once and for all.

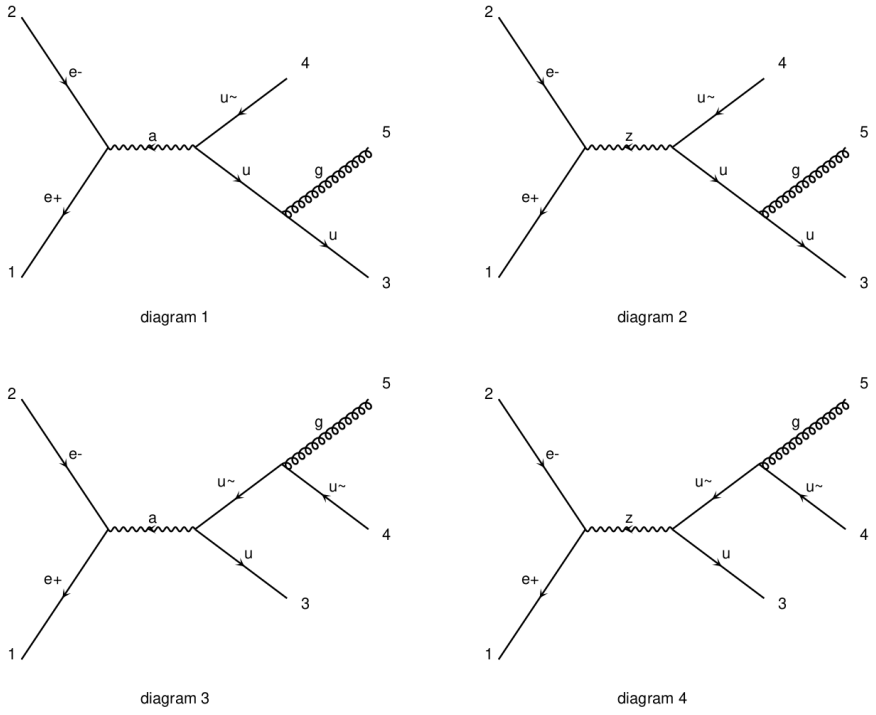


Figure 5.2: Diagrams for the process $e^+e^- \to u\overline{u}g$.

MADGRAPH5, the version used in the CMS Collaboration to generate simulations for the Run-II dataset studies, is written in Python. As a result, this new implementation greatly facilitates the selective use of modules and the addition of new features. While MADGRAPH5 performs all the necessary leading order (LO) calculations, its extension MADGRAPH5_AMC@NLO extends the accuracy of the predictions contained in the simulations to the next-to-leading order

(NLO) [54]. It requires a theoretical model and a set of process-independent building blocks, where the first is equivalent to the Lagrangian of the theory plus its parameters, such as couplings and masses. As a result, MADGRAPH generates all the Feynman diagrams for the process necessary to evaluate the matrix element at a given point in phase space [166]. Such calculations are then used to predict physical observables with different perturbative accuracies and final state descriptions. They are listed below in order of increasing complexity:

1. fLO: this is a tree- and parton-level calculation, where the exponents of the coupling constants are the smallest for which a scattering amplitude is non-zero. No shower is involved, and the observables are reconstructed using the very particles that appear in the matrix elements.

2. fNLO: the same as fLO, except that the perturbative accuracy is NLO, including both tree-level and one-loop matrix elements.

3. LO+PS: uses the matrix elements of an fLO calculation but fits them to parton showers (usually done with PYTHIA). Therefore, the observables must be reconstructed using the particles resulting from the Monte-Carlo simulation.

4. NLO+PS: Same as LO+PS, except that the underlying calculation is an NLO one instead of an LO one.

5. MLM-merged: combines multiple LO+PS samples that differ in final state multiplicities.

6. FxFx-merged: combines multiple NLO+PS samples that differ by final-state multiplicities.

Tree-level amplitude computation is a crucial aspect of matrix-level physics simulations since all matrix elements used in LO and NLO calculations are effectively constructed using tree-level techniques [167, 168]. In particular, the algorithm used in MADGRAPH to determine the Feynman diagrams takes advantage of the fact that it is possible to "factorize" diagrams efficiently, so that if a particular substructure appears in several diagrams, it only needs to be calculated once, speeding up the calculation considerably.

Generating large samples of events for experimental analysis can be a very time-consuming operation, especially if it involves a full simulation of the detector response. Therefore, it is convenient to apply corrections or study the theoretical uncertainties and different modelizations by using reweighting techniques whenever possible. In MADGRAPH, reweighting is implemented on an event-by-event basis, as this is the most accurate way, at least in the limit of a large number of events. Starting with a set of hard events, their weights can be re-scaled as follows:

$$w_{new} = r w_{old} \tag{5.1}$$

without modifying their kinematics.

The practical implementation of this equation depends on the type of problem one wishes to solve. For the studies presented in this thesis, the most compelling is the case where one is interested in studying the effects of changing the modeling of a process, with the only constraint that its initial and final states be the same. Therefore, one can simply write:

$$r = \frac{|\mathscr{A}_{new}|^2}{|\mathscr{A}_{old}|^2} \tag{5.2}$$

where $|\mathscr{A}_{new}|$ and $|\mathscr{A}_{old}|$ are the modified and original scattering amplitudes, respectively. In particular, with this approach, it is possible to evaluate the effect of an arbitrary number of EFT operators using only a sample of simulated events and applying the reweighting factor r.

5.2 Showering and hadronization: Pythia

The PYTHIA package can be applied to a wide range of phenomenological problems in particle physics and related problems in astroparticle, nuclear, and neutrino physics [164]. Historically, the core of PYTHIA is the Lund string model of hadronization, based on a perturbative approach. This model is most appropriate when the invariant masses of the hadronizing systems are above 10 GeV or so. For systems with lower masses, the model is less reliable because the perturbative framework for the QCD computations breaks down.

Internally, the latest version of PYTHIA is structurally divided into three main parts: process level, parton level, and hadron level. The process level represents the hard scattering process, including the production of short-lived resonances. The hard process is typically described perturbatively, with a limited number of particles, typically at high energy scales. The parton level includes initial and final state radiation, where different shower models are available. Multiparton interactions are also included at this stage, along with the treatment of beam remnants and the possibility of color reconnection. At the end of the parton level, the event represents a realistic partonic structure, including jets and the description of the underlying event. The hadronic level then takes care of the QCD confinement of the partons into color singlet systems. In PYTHIA 8.3, the hadronization is described by the fragmentation of QCD strings into hadrons. In addition, other aspects such as the decay of unstable hadrons and hadron rescattering are treated at the hadron level. The physical models of hadronization are typically non-perturbative and thus require modeling and parameter tuning. The output of the hadron level is then a realistic event as it can be observed in a detector.

In addition to this general structure, a significant number of shared objects and crosstalk are passed between these levels: PDFs are relevant in both the process level and the ISR, and the matching and merging engine works at the interface between parton showers and the process level, and the “Info” object is used in all levels to store and access central information.

Currently, in the CMS Collaboration, PYTHIA in version 8.3 is used on top of external matrix element codes providing the hard interactions, e.g. via the Les Houches interfaces, to add parton showers, underlying events, particle decays, and hadronization. Some details of the parton and hadron levels are given below.

5.2.1 Parton shower

The energetic pp collisions at the LHC may present five to ten separated jets, each with a very complex substructure associated with the perturbative production of increasingly close patrons. Such a fractal nature is expected to continue down to the hadronization scale, just below 1 GeV. At this scale, the event may contain up to a hundred partons, even if the full partonic structure is masked by the subsequent non-perturbative hadronization process. It is not feasible to perform matrix element calculations to describe such complicated event topologies. The standard approach is therefore to start from a matrix element calculation with only a few well-separated partons and then apply a parton shower.

Parton showers attempt to describe how a basic hard process is dressed up by emissions at successive resolution scales to give an approximate but realistic picture of the structure of all possible partonic states. Such a shower is constructed recursively from the large scale of the hard process down to a lower cutoff at about the hadronization scale. They are present in QCD, QED, and the EW theory above the symmetry-breaking scale, and are based on a generic scheme that can be applied to any process of interest and is derived from the DGLAP evolution equations [169, 170]. Modern showers are designed to cover the entire phase space as well as possible [164]. Showers are divided into two types, ISR and FSR, which occur before and after the hard process, respectively. LHC processes typically contain both ISR and FSR.

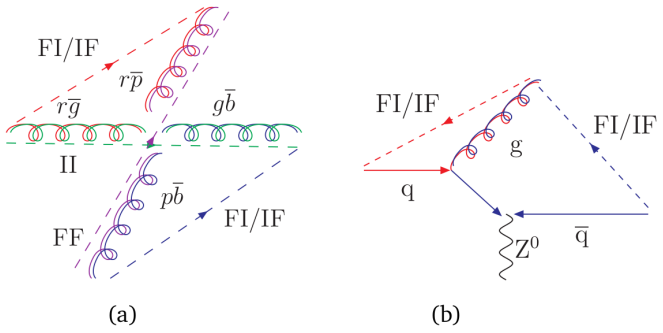


Figure 5.3: (a) Color flow for the process $gg \to gg$. The dashed lines represent the color lines between the dipole ends. The dipole type is indicated. (b) Same diagram for $q\bar{q} \to Zg$.

Since an isolated parton cannot branch due to the conservation of energy and momentum, a gluon g emission can be seen as jointly emitted by a $q\bar{q}$ pair, with q and $\bar{q}$ having opposite and compensating colors, forming a color dipole. Hence, this phenomenon can be called "dipole radiation". This picture generalizes to the subsequent emission of more gluons, and the recoil is distributed within each dipole. The full evolution in an event requires special calculations for the total color flux and the resulting set of radiating dipoles. In hadronic collisions, the dipole pattern can be quite complicated, as shown in Fig. 5.3. All possible configurations are taken into account to model every possible parton shower. The recoil can be modeled separately for ISR and FSR showers as carried by all final state particles (global) or by a single particle (dipole). The latter is preferred in processes like VBS because it relates the radiation pattern to color correlations and avoids overestimating the emission rate of the involved partons.

5.2.2 Hadronization

Hadronization, or fragmentation, is the process by which the last outgoing colored partons become colorless hadrons. This transition is non-perturbative and must be handled by phenomenological models. As mentioned above, in PYTHIA 8.3 it is based on the Lund string model [171].

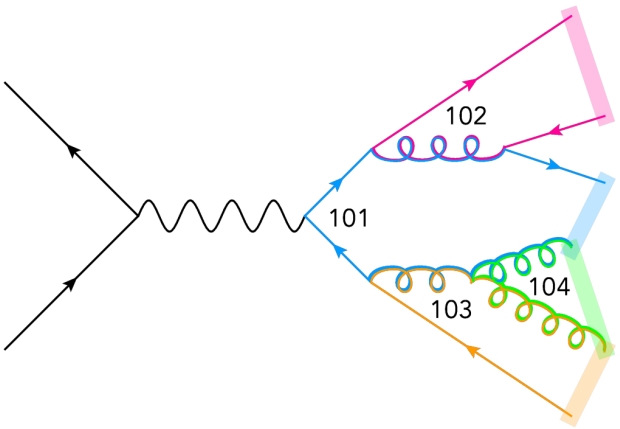


Figure 5.4: Illustration of the LC color flow in a simple $e^+e^- \to q\bar{q}$ plus shower event. The shaded regions represent the resulting unique LC string topology.

The strings are color dipoles represented as flux tubes with potential energy increasing linearly with relative distance. This model allows to "produce" the on-shell mesons and baryons

with an iterative procedure taking into account the four-momentum conservation. To properly calculate and model hadronization, a simplified set of color flow rules is used to determine between which partons confining potentials should arise. In the context of the string model, this determines a unique string topology that sets the stage for the subsequent hadronization and respects the conservation of color charges. An example of the application of this procedure is shown in Fig. 5.4 for a $e^+e^- \to q\bar{q}$ event.

5.3 Simulation of events with W' boson production

In order to properly model the W' contribution studied in the analysis reported in Ch. 6 and introduced in Sec. 2.1, the MADGRAPH5_MC@NLO framework shown above has been used. The model used is an extension of the SM, including additional interaction of fermions with a W' boson following the lowest order effective Lagrangian, as in Eq. 2.2. This model accounts for the fully differential production and decay of a W' boson, with arbitrary vector and axial-vector couplings, to any final state at the next-to-leading order in QCD. It allows a complete factorization of next-to-leading order couplings both in the partial width of the W' boson and in the full two-to-two cross-section. The model is fully described in Refs. [87, 172]. The process studied is $pp \to W' \to tb$, where the t quark decays in MADSPIN [173] according to the decay chain $t \to Wb$.

Different scenarios are taken into account:

- pure LH without SM interference;
- pure RH;
- LH plus SM interference;
- LH-RH equally mixed couplings plus SM interference.

In this way, it is possible to investigate a comprehensive spectra of the W' couplings to the RH and LH fermions, as well as the interplay with the SM counterpart.

Since the ratio $\Gamma_{W'}/m_{W'}$ is a free parameter of the model, different values are examined. In the considered model, the couplings, and thus the partial widths, are not varied together with the total width. In this interpretation this can be attributed to the presence of additional unconsidered decays, as several models predict W' bosons decaying to additional possible new particles [174, 175]. This results in a smaller branching fraction of W' to t quarks, and thus in a smaller overall cross-section times branching fraction for the process $pp \to W' \to tb$. This approach was also implemented, for example, for Z' searches reported in Ref. [176].

For the narrow width case, $\Gamma_{W'}/m_{W'} = 1\%$ was chosen, and the masses range from 2 TeV to 6 TeV in steps of 200 GeV were simulated. The lower limit is chosen according to the currently available limits [102]. In Tab. 5.1 the values of the cross-sections times the branching fractions calculated with the MADGRAPH5_MC@NLO generator for the process $pp \to W' \to tb \to Wbb$ are given for the narrow width case.

In the narrow width case, scenarios corresponding to chirality combinations different from the ones generated are considered by following the procedure described in Ref. [177]. Only in this case, indeed, for an arbitrary combination of α_R and α_L, signal distributions, yields, and cross-sections can be obtained as a linear combination of the ones from LH, RH, and LR W' samples and the SM single-top $s-$channel sample. For example, the cross-section of the investigated process for arbitrary values of α_R and α_L can be obtained with a simple linear combination of the ones for single top $s-$channel (σ_{SM}), pure LH (σ_L), pure RH (σ_R), and equally mixed LH and

RH (σ_{LR}) with $\alpha_R = \alpha_L = 1/\sqrt{2}$ [86]:

$$\sigma = (1-\alpha_L^2)\sigma_{SM} + \frac{1}{\alpha_L^2+\alpha_R^2}\left[\alpha_L^2(\alpha_L^2-\alpha_R^2)\sigma_L + \alpha_R^2(\alpha_R^2-\alpha_L^2)\sigma_R + 4\alpha_L^2\alpha_R^2\sigma_{LR} - 2\alpha_L^2\alpha_R^2\sigma_{SM}\right] \quad (5.3)$$

representing a valid approximation of the more general formula reported in Eq. 2.3. In the study described in Ch. 6, values of α_R and α_L between 0 and 1 are considered in steps of 0.1.

For the wide width case, three different hypotheses are made: 10%, 20%, and 30%. Again, the masses range from 2 to 6 TeV, but in steps of 400 GeV. The Tab. 5.2 shows the values of the cross-sections times the branching fractions calculated with the MADGRAPH5_MC@NLO generator for the process $pp \to W' \to tb \to Wbb$ for the wide width cases.

The mass range proposed for the study in Ch. 6 aims at extending the range investigated by previous analyses performing similar searches [178, 179], ending up to $m_{W'} = 4$ TeV.

Fig. 5.5 shows representative parton-level mass distributions of the tb quark pair for narrow width and large width samples, for LH and RH chirality, respectively, for a W' mass hypothesis of 3.6 TeV. For the LH hypothesis, the region of the W' boson mass spectrum below 2 TeV is dominated by the SM s-channel production of a tb quark pair. The width of the W' affects the reconstructed mass distribution, resulting in a broader peak and an asymmetry favoring lower values. This is quite visible for the cases with large decay width, where the tail towards small masses is dominant because of off-shell W' production, enhanced by rapidly increasing parton density functions for decreasing partonic momentum fractions. This effect is especially visible in the RH case, while for LH samples it is subdominant with respect to the effect of the SM contribution.

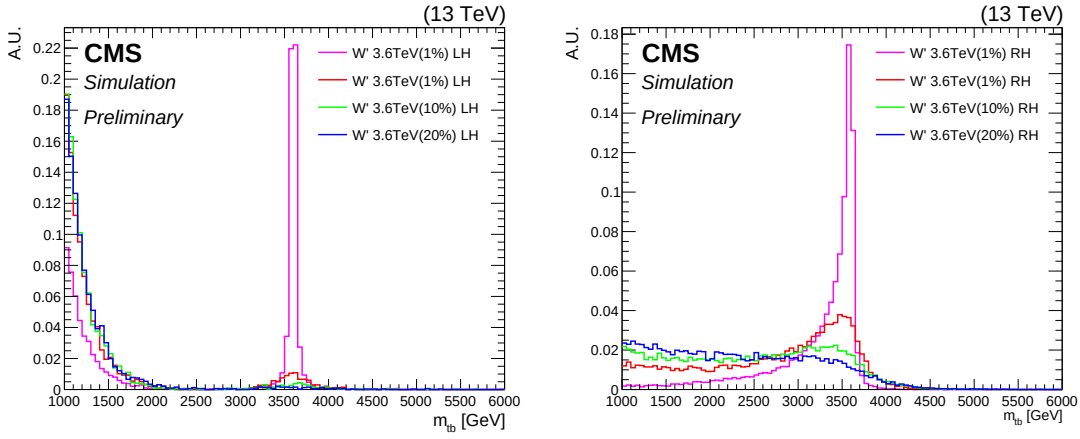


Figure 5.5: Representative parton-level distributions of the invariant mass of the tb quark pair for LH (left), RH (right) W' boson, with relative width $\Gamma_{W'}/m_{W'}$ of 1%, 10%, 20%, and 30 % of the particle mass $m_{W'}$ for a W' mass hypothesis of $m_{W'} = 3.6$ TeV.

Mass [GeV]	Width [GeV]	Cross-section × BR (pb)			
		LH no SM	RH	LH with SM	LH+RH
2000	20	1.342	1.397	1.375	1.462
2200	22	0.8111	0.8439	0.9087	0.9701
2400	24	0.5005	0.5203	0.6308	0.6756
2600	26	0.3115	0.3236	0.465	0.5016
2800	28	0.1974	0.2049	0.365	0.3945
3000	30	0.1271	0.1318	0.3031	0.3277
3200	32	0.08254	0.08553	0.266	0.2863
3400	34	0.0544	0.05633	0.2434	0.2612
3600	36	0.03624	0.0375	0.2298	0.2454
3800	38	0.02449	0.02533	0.2214	0.2355
4000	40	0.01679	0.01736	0.2168	0.2293
4200	42	0.01161	0.01208	0.2141	0.2255
4400	44	0.008501	0.00879	0.2131	0.223
4600	46	0.006172	0.006384	0.2125	0.2217
4800	48	0.004538	0.004696	0.2125	0.221
5000	50	0.003381	0.003501	0.2128	0.2204
5200	52	0.00254	0.002633	0.2131	0.2202
5400	54	0.001929	0.002008	0.2139	0.2201
5600	56	0.001476	0.001533	0.2142	0.22
5800	58	0.001138	0.001182	0.2148	0.2201
6000	60	0.0008807	0.0009153	0.2154	0.2203

Table 5.1: Cross-section obtained with the MADGRAPH5_MC@NLO generator for narrow widths (1%) W' boson for different mass hypotheses.

Mass [GeV]	Width [GeV]	Cross-section × BR (pb)			
		LH no SM	RH	LH with SM	LH+RH
2000	200	0.148	0.154	0.2508	0.3059
	400	0.07863	0.08177	0.1953	0.2436
	600	0.05403	0.05617	0.1825	0.2254
2400	240	0.05875	0.06106		
	480	0.03287	0.03416		
	720	0.02314	0.02404		
2800	280	0.02556	0.02655	0.1969	0.2222
	560	0.01496	0.01554	0.1897	0.2132
	840	0.01115	0.01158	0.1898	0.2115
3200	320	0.01197	0.01242		
	640	0.007494	0.007778		
	960	0.005784	0.006004		
3600	360	0.005971	0.006191	0.2004	0.215
	720	0.004044	0.004195	0.1998	0.2139
	1080	0.003249	0.003372	0.2009	0.214
4000	400	0.003271	0.00339		
	800	0.002355	0.002443		
	1200	0.00195	0.002024		
4400	440	0.001899	0.001967	0.2069	0.2165
	880	0.001437	0.001491	0.2072	0.2165
	1320	0.001232	0.001278	0.2081	0.2168
4800	480	0.001173	0.001217		
	960	0.00094	0.0009753		
	1440	0.0008208	0.0008519		
5200	520	0.0007642	0.0007929	0.2116	0.2185
	1040	0.0006335	0.0006573	0.212	0.2185
	1560	0.0005667	0.0005882	0.2129	0.2187
5600	560	0.000518	0.0005376		
	1120	0.000449	0.000466		
	1680	0.0004055	0.0004209		
6000	600	0.0003715	0.0003858	0.2148	0.2198
	1200	0.0003255	0.000338	0.2153	0.2198
	1800	0.0002973	0.0003086	0.2155	0.2198

Table 5.2: Cross-section obtained with the MADGRAPH5_MC@NLO generator for wide widths (10%, 20%, and 30%) W' boson for different mass hypotheses.

5.4 Simulation of Vector Boson Scattering processes and related SMEFT contributions

In the work presented in Ch. 7, the vector boson scattering between a W pair of the same sign (ssWW VBS), with $\ell\tau_h\nu_\ell\nu_\tau$ in the final state, represents the signal under study. In particular, different contributions are considered, first of all, the SM process and the SMEFT operators responsible for the anomalous W boson couplings.

The SM contribution is simulated following the path established in the previous analysis of the CMS Collaboration, which investigated the VBS of W pairs of the same sign with a fully leptonic final state, both for the pure EW process and for the irreducible QCD background, as illustrated in Sec. 1.6. The interference is not taken into account, since it gives a negligible contribution [116]. As a consequence, the VBS signal samples are simulated at Leading Order (LO) with the MADGRAPH5 version 2.6.5 generator, requiring a final state with the $\ell\nu_\ell, \tau\nu_\tau$ pairs from the decays of the two W bosons.

The starting point for the simulation of the dim-6 and dim-8 EFT operators is the setup for the simulation of the SM signal, to include all the operators listed in Sec. 2.3. For the dim-6 operators, the effective lagrangian for MADGRAPH is defined with the SMEFTSIM package [180, 181], while the dim-8 operators are implemented with a special FEYNRULES set [182]. In both cases, the reweighting procedure explained in Eq 5.2 is used to take into account the contribution of all operators to face the computational and memory power limitations. The version of MADGRAPH used to simulate the EFT contributions is 2.6.5.

5.5 Data-MC corrections

Simulations must be corrected regularly to account for various experimental effects that change from year to year due to different experimental conditions. For many of the corrections, scaling factors are provided by the CMS Collaboration and are applied with the same procedure in each analysis. In the following, the corrections implemented in the studies reported in Ch. 6 and Ch. 7 are described. The others, derived and applied specifically for the analyses presented in the aforementioned Chapters, are described in separate Sections.

HEM15/16 issue in 2018

During the final data-taking period of 2018, two sections of the hadron endcap calorimeter ceased to function. To account for this, events in these runs are vetoed if an electron or narrow jet, as defined in Sec 4.3 and Sec. 4.4, falls in the region $-3.2 < \eta < -1.3$ and $-1.57 < \phi < -0.87$, respectively. Events in the MC samples for the 2018 conditions where an electron or jet falls in this region are weighted to emulate the actual luminosity recorded with those modules active. This corresponds to about 35% of the total 2018 sample. Figure 5.6 shows the effect of such a correction on the electron angular distributions, with the ones on the bottom that are fixed.

Prefiring in 2016 and 2017

In the 2016 and 2017 data acquisition periods, some trigger primitives in ECAL were associated with the wrong bundle crossing. As a consequence of the trigger rules vetoing subsequent events, this could cause a good physics event to self-veto if a trigger was fired in the region of interest. A similar effect is present in the muon system, where the assignment of muon candidates can be wrong due to the limited time resolution of the muon detectors. This effect was most pronounced in 2016 but is non-zero for both 2017 and 2018. The prefire probabilities

have been measured per photon, jet, and muon by the dedicated CMS subgroup. These can be used to calculate an event weight to be applied to each Monte-Carlo simulated event to reproduce the prefiring effect observed in the data:

$$w = 1 - P(\text{prefiring}) = \prod_{i=\text{photons, muons, jets}} \left[1 - \varepsilon_i^{prefiring}(\eta, p_T)\right] \tag{5.4}$$

Events were corrected for this effect with a per-event weight, and the uncertainties of the measured per-jet, per-photon efficiencies were propagated through this chain.

Pile-up reweighing

Monte-Carlo samples are generated with distributions for the number of pile-up interactions to roughly cover the conditions expected for each data-taking period. Although the primary vertex reconstruction is efficient and well-behaved up to high levels of pile-up, the final distribution for the number of reconstructed primary vertices is still sensitive to the details of the primary vertex reconstruction and to differences in the underlying event in data vs the Monte-Carlo samples. Additionally, there is the potentially larger effect that the distribution for the number of reconstructed vertices can be biased by the offline event selection criteria and the trigger. As a consequence, the number of primary vertices observed in the simulations does not match that observed in the data, and the number of simulated pile-up interactions is then reweighed to match the number of primary vertices observed in the data, as recommended by the CMS dedicated subgroup.

Simulated top p_T reweighting

During the Run I and Run II data-taking periods of the LHC, it was observed that the spectra of the top quark p_Ts in the data were significantly softer than those predicted by the various simulations based on either LO or NLO matrix elements linked to parton showers. To properly address this issue, the dedicated CMS subgroup provided reweighting functions to be applied to processes containing a $t\bar{t}$ quark pair at the matrix element generation level [183].

$b-$tagging efficiencies

To correct the $b-$tagging discriminant distributions in the Monte-Carlo sample to match those in the data, a reweighting method is applied. This method works for arbitrary combinations of $b-$tagged and mistagged jets, it does not require the use of discriminant reshaping, and the goal of this method is to predict the correct event yield in the data by changing only the weight of the selected Monte-Carlo events, with the events not passing the selection that are not needed to be added back.

In particular, they are calculated by properly combining scale factors (SF_i) and efficiencies (ε); the former are provided by the dedicated CMS subgroup for each working point of the available $b-$taggers, while the latter have to be specifically estimated for each analysis. The probability of a given jet configuration in the Monte-Carlo simulation or the data is defined as follows:

$$P(\text{MC}) = \prod_{i=\text{tagged}} \varepsilon_i \prod_{j=\text{not tagged}} (1 - \varepsilon_j) \tag{5.5}$$

$$P(\text{DATA}) = \prod_{i=\text{tagged}} \text{SF}_i \varepsilon_i \prod_{j=\text{not tagged}} (1 - \text{SF}_j \varepsilon_j) \tag{5.6}$$

The event weight is calculated:

$$w = \frac{P(\text{DATA})}{P(\text{MC})} \tag{5.7}$$

The resulting $b-$tagging and mistagging weights are then applied to both signal and background samples, both when $b-$tagging is used to select events in MC-simulated samples and when it is used to veto events with $b-$jets.

Trigger efficiencies and scale factors

The trigger rules available in CMS Collaboration present different efficiencies in the Monte-Carlo simulated samples and the data, so special scale factors are needed to properly reproduce the data behavior with the Monte-Carlo samples.

For the analysis reported in Ch. 6, the trigger efficiencies and scale factors are specifically calculated and illustrated in a dedicated Section.

For the study presented in Ch. 7, the efficiencies and scale factors for the isolated muon triggers have been calculated by the dedicated CMS group and consistently implemented in the analysis workflow; those for the isolated electron triggers have been specifically estimated for the peculiar final state and will be illustrated in a separate Section.

Light lepton reconstruction, isolation and identification efficiencies, and scale factors

The efficiencies associated with the reconstruction, isolation, and identification algorithms applied to the muons and electrons (see Sec. 4.2 and Sec. 4.3, respectively) are different between Monte-Carlo samples and data, and thus scaling factors are needed to scale the distributions extracted from Monte-Carlo samples to what is observed in the data. These corrections, already computed in the CMS Collaboration for the muons, have been consistently applied in the analyses presented in Ch 6 and Ch 7; for the electrons, for the study in Ch. 6 it was possible to use the corrections already computed in the CMS Collaboration, for the one described in Sec. 7 this was possible only for the reconstruction and identification scale factors, while those for the isolation algorithm were obtained specifically for this analysis together with the SFs for the single electron trigger.

Hadronic τ lepton reconstruction and identification

Similar to the light leptons, the distributions with hadronic taus extracted from Monte-Carlo samples need to be corrected to account for discrepancies with the data concerning the reconstruction with the HPS algorithm and the identification with the DeepTau discriminators. The dedicated CMS subgroup provides the latest version of the scale factors associated with the HPS algorithm and the three DeepTau discriminators, and they are consistently applied in the VBS analysis described in Ch. 7.

Hadronic jets identification

For the tight ID identification criteria set, there is no need to implement scale factors since the efficiency is close to 100% in both Monte-Carlo simulations and data. For the PileupJet ID criteria, instead, the scale factors are calculated and distributed by the dedicated CMS subgroup and applied with the same method described for the $b-$tagging/mistagging scale factors.

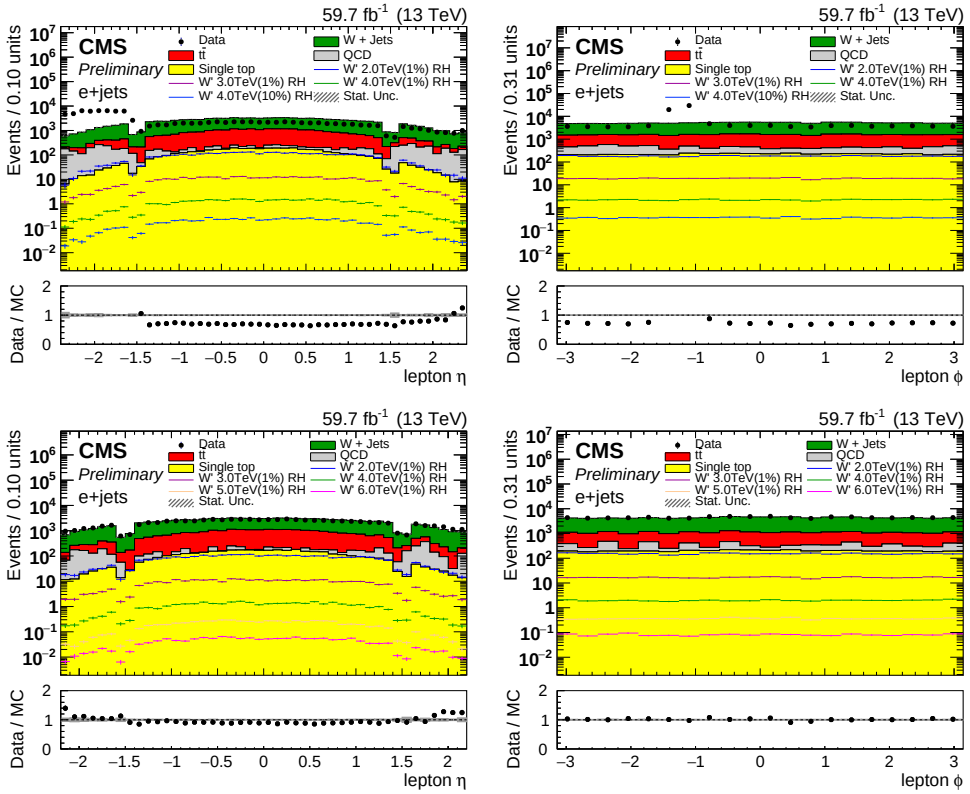


Figure 5.6: Distribution of electron η (left) and ϕ (right) before (top) and after (bottom) applying the correction described in Sec. 5.5, as obtained in the study reported in Ch. 6.

5.6 Systematic uncertainties

The modeling of Monte-Carlo simulations introduces some irreducible uncertainties that affect the power of statistical analysis acting in the different steps of the simulations, and are considered systematic uncertainties, or "systematics". They are ultimately implemented in the statistical analysis as nuisance parameters and can affect both the shape and the yield of the distributions used for signal extraction. Their magnitude is evaluated by shifting the distributions of interest from $+1\sigma$ (-1σ) up (down) in accordance with the statistical distribution of the associated nuisance parameter. They are usually classified into experimental and theoretical systematics: the former are due to inaccuracies in the simulation of the detector response or to effects observed in the data that are not included in the Monte-Carlo simulations, and the latter arise from the intrinsic uncertainties affecting the theoretical prediction used to generate the simulation of a given process. The experimental systematics are essentially the statistical fluctuations of the corrections presented in Sec. 5.5 and can be estimated together with the corrections as their statistical uncertainties. Theoretical uncertainties have usually a less straightforward statistical interpretation and introduce more assumptions, and are usually the main source of the total uncertainty affecting the signal extraction. Their magnitudes can be assessed by varying the parameters of the theoretical modeling of the simulations in accordance with their current uncertainty.

In the following, the experimental systematic uncertainties, which are not related to any of the corrections listed in the previous Section, are described together with the theoretical ones. They are included in the studies described in detail in Ch. 6 and Ch. 7.

Luminosity

The luminosity of the collected data is affected by intrinsic statistical uncertainty. As recommended by the CMS Collaboration, a relative uncertainty of 1.6% is applied to the integrated luminosities for 2016, 2017, and 2018. This uncertainty is considered to affect only the yields and is treated as uncorrelated over the three years of the Run II data-taking period [184–186].

Jet Energy Scale and Resolution (JES and JER)

In simulated events, the reconstructed four-momenta of all hadronic jets are simultaneously varied according to the $\eta-$ and p_T- dependent uncertainties in the jet energy scale (JES) described in Sec. 4.4. These variations are then propagated to the $\vec{p}_T^{\,\text{miss}}$. In addition, the same procedure is applied to properly evaluate the systematic effect due to differences in jet energy resolution (JER) between data and simulations.

Unclustered energy

The energy deposits in the calorimeters that are not used by Particle Flow in any reconstructed object constitute the so-called unclustered energy. Since the missing transverse momentum $\vec{p}_{\mathrm{T}}$ is extracted from the transverse imbalance of the total reconstructed momentum, the amount of unclustered energy indirectly affects this measurement. The systematic effects due to the uncertainties on the this quantity are implemented similarly to what is done for JER and JES.

Tau Energy Scale uncertainties (TES and FES)

As mentioned in Sec. 4.5, according to the real nature of the generated particle associated with a given reconstructed τ_h in simulated events, corrections to the energy of each reconstructed hadronic tau are imposed via the Tau/Fake Energy Scale (TES/FES) illustrated in Sec. 4.5. They introduce systematic uncertainties due to the corresponding statistical uncertainties and are evaluated in a similar way to the identification of the hadronic tau.

QCD scale

Uncertainties on the total yield for background and signal processes are derived by varying the renormalization and factorization scale, μ_R and μ_F, at the matrix element (ME) level, and discussed in the introduction of this Chapter. The total uncertainty can be estimated considering the envelope of the combined effect of upward and downward variations with respect to the nominal expectation. It is considered uncorrelated among the different classes of processes and correlated among years.

Parton Distribution Functions uncertainties

The Monte-Carlo samples are generated using a default Parton Distribution Function (PDF) set, depending on the features of the physics process that is simulated. To take into account the uncertainty arising from the choice of a particular PDF set, the effects of 103 alternative ones are also evaluated during the simulations, with their deviations from the nominal one evaluated using the reweighting procedure explained in Sec. 5 and stored in dedicated event weights. These can be used to extract alternative kinematic distributions and to estimate the systematic uncertainty arising from the choice of a particular set of PDFs. The complete procedure is implemented according to the guidelines of the PDF4LHC group [187].

Parton showering uncertainties

Among the possible theoretical uncertainties affecting the simulated events, there are those related to the QCD-induced parton shower modeling, i.e. the parton showering. They are divided into Initial State Radiation and Final State Radiation and are estimated in a similar way as described for the QCD scale uncertainty.

Limited size of Monte-Carlo simulated event samples

The limited size of the simulated event samples, the MC statistics for short, introduces an irreducible source of statistical uncertainty into the statistical analysis, limiting the precision of the modeling, and is therefore considered a systematic uncertainty. It is evaluated for each bin using the Barlow-Beeston “lite” method [188, 189].

Part III

Searches for New Physics with vector bosons

Chapter 6

Search for a new W' boson decaying in third generation quarks

As shown in Sec. 2.1, the SM is not sufficient to explain the current picture of fundamental particle physics, and several other models have emerged to provide an explanation of the open issues within the SM and introduce new particles in the LHC energy range. Among these models, those that predict the existence of new massive bosons, called Z' and W', with properties similar to the SM electroweak bosons, are particularly interesting. These bosons can be detected by the CMS experiment by performing a "bump hunt" search over a continuous background in the case of a narrow-width resonance. Many such studies have already been performed at the LHC, but no significant excess has been found. There remains the possibility that the New Physics does not manifest itself as a clear peak in the invariant mass distribution, for example in the case of a broad-width resonance. In this case, the bump-hunting technique is not optimal, as the peak may be smeared, becoming indistinguishable from a smooth deviation from the SM spectrum, as reported in Ref. [176]. For such scenarios, the study of the couplings of these new particles with the SM ones and the study of the spin-angle properties of the decay products can improve the sensitivity of the search. A particularly important role in this context is played by the t quark, which on the one hand has a strong coupling to the W' and Z' bosons due to its large mass in many BSM scenarios, and on the other hand has a very distinct signature due to its decay chain, which can be fully reconstructed in the detector.

This analysis [1] focuses on the leptonic decay chains of the hypothetical W' boson, i.e. $qq' \to W' \to tb \to Wbb \to \ell\nu_\ell bb$, where ℓ stands for e and μ, promptly produced by the W decay or coming from the leptonic decay of a τ lepton produced in the t decay chain. This study aims to investigate the properties of the final state and to obtain stringent exclusion limits for the existence of this new particle, in the context of the Sequential Standard Model introduced in Sec. 2.1 and in Sec. 5.3.

6.1 Signal and background processes

One of the advantages of a hadron collider in the search for New Physics is that the momentum of the colliding partons is not fixed. Thus, a wide range of masses for new particles is accessible, and the predictions of many models can be tested. At the LHC, the main production channel for a real, i.e. on-shell, W' boson is the $s-$channel. This process allows the mass of the hypothesized

particle to be probed from its decay products. One of the most promising ways to search for W' is through its decays via third-generation quarks, namely $W' \to tb$.

The representative Feynman diagram at LO for the process under study is shown in Fig. 6.1. This channel has a relatively small QCD multijet background compared to the channel $W' \to \mathrm{ud}, \mathrm{cs}$.

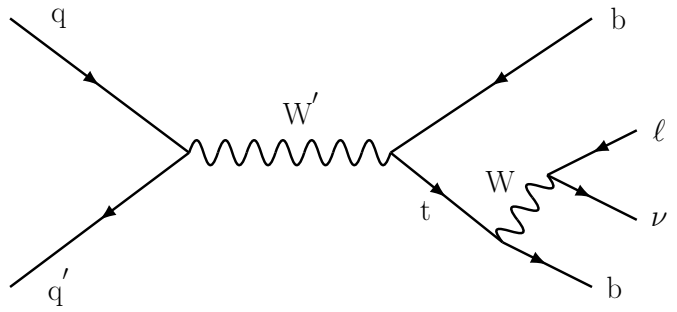


Figure 6.1: Leading Feynman diagram for the W' production in the $s-$channel.

As fully introduced in Sec. 5.3, the decay width of the W' boson could be significant compared to its mass, resulting in signatures that could escape standard searches. For the first time, this analysis probes relative W' widths ($\Gamma_{W'}/m_{W'}$) of 1%, 10%, 20% and 30%. In addition, cases where the chirality of the W' boson is left-handed (LH), right-handed (RH), or a combination of the two are considered, together with the interference with the SM production of single top quarks in the s channel.

The main sources of background that present a final state that mimics that expected from W' decay are the following

- $t\bar{t}$: The production of a top quark-antiquark pair is the main background source. Events where one top quark has a lepton in the decay chain and the other decays to hadronic final states closely mimic the signal signature. The presence of two $b-$jets in the final state is another challenge to discriminate this background. Additional contamination comes from dileptonic decays where a lepton is not reconstructed or is out of the analysis acceptance.

- W+jets: Processes where a W boson is produced in association with quarks, especially heavy flavor quarks, form the second main background together with $t\bar{t}$. The leptonic decay chain of the W boson is the same as in the signal processes, and if $b-$jets are reconstructed in the same event, the final topology of the signal matches the content of the physical objects. An important feature is the absence of the top quark, which can be exploited to discriminate this type of background.

- QCD-induced multijet production (QCD): strong interaction processes producing a quark-antiquark pair are also a background source, in very rare configurations where a non-prompt lepton, for example from a pion or from a B hadron decay chain, is reconstructed and a large amount of energy is lost due to very energetic neutrinos. These events are rarely present in the same kinematic region of interest for the signal, this rarity is only balanced by the very large production cross-section. The reconstructed invariant mass of the t quark is a powerful quantity to discriminate events from this background against the signal.

- Single t quark processes: Both $t-$channel and tW associated productions have the same signature as the signal. The $s-$channel is instead the same process as the signal, but the interaction is via a virtual W boson.

- Other minor backgrounds, such as double vector boson production, are neglected because their production cross-sections are much smaller than the processes listed above, and they could contribute to the final selection only in very rare final state configurations.

6.2 CMS dataset

The analysis uses p-p collision data at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 138 fb^{-1}, collected from 2016 to 2018 by the CMS experiment at the LHC. They are collected with triggers requiring either a muon, an electron, or large amounts of hadronic activity H_T in the final state. Tab. 6.1 shows the list of primary datasets used in the analysis.

Dataset	Year	Integrated luminosity (fb^{-1})
≥ 1 muon, ≥ 1 electron, or significant H_T	2016	36.3
≥ 1 muon, ≥ 1 electron, or significant H_T	2017	41.5
≥ 1 muon, ≥ 1 electron, or significant H_T	2018	59.8

Table 6.1: List of primary datasets of p-p collision data produced at $\sqrt{s} = 13$ TeV and collected by CMS from 2016 to 2018 used in the analysis. The data collected correspond to an integrated luminosity of 138 fb^{-1}.

6.3 Monte-Carlo simulated samples

Monte-Carlo event generators are used to simulate signal and background samples. For the signal, full details are given in Sec. 5.3.

The single top quark $t-$channel events are generated at next-to-leading order (NLO) in QCD with POWHEG 2.0 [55–57]. The four-flavor scheme [190] is used for events with the V_{tb} vertex in production, while the five-flavor scheme [191] is used for events with a V_{td} or V_{ts} vertex in production. Top quark decays are simulated with MADSPIN [173].

The $t\bar{t}$ background process [192], the associated t quark and W boson production are simulated with POWHEG in the five-flavor scheme [193]. Single t quark s channel events (t, s-ch) are simulated with MADGRAPH5_MC@NLO 2.2.2 [54] at NLO. For all simulations involving a t quark, the value of the t quark mass used in the simulated samples is 172.5 GeV, and PYTHIA 8. 180 [194] with the tune CUETP8M1 [195] is used to simulate the parton shower, the quark hadronization, and the underlying event, except for $t\bar{t}$ where the tune CUETPM2T4 is used. Such a tune has been derived specifically for top quark-antiquark pair production in CMS [196].

Simulated event samples with a W boson produced in conjunction with jets (W+jets) are generated using MADGRAPH5_MC@NLO 2.2.2. For these processes, events are simulated with up to two additional partons emitted in hard scattering, and the FxFx merging scheme [197] is used to avoid double counting with parton emissions generated in parton showering. Simulated QCD multijet events are generated at LO with PYTHIA 8.180 as a function of the HT activity of the event. For 2017 and 2018 the tune CP5 [198] is used for all samples, while for 2016 the tune CUETP8M1 [199] is used.

The default PDF parameterization used in all simulations is NNPDF3.0 [200] in LO or NLO QCD, with the order matching that of the matrix element calculation. All generated events undergo a full simulation of the detector response according to the model of the CMS detector within GEANT4 [201]. Additional p-p interactions within the same or nearby pile-up are included

in the simulation with the same distribution as observed in the data. Except for the QCD multijet process, which is obtained from a fit to the data, all simulated samples are normalized to the expected cross-sections. The details of the simulations for the samples used in the analysis are summarized in Tab. 6.2.

Process	Matrix Element	Matching Method	Tune	Cross-section × BR (pb)
pp→ $t\bar{b}$	POWHEG-MADSPIN	-	CUETP8M1/CP5	136.02 (NLO)
pp→ $\bar{t}b$	POWHEG-MADSPIN	-	CUETP8M1/CP5	80.95 (NLO)
pp→tW	POWHEG	-	CUETP8M1/CP5	35.6 (NLO)
pp→ $\bar{t}$W	POWHEG	-	CUETP8M1/CP5	35.6 (NLO)
pp→tb→ $\ell\nu$bb (s-ch.)	MC@NLO	-	CUETP8M1/CP5	10.32 (NLO)
pp→Wq→ $\ell\nu$j $70 < H_T < 100$	MC@NLO	FxFx	CUETP8M1/CP5	1353.0 * 1.21
pp→Wq→ $\ell\nu$j $100 < H_T < 200$	MC@NLO	FxFx	CUETP8M1/CP5	1345 * 1.21
pp→Wq→ $\ell\nu$j $200 < H_T < 400$	MC@NLO	FxFx	CUETP8M1/CP5	359.7 * 1.21
pp→Wq→ $\ell\nu$j $400 < H_T < 600$	MC@NLO	FxFx	CUETP8M1/CP5	48.91 * 1.21
pp→Wq→ $\ell\nu$j $600 < H_T < 800$	MC@NLO	FxFx	CUETP8M1/CP5	12.05 * 1.21
pp→Wq→ $\ell\nu$j $800 < H_T < 1200$	MC@NLO	FxFx	CUETP8M1/CP5	5.501 * 1.21
pp→Wq→ $\ell\nu$j $1200 < H_T < 2500$	MC@NLO	FxFx	CUETP8M1/CP5	1.329 * 1.21
pp→Wq→ $\ell\nu$j $H_T > 2500$	MC@NLO	FxFx	CUETP8M1/CP5	0.03216 * 1.21
pp→ $t\bar{t}$	POWHEG	-	CUETP8M2T4/CP5	831.8 (NLO)
pp→ $t\bar{t}$ $700 < m_{t\bar{t}} < 1000$	POWHEG	-	CUETP8M2T4/CP5	80.5 (NLO)
pp→ $t\bar{t}$ $m_{t\bar{t}} > 1000$	POWHEG	-	CUETP8M2T4/CP5	21.3 (NLO)
pp→$q\bar{q}$ $300 < H_T < 500$	MC@NLO	-	CUETP8M2T4/CP5	347700 (NLO)
pp→$q\bar{q}$ $500 < H_T < 700$	MC@NLO	-	CUETP8M2T4/CP5	32100 (NLO)
pp→$q\bar{q}$ $700 < H_T < 1000$	MC@NLO	-	CUETP8M2T4/CP5	6831 (NLO)
pp→$q\bar{q}$ $1000 < H_T < 1500$	MC@NLO	-	CUETP8M2T4/CP5	1207 (NLO)
pp→$q\bar{q}$ $1500 < H_T < 2000$	MC@NLO	-	CUETP8M2T4/CP5	119.9 (NLO)
pp→$q\bar{q}$ $H_T > 2000$	MC@NLO	-	CUETP8M2T4/CP5	25.24 (NLO)

Table 6.2: Simulated samples used for the three years in the analysis.

6.4 Physics objects and selections

To define a sample enriched in signal events, the first set of selection requirements is designed to identify the final state objects that could come from a W' boson candidate decay. Events are retained if their final state has exactly one light lepton, i.e., muon or electron, at least two jets, and some amount of missing transverse momentum attributed to the invisible neutrino. Events are selected when a primary vertex is reconstructed and the noise filters are passed as reported in Sec. 4.6. All physics objects and their selection requirements are summarized in Tab. 6.4.

6.4.1 Trigger

Events are required to pass a combination of electron, photon, muon, or hadronic triggers to take advantage of the presence of a muon, an electron, or large amounts of hadronic activity in a $W'-$like event. According to this criterion, muons reconstructed by the offline trigger algorithm are selected if they pass a 50 GeV threshold in p_T, while electrons passing a higher threshold of 115 GeV are selected. In addition, triggers with isolation criteria and lower p_T thresholds are used to recover lower p_T electrons.

The hadronic triggers use the quantity H_T, defined as the transverse component of the total four-momentum of the reconstructed jets in the event, as a quantitative measure of the hadronic activity. All used trigger paths[1] are listed in Tab. 6.3.

If the trigger for high-momentum electrons is used, it must be implemented together with the corresponding paths for photon selection, since the electrons with high transverse momentum can be reconstructed as photons by the online HLT algorithm.

To account for differences between simulation and data, trigger scale factors are derived ad hoc for this analysis, and the procedure is described in Sec. 6.4.5.

[1] A trigger path is defined as a particular set of requirements applied to physics objects by the HLT software.

Purpose	Path	Year
High p_{T} electron	1 MVA WP80 NoIso electron with $p_T > 115$ GeV or 1 isolated photon with $p_T > 175$ GeV	2016
	1 MVA WP80 NoIso electron with $p_T > 115$ GeV or 1 isolated photon with $p_T > 200$ GeV	2017C to F, 2018
Low p_{T} electron	1 MVA WP80 NoIso electron with $p_T > 27$ GeV	2016
	1 MVA WP80 NoIso electron with $p_T > 35$ GeV	2017, 2018
Muon	1 tight ID muon with $p_T > 50$ GeV or 1 Tracker muon with $p_T > 50$ GeV	2016
	1 tight ID muon with $p_T > 50$ GeV	2017B
	1 tight ID muon with $p_T > 50$ GeV or 1 Tracker muon with $p_T > 50$ GeV	2017C to F, 2018
HT	$H_T > 800$ GeV	2016B to F
	$H_T > 900$ GeV	2016G
	$H_T > 780$ GeV	2017, 2018

Table 6.3: Complete list of trigger paths required for the analysis.

6.4.2 Electrons and muons

Events with exactly one electron with $p_{\mathrm{T}} > 50$ GeV, $|\eta| < 2.2$ and passing the MVA NoIso WP90 as defined in Sec. 4.3 are selected. In addition, electrons in the barrel endcap gap at $1.4442 < |\eta| < 1.566$ are excluded from the selection. To distinguish prompt leptons from those coming from secondary decay chains, which are usually included in a beam, isolation is an important tool. However, the electrons produced by the W' decay chain come from boosted objects, and, as a consequence, they often overlap with a jet produced in the same decay chain. For this reason, a standard isolation cut as described in Sec. 4.2 for muons and in Sec. 4.3 for electrons would fail to properly discriminate the W' signal from the competing backgrounds. To overcome this limitation, a special isolation criterion called "mini-isolation" is adopted to recover the signal efficiency when leptons are produced in the decay chain of a boosted object. It is constructed by defining a cone whose radius R varies with the p_{T} of the lepton as follows:

- R = 0.2 for lepton with $p_{\mathrm{T}} \leq 50$ GeV,
- R = 10 GeV/p_{T} for lepton with $50 < p_{\mathrm{T}} < 200$ GeV,
- R = 0.05 for lepton with $p_{\mathrm{T}} \geq 200$ GeV.

The isolation is then computed for a given lepton in this cone defined on an event basis, taking into account its energy to reduce the probability of overlap for more boosted topologies. For this analysis, electrons are required to have mini isolation $I^{\mathrm{e}}_{\mathrm{mini}} < 0.1$, according to the studies reported in Appendix A.

Analogous selection criteria are applied to single muon events. The latter is required to have $|\eta| < 2.4$ and to be reconstructed as a tight muon by the PF algorithm (see Sec. 4.2). In addition, an isolated muon with $p_{\mathrm{T}} > 55$ GeV is selected to achieve the highest trigger efficiency. Similar to electrons, muon isolation requirements are applied using the mini-isolation shown above, with $I^{\mu}_{\mathrm{mini}} < 0.1$.

Veto on additional leptons

Events containing additional leptons besides either a muon or an electron, selected according to the criteria defined above, are excluded from the selection, since they may increase the contamination from the dileptonic $t\bar{t}$ background. Veto muons are defined as passing the loose identification criteria defined in Sec. 4.3 with $p_{\mathrm{T}} > 30$ GeV, and $I^{\mu}_{\mathrm{mini}} < 0.4$. They are also required not to pass the narrow WP so as not to overlap with the corresponding selection rule. Similarly, additional electrons that pass the MVA NoIso WPLoose but not the MVA NoIso WP90 with $p_{\mathrm{T}} > 35$ GeV, $|\eta| < 2.2$, and $I^{\mu}_{\mathrm{mini}} < 0.4$ are vetoed.

6.4.3 Jets

Jets are reconstructed by the PF algorithm using the anti-k_{T} algorithm as described in Sec. 4.4. Two types of jet definitions are used in this analysis:

- jets clustered with a cone parameter of $R = 0.4$ (AK4 jets) are considered to search for $b-$jets coming from the W' decay vertex, or the jets coming from the t quark decay chain;
- jets clustered with a cone parameter of $R = 0.8$ (AK8 jets) are also used first to perform a loose pre-selection, and then to define an additional veto on hadronic decaying t quarks coming from SM backgrounds.

Jets in both collections must pass a set of identification criteria corresponding to the "tight" working point. AK4 jets are considered for analysis with $|\eta| < 2.4$ and it is required that at least two of them have $p_{\mathrm{T}} > 100$ GeV. The presence of at least 2 AK8 jets is also required to reduce contamination from low energy $t\bar{t}$ and QCD events while maintaining a high signal efficiency of 99.8%. This rule exploits, first, the property of the top decay products to be highly boosted in the searched signal process and therefore to be collected in a single AK8 jet, and, second, the expected property of the b quark coming from W' decay to transport a large amount of energy and induce a large-area hadronic shower.

$b-$tagging

In the final state of the signal, the two jets from the W' boson and top quark decay vertices come from the hadronization of a $b-$quark pair. Consequently, $b-$tagging plays a key role in increasing the sensitivity of the signal.

A $b-$tagging procedure is applied to the AK4 jets with $|\eta| < 2.4$, and the performances in terms of signal sensitivity of the DeepCSV and DeepFlavor algorithms have been thoroughly investigated (see Appendix A). The threshold chosen to label jets as $b-$jets is the "Medium" operating point of the DeepFlavor discriminator, corresponding to a score > 0.2770 for 2016, 2017, and 2018. This choice induces a false alarm rate of about 1%. In Appendix A, more details are also given for the efficiency and mistag working points as a function of the candidate $b-$jet transverse momentum.

To account for the different $b-$tagging and mistagging efficiencies in the data and MC, scale factors for b-jets, c-jets, and light jets are used to reweight MC events using the procedure illustrated in Sec. 5.5.

6.4.4 Missing transverse momentum

If no neutral particles other than SM neutrinos are expected to be produced and then escape the detector without leaving a trace, they represent the only contribution to the missing transverse momentum. In particular, the signature of the search signal events is expected to have a significant amount of p_{T}^{miss} due to the large mass of the W' boson. A requirement on this quantity can therefore help to reduce the contamination from QCD background processes, mainly $b\bar{b}$ and $c\bar{c}$, which are expected to have lower p_{T}^{miss} due to non-prompt low-energy neutrinos and resolution effects. Therefore, $p_{\mathrm{T}}^{miss} > 120$ GeV is required.

6.4.5 Data-MC corrections

The simulation is corrected to account for several experimental effects that change from year to year due to the changing experimental conditions. Many of the corrections are provided directly by the CMS Collaboration, as reported in Sec. 5.5. Corrections derived specifically for this analysis are described below.

Object	Requirements	Counts	
		μ	e
Tight muon	$p_{\mathrm{T}} > 55$ GeV, $\lvert\eta\rvert < 2.4$, $I^{\mu}_{\mathrm{mini}} < 0.1$	1	0
Loose muon	$p_{\mathrm{T}} > 30$ GeV, $\lvert\eta\rvert < 2.4$, $I^{\mu}_{\mathrm{mini}} < 0.4$	0	0
MVA WP80 NoIso electron	$p_{\mathrm{T}} > 50$ GeV, $\lvert\eta\rvert < 2.2$, $I^{\mathrm{e}}_{\mathrm{mini}} < 0.1$	0	1
MVA WPLoose NoIso electron	$p_{\mathrm{T}} > 35$ GeV, $\lvert\eta\rvert < 2.2$, $I^{\mathrm{e}}_{\mathrm{mini}} < 0.4$	0	0
p_{T}^{miss}	$p_{\mathrm{T}} > 50$ GeV		
Jet	$p_{\mathrm{T}} > 100$ GeV, $\lvert\eta\rvert < 2.4$	≥ 2	
$b-$jet	$p_{\mathrm{T}} > 100$ GeV, $\lvert\eta\rvert < 2.4$, DeepFlv ≥ 0.2770	≥ 0	

Table 6.4: Summary of all defined physics objects, their corresponding selection requirements, and the number of candidates required for the muon and electron channels.

Trigger efficiency measurement

The trigger efficiency has been measured by considering a sample orthogonal to the one used for this analysis, requiring:

- 1 tight muon and 1 tight electron
- at least 2 jets with $p_{\mathrm{T}} > 100$ GeV and $\lvert\eta\rvert < 2.4$

This selection is applied to data samples described in Table 6.1, as well as to dilepton $\mathrm{t\bar{t}}$ samples. The efficiency is then derived as:

$$\varepsilon = \frac{N_{pass}}{N_{total}}, \tag{6.1}$$

where for:

- Muon trigger efficiency: N_{total} is defined as the number of events in which the electron or photon triggers are fired and the electron has $p_{\mathrm{T}} > 50$ GeV and N_{pass} is found by additionally requiring the muon or HT trigger to be fired as well;
- Electron trigger efficiency: N_{total} is defined as the number of events in which the muon trigger is fired and the muon has $p_{\mathrm{T}} > 55$ GeV and N_{pass} is found by additionally requiring the electron or photon or HT trigger to be fired as well.

For this measurement, dedicated samples of di-leptonic $t\bar{t}$ MC are used, while for the data the same samples used for the analysis have been considered. To have the most precise tuning of the MC samples to reproduce the data, the trigger efficiency has been derived with p_{T}-$\lvert\eta\rvert$ dependence for muons and $\lvert\eta\rvert$-p_{T} for electrons. The different treatment for muons and electrons is done to be consistent with other corrections provided by the CMS Collaboration. From Fig. 6.2 to Fig.6.4 the trigger efficiencies and scale factors for the three years are shown.

The physics objects selected according to the rules above represent the ingredients needed to reconstruct the W boson and the t quark in the first place, and eventually the W' boson candidate. The events are then classified according to the number of $b-$tagged jets, and whether one of them is the jet used to reconstruct the t quark or the W' boson. Finally, several regions are defined to perform the signal extraction, with additional requirements to further reduce the background contamination in the regions enriched in signal events.

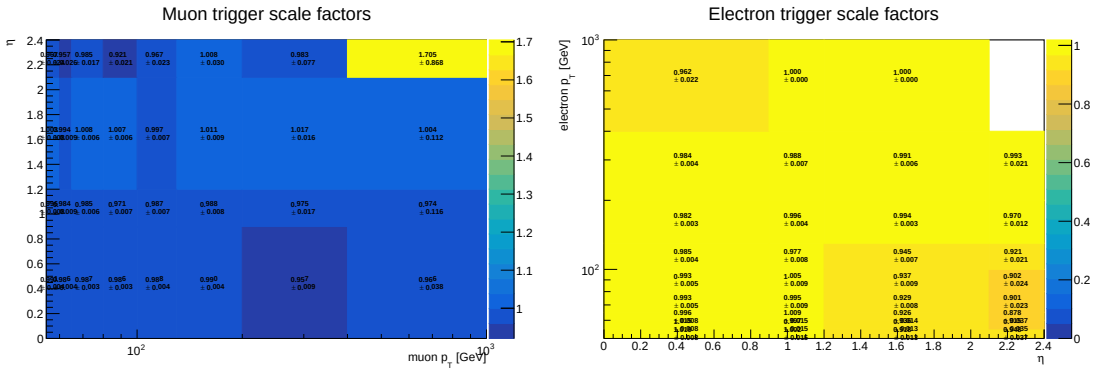


Figure 6.2: Trigger scale factor in 2016 for muons (left), and electrons (right).

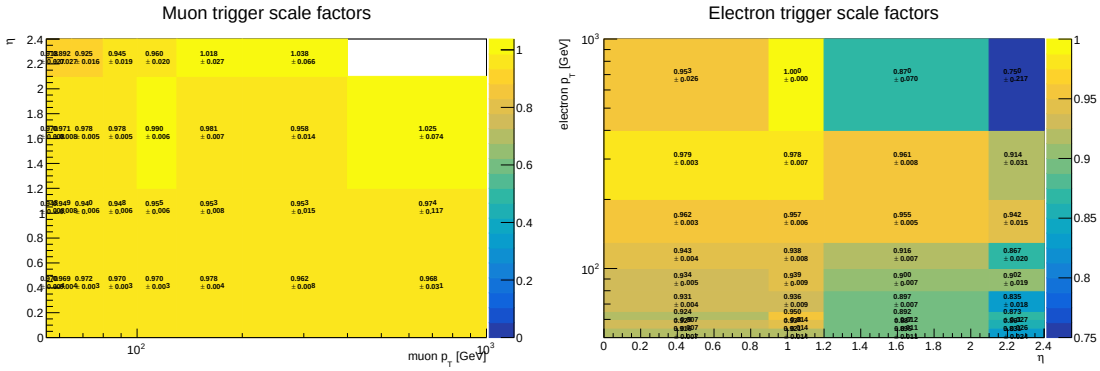


Figure 6.3: Trigger scale factor in 2017 for muons (left), and electrons (right).

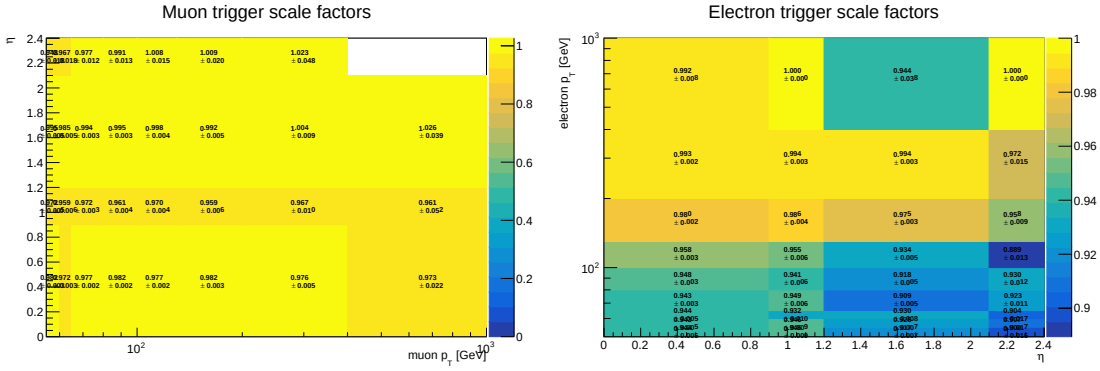


Figure 6.4: Trigger scale factor in 2018 for muons (left), and electrons (right).

6.5 Event reconstruction

In order to fully reconstruct a suitable W' candidate, it is necessary to optimally exploit all the information stored in the selected events with the goal the reconstruct the full W' decay chain of interest. In the following, an original procedure to obtain a W' candidate for each event, totally relying on the kinematic properties of such a decay process, is illustrated.

6.5.1 W boson reconstruction

The first step in the reconstruction of the t quark from its decay products is the reconstruction of the W boson produced by the decay. Assuming that the x and y components of the missing transverse energy are entirely due to the escaping neutrino, the invariant mass of the system $\ell\nu$ present in the final state can be constrained to be equal to the nominal W boson mass value:

$$m_{\mathrm{W}}^2 = \left(E_\ell + \sqrt{p_{\mathrm{T}}^{miss\,2} + p_{z,\nu}^2}\right)^2 - \left(\vec{p}_{\mathrm{T},\ell}^{\,2} + p_{\mathrm{T}}^{miss\,2}\right)^2 - \left(p_{z,\ell}^2 + p_{z,\nu}^2\right)^2. \tag{6.2}$$

where m_{W}^2 is the W boson mass, E_ℓ, $p_{\mathrm{T},\ell}$ and $p_{z,\ell}$ are the energy, transverse and longitudinal lepton momentum components, respectively, while the same quantities with the subscript ν refer to the neutrino components. In this expression, the only unknown quantity is the neutrino momentum component z, $p_{z,\nu}$, and so it can be extracted by solving the second-order equation. If the lepton mass is neglected, two solutions are obtained:

$$p_{z,\nu} = \frac{\Lambda \cdot p_{z,\ell}}{p_{\mathrm{T},\ell}^2} \pm \sqrt{\frac{\Lambda^2 \cdot p_{z,\ell}^2}{p_{\mathrm{T},\ell}^4} - \frac{E_\ell^2 \cdot p_{\mathrm{T}}^{miss\,2} - \Lambda^2}{p_{\mathrm{T},\ell}^2}} \tag{6.3}$$

where:

$$\Lambda = \frac{m_{\mathrm{W}}^2}{2} + \vec{p}_{\mathrm{T},\ell} \cdot \vec{p}_{\mathrm{T}}^{miss} \tag{6.4}$$

In 75% of the cases, this leads to two real solutions for $p_{z,\nu}$. Introducing the following quantity:

$$\chi_{m_W} = m_W - \overline{m}_{Wi} \tag{6.5}$$

where $\overline{m}_{Wi}$ is the mass of the reconstructed W boson with the $i-$th solution for $p_{z,\nu}$, then the p_{z,ν_i} value that minimizes χ_{m_W} is chosen.

For the remaining 25% of events the discriminant in Eq. 6.3 becomes negative, or equivalently the reconstructed W transverse mass m_{T}^W is greater than m_W, leading to complex solutions for $p_{z,\nu}$ due to the finite p_{T}^{miss} resolution. The lepton momentum resolution and the finite W intrinsic width give negligible contributions to this effect. To deal with this situation, the imaginary component is canceled by modifying $p_{x,\nu}$ and $p_{y,\nu}$ so that $m_{\mathrm{T}}^W = m_{\mathrm{W}}$, while still respecting the m_{W} constraint in Eq. 6.2. This is achieved by imposing that the determinant, and thus the square root term in Eq. 6.3, is zero.

This condition gives a quadratic relation between $p_{x,\nu}$ and $p_{y,\nu}$, which gives two possible relations between the two transverse components:

$$p_{y,\nu} = \frac{m_W^2 p_{x,\ell} + 2p_{x,\ell}p_{y,\ell}p_{x,\nu} \pm m_W p_{T,\ell}\sqrt{m_W^2 - 4p_{x,\nu}p_{x,\nu}}}{2p_{x,\ell}^2} \tag{6.6}$$

and one remaining degree of freedom.

Using the Eq. 6.6, the $p_{x\nu}, p_{y\nu}$ pair minimizes the distance between $p_{T,\nu}$ and the observed

p_{T}^{miss}, defined as:

$$\Delta^2 = (p^{miss}{}_x - p_{x,\nu})^2 + (p^{miss}{}_y - p_{y,\nu})^2 \tag{6.7}$$

is retained and used instead of $p_{\mathrm{T}}^{miss}{}_x, p_{\mathrm{T}}^{miss}{}_y$ in Eq. 6.2 to finally obtain a solution for $p_{z,\nu}$. This strategy returns the closest solution (i.e. it has a negligible difference with) to the measured p_{T} in the 60% of cases with negative discriminant.

After obtaining the full four-momentum of the escaping neutrino, the W boson is reconstructed as the vectorial sum of the four-momenta of the neutrino and the charged lepton.

6.5.2 t quark reconstruction

After reconstructing both the unobserved neutrino and the W boson, the next step is to evaluate the t quark four-momentum. For this purpose, it is crucial to choose the AK4 jet, called $\mathrm{jet_{top}}$, to couple with the reconstructed W boson. Taking into account the characteristics of a t quark decay and the energy scale of the process from which it comes, the following three criteria are used in parallel for the coupling:

- **mass criterion**: since the W boson and the candidate jet are products of the t quark decay, one expects the invariant mass of the sum of the corresponding four moments to be as close as possible to the nominal value of the top quark mass, taken as 172.5 GeV. The $\mathrm{jet_{top}}$ can be chosen as the one that minimizes the quantity:

  $$\chi_{m_{\mathrm{t}}} = |m_{\mathrm{t}} - m_{\mathrm{t},i}| \tag{6.8}$$

  where $m_{\mathrm{t},i}$ is the invariant mass of the system consisting of the reconstructed W boson and the $i-$th AK4 jet. The effects of experimental and theoretical uncertainties on the t quark mass are expected to be orders of magnitude smaller than the experimental resolution of the t quark mismatching effects.

- **closest criterion**: The jet produced by the b quark is expected to be very close to the lepton produced in semileptonic top quark decay. Then the t quark four-momentum can be obtained by combining the reconstructed W and the jet closest in ΔR to the lepton.

- **subleading criterion**: Considering the production and decay chain investigated in this study, the jets originate from either one of the two b quarks produced in the W' decay final state. Given the energy scales of the hypothesized W' decay chain, the b jets are expected to be the two highest p_{T} jets in the event. In particular, the leading AK4 jet in p_{T} is expected to be produced by the b quark coming from the W' boson, the second (i.e. subleading) one by the b quark coming from the t quark decay. As a consequence, the subleading AK4 jet in p_T is chosen to properly reconstruct the t quark.

If the same AK4 jet satisfies at least two criteria, it is selected to reconstruct the top quark four-momentum. Instead, if three different AK4 jets are selected by the criteria, the one indicated by the mass criterion is selected as $\mathrm{jet_{top}}$.

6.5.3 W' candidate reconstruction

To reconstruct the W' candidate in the event, another AK4 jet is needed, called $\mathrm{jet}_{W'}$. This AK4 jet is selected as the one with the highest p_{T} in the jet collection after the $\mathrm{jet_{top}}$ is selected. The jet collection in which the $\mathrm{jet_{top}}$ and $\mathrm{jet}_{W'}$ are selected will only consist of AK4 $b-$jets if the event has at least two AK4 $b-$jets. Otherwise, the jet collection is composed of all AK4 jets in the event, regardless of their $b-$tagging values. This choice is motivated by the final state of the signal process, in which both jets should be $b-$jets. Thus, if a signal event has at least two AK4

jets that meet the $b-$tagging requirements, and one of them is not selected as the jet_{top}, this is much more likely to happen because it fails the selection criteria rather than the $b-$tagging algorithm.

6.6 Event categorization

An additional quantity that is defined and useful for distinguishing the signal from the background is the soft−drop mass of the AK8 jet associated with the $\text{jet}_{W'}$, called m_{SD}^{AK8}. To define it, the AK8 jet closest to the $\text{jet}_{W'}$ is the one with the smallest angular distance in the $\eta - \varphi$ plane, namely $\Delta R(\text{jet}_{W'}\text{ , AK8 jet})$, to the $\text{jet}_{W'}$. If $\Delta R(\text{jet}_{W'}\text{ , AK8 jet}) < 0.4$, the AK8 jet is called a matched AK8 jet (matchAK8).

The events are therefore divided into categories depending on whether the selected AK4 jets are also $b-$tagged. The categories are

- SR2B: at least two $b-$jets are reconstructed in the event, and both jet_{top} and $\text{jet}_{W'}$ are $b-$tagged jets; this is the main signal region.
- SRT: only 1 jet in the event is reconstructed as a $b-$jet, and it is the jet_{top}.
- SRW: the $\text{jet}_{W'}$ is a $b-$jet and it is the only $b-$jet in the event.
- CR0B: no $b-$jet is reconstructed in the event; this category has low signal contamination and is mainly used as a validation region for the analysis strategy.

Signal events may end up in the secondary control region CR1B if both jet_{top} and $\text{jet}_{W'}$ are not $b-$tagged, but an additional jet in the events identified as a $b-$jet was not selected as either jet_{top} or $\text{jet}_{W'}$. However, this is a relatively rare event. Since the composition of the backgrounds in this region is different from that in the signal regions defined above and the signal is very low, it is not used in the analysis. The category definitions are summarized in Tab. 6.5 for reference.

number of $b-$jets	jet_{top} is $b-$jet	$\text{jet}_{W'}$ is b-jet	type of region
	Control regions		
0	no	no	Control region (CR0B)
1	no	no	1-bjet control region (CR1B)
	Secondary signal regions		
1	yes	no	top jet SR (SRT)
1	no	yes	W' jet SR (SRW)
	Main signal region		
≥ 2	yes	yes	Signal region (SR2B)

Table 6.5: Event categories defined in terms of the total number of reconstructed $b-$jets and the $b-$jettiness of the AK4 jets used to reconstruct the top quark and W' boson candidates.

6.6.1 Region definitions

After the event reconstruction and categorization, a constraint on the mass of the reconstructed top quark is applied to reduce the background contamination. In particular, the requirement $m_t > 120$ GeV significantly reduces the QCD background, while the requirement $m_t < 220$ GeV helps to reject W+jets events. This also allows removing the t quark candidates from $t\bar{t}$ semileptonic

events where either the $b-$jet association failed, or from rare dileptonic events that happened to pass the lepton veto. To further reduce the number of $t\bar{t}$ semileptonic events that pass the reconstruction stage, an additional constraint on m_{SD}^{AK8} is applied if the match succeeds. In this case, m_{SD}^{AK8} must be less than 60 GeV, otherwise, the event is discarded. This is motivated by the observation that the soft-drop mass of an AK8 jet from a t hadronic quark decay is expected to be centered around the nominal t quark mass. By requiring low values of the soft-drop mass of the AK8 jet associated with the decay of the W' boson, signal purity is improved.

From Fig. 6.5 to Fig. 6.8 the W' mass for the CR0B and the three SRs are shown after applying the W+jets and the $t\bar{t}$ vetoes. As can be seen from these plots, the most relevant backgrounds are $t\bar{t}$ and W+jets, as well as the single-top quark productions in SR2B. In particular, the $t\bar{t}$ background is more abundant in the SRT and SR2B categories, while the W+jets is more abundant in the SRW category. Since there is a $b-$tagged jet in the SRW category, the W+jets background in this category is predominantly composed of a W boson associated with c or b quarks. For this reason, this contribution cannot be extracted from the CR0B category, where the dominant contribution comes from W boson production associated with light flavor jets.

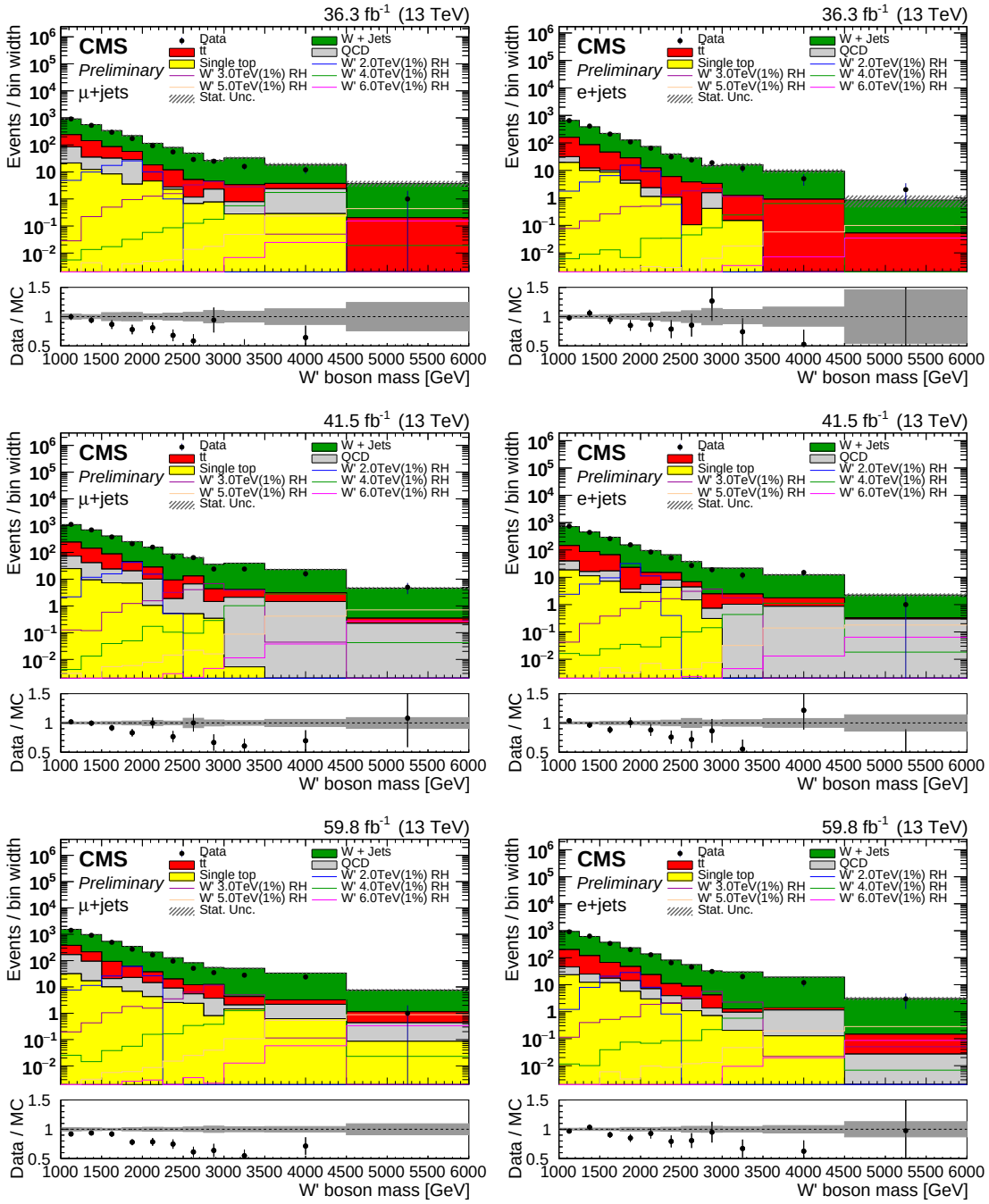


Figure 6.5: W' reconstructed mass for muon (left) and electron (right) channel in the CR0B region in 2016 (top), 2017 (mid), and 2018 (bottom).

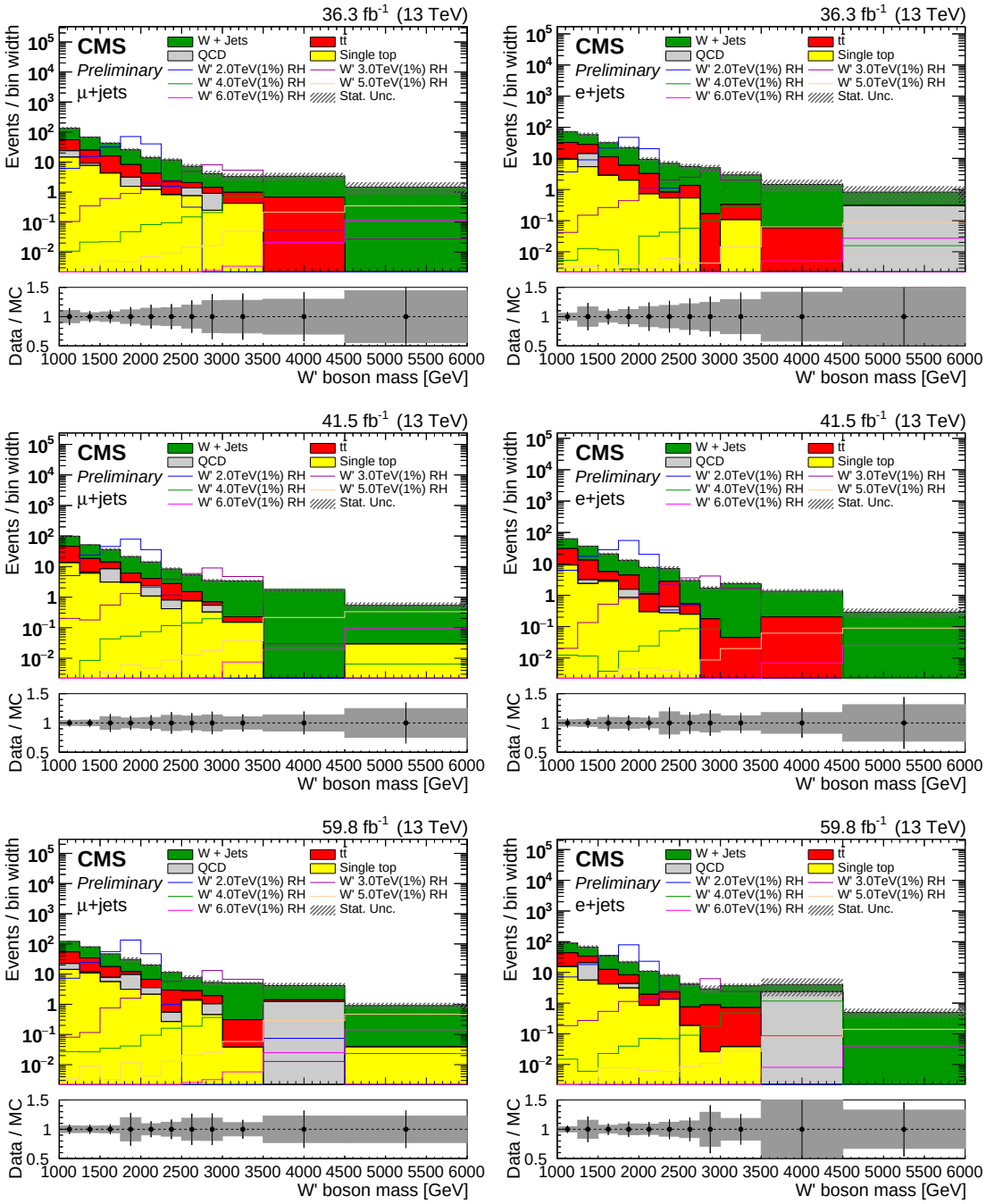


Figure 6.6: W' reconstructed mass for muon (left) and electron (right) channel in the SRW region in 2016 (top), 2017 (mid), and 2018 (bottom).

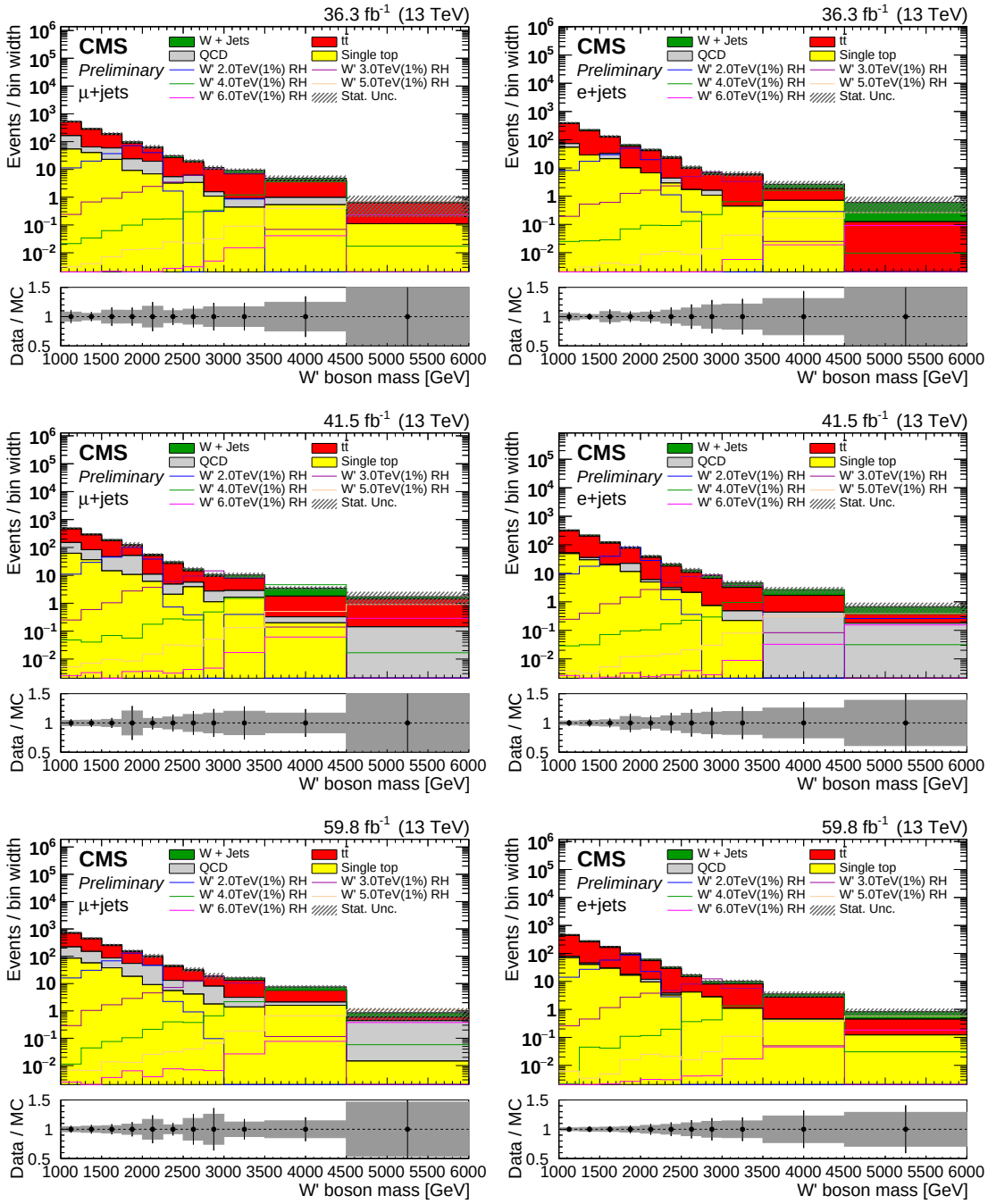


Figure 6.7: W' reconstructed mass for muon (left) and electron (right) channel in the SRT region in 2016 (top), 2017 (mid), and 2018 (bottom).

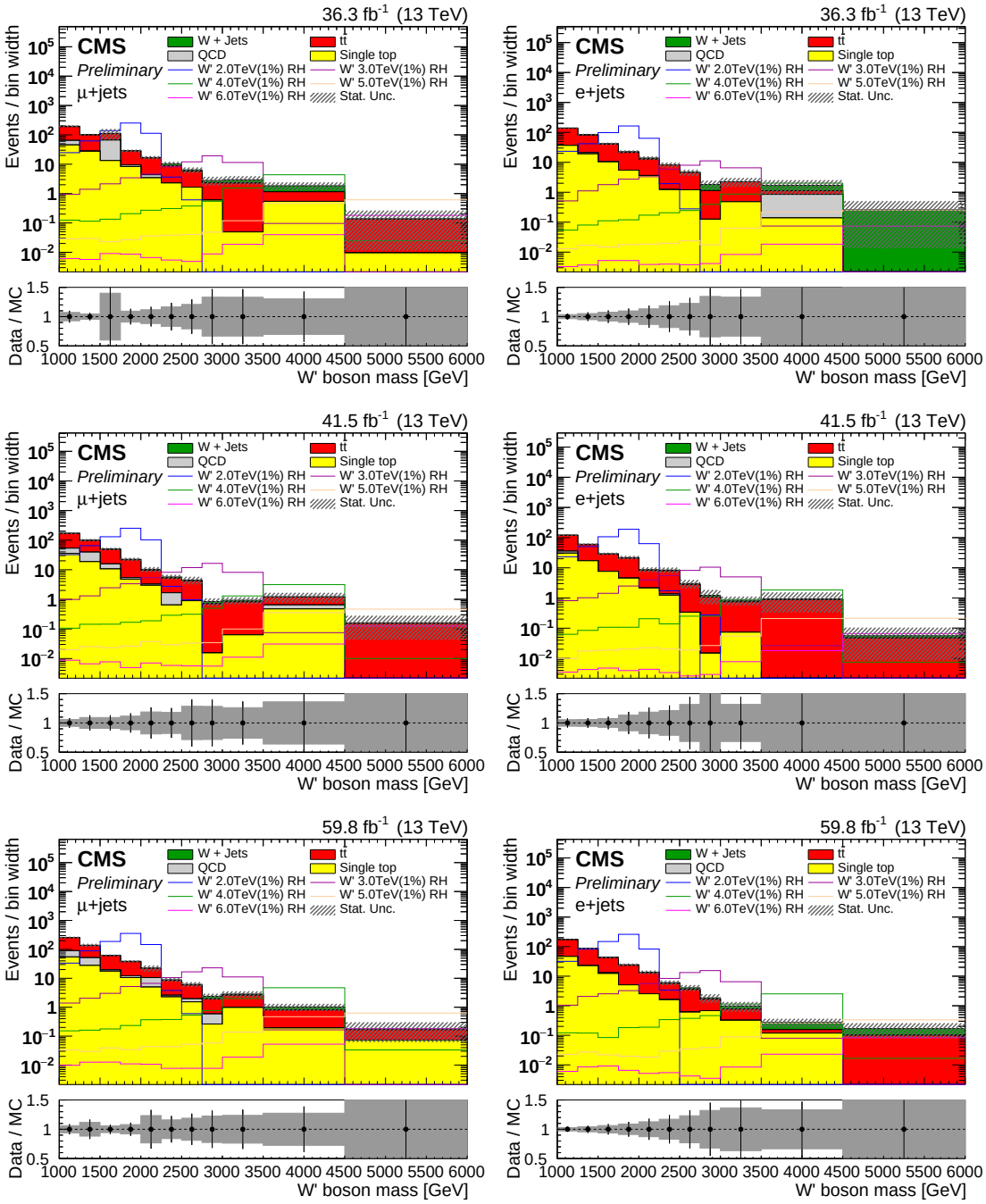


Figure 6.8: W' reconstructed mass for muon (left) and electron (right) channel in the SR2B region in 2016 (top), 2017 (mid), and 2018 (bottom).

6.7 Data-driven background extraction

In a very energetic phase space like the one considered in this search, Monte-Carlo simulations could not be accurate and reliable enough. The reasons behind this could be many, for example, the lack of the NLO corrections that might change the expected shape significantly, like in the case of the W+jets background [202]. An additional challenge is represented by the limited number of simulated Monte-Carlo events in such a phase space and the consequent worsening of the uncertainty once reweighting corrections are applied. For the W+jets background another limiting factor is the different composition in terms of jet flavor, which would require dedicated studies in very similar phase space to assess the same $\mathrm{c}-$ and $b-$jet multiplicities. For these reasons, a shape-based data-driven method for the derivation of the most relevant backgrounds. The advantages of this approach are:

- compensate for potential mismodelling;
- robust against Monte-Carlo statistical issues;
- robust against $b-$tagging and other experimental sources of uncertainty.

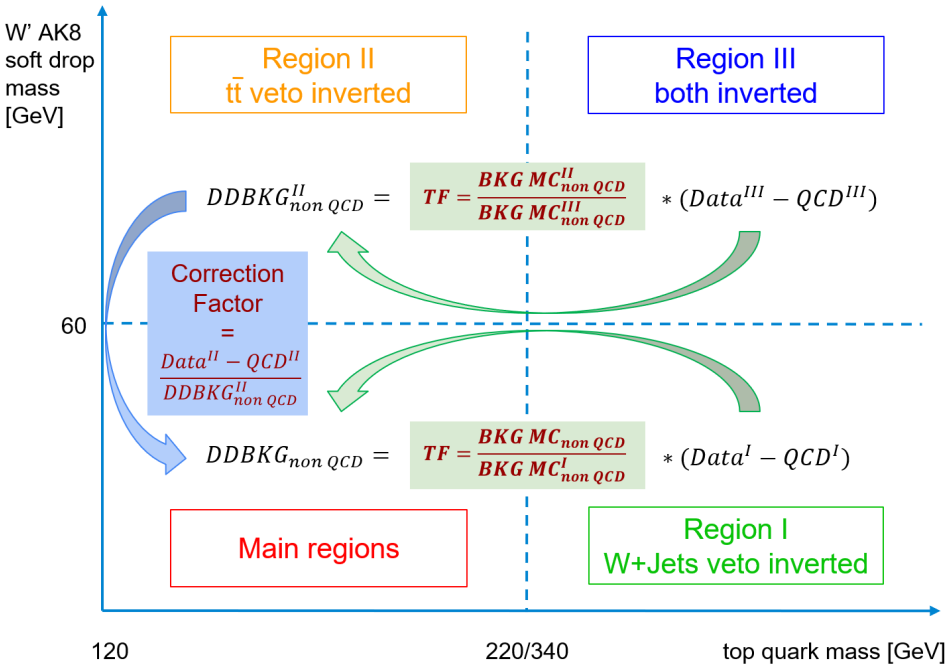


Figure 6.9: Sketch of the subregions definition and the scope of each one of them for the data-driven background extraction procedure.

A prime requirement for the application of the extraction procedure is to find one or more variables that are uncorrelated or, if anything, mildly correlated to the kinematic of the selected topology. As pointed out in the previous Section, in this search the chosen variables for the definition of main regions and subregions are the reconstructed t quark mass and the soft-drop mass of the AK8 jet associated with the $\mathrm{jet}_{W'}$.

For this purpose, additional control regions, also called subregions, are defined by applying requirements to the reconstructed m_t and/or m^{AK8}_{SD} to veto the signal. For CR0B, SRT, and SRW, inverting one of the two cuts is sufficient to deplete the subregions of the signal. Subregions with $m^{AK8}_{SD} > 60$ GeV, called subregions-II and III, turn out to have limited statistics for the W+jets background, and for this reason, they cannot be used to derive the background estimate for the

main regions. Instead, they are used to derive a correction and a cross-check of the background extraction procedure, which will be explained in the next Section. Conversely, the regions with $m_{SD}^{AK8} < 60$ GeV and $m_t > 220$ GeV, henceforth referred to as subregions-I, are more reliable for W+jets, $t\bar{t}$ and single-top (ST) quark production and will be used in the background studies. Fig. 6.9 reports a sketch with the definition of the subregions and their use in this search. The details of the extraction method are given below.

6.7.1 Background estimation

The background estimation procedure consists of comparing the shapes of both W+jets, $t\bar{t}$, and ST together in the main regions and their respective subregion with the $m_{top,best}$ cut inverted. The ratio of the W' mass spectra in one of the main regions to the corresponding subregions-I, as reconstructed with the Monte-Carlo samples, is obtained. This ratio is then fitted to obtain a transfer function, TF. The data-driven background is obtained from the data after subtracting the non-relevant background, i.e. the QCD multijet production, via its Monte-Carlo estimates in subregions-I and transferred to the main region using the TF. This allows the prediction of the total background spectrum in the signal regions from the data of subregions-I. The procedure can be summarized in formulas as follows:

$$\mathrm{DDBKG_{nonQCD}} = TF(\mathrm{Data}^I - \mathrm{QCD\,MC}^I), \tag{6.9}$$

where:

$$TF = \frac{\mathrm{BKG\,MC_{nonQCD}}}{\mathrm{BKG\,MC_{nonQCD}}^I}. \tag{6.10}$$

where "DDBKG" represents the results of the data-driven extraction. The data-driven background extraction is applied separately for the muon and electron channels. For each of these channels, the data collected during the three years of data acquisition are summed before applying this procedure. Fig. 6.10 shows the ratio of the reconstructed W' mass between each category and its corresponding subregion-I, the analytic form for the fit function, and the fit results. The fit function is $ae^{bx} + c + dx$ for both $t\bar{t}$ and W+jets dominated subregions. The muon and electron channels show similar behavior, although not quantitatively the same due to a difference in the selections: subregions-I contain lower-p_{T} leptons, and the trigger efficiency for low-p_{T} electrons is greater with respect to the muons, see also Sec. 6.4.5.

Before applying the TF, the $m_{W'}$ data spectra in subregions-I are fitted with a function also of the form $ae^{bx} + c + dx$, and the number of events in each bin is evaluated from the function integral. Statistical uncertainties on this fit are propagated to the corresponding data-driven distribution in the main regions. This method is preferred to simply use the data histogram as a template because it allows for smoothing variations in the tail of the distribution and reduces bias in the fit. The Monte-Carlo simulated QCD multijet events (QCD MC) are then subtracted from the resulting histogram in the main regions. Fig. 6.11 shows this fit of the data in the different subregions-I to the function in the different subregions for muons or electrons. The χ^2 of the fit is led by the first bins, which have much higher statistics but are less important for signal extraction. Therefore, alternative functions where the fit starts after the first bin and an alternative fit model based on a Landau function are considered for the study of the uncertainties related to this estimation procedure.

In addition, the functional dependence from the W+jets, $t\bar{t}$, and ST composition is added as a nuisance parameter in the statistical fit to extract the sensitivity to the search signal, as well as an additional modeling uncertainty, as reported in the following.

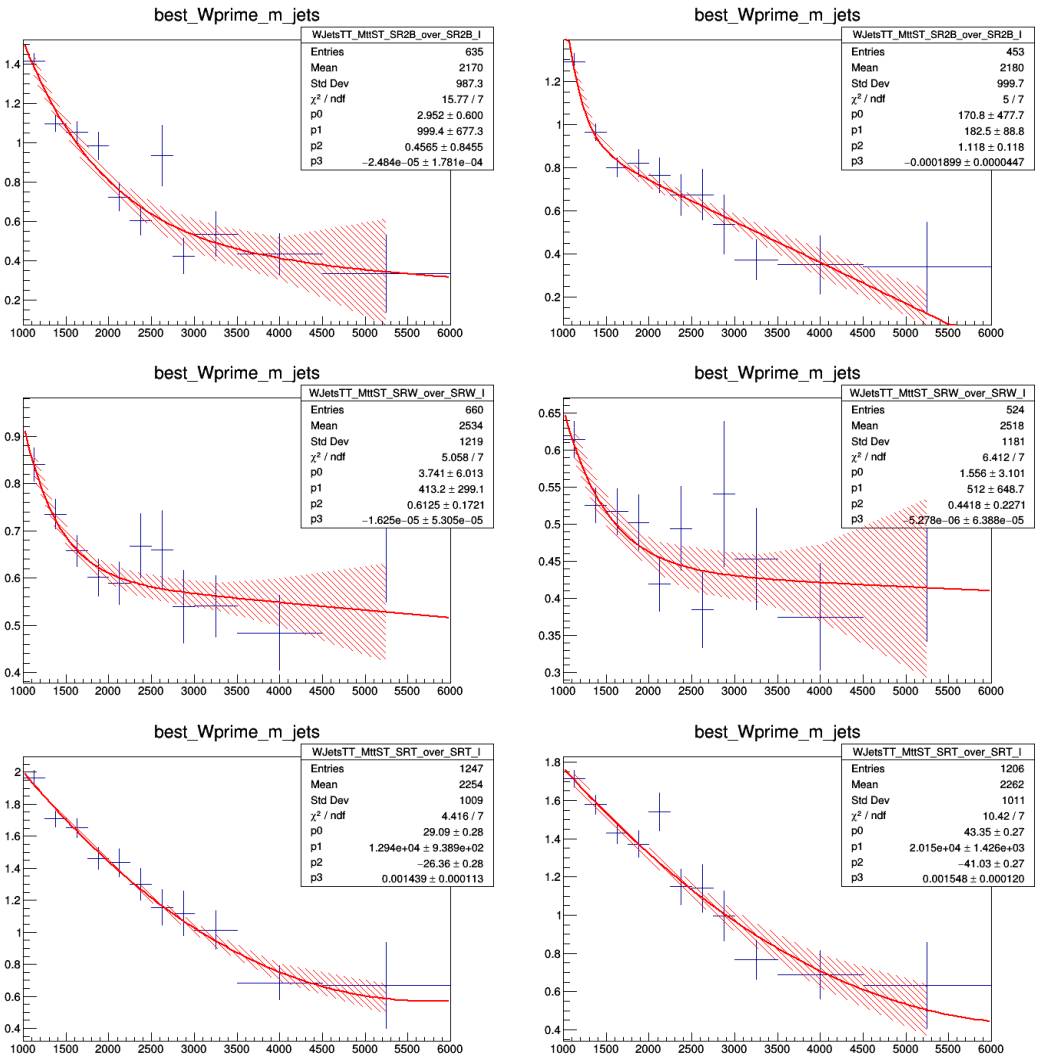


Figure 6.10: Ratio of the reconstructed W' mass between each category and its corresponding subregion-I, the analytic form for the fitted function, and the fit results for muon (left) and electron (right) for SR2B (top), SRW (middle), and SRT (bottom). On the $x-$ axis: the reconstructed mass of the W' boson in GeV.

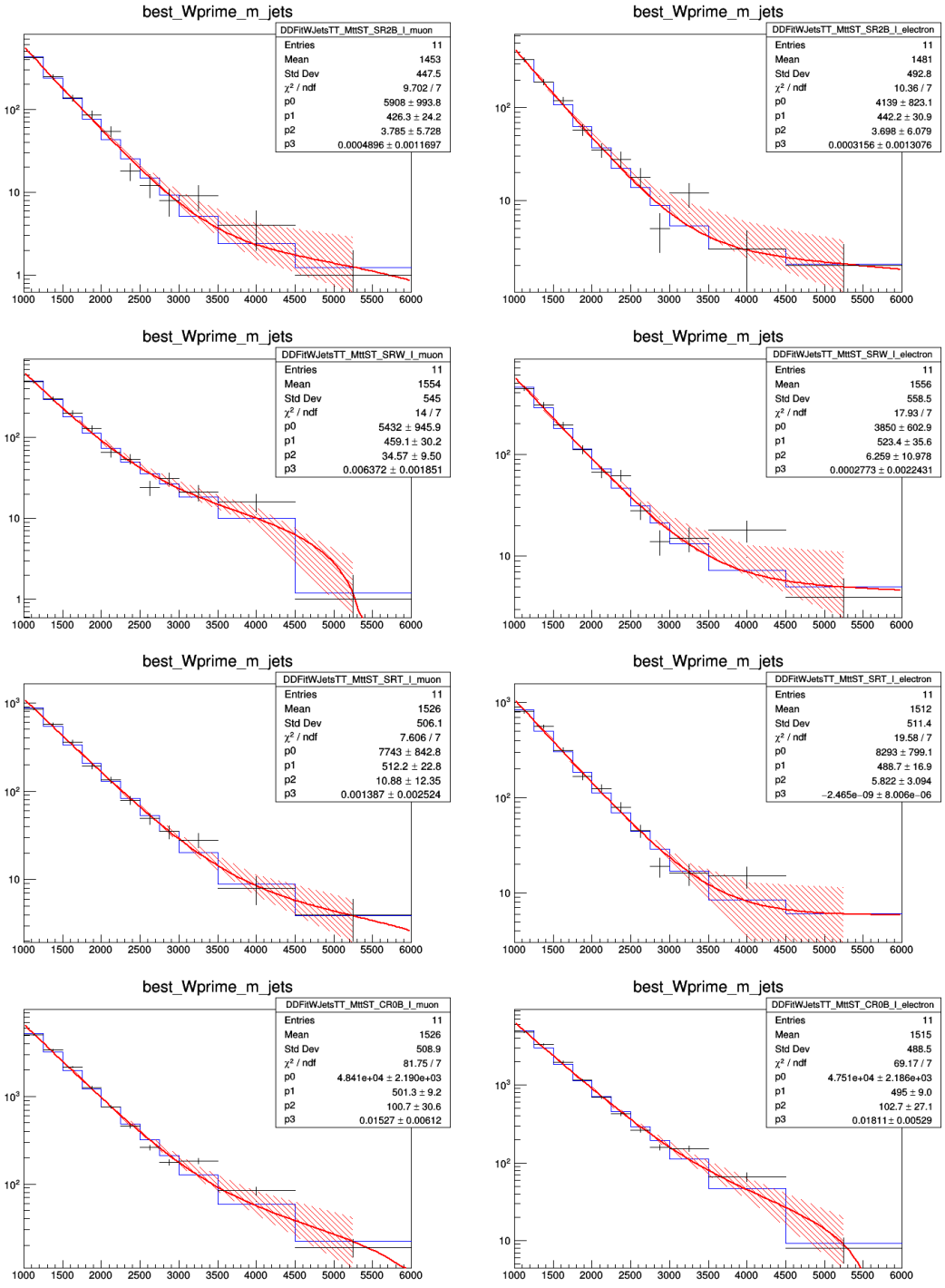


Figure 6.11: Fit to data in subregions-I after subtracting the QCD MC, for SR2B (first row), SRW (second row), SRT (third row), and CR0B (fourth row) for muons (left) and electrons (right). On the $x-$ axis: reconstructed mass of the W' boson in GeV.

6.7.2 Background estimation uncertainties and cross-checks

The data-driven extraction has four potential sources of uncertainty:

1. statistical uncertainties in data subregions-I;
2. statistical uncertainty on the simulations used for the TF;
3. modeling uncertainty on the fits in subregions-I;
4. modeling uncertainty on the TF.

The latter is the most important: since the TF are taken from simulations, they are sensitive not only to the choice of function made for the fit but also to possible mismodeling in the simulations. First, the statistical and modeling uncertainties in the function choice are investigated, and then the uncertainties in the background composition are estimated and additional cross-checks in the data are performed. In both the TF and subregions-I fits, the statistical uncertainties are evaluated from the fit parameters. A single uncertainty is derived by considering the full correlation matrix of the fit parameters.

Modeling uncertainties for the choice of fit functions for the subregions-I and TF are considered by taking the full difference with an alternative modeling function, where for W' masses below 1.5 TeV it corresponds to the histograms without fitting them, and for values above 1.5 TeV, it is represented by the result of the fit to the exponential+linear function, as described in Sec. A.2.1. This uncertainty takes into account the variations induced by choosing a different starting point for the functional form adopted to extract the nominal contributions of the background in the signal regions. To account for different possible trends in the data localized in the high $m_{W'}$ portion of the spectrum, additional modeling uncertainties are considered in the choice of fit functions for the subregions-I by using a Landau as the fit function and by taking the full difference with the nominal function used for background extraction in the main regions.

Modeling uncertainties for the subregions-I fit are also evaluated by varying the background composition, as explained in Sec. A.2.2. A second type of modeling uncertainty is considered by performing the procedure on independent subregions rich in both $t\bar{t}$ and W+jets, and deriving a correction factor CF to the prediction directly from the data, as explained in Sec. A.2.4.

Further details on the estimation of these systematic uncertainties are fully reported in Appendix A.2 and summarized in Sec. 6.8.1.

6.8 Systematic uncertainties

All the systematic uncertainties introduced in Sec. 5.6 are included in the statistical inference that will be outlined in Sec. 6.9. In addition, some uncertainties specific to this study are considered and illustrated below. In general, the systematic uncertainties affect both the shape and the yield of the reconstructed W' distribution, unless otherwise stated in the description. In Tab. 6.6 all the systematic uncertainties included are listed, together with the correlation among years and their effect.

6.8.1 Data-driven background estimation uncertainties

The uncertainties of the background estimation procedure are described in detail in Sec. A.2. Here the different components are summarized. The following sources are considered:

- **Alt**: Alternative shapes of the exponential+linear function are also taken as uncertainties by varying the starting point of the fit to the data in the subregions-I. The fit is repeated

starting at 1.5 TeV and is used to infer only the number of events in the corresponding region. The rest of the distribution is taken directly from the histogram of the data. This is done to reduce the dependence of the fit on the low-mass region, which has the highest statistics and to separate the low-mass regime from the high-mass regime in the background. This uncertainty is then symmetrized to obtain a two-sided variation by taking into account the full difference from the nominal estimate.

- **Alt2**: Alternative shapes of the expo+linear function are also taken as uncertainties by varying the modeling of the fit to the data in subregions-I with a Landau function. The fit is repeated from the beginning and is used to infer only the number of events. The systematic is taken into account by taking the difference between the Landau description of the data with respect to the expo+linear one, to cover possible correlation effects between low and high $m_{W'}$ regions that are not correctly modeled by the nominal function. This uncertainty is then symmetrized to obtain a two-sided variation by taking into account the full difference from the nominal estimate.

- **CR**: As explained in Appendix A.2.4, a correction factor CF is introduced to account for the effect of different jet flavor compositions in the final states on the transfer function. An uncertainty on this CF is applied by considering the full difference between the final distribution before and after the correction is applied. This uncertainty is then symmetrized by considering the full difference from the nominal estimate.

- **TT, W+jets, ST**: The background composition in the MC samples is varied to account for possible mismodelling. The variation is by a factor of 1.8 down and 0.2 up for each background, separately. The entire procedure is repeated by varying each background in all regions, deriving a new TF between a main region and the related subregion-I, and applying the TF to the subregion-I distribution from the data once the QCD MC contribution is substracted.

- **TF**: Statistical uncertainties on the fits to the transfer functions are propagated to the final form. The uncertainty on the parameters obtained from the fit is propagated to the predicted number of events, taking into account the correlation between the parameters, to give a single uncertainty.

- **AltTF**: Alternative shapes of the transfer function are taken as uncertainties by varying the starting point of the fit to the main region over subregion-I ratio in the MC samples. The fit is repeated starting at 1.5 TeV and is used to derive only the ratio in the corresponding range. The rest of the distribution is taken directly from the ratio of the number of events in the MC histograms. This reduces the dependence of the fit on the low-mass region, which has the highest statistics and separates the low-mass regime from the high-mass regime in the background. This uncertainty is then symmetrized to obtain a two-sided variation by taking into account the full difference from the nominal estimate.

- **DD**: Statistical uncertainties on the fits to the subregions-I are propagated to the final shape. The uncertainty on the parameters obtained from the fit is propagated to the predicted number of events, taking into account the correlation between the parameters, so that a single uncertainty is obtained.

- **Background Rate**: In addition, to disentangle the normalization and shape effect of the composition uncertainty, the total background yield for each region is left free to float without any constraint.

Fig. 6.12 shows an example of the magnitude of the different systematic uncertainties for the signal, QCD, and data-driven background for muon and electron channel for the SR2B.

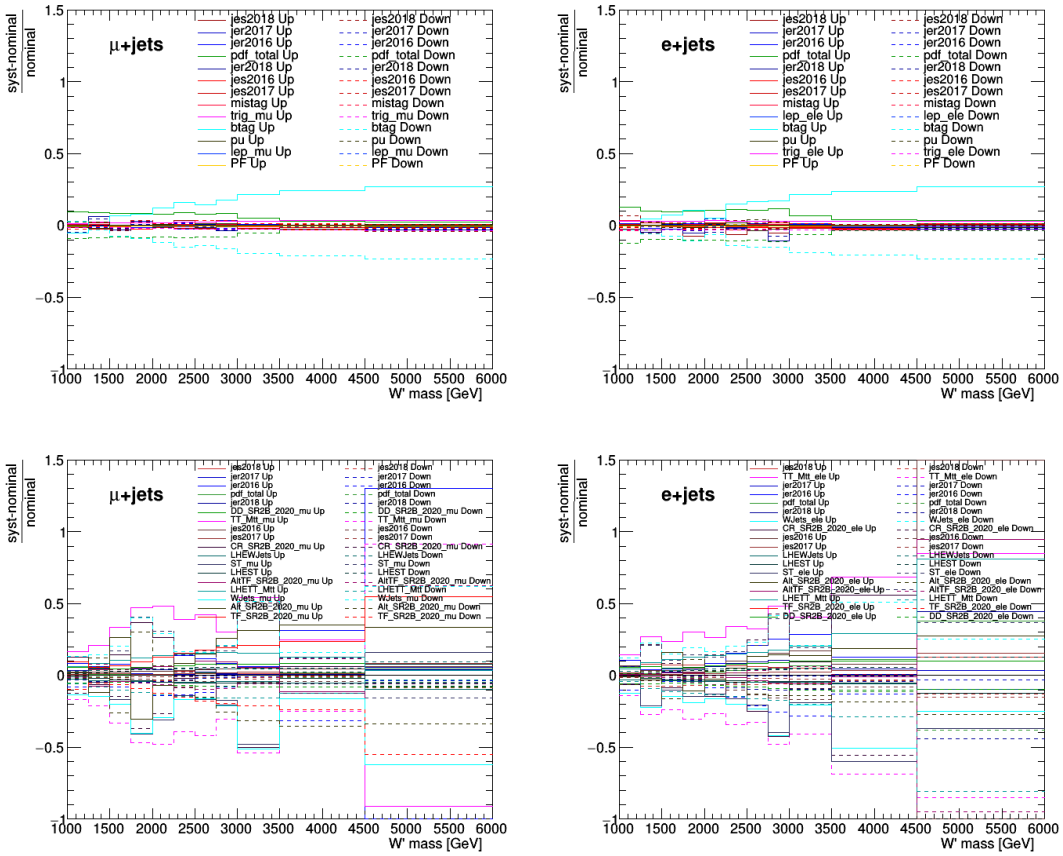


Figure 6.12: Pre-fit magnitude of the systematic uncertainties affecting a signal sample (5 TeV) (top row) and for the data-driven background (bottom row) in the muon (left) and electron (right) in the SR2B region.

Source	Process affected	Year correlation	Channel correlation
Luminosity	All	Yes	Yes
Pileup	signal, QCD	Yes	Yes
$b-$tagging and mistagging	signal, QCD	Yes	Yes
Trigger efficiency	signal, QCD	Yes	No
Lepton ID and MiniIso efficiency	signal, QCD	Yes	No
Prefiring	signal	No	Yes
Jet energy scale and resolution	signal, DDbkg	No	Yes
PDF uncertainties	signal, DDbkg	Yes	Yes
QCD scale uncertainty	QCD, DDbkg	No	Yes
QCD background yield	QCD	Yes	Yes
Transfer function uncertainty	DDbkg	—	No
Transfer function fit starting point uncertainty	DDbkg	—	No
Exponential+linear function uncertainty	DDbkg	—	No
Exponential+linear fit starting point uncertainty	DDbkg	—	No
Alternative Landau function uncertainty	DDbkg	—	No
ST background composition	DDbkg	—	No
$t\bar{t}$ background composition	DDbkg	—	No
W+jets background composition	DDbkg	—	No
Background rate	DDbkg	—	No
QCD MC statistical uncertainty	QCD	—	No

Table 6.6: Summary of systematic uncertainties. The second column indicates which estimates are affected by a particular uncertainty. The third column indicates whether a systematic uncertainty is considered to be fully correlated over the years of data collection or not. The fourth column indicates whether the uncertainty is correlated across lepton channels (muon or electron). Data-driven uncertainties are uncorrelated across fit regions, while experimental uncertainties are correlated across fit regions, and all except luminosity affect both the shape and rate of the distributions.

6.9 Statistical inference

After the requirements described in Sec. 6.6 have been applied, the W' reconstructed mass shows good discrimination between signal and background in both the signal and control regions. This variable was used in the simultaneous maximum likelihood fit for limit extraction implemented with the CombinedLimit tool [203] developed by CMS for the 2012 Higgs discovery. The current fit configuration includes the three SRT, SRW, and SR2B for both the muon and electron channels. The parameter r represents the signal strength for the signal process considered for the different hypothesized masses, defined as the ratio between the observed cross-section and the one expected from the theory.

6.9.1 Closure tests for background estimation

The data-driven background extraction is cross-checked by performing the fit in a control region to assess whether the uncertainties are sufficient directly on the data. The background model is taken as the prediction derived in Sec. 6.7. The W+jets, $t\bar{t}$, and ST estimates modeled by the data-driven background extraction are first verified by checking the distributions for the W' mass before and after the fit in the CR0B region. They are shown in Fig. 6.13 and demonstrate the reliability of the data-driven background estimate in a region completely depleted of signal contribution. In addition, a goodness-of-fit (GoF) test is performed with the correction procedure described in Sec. 6.8.1, using as the test statistic a likelihood ratio introduced by the "saturated model" [204]. Fig. 6.14 shows the results of the GoF test on the CR0B category observed with the full Run II dataset, demonstrating compatibility with the background-only hypothesis when the data-driven background estimation is used.

6.9.2 Statistical tests for the signal extraction

Several tests are performed to ensure that the statistical procedure applied to the signal regions is robust. First, the effect of systematic uncertainties treated as nuisance parameters, and the constraints on them from the fit are determined. The bias in the absence of a signal and the bias in the presence of a visible signal is also checked.

The impact of the nuisance parameters on the signal extraction is obtained by performing the fit on an Asimov data set based on the a-posteriori expectation obtained in the background-only scenario. In other words, all observables are set to the values obtained by fitting the background-only hypothesis in the fit region of interest. This check is performed in the blinded signal fit regions using the data-driven prediction as the background model.

The impact is evaluated by performing the full fit with each nuisance parameter frozen one at a time to its ± 1 standard deviation values as measured by the initial fit to the data. The change in the value of the signal strength r at each new minimum quantifies the effect on r of varying each perturbation parameter by ± 1 standard deviation. The pull gives the constraint on that particular perturbation from the data. A similar procedure is used to evaluate these effects with an unblinded fit to the data in the CR0B region.

Pseudo-experiments were generated either in the pure background hypothesis or by injecting different amounts of signal. Fig. 6.15 shows the bias of the fit for the hypothesis of $m_{W'} = 4$ TeV assuming 0 signal strength, a signal strength of 0.1, 0.32, and 0.6, corresponding to significance values of 1, 3, and 5 standard deviations, respectively. This test is performed in the blinded signal-fit regions using the data-driven prediction as the background model and confirms the absence of any bias in the measurement of the signal strength r induced by the background estimation.

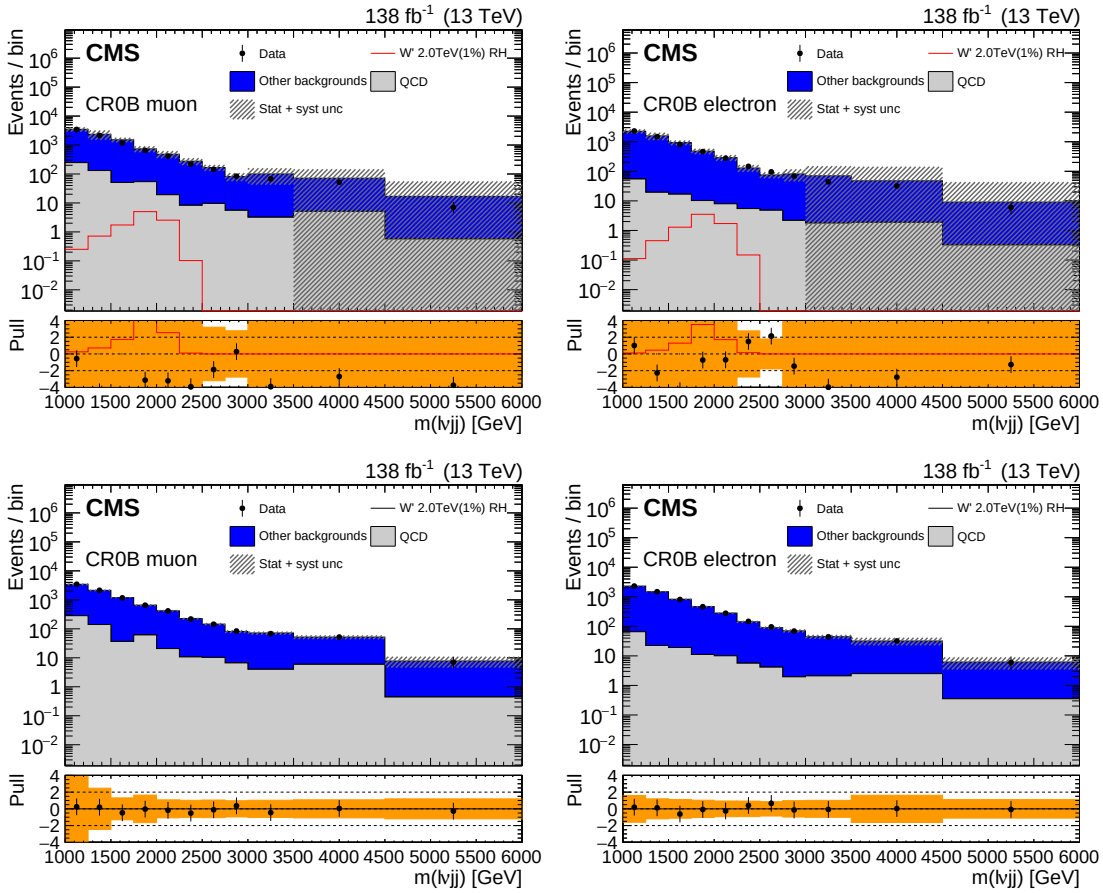


Figure 6.13: Distributions of the postfit W' boson mass in the CR0B control region for muons (left) or electrons (right).

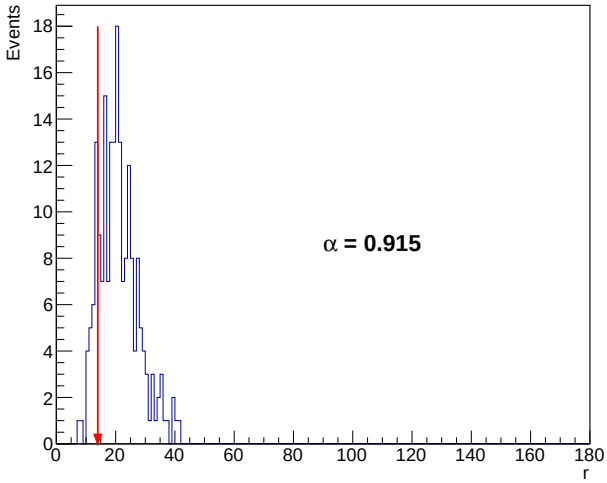


Figure 6.14: Results of the goodness of fit test performed in the CR0B category on the Run II dataset. α represents the $p-$value for the observed value of the test statistic r under study [204].

Fig. 6.16 shows an example of the impact plot obtained from the fit for a hypothesized mass of the W' of 3.8 TeV in the CR0B region (top) and in the three blinded SRs (bottom). The results are not significantly different between the two fits, showing the consistency of the background estimation performed for the analysis. The main uncertainties are the ones for the functional form of the fit to data (Alt, Alt2) and the relative background composition of the subregions-I (CR), the modeling of the main contributions to the background with the MC samples used to evaluate the TF, JES, JER, and the MC statistical uncertainties on the distributions for the QCD multijet process.

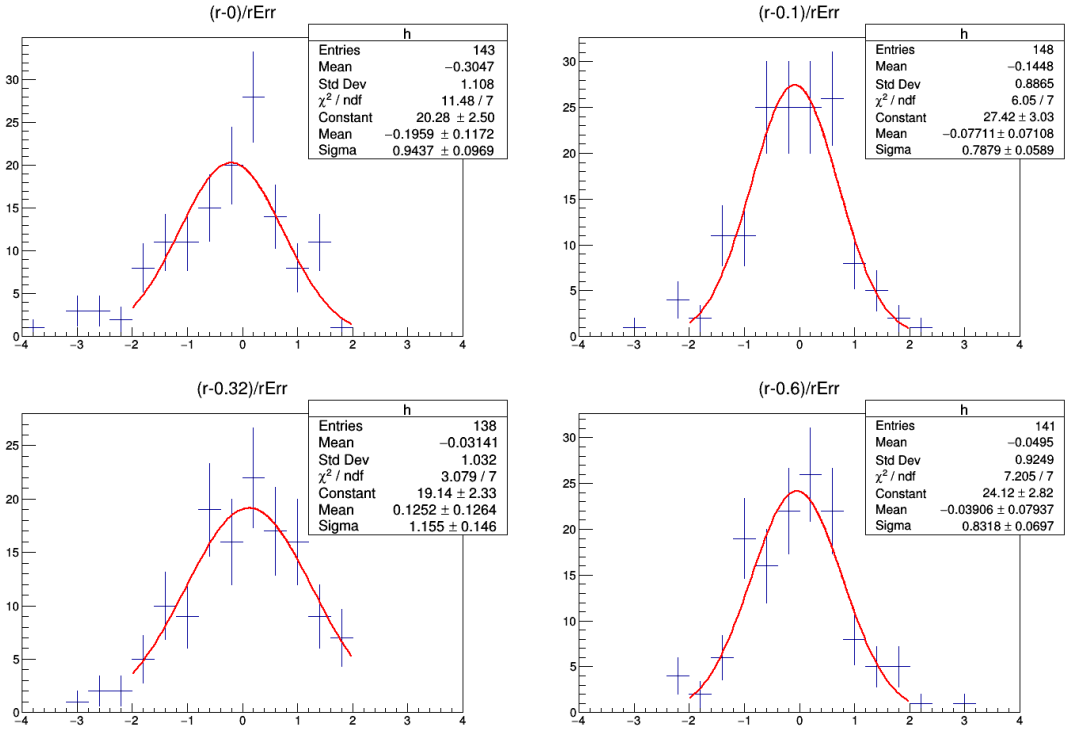


Figure 6.15: Results of the bias test performed on the signal sample at $m_{W'} = 4$ TeV with 1% width in the background only hypothesis (top left), and different amount of signal equal to 0.1 (top right), 0.32 (bottom left), and 0.6 (bottom right). The tests show that there is no significant bias.

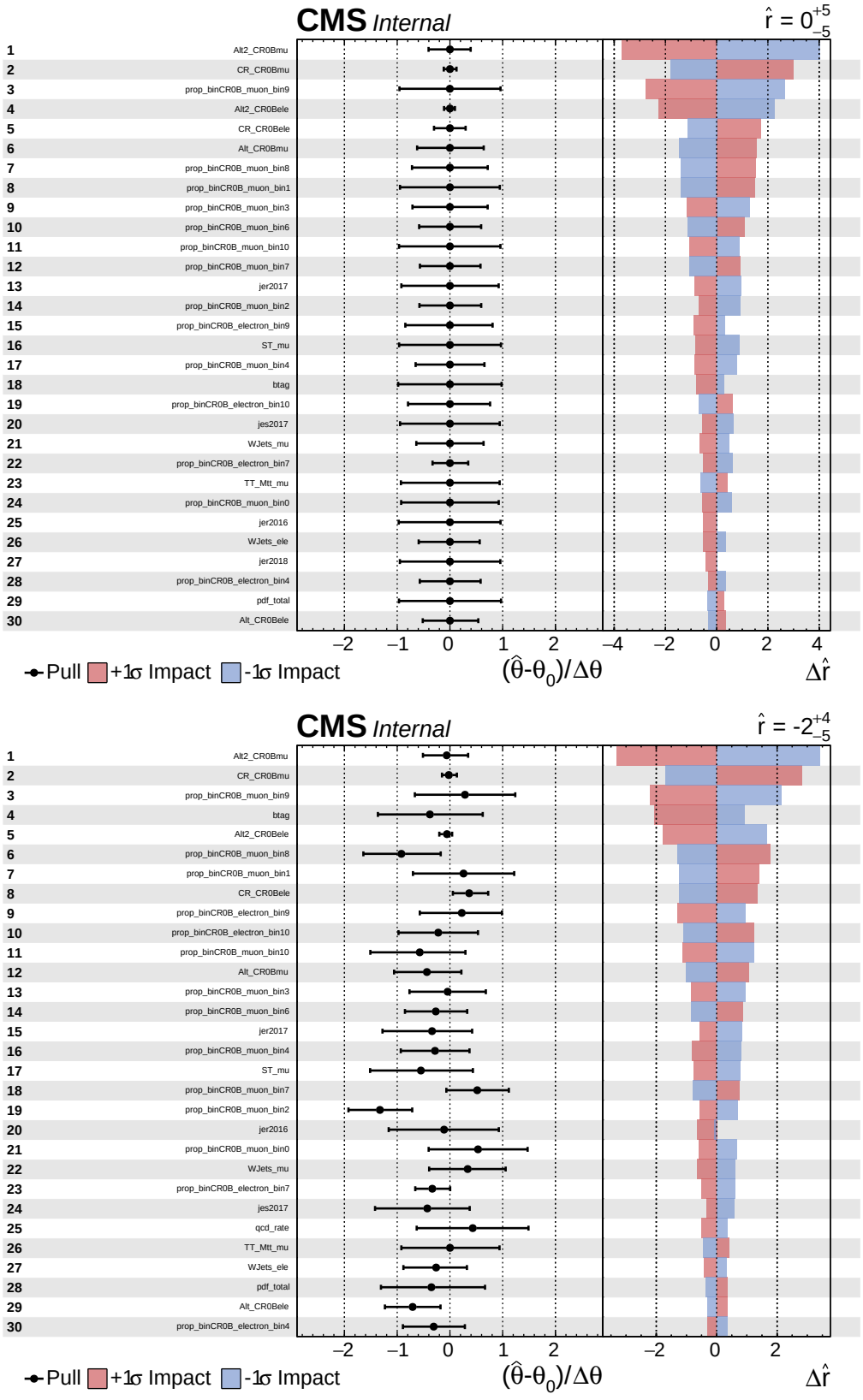


Figure 6.16: A-posteriori impact plot showing the list of systematics used in the fit in the CR0B region with an Asimov data set on top (bottom), the pull with respect to their pre-fit value, and the impact of the final uncertainty for a mass of 3.8 TeV. The label "prop_bin[region]_[channel+year]_[bin $i-$th]" indicates the perturbations associated with the MC bin-by-bin statistical uncertainty for each region, channel, year, and bin of the fitted distributions.

6.10 Results

The simultaneous maximum likelihood fit described in Sec. 6.10 is performed on the 138 fb^{-1}data set at a center-of-mass energy of 13 TeV. Upper bounds on the production cross-section are extracted by following the CLs prescription using the asymptotic approximation [205–207]. Cross-section limits are derived for each value of W' mass, width, and chirality, and compared with the prediction obtained as described in Sec. 5.2. Fig. 6.17 and Fig. 6.18 show for the RH and LH hypotheses, respectively, the upper bounds of the cross-section versus the W' boson mass in the scenarios with $\Gamma_{W'}/m_{W'}$ = 1%, 10%, 20%, 30%. LH and RH W' bosons with masses below 3.9 and 4.3 TeV are thus excluded with 95% confidence in the approximation of the narrow width of the new particle.

In addition, the observed results show a mild local excess in the range of 3.4−4.4 TeV, with a peak at 3.8 TeV with a significance of 2.6 σ. The excess is not observed in the CR, and there is no evidence of error in the statistical tests performed and reported in 6.9. A look-elsewhere-effect test [208] is performed on the results of the fit, and the probability of significance greater than 2.6σ is 1.5%, corresponding to a global significance of 2.0σ.

Following the procedure explained in Sec. 5.4 and summarized in the Eq. 5.3, observed 95% CL upper limits on the production cross section for a generalized left-right coupling of W' boson to t and b quarks are obtained in the narrow width hypothesis, for masses of the W' boson from 2.0 TeV up to 6.0 TeV in steps of 800 GeV. The limits expected and observed for masses of the W' boson from 2.0 TeV up to 6.0 TeV are reported from Fig. 6.19 to Fig. 6.20. Similarly, lower limits on the W' boson mass are also obtained for all the generalized left-right coupling of W' boson to t and b quarks tested. Fig. 6.21 shows the expected and observed limits on the mass of the W' boson.

Fig. 6.22 reports the postfit distributions for the fit at the same value of $m_{W'}$ in the 1% width hypothesis, with the signal strength fixed to the postfit results, i.e. $r = 0.29^{+0.13}_{-0.11}$. In the end, Fig. 6.23 reports the impacts obtained for the fit at $m_{W'} = 3.8$ TeV.

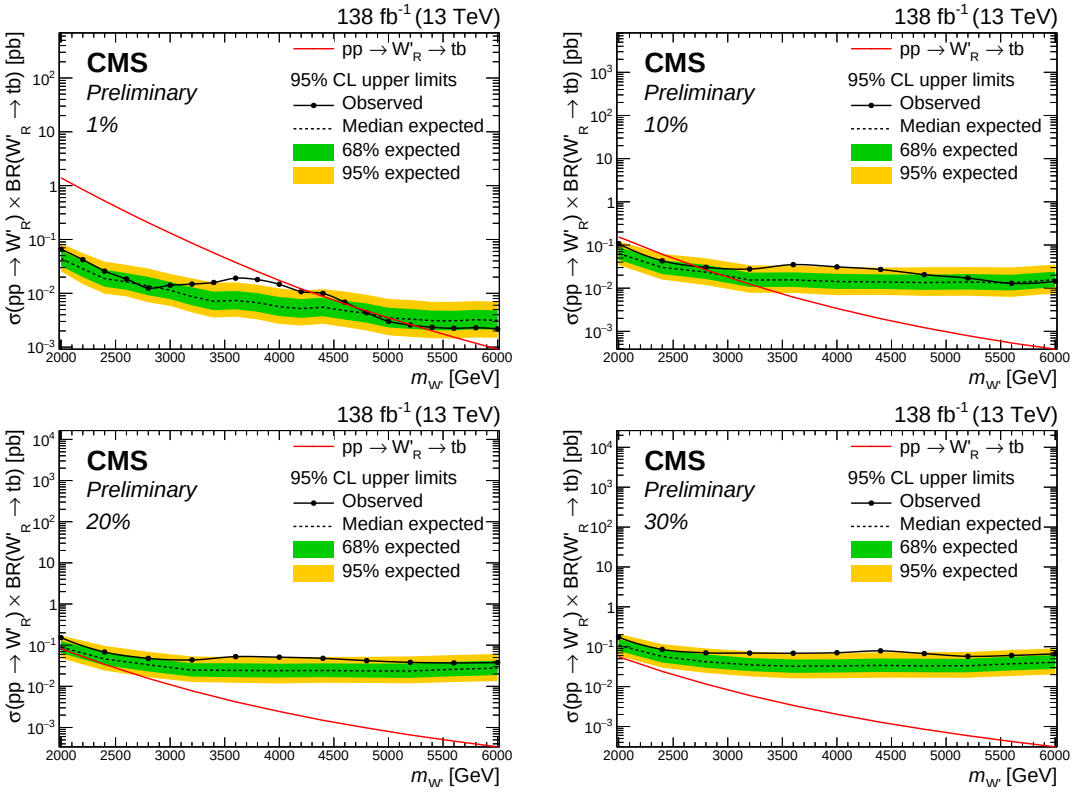


Figure 6.17: Results obtained by this analysis for the full Run-II dataset in the hypothesis of an RH W' with a width corresponding to 1% (top left), 10% (top right), 20% (bottom left) and 30% (bottom right) of its mass.

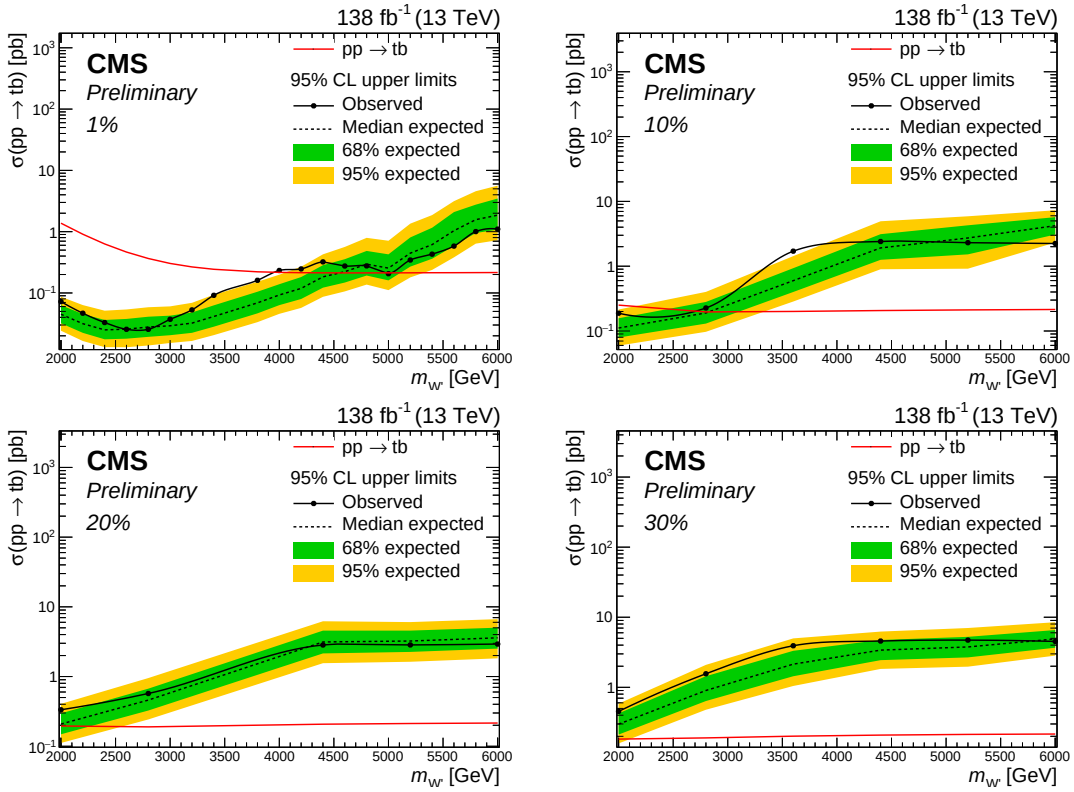


Figure 6.18: Results obtained by this analysis for the full Run-II dataset in the hypothesis of an LH W', including interference with the Standard Model, with a width corresponding to 1% (top left), 10% (top right), 20% (bottom left) and 30% (bottom right) of its mass.

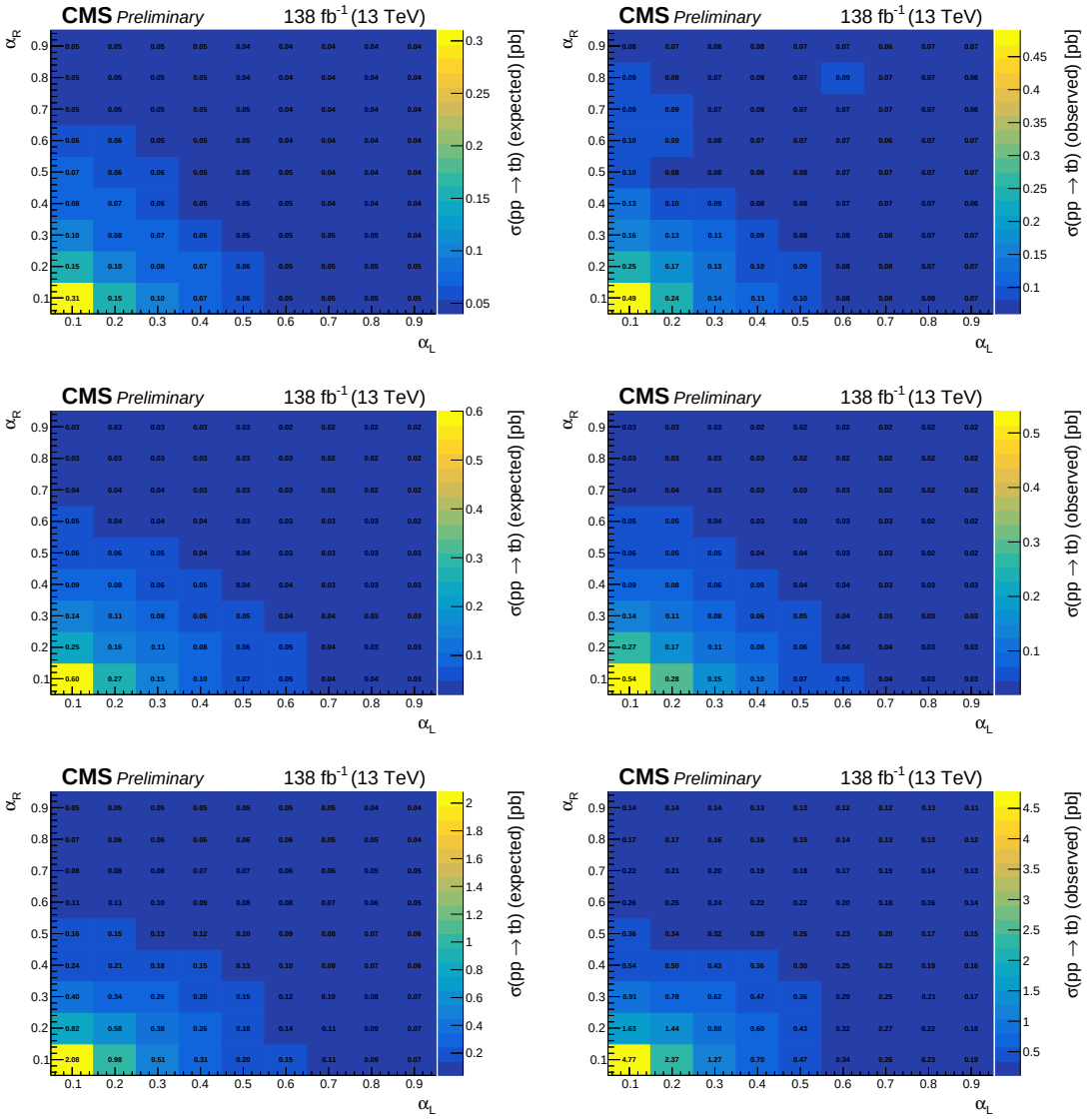


Figure 6.19: Expected (left) and observed (right) 95% CL upper limit on the production cross section for a generalized left-right coupling of W' boson to t and b quarks for a mass of the W' boson of 2.0 TeV (top), 2.8 TeV (middle), and of 3.6.0 TeV (bottom).

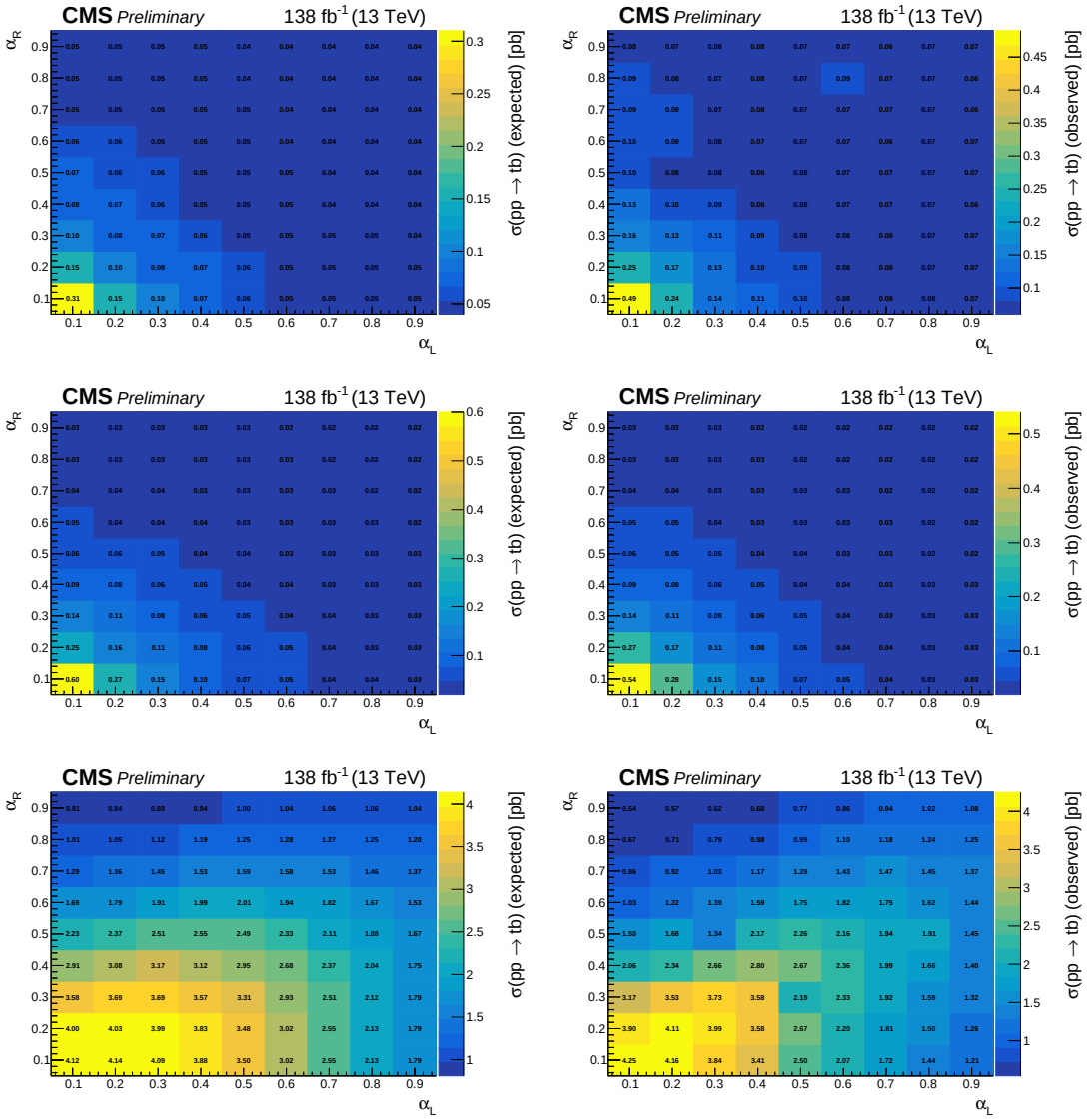


Figure 6.20: Expected (left) and observed (right) 95% CL upper limit on the production cross section for a generalized left-right coupling of W' boson to t and b quarks for a mass of the W' boson of 4.4 TeV (top), 5.2 TeV (middle), and of 6.0 TeV (bottom).

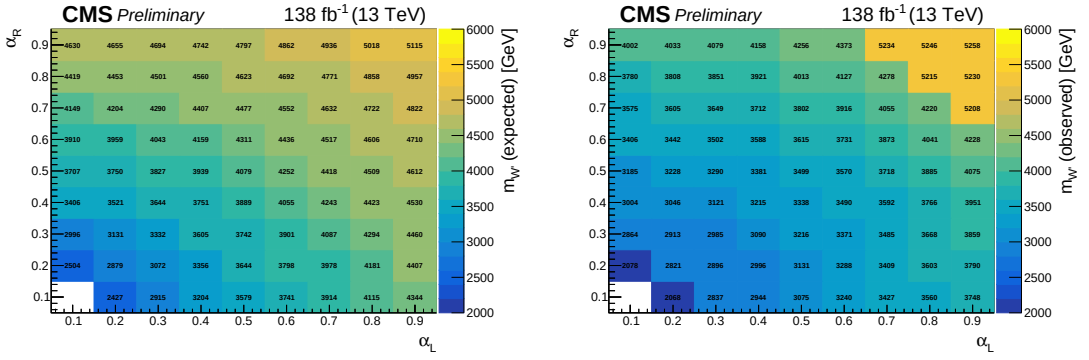


Figure 6.21: Expected (left) and observed (right) 95% CL lower limit on the W' boson mass for a generalized left-right coupling of W' boson to t and b quarks.

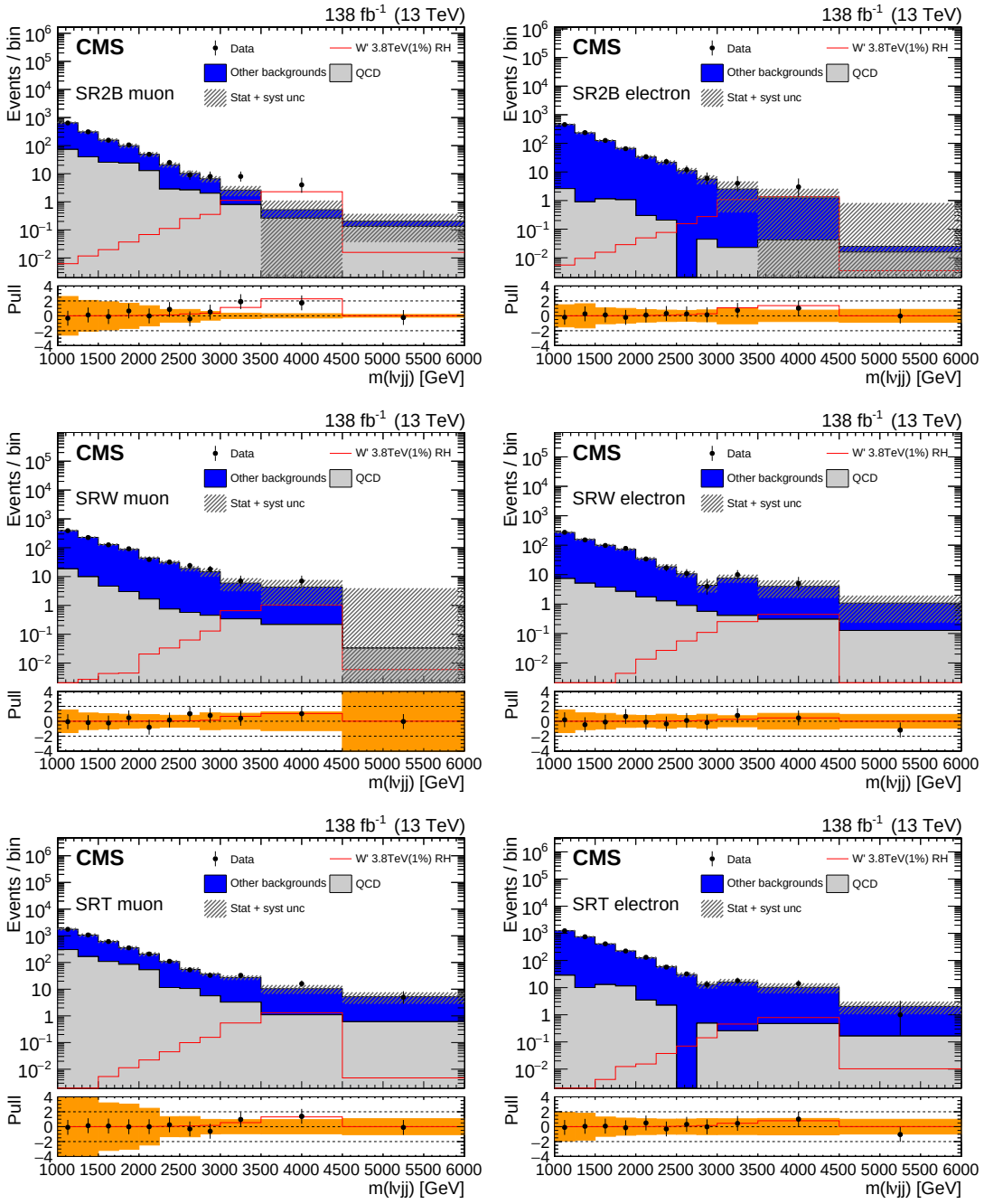


Figure 6.22: Postfit W' candidate mass distributions for the SR2B (top), SRW (middle), and SRT (bottom) categories the full Run II datasets for muon (left) and electron (right) channels.

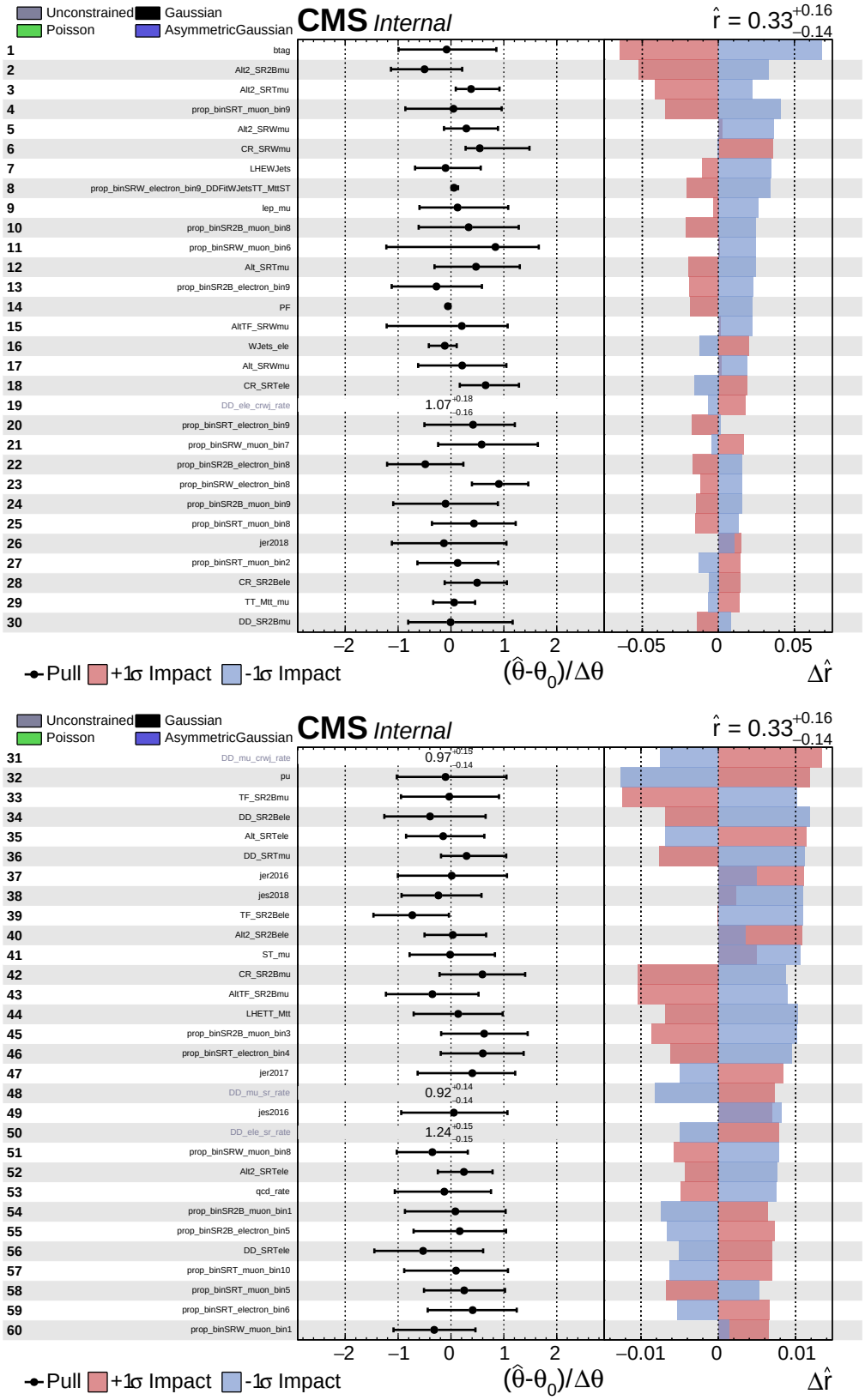


Figure 6.23: Impacts of the variation of systematic uncertainties on the final results of the fit in the SR categories for the 3.8 W' mass point.

Chapter 7

Same−sign W pair Vector Boson Scattering with hadronic tau in final state: SM and EFT studies

As illustrated in Sec. 1.6, the vector boson scattering (VBS) processes play a special role in the EWSB, since the unitarity of the longitudinal polarization scattering of vector bosons is only given thanks to the Feynman diagrams involving the Higgs field. Therefore, for any deviation from the SM couplings of the Higgs with the vector bosons, the VBS cross-section would diverge with increasing energy. The VBS is thus a probe to study the interplay between the Higgs sector and the electroweak one. Nevertheless, the divergence of the cross-section for VBS processes raises unitarity problems [209], giving great importance to the method chosen to recover the unitarity of the process cross-section at higher energies when calculating deviations from the SM. Among all the VBS processes, the scattering of a W pair of the same sign (ssWW VBS for short) is one of the most favored channels, since it is one of the channels with the highest cross-section and the largest cross-section ratio between electroweak (EW) and QCD production.

In the analysis presented in this Chapter, a new and so far unexplored final state is investigated by studying the decay of one of the scattered W bosons into a τ lepton, with the latter eventually decaying into hadrons (hadronic tau, τ_h). The final state thus consists of a charged light lepton $\ell = e, \mu$, the corresponding neutrino ν_ℓ, a τ_h, the corresponding ν_τ, and two jets produced by the quarks that emitted the W boson pair. The τ lepton is included for the first time in the study of VBS processes and represents an important probe of BSM physics due to its large mass and a preferential coupling to the Higgs boson (see Sec. 1.5).

A measurement of the process is proposed, as well as the evaluation of the sensitivity to indirect New Physics effects, by implementing the Standard Model Effective Field Theory (SMEFT) framework described in detail in Sec. 2.2. For this purpose, both a cut-based analysis and a machine-learning (ML) approach for the identification of the signal with the searched final state are presented to analyze the data collected by the CMS experiment during the Run-II data acquisition period, corresponding to an integrated luminosity of 138 fb^{-1}.

7.1 Signal and background processes

The main focus of this study is the scattering of a same-charge (or same-sign) W boson pair initiated by a quark pair, with one W decaying into a $\ell\nu_\ell$ pair, the other into a tau-flavored $\tau_h\nu_\ell$ pair, where τ_h denotes a hadronically decaying tau lepton. The case where the $\ell\nu$ pair is

produced in the decay of another τ lepton is also considered, extending the study to the final state $\tau_h\tau_\ell\nu_\tau\nu_\ell$.

As described in detail in Sec. 1.6, the characteristic kinematic properties of VBS events are high values of the invariant mass of the final state jets, m_{jj}, and well-defined angular separation, $\Delta\eta_{jj}$. Moreover, since no quark is produced by the scattered W pair, no hadronic activity is expected in the spatial region between the $\ell\nu_\ell$ and $\tau\nu_\tau$ pairs, and this feature, often called the "rapidity gap", can be exploited to increase the sensitivity to the signal.

In Fig. 7.1 the LO Feynman diagrams contributing to the proposed signal are shown: starting from the left, the first three are the pure electroweak (EW) contributions, while the one on the right is the irreducible contribution including QCD interaction vertices.

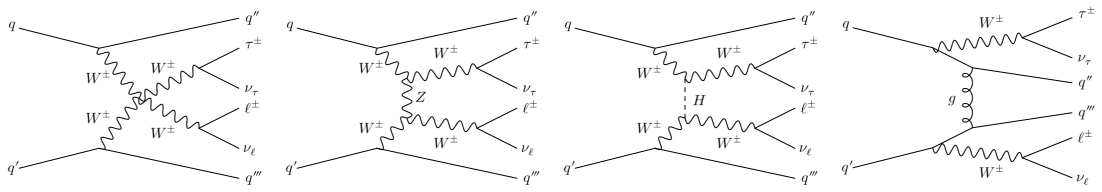


Figure 7.1: Feynman diagrams for purely electroweak VBS processes (starting from the left, the first three diagrams) and through QCD production (the last diagram to the right).

Following the convention of VBS studies, the pure EW diagrams are considered as the signal, while the QCD-mediated one is considered as the irreducible background. As pointed out in Sec. 1.6, this is feasible as long as only the LO is considered, given that at higher order interference between pure EW diagrams and QCD-mediated processes appears. Such terms however can be shown to give a negligible contribution [116], and are therefore not considered.

When the SMEFT framework is introduced to explore the indirect effects of New Physics, the modifications induced by the dim-6 and dim-8 operators illustrated in Sec. 1.5 enter the vertices of the SM LO diagrams, as shown in Fig. 2.4 and Fig. 2.5, respectively.

Any SM process that could mimic the final states represents a potential source of background for this analysis. The processes that make a non-negligible contribution as background in the final state are listed below in order of importance.

- **Non-prompt (or fake) leptons and taus**: events containing hadronic jets erroneously reconstructed as charged light leptons or hadronic taus are the dominant source of the background and are called fake or non-prompt leptons/taus (for short "non-prompt/fake leptons"). They are produced mainly by QCD-mediated multijet, associated W+jets, and hadronic and semileptonic $t\bar{t}$ pair production.
- **Dileptonic** $t\bar{t}$: the dileptonic channel decay of the $t\bar{t}$ quark pair production is the only one of this process to mimic the signal final state by having both the light lepton and the tau being prompt. The two b quarks and the oppositely charged lepton-tau pair produced by the decay of the t quarks are the main features that allow distinguishing this background source from the VBS signal.
- Other processes producing lepton-tau pairs of opposite sign (briefly, **Opposite Sign**): this category includes any process with an opposite-charge $\ell\tau_h$ pair other than Drell-Yan + jets and dileptonic $t\bar{t}$. This contribution comes from events where the charge of one of the leptons is misidentified and is mainly due to $WW \to \ell\tau_h\nu\nu$, vector boson fusion $WWqq' \to Hqq' \to \tau_\ell\tau_h qq'$, and $H \to \tau_\ell\tau_h$ (τ_ℓ is the leptonically decaying τ lepton).
- Z/γ^*+ jets (**Drell-Yan**): this type of event mainly contributes to the $\tau_h\tau_\ell$ channel decay when the charge of either ℓ or τ_h is misconstrued. Its contribution can be reduced both by requiring that ℓ and τ have the same charges and by requiring significant p_{T}^{miss}.

- **Triboson** production: $WWW, WWZ, WZZ, ZZZ,$ and $WW\gamma$ production processes contribute when at least one of the three vector bosons decays into a $\tau_h \nu_\tau$ pair, another into a light $\ell\nu_\ell$ pair, and the third into a qq' quark pair that exactly mimics the final state under study.
- **ZZ** diboson production: all processes involving ZZ production with at least one charged lepton in the final state can give a contribution to the final state if one Z decays into a $\tau_h\tau_\ell$ pair and the other Z decays into a qq' pair.
- **WZ** diboson production: The WZ pair production can enter the studied final states if W decays into hadrons and Z decays into a $\tau_h\tau_\ell$ pair.
- Associated production of a t quark, a vector boson V and another quark X (briefly, **tVX**): this broad class of processes gives a small contribution that can be mitigated by applying a veto on the presence of $b-$jets.
- **V**γ associated production: this process can mimic the final state of the signal in a way similar to ZZ/WZ diboson productions.

Together with the requirements on m_{jj} and $\Delta\eta_{jj}$ that are applied to discriminate the background sources listed above from the VBS signal, a requirement on a significant amount of p_{T}^{miss} in an event is very effective to further deplete a kinematic region of background events and make it enriched in the signal ones.

All background sources are estimated with dedicated Monte-Carlo simulations, as illustrated in Sec. 7.3, except for the non-prompt leptons contribution. The latter is estimated by a dedicated data-driven method, which will be presented in Sec. 7.6.

7.2 CMS data analysis

As for the study presented in Ch. 6, the data analyzed are the p-p collisions at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 138 fb^{-1}, recorded from 2016 to 2018 by the CMS experiment at the LHC.

Dataset	Run	Integrated luminosity (fb^{-1})
≥ 1 muon, ≥ 1 electron, or significant H_T	2016 (APV)	19.4
≥ 1 muon, ≥ 1 electron, or significant H_T	2016	16.5
≥ 1 muon, ≥ 1 electron, or significant H_T	2017	41.5
≥ 1 muon, ≥ 1 electron, or significant H_T	2018	59.9

Table 7.1: List of primary datasets of p-p collision data produced by $\sqrt{s} = 13$ TeV and collected by CMS from 2016 to 2018 used in the analysis. The data collected correspond to an integrated luminosity of 138 fb^{-1}.

The CMS production campaign of processed data and associated Monte-Carlo simulations used for this study is different from the one implemented in the search for a W' presented in Ch. 6 since it was released when that study was well advanced and presents an overall improved quality of processing of the information recorded by the detector. On this occasion, the 2016 dataset is split into two parts, 2016APV and 2016, in order to properly take into account the saturation effects in the preamplifier of the CMS Tracker Strip APV chip. The first subdataset represents 54% of the total 2016 dataset and the second consists of the remaining 46% of the 2016 events. For the measurement of the ssWW VBS process presented in the previous Sections and for the investigation of indirect New Physics effects, data are collected

with triggers requiring either a muon or an electron. The implementation of additional triggers requiring the presence of an isolated hadronic tau or a $\tau\ell$ pair has been investigated but has resulted in negligible contributions in terms of the number of events in the data sample. The data sample collected with triggers requiring hadronic activity in the final state is also used, but only to estimate the "fake ratios" for e, μ, τ_h, which is the main ingredient of the data-driven method used to properly estimate the non-prompt background, as will be explained in Sec. 7.5. Tab. 7.1 shows the list of primary data sets used in the analysis.

7.3 Monte-Carlo simulated samples

Except for the background events including non-prompt leptons and taus, all signal and background processes are simulated with appropriate Monte-Carlo event generators. For the signals, both SM and with EFT contributions, full details are given in Sec. 5.4.

The Monte-Carlo simulated samples needed to properly cover all background sources are given in Tab. 7.2, together with the event generators used and the corresponding cross-section accuracy. In general, MADGRAPH_AMC@NLO, POWHEG, PYTHIA8, and MCFM [210] are used for matrix element generation, i.e. for the hard processes. For all Monte-Carlo simulations, the default parameterization of the PDFs is NNPDF3.1 [211] at LO, NLO, or NNLO QCD, with the order matching that of the matrix element calculation. In addition, PYTHIA version 8.306 is used for the parton shower, quark hadronization, and the underlying event with the CP5 tune [212]. The FxFx and MLM merging schemes are used when necessary to avoid double counting of final state partons with those generated in the parton shower.

7.3.1 Electroweak and QCD higher order corrections for VBS simulations

Higher-order electroweak corrections are applied to the VBS ssWW signal processes, leading to a reduction in yields of about 10%. These corrections are taken from [213, 214]. These corrections have been measured for inclusive VBS ssWW processes as a function of the invariant mass of the generated VBS jets. The same type of correction is applied to the irreducible background represented by QCD-mediated ssWW VBS. Both are shown in Fig. 7.2.

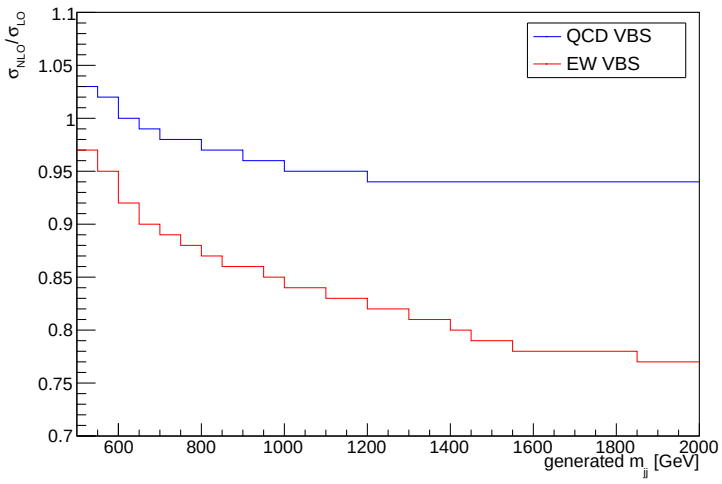


Figure 7.2: Ratio of leading-order to next-to-leading order corrections as a function of generated m_{jj} for EW and QCD ssWW VBS processes.

Group	Process	Matrix element	Matching method	Cross-section × BR (pb)
Signal	$pp \to WWqq' \to 2l2\nu_l2q$ (only EW)	MadGraph	–	0.02870 (LO)
	$pp \to WWqq' \to 2l2\nu_l2q$ (QCD)	MadGraph	–	0.02227 (LO)
	$pp \to WWqq' \to 2l2\nu_l2q$ with dim-6 EFT	MadGraph + SMEFTSim	–	depending on EFT operator (LO)
	$pp \to WWqq' \to 2l2\nu_l2q$ with dim-8 EFT	MadGraph + UFO	–	depending on EFT operator (LO)
	$pp \to WWqq' \to 2l2\nu_l2q$ with $c_W/\Lambda^{-2} = 1.0$	MadGraph + SMEFTSim	–	0.00100 (int.) + 0.01388 (pure EFT) (LO)
	$pp \to WWqq' \to 2l2\nu_l2q\, c_{HW}/\Lambda^{-2} = 1.0$	MadGraph + SMEFTSim	–	0.00051 (int.) + 0.00014 (pure EFT) (LO)
Dileptonic $t\bar{t}$	$pp \to t\bar{t} \to 2l2\nu_l2q$	POWHEG	–	72.1 (NNLO)
Opposite Sign	$pp \to WW \to 2l2\nu_l$	POWHEG	–	12.178 (NLO)
	$pp \to gg \to WW2l2\nu_l$	MCFM	–	2.129 (LO)
	$pp \to tW$	POWHEG	–	35.85 (LO)
	$pp \to \bar{t}W$	POWHEG	–	35.85 (LO)
	$pp \to ggH \to WW \to 2l2\nu_l$ (H $g-$fusion production)	POWHEG	–	1.0315 (LO)
	$pp \to ggH \to ZZ \to 4l$ (H $g-$fusion production)	POWHEG	–	0.0118 (LO)
	$pp \to ggH \to 2\tau$ (H $g-$fusion production)	POWHEG	–	0.237 (LO)
	$pp \to Hqq' \to WWqq' \to 2\ell2\nu_\ell$ + hadrons (H VBF production)	POWHEG	–	0.0896 (LO)
	$pp \to Hqq' \to 2\tau$ + hadrons (H VBF production)	POWHEG	–	2.7757 (LO)
	$pp \to ttH$	POWHEG	–	0.2120 (LO)
	$pp \to VH(V = Z,W)$	MadGraph_aMC@NLO + MadSpin	FxFx	0.952 (NLO)
Drell-Yan	$pp \to Z/\gamma^* \to 2lm_{ll} > 50$ GeV	MadGraph_aMC@NLO	FxFx	7181.0 (NLO)
WZ	$pp \to WZ$	PYTHIA	–	47.13 (LO)
ZZ	$pp \to ZZ \to 2l2\nu_l$	POWHEG	–	0.9738 (NLO)
	$pp \to ZZ \to 4l$	POWHEG	–	13.74 (NLO)
	$pp \to gg \to ZZ \to 4l$	MCFM	–	0.004780 (LO)
Triboson	$pp \to WWW$	MadGraph_aMC@NLO	–	0.2086 (NLO)
	$pp \to WWZ$	MadGraph_aMC@NLO	–	0.1651 (NLO)
	$pp \to WZZ$	MadGraph_aMC@NLO	–	0.05565 (NLO)
	$pp \to ZZZ$	MadGraph_aMC@NLO	–	0.01398 (NLO)
	$pp \to WW\gamma$	MadGraph_aMC@NLO	–	0.2147 (NLO)
tVX	$pp \to ttW \to$ hadrons	MadGraph_aMC@NLO + MadSpin	FxFx	0.4062 (NLO)
	$pp \to ttW \to$ leptons	MadGraph_aMC@NLO + MadSpin	FxFx	0.2160 (NLO)
	$pp \to ttZ \to$ hadrons	MadGraph_aMC@NLO	–	0.5297 (NLO)
	$pp \to ttZ \to$ leptons	MadGraph_aMC@NLO	–	0.2529 (NLO)
	$pp \to tt\gamma$	MadGraph_aMC@NLO + MadSpin	FxFx	3.6970 (NLO)
	$pp \to tqZ$	MadGraph_aMC@NLO	–	0.0758 (NLO)
Vγ	$pp \to Z\gamma \to \gamma 2l+$ hadrons	MadGraph_aMC@NLO	FxFx	51.1 (NLO)
	$pp \to W\gamma \to \gamma l\nu_l+$ hadrons	MadGraph_aMC@NLO	FxFx	191.3 (NLO)
W+jets	$pp \to Wq \to l\nu_l q 70 < h_T < 100$ GeV	MadGraph_aMC@NLO	MLM	1529.44 (NNLO)
	$pp \to Wq \to l\nu_l q 100 < h_T < 200$ GeV	MadGraph_aMC@NLO	MLM	1627.45 (NNLO)
	$pp \to Wq \to l\nu_l q 200 < h_T < 400$ GeV	MadGraph_aMC@NLO	MLM	435.237 (NNLO)
	$pp \to Wq \to l\nu_l q 400 < h_T < 600$ GeV	MadGraph_aMC@NLO	MLM	59.1811 (NNLO)
	$pp \to Wq \to l\nu_l q 600 < h_T < 800$ GeV	MadGraph_aMC@NLO	MLM	14.5805 (NNLO)
	$pp \to Wq \to l\nu_l q 800 < h_T < 1200$ GeV	MadGraph_aMC@NLO	MLM	6.65621 (NNLO)
	$pp \to Wq \to l\nu_l q 1200 < h_T < 2500$ GeV	MadGraph_aMC@NLO	MLM	1.60809 (NNLO)
	$pp \to Wq \to l\nu_l q h_T > 2500$ GeV	MadGraph_aMC@NLO	MLM	0.0389136 (NNLO)
Hadronic $t\bar{t}$	$pp \to t\bar{t} \to 4ql\nu_l$	POWHEG	–	365.34 (NNLO)
	$pp \to t\bar{t} \to 6q$	POWHEG	–	377.96 (NNLO)

Table 7.2: Simulated samples used for the three years in the analysis, where $l = e, \mu, \tau$ and $\ell = e, \mu$.

7.4 Physics objects and selections

The first set of selection requirements aims at identifying the final state objects compatible with the topology of the investigated VBS process with the desired final state and at defining a sample enriched in signal events. Indeed, an event is retained if in the final state, there is exactly one light lepton, i.e. muon or electron, one hadronic tau, at least two well-separated jets with high invariant mass, and an amount of missing transverse momentum attributed to the invisible neutrinos, following what is described in Sec. 1.6. The event is therefore selected when a primary vertex is reconstructed and the noise filters are passed, as described in Sec. 4.6. All physics objects and their selection requirements are summarized in Tab. 7.4.

For the physics object, a "tight" and "loose" definition is given for electrons, muons, and hadronic taus, as the former is used to preserve the events used in the statistical analysis, and the latter is used for the data-driven extraction of the non-prompt leptons and hadronic taus.

7.4.1 Triggers

The events used for the statistical analysis are required to pass through a combination of electrons and muons that exploit the topology and kinematics of the VBS process under study. According to this criterion, isolated muons reconstructed by the offline trigger algorithm are selected if their $p_{\mathrm{T}} > 24\,(27)$ GeV in the 2016 and 2018 (2017) datasets, while isolated electrons reconstructed similarly present $p_{\mathrm{T}} > 27\,(32)$ GeV in the 2016 (2017 and 2018) datasets.

The data-driven method implemented in this analysis for the estimation of non-prompt hadronic tau, electrons, and muons (see Sec. 7.6) requires a data sample enriched in events with a substantial amount of hadronic jets and a small amount of missing transverse momentum. Moreover, this sample should be orthogonal while staying as similar as possible in terms of event features. As a consequence, only events that, according to the trigger algorithms, present a hadronic jet with $p_{\mathrm{T}} > 40$ GeV or a hadronic activity $H_T > 125$ GeV are retained to populate the desired QCD-enriched sample. The full list of trigger paths used in this analysis is given in Tab. 7.3.

Sample	Purpose	Path	Year
Event reconstruction	Isolated electron	1 tight ID electron with $p_T > 27$ GeV	2016
		1 tight ID electron with $p_T > 32$ GeV	2017 and 2018
	Isolated muon	1 tight ID muon or 1 tracker muon with $p_T > 24$ GeV	2016
		1 tight ID muon with $p_T > 27$ GeV	2017
		1 tight ID muon with $p_T > 24$ GeV	2018
QCD-enriched	Hadronic activity	1 PF jet with $p_T > 40GeV$ or $H_T > 125$ GeV	2016
		1 PF jet with $p_T > 40GeV$ or $H_T > 180$ GeV	2017 and 2018

Table 7.3: Complete list of trigger paths required for the analysis, both for the event reconstruction and the implementation of the non-prompt data-driven estimation method described in Sec. 7.6.

7.4.2 Electrons and muons

The electrons and muons used in this analysis are either prompt or produced but τ leptonic decays, the latter having a p_{T} softer than the former.

Electrons within the pseudorapidity region $|\eta| < 2.5$ are considered, excluding the barrel endcap gap corresponding to the region $1.4442 < |\eta| < 1.566$. Electrons are divided into two categories, "tight" and "loose", using the identification criteria p_{T} and the relative isolation I^{ℓ}_{rel} within a cone of $\Delta R = 0.3$. An electron is labeled tight if it passes the MVAIsoWP90 identification

cut with $I^{\ell}_{rel} < 0.08$ and $p_{\mathrm{T}} > 30\,(35)$ GeV for 2016 (2017 and 2018) data; instead, it is labeled loose if it passes the MVA Iso WPL criteria, with $I^{\ell}_{rel} < 0.20$ and $p_{\mathrm{T}} > 15$ GeV.

Only muons with $|\eta| < 2.5$ are considered in this study. Similar to electrons, muons are divided into tight and loose categories. A muon is labeled "tight" if it satisfies the tight identification criteria defined in Sec. 4.2, with $I^{\ell}_{rel} < 0.15$ and $p_{\mathrm{T}} > 30$ GeV; it is labeled "loose" if it satisfies the loose identification criteria defined in Tab. 4.1, with $I^{\ell}_{rel} < 0.4$ and $p_{\mathrm{T}} > 15$ GeV.

For each event, either electrons or muons are selected, depending on the trigger paths fired in the event and the flavor of the leading lepton, i.e. the lepton with the highest p_T.

7.4.3 Hadronic tau leptons

In this analysis, τ leptons that decay into hadrons, τ_h, play a crucial role. Therefore, their optimal identification and reconstruction are important for the study.

Only τ_hs with $p_{\mathrm{T}} > 30$ GeV and $|\eta| < 2.3$ are considered in the analysis workflow.

The DeepTau discriminators illustrated in Sec. 4.5, namely DeepTauVsjet, DeepTauVsmu, and DeepTauVse, are implemented to correctly identify the τ_h against jets, muons, and electrons, respectively. A dedicated study of the three discriminators was performed to find the optimal combination of WPs to increase the signal sensitivity, which represents the "tight" selection for the hadronic taus:

- DeepTauVsjet VeryTight;
- DeepTauVsmu Tight;
- DeepTauVse VeryLoose.

The "loose" selection criteria are defined using the same DeepTauVsmu and DeepTauvse WPs implemented for the "tight" definition, while the DeepTauvsjet WP is selected as Medium for each data collection period and final state.

τ_h Decay Modes

As reported in Sec. 4.5, the reconstructed hadronic tau is classified according to the number of charged hadrons and neutral pions, called "tau decay modes" (tau DM for short):

- 0: 1 charged hadron, 0 neutral pions;
- 1: 1 charged hadron, 1 neutral pion;
- 2: 1 charged hadron, 2 neutral pions;
- 5: 2 charged hadrons, 0 neutral pions;
- 6: 2 charged hadrons, 1 neutral pion;
- 10: 3 charged hadrons, 0 neutral pions;
- 11: 3 charged hadrons, 0 neutral pions.

For the DM set used with the version of the DeepTau algorithms implemented in this analysis, it is recommended to reject values 5 and 6, as they are expected to induce an anomalous background increase with respect to the data.

In this analysis, the anomalous background mentioned above is observed, as expected for these DMs, in all control regions defined in Sec. 7.5 (see Fig. 7.3). Since excluding the "dangerous" τ_h DMs does not affect the overall sensitivity, they are rejected.

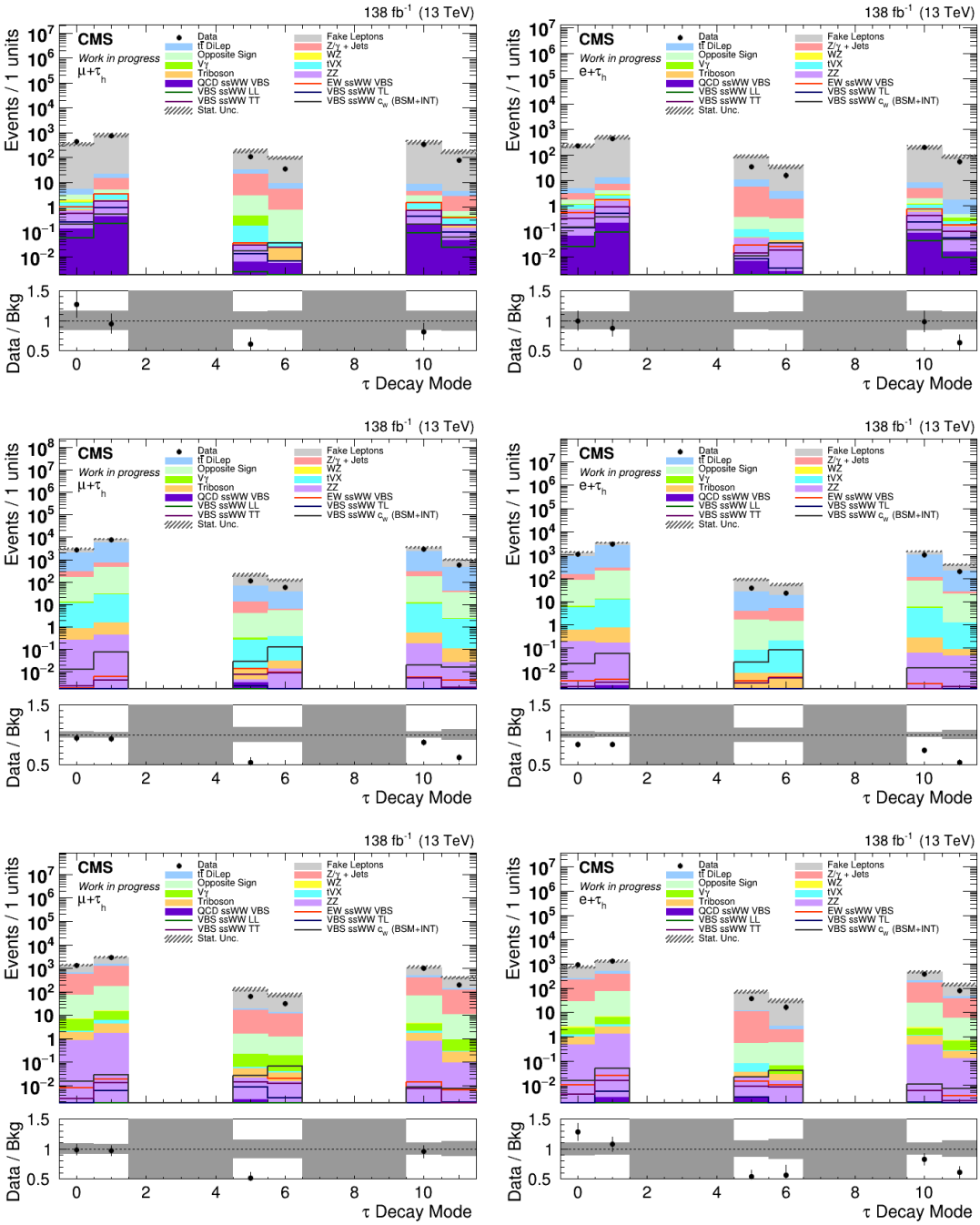


Figure 7.3: Prefit τ DMs distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) for the Run II samples in the $e\tau_h$ (right) and $\mu\tau_h$ (left) channels.

7.4.4 Jets

The reconstructed AK4 jets, as defined in Sec. 4.4, are used to identify the forward jet pair produced by the partons emitting the W boson pair, and are hereafter referred to as *VBS jets*. The contamination from the pile-up events in the AK4 jets is subtracted using the CHS algorithm.

Only jets with $p_T > 30$ GeV and $|\eta| < 5.0$ are considered consistent with the event topology of a VBS process and are required to pass the *tight jetID* identification criteria. For jets with $p_T < 50$ GeV, the Pileup JetID "tight" WP is required in addition to further rejecting jets from the pile-up.

b-tagged jets

One of the main sources of background for the current search is $t\bar{t}$ production. To reduce this contribution, events with at least one jet identified as a $b-$jet are vetoed using a $b-$*veto*. For $b-$tagging, both the DeepCSV and DeepFlavor algorithms illustrated in Sec. 4.4 were considered, estimating their $b-$veto performance for different working points (WPs). For both taggers, the WPs are defined as

- Loose, with misidentification probability 0.1;
- Medium, with misidentification probability 0.01;
- Tight, with misidentification probability 0.001.

The $b-$veto performances are estimated using the figure of merit $S/\sqrt{S+B}$, where S (B) is the number of events in the signal SM (including $t\bar{t}$ production) process that passes the $b-$veto, to find the $b-$veto algorithm and WP with the highest rejection power. According to the results shown in Fig. 7.4, DeepFlavour is chosen at the "Medium" WP to define the $b-$veto. A $b-$veto based on "loose" DeepFlavour WP, called *loose* $b-$*veto* for short, is considered to further deplete the $t\bar{t}$ production contribution if necessary.

The $b-$veto is applied by counting $b-$tagged jets (at the selected DeepFlavour WP) among the jets classified as "tight" in an event.

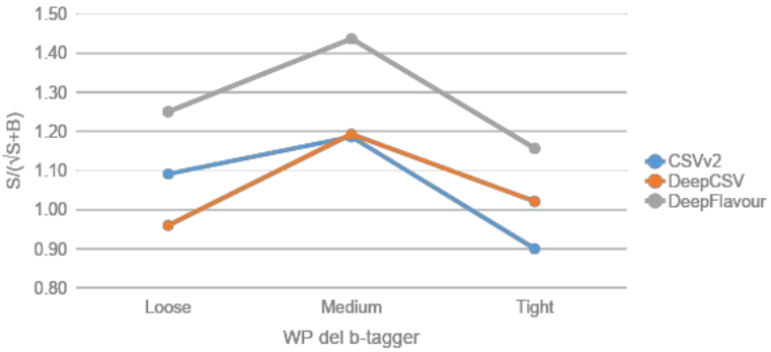


Figure 7.4: Signal significance after applying the $b-$veto for different $b-$taggers and different working points.

7.4.5 Missing Transverse Energy (MET)

The missing transverse energy $\vec{p}_T^{\,\text{miss}}$ is used to define regions of different compositions in signal and background, and to calculate kinematic quantities useful for separating the signals from the background.

Object	Requirements
Tight muon	$p_{\mathrm{T}} > 30$ GeV, $\lvert\eta\rvert < 2.5$, $I^{\mu}_{rel} < 0.15$
Loose muon	$p_{\mathrm{T}} > 15$ GeV, $\lvert\eta\rvert < 2.5$, $I^{\mu}_{rel} < 0.4$
MVA WP80 NoIso electron	$p_{\mathrm{T}} > 30(35)$ GeV, $\lvert\eta\rvert < 2.5$, $I^{e}_{rel} < 0.08$
MVA WPLoose NoIso electron	$p_{\mathrm{T}} > 15$ GeV, $\lvert\eta\rvert < 2.5$, $I^{e}_{rel} < 0.20$
Tight hadronic tau	$p_{\mathrm{T}} > 30$ GeV, $\lvert\eta\rvert < 2.3$, DeepTauVsjet VT, DeepTauVsmu T, DeepTauVse VL
Loose hadronic tau	$p_{\mathrm{T}} > 30$ GeV, $\lvert\eta\rvert < 2.3$, DeepTauVsjet M, DeepTauVsmu T, DeepTauVse VL
Jet	$p_{\mathrm{T}} > 30$ GeV, $\lvert\eta\rvert < 5.0$, TightID and PileupJet ID tight WP for $p_T < 50$ GeV
$b-$jet	$p_{\mathrm{T}} > 30$ GeV, $\lvert\eta\rvert < 5.0$, DeepJet M

Table 7.4: Summary of all the physics objects defined, their corresponding selection requirements, and the number of candidates required for the muon and electron channels.

7.4.6 Data-MC corrections

The simulation is corrected to account for several experimental effects that change from year to year due to the changing experimental conditions. Many of the corrections are provided directly by the CMS Collaboration, as reported in Sec. 5.5. Corrections derived specifically for this analysis are described below.

Trigger efficiency

Differences between the simulations and the data in the trigger efficiency are taken into account for the event reconstruction sample with dedicated trigger scale factors (SFs). For the triggers that select events with at least one isolated muon, the SFs are provided directly by the CMS Collaboration and implemented consistently. For those that select events with at least one isolated electron, SFs have been derived specifically for this study using the tag-and-probe procedure recommended by the CMS Collaboration.

The tag-and-probe procedure aims at measuring the efficiency of any object reconstruction or identification from data by exploiting resonances decaying to two of such objects, such as the Z boson or the J/Ψ meson, according to the following scheme:

- resonances are reconstructed as pairs with one leg passing a tight identification (tag) and one passing a loose identification (probe);
- "passing probes" are defined according to what is the efficiency to be measured;
- the "tag + passing probe" and "tag + failing probe" distributions are fit separately with a signal + background model;
- the efficiency is calculated from the ratio of the signal yields in the two distributions above;
- the procedure is repeated in bins of probe variables, usually p_{T} and η, to compute efficiency histograms as a function of these variables.

Specifically, this procedure is implemented by defining the tag as the electron that passes all of the corresponding tight requirements of this analysis, including the trigger requirements, and the probe as the electron that passes the looser requirements and does not necessarily pass the trigger selection. After measuring the efficiencies for the trigger chains both in the data and in the Monte-Carlo simulations as a function of the probe electron p_{T} and η, the SFs are obtained as the ratio between the data and simulation efficiencies. This procedure is repeated for each year of the Run II data period, and the SFs subsequently used for events in single-electron Monte-Carlo simulations are shown in Fig. 7.5.

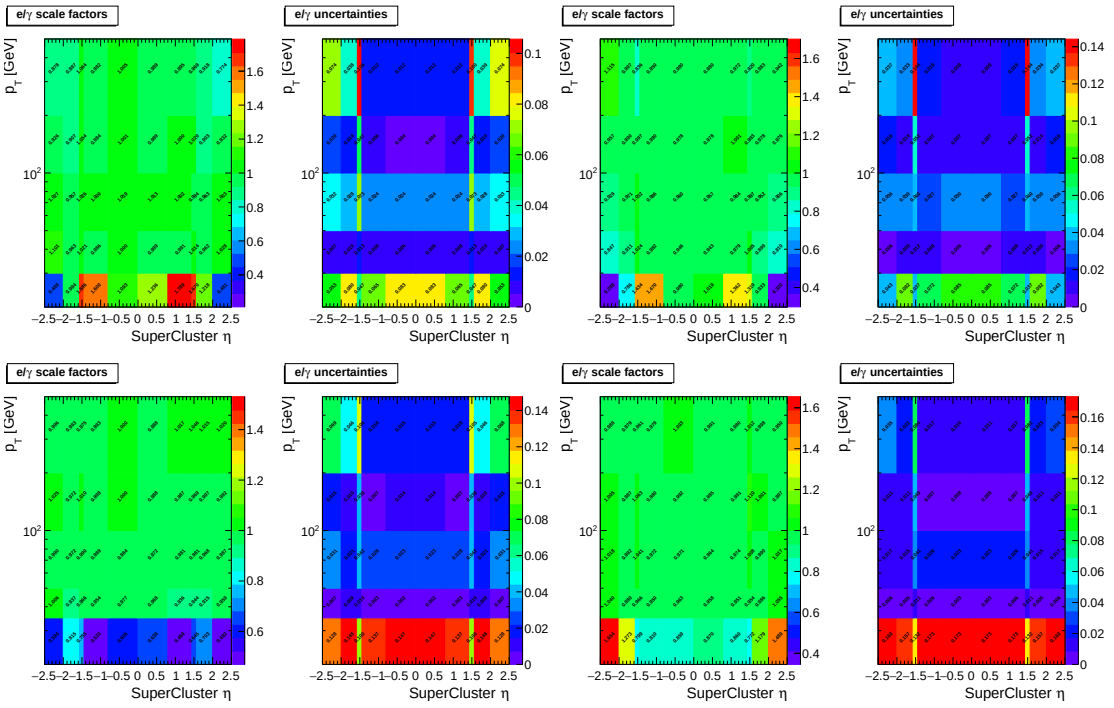


Figure 7.5: HLT Scale Factors and corresponding statistic+systematic uncertainties for 2016APV (top left), 2016 (top right), 2017 (bottom left), and 2018 (bottom right) datasets.

7.5 Event reconstruction and categorization

A first selection is imposed on the physical objects that are reconstructed in the event. Two different sets of requirements are applied for two different purposes, according to the definitions introduced above:

1. **VBS event reconstruction**
   - A "tight" isolated charged lepton ℓ (e or μ);
   - no additional loose charged leptons;
   - one " tight " isolated τ_h;
   - no additional "loose" τ_h;
   - at least two "tight" AK4 jets with a pseudorapidity separation of $|\Delta\eta_{jj}| > 2.5$.
2. **QCD-enriched control region for data-driven background estimation**
   - one "loose" charged lepton ℓ or "loose" τ_h;
   - no additional "loose" ℓ or τ_h;
   - $\vec{p}_T^{\,\text{miss}} < 50$ GeV;
   - $m_T(\ell, \vec{p}_T^{\,\text{miss}}) < 50$ GeV.

In the last bullet $m_T(\ell, \vec{p}_T^{\,\text{miss}})$ is the *transverse mass* of the system ℓ and $\vec{p}_T^{\,\text{miss}}$, defined as:

$$m_T(\ell, \vec{p}_T^{\,\text{miss}}) = \sqrt{2 p_T^{\ell} p_T^{\text{miss}} \left[1 - \cos\Delta\phi\right]}. \tag{7.1}$$

A dedicated study was performed to verify which is the optimal event selection by using simulated signal events. The various possible selection rules were compared with the event re-

construction performed using the Monte-Carlo truth information, which allows the reconstruction of the generated hard process:

1. select first the ℓ, τ_h pair and then the VBS jet pair as the leading ones in the "tight" jet collection;
2. select first the ℓ, τ_h pair and then the VBS jet pair as the "tight" jet pair with the highest invariant mass;
3. select first the VBS jet pair as the leading ones in the "tight" jet collection and then the ℓ, τ_h pair;
4. select first the VBS jet pair as the "tight" jet pair with the highest invariant mass and then the ℓ, τ_h pair.

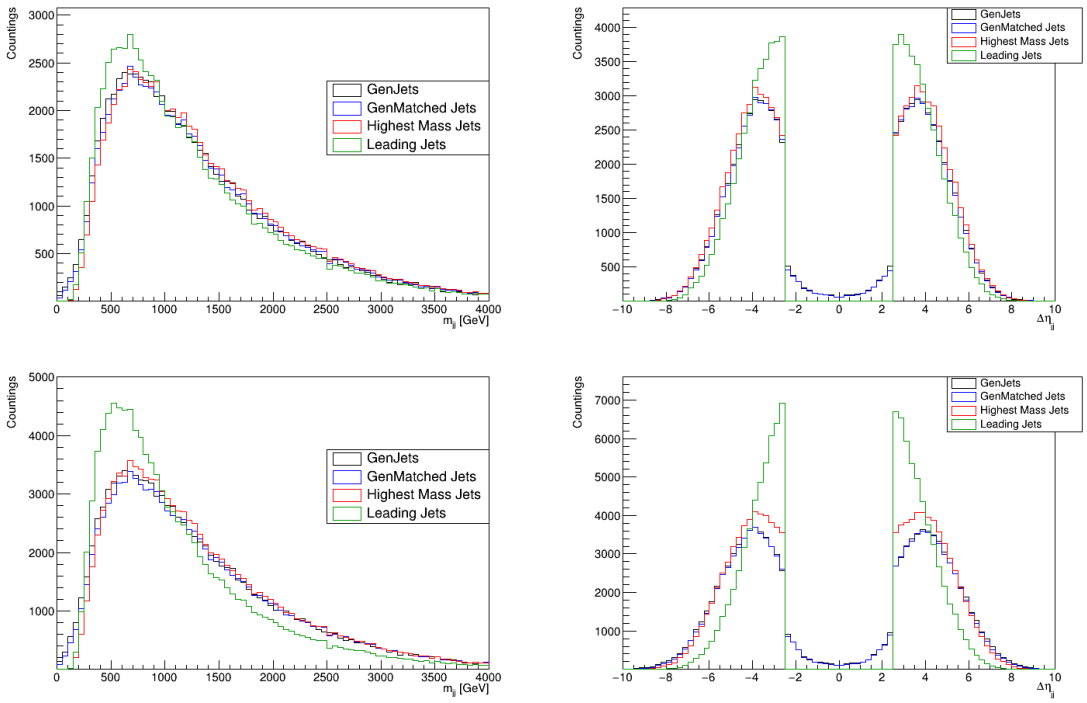


Figure 7.6: Top: dijet-invariant mass and pseudorapidity separation distributions for events after selecting first the VBS jet pair and then the $\ell\tau_h$ pair. Bottom: Similar distributions with the order of event selection reversed, i.e. ℓ and τ_h are selected first, and then the VBS jet pair. The color scheme depends on the algorithm used to identify the VBS jet pair; black is associated with generated jets (MC truth matching), the others with reconstructed ones.

In general, the following requirements apply to all of the strategies listed above:

- the VBS jet pairs are required to have a pseudorapidity separation of $|\Delta\eta_{jj}| > 2.5$;
- only light leptons and hadronic pairs that do not overlap within and have an angular opening $\Delta R < 0.4$ with any of the jets in the candidate VBS pair are considered;
- the leading ℓ, according to the flavor of the lepton that fired the trigger chains, and the leading τ_h in p_T are selected, requiring that they do not overlap within a cone of $\Delta R < 0.5$.

The distributions of the dijet invariant mass and pseudorapidity separation for the different strategies are shown in Fig. 7.6 and compared with those obtained with jets generated with

PYTHIA from the final-state VBS quark pair ("GenJets") and those obtained with the detector-level reconstructed jets matched to the VBS-generated jet pair ("GenMatched Jets"). This comparison shows that the optimal reconstruction strategy is strategy 4 since it reconstructs the distributions closest to the generated spectra. This event selection allows finding the best signal VBS candidates among all reconstructed physical objects to be associated with the particles in the considered final state.

Once the event is reconstructed, several quantities are introduced to exploit the kinematic properties of the studied signals. A variable measuring the centrality of the leptons with respect to the selected VBS jets, the "Zeppenfeld variable" [58], is constructed using the pseudorapidity η of one of the charged leptons, $l = \ell, \tau_h$, and the two VBS jets, j_1 and j_2:

$$z_l = \frac{\eta_l - \left| \frac{\eta_{j_1} + \eta_{j_2}}{2} \right|}{\eta_{j_1} - \eta_{j_2}}. \tag{7.2}$$

that is referred to as z_ℓ if the lepton is a muon or an electron, or z_τ if it is a hadronic tau. The event Zeppenfeld z_{ev} can be obtained by averaging z_ℓ and z_{τ_h}.

Additional kinematic variables can be defined to separate the SM contribution from the effects induced by the EFT operators studied in the VBS process. Two transverse masses are introduced for the $\ell, \tau_h, \vec{p}_T^{\,\text{miss}}$ system [215]:

$$M_{1\text{T}} = \left(\sqrt{M_{\tau l}^2 + p_{T\,\tau l}^2} + \vec{p}_T^{\,\text{miss}} \right)^2 - \left| \vec{p}_{T\,\tau l} + \vec{p}_T^{\,\text{miss}} \right|^2 \tag{7.3}$$

$$M_{\circ 1} = \left(p_{T\,\tau} + p_{T\,l} + \vec{p}_T^{\,\text{miss}} \right)^2 - \left| \vec{p}_{T\,\tau l} + \vec{p}_T^{\,\text{miss}} \right|^2 \tag{7.4}$$

The distributions of these quantities are compared in Fig. 7.7 for the SM contribution, separated by W boson polarization states, and the addition of the effects of the EFT dim-6 and dim-8 operators, as obtained by the corresponding simulations.

These quantities are, among others, used to develop the signal-vs-background discriminators based on the Deep Neural Network model that will be outlined in Sec. 7.7.

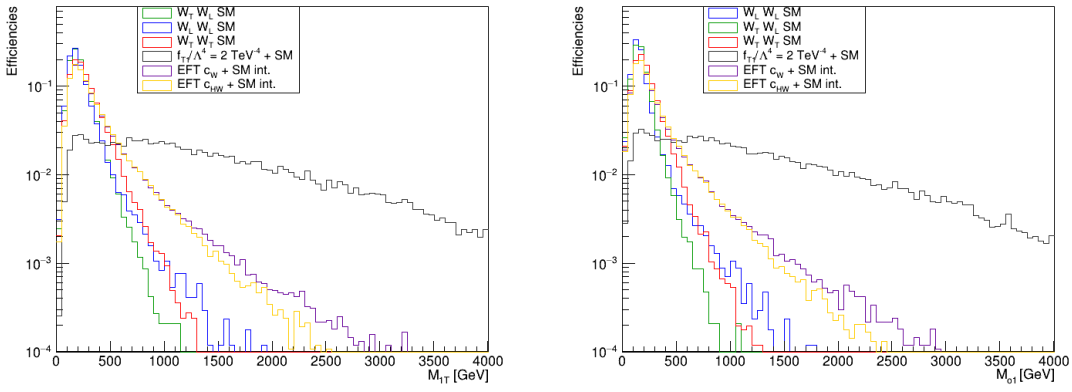


Figure 7.7: M_{1T} and M_{o1} distributions for different contributions to VBS processes, normalized to unit areas.

The reconstructed events are categorized into regions with different relative signal and background compositions to properly evaluate the estimation of the main background sources as well as the SM and EFT contributions to the VBS signal. They are defined as reported in the following list.

- **Signal Region:** reconstructed events with the same charge $\ell\tau_h$ pair, $p_{\mathrm{T}}^{miss} > 50$ GeV, $m_{jj} > 500$ GeV and passing the $b-$veto are collected in this signal-enriched category, dedicated to making the signal appear above the backgrounds.
- **Fakes Control Region:** the Fakes CR is intended to test and validate the data-driven non-prompt lepton background estimated with the data-driven "fakeable" method described in Sec. 7.6. All reconstructed events with a same-charge $\ell\tau_h$ pair and $p_{\mathrm{T}}^{miss} < 50$ GeV belong to this category.
- $t\bar{t}$ **Control Region:** the CR $t\bar{t}$ is introduced to evaluate the consistency of Monte-Carlo simulation predictions for dileptonic $t\bar{t}$ pair production. In this category, reconstructed events with an opposite-charge $\ell\tau_h$ pair, $p_{\mathrm{T}}^{miss} > 50$ GeV, and failing the $b-$veto request are collected.
- **Opposite Sign Control Region:** the Opposite Sign CR is intended to evaluate the reliability of Monte-Carlo simulations for the Drell-Yan Z/γ and all processes presenting a ℓ and a τ_h produced with opposite charges. It is populated by reconstructed events with opposite-charge $\ell\tau_h$ pairs and passing the loose $b-$veto, and then it is depleted of $t\bar{t}$ pair production events.
- **QCD-enriched region:** the QCD-enriched region is introduced to perform the first step of the data-driven non-prompt lepton background estimation, illustrated in Sec. 7.6.

The regions defined above are shown synthetically in Fig. 7.8.

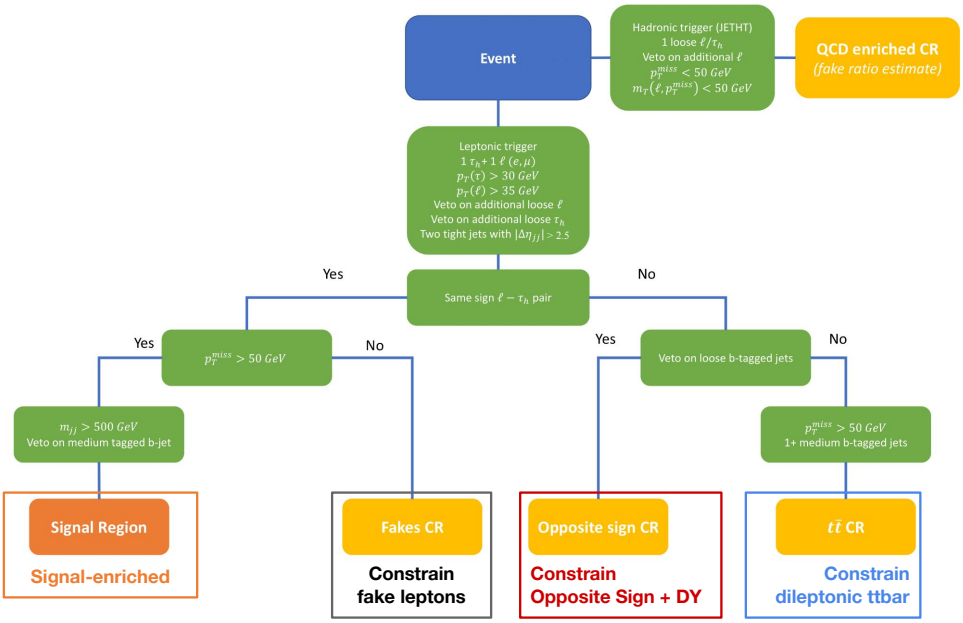


Figure 7.8: Schematic representation of the definitions of Signal and Control Regions (CRs), together with the CR used for the Fake Rate Estimate.

7.6 Data-driven non-prompt background estimation

The non-prompt leptons and taus introduced in Sec. 7.1 represent the dominant source of background for this analysis since they are events in which a hadronic jet is erroneously reconstructed as either e/μ or τ_h. They are mainly QCD-mediated multijet (briefly, QCD), W+jets, and hadronic and semileptonic $t\bar{t}$ production processes, listed in order of importance.

Since the available Monte-Carlo simulations, especially for the first two contributions, are not reliable enough to properly model the non-prompt leptons, data-driven methods are preferred to estimate this particular background source. For this analysis, the data-driven *fakeable method* described in detail in [216] is implemented. The basic idea of this method is to select a sample enriched in the non-prompt background and measure the probability for a non-prompt lepton or hadronic tau as a prompt one, called the "fake ratio", which is then used to estimate the contribution of this background in the main regions. The method is data-driven because the QCD-enriched sample is selected in the data and the fake ratio is also applied to the data.

First, a QCD-enriched sample with the set of hadronic trigger paths reported in Tab. 7.3 is extracted from the data sets enriched with hadronic activity, considering only the events entering the QCD-enriched regions to reduce the number of events with real leptons and hadronic taus. The remaining events with prompt leptons due to W+jets, DY+jets, hadronic and semileptonic $t\bar{t}$, and ZZ processes entering the QCD-enriched sample are subtracted using the corresponding simulations reported in Tab. 7.2.

Since there is no way to understand whether a reconstructed lepton is either an actual prompt lepton or a misreconstructed hadronic jet when using only detector-level information, an association between the sets of identification/isolation criteria for reconstructed leptons defined in Sec. 7.4 and their actual origin is required. The set of leptons satisfying the loose selection rules is called the "fakeable objects" or "denominators" and is defined separately for e, μ, τ_h. These objects are considered to be the non-prompt leptons, while the charged leptons that pass the corresponding "tight" selection are considered to be prompt and are also called "numerators".

The efficiency for these "loose" objects to pass the tight selection criteria is measured in data, and it is considered as the probability of a non-prompt lepton passing the tight selection and then entering the main regions. It is called the $e/\mu/\tau_h$ fake ratio $\varepsilon_{\text{fake}}$ and is parametrized as a function of the lepton p_T and η, taking into account any dependence on the kinematics or topology of an event. A "loose" object that does not pass the tight criteria is called "loose-not-tight" in the following.

$$\varepsilon_{\text{fake}} = \frac{N_{\text{tight}}}{N_{\text{loose}}}$$

The fake ratios extracted for each era of the Run II dataset analyzed in this study are shown from Fig. 7.9 to Fig. 7.14. In general, it can be seen that the fake ratios increase with the transverse momentum p_T of the considered lepton since the reconstruction and identification of inefficiencies increase. For muons, they increase with the absolute value of the pseudorapidity, while for electrons and hadrons, the opposite behavior is observed. For the former, the observed trend is justified by the fact that the identification efficiency is 99% in the whole pseudorapidity range (see Sec. 4.2), but the misreconstruction of hadronic jets as muons is more important at high values of $|\eta|$; for the latter, on the contrary, it is because the identification efficiency increases in the endcap regions (see Sec. 4.3 and Sec. 4.5), and thus, with respect to what happens in the barrel region, it is less likely that a jet misreconstructed as either a "loose" electron or a "loose" hadronic tau will also pass the corresponding "tight" tests.

These fake ratios are then used to compute the weights needed to extrapolate the non-prompt background yields from a region populated by reconstructed events with at least one "loose-not-tight" denominator lepton to the corresponding main region populated by events with both charged leptons labeled "tight". If there is only one "loose-not-tight" lepton, the event is labeled "1L", if both leptons are "loose-not-tight", the event is labeled "2L".

The extrapolation is performed using the "1L" and "2L" event samples. For each event in these samples, a weight is assigned to each "loose-not-tight" object, starting from the corresponding

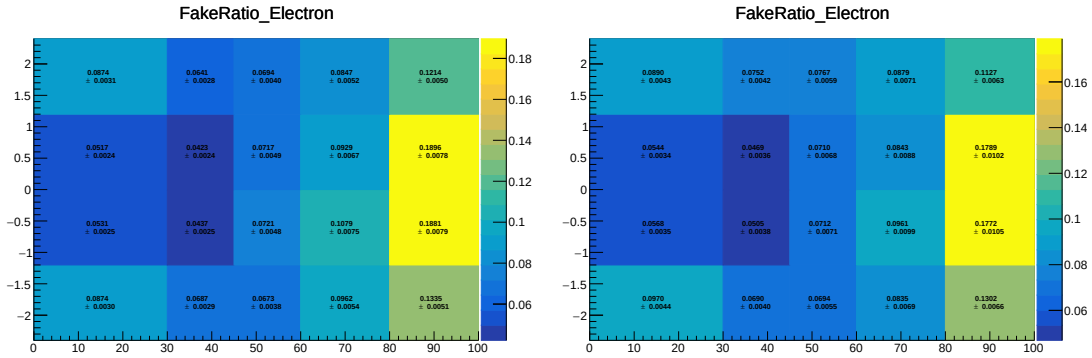


Figure 7.9: Electron fake ratio as function of p_T, η for 2016 (APV) (left) and 2016 (right) datasets.

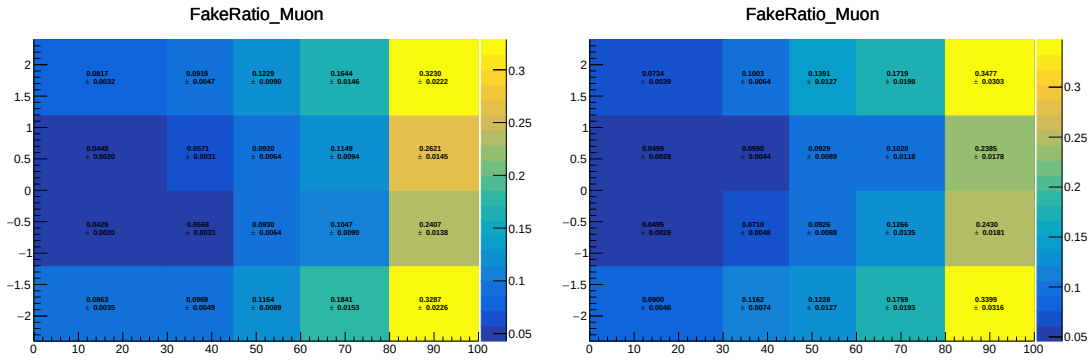


Figure 7.10: Muon fake ratio as function of p_T, η for 2016 (APV) (left) and 2016 (right) datasets.

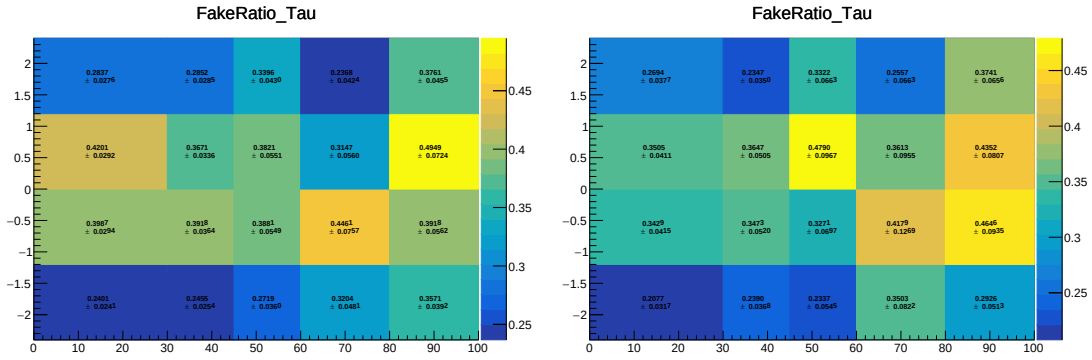


Figure 7.11: τ_h fake ratio as function of p_T, η for 2016 (APV) (left) and 2016 (right) datasets.

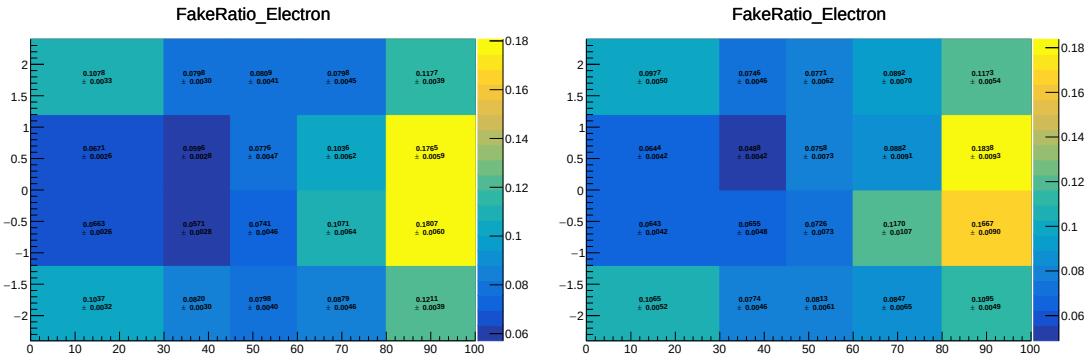


Figure 7.12: Electron fake ratio as a function of p_T, η for 2017 (left) and 2018 (right) datasets.

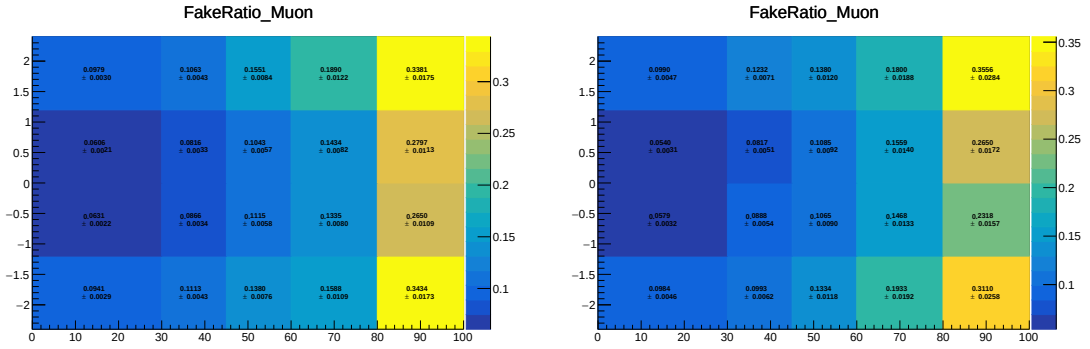


Figure 7.13: Muon fake ratio as a function of p_T, η for 2017 (left) and 2018 (right) datasets.

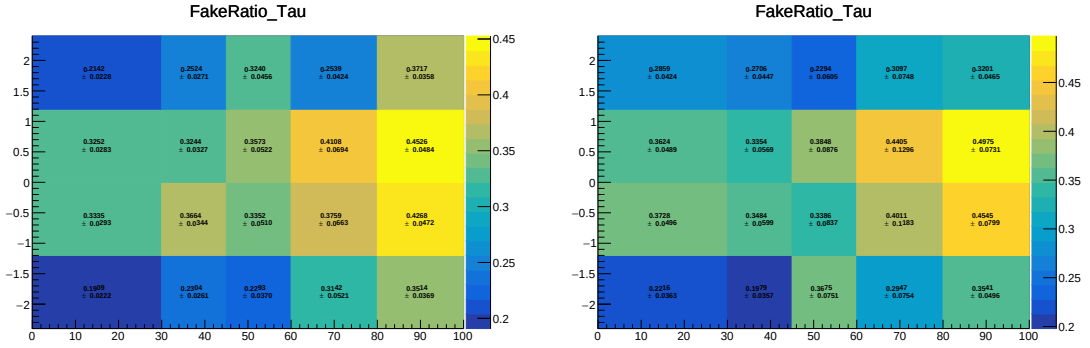


Figure 7.14: τ_h fake ratio as a function of p_T, η for 2017 (left) and 2018 (right) datasets.

value of the fake ratio $\varepsilon_{\text{fake}}(p_T, \eta)$:

$$w(p_T, \eta) = \frac{\varepsilon_{\text{fake}}(p_T, \eta)}{1 - \varepsilon_{\text{fake}}(p_T, \eta)}, \tag{7.5}$$

The weight of the "non-prompt lepton" event given by the product of the weights assigned to each loose lepton is finally taken as positive if the event falls in the 1L category, and negative if it falls in the 2L category. This procedure protects the estimate from double counting. The sum of the "non-prompt lepton" event weights represents the estimate of the non-prompt leptons in the "tight+tight" (i.e., main) region. This procedure is applied to estimate the non-prompt lepton and taus contribution in all regions defined in Sec. 7.5.

The data-driven estimation of the non-prompt background is considered to be affected by two sources of uncertainties. The first one is due to the limited size of the "1L" and "2L" samples, and it is implemented as the bin-by-bin statistical uncertainty of the related templates. The second one is represented by a conservative uncertainty of 30% per lepton flavor that takes into account all the possible uncertainties related to the determination of the non-prompt contribution with the fakeable method, and is considered to affect the overall yield of the corresponding templates.

Non-prompt background closure-test

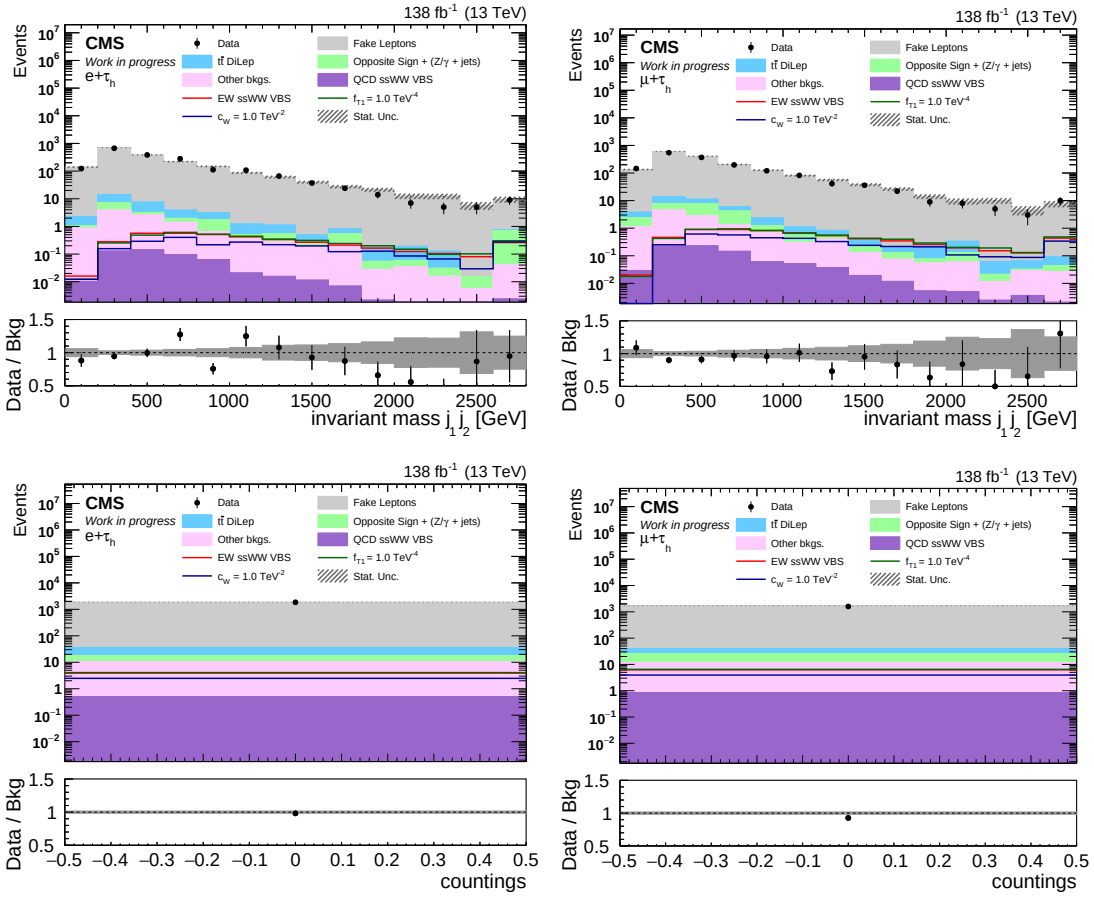


Figure 7.15: The distributions of VBS jet pair invariant mass m_{jj} (top) and countings (bottom) extracted with the full Run II dataset in the Fakes CR for the $e + \tau_h$ (left) and the $\mu + \tau_h$ (right) channels.

To validate the non-prompt lepton estimate obtained with the above method, this background

is estimated in the Fakes CR (see Fig. 7.8). Since in this region the data-background agreement is good within the statistical uncertainty and the conservative 30% uncertainty on total yield, the estimation procedure presented above is considered reliable for use in the other regions. Distributions for the VBS jet pair invariant mass m_{jj} and the countings extracted with the full Run II dataset in the Fakes CR are shown in Fig. 7.15.

7.7 Machine-Learning discriminators

Once the event categorization and the contributions to the signal and background processes are estimated, it is important to correctly discriminate signal events from background events to further improve the sensitivity of the study. For this purpose, several ML models are developed and finally implemented as signal-versus-background discriminators in the statistical analysis designed in Sec. 7.9.

7.7.1 Signal-vs-background ML discriminator

The separation between the different contributions to the signal and the main background sources can be improved by using ML-based multivariate discriminators, and their outputs can be implemented in the statistical analysis of the data. Three Deep Neural Networks (DNNs) are designed and optimized for this purpose, the first one being dedicated to discriminate the SM contribution to the VBS process under study, called SM DNN, while the others deal with the EFT contributions. Specifically, one DNN is outlined to discriminate the effects induced by the dim-6 SMEFT operators against the SM contribution and the backgrounds, and the other one is devoted to the same with respect to the effects arising from the dim-8 SMEFT operators. It is worth noting that this is the first study in CMS to implement DNN-based discriminators in EFT studies.

The model building and training procedure is performed using the Python Keras library [217] on a Tensorflow backend, while the Python Scikit-Learn library is used to preprocess the data fed into the models. Their hyperparameters are optimized using Bayesian optimization as implemented in the Python Scikit-Optimize library [218], and the feature importance is estimated thanks to the SHAP library [219]. Specifically, SHAP is the abbreviation for "SHapley Additive exPlanations", which is a game-theoretic approach to explain the output of any ML model, where the absolute value of the SHAP values computed on 100 randomly selected test events is used to rank the features.

The development procedure is the same for all DNNs introduced above. First, a subset of the kinematic variables is identified to reduce the number of features fed to the model while maintaining the necessary discriminative power. The training, validation, and test sets are created by extracting the 60%, 20%, and 20% of the Run II Monte-Carlo simulated events that populate the signal region, respectively. In particular, for the simulations of the processes that produce non-prompt leptons, i.e. W+jets, hadronic or semi-leptonic $t\bar{t}$, and Z/γ+jets, all events in the SR that satisfy the loose lepton selection rules are collected, while for those of the other background and signal sources, the tight lepton selection rules are applied. Then, for each background contribution, the events are weighted to normalize the total yield to the corresponding nominal cross-sections and luminosities of the Run II data set. In a further step, the sum of all background events is normalized to the same yield of the signal sample to achieve balanced training between signal and background, and the values of the input features are standardized to avoid any scale-related bias. Finally, the DNN models are trained and optimized.

The performance is evaluated by making use of the overtraining plots and the ROC curves. In the following, the three DNNs are described in their characteristics. The prefit distributions of

the kinematic quantities injected as input features for the training of the DNNs are shown in all CRs and SR for both $e+\tau_h$ and $\mu+\tau_h$ channels in Appendix B.1, as well the prefit distributions for the outputs of the three DNN-based discriminators.

SM DNN

For the DNN discriminating the SM contribution to the VBS process under study from the backgrounds, the Monte-Carlo samples used as the signal is the ssWW EW VBS with polarized W bosons to avoid the use of the same signal events eventually injected in the statistical analysis. The background processes are all estimated by Monte-Carlo simulations and are, in order of importance, $t\bar{t}$, $W+$ jets, Opposite Sign, Drell-Yan Z/γ^*+jets, and ZZ.

Nine input variables are used to capture the signal features:

- VBS jet pair invariant mass m_{jj};
- transverse mass of the system $(\ell, \vec{p}_T^{\text{miss}})$;
- transverse masses m_{1T}, m_{o1}
- p_T of both VBS jets;
- τ_h p_T;
- lepton p_T;
- p_{T} of the leading track of the jet associated with the selected τ_h over the τ_h p_{T}.

A summary of the SM DNN performance is shown in Fig. 7.16.

dim-6 DNN

A DNN to discriminate the effects induced in the ssWW VBS process by the SMEFT dim-6 operators is trained using as the signal the Monte-Carlo simulations of the insertion of the dim-6 $\mathcal{O}_{HW}$ and $\mathcal{O}_W$ operators. Both the SM interference and the pure "quadratic" EFT contributions are considered. These two contributions are balanced against each other, similar to what is done to balance the total background and signal yields. The pure SM contribution in the VBS signal and the background samples listed above for the SM DNN model are injected as background for the dim6 DNN training.

In addition, 23 different input variables are considered:

- VBS jet pair invariant mass m_{jj} and their separation in ϕ;
- transverse mass of the system $(\ell, \vec{p}_T^{\text{miss}})$;
- transverse masses m_{1T}, m_{o1};
- p_T and mass of both VBS jets;
- lepton and τ_h p_T and η;
- event Zeppenfeld value, divided by the $\Delta\eta$ separation among the VBS jets;
- p_{T} of the leading track of the jet associated with the selected τ_h over τ_h p_{T};
- component of lepton and τ_h p_T perpendicular to leading and subleading jets direction, p_{T}^{rel};
- distance of lepton and τ_h to leading and subleading jets direction in the ϕ plane.

Performance summary distributions are shown in Fig. 7.17.

dim-8 DNN

Similar to the training of the dim-6 DNN, the signal for the dim-8 DNN is represented by a set of 18 collections of simulated events for the VBS process with EFT-induced effects, each of which includes the effects of inserting a specific dim-8 SMEFT operator. The collections have the same number of simulated events, taking into account the SM interference and the pure EFT contributions. This is possible using the EFT reweighting weights for both positive and negative Wilson coefficient values associated with a particular operator, implemented in the simulations for the SMEFT dim-8 operators reported in Tab. 7.2. Since the EFT reweighting weight of a given operator depends on the value of the corresponding Wilson coefficient, the absolute values are chosen to be higher than the 95% CL limits obtained by fitting with M_{o1} (see Sec. 7.10).

The background samples are injected in the same way illustrated for the dim-6 DNN model. For the dim-8 DNN training, 11 quantities are used as input features:

- transverse mass of the system ($\tau_h, \vec{p}_T^{\,\text{miss}}$);
- transverse mass of the system ($\ell, \vec{p}_T^{\,\text{miss}}$);
- transverse mass of the system ($\ell, \tau_h, \vec{p}_T^{\,\text{miss}}$);
- transverse masses m_{1T}, m_{o1};
- p_T and mass of both VBS jets;
- lepton and τ_h p_T;

The distributions summarizing the performances of the dim-8 DNN are shown in Fig. 7.18.

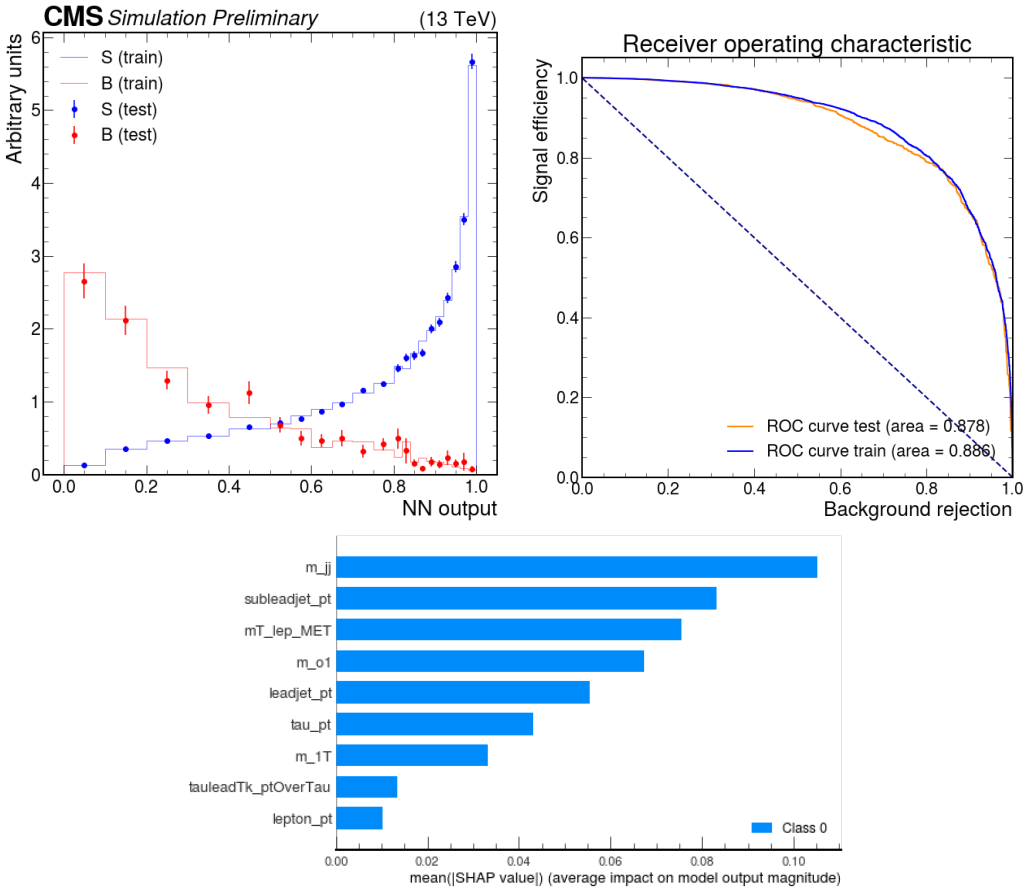


Figure 7.16: SM DNN discriminator output on training and test sets (top left, bin content is divided by bin width and then by histogram integral), ROC on test and training sets (top right), feature importance evaluated with the SHAP algorithm (bottom).

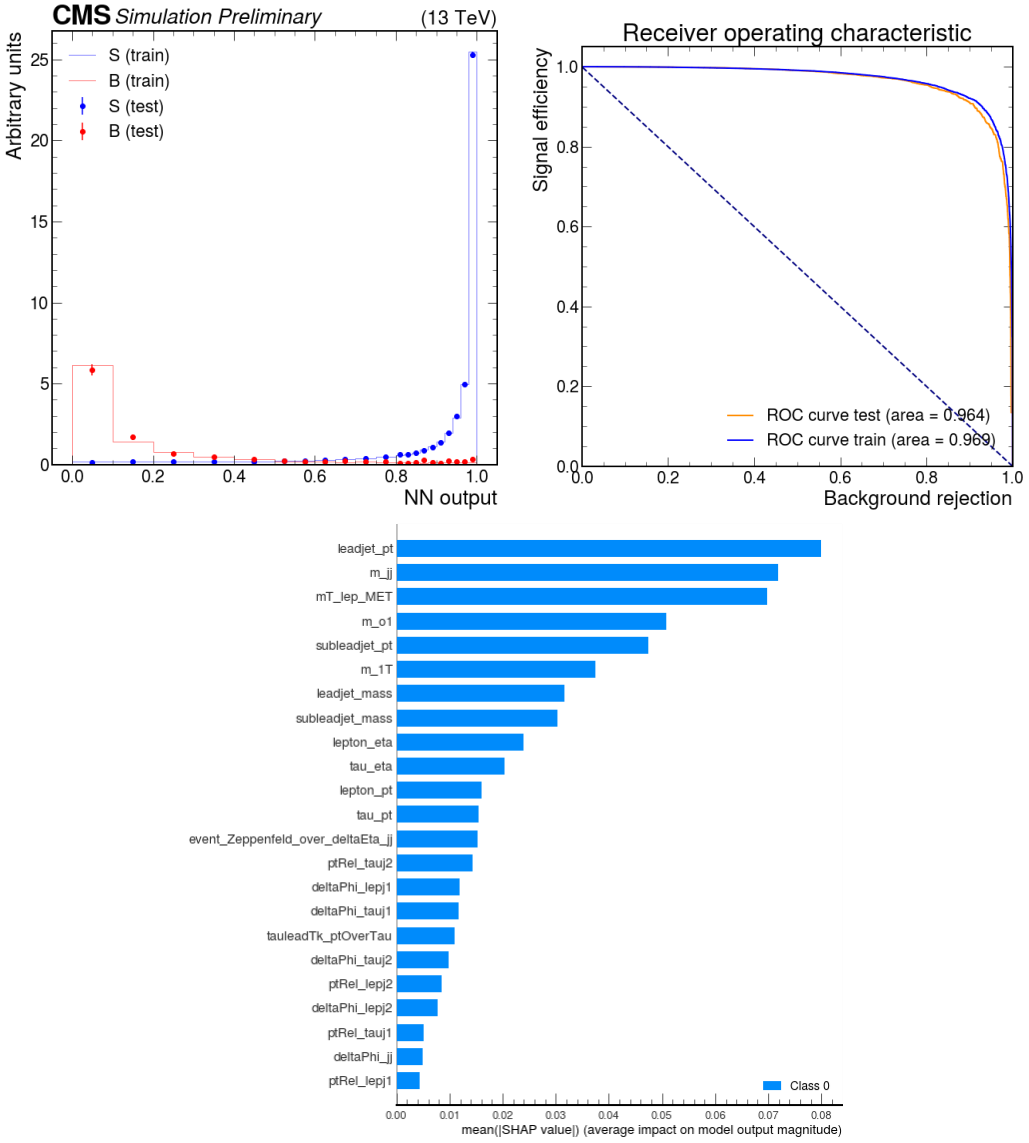


Figure 7.17: dim-6 DNN discriminator output on training and test sets (top left, bin content is divided by bin width and then by histogram integral), ROC on test and training sets (top right), feature importance evaluated with the SHAP algorithm (bottom)

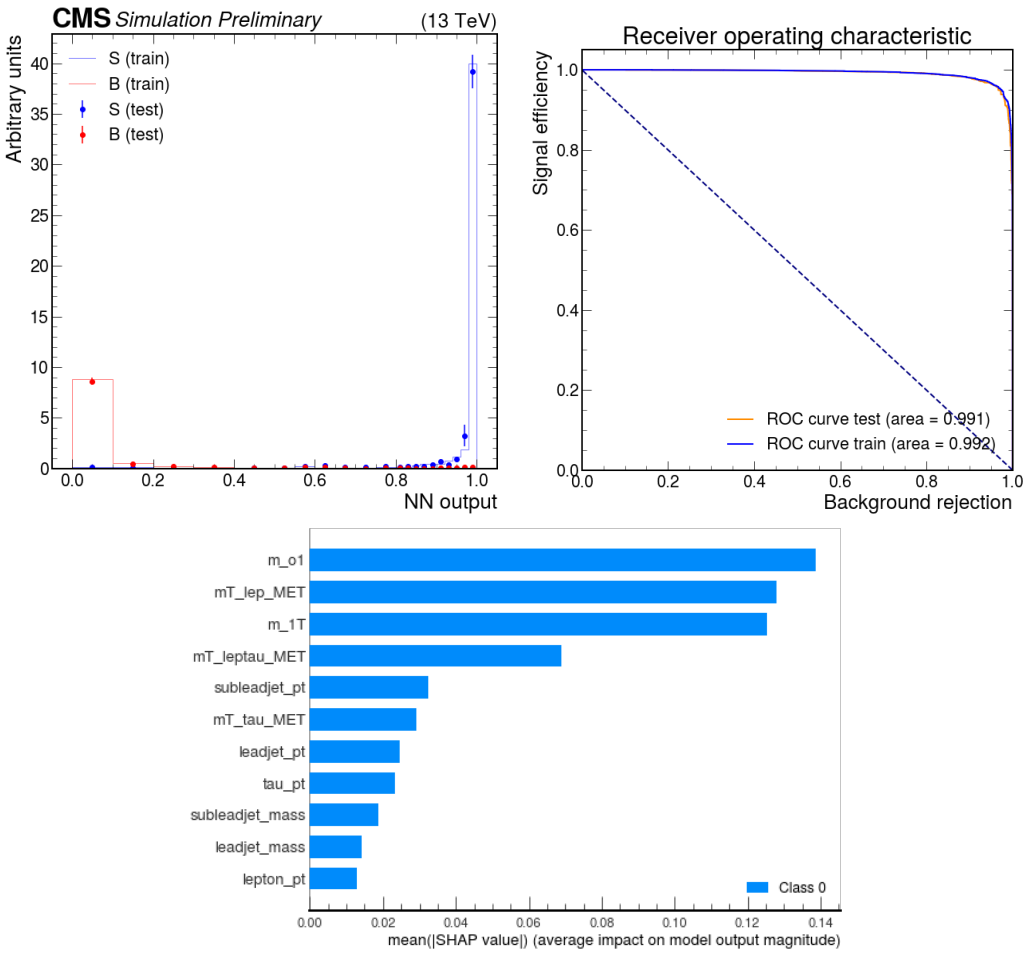


Figure 7.18: dim−8 DNN discriminator output on training and test sets (top left, bin content is divided by bin width and then by histogram integral), ROC on test and training sets (top right), feature importance evaluated with the SHAP algorithm (bottom)

7.8 Systematic uncertainties

All of the systematic uncertainties introduced in Sec. 5.6 are included in the statistical analyses described in Sec. 7.9 and Sec. 7.10. In addition, some uncertainties specific to this study are considered and illustrated below. Tab. 7.5 lists all included systematic uncertainties, together with the correlation between years and their effect. Since the main sources of systematic uncertainty affecting the different contributions to the ssWW VBS signal come from the theoretical modeling, in particular from the PDF and QCD scales, their upward and downward systematic effects are compared with the nominal distributions of the DNN discriminator output, presented in Sec. 7.7, from Fig. 7.19 to Fig. 7.21. Besides showing the magnitude of the variations induced by these two theoretical uncertainties on the signal distributions, it is clear from these plots that the QCD scale uncertainty becomes more relevant for the values of the output of the DNN discriminators where the signals are more relevant.

7.8.1 Uncertainties on data-driven non-prompt leptons estimate

The data-driven estimate of the non-prompt leptons comes with an irreducible uncertainty due to the limited statistics of the samples used to extrapolate it in the main regions, as well as with the uncertainties arising during the data-driven procedure of the "fakeable method". As explained in Sec. 7.6, the first is treated as a bin-by-bin statistical uncertainty, similar to what is done for the limited size of the Monte-Carlo samples, and the second is taken into account with a flavor-dependent conservative 30% uncertainty imposed on the total yields of the non-prompt background estimate for each year of the Run II data-taking period introduced in the statistical analysis. This approach turns out to take into account all possible sources of uncertainty affecting this type of background, as can be verified from the systematic impact studies reported in Fig. 7.23.

7.8.2 Normalization of Opposite Sign, Z/γ^*+jets, and dileptonic $t\bar{t}$ estimates

The contributions to the total background from Drell-Yan events and the minor processes collectively called "Opposite Sign" are injected into the statistical study fit as a single process to reduce the effect of the statistical fluctuations induced by the small unweighted number of events with the charge of ℓ and τ_h misreconstrued entering the signal region. The total normalization is left free and the corresponding uncertainty is included as an additional nuisance parameter, one per year of Run II. Similarly, the normalization of the distributions of the dileptonic $t\bar{t}$ contribution is left free, with a separate nuisance parameter introduced to account for its uncertainty. Both normalization uncertainties are treated as uncorrelated over years.

7.8.3 Misidentification of the charges of the lepton and the hadronic tau

According to the previous studies performed in the CMS Collaboration on this topic, the misidentification of the lepton charges may only be relevant for the reconstructed electrons. To account for this inefficiency in reconstructing the charge of the electrons, a flat conservative uncertainty of 15% is applied to the distributions of the background processes with an opposite-charge $\ell\tau_h$ pair in the final state in the Signal Region, according to the studies reported in [220], and uncorrelated among years.

7.8.4 ssWW VBS LO-to-NLO corrections uncertainties

The LO-to-NLO corrections for the SM contribution to the ssWW VBS signal introduced in Sec. 7.3 carry a statistical uncertainty. This is propagated into statistical analyses with a dedicated nuisance parameter, similar to what is done for the experimental scale factors, without correlation between years.

Systematic	Year Correlation	Type
Luminosity	no	normalization
Non-prompt leptons estimation unc.	no	normalization
$t\bar{t}$ normalization	no	normalization
Z/γ+jets and OS normalization	no	normalization
MC and non-prompt leptons stat. unc.	no	shape
Pre-firing	yes	shape
Pileup corrections	yes	shape
Pileup-JetID corrections	no	shape
$b-$tagging efficiency	yes	shape
mistagging efficiency	yes	shape
e charge misidentification	no	shape
Lepton reconstruction, isolation, and trigger corrections	no	shape
DeepTau algorithm corrections	no	shape
Tau Energy Scale	yes	shape
PDF uncertainty	yes	shape
QCD scale uncertainty	yes	shape
Parton showering - ISR	yes	shape
Parton showering - FSR	yes	shape
JES	no	shape
JER	no	shape
MET unclustered energy	yes	shape
LO-to-NLO VBS corrections	no	shape

Table 7.5: List of systematics included in the statistical analysis with their correlation among years and type, as defined in Sec. 5.6

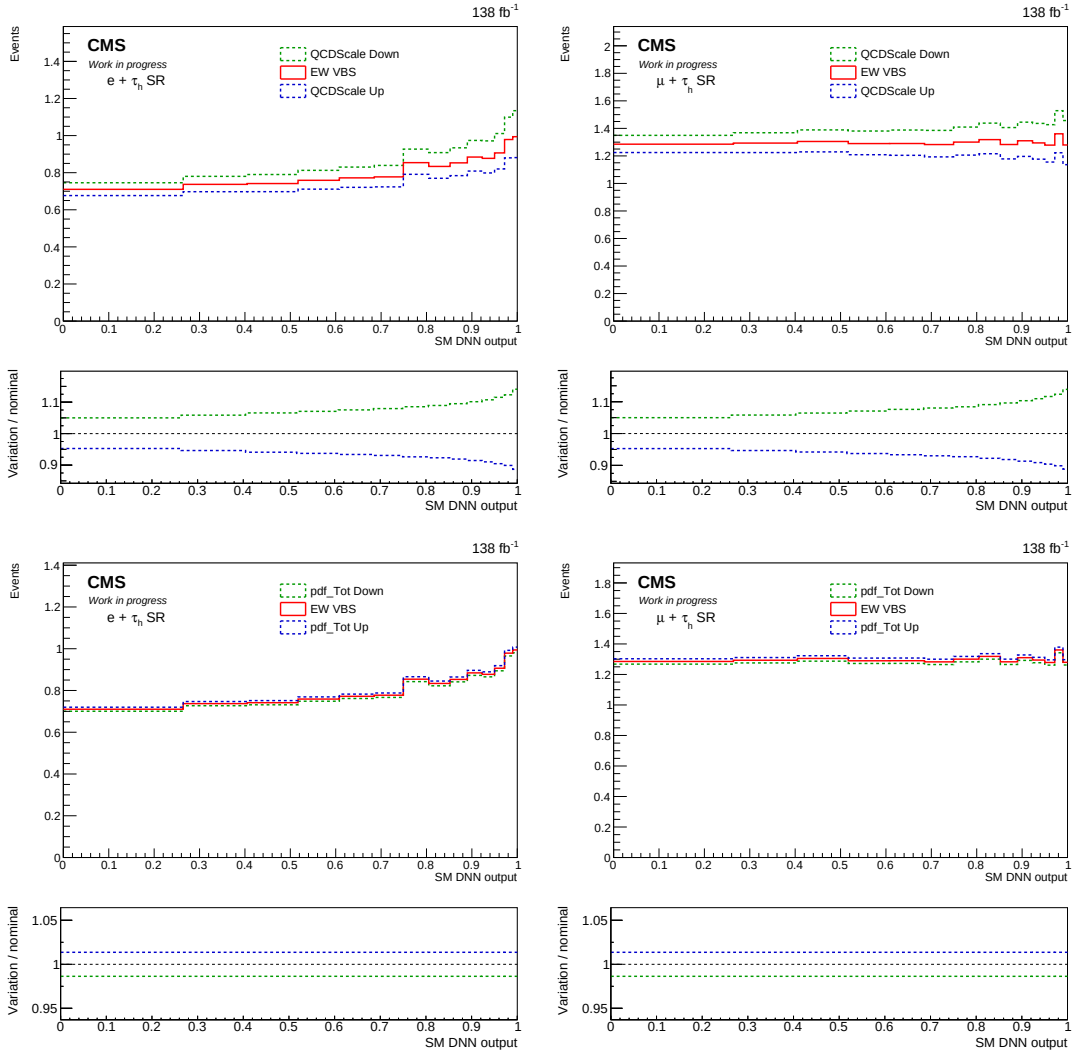


Figure 7.19: Upward and downward variations induced in the SM DNN output distributions for the EW VBS signal in the $e + \tau_h$ (left) and $\mu + \tau_h$ (right) channel by the QCD scale (top) and PDF (bottom) theoretical uncertainties.

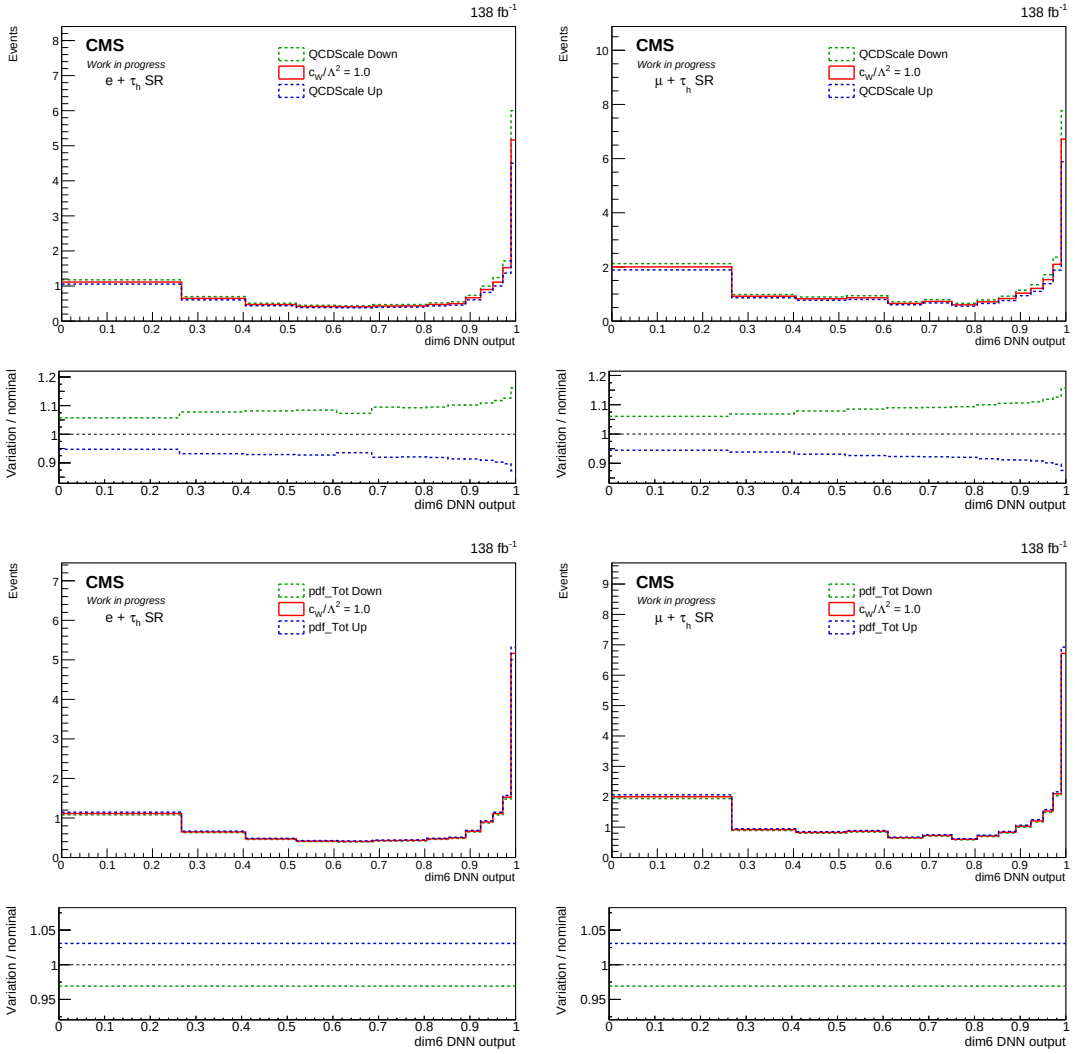


Figure 7.20: Upward and downward variations induced in the SM DNN output distributions for the dim−6 Q_W EFT operator (with $c_W/\Lambda^2 = 1.0$) in the $e + \tau_h$ (left) and $\mu + \tau_h$ (right) channel by the QCD scale (top) and PDF (bottom) theoretical uncertainties.

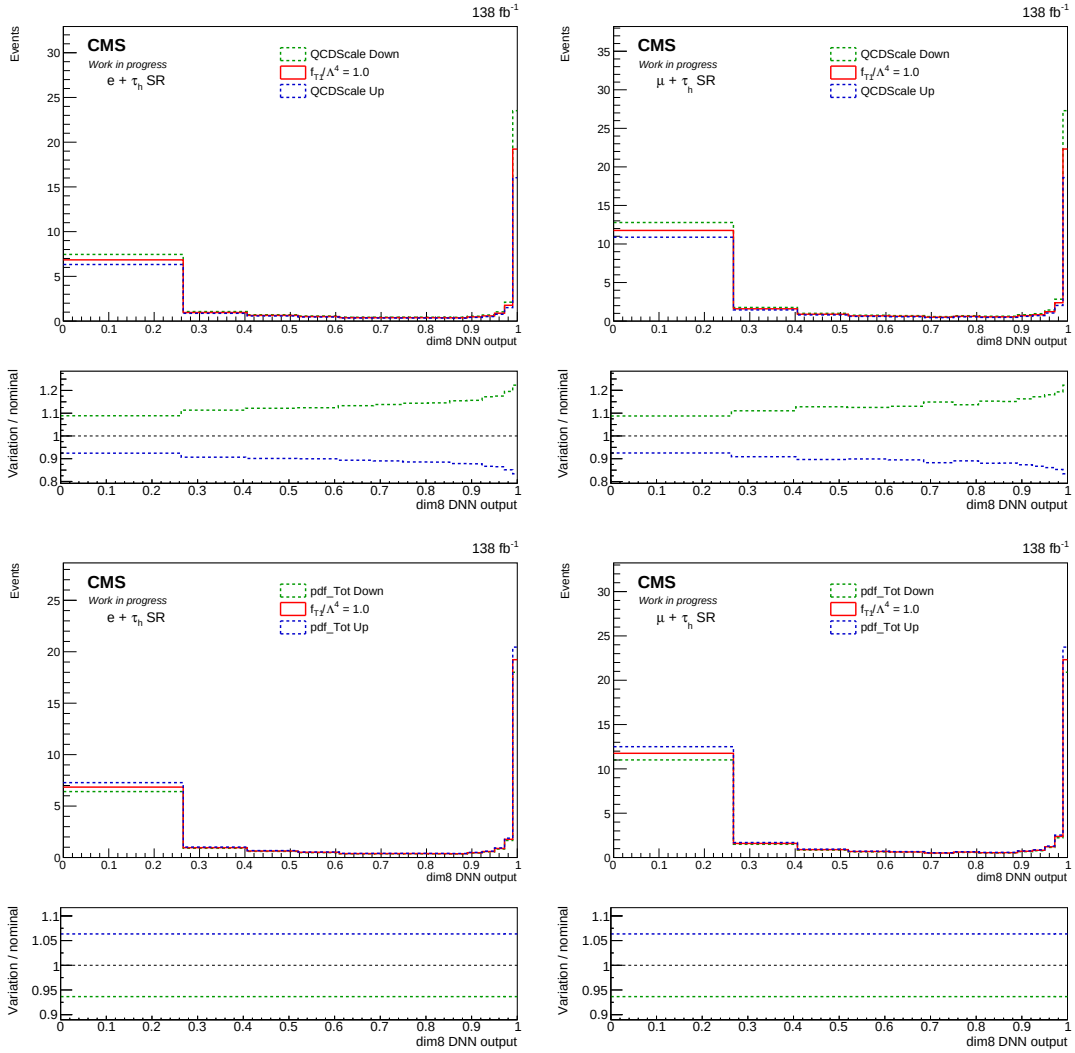


Figure 7.21: Upward and downward variations induced in the SM DNN output distributions for the dim-8 O_{T1} EFT operator (with $f_{T1}/\Lambda^4 = 1.0$) in the $e + \tau_h$ (left) and $\mu + \tau_h$ (right) channel by the QCD scale (top) and PDF (bottom) theoretical uncertainties.

7.9 Search for the production of ssWW VBS

In order to properly exploit the signal-versus-background discrimination power of the SM DNN model, an appropriate binning must be determined to generate the DNN output distribution templates to be injected into the fits. For this study, the signal flattening strategy introduced in [53] for analogous BDT models has been implemented to find a binning such that the DNN output distribution for the signal has the same weight for each bin and, consequently, a flat behavior in the SR. For each signal, three different signal-flattening binnings are tested: broad with 14 bins, medium with 18 bins, and fine with 22 bins. Considering that each bin must contain a significant amount of the total expected signal and background events, the broad signal-flattening binning was chosen to inject the SM DNN distributions into the statistical fit.

A simultaneous maximum likelihood fit is performed on the 138 fb^{-1} data set collected by CMS in the Run II data collection period in the SR, $t\bar{t}$ CR, and Opposite Sign CR, implementing the CombineLimit tool. This study is considered preliminary as the SR is blinded to the data according to the current status of the internal approval procedure of the analysis in the CMS Collaboration. The sensitivity to the signal is evaluated with the significance of the EW and EW + QCD ssWW VBS processes with a hadronic tau in the hypothesis of only SM background, exploiting the asymptotic limit of Wilks' theorem [221]. To validate the results obtained by relying on the asymptotic limit, the same estimate is computed without using the latter and generating frequentist toy samples for the signal and the background, taking into account their statistical fluctuations. From now on, this strategy will be called the "toy method" and does not suffer from the possible limited number of events in the bins, providing a cross-check of the asymptotic limit results. The relevance of the signal is measured by the signal strength, defined as the ratio among the observed cross-section and the one expected from the theory, denoted by μ_S or r.

In Fig. 7.22 the postfit distributions of the SM DNN output are shown in all regions injected into the fit and for both the $e+\tau_h$ and $\mu+\tau_h$ channels, showing only the most sensitive bins. The total statistical and systematic uncertainty as injected into the fit (prefit) is compared with the total constraint on the various sources of uncertainty, called the postfit uncertainty. The statistical fit can significantly reduce the magnitude of the total uncertainty. In addition, the effect of each uncertainty on the likelihood is disentangled to show the importance of each of them, as reported in Tab. 7.6. This procedure is commonly called uncertainty decomposition. Finally, the fluctuations induced by each uncertainty on the parameter of interest are estimated by varying the corresponding perturbation parameters (briefly referred to as "impacts"), and the constraints on them resulting from the fit in the case of the SM EW ssWW VBS measurement are determined and shown in Fig. 7.23.

All of these statistical tests are performed considering the preliminary fit to the data shown above, using both signal and background expectations. They show that the main source of uncertainty, apart from the statistical uncertainty, is due to the theory, in particular, the QCD scale and the PDF for the signal distributions, together with the bin-by-bin uncertainties affecting the background estimates. The main uncertainties are not significantly overconstrained, i.e. the prefitting estimate of the uncertainties does not differ significantly from the constraints on the corresponding perturbative parameters estimated by the fitting procedure. This indicates that the uncertainties are not overestimated, confirming the goodness of the procedure implemented to estimate them before the statistical fit.

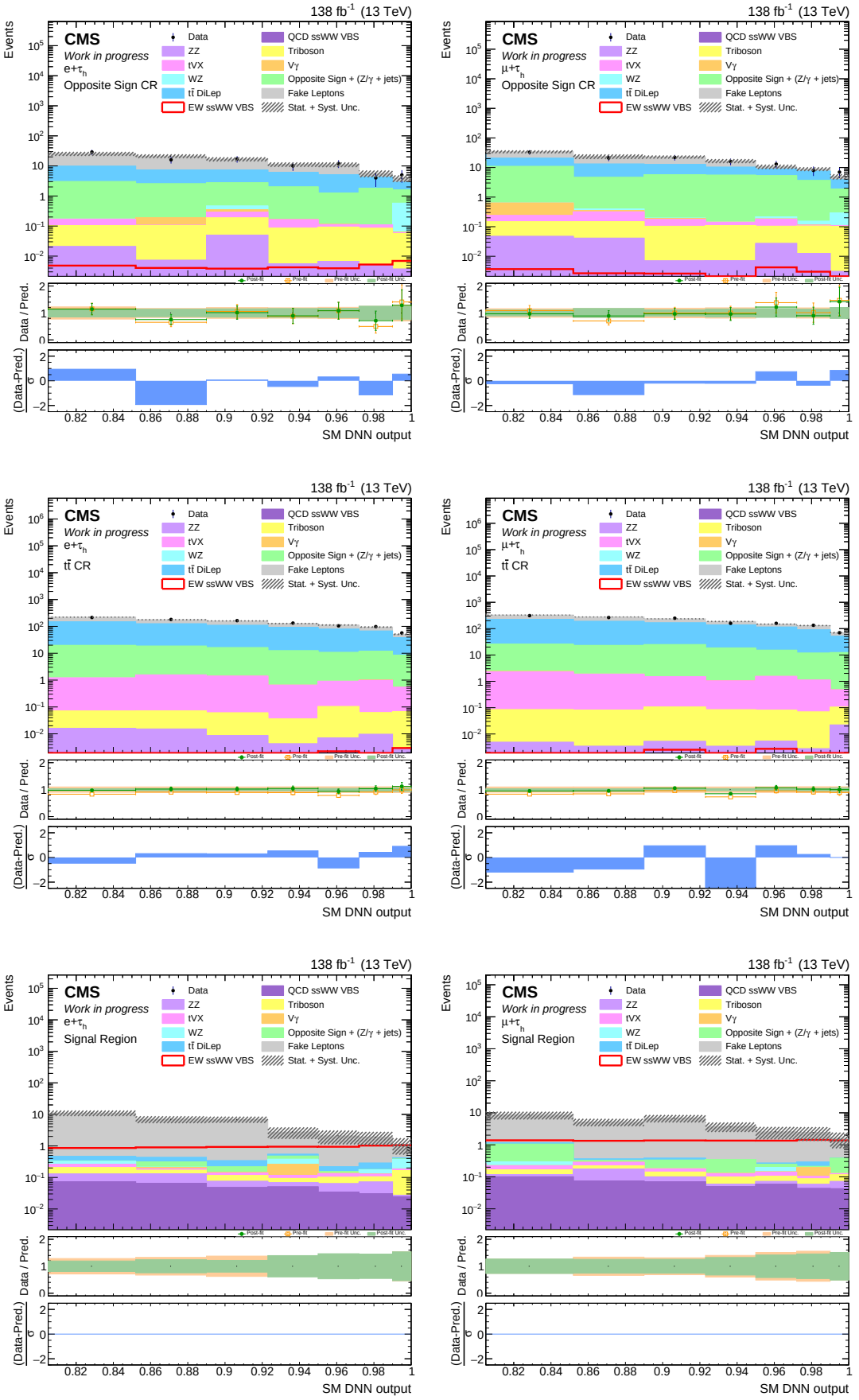


Figure 7.22: Postfit distributions for the SM DNN output in Opposite Sign CR (top), $t\bar{t}$ CR (center) and Signal Region (bottom), and the $e+\tau_h$ (left) and $\mu+\tau_h$ channels (right).
The prefit (light yellow) and postfit (light green) data-to-background agreements and the statistical pulls are shown in the two panels on the bottom of each plot.

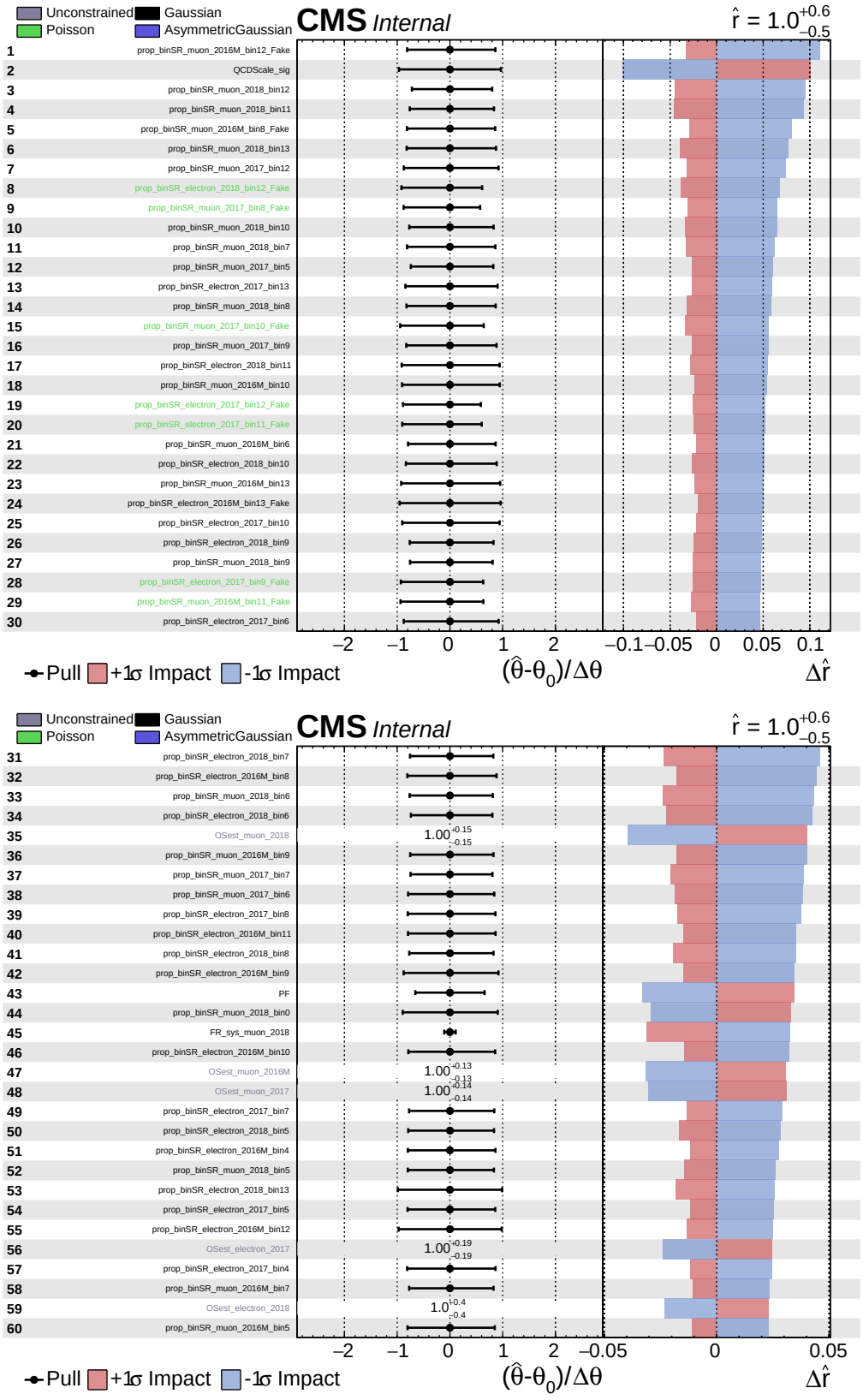


Figure 7.23: Impacts of the most relevant systematic uncertainties for the statistical fit with EW ssWW VBS as the signal, as defined above. The entries prop_bin[region]_[channel+year]_[bin i − th] represent the bin-by-bin statistic uncertainty assigned to all the distributions injected in the fit.

Uncertainty source	$+1\sigma$	-1σ
Theory (PDF, QCD-scale, ISR, and FSR)	$+0.158$	-0.068
Non-prompt estimation	$+0.134$	-0.105
$t\bar{t}$ normalization	$+0.083$	-0.035
Opposite sign + $Z/\gamma+$ jets normalization	$+0.015$	-0.009
Prefiring	$+0.044$	-0.019
Luminosity	$+0.031$	-0.027
$b-$tagging and mistagging	$+0.023$	-0.014
Jet energy scale and resolution, Pile-up jet ID	$+0.001$	-0.006
Pile-up	$+0.016$	-0.003
LO-to-NLO VBS corrections	$+0.015$	-0.005
Unclustered energy	$+0.002$	-0.003
Hadronic tau energy scale and DeepTau	$+0.009$	-0.004
Charge misidentification	$+0.010$	-0.009
Lepton reco, ID, Iso	$+0.001$	-0.001
Bin-by-bin statistics	$+0.247$	-0.221
Data statistics	$+0.467$	-0.417
Total	$+0.577$	-0.491

Table 7.6: Magnitudes of the statistical and systematic uncertainties as extracted by the preliminary fit with the EW ssWW VBS as the signal, expressed in terms of upward ($+1\sigma$) and downward (-1σ) variations of the signal strength.

7.9.1 Measurement of ssWW VBS with one hadronic tau in final state

From the preliminary simultaneous maximum likelihood fit described in Sec. 7.9, preliminary estimates of the expected significance for the EW and EW + QCD ssWW VBS are extracted and shown in Tab. 7.7. The SM DNN discriminator was found to be more powerful than the most sensitive kinematic quantity, the transverse mass M_{o1}.

Method	EW ssWW VBS		EW + QCD ssWW VBS	
	M_{o1}	SM DNN	M_{o1}	SM DNN
Asymptotic Limit	0.923	2.36	0.994	2.52
Toy method	0.942 ± 0.010	$2.23^{+0.08}_{-0.07}$	1.017 ± 0.011	$2.37^{+0.09}_{-0.08}$

Table 7.7: Expected significancies for EW and EW + QCD ssWW VBS signals in the hypothesis of only background contribution. Results extracted with the Asymptotic Limit and the Toy method are reported, the latter also with statistical uncertainty.

In addition, this preliminary result points to possible evidence if the statistical fit to the data is performed also in the signal region. This would be the first evidence ever for the ssWW VBS process with one of the two W bosons producing a hadronic tau lepton.

7.10 Limits on SM Effective Field Theory Operators

The dim-6 and dim-8 DNN discriminators are implemented in the statistical analysis to measure the sensitivity to the anomalous couplings induced separately by the dim-6 and dim-8 SMEFT operators, respectively, with the same signal-flattening binning strategy described in Sec. 7.9. The binning considered is, in fact, the same as that used to extract the distributions for the SM DNN discriminator, since the goal of the statistical study explained here is to limit any anomalous effect that appears on top of the SM contribution. To study the combined effect of all possible SMEFT operator pairs, the distributions of the transverse mass M_{o1} are used to obtain results that are consistent between each pair. In both cases, the setup of the fit and the data set are shared with the measurement of ssWW VBS with one hadronic tau in the final state reported in Sec. 7.9. It should be noted that the SR is blind to the data in these studies as well.

The maximum likelihood fit is based on the considerations reported in Sec. 2.2.3: the starting point is represented by Eq. 2.11. When the effect of a single SMEFT operator is considered, the expected number of events is:

$$
\begin{aligned}
N &= N_{SM} + \frac{c_\alpha}{\Lambda^2} N_\alpha^{int} + \frac{c_\alpha^2}{\Lambda^4} N_\alpha^{pure} && \text{one dim}-6 \text{ SMEFT operator} \\
N &= N_{SM} + \frac{f_i}{\Lambda^4} N_i^{int} + \frac{f_i^2}{\Lambda^8} N_i^{pure} && \text{one dim}-8 \text{ SMEFT operator}
\end{aligned}
\tag{7.6}
$$

where c_α and f_i are the Wilson coefficients for the dim-6 operator Q_α and the dim-8 operator $\mathcal{O}_i$, respectively, N^{int} is the expected number of events due to the interference between the SM and the considered operator (i.e. linear contribution), N^{pure} is the expected number of events due to the single effect of the considered operator (i.e. quadratic contribution). By extracting separately the distributions of the three contributions to N, it is possible to perform a statistical analysis of the data with the CombinedLimit tool, modeled with Eq. 7.6, where the Wilson coefficient is the parameter of interest to fit.

For this purpose, the test statistic q is introduced as follows:

$$q = -2\ln\frac{\mathscr{L}_{s+b}}{\mathscr{L}_s} \tag{7.7}$$

where $\mathscr{L}_{s+b}$ and $\mathscr{L}_b$ are the likelihoods evaluated for the signal-plus-background and background only hypotheses, respectively. This test statistic is usually referred to as the "profiled likelihood ratio" and can also be denoted by $-2\ln\Delta\mathscr{L}$. The expected distribution of q as a function of the Wilson coefficient under consideration is thus estimated to extract the most probable value, the one-dimensional 1σ and 2σ confidence level (CL) intervals for each c_α and f_i. Similarly, when the effects of two operators are combined, Eq. 2.11 takes the following form when all terms up to order Λ^{-4} are retained:

$$N = N_{SM} + \frac{c_\alpha}{\Lambda^2}N_\alpha^{int} + \frac{c_\beta}{\Lambda^2}N_\beta^{int} + \frac{c_\alpha^2}{\Lambda^4}N_\alpha^{pure} + \frac{c_\beta^2}{\Lambda^4}N_\beta^{pure} + \frac{c_\alpha c_\beta}{\Lambda^4}N_{\alpha\beta}^{mix} \qquad \text{dim}-6/\text{dim}-6 \text{ operator pair} \tag{7.8}$$

$$N = N_{SM} + \frac{c_\alpha}{\Lambda^2}N_\alpha^{int} + \frac{c_\alpha^2}{\Lambda^4}N_\alpha^{pure} + \frac{f_i}{\Lambda^4}N_i^{int} \qquad \text{dim}-6/\text{dim}-8 \text{ operator pair} \tag{7.9}$$

With a dedicated statistical fit, the two-dimensional shape of $-2\ln\Delta\mathscr{L}$ as a function of c_α, c_β or c_α, f_i can be estimated to extract the most probable value, 1σ and 2σ confidence level 2D contours for any possible dim-6/dim-6 and dim-6/dim-8 Wilson coefficient pair. These studies are applied to all dim-6 and dim-6 SMEFT operators introduced in Sec. 2.2.

In Fig. 7.24 and Fig. 7.25, selections of the M_{o1} distributions for the "linear" and "quadratic" contributions for the dim-6 and dim-8 SMEFT operators are shown separately for the $e+\tau_h$ and $\mu+\tau_h$ channels, respectively. The complete list of all such distributions is reported in Appendix B.3. These distributions show that, depending on the structure of the SMEFT operators, each of them gives a different composition in the linear and quadratic contributions. Since the former is due to the interference of the SMEFT operator insertions with the SM-only ssWW VBS Feynman diagrams, it appears at intermediate energies and can also take negative values, tending to mimic the SM-only contribution; the latter instead gives the pure contribution of an operator at the high energy scale, significantly enhancing the tails of the kinematic distributions. In general, the contribution of an operator can be relevant already at small values of the corresponding Wilson coefficients when the quadratic term is dominant with respect to the linear one, as is the case for the dim-6 four-quark $\mathscr{O}_{qq}$ operators and the pure bosonic Q_W one (affecting only the W boson gauge self-coupling), since the latter does not provide any cancellation and the high-energy tail is significantly enhanced compared to the SM-only contribution. However, this behavior can easily lead to deviations in unphysical values, such as very high values of the corresponding Wilson coefficient. This tendency can be tamed by the linear counterpart only for very small values of the Wilson coefficient, considering that for a given Wilson coefficient the interference scales linearly and the pure EFT contribution scales quadratically. This reasoning also holds when the effects of two operators are combined, and similar divergences can be observed when mutual interference is not relevant, for example when the two operators act on separate vertices. On the contrary, for an operator with a linear contribution comparable to or even more important than the quadratic one, it is more difficult to constrain its effect, but the divergences can be kept under control since the interference can cancel the high-energy tail enhancement driven by the quadratic counterpart. In particular, the linear term dominates over the quadratic one when the associated operator directly modifies the electroweak input parameters of the theory, such as $Q_{ll}^{(1)}$ in the top row of Fig. 7.24; it results comparable to the quadratic contribution when an operator is constructed with the Higgs

fields (i.e, the dim−6 Q_{H*} and dim−8 $\mathcal{O}_{M*}, \mathcal{O}_{S*}$ operators), due to the properties of the Higgs mechanism. These properties, listed above for the linear and quadratic contributions of the SMEFT operators under study, will be particularly important for interpreting the results of the one- and two-dimensional SMEFT fits.

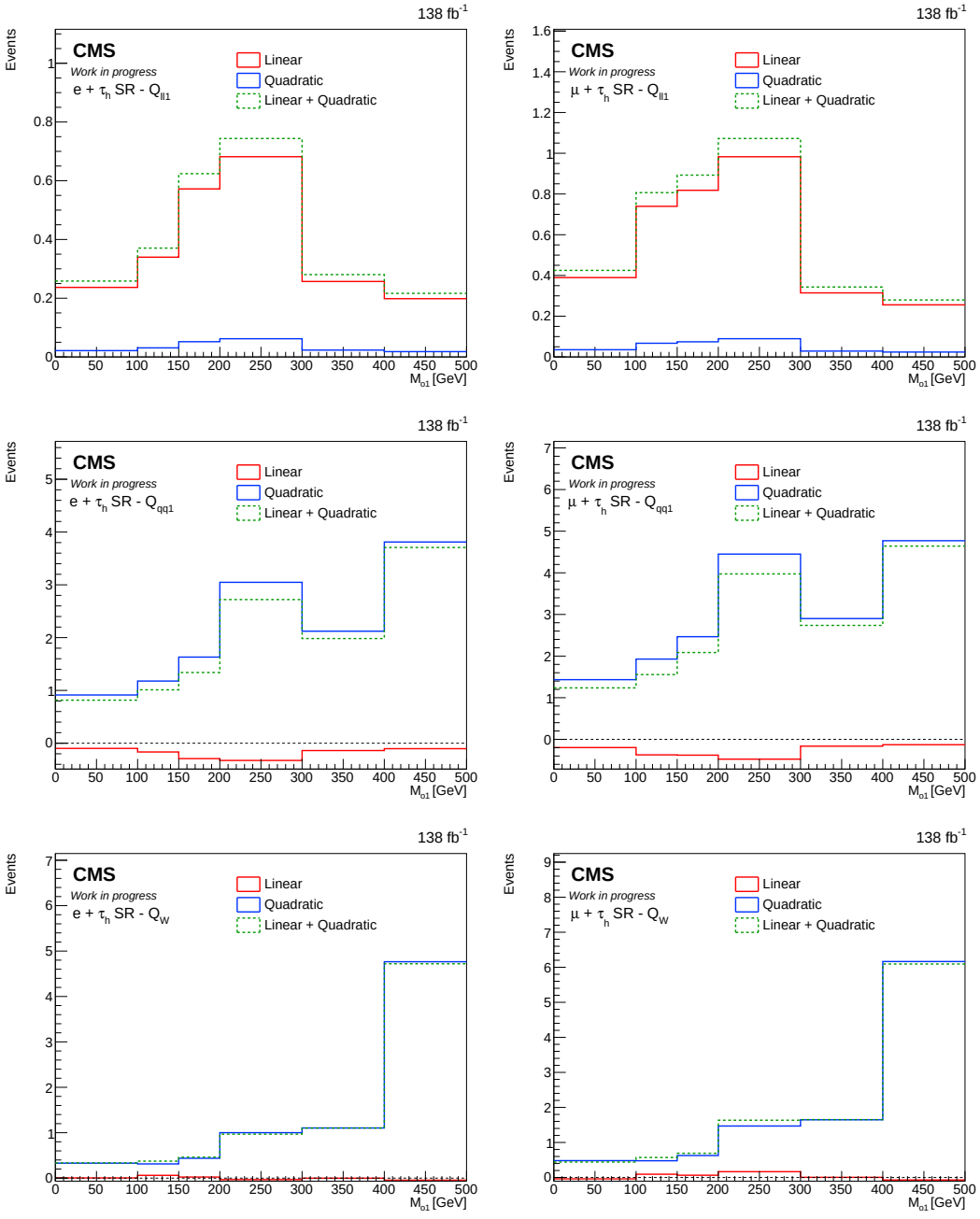


Figure 7.24: M_{o1} distributions for the "linear" and "quadratic", together with their sum, for the operator $Q_{ll}^{(1)}, Q_{qq}^{(1)}, Q_W$ with the full Run II dataset.

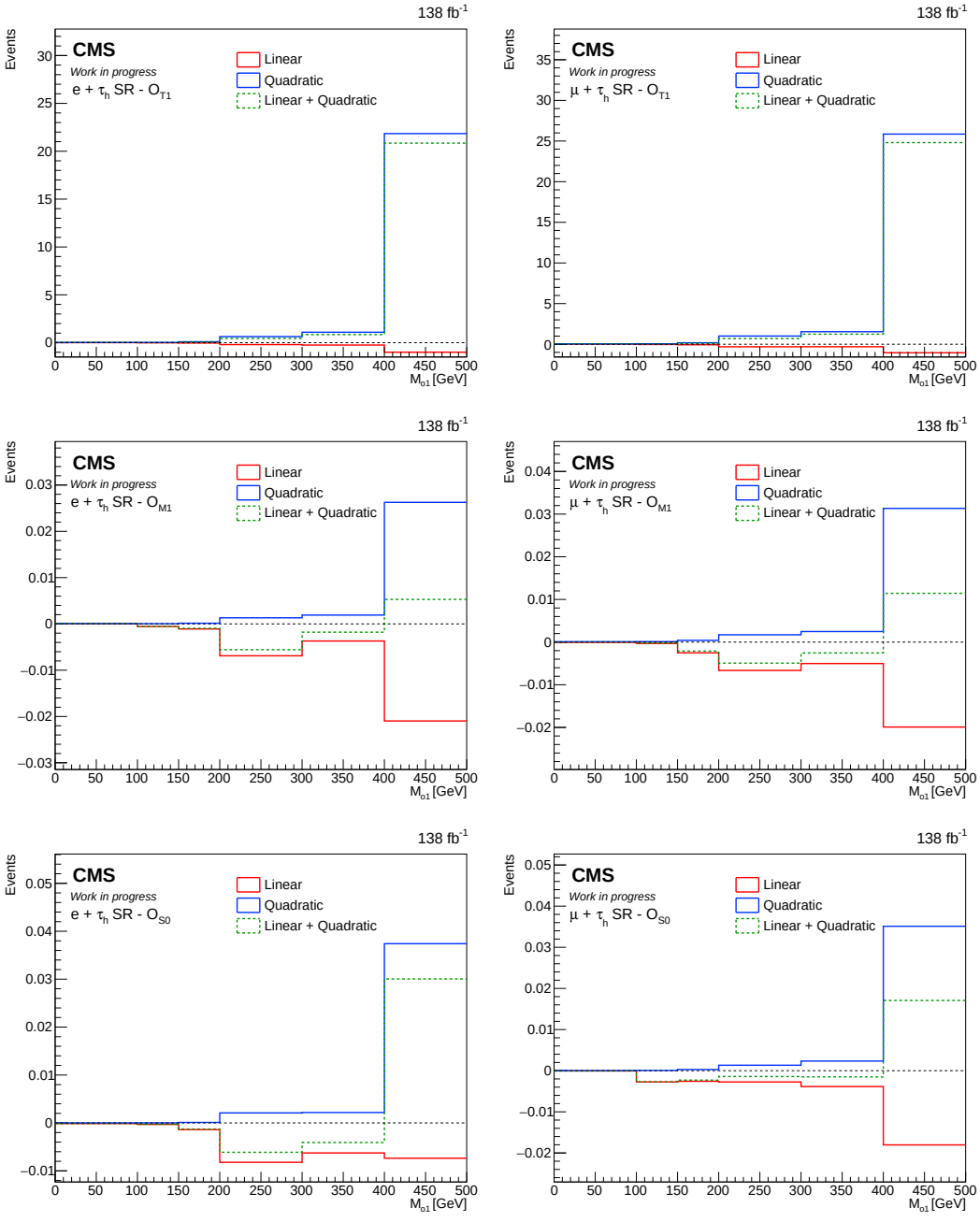


Figure 7.25: M_{o1} for the "linear" and "quadratic", together with their sum, for the operator $\mathcal{O}_{T1}, \mathcal{O}_{M1}, \mathcal{O}_{S0}$ with the full Run II dataset.

The uncertainty breaking is also performed for the one-dimensional SMEFT fits, as reported in Tab. 7.8 for the dim-6 Q_W operator and in Tab. 7.9 for the dim-8 $\mathcal{O}_{T1}$ operator. The bin-by-bin and theoretical uncertainties are the main sources, apart from the limited size of the dataset, similarly to what is reported for the SM ssWW VBS measurement in Tab. 7.6.

Uncertainty source	$+1\sigma$	-1σ
Theory (PDF, QCD-scale, ISR, and FSR)	$+0.074$	-0.074
Non-prompt estimation	$+0.029$	-0.022
$t\bar{t}$ normalization	$+0.031$	-0.031
Opposite sign + Z/γ+ jets normalization	$+0.005$	-0.005
Prefiring	$+0.010$	-0.010
Luminosity	$+0.007$	-0.006
$b-$tagging and mistagging	$+0.010$	-0.010
Jet energy scale and resolution, Pile-up jet ID	$+0.001$	-0.001
Pile-up	$+0.002$	-0.003
LO-to-NLO VBS corrections	$+0.003$	-0.003
Unclustered energy	$+0.002$	-0.002
Hadronic tau energy scale and DeepTau	$+0.001$	-0.001
Charge misidentification	$+0.003$	-0.003
Lepton reco, ID, Iso	$+0.001$	-0.001
Bin-by-bin statistics	$+0.238$	-0.236
Data statistics	$+0.541$	-0.530
Total	$+0.598$	-0.586

Table 7.8: Magnitudes of the statistical and systematic uncertainties as extracted by the preliminary fit with the dim-6 Q_W operator as signal, expressed in terms of upward ($+1\sigma$) and downward (-1σ) variations of the signal strength.

Uncertainty source	$+1\sigma$	-1σ
Theory (PDF, QCD-scale, ISR, and FSR)	$+0.057$	-0.052
Non-prompt estimation	$+0.005$	-0.006
$t\bar{t}$ normalization	$+0.020$	-0.019
Opposite sign + Z/γ+ jets normalization	$+0.003$	-0.002
Prefiring	$+0.006$	-0.006
Luminosity	$+0.003$	-0.004
$b-$tagging and mistagging	$+0.006$	-0.005
Jet energy scale and resolution, Pile-up jet ID	$+0.001$	-0.001
Pile-up	$+0.003$	-0.002
LO-to-NLO VBS corrections	$+0.000$	-0.000
Unclustered energy	$+0.000$	-0.001
Hadronic tau energy scale and DeepTau	$+0.001$	-0.000
Charge misidentification	$+0.001$	-0.001
Lepton reco, ID, Iso	$+0.001$	-0.000
Bin-by-bin statistics	$+0.127$	-0.119
Data statistics	$+0.299$	-0.256
Total	$+0.331$	-0.288

Table 7.9: Magnitudes of the statistical and systematic uncertainties as extracted by the preliminary fit with the dim-8 $\mathcal{O}_{T1}$ operator as signal, expressed in terms of upward ($+1\sigma$) and downward (-1σ) variations of the signal strength.

7.10.1 One-dimensional constraints on SMEFT dim-6 and dim-8 Wilson coefficients

With the strategy outlined above, the expected 1σ and 2σ confidence level intervals on the Wilson coefficients of the dim-6 and dim-8 SMEFT operators under study are extracted and reported in Tab. 7.10, together with the corresponding minima. The results show again that the dim-6 and dim-8 DNN discriminators are more sensitive to the possible anomalous couplings than the transverse mass M_{o1}. Since this is the first time that the effects of the dim-6 operators, as defined in the Warsaw basis, are investigated in a VBS analysis, no direct comparison with previously observed results is shown. For the dim-8 operators, the expected results are compared with the observed ones obtained in [220], where their effect is constrained in ssWW and WZ VBS with a final state consisting of two light-charged leptons and two neutrinos, together with the two VBS jet pairs. It is important to note first of all that no bias is induced by performing the fits with the dim-6/dim-8 DNN discriminators instead of the kinematic quantity M_{o1}, since the effect of the former is only to further constrain the CL intervals extracted with the latter, without introducing any systematic shift.

As expected, it can be seen that the SMEFT operators associated with the most constrained Wilson coefficients in this analysis are those that present a predominant quadratic contribution. In fact, for the dim-6 operators, this analysis shows to be more sensitive to the pure fermionic, the mixed built with Higgs and fermionic fields (especially when the charged current is a SU(2) triplet), and the pure bosonic Q_W operators. The operators with a more pronounced linear contribution are expected to have a double-minimum structure in the form of $-2\ln\Delta\mathscr{L}$, since it is odd with respect to the associated Wilson coefficient, such as $Q_{ll}^{(1)}, Q_{Hl}^{(3)}$. If the sensitivity of one of these operators is such that it can be resolved at $\pm1\sigma$ CL, two separate intervals at the same CL are extracted, each centered on one of the minima; otherwise, a single interval not centered on and containing both minima is obtained.

For the dim-8 operators, the comparison between the M_{o1} distributions obtained with the 2σ CL interval extremes presented in [220] and those of the SM EW ssWW VBS shows that the intrinsic sensitivity of the channel considered in this study to the dim-8 effects differs from the sensitivity of [220] on operator basis. In fact, it can be observed that, moving from the transverse operators $\mathcal{O}_{T0,1,2}$ to the mixed operators $\mathcal{O}_{M0,1,7}$ and then to the longitudinal operators $\mathcal{O}_{S0,1}$, the contribution of the dim-8 operator in the tail closes the gap with the pure SM contribution and finally starts to rise above it. This behavior is also responsible for the ratios between the observed CL intervals in [220] and the expected ones extracted in this analysis. The conclusion is that the studied final state is more sensitive to anomalous couplings involving the Higgs field than the ssWW VBS process with a leptonic final state.

To get a better insight into the sensitivities to the different SMEFT operators just described, in Fig. 7.26 and Fig. 7.27 the M_{o1} distributions are shown for a selection of the dim-6 and dim-8 operators under study, respectively, according to Eq. 7.6, using for the Wilson coefficients the values corresponding to the extremes of the 2σ ($\pm2\sigma$) confidence level intervals extracted with the dim-6/dim-8 DNN discriminator. For the dim-8 operators, similar distributions obtained with the results reported in [220] are shown for comparison. Such distributions for the full sets of dim-6 and dim-8 operators can be found in Appendix B.3.

Wilson coefficient		Minimum(a) position(s)	1σ CL interval(s) M_{o1}	1σ CL interval(s) DNN	2σ CL interval M_{o1}	2σ CL interval DNN	current best limit
dim−6	$c_{ll}^{(1)}$	$-10.9, 0.0$	$[-14.8, -5.43] \cup [-5.43, 3.73]$	$[-13.3, -7.64] \cup [-2.89, 2.12]$	$[-17.7, 6.47]$	$[-15.4, 4.02]$	
	$c_{qq}^{(1)}$	0.0	$[-0.629, 0.665]$	$[-0.402, 0.445]$	$[-0.927, 0.964]$	$[-0.620, 0.665]$	
	$c_{qq}^{(1,1)}$	0.0	$[-0.042, 0.041]$	$[-0.033, 0.033]$	$[-0.075, 0.072]$	$[-0.060, 0.059]$	
	$c_{qq}^{(3)}$	0.0	$[-0.041, 0.040]$	$[-0.033, 0.033]$	$[-0.075, 0.072]$	$[-0.060, 0.058]$	
	$c_{qq}^{(3,1)}$	0.0	$[-0.124, 0.121]$	$[-0.098, 0.104]$	$[-0.224, 0.203]$	$[-0.176, 0.172]$	
	c_W	0.0	$[-0.586, 0.598]$	$[-0.448, 0.456]$	$[-0.869, 0.880]$	$[-0.692, 0.697]$	
	c_{HW}	0.0	$[-5.88, 5.58]$	$[-4.79, 4.25]$	$[-8.67, 8.34]$	$[-7.37, 6.79]$	
	c_{HWB}	0.0	$[-74.4, 98.7]$	$[-61.7, 88.5]$	$[-128, 156]$	$[-115, 143]$	
	$c_{H\Box}$	0.0	$[-14.5, 15.1]$	$[-11.3, 11.9]$	$[-23.2, 23.8]$	$[-19.3, 19.9]$	
	c_{HD}	0.0	$[-26.6, 30.5]$	$[-20.1, 23.9]$	$[-43.7, 48.8]$	$[-35.5, 41.0]$	
	$c_{Hl}^{(1)}$	0.0	$[-31.8, 31.2]$	$[-24.6, 24.1]$	$[-56.4, 54.9]$	$[-55.3, 53.0]$	
	$c_{Hl}^{(3)}$	$0.0, 7.8$	$[-2.77, 10.6]$	$[-1.58, 2.10] \cup [5.86, 9.60]$	$[-4.78, 12.7]$	$[-2.98, 11.1]$	
	$c_{Hq}^{(1)}$	0.0	$[-4.12, 4.17]$	$[-3.09, 3.05]$	$[-6.14, 6.17]$	$[-4.80, 4.73]$	
	$c_{Hq}^{(3)}$	0.0	$[-2.16, 1.82]$	$[-1.54, 1.26]$	$[-3.21, 2.76]$	$[-2.33, 1.98]$	
dim−8	f_{T0}	0.0	$[-0.697, 0.742]$	$[-0.463, 0.492]$	$[-1.04, 1.09]$	$[-0.748, 0.778]$	$[-0.28, 0.31]$
	f_{T1}	0.0	$[-0.287, 0.331]$	$[-0.193, 0.224]$	$[-0.435, 0.479]$	$[-0.317, 0.349]$	$[-0.12, 0.15]$
	f_{T2}	0.0	$[-0.789, 0.973]$	$[-0.541, 0.689]$	$[-1.21, 1.39]$	$[-0.902, 1.06]$	$[-0.38, 0.50]$
	f_{M0}	0.0	$[-6.98, 6.83]$	$[-4.59, 4.47]$	$[-10.5, 10.3]$	$[-7.36, 7.24]$	$[-3.0, 3.2]$
	f_{M1}	0.0	$[-8.86, 9.56]$	$[-5.87, 6.49]$	$[-13.6, 14.3]$	$[-9.67, 10.3]$	$[-4.7, 4.7]$
	f_{M7}	0.0	$[-14.9, 13.9]$	$[-9.93, 9.05]$	$[-22.7, 21.6]$	$[-15.9, 15.0]$	$[-6.7, 7.0]$
	f_{S0}	0.0	$[-7.92, 8.28]$	$[-5.16, 5.49]$	$[-12.2, 12.6]$	$[-8.53, 8.84]$	$[-6.0, 6.4]$
	f_{S1}	0.0	$[-29.6, 31.0]$	$[-19.0, 20.0]$	$[-49.4, 50.9]$	$[-19.0, 20.0]$	$[-18, 19]$
	f_{S2}	0.0	$[-29.6, 31.0]$	$[-19.0, 20.0]$	$[-49.4, 50.9]$	$[-19.0, 20.0]$	$[-18, 19]$

Table 7.10: 1σ and 2σ confidence level (CL) intervals on the Wilson coefficients associated to the SMEFT dim-6 and dim-8 under investigation extracted with the fit procedure illustrated in Sec. 7.10. For the dim-8 Wilson coefficients, the results are compared to the observed ones presented in [220].

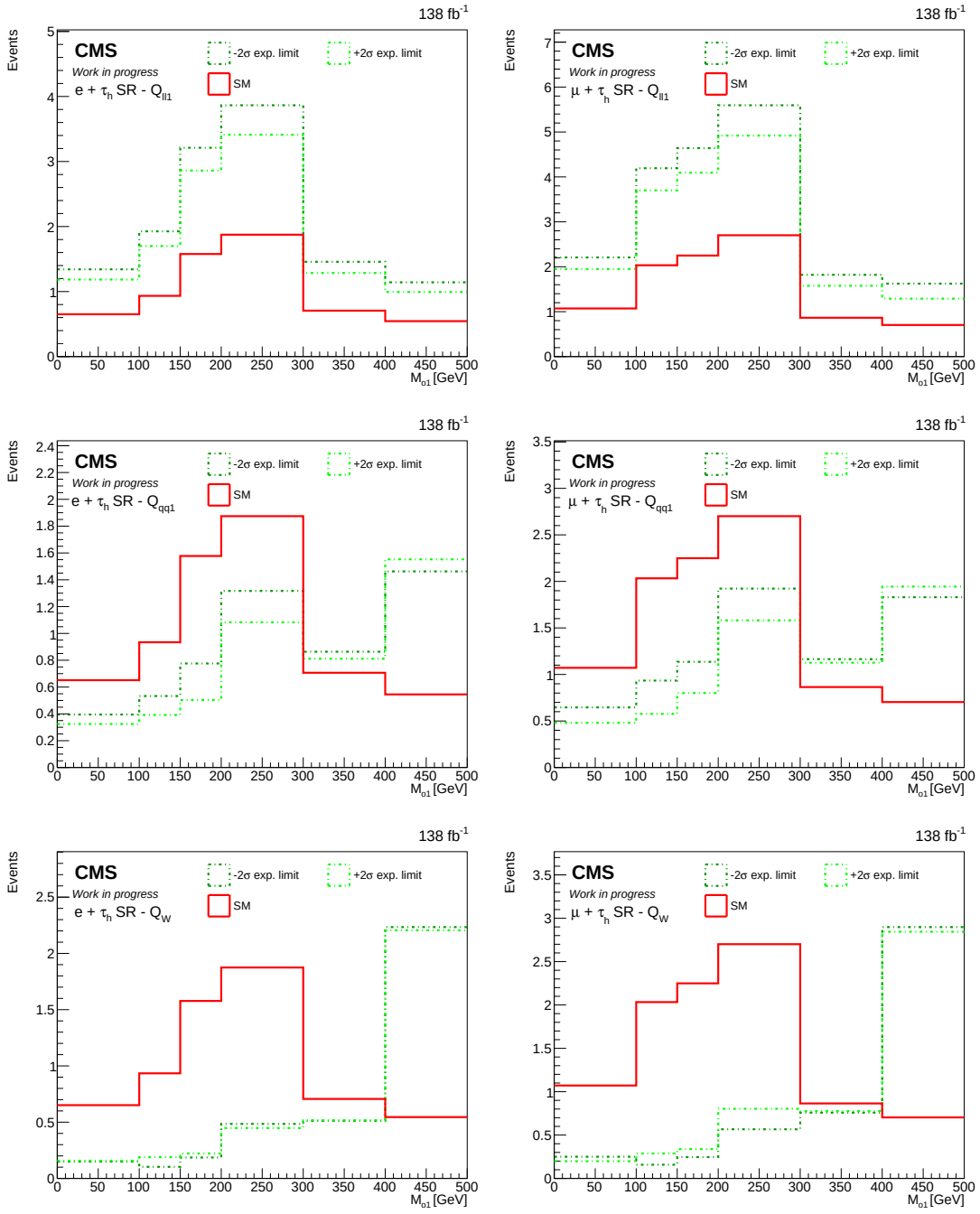


Figure 7.26: M_{o1} distributions in SR for EW ssWW VBS (red) and the SMEFT dim-6 operators $Q_{ll}^{(1)}, Q_{qq}^{(1)}$, when the Wilson coefficients equal the upper ($+2\sigma$) and lower (-2σ) extremes of 2σ CL intervals from this analysis (green) reported in Tab. 7.10.

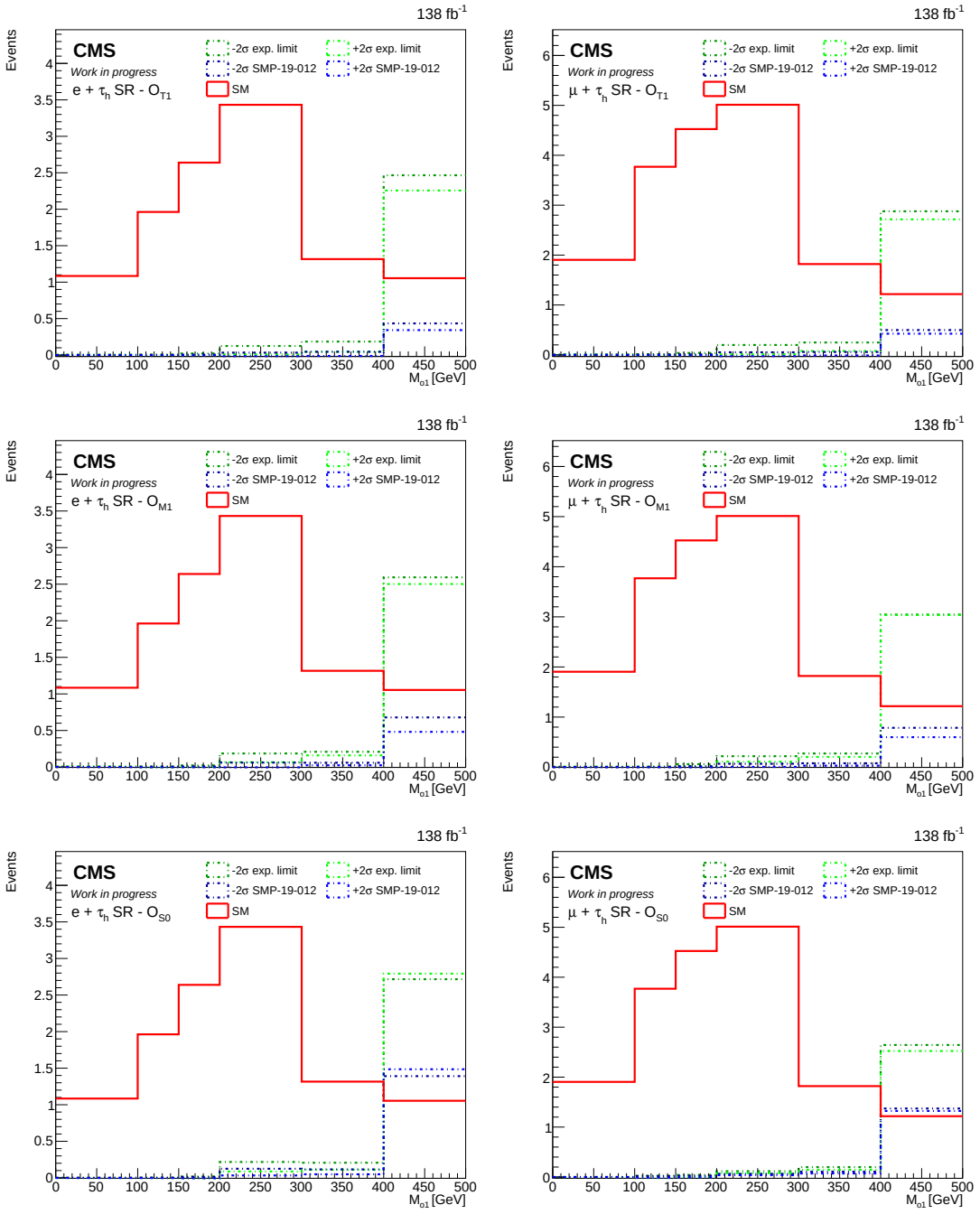


Figure 7.27: M_{o1} distributions in SR for EW ssWW VBS (red) and the SMEFT dim−8 operators $\mathcal{O}_{T1}, \mathcal{O}_{M1}, \mathcal{O}_{S0}$, when the Wilson coefficients equal the upper ($+2\sigma$) and lower (-2σ) extremes of 2σ CL intervals from this analysis (green) and [220] (blue) reported in Tab. 7.10.

7.10.2 Two-dimensional constraints on SMEFT dim-6 and dim-8 Wilson coefficients

The combined effect of two SMEFT operators at the same time is investigated by the dedicated statistical fit shown at the beginning of this Section. The preliminary results, useful for highlighting their main features, are presented from Tab. 7.11 to Tab. 7.13 for a selection of dim-6/dim-6 operator pairs, and in Tab. 7.15 and Tab. 7.16 for a selection of dim-6/dim-8 operator pairs as the expected two-dimensional distributions of $-2\ln\Delta\mathscr{L}$. These selections are useful to highlight the main features of this study in the following, and the full collection of two-dimensional bounds on the SMEFT operator pairs is reported in Appendix B.4. A comparison between the one-dimensional limits reported in Tab. 7.10 and the results presented in this paragraph can be made by considering the points on the 1σ and 2σ two-dimensional contours with one of the two Wilson coefficients null.

Focusing first on the results for the dim-6/dim-6 pairs, the combination of the effects of two operators, modeled with the Eq. 7.8, leads to a broadening of the 1σ and 2σ intervals extracted with the corresponding one-dimensional fit, depending on which operator is used. In general, the contours have an elliptical shape with correlation effects for operators acting on the same vertex and a non-negligible linear term. Thus, in this case, the potential divergences caused by the quadratic terms are balanced by the interference terms. The prevalence of the linear terms for $Q_{Hl}^{(3)}$ and $Q_{ll}^{(1)}$ leads to the double-minima shape of the $-2\ln\Delta\mathscr{L}$ for the operator pairs in which they are included. In addition, some flat directions can be detected as diagonal bands in the two-dimensional distributions, such as for the pairs $c_{qq}^{(3)}$ vs. $c_{qq}^{(1,1)}$, $c_{ll}^{(1)}$ vs. $c_{Hl}^{(3)}$, $c_{Hq}^{(3),(1)}$ vs. $c_{Hl}^{(1)}$, c_{HD} vs. c_W. In these directions, the sensitivity to anomalous couplings is very poor.

Finally, particular attention must be paid to the $-2\ln\Delta\mathscr{L}$ distributions for the pairs involving the four-quark operators $Q_{qq}^{(3)}, Q_{qq}^{(1,1)}, Q_{qq}^{(3,1)}$. As pointed out in the previous paragraph, these operators have a dominant quadratic contribution and a negligible linear one, a fact that can easily lead to divergences. The remaining four-quark operator, $Q_{qq}^{(1)}$, instead presents a non-negligible linear contribution, even if the quadratic one is dominant. These features lead to the $-2\ln\Delta\mathscr{L}$ shapes extracted for the pair involving this class of dim-6 operators: when they are combined among themselves, Q_{Hq} operators, or Q_W operator, the possible divergence is resolved by their mutual interference; when they are combined with other operators, the two-dimensional contours "explode", except for the operator $Q_{qq}^{(1)}$.

Regarding the two-dimensional $-2\ln\Delta\mathscr{L}$ shapes extracted for the dim-6/dim-8 operator pairs, the addition of a linear dim-8 contribution to the total contribution of a dim-6 operator, modeled with the Eq. 7.9, leads to a slight broadening of the one-dimensional 1σ and 2σ intervals for the dim-6 operators. The range of values does not vary significantly from one dim-8 operator to another. The shape of the contours depends mainly on the dim-6 operator included in the pair, with the orientation depending on the interplay between the dim-8 linear and dim-6 quadratic contributions. It is worth noting that the $-2\ln\Delta\mathscr{L}$ extracted for the dim-6/dim-8 are symmetric with respect to the dim-6 Wilson coefficients. Using Eq. 7.9, this can be understood by noting that the additional dim-8 linear term offsets the dim-6 quadratic contribution, and thus the dim-6 linear term becomes the one driving the dependence of $-2\ln\Delta\mathscr{L}$ on the dim-6 Wilson coefficient. These considerations lead to the conclusion that, first of all, the one-dimensional constraints on the dim-8 Wilson coefficients are significantly biased by neglecting the contributions of the dim-6 operators arising in the same physical process and at the same or smaller order of power of Λ_{NP}, since the region constrained in the dim-8 direction by the combined dim-6/dim-8 does not differ from one dim-8 Wilson coefficient to another. Conversely, the linear term of each dim-8 operator reduces the importance of any dim-6 quadratic contribution, helping to make the linear term the driving one for determining the constraints on the dim-6 Wilson coefficients.

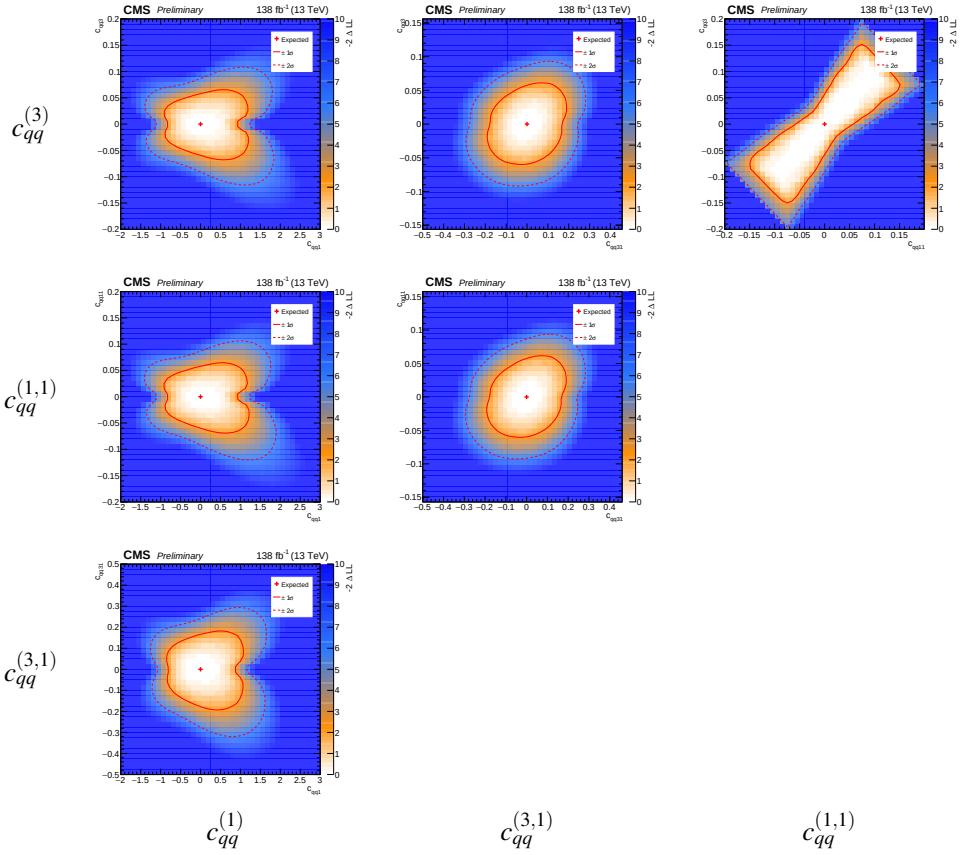


Table 7.11: Selection of the expected 2D $-2\ln\Delta\mathscr{L}$ as functions of the reported pure fermionic dim-6 Wilson coefficient pairs. 1σ and 2σ CL contours are superimposed.

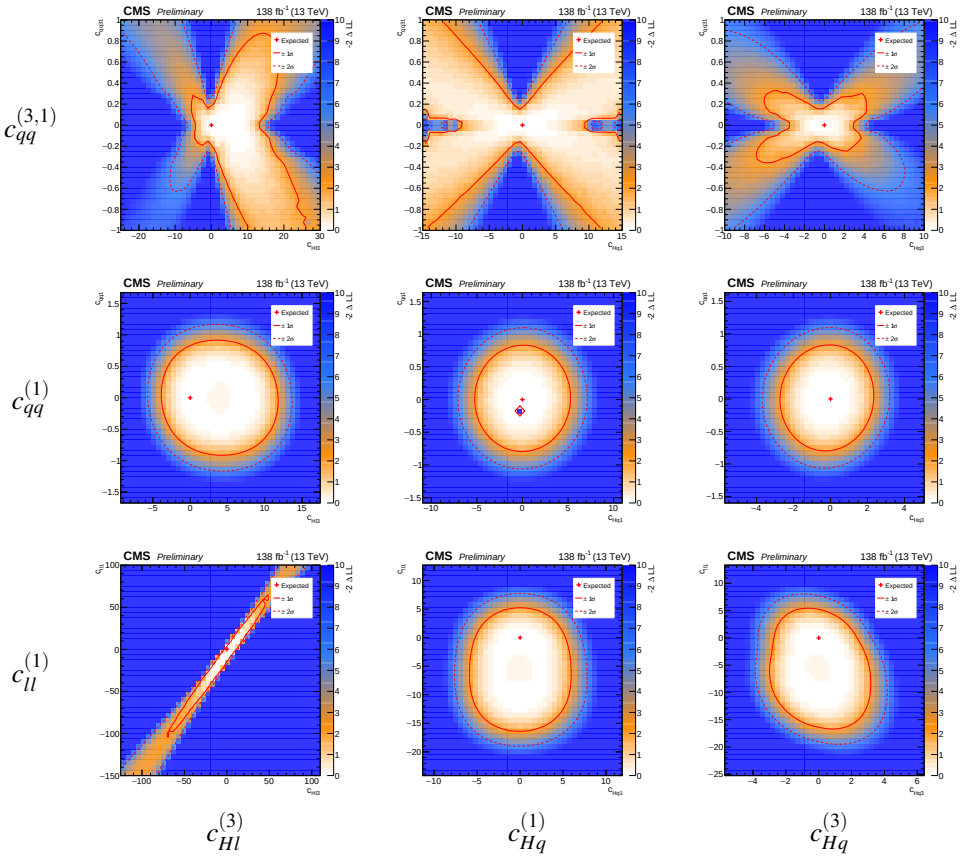


Table 7.12: Selection of the expected 2D $-2\ln\Delta\mathscr{L}$ as functions of the dim-6 pure fermionic versus mixed dim-6 Wilson coefficient pairs. 1σ and 2σ CL contours are superimposed.

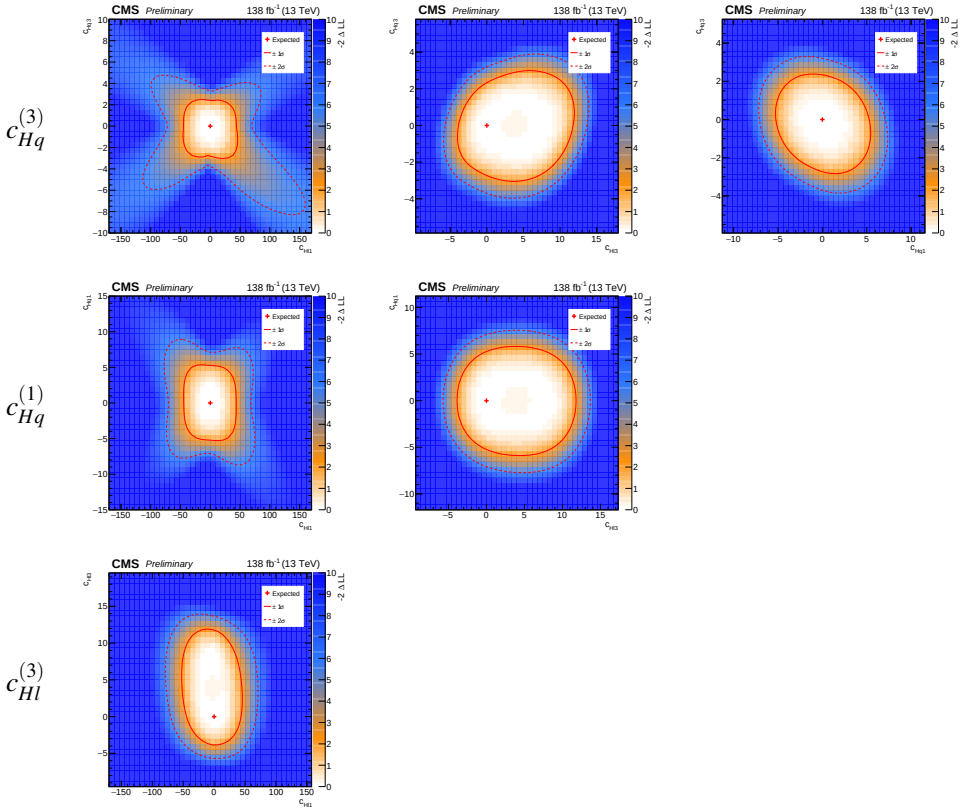


Table 7.13: Selection of the expected 2D $-2\ln\Delta\mathscr{L}$ as functions of the mixed dim-6 Wilson coefficient pairs. 1σ and 2σ CL contours are superimposed.

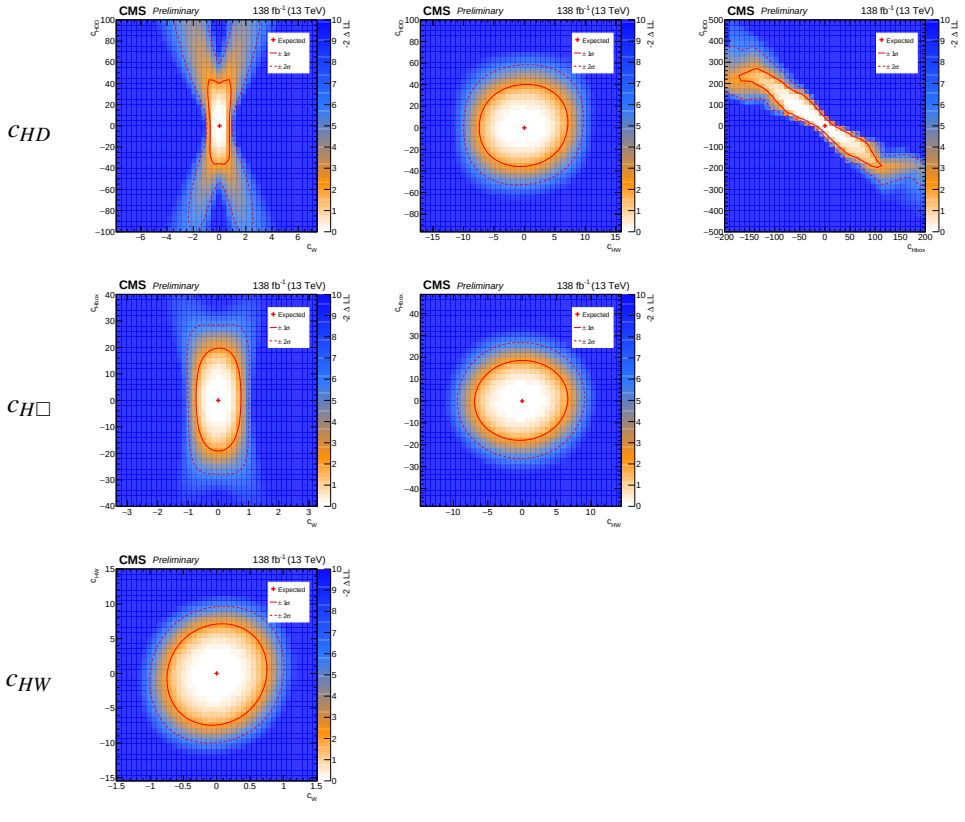


Table 7.14: Selection of the expected two-dimensional $-2\ln\Delta\mathscr{L}$ as functions of the pure bosonic dim-6 Wilson coefficient pairs. 1σ and 2σ CL contours are superimposed.

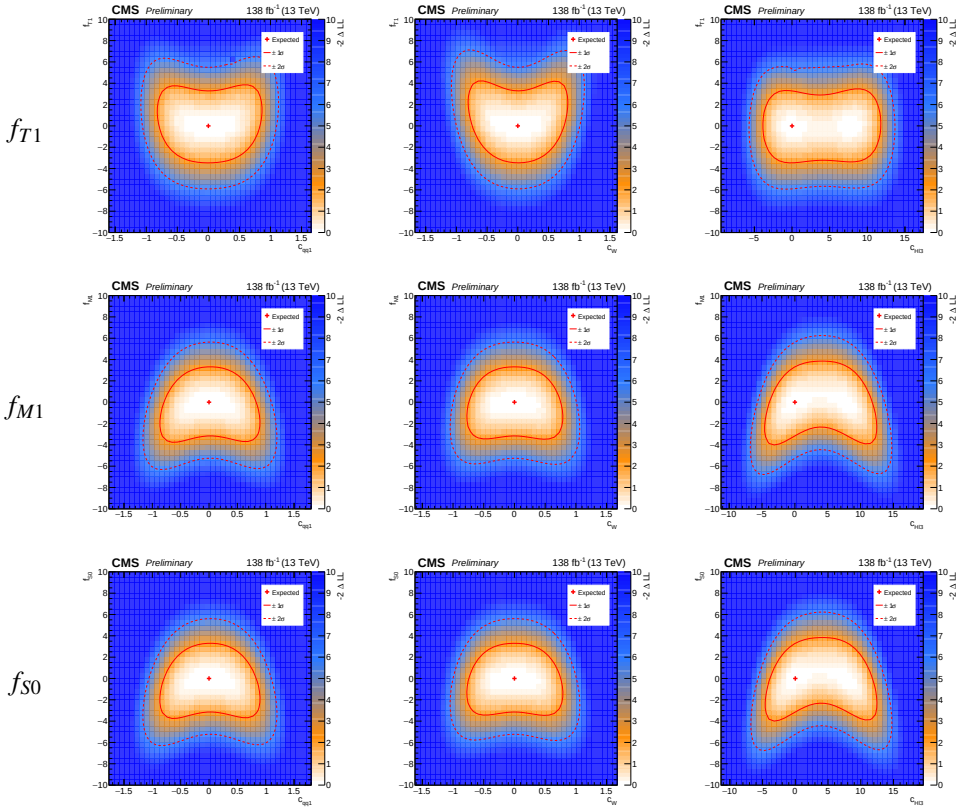


Table 7.15: Selection of the expected 2D $-2\ln\Delta\mathscr{L}$ as functions of the dim−6/dim−8 Wilson coefficient pairs. 1σ and 2σ CL contours are superimposed.

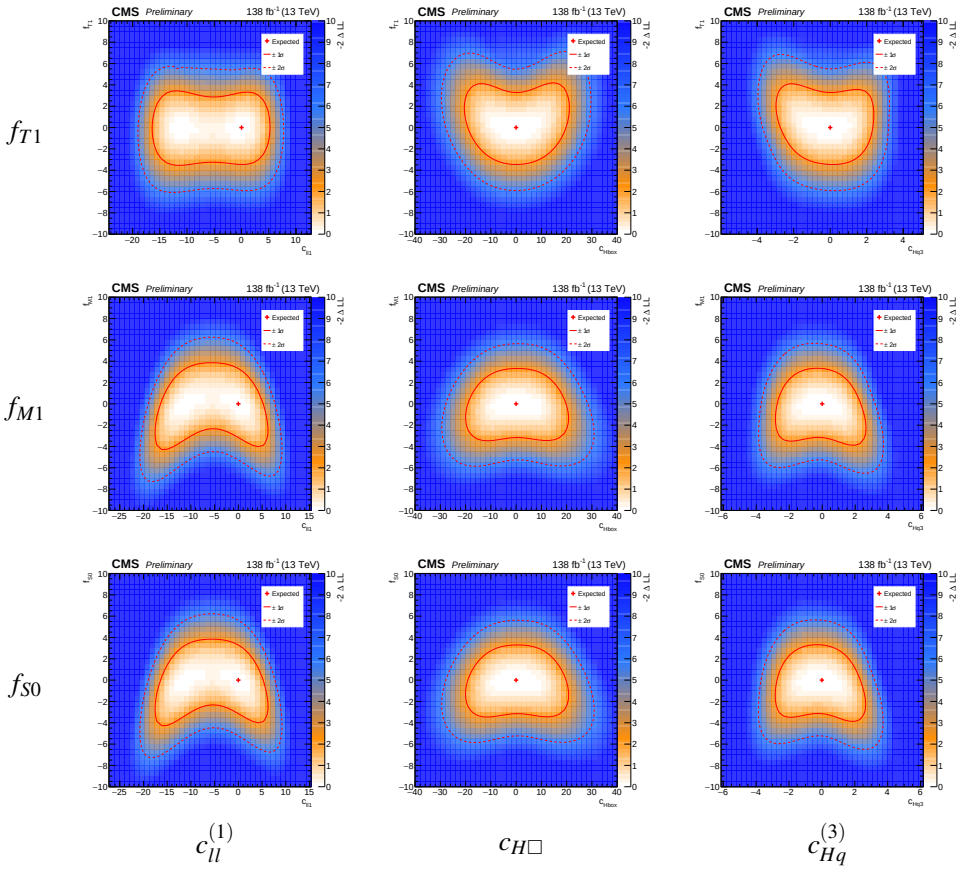


Table 7.16: Selection of the expected 2D $-2\ln\Delta\mathscr{L}$ as functions of the dim−6/dim−8 Wilson coefficient pairs. 1σ and 2σ CL contours are superimposed.

Conclusion

In this work, two different studies with the same goal to search for New Physics signals within the electroweak sector of the SM of particle physics are presented, exploiting the proton-proton collisions with center-of-mass energy $\sqrt{s} = 13$ TeV collected by the CMS experiment at the LHC from 2016 to 2018, and corresponding to an integrated luminosity of 138 fb^{-1}.

First, a search for a new heavy vector boson W' decaying in top-bottom quark pair, with the top quark decaying to another bottom quark and a leptonic pair $\ell\nu_l\,(\ell = e, \mu)$ is performed. To consider an as large as possible array of New Physics scenarios, several hypotheses for the new W' boson are investigated, depending on the mass and the decay width of the new resonance, as well as the coupling to left-handed (LH) and right-handed (RH) fermions.

The selection requires the presence of exactly one lepton, a significant amount of transverse missing momentum, and at least two jets. Requirements are applied to these objects to reject low-energy backgrounds and, in particular, QCD-induced multijet events. To reduce the combinatorial background and best exploit the signal topology, the top quark is reconstructed starting from the final state objects using only kinematic criteria aiming at the correct jet assignment.

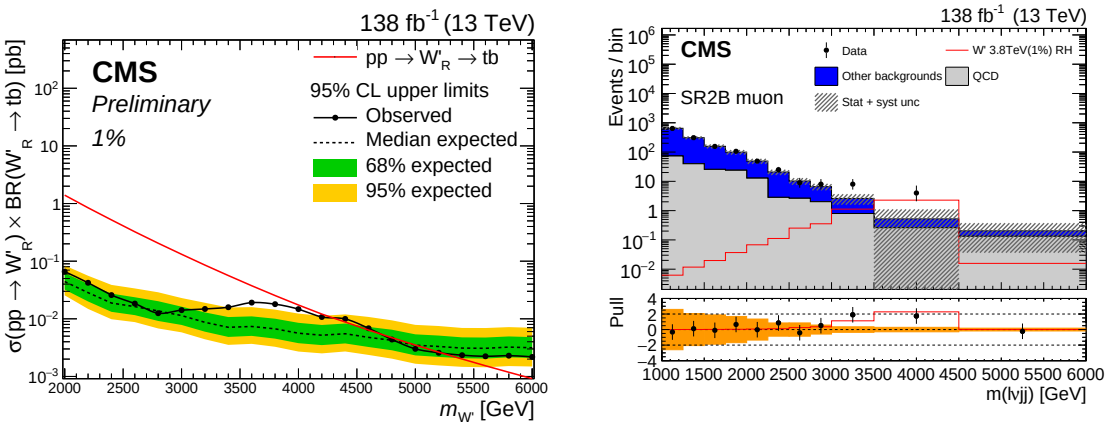


Figure C.1: Left, observed 95% CL upper limit on the cross-section for the W' production (solid black line), together with the corresponding expected limit in the hypothesis of background-only contribution (dashed blue line) and theoretical prediction (solid red line); the green and yellow stripes represent, respectively, the 1σ and 2σ for the expected upper limit due to the overall uncertainty. Right, the postfit stacked distributions for the reconstructed W' mass in the most sensitive region for the SM background processes and the events observed in data, with a 2.6σ excess centered among $m_{W'} = 3.8$ TeV; the distribution expected for a signal generated by a hypothetical W' boson with mass equals to 3.8 TeV is superimposed (red solid line).

The events are categorized depending on whether the jets are tagged as coming from b quarks. After the categories are defined, additional requirements are applied to the top quark

mass and the mass of the wide jet including the narrow jet used in the W' reconstruction. These requirements are also inverted to obtain control regions from which the main backgrounds are extracted and employ dedicated transfer functions used in the signal regions.

Performing extended maximum likelihood fits for the various W' scenarios under investigation, this analysis was able to improve the previous searches excluding LH and RH W' bosons with masses below 3.9 and 4.3 TeV, respectively, and extending the search to hypothesized masses up to 6 TeV. The largest deviation from the SM expectations results in a mild local excess of 2.6 standard deviations in correspondence of $m_{W'} = 3.8$ TeV for an RH W' in the narrow width hypothesis. In Fig. C.1 the observed upper limit on the cross-section production for this scenario is shown on the left, as well as the postfit distributions extracted in the most sensitive signal region showing the excess mentioned above on the right.

In the second investigation, hints of New Physics are searched for in the context of the precision measurements of SM processes involving the Higgs and the electroweak sectors very susceptible to any possible deviation from the SM expectations arising from New Physics phenomena, even in the hypothesis that new resonances cannot be observed with the current experiments. In particular, a search is performed for a same-charge W boson pair through VBS processes (ssWW VBS) and related anomalous couplings, by exploiting for the first time the decay of one of the W bosons to a τ lepton and its subsequent hadronic decay (hadronic tau, τ_h), resulting in an original final state consisting of $\ell\tau_h\nu_\ell\nu_\tau jj$.

The possible anomalies induced by indirect New Physics effects in the boson-boson and boson-fermion couplings are investigated and interpreted with the agnostic approach of the SM Effective Field Theory (SMEFT), introducing operators constructed with SM fields with dimension six (dim-6 operators) and eight (dim-8 operators) and their magnitudes expressed via the Wilson coefficients. It is the first time that SMEFT operators with different dimensions are considered with the same channel.

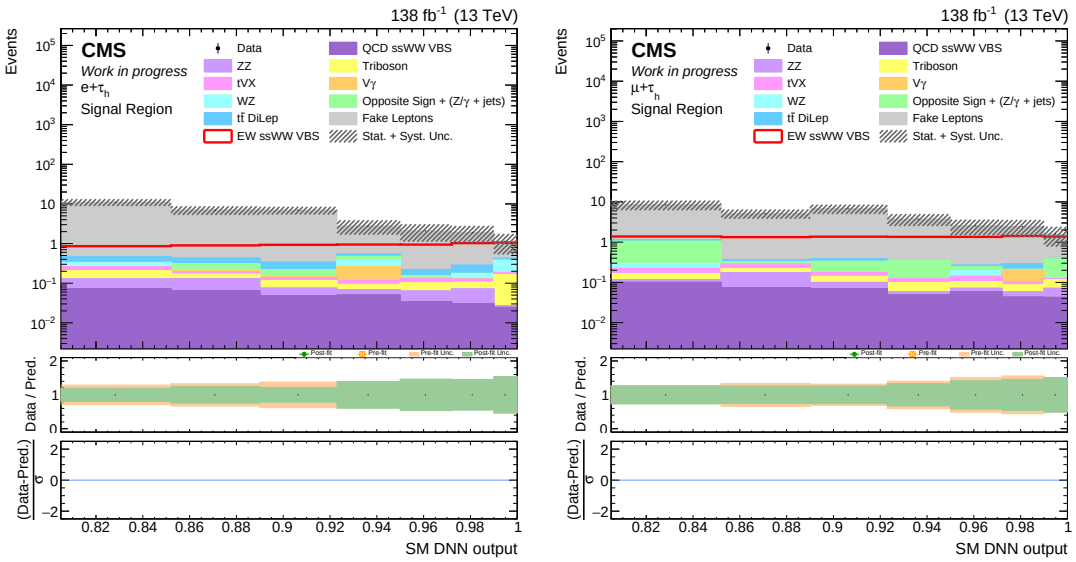


Figure C.2: Postfit distributions for the SM DNN output implemented in the ssWW VBS search in the signal region, blinded to data, in the $e + \tau_h$ (left) and $\mu + \tau_h$ (right) channels. The prefit (light yellow) and postfit (light green) data-to-background agreements and the statistical pulls are shown in the two panels at the bottom of each plot.

The first selection requires the presence of one electron or muon, one hadronic tau, and two energetic hadronic jets compatible with the signature of the ssWW VBS events. Kinematic quantities capable to separate the signal events from the background are computed, introducing

two new transverse masses for the leptonic system, which are M_{o1} and M_{1T}. Defining a collection of events presenting hadronic jets misreconstructed as an electron, muon, or hadronic taus, a data-driven method is applied to estimate the main source of background for this study, represented by events with at least one non-prompt lepton or non-prompt hadronic tau. The events are then categorized to define a VBS signal region and additional control regions, with the latter ones employed to validate the estimation of the non-prompt leptons and taus, as well as constrain the remaining relevant sources of background.

To enhance the sensitive power of the study, three deep neural networks (DNNs) are developed with the simulated events falling in the signal region to discriminate between the SM background processes and the three types of signals, SMEFT dim-6, and SMEFT dim-8. The corresponding output probabilities, together with the transverse mass M_{o1}, represents thus the basis of the statistical inference procedure, preliminarily applied with the data from proton-proton collisions with a center-of-mass energy of $\sqrt{s} = 13$ TeV, corresponding to 138 fb^{-1}, collected by the CMS experiment at the LHC from 2016 to 2018.

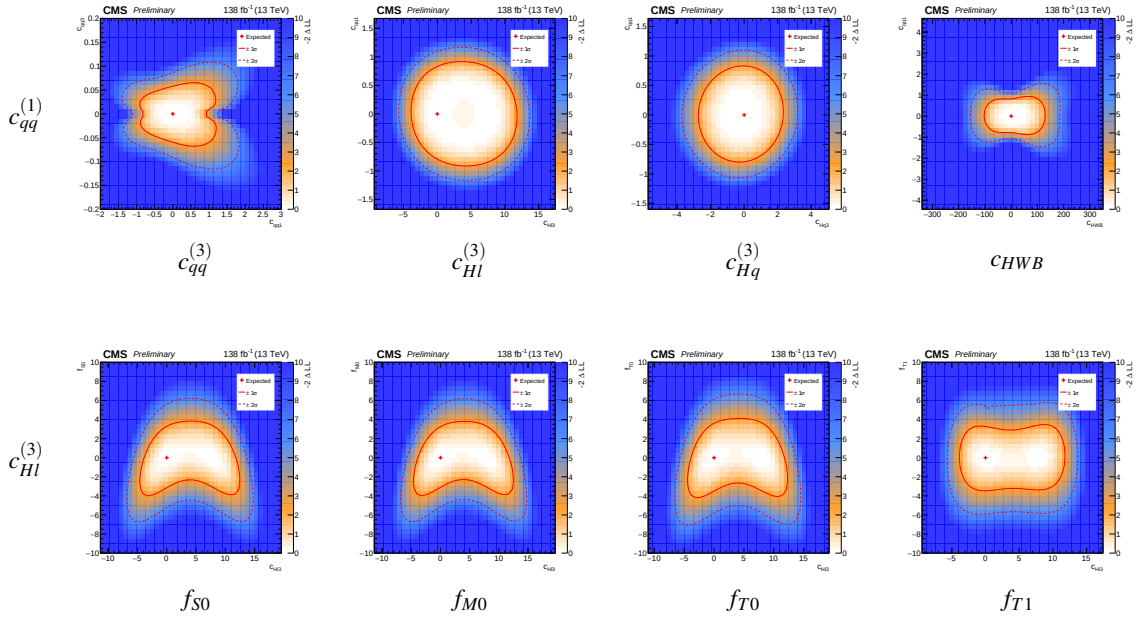


Table C.1: Selection of the two-dimensional 95% and 68% CL contours extracted for the SMEFT operator dim-6/dim-6 pairs (top) and dim-6/dim-8 pairs (bottom).

The preliminary results obtained keeping the signal region blinded to data are promising, with an expected significance of 2.5 standard deviations with respect to the background-only hypothesis for the SM ssWW VBS processes with one hadronic tau in the final state, possibly leading to observed evidence. In Fig. C.2, the postfit distributions for the values of the SM DNN output that are most sensitive to the SM ssWW VBS signal are shown for the $e + \tau_h$ and $\mu + \tau_h$ channels. In addition, expected limits at 1 and 2 standard deviations confidence levels are extracted separately for the Wilson coefficients of the SMEFT operators under investigation. This study represents the first one to extract constraints on the Wilson coefficients related to dim-6 operators investigating VBS processes. In the end, two-dimensional constraints at 1 and 2 standard deviations confidence levels are obtained for all the possible dim-6/dim-6 and dim-6/dim-8 SMEFT operator pairs, examining for the first time the combined effect of two dim-6 SMEFT operators in VBS and, for the first time in an experimental search, the combined effect of a dim-6/dim-8 SMEFT operator pair. A selection of the related results is reported in Tab. C.1. The results of the two-dimensional SMEFT investigations are particularly important to understand the interplay among SMEFT operators with different dimensions acting in the same

process, measuring the impact of a lower-dimension operator effects on the sensitive power to the anomalies induced by higher-dimension operators, and vice-versa.

In conclusion, in this thesis a comprehensive and original search for New Physics phenomena in a model-independent way is devised fully exploiting the LHC potential and the features of the electroweak sector of particle physics SM, paving the way to more general searches for new resonances and the systematic exploration and interpretation of low-energy anomalous couplings.

Appendix $\mathcal{A}$

Dedicated studies for W' boson search

A.1 Kinematic distributions

In the following, several kinematic distributions at the selection level are shown.

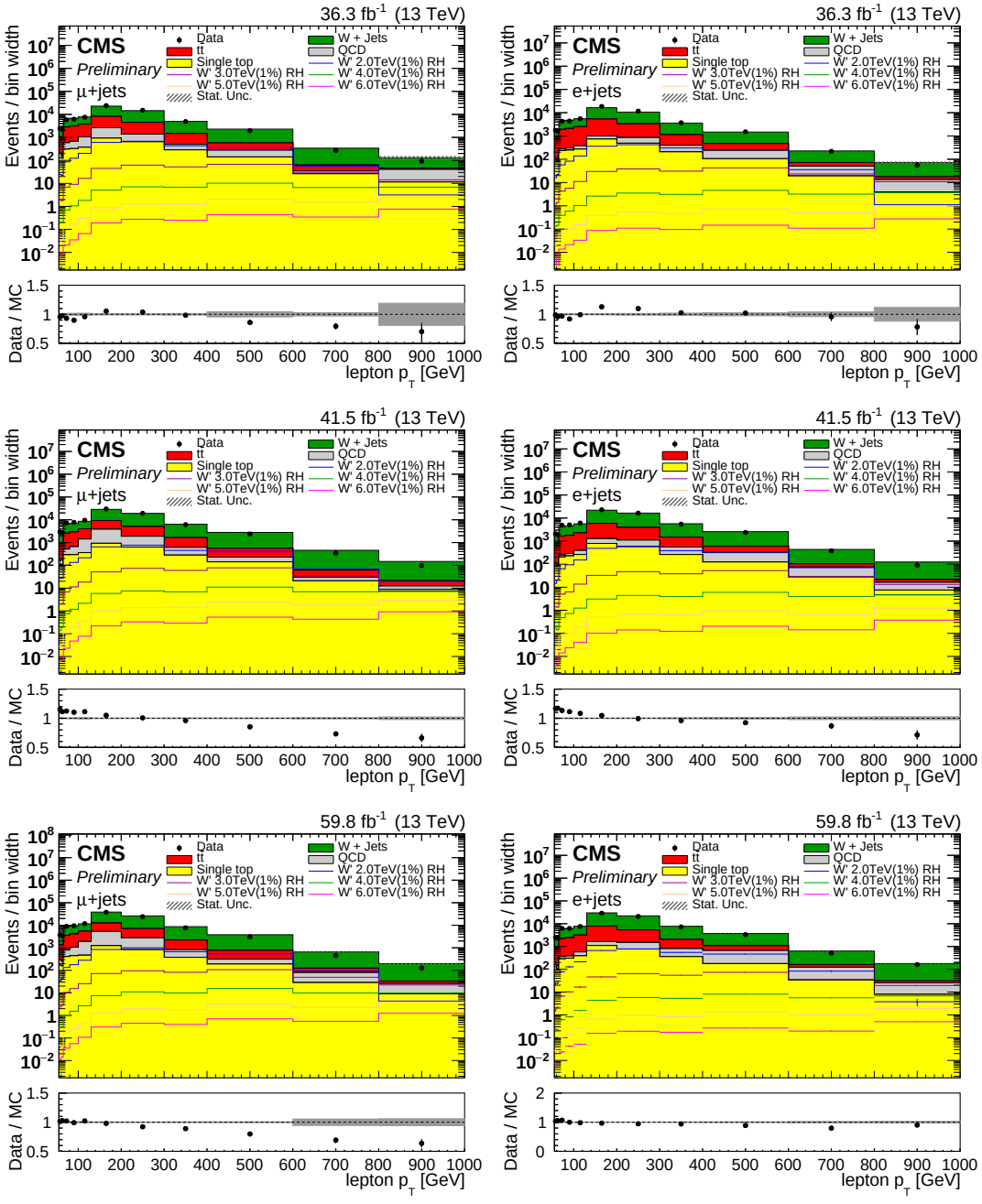


Figure A.1.1: Lepton p_{T} for muons (left) and electrons (right) in 2016 (top), 2017 (mid), and 2018 (bottom).

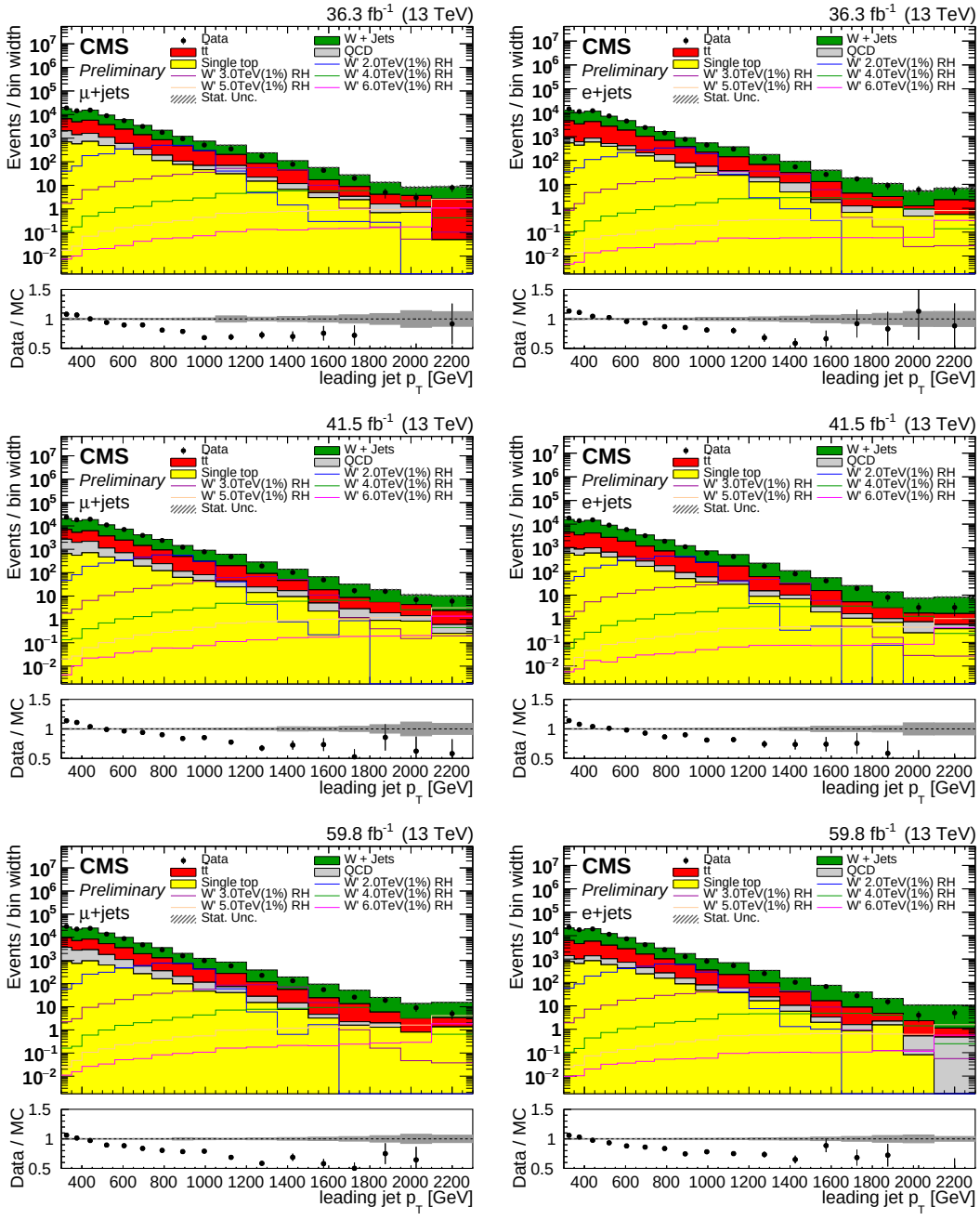


Figure A.1.2: Leading jet p_{T} for muons (left) and electrons (right) in 2016 (top), 2017 (mid), and 2018 (bottom).

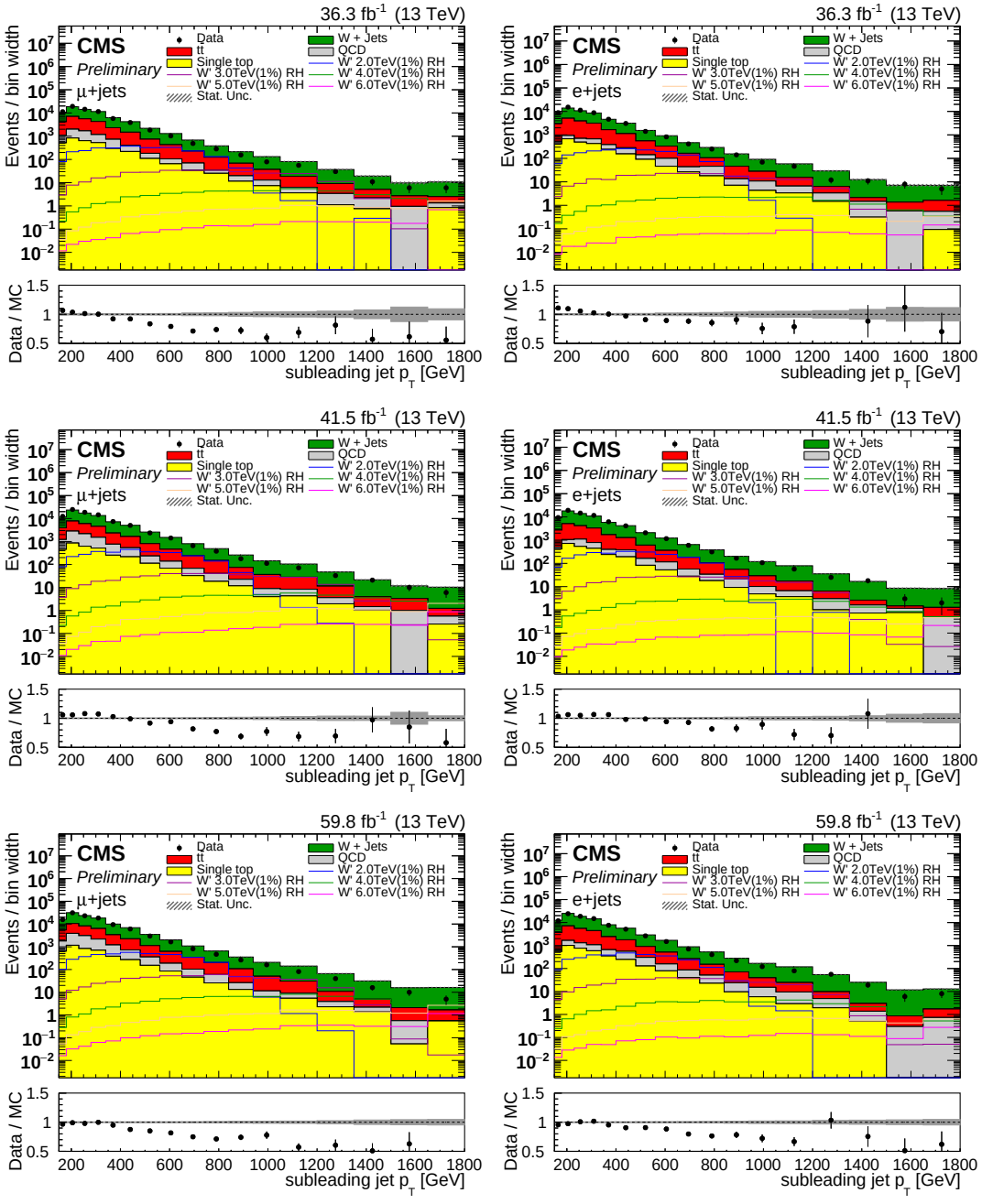


Figure A.1.3: Subleading jet p_{T} for muons (left) and electrons (right) in 2016 (top), 2017 (mid), and 2018 (bottom).

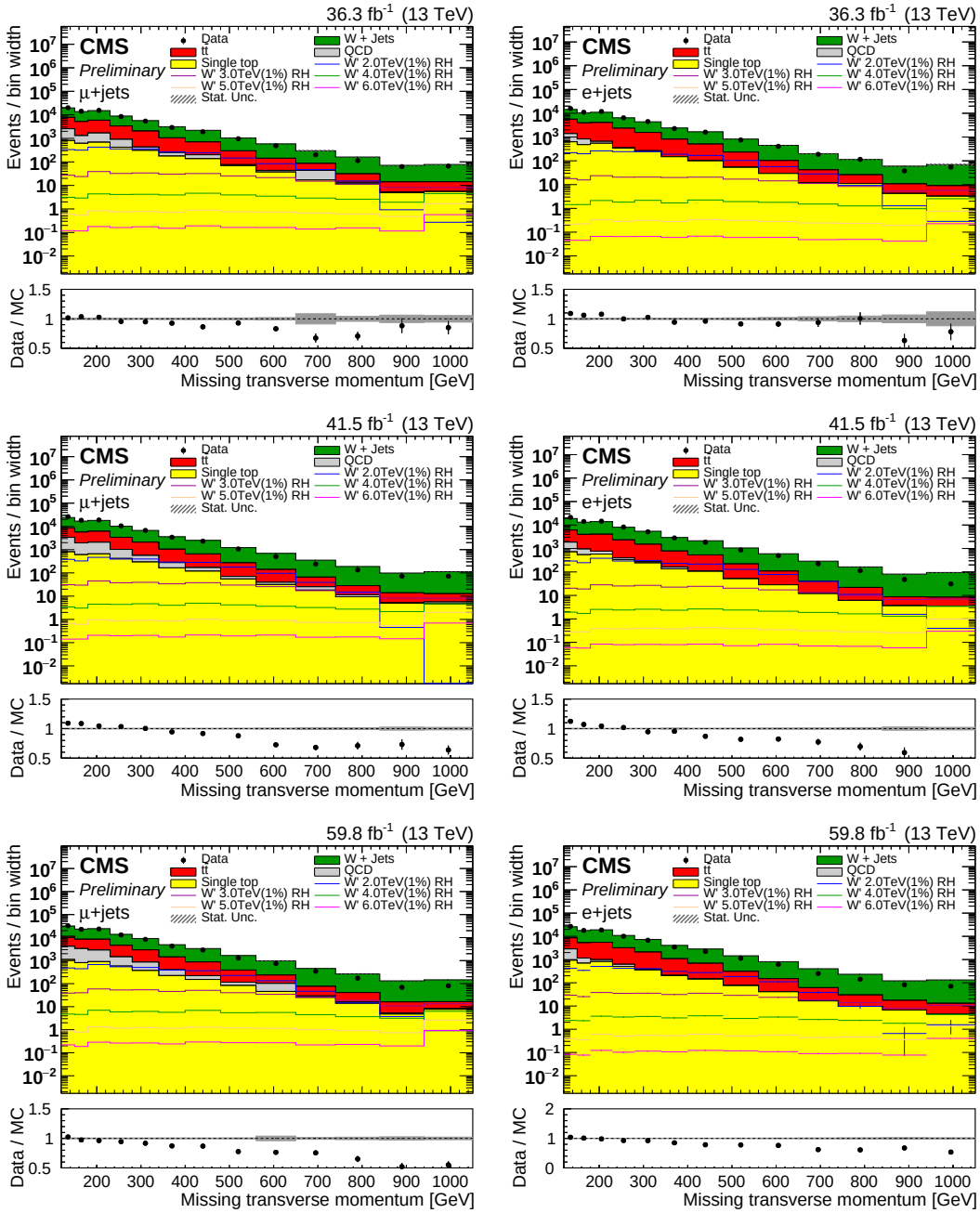


Figure A.1.4: MET for muons (left) and electrons (right) in 2016 (top), 2017 (mid), and 2018 (bottom).

A.2 Evaluation of the data-driven background estimation uncertainties and additional checks

A.2.1 Fit modelling uncertainties

Alternative shapes of the TF and the fit function to the subregions-I are taken as uncertainties by varying the starting point of the fit to the main region over subregion-I ratio in Monte-Carlo for the TF, and to the subregion-I region for the subregion-I fit. The fit is repeated starting from 1.5 TeV and is used to derive only the ratio in the corresponding range. The rest of the distribution is taken from the ratio of the number of events in the two MC histograms directly. This is done in order to reduce the dependency of the fit from the low-mass region, which has the highest statistics, and to separate the low-mass regime from the high-mass regime in the background. This uncertainty is then symmetrized, by considering the full difference with the nominal, but with the opposite sign. Fig. A.2.5 shows the TF fit starting from 1.5 TeV, while Fig. A.2.6 shows the fit to data in the subregion-I region starting from 1.5 TeV.

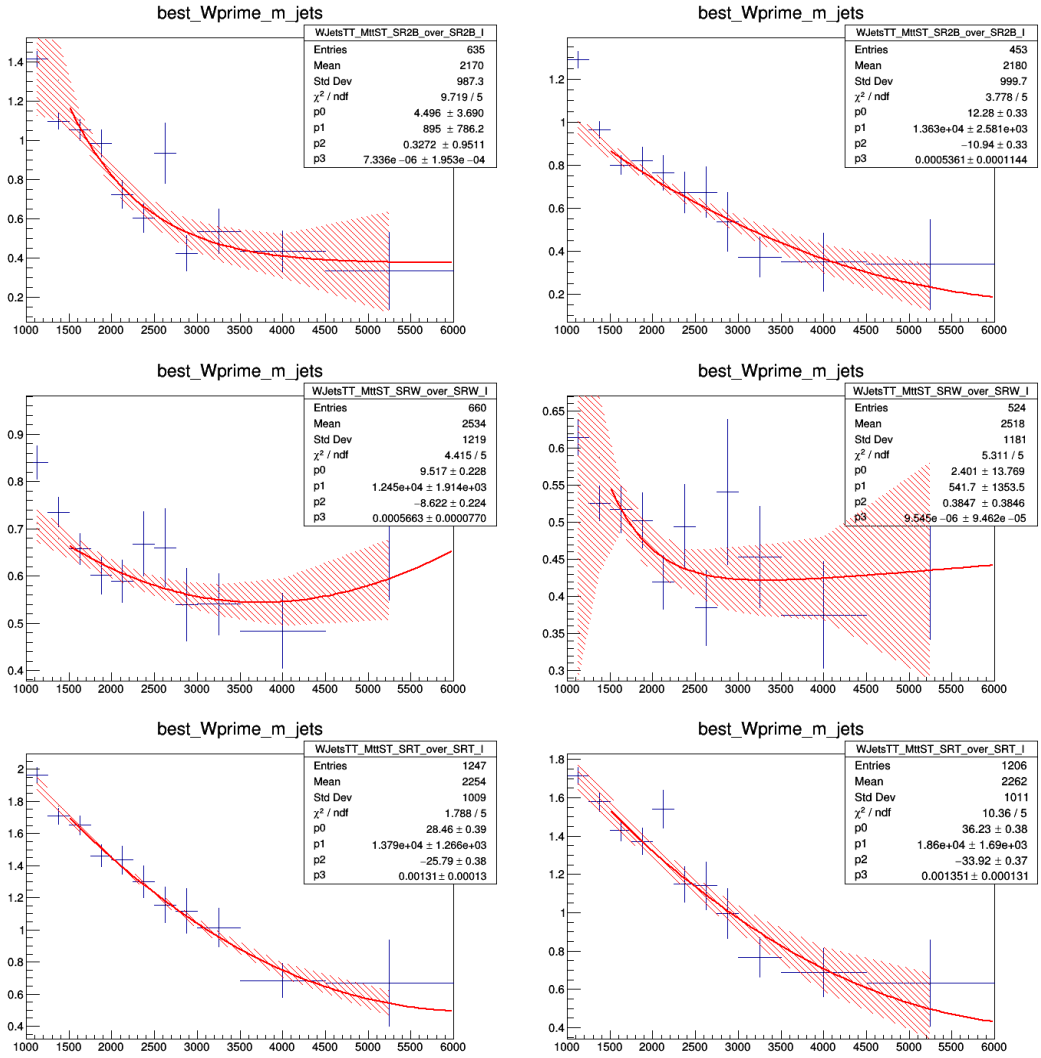


Figure A.2.5: Ratio of the shapes between each category and its corresponding region I, the analytic form for the function fitted and the fit results for muon (left) and electron (right) for SR2B (top), SRW (middle), and SRT (bottom). On the $x-$axis, reconstructed mass of the W' boson in GeV.

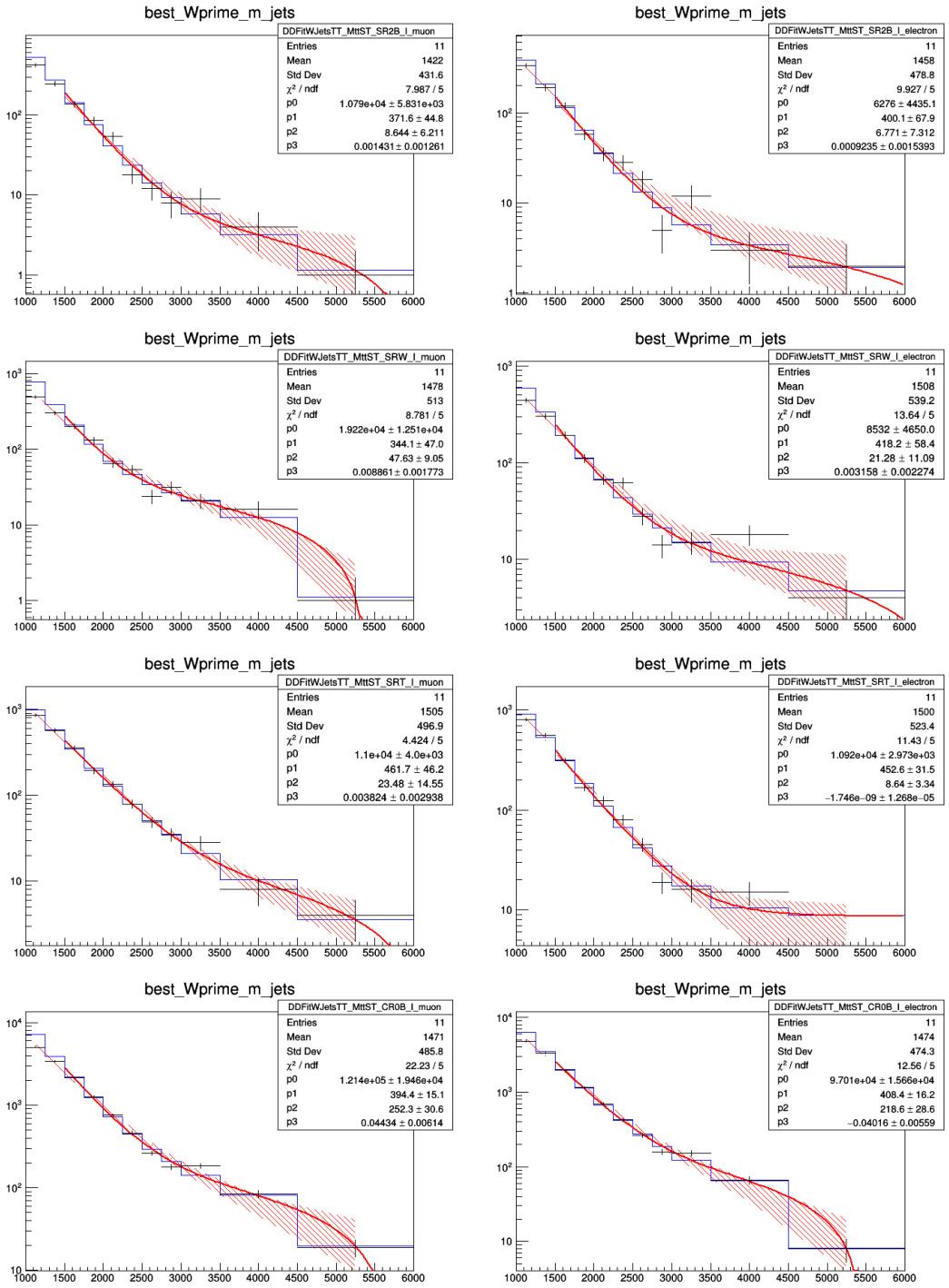


Figure A.2.6: Fit to data fit in the region I, for SR2B (first row), SRW (second row), SRT (third row), and CR0B (fourth row) for muons (left), and electrons (right). The black histogram is data, the blue histogram is the integral of the red line. On the $x-$axis, reconstructed mass of the W' boson in GeV.

A.2.2 Background composition uncertainties

Predicted rates and shapes have a dependence on the relative fraction of backgrounds in the subregions-I and in the respective signal region.

Starting from Eq. 6.9 and Eq. 6.10, the dependence of the bias on the estimated number of

background events can be derived as a function of the potential differences between data and simulation in terms of background composition. In the following, a derivation of the effect of a potential mismodeling of a certain component of the background is formalized.

For each background, indicated with a generic x (that can be $t\bar{t}$, W+jets, or single-top) the following quantities are defined:

- n_x^{MC}: number of MC event for the component x in a main region;
- m_x^{MC}: number of MC event for the component x in a subregion-I;
- n_x^{Data}: number of event in data for the component x in a main region (assumed to be known);
- m_x^{Data}: number of event in data for the component x in a subregion-I (assumed to be known);
- $n_{non\,QCD}^{MC}$: number of MC events for all the background, except for QCD, in an main region;
- $m_{non\,QCD}^{MC}$: number of MC events for all the background, except for QCD, in an subregion-I;
- $n_{non\,QCD}^{Data}$: number of events in data for all the background, subtracting QCD MC, in an main region (unknown/blinded);
- $m_{non\,QCD}^{Data}$: number of events in data for all the background, subtracting for QCD MC, in an subregion-I (known);
- a_x: fraction of events in the main region, i.e. $\frac{n_x^{MC}}{n_{non\,QCD}^{MC}}$;
- b_x: fraction of events in the subregion-I, i.e. $\frac{m_x^{MC}}{m_{non\,QCD}^{MC}}$;
- γ_x: ratio between amount of background x in data and background x in MC, i.e. n_x^{Data}/n_x^{MC};
- Δ_x = relative excess of x in data, i.e. $\frac{n_x^{Data}-n_x^{MC}}{n_x^{MC}} = \gamma_x - 1$.

Assuming one by one the backgrounds are mismodeled by a factor γ_x, the total number of data events can be written as:

$$\begin{aligned} m_{non\,QCD}^{Data} &= m_x^{MC}\gamma_x + (m_{non\,QCD}^{MC} - m_x^{Data}) = \\ &= m_{non\,QCD}^{MC}(b_x\gamma_x + (1-b_x)) = m_{non\,QCD}^{MC}(1+b_x\Delta_x) \\ n_{non\,QCD}^{Data} &= n_x^{Data}\gamma_x + (n_{non\,QCD}^{MC} - n_x^{Data}) = \\ &= n_{non\,QCD}^{MC}(a_x\gamma_x + (1-a_x)) = n_{non\,QCD}^{MC}(1+a_x\Delta_x) \end{aligned} \tag{A.1}$$

The prediction $n_{non\,QCD}^{Pred}$ for $n_{non\,QCD}^{Data}$ is obtained as:

$$\begin{aligned} n_{non\,QCD}^{Pred} &= m_{non\,QCD}^{Data} \cdot \frac{n_{non\,QCD}^{MC}}{m_{non\,QCD}^{MC}} = \\ &= m_{non\,QCD}^{MC} \cdot (b_x \cdot \gamma_x + (1-b_x)) \cdot \frac{n_{non\,QCD}^{MC}}{m_{non\,QCD}^{MC}} = \\ &= n_{non\,QCD}^{MC} \cdot (1+b_x\Delta_x) \end{aligned} \tag{A.2}$$

Now multiplying and dividing by $1+a_x\Delta_x$

$$n^{MC}_{non\,QCD}\cdot(1+a_x\Delta_x)\cdot\frac{(1+b_x\Delta_x)}{(1+a_x\Delta_x)}=n^{Data}_{non\,QCD}\cdot\frac{1+b_x\Delta_x}{1+a_x\Delta_x} \tag{A.3}$$

Therefore the bias on the prediction in the signal region, parameterized the Δ_x assuming all misprediction originates from component x, (while simultaneously $\Delta_y=\Delta_z=0$), is given by the quantity:

$$R_x=\frac{1+a_x\Delta_x}{1+b_x\Delta_x} \tag{A.4}$$

Fig. A.2.7 shows the size of the bias as a function of Δ_x for various channels in the SR2B for muons and electrons. The range of R_x is $]-1,5]$, since -1 corresponds to 0 events of that particular background x in data. A value of $\Delta_x=5$ means data has 6 times, corresponding to $(1+\Delta_x)$, more background x than predicted in MC.

The first that is done is to control whether this range of variation is compatible with data, e.g. in the SR2B-I. Considering the actual number of events in the subregion-I, one can estimate the value of Δ_x as:

$$\Delta_x=(m^{Data}_{non\,QCD}+QCD)-(m^{MC}_{non\,QCD}+QCD) \tag{A.5}$$

i.e. assigning the entire difference between data and simulation to the background x.

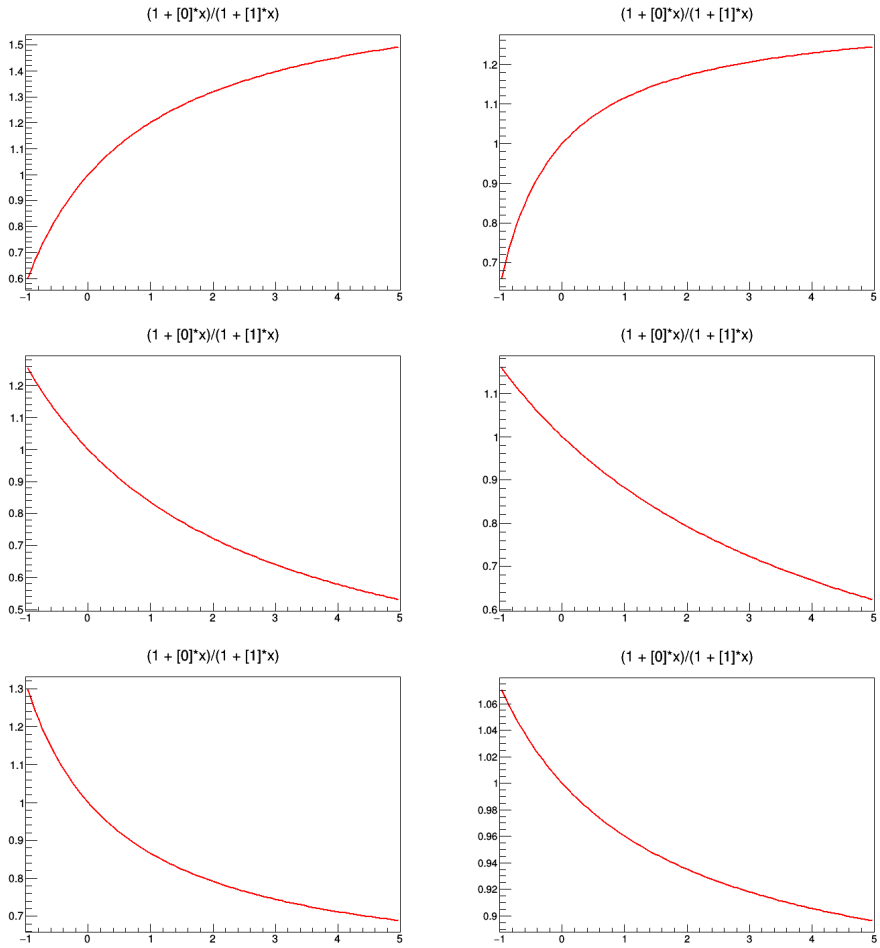


Figure A.2.7: Estimated bias on the prediction in the SR2B region as a function of the difference between data and simulation for muons (left) and electrons (right), and $t\bar{t}$ (top), W+jets (middle), single top quark (bottom). On the $x-$axis, Δ_x.

A.2.3 Effect of systematic uncertainties

A systematic uncertainty is added to take into account the potential bias of the different background compositions between the numerator and the denominator on the TF. This uncertainty can therefore effectively fit the Δ_x, fixing the bias. If we call σ_x the fraction of events by which the rate of the background x might be mismodeled, we have a modified total number of events and modified transfer function TF':

$$\begin{aligned}
n^{Data}_{non\,QCD} &= n^{MC}_{x}\gamma_x + (n^{MC}_{non\,QCD} - n_x) = n^{MC}_{non\,QCD}(a_x\gamma_x + (1-a_x)) = n^{MC}_{non\,QCD}(1+a_x\sigma_x)\\
m^{Data}_{non\,QCD} &= m^{MC}_{x}\gamma_x + (m^{MC}_{non\,QCD} - m_x) = m^{MC}_{non\,QCD}(b_x\gamma_x + (1-b_x)) = m^{MC}_{non\,QCD}(1+b_x\sigma_x)\\
TF' &= \frac{n^{Data}_{non\,QCD}}{m^{Data}_{non\,QCD}} = \frac{n^{MC}_{non\,QCD}(1+a_x\sigma_x)}{m^{MC}_{non\,QCD}(1+b_x\sigma_x)} = \frac{(1+a_x\sigma_x)}{(1+b_x\sigma_x)}TF\\
n^{Pred}_{non\,QCD} &= m^{Data}_{non\,QCD}TF'
\end{aligned} \tag{A.6}$$

The goal now is to rewrite the relation in Eq. A.6 by expliciting $n^{Data}_{non\,QCD}$ as a function of $n^{MC}_{non\,QCD}, \Delta_x, \sigma_x$. Performing the substitution:

$$m^{Data}_{non\,QCD} = m^{MC}_{non\,QCD}(1+b_x\Delta_x) \tag{A.7}$$

and then multiplying and dividing by $1+a_x\Delta_x$:

$$\begin{aligned}
n^{Pred}_{non\,QCD} &= m^{MC}_{non\,QCD}\frac{(1+b_x\Delta_x)}{(1+a_x\Delta_x)}(1+a_x\Delta_x)TF' =\\
&= m^{MC}_{non\,QCD}\frac{n^{MC}_{non\,QCD}}{m^{MC}_{non\,QCD}}(1+a_x\Delta_x)\frac{1+b_x\Delta_x}{1+a_x\Delta_x}\frac{1+a_x\sigma_x}{1+b_x\sigma_x}TF
\end{aligned} \tag{A.8}$$

Finally:

$$n^{Pred}_{non\,QCD} = n^{MC}_{non\,QCD}(1+a\Delta_x)\frac{1+b_x\Delta_x}{1+a_x\Delta_x}\frac{1+a_x\sigma_x}{1+b_x\sigma_x}TF = n^{Data}_{non\,QCD}\frac{1+b_x\Delta_x}{1+a_x\Delta_x}\frac{1+a_x\sigma_x}{1+b_x\sigma_x}TF \tag{A.9}$$

Therefore a formula that is similar to Eq. A.4 is obtained for the R, but with a mitigating factor:

$$R = \frac{1+a_x\Delta_x}{1+b_x\Delta_x}\frac{1+b_x\sigma_x}{1+a_x\sigma_x} \tag{A.10}$$

If $\sigma_x = \Delta_x$ the bias is removed.

As a consequence, if a prefit value for σ_x^{prefit} larger than Δ_x is chosen, the fit will be able to adapt to the values obtained in the data. If the σ_x^{prefit} is smaller than Δ_x, the postfit value will not be able to fully remove the bias. This means that systematic uncertainties quoted above for the different backgrounds can effectively fully compensate the bias safely in a certain range of Δ_x, and partially compensate the bias in a wider range. To assess the size of the compensation, the formula in Eq. A.10 is used for different values of σ_x. This effect is shown in 3 steps.

First, the overall effect of a reduction over the range of Δ_x is displayed. In Fig. A.2.8 what happens when σ_x is a particular fraction of Δ_x is shown. Each curve represents a different hypothesis, for example, the 0.5 curve is a case where $\sigma_x = 0.5\Delta_x$ for each value of Δ_x.

Hence, additional uncertainty sources are considered, one per each background, by varying each by a factor large enough to mitigate any potential discrepancy that is seen in the corresponding subregion-I. From Tab. A.2.1, A.2.2, one can see that the total excess in the W+jets-dominated regions (SRW-I, CR0B, and CR0B-I) is below 20%, and in the $t\bar{t}$ region (SRT-I and SR2B-I) it is below 30%. To check the potential discrepancies differentially in $m_{W'}$, the same

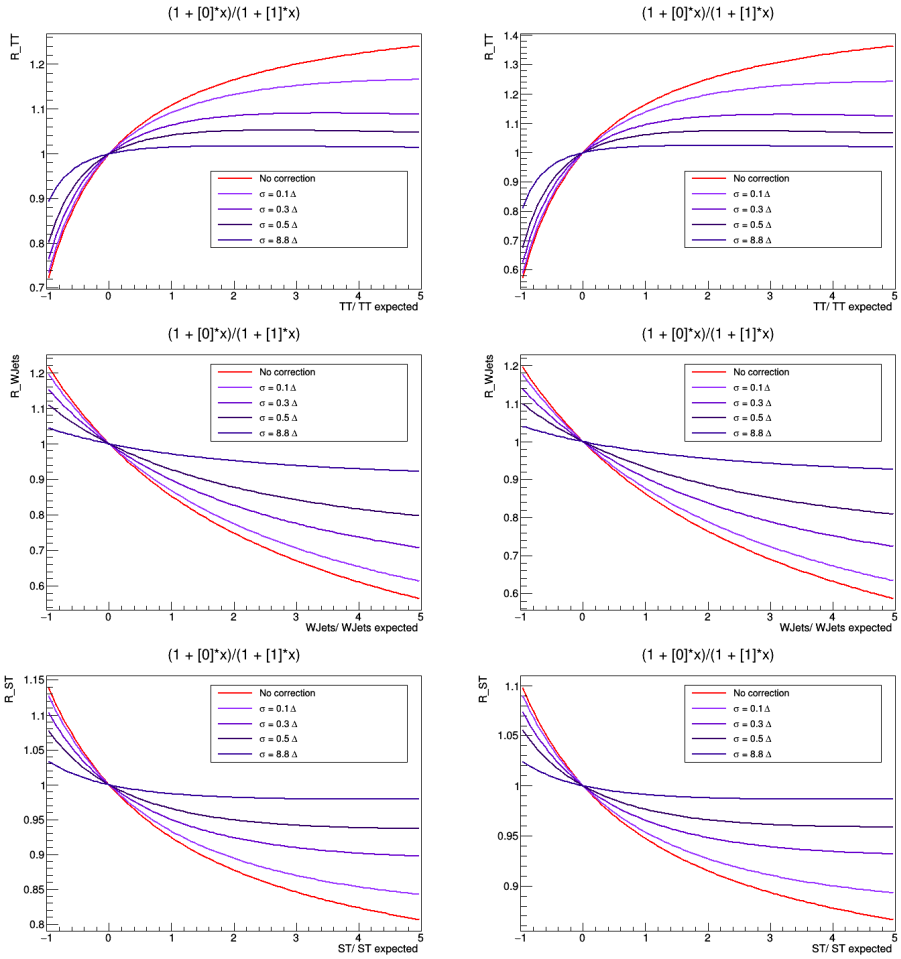


Figure A.2.8: Estimated bias on the prediction in the SR2B region as a function of the difference between data and simulation for muons (left) and electrons (right), and $t\bar{t}$ (top), W+jets (middle), single top quark (bottom). On the $x-$axis, Δ_x.

quantities are shown divided in three $m_{W'}$ ranges in Tab. A.2.3, A.2.4. One can see here that the maximum discrepancy reached is about 150% in the higher-mass, low statistics bins.

Fig. A.2.9 shows the behavior of the bias without any value of σ and with a σ able to cover a given fraction of the Δ_x. One can see that even covering half the bias can usually result in a bias below 10%. It takes a_x and b_x from the SR2B of Tab. A.2.1, A.2.2, and a σ_x that is a fraction of Δ_x.

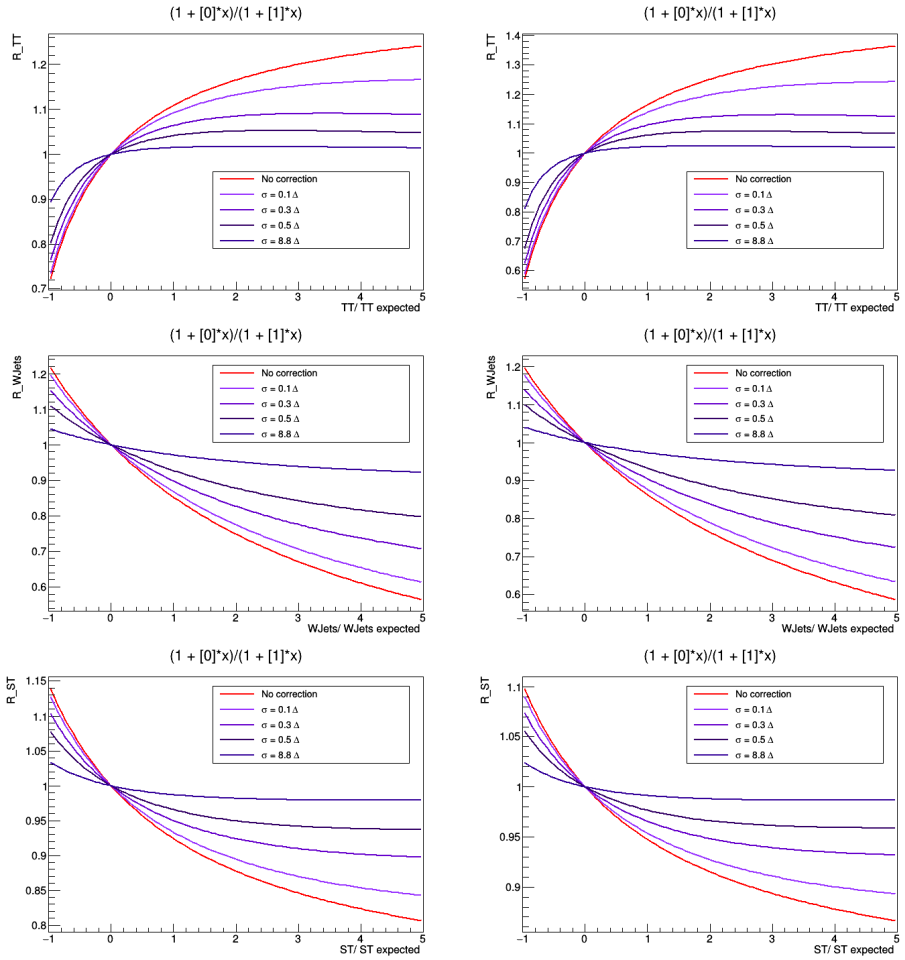


Figure A.2.9: Estimated bias on the prediction as a function of the difference between data and simulation in the SR2B for muons (left) and electrons (right), and $t\bar{t}$ (top), W+jets (middle), single top quark (bottom). Δ_x is taken form the SR2B in Tables A.2.1, A.2.2. The lines correspond to no σ_x (red), a and σ_x of 10%, 30%, 50%, and 80% of the Data-MC discrepancy. On the $x-$axis, Δ_x

In order to make sure that the maximum possible difference is covered, three separate uncertainties of $\pm 80\%$ of the backgrounds are assigned, which can fully cover a $t\bar{t}$, W+jets, or single-top that is up to 160% of the SM expectation, and to reduce to as low as 10% an eventual bias due to a signal being up to 260% of its SM expectation. Such a number can be seen as the value of the $\sigma_x = 0.5\Delta_x$ curve for $\Delta_x = 2.6$. The effect of the uncertainty in the SR2B can be observed in Fig. A.2.10 for the three backgrounds, where the central value of the bias is shown in red, the dotted red line represents the value of the bias obtained by fixing $\sigma_{prefit} = \pm 0.8$, i.e. the envelope of variations that can be covered by the systematic variation used. Finally, in blue it is shown the expected postfit value when considering $\sigma(\Delta)$ fit from data, assuming that the maximum is given by the envelope, i.e. $\sigma \leq \sigma_{prefit}$.

In order to make sure that the maximum possible difference is covered, three separate un-

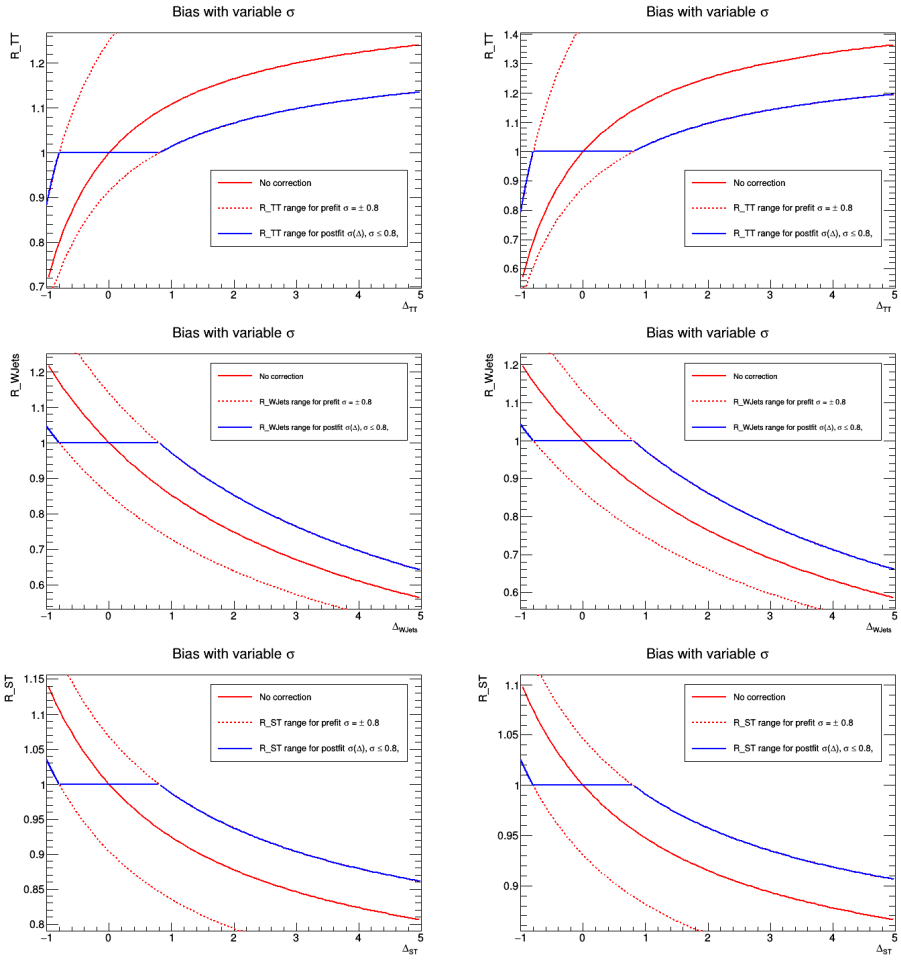


Figure A.2.10: Estimated bias on the prediction taking into account the prefit (dotted red) and postfit (blue) effect of the uncertainty on the single background, plotted as a function of the difference between data and simulation in the SR2B for muons (left) and electrons (right), and $t\bar{t}$ (top), W+jets (middle), single top quark (bottom). Δ_x is taken from the SR2B in Tables A.2.1, A.2.2.

certainties of $\pm 80\%$ of the backgrounds are assigned, which can fully cover a $t\bar{t}$, W+jets, or single-top that is up to 160% of the SM expectation, and to reduce to as low as 10% an eventual bias due to a signal being up to 260% of its SM expectation. To disentangle the rate and shape effect of the composition, the total background yield for each region is left free to float without any constraint. In addition, to evaluate any further effect in data, an additional correction is derived when comparing between data and simulation in the subregions II and III.

Table A.2.1: Number of background events from the MC simulations and their realtive abundance in the main region and in the region I. In Region I also the ratio between the number of events for ech background N divided by the difference between Data and total MC prediction ($|Data - TotMC|$) is given. This is meant to indicate how large the excess or defect would be if attributed to the single process in question.

Lepton: electron	main region		control region		
Process	Events $\pm$ Unc.	Abundance(%)	Events. $\pm$ Unc.	Abundance(%)	$\frac{N}{\lvert Data-TotMC\rvert}$
	CR0B		CR0B_I		
Tot MC	5990 ± 40	100	12959 ± 51	100	
Data	blinded		12954 ± 114	100	
$t\bar{t}$	914 ± 19	15	1084 ± 20	8	0.06
W+jets	4751 ± 27	79	11487 ± 42	89	0.01
ST	153 ± 8	3	147 ± 7	1	0.47
QCD	124 ± 21	2	112 ± 19	1	0.62
	SRW		SRW_I		
Tot MC	619 ± 17	100	1143 ± 21	100	
Data	blinded		1234 ± 35	108	
$t\bar{t}$	151 ± 6	24	148 ± 6	13	0.77
W+jets	364 ± 8	59	800 ± 11	70	0.14
ST	68 ± 4	11	134 ± 6	12	0.84
QCD	27 ± 13	4	47 ± 16	4	2.43
	SRT		SRT_I		
Tot MC	2962 ± 38	100	1902 ± 24	100	
Data	blinded		2136 ± 46	112	
$t\bar{t}$	2185 ± 26	74	1029 ± 18	54	0.24
W+jets	269 ± 6	9	646 ± 10	34	0.39
ST	423 ± 12	14	193 ± 7	10	1.29
QCD	73 ± 24	2	18 ± 11	1	13.96
	SR2B		SR2B_I		
Tot MC	936 ± 14	100	939 ± 14	100	
Data	blinded		789 ± 28	84	
$t\bar{t}$	660 ± 11	71	435 ± 7	46	0.33
W+jets	30 ± 2	3	184 ± 5	20	0.79
ST	235 ± 7	25	300 ± 9	32	0.48
QCD	8 ± 5	1	13 ± 7	1	11.21

Table A.2.2: Number of background events from the MC simulations and their realtive abundance in the main region and in the region I. In Region I also the ratio between the number of events for ech background N divided by the difference between Data and total MC prediction ($|Data - TotMC|$) is given. This is meant to indicate how large the excess or defect would be if attributed to the single process in question.

Lepton: muon	main region		control region		
Process	Events $\pm$ Unc.	Abundance(%)	Events. $\pm$ Unc.	Abundance(%)	$\frac{N}{\lvert Data-TotMC\rvert}$
	CR0B		CR0B_I		
Tot MC	9215 $\pm$ 81	100	14281 $\pm$ 58	100	
Data	blinded		13709 $\pm$ 118	96	
$t\bar{t}$	1266 $\pm$ 23	14	1127 $\pm$ 21	8	0.42
W+jets	7145 $\pm$ 33	78	12715 $\pm$ 44	89	0.04
ST	182 $\pm$ 8	2	163 $\pm$ 8	1	2.88
QCD	530 $\pm$ 69	6	118 $\pm$ 31	1	3.98
	SRW		SRW_I		
Tot MC	891 $\pm$ 20	100	1183 $\pm$ 16	100	
Data	blinded		1317 $\pm$ 37	111	
$t\bar{t}$	201 $\pm$ 7	23	156 $\pm$ 7	13	0.96
W+jets	537 $\pm$ 13	60	868 $\pm$ 12	73	0.17
ST	98 $\pm$ 5	11	129 $\pm$ 6	11	1.17
QCD	42 $\pm$ 13	5	8 $\pm$ 5	1	18.35
	SRT		SRT_I		
Tot MC	4390 $\pm$ 82	100	2147 $\pm$ 25	100	
Data	blinded		2300 $\pm$ 48	107	
$t\bar{t}$	2717 $\pm$ 29	62	1188 $\pm$ 20	55	0.14
W+jets	358 $\pm$ 8	8	703 $\pm$ 10	33	0.23
ST	498 $\pm$ 12	11	206 $\pm$ 8	10	0.80
QCD	797 $\pm$ 76	18	22 $\pm$ 7	1	7.34
	SR2B		SR2B_I		
Tot MC	1391 $\pm$ 55	100	1071 $\pm$ 14	100	
Data	blinded		992 $\pm$ 32	93	
$t\bar{t}$	857 $\pm$ 13	62	490 $\pm$ 8	46	0.15
W+jets	41 $\pm$ 3	3	222 $\pm$ 6	21	0.33
ST	298 $\pm$ 9	21	336 $\pm$ 9	31	0.22
QCD	190 $\pm$ 53	14	13 $\pm$ 4	1	5.69

Table A.2.3: Number of background events from the MC simulations and their realtive abundance in the main region and in the region I divided in bins. In Region I also the ratio between the number of events for ech background N divided by the difference between Data and total MC prediction ($|Data - TotMC|$) is given. This is meant to indicate how large the excess or defect would be if attributed to the single process in question.

Lepton: electron	main region		control region		
Process	Events $\pm$ Unc.	Abundance(%)	Events $\pm$ Unc.	Abundance(%)	$\frac{\lvert N\rvert}{\lvert Data-TotMC\rvert}$
	SRW		SRW_I		
Tot MC, bin [1000,1750]	473 $\pm$ 16	100	825 $\pm$ 19	100	
Tot MC, bin [1750,2500]	107 $\pm$ 5	100	232 $\pm$ 9	100	
Tot MC, bin $\geq$2500	40 $\pm$ 3	100	86 $\pm$ 4	100	
$t\bar{t}$, bin [1000,1750]	129 $\pm$ 6	27	125 $\pm$ 6	15	0.89$\pm$ 0.29
$t\bar{t}$, bin [1750,2500]	19 $\pm$ 2	18	20 $\pm$ 2	9	0.42$\pm$ 0.88
$t\bar{t}$, bin $\geq$2500	4 $\pm$ 1	10	3 $\pm$ 1	4	2.22$\pm$ 3.08
W+jets, bin [1000,1750]	262 $\pm$ 6	55	558 $\pm$ 10	68	0.20$\pm$ 0.06
W+jets, bin [1750,2500]	76 $\pm$ 3	71	179 $\pm$ 5	77	0.05$\pm$ 0.09
W+jets, bin $\geq$2500	32 $\pm$ 2	80	76 $\pm$ 3	89	0.09$\pm$ 0.12
ST, bin [1000,1750]	57 $\pm$ 4	12	109 $\pm$ 6	13	1.02$\pm$ 0.33
ST, bin [1750,2500]	10 $\pm$ 2	9	20 $\pm$ 2	9	0.42$\pm$ 0.88
ST, bin $\geq$2500	1 $\pm$ 0	3	5 $\pm$ 1	6	1.29$\pm$ 1.78
QCD, bin [1000,1750]	25 $\pm$ 13	5	33 $\pm$ 14	4	3.35$\pm$ 1.74
QCD, bin [1750,2500]	2 $\pm$ 1	2	12 $\pm$ 6	5	0.71$\pm$ 1.44
QCD, bin $\geq$2500	3 $\pm$ 2	7	1 $\pm$ 1	2	5.25$\pm$ 7.67
	SRT		SRT_I		
Tot MC, bin [1000,1750]	2384 $\pm$ 35	100	1458 $\pm$ 22	100	
Tot MC, bin [1750,2500]	481 $\pm$ 14	100	337 $\pm$ 9	100	
Tot MC, bin $\geq$2500	98 $\pm$ 5	100	106 $\pm$ 5	100	
$t\bar{t}$, bin [1000,1750]	1788 $\pm$ 23	75	830 $\pm$ 16	57	0.27$\pm$ 0.05
$t\bar{t}$, bin [1750,2500]	340 $\pm$ 10	71	162 $\pm$ 7	48	0.22$\pm$ 0.12
$t\bar{t}$, bin $\geq$2500	63 $\pm$ 4	64	42 $\pm$ 3	39	0.27$\pm$ 0.24
W+jets, bin [1000,1750]	201 $\pm$ 6	8	462 $\pm$ 9	32	0.49$\pm$ 0.10
W+jets, bin [1750,2500]	54 $\pm$ 3	11	138 $\pm$ 4	41	0.26$\pm$ 0.15
W+jets, bin $\geq$2500	18 $\pm$ 1	18	56 $\pm$ 2	52	0.20$\pm$ 0.19
ST, bin [1000,1750]	341 $\pm$ 10	14	152 $\pm$ 7	10	1.48$\pm$ 0.31
ST, bin [1750,2500]	68 $\pm$ 4	14	36 $\pm$ 3	11	1.00$\pm$ 0.59
ST, bin $\geq$2500	16 $\pm$ 2	16	5 $\pm$ 1	5	2.22$\pm$ 2.13
QCD, bin [1000,1750]	54 $\pm$ 23	2	14 $\pm$ 10	1	16.16$\pm$ 12.61
QCD, bin [1750,2500]	18 $\pm$ 8	4	1 $\pm$ 1	0	31.86$\pm$ 27.42
QCD, bin $\geq$2500	1 $\pm$ 1	1	4 $\pm$ 1	3	3.06$\pm$ 3.12
	SR2B		SR2B_I		
Tot MC, bin [1000,1750]	784 $\pm$ 13	100	727 $\pm$ 13	100	
Tot MC, bin [1750,2500]	128 $\pm$ 6	100	166 $\pm$ 6	100	
Tot MC, bin $\geq$2500	24 $\pm$ 2	100	46 $\pm$ 3	100	
$t\bar{t}$, bin [1000,1750]	552 $\pm$ 10	70	365 $\pm$ 6	50	0.26$\pm$ 0.08
$t\bar{t}$, bin [1750,2500]	93 $\pm$ 5	73	59 $\pm$ 3	35	0.77$\pm$ 0.21
$t\bar{t}$, bin $\geq$2500	16 $\pm$ 2	67	12 $\pm$ 1	26	0.54$\pm$ 0.57
W+jets, bin [1000,1750]	21 $\pm$ 2	3	131 $\pm$ 4	18	0.71$\pm$ 0.21
W+jets, bin [1750,2500]	7 $\pm$ 1	5	40 $\pm$ 2	24	1.14$\pm$ 0.31
W+jets, bin $\geq$2500	3 $\pm$ 1	13	16 $\pm$ 1	35	0.40$\pm$ 0.42
ST, bin [1000,1750]	203 $\pm$ 7	26	219 $\pm$ 8	30	0.43$\pm$ 0.12
ST, bin [1750,2500]	27 $\pm$ 3	21	67 $\pm$ 4	40	0.68$\pm$ 0.18
ST, bin $\geq$2500	4 $\pm$ 1	17	17 $\pm$ 2	37	0.38$\pm$ 0.39
QCD, bin [1000,1750]	8 $\pm$ 5	1	11 $\pm$ 7	1	8.89$\pm$ 6.27
QCD, bin [1750,2500]	0 $\pm$ 0	0	1 $\pm$ 1	1	36.09$\pm$ 21.57
QCD, bin $\geq$2500	1 $\pm$ 1	3	1 $\pm$ 1	3	4.52$\pm$ 5.32

Table A.2.4: Number of background events from the MC simulations and their realtive abundance in the main region and in the region I divided in bins. In Region I also the ratio between the number of events for ech background N divided by the difference between Data and total MC prediction ($|Data - TotMC|$) is given. This is meant to indicate how large the excess or defect would be if attributed to the single process in question.

Lepton: muon	main region		control region		
Process	Events $\pm$ Unc.	Abundance(%)	Events $\pm$ Unc.	Abundance(%)	$\frac{\lvert N\rvert}{\lvert Data-TotMC\rvert}$
	SRW		SRW_I		
Tot MC, bin [1000,1750]	678 $\pm$ 18	100	850 $\pm$ 14	100	
Tot MC, bin [1750,2500]	157 $\pm$ 8	100	239 $\pm$ 7	100	
Tot MC, bin $\geq$2500	57 $\pm$ 3	100	94 $\pm$ 3	100	
$t\bar{t}$, bin [1000,1750]	172 $\pm$ 6	25	134 $\pm$ 6	16	1.06$\pm$ 0.26
$t\bar{t}$, bin [1750,2500]	24 $\pm$ 3	15	18 $\pm$ 2	8	0.53$\pm$ 0.94
$t\bar{t}$, bin $\geq$2500	6 $\pm$ 1	11	5 $\pm$ 1	5	0.19$\pm$ 2.13
W+jets, bin [1000,1750]	397 $\pm$ 12	59	604 $\pm$ 10	71	0.24$\pm$ 0.05
W+jets, bin [1750,2500]	107 $\pm$ 4	69	198 $\pm$ 6	83	0.05$\pm$ 0.08
W+jets, bin $\geq$2500	43 $\pm$ 2	75	86 $\pm$ 3	92	0.01$\pm$ 0.11
ST, bin [1000,1750]	80 $\pm$ 5	12	107 $\pm$ 5	13	1.32$\pm$ 0.32
ST, bin [1750,2500]	14 $\pm$ 2	9	21 $\pm$ 2	9	0.47$\pm$ 0.82
ST, bin $\geq$2500	5 $\pm$ 1	8	2 $\pm$ 1	2	0.45$\pm$ 5.01
QCD, bin [1000,1750]	28 $\pm$ 12	4	5 $\pm$ 5	1	27.18$\pm$ 24.47
QCD, bin [1750,2500]	12 $\pm$ 6	8	2 $\pm$ 1	1	3.87$\pm$ 7.03
QCD, bin $\geq$2500	3 $\pm$ 1	6	1 $\pm$ 0	1	1.27$\pm$ 14.34
	SRT		SRT_I		
Tot MC, bin [1000,1750]	3504 $\pm$ 72	100	1606 $\pm$ 22	100	
Tot MC, bin [1750,2500]	719 $\pm$ 38	100	400 $\pm$ 10	100	
Tot MC, bin $\geq$2500	167 $\pm$ 10	100	141 $\pm$ 6	100	
$t\bar{t}$, bin [1000,1750]	2213 $\pm$ 26	63	932 $\pm$ 17	58	0.19$\pm$ 0.05
$t\bar{t}$, bin [1750,2500]	421 $\pm$ 11	58	206 $\pm$ 8	51	0.03$\pm$ 0.10
$t\bar{t}$, bin $\geq$2500	94 $\pm$ 5	56	58 $\pm$ 4	42	0.28$\pm$ 0.20
W+jets, bin [1000,1750]	269 $\pm$ 7	8	500 $\pm$ 8	31	0.35$\pm$ 0.09
W+jets, bin [1750,2500]	69 $\pm$ 3	10	151 $\pm$ 4	38	0.04$\pm$ 0.15
W+jets, bin $\geq$2500	26 $\pm$ 2	15	68 $\pm$ 3	48	0.25$\pm$ 0.18
ST, bin [1000,1750]	409 $\pm$ 11	12	160 $\pm$ 7	10	1.10$\pm$ 0.30
ST, bin [1750,2500]	71 $\pm$ 4	10	37 $\pm$ 3	9	0.15$\pm$ 0.61
ST, bin $\geq$2500	21 $\pm$ 3	13	11 $\pm$ 2	8	1.49$\pm$ 1.13
QCD, bin [1000,1750]	613 $\pm$ 66	18	14 $\pm$ 7	1	12.27$\pm$ 6.55
QCD, bin [1750,2500]	159 $\pm$ 36	22	7 $\pm$ 2	2	0.83$\pm$ 3.27
QCD, bin $\geq$2500	26 $\pm$ 8	16	3 $\pm$ 1	2	4.78$\pm$ 4.02
	SR2B		SR2B_I		
Tot MC, bin [1000,1750]	1192 $\pm$ 55	100	828 $\pm$ 12	100	
Tot MC, bin [1750,2500]	164 $\pm$ 8	100	188 $\pm$ 6	100	
Tot MC, bin $\geq$2500	34 $\pm$ 3	100	55 $\pm$ 3	100	
$t\bar{t}$, bin [1000,1750]	732 $\pm$ 12	61	411 $\pm$ 7	50	0.06$\pm$ 0.07
$t\bar{t}$, bin [1750,2500]	106 $\pm$ 5	64	67 $\pm$ 3	36	0.45$\pm$ 0.20
$t\bar{t}$, bin $\geq$2500	22 $\pm$ 2	63	15 $\pm$ 2	26	1.46$\pm$ 0.47
W+jets, bin [1000,1750]	31 $\pm$ 2	3	156 $\pm$ 5	19	0.15$\pm$ 0.20
W+jets, bin [1750,2500]	7 $\pm$ 1	4	52 $\pm$ 3	28	0.58$\pm$ 0.26
W+jets, bin $\geq$2500	4 $\pm$ 1	12	19 $\pm$ 1	34	1.12$\pm$ 0.35
ST, bin [1000,1750]	251 $\pm$ 8	21	252 $\pm$ 8	30	0.09$\pm$ 0.12
ST, bin [1750,2500]	40 $\pm$ 3	25	67 $\pm$ 4	35	0.45$\pm$ 0.20
ST, bin $\geq$2500	7 $\pm$ 1	22	21 $\pm$ 2	38	1.02$\pm$ 0.32
QCD, bin [1000,1750]	178 $\pm$ 53	15	10 $\pm$ 4	1	2.22$\pm$ 3.07
QCD, bin [1750,2500]	11 $\pm$ 5	7	2 $\pm$ 1	1	16.28$\pm$ 13.14
QCD, bin $\geq$2500	1 $\pm$ 1	3	1 $\pm$ 1	2	24.83$\pm$ 17.00

A.2.4 Cross checks in the subregions II and III

By inverting the requirement on the $m_{SD,AK8}$, one can define the two additional subregions per each main region and for the CR0B. Subregion-II is defined by requiring $m_{SD,AK8} > 60$ GeV and $120 < m_{top,best} < 220$ GeV. Subregion-III is defined by requiring $m_{SD,AK8} > 60$ GeV and $m_{top,best} > 220$ GeV. The extraction procedure is further validated in this subregion, and can also provide a further correction to the procedure in case there is a bin-by-bin dependence on the background composition that is not covered by the other uncertainties previously defined in this chapter. The procedure is the same as for Sec. 6.7.1. The equivalent of Eq. 6.9 writes:

$$\mathrm{DDBKG_{nonQCD}}^{II} = TF(\mathrm{Data}^{III} - \mathrm{QCD}^{III}), \tag{A.11}$$

where:

$$TF = \frac{\mathrm{BKG\,MC_{nonQCD}}^{II}}{\mathrm{BKG\,MC_{nonQCD}}^{III}}. \tag{A.12}$$

Data in subregion-III are fitted with a function of the form $a\mathrm{e}^{bx} + c + dx$. Fig. A.2.11 shows the fit to the data in region III. Fig. A.2.12 shows the transfer functions from region III to region II. A correction factor (CF) is then derived:

$$CF_{\mathrm{main\,region}} = \frac{\mathrm{Data}^{II} - \mathrm{QCD}^{II}}{\mathrm{DDBKG_{nonQCD}}^{II}} \tag{A.13}$$

And finally applied to the SR as follows:

$$\mathrm{DDBKG_{nonQCD}}^{corr} = CF_{\mathrm{main\,region}} \times \mathrm{DDBKG}_{nonQCD} \tag{A.14}$$

This is not used as the central value, but rather the difference between this and the central value is used as conservative systematic uncertainty, to take into account any potential differences between the two functions. In the case of SR2B, since there is not enough statistics to constitute a meaningful control region, and the composition is similar to SRT, the CF from SRT-II/SRT-III is taken.

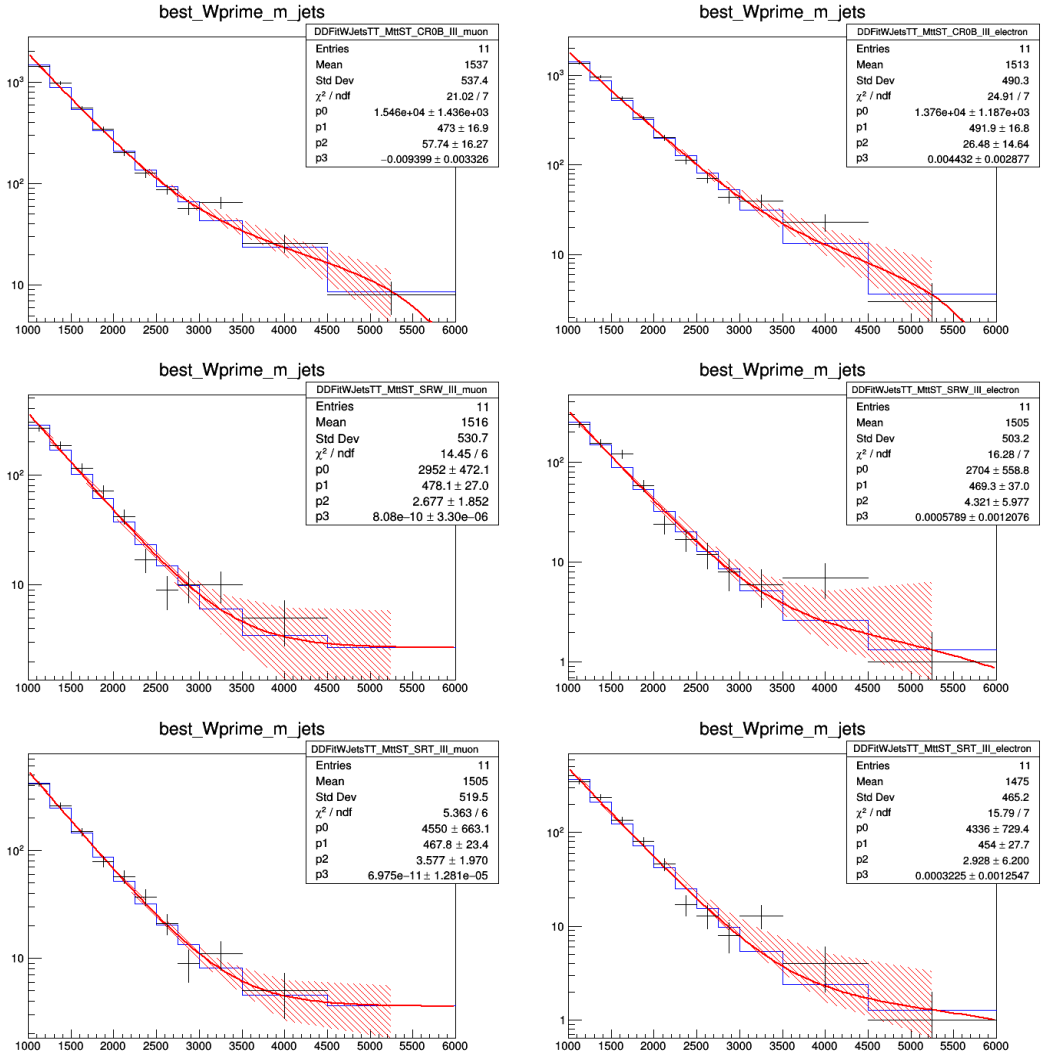


Figure A.2.11: Fit to data fit in the region III, for CR0B (first row), SRW (second row), and SRT (third row), for muons (left), and electrons (right). The black histogram is data, the blue histogram is the integral of the red line. On the $x-$axis, reconstructed mass of the W' boson in GeV.

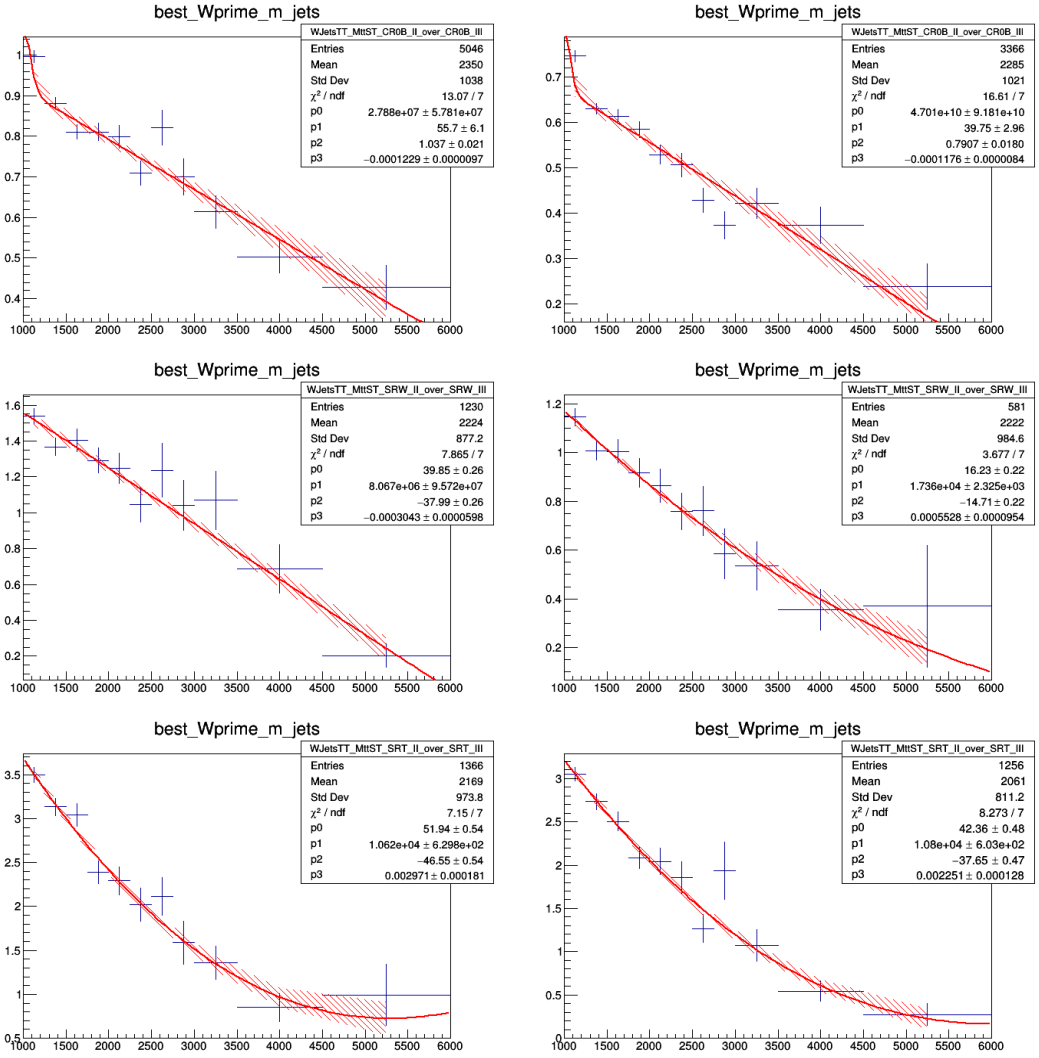


Figure A.2.12: Ratio of the distributions of $m_{W'}$ between subregion-II and subregion-III, the analytic form for the function fitted and the fit results for muon (left) and electron (right) for CR0B (first row), SRW (second row), and SRT (third row). On the $x-$axis, reconstructed mass of the W' boson in GeV.

Appendix $\mathcal{B}$

Additional plots of interest for the ssWW VBS with one hadronic tau in final state studies

B.1 Prefit distributions

In the following, the distributions of all the quantities injected as input features in the DNN discriminators explained in Sec. 7.7 is shown in the $t\bar{t}$, Opposite Sign and Fakes control regions, and in the Signal Region from Fig. B.1.1 to Fig. B.1.24. The SM, dim-6, and dim-8 DNN output distributions are also shown for the same regions from Fig. B.1.25 to Fig B.1.27.

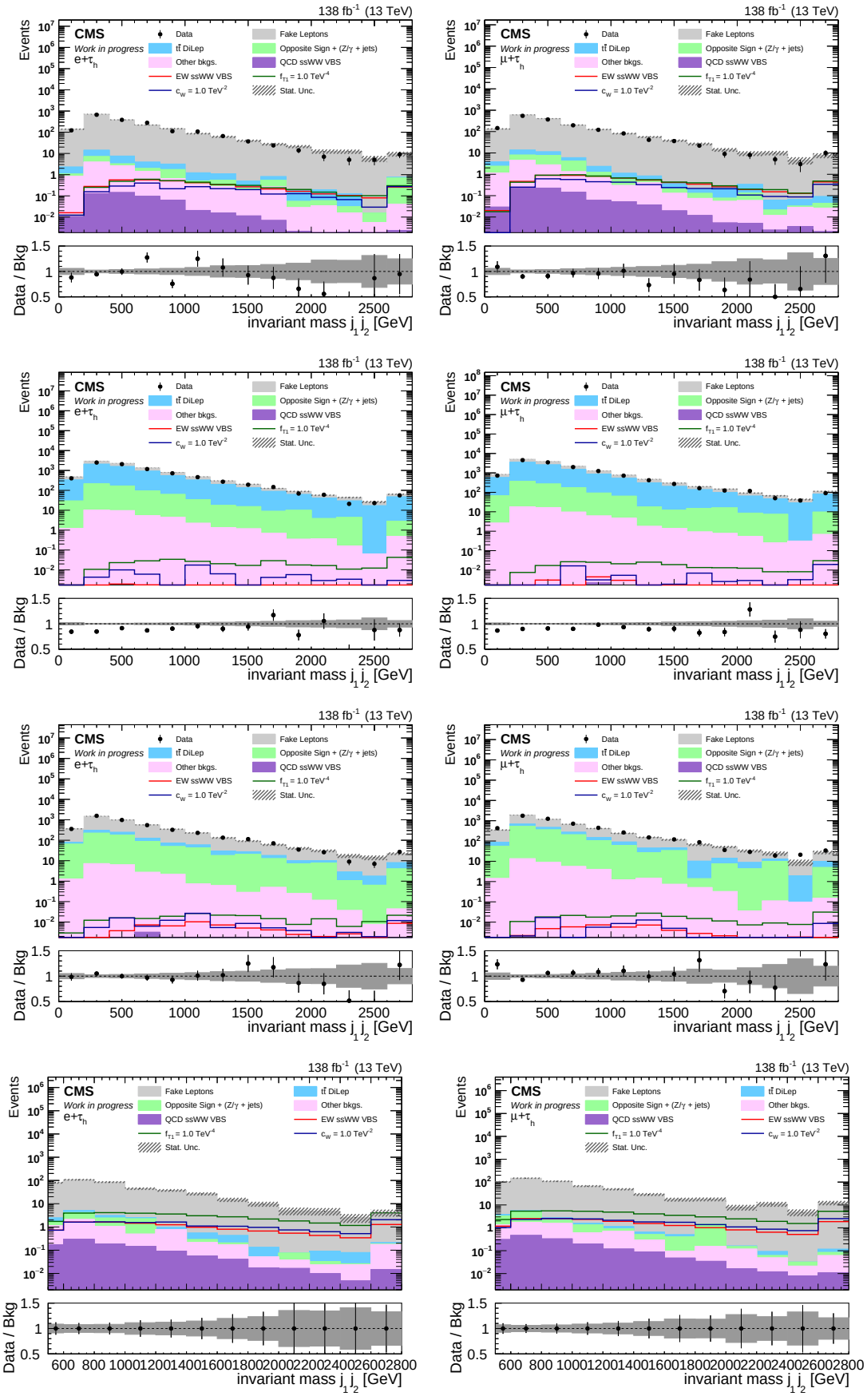


Figure B.1.1: Prefit m_{jj} distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

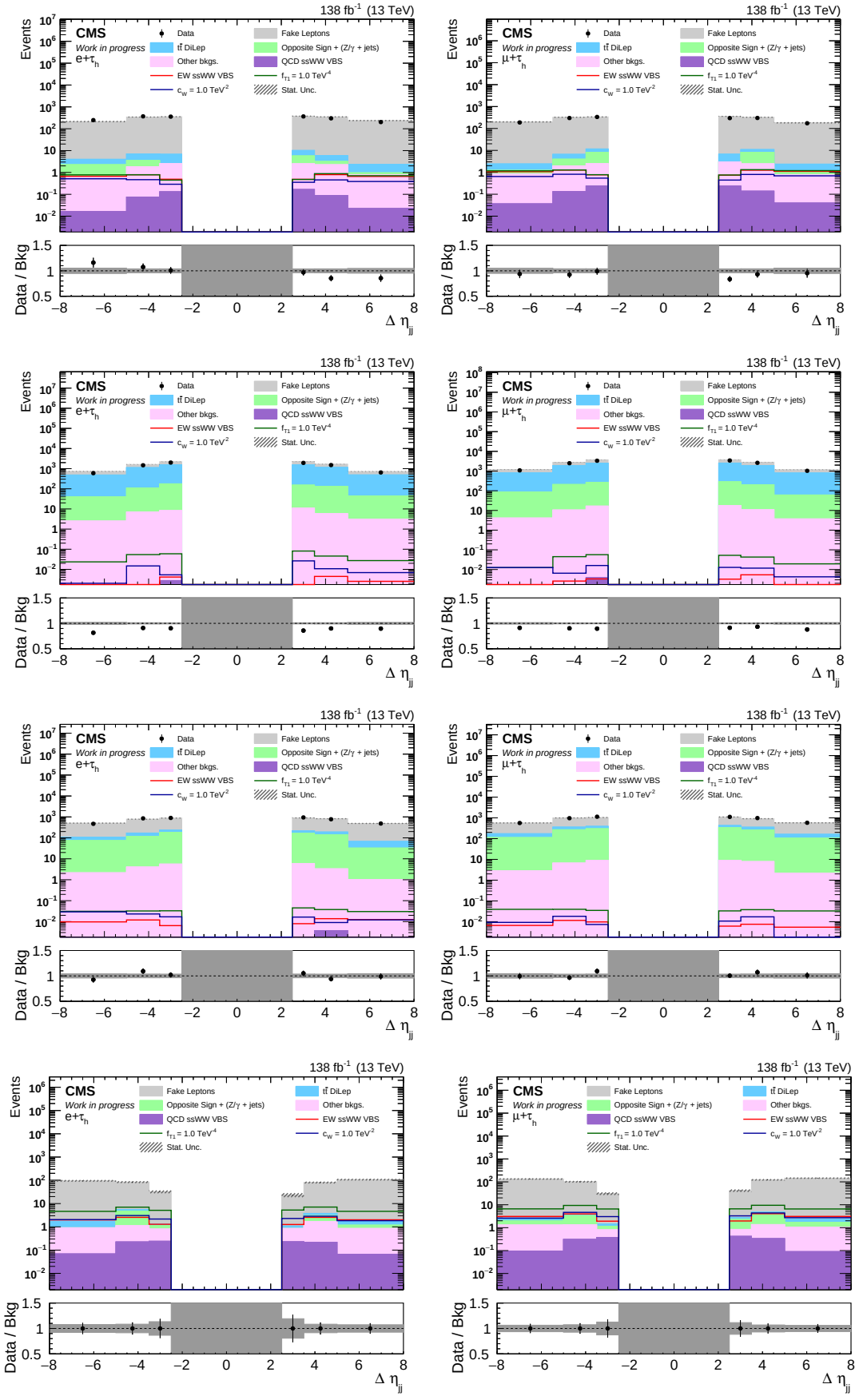


Figure B.1.2: Prefit $\Delta\eta_{jj}$ distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

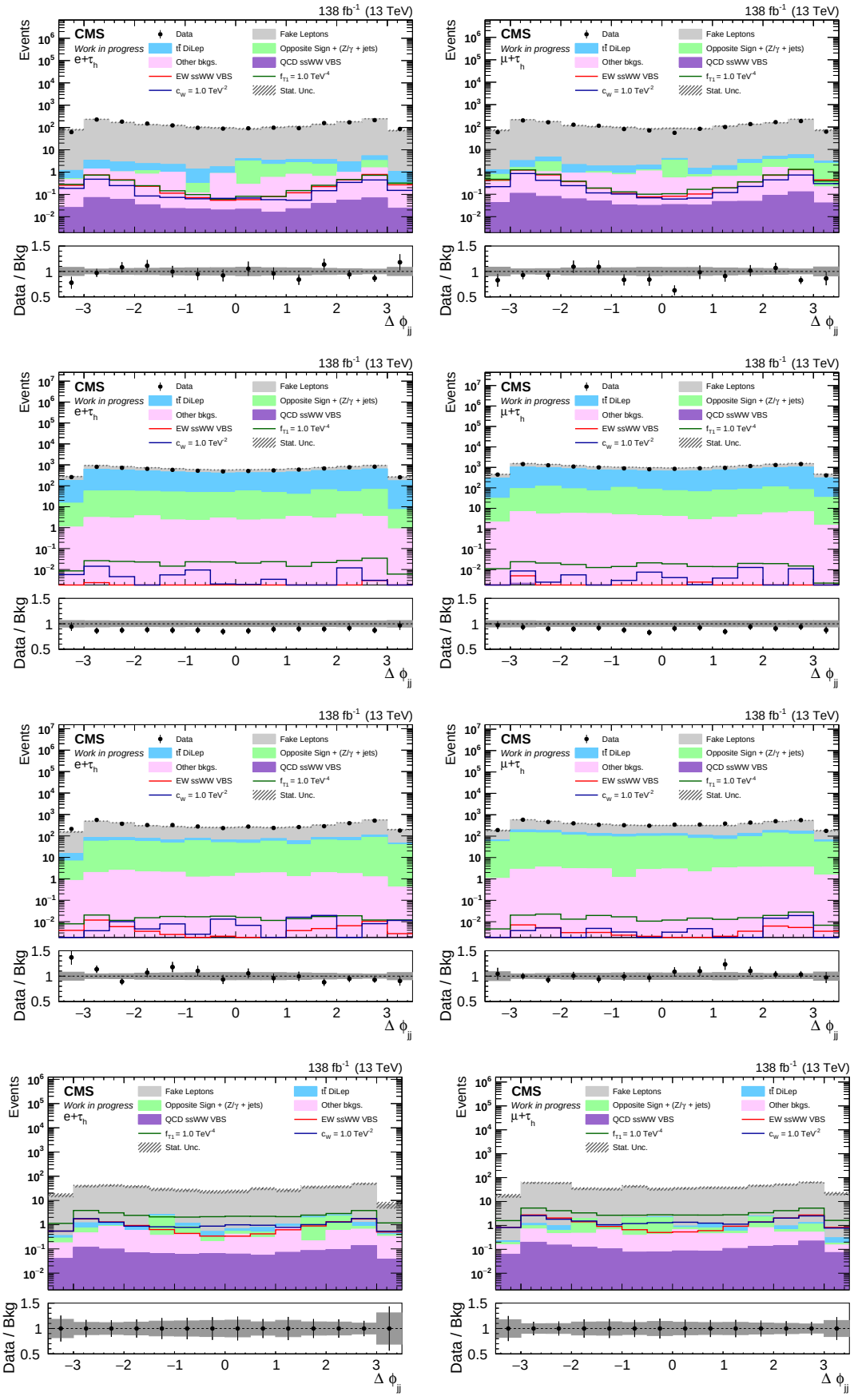


Figure B.1.3: Prefit $\Delta\phi_{jj}$ distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

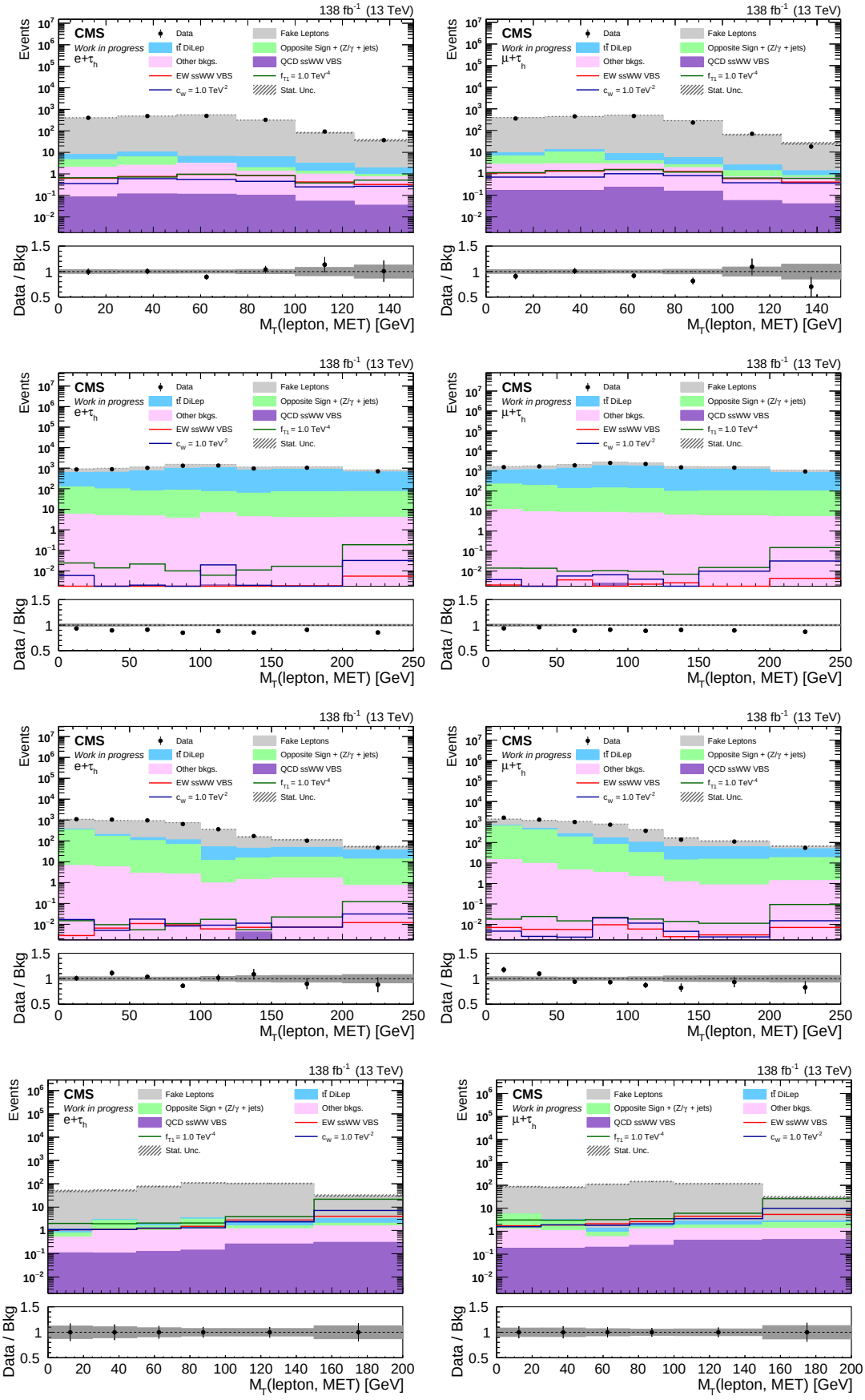


Figure B.1.4: Prefit $m_T(\ell, p_T^{miss})$ distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

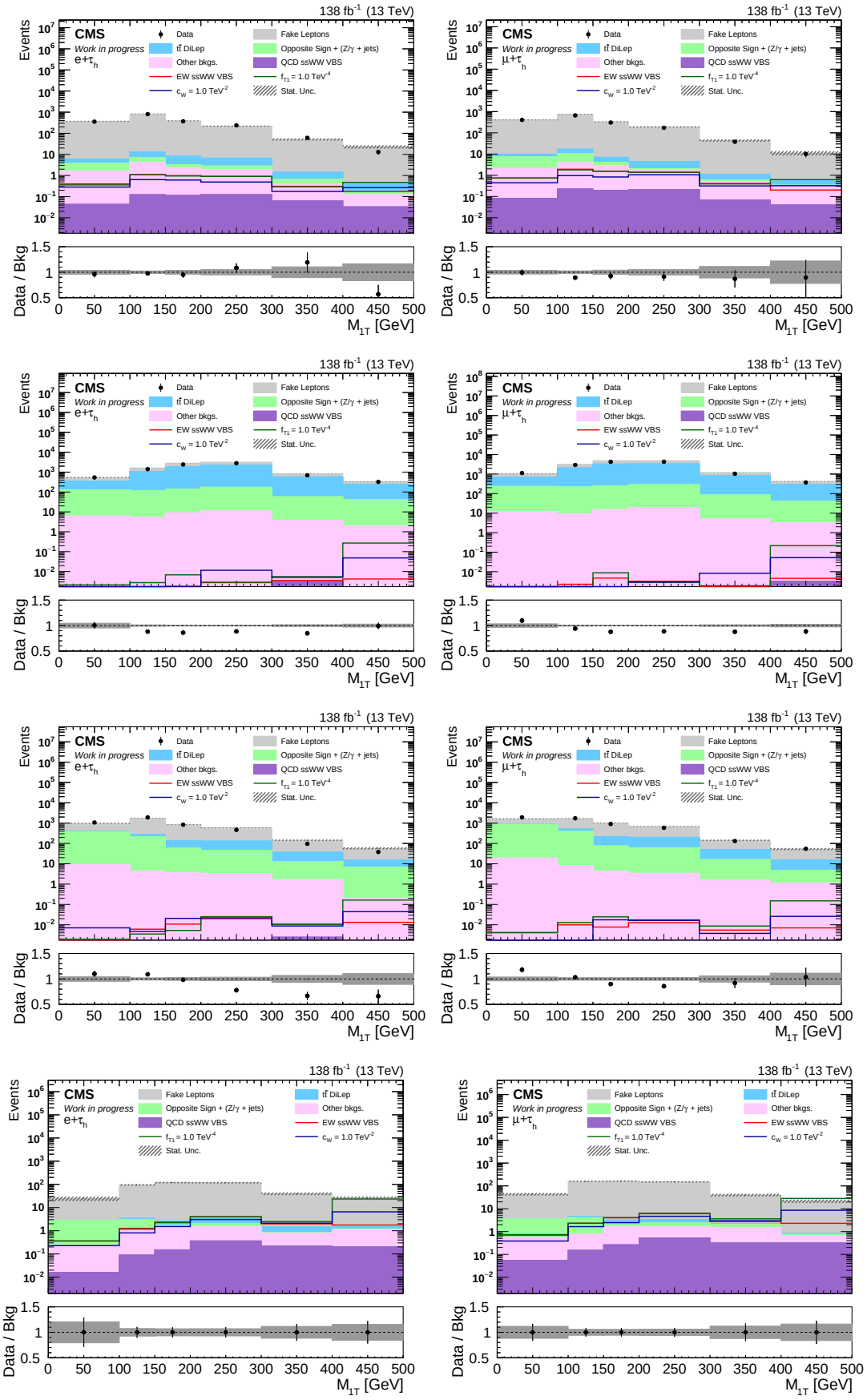


Figure B.1.5: Prefit m_{1T} distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

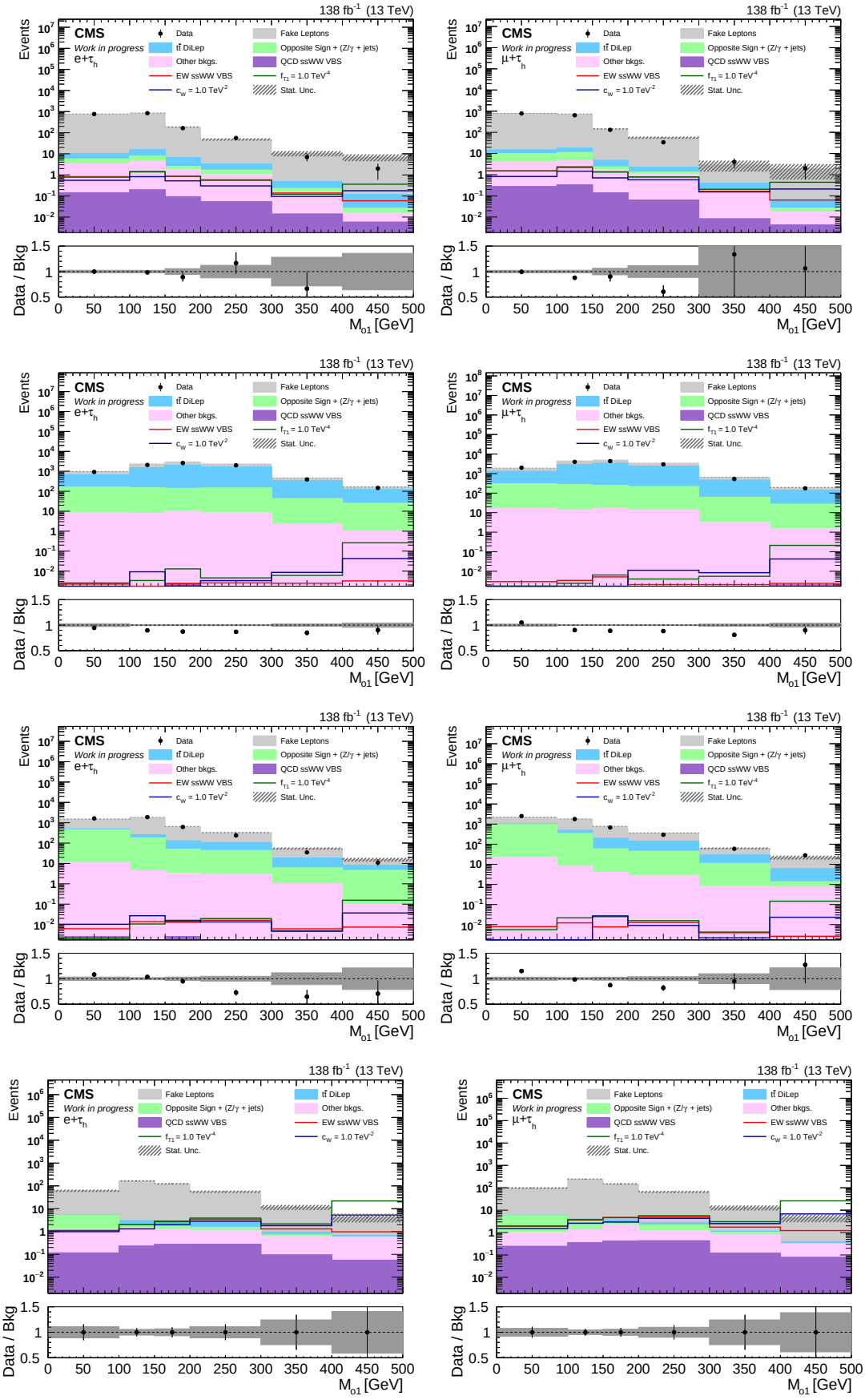


Figure B.1.6: Prefit m_{o1} distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

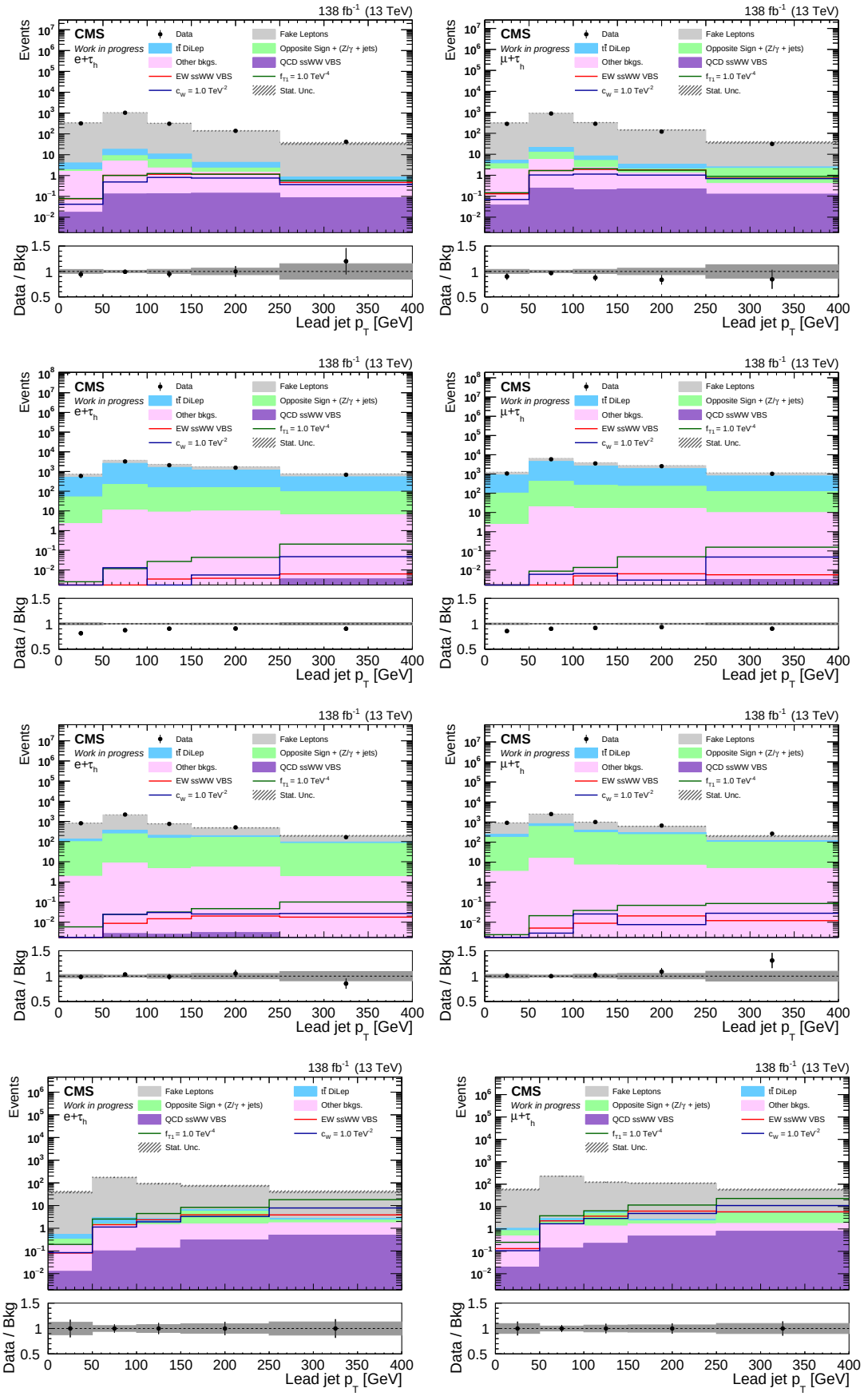


Figure B.1.7: Prefit leading jet p_T distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

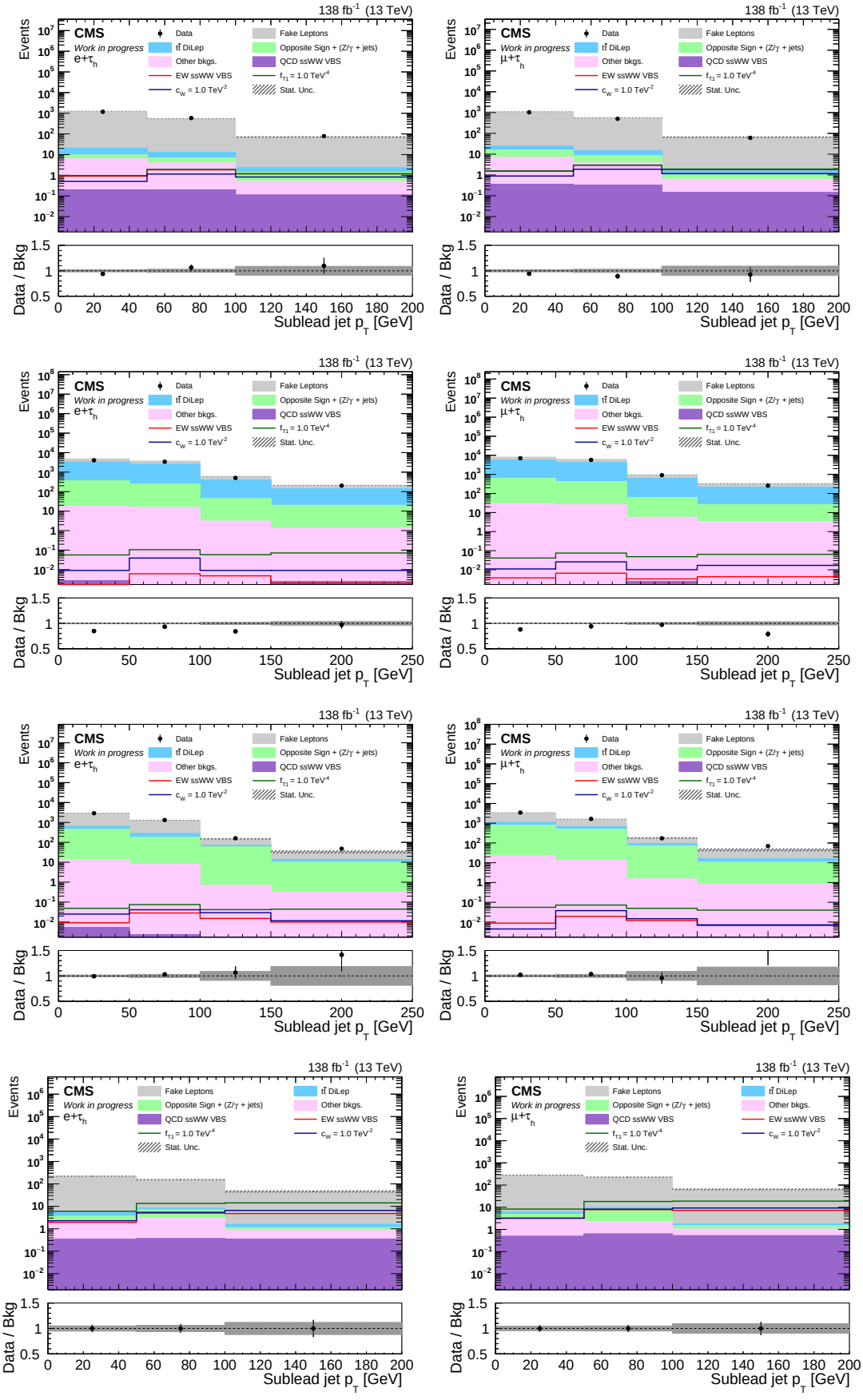


Figure B.1.8: Prefit Subleading jet p_T distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

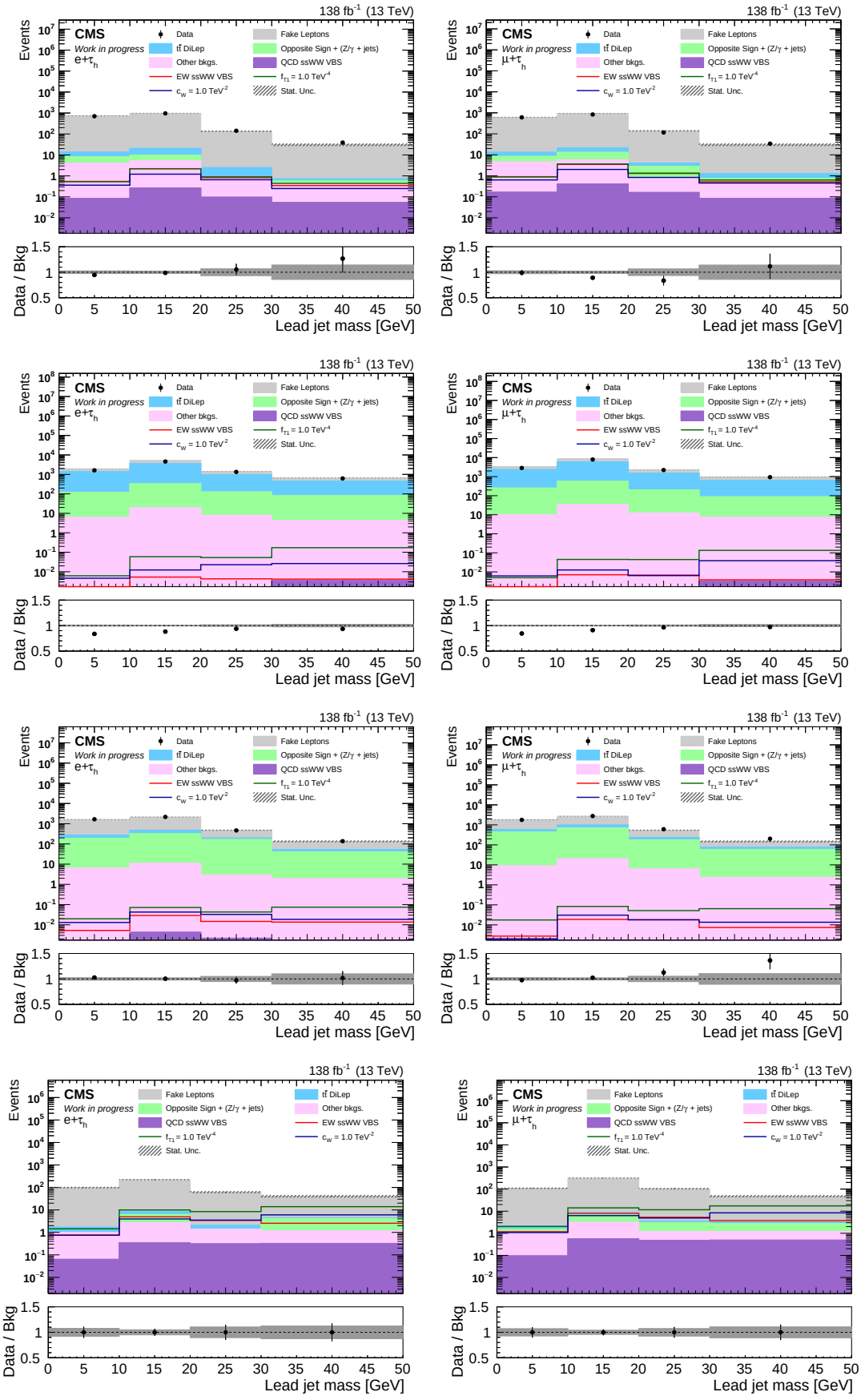


Figure B.1.9: Prefit leading jet mass distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

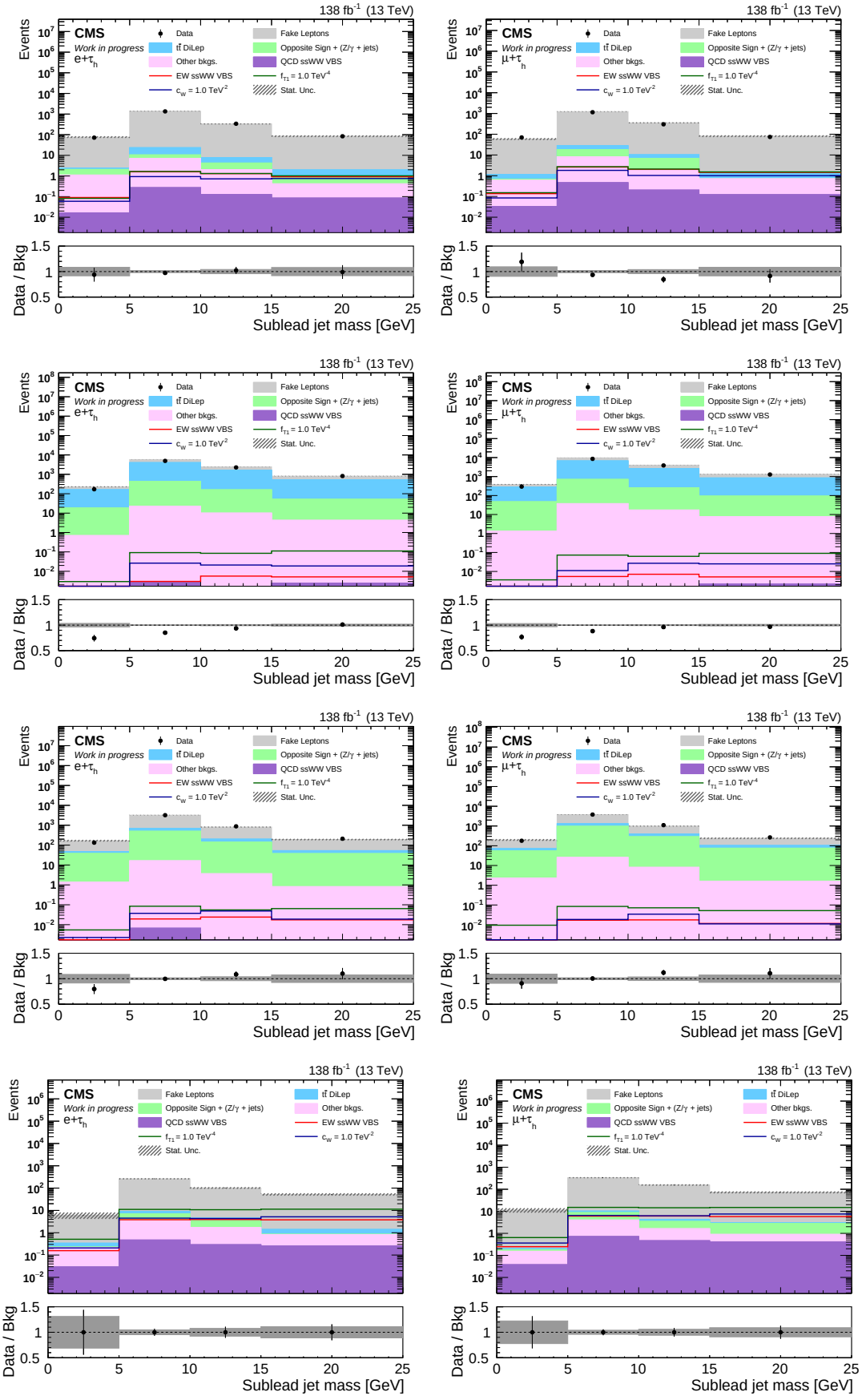


Figure B.1.10: Prefit Subleading jet mass distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

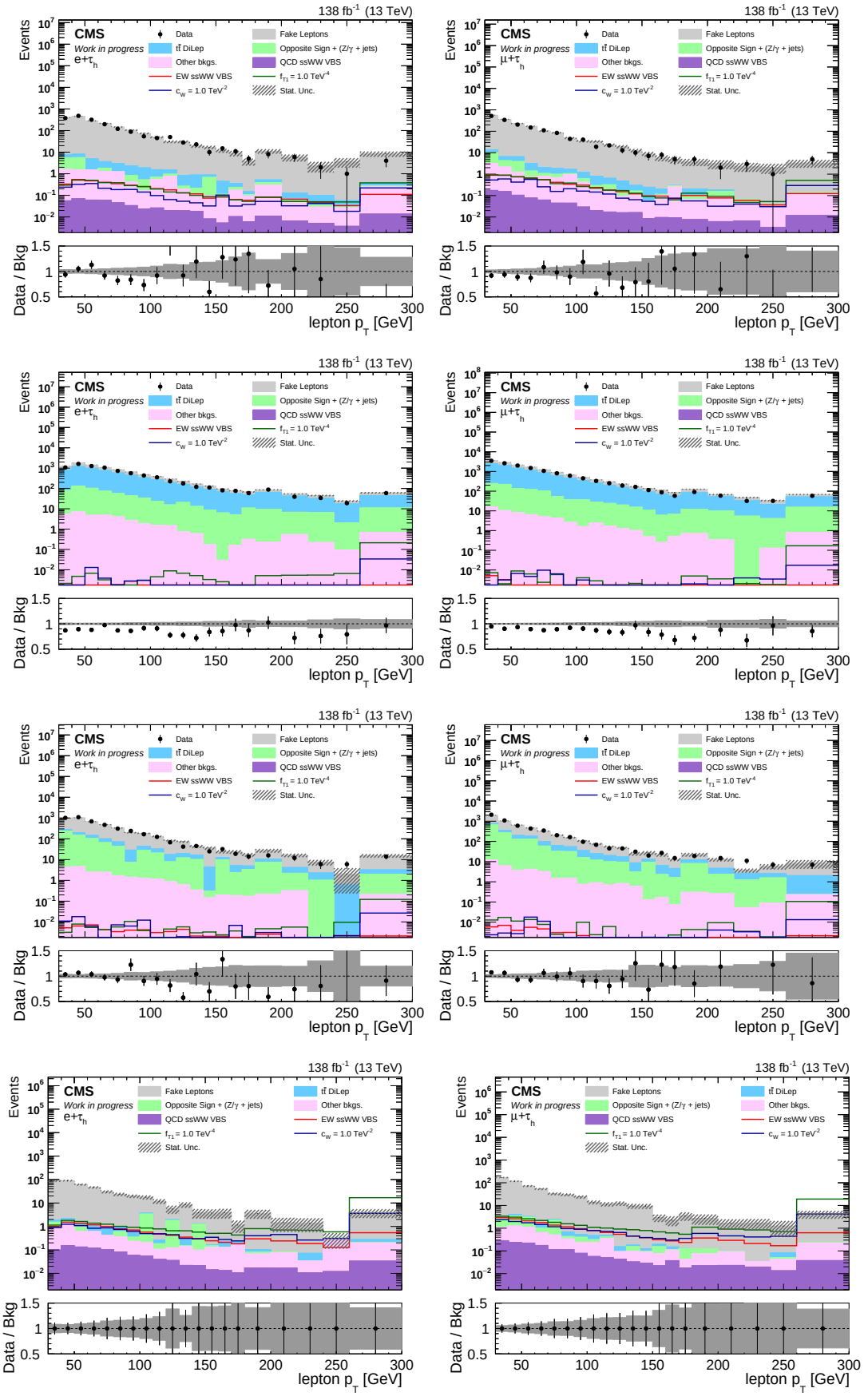


Figure B.1.11: Prefit lepton p_T distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

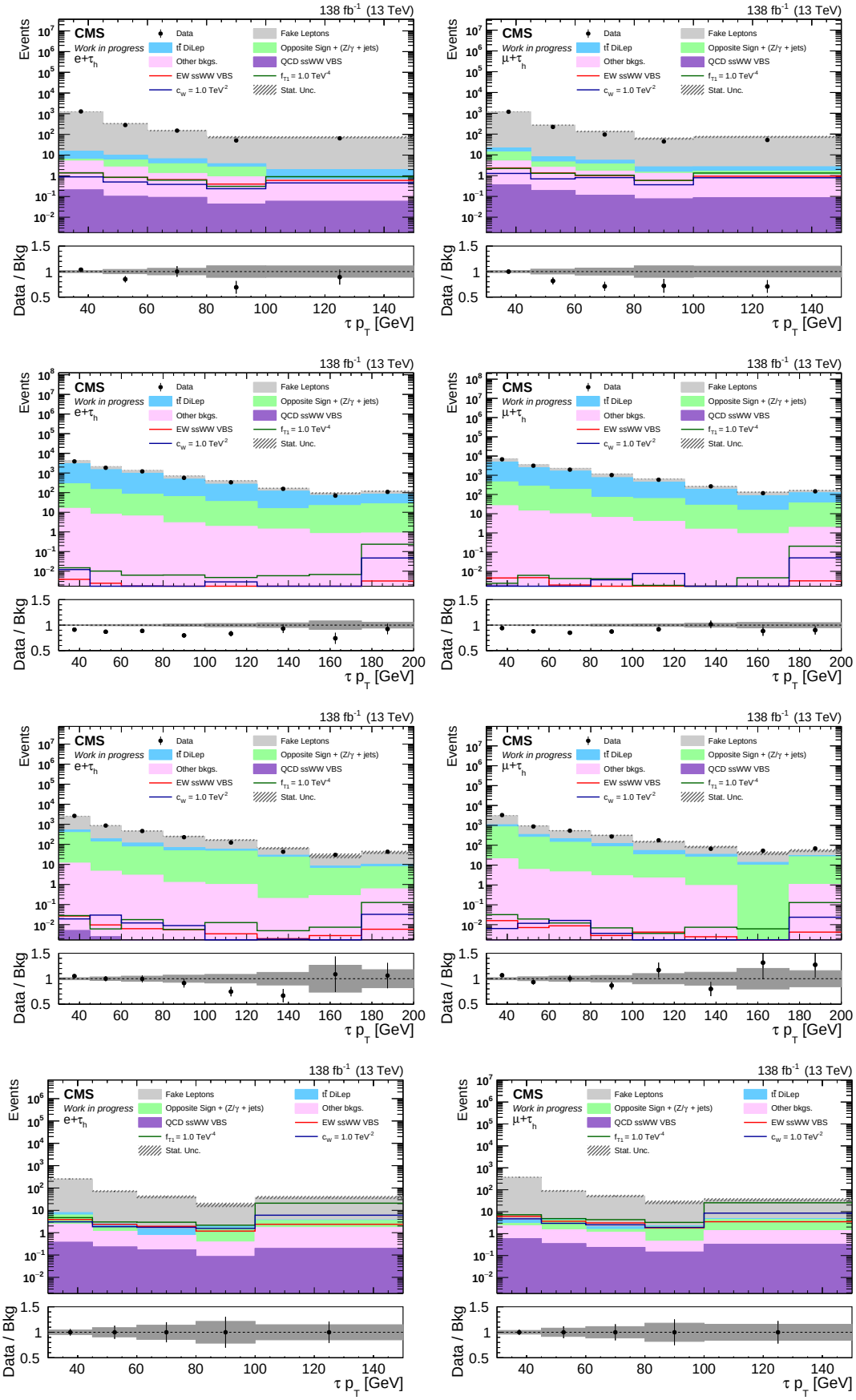


Figure B.1.12: Prefit τ_h p_T distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

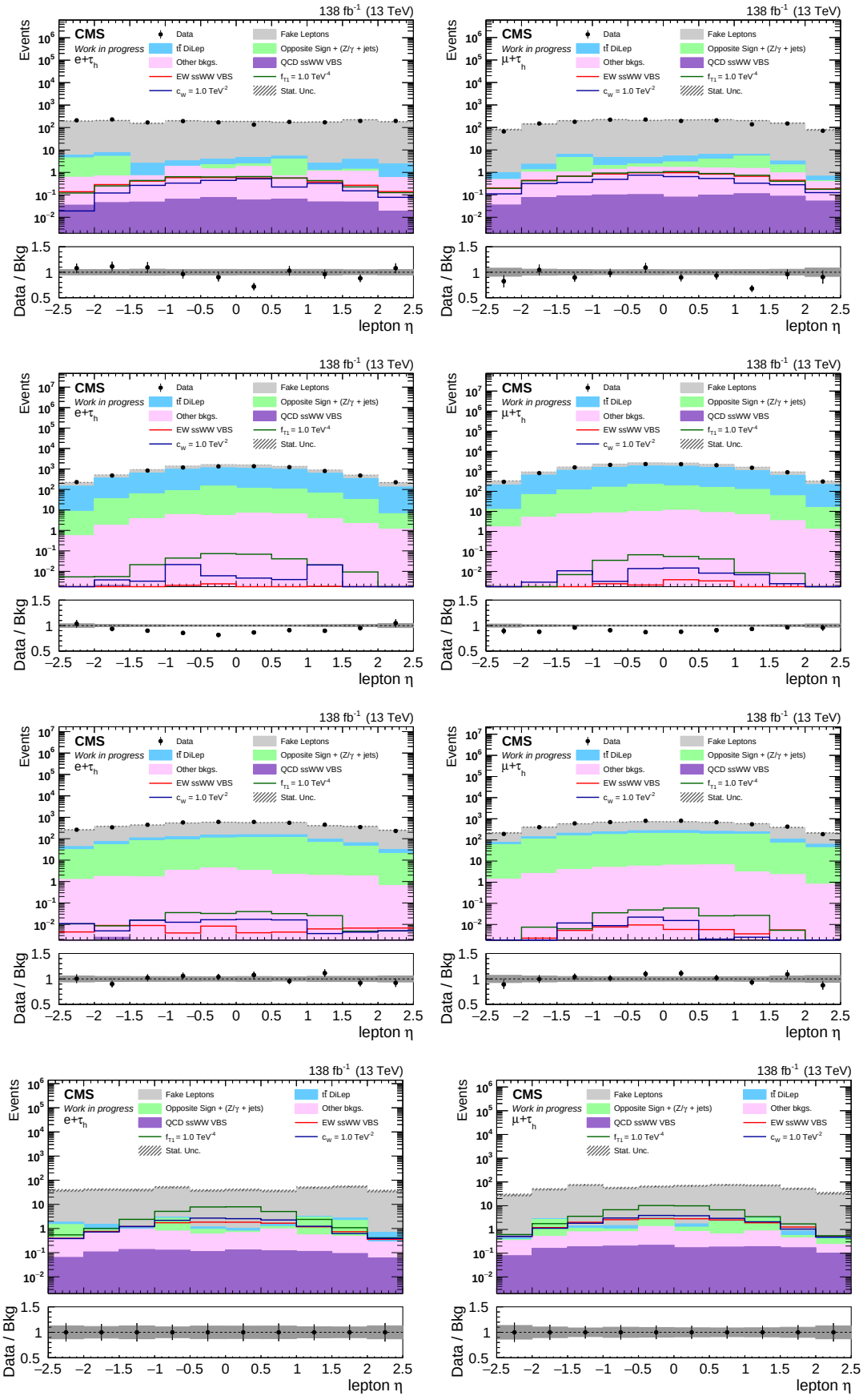


Figure B.1.13: Prefit lepton η distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

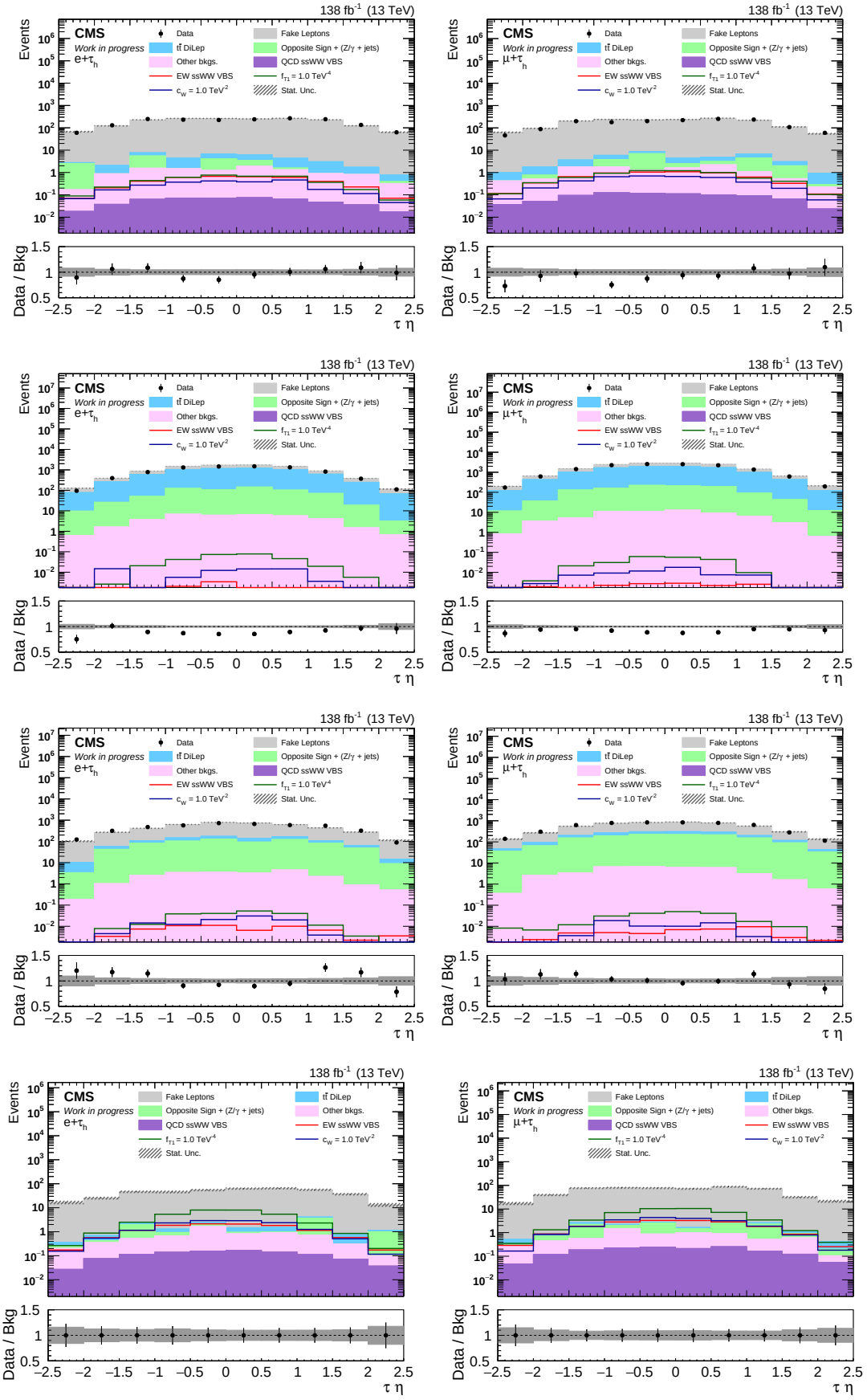


Figure B.1.14: Prefit τ_h η distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

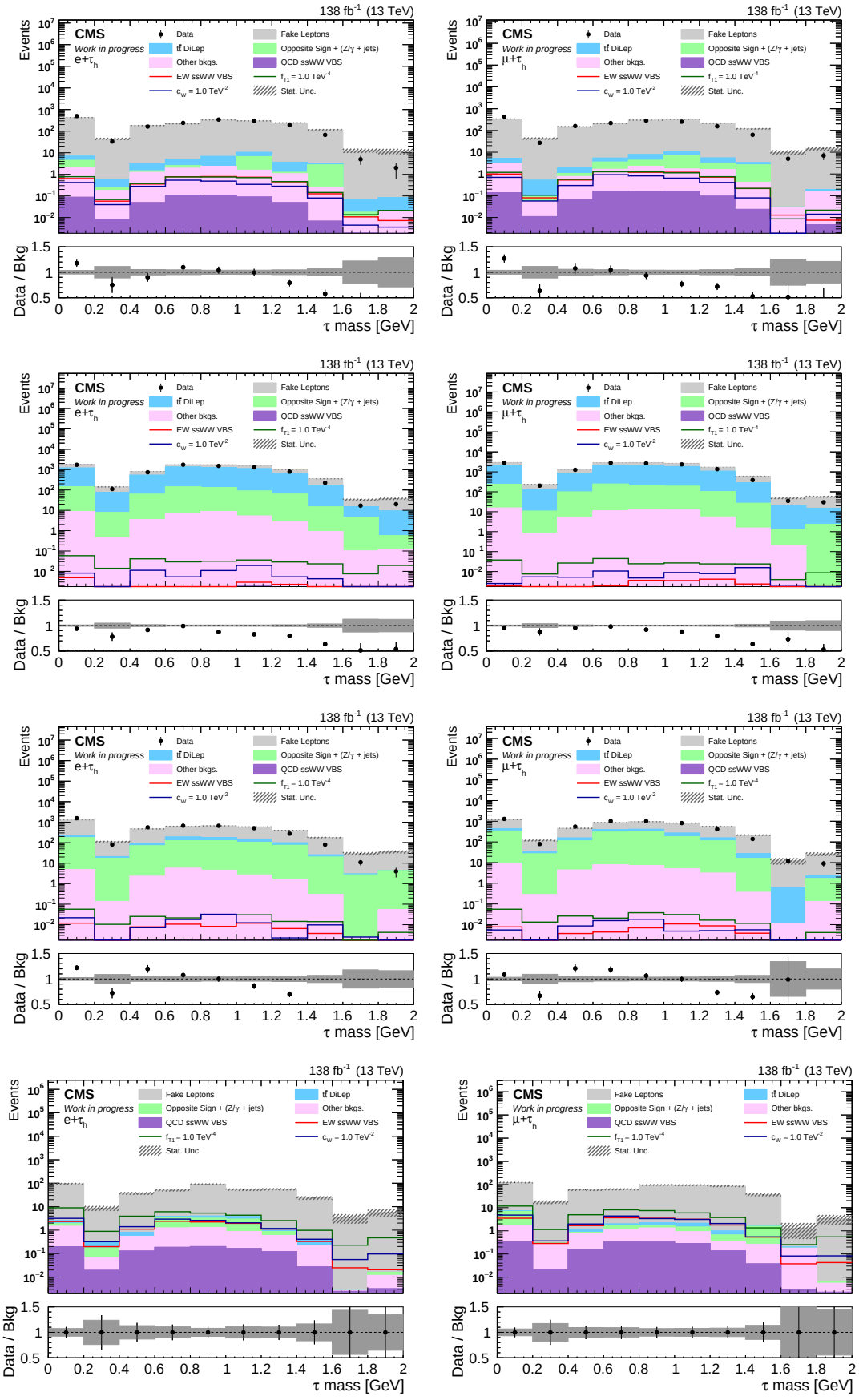


Figure B.1.15: Prefit τ_h mass distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

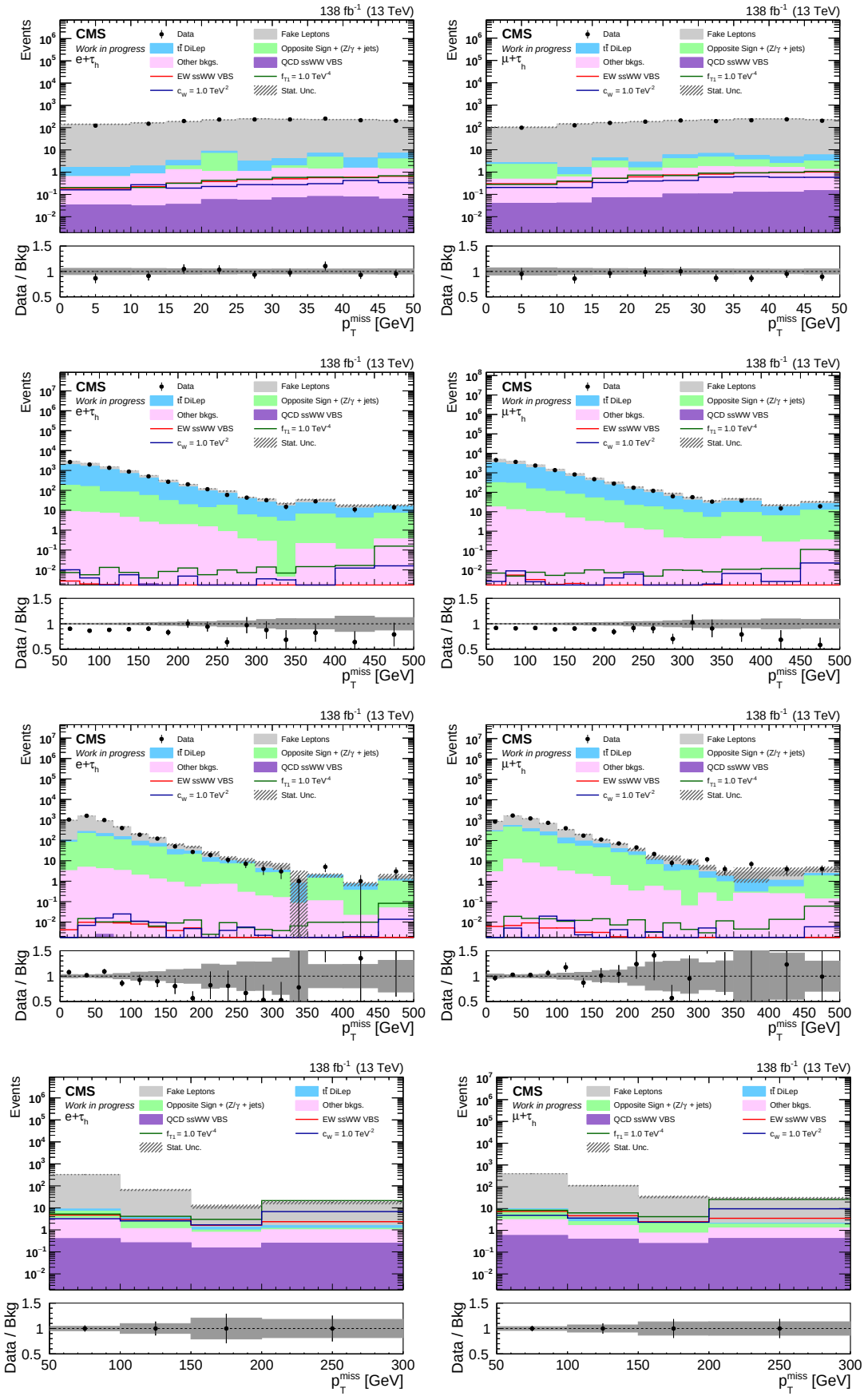


Figure B.1.16: Prefit p_T^{miss} distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

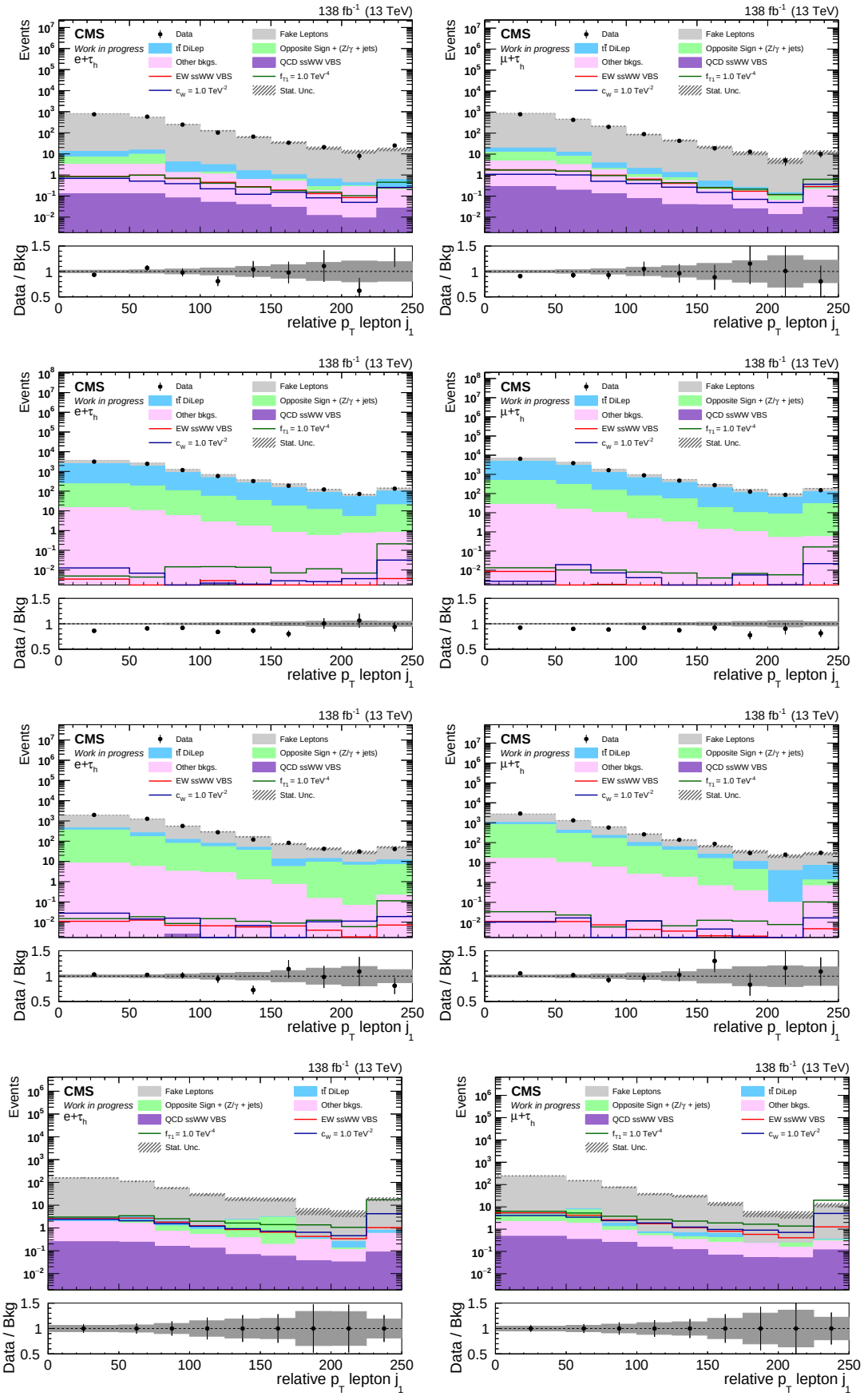


Figure B.1.17: Prefit $p_T^{rel}(\ell$, leading jet) distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

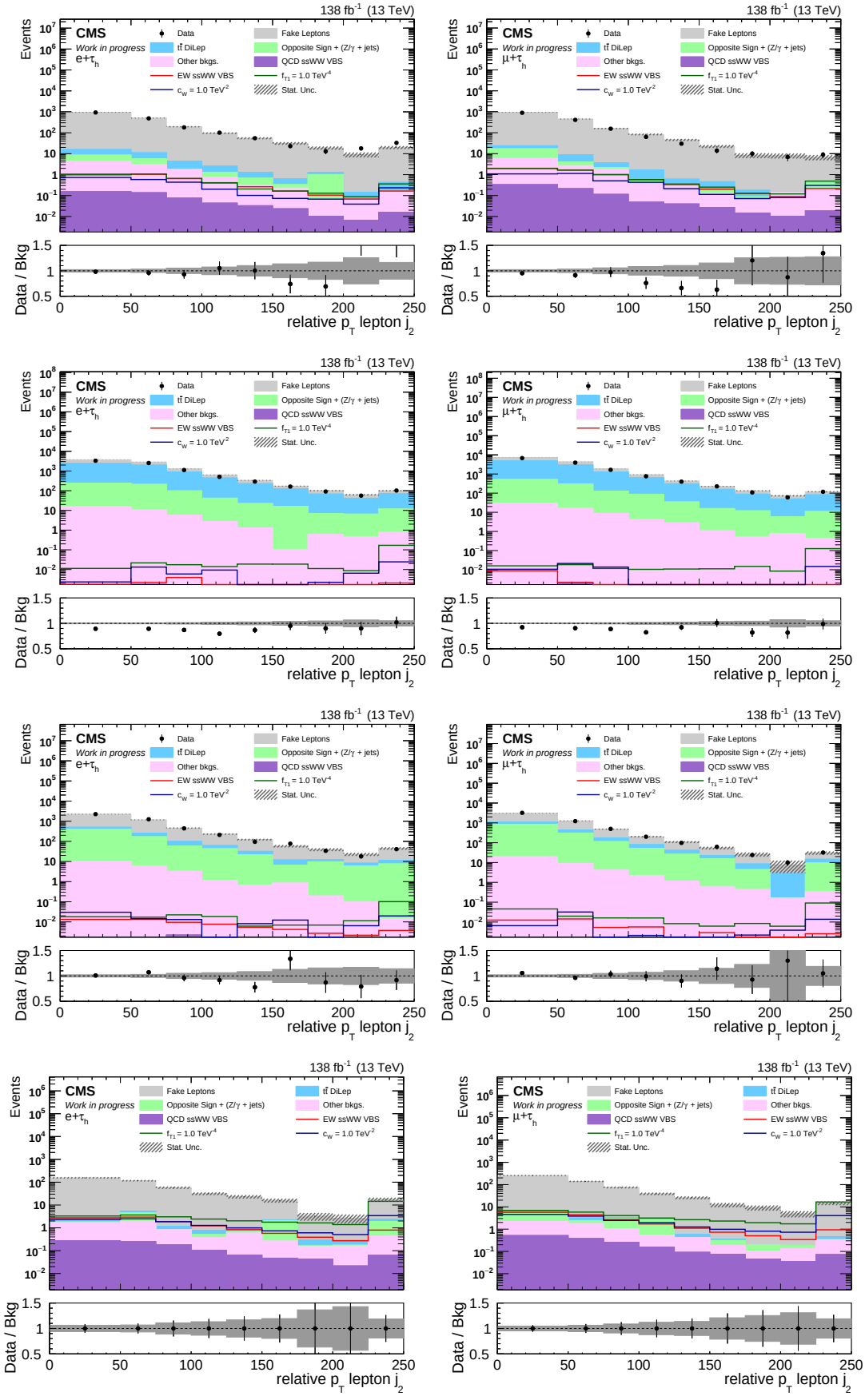


Figure B.1.18: Prefit $p_T^{rel}(\ell$, subleading jet) distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

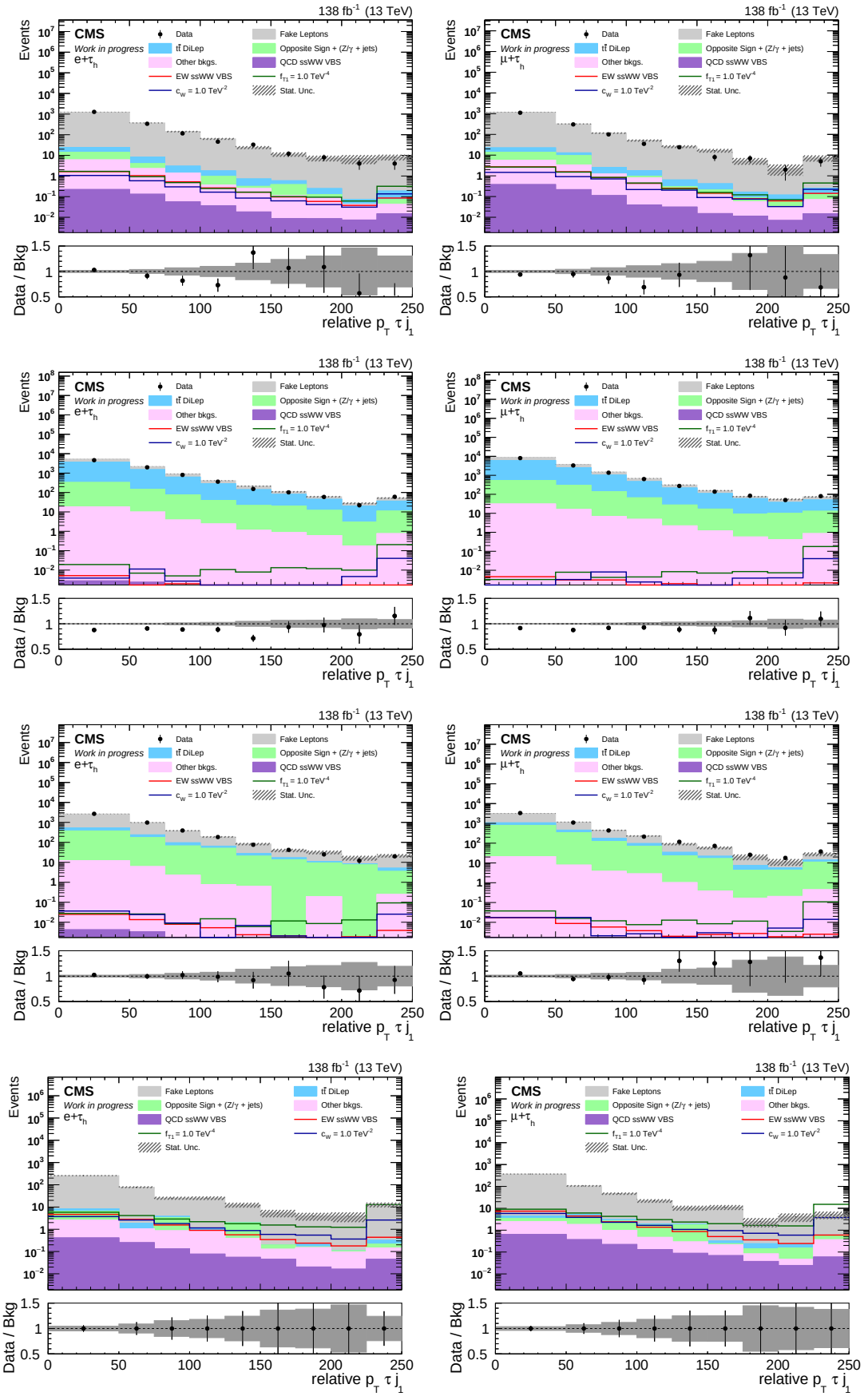


Figure B.1.19: Prefit $p_T^{rel}(\tau_h$, subleading jet) distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

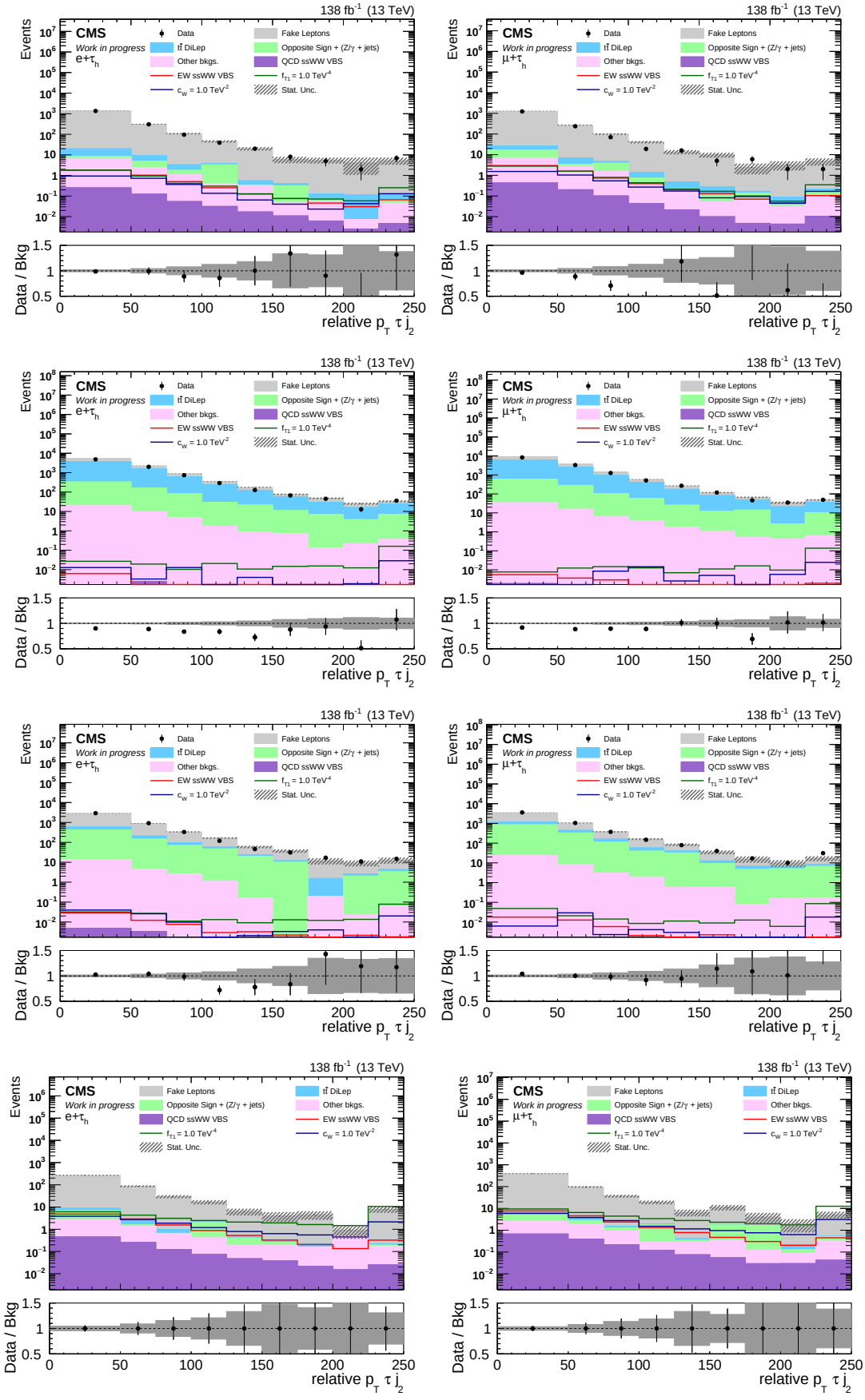


Figure B.1.20: Prefit $p_T^{rel}(\tau_h$, subleading jet) distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

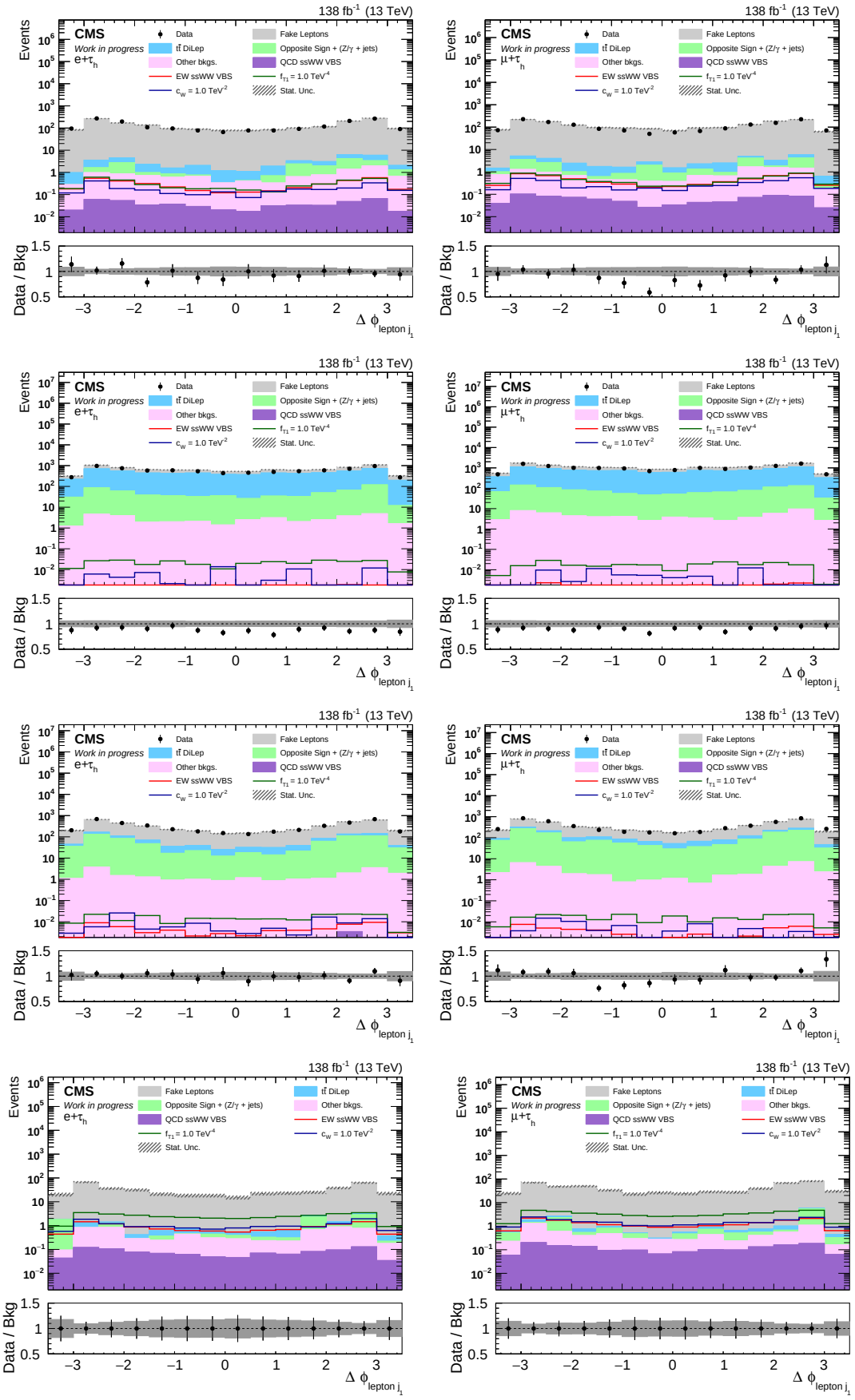


Figure B.1.21: Prefit $\Delta\phi(\ell, j_1)$ distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

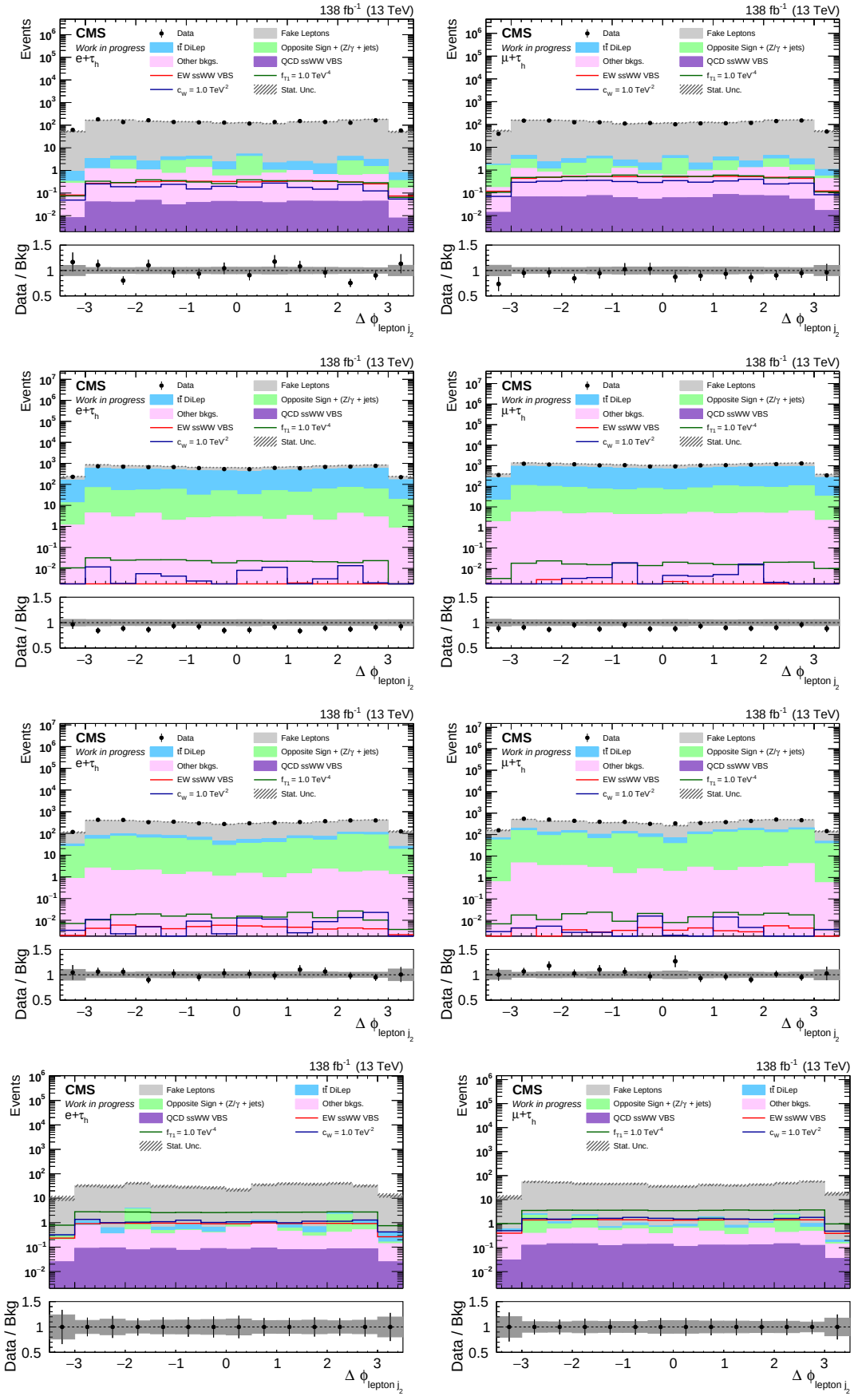


Figure B.1.22: Prefit $\Delta\phi(\ell, j_2)$ distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

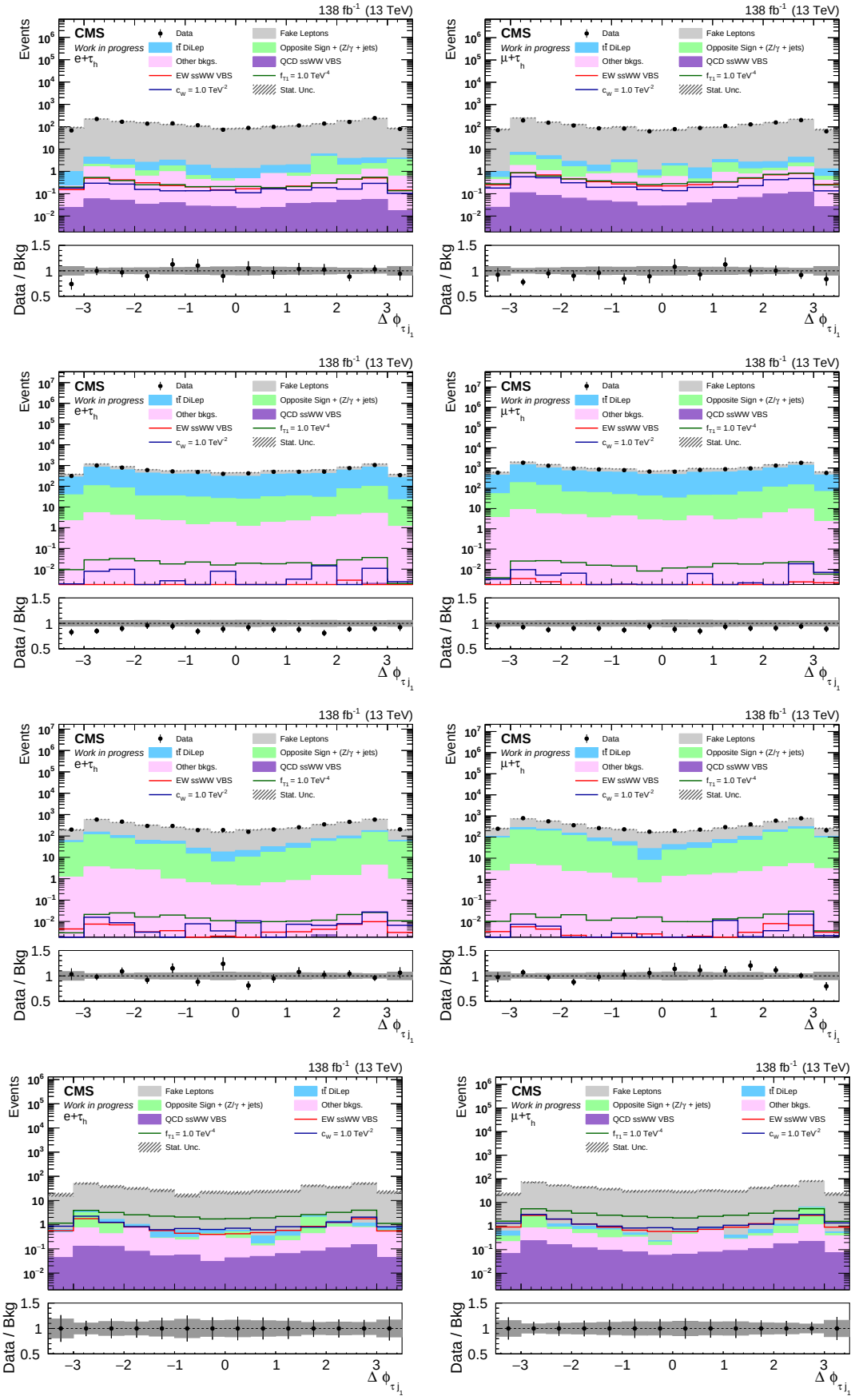


Figure B.1.23: Prefit $\Delta\phi(\tau_h, j_1)$ distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

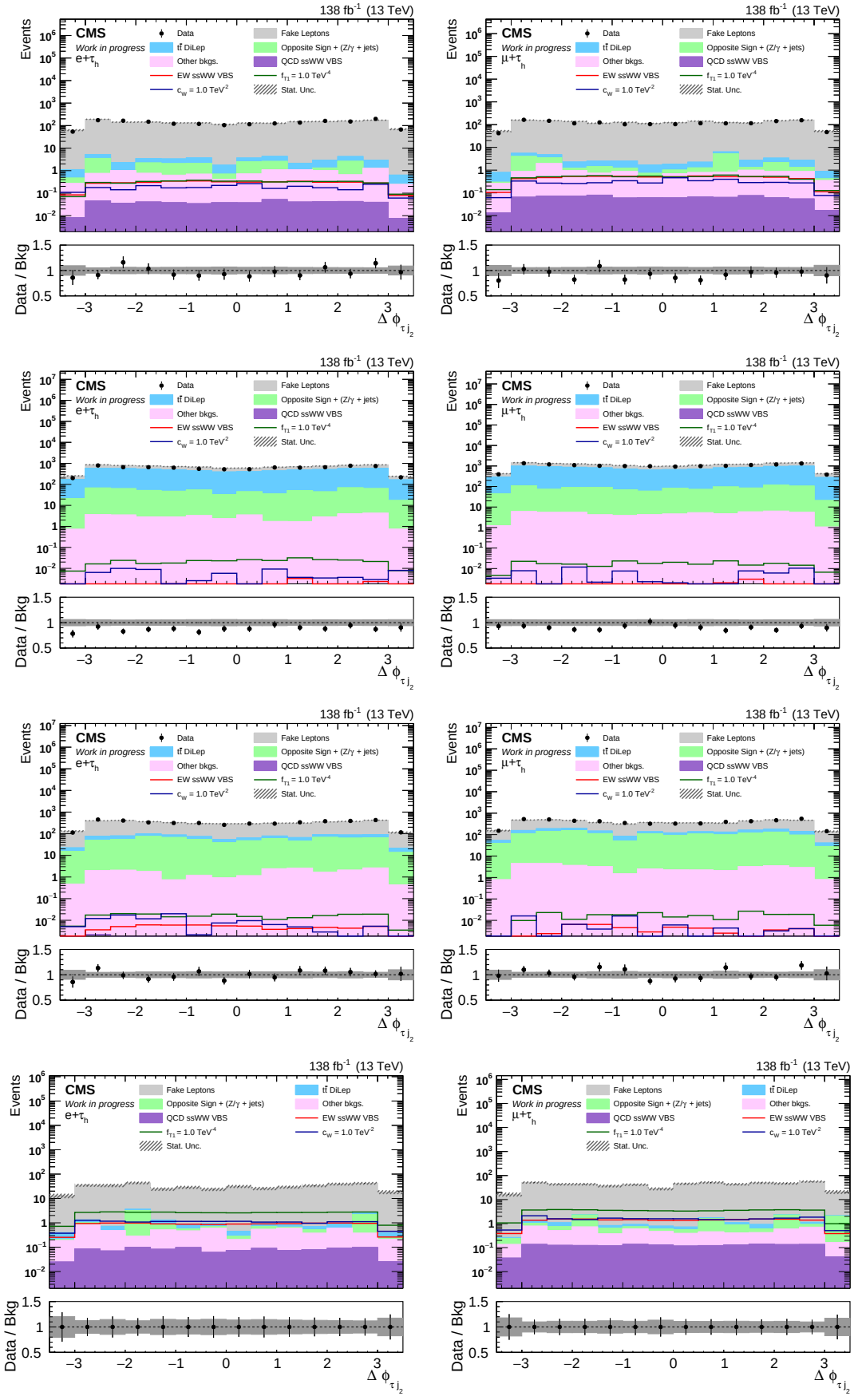


Figure B.1.24: Prefit $\Delta\phi(\tau_h, j_2)$ distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels.

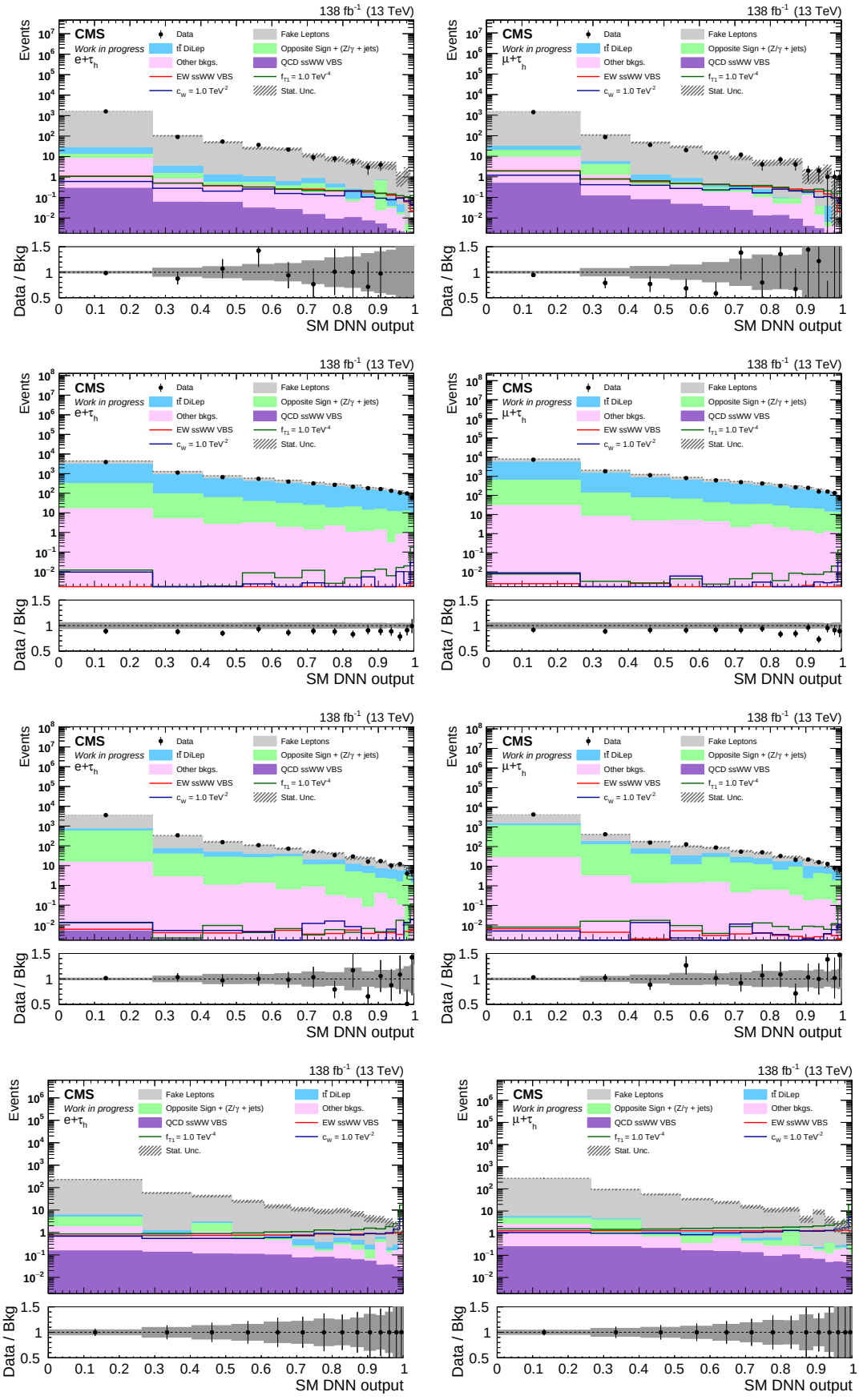


Figure B.1.25: Prefit SM DNN distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels. broad signal-flattening binning is used.

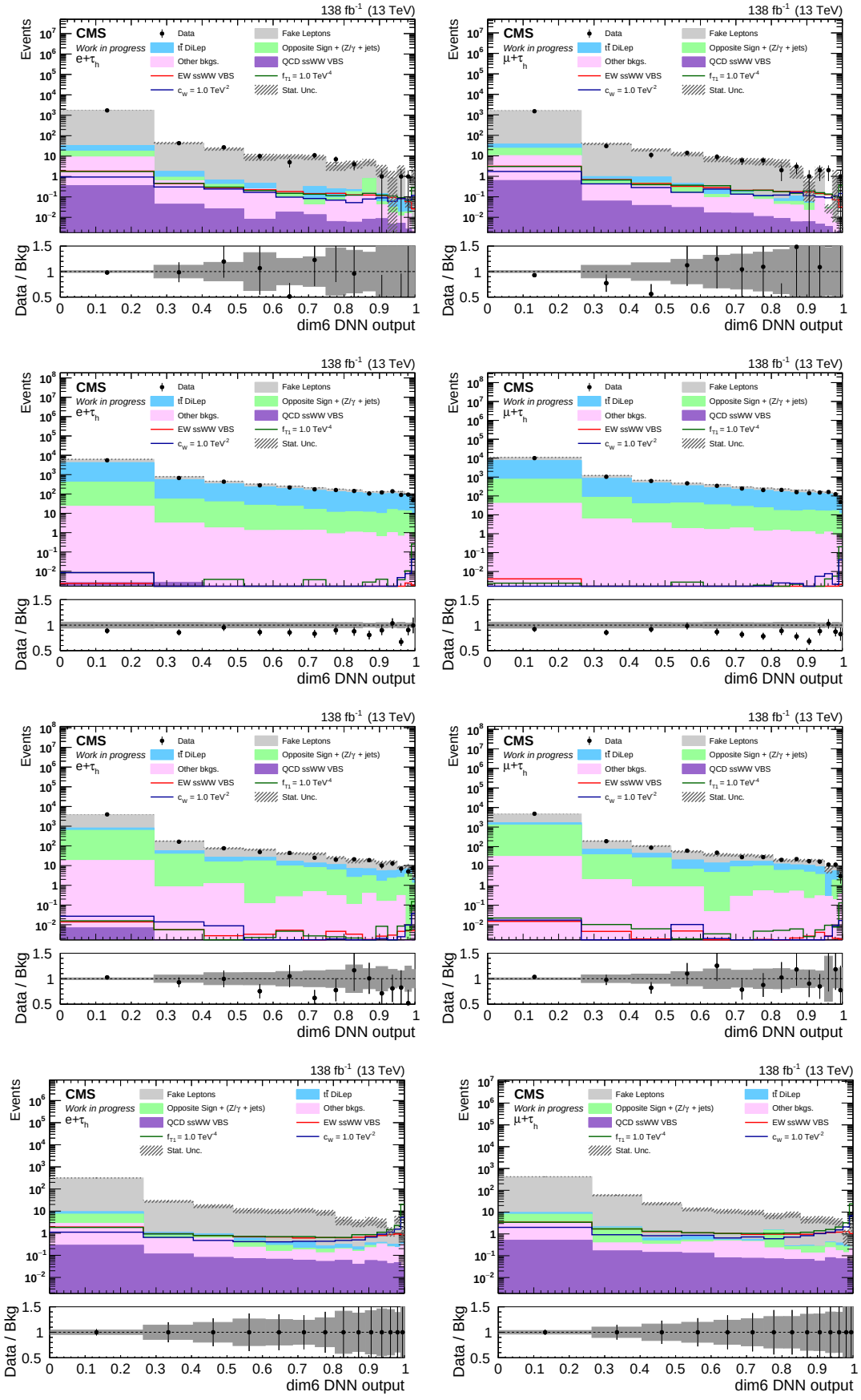


Figure B.1.26: Prefit SM DNN distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels. broad signal-flattening binning is used.

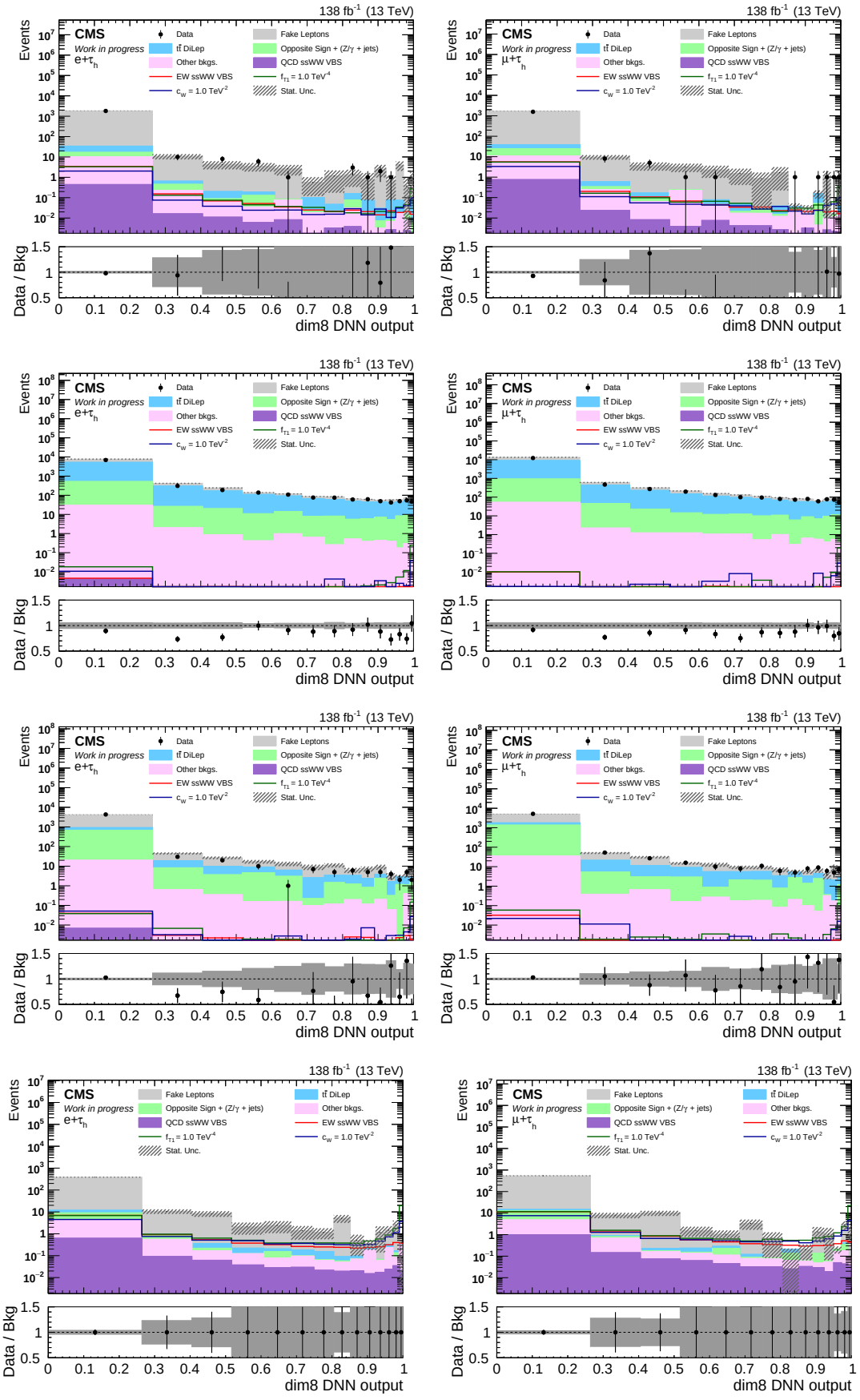


Figure B.1.27: Prefit SM DNN distributions in the three CRs (from top to bottom, Fakes, $t\bar{t}$, Opposite Sign) and in the SR (bottom row) for the Run II samples in the $e\tau_h$ (left) and $\mu\tau_h$ (right) channels. broad signal-flattening binning is used.

B.2 Linear and quadratic terms of the SMEFT operators

In the following, the M_{o1} distributions for the "linear" and "quadratic" contributions for each operator are shown, separately for the $e + \tau_h$ and $\mu + \tau_h$ channels.

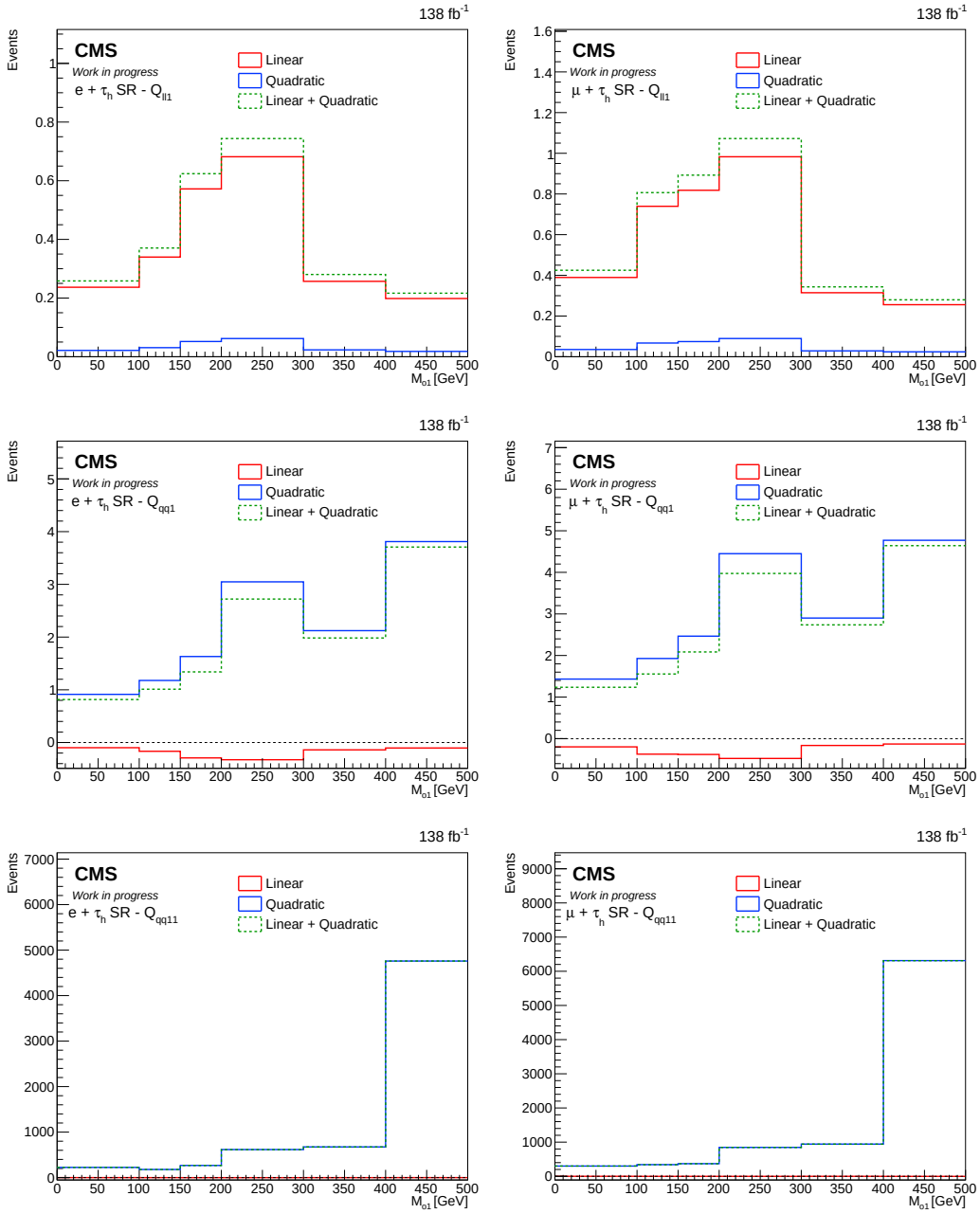


Figure B.2.1: M_{o1} distributions for the "linear" and "quadratic", together with their sum, for the operator $Q_{ll}^{(1)}, Q_{qq}^{(1)}, Q_{qq}^{(1,1)}$ with the full Run II dataset.

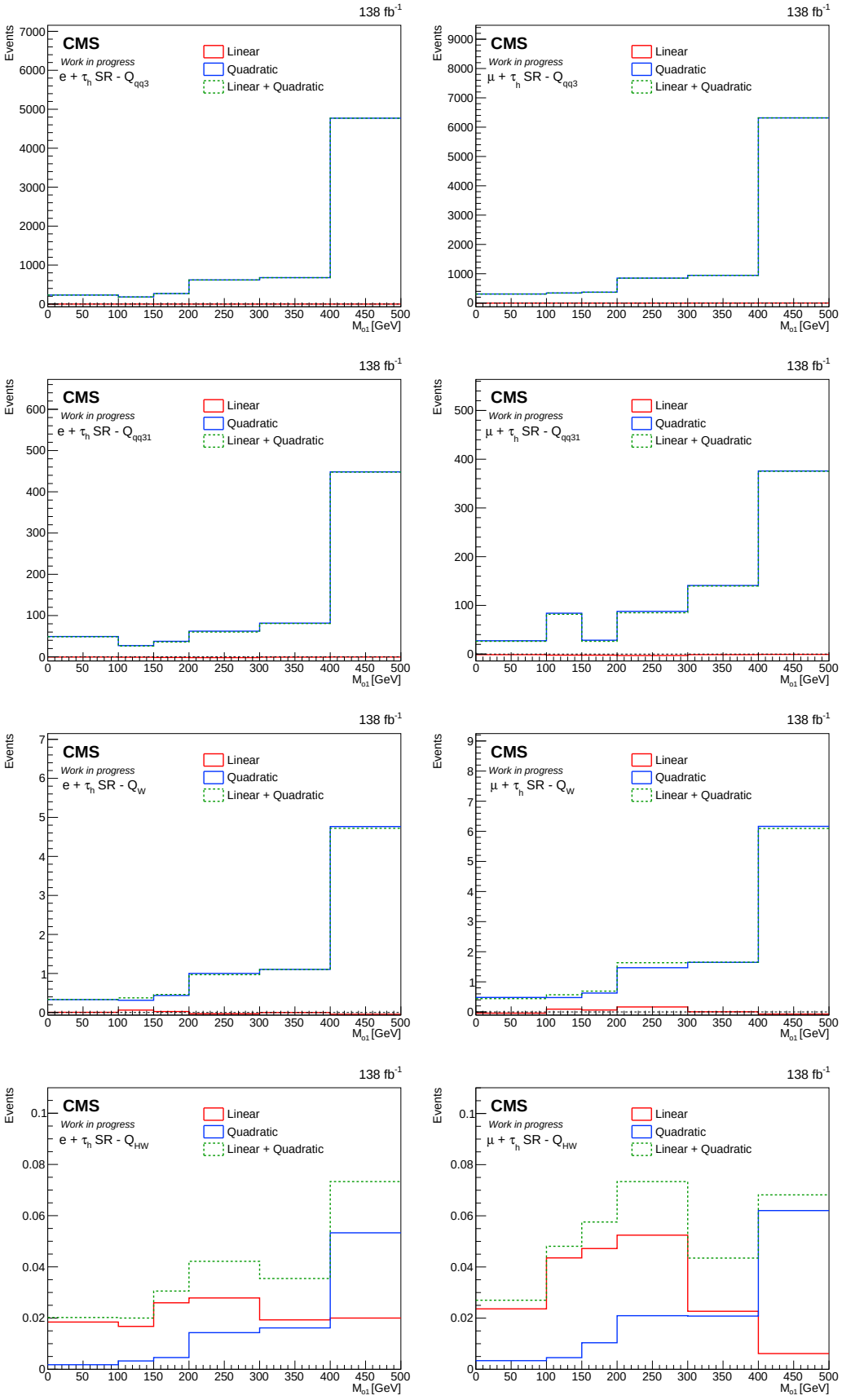


Figure B.2.2: M_{o1} distributions for the "linear" and "quadratic", together with their sum, for the operator $Q_{qq}^{(3)}, Q_{qq}^{(3,1)}, Q_W, Q_{HW}$ with the full Run II dataset.

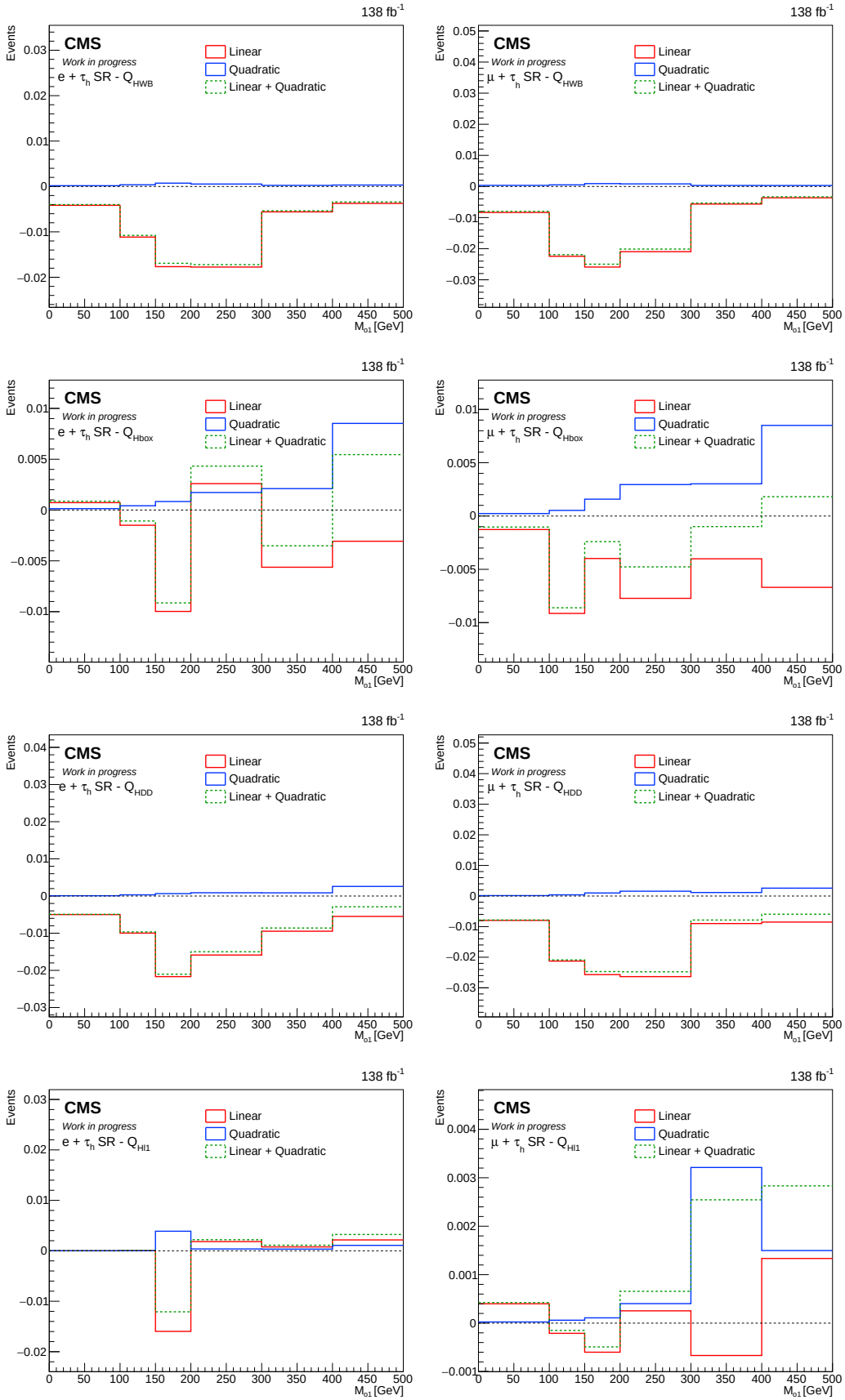


Figure B.2.3: M_{o1} distributions for the "linear" and "quadratic", together with their sum, for the operator $Q_{HWB}, Q_{H\Box}, Q_{HD}, Q_{Hl}^{(1)}$ with the full Run II dataset.

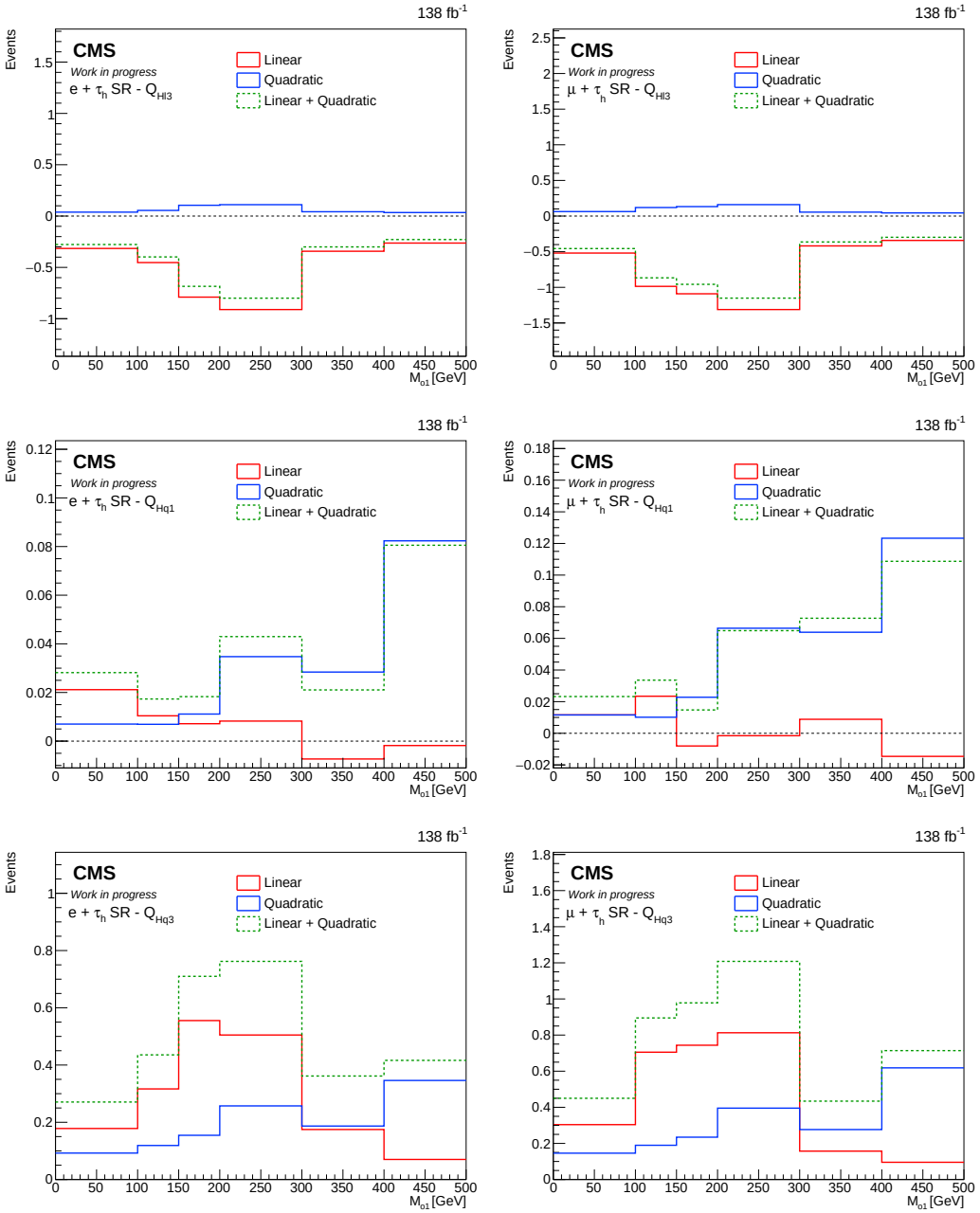


Figure B.2.4: M_{o1} distributions for the "linear" and "quadratic", together with their sum, for the operator $Q_{Hl}^{(3)}, Q_{Hq}^{(1)}, Q_{Hq}^{(3)}$ with the full Run II dataset.

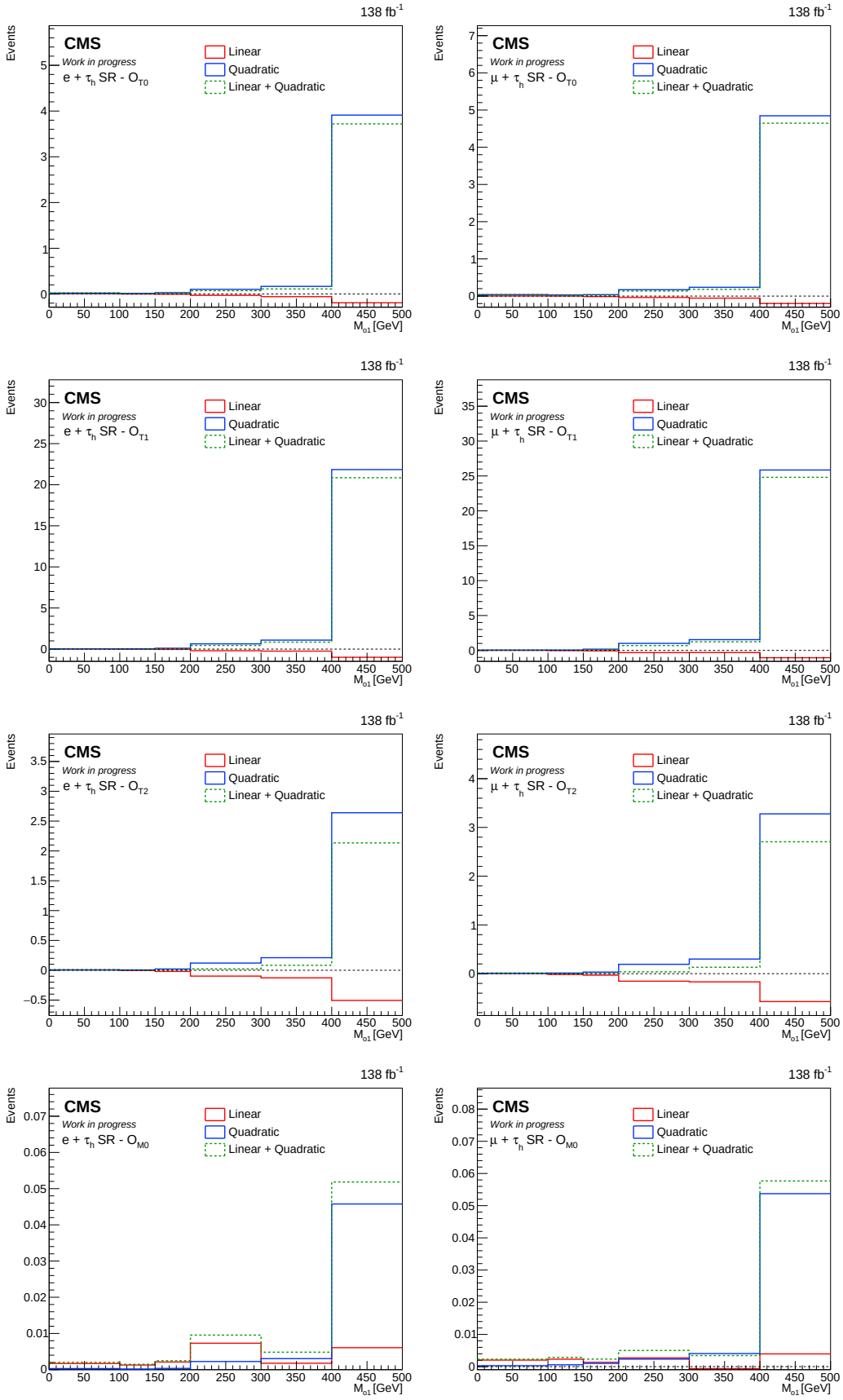


Figure B.2.5: M_{o1} for the "linear" and "quadratic", together with their sum, for the operator $\mathcal{O}_{T0}, \mathcal{O}_{T1}, \mathcal{O}_{T2}, \mathcal{O}_{M0}$ with the full Run II dataset.

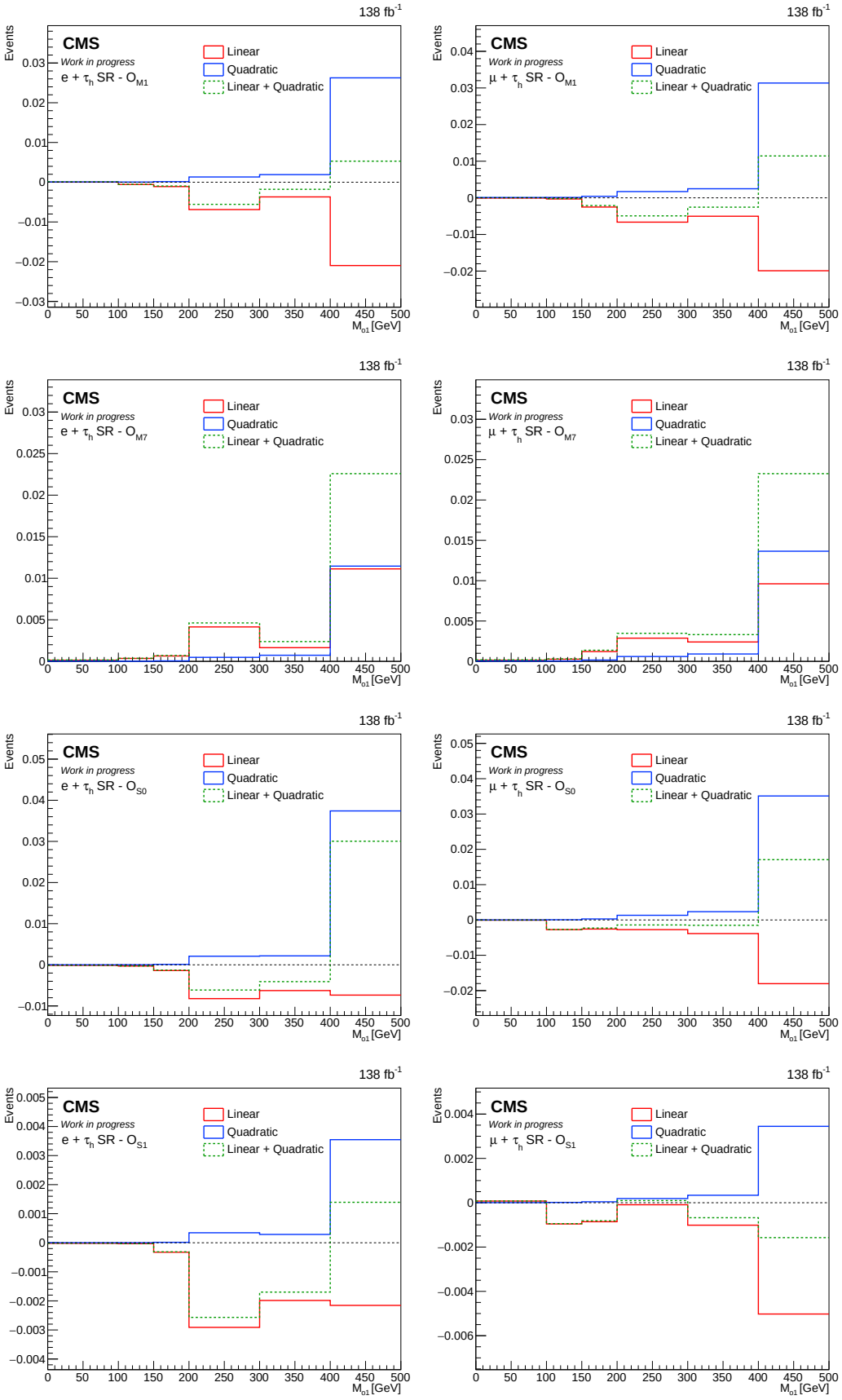


Figure B.2.6: M_{o1} for the "linear" and "quadratic", together with their sum, for the operator $\mathcal{O}_{M1}, \mathcal{O}_{M7}, \mathcal{O}_{S0}, \mathcal{O}_{S1}$ with the full Run II dataset.

B.3 Comparison between the M_{o1} distributions for the SM ssWW VBS signal and SMEFT operators with Wilson coefficients equals to expected limits

In the following, the M_{o1} distributions for each operator under study are shown according to the Eq. 7.6, using for the Wilson coefficients the values corresponding to the extremes of the 2σ ($\pm 2\sigma$) confidence level intervals extracted with dim-6/dim-8 DNN discriminator. For the dim-8 operators, similar distributions obtained with the results reported in [220] are shown for comparison. The distributions for the dim-6 operator $\mathcal{O}_{S2}$ are not shown, since are identical to the ones extracted for the dim-8 operator $\mathcal{O}_{S1}$.

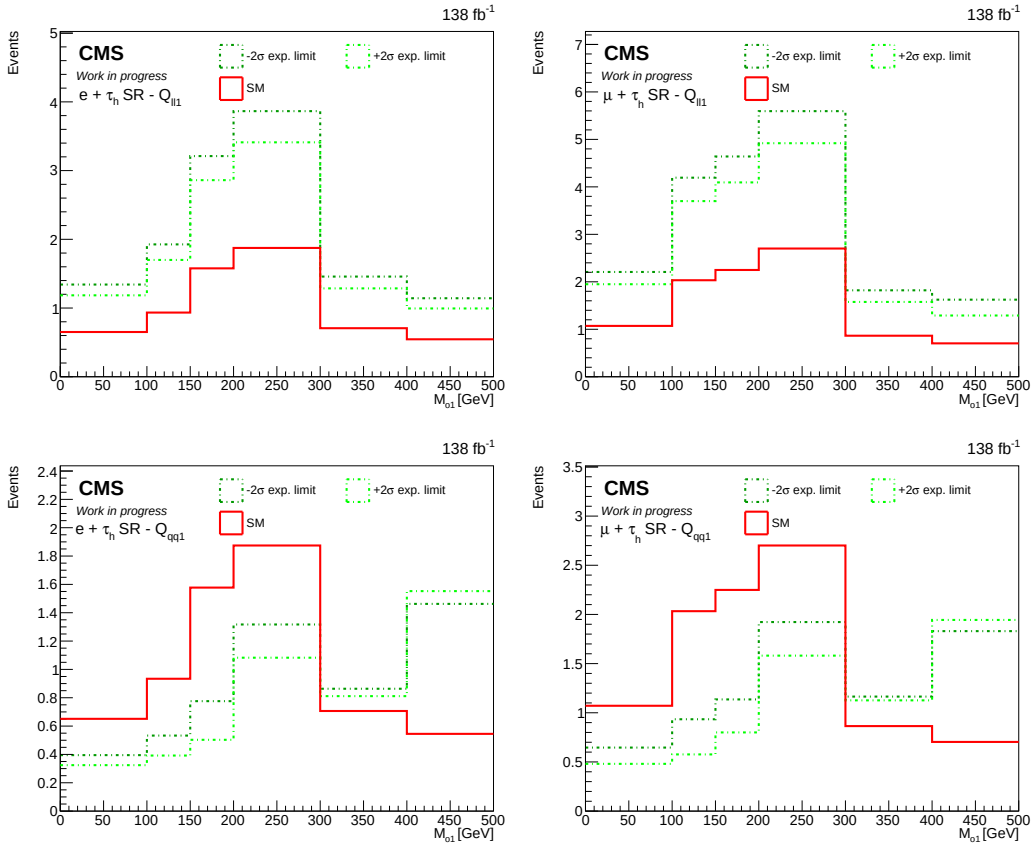


Figure B.3.1: M_{o1} distributions in SR for EW ssWW VBS (red) and the SMEFT dim-6 operators $Q_{ll}^{(1)}, Q_{qq}^{(1)}$, when the Wilson coefficients equal the upper ($+2\sigma$) and lower (-2σ) extremes of 2σ CL intervals from this analysis (green) reported in Tab. 7.10.

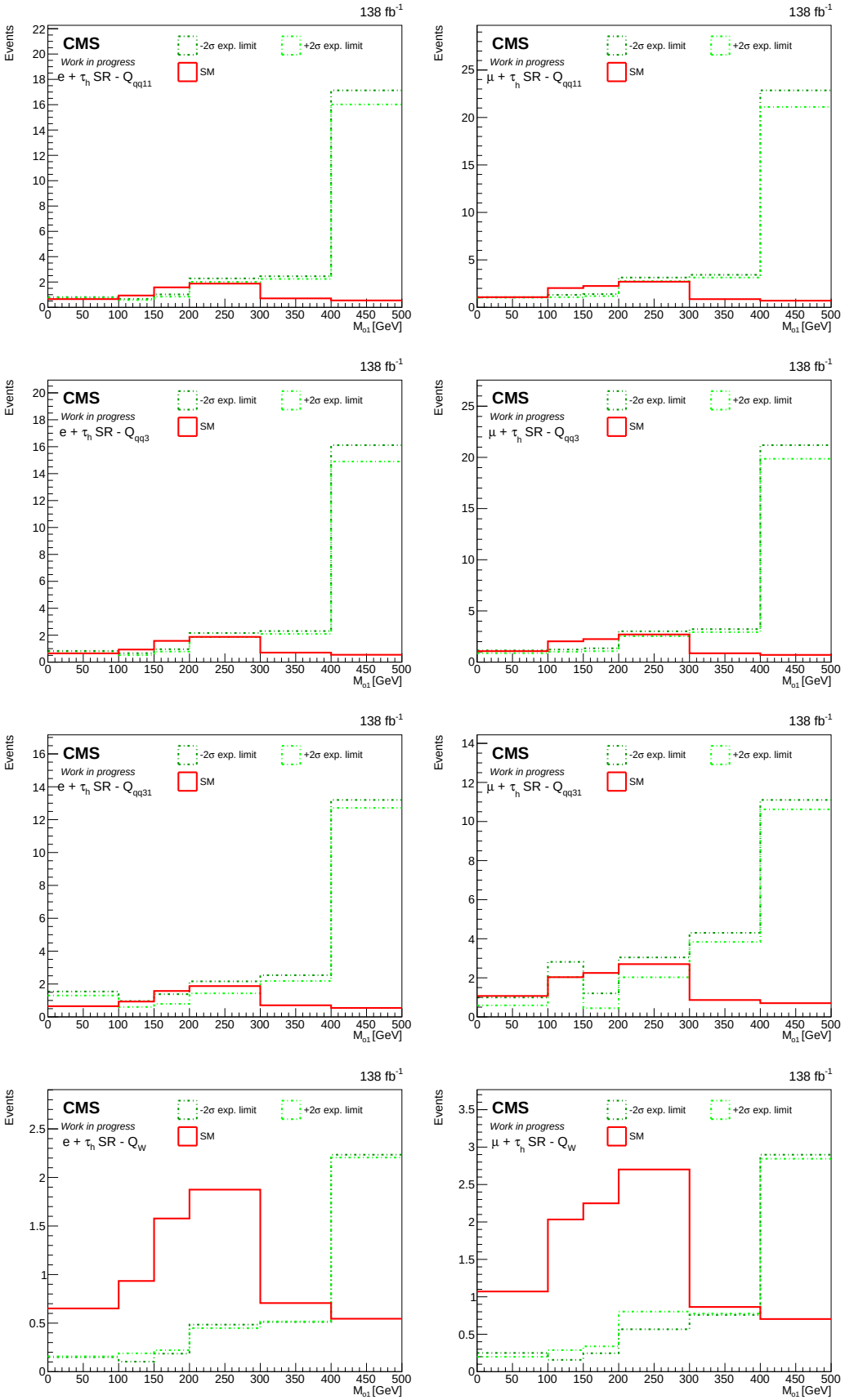


Figure B.3.2: M_{o1} distributions in SR for EW ssWW VBS (red) and the SMEFT dim-6 operators $Q_{qq}^{(1,1)}, Q_{qq}^{(3)}, Q_{qq}^{(3,1)}, Q_W$, when the Wilson coefficients equal the upper ($+2\sigma$) and lower (-2σ) extremes of 2σ CL intervals from this analysis (green) reported in Tab. 7.10.

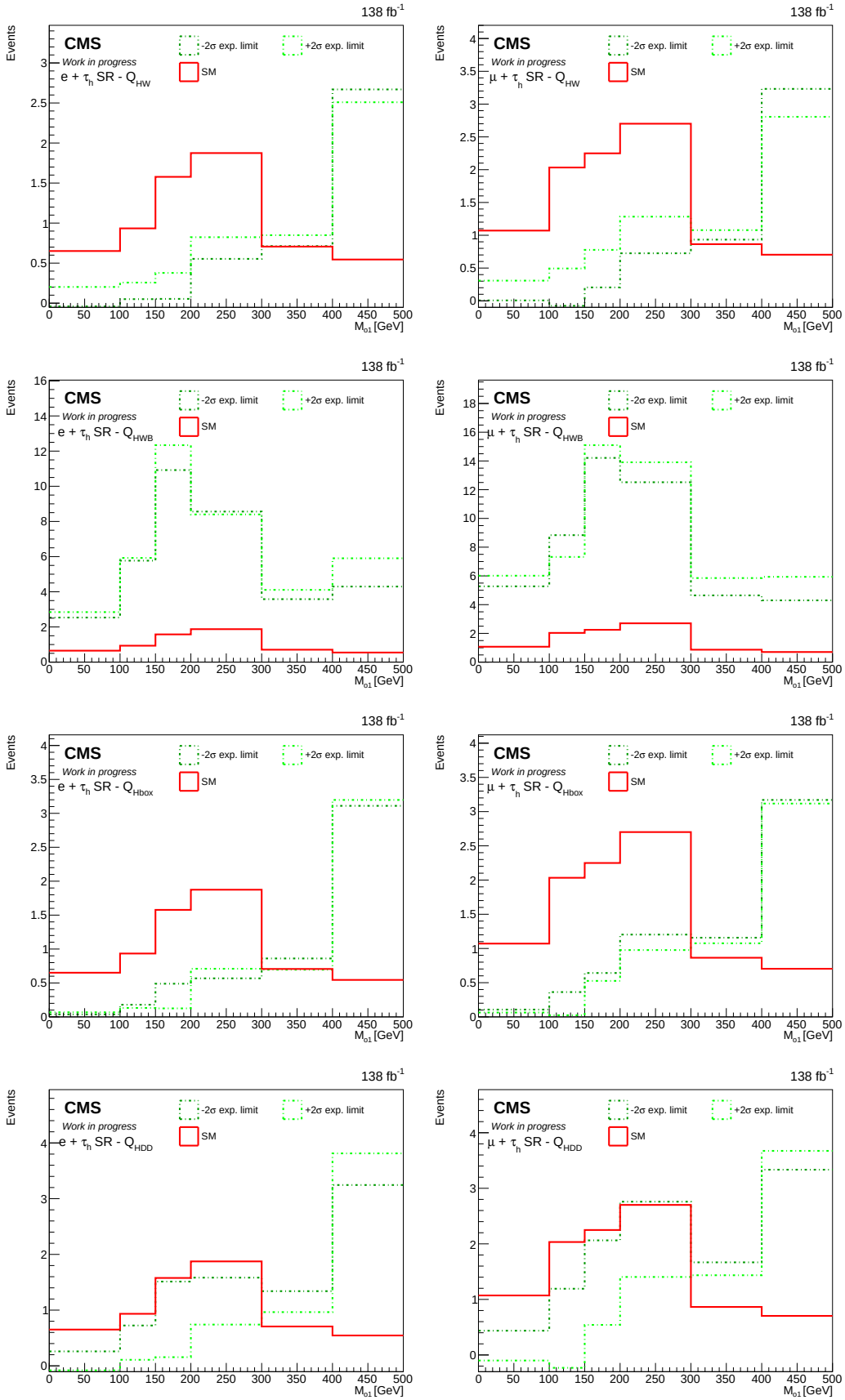


Figure B.3.3: M_{o1} distributions in SR for EW ssWW VBS (red) and the SMEFT dim-6 operators $Q_{HW}, Q_{HWB}, Q_{H\Box}, Q_{HD}$, when the Wilson coefficients equal the upper ($+2\sigma$) and lower (-2σ) extremes of 2σ CL intervals from this analysis (green) reported in Tab. 7.10.

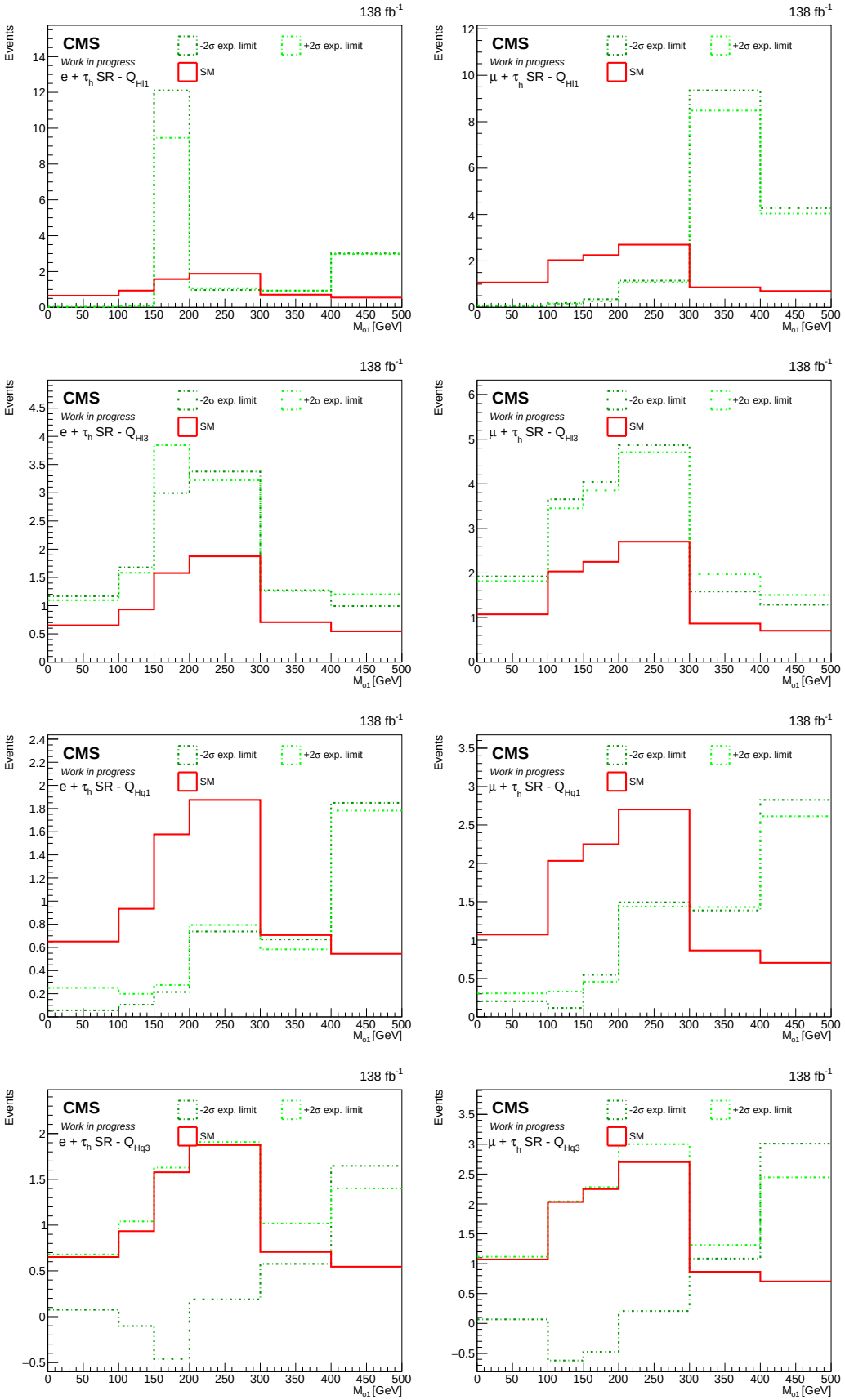


Figure B.3.4: M_{o1} distributions in SR for EW ssWW VBS (red) and the SMEFT dim-6 operators $Q_{Hl}^{(1)}, Q_{Hl}^{(3)}, Q_{Hq}^{(1)}, Q_{Hq}^{(3)}$, when the Wilson coefficients equal the upper ($+2\sigma$) and lower (-2σ) extremes of 2σ CL intervals from this analysis (green) reported in Tab. 7.10.

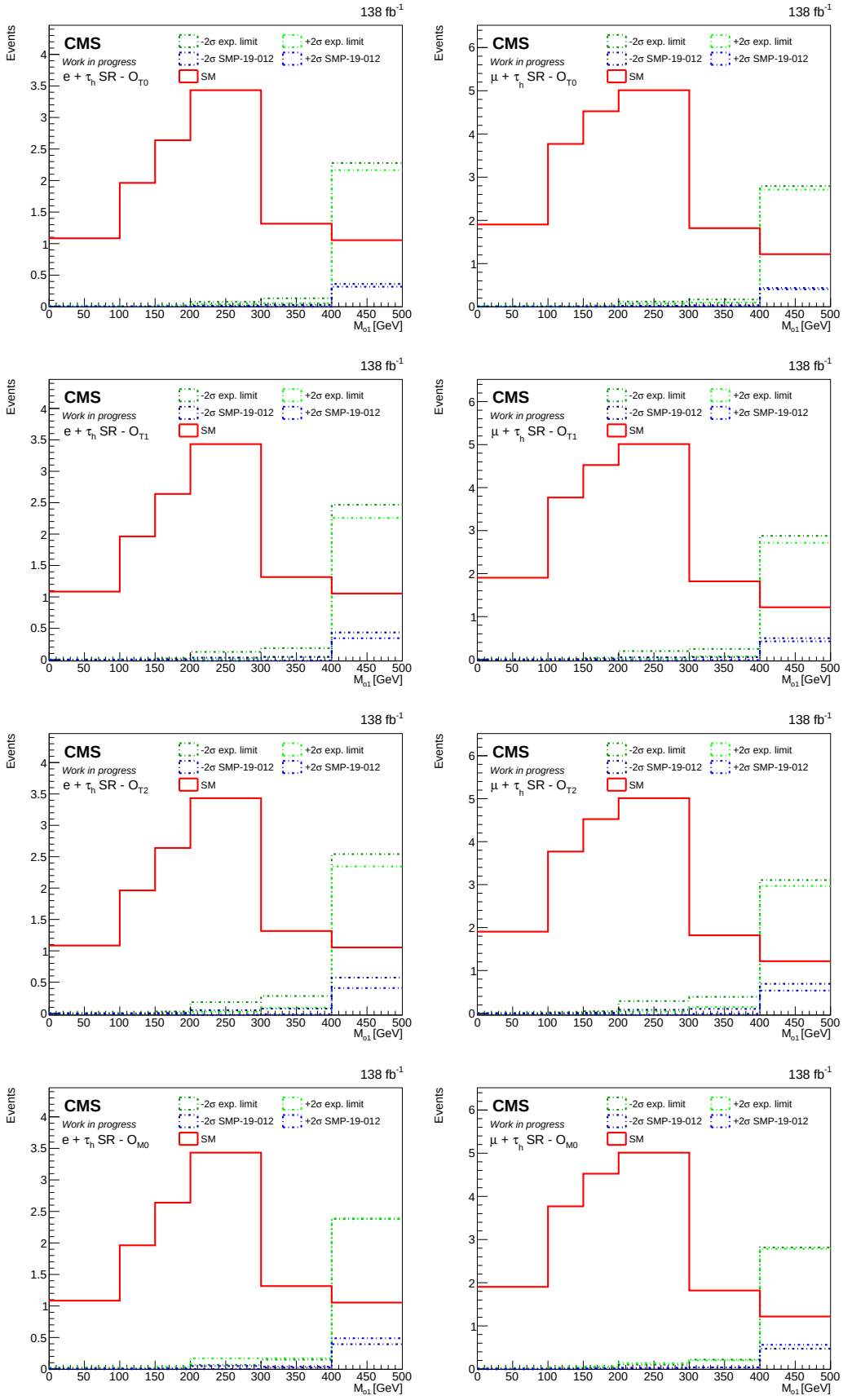


Figure B.3.5: M_{o1} distributions in SR for EW ssWW VBS (red) and the SMEFT dim-8 operators $\mathcal{O}_{T0}, \mathcal{O}_{T1}, \mathcal{O}_{T2}, \mathcal{O}_{M0}$, when the Wilson coefficients equal the upper ($+2\sigma$) and lower (-2σ) extremes of 2σ CL intervals from this analysis (green) and [220] (blue) reported in Tab. 7.10.

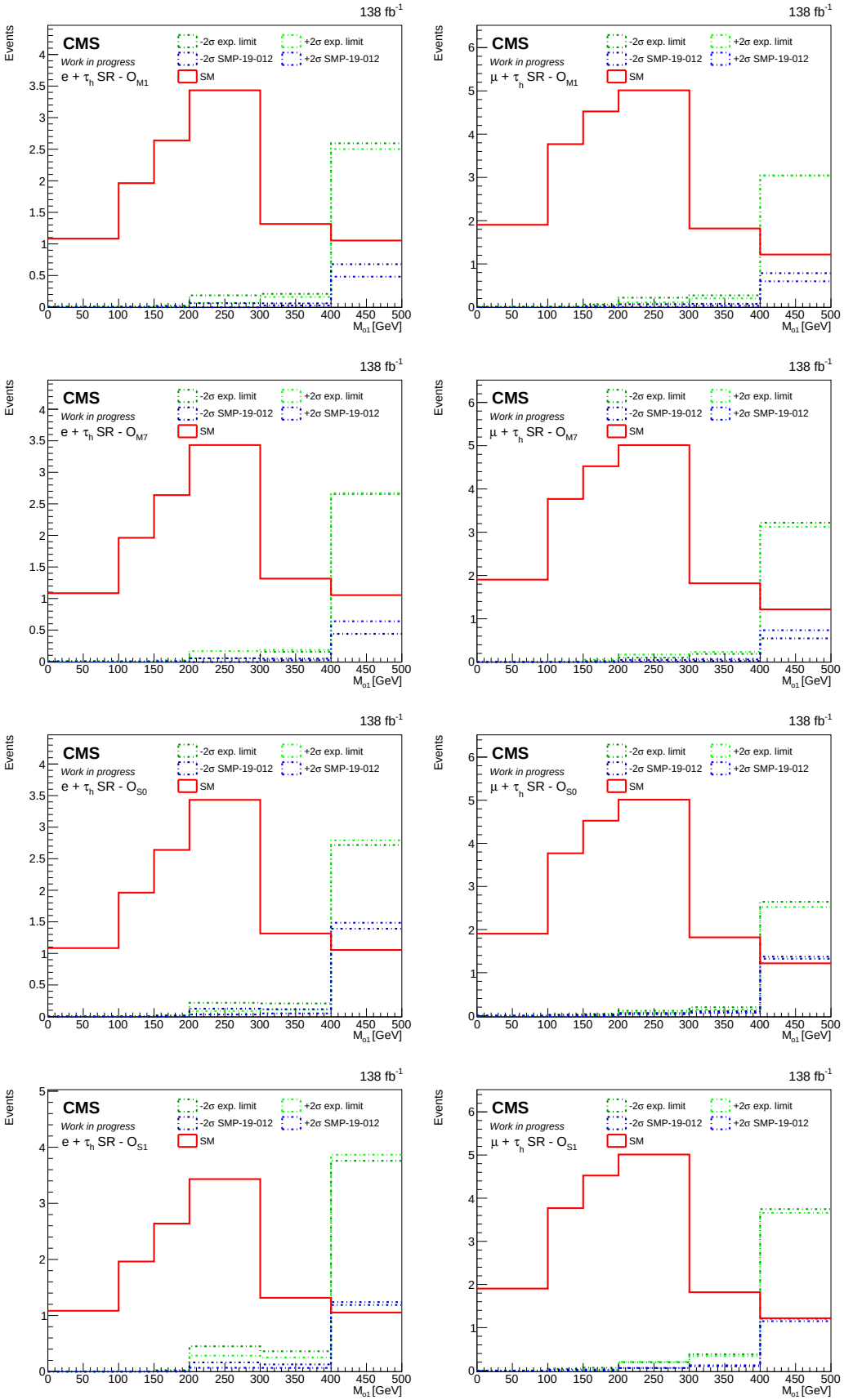


Figure B.3.6: M_{o1} distributions in SR for EW ssWW VBS (red) and the SMEFT dim-8 operators $\mathcal{O}_{M1}, \mathcal{O}_{M7}, \mathcal{O}_{S0}, \mathcal{O}_{S1}$, when the Wilson coefficients equal the upper ($+2\sigma$) and lower (-2σ) extremes of 2σ CL intervals from this analysis (green) and [220] (blue) reported in Tab. 7.10.

B.4 Full list of two-dimensional constraints on SMEFT operator pairs

In the following, the preliminary results for the two-dimensional fits for all the possible SMEFT operator pairs are shown as the expected two-dimensional distributions of $-2\ln\Delta\mathscr{L}$.

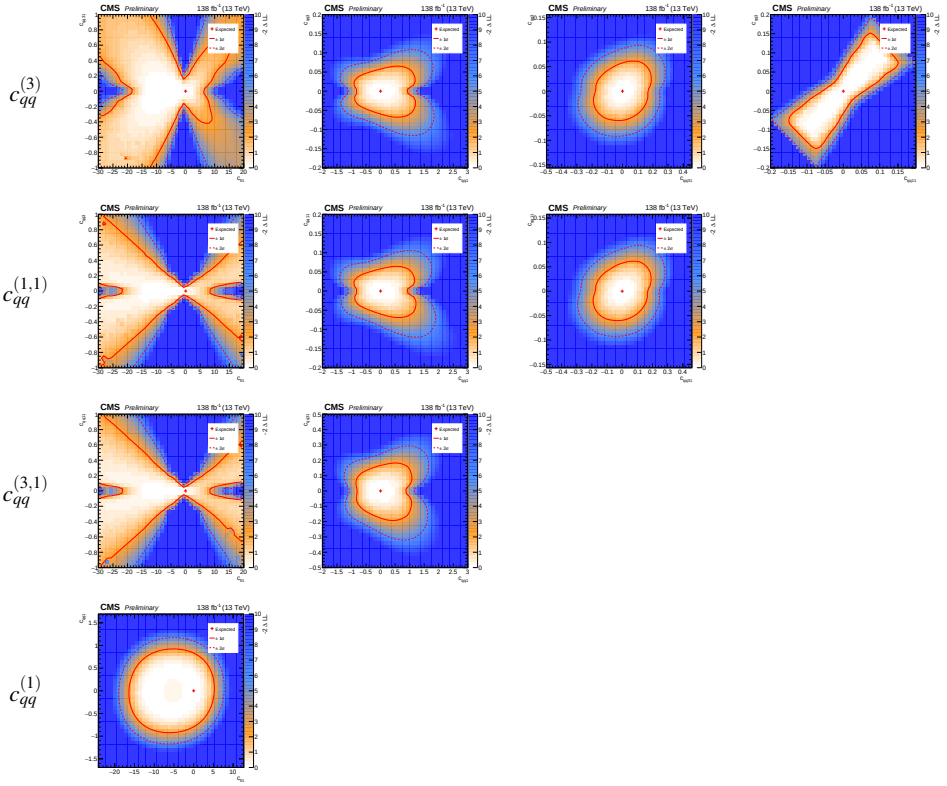


Table B.4.1: Grid of the expected 2D $-2\ln\Delta\mathscr{L}$ as functions of the reported dim-6 Wilson coefficient pairs. 1σ and 2σ CL contours are superimposed

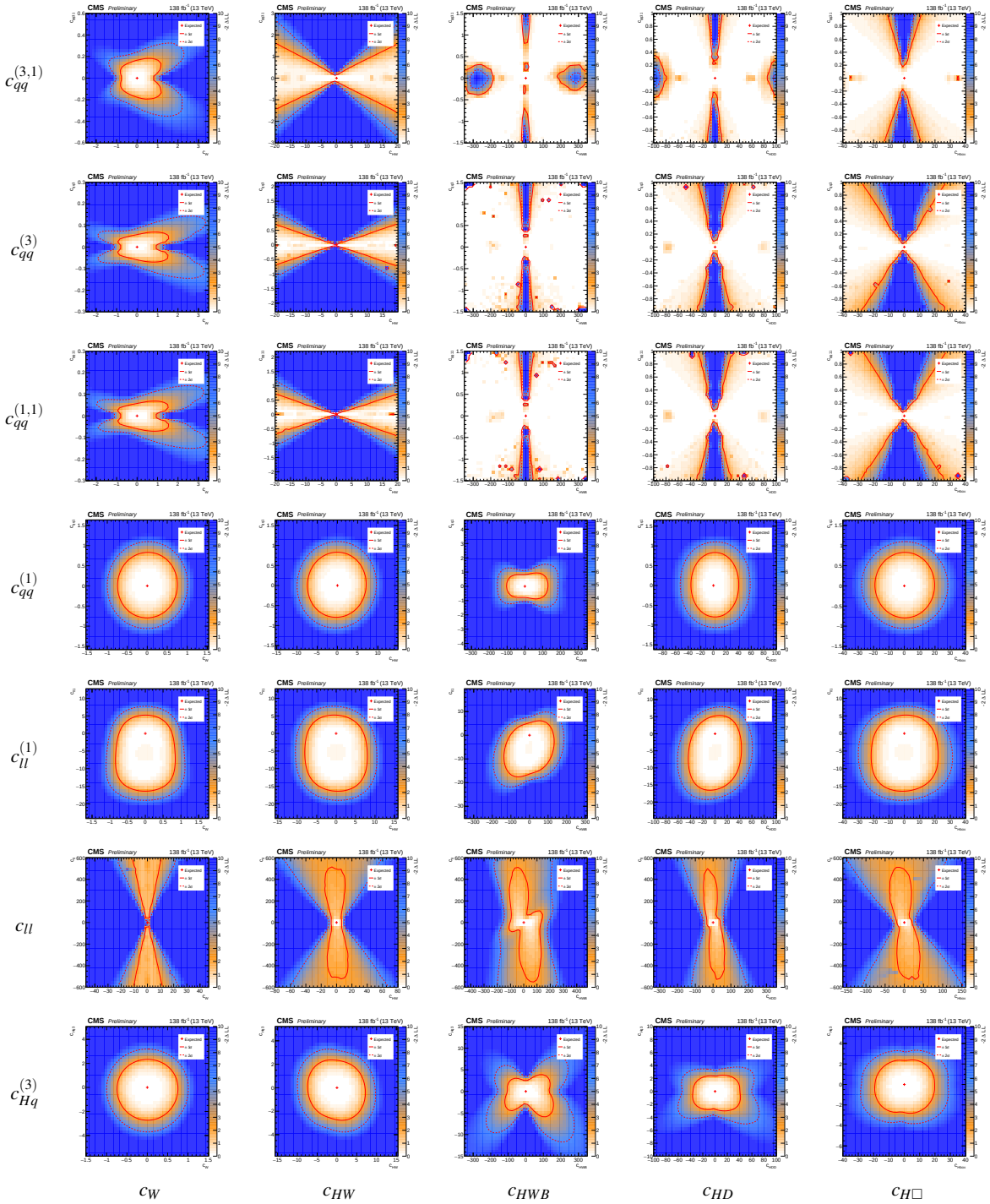


Table B.4.2: Grid of the expected 2D $-2\ln\Delta\mathscr{L}$ as functions of the reported dim-6 Wilson coefficient pairs. 1σ and 2σ CL contours are superimposed

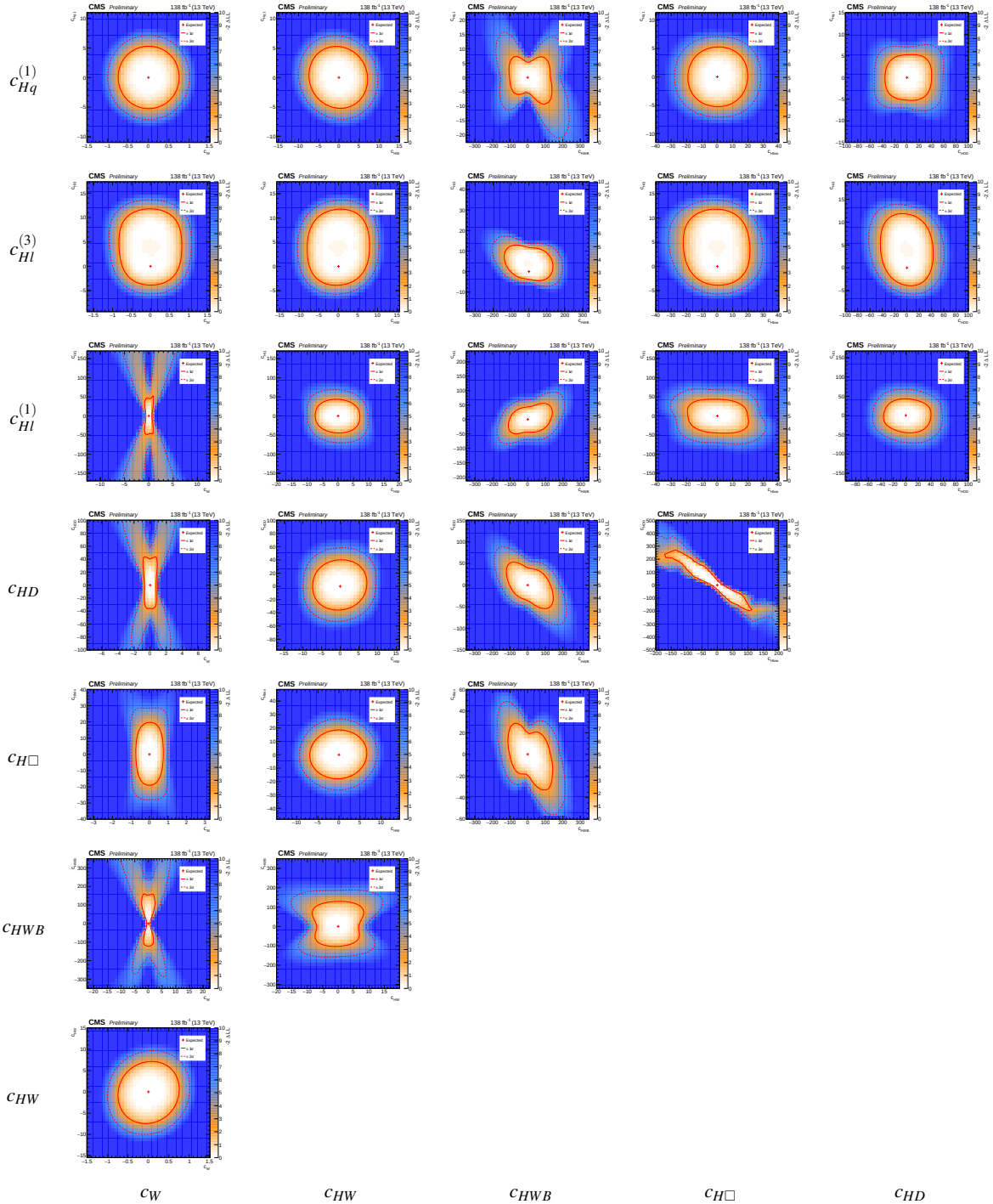


Table B.4.3: Grid of the expected 2D $-2\ln\Delta\mathscr{L}$ as functions of the reported dim-6 Wilson coefficient pairs. 1σ and 2σ CL contours are superimposed

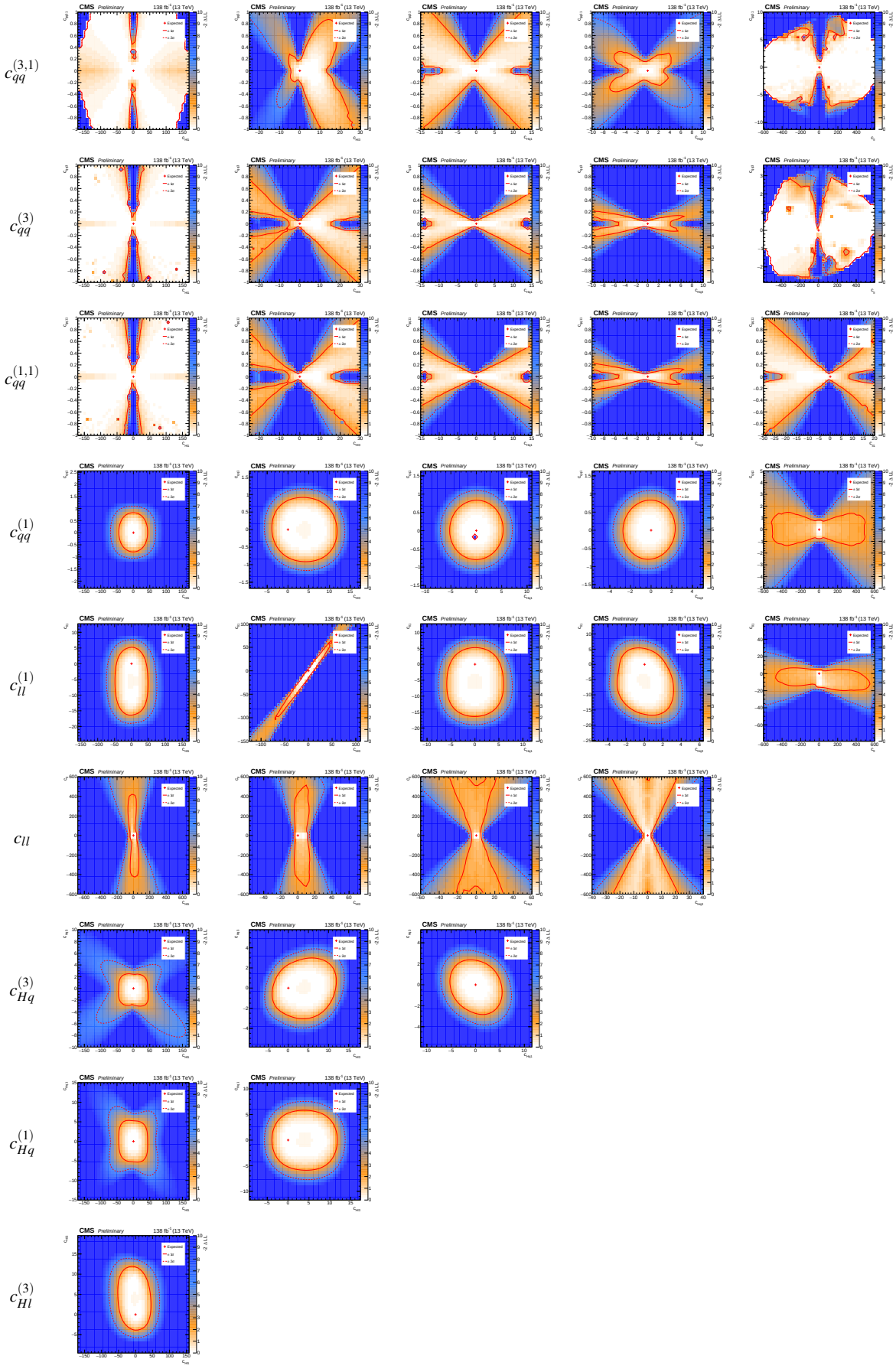


Table B.4.4: Grid of the expected 2D $-2\ln\Delta\mathscr{L}$ as functions of the reported dim-6 Wilson coefficient pairs. 1σ and 2σ CL contours are superimposed

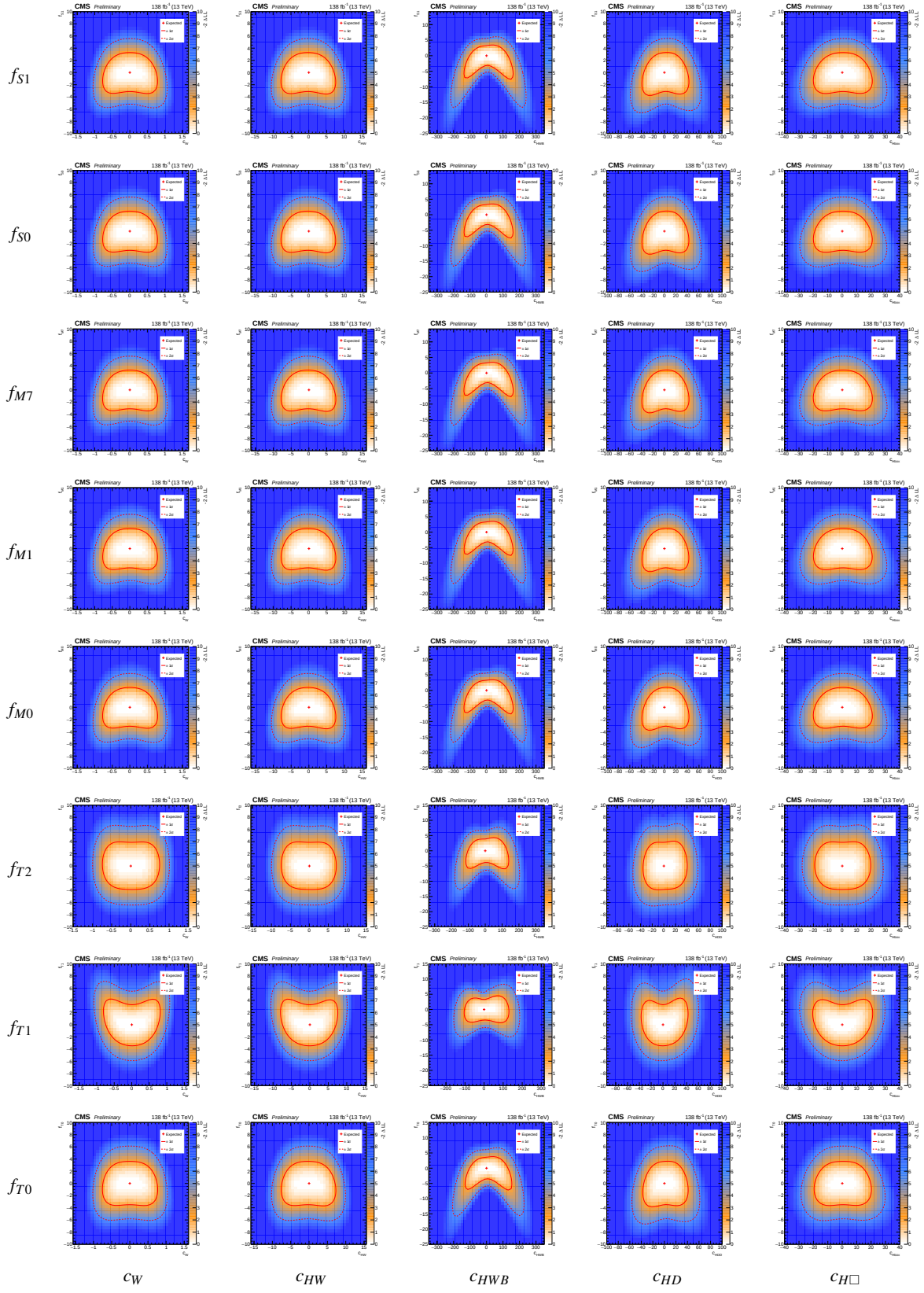


Table B.4.5: Grid of the expected 2D $-2\ln\Delta\mathscr{L}$ as functions of the reported dim-6/dim-8 Wilson coefficient pairs. 1σ and 2σ CL contours are superimposed

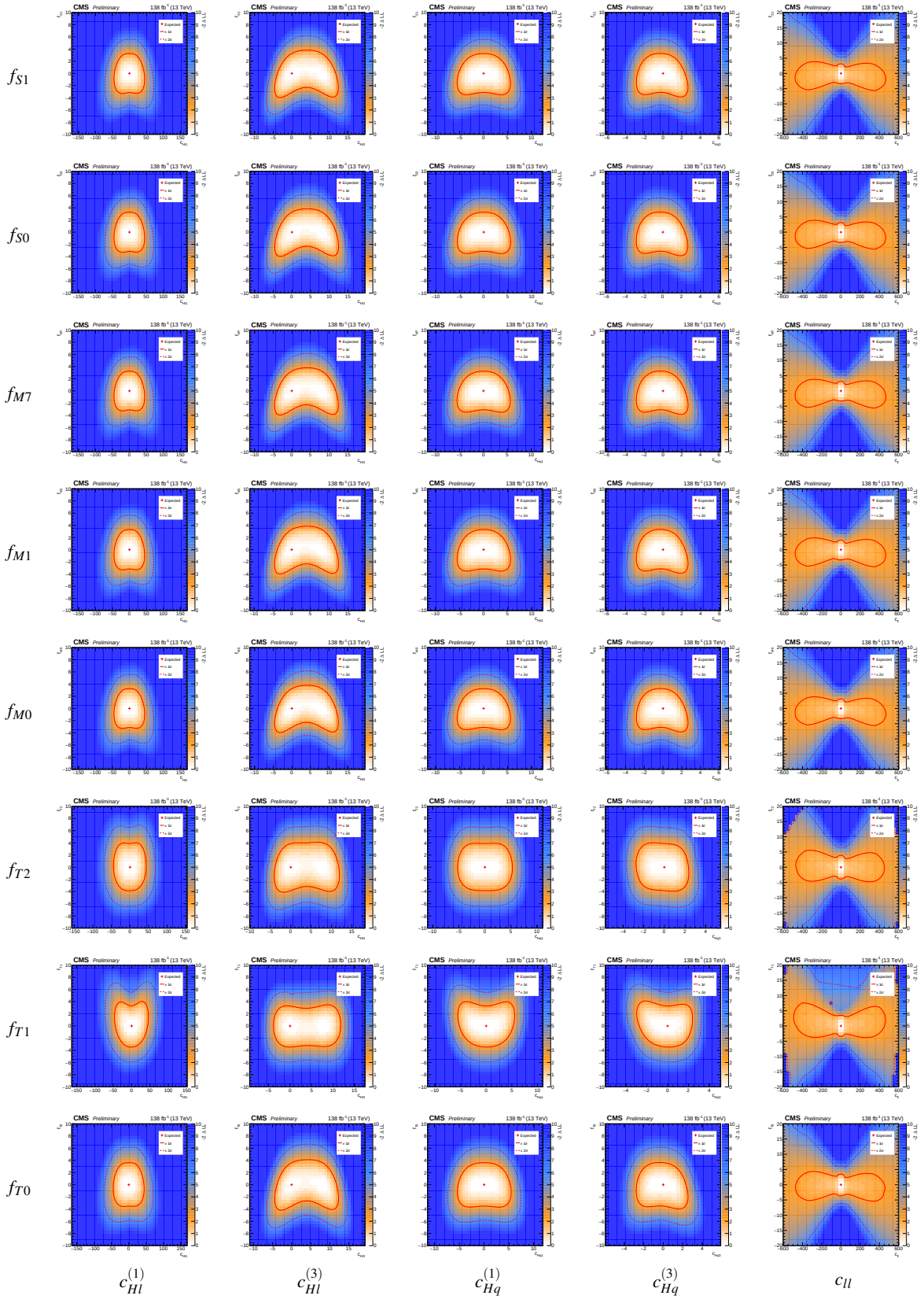


Table B.4.6: Grid of the expected 2D $-2\ln\Delta\mathscr{L}$ as functions of the reported dim-6/dim-8 Wilson coefficient pairs. 1σ and 2σ CL contours are superimposed

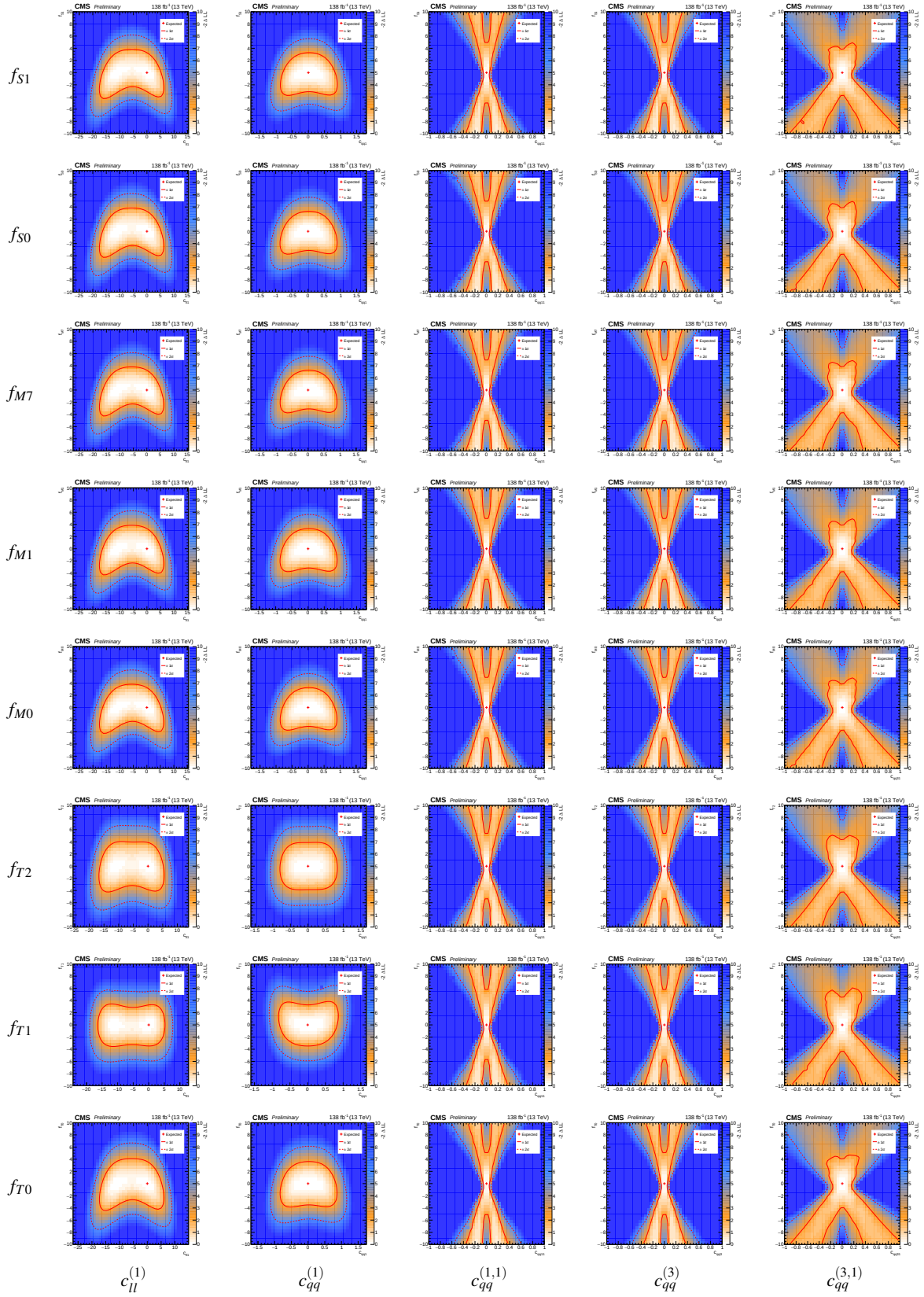


Table B.4.7: Grid of the expected 2D $-2\ln\Delta\mathscr{L}$ as functions of the reported dim-6/dim-8 Wilson coefficient pairs. 1σ and 2σ CL contours are superimposed

Bibliography

[1] *Search for W' bosons decaying to a top and a bottom quark in leptonic final states at* $\sqrt{s} = 13\text{TeV}$. Tech. rep. Geneva: CERN, 2023. URL: http://cds.cern.ch/record/2853340.

[2] S. L. Glashow. "Partial Symmetries of Weak Interactions". In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

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

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

[5] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 321–323. DOI: 10.1103/PhysRevLett.13.321.

[6] Peter W. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Phys. Lett.* 12 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9.

[7] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 585–587. DOI: 10.1103/PhysRevLett.13.585.

[8] G. Arnison et al. "Experimental Observation of Isolated Large Transverse Energy Electrons with Associated Missing Energy at $\sqrt{s} = 540$ GeV". In: *Phys. Lett. B* 122 (1983), pp. 103–116. DOI: 10.1016/0370-2693(83)91177-2.

[9] M. Banner et al. "Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN anti-p p Collider". In: *Phys. Lett. B* 122 (1983), pp. 476–485. DOI: 10.1016/0370-2693(83)91605-2.

[10] P. Bagnaia et al. "Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p}p$ Collider". In: *Phys. Lett. B* 129 (1983), pp. 130–140. DOI: 10.1016/0370-2693(83)90744-X.

[11] S. Abachi et al. "Observation of the top quark". In: *Phys. Rev. Lett.* 74 (1995), pp. 2632–2637. DOI: 10.1103/PhysRevLett.74.2632. arXiv: hep-ex/9503003.

[12] 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].

[13] Serguei Chatrchyan et al. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[14] G. Arnison et al. “Experimental Observation of Lepton Pairs of Invariant Mass Around 95 GeV/c^2 at the CERN SPS Collider”. In: *Phys. Lett.* 126B (1983). [,7.55(1983)], pp. 398–410. DOI: 10.1016/0370-2693(83)90188-0.

[15] Andreas S. Kronfeld et al. “Lattice QCD and Particle Physics”. In: (July 2022). arXiv: 2207.07641 [hep-lat].

[16] 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. URL: http://www.sciencedirect.com/science/article/pii/0029558261904692.

[17] A. Salam and N. Svartholm. “Elementary Particle Theory. Relativistic Groups and Analyticity. Proceedings of the eighth Nobel Symposium, Aspenäsgården, Lerum, Sweden, May 1968. Ed. Interscience (Wiley), New York, and Almqvist and Wiksell, Stockholm, 1969. 400 pp., il...” In: *Science* 168.3936 (1970), pp. 1196–1197. ISSN: 0036-8075. DOI: 10.1126/science.168.3936.1196-a. eprint: https://science.sciencemag.org/content/168/3936/1196.2.full.pdf. URL: https://science.sciencemag.org/content/168/3936/1196.2.

[18] R. L. Workman and Others. “Review of Particle Physics”. In: *PTEP* 2022 (2022), p. 083C01. DOI: 10.1093/ptep/ptac097.

[19] Yung-Su Tsai. “Decay Correlations of Heavy Leptons in $e^+ + e^- \to l^+ + l^-$”. In: *Phys. Rev. D* 4 (9 1971), pp. 2821–2837. DOI: 10.1103/PhysRevD.4.2821. URL: https://link.aps.org/doi/10.1103/PhysRevD.4.2821.

[20] Yung-Su Tsai. “Erratum: Decay correlations of heavy leptons in $e^+ + e^- \to l^+ + l^-$”. In: *Phys. Rev. D* 13 (3 1976), pp. 771–771. DOI: 10.1103/PhysRevD.13.771. URL: https://link.aps.org/doi/10.1103/PhysRevD.13.771.

[21] David London and Joaquim Matias. “B Flavour Anomalies: 2021 Theoretical Status Report”. In: *Ann. Rev. Nucl. Part. Sci.* 72 (2022), pp. 37–68. DOI: 10.1146/annurev-nucl-102020-090209. arXiv: 2110.13270 [hep-ph].

[22] P. A. Zyla et al. “Review of Particle Physics”. In: *PTEP* 2020.8 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.

[23] R. Sekhar Chivukula and Howard Georgi. “Composite Technicolor Standard Model”. In: *Phys. Lett. B* 188 (1987), pp. 99–104. DOI: 10.1016/0370-2693(87)90713-1.

[24] W. J. Marciano and A. Sirlin. “Electroweak Radiative Corrections to tau Decay”. In: *Phys. Rev. Lett.* 61 (1988), pp. 1815–1818. DOI: 10.1103/PhysRevLett.61.1815.

[25] Georges Aad et al. “Test of the universality of τ and μ lepton couplings in W-boson decays with the ATLAS detector”. In: *Nature Phys.* 17.7 (2021), pp. 813–818. DOI: 10.1038/s41567-021-01236-w. arXiv: 2007.14040 [hep-ex].

[26] Yasmine Sara Amhis et al. “Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018”. In: *Eur. Phys. J. C* 81.3 (2021), p. 226. DOI: 10.1140/epjc/s10052-020-8156-7. arXiv: 1909.12524 [hep-ex].

[27] A. M. Baldini et al. “Search for the lepton flavour violating decay $\mu^+ \to e^+\gamma$ with the full dataset of the MEG experiment”. In: *Eur. Phys. J. C* 76.8 (2016), p. 434. DOI: 10.1140/epjc/s10052-016-4271-x. arXiv: 1605.05081 [hep-ex].

[28] Alejandro Celis, Vincenzo Cirigliano, and Emilie Passemar. “Model-discriminating power of lepton flavor violating τ decays”. In: *Phys. Rev. D* 89.9 (2014), p. 095014. DOI: 10.1103/PhysRevD.89.095014. arXiv: 1403.5781 [hep-ph].

[29] Tomas Husek, Kevin Monsalvez-Pozo, and Jorge Portoles. “Lepton-flavour violation in hadronic tau decays and $\mu - \tau$ conversion in nuclei”. In: *JHEP* 01 (2021), p. 059. DOI: 10.1007/JHEP01(2021)059. arXiv: 2009.10428 [hep-ph].

[30] 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].

[31] Roberto Covarelli, Mathieu Pellen, and Marco Zaro. “Vector-Boson scattering at the LHC: Unraveling the electroweak sector”. In: *Int. J. Mod. Phys. A* 36.16 (2021), p. 2130009. DOI: 10.1142/S0217751X2130009X. arXiv: 2102.10991 [hep-ph].

[32] Benedikt Biedermann, Ansgar Denner, and Mathieu Pellen. “Complete NLO corrections to W^+W^+ scattering and its irreducible background at the LHC”. In: *JHEP* 10 (2017), p. 124. DOI: 10.1007/JHEP10(2017)124. arXiv: 1708.00268 [hep-ph].

[33] Sally Dawson. “The effective W approximation”. In: *Nuclear Physics B* 249.1 (1985), pp. 42–60. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(85)90038-0. URL: https://www.sciencedirect.com/science/article/pii/0550321385900380.

[34] M.J. Duncan, G.L. Kane, and W.W. Repko. “WW physics at future colliders”. In: *Nuclear Physics B* 272.3 (1986), pp. 517–559. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(86)90234-8. URL: https://www.sciencedirect.com/science/article/pii/0550321386902348.

[35] R.N. Cahn and Sally Dawson. “Production of very massive Higgs bosons”. In: *Physics Letters B* 136.3 (1984), pp. 196–200. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(84)91180-8. URL: https://www.sciencedirect.com/science/article/pii/0370269384911808.

[36] I. Kuss and H. Spiesberger. “Luminosities for vector-boson-vector-boson scattering at high energy colliders”. In: *Phys. Rev. D* 53 (11 1996), pp. 6078–6093. DOI: 10.1103/PhysRevD.53.6078. URL: https://link.aps.org/doi/10.1103/PhysRevD.53.6078.

[37] E. Accomando, Ansgar Denner, and S. Pozzorini. “Logarithmic electroweak corrections to $e^+e^- \to \nu_e\bar{\nu}_e W^+W^-$”. In: *JHEP* 03 (2007), p. 078. DOI: 10.1088/1126-6708/2007/03/078. arXiv: hep-ph/0611289.

[38] Alessandro Ballestrero et al. “Precise predictions for same-sign W-boson scattering at the LHC”. In: *Eur. Phys. J. C* 78.8 (2018), p. 671. DOI: 10.1140/epjc/s10052-018-6136-y. arXiv: 1803.07943 [hep-ph].

[39] T. Figy, D. Zeppenfeld, and C. Oleari. “Next-to-leading order jet distributions for Higgs boson production via weak-boson fusion”. In: *Phys. Rev. D* 68 (7 2003), p. 073005. DOI: 10.1103/PhysRevD.68.073005. URL: https://link.aps.org/doi/10.1103/PhysRevD.68.073005.

[40] Benjamin W. Lee, C. Quigg, and H. B. Thacker. “Strength of Weak Interactions at Very High Energies and the Higgs Boson Mass”. In: *Phys. Rev. Lett.* 38 (16 1977), pp. 883–885. DOI: 10.1103/PhysRevLett.38.883. URL: https://link.aps.org/doi/10.1103/PhysRevLett.38.883.

[41] Benjamin W. Lee, C. Quigg, and H. B. Thacker. “Weak interactions at very high energies: The role of the Higgs-boson mass”. In: *Phys. Rev. D* 16 (5 1977), pp. 1519–1531. DOI: 10.1103/PhysRevD.16.1519. URL: https://link.aps.org/doi/10.1103/PhysRevD.16.1519.

[42] Michael S. Chanowitz and Mary K. Gaillard. “Multiple production of W and Z as a signal of new strong interactions”. In: *Physics Letters B* 142.1 (1984), pp. 85–90. ISSN: 0370-2693. DOI: https://doi.org/10.1016/0370-2693(84)91141-9. URL: https://www.sciencedirect.com/science/article/pii/0370269384911419.

[43] Michael S. Chanowitz and Mary K. Gaillard. “The TeV physics of strongly interacting W’s and Z’s”. In: *Nuclear Physics B* 261 (1985), pp. 379–431. ISSN: 0550-3213. DOI: https://doi.org/10.1016/0550-3213(85)90580-2. URL: https://www.sciencedirect.com/science/article/pii/0550321385905802.

[44] Alessandro Ballestrero, Ezio Maina, and Giovanni Pelliccioli. “W boson polarization in vector boson scattering at the LHC”. In: *JHEP* 03 (2018), p. 170. DOI: 10.1007/JHEP03(2018)170. arXiv: 1710.09339 [hep-ph].

[45] Alessandro Ballestrero, Ezio Maina, and Giovanni Pelliccioli. “Different polarization definitions in same-sign WW scattering at the LHC”. In: *Phys. Lett. B* 811 (2020), p. 135856. DOI: 10.1016/j.physletb.2020.135856. arXiv: 2007.07133 [hep-ph].

[46] Diogo Buarque Franzosi et al. “Automated predictions from polarized matrix elements”. In: *JHEP* 04 (2020), p. 082. DOI: 10.1007/JHEP04(2020)082. arXiv: 1912.01725 [hep-ph].

[47] Brando Bellazzini, Csaba Csáki, and Javi Serra. “Composite Higgses”. In: *Eur. Phys. J. C* 74.5 (2014), p. 2766. DOI: 10.1140/epjc/s10052-014-2766-x. arXiv: 1401.2457 [hep-ph].

[48] Giuliano Panico and Andrea Wulzer. *The Composite Nambu-Goldstone Higgs*. Vol. 913. Springer, 2016. DOI: 10.1007/978-3-319-22617-0. arXiv: 1506.01961 [hep-ph].

[49] Armen Tumasyan et al. “Evidence for WW/WZ vector boson scattering in the decay channel $\ell\nu$qq produced in association with two jets in proton-proton collisions at s=13 TeV”. In: *Phys. Lett. B* 834 (2022), p. 137438. DOI: 10.1016/j.physletb.2022.137438. arXiv: 2112.05259 [hep-ex].

[50] Michele Gallinaro et al. “Beyond the Standard Model in Vector Boson Scattering Signatures”. In: *International Workshop on BSM models in Vector Boson Scattering processes*. May 2020. arXiv: 2005.09889 [hep-ph].

[51] Morad Aaboud et al. “Observation of electroweak production of a same-sign W boson pair in association with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Phys. Rev. Lett.* 123.16 (2019), p. 161801. DOI: 10.1103/PhysRevLett.123.161801. arXiv: 1906.03203 [hep-ex].

[52] Albert M Sirunyan et al. “Observation of electroweak production of same-sign W boson pairs in the two jet and two same-sign lepton final state in proton-proton collisions at $\sqrt{s} =$ 13 TeV”. In: *Phys. Rev. Lett.* 120.8 (2018), p. 081801. DOI: 10.1103/PhysRevLett.120.081801. arXiv: 1709.05822 [hep-ex].

[53] Albert M Sirunyan et al. “Measurements of production cross sections of WZ and same-sign WW boson pairs in association with two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Lett. B* 809 (2020), p. 135710. DOI: 10.1016/j.physletb.2020.135710. arXiv: 2005.01173 [hep-ex].

[54] 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].

[55] 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.

[56] 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].

[57] 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].

[58] David L. Rainwater, R. Szalapski, and D. Zeppenfeld. “Probing color singlet exchange in Z + two jet events at the CERN LHC”. In: *Phys. Rev. D* 54 (1996), pp. 6680–6689. DOI: 10.1103/PhysRevD.54.6680. arXiv: hep-ph/9605444.

[59] Albert M Sirunyan et al. “Measurements of production cross sections of polarized same-sign W boson pairs in association with two jets in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Lett. B* 812 (2021), p. 136018. DOI: 10.1016/j.physletb.2020.136018. arXiv: 2009.09429 [hep-ex].

[60] *Study of $W^{\pm}W^{\pm}$ production via vector boson scattering at the HL-LHC with the upgraded CMS detector*. Tech. rep. Geneva: CERN, 2018. URL: http://cds.cern.ch/record/2646870.

[61] *Prospects for the measurement of the $W^{\pm}W^{\pm}$ scattering cross section and extraction of the longitudinal scattering component in pp collisions at the High-Luminosity LHC with the ATLAS experiment*. Tech. rep. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2018-052. Geneva: CERN, 2018. URL: http://cds.cern.ch/record/2652447.

[62] Joseph Polchinski. *String Theory*. Vol. 1. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 1998. DOI: 10.1017/CBO9780511816079.

[63] Djuna Croon et al. “GUT Physics in the era of the LHC”. In: *Front. in Phys.* 7 (2019), p. 76. DOI: 10.3389/fphy.2019.00076. arXiv: 1903.04977 [hep-ph].

[64] Jonathan L. Feng. “Naturalness and the Status of Supersymmetry”. In: *Ann. Rev. Nucl. Part. Sci.* 63 (2013), pp. 351–382. DOI: 10.1146/annurev-nucl-102010-130447. arXiv: 1302.6587 [hep-ph].

[65] Gabriele Ferretti. “Gauge theories of Partial Compositeness: Scenarios for Run-II of the LHC”. In: *JHEP* 06 (2016), p. 107. DOI: 10.1007/JHEP06(2016)107. arXiv: 1604.06467 [hep-ph].

[66] M. Tanabashi et al. “Review of Particle Physics”. In: *Phys. Rev. D* 98 (3 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001.

[67] Nima Arkani-Hamed, Savas Dimopoulos, and G. R. Dvali. “The Hierarchy problem and new dimensions at a millimeter”. In: *Phys. Lett. B* 429 (1998), pp. 263–272. DOI: 10.1016/S0370-2693(98)00466-3. arXiv: hep-ph/9803315.

[68] Lisa Randall and Raman Sundrum. “A Large mass hierarchy from a small extra dimension”. In: *Phys. Rev. Lett.* 83 (1999), pp. 3370–3373. DOI: 10.1103/PhysRevLett.83.3370. arXiv: hep-ph/9905221.

[69] Jan Kretzschmar. “Searches for extra dimensions with the ATLAS and CMS detectors”. In: *Nuclear and Particle Physics Proceedings* 273-275 (Apr. 2016), pp. 541–545. DOI: 10.1016/j.nuclphysbps.2015.09.080.

[70] Oliver Witzel. “Review on Composite Higgs Models”. In: *PoS* LATTICE2018 (2019), p. 006. DOI: 10.22323/1.334.0006. arXiv: 1901.08216 [hep-lat].

[71] G. Senjanovic and Rabindra N. Mohapatra. “Exact Left-Right Symmetry and Spontaneous Violation of Parity”. In: *Phys. Rev.* D12 (1975), p. 1502. DOI: 10.1103/PhysRevD.12.1502.

[72] W. Grimus. “Introduction to left-right symmetric models”. In: *Elementary particle physics. Proceedings, 4th Hellenic School, Corfu, Greece, September 2-20, 1992. 1&2*. 1993, pp. 619–632. URL: http://lss.fnal.gov/archive/other/uwthph-1993-10.pdf.

[73] Guido Altarelli, B. Mele, and M. Ruiz-Altaba. “Searching for New Heavy Vector Bosons in $p\bar{p}$ Colliders”. In: *Z. Phys. C* 45 (1989). [Erratum: Z.Phys.C 47, 676 (1990)], p. 109. DOI: 10.1007/BF01556677.

[74] David J. Muller and Satyanarayan Nandi. “Top flavor: A Separate SU(2) for the third family”. In: *Phys. Lett. B* 383 (1996), pp. 345–350. DOI: 10.1016/0370-2693(96)00745-9. arXiv: hep-ph/9602390.

[75] Yoichiro Nambu and G. Jona-Lasinio. “Dynamical model of elementary particles based on an analogy with superconductivity I.” In: *Phys. Rev.* 122 (1961). Ed. by T. Eguchi, pp. 345–358. DOI: 10.1103/PhysRev.122.345.

[76] Yoichiro Nambu and G. Jona-Lasinio. “Dynamical model of elementary particles based on an analogy with superconductivity II.” In: *Phys. Rev.* 124 (1961). Ed. by T. Eguchi, pp. 246–254. DOI: 10.1103/PhysRev.124.246.

[77] Roberta Calabrese et al. “The TopFlavor scheme in the context of W' searches at LHC”. In: (Apr. 2021). arXiv: 2104.06720 [hep-ph].

[78] Racha Cheaib. “Overview of R(D) and R(D*)”. In: *20th Conference on Flavor Physics and CP Violation*. July 2022. arXiv: 2207.14146 [hep-ex].

[79] Pouya Asadi, Matthew R. Buckley, and David Shih. “It’s all right(-handed neutrinos): a new W$^{'}$ model for the R_{D^*} anomaly”. In: *JHEP* 09 (2018), p. 010. DOI: 10.1007/JHEP09(2018)010. arXiv: 1804.04135 [hep-ph].

[80] Hsin-Chia Cheng and Ian Low. “TeV symmetry and the little hierarchy problem”. In: *JHEP* 09 (2003), p. 051. DOI: 10.1088/1126-6708/2003/09/051. arXiv: hep-ph/0308199.

[81] K. S. Babu, Xiao-Gang He, and Ernest Ma. “New Supersymmetric Left-Right Gauge Model: Higgs Boson Structure and Neutral Current Analysis”. In: *Phys. Rev. D* 36 (1987), p. 878. DOI: 10.1103/PhysRevD.36.878.

[82] Howard Georgi, Elizabeth Ellen Jenkins, and Elizabeth H. Simmons. “The Ununified Standard Model”. In: *Nucl. Phys. B* 331 (1990), pp. 541–555. DOI: 10.1016/0550-3213(90)90083-P.

[83] Xiao-yuan Li and Ernest Ma. “Gauge model of generation nonuniversality reexamined”. In: *J. Phys. G* 19 (1993), pp. 1265–1278. DOI: 10.1088/0954-3899/19/9/006. arXiv: hep-ph/9208210.

[84] Andrea Donini et al. “Phenomenological aspects of a fermiophobic SU(2) x SU(2) x U(1) extension of the standard model”. In: *Nucl. Phys. B* 507 (1997), pp. 51–90. DOI: 10.1016/S0550-3213(97)00537-3. arXiv: hep-ph/9705450.

[85] R. Sekhar Chivukula et al. “A Three Site Higgsless Model”. In: *Phys. Rev. D* 74 (2006), p. 075011. DOI: 10.1103/PhysRevD.74.075011. arXiv: hep-ph/0607124.

[86] E. Boos et al. “Interference between W' and W in single-top quark production processes”. In: *Phys. Lett. B* 655 (2007), pp. 245–250. DOI: 10.1016/j.physletb.2007.03.064. arXiv: hep-ph/0610080.

[87] Zack Sullivan. "Fully Differential W' Production and Decay at Next-to-Leading Order in QCD". In: *Phys. Rev. D* 66 (2002), p. 075011. DOI: 10.1103/PhysRevD.66.075011. arXiv: hep-ph/0207290.

[88] Albert M Sirunyan et al. "Search for high-mass resonances in final states with a lepton and missing transverse momentum at $\sqrt{s} = 13$ TeV". In: *JHEP* 06 (2018), p. 128. DOI: 10.1007/JHEP06(2018)128. arXiv: 1803.11133 [hep-ex].

[89] Georges Aad et al. "Search for a heavy charged boson in events with a charged lepton and missing transverse momentum from pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Rev. D* 100.5 (2019), p. 052013. DOI: 10.1103/PhysRevD.100.052013. arXiv: 1906.05609 [hep-ex].

[90] Morad Aaboud et al. "Search for High-Mass Resonances Decaying to $\tau\nu$ in pp Collisions at $\sqrt{s}$=13 TeV with the ATLAS Detector". In: *Phys. Rev. Lett.* 120.16 (2018), p. 161802. DOI: 10.1103/PhysRevLett.120.161802. arXiv: 1801.06992 [hep-ex].

[91] Albert M Sirunyan et al. "Search for a W' boson decaying to a τ lepton and a neutrino in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Lett. B* 792 (2019), pp. 107–131. DOI: 10.1016/j.physletb.2019.01.069. arXiv: 1807.11421 [hep-ex].

[92] Georges Aad et al. "Search for dijet resonances in events with an isolated charged lepton using $\sqrt{s} = 13$ TeV proton-proton collision data collected by the ATLAS detector". In: *JHEP* 06 (2020), p. 151. DOI: 10.1007/JHEP06(2020)151. arXiv: 2002.11325 [hep-ex].

[93] Morad Aaboud et al. "Search for new phenomena in dijet events using 37 fb^{-1} of pp collision data collected at $\sqrt{s}$ =13 TeV with the ATLAS detector". In: *Phys. Rev. D* 96.5 (2017), p. 052004. DOI: 10.1103/PhysRevD.96.052004. arXiv: 1703.09127 [hep-ex].

[94] Albert M Sirunyan et al. "Search for high mass dijet resonances with a new background prediction method in proton-proton collisions at $\sqrt{s} =$ 13 TeV". In: *JHEP* 05 (2020), p. 033. DOI: 10.1007/JHEP05(2020)033. arXiv: 1911.03947 [hep-ex].

[95] Morad Aaboud et al. "Search for heavy resonances decaying to a W or Z boson and a Higgs boson in the $q\bar{q}^{(\prime)}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Lett. B* 774 (2017), pp. 494–515. DOI: 10.1016/j.physletb.2017.09.066. arXiv: 1707.06958 [hep-ex].

[96] Morad Aaboud et al. "Search for heavy resonances decaying into a W or Z boson and a Higgs boson in final states with leptons and b-jets in 36 fb^{-1} of $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector". In: *JHEP* 03 (2018). [Erratum: JHEP 11, 051 (2018)], p. 174. DOI: 10.1007/JHEP03(2018)174. arXiv: 1712.06518 [hep-ex].

[97] Albert M. Sirunyan et al. "Search for heavy resonances decaying into a vector boson and a Higgs boson in final states with charged leptons, neutrinos and b quarks at $\sqrt{s} = 13$ TeV". In: *JHEP* 11 (2018), p. 172. DOI: 10.1007/JHEP11(2018)172. arXiv: 1807.02826 [hep-ex].

[98] Albert M Sirunyan et al. "Search for heavy resonances that decay into a vector boson and a Higgs boson in hadronic final states at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 77.9 (2017), p. 636. DOI: 10.1140/epjc/s10052-017-5192-z. arXiv: 1707.01303 [hep-ex].

[99] Morad Aaboud et al. "Search for vector-boson resonances decaying to a top quark and bottom quark in the lepton plus jets final state in pp collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector". In: *Phys. Lett. B* 788 (2019), pp. 347–370. DOI: 10.1016/j.physletb.2018.11.032. arXiv: 1807.10473 [hep-ex].

[100] Morad Aaboud et al. "Search for $W' \to tb$ decays in the hadronic final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Lett. B* 781 (2018), pp. 327–348. DOI: 10.1016/j.physletb.2018.03.036. arXiv: 1801.07893 [hep-ex].

[101] Albert M Sirunyan et al. "Search for W' bosons decaying to a top and a bottom quark at $\sqrt{s}$ =13 TeV in the hadronic final state". In: (Apr. 2021). arXiv: 2104.04831 [hep-ex].

[102] Albert M Sirunyan et al. "Search for heavy resonances decaying to a top quark and a bottom quark in the lepton+jets final state in proton–proton collisions at 13 TeV". In: *Phys. Lett. B* 777 (2018), pp. 39–63. DOI: 10.1016/j.physletb.2017.12.006. arXiv: 1708.08539 [hep-ex].

[103] Steven Weinberg. "Phenomenological Lagrangians". In: *Physica A* 96.1-2 (1979). Ed. by S. Deser, pp. 327–340. DOI: 10.1016/0378-4371(79)90223-1.

[104] Ashok Das and Jnanadeva Maharana. "Phenomenological implications of S-duality symmetry". In: *Phys. Lett. B* 699 (2011), pp. 264–270. DOI: 10.1016/j.physletb.2011.04.005. arXiv: 1103.4247 [hep-ph].

[105] Steven Weinberg. "Baryon- and Lepton-Nonconserving Processes". In: *Phys. Rev. Lett.* 43 (21 1979), pp. 1566–1570. DOI: 10.1103/PhysRevLett.43.1566. URL: https://link.aps.org/doi/10.1103/PhysRevLett.43.1566.

[106] Chung Ngoc Leung, S. T. Love, and S. Rao. "Low-Energy Manifestations of a New Interaction Scale: Operator Analysis". In: *Z. Phys. C* 31 (1986), p. 433. DOI: 10.1007/BF01588041.

[107] W. Buchmuller and D. Wyler. "Effective Lagrangian Analysis of New Interactions and Flavor Conservation". In: *Nucl. Phys. B* 268 (1986), pp. 621–653. DOI: 10.1016/0550-3213(86)90262-2.

[108] Eduardo da Silva Almeida, O. J. P. Éboli, and M. C. Gonzalez–Garcia. "Unitarity constraints on anomalous quartic couplings". In: *Phys. Rev. D* 101.11 (2020), p. 113003. DOI: 10.1103/PhysRevD.101.113003. arXiv: 2004.05174 [hep-ph].

[109] H. Georgi. "Effective field theory". In: *Ann. Rev. Nucl. Part. Sci.* 43 (1993), pp. 209–252. DOI: 10.1146/annurev.ns.43.120193.001233.

[110] Riccardo Bellan et al. "A sensitivity study of VBS and diboson WW to dimension-6 EFT operators at the LHC". In: *JHEP* 05 (2022), p. 039. DOI: 10.1007/JHEP05(2022)039. arXiv: 2108.03199 [hep-ph].

[111] C. Arzt, M. B. Einhorn, and J. Wudka. "Patterns of deviation from the standard model". In: *Nucl. Phys. B* 433 (1995), pp. 41–66. DOI: 10.1016/0550-3213(94)00336-D. arXiv: hep-ph/9405214.

[112] O. J. P. Éboli and M. C. Gonzalez-Garcia. "Classifying the bosonic quartic couplings". In: *Phys. Rev. D* 93.9 (2016), p. 093013. DOI: 10.1103/PhysRevD.93.093013. arXiv: 1604.03555 [hep-ph].

[113] Jan Kalinowski et al. "Same-sign WW scattering at the LHC: can we discover BSM effects before discovering new states?" In: *Eur. Phys. J. C* 78.5 (2018), p. 403. DOI: 10.1140/epjc/s10052-018-5885-y. arXiv: 1802.02366 [hep-ph].

[114] Celine Degrande et al. "Effective Field Theory: A Modern Approach to Anomalous Couplings". In: *Annals Phys.* 335 (2013), pp. 21–32. DOI: 10.1016/j.aop.2013.04.016. arXiv: 1205.4231 [hep-ph].

[115] CMS Collaboration. “Search for anomalous electroweak production of vector boson pairs in association with two jets in proton-proton collisions at 13 TeV”. In: *Physics Letters B* 798 (2019), p. 134985. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2019.134985. URL: https://www.sciencedirect.com/science/article/pii/S0370269319307075.

[116] CMS Collaboration. “Measurements of production cross sections of polarized same-sign W boson pairs in association with two jets in proton-proton collisions at s=13 TeV”. In: *Physics Letters B* 812 (2021), p. 136018. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2020.136018. URL: https://www.sciencedirect.com/science/article/pii/S0370269320308212.

[117] Hemdon, M. and others. *Limits on anomalous triple and quartic gauge couplings*. Twiki. 2020. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSMPaTGC.

[118] Georges Aad et al. “Differential cross-section measurements for the electroweak production of dijets in association with a Z boson in proton–proton collisions at ATLAS”. In: *Eur. Phys. J. C* 81.2 (2021), p. 163. DOI: 10.1140/epjc/s10052-020-08734-w. arXiv: 2006.15458 [hep-ex].

[119] K. Arnold et al. “VBFNLO: A Parton level Monte Carlo for processes with electroweak bosons”. In: *Comput. Phys. Commun.* 180 (2009), pp. 1661–1670. DOI: 10.1016/j.cpc.2009.03.006. arXiv: 0811.4559 [hep-ph].

[120] Raquel Gomez-Ambrosio. “Studies of Dimension-Six EFT effects in Vector Boson Scattering”. In: *Eur. Phys. J. C* 79.5 (2019), p. 389. DOI: 10.1140/epjc/s10052-019-6893-2. arXiv: 1809.04189 [hep-ph].

[121] Athanasios Dedes, Paweł Kozów, and Michał Szleper. “Standard model EFT effects in vector-boson scattering at the LHC”. In: *Phys. Rev. D* 104.1 (2021), p. 013003. DOI: 10.1103/PhysRevD.104.013003. arXiv: 2011.07367 [hep-ph].

[122] Jacob J. Ethier et al. “SMEFT analysis of vector boson scattering and diboson data from the LHC Run II”. In: *Eur. Phys. J. C* 81.6 (2021), p. 560. DOI: 10.1140/epjc/s10052-021-09347-7. arXiv: 2101.03180 [hep-ph].

[123] 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. URL: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08001.

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

[125] G. Aad et al. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

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

[127] 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.

[128] “High-Luminosity Large Hadron Collider (HL-LHC): Technical Design Report V. 0.1”. In: 4/2017 (2017). Ed. by G. Apollinari et al. DOI: 10.23731/CYRM-2017-004.

[129] V Karimäki et al. *The CMS tracker system project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/368412.

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

[131] Serguei Chatrchyan et al. “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].

[132] R. Fruhwirth. “Application of Kalman filtering to track and vertex fitting”. In: *Nucl. Instrum. Meth. A* 262 (1987), pp. 444–450. DOI: 10.1016/0168-9002(87)90887-4.

[133] *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/349375.

[134] *The CMS hadron calorimeter project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/357153.

[135] *The CMS magnet project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: http://cds.cern.ch/record/331056.

[136] Giovanni Abbiendi. “The CMS muon system in Run2: preparation, status and first results”. In: *PoS* EPS-HEP2015 (2015). Comments: 8 pages, 13 figures, to be published in the proceedings of EPS-HEP 2015, p. 237. DOI: 10.22323/1.234.0237. arXiv: 1510.05424. URL: https://cds.cern.ch/record/2061199.

[137] A Tapper and Darin Acosta. *CMS Technical Design Report for the Level-1 Trigger Upgrade*. Tech. rep. CERN-LHCC-2013-011. CMS-TDR-12. June 2013. URL: http://cds.cern.ch/record/1556311.

[138] Sergio Cittolin, Attila Rácz, and Paris Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical design report. CMS. Geneva: CERN, 2002. URL: http://cds.cern.ch/record/578006.

[139] A. M. Sirunyan et al. “Particle-flow reconstruction and global event description with the CMS detector”. In: *JINST* 12.10 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[140] Rudolf Fruehwirth et al. *Data analysis techniques for high-energy physics; 2nd ed.* Cambridge monographs on particle physics, nuclear physics, and cosmology. Cambridge: Cambridge Univ. Press, 2000. URL: https://cds.cern.ch/record/429722.

[141] K. Rose. “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2210–2239. DOI: 10.1109/5.726788.

[142] R Fruehwirth, Wolfgang Waltenberger, and Pascal Vanlaer. *Adaptive Vertex Fitting*. Tech. rep. Geneva: CERN, 2007. URL: https://cds.cern.ch/record/1027031.

[143] “Muon tracking performance in the CMS Run-2 Legacy data using the tag-and-probe technique”. In: (2020). URL: https://cds.cern.ch/record/2724492.

[144] A. M. Sirunyan et al. “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *JINST* 13.06 (2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det].

[145] W. Adam et al. “Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC”. In: *eConf* C0303241 (2003), TULT009. DOI: 10.1088/0954-3899/31/9/N01. arXiv: physics/0306087.

[146] Albert M Sirunyan et al. “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”. In: *JINST* 16.05 (2021), P05014. DOI: 10.1088/1748-0221/16/05/P05014. arXiv: 2012.06888 [hep-ex].

[147] Gavin P. Salam. "Towards Jetography". In: *Eur. Phys. J. C* 67 (2010), pp. 637–686. DOI: 10.1140/epjc/s10052-010-1314-6. arXiv: 0906.1833 [hep-ph].

[148] 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].

[149] 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].

[150] *Pileup Removal Algorithms*. Tech. rep. CMS-PAS-JME-14-001. Geneva: CERN, 2014. URL: http://cds.cern.ch/record/1751454.

[151] Daniele Bertolini et al. "Pileup Per Particle Identification". In: *JHEP* 10 (2014), p. 059. DOI: 10.1007/JHEP10(2014)059. arXiv: 1407.6013 [hep-ph].

[152] *Jet algorithms performance in 13 TeV data*. Tech. rep. Geneva: CERN, 2017. URL: http://cds.cern.ch/record/2256875.

[153] "Performance of the pile up jet identification in CMS for Run 2". In: (2020). URL: https://cds.cern.ch/record/2715906.

[154] Vardan Khachatryan et al. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *JINST* 12.02 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].

[155] *Identification of b quark jets at the CMS Experiment in the LHC Run 2*. Tech. rep. CMS-PAS-BTV-15-001. Geneva: CERN, 2016. URL: http://cds.cern.ch/record/2138504.

[156] A. M. Sirunyan et al. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *JINST* 13.05 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det].

[157] Emil Bols et al. "Jet Flavour Classification Using DeepJet". In: *JINST* 15.12 (2020), P12012. DOI: 10.1088/1748-0221/15/12/P12012. arXiv: 2008.10519 [hep-ex].

[158] A. M. Sirunyan et al. "Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV". In: *JINST* 13.10 (2018), P10005. DOI: 10.1088/1748-0221/13/10/P10005. arXiv: 1809.02816 [hep-ex].

[159] "Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector". In: (Aug. 2016).

[160] Frederick James and Lorenzo Moneta. "Review of High-Quality Random Number Generators". In: *Comput. Softw. Big Sci.* 4.1 (2020), p. 2. DOI: 10.1007/s41781-019-0034-3. arXiv: 1903.01247 [physics.comp-ph].

[161] George Marsaglia, B. Narasimhan, and Arif Zaman. "A random number generator for PC's". In: *Computer Physics Communications* 60.3 (1990), pp. 345–349. ISSN: 0010-4655. DOI: https://doi.org/10.1016/0010-4655(90)90033-W. URL: https://www.sciencedirect.com/science/article/pii/001046559090033W.

[162] Bo Andersson, G. Gustafson, and B. Soderberg. "A General Model for Jet Fragmentation". In: *Z. Phys. C* 20 (1983), p. 317. DOI: 10.1007/BF01407824.

[163] Torbjorn Sjostrand. "Jet Fragmentation of Nearby Partons". In: *Nucl. Phys. B* 248 (1984), pp. 469–502. DOI: 10.1016/0550-3213(84)90607-2.

[164] Christian Bierlich et al. "A comprehensive guide to the physics and usage of PYTHIA 8.3". In: (Mar. 2022). DOI: 10.21468/SciPostPhysCodeb.8. arXiv: 2203.11601 [hep-ph].

[165] T. Stelzer and W. F. Long. "Automatic generation of tree level helicity amplitudes". In: *Comput. Phys. Commun.* 81 (1994), pp. 357–371. DOI: 10.1016/0010-4655(94)90084-1. arXiv: hep-ph/9401258.

[166] Fabio Maltoni and Tim Stelzer. "MadEvent: Automatic event generation with MadGraph". In: *JHEP* 02 (2003), p. 027. DOI: 10.1088/1126-6708/2003/02/027. arXiv: hep-ph/0208156.

[167] Kaoru Hagiwara and D. Zeppenfeld. "Helicity Amplitudes for Heavy Lepton Production in e+ e- Annihilation". In: *Nucl. Phys. B* 274 (1986), pp. 1–32. DOI: 10.1016/0550-3213(86)90615-2.

[168] F. Maltoni et al. "Color Flow Decomposition of QCD Amplitudes". In: *Phys. Rev. D* 67 (2003), p. 014026. DOI: 10.1103/PhysRevD.67.014026. arXiv: hep-ph/0209271.

[169] Yuri L. Dokshitzer. "Calculation of the structure functions for deep inelastic scattering and e+ e- annihilation by perturbation theory in quantum chromodynamics." In: *Sov. Phys. JETP* 46 (1977), pp. 641–653.

[170] Guido Altarelli and G. Parisi. "Asymptotic Freedom in Parton Language". In: *Nucl. Phys. B* 126 (1977), pp. 298–318. DOI: 10.1016/0550-3213(77)90384-4.

[171] Silvia Ferreres-Solé and Torbjörn Sjöstrand. "The space–time structure of hadronization in the Lund model". In: *Eur. Phys. J. C* 78.11 (2018), p. 983. DOI: 10.1140/epjc/s10052-018-6459-8. arXiv: 1808.04619 [hep-ph].

[172] Daniel Duffty and Zack Sullivan. "Model independent reach for W-prime bosons at the LHC". In: *Phys. Rev. D* 86 (2012), p. 075018. DOI: 10.1103/PhysRevD.86.075018. arXiv: 1208.4858 [hep-ph].

[173] Pierre Artoisenet et al. "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations". In: *JHEP* 03 (2013), p. 015. DOI: 10.1007/JHEP03(2013)015. arXiv: 1212.3460 [hep-ph].

[174] Natascia Vignaroli. "New W' signals at the LHC". In: *Phys. Rev. D* 89.9 (2014), p. 095027. DOI: 10.1103/PhysRevD.89.095027. arXiv: 1404.5558 [hep-ph].

[175] Daniele Barducci and Cédric Delaunay. "Bounding wide composite vector resonances at the LHC". In: *JHEP* 02 (2016), p. 055. DOI: 10.1007/JHEP02(2016)055. arXiv: 1511.01101 [hep-ph].

[176] Albert M Sirunyan et al. "Search for resonant $t\bar{t}$ production in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 04 (2019), p. 031. DOI: 10.1007/JHEP04(2019)031. arXiv: 1810.05905 [hep-ex].

[177] Serguei Chatrchyan et al. "Search for a W' boson decaying to a bottom quark and a top quark in pp collisions at $\sqrt{s} = 7$ TeV". In: *Phys. Lett. B* 718 (2013), p. 1229. DOI: 10.1016/j.physletb.2012.12.008. arXiv: 1208.0956 [hep-ex].

[178] Albert M Sirunyan et al. "Search for heavy resonances decaying to a top quark and a bottom quark in the lepton+jets final state in proton–proton collisions at 13 TeV". In: *Phys. Lett. B* 777 (2018), pp. 39–63. DOI: 10.1016/j.physletb.2017.12.006. arXiv: 1708.08539 [hep-ex].

[179] Albert M Sirunyan et al. "Search for W' bosons decaying to a top and a bottom quark at s=13TeV in the hadronic final state". In: *Phys. Lett. B* 820 (2021), p. 136535. DOI: 10.1016/j.physletb.2021.136535. arXiv: 2104.04831 [hep-ex].

[180] Ilaria Brivio, Yun Jiang, and Michael Trott. "The SMEFTsim package, theory and tools". In: *JHEP* 12 (2017), p. 070. DOI: 10.1007/JHEP12(2017)070. arXiv: 1709.06492 [hep-ph].

[181] 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].

[182] Eboli, O.J.P. and Gonzalez-Garcia, M.C. *Anomalous Gauge Coupling*. Twiki. 2021. URL: https://feynrules.irmp.ucl.ac.be/wiki/AnomalousGaugeCoupling.

[183] Michal Czakon et al. "Top-pair production at the LHC through NNLO QCD and NLO EW". In: *JHEP* 10 (2017), p. 186. DOI: 10.1007/JHEP10(2017)186. arXiv: 1705.04105 [hep-ph].

[184] A. M. Sirunyan et al. "Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS". In: *Eur. Phys. J. C* 81 (2021), p. 800. DOI: 10.1140/epjc/s10052-021-09538-2. arXiv: 2104.01927 [hep-ex].

[185] 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: http://cds.cern.ch/record/2621960.

[186] 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: http://cds.cern.ch/record/2676164.

[187] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *Journal of Physics G: Nuclear and Particle Physics* 43.2 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. URL: https://doi.org/10.1088/0954-3899/43/2/023001.

[188] Roger J. Barlow and Christine Beeston. "Fitting using finite Monte Carlo samples". In: *Comput. Phys. Commun.* 77 (1993), pp. 219–228. DOI: 10.1016/0010-4655(93)90005-W.

[189] J. S. Conway. "Incorporating Nuisance Parameters in Likelihoods for Multisource Spectra". In: *PHYSTAT 2011*. 2011, pp. 115–120. DOI: 10.5170/CERN-2011-006.115. arXiv: 1103.0354 [physics.data-an].

[190] Rikkert Frederix, Emanuele Re, and Paolo Torrielli. "Single-top t-channel hadroproduction in the four-flavour scheme with POWHEG and aMC@NLO". In: *JHEP* 09 (2012), p. 130. DOI: 10.1007/JHEP09(2012)130. arXiv: 1207.5391 [hep-ph].

[191] Simone Alioli et al. "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions". In: *JHEP* 09 (2009). [Erratum: JHEP 02, 011 (2010)], p. 111. DOI: 10.1088/1126-6708/2009/09/111. arXiv: 0907.4076 [hep-ph].

[192] Stefano Frixione, Paolo Nason, and Giovanni Ridolfi. "A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction". In: *JHEP* 09 (2007), p. 126. DOI: 10.1088/1126-6708/2007/09/126. arXiv: 0707.3088 [hep-ph].

[193] Emanuele Re. "Single-top Wt-channel production matched with parton showers using the POWHEG method". In: *Eur. Phys. J. C* 71 (2011), p. 1547. DOI: 10.1140/epjc/s10052-011-1547-z. arXiv: 1009.2450 [hep-ph].

[194] 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].

[195] 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].

[196] *Investigations of the impact of the parton shower tuning in Pythia 8 in the modelling of* $t\bar{t}$ *at* $\sqrt{s} = 8$ *and 13 TeV*. Tech. rep. CMS-PAS-TOP-16-021. Geneva: CERN, 2016. URL: http://cds.cern.ch/record/2235192.

[197] 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].

[198] Albert M Sirunyan et al. "Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements". In: *Eur. Phys. J. C* 80.1 (2020), p. 4. DOI: 10.1140/epjc/s10052-019-7499-4. arXiv: 1903.12179 [hep-ex].

[199] Peter Skands, Stefano Carrazza, and Juan Rojo. "Tuning PYTHIA 8.1: the Monash 2013 Tune". In: *Eur. Phys. J. C* 74.8 (2014), p. 3024. DOI: 10.1140/epjc/s10052-014-3024-y. arXiv: 1404.5630 [hep-ph].

[200] Richard D. Ball et al. "Reweighting and Unweighting of Parton Distributions and the LHC W lepton asymmetry data". In: *Nucl. Phys. B* 855 (2012), pp. 608–638. DOI: 10.1016/j.nuclphysb.2011.10.018. arXiv: 1108.1758 [hep-ph].

[201] 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.

[202] S. Kallweit et al. "NLO QCD+EW predictions for V + jets including off-shell vector-boson decays and multijet merging". In: *Journal of High Energy Physics* 2016.4 (2016), pp. 1–51. DOI: 10.1007/jhep04(2016)021. URL: https://doi.org/10.1007\%2Fjhep04\%282016\%29021.

[203] CMS Collaboration. *HiggsAnalysis-CombinedLimit*. https://cms-analysis.github.io/HiggsAnalysis-CombinedLimit/. 2021.

[204] Robert D. Cousins. "Generalization of Chisquare Goodness-ofFit Test for Binned Data Using Saturated Models , with Application to Histograms". In: 2013.

[205] Thomas Junk. "Confidence level computation for combining searches with small statistics". In: *Nucl. Instrum. Meth. A* 434 (1999), pp. 435–443. DOI: 10.1016/S0168-9002(99)00498-2. arXiv: hep-ex/9902006.

[206] Alexander L. Read. "Presentation of search results: The CL(s) technique". In: *J. Phys. G* 28 (2002). Ed. by M. R. Whalley and L. Lyons, pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313.

[207] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *Eur. Phys. J. C* 71 (2011). [Erratum: Eur.Phys.J.C 73, 2501 (2013)], p. 1554. DOI: 10.1140/epjc/s10052-011-1554-0. arXiv: 1007.1727 [physics.data-an].

[208] Eilam Gross and Ofer Vitells. "Trial factors for the look elsewhere effect in high energy physics". In: *Eur. Phys. J. C* 70 (2010), pp. 525–530. DOI: 10.1140/epjc/s10052-010-1470-8. arXiv: 1005.1891 [physics.data-an].

[209] Claudia Garcia-Garcia, Maria Herrero, and Roberto A. Morales. "Unitarization effects in EFT predictions of WZ scattering at the LHC". In: *Phys. Rev. D* 100.9 (2019), p. 096003. DOI: 10.1103/PhysRevD.100.096003. arXiv: 1907.06668 [hep-ph].

[210] John M. Campbell, R. Keith Ellis, and Ciaran Williams. "Vector boson pair production at the LHC". In: *JHEP* 07 (2011), p. 018. DOI: 10.1007/JHEP07(2011)018. arXiv: 1105.0020 [hep-ph].

[211] Richard D. Ball et al. "Parton distributions from high-precision collider data". In: *Eur. Phys. J. C* 77.10 (2017), p. 663. DOI: 10.1140/epjc/s10052-017-5199-5. arXiv: 1706.00428 [hep-ph].

[212] Albert M Sirunyan et al. “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements”. In: *Eur. Phys. J. C* 80.1 (2020), p. 4. DOI: 10.1140/epjc/s10052-019-7499-4. arXiv: 1903.12179 [hep-ex].

[213] Benedikt Biedermann, Ansgar Denner, and Mathieu Pellen. “Large Electroweak Corrections to Vector-Boson Scattering at the Large Hadron Collider”. In: *Phys. Rev. Lett.* 118 (26 2017), p. 261801. DOI: 10.1103/PhysRevLett.118.261801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.118.261801.

[214] Benedikt Biedermann, Ansgar Denner, and Mathieu Pellen. “Complete NLO corrections to W^+W^+ scattering and its irreducible background at the LHC”. In: *Journal of High Energy Physics* 2017.10 (2017). DOI: 10.1007/jhep10(2017)124. URL: https://doi.org/10.1007\%2Fjhep10\%282017\%29124.

[215] A. J. Barr et al. “Guide to transverse projections and mass-constraining variables”. In: *Phys. Rev. D* 84 (9 2011), p. 095031.

[216] Serguei Chatrchyan et al. “Measurement of Higgs Boson Production and Properties in the WW Decay Channel with Leptonic Final States”. In: *JHEP* 01 (2014), p. 096. DOI: 10.1007/JHEP01(2014)096. arXiv: 1312.1129 [hep-ex].

[217] François Chollet. *keras*. https://github.com/fchollet/keras. 2015.

[218] *Scikit-Optimize*. URL: https://scikit-optimize.github.io/stable/.

[219] *SHAP*. URL: https://shap.readthedocs.io/en/latest/index.html.

[220] The CMS Collaboration. “Measurements of production cross sections of WZ and same-sign WW boson pairs in association with two jets in proton-proton collisions at s=13TeV”. In: *Physics Letters B* 809 (2020), p. 135710. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2020.135710. URL: https://www.sciencedirect.com/science/article/pii/S037026932030513X.

[221] Jianqing Fan, Chunming Zhang, and Jian Zhang. “Generalized Likelihood Ratio Statistics and Wilks Phenomenon”. In: *The Annals of Statistics* 29.1 (2001), pp. 153–193. ISSN: 00905364. URL: http://www.jstor.org/stable/2674021 (visited on 01/14/2023).

Acknowledgements

And this thesis ends my career as an university student. It has been a long journey, during which I have met many people, from professors to colleagues to young researchers.

I would like to thank, among others, Orso, who introduced me to the world of experimental particle physics as a supervisor of Master's and Doctoral degrees. To me, he is a mentor in many aspects of my education, willing to provide his experience and expertise to untangle a very multifaceted and complex environment, as well as friendly and professional. I could hardly have found a better reference person to start my career as a researcher. I would like to thank Livio, who provided me with the opportunity to explore an area unknown to me, to travel around Europe for my activities, and to meet new people and environments different from those in which I grew up. My heartfelt thanks also go to Prof. Carlotti, who was able to skillfully coordinate the activities of my doctoral cycle despite the great challenges posed by the pandemic in just the last three years. Obrigado to Michele, who introduced me to tau physics and put his long-standing experience at the disposal of the successful VBS analysis, believing in it even in the most complex moments. A thought also goes to Alessandro, who introduced me to the Tracker DQM, a dark world indispensable for research activities at CMS.

Passing to people closer to my own age, I feel particularly grateful to Agostino, who was the first to welcome me into a CMS lab and provide me with the means to unravel the world of data analysis for a high-energy physics experiment, to Matteo (the elder) and Valentina, who helped and guided me in facing the challenges posed by a PhD developed in the context of a very large experiment; credit for the results achieved in these years goes also to all of them.

My doctoral journey was not isolated, but shared with several colleagues from the Italian community at CERN, even when the pandemic made it complicated to share this experience more closely. I cannot but thank Tommaso, with whom I shared both the admission competition interview and the discussion of the doctoral thesis, as well as the whole "PhD in CMS" package. His support was essential both for the shared research activities and for orienting in the Perugia environment. Among others, special thoughts go to Matteo (the younger) for the moments shared between Perugia and Geneva, to Luca and Renato who, after the adventures they shared for the laboratory courses, helped me maintain connections with my home community, and to Tommaso in Bologna for providing me with support in the CMS jungle of MC simulations and sharing my first live Formula 1 experience.

There have been several people in this journey, outside the academia, whose prerogatives and talents have supported and helped to further my efforts.

First of all, immeasurable thanks go to Erika, who shared the doctoral period with me very closely, as well as the pandemic period. Her support and closeness in the most complicated moments, as well as in the most exciting ones, is and has been essential to preserve me from getting hopelessly discouraged and anxious, reminding me whenever necessary why I had embarked on this path and what my goals and desires were.

Together we have learned what it means to build a relationship based on living together, and I will always strive to provide her with the same support and love gifted by her.

I also want to thank my family in Naples, who always gave me the right freedom to take the path I felt was most mine, without having to feel subjected to choosing more by following social dynamics than my own inclinations and inspirations. In particular, I thank my brother Simone, whose way of acting and thinking has always helped me not to fossilize on my mental structures, and with whom I share several interests. And I thank my parents Eliana and Pino, who taught me to be open-minded and rational, not forgetting the importance of cultivating one's own identity, preserving one's origins, not putting social conventions before the needs and desires of human beings, being always respectful and empathetic toward other people's life experiences.

In addition, I would like to thank all the relatives and friends who have stood by me over the years with their kind affection.

At this particular moment, a thought goes out to all my peers whom I have crossed paths with during my time at university, both in Naples and Perugia and on social networks, people with whom I share and have shared my desire to find in the University the most suitable training to pursue one's goals and professional fulfillment. Unfortunately, I have learned how the university environment, rather than providing the means to grow and form oneself respecting the individualities, talents, and needs of each one, too often represents an ultra-competitive, debasing, and disrespectful environment towards the people who spend and devote most of their lives and time there, giving more importance to performing than educating. Several of my peers, surrounded also by social and family contexts that reinforce certain perverse and destructive mentalities, begin to badly suffer, without having the possibility and means to heal, even going in the end to extreme gestures. We cannot be silent accomplices to such a system, and that is why I would like to invite anyone reading these lines to pay more attention to certain dynamics, not to go along with them, and to make their own contribution to not leave those who suffer and are in extreme difficulty isolated, and to promote a more serene and purposeful social approach to university and psychotherapy. I feel extremely fortunate to have found myself in social contexts and networks that would help me face certain challenges in academia with the right serenity, because too many people have not had this sort of privilege.

In the very end, I thank INFN and CERN for giving me the material opportunity to conduct this PhD and the successful research activities I have presented in this thesis.

And I thank me, myself and I.

Kindly,